Volume 11 Number 17 1983 Nucleic Acids Research

Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA

Liang Hu Qu, Bernard Michot and Jean-Pierre Bachellerie*

Centre de Recherches de Biochimie et de Génétique Cellulaires du CNRS, 118, route de Narbonne, 31062 Toulouse Cédex, France

Received 28 June 1983; Revised and Accepted 5 August 1983

ABSTRACT

We have developed a combined approach for probing native structures in large RNAs.

In the first method, after digestion with a structure specific nuclease, accessible sites are mapped at sequence resolution along the entire RNA molecule which is used as a template for the reverse transcriptase elongation of a 5'end labelled selected primer (coding strand of a small restriction fragment of the cloned gene). This method circumvents any prior end-labelling of RNA, a technique with major limitations for large RNAs.

In the second approach, a rapid "heterologous" sequencing can be easily applied to definite domains of an RNA molecule in a variety of species (or individuals), without additional DNA cloning nor end-labelling of RNA. By taking advantage of the presence of evolutionary conserved tracts within an RNA sequence, it allows a rapid analysis of RNA folding patterns in terms of phylogenetic comparisons : when located within such a conserved tract, selected restriction fragments from a cloned gene can be used as heterologous primers for sequencing the upstream divergent region in RNAs of other species by currently available technology, i.e. reverse transcriptase elongation in the presence of chain terminator dideoxynucleotides.

INTRODUCTION

The precise mapping of secondary and tertiary structures in RNA sequences is of fundamental importance for a detailed understanding of the role of these structural features in RNA functions.

Effective structure mapping methods for small RNA molecules like tRNAs have been reported using structure-specific enzymes, such as single-strand specific S1 nuclease (1) or double-strand specific ribonuclease V1 from cobra venom (2) : by using a terminally labelled RNA, identification of reactive sites is equated to measuring the distance of the RNA breaks from the terminal label. A variety of chemical reagents may also serve as sensitive probes of RNA structure when their reactivity is strongly dependent upon local conformation around target nucleotides (3-7). However, mapping chemically reactive sites may represent an exceedingly difficult task, even for moderately complex RNA molecules, unless a strand scission can be gene-

rated at the target site (7) as in the case of enzymatic probes.

The structure mapping methods reported so far (1, 2, 7) suffer from a series of limitations, all inherent to the necessity of a prior end-labelling of RNA before structure probing, which hamper their application to long molecules such as mRNAs and rRNAs. Firstly, native RNA structure may be irreversibly disrupted throughout the labelling procedure and it cannot be probed in the form of a native RNP complex. Secondly, little or no information can be gained for RNA domains located several hundred nucleotides away from terminal label, despite improved gel separation techniques (8). Thirdly, the obtention of an homogeneous terminal label may prove particularly difficult for the more labile long RNA molecules and is often complicated by some intrinsic terminal heterogeneity.

This paper presents a general method for mapping RNA structure at sequence resolution over the entire length of long RNA molecules. Like the above mentioned techniques (1, 2, 7), it involves the generation of structure-specific cleavages along RNA but is clearly distinct in the way cleavage sites are mapped, since no prior end-labelling of RNA is required.

For identifying secondary structure features, a more indirect but complementary approach relies upon comparative sequence analyses on different homologs. It has proven powerful in establishing structure models for tRNA (9), 5S RNA (10) and E. coli 16S rRNA (11). It obviously requires that a collection of homolog RNA sequences of sufficient diversity be available, a condition which cannot be generally fulfilled for a given RNA type by sole access to already published data. We therefore have designed a second approach which obviates this limitation since a rapid "heterologous" RNA sequencing can be performed on a wide range of species, provided some sequence tracts have been conserved among these species. Since these two approaches both rely on the reverse transcriptase extension of selected DNA primers they can be easily associated in the same experiment and both primary sequence and secondary structure features of RNA can be directly read off the same gel in a variety of species.

Application of this double approach to the 5'terminal domain of eukaryotic 28S rRNA is described.

MATERIALS AND METHODS

DNA primer :

The 3.7 kb EcoRI-BamHI fragment of mouse rDNA containing the 3'terminal domain of 18S rRNA, internal transcribed spacers, 5.8S rRNA and the

5'terminal domain of 28S rRNA was inserted into the (EcoRI + BamHI) cleaved plasmid pBR 322 giving rise to a pMEB 3 recombinant plasmid. Its restriction map and primary sequence have been reported elsewhere (12, 13, 14).

The 44 bp HinfI fragment of mouse rDNA coding for segment 79-122 of mouse 28S rRNA (positions from 5'terminus) is located within a region of very high sequence homology between yeast and mouse (12). It was obtained by a direct digestion of total pMEB 3 DNA followed by dephosphorylation with calf intestine alkaline phosphatase and preparative electrophoresis onto a 6 % acrylamide gel (Bis/Acrylamide = 1/30) in 50 mM Tris, 50 mM Boric Acid, pH 8.3, 1 mM EDTA. After elution from gel and repurification through a DEAE column, the double-stranded fragment was 5' end labelled, using (γ-^{32}P) ATP and T4 polynucleotide kinase. Strand separation was carried out by gel electrophoresis onto a 12 % acrylamide gel (Bis/Acrylamide = 1/50) immediately following a heat denaturation (90°C, 2 min), in the presence of 50 % DMSO. After elution from gel, coding strand was further purified by DEAE chromatography. All these steps were essentially performed according to Maxam and Gilbert (15). Specific activities of purified DNA primers were usually in the range 10^5-10^6 cpm/µmole.

Isolation of RNA :

. "Native" RNA : cytoplasmic ribosomes were isolated from Chinese hamster ovary cells and from mouse LY5178 cells grown in culture as described previously (16). Ribosomal RNA was extracted in 0.2M NaCl, 50 mM Tris-HCl, pH 7.4, 10 mM EDTA, 0.3 % SDS by phenol (30 min., room temperature). Aqueous phase was reextracted twice by phenol (15 min., 4°C) before ether extractions and ethanol precipitations. Separation of 28S rRNA from 18S rRNA was achieved by ultracentrifugation onto a 0.5M - 1M sucrose gradient in 0.2M NaCl, 1 mM EDTA pH 7.4, 0.1 % SDS.

