Published online August 27, 2004

4646–4656 *Nucleic Acids Research, 2004, Vol. 32, No. 15*
doi:10.1093/nar/gkh804

Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development

Anders Aspegren, Andrea Hinas, Pontus Larsson[1], Anders Larsson[1] and Fredrik Söderbom*

Department of Molecular Biology, Biomedical Center, Swedish University of Agricultural Sciences, Box 590, S-75124 Uppsala, Sweden and [1]Institute of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, S-75124 Uppsala, Sweden

Received May 26, 2004; Revised and Accepted August 17, 2004 DDBJ/EMBL/GenBank accession nos[+]

ABSTRACT

The quest for non-coding RNAs (ncRNAs) in the last few years has revealed a surprisingly large number of small RNAs belonging to previously known as well as entirely novel classes. Computational and experimental approaches have uncovered new ncRNAs in all kingdoms of life. In this work, we used a shotgun cloning approach to construct full-length cDNA libraries of small RNAs from the eukaryotic model organism *Dictyostelium discoideum*. Interestingly, two entirely novel classes of RNAs were identified of which one is developmentally regulated. The RNAs within each class share conserved 5′- and 3′-termini that can potentially form stem structures. RNAs of both classes show predominantly cytoplasmic localization. In addition, based on conserved structure and/or sequence motifs, several of the identified ncRNAs could be divided into classes known from other organisms, e.g. 18 small nucleolar RNA candidates (17 box C/D, of which a few are developmentally regulated, and one box H/ACA). Two ncRNAs showed a high degree of similarity to the small nuclear U2 RNA and signal recognition particle RNA (SRP RNA), respectively. Furthermore, the majority of the regions upstream of the sequences encoding the isolated RNAs share conserved motifs that may constitute new promoter elements.

INTRODUCTION

Several experimental screens for RNAs in the size range of 50–500 nt have recently identified a large number of novel non-coding RNAs (ncRNAs) in model organisms as diverse in complexity as *Escherichia coli* and mouse (1–5). The RNAs that are frequently found in these screens are small nuclear RNAs (snRNAs) and SRP RNA (signal recognition particle RNA). One surprising observation was the great number of small nucleolar RNAs (snoRNAs) isolated from eukaryotic cells. snoRNAs are divided into two sub-classes based on sequence motifs, box C/D and box H/ACA that guide 2′-*O*-ribose methylation and pseudouridylation of RNAs, respectively. snoRNAs bind to their target RNA via antisense interactions of complementary sequences. Initially, the targets for snoRNA-mediated modifications appeared to be restricted to rRNA, but the target repertoire has been extended to include snRNAs and, most likely, transfer RNAs (tRNAs) [(6) and references therein]. The growing number of orphan snoRNAs, i.e. those that lack sequence complementarity to either of these classes of RNAs, has suggested additional targets, e.g. mRNAs [for reviews see (6,7)].

Very little is known about the prevalence of ncRNAs in the eukaryotic model organism *Dictyostelium discoideum*. This genetically tractable protist, also called social amoeba, inhabits the forest floor. Starvation initiates a developmental program during which single cells aggregate and develop as a multicellular organism. The final developmental structure is a stalk topped with a ball of spores (8,9). Evolutionarily, *Dictyostelium* belongs to the group of mycetozoans which branched out after plants but before metazoans and fungi (10,11). Prior to this study, only a few small ncRNAs, besides tRNAs and rRNAs, had been isolated from *Dictyostelium* (12–15). One of these, characterized as U3 snoRNA (16), was isolated from nuclear RNA.

Dictyostelium is of evolutionary interest with respect to its life cycle, unicellular growth but multicellular development, as well as its intermediate niche between animals, plants and lower eukaryotes. These features, and the fact that very little is known about ncRNAs in *Dictyostelium*, prompted us to isolate small RNAs from this organism in order to better understand their abundance and function from an evolutionary and developmental perspective. Full-length cDNA libraries were generated, representing small RNAs from developing cells.

*To whom correspondence should be addressed. Tel: +46 18 471 4901; Fax: +46 18 536971; Email: Fredrik.Soderbom@molbio.slu.se
Present address:
Anders Aspegren, AstraZeneca, R&D, Department of Molecular Pharmacology, Pepparedsleden 1, S-431 83 Mölndal, Sweden

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

[+]AJ698945–AJ698946, AJ703792–AJ703794, AJ699367–AJ699397

Novel small RNAs, snoRNA, snRNA and SRP RNAs were identified and analyzed with respect to the developmental expression, genomic location and putative promoter sequences. The predicted rRNA target for one of the isolated snoRNAs was shown to be methylated. Interestingly, we identified two entirely novel classes of RNA, to our knowledge not known from any other organism. These RNAs are mainly located in the cytoplasm. Furthermore, we showed that one class is developmentally regulated.

MATERIALS AND METHODS

New sequences

Isolated RNA sequences and their accession numbers are as follows: DdR-1 (*Dictyostelium discoideum* RNA) and DdR-2; AJ698945, AJ698946, DdR-3–DdR-5; AJ703792–AJ703794, and DdR-6 to DdR-36; AJ699367 to AJ699397.

Oligonucleotides

DNA oligonucleotides (Invitrogen) and RNA oligonucleotide (Dharmacon Research) used are listed in Supplementary Material (Table S2).

Growth and development

The *Dictyostelium* strain AX4 (17) was grown in HL5 medium and synchronously developed on nitrocellulose filters supported by buffer-saturated pads (18).

RNA preparation and cDNA library construction

Total RNA was extracted by the TRIzol method (Gibco-BRL) from 10^8 cells developed for 16 h and subsequently size fractionated on denaturing 8% polyacrylamide gels (7 M urea, 1× TBE buffer), eluted and poly(C) tailed (5,19). In order to clone full-length RNA of both primary transcripts and processed RNA, GeneRacer™ Kit (Invitrogen) was used essentially according to the manufacturer's recommendation but with modifications as described. Briefly, the C-tailed RNA was treated with tobacco acid pyrophosphatase (TAP) (Epicentre Technologies) before the provided RNA oligonucleotide was ligated to the 5′ end in the presence of 10% dimethyl sulfoxide (DMSO). CIP treatment was omitted. Reverse transcription was performed at 42 and 50°C for 30 min each, followed by inactivation of the reaction for 15 min at 70°C. A mixture of three different oligonucleotides (0.8 μM each) was used as primers for reverse transcription; Generacer Oligo dC3′A, Generacer Oligo dC3′C, and Generacer Oligo dC3′T. The oligonucleotides were identical except for the 3′ most nucleotide, which was either A, C or T, in order to define the very 3′ end of the RNA providing that the last nucleotide was not a C (Table S2). For PCR amplification, 2 μl of the cDNA reactions were used as template together with 5′ and 3′ primers provided by the kit. PCR products (150–500 nt library) were cleaved with ClaI (Amersham Biosciences) to reduce the number of 5.8S rDNA-containing inserts. PCR products were separated on 1% agarose (Gibco BRL) gels, and the smear of DNA was excised and gel-eluted (S.N.A.P.™ Columns, Invitrogen), ethanol precipitated and subsequently cloned into pCR-2.1 TOPO vector (Invitrogen).

Colony hybridization and PCR

To reduce the number of cDNAs representing rRNA, e.g. 5S rRNA (50–150 nt library) and 5.8S rRNA (150–500 nt library), colony hybridizations using ^{32}P-end-labeled oligonucleotides were performed. Colonies were transferred to Hybond-N nylon membranes (Amersham Biosciences) and treated according to the manufacturer's recommendations. DNA was immobilized by ultraviolet crosslinking using a Bio-Rad crosslinker. The membranes were analyzed as described under 'northern analysis' and colonies harboring cDNAs of rRNA were excluded from further analysis.

Colony PCR with GeneRacer 5′ Primer and GeneRacer 3′ Primer, provided in the GeneRacer Kit, was performed on the 150–500 nt library. PCR products generated from cDNA of RNA sized 150 nt or longer were treated with ClaI and PstI (Amersham Biosciences) to exclude clones that still harbored cDNA of rRNA, e.g. 5.8S and 17S fragments, respectively. (This treatment will also digest cDNA from RNA other than rRNA that harbors restriction sites for ClaI and PstI.) Plasmids were prepared (Qiagen Miniprep Kit, Qiagen) from the remaining colonies (that passed these tests) and the cDNA inserts were sequenced on an ABI 377 DNA Sequencer (Applied Biosystems). About 10 cDNAs were sequenced between each screening/selection step to validate the procedures.

