Asymmetric Methylation in the Hypermethylated CpG Promoter Region of the Human L1 Retrotransposon*

(Received for publication, June 10, 1996, and in revised form, October 23, 1996)

David M. Woodcock‡, Celine B. Lawler, Martha E. Linsenmeyer, Judith P. Doherty, and William D. Warren

From the Sir Donald and Lady Trescowthick Research Laboratories, Peter MacCallum Cancer Institute, Locked Bag No. 1, A’Beckett Street, Melbourne, Victoria 3000, Australia

We have investigated the function and sequence specificity of DNA methylation in the hypermethylated CpG island promoter region of the endogenous human LINE-1 (L1) retrotransposon family. In nontransformed human embryonic fibroblasts, inhibition of DNA methylation with 5-aza-deoxycytidine induced a greater than 4-fold increase in transcription from potentially functional L1 elements without increasing the transcription level of the majority of degenerate elements, implicating hypermethylation in the repression of L1 activity. Using bisulfite genomic sequencing to assess the pattern of methylation in a subset of nondegenerate L1 elements, we found 29 sites within a 460-base pair region of the noncoding (top) DNA strand of the L1 promoter in which cytosine methylation was maintained with high efficiency. Of these, 25 were at CG dinucleotides and four were in non-CG sites. When the methylation sites were analyzed for the complementary (bottom) strand, the only highly conserved sites of methylation were in CG dinucleotides. Several of these sites of CG methylation in the bottom (coding) strand were at positions where top (noncoding) strand-derived sequences were unmethylated, suggesting that these sites might be maintained in a hemi-methylated state. Hence, there is a subset of human L1 elements in which methylation is efficiently maintained in asymmetric non-CG sites and further that this non-CG methylation may be part of a wider phenomenon involving hemi-methylation at CG dinucleotides. Maintenance of asymmetric methylation at non-CG sites (and possibly at hemi-methylated CG dinucleotides) could be through a novel DNA methyltransferase activity. Alternatively, the promoter region of L1 elements may be induced by factor binding to form some type of secondary structure that presents as a highly efficient substrate for de novo methylation.

Five to ten percent of the human genome is derived from one transposable element family, the L1 or LINE-1 family (1, 2), which belongs to the non-LTR retrotransposon class of elements. The majority of human L1 elements are inactive degenerate remnants, some are clearly functional, as the majority of human L1 elements are inactive degenerate remnants that are spread widely among eukaryotes. Although the majority of human L1 elements are inactive degenerate remnants, some are clearly functional, as de novo insertion of L1 elements have been documented in the germ line of both humans (3–5) and mice (6) as well as in somatic (tumor) cells (7). Nondegenerate full-length mammalian L1 elements are some 6 kilobases in length and contain two evolutionarily conserved open reading frames, the second of which encodes a reverse transcriptase (8, 9) with intrinsic RNase-H and AP endonuclease-like activity (10, 11). The promoter responsible for the full-length L1 transcript has been shown to be located within the 5’ end of the element and downstream from the transcriptional start site (12). This promoter region also contains a strong binding site for the ubiquitous transcription factor, YY1, which has recently been shown to be the nuclear matrix-associated protein, NMP-1 (13). In contrast to regions normally associated with the nuclear matrix that are high in AT bases (14), the 5’ end of the human L1 element is an atypical CpG island in that it is very heavily methylated in DNA from both the human embryonic fibroblast culture and the bisulfite genomic sequencing technique used in the experiments reported here in stimulated normal human peripheral lymphocytes in short term culture (15, 16). It seems plausible that the hypermethylation of the L1 promoter region is a cellular response to repress a genetic element that could be potentially very damaging if actively transcribed.

