2204–2211 *Nucleic Acids Research, 1996, Vol. 24, No. 12*

Common DNA structural features exhibited by eukaryotic ribosomal gene promoters

Monique Marilley* and Philippe Pasero

Laboratoire de Génétique GDR CNRS 976, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 5, France

Received April 22, 1996; Accepted May 3, 1996

ABSTRACT

Nucleotide sequences of DNA regions containing eukaryotic ribosomal promoters were analysed using strategies designed to reveal sequence-directed structural features. DNA curvature, duplex stability and pattern of twist angle variation were studied by computer modelling. Although ribosomal promoters are known to lack sequence homology (unless very closely related species are considered), investigation of these structural characteristics uncovered striking homologies in all the taxonomic groups examined so far. This wide conservation of DNA structures, while DNA sequence is not conserved, suggests that the determined structures are fundamental for ribosomal promoter function. Moreover, this result agrees well with the recent observations showing that RNA polymerase I transcription factors have not evolved as intensively as previously suspected.

INTRODUCTION

Unlike RNA polymerases II and III, RNA polymerase I is involved in the synthesis of a sole product, pre-ribosomal RNA. Consequently, it requires recognition of only one kind of starting signal for the expression of hundreds of gene units. It is highly regulated to be responsive to both general metabolism (e.g. growth rate) and to specific environmental challenges (see 1–3 for reviews). Surprisingly, systematic analyses of the nucleotide sequence around the origins of transcription of rDNA in different organisms has revealed no common pattern of nucleotide sequences (4–6). Moreover, the RNA polymerase I transcription system appears to diverge considerably between organisms. Ribosomal transcription is generally specific to taxonomic orders, the promoter of one group not being recognized by the transcription factors of another. This disparity of RNA polymerase I promoter sequences is apparently in agreement with this order of species specificity (6,7).

However, several lines of evidence now suggest the existence of a common organization of all the promoters (7–25). According to these results, a ribosomal promoter consists of essentially two domains. There is a 'proximal promoter domain' (also called the minimal or core promoter) of ~45 bp, which includes the start point of transcription and is absolutely required for determining the accuracy of initiation, and an 'upstream promoter domain' or upstream control element (UCE), mapping at about –150 bp relative to the transcription start site. Moreover, recent findings indicate that the RNA polymerase I transcription system has not diverged as intensively as first appeared. For example, one ribosomal transcription factor, the upstream binding factor (UBF) was found in human and also in *Xenopus*, mouse and rat (26–30). Mouse UBF and human UBF (and other transcription factors) were found to be functional on either the mouse or human promoter (28). A simple change of half a helical turn is also able to convert a *Xenopus laevis* promoter into a highly active mouse promoter (18). This example implies that proteins as well as DNA from both species share homologies, despite the divergence between the rRNA promoters. These apparently conflicting results can be easily explained if, as recently proposed for polymerase II transcription (31), the spatial organization of ribosomal gene promoters plays an important role in species specificity.

General studies on transcription have focused on protein–protein interactions as playing a critical role both in promoter recognition and in regulation of transcription. These studies have also pointed, as indicated before, to an additional mechanism, the assembly of a stereospecific nucleoprotein complex. This process requires proteins that bind to DNA in a sequence-specific manner, but function as architectural components. Thus, a different spatial organization of modular elements might induce species specificity.

Initiation involves sequence-specific binding of transcriptional factors to DNA. Stereospecific assembly of the nucleoprotein complex requires, in addition, that DNA structures facilitate, or at least allow, the architectural complex to be built. Because they are likely to represent a physical support for promoter activation, it is of major interest to test the hypothesis that specific structural features might be present within a promoter domain and conserved throughout all taxonomic groups.

The number of works focusing on the potential role of DNA structure in the maintenance of a specific function is now increasing. It has been shown that sequence-directed bending of DNA causes local variations in the structure of genomes (32). Bent helices are characteristics of some promoters and of other regulatory regions (for reviews see 33,34). Moreover, the basic rules of DNA curvature are now well enough established to render this parameter directly accessible to analysis on the basis

* To whom correspondence should be addressed

of the DNA sequence (32,35,36 and references therein). Direct examination of the nucleotide sequence has also proven to be valuable for the study of other structural parameters of the helix, such as duplex unwinding elements (37,38), variations of twist angle (39,40) and variations of groove size (41).

In this context, we have chosen to analyse these widely studied structural parameters (DNA curvature, helical stability and unusual variations in twist angle values) instead of directly comparing the various DNA sequences, as has been done in the past. This allowed us, as a contribution to the understanding of the role of DNA structure in the function of eukaryotic ribosomal promoters, to realize a comparative study of sequence-directed structural features and to examine how they can reflect specific structural properties.

Our results confirm a basic conserved organization of ribosomal promoters into domains. We show that these domains may be distinguished on the basis of their DNA structural features. These results support the existence of a modular organization of the ribosomal gene transcription apparatus and underline the importance of the spatial organization of the underlying DNA. Because of the conservation of these structural features from lower plants to human while the DNA sequence is not conserved, the results support the view of a structural code for DNA regulation sequences. This code should correspond to DNA structures necessary to provide a physical support for the transcription machinery.

MATERIALS AND METHODS

Sequence homologies

Nucleotide sequences are from the GenBank/EMBL database: *Homo sapiens* (X01547), *Rattus norvegicus* (X00677, K01588, M12030), *Xenopus laevis* (J01005), *Xenopus borealis* (X05263, Y00132, X00184), *Drosophila melanogaster* (X02210), *Paracentrotus lividus* (X63234), *Tetrahymena pyriformis* (J01212, M10096), *Dictyostelium discoideum* (X00601), *Arabidopsis thaliana* (X15550), *Pisum sativum* (X52575), *Triticum aestivum* (X07841), *Zea mays* (X03990) and *Physarum polycephalum* (42) The program Geneworks was used to find the best sequence alignments and to calculate the percentage homology between the DNA molecules analysed here.

