N6-methyladenine:

The other methylated base of DNA.

David Ratel, Jean-Luc Ravanat¤, François Berger and Didier Wion*.

INSERM U318, CHU Michallon, 38043 Grenoble cedex 09, France ; ¤ Laboratoire Lesions des Acides Nucleiques Service de Chimie Inorganique et Biologique, FRE 2600, Departement de Recherche Fondmentale sur la Matiere Condensée, CEA Grenoble, France.

Corresponding author : didier.wion@ujf-grenoble.fr

Abstract.

DNA methylation is an epigenetic mechanism involved in many biological functions in prokaryotes and eukaryotes. Contrary to mammalian DNA, which is thought to contain only 5-methylcytosine (m5C), bacterial DNA contains two additional methylated bases, namely N6-methyladenine (m6A), and a more recently discovered minor base N4-methylcytosine (m4C). These modified bases are involved in the protection of bacterial DNA from the action of specific endonucleases via the host-specific restriction-modification system which is regarded as a defense mechanism against bacteriophage infection. However, if the main function of m5C and m4C in bacteria is the protection against restriction enzymes, the roles of m6A are multiple and include for example the regulation of virulence and the control of many bacterial DNA functions such as the replication, repair, expression and transposition of DNA. Hence, in regard to the multiple roles of m6A in bacteria, and to the well known tendency for m5C to deaminate in thymine, the selection of the mutagenic m5C instead of m6A in

mammals as the only methylated base may seem surprising. However, even if adenine methylation is usually considered as a bacterial DNA feature, the presence of m6A is not restricted to prokaryotic DNA since this methylated base has been found in protist and plant DNAs. Furthermore, indirect evidence suggests the presence of m6A in mammal DNA, raising the possibility that this base has remained undetected due to the low sensitivity of the analytical methods used. This points to the importance to consider m6A as the sixth element of DNA.

Keywords: DNA methylation; epigenetics; N6-methyladenine; 5-methylcytosine.

I. Introduction.

Although four bases are required for DNA synthesis, DNA contains several additional methylated bases, namely 5-methylcytosine, N6-methyladenine, N4-methylcytosine, which result from the post-replicative modification of DNA by DNA methylases (1-3). However, if the most popular modified base is 5-methylcytosine (m5C), recent data point to the biological importance of N6-methyladenine (m6A) as the other methylated base. Thus, adenine methylation is essential for the viability of several bacteria (4-9). Moreover, accumulating evidence suggests that the presence of m6A is not limited to eubacterial DNA but also occurs in at least some archaeabacteria and eukaryotic cells where its role remains largely unknown (Fig. 1) (10-14) . This review focuses on m6A as the sixth base of DNA and aims to be a source of information and inspiration for the development of new ideas and hypotheses on the possible functions of m6A in eukaryotic cells.

II. DNA adenine methylation in prokaryotes.

In bacterial DNA the function of adenine methylation has historically been associated with the protection of DNA from the action of specific endonucleases via the host-specific restriction/modification system. In this system, which is regarded as a defense mechanism against bacteriophage infection (15,16), cytosine and adenine methylation of bacterial DNA protects it from the action of the corresponding restriction endonuclease, whereas unmethylated sites of foreign DNA such as bacteriophage DNA are cleaved (Fig. 2A). Interestingly, DNA adenine methyltransferases can also be encoded by viral DNA as has been shown for bacteriophages T4, Mx8 (17, 18) and for the archeal viruses ϕ.Ch1 and SNDV (19, 20). In addition to these DNA adenine methyltransferases related to the restriction/modification system, there are at least two DNA adenine methyltransferases which

