RglB facilitated cloning of highly methylated eukaryotic DNA: the human L1 transposon, plant DNA, and DNA methylated *in vitro* with human DNA methyltransferase

D.M.Woodcock*, P.J.Crowther, W.P.Diver, M.Graham¹, C.Bateman¹, D.J.Baker² and S.S.Smith²

Molecular Science Group, Peter MacCallum Cancer Institute, 481 Little Lonsdale St., Melbourne, Victoria 3000, Australia, ¹Calgene Pacific, PO Box 53, Ivanhoe, Victoria 3079, Australia and ²Molecular Surgery, City of Hope National Medical Center, Duarte, CA 91010, USA

Received February 4, 1988; Revised and Accepted March 28, 1988

ABSTRACT

In vitro methylation of Bluescribe plasmid DNA (pBS) with human placental DNA methyltransferase to 6% 5-methylcytosine (mC) reduced transformation efficiencies in rglB⁺ host strains C600 and DS410 by almost 2 orders of magnitude. By contrast, the rglB⁻ derivative of DS410 showed no reduction in transformation efficiency with methylation while the rglB⁻ derivative of C600 was partially tolerant to methylation. Further, we show that the 1.8 kilobase (kb) and 1.2 kb KpnI fragments derived from the human L1 repeat have respectively 18.3% and 2.3% mC in vivo. Using these hyper- and hypo-methylated genomic segments ligated into the pBS plasmid, transformants with the highly methylated 1.8 kb L1 insert were recovered at 17 to 40 fold higher frequency with the rglB⁻ host strains than with the rglB⁺ hosts. In addition, recombinant phage (lambda 2001) containing inserts of plant genomic DNA with 26.7% mC (from Petunia hybrida) when plated on rglB⁻ hosts gave titres up to 222 times higher than on the rglB⁺ strains.

INTRODUCTION:

Recently Raleigh and Wilson (1) and also Noyer-Weidner et al (2) have shown that the RglA and RglB genes of E.coli function not just to restrict phage containing non-glycosylated 5-hydroxy-methylcytosine in their genomes but also recognise 5-methylcytosine (mC). Wild-type E.coli are rglA⁺ rglB⁺ while most strains used as hosts in molecular cloning are either rglA⁺ rglB⁻ or rglA⁻ rglB⁺. The presence of either of these activities reduced transformation and plating efficiencies of plasmid and phage with appropriately methylated DNA by one to two orders of magnitude (1,2). The specificities of the presumptive nucleases encoded by these genes have not been fully elucidated. Transformation with cloned bacterial methylase genes has shown the RglB function to be of very broad specificity, recognising a sequence including GmC followed by any of the four bases (1). The RglA function is as yet only defined by the HpaII methylase (CmCGG) (1).

Using plasmid transformation, we have investigated the effect of the RglB nuclease on the transformation efficiency of plasmid DNA methylated in vitro with the human DNA methyltransferase. Secondly, we have determined the transformation frequencies in rglB⁺ and rglB⁻ strains of unmethylated plasmids containing inserts of human genomic segments derived from the L1 repeat (3-12), one of which fragments we show to have a very high methylation levels in vivo compared to the genomic average and another of which we show to be normally hypomethylated. These data are relevant not just to recovery of recombinant clones from more highly methylated regions of mammalian genomes but also for the biology of L1 repeats.

As well as investigating the effect of highly methylated inserts on plasmid transformation, we have examined the effect of RglB on cloning in lambda using genomic DNA from the plant Petunia hybrida which we show has 26.7% mC. (In general, plant genomes are methylated to much higher levels than mammalian genomes.) Assaying the titres of primary packagings of plant genomic DNA in a lambda vector in rglB⁺ and rglB⁻ hosts demonstrates the relevance of host RglB function to the recovery of phage recombinants with highly methylated genomic inserts.

METHODS:

Bacterial Strains

Strains C600, TC600, DS410 and TC410 were obtained from M. Noyer Weidner. A further isolate of C600 was obtained from J. Pittard for comparison purposes. The genotypes and derivations of these strains are detailed in ref. 2. C600 and DS410 are rglA⁻ rglB⁺ while TC600 and TC410 are their respective rglB⁻ derivatives (2). DS410 and TC410 are also minA minB minicell producing strains.

In Vitro Methylation and Bacterial Transformation

The DNA (cytosine-5) methyltransferase, purified from human placenta as previously described (13), was used to methylate Bluescribe (pBS M13+) DNA (Stratagene) in vitro. Each reaction included 87 units of enzyme, 0.5 µg plasmid DNA, 0.237 µg S-adenosylmethionine (SAM) (Sigma) in a 100 µl reaction volume. Plasmid DNA was incubated for 30 minutes to 16 hours in the reaction mixture, after which time 6.02% of deoxycytidines were methylated (25.03% C in pBS DNA). Unmethylated control plasmid was incubated either for 16 hours in the reaction mix but without enzyme or incubated for 60 minutes with enzyme in a reaction mixture without SAM. Reactions were terminated by the addition of 100 µl of 0.1M Tris-acetate buffer pH 8, extraction with phenol/chloroform/isoamyl-alcohol (50/24/1) and the removal of the top 140 µl of the aqueous phase. The extent of substitution at each time point was estimated in parallel reactions in which 10% of the unlabelled SAM was replaced with [³H]SAM (Amersham), followed by precipitation with TCA.

