INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

.

Identification of proteins interacting with lupin and *Arabidopsis* tRNA

nucleotidyltransferases

Jun Gu

A Thesis

in

The Department

of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

March 2000

Canada

ABSTRACT

Identification of proteins interacting with lupin and *Arabidopsis* tRNA nucleotidyltransferases

Jun Gu

ATP (CTP):tRNA nucleotidyltransferase catalyzes the addition of CMP and AMP residues to the 3' ends of tRNAs that have an incomplete CCA sequence. In eukaryotic cells, nuclear, mitochondrial and chloroplast genomes all encode tRNAs. Therefore, although tRNA nucleotidyltransferase is synthesized in the cytosol, it must function in multiple organelles and be targeted to these locations from its site of synthesis. In general, while much is known about the targeting information contained in proteins, less is known about the accessory factors involved in targeting. To identify proteins which could interact with lupin or *Arabidopsis* tRNA nucleotidyltransferase, the yeast two-hybrid system was employed to screen an *Arabidopsis* cDNA library. As a result, 11 cDNA clones have been isolated belonging to four different classes of cDNAs. Representatives from each class were sequenced in their entirety. Sequence analysis of these cDNAs revealed a major open reading frame for each cDNA. The predicted amino acid sequence of each cDNA was used to search the GenBank and *Arabidopsis* databases. One group of clones could encode a 615-amino acid polypeptide showing 54% identity with the *Arabidopsis* kinesin light chain. Another group could encode a 129-amino acid polypeptide which shows 39% identity with the human KE2 protein. A third group could encode a 180-amino acid polypeptide which shows 58% identity with cucumber raffinose

synthase and finally one clone could encode a 188-amino acid polypeptide which shows

37% identity with a hypothetical protein from *Synechocystis* sp. Mapping of the regions

of tRNA nucleotidyltransferase required for interaction demonstrated that the amino-

terminal extensions of the tRNA nucleotidyltransferases are required for interaction with

all four of these clones.

ACKNOWLEDGMENTS

DEDICATION

I would like to dedicate this thesis to my beloved wife, Dan Yang, who always encouraged and supported me during my academic endeavour.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

aa   -Amino acid

bp   -Base pairs

DTT   -Dithiothreitol

dH_2O -Distilled water

EDTA   -Ethylenediaminetetraacetic acid

g   -Gram

His   -Histidine

kDa   -Kilo Dalton

LB   -Luria Broth

Leu   -Leucine

min   -Minute

OD   -Optical density

PCR   -Polymerase chain reaction

PEG   -Polyethyleneglycol

rpm   -Revolutions per minute

RT   -Room temperature

SC   -Synthetic complete

TBE   -Tris / Borate / EDTA

TE   -Tris / EDTA

Trp   -Tryptophan

X-Gal   -5-bromo-4-chloro-3-indolyl-β- galactoside

YPD   -Yeast extract / peptone / dextrose

INTRODUCTION

Transfer ribonucleic acids (tRNAs) are low molecular weight RNA molecules with known three-dimensional structure. They play an essential role by acting as adaptors in protein synthesis, and so, are required for the growth of all cells. Transfer RNAs are encoded by tRNA genes and are transcribed as precursors. The precursor tRNAs contain the complete tRNA primary sequence as well as additional residues at both the 5' and 3' ends (Deutscher, 1984). Before tRNAs can function in protein synthesis they have to be processed from these precursors into a mature form. The maturation of a functional tRNA involves a number of steps, e.g. 5' leader and 3' trailer sequences must be removed by processing enzymes, numerous base modifications must be carried out and a 3'-terminal cytidine-cytidine-adenosine addition by tRNA nucleotidyltransferase must occur on tRNAs lacking this sequence. In addition, some tRNAs contain introns that must be removed (Deutscher, 1990). Because many steps of tRNA maturation take place in the nucleus (Deutscher, 1984) whereas protein synthesis takes place in the cytosol, tRNA maturation enzymes have to be imported from the cytosol (where they are synthesized) into the nucleus to function in tRNA maturation. In addition, both mitochondrial and chloroplast genomes encode transfer RNAs which function in organellar protein synthesis. The maturation of organellar tRNAs occurs inside these organelles, and tRNA processing activities in both mitochondrial (Hanic-Joyce and Gray, 1990) and chloroplast (Greenberg and Hallick, 1986) extracts have been reported. Although mitochondria and chloroplasts contain their own genomes, they encode only a few proteins. The majority of mitochondrial and chloroplast proteins, including those involved in tRNA biogenesis, are encoded by nuclear genes, synthesized on cytosolic ribosomes and subsequently

1

imported into these organelles. I am interested in one specific tRNA maturation enzyme, ATP (CTP): tRNA nucleotidyltransferase which has been found in the nucleus, mitochondrion and chloroplast.

I. Transfer RNA nucleotidyltransferase

All tRNAs need a 3'-terminal CCA sequence to function as the amino acid attachment site in protein synthesis. The incorporation of CMP and AMP residues into tRNAs that have an incomplete CCA sequence is catalyzed by the enzyme ATP (CTP): tRNA nucleotidyltransferase. Among tRNA maturation enzymes in *Escherichia. coli*, tRNA nucleotidyltransferase is one of the best characterized (Williams and Schofield, 1977). Among eukaryotes, tRNA nucleotidyltransferase has been well characterized in yeast (Chen *et al.*, 1990), housefly (Poblete *et al.*, 1977), rabbit (Masiakowski and Deutscher, 1980), rat (Mukerji and Deutscher, 1972), wheat (Dullin *et al.*, 1975; Vicaretti and Joyce, 1999) and lupin (Cudny *et al.*, 1978; Shanmugam *et al.*, 1996).

In *E. coli*, all tRNA genes encode the 3' CCA end sequence (Komine *et al.*, 1990), so tRNA nucleotidyltransferase activity is not essential in this organism (Zhu *et al.*, 1986), although slower growth rates of tRNA nucleotidyltransferase mutants (Deutscher *et al.*,1974) suggest a repair function for this enzyme. In contrast, eukaryotic nuclear, mitochondrial and chloroplast tRNA genes lack a complete 3'-terminal CCA sequence (Sprinzl *et al.*, 1998) so that tRNA nucleotidyltransferase performs an essential role in tRNA biosynthesis in eukaryotes. In yeast, a single nuclear gene, *CCA1*, encodes the nuclear, mitochondrial and cytosolic forms of this enzyme (Chen *et al.*, 1992). A yeast strain carrying a temperature-sensitive mutation in the *CCA1* gene (*cca1-1*) cannot grow at the non-permissive temperature (37 ^{0}C) and CCA addition to both cytosolic and

2

mitochondrial tRNAs is affected (Chen *et al.*, 1992), indicating that the products of this gene are required for CCA addition in the cytosol and mitochondrion. Sequence analysis of the yeast *CCA1* gene reveals three in-frame start codons at amino acids 1, 10 and 18 of the predicted open reading frame (Chen *et al.*, 1990), which suggests a way to make multiple forms of the gene products from this single gene (Martin and Hopper, 1994). Cca1p-I (produced from the first start codon) is located primarily in mitochondria with minor activity detected in the nucleus and cytosol. In contrast, Cca1p-III (produced from the third start codon) is located only in the nucleus and cytosol with no enzyme activity found in mitochondria. Finally, Cca1p-II (produced from the second start codon) is located primarily in the nucleus and cytosol with a small amount imported into mitochondria (Chen *et al.*, 1992; Wolfe *et al.*,1994). When both the first and second start codons were eliminated by site-directed mutagenesis and the constructs were introduced into a yeast strain carrying a temperature-sensitive mutation in its genomic *CCA1* gene. these cells did not grow on a non-fermentable carbon source (glycerol) at the non-permissive temperature. but grew well on a fermentable carbon source (glucose) under the same conditions (Chen *et al.*, 1992). These results strongly suggest that (1) the mitochondrial and nuclear/cytoplasmic forms of yeast tRNA nucleotidyltransferase are encoded by a single gene. (2) the first 17 amino acids (upstream of the third start codon) are not essential for enzyme activity, and (3) the amino-terminal extension of yeast tRNA nucleotidyltransferase functions as a mitochondrial targeting signal. As yeast tRNAs have complete CCA ends even when blocked from export from the nucleus (Knapp *et al.*, 1978). tRNA nucleotidyltransferase also must be imported into the nucleus. suggesting

that this enzyme also contains a nuclear localization signal, although it has not been identified as yet.

In our lab, a major effort has been devoted to studying the targeting of tRNA nucleotidyltransferase in plants. Plant cells contain an extra organelle, the chloroplast, not present in yeast, which adds complexity to protein trafficking in these cells. Like the nuclear and mitochondrial genomes, the chloroplast genome encodes its own tRNAs, such that tRNA nucleotidyltransferase must function there also. Before this work, tRNA nucleotidyltransferase from lupin had already been purified and characterized in our lab (Shanmugam et al., 1996). As in yeast, the protein predicted from the cDNA encoding lupin tRNA nucleotidyltransferase contains additional amino-terminal sequence not present in the E. coli protein. The predicted lupin enzyme shows high sequence similarity with the yeast enzyme (Figure 1). Sequence analysis of the lupin tRNA nucleotidyltransferase cDNA revealed two potential in-frame start codons coding for amino acids 1 and 58. Depending on which start codon is used in translation, proteins produced will have, or lack, a 57 amino acid amino-terminal extension. These 57 amino acids have characteristics of both mitochondrial and chloroplast targeting signals (Von Heijne et al., 1989), and are not essential for enzyme activity (Shanmugam et al., 1996). These observations suggest that these amino-terminal amino acids of the lupin enzyme could encode an organellar targeting signal. The lupin enzyme containing these amino acids was able to complement the yeast cca1-1 mutant for growth on a fermentable carbon source but not for growth on a non-fermentable carbon source (Shanmugam et al., 1996). There are a number of possible explanations for the lack of growth on the non-fermentable carbon source. Firstly, these amino-terminal amino acids may not encode a

```
Lupin    MRLSFKTVTN VVVVLPRGRT RSIINFTLFP TIT.SNLVLH PLLRTPKTPS FHSSLSSPMS SHKVRDNIQL
Yeast    .......... ....MLRSTI SLLMNSAAQK TMTNSNFVLN A......... .......... .....PKITL
E.coli   .......... .......... .......... .......... .......... .......... ..........
                        +          +        + + ++ ++                          + +


Lupin    SDVEKRIFDR LLATLRFFNL QTH....... LRVA[GGWVRD KLLG]KECYDI D.IALDKMMG TEFVDKVREY
Yeast    TKVEQNICNL LNDYTDLYNQ KYHNKPEPLT LRIT[GGWVRD KLLG]QGSHDL D.IAINVMSG EQFATGLNEY
E.coli   .......... .......... .......MK IYLV[GGAVRD ALLG]PVKDR DWVVVGSTPQ EMLDAGYQQV
           ++  +      +           +          +    ++ *  +*** +***     *     * ++    + +  + +   + ++


Lupin    L....LSIGE EAQGVCVIES NPDQSKHLET ARMRLFDMWI DFVNLRSEEY TDNSRIPSMQ RFG[TPEEDAY
Yeast    LQQHYAKYGA KPHNIHKIDK NPEKSKHLET ATTKLFGVEV DFVNLRSEKY TELSRIPKY. CFG[TPEEDAL
E.coli   GRDFPVFLHP QTHEEYALAR TERKS..... .......... ......GSGY TGFTCYAAPD ..V[FLEDDLK]
           +           +          +        +      ++ +*+++++ +      ++        +++++++  *  *     +++        ++*+++++


Lupin    [RRDLTINSLF] YNINTDSVED FTKRGISDLK SGKIVTPLPP KAT[FLDDPLR VVRAIRFGAR]F...EFTLDE
Yeast    [RRDATLNALF] YNIHKGEVED FTKRGLQDLK DGVLRTPLPA KOT[FLDDPLR VLRLIRFASR]F...NFTIDP
E.coli   [RRDLTINALA] QDDNGEIIDP YN..GLGDLQ NRLLRHVSPA ...[FGEDPLR VLRVARFAAR]YAHLGFRIAD
           ***-*+*+*+  +-++   +++  ++++*+ **+   +  ++++*++ +  +*+++****  *+* +**++*  +      *+++


Lupin    DLKQAAACDE VKDALAAKIS RERIGTEIDL MISGNQPVKA MTYI.....C DLTI....FW IVFSLPPTFE
Yeast    EVMAEMGDPQ INVAFNSKIS RERVGVEMEK ILVGPTPLLA LQLIQRAHLE NVIF....FW HNDSSVVKFN
E.coli   ETLALMREMT HAGELE.HLT PERVWKETES ALTTRNPQVF FQVLRDCGAL RVLFPEIDAL FGVPAPAKWH
           +   +  +       ++    +++  +**++  *  +    + +    *   +    + +          + +       ++       + + ++


Lupin    PAISDGCERL CISQLDISWN LIHLLGKTTF TDEQRRLTLY AAMFLPLRNT IYRE[KKAKKV] PV.[VNYIFRE
Yeast    EENCQDMDK. .......... INHVYNDNIL NSHLKSFIEL YPMFLE.KLP ILRE[KIGRSP] GFQQNFILSA
E.coli   PEIDTGIHTL ......MTLS MAAMLSPQV. ..DVRFATLC HDLGKGLTPP ELWPRHHGHG PAGVKLV..E
           +++ +   +        + +      +   ++      +++ +    +   +++++      +  ++ + +


Lupin    [SLKRKAKDPE] TVLDLHRASN KFLSLIPCLV ......SNED VQIVGH.... CODWMTE..L IDVPVSSRV.
Yeast    [ILSPMA.NLQ] IIGNPKKKIN NLVSVTESIV KEGLKLSKND AAVIAKTVDS ICSYEEI..L AKFADRSQLK
E.coli   [QLCQRLRVPN] EIRDLARLVA EFHDLIHTFP ML.......N PKTIVKLFDS IDAWRKPQRV EQLALTSEAD
            *   +  +     +  ++ +   +   + +++    +             + +      + +  ++ +  +          +      +  +


Lupin    .RVLTGFLLR ELRDFWRVAL LISILLHPID VNDTEDESSQ LSKRRDLFNT VENSVIKLGD EKVWDVKQLI
Yeast    .KSEIGIFLR NFNGEWETAH FASLSDAFLK IPKLETKKIE L..LFQNYNE FYSYIFDNNL NNCHELKPIV
E.coli   VRGRTGFESA DYPQGRW... .......... .......... LREAWEVAQS VPTKAVVEAG FKGVEIREEL
           +  +*+ ++       +  +      +             +         *    +  +          +   +  + +


Lupin    NGKDVMSVLQ LKGGPMVKEW LDKAMACNLP IPQELQRNVL IG........ .
Yeast    DGKQMAKLLQ MKPGPWLGKI NNEAIRWQFD NPTGTDQELI THLKAILPKY L
E.coli   TRRRIAAVAS WKEQRCPKPE .......... .......... .......... .
           ++  +  +++  *  ++   +        +          +
```

Figure 1. Similarity of the lupin, yeast and *E.coli* tRNA nucleotidyltransferase amino acid sequences (Shanmugam *et al.*, 1996). Standard one letter abbreviations are used for the amino acids. * indicates an amino acid conserved in all three sequences. + indicates an amino acid conserved in two of the three sequences. . indicates a gap introduced to optimize alignments. Three regions of high amino acid sequence identity are boxed and a potential nuclear localization sequence in the lupin sequence is underlined.

mitochondrial targeting signal. If this is the case it is not surprising that the lupin enzyme is not imported into yeast mitochondria. Secondly, the lupin enzyme could contain an organellar targeting signal but this targeting signal is not suitable for transport into yeast mitochondria. Thirdly, the lupin enzyme requires proteins for targeting to mitochondria that are not present in the yeast cytosol. Fourthly, the enzyme does contain an organellar targeting signal and is imported into yeast mitochondria, but does not function there. As we can not distinguish between these possibilities as yet, whether this protein is targeted to mitochondria or chloroplast in lupin is still unclear. Interestingly, sequence analysis of the lupin tRNA nucleotidyltransferase revealed the consensus sequence for a classical nuclear localization signal (Figure 1) (Shanmugam *et al.*, 1996). Therefore, based on sequence analysis, it is possible that as in yeast, a single nuclear gene encodes multiple forms of lupin tRNA nucleotidyltransferase that function in different cellular compartments. The targeting and intracellular localization of the lupin tRNA nucleotidyltransferase are being addressed by others in our lab.

Recently, both the gene and cDNA encoding an *Arabidopsis* tRNA nucleotidyltransferase have been identified. The gene sequence arose from the *Arabidopsis* genome sequencing project while the cDNA was identified from the *Arabidopsis* EST library and sequenced completely in this work. The protein predicted from the *Arabidopsis* cDNA sequence shows high sequence similarity with the lupin tRNA nucleotidyltransferase (Figure 2). Sequence analysis suggests that like the lupin enzyme, the *Arabidopsis* tRNA nucleotidyltransferase could contain a putative amino-terminal organellar targeting signal and a potential nuclear localization signal so that

```
CCAARA   - MRLSSLPINTLINLPKSLFLISPFRFRNLNRSLTVASRISSTLLRVSGVSSRPCG   55
           XXXX      x   x xx  X X x        x  X  xX      x x
LUPCCA   - MRLS------FKTVTNVVVVLPRGRTRSIINFTLFPTITSNLVLHPLLRTPKTPS   49

CCAARA   - YWPSTNAAMTNVGEEDKQSIPSIELKENIELTDKERKIFDRLLSTLRYCNLDTQL  110
           x  X  x Xx         X  xxxXX XxX XxxXXXXXXxXXXx XX X X
LUPCCA   - FHSSLSSPMS-----------SHKVRDNIQLSDVEKRIFDRLLATLRFFNLQTHL   93

CCAARA   - RVAGGWVRDKLLGKESDDIDIAIDNMSGSEFLDKFKEYLSSRDEEVQGDTVIERN  165
           XXXXXXXXXXXXXXX XXXXXxX X XxXXxXX xXXX X  XX XX  XXX X
LUPCCA   - RVAGGWVRDKLLGKECYDIDIALDKMMGTEFVDKVREYLLSIGEEAQGVCVIESN  148

CCAARA   - PDQSKHLETAKLRIYDQWIDFFNLRSEEYTENSRIPTM-KFGTAKDDAFRRDLTI  219
           XXXXXXXXXXXxxXxxX XXXX XXXXXXXXxXXXXXxX xXXX  xXXxXXXXXX
LUPCCA   - PDQSKHLETARMRLFDMWIDFVNLRSEEYTDNSRIPSMQRFGTPEEDAYRRDLTI  203

CCAARA   - NSLFYNINSGAVEDLTERGIDDLKSGKIVTPLPAKATFLDDPLRVLRAVRFGARF  274
           XXXXXXXXx xXXX X XXX XXXXXXXXXXXX XXXXXXXXXxXXxXXXXX
LUPCCA   - NSLFYNINTDSVEDFTKRGISDLKSGKIVTPLPPKATFLDDPLRVVRAIRFGARF  258

CCAARA   - GFTLDEELKEAASSEEVRVALGEKISRERIGNEIDLMISGNGPVSAVTYLSDLKL  329
           XXXXXxXXX XXx xXXx XX  XXXXXXXX XXXXXXXX XX XxXXx XX x
LUPCCA   - EFTLDEDLKQAAACDEVKDALAAKISRERIGTEIDLMISGNQPVKAMTYICDLTI  313

CCAARA   - FSVVFALPSSAEPSPPENCGSLSQSYLEAMWSLLKTPRPGKFSVNKDGLLSMLLC  384
           X xXXxXX x XXx  x X  X  X Xx  X Xx      Xx      X
LUPCCA   - FWIVFSLPPTFEPAISDGCERLCISQLDISWNLIHLLGKTTFTDEQRRLTLYAAM  368

CCAARA   - FSLLGKLYTRTLRANRFLLSTTFFKFSMKRKTSDAETVMNIHQTTERFRSLIPSL  439
           X X    X  xX x  x    Xx XxXXXx X XXXx xX xx xX XXXX X
LUPCCA   - FLPLRNTIYREKKAKKVPVVNYIFRESLKRKAKDPETVLDLHRASNKFLSLIPCL  423

CCAARA   - EVKKDVELDELTWAADILEHWKSITLNEPVIPATSKIRVLTGFLLRDIKDFWRVS  494
              XX x    X xxxxx              xX xXxxXXXXXXXXXXxxxXXXXXx
LUPCCA   - VSNEDVQIVGHDWMTELID-----------VPVSSRVRVLTGFLLRELRDFWRVA  467

CCAARA   - LLTSLLLSATVDGSNDHQDIGQLDFQLERMRETYLTVEATIHELGLDKIWDAKPL  549
           XX XxXX      x          XX x Xx x XXX xx   XXXxXxXX X X
LUPCCA   - LLISILLHPIDVNDTEDESS-----QLSKRRDLFNTVENSVIKLGLEKVWDVKQL  517

CCAARA   - VNGREIMQIAELKGGSRLIREWQQKLLTWQLAYPNGTAEECKEWMRDIKAKRQRI  604
           xXXxxxX x  XXXX   xxXX X xx xX  Xx              x   X
LUPCCA   - INGKDVMSVLQLKGGPM-VKEWLDKAMACNLPIPQELQ-----------RNVLI  559

CCAARA   - E  605

LUPCCA   - G  560
```

7

Figure 2: Similarity of the tRNA nucleotidyltransferase amino acid sequences of lupin (LUPCCA) and *Arabidopsis* (CCAARA). The *Arabidopsis* sequence starts from ATG6 and the lupin sequence starts from ATG1. Standard one letter abbreviations are used for the amino acids. The character used to show that two aligned residues are identical is 'X' and the character used to show that two aligned residues are similar is 'x'. - indicates a gap introduced to optimize alignments. The potential nuclear localization sequence in the lupin sequence is underlined.

the targeting of this enzyme may be similar to that predicted in lupin and shown in yeast. Again, the targeting and intracellular localization of this enzyme are being studied by others in our lab.

My research has concentrated on a search for proteins which interact with the lupin and/or *Arabidopsis* tRNA nucleotidyltransferases and which may include components of the transport machinery responsible for targeting these enzymes to their final intracellular destinations, be they the nucleus, mitochondria, or chloroplasts. In addition, these studies may identify other proteins which interact with tRNA nucleotidyltransferase but whose functions are still unknown. Because the mechanisms of protein targeting are becoming well established, I will focus on these mechanisms initially with emphasis on targeting signals and components of the transport machinery.

II. Nuclear localization of proteins

In eukaryotic cells, the nucleus is surrounded by the nuclear envelope, which includes both inner and outer membranes. Macromolecules such as proteins and RNAs move not only into, but also out of the nucleus through the nuclear pore complex (NPC), which forms the channel for bidirectional traffic (for review, see Heese-Peck and Raikhel, 1998). Protein import into the nucleus is dependent on specific targeting signals called nuclear localization signals (NLSs), which are found in most nuclear proteins (Dingwall and Laskey, 1991; Nigg, 1997). Classical nuclear localization signals are generally characterized by one or more clusters of basic amino acids that are found within the nuclear targeted protein itself (Gorlich and Mattaj, 1996). Most of the identified NLSs can be categorized into one of three classes (Nigg, 1997). Nucleoplasmin, the first

9

protein in which an NLS was experimentally demonstrated, contains a bipartite NLS (KRPAATKKAGQAKKKK) (Dingwall *et al.*, 1982). The bipartite signals are usually composed of two regions of basic amino acids separated by a spacer (Gorlich and Mattaj, 1996). This is the type of potential NLS identified in lupin tRNA nucleotidyltransferase. The NLS of the large T antigen of simian virus 40 (SV40). PKKKRKV, was initially found by point mutations that mistargeted this protein to the cytoplasm (Kalderon *et al.*, 1984) and was later defined as a monopartite NLS which possesses a single short region rich in basic amino acids (Gorlich and Mattaj, 1996). Finally, the yeast Mat α2-like NLS (KIPIK) consists of a short hydrophobic region that contains one or more basic amino acids (Hall *et al.*, 1984). Although there appear to be multiple different types of targeting signals, they seem to be functionally conserved among different organisms. For example, the SV40 large T-antigen NLS and bipartite NLSs from animals function in plants (Hicks and Raikhel, 1993), and the yeast Mat α2 NLS also functions in plants (Hicks *et al.*, 1995). However, there are some species-specific differences too, as the yeast Mat α2 NLS does not work in animals (Chelsky *et al.*, 1989).

The nuclear protein import process has been well studied in animals, plants and yeast. The basic mechanism appears to be conserved among these organisms (Heese-Peck and Raikhel, 1998; Smith *et al.*, 1997). It involves the NLS-dependent docking of import substrate at the NPC followed by an energy-dependent translocation through the NPC (Gorlich, 1998). The great development in studying nuclear import has been achieved by using an *in vitro* import system. This system is based on digitonin-permeabilized mammalian cells that maintain import-competent nuclei and intact nuclear envelopes, but which lose most cytosolic proteins (Adam *et al.*, 1990). By using this *in*

vitro system, some soluble cytosolic factors which mediate nuclear import in animals

have been characterized. These include importin α (Adam and Gerace, 1991), importin β

(Gorlich *et al.*, 1995a), NTF2 (nuclear transport factor2) (Paschal and Gerace, 1995), Ran

(small GTPase) (Moore and Blobel, 1993), and Ran-binding proteins such as RanBP5

(Deane *et al.*, 1997) and RanBP7 (Gorlich *et al.* 1997). The import process is illustrated

in Figure 3. The initial cytoplasmic event in nuclear import is the binding of the NLS-

containing protein to the importin α/β heterodimer to form a trimeric complex (Gorlich *et*

al., 1995b). Importin α provides the NLS-binding site (Adam and Adam, 1994), whereas

importin β enhances the binding affinity and mediates the docking of the trimer to the

NPC (Gorlich *et al.*, 1995a). The interaction between importins α and β in nuclear

protein import has been demonstrated in several ways including the *in vitro* import

system described above and the yeast two-hybrid system (Cortes *et al.*,1994).