. Denaturated RNA : When the analysis was restricted to the determination of primary sequence data, RNA was directly extracted from unfractionated cells in denaturing conditions, by 3M LiCl, 6M Urea according to Auffrey and Rougeon (17) as modified by Le Meur et al. (18). In earlier experiments, primer extension was carried out on purified 28S rRNA after ultracentrifugation onto a 0.5M-1M sucrose gradient in 20 mM NaCl, 1 mM EDTA pH 7.4, 0.1 % SDS. We found later that the quality of the sequence determination was not generally affected by omission of this purification step. Accordingly, for most species analyzed in this study, primer extension was directly carried out with total cellular RNA.

- Structure probing :

After ethanol precipitation, 28S rRNA was redissolved in 0.1M NaCl,2 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.6 and allowed to stand for 10 min. at 45°C. Partial digestion with S1 endonuclease was carried out in 0.2M NaCl, 50 mM Na acetate pH 5.0, 1 mM $ZnCl_2$ at 4°C with increasing amounts of enzyme. Digestion was stopped by adding 10 mM EDTA, 1 % SDS before phenol and ether extractions and ethanol precipitations.

Hybridization with DNA primer and reverse transcriptase elongation were derived from Youvan and Hearst (19).

The DNA primer was redissolved in H_2O, heat denatured (100°C, 3 min) and quickly chilled immediately before mixing with the RNA solution in H_2O which had been preequilibrated at 65°C. After 10 min. at 65°C, the mixture was made 50 mM Tris-HCl, pH 8.3, 6 mM $MgCl_2$, 40 mM KCl and allowed to stand 7 additional minutes at 65°C, before a slow cooling down to room temperature. The four normal deoxyribonucleoside triphosphates (dNTPs) were added at 200 µM. Reaction mixtures (15 µl) contained 15 µg (10 pmoles) of 28S rRNA and 0.7 pmole of 5'(^{32}P) DNA primer. Seven units of AMV reverse transcriptase (a generous gift of J.W. Beard) were added to each mixture and elongation of primer was carried out through a 30 min. incubation at 37°C. Reaction was stopped by adding 4 µl of 3 % SDS, 100 mM EDTA. After phenol extraction and ethanol precipitation, the precipitate was redissolved in 0.3M KOH and complete hydrolysis of the RNA template was achieved by an overnight incubation at 37°C. After neutralization with acetic acid, the solution was diluted with 8 volumes of 0.3M Na acetate pH 5.0 and DNA was recovered by ethanol precipitation. It was redissolved in 80 % formamide, 10 mM NaOH, 1 mM EDTA, 0.1 % xylene cyanol, 0.1 % bromphenol blue. After heat denaturation (90°C, 2 min) and quick-chilling, aliquots were analyzed according to Maxam and Gilbert (15) by electrophoresis on 6-8 % acrylamide/7M urea gels (dimensions : 38 cm x 23 cm x 0.04 cm) in 50 mM Tris boric acid, 1 mM EDTA.

- Heterologous sequencing :

Hybridization of RNA with DNA primer was performed as described in "Structure probing". Elongation by reverse transcriptase was carried out as above except for the presence of 2', 3'-dideoxyribonucleoside triphosphates (ddNTPs) (Boehringer-Mannheim). The dNTP concentrations were 100 µM for each nucleotide other than the ddNTP. For each base-specific reaction, the standard ddNTP concentrations were as follows : ddA : 5 µM, ddG : 5 µM, ddC : 2.5 µM, ddT : 5 µM with the concentration of the corresponding dNTP kept as 25 µM. Concentrations of ddNTPs were occasionally increased for an improved

resolution at nucleotide positions proximal to the DNA primer (see Discussion). Each mixture contained 0.3-3 pmoles 28S rRNA and 0.05-0.5 pmole 5' (^{32}P) DNA primer in 10 µl. Reactions were stopped and samples processed for analysis on polyacrylamide gels as described above.

RESULTS

Experimental strategy

Our strategy for probing RNA structure is depicted in Fig. 1, as compared with previous methods involving an end-labelling of RNA. A flow diagram summarizing the different experimental steps involved in this tech-

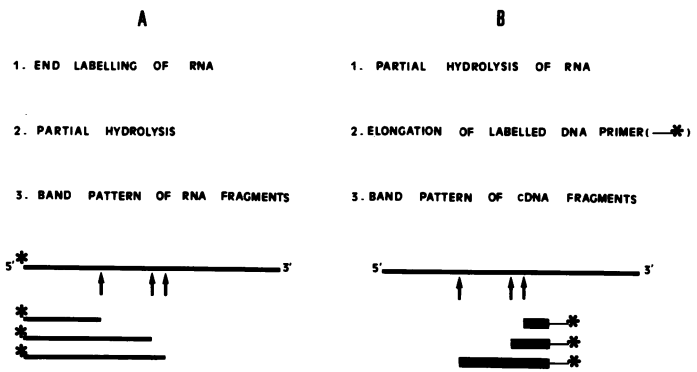


Figure 1 : Strategies for mapping higher order structures in long RNAs at sequence resolution

(A) The now classical approach involves a prior end-labelling (5' or 3') of RNA (step 1) which is followed by a partial attack by a structure-specific probe (step 2), either chemical (7) or enzymatic (1, 2), generating cleavages of phosphodiester linkages at discrete positions in RNA. Identification of accessible sites along the primary sequence is achieved (step 3) by sizing resulting end-labelled RNA fragments on polyacrylamide gels in parallel with a set of base-specific cleavage reactions (23) performed on intact end-labelled RNA.

