Volume 11 Number 11 1983 Nucleic Acids Research

Effect of CpG methylation on *Msp* I

Elana Keshet and Howard Cedar

Department of Molecular Biology, The Hebrew University-Hadassah Medical School, P.O.B. 1172, Jerusalem, Israel

Received 28 March 1983; Accepted 11 May 1983

ABSTRACT

The restriction enzyme Msp I is inhibited by the presence of a methyl moiety at the external cytosine of the sequence CCGG, but is generally unaffected by methylation at the internal cytosine. At specific subsets of this sequence such as the hexanucleotide CCGGCC, however, methylation of the internal cytosine strongly inhibits Msp I digestion, leading to artifacts in the interpretation of DNA methylation analyses. Our results show, for instance, that the CCGG site at the 5' end of the human γ globin gene, which was thought to be methylated at both the internal and external cytosines, is actually methylated only at the internal CpG residue.

INTRODUCTION

Although almost all methylated cytosine moieties in animal cells are found in the dinucleotide sequence CpG (1), several studies indicate that there may be some rare methylation of other nucleotide sequences (2-5). One of the accepted methods of assaying methylation at CpC residues is by the use of the restriction enzyme Msp I which recognizes the sequence CCGG and will not cleave if the external cytosine is methylated (6,7). Using this technique, it was shown that a unique Msp I site in the γ^A and γ^G human globin genes as well as similar sites in other genes are methylated at CpC moieties (7-10). While studying the mechanism of inheritance of these methyl groups it was found that at certain specific CCGG sequences cleavage by Msp I is strongly inhibited by methylation at the internal cytosine alone. The fact that Msp I can be affected by methylation at CpG residues casts doubt on experiments which rely on the specificity of this enzyme to demonstrate CpC methylation. Using this information it is possible to show conclusively that in human tissues the unique CCGG site at the 5' end of the γ globin gene is indeed methylated at the internal cytosine residue but unmethylated at the external cytosine.

© IRL Press Limited, Oxford, England.

MATERIALS AND METHODS

Human leukocytes were isolated from blood samples (11) and DNA isolated by treatment with proteinase K, subsequent phenol and chloroform extractions and ethanol precipitation (12). Plasmid pHγ1 containing the Hind III fragment of the γ^A globin gene inserted into pBR322 was obtained from R. Flavell. Plasmid DNAs and their derivatives were propagated in Escherichia coli K-12 and purified by the method of Clewell (13). Bacteriophage M13 replicative form (RF) DNAs were prepared as described (14).

Hpa II DNA methylase was obtained from Biolabs (15). Plasmid DNA was incubated at 50μg/ml in 50mM Tris pH 7.5, 5μM S-adenosylmethionine, 5mM dithiothreitol and 10mM EDTA at 37°C overnight with a saturating amount of enzyme (as determined experimentally). The reaction was extracted with phenol and chloroform and concentrated by ethanol precipitation (16). DNA methylated *in vitro* was tested for methylation at the Hpa II sites by restriction digestion analysis. In every case methylation was over 98% complete and localized to the internal cytosine of the CCGG sites (16). Although Msp I cuts this DNA at most methylated CCGG sites, specific sites remain relatively resistant to Msp I digestion. In order to determine whether these specific sites are uniquely modified at the external cytosine residue, fragments resulting from cleavage at one of these sites in methylated pHγ1 DNA were extracted from the gel following electrophoresis. Since Msp I cuts the CCGG site between the two cytosine residues the following sticky ends were obtained $\genfrac{}{}{0pt}{}{\text{C}}{\text{GGC}}$. These ends were labelled with ^{32}PαdCTP using the Klenow fraction of *E. coli* DNA polymerase I (17) and the methylation state of the external cytosine could then be determined by nearest neighbor analysis and thin layer chromatography (18).

Restriction enzyme digestions were carried out under the conditions recommended by the suppliers (New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim). Reactions were terminated by the addition of EDTA and the products electrophoresed on agarose slab gels. For blot hybridization analysis DNA fragments were transferred to nitrocellulose sheets, hybridized and washed as described (19). The 3.3kb Hind III fragment containing the γ globin gene was nick translated with [α^{32}P]dCTP (New England Nuclear) to a level of 3-5x10^8 cpm/μg (12).