lack a cognate restriction enzyme. These DNA methylases named Dam and CcrM, methylate GATC and GANTC sequences respectively. They differ in their distribution since Dam methylation is found primarily in the members of the gamma division of Protobacteria and in some Archeabacteria (21), whereas CcrM is relatively widespread in the eubacteria alpha division (22, 23). In *E. coli*, Dam is involved in the replication, mismatch repair and transposition of DNA (Fig. 2B) (24, 25), and in the control of gene expression (26). Hence, Dam mutant are characterized by a pleiotropic phenotype including for example enhanced sensitivity to DNA-damaging agents, higher mutability and increased recombination frequency. However, *E. coli* lacking Dam activity are viable, but *dam* is an essential gene in *Vibrio cholerae* and *Yersinia pseudotubercolosis* (5). The other known DNA adenine methyltransferase lacking a corresponding restriction endonuclease is CcrM. CcrM is a DNA methyltransferase originally described in *Caulobacter cresecentus* as a "cell cycle-regulated methyltransferase" (26). It is essential for viability in *Caulobacter crescentus*, *Rhizobium meliloti*, *Agrobacterium tumefaciens*, *Brucella abortus* (4, 7, 8). Like Dam, CcrM regulates gene expression (6), and control the initiation of DNA replication (23, 28). In addition to their multiple functions which have led Dam and CcrM to be considered as "cell cycle regulators" (23), these DNA adenine methyltransferases are also involved in bacterial pathogenicity as they control virulence gene expression and secretion of virulence determinants (Fig. 2C) (5, 29-32). These points are of particular interest since they suggest that DNA adenine methylation could be a new target for antibiotics (33, 34).

III. DNA adenine methylation in eukaryotes.

M6A in protist DNA.

It has long been known that m6A is present in DNA from several unicellular eukaryotes, including members of the genera Chlamydomonas, Chlorella, Oxytricha, Paramecium,

Tetrahymena (Fig. 2) (14, 35-41). Moreover, genes encoding both restriction endonucleases and their cognate adenine DNA methyltransferases have been found in Chlorella viruses, suggesting a role for adenine methylation during viral infection (42, 43). A main feature of ciliate protozoans such as Paramecia and Tetrhaymena, is the presence in the same cell of two nuclei, a diploid germ line micronucleus and a polyploid somatic macronucleus whose transcription supports cell growth, differentiation and proliferation (41). Interestingly, m6A is detected in macronucleus DNA only. Unfortunately, the exact functions of this adenine methylation are far less understood in protist than in bacteria. Studies performed on Tetrahymena have demonstrated that *de novo* adenine methylation of DNA is not random, but occurs according to a specific pattern and preferentially in linker DNA (38, 40, 44). Morever, it has been reported that in *Physarum flavicomumrecent*, cyst DNA but not growing cell DNA contains m6A (45), whilst recent data suggest that in Paramecium, adenine methylation could be involved in the excision of internal eliminated sequences (IESs) (46).

M6A in plant and animal DNA

The general assumption that m6A is not found in the DNA of higher eukaryotes originates from experiments performed more than thirty years ago which had a detection limit around 0.1% - 0.01% (47-49) and which detected m5C as the only methylated base. This selection of m5C instead of m6A to control crucial regulatory biological processes such as genomic imprinting, X-chromosome inactivation, gene expression and embryonic development may seem surprising if we consider the multiple roles of m6A in bacteria, and the well known tendency for m5C to deaminate in thymine (50). A possible explanation lies in the fact that most of the m5C in mammalian DNA is found in transposons (51), a finding that has led to the suggestion that the presence of m5C in mammalian DNA could provide a host defense mechanism against parasitic DNA through the repressive effects of m5C on gene expression and through the accumulation of mutations resulting from the spontaneous deamination of

m5C in thymine (52). Hence, the large amount of transposons (>45%) found in human DNA could account for the relative abundance of m5C which could have masked, by its overrepresentation, the presence of small amounts of m6A. In this respect it is noteworthy that experiments designed to determine the base composition of mammalian DNA were, in addition to their low sensitivity, performed on a limited number of tissue or cell samples. Consequently the occurrence of m6A during development, programmed cell death, aging or in pathologic conditions such as proliferative, degenerative or infectious diseases has not been extensively investigated. Furthermore, and in spite of common opinion, several data have reported the presence of m6A in the DNA of higher eukaryotes. Thus, m6A has been detected in plastid, mitochondrial and nuclear plant DNA (10, 13, 53-55), and in mosquito DNA (56). Regarding the presence of m6A in mammalian DNA, indirect evidence obtained using restriction enzymes sensitive to adenine methylation suggests the presence of m6A in the mouse Myo-D1 gene and in the rat type 2 steroid 5α-reductase gene (11, 12). In the case of the rat type 2 steroid 5α-reductase gene, the restriction pattern is correlated to its expression (11). Unfortunately, direct evidence for the presence of m6A in mammals, based on its physical detection by mass spectrometry, is still lacking. Indeed, it seems that the relative high abundance of 5mC in mammalian DNA has focused attention on the role of m5C to the detriment of the m6A quest. In this regard, it should be pointed out that at an overall content of <0.001% m6A can be biologically significant if it occurs, for example, in a regulated fashion on specific mammalian gene regulatory elements (Fig. 2D). Hence, according to the size of the human genome (3.3×10^9 bp) and to the number of genes ($\sim 30 \times 10^3$), the presence of a few hundred m6A can be sufficient to play a crucial role in the control of biological processes such as cell differentiation or morphogenesis. Interestingly, evidence suggests that m6A does affect the regulation of gene expression in mammalian cells. Thus, in mammalian cells, the artificial presence of m6A can affect the binding of a nuclear factor to its responsive

