PhD thesis

Biochemical and structural studies of the cellular mRNA export factor TAP, and its interaction with the viral CTE-RNA

ERIKA LIKER

Supervisor: Dr. Elena Conti

European Molecular Biology Laboratory
Heidelberg, Germany

2002

TABLE OF CONTENTS

ABBREVIATIONS

CTE	Constitutive Transport Element
EJC	Exon-exon Junction Complex
GST	Glutathione S-Transferase
LRR	Leucine Rich Repeat
MAD	Multiple Anomalous Dispersion
MIR	Multiple Isomorphus Replacement
MR	Molecular Replacement
NES	Nuclear Export Signal
NLS	Nuclear Localization Signal
NPC	Nuclear Pore Complex
RNP	Ribonucleoprotein
TAP	TIP Associating Protein
UBA	Ubiquitin-associated

1. INTRODUCTION

The compartmentalization of eukaryotic cells gives rise to a need for the intracellular transport of macromolecules. Specialized systems have evolved that allow proteins to be imported into membrane-bound organelles such as mitochondria, chloroplasts, lysosomes, the endoplasmic reticulum (ER), and nuclei. Nuclear transport is unusual in that it is bi-directional, with both import into the organelle and export out of it are major processes. All nuclear proteins are made in the cytoplasm and must be imported into the nucleus. In cells with an open mitosis these proteins must be reimported after each nuclear division. On the other hand, cellular RNAs are transcribed in the nucleus and have to be exported to the cytoplasm for protein synthesis. Many macromolecules shuttle continuously between the nucleus and the cytoplasm. In total, this gives rise to an enormous level of nucleocytoplasmic traffic (Görlich and Mattaj, 1996).

The nucleus is surrounded by the nuclear envelope (NE), a double membrane that is continuous with the ER. There are large assemblies termed nuclear pore complexes (NPCs), which form channels in the NE. These NPCs are the sole sites of exchange between nucleus and cytoplasm during interphase (Feldherr et al., 1984). The NPCs are huge structures of about 125 million Daltons in vertebrates through which all nucleocytoplasmic transport is thought to occur. NPCs are composed of between 50 and 100 distinct polypeptides (Rout et al., 2000) that are generically called nucleoporins. NPCs are characterized by an eightfold symmetry. With the exception of the largely membrane-imbedded central core structure, NPCs also contain cytoplasmic and nuclear extensions that form cytoplasmic filaments and nuclear basket, respectively. Molecules of up to approximately 9 nm in diameter, corresponding to a globular protein of approximately 60 kDa, can in principle enter or leave the nucleus by diffusion through the NPC, although in practice very few proteins and no

known RNAs do so. Rather, nucleocytoplasmic transport is an active, signal-mediated process.

Active transport between the nucleus and the cytoplasm involves primarily four classes of macromolecules: substrates (cargoes), adaptors, regulators and receptors. Cargoes are recognized by the nucleocytoplasmic transport machinery via signals (nuclear localization signals, NLSs, and nuclear export signals, NESs). The signals in many protein cargoes and in some RNP cargoes are recognized by one or more members of the nuclear transport receptor family. Members of this protein family have been given various names, including karyopherins, importins, transportins, exportins, etc, (Mattaj and Englmeier, 1998). These receptors are generally large (90-130 kDa) acidic proteins sharing 15-25% sequence identity. They all have an N-terminal RanGTP-binding domain, a C-terminal cargo-binding domain, and the capacity to bind components of the NPC.

Some transport substrates bind directly to an import or an export receptor, while others require one or more adaptors to mediate formation of a receptor-substrate complex. In the simplest case, the signal is bound by a transport receptor on one side of the NPC, translocated through the NPC via receptor-mediated interactions, and released on the other side. The empty receptor is then recycled to the original compartment and restored to a form competent for signal binding. A common and slightly more complicated variant of this scheme involves the use of adaptors. In these cases, substrate-receptor interactions are not direct but are mediated by one or more adaptor proteins. Here not only the receptor but also the adaptor needs to be recycled after transport of the substrate and therefore also shuttles between the nuclear and cytoplasmic compartments.

A characteristic of nucleocytoplasmic transport is directionality. Substrates are moved either into the nucleus or out of it. Ran, a small GTPase, is thought to be critical for this directionality. Like other GTPases, Ran needs regulators for its activity. The regulators either

stimulate Ran to hydrolyze GTP (RanGAP) (Becker et al., 1995; Bischoff et al., 1994) or to release the resultant GDP and rebind GTP (RanGEF) (Klebe et al., 1995). Ran's regulators are distributed asymmetrically across the NE such that GTPase activity is favored in the cytoplasm and GDP/GTP exchange in the nucleus. This predicts a high RanGTP concentration in the nucleus and a low cytoplasmic one. Both import receptors and export receptors bind to RanGTP and exit the nucleus in association with it. However, RanGTP regulates cargo loading and unloading differently in these two classes of receptors. Export receptors bind their cargo as part of a RanGTP-containing complex within the nucleus. These receptors dissociate from the cargo after GTP hydrolysis in the cytosol. By contrast, import receptors associate with their cargo in cytosolic complexes that do not contain RanGTP, and these complexes dissociate upon Ran-GTP binding in the nucleus (Fig.1).

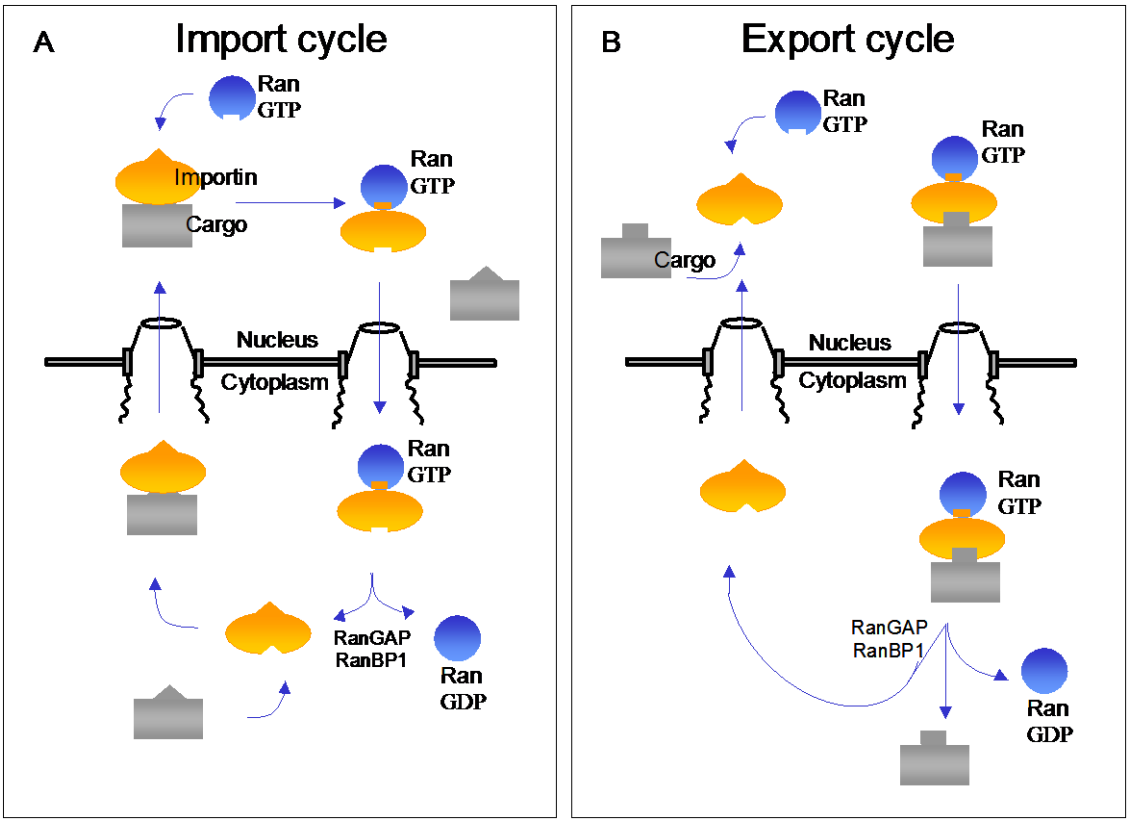


Fig. 1
Nucleocytoplasmic transport. (**A**) Import cycle (**B**) Export cycle.

1.1 NUCLEAR IMPORT

Key issues in nuclear import include substrate recognition, targeting of the transport complex to the NPC, vectorial movement through the NPC and substrate dissociation in the nucleus. The direction of transport through the NPC is determined by a signal. The nuclear localization signal (NLS) directs proteins into the nucleus. While targeting sequences for mitochondrial or ER import are cleaved after the transit has occurred, nuclear targeting signals are not cleaved and they are permanent part of the mature protein.

Importin α is an adaptor protein that recognizes the first discovered or classical NLS, which is characterized by one or two stretches of basic residues (Pemberton et al., 1998). Importin α also interacts with importin β and together they import proteins containing the classical NLS. Over the past several years, many other import pathways have been identified, all involving transport substrates with NLSs distinct from the classical sequence (Pemberton et al., 1998). Non-classical NLSs have diverse amino acid sequences that bind directly and specifically to the different importin β homologues that constitute the importin β family. Unlike the classical NLS, which interacts with importin β indirectly through an adaptor protein, almost all other NLSs bind directly to their specific importin βs.

After a complex being formed between the NLS of a protein and importin β (directly or through an adaptor) the β component docks the complex to distinct sites along the fibrils of the nuclear pore complex (NPC). After translocation at the nucleoplasmic side β releases its cargo by interacting with the small GTPase Ran in its GTP-bound form. The NLS-containing protein diffuses into the nucleus to exert its function while the soluble components of the transport machinery are recycled back to the cytoplasm for a new import cycle (Fig.1).

1.2 RNA EXPORT

RNA export is the process by which RNAs are transported to the cytoplasm after synthesis, processing, and RNP assembly within the nucleus. Specific factors mediate the nuclear export of different classes of RNAs. As originally shown by microinjection experiments in *Xenopus* oocytes, uracyl-rich small nuclear RNA (U snRNA), transfer RNA (tRNA), messenger RNA (mRNA) and ribosomal RNA (rRNA) do not compete for export, suggesting they access distinct pathways (Jarmolowski et al., 1994). The quest for the identification of the molecular mechanisms underlying cellular RNA nuclear export pathways is still in progress, but several specific factors have been recently identified. The transport of tRNAs and U snRNAs is mediated by two members of the exportin/karyopherin β receptor family, namely Xpo-t and Crm1 (Kutay et al., 1998; Arts et al., 1998; Fornerod et al., 1997; Stade et al., 1997; Fukuda et al., 1997). This family of transport factors includes receptors that have distinct binding specificity but share a similar mechanistic and structural framework (Görlich, 1997; Mattaj and Conti, 1999, and references therein). Exportins recognize a specific export signal on the macromolecule to be translocated and associate with it in a ternary complex with RanGTP. Upon reaching the cytoplasm, nucleotide hydrolysis promotes the dissociation of the ternary complex and release of the cargo. In the case of tRNA nuclear export, the karyopherin β receptor Xpo-t binds the nucleic acid cargo by recognizing particular structural features common to mature tRNAs (Arts et al., 1998). In the case of spliceosomal U snRNA export, the interaction with the Crm1 transport factor is not direct but mediated by an adapter, CBC, the cap binding complex that recognizes the cap structure characteristic of U snRNAs mature for export (Izaurralde et al., 1995; Fornerod et al., 1997). Unexpectedly in this context, the 5' m^7G cap structure that is crucial for U snRNA export is not essential for mRNA export. Moreover, genetic screens for mRNA export factors have not identified proteins related to the karyopherin β protein family.

1.3 CELLULAR mRNA EXPORT

The export of mRNA is probably the most complex and less characterized RNA export pathway in the field. Unlike the transport processes mentioned above, a karyopherin family member that functions in general mRNA export has not been identified. Until recently, the primary candidates for mRNA export factors were the highly abundant hnRNP proteins, some of which shuttle between the nucleus and the cytoplasm. The shuttling hnRNP proteins were proposed to bind to mRNAs, and mediate their export. Recent data, however, have strongly implicated other proteins, Mex67 in yeast and its conserved metazoan homologue, TAP in the mRNA export pathway. Mex67 was first detected in yeast via a synthetic lethal screen with Nup85, a nucleoporin that functions in mRNA export (Segref et al., 1997). Evidence that TAP is a key player in mRNA export in metazoans was first provided by studies of the constitutive transport element (CTE), an RNA element required for export of unspliced genomic retroviral RNA (Grüter et al., 1998). These studies revealed that TAP binds directly to the CTE and mediates its export (Grüter et al., 1998). Moreover, excess CTE blocks mRNA export, presumably by titrating TAP. Mex67 forms a heterodimer with a small protein designated Mtr2 (Kadowaki et al., 1994), and this Mex67-Mtr2 heterodimer is essential for mRNA export (Santos-Rosa et al., 1998). The metazoan ortholog of Mtr2 is known as p15 or NXT. It was shown that the TAP-p15 heterodimer directly stimulates the export of cellular mRNAs, confirming its role in mRNA export (Katahira et al., 1999; Braun et al., 2001). Both Mex67-Mtr2 and TAP-p15 also shuttle between the nucleus and the cytoplasm (Katahira et al., 1999; Schmitt and Gerace, 2001). This observation, together with the finding that the conserved heterodimer interacts directly with the nuclear pore complex and (directly or indirectly) with the mRNA export cargo, strongly suggests that this factor is a

general mRNA export receptor. In contrast to karyopherins, the Mex67/TAP-Mtr2/p15 heterodimer does not require the GTPase Ran (Clouse et al., 2001).

Studies over the past several years indicate that there is extensive coupling between the different steps in gene expression. Consistent with this emerging concept, mRNA export is thought being coupled to transcription, pre-mRNA splicing, and other steps in gene expression. It has been proposed that assembly of the spliceosome causes retention of pre-mRNA in the nucleus (Chang and Sharp, 1989). However, release from retention factors does not seem to be sufficient for mRNA export. For some mRNAs at least, splicing enhances the efficiency of export compared with that of unspliced mRNAs that are transcribed from intronless complementary DNAs. This implies a link between splicing and export. Removal of a single intron from a two-intron-containing mRNA is sufficient for its efficient export, which indicates that mRNAs acquire a positively acting factor as a result of splicing (Kataoka et al., 2000). Ultraviolet crosslinking after *in vivo* splicing showed that several proteins become associated with spliced mRNAs (Le Hir et al., 2000). These proteins are organized into a complex, called the exon-exon junction complex (EJC), and are located about 20 nucleotides upstream of these junctions. The known components of the EJC include at least six proteins – SRm160, RNPS1, Aly/REF, Y14, magoh and Upf3.

Aly/REF, the metazoan homologue of the yeast mRNA export factor Yra1p, is recruited to mRNP complexes generated by splicing. Aly shuttles between the nucleus and the cytoplasm, and an excess of recombinant Aly increases both the rate and efficiency of mRNA export in vivo (Zhou et al., 2000). The REF family of evolutionarily conserved heterogeneous ribonucleoprotein (hnRNP)-like proteins consists of one central-RNP-type RNA binding domain flanked by Arg-Gly-rich regions of variable length. Members of this protein family bind directly to RNA and the mRNA export factor TAP/Mex67p. It has been suggested that they facilitate the recruitment of TAP/Mex67p to cellular mRNPs (Rodrigues et al., 2001).

Other EJC proteins – including Y14, magoh and Upf3 – also bind TAP, so it is likely that Aly/REF is not the only factor that can recruit TAP to mRNAs (Kataoka et al., 2001; Lykke-Andersen et al., 2001). So, by including a high local concentration of several TAP-binding proteins, the EJC could recruit several TAP molecules and enhance the export of spliced mRNAs.

Throughout the pathway of gene expression, numerous mechanisms exist to ensure the quality control of mRNA and hence the final protein product. Nonsense-mediated decay (NMD) is a conserved mechanism used to detect and selectively degrade mRNAs that contain premature stop codons. EJC contains components of the conserved export machinery as well as factors involved in NMD (Kim et al., 2001; Le Hir et al., 2001; Lykke-Andersen et al., 2001). Of most significance to NMD are a novel protein designated Y14, a splicing factor RNPS1, and hUpf3. All three of these proteins shuttle between the nucleus and the cytoplasm (Kataoka et al., 2000; Lykke-Andersen et al., 2000, 2001). Y14, hUpf3, Aly and TAP-p15 are thought to form a stable protein-protein complex (Kim et al., 2001). Thus, this complex functions in both mRNA export and NMD to transmit a positional mark of the exon-exon junction from the nucleus to the cytoplasm.

1.4 VIRAL mRNA EXPORT

Retroviral replication requires the nuclear export and cytoplasmic translation of both incompletely spliced and fully spliced forms of the initial, genom-length viral transcript. Thus, in retroviruses, the initial transcript leaves the nucleus as unspliced, incompletely spliced and multiply spliced mRNAs. The ability of retroviruses to express cytoplasmic mRNAs that retain one or more introns contrasts sharply with the pattern observed for cellular mRNAs. While cellular mRNAs may be subject to alternative splicing, they are nevertheless

exported from the nucleus in a fully spliced form. Indeed, nuclear export of RNAs via the mRNA export pathway is highly selective, so that incompletely spliced RNAs and introns are actively retained in the nucleus while only fully spliced mRNAs are able to reach the cytoplasm (Legrain and Rosbash, 1989; Fischer et al., 1994). A subset of splicing factors, termed commitment factors, are believed to be responsible for both the nuclear retention of incompletely spliced mRNAs and for initial steps in the recognition and definition of intronic sequences. The rationale for such a nuclear retention activity is obvious - if it did not exist, the nuclear export and cytoplasmic translation of pre-mRNAs would lead to the synthesis of nonsense proteins. Intron-containing retroviral mRNAs are exported from the nucleus via export pathways that are at least in part distinct from the nuclear export pathway used by fully spliced cellular and retroviral mRNAs. The study of these alternative nuclear export pathways had led to considerable insights into retroviral nuclear RNA export in particular and cellular nucleocytoplasmic transport pathways in general.

Among retroviruses, the HIV-1 retrovirus recently has been studied very extensively. Like other retroviruses, HIV-1 produces a single transcript with alternatively spliced forms, which at various points of its life cycle must be exported to the cytoplasm of the host cell for translation or packaging. Since eukaryotic cells retain their unspliced pre-mRNAs in the nucleus, a key question is how HIV-1 can circumvent the cellular retention machinery and export all its different mRNAs. The observation that the viral protein Rev is able to activate the cytoplasmic expression of incompletely spliced HIV-1 mRNA without further inhibiting the already low efficiency of HIV-1 RNA splicing led to the proposal that Rev functioned as a sequence specific nuclear RNA export factor (Felber et al., 1989; Hammarskjöld et al., 1989). The sequence specificity of Rev function is mediated by a *cis*-acting viral RNA stem loop structure, which is located in the *env* gene and is termed the Rev response element (RRE) (Malim et al., 1989). Rev directly interacts with a specific loop in the RRE via an arginine-

rich RNA binding motif that also serves as the Rev nuclear localization signal (NLS) (Zapp and Green, 1989; Malim et al., 1990). Rev functions as an adaptor to bridge the viral mRNA to the cellular export factor Crm1, which belongs to the Importin β family of nuclear transport receptors (Fornerod et al, 1997; Neville et al., 1997). The interaction between Rev and Crm1 is mediated by a leucine-rich activation domain located between residues 75 and 84 in the 116 amino acid Rev protein. This activation domain is in fact a nuclear export signal (NES) and it can induce the efficient nuclear export of linked substrate proteins (Fischer et al., 1995; Wen et al., 1995).

Importantly, the nuclear RNA export pathway accessed by Rev is distinct from the pathway that mediates the export of cellular mRNAs. Evidence demonstrating that was first reported in the *Xenopus* oocyte system. The Rev NES has been found to inhibit Rev function as well as 5S rRNA and U snRNA export but not mRNA export (Fischer et al., 1994). Furthermore, the drug leptomycin B, which is a specific inhibitor of Crm1, has been shown to inhibit Rev function but not mRNA export in higher eukaryotic cells (Fornerod et al., 1997). Therefore, it is apparent that the Rev-Crm1 pathway is distinct from the mRNA export pathway, although Crm1 is likely to be important for both 5S rRNA and U snRNA export.

The requirement of a viral Rev-like protein for export of nonspliced RNAs is bypassed by simple retroviruses, which do not encode regulatory proteins. The genomic RNAs of simple type D retroviruses bear *cis*-acting RNA elements that can functionally replace Rev or RRE in infectious HIV-1 clones by constitutively exporting unspliced or singly-spliced retroviral RNA (Bray et al., 1994; Zolotukhin et al., 1994; Tabernero et al., 1997). The *cis*-acting constitutive transport element (CTE) of the simian type D retrovirus and of the Mason-Pfizer monkey virus directly interact with the cellular factor TAP (Grüter et al., 1998; Kang et al., 1999).

The viral CTE RNA export pathway is shared with cellular mRNAs (Saavedra et al., 1997; Pasquinelli et al., 1997; Grüter et al., 1998). However, several lines of evidence suggests, that the mode of interaction of TAP with viral and cellular RNAs is likely to be different. First, most cellular RNAs do not contain sequences similar to the CTE stem loop structure, which is characterized by two conserved 12-nucleotide loops that form an almost perfect mirror image of one another (Tabernero et al., 1996; Ernst et al., 1997b) The binding of CTE is three orders of magnitude better than to mRNA (Braun et al., 1999), despite the fact that TAP shows general RNA binding capabilities (Braun et al., 1999; Katahira et al., 1999). Furthermore, titration of TAP by excess of CTE prevents cellular mRNA from exiting the nucleus of *Xenopus* oocytes whereas an excess of mRNA does not interfere with CTE export (Saavedra et al., 1997; Pasquinelli et al., 1997; Grüter et al., 1998). Finally, TAP stimulates the nuclear export of CTE in *Xenopus* oocytes but not of cellular mRNAs, indicating the cellular factor is not limiting in the latter process (Grüter et al., 1998). These lines of evidence all pointed to the suggestion that the CTE-RNA may access the mRNA cellular pathway at a late step and directly interacts with the export factor TAP, while the interaction of cellular mRNAs with TAP is likely to be more complex and mediated by many adaptor proteins (Reed and Hurt, 2002).

1.5 TAP

TAP is the metazoan homologue of the yeast mRNA export factor Mex67p and belongs to the conserved family of NXF proteins (Herold et al., 2000). To date knockout experiments in yeast, *Drosophila* and *C. elegans* have demonstrated that NXF family members are essential mRNA export factors. The human NXF1 protein TAP has also been

implicated in the nuclear export of unspliced genomic RNA of simple simian retroviruses (Grüter et al., 1998). Human TAP is a multidomain protein of 70 kDa molecular weight.

