Nucleic Acids Research, Vol. 20, No. 21 5713–5718

Intramolecular triplex potential sequence within a gene down regulates its expression *in vivo*

Partha S.Sarkar[1] and Samir K.Brahmachari[1,2,*]
[1]Molecular Biophysics Unit and [2]Centre for Genetic Engineering, Indian Institute of Science, Bangalore 560 012, India

Received July 15, 1992; Revised and Accepted October 6, 1992

ABSTRACT

Polypurine/polypyrimidine sequences have been shown to adopt intramolecular triple helix structures under torsional stress and/or at low pH. Such sequences have been observed within the the regulatory as well as the coding regions of several genes and the involvement of triple helical structure adopted by these sequences in transcriptional control has heen speculated. Taking advantage of codon degeneracy we have engineered a 38 bp long intramolecular triple helix potential polypurine/ polypyrimidine sequence motif between the 37th and 50th codons of β-galactosidase gene in the plasmid pBluescriptIISK+ to investigate whether *in vivo E.coli* RNA polymerase would transcribe sequence motifs adopting triple helix structure, when present within the coding region of the gene. *E.coli* JM109 cells transformed with this construct pSBT1, exhibited 80% inhibition of β-galactosidase expression compared to another construct pSBmT12 made using less preferred codons for identical amino acid sequence, but lacking the polypurine/polypyrimidine sequence motif. Truncated β-galactosidase transcripts were observed for pSBT1 but not for pSBmT12. Here we report that a putative triple helix potential sequence within a gene can down regulate its expression by partially blocking the transcription elongation *in vivo*.

INTRODUCTION

In recent years evidences have accumulated which suggest that altered DNA structures could significantly influence RNA polymerase ensemble and template DNA interaction resulting in altered level of gene expression (1–4). Defined sequences within a gene adopting left handed Z-DNA (5) and cruciform structure (6,7) have been shown to block transcription elongation *in vitro*. Insertion of a sequence with a potential to extrude into cruciform structure within the promoter region of tet^R gene was found to inhibit transcription initiation *in vivo* as well (8). A DNA sequence capable of adopting intrinsic bend structure has been shown to functionally replace a transcriptional activator known to induce a bend at its binding locus in *E.coli* (9).

Polypurine/polypyrimidine sequences have been observed in the vicinity of promoters and within the coding regions of several genes such as *c-myc*, chicken α2(I) collagen, *Drosophila hsp 26* and *hsp 70*, SV 40 late genes and β-globin genes (2–4,10–13). Several evidences indicate that such sequences under superhelical stress and/or at low pH can adopt intramolecular triple stranded DNA structure where half of the pyrimidine strand is unpaired and forms Hoogsteen base pairs with the remaining duplex and the other half of the purine strand remains in an unstructured single stranded conformation (2–4,14,15). Formation of intramolecular triple helical structure with purine.purine. pyrimidine hydrogen bonding in the base triad has also been reported for poly(dG).poly(dG).poly(dC) sequences at pH range of 5–8 and in the presence of bivalent metal ions such as Mg^{2+} and Zn^{2+} *in vitro* (16,17). The formation of intramolecular triple helix structure has been studied extensively by pH dependent two dimensional chloroquine gel electrophoresis in the presence or absence of bivalent metal ions *in vitro* (16,18). The role of polypurine/polypyrimidine sequences potential to adopt triple stranded DNA structure in various biological processes has been speculated (2–4) but the precise role and mechanism by which these sequences function are poorly understood. Recently triple helix structure has attracted considerable attention because of its potential use as antisense DNA (19) and as a tool for chromosome mapping (20). Upstream poly purine/polypyrimidine sequences with potential to adopt intramolecular triple helix structure have been implicated to play an important role in the transcription of various genes. The −125 region of c-*myc* gene has been shown to adopt intramolecular triple helix structure *in vitro* the deletion of which significantly reduced the expression of the *c-myc* gene (21,22).

Triplex forming oligonucleotides have been shown to bind upstream as well as downstream of promoters and inhibit transcription of genes *in vitro* (22–25). Such targeted oligonucleotides were found to recognize the poly(pu)/poly(py) sequence regions in the vicinity of promoters of *c-myc* gene and inhibit transcription of the gene in a cell free system (25) and also in HeLa cells, following uptake by the nuclei (26). Poly(dG)/poly(dC) sequences when present upstream of a promoter have been observed to augment the expression of the gene in eukaryotic systems (27). Upstream poly(pu)/poly(py)

* To whom correspondence should be addressed

sequences capable of adopting triplex structure have been found to be critical for the transcription of *Drosophila hsp26 gene* (28). However, no systematic study has been undertaken to investigate whether such sequences with a potential to form triple helix structure present within genes have any regulatory role in transcription elongation *in vivo* although such sequence motifs have been observed within the coding regions of various genes (2−4,29).