Translocation of the trimeric complex through the NPC requires GTP hydrolysis by Ran

and appears to be facilitated by NTF2, Ran-binding proteins as well as some components

of the NPC (Heese-Peck and Raikhel, 1998). Ran, a small GTPase, cycles between Ran-

GTP and Ran-GDP during nuclear transport. These two forms of Ran bind to different

components of the transport machinery to mediate multiple protein-protein interactions

important for translocation of the trimeric complex through the NPC (Koepp and Silver,

1996). Furthermore, GTP hydrolysis by Ran provides much of the energy required for

nuclear transport (Melchior *et al.*, 1993). Ran activity is mediated by Ran-binding

proteins (Moore and Blobel, 1993). Using an *in vitro* import system, NTF2 was found to

mediate nuclear protein import by interacting with the NPC protein p62 (Paschal and

Gerace, 1995). After translocation, both importin α and β are released and then recycled
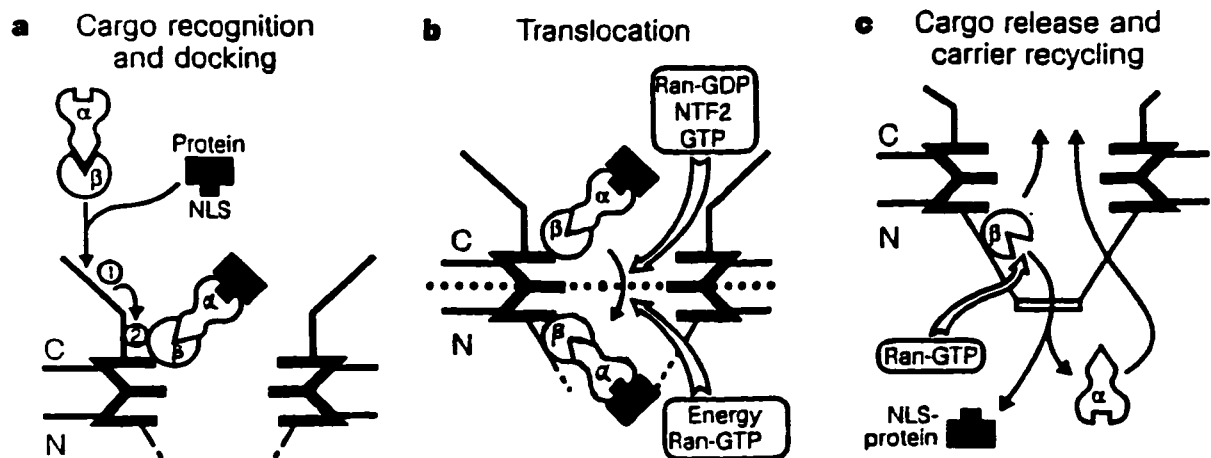
Figure 3: Model for protein import into the nucleus (Nigg, 1997). **a.** NLS protein recognition by an importin α/β heterodimer to form a trimeric complex and docking to the NPC in a two-step reaction. **b,** Translocation of the trimeric complex through the NPC is energy-dependent, requires Ran, NTF2, and Ran-binding proteins as well as some components of the NPC. **c,** After translocation, both importin α and β are released and return to the cytoplasm separately to start a new cycle of import.

to the cytoplasm separately to start a new round of protein import (Gorlich, 1998).

Cytosolic chaperone Hsp70 is involved in the import of some nuclear proteins to confer

upon them an import-competent state (Shi and Thomas, 1992). However, Hsp70 is not

an organelle-specific import factor, as it is also involved in endoplasmic reticular

secretion and mitochondrial and chloroplast targeting (Deshaies *et al.*, 1988; Endo *et al.*,

1996).

Some components of the NPC, nucleoporins, have been identified to serve as

binding sites for the NLS-dependent import complex at the NPC. These nucleoporins are

usually characterized by FXFG or GLFG repeat motifs (Heese-Peck and Raikhel, 1998).

Antibodies against the FXFG domain inhibit nuclear import *in vivo* in *Xenopus* oocytes

(Featherstone *et al.*, 1988). Moreover, *in vitro* binding experiments demonstrated that

FXFG-containing proteins p62 (Starr *et al.*, 1990), Nup153 (Sukegawa and Blobel,

1993), and Nup214 (Kraemer *et al.*, 1994) bind nuclear import substrates in *Xenopus*

oocytes. Yeast Nup2 and vertebrate Nup358 were shown to interact with Ran in the yeast

two-hybrid system (Dingwall *et al.*, 1995). The GLFG region of Nup116 is essential for

cell viability and deletion of the Nup116-GLFG region inhibits nuclear protein import *in*

vivo (Iovine *et al.*, 1995). Furthermore, Nup116-GLFG interacts in the yeast two-hybrid

system and in ligand blot assays with Kap95, the yeast homologue of vetebrate importin

β (Iovine *et al.*, 1995). These experiments indicates that there are a large number of

proteins, both soluble and membrane-bound, involved in targeting of proteins to the

nucleus.

In plants, nuclear protein targeting appears to utilize an import machinery similar to

that described in animals and fungi. However, plant nuclear import has shown some

unique features (Hicks *et al.*, 1996). Recently, a vertebrate importin α homolog from *Arabidopsis* (At-IMPα) has been identified and characterized (Smith *et al.*, 1997). Using an *in vitro* binding/co-immunoprecipitation assay, At-IMPα was demonstrated to bind three classes of NLSs (the bipartite NLS, SV40 T-antigen NLS and Mat α2-like NLS), and the binding has been proven to be specific (Smith *et al.*, 1997). Plant importin α is localized to both the cytoplasm and the nucleus with a high concentration at the nuclear envelope (Smith *et al.*, 1997), suggesting that plant nuclei possess a NLS-binding site at the NPC. Further evidence to support this is that three classes of synthetic peptides corresponding to the SV40 large T antigen NLS, bipartite NLS from the endogenous transcription factor Opaque 2 and Mat α2-like NLS from the endogenous transcription factor R, were demonstrated to bind to purified plant nuclei at a single low-affinity site that is tightly associated with the nuclear envelope and NPC (Hicks *et al.*, 1995). These peptides bound to purified nuclei even in the absence of any added cytosolic fractions in the *in vitro* import experiment using permeabilized tobacco protoplasts (Hicks *et al.*, 1996). In addition, four NLS binding proteins were identified in the purified plant nuclei by a cross-linking approach using a bipartite NLS from Opaque-2 (Hicks and Raikhel, 1995). These results suggest that unlike animals and fungi some components of NLS recognition in plants are located at the NPC.

In eukaryotic cells, many steps of tRNA maturation, including 3' end CCA addition, occur in the nucleus, whereas protein synthesis takes place in the cytosol. One aim of this work is to identify proteins, such as soluble factors or NPC components, which are involved in targeting tRNA nucleotidyltransferase to the nucleus in plants.

III. Mitochondrial and chloroplast targeting signals

As mentioned previously, plant mitochondria and chloroplasts contain their own genomes encoding tRNAs for organellar protein synthesis. However, these organellar genomes have a limited coding capacity such that most mitochondrial and chloroplast proteins are encoded in the nuclear genome and are synthesized as precursors on cytosolic ribosomes. The precursors usually contain amino-terminal extensions as targeting signals which are necessary to direct these proteins to the appropriate organelles (Neupert, 1997; Keegstra and Cline, 1999). In most cases, the targeting peptides are removed by specific proteases during or shortly after import (Glick and Schatz, 1991). The mitochondrial and chloroplast targeting signals have some intriguing structural similarities. e.g., both are rich in basic and hydroxylated amino acid residues and lack acidic amino acid residues (Von Heijne et al., 1989). However, their predicted secondary structures differ. While the mitochondrial targeting peptides (mTPs) are predicted to form amphiphilic α-helices (Von Heijne et al., 1989), the chloroplast transit peptides (cTPs) are usually composed of three regions: 1) an amino-terminal region lacking proline. glycine or charged residues, 2) a central region rich in basic and hydroxylated residues. and 3) a carboxy-terminal region that is predicted to form an amphiphilic β-sheet (Von Heijne et al., 1989). Analysis of both classes of targeting peptides revealed that even within one class there is little conservation at the primary sequence level either in composition or in length (Von Heijne et al., 1989).

IV. Mitochondrial import machinery

Mitochondria contain an outer membrane, an inner membrane, a compartment

between these two membranes (intermembrane space), and an internal compartment

(matrix). Protein import into mitochondria is illustrated in Figure 4. It is a multi-step

process: preproteins synthesized on cytosolic ribosomes are targeted to the surface of

mitochondria, translocated through the mitochondrial membranes, and sorted to the

different mitochondrial subcompartments (for review, see Neupert, 1997). Cytosolic

factors which recognize and interact with the targeting sequences, help to deliver

preproteins to mitochondria (Hartl, 1996). Newly synthesized preproteins in the cytosol

may first interact with cytosolic factors which help them to attain import-competent

conformations and deliver them to the surface of mitochondria, although in some cases,

preproteins can be recognized by membrane receptors directly (Becker *et al.*, 1992;

Mayer et al., 1995). Several cytosolic factors have already been identified and

characterized including cytosolic chaperone Hsp70, a 70 kDa heat shock protein,

involved in protein import into mitochondria as well as into other organelles (Deshaies *et*

al., 1988). Hsp70 is generally believed to unfold preproteins so as to keep them in an

import-competent state (Deshaies *et al.*, 1988; Endo *et al.*, 1996). Another cytosolic

chaperone, mitochondrial import stimulation factor (MSF) consisting of 30- and 32- kDa

subunits, was first purified from rat liver cytosol and was found to stimulate

mitochondrial preprotein import (Hachiya *et al.*, 1993). Wheat germ lysate synthesized

mitochondrial preproteins are generally in aggregated states and are not imported or are

poorly imported into isolated mitochondria (Murakami *et al.*, 1988). MSF could unfold

the aggregated adrenodoxin precursor synthesized in wheat germ lysate and stimulate its
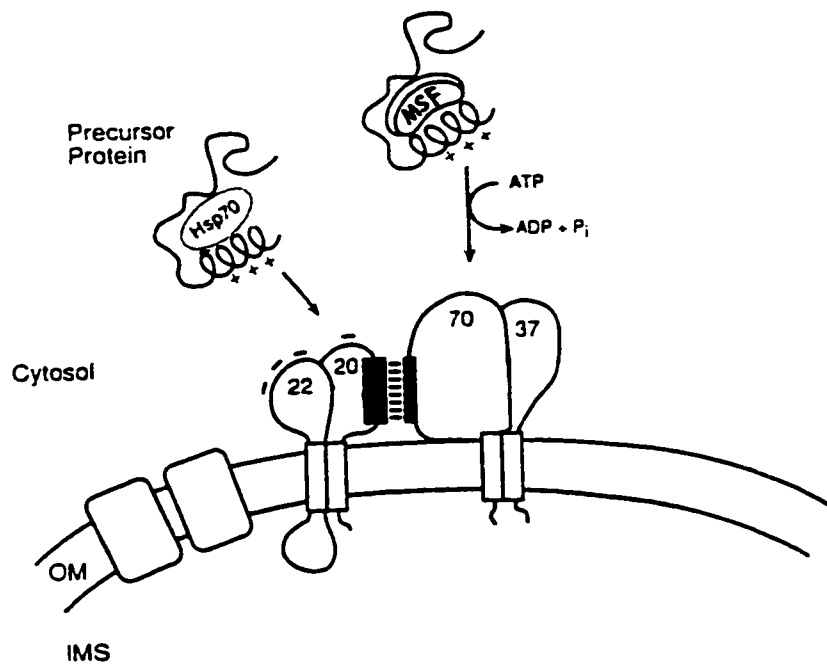
Figure 4: Model for protein import into mitochondria (Haucke *et al.*, 1996). Targeting of newly synthesized precursor proteins from cytosol to mitochondria is proposed to occur via two different pathways: One pathway depends on the MSF and ATP, and the other pathway depends on Hsp70 and is ATP independent. The yeast mitochondrial outer membrane contains at least four membrane receptors: Tom70, Tom37, Tom22 and Tom20. These receptors function as two sub-complexes: Tom70/Tom37 and Tom20/Tom22, corresponding to two transport pathways for mitochondrial precursor proteins. OM, mitochondrial outer membrane. + and -, positively and negatively charged amino acids.

in vitro import into mitochondria in an ATP dependent manner (Hachiya *et al.*, 1994).

Unlike Hsp70, MSF is a mitochondria-specific import chaperone and the stimulation of

in vitro import into rat mitochondria by MSF does not require the presence of Hsp70. In

yeast, two other cytosolic factors, Mft52 (Cartwright *et al.*, 1997) and nascent

polypeptide-associated complex (NAC) (George *et al.*, 1998) have been identified as

being involved in protein import into mitochondria. Both have been shown to deliver

CoxIV-DHFR, in which the N-terminal targeting sequence of CoxIV is fused to the

reporter dihydrofolate reductase, into yeast mitochondria *in vitro* (Cartwright *et al.*, 1997;

George *et al.*, 1998). Loss of NAC leads to a defect in mitochondrial targeting of CoxIV-

DHFR *in vivo*, and if both NAC and Mft52 are missing from yeast cells, inefficient

targeting of mitochondrial precursors leads to defects in both mitochondrial function and

morphology (George *et al.*, 1998).

In addition to cytosolic factors, import of proteins into mitochondria requires the

specific recognition of target signals by receptor components in the mitochondrial outer

membrane. Studies in yeast have identified several such membrane receptors: Tom70,

Tom37, Tom22 and Tom20 (Tom indicates the translocases in the mitochondrial outer

membrane and the numbers indicate molecular masses in kilodaltons) (for review, see

Neupert, 1997). These receptors function as two sub-complexes in the yeast

mitochondrial outer membrane: Tom37/Tom70 and Tom20/Tom22, corresponding to two

transport pathways for mitochondrial precursors. Different precursor proteins may use

different sub-complexes as import receptors. *In vitro* import experiments indicated that

the Tom37/Tom70 complex promotes the import of the ADP/ATP carrier, cytochrome c_1

and F_1-ATPase β subunit, but not import of an artificial fusion precursor containing

DHFR as a passenger (Hines *et al.*, 1990), whereas Tom20/Tom22 interacts with almost

all mitochondrial precursors including artificial fusion proteins (Moczko *et al.*, 1994).

These receptors show partial overlap in function. For example, yeast cells lacking

Tom20 are respiratory deficient, but this defect can be compensated for by

overexpression of Tom70 (Ramage *et al.*, 1993). However, Tom22 has been shown to be

essential for mitochondrial protein import and viability of yeast cells (Honlinger *et al.*,

1995). Both Tom20 and Tom22 possess a high abundance of negatively charged residues

on their cytosolic domains (Kiebler *et al.*, 1993). Experiments using fusion proteins such

as F_1F_0-ATPase subunit 9 fused to DHFR (Su9-DHFR) for mitochondrial import showed

that binding between Tom20/Tom22 and mitochondrial precursors is strongly mediated

by salt-sensitive electrostatic interactions, strongly suggesting that Tom20/Tom22,

through their cytosolic exposed acidic domains, interact with basic and amphiphilic

mitochondrial targeting signals (Haucke *et al.*, 1995). Recognition of preproteins by the

Tom37/Tom70 sub-complex may involve the mature part of mitochondrial precursors

(Haucke *et al.*, 1995). In contrast, less is known about mitochondrial outer membrane

receptors in plants. Until now, only a 23 kDa outer membrane protein from potato,

termed pTom20, has been identified as a membrane receptor for mitochondrial precursors

(Heins and Schmitz, 1996). Pretreatment of isolated potato mitochondria with antibody

against pTom20 reduced their import capacity by 30-40% (Heins and Schmitz, 1996). A

70 kDa heat-shock protein has been found in the mitochondrial outer membrane in plants,

although its function in protein import is still poorly understood (Mooney and Harmey,

1996). Some evidence suggests that protein import into mitochondria is conserved

between plants and yeast, as some plant proteins are efficiently imported into yeast

mitochondria (Chaumont *et al.*, 1990) and some yeast proteins are imported and correctly

processed in plant mitochondria (Schmitz and Lonsdale, 1989).

V. Chloroplast import machinery

In addition to mitochondria, plant cells contain chloroplasts to which proteins have

to be transported. The chloroplast import machinery has been well studied in the past and

several cytosolic factors and membrane receptors have been identified (for review, see

Keegstra and Cline, 1999). A model of the chloroplast import process is shown in Figure

5 (Soll and Tien, 1998). Like mitochondria, chloroplasts import most of their proteins

from the cytosol as precursor proteins which contain an amino-terminal targeting

sequence. Cytosolic chaperone Hsp70 is found to stimulate chloroplast import of some

precursors (Waegemann *et al.*, 1990). Interestingly, it has been reported that chloroplast

but not mitochondrial precursor proteins are phosphorylated by a cytosolic protein kinase

in an ATP dependent manner (Waegemann and Soll. 1996). The phosphorylation occurs

on serine or threonine residues within the targeting sequence such that phosphorylated

precursors are translocation-incompetent. Dephosphorylation is required to complete the

translocation process. Protein import into chloroplasts is initiated by specific binding of

the precursor proteins to the surface of chloroplasts followed by translocation of

precursor proteins across both the outer and inner envelope membranes (Keegstra and

Cline. 1999). Outer envelope membranes isolated from pea chloroplast could bind

precursor proteins (Soll and Waegemann. 1992), indicating a role in chloroplast import.

The outer envelope membrane contains at least three transport components: Toc86,

Toc75 and Toc34 (Toc indicates the tranlocases in the outer chloroplast membrane and
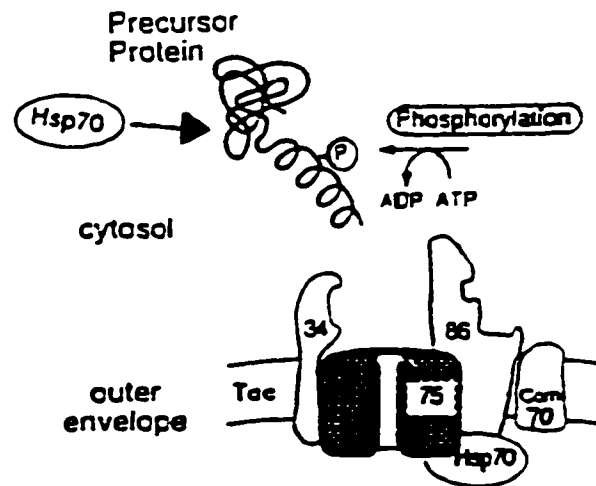
Figure 5: Model for protein import into chloroplast (Soll and Tien, 1998). Chloroplast precursor proteins are phosphorylated by a cytosolic protein kinase. The chloroplast outer envelope membrane contains at least three transport components: Toc86, Toc75 and Toc34, which form a trimeric complex in the outer membrane. Protein import into chloroplasts is initiated by specific binding of the precursor proteins to the surface of chloroplasts followed by translocation of precursor proteins across both the outer and inner envelope membranes.

the numbers indicate molecular masses in kilodaltons), which form a trimeric complex in the outer membrane (Waegemann and Soll, 1991). Toc86 is a GTP-binding protein and is highly protease sensitive (Kessler et al., 1994). When purified pea chloroplasts are incubated with the precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase oxygenase (preSSU) in label transfer cross-linking experiments, Toc86 is strongly labelled in an ATP-independent manner (Perry and Keegstra, 1994). In addition, binding of preSSU to chloroplasts is inhibited by the Fab fragment of anti-Toc86 antibody (Hirsch et al., 1994), suggesting that Toc86 serves as the main receptor in the chloroplast outer envelope membrane. Toc75 is proposed to form the translocation channel in the outer envelope membrane, as heterologously expressed and purified Toc75, reconstituted into liposomes, can form a voltage-gated transmembrane channel (Hinnah et al., 1997). Further experiments demonstrated that Toc75 interacts specifically with the transit sequence during the import process (Ma et al., 1996). Toc34 is also a GTP binding protein proposed to have regulatory functions in the early stage of preprotein translocation across the chloroplast envelope (Kessler et al., 1994). Recently, a direct cross-link interaction between Toc34 and preproteins destined for the chloroplast has been demonstrated in the absence of added ATP or GTP (Kouranov and Schnell, 1997). This observation suggests that Toc34 participates with Toc86 in the initial interaction of preproteins with the outer envelope membrane. Toc86 shows strong sequence similarity with Toc34 in its cytosolically exposed N-terminus and both of these proteins are located in close physical proximity to Toc75 (Ma et al., 1996). In addition, two Hsp70 homologues are found in the outer envelope membrane and are also thought

22

to be involved in protein translocation across the outer membrane (Schnell *et al.*, 1994;

Kourtz and Ko, 1997)

VI. Use of yeast two-hybrid system to detect protein-protein interaction

Protein-protein interactions have attracted much attention because they form the

basis of a wide variety of biochemical reactions and play an important role in protein

transport. The identification of proteins which interact with a known protein is essential

to elucidate the regulation and function of that protein. A number of biochemical and

genetic approaches have been developed to identify genes encoding interacting proteins

including co-immunoprecipitation, cross-linking and the yeast two-hybrid system.

The yeast two-hybrid system is a genetic assay to detect protein-protein interactions

in vivo (Fields and Song, 1989; Durfee *et al.*, 1993). It is based on the fact that some

yeast transcription factors, such as Gal4 protein, consist of two physically separable

domains: a DNA binding domain (DB) and a transcription activation domain (AT). The

DNA binding domain targets the transcription factor to the specific upstream activation

sequence (UAS) whereas the activation domain serves to initiate transcription of the

downstream genes. Neither of these domains alone is able to activate transcription, but

each domain continues to function when fused to other proteins.

In the yeast two-hybrid system, sequences encoding two domains of the Gal4p

transcription factor have been cloned into two different vectors: a DNA binding domain

vector and an activation domain vector. The cDNA sequence coding for the protein of

interest (the bait, in my case the lupin or *Arabidopsis* tRNA nucleotidyltransferase) is

fused to DB, whereas cDNAs coding for potential interacting proteins (in my case an

Arabidopsis cDNA library) are introduced into the AT vector. These two vectors are then co-transformed into a yeast strain containing the reporter genes, *lacZ* and *HIS3*, under the control of the UAS. If the interaction between the two proteins of interest occurs, the DB and AT domains are brought into close proximity and a functional transcription activator is reconstituted (Figure 6), resulting in transcription of the reporter genes. Expression of the *lacZ* gene can be detected using filter or liquid assays for β-galactosidase activity. The two-hybrid system not only allows for identification of proteins that interact, but also can be used to define and/or test the domain/residues necessary for the interaction of two proteins (Li and Fields, 1993). In the last few years, a large number of proteins from a variety of organisms have been identified using this method, including some cytosolic factors (Haizel *et al.*, 1997) and membrane receptors (Dingwall *et al.*, 1995), which are essential for protein transport.

