Sequence motifs that distinguish ATP(CTP):tRNA nucleotidyl transferases from eubacterial poly(A) polymerases

GEORGES MARTIN and WALTER KELLER
Department of Cell Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

ABSTRACT
ATP(CTP):tRNA nucleotidyl transferases, tRNA maturing enzymes found in all organisms, and eubacterial poly(A) polymerases, enzymes involved in mRNA degradation, are so similar that until now their biochemical functions could not be distinguished by their amino acid sequence. BLAST searches and analysis with the program “Sequence Space” for the prediction of functional residues revealed sequence motifs which define these two protein families. One of the poly(A) polymerase defining motifs specifies a structure that we propose to function in binding the 3′ terminus of the RNA substrate. Similar motifs are found in other homopolyribonucleotidyl transferases. Phylogenetic classification of nucleotidyl transferases from sequenced genomes reveals that eubacterial poly(A) polymerases have evolved relatively recently and are found only in a small group of bacteria and surprisingly also in plants, where they may function in organelles.

Keywords: nucleotidyl transferase; ATP(CTP):tRNA nucleotidyl transferase; poly(A) polymerase; RNA modification; RNA processing

INTRODUCTION
ATP(CTP):tRNA nucleotidyltransferases (we also use the synonyms CCA-adding enzymes, CCA transferases, or the abbreviation CCAtrs) are nucleotidyl transferases (Ntrs) responsible for the synthesis or repair of the 3′ terminal CCA sequence of tRNA molecules. They are coded by essential genes in almost all organisms and are able to add the three ribonucleotides C, C, and A to a tRNA’s 3′ end in a sequential order. CCAtrs were also identified and cloned from archaea (Yue et al. 1996; Seth et al. 2002). In addition, it was shown recently that in some eubacteria, two separate nucleotidyl transferases collaborate to build or repair a CCA tag by adding either CC or A to the 3′ ends of tRNA precursors (Tomita and Weiner 2001, 2002). Crystal structures of eubacterial, archaeal, and human CCAtrs have been solved recently (Li et al. 2002; Augustin et al. 2003; Okabe et al. 2003; Xiong et al. 2003).

Eubacterial poly(A) polymerases (eubPAPs, to distinguish them from eukaryotic poly(A) polymerases or PAPs) are RNA polymerases that add multiple AMPs to the 3′ ends of messenger RNAs. These poly(A) tails promote the degradation of the attached mRNAs by 3′-5′-exonucleases (Symmons et al. 2002). In Escherichia coli, eubPAP (Cao and Sarkar 1992) is a nonessential enzyme and its function can be replaced by polynucleotide phosphorylase (PNP; Mohanty and Kushner 2000), an enzyme which can both synthesize and degrade poly(A). PNP was also found to be responsible for poly(A) addition to mRNAs in spinach chloroplasts and cyanobacteria (Yehudai-Resheff et al. 2001; Rott et al. 2003).

eubPAPs and CCAtrs belong to a superfamily of nucleotidyl transferases (Martin and Keller 1996; Aravind and Koonin 1999), members of which share sequence homology mainly in the catalytic domain. They have been divided into class I and II according to specific sequence motifs in the catalytic domain (Yue et al. 1996). All class I Ntrs (Yue et al. 1996) which share the same active site signature, including archaeal CCAtrs, eukaryotic, nuclear, and regulatory poly(A) polymerases, and related enzymes such as DNA polymerase β (Pol β), terminal deoxynucleotidyl transferase (TdT), and 2′-5′ oligo(A) synthase, must share a common ancestor (Holm and Sander 1995; Martin and Keller 1996; Yue et al. 1996; Aravind and Koonin 1999). Class II Ntrs must have branched very early from class I Ntrs (Yue...
et al. 1996), and the two classes may even have evolved independently twice (Aravind and Koonin 1999). This is very likely if one considers the fact that the crystal structures of class I and class II CCAtrs differ quite extensively (Li et al. 2002; Augustin et al. 2003; Okabe et al. 2003; Xiong et al. 2003). For example, one of the obvious differences between the two classes of enzymes is that the helix that interacts with the phosphates of the incoming nucleotides has a single turn in class I Ntrs and two turns in class II enzymes (see below).