To delineate the role of sequences with a potential to adopt different unusual structures in transcription elongation control, we have developed a method for generating and introducing structural cassettes within a gene by exploiting codon degeneracy of different amino acids. This approach was used to investigate the regulatory role of such defined sequences with unusual secondary structural motifs, if any, in transcription elongation control *in vivo* (30). Earlier using this approach we have shown that sequences which have the potential to adopt cruciform structure could act as transcriptional block *in vivo* (31). Using a similar approach we have been able to synthesise and introduce a stretch of polypurine/ polypyrimidine sequence with mirror symmetry, with a potential to form intramolecular triple helix structure within the β-galactosidase gene in plasmid pBluescriptIISK+. The work reported here demonstrates that the presence of triplex potential sequences can down regulate gene expression when present within a gene and the down regulation occurs at the transcription elongation process.

MATERIALS AND METHODS

DNA duplexes

Individual strand of duplex T1 and duplex mT12 (fig 1a) were synthesised on Pharmacia GA plus automated DNA synthesizer using phosphoramidite chemistry and purified on 16% denaturing polyacrylamide gel containing 7M urea. The gel eluted oligonucleotides were desalted on NAP columns (Pharmacia).The complimentary strands were mixed together in equimolar ratios and annealed by heating at 90°C for 3 minutes and slowly cooling to room temperature in presence of 50 mM NaCl. The oligoduplexes were phosphorylated with dATP and T4 polynucleotide kinase (Bangalore Genei,India). The phosphorylated duplexes were desalted by passing through sephadex G-15 spun columns.

Plasmids

The plasmid pBluescriptIISK+ was digested with BamHI and ApaI to generate BamHI-ApaI sticky ends. The larger linear fragment was electroeluted from agarose gel, cleaned by phenol-chloroform extraction and ethanol precipitation and dephosphorylated with calf intestinal phosphatatase (CIP) (Pharmacia). The phosphorylated oligoduplexes were ligated with dephosphorylated pBluescriptIISK+ BamH1-Apa1 vector with T4 DNA ligase (Bangalore Genei,India). Ligation was done at 16°C for 12 hours with a vector to insert ratio of 2:1. All restriction enzymes are from New England Biolabs. Duplex T1 (fig 1) was cloned into the BamHI-ApaI site of pBluescriptIISK+ vector to construct pSBT1. Similarly duplex mT12 (fig 1) was cloned into the same vector to construct pSBmT12. All the plasmids for structural analysis were prepared by alkali lysis procedure and purified by centrifugation to equilibrium in cesium chloride-ethidium bromide gradient. Sequencing of the plasmids were done using dideoxy chain termination method to confirm that all the plasmids carried the correct sequence.

Whole cell β-galactosidase assay

Whole cell β-galactosidase activity was assayed as described by Miller (32). In brief, the *E.coli* JM109 cells carrying plasmids pSBT1 pSBmT12 and pBluescriptIISK+ were grown for 5 hours and IPTG was added to a final concentration of 2mM to induce the β-galactosidase gene. Aliquots of the culture were taken at different time intervals after IPTG induction (i.e. zero time) till the cell density reached 1.6(OD 660). The culture was shaken vigorously after adding a few drops of toluene to permeabilize the cells and then β-galactosidase activity of each aliquot was assayed using ONPG as the substrate. The reaction mixture was incubated at 37°C for 15 minutes and 1M sodium carbonate solution was added to stop the reaction. The optical density of each aliquot at 420 nm, 550 nm and 600 nm were measured. The β-galactosidase activity was calculated as described previously (30,31) using the following equation and OD_{420} was plotted against time after normalizing the cell density for all the cases.

$$\text{Activity} = 1000 \times (OD_{420} - 1.75 \times OD_{550})/t.v.OD_{600}$$

t = time of reaction in minutes.

v = volume of culture used in the assay in ml.

The template plasmid concentration in each of the aliquot was found to be identical. The results shown here are average values of at least three sets of experiments.

2D chloroquine gel electrophoresis

Topoisomers of constructs pSBT1, pSBmT12 and pBluescript IISK+ were prepared as described previously (33). In brief, the plasmids were first incubated with different concentrations of EtBr (5−100 μM) and then relaxing with wheat germ topoisomerase I. EtBr was removed by phenol-chloroform extraction. Topoisomers of the plasmids were mixed and analysed on 1.6% agarose gel. The first dimension was run in 20mM Tris-acetate of different pH as indicated in fig 4 both in presence and absence of 2mM Mg^{2+}. After the first dimension the gel was equilibrated in 40mM Tris-acetate (pH 8), 1mM EDTA containing 1.3 μg/ml chloroquine. The second dimension was run in 40 mM tris acetate buffer (pH 8), 1mM EDTA containing 1.3 μg/ml chloroquine and perpendicular to the first dimension. Chloroquine was removed by washing the gel in tris acetate buffer and stained with EtBr.