VII. This work

As discussed above, a large number of proteins are involved in directing a protein to its final cellular location be that nucleus, mitochondrion or chloroplast. The enzyme, tRNA nucleotidyltransferase, is required for tRNA maturation in multiple compartments in plants. Based on sequence analysis of the proteins predicted from the lupin and *Arabidopsis* cDNAs, it appears that they contain sequences characteristic of both nuclear and mitochondrial and/or chloroplast targeting signals. Assuming that any of these targeting signals are real, there should be proteins which recognize and interact with them. In addition, there may be other proteins that interact with tRNA nucleotidyltransferase for some as yet unknown purpose distinct from targeting. In this
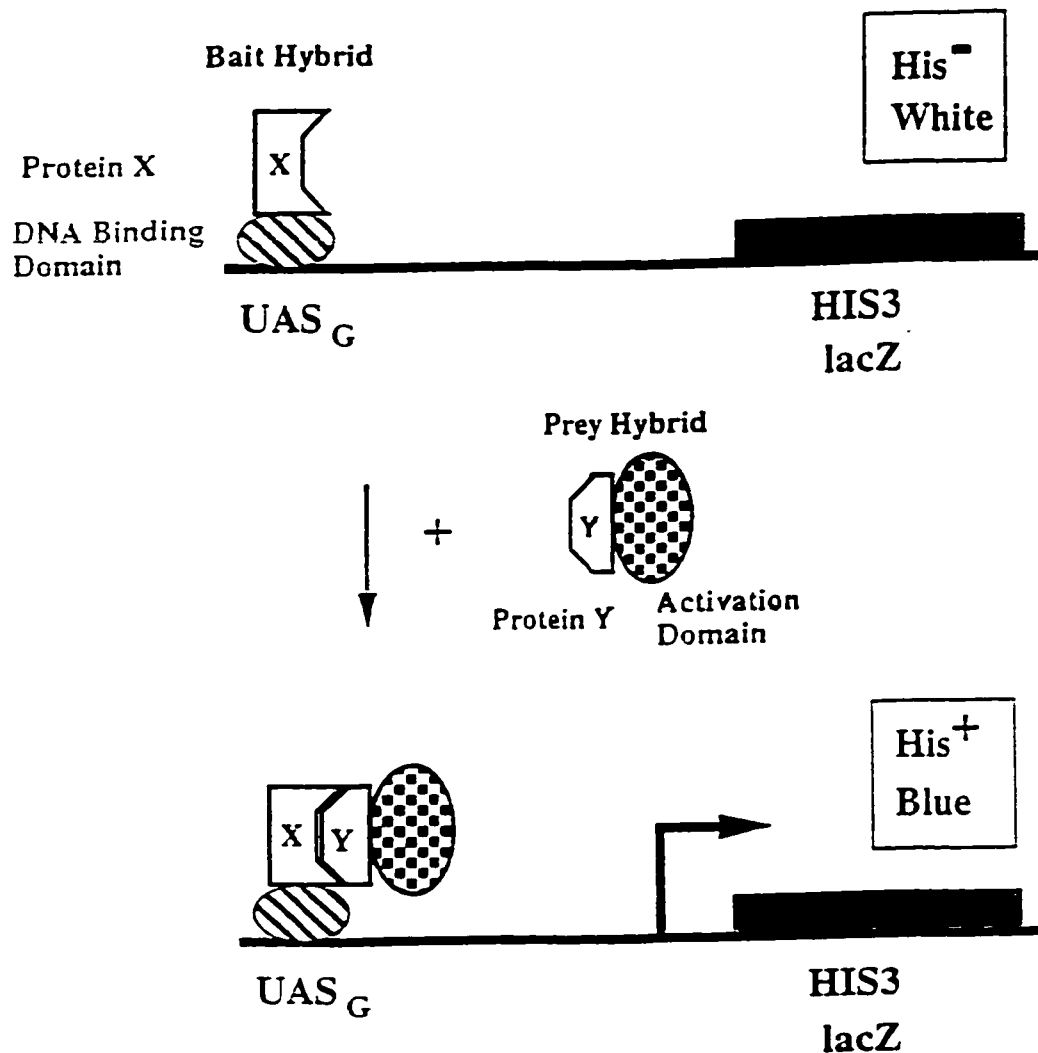
Figure 6: A schematic representation of the two-hybrid system: Two hybrid proteins are generated. The bait hybrid consists of protein X fused to a DNA binding domain while the prey hybrid consists of protein Y fused to an activation domain. The DNA binding domain binds to the specific upstream sequence (UAS). Interaction between protein X and Y brings the transcription activation domain to the UAS and activates expression of reporter gene (*HIS3* and *lacZ*) (Bai and Elledge, 1996).

work, I employed the yeast two-hybrid system (Fields and Song, 1989) to screen an *Arabidopsis* cDNA library, to find proteins which interact with lupin or *Arabidopsis* tRNA nucleotidyltransferase. The cDNA clones isolated by this process were sequenced and their potential functions predicted.

MATERIALS AND METHODS

I. Strains and growth media

Saccharomyces cerevisiae strains HF7C and NT33-5, Escherichia coli strains XL2-Blue and BNN132 were used in this work. Their genotypes are listed in Table 1. Yeast strain HF7C was kindly supplied by Pascale Gaudet. Yeast strain NT33-5 was constructed in our lab by Dr. Pam Hanic-Joyce. E. coli strain BNN132 was supplied with the Arabidopsis cDNA library by the Arabidopsis Biological Resource Center at the Ohio State University, and XL2-Blue was purchased from Stratagene. Growth media used in this work are listed in Table 2.

II. Construction of plasmids

The DNA binding domain vector pAS2 (Figure 7) was supplied by the Arabidopsis Biological Resource Center at the Ohio State University. This vector contains TRP1, a 2μ origin. and the ADH promoter driving expression of the GAL4 DNA-binding domain (amino acids 1-147) fused to a polylinker. Inserts of interest were cloned into the NdeI and SalI sites of the polylinker to construct the bait plasmids. All the constructs are listed in Table 3.

III. Generation of inserts

A) Polymerase chain reaction (PCR)

Templates and primers used to PCR-amplify the DNA fragments for cloning into plasmid pAS2 are listed in Table 3 and Table 4. All of the 5' end primers contain an

Table 1: Strains

Strain	Organism	Genotype	Reference
BNN132	*E. coli*	*endA1 gyrA96 thi hsdR17 supE44 relA1 Δ(lac-proAB) (F' traD36 proAB⁺ lacI^qZΔM15)*	Elledge *et al.*, 1991
XL2-blue	*E. coli*	*recA1 endA1 gyrA96 thi hsdR17 relA1 supE44 lac (F' proAB⁺ lacI^qZΔM15 Tn10 Tet^r Amy Cam^r)[a]*	Stratagene catalog 1999
HF7C	*S. cerevisiae*	MATa *ura3-52 his3-200 lys2-801 ade2-101 trp1-901 leu2-3,112 gal4-542 gal80-538* LYS2::GAL1-HIS3 URA3::(GAL4 17-mers)$_3$ CYC-lacZ	Fellotter *et al.*, 1994
NT33-5	*S. cerevisiae*	MATa *ccal-1 ade2-101 ura3-52 trp1-1 his3-200 lys2 leu2-3,112*	Shanmugam *et al.*, 1996

Table 2: Growth media

Organism	Growth medium	Recipe	References
E. coli	L-Broth	1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl (plates: 1.5% agar)	Sambrook *et al.*, 1989
E. coli	SOC	2% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.05% NaCl, 0.25 mM KCl, 0.01 M MgCl$_2$, 20 mM glucose.	Sambrook *et al.*, 1989
S. cerevisiae	YPD	1% Bacto-yeast extract, 2% peptone, 2% dextrose. (plates: 2% agar)	Sherman, 1991
S. cerevisiae	SC (synthetic complete)	20 µg/ml adenine, 20 µg/ml arginine, 100 µg/ml aspartic acid, 20 µg/ml histidine, 60 µg/ml leucine, 30 µg/ml lysine 20 µg/ml methionine, 50 µg/ml phenylalanine, 200 µg/ml threonine, 40 µg/ml tryptophan, 30 µg/ml tyrosine, 20 µg/ml uracil, 2% dextrose, 0.67% YNB* without amino acids (plates: 2% agar)	Bai and Elledge, 1996

* Yeast Nitrogen Base
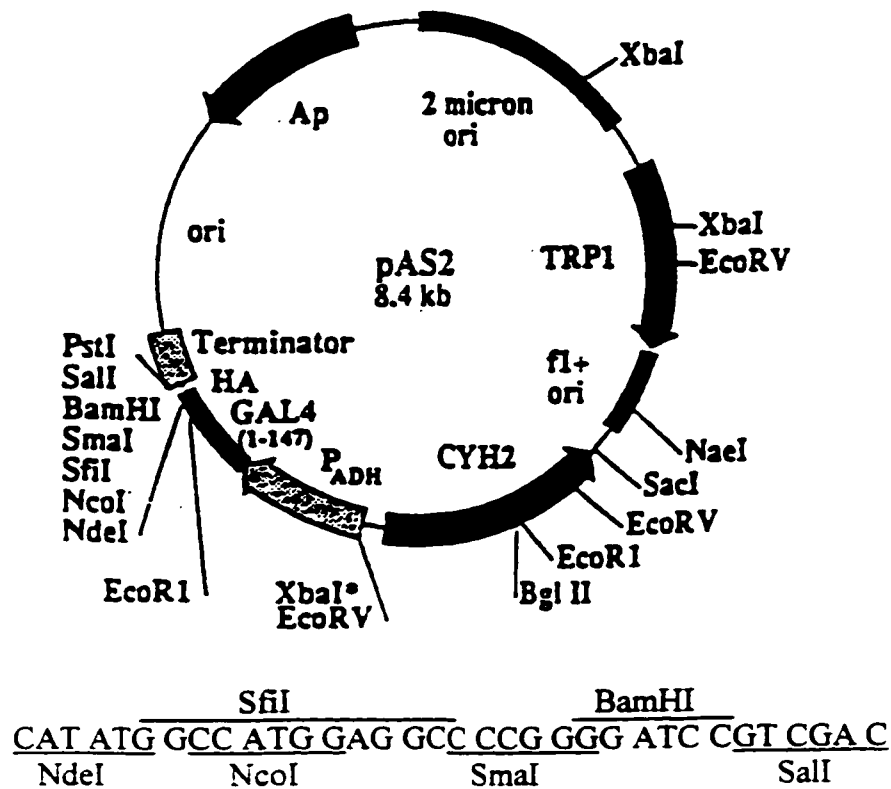
Figure 7: Map of plasmid pAS2 (Durfee *et al.*, 1993). This plasmid contains the sequences encoding the GAL4 DNA binding domain followed by a polylinker to facilitate the construction of bait hybrids. The reading frame of the polylinker is shown in coding triplets. In this work, cDNAs encoding the proteins of interest were cloned into the *Nde*I and *Sal*I sites of pAS2 to create the bait plasmids.

Table 3: Bait plasmids constructed in DNA binding domain vector pAS2

Plasmid	Insert	Insert size	PCR primers		PCR Templates
			Forward	Reverse	
pAS2-lupincca1	full length lupin *CCA* starting from ATG1	1900 bp	—	—	—
pAS2-lupincca2	full length lupin *CCA* cDNA starting from ATG58	1520 bp	CCA2NDE1	CCA3SAL1A	pET32a-lupincca[a]
pAS2-lupincca3	lupin *CCA* cDNA fragment from ATG1 to base 1080 (amino acid 344)	1050 bp	PET5	0.5CCA3A	pET32a-lupincca
pAS2-lupincca4	lupin *CCA* cDNA fragment from base 1060 (amino acid 342) to the stop codon	665 bp	CCAMNDE1	CCA3SAL1A	pET32a-lupincca
pAS2-aracca3	full length *Arabidopsis CCA* cDNA starting from ATG69	1690 bp	ARACCA3	ARACCA4	pZL1-aracca[b]
pAS2-aracca2	full length *Arabidopsis CCA* cDNA starting from ATG6	1880 bp	ARACCA5	ARACCA4	pZL1-aracca
pAS2-aracca1	full length *Arabidopsis CCA* cDNA starting from ATG1	1895 bp	ARAATG1EN	ARACCA4	pAS2-aracca2[c]

a: Plasmid pET32a-lupincca containing the lupin *CCA* cDNA was constructed by Fouad Karam in our lab.

b: Plasmid pZL1-aracca (122M21T7) containing the *Arabidopsis CCA* cDNA was supplied from the *Arabidopsis* EST Sequencing project by the *Arabidopsis* Biological Resource Center at the Ohio State University.

c: Plasmid pAS2-aracca2 was constructed in this work.

Table 4: Primers designed for PCR and DNA sequencing

A) PCR

Primer[a]	Sequence	Use	restriction site
CCA2NDE1	5'CCTCA*CATATG*TCT TCACACAAGG3'	5' end PCR primer for lupincca2	*Nde*I
PET5	5'CAGAAT*CATATG*A GACTAA3'	5' end PCR primer for lupincca3	*Nde*I
CCAMNDE1	5'TCTCAATTG*CATAT GT*CATGG3'	5' end PCR primer for lupincca4	*Nde*I
0.5CCA3A	5'CTAG*GTCGAC*TCAA AGGTTCCATGAGAT3'	3' end PCR primer for lupincca3	*Sal*I
CCA3SAL1A	5'TTGG*GTCGAC*TCAA CCAATCAAGAC3'	3' end PCR primer for lupincca2 and 4	*Sal*I
ARACCA3	5'GAACGCG*CATATG*A CGAATGTTGGA3'	5' end PCR primer for aracca3	*Nde*I
ARACCA5	5'TACTAAAA*CATATG* AGACTGTCTTCTCTTC CGATCAACACTCTCAT AAATCTCCCAAAATCT CTCTTTCTAATTTCTCC TT3'	5' end PCR primer for aracca2	*Nde*I
ARAATG 1EN	5'CGGAATT*CATATG*A TACTAAAAACCATGA GACTGTCTTCTC3'	5' end PCR primer for aracca1	*Nde*I
ARACCA4	5'ATTA*GTCGAC*AATG TTAGTGGA3'	3' end PCR primer for aracca1, 2 and 3	*Sal*I
CCA5SP	5'GAAACAG*ACTAGT*A TGAGAC3'	5' end PCR primer as positive control	*Spe*I
3'CCASAL	5'CCAACT*GTCGAC*TC AATGAT3'	3' end PCR primer as positive control	*Sal*I

33

B) DNA sequencing

Primer[b]	Sequence	Use
LibrA	5'GCGTTTGGAATCA CTACAGG3'	5' end primer for sequencing of cDNA clones
LibrB	5'GCACGATGCACAG TTGAAGT3'	3' end primer for sequencing of cDNA clones

a: Primers used for PCR were purchased from BioCorp Inc. (Montreal).
b: Primers used for DNA sequencing were kindly supplied by Pascale Gaudet.

*Nde*I site and all of the 3' end primers contain a *Sal*I site. The PCR reactions were performed in a PERKIN ELMER DNA thermal cycler. The reaction mixture contained 100 pmols of each primer, 10 ng of template DNA, 200 μM dNTPs, 1.5 mM MgCl$_2$, 1 μl of Taq DNA polymerase (MBI) and 1×Taq polymerase buffer in a final volume of 50 μl adjusted with water. The mixture was overlaid with 50 μl of mineral oil to avoid evaporation at high temperature. The PCR reaction was started at 95^{0}C for 5 min, followed by 30 cycles of denaturation at 95^{0}C for 1 min, annealing at 35^{0}C for 1 min, and extension at 72^{0}C for 1 min. The last cycle was finished at 72^{0}C for 10 min.

B) Purification of PCR products

After 5 μl of each PCR reaction mix had been set aside for agarose gel electrophoresis to confirm the sizes of PCR products generated, the remaining 45 μl were transferred to an Eppendorf tube containing 45 μl of water and 10 μl of 3 M sodium acetate. The mixture was subjected to phenol extraction twice with equal volumes of phenol and ether extraction twice with equal volumes of ether. The DNA present in the aqueous phase was precipitated by adding two volumes of 99% ethanol. The samples were placed at –70^{0}C for 30 min and centrifuged at 14 000 rpm in an Eppendorf microfuge for 30 min at 4^{0}C. The pellet was washed with 70% ethanol. desiccated, and resuspended in 10 μl of sterile water.

C) Restriction digestion

The restriction enzymes used in this work were supplied by MBI or Promega. The PCR products and plasmid pAS2 were digested with restriction enzymes *Nde*I and *Sal*I as

recommended by the supplier. The digestion mixtures were separated by 1.2% agarose gel electrophoresis at 80 volts for 2 hours.

D) Purification of DNA fragments from agarose gel

DNA fragments were purified from agarose gels by the phenol freeze-thaw method (Bewsey *et al.*, 1991). Briefly, a small piece of agarose gel containing the DNA fragment of interest was cut out and crushed in a 1.5 ml Eppendorf tube. 400 µl of phenol was added to this tube which was vortexed vigorously and placed at -70^0C for 30 min, followed by thawing at 37^0C for 10 min. Another 400 µl of phenol was added, the sample was vortexed, placed again at -70^0C for 30 min. and thawed at 37^0C for 10 min. To this, 100 µl of water and 50 µl of 3 M sodium acetate were added, the sample was vortexed and centrifuged at 14 000 rpm for 5 min at room temperature. The aqueous phase was collected. DNA was further extracted twice with phenol and twice with ether, precipitated with ethanol as described above, and resuspended in 10 µl of sterile water.

E) Ligation

Ligations were performed according to instructions supplied with the T4 DNA ligase (MBI). In each case, 300-500 ng of insert and vector were mixed with 0.5 µl of ligase in a total volume of 20 µl of 1×ligase buffer and incubated at 4^0C overnight.

IV. Plasmid DNA preparation

A) Preparation of competent *E. coli* cells and CaCl₂-mediated transformation

Competent cells were prepared according to the protocol described by Sambrook *et al.* (1989). A single colony of *E. coli* strain XL2-Blue (Table 1) was inoculated into 5 ml of LB medium (Table 2) and grown at 37^0C overnight with shaking at 225 rpm. The next morning, 2.5 ml of overnight culture were transferred to 200 ml of LB and incubated at 37^0C with shaking to reach an OD_{600} of 0.6-0.7. The cells were chilled on ice for 10 min, and harvested by centrifugation at 5000 rpm for 15 min at 4^0C using a JA-14 rotor in a Beckman J2HS centrifuge. The pellet was resuspended in 100 ml of 0.1 $MgCl_2$ and centrifuged again as above. The pellet was resuspended in 100 ml of 0.1 $MgCl_2$ again, left on ice for 25 min, and centrifuged again. The cells were resuspended in a 10 ml solution of 0.1M $CaCl_2$, 14% glycerol. Aliquots were incubated at 4^0C overnight and stored at -70^0C.

Transformation was achieved by mixing 5 µl of ligation mixture or 100 ng of plasmid DNA with 75 µl of competent XL2-Blue cells in a sterile Eppendorf tube. The mixture was left on ice for 15 min. The transformed cells were plated on LB agar plates containing ampicillin at 50 µg/ml and incubated overnight at 37^0C.

B) Transformation of *E. coli* by electroporation

High voltage electroporation was used when high transformation efficiency was needed. Electrocompetent cells were prepared as follows: 200 ml of low-salt LB (Table 2 but with 0.5% NaCl) were inoculated with 4 ml of a fresh overnight culture of *E. coli*

XL2-Blue and incubated at 37^0C with shaking to reach an OD_{600} of 0.5-0.7. The cells were chilled for 20 min and centrifuged at 5000 rpm for 15 min at 4^0C. The pellet was resuspended once in 200 ml of cold water, once in 100 ml of cold water, and once in 4 ml of 10% glycerol with the cells collected by centrifugation as above at each step. The cells were finally resuspended in 10% glycerol to a final volume of 1 ml and stored at -70^0C in 50 μl aliquots.

Library plasmids isolated from yeast HF7C cells were transformed into electro-competent XL2-Blue cells by electroporation. This was performed using a Gene Pulser apparatus (BioRad). The Gene Pulser was set at 25 μF and 2.5 kV, and the Pulser Controller was set to 200 Ω. In an ice-cold 0.2 cm electroporation cuvette, 50 μl of competent cells were mixed with 100 ng of plasmid DNA. The cuvette was placed in the apparatus, and the mixture was pulsed once. Next, 1 ml of SOC (Table 2) was added immediately to the cuvette. The cell suspension was transferred to a 1.5 ml Eppendorf tube and incubated at 37^0C for 1 hour. Aliquots were plated on LB plates with ampicillin (50 μg/ml) and incubated at 37^0C overnight.

C) Plasmid preparation

Fisher brand screw cap tubes (50 ml) containing 5 ml of LB with ampicillin at 50 μg/ml were inoculated with a single colony from transformation and incubated overnight at 37^0C with shaking at 225 rpm. Plasmid DNA was isolated from the cells using the alkaline lysis procedure described in Promega's Protocols and Applications Guide (1991). Plasmid DNA was dissolved in sterile water and stored at -20^0C.

V. Yeast two-hybrid system

A) Conversion of a λ-ACT cDNA library into a plasmid library

The *Arabidopsis* λ ACT cDNA library was constructed from total mRNA of 3-day old etiolated *Arabidopsis* seedlings by Kim and Theologis (1997) and supplied by the *Arabidopsis* Biological Resource Center at the Ohio State University. The total library contains 36×10^6 independent recombinants. The size distribution of cDNA clones is 0.6-2.5 kb with an average size of 1 kb. Conversion of the λ ACT phage vector to plasmid was performed according to the instruction supplied with the cDNA library. A single colony of *E. coli* BNN132 cells (Table 1) was inoculated into 2 ml of LB containing 0.2% maltose and incubated overnight at 37^0C with shaking at 225 rpm. Cells were centrifuged and resuspended in 2 ml of 10 mM $MgCl_2$. The phage (10^8) were incubated with 1 ml of this culture for 30 min at 30^0C without shaking. 2 ml of LB were added and the cells were incubated with shaking for 1 hour at 30^0C. Subsequently. 200 µl of cells were plated on each of 10 150-mm LB plates with ampicillin (50 µg/ml) and 0.2% glucose and incubated at 37^0C overnight. The ampicillin-resistant colonies were scraped from these plates. transferred into 3 l of LB with 50 µg/ml ampicillin, and grown overnight at 37^0C with shaking at 225 rpm. Plasmid DNA was prepared and purified from this culture by the method of alkaline lysis as described before with minor modifications. After desiccation and resuspension in 500 µl of 1×TE (10 mM Tris-HCl pH8. 1 mM EDTA), the plasmid DNA was subjected to phenol extraction twice followed by ether extraction twice. DNA was precipitated by adding one tenth volume of 3 M sodium acetate and two volumes of 99% ethanol. This was placed at -70^0C for 30 min

and centrifuged at 14 000 rpm for 30 min at 4°C. The pellet was washed with 70% ethanol, desiccated, resuspended in 200 μl of sterile 1×TE and stored at -20°C.

The plasmid pACT (Figure 8), excised from λ ACT, contains the ColE1 origin of replication and *bla* gene for selection in *E. coli*, and *LEU2*, 2μ origin, and the ADH promoter sequence for selection, replication, and expression in *Saccharomyces cerevisiae*. In addition, it contains the sequence for the GAL4 activation domain fused to a polylinker containing an *Xho*I site into which the cDNAs have been inserted.

B) Preparation of competent yeast cells

Competent yeast HF7C (Table 1) cells were prepared as follows: A single colony of yeast strain HF7C was inoculated into 5 ml of YPD medium (Table 2) and grown at 30°C overnight with shaking at 225 rpm. This culture was diluted in 50 ml of YPD to an OD$_{600}$ of 0.4-0.5 and incubated at 30°C with shaking to reach an OD$_{600}$ of 1.5-2.3. The cells were harvested by centrifugation at 5000 rpm for 5 min at room temperature using a JA14 rotor in a Beckmen J2HS centrifuge. The pellet was washed with 20 ml of sterile water and centrifuged again as before. Cells were resuspended in 1 ml of sterile 100 mM lithium acetate and transferred to a sterile Eppendorf tube. After centrifugation at 14 000 rpm, the cells were resuspended in sterile 100 mM lithium acetate to a final volume of 500 μl and incubated at 30°C for 15-30 min without shaking. Cells were stored at –70°C in 50 μl aliquots after adding glycerol to 15%.
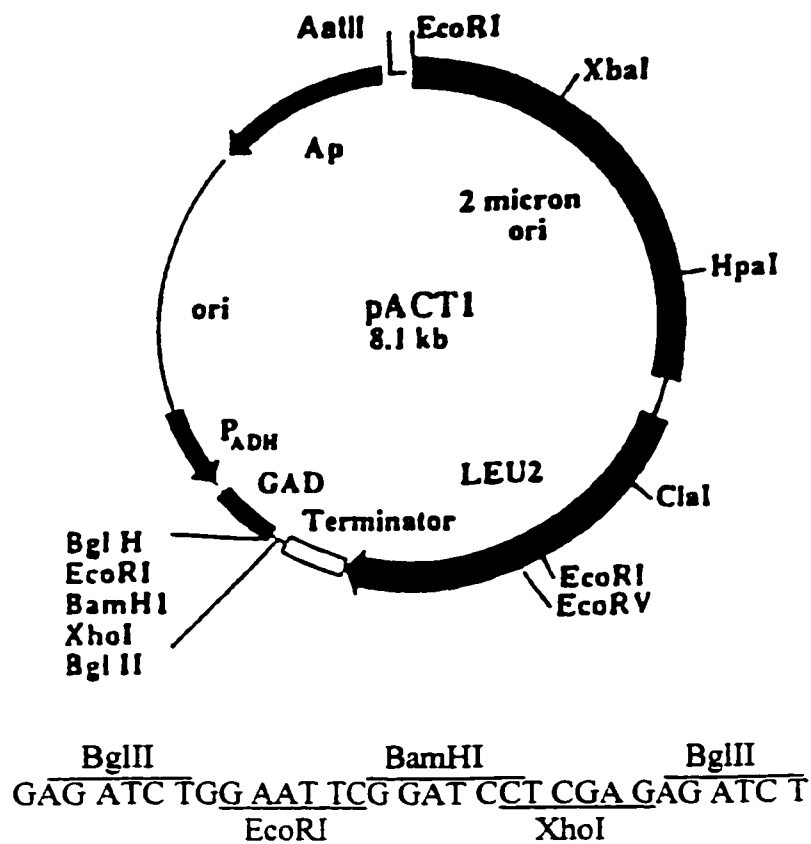
Figure 8: Map of plasmid pACT1 (Durfee *et al.*, 1993). This plasmid contains the sequences encoding the GAL4 transcription activation domain fused to a polylinker. The reading frame of the polylinker is shown in coding triplets. The *Arabidopsis* cDNA library was constructed in the *Xho*I site of the polylinker.

C) Yeast transformation

Yeast transformation was carried out using the lithium acetate procedure developed by Ito *et al.*, (1983) and modified by Gietz *et al.* (1992). A 50 µl aliquot of competent yeast cells prepared as above was centrifuged at 14 000 rpm for 5 seconds in a sterile Eppendorf tube and the resulting cell pellet was resuspended in 240 µl of 50% PEG (MW3350), to which 36 µl of 1.0 M lithium acetate, 50 µg of single-stranded herring sperm carrier DNA and 5 µg of plasmid DNA were added sequentially. The mixture was vortexed gently, incubated at 30^0C for 30 min and then heat shocked at 42^0C for 20 min. The transformed cells were centrifuged at 14 000 rpm, the pellet was resuspended in 400 µl of sterile water and 150 µl of this was plated on synthetic complete medium (Table 2) lacking appropriate amino acids and incubated at 30^0C until colonies appeared.

