Corsaro, D., Valassina, M., Venditti, D., Venard, V., Le Faou, A., and Valensin, P.E. (1999) Multiplex PCR for rapid and differential diagnosis of *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* in respiratory infections. *Diagnosis in Microbiology and Infectious Disease* **35**, 105–8. The use of PCR to detect infectious organisms as a cause of asthma.

Dean, F.B., Hosono, S., Fang, L., Wu, X. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences, USA* **99**, 5261–6. The isothermal method MDA for amplifying DNA which can be done overnight at 30°C.

Gold, B., Bergeron, J., Lachtermacher-Triunfol, M., and Dean, M. (2001) Human duplex sex determination PCR. *Biotechniques* **31**, 28–35. The use of PCR to identify male *versus* female DNA samples using X or Y chromosomes.

Koetz, K., Bryl, E., Spickschen, K., O'Fallon, M., Goronzy, J.J., and Weyand, C.M. (2000) T-cell homeostasis in patients with rheumatoid arthritis. *Proceedings of the National Academy of Sciences, USA* **97**, 9203–8. The use of PCR to monitor dysfunction of the immune system in rheumatoid arthritis.

Nuovo, G.J. (2000) *In situ* strand displacement amplification: an improved technique for the detection of low-copy nucleic acids. *Diagnostic Molecular Pathology* **9**, 195–202. Tiny amounts of DNA may be amplified at 55°C by an isothermal method known as SDA.

DNA sequencing and array technologies

Cantor, C.R. and Smith, C.L. (1999) *Genomics: the science and technology behind the human genome project*. John Wiley and Sons, New York. A summary of methods used for the sequence analysis of entire genomes.

Elkin, C.J., Richardson, P.M., Fourcade, H.M., Hammon, N.M. *et al.* (2001) High-throughput plasmid purification for capillary sequencing. *Genome Research* **11**, 1269–74. Methods for sequencing 18 million bases of DNA per day at one facility in the USA.

Gupta, V., Cherkassky, A., Chatis, P., Joseph, R. *et al.* (2003) Directly labeled mRNA produces highly precise and unbiased differential gene expression data. *Nucleic Acids Research* **31**, e13. Messenger-RNA for microarray applications may be labeled for detection either by enzymes or chemicals.

Petricoin, E.F., Hackett, J.L., Lesko, L.J., Puri, R.K. *et al.* (2002) Medical applications of microarray technologies: a regulatory science perspective. *Nature Genetics Supplement* **32**, 474–9. A summary of DNA micro-array technology.

Rosenblum, B.B., Lee, L.G., Spurgeon, S.L., Khan, S.H. *et al.* (1997) New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Research* **25**, 4500–4. Synthesis and testing of di-deoxynucleotides to which fluorescent dyes have been attached for sequencing purposes.

Single-nucleotide-polymorphisms

Jordan, B., Charest, A., Dowd, J.F., Blumenstiel, J.P. *et al.* (2002) Genome complexity reduction for SNP genotyping analysis. *Proceedings of the National Academy of Sciences, USA* **99**, 2942–7. The use of semi-random PCR to analyse single-nucleotide-polymorphisms (SNPs).

McClay, J.L., Sugden, K., Koch, H.G., Higuchi, S., and Craig, I.W. (2002) High-throughout SNP genotyping by fluorescent competitive allele-specific polymerase chain reaction. *Analytical Biochemistry* **301**, 200–6. Inexpensive analysis of SNPs using two PCR primers of different sequence and colour.

Pusch, W., Wurmbach, J.H., Thiele, H., and Kostrzewa, M. (2002) MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* **3**, 537–48. A summary of time-of-flight mass spectroscopy as applied to detection of DNA single-nucleotide-polymorphisms.

Shastry, B.S. (2002) SNP alleles in human disease and evolution. *Journal of Human Genetics* **47**, 561–6. A review of the latest progress on SNPs, and their implications for medical diagnosis.