DNA curvature

The algorithm for calculating DNA bending from nucleotide sequences was published by Eckdahl and Anderson (43). Three-dimensional co-ordinates of the helical axis are calculated along the sequence as previously described (44) using the parameters of the wedge model for bent DNA from Ulanovsky and Trifonov (45), Bolshoy *et al.* (46) and de Santis *et al.* (47). The magnitude of DNA bending on curvature maps is expressed as the ENDS ratio, which is defined as the ratio of the contour length of a segment of the helical axis to the shortest distance between its ends. ENDS ratios were computed at a window width of 200 nt and with a window step of 10 nt to allow comparison of the results with the data of Anderson and co-workers (32). High resolution analysis within curved regions was performed with a window width of 30 nt and a 1 nt step. This window size was chosen to be large compared to the helix pitch so that very local variations are not taken into account but remains small in comparison with promoter size, which is only ~150 nt, and far less (~45 nt) for the core promoter.

H.sapiens	*R.norvegicus*	*X.laevis*	*X.borealis*	*D.melanogaster*	*P.lividus*	*T.pyriformis*	*D.discoideum*	*A.thaliana*	*P.sativum*	*T.aestivum*	*Z.mays*	*P.polycephalum*	%
100	42	43	26	27	34	25	30	27	12	16	24	41	*H.sapiens*
	100	33	32	41	23	12	33	37	32	17	25	13	*R.norvegicus*
		100	79	33	29	41	15	38	39	23	35	42	*X.laevis*
			100	34	25	31	33	36	42	41	43	35	*X.borealis*
				100	30	14	13	17	42	19	18	40	*D.melanogaster*
					100	15	22	15	21	30	30	17	*P.lividus*
						100	35	38	29	32	45	20	*T.pyriformis*
							100	10	13	27	10	41	*D.discoideum*
								100	46	27	41	36	*A.thaliana*
									100	48	47	35	*P.sativum*
										100	55	30	*T.aestivum*
											100	24	*Z.mays*
												100	*P.polycephalum*

Figure 1. Comparison of 13 eukaryotic ribosomal gene promoters by sequence alignment. Nucleotide sequences of 80 bp (corresponding to positions –50 to +30 bp from the start point of transcription) were analysed using the Geneworks sequence alignment program and compared two-by-two. These sequences contain gene core promoters from *H.sapiens, R.norvegicus, X.laevis, X.borealis, D.melanogaster, P.lividus, T.pyriformis, D.discoideum, A.thaliana, P.sativum, T.aestivum, Z.mays* and *P.polycephalum* Percentage identities are indicated. Unless closely related, nucleotide sequences show very little sequence homology.

DNA duplex stability and twist angle pattern of variation

The thermodynamic library of Breslauer *et al.* (48) characterizing all 10 Watson–Crick nearest-neighbour interactions in DNA was used to calculate DNA duplex stability. These thermodynamic data provide an experimental basis for predicting the stability (ΔG) of any DNA duplex region by inspection of its primary sequence. We have developed a computer program similar to the Thermodyn program of Kowalski (38) to calculate the mean sliding ΔG for the chosen size of DNA segment to be studied. Each calculated value takes into account the contribution of the surrounding nucleotides. Here, values refer to the disruption of the interaction in an existing duplex at 1 M NaCl, 25°C and pH 7.

Variations of the twist angle were mapped as described for the calculation of duplex stability. Twist angle values were taken from Kabsch *et al.* (39) and from de Santis *et al.* (47).

Calculations were made with the PACS DNA program developed in our laboratory and already exploited in various nucleic acids studies (44,49,50,51).

RESULTS

Eukaryotic ribosomal promoters display a low level of sequence homology

The level of sequence homology in a wide range of ribosomal promoters was calculated by nucleotide sequence alignment using the Geneworks program. These promoters were taken from mammals (*H.sapiens*, *R.norvegicus*), amphibians (*X.laevis*, *X.borealis*), echinoderms (*P.lividus*), insects (*D.melanogaster*), protozoans (*T.pyriformis, D.discoideum*), from dicotyledonous