(B) Our alternate approach : RNA is directly submitted to the action of the structure-specific probe, without a prior end-labelling, thus allowing the analysis of native RNA, which can be organized in a RNP structure (step 1). After cleavage, RNA is repurified and hybridized with a 5' end-labelled DNA primer (coding strand of a restriction fragment from the cloned gene). The identification of accessible sites is carried out through a reverse transcriptase elongation of the labelled primer (step 2). The discrete pattern of elongated labelled DNA fragments which is generated (each extending exactly down to a cleavage site) is sized by gel electrophoresis (step 3) in parallel with a set of the four base-specific reactions, i.e. reverse-transcriptase elongations of the same primer using intact RNA as a template, in the presence of chain terminator dideoxynucleotides (21).

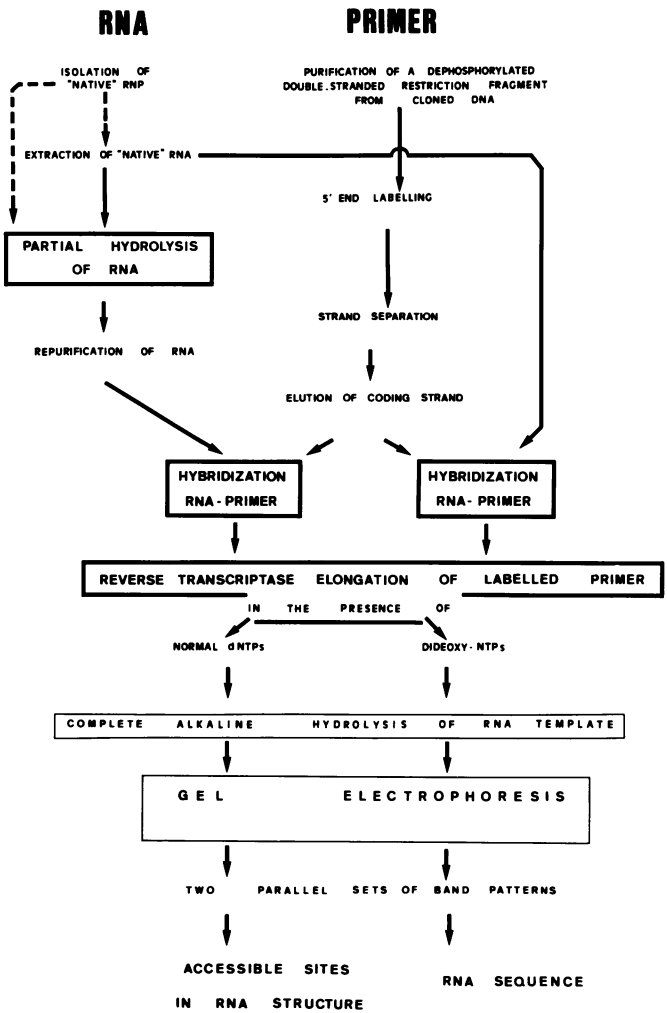


Figure 2 : Flow diagram summarizing the structure mapping technique.
Since partial RNA hydrolysis can also be carried out on RNP instead of naked RNA (broken arrow), patterns of accessible sites in both cases can be compared.

nique is shown in Fig. 2.This alternative approach, devised mainly for studying long RNA molecules, involves the synthesis and subsequent analysis of cDNAs copied from RNA template by reverse transcriptase. It has three major advantages over the methods available so far :

- firstly, probing can be carried out on a native structure since the disruptive prior end-labelling of RNA is no longer required. Hence, this ap-

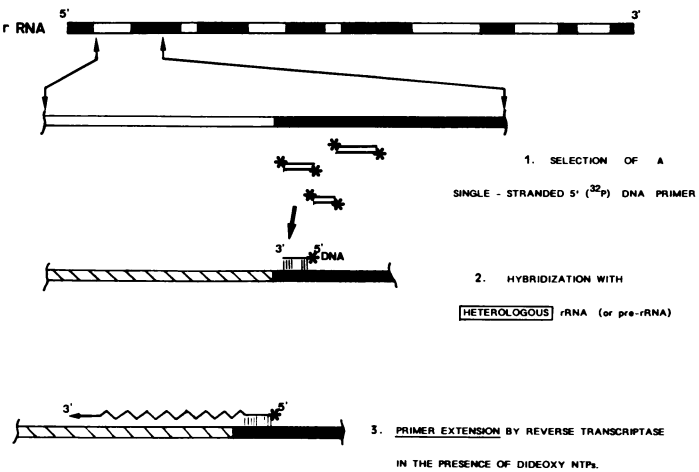


Figure 3 : Outlines of the experimental procedure for "heterologous" RNA sequencing.
Black boxes along rRNA molecule refer to highly conserved sequence tracts which are interspersed with more divergent regions (open boxes).

proach has the potential for generating detailed information about the topological organization of an RNA molecule within an RNP complex.
- secondly, the precise mapping of accessible sites is no longer restricted to regions proximal of termini as for pre-labelled RNA but the entire RNA molecule can be probed at sequence resolution, since multiple DNA primers can be selected, with convenient locations along the RNA sequence.
- last, but not least, this technology provides the basis for studying evolutionary changes in RNA structure through comparative analyses on more or less distant eukaryotes. This method takes advantage of the presence, along an RNA molecule, of sequence tracts which have been strongly conserved across a more or less broad phylogenetic range. The characteristic pattern of highly conserved areas interspersed with divergent tracts observed in rRNAs of distant eukaryotes (20, 12, 14, 8) can be utilized for performing both RNA sequencing and RNA structure probing in a wide array of eukaryotic species as schematized in Fig. 3 : when located within a conserved tract (either entirely or at least through its 15-20 3'terminal nucleotides), a DNA primer obtained from one species can also act as an "heterologous" primer for reverse transcription of other rRNAs. Primary sequence of RNA is then classically determined by primer extension in the presence of chain terminator dideoxynucleotides (21). Accordingly, the potent approach of

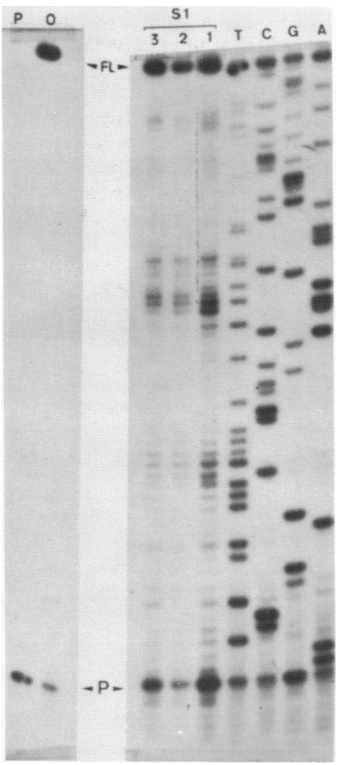


Figure 4 : Structure mapping in the 5'terminal domain of 28S rRNA.

