The EMBO Journal vol.4 no.7 pp.1839-1845, 1985

Nuclease S1-sensitive sites in multigene families: human U2 small nuclear RNA genes

Han Htun, Elsebet Lund, Gunnar Westin[1], Ulf Pettersson[1] and James E.Dahlberg

Department of Physiological Chemistry, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA and [1]Medical Genetics and Microbiology, Box 589, Biomedical Center, S-751 23 Uppsala, Sweden

Communicated by U.Pettersson

We show here that human U2 small nuclear RNA genes contain a 'strong nuclease S1 cleavage site' (SNS1 site), a sequence that is very sensitive to digestion by nuclease S1. This site is located 0.50-0.65 kb downstream of the U2 RNA coding region. It comprises a 0.15-kb region in which (dC-dT)$_n$:(dA-dG)$_n$ co-polymeric stretches represent >90% of the sequence. Nuclease S1 is able to excise unit length repeats of the human U2 RNA genes both from cloned fragments and total human genomic DNA. The precise locations of the cleavage sites are dependent on the superhelicity of the substrate DNA. In negatively supercoiled substrates, cleavages are distributed over the entire 0.15-kb region, but in linearized substrates, they occur within a more limited region, mainly at the boundary of the SNS1 site closest to the human U2 RNA coding region. Nuclease S1 cleavage of negatively supercoiled substrates occurs at pHs as high as 7.0; in contrast, cleavage of linearized substrates requires a pH <5.0, indicating that supercoiling contributes to the sensitivity of this site. Mung bean nuclease gives results similar to that observed with nuclease S1.

Key words: human U2 snRNA genes/nuclease S1-sensitive site/(dC-dT).(dA-dG) co-polymer/gene homogenization/multigene families

Introduction

Small nuclear RNAs are abundant, conserved RNAs found within ribonucleoprotein particles in essentially all eukaryotes (for a review, see Busch *et al.*, 1982). Examination of the genes for two of the small nuclear RNAs, U1 and U2 RNA, reveals that in many species each of these RNAs is encoded by a multigene family. In several instances, members within each multigene family show a high degree of sequence homology extending well outside the RNA coding region of the genes (Manser and Gesteland, 1982; Denison and Weiner, 1982; Card *et al.*, 1982; Mattaj and Zeller, 1983; Lund and Dahlberg, 1984; Van Arsdell and Weiner, 1984; Htun *et al.*, 1984; Westin *et al.*, 1984a; Lund *et al.* 1984; Krol *et al.*, 1985). In the case of human U1 and U2 RNA genes, sequence homologies between the individual genes extend several kilobases (kb) upstream and downstream of the RNA coding region (Manser and Gesteland, 1982; Htun *et al.*, 1984; Van Arsdell and Weiner, 1984; Westin *et al.*, 1984a). It is unclear why at least 5-6 kb are conserved around genes that are only ~0.4-0.5 kb long, as defined by transcriptional studies in frog oocytes and in mammalian cells (Skuzeski *et al.*, 1984; Westin *et al.*, 1984b; Schenborn *et al.*, 1985; Ares *et al.*, 1985; J.M.Skuzeski, J.T.Murphy, T.Steinberg, E.Lund and J.E.Dahlberg, in preparation).

Using the single strand-specific nuclease S1 (Vogt, 1973), we probed the structure of human U1 RNA genes and found a 'strong nuclease S1 cleavage site' (SNS1 site) within (dC-dT)$_n$:(dA-dG)$_n$ co-polymeric sequences 1.8 kb downstream of the human U1 RNA coding region (Htun *et al.*, 1984). We proposed that the (dC-dT)$_n$:(dA-dG)$_n$ SNS1 site reflects a feature common to multigene families, important perhaps for the maintenance of homogeneity.

We have now extended our studies with nuclease S1 to analyze human U2 RNA genes. These genes are repeated in the genome ~20-40 times and are arranged as tandem arrays of 6.2-kb repeating units (Van Arsdell and Weiner, 1984; Westin *et al.*, 1984a). We report here that human U2 RNA genes contain an SNS1 site. This site is found once within the repeat unit from 0.50 to 0.65 kb downstream of the RNA coding region. Although the SNS1 sites of both the U1 and U2 RNA genes are composed of (dC-dT)$_n$:(dA-dG)$_n$ (Htun *et al.*, 1984; this work), the human U2 RNA genes' SNS1 sites are longer, with several stretches of (dC-dT)$_n$:(dA-dG)$_n$ co-polymeric sequences constituting >90% of the sequence in a 0.15-kb region.

Results

Identification and mapping of a strong nuclease S1 site in a 4.4-kb PstI fragment of a cloned human U2 RNA gene

We probed the physical structure of cloned human U2 RNA genes (Figure 1A) using the single strand-specific nuclease S1. When a negatively supercoiled plasmid DNA containing two-thirds of a gene repeat unit (pU2.6/1) was treated with nuclease S1, all the molecules were cleaved to yield either linear or relaxed-circular forms (Figure 1B, lanes 2 and 3). Because practically all the DNA molecules are cut by nuclease S1 at a specific site, we call this site a 'strong nuclease S1 cleavage site' or SNS1 site, in keeping with our earlier nomenclature (Htun *et al.*, 1984).

The site of nuclease S1 cleavage in pU2.6/1 DNA was mapped using several different restriction enzymes. As shown in Figure 1B (lanes 4-7) and illustrated in the top half of Figure 1A, the sizes of the fragments produced by digestion with *Hind*III and *Pst*I allowed us to deduce that the SNS1 site was ~0.5 kb downstream of the 5' end of the U2 RNA coding region. Other restriction enzyme digests verified the location of the SNS1 site (data not shown).

To determine if additional SNS1 sites existed in the human U2 RNA gene repeat units, we treated pU2.6/A DNA, which contained one-and-one-half repeat units of the human U2 RNA gene, with nuclease S1. Nuclease S1 treatment of the 14-kb supercoiled plasmid DNA generated one 6.2-kb repeat unit length fragment of the human U2 RNA gene plus the other 7.8-kb fragment, indicating that each repeat unit contained only one site; in addition, a small amount of the full-length linear fragment, formed by cleavage at only one of the two SNS1 sites, was obtained (Figure 1C, lane 3). Re-digestion of the nuclease S1 cleavage products with *Eco*RI, an enzyme that cuts at the junction bet-

ween the human and the vector DNAs (but not within the human DNA insert), left the 6.2-kb fragment intact but cleaved out the remainder of the insert from the vector (Figure 1C, lane 5). Digestion with *Kpn*I, which cuts once in the repeat unit (cf. Figure 1A), showed that the nuclease S1-generated 6.2-kb fragment contained a single *Kpn*I site and again placed the SNS1 site 0.50 kb downstream of the start of the RNA coding region (Figure 1C, lane 6). Thus, only one SNS1 site exists in each repeat unit of the human U2 RNA gene.

