Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved *cis*-Regulatory Elements in Human and Related Species

Anjali Verma,[||,†] Kangkan Halder,[||,†] Rashi Halder,[‡] Vinod Kumar Yadav,[‡] Pooja Rawal,[‡] Ram Krishna Thakur,[†] Farhan Mohd,[§] Abhay Sharma,[§] and Shantanu Chowdhury*[,†,‡]

Proteomics and Structural Biology Unit, G. N. Ramachandran Knowledge Centre for Genome Informatics and, Comparative Genomics and Gene Expression Unit, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India

Received April 20, 2008

Using a combination of in silico and experimental approaches, we present evidence that the G-quadruplex (G4) motif (an alternative higher-order DNA conformation) has regulatory potential. Genome-wide analyses of 99980 human, chimpanzee, mouse, and rat promoters showed enrichment of sequence with potential to adopt G4 (potential G4 or PG4) motifs near transcription start sites (TSS; $P < 0.0001$), supporting earlier findings. Interestingly, we found >700 orthologously related promoters in human, mouse, and rat conserve PG4 motif(s). The corresponding genes have enriched (z score > 4.0) tissue-specific expression in 75 of 79 human tissues and are significantly overrepresented in signaling and regulation of cell-cycle ($P < 10(-05)$). This is supported by results from whole genome expression experiments in human HeLa S3 cells following treatment with TMPyP4 [5,10,15,20-tetra(*N*-methyl-4-pyridyl) porphine chloride], which is known to bind the G4 motif inside cells. Our results implicate G4-motif mediated regulation as a more general mode of transcription control than currently appreciated.

Introduction

To understand the transcriptional network of a genome, characterization of all encoded regulatory elements is a fundamental prerequisite. Although specific nucleotide arrangements are established as regulatory elements at the sequence level,[1] the role of their potential to form alternate conformation is relatively unexplored in a regulatory context. The role of non-B-DNA structures has been implicated in recombination, replication, and regulation of gene expression in particular[2−6] both in prokaryotes (reviewed in ref 7) and eukaryotes (reviewed in refs 3, 8). Systematic prediction and determination of regulatory elements at the sequence level have been addressed in many studies,[1,9,10] however, very few have investigated DNA structure in a similar fashion.[11] Herein, we focus on searching and investigating the role of a particular type of non-B DNA motif, the G-quadruplex (G4[a]) motif, as a structural regulatory element.

G4 motifs are structural conformations composed of consecutive stacking of coplanar array of four cyclic reverse Hoogsteen hydrogen bonded guanines (G-tetrads, Figure 1A).[12,13] It is highly polymorphic due to the multiple combinations possible with respect to strand orientation and loop composition (reviewed in ref 14). These conformations were first reported in telomeres[15] and subsequently in other genomic regions, viz. immunoglobin heavy chain switch regions,[16] G-rich minisatellites,[17,18] and rDNA.[19] Apart from genome-wide prevalence,[20,21] several gene promoters like β-globin,[22] retinoblastoma susceptibility genes,[23] the insulin gene,[24] adenovirus serotype 2,[25] PDGF,[26] *c-KIT*,[27] hypoxia inducible factor 1-alpha,[28] *BCL*-2,[29] and *c-MYC*[30,31] harbor PG4 motifs. Recent evidence of in vivo presence of G4 motif in the telomeres of *Stylonychia* nanochromosomes[32] and the distinct role of RNA quadruplexes in translation modulation[33,34] strongly suggest the possibility of G4 motifs as functional elements. Moreover, telomere shortening induced by G4 motif specific molecules[35,36] and repression of *c-MYC* on stabilization of the PG4 motif in its promoter (using the G4 binding ligand TMPyP4)[37] and, conversely, overexpression in case of site-specific mutations that destabilized a PG4 motif within the *c-MYC* promoter[30] indicated the functional role of G4 motif in transcription.

In genome-wide studies, enrichment of PG4 motifs in promoters of mammalian genomes including human,[38−41] chicken,[42] and several bacteria[43] has been observed, suggesting widespread regulatory influence of G4 motifs. Keeping this in view, we asked whether G4 motifs present structural *cis*-regulatory sites. This was tested in a genome-wide context using a combination of bioinformatics and experimental methods wherein conservation within related species and other characteristics that are hallmarks of regulatory elements, e.g., position with respect to TSS, association with core promoter elements, and tissue-specific expression were analyzed. This was followed by experimentally observing whole genome expression of cells under the influence of a G4-interacting ligand and analyzing results in the context of G4 motif-mediated regulation. Our results support DNA structure in the form of G4 motifs as *cis*-regulatory sites with wide-ranging biological implications.

Materials and Methods

Sequence Retrieval and Analysis. The 10 kb regions centered at TSS were retrieved from NCBI Genbank files (ftp://ftp.ncbi.nih.gov/genomes/) build 35 for Human, build 1 for chimpanzee, build 34 for mouse, and build 3 for rat. CpG island and "intron" data were obtained from http://genome.ucsc.edu/cgi-bin/hgTables. Potential G4 (PG4) motif forming sequences (with stem size three (Figure 1A)) were searched with a customized algorithm as

* To whom correspondence should be addressed. Phone: 91 11 2766 6157. Fax: 91 11 2766 7471. E-mail: shantanuc@igib.res.in.

† Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR.

‡ G. N. Ramachandran Knowledge Centre for Genome Informatics, Institute of Genomics and Integrative Biology, CSIR.

§ Comparative Genomics and Gene Expression Unit, Institute of Genomics and Integrative Biology, CSIR.

|| These authors contributed equally to this work

[a] Abbreviations: G4, G-quadruplex; PG4, potential G-quadruplex; TMPyP4, 5,10,15,20-tetra(*N*-methyl-4-pyridyl) porphine chloride; TSS, transcription start site.