We were interested in (a) determining whether the hypermethylation of the promoter region of L1 elements might be involved in inactivating potentially functional elements and (b) determining the sequence specificity of this hypermethylation. To these ends, we have employed a sensitive RNase protection assay that allows the quantitation of transcription from a subset of nondegenerate, potentially functional elements in the presence of a high background of transcription from degenerate elements (17) and the bisulfite genomic sequencing technique that generates a positive signal for sites of DNA methylation in individual strands from genomic DNA (18). In this latter procedure, denatured genomic DNA is modified by reaction with bisulfite under conditions that convert all unmethylated cytosines to uracils (18, 19) and PCR1 primers specific for either the “top” or “bottom” DNA strand to amplify the sequence of interest. Using this approach we have been able to identify conserved sites of methylation in both CG and non-CG sites within the L1 promoter region.

EXPERIMENTAL PROCEDURES

Nontransformed human embryonic fibroblast culture (HEF) and the human teratocarcinoma cell line Ntera2D1 cells were maintained as described previously (17). To assess the effects of demethylation on the levels of L1-specific RNAs, cultures of HEF and human teratocarcinoma Ntera2D1 cells were grown in minimum Eagle’s medium-a and treated with 5-aza-dC (Sigma) at a concentration of 1 μM, which was previously

---

* This work was performed with the aid of a grant (to D. M. W.) from the National Health and Medical Research Council of Australia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Peter MacCallum Cancer Institute, Locked Bag No. 1, A’Beckett St., Melbourne, Victoria 3000, Australia. Tel.: +61 (3) 9656 1359; Fax: +61 (3) 9656 1411; Email: woodcock@petermac.unimelb.edu.au.

1 The abbreviations used are: PCR, polymerase chain reaction; 5-aza-dC: 5-aza-2’-deoxycytidine; HEF, human embryonic fibroblasts; M.SssI, SssI methylase.
shown to almost totally inhibit post-replicative DNA methylation (20). 5-AzadC was added to the media when the cell monolayer was approximately 30–40% confluent and incubated at 37 °C until the cultures reached 80–90% confluence. The media was changed every 24 h since 5-azadC has a limited half-life in culture media (21). RNA was extracted using the guanidine lysis method (22). Slot blot quantification of total L1 transcription was performed by RNA immobilization on Bio-Rad Zetaprobe membrane followed by hybridization with 32P-labeled probe (23). RNase protection experiments for the detection of specific L1 transcripts was performed using published procedures (17), and hybridization signals were quantified using a Molecular Dynamics PhosphorImager.

DNA from HEF cells was modified by bisulfite using standard procedures (18), except that the modification reaction was performed at 55 °C and was cycled up to 94 °C for 5 min every 3 h for varying times ranging up to 48 h (24), although it was later found that bisulfite modification was essentially complete in 5–6 h when cycled every hour. Bisulfite-modified DNA was recovered using Wizard resin (Promega) and desulfonated (18) prior to PCR amplification. PCR primers, designed so as to be able to amplify from the top (noncoding) strand of both bisulfite-modified and native (unmodified) DNAs, were 5'-GGGAGGAGGAGGTTAAGATTGTCATG-3' and 5'-CTCCACACCATCATCAAACT-3' and correspond to bases 1–22 and 505–484 of our previously derived L1 consensus (GenBank accession X58075) (16), respectively. For the modification of the bottom (coding) strand, primers corresponding to bases 30–50 and 464–441 of the L1 consensus, were used. These primers were designed to allow the complement of the bisulfite-modified form of the consensus generated from the L1 sequences amplified from native DNA using the top strand primers containing the pattern of methylation observed in the top strand clones from bisulfite-modified DNA. PCR-generated DNA fragments were ligated into the pGEM-T cloning vector (Promega) and electroporated into DH12S cells (Life Technologies, Inc.) using standard protocols. By cloning PCR products amplified from bisulfite-modified DNA using this approach, we generally obtained L1 element-derived insert DNA in 30–70% of randomly selected white colonies. We found that transformation by electroporation was essential for the success of this technique as standard chemical transformation protocols yielded L1 inserts in only a few percent of white colonies, resulting in an unacceptable cloning bias. Plasmid clones were sequenced using either the Pharmacia ALF or the ABI 377 system using forward and reverse M13 primers giving complete or almost complete sequence from both strands. Sequence data analyses were performed using the GCG program suite (University of Wisconsin version 8).