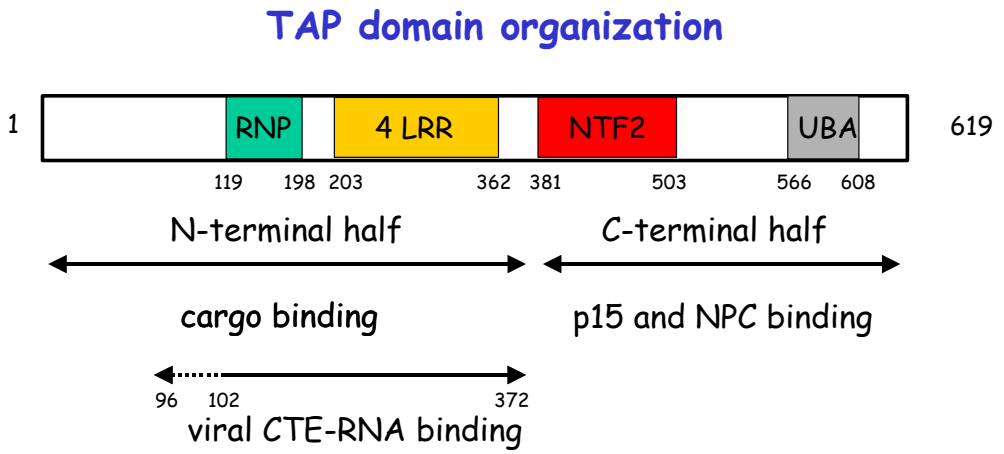


Fig. 2

TAP domain organization
1
RNP
4 LRR
NTF2
UBA
619
N-terminal half
C-terminal half
cargo binding
viral CTE RNA (direct)
cellular hnRNP-like proteins
p15 and NPC binding

Domain structure of human TAP. The domain mapping of the C-terminal region is as previously described (of the RNP and the LR in; LRR – Leucine Rich Repeat; NTF2 – Nuclear Transport Factor 2; UBA – Ubiquitine-associated

It contains an N-terminal cargo-binding domain and a C-terminal domain that 7functions in NPC translocation. The C-terminal portion binds to a subset nucleoporins that line the nuclear pore complex. In particular, several nucleoporins that line the NPC have been shown to interact with the carboxy terminal fragment (residues 508 – 619), which contains a putative ubiquitin-associated (UBA) domain (Bachi et al., 2000; Suyama et al., 2000). The upstream region (residues 371-551) contains a domain homologous to the RanGDP-binding transport factor NTF2.

The N-terminal half of TAP (residues 1-372) has been found to bind to several types of cargoes, including CTE RNA and RNA binding proteins such as E1B-AP5 (Bachi et al., 2000) and members of the Yra1/REF family of hnRNP proteins (Strasser and Hurt, 2000; Stutz et al., 2000). The binding sites are overlapping but not identical, with the first 60 residues being dispensable for CTE binding but important for binding to hnRNP-like proteins

(Braun et al., 1999; Bachi et al., 2000). The leucine-rich-repeat (LRR) region of TAP is essential for binding the CTE viral RNA and also affects cellular mRNA export when mutated (Braun et al., 1999).

In this study I try to get closer to determine the molecular details of CTE-TAP binding. The minimal CTE-binding domain of TAP was determined and this protein fragment was expressed, purified and crystallized. The structure of the CTE-binding domain was solved, revealing the presence of RNP and LRR domains. Mutational experiments confirmed these structural domains and determined the most important parts of these domains in RNA binding. With systematic modifications of CTE-RNA we found a relatively short but still functionally active RNA fragment what can be suitable in the future for crystallization, either alone or in complex with CTE-RNA.

2. MATERIALS AND METHODS

2.1 BASIC INTRODUCTION INTO X-RAY CRYSTALLOGRAPHY

2.2.1 Background of X-ray crystallography

X-ray crystallography is used to determine the structure of macromolecules and in particular their interactions (with substrates, inhibitors, effectors, etc.) at nearly the atomic level. X-ray crystallography exploits a physical phenomenon, diffraction. It is a similar principle to that used in a light microscope, where the magnified image of an object is obtained by illuminating the object with visible light, which is scattered (diffracted) by the sample and then recombined by using lenses. For an object to diffract, the wavelength of the incident radiation has to be no larger than the object. If we want to know atomic details, we need radiation with a wavelength comparable to the distance between bonded atoms in protein molecules (1.5 Å). Electromagnetic radiation with this wavelength falls in the X-ray range.

X-rays are scattered by electrons. A single protein molecule is a weak diffractor of X-rays, since it contains light atoms (C,N,O,H) with only few electrons. We use crystals, which contain many protein molecules (about 10^{15}) repeated with an identical orientation in a regular arrangement so that they diffract identically and their diffracted rays augment each other along certain directions to give detectable diffracted beams.

A crystal is built by an array of molecules arranged in a pattern that repeats in three dimensions in a periodic contiguous way. Crystals can be classified in a total of 230 different space groups. The space group of a crystal is a group of symmetry elements consistent with an infinitely extended, regularly repeated pattern. Since crystals of biological macromolecules consist of chiral molecular species, the application of mirror planes and inversion centers is not allowed and the number of possible space groups is limited to 65. The basic building

block of a crystal is the unit cell which is defined by three axes a, b and c and three angles α, β and γ. The unit cell is the scattering unit of the crystal. A crystal can be thought of being made up of unit cells repeated in three dimensions. Depending on the space group and on the symmetry elements that characterize it, the unit cell can contain one or several asymmetric units. The asymmetric unit is the unique portion of the unit cell that is not related to any other portion by crystal symmetry and that generates the contents of the unit cell when acted on by symmetry elements.

The first interpretation of X-ray diffraction patterns from crystals was given by Bragg in 1913. Reflections are treated as if they occurred from an imaginary set of planes through the crystal. The angles of diffraction obey Bragg's law,

$$2d \sin \Theta_n = n\lambda$$

where d is the interplanar spacing, Θ_n is the diffraction angle for the n^{th} order reflection and λ is the wavelength. The lattice planes are defined by the Miller indices h, k, l. Every reflection can be assigned distinct Miller indices h, k and l. The resolution at which a crystal structure is determined is equivalent to the minimum interplanar spacing that can be observed. The smaller the interplanar spacing, the larger the diffraction angle and the higher the resolution.

X-ray waves undergo an electromagnetic interaction with the electrons of the molecules in the crystal. The resultant wave scattered by a crystal in a particular direction defined by indices h,k,l is called the structure factor F(hkl). Unfortunately, there are no lenses that are able to focus the scattered X-rays in the same way as you can use glass lenses for the scattered visible radiation in a light microscope or magnets in an electron microscope. Instead, we use a 'mathematical lens', the Fourier transformation in order to obtain the magnified image of the scattering object, that is the distribution of electrons in the molecule, the so-

called electron density map. To calculate the electron density distribution within the unit cell, one needs to know the structure factors. The structure factor is a complex number and consists of the structure factor amplitude F(hkl) and the phase angle ϕ(hkl). The amplitudes can be obtained by measuring the intensities of the reflections, however the structure factor phases cannot be measured directly. Without the phase information, electron density cannot be calculated. This is the phase problem, which plays a central role in crystallography.

Various methods exist to solve the phase problem. The most common phasing methods are molecular replacement (MR) in cases where a sufficiently homologous structure is known or multiple isomorphous replacement (MIR) and multiple anomalous dispersion (MAD) where no similar structure information is available.

In case of molecular replacement (MR), if the structure of a homologous protein is known, we can use its coordinates to calculate an initial estimate of the phases. In this method the homologous probe structure is placed into the unit cell of the unknown structure. The success of this technique depends largely on the structural identity between the probe and the unknown molecule. The MR method can be applied when the probe shares at least 25% sequence identity with the unknown structure.

The multiple isomorphous replacement (MIR) method relies on the changes in intensities caused by a heavy atom bound to the crystalline protein molecule (heavy atom derivative) as compared to the parent (native) diffraction pattern. Estimates of the phases can be derived from the differences between the amplitudes of the derivative and the native diffraction data. The success of the method relies on the isomorphism between native and derivative structures: the two structures should differ only in the presence of the heavy atom at a position previously occupied by solvent. Heavy-atom derivatives are prepared by soaking the native crystals in a buffer containing a heavy-metal compound (mercury, platinum, uranium, etc.), or by cocrystallizing a protein with it. Usually at least two isomorphous

derivatives are required for a reliable phase determination, but often multiple derivatives are needed.

Multiple anomalous dispersion (MAD) exploits the changes in the diffraction pattern obtained when using different wavelengths around the absorption edge of certain atoms (anomalous scatterer) in the crystal. Different anomalous signals are collected by carefully choosing different wavelengths around the absorption edge of the compound. Apart from soaking a heavy metal in the crystal, particularly successful is the use of selenium as anomalous scatterer. Incorporation of selenium in the protein is achieved by supplying selenomethionine to a methionine-auxotropic expression system. The disadvantage of the method is that the signal is small in respect to the intensities (few percentages), and the data have to be very accurate. Obviously the selenomethionine protein has to be crystallizable. The advantage is that collecting the three or four needed data sets from a single crystal eliminates the problem of non-isomorphism.

Once good enough phases are available, the electron density map can be calculated and interpreted. A protein model built into an experimentally determined electron density map contains errors. Refinement is used to change the model such that it fits the experimental data the best. This is done by minimizing the difference between the observed data and the same quantities that can be calculated from the atomic positions of the model. The parameters we change in the minimization are the position of all the atoms in the structure (x,y,z) and also their temperature factors (B), which give a measure of how much the atom oscillates from the defined position specified in the model. Refinement is monitored by the so-called R-factor (reliability factor). The R-factor ($\Sigma/F_{obs}-F_{calc}/\Sigma\ F_{obs}$) indicates the agreement between the calculated structure factors (F_{calc}) and observed sructure factors (F_{obs}). The free R-factor is a much better estimate of the quality of the model, since it is an unbiased cross-validation. It refers to the R-factor calculated with a random portion of the data (2-10%) that has not been

included in the refinement. The R_{free} of a refined model has to be lower than 30% at ~2.8 Å resolution.

2.1.2 General principles and techniques for crystallization

The rate-limiting step in any crystallography project is the production of crystals that diffract to high resolution. Biological macromolecule crystallization has been regarded for many years as an art rather than a science, due to unpredictable and often irreproducible results.

Crystallization consists of two stages: screening, where we search for crystals of any quality, and optimization, where conditions are fine-tuned to improve crystal quality until the crystals are useful for data collection.

Crystals form from a supersaturated solution, in which there are more molecules in solution than allowed in its equilibrium states (solubility). The molecules will thus tend to be excluded from the solution and form a solid state, which can either be amorphous (precipitate) or ordered (crystal). To bring about supersaturation, a precipitant is added to the solution to compete with protein molecules for water. The most common technique is to use a hanging drop containing the concentrated protein in a few μl droplet and to slowly dehydrate it by vapor diffusion against a reservoir of 1ml containing the precipitant. The concentration needed to obtain crystals is supraphysiological and depends on the solubility of the protein, but usually ranges between 5 and 50 mg/ml.

Protein crystallization is a complex, multiequilibrium process, which can be divided into three distinct stages; nucleation, crystal growth and cessation of growth.

Nucleation

Homogenous nucleation is the spontaneous formation of solute nuclei in a supersaturated solution. Prior to nucleation, protein monomers are in equilibrium with various kinds of aggregates. In the metastable region of supersaturation, the aggregates are unstable and dissolve as rapidly as they are formed, whereas in the labile region, the formation of an aggregate large enough to be stable (a critical nucleus) is possible. At low levels of supersaturation, the formation of stable nuclei will occur at a slow rate. In practice, if the nucleation rate is so slow that no nuclei form in a week in a sample of a few microliters, the supersaturated solution may be considered metastable. As supersaturation increases, nucleation becomes rapid. At the other extreme, if many nuclei form within a second, the material may be considered as a precipitate.

For a protein to crystallize, there must be a regular pattern of molecular interactions. If there is no such regular pattern of interactions, material deposited from a supersaturated solution will form a non-crystalline aggregate (or precipitate).

Heterogeneous nucleation is the formation of solute nuclei on foreign substrates. This may happen on the wall of the containing vessel, on a crystalline surface on a foreign particle on the surface of an epoxy resin or on a biological contaminant. The activation energy for heterogeneous nucleation is less than for homogeneous nucleation, due to an attraction between the solute and the nucleant, so that heterogeneous nucleation occurs at lower supersaturation.

Crystal growth

The process of crystal growth consists of two steps. First is the transport of molecules from the bulk of the solution to the crystal/solution interface and the second is the attachment of the molecule onto the crystal.

Cessation of growth

Protein crystals often cease to grow after reaching a certain size that is often insufficient for X-ray diffraction studies. It is postulated that the incorporation of the impurities (the poisoning of attachment sites) is responsible for the growth termination.

Factors affecting the solubility of proteins

Protein crystallization depends on decreasing the molecule's solubility in solution to favor formation of a solid crystalline phase. Consequently, alteration of any solution property that affects protein solubility can be used to induce crystallization. In most cases, protein crystallization solutions include precipitants such as salts, alcohols or polymers like polyethylene glycol (PEG) to lower protein solubility.

Some proteins are poorly soluble in pure water, but do dissolve if a small amount of salt is added. This increase in protein solubility at low ionic strength is called "salting in". This phenomenon is explained by non-specific electrostatic interactions between the charged protein molecule and the ionic species of the salt. This salting-in effect, when used in reverse, can be used as a crystallization method. A protein that is soluble in an ionic salt is dialyzed against distilled water or a low ionic strength buffer to remove ions and reduce the protein solubility.

The opposite effect, used very frequently in protein crystallization, is "salting out", where the solubility of the protein decreases as the salt concentration increases. Salting-out of a protein from solutions of high ionic strength is due to competition between the salt ions and polyionic protein molecules for water.

The addition of polar organic solvents may produce precipitation or crystallization of proteins in a similar way to salts. The polar groups compete to some extent like salt for water molecules, but also reduce the dielectric constant of the medium, thereby increasing the

effective strength of electrostatic forces that attract protein molecules to each other. Polymers such as PEG also dehydrate proteins in solution and alter the dielectric properties like organic solvents.

Combinations of different precipitants are sometimes used to manipulate the attractive and repulsive non-specific interactions in a protein solution to favor crystallization.

The solubility of a protein is strongly influenced by variation of pH. At a given temperature and ionic strength, a solubility minimum is determined for a protein at some specific pH, usually the isoelectric point of the protein (pI), since at this pH there are no repulsive electrostatic forces between the molecules and no preferential electrostatic interactions with the ions in the solvent. At lower pH, there is a surplus of positively charged groups localized on the surface of the protein, and at higher pH, a surplus of negative charges is present. These modify the interactions of the protein with any charged (or polar) precipitating agent and also influence the dielectric properties of he crystallizing medium. The choice of buffer to control the pH is very important in protein crystallization. At exactly the same pH, the protein solubility can be very different depending on the nature and concentration of the buffer.

Most proteins vary in solubility as a function of temperature. At high ionic strength proteins are generally less soluble at 25°C than at 4°C. The temperature coefficient of solubility is often positive at low ionic strength, the solubility increases with increasing temperature. Generalizations concerning the temperature coefficient of solubility for proteins are difficult to make. Whether increase or decrease of temperature brings the protein solution to saturation depends on the particular protein and experimental conditions. Crystallization has been reported to occur for proteins over the entire range from 0 °C to 40°C although it usually conducted at 4°C or at room temperature.

2.1.3 Crystallization strategy

Protein purity

The first consideration when crystallizing a protein is its purity. Purity plays a critical role in crystallization and lack of its control explains many non-reproducible results. Crystallization relies on the cooperative addition of identical components to form a growing crystal lattice. If the molecules in the protein preparation are not identical, packing defects are introduced on the surface of the crystal. This results in termination of crystal growth or limitation of crystal quality. Protein purification is currently regarded as perhaps the single most important step in the crystallization of a macromolecule, as we need mg quantities of a stable, properly folded and homogeneous protein.

In order to grow high quality crystals it is important that the protein is pure, not only in removal of contaminants, but also in terms of structural microheterogeneities. There are different kinds of molecular heterogeneity one should worry about. First is chemical heterogeneity: the sample should contain different proteins only in a small percentage as judged from SDS gels. Protein truncations due to proteolysis are of particular menace since the fragment has some but not all the characteristics of the intact molecule, and termination of growth will be a likely outcome. Even more dangerous is microheterogeneity of proteins having otherwise the same amino-acid sequence. One example of microheterogeneity is aggregation, which results in the presence of different oligomeric states. Gel filtration chromatography, and more definitely dynamic light scattering and analytical ultra-centrifugation are used to detect it. Post-translational modifications are a second example of microheterogeneity, and include different types and extents of phosphorylation, acetylation or glycosylation. They can be detected with native gels and isoelectric focusing. Various enzymes can be used to remove the added chemical group, or the protein can be expressed in

bacterial expression systems, which cannot do these modifications. However, a particular modification and its extent are usually critical for function. Even if the modification in itself is very small, it may result in conformational heterogeneity.

Protein stability

Another crucial factor is the stability of the protein under storage conditions. Reducing agents such as dithiothreitol may be required to maintain sulphydryl groups in a reduced state and methal chelators, for example, EDTA may be added to sequester reactive metals that could bind to the protein.

Screening

Most attempts to crystallize proteins have traditionally relied upon setting large numbers of crystallization trials to determine what combination of pH, precipitant type, concentration, temperature etc. would yield some form of crystals. Subsequent refinement of conditions would then be used to optimize crystal growth.

There are some crystal screens available commercially, like the sparse matrix screen available from Hampton Research (USA). This screening method introduced by Jancarik & Kim (1991), based upon data contained in the Biological Macromolecule Crystallization Database. This strategy is designed to evaluate a large number of pH and precipitant combinations with a limited amount of protein sample. The main variables in this method are pH and buffer composition, additives, and precipitating agents. After Crystal screen I the range of screens available from Hampton Research has been recently expanded to include Crystal screen II (an extended sparse matrix), various grid screens (i.e. varying two parameters such as pH and precipitant simultaneously), a detergent screening kit, a nucleic acid screening kit and a membrane protein crystallization screening kit. Other laboratories

have published details of their own screening protocols (Scott et al., 1995; Cherezov et al., 2001).

Optimization

An initial strategy is to design a screen bracketing the reagent conditions which produced the crystals with finer increments of precipitant. The pH may also be optimized at this stage and alternative buffers tried. Different salts may be evaluated, combinations of salts and polymers may also be investigated. Changing the molecular weight of the PEG used as precipitant or using a modified PEG may yield better crystals. Another crystallization method may be used to achieve supersaturation by a different route. Additives are often used to improve the morphology or order of crystals. In some cases these are chosen on a rational basis i.e. inhibitors, coenzymes or ligands which may bind specifically to the protein and alter its conformational state. Another class of additives introduces a lattice of strong polar interactions between molecules within crystals. Detergents, especially β-octylglucoside have been widely used as additives for crystallization of both soluble and membrane proteins and are particularly useful when sample aggregation is a problem due to non-specific hydrofobic interactions. Seeding is frequently a useful technique for optimizing crystal growth. Crystal seeds obtained from earlier trials are introduced into protein droplets at a lower supersaturation than that required for nucleation.

2.2 METHODS

2.2.1 Methods used for TAP cloning, expression and purification

2.2.1.1 Plasmid cloning for TAP expression

Our aim was to crystallize TAP protein, and different TAP truncations in order to get crystals for X-ray crystallographic analysis. We have chosen the protein expression method in *E. coli* for that. We designed plasmid constructs for TAP protein expression with a GST tag, to purify the protein easier. In earlier studies Elisa Izaurralde constructed a plasmid for expressing full length GST-TAP 1-619, (pGEX-cs-TAP 1-619) using NcoI-BamHI restriction sites into a TEV cleavable GST expression vector (Braun et al., 1999). We further created plasmids for TAP expression encoding the truncated proteins, either for crystallize them, or for mutational studies to further investigate the role of TAP in binding to CTE.

Constructs encoding GST fusions of TAP fragments 1-118, 96-198, 61-372, 96-372 and 102-372 were subcloned into pGEX-cs vector (Parks et al, 1994; Fig. 3) with PCR, using NcoI-BamHI restriction sites.

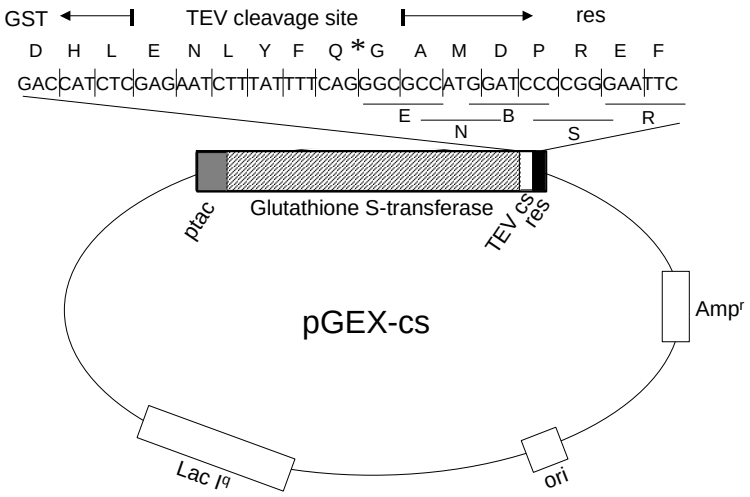


Fig. 3
pGEX-cs vector. The plasmid backbone is derived from pGEX-1 (Amersham Biosciences). TEV cleavage site and restriction sites (res) are marked. Abbreviations: E, EheI; N, NcoI; B, BamHI; S, SmaI; R, EcoRI.

The original pGEX-cs-TAP 1-619 construct served as a template for PCR. Fragment 199-372 was cloned into pQE-60 vector (Qiagen) with an uncleavable c-terminal His-tag, between NcoI-BamHI sites, while the GST fusion Mex67p 1-100 construct is in pGEX-5x-2 (Amersham Biosciences). All constructs were confirmed by restriction mapping and sequencing around the cloning sites. The ones we used for crystallization were fully sequenced before expression.

Site directed mutagenesis was performed using the Quick-change mutagenesis system (Stratagene), and confirmed by restriction mapping and sequencing. The site directed mutant constructs what were used for protein expression were the following:

pGEX-cs-TAP ΔRNP (Δ118-198); TAP R128E,K129E; TAP D228K; TAP E318R,E319R.

2.2.1.2 Mutagenesis of constructs used for *in vitro* translation

All mutations were introduced using an oligonucleotide-directed *in vitro* mutagenesis system from Stratagene (Quick-change site-directed mutagenesis) following the instructions of the manufacturer. Mutants generated for in vitro translation on pBSSK-TAP 1-619 are listed below. The *in vitro* translated proteins from these constructs were used in gelshift experiments.