Native 28S rRNA from hamster cells (C.H.O.) was probed for S1 nuclease accessible sites according to the procedure summarized in Fig. 2. The 5' (^{32}P) primer (lane P) was the coding strand of a 44 bp long Hinf I fragment from pMEB3 mouse ribosomal DNA recombinant. It codes for RNA segment 79-122 (positions from 5'end) in mouse 28S rRNA. Primer was elongated up to the 5'terminus of 28S rRNA (FL : full length). Template was either intact RNA (lane 0) or S1 digested rRNA : lanes 1 - 2 - 3 correspond respectively to 10 U, 20 U and 30 units nuclease S1 per µg RNA (30 min., 4°C). Elongations in the presence of ddNTPs (lanes T, C, G, A) were carried out on intact rRNA. (Electrophoresis on 8 % acrylamide-/7M urea gel).

comparative sequence analysis for derivation of secondary structure models (9-11) can be easily applied in conjunction with the direct identification of sites accessible to structure-specific probes.

Partial cleavage of hamster 28S rRNA with S1 nuclease

Fig. 4 demonstrates the potential of the present reverse transcriptase approach to identify S1 nuclease susceptible phosphodiester bonds in RNA at sequence resolution. The 5'terminal region of hamster 28S rRNA has been probed with a 5'(^{32}P) labelled 44 nucleotides long single-stranded DNA primer, generated by Hinf I restriction of cloned rDNA. Its location is shown in Fig. 5 : its 3'terminal nucleotide corresponds to position 79 in hamster 28S rRNA (from 5'terminus).

Since a primary sequence pattern is generated in parallel by elongation in the presence of chain terminator dideoxynucleotides, exact positions of S1 accessible sites are unequivocally identified along the sequence. Identical patterns of cleavage are obtained over a large range of enzyme to substrate ratios (with S1 nuclease generating less than one break per hun-

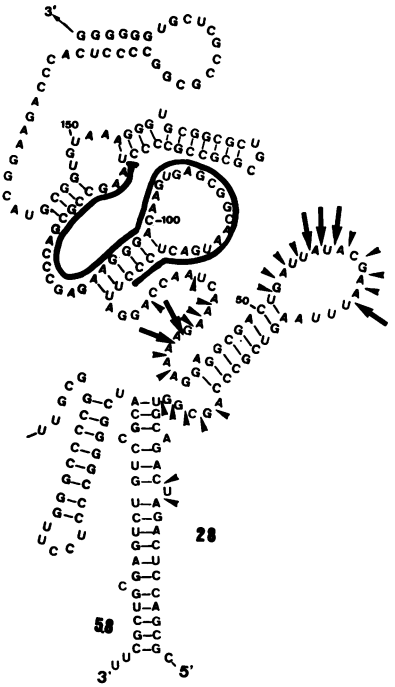


Figure 5 : Secondary structure model of the 5'terminal domain of 28S rRNA in mammals.

The region complementary to mouse DNA primer is overlined. Arrows indicate phosphodiester bonds more susceptible to cleavage by S1 nuclease, while arrowheads denote positions accessible to a lesser extent. Beyond nucleotide position 78 (from 5'-terminus), the sequence shown corresponds to mouse's (12). For segment 1-78, mammals have a common sequence (see Table 1). The same holds true for entire 5.8S rRNA (24).

dred nucleotides). These incomplete reverse transcripts do arise from S1 nuclease action, as clearly shown by their absence in the control elongation reaction performed on undigested rRNA.

It is important to note that, in this experiment, the structure of hamster 28S rRNA has been probed not with an hamster rDNA primer but with a mouse rDNA fragment. Fig. 4 therefore shows that both primary sequence and secondary structures features in RNA of different species can be rapidly determined, with a direct "read-off" from the same gel, by using heterologous DNA primers located in conserved sequence tracts. No sequence data were available so far for hamster rDNA : Fig. 4 experiment indicates that the sequence of the 78 5'-terminal nucleotides of hamster 28S rRNA is perfectly identical to its mouse counterpart (12). Results of the structure probing experiment are summarized in Fig. 5 using the secondary structure folding pattern of this region in mouse rRNA, recently derived (12) from comparative sequence analysis with yeast and procaryotes (22). It is worthwhile remembering that this 5'terminal region of eukaryotic large rRNA is involved in a strong base-pairing with the 3'terminal region of 5.8S rRNA, as discussed in the following section. It is clearly apparent that the loca-