Plasmid DNA containing a truncated human L1 element (clone pA41) (16) was added to HEF genomic DNA prior to each modification reaction as an internal control. PCR amplification of these control sequences was performed as above using the primer corresponding to bases 505–484 of the L1 consensus in combination with a primer (5'-GGGAGGAGGAGGTTAAGATTGTCATG-3') specific for the flanking plasmid vector DNA sequence. Of 20 cloned sequences amplified from this control plasmid (derived from eight separate modification reactions), a total of three unmodified cytosines were found, indicating an overall cytosine conversion efficiency greater than 99.9%. This total excludes cytosine methylation in dcm sites (the inner C of CG(A/T)GG) observed in otherwise fully bisulfite-modified sequences amplified from pA41 plasmid grown in the dcm- Escherichia coli host DH12S. When the pA41 plasmid was grown in the dcm+ host strain GM2163 (25), no dcm methylation was observed in the PCR-amplified control sequences.

For the control experiments to test for bias in the cloning step, pA41 DNA was methylated at all CG dinucleotides with M.SsIl and mixed with an equal amount of unmethylated pA41 DNA before bisulfite modification. PCR amplification was performed with the 505–484 consensus primer used for both genomic L1 products and pA41 controls in combination with a mixture of two plasmid-specific primers (equal quantities of the primer used above for amplification of unmethylated pA41 in combination with a primer 5'-TACTAGGGCGAATTCGAGCCTGG-3'-specific for the M.SsIl methylated form of the flanking plasmid vector DNA sequence).

The full sets of data from all of the sequences analyzed in this study are available on request from the corresponding author via anonymous FTP.

RESULTS

To test whether DNA methylation plays a role in maintaining potentially functional L1 elements in a quiescent state, both nontransformed (limited lifespan) human embryonic fibroblasts (HEF) cells and a human teratocarcinoma cell line (Ntera2D1, in which L1 elements are known to be transcriptionally activated (26)) were treated with the inhibitor of post-replicative DNA methylation, 5-azadC. The effect of 5-azadC on the levels of L1 element transcription was quantitated by RNase protection assay using a probe that only detects transcripts from nondegenerate, potentially functional elements (17). Treatment of HEF cells with 5-azadC resulted in a >4-fold increase in RNA transcripts from potentially functional L1 elements, whereas 5-azadC treatment of NteraD1 cells, in which L1 transcripts are already relatively abundant, resulted in a >2-fold increase (Fig. 1). Although, with the majority of human cell types, the overwhelming majority of L1 transcripts are from nonspecific readthrough from the promoters of other genes, in Ntera2D1 cells (one of only a few cell lines in which discrete L1 transcripts on Northern analysis can be detected), primer extension experiments show a predominant start site for transcription coinciding with the 5'-end of L1 elements (17, 26). To test whether this increase in L1 RNA protection products from transcripts corresponding closely to functional L1 elements might be the result of a general stimulation of non-specific transcription through L1 elements, total levels of sense transcripts from nondegenerate, potentially functional elements were determined using slot blot hybridization with total cellular RNA. The sense strand probe allowed quantitation of transcripts from all L1 elements (not just those conforming closely to the consensus for functional elements), whereas the antisense probe only detects so-called “junk” L1 transcripts produced by read-through from nearby promoters and is thus a control for nonspecific activation of transcription throughout the genome. These data, shown in Fig. 1, indicate that 5-azadC-induced DNA demethylation causes specific transcription from potentially functional (non-degenerate) L1 elements but has no significant effect on the level of overall transcription or total L1 element transcription in either cell type.

To elucidate the sequence specificity of DNA methylation in the promoter region of potentially functional human L1 elements, we employed the bisulfite genomic sequencing method that allows unambiguous identification of sites of methylated
cytosines in clones derived from individual DNA strands. In this procedure, genomic DNA is modified with bisulfite under conditions where unmethylated cytosines are converted to uracils (18, 19). Individual L1 strands are then amplified by PCR. (The two strands are no longer complementary after conversion of unmethylated cytosines to uracils.) Following the PCR amplification, uracils derived from the unmethylated cytosines are replaced by thymines, and the size-fractionated PCR products are cloned into a plasmid vector and the DNA sequence of individual clones determined.

