Ann. Ist. Super. Sanità
Vol. 25, N. 1 (1989), pp. 149-154

DIFFERENT PATTERNS OF G/T AND A/C MISMATCH REPAIR IN SIMIAN CELLS CORRESPOND TO THE SPECIFICITY OF A MISMATCH BINDING PROTEIN ISOLATED FROM SIMIAN AND HELA CELLS

T.C. BROWN, N. CORMAN, B.B. RUDKIN and J. JIRICNY

Friedrich Miescher Institute, Basel, Switzerland

Summary. - *The two transition mispairs G/T and A/C are corrected with different efficiencies and specificities in CV-1 African green monkey kidney cells. G/T mispairs are corrected with 99% efficiency and almost always in favor of guanine, while A/C mispairs exhibit a 31% uncorrected sector and are otherwise randomly corrected. The higher correction efficiency and bias for the G/T mispair can be correlated with the substrate specificity of a protein, found in HeLa and CV-1 cell extracts, that binds selectively to oligonucleotide duplexes containing G/T mispairs. No binding can be detected to duplexes containing other mispairs, or to homoduplexes.*

Riassunto (Diversi meccanismi di riparazione dell'appaiamento errato G/T e A/C dipendono dalla specificità di una proteina isolata da cellule di scimmia ed HeLa che si lega alla coppia errata). - *L'appaiamento errato di basi del tipo transizione G/T e A/C è corretto con differente efficienza e specificità in cellule CV-1 derivate dal rene della scimmia verde africana. L'appaiamento errato G/T è corretto con una efficienza del 99% quasi sempre in favore della guanina, mentre l'appaiamento A/C non è corretto nel 31% dei casi ed il processo di correzione avviene casualmente. La più alta efficienza e preferenzialità di correzione della coppia G/T può essere correlata con la specificità di substrato di una proteina, identificata in estratti cellulari di HeLa e CV-1, che si lega selettivamente agli oligonucleotidi a doppio filamento contenenti la coppia scorretta G/T. Nessun legame si può osservare con oligonucleotidi che contengono altri appaiamenti errati o hanno una sequenza corretta.*

Introduction

Mismatched bases are formed in mammalian cells during recombination of homologous but nonidentical sequences, occur as errors of DNA replication, and arise when 5-methylcytosine spontaneously deaminates to form thymine, creating a G/T mispair [1]. The efficiency and specificity of mismatch correction therefore influences many different genetic events including gene conversion [2-4], homogenization of repeated sequence families [5], generation of antibody diversity [6, 7], DNA replication fidelity [8, 9], and stabilization of 5-methylcytosine residues [10]. Because mismatched heteroduplexes arise as a consequence of different genetic processes, it is reasonable to suppose that they will not always be repaired in exactly the same way. Rather, patterns of repair may differ according to the genetic context in which mispairs occur.

Studies of mismatch repair in procaryotes confirm that repair patterns differ according to the circumstances of mismatch formation. Thus, mispairs formed as errors of DNA replication are corrected in favor of the parental strand [11], mismatched heteroduplexes formed during recombination are corrected randomly, with extensive co-repair of separated markers [12, 13], and G/T mispairs arising through 5-methylcytosine deamination are restored to G/C pairs [14-17]. In *Escherichia coli (E. coli)*, these dissimilar patterns of mismatch repair are mediated by repair pathways requiring different, although overlapping, sets of gene products [1].

We have developed a way of introducing specific mispairs into the genome of Simian Virus 40 (SV40) and of determining the fate of the mispaired bases in simian cells [10]. Mindful of the possible correlation between mismatch repair patterns and mismatch origin, we have sought evidence for differences in the correction of the two transition mispairs G/T and A/C. Hydrolytic deamination of 5-methylcytosine gives rise only to G/T, while both G/T and A/C can occur through recombination or during DNA replication. We recently reported that G/T mispairs are corrected with 99% efficiency and mostly in favor of guanine [10]. We proposed that this efficient and biased pattern of correction was mediated by a G/T-specific repair pathway that restores G/C pairs lost through 5-methylcytosine deamination. We are now

able to substantiate this hypothesis in two ways. First, we show that A/C mispairs are corrected randomly and relatively inefficiently. Second, we have identified a protein, present in mammalian cell nuclei, that binds specifically to DNA duplexes containing a G/T, but not an A/C, mispair.

