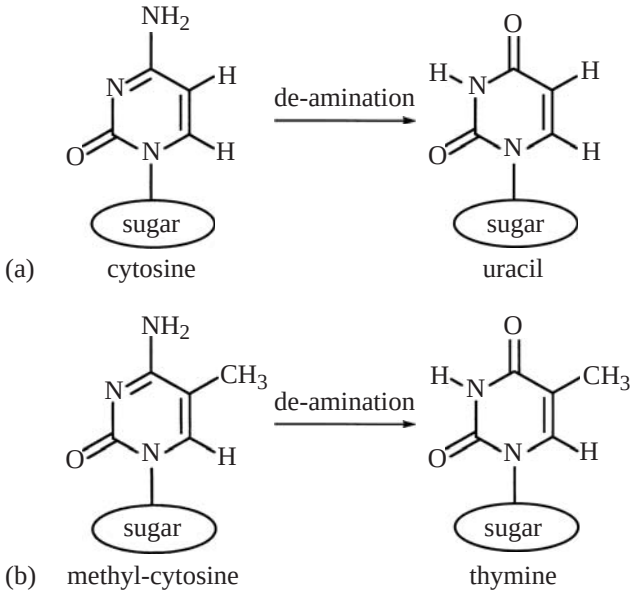
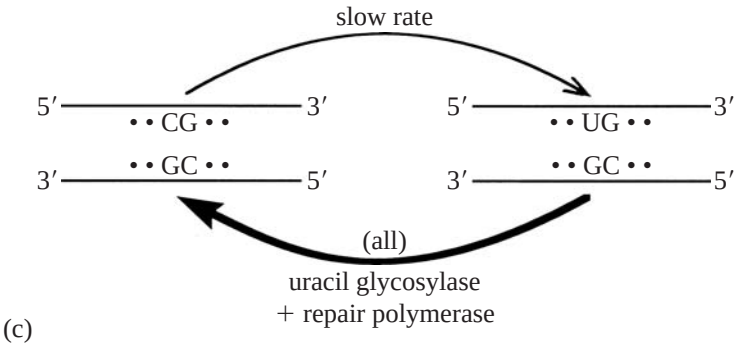


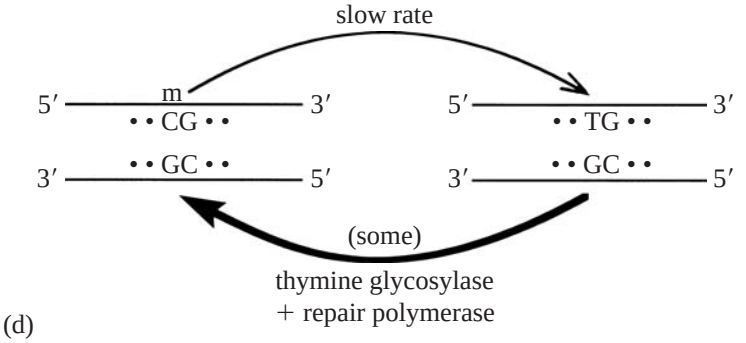
Figure 11.7 Two processes of chemical decay, followed by enzymatic repair, influence the di-nucleotide step composition of animal and plant genomes. In (a) we see that cytosine de-aminates continually to uracil at a tiny but finite rate. In (b) we see similarly that 5-methyl-cytosine de-aminates continually but slowly to thymine. In (c) we see that a change of step CG to UG may be fixed by a repair enzyme called 'uracil glycosylase', that seeks any aberrant uracil U bases (which are not normally present in DNA) and cuts them out entirely. A repair polymerase then restores the missing base. But (d) changes of the kind m-CG step to TG cannot be fixed so efficiently, since thymine T is a normal DNA base. Another repair enzyme 'thymine glycosylase' removes certain m-CG to TG step changes, based on recognition of a G–T base-pair mismatch between the mutated TG sequence and its partner CG on the other strand; yet this latter process is not entirely efficient, and so some m-CG to TG step changes will accumulate, despite repair. The overall result will be to deplete CG sequences from animal and plant genomes, wherever cytosine bases might typically be methylated.