The top strand PCR primers employed in this study were designed to match our previously derived consensus sequence for the 5' end of the human L1 element (accession no. X58075) (16). This consensus was compiled from a series of random human L1 clones made using optimally methylation-tolerant (mcr-) E. coli host strains so as to eliminate cloning bias. Within the region amplified with these primers, this PCR-derived consensus is 98.8% identical with the sequence of the L1.2b element that was the direct progenitor of a de novo L1 insertion in the factor VIII gene in the germ line of a hemophilic (27). We were thus aiming to determine the methylation status of potentially functional L1 elements such as those whose transcriptional activity would be detected in the RNase protection assay.

The positions of the top strand L1 primers were chosen on the basis of being high in guanines and low in cytosines. Incorporation of appropriate degeneracies at the single cytosine within each of these PCR primers allowed us to use the same primer combination to amplify nondegenerate L1 sequences from either bisulfite-modified or native (unmodified) genomic DNA. While this primer design limited what regions could be targeted for amplification, it avoided any selective amplification of a minority of unusually modified sequences based on some presumption of methylation status. Furthermore, the cloning strategy employed (see below) allowed the unbiased recovery of clones from PCR products from both methylated and unmethylated forms of the target sequence with high efficiency.

Twenty-five L1 clones from native (not bisulfite-modified) DNA isolated from HEF cells were sequenced. A representative selection of these sequences have been deposited in GenBank (accessions U68333 to U68349 inclusive). Fig. 2A shows an alignment of a portion of these sequences (bases 161–210) that is indicative of the features of the total sequence alignment. None of the sequences were identical, indicating that they were derived from different individual L1 elements. From these sequences, we were able to generate a consensus sequence for that subset of L1 elements amenable to using these PCR primers. We later compared this consensus from clones from unmodified DNA with clones from bisulfite-modified DNA. Sequences amplified from the bottom L1 strand are present in this region.

Data shown are for a representative section (bases 161–210) of the total length of each sequence. Sites of methylation in CG dinucleotides (circles) (both symmetric and asymmetric) and one of the sites of non-CG methylation (triangle) are present in this region. A, sequences amplified from native (not bisulfite-modified) DNA; B, top strand sequences amplified from bisulfite-modified DNA; and C, bottom strand sequences amplified from bisulfite-modified DNA. Sequences amplified from the bottom L1 strand are presented in a reverse-complemented orientation (bisulfite-resistant cytosines appear as guanines).

Fig. 2. Alignment of PCR-amplified sequences from the 5’ promoter region of human L1 elements. Data shown are for a representative section (bases 161–210) of the total length of each sequence. Sites of methylation in CG dinucleotides (circles) (both symmetric and asymmetric) and one of the sites of non-CG methylation (triangle) are present in this region. A, sequences amplified from native (not bisulfite-modified) DNA; B, top strand sequences amplified from bisulfite-modified DNA; and C, bottom strand sequences amplified from bisulfite-modified DNA. Sequences amplified from the bottom L1 strand are presented in a reverse-complemented orientation (bisulfite-resistant cytosines appear as guanines).
pairs from a site of non-CG methylation (Fig. 2, bases 188 and 183, respectively). Fig. 4 shows aligned sequence electropherograms from representative clones derived from bisulfite-modified human DNA, native DNA, and unmethylated control plasmid from a region containing putative sites of methylation in both CG and non-CG contexts (bases 399–452). This illustrates complete conversion of cytosines in the L1 plasmid control and unambiguous residual cytosines (sites of methylation) in both CG and non-CG dinucleotides in this region of the clone from bisulfite-modified genomic DNA. The sequence contexts of all four of the non-CG methylated bases are presented in Table I. Two of the 9 clones also contained a single cytosine outside the conserved unreactive sites, suggesting that our overall cytosine conversion efficiency at these other sites was at least 99.8% (although these 2 sites equally might represent actual sites of DNA methylation).

