Summary of PhD thesis

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

by

ERIKA LIKER

Supervisor: Dr. Elena Conti

European Molecular Biology Laboratory
Heidelberg, Germany

2002

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.

METHODS

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 resin (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.

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. 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.

Gel-shift with labeled RNA

For native gel assays with labeled RNA, a 224 nucleotide long CTE-RNA probe was employed. 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, 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.

Gel-shift with unlabeled RNA

In order to mimic crystallization conditions, RNA binding assays were done without labeling 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. Gels were visualized with toluidine blue, occasionally double stained with toluidine blue for RNA and Coomassie Brilliant Blue for proteins.

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. 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.

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. 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.

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.

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.