RESULTS AND DISCUSSION

Search for protein sequences with homology to bona fide eubacterial poly(A) polymerases and CCA transferases

We first conducted an extensive database search with the BLAST program for protein sequences with homology to experimentally verified CCAtrs or eubPAPs as query. Sequences with a probability threshold (E-value) of better than $10^{-7}$ were included for further analysis, although most positives had E-values of better than $10^{-10}$. Resulting multiple sequence alignments were then subjected to analysis by the computer program “Sequence Space,” which allows the determination of residues which should specify functional differences between protein families (Casari et al. 1995). This program designates the entire protein and sequence residues as vectors in a generalised sequence space. Projection of these vectors onto a lower-dimensional space reveals protein subfamilies as clusters and highlights characteristic residues.

We expected the Sequence Space program to identify conserved motifs that were characteristic for either CCAtrs or eubPAPs. Figure 1A depicts an example of an alignment displayed in the viewer window of the Sequence Space program, where motif-specific residues are highlighted in red (columns) and rows depict sequences (in blue) that contain motifs complying with eubPAP criteria (summarized in Table 1). Other Sequence Space viewer windows display either the spatial distribution of individual residues (Fig. 1B) or of the protein sequences (Fig. 1C). The figures illustrate how residues specific for CCAtrs (Fig. 1B) and a cluster of protein sequences of CCAtrs (Fig. 1C) are located to the left in both windows, whereas eubPAP-specific residues and sequences form clusters to the right in these windows. In a further step we systematically searched the entire collection of sequenced genome databases for the presence of eubPAPs and CCAtrs (for details see Materials and Methods). Table 1 contains a list of sequence signatures that were applied for the classification of positives, and the results are listed in Table 2.

Two distinct signatures emerged in the Sequence Space analysis (Table 1): The first corresponded to amino acids 108–118 of the E. coli eubPAP and had the consensus [LIV]-[LIV]-G-[RK]-[RK]-F-x-[LIV]-h-[HLQ]-[LIV], where x is any and h is a hydrophobic residue. This consensus signature is only found in bona fide eubPAPs but not in CCAtrs. Because there are no crystal structures of eubPAPs available, we tried to locate this motif in the closely related structures of bacterial (Li et al. 2002) and human CCAtr (Augustin et al. 2003). The corresponding sequence in these CCAtrs forms a turn or loop of variable size followed by β-strand 4 or 3, respectively. Since no structures of CCAtrs in complex with a primer are known, we superimposed the catalytic region of the class I enzyme Pol B including a primer (Sawaya et al. 1997) with the corresponding structure of the B. steatorrhophilus CCAtr (Fig. 2A). The resulting model shows a close contact between the terminal three nucleotides of the primer and the loop between β-strands 3 and 4 (“loop 3/4”) of the CCAtr structure.