RESULTS

The plasmid pHγ1 contains the human genomic γ^A globin gene inserted into pBR322 at the Hind III site (figure 1). In order to study the effect of DNA methylation on cleavage by the restriction enzyme Msp I, this plasmid was

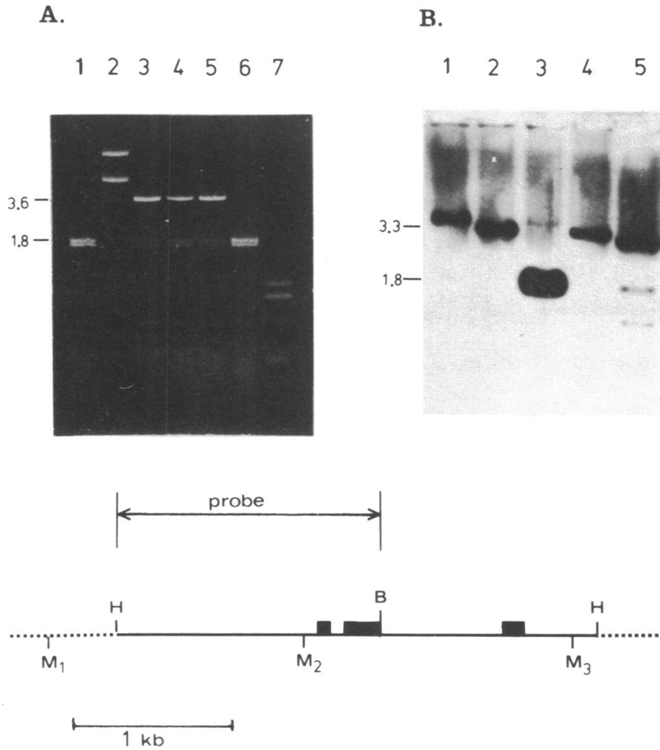

Fig. 1. Msp I digestion of methylated pHγ1. A. pHγ1 was methylated in vitro using the Hpa II methylase as described in Materials and Methods. 1.5μg of unmethylated (lane 1) or methylated pHγ1 DNA (lanes 2-6) were digested with 6 units (lanes 1 and 3), 24 units (lane 4), 48 units (lane 5) or 480 units (lane 6) of Msp I for 1 hr at 37°C and electrophoresed on a 1% agarose gel. Undigested pHγ1 DNA is shown in lane 2 and Hae III digested ØX174 DNA in lane 7. B. 20μg mouse DNA was mixed with 80pg pHγ1 DNA premethylated with Hpa II methylase in vitro. The reconstituted mixture was digested with 40 units of the indicated enzymes for 2 hr at 37°C. The samples were then electrophoretically resolved, blotted and probed with the nick translated Hind III/BamHI fragment. DNA was treated with Msp I (lane 1), Hind III (lane 2), Hind III + BamHI (lane 3), Hind III + Hpa II (lane 4) or Hind III + Msp I (lane 5). The schematic shows the pHγ1 plasmid with the restriction enzyme sites for Hind III (H), BamHI (B) and Msp I (M). The blocked areas represent the gene exons and the dotted line shows the pBR322 DNA. The two bands (in A) which migrate at about 1.8kb represent the digestion products M1/M2 and M2/M3, while the 3.6kb fragment results from a lack of digestion at M2. The pBR322 digestion products appear at the bottom of the gel, the largest fragment being 527kb. In part B the 3.3kb Hind III fragment and the 1.8kb Hind III/BamHI fragment are indicated. In the map of the γ globin gene the 5' end is on the left side.