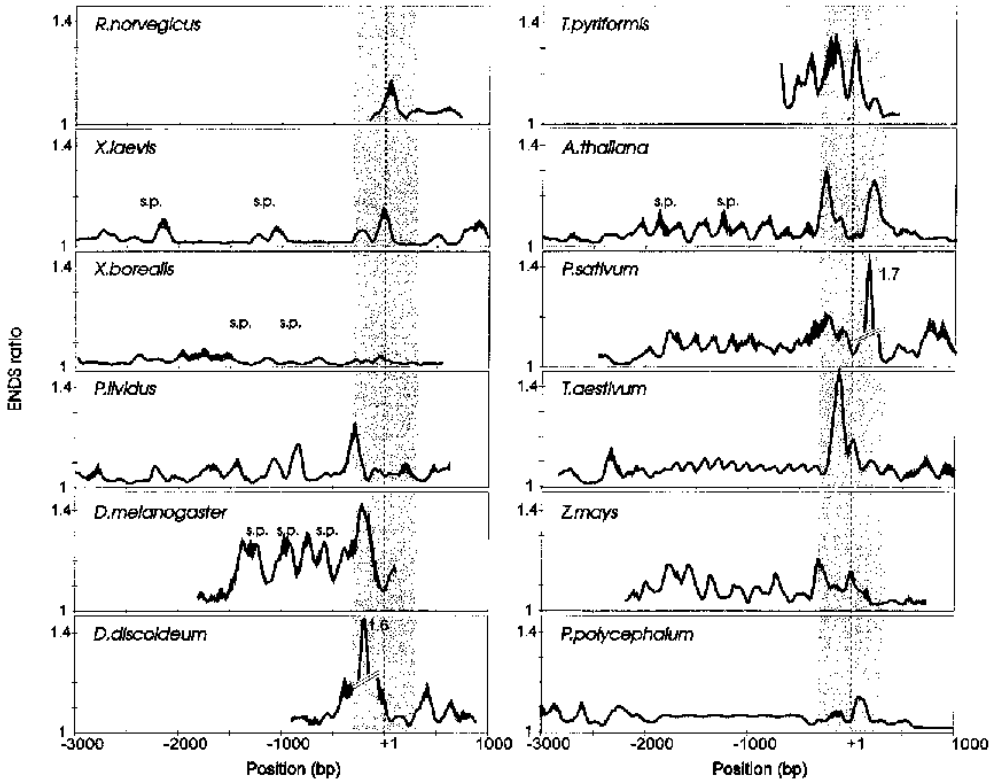

Figure 2. Curvature maps of ribosomal gene promoters. Nucleotide sequences ranging from mammalian to lower plant were analysed by computer modelling to reveal sequence-directed curvature. In each case, the three-dimensional helical path of the molecule was calculated using the models of Trifonov (46) and de Santis (47) and mean curvature maps are shown. ENDS ratios were computed at a 200 bp window size and a 10 bp step as described in Materials and Methods. Standard deviations are represented by vertical bars. Positions of nucleotides are relative to the start point of transcription (dashed line). Regions containing the ribosomal promoter are boxed. Positions of spacer promoters (SP) are indicated. ENDS ratios for random sequences with different G+C content are as follows: 100% G+C, 1.002± 0.000; 80% G+C, 1.015 ± 0.003; 60% G+C, 1.046 ± 0.013; 40% G+C, 1.080 ± 0.021; 20% G+C, 1.086 ± 0.025; 0% G+C, 1.105 ± 0.038.

(*A.thaliana*, *P.sativum*) and monocotyledonous (*T.aestivum*, *Z.mays*) plants and from myxomycetes (*P.polycephalum*). Except in closely related species, like *X.laevis* and *X.borealis*, we saw, as was already known, that the nucleotide sequences containing the ribosomal promoter region have a very low degree of homology between them (data not shown). Figure 1 focuses on the scores registered in the core promoter sequence, since this region is expected to contain the more conserved sequences. It is clear that, even within the minimal promoter region, the nucleotide sequences display a very low degree of homology.

Ribosomal gene promoters are localized within a curved region of the intergenic spacer

An increasing number of works are focusing on the presence of intrinsically curved DNA in regulatory regions (33). In order to test whether bent DNA is also an important structure of the ribosomal promoter we have analysed the structure of the corresponding DNA fragments by computer modelling. Independent wedge models of DNA curvature, like those of Trifonov and of de Santis (45–47,52), were used in this study. These models were shown to be reliable for the prediction of electrophoretic retardation and circularization and were also used for theoretical prediction of nucleosome positioning (36,43,44,47,50,51, 53–59).

Figure 2 shows an analysis of DNA curvature within the intergenic ribosomal spacer of sequences from both the animal and vegetal kingdoms. When available these nucleotide sequence analyses span from 3 kb upstream to 1 kb downstream of the transcription initiation site. This allows not only fine analysis of the ribosomal gene promoter, but also of a large region of the non-transcribed spacer (NTS) containing the spacer promoters. The magnitude of bending is expressed as the ENDS ratio and was computed for a window size of 200 bp, thus allowing a comparison with the values reported by Van Wye *et al.* (32). This analysis allowed us to determine whether curved elements are frequent or unusual features in the surroundings of the promoter. As a comparison, Van Wye *et al.*, in their analysis of the GenBank/EMBL database, defined values above 1.5 as strong bending elements (for example, the bent motif associated with the yeast ARS1 has a value of 1.54). Figure 2 shows that most of the ribosomal gene promoters display a significant DNA curvature. The deflection of the helix axis is notably stronger in some species (about 1.7 for *P.sativum*, 1.6 in *D.discoideum*) than in others. This observation is consistent with the species-dependent pattern of bending described by Van Wye. Promoters with a high G+C content have low bending scores, thus resembling on this point bacterial G+C-rich ribosomal promoters (32).

Association of a ribosomal gene promoter with DNA curvature was previously reported in the *P.polycephalum* polymerase I promoter (60). Our analysis confirms this result and shows that the occurrence of a bending structure within the polymerase I promoter region is more likely to be the rule than the exception.

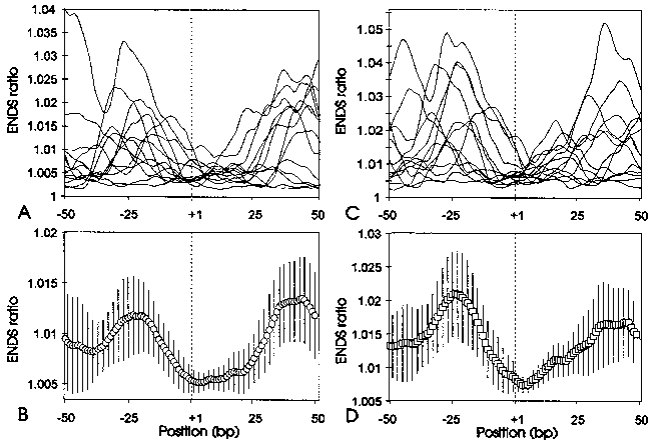

Figure 3. High resolution curvature map of eukaryotic rRNA promoters. Nucleotide sequences from *H.sapiens, R.norvegicus, X.laevis, X.borealis, D.melanogaster, P.lividus, T.pyriformis, D.discoideum, A.thaliana, P.sativum, T.aestivum, Z.mays* and *P.polycephalum* were analysed at a 30 bp window size as previously described. (**A** and **B**) Calculated using de Santis values (47). (**C** and **D**) Calculated using the Trifonov model. Detailed curvature maps of nucleotide sequence surrounding the transcription start site are shown for the 13 species. To facilitate the representation, mean values are spotted instead of individual curves (B and D). Standard deviation values (vertical bars) are indicated. The lower variations are registered for positions –1 to +10.

Highly conserved structural elements in the vicinity of the transcription start point

Because 200 bp is large compared to the size of a promoter (the core promoter is reported to be ~45 bp) we used a smaller window size (30 bp) to investigate more precisely the organization of rRNA promoters. It should be noted that the ENDs ratio values calculated here cannot be directly compared with previous ones (since bending results from the cumulative contribution of small curvatures in phase with the helix pitch, the ENDS ratio depends on the window size). Using a smaller window size allows us to separately visualize these small curvatures and enables us to see small stretches of structural elements otherwise undetected. Moreover, since many available DNA sequences are often limited to the nucleotide sequence of the promoter, decreasing the window size allows the analysis of a larger number of promoters. Figure 3 shows that several minor bending elements are involved in the three-dimensional shape of the promoter. A segment of non-curved DNA is also observed around the transcription start. Strikingly, this straight motif is highly conserved in evolution, as indicated by its low standard deviation in the averaged curves (B and D). This is clearly visible whatever the model used for structural prediction. It is worthwhile noting that it is essentially the structure and not the sequence that is conserved around the initiation site. Although a 13 bp conserved region surrounding the transcription origin was found among *Xenopus* species *X.laevis, X.borealis* and *X.clivii* (61), no significant homologies were detected in more distantly related organisms (human, mouse, *Xenopus, Drosophila* and *Tetrahymena*) (4).