Mutants generated to probe the interaction surface of the TAP RNP domain:

TAP K121A; TAP R128A,K129A; TAP R128E,K129E; TAP D131A,K129A;

TAP K132E; TAP A133Q,S137Q; TAP E155A,R158A; TAP N156A,T157A;

TAP A174Q,N176A;

Mutants generated to probe the interaction surface of the LLR domain:

TAPK213A,K218A; TAP D228A; TAP D228K; TAP K230A;

TAP R307A,K311A; TAP K311E,K313E; TAP E319A; TAP E318R,E319R; TAP D323K; TAP K347A; TAP K347E; TAP ΔRNP.

2.2.1.3 TAP expression and purification

TAP 102-372 was overexpressed in *E. coli* BL21(DE3) as glutathione S-transferase (GST) fusion protein linked by a Tev protease cleavage site. Cultures were grown at 37°C in LB medium and induced at an OD_{600} of 0.7 with 0.5 mM IPTG for 4 hours. Cells from 6 l cultures were resuspended in 150 ml buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM EDTA) with 0.5 mg/ml lysozyme and 1mM PMSF and stored at –80°C. After thawing, cells were lysed by sonication and the insoluble material was removed by centrifugation at 17,000 rpm. The soluble fraction was passed through a 5 μm filter and loaded onto a Glutathione column (Amersham Biosciences) equilibrated with buffer A. GST-TAP was eluted with reduced glutathione and dialyzed overnight against buffer A containing 4 mM β-mercaptoethanol. The fusion protein was cleaved by incubation with Tev protease at 4°C for 36 hours, and the mixture loaded on a cation-exchange column (Macroprep HiS, Bio-Rad) equilibrated with buffer A. Cleaved GST was washed away in the flow-through and pure TAP was eluted with a linear gradient of buffer A containing 1 M NaCl. More than 10 mg of pure protein per liter of starting *E. coli* culture were obtained and stored at 40 mg/ml in buffer B (20 mM Hepes pH 7.0, 100 mM NaCl, 10% glycerol). Other TAP fragments were expressed and purified with the same procedure.

The seleno-methionine variant of TAP 102-372 was produced by overexpression in B834(DE3)pLysS in minimal medium in the presence of 1mM $MgSO_4$, 0.02 mM $CaCl_2$, 0.2%

glucose, 1 mg/l thiamine, and supplemented with the 19 amino acids (excluding methionine) at 40 mg/l and L-seleno-methionine (Acros) at 50mg/l. The selenomethionine protein was purified with a similar protocol as used for the native TAP fragment, the major difference being a decrease in NaCl content (to 20mM) in the cation-exchange wash buffer to allow efficient binding of the selenomethionine protein to the resin. Analysis by mass spectrometry confirmed the full incorporation of the 4 seleno-methionines in the TAP preparation (data not shown).

2.2.1.4 TEV preparation

TAP fragments were expressed and purified with a GST-tag, which was cleaved with TEV protease. The necessity to cleave the tag with TEV protease at 4°C to prevent protein degradation demanded a very high amount of protease. Thus, the production and purification of TEV was preferred to the commercially available source (Gibco BRL).

The TEV protease has an apparent molecular mass of 29 kDa. The expression of a recombinant form of TEV in *E.coli* and its purification via histidine tag was previously reported (Parks et al., 1994). The protease activity was characterised and optimal cleavage conditions established, as reported in the Gibco BRL manual. TEV has a seven amino acid recognition site (Glu-Asn-Leu-Tyr-Phe-Gln*Gly, where * indicates the position of the peptide bond that is cleaved). This confers to the protein a high specificity and makes its use very convenient when dealing with large proteins.

TEV protease was expressed in BL21(DE3) as a His-tagged protein. Cultures were grown at 37°C in LB medium and induced at an OD_{600} of 0.5 with 0.5 mM IPTG at room tempetrature for 8 hours. Cells were harvested then resuspended in lysis buffer (20 mM Tris, pH 8.0, 500 mM NaCl, 10 μg/ml aprotinin, 1 mM PMSF), and lysed using french press. Cell

lysate was centrifuged at 15,000 rpm for 45 minutes to pellet cell debris. His-TEV was purified on a NTA-Ni column, and eluted with a linear gradient of imidazole from 20 mM to 1 M. The protein was further purified on a cation exchange HiS column.

2.2.1.5 *In vitro* translation

For generation of ^{35}S-labeled *in vitro* translated proteins, the combined *in vitro* transcription-translation (TnT) kit from Promega was used. Reactions were carried out at 30°C for two hours. Plasmid pBSSK encoding wild type TAP or various TAP mutants (listed earlier) were used as templates. Translation was checked by SDS-PAGE and subsequent autoradiography.

2.2.2 Methods used for RNA transcription

2.2.2.1 Cloning constructs for CTE-RNA transcription

DNA sequences corresponding to the desired CTE-RNA sequences were cloned into pBluescript-KSII+ (pBS-KSII+) vector at BssH II sites. The inserts were made by annealing synthetic oligonucleotides designed such, that after annealing the two strands, they gave BssH II sites at both ends. Constructs with the right orientation were chosen by PCR and confirmed by sequencing.

The advantage using these sites for cloning is that one BssH II site is located upstream of the T7 site in pBluescript, therefore using oligonucleotides containing the BssH II site, followed by the T7 site and the desired RNA sequence right downstream of it, we can avoid having nucleotides not related to CTE-RNA after transcription.

Cloning of the pBS-CTE-HH construct for RNA transcription with hammerhead ribozymes was done similarly. Two synthetic oligonucleotides corresponding to the sequence of CTE-RNA with hammerhead ribozyme sequences at both ends were annealed, and cloned into pBS-KSII+ vector at BamHI, EcoRI sites. The construct was confirmed by sequencing. Details of the sequence and the cleavage sites of the hammerhead ribozyme are given later (Results and Discussion).

2.2.2.2 General considerations for work with RNA

RNases are ubiquitous and have to be a constant concern when working with RNA molecules. Eucaryotic RNases are present in human fingerprints, for example, and bacterial RNases have to be removed from recombinant protein if it is going to be incubated with RNA. Therefore a general rule, common glassware was avoided, and solutions were prepared fresh in plasticware and sterile filtered. If glassware had to be used it was thoroughly rinsed with ethanol before. Tubing and columns for chromatography were extensively washed with 20 % ethanol before use. Gloves were worn at all time during bench work.

2.2.2.3 *In vitro* transcription of RNA

For run-off *in vitro* transcription with T7 RNA polymerase, plasmids were linearized with HindIII, phenol extracted and ethanol precipitated. The transcription reaction was done in the presence of 40mM Tris-HCl (pH 8.0), 5mM DTT, 1mM spermidine, 0.01% Triton X-100, 28mM $MgCl_2$, 4mM of each NTPs, 100μM linearized DNA template and T7 RNA polymerase in empirically determined amount. Neutralized nucleotides (Sigma) were stored

as 100mM solutions at –80°C. T7 RNA polymerase was expressed and purified by the Protein Expression Unit of EMBL.

The transcription reaction was incubated at 37°C for two hours, then the template DNA was digested with RNase free DNase I (Roche). 0.1 volume of 0.5M EDTA was mixed in to dissolve the magnesium pyrophosphate precipitate, then the sample was phenol extracted and desalted on a disposable PD10 column (Amersham Biosciences). Finally, the samples were concentrated by evaporation in a Speed Vac Concentrator to achieve the desired concentration of RNA. The transcribed RNA was checked on a 6 or 8% denaturing 8M urea acrylamide gel and on a 5% native polyacrylamide gel.

2.2.2.4 *In vitro* transcription with hammerhead ribozyme

Specially designed hammerhead ribozymes (Price et al., 1995) auto-cleave themselves off co-transcriptionally. The transcription reaction was done the same way as it was described before. After 2h at 37°C the transcription products were heated twice to 65°C and slowly cooled back to 37°C in order to drive the ribozyme cleavage reaction to completion. 0.1 volume of 0.5M EDTA was added to dissolve the pyrophosphate precipitate, and one volume of 8M urea was added before directly loading the reaction mixture into a single well on a denaturing (8M urea) 6% or 8% polyacrylamide gel.

For fractionation of the reaction products, the gels were run at constant 40W in standard 1xTBE buffer. The RNA was visualized with UV shadowing (312nm light over thin layer chromatography plates containing the fluorescent dye F_{254} (Merck)), and the corresponding gel slices were excised. The RNA was electroeluted in an elutrap (Schleicher & Schuell) at 4°C with 120mA for 6-7 hours. The elution was followed by measuring the

OD_{260} of the eluted fractions. The RNA containing fractions were pooled, and desalted on a disposable PD10 column, then concentrated in a Speed Vac Concentrator.

2.2.4 Assays

2.2.4.1 SDS polyacrylamide gel electrophoresis

SDS polyacrylamide electrophoresis was done using a minigel system (BioRad). General procedures were as described (Laemmli, 1970).

2.2.4.1 Limited proteolysis

Limited proteolysis was done in order to determine stable protein fragments. The idea behind is that stable compact protein domains are protected from proteases if we use them for a limited amount of time and concentrations. In our experiments we used subtilizin and trypsin. Subtilisin cuts at small hydrophobic residues, it is a generally favorite protease: it is not very specific and not very potent, which is ideal if somebody wants to cut floppy domain boundaries and not a specific residue in an exposed loop. Trypsin cuts at Arg and Lys residues.

Reactions were carried out the following way: 15µl protein (0.6 mg/ml in 20mM HEPES pH 7.5, 50 mM NaCl) and 4µl protease (different dilutions: 1/1000, 1/100, 1/10 of the 1mg/ml protease solution) were mixed and incubated 30 minutes on ice, then the reaction was killed with 1µl PMSF (100mM). After adding 5 µl of SDS gel sample buffer, the rection mixture was heated at 95 °C for 5 minutes, then loaded on a 15% SDS protein gel. The fragments were determined with N-terminal sequencing after blotting the gel on a PVDF membrane or with MALDI mass spec.

2.2.4.2 Native gel assay

Native RNA gels (20 cm x 20 cm x 0.1 cm) were cast in 50 mM Tris-acetate (pH 7.5) and 10 mM magnesium acetate using 5 % acrylamide (acrylamide /bisacrylamide, 19:1). Gels were prerun for 20 minutes at 4V/cm in the cold room (4 °C). To check conformational homogeneity, 2.5 μg RNA was annealed in 10 μl of binding buffer (10 mM HEPES pH 7.9, 50 mM KCl, 10% glycerol, 0.5 mM DTT, 10 mM $MgCl_2$), and loaded on the gel. Electrophoresis was carried out for 16 hours at constant 4 V/cm in 50 mM Tris-acetate (pH 7.5) and 10 mM magnesium acetate buffer. Gels were stained for RNA with toluidine blue.

2.2.4.3 RNase assay

In order to follow the course of RNase elimination during protein purification, a simple test for RNase activity was devised. 5 μg of pure CTE RNA was incubated for one week at 4°C and room temperature in binding buffer (10 mM HEPES pH 7.9, 50 mM KCl, 10% glycerol, 0.5 mM DTT, 10 mM $MgCl_2$) with excess protein samples taken at different stages of the purification procedure. Samples were analyzed on denaturing polyacrylamide gels in parallel to control samples without protein. Denaturing gels (20 cm x 20 cm x 0.1 cm) containing 8 M urea were cast in standard 1 x TBE (Sambrook et al., 1989) using 8 % acrylamide (acrylamide/bisacrylamide, 19:1). They were run hot at 20 W to check the covalent purity of the RNA samples. Gels were stained with toluidine blue and dried.

2.2.4.4 *In vitro* RNA binding assay

In vitro RNA binding was checked with an electrophoretic mobility shift assay. The basis of the electrophoretic mobility shift assay is the change in the electrophoretic mobility of a nucleic acid molecule upon binding to a protein. The unbound nucleic acid, in our case the RNA has a characteristic electrophoretic mobility on a native gel. If the RNA is bound by one or more proteins, the movement of the RNA through the gel is retarded. This gives rise to a characteristic shift in the position of the RNA band on the gel. In our case, this assay was done by two ways, with labeled RNA in the presence of nonspecific competitors, in order to check the RNA binding affinity of the truncated and mutated proteins, and with unlabeled RNA, without any nonspecific competitor molecule, using only the pure RNA and pure protein, in order to mimic the crystallization conditions.

2.2.4.4.1 Gel-shift with labeled RNA

For native gel assays with labeled RNA, a 224 nucleotide long CTE-RNA probe was employed. Synthesis and purification of the RNA probe and of unlabeled competitor RNAs was as described previously (Izaurralde et al., 1992). Binding reactions were performed with *in vitro* translated proteins in the presence of competitor tRNA (300ng/µl), herring sperm single-stranded DNA (30ng/µl) and M36 RNA (0.5µg/µl). M36 has the same secondary structure as the CTE RNA, but due to several nucleotide changes, it is unable to bind TAP (Grüter et al., 1998), therefore, it is the most suited nonspecific RNA competitor. Reactions were carried out in binding buffer (10mM HEPES, pH7.9, 50mM KCl, 5mM NaCl, 0.1mMEDTA, 10% glycerol, 0.5mM dithiothreitol and 0.025% NP-40). Final sample volumes were 10µl. After 30 min at room temperature, 1µl of a solution containing 0.2 mg/ml

of heparin and 0.05% bromophenol blue was added to the reaction mixtures, and incubation was continued for another 10 min at room temperature. Samples were applied to a 5% non-denaturing polyacrylamide gel (19:1 acryl:bisacryl ratio). Electrophoresis was carried out at constant voltage at 17 V/cm at 4°C in 0.5x TBE buffer. Complexes were visualized by autoradiography.

2.2.4.4.2 Gel-shift with unlabeled RNA

In order to mimic crystallization conditions, RNA binding assays were done without labelling the RNA, and without any additional competitors. In this case we can check the real binding stochiometry of the complex formation. For a binding assay, 2.5 μg CTE RNA was incubated with different TAP protein fragments at 1:1, 1:2 and 1:4 molar ratio in binding buffer (10 mM HEPES pH 7.9, 50 mM KCl, 10% glycerol, 0.5 mM DTT, 10 mM $MgCl_2$) for 30 minutes at room temperature. Samples were loaded directly on a 5 % native acrylamide gel and run as it was described before (native gel assay). Gels were visualized with toluidine blue, occasionally double stained with toluidine blue for RNA and Coomassie Brilliant Blue for proteins.

2.3 STRUCTURE DETERMINATION

2.3.1 Crystallization and data collection

TAP 102-372 was crystallized by vapor diffusion at 4°C, after mixing the protein solution at 20 mg/ml with an equal volume of the well solution containing 100 mM cacodylate pH 6.8, 18 % (w/v) polyethylene glycol (PEG) 8000 and 20 mM EDTA. Needle-shaped crystals appeared in one week and typically grew to a size of 50 x 50 x 400 μm.

Crystals were cryo-protected in 100 mM cacodylate pH 6.8, 12 % (w/v) PEG 8000, 10 mM EDTA, 20% glycerol and flash-frozen in liquid nitrogen-cooled propane. The crystals are in space group $P4_32_12$ ($a = b = 96.4$ Å, $c = 152.2$ Å) with two molecules per asymmetric unit and 57% solvent content (Matthews, 1968). Only weak diffraction to low resolution could be observed using in-house CuKα X-rays, while synchrotron radiation allowed to measure reflections to better than 3 Å resolution routinely. The data were processed with the Denzo/HKL package (Otwinowski and Minor, 1997).

Seleno-methionine substituted TAP 102-372 was crystallized under similar conditions, occasionally yielding bipyramid-shaped tetragonal crystals with a larger unit cell. They are in space group $P4_12_12$ ($a = b = 139.9$ Å, $c = 206.7$ Å) with 4 molecules in the asymmetric unit and 70% of solvent (Matthews, 1968). After exchange of the arsenate-based buffer with Bis-Tris propane in the stabilizing and harvesting solution, a MAD experiment was recorded around the absorption edge of selenium at the ESRF ID14-4 beamline. Three data sets at 3.5 Å, 3.5 Å and 3.15 Å resolution were sequentially collected at the peak wavelength (12666 eV), at the inflection (12661 eV) and at the high-energy remote (13200 eV) wavelengths.

2.3.2 Structure determination and refinement: the large tetragonal crystal form

An extensive heavy-atom search with the crystals with smaller tetragonal cell resulted in no useful derivatives due to severe non-isomorphism problems. Phasing by SeMet MAD with the bipyramid-shaped crystals with the larger unit cell was instead successful. MAD phasing was calculated as a special case of MIR, where the data set at the high-energy remote (λ3) was used as native and the other wavelength data were used as individual derivatives. The resulting electron density map allowed the identification of the asymmetric unit content,

which consists of four large LRR domains, two small RNP domains and another small domain that is partially disordered on a crystallographic two-fold.

The model was built with the program O (Jones et al., 1991) and refined using the maximum likelihood target in the program CNS (Brünger, 1998) against the high-energy remote data set extending to 3.15 Å resolution. The progress of the refinement was judged by monitoring the agreement between calculated and observed structure factors for the excluded reflections (*R*free%).

The final model consists of residues 205 to 362 of each LRR domain and of residues 119 to 198 of each RNP domain. The R_{free} is 30.3% at 3.15 Å resolution.

2.3.3 Structure determination and refinement: the small tetragonal crystal form

Phasing of the needle-shaped small tetragonal crystal form was achieved using the molecular replacement method with the program AmoRe (CCP4, 1994) and the LRR domain as search model. The model has been refined to 2.9 Å resolution to a *R*free of 27.8% and includes two LRR domains (residues 203 - 362), and one partially disordered RNP domain (residues 123 – 191). After superposition, the LRR domains superpose with each other with a rms separation of corresponding α-carbon atoms of 0.28 Å, and a rms separation of less than 0.6 Å when comparing with the four LRR domains in the large tetragonal crystal form.

3. RESULTS AND DISCUSSION

TAP has been shown to directly bind to CTE, a constitutive transport element of simple type retroviruses, and to promote the nuclear export of CTE-containing transcripts (Grüter et al., 1998). Our goal was to crystallize the CTE-TAP complex, in order to understand the nature of the RNA-protein interaction, and ultimately how it is exported.

Full length SRV-1 CTE is a 173 nucleotide-long RNA that folds into an extended RNA stem loop structure (Fig. 4). It contains two conserved internal loops, A and B, and an AAGA bulge adjacent to loop A. The loops are arranged in mirror symmetry on the RNA element. Previous analysis of the structural requirements for CTE function led to the prediction that the internal loops represent the interaction sites for cellular CTE-binding proteins. Their symmetric arrangement suggests duplication of the binding site (Tabernero et al., 1996, 1997; Ernst et al., 1997a, 1997b). The most plausible model is that TAP binds CTE at the conserved loops and thus exports the non-spliced viral CTE-RNA from the nucleus. X-ray crystallography is a very powerful tool to understand the molecular basis for this interaction.

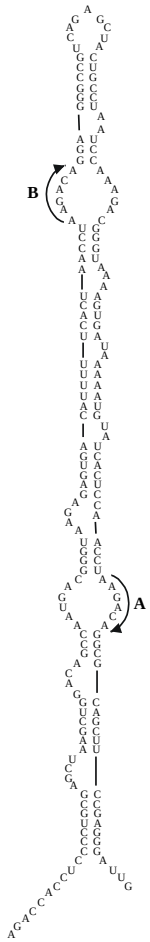


Fig. 4
Computer-predicted model of the secondary structure of the SRV1-CTE

The crystallization of such RNA-protein complex is rather difficult task. The strategy we chose was to identify the smallest RNA fragment that is still able to bind TAP, and to identify the smallest, but still stable protein domain that is able to bind the CTE-RNA. To this

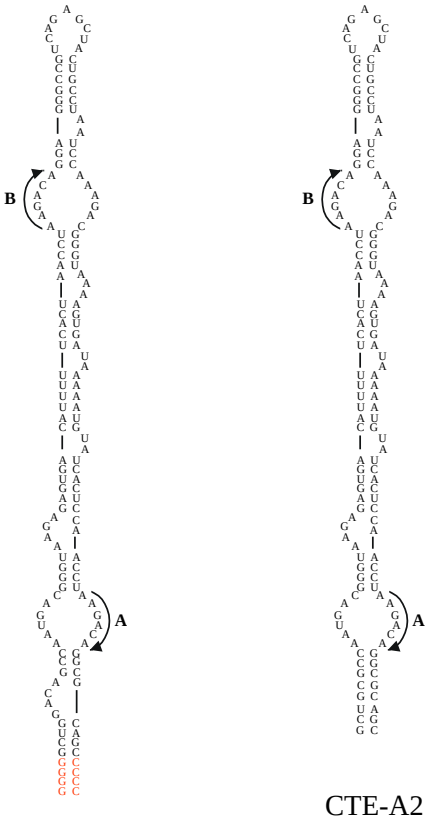
end we had to modify both the RNA and the protein systematically in order to determine the minimal RNA-protein complex to work with for crystallization. Since to get crystals from an RNA-protein complex can be difficult and time consuming, we considered first crystallizing the protein and the RNA alone, too. Although this cannot tell us everything about the interaction between them, it can still be very informative, both in providing a structural framework and later can help to determine the structure of the complex.

3.1 STUDIES ON THE CTE-RNA

3.1.1 Design of the CTE domains suitable for crystallization.

In our experiments we used *in vitro* transcribed CTE-RNA (see Materials and Methods). Large RNAs should be excluded because of their flexibility that generates conformational heterogeneity not compatible with crystallization. Smaller RNAs with more compact structures (with 75-120 nt) were the first natural nucleic acids to be crystallized (Abdel-Meguid et al., 1983; Dock et al., 1984). Since the 173 nt full length CTE-RNA is too long for crystallization, we designed shorter fragments, and tested whether they are still capable to bind TAP. We wanted to determine the minimal functional domain of the CTE-RNA. In previous studies (Grüter et al., 1998) it was shown that the possible interaction sites with the TAP protein are the loops A and B. We decided to modify CTE systematically, either to keep both loops but shorten the flanking regions, or to make CTE truncations keeping only one interaction loop.

First we designed two constructs where both the A and B loops are retained, but the RNA is a bit shorter than the full length CTE (CTE-A1 and CTE-A2) (Fig. 5).

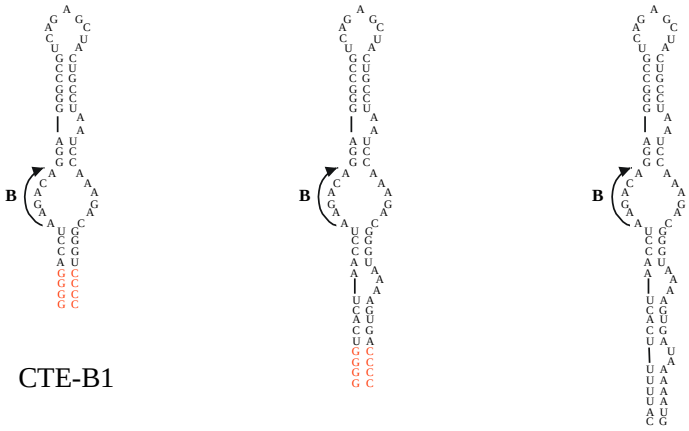


CTE-A1 CTE-A2

Fig. 5
CTE-A1 and CTE-A2 RNA constructs. The nucleotides highlighted in red are modified.

