Nucleic Acids Research, Vol. 20, No. 5 975–982

The initiation accuracy of the SV40 early transcription is determined by the functional domains of two TATA elements

Marc Pauly, Michèle Treger[1], Eric Westhof[1] and Pierre Chambon[2]
Laboratoire de Recherche sur le Cancer et les Maladies du Sang, Zone industrielle Grasbusch, L-3370 Leudelange, Luxembourg, [1]Laboratoire de Cristallographie Biologique du CNRS, Institut de Biologie Moléculaire et Cellulaire, rue René Descartes 15, F-67084 Strasbourg Cédex and [2]Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS/ Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, rue Humann 11, F-67085 Strasbourg Cédex, France

Received January 16, 1992; Accepted January 30, 1992

ABSTRACT

To locate the boundaries of the TATA element in the SV40 early promoter, point mutations have been constructed such as to cover the whole T+A-rich region of the replication origin. The effects of these mutations on the rate of transcription *in vivo* show that this region actually contains two TATA elements I and II, each independently directing the accurate initiation of transcription from a specified set of start sites, EES1 and EES2, respectively. The sequence of TATA element I fits best with the compiled 'consensus' sequence found in eukaryotic gene promoters and is the most efficient in directing transcription initiation. Mutations which improve this fit can still increase the rate of transcription, confirming the theory of a correlation between the nucleotide sequence of a TATA element and its functional efficiency. Moreover, some mutations which simultaneously modify the angle of DNA curvature in the T+A-rich promoter region and the rate of transcription reveal a correlation between DNA bending and transcription initiation.

INTRODUCTION

Over the last years, recombinant DNA technology greatly improved our understanding of the molecular mechanisms underlying the regulation of gene expression. Several DNA elements in prokaryotic and eukaryotic gene promoters were found to be implicated in the control of transcription initiation (1, 2, 3). The compilation of DNA sequences of a large number of eukaryotic promoters revealed a strongly conserved element with the 'consensus' sequence 5'-TATAAA-3' (4). This 'TATA box' or 'Goldberg-Hogness box' is located about 30 nucleotides upstream of the mRNA start site of most protein encoding genes transcribed by RNA polymerase B (or II) (5). Site-directed mutagenesis studies, both *in vivo* and *in vitro*, showed that a TATA element functions as a positioning or selecting element, directing the transcription machinery to initiate transcription accurately on only one or a few defined start sites (5).

In the early promoter of the simian vacuolating virus 40 (SV40), such a DNA element controls the accurate initiation of transcription *in vivo* (6, 7, 8) and *in vitro* (9) from the early-early start sites (EES), at positions 5230 to 5239 (Figure 1), during the early phase of the viral lytic cycle. Located around position 20 in the T+A-rich region of the SV40 replication origin (Figure 1), its sequence 5'-TATTTA-3' is slightly diverging from the consensus sequence. Deletion analyses revealed that this TATA element is dispensable for quantitatively efficient transcription *in vivo* (6, 7, 10) and *in vitro* (9, 11), but that the transcription initiates at multiple, heterogeneous sites in its absence. Two further SV40 promoter elements are, on the contrary, indispensable for efficient transcription, at least *in vivo*: the upstream element or 21 bp repeats (position 37 to 107, Figure 1)(10, 12, 13) and the enhancer or 72 bp repeats (position 107 to 250)(6, 10, 14, 15). Early in infection, large T tumor antigen binds to three domains covering the replication origin and the downstream early promoter, providing a nucleoproteic matrix for the onset of viral DNA replication and simultaneously repressing its own synthesis by a negative autoregulation mechanism (16). As a result, transcription initiation shifts from the EES downstream of the TATA element, to the late-early start sites (LES) upstream of this element (LES1: positions 34, 35; LES2: positions 30, 31; LES3: positions 21, 22)(8, 17, Figure 1). Whereas, before repression, the early TATA element directs selective transcription initiation from the EES, some weak TATA-like substitute elements might account for an accurate transcription initiation from the new sites (6). The early-early and late-early promoters overlap in the 21 bp and 72 bp-repeat regions (8, 12).

In the present study, oligonucleotide site-directed mutagenesis was used to delineate the boundaries of the functional domains of the SV40 early TATA element and to analyse the relationship between its nucleotide sequence and its functional efficiency *in vivo*. As the presence of the T+A-rich sequence causes DNA bending in the SV40 replication origin and early promoter region

(18, 19), we also investigated the relationship between the modifications of DNA curvature induced by these mutations and the rate of transcription initiation.

MATERIALS AND METHODS

Construction of mutant recombinants (Figure 1)

All recombinant plasmids are similar to pSEG0 (12), in that they contain the SV40 early promoter region (from HpaII, position 346, to HindIII, position 5171) inserted upstream from a promoter-less rabbit β-globin gene (position −9 to +1650) and a complete rabbit β-globin gene (including its own promoter region), acting as an internal 'reference' gene. All of them harbour a single 72 bp sequence in the SV40 enhancer region. The A-series and E-series mutant recombinants contain a genetically engineered SalI site at position 32 (13). In the C-series, the SalI site at position 32 is absent, except for pSEG C2, C6 and C10, but all these recombinants harbour the artificial BamHI site created at position 101 (12). Introducing these artificial restriction sites in the SV40 promoter region has no effect on the rate of transcription initiation (12, 13). The A-series and E-series mutants were constructed by shot-gun ligation using synthetic oligodeoxynucleotides (21). The T+A-rich region of the SV40 promoter between the SfiI site (position 0) and the SalI site (position 32) was reconstructed using complementary oligodeoxynucleotides with phosphorylated protruding 5'-ends. One defined set of oligodeoxynucleotides with a given mutated sequence was annealed and ligated for each mutant to be constructed. pSEG A0, the reference 'wild type' recombinant of the A-series and E-series mutants, harbours the SV40 wild type sequence with a single 72 bp enhancer sequence and the artificial SalI site at position 32. In the C-series mutants, pSEG C1 and C3 were constructed using plasmid pSEG (12, 13) and mismatch-primer mutagenesis as described in (21). pSEG C2 was constructed as the A-series mutants (see above). pSEG C4, C5, C6 and C7 with mutated SV40 enhancer sequences were obtained by ligating the purified, small KpnI−BamHI SV40 fragment of recombinant pA93 (15) into the purified, large fragment of pSEG, pSEG C1, C2 and C3, respectively, cut with KpnI and BamHI. The same procedure was followed to yield pSEG C8, C9, C10 and C11 from pA94 (15). Recombinant DNAs were grown in E. coli HB101 recA- bacteria and prepared as described elsewhere (13). All were monitored for possible rearrangements by restriction enzyme analysis and checked for their correct sequence by dideoxynucleotide sequencing of both strands using double-stranded plasmid DNA as template and synthetic oligodeoxynucleotide primers (22).