To analyze the sites of genomic methylation in the bottom (coding) strand of L1 elements, new PCR primers were designed to amplify preferentially from the bisulfite-modified L1 sequences belonging to the same subset of elements that were amplified using the top strand primers. The sequence of these “bottom” strand primers was based on the PCR-derived consensus sequence determined from unmodified DNA and from the conserved top strand methylation sites determined from modified HEF DNA as described above. Using bisulfite-modified HEF DNA as template, we used these primers to PCR-amplify and clone L1 bottom strand sequences from which the base sequence of 19 individual clones was determined. Detailed sequence data are shown for the same representative portion (bases 161–210) of these sequences (Fig. 2C). However, for easier comparison with the other sequence data, the sequences from these modified bottom strands are shown as their complementary top strand sequence. Hence, residual (methylated) cytosines in the bottom strands thus appear as the complementary guanines in the sequences as presented.

The occurrence of bisulfite-resistant cytosines in these bottom strand L1 sequences was more variable than was observed with the bisulfite-modified top strand data (Figs. 2 and 3). All 23 of the methylated CG sites in the top strand within the region amplified by the bottom strand primers had corresponding methylation CG sites in the bottom strand (i.e., were symmetrically methylated). Taking the set of bottom strand clones as a whole, there were on average 1.5 sites of non-CG methylation per clone. There were no consistently methylated non-CG sites identified with the possible exception of bases 432,433 where 2/19 clones had 5'-mCmCAG-3' sequences (as read on the bottom strand). However, there were several CG sites in which 30–70% of bottom strand sequences had bisulfite-resistant cytosines (shown as the complementary guanines in Fig. 2) for which no methylated cytosines were observed in CG dinucleotides in the top strand data (Figs. 2 and 3).

To investigate the possibility that our data are affected by selection bias introduced by the cloning procedures used to isolate L1 sequences isolated from bisulfite-modified DNA (and, in particular whether we were preferentially losing high A/T containing clones from unmethylated L1 elements), we performed a control experiment in which DNA from the truncated L1 clone (pA41) was methylated at all CG dinucleotides with M.SssI, mixed with an equal amount of unmethylated pA41 plasmid DNA, and then modified with bisulfite and simultaneously amplified by PCR from both the methylated and unmethylated forms of the sequence (see “Experimental Procedures”). After cloning of the mixed PCR product in pGEM-T, 20
Specific Non-CG Methylation in the Human Genome

TABLE I

Base assignments from sequencing at consistently unmodified cytosines in non-CG sites in clones from unmodified and bisulfite-modified human L1 elements

| Site I | Site II | Site III |
|--------|---------|----------|
| Bisulfite-modified | Sequence (104–124 bp)* | No. of clones | Bisulfite-modified | Sequence (178–198 bp) | No. of clones | Bisulfite-modified | Sequence (399–423 bp) | No. of clones |
| No | gggtccatct–Cacttagggagt | 7 | No | ca-gggccagggCattgcctcac | 9 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 3 | No | ca-gggtcagggCattgcctcac | 2 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 3 | No | ca-gggtcagggCattgcctcac | 2 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 2 | No | ca-gggccagggCattgcctcac | 1 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 1 | No | cagggccagggCattgcctcac | 1 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 1 | No | cagggccagggCattgcctcac | 1 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 1 | No | cagggccagggCattgcctcac | 1 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 1 | No | cagggccagggCattgcctcac | 1 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| No | gggtccatct–Cacttagggagt | 1 | No | cagggccagggCattgcctcac | 1 | No | agttcagagatCaa–Tctgaacggcg | 2 |
| Yes | gggtttattt–Cattagggagt | 7 | Yes | ta-gggtcagggCatttccttat | 5 | No | agttcagagatCaa–Tctgaacggcg | 1 |
| Yes | gggtttattt–Cattagggagt | 1 | No | ta-gggtcagggCatttccttat | 5 | No | agttcagagatCaa–Tctgaacggcg | 1 |
| Yes | gggtttattt–Cattagggagt | 1 | No | ta-gggtcagggCatttccttat | 5 | No | agttcagagatCaa–Tctgaacggcg | 1 |
| Yes | gggtttattt–Cattagggagt | 1 | No | ta-gggtcagggCatttccttat | 5 | No | agttcagagatCaa–Tctgaacggcg | 1 |
| Yes | gggtttattt–Cattagggagt | 1 | No | ta-gggtcagggCatttccttat | 5 | No | agttcagagatCaa–Tctgaacggcg | 1 |
| Yes | gggtttattt–Cattagggagt | 1 | No | ta-gggtcagggCatttccttat | 5 | No | agttcagagatCaa–Tctgaacggcg | 1 |
| Consensus | gggtccatct–Cacttagggagt | cagggccagggCattgcctcac | agttcagagatCaa–Tctgaacggcg |