10.1021/jm800448a CCC: $40.75
Published on Web 09/04/2008

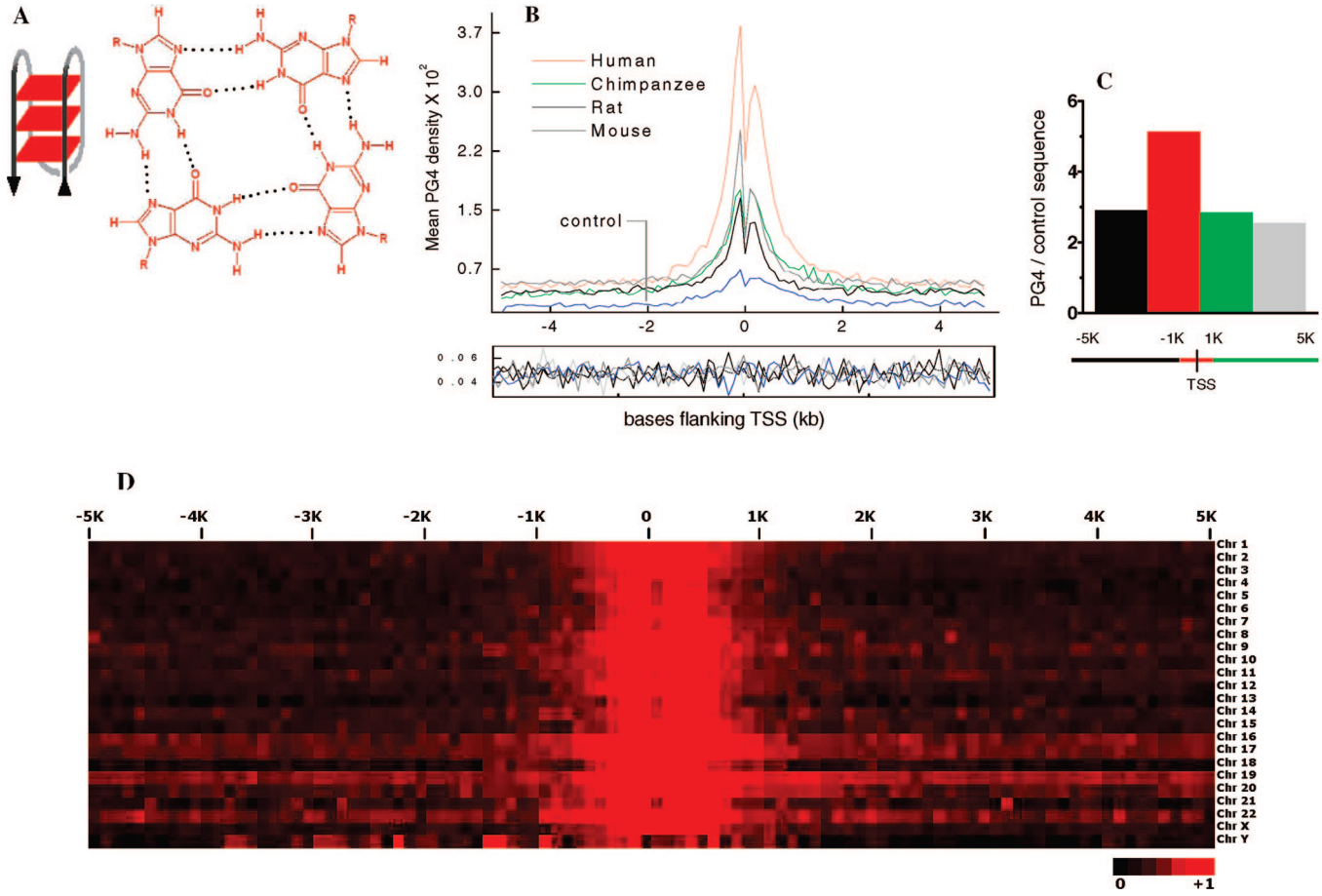

Figure 1. Enrichment of PG4 motifs near transcription start sites (TSS). (A) Schematic representation of G4 motif: red planes show three G-tetrads constituting the stem (left panel) and hydrogen bonding scheme of guanines forming a G-tetrad is shown in right panel. (B) Upper panel: distribution of PG4 motifs in ± 5 kb region flanking TSS and a control pattern (in human) that will not adopt G4-motif; PG4 motif density in 100 base windows (ratio of number of nucleotides that contribute to PG4 in each window) averaged over the number of promoters in the respective genomes is shown. Lower panel: distribution of PG4 motifs after "shuffling" the above human sequences 10000 times, repeated five times. (C) Distribution of PG4 motifs as a ratio of the control pattern in ± 1 kb centered at TSS (red), 4 kb region beyond (upstream (black) and downstream (green)), and >129 Mb of the last two introns (gray) within the human genome. Statistical comparisons were done in pairwise mode for the different regions and significance tested using the two-tailed nonparametric signed Wilcoxon test. (D) Relative enrichment of PG4 motifs near TSS across all chromosomes. Heat map showing chromosome wise PG4 density in 100 base windows in the human genome (averaged over all genes in respective chromosomes).

Table 1. Distribution of PG4 Motifs Near Transcription Start Sites (TSS)

		10 kb centered at TSS		±1 kb centered at TSS	
	ORFs studied	promoters with at least one motif	no. of PG4 motifs	promoters with at least one PG4 motif (% of ORFs studied)	total no. of PG4 motifs (% of motifs in 10 kb)
human	25706	19809	89455	14267 (55.5)	38757 (43.3)
mouse	26360	22144	73095	12701 (48.2)	24460 (33.5)
rat	24758	19537	54560	9923 (40.1)	16948 (31.1)
chimpanzee	23156	15498	56534	9469 (40.9)	20140 (35.6)

described earlier[43] within 10 kb, centered at annotated transcription start sites (TSS) of 25706 human, 23156 chimpanzee, 26360 mouse, and 24758 rat genes (Table 1). Briefly, we adopted a general pattern G_n-N_{L1}-G_n-N_{L2}-G_n-N_{L3}-G_n where G is guanine and N is any nucleotide including G. The PG4 stem (number of consecutive G-tetrad stacking) is given by n, which could be varied from 2−6, maintaining a constant n within a single motif while the number of nucleotides within loops (L1, L2 and L3) could vary from 1−7. The program was rerun with cytosine (C) instead of guanine (G) to identify motifs on the complementary strand and later appropriately corrected for strand orientation before mapping their position in the context of genes. The motif finding algorithm is conceptually similar to ones published earlier.[20,21]

We restricted our program to a stem size of 3 and loop length of 1−7 after considering the following points. PG4 motifs with larger stem sizes (4, 5, and 6) were relatively few (Supporting Information Table S1). PG4 motifs with stem size 3 were selected for our study because most of the in vitro characterizations and experiments with G4 motifs within the human genome (telomeric and other genomic sequences) studied have stem size 3.[20,21] The loop length was restrained to a maximum of seven nucleotides arbitrarily for practical reasons. An unrestrained loop length would make searching difficult, moreover, we found that PG4 motifs exist as short nucleic acids (length between 10−39 bases was predominant (data not shown)), which is also supported by earlier evidence.[43]

Using this algorithm, we found that PG4 motifs were enriched within a ± 1 kb region centered at TSS for human, chimpanzee, mouse, and rat genomes compared to the rest of genomic regions. We collated this information in a database EuQuad, which allows

searching for PG4 motifs across human, chimpanzee, mouse, and rat promoters with relative ease (http://quadbase.igib.res.in[44]).

Analysis of Motif Conservation in "Orthologous" Promoters. Using the human gene names that have at least one PG4 motif within ± 2 kb centered at TSS, we searched for the orthologous gene name in rat and mouse using KEGG database.[45] We retrieved 62509 mouse− and 24930 rat−human promoter pairs. Then for each gene pair, a PG4 motif was searched within 200 bp on the either side in the corresponding rat and mouse orthologous promoter with respect to the human PG4 motif position. All PG4 motifs identified within human, chimpanzee, mouse, and rat promoters along with PG4 motifs that are conserved across multiple species is available from the database at http://quadbase.igib.res.in.[44]

Tissue-Specific Enrichment of Gene-Set with PG4 Motifs in Promoters. Tissue-specificity of genes harboring motifs within putative promoters was checked in 79 human tissues.[46] Analysis was largely based on a previously described method.[1] The expression data of each gene across all tissues was first normalized to be mean 0 and variance 1 before ranking them as per their normalized expression level in each tissue, hence generating 79 tissue-specific ranked gene lists. We generated four different sets of genes (S1−S4) for evaluation of expression enrichment (details in text) and significance analysis.