In vivo transcription assay of recombinants

Transfection of recombinant plasmid DNAs into HeLa cells by the calcium phosphate coprecipitation technique and culture of these cells were performed as described previously (8, 12, 13). Total RNA was extracted 48 hours after transfection as described elsewhere (8). RNA initiated from the SV40 early promoter was measured by quantitative S1 nuclease mapping using two probes as indicated in (12, 13). The amount of RNA initiated from the reference β-globin gene (GLOB band in Figure 2) and the total amount of RNA initiated from the SV40 early promoter region (EP band in Figure 2) was determined with probe A. Probe B was used to accurately map and quantitate RNA initiated from the EES and LES of the SV40 early promoter. Probe A corresponds to the (32P) 5'-end-labelled coding strand of the BstNI fragment of the rabbit β-globin gene (position +138 to −84) and probe B to the (32P) 5'-end-labelled SV40 early coding strand of the HindIII to ClaI fragment of pSEG A0 (for A- and E-series recombinants) or pSEG (for C-series recombinants)(20). A 'homologous' probe was prepared from the corresponding pSEG mutant recombinant. Degradation of the RNAs initiated from the late-early start sites upstream from the mutation site by S1 nuclease, occurring in these cases with the 'heterologous' B probe, could thus be avoided and correct quantitation of these RNAs performed, as for the other RNA species.

DNA bending analysis

Bending analysis by 3 D-modeling was performed on double-stranded DNA fragments extending from nucleotide position 5230 to 42 in the early SV40 promoter region of the different recombinants. An interactive modeling program for DNA was used (23) and the calculated structure displayed on an Evans and Sutherland PS300 graphic station. Sequence-dependent structural parameters were quantitated according with the following model (24): no roll and a twist angle of 36° for 5'-AA-3' steps; a negative roll and a low twist angle for 5'-AT-3' steps (−5° and 32°, respectively); a positive roll and a high twist angle for 5'-TA-3' steps (10° and 40°, respectively). Bending analysis by electrophoretic mobility shift assays were performed as described (19) on double-stranded restriction DNA fragments generated by enzymatic digestion of recombinant plasmid DNA with HindIII (position 5171) and SphI (position 128) in the early SV40 promoter region.

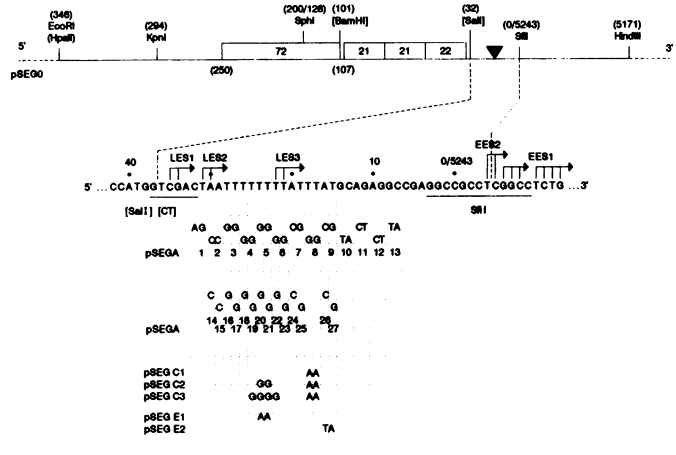

Figure 1. Structural organization and sequences of the SV40 early promoter region in the plasmid recombinants pSEG. The organization of the SV40 early promoter region in recombinant pSEG 0 between the HindIII site (position 5171) and the destroyed HpaII site (in parentheses, position 346) is shown at the top. Some other key natural or engineered (SalI and BamHI, in brakets) restriction sites are indicated with their nucleotide position (according to the BBB numbering system (25)) in parentheses. The location of the TATA box region is featured by a black triangle. The boxed regions symbolize the two 21 bp perfect and the single 22 bp imperfect repeated sequences of the upstream element, and the single 72 bp repeated sequence of the enhancer element contained in these recombinants. The nucleotide sequence of the noncoding strand in the SV40 early TATA region from position 5230 to 42 is detailed below. The locations of the natural SfiI site (position 0/5243) and of the engineered SalI site (position 32) are indicated (underlined). This SalI site is present in the A- and E-series recombinants and in pSEG C2, C6 and C10, the artificial BamHI site (position 101) exists in the C-series recombinants. The positions of the early-early start sites (EES1 and EES2) and of the late-early start sites (LES 1, LES 2 and LES 3) are represented by arrows. The sequence point-mutated in the different mutants is given below that sequence.