DNA duplex stability

The thermodynamic stability of double-strand DNA molecules is sequence dependent. Not only GC% but also nearest-neighbour interactions between the DNA bases are involved in DNA duplex stability. Breslauer *et al.* (48) have characterized all the 10 possible interactions in a Watson–Crick DNA duplex structure. They have also shown that the stability of the duplex structure can be considered to be the sum of its nucleotide nearest-neighbour interactions. Their data are used here to predict the relative stability of local domains within the DNA region containing the ribosomal gene promoter. The approach is very similar to the one realized by Umek and Kowalski (37,38,62) for characterizing duplex unwinding elements (DUE) in replication origins.

Although the overall ΔG value appears to vary widely from one species to another (*Xenopus* rDNA is G+C rich, but *Drosophila* and *Tetrahymena* have low G+C contents) it is noticeable (Fig. 4) that all the studied sequences show a decrease in ΔG values (and G+C%) in the region of the promoter. The extent of this decrease may vary somewhat, but these values are always below the average ΔG of the surrounding sequences. It is also remarkable that this region of low ΔG is often flanked by a downstream stable domain (Fig. 4b). As shown in the lower graph of Figure 6, where a smaller window size (30 bp) is used, the decrease in ΔG values is due to a sharp decrease occurring essentially within the core promoter domain and within the UCE. It is worth mentioning that the transcription initiation site is localized at a transition zone between minimal and maximal ΔG values. This 'barrier of ΔG' is apparently conserved throughout evolution, even in the absence of sequence homology.

Patterns of twist angle variation along the promoter sequence

Sequence-dependent variations in conformational parameters such as helical twist (and other helicoidal parameters) contribute to the overall three-dimensional shape of the DNA surface and presumably to the ability of DNA binding proteins to recognize specific sequences (35). Recently, MacLeod (40) has provided evidence for a so-called pyrimidine sandwich element (PSE) which seems to play an important role in the interaction of *trans*-acting factors with DNA control regions. This shows that sequence-dependent variation in the pattern of the twist may be an important structural feature involved in specific DNA–protein recognition and may play an important function in transcription control.

Kabsch *et al.* (39) have shown that an angle larger than average usually tends to be compensated for by a smaller angle in the immediately following dinucleotide. Sequence-directed variation in the twist angle tends to prevent accumulation of over- or undertwisting along a DNA molecule. As a consequence, the structure of the B-DNA backbone typically shows a gentle zig-zag of plus or minus a few degrees. In order to detect local anomalies that may reflect some unusual structure, we have averaged successive twist angle values and followed their variations along the molecules. A 200 bp window size was chosen to focus on variations of large amplitude.

However, since the new evaluation of twist angles by de Santis *et al.* (47) resulted in values largely different from those of Kabsch, we used both sets of values and compared the two results. Figure 5 shows a comparison of this twist angle pattern in *X.laevis, X.borealis, P.lividus, D.melanogaster, D.discoideum, A.thaliana, P.sativum, T.aestivum, Z.mays* and *P.polycephalum.* The choice of these species was dictated by the size of the available sequences. Comparing long nucleotide sequences

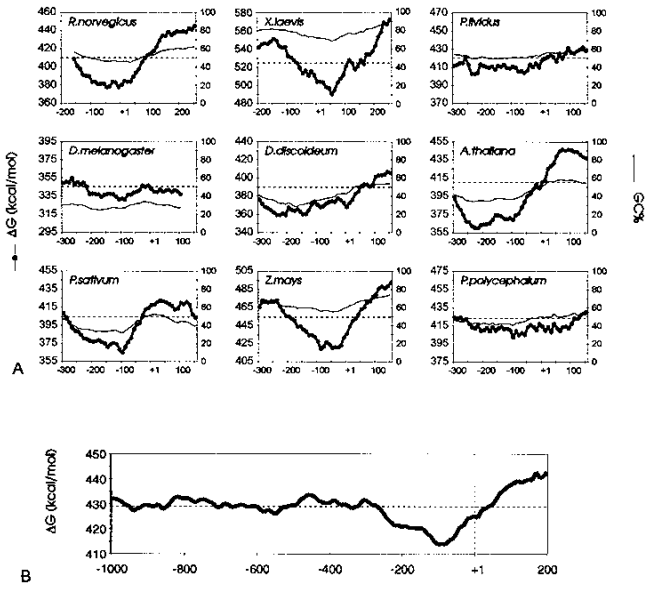

Figure 4. Variation of duplex stability along eukaryotic rRNA promoter regions. (**A**) Individual Δ*G* and G+C content maps (Δ*G* curves, filled squares; G+C content, solid line). Examples from mammals, amphibians, insects, echinoderms and dicotyledonous and monocotyledonous plants and myxomycetes are given. They are respectively: *R.norvegicus, X.laevis, P.lividus, D.melanogaster, D.discoideum, A.thaliana, P.sativum, Z.mays* and *P.polycephalum.* Mean Δ*G* values (dashed line) were calculated from surrounding nucleotide sequences (as far as possible –500 to +500 bp from the start point of transcription). Δ*G* is calculated as the sum of nearest-neighbour interaction values for a 200 bp window sliding along the sequence. Δ*G* values for randomly generated sequences (200 bp) with different G+C content are as follows: 100% G+C, 644 ± 0.7; 80% G+C, 527 ± 3.9; 60% G+C, 436 ± 3.8; 40% G+C, 368 ± 3.6; 20% G+C, 327 ± 3.1; 0% G+C, 310 ± 1.6. (**B**) Mean curve of Δ*G* variations in promoter regions and surrounding sequences. Large sized sequences were analysed: *X.laevis, X.borealis, P.lividus, D.discoideum, A.thaliana, P.sativum, T.aestivum, Z.mays* and *P.polycephalum.*