D) Yeast transformation to screen cDNA library

The method used for large scale library screening was described by Schiestl and Gietz (1989). Yeast HF7C cells hosting the bait plasmid were inoculated into 50 ml of SC-Trp medium and grown overnight at 30^0C with shaking at 225 rpm. This culture was transferred to 300 ml of pre-warmed YPD medium in a Fernbach flask to produce an OD_{600} of 0.1-0.2 and incubated at 30^0C with shaking at 225 rpm to an OD_{600} of 0.5-0.7. The cells were harvested by centrifugation at 5000 rpm using the JA-14 rotor in a Beckman J2HS centrifuge for 5 min at room temperature. All pellets were washed with 25 ml of sterile water and then pooled together in one centrifuge bottle, followed by centrifugation again as above. Cells were resuspended in sterile 100 mM lithium acetate to a final volume of 1.5 ml. In a sterile 50 ml tube, 6 ml of 40% PEG (MW3350)/0.1 M

LiAc/1×TE, 1 ml of cells, 25 µg of library DNA and 2 mg of single-stranded carrier

DNA were added sequentially, and vortexed vigorously. The mixture was incubated at

30^0C for 30 min with shaking at 200 rpm. Next, 700 µl of dimethylsulfoxide were added

and mixed gently, the cells were heat shocked at 42^0C for 15 min, chilled on ice for 2 min

and centrifuged at 5000 rpm for 5 min. The supernatant was discarded, the cell pellet

was resuspended in 50 ml of SC-Trp, -Leu, -His liquid medium in a sterile 250 ml flask

and incubated at 30^0C with shaking for 1 hour. The cells were centrifuged again as

before, the pellet was resuspended in a final volume of 7.5 ml of 1×TE and 500 µl of

cells were plated on each 150 mm SC-Trp, -Leu, -His plate and incubated at 30^0C for 5-7

days. The omission of tryptophan and leucine selected for transformation with the two

plasmids, bait and prey, while the omission of histidine selected for expression of the

reporter gene *HIS3*. To estimate transformation efficiency, 1/100 and 1/200 dilutions of

the transformation mixture were plated on SC-Trp, -Leu plates and incubated at 30^0C.

E) β-galactosidase filter assays

The filter assays for β-galactosidase activity were performed according to the

procedure described by Guarente (1983). In brief, a sterile 12.5 cm filter (Whatman No.

1) was laid onto a plate of yeast colonies. The filter was marked asymmetrically, lifted,

and submerged in liquid nitrogen for 5-10 seconds. The thawed filter was placed, colony

side up, over another 12.5 cm filter presoaked with 3 ml of a solution containing 100 ml

of Z-buffer (16.1 g/l $Na_2HPO_4.7H_2O$, 5.5 g/l $NaH_2PO_4.H_2O$, 0.75 g/l KCl, 0.246 g/l

$MgSO_4.7H_2O$, pH7), 0.27 ml of β-mercaptoethanol, and 3.0 ml of X-gal (10 mg/ml of 5-

bromo-4-chloro-3-indolyl-β-D-galactoside in N,N-dimethylformamide) in a Petri dish.

This was incubated at 30°C for the development of color. Colonies that turned blue within 24 hours were picked from the original plates or filters, streaked on SC-Trp, -Leu, -His plates, and tested again for β-galactosidase activity along with both negative and positive controls.

F) Isolation of library plasmids from yeast colonies

Plasmid DNA was extracted from double positive (His$^+$, lacZ$^+$) yeast colonies using the protocol described by Hoffman and Winston (1987). A single yeast colony was picked from a plate to inoculate 5 ml of SC-Leu medium and grown overnight at 30°C with shaking at 225 rpm. Cells from 1.5 ml of overnight culture were harvested by centrifugation at 14 000 rpm. The pellet was resuspended in 10 μl of sterile water, and mixed with 100 μl of lysis buffer (2.5 M LiCl. 50 mM Tris-HCl pH8.0, 4% Triton X-100 and 62.5 mM EDTA). An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to this mixture along with 0.2 g of acid-washed glass beads. This was vortexed vigorously for 2 min and centrifugation at 14 000 rpm for 1 min. The aqueous phase was transferred to a fresh microfuge tube. Plasmid DNA was precipitated with 2.5 volumes of 95% ethanol and placed at –20°C for 20 min. After centrifugation at 14 000 rpm for 15 min, the pellet was washed with 70% ethanol, desiccated, and resuspended in 20 μl of sterile water.

VI. Identification of cDNA clones in library plasmids

A) Yeast two-hybrid assay

Library plasmids containing inserts of interest were co-transformed into yeast HF7C with bait plasmids listed in Table 3 to eliminate false positives and to define possible domains responsible for interaction. The transformants were tested for β-galactosidase expression as described before.

B) DNA sequencing

DNA sequencing was performed using the T7 DNA sequencing kit Version 2.0 according to the manufacturer's protocol (USB, Pharmacia). Primers used for sequencing are listed in Table 4. In brief, 1.0 μl of 2 M NaOH and 1.0 μl of 2 mM EDTA were added to 10 μl of DNA (4-5 μg) dissolved in water. The denaturing mixture was vortexed, centrifuged briefly, and left at room temperature for 30 min. To this, 1.2 μl of 3 M sodium acetate and 35 μl of 99% ethanol were added. The mixture was placed at -70^0C for 30 min. DNA was pelleted by centrifugation at 14 000 rpm for 30 min at 4^0C, washed with 70% ethanol, dried and resuspended in 7 μl of water. Annealing was achieved by adding 2 μl of 5X sequencing buffer and 1 μl of primer (1 pmol/μl) to the denatured DNA in an Eppendorf tube. The mixture was heated at 65^0C for 2 min and cooled slowly to 30^0C. The labelling reaction mix was prepared according to the manufacturer's instructions and added to 10 μl of the annealed template-primer. This mixture was incubated at room temperature for 2 min and 3.5 μl aliquots were transferred to the Eppendorf tubes containing 2.5 μl of termination mix (pre-warmed at 37^0C).

45

Incubation was continued for another 3 min and the reaction was stopped by adding 4 µl of stop solution. Samples were boiled 2 min prior to loading onto a 6% acrylamide sequencing gel (7M urea, 5.7% acrylamide, 0.3% bis-acrylamide, 1X TBE (50 mM Tris, 50 mM Boric acid, 1 mM EDTA)). Sequencing gels were prerun for 30 min at 2000 V. Samples (3 µl) were loaded onto the sequencing gel. Electrophoresis was carried out at a constant voltage of 2000 V for 5-16 hours. After electrophoresis, gels were dried under vacuum for 2 hours at 80°C and then exposed to X-ray film (Fuji).

Additional DNA sequencing was carried out at the Center for Structural and Functional Genomics at Concordia University using the Beckman CEQ 2000 DNA sequencing machine according to the manufacture's instructions without any modifications or commercially by BioS&T (Lachine, Quebec).

RESULTS

I. Sequencing of cDNA encoding *Arabidopsis* tRNA nucleotidyltransferase

A cDNA (122M21T7) encoding *Arabidopsis* tRNA nucleotidyltransferase in the vector pZL1 was supplied by the *Arabidopsis* Biological Resource Center at Ohio State. This cDNA was sequenced completely in this work. Translation of this cDNA showed an open reading frame of 1764 bp starting from nucleotide 1, extending to a TGA at nucleotide 1765, and potentially encoding 588 amino acids (Figure 9). Because the open reading frame of this cDNA extends for more than 40 amino acids upstream of the first in-frame start codon, we then compared it with the *Arabidopsis CCA* gene sequence from the *Arabidopsis* database. Figure 10 shows part of an alignment of the *Arabidopsis* cDNA and gene sequences. This comparison reveals that the open reading frame can be extended upstream of the cDNA so that it contains two additional in-frame start codons (referred to as ATG1, ATG6). The first in-frame ATG encoded by the cDNA sequence corresponds to ATG69 in the extended open reading frame. Stop codons in all three reading frames are found upstream of the first potential start codon. A potential TATA motif required for transcription initiation (Mukumoto *et al.*, 1993) is also found upstream of the first start codon (Figure 10).

II. PCR

The polymerase chain reaction was performed to amplify DNA fragments containing appropriate restriction sites for insersion into the bait plasmid. The PCR products were electrophoresed on a 1.2% agarose gel and showed the expected products.

47

```
                          CTCTTTCTAATTTCTCCTTTTCGCTTCCGTAACCTTAATCGTTCCCTTA    49
                            L  F  L  I  S  P  F  R  F  R  N  L  N  R  S  L
CCGTCGCGTCTCGCATTTCATCAACGCTTCTTCGCGTTTCCGGTGTCTCGTCTCGTCCGT   109
  T  V  A  S  R  I  S  S  T  L  L  R  V  S  G  V  S  S  R  P
GTGGTTACTGGTTCTCGACGAACGCGGCGATGACGAATGTTGGAGAGGAAGATAAGCAAT   169
  C  G  Y  W  F  S  T  N  A  A  M  T  N  V  G  E  E  D  K  Q
CGATTCCTTCAATCGAATTGAAGGAGAATATTGAGCTCACTGACAAAGAACGGAAAATTT   229
  S  I  P  S  I  E  L  K  E  N  I  E  L  T  D  K  E  R  K  I
TCGATCGGCTTCTTAGTACTCTCCGTTACTGTAACCTTGACACACAGCTTCGTGTCGCCG   289
  F  D  R  L  L  S  T  L  R  Y  C  N  L  D  T  Q  L  R  V  A
GTGGTTGGGTCCGTGATAAGCTGCTAGGGAAAGAGAGTGATGATATTGATATAGCTATTG   349
  G  G  W  V  R  D  K  L  L  G  K  E  S  D  D  I  D  I  A  I
ATAATATGTCTGGAAGTGAATTTCTTGATAAGTTTAAGGAATATTTATCTTCGAGAGATG   409
  D  N  M  S  G  S  E  F  L  D  K  F  K  E  Y  L  S  S  R  D
AAGAAGTACAAGGCGACACTGTTATTGAAAGGAACCCTGACCAATCTAAACACTTGGAGA   469
  E  E  V  Q  G  D  T  V  I  E  R  N  P  D  Q  S  K  H  L  E
CAGCAAAGTTGCGGATTTATGACCAGTGGATAGATTTTGTTAACTTGAGAAGTGAAGAAT   529
  T  A  K  L  R  I  Y  D  Q  W  I  D  F  V  N  L  R  S  E  E
ACACAGAGAATAGTCGTATTCCTACAATGAAATTTGGCACCGCAAAAGATGATGCTTTTC   589
  Y  T  E  N  S  R  I  P  T  M  K  F  G  T  A  K  D  D  A  F
GTAGAGATTTAACCATAAATAGTTTATTCTACAATATCAACTCTGGCGCAGTAGAAGACC   649
  R  R  D  L  T  I  N  S  L  F  Y  N  I  N  S  G  A  V  E  D
TTACAGAAAGAGGCATCGATGACCTCAAGTCCGGAAAAATTGTTACACCCTTGCCTGCAA   709
  L  T  E  R  G  I  D  D  L  K  S  G  K  I  V  T  P  L  P  A
AAGCCACATTTTTGGATGATCCTTTGCGAGTTCTCCGGGCTGTTCGCTTTGGTGCAAGGT   769
  K  A  T  F  L  D  D  P  L  R  V  L  R  A  V  R  F  G  A  R
TTGGTTTTACTTTGGATGAAGAACTAAAAGAAGCTGCTTCAAGTGAGGAAGTAAGGGTAG   829
  F  G  F  T  L  D  E  E  L  K  E  A  A  S  S  E  E  V  R  V
CTCTTGGAGAAAAAATTAGCAGAGAGCGGATTGGGAATGAGATTGACCTGATGATATCTG   889
  A  L  G  E  K  I  S  R  E  R  I  G  N  E  I  D  L  M  I  S
GGAATGGACCAGTTTCAGCGGTTACCTATCTTTCTGACCTGAAACTATTTAGTGTTGTCT   949
  G  N  G  P  V  S  A  V  T  Y  L  S  D  L  K  L  F  S  V  V
TTGCCCTTCCCTCTTCAGCTGAGCCTTCACCACCAGAAAATTGTGGCAGCCTCTCCCAAT  1009
  F  A  L  P  S  S  A  E  P  S  P  P  E  N  C  G  S  L  S  Q
CCTACTTGGAAGCTATGTGGAGCTTACTTAAAACACCTCGGCCAGGAAAGTTCAGTGGTG  1069
  S  Y  L  E  A  M  W  S  L  L  K  T  P  R  P  G  K  F  S  G
AACAAAGACGGCTTGCTCTCTATGCTGCTATGTTTCTCCCTTTTAGGAAAACTGTATACA  1129
  E  Q  R  R  L  A  L  Y  A  A  M  F  L  P  F  R  K  T  V  Y
AGGACACTAAGGGCAAATCGATTCCTGTTGTCAACCACATTTTTAAATTCTCCATGAAGA  1189
  K  D  T  K  G  K  S  I  P  V  V  N  H  I  F  K  F  S  M  K
GGAAGACCAGTGACGCTGAAACAGTTATGAATATACATCAGACCACGGAGAGATTTCGTT  1249
  R  K  T  S  D  A  E  T  V  M  N  I  H  Q  T  T  E  R  F  R
CGTTGATTCCTTCCCTTGAAGTGAAGAAAGATGTAGAACTGGACGAGCTAACTTGGGCCG  1309
  S  L  I  P  S  L  E  V  K  K  D  V  E  L  D  E  L  T  W  A
CTGATATTCTTGAGCATTGGAAATCCATCACCCTTAATGAGCCAGTGATCCCAGCAACTT  1369
  A  D  I  L  E  H  W  K  S  I  T  L  N  E  P  V  I  P  A  T
CAAAAATAAGAGTACTCACAGGGTTTCTACTTAGAGATATTAAAGACTTCTGGCGTGTTT  1429
  S  K  I  R  V  L  T  G  F  L  L  R  D  I  K  D  F  W  R  V
CTCTGTTAACATCTTTGTTATTGTCTGCTACGGTTGACGGTAGCAACGATCATCAAGATA  1489
  S  L  L  T  S  L  L  L  S  A  T  V  D  G  S  N  D  H  Q  D
TCGGGCAGCTGGACTTCCAACTGGAAAGGATGCGAGAAACTTATCTGACAGTTGAGGCCA  1549
  I  G  Q  L  D  F  Q  L  E  R  M  R  E  T  Y  L  T  V  E  A
CCATTCATGAACTGGGTCTTGATAAGATCTGGGATGCAAAGCCTCTAGTCAATGGCAGGG  1609
  T  I  H  E  L  G  L  D  K  I  W  D  A  K  P  L  V  N  G  R
```

```
AGATAATGCAAATTGCTGAGCTGAAGGGAGGATCCCGGCTCATCCGTGAATGGCAACAGA 1669
E   I   M   Q   I   A   E   L   K   G   G   S   R   L   I   R   E   W   Q   Q
AGCTGCTCACTTGGCAGCTGGCTTACCCCAACGGCACCGCAGAGGAGTGCAAGGAGTGGA 1729
K   L   L   T   W   Q   L   A   Y   P   N   G   T   A   E   E   C   K   E   W
TGAGAGATATCAAAGCTAAACGACAAAGGATAGAGTGACTTAAGAATCTAGTGGCTAGTT 1789
M   R   D   I   K   A   K   R   Q   R   I   E   *   L   K   N   L   V   A   S
GTGCTGTGTAGTCTTATAAAAATCCACTAACATTGTAGATTAATTTAAACTCAGTTTTGA 1849
C   A   V   *   S   Y   K   N   P   L   T   L   *   I   N   L   N   S   V   L
CAAAAAAAAAAAAAAAAAG
```

Figure 9: Nucleotide sequence and predicted amino acid sequence of a cDNA encoding *Aradibopsis* tRNA nucleotidyltransferase. Standard one letter abbreviations for amino acids are used. The first in-frame ATG is bold. * indicates in-frame stop codons. The putative polyadenylation signal is underlined. The first nucleotide at the 5' end and the last nucleotide at the 3' end are linked to the adaptor sequences.

49

```
CCAGENE    -                A TGA GTC GGG ATG TGT TTT AAT TTG TGC TAA   31
CCAGENE    - TAC AAA TTT GTG TTG TAG TCT CTC TTT GTA CGT ACG GGT   70
CCAGENE    - GAT CGC TAC TTT TGT TGG TAT ACA TGA TCT AAC GCA TGC  109
CCAGENE    - GTG TAA ATC TAT TTA AAG ACC TAT ACA TAG TTT CGT AGG  148
CCAGENE    - TTT CAC GTA AAG TCT TCA ATG GTT GAA CTA AAT AGT GTT  187
CCAGENE    - ACT CAT AAT CAC CTA TCC ATT TGC ACC AGA TGA AAG AGA  226
CCAGENE    - CGT ATT CAT TTT CTT TTC CCG CTT AAA ATT ATA TAT TGA  265
CCAGENE    - ATT GAT CTA AAT AAC CGA TTC ACC AAT TGT ATA TCT TAG  304
CCAGENE    - TAA AGG AAA TTG GGC CAC TAT TGG GCC TAA TAA CCA TTC  343
CCAGENE    - TAA CTT TTA CGG ATC CGA CTT AGC CCA AAC ACA TCG CCG  382
CCAGENE    - GTT CAG AAC GCT GCC GTA AAT CGC TGC ATC GGA AAA AGC  421
CCAGENE    - ACG ATG ATA CTA AAA ACC ATG AGA CTG TCT TCT CTT CCG  460
CCAGENE    - ATC AAC ACT CTC ATA AAT CTC CCA AAA TCT CTC TTT CTA  499
                                                   ||| ||| |||
CCASEQ     -                                         CTC TTT CTA    9

CCAGENE    - ATTTCTCCTTTTCGCTTCCGTAACCTTAATCGTTCCCTTACCGTCGCGTCT  550
             |||||||||||||||||||||||||||||||||||||||||||||||||||
CCASEQ     - ATTTCTCCTTTTCGCTTCCGTAACCTTAATCGTTCCCTTACCGTCGCGTCT   60

CCAGENE    - CGCATTCATCAACGCTTCTTCGCGTTTCCGGTGTCTCGTCTCGTCCGTGT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||
CCASEQ     - CGCATTCATCAACGCTTCTTCGCGTTTCCGGTGTCTCGTCTCGTCCGTGT  110

CCAGENE    - GGTTACTGGTTCTCGACGAACGCGGCGATGACGAATGTTGGAGAGGAAGAT  650
             |||||||||||||||||||||||||||||||||||||||||||||||||||
CCASEQ     - GGTTACTGGTTCTCGACGAACGCGGCGATGACGAATGTTGGAGAGGAAGAT  160

CCAGENE    - AAGCAATCGATTCCTTCAATCGAATTGAAGGAGAATATTGAGCTCACTGAC  700
             |||||||||||||||||||||||||||||||||||||||||||||||||||
CCASEQ     - AAGCAATCGATTCCTTCAATCGAATTGAAGGAGAATATTGAGCTCACTGAC  210

CCAGENE    - AAAGAACGGAAAATTTTCGATCGGCTTCTTAGTACTCTCCGTTACTGTAAC  750
             |||||||||||||||||||||||||||||||||||||||||||||||||||
CCASEQ     - AAAGAACGGAAAATTTTCGATCGGCTTCTTAGTACTCTCCGTTACTGTAAC  260

CCAGENE    - CTTGACACACAGCTTCGTGTCGCCGGTGGTTGGGTCCGTGATAAG
             |||||||||||||||||||||||||||||||||||||||||||||
CCASEQ     - CTTGACACACAGCTTCGTGTCGCCGGTGGTTGGGTCCGTGATAAG
```

Figure 10: Alignment of part of the *Arabidopsis* gene (CCAGENE) and cDNA (CCASEQ) sequences. The gene sequence shown here starts from a nucleotide which is 490 nucleotides upstream of the beginning of the cDNA sequence and ends at the first likely intron downstream of the beginning of the cDNA sequence. The gene sequence upstream of the cDNA sequence is shown in coding triplets in frame with the predicted cDNA open reading frame. Three in-frame ATGs in the gene sequence and one in-frame ATG in the cDNA sequence are bold. Stop codons upstream of the first start codon in the gene sequence are underlined. The TATA-like motif of the probable promoter region is double underlined.

50

The approximate sizes of PCR products lupincca2, lupincca3, and lupincca4 are 1500 bp, 1000 bp, and 650 bp (Figure 11), while the expected sizes were 1532 bp, 1061 bp, 682 bp, respectively. The approximate sizes of PCR products aracca1, aracca2 and aracca3 are 1900 bp, 1900 bp, and 1700 bp (Figures 12 and 13), while the expected sizes were 1921 bp, 1905 bp, 1705 bp, respectively.

III. Construction of bait plasmids

The bait plasmids constructed in this work are list in Table 3.

IV. Complementation of the yeast *CCA1* mutation

Yeast strain NT33-5 carries a temperature-sensitive mutation in the tRNA nucleotidyltransferase gene. To determine whether the constructed plasmids pAS2-lupincca1, pAS2-aracca2 and pAS2-aracca3 could express functional proteins as GAL4 fusions to complement this mutation, they were transformed individually into NT33-5 cells and grown on synthetic complete medium lacking tryptophan (SC-Trp) at the permissive temperature (20^0C). Plasmid pAS2 without insert was also transformed into NT33-5 as a negative control. The absence of tryptophan selected for the transformed plasmids. The single colonies were patched onto new SC-Trp plates, grown at 20^0C, and then replica-plated on two sets of YPD and SC-Trp plates. One set of plates was incubated at the non-permissive temperature of 37^0C, while the other was incubated at the permissive temperature of 20^0C.

As shown in Table 5, cells containing the inserts of lupincca1, aracca2, or aracca3 were able to grow on both YPD and SC-Trp plates at both 20^0C and 37^0C. However,
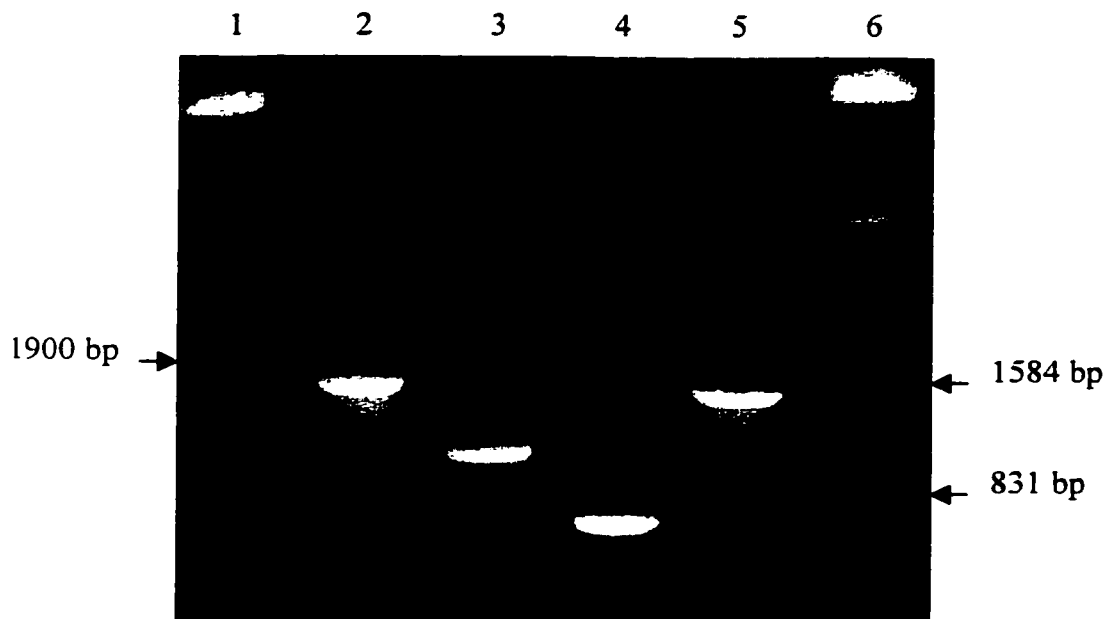
Figure 11: PCR products lupincca2, lupincca3 and lupincca4 separated on a 1.2% agarose gel. For all PCR reactions, pET32a-lupincca1was used as template.

Lane 1 and 6: λ EcoRI-HindIII DNA marker.

Lane 2: Full length lupin cDNA amplified with primers CCA5SP and 3'CCASAL (Table 4) as positive control.

Lane 3: lupincca3 amplified with primers PET5 and 0.5CCA3A (Table 4).

Lane 4: lupincca4 amplified with primers CCAMNDE1 and CCA3SAL1A (Table 4).