In these RNAs we removed the probably flexible 5' and 3' ends. We designed these constructs for *in vitro* transcription in pBluescript vector not to have non-specific nucleotides from the vector, after transcription. The original full length CTE construct for transcription (gift from Elisa Izaurralde) works nicely in assays where small amounts of labeled RNA is used and excess of non-specific competitor is added, but it's not usable in gel-shift assays in the absence of competitors. When we want to mimic crystallization conditions it is absolutely required not to have any disturbing agents like competitor molecules. In crystallization experiments it is essential to have just the RNA and the protein in stochiometrical ratios dissolved in a simple binding buffer (see Materials and Methods). The CTE-A1 and CTE-A2 constructs include both the A and B loops, just like in the wild type SRV-1 CTE-RNA. The only difference between the two RNA constructs is that while in CTE-A1 the small bulge next to loop A is kept, in CTA-A2 it is simply closed, by modifying some nucleotides (Fig. 5).

The CTE-B series of constructs only contains loop B, and the major difference among them is the length of the RNA surrounding the loop (Fig. 6). In CTE-B1 we kept few base pairs closing loop B, and stabilized the RNA by adding four GC base pairs in order to avoid



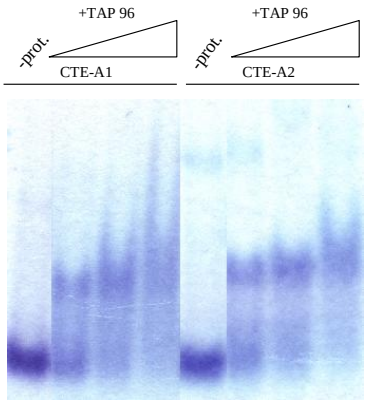
CTE-B1 CTE-B2 CTE-B3

Fig. 6
CTE-B1, CTE-B2 and CTE-B3 RNA constructs. The nucleotides highlighted in red are modified.

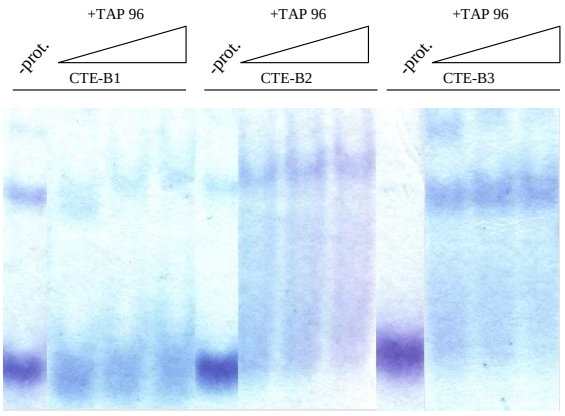
loop opening. CTE-B2 is longer than B1, and also contains four GC pairs to stabilize the RNA. CTE-B3 is the longest among the CTE-B mutants, in this case we assumed it was not necessary to stabilize the construct with additional nucleotides.

In order to test the TAP binding capability of the different RNA constructs we performed gel-shift assays, using non-labeled *in vitro* transcribed RNAs and TAP fragments in different molar ratios (see Materials and Methods). We have tested 1:1, 1:2 and 1:4 RNA:protein molar ratios. As it's shown on Fig. 7A, both CTE-A1 and CTE-A2 can bind TAP in a similar manner. Thus, these modifications of the wild type CTE do not significantly affect TAP binding.

Gel-shifts with the CTE-B constructs reveal (Fig. 7B) that CTE-B2 and CTE-B3 can also bind TAP efficiently. With CTE-B1 complex formation can be achieved while a



A



B

Fig. 7
Electrophoretic mobility shift assay showing the different affinity of TAP binding to (**A**) CTE-A RNA constructs, (**B**) to CTE-B RNAs. RNA:protein ratios were 1:1, 1:2, 1:4.

significant amount of unbound RNA is still present. We can draw two important conclusions after these experiments: (1) an RNA with only one loop still binds TAP, (2) among the CTE-B RNA constructs CTE-B2 is the shortest that is still capable of TAP binding. Another interesting observation we made is that while for the one-loop RNA we need 1:1 RNA-protein ratio to form a complex with TAP, for the two-loop RNA this amount of protein is not enough. It appears that we need twice as much protein to shift all the RNA that is present in the binding reaction. This is consistent with the previous model of the two loop constructs having two interaction sites per RNA molecule.

These native gels also show that RNA without protein tends to dimerize. To minimize this effect we adapted an annealing procedure, heating up the RNA to 65 °C and letting it cool down slowly. However, this annealing method was not efficient to get monomeric RNA. This is not an uncommon problem in RNA crystallization, with a conformationally homogeneous monomer RNA population it is not uncommon to obtain the structure of a dimer (Weichenrieder et. al., 2000; Jovine et al., 2000). This is specially the case for stem-loop RNA structures, like the CTE-RNA.

Since the conclusion from the gelshifts was that CTE-B2 construct is a good candidate to work with, we decided to modify it further, in order to find the best RNA for

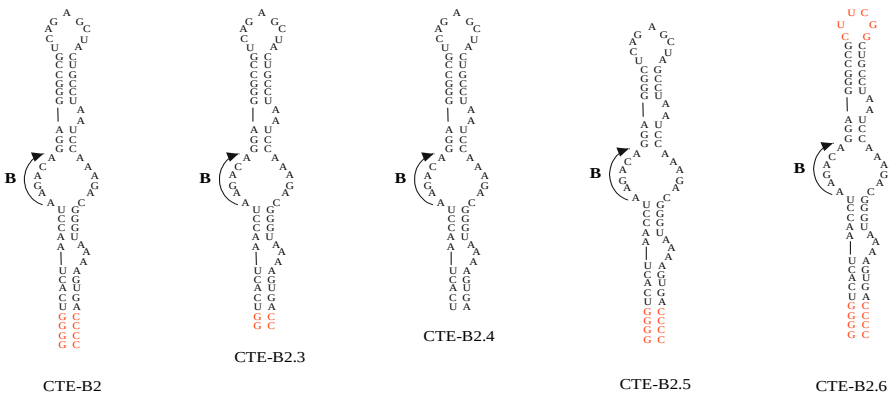


Fig. 8
CTE-B2, CTE-B2.3, CTE-B2.4, CTE-B2.5 and CTE-B2.6 RNA constructs. The nucleotides highlighted in red are modified.

crystallization. Fig. 8 shows a second set of CTE-B2 RNA constructs that we have designed, cloned and *in vitro* transcribed.

First we checked the effect of removal of the extra GC basepairs what we have added to CTE-B2 before, in order to stabilize the structure (CTE-B2.3 and CTE-B2.4). Then we tried to shorten the distance between loop B and the hairpin structure (CTE-B2.5). Finally, we replaced the end loop with a tetraloop (UUCG) that was shown to be thermodynamically a stable RNA conformation (Merouech and Chow, 1999). The strategy in all cases was to further shorten the RNA, to make it more suitable for crystallization. Fig. 9 shows the results of gel-shifts with these new constructs. The conclusion from this experiment was that we can remove the GC basepairs without any effect on protein binding (CTE-B2.3, CTE-B2.4), but shortening the stem loop (CTE-B2.5) or changing the endloop for a tetraloop (CTE-B2.6) abolishes binding.

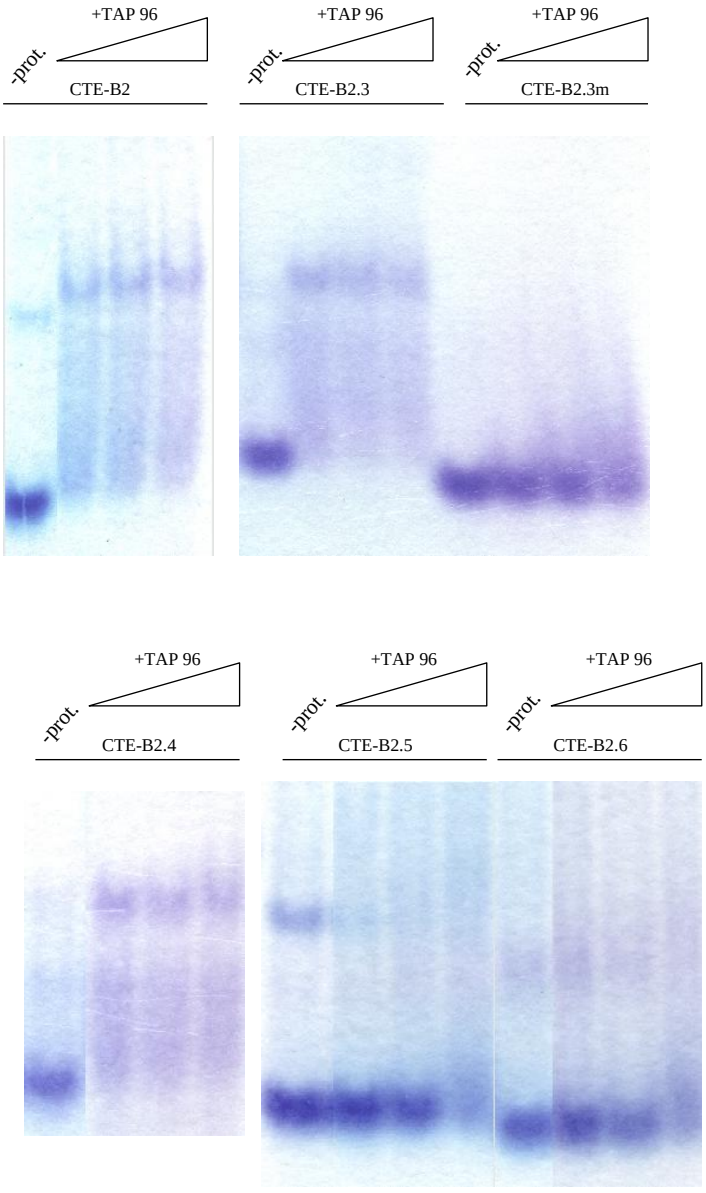


Fig. 9
Electrophoretic mobility shift assay showing the affinity of TAP binding to different CTE-B RNAs. RNA:protein ratios were 1:1, 1:2, 1:4.

Among the CTE constructs what we designed and tested, CTE-B2.4 is the shortest still retaining TAP binding ability. A mutant (CTE-B2.3m) with the sequence as CTE-B2.3, but for one point mutation in B loop shows nicely that loop B is required to be the interaction site for TAP, since the RNA with the mutated loop sequence is not able to bind the protein.

3.1.2 RNA transcription with "ribozyme technology"

Transcription of the CTE-RNAs described in the previous chapter was performed with simple run-off transcription. In this case an oligonucleotide encoding the desired RNA was cloned into a high copy number plasmid (pBluescript-KSII+) downstream of a T7 promoter. Large amount of plasmid DNA was isolated and then linearized to obtain template DNA for *in vitro* transcription. When a template plasmid is linearized with a restriction enzyme overhanging nucleotides limit the choice of 3' sequences. In addition, RNA polymerase often adds one or more extra nucleotides to the 3' end of the run-off transcript in a template-independent manner. Therefore, the product of run-off transcription is often very heterogeneous. A heterogeneous RNA is a serious drawback for crystallization as the sequence of the RNA is often critical. In order to avoid these problems, the desired sequence can be co-transcribed with flanking sequences which fold into appropriate self-cleaving ribozymes (Price et al., 1995; Ferré d'Amare and Doudna, 1996). A common choice is the hammerhead ribozyme. The hammerhead ribozyme was identified as a sequence sufficient for self-cleavage in satellite RNAs of certain plant viruses (Prody et al., 1986; Hutchins et al., 1986), and catalyses a transesterification reaction generating a cyclic 2',3'-phosphodiester and a free 5'-hydroxyl terminus of the product RNAs. The hammerhead ribozyme is the best choice to generate 5' terminus of the desired RNA sequence because of the absence of

sequence restrictions 3' to the cleavage site. However, the cleavage site must be preceded by the dinucleotide UX on the 5' side (X is any nucleotide other than G).

In our case we have chosen a ribozyme strategy to transcribe the CTE-RNA. Hammerhead ribozymes were used at both the 5' and 3' ends of the CTE-RNA to avoid having additional nucleotides as a result of the vector used for cloning the CTE and additional nucleotides generated by run-off transcription. Fig. 10A shows the sequence of the CTE-RNA generated by a double cis-acting hammerhead ribozyme (Fig. 10B).

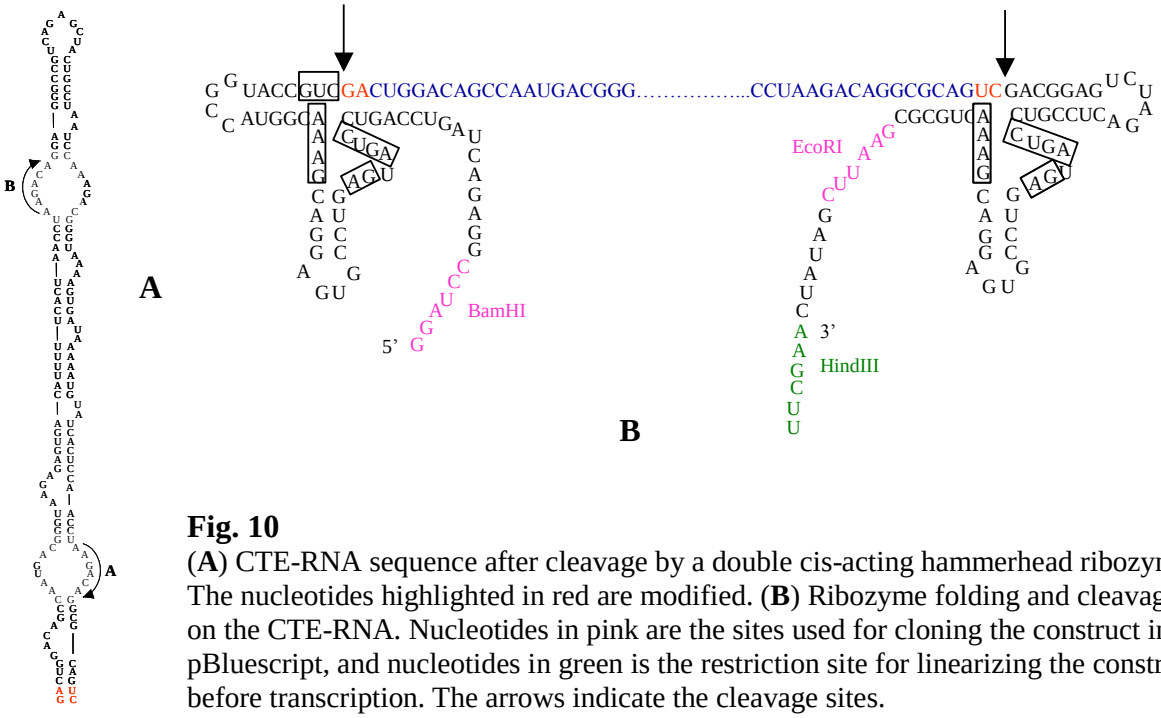


Fig. 10
(**A**) CTE-RNA sequence after cleavage by a double cis-acting hammerhead ribozyme. The nucleotides highlighted in red are modified. (**B**) Ribozyme folding and cleavage on the CTE-RNA. Nucleotides in pink are the sites used for cloning the construct into pBluescript, and nucleotides in green is the restriction site for linearizing the construct before transcription. The arrows indicate the cleavage sites.

The arrows indicate the cleavage sites of the ribozyme and the nucleotides in boxes show the nucleotides essential for ribozyme activity. Some nucleotides were changed (highlighted in red) for appropriate ribozyme activity at the 3' end and therefore at 5' end for getting basepairing in the resulting CTE-RNA. Nucleotides in pink indicate the cloning sites in the construct, and nucleotides in green show the restriction site used for linearization of the construct prior transcription.

During co-transcription of the CTE-RNA with the two ribozymes (see Materials and Methods) the CTE-RNA and M45, a mutant one-loop CTE-RNA (Grüter et al., 1998) are cleaved out, and the resulting products are visible on a denaturing gel (Fig. 11A).

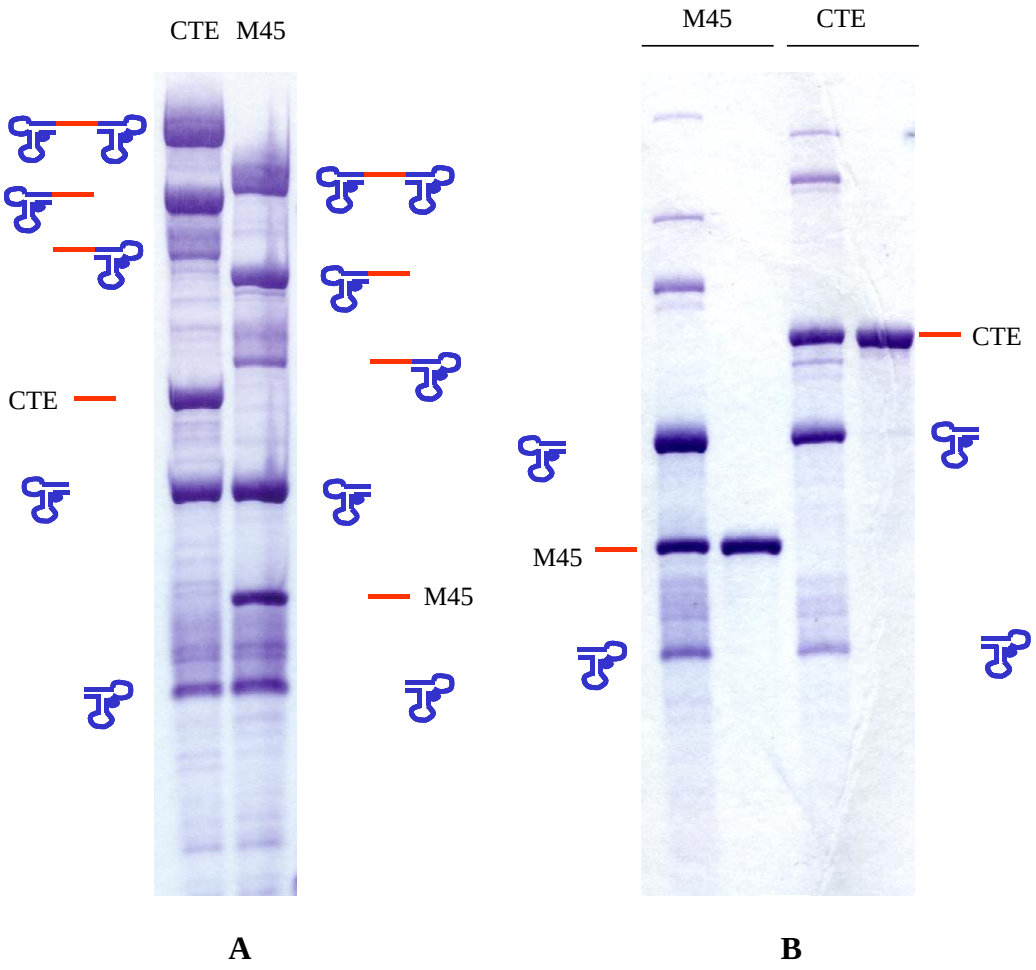


Fig. 11
CTE and M45 RNA fragments after cleavage with hammerhead ribozyme on a denaturing polyacrylamide gel (**A**) after the first cleavage reaction and (**B**) after the second cleavage reaction. The symbols indicate the cleaved or partially cleaved RNA products.

Since the cleavage efficiency was not 100%, during large-scale transcription, we cut out the uncleaved CTE-ribozyme product from the preparative denaturing urea-acrylamide gel, electro-eluted it and performed a second cleavage reaction. With this method we achieved 90 % cleavage (Fig. 11B) of combining the two CTE-RNA fractions purified from the

original transcription-cleavage and the second cleavage reaction. As expected, the CTE-RNA transcribed from a ribozyme construct can efficiently bind TAP (Fig. 12).

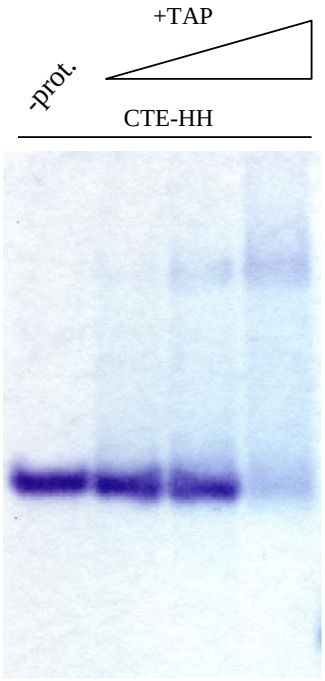


Fig. 12
The hammerhead ribozyme-cleaved CTE-RNA can bind TAP.

Taken together, our conclusion from the RNA experiments is that the CTE-B2.4 construct is the best candidate to start crystallization experiments, because it is the smallest RNA that binds TAP. RNA transcription can be best done with double hammerhead ribozyme constructs to avoid heterogeneous 3' ends and 5' sequence limitations due to transcription start nucleotides.

3.2 CRYSTALLIZATION OF TAP

3.2.1 Expression and purification of TAP

Crystallization requires a large amount of very pure protein. To obtain sufficient amount of protein, the generally used strategy is to clone the cDNA into a bacterial expression vector and overexpress it in an *E.coli* strain. To this end, we have chosen an expression vector containing an N-terminal GST tag that is cleavable with TEV protease (pGEX-cs) (Parks et al 1994). The GST tag ensures that the protein can be easily and quickly purified from the whole cell lysate by affinity chromatography. After transforming *E. coli* BL21(DE3) strain with

TAP constructs we grew cells at 37 °C in LB medium and then induced at an OD_{600} of 0.7 with 0.5 mM IPTG for 4 hours. When we loaded the induced and uninduced whole bacterial cells on a Coomassie stained SDS gel we could see relatively big amount of expressed TAP proteins (Fig. 13).