The unlabelled plasmid DNAs were further purified with RNase A and T₁ (Boehringer), digested with Proteinase K (Boeh-

ringer) and extracted three times with phenol/chloroform/octanol (50/24/1), twice with chloroform/octanol (24/1), and ethanol precipitated. 10% of each plasmid DNA sample was digested with HindIII under conditions recommended by the supplier (Boehringer) and electrophoresed on a 1% agarose gel. The methylated plasmid DNA samples showed no sign of degradation (not illustrated). Also the recovery of plasmid DNA was uniform in all samples.

Bacterial transformation was by the high efficiency method of Hanahan (14) with < 1 ng plasmid DNA used in each transformation.

Methylation Levels in the Human L1 Repeat

Two different types of primary cultures of untransformed normal human cells were employed to produce independent estimations of relative methylation levels in different regions of human L1 repetitive elements. Normal human peripheral lymphocytes (HPL cells) were initially stimulated with phytohaemagglutinin and maintained in rapid growth with a complex culture medium containing interleukin-2 (15). HPL cells were in culture for no more than 10 days before labelling. HEF cells were a culture of human embryonic fibroblasts which exhibited the Hayflick-type response of limited lifetime in culture. HEF cells were grown in alpha-MEM with 10% foetal calf serum (Flow).

Cells were labelled for 6 hours with [^3H]thymidine (dThd) to label total DNA or with [^3H]methionine (Amersham). In the latter instance, the labelling medium was MEM minus methionine (Flow) supplemented with 20mM sodium formate (to suppress labelling of the C-1 pool) as described in detail previously (16). Under these latter conditions, ^3H incorporation into DNA was specific for the methyl groups incorporated into the 5-position of cytosine in DNA as determined by high performance liquid chromatography (HPLC) of the DNA nucleosides (16) (not illustrated). DNA was purified using Proteinase K digestion, phenol/chloroform/octanol extractions, RNase A and T₁ digestions, followed by a further Proteinase K digestion and extractions. The purified DNAs were digested with KpnI (5 units of enzyme per μg DNA) for 16 hours at 37°C under the recommended conditions (Pharmacia). Fragments were separated on a 1% agarose gel which was subsequently soaked in 2 changes of 2 litres of ethanol, then in 3% w/v 2,5-diphenyloxazole (PPO) in ethanol. The gel was then soaked in distilled H₂O to precipitate the PPO and dried under vacuum. The dried gels were fluorographed at -70°C with Kodak XAR-5 film which had been preflashed to bring the film into the range for a linear response (17). Exposure was for 8 days with the [^3H]dThd labelled DNA and 13 weeks for the [^3H]methionine labelling.

For direct HPLC determinations of deoxycytidine methylation levels in human KpnI L1 fragments, DNA was isolated from a sample of human spleen. We had previously estimated that this DNA had $4.08 \pm 0.16\%$ of deoxycytidines methylated (16). The prominent 1.2 kb and 1.5 kb bands produced by KpnI digestion of human DNA were isolated from preparative agarose gels using NA45 membranes (Schleicher and Schuell). This method allowed isolation of very narrow molecular weight distributions around the prominent L1-derived KpnI peaks. These 1.2kb and 1.5kb KpnI size fractions were enzymatically hydrolysed to deoxynucleosides and the level of deoxycytidine methylation estimated using HPLC separation on an Alltech C18 column as described (16). HPLC estimations of

methylation in the 1.8kb KpnI fraction were not performed owing to technical difficulties in obtaining sufficient material and, more importantly, the high level of background fragments in this region (Figure 2c).

Transformation with Recombinant Plasmids with Human L1 Inserts

Bluescribe (pBS M13+) plasmid DNA was cut with KpnI and treated with alkaline phosphatase (18). 20 ng vector DNA was ligated with a three fold molar excess of either the 1.2kb or 1.8kb KpnI size fraction for 3 to 4 hours at 16°C. The bacterial strains were transformed with less than 10ng plasmid DNA using the Hanahan method (14) and selected on ampicillin plates. No blue/white selection was employed. Background levels of non-recombinant ampicillin-resistant colonies were 0 to 15% of total colonies as estimated by the number of colonies obtained by transformations with ligations of the cut and phosphatased vector alone.

A minimum of 100 ampicillin-resistant colonies obtained with each ligation were transferred individually as a short streak on fresh ampicillin/LB agar plates, incubated overnight, and subsequently transferred to nitrocellulose (19) and screened for human L1 inserts using ³²P-probes labelled by random priming. This procedure allowed unambiguous identification of L1-positive colonies. The probes were the purified inserts from pBK(1.2)11 and pBK(1.8)26 (20) which contain "KpnI" inserts of the 1.2kb and 1.8kb human L1 fragments (provided by Dr J.J. Maio).

A random selection of colonies which appeared positive for the 1.8kb and 1.2kb L1 inserts were grown overnight in LB with ampicillin and "minipreps" of plasmid DNA (18) were cut with Asp718 (Boehringer) (a more reliable isoschizomer of KpnI) and after agarose gel electrophoresis transferred to nitrocellulose filters by the method described by Davies *et al* (21). Filters were hybridized with ³²P-probes for the 1.8kb or 1.2kb L1 fragments and autoradiographed with XAR-5 film.