Shi, M.M. (2002) Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes. *American Journal of Pharmacogenomics* **2**, 197–205. The latest techniques used to measure slight genetic variation in human patients.

Triplet expansions and microsatellites

Chang, Y.C., Ho, C.L., Chen, H.H., Chang, T.T. *et al.* (2002) Molecular diagnosis of primary liver cancer by microsatellite DNA analysis in the serum. *British Journal of Cancer* **87**, 1449–53. Over 100 different microsatellites were screened to diagnose liver cancer from tumor DNA in blood.

Girardet, A., Hamamah, S., Anahory, T., Dechaud, H. *et al.* (2003) First preimplantation genetic diagnosis of hereditary retinoblastoma using informative microsatellite markers. *Molecular and Human Reproduction* **9**, 111–6. Microsatellites may be used to avoid disease when couples choose to have a child by *in vitro* fertilization.

Heidenfelder, B.L., Makhov, A.M., and Topal, M.D. (2003) Hairpin formation in Friedreich's ataxia triplet repeat expansion. *Journal of Biological Chemistry* **278**, 2425–31. Even GAA and TTC repeats form stable hairpins, presumably by non-classical G–A or T–C base-pairs.

Jin, P. and Warren, S.T. (2003) New insights into fragile-X syndrome: from molecules to neurobehaviors. *Trends in Biochemical Science* **28**, 152–8. A general review of fragile-X syndrome and its biochemistry.

Sinden, R.R., Potaman, V.N., Oussatcheva, E.A., Pearson, C.E. *et al.* (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *Journal of Bioscience* **27** (supplement), 53–65. An overview of triplet expansion in terms of DNA structural dynamics.

Sutherland, G.R. and Baker, E. (2000) The clinical significance of fragile sites on human chromosomes. *Clinical Genetics* **58**, 157–161. A clinical review of fragile-X syndrome and other fragile chromosomal sites.

DNA imprinting and regulatory RNA

Beletskii, A., Hong, Y.-K., Pehrson, J. Egholm, M., and Strauss, W.M. (2001) PNA interference mapping demonstrates functional domains in the noncoding RNA *Xist*. *Proceedings of the National Academy of Sciences, USA* **98**, 9215–20. The repetitive "C" domain of *Xist* RNA mediates its binding to the X-chromosome.

Jouvenot, Y., Ginjala, V., Zhang, L., Liu, P.Q. *et al.* (2003) Targeted regulation of imprinted genes by synthetic zinc-finger transcription factors. *Gene Therapy* **10**, 513–22. A zinc-finger protein can carry a transcription activator to silence genes in cultured cells, and thereby reverse repression by imprinting.

Lee, J.T. (2003) Molecular links between X-inactivation and autosomal imprinting. *Current Biology* **13**, R242–54. The latest molecular and genetic data on imprinting of sex and non-sex chromosomes.

Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Reports* **2**, 986–991. A survey of non-coding RNA molecules and their possible biological functions.

Polychronakos, C. and Kukuvitis, A. (2002) Parental genomic imprinting in endocrinopathies. *European Journal of Endocrinology* **147**, 561–9. A summary of human genetic diseases caused by faulty imprinting.

Rougeulle, C. and Heard, E. (2002) Antisense RNA in imprinting: spreading silence through *Air*. *Trends in Genetics* **18**, 434–7. A long non-coding RNA called *Air* controls imprinting of a gene cluster on a non-sex chromosome.

Szittya, G., Silhavy, D., Molnar, A., Havelda, Z. *et al.* (2003) Low temperature inhibits RNA silencing-mediated defence by the control of *si*-RNA generation. *EMBO Journal* **22**, 633–40. Small, interfering RNA molecules of size 21 to 26 bases can protect plants against viral infection.

Viral and non-viral gene therapy

Hirata, R.K. and Russell, D.W. (2000) Design and packaging of adeno-associated virus gene-targeting vectors. *Journal of Virology* **74**, 4612–20. Gene correction at frequencies of 1% in cultured cells, using AAV to deliver single-stranded DNA into the cell nucleus.