This gel shows the expression of three different TAP fragments with a GST tag, which can bind CTE. The expression level is very high, as it's visible on a gel (Fig. 14) showing the fractions that we got from glutathion sepharose column purification (see Materials and Methods). From 6 liters of culture we collected 5ml fractions, and from the fractions 10µls

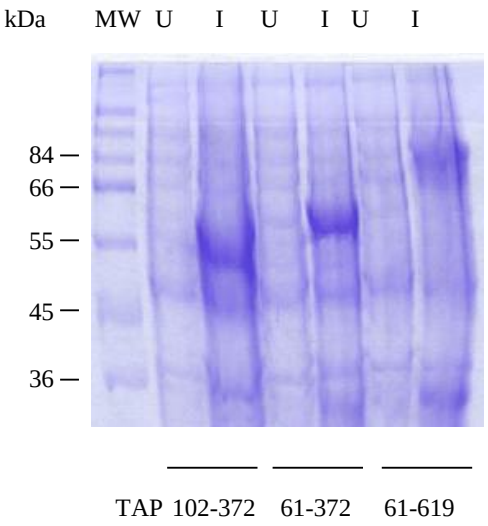


Fig. 13
Expression of TAP 102-372, 61-372 and 61-619 protein fragments in induced (I) versus uninduced (U) bacterial cell.

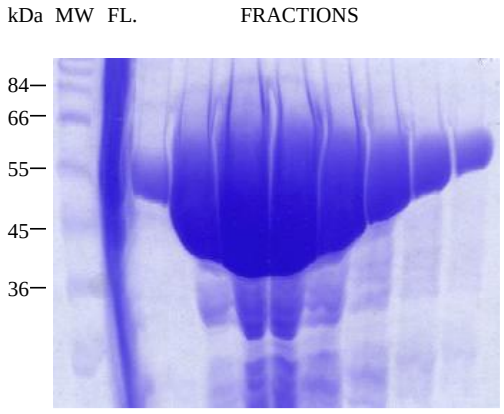


Fig. 14
The amount of theTAP102-372 protein in the different fractions after glutathion sepharose column purification. The flowthrough is indicated as FL.

were loaded on a 10 % SDS polyacrylamide gel. This gel is heavily overloaded, but it nicely shows how efficient the TAP overexpression was.

The fusion protein then was cleaved with TEV protease (36 hrs at 4 °C). In order to separate the cleaved GST tag from TAP, a second purification step was necessary. For that purpose a HiQ anion exchange or a HiS cation-exchange column can be used efficiently. Since our major goal was to use the purified TAP for complex formation with RNA, it was a

strict requirement to remove all RNases from the protein prep. After comparing the purification profile of TAP and the behavior of RNase A on these columns (Fig. 15), we decided to use the HiS cation-exchange column, since this can efficiently remove the GST tag and the RNase A together from the TAP-GST mixture. GST and RNase A can be found in the flowthrough and TAP can be eluted from the column with a linear gradient of buffer A (see Materials and Methods) containing 1 M NaCl.

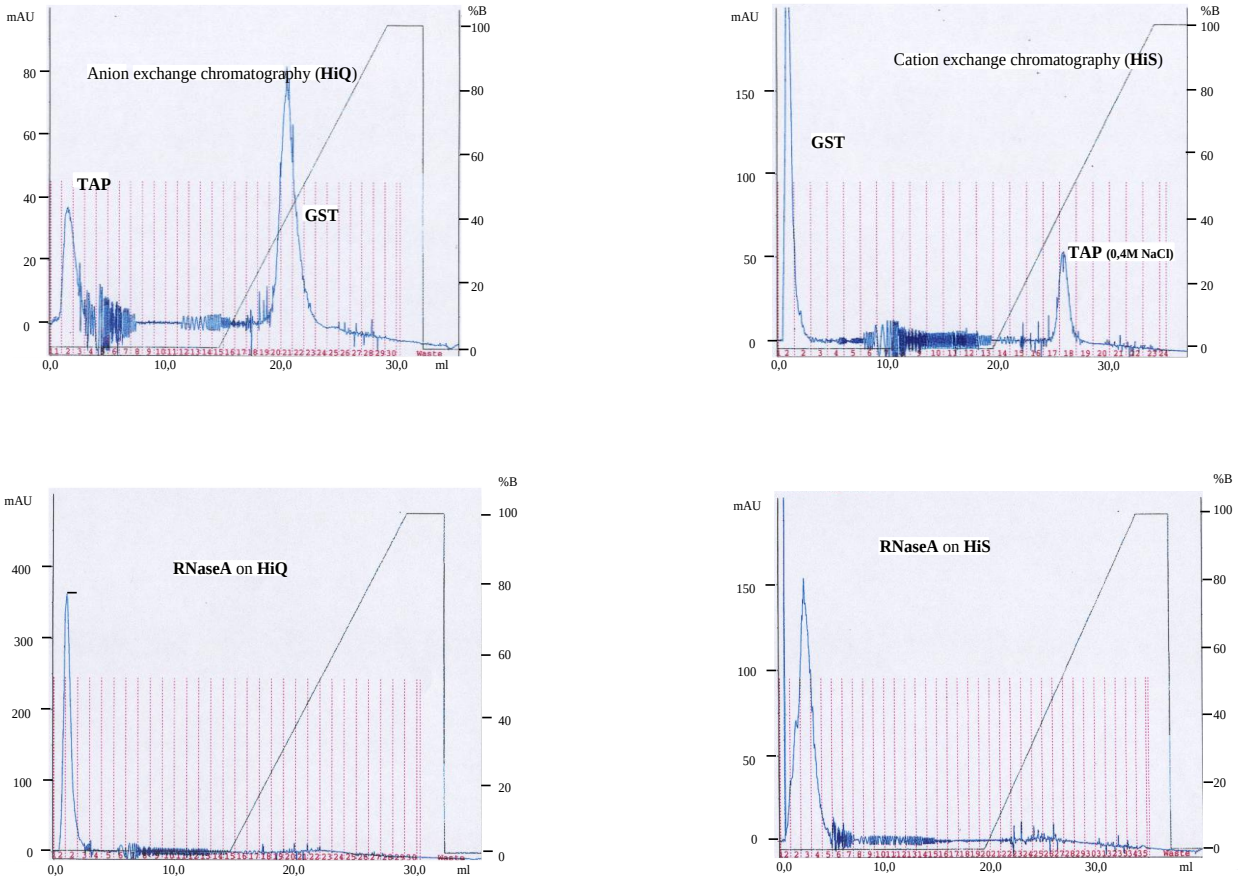


Fig. 15
Purification of cleaved TAP from GST by ion exchange chromatography.

More than 10 mg of pure protein per liter of starting *E. coli* culture were obtained and stored at 40 mg/ml concentration in buffer B (see Materials and Methods). All the expressed TAP fragments were purified similarly.

On Fig. 13 and Fig. 16 it is visible that we expressed three different fragments of TAP: TAP 61-619, TAP 61-372 and TAP 102-372. TAP 61-619 was originally thought to be the full length protein (Yoon et al., 1997), and this fragment can bind CTE as efficiently as TAP 1-619 (Braun et al., 1999). Moreover, the export functions of the truncated protein cannot be distinguished from that of TAP 1-619. Previous domain mapping studies of TAP (Braun et al., 1999) revealed that the 61-372 fragment is able to bind the CTE-RNA directly, and within this domain, amino acids 102-372 define the minimal CTE-binding domain. Nevertheless, we should mention that TAP 102-372 appeared to bind RNA more efficiently when fused to GST (Braun et al., 1999).

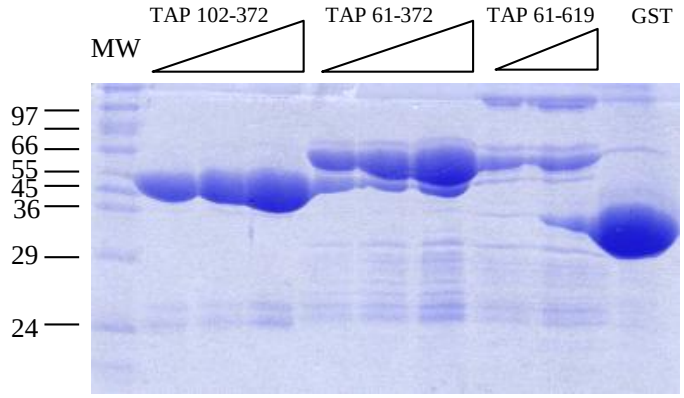


Fig. 16
Expression of different TAP protein fragments as GST fusion.

The prerequisite for protein crystallization is the obtainment of a stable and homogenous protein fragment in milligram amounts. When we expressed the three fragments mentioned above, we could conclude that TAP 61-619 expression didn't result in sufficient amounts of the protein, and the protein had limited solubility. TAP 61-372 was proteolitically sensitive, while 102-372 looked rather stable (Fig. 16).

To determine stable fragments of a protein, a limited proteolysis experiment can be very informative. The idea behind is that proteases can cleave a floppy end of a protein or cut it between domains. If proteases are used in a proper dilution, they cannot access a compact protein domain. In this case, adding trypsin (cuts at Lys, Arg residues) and subtilisin (cuts at small hydrophobic residues) in proper dilutions, we could get similar digestion pattern (Fig.

17 A,B). TAP 61-372 gave a stable fragment running similarly as TAP 102-372, and the 61-619 fragment first degraded to 61-372 than to 102-372 fragments. To identify the fragments precisely, we used MALDI mass spectrometry and N-terminal sequencing. In fact, it turned out with N-terminal sequencing, that the most stable fragment given by proteolysis is longer few residues than TAP 102-372. This stable fragment was TAP 96-372. Since preliminary crystallization experiments gave worse results with TAP 96-372, rather than with fragment 102-372, we started to work with the latter. As TAP 102-372 is functionally active (Braun et al., 1999), we continued setting up crystallization trays with this protein fragment.

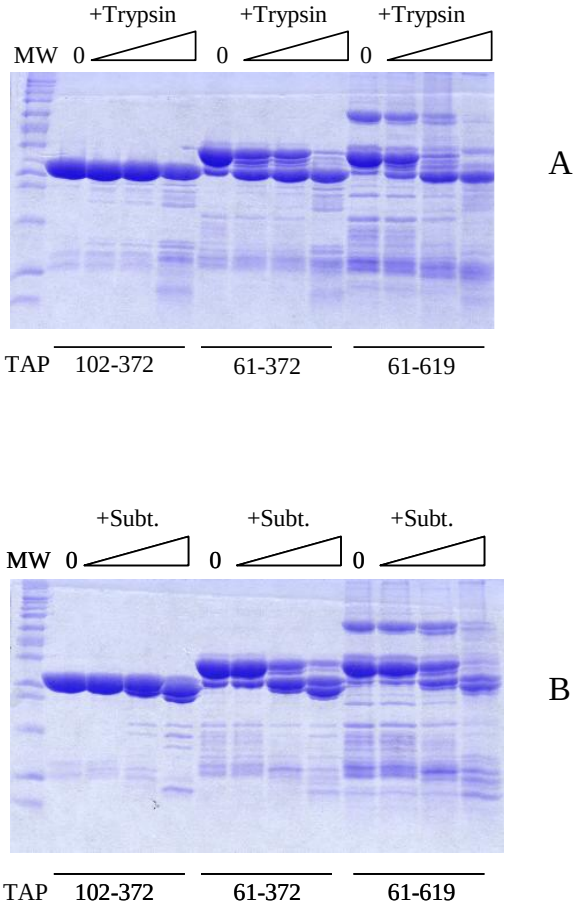


Fig. 17
Limited proteolysis experiment using increasing amount of (**A**) trypsin and (**B**) subtilisin to digest the different TAP protein fragments indicated. After proteolysis the protein samples were resolved on SDS-PAGE

3.2.2 Crystallization of TAP 102-372

After we purified a relatively large amount of pure protein, we started to explore good crystallization conditions. We used the hanging drop method. In this method, the hanging

drop contains the concentrated protein in a few µl droplet and it slowly dehydrates by vapor diffusion against a reservoir containing a precipitant at higher concentration. In general, setting large numbers of crystallization trials determines what combination of pH, precipitant type, concentration, temperature etc. would yield some form of crystals. Subsequent refinement of conditions is then used to optimize crystal growth. (More details are in Materials and Methods.)

A commercially available crystal screen, the sparse matrix screen (Hampton Research) gave initial crystals, which were than optimized by tuning the conditions. The final crystallization condition of TAP 102-372 was vapor diffusion at 4°C by mixing the protein solution at 20 mg/ml with an equal volume of the well solution containing 100 mM cacodylate pH 6.8, 18 % (w/v) polyethylene glycol (PEG) 8000 and 20 mM EDTA. Needle-shaped crystals appeared in one week and typically grew to a size of 50 x 50 x 400 µm (Materials and Methods).

3.2.3 Structural overview

The three-dimensional structure of the minimal CTE-binding fragment of human TAP 102–372 was determined at 3.15 and 2.90 Å resolution in two different crystal forms, which contain four and two independent protein molecules per asymmetric unit, respectively (see Materials and Methods). The structure was initially solved in the large tetragonal crystal form by multiple wavelength anomalous dispersion (MAD) methods with selenomethionine-substituted protein (For expression and purification of the selenomethionine protein see Materials and Methods). The model was built and refined with the aid of multi-domain non-crystallographic symmetry (NCS) averaging and used to determine the structure in the small tetragonal crystal form by molecular replacement techniques. The structures have been

refined to an R_{free} of 27.8 and 30.3% in the 2.90 and 3.15 Å resolution crystal forms, respectively. Since in this thesis I don't focus on the structure determination technique based on the X-ray diffraction data, more details are not given here. (For more details see Liker et al., 2000.)

As a result from structure determination we can conclude that the minimal CTE-binding domain of TAP comprises two tandem globular domains and an N-terminal disordered flexible region. The N-terminal domain has a ribonucleoprotein (RNP) fold, which is one of the most common RNA-binding modules. The C-terminal domain folds into a leucin rich repeat (LRR) containing protein. Below I'll present and discuss the structural data and functional studies with structure-based mutants.

3.2.3.1 The N-terminal domain: an RNP fold with non-canonical sequence motifs

The region of human TAP encompassing residues 119–198 folds into a four-stranded antiparallel β-sheet with a pronounced right-handed twist (Fig. 18A).

Two perpendicularly oriented α-helices pack against one side of the β-sheet, whereas the other side is exposed to solvent. The C-terminal helix (α2) of the domain is connected to the C-terminal strand (β4) by an extended loop containing two small antiparallel β-strands (Fig. 18B). The βαββαβ topology of this 80-residue domain is similar to the secondary structure elements of a prototype RNP domain (Burd and Dreyfuss, 1994; Varani and Nagai, 1998). Indeed, structural similarity searches of the Protein Data Bank using the program DALI (Holm and Sander, 1993) indicated statistically significant similarities with several RNP domains including the spliceosomal components U1A and U2B", and the single-stranded RNA-binding proteins Sex-lethal (Sxl) and poly(A)-binding protein Pab1 (Oubridge et al., 1994; Price et al., 1998; Deo et al., 1999; Handa et al., 1999). A structure-based sequence alignment between human TAP and the RNP domains of Sxl, Pab1, U1A and U2B" is shown

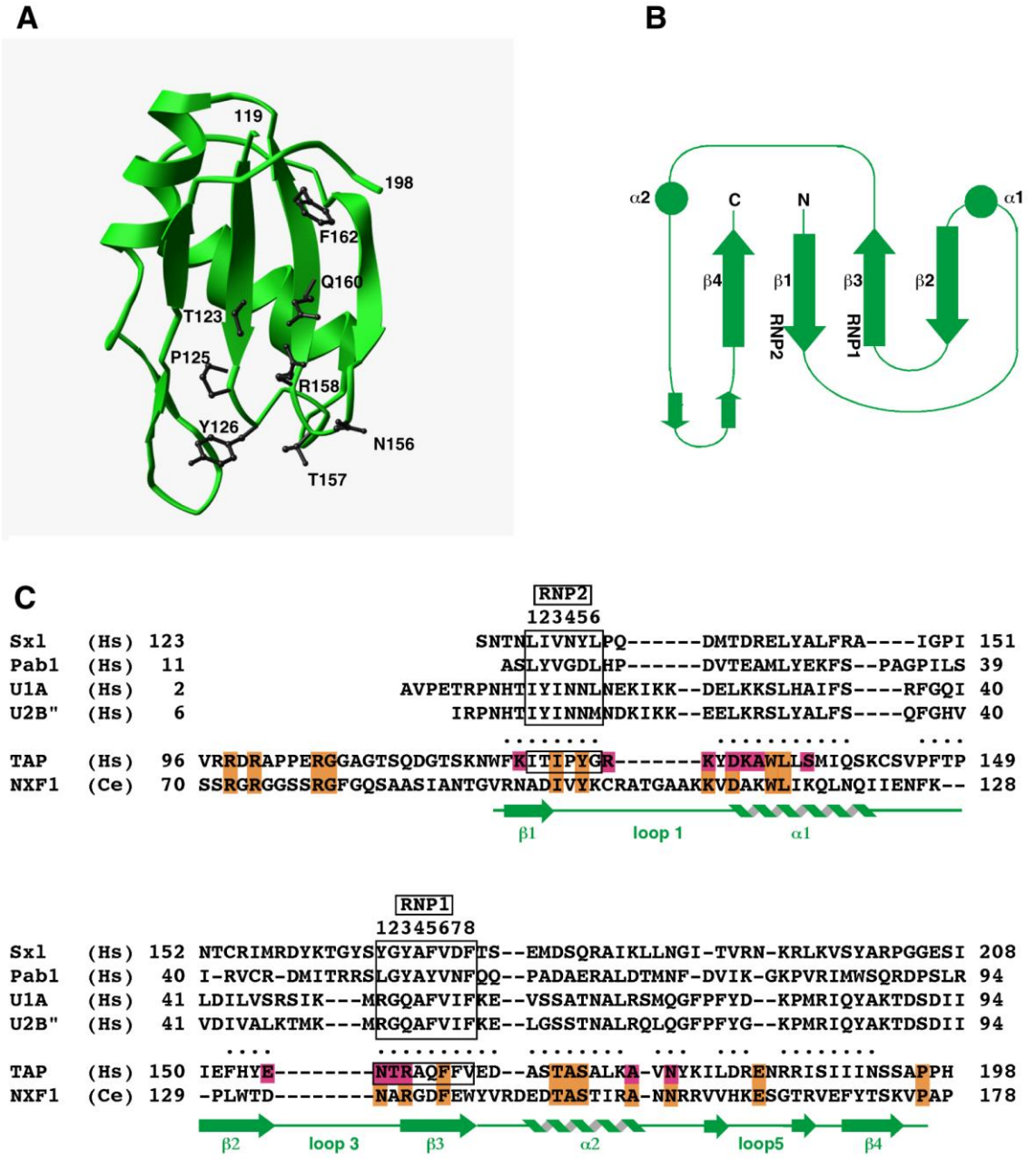


Fig. 18
The ribonucleoprotein (RNP) domain of human TAP. (**A**) A ribbon diagram showing in black the side chains of residues that are solvent exposed on the ß-sheet platform and are part of the RNP1 and RNP2 consensus sequences of the domain. (**B**) Topological arrangement of the secondary structural elements. The circles represent α-helices and the arrows ß-strands. (**C**) Structure-based amino acid sequence alignment of the N-terminal domain of human TAP with the RNP domains of Sxl, Pab1 and the spliceosomal proteins U1A and U2B". The dots indicate residues in TAP that lie within 3.0 Å of corresponding RNP C_α atoms after optimal superposition of the domains. Residues that have been mutated and analyzed for function are highlighted in pink, while conserved residues are orange.

in Fig. 18C and reveals only one identity (Ala159 in β3) between equivalent residues in the RNP domains. In the light of the extremely limited sequence homology, the structural similarity is remarkable. The alignment shows that the RNP domain of human TAP lacks canonical RNP1 and RNP2 consensus sequence motifs, but maintains a loose sequence homology for the structurally important hydrophobic residues that are interspersed at particular positions in the sequence. Similarly, a cryptic RNA-binding domain has recently been revealed in the crystal structure of poly(A) polymerase (Martin et al., 2000).

The RNP1 sequence is defined by eight amino acid residues with consensus $(K,R)_{OUT}$-G-$(F,Y)_{OUT}$-$(G,A)_{IN}$-F_{OUT}-V_{IN}-x_{OUT}-$(F,Y)_{IN}$ (where x is any amino acid, and IN and OUT refer to the amino acid side chain pointing either into the core or out to the solvent). The canonical RNP2 sequence consists of the six residues $(L,I)_{IN}$-$(Y,F)_{OUT}$-$(V,I)_{IN}$-$(G,N)_{OUT}$-$(G,N)_{OUT}$-$(L,M)_{IN}$. As originally shown in the structure of U1A (Nagai et al., 1990), the two sequence motifs are juxtaposed on the two central β-strands β3 and β1 and serve both a structural and a functional role. The residues at positions 4, 6 and 8 of RNP1 and at positions 1, 3 and 6 of RNP2 point to the interior of the domain and are crucial for the formation of the hydrophobic core by packing with residues of the two α-helices. The charged and aromatic side chains typically present at the other positions are exposed to solvent (these residues are shown in black in Fig. 18A) and are usually involved in contacting the RNA (Oubridge et al., 1994; Price et al., 1998; Deo et al., 1999; Handa et al., 1999).

At the equivalent positions in the structure of TAP, the octameric RNP1 sequence N_{OUT}-T_{OUT}-R_{OUT}-A_{IN}-Q_{OUT}-F_{IN}-F_{OUT}-V_{IN} (residues 156–163) and the hexameric RNP2 sequence I_{IN}-T_{OUT}-I_{IN}-P_{OUT}-Y_{OUT}-G_{IN} (residues 122–127) deviate significantly from the consensus. In RNP2, a glycine is found in place of the hydrophobic amino acid trademark of position 6, while the solvent-exposed residue at position 4 is an unusual proline, whose occurrence prematurely ends the β-strand conformation of the RNP2 motif. Furthermore, in

RNP1 the conserved glycine at position 2 is replaced by a threonine. Not surprisingly, this is at the loop region displaying dihedral angles that are unfavorable for non-glycine amino acids.

3.2.3.2 The C-terminal domain: an LRR fold structurally homologous to U2A'

The C-terminal domain of the CTE-binding fragment spans residues 203–362. It includes not only the four tandem LRRs that had been detected previously by sequence analysis (Segref et al., 1997), but also the regions that flank the repeats forming a single structural unit (Fig. 21A–D). The domain features a concave surface lined by β-strands and an outer surface formed by α-helices running roughly antiparallel to the strands. The individual repeats of the LRR correspond to β-α units analogous to those originally reported for the ribonuclease inhibitor structure (Kobe and Deisenhofer, 1993). Each repeat consists of 26 amino acids with the consensus sequence L_{IN}-x_{OUT}-L_{IN}-x_{OUT}-L_{IN}-x_{OUT}-x_{OUT}-N_{IN} (where x, IN and OUT are as defined above). As in other LRR structures (Kobe and Deisenhofer, 1995), the consensus sequence of the repeats maps onto the β-strands, which are relatively short and followed by a characteristic turn. The leucine side chains associated with the β-strands are involved in stacking interactions that are crucial for maintaining the hydrophobic core of the domain. The asparagine residues in the following turn hydrogen-bond to the backbone of the β-strands and contribute to the flatness of the β-sheet, which lacks the typical twist. The variable residues of the consensus sequence are instead exposed to solvent.