Plant Genomic Library Construction and Plating

Petunia hybrida genomic DNA was prepared by a method based on that of Goldberg *et al* (22). This involved DNA extraction from isolated nuclei and banding in CsCl gradients, resulting in highly purified DNA with minimal contamination with chloroplast DNA. A partial Sau3A digest of this DNA was fractionated by centrifugation on a 10 to 30% glycerol gradient at 30,000 rpm at 4°C for 16 hours. The 16 to 20kb fraction was ligated at a 1/1 molar ratio with BamHI cut, phosphatase-treated lambda 2001 arms (Stratagene) and packaged using the Promega packaging extract. The lambda 2001 arms were also ligated with a control, recombinant-derived insert (pK BamHI; Stratagene) and similarly packaged. The phage were plated on the four bacterial strains by the method described in ref.(18). The titres with the different hosts were also determined with the same *Petunia* genomic library in lambda 2001 which had previously been amplified in C600Hf1 (Promega). This library had a background of < 1% estimated from an insert minus ligation. In order to analyse the library, DNA was prepared from 12 individual plaques and examined by restriction analysis: all phage showed different restriction patterns. The amplified library was prepared from 3 x 10⁵ plaques from a plating on the Promega C600Hf1 which had been found empirically to give a high plating efficiency with the primary packaging.

RESULTS:

Plasmid pBS DNA was methylated in vitro with DNA (cytosine-5) methyltransferase from human placenta (13) resulting in up to 6.02% of cytosines being methylated. Studies into the action of the human DNA methyltransferase in vitro are consistent with this enzyme exhibiting similar specificities during de novo methylation on an unmethylated substrate and maintenance methylation at hemimethylated sites (23). In vitro, the pattern of methylation applied to DNA by this enzyme is strongly influenced by primary DNA sequence (23) and perturbations in DNA structure (24). In the absence of a strong structural cue (e.g. an mC residue or a mispaired base), the enzyme generally methylates at cytosines in CG dinucleotides at a moderate rate, and at cytosines in other sequence contexts at slower but clearly measurable rates (24). Recoveries of plasmid after the in vitro methylation reaction were very uniform with no sign of degradation of the plasmid DNA (not illustrated). The rglB⁺ hosts C600 and DS410 together with their rglB⁻ derivatives were transformed by the pBS DNAs in which from 0% to 6.02% of deoxycytidines were methylated (Figure 1). Transformation was by high efficiency Hanahan procedure (14). Both the rglB⁺ strains transformed with the most highly methylated pBS DNA at only a few per cent of the efficiency of the unmethylated pBS. This is unlikely to be due to any cryptic damage to the plasmid by the methylase enzyme or any contaminating protein in that the pBS DNA incubated for 60 minutes with the full reaction mixture but without S-adenosylmethionine transformed each of the strains with the same efficiency as the control pBS which had been incubated without enzyme. Transformation efficiencies were characteristic of the host strain. However the RglB status had no appreciable or consistent effect on transformation efficiencies with unmethylated plasmid in that, in the experiment illustrated in Figure 1, C600 and TC600 gave 2.8 and 2.7 X 10⁷ transformants/μg pBS DNA. With DS410 and TC410, results were consistently 4 to 5 times lower with, for example, 0.66 and 0.40 x 10⁷ transformants/μg respectively in this experiment. These represent conservative estimations of transformation efficiencies in that these are calculated on the assumption that there was a 100% yield of plasmid after the methylation reaction and subsequent repurification/precipitation steps. However, it should be noted that the yield of plasmid was very similar for all samples (not illustrated). The lower transformation efficiencies of the DS410/TC410 strains is probably due to these cells carrying the minA minB mutations (25) which result in these cells continually producing non-viable minicells.

With even less than 1% of cytosines methylated, transform-

Figure 1

Effect of DNA methylation on transformation efficiency with rglB⁺ and rglB⁻ strains. pBS DNA methylated *in vitro* with human DNA methyltransferase was transformed into (A) DS410 and TC410 and (B) C600 and TC600. In each case, the squares with the solid line represent the response of the rglB⁺ parental strain and the crosses with the broken line represent its rglB⁻ derivative.

ation of the rglB⁺ strains by the methylated pBS was found to be significantly reduced (Figure 1). Both rglB⁺ hosts showed only 2 to 5% of the control transformation efficiencies with the pBS methylated to 6.02% mC. Transformation efficiencies for both C600 and DS410 remained reasonably constant from 3.5% to 6% mC, suggesting that methylation, while drastically reducing transformation efficiency, was not an absolute bar to the survival and replication of the plasmids.

While both rglB⁺ strains were similarly effected by methyl-

ation, TC410 (Figure 1A) was unaffected by methylation of the plasmid DNA and indeed seems to transform with higher efficiency with more highly methylated plasmid (observed in 5 of 5 experiments). However transformation efficiencies with TC600 fell with increased methylation but were higher (by 2 to 8 times) than the rglB⁺ parental strain (Figure 1B). To examine the reason for this, all 4 strains were transformed with unmethylated pBS (as a control for relative transformation efficiencies) and (separately) with the plasmid pBN16 which carries the methylase gene from the ϕ 3T temperate phage of Bacillus subtilis (26). This plasmid will only transform rglB⁻ strains. (The specificity of this methylase is GGmCC and GmCNGC (2).) While both TC600 and TC410 could be transformed with good efficiency with pBN16, the efficiency with TC410, when corrected for overall transformation efficiencies, was 19 times higher than that observed with TC600. Transformation efficiencies with rglB⁺ C600 and DS410 with pBN16 were a minimum of 2 to 3 order of magnitude lower in that no transformants were recovered with these strains in this experiment. Hence, while TC600 is rglB⁻ by this criterion, it appears to retain partial RglB function when compared to TC410. However it is not possible to rule out absolutely that some other known or unknown mutation present in the DS410 background unrelated to RglB is responsible for this difference in tolerance to methylation.

