Research article

Poly purine.pyrimidine sequences upstream of the beta-galactosidase gene affect gene expression in Saccharomyces cerevisiae
Amit K Maiti* and Samir K Brahmachari

Address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India
E-mail: Amit K Maiti* - amit@dplanet.ch; Samir K Brahmachari - skb@cbt.res.in
*Corresponding author

Abstract

**Background:** Poly purine.pyrimidine sequences have the potential to adopt intramolecular triplex structures and are overrepresented upstream of genes in eukaryotes. These sequences may regulate gene expression by modulating the interaction of transcription factors with DNA sequences upstream of genes.

**Results:** A poly purine.pyrimidine sequence with the potential to adopt an intramolecular triplex DNA structure was designed. The sequence was inserted within a nucleosome positioned upstream of the \( \beta \)-galactosidase gene in yeast, Saccharomyces cerevisiae, between the cycl promoter and gal |0| Upstream Activating Sequences (UASg). Upon derepression with galactose, \( \beta \)-galactosidase gene expression is reduced 12-fold in cells carrying single copy poly purine.pyrimidine sequences. This reduction in expression is correlated with reduced transcription. Furthermore, we show that plasmids carrying a poly purine.pyrimidine sequence are not specifically lost from yeast cells.

**Conclusion:** We propose that a poly purine.pyrimidine sequence upstream of a gene affects transcription. Plasmids carrying this sequence are not specifically lost from cells and thus no additional effort is needed for the replication of these sequences in eukaryotic cells.

Background

Stretches of poly purine.pyrimidine (poly pur.pyr) sequences are overrepresented in the eukaryotic genome [1,2] and often positioned upstream of genes [3,4]. Many eukaryotic genes contain a poly pur.pyr box upstream of the coding region [5]. Transcription factors may interact specifically with those regions as occurs upstream of the Drosophila hsp26 gene [6,7] where transcription is enhanced during a heat shock.

Under torsional stress poly pur.pyr sequences have the potential to adopt an intramolecular triplex formation as HY-3 and H-Y5 [8–10]. A proposed triplex model fits with the regular poly pur.pyr tract with mirror symmetry [11] around a loop sequence. Shimizu et al. [10] have shown that base composition in loop sequences affects intramolecular triplex formation while orientation of the insert in supercoiled plasmid has no effect on isomerisation of the two types of triplex DNA. Nonmirror symmetric poly pur.pyr sequences can also adopt an intramolecular triplex structure [12].

The *in vivo* existence of poly pur.pyr sequences as an intramolecular triplex has been detected in *Escherichia*...
coli [13,14]. Modulation of gene expression by insertion of designed poly pur.pyr sequences within the E. coli genes [15,16] and upstream of a gene in mouse cells has been demonstrated [17]. Poly pur.pyr sequences in active regions of transcription form a DNA-RNA hybrid triplex and make the template transcriptionally inert [18]. Poly homo purine.pyrimidine sequences stop in vitro replication and pause in vivo replication when cloned into a plasmid under the control of the SV40 promoter [19]. Peleg et al. [20] showed that the unwinding activity of the SV40 large T-antigen helicase is decreased in the presence of a triplex structure in vitro. When the triplex structure is stabilized with BeP1 [benzo(c) pyridoindole], replication elongation is inhibited in plasmids by prokaryotic DNA polymerase [16]. Friedreich Ataxia-associated poly pur.pyr expansion (GAA.TTC) when cloned in plasmids, shows inhibition of transcription and replication in cos-7 cells [21]. However, structural aspects of the poly pur.pyr sequences as intramolecular triplex DNA and their role in different biological processes in transcription regulation, replication, recombination and interaction with the nucleosome in vivo have not been fully determined [3].

A systematic search for poly pur.pyr sequences within yeast chromosome III indicates that they frequently occur upstream of yeast genes [4]. The presence of these sequences in the promoter regions suggests they are involved in gene regulation. Here, we examine the effect of poly pur.pyr sequences on gene regulation in a position (UASg), upstream of the β-galactosidase gene where enhancer proteins are known to disrupt nucleosomes to activate transcription initiation. The E. coli (β-galactosidase gene is under the control of the upstream activating sequences of gall-gal 10 and the cycl promoter (Fig. 1). Yeast cells carrying this plasmid [22] grow normally within medium supplemented with glucose as the carbon source and do not express (β-galactosidase (repressed state). When cells are transferred to a medium containing galactose (derepression), two proteins, Gal4 and Gal80, are expressed from chromosomal genes, disrupt nucleosomes, bind at the UASg and activate transcription [23,24]. We inserted poly pur.pyr sequences and a duplex control sequence within a positional nucleosome at the XhoI site between the UASg and the cycl promoter.