All of a sudden this discussion has taken an unpleasant turn, because cytosine methylation seems to make us old and sick, without any possible cure! Why should such a hazardous system have evolved? If the fly or worm can do without cytosine methylation, why cannot we? When we study this question more closely, we can see that the biochemical system of DNA methylation, when combined with a decay of methyl-cytosine to thymine, actually serves as an excellent defence against foreign invading DNA; which was probably essential during our long evolutionary history.

For example, the human race (or ancestors thereof) have been invaded roughly 10 000 times during the past fifty million years,



(c)



(d)

with foreign DNA elements known as 'mariner transposons', that inhabit specialized, single-cell micro-organisms known as 'trypanosomes', which in turn live inside insects such as the tsetse fly. Now when an infected tsetse fly bites a human, it will transfer some of its trypanosomes into the human bloodstream, where they can multiply and cause illnesses such as 'sleeping sickness' and 'Chagas disease'. Once in the blood, those trypanosomes can also infect sex cells (usually sperm) with their own mobile DNA (or fly DNA), and thereby transfer mariner transposons into the human germline, and consequently into subsequent generations!

How could our ancestors have defended themselves against such ancient DNA invasions, which now make up an astonishing 45% of the human genome? They defended themselves largely by means of a methylation defence system, which works as follows. Since most invading DNA elements remain high in CG dinucleotide content, because they come from micro-organisms that lack cytosine methylation, the cells of our ancestors could add methyl groups almost immediately to any unprotected CG within the foreign DNA, as shown in Fig. 11.8(a) and (b). Those cells could thereby silence any foreign genes, that might be responsible for inserting or 'jumping' foreign DNA within a chromosome, or else for making an RNA copy to help the foreign DNA to reproduce.

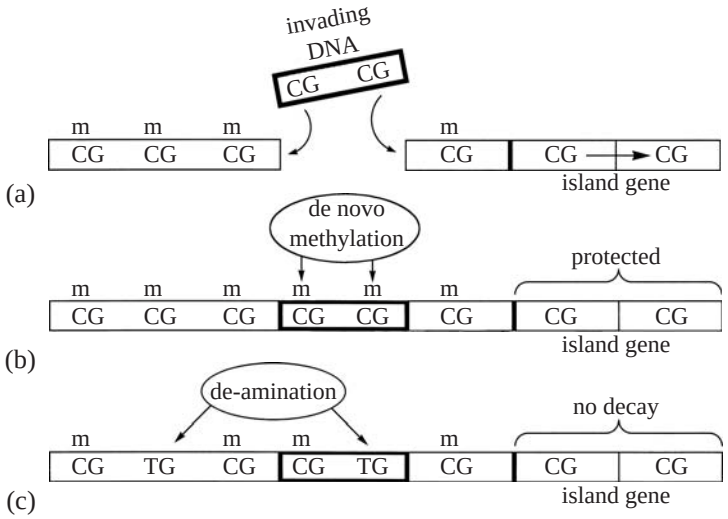


Figure 11.8 Cytosine methylation actually serves as a system of defence against invading foreign DNA, say from transposons or retrotransposons. In (a) we see that the invading DNA contains typically unmethylated CG sequences, especially if it comes from some simple animal such as a worm or fly. On the right, we see also a 'CG island' and its associated gene that remain typically unmethylated. Next in (b) we see that *de novo* methylases such as Dnmt3a or 3b will quickly add new methyl groups to the foreign DNA, whereas any nearby CG island will remain protected from methylation by bound proteins. Then in (c) we see that all methylated CG steps will undergo slow but continual deamination to TG, both in the animal genome and also in the foreign DNA; but not at any unmethylated CG island nearby. Hence foreign genes will be silenced immediately but reversibly by cellular methylases; and then will be silenced slowly but permanently by methyl-CG to TG mutations in their coding regions.