The strong nuclease S1 cleavage site of the human U2 RNA gene spans ~0.15 kbp

The location of the SNS1 site in supercoiled pU2.6/1 DNA was determined more precisely by labeling the nuclease S1-generated ends with ^{32}P and recutting the DNA with *Acc*I, an enzyme that cuts within a few hundred base pairs on either side of the SNS1 site (Figure 1A). When the DNA was labeled with ^{32}P after nuclease S1 treatment (Figure 2A, lane 3), label was present both in low mol. wt. DNA (L-DNA) and high mol. wt. DNA (H-DNA). When the same DNA was labeled without any prior nuclease S1 treatment, label was present only in high mol. wt. DNA (lane 2).

Treatment of end-labeled, gel-purified H-DNA with *Acc*I released two classes of labeled redigestion products (Figure 2B). From the relative distances of the *Acc*I sites upstream and downstream of the SNS1 site (cf. Figure 1A), we deduced that the larger size class (H-1) corresponded to fragments derived from sequences between the SNS1 site and the downstream *Acc*I site; these fragments typically ranged in size from 360 to 550 bp. Likewise, we deduced that the smaller size class (H-2) contained fragments extending from the SNS1 site to the *Acc*I site located between the U2 RNA coding region and the SNS1 site; these fragments ranged in size from 150 to 300 bp. As discussed below, 5′ end nucleotide analysis confirmed these assignments.

The size range of fragments obtained by the double digestions shown in Figure 2B indicated that nuclease S1 could cleave the supercoiled DNA at a number of places over a range of ~150 nucleotides. The release of fragments ranging up to ~0.15 kb (L-DNA) upon treatment of pU2.6/1 DNA with nuclease S1 alone (Figure 2A, lane 3) is consistent with that conclusion.

DNA sequence analysis of the SNS1 site revealed that the region of nuclease S1 cleavage contained several $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ co-polymeric stretches distributed over a range of ~0.15 kb (Figure 3). We have shown previously that such copolymeric sequences are sensitive to nuclease S1 under appropriate conditions (Htun *et al.*, 1984).

Nuclease S1 cleaves within $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ co-polymeric stretches at the SNS1 site

Analysis of the 5′ end nucleotides generated by nuclease S1 treatment of pU2.6/1 DNA showed that the nuclease cleaved in $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ co-polymer stretches. Figure 2C shows the results of the 5′ end nucleotide analysis using the H-1 and H-2 fragments indicated by the arrows in Figure 2B. In fragments derived from the region between the SNS1 site and the right-hand *Acc*I site (H-1), only dCMP or dTMP were at the 5′ ends generated by nuclease S1 (panel H-1). In contrast, nuclease S1-generated

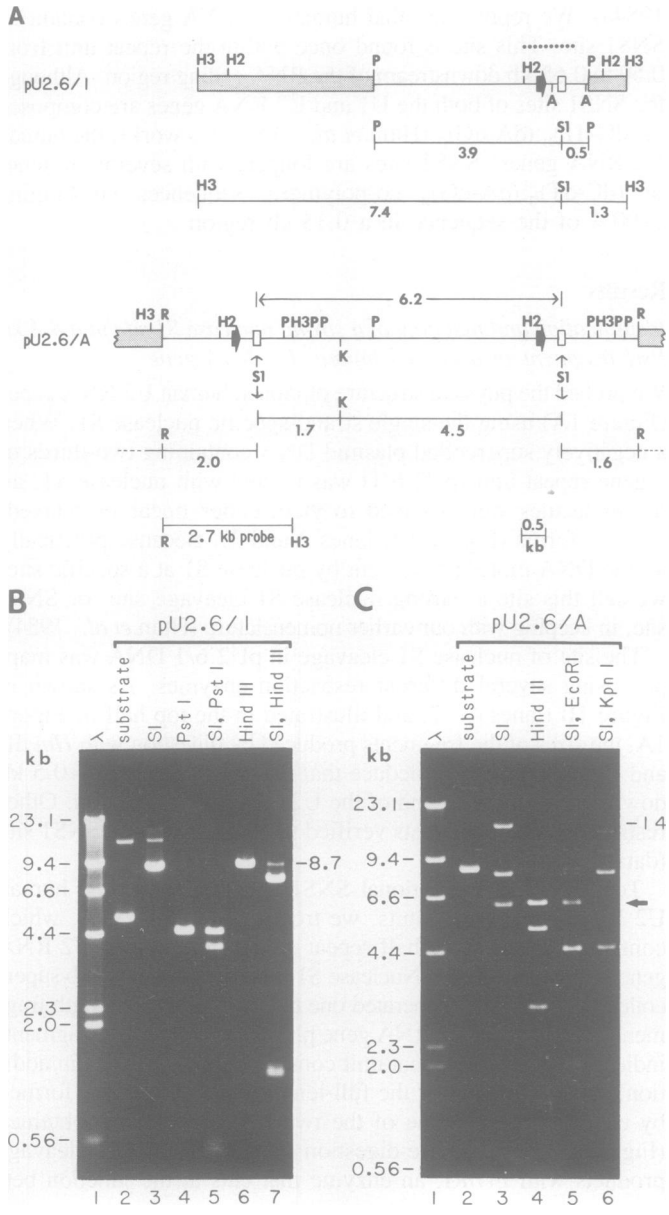
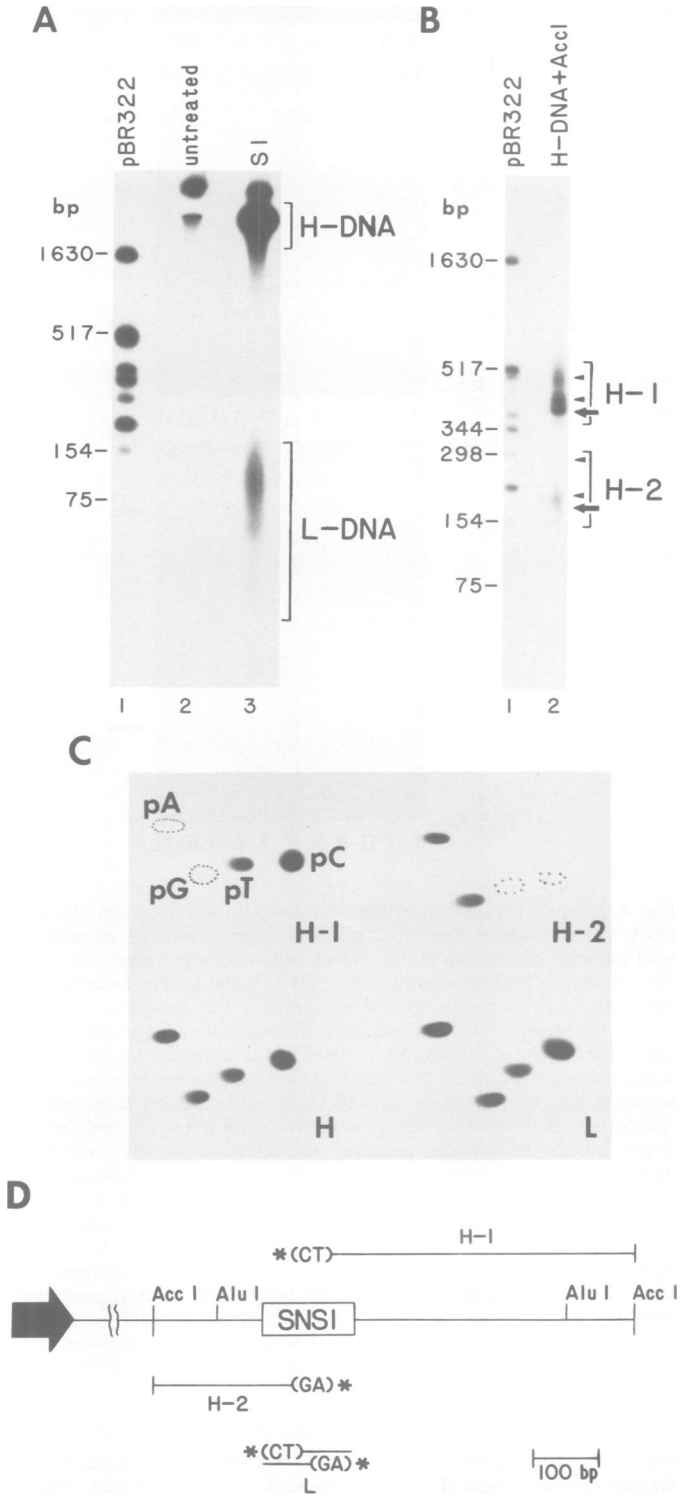