Figure 1
A β-galactosidase expression system in yeast. The E. coli β-galactosidase gene is under the control of gall-gal 10 UASg and the cycl promoter. In glucose medium, this gene is not expressed and remains repressed, while activated in galactose medium by the expression of chromosomal proteins Gal4 and Gal80. The Gal4-Gal80 complex disrupts nucleosomes and binds DNA at UASg to activate transcription.

Results
Poly pur.pyr sequences affect gene expression in yeast cells
A 61-mer Xho1-Xho1 fragment containing a 58-mer poly pur.pyr sequence with a loop sequence at the center capable of adopting a triplex DNA structure was designed (Fig. 2). The fragment was cloned (single copy, pAMTS61; multiple copies, pAMTS61a, pAMTS61b, pAMTM61; Fig. 3a) into the Xho1 site of pLGSD36. The latter differs from the original pLGDS5 [22,25] in having a polylinker. A random sequence of the same length that will take up a normal duplex structure was designed and included in the study (pAMDC61).

We have taken advantage of the induction system of the β-galactosidase gene under the control of gall-gal 10 UASg. Upon derepression with galactose, enzyme expression is reduced about 12–50 fold (Fig. 3b and table 3c) in cells carrying single to multiple copies of poly pur.pyr inserts containing plasmids compared with duplex-carrying plasmids. No marked differences were observed in the enzyme activities in cells containing pLGDS36 (480 units) and pAMDC61 (485 units) (i.e. insertion of extra DNA sequences of the same length as the poly pur.pyr sequence or separation of adjacent DNA sequences around the Xho1 site has no effect on β-galactosidase expression). Therefore, the presence of the poly pur.pyr sequence in the plasmid causes a reduction in β-galactosidase expression during derepression.
Reduced gene expression is correlated with reduced transcription

After derepression with galactose, transcripts of the β-galactosidase gene were analyzed and quantitated from yeast cells carrying pAMTS61 and pAMDC61. Cells were aliquoted during the expression measurement (Table 3c) at OD$_{600}$ = 1. For northern blot analysis, total cellular RNA was hybridized with the 4.8 kb EcoR1 fragment carrying the ura3 and lacZ genes. There was reduced β-galactosidase (lacZ) transcript (Fig. 4b lane 2) in cells carrying pAMTS61 compared with cells carrying pAMDC61 (lane 1). There was 10-fold less transcript in cells carrying pAMTS61 (single copy poly pur.pyr sequence lacZ band/ura3 band = 0.05) than those harboring pAMDC61 (normal duplex lacZ band/ura3 band = 0.5). There was about a 12-fold reduction in β-galactosidase expression in cells carrying the poly pur.pyr sequence (42 units) compared with those carrying the duplex (485 units). Measurements of protein and transcript indicate that reduced expression in cells carrying the poly pur.pyr sequence is almost correlated with reduced transcription.

Poly pur.pyr insert carrying plasmids are not specifically lost from the cells

To determine whether plasmids carrying the poly pur.pyr sequence are specifically lost from cells after subsequent generations of growth in nonselective medium, plasmid stabilities and copy numbers were measured. Yeast cells do not maintain these minichromosome vectors in 100% efficiency and typically 80–90% of cells carry these plasmids in high copy number (20–30 copies/cell) after transformation. Cells lose plasmids slowly during subsequent generations of growth in nonselective medium. However, it has been shown that when replication is impaired at the plasmid origin, copy number/cell rapidly decreases during subsequent culture (within 10–25 generations) and cells completely lose the plasmid [26]. The percentage of cells maintaining plasmids is denoted as % stability. The % stability of each plasmid could be quantitated by the loss of a marker (ura3) from the cells, which will not grow in minimal medium devoid of uracil (CM-URA).