By this system, most CG sequences in the human chromosome would also become methylated and inactive. However, the human genome is designed so that it contains a large number of 'CG islands', where CG sequences remain protected from methylation by tightly-bound proteins which facilitate transcription. Almost all of those CG islands remain protected from methylation, so that they can specify active genes where needed: see Fig. 11.8(b).

Next, once any invading DNA has been silenced by CG methylation, all of its coding regions will begin slowly to decay, due to continual deamination of methyl-CG steps to TG without repair as shown in Fig. 11.8(c). Hence most foreign genes will become mutated and inactive within just a few generations. One can study in fact various foreign genes within the human genome, and estimate their original date of invasion by the ratio of CG-to-TG steps, just as for a radioactive half-life in physics.

Somewhat amazingly, DNA oligonucleotides which have been tested in humans for medical purposes (see Appendix 3) can sometimes induce an immune response in human blood, if they happen to contain one or more unmethylated CG step or steps. The human body seems to be designed so that it possesses several layers of defense against invading DNA from micro-organisms: whether by an immunological response at the whole-organism level; or by methylation response within individual cells.

In summary, we have learned here how cytosine DNA methylation works biochemically in humans and plants, through a variety of different enzymes, and through specific protein-to-protein interactions at a chromosomal level. Also, we have discussed its great importance in biological or medical phenomena such as embryonic development, aging or cancer. How then might specific patterns of DNA methylation be established in early embryos? And how might specific patterns of methylation be changed in a controlled fashion, so as to alleviate aging or cancer in adults?

No one knows today how new patterns of DNA methylation might be laid down in early embryos; we can only speculate over the possibilites. We know that the male pro-nucleus loses most of its methyl-cytosine (except at imprinted regions) just after fertilization; and that the female pro-nucleus loses some of its methyl-cytosine (except at imprinted regions) due to a limited number of cell divisions, before any methylase becomes active. But once expressed, the *de novo* enzymes Dnmt3a and Dnmt3b are somehow directed to add new methyl groups within a growing embryo, with great specificity across the entire genome, so as to create different tissues. Not all of the old methyl groups would be lost from the original sperm or egg DNA, especially within imprinted regions and on female-origin chromosomes. Perhaps those prior patterns form the initial loci of further *de novo* methylation?

Other epigenetic factors, say histone acetylation or methylation, might direct such *de novo* DNA methylation to certain regions of the genome but not others. Another contributing factor might be small RNA molecules that serve as 'guides' for *de novo* cytosine methylation in animals as well as plants. We know that small fragments of RNA can guide 2′-ribose methylation of ribosomal RNA, and can guide the degradation of messenger-RNA by nucleases (see Appendix 3). Also, long, repetitive RNA such as *Xist* or *Air* might induce *de novo* cytosine methylation all across the genome, as well as at the X-chromosome and imprinting sites.

Finally, how might one alter specific patterns of methylation for medical purposes, for example to alleviate aging or cancer? The de-methylating agent 5-aza-cytosine is too toxic to give to humans,

although a safer chemical known as 'valproate' shows some activity of that kind. Overall methylase activity may be reduced by using 'anti-sense' oligomers which degrade the messenger-RNA for Dnmt1; while methyl groups may perhaps be added to specific genes by using oligomers that contain 5-methyl-cytosine. Some workers have used anti-sense oligomers to degrade RNA molecules such as *Xist*, which control methylation imprinting, and that show their own anti-sense in living cells called '*Tsix*'. Most work here, however promising, still seems to be a long way from clinical applications.

To conclude: if as a student you have reached the end of this Chapter, and have also read the other ten Chapters in this book, you should have gained much knowledge concerning the structure of DNA and its role in biology. Scientists of the past were no less clever than the ones today, yet they had less technology by which to do good experiments. Also they perceived the world differently in each era: many of their well-established ideas turned out to be faulty! When contemplating your own future in science, therefore, be sure to keep an open mind. Also try to remember what Aldous Huxley wrote about scientific research in 1960: 'What we perceive depends on the conceptual lattice through which it has been filtered'.

Further Reading and Bibliography

Early hypotheses, recent reviews

Holliday, R. and Pugh, J.E. (1975) DNA modification mechanisms and gene activity during development. *Science* **187**, 226–32. An early article suggesting the importance of cytosine methylation for gene activity.

Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics Supplement* **33**, 245–54. A summary of the many factors and/or proteins which influence patterns of cytosine methylation.

Millar, D.S., Holliday, R., and Grigg, G.W. (2003) Five not four: history and significance of the fifth base. In *The Epigenome: Molecular Hide and Seek*, eds. S. Beck and A. Olek, Wiley Academic Press, New York, pp. 1–18. A historical review of 5-methyl-cytosine and its role in biology.

Riggs, A.D. (1975) X inactivation, differentiation and DNA methylation. *Cytogenetics and Cellular Genetics* **14**, 9–25. An early article suggesting the importance of cytosine methylation for X-chromosome inactivation.

Turner, B.M. (2001) *Chromatin and Gene Regulation*: *Molecular Mechanisms in Epigenetics* Blackwell Science, Oxford. A survey of chromatin as it relates to epigenetic phenomena such as DNA or histone methylation.

Chemical or enzymatic methods of methylation analysis

Fraga, M.F. and Esteller, M. (2002) DNA methylation: a profile of methods and applications. *Biotechniques* **33**, 632–49. A survey of methods used to analyze cytosine methylation at specific sites in chromosomal DNA.

Gonzalgo, M.L. and Jones, P.A. (2002) Quantitative methylation analysis using methylation-sensitive single-nucleotide primer extension. *Methods* **27**, 128–33. Extension of a short primer by DNA polymerase on bisulfite-treated DNA can determine whether any position contains 5-methyl-cytosine or else normal cytosine.

Grigg, G. and Clark, S. (1994) Sequencing 5-methyl-cytosine residues in genomic DNA. *Bioessays* **16**, 431–6. Invention of the bisulfite method for determining whether any base might be cytosine or 5-methyl-cytosine.

Li, J., Protopopov, A., Wang, F., Senchenko, V. *et al.* (2002) *Not* I subtraction and *Not* I-specific microarrays to detect copy number and methylation changes in whole genomes. *Proceedings of the National Academy of Sciences, USA* **99**, 10724–9. An enzymatic method for profiling normal *versus* cancerous cells, based on the sensitivity of *Not* I (GCGGCCGC) to methylation at CG steps.

Millar, D.S., Warnecke, P.M., Melki, J.R., and Clark, S.J. (2002) Methylation sequencing from limiting DNA: embryonic, fixed and microdissected cells. *Methods* **27**, 108–13. Procedures for bisulfite analysis of cytosine methylation from small quantities of DNA.

Cytosine methylation in whole genomes

Amoreira, C., Hindermann, W., and Grunau, C. (2003) An improved version of the DNA methylation database. *Nucleic Acids Research* **31**, 75–7. An Internet-based compilation of all data regarding cytosine methylation at **http://www.methdb.net.**

Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–13. A small fraction of human DNA consists of small CG-rich regions known as 'islands', which usually remain unmethylated and may contain the regulatory sequence for a nearby gene.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V. *et al.* (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase Dnmt3a. *Proceedings of the National Academy of Sciences, USA* **97**, 5237–42. DNA sequences other than CG may contain 5-methyl-cytosine in embryonic stem cells, where Dnmt3a adds methyl groups occasionally to CA and CT.

Tomso, D.J. and Bell, D.A. (2003) Sequence content at human SNPs: over-representation of the CG dinucleotide at polymorphic sites, but suppression of variation in CpG islands. *Journal of Molecular Biology* **327**, 303–8. CG dinucleotides are more polymorphic than others at general methylated sites in human DNA, due to chemical instability without repair caused by deamination; but are less polymorphic in 'CG islands', where cytosine bases remain unmethylated.

Cytosine methylation in cancer, embryonic development and cloning

Cezar, G.G., Bartolomei, M.S., Forsberg, E.J., First, N.L. *et al.* (2003) Genome-wide epigenetic alterations in cloned bovine fetuses. *Biology of Reproduction* **68**, 1009–14. Cattle embryos which are prepared by nuclear transfer show reduced levels of cytosine methylation, so as to cause abortion.