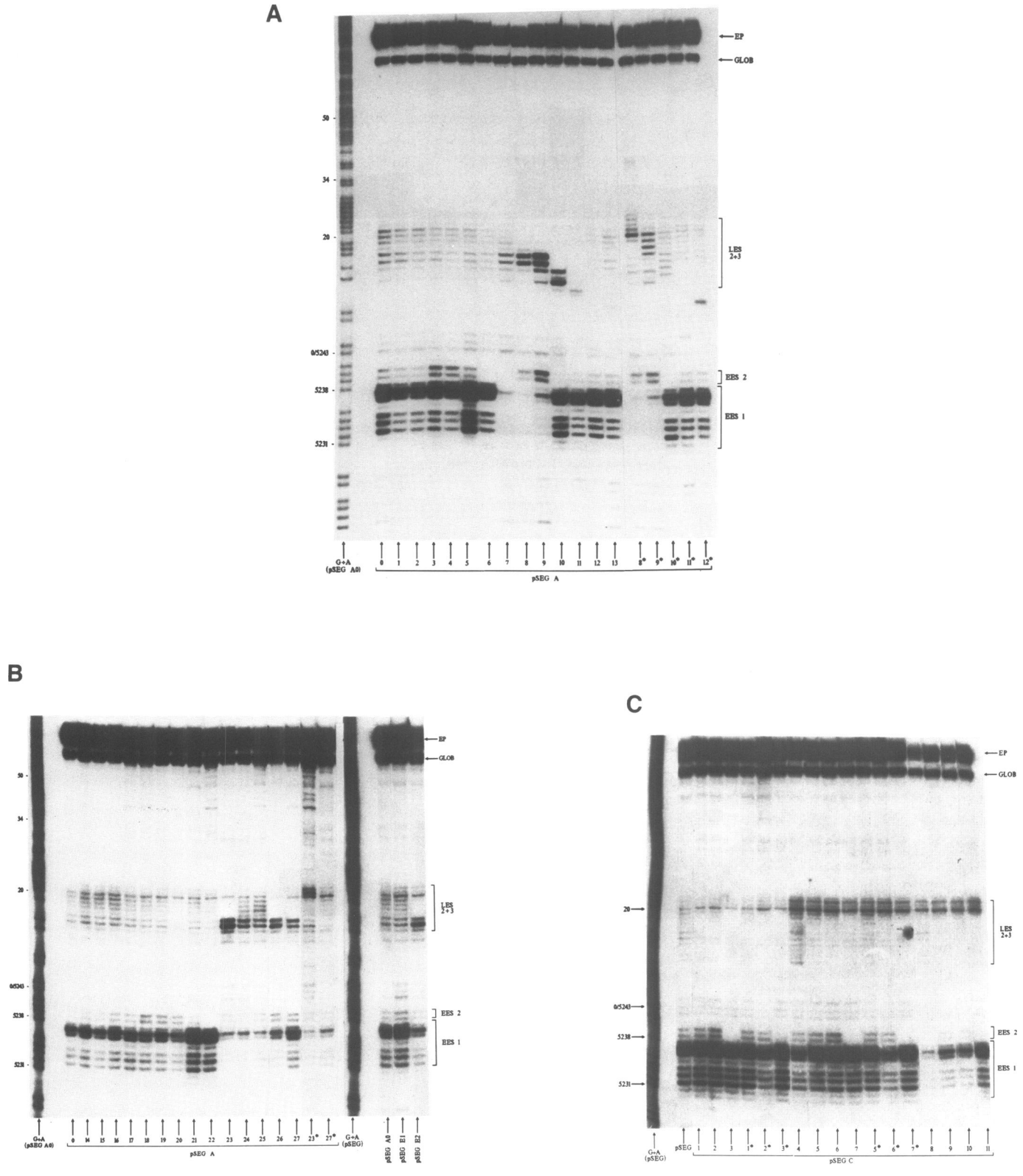

Figure 2. Quantitative S1 nuclease mapping analysis. HeLa cells were transfected with plasmid recombinant DNAs as indicated at the bottom of each lane. 10 μg (80 μg in the case of pSEG C4-C11) of total RNA was analysed by quantitative S1 nuclease mapping as described in Materials and Methods. Electrophoresis of S1-resistant DNA fragments was performed on an 8% polyacrylamid denaturating gel (urea: 8M), autoradiographs of which are shown. EP indicates the S1 nuclease-resistant DNA fragment that arises from the homology between probe A and the SV40-globin hybrid transcripts. GLOB corresponds to the S1 nuclease-resistant DNA fragment of probe A expected for RNA initiated at the start-site of the reference β-globin gene contained in these recombinants. The S1 nuclease-resistant bands EES1, EES2 and LES 2+3 correspond to RNA initiated from the early-early and the late-early start sites, respectively. The lanes labelled (*) correspond to mutants whose RNA was analysed using a 'homologous' B probe, as described in Materials and Methods. Numbers on the left-hand side pointing to the G+A sequence ladder of probe B, treated by base-specific chemical cleavage reactions according to the method of Maxam and Gilbert (26), indicate the SV40 nucleotide positions using the BBB numbering system (25).