A motif forming a similar loop 3/4 structure has been proposed to be important in primer binding in class I eukaryotic PAPs and in terminal uridylyl transferases (Keller and Martin 2002). Therefore, it appears that two groups of Ntrs possess a loop 3/4 motif with a strongly conserved sequence signature, whereas most of the remaining Ntrs have a loop 3/4 whose amino acid sequence is much less well conserved. Close inspection of these enzymes reveals that the two groups of proteins consist of Ntrs that synthesise homopolynucleotides (hprNtrs) and that the two groups belong to either class I or class II Ntrs. We designate these two types of loop 3/4 motifs according to the two classes of Ntrs “hprNS-I” and “hprNS-II”, for “homopolyribonucleotidyl transferase signature I and II.” An alignment with a representative selection of sequences belonging to the two classes of loop 3/4 motifs is shown in Figure 2B. All motifs contain a confirmed or predicted loop 3/4 structure about 14 residues upstream of the third catalytic Glu or Asp residue, except for terminal uridylyl transferase from T. brucei, where the distance is larger. The crystal structures of five proteins in the list are solved and result in a consensus of structural elements (Fig. 2B, bottom). Therefore, we predict that the group of hprNS-II proteins (the eubPAPs) contain a loop structure at the same position. Although sequences between groups are not strongly conserved for reasons of group specificity, hydrophobic residues or Pro/Gly are frequently conserved and are an indication for structure conservation. Loop 3/4 structures often feature residues with a ring-containing side chain (Phe or His) or the basic residues Lys or Arg. The conserved features may be specific for the type of primer involved and could be an indication that these side chains engage in stacking interactions with the primer via the ring or the aliphatic central part of lysine and arginine. Mutagenesis of Phe153, located on the loop, and Val156 and Lys158 on strand 4 of the hprNS-I of bovine PAP resulted in a strongly increased $K_{m}$ when titrated with oligo (A)$_{15}$ primer, an indication that several residues of the hprNS-I are involved in primer binding (G. Martin and W. Keller, unpubl.). A domain close to
FIGURE 1. Analysis of multiple sequence alignments with the program Sequence Space. (A) Part of a multiple sequence alignment of the Java viewer window of the Sequence Space program with eubPAPs and CCAtrs from a BLAST output with *E. coli* eubPAP sequence as query (top line). Only columns of the MSF alignment are displayed which contain the catalytic site region and motifs of interest. Some blocks of rows with very similar proteins were left out. Numbers at the top correspond to residues in the *E. coli* eubPAP protein sequence (Swiss-Prot accession pcnb_ecoli). Rows in blue indicate predicted eubPAP sequences, and rows in red highlight conserved residues in the upstream and the downstream motif. Swiss-Prot or TrEMBL accessions are listed to the left. At the bottom, asterisks (*) mark the three catalytic Asp or Glu residues, and # marks other invariant residues in Ntrs; $ indicates residues that belong to the upstream and downstream motifs, and the color code red is given for eubPAP-specific residues and black for residues invariant in both CCAtrs and eubPAPs. Rows in blue indicate predicted eubPAP sequences, and rows in red highlight conserved residues in the upstream and the downstream motif. Swiss-Prot or TrEMBL accessions are listed to the left. At the bottom, asterisks (*) mark the three catalytic Asp or Glu residues, and # marks other invariant residues in Ntrs; $ indicates residues that belong to the upstream and downstream motifs, and the color code red is given for eubPAP-specific residues and black for residues invariant in both CCAtrs and eubPAPs. (B) Display of the residues window of the Sequence Space Java viewer with a 3D view of the results of the alignment in A. To the left, the red * (circled and named E/251) indicates the position of the residues only found in CCAtrs corresponding to residue 251 in *E. coli* eubPAP in A. (C) The protein window of the Java viewer displays the proteins from the alignment in red if they contain a eubPAP-specific motif (blue rows in A), found in the same location in 3D as eubPAP-specific residues in B.

www.rnajournal.org 901
the catalytic site (which most likely corresponds to an hprNS-I structure) was also found to be responsible for binding the three terminal nucleotides of the primer in yeast poly(A) polymerase (Zhelkovsky et al. 1998). The combined results suggest that the hprNS-I and hprNS-II motifs are responsible for binding of the 3′ end of the RNA substrate.

It is not clear whether or not the loop 3/4 structures in proteins that do not belong to the hprNtrs are also involved in binding of the RNA 3′ terminus. However, in the B. stearothermophilus CCAtr structure, a primer was modeled into the space between loop 3/4 and residue R194 on the opposite helix J, and it was proposed that the state of the primer strand could be communicated via R194 and helix J to helix G, where protein templating would then switch from C to A recognition (Li et al. 2002). Furthermore, it has been postulated that all DNA and RNA polymerases catalyze nucleotide addition by a unified mechanism, and that catalysis can only proceed when the two metal ions and the 3′ terminus of the primer strand are positioned in a unique orientation (Steitz et al. 1994). Thus, our prediction of the position of the 3′ terminus of the RNA substrate in CCAtrs and eubPAPs conforms to this general principle. Superposition of class I and class II Ntrs (Fig. 2A) illustrates this conservation of the catalytic domain structure. The loop 3/4 has a different angle in the two proteins to accomodate the either single- or double-stranded nucleic acid substrates. Docking experiments have suggested that the axes of tRNA or a tRNA minihelix do not enter the active site from the same direction in class I and class II CCAtrs (Okabe et al. 2003; Xiong et al. 2003), although this does not necessarily imply that the orientations of the 3′ terminal nucleotides differ. In eubPAPs and other nucleotidyl transferases that synthesize long polynucleotides, the 3′ terminal nucleotides of the primer are presumably linearly stacked on each other. Therefore, a primer binding structure in these enzymes may facilitate rapid translocation of the emerging polynucleotide chain. In contrast, in the case of CCAtrs the primer’s 3′ end is held in a fixed position to force scrunching of the growing CCA tail into a mold or cavity of the catalytic site (Shi et al. 1998). In summary, we assume that there are specific differences between the loop 3/4 structures in CCAtrs and in eubPAPs, and that the upstream sequence signature that we detected in eubPAPs corresponds to a specific loop 3/4 structure that is involved in primer binding and is present only in polyribonucleotide polymerases.