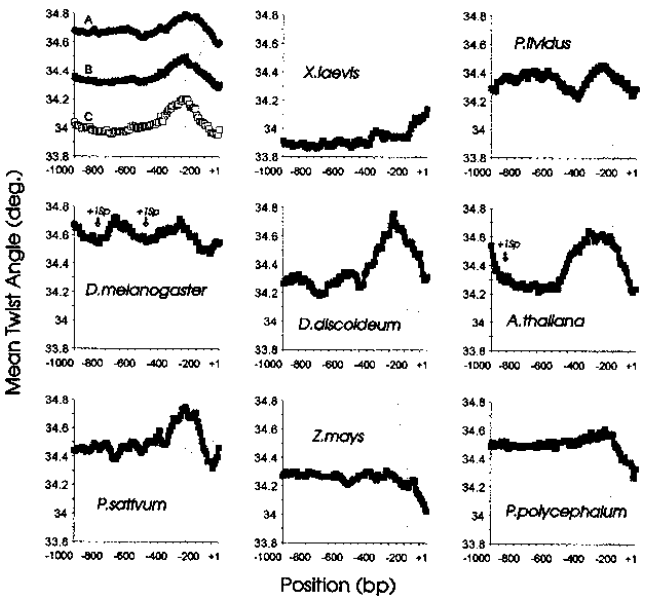

Figure 5. Patterns of twist angle variation in ribosomal gene promoters and surrounding nucleotide sequences (–900 to +50). Mean twist angle map calculated for *X.laevis, X.borealis, P.lividus, D.melanogaster, D.discoideum, A.thaliana, P.sativum, T.aestivum, Z.mays* and *P.polycephalum* promoters. (A) Calculated with de Santis values (47) (filled circle). (B) Mean curve of twist angle variations(filled square). (C) Calculated using the values Kabsch *et al.* (39) (empty square). Individual maps are shown for *X.laevis, P.lividus, D.melanogaster, D.discoideum, A.thaliana, P.sativum, Z.mays* and *P.polycephalum.* Mean values for randomly generated sequences with different G+C content are: 100% G+C, 33.77 ± 0.01; 80% G+C, 33.96 ± 0.03; 60% G+C, 34.19 ± 0.04; 40% G+C, 34.44 ± 0.03; 20% G+C, 34.71 ± 0.03; 0% G+C, 35.02 ± 0.02. A motif showing a gradual decrease in the mean twist angle corresponding to the region of the gene promoter is boxed. Sp, Spacer promoter. Window size 200 bp.

allows us to see to what extent the observed patterns are different from those of neighbouring nucleotide sequences. Although there is a lack of sequence homology and these sequences display an extremely different G+C%, we can see in the figure that the profiles of twist variation, characterized by a successive accumulation of over- and undertwisting, are strikingly similar irrespective of whether Kabsch or de Santis values of twist angles are used (Fig. 5A and C). Moreover, in most cases a sharp and continuous decrease of twist value is observed in the 200 nt sequence which includes the gene promoter. A possible exception to this rule is *X.laevis.* However, it is possible to detect a similar event in Figure 6 (where a smaller window size is used), but the decrease is visible only in the region of the core promoter.

Finally, it is worthwhile noting that in two cases, *D.melanogaster* and *A.thaliana*, the same characteristic pattern of variation is observed in other places than in the gene promoter region. Interestingly, they were found to be associated with the spacer promoters.

Various domains in the promoter can be detected by the presence of various structural features

Refining the analysis (30 bp window size) allows us to detail the structure of the promoter. Figure 6 shows a structural map of the *X.laevis* ribosomal promoter compared to a similar analysis made on 12 promoters taken from a wide range of taxonomic groups. The *X.laevis* promoter is omitted from this last analysis to avoid any interference with the result. Since the *X.laevis* promoter is one of the most extensively studied (16–20), this allows us to position structural elements relative to the nucleotide sequences important for promoter function. It is worthwhile mentioning that the different functional domains of the *X.laevis* promoter (core, UCE, enhancer homologue) correspond to regions containing specific structures. Although very different base compositions may be encountered in nucleotide sequences, *X.laevis* is by far less A+T-rich (only 15%), comparison of the two types of graph reveals common structural features. This reinforce the previous result that the sequences have developed equivalent structural characteristics for assuming promoter function. This is particularly clear for the core promoter region, where we can observe exactly the same pattern of variation (curvature, twist and helical stability) in *Xenopus* and in the set of 12 promoters. This result is indicative of a high structural conservation of this promoter domain throughout evolution (while sequence is not, as previously shown in Fig. 1). Although some species variation might be observed when individual patterns are considered, strong analogies are found throughout the promoter which allow us to distinguish different regions. Reeder (17), using linker scanner mutagenesis, concluded that the *X.laevis* promoter is composed of three domains, one of which is an enhancer element. All three domains are visible in our analysis and, although an enhancer element is