Clark, S.J. and Melki, J.M. (2002) DNA methylation and gene silencing in cancer: which is the guilty party? *Oncogene* **21**, 5380–7. Gene silencing prior to cytosine methylation may be the cause of human cancers.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L. *et al.* (2003) Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489–92. Transgenic mice with low but non-lethal amounts of Dnmt1 acquire chromosomal abnormalities and cancer at an early age.

Hardeland, U., Bentele, M., Lettieri, T., Steinancher, R. *et al.* (2001) Thymine DNA glycosylase. *Progress in Nucleic Acid Research and Molecular Biology* **68**, 235–53. Thymine glycosylase repairs mutations (methyl-CG to TG), along with uracil glycosylase (CG to UG).

Pfeifer, G.P. and Denissenko, M.F. (1998) Formation and repair of DNA lesions in the p53 gene: relation to cancer mutations? *Environmental and Molecular Mutagenesis* **31**, 197–205. Some of the mutations that cause cancer are biochemical (methyl-CG to TG), whereas others are environmental (sunlight, carcinogens).

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental Biology* **241**, 172–82. Loss of methylation in the male pro-nuclear DNA proceeds within six hours after fertilization, perhaps because the de-methylase is sperm-specific; whereas the female pro-nuclear DNA remains methylated initially, followed by passive loss of methylation due to cell division; finally the methylases Dnmt1, Dnmt3a and Dnmt3b become active at a later stage.

Pharmaceutical strategies involving cytosine methylation

Goffin, J. and Eisenhauer, E. (2002) DNA methyltransferase inhibitors – state of the art. *Annals of Oncology* **13**, 1699–716. 5-azacytidine and 5-aza-2′deoxycytidine which inhibit methylases, as well as an antisense oligonucleotide which binds to the messenger-RNA for Dnmt1, are being tested as anti-cancer drugs.

Holliday, R. and Ho, T. (2002) DNA methylation and epigenetic inheritance. *Methods* **27**, 179–83. Certain genes may be activated using 5-aza-cytidine (a methylase inhibitor), or else may be silenced using 5-methyl-cytosine triphosphate (when incorporated during replication).

Krieg, A.M. (1999) Mechanisms and applications of immune stimulatory CpG oligonucleotides. *Biochimica et Biophysica Acta* **1489**, 107–16.

Certain DNA oligonucleotides will produce an immune response in humans, if they contain unmethylated CG sequences that resemble DNA of microbial origin.

Yao, X., Hu, J.F., Daniels, M., Shiran, H. *et al.* (2003) A methylated oligonucleotide inhibits IGF2 expression and enhances survival in a model of hepatocellular carcinoma. *Journal of Clinical Investigation* **111**, 265–73. A DNA oligonucleotide containing 5-methyl-cytosine may induce altered DNA methylation at specific sites in cells.

Enzymes or proteins which regulate cytosine methylation

Cervoni, N. and Szyf, M. (2001) Demethylase activity is directed by histone acetylation. *Journal of Biological Chemistry* **276**, 40778–87. Increased acetylation of histones leads to active de-methylation of nearby DNA, although the enzyme responsible has not been characterized.

Chedin, F., Lieber, M.R., and Hsieh, C.-L. (2002) The DNA methyltransferase-like protein Dnmt3L stimulates *de novo* methylation by Dnmt3a. *Proceedings of the National Academy of Sciences, USA* **99**, 16916–21. The protein Dnmt3L, which is responsible for methylation 'imprints' in the female genome, stimulates *de novo* cytosine methylation by Dnmt3a but not by Dnmt3b.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X. *et al.* (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *Journal of Biological Chemistry* **278**, 4035–40. The mammalian protein MeCP2, which binds specifically to CG-methylated DNA, also binds specifically (perhaps through Dnmt1) to histone de-acetylases and histone methylases.