With the help of the Sequence Space program we identified a second eubPAP- or CCAtr-specific motif about 125 amino acids downstream of the eubPAP-specific motif, conforming to either the consensus ERxxxExxxhh or sRxxxExxxhh (“s” is a small residue, “h” is a hydrophobic and “x” is any amino acid), here termed “downstream motif.” Interestingly, the two variants of this motif could in most cases be assigned specifically to either the CCAtrs (ERxxxExxxhh) or to the eubPAP group (sRxxxExxxhh; Fig. 1A; Table 1). Arg194 of motif E, described in the B. stearothermophilus CCAtr structure (the R at the second position in the two motifs) was suggested to play a critical role in the templating specificity to generate CCA (Li et al. 2002). In addition, the CCAtr-specific residue Glu193 (E/251 in Fig. 1B) may also have a role in the nucleotide selection mechanism for either CC or A. Analysis of appropriate mutants could shed light on this question.

Some early branching eubacteria employ two separate enzymes to collaboratively add either CC or A to tRNAs (Tomita and Weiner 2001, 2002). These enzymes conform to a modified rule, whereby CC-adding proteins contain an ERhxxExxxhh motif and A-adding proteins carry an sRhxxExxxhh signature. Thus, a glutamic acid at the N-terminal end of the downstream motif is only present if the enzyme adds CC or CCA to tRNA precursors, whereas if only A has to be added as in eubPAPs or A-adding tRNA-specific Ntrs, Glu is replaced by a small inert residue. However, there are also exceptions to this rule: For example, the Thermotoga maritima nucleotidyl transferase contains the downstream motif PRxxxExxxhh (a signature for eubPAP or A-adding enzymes) but has recently been determined to be a CCAtr (Tomita and Weiner 2001). Therefore, the fact that the Glu in the downstream motif is replaced by a Pro in Thermotoga CCAtr could be an indication that this residue is not essential for catalysis. In addition, the α-proteobacterium Buchnera aphidicola and the early branching protist Giardia lamblia have Glu replaced by the functional analogs Asn or Gln in the downstream motif (Table 2), and also several protozoa carry small residues at this position. In summary, our rule strictly applies to the upstream eubPAP-specific signature, where no exceptions were found so far. Although the downstream motif is less well conserved, it is still useful to distinguish between CC- and A-adding enzymes.

**Phylogenetic assignment of eubacterial poly(A) polymerases and CCA transferases**

Integrating our results into a phylogenetic tree based on rRNA sequence analysis (Olsen et al. 1994; Woese 2002)
**TABLE 2.** Phylogenetic distribution of eubPAPs, CCAtrs, CC- and A-adding enzymes

<table>
<thead>
<tr>
<th>Organism</th>
<th>eubPAP Accession</th>
<th>CCAtr Accession</th>
<th>CC-adding Accession</th>
<th>A-adding Accession</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertebrata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemiptera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protozoa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinobacteria, γ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other prokaryotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanobacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanobacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prochlorococaceae A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Distribution of Ntr in a phylogenetic tree based on rRNA sequence analysis (distances not drawn to scale; Olsen et al. 1994; Woese 2002).**

Accession numbers in bold and framed indicate tested proteins (see references). Numbers in parentheses indicate the total number of organisms of the same genus that contain an identical type of Ntr.