A database search with the program DALI (Holm and Sander, 1993) showed that among the known LRR structures, the TAP 203–362 domain shares remarkable similarity with the spliceosomal protein U2A', which contains five tandem LRRs (Price et al., 1998). The similarity extends beyond the expected structural homology at the LRR motifs. In particular, the C-terminal region flanking the hydrophobic residues of the last LRR motif of

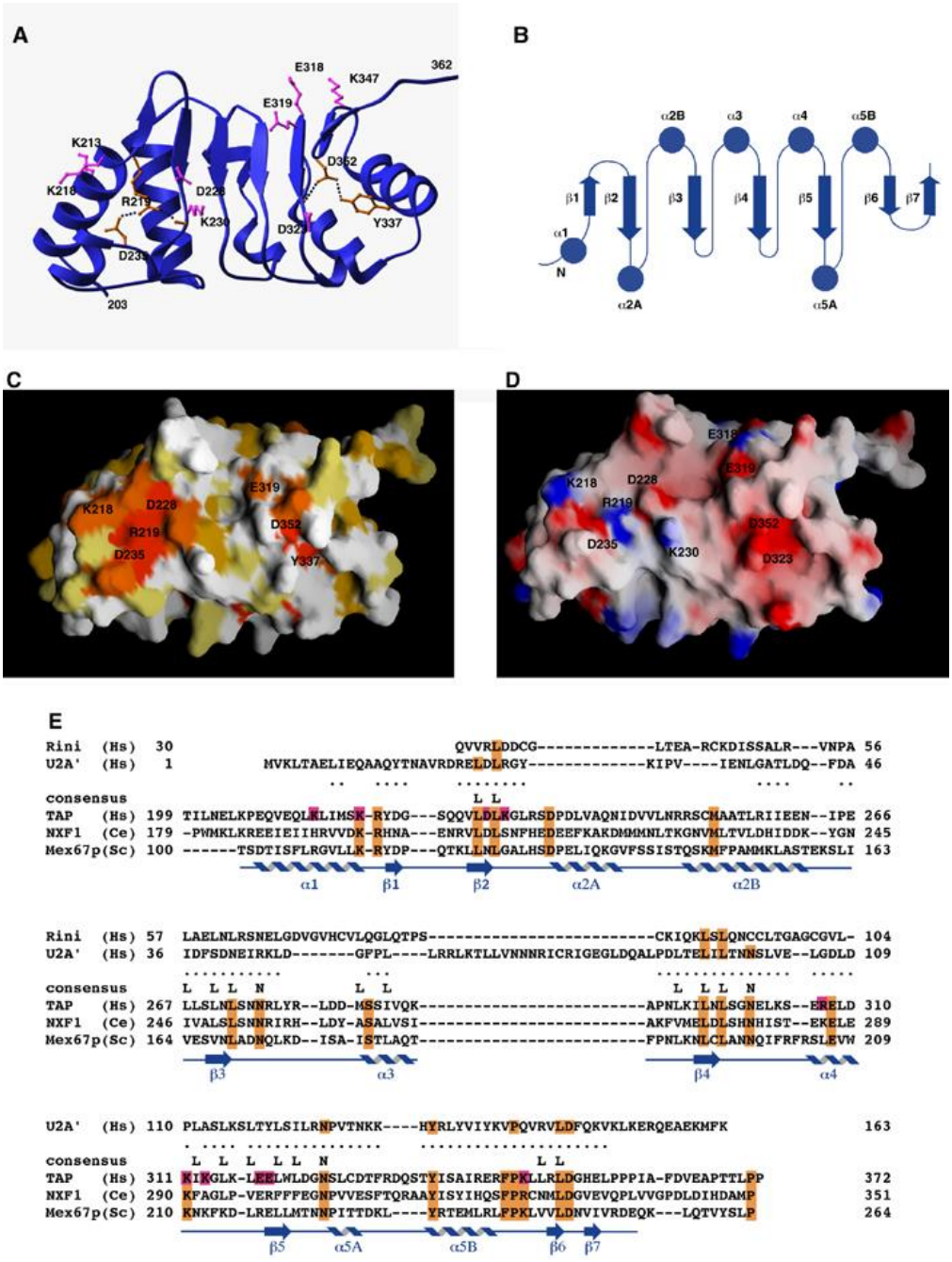


Fig. 19

The LRR domain of human TAP. (**A**) Ribbon diagram showing the concave ß-sheet surface of the domain. (**B**) Topology diagram of the secondary structural elements, with their sequential numbering indicated. (**C**) Surface representation of the concave ß-sheet face of the LRR domain shown in a similar orientation to (A). The surface is colored according to sequence conservation within the NXF protein family, with a color gradient ranging from white (no sequence identity) to dark orange (100% sequence identity). (**D**) Electrostatic surface representation of the concave ß-sheet face of the LRR domain. Positive potential is colored blue and negative potential red. (**E**) Structure-based amino acid sequence alignment of the LRR domain of human TAP with the ribonuclease inhibitor (Rini) and the spliceosomal protein U2A'. The dots indicate residues in TAP that lie within 3.0 Å of corresponding LRR $C\alpha$ atoms after optimal superposition of the domains. Highlighted in orange are conserved residues in the NXF family, and pink represents residues that have been mutated and analyzed for function.

TAP is anchored by the same set of hydrogen-bonding interactions found in U2A'. An invariant aspartic acid (Asp352 in TAP, 146 in U2A') engages its side-chain carbonyl oxygens in two hydrogen bonds with a main-chain nitrogen and with a tyrosine side chain (Tyr337 in TAP, 131 in U2A') (Fig. 19A and E). The N-terminal flanking region similarly forms a hydrophilic shield on the buried hydrophobic residues of the first repeat and is structured by a salt bridge between two conserved and buried charged residues (Arg219 and Asp235) that also hydrogen-bond to a main-chain carbonyl oxygen (Fig. 19A). Helix α2A in the N-terminal flanking region (Fig. 19B) is the most variable region of the molecule, displaying some conformational flexibility when comparing the six independent molecules in the crystals.

3.2.4 General and specific RNA-binding activity of TAP domains

The general RNA-binding ability of TAP fragments was tested by an electrophoretic gel-mobility retardation assay using purified recombinant protein fragments and a labeled, non-specific single-stranded RNA probe derived from pBluescribe (Fig. 20A). Binding of the RNA probe to the TAP fragment 96–372 is as efficient as to the full-length protein (Fig. 20A, lanes 2–6). The 96–372 fragment includes the RNP domain (119–198), the LRR domain (203–362) and an N-terminal region whose sequence (Fig. 18C) and lack of electron density suggest flexibility in the absence of CTE-RNA. Addition of seven residues N-terminal to the minimal CTE-binding fragment 102–372 (Braun et al., 1999) allows binding to the CTE-RNA independently of the presence of the glutathione S-transferase (GST) fusion protein used for expression and purification (data not shown), and was therefore used for biochemical experiments. Upon testing the two domains separately, binding to the RNA probe is observed

only with the N-terminal RNP-containing fragment (96–198) (Fig. 20A, lanes 7–10), and not with the C-terminal LRR-containing fragment (199–372) (Fig. 20A, lanes 11–14). The N-terminal region of human TAP upstream of the RNP domain (residues 1–118) does not exhibit detectable RNA-binding capabilities (Fig. 20B, lane 5). The non-canonical RNP of TAP is therefore a *bona fide* RNA-binding domain despite having divergent sequence motifs. Its basal level of RNA-binding activity is not dependent simply on the presence of positively charged surface residues, as the LRR domain also features an electropositive surface area (Fig. 19D) but no

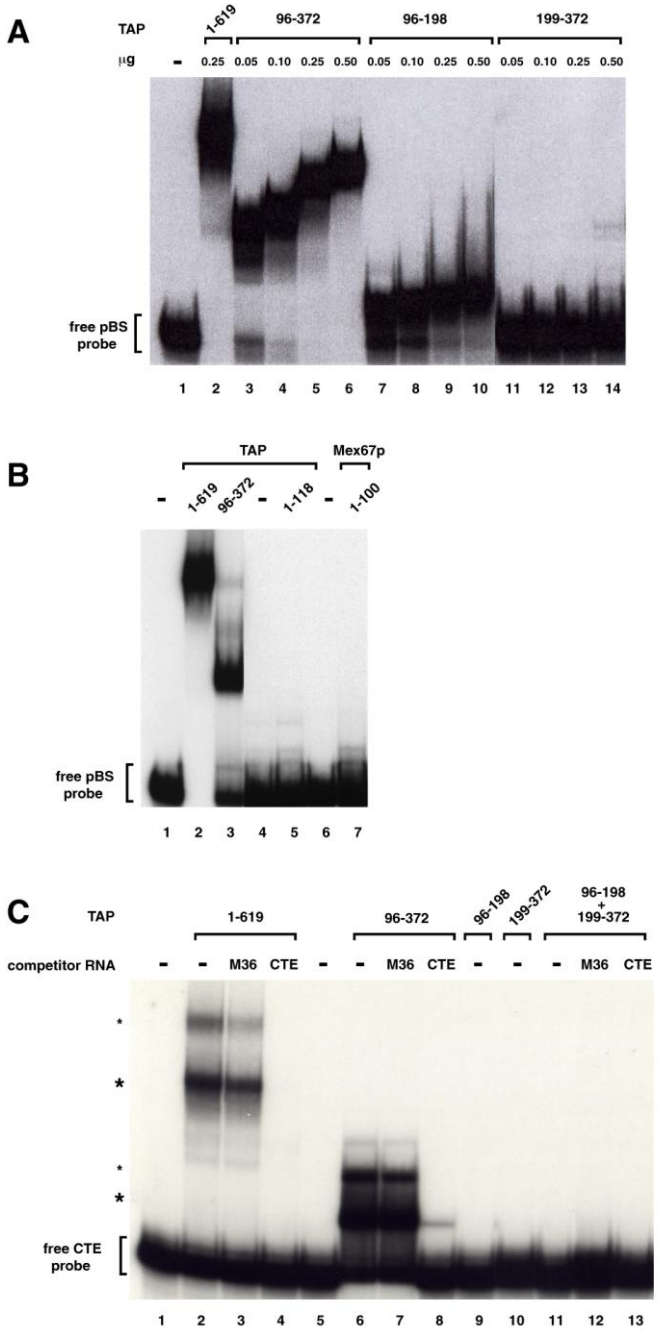


Fig. 20
Electrophoretic mobility retardation assay shows the RNA-binding properties of TAP fragments. (**A**) The RNP domain exhibits general RNA binding affinity, while the LRR domain does not. (**B**) The N-terminal region of human TAP upstream of the RNP domain does not exhibit RNA-binding capabilities. (**C**) The RNP domain requires the LRR domain in *cis* for specific binding to the CTE RNA. (Big asterisks show the TAP–CTE RNA complexes, small asterisks may represent two molecules of TAP bound to the CTE-RNA.)

general RNA binding can be detected in this case.

Sequence comparison of members of the NXF family indicates that the RNP is one of the least conserved domains within the protein. It is present in *C.elegans* and *Drosophila melanogaster* NXF1, as well as in all human NXF proteins (Fig. 18C; Herold et al., 2000), but it might not be present in *Saccharomyces cerevisiae* Mex67p. The region upstream of the LRR domain in *S.cerevisiae* Mex67p is half the size of that in human TAP and this N-terminal 100-residue fragment of Mex67p is unable to bind to the general RNA probe (Fig. 20B, lane 7). Additional protein regions are likely to be required for the *in vitro* RNA binding reported for the *S.cerevisiae* Mex67p–Mtr2p complex (Santos-Rosa et al., 1998).

The specific CTE-RNA-binding ability of TAP fragments was tested using a CTE probe in an *in vitro* binding assay similar to that described above. The full-length protein and the 96–372 fragment show high affinity in CTE recognition, both forming a complex even in the presence of high concentrations of competitor RNA (Fig. 20C, lanes 2–8). Formation of these complexes is specifically competed by the presence of unlabeled CTE-RNA but not by M36 RNA, a CTE derivative that does not bind TAP (Grüter et al., 1998) (Fig. 20C, lanes 3, 4, 7 and 8). Under the same conditions, the isolated RNP or LRR domains do not specifically interact with the CTE-RNA (Fig. 20C, lanes 9 and 10). Analogously, no specific binding is observed when the two domains are added in *trans* (Fig. 20C, lanes 11–13). It is unclear at present whether the requirement of the two domains in *cis* might be due to an effective increase in their local concentration when covalently attached to each other or to sequence-specific RNA contacts of the amino acids in the linker that connects them. On the shifts with the full length TAP two complexes are seen. The upper complexes (small asterisks) may represent two molecules of TAP bound to the CTE-RNA, according to the previous suggestion that in the CTE-RNA both loops are possible interaction sites (Tabernero et al., 1996).

The requirement of both the RNP and LRR domains of TAP for cooperative binding to CTE-RNA, their RNA-binding abilities and their structural homology to U2B" and U2A', all point to a remarkable similarity with the spliceosomal U2B"–U2A' complex (Fig. 21). The spliceosomal U2B" RNP protein is not able to recognize specifically the cognate U2 snRNA

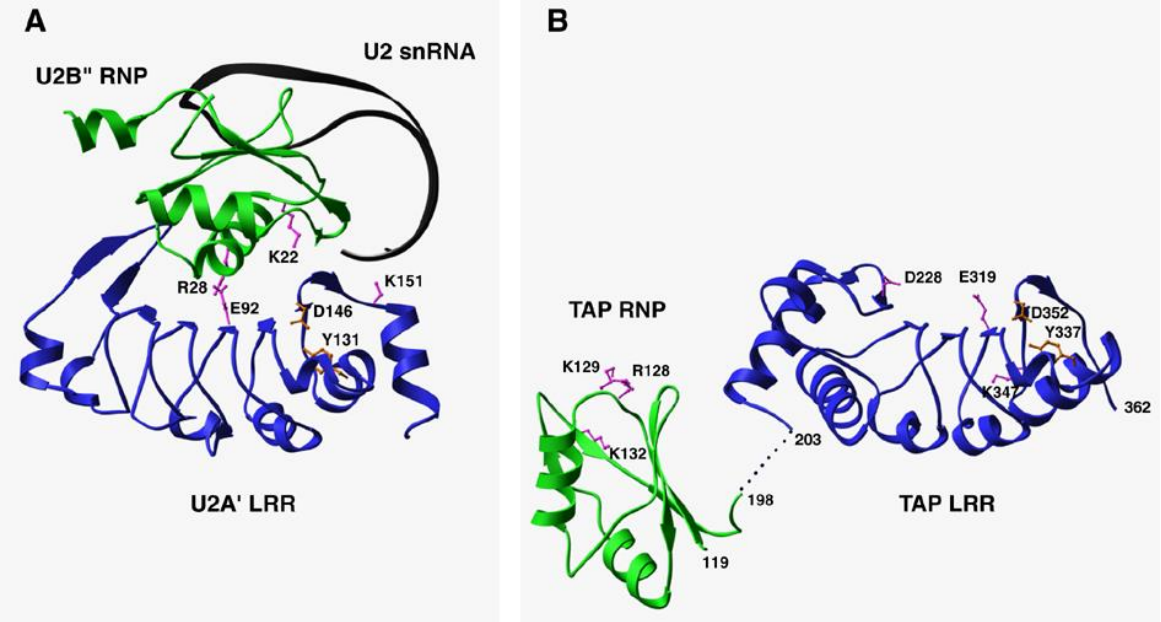


Fig. 21
Structural homology between TAP and the spliceosomal U2B"–U2A' complex. (**A**) The U2B" RNP domain (green) and the U2A' LRR domain (blue) of the spliceosomal complex with the cognate RNA (in black). **(B**) The RNP domain (green) and LRR domain (in blue) of TAP in one copy in the asymmetric unit of the crystals. In the absence of RNA, the relative orientation of the two domains is likely to be due to crystal packing.

on its own, as instead is the case for the recognition of U1 snRNA by the homologous U1A RNP protein (Scherly et al., 1990a,b). Although the U2A' LRR protein does not detectably bind to RNA *in vitro* (Scherly et al., 1990a), its association with U2B" is essential in increasing the protein–RNA contacts (Price et al., 1998). The concave β-sheet surface of the LRR domain associates with the α1 helix of the RNP domain, allowing the U2 snRNA to bind both to the β-sheet platform of the RNP and to a positively charged patch at the C-terminal edge of the U2A' LRR. A few intersubunit salt bridges are critical for the interaction between U2B" and U2A', which is impaired even by conservative mutations of the charged residues involved in these ionic interactions (such as Lys to Arg or Asp to Glu) (Scherly et al.,

1990a). To determine whether the RNP and LRR domains of TAP are involved in a similar mode of macromolecular recognition with the cognate CTE-RNA, selected site-directed mutants were tested in *in vitro* and *in vivo* functional assays as described below.

3.2.5 Identification of macromolecular interaction surfaces

The surfaces of the RNP and LRR domains of TAP established from the structure were probed at multiple positions to identify potential sites of macromolecular interactions.

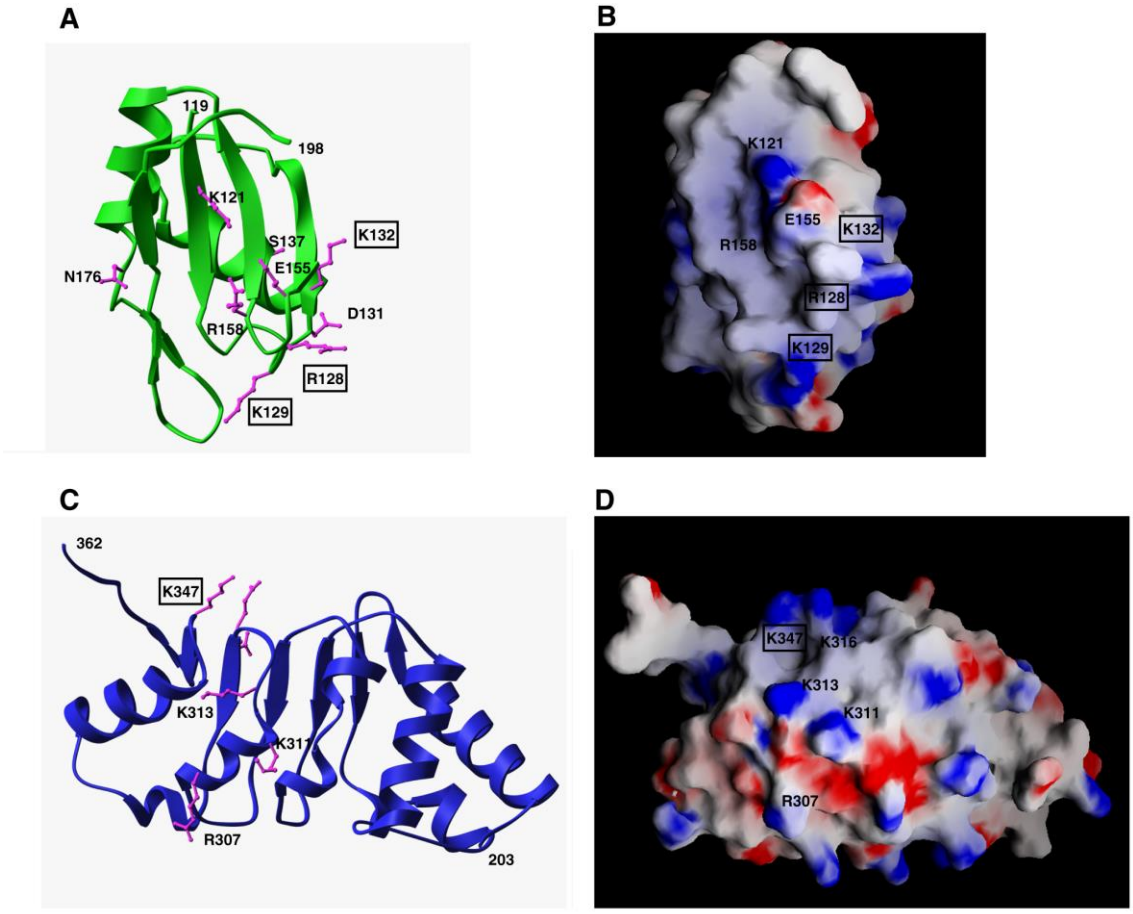


Fig. 22
Identification of putative macromolecular interaction surfaces. (**A**) Ribbon diagram of the RNP domain in green with the side chains of mutated residues in pink. (**B**) Surface representation of the RNP domain in a similar orientation to (A). The surface is colored according to electrostatic potential, with blue indicating positively charged areas and red indicating negatively charged areas. (**C**) Structure of the LRR domain viewed towards the convex α-helical surface. Residues at this surface that have been mutated are shown in pink.(**D**) Electrostatic surface of the convex outer surface of the LRR domain viewed as in (C), and colored as in (B).

Site-directed mutants were selected according to their amino acid type, phylogenetic conservation and location on the surface of the domains (Fig. 19 and 22). The mutant proteins were synthesized *in vitro* and their ability to bind CTE-RNA was assayed in the context of the full-length protein by an electrophoretic gel mobility retardation assay (Fig. 23A). Interestingly, the CTE–TAP interaction is disrupted only by a few reverse-charged mutations in the RNP and LRR domains. More conservative substitutions of these residues to alanines does not detectably affect CTE binding (Fig. 23A, compare lanes 4 and 5, or 21 and 22), suggesting that recognition is more likely mediated by multiple rather than a few crucial interactions. Although drastic, the effects of reverse-charged substitutions are specific, since at other positions they do not affect binding (Fig. 23A, lanes 14 and 20).

The positively charged edge of the RNP β-sheet platform is essential for CTE binding. A double mutation of Arg128 and Lys129 (Fig. 22A and B) to negatively charged residues abolishes CTE recognition (Fig. 23A, lane 5). In the structure of U2B", a basic residue (Lys22) is similarly located in loop 1 (Fig. 18C and 21) and interacts with RNA phosphate groups (Price et al., 1998). Deleterious effects in CTE binding are also produced by a reverse-charge mutation of Lys132 (Fig. 23A, lane 7) in the RNP α1 helix pointing towards the side of the β-sheet platform (Fig. 22A and B). No effect is observed when substituting Ala133 and Ser137 (Fig. 23A, lane 8), which from helix α1 point towards the back of the β-sheet platform. Ala133 is in the equivalent structural position to Arg28 in U2B" (Fig. 18C), a crucial residue for the interaction with U2A' (Fig. 21). It is unclear in this context whether Lys132 in the TAP RNP domain is involved in protein–RNA interactions or protein–protein interactions with the LRR domain.

Unexpectedly, substitution of several amino acids at the concave β-sheet surface of the LRR domain does not prevent CTE-RNA recognition. In particular, a reverse-charge mutation of Asp228 results in equally strong CTE binding (Fig. 23A, lane 14).