If replication of poly pur.pyr sequences in the plasmid were affected, a reduction in stability as well as in copy number would be expected [26] typically within 15–20 generations of growth. Figure 4a shows little difference in stabilities. Almost equal stability has been observed in both the cases after 20 generations of growth (i.e. cells are not preferentially loosing plasmids carrying a poly pur.pyr sequence). The copy number (Fig. 4c) of plasmids carrying a poly pur.pyr sequence (copy number, 2.6) compared with those carrying a duplex (copy number, 2.2) after 20 generations of growth does not differ significantly. If replication of the plasmid is impaired, a marked decrease in copy number of plasmids carrying a poly pur.pyr sequence would have been observed. These results suggest that plasmid replication may not be significantly reduced or affected by the presence of a poly pur.pyr sequence.
It has been reported that poly pur.pyr sequences within genes down-regulate gene expression in *E. coli* [15]. A homogenous d (CT) n. d (AG) n stretch (which does not form an intramolecular triplex in *vivo*) together with Heat Shock Elements (HSE) upstream of the *Drosophila hsp26* gene plays an integral role in resetting chromatin structure [6,7]. It is possible that heat shock facilitates nucleosome disruption from tightly bound dCT.dAG stretch by GAGA transcription factor [30] and transcription initiation occurs upon binding with Heat Shock Factors (HSF). The fine-tuning of eukaryotic gene regulation also depend upon the DNA sequences located upstream of genes. The decreased transcription, observed in yeast cells, in this study, may be due to the presence of poly pur.pyr sequences that might bind more tightly to a positional nucleosome(s) present at the junction of UASg and the cycl1 promoter [23] than the control DNA sequences. This binding of nucleosome with a poly pur.pyr sequence may inhibit the binding of protein factors [24,31] on the DNA. Here, poly pur.pyr and duplex sequences are inserted in DNA where a positional nucleosome is reorganizable and is normally disrupted before transcription initiation by the Gal4-Gal80 complex. We analyzed the fate of this nucleosome in *vivo* during repression with glucose and derepression with galactose in both duplex- and poly pur.pyr-carrying cells (unpublished observation). Preliminary results suggest that the single copy poly pur.pyr sequence binds with the positional nucleosome with higher affinity than the normal duplex sequence and inhibits disruption by the Gal4-Gal80 complex before transcription initiation. Also, in multiple tandem copies the poly pur.pyr sequence (pAMTM61a) keeps the nucleosome binding and interestingly disruption of nucleosome occurs from the inverted orientation of the poly pur.pyr sequence (pAMTM61) at the other end. From these data, the reduced gene expression may be explained in the context of nucleosome binding with the poly pur.pyr sequence and a model is presented (Fig. 5). However, this observation is contrary to the evidence that an intermolecular triplex repels the nucleosome in *vivo* [32]. We believe that attraction or repulsion of the nucleosome in *vivo* depends on the different poly pur.pyr sequences and their conformation at the nucleosomal surface (i.e. as an intramolecular triplex). This sequence forms an intramolecular triplex in *vivo* [27], however its conformation in *vivo* requires further investigation.

No large decrease in stability and copy number of plasmids carrying the poly pur.pyr sequence was observed during subsequent generations of growth suggesting that the presence of this sequence does not affect replication in eukaryotic cells unlike the observation of Duval-Valentine *et al.* [16] where prokaryotic DNA polymerases pause during replication in *vivo*. This supports the sug-
gestion that no extra effort is needed by DNA polymerases to replicate such sequences that are abundant in the genome of eukaryotes. A plausible explanation is that the intramolecular triplex structure does not exist on naked DNA in eukaryotes rather such looped structure exists on the nucleosomal surface [33] and as replication proceeds, the nucleosome makes way [34] for the DNA polymerase, the intramolecular triplex structure resolves into duplex DNA and facilitates replication. It is possible that extra replication machinery in eukaryotic cells could resolve the intramolecular structure during replication. It has been observed that the SV40 large T-antigen has the unwinding capacity of intramolecular triplex during replication [35] in vitro and replication termination in poly dT tract is relieved by mitochondrial single-strand binding proteins [36]. It would be interesting to investi-
growth in galactose. (β-galactosidase expression was measured as described by Reynolds and Lundablad [37] and expressed in Miller's unit [38]. An aliquot of culture was plated on complete medium (YEPD) and replica plated onto minimal medium (CM-URA) to determine the percentage of cells carrying plasmids (% stability). Actual (β-galactosidase activities were obtained by multiplying average (β-galactosidase activity with the stability factor of each plasmid).

**Stability and copy number determination**

Plasmid stabilities and copy numbers were determined essentially as described earlier [26]. Transformants carrying plasmids containing the poly pur.pyr sequence and the duplex sequence were grown in YEPD (rich medium) from a single colony to saturation and plated on YEPD plates after dilution. After sufficient growth, plates were replica plated onto CM-URA synthetic medium.

% Stability = (Number of colonies grown in CM-URA / number of colonies grown in YEPD) \times 100.