element (57), decrease the activity of adenoviral E1A promoter (58), or generate a steroid hormone response element (59). Incidentally, this point may be of special concern since all the plasmids currently used in transient gene expression experiments are subjected to adenine methylation as a consequence of the bacterial Dam and CcrM activities of *E. coli* (60, 61). In addition, several reports have also demonstrated the influence of m6A on the activity of plant gene promoters (62-64), while addition of m6A to mammalian cell cultures induces cell differentiation in several cell lines (65, 66).

III. Perspective and concluding remarks.

The fact that the essential roles played by DNA adenine methyltransferases in bacterial viability and virulence have been recognized only recently underscores the importance to investigate the presence and biological functions of adenine methylation in eukaryotic DNA. Thus, a precise knowledge of the adenine methylation status in human DNA is of crucial concern for the development of new antibiotics targeting bacterial DNA adenine methyltransferases. Moreover, the possibility that DNA adenine methylation could occur during the life-cycle of some parasites and virus should also be reconsidered, since, for example, the reported lack of m6A in the DNA of *Plasmodium falciparum* or adenovirus has been based on the use of low sensitive methods (67, 68). In this regard, if we speculate that one of the functions of adenine methylation in eukaryotic cells is a defense mechanism which marks foreign DNA in order to ensure its degradation, then, preparations of viral DNA used to quantify the presence of m6A are inevitably made of molecules which have escaped this methylation/degradation process and which are therefore described as unmethylated. Another possible function of adenine methylation could be to mark the immortal DNA strand suggested to be present in adult stem cells that divide by asymmetric mitosis (69-71). Thus, if we consider that the object of the immortal DNA strand is to protect the genome of stem cells

from mutations, then m6A is a conceivable alternative to the use of the mutagenic m5C (50). Hence, the presence of m6A in a very rare population of cells could provide another explanation of the difficulty in detecting m6A in mammals. A prerequisite to the "natural" methylation of adenine in mammalian DNA is the existence of at least one adenine-N^6-DNA methyltransferase gene in mammalian genomes. In contrast, to cytosine DNA methyltransferases, which belong to a family of conserved enzymes, bacterial adenine-N^6 methyltransferase are much more heterogeneous. Besides a weakly conserved F_G_G amino acid motif shared by all Mtases, DNA adenine-methyltransferases only contain one moderately conserved (D/N)PP(Y/F) motif (72). On the basis of computer analysis several putative DNA adenine methyltransferases have been identified in human and murine genomes (66, 73). Whether these genes encode true DNA adenine methylases or are the fossils of a restriction-modification system present either in the putative archea-like ancestor of eukaryotic cells or ancestral bacterial endosymbiont at the origin of mitochondria should warrant further investigations.

Acknowledgements: We thank Pr. AL. Benabid for his support and Dr. K. Herbert for critical reading of the manuscript. DW is supported by the Fondation Jérôme Lejeune

Fig. 1. Schematic representation of the phylogenetic distribution of m6A in DNA.