Huang, N., Banavali, N.K., and MacKerell, A.D. Jr. (2003) Protein-facilitated base-flipping in DNA by cytosine-5-methyl-transferase. *Proceedings of the National Academy of Sciences, USA* **100**, 68–73. A cytosine base may flip out from a double helix to become methylated by M.*Hha* I from bacteria, which resembles Dnmt2 from humans.

Kimura, H. and Shiota, K. (2003) Methyl-CG-binding protein MeCP2 is a target for the maintenance DNA methyltransferase Dnmt1. *Journal of Biological Chemistry* **278**, 4806–12. The mammalian protein MeCP2, which binds specifically to CG-methylated DNA, also binds specifically to Dmnt1.

DNA methylation in plants, fungi, flies and bacteria

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002) RNA-directed DNA methylation in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **99**, 16499–506. Short double-stranded RNA of 21 to 24 nucleotides in plants can degrade messenger-RNA in the cytoplasm, or induce *de novo* methylation of DNA in the

nucleus: every cytosine in the RNA-DNA is methylated by an unknown enzyme.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing. *EMBO Journal* **21**, 4671–9. Short RNA of 21 nucleotides guides degradation of messenger-RNA in the cytoplasm, while short RNA of 25 nucleotides guides cytosine methylation of DNA in the nucleus.

Kato, M., Miura, A., Bender, J., Jacobsen, S.E. *et al.* (2003) Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Current Biology* **13**, 421–6. Two maintenance methylases MET1 for CG or CMT3 for CNG act to suppress 'jumping' of DNA transposons in the plant *Arabidopsis thaliana*.

Tamaru, H. and Selker, E.U. (2003) Synthesis of signals for *de novo* DNA methylation in *Neurospora crassa*. *Molecular and Cellular Biology* **23**, 2379–94. Repeat sequences composed of the easily unwound $(TAAA)_n$ or $(TTAA)_n$ induce *de novo* methylation at nearby CG sequences in a bread mould.

Tamaru, H., Zhang, X., McMillen, D., Singh, P.B. *et al.* (2003) Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nature Genetics* **34**, 75–9. Adding methyl groups to lysine 9 of histone H3 is associated with chromatin condensation (in yeast or flies) or cytosine methylation (in mould, mammals or plants).

Urig, S., Gowher, H., Hermann, A., Beck, C. *et al.* (2002) The *Escherichia coli dam* DNA methyltransferase modifies DNA in a highly processive reaction. *Journal of Molecular Biology* **319**, 1085–96. N6-adenine methylase from bacteria controls replication, gene expression and mismatch-repair by modifying a sequence GATC.

Postscript

Our book is a small one, but the investigation of DNA is clearly a large subject, and potentially very large indeed. We have included eleven chapters on what we consider to be the most fundamental and well-understood aspects of DNA research; but we have omitted several potential chapters that might have dealt with other, less well-understood subjects such as: (a) the chemical theory of base-stacking in the context of DNA bound to antibiotics or proteins; (b) 'recombination' of the DNA molecule by various enzymes which cross-over and then re-connect its two strands; (c) the possible use of DNA for constructing tiny molecular machines or novel computational devices, generally known as 'nanotechnology'; and (d) the impact of genomic sequencing on our understanding of molecular evolution, and the apparent historical relationships among many different forms of life on Earth.

The chemical theory of base-stacking (a) in free DNA is discussed briefly in Appendix 2, with special reference to the recent work of Chris Hunter. We anticipate that other aspects of DNA chemical theory will soon be better understood, particularly in the context of its specific interactions with proteins, since new protein-DNA X-ray structures are now appearing almost weekly.

With regard to (b) recombination, it should be emphasized that the long-range ordering of base-pairs along any chromosomal DNA molecule need not always be rigidly fixed, but can change somewhat due to various enzymatic processes, as the cell grows and divides. Such *recombinatorial* changes may occur even before the fertilization of an egg cell by a sperm, when homologous pairs of chromosomes join together to make the precursor to a sex cell. Recombination of a type called 'V-D-J' is also very important for generating immunological diversity in adult human cells, so as to achieve T-cell or antibody recognition of foreign particles. (Changes of base