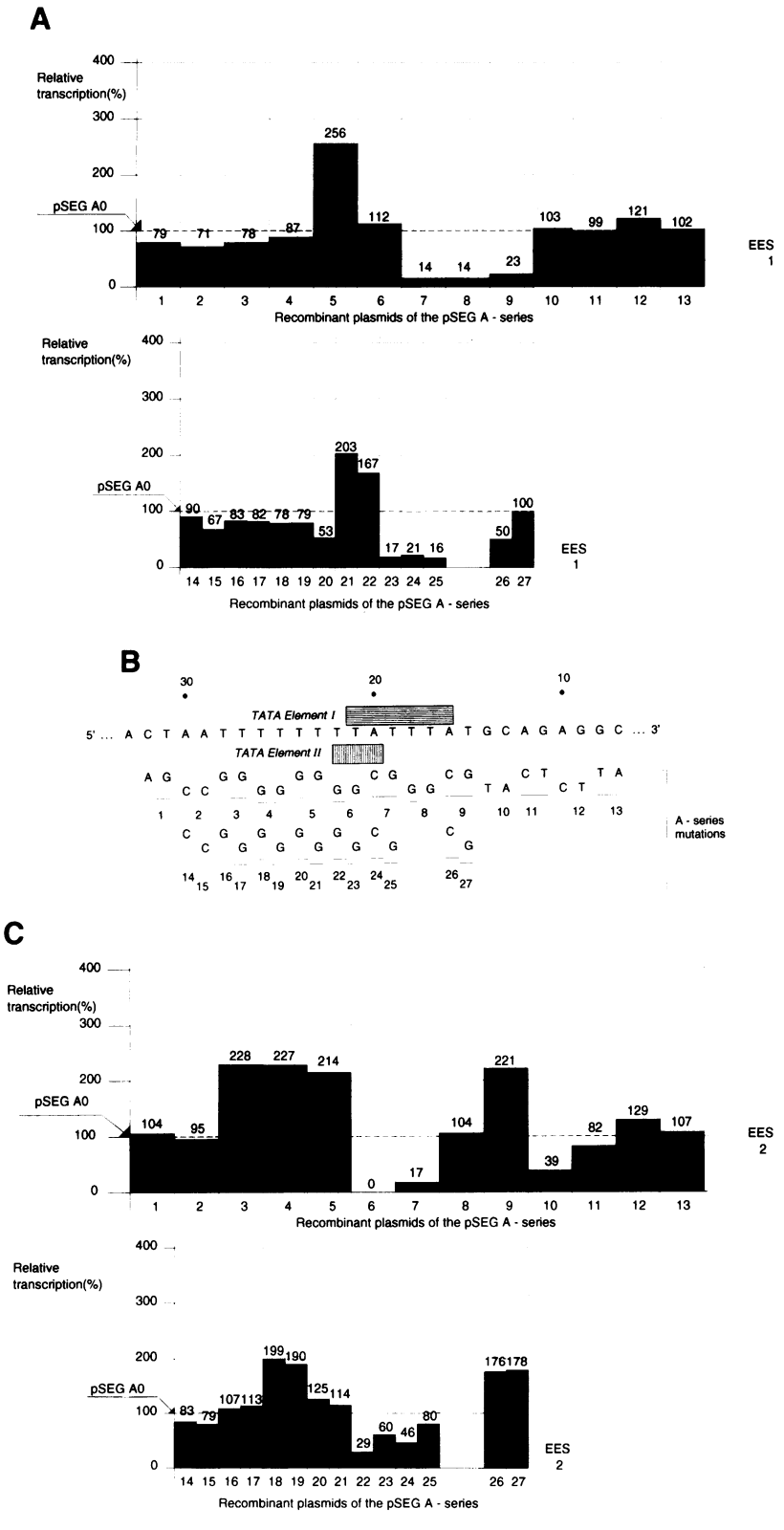

Figure 3. Diagramatic representation of the relative *in vivo* transcription values of the pSEG A-series recombinants and deduction of the location of the functional domains in the SV40 early TATA region. The relative amount of RNA initiated at the EES1 and EES2 after transfection of various recombinants of the pSEG A- and C-series into HeLa cells was estimated by scanning autoradiograms (similar to those shown in Figure 2) from several quantitative S1 nuclease mapping experiments using different plasmid preparations. After correction for transcription from the reference β-globin gene (12, 13), the results were expressed relative to pSEG A0, taken as 100%. The number of the mutant recombinant is indicated in each case on the abscissa and the corresponding relative transcription value is given on the ordinate. Latter value is also shown above each mutant number. The nucleotide sequence of the noncoding strand in the SV40 early TATA region from position 7 to 33 of pSEG A0 and the sequence point-mutated in the various mutant recombinants of the A-series are indicated. The location of the functional domain of the TATA element I directing the EES1 is represented by a horizontally hatched box, that of the functional domain of the TATA II element directing the EES2 by a vertically hatched box.

RESULTS

Effects of point-mutations in the SV40 early TATA region on transcription initiation *in vivo*

Effects of the A-series point-mutations. The A-series recombinants encompass double and single transversional point-mutations covering the whole T+A-rich region of the SV40 early promoter and replication origin (Figure 1). Our transcription assays with these recombinants show that there are two independant groups of EES: a first group of major start sites EES1 (positions 5230−5233, 5235−5237; Figure 1) and a second group of minor start sites EES2, located immediately upstream of the previous ones (positions 5238, 5239). The rate of transcription from the EES2 is only 10% of that from the EES1. Mutations located within a domain extending from position 16 to 21 drastically decrease transcription initiation from the EES1, as compared to 'wild type' recombinant pSEG A0 (Figures 2, 3). We thus concluded that this region is the functional domain of a major TATA element I, directing efficient accurate transcription from the EES1. However, changing Ts into Gs in positions 22 and 23 immediately upstream of this domain strongly stimulates transcription from the EES1. Mutations in a domain extending from position 20 to 22 drastically inhibit the initiation of transcription from the EES2. This region thus appears as the functional domain of a minor TATA element II, playing an important role in controling accurate initiation from the EES2. Both TATA elements overlap at positions 20 and 21, as mutations within these positions decrease both transcription from the EES1 and EES2. Again, changing thymines into guanines in positions 15, 16 downstream and 25, 26 upstream of TATA element II drastically stimulate transcription from the EES2 (Figures 2, 3).

The rate of transcription initiation from the LES2+3 sites is only little affected by the A-series mutations: in most cases, it is not very different from the wild type level (Figure 2). Nevertheless, an increase can be noticed in the case of a few mutants showing a lower rate of transcription from the EES1, as for mutants pSEG A8, A9 and A23, with 137%, 143% and 139%, respectively, of pSEGA0. This increase is probably due, at least in part, to a compensation of the low transcription from the EES by a higher transcription from the LES (8). On the contrary, double point-mutations in positions 9 to 12 induce a decrease of transcription from the LES2+3, as for pSEG A11, A12, with 39% and 44%, respectively, of wild type transcription (Figure 2). Furthermore, there is a preferential use of certain start sites of the LES2+3 group in the case of mutations extending from position 13 to 18 (see pSEG A8, A9, and A10 on Figure 2, homologous probes used). It is however impossible to further investigate these phenomena by the studies presented herein. In general, as a result of the compensation, mutations in the SV40 early TATA region induce no major changes in the overall rate of SV40 early transcription.