* bp, base pair(s).

randomly selected clones with the correct sized inserts were sequenced. Of these clones, 10 were found to be derived from M.SssI-modified pA41 DNA and 10 from unmethylated pA41, indicating that there is no selection against sequences derived from the unmethylated form of a target sequence amplified and cloned under these conditions.

DISCUSSION

The principle model for the replication of mammalian DNA methylation has been that specific patterns of cytosine methylation will only be stably maintained following cell division if methylation occurs in CG dinucleotides where the methylcytosine in the parental strand acts as a template for the addition of a methyl group to the palindromic C in the daughter strand (36). Also, it has been shown in mammalian cells that there can be active demethylation of DNA (37–39), although the actual process involved remains unclear at this time. Hence, the simple model for the maintenance of DNA methylation in mammalian cells (28, 29) is likely to require some elaboration, at least for some mammalian sequence elements.

Our data indicate that, in the top strand of some human L1 elements, there are several sites in which asymmetric methylation is faithfully maintained in non-CG sites within the hypermethylated CpG island region of the L1 promoter. When we attempted to amplify selectively the bottom strands from this same subset of L1 elements using primers specific for the bisulfite-modified complement of the top strand clones, we found no consistently methylated non-CG sites in bottom strand-derived clones but several conserved sites of methylation at CG dinucleotides where there was no corresponding site of methylation in top strand-derived clones. Moreover, we observed two sites in the top strand in which a well maintained methylcytosine was present while, in bottom strand clones, a methylcytosine was only present in a small minority of clones (10 or 15%) (Fig. 2). These data indicate (i) that there are specific sites of asymmetric methylation in non-CG sites at least some L1 elements and (ii) that this may be part of a wider phenomenon involving hemi-methylation in CG sites in both top and bottom strands of these elements.

Reports of hemi-methylation in the DNA of eukaryotic cells are not unprecedented. In addition to that of Toth et al. (32) of a hemi-methylated site in integrated adenovirus DNA in a mammalian cell line, hemi-methylation has also been detected in plant DNA (40). This was detected by genomic sequencing of a transgene in the context of gene silencing, although the mechanisms and functions of DNA methylation in plants may be significantly different from mammalian cells.

There are several potential sources of artifacts in relation to (a) the bisulfite genomic sequencing method and (b) the complexity of amplifying from multicopy gene families. The potential sources of artifacts and the measures taken to avoid them...
are as follows.

Incomplete Bisulfite Modification—The residual cytosines in non-CG sites in the top strand-derived sequences could be artifacts generated by the bisulfite modification process. We believe that this is unlikely as we have carefully controlled for partial bisulfite modification by adding plasmid DNA containing sequences of a truncated L1 element to the HEF genomic DNA prior to each bisulfite modification reaction. The top and bottom strand L1 sequences reported here were amplified from bisulfite modification reactions where control DNA clones were found to be fully modified. Also, there were no signs of intrinsic unreactivity at these or any other sites in the control clones.