Lane 5: lupincca2 amplified with primers CCA2NDE1 and CCA3SAL1A (Table 4).
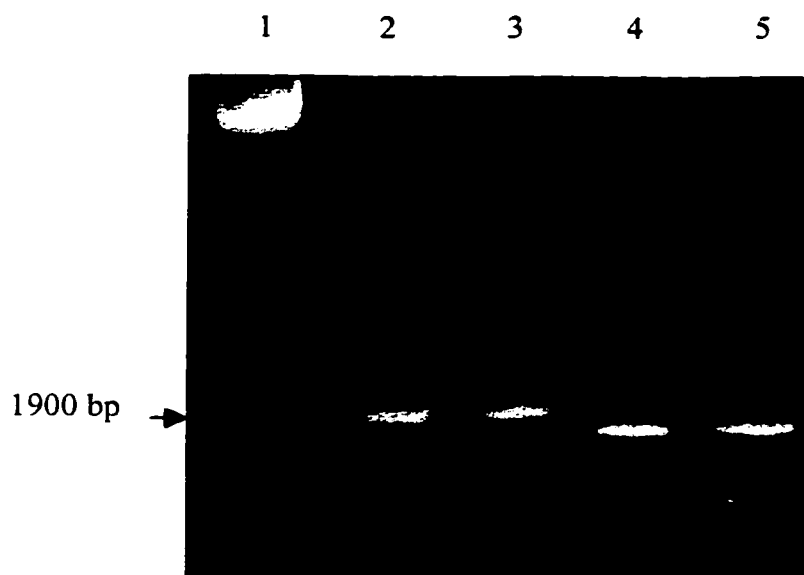
Figure 12: PCR products aracca2 and aracca3 separated on a 1.2% agarose gel.
Lane   1: λ EcoRI-HindIII DNA marker
Lane 2, 3: aracca2. PCR template pZL1-aracca amplified with primers ARACCA5
            and ARACCA4 (Table 4).
Lane 4, 5: aracca3. PCR template pZL1-aracca amplified with primers ARACCA3
            and ARACCA4 (Table 4).
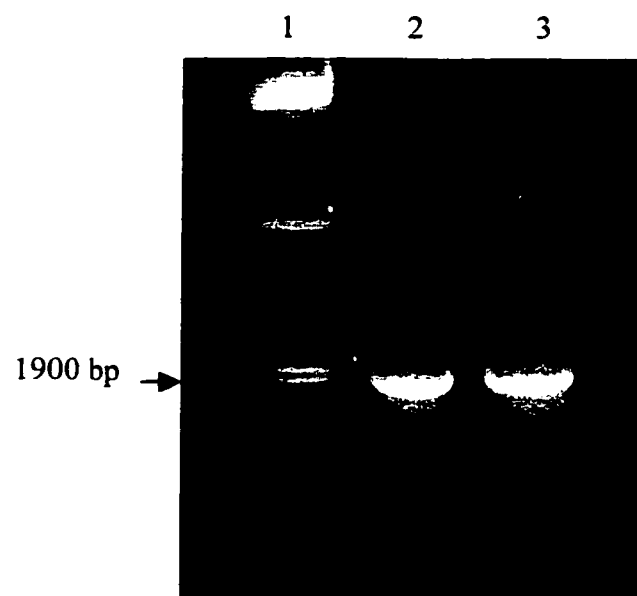
Figure 13: PCR product aracca1 separated on a 1.2% agarose gel.
Lane   1: λ *EcoRI-Hind*III DNA marker
Lane 2, 3: aracca1. PCR template pAS2-aracca2 amplified with
           primers ARAATG1EN and ARACCA4 (Table 4).

Table 5: Complementation of the yeast *ccal-1* mutation

Plasmid	20⁰C		37⁰C	
	YPD	SC-Trp	YPD	SC-Trp
pAS2	+	+	--	--
pAS2-lupinccal	+	+	+	+
pAS2-aracca2	+	+	+	+
pAS2-aracca3	+	+	+	+

Plasmids pAS2, pAS2-lupinccal, pAS2-aracca2, or pAS2-aracca3 were transformed individually into yeast strain NT33-5 carrying a temperature-sensitive mutation in the *CCA1* gene and grown on SC-Trp plates at the permissive temperature (20⁰C). Single colonies were patched onto new SC-Trp plates, grown at 20⁰C again, and then replica-plated onto two sets of YPD and SC-Trp plates. One set was incubated at the non-permissive temperature (37⁰C) and the other at the permissive temperature (20⁰C).

+: Growth. --: No growth.

cells harbouring plasmid pAS2 without insert can not grow at 37 ^{0}C on either YPD or SC-Trp plate. Therefore, all three plasmid constructs are able to express proteins capable of complementing the temperature-sensitive mutation in the yeast *CCA1* gene.

V. Yeast two-hybrid system

To identify proteins which could interact with lupin or *Arabidopsis* tRNA nucleotidyltransferase, the yeast two-hybrid system was employed to screen the *Arabidopsis* cDNA library.

A) Screening the cDNA library with bait plasmid pAS2-lupinccal

Lupinccal contains a full length lupin *CCA* cDNA starting from ATG1 (Table 3). Using pAS2-lupinccal which expresses the GAL4 DNA-binding domain fusion protein as a bait plasmid, approximately 22×10^6 cDNA clones from the *Arabidopsis* cDNA library were screened on synthetic complete medium (Table 1) lacking leucine, tryptophan and histidine (SC-Trp, -Leu, -His). Colonies grew within 3-5 days, indicating expression of the reporter gene *HIS3*. Lift assays using β-galactosidase yielded blue colonies if the other reporter gene, *lacZ*, was expressed. This double screen identified 13 colonies which express both the *HIS3* and *lacZ* genes. These colonies were streaked for single colonies and were positive again in the second β-galactosidase lift assay. Library plasmid DNAs were extracted from all 13 positives. They were co-transformed individually into yeast HF7C with pAS2-lupinccal. However, repeating the β-galactosidase assay yielded only one positive, clone 108 (Table 6A).

Table 6: Yeast two-hybrid results[a]

A. Library screening with plasmid pAS2-lupinccal

cDNA clone	Library screening with pAS2-lupinccal	Second test with pAS2-lupinccal[b]	Test with pAS2-lupincca2[b]	Test with pAS2-aracca2[b]
108	Blue[c] ++	Blue ++	--	--

B. Library screening with plasmid pAS2-aracca2

cDNA clone	Library screening with pAS2-aracca2	Second test with pAS2-aracca2[b]	Test with pAS2-araccal[b]	Test with pAS2-aracca3[b]
14A	Blue ++	Blue ++	Blue ++	--
19A	Blue ++	Blue ++	Blue ++	--
22A	Blue +	Blue +	Blue +	--
23A	Blue ++	Blue ++	Blue ++	--
27A	Blue +	Blue +	Blue +	--
39A	Blue +	Blue +	Blue +	--
49A	Blue ++	Blue ++	Blue ++	--
54A	Blue +	Blue +	Blue +	--
68A	Blue ++	Blue ++	Blue ++	--
75C	Blue ++	Blue ++	Blue ++	--

a: Yeast strain HF7C was used as host strain in this work.

b: Library plasmids were extracted from the yeast colonies expressing *HIS3* and *lacZ* genes and co-transformed individually into yeast HF7C with the bait plasmid indicated.

c: Expression of *lacZ* gene resulted in blue colonies in the β-galactosidase lift assay.

++: Bright blue, strong interaction. +: Blue, medium interaction. --: No interaction.

B) Screening the cDNA library with bait plasmid pAS2-aracca2

Aracca2 contains the full length *Arabidopsis CCA* cDNA plus the sequence coding

for 63 additional amino acids from ATG6 to ATG69 (Table 3). Of a total of 23 ×10^6

transformants screened on SC-Trp, -Leu, -His plates with the bait plasmid pAS2-aracca2,

11 double positive (*HIS*+ and *lacZ*+) colonies were obtained. They were streaked for

single colonies and subjected to β-galactosidase assay again. Library plasmids were

extracted from all 11 putative positives and co-transformed individually into yeast HF7C

with pAS2-aracca2. The β-galactosidase assay yielded 10 *His*+ and *lacZ*+ clones (Table

6B).

In summary, 11 putative positive cDNA clones were obtained using bait plasmids

pAS2-lupinccal and pAS2-aracca2 to screen the *Arabidopsis* cDNA library (Table 6).

Clones 14A, 19A, 23A, 49A, 68A, 75C and 108 appear to interact strongly with the bait

hybrids as indicated by bright blue colonies in the β-galactosidase lift assay (indicated

with ++ in Table 6), whereas clones 22A, 27A, 39A and 54A appear to interact with bait

hybrids less strongly (indicated with + in Table 6).

C) Screening putative positive cDNA clones with other bait plasmids

Using bait plasmid pAS2-lupinccal to screen the *Arabidopsis* cDNA library

resulted in only one putative positive, clone 108. To determine whether clone 108 could

interact with bait plasmid pAS2-aracca2 in the yeast two-hybrid system, it was co-

transformed into yeast HF7C with pAS2-aracca2. However, it failed to activate

expression of the reporter genes (Table 6A). Clone 108 was also co-transformed into

yeast HF7C with plasmid pAS2-lupincca2, pAS2-lupincca3 or pAS2-lupincca4

individually. Lupincca2 contains a lupin cDNA starting from ATG58 (Table3), whereas

lupincca3 and lupincca4 are truncated lupin cDNA fragments from ATG1 to base1080

(amino acid 344) and from base 1060 (amino acid 342) to the stop codon, respectively

(Table 3). However, all of these gave negative results in the yeast two-hybrid system.

Screening the cDNA library with bait plasmid pAS2-aracca2 resulted in 10 putative

positive cDNA clones. They were co-transformed individually into yeast HF7C with bait

plasmid pAS2-lupincca1, but all of them gave negative results in the yeast two-hybrid

system.

The 10 putative positive clones identified from library screening with pAS2-

aracca2 were tested individually for hybridization with bait plasmids pAS2-aracca1 or

pAS2-aracca3 in the yeast two-hybrid system. Aracca1 and aracca3 contain *Arabidopsis*

CC4 cDNAs starting from ATG1 and ATG69, respectively (Table 3). Transformants

were assayed for β-galactosidase expression and the results of these are summarized in

Table 6B. When plasmid pAS2-aracca1 was used as bait, all of the 10 cDNA clones

grew on selective medium lacking histidine and expressed *lacZ* which turned the color of

the colonies to blue in the β-galactosidase lift assay. The results matched those using

pAS2-aracca2 as bait (Table 6B). However, when pAS2-aracca3 was used as a bait, all

of the 10 cDNA clones failed to grow under histidine selection and tested negative for β-

galactosidase expression (Table 6B). As the negative control, all of the cDNA clones

were transformed individually into yeast HF7C alone or co-transformed with plasmid

pAS2 without insert. None of these transformants could express the reporter genes.

VI. Analysis of putative positives

A) Restriction digestion

Restriction digestion analysis was used to establish whether or not the same cDNA

clone had been independently isolated more than once. Because the cDNAs were

inserted into the *Xho*I site of plasmid pACT in the cDNA library, library plasmids were

digested with restriction enzyme *Xho*I and separated by agarose gel electrophoresis. As a

result, the 11 cDNA clones were classified into 4 groups based on the insert sizes (Table

7). Clones 19A, 23A, 68A and 75C which interacted strongly with the bait hybrid have

an insert size of 650 bp. Clones 14A and 49A which also interacted strongly with the bait

hybrid have an insert size of 1900 bp. Clones 22A, 27A, 39A and 54A which interacted

with the bait hybrid less strongly have an insert size of 700 bp. Clone 108 is 850 bp in

size. In addition, the plasmids were also analysed by *Hind*III digestion and the same

grouping results were obtained (data not shown).

B) Sequencing of cDNA clones

Preliminary sequencing of the cDNA clones from their 5' ends was performed using

the primer LibrA (Table 3). Analysis of sequencing results identified the cDNAs as

belonging to 4 different classes, confirming the groupings in Table 7. Both clones 14A

and 49A in Group D were sequenced from their 5' ends, and their nucleotide sequences

are identical over the region examined. Sequencing of clones 19A, 23A, 68A and 75C in

Group C from their 5' ends revealed that their sequences were identical over the region

examined. Two of four members in Group B, clones 22A and 54A, were also sequenced

from their 5' ends and shown to have the same sequence. Clone 108 was also sequenced

Table 7: Classification of cDNA clones by restriction digestion

Group	Clones	Insert size[a] (bp)
A	108	850
B	22A, 27A, 39A, 54A	700
C	19A, 23A, 68A, 75C	650
D	14A, 49A	1900

a: Approximate insert size was determined from agarose gel electrophoresis after restriction digestion with *Xho*I.

from its 5' end and shown to be different from the other 3 sets of clones. Representatives from each group, 14A, 19A, 22A and 108, were sequenced in their entirety. Their nucleotide sequences and predicted amino acid sequences are reported in Figures 14, 15, 16 and 17, respectively. Analysis of the predicted amino acid sequences revealed a major open reading frame for each cDNA clone (bold in Figures 14, 15, 16 and 17) which is in the same frame with the GAL4 open reading frame. The predicted amino acid sequences of each of these cDNA clones were used to search the GenBank and *Arabidopsis* databases with the BLASTX program (Gish and States,1993) which compares the query sequence to existing protein sequences in the databases. The results are shown in Table 8.

```
CGTTTGTGAAATGGAGAGTTCTGATGATCAGTCTCCGTCCAGGGCGAGTTTTATTTCGTA   60
  R  L  -  N  G  E  F  -  -  S  V  S  V  Q  G  E  F  Y  F  V
     V  C  E  M  E  S  S  D  D  Q  S  P  S  R  A  S  F  I  S  Y
  F  V  K  W  R  V  L  M  I  S  L  R  P  G  R  V  L  F  R
TGGAGCTGAATCTAGGATTGATTTGGAGTTGAGGCATTTGGTTGGAGATGTAGGAGAAGA  120
  W  S  -  I  -  D  -  F  G  V  E  A  F  G  W  R  C  R  R  R
     G  A  E  S  R  I  D  L  E  L  R  H  L  V  G  D  V  G  E  E
  M  E  L  N  L  G  L  I  W  S  -  G  I  W  L  E  M  -  E  K
AGGAGAGAGCAAGAAAGAGATTATCTTGGAGAAGAAGGAAGAAAGTAATGGAGAAGGTAG  180
  R  R  E  Q  E  R  D  Y  L  G  E  E  G  R  K  -  W  R  R  -
     G  E  S  K  K  E  I  I  L  E  K  K  E  E  S  N  G  E  G  S
  K  E  R  A  R  K  R  L  S  W  R  R  R  R  K  K  V  M  E  K  V
CTTGAGTCAGAAGAAACCTCTTTCGAATGGTAAGAAGGTTGCGAAAACGAGTCCTAATAA  240
  L  E  S  E  E  T  S  F  E  W  -  E  G  C  E  N  E  S  -  -
     L  S  Q  K  K  P  L  S  N  G  K  K  V  A  K  T  S  P  N  N
  A  -  V  R  R  N  L  F  R  M  V  R  R  L  R  K  R  V  L  I
TCCGAAGATGCCGGGTTCGAGGATCTCTTCGAGGAAATCACCGGATTTGGGGAAGGTTTC  300
  S  E  D  A  G  F  E  D  L  F  E  E  I  T  G  F  G  E  G  F
     P  K  M  P  G  S  R  I  S  S  R  K  S  P  D  L  G  K  V  S
  I  R  R  C  R  V  R  G  S  L  R  G  N  H  R  I  W  G  R  F
TGTTGATGAGGAGAGTCCTGAGCTTGGTGTGGTGTTGTTGAAGCAAGCCAGAGAGTTGGT  360
  C  -  -  G  E  S  -  A  W  C  G  V  V  E  A  S  Q  R  V  G
     V  D  E  E  S  P  E  L  G  V  V  L  L  K  Q  A  R  E  L  V
  L  L  M  R  R  V  L  S  L  V  W  C  C  -  S  K  P  E  S  W
TTCTTCAGGTGAGAATTTAAACAAGGCTCTTGATTTAGCGTTGAGAGCGGTGAAAGTGTT  420
  F  F  R  -  E  F  K  Q  G  S  -  F  S  V  E  S  G  E  S  V
     S  S  G  E  N  L  N  K  A  L  D  L  A  L  R  A  V  K  V  F
  F  L  Q  V  R  I  -  T  R  L  L  I  -  R  -  E  R  -  K  C
TGAGAAATGTGGGGAAGGAGAGAAACAACTTGGGTTGAATTTGGTGATGTCTTTGCATAT  480
  -  E  M  W  G  R  R  E  T  T  W  V  E  F  G  D  V  F  A  Y
     E  K  C  G  E  G  E  K  Q  L  G  L  N  L  V  M  S  L  H  I
  L  R  N  V  G  K  E  R  N  N  L  G  -  I  W  -  C  L  C  I
CTTGGCAGCGATATATGCAGGGTTAGGGAGATACAATGACGCTGTGCCTGTCCTTGAACG  540
  L  G  S  D  I  C  R  V  R  E  I  Q  -  R  C  A  C  P  -  T
     L  A  A  I  Y  A  G  L  G  R  Y  N  D  A  V  P  V  L  E  R
  S  W  Q  R  Y  M  Q  G  -  G  D  T  M  T  L  C  L  S  L  N
TTCTATCGAGATACCGATGATTGAAGATGGTGAGGATCACGCGTTGGCGAAATTTGCAGG  600
  F  Y  R  D  T  D  D  -  R  W  -  G  S  R  V  G  E  I  C  R
     S  I  E  I  P  M  I  E  D  G  E  D  H  A  L  A  K  F  A  G
  V  L  S  R  Y  R  -  L  K  M  V  R  I  T  R  W  R  N  L  Q
GTGTATGCAGCTTGGTGACATGTATGGTTTGATGGGTCAAGTTGAAAATTCGATTATGCT  660
  V  Y  A  A  W  -  H  V  W  F  D  G  S  S  -  K  F  D  Y  A
     C  M  Q  L  G  D  M  Y  G  L  M  G  Q  V  E  N  S  I  M  L
  G  V  C  S  L  V  T  C  M  V  -  W  V  K  L  K  I  R  L  C
ATACACCGCGGGTTTGGAGATTCAAAGACAAGTTCTTGGAGAATCGGATGCACGAGTTGG  720
  I  H  R  G  F  G  D  S  K  T  S  S  W  R  I  G  C  T  S  W
     Y  T  A  G  L  E  I  Q  R  Q  V  L  G  E  S  D  A  R  V  G
  Y  T  P  R  V  W  R  F  K  D  K  F  L  E  N  R  M  H  E  L
CGAAACTTGTAGGTACTTAGCGGAAGCTCATGTTCAAGCAATGCAGTTTGAAGAAGCCTC  780
  R  N  L  -  V  L  S  G  S  S  C  S  S  N  A  V  -  R  S  L
     E  T  C  R  Y  L  A  E  A  H  V  Q  A  M  Q  F  E  E  A  S
  A  K  L  V  G  T  -  R  K  L  M  F  K  Q  C  S  L  K  K  P
```

63

```
AAGGCTTTGTCAAATGGCATTGGACATTCACAAAGAAAACGGTGCTGCCGCTACAGCTTC    840
  K  A  L  S  N  G  I  G  H  S  Q  R  K  R  C  C  R  Y  S  F
 R  L  C  Q  M  A  L  D  I  H  K  E  N  G  A  A  A  T  A  S
Q  G  F  V  K  W  H  W  T  F  T  K  K  T  V  L  P  L  Q  L
TATTGAAGAAGCTGCGGATAGAAAACTGATGGGACTTATCTGCGATGCGAAAGGTGATTA    900
  Y  -  R  S  C  G  -  K  T  D  G  T  Y  L  R  C  E  R  -  L
 I  E  E  A  A  D  R  K  L  M  G  L  I  C  D  A  K  G  D  Y
L  L  K  K  L  R  I  E  N  -  W  D  L  S  A  M  R  K  V  I
TGAAGTTGCGCTTGAGCATTATGTTTTAGCGAGCATGGCAATGTCATCTCAGAACCATAG    960
  -  S  C  A  -  A  L  C  F  S  E  H  G  N  V  I  S  E  P  -
 E  V  A  L  E  H  Y  V  L  A  S  M  A  M  S  S  Q  N  H  R
M  K  L  R  L  S  I  M  F  -  R  A  W  Q  C  H  L  R  T  I
AGAAGATGTTGCTGCTGTAGATTGCAGTATTGGTGATGCTTACATGTCGCTAGCTAGGTT   1020
  R  R  C  C  C  C  R  L  Q  Y  W  -  C  L  H  V  A  S  -  V
 E  D  V  A  A  V  D  C  S  I  G  D  A  Y  M  S  L  A  R  F
E  K  M  L  L  L  -  I  A  V  L  V  M  L  T  C  R  -  L  G
TGATGAAGCGATATTCGCATACCAGAAGGCTTTAGCTGTGTTTAAGCAAGGAAAAGGTGA   1080
  -  -  S  D  I  R  I  P  E  G  F  S  C  V  -  A  R  K  R  -
 D  E  A  I  F  A  Y  Q  K  A  L  A  V  F  K  Q  G  K  G  E
L  M  K  R  Y  S  H  T  R  R  L  -  L  C  L  S  K  E  K  V
GACTCATTCCTCTGTTGCTTTGGTTTACGTTAGGCTTGCTGACTTGTACAACAAGATTGG   1140
  D  S  F  L  C  C  F  G  L  R  -  A  C  -  L  V  Q  Q  D  W
 T  H  S  S  V  A  L  V  Y  V  R  L  A  D  L  Y  N  K  I  G
R  L  I  P  L  L  L  W  F  T  L  G  L  L  T  C  T  T  R  L
GAAAACACGCGATTCCAAATCGTACTGCGAAAACGCTCTCAAGATTTACCTAAAACCGAC   1200
  E  N  T  R  F  Q  I  V  L  R  K  R  S  Q  D  L  P  K  T  D
 K  T  R  D  S  K  S  Y  C  E  N  A  L  K  I  Y  L  K  P  T
G  K  H  A  I  P  N  R  T  A  K  T  L  S  R  F  T  -  N  R
TCCAGGAACTCCTATGGAAGAGGTTGCCGACTGGTTTTATAGAGATTTCTGCGATATATCA   1260
  S  R  N  S  Y  G  R  G  C  D  W  F  Y  R  D  F  C  D  I  S
 P  G  T  P  M  E  E  V  A  T  G  F  I  E  I  S  A  I  Y  Q
L  Q  E  L  L  W  K  R  L  R  L  V  L  -  R  F  L  R  Y  I
GTCAATGAATGAACTTGATCAAGCTCTTAAGTTGTTGAGACGGGCGTTGAAGATATATGC   1320
  V  N  E  -  T  -  S  S  S  -  V  V  E  T  G  V  E  D  I  C
 S  M  N  E  L  D  Q  A  L  K  L  L  R  R  A  L  K  I  Y  A
S  Q  -  M  N  L  I  K  L  L  S  C  -  D  G  R  -  R  Y  M
AAACGCTCCAGGTCAACAGAACACGATTGCAGGTATTGAAGCTCAAATGGGTGTGGTTAC   1380
  K  R  S  R  S  T  E  H  D  C  R  Y  -  S  S  N  G  C  G  Y
 N  A  P  G  Q  Q  N  T  I  A  G  I  E  A  Q  M  G  V  V  T
Q  T  L  Q  V  N  R  T  R  L  Q  V  L  K  L  K  W  V  W  L
TTATATGATGGGAAATTACTCCGAGTCTACGACATATTCAAGAGCGCGATATCAAAGTTC   1440
  L  Y  D  G  K  L  L  R  V  Y  D  I  F  K  S  A  I  S  K  E
 Y  M  M  G  N  Y  S  E  S  T  T  Y  S  R  A  R  Y  Q  S  S
L  I  -  W  E  I  T  P  S  L  R  H  I  Q  E  R  D  I  K  V
CGCAATAGCGAGAGAAAAAACCGCTCTTTTCGGGATTGCTTTGAATCAGATGGGACTAGC   1500
  R  N  S  E  R  K  N  R  S  F  R  D  C  F  E  S  D  G  T  S
 A  I  A  R  E  K  T  A  L  F  G  I  A  L  N  Q  M  G  L  A
P  Q  -  R  E  K  K  P  L  F  S  G  L  L  -  I  R  W  D  -
TTGCGTTCAGCGTTACGCGATCAATGAAGCTGCAGATTTGTTTGAAGAAGCGAGAACTAT   1560
  L  R  S  A  L  R  D  Q  -  S  C  R  F  V  -  R  S  E  N  Y
 C  V  Q  R  Y  A  I  N  E  A  A  D  L  F  E  E  A  R  T  I
L  A  F  S  V  T  R  S  M  K  L  Q  I  C  L  K  K  R  E  L
TCTTGAGAAAGAGTGTGGACCGTACCATCCCGATACATTAGCGGTTTACAGTAACCTTGC   1620
  S  -  E  R  V  W  T  V  P  S  R  Y  I  S  G  L  Q  -  P  C
 L  E  K  E  C  G  P  Y  H  P  D  T  L  A  V  Y  S  N  L  A
F  L  R  K  S  V  D  R  T  I  P  I  H  -  R  F  T  V  T  L
```

```
TGGAACGTACGACGCAATGGGAAGGTTAGATGATGCTATAGAGATATTGGAGTATGTTGT   1680
 W  N  V  R  R  N  G  K  V  R  -  C  Y  R  D  I  G  V  C  C
  G  T  Y  D  A  M  G  R  L  D  D  A  I  E  I  L  E  Y  V  V
L  E  R  T  T  Q  W  E  G  -  M  M  L  -  R  Y  W  S  M  L
TGGGACAAGAGAAGAGAAGCTCGGGACGGCAAATCCCGAGGTTGAGGACGAGAAGCAGAG   1740
 W  D  K  R  R  E  A  R  D  G  K  S  R  G  -  G  R  E  A  E
  G  T  R  E  E  K  L  G  T  A  N  P  E  V  E  D  E  K  Q  R
L  G  Q  E  K  R  S  S  G  R  Q  I  P  R  L  R  T  R  S  R
GCTTGCTGCGTTGTTGAAAGAAGCTGGAAGAGGAAGGAGTAAAAGAAACAGAGCACTTCT   1800
 A  C  C  V  V  E  R  S  W  K  R  K  E  -  K  K  Q  S  T  S
  L  A  A  L  L  K  E  A  G  R  G  R  S  K  R  N  R  A  L  L
G  L  L  R  C  -  K  K  L  E  E  E  G  V  K  E  T  E  H  F
TACCCTTTTGGACAACAACCCTGAAATTGCAAATGGACAGAGACCGGTTTATTGATTAAC   1860
 Y  P  F  G  Q  Q  P  -  N  C  K  W  T  E  T  G  L  L  I  N
  T  L  L  D  N  N  P  E  I  A  N  G  Q  R  P  V  Y  -  L  T
L  P  F  W  T  T  T  L  K  L  Q  M  D  R  D  R  F  I  D  -
CGGATTATTGTCCGGTCTAAATTGTATATATATACAAAATACAAGTTCGAATACGGTACT   1920
 R  I  I  V  R  S  K  L  Y  I  Y  T  K  Y  K  F  E  Y  G  T
  G  L  L  S  G  L  N  C  I  Y  I  Q  N  T  S  S  N  T  V  L
P  D  Y  C  P  V  -  I  V  Y  I  Y  K  I  Q  V  R  I  R  Y
TACACGGTCGAGGGTAAAT
 Y  T  V  E  G  K
  T  R  S  R  V  N
L  H  G  R  G  -
```

Figure 14: Nucleotide sequence and predicted amino acid sequences in three reading frames of clone 14A. Standard one letter abbreviations are used for the amino acids. The predicted open reading frame is shown in bold. – indicates stop codons. The first nucleotide at the 5' end and the last nucleotide at the 3' end are linked to the adaptor sequences.