Fig. 1. Mapping of the 'strong nuclease S1 cleavage site' (SNS1 site) downstream of the human U2 RNA genes in the U2/6 locus. (**A**) Restriction cleavage site maps of pU2.6/1 and pU2.6/A DNAs. Human DNA inserts in pU2.6/1 and pU2.6/A are represented by a thin line; vector DNA sequences are indicated by stippled boxes. The U2 RNA coding region is marked by a solid horizontal arrow. The SNS1 site is indicated by an open box. The *Pst*I and *Hind*III fragments sensitive to nuclease S1 are represented by lines below the map of pU2.6/1, with the sizes of the fragments generated by nuclease S1 indicated in kbp. The 7.8-kb and the 6.2-kb fragments generated by nuclease S1 digestion of pU2.6/A DNA are represented by lines below the map of pU2.6/A. Abbrevations for restriction enzymes are: A = *Acc*I; H2 = *Hinc*II; H3 = *Hind*III; K = *Kpn*I; P = *Pst*I; and R = *Eco*RI. The vector DNA sequences of pU2.6/A are not to scale. The probe used in Figure 6 is indicated below the map for pU2.6/A DNA. The dotted line above pU2.6/1 indicates the region whose sequence is reported in Figure 3. (**B**) Mapping of the SNS1 site. Plasmid pU2.6/1 DNA was not treated (**lanes 2, 4** and **6**) or treated with nuclease S1 (**lanes 3, 5** and **7**) prior to digestion with no enzyme (**lanes 2** and **3**), *Pst*I (**lanes 4** and **5**) or *Hind*III (**lanes 6** and **7**). Samples were analyzed in a 0.8% agarose gel and DNA fragments were visualized by ethidium bromide staining. **Lane 1** shows size markers of *Hind*III-cleaved lambda DNA; numbers on the left indicate fragment sizes in kbp. Position of the 8.7-kb full-length linear DNA fragment is indicated on the right. (**C**) Excision of one complete repeat unit of human U2 RNA gene by nuclease S1. Plasmid pU2.6/A DNA was not treated (**lane 2**) or treated with only nuclease S1 (**lane 3**), or nuclease S1 followed by digestion with *Eco*RI (**lane 5**) or *Kpn*I (**lane 6**). **Lanes 1** and **4** contain *Hind*III-digested lambda DNA or pU2.6/A DNA, respectively, as markers. DNA fragments were separated in a 0.8% agarose gel and visualized by ethidium bromide staining. The arrow on the right indicates the mobility of the 6.2-kb repeat unit of the human U2 RNA gene; the position of the 14-kb full-length, linear DNA in **lanes 3** and **6** is indicated.

fragments derived from the regions to the left of the SNS1 site (H-2) contained only dAMP or dGMP as 5′-terminal nucleotides (panel H-2). Similar analyses were carried out for the other fragments indicated by arrow-heads in Figure 2B; all fragments within the brackets 'H-1' or 'H-2' contained only pyrimidines or only purines at their 5′ ends, respectively (data not shown). Furthermore, analysis of the H-1 and H-2 fragments by high resolution polyacrylamide gel electrophoresis confirmed that nuclease S1 cleaved supercoiled pU2.6/1 DNA throughout the 0.15-kb $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ region (data not shown). These results are summarized diagrammatically in Figure 2D. When total H-DNA or L-DNAs were examined (Figure 2C, panels H and L), mixtures of all four mononucleotides were present at the 5′ end as expected, since these fragments contained nuclease S1-generated ends of both DNA strands (cf. Figure 2D).