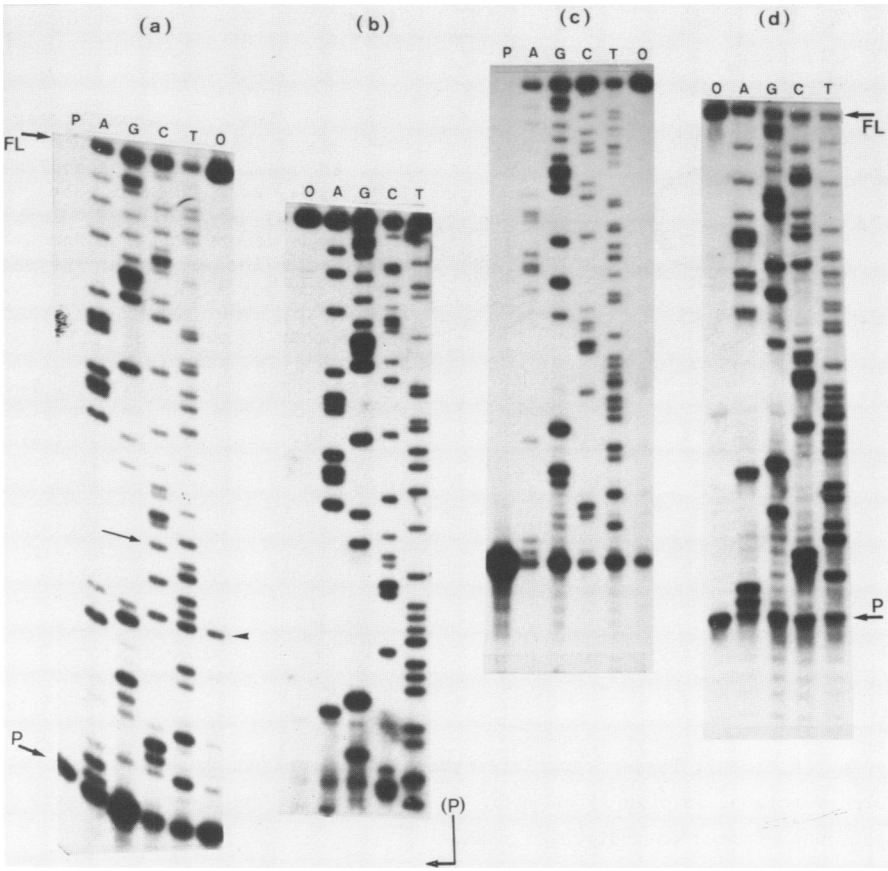
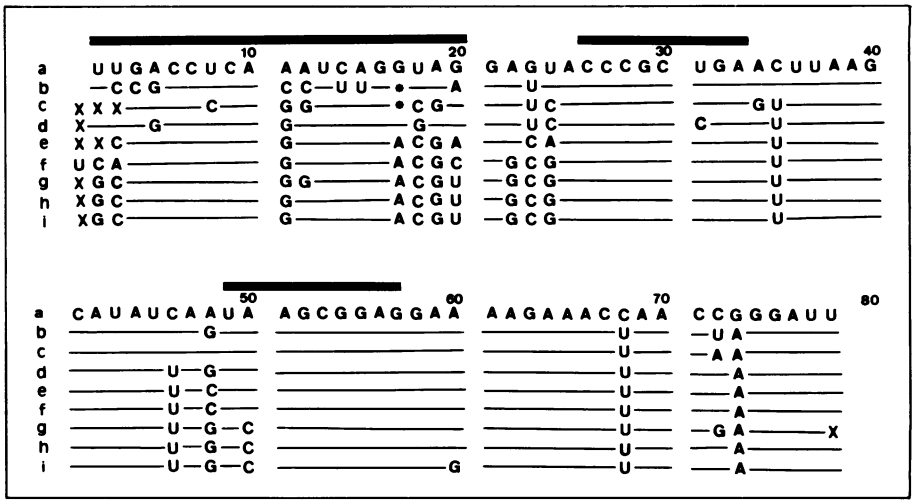


Figure 6 : Interspecies RNA sequencing using a "conserved" DNA primer from mouse rDNA

The primary sequence of the 5'terminal domain of the large rRNA was determined for distantly related eukaryotes through elongation of the same 44 nucleotides long 5' (^{32}p) DNA primer as used in Fig. 4 experiment : a) Human HeLa cells (arrow points to a human-specific mutation within an otherwise invariant sequence tract); b) Chick,Gallus gallus; c) Rainbow trout, Salmo gardneri R.; d) Turnip, Brassica napus.

tion of all S1 accessible phosphodiester bonds is in full agreement with the folding model: cleavage sites are all located in single-stranded regions or adjacent to a bulge nucleotide while all nucleotides engaged in base-paired interactions remain untouched. It is interesting to note that some nucleotide positions located in unpaired regions (such as area 34-37) in the model are not accessible to the enzyme, which could be indicative of tertiary interactions which remain to be identified.

Table 1 : Sequence of the 5'terminus of 28S (26S) rRNA in eukaryotes



All available sequences for the 5'terminal 80-odd nucleotides of eukaryotic large rRNA have been compared to yeast (22). The only written bases correspond to substitutions. Deletions are denoted by a star. X stands for an unidentified nucleotide (It must be remembered that the 5'terminal nucleotide in RNA template cannot be identified in this sequencing approach since chain termination at this position also occurs after addition of a normal dNTP). For underlined species, the sequence determination was carried out by us using the presently described reverse-transcriptase method (as shown in Fig. 6). Areas involved in potential stem structures (according to Fig. 5) are overlined by a bar.
a : yeast, S. cerevisiae (22) - b : Dictyostelium discoideum / our determination confirms and extends preliminary data by others (26)/. c : Turnip, Brassica napus ; d : Bombyx mori; e: Rainbow Trout, Salmo gardneri R.; f) Xenopus laevis (27); g : Chick, Gallus gallus; h : common to the following mammals : mouse (12), rat (28), hamster (CHO cells), rabbit and monkey (MK2 cells); i : human HeLa cells;

Sequencing of the 5'terminal domain of 28S rRNA in other eukaryotes

Owing to its location within a highly conserved region of 28S rRNA gene, the same 5' (^{32}P) DNA primer obtained from recombinant mouse rDNA has been used to derive the primary sequence of the 5'terminal region of large rRNA in a variety of eukaryotic species, across a very broad phylogenetic range, through reverse transcriptase elongation on RNA templates. Sequencing gels for a plant, a fish, a bird and a mammal are shown in Fig. 6. Results of our sequence determinations on ten eukaryotic species are summarized in Table 1, together with the previously published sequences for this region.It is worthwhile mentioning that, for all species, no sequence heterogeneity in this area of 28S rRNA was apparent. While the degree of se-