65

```
GAAGAAGCAGACGACGACGACGAAGATGAGTTCATCGACTGTTCGTGATTTGCAGAGAGA   60
  E  E  A  D  D  D  D  E  D  E  F  I  D  C  S  -  F  A  E  R
 K  K  Q  T  T  T  T  K  M  S  S  S  T  V  R  D  L  Q  R  D
   R  S  R  R  R  R  R  R  -  V  H  R  L  F  V  I  C  R  E
CCTAGAAAACAAGGCTAATGATCTCGGCAAAATCCAGAAAGATATCGGGAAGAATCATCA  120
  P  R  K  Q  G  -  -  S  R  Q  N  P  E  R  Y  R  E  E  S  S
 L  E  N  K  A  N  D  L  G  K  I  Q  K  D  I  G  K  N  H  Q
   T  -  K  T  R  L  M  I  S  A  K  S  R  K  I  S  G  R  I  I
GCTGAGAAAGAAGTATACGATCCAACTCGGTGAGAATGAGCTCGTGCTTAAGGAGTTAGA  180
  A  E  K  E  V  Y  D  P  T  R  -  E  -  A  R  A  -  G  V  R
 L  R  K  K  Y  T  I  Q  L  G  E  N  E  L  V  L  K  E  L  D
   S  -  E  R  S  I  R  S  N  S  V  R  M  S  S  C  L  R  S  -
TTTGCTGGAAGAAGATGCTAACGTCTACAAATTGATTGGTCCGGTGTTAGTGAAGCAGGA  240
  F  A  G  R  R  C  -  R  L  Q  I  D  W  S  G  V  S  E  A  G
 L  L  E  E  D  A  N  V  Y  K  L  I  G  P  V  L  V  K  Q  D
   I  C  W  K  K  M  L  T  S  T  N  -  L  V  R  C  -  -  S  R
TTTGGCTGAAGCTAATGCAAATGTTCGTAAACGAATTGAGTACATCTCTGCTGAATTGAA  300
  F  G  -  S  -  C  K  C  S  -  T  N  -  V  H  L  C  -  I  E
 L  A  E  A  N  A  N  V  R  K  R  I  E  Y  I  S  A  E  L  K
   I  W  L  K  L  M  Q  M  F  V  N  E  L  S  T  S  L  L  N  -
ACGGCTTGATGCAATTCTTCAAGATATGGAAGAAAAACAAAACAACAAGAGAGAAACGAT  360
  T  A  -  C  N  S  S  R  Y  G  R  K  T  K  Q  Q  E  R  N  D
 R  L  D  A  I  L  Q  D  M  E  E  K  Q  N  N  K  R  E  T  I
   N  G  L  M  Q  F  F  K  I  W  K  K  N  K  T  T  R  E  K  R
AATGAAGTTACAACAAAGGCTACAAACTATCCAGGCAGGAAAAGCAAAGGCTTGATTGAG  420
  N  E  V  T  T  K  A  T  N  Y  P  G  R  K  S  K  G  L  I  E
 M  K  L  Q  Q  R  L  Q  T  I  Q  A  G  K  A  K  A  -  L  S
   -  -  S  Y  N  K  G  Y  K  L  S  R  Q  E  K  Q  R  L  D  -
CTACTCAGTTGCAACCACACTTGTTAAACCCTTTCCACTGTGCTCGGCCTTGTCTCAATG  480
  L  L  S  C  N  H  T  C  -  T  L  S  T  V  L  G  L  V  S  M
 Y  S  V  A  T  T  L  V  K  P  F  P  L  C  S  A  L  S  Q  W
   A  T  Q  L  Q  P  H  L  L  N  P  F  H  C  A  R  P  C  L  N
GTGTCATATTCAGTTACTGAGGGGTTCTTGTGGTTTCAGTTTATTGGGTGATTGATGTCTC  540
  V  S  Y  S  V  T  E  G  S  C  G  F  S  L  L  G  D  -  C  L
 C  H  I  Q  L  L  R  V  L  V  V  S  V  Y  W  V  I  D  V  S
   G  V  I  F  S  Y  -  G  F  L  W  F  Q  F  I  G  -  L  M  S
TTTGTGTCACTAATCTTTTTTTTGTTGACAACTAGACTTAAGTGTGTAACTTATACATGAT  600
  F  V  S  L  I  F  F  C  -  Q  L  D  L  S  V  -  L  I  H  D
 L  C  H  -  S  F  F  V  D  N  -  T  -  V  C  N  L  Y  M  I
   L  C  V  T  N  L  F  L  L  T  T  R  L  K  C  V  T  Y  T  -
CTTGTCGCCCTTTATGGAGTTATGAGTTTGGAGGGTCAAT
  L  V  A  L  Y  G  V  M  S  L  E  G  Q
 L  S  P  F  M  E  L  -  V  W  R  V  N
   S  C  R  P  L  W  S  Y  E  F  G  G  S
```

Figure 15: Nucleotide sequence and predicted amino acid sequences in three reading frames of clone 19A. Standard one letter abbreviations are used for the amino acids. The predicted open reading frame is shown in bold. – indicates stop codons. The first nucleotide at the 5' end and the last nucleotide at the 3' end are linked to the adaptor sequences.

```
TGGAGCATTCAACTGTCAAGGAGGAGGATGGTGCAGAGAAACCAGACGTAACCAATGTTT    60
 W  S  I  Q  L  S  R  R  R  M  V  Q  R  N  Q  T  -  P  M  F
  G  A  F  N  C  Q  G  G  G  W  C  R  E  T  R  R  N  Q  C  F
   E  H  S  T  V  K  E  E  D  G  A  E  K  P  D  V  T  N  V
CTCTGAATGCGTCAACACGTTAACCGCCACCACAAGCCCTAAAGACGTTGAATGGAACAG   120
 L  -  M  R  Q  H  V  N  R  H  H  K  P  -  R  R  -  M  E  Q
  S  E  C  V  N  T  L  T  A  T  T  S  P  K  D  V  E  W  N  S
   S  L  N  A  S  T  R  -  P  P  P  Q  A  L  K  T  L  N  G  T
TGGAAGCAGCCCAATCTCCATTGCAAACGTTGAAGAGTTTGCTTTGTTCTTGTCTCAATC   180
 W  K  Q  P  N  L  H  C  K  R  -  R  V  C  F  V  L  V  S  I
  G  S  S  P  I  S  I  A  N  V  E  E  F  A  L  F  L  S  Q  S
   V  E  A  A  Q  S  P  L  Q  T  L  K  S  L  L  C  S  C  L  N
CAAGAAGCTTTTGTTGTCTGGACTAAACGATGATCTTGAGCTGACTTTGGAGCCTTTCAA   240
 Q  E  A  F  V  V  W  T  K  R  -  S  -  A  D  F  G  A  F  Q
  K  K  L  L  L  S  G  L  N  D  D  L  E  L  T  L  E  P  F  K
   P  R  S  F  C  C  L  D  -  T  M  I  L  S  -  L  W  S  L  S
GTTTGAGCTGATCACTGTCTCTCCTGTTGTGACCATTGAGGGTAATTCAGTCCGGTTTGC   300
 V  -  A  D  H  C  L  S  C  D  H  -  G  -  F  S  P  V  C
  F  E  L  I  T  V  S  P  V  V  T  I  E  G  N  S  V  R  F  A
   S  L  S  -  S  L  S  L  L  L  -  P  L  R  V  I  Q  S  G  L
TCCGATTGGACTGGTTAACATGCTAAACACAAGCGGTGCGATCCGGTCTTTGGTGTATAA   360
 S  D  W  T  G  -  H  A  K  H  K  R  C  D  P  V  F  G  V  -
  P  I  G  L  V  N  M  L  N  T  S  G  A  I  R  S  L  V  Y  N
   L  R  L  D  W  L  T  C  -  T  Q  A  V  R  S  G  L  W  C  I
TGATGAATCTGTTGAGGTCGGAGTTTTTGGTGCCGGAGAGTTTAGGGTTTATGCATCGAA   420
 -  -  I  C  -  G  R  S  F  W  C  R  R  V  -  G  L  C  I  E
  D  E  S  V  E  V  G  V  F  G  A  G  E  F  R  V  Y  A  S  K
   M  M  N  L  L  R  S  E  F  L  V  P  E  S  L  G  F  M  H  R
AAAACCTGTGAGCTGCTTAATTGATGGTGAAGTTGTTGAGTTTGGTTATGAAGACTCAAT   480
 K  T  C  E  L  L  N  -  W  -  S  C  -  V  W  L  -  R  L  N
  K  P  V  S  C  L  I  D  G  E  V  V  E  F  G  Y  E  D  S  M
   K  N  L  -  A  A  -  L  M  V  K  L  L  S  L  V  M  K  T  Q
GGTGATGGTGCAAGTCCCTTGGTCTGGTCCAGATGGTTTGTCTTCTATTCAGTATTTGTT   540
 G  D  G  A  S  P  L  V  W  S  R  W  F  V  F  Y  S  V  F  V
  V  M  V  Q  V  P  W  S  G  P  D  G  L  S  S  I  Q  Y  L  F
   W  -  W  C  K  S  L  G  L  V  Q  M  V  C  L  L  F  S  I  C
TTAGAGATTTTGAAGGTCCTTATTTGTTTCCTTCTTAACTCCGTACCGTTCGTTGAGTTT   600
 L  E  I  L  K  V  L  I  C  F  L  L  N  S  V  P  F  V  E  F
  -  R  F  -  R  S  L  F  V  S  F  L  T  P  Y  R  S  L  S  F
   F  R  D  F  E  G  P  Y  L  F  P  S  -  L  R  T  V  R  -  V
TTGAAAGAACTCCGAGCCGTTTTTTATATGGTTTCTGAGAGTAAACAATTTATGTTGGTT   660
 L  K  E  L  R  A  V  F  Y  M  V  S  E  S  K  Q  F  M  L  V
  -  K  N  S  E  P  F  F  I  W  F  L  R  V  N  N  L  C  W  F
   F  E  R  T  P  S  R  F  L  Y  G  F  -  E  -  T  I  Y  V  G
TTTTTTTGTCAAACATTTATGTTGGCTTTTTTAAAAAAAAAAAGGTAAAT
 F  F  C  Q  T  F  M  L  A  F  L  K  K  K  G  K
  F  F  V  K  H  L  C  W  L  F  -  K  K  K  V  N
   F  F  L  S  N  I  Y  V  G  F  F  K  K  K  R  -
```

Figure 16: Nucleotide sequence and predicted amino acid sequences in three reading frames of clone 22A. Standard one letter abbreviations are used for the amino acids. The predicted open reading frame is shown in bold. – indicates stop codons. The first nucleotide at the 5' end and the last nucleotide at the 3' end are linked to the adaptor sequences.

```
GTCGTCAAAGATGGTGAAGTCTTCAGAAATCGTCGGAAAACTTAATCTCAGGGCACACCA   60
 V  V  K  D  G  E  V  F  R  N  R  R  K  T  -  S  Q  G  T  P
S  S  K  M  V  K  S  S  E  I  V  G  K  L  N  L  R  A  H  Q
  R  Q  R  W  -  S  L  Q  K  S  S  E  N  L  I  S  G  H  T
AGAAGGAGGCTTTTTCGCTGAAACTTTCAGAGATTCTTCTGTTTTCCTCTCCACCTCTCA  120
 R  R  R  L  F  R  -  N  F  Q  R  F  F  C  F  P  L  H  L  S
E  G  G  F  F  A  E  T  F  R  D  S  S  V  F  L  S  T  S  Q
  K  K  E  A  F  S  L  K  L  S  E  I  L  L  F  S  S  P  P  L
GCTCCCTCCAACCTTTAAGGTGGACATAGCTGTGAGCACAAGTATCTACTTCTTGCTTCC  180
 A  P  S  N  L  -  G  G  H  S  C  E  H  K  Y  L  L  L  A  S
L  P  P  T  F  K  V  D  I  A  V  S  T  S  I  Y  F  L  L  P
  S  S  L  Q  P  L  R  W  T  -  L  -  A  Q  V  S  T  S  C  F
ATCTGGGAGCGTTTCTCGTCTTCATCGCATACCAATGGCTGAAACCTGGCACTTTTATTT  240
 I  W  E  R  F  S  S  S  H  T  N  G  -  N  L  A  L  L  F
S  G  S  V  S  R  L  H  R  I  P  M  A  E  T  W  H  F  Y  L
  H  L  G  A  F  L  V  F  I  A  Y  Q  W  L  K  P  G  T  F  I
GGGTGAACCCCTTACTGTAGTGGAACTCTACGATGATGGGAAGTTGAAATTCACATGTCT  300
 G  -  T  P  Y  C  S  G  T  L  R  -  W  E  V  E  I  H  M  S
G  E  P  L  T  V  V  E  L  Y  D  D  G  K  L  K  F  T  C  L
  W  V  N  P  L  L  -  W  N  S  T  M  M  G  S  -  N  S  H  V
TGGTCCTGACTTGTTTGAAGGTGATCAGAAGCCTCAGTACACGGTCCCTCCAAACGTCTG  360
 W  S  -  L  V  -  R  -  S  E  A  S  V  H  G  P  S  K  R  L
G  P  D  L  F  E  G  D  Q  K  P  Q  Y  T  V  P  P  N  V  W
  L  V  L  T  C  L  K  V  I  R  S  L  S  T  R  S  L  Q  T  S
GTTTGGTTCGTTTCCTACAAAGGATGTTCATTTTTCTCAAGACGGGGCTCTGCTTAAAGC  420
 V  W  F  V  S  Y  K  G  C  S  F  F  S  R  R  G  S  A  -  S
F  G  S  F  P  T  K  D  V  H  F  S  Q  D  G  A  L  L  K  A
  G  L  V  R  F  L  Q  R  M  F  I  F  L  K  T  G  L  C  L  K
TGAGGCTAGAGACTCTGAGAACCATTTCTCTCTTGTTGGCTGCACCTGCGCTCCTGCTTT  480
 -  G  -  R  L  -  E  P  F  L  S  C  W  L  H  L  R  S  C  F
E  A  R  D  S  E  N  H  F  S  L  V  G  C  T  C  A  P  A  F
  L  R  L  E  T  L  R  T  I  S  L  L  L  A  A  P  A  L  L  L
CCAGTTCGAGGACTTTGAGCTCGCGAAACGCTCTGATCTCTTATCAAGGTTTCCGCAACA  540
 P  V  R  G  L  -  A  R  E  T  L  -  S  L  I  K  V  S  A  T
Q  F  E  D  F  E  L  A  K  R  S  D  L  L  S  R  F  P  Q  H
  S  S  S  R  T  L  S  S  R  N  A  L  I  S  Y  Q  G  F  R  N
CGAGTCACTCATCACAATGCTATCTTACCCGGAGTGAAAGGGCCTACCTAGAGGAGACGC  600
 R  V  T  H  H  N  A  I  L  P  G  V  K  G  P  T  -  R  R  R
E  S  L  I  T  M  L  S  Y  P  E  -  K  G  L  P  R  G  D  A
  T  S  H  S  S  Q  C  Y  L  T  R  S  E  R  A  Y  L  E  E  T
CTTATTGGCTCAATGGTTTTTGGTAACTACTAAGAGAAGGATACTATGTTTGCTTTGTTA  660
 L  I  G  S  M  V  F  G  N  Y  -  E  K  D  T  M  F  A  L  L
L  L  A  Q  W  F  L  V  T  T  K  R  R  I  L  C  L  L  C  Y
  P  Y  W  L  N  G  F  W  -  L  L  R  E  G  Y  Y  V  C  F  V
CTTCAACTTTTCATATGAATCATACAAAGGTTGTAACTGTACTTTGAGTTCACTTTTCAT  720
 L  Q  L  F  I  -  I  I  Q  R  L  -  L  Y  F  E  F  T  F  H
F  N  F  S  Y  E  S  Y  K  G  C  N  C  T  L  S  S  L  F  I
  T  S  T  F  H  M  N  H  T  K  V  V  T  V  L  -  V  H  F  S
CTTTGTATCTTTGTTTTTACTGAATTGTGATTGATCAAATAAATGGCCAGTAAAGAACAG  780
 L  C  I  F  V  F  T  E  L  -  L  I  K  -  M  A  S  K  E  Q
F  V  S  L  F  L  L  N  C  D  -  S  N  K  W  P  V  K  N  R
  S  L  Y  L  C  F  Y  -  I  V  I  D  Q  I  N  G  Q  -  R  T
```

```
GCTCAACAATTTCATTCCATAAACAGAGCAGAGGATCTGCAACGACTAAGTTCTGAGGAA    840
 A  Q  Q  F  H  S  I  N  R  A  E  D  L  Q  R  L  S  S  E  E
   L  N  N  F  I  P  -  T  E  Q  R  I  C  N  D  -  V  L  R  K
 G  S  T  I  S  F  H  K  Q  S  R  G  S  A  T  T  K  F  -  G
AAAAAAAAAAAAAAGGG
  K  K  K  K  K
   K  K  K  K  R
 K  K  K  K  K  G
```

Figure 17: Nucleotide sequence and predicted amino acid sequences in three reading frames of clone 108. Standard one letter abbreviations are used for the amino acids. The predicted open reading frame is shown in bold. − indicates stop codons. The first nucleotide at the 5' end and the last nucleotide at the 3' end are linked to the adaptor sequences.

Table 8: Results of searches in the GenBank and *Arabidopsis* databases with BLASTX

cDNA clone	amino acids	Identities	Organism	Predicted protein	Accession No.
14A	615 aa	54%/606 aa	*Arabidopsis*	Kinesin light chain	AC004557
19A	129 aa	39%/121 aa	Human	Protein KE2	O15212
22A	180 aa	58%/178 aa	Cucumber	Raffinose synthase	AF073744
108	188 aa	37%/178 aa	*Synechocystis* sp.	Hypothetical protein	D90901

DISCUSSION

The enzyme, ATP (CTP):tRNA nucleotidyltransferase, catalyzes the addition of CMP and AMP residues to tRNAs that have an incomplete CCA sequence. In *Saccharomyces cerevisiae*, tRNA nucleotidyltransferase, produced from a single nuclear gene has been shown to function in multiple subcellular locations (Chen *et al.*, 1992). Preliminary data from our group suggests that this may also be the case in other organisms including plants. However, how this enzyme is targeted to different subcellular locations is still poorly understood. This work was initiated to identify proteins from plants which interact with tRNA nucleotidyltransferase. Protein-protein interactions form the basis of a wide variety of biochemical reactions and play an important role in protein transport. The identification of proteins which interact with a known protein is essential to elucidate the regulation and function of that protein. For these reasons, I employed the yeast two-hybrid system to screen an *Arabidopsis* cDNA library. to identify proteins that interact with lupin or *Arabidopsis* tRNA nucleotidyltransferase.

I. Co-targeting of proteins

Proteins which are shared between multiple subcellular locations are often the result of different transcription initiation sites being used in a gene which contains multiple in-frame start codons (Small *et al.*, 1998). Numerous genes have been identified which contain two or more in-frame ATGs at their 5' ends, which might correspond to the initiation codons used to initiate translation of functional proteins. If the sequence

71

between the two ATGs encodes an organellar targeting sequence, then the protein synthesized from the upstream ATG will contain this targeting information while protein initiated from the downstream ATG will not. The first actual example of co-targeting of a plant protein was shown for pea glutathione reductase (pGR) which is found in the cytosol, mitochondria and chloroplast, but is encoded by a single nuclear gene (Creissen *et al.*, 1995). Characterization of a full-length pGR cDNA revealed two potential in-frame start codons near the amino terminus of the predicted protein. The region between these two start codons was shown to code for a sequence required for organellar targeting, as expression of the full-length cDNA which encodes the protein with the amino-terminal targeting sequence resulted in substantially elevated pGR activities in both chloroplasts and motochondria in transgenic tobacco. Moreover, expression of a fusion protein of the amino-terminal region of pGR and phosphinothricin acetyl transferase also resulted in targeting of the foreign protein to both chloroplasts and mitochondria. Protein produced from the downstream start codon without the amino-terminal targeting sequence may contribute to cytosolic enzyme activity (Creissen *et al.*, 1995). More recently, it was found that genes encoding *Arabidopsis* alanyl-tRNA synthetase (Mireau *et al.*, 1996), histidyl-tRNA synthetase (Akashi *et al.*, 1998) and methionyl-tRNA synthetase (Menand *et al.*, 1998) all contain two in-frame start codons. In each case, the amino acids between the two start codons encode an organellar targeting signal such that the large form of the enzyme is directed to the organelle while the short form functions in the cytosol. These examples suggest that some plant proteins which are shared between multiple subcellular compartments are encoded by a single gene.

II. Lupin and *Arabidopsis* tRNA nucleotidyltransferases

My research focused on tRNA nucleotidyltransferase. In yeast, this enzyme is encoded by the *CCA1* gene which contains three in-frame start codons (Aebi *et al.*, 1990). Proteins produced from transcripts which initiate upstream of the first start codon are targeted to mitochondria while proteins produced from transcripts which initiate after the first start codon provide the nucleocytosolic enzyme activities (Martin and Hopper, 1994). The additional amino acids resulting from translation of the longer transcripts function as a mitochondrial targeting signal. Before this work, lupin tRNA nucleotidyltransferase had already been purified and characterized in our lab (Shanmugam *et al.*, 1996). Sequence analysis of a lupin tRNA nucleotidyltransferase cDNA revealed two in-frame start codons which could code for amino acids 1 and 58, respectively (Figure 1) (Shanmugam *et al.*, 1996). The first 57 amino acids have characteristics of both mitochondrial and chloroplast targeting signals and are not essential for enzyme activity (Shanmugam *et al.*, 1996). These observations suggest that the first 57 amino acids of the lupin enzyme might potentially encode an organellar targeting signal. In addition, the lupin tRNA nucleotidyltransferase contains a classical nuclear localization signal consensus sequence (Figure 1) (Shanmugam *et al.*, 1996). Therefore, it is possible that in lupin as in yeast, a single nuclear gene encodes multiple forms of tRNA nucleotidyltransferase which might function in different subcellular locations.

A cDNA (122M21T7) encoding *Arabidopsis* tRNA nucleotidyltransferase which was identified from the *Arabidopsis* EST library was sequenced completely in this work

(Figure 9). Assuming that translation begins at the first AUG, translation of this cDNA showed an open reading frame of 1626 bp starting from an ATG at position 139 in the cDNA, extending to a TGA at position 1765 in the cDNA, and potentially encoding a protein of 542 amino acids (Figure 9). The size of the protein predicted is in good agreement with the yeast and lupin enzymes which are predicted to contain 546 amino acids (Aebi *et al.*, 1990) and 560 amino acids (Shanmugam *et al.*, 1996), respectively. Sequence analysis of this cDNA showed no stop codons upstream of the first in-frame ATG, indicating that this cDNA may not represent the full-length transcript from the *Arabidopsis* gene. When this cDNA sequence was compared with the *Arabidopsis* gene sequence generated from the *Arabidopsis* genome sequencing project, it was found that the open reading frame of the cDNA could be extended further upstream so that it contains two additional in-frame ATGs (ATG1 and ATG6) (Figure 10). The first in-frame start codon of the cDNA corresponds to ATG69 in the extended open reading frame. Moreover, stop codons in all three frames were found upstream of ATG1 in the gene sequence (Figure 10). In addition, sequences between the most proximal stop codon upstream of the first ATG and the beginning of the cloned cDNA sequence lack a typical intron-exon junction sequence (a T-rich region followed by AG) (Snyder and Storno, 1995). Taken together, these data suggest the possibility that a full-length cDNA might start from ATG1 and contain three start codons at amino acids 1, 6 and 69. The first start codon in the gene sequence is preceded by a TATATA sequence (Figure 10), which is proposed to be the core element of plant promoters (Mukumoto *et al.*, 1993). Since the TATA element is usually located 30-40 nucleotides upstream from the transcription initiation site (Mukumoto *et al.*, 1993), transcription probably begins upstream of ATG1.