Preparation of mismatched SV40 DNA

We constructed SV40 DNA with specific mispairs by replacing a 21 bp sequence between the *Bst*XI and *Taq*I restriction sites (Fig. 1) with mismatched synthetic 12 bp duplexes [10]. DNA from wild type SV40 (strain 776) was digested with *Bst*XI and then with *Taq*I. After each digestion linear DNA was isolated after electrophoresis to exclude partially digested circular molecules and to eliminate the 21 bp fragment between these restriction sites. Cleavage with these two enzymes produced linear DNA with noncompatible sticky ends (Fig. 1). We then ligated aliquots of this DNA to the duplexes shown in Table 1. Each synthetic oligonucleotide of a

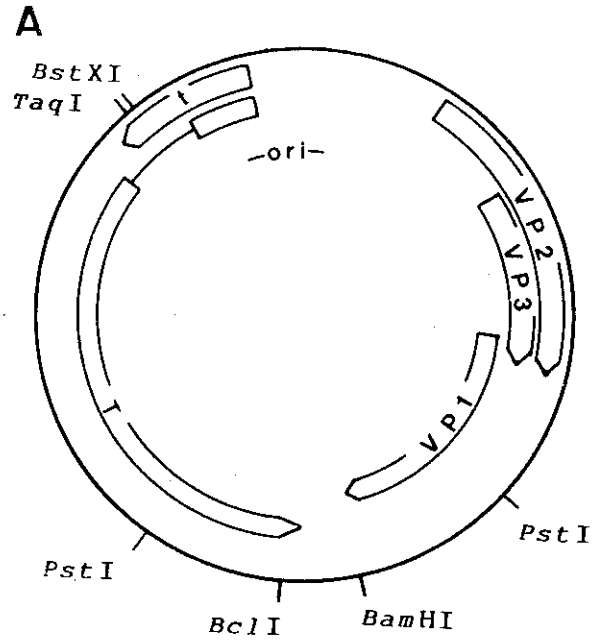

Fig. 1. - Construction of SV40 with mismatched base pairs. A: map of the SV40 genome showing restriction sites relevant to this study. B: sequences surrounding the *Bst*XI and *Taq*I restriction sites. Arrows indicate sites of cleavage.

Table 1. - *Heteroduplexes ligated into SV40 DNA between* BstXI *and* TaqI. *Each strand of the heteroduplex contains a restriction enzyme recognition sequence as indicated*

Mispair	Duplex
GT1	*Bam*HI CGTGATC^{G}GATCCCACAA ACTAG$_{T}$CTAGGG *Bcl*I
GT2	*Bcl*I CGGGGATC^{T}GATCAACAA CCCTAG$_{G}$CTAGT *Bam*HI
GT3	*Cla*I CGGATCGA^{T}TCGAGACAA CTAGCT$_{G}$AGCTC *Xho*I
AC1	*Bgl*II CGCGATC^{A}GATCTCACAA GCTAG$_{C}$CTAGAG *Pvu*I
AC2	*Bam*HI CGGGATC^{C}GATCACACAA CCTAG$_{A}$CTAGTG *Bcl*I

duplex pair contained a different restriction enzyme recognition sequence. Duplexes were mismatched at the 5' terminal bases of the two different restriction sites. Single-stranded ends complementary to the sticky ends of the viral DNA ensured efficient ligation in a defined orientation.