Hirata, R., Chamberlain, J., Dong, R., and Russell, D.W. (2002) Targeted transgene insertion into human chromosomes by adeno-associated viral vectors. *Nature Biotechnology* **20**, 735–8. Gene addition at frequencies of 1% in cultured cells, using AAV to deliver single-stranded DNA into the cell nucleus.

Liu, L., Rice, M.C., and Kmiec, E.B. (2001) *In vivo* gene repair of point and frameshift mutations directed by chimeric RNA/DNA oligonucleotides and modified single-stranded nucleotides. *Nucleic Acids Research* **29**, 4238–50. The use of electroporated oligonucleotides to repair a variety of genetic mutations in yeast.

Niidome, T. and Huang, L. (2002) Gene therapy progress and prospects: nonviral vectors. *Gene Therapy* **9**, 1647–52. Physical or chemical methods which may be used to add new DNA to human cells or patients.

Recchia, A., Parks, R.J., Lamartina, S., Toniatti, C. *et al.* (1999) Site-specific integration by a hybrid adenovirus/adeno-associated virus vector. *Proceedings of the National Academy of Sciences, USA* **96**, 2615–20. Two adenoviruses, one carrying AAV and the other expressing *Rep78* protein, can add a new gene to a specific site on human chromosome 19.

Shayakhmetov, D.M., Carlson, C.A., Stecher, H., Li, Q., Stamatoyannopoulos, G., and Lieber, A. (2002) A high-capacity hybrid adenovirus/adeno-associated virus for stable transduction of human hematopoietic cells. *Journal of Virology* **76**, 1135–43. A mixed adenovirus/AAV vector can add a large globin gene permanently within cells.

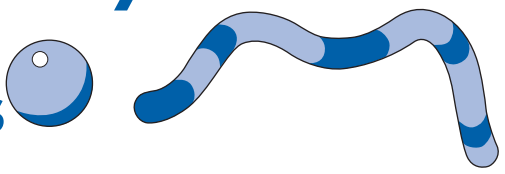
Thorpe, P., Stevenson, B.J., and Porteous, D.J. (2002) Optimising gene repair strategies in cell culture. *Gene Therapy* **9**, 700–2. The use of lipofected single-stranded DNA to repair a point mutation in cultured cells.

Web-based Resources

Introduction to polymorphisms, and data-bases for SNPs
http://www-hto.usc.edu/~cbmp/2001/SNP

CHAPTER 11

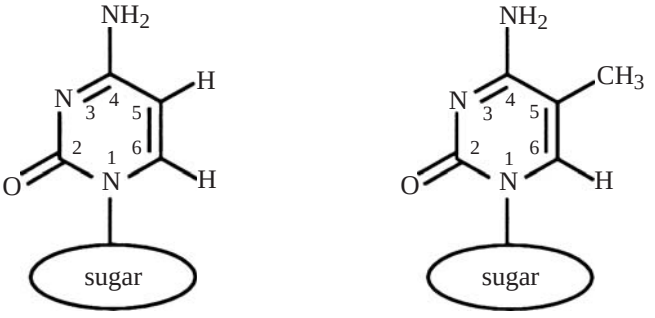
Cytosine Methylation and DNA Epigenetics