Many efficient plant promoters have this TATA element including genes for the chlorophyll a/b-binding protein in petunia and genes for alcohol dehydrogenase in maize and *Arabidopsis* (Mukumoto *et al.*, 1993). If proteins are produced from ATG1 or ATG6, they will contain an amino-terminal sequence, lacking from proteins produced from ATG69. Like the predicted amino-teminal region of the lupin and yeast proteins, the amino-terminal portion of the predicted *Arabidopsis* protein contains many basic and hydroxylated amino acids characteristic of both mitochondrial and chloroplast targeting signals. In this work, *Arabidopsis* cDNAs aracca2 (starting from ATG6) and aracca3 (starting from ATG69) were PCR amplified (Figure 12) and cloned into two-hybrid bait plasmid pAS2, separately. When the resulting plasmids, pAS2-aracca2 and pAS2-aracca3, were introduced into yeast strain NT33-5 which carries a temperature-sensitive mutation in the yeast *CCA1* gene, they both complemented this mutation (Table 5), indicating that the amino acids between ATG6 and ATG69 are not essential for enzyme activity. As in *S. cerevisiae* (Chen *et al.*, 1992), these additional amino-terminal amino acids encode a potential organellar targeting signal. The amino acid sequence predicted from the *Arabidopsis* cDNA showed high levels of similarity with the sequence predicted for the lupin tRNA nucleotidyltransferase (Figure 2). However, the similarity was low for the amino-terminal regions. This may not be surprising as amino-terminal targeting signals generally lack a conserved primary sequence (Von Heijne *et al.*, 1989). The spacing between ATG6 and ATG69 (62 amino acids) in the *Arabidopsis* tRNA nucleotidyltransferase is very similar to that between ATG1 and ATG58 (56 amino acids) in the lupin tRNA nucleotidyltransferase. In addition, the predicted lupin protein contains a potential nuclear localization signal (Shanmugam *et al.*, 1996), two stretches

75

of basic amino acids separated by a spacer of about 10 amino acids (Figure 2). The predicted *Arabidopsis* protein also contains some basic amino acids in this region (Figure 2). Therefore, plasmids pAS2-lupinccal and pAS2-aracca2 were constructed in this work and used as baits to screen an *Arabidopsis* cDNA library in the yeast two-hybrid system with a major objective of identifying proteins that may interact with these potential targeting signals.

III. The yeast two-hybrid system

The yeast two-hybrid system was devised to identify *in vivo* genes encoding proteins that physically associate with a protein of interest. In contrast to biochemical methods detecting protein-protein interaction, this system is based on a yeast genetic assay in which the interaction of two proteins is measured by the reconstitution of a functional transcription activator in yeast. A major advantage of using the yeast two-hybrid system to screen cDNA libraries is that the cDNAs encoding the interacting proteins are immediately available for further studies. Furthermore, the two-hybrid system is highly sensitive to weak or transient interactions that cannot be detected by other methods (Estojak *et al.*, 1995). However, when using this method, the proteins tested must be able to retain their functions and fold into their native conformations in the yeast cell. As this system requires expression of the proteins as GAL4 fusion hybrids, this may interfere with the interactions.

In this work, I employed the yeast two-hybrid system to screen the *Arabidopsis* cDNA library with the bait plasmids pAS2-lupinccal and pAS2-aracca2. As mentioned previously, proteins produced from plasmids pAS2-lupinccal and pAS2-aracca2 could

complement temperature-sensitive mutations in the yeast *CCA1* gene, indicating that the proteins produced are functional when expressed as GAL4 fusions. The whole cDNA library contains 36×10^6 independent recombinants. More than 22×10^6 transformants were screened with each bait plasmid. As a result, ten positive cDNA clones were isolated when the *Arabidopsis* tRNA nucleotidyltransferase was used as bait and one additional cDNA clone was isolated with the lupin tRNA nucleotidyltrasferase bait hybrid.

When screening a cDNA library with the yeast two-hybrid system, false positives may arise from activation of a reporter gene due to spurious interactions between the transcription activation hybrid and the promoter region of the reporter gene (Bartel *et al.*, 1993). A number of proteins are commonly encountered as false positives in the yeast two-hybrid system. These include heat shock proteins, ribosomal proteins, cytochrome oxidase, cytoskeletal proteins, proteasome subunits and tRNA synthases (Serebriiskii *et al.*, 2000). To eliminate false positives, all 11 cDNA clones were transformed individually into HF7C alone (data not shown) or with bait plasmid pAS2 without insert. None of these was capable of activating transcription of the reporter genes (Table 6) through direct interaction with the promoter region of the reporter genes or with the GAL4 DNA-binding domain, suggesting that they were not false positives. Therefore, further analysis of these cDNA clones was carried out.

Restriction digestion and preliminary DNA sequencing from their 5' ends demonstrated that the 11 positive clones were classified into 4 groups representing 4 different cDNAs. Clones 14A and 49A which interacted strongly with the *Arabidopsis* tRNA nucleotidyltransferase contain identical cDNAs and have an insert size of 1941 bp

(Figure 14). Clones 19A, 23A, 68A and 75C which also interacted strongly with the

Arabidopsis tRNA nucleotidyltransferase contain identical cDNAs and have an insert size

of 640 bp (Figure 15). Clones 22A, 27A, 39A and 54A which interacted with the

Arabidopsis tRNA nucleotidyltransferase less strongly contain identical cDNAs and have

an insert size of 709 bp (Figure 16). Finally, clone 108 which interacted strongly with the

lupin tRNA nucleotidyltransferase was grouped separately and has an insert size of 857

bp (Figure 17). The strength of interaction described above is based on the color of

colonies in the β-galactosidase lift assay. Estojak *et al.* (1995) found that protein-protein

interactions detected in the yeast two-hybrid system can be discriminated between high-,

intermediate-, or low-affinity based on the relative ability to activate transcription of a

reporter gene. When the interaction was of high affinity, the reporter gene was strongly

transcribed, resulting in dark blue colonies in the β-galactosidase lift assay, whereas low-

affinity interaction only resulted in slight blue colonies in the β-galactosidase lift assay.

The 10 positive cDNA clones isolated by interactions with the bait plasmid pAS2-

aracca2 (containing ATG6) also interacted with the bait plasmid pAS2-aracca1

(containing ATG1) but did not interact with the bait plasmid pAS2-aracca3 (containing

only ATG69) in the yeast two-hybrid system (Table 6), clone 108 interacted with the bait

plasmid pAS2-lupincca1 (containing ATG1) but did not interact with the bait plasmid

pAS2-lupincca2 (containing ATG58) in the yeast two-hybrid system (Table 6). As

proteins produced from bait plasmids pAS2-lupincca1, pAS2-aracca1 and pAS2-aracca2

contained the amino-terminal extensions, and proteins produced from bait plasmids

pAS2-lupincca2 and pAS2-aracca3 did not, these results suggest that the amino-terminal

extensions of the lupin and *Arabidopsis* tRNA nucleotidyltransferases are required for the

interactions. However, when the cDNA coding for the amino-terminal 344 amino acids of the lupin protein was used as bait no interaction was evident (data not shown), suggesting that more information than just that encoded in the region between the start codons is required for the interactions. Moreover, it is interesting that the proteins identified as interacting with the *Arabidopsis* tRNA nucleotidyltransferase did not interact with the lupin tRNA nucleotidyltransferase or *vice versa*. This would argue against these interactions resulting from non-specific interactions with the conserved region of the lupin and *Arabidopsis* enzymes which share greater than 70% sequence similarity (Figure 2) and would be expected to have very similar higher order structures. In contrast. the amino-terminal regions of the *Arabidopsis* and lupin tRNA nucleotidyltransferases share very low primary sequence similarity (Figure 2). Since these regions were required for their respective interactions (Table 6) and are not essential for enzyme activities (Table 5), perhaps they are primarily responsible for the different reactions detected in this work.

IV. Characterization of the cDNA clones

Clones 14A. 19A. 22A and 108 representing the four different classes of cDNAs isolated from the cDNA library were sequenced completely. The nucleotide sequences and the predicted amino acid sequences are reported in Figures 14, 15, 16 and 17, respectively. Analysis of the predicted amino acid sequences revealed a major open reading frame for each cDNA clone in the same open reading frame as the sequence coding for the GAL4 activation domain.

First I will discuss the proteins which showed the stronger interaction in the two-hybrid assay and which appear to be the most likely candidates for biologically significant interactions with the tRNA nucleotidyltransferases. Clone 14A could encode a 615-amino acid polypeptide (p14A) from the first in-frame start codon at nucleotide 11 to the first in-frame stop codon at nucleotide 1857 (Figure 14). After searches of the GenBank and *Arabidopsis* databases with BLASTX (Gish and States, 1993), it was found that the predicted protein shows 54% identity over its complete length to the predicted *Arabidopsis* kinesin light chain which contains 707 amino acids (GenBank Accession No. AC004557) (Figure 18). While it is evident that this cDNA does not encode the kinesin light chain, it may encode a related protein. Indeed, kinesins represent a large family of proteins in yeast (Saunders *et al.*, 1995) and mammals (Sack *et al.*, 1999), and this result suggests there will also be multiple kinesin genes in plants.

Kinesins are motor proteins that utilize ATP hydrolysis to drive the transport of vesicles or organelles towards the plus end of microtubules. They are found in a variety of cells and tissues and function during both interphase and mitosis (Bloom and Endow, 1995). Native kinesin is a heterotetramer composed of two heavy chains (KHC) and two light chains (KLC) (Bloom *et al.*, 1988). The molecular masses vary in different species from 110 to 130 kDa for the KHCs and from 50 to 70 kDa for KLCs (Bloom *et al.*, 1988). KLC structure is highly conserved among species. The amino-terminal region of KLC is predicted to form an alpha-helical coiled coil that is necessary and sufficient for interaction with KHC (Gauger and Goldstein, 1993). The carboxy-terminal region of KLC is largely made up of six tetratricopeptide repeats (TPR) (Gindhart and Goldstein, 1996) that are protein-protein interaction motifs identified in a diverse group of proteins.

```
Query:    1    VCEMESSDDQSPSRASFISYGAESRIDLELRHLV------------GDVGEEGESKKEI
               +CE++SS+ QSPSR SF SYG ES+ID +L+HL            GD  E
Sbjct:   85    LCELQSSN-QSPSRQSFGSYGDESKIDSDLQHLALGEMRDIDILEDEGDEDEVAKPEEFD

Query:   48    ILEKKEESNGEGSLSQKKPLSNGKKVAKTSPNNPKMPGSRISSRKSPDLGKVSVDEESPE
               + E       +  +  K V K++    M   ++    K  + G        E+ E
Sbjct:  144    VKSNSSNLDLEVMPRDMEKQTGKKNVTKSNVGVGGMRKKKVGGTKLQN-GNEEPSSENVE

Query:  108    LGVVLLKQARELVSSGENLNKALDLALRAVKVFEKCGEGEKQLGLNLVMSLHILAAIYAG
               L   LL QAR LVSSG++ +KAL+L  RA K+FE   E  K   L  +M LH+ AA++
Sbjct:  203    LARFLLNQARNLVSSGDSTHKALELTHRAAKLFEASAENGKPC-LEWIMCLHVTAAVHCK

Query:  168    LGRYNDAVPVLERSIEIPMIEDGEDHALAKFAGCMQLGDMYGLMGQVENSIMLYTAGLEI
               L   YN+A+PVL+RS+EIP++E+GE+HALAKFAG MQLGD Y ++GQ+E+SI  YT GL I
Sbjct:  262    LKEYNEAIPVLQRSVEIPVVEEGEEHALAKFAGLMQLGDTYAMVGQLESSISCYTEGLNI

Query:  228    QRQVLGESDARVGETCRYLAEAHVQAMQFEEASRLCQMALDIHKENGAAATASIEEAADR
               Q++VLGE+D RVGETCRYLAEA VQA++F+EA ++C+ AL IH+E+G            DR
Sbjct:  322    QKKVLGENDPRVGETCRYLAEALVQALRFDEAQQVCETALSIHRESGLPGSIAEAA--DR

Query:  288    KLMGLICDAKGDYEVALEHYVLASMAMSSQNHREDVAAVDCSIGDAYMSLARFDEAIFAY
               +LMGLIC+ KGD+E ALEH VLASMAM++     +VA VD SIGD+Y+SL+RFDEAI AY
Sbjct:  380    RLMGLICETKGDHENALEHLVLASMAMAANGQESEVAFVDTSIGDSYLSLSRFDEAICAY

Query:  348    QKALAVFKQGKGETHSSVALVYVRLADLYNKIGKTRDSKSYCENALKIYLKPTPGTPMEE
               QK+L   K  KGE H +V  VY+RLADLYN+ GK R++KSYCENAL+IY          EE
Sbjct:  440    QKSLTALKTAKGENHPAVGSVYIRLADLYNRTGKVREAKSYCENALRIYESHNLEISPEE

Query:  408    VATGFIEISAIYQSMNELDQALKLLRRALKIYANAPGCQNTIAGIEAQMGVVTYMMGNYS
               +A+G  +IS I +SMNE++QA+ LL++ALKIYA++PGQ+  IAGIEAQMGV+ YMMG Y
Sbjct:  500    IASGLTDISVICESMNEVEQAITLLQKALKIYADSPGQKIMIAGIEAQMGVLYYMMGKYM

Query:  468    ES-TTYSRARYQSSAIAREKTALFGIALNQMGLACVQRYAINEAADLFEEARTILEKECG
               ES  T+  A  +  A   +++ FGIALNQMGLAC+Q  AI EA +LFEEA+ ILE+ECG
Sbjct:  560    ESYNTFKSAISKLRATGKKQSTFFGIALNQMGLACIQLDAIEEAVELFEEAKCILEQECG

Query:  528    PYHPDTLAVYSNLAGTYDAMGRLDDAIEILEYVVGTREEKLGTANPEVEDEKQRLAALLK
               PYHP+TL +YSNLAG YDA+GRLDDAI++L +VVG REEKLGTANP  EDEK+RLA LLK
Sbjct:  620    PYHPETLGLYSNLAGAYDAIGRLDDAIKLLGHVVGVREEKLGTANPVTEDEKRRLAQLLK

Query:  588    EAGRGRSKRNRALLTLLDNN
               EAG     ++ ++L TL+D++
Sbjct:  680    EAGNVTGRKAKSLKTLIDSD
```

Figure 18: Similarity of p14A (Query) and *Arabidopsis* kinesin light chain (Sbjct, GenBank Accession No. AC004557) amino acid sequences. Standard one letter abbreviations are used for the amino acids. Identical amino acids in the two sequences are indicated with the same letter used for the amino acid. + indicates a similar amino acid in the two sequences. - indicates a gap introduced to optimize alignments. Numbers indicate the amino acid positions.

In fact, p14A shows a high level of conservation in this region. The TPR motif is a degenerate conserved sequence of 34 amino acid residues that is often arranged in tandem arrays and predicted to form an α-helix (Gobel and Yanagida, 1991). It was first identified in *CDC23*, *CDC16*, and *NUC2*, three cell cycle genes required for completion of mitosis in yeast (Sikorski *et al.*, 1990). Subsequent studies have revealed that TPRs exist in multiple copies in a diverse group of proteins that function in mitosis, transcription, RNA splicing, protein import, protein folding, and serine/threonine phosphorylation (Gobel and Yanagida, 1991).

Structural analyses and mutational studies of the TPR motif suggest that it may serve as an interface to mediate protein-protein interactions (Haucke *et al.*, 1996). For example, protein import into yeast mitochondria is mediated by four outer membrane proteins which function as two receptor subcomplexes: Tom37/Tom70 and Tom20/Tom22 (see introduction). The cytosolic domains of Tom37 and Tom20 each contain one TPR motif, whereas the cytosolic domain of Tom70 contains seven such motifs (Gobel and Yanagida, 1991). The TPR motifs of Tom70 are functionally important, as a C-terminally truncated Tom70 lacking most of its TPR motifs was inactive as an import receptor (Gobel and Yanagida, 1991). Mutation of the TPR motif of Tom20 impaired protein import into mitochondria (Haucke *et al.*, 1996). Thus, we can speculate that p14A interacts with the *Arabidopsis* tRNA nucleotidyltransferase via its TPR motifs. Because both the TPR motif and kinesin are associated with intracellular transport, clone 14A is a good candidate for further study. However, as cytoskeletal proteins frequently appeared as false positives in the yeast two-hybrid system (Serebriiskii *et al.*, 2000), the interaction between p14A and the *Arabidopsis* tRNA

nucleotidyltransferase detected in this work needs to be demonstrated by *in vitro*

biochemical methods such as immunoprecipitation or cross-linking or by *in vivo* studies.

Clone 19A could encode a 129-amino acid polypeptide (p19A) from the first in-

frame start codon at nucleotide 26 to the first in-frame stop codon at nucleotide 413

(Figure 15). The predicted protein shows 39% identity over its complete length to the

129-amino acid human KE2 protein (GenBank Accession No. O15212) (Figure 19). No

matches were found in the *Arabidopsis* database. This suggests that clone 19A may

represent a full length cDNA coding for the *Arabidopsis* homologue of the human KE2

protein. The function of the human KE2 protein is still unknown, however, the predicted

primary sequences of KE2 proteins are conserved between human, mouse and yeast, *i.e.*,

47% identity between yeast and mouse proteins which are predicted to contain 114 amino

acids and 127 amino acids, respectively (Shang *et al.*, 1994). Experiments preformed by

Geissler *et al.* (1998) showed that yeast KE2 protein promotes the formation of functional

α- and γ-tubulin, the subunits of microtubules. The authors predicted that it may function

in protein folding and assembly processes, possibly as a molecular chaperone. Deletion

of the *KE2* gene causes microtubule defects in yeast cells, resulting in a super-sensitive

phenotype towards the microtubule-depolymerizing drug benomyl and synthetical

lethality with a mutated yeast γ-tubulin (Geissler *et al.* 1998). However, these

phenotypes were complemented by the mouse homologue, indicating that the function of

the KE2 protein is conserved between yeast and mammals. Therefore, it is possible that

it is also conserved in plants. Clone 19A may encode an *Arabidopsis* homologue of the

human and yeast KE2 proteins. If the yeast KE2 protein functions as a molecular

chaperone, it is not surprising that its *Arabidopsis* homologue can interact with tRNA

```
Query:  17  LQRDLENKANDLGKIQKDIGKNHQLRKKYTIQLGENELVLKELDLLEEDANVYKLIGPVL
            +Q+ L+ +      ++QKD+ K+   R+K   QL EN  V   EL LL      V+KL+GPVL
Sbjct:  5   IQKKLQGEVEKYQQLQKDLSKSMSGRQKLEAQLTENNIVKEELALLDGSNVVFKLLGPVL

Query:  77  VKQDLAEANANVRKRIEYISAELKRLDAILQDMEEKQNNKRETIMKLQQRLQTIQAGKAK
            VKQ+L EA A V KR++YI+AE+KR ++ L+D+E +    +RET+ +LQQ  Q  QA KA
Sbjct:  65  VKQELGEARATVGKRLDYITAEIKRYESQLRDLERQSEQQRETLAQLQQEFQRAQAAKAG

Query:  137 A
            A
Sbjct:  125 A
```

Figure 19: Similarity of p19A (Query) and human KE2 protein (Sbjct, GenBank Accession No. O15212) amino acid sequences. Standard one letter abbreviations are used for the amino acids. Identical amino acids in the two sequences are indicated with the same letter used for the amino acid. + indicates a similar amino acid in the two senquences. Numbers indicate the amino acid positions.

nucleotidyltransferase, as molecular chaperones such as Hsp70, Com70 and MSF have already been shown to be involved in protein folding, assembly and intracellular trafficking (Hachiya *et al.*, 1993). Again, the association of the KE2 protein with cytoskeletal elements may make the specificity of interaction suspect and further experiments will be required to determine the significance of the interaction of the KE2 protein with the tRNA nucleotidyltransferase.

Clone 22A could encode a 180-amino acid polypeptide (p22A) from the beginning of the cDNA to the first in-frame stop codon at nucleotide 542 (Figure 16). The predicted protein shows 58% identity to the C-terminal 178 amino acids of cucumber raffinose synthase which contains 784 amino acids (GenBank Accession No. AF073744) (Figure 20), suggesting that if this clone codes for *Arabidopsis* raffinose synthase it represents an incomplete cDNA. Raffinose synthase (EC2.4.1.82) catalyzes the transfer of a galactosyl residue from galactinol to sucrose to form raffinose. It is one of the key enzymes in the biosynthesis of the raffinose family of oligosaccharides. However, since no work has been carried out as yet to define the role of any of its functional domains, there is no explanation as to why this enzyme or a protein with a similar domain to this C-terminal region would interact with the tRNA nucleotidyltransferase.

Finally, clone 108 could encode a 188-amino acid polypeptide (p108) from the first in-frame start codon at position 11 to the first in-frame stop codon at position 575 (Figure 17). When the predicted protein sequence was used to search the GenBank and *Arabidopsis* databases, the highest score of match found in the Genbank database was a 164-amino acid hypothetical protein from *Synechocystis* sp. with only 37% identity (GenBank Accession No. D90901) (Figure 21). No matches were found in the

85

```
Query:   1    GAFNCQGGGWCRETRRNQCFSECVNTLTATTSPKDVEWNSGSSPISIANVEEFALFLSQS
              GAFNCQGGGWCRETRRNQCFS-    +T+ T+PKD+EW+SG +PISI  V+ FAL L Q
Sbjct: 599    GAFNCQGGGWCRETRRNQCFSQYSKRVTSKTNPKDIEWHSGENPISIEGVKTFALYLYQA

Query:  61    KKLLLSGLNDDLELTLEPFKFELITVSPVVTIEGNSVRFAPIGLVNMLNTSGAIRSLVYN
              KKLL S    DL +PF+FELITVSPV +   S+ FAPIGLVNMLNTSGAI+S+ Y+
Sbjct: 659    KKLILSKPSQDLDIALDPFEFELITVSPVTKLIQTSLHFAPIGLVNMLNTSGAIQSVDYD

Query: 121    DE--SVEVGVFGAGEFRVYASKKPVSCLIDGEVVEFGY-EDSMVMVQVPW---SGPDGLS
              D+  SVE+GV G GE RV+ASKKP +C IDGE V F Y +D MV+VQVPW   S  G+S
Sbjct: 719    DDLSSVEIGVKGCGEMRVFASKKPRACRIDGEDVGFKYDQDQMVVVQVPWPIDSSSGGIS

Query: 175    SIQY
              I+Y
Sbjct: 779    VIEY
```

Figure 20: Similarity of p22A (Query) and cucumber raffinose synthase (Sbjct, GenBank Accession No. AF073744) amino acid sequences. Standard one letter abbreviations are used for the amino acids. Identical amino acids in the two sequences are indicated with the same letter used for the amino acid. + indicates a similar amino acid in the two senquences. - indicates a gap introduced to optimize alignments. Numbers indicate the amino acid positions.

```
Query:  11   VGKLNLRAHQEGGFFAETFRDSSVFLSTSQLPPTFKVDIAVSTSIYFLLPSGSVSRLHRI
             V KL L  H EGGF+ ET+R S V  + +      F    V T IYFL+   + S  HRI
Sbjct:  9    VEKLELLPHPEGGFYKETYR-SPVEANFAG----FDGSRNVVTGIYFLMTKDNFSAFHRI

Query:  71   PMAETWHFYLGEPLTVVELYDDGKLKFTCLGPDLFEGDQKPQYTVPPNVWFGSFPTKDVH
             E WHFY G+ L +       G+L+   LG DL +G+   PQ  V   + WF S
Sbjct:  64   KSDEMWHFYAGDSLEIYWFSPQGELEVISLGLDLEKGEV-PQAVVLKDGWFAS------

Query:  131  FSQDGALLKAEARDSENHFSLVGCTCAPAFQFEDFELAKRSDLLSRFPQHESLITMLS
                    +    ++LVGCT AP F F+DFELA R DLL  +P+   +I   L+
Sbjct:  116  ----------RVKNGGDYALVGCTVAPGFDFQDFELANREDLLKLYPRSAEIINQLT
```

Figure 21: Similarity of p108 (Query) and a hypothetical protein from *Synechocystis* sp. (Sbjct, GenBank Accession No. D90901) amino acid sequences. Standard one letter abbreviations are used for the amino acids. Identical amino acids in the two sequences are indicated with the same letter used for the amino acid. + indicates a similar amino acid in the two senquences. - indicates a gap introduced to optimize alignments. Numbers indicate the amino acid positions.

Arabidopsis database. Therefore, clone 108 encodes a protein of unknown function.