Northern blotting of β-galactosidase transcripts with radiolabelled lacZ′

E.coli JM 109 cells harbouring pSBT1, pSBmT12 and pBluescriptIISK+ were grown and β-galactosidase gene was induced by IPTG. Total RNA from the cells was isolated by hot phenol method (when the cell density reached 0.8−0.9) (34). Equal amount of RNA was loaded on 5% denaturing polyacrylamide gel and electrophoresed to resolve the various mRNA products. The RNA was transferred to nylon membrane (SIGMA) and UV cross-linked with the membrane. The plasmid pBluescriptIISK+ was digested with PvuII and the 448 nucleotide long lacZ′ fragment was gel eluted, nick translated with [α-32 P] dATP and used to probe the β-galactosidase transcripts. Hybridization and washing of the membrane was done as described in Sambrook et al (35) and the membrane was exposed to X-ray film for autoradiography. HinfI digest of pUC19 was radiolabelled and used as size marker.

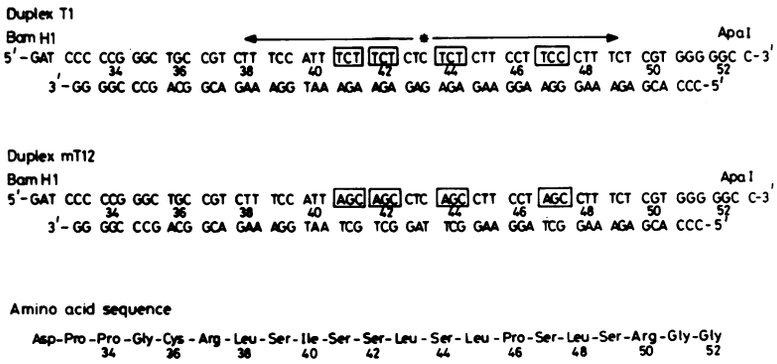

Figure 1. The synthetic structural cassettes used to investigate the role of poly(pu)/poly(py) sequences in the regulation of gene expression *in vivo*. (a) The duplex T1 is the degenerate sequence, generated by the computer program containing 38 bp poly(pu)/poly(py) sequence with mirror symmetry potential to adopt intramolecular triple helix structure. (b) The duplex mT12 is the variant of duplex T1 where triplets at position 41,42,44 and 47 which code for serine are substituted by the triplet AGC which break the poly(pu)/poly(py) motif and the mirror symmetry of the sequence. The triplets boxed in duplex T1 are replaced by triplets AGC in duplex mT12. Individual strands of each duplex were synthesised and annealed to generate duplexes T1 and mT12 and designed in such a way to generate BamHI and ApaI sticky ends for cloning at the same sites in plasmid pBluescriptIISK+.

RESULTS AND DISCUSSION

Strategy to design genes carrying triplex potential structural cassettes and *in vivo* expression study

We developed a computer program 'Gencheck' to design structural cassettes with a potential to adopt unusual DNA structures and to introduce such structural cassettes within genes in order to study their effect on the gene expression *in vivo*. This program takes a DNA sequence as input and generates all possible degenerate sequences taking advantage of codon degeneracy thus keeping the amino acid sequence unaltered. It also searches for the presence of structural motifs using specific knowledge based sequence criteria for unusual DNA structures. While looking for triple helix potential sequences from the pool of degenerate sequences the program also reports for the presence of triplex potential sequences with mismatches in the base triads if there is no perfect poly(pu)/poly(py) stretch with a mirror symmetry in the generated sequences. Duplex T1 is such a degenerate sequence (fig 1) generated by the program taking the BamHI-ApaI fragment of plasmid pBluescriptIISK+ as the input sequence corresponding to 31st to 51st amino acid of the β-galactosidase gene while searching for the presence of sequences with the potential to adopt intramolecular triplex structure. The duplex T1 has a stretch of 38 bp poly(pu)/poly(py) sequence with a mirror symmetry (indicated by *). The synthetic duplex T1 can adopt intramolecular triplex structure in supercoiled plasmid where either the pyrimidine strand or the purine strand can fold back into the major groove of the remaining duplex DNA to form pyrimidine.pyrimidine.purine tupe (fig 2a) or purine.purine.pyrimidine type (fig 2b) of triplex structure respectively. To obtain a required sequence length with a mirror symmetry it became necessary to change amino acids of β-galactosidase gene fragment at position 38,45 and 48 from Asn,Ileu and Thr to leu. Since it is known that at least the first 27 amino acids of the β-galactosidase gene are quite flexible and can be replaced by a wide variety of amino acids without affecting the activity of the β-galactosidase gene (36, 37), this region of the β-galactosidase gene was selected for introducing triplex potential sequence by replacing the BamHI-ApaI fragment with