Effects of the E-series point-mutations. As outlined above, there are two sites upstream and downstream of the SV40 early TATA region where the change of even one single thymine into a guanine has a drastic up-effect on transcription initiation *in vivo*. In order to further investigate this phenomenon, we constructed mutants pSEG E1 and pSEG E2. In pSEG E1, two adenines were substituted by two thymines in position 23 and 24 on the noncoding strand, while in pSEG E2, the dinucleotide 5′-AT-3′ in position 15 and 16 on the noncoding strand was changed into 5′-TA-3′ (Figure 1). In the case of pSEG E1, the level of transcription from the EES1 and the EES2 is the same as wild type. This observation strongly contrasts with the drastic increase obtained above with mutant pSEG A5. Transcription from the LES2+3 is 76% and about the same as with pSEG A5. In mutant pSEG E2, transcription from the EES1 is 30% and shows no difference, when compared to pSEG A9, but transcription from the EES2 and from the LES2+3 is reduced to 49% and 62%, respectively, while it was increased for pSEG A9. Thus, we concluded that the stimulation of transcription observed in these cases is due to a change in the DNA conformation, rather than to the alteration of a functional domain.

Effects of the C-series point-mutations. The C-series encompass several point-mutations in the major TATA element I in order to analyse the relationship existing between its nucleotide sequence and its function. Mutant recombinant pSEG C1 has a double point mutation changing 5′-TT-3′ into 5′-AA-3′ at positions 18 and 17 (noncoding strand), thus creating a typical consensus sequence of an eukaryotic TATA element (Figure 1). Mutant recombinants pSEG C5 and C9 have the same sequence as pSEG C1, but they harbour, in addition, a defective enhancer element in the SV40 early promoter region, the activity of which was reduced by point-mutations introduced in the functionally relevant domains of this element (15). Therefore, the enhancer activity in pSEG C5 is only 8% of the wild type activity and in pSEG C9, it is only 0,6%. This destruction appeared as necessary to prevent the 'masking' of the effects observed with the C-series mutations by the much stronger enhancer activity (see below). Mutant recombinants pSEG C2 and pSEG C3 have, in addition to the previous consensus sequence, two (position 23, 24) or four (position 22 to 25) guanines, respectively, on the non-coding strand upstream of the TATA sequence, as some other efficient eukaryotic TATA elements. Again, mutant recombinants pSEG C6, C7 and C10, C11 had to be constructed with these same mutations, but with deficient enhancer elements as above.

Table 1. Comparison of the r_L, R^4_L and relative transcription values of the pSEG recombinants. Recombinant numbers are given in the leftmost column. r_L represents the ratio of the length of the double helix to the end-to-end distance and was calculated in each case for the corresponding 3D DNA model. R^4_L represents the electrophoretic mobility shift ratio of double-stranded DNA fragments covering the SV40 early TATA region of the different recombinants. The apparent size of each fragment was deduced from its position on an electrophoretic gel like those shown in Figure 4, by comparison to a size marker. The values of the electrophoretic mobility shift ratio R^4_L were obtained in each case by calculating the ratio of the apparent size to the actual size of the fragment. The relative *in vivo* transcription values from EES1 of the pSEG recombinants are indicated in the rightmost column.

Recombinant	r_L	R^4_L	Relative Transcription from EES1 [%]
pSEG A0/pSEG C0	1.008	1.035	100
pSEG A5	1.021	1.140	256
pSEG A20	1.012	1.100	53
pSEG A21	1.021	1.125	203
pSEG E1	1.014	1.065	90
pSEG A6	1.024	1.100	112
pSEG A22	1.020	1.075	167
pSEG A23	1.016	1.125	17
pSEG A7	1.008	1.035	14
pSEG C1	1.009	1.035	117
pSEG C2	1.020	1.115	201
pSEG C3	1.027	1.140	109