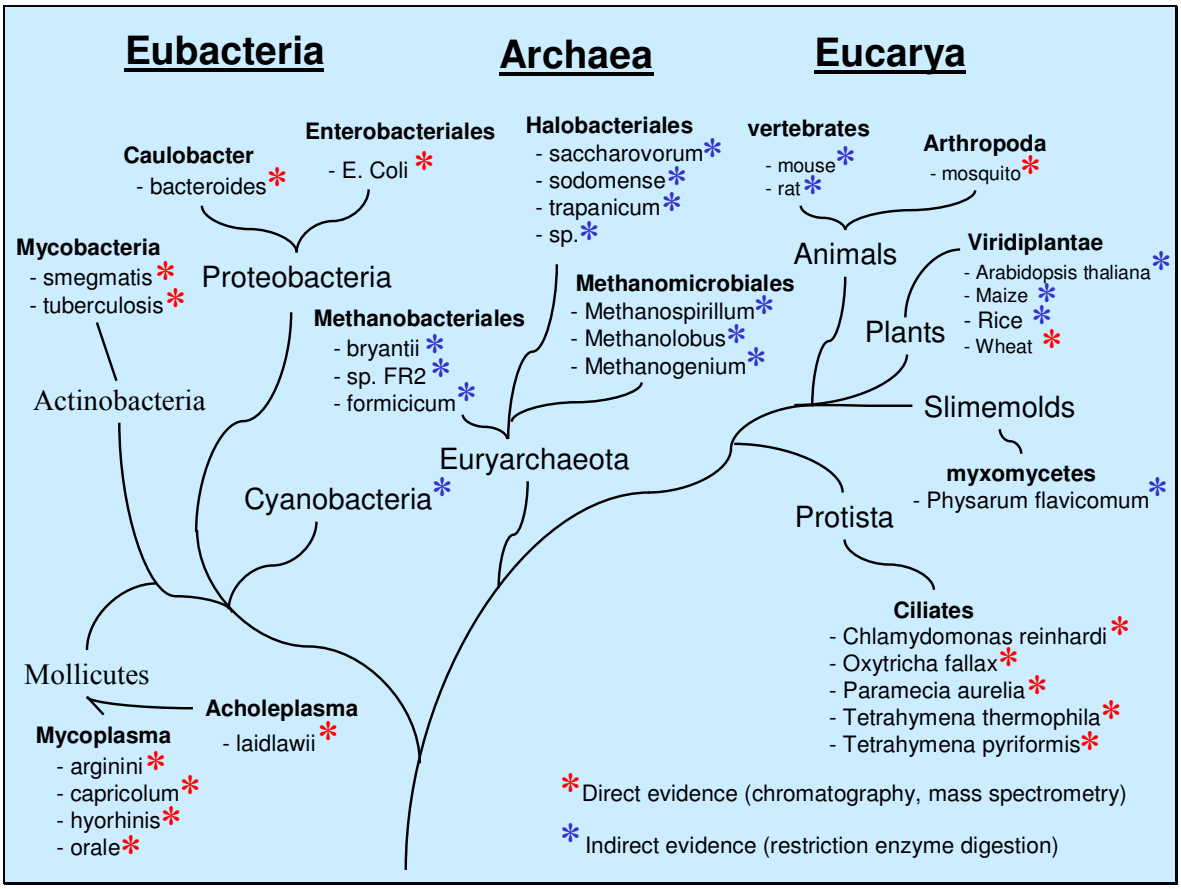
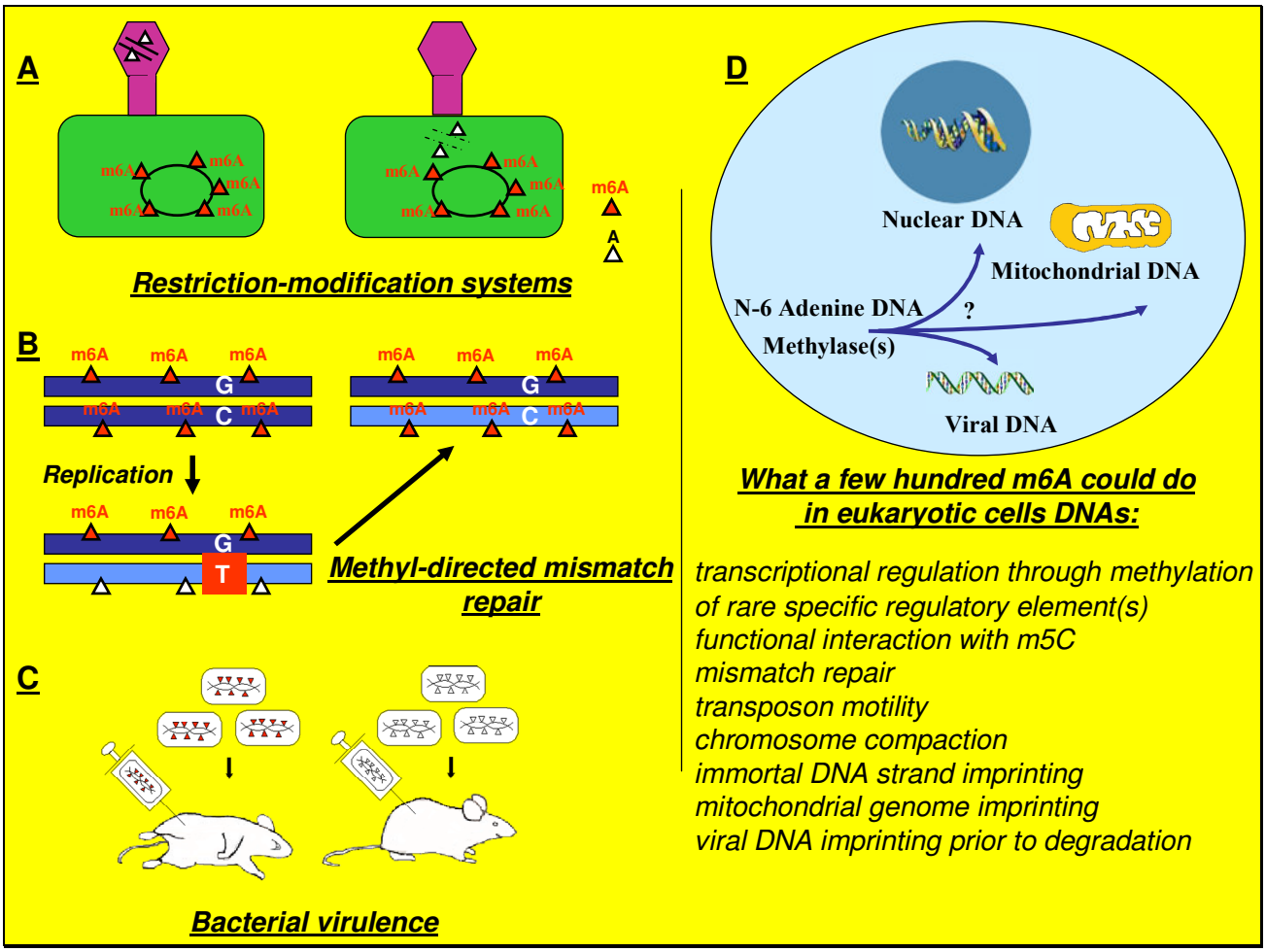