Northern analysis

DNase-treated RNA samples (20 μg) were denatured for 5 min at 95°C in loading buffer containing 46% formamide, 8.5 mM EDTA, 0.013% bromophenol blue and 0.013% xylene cyanol, separated on denaturing 8% polyacrylamide gels (7 M urea, 1× TBE) and transferred to Hybond-N+ (Amersham Biosciences) nylon membranes by electroblotting. The RNA was immobilized as described above, followed by hybridization with specific ^{32}P-end-labeled primers. Hybridization signals were analyzed by a Phosphor Imager (Molecular Dynamics). To ensure equal loading, aliquots of the RNA samples were analyzed by agarose gel electrophoresis followed by ethidium bromide staining before PAGE. RNAs that demonstrated ⩾2-fold difference in expression, after normalization to 5.8S rRNA hybridization signals, were regarded as developmentally regulated. Developmental regulation was confirmed by repeating the northern blot analysis at least once.

Cellular localization of Class I and Class II RNA

Nuclei were prepared from growing cells as described by Crowley *et al*. (20) and nuclear RNA was prepared by the TRIzol method. Nuclear and total RNA, 3.7 and 20 μg, respectctively (to give equal U2 hybridization signals), was separated on a denaturing 10% polyacrylamide gel (7 M urea, 1× TBE) and subsequently transferred to nylon membranes, immobilized and hybridized as described above. Hybridization oligo AH-tRNA-Arg was used to detect $tRNA^{Arg}$. End-labeled MspI-digested pUC19 DNA (MBI Fermentas) and Decade Marker (Ambion) were used as size markers.

Computational analysis

The genomic locations of the RNA encoding regions were determined by BLASTN searches (http://www.dictybase.org)

and (http://www.sanger.ac.uk/Projects/D_discoideum/) as of April 12, 2004.

The *Dictyostelium* rRNA genes (21) and the U2 snRNA homolog (this study) were used for target search for the isolated box C/D snoRNA candidates. Potential methylation targets for the candidate snoRNAs were identified by searching for regions complementary to the guide sequence upstream of presumed boxes D and D′. Duplexes were required to consist of a minimum of nine base pairs. The number of GC base pairs, mismatches, and bulges were assessed in accordance with what was established by Cavaille *et al*. (22). The snoRNA sequence motifs were identified by the established consensus sequences (23) (5′-CUGA-3′ for box D and 5′-UGAUGA-3′ for box C). The box D′ and box C′ motifs were identified by allowing one and two deviations from the consensus, respectively. The predicted 2′-*O*-ribose methylated nucleotide in the target sequence, located in the predicted duplex 5 nt upstream from box D or box D′, was subsequently compared to conserved methylation sites in other eukaryotes, i.e. human, *Arabidopsis*, *Caenorhabditis elegans* and yeast [see Figure S2 and (24–28)]. Secondary structure predictions were performed using RNA mfold version 2.3, temperature set to 22°C (29,30) and RnaViz2 was used for the representation of RNA structures (31).

Determination of methylation sites

rRNA 2′-*O*-ribose methylation was determined by primer extension analysis at low dNTP concentration essentially as described by Lowe and Eddy (28). For each reaction, 0.3 pmol of ^{32}P-end-labeled primer RT17SA was annealed to 0.8 μg of total RNA or 10 ng of the corresponding *in vitro*-transcribed RNA at 57°C for 4 min after an initial denaturation at 95°C for 1 min. Primer extension reactions were carried out in the presence of 10% DMSO, 1 U/μl of AMV reverse transcriptase (USB) and 1 mM or 0.004 mM dNTP at 37°C for 30 min. The reaction was terminated by the addition of 0.6% SDS and the RNA hydrolyzed in 0.3 M KOH for 3 min at 95°C followed by incubation at 37°C for 16 h. After ethanol precipitation, primer extension products were analyzed on a 10% polyacrylamide gel (7 M urea, 1× TBE). Sequencing reactions were carried out on *in vitro*-transcribed RNA using 6.5 U/μl of Superscript II reverse transcriptase at 42°C for 30 min. *In vitro*-transcribed RNA was generated through the MEGAscript system (Ambion) on a template obtained by PCR using primers T717SA and 17SA.

5′, 3′ RACE and RT–PCR

Determination of the U2 snRNA homolog termini was carried out essentially as described by Argaman *et al*. (32) but with a few modifications [see also Bensing *et al*. (33)]. Briefly, 5′ RACE was performed on 15 μg of total RNA from growing cells (DNase treated 2×), incubated with TAP to convert the 5′ triphosphate of primary transcripts to monophosphates. The RNA samples were incubated with GeneRacer RNA oligo, T4 RNA ligase and subsequently PCR-amplified using oligonucleotides GeneRacer 5′Primer and the gene-specific primer U2-3′RT. Cycling conditions were 95°C for 8 min; 5 cycles of 95°C for 30 s, 62–57°C for 40 s, 72°C for 40 s; 35 cycles of 95°C for 30 s, 57°C for 40 s, 72°C for 40 s and 72°C for 10 min. Control samples were incubated under identical conditions but with the exclusion of TAP. 3′ RACE was performed as described previously (32) but using the gene-specific primer U2-5′ REVERSE in PCR amplification as described above. The PCR products were extracted from agarose gels, cloned and sequenced.

The cluster analysis was performed on total RNA (DNase treated 2×) isolated from growing cells. The RNA samples (2 μg) were incubated with 20 pmol of primers (DDSNO1, DDsno2, DDsno3 and DDR-5 clust) followed by reverse transcription reactions (see above). Control reactions were incubated under identical conditions where the RNA, the gene-specific primer or reverse transcriptase had been excluded. PCR was carried out using 2 μl of the reverse transcription reactions and 20 pmol of primers specific for the snoRNA genes within each predicted cluster, i.e. cluster DdR-9 and DdR-1, primers DDSNO1 plus Ddsno9clust; cluster DdR-13 and DdR-2, primers DDsno2 plus Ddsno13clust; cluster DdR-18 and DdR-4, primers DDsno3 plus Ddsno18clust; and cluster DdR-16 and DdR-5, primers DdR5clust plus DdR16clust. The reaction conditions were as described above. Cycling conditions were: 95°C for 9 min; 5 cycles of 95°C for 1 min, 60–55°C for 40 s, 72°C for 1 min; 30 cycles of 95°C for 1 min, 55°C for 40 s, 72°C for 1 min and 72°C for 10 min.

RESULTS

Construction and analysis of cDNA libraries

In order to identify small RNAs present during *Dictyostelium* development, cDNA libraries representing full-length RNAs were constructed from cells developed for 16 h (slug stage). At this time of development, cells have differentiated into two main cell types, pre-spore and pre-stalk cells, and sorted out to defined regions (8). Total RNA was isolated and size-fractionated (50–150 and 150–500 nt, respectively), followed by the addition of a C-tail to the 3′ ends using poly(A) polymerase, basically according to the method developed by Hüttenhofer *et al*. (5) (see also Materials and Methods). To ensure that the cDNA libraries represented true 5′ ends of both primary and processed RNAs, RNA oligonucleotides were ligated to the 5′ ends of the C-tailed RNA after treatment with TAP. The RNA was reversely transcribed, PCR amplified and subsequently cloned. The cDNA libraries constructed from 50–150 and 150–500 nt RNA isolated from cells developed for 16 h are denoted as 1650 and 16150 cDNA libraries, respectively.

A total number of 635 colonies from the 1650 cDNA libraries were analyzed (Figure S1). Inserts represented in the cDNA libraries were either derived from rRNA or were novel RNAs. Of the cDNA clones (34 unique), 132 contained new sequences that could not be assigned to any previously known RNA in *Dictyostelium* (Figure S1 and Table S1). Interestingly, two novel classes of abundant RNAs were identified, constituting 14% (16 unique sequences) of the total cDNA clones. About 7% (18 unique sequences) of the total cDNA clones were classified as putative snoRNAs based on conserved structure and/or sequence motifs.

From the 16150 cDNA library, 350 colonies were randomly selected and seven of these contained new sequences (two unique) classified as SRP RNA and U2 snRNA homologs based on structure and sequence motifs.

Taken together, 36 novel and unique RNA sequences were identified that could be divided into five different groups. The expression, sizes and developmental regulation were analyzed by northern blot analysis of total RNA isolated from cells in different stages of development. Only RNAs that showed ⩾2-fold difference in expression, determined by internal standard (5.8S rRNA), were considered to be developmentally regulated.