The above data relates to methylation introduced within the 3.2 kb length of the plasmid DNA, most of which is required for plasmid survival within the host. We have further investigated the effect of introducing an insert of a genomic DNA segment which in vivo contains a relatively high or relatively low level of methylation into the unmethylated plasmid. The genomic fragments employed were the prominent 1.2 kb and 1.8 kb fragments produced by KpnI digestion of human DNA which are derived from the L1 repeat. We show below that there is an overall eight fold difference in the proportion of cytosines methylated in these two genomic fragments.

The full length human L1 element is thought to be over 6 kb in length and to contain a series of conserved KpnI sites as illustrated in Figure 2a (11,12). The A-rich tract at the 3' end is some 1.4 kb 3' to the last conserved KpnI site. Hence cleavage of human DNA with KpnI produces prominent KpnI fragments of 1.8kb, 1.5kb, and 1.2kb (Figure 2). The 1.2 kb fragment from towards the 3' end of the L1 repeat is present at a 4 to 5 fold higher molar ratio than the 1.8 kb KpnI region from the 5' end (11,12). The methylation levels in the different regions of the human L1 element were estimated by combining direct HPLC

Figure 2

(a): Current consensus structure of a full length L1 repeat from the human genome (11,12). Conserved KpnI sites are indicated (K) together with the prominent 1.8 kb, 1.5 kb, and 1.2 kb L1-derived restriction fragments observed in KpnI digests of human DNA. The 3' poly-A region is indicated by the wavy line. There are no conserved KpnI sites at the 5' or 3' ends.

(b): Distribution of (A) total DNA and (B) methylated sites in the KpnI fragments of DNA of HEF cells after agarose gel electrophoresis (autoradiograms of the parallel tracks of (A) total and (B) mC-labelled DNA). The numbers to the left of the autoradiograms indicate the positions of the size markers (in kilobase pairs). On the right, the positions of the prominent 1.2, 1.5, and 1.8 kb KpnI L1 fragments are shown. Also indicated is another relatively heavily methylated band at 5 kb. This latter band in the mC labelled track is of unknown origin and significance. It may or may not be an L1-related sequence.

(c): Densitometric scans of the lower molecular weight regions of the [^3H]dThd-labelled track of this autoradiogram. The positions of the 1.2, 1.5, and 1.8 kb KpnI bands are indicated. Tight molecular weight size fractions isolated from each of the peaks would have included some fragments from the background smear of mainly unrelated fragments. This background is indicated in the figure by differential cross-hatching and was estimated to comprise 10% of the total fragments in the 1.2 kb fraction but to be 70% of the 1.8 kb fraction. This was due both to the 4 to 5 fold lower molar representation of the 5' end of the L1 elements (12) and to the rising background at the higher molecular weight.

estimates of mC levels in isolated genomic size fractions with autoradiographic data from in vivo labelling of human DNA. Non-transformed human embryonic fibroblasts (HEF cells) and stimulated normal human peripheral lymphocytes (HPL cells) were labelled in vivo with either [^3H]thymidine (dThd) (for total DNA) or with [me ^3H]methionine, the latter under conditions in which labelling in the DNA was specific for the methyl group of mC (16). These DNAs were purified, digested with KpnI, the fragments separated on an agarose gel which, after impregnation with scintillant, was fluorographed with preflashed film (17) (Figure 2b). Densitometric scanning of the autoradiograms showed that the 1.8kb peak had, on average for the 2 cell types, some 8 fold higher level of methylation than the 1.2 kb peak (Table I). Note that further cutting with a series of restriction enzymes of the isolated 1.2 kb and 1.8 kb size fractions produced a series of discrete fragments, consistent with each of these peaks being derived from a single major repetitive element (not illustrated).

The level of mC in the 1.2 kb and 1.5 kb KpnI L1 fragments was determined directly using HPLC analysis (16). These fragments were isolated from preparative agarose gels using NA45 paper which allowed a tight size cut. HPLC analysis of DNA nucleosides (Table I) showed that the 1.2 kb fraction had some 2.27% mC and was thus hypomethylated compared to the genomic average of 4.08%. As can be seen from the scans of the [^3H]dThd-labelled track (Figure 2c), the size fraction taken for the isolation of each of the prominent L1 fragments will contain some unrelated DNA fragments from the background "smear" from the

Table I
In Vivo Methylation in Human L1 Regions

	KpnI L1 Fractions		
	1.2 kb	1.5 kb	1.8 kb
<u>HEF Cells</u>			
Molar Ratios	1	0.61	0.19
(relative to 1.2 kb)			
Methylation	1	1.62	9.67
(relative to 1.2 kb)			
<u>HPL Cells</u>			
Molar Ratios	1	0.53	0.24
(relative to 1.2 kb)			
Methylation	1	1.70	6.44
(relative to 1.2 kb)			
<u>Human Spleen DNA</u>			
%C Methylation	2.27± 0.32%	3.4± 0.14%	18.3%
(HPLC)	(n=4)	(n=4)	(calculated)
Methylation (by HPLC)	1	1.53	
(relative to 1.2 kb)			