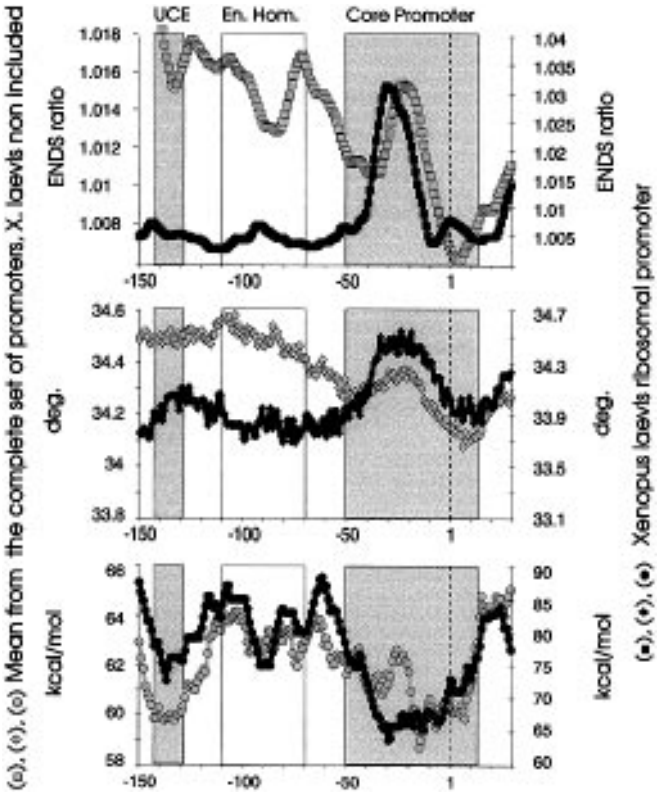

Figure 6. Conserved structural features of eukaryotic ribosomal gene promoters. Detailed analysis of DNA curvature, duplex stability and patterns of twist variation are shown for *X.laevis* (filled symbols) and for the set of eukaryotic promoters, from which *X.laevis* is omitted (shaded symbols). Curvature maps are represented as rectangles. Variations in duplex stability along the molecules correspond to circles. Patterns of twist angle variation correspond to diamonds. The window size is 30 bp. Promoter domains from *X.laevis* are represented by shaded boxes (16,17,20). The promoter region showing some similarities with the *X.laevis* enhancer homologous region is indicated by an empty box. All these regions of the *X.laevis* promoter are clearly distinguishable on the grounds of their structural parameters. All three structural parameters show a similar pattern of variation in a region which co-maps with the *X.laevis* core promoter.

not described in other promoters than *Xenopus*, an intermediary structure possibly equivalent to it is observed in the set of 12 promoters.

The organization of structural elements within promoter domains supports the proposal of Reeder (17) that the promoter functions as a set of interacting domains, but also suggests that these domains are interdependent.

DISCUSSION

Transcriptional activation requires the ordered assembly of large multiprotein–DNA complexes. Important progress has been made in the identification and purification of eukaryotic transcription factors, mainly dealing with RNA polymerase II. Analysis of the process of initiation complex assembly has determined the complex nature of protein–DNA interactions and one remarkable outcome of this research is that DNA not only contains information for binding cognate regulators but also has intrinsic structural properties playing an active role in transcription initiation (for a review see 31). We report here an analysis of the intrinsic structural features of rRNA promoters and discuss the possible functional involvement of DNA structure in ribosomal gene transcription.

DNA curvature and promoter function

Several recent studies have shown that sequence-directed curvature and protein-mediated DNA bending play a key role in the regulation of gene expression (for reviews see 33–35,63,64). The first evidence of an intrinsic bending associated with the activity of a bacterial promoter was obtained in 1984 (65) and the participation of bent DNA at nearly all the stages of prokaryotic transcription is now well documented (34). Although rarely investigated in eukaryotes, intrinsically bent DNA has been described in association with polymerase II promoters (66) and also detected in one polymerase I promoter, the ribosomal gene promoter from *Physarum* (60). Here, we have found intrinsic bending as a constant component of the ribosomal promoter. It is thus very likely that this DNA feature is a rule rather than an exception in eukaryotic promoters.

The precise function of DNA curvature associated with ribosomal promoters has now to be investigated, as was previously done with prokaryotic systems. Nevertheless, it is worthwhile noting that this structural element may be involved in a large spectrum of functions. To give a more precise idea of this variety it is important to note the following. (i) Intrinsic bending may be involved in protein docking and/or in the wrapping of DNA around proteins, which may thus be energetically favoured (see 67 for a review). (ii) It may also contribute to the formation of DNA-specific binding sites by modifying the groove shape, allowing the exposure of residues that are to interact with the cognate protein (41, see 68 for a review). (iii) Bent elements are likely to affect chromatin structure around the start site and should constitute a preferential binding site for HMG box proteins like UBF (69–73). (iv) Bent DNA may also participate in local base pair opening when destabilized by torsional stress or protein binding. In turn, once a base pair is disrupted, unstacking creates a flexible joint which is very easily bent (74). (v) DNA intrinsic bending determines the three-dimensional helix path. The three-dimensional organization of DNA together with its flexibility are very important inherent properties that must be accommodated in the assembly of the stereonucleoprotein structure of the active promoter (31, see 75 for a review). (vi) The relative orientation and phasing of bent elements may be modified by variation in the superhelicity of the DNA, thus affecting the amplitude of the curvature and, by the way, promoter activity (76).

Finally, it must be stressed that the same curvature is simultaneously involved in several of these processes. Therefore, the assembly of an active promoter must be viewed as a global and dynamic process involving the overall structure of both proteins and DNA.

ΔG and the energy requirement for strand separation during transcription initiation

The process of initiation is dependent on localized melting of the DNA double helix by the transcription complex. Several structural characteristics of the DNA molecule are known to facilitate base pair opening. Among them, DNA topology has been shown to affect the thermal requirement for strand separation. DNA bending described above is also known to lead to a significant decrease in the opening energy. This is explained

by a simultaneous lowering of the unstacking energy and by the accumulation of energy within the sugar–phosphate backbone, which may be further released to open the DNA (74). In addition, DNA supercoiling and curved DNA associated with the promoter may cooperate to induce a localized melting of the duplex (35,77,78). Kowalski and co-workers have recently shown that DNA unwinding elements are associated with some prokaryotic promoters (the β-lactamase gene) and that torsional stress alone is sufficient to locally unwind the DNA, even in the absence of initiation proteins (38). The general Δ*G* decrease that we observed here within the promoter region of the rRNA genes locally lowers the energy required for strand separation. It should be noticed that although A+T- and G+C-rich promoters do not have the same energy requirement for DNA unwinding, they have the same necessity to open it in a well-defined region. Here we observed that Δ*G* profiles are very similar in A+T- and G+C-rich sequences and clearly indicate the position of the promoter region.

The sharp Δ*G* increase 3′ of the promoter is also an interesting point to discuss (although the position and the amplitude of this peak may vary somewhat with the species). A high Δ*G* might be related to either the specificity of initiation, through stabilization of the initiation complex and/or to promoter clearance. It is worthwhile observing that the region of high Δ*G* values overlaps the 3′ region of straight DNA that surrounds the start of transcription. Base pair opening in the same sequence of straight DNA needs far higher energy than in bent DNA (74), so an increase in Δ*G* value and non-curved DNA might cooperate in positioning the region of double-strand opening. This might play a role in the accuracy of initiation.