As mentioned in the introduction, many proteins have already been identified as being involved in protein targeting. However, none of these was isolated in this work through their interaction with the lupin or *Arabidopsis* tRNA nucleotidyltransferase. One possible reason is that while the whole *Arabidopsis* cDNA library contains 36×10^6 independent recombinants (according to the instructions supplied with the cDNA library), only 22×10^6 transformants were screened with each bait plasmid, so roughly 40% of the library was not screened. Moreover, some clones in the library may have been lost due to repeated amplification of the library both at the *Arabidopsis* Biological Resource Center and here. However, the formal possibility still exists that the lupin and *Arabidopsis* tRNA nucleotidyltransferases encoded by the cDNAs used in this work are cytosolic enzymes and therefore would not be expected to interact with organellar targeting factors.

V. Conclusions

In this work, a cDNA (122M21T7) encoding *Arabidopsis* tRNA nucleotidyltransferase was sequenced completely. The predicted translation of this cDNA showed an open reading frame of 1764 bp starting from nucleotide 1, extending to a TGA at position 1765, and potentially encoding 588 amino acids with the first in-frame start codon at amino acid 47. When compared with the *Arabidopsis CCA* gene sequence from the *Arabidopsis* database, it was found that the open reading frame of the cDNA can be extended further upstream so that it contains three in-frame start codons at amino acids 1, 6 and 69. The ability of the *Arabidopsis* cDNAs aracca2 (beginning at ATG6) and aracca3 (beginning at ATG69) to produce functional enzymes as GAL4 fusion proteins

was shown by complementation of a temperature-sensitive mutation in the yeast *CCA1* gene, indicating that the amino acids between ATG6 and ATG69 are not essential for enzyme activity. After screening an *Arabidopsis* cDNA library in the yeast two-hybrid system with the lupin and *Arabidopsis* tRNA nucleotidyltransferases as baits, four different cDNAs were isolated, and then sequenced completely. The potential functions of the predicted proteins encoded by these cDNAs were predicted after searching the GenBank and *Arabidopsis* databases. Two of these proteins (p14A and p19A) appear to be good candidates for further studies as their potential homologues in the databases are predicted to be involved in protein transport or folding and assembly. The reasons for the interactions of the other two proteins with tRNA nucleotidyltransferases remain to be discovered.

REFERENCES

Adam, E.J. and Adam, S.A. (1994) Identification of cytosolic factors required for nuclear sequence-mediated binding to the nuclear envelope. Journal of Cell Biology 125: 547-555.

Adam, S.A. and Gerace, L. (1991) Cytosolic proteins that specifically bind nuclear location signals are receptors for nuclear import. Cell 66: 837-847.

Adam, S.A., Sterne-Marr, R. and Gerace, L. (1990) Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors. Journal of Cell Biology 111: 807-816.

Aebi, M., Kirchner, G., Chen, J., Vijayraghavan, U., Jacobson, A., Martin, N.C. and Abelson, J. (1990) Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. Journal of Biological Chemistry 265: 16216-16220.

Akashi, K., Grandjean,O and Small, I. (1998) Potential dual targeting of an *Arabidopsis* archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. FEBS Letters 431: 39-44

Bai, C. and Elledge, S.J. (1996) Gene identification using the yeast two-hybrid system. Methods in Enzymology 273: 331-347.

Bartel, P.L., Chien, C.T., Sternglanz, R. and Fields, S. (1993) Elimination of false positives that arise in using the two-hybrid system. Biotechniques 14: 920-924.

Becker, K., Guiard, B., Rassow, J., Sollner, T. and Pfanner, N. (1992) Targeting of a chemically pure preprotein to mitochondria does not require the addition of a cytosolic signal recognition factor. Journal of Biological Chemistry 267: 5637-5643.

Bewsey, K.E., Johnson, M.E. and Huff, J.P. (1991) Rapid isolation and purification of DNA from agarose gels: The phenol-freeze-fracture method. Biotechniques 10: 724-725.

Bloom, G., Wagner, M., Pfister, K. and Brady. S. (1988) Native structure and physical properties of bovine brain kinesin and identification of the ATP-binding subunit polypeptide. Biochemistry 27: 3409-3416.

Bloom, G.S. and Endow, S.A. (1995) Motor proteins 1: kinesins. Protein Profile 1: 1059-1088.

Cartwright, P., Beilharz, T., Hansen, P., Garrett, J. and Lithgow, T. (1997) Mft52, an acid-bristle protein in the cytosol that delivers precursor proteins to yeast mitochondria. Journal of Biological Chemistry 272: 5320-5325.

Chaumont, F., O'Riordan, V. and Boutry, M. (1990) Protein transport into mitochondria is conserved between plant and yeast species. Journal of Biological Chemistry 265: 16856-16862.

Chelsky, D., Ralph, R. and Jonak, G. (1989) Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Molecular Cell Biology 9: 2487-2492.

Chen, J.Y., Kirchner, G., Aebi, M. and Martin, N.C. (1990) Purification and properties of yeast ATP (CTP):tRNA nucleotidyltransferase from wild type and overproducing cells. Journal of Biological Chemistry 265: 16221-16224.

Chen, J.Y., Joyce, P.B.M., Wolfe, C.L., Steffen, M.C. and Martin, N.C. (1992) Cytoplasmic and mitochondrial tRNA nucleotidyltransferase activities are derived from the same gene in the yeast Saccharomyces cerevisiae. Journal of Biological Chemistry 267: 14879-14883.

Cortes, P., Ye, Z.S. and Baltimore, D. (1994) RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1. Proceedings of the National Academy of Sciences USA 91: 7633-7637.

Creissen, G., Reynolds, H., Xue, Y. and Mullineaux, P. (1995) Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplast and mitochondria in transgenic tobacco. The Plant Journal 8: 167-175.

Cudny, H., Pietrzak, M. and Kaczkowski, J. (1978) Plant tRNA nucleotidyltransferase: I. Isolation and purification of tRNA nucleotidyltransferase from Lupinus luteus seeds. Planta 142: 23-27.

Deane, R., Schafer, W., Zimmermann, H.P., Mueller, L., Gorlich, D., Prehn, S., Ponstingl, H. and Bischoff FR. (1997) Ran-binding protein 5 (RanBP5) is related to the nuclear transport factor importin-beta but interacts differently with RanBP1. Molecular and Cellular Biology 17: 5087-96.

Deshaies, R., Koch, B., Werner-Washburne, M., Craig, E. and Schekman, R. (1988) A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. Nature 332: 800-805.

Deutscher, M.P. (1984) Processing of tRNA in prokaryotes and eukaryotes. Critical Reviews in Biochemistry 17: 45-71.

Deutscher, M.P. (1990) Ribonucleases, tRNA nucleotidyltransferase, and the 3' processing of tRNA. Nucleic Acids Research 39: 209-240.

Deutscher, M., Foulds, J. and McClain, W.H. (1974) Transfer ribonucleic acid nucleotidyltransferase plays an essential role in the normal growth of *Escherichia coli* and in the biosynthesis of some bacteriophage T4 transfer ribonucleic acids. Journal of Biological Chemistry 249: 6696-6699.

Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences-a consensus? Trends in Biochemical Sciences 16: 478-481.

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

Dingwall, C., Kandels-Lewis, S. and Seraphin, B. (1995) A family of Ran binding proteins that includes nucleoporins. Proceedings of the National Academy of Sciences USA 92: 7525-7529.

Dullin, P., Fabisz-kajiwska, A. and Walerych, W. (1975) Isolation and properties of tRNA nucleotidyltransferase from wheat embryos. Acta Biochimica Polonica 22: 279-289.

Durfee, T., Becherer, K., Chen, P., Yeh, S., Yang, Y., Kilbum, A., Lee, W. and Elledge, S. (1993) The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Gene & Development 7: 555-569.

Elledge, S., Mulligan, J., Ramer, S., Spottswood, M. and Davis, R. (1991) λYES: a multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. Proceedings of the National Academy of Sciences USA 88: 1731-1735.

Endo, T., Mitsui, S., Nakai, M. and Roise, D. (1996) Binding of mitochondrial presequences to yeast cytosolic heat shock protein 70 depends on the amphiphilicity of the presequence. Journal of Biological Chemistry 271: 4161-4167.

Estojak, J., Brent, R. and Golemis, E. (1995) Correlation of two-hybrid affinity data with *in vitro* measurements. Molecular and Cellular Biology 15: 5820-5829.

Featherstone, C., Darby, M.K. and Gerace, L. (1988) A monoclonal antibody against the nuclear pore complex inhibits nucleocytoplasmic transport of protein and RNA *in vivo*. Journal of Cell Biology 107: 1289-1297.

Fellotter, H., Hannon, G., Ruddle, C. and Beach, D. (1994) Construction of an improved host strain for two hybrid screening. Nucleic Acids Research 22: 1502-1503.

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. Nature 340: 245-247.

Gauger, A. and Goldstein, L. (1993) The Drosophila kinesin light chain. Primary structure and interaction with kinesin heavy chain. Journal of Biological Chemistry 268: 13657-13666.

Geissler, S., Siegers, K. and Schiebel, E. (1998) A novel protein complex promoting formation of functional alpha- and gamma-tubulin. EMBO Journal 17: 952-966.

George, R., Beddoe, T., Landl, K. and Lithgow,T. (1998) The yeast nascent polypeptide-associated complex initiates protein targeting to mitochondria *in vivo*. Proceedings of the National Academy of Sciences USA 95: 2296-2301.

Gietz, D., Jean, A., Woods, R.A. and Schiestl, R.H. (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Research 20: 1425.

Gindhart, J. and Goldstein, L. (1996) Tetratrico peptide repeats are present in the kinesin light chain. Trends in Biochemical Sciences 21: 52-53.

Gish, W. and States, D. (1993) Identification of protein coding regions by database similarity search. Nature Genetics 3: 266-272.

Glick, B. and Schatz, G. (1991) Import of proteins into mitochondria. Annual Review of Genetics 25: 21-44.

Gobel, M. and Yanagida, M. (1991) The TPR snap helix: a novel protein repeat motif from mitosis to transcription. Trends in Biochemical Sciences 16: 173-177.

Gorlich, D. (1998) Transport into and out of the cell nucleus. EMBO Journal 17: 2721-2727.

Gorlich, D. and Mattaj, L. (1996) Nucleocytoplasmic transport. Science 271: 1513-1518.

Gorlich, D., Vogel, F., Mills, A., Hartman, E. and Laskey, R. (1995a) Distinct function for the two importin subunits in nuclear protein import. Nature 377: 246-248.

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

Gorlich D, Dabrowski M, Bischoff FR, Kutay U, Bork P, Hartmann E, Prehn, S. and Izaurralde E. (1997) A novel class of RanGTP binding proteins. Journal of Cell Biology 138: 65-80.

Greenberg, B.M. and Hallick, R.B. (1986) Accurate transcription and processing of 19 *Euglena* chloroplast tRNAs in a *Euglena* soluble extract. Plant Molecular Biology 6: 89-100.

Guarente, L. (1983) Yeast promoters and *lacZ* fusions designed to study expression of cloned genes in yeast. Methods in Enzymology 101: 181-191.

Hachiya, N., Komiya, T., Sakasegawa, Y., Sakaguchi, M., Mihara, K. and Omura, T. (1993) A mitochondrial import factor purified from rat liver cytosol is an ATP-dependent conformational modulator for precursor proteins. EMBO Journal 12: 1579-1586.

Hachiya, N., Komiya, T., Alam, R., Iwahashi, J., Sakaguchi, M., Omura, T. and Mihara, K. (1994) MSF, a novel cytoplasmic chaperone which functions in precursor targeting to mitochondria. EMBO Journal 13: 5146-5154.

Haizel, T., Merkle, T., Pay, A., Fejes, E. and Nagy, F. (1997) Characterization of proteins that interact with the GTP-bound form of the regulatory GTPase Ran in *Arabidopsis*. The Plant Journal 11: 93-103.

Hall, M.N., Hereford, L. and Herskowitz, I. (1984) Targeting of *E. coli* β-galactosidase to the nucleus in yeast. Cell 36: 1057-1065.

Hanic-Joyce, P. and Gray, M.W. (1990) Processing of transfer RNA precursors in a wheat mitochondrial extract. Journal of Biological Chemistry 265: 13782-13791.

Hartl, F.U. (1996) Molecular chaperones in cellular protein folding. Nature 381: 571-579.

Haucke, V., Lithgow, T., Rospert, S., Hahne, K. and Schatz, G. (1995) The yeast mitochondrial protein import receptor Mas20p binds precursor proteins through electrostatic interaction with the positively charged presequence. Journal of Biological Chemistry 270: 5565-5570.

Haucke, V., Horst, M., Schatz, G. and Lithgow, T. (1996) The Mas20p and Mas70p subunits of the protein import receptor of yeast mitochondria interact via the tetratricopeptide repeat motif in Mas20p: evidence for a single hetero-oligomeric receptor. EMBO Journal 15: 1231-1237.

Heese-Peck, A. and Raikhel, N.V. (1998) The nuclear pore complex. Plant Molecular Biology 38: 145-162.

Heins, L. and Schmitz, U.K. (1996) A receptor for protein import into potato mitochondria. The Plant Journal 9: 829-839.

Hicks, G.R. and Raikhel, N.V. (1993) Specific binding of nuclear localization sequences to plant nuclei. The Plant Cell 5: 983-994.

Hicks, G.R. and Raikhel, N.V. (1995) Nuclear localization signal binding proteins in higher plant nuclei. Proceedings of the National Academy of Sciences USA 92: 734-738.

94

Hicks, G.R., Smith, M.S., Shieh, M. and Raikhel, N.V. (1995) Three classes of nuclear import signals bind to plant nuclei. Plant Physiology 107: 1055-1058.

Hicks, G.R., Smith, M.S., Lobreaux, S. and Raikhel, N.V. (1996) Nuclear import in permeabilized protoplasts from higher plants has unique features. The Plant Cell 8: 1337-1352.

Hines, V., Brandt, A., Griffiths, G., Horstmann, H., Brutsch, H. and Schatz, G. (1990) Protein import into yeast mitochondria is accelerated by the outer membrane protein MAS70. EMBO Journal 9: 3191-3200.

Hinnah, S., Hill, K., Wagner, R., Schlicher, T. and Soll, J. (1997) Reconstitution of a chloroplast protein import channel. EMBO Journal 16: 7351-7360.

Hirsch, S., Muckel, E., Heemeyer, F., Heijne, G.V. and Soll, J. (1994) A receptor component of the chloroplast protein translocation machinery. Science 266: 1989-1992.

Hoffman, C.S. and Winston, F. (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57: 267-272.

Honlinger, A., Kubrich, M., Moczko, M., Gartner, F., Mallet, L., Bussereau, F., Eckerskorn, C., Lottspeich, K., Jacquet, M. and Pfanner, N. (1995) The mitochondrial receptor complex: Mom22 is essential for cell viability and directly interacts with preproteins. Molecular and Cellular Biology 15: 3382-3389.

Iovine, M.K., Watkins, J.L. and Wente, S.R. (1995) The GLFG repetitive region of the nucleoporin Nup116p interacts with Kap95p, an essential yeast nuclear import factor. Journal of Cell Biology 131: 1699-1713.

Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. Journal of Bacteriology 153: 163-168.

Kalderon, D., Rcihardson, W.D., Markham, A.F. and Smith, A.E. (1984) Sequence requirements for nuclear location of simian virus 40 large T antigen. Nature 311: 33-8.

Keegstra, K. and Cline, K. (1999) Protein import and routing systems of chloroplasts. The Plant Cell 11: 557-570.

Kessler, F., Blobel, G., Patel, H.A. and Schnell, D.J. (1994) Identification of two GTP-binding proteins in the chloroplast protein import machinery. Science 266: 1035-1039.

Kiebler, M., Keil, P., Schneider, H., Pfanner, N. and Neupert, W. (1993) The mitochondrial receptor complex: a central role of MOM22 in mediating preprotein transfer from receptors to the general insertion pore. Cell 74: 483-492.

Kim, J., Harter, K. and Theologis, A. (1997) Protein-protein interactions among the Aux/IAA proteins. Proceedings of the National Academy of Sciences USA 94: 11786-11791.

Knapp, G., Bechmann, J.S., Jhonson, P.F., Fuhrman, S.A. and Abelson, J. (1978) Transcription and processing of intervening sequences in yeast tRNA genes. Cell 14: 221-236.

Koepp, D.M. and Silver, P.A. (1996) A GTPase controlling nuclear trafficking: running the right way or walking randomly? Cell 87: 1-4.

Komine, Y., Adachi, T., Inokuchi,H. and Ozeki, H. (1990) Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. Journal of Molecular Biology 212: 579-598.

Kouranov, A. and Schnell, D.J. (1997) Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. Journal of Cell Biology 139: 1677-1685.

Kourtz, L. and Ko, K. (1997) The early stage of chloroplast protein import involves Com70. Journal of Biological Chemistry 272: 2808-2813.

Kraemer, D., Wozniak, R.W., Blobel, G. and Radu, A. (1994) The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. Proceedings of the National Academy of Sciences USA 91: 1519-1523.

Li, B. and Fields, S. (1993) Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. FASEB Journal 7: 957-963.

Martin, N.C. and Hopper, A.K. (1994) How single genes provide tRNA processing enzymes to mitochondria, nuclei and the cytosol. Biochimie 76: 1161-1167.

Masiakowski, P. and Deutscher, M.P. (1980) Dissection of the active site of rabbit liver tRNA nucleotidyltransferase. Specificity and properties of the tRNA and acceptor subsites determined with model acceptor substrates. Journal of Biological Chemistry 255: 1233-1239.

Ma, Y.K., Kouranov, A., LaSala, S.E. and Schnell, D.J. (1996) Two components of the chloroplast protein import apparatus, IAP86 and IAP75, interact with the transit sequence during the recognition and translocation of presursor proteins at the outer envelope. Journal of Cell Biology 134: 315-327.

Mayer, D., Nargang, F.E., Neupert, W. and Lill, R. (1995) MOM22 is a receptor for mitochondrial targeting sequences and cooperates with MOM19. EMBO Journal 14: 4204-4211.

Melchior, F., Paschal, B., Evans, E. and Gerace, L. (1993) Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. Journal of Cell Biology 123: 1649-1659.

Menand, B., Drouard, L., Sakamoto, W., Dietrich, A. and Wintz, H. (1998) A single gene of chloroplast origin codes for mitochondrial and chloroplastic metnionyl-tRNA synthetase in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences USA 95: 11041-11019.

Mireau, H., Lancelin, D. and Small, I. (1996) The same *Arabidopsis* gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetase. The Plant Cell 8: 1027-1039.

Moczko, M., Ehmann, B., Gartner, F., Honlinger, A., Schafer, E. and Pfanner, N. (1994) Deletion of the receptor MOM19 strongly impairs import of cleavable preproteins into *Saccharomyces cerevisiae* mitochondria. Journal of Biological Chemistry 269: 9045-9051.

Mooney, B. and Harmey, M.A. (1996) The occurrence of hsp70 in the outer membrane of plant mitochondria. Biochemical and Biophysical Research Communications 218: 309-313.

Moore, M.S. and Blobel, G. (1993) The GTP-binding protein Ran is required for protein import into the nucleus. Nature 365: 661-663.

Mukerji, S.K. and Deutscher, M.P. (1972) Reaction at the 3' terminus of transfer ribonucleic acid: Subcellular localization and evidence for a mitochondrial transfer ribonucleic acid nucleotidyltransferase. Journal of Biological Chemistry 247: 481-488.

Murakami, H., Pain, D. and Blobel, G. (1988) 70-kD heat shock-related protein is one of at least two distinct cytosolic factors stimulating protein import into mitochondria. Journal of Cell Biology 107: 2051-2057.

Mukumoto, F., Hirose, S., Imaseki, H. and Yamazaki, K. (1993) DNA sequence requirement of a TATA element-binding protein from *Arabidopsis* for transcription *in vivo*. Plant Molecular Biology 23: 995-1003.

Neupert, W. (1997) Protein import into mitochondria. Annual Review of Biochemistry. 66: 863-917.

Nigg, E.A. (1997) Nucleocytoplasmic transport: signals, mechanisms and regulation. Nature 386: 779-787.

Paschal, B.M. and Gerace, L. (1995) Identification of NTF2, a cytosolic factor for nuclear import that interacts with nuclear pore complex protein p62. Journal of Cell Biology 129: 925-937.

Perry, S.E. and Keegstra, K. (1994) Envelope membrane proteins that interact with chloroplastic precursor proteins. The Plant Cell 6: 93-105.

Poblete. P., Jedlicky. E. and Litvak. S. (1977) Purification and properties of tRNA nucleotidyltransferase from *Musca domestica*. Biochimica et Biophysica Acta 476: 333-341.

Ramage, L., Junne, T., Hahne, K., Lithgow, T. and Schatz, G. (1993) Functional cooperation of mitochondrial protein import receptors in yeast. EMBO Journal 12: 4115-4123.

Sack. S., Kull. J. and Mandelkow, E. (1999) Motor proteins of the kinesin family. European Journal of Biochemistry 262: 1-11.

Sambrook, J., Fristsch. E.F. and Maniatis, T. (1989) Molecular Cloning: A laboratory manual. Second edition, Cold Spring Harbor, NY.

Saunders. W.S., Koshland, D., Eshel, D., Gibbons, I. and Hoyt, M. (1995) *Saccharomyces cerevisiae* kinesin- and dynein-related proteins required for anaphase chromosome segregation. Journal of Cell Biology 128:617-624

Schiestl. R.H. and Gietz, R.D. (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Current Genetics 16: 339-346.

Schmitz, U. and Lonsdale, D.M. (1989) A yeast mitochondrial presequence functions as a signal for targeting to plant mitochondria *in vivo*. The Plant Cell 1: 783-791.

Schnell, D.J., Kessler, F. and Blobel, G. (1994) Isolation of components of the chloroplast protein import machinery. Science 266: 1007-1012.

Serebriiskii, I., Estojak, J., Berman, M. and Golemis, E. (2000) Approachs to detecting false positives in yeast two-hybrid system. Biotechniques 28: 328-336.

Shang, H., Wong, S., Tan, H. and Wu, M. (1994) YKE2, a yeast nuclear gene encoding a protein showing homology to mouse KE2 and containing a putative leucine-zipper motif. Gene 151: 197-201.

Shanmugam, K., Hanic-Joyce, P.J. and Joyce, P.B.M. (1996) Purification and characterization of a tRNA nucleotidyltransferase from *Lupinus albus* and functional complementation of a yeast mutation by the corresponding cDNA. Plant Molecular Biology 30: 281-295.

Sherman, F. (1991) Getting started with yeast. Methods in Enzymology 194: 3-21.

Shi, Y. and Thomas, J.O. (1992) The transport of proteins into the nucleus requires the 70-kilodalton heat shock protein or its cytosolic cognate. Molecular Cell Biology 12: 2186-2192.

Sikorski, R., Boguski, M., Goebl, M. and Hieter, P. (1990) A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell 60: 307-317.

Small, I., Wintz, H., Akaski, K. and Mireau, H. (1998) Two birds with one stone: genes that encode products targeted to two or more compartments. Plant Molecular Biology 38: 265-277.

Smith, M.S., Hicks, G.R. and Raikhel, N.V. (1997) Importin α from *Arabidopsis thaliana* is a nuclear import receptor that recognizes three classes of import signals. Plant Physiology 114: 411-417.

Snyder, E. and Storno, G. (1995) Identification of protein coding regions in genomic DNA. Journal of Molecular Biology 248: 1-18.

Soll, J. and Waegemann, K. (1992) A functionally active protein import complex from chloroplasts. The Plant Journal 2: 253-256.

Soll, J. and Tien, R. (1998) Protein translocation into and across the chloroplastic envelope membranes. Plant Molecular Biology 38: 191-207.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Research 26: 148-153.

Starr, M.C. D'Onofrio, M., Park, M.C. and Hanover, J.A. (1990) Primary sequence and heterologous expression of nuclear pore glycoprotein p62. Journal of Cell Biology 110: 1861-1871.

Sukegawa, J. and Blobel, G. (1993) A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. Cell 72: 29-38.

Vicaretti, R. and Joyce, P.B.M. (1999) Comparison of mitochondrial and cytosolic tRNA nucleotidyltransferases from *Triticum aestivum*. Canadia Journal of Botany 77: 230-239.

Von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Domain structure of mitochondrial and chloroplast targeting peptides. European Journal of Biochemistry 180: 535-545.

Waegemann, K. and Soll, J. (1991) Characterization of the protein import apparatus in isolated outer envelopes of chloroplasts. The Plant Journal 1: 149-158.

Waegemann, K. and Soll, J. (1996) Phosphorylation of the transit sequence of chloroplast precursor proteins. Journal of Biological Chemistry 271: 6545-6554.

Waegemann, K., Paulsen, H. and Soll, J. (1990) Translocation of proteins into chloroplasts requires cytosolic factors to obtain import competence. FEBS Letters 261: 89-92.

Willims, K.R. and Schofield, P. (1977) Kinetic mechanism of tRNA nucleotidyltransferase from *Escherichia coli*. Journal of Biological Chemistry 252: 5589-5597.

Wolfe, C.L., Lou, L., Hopper, A.K. and Martin, N.C. (1994) Interplay of heterogeneous transcriptional start sites and translational selection of AUGs dictate the production of mitochondrial and cytosolic/nuclear tRNA nucleotidyltransferase from the same gene in yeast. Journal of Biological Chemistry 269: 13361-13366.

Zhu, L., Cudny, H. and Deutscher, M.P. (1986) A Mutation in *Escherichia coli* tRNA nucleotidyltransferase that affects only AMP incorporation is in a sequence often associated with nucleotide-binding proteins. Journal of Biological Chemistry 261: 14875-14877.