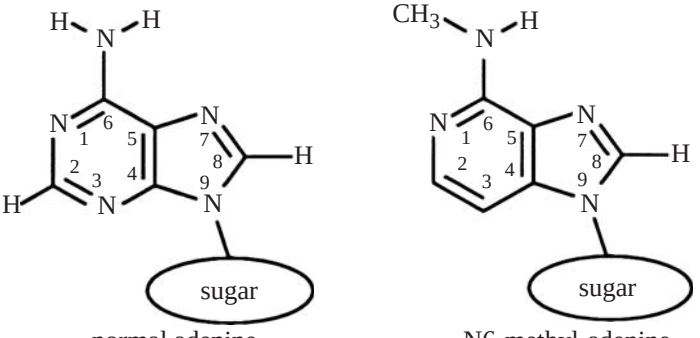
In Chapter 10 we discussed many aspects of 'normal genetics' and its relation to modern medicine. Normal genetics comes about mainly from the pairing of two bases across any double helix, as A with T or G with C in a Watson–Crick fashion. We also pointed out that those same DNA bases may become modified reversibly to different chemical forms by cellular enzymes known as 'methylases'. Once we add DNA methylation to our list of unexpected phenomena which have been observed in biology, we enter another realm of science known as 'epi-genetics', where 'epi' means here 'outside of' or 'in addition to'.

Thus, epigenetics describes an unexpected extension of classical 20th-century genetics, where information may be stored not only within the primary sequence of DNA, i.e. A, T, C or G; but also within small chemical changes to that primary sequence, for example by adding a carbon atom and its associated hydrogens to any base ring. For all animals and plants on Earth, it turns out that methylation of cytosine C bases will provide the key to understanding how most epigenetic effects work: for example the sex-cell-specific 'imprinting' of genetic inheritance between different generations; or the low incidence of CG dinucleotides in any genome from animals or plants; or a subtle means of defence whereby animals or plants can deal with the unwanted invasion of foreign DNA. But before we introduce such advanced topics, we first need to describe the 'nuts and bolts' of DNA methylation, and something of the brief history of the subject.

The most common forms of methylated DNA are shown in Fig. 11.1(a) and (b). Cytosine may be modified reversibly by adding a methyl group CH_3 to its carbon 5-position; while adenine may be



(a) normal cytosine 5-methyl-cytosine



(b) normal adenine N6-methyl-adenine

	most animals	worm, fly	plants	bacteria
5-methyl cytosine	CG	–	CG, CNG	CC A/T GG and others
N6-methyl adenine	–	–	–	GATC and others

(c) major sites of DNA methylation

Figure 11.1 Two common forms of methylated DNA bases. In (a) we show the structures of normal cytosine and 5-methyl-cytosine. In (b) we show the structures of normal adenine and N6-methyl-adenine, the latter of which is found in *E. coli* or other bacteria at sequences *dam* Gm-ATC; *M. Eco R*I GAm-ATTC or *M. Taq* I TCGm-A. The table in (c) gives a simplified summary of the major sites of DNA methylation in different, representative organisms.

modified by adding a methyl group CH_3 to its nitrogen 6-position. Other modifications, such as 5-hydroxy-methyl-cytosine (add CH_2-OH) or 8-hydroxy-guanine (add OH), are found only in bacterial viruses, or in oxidatively damaged DNA.

These methylated bases are usually found at specific sequences in DNA, and not just at any general location. For example, as shown

in Fig. 11.1(c), 5-methyl-cytosine is found mainly at the dinucleotide CG for animal DNA, or at the two short sequences CG or CNG for plant DNA, where N can be any base. (The CG dinucleotide is often referred to in the literature as 'CpG', where p stands for the phosphate group; see Chapter 1). Here a special enzyme methylates nearby cytosine bases on both strands. We do not see cytosine methylation at other sequences (that is, CC, CA or CT) except in special cases, where slight amounts of methyl-CA or methyl-CT may be found in fly embryos, or in mammalian embryonic stem cells. The adenine base may also be methylated, as shown in Fig. 11.1(b). N6-methyl-adenine is found usually at the tetranucleotide GATC in many bacteria including *E. coli*, as well as in other minor locations: it is important for bacterial biochemistry, but we shall not discuss adenine methylation further here. Finally, certain bacteria and viruses often have special enzymes that add 5-methyl-cytosine at very precise locations, for example at CCAGG or CCTGG for the *dcm* methylase of *E. coli*; but those enzymes are not found in animals or plants.