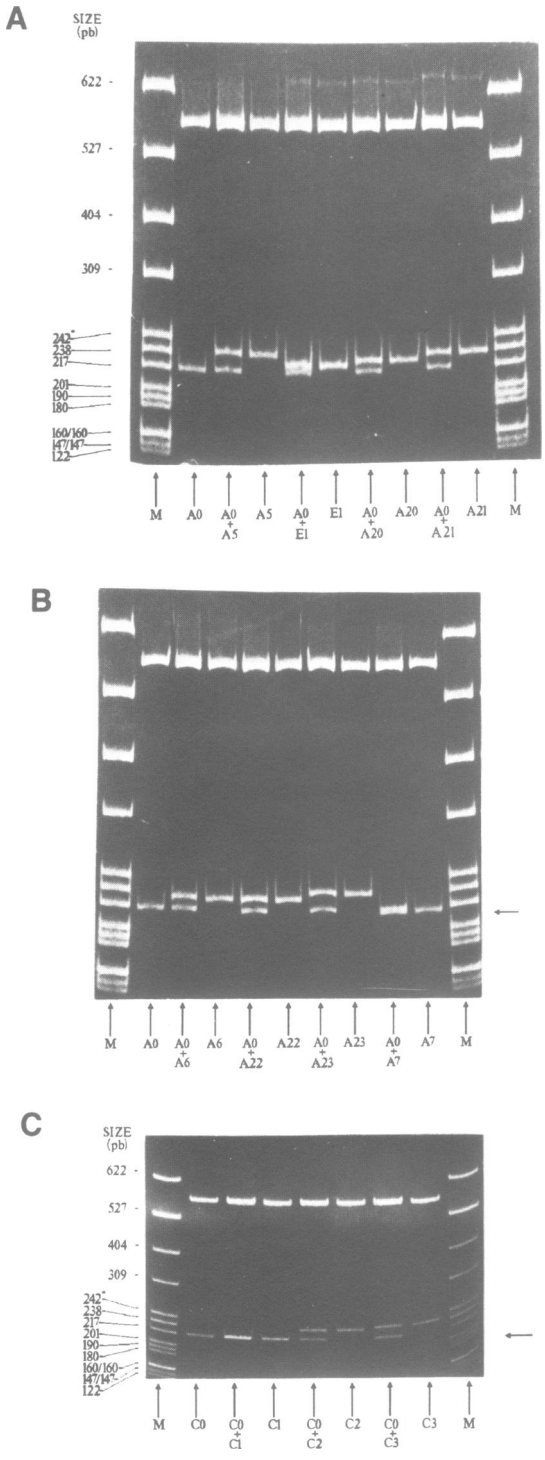

Figure 4. Electrophoretic mobility studies on double-stranded DNA fragments covering the SV40 early TATA region of the pSEG recombinants. 1,5 μg of digested DNA was run on a 12% polyacrylamid gel at 4°C as described in Materials and Methods. Recombinant numbers are indicated at the bottom of each lane (A0: recombinant pSEG A0; A5: mutant recombinant pSEG A5; A0+A5: equimolar mixture of DNA fragments prepared from pSEG A0 and pSEG A5 recombinant digests; etc...). The position of the double-stranded DNA fragment of the pSEG A0 or the pSEG (i.e. pSEG C0) recombinant taken as a 'wild type' reference, is indicated by an arrow on the right-handed side. This fragment has an actual size of 200 bp, but migrates, as a consequence of DNA bending, like an apparently larger fragment of 207 bp. The positions of different size markers (lane M), a digest of pBR322 by HpaII, are indicated in base-pairs on the left-handed side. All double-stranded DNA fragments of this marker are straight, except for the 242 bp fragment, labelled by a star, which shows a slower mobility than expected for its size and is probably bent, as it contains a stretch of six adenines in its sequence (19).

In spite of their consensus sequence, pSEG C1 and C3 do not indeed show an increase of the initiation rate of transcription from the EES1 (Figure 2, Table 1). Only mutant pSEG C2, similar to the previous pSEG A5, exhibits an increase of transcription from EES1. The stimulation of transcription initiation from the EES1, due to the creation of a consensus sequence in the functional domain of TATA I clearly appears the more the enhancer activity is reduced. This is observed with mutants pSEG C5, C9 with an average rate of 193% and 266%, when compared to pSEG C4 and pSEGC8, respectively, taken as reference (Figure 2). pSEG C4 and pSEG C8 harbour the same altered enhancer sequence and there enhancer activity is also the same. The transcription from EES1 is even higher when several guanines are present, in addition, upstream of TATA element I, as in pSEG C6, C7 and C10, C11, with respectively 252%, 194% and 315%, 358%, in average.

Effects of point-mutations on DNA bending in the SV40 early TATA region

To analyze the effects of the E-series mutations on the initiation of transcription and to determine if these and some other mutations could have an effect on DNA conformation in the SV40 early promoter region, we performed a serie of biophysical studies. From previous studies, it was known that the T+A-rich region of the SV40 early promoter exhibited DNA bending (18, 19). The underlying mechanism of this DNA bending is still unknown, but the stretch of adenines immediately upstream of the SV40 early TATA elements seems to be implicated, and their presence would even be the direct cause generating DNA curvature (18).

To examine how DNA bending is affected by point-mutations in the SV40 early TATA region, we first used a theoretical approach by 3D-DNA modeling as described in Materials and Methods. According to the predictions provided by these models, the DNA in the wild type SV40 early promoter region shows a strong net bend which is centered around position 23, immediately upstream of the functional domains of TATA elements I and II. Some of the point-mutations in the SV40 early TATA region increase the bent angle, compared to that of a 'wild type' sequence. The most drastic increase is observed for mutants pSEG A5, A21, A6, A22, C2 and C3, compared to pSEG A0, taken as 'wild type reference' (Table 1). In all these cases, the mutations are in a region located around position 23. The r_L ratio has intermediate values for mutant pSEG A20, for a change of a T into G at position 24 and for pSEG A23, with the change of a T into a G at position 21, on the noncoding strand. The same is true for mutant pSEG E1, showing a moderate increase of DNA curvature compared to pSEG A0 for two Ts changed into two As at positions 23 and 24 of the same strand. Almost no effect is observed with mutants pSEG A7, with a change of a dinucleotide 5'-AT-3' into 5'-CG-3' in positions 19 and 20 of the noncoding strand and pSEG C1 with two Ts changed into two As at positions 17 and 18.

As a second, empirical approach evaluating DNA bending, we used an electrophoretic mobility shift assay as described in Materials and Methods and in Figure 4. Bending of a DNA fragment affects its electrophoretic mobility on a gel (28). The mobility shift of double-stranded DNA fragments covering the SV40 early promoter region of pSEG recombinants was monitored at a constant temperature of 4°C. The R^4_L ratio of the variation of electrophoretic mobility at 4°C was then calculated as indicated in Table 1. This ratio represents a direct, empirical measure of the bent angle. In the case of recombinant

pSEG A0, taken as 'wild type' reference, the R^4_L ratio has a value of 1,035 and is thus greater than 1, which indicates the existence of a DNA curvature already in the 'wild type' SV40 promoter region (Table 1 and Figure 4). As expected from above, a striking increase of the R^4_L ratio values is observed for mutations within positions 21 to 24 as with pSEG A5, A21, A23 and pSEG C2, C3, compared to pSEG A0. Mutants pSEG E1, A6, A20 and A22 show moderate increases of the R^4_L ratios. Again, and as also shown in (18), we noticed that mutations further downstream or upstream of position 23, and especially beyond the polyadenylic sequence, induce no major modification of DNA curvature. This is the case for pSEG, pSEG A7 and pSEG C1 (Table 1), pSEG A9, E2, as well as pSEG A4, A18, A19 (not shown). As there is no difference between pSEG and pSEG A0, the creation of the SalI site at position 32 in the latter has no influence on DNA bending of the SV40 early promoter. From these results, one can conclude and confirm that the DNA sequence directing the curvature in this region is the polyadenylic stretch extending from positions 21 to 28 upstream of the functional domains of the SV40 early TATA elements I and II and that positions 21 to 24 are the most relevant sites in this view. A comparison of the R^4_L ratios to the previously calculated r_L ratios shows a high correlation between the theoretical and the empirical data reflecting DNA bending, as obtained in our study. A linear regression calculation provided a correlation coefficient of 0,81.