Table 2 : Divergence among 5'terminal 28S rRNA sequence

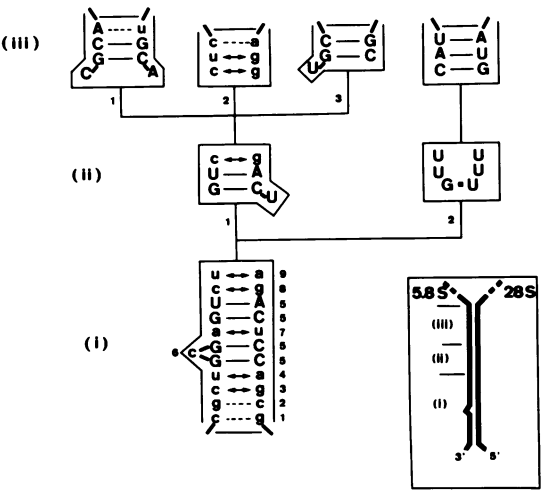
	Yeast	Dictyost.	Turnip	Bombyx	Trout	Xenopus	Chick	Mammals
Yeast	•	13.5	12.5	11	12	13	16	14
Dictyost.	13.5	•	15	13.5	14.5	18.5	17.5	17.5
Turnip	12.5	15	•	10.5	9.5	10.5	10.5	11.5
Bombyx	11	13.5	10.5	•	10	12	13	12
Trout	12	14.5	9.5	10	•	3	7	5
Xenopus	13	18.5	10.5	12	3	•	5	3
Chick	16	17.5	10.5	13	7	5	•	2
Mammals	14	17.5	11.5	12	5	3	2	•

For all pairs of eukaryotes listed in Table 1, the number of homologous positions containing different nucleotides is given (a gap vs. a nucleotide was scored 0.5). Due to a size heterogeneity and presence of still unidentified nucleotides at the 5'terminus, the analysis was restricted to positions 4-78 (numbering as in Table 1). "Mammals" stands for the species listed in Table 1, h.

quence homology between all species in this region is high enough to make all alignments unambiguous, there is however sufficient sequence diversity to allow effective comparative analyses, both in terms of phylogenetic relations between distant eukaryotes and of secondary structure folding of RNA. Differences between pairs of all eukaryotic 28S rRNA sequence listed in Table 1 have been scored (Table 2). It is noteworthy that Dictyostelium discoideum sequence shows the lowest homology with all other eukaryotic sequences, a result in agreement with previous findings by molecular phylogeny indicating that this slime mold has diverged from the eukaryotic mainstream at the earliest known branch (29, 26).

Results of a comparative analysis of these sequence data in terms of secondary structure of the region of 28S rRNA interacting with 3'terminal domain of 5.8S rRNA are summarized in Fig. 7. Despite a number of mutations in these regions of the molecules among all these species, secondary structure features have been conserved to a very high extent. A substantial core of this stem (Fig. 7, i) has remained unchanged through a number of compensatory base changes. It is remarkable that a bulge nucleotide (Fig. 7, i6)

Figure 7 : Base-paired interaction between the 3'terminus of 5.8S rRNA and the 5'terminus of 28S rRNA among phylogenetically distant eukaryotes.



All species listed in Table 1 were compared using available 5.8S rRNA sequences (24, 25). For turnip, we used Vicia faba 5.8S rRNA sequence, considering the very high conservation observed between two very distant plant sequences (24). Drosophila melanogaster was also analyzed after personal communication of 26S rRNA sequence data (B. Jacq)
This stem appears made up of three parts (inset). While in (i) the consensus structure applies to all eukaryotes, in (ii) Dictyostelium discoideum -branch 2 - differs from all other eukaryotes. In (iii), alternate structures are detected for vertebrates - (1) -, yeast and insects -(2) -, and a plant - (3) -respectively.

In each group or subgroup, a capital letter corresponds to a common nucleotide and a lower-case letter to a variable nucleotide. Base-pairings maintained through compensatory base-changes within a group are indicated by a two-head arrow. Base-pairings which are not always present within a group are shown by a broken line.

Notes:

(i) The sequence shown is for mouse

- 1 : no base-pairing in yeast, Dictyostelium, turnip, xenopus. A compensatory change is observed for Drosophila (UA instead of CG). GU pair in Bombyx mori.
- 2 : no base-pairing in turnip and xenopus. A compensatory change is observed in yeast (AU instead of GC). GU pair in Bombyx mori.
- 3 : Compensatory changes in Dictyostelium (GC) and Drosophila (UA)
- 4 : Compensatory change in Dictyostelium (UG) and Bombyx (CG)
- 5 : conserved pairs in all species with identical nucleotides.
- 6 : variable nucleotide (even between vertebrates) always present at this position as a bulge.
- 7 : Compensatory change in turnip(GC instead of AU).
- 8 : Compensatory changes in yeast, Drosophila (UA) and Dictyostelium (GC).
- 9 : Compensatory changes : Dictyostelium, Drosophila (GC), chick, turnip (CG).

In (ii)1, sequence shown is for mouse. In (iii), sequences shown are for mouse -1-, Bombyx -2- and Turnip -3 .

is always present at this position, surrounded by two GC pairs, a structure which could be involved in a basic recognition process in eukaryotic ribosome. In other parts of this complex stem, some structural variance has appeared at definite positions among different phylogenic branches (Fig. 7, ii and iii).

DISCUSSION

The distinctive characteristic of this approach is its potential for identifying the status of individual nucleotides, with reference to their presence within single-stranded loops or double helical stems and/or binding sites with effector proteins, along the entire length of long RNA molecules, instead of being restricted to their terminus-proximal regions. It has been essentially designed by the combination of two previously available techniques, i.e. the partial hydrolysis of RNA with structure-specific probes (1, 2) and the reverse transcriptase mapping of RNA 5' termini (30). Sites of phosphodiester cleavages in native RNA are identified by reverse transcriptase extension of a DNA primer, using as a template partially hydrolyzed RNA instead of intact RNA. The length of resulting cDNA corresponds to the distance from the 5' end of the primer DNA to the nucleotide immediately downstream the cleaved phosphodiester bond in RNA.