Relative methylation levels and molar ratios of the 1.2, 1.5, and 1.8 kb peaks in the distributions in the autoradiograms of in vivo labelled DNA from HEF and HPL cells and the fraction of deoxycytidines methylated in the 1.2 and 1.5 kb KpnI fractions isolated from human spleen DNA as quantitated by HPLC of nucleosides. The areas under the peaks on the scans of the autoradiograms were calculated by (height) x (width at half height). The molar ratios of the peaks were calculated by dividing the area by the fragment size, and are expressed relative to the value calculated for the 1.2 kb peak. The relative methylation levels are the areas of the peaks in the mC track divided by areas of the peaks in the total DNA label track, and, for simplicity, these values are again expressed relative to the value calculated for the 1.2 kb peak. The fractions of deoxycytidines methylated in the 1.2 kb and 1.5 kb size fractions from KpnI cut human spleen DNA were estimated after enzymatic hydrolysis to nucleosides and HPLC on a C-18 column. Results are expressed ± the standard error of the mean and "n" denotes the number of independent estimations. The level of methylation could not be estimated by this method for the 1.8 kb L1 fragment due to 70% of this size fraction being composed of unrelated "background" fragments (Figure 2c). The level of methylation in the 1.8 kb L1 region was calculated from the HPLC-determined methylation level in the 1.2 kb fraction (which had the minimum of background) (Figure 2c) and the average of the excess of methylation in the 1.8 kb peak compared to the 1.2 kb L1 peak as determined by autoradiography with HEF and HPL cell DNAs.

random positioning of KpnI sites in other regions of the genome. The 1.2 kb fraction was estimated from the [³H]dThd scan to contain 90% of DNA from the repetitive element-derived peak and

Table II
Percentages of Transformants Positive for L1 Inserts

	1.2kb (2.3% <u>mC</u>)		1.8kb (18.3% <u>mC</u>)	
	<u>Exp 1</u>	<u>Exp 2</u>	<u>Exp 1</u>	<u>Exp 2</u>
C600 (<u>rglB</u> ⁺)	37%	44%	2%	0.5%
TC600 (<u>rglB</u> ⁻)	86%	68%	20%	23%
DS410 (<u>rglB</u> ⁺)	53%	24%	0%	0.9%
TC410 (<u>rglB</u> ⁻)	93%	66%	21%	15%

The 1.2 kb and 1.8 kb KpnI size fractions from human spleen DNA were ligated to KpnI cut, alkaline phosphatase treated pBS plasmid and used to transform the 4 different strains by the Hanahan procedure (14). A minimum of 100 ampicillin resistant colonies from each of the transformations were transferred by sterile toothpick as a short streak on a large LB plate. After overnight growth, the colonies were screened using the Maio probes (20) for the 1.8 kb and 1.2 kb KpnI fragments from the human L1 repeat. 94% to 100% of the transferred colonies grew in each case. In Experiment 2, 100 colonies were scored from transformations with each of two independently maintained isolates of the C600 strain (from M. Noyer-Weidner, Berlin, and J. Pittard, Melbourne). Since the results with the two different C600 isolates were virtually identical, their data have been combined for this Table. Results are expressed as the percentage of the colonies which grew which were positive for the appropriate human L1 fragment.

only 10% from the background (Figure 2c). By contrast, in the 1.8 kb size fraction, some 70% of DNA fragments were from the background. Combining the HPLC estimate of methylation level in the 1.2 kb fraction (which contains the minimum amount of background fragments) with the average of the increments in methylation in the 1.8 kb KpnI peak compared to the 1.2 kb peak found with the HEF and HPL cell DNAs, we estimate that some 18.3% of cytosines are methylated in the 1.8 kb L1 segment. This is more than 4 times the genomic average of 4.08% mC. Note that the maximum attainable level of methylation in any given segment of the human genome is not limited by the number of 5'-CG-3' dinucleotides present in that we have shown that, even in total human spleen DNA, >50% of total methylation is present in dinucleotides other than CG (27).

The 1.2 kb and 1.8 kb KpnI L1 segments were used as a model system to examine the effect of differential methylation levels in inserts ligated into an unmethylated plasmid on the efficiency of recovery of recombinants. Because the KpnI fragments from the genomic background present in each of the 1.2 kb and 1.8 kb size

fractions (Figure 2c), completely random cloning of genomic fragments should result in 90% of transformants from the 1.2 kb ligation and 30% from the 1.8 kb ligation being positive for L1 inserts. Strains C600, TC600, DS410 and TC410 were each transformed by the Hanahan procedure (14) with ligations of the 1.2 kb and 1.8 kb KpnI size fractions into pBS plasmid which had been KpnI cut and treated with alkaline phosphatase (18). A minimum of 100 ampicillin-resistant colonies of each strain from each of the 2 ligations were transferred individually to fresh selective plates and, after overnight growth, transferred to nitrocellulose (19). The filters were probed with 1.2 kb and 1.8 kb "KpnI repeats" cloned from human L1 elements by Shafit-Zagardo *et al* (20). The results of two independent experiments in which a total of >1700 individual transformants were screened for either the 1.2 kb or 1.8 kb L1 insert are presented in Table II.