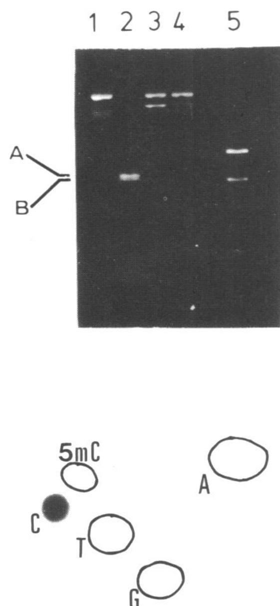

Fig. 2. Hpa II methylase methylates only the internal cytosine in pHγ1 DNA. pHγ1 DNA was methylated *in vitro* by either Hpa II methylase (lanes 1,2) or Msp I methylase (lanes 3,4). 1.5μg of each DNA was digested with 150 units of Msp I (lanes 2 and 4) for 1 hr at 37°C and electrophoresed on a 1% agarose gel. Undigested DNA's were run in parallel (lanes 1 and 3). ØX174 DNA digested with Msp I served as a size marker (lane 5). Fragments A and B were extracted from the gel and end labelled, as specified in Materials and Methods. The labelled DNA was then enzymatically digested to nucleoside-3'-monophosphates, separated by two dimensional thin-layer chromatography and autoradiographed (18). The migration of the various nucleotides are identified in the figure: (A) dAMP; (G) dGMP; (T) dTMP; (C)dCMP; (5mC) 5 methyl-dCMP.

methylated *in vitro* using Hpa II methylase, an enzyme which methylates all CCGG sites at their internal cytosine residue (16). When the methylated pHγ1 was cleaved with Msp I at an enzyme to DNA ratio of 1 unit/μg DNA, almost all of the CCGG sites within the plasmid were digested to completion (lower bands in figure 2), but the Msp I site at the 5' end of the gene insert remained uncleaved. Even at very high enzyme to DNA ratios this site remained resistant to cutting and only at a ratio greater than 150:1 was this unique site digested to completion (see figure 1).

This experiment strongly suggested that methylation of the internal cytosine of this unique CCGG site may have an inhibiting effect on Msp I cleavage. An alternate interpretation of these results is that inhibition may have been caused by the fortuitous methylation of the external cytosine of this particular site by some activity in the Hpa II methylase preparation. In order to rule out this possibility the external cytosine residue of this γ globin CCGG site in Hpa II methylated DNA was analyzed for methylation by thin layer chromatography. To this end the cleavage bands A and B (figure 2) were extracted from the gel (20) and the Msp I sticky ends were labelled with [α^{32}P]dCTP using the DNA polymerase Klenow fraction. The resulting labelled DNA was then subjected to nearest neighbor analysis in order to determine the methylation state of the external cytosine adjacent to the ^{32}P label. Analysis of the

labelled nucleotides by two dimensional thin layer chromatography revealed that this cytosine was completely unmodified (figure 2). Since both bands A and B contained cytosine and not 5-methylcytosine at their 3' ends, we may conclude that both external cytosines of this CCGG site were not methylated by Hpa II methylase in vitro. Further evidence suggesting that the external cytosine of this site is not methylated was obtained from experiments in which the pHγ1 DNA was methylated in vitro with the Msp I methylase. This enzyme, unlike the Hpa II methylase, modifies both the internal and external cytosine residues of every CCGG site (6). When this DNA was exposed to Msp I, none of the CCGG sites were cleaved by Msp I even at enzyme to DNA ratios of 200:1.

The Msp I site located at the 5' end of both human γ globin genes is imbedded in a highly GC rich region (21). The nucleotide sequence in the immediate vicinity of this site is as follows: GGGGCCGGCGGC. As can be seen in figure 1 other CCGG containing sequences are apparently cut normally by Msp I even when methylated at the internal cytosine residue, but the resolution obtained in this experiment is not sufficient to rule out abherrant cleavage of other sites. In order to search for other CCGG sites which are relatively resistant to Msp I digestion, pBR322 and M13 RF DNA were cleaved with moderate concentrations of Msp I, and the resulting fragments were end-labelled and analyzed by gel electrophoresis. At relatively high enzyme to DNA ratios most of the CCGG sites were cleaved in both DNA preparations, but several distinct partial digestion products were also identified. In pBR322 two such bands corresponding to 282bp and 24bp were identified and in M13 DNA two bands of molecular size 974bp and 603bp were observed. In each of these cases the partial product could be traced to a lack of cleavage at a particular site within the DNA, and all of these sites contained the concensus sequence CCGGCC. It should be noted that other CCGG containing sequences are also cut poorly by Msp I when the internal cytosine is methylated. Thus at intermediate ratios of DNA the partial digests which are obtained do not show a random distribution (figure 3). The pBR322 digest, for example, shows additional preferred partials at about 90, 225 and 250bp. These bands can be traced to a lack of digestion at specific CCGG sites. Although we found no common denominator which characterizes all of these inhibited sites, they all contain clusters of 5-6 cytosine or guanosine nucleotides in the immediate vicinity of the recognition sequence. Methylation at the internal cytosine of CCGG sites obviously produces varying degrees of inhibition depending on the sequences flanking these regions, but the sequence CCGGCC shows the most striking inhibitory effect.