Enrichment of expression of each gene set ***S*** in a particular tissue and its significance was analyzed from the ranked list of genes ***T*** specific for that tissue after evaluating the nonrandomness of ranks of ***S*** within ***T***, using the Mann−Whitney rank sum statistic. After summing the ranks of ***S*** in list ***T***, we tested the significance of this rank sum against the rank sum of control set (all genes in ***T***, excluding ***S***). If μ and σ^2 are the mean and variance of the control set than enrichment (z-score) of ***S*** is given by $(\mu - S)/\sigma$, which measures enrichment in terms of number of standard deviations away from the mean of the control set. A z-score of ≥4.0 is considered to be significant in the present study.

Association of PG4 Motifs with Core Promoter Elements. 14252 human promoters (with at least one core promoter element and one PG4 motif) were considered and the percentage of PG4 motifs associated with core promoter elements checked, association was defined as overlapping and/or presence of core promoter element within 20 bases flanking (either side) each PG4 motif. Statistical significance of association was analyzed as detailed before from estimation of random expectancy of co-occurrence.[47] Briefly, the expected frequency of occurrence of sequence elements within an interdistance of m bp in a length of n bp sequence was computed within a reference promoter data set of size N. Then the actual occurrence frequency of each core promoter element in the same data set N was found along with the frequency of co-occurrence within m bp. The statistical significance in discrepancy of observed versus expected frequencies was estimated using degree of freedom 1 to exclude false positives (Bonferroni correction). Significance of the enrichment of PG4 motifs within CpG islands was tested by first calculating the expected number of PG4 motif occurrences within each CpG island and then comparing this with actual number of occurrences. Statistical significance of the discrepancy was analyzed using a simple χ-square test.

Gene Expression and Significance Analysis. HeLa S3 and A549 cells (from NCCS, Pune, India) were maintained under recommended conditions and treated with 100 μM TMPyP4 (Porphyrin Systems GbR, Germany) for 24 or 48 h prior to RNA isolation and hybridization on cDNA microarray slides (Human-19k, University Health Network, Canada). All treatments were checked for *c-MYC* repression due to TMPyP4 treatment by luciferase expression (Promega) using Del-4 plasmid (as reported previously,[30] Del-4 harbors −850 bp of *c-MYC* sequence with two major promoters P1 and P2 in a luciferase reporter cassette; a gift from Bert Vogelstein) and quantitative RT-PCR before hybridization. Cell viability at both time points was checked (MTT assay; 96.5% and 87.4% cells were viable at 24 and 48 h, respectively (average of experiments in triplicate)). Microarray experiments (in HeLa S3 cells only) were done in quadruplicate at each time point (biological replicates). SAM was run on normalized data ($\log_2$ ratio). Genes were retrieved at ≤20% false discovery rate (FDR). Gene Ontology analysis was done using GO ToolBox (http://www.geneontology.org).[48] HPRD (http://www.hprd.org/) was used for determining the multiple protein partners. We used quantitative RT-PCR (ABI 7900 from Applied Biosystems) to independently determine the level of gene expression using SYBR Green Assay in HeLa S3 and A549 cells with β−2-microglobulin as the endogenous control.

Further details of methods used in this study are provided as Supporting Information.

Results

First, we took a closer look at the prevalence of PG4 motifs near TSS in order discern any strand or chromosome specificity of G4 motif presence and also to confirm that the observed enrichment was due to sequence with structure forming potential. Two different methods were used to rule out artifactual prevalence of the G4 motif. First, we independently shuffled each of the 25706 human regions (±5 kb centered at TSS), 10000 times, keeping the nucleotide composition the same and searched for PG4 motifs. A random distribution was found with ∼10-fold decrease in overall PG4 density relative to actual occurrence (Figure 1B, lower panel). Second, to check whether the observed enrichment of PG4 motifs within ± 1 kb region centered at TSS was unique to motif forming patterns, we analyzed the distribution of a control pattern within the human genome (Figure 1B). The control sequence had three G-runs instead of four required for quadruplex formation and hence would not be able to adopt structure (see Supporting Information). We found ∼5-fold increase in PG4 density relative to the control-sequence density near TSS (±1 kb), supporting significance of the structural form (Figure 1B,C; $P < 0.0001$). About 2.5-fold enrichment of PG4 motif density relative to the control sequence was observed in regions beyond 1 kb of TSS (up to 5 kb upstream/downstream) and also when compared with sequence obtained from last two introns of human genes (Figure 1C, $P < 0.0001$). We selected more than 129 Mb of sequence from the last two introns to use as control regions as they are largely devoid of regulatory elements.[1] It is interesting to consider the control sequence vis-à-vis the PG4 motif. Both the control sequence and PG4 motif are expected to be influenced by GC-content by an approximately equivalent factor, therefore normalizing for the control sequence rules out any bias due to GC-content. This underscores the significance of PG4 occurrence in regions immediate to TSS.

Interestingly, we noted a sharp decrease in PG4 density at the TSS. Further analysis showed 480 human genes harbored a PG4 motif at the TSS, and these genes were over-represented in regulation of cell cycle (Gene Ontology (GO) analysis; $P < 0.005$). Interestingly, ∼77% (19809) and ∼55.5% (14267) of human promoters harbored at least one PG4 motif within the ±5 or ±1 kb region (with respect to TSS), respectively (Table 1). Similar prevalence was also found in the other genomes studied (Table 1). To check whether the PG4 motif enrichment had any chromosome or strand-specific characteristics, distribution of PG4 motif across all chromosomes in both strands was analyzed. PG4 motif enrichment near TSS was a general feature observed across all chromosomes (Figure 1D) and in both the strands (Supporting Information Figure S1). Similar observations were made in chimpanzee, rat, and mouse genomes (data not shown).

PG4 Motifs are Conserved in Promoters. To infer regulatory significance of G4 motif(s), which is a higher-order conformation, we searched for conservation of sequences that would form the structural PG4 motif but not necessarily have

identical sequence. For every PG4 motif present in ±2 kb of human TSS, corresponding rat and mouse promoters were searched for PG4 motif within ±200 bp regions with respect to the human PG4 motif (Methods). This window size was carefully selected based on a recent whole genome comparative study of human, mouse, chimpanzee, and dog, which showed that generally regulatory elements were conserved within a window of ±200−250 bases.[1] We found 1414 PG4 motifs within 773 human promoters that were conserved in corresponding promoters of orthologous mouse and rat genes. We reasoned that the set of 773 promoters harboring one or more conserved PG4 motifs had the maximum likelihood of being relevant in the context of G4 motif-mediated transcription and tested this aspect for characteristics of *cis*-regulatory sites in multiple ways.