The majority of the inserts represented in the cDNA libraries were derived from full-length RNA since the length of the inserts in most cases is in concurrence with their sizes as estimated by northern blot analysis. Furthermore, the exact identity of the termini of U2 snRNA was determined by RACE experiments and were identical to the ends of the U2 snRNA represented in the cDNA library (see below).

The genomic location of the identified RNA genes was analyzed by BLASTN search of the recently released *Dictyostelium* genome sequence (http://www.dictybase.org). The great majority of the novel RNAs was encoded from regions located between predicted genes (see Supplementary Table S1). Interestingly, the chromosomal regions encoding the isolated RNAs have almost exclusively a higher G/C content than the flanking DNA sequences (Table S1), similar to the higher G/C content of open reading frames of protein-encoding genes when compared to that of their flanking regions (11).

Two novel classes of small RNAs—developmental expression and cellular localization

The most abundant new RNAs identified in the screen were divided into two separate classes. RNAs of similar sequences are, to our knowledge, not known from any other organism; BLASTN searches failed to reveal any homologs. Although functions and putative associated proteins are yet unknown, the strong sequence and the predicted structural conservation of the RNAs within each class may indicate a related biological role (see below). Class I constituted ∼12% of the total RNA represented in the 1650 cDNA library (Figure S1). Fourteen unique sequences were isolated, DdR-21 to DdR-34, predicted to be encoded from 17 loci (Figure 1A and Table S1). Interestingly, the majority of these RNAs (55–65 nt long) share 5′ and 3′ sequence elements, 16 and 8 nt, respectively. Parts of the conserved sequences are complementary, predicted to form a stem structure. Only two of the sequences (DdR-31 and DdR-34) differed from these consensus sequences and showed single nucleotide substitutions in the conserved 5′ motifs (Figures 1A and 3A). The DdR-34 cDNA most likely represents a 3′ end truncated version of a Class I RNA since the conserved 5′ end is present in the cDNA. Furthermore, the conserved 3′ sequence element is situated in the genome sequence just downstream from the region encoding DdR-34. In addition, the A to G substitution in

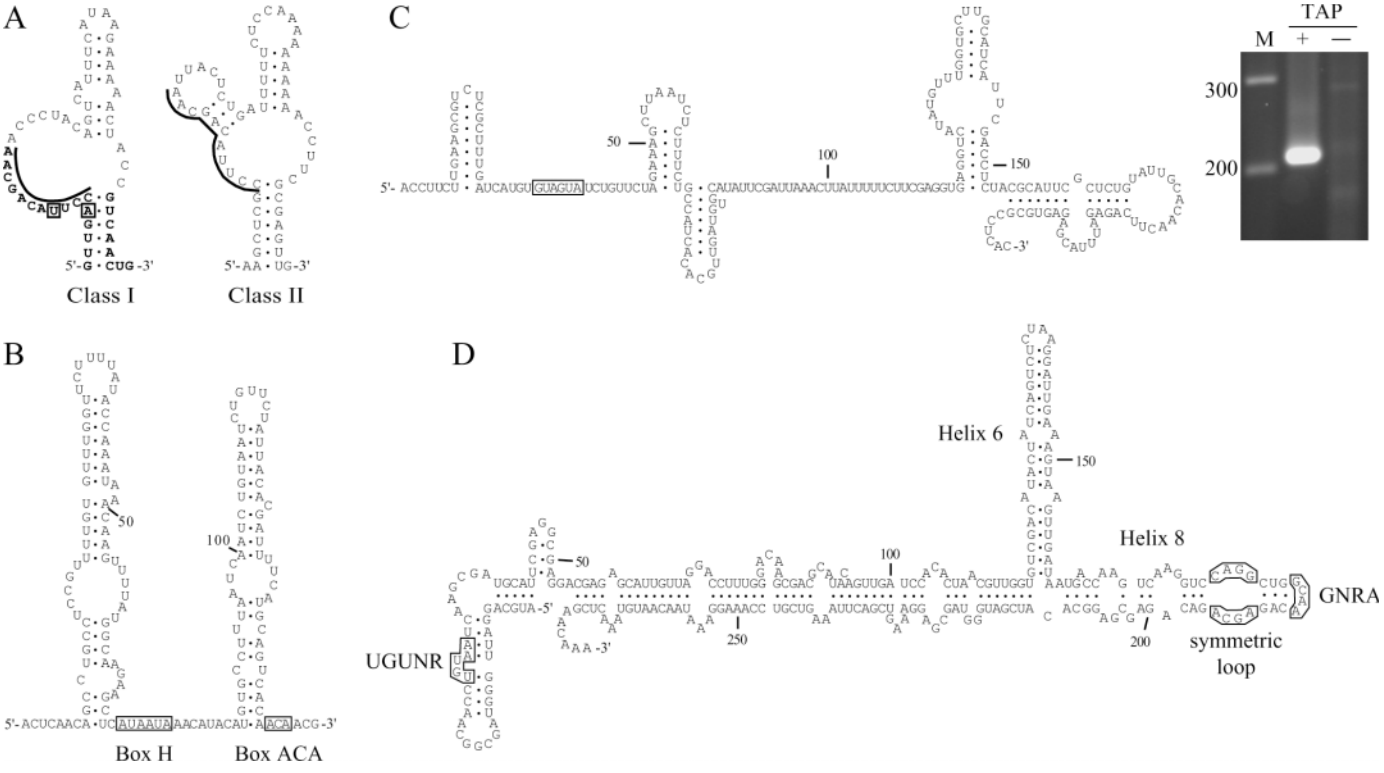

Figure 1. Structures of *Dictyostelium* ncRNAs as predicted by Mfold version 2.3 at 22°C. (**A**) Sequence and structure comparison of Class I (DdR-21) and Class II (DdR-35) RNAs. Sequences shared by the Class I RNAs are shown in boldface. Boxed nucleotides represent deviations from the consensus sequence (DdR-31: A to G substitution, DdR-34: U to C substitution). The solid line represents the sequence motif present in both Class I and Class II RNAs. (**B**) Box H/ACA snoRNA DdR-18. Boxed nucleotides represent conserved boxes H and ACA. (**C**) Left, predicted structure of the *Dictyostelium* U2 homolog DdR-19, as derived from the human U2 structure (37). The boxed sequence represents nucleotides complementary to the pre-mRNA splicing branch site. Right, RACE mapping of the U2 5′ end with and without prior treatment with TAP, respectively, indicating a non-processed 5′ triphosphate. (**D**) Predicted structure of SRP RNA DdR-20. Conserved motifs UGUNR (A), the symmetric loop (B) and GNRA (C) are indicated, as well as the conserved helices 6 and 8.

the conserved sequence of DdR-31 is predicted to maintain the base pairing (Figure 1A).

In order to investigate the expression profile of the novel RNAs throughout development, total RNA was isolated from growing (single) cells, cells developed for 16 h (slugs) and terminally differentiated cells (fruiting bodies) developed for 24 h, followed by northern blot analysis. Interestingly, the expression of Class I RNA is down-regulated during development (Figure 2). The expression profile was analyzed using probes that specifically recognize DdR-21 or recognize all Class I RNA molecules by hybridizing to the conserved 5′ sequence element (Figure 3A). Moreover, BLASTN search of the *Dictyostelium* genome revealed 24 additional regions homologous to the sequences encoding Class I RNA, e.g. they all share the conserved 5′ and 3′ sequence elements and are of similar length (data not shown).

The second class of novel small RNAs, Class II, identified in the 1650 cDNA library was present as two unique sequences, DdR-35 and DdR-36, 59 and 60 nt, respectively (Table S1). Northern blot analysis demonstrated expression in all the developmental stages tested (Figure 2). Due to their high degree of sequence similarity (92%) both RNAs were recognized by the same probe. The 5′ and 3′ ends have the potential to form a stem structure of similar length as is the case for the Class I RNAs (Figure 1A). Interestingly, Class I and Class II RNAs share a conserved sequence motif, located at similar positions in the predicted secondary structures. Alternative folding of the different RNAs still positions the conserved sequence in 'open', less-structured regions (data not shown).