We isolated circular, form II viral DNA and used it to transfect host CV-1 simian cells. Transfection produced plaques, each plaque corresponding to a productive infection initiated by a single viral DNA molecule. Duplexes were designed so that correction of the mispair in the host cell would create distinct restriction sites. Noncorrection, or viral DNA replication before repair occurred, would generate a mixture of viral DNA molecules, some with one restriction site and some with the other. Repair patterns could therefore be determined by diagnostic restriction analysis of DNA derived from individual plaques.

G/T and A/C mispairs are corrected with different efficiencies and specificities

We transfected CV-1 cells with SV40 DNA modified to contain duplex GT1, GT2 or GT3. Digestion of SV40 DNA derived from 347 plaques indicated that 314

mispairs were corrected to G/C (90.5%), 29 to A/T (8.4%), and 4 were uncorrected (1.2%). High repair efficiency and strong G/C bias were observed in both GT1 and GT2, indicating that correction patterns do not depend on the orientation of the mispair within the SV40 genome, and in GT3, where the G/T mispair is present in a different sequence context.

Transfection of CV-1 cells with SV40 DNA modified to contain duplex AC1 or AC2 produced 91 plaques. Restriction analysis of DNA derived from these plaques indicated that 39 mispairs were corrected to G/C (42.9%), 24 were corrected to A/T (26.4%) and 28 were uncorrected (30.8%). As for G/T mispairs, correction patterns observed for A/C mispairs were about the same regardless of the orientation of the mismatched bases in the viral genome.

Repair data for G/T and A/C mispairs are summarized in Fig. 2. G/T mispairs were corrected with extremely high efficiency and mostly in favor of guanine. In contrast, A/C mispairs exhibited an uncorrected sector of 31%, and little repair bias.

Detection of a G/T-specific binding protein

In an effort to determine whether the efficient, biased correction of G/T mispairs is mediated by a specific repair pathway, we scanned HeLa cell extracts for proteins that bound to oligonucleotide duplexes containing G/T or A/C mispairs.

According to the rationale of these experiments, detection of one or more proteins binding exclusively to G/T-mismatched duplexes would attest to a repair pathway specific for this mispair.

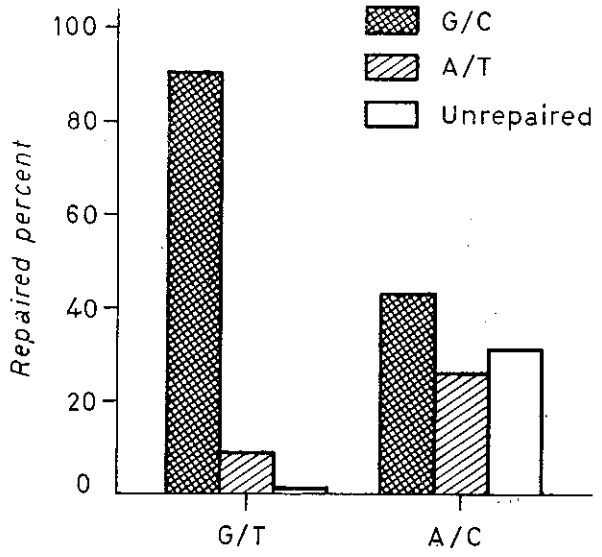

Fig. 2. - Correction patterns of G/T and A/C mispairs in CV-1 cells.

Binding substrates were prepared by annealing the synthetic 5'-^{32}P-labeled 34-mer 5'-AATTCCCGGGGAT CCGTCRACCTGCAGCCAAGCT-3' (**R** = **G** or **A**) to the respective unlabelled complementary strand 5'-AGCTTGGCTGCAGGTYGACGGATCCCCGGGAAT T-3' (**Y** = **T** or **C**) to yield homoduplexes G/C and A/T, and heteroduplexes G/T and A/C. HeLa whole cell extracts, prepared by the method of Manley [18], were incubated with the labeled oligonucleotide duplexes according to the procedure of Fried and Crothers [19]. The binding reactions were allowed to proceed at room temperature for 30 minutes. Protein-bound duplexes were resolved from unbound duplexes using 6% non-denaturing polyacrylamide gels. Fig. 3 shows the autoradiograph of the dried gel. This result indicates that the HeLa cell extract contains at least two factors that bind selectively to the duplex containing the G/T mispair. Similar results have been obtained using CV-1 cell extracts (results not shown).