Plasmid copy number was determined by quantifying ura3, which is present in single copy in yeast genome and presence of it in the multicopy plasmids. Total cellular RNA was isolated by the hot phenol method [39] from each culture (grown in selective media with 2% galactose) of cells carrying pAMTS61 and pAMDC61. An equal amount of RNA was run in a MOPS-formaldehyde agarose gel and transferred to a nylon membrane with 20 × SSC. A 4.8 kb EcoRI fragment carrying the entire lacZ and ura3 genes of pLGSD36 was labeled with alpha-32P dATP by random priming and used as a probe for hybridization at 42°C in 5 × SSC, 50% formamide. Filters were washed twice with 2 × SSC (37°C) for 15 min and twice with 0.5 × SSC (42°C) and autoradiographed. Autoradiograms were scanned and the ratio of the intensities of lacZ transcripts to ura3 transcripts was calculated.

**Isolation of total RNA and northern hybridisation**

Total cellular RNA was isolated by the hot phenol method [39] from each culture (grown in selective media with 2% galactose) of cells carrying pAMTS61 and pAMDC61. An equal amount of RNA was run in a MOPS-formaldehyde agarose gel and transferred to a nylon membrane with 20 × SSC. A 4.8 kb EcoRI fragment carrying the entire lacZ and ura3 genes of pLGSD36 was labeled with alpha-32P dATP by random priming and used as a probe for hybridization at 42°C in 5 × SSC, 50% formamide. Filters were washed twice with 2 × SSC (37°C) for 15 min and twice with 0.5 × SSC (42°C) and autoradiographed. Autoradiograms were scanned and the ratio of the intensities of lacZ transcripts to ura3 transcripts was calculated.

**Acknowledgement**

We thank Dr. P. S. Sarkar for designing poly pur.pyr sequences. Dr. L. Guarante for generous gift of plasmid, pLGSD5 and K. Usha for synthesizing oligonucleotides. We are thankful to “Sciencemanager” for careful editing of this manuscript. AKM is supported by DBT postdoctoral fellowship (R (IA) RA (NBTB/ MBUAKM/92-777). SKB is supported by a grant from DBT, Govt of India. We are grateful to anonymous reviewers for their helpful comments to improve this manuscript.