In order to investigate the cellular localization of the two novel classes, RNA was isolated from nuclei and from total cell extract. As controls for the quality of the nuclear preparations, we probed for the U2 snRNA homolog identified in this study (nuclear; see below) and tRNA (cytoplasmic), respectively. Loading of RNA was chosen so that U2 signals were equal, thus facilitating an evaluation of the relative intensities of the other RNA bands. Quantitation of band intensities

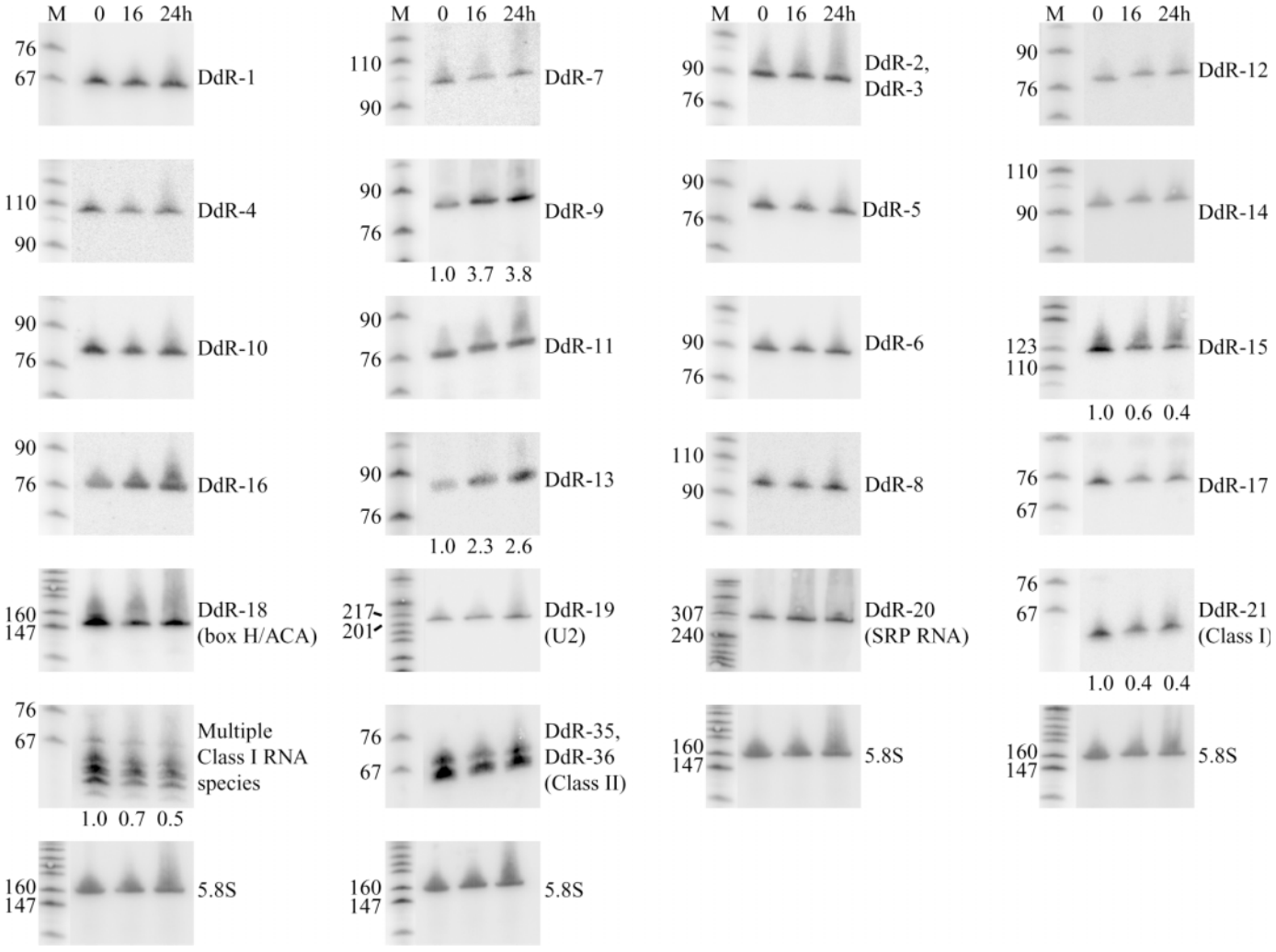

Figure 2. Expression patterns of small RNAs during *Dictyostelium* development as shown by northern blot analysis. Time points represent hours of development where 0 h designates growing cells. Size markers (M) are shown for reference. Developmental regulation of Class I RNA expression was shown both by using a probe complementary to the conserved 5′ end (multiple Class I RNA species) and by a probe detecting DdR-21 specifically. The increase/decrease in expression of developmentally regulated RNAs (after normalization to 5.8S rRNA) is shown below the panels.

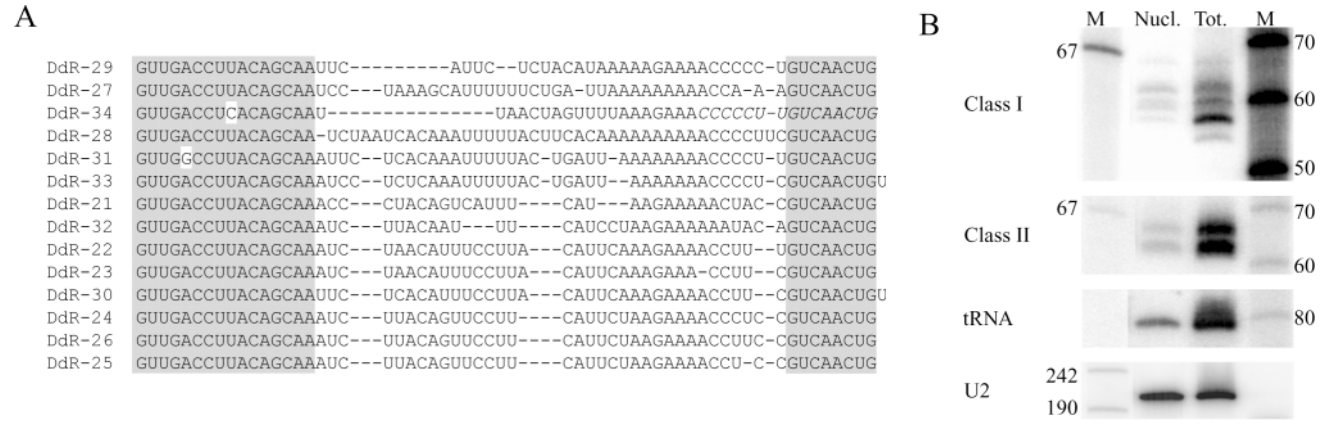

Figure 3. (**A**) Alignment of isolated Class I RNAs. Shaded nucleotides represent conserved 5′ and 3′ sequences. The 3′ end of DdR-34 (shown in italics) was obtained from dictyBase. Nucleotides deviating from the 5′ consensus sequence are presented on white background. (**B**) Class I and Class II RNAs are primarily cytoplasmic, as shown by northern blot analysis of nuclear and total RNA. Each panel shows the same membrane successively probed for the different RNAs. RNA amounts loaded from the two preparations were adjusted to give approximately equal U2 signals, to facilitate comparison. The probe used for Class I hybridizes to the conserved 5′ sequence, and the probe for Class II recognizes both DdR-35 and DdR-36 (see Results). Signals were quantified after normalization to the nuclear control U2 snRNA. The estimated fractions located in the cytoplasm were tRNA, 91%; Class I, 93%; Class II, 95%. Size markers (M), MspI-digested pUC19 DNA (left) and RNA Decade marker (right).

of the northern blot in Figure 3B showed that the majority of U2 RNA is located in the nucleus while ∼90% of the $tRNA^{Arg}$ fraction is found in the cytoplasm. Interestingly, both Class I and Class II RNAs are predominantly localized in the cytoplasm, even to a higher degree than $tRNA^{Arg}$.

Novel snoRNA candidates—developmental expression, polycistronic organization and targets

Box C/D snoRNAs. Based on conserved sequence and structure motifs, 18 unique snoRNA candidates were identified in the screen. Several of these were represented more than once in the cDNA library (Figure S1 and Table S1). The majority (17/18) belongs to the box C/D family and contains the hallmark motifs, box C and D elements, located at their 5′ and 3′ ends, respectively [Figure 4 and Bachellerie *et al*. (6)]. Furthermore, the less conserved C′ and D′ motifs were present in most of the snoRNA sequences. The snoRNA candidates lack the inverted repeats positioned at 5′ and 3′ of box C and D, which are frequently present in snoRNAs from other organisms where they form a typical stem structure. However, this stem is not ubiquitously present in snoRNAs identified in other eukaryotes and can be compensated for by external or internal stem structures (34). The majority of the snoRNAs isolated in this study are flanked by inverted repeat sequences that might form stem structures, thereby compensating for the lack of a canonical terminal stem (data not shown). The absence of complementary nucleotides is not an artifact caused by the construction of the cDNA library since the majority of the sequenced cDNA seems to represent full-length RNA (see northern blot analysis in Figure 2).