The rglB⁻ strains gave, on average, close to the expected frequency of transformants with the 1.2 kb L1 insert (a mean of 77% for TC600 and 80% for TC410 for the two experiments against the expected 90%) but only about 2/3 of the expected 30% for the 1.8 kb L1 insert (22% for TC600 and 18% for TC410). However, for the rglB⁺ strains, the recovery of recombinants with the highly methylated 1.8 kb L1 insert were very much lower than the expected 30% with a mean of 1.25% positive colonies with C600 and 0.45% with DS410. Furthermore, the recovery of recombinants for the rglB⁺ hosts for the 1.2 kb L1 segment were only half that observed with the rglB⁻ strains with an average over 2 experiments of 41% positives with C600 and 39% with DS410. Thus, the recovery of transformants positive for highly methylated inserts is severely depressed in rglB⁺ hosts and the recovery of transformants with even relatively hypomethylated genomic segments is significantly below what would be expected for random cloning. In contrast, rglB⁻ strains allow apparently much more random recovery of recombinants from the different regions of human L1 elements.

The identity of a random sample of these transformants was checked using Southern analysis of "minipreps". Ten out of 11 colonies which were scored as positive for the 1.8 kb L1 insert contained plasmids with a 1.8 kb insert which hybridized with the 1.8 kb L1 probe. For the 1.2 kb positive recombinant colonies, 5 of 5 from which plasmid DNA was recovered from the "minipreps" contained a 1.2 kb insert which hybridized to the 1.2 kb L1 probe. With the TC410 recombinants, we found that the plasmid was readily lost from an overnight liquid culture in spite of the use of a high concentration of ampicillin (100 µg/ml). This is most probably a result of the minA minB genotype of this strain.

Plasmids in TC600 were found to be stably maintained. Hence the frequency of recovery of positive transformants with the TC410 host must be considered a minimum estimate due to the propensity of this strain to being overgrown by cells which have lost the plasmid. Consistent with this, DS410 and TC410 transformants gave noticeably weaker signals in the colony screening than the C600 and TC600 transformants (not illustrated).

The similar recoveries of 1.8 kb L1 recombinants with the TC600 and the TC410 strains (Table II) appears inconsistent with the only partial tolerance to methylation observed with the pBS methylated in vitro (Figure 1). This could be due to two possible mechanisms. Firstly, apparently "leaky" rglB⁻ strains such as TC600 may be more tolerant to methylation in an insert than to methylation throughout the plasmid genome in regions essential for plasmid survival and replication. Secondly, we may be significantly underscoring the number of primary transformants in TC410 positive for the L1 insert due to the minA minB genotype of these cells. As with our difficulties in recovering plasmid in "minipreps" from TC410 transformants, it appears that the plasmid containing the ampicillin resistance gene is frequently segregated into minicells. These minicells continue to produce beta-lactamase, allowing the growth of, and subsequent overgrowth

Table III
Plating Efficiencies (pfu/ml) of Lambda Recombinants
on Different Host Strains

Host Strain	Packaged Petunia Library	Packaged Control Insert	Amplified Petunia Library
C600	5.6 x 10 ²	2.2 x 10 ⁵	2.9 x 10 ⁹
TC600	46 x 10 ²	2.4 x 10 ⁵	2.1 x 10 ⁹
DS410	9 x 10 ²	1.7 x 10 ⁵	3.1 x 10 ⁹
TC410	2000 x 10 ²	2.0 x 10 ⁵	3.0 x 10 ⁹

Titre of phage in pfu/ml on the different host strains. The phage samples tested were freshly packaged ligations of lambda 2001 arms with Petunia genomic DNA and also freshly packaged ligations of lambda 2001 arms with an unmethylated control insert, pK BamHI. A previously amplified genomic library in lambda 2001 with Petunia DNA inserts was also plated on all 4 hosts. The rglB⁺ strains and their rglB⁻ derivatives gave virtually identical titres with both the controls. However, both rglB⁻ strains gave higher titres (8 times in one instance, 222 times in the other) than the rglB⁺ parental strain with the freshly packaged lambda 2001 containing inserts of highly methylated plant genomic DNA.

by, cells which have lost the plasmid, resulting in an L1 positive colony being scored as a negative after a second round of growth. Indeed both of these mechanisms might be acting in concert to give apparently equal transformation efficiencies in the TC600 and TC410 strains with plasmids containing the highly methylated insert.

The two pairs of rglB⁺/rglB⁻ strains were also tested as to their relative efficiencies as hosts for lambda recombinants containing highly methylated inserts or with unmethylated inserts. We estimate by HPLC analysis that genomic DNA from the plant Petunia hybrida has on average 26.7% of deoxycytidines methylated, several times the 3 to 4% mC typically found in mammalian genomes. Plating efficiencies were compared on the 4 bacterial strains with (1) the primary packaging of this highly methylated plant genomic DNA in lambda 2001, (2) the primary packaging of lambda 2001 with an unmethylated (recombinant-derived) insert, and (3) an amplified genomic library containing inserts from the same plant DNA in which the original plant pattern of methylation would have been lost through growth on C600Hfl (Table III). The titre of phage was equivalent on all 4 strains with the previously amplified library and also with the packaged lambda 2001 ligated with the unmethylated control insert. The plating efficiencies of the primary plant library were similar for the two rglB⁺ hosts but were higher on both rglB⁻ strains (8.2 and 222 times higher in TC600 and TC410 respectively compared to their rglB⁺ parental strains). The higher titre with TC410 parallels the higher transformation efficiency of this strain with methylated plasmid DNA (Figure 1). The relatively small increment of titre with TC600 compared to TC410 is probably the result of the TC600 being a leaky rglB⁻ mutant.