Cleavage by single strand-specific nucleases at the SNS1 site can occur at neutral pH in negatively supercoiled but not in linearized DNAs

To examine the nuclease sensitivity of the SNS1 site under less acidic conditions, we treated supercoiled pU2.6/1 DNA at various pHs with nuclease S1 and another single strand-specific nuclease, mung bean nuclease. Both nucleases cleaved the supercoiled DNA at pHs in the range of pH 4.5 to pH 7.0 but not at pH 8.0 (Figure 4A); and in all cases, the site of cleavage mapped to the SNS1 site (Figure 4B). However, in linearized pU2.6/1 DNA, both nucleases cleaved at the SNS1 site only when the pH was <5.0. Thus, when the pH is near neutrality, single strand-specific nucleases can cleave at the SNS1 site of pU2.6/1 DNA only if it is negatively supercoiled.

Nuclease S1 cleaves mainly at the left-hand boundary of the SNS1 site in linearized DNAs

Since the human U2 RNA gene's SNS1 site is composed of many stretches of $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ co-polymeric sequences spanning ~ 0.15 kb (cf. Figure 3), it is possible that various lengths of co-polymeric sequences within the SNS1 site could require more or less negative superhelical tension for nuclease S1 sensitivity. Thus, one might expect to find differences in the nuclease S1 cleavage pattern within the SNS1 site depending on whether or not the DNA is supercoiled.

As shown in Figure 5A, the pattern of nuclease S1-generated fragments differed, depending on whether linear (lane 2) or supercoiled (lane 3) DNAs were treated with nuclease S1. From the size-classes of the H-1 fragments, it appears that in linear DNAs, the major site of cleavage is $\sim 10-20$ bp from the left-hand boundary of the SNS1 site, but in supercoiled DNAs, additional cleavages are observed within the 0.15-kb SNS1 site (cf. Figure 5B). Thus, nuclease S1 sensitivity at the left-hand boundary of

Fig. 2. Characterization of DNA fragments labeled at the nuclease S1-generated ends. (**A**) Low mol. wt. DNAs released by nuclease S1 cleavage at the SNS1 site. Supercoiled pU2.6/1 DNA was not treated (**lane 2**) or treated (**lane 3**) with nuclease S1 for 10 min, labeled at the 5′ end with ^{32}P, and then separated in a 6% (60:1) polyacrylamide gel. The sample in **lane 2** contained five times the amount of DNA as that in **lane 3. Lane 1** contained size markers of ^{32}P-labeled *Hinf*I-digested pBR322 DNA. An autoradiogram is shown. (**B**) Restriction enzyme analysis of the high mol. wt. nuclease S1-generated products. High mol. wt. pU2.6/1 DNA labeled at the nuclease S1-generated ends (H-DNA in **lane 3** of **A**) was digested with *Acc*I and separated in a 6% (30:0.8) polyacrylamide gel. The resultant autoradiogram is shown (**lane 2**). Marker DNAs are *Hinf*I-digested pBR322 DNA (**lane 1**). The arrows and arrow-heads point to discrete bands present in **lane 2** which were further analyzed. The top bracket (H-1) and the bottom bracket (H-2) indicate fragments derived from the right- and left-hand sides of the SNS1 site, respectively (cf. **D** and Figure 3). (**C**) 5′ terminal nucleotides of nuclease S1-generated fragments. DNA fragments indicated in **A** and **B** were digested with DNAse I and snake venom phosphodiesterase, and the resulting 5′-labeled mononucleotides were analyzed. The first dimension of the two-dimensional chromatograms was from bottom to top using solvent (a) and the second dimension was from left to right using solvent (c) of Silberklang *et al.* (1979). The resultant autoradiograms are shown; dotted areas mark the positions of unlabeled deoxynucleoside-5′-monophosphate markers. (**D**) Schematic representation of the high mol. wt. fragments generated by nuclease S1 digestion of pU2.6/1 DNA. The lines above and below the map represent the two 5′ end-labeled (*) DNA strands extending from the $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ SNS1 site to the *Acc*I cleavage sites (cf. **A**–**C**). The deduced origin of the nuclease S1-generated L-DNA, which is 5′ end-labeled in both strands, is also shown.

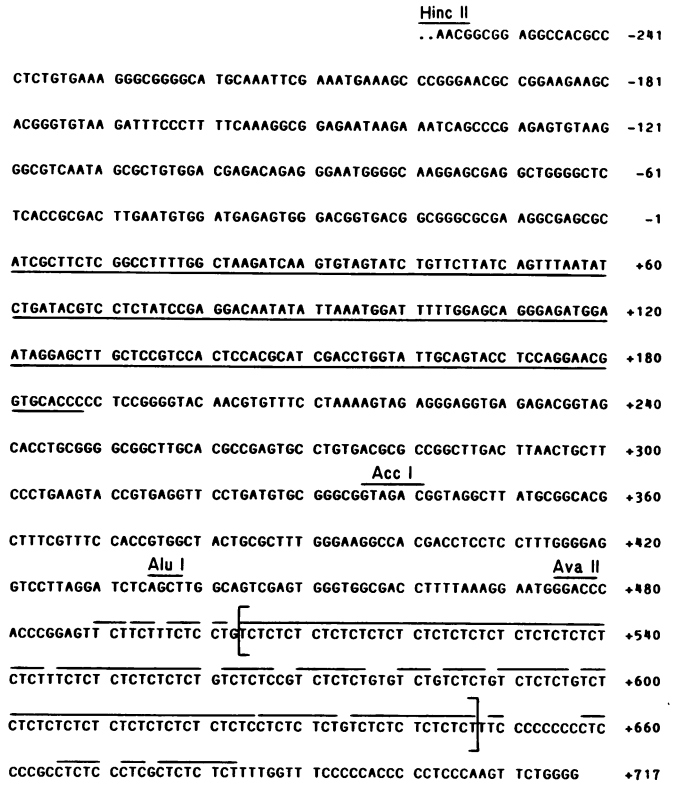


Fig. 3. Nucleotide sequence of the region of the human U2 RNA gene that contains the SNS1 site. The brackets enclose the SNS1 site defined by cleavage in supercoiled substrates. Stretches of $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ copolymeric sequences in and around the SNS1 site are overlined. The human U2 RNA coding region is underlined and the position of the *Acc*I and *Alu*I sites between the U2 RNA coding region and the SNS1 site is indicated. The other *Acc*I and *Alu*I sites discussed in the text are located around positions +1070 and +965, respectively. Sequences upstream of position +543 were determined previously (Hammarström *et al.*, 1984; Westin *et al.*, 1984a, 1984b) and are as previously reported except at positions +489 and +504 where T replaces A (Hammarström *et al.*, 1984). Sequences downstream of position +543 were determined by Maxam and Gilbert sequencing from the *Ava*II cleavage site at position +475. Only the strand containing polypyrimidines is shown.

the SNS1 site is independent of superhelical tension; whereas, cleavage at other regions of the SNS1 site is dependent on negative superhelical tension (Figure 5B).