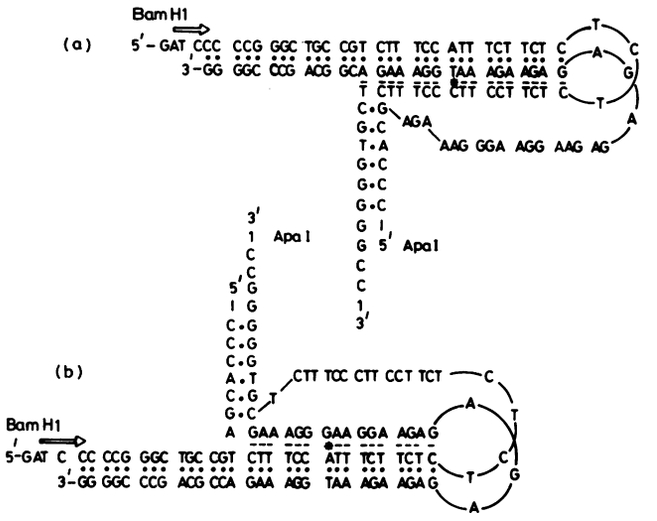

Figure 2. The two possible intramolecular triple helix structure which the synthetic duplex T1 can adopt in supercoiled plasmid pSBT1. The purine and pyrimidine motifs are drawn according to the proposed models for poly(pu)/poly(py) sequences. Watson–Crick base pairing and Hoogsteen base pairing are shown by (...) and (– – –) respectively. The * indicates the mismatch in base pairing in the base triad formation. The arrow indicates the direction of transcription (a) Pyrimidine motif of intramolecular triplex structure. (b) The purine motif of intramolecular triplex structure. In both the cases the more stable isomer is shown.

the synthetic duplex T1 maintaining the proper reading frame of the β-galactosidase gene. In designing the required sequence, codon usage was also taken care of to select the codons which are very frequently used by the *E.coli* mRNAs (38). *E.coli* mRNAs show considerable codon bias like any other organism, where some codons are prefered over other codons. For example, in highly expressed genes the usage frequency for the codons UCU, UCC, AGC, UCG, AGU and UCA that code for serine are 0.361, 0.338, 0.191, 0.046, 0.038 and 0.023 respectively (38). It is known that the translation efficiency of a given gene goes down if poorer codons are used. In designing duplex T1

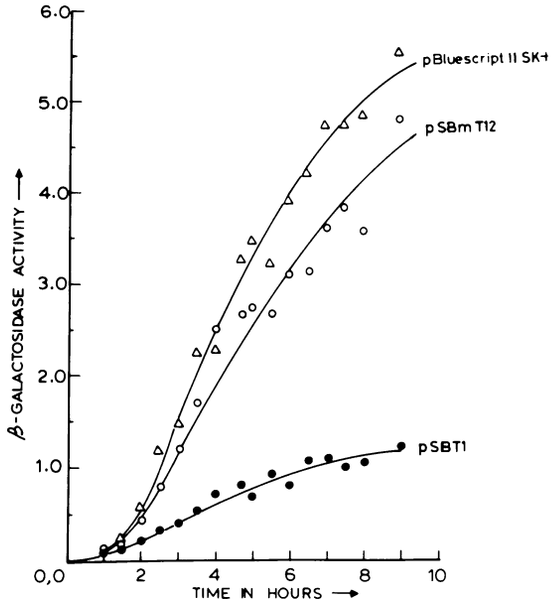

Figure 3. *In vivo* assay of β-galactosidase activity of *E.coli* cells harbouring the construct pSBT1 (●), pSBmT12 (○) *and pBluescriptIISK+* (△). β-galactosidase activity (normalized OD_{420}) is plotted against time in hours.