Northern blot analysis revealed that at least two of the RNAs, DdR-9 and DdR-13, are up-regulated during development, while DdR-15 is down-regulated (Figure 2). Furthermore, the northern blot analysis confirmed that all the candidate snoRNAs represented in the library were expressed, although the sequences of DdR-2 and DdR-3 (95% identical) could not be distinguished as separate signals (Figures 2 and 4). For one of the isolated RNAs, DdR-14, the length of the cDNA insert is shorter than expected when compared to the hybridization signal obtained by northern blot. The smaller cDNA most likely represents a 5′ end truncated version of a putative C/D box snoRNA since a D box is present at the 3′ end of the insert. Furthermore, a C box motif is situated in the genome sequence just upstream of the region encoding DdR-14 (Figures 2 and 4).

In an attempt to locate targets for the identified box C/D snoRNA candidates, *Dictyostelium* rRNA and U2 snRNA (see below) were searched for complementarities to sequences upstream of box D and box D′ (see Materials and Methods). This interaction guides 2′-*O*-ribose methylation of the nucleotide paired to the fifth nucleotide upstream of these motifs (35,36). Possible targets for most of the box C/D snoRNA candidates were identified. Most of these correspond to methylation sites conserved in other organisms [see Figure S2 and (24–28)]. Interestingly, DdR-9 and DdR-13, which are both up-regulated during development, are predicted to target methylation of the same conserved nucleotide.

An established procedure was used to analyze 2′-*O*-ribose methylation of a predicted target, conserved in *Saccharomyces* (28) and human (24), for one of the snoRNA candidates. Briefly, primer extension analysis performed in low dNTP concentration induces termination at methylated nucleotides [see Materials and Methods and (28)]. Methylation of one of the nucleotides predicted to be targeted by DdR-1 was verified (Figure 5). The predicted extension product was accumulated in low dNTP concentration when total RNA extracted from cells was used as template, while this product was absent when the template was unmethylated *in vitro*-transcribed RNA. This strongly indicates that the predicted target for DdR-1 is 2′-*O*-ribose methylated. The 1 nt larger product is most likely a termination product caused by structural features or fragmented template RNA, given that this product is also visible when *in vitro*-transcribed RNA is used as template.

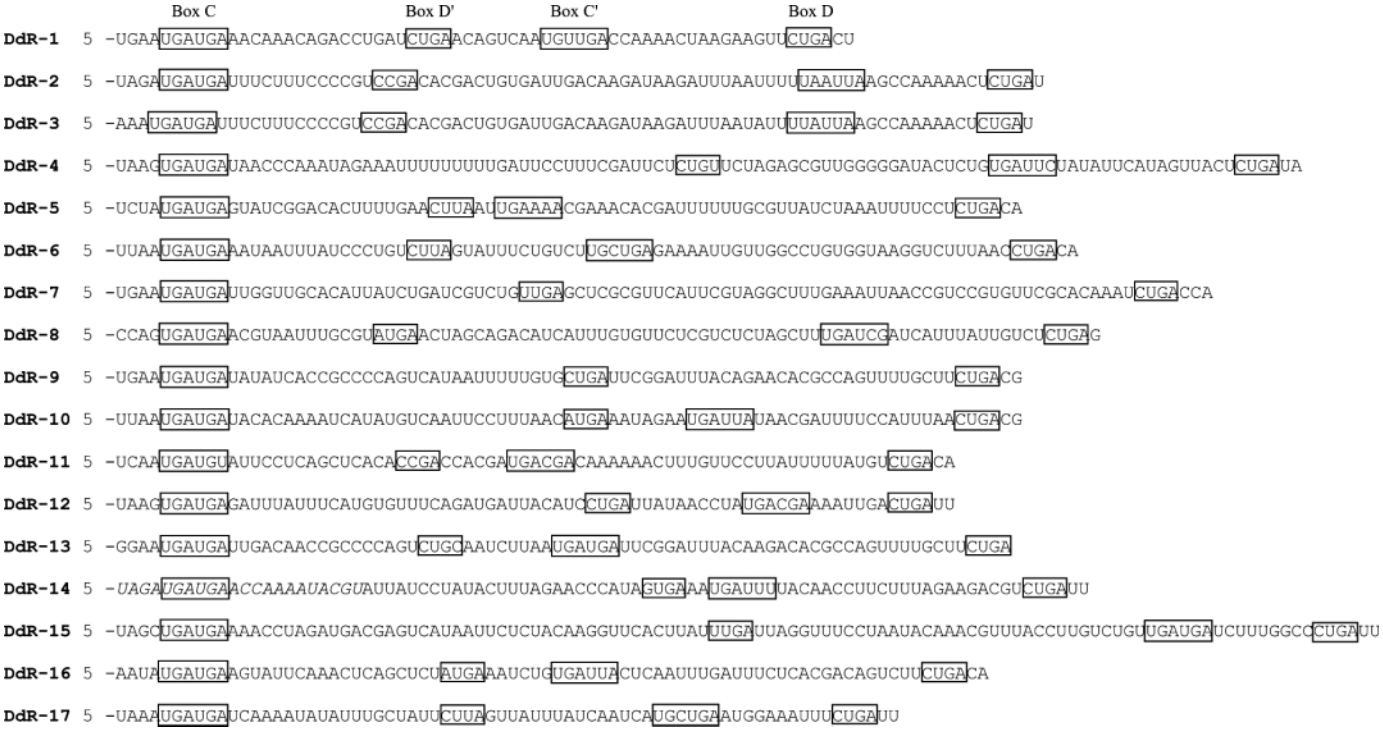

Figure 4. Sequences and conserved elements of *Dictyostelium* box C/D snoRNAs DdR-1 to DdR-17. Conserved boxes C, D′, C′ and D are shown as boxed nucleotides (Materials and Methods). The 5′ end of DdR-14 (shown in italics) was obtained from dictyBase.

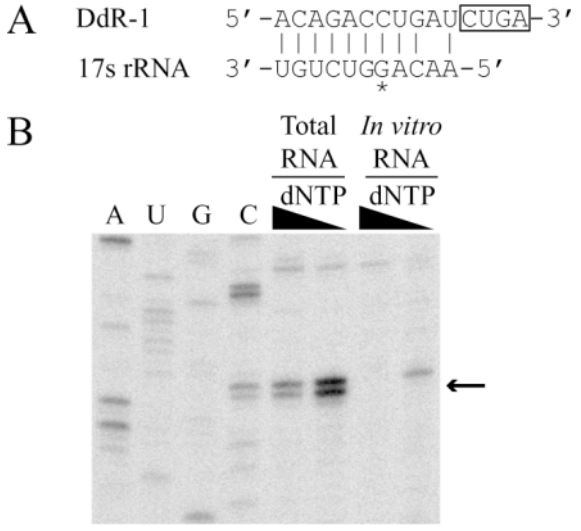

Figure 5. Confirmation of predicted target for box C/D snoRNA DdR-1. (**A**) Predicted base pair interaction between DdR-1 and its 17S rRNA target sequence. Boxed nucleotides represent the predicted D′ box and the asterisk indicates the methylated nucleotide Gm1501. (**B**) Experimental verification of predicted 17S rRNA 2′-*O*-ribose methylation site by primer extension analysis in the presence of high- and low concentration of dNTP, on total and *in vitro*-transcribed RNA, respectively. The reverse transcription termination caused by rRNA methylation is indicated by an arrow. Sequencing reactions on *in vitro*-transcribed RNA were run in parallel.

The genomic location of the putative box C/D snoRNA genes was analyzed. The great majority was predicted to be located in intergenic regions (Table S1). The genes for DdR-5 and DdR-16 were present on a contig not included in the annotated genome sequences. Furthermore, several of the snoRNA-like genes are located in clusters, i.e. DdR-1 and DdR-9, DdR-5 and DdR-16, DdR-2 and DdR-13, and DdR-4 and DdR-18, separated by ∼140–590 nt (see Table S1). The expression of polycistronic RNAs from three of the clusters was confirmed by RT–PCR (Figure 6). Whether DdR-5 and DdR-16 are processed from one transcript is doubtful since RT–PCR failed to yield a signal (data not shown). In addition, ∼200 nt separate DdR-11 from an almost identical genomic copy identified by BLASTN search of the *Dictyostelium* genome sequence (data not shown).