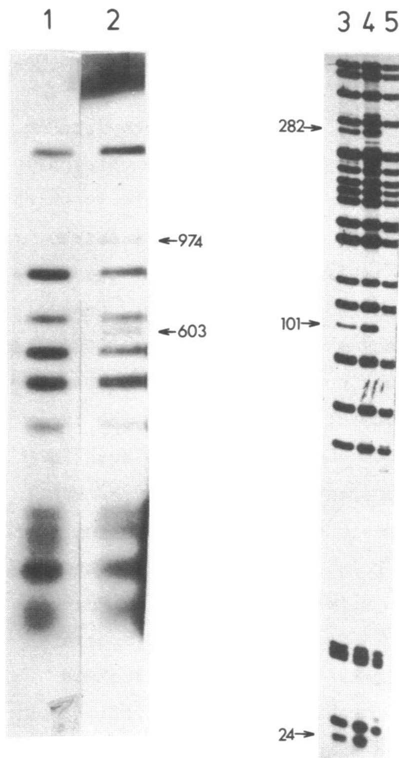

Fig. 3. Msp I sequence specificity of methylated CCGG sites. 1. 1.5μg M13-mp8 DNA and pBR322 DNA were methylated with Hpa II methylase and digested with indicated amounts of Msp I. The samples were end-labelled with [α^{32}P]dCTP using the DNA polymerase Klenow subunit. The M13 DNA digest was electrophoresed on a 2% agarose gel, dried and autoradiographed. Lane 1 shows the Msp I digestion pattern of unmethylated M13 mp8 DNA and lane 2 the pattern obtained with the Hpa II methylated DNA at an enzyme to DNA ratio of 24 units/μg for 1 hr at 37°C.
The pBR-DNA digest was electrophoresed on a 6% acrylamide gel, baked and autoradiographed. The Msp I digestion pattern (1 unit/μg) of the unmethylated pBR322 DNA is shown in lane 5 and that of the methylated plasmid in lane 4 (2 units/μg) and lane 3 (3 units/μg). The expected fragment sizes were obtained from sequencing data for these DNA samples (22,23).

The human γ globin genes are actively expressed in fetal liver, but are turned off and remain inactive in all adult tissues. In keeping with this, it was found that the CCGG site at the 5' end of the γ globin gene is undermethylated in fetal liver, but completely methylated at both cytosine residues in all human adult tissues tested (8). These results were suggested by restriction enzyme analysis showing that this site exists and is cut by Msp I in the cloned plasmid DNA but is uncuttable by both Msp I and Hpa II in genomic DNA from various tissues (7). In light of the results presented in figure 1, suggesting that Msp I may be inhibited by methylation at the internal cytosine, an experiment was designed to determine the true methylation state of this unique CCGG site at the 5' end of the γ globin gene. To this end several hundred micrograms of total genomic DNA from human lymphocytes was digested to completion with the enzyme Hind III and all of the DNA corresponding to the size class of 3.3kb was extracted from the gel. This extracted DNA was then subjected to Msp I digestion at an enzyme to DNA ratio of 150:1 and the resulting cleavage products were analyzed by gel electrophoresis and blot hybridi-

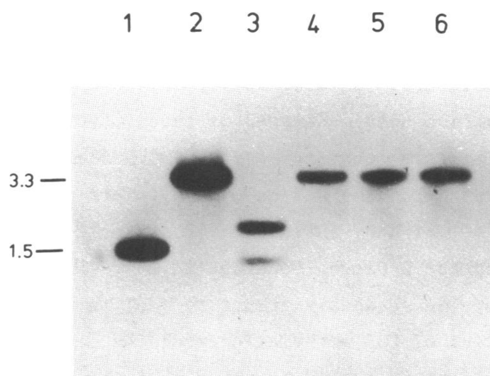

Fig. 4. Digestion of γ globin by Msp I. Human lymphocyte DNA was digested to completion with Hind III, run on a 1% agarose gel and the fraction corresponding to the 3.3kb size range was extracted from the gel. 1μg samples of this purified fraction were digested with various restriction enzymes blotted and hybridized using the nick translated Hind III/Hind III pHγ1 probe. Digestions were carried out at 37°C for 1 hr with 150 units Msp I (lane 3), 6 units Msp I (lane 4), 6 units Hpa II (lane 5) or no enzyme (lane 6). 60pg of the 1.5kb Hind III/BamHI fragment of pHγ1 (lane 1) and the 3.3kb Hind III fragment (lane 2) served as size markers (see figure 1). It should be noted that the human DNA used in this experiment lacked the Msp I site M3. Due to the pre-selection of the 3.3kb Hind III cut DNA, only the γ^A gene is visualized in this experiment.