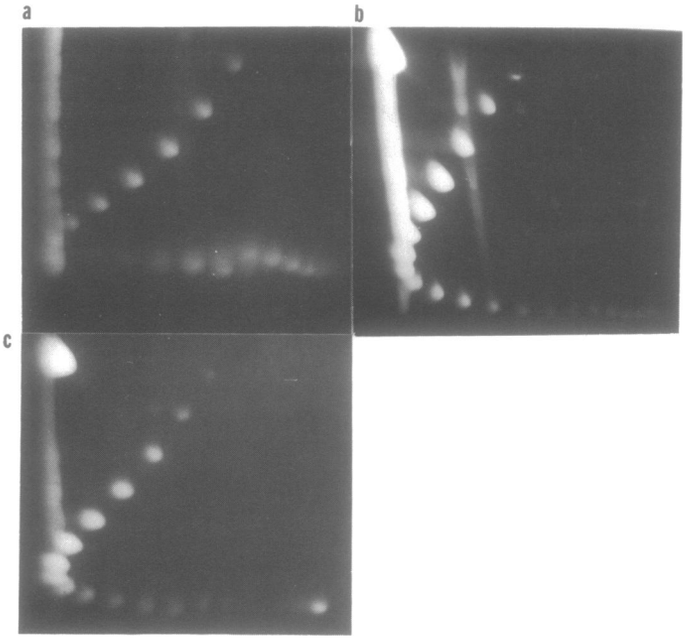

Figure 4. 2D chloroquine gel topoisomer profile for the constructs (**a**) pSBT1 at pH 5 in the presence of 2mM Mg^{+2} ion where a break in topoisomer profile indicates the presence of unwound structure in supercoiled plasmid at low pH (**b**) pSBT1 at pH 5 in the absence of Mg^{+2} ion where the similar break in topoisomer profile is observed (**c**) pSBmT12 at pH 5 and in the presence of 2 mM Mg^{+2} ions which does not show any such break in topoisomer profile under similar experimental conditions.

four UCU, one UCA and one AGC in the parent plasmid coding for six serines ısequence were changed to four UCU and two UCC codons respectively as they were known to have a higher usage frequency.

The plasmid pSBT1 carrying the triplex potential sequence motif was transformed into *E.coli* JM109 cells to study the influence of intramolecular triplex potential sequences on gene expression *in vivo*. When plated on LB-agar plate containing X-gal and IPTG, the pSBT1 transformants developed as light blue colonies as those of the parent plasmid pBluescriptIISK+. To quantitate the β-galactosidase expression the whole cell β-galactosidase activity was measured for cells harbouring pBluescriptIISK+ and pSBT1. The cells carrying pSBT1 showed 80% inhibition of β-galactosidase expression compared to the cells carrying the parent plasmid pBluescriptIISK+ (fig 3).

The reduced β-galactosidase expression in the case of pSBT1 might have resulted either due to poorer codon usage or due to poly(pu)/poly(py) sequence adopting an altered structure which cannot be efficiently transcribed by *E.coli* RNA polymerase *in vivo*. The first argument of reduced gene expression due to poorer codon usage can be ruled out as preferred codons were used to construct pSBT1. Therefore it is likely that the poly(pu)/poly(py) sequence in pSBT1 adopts a structure which acts as transcriptional block *in vivo*.

To substantiate our proposition we selected another degenerate duplex (duplex mT12) which is a variant of duplex T1 where TCT triplets at positions 41,42 and 44 and TCC triplet at position 47 for serine were replaced by AGC triplets to break the poly(pu)/poly(py) sequence motif (fig 1). Since AGC is a less preferred codon (codon usage frequency 0.191) compared to UCU and UCC (codon usage frequencies are 0.361 and 0.338 respectively) for serine residue, such a replacement was expected to further decrease the β-galactosidase expression if codon usage alone was responsible for the five fold decrease in β-galactosidase expression in pSBT1. Duplex mT12 lacks the poly(pu)/poly(py) sequence motif and thus fails to meet the sequence requirements for an intramolecular triple helix structure eventhough it codes for same amino acid sequence as pSBT1. The construct pSBmT12 was transformed into *E.coli* JM109 cells and the transformants developed as dark blue colonies when plated on X-gal-IPTG plate whereas the pSBT1 transformants developed as light blue colonies. When β-galactosidase activity was measured the cells harbouring pSBmT12 showed five fold higher β-galactosidase expression as compared to those containing the construct pSBT1 (fig 3) eventhough less preferred codons were used in the case of pSBmT12. Moreover, in both the cases the β-galactosidase gene is under the control of identical promoters. The β-galactosidase expression of pSBmT12 was similar to that of the parent plasmid pBluescriptIISK+ as shown in figure 3. This observation rules out the possibility that the reduced gene expression for pSBT1 due to poor codon usage and substantiates the proposition that the poly(pu)/poly(py) sequence within the β-galactosidase gene adopts a structure which cannot be transcribed by *E.coli* RNA polymerase efficiently.