Box H/ACA snoRNA. One of the isolated RNAs, DdR-18, shares the hallmark sequence and structure motifs of box H/ACA snoRNAs (Figure 1B). Two predicted stem–loop structures are separated by a hinge that harbors the conserved box H motif (ANANNA). The conserved ACA motif is present at the expected location in the 3′ end of the RNA. No obvious pseudouridylation targets were found in rRNA or U2 snRNA (Dr A. Poole, personal communication). DdR-18 and the C/D box snoRNA DdR-4 genes are clustered and transcribed as a polycistronic pre-snoRNA transcript (Figure 6 and Table S1).

U2 snRNA homolog

An snRNA homolog, U2, was identified in the screen. This RNA, DdR-19, shares motifs commonly found in eukaryotic U2 snRNAs (37) as indicated in Figure 1C, is expressed throughout development (Figure 2), and is located in the nucleus (Figure 3A). The reported branch site sequences in *Dictyostelium* introns (38) are predicted to interact with the conserved sequence motif (Figure 1C). RACE experiments were carried out to identify the 5′ and 3′ ends and to determine

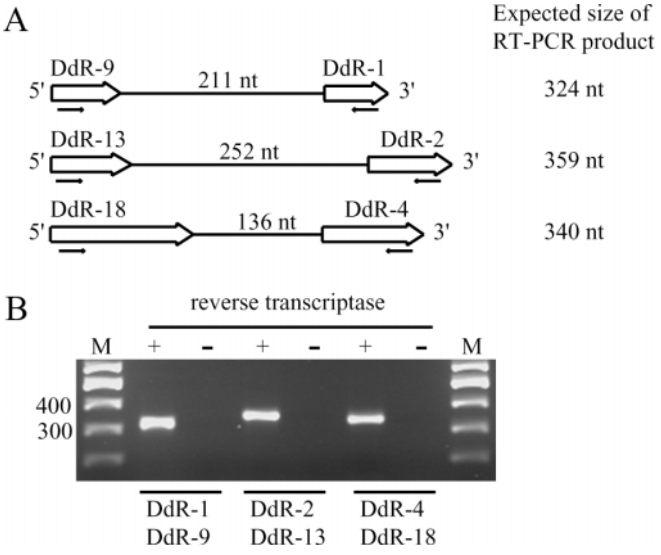

Figure 6. Cluster analysis. (**A**) Schematic drawing of snoRNA pairs located in clusters. Open arrows represent snoRNAs, and small solid arrows indicate location of primers used for the analysis. The distance between each snoRNA is indicated. The clusters are drawn to scale. (**B**) RT–PCR products run on 2% agarose gel. The PCR signals were dependent on previous treatment of the RNA with reverse transcriptase (+). No signals were detected in the DNA contamination controls (−).

whether DdR-19 is a primary transcript (see Materials and Methods). The results showed that the 5′ and 3′ ends of the U2 homolog are identical to the ones observed in the cDNAs, and hence the full-length RNA was cloned. Moreover, the analysis also proved that the RNA is a primary transcript (Figure 1C) and thus, transcribed from its own promoter (see below).

Three genes on chromosome 1 are predicted to encode the U2 homolog (one of the predicted genes has one nucleotide mismatch as compared to the isolated RNA). Two identical sequence motifs, that could constitute promoter sequences, are located upstream of the predicted RNA genes (see below). In addition, BLASTN search of the *Dictyostelium* genome revealed four additional chromosomal regions that showed similarity to the U2 like cDNA. These might represent pseudogenes, since putative promoter elements are missing or degenerated (data not shown).

SRP RNA homolog

One of the isolated RNAs, DdR-20, shows homology to SRP RNA and is expressed at all the developmental stages tested (Figure 2). The highly conserved helices 6 and 8 are present as well as conserved sequence motifs, e.g. the GNRA loop in helix 8 [Figure 1D and reviewed in (39)]. The Alu domain contains the consensus UGUNR motif but does not seem to fold into the two hairpin structures commonly present in SRP RNAs from archaea and eukaryotes. Interestingly, high variability of Alu domains has recently been reported for SRP RNAs in protozoa and fungi (40).

An additional copy with ∼88% sequence identity to the DdR-20 encoding region is present in the genome and a putative promoter sequence is located upstream from the predicted RNA coding sequence (see below). Whether this copy is expressed or not is unknown.

Putative promoter sequences

The genomic sequences upstream of the RNA-encoding regions were analyzed for sequence similarities. Since the U2 homolog, DdR-19, is a primary transcript (see above), promoter elements should be located upstream of the determined start site of transcription. The 5′ genomic sequences (100 nt) flanking the three U2 snRNA regions were aligned (Figure 7). Two motifs are present in all three regions, positioned approximately at the same distance from the verified/predicted start of transcription. The B element is located almost immediately upstream of the putative RNA-encoding sequences, while element A is centered at −58 or −60. The sequence of the A element and its distance from the start of transcription resembles snRNA promoter elements in *Arabidopsis* (41).

Interestingly, a sequence almost identical to element A is positioned in front of the regions encoding SRP RNA (DdR-20) and its similar copy. Furthermore, the distance between the predicted start of the SRP RNA candidate genes and the A elements are 60 nt, as for the U2 candidates (Figure 7).

The genomic region encoding Class II RNAs, DdR-35 and 36, obey to the same rules. The A motif is identical to the one in front of the candidate SRP RNA genes and positioned at 61 nt from the putative transcription start.

Analysis of the 5′ flanking regions of Class I RNA-encoding sequences revealed two consensus motifs centered at about −76 and −15 nt, respectively. The −76 nt sequence is highly similar to the A-element described above (Figure 7). Interestingly, these two elements are positioned at −93 and −29, respectively, from the chromosomal DNA encoding DdR-21 and DdR-32; hence, basically maintaining the distance between the motifs. This may indicate that DdR-21 and DdR-32 are transcribed as longer precursors, which subsequently are processed to their final size as observed in the cDNA library and by northern blot analysis.

DISCUSSION

The intriguing life cycle of *Dictyostelium* and its interesting evolutionary position have rendered this organism a unique place in research of fundamental aspects of different biological processes. Although *Dictyostelium* is widely used as a model organism, next to nothing is known about the presence and hence the function of endogenous ncRNAs (apart from tRNA and rRNA) in this organism. However, the first experimentally characterized antisense RNA in eukaryotes was discovered in *Dictyostelium* (42) and a few small ncRNAs have been isolated (12–16). In order to investigate the presence of small RNAs in developing *Dictyostelium*, we used shotgun cloning on size-selected RNAs (50–500 nt) to create full-length cDNA libraries. Similar approaches, termed experimental RNomics, have successfully been applied on a number of different organisms (2–5,43,44). The method presented here differs from the experimental RNomics approach, in that full-length cDNAs are generated by ligating an RNA oligo to RNA that has been subjected to TAP treatment before reverse transcription and cloning (see Materials and Methods). Thus, cDNA