**References**

1. Tripathy J, Brahmachari SK: Distribution of simple repetitive (TG/CA) n and (CT/AG) n sequences in human and Rodent genomes. J Biol Chem 1992, 267:387-397
2. Behe MJ: An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. Nucleic Acids Res 1995, 23:689-95
3. Wells RD, Collmer DA, Hanvey JC, Shimizu M, Wohlrab F: The chemistry and biology of unusual DNA structures adopted by oligopurine. oligopyrimidine sequences. FASEB J 1988, 2:2399-49
4. Raghavan S, Burma PK, Brahmachari SK: Positional preferences of polypurine/polypyrimidine tracts in Saccharomyces cerevisiae genome: implications for cis regulation of gene expression. J Mol Evol 1997, 45:485-98
5. De Martinoff G, Pohl V, Mercken L, van Ommen GJ, Vassert G: Structural organization of the bovine thyroglobulin gene and its S’-flanking region. Eur J Biochem 1987, 164:591-599
6. Glaser RL, Thomas GM, Siegfried E, Elgin SC, Lis JT: Optimal heat-induced expression of the Drosophila hs26 gene requires a promoter sequence containing (CT) n. (GA) n repeats. J Mol Biol 1990, 211:751-61
7. Lu Q, Wallrath LL, Granok H, Elgin SC: (CT) n (GA) n repeats and heat-shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hs26 gene. Mol Cell Biol 1993, 3:2802-2814
8. Huyn H, Dahlberg JE: Topology and formation of triple-stranded H-DNA. Science 1989, 243:1571-6
9. Johnston BH: The S1-sensitive form of d(C-T)n.d(A-G)n: chemical evidence for a three-stranded structure in plasmids. Science 1988, 241:1800-4
10. Shimizu M, Kubo K, Matsumoto U, Shindo H: The loop sequence plays crucial roles for isomerization of intramolecular DNA triplexes in supercoiled plasmids. Mol Cell Biol 1994, 14:185-97
11. Mirkin SL, Lymanchev VI, Drushlyak KN, Dobrynin VN, Filipov SA, Frank-Kamenetskii MD: DNA H-form requires a homopurine-homopyrimidine mirror repeat. Nature 1987, 330:495-7
12. Kysic J: An intramolecular triplex structure from non mirror repeated sequence containing both pyruv.ppy and puppy.py triads. J Mol Biol 1995, 245:499-507
13. Karlovsky P, Pecinka P, Vojtiskova M, Makaturova E, Palecek E: Protonated triplex DNA in E. coli cells as detected by chemical probing. FEBS Lett 1990, 274:39-42
14. Kohwi Y, Maltoshnov SR, Kohwi-Shigematsu T: Intramolecular dG.dG.dC triplex detected in Escherichia coli cells. J Mol Biol 1992, 223:817-822
15. Sarkar PS, Brahmachari SK: Intramolecular triplex potential sequence within a gene downregulates its expression in vivo. Nucleic Acids Res 1992, 20:5713-5718
16. Duval-Valentin G, de Bizemont T, Takasugi M, Mergny JL, Bisagni E, Helene C: Triple-helix specific ligands stabilize H-DNA coformation. J Mol Biol 1995, 247:847-56
17. Kohwi Y, Kohwi-Shigematsu T: Altered gene expression correlates with DNA structure. Genes Dev 1991, 5:2547-54
18. Reaban ME, Griffin JA: Induction of RNA stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature 1990, 348:342-344
19. Rao BS: Regulation of DNA replication by homopurine/homopyrimidine sequences. Mol Cell Biochem 1996, 156:163-8
20. Peleg M, Kopel V, Borowiec JA, Manor H: Formation of DNA triple helices inhibits DNA unwinding by the SV40 large T-antigen helicase. Nucleic Acids Res 1995, 23:1299-302
21. Ohshima K, Montermini L, Walls RD, Pandolfo M: Inhibitory effects of expanded GAAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J Biol Chem 1998, 273:14588-95
22. Guarante L, Yocum RR, Sifford P, A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc Natl Acad Sci USA 1982, 79:7410-4
23. Fedor MJ, Lue NF, Kornberg RD: Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast. J Mol Biol 1988, 204:109-27
24. Morse RH: Nucleosome disruption by transcription factor binding in yeast. Science 1993, 262:1563-6
25. Guarente L, Ptashne M: Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. Proc Natl Acad Sci USA 1981, 78:2199-203
26. Maji AK, Sinha P: The mcm2 mutation of yeast affects replication, rather than segregation or amplification of the two micron plasmid. J Mol Biol 1992, 224:545-558
27. Brahmachari SK, Meera G, Sarkar PS, Balagurumoorthy P, Tripathi J, Raghavan S, Shaligram U, Pataskar S: Simple repetitive sequences in the genome: structure and functional significance. Electrophoresis 1995, 16:1705-14
28. Yagi G: Paranemic structures of DNA and their role in DNA unwinding. Crit Rev Biochem & Mol Biol 1991, 26:475-559
29. Brahmachari SK, Sarkar PS, Raghavan S, Narayan M, Maji AK: Polyurine/polypyrimidine sequences as cis-acting transcriptional regulators. Gene 1997, 190:17-26
30. Tsukiyama T, Becker PB, Wu C: ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 1994, 367:525-32
31. Chasan DI, Lue NF, Buchanan AR, LaPointe JW, Lorch Y, Kornberg RD: A yeast protein that influences the chromatin structure of UASg and functions as a powerful auxiliary gene activator. Genes Dev 1990, 4:303-14
32. Westin L, Blomquist P, Milligan JF, Wrangel O: Triple helix DNA alters nucleosomal histone-DNA interactions and acts as a nucleosome barrier. Nucleic Acids Res 1995, 23:184-91
33. Larsen A, Weintrab H: An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell 1982, 2:609-622
34. Bonne-Andrea C, Wong ML, Alberts BM: In vitro replication through nucleosomes without histone displacement. Nature 1990, 343:719-26
35. Kopel V, Pozner A, Baran N, Manor H: Unwinding of the third strand of a DNA triple helix, a novel activity of the SV40 large T-antigen helicase. Nucleic Acids Res 1996, 24:330-5
36. Mikhailov VS, Bogenhagen DF: Termination within oligo(dT) tracts in template DNA by DNA polymerase gamma occurs with formation of a DNA triplex structure and is relieved by mitochondrial single-stranded DNA-binding protein. J Biol Chem 1996, 271:30774-80
37. Reynolds A, Lundablad V: In current protocols in Molecular Biology (ed. Ausubel et al; Green publishing associates, inc and John willey and Sons 1993
38. Miller JH: experiments in molecular Genetics (Cold spring harbour laboratory Press, Cold spring harbour, NY), 1972
39. Collart M, Oliviero S: Preparation of yeast RNA. In current protocols in Molecular Biology (ed. Ausubel et al; Green publishing associates, inc and John willey and Sons 1993