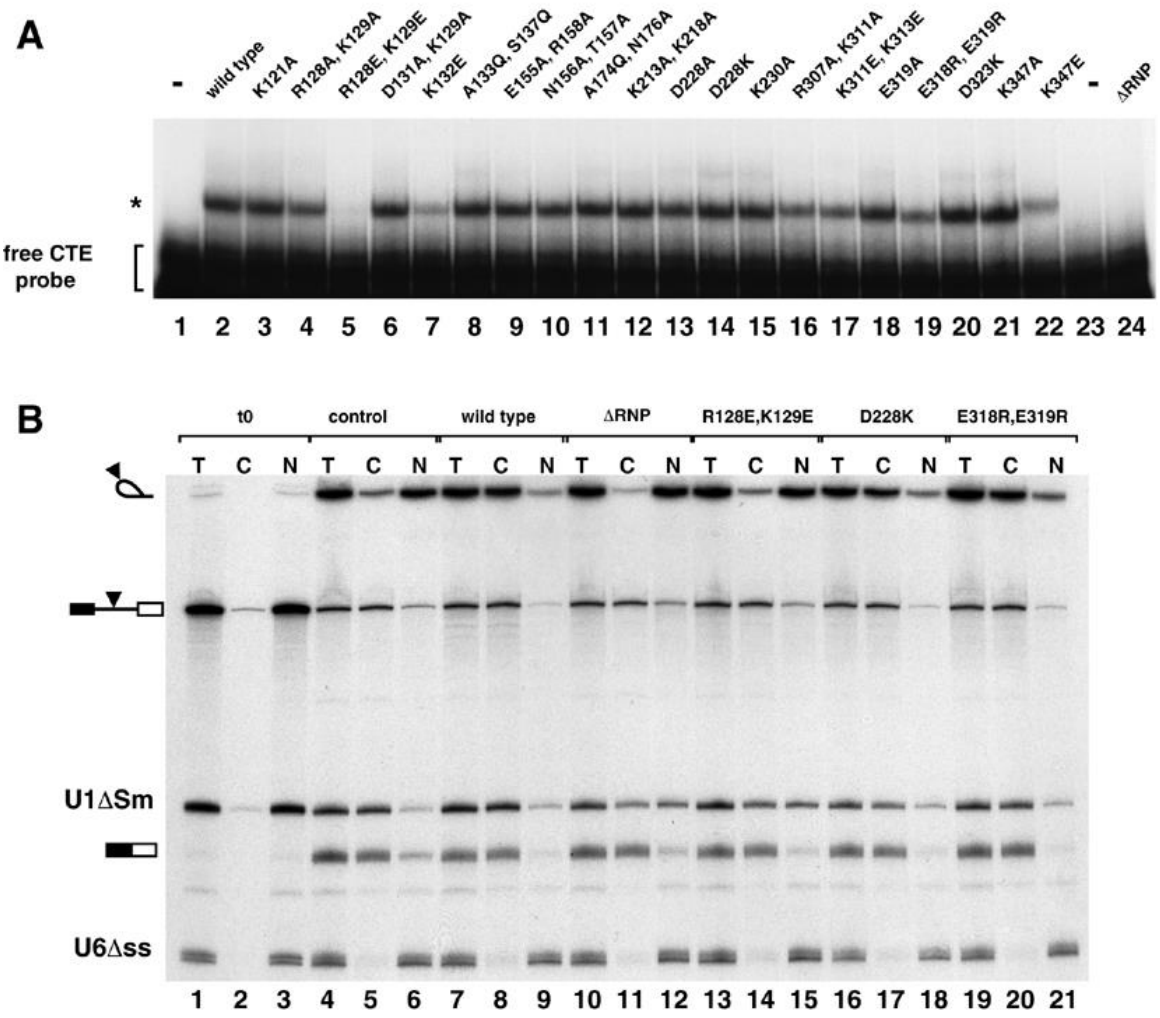


Fig. 23
In vitro and *in vivo* functional studies with structure-based mutants. (**A**) A gel-mobility retardation assay was performed with a labeled CTE RNA probe and *in vitro* translated mutant proteins. (**B**) *Xenopus* oocyte nuclei were injected with a mixture of *in vitro* transcribed ^{32}P-labeled U1ΔSm RNA, U6Δss RNA and a precursor RNA containing the SRV-1 CTE inserted at the intron (Ad-CTE). Purified recombinant GST–TAP and various TAP mutants (8 μM) were included in the injection mixtures as indicated. RNA samples from total oocytes (T), cytoplasmic (C) and nuclear (N) fractions were collected immediately after injection (lanes 1–3) or 2.5 h after injection (lanes 4–21). Products of the splicing reaction were resolved on 10% acrylamide–7 M urea denaturing gels.

The solvent-exposed Asp228 is conserved within the NXF family and is located in the N-

terminal region flanking the first repeat (Fig. 19A, C and D). The two neighboring conserved residues Arg219 and Asp235 (Fig. 19C) play a structural role and their mutation was therefore avoided not to compromise the structural integrity of the protein. No effect in CTE recognition is observed upon mutating other surface residues (Fig. 23A, lanes 12, 15 and 20), while a slight decrease in affinity is produced by the reverse-charge of Glu318 and Glu319 (Fig. 23, lane 19) at the edge of the concave surface (Fig. 19). The most dramatic effect within the LRR domain is caused by mutation of the well-conserved Lys347 to an acidic residue (Fig. 23, lane 22). Lys347 is located in the C-terminal region flanking the last repeat (Fig. 19A or 21C) and lies within an electropositive patch at the edge of the outer convex surface (Fig. 22D). The importance of this region is reinforced by the conservation of the interactions that play a structural role in TAP (Asp352, Tyr337) and remarkably also in U2A' (Asp146, Tyr131). In the C-terminal region of U2A', although not in an identical position, a positively charged residue (Lys151) is crucial for the recognition of the U2 snRNA stem–loop (Fig. 22).

Selected mutants that either do or do not show *in vitro* CTE-binding abilities were tested for stimulation of CTE-dependent nuclear export *in vivo*. Purified recombinant proteins were injected into Xenopus oocyte nuclei together with a mixture of three labeled RNAs (Saavedra et al., 1997). This mixture consisted of an adenovirus-derived precursor mRNA bearing the CTE in the intron, U6Δss RNA and U1ΔSm RNA. Δss RNA is neither imported nor exported from the nucleus and serves as a control for accurate nuclear injection (Vankan et al., 1992). U1ΔSm is exported by a different pathway than the CTE-RNA (Pasquinelli et al., 1997; Saavedra et al., 1997), and serves as a control for the specificity of the recombinant proteins. Immediately after injection all RNAs were found in the nuclear fraction (Fig. 23B, lanes 1–3). Following 2.5 h of incubation, splicing of the precursor RNA was complete and 30% of the resulting CTE-bearing intron lariat was found in the cytoplasmic fraction

(Fig. 23B, lanes 4–6). Coinjection in oocytes of wild-type TAP stimulated the export of the intron lariat, ~95% of which reached the cytoplasm (Fig. 23B, lanes 7–9). The mutant proteins ΔRNP and R128E,K129E, which are unable to bind to CTE *in vitro*, did not stimulate the export of the intron lariat (Fig. 23B, lanes 11, 12, 14 and 15). In contrast, efficient export to the cytoplasm was observed when injecting the mutant proteins D228K and E318R,E319R (Fig. 23B, lanes 17, 18, 20 and 21), which retain CTE-binding activity *in vitro*. Export of the spliced Ad mRNA was not affected (Fig. 21B, compare lanes 7–12 with 4–6) while we observed some non-specific effect on U1ΔSm RNA export. Thus, the *in vivo* behavior of the mutants tested is in agreement with the *in vitro* results.

4. CONCLUSIONS

The export of CTE-bearing retroviral RNAs to the host cytoplasm is achieved by their direct interaction with a fragment of the cellular protein TAP. The aim of this work is to obtain molecular insights in how the TAP-CTE interaction is achieved. To this end we solved the X-ray structure of the minimal CTE interaction domain of TAP, and used this structural information to probe the TAP interaction surface by mutagenesis.

To summarize the results, we can conclude that the TAP minimal CTE binding domain includes two independent globular domains. The N-terminal domain folds and functions as a *bona fide* RNP domain, despite lacking the canonical conserved sequence motifs. The C-terminal domain is an LRR-containing protein that does not show general RNA-binding activity but is required for specific binding to CTE-RNA. The two independent domains have similar structural and biochemical properties to the U2B'' and U2A' components of the spliceosomal complex. Functional studies with structure-based mutants indicate that positively charged residues at the β-sheet platform of the RNP are likely to be involved in RNA binding, similarly to U2B'' and to canonical RNP proteins in general. A residue identified on a helix at the back of the RNP platform plays an important role, either in RNP-LRR or RNP-RNA interactions. A positively charged patch at the outer convex surface of the LRR domain might also be involved in RNA binding, conferring specificity to CTE-RNA recognition similarly to the positively charged patch on the U2A' surface in U2 snRNA stem recognition. Despite these similarities, it is conceivable that the roles of the TAP and U2A' LRRs in specific RNA binding are at least partly divergent. The most obvious difference is that recognition of the CTE-RNA requires the RNP and LRR domains to be present *in cis* in the same polypeptide together with the N-terminal flexible region, while U2B'' and U2A' function as separate proteins. Furthermore, substitution of residues along the

conserved concave surface of the LRR domain of TAP does not disrupt CTE binding *in vitro* or its export *in vivo*, as would be expected for a U2B''-U2A' mode of protein-protein interaction.

Molecular details of how the CTE-RNA interacts with TAP can only obtained by determining the crystal structure of the complex. Given the size of the RNA, this is a long-term project which goes beyond the scope and time frame of this work. As part of this goal, we identified the smallest CTE-RNA fragment that is suitable for crystallization and that is still active in TAP binding. We showed that a double hammerhead ribozyme strategy can be used to transcribe RNA with homogenous 3' ends and without any unspecific nucleotides at the 5' end. These results are currently being used to pursue structural studies of the CTE-RNA TAP complex.

5. REFERENCES

Abdel-Meguid, S.S., Moore, P.B. and Steitz, T.A. (1983)
Crystallization of a ribonuclease-resistant fragment of *Escherichia coli* 5S ribosomal RNA and its complex with protein L25. *J. Mol. Biol.* **171,** 207.

Arts, G.-J., Fornerod, M. and Mattaj, I.W. (1998)
Identification of a nuclear export receptor for tRNA. *Curr. Biol.* **8**, 305-314.

Bachi, A., Braun, I.C., Rodrigues, J.P., Panté, N., Ribbeck, K., von Kobbe, C., Kutay, U., Wilm, M., Görlich, D., Carmo-Fonseca, M., and Izaurralde, E. (2000)
The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA* **6**, 136-158.

Becker, J., Melchior, F., Gerke, V., Bischoff, F.R., Ponstingl, H. and Wittinghofer, A. (1995)
RNA1 encodes a GTPase-activating protein specific for Gsp1, the Ran/TC4 homologue of *Saccharoyces cerevisiae. J. Biol. Chem.* **270,** 11860-11865.

Bischoff, F.R., Klebe, C., Kretschmer, J., Wittinghofer, A. and Postingl, H. (1994)
RanGAP1 includes GTPase activity of nuclear ras-related Ran. *Proc. Natl. Acad. Sci. USA* **91,** 2587-2591.

Black, B.E., Lévesque, L., Holaska, J.M., Wood, T.C. and Paschal, B. (1999)
Identification of an NTF2-related factor that binds Ran-GTP and regulates nuclear protein export. *Mol. Cell. Biol.* **19**, 8616-8624.

Braun, I.C., Rohrbach, E., Schmitt, C. and Izaurralde, E. (1999)
TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. *EMBO J.* **18**, 1953-1965.

Braun, I.C., Herold, A., Rode, M., Conti, E. and Izaurralde, E. (2001)
Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export. *J. Biol. Chem.* **276**, 20536-20543.

Bray, M., Prasad, S., Dubay, J.W., Hunter, E., Jeang, K.T., Rekosh, D. and Hammarskjöld ML. (1994)
A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc. Natl. Acad. Sci.* USA **91,** 1256-1260.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998)
Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.

Burd, C.G. and Dreyfuss, G. (1994)
Conserved structures and diversity of functions in RNA-binding proteins. *Science*, **265**, 615–621

Chang, D.D. and Sharp, P.A. (1989)
Regulation by HIV Rev depends upon recognition of splice sites. *Cell* **59,** 789-795.

Cherezov, V., Fersi, H. and Caffrey, M. (2001)
Crystallization Screens: Compatibility with the Lipidic Cubic Phase for *in Meso* Crystallization of Membrane Proteins. *Biophys J.* **81,** 225-242.

CCP4 (1994)
Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50,** 760-776.

Chi, N.C., Adam, E.J. and Adam, S.A. (1995)
Sequence and characterization of cytoplasmic nuclear protein import factor p97. *J. Cell Biol.* **130**, 265-274.

Clouse, K.N., Luo, M.J., Zhou, Z. and Reed, R. (2001)
A Ran-independent pathway for export of spliced mRNA. *Nat. Cell Biol*. **3,** 97-99.

Cullen, B.R., (1998)
Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249**, 203-210.

Deo, R.C., Bonanno, J.B., Sonenberg, N. and Burley, S.K. (1999)
Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845.

Dock, A-C., Lorber, B., Moras, D., Pixa, G., Thierry, J-C. And Giegé, R. (1984)
Crystallization of transfer ribonucleic acids. *Biochimie* **66,** 179.

Dingwall, C., Sharnick, S.V. and Laskey, R.A. (1982)
A poypeptide domain that specifies migration into the nucleus. *Cell* **30,** 449-458.

Dingwall, C. and Laskey, R.A. (1991)
Nuclear targeting sequences – a consensus? *Trends Biol. Sci.* **16**, 178-181.

Ernst, R.K., Bray, M., Rekosh, D. and Hammarskjold, M.-L. (1997a)
A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Mol. Cell. Biol.* **17**, 135-144.

Ernst, R.K., Bray, M., Rekosh, D. and Hammarskjold, M.-L. (1997b)
Secondary structure and mutational analysis of the Mason–Pfizer monkey virus RNA constitutive transport element. *RNA* **3**, 210–222.

Felber, B.K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T. and Pavlakis, G.N. (1989)
Rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci.* USA **86,** 1495-1499.

Feldherr, C.M., Kallenbach, E. and Schultz, N. (1984)
Movement of karyophilic protein through the nuclear pores of oocytes. *J. Cell Biol.* **99,** 2216-2222.

Ferré d'Amare, A.R. and Doudna, J.A. (1996)
Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res.* **24,** 977-978.

Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W. and Lührmann, R. (1995)
The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82,** 475-483.

Fischer, U., Meyer, S., Teufel, M., Heckel, C., Lührmann, R. and Rautmann, G. (1994)
Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. *EMBO J* **13**, 4105-4112.

Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I.W. (1997)
CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90,** 1051-1060.

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M. and Nishida, E. (1997)
CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308-311.

Görlich, D., Kostka, S., Kraft, R., Dingwall, C., Laskey, R.A., Hartmann, E. and Prehn, S., (1995)
Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. *Curr. Biol.* **5**, 383-392.

Görlich, D. and Mattaj, I.W. (1996)
Nucleocytoplasmic Transport. *Science* **271**, 1513-1519.

Görlich, D. (1997)
Nuclear protein import. *Curr. Opin. Cell Biol*. **9(3),** 412-419.

Görlich, D. and Kutay, U. (1999)
Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 607-660.

Grüter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., and Izaurralde, E. (1998)
TAP, the human homologue of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell* **1**, 649-659.

Hammarskjöld, M.-L, Heimer, J., Hammarskjöld, B., Sangwan, I., Albert, L. and Rekosh, D. (1989)
Regulation of human immunodeficiency virus env expression by the rev gene product. *J. Virol.* **63,** 1959-1966.

Hammarskjöld,M.-L. and Dahlberg, J.E. (1997)
The constitutive transport element (CTE) of Mason–Pfizer monkey virus (MPMV) accesses an RNA export pathway utilized by cellular messenger RNAs. *EMBO J.* **16**, 7500–7510.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999)
Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585.

Herold, A., Suyama, M., Rodrigues, J.P., Braun, I., Kutay, U., Carmo-Fonseca, M., Bork, P. and Izaurralde, E. (2000)
TAP/NXF1 belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Mol. Cell. Biol.* **20**, 8996-9008.

Hieda, M., Tachibana, T., Yokoya, F., Kose, S., Imamoto, N. and Yoneda, Y. (1999)
A monoclonal antibody to the COOH-terminal acidic portion of ran inhibits both the recycling of ran and nuclear protein import in living cells. *J. Cell Biol.* **144,** 645-655.

Holm, L. and Sander, C. (1993)
Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

Hutchins, C.J., Rathjen, P.D., Forster, A.C. and Symons, R.H. (1986)
Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucl. Acids Res.* **14,** 3627-3640.

Imamoto, N., Shimamoto, T., Kose, S., Takao, T., Tachibana, T., Matsubae, M., Sekimoto, T., Shimonishi, Y. and Yoneda, Y. (1995)
The nuclear pore-targeting complex binds to nuclear pores after association with karyophile. *FEBS Lett.* **368**, 415-419.

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

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C. and Mattaj, I.W. (1995)
A cap-binding protein complex mediating U snRNA export. *Nature* **376,** 709-12

Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I.W. and Görlich, D. (1997)
The asymmetric distribution of the constituents of the Ran system is esential for transport into and out the nucleus. *EMBO J.* **16,** 6535-6547.

Jancarik, J. and Kim, S-H. (1991)
Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24,** 409-411.

Jarmolowski, A., Boelens, W.C., Izaurralde, E. and Mattaj, I.W. (1994)
Nuclear export of different classes of RNA is mediated by specific factors. *J. Cell Biol.* **124,** 627-635.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991)
Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A.* **47**, 110-119,

Jovine, L., Hainzl, T., Oubridge, C., Scott, W.G., Li, J., Sixma, T.K., Wonacott, A., Skarzynski, T. And Nagai, K. (2000)
Crystal structure of the ffh and EF-G binding sites in the conserved domain IV of Escherichia coli 4.5S RNA. *Structure Fold Des.* **15**, 527-40.

Kadowaki, T., Hitomi. M., Chen, S. and Tartakoff, A.M. (1994)
Nuclear mRNA accumulation causes nucleolar fragmentation in yeast mtr2 mutant. *Mol. Biol. Cell* **5,** 1253-1263.

Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984)
A short amino acid sequence able to specify nuclear location.
Cell **39**, 499-509.

Kang, Y., Bogerd, H.P., Yang, J. and Cullen, B.R. (1999)
Analysis of the RNA binding specificity of the human tap protein, a constitutive transport element-specific nuclear RNA export factor. *Virology* **262,** 200-209.

Katahira, J., Strasser, K., Podtelejnikov, A., Mann, M., Jung, J.U. and Hurt, E. (1999)
The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *EMBO J.* **18**, 2593-2609.

Kataoka, N., Diem, M.D., Kim, V.M., Yong, J. and Dreyfuss, G. (2001)
Magoh, a human homolog of Drosophila mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J.* **20,** 6424-6433.

Kataoka, N., Yong, J., Kim, V.M., Velazquez, F., Perkinson, R.A., Wang, F. and Dreyfuss, G. (2000)
Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* **6,** 673-682.

Kim, V.N., Kataoka, N. and Dreyfuss, G. (2001)
Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* **293,** 1832-1836.

Klebe, C., Prinz, H., Wittinghofer, A. and Goody, R.S. (1995)
The kinetic mechanism of Ran-nucleotide exchange catalyzed by RCC1. *Biochemistry* **34,** 12543-12552.

Kobe, B. and Deisenhofer, J. (1993)
Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**, 751–756.

Kobe, B. and Deisenhofer, J. (1995)
Proteins with leucine-rich repeats. *Curr. Opin. Struct. Biol.* **5**, 409–416.

Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R. and Görlich, D. (1997)
Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell* **90,** 1061-1071.

Kutay, U., Lipowsky, G., Izaurralde, E., Bischoff, R.F., Schwarzmaier, P., Hartmann, E. and Görlich, D. (1998)
Identification of a tRNA-specific export receptor. *Mol. Cell* **1**, 359-369.

Lanford, R.E. and Butel, J.S. (1984)
Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell* **37**, 801-813.

Laemmli, U.K. (1970)
Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680-685.

Legrain, P. and Rosbash, M. (1989)
Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **57,** 573-583.

Le Hir, H., Moore, M.J. and Maquat, L.E. (2000)
Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* **14,** 1098-1108.

Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. (2001)
The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20,** 4987-4997.

Liker, E., Fernandez, E., Izaurralde, E. and Conti, E. (2000)
The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. *EMBO J.* **19,** 5587-5598.

Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2001)
Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293,** 1836-1839.

Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2000)
Human Upf proteins target an mRNA for nonsence-mediated decay when bound downstream of a termination codon. *Cell* **103,** 1121-1131.

Malim, M.H. and Cullen, B.R. (1993)
Rev and the fate of pre-mRNA in the nucleus: Implications for the regulation of RNA processing in eukryotes. *Mol. Cell. Biol.* **13,** 6180-6189.

Malim, M.H., Hauber, J., Le, S-Y., Maizel, J.V. and Cullen, B.R. (1989)
The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338,** 254-257.

Malim, M.H., Tiley, L.S., McCarn., D.F., Rusche., J.R., Hauber, J. and Cullar, B.R. (1990)
HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **60,** 675-683.

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

Mattaj, I.W. and Englmeier, L. (1998)
Nucleocytoplasmic transport: the soluble phase. *Annu Rev. Biochem.* **67**, 265-306.

Mattaj, I.W. and Conti, E. (1999)
Cell biology: Snail mail to the nucleus. *Nature* **399,** 208 – 210.

Matthews, B.W. (1968)
Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

Meroueh, M. and Chow, C. (1999)
Thermodynamics of RNA hairpins containing single internal mismatches. *Nucleic Acids Res.* **27,** 1118-1125.

Meyer, B.E. and Malim, M.H. (1994)
The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev.* **8,** 1538-1547.

Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990)
Crystal structure of the RNA binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515–520.

Nakielny, S., Siomi, N.C., Siomi, H., Michael, W.M., Pollard, V. et al. (1996)
Transportin: nuclear transport receptor of a novel nuclear protein import pathway. *Exp. Cell Res.* **229**, 261-266.

Nakielny, S. and Dreyfuss, G. (1999)
Transport of proteins and RNAs in and out of the nucleus. *Cell* **99**, 677-690.

Neville, M., Stutz, F., Lee, L., Davis, L.I. and Rosbash, M. (1997)
The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.* **7,** 767-775.

Otwinowski, Z. and Minor, W. (1997)
Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994)
Crystal structure at 1.92Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.