Fig.2. Some established and putative functions of m6A in DNA.



A) Restriction-modifications systems defend bacteria from invasions by viral DNA. These systems are composed of a bacterial restriction enzyme which cuts a specific sequence in phage DNA, and a cognate adenine or cytosine DNA methyltransferase which recognizes the same target site that the restriction enzyme and protects bacterial DNA from its own restriction enzyme(s). B) In E. coli, adenine methylation is also used to discriminate the transiently unmethylated newly synthesized complementary strand during DNA replication. This discrimination is the basis of the methyl-directed mismatch repair which makes correction on the newly synthesized strand only. C)

Alteration in the levels of DNA adenine methylation attenuates the virulence of a number of pathogens. D) Speculative representation of the biological functions of putative eucaryotic N-6 adenine-specific DNA methylase(s).

References.

1. J. Doskocil, Z. Sormo'Va, Biochim Biophys Acta 95, 513-5 (Mar 15, 1965).
2. D. B. Dunn, J. D. Smith, Biochem J 68, 627-36 (Apr, 1958).
3. M. Ehrlich et al., Nucleic Acids Res 13, 1399-412 (Feb 25, 1985).
4. L. S. Kahng, L. Shapiro, J Bacteriol 183, 3065-75 (May, 2001).
5. S. M. Julio et al., Infect Immun 69, 7610-5 (Dec, 2001).
6. A. Reisenauer, L. Shapiro, Embo J 21, 4969-77 (Sep 16, 2002).
7. G. T. Robertson et al., J Bacteriol 182, 3482-9 (Jun, 2000).
8. R. Wright, C. Stephens, L. Shapiro, J Bacteriol 179, 5869-77 (Sep, 1997).
9. C. Stephens, A. Reisenauer, R. Wright, L. Shapiro, Proc Natl Acad Sci U S A 93, 1210-4 (Feb 6, 1996).
10. V. V. Ashapkin, L. I. Kutueva, B. F. Vanyushin, FEBS Lett 532, 367-72 (Dec 18, 2002).
11. E. M. Reyes, I. Camacho-Arroyo, G. Nava, M. A. Cerbon, J Androl 18, 372-7 (Jul-Aug, 1997).
12. P. H. Kay et al., Gene 151, 89-95 (Dec 30, 1994).
13. J. A. Pintor-Toro, Biochem Biophys Res Commun 147, 1082-7 (Sep 30, 1987).
14. D. J. Cummings, A. Tait, J. M. Goddard, Biochim Biophys Acta 374, 1-11 (Nov 20, 1974).
15. S. E. Luria, M. L. Human, J Bacteriol 64, 557-69 (Oct, 1952).
16. G. Bertani, J. J. Weigle, J Bacteriol 65, 113-21 (Feb, 1953).
17. S. L. Schlagman, S. Hattman, Gene 22, 139-56 (May-Jun, 1983).