Moreover, through the same technology, a phylogenetic comparative analysis can be rapidly carried out, both in terms of primary sequence determination (32) and of mapping of "native" RNA structural features.Reliable RNA sequence data can be easily obtained from a variety of species by using whole unfractionated cellular RNA without DNA cloning nor RNA end-labelling. Once a suitable DNA primer has been selected and prepared, long RNA tracts (up to 300 nucleotides long) can be sequenced on several different species by the same worker within 3 days (including RNA isolation). Moreover, the sequence is directly read from RNA and not from a cloned, possibly non-functional, gene fragment. Accordingly, variants can be rapidly detected within a repeated gene family and assayed for potential differential expression at various stages of growth or differenciation. This approach should also allow a rapid screening of RNA sequence diversity within a population. Although rRNA may represent an extreme case for sequence conservation during evolution, the pattern of interspersion of strongly homologous tracts with more divergent sequences could also apply, in a broader sense, to a variety of structural genes, provided phylogenetic distances are short enough. For example, long blocks of perfect homology

are present in α-globin and β-globin mRNAs in mouse, rabbit and human (33-35) which could provide "multitarget" primers, obtained either from recombinant DNA or through in vitro synthesis of oligonucleotides.

In both approaches, the quality of the results is basically dependent upon optimizing several parameters :

RNA template - Hybridization: Ideally, the extent of specific cleavages by the structural probe should be maintained low so as to reduce the proportion of secondary cuttings which could be not representative of the initial RNA conformation. Preexisting nicks in RNA must be infrequent as compared with specific cleavages : they are identified in a control reaction carried out on "intact" RNA (i.e. not submitted to partial digestion). As for the sequence determination, ambiguity at some positions due to preexisting nicks (see Fig. 6, left, arrow-head for example) can generally be solved through adjustments in ddNTP concentrations, so as to increase the relative intensity of the base-specific band at these positions. Conditions for hybridization (temperature, duration, RNA concentration and molar ratio of RNA to DNA primer) have been selected for minimizing RNA nicking. It is noteworthy that, due to a prior strand-separation of the primer, these conditions need not be adjusted for each primer in order to overcome DNA renaturation.

Selection of DNA primer :A primer must meet the following criteria :
- location in the vicinity (downstream) of the RNA domain to probe.
- small size (30-60 nucleotides is optimal, but 150 nucleotides long primers are still useful), since the portion of sequence that can be resolved is decreased in proportion.
- facility of isolation by restriction from cloned recombinant DNA. Moreover, a location within a conserved sequence tract is required if a comparative structural analysis with other eukaryotic species is to be carried out. Clones of mouse rDNA recombinants provide a collection of such "multitarget" primers which can be used to probe the entire 18S and 28S rRNA molecules in most eukaryotes since blocks of very high homology with yeast (36, 22) are not restricted to the 3' domain of 18S rRNA (14) and the 5' domain of 28S rRNA (12) but are also present in other regions of mouse large rRNAs as revealed by complete sequence determination (Raynal, Hassouna, Michot and Bachellerie, in preparation).

Obviously base-pairing of primer to RNA template needs not to be perfect along the entire hybrid. A substantial number of mismatches can be tolerated unless located in the immediate vicinity of the 3'terminus of

primer, which could either prevent elongation or generate differently phased multiple cDNA band patterns.

It is important to note that radioactive reverse transcripts are generated by using a pre-labelled primer rather than performing elongation in the presence of (α -^{32}P) dNTPs. This ensures that all labelled bands do originate from this unique primer, thus avoiding artefacts that could arise either from a self-primed reverse-transcription of RNA, a reaction that has been shown to take place on U3 snRNA (37) and also on E. coli 16S rRNA (Youvan, Bachellerie and Hearst,unpublished results), or from elongation onto cDNA template of RNA primers generated by the integral RNase H activity of the reverse transcriptase, if the enzyme action is not strictly processive (38, 39).

Elongation of DNA primer : It is essential that all labelled bands actually correspond to cDNAs extending exactly to the 5' end of RNA fragments. In line with previous studies on high molecular weight SV40 cRNA (40) and E. coli 16S rRNA (19), our analyses on mouse 28S rRNA indicate that premature termination of reverse transcription is an infrequent event.

If base modifications altering Watson-Crick base-pairing can interfere with the polymerization reaction catalyzed by AMV reverse transcriptase (41, 42, 19), stable double helical stems such as present in E. coli 16S rRNA (11) do not significantly attenuate the progression of the enzyme (19). Although more extensive secondary structures in other RNAs may eventually prove difficult for the enzyme to traverse, only very minor kinetic pauses in the reverse transcription of mouse 28S rRNA have been observed so far that could be correlated to strong secondary structures (our unpublished results).Such weak pauses cannot usually interfere seriously with the identification of genuine termination points : the relative intensity of such bands can be decreased to an acceptable level by increasing the reverse transcriptase concentration, a result in line with previous reports indicating that the mechanism of action of reverse transcriptase must be at least partially distributive (38, 39).

Large rRNA and phylogeny : As for phylogenic taxonomy of eukaryotes, the sequences of large rRNAs should provide a versatile and sensitive indicator, which unlike 5SrRNA sequences (43, 44), can be useful for moderately or even closely related species. Large rRNA structure in eukaryotes can be considered as made up by the juxtaposition of a series of definite segments of different sizes which represent a wide spectrum in their rates of nucleotide substitution during evolution. Although highly conserved segments

such as the one analyzed in Table 1 are obviously useful for establishing phylogenetic relationships between very distant eukaryotes, other segments can be selected for studying a much narrower phylogenetic range, due to their more rapid divergence (our unpublished results).