Accession numbers of the type NP_xxxxxx are from completed genomes; accessions ZP_xxxxxx are from unfinished genomes.

www.rnajournal.org 903
which was adapted to include selected organisms according to the results of database searches, disclosed several new findings (Table 2). First, eubPAPs are only detected in the \(H.9252\), \(H.9253\), and \(H.9254\) subdivisions of proteobacteria and in some Chlamydiales and Spirochaetales but were not found in the \(H.9251\) and \(H.9255\)-proteobacteria subdivisions. This might indicate that eubPAPs evolved in an ancestor of these closely related bacteria and may have been lost in more derived lineages. With the exception of plants, no eubPAPs could be detected in archaea and eukaryotes. All Gram-positive bacteria and eubacteria that have diverged before the Gram-positives do not contain eubPAPs. The parasitic bacteria \(Mycoplasma\) and \(Ureaplasma\) (six genomes) remain the only organisms with no CCA transferase or eubPAP (Mushegian and Koonin 1996). These bacteria contain a minimal gene set, and all tRNA genes code for \(\delta\)-terminal CCA. Nevertheless, it is surprising that \(Mycoplasma\) does not need a CCA transferase, because many other organisms also code for CCA in their tRNA genes but do require a CCA-transferase as tRNA repair enzymes.

Early lineages of eubacteria, and in particular the cyanobacteria, carry two different enzymes for the synthesis or repair of the tRNAs' CCA ends (Tomita and Weiner 2001, 2002). The \(T.9251\) and \(T.9255\) lineages were among the first to acquire a single CCAtr. However, a few Gram-positives and even one of the proteobacteria kept a system with separate CC- and A-adding enzymes. Because eubPAPs are found in eubacteria and in plants, two separate branches of the phylogenetic tree (Table 2), we can consider a scenario for their origin where a first eubPAP was derived in a common ancestor of proteobacteria and Chlamydiales and Spirochaetales. Some bacterial species or entire groups, such as the \(H.9251\) and \(H.9255\)-proteobacteria, eventually lost eubPAP. If loss of eubPAP is “easy” for the cell, then many losses can be accepted on parsimony grounds, considering the fact that other enzymes, such as polynucleotide phosphorylase (PNP), take over the function of eubPAP (Yehudai-Resheff et al. 2001; Rott et al. 2003). In a second event, eubPAPs and CCA-adding enzymes were transmitted from eubacteria to an early eukaryote by endosymbiosis. Endosymbiosis of \(\alpha\)-proteobacteria is generally considered to be the origin of mitochondria, whereas chlo-
roplasts are thought to originate from the endosymbiosis of cyanobacteria (Burger et al. 2003).

Do the plant eubPAPs indeed descend from eubPAPs of a proteobacterial ancestor, or do they originate from plant CCAtrs? Sequence relationships between eubPAPs and CCAtrs from plants and eubPAPs of proteobacteria (as displayed in Fig. 1C) reveal that, although plant eubPAPs are close to a cluster of CCAtrs, they share a cluster with eubPAPs of \textit{Spirochaetales}, a lineage which has branched before the proteobacteria, and also with \textit{H. sapiens}-proteobacteria. It is therefore likely that plants inherited eubPAPs from eubacteria and that the enzymes were not reinvented by conversion of plant CCAtrs. Furthermore, CC- and A-adding enzymes were found in the same cluster as CCAtrs in the Sequence Space protein viewer window (data not shown).

We found that all known plant genomes contain at least one eubPAP and one CCAtr. For instance, we identified cDNA sequences for one CCAtr and four eubPAPs, each coded by different genes on nuclear chromosomes in the \textit{Arabidopsis thaliana} databases. Interestingly, the prediction programs for sorting signals Predotar (http://www.inra.fr/predotar/), TargetP (Emanuelsson et al. 2000), Mitoprot (Claros and Vincens 1996), and iPSORT (Bannai et al. 2002) revealed that all eubPAPs and CCAtrs tested and listed in Table 3 are predicted to contain either mitochondrial or chloroplast targeting sequences, an indication for transport to these organelles. In the yeast \textit{Saccharomyces cerevisiae}, protein products of a single CCAtr gene were found to be targeted to the nucleus, the cytosol, and to mitochondria (Chen et al. 1992), and mammalian CCAtrs were also found to be imported into mitochondria (Nagaie et al. 2001; Reichert et al. 2001). Interestingly, these enzymes are coded not in the mitochondrial or plastid genomes but rather in the nucleus. It has been reported that $\sim$18% of protein-coding genes in the \textit{Arabidopsis} nuclear genome are derived from cyanobacteria but that gene origin and compartmentation do not strictly correlate (Martin et al. 2002). It is an intriguing possibility that the eubPAPs of proteobacterial origin in \textit{Arabidopsis} (and possibly other plants) which are encoded in the nucleus, are targeted to the cytoplasm, to plastids, or to mitochondria.