Importance of selective G/T correction

Spontaneous deamination of cytosine is thought to occur at a rate of 100 per mammalian cell genome per day [20]. 5-Methylcytosine deaminates about 2.5 times more rapidly than cytosine at neutral pH [21]. Cells in which 5% of cytosines are methylated would therefore accumulate 12 G/T mispairs per day, or 2000 per year. Gradual mutational loss of 5-methylcytosine could seriously affect cell behavior because these modified bases are, as a class, particularly crucial DNA residues implicated in gene regulation [22-24], differentiation [25] and tumorigenesis [26-31]. Our results attest to a specific mismatch repair pathway that stabilizes the methylation pattern of mammalian cellular DNA by restoring G/C pairs whenever G/T mispairs arise through the deamination of 5-methylcytosine.

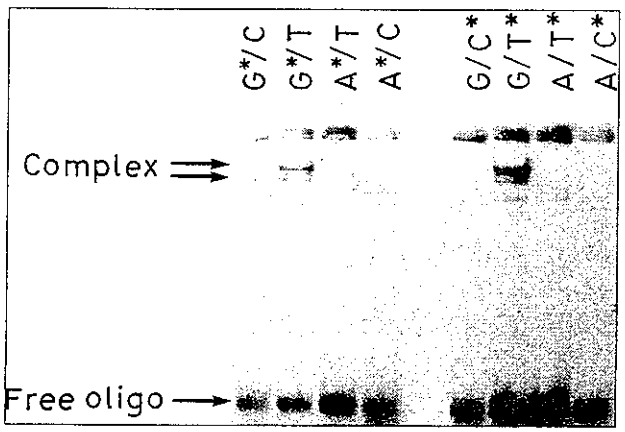

Fig. 3. - Binding of protein factors contained in HeLa whole cell extract to labeled oligonucleotide duplexes G/C, G/T, A/T and A/C. The duplexes were labeled either in the top strand or in the bottom strand. The asterisk denotes the labeled strand.

Selective correction of G/T mispairs replaces thymine with cytosine rather than 5-methylcytosine. Subsequent methylation of the restored cytosine depends on the specificity of DNA methyltransferase, and would be governed by the sequence [32] and structure [33] of DNA at the site of repair.

Despite the specificity of G/T mismatch repair in favor of guanine, it is likely that deamination of 5-methylcytosine occasionally leads to its loss. 5-Methylcytosine occurs mainly, if not exclusively, at ^{m}CpG dinucleotides (where ^{m}C is 5-methylcytosine) [34, 35]. The rarity [36] and instability [37] of CpG dinucleotides in mammalian cellular DNA indicates that ^{m}CpG dinucleotides sustain high mutation rates [38]. Certainly in *E. coli*, specific correction of G/T mispairs at presumptive sites of cytosine methylation [14-17] does not wholly abolish mutations attributable to 5-methylcytosine deamination [39]. Thus, while our results indicate that specific correction of G/T mispairs protects cells from loss of 5-methylcytosine, protection is probably not complete.

In light of the findings presented here, our observation that 5%-10% of G/T mispairs are corrected to A/T pairs is most easily explained by supposing that G/T mispairs are subject to two correction pathways. One, which we believe acts exclusively on G/T, is probably highly specific for correction in favor of G/C. The proteins found to bind selectively to G/T mismatched heteroduplexes may play a role in this specific pathway. A second pathway, which may act on all mispairs upon their occurrence during DNA replication or recombination, may be more random and may account for the majority of repair events in favor of thymine. Our results do not allow us to estimate what proportion of G/T mispairs in transfected SV40 DNA might be addressed by these two hypothetical pathways.