Unusual pattern of variation of the twist angle

Twist angle variations in the ribosomal gene promoters analysed here are seen to follow an unusual pattern when compared to surrounding sequences. This unusual variation may lead to local variations in groove shape. Moreover, it should be noted that the twist angle together with the other helical angles also determine the three-dimensional helix path. Thus, this particular pattern of variation might be associated with some basic three-dimensional organization that will be discussed later.

Common structural features and species variations

Earlier investigations to decipher the sequences necessary for transcription initiation have revealed very little (if any) sequence homology between species. Even related species (mammalian sequences) were shown to be no more similar than two random segments with the same base composition (4). However, as stressed by Treco *et al.* (4), the rapid divergence of the sequences between species does not mean that they do not serve a critical function, but only that a variety of nucleotide sequences can carry out the same function equally well. The rapid evolution might also be explained by fixation of selectively advantageous mutations. Our results are in complete agreement with both explanations. Indeed, we observed that promoter regions possess intrinsic structural properties which may play an active role in the transcription process.

Previous reports have pointed out the importance of sequence variation in RNA polymerase I promoters and underlined the fact that this variation implies a concomitant evolution of the proteins of the initiation complex. However, these proteins appeared to have evolved less intensively than expected (28,79). Therefore, our observation that a high level of sequence variation does not necessarily imply important changes in the structure of the ribosomal promoter (especially within the core promoter) is more in agreement with recent results showing that numerous transcription factors (UBF, TBP, TIF IA and TIF IC) are interchangeable between species. This observation underscores the importance of transcription factors in determining promoter selectivity (79) and again gives strength to the hypothesis that species specificity may arise from stereoassembly of the nucleoprotein complex.

Because of its lack of sequence homology, the ribosomal promoter might be useful in identifying the sequence-directed structural features that are fundamental for promoter function. Here, we show that computer modelling analysis is a valid approach to identify structurally active sites in rRNA promoters. These sites can be specifically modified by mutagenesis, thus providing an additional experimental approach for investigating the complex puzzle of transcriptional activation.

ACKNOWLEDGEMENTS

This work was supported by a research grant from the Association française pour la Recherche contre le Cancer (ARC). We thank A. Humbert for helpful discussions and suggestions concerning DNA modelling. We gratefully thank Dr S. Granjeaud and C. Blettry for their help in writing new programs.

REFERENCES

1 Reeder,R.H. (1990) *Trends Genet.*, **6**, 390–395.
2 Sollner-Webb.B. and Tower,J. (1986) *Annu. Rev. Biochem.*, **55**, 801–830.
3 Sollner-Webb,B. and Mougey,E.B. (1991) *Trends Biochem. Sci.*, **16**, 58–62.
4 Treco,D., Brownell,E. and Arnheim,N. (1982) *The Cell Nucleus.* Academic Press, New York, NY, Vol. XII, pp. 101–126.
5 Sommerville,J. (1984) *Nature*, **310**, 189–190.
6 Moss,T., Mitchelson,K. and de Winter,R.F.J. (1985) *Oxford Surv. Eukaryotic Genes*, **2**, 207-250.
7 Grummt,I., Roth,E. and Paule,M.R. (1982) *Nature*, **296**, 173–174.
8 Learned,R.M., Smale,S.T., Haltiner,M.M. and Tjan,R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3558–3562.
9 Haltiner,M.M., Smale,S.T. and Tjan,R. (1986) *Mol. Cell. Biol.*, **6**, 227–235.
10 Jones HaltinerM., Learned,R.M. and Tjan,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 669–673.
11 Grummt,I. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6908–6911.
12 Yamamoto,O., Takakusa,N., Mishima,Y., Kominami,R. and Muramatsu,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 299–303.
13 Miller,K.G., Tower,J. and Sollner-Webb,B. (1985) *Mol. Cell. Biol.*, **5**, 554–562.
14 Clos,J., Normann,A., Öhrlein,A. and Grummt,I. (1986) *Nucleic Acids Res.*, **14**, 7581–7595.
15 Henderson,S.L. and Sollner-Webb,B. (1990) *Mol. Cell. Biol.*, **10**, 4970–4973.
16 Windle,J.J. and Sollner-Webb, B. (1986) *Mol. Cell. Biol.*, **6**, 4585–4593.
17 Reeder,R.H., Pennock,D., McStay,B., Roan,J., Tolentino,E. and Walker,P. (1987) *Nucleic Acids Res.*, **15**, 7429–7441.
18 Pape,L.K., Windle,J.J. and Sollner-Webb,B. (1990) *Genes Dev.*, **4**, 52–62.
19 Firek,S., Read,C., Smith,D.R. and Moss,T. (1990) *Nucleic Acids Res.*, **18**, 105–109.
20 Read,C., Larose,A.-M., Leblanc,B., Bannister,A.J., Firek,S., Smith,D.R. and Moss,T. (1992) *J. Biol. Chem.*, **267**, 10961–10967.
21 Kohorn,B.D. and Rae,P.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3265–3268.
22 Kohorn, B.D. and Rae,P.M. (1983) *Nature*, **304**, 179–181.
23 Kownin,P., Lida,C.T., Brown-Shimer,S. and Paule,M.R. (1985) *Nucleic Acids Res.*, **13**, 6237–6248.
24 Kownin,P., Bateman,E. and Paule,M.R. (1988) *Mol. Cell. Biol.*, **8**, 747–753.
25 Musters,W., Knol,J., Dekker,A.F. H., van Heerikhuisen,H. and Planta,R.J. (1989) *Nucleic Acids Res.*, **17**, 9661–9678.