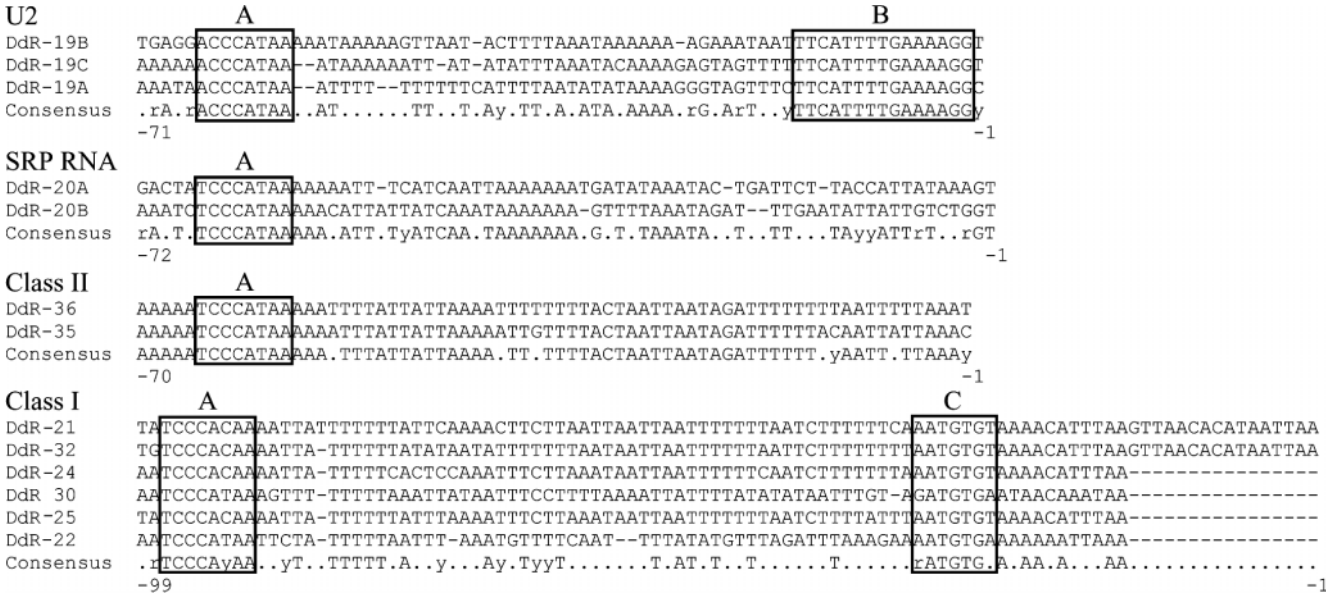

Figure 7. Putative *Dictyostelium* promoter elements. Alignments of genomic sequences present upstream from regions encoding U2 snRNA, SRP RNA, Class II RNAs and Class I RNAs (a few representative sequences shown). Putative promoter elements A, B and C are indicated. Numbers represent distance from the 5′ end of the coding regions.

representing true 5′ ends (and 3′ ends) from primary as well as processed transcripts are obtained.

In the cDNA libraries, we identified 36 new and unique sequences. Based on sequence and/or structure homology, 20 of these could be classified as box C/D and box H/ACA snoRNAs, SRP RNA and U2 snRNA. Interestingly, the remainder of the isolated RNAs belonged to two novel classes of RNAs. The absence of expected tRNA clones is most likely due to highly structured 5′ and 3′ ends which will influence the ligation of RNA-oligos and C-tailing, respectively. Most of the investigated new RNAs represented in the cDNA library were full-length clones as estimated by northern blot analysis.

The most unexpected result was the identification of two novel classes of RNA. These short RNAs (55–65 nt) appear to be abundant and share an 11 nt sequence motif located at similar positions in all the RNAs. Furthermore, within each class, sequences at the 5′ and 3′ ends are conserved. Although these nucleotide sequences differ between the different classes, they are predicted to form stem structures of similar length located at corresponding positions in both classes of RNAs. Interestingly, Class I RNAs are down-regulated during development (Figure 2), which may suggest functional involvement during development. Furthermore, their abundance as well as number of different sequences (see Results) could indicate several different targets. Intriguingly, both Class I and Class II RNAs are localized in the cytoplasm (Figure 3B). These features taken together—developmental regulation, cytoplasmic localization and relatively short sequences with stem forming potential—are somewhat reminiscent of pre-micro/microRNAs [reviewed in (45)]. Whether these new classes of RNAs have a similar regulatory function is yet unknown; this possibility is currently being pursued.

All the 17 box C/D snoRNA candidates identified share the canonical box C and D motifs but lack complementary nucleotides at their ends and can therefore not form the commonly found 5′- and 3′-terminal stem structure [Figure 4 and (6)]. However, this motif is not ubiquitous since the absence of complementary 5′ and 3′ nucleotides has been reported in a number of cases, [e.g. (46–48)]. Furthermore, the sequences flanking the majority of the isolated snoRNAs are predicted to form helices and thereby compensate for the lack of terminal stems (34). The failure to isolate additional box C/D snoRNA candidates with terminal stem structures, may be a result of the cloning method as discussed above. Interestingly, the expression of at least three of the snoRNA candidates is developmentally controlled; DdR-15 is down-regulated while DdR-9 and DdR-13 are up-regulated (Figure 2). Developmental regulation of snoRNAs has previously been reported in *Drosophila* (3). Whether differential expression of snoRNAs is of importance for the development is not known. We are currently creating gene disruptions of the developmentally regulated snoRNA genes in *Dictyostelium*.

DdR-18 RNA carries the hallmarks of a bona fide box H/ACA snoRNA (Figure 1B), and hence both classes of snoRNAs seem to be represented in *Dictyostelium*. The reason for the isolation of only one box H/ACA snoRNA lies probably in an extensive RNA secondary structure (see above). Neither 2′-*O*-ribose methylation nor pseudouridylation of nucleotides in rRNA have been investigated in *Dictyostelium*. Nevertheless, several putative targets for the box C/D snoRNA candidates were predicted (Figure S2) and methylation of one rRNA nucleotide, predicted to be guided by DdR-1, was experimentally supported (Figure 5). No obvious target for the box H/ACA snoRNA was predicted. Except for the U2 candidate isolated in this study, the other group of main targets for snoRNAs, snRNAs, have not been identified in *Dictyostelium*. At this point, targets for the isolated orphan snoRNAs remain elusive.

Several of the isolated snoRNA candidates are processed from polycistronic clusters (Figure 6). This is in accordance with the genomic organization of snoRNA genes in, e.g. trypanosomes (49), *Arabidopsis* and yeast (50,51).

Besides function, another question that arises is how the expression of small RNAs is regulated. This study presents several small RNAs that are down- or up-regulated during the development but it is yet unknown whether this occurs at the transcriptional or post-transcriptional level. Furthermore, very little is known about promoter sequences in *Dictyostelium*. Since the cloning protocol employed identified bona fide 5′ start nucleotides of RNAs, this permitted a search for conserved sequence elements in the upstream flanking region of the identified RNA-encoding sequences (Figure 7). One consensus motif found upstream of the U2-encoding sequence is also present in front of the SRP RNA-, Class I- and Class II-encoding regions, as well as in front some of the candidate snoRNA genes (data not shown). This sequence is similar to the USE promoter element located upstream of snRNA and 7SL RNA (SRP RNA) genes in *Arabidopsis* (41,52). The significant similarity between these motifs in *Dictyostelium*, their position and the high degree of cytidines in the otherwise A/T-rich intergenic sequences [see Table S1 and (11)] strongly indicate new promoter elements. Gene fusions and mutational analysis will be performed to test promoter activities and search for putative, developmentally regulated elements in the control of RNA genes, e.g. Class I genes.

In the last few years, discoveries of numerous new small RNAs in a variety of organisms have extended the ncRNA family dramatically (51,53). The identification of small RNAs in *Dictyostelium* revealed classes of RNAs previously known from other organisms but also novel classes of small RNAs. The challenge will now be to assign functions to these new classes of RNAs and to explore whether these RNAs are present in other organisms or are specific for *Dictyostelium*.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Gerhart Wagner for critically reading this manuscript and to Mats Ensterö and Gerhart Wagner for helpful discussions. We also thank dictyBase (http://dictybase.org) and the Dictyostelium Sequencing Center Consortium. This work was supported by grants from the Swedish Research Council, the Wallenberg Foundation and Carl Tryggers Foundation to F.S.