We have previously noted that inflexible bias in the correction of G/T mispairs to G/C would be mutagenic, since resolution of mispairs formed by incorporation of guanine opposite thymine during DNA replication would favor fixation of the mutation [10]. This difficulty is overcome by supposing that the G/T mismatch repair pathway specific for establishment of G/C does not act on newly replicated DNA, but is supplanted by a pathway correcting mispairs in favor of the parental strand [8]. According to this hypothesis, mismatch repair patterns in mammalian cells and in *E. coli* would share at least one point of similarity: correction of G/T mispairs resulting from replication error and from 5-methylcytosine deamination would be addressed by different, though overlapping, repair pathways.

Review submitted on invitation by the Editorial Board of the *Annali*. Accepted for publication: October 1988.

REFERENCES

1. MODRICH, P. 1987. DNA mismatch correction. *Annu. Rev. Biochem.* **56**: 435-466.
2. HOLLIDAY, R. 1974. Molecular aspects of genetic exchange and gene conversion. *Genetics* **78**: 273-287.
3. WHITE, J.H., LUSNAK, K. & FOGEL, S. 1985. Mismatch-specific post-meiotic segregation frequency in yeast suggests a heteroduplex recombination intermediate. *Nature* **315**: 350-352.
4. KOURILSKY, P. 1986. Molecular mechanisms for gene conversion in higher cells. *Trends Genet.* **2**: 60-63.
5. DOVER, G.A. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet.* **2**: 159-165.
6. BALTIMORE, D. 1974. Is terminal deoxynucleotidyl transferase a somatic mutagen in lymphocytes? *Nature* **248**: 409-411.
7. KUNKEL, T.A., GOPINATHAN, K.P., DUBE, D.K., SNOW, E.T. & LOEB, L.A. 1986. Rearrangements of DNA mediated by terminal transferase. *Proc. Natl. Acad. Sci. USA* **83**: 1867-1871.
8. HARE, J.T. & TAYLOR, H. 1985. One role of DNA methylation in vertebrate cells is strand discrimination in mismatch repair. *Proc. Natl. Acad. Sci. USA* **82**: 7350-7354.
9. REYLAND, M.E. & LOEB, L.A. 1987. On the fidelity of DNA replication. *J. Biol. Chem.* **262**: 10824-10830.
10. BROWN, T.C. & JIRICNY, J. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* **50**: 945-950.
11. RADMAN, M. & WAGNER, R. 1986. Mismatch repair in *Escherichia coli*. *Annu. Rev. Genet.* **20**: 523-538.
12. FISCHEL, R.A., SIEGEL, E.C. & KOLODNER, R. 1983. The identification of two repair pathways for mismatched nucleotides. *UCLA Symp. Mol. Cell. Biol.* **11**: 309-326.
13. FISHEL, R.A., SIEGEL, E.C. & KOLODNER, R. 1986. Gene conversion in *Escherichia coli*. Resolution of heteroallelic mismatched nucleotides by co-repair. *J. Mol. Biol.* **188**: 147-157.

14. LIEB, M. Recombination in the lamba repressor gene: evidence that very short patch (VSP) mismatch correction restores a specific sequence. *Mol. Gen. Genet.* **199**: 465-470.

15. LIEB, M., ALLEN, E. & READ, D. 1986. Very short patch mismatch repair in phage lambda: repair sites and length of repair tracts. *Genetics* **144**: 1041-1060.

16. JONES, M., WAGNER, R. & RADMAN, M. 1987. Mismatch repair of deaminated 5-methyl-cytosine. *J. Mol. Biol.* **194**: 155-159.

17. ZELL, M.J. & FRITZ, H.J. 1987. DNA mismatch repair in *Escherichia coli* counteracting the hydrolytic deamination of 5-methyl-cytosine residues. *EMBO J.* **6**: 1809-1815.

18. MANLEY, J.L. 1984. Transcription of eukaryotic genes in a whole-cell extract. In: *Transcription and translation, a practical approach.* B.D. Hames & S.J. Higgins (Eds). IRL Press, Oxford. pp. 71-80.