DISCUSSION:

Several percent of mammalian genomes are refractory to molecular cloning in lambda vectors (28,29). While much of this is likely to be due to such features in the genomic DNA as palindromic sequences which have the potential of forming regions of secondary structure incompatible with phage replication (29), our data demonstrate that another impediment to the recovery of recombinants of genomic DNA is the in vivo level of methylation. Our data suggest that a high methylation level is probably not an absolute bar to the recovery of recombinants in rglB⁺ hosts (such as are many commonly used strains such as C600 and LE392 (1)), but rather acts to reduce significantly the probability of recovery of recombinant clones from more highly methylated

genomic regions. For a low copy number or unique sequence, this might in practical terms mean the difference between success or failure in the recovery of an appropriate recombinant clone. A corollary of this is that regions of the genome with intrinsically low methylation levels in vivo will be overrepresented in genomic libraries generated using many of the commonly used host strains. Further, our data shows that it is not sufficient simply to use any rglB⁻ host strain in that the two rglB⁻ strains employed in this study showed a major difference in tolerance to methylation. However the use of even a "leaky" rglB⁻ mutant such as TC600 allows a marked improvement in recovery of a highly methylated region of the human genome (the 5' end of the L1 repeat).

The L1 repeat (or LINE) is the most prominent family of long interspersed repetitive elements in the human genome (3). Evolutionarily related L1 repeats are present in all mammalian (and indeed marsupial) genomes examined to date (4). L1 elements are mobile genetic elements, at least in the evolutionary sense, in that there have been a minimum of 50,000 deletions or insertion events involving L1 elements in the 2 million years since the divergence of the human and chimpanzee lineages (5). These elements are thought to be retroposons (i.e. mobile via an RNA intermediate). The 3' end frequently contains a poly-A tract and the other (5') end is absent from the majority of the members of the family with truncation occurring at variable distances from the A-rich end (6,7). This overrepresentation of 3' end fragments means that at least 75% to 80% of the 1.2 kb KpnI fragments must be derived from truncated, pseudogene-like members of the L1 family. This truncation has been suggested to be due to variable extents of reverse transcription from full length L1 transcripts. Strong support for this model comes from the observation that L1 elements from several mammalian species contain a conserved open reading frame which potentially encodes a protein with significant homology with known reverse transcriptases (8,9,10).

As to the significance of the in vivo methylation levels of the different regions of the human L1 retroposons, there are two different interpretations of our data. The high level of methylation of 1.8 kb KpnI fragments from the 5' end of full length (and hence potentially functional) L1 elements might have the role in vivo in maintaining these elements in a quiescent state, thus preventing any destabilization of the genome resulting from bursts of transposition events. Alternatively, low methylation levels might be characteristic of the non-functional, pseudogene-like members of the L1 family (2.27% mC in the 1.2 kb KpnI region compared to the genomic average of 4.08% mC). Mobilization of

genetic elements via an RNA intermediate includes two steps (transcription and reverse transcription) in which any original pattern of methylation would be lost. Because of the 4 to 5 fold higher molar representation of the 1.2 kb L1 fragments compared to the 5' region (12), at least 75% to 80% of the 1.2 kb fragments must be derived from truncated members of the L1 family. It is possible that full length L1 elements which are potentially functional may have a high level of methylation throughout their entire sequence. If this latter possibility were correct, molecular cloning of human L1 elements using many *E.coli* hosts will result in the preferential cloning of the members of the L1 family with low methylation levels in vivo, i.e. the non-functional pseudogene members of the family. Hence those members of this repeat family which have accumulated the most mutations (base substitutions, deletions and insertions) may be inadvertently being preferentially cloned in many instances because of the choice of host strain. There is a high level of base sequence alterations amongst recombinant clones from homologous regions of L1 repeats from a given mammalian genome (3,8,9). However limited genomic sequencing of a region of the mouse L1 repeat indicated that there was less variation in this L1 segment in the mouse genome than would be indicated from the sequence variation between the recombinant inserts recovered from this region (30). This would be consistent with the second of the two possibilities considered above. This hypothesis is currently being tested directly with L1 clones derived using *rglB*⁻ and *rglB*⁺ hosts.