In the case of mutants pSEG A5, A21, A22 and E1, we found a correlationship between the bend angle and the efficiency of transcription initiation *in vivo* (Table 1, Figure 5). Mutations in a region extending from position 22 to 24 upstream of the functional domains of the SV40 early TATA elements have a dramatic effect on both the net DNA curvature of that region and the rate of transcription initiation from the major EES1 sites. Thus, changing thymines in guanines on the noncoding strand in positions 23 and 24, essentially, induces a strong increase in the net DNA bend of the SV40 early promoter region and drastically stimulates, in parallel, the transcription initiation from the EES1. Changing the two Ts in positions 23 and 24 into two As, as for pSEG E1, induces a far lesser increase of the DNA bend and a transcription rate from the EES1 which is about that of the wild type. In the cases of the other mutants, or for transcription from EES2 and LES2+3, it was impossible to deduce a clear relationship between DNA bending and transcription efficiency, as the transcription initiation was simultaneously influenced by other factors, mainly by the alteration of the nucleotide sequence *per se*.

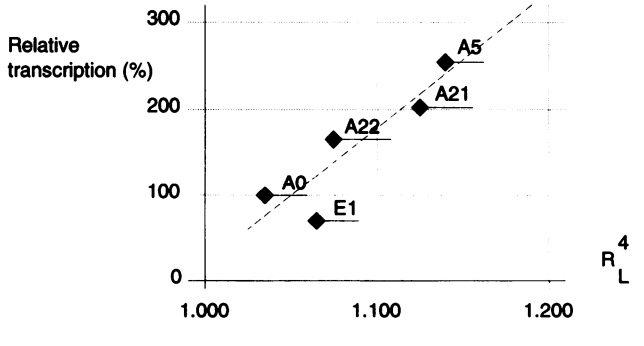

Figure 5. Correlation between the rate of SV40 early trancription from the EES1 and the DNA net bent angle. Taken from table 1, R^4_L ratios are given on the abscissa, relative *in vivo* transcription values on the ordinate. Relevant recombinant numbers are indicated on the diagram.

DISCUSSION

Structural and functional organization of the TATA region in the SV40 early promoter

According to previous *in vivo* studies (6, 7, 8), our analysis by site-directed mutagenesis confirms the functional role of the SV40 early TATA sequence as that of a 'start site selector element', which correctly positions and directs the transcription machinery to preferentially and accurately initiate at the viral early-early start sites. Furthermore, as shown by our results, there are two TATA elements located in the SV40 early T+A-rich region, each controling initiation of transcription from two, independent sets of early-early start sites. Thus, TATA element I directs transcription from a major set of EES1, and pseudo- TATA element II from a minor set of EES2. We could locate the relevant functional domains of these TATA elements to position 16 to 21 for TATA element I and position 20 to 22 for TATA element II. The functional domains of TATA element I and II overlap at position 20 and 21.

An important feature of a TATA element is its nucleotide sequence. This is emphasized by mutant recombinants of the C-series which show that the functional efficiency *in vivo* of the SV40 early TATA element I can be improved by introducing a nucleotide sequence closer to the consensus into its functional domain. Such an improvement of the functional efficiency could also be induced in the case of the SV40 late TATA element. Converting the sequence 3'-TACCTAA-5' (position 296 to 302) to sequences closer to the consensus sequence results in a strong increase of transcription initiation from the major late site at position 325, at least *in vitro* (27).

Relationship between DNA bending and the initiation of SV40 early transcription *in vivo*

Several recent investigations showed that some nucleotide sequences can induce a DNA curvature and that DNA fragments harbouring such a structure exhibit an abnormal mobility on electrophoresis gels (28, 29, 30, 34). In all cases has a stretch of adenines been discovered, mostly repeated periodically on the same side of the DNA helix, and most probably responsible for inducing an overall bent by deviation of the normal DNA axis. The exact mechanism of this bending has yet remained unclear, but several models have since been proposed which try to give an explanation to this biophysical phenomenon (29, 34).

In the case of simian virus SV40, it was shown that there is a relationship existing between the initiation of viral DNA replication *in vivo* and DNA bending in the SV40 early promoter region, which also serves as replication origin (18). Different point-mutations in the T+A-rich region affect the angle of the net bend and change, in parallel, the efficiency of viral replication. Possibly, a correct and properly localized curved conformation in the DNA of the replication origin region could be recognized as a specific structural binding signal by those protein factors implicated in the replication process (18). The stretch of eight As, in which the mutations were introduced in these cases, and extending from position 21 to 28, was thought to be responsible for inducing this DNA bending (18, 19). Some reports showed that sequence distributions associated with DNA curvature are found upstream of strong promoters (31), and that a loss of curvature can lead to a nearly complete loss of promoter activity

(32), suggesting that DNA bending might play an important role in regulation mechanisms of transcription.

We examined this possibility in the present study, knowing that the T+A-region of the core replication origin of SV40 is an important part of the viral early promoter, as it harbours the TATA elements of this promoter. Moreover, a LOB protein which specifically binds and induces or amplifies a DNA bend in the T+A-rich late origin region has been isolated and thought to play a role either in replication and/or transcription initiation (33). Possibly, this bend could be required for an appropriate binding of this protein to the DNA and for fulfilling its regulatory function. Our results confirm the existence of a sequence-directed DNA bending in the T+A-rich region, centered around position 23, and directed by the polyadenylic sequence from positions 21 to 28. Furthermore, they suggest that there is, in some cases at least, a correlation between DNA bending and the efficiency of transcription initiation. Mutations in this polyadenylic sequence induce a variation of the angle of the DNA net bend in that region and of the *in vivo* transcription rate from the EES1 sites as well. Thus, the increase of the DNA curvature existing in the SV40 early promoter region might stimulate early transcription initiation from these sites. One of the biological roles of DNA bending in the control of the transcription initiation could consist in facilitating the binding of transcription factors to DNA in the vicinity of promoter elements. The curved DNA could be recognized as a special signal for binding, and the formation of a stable nucleoproteic complex could be improved by wrapping of the proteins in a bent DNA structure.