In conclusion, the sequence signatures reported here are useful to assign functions to nucleotidyl transferase sequences present in existing sequence databases and those emerging from the rapidly growing number of new genome-sequencing projects.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Species</th>
<th>Protein</th>
<th>Predotar</th>
<th>TargetP</th>
<th>Mitoprot II</th>
<th>iPSORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_173680</td>
<td>\textit{A. thaliana}</td>
<td>CCAtr</td>
<td>---</td>
<td>Plastid</td>
<td>0.9688</td>
<td>Plastid</td>
</tr>
<tr>
<td>NP_197758</td>
<td>\textit{A. thaliana}</td>
<td>eubPAP</td>
<td>Mito</td>
<td>Mito</td>
<td>0.7029</td>
<td>Mito</td>
</tr>
<tr>
<td>NP_174130</td>
<td>\textit{A. thaliana}</td>
<td>eubPAP</td>
<td>Plastid</td>
<td>Mito</td>
<td>0.648</td>
<td>Plastid</td>
</tr>
<tr>
<td>NP_190452</td>
<td>\textit{A. thaliana}</td>
<td>eubPAP</td>
<td>Mito</td>
<td>Sig-Pep</td>
<td>0.7269</td>
<td>Sig-Pep</td>
</tr>
<tr>
<td>AAM22707</td>
<td>\textit{Oryza sativa}</td>
<td>eubPAP</td>
<td>Mito</td>
<td>Mito</td>
<td>0.946</td>
<td>Plastid</td>
</tr>
<tr>
<td>P21269</td>
<td>\textit{S. cerevisiae}</td>
<td>CCAtr</td>
<td>Mito</td>
<td>Mito</td>
<td>0.9561</td>
<td>Mito</td>
</tr>
<tr>
<td>Q96Q11</td>
<td>\textit{H. sapiens}</td>
<td>CCAtr</td>
<td>Mito</td>
<td>Mito</td>
<td>0.9991</td>
<td>Mito</td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p25500</td>
<td>\textit{Bos taurus}</td>
<td>nucPAP-\textalpha{}</td>
<td>---</td>
<td>---</td>
<td>0.0333</td>
<td>---</td>
</tr>
<tr>
<td>Q10569</td>
<td>\textit{Bos taurus}</td>
<td>CPSF-160</td>
<td>Mito</td>
<td>---</td>
<td>0.0391</td>
<td>---</td>
</tr>
<tr>
<td>P79101</td>
<td>\textit{Bos taurus}</td>
<td>CPSF-73</td>
<td>---</td>
<td>---</td>
<td>0.0721</td>
<td>---</td>
</tr>
</tbody>
</table>

Mito, mitochondrial; Sig-Pep, signal peptide. As controls for proteins that are not expected to contain target peptides, sequences of bovine nuclear poly(A) polymerase \textalpha{} (nucPAP-\textalpha{}) and two subunits of the cleavage and polyadenylation specificity factor CPSF (CPSF-73 and CPSF-160) were submitted.

**MATERIALS AND METHODS**

BLAST searches were conducted in the genome databases (http://www.ncbi.nlm.nih.gov/ sutils/genom_table.cgi) or other nonredundant or EST databases at the Swiss EMNet node (Falquet et al. 2003) at http://www.ch.embnet.org/software/BottomBLASTadvanced.html or at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). The Sequence Space program (Casari et al. 1995) was downloaded from http://industry.ebi.ac.uk/SeqSpace/. For Table 2, a phylogenetic tree describing archaea and bacteria (Olsen et al. 1994) was modified according to results from database searches and was extended to include the eucarya according to the universal phylogenetic tree determined by rRNA sequence analysis (Woese 2002).

**ACKNOWLEDGMENTS**

We thank Lorenza Bordoli for help with the Sequence Space program and for critically reading the manuscript, David Rand (Brown Univ.) for his invaluable expert advice, Stepanka Vana- cova for improving the manuscript, and an anonymous referee for many useful suggestions. This work has been supported by the University of Basel and the Swiss National Science Fund.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby...
marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received November 21, 2003; accepted February 26, 2004.

REFERENCES


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