In summary, from a practical point of view, neither of the *rglB*⁻ strains used in this study represent ideal hosts for recombinant cloning. The TC600 permits ready recovery of recombinant plasmids but is only partially tolerant to cytosine methylation. However this strain may be more tolerant to methylation of an insert than to methylation of the whole plasmid. The second strain, TC410, appears to be totally tolerant to methylation but, probably because of its being also *minA minB* (25), it is difficult to recover plasmids from this host. However, when cloning highly methylated genomic inserts in lambda vectors and during construction of eukaryotic genomic libraries, this strain may allow the generation of a higher titre of more representative phage clones and may be worthy of immediate practical application. These results are relevant to the construction of all eukaryotic genomic libraries in that host *RglB* (and probably *RglA*) function have a significant effect on both the efficiency of recovery of recombinants and, possibly even more importantly, the use of a *rgl*⁺ host is likely to lead to bias in the representation of sequences. This applies to mammalian libraries and

especially to plant libraries. In the latter instance, because chloroplast DNA is unmethylated (31), recombinants containing chloroplast-derived inserts are likely to be grossly overrepresented in plant libraries plated on most of the commonly used E.coli host strains.

Acknowledgement:

We would like to thank Dr Mario Noyer-Weidner and Professor J. Pittard for bacterial strains used in this study, Dr Noyer-Weidner for the pBN16 plasmid, Professor Alex Morley for the stimulated human lymphocytes, Dr T.R. Bradley for the HEF cell culture, Dr J.J. Maio for the human L1 clones, and Mrs Margaret Tsakmakis for typing the manuscript. This work was supported in part by Grants to S.S.S. from the Council for Tobacco Research (1571A) and the National Institutes of Health (GM-32863).

*To whom correspondence should be addressed

References:

1. Raleigh, E.A. and Wilson, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9070-9074.
2. Noyer-Weidner, M., Diaz, R. and Reiners, L. (1986) *Mol. Gen. Genet.* 205, 469-475.
3. Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) *Ann. Rev. Biochem.* 55, 631-661.
4. Burton, F.H., Loeb, D.D., Voliva, C.F., Martin, S.L., Edgell, M.H. and Hutchinson, C.A. (1986) *J. Mol. Biol.* 187, 291-304.
5. Hwu, H.R., Roberts, J.W., Davidson, E.H. and Britten, R.J. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3875-3879.
6. Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchinson, C.A. and Edgell, M.H. (1983) *Nucleic Acids Res.* 11, 8847-8859.
7. Soares, M.B., Schon, E. and Efstratiadis, E. (1985) *J. Mol. Evol.* 22, 117-133.
8. Fanning, T. and Singer, M. (1987) *Nucleic Acids Res.* 15, 2251-2260.
9. Loeb, D.D., Padgett, R.W., Hardies, S.C., Shehee, W.R., Comer, M.B., Edgell, M.H. and Hutchinson, C.A. (1986) *Mol. Cell. Biol.* 6, 168-182.
10. Hattori, M., Kuhara, S., Takenada, O and Sakaki, Y. (1986) *Nature* 321, 625-628.
11. Singer, M.F. and Skowronski, J. (1985) *Trends Biochem. Sci.* 10, 119-122.
12. Grimaldi, G., Skowronski, J. and Singer, M.F. (1984) *EMBO J.* 3, 1753-1759.
13. Zucker, K.E., Riggs, A.D. and Smith, S.S. (1985) *J. Cell. Biochem.* 29, 337-349.
14. Hanahan, D. (1983) *J. Mol. Biol.* 166, 557-580.
15. Morley, A.A., Trainor, K.J., Seshadri, R. and Ryall, R.G. (1983) *Nature* 302, 155-157.
16. Woodcock, D.M., Crowther, P.J., Simmons, D.L. and Cooper,

- I.A. (1984) *Biochim. Biophys. Acta* 783,, 227-233.
17. Laskey, R.A. and Mills, A.D. (1975) *Eur. J. Biochem.* 56, 335-341.
 18. Maniatis, T., Fritsch, E. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratory, Cold Spring Harbour, New York).
 19. Nicholls, R.D., Hill, A.V.S., Clegg, J.B. and Higgs, D.R. (1986) *Nucleic Acids Res.* 13, 7572-7578.
 20. Shafit-Zagardo, B., Brown, F.L., Maio, J.J., and Adams, J.W. (1982) *Gene* 20, 397-407.
 21. Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) *Basic Methods in Molecular Biology* (Elsevier, New York).
 22. Goldberg, R.B., Hoschek, G. and Vodkin, L.D. (1983) *Cell* 33, 465-475.
 23. Ward, C., Bolden, A., Nalin, C.M. and Weissbach, A. (1987) *J. Biol. Chem.* 262, 11507-11063.
 24. Smith, S.S., Hardy, T.A. and Baker, D.J. (1987) *Nucleic Acids Res.* 15, 6899-6919.
 25. Dougan, G. and Sherratt, D. (1977) *Mol. Gen. Genet.* 151, 151-160.
 26. Noyer-Weidner, M., Jentsch, S., Kupsch, J., Bergbauer, M. and Trautner, T.A. (1985) *Gene* 35, 143-150.
 27. Woodcock, D.M., Crowther, P.J. and Diver, W.P. (1987) *Biochem. Biophys. Res. Commun.* 145, 888-894.
 28. Wyman, A.R., Wolfe, L.B. and Botstein, D. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2880-2884.
 29. Shurvington, C.E., Stahl, M.M. and Stahl, F.W. (1987) *Proc. Natl. Acad. Sci USA* 84, 1624-1628.
 30. Vizard, D.L. and Yarsa, J. (1985) *Nucleic Acids Res.* 13, 473-484.
 31. Tewari, K.K. and Wildman, S.G. (1966) *Science* 153, 1269-1271.