26 Bell,S.P., Pikaard,C.S., Reeder,R.H. and Tjan,R. (1989) *Cell*, **59**, 489–497.
27 Pikaard,C.S., McStay,B., Schultz,M.C., Bell,S.P. and Reeder,H. (1989) *Genes Dev.*, **3**, 1779–1788.
28 Bell,S.P., Jantzen,H.-M. and Tjan,R. (1990) *Genes Dev.*, **4**, 943–954.
29 Smith,S.D., Oriahi,E., Lowe,D., Yang-Yen,H.-F., O'Mahony,D., Rose,K., Chen,K. and Rothblum,L.I. (1990) *Mol. Cell. Biol.*, **10**, 3105–3116.
30 Pikaard,C.S., Smith,S.D., Reeder,R.H. and Rothblum,L.I. (1990) *Mol. Cell. Biol.*, **10**, 3810–3812.
31 Tjan,R. and Maniatis,T. (1994) *Cell*, **77**, 5–8.
32 Van Wye,J.D., Bronson,E.C. and Anderson,J.N. (1991) *Nucleic Acids Res.*, **19**, 5253–5261.
33 Hagerman,P.J. (1990) *Annu. Rev. Biochem.*, **59**, 755–781.
34 Perez-Martin,J., Rojo,F. and de Lorenzo,V. (1994) *Microbiol. Rev.*, **58**, 268–290.
35 Calladine,C.R. and Drew,H.R. (1992) *Understanding DNA.* Academic Press. New York, NY.
36 Haran,T.E., Kahn,J.D. and Crothers,D.M. (1994) *J. Mol. Biol.*, **244**, 135–143.
37 Umek,R.M. and Kowalski,D. (1987) *Nucleic Acids Res.*, **15**, 4467–4480.
38 Natale,D.A., Umek,R.M. and Kowalski,D. (1993) *Nucleic Acids Res.*, **21**, 555–560.
39 Kabsch,W., Sander,C. and Trifonov,E.N. (1982) *Nucleic Acids Res.*, **10**, 1097–1104.
40 MacLeod,M.C. (1993) *Nucleic Acids Res.*, **21**, 1439–1447.
41 Boutonnet,N., Hui,X. and Zakrzewska,K. (1993) *Biopolymers*, **33**, 479–490.
42 Ferris,P.J. (1985) *Gene*, **39**, 203–211.
43 Eckdahl,T.T. and Anderson,J.N. (1987) *Nucleic Acids Res.*, **18**, 1609–1612.
44 Pasero,P., Sjakste,N., Blettry,C., Got,C. and Marilley,M. (1993) *Nucleic Acids Res.*, **21**, 4703–4710.
45 Ulanovsky,L.E. and Trifonov,E.N. (1987) *Nature*, **326**, 720–722.
46 Bolshoy,A., McNamara,P., Harrington,R.E. and Trifonov,E.N. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 2312–2316.
47 De Santis,P., Fuà,M., Palleschi.,A. and Savino,M. (1993) *Biophys.Chem.*, **46**, 193–204.
48 Breslauer,K.J., Frank,R., Blöcker,H. and Marky,L.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3746–3750.
49 Marilley,M., Pasero,P., Humbert,A., Granjeaud,S., Dayez,M., Pierrisnard,R. and Jordan,B. (1994) *Microsc. Microanal. Microstruct.*, **5**, 47–56.
50 Pasero,P., Blettry,C., Marilley,M., Jordan,B., Granjeaud,S., Dayez,M. and Humbert,A. (1994) *J. Vac. Sci. Technol.*, **B12**, 1521–1525.
51 Delabre,M-L., Pasero,P., Marilley,M., Bougis,P. (1995) *Biochemistry*, **34**, 6729–6736.
52 de Santis,P., Palleschi,A., Savino,M. and Scipioni,A. (1988) *Biophys. Chem.*, **32**, 305–317.
53 Griffith,J., Bleyman,M., Rauch,C.A., Kitchin,P.A. and Englund,P.T. (1986) *Cell*, **46**, 717–724.
54 Muzard,G., Théveny,B. and Révet,B. (1990) *EMBO J.*, **9**, 1289–1298.
55 Ulanovsky,L., Bodner,M., Trifonov, E.N. and Choder,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 862–866.
56 Zahn,K. and Blattner,Z.K. (1987) *Science*, **236**, 416–422.
57 Boffeli,D., de Santis,P., Palleschi,A., Risuelo,G. and Savino,M. (1992) *FEBS Lett.*, **300**, 175–178.
58 de Santis,P., Palleschi,A., Savino,M. and Scipioni,A. (1992) *Biophys. Chem.*, **42**, 147–152.
59 Nair,T.M. and Kulkarni,B.D. (1994) *Biophys. Chem.*, **48**, 383–393.
60 Schroth,G.P., Siino,J.S., Cooney,C.A., Th'ng,J.P.H., Ho,P.H.and Bradbury,E.M. (1992) *J. Biol. Chem.*, **267**, 9958–9964.
61 Bach,R., Allet,B. and Crippa,M. (1981) *Nucleic Acids Res.*, **20**, 5311–5330.
62 Umek,R.M. and Kowalski,D. (1988) *Cell*, **52**, 559–567.
63 Travers,A.A. (1992) *Curr. Opin. Struct. Biol.*, **2**, 71–77.
64 van der Vliet,P.C. and Verrijer,C.P. (1993) *BioEssays*, **15**, 25–32.
65 Bossi,L. and Smith,D.M. (1984) *Cell*, **39**, 643–652.
66 Wada-Kiyama,Y. and Kiyama,R. (1994) *J. Biol. Chem.*, **269**, 22238–22244.
67 Travers,A.A. (1989) *Annu. Rev. Biochem.*, **58**, 427–452.
68 Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
69 Kornberg,R.D. and Lorch,Y. (1992) *Annu. Rev. Cell Biol.*, **8**, 563–587.
70 Travers,A.A. (1994) *BioEssays*, **16**, 657–662.
71 Paranjape,S.M., Kamakala,R.T. and Kadanoga,J.T. (1994) *Annu. Rev. Biochem.*, **63**, 266–297.
72 Lilley,D.M. (1992) *Nature*, **357**, 282–283.
73 Bazett-Jones,D.P., Leblanc,B., Herfort,M. and Moss,T. (1994) *Science*, **264**, 1134–1137.
74 Ramstein,J. and Lavery,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7231–7235.
75 Travers,A.A. (1990) *Cell*, **60**, 177–180.
76 McAllister,C.F. and Achberger,E.C. (1989) *J. Biol. Chem.*, **264**, 10451–10456.
77 Bauer,B.R., Lund,R.A. and White,J.H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 833–837.
78 Vologodskii,A.V. and Cozzarelli,N.R. (1994) *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 609–643.
79 Rudloff,U., Stunnenberg,H.G., Keaveney,M. and Grummt,I. (1994) *J. Mol. Biol.*, **243**, 840–845.