18. V. Magrini et al., J Bacteriol 179, 4254-63 (Jul, 1997).

19. U. Baranyi, R. Klein, W. Lubitz, D. H. Kruger, A. Witte, Mol Microbiol 35, 1168-79 (Mar, 2000).

20. H. P. Arnold, U. Ziese, W. Zillig, Virology 272, 409-16 (Jul 5, 2000).

21. D. Lodwick, H. N. Ross, J. E. Harris, J. W. Almond, W. D. Grant, J Gen Microbiol 132, 3055-9 (Nov, 1986).

22. T. Barbeyron, K. Kean, P. Forterre, J Bacteriol 160, 586-90 (Nov, 1984).

23. A. Reisenauer, L. S. Kahng, S. McCollum, L. Shapiro, J Bacteriol 181, 5135-9 (Sep, 1999).

24. F. Barras, M. G. Marinus, Trends Genet 5, 139-43 (May, 1989).

25. A. Lobner-Olesen, O. Skovgaard, M. G. Marinus, Curr Opin Microbiol 8, 154-60 (Apr, 2005).

26. T. Oshima et al., Mol Microbiol 45, 673-95 (Aug, 2002).

27. G. Zweiger, G. Marczynski, L. Shapiro, J Mol Biol 235, 472-85 (Jan 14, 1994).

28. E. Boye, A. Lobner-Olesen, Cell 62, 981-9 (Sep 7, 1990).

29. B. A. Braaten, X. Nou, L. S. Kaltenbach, D. A. Low, Cell 76, 577-88 (Feb 11, 1994).

30. D. M. Heithoff, R. L. Sinsheimer, D. A. Low, M. J. Mahan, Science 284, 967-70 (May 7, 1999).

31. F. Garcia-Del Portillo, M. G. Pucciarelli, J. Casadesus, Proc Natl Acad Sci U S A 96, 11578-83 (Sep 28, 1999).

32. D. A. Low, N. J. Weyand, M. J. Mahan, Infect Immun 69, 7197-204 (Dec, 2001).

33. N. Mashhoon et al., J Biol Chem 279, 52075-81 (Dec 10, 2004).

34. D. C. Wahnon, V. K. Shier, S. J. Benkovic, J Am Chem Soc 123, 976-7 (Feb 7, 2001).

35. S. Hattman, C. Kenny, L. Berger, K. Pratt, J Bacteriol 135, 1156-7 (Sep, 1978).

36. M. A. Gorovsky, S. Hattman, G. L. Pleger, J Cell Biol 56, 697-701 (Mar, 1973).

37. P. M. Rae, B. B. Spear, Proc Natl Acad Sci U S A 75, 4992-6 (Oct, 1978).

38. K. Pratt, S. Hattman, Mol Cell Biol 1, 600-8 (Jul, 1981).

39. E. E. Capowski, J. M. Wells, G. S. Harrison, K. M. Karrer, Mol Cell Biol 9, 2598-605 (Jun, 1989).

40. G. S. Harrison, R. C. Findly, K. M. Karrer, Mol Cell Biol 6, 2364-70 (Jul, 1986).

41. J. C. Gutierrez, S. Callejas, S. Borniquel, A. Martin-Gonzalez, Int Microbiol 3, 139-46 (Sep, 2000).

42. Y. Zhang, M. Nelson, J. W. Nietfeldt, D. E. Burbank, J. L. Van Etten, Nucleic Acids Res 20, 5351-6 (Oct 25, 1992).

43. Q. Que et al., Gene 190, 237-44 (May 6, 1997).

44. K. M. Karrer, T. A. VanNuland, Nucleic Acids Res 30, 1364-70 (Mar 15, 2002).