In vitro structural transition: 2D chloroquine gel electrophoresis

We carried out 2D chloroquine gel elctrophoresis to investigate whether *in vitro* and under physiological ion concentration, the designed poly(pu)/poly(py) sequence adopts triple helix structure in the construct pSBT1 and is absent in the case of the sequence in pSBmT12 as expected from the β-galactosidase expression study. Two-dimensional chloroquine gel electrophoresis under various pH and metal ion concentration has extensively been used to study the structural transition of a segment of poly(pu)/poly(py) DNA sequence from right handed B DNA to intramolecular triple

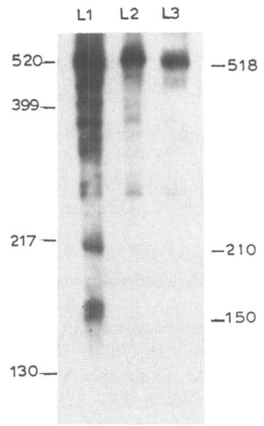

Figure 5. RNA northern blot analysis for total RNA isolated from *E.coli* JM109 cells harbouring pSBT1 (lane 1), pSBmT12 (lane 2) and parent plasmid pBluescriptIISK+ (lane 3) The size marker is HinfI digested pUC19.

helix structure in supercoiled plasmid under physiological superhelical density (18). Here the first dimension is run at low pH since triple helix is stabilized at low pH. Subsequently the gel is equilibrated and run in the second dimension at a higher pH and in the presence of chloroquine to destabilize the triplex structure and hence, to detect structural transition from an unwound triplex to B DNA form. The structural transition is associated with a loss of supercoiling which results in a break in the topoisomer profile. Figure 4 shows the 2D gel electrophoresis data for the construct pSBT1 and pSBmT12 at different hydrogen ion concentrations and in the presence and absence of bivalent metal ions (described in materials and method). Electrophoregram in figure 4 shows a break in topoisomer profile for pSBT1 at pH 5 both in the presence (fig 4a) and absence of Mg^{2+} ions (fig 4b). The construct pSBT1 also showed a detectable break in 2D gel at pH 6.8 in presence of 2mM Mg^{2+} ions but did not show such break at pH greater than 6.8, even in the presence of Mg^{2+} ions (data not shown). The construct pSBmT12 did not show any break in 2D gel topoisomer profile under identical experimental conditions (fig 4c). These experiments suggest that at physiological pH and in presence of 2mM Mg^{2+} poly(pu)/poly(py) sequence present in the plasmid pSBT1 adopt about one and half turn of altered DNA structure. The break in the topoisomer profile at low pH in presence of Mg^{2+} ion in 2D chloroquine gel has been shown to be a characteristic feature for the formation of intramolecular triple helix structure under physiological superhelical density (14). However, our experiments do not differentiate whether the triplex structure is purine. purine.pyrimidine or purine.pyrimidine. pyrimidine under the above experimental conditions. The parent plasmid pBluescriptIISK+ and the clone pSBmT12 carrying the synthetic duplex mT12 did not show the presence of unwound structure under identical experimental conditions.

RNA Northern blot analysis of the β-galactosidase transcript

To investigate whether the down regulation of β-galactosidase gene expression occurs during transcription elongation, RNA Northern blot analysis of the β-galactosidase transcripts was done to examine the length of β-galactosidase transcripts for construct pSBT1, pSBmT12 and the parent plasmid pBluescriptIISK+. Figure 5 shows the autoradiogram of the Northern blot analysis of the total RNA isolated from cells harbouring pSBT1, pSBmT12 and the parent plasmid pBluescriptIISK+. The Northern analysis of the transcripts isolated from cells harbouring pSBT1 showed the presence of truncated transcripts including a transcript of about 150 nucleotide, in addition to the expected full length transcript of β-galactosidase gene when radiolabelled lacZ′ fragment was used as probe (fig 5). However, the cells harbouring plasmid pSBmT12 did not show any such aborted transcripts. The 150 nucleotides long aborted transcript observed for pSBT1,accounts for the distance between the transcription start site of the β-galactosidase gene (854) and the first nucleotide (706) of the triple helix potential poly(pu)/poly(py) sequence segment. Our RNA bloting analysis (figure 5) showed the presence of aborted transcripts other than the truncated 150 nucleotide long transcript only for pSBT1. The detection of aborted β-galactosidase transcripts clearly suggests that someway triple helix potential polypurine/polypyrimidine sequences present in pSBT1 is causing transcription elongation block. However, 80% reduction in β-galactosidase expression observed in the case of pSBT1 can not be directly correlated with the intensity of the aborted transcript bands comapred to band intensity of the full length transcript as seen in the figure 5. This could be either due to the instability of the truncated transcripts or due to the use of nick translated lacZ′ fragment as probe which is likely to hybridize more efficiently with the full length transcript than the smaller truncated transcript. Quantitation of the level of the truncated transcripts using 5′ labelled oligonucleotide probe corresponding to the upstream of the poly(pu)/poly(py) sequence was not possible due to weak hybridization signal.This RNA blotting experiments demonstrate that poly(pu)/poly(py) sequence present within a gene can down regulate its expression and the regulation indeed occurs at the level of transcription elongation where the movement of RNA polymerase *ensemble* over the DNA template is blocked by the structure adopted by the poly(pu)/poly(py) sequences.