Now it is not only the bases in DNA which may be chemically modified, since some of the amino acids in proteins may be modified reversibly as well, by enzymes known as 'protein acetylases' or 'protein methylases'. Such modifications typically involve a change of normal lysine to N-acetyl-lysine (add an acetyl CO-CH_3), or else to di-methyl or tri-methyl-lysine (add two or three methyl CH_3), within the abundant histone proteins of eukaryotic chromatin. Those histone proteins wrap long pieces of DNA around themselves into compact ball-like structures known as nucleosomes, as described in Chapter 7. Hence a change of lysine chemical structure within histones might perhaps influence the tightness of DNA wrapping around these proteins; and it might perhaps affect the entire structure of the chromatin, in particular chromosomal locations, and thereby influence the activities of specific genes. We shall return to that subject below, in a discussion of DNA methylases and their interactions with enzymes that modify histones.

Biologists had been aware of the existence of 5-methyl-cytosine within DNA for nearly fifty years; but such slight variations of DNA content were not regarded as significant. One reason for dismissing the potential significance of DNA methylation in biology was that simple organisms such as worm, fly or yeast contain little or no 5-methyl-cytosine. Nevertheless those organisms eat, move, reproduce and thrive. What is good enough for the worm must be good enough for humans!

The inadequacy of such a simple-minded philosophy did not become clear until 1975, when two workers, Art Riggs and Robin

Holliday, suggested independently that 5-methyl-cytosine could play an important role in the normal biology of humans or other mammals. Riggs suggested that 5-methyl-cytosine would prove important for X-chromosome inactivation in the early embryo – where only one female X-chromosome remains active, while the other becomes repressed, as mentioned in Chapter 10. Holliday suggested that 5-methyl-cytosine would prove important for regulating gene expression in the development of a complex organism.

Back then there was hardly any interest in the field, especially when compared with the explosion of interest that we have seen recently. For example, Riggs would visit the laboratory of one of the authors of this book (H.R.D.) in the late 1970's, every Wednesday for lunch; and one day the following conversation ensued. 'Should I work on a crystal structure of DNA with 5-methyl-cytosine?' asked the young student. 'Not many people would be interested, so it might hurt your career. But it would interest me!' – Art Riggs, 1978.

In fact, research before 1990 in the field of cytosine methylation was often not particularly incisive, because it was limited to only a few simple techniques. Most commonly, total genomic DNA would be digested with either of two bacterial enzymes, *Hpa* II or *Msp* I. The enzyme *Hpa* II cuts only unmethylated CCGG, whereas the enzyme *Msp* I will cut both unmethylated and methylated CCGG equally; see Fig. 11.2(a) and its legend. Then the products of such digestion would be applied to an agarose gel, where they could be separated by size. Next, all DNA fragments from the gel would be transferred to a nylon filter, where they could be visualized by hybridizing to some 'radioactive probe' which would be made from some small part of the genome. (Such DNA filter-hybridization is known as a 'Southern blot', after Ed Southern who invented it.) When the pattern of radioactive bands was examined afterwards, scientists could determine which bases near some particular gene might contain normal cytosine or else 5-methyl-cytosine.

As a variant of that same technique, genomic DNA samples from various organisms could be digested with *Hpa* II or *Msp* I, and then applied, as above, to an agarose gel. Next, all of the many different DNA fragments could be studied in total by staining with ethidium bromide dye, rather than by using a sequence-specific probe. The unexpected finding from such work was that a small (1%) fraction of vertebrate or animal DNA seemed to contain many unmethylated CG sequences clustered together, over short regions of 500 to 1000 base-pairs: see Fig. 11.2(b). Those short regions were therefore called 'CG islands', and they were detected first by their sensitivities to the enzyme *Hpa* II, which cuts only unmethylated CCGG.