This conclusion is supported by examination of the H-2 fragments; for both linear and supercoiled DNAs, the distribution of these fragments is similar (Figure 5A), with cleavages occurring 10–50 bp into the SNS1 site from the left-hand boundary. This apparent similarity in the distribution of the H-2 fragments is to be expected, since larger fragments generated by cleavage of superhelical DNA in more distal parts of the SNS1 site would, after relaxation, still be sensitive to nuclease S1 at the left-hand boundary (cf. Figure 5B). The release of small DNA fragments (20–150 bp long) from supercoiled DNAs (L-DNA in Figures 2A and D) but not from linear DNAs (data not shown) is consistent with multiple sites of cleavage in the former, but not the latter DNA.

Nuclease S1 can excise unit repeats of human U2 RNA genes from total chromosomal DNA

Since cleavage at SNS1 sites occurs even in relaxed DNAs (Htun *et al.*, 1984; this work), it should be possible to detect SNS1 sites in genomic DNA. As expected, when human genomic DNA was treated with nuclease S1, 6.2-kb repeat units of human U2 RNA genes were excised (Figure 6, lanes 2–5). Redigestion

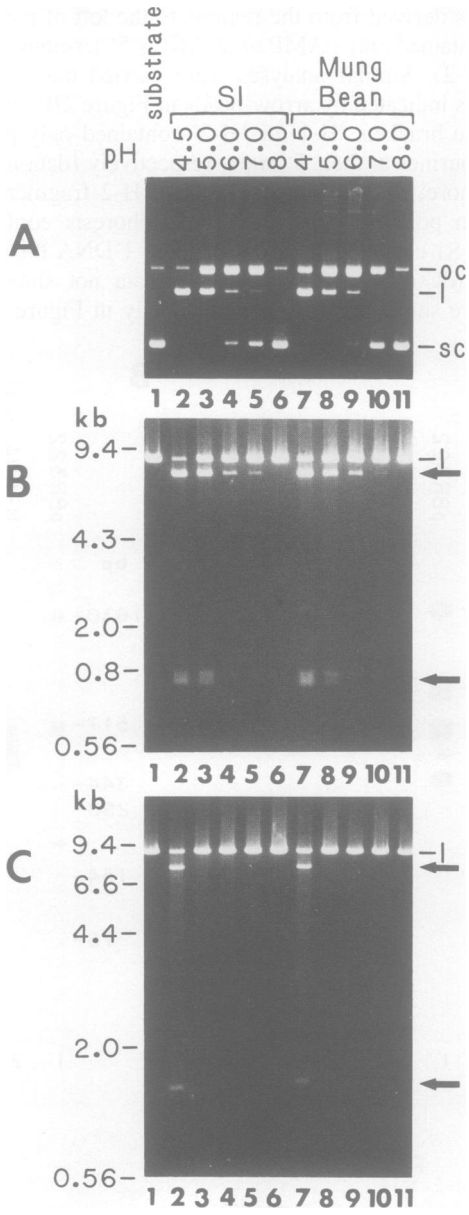


Fig. 4. Effect of pH on the cleavage of supercoiled and linearized pU2.6/1 DNA. (**A**) Cleavage of supercoiled DNA by either nuclease S1 or mung bean nuclease. Supercoiled pU2.6/1 DNA (substrate) was treated with the enzymes at pHs ranging from pH 4.5 to pH 8.0. The various forms of the DNAs are indicated on the right: sc = supercoils; l = linears; oc = open circles. (**B**) Mapping of single strand-specific nuclease cleavage sites in supercoiled DNAs at various pHs. Treated DNAs shown in **A** were redigested with *Hind*III and analyzed by agarose gel electrophoresis. The sample in each numbered lane corresponds to the comparably numbered sample as in **A**. Fragment sizes are indicated on the left in kbp and the arrows indicate the fragments generated by cleavage at the SNS1 site. The doublets seen in the lower half of the gel are a result of predominant cleavages at the left- and right-hand boundaries of the SNS1 site (cf. Figure 5B). (**C**) Effect of pH on cleavage of linearized DNA. *Hind*III-linearized pU2.6/1 DNA (substrate) was treated with nuclease S1 or mung bean nuclease at different pHs (as in **A**) and the cleavage products separated. The arrows mark the positions of the fragments produced by cleavage with the single strand-specific nuclease at the SNS1 site. Reaction conditions were as in **A** except for the omission of NaCl. DNA fragments were separated in a 0.8% agarose gel and the patterns of ethidium bromide stained DNA fragments are shown.

of the nuclease S1-treated DNA with *Hind*III mapped the SNS1 site to the same place in genomic DNA as we had determined for the cloned genes (Figure 6, lanes 6–9; Figure 1A). Thus,