45. C. M. Zhu, H. R. Henney, Jr., Biochem Cell Biol 68, 944-8 (Jun, 1990).

46. S. Nakayama, H. Endoh, Genes Genet Syst 78, 391-8 (Dec, 2003).

47. B. F. Vanyushin, S. G. Tkacheva, A. N. Belozersky, Nature 225, 948-9 (Mar 7, 1970).

48. P. D. Lawley, A. R. Crathorn, S. A. Shah, B. A. Smith, Biochem J 128, 133-8 (Jun, 1972).

49. U. Gunthert, M. Schweiger, M. Stupp, W. Doerfler, Proc Natl Acad Sci U S A 73, 3923-7 (Nov, 1976).

50. A. Poole, D. Penny, B. M. Sjoberg, Nat Rev Mol Cell Biol 2, 147-51 (Feb, 2001).

51. T. H. Bestor, Hum Mol Genet 9, 2395-402 (Oct, 2000).

52. J. A. Yoder, C. P. Walsh, T. H. Bestor, Trends Genet 13, 335-40 (Aug, 1997).

53. B. F. Vanyushin, N. I. Alexandrushkina, M. D. Kirnos, FEBS Lett 233, 397-9 (1988).

54. M. S. Dhar, V. V. Pethe, V. S. Gupta, P. K. Ranjekar, Theor Appl Genet 80, 402-8 (1990).

55. J. Ngernprasirtsiri, T. Akazawa, Eur J Biochem 194, 513-20 (Dec 12, 1990).

56. R. L. Adams, E. L. McKay, L. M. Craig, R. H. Burdon, Biochim Biophys Acta 563, 72-81 (Jun 20, 1979).

57. F. Tronche, A. Rollier, I. Bach, M. C. Weiss, M. Yaniv, Mol Cell Biol 9, 4759-66 (Nov, 1989).

58. D. Knebel, W. Doerfler, J Mol Biol 189, 371-5 (May 20, 1986).

59. M. Truss, J. Bartsch, G. Chalepakis, M. Beato, Nucleic Acids Res 20, 1483-6 (Apr 11, 1992).

60. F. Berger, C. Canova, A. L. Benabid, D. Wion, Nat Biotechnol 17, 517 (Jun, 1999).

61. S. Allamane et al., Biochem Biophys Res Commun 276, 1261-4 (Oct 5, 2000).

62. K. Sugimoto, S. Takeda, H. Hirochika, Plant J 36, 550-64 (Nov, 2003).

63. J. C. Rogers, S. W. Rogers, Plant J 7, 221-33 (Feb, 1995).

64. M. W. Graham, P. J. Larkin, Transgenic Res 4, 324-31 (Sep, 1995).

65. D. Ratel et al., Biochem Biophys Res Commun 285, 800-5 (Jul 20, 2001).

66. M. P. Charles et al., Biochem Biophys Res Commun 314, 476-82 (Feb 6, 2004).

67. Y. Pollack, N. Kogan, J. Golenser, Exp Parasitol 72, 339-44 (May, 1991).

68. U. Wienhues, W. Doerfler, J Virol 56, 320-4 (Oct, 1985).

69. J. Cairns, Nature 255, 197-200 (May 15, 1975).

70. J. R. Merok, J. A. Lansita, J. R. Tunstead, J. L. Sherley, Cancer Res 62, 6791-5 (Dec 1, 2002).

71. C. S. Potten, G. Owen, D. Booth, J Cell Sci 115, 2381-8 (Jun 1, 2002).

72. A. Jeltsch, F. Christ, M. Fatemi, M. Roth, J Biol Chem 274, 19538-44 (Jul 9, 1999).

73. B. Y. Shorning, B. F. Vanyushin, Biochemistry (Mosc) 66, 753-62 (Jul, 2001).