Parks, T.D., Leuther, K.K., Howard, E.D., Johnston, S.A. and Dougherty, W.G. (1994)
Release of proteins and peptides from fusion proteins with a recombinant plant virus proteinase. *Anal. Biochem.* **216**, 413-417.

Pasquinelli, A.E., Ernst, R.K., Lund, E., Grimm, C., Zapp, M.L., Rekosh, D., Hammarskjold, M.-L. and Dahlberg, J.E. (1997)
The constitutive transport element (CTE) of Mason–Pfizer monkey virus (MPMV) accesses an RNA export pathway utilized by cellular messenger RNAs. *EMBO J.* **16**, 7500–7510.

Pemberton, L.F., Blobel, G. and Rosenblum, J.S. (1998)
Transport routes through the nuclear pore complex. *Curr. Opin. Cell Biol.* **10,** 392-399.

Pollard, V.W., Michael, W.M., Nakielny, S., Siomi, M.C., Wang, F. and Dreyfuss, G. (1996)
A novel receptor-mediated nuclear protein import pathway. *Cell* **86**, 985-994.

Price, S.R., Ito, N., Oubridge, C., Avis, J.M. and Nagai, K. (1995)
Crystallization of RNA-protein complexes I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249,** 398-408.

Price, S.R., Evans, P.R. and Nagai, K. (1998)
Crystal structure of the spliceosomal U2B"–U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650.

Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R. and Bruening, G. (1986)
Autolytic processing of dimeric plant satellite RNA. *Science* **231,** 1577-1580.

Radu, A., Blobel, G. and Moore, M.S. (1995)
Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. *Proc. Natl. Acad. Sci. USA* **92**, 1769-1773.

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

Reichelt, R., Holzenburg, A., Buhle, E.L. Jr., Jarnik, M., Engel, A., Aebi, U. (1990)
Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *J. Cell Biol.* **110**, 883-894.

Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991)
Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615-623.

Rodrigues, J.P., Rode, M., Gatfield, D., Blencowe, B., Carmo-Fonseca, M. and Izaurralde, E. (2001)
REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc. Natl. Acad. Sci*. **98**, 1030-1035.

Rout, M.P. and Blobel, G. (1993)
Isolation of the yeast nuclear pore complex. *J. Cell Biol.* **123**, 771-783.

Rout, M.P, Aitchison, J.D., Suprapto, A.., Hjertaas, K., Zhao, Y. and Chait, B.T. (2000)
The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol*. **148**, 635-651.

Saavedra,C., Felber,B.K. and Izaurralde,E. (1997)
The simian retrovirus-1 constitutive transport element (CTE), unlike HIV-1 RRE, utilizes factors required for cellular RNA export. *Curr. Biol.* **7**, 619–628.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989)
Molecular cloning: A laboratory manual (New York: Cold Spring Harbor Laboratory Press)

Santos-Rosa, H., Moreno, H., Simos, G., Segref, A., Fahrenkrog, B., Panté, N. and Hurt, E. (1998)
Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores. *Mol. Cell. Biol.* **18**, 6826–6838.

Scherly, D., Boelens, W., Datham, N.A., van Venrooij, W.J. and Mattaj, I.W. (1990a)
Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B" and their cognate RNAs. *Nature* **345**, 502–506.

Scherly, D., Dathan, N.A., Boelens, W., van Venrooij, W.J. and Mattaj, I.W. (1990b)
The U2B" RNP motif as a site of protein–protein interaction. *EMBO J.* **9**, 3675–3681.

Schmitt, I. and Gerace, L. (2001)
In vitro analysis of nuclear transport mediated by the C-terminal shuttle domain of TAP. *J. Biol. Chem.* **276,** 42355-42363.

Scott, W.G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J. and Klug, A. (1995)
Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.* **250,** 327-332.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Luhrmann, R. and Hurt, E. (1997)
Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)$^+$ RNA and nuclear pores. *EMBO J.* **16**, 3256–3271.

Stade, K., Ford, C.S., Guthrie, C., and Weis, K. (1997)
Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041-1050.

Strasser, K. and Hurt E.C. (2000)
Yra1p, a conserved nuclear RNA binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J.* **19,** 420-430.

Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Séraphin, B., Wilm, M., Bork, P. and Izaurralde, E. (2000)
FEF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**, 638-650.

Suyama, M., Doerks, T., Braun, I.C., Sattler, M., Izaurralde, I. and Bork, P. (2000)
Prediction of structural domains of TAP reveals details of its interaction with p15 and nucleoporins. *EMBO Repts.* **1**, 53-58.

Tabernero, C., Zolotukhin, A.S., Valentin, A., Pavlakis, G.N. and Felber, B.K. (1996)
The posttranscriptional control element of the simian retrovirus type 1 forms an extensive RNA secondary structure necessary for its function. *J. Virol.* **70**, 5998–6011.

Tabernero, C., Zolotukhin, A.S., Bear, J., Schneider, R., Karsenty,G. and Felber, B.K. (1997)
Identification of an RNA sequence within an intracisternal-A particle element able to replace Rev-mediated posttranscriptional regulation of human immunodeficiency virus type 1. *J. Virol.* **71**, 95-101

Vankan, P., McGuigan, C. and Mattaj, I.W. (1992)
Roles of U4 and U6 snRNAs in the assembly of splicing complexes. *EMBO J.* **11**, 335–343.

Varani, G. and Nagai, K. (1998)
RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 407–445.

Weichenrieder, O., Wild, K., Strub, K. and Cusack, S. (2000)
Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* **408,** 167-173.

Wen, W., Meinkoth, J.L., Tsien, R.Y. and Taylor, S.S. (1995)
Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82,** 463-473.

Yoon, D.W., Lee, H., Seol, W., DeMaria, M., Rosezweig, M. and Jung, J.U. (1997)
Tap: a novel cellular protein that interacts with tip of herpes virus saimiri and induces lymphocyte aggregation. *Immunity* **6**, 571-582

Zapp, M.L. and Green, M.R. (1989)
Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature* **342,** 714-716.

Zhou, Z.L., Luo, M., Strasser, K., Katahira, J., Hurt, E. and Reed, R. (2000)
The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407,** 401-405.

Zolotukhin, A.S., Valentin, A., Pavlakis, G.N. and Felber, B.K. (1994)
Continuous propagation of RRE(-) and Rev(-)RRE(-) human immunodeficiency virus type 1 molecular clones containing a cis-acting element of simian retrovirus type 1 in human peripheral blood lymphocytes. *J Virol* **68,** 7944-7952.

LIST OF PUBLICATIONS

THIS THESIS IS BASED ON THE FOLLOWING PUBLICATION:

Liker, E., Fernandez, E., Izaurralde, E. and Conti E. (2000) The structure of the mRNA export factor TAP reveals a *cis* arrangement of a non-canonical RNP domain and an LRR domain. *EMBO J.* **19.** 5587-5598.

OTHER PUBLICATIONS

Busheva, M., Garab, G., **Liker**, **E.,** Toth, Zs., Szell, M. and Nagy, F. (1991)
Diurnal fluctuations in the content and functional properties of the light harvesting chlorophyll a/b complex in thylakoid membranes. *Plant Physiol.* **95,** 997-1003.

Liker, E. and Garab, G. (1995)
Diurnal fluctuation in the composition of chlorophyll a/b light harvesting antenna of photosystem 2 in young wheat leaves. *Physiologia Plantarum* **93**, 187-190.

CONFERENCE ABSTRACTS

Istokovics, A., Lajko, F., **Liker**, **E.,** Barzda, V., Simidjiev, I. and Garab, G. (1992) Inhibition of the light induced reversible structural rearrangements of the macrodomains and the phosphorylation of membranes by quinone antagonistics. *Research in Photosynthesis*, Kluwer Academic Publishers, Vol. II. (ed. Norio Murata) pp. 631-634.

Liker, **E.,** Busheva, M., Nagy, F. and Garab, G. (1993)
Diurnal fluctuation in the structure and function of LHCII in wheat thylakoids. 11th International Biophysics Congress July 25-30, 1993 Budapest, Hungary, Book of Abstracts pp. 193.

Liker, **E.,** Cheng, L., Garab, G. and Allen, J.F. (1995)
Sensitivity of the phosphorylation of different phosphoproteins to Q_o and Q_i site inhibitors of the cytochrome b_6/f complex in chloroplast thylakoids. in *Photosynthesis: from Light to Biosphere* (ed: Mathis, P.) Vol.I. pp. 85-89. Kluwer Academic Publishers, Dordrecht.

ACKNOWLEDGEMENT

I would like to thank Dr. Elena Conti that I could do this work in her lab. I got a lot of independence, a lot of trust and a lot of help from her. In her lab I have learned a lot of new things, she was always ready to teach, explain and discuss. I will not forget the special atmosphere, when the two of us arrived to Heidelberg and started a new lab at EMBL. Many thanks for her support in writing this thesis, her comments and discussions.

I would like to thank all people in the lab, Elena Fernandez, Noemi Fukuhara and Judith Ebert for their help, support and friendship. They were very good colleges, together we were a very happy "women-lab".

I am very grateful to Dr. Elisa Izaurralde, to our collaborator, who helped us a lot in our work, providing vectors, constructs, samples and advices. Special thanks for the experiments she made in order to support our results.

I am very grateful to Dr. Matti Saraste that he gave opportunity to work in his department, in the Structural Biology Program in EMBL. I am very sad because of his death, I will have always good memories about him.

I am also very grateful to Dr. Andor Udvardy and Dr. István Kiss for reading the thesis, and for their useful comments. Many thanks for Dr. János Gausz for encouraging me to write this thesis.

I would like to thank Tamás Ábrahám and Mária Bodócsi for their support, encouragement and for the good atmosphere at Biocenter Ltd. that helped me a lot in writing the thesis and finishing the PhD procedure.

Last, but not least, I would like to thank my family for patience, love and support during my work in Heidelberg, and during writing the thesis in Szeged.

SUMMARY OF PHD THESIS

INTRODUCTION

Eucaryotic RNAs are transported from their site of transcription in the nucleus to their site of function in the cytoplasm via interaction with factors that recognize and translocate individual RNA cargoes. Specific factors mediate the nuclear export of different classes of RNA. U snRNA, tRNA, mRNA and rRNA do not compete for export, suggesting that they access distinct export pathways.

Nuclear export of cellular mRNAs is highly selective, as usually only fully processed RNAs are exported. Incompletely spliced pre-mRNAs and excised introns are actively retained in the nucleus. In contrast, retroviruses need to export unspliced and partially spliced RNA transcripts to the host cytoplasm for viral replication to occur. To overcome nuclear retention retroviruses encode *cis*-acting RNA elements that function as export signals for unspliced RNAs and interact with the cellular transport machinery. One such element, the *cis*-acting constitutive transport element (CTE) of simian type D retroviruses interacts directly with the cellular factor TAP.

Human TAP is a multidomain protein of 70 kDa molecular weight. The C-terminal portion interacts with components of the nuclear pore complex, while the N-terminal half of TAP (residues 1-372) binds the CTE-RNA and several RNA binding proteins. The CTE-RNA folds into an extended stem-loop structure comprising two identical internal loops that are arranged in mirror symmetry on the RNA element. The internal loops are conserved and are the interaction sites for the cellular protein TAP. Besides being implicated in the nuclear export of unspliced genomic RNA of simian type D retroviruses, TAP also plays a role in cellular mRNA export. However, several lines of evidence suggest that the mode of

interaction of TAP with viral and cellular RNAs is different. While binding to the retroviral RNA is direct, binding to the cellular mRNA is likely not direct but mediated by adapter proteins.

OBJECTIVES

The objective of this work is to elucidate the determinants of TAP – CTE-RNA recognition at the molecular level, using both biochemical and biophysical approaches. For this, X-ray crystallography is a very powerful tool, which entails several steps. The project is extensively biochemical at the outset, to assess the minimal domain of the protein that is able to bind the cognate RNA with affinity and specificity comparable to the wild-type protein. A suitable expression system for protein production and protocols for protein purification are established, to yield the minimal CTE-binding domain of TAP in the quantities and homogeneity required for structural studies. The use of a GST fusion *E.coli* expression vector with a Tev protease cleavage site, and the combination of affinity and ion-exchange chromatography results in mg quantities of pure protein per liter of bacterial cell. Crystallization experiments are subsequently carried out, screening hundreds of different conditions. Once crystals that diffract to high enough resolution are reproducibly obtained, the structure is determined, particularly using synchrotron radiation. The structural information guides mutagenesis studies of selected protein residues to gain molecular insight on how the macromolecule functions.

RESULTS AND CONCLUSIONS

The export of CTE-bearing retroviral RNAs to the host cytoplasm is achieved by their direct interaction with a fragment of the cellular protein TAP. The aim of this work is to obtain molecular insights in how the TAP-CTE interaction is achieved. To this end we solved the X-ray structure of the minimal CTE interaction domain of TAP, and used this structural information to probe the TAP interaction surface by mutagenesis.

To summarize the results, we can conclude that the TAP minimal CTE binding domain includes two independent globular domains. The N-terminal domain folds and functions as a *bona fide* RNP domain, despite lacking the canonical conserved sequence motifs. The C-terminal domain is an LRR-containing protein that does not show general RNA-binding activity but is required for specific binding to CTE-RNA. The two independent domains have similar structural and biochemical properties to the U2B'' and U2A' components of the spliceosomal complex. Functional studies with structure-based mutants indicate that positively charged residues at the β-sheet platform of the RNP are likely to be involved in RNA binding, similarly to U2B'' and to canonical RNP proteins in general. A residue identified on a helix at the back of the RNP platform plays an important role, either in RNP-LRR or RNP-RNA interactions. A positively charged patch at the outer convex surface of the LRR domain might also be involved in RNA binding, conferring specificity to CTE-RNA recognition similarly to the positively charged patch on the U2A' surface in U2 snRNA stem recognition. Despite these similarities, it is conceivable that the roles of the TAP and U2A' LRRs in specific RNA binding are at least partly divergent. The most obvious difference is that recognition of the CTE-RNA requires the RNP and LRR domains to be present in *cis* in the same polypeptide together with the N-terminal flexible region, while U2B'' and U2A' function as separate proteins. Furthermore, substitution of residues along the

conserved concave surface of the LRR domain of TAP does not disrupt CTE binding *in vitro* or its export *in vivo*, as would be expected for a U2B''-U2A' mode of protein-protein interaction.

Molecular details of how the CTE-RNA interacts with TAP can only be obtained by determining the crystal structure of the complex. Given the size of the RNA, this is a long-term project, which goes beyond the scope and time frame of this work. As part of this goal, we identified the smallest CTE-RNA fragment that is suitable for crystallization and that is still active in TAP binding. We showed that a double hammerhead ribozyme strategy can be used to transcribe RNA with homogenous 3' ends and without any unspecific nucleotides at the 5' end. These results are currently being used to pursue structural studies of the CTE-RNA TAP complex.

A PHD ÉRTEKEZÉS ÖSSZEFOGLALÓJA

BEVEZETÉS

Az eukarióta RNS-ek a sejtmagból, a transzkripció helyéről a citoplazmába, a funkció helyére transzportálódnak. Ez a transzport olyan faktorok közreműködésével jön létre, amelyek egyedi RNS szállítmányokat ismernek fel, és exportálnak. A különböző osztályokba tartozó RNS-ek nukleáris exportját specifikus faktorok irányítják. Az U snRNS-ek, a tRNS-ek, mRNS-ek és rRNS-ek nem versenyeznek egymással az export terén, ami arra enged következtetni, hogy a különböző típusú RNS-ek különböző export utakat járnak be.

A sejtek mRNS-einek nukleáris exportja igen szelektív folyamat, mivel csak a teljesen processzálódott RNS-ek exportálódnak. A splicingon nem teljes mértékben átesett pre-mRNS-ek és a kivágódott intronok a sejtmagban maradnak. Ezzel ellentétben retrovírusok esetén ahhoz, hogy a virális replikáció megtörténhessen, szükség van a splicing folyamaton át nem esett vagy csak részleges intron kivágódáson túljutott RNS transzkriptumoknak a gazdasejt citoplazmájába történő exportjára is. A retrovírusok ahhoz, hogy legyőzzék ezt a nukleáris retenciót, olyan *cis* RNS elemeket kódolnak, amelyek a nem érett RNS számára export szignálként funkcionálnak, és a sejt transzport rendszeréhez kapcsolódnak. Az egyik ilyen elem, a simian típusú D retrovírusok *cis* konstitutív transzport eleme (CTE) közvetlenül kapcsolódik a celluláris TAP fehérjéhez.

A humán TAP egy 70 kDa-os multidomén protein. A C-terminális rész a nukleáris pórus komplex komponenseihez kötődik, míg az N-terminális rész a CTE-RNS-hez és különböző RNS kötő fehérjékhez kapcsolódik. A CTE-RNS egy kiterjedt stem-loop szerkezettel rendelkezik, amely két belső, egymással megegyező hurkot foglal magába. Ezek a hurkok tükör szimmetrikusan helyezkednek el az RNS elemen. A belső hurkok szekvenciája

konzervált, és ezek a hurkok szolgálnak a celluláris TAP fehérje interakciós helyéül. A simian típusú D retrvírusok intron kivágódáson át nem esett genomiális RNS-ének nukleáris exportjában beöltött szerepén túl a TAP fehérje a celluláris mRNS exportjában is közreműködik. Különböző kísérletek bizonyítják azonban, hogy a TAP interakciója a virális és celluláris RNS-ekkel különböző: míg a retrovírus RNS-hez közvetlenül kötődik, addig a celluláris RNS-hez valószínűleg adapter molekulákon keresztül kapcsolódik.

CÉLKITŰZÉSEK

Munkám során a TAP – CTE-RNS interakció molekuláris szintű elemeinek meghatározását tűztem ki célul biokémiai és biofizikai módszerekkel. Ehhez kitűnő segítséget nyújt a Röntgen krisztallográfia. Ahhoz, hogy a Röntgen krisztallográfiához megfelelő vizsgálati anyagot szolgáltassunk, előzetesen biokémiai módszerekkel meg kell találnunk a fehérjének azt a minimális doménjét, amely a vad típusú fehérjével összevethető mértékben és specificitással képes az RNS-t megkötni. Meghatároztuk a legmegfelelőbb expressziós rendszert és fehérje tisztítási módszert, melyek segítségével a TAP fehérje minimális CTE-kötő doménjét kellő mennyiségben és tisztaságban képesek voltunk előállítani. A Tev proteáz hasító hellyel rendelkező GST fúziós *E. coli* expressziós vektor használatával, az affinitás és ioncserélő kromatográfia alkalmazásával képesek voltunk a baktérium sejtekből mg-nyi mennyiségű tiszta proteint előállítani literenként. A megfelelő mennyiségű és tisztaságú fehérje kinyerése után a kristályosítási kísérletek következtek, több száz kristályosítási körülmény kipróbálásával. Miután sikerült olyan kristályokat kapni, amelyek megfelelő felbontásban diffraktáltak, a szinkrotronban nyert diffrakciós adatokból a szerkezetet

meghatároztuk. A szerkezeti információk segítségével mutagenezis vizsgálatok alapján igyekeztünk betekintést nyerni a funkció molekuláris alapjaiba.

AZ EREDMÉNYEK ÉS ÖSSZEFOGLALÁSUK

A CTE-t magában hordozó retrovírus RNS gazdasejtbe történő exportja a celluláris TAP fehérje fragmentumával való közvetlen kapcsolaton keresztül valósul meg. Ennek a munkának a célja az volt, hogy betekintést nyerjünk a TAP-CTE kapcsolat molekuláris alapjaiba. E cél megvalósítása érdekében meghatároztuk a TAP minimális CTE-kötő doménjének Röntgen krisztallográfiás szerkezetét, és ezeknek a szerkezeti adatoknak a segítségével mutagenezis vizsgálatokkal teszteltük a TAP interakciós felszínét.

Az eredményeket összefoglalva arra a következtetésre juthatunk, hogy a TAP minimális CTE-kötő doménje két egymástól független globuláris domént tartalmaz. Az N-terminális domén úgy hajtogatódik és funkcionál, mint egy RNP domén, annak ellenére, hogy nem rendelkezik kanonikus konzervált szekvencia motívumokkal. A C-terminális domén egy LRR-t tartalmazó fehérje fragmentum, amely nem mutat általános RNS-kötő aktivitást, de jelenlétére szükség van a CTE-RNS-hez való specifikus kötődéshez. A két független doménnak a spliceosome komplex U2B'' és U2A' komponenseihez hasonló szerkezeti és biokémiai tulajdonságai vannak. A szerkezetre alapozott mutációk vizsgálatával kimutattuk, hogy az RNP β-sheet platformján helyet foglaló pozitív aminosavak valószínűleg szerepet játszanak az RNS kötésben, az U2B'' és a kanokikus RNP fehérjékhez hasonlóan. Egy, az RNP platform mögött elhelyezkedő hélixen azonosított aminosav fontos szerepet játszik vagy az RNP-LRR, vagy pedig az RNP-RNS interakcióban. Az LRR domén külső konvex felületén található pozitív töltésű terület valószínűleg szintén szerepet játszik az RNS kötésben, ami

arra utal, hogy a CTE-RNS felismerés specifikus, hasonlóan ahogy az U2A' fehérje pozitív töltésű része felismeri az U2 snRNS-t. Ezen hasonlóságok ellenére elképzelhető, hogy a TAP és U2A' LRR doménjének specifikus RNS kötésben betöltött szerepe bizonyos mértékben különbözik. A legnyilvánvalóbb különbség az, hogy míg a CTE-RNS felismeréshez az RNP és LRR domének a flexibilis N-terminális régióval együtt egy polipeptiden helyezkednek el, addig az U2B'' és az U2A' proteinek különálló fehérjeként funkcionálnak. Ezen kívül, a TAP fehérje konkáv felületén elhelyezkedő konzervált aminosavak mutációja nincs hatással az *in vitro* CTE kötésre és az *in vivo* exportra, mint ahogy ezt egy U2B''-U2A' típusú fehérje-fehérje interakciónál elvárnánk.

Ahhoz, hogy a CTE-RNS – TAP interakció molekuláris részleteibe betekintést nyerhessünk, a komplex kristály szerkezetét kell meghatároznunk. Az RNS mérete miatt ez egy hosszú távú project, és túlmutat ennek a munkának a hatáskörén és időintervallumán. A teljes project részeként meghatároztuk azt a legkisebb CTE-RNS fragmentumot, ami kristályosításra alkalmas, de még mindig rendelkezik TAP kötő aktivitással. Megmutattuk, hogy a kettős „hammerhead" ribozim stratégia alkalmas arra, hogy a transzkripciót úgy hajtsuk végre, hogy az RNS 3' vége homogén legyen, és az 5' végén elkerüljük a nem specifikus nukleotidok jelenlétét. A heidelbergi csoportban a CTE-RNS – TAP komplex szerkezeti vizsgálata jelenleg is folytatódik ezeknek az eredményeknek a felhasználásával.