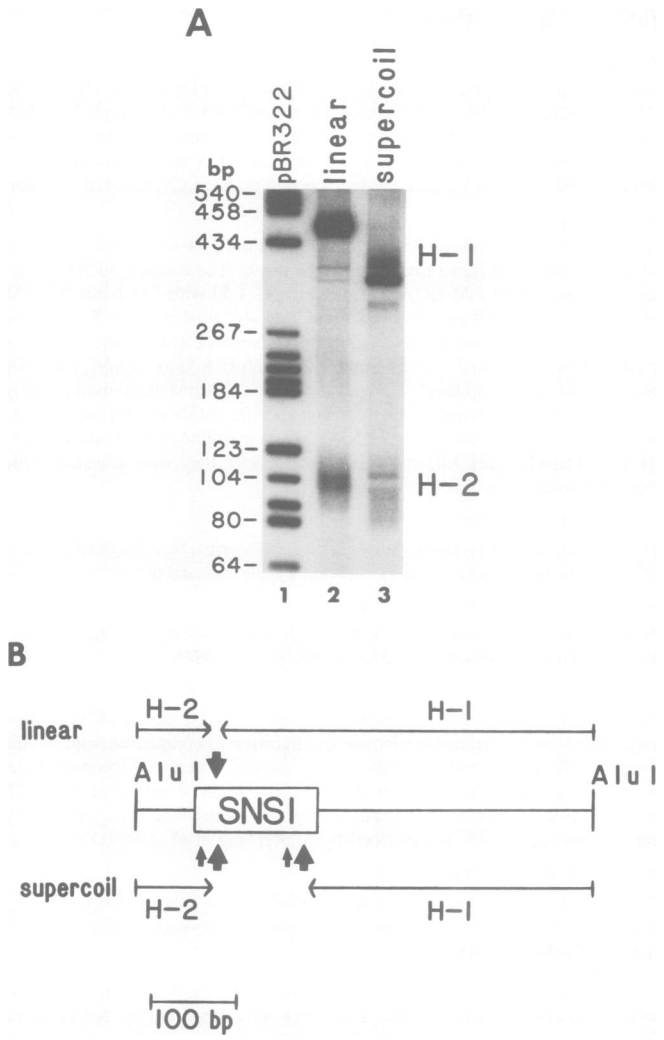


Fig. 5. Nuclease S1 cleavage within the SNS1 site of negatively supercoiled *versus* relaxed pU2.6/1 DNAs. (**A**) Restriction enzyme analysis of the high mol. wt. nuclease S1-generated products of supercoiled and linear DNAs. Supercoiled and *Hind*III-linearized pU2.6/1 DNA samples in 30 mM acetic acid (pH 4.35) and 30 mM NaCl were treated with nuclease S1 for 15 min. The nuclease S1-treated DNAs were then labeled at the 5′ end with ^{32}P and the high mol. wt. DNAs (H-DNAs) were purified and digested with *Alu*I (cf. Figure 2). An autoradiogram of the resultant labeled fragments from linear (**lane 2**) and supercoiled (**lane 3**) substrates separated in a 6% (30:0.8) polyacrylamide gel is shown. The assignments of H-1 and H-2 DNA fragments were verified by 5′-terminal nucleotide analysis as in Figure 2C. Mol. wt. markers of pBR322 DNA cleaved with *Hae*III are shown in **lane 1** and some sizes indicated in bp. (**B**) Schematic representation of nuclease S1 cleavage in negatively supercoiled and linear DNAs. The 530-bp *Alu*I fragment, containing the SNS1 site (open box), is shown. The major sites of cleavages (cf. **A**) are indicated by thick vertical arrows; the predominant minor sites are indicated by smaller vertical arrows. Cleavages in linear and supercoiled DNAs are indicated above and below the map, respectively.

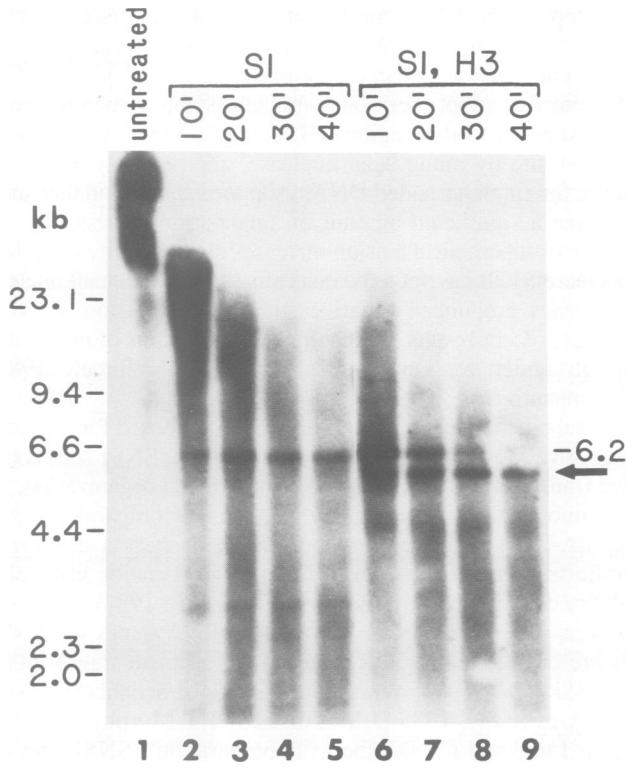


Fig. 6. Excision of human U2 RNA gene repeat units from genomic DNA by nuclease S1. Samples of 20 μg of HeLa cell DNA were either not treated (**lane 1**) or treated with nuclease S1, in standard nuclease S1 buffer modified to 10 mM NaCl, pH 4.35, for 10 min (**lanes 2** and **6**), 20 min (**lanes 3** and **7**), 30 min (**lanes 4** and **8**), or 40 min (**lanes 5** and **9**). After nuclease S1 treatment, half of each sample was further digested with *Hind*III (**lanes 6–9**). DNA fragments, separated in a 0.6% agarose gel, were transferred to nitrocellulose filter paper (Southern, 1975) and hybridized to a ^{32}P-labeled 2.7-kb *Hind*III-*Eco*RI fragment, containing the human U2 RNA gene and the associated SNS1 site (see Figure 1A). The resultant autoradiogram is shown. The position of the 6.2-kb human U2 RNA gene repeat unit is indicated; the arrow points to the position of the nuclease S1-*Hind*III product arising from cleavage at the SNS1 site.

we conclude that many, and probably all, human U2 RNA gene repeat units have an SNS1 site.

Discussion

We report that a 'strong nuclease S1 cleavage site' or SNS1 site, consisting of several $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ stretches is found 0.50–0.65 kb downstream of the 5′ end of human U2 RNA coding regions. Cleavage at the SNS1 site in supercoiled pU2.6/1 DNA is relatively independent of pH but cleavage of linearized substrates requires a pH below 5.0. Mung bean nuclease, another single strand-specific nuclease, produces similar results.

Within the SNS1 site of supercoiled pU2.6/1 DNA, cleavages are distributed over a rather broad range of ~150 nucleotides. In contrast, the distribution of cleavages in linearized substrates is more limited, being localized to a region that contains a stretch of 20 (dC-dT):(dA-dG) repeats. In linearized DNA, few cleavages are observed at the next largest uninterrupted stretch, $(dC\text{-}dT)_{14}$:$(dA\text{-}dG)_{14}$. Thus, nuclease S1 cleavage at $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ stretches in linearized DNA appears to be a function of the length of the $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ stretches, as we proposed earlier (Htun *et al.*, 1984). Additionally, the base composition or sequence of nucleotides surrounding such stretches could influence the position or efficiency of cleavage by nuclease S1.