First, we looked into the positional preference (with respect to TSS) as usually found for *cis*-regulatory elements. The 1414 conserved PG4 motifs observed within 773 human promoters were considered for this analysis. The respective positions of the 1414 PG4 motifs with respect to the TSS in each promoter was compared with positions of all other conserved PG4 motifs (in other promoters) in order to find if there are any preferred positions that are evident across many promoters. Figure 2A was the result of a cluster analysis, which shows distinct sets of promoters with preferred positions for PG4 motifs in a ±2 kb region flanking TSS across 773 human promoters. A detailed figure with all gene names is available online (http://quadbase.igib.res.in/conservedPG4.html). Enrichment of the conserved PG4 motifs near TSS was also observed in this case (upper panel, Figure 2A), consistent with the enrichment for all PG4 motifs near TSS observed earlier (Figure 1B). This analysis clearly indicated preferred positions for conserved PG4 motifs in most cases. We reasoned that the conserved PG4 motifs were most likely to be functionally relevant and checked the functional classes represented by the 773 human genes with conserved PG4 motif in promoter using Gene Ontology (GO) database.[49] Interestingly, significant over-representation ($P < 0.05$; after Bonferroni correction for multiple hypothesis testing) was observed in several important biological processes related to signaling, including cellular communication, regulation of cell-cycle, and phosphate/protein metabolism. All enriched GO categories are given in Supporting Information Table S2.

Genes with Conserved PG4 Motifs in Promoters Show Tissue-Specific Expression. Next, to evaluate the regulatory potential of PG4 motifs, we analyzed expression of the above set of 773 genes in human tissues. This was largely based on the understanding that regulatory control of a group of genes by a specific motif is likely to result in significantly enriched (or altered) expression response (either up or down regulation) within specific tissues relative to other randomly picked genes.[1] Using gene expression data from 79 human tissues,[46] we observed significantly enriched expression-response (z-score > 4.0, Methods) for the above gene set of 773 genes (S1) in most tissues (75/79, Figure 2B). We also tested a gene set S2, having at least one PG4 motif within ±1 kb centered at TSS (both conserved and nonconserved). Surprisingly, 66 out of 79 tissues exhibited significantly enriched expression for gene set S2 (Figure 2B), suggesting that even "nonconserved" PG4 motifs may be functionally relevant. Control gene sets, with PG4 motif occurrences beyond ±2 kb of TSS (S3) or with no motif(s) within ±2 kb centered at TSS (S4) showed no significant change in gene expression (Figure 2B, Methods).

PG4 Motifs are Associated with Core Promoter Elements. Enrichment of PG4 motifs near TSS and its putative regulatory potential prompted us to analyze association of PG4 motifs with known core promoter elements and also with CpG islands, which are known to be enriched in regulatory elements. Five core promoter elements were selected for this study, INR, DPE, BRE, CCAT, and TATA,[50] and the co-occurrence frequency and its statistical significance estimated (Methods). We found that DPE, BRE, and INR were significantly associated with the motif ($P < 0.005$) and 46.8% of PG4 motifs present on CpG containing promoters were found within CpG islands ($P < 0.0001$). Significant association with TATA box was not observed, which may suggest a preference of the PG4 motifs for TATA-less and DPE-enriched promoters; similar (∼10%) association with both DPE and INR is consistent with the finding that DPE/INR function cooperatively in the absence of TATA box.[51] Notably, a recent study also observed that active human promoters were not significantly associated with the TATA box element.[52]

PG4 Motifs Affect Gene Expression Genome-Wide. To test our prediction that G4 motif(s) may be involved in relatively widespread regulatory activity, we experimentally analyzed the effect of PG4 motifs on genome-wide expression. To do this, we selected human cervical carcinoma cells, HeLa S3, and treated with the intracellular G4-binding molecule TMPyP4[30,53,54] as a case study. We reasoned that this would allow us to largely distinguish the effect of G4 motifs, as intracellular ligand-induced stabilization/destabilization of motifs is expected to lead to perturbation of gene expression. The cationic porphyrin TMPyP4 has been studied extensively for its G4 motif binding.[54−60] However, on the other hand, TMPyP4 is also known to bind double-strand DNA. At least two lines of evidence indicated that TMPyP4 was sufficiently specific toward G4 motifs inside cells. First, direct and specific effect of TMPyP4 on G4 motif-mediated suppression of *c-MYC*[30] and recently PDGF-A[61] has been reported. It was demonstrated that on introducing specific nucleotide substitutions that disrupted the G4 motif in the promoter of *c-MYC*, the suppressive effect of TMPyP4 was lost.[30] Second, it was shown that TMPyP2 a close homologue of TMPyP4, which had lower affinity toward G4 motifs but not double-strand DNA did not suppress *c-MYC* expression.[30] Taken together, these studies demonstrated that the effect of TMPyP4 under intracellular conditions could be reasonably specific for G4 motifs. Furthermore, this is by far the only molecule that has been reported to stabilize G4 motifs intracellullarly, leading to change in gene transcription. Keeping these in mind, we selected TMPyP4 for testing genome-wide regulatory role of G4 motifs.

For all experiments reported here, we measured *c-MYC* levels by quantitative RT-PCR and luciferase reporter assays before hybridization on microarray slides and suppression was noted in all cases, consistent with earlier observations (Figure 3A).[30,37] We also used the concentration of TMPyP4 (100 μM) that was used earlier to report TMPyP4-mediated suppression of *c-MYC* promoter activity.[30] Many significantly differentially expressed genes (≤20% false discovery rate (FDR), using significant analysis of microarray (SAM, Methods[62]), harbored one or more PG4 motifs in ±5 kb centered at TSS (67/93 and 46/69 genes at 24 and 48 h, respectively; see Supporting Information Table S3), and 21 genes were common among the ones down-regulated at 24 and 48 h, confirming the reliability of expression profiles (Figure 3B,C). To further confirm the differential expression analysis from cDNA microarray, quantitative RT-PCR was performed on 20 randomly selected genes. It was observed that 18 of 20 cases (Figure 3D) show gene expression levels similar to those obtained from microarray analysis (while