REFERENCES

1. Vogel,J., Bartels,V., Tang,T.H., Churakov,G., Slagter-Jäger,J.G., Hüttenhofer,A. and Wagner,E.G.H. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
2. Marker,C., Zemann,A., Terhorst,T., Kiefmann,M., Kastenmayer,J.P., Green,P., Bachellerie,J.P., Brosius,J. and Hüttenhofer,A. (2002) Experimental RNomics. identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
3. Yuan,G., Klambt,C., Bachellerie,J.P., Brosius,J. and Hüttenhofer,A. (2003) RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, **31**, 2495–2507.
4. Tang,T.H., Bachellerie,J.P., Rozhdestvensky,T., Bortolin,M.L., Huber,H., Drungowski,M., Elge,T., Brosius,J. and Hüttenhofer,A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
5. Hüttenhofer,A., Kiefmann,M., Meier-Ewert,S., O'Brien,J., Lehrach,H., Bachellerie,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
6. Bachellerie,J.P., Cavaille,J. and Hüttenhofer,A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
7. Kiss,T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
8. Kessin,R.H. (2001) *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity*. Cambridge University Press, Cambridge, UK.
9. Kuspa,A., Sucgang,R. and Shaulsky,G. (2001) The promise of a protist: the *Dictyostelium* genome project. *Funct. Integr. Genomics*, **1**, 279–293.
10. Baldauf,S.L., Roger,A.J., Wenk-Siefert,I. and Doolittle,W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, **290**, 972–977.
11. Glockner,G., Eichinger,L., Szafranski,K., Pachebat,J.A., Bankier,A.T., Dear,P.H., Lehmann,R., Baumgart,C., Parra,G., Abril,J.F. *et al.* (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature*, **418**, 79–85.
12. Wise,J.A. and Weiner,A.M. (1981) The small nuclear RNAs of the cellular slime mold *Dictyostelium discoideum*. *J. Biol. Chem.*, **256**, 956–963.
13. Kaneda,S., Gotoh,O., Seno,T. and Takeishi,K. (1983) Nucleotide sequence of *Dictyostelium* small nuclear RNA Dd8 not homologous to any other sequenced small nuclear RNA. *J. Biol. Chem.*, **258**, 10606–10613.
14. Takeishi,K. and Kaneda,S. (1981) Isolation and characterization of small nuclear RNAs from *Dictyostelium discoideum*. *J. Biochem.*, **90**, 299–308.
15. Pi,M., Morio,T., Urushihara,H. and Tanaka,Y. (1998) Characterization of a novel small RNA encoded by *Dictyostelium discoideum* mitochondrial DNA. *Mol. Gen. Genet.*, **257**, 124–131.
16. Wise,J.A. and Weiner,A.M. (1980) *Dictyostelium* small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. *Cell*, **22**, 109–118.
17. Knecht,D.A., Cohen,S.M., Loomis,W.F. and Lodish,H.F. (1986) Developmental regulation of *Dictyostelium discoideum* actin gene fusions carried on low-copy and high-copy transformation vectors. *Mol. Cell. Biol.*, **6**, 3973–3983.
18. Sussman,M. (1987) Cultivation and synchronous morphogenesis of *Dictyostelium* under controlled experimental conditions. *Methods Cell Biol.*, **28**, 9–29.
19. DeChiara,T.M. and Brosius,J. (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *Proc. Natl Acad. Sci. USA*, **84**, 2624–2628.
20. Crowley,T.E., Nellen,W., Gomer,R.H. and Firtel,R.A. (1985) Phenocopy of discoidin I-minus mutants by antisense transformation in *Dictyostelium*. *Cell*, **43**, 633–641.
21. Sucgang,R., Chen,G., Liu,W., Lindsay,R., Lu,J., Muzny,D., Shaulsky,G., Loomis,W., Gibbs,R. and Kuspa,A. (2003) Sequence and structure of the extrachromosomal palindrome encoding the ribosomal RNA genes in *Dictyostelium*. *Nucleic Acids Res.*, **31**, 2361–2368.
22. Cavaille,J. and Bachellerie,J.P. (1998) snoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res.*, **26**, 1576–1587.
23. Kiss-Laszlo,Z., Henry,Y. and Kiss,T. (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.*, **17**, 797–807.
24. Maden,B.E. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **39**, 241–303.
25. Rebane,A., Roomere,H. and Metspalu,A. (2002) Locations of several novel 2′-O-methylated nucleotides in human 28S rRNA. *BMC Mol. Biol.*, **3**, 1.
26. Higa,S., Maeda,N., Kenmochi,N. and Tanaka,T. (2002) Location of 2′-*O*-methyl nucleotides in 26S rRNA and methylation guide snoRNAs

in *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun*., **297**, 1344–1349.
27. Brown,J.W., Echeverria,M., Qu,L.H., Lowe,T.M., Bachellerie,J.P., Hüttenhofer,A., Kastenmayer,J.P., Green,P.J., Shaw,P. and Marshall,D.F. (2003) Plant snoRNA database. *Nucleic Acids Res*., **31**, 432–435.
28. Lowe,T.M. and Eddy,S.R. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168–1171.
29. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*., **31**, 3406–3415.
30. Walter,A.E., Turner,D.H., Kim,J., Lyttle,M.H., Muller,P., Mathews,D.H. and Zuker,M. (1994) Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl Acad. Sci. USA*, **91**, 9218–9222.
31. De Rijk,P., Wuyts,J. and De Wachter,R. (2003) RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics*, **19**, 299–300.
32. Argaman,L., Hershberg,R., Vogel,J., Bejerano,G., Wagner,E.G.H., Margalit,H. and Altuvia,S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol*, **11**, 941–950.
33. Bensing,B.A., Meyer,B.J. and Dunny,G.M. (1996) Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. *Proc. Natl Acad. Sci. USA*, **93**, 7794–7799.
34. Darzacq,X. and Kiss,T. (2000) Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5′,3′-terminal stem structure. *Mol. Cell. Biol*., **20**, 4522–4531.
35. Nicoloso,M., Qu,L.H., Michot,B. and Bachellerie,J.P. (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2′-O-ribose methylation of rRNAs. *J. Mol. Biol*., **260**, 178–195.
36. Kiss-Laszlo,Z., Henry,Y., Bachellerie,J.P., Caizergues-Ferrer,M. and Kiss,T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
37. Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) Splicing of precursors to mRNAs by the spliceosomes. In Gesteland,R.F., Cech,T. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.
38. Rivero,F. (2002) mRNA processing in *Dictyostelium*: sequence requirements for termination and splicing. *Protist*, **153**, 169–176.
39. Nagai,K., Oubridge,C., Kuglstatter,A., Menichelli,E., Isel,C. and Jovine,L. (2003) Structure, function and evolution of the signal recognition particle. *EMBO J*., **22**, 3479–3485.
40. Rosenblad,M.A., Zwieb,C. and Samuelsson,T. (2004) Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. *BMC Genomics*, **5**, 5.
41. Vankan,P., Edoh,D. and Filipowicz,W. (1988) Structure and expression of the U5 snRNA gene of *Arabidopsis thaliana*. Conserved upstream sequence elements in plant U-RNA genes. *Nucleic Acids Res*., **16**, 10425–10440.
42. Hildebrandt,M. and Nellen,W. (1992) Differential antisense transcription from the *Dictyostelium* EB4 gene locus: implications on antisense-mediated regulation of mRNA stability. *Cell*, **69**, 197–204.
43. Cavaille,J., Buiting,K., Kiefmann,M., Lalande,M., Brannan,C.I., Horsthemke,B., Bachellerie,J.P., Brosius,J. and Hüttenhofer,A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
44. Tang,T.H., Rozhdestvensky,T.S., d'Orval,B.C., Bortolin,M.L., Huber,H., Charpentier,B., Branlant,C., Bachellerie,J.P., Brosius,J. and Hüttenhofer,A. (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res*., **30**, 921–930.
45. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
46. Qu,L.H., Henras,A., Lu,Y.J., Zhou,H., Zhou,W.X., Zhu,Y.Q., Zhao,J., Henry,Y., Caizergues-Ferrer,M. and Bachellerie,J.P. (1999) Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol. Cell Biol*., **19**, 1144–1158.
47. Dunbar,D.A., Wormsley,S., Lowe,T.M. and Baserga,S.J. (2000) Fibrillarin-associated box C/D small nucleolar RNAs in *Trypanosoma brucei*. Sequence conservation and implications for 2′-*O*-ribose methylation of rRNA. *J. Biol. Chem*., **275**, 14767–14776.
48. Bachellerie,J.P., Michot,B., Nicoloso,M., Balakin,A., Ni,J. and Fournier,M.J. (1995) Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem. Sci*., **20**, 261–264.
49. Dunbar,D.A., Chen,A.A., Wormsley,S. and Baserga,S.J. (2000) The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. *Nucleic Acids Res*., **28**, 2855–2861.
50. Brown,J.W., Echeverria,M. and Qu,L.H. (2003) Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci*., **8**, 42–49.
51. Hüttenhofer,A., Brosius,J. and Bachellerie,J.P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol*., **6**, 835–843.
52. Heard,D.J., Filipowicz,W., Marques,J.P., Palme,K. and Gualberto,J.M. (1995) An upstream U-snRNA gene-like promoter is required for transcription of the *Arabidopsis thaliana* 7SL RNA gene. *Nucleic Acids Res*., **23**, 1970–1976.
53. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet*., **2**, 919–929.