zation analysis (figure 4). The results clearly show that this site remains uncut at relatively low concentrations of Msp I or Hpa II but is totally cleaved by a large enzyme excess suggesting that this site is indeed methylated at the internal cytosine but is totally unmethylated at the external cytosine residue.

DISCUSSION

The enzyme Msp I which recognizes and cleaves the tetranucleotide sequence CCGG is known to be prevented from cleavage by methylation at the external cytosine, but generally unaffected by methylation at the internal cytosine. It is this fact which makes this enzyme useful for measuring CpG methylation in animal cells when used in conjunction with the isoschizomer Hpa II. The results of this research show that Msp I is also inhibited by methylation at the internal cytosine at certain specific CCGG sites. One example of such a sequence is CCGGCC (equivalent to GGCCGG) which appears in the 5' promoter region of the human γ globin gene and is found several times in M13 phage DNA and pBR322 plasmid DNA. Methylation of the internal cytosine of

this sequence drastically reduces the cutting efficiency of Msp I although large enzyme excess overcomes this inhibition. CpG methylation of other CCGG sequences within pBR322 also affects, to varying degrees, the ability of Msp I to cleave its recognition site. Several of these CCGG sites which were affected by CpG methylation were found to have G-C rich flanking sequences suggesting that local hydrogen bonding may have some effect on the ability of Msp I to cut methylated DNA. It has, in fact, been suggested that methyl moieties themselves may strengthen hydrogen bonding (24,25). Msp I recognition may indeed be determined by DNA secondary structure (26) in which case one would predict that the effect of CpG methylation would be dependent on the flanking sequences. In this light it should be noted that the molecular structure of crystals of the tetranucleotide sequence CCGG (27) differs from that of the sequence GGCCGGCC (26) as determined by X-ray diffraction studies. In both cases the DNA structure differed from the standard B-form. We have not completely characterized the nature of the flanking sequences which causes this effect, and it is possible that other methylations at CpG moieties in the region of the CCGG sites also affect the activity of Msp I. M13 RF hemimethylated et every cytosine residue is resistant to digestion by many enzymes which are sensitive to methylated cytosine in their restriction site (28). In addition, we have shown in our laboratory that other enzymes such as Xho I, EcoR I and Hind III which normally recognize 6-methyladenine in their restriction site are also inhibited from cleaving the hemimethylated M13 molecule, suggesting that 5-methylcytosine either within the site itself or in its flanking sequences may have an affect on the specificity of these enzymes.

It is important when using restriction enzymes on animal cell DNA to keep in mind the possible effect of flanking sequences and the effect of flanking or adjacent methylations, since these factors can alter the digestion pattern at specific sites. A case in point is the CCGG site located 50 nucleotides upstream from the transcription start of the human γ globin gene. Although it was thought that this site is methylated at both cytosine residues, our experiments indicate that only the CpG site is indeed modified. Furthermore, additional experiments in which cellular DNA was digested in the presence of a plasmid DNA internal standard showed that the cleavage of the cellular DNA lagged behind that of the plasmid. This suggests that many Msp I sites require large enzyme to DNA excess for complete digestion.

It is well accepted that the large majority of methylated cytosines are located at CpG residues, but some biochemical and restriction enzyme analyses suggest, but do not prove, that CpC sites may be infrequently modified (5).