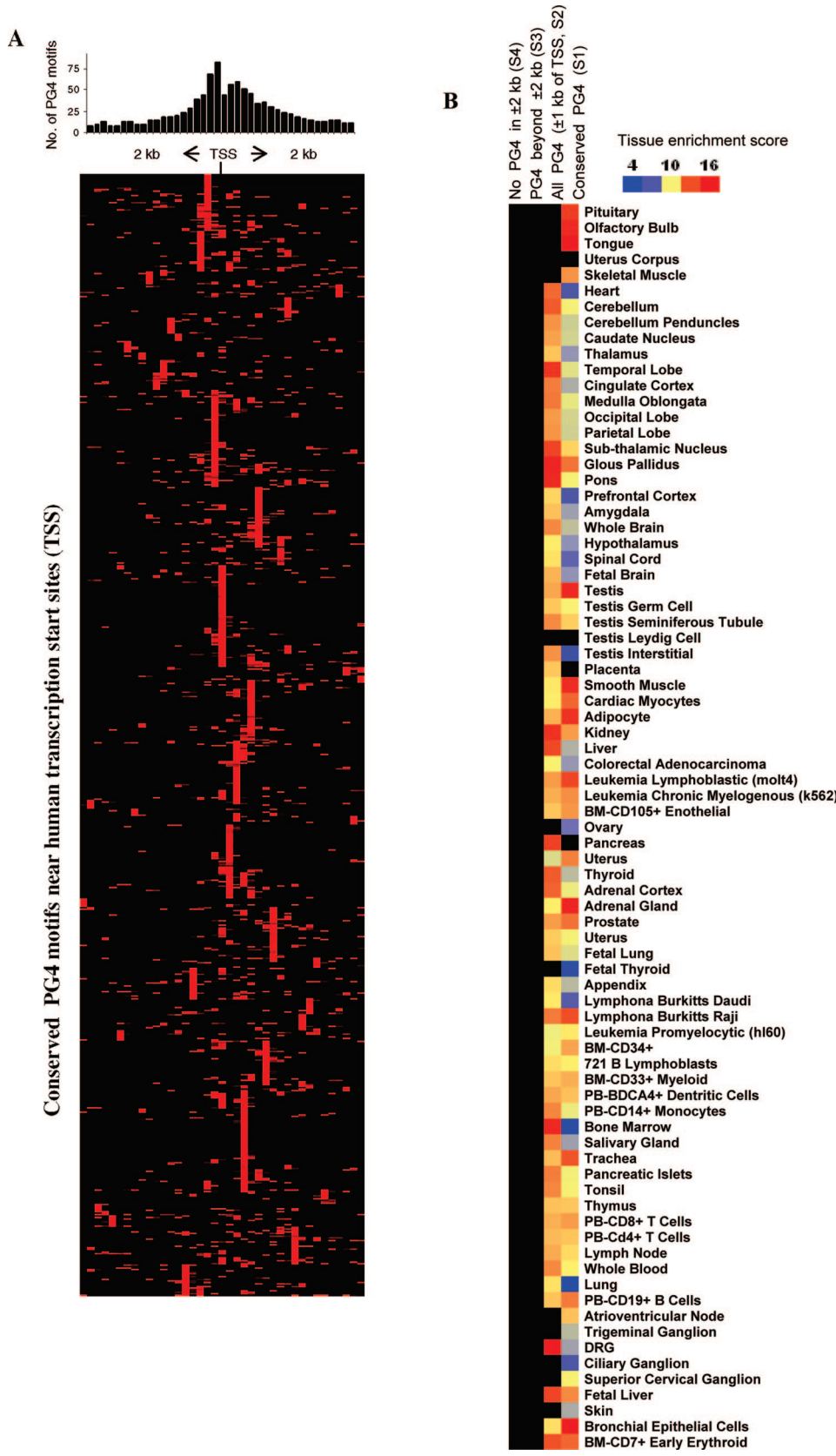

Figure 2. Functional relevance of PG4 motifs within promoters. (A) Conserved human PG4 motifs occur at preferred distances with respect to the TSS. Clustering of PG4 motifs in the array of 773 human promoters harboring 1414 PG4 motifs that are conserved with "orthologous" mouse and rat promoters (lower panel, red boxes indicate PG4 motif(s) in 100 bp window and each row represents one promoter). Upper panel: conserved PG4 motifs are enriched near TSS. Number of PG4 motifs found in each 100 bp window for all the 773 promoters is shown. (B) Genes harboring conserved PG4 motifs in promoters show enriched tissue-specific expression. Statistical enrichment (z score > 4.0) is represented in pseudocolor for different gene-sets (S1−S4) in 79 human tissues; S1: genes considered in (A); S2: genes with at least one PG4 motif (both conserved and nonconserved) within ±1 kb centered at TSS; S3: genes with PG4 motif beyond ±2 kb of TSS (up to 5 kb upstream and downstream); S4: no PG4 motif occurrence within ± 2 kb centered at TSS.