Biological significance

Transcription elongation control has been observed in case of several genes such as *c-myc, c-fos, Drosophila hsp70* etc but whether any DNA secondary structures are responsible for such an abnormal transcription termination is still unclear, even though sequence analysis done so far shows the distribution of sequence motifs capable of adopting unusual DNA structures (2−4,29,39).

Our results using β-galactosidase gene as a model system for the first time show that poly(pu)/poly(py) sequence motifs within the gene can act as transcriptional regulators by partially blocking transcription elongation *in vivo* inspite of the fact that the triple helix structure can get destabilized due to the positive supercoils generated ahead of the transcription *ensemble* and the triplex can flip over to the transcribable B-DNA form. Hence we propose that a stretch of poly(pu)/poly(py) sequence within a gene can establish a dynamic equilibrium between the non-transcribable triplex structure which can physically block the movement of RNA polymerase over the template DNA and the transcribable B-DNA form. This would lead to modulation of transcription elongation resulting in the formation of truncated transcripts and an overall reduction of transcription as observed in the case reported here.

The effect of intramolecular triplex formation on transcription will not only depend on the stability of such structures *in vivo* but also on the location of such sequence elements with respect to the transcription start site. In figure 2 we have shown a schematic diagram of purine motif and the pyrimidine motif of

triplex structure which the duplex T1 can adopt when present in supercoiled plasmid. In the pyrimidine motif half of the pyrimidine strand folds back into the major groove of the remaining WC duplex and forms a triple helix by Hoogsteen base pairing and the purine strand loops out. In this model half of the purine strand is involved in the triplex structure formation and the other half remains in an unstructured single stranded conformation. In the purine motif of triplex formation half of the polypurine strand folds back into the major groove of the remaining WC duplex forming triplex structure and the polypyrimidine strand loops out (2−4,14−15). This structure has been found to be stable in the pH range of 5−8 in presence of bivalent metal ions.

The effect of intramolecular triple helix structure on transcription will also depend on the strand which is being transcribed (fig 2). Here we outline a probable mechanism by which intramolecular triple helix models for the poly(pu)/poly(py) sequences can differentially inhibit transcription elongation. In pyrimidine motifs (fig 2a) the RNP will encounter the 'triplex knot' once, as half of the polypurine strand (sense strand) is involved in the triplex, while in the purine motifs (fig 2b) the whole polypurine stretch is involved in triplex formation so the RNP will find purine motifs as stronger transcriptional block compared to pyrimidine motifs.

Several triple helix binding proteins have been identified (40−42) including ribonucleoproteins containing a RNA component which were found to interact with triplex potential region of the *c- myc* gene and modulate *c-myc* gene expression (21,40). Binding of these protein factors may be the first step towards the nucleation of triple helix formation *in vivo* making the *in vivo* structural transition a flexible process.

CONCLUSIONS

The results reported here demonstrate that polypurine/ polypyrimidine sequences present within genes can down regulate gene expression *in vivo* and the regulation occurs at the transcription elongation step where the stretch of poly(pu)/ poly(py) sequence adopt structure which destabilize the assembly of RNA polymerase on DNA while transcribing the template DNA. It would be interesting to investigate the effect of such triple helix potential sequences when present both downstream and upstream of genes to understand how local supercoil induced DNA structure influence gene expression in eukaryotic systems where DNA is wrapped around nucleosomes.

ACKNOWLEDGEMENTS

The authors acknowledge Department of Biotechnology (India), for financial support, P.Balagurumoorthy for oligonucleotide synthesis, Dr. P.Burma for valuable suggestions on experimental work, Dr.Vani Brahmachari and R.Bagga for critical comments on the manuscript. PSS is senior research fellow of Council of Scientific and Industrial Research (India).