Selective PCR Amplification—Depending on assumptions made as to what sites are methylated, primers could be selectively amplifying from a sub-group of sequences with a particular methylation status. We have designed primers that allow PCR amplification from both modified and unmodified versions of target sequences to avoid this bias.

Selective Recovery of Clones from Methylated Sequences—As the bisulfite modification process produces sequences high in AT content, propagation of clones from originally unmethylated sequences might be selected against E. coli. This is unlikely under our experimental conditions since we found that, using high efficiency electrocompetent E. coli hosts, clones from unmethylated and methylated forms of a target sequence were recovered with equal efficiency.

Selective Cloning of Different Sub-families of L1 Elements—For the initial amplification of top strand sequences, we specifically attempted to determine the methylation status of L1 elements that were close to the consensus for a functional transposon. The subsequent design of the primers for the bottom strand amplification was based upon the sequences and methylation sites determined from the top strand data and should have biased amplification toward DNA strands complementary to these sequences. When dealing with multicyclop families of related sequence elements, it is not possible to eliminate the possibility that clones were amplified from different sub-families of elements, despite all attempts to bias PCR amplification toward the strands from a specific sub-group of elements. Although bisulfite modification-induced conversion of unmethylated cytosines to thymines limits the power of comparisons, alignment of the sequences from the modified top and bottom strands with native DNA clones (Fig. 2) is consistent with the top and bottom strand data being derived from extremely similar sequences. L1 elements have been shown to undergo de novo methylation in a uniform and concerted manner during differentiation both in vivo (41, 42) and in vitro (43). Extremely similar sequences would thus be expected to acquire very similar methylation patterns.

As to other explanations for the asymmetric non-CG methylation in this region of nondegenerate L1 elements, this is unlikely to have resulted from some form of highly efficient de novo methylation induced by hypermethylation of this region since 14% of CG dinucleotides in the top strand remained consistently unmethylated. It would seem improbable that, with the documented preference of the mammalian DNA methyltransferase for CG dinucleotides for both maintenance and de novo methylation (35), this enzyme would methylate de novo with very high efficiency at some nonpreferred substrates (non-CG sites), whereas some adjacent CG dinucleotides remained unmethylated. Furthermore, there does not appear to be anything special about the L1 sequence as a substrate for DNA methyltransferase per se. When the L1 sequence from one of the unmodified L1 clones (5μm52, Fig. 2A) corresponding closely to the native PCR-derived consensus was methylated in vitro with the human placental DNA methyltransferase and the sites of methylation subsequently determined by bisulfite genomic sequencing, it was found that this sequence was a poor substrate for de novo methylation with only 1 to 4 (apparently randomly) methylated sites being present in each of the clones analyzed.2

It has been demonstrated that human DNA methyltransferase will methylate de novo with high efficiency at cytosines in regions of secondary structure (44) or in single-stranded DNA (45). At such sites, the cytosine might be mimicking the structure of a reaction intermediate for methylation, particularly if the cytosine is partially extruded from the DNA duplex (46). Examination of the regions surrounding the sites of non-CG methylation in the L1 element failed to reveal any obvious hairpin or stem-loop structures. However, the close apposition of a non-CG methylation site in the top strand and an apparently hemi-methylated site in the bottom strand (positions 188 and 183, respectively) suggests that, if there were slippage between the DNA strands in this region, this could generate a functionally symmetrically methylated site (Fig. 5). Such a “dogleg” structure in a region close to a known binding site of a nuclear matrix-associated transcription factor might simply be viewed as an extreme example of DNA bending that is increasingly being shown to be induced by transcription factor binding, including YY1 (NMP-1) (47, 48).

Alternatively, this “unusual” specificity of methylation might be due to some as yet undocumented accessory protein(s) that acts to alter the specificity of the mammalian DNA methyltransferase or it could be due to another DNA methyltransferase in human cells that methylates specifically in heterochromatic “structural” regions of the genome rather than in the more open regions containing coding sequences. A precedent for this might be the maintenance of methylation in CNG sites in transfected plasmid DNA sequences in mouse cell lines (19).