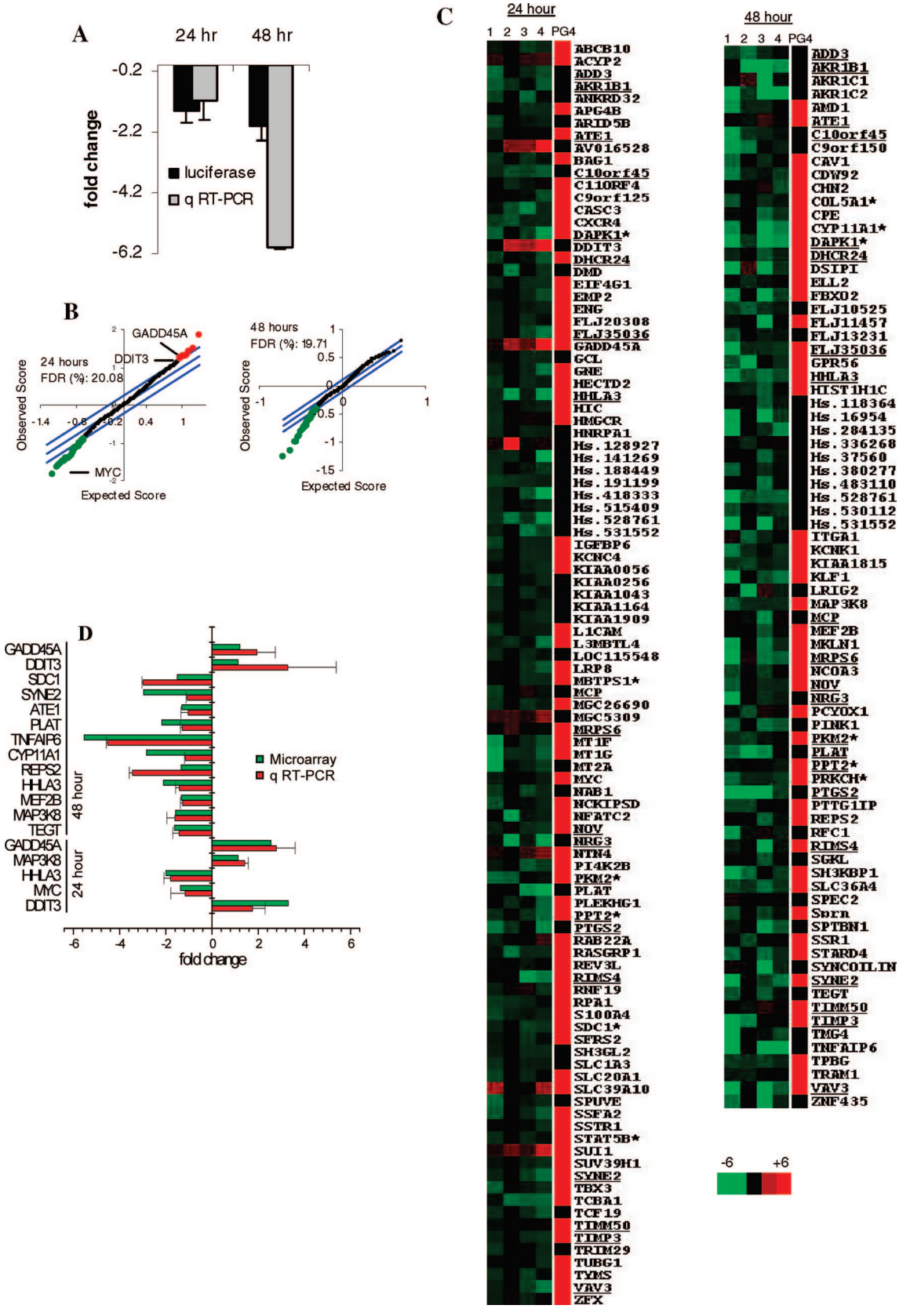

Figure 3. PG4 motifs induce gene expression changes. (A) *c-MYC* repression with 100 μM TMPyP4 was observed using luciferase expression (in Del-4 constructs, see Methods) and quantitative RT-PCR. Fold change is with respect to untreated cells. (B, C) Association of PG4 motifs with gene expression. (B) Greater or lower signal in observed data over expected signal in permuted data indicates selective amplification (in red) or suppression (in green) respectively, on treatment with TMPyP4. (C) Relative expression of probes with significant response (<20% FDR; see Supporting Information Table S3) in four replicates (columns 1−4). Presence (red) or absence (black; also not determined in case of transcripts with Hs IDs) of PG4 motif within 2 kb centered at TSS is shown. Genes with promoters harboring motifs that are conserved in mouse and rat (*) and are common (underlined) in the 24 and 48 h list are indicated. (D) Validation of microarray results by quantitative RT-PCR. All selected genes (16) were below 20% FDR at the respective time point except *MAP3K8* (at 24 h) and *SDC1*, *DDIT3*, and *GADD45A* (at 48 h).

for *AKR1B1* and *NOV* at 48 h, it was different); 10 of 12 reactions tested gave similar results in A549 (lung carcinoma) cells also (Supporting Information Table S4). We did not find any clear statistical enrichment (observed over expected) of genes with PG4 motifs in promoters among the genes that responded to TMPyP4 treatment. This is not surprising given that >75% of human genes have one or more PG4 motif(s) within ±5 kb of TSS (>50% on considering ±1 kb (Table 1)), which accounts for a substantially high number of genes that are expected to be affected by chance. Therefore, inference from statistical enrichment would be limited. Nevertheless, to gain an initial understanding of the biological implications and also to test the bioinformatics predictions indicating genome-wide regulatory role, we analyzed the biological implications of gene expression change after considering a reasonably conservative set of genes (≤20% FDR).

Biological Implications of Genes Regulated by PG4 Motifs. To estimate the overall effect of genes that are up or down-regulated in the microarray, we analyzed genes whose protein products were either transcription factors or transcription regulatory factors and/or have multiple binding partners. All these genes responded at ≤20% FDR and harbored PG4 motifs within their putative promoters. The widespread biological implications of affecting the expression of these genes is discussed below (also see Supporting Information for more details). Three genes responded at both 24 and 48 h, and also harbor orthologously conserved PG4 motifs, across human, mouse, and rat, within ±1 kb from TSS: (a) death associated protein kinase 1 (*DAPK1*), a positive regulator of interferon gamma induced apoptosis was most significant (0% FDR at both time points), (b) muscle specific pyruvate kinase 2 (*PKM2*; 6.3 and 16.4% FDR at 24 and 48 h, respectively), and (c) a lysosomal factor involved in fatty acid metabolism palmitoyl protein thioesterase 2 (*PPT2*; 0 and 3.84% FDR at 24 and 48 h, respectively). At 24 h, five genes coding for transcription factors; *STAT5B*, *HIC*, *TBX3*, *TCF19*, and *c-MYC* were down-regulated. STAT5B is a transcriptional activator, which mediates signal transduction triggered by IL2, IL4, CSF1, and other growth factors. TBX3 is a transcription repressor involved in limb development; TCF19 is required for late cell cycle progression, while HIC interacts with cyclin T1 and modulates P-TEFb-dependent transcription and MYC, which plays a critical role in cell cycle progression, apoptosis, and cellular transformation. All of these genes harbored one or more PG4 motifs within ±1kb of TSS. Three transcription regulatory factors were down-regulated at 24 h; *NFATC2*, *TRIM29*, and the X-linked zinc finger protein *ZFX*, which had PG4 motifs within either ±1 or ±5 kb (*TRIM29*) of TSS.

At 48 h, several down-regulated genes harboring one or more PG4 motifs in their promoters (within ±1 kb from TSS) are notable. The transcription factors: *ELL2*, *KLF1*, and *MEF2B* were down-regulated. Elongation factor RNA polymerase II, 2 (ELL2), which can increase the catalytic rate of RNA polymerase II transcription by suppressing transient pausing by the polymerase at multiple sites along the DNA and is frequently translocated in acute myeloid leukemia; Kruppel-like factor 1 (KLF1), which regulates the adult beta globin locus and the MADS box transcription enhancer 2 (MEF2B), which is involved in regulation of muscle specific genes. Three transcription regulatory proteins were down-regulated: delta sleep inducing peptide, immunoreactor (*DSIPI),* nuclear receptor coactivator 3 (*NCOA3*), and zinc finger protein (*ZNF435*). DSIPI has key role in anti-inflammatory and immunosuppressive effects; the nuclear receptor coactivator, NCOA3, which recruits p300/CBP-associated factor and CREB binding protein for transcriptional activation and the zinc finger protein ZNF435, with putative transcription regulatory activity. Furthermore, to identify "hub proteins" (factors that interact with multiple partners[63]) that are differentially expressed in our microarray, we focused on transcription regulatory factors present within our data set. Among the genes that are down-regulated at both the time-points, the protein products of *NOV*, overexpressed in nephroblastoma, has a putative role in cell adhesion and migration and binds to 15 other proteins. The protein product of oncogene *VAV3* interacts with 12 binding partners and participates in signal transduction processes. The tissue plasminogen factor, PLAT, and VAV3 interact with 12 other proteins (Supporting Information).

Discussion

Sequence analysis was used to test the possibility of any bias in observed PG4 enrichment near TSS due to sequence artifacts. Over-representation of genes harboring PG4 motifs within promoters in certain functional categories (Supporting Information Table S2) precluded this possibility because random presence was expected to show even distribution across categories. Several other observations supported statistical significance over random occurrence (Supporting Information). Next, we designed several bioinformatics experiments to test the relevance of enrichment of PG4 motifs in promoters in the context of regulatory activity. Finally, we extended our findings to whole genome expression studies and show wide-ranging biological implications of genes that could be regulated by G4 motifs. We found over 700 promoters of orthologous genes harbored conserved PG4 motifs. Genes harboring conserved PG4 motifs in putative promoters showed tissue-specific expression patterns supporting regulatory activity of the G4 motif and were also over-represented in several important functional classes related to cellular communication and signaling. Molecular function of the G4 motif as a *cis*-regulatory element was supported by widespread changes in gene expression observed in whole genome array experiments when treated with the G4 motif-binding ligand TMPyP4, lending support to genome-wide predictions.