The precise reason why $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ sequences are sensitive to nuclease S1 is unknown. Nuclease S1 cleaves negatively supercoiled DNAs in closely-spaced inverted repeats (Lilley, 1980; Panayotatos and Wells, 1981), in direct repeats (Hentschel, 1982; Mace *et al.*, 1983; Dybvig *et al.*, 1983; Htun *et al.*, 1984), at junctions between B-DNA and Z-DNA (Singleton *et al.*, 1982), in polypyrimidine.polypurine stretches (Shen, 1983; Nickol and Felsenfeld, 1983; Schon *et al.*, 1983; Evans *et al.*, 1984; Htun *et al.*, 1984), and in A-T rich regions (Hofstetter *et al.*, 1976). Presumably, such sequences contain single-stranded regions or

they adopt other DNA conformations that are susceptible to cleavage by the nuclease. We feel that under negative superhelical tension and mildly acidic conditions, the $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ SNS1 site can adopt thermodynamically stable structures containing single-stranded regions (Htun *et al.*, 1984). Cleavage at the SNS1 site by mung bean nuclease, another enzyme that is specific for single-stranded DNAs, supports the notion that such sites have a significant amount of single-strandedness.

Negative superhelical tension increases the sensitivity of DNAs to nuclease S1. It has yet to be determined whether small nuclear RNA genes are under negative superhelical tension *in vivo*. However, recent results suggest that active regions of chromatin may be under torsional stress (Ryoji and Worcel, 1984; Villeponteau *et al.*, 1984).

Several biochemical functions have been proposed for nuclease S1-sensitive sites, but none has been established. SNS1 sites could affect transcription, chromosome structure and organization, or the homogeneity of multigene families. Comparable $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ sequences are present in or near a cell type-specific transcription enhancer for a class II major histocompatibility complex gene of mouse (Gillies *et al.*, 1984). We note, however, that in the human U1 and U2 RNA genes the SNS1 sites are outside the regions required for efficient transcription (Skuzeski *et al.*, 1984; Westin *et al.*, 1984b; Schenborn *et al.*, 1985; Ares *et al.*, 1985; J.M.Skuzeski, J.T.Murphy, T.Steinberg, E.Lund and J.E.Dahlberg, in preparation). SNS1 sites do not appear to play a structural role in the attachment of DNA to the chromosome scaffold, since the $(dC\text{-}dT)_{10}$:$(dA\text{-}dG)_{10}$ of *Drosophila melanogaster* histone genes (Glikin *et al.*, 1983) is not associated with that structure (Mirkovitch *et al.*, 1984). It is possible that the $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ SNS1 sites associated with human U1 and U2 RNA genes represents a remnant of a primitive DNA sequence (Ohno and Epplen, 1983) or were created during gene amplification from an archetypal gene to generate a multigene family; if that were the case, these SNS1 sites need not at present have a function.

The DNA secondary structure formed at the SNS1 site could have an important role in the evolution of these small nuclear RNA genes. Previous studies have shown that some nuclease S1-sensitive structures resemble intermediates in genetic recombination (Lilley and Kemper, 1984) and that the border of some gene conversion units (Slightom *et al.*, 1980; Shen *et al.*, 1981) exhibit nuclease S1 sensitivity (Killpartrick *et al.*, 1984; Stoeckert *et al.*, 1984). As such, we feel that the SNS1 site could participate in promoting recombination events. Locally denatured regions might be nucleation points for genetic exchanges leading to gene conversion and subsequent homogenization of multigene families (Htun *et al.*, 1984). $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ SNS1 sites would be good candidates for such nucleation points, since in negatively supercoiled DNA, we have shown that they readily undergo denaturation near physiological pH. In support of the notion that $(dC\text{-}dT)_n$:$(dA\text{-}dG)_n$ SNS1 site may promote genetic exchange is their occurrence in many members of multigene families. Recently, we discovered that the repeat unit of *Xenopus laevis* embryonic U1 RNA genes also has an SNS1 site composed of $(dC\text{-}dC\text{-}dT)_8$:$(dG\text{-}dG\text{-}dA)_8$ (Krol *et al.*, 1985).

One prediction of the model for a role of SNS1 sites in gene conversion is that members of multigene families that lack these sites would tend to evolve or drift away from the other members. In this regard, it is interesting to note that the sequence homology between a human U2 RNA pseudogene, U2/7, and the true U2 RNA genes ends abruptly just before the SNS1 site of the true genes (Hammarström *et al.*, 1984).

Materials and methods

Plasmid

The maps of plasmids pU2.6/1 and pU2.6/A are shown in Figure 1A. The plasmid pU2.6/1 contains a 4.4-kb *Pst*I fragment of human DNA with a single U2 RNA gene cloned in pBR322 (Westin *et al.*, 1984a). The plasmid pU2.6/A contains an insert of ~10 kb of human DNA with two U2 RNA genes within one-and-one-half repeat units of a U2 gene cluster cloned in pUN121 (Westin *et al.*, 1984b).

Nuclease treatment

Single strand-specific nuclease S1 (Vogt, 1973) and mung bean nuclease (Kowalski *et al.*, 1976) were purchased from P-L Biochemicals. The standard nuclease reaction condition was 30 mM HOAc (adjusted to pH 4.53 with Tris base), 80 mM NaCl, 1 mM $ZnCl_2$, 50 μg/ml of DNA and 25 units/ml of nuclease S1 or mung bean nuclease; incubation was at 37°C for 60 min except where indicated. After the addition of EDTA to 5 – 10 mM and Tris-HCl (pH 8.5) to 30 mM, the DNA was extracted and precipitated. The various buffers (30 mM each) used to study the effect of pH on cleavage by single strand-specific nucleases (Figure 4) were: acetic acid (pH 4.5), acetic acid (pH 5.0), 2-(n-morpholino)-ethane sulfonic acid (pH 6.0), Tris-HCl (pH 7.0) and Tris-HCl (pH 8.0); they were adjusted to the proper pH using Tris-base.

Restriction enzyme digestion

Restriction enzymes from Bethesda Research Laboratories, New England Biolabs, and Promega Biotec were used as directed by the manufacturers.