Whatever the mechanistic origin of this asymmetric methylation in non-CG sites (and possibly also sites of hemi-methylation), the high methylation levels in the region of the promoter domain of the human L1 element seem most likely to be involved in repressing the transcription of these retrotransposons in normal cells. The sequestration of such potentially damaging genetic elements in heterochromatin could confer a significant selective advantage to the cell and to the organism as a whole. It remains to be determined what are the relative contributions of “conventional” CG methylation, non-CG (asymmetric) methylation, and DNA secondary structure to this process.

Acknowledgment—We thank Dr. S. S. Smith for the gift of human DNA methyltransferase.

REFERENCES

1. Hutchison, C. A., III, Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. (1989) in Mobile DNA (Berg, D. E., and Howe, M. M., eds) pp. 393–636.

2. M. Linsenmeyer, unpublished observations.
American Society for Microbiology, Washington, D.C.

1. Singer, M.F., Krek, V., McMullin, J.P., Swergold, G.D., and Thayer, R.E. (1993) Gene 135, 183–188
2. Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988) Nature 332, 164–166
3. Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Minami, R., Nakamura, H., and Matsuo, M. (1993) J. Clin. Invest. 91, 1862–1867
4. Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994) Nat. Genet. 7, 136–142
5. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1994) Cancer Res. 54, 643–645
6. Woodcock, D.M., Crowther, P.J., Doherty, J.P., Orian, J.M., Brandon, M.R., and Adams, T.E. (1995) Transgenic Res. 4, 114–122
7. Smith, S.S., Lingeman, R.G., and Kaplan, B.E. (1992) Biochemistry 31, 850–854
8. Lei, H., Oh, S.P., Okano, M., Jutterman, A., Goss, K.A., Jaenisch, R., and Li, E. (1996) Development 122, 3195–3205
9. Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A., and Cedar, H. (1991) Nature 351, 239–241
10. Sanford, J.P., Clark, H.J., Chapman, V.M., and Rossant, J. (1987) Genes Dev. 1, 1039–1046
11. Br. J. Cancer 65, 471–475
12. Mol. Cell. Biol. 14, 5636–5644
13. Proc. Natl. Acad. Sci. U.S.A. 82, 6050–6054
14. Proc. Natl. Acad. Sci. U.S.A. 87, 8252–8256
15. Proc. Natl. Acad. Sci. U.S.A. 91, 10502–10506
16. Biochemistry 31, 9–25
17. J. Mol. Biol. 261, 8325–8326
18. Science 254, 1805–1808
19. J. Biol. Chem. 270, 24623–24626
20. Proc. Natl. Acad. Sci. U.S.A. 91, 10562–10566
21. Biophys. Res. Commun. 145, 888–894
22. Nucleic Acids Res. 14, 4353–4367
23. Nucleic Acids Res. 2395–2401
24. Nucleic Acids Res. 183–188
25. Nucleic Acids Res. 3469–3478
26. Nucleic Acids Res. 3469–3478
27. Nucleic Acids Res. 3469–3478
28. Nucleic Acids Res. 3469–3478
29. Nucleic Acids Res. 3469–3478
30. Nucleic Acids Res. 3469–3478
31. Nucleic Acids Res. 3469–3478
32. Nucleic Acids Res. 3469–3478
33. Nucleic Acids Res. 3469–3478
34. Nucleic Acids Res. 3469–3478
35. Nucleic Acids Res. 3469–3478
36. Nucleic Acids Res. 3469–3478
37. Nucleic Acids Res. 3469–3478
38. Nucleic Acids Res. 3469–3478
39. Nucleic Acids Res. 3469–3478
40. Nucleic Acids Res. 3469–3478
41. Nucleic Acids Res. 3469–3478
42. Nucleic Acids Res. 3469–3478
43. Nucleic Acids Res. 3469–3478
44. Nucleic Acids Res. 3469–3478
45. Nucleic Acids Res. 3469–3478