Several approaches suggest that PG4 motifs show general characteristics of regulatory elements, particularly, preferred positions of conserved PG4 motifs with respect to TSS (positional clustering) and association with core promoter elements. On the other hand, it may be argued that this could be due to conserved occurrence of any G-rich transcription factor binding site and may not be necessarily due to a G4 motif. For example, conservation of G-rich SP1 sites could show such global patterns. Our findings do not exclude this possibility, it only indicates that there is a possibility that some of these sites could in effect present structural elements, which gains support from the observation that sequence patterns with specific substitutions that exclude structure formation (and otherwise are G-rich) show almost no enrichment near TSS (Figure 1B). Indeed, two recent studies observed that many of the SP1 sites upstream of TSS could contribute to G4 motifs.[41,64] This also raises the interesting possibility that formation of structural motifs may disrupt SP1 binding and therefore affect transcription.

We noted widespread suppression of gene expression on treatment of HeLa S3 cells with TMPyP4 (Figure 3C). It is interesting to consider this result in light of recent bioinformatics analyses correlating presence of G4 motifs downstream (500 bp) of TSS to enhanced gene expression, where authors postulated that G4 motif-mediated stabilization of the transcrip-

tion bubble leads to higher transcription rate.[40] Ligand-induced stabilization of G4 motifs, on the other hand, could lead to "jamming" of the transcription bubble resulting in stalling of RNAP II and therefore aborted transcription. This is in agreement with our experimental observation where we noted suppression of gene expression. Another recent study noted that enriched occurrence of sequence with potential to form G4 motifs was a conserved phenomenon within the first introns in human, mouse, chicken, frog, and zebra fish genomes.[64] Although in overall agreement with our findings, this study analyzed all intronic sequences of a particular genome at a time to derive the enrichment unlike the pairwise comparison of orthologously related genes considered by us.

Results presented herein (Table 1) show that more than 50% of the human promoters harbor one or more PG4 motifs in ±1 kb centered at TSS (Table 1). This is in line with previous findings.[38,39] However, the number of motifs actually rendering function may be limited because intramolecular G4 motif formation requires strand separation of the double-strand DNA. Therefore, in a chromosomal context, it may be envisaged that higher-order organization of chromosomal DNA, i.e., chromatin, is an important factor regulating formation/deformation of G4 motifs. Formation of non-B DNA motifs when chromosomal DNA is under physical stress and the consequent regulatory role of such motifs have been demonstrated for several operons in *Escherichia coli*.[7] In an earlier study, our results predicted role of G4 motifs in *E. coli* gene regulation when the genome is supercoiled, which increases the propensity for strand separation.[43] Keeping these in mind, to select PG4 motifs with maximum likelihood of having functional role, we looked for conservation in related species. Perhaps surprisingly, more than 700 such promoters were found. We noted that these promoters showed position-specific presence of PG4 motifs with respect to TSS within the human genome (Figure 2A), which was not observed when all PG4 motifs found in human promoters were taken (Figure 1D), supporting our strategy for selection of biologically meaningful G4 motifs and underscoring the importance of conserved PG4 motifs.

In conclusion, the current results envisage substantial implications of G4 motifs as *cis*-regulatory factors on gene expression. It is tempting to propose an additional/alternative form of regulation that involves formation/deformation of the G4 motif coupled with the state of chromatin (which must locally uncoil (or melt) and become single stranded for replication, repair, recombination and transcription), thereby leading to transcriptional activation/silencing of specific genes. Interestingly, a recent paper from the Levens' laboratory shows that formation of non-B DNA structures could directly result from supercoiling induced during transcription. The non-B structural forms, in turn, recruit structure-specific regulatory factors for specific gene regulatory control.[65] These observations further support the genome-wide implications of results presented herein.

Acknowledgment. We thank Munia Ganguli for discussion and comments on the manuscript and Swapan K. Das for suggestions with statistical analysis. This research was supported by fellowships (A.V., K.H., and F.M.) and grants to S.C. by CSIR. R.T. and R.H. are supported by fellowships from UGC and ICMR, respectively. We thank Sangeeta Thatai (LabIndia) for technical help. We also thank the reviewers for helping us to improve the manuscript. The microarray data sets are available from GEO (Gene Expression Omnibus) under accession number GSE5226.

Supporting Information Available: Details of control sequence design, statistical significance of identified motifs, and biological implication of findings including figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Xie, X.; Lu, J.; Kulbokas, E. J.; Golub, T. R.; Mootha, V.; Lindblad-Toh, K.; Lander, E. S.; Kellis, M. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **2005**, *434*, 338–345.
(2) Wells, R. D. Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* **2007**, *32*, 271–278.
(3) Bacolla, A.; Wells, R. D. Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* **2004**, *279*, 47411–47414.
(4) Pedersen, A. G.; Jensen, L. J.; Brunak, S.; Staerfeldt, H. H.; Ussery, D. W. A DNA structural atlas for *Escherichia coli*. *J. Mol. Biol.* **2000**, *299*, 907–930.
(5) Perez-Martin, J.; de Lorenzo, V. Clues and consequences of DNA bending in transcription. *Annu. Rev. Microbiol.* **1997**, *51*, 593–628.
(6) Sinden, R. R. *DNA: Structure and Function*; Academic Press: San Diego, CA, 1994.
(7) Hatfield, G. W.; Benham, C. J. DNA topology-mediated control of global gene expression in Escherichia coli. *Annu. Rev. Genet.* **2002**, *36*, 175–203.
(8) Rich, A.; Zhang, S. Timeline: Z-DNA: the long road to biological function. *Nat. Rev. Genet.* **2003**, *4*, 566–572.
(9) Beer, M. A.; Tavazoie, S. Predicting gene expression from sequence. *Cell* **2004**, *117*, 185–198.
(10) Wasserman, W. W.; Palumbo, M.; Thompson, W.; Fickett, J. W.; Lawrence, C. E. Human−mouse genome comparisons to locate regulatory sites. *Nat. Genet.* **2000**, *26*, 225–228.
(11) Florquin, K.; Saeys, Y.; Degroeve, S.; Rouze, P.; Van de, P. Y. Large-scale structural analysis of the core promoter in mammalian and plant genomes. *Nucleic Acids Res.* **2005**, *33*, 4255–4264.
(12) Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, 364–366.
(13) Balagurumoorthy, P.; Brahmachari, S. K. Structure and stability of human telomeric sequence. *J. Biol. Chem.* **1994**, *269*, 21858–21869.
(14) Simonsson, T. G-quadruplex DNA structures: variations on a theme 34. *Biol. Chem.* **2001**, *382*, 621–628.
(15) Jin, R. Z.; Breslauer, K. J.; Jones, R. A.; Gaffney, B. L. Tetraplex formation of a guanine-containing nonameric DNA fragment. *Science* **1990**, *250*, 543–546.
(16) Dunnick, W.; Hertz, G. Z.; Scappino, L.; Gritzmacher, C. DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.* **1993**, *21*, 365–372.
(17) Jeffreys, A. J.; Royle, N. J.; Wilson, V.; Wong, Z. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **1988**, *332*, 278–281.
(18) Weitzmann, M. N.; Woodford, K. J.; Usdin, K. DNA secondary structures and the evolution of hypervariable tandem arrays. *J. Biol. Chem.* **1997**, *272*, 9517–9523.
(19) Hanakahi, L. A.; Sun, H.; Maizels, N. High affinity interactions of nucleolin with G−G-paired rDNA. *J. Biol. Chem.* **1999**, *274*, 15908–15912.
(20) Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
(21) Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005**, *33*, 2901–2907.
(22) Howell, R. M.; Woodford, K. J.; Weitzmann, M. N.; Usdin, K. The chicken beta-globin gene promoter forms a novel "cinched" tetrahelical structure. *J. Biol. Chem.* **1996**, *271*, 5208–5214.
(23) Murchie, A. I.; Lilley, D. M. Retinoblastoma susceptibility genes contain 5′ sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids Res.* **1992**, *20*, 49–53.
(24) Catasti, P.; Chen, X.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. Structure−function correlations of the insulin-linked polymorphic region. *J. Mol. Biol.* **1996**, *264*, 534–545.
(25) Kilpatrick, M. W.; Torri, A.; Kang, D. S.; Engler, J. A.; Wells, R. D. Unusual DNA structures in the adenovirus genome. *J. Biol. Chem.* **1986**, *261*, 11350–11354.
(26) Ma, D.; Xing, Z.; Liu, B.; Pedigo, N. G.; Zimmer, S. G.; Bai, Z.; Postel, E. H.; Kaetzel, D. M. NM23-H1 and NM23-H2 repress transcriptional activities of nuclease-hypersensitive elements in the platelet-derived growth factor-A promoter. *J. Biol. Chem.* **2002**, *277*, 1560–1567.
(27) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative

DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589.
(28) De, A. R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* **2005**, *44*, 16341–16350.
(29) Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **2006**, *128*, 5404–5415.
(30) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593–11598.
(31) Simonsson, T.; Pecinka, P.; Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **1998**, *26*, 1167–1172.
(32) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847–854.
(33) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **2007**, *3*, 218.
(34) Wieland, M.; Hartig, J. S. RNA quadruplex-based modulation of gene expression. *Chem. Biol.* **2007**, *14*, 757–763.
(35) Riou, J. F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Megnin-Chanet, F.; Helene, C.; Mergny, J. L. Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2672–2677.
(36) Zhou, J. M.; Zhu, X. F.; Lu, Y. J.; Deng, R.; Huang, Z. S.; Mei, Y. P.; Wang, Y.; Huang, W. L.; Liu, Z. C.; Gu, L. Q.; Zeng, Y. X. Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines. *Oncogene* **2006**, *25*, 503–511.
(37) Grand, C. L.; Han, H.; Munoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo. *Mol. Cancer Ther.* **2002**, *1*, 565–573.
(38) Huppert, J. L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–413.
(39) Zhao, Y.; Du, Z.; Li, N. Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett.* **2007**, *581*, 1951–1956.
(40) Du, Z.; Zhao, Y.; Li, N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.* **2008**, *18*, 233–241.
(41) Todd, A. K.; Neidle, S. The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res.* **2008**, *36*, 2700–2704.
(42) Du, Z.; Kong, P.; Gao, Y.; Li, N. Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem. Biophys. Res. Commun.* **2007**, *354*, 1067–1070.
(43) Rawal, P.; Kummarasetti, V. B.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S. K.; Chowdhury, S. Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.* **2006**, *16*, 644–655.
(44) Yadav, V. K.; Abraham, J. K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. QuadBase: genome-wide database of G4 DNA: occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* **2008**, D381.
(45) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
(46) Su, A. I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K. A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; Cooke, M. P.; Walker, J. R.; Hogenesch, J. B. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6062–6067.
(47) Qiu, P.; Ding, W.; Jiang, Y.; Greene, J. R.; Wang, L. Computational analysis of composite regulatory elements. *Mamm. Genome* **2002**, *13*, 327–332.
(48) Martin, D.; Brun, C.; Remy, E.; Mouren, P.; Thieffry, D.; Jacq, B. GOToolBox: functional analysis of gene datasets based on gene ontology. *Genome Biol.* **2004**, *5*, R101.
(49) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29.
(50) Butler, J. E.; Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **2002**, *16*, 2583–2592.
(51) Kutach, A. K.; Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **2000**, *20*, 4754–4764.
(52) Kim, T. H.; Barrera, L. O.; Zheng, M.; Qu, C.; Singer, M. A.; Richmond, T. A.; Wu, Y.; Green, R. D.; Ren, B. A high-resolution map of active promoters in the human genome. *Nature* **2005**, *436*, 876–880.
(53) Izbicka, E.; Wheelhouse, R. T.; Raymond, E.; Davidson, K. K.; Lawrence, R. A.; Sun, D.; Windle, B. E.; Hurley, L. H.; Von Hoff, D. D. Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.* **1999**, *59*, 639–644.
(54) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709.
(55) Weisman-Shomer, P.; Cohen, E.; Hershco, I.; Khateb, S.; Wolfovitz-Barchad, O.; Hurley, L. H.; Fry, M. The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res.* **2003**, *31*, 3963–3970.
(56) Kim, M. Y.; Gleason-Guzman, M.; Izbicka, E.; Nishioka, D.; Hurley, L. H. The Different Biological Effects of Telomestatin and TMPyP4 Can Be Attributed to Their Selectivity for Interaction with Intramolecular or Intermolecular G-Quadruplex Structures. *Cancer Res.* **2003**, *63*, 3247–3256.
(57) Seenisamy, J.; Dexheimer, T.; Rezler, E.; Grand, C.; Bashyam, S.; Hurley, L. The differential effects of TMPyP4, TMPyP2, Se2SAP and telomestatin on c-MYC gene expression can be explained by their differential interactions with the G-quadruplex in the c-MYC silencer element. *AACR Meet. Abstr.* **2004**, *2004*, 531–53a.
(58) Wei, C.; Jia, G.; Yuan, J.; Feng, Z.; Li, C. A spectroscopic study on the interactions of porphyrin with G-quadruplex DNAs. *Biochemistry* **2006**, *45*, 6681–6691.
(59) Shammas, M. A.; Shmookler Reis, R. J.; Akiyama, M.; Koley, H.; Chauhan, D.; Hideshima, T.; Goyal, R. K.; Hurley, L. H.; Anderson, K. C.; Munshi, N. C. Telomerase inhibition and cell growth arrest by G-quadruplex interactive agent in multiple myeloma. *Mol. Cancer Ther.* **2003**, *2*, 825–833.
(60) Freyer, M. W.; Buscaglia, R.; Kaplan, K.; Cashman, D.; Hurley, L. H.; Lewis, E. A. Biophysical Studies of the c-MYC NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin. *Biophys. J.* **2007**, *92*, 2007–2015.
(61) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.* **2007**, *35*, 7698.
(62) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5116–5121.
(63) Han, J. D.; Bertin, N.; Hao, T.; Goldberg, D. S.; Berriz, G. F.; Zhang, L. V.; Dupuy, D.; Walhout, A. J.; Cusick, M. E.; Roth, F. P.; Vidal, M. Evidence for dynamically organized modularity in the yeast protein−protein interaction network. *Nature* **2004**, *430*, 88–93.
(64) Eddy, J.; Maizels, N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* **2008**, *36*, 1321–1333.
(65) Kouzine, F.; Sanford, S.; Elisha-Feil, Z.; Levens, D. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.* **2008**, *15*, 146–154.

JM800448A