Preparation of end-labeled DNA fragments

DNA fragments were labeled at their 5′ ends using [γ-^{32}P]ATP and T4 polynucleotide kinase as described earlier (Htun *et al.*, 1984).

5′ end nucleotide analysis

Gel-purified fragments labeled with ^{32}P at the nuclease S1-generated 5′ ends were analyzed for the 5′-terminal nucleotide by digestion to completion with bovine pancreatic DNase I (Sigma) and snake venom phosphodiesterase. Mononucleotides from the digested DNAs were mixed with 2.5 μg each of unlabeled 5′-dNMP marker (P-L Biochemicals) and separated by two dimensional t.l.c., using solvent systems (a) and (c) as described by Silberklang *et al.* (1979).

DNA sequence determinations

DNA sequencing was by the chemical method (Maxam and Gilbert, 1977).

Acknowledgements

We thank Maria Nierras, Philip Rice, Alain Krol, and Frank Johnston for many useful discussions and comments. This work was supported by N.I.H. grant GM-30220 and N.S.F. grant PCM-8309618 to J.E.D. and E.L. and by grants from the Swedish Medical Research Council and the Swedish National Board for Technical Development to U.P.

References

Ares,M., Mangin,M. and Weiner,A.M. (1985) *Mol. Cell. Biol.*, in press.
Busch,H., Reddy,R., Rothblum,L. and Choi,Y.C. (1982) *Annu. Rev. Biochem.*, **51**, 617-654.
Card,C.O., Morris,G.F., Brown,D.T. and Marzluff,W.F. (1982) *Nucleic Acids Res.*, **10**, 7677-7688.
Denison,R.A. and Weiner,A.M. (1982) *Mol. Cell. Biol.*, **2**, 815-828.
Dybvig,K., Clark,C.D., Aliperti,G. and Schlesinger,M.J. (1983) *Nucleic Acids Res.*, **11**, 8495-8508.
Evans,T., Schon,E., Gora-Maslak,G., Patterson,J. and Efstratiadis,A. (1984) *Nucleic Acids Res.*, **12**, 8043-8058.
Gillies,S.D., Folson,V. and Tonegawa,S. (1984) *Nature*, **310**, 594-597.
Glikin,G.C., Gargiulo,G., Rena-Descalzi,L. and Worcel,A. (1983) *Nature*, **303**, 770-774.
Hammarström,K., Westin,G., Bark,C., Zabielski,J. and Pettersson, U. (1984) *J. Mol. Biol.*, **179**, 157-169.
Hentschel,C.C. (1982) *Nature*, **295**, 714-716.
Hofstetter,H., Schambock,A., Van Den Berg,J. and Weissmann,C. (1976) *Biochim. Biophys. Acta*, **454**, 587-591.
Htun,H., Lund,E. and Dahlberg,J.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7288-7292.
Kilpatrick,M.W., Klysik,J., Singleton,C.K., Zarling,D.A., Jovin,T.M., Hanau,L.H., Erlanger,B.F. and Wells,R.D. (1984) *J. Biol. Chem.*, **259**, 7268-7274.
Kowalski,D., Kroeker,W.D. and Laskowski,Sr.,M. (1976) *Biochemistry (Wash)*, **15**, 4457-4463.
Krol,A., Lund,E. and Dahlberg,J.E. (1985) *EMBO J.*, **4**, 1529-1535.
Lilley,D.M.J. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6468-6472.
Lilley,D.M.J. and Kemper,B. (1984) *Cell*, **36**, 413-422.

Lund,E. and Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 2013-2021.
Lund,E., Dahlberg,J.E. and Forbes,D.J. (1984) *Mol. Cell. Biol.*, **4**, 2580-2586.
Mace,H.A.F., Pelham,H.R.B. and Travers,A.A. (1983) *Nature*, **304**, 555-557.
Manser,T. and Gesteland,R.F. (1982) *Cell*, **29**, 257-264.
Mattaj,I.W. and Zeller,R. (1983) *EMBO J.*, **2**, 1883-1891.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
Mirkovitch,J., Mirault,M.-E. and Laemmli,U.K. (1984) *Cell*, **39**, 223-232.
Nickol,J.M. and Felsenfeld,G. (1983) *Cell*, **35**, 467-477.
Ohno,S. and Epplen,J.T. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3391-3395.
Panayotatos,N. and Wells,R.D. (1981) *Nature*, **289**, 466-470.
Ryoji,M. and Worcel,A. (1984) *Cell*, **37**, 21-32.
Schenborn,E.T., Lund,E., Mitchen,J.L. and Dahlberg,J.E. (1985) *Mol. Cell. Biol.*, **5**, 1318-1326.
Schon,E., Evans,T., Welsch,J. and Efstradiadis,A. (1983) *Cell*, **35**, 837-848.
Shen,C.-K.J. (1983) *Nucleic Acids Res.*, **11**, 7899-7910.
Shen,S.-H., Slightom,J.L. and Smithies,O. (1981) *Cell*, **26**, 191-203.
Silberklang,M., Gillum,A.M. and Raj Bhandary,U.L. (1979) *Methods Enzymol.*, **59**, 58-109.
Singleton,C., Keysik,J., Stirdivant,S.M. and Wells,R.D. (1982) *Nature*, **299**, 312-316.
Skuzeski,J.M., Lund,E., Murphy,J.T., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 8345-8352.
Slightom,J.L., Blechl,A.E. and Smithies,O. (1980) *Cell*, **21**, 627-638.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Stoeckert,C.J.,Jr., Collins,F.S. and Weissman,S.M. (1984) *Nucleic Acids Res.*, **12**, 4469-4479.
Van Arsdell,S.W. and Weiner,A.M. (1984) *Mol. Cell. Biol.*, **4**, 492-499.
Villeponteau,B., Lundell,M. and Martinson,H. (1984) *Cell*, **39**, 469-478.
Vogt,V.M. (1973) *Eur. J. Biochem.*, **33**, 192-200.
Westin,G., Zabielski,J., Hammarström,K., Monstein,H.J., Bark,C. and Pettersson,U. (1984a) *Proc. Natl. Acad. Sci. USA*, **81**, 3811-3815.
Westin,G., Lund,E., Murphy,J.T., Pettersson,U. and Dahlberg,J.E. (1984b) *EMBO J.*, **3**, 3295-3302.

Received on 29 April 1985