Conclusion : Although by itself the direct structure mapping method we have described cannot identify which nucleotides are base-paired to which nucleotides, new information can be obtained for selecting among potential alternate RNA-RNA interactions by associating this technique with extensive comparative sequence analyses carried out through the rapid heterologous sequencing approach. Together with currently available techniques, such as psoralen cross-linking (45) or direct isolation of RNA duplexes (46), this combined approach must represent a powerful tool for deriving a thorough and precise picture of complex RNA structural organization in solution.

Acknowledgments : We are grateful to M. Nicoloso for invaluable help in isolating most RNAs analyzed in this study. The assistance of J. Feliu in the isolation of plasmid DNA was highly appreciated. We thank B. Felenbock, A. Garel and P. Jeanteur for providing us with Dictyostelium discoideum, Bombyx mori and human HeLa poly A- RNA respectively. We also thank B. Jacq for communication of sequence data prior to publication. Encouragements of Pr. J.P. Zalta were appreciated. This work was supported by a grant from C.N.R.S. (ATP : "Biologie Moléculaire du Gène"), by A.D.R.C (contract n° 6107), by a fellowship from Ligue Nationale Française contre le Cancer to B.M. and by a scholarship from the People's Republic of China to L. H. Q.

*To whom correspondence should be addressed

REFERENCES

1. Wurst, R.M., Vournakis, J.N. and Maxam, A.M. (1978) Biochemistry, 17, 4493-4499
2. Lockard, R.E. and Kumar, A. (1981) Nucleic Acids Res. 9, 5125-5140
3. Noller, H.F. (1980) in Ribosomes, J. Davies, K. Davis, L. Kahan and M. Nomura (eds) University Park Press, Baltimore, pp. 3-22
4. Farber, N. and Cantor, C.R. (1981) J. Mol. Biol. 146, 223-239
5. Wagner, R. and Garrett, R.A. (1978) Nucleic Acids Res. 5, 4065-4075
6. Bachellerie, J.P. and Hearst, J.E. (1982) Biochemistry, 21, 1357-1363
7. Peattie, D.A. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 4679-4682
8. Lockard, R.E., Connaughton, J.F. and Kumar, A. (1982) Nucleic Acids Res. 10, 3445-3457
9. Holley, R.W. (1965) Science, 147, 1462
10. Fox, G. and Woese, C.R. (1975) Nature (London) 256, 505-507
11. Noller, H.F. and Woese, C.R. (1981) Science, 212, 403-411
12. Michot, B., Bachellerie, J.P. and Raynal, F. (1982) Nucleic Acids Res. 10, 5273-5283

13. Michot, B., Bachellerie, J.P. and Raynal, F. (1983) Nucleic Acids Res.11, 3375-3391
14. Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 142, 260-266
15. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560
16. Caboche, M. and Bachellerie, J.P. (1977) Eur. J. Biochem. 74, 19-29
17. Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303-314
18. Le Meur, M., Glanville, N., Mandel, J.L., Gerlinger, P., Palmiter, R. and Chambon, P. (1981) Cell, 23, 561-571
19. Youvan, D.C. and Hearst, J.E. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3751-3754
20. Salim, M. and Maden, B.E.H. (1981) Nature (London) 291, 205-208
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467
22. Veldman, G.M., Klootwijk, J., De Regt, V.C.H.F., Planta, R.J., Branlant, C.,Krol, A. and Ebel, J.P. (1981) Nucleic Acids Res., 9, 6935-6952
23. Peattie, D.A. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1760-1764
24. Erdman, V.A. (1982) Nucleic Acids Res. 10, n° 2, r93-115
25. Fujiwara, H., Kawata, Y. and Ishikawa, H. (1982) Nucleic Acids Res. 10, 2415-2418
26. Olsen, G.J. and Sogin, M.L. (1982) Biochemistry, 21, 2335-2343
27. Hall, L.M.C. and Maden, B.E.H. (1980) Nucleic Acids Res., 8, 5993-6005
28. Subrahmanyam, C.S., Cassidy, B., Busch, H. and Rothblum, L.I. (1982) Nucleic Acids Res., 10, 3667-3680
29. Zuckerkandl, E. and Pauling, L. (1965) J. Theor. Biol. 8, 357-366
30. Bina-Stein, M., Thoren, M., Salzman, N. and Thompson, J. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 731-735
31. Ref. deleted.
32. Zimmern, D. and Kaesberg, P. (1978) Proc. Natl.Acad. Sci. U.S.A. 75, 4257-4261
33. Efstratiadis, A., Kafatos, F.C. and Maniatis, T. (1977) Cell, 10, 571-585
34. Marotta, C.A., Wilson, J.T., Forget, B.G. and Weissman, S.M. (1977) J. Biol. Chem., 252, 5040-5053
35. Pavlakis, G.N., Lockard, R.E., Vamvakopoulos, N., Rieser, L., Rajbhandary, U.L. and Vournakis, J.N. (1980) Cell 19, 91-102
36. Rubstov, P.M., Musakhanov, M.M., Zakhariev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucleic Acids Res. 8, 5779-5794
37. Bernstein, L.B., Mount, S.M. and Weiner, A.M. (1983) Cell 32, 461-472
38. Dube, D.K. and Loeb, A.L. (1976) Biochemistry 15, 3605-3611
39. Friedman, E.Y. and Rosbach, M. (1977) Nucleic Acids Res., 4, 3455-3471
40. Ghosh, P.K., Reddy, V.B., Piatak, M., Lebowitz, P. and Weissman, S.M. (1980) Methods in Enzymol. 65, 580-595
41. Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell, 13, 551-563
42. Youvan, D.C. and Hearst, J.E. (1981) Nucleic Acids Res., 9, 1723-1741
43. Osawa, S. and Hori, H. (1980) in Ribosomes, J. Davies, K. Davis, L. Kahan and M. Nomura (eds) University Park Press, Baltimore, pp. 333-355
44. Küntzel, H., Heidrich, M. and Piechulla, B. (1981) Nucleic Acids Res., 9, 1451-1461
45. Thompson, J.F. and Hearst, J.E. (1983) Cell 32, 1355-1365
46. Ross, A. and Brimacombe, R. (1979) Nature (London) 281, 271-276