REFERENCES

1. Proudfoot,N.J. (1989) *Trends Biochem.Sci.* **14**, 105
2. Wells,R.D., Collier,D.A., Hanvey,J.C., Shimizu,M. and Wholrab, F.(1988) *FASEB J.* **2**, 401−407.
3. Yagil,G.(1991) *Critical Reviews in Biochemistry and Molecular Biology* **26**, 576−559
4. Palecek,E.(1991) *Critical Reviews in Biochemistry and Molecular Biology* **26**, 151−226.
5. Lawrence,J.P. and Wang, J.C. (1985) *Cell* **40,** 129−137.
6. Bagga,R.,Ramesh,N. and Brahmachari,S.K.(1990) *Nucl.Acids.Res.*, **18**, 3363−3369.
7. Waga, S.,Mizuno,S. and Yoshida,M. (1990) *J.Biol.Chem.*, **265** 19424−19428
8. Horowitz, M.S. and Loeb, L.A. (1988) *Science*, **241**,703−705.
9. Branco,L., Kotlarz,D., Kolb,A., Diecman,S. and Buc,H.(1989) *EMBO J.* **8**, 4289−4296.
10. Hentschel,C.C.(1982) *Nature* **295**,714−716
11. Birnboin, H.C.,Sederoff,R.R.and Paterson, M.C.(1979) *Eur.J. Biochem.* **98**,301−307.
12. Manor,H.,Sridhara, R.B.and Martin,R.G.(1984) *J.Mol.Evol.* **27**, 96−101.
13. Behe,M.(1987) *Biochemistry* **26**, 7870−7875
14. Voloshin,O.N.,Mirkin,S.M.,Lyamichev,V.I.,Belotserkovskii,and Frank-Kamenetskii,M.D.(1988) *Nature* **333**, 475−476
15. Kohwi,Y. and Shigematsu,T.K. (1988) *Proc.Natl.Acad.Sci.***85**, 3781−3785.
16. Bernues, J., Beltran,R., Casasnovas, J.M. and Azorin,F. (1989) *EMBO J.* **8**, 2087−2094.
17. Kohwi, Y.,Malkhosyam, S.R.,and Kohwi-Shigematsu, K.(1992) *J.Mol.Biol.* **223**, 817−822.
18. Lyamichev, V.I., Mirkin,S.M.,Frank-Kamenetskii,M.D.(1985) *J.Biomol.Str.Dyn.* **3,** 327−338.
19. Toulme,J.J. and Helene,C.(1988) *Gene* **72**, 51−58.
20. Strobel,S.A.and Dervan,P.B.(1990) *Science* **249**,73−75.
21. Kinniburge, A. J. (1989) *Nucl.Acids. Res.* **17**, 7771−7778.
22. Coony, M.,Czernuszewick,G.,Postel,E.H.,Flint,S.J. and Hogan,M.E. (1988) *Science* **241,** 456−459
23. Young,S.L., Krawczyle, S.H., Matleucci,M.D. and Toole,J. (1991) *Proc.Natl.Acad.Sci.(USA)* **88**, 10023−10026.
24. Nalentin,G.,Thuong,N.T.and Helene,C.,(1992) *Proc.Natl.Acad.Sci. (USA)* **89**, 504−508.
25. Maher III,L.J.,Dervan,P.B.,and Wold,B.(1992) *Biochemistry* **21**,70−81
26. Postel,E.H., Flint,S.J.,Kessler,D.J. and Hogan,M.E.(1991) *Proc.Nal.Acad.Sci.(USA)* **88,** 8227−8231
27. Kohwi,Y. and Kohwi-Shigematsu,T. (1991) *Genes and Dev.* **5**, 2547−2554.
28. Glaser, R. G., Thomas,G.H., Siegfried, E.,Elgin,S.E.R and Lis,J.T. (1990) *J.Mol.Biol.* **211**, 751−761.
29. Tripathi,J. and Brahmachari, S.K.(1991) *J.Biomol.Str.Dyn.***9**, 387−397
30. Sarkar, P.S.,Bagga R.,Balagurumoorthy, P. and Brahmachari, S.K.,(1991) *Curr. Sci.* **61**,586−591.
31. Brahmachari, S.K.,Sarkar, P.S.,Bagga,R.,Balagurumoorthy,P. and Burma,P.(1991) *Nucl.Acids.Res.Symp,Ser* **24**,163−166
32. Miller,J.H. (1972) *Experiments in Molecular Genetics* Cold Spring Harbour Laboratory,CSH.
33. Brahmachari,S.K.,Mishra,R.,Bagga,R.and Ramesh,N.,(1989) *Nucl. Acids. Res.* **17**, 7273−7281 34. Ebina,Y. and Nakazawa, A.,(1983) *J.Biol.Chem.* **258**,7072−7078
35. Sambrook,J., Fritsch,E.F., Maniatis,T., eds.(1989) *Molecular Cloning: A Laboratory Manual*,Cold Spring Harbour Laboratory Press,CSH.
36. Berman,M.L.(1983) *Biotechniques* **7**,576−579
37. Andree,V., Fowlev, and Zabin, F.(1983) *J.Biol.Chem.* **258,** 14354−14358.
38. Grosjean,H. and Feirs,W.(1982) *Gene* **18,** 199−209.
39. Bentley,D.L. and Groudine,M.(1988) *Cell* **52**, 245-
40. Davis, T.L., Firulli,A.B., Kinniburge,A.J. (1989) *Proc.Natl. Acad.Sci.(USA)* **86**, 9682−9686.
41. Lewis,C.D.,Clark,S.P.,Felsenfeld,G.(1988) *Genes.Dev.***2**, 863−873.
42. Kiyama,R. and Camerini-Otero,R.D.,(1991) *Proc.Natl.Acad.Sci. (USA)*, **88**, 10450−10454.