Our results indicate that one cannot rely on data obtained from Msp I analysis. Using a sensitive nearest neighbor analysis assay of animal cell DNA (18) it can be shown that each of the dinucleotides CpC, CpT and CpA are less than 0.5% methylated. While isolated mammalian methylases modify mainly CpG residues, these enzymes are capable of modifying other cytosine containing dinucleotides at a very low efficiency (A. Razin, personal communication). When DNA methylated at every cytosine was inserted into L-cells by DNA mediated gene transfer, only CpG methylations were inherited to future generations of cells (12). Thus, even if some infrequent CpC methylations were present in animal cell DNA these would probably be removed by selective methylation of CpG sites by the maintenance methylase. If other CpC methylations are indeed present and inheritable in animal cells, these must be associated with highly specific flanking sequences. It should be noted that CpC methylations do exist in higher plant cells (29) and are localized to CCG residues. This type of methylation is probably present in plant cells but not in animal cells since only plants have a special mechanism for inheriting such methyl moieties which are found in the symmetrical trinucleotide prototype sequence CXG. The inheritance mechanisms in animal cells appear to be limited to the symmetrically methylated dinculeotide sequence CpG.

ACKNOWLEDGEMENTS

We thank Drs. R. Flavell and M. Busslinger for providing the pHY1 plasmid used in this study. This research was supported by grants from the N.I.H. and from the U.S.-Israel Binational Science Foundation.

REFERENCES

1. Razin, A. and Riggs, A.D. (1980) Science 210, 604-610.
2. Sneider, T.W. (1971) J. Biol. Chem. 246, 4774-4783.
3. Sneider, T.W. (1972) J. Biol. Chem. 247, 2872-2875.
4. Salomon, R. and Kaye, A.M. (1970) Biochim. Biophys. Acta 204, 340-351.
5. Sneider, T.W. (1980) Nucleic Acids Res. 8, 3829-3840.
6. Jentsch, S., Günthert, V. and Trautner, T.A. (1981) Nucleic Acids Res. 9, 2753-2759.
7. van der Ploeg, L.H.T., Groffen, J. and Flavell, R.A. (1980) Nucleic Acids Res. 8, 4563-4574.
8. van der Ploeg, L.H.T. and Flavell, R.A. (1980) Cell 19, 947-958.
9. McKeon, L., Ohkubo, H., Pastan, I. and Crombrugghe, B. (1982) Cell 29, 203-210.
10. Haigh, L.S., Owens, B.B., Hellewell, S. and Ingram, V.M. (1982) Proc. Natl. Acad. Sci. (USA) 79, 5332-5336.
11. Kankel, L.M., Smith, K.D., Boyer, S.H., Borgaonkar, D.S., Wachtel, S.S. Miller, O.J., Breg, W.R., Jones, H.W. Jr. and Rary, J.M. (1977) Proc. Natl. Acad. Sci. (USA) 74, 1245-1249.
12. Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A. and Cedar, H. (1982) Proc.

Natl. Acad. Sci. (USA) 79, 61-65.
13. Clewell, D.B. (1972) J. Bacteriol. 110, 667-676.
14. Messing, J., Gronenborn, B., Muller-Hill, B., Hofschneider, P.H. (1977) Proc. Natl. Acad. Sci. (USA) 74, 3642-3646.
15. Mann, M.B. and Smith, H.O. (1977) Nucleic Acids Res. 4, 4211-4221.
16. Quint, A. and Cedar, H. (1981) Nucleic Acids Res. 9, 633-646.
17. Drouin, J. (1980) J. Mol. Biol. 140, 15-34.
18. Gruenbaum, Y., Stein, R., Cedar, H. and Razin, A. (1981) FEBS Lett. 124, 67-71.
19. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. (USA) 77, 5201-5205.
20. Vogelstein, B. and Gillespie, D. (1979) Proc. Natl. Acad. Sci. (USA) 76, 615-619.
21. Shen, S., Slighton, J.L. and Smithies, O. (1981) Cell 26, 191-203.
22. van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Gene 11, 129-148.
23. Sutcliffe, J.G. (1979) Cold Spring Harbor Symposium 43, 77-90.
24. Ehrlich, M., Ehrlich, K. and Mayo, J.A. (1975) Biochim. Biophys. Acta 395, 109-119.
25. Gill, J.E., Mazrimas, J.A. and Bishop, C.C. Jr. (1974) Biochim. Biophys. Acta 335, 330-348.
26. Wang, A.H.J., Fujii, S., van Boom, J.H. and Rich, A. (1982) Proc. Natl. Acad. Sci. (USA) 79, 3968-3972.
27. Conner, B.N., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R.E. (1982) Nature 295, 294-299.
28. Gruenbaum, Y., Cedar, H. and Razin, A. (1981) Nucleic Acids Res. 9, 2509-2515.
29. Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) Nature (LOndon) 292, 860-862.