Moreover, the existence of a curved DNA could facilate the denaturation process and 'opening' of the DNA double helix by inducing a 'torsional stress'. It has also been shown that proteins stimulating transcription induce a stronger DNA curvature at their fixation site (34). Finally, an increased DNA bend could also bring in close vicinity, in a 'loop conformation', two or several important DNA promoter regions, which are normally separated over long distances. Thus, they could interact for binding of crucial transcription factors.

In all cases, the results of the present studies show that the biochemical and biophysical properties of the TATA sequence of the SV40 early promoter region interact in a most complex manner to induce an accurate and efficient transcription initiation from the viral early-early start sites.

ACKNOWLEDGEMENTS

We would like to thank M. Wintzerith, A. Staub and H. Matthes for useful discussions and advice about oligodeoxynucleotide synthesis and M13 bacteriophage cloning techniques, as well as H. Barrera-Saldana, K. Takahashi and J. Perez-Mutul about RNA preparation and S1 nuclease mapping techniques. We would also like to thank H. Barrera-Saldana for the generous gift of the pSEG and pSEG0 recombinants and M. Zenke of the pA93 and pA94 recombinants. Further, we would like to thank M. Acker and her staff for preparations during cell culture and G. Pauly and B. Boulay for help in illustrations. This work was supported by the CNRS (ATP 3582), INSERM (PRC 124026), the Fondation pour la Recherche Médicale, the Ministère de la Recherche et de la Technologie (82V1283) and the Association pour le Développement de la Recherche sur le Cancer. M. Pauly was recipient of fellowships from the Faculté de Médecine, Université Louis Pasteur in Strasbourg and the Ministère de l'Education Nationale et de la Jeunesse in Luxembourg.

REFERENCES

1. Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A. (1985) *Annual Reviews Genet.*, **19**, 355–387.
2. Wasylyk, B. (1985) In Reeck, G.R., Goodwin, G.H., Puigdomènech, P. (ed.), Chromosomal proteins and gene expression. Plenum Press, New York, London, pp. 103–119.
3. Wasylyk, B. (1988) *CRC Critical Reviews Biochem.*, **23**, 77–120.
4. Bucher, P. and Trifonov, E.N. (1986) *Nucleic Acids Res.*, **14**, 10009–10026.
5. Breathnach, R. and Chambon, P. (1981) *Annual Reviews Biochem.*, 50, 349–383.
6. Benoist, C. and Chambon, P. (1981) *Nature*, **290**, 304–310.
7. Ghosh, P.K., Lebowitz, P., Frisque, R.J. and Gluzman, Y. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 100–104.
8. Wasylyk, B., Wasylyk, C., Matthes, H., Wintzerith, M. and Chambon, P. (1983) *EMBO J.*, **2**, 1605–1611.
9. Mathis, D.J. and Chambon, P. (1981) *Nature*, **290**, 310–315.
10. Fromm, M. and Berg, P. (1983a) *J. Mol. Appl. Genet..*, **2**, 127–135.
11. Lebowitz, P. and Ghosh, P.K. (1982). *J. Virol.*, **41**, 449–461.
12. Baty, D., Barrera-Saldana, H.A., Everett, R.D., Vigneron, M. and Chambon, P. (1984) *Nucleic Acids Res.*, **12**, 915–932.
13. Barrera-Saldana, H., Takahashi, K., Vigneron, M., Wildeman, A., Davidson, I. and Chambon, P. (1985) *EMBO J.*, **4**, 3839–3849.
14. Fromm, M. and Berg, P. (1983b) *J. Mol. Appl. Genet.*, **2**, 127–135.
15. Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) *EMBO J*, **5**, 387–397.
16. Myers, R.M., Rio, D.C., Robbins, A.K. and Tjian, R. (1981) *Cell*, **25**, 373–384.
17. Hansen, U., Tenen, D.G., Livingston, D.M. and Sharp, P.A. (1981) *Cell*, **27**, 603–612.
18. Deb, S., DeLucia, A.L., Koff, A., Tsui, S. and Tegtmeyer, P. (1986) *Mol. Cell. Biol.*, **6**, 4578–4584.
19. Hertz, G.Z., Young, M.R. and Mertz, J.E. (1987) *J. Virol.*, **61**, 2322–2325.
20. Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, W.M. and Chambon, P. (1986) *Nature*, **319**, 121–126.
21. Grundström, T., Zenke, W.M., Wintzerith, M., Matthes, H.W.D., Staub, A. and Chambon, P. (1985) *Nucleic Acids Res.*, **13**, 3305–3316.
22. Wallace, R.B., Johnson, M.J., Suggs, S.V., Miyoshi, K., Bhatt, R. and Itakura, K. (1981) *Gene*, **16**, 21–26.
23. Treger, M. and Westhof, E. (1987) *J. Mol. Graph.*, **5**, 178–183.
24. Westhof, E. and Beveridge, D.L. (1990) *Water Science Reviews*, **5**, 24–186.
25. Tooze, J. (ed.) (1982) DNA tumor viruses. Cold Spring Harbor Laboratory Press, New York.
26. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.
27. Nandi, A., Das, G. and Salzman, N.P. (1985) *Mol. Cell. Biol.*, **5**, 591–594.
28. Wu, H.-M. and Crothers, D.M. (1984) *Nature*, **308**, 509–513.
29. Lilley, D. (1986) *Nature*, **320**, 487–488.
30. Travers, A. (1989) *Nature*, **341**, 184–185.
31. Plaskon, R.R. and Wartell, R.M. (1987) *Nucleic Acids Res.*, **15**, 785–796.
32. Collis, C.M., Molloy, P.L., Both, G.W. and Drew, H.R. (1989) *Nucleic Acids Res.*, **17**, 9447–9468.
33. Baur, C.-P. and Knippers, R. (1988) *J. Mol. Biol.*, **203**, 1009–1019.
34. Travers, A.A. (1990) *Cell*, **60**, 177–180.