19. FRIED, M.G. & CROTHERS, D.M. 1983. CAP and RNA polymerase interactions with the *lac* promotor: binding stoichiometry and long range effects. *Nucleic Acids Res.* **11**: 141-158.

20. LINDHAL, T. 1982. DNA repair enzymes. *Annu. Rev. Biochem.* **51**: 61-87.

21. WANG, R.Y.H., KUO, K.C., GEHRKE, C.W., HUANG, L.H. & EHRLICH, M. 1982. Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA. *Biochim. Biophys. Acta* **697**: 371-377.

22. DOERFLER, W. 1983. DNA methylation and gene activity. *Annu. Rev. Biochem.* **52**: 93-124.

23. BIRD, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209-213.

24. SALUZ, H.P., JIRICNY, J. & JOST, J.P. 1986. Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA methylation of the overlapping estradiol-glucocorticoid receptor binding sites and the rate of avian vitellogenin synthesis. *Proc. Natl. Acad. Sci. USA* **83**: 7167-7171.

25. RAZIN, A., SZYF, M., KAFRI, T., ROLL, M., GILOH, H., SCARPA, S., CAROTTI, D. & CANTONI, G.L. 1986. Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA methylation during differentiation. *Proc. Natl. Acad. Sci. USA* **83**: 2827-2831.

26. HOLLIDAY, R. 1979. A new theory of carcinogenesis. *Br. J. Cancer* **40**: 513-522.

27. RIGGS, A.D. & JONES, P.A. 1983. 5-methylcytosine, gene regulation and cancer. *Adv. Cancer Res.* **40**: 1-30.

28. KASTAN, M.B., GOWANS, B.J. & LIEBERMAN, M.W. 1982. Methylation of deoxycytidine by excision-repair synthesis of DNA. *Cell* **30**: 509-516.

29. WILSON, V.L. & JONES, P.A. 1983. Inhibition of DNA methylation by chemical carcinogens *in vitro*. *Cell* **32**: 239-246.

30. GAMA-SOZA, M.A., SLAGEL, V.A., TREWYN, R.W., OXENHANDLER, R., KUO, K.C., GEHRKE, C.W. & EHRLICH, M. 1983. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11**: 6883-6894.

31. BARR, F.G., RAJAGOPALAN, S., MACARTHUR, C.A. & LIEBERMAN, M.W. 1986. Genomic hypomethylation and far 5' sequence alterations are associated with carcinogen-induced activation of the hamster thymidine kinase gene. *Mol. Cell. Biol.* **6**: 3023-3033.

32. BOLDEN, A.H., NALIN, C.M., WARD, C.A., POONIAN, M.S. & WEISBACH, A. 1986. Primary sequence determines sites of maintenance and *de novo* methylation by mammalian DNA methyltranferase. *Mol. Cell. Biol.* **6**: 1135-1140.

33. BESTOR, T. 1987. Supercoiling-dependent sequence specificity of mammalian DNA tranferase. *Nucleic Acids Res.* **15**: 3835-3843.

34. BIRD, A. 1978. Use of restriction enzymes to study eucaryotic DNA methylation. II. The symmetry of methylated sites supports semiconservative copying of the merthylation patterns. *J. Mol. Biol.* **118**: 49-60.

35. CEDAR, H., SOLANGE, A., GLASER, G. & RAZIN, A. 1979. Direct detection of methylated cytosine in DNA by use of the restriction enzyme *Msp*I. *Nucleic Acids Res.* **6**: 2125-2132.

36. BIRD, A.P. 1980. DNA methylation and the frequency of CpG in animal cells. *Nucleic Acids Res.* **8**: 1499-1504.

37. BARKER, D., SCHAFER, M. & WHITE, R. 1984. Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* **36**: 131-138.

38. BIRD, A.P. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342-347.

39. COULONDRE, C., MILLER, J.H., FARABAUGH, P.J. & GILBERT, W. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* **274**: 775-780.