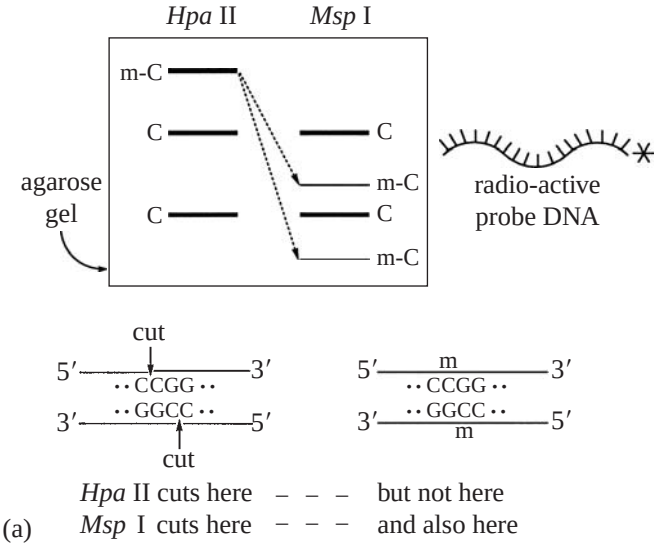
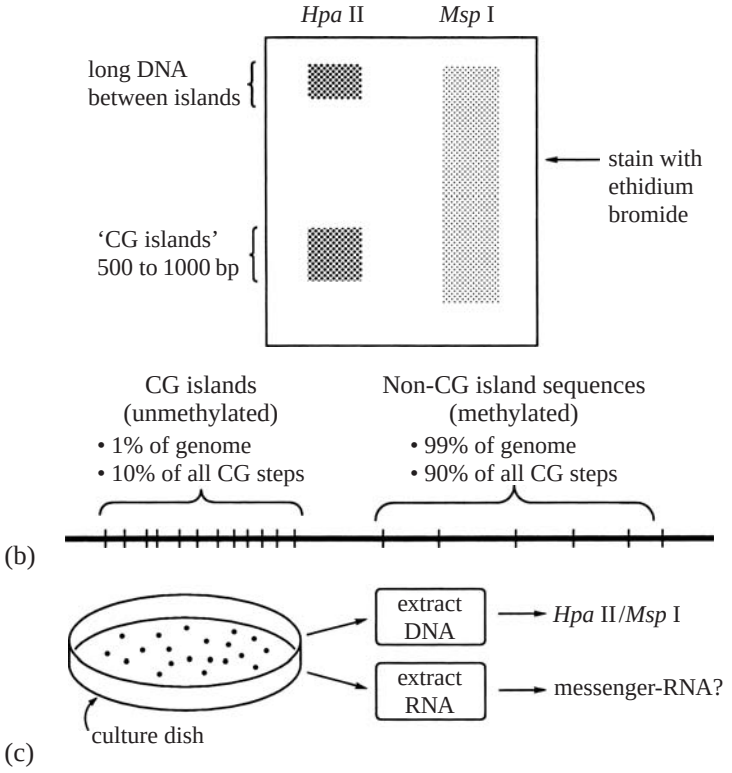


Figure 11.2 Three common kinds of experiment to study methylated DNA that were performed in the 1980s. In (a) we show an idealized agarose gel, which has been used to separate by size two different samples of genomic DNA; one of which has been digested with *Hpa* II, and the other with *Msp* I. After the gel has been run, it is treated with radioactive DNA of specific sequence which attaches itself to the few DNA fragments in the gel with complementary sequence. Here, three fragments have been picked out by the radioactive probe, and one can see that a large band 'm-C' on the left becomes split into two small bands 'm-C' on the right, because *Hpa* II cannot cut a methylated sequence Cm-CGG whereas *Msp* I can – as shown in the lower part of the diagram. In (b) we show the same kind of experiment using an agarose gel which has been stained with ethidium bromide, rather than hybridized to a specific radioactive probe. The genomic DNA contains many sequences CCGG, most of which are methylated, but some of which are not. The *Msp* I enzyme cuts all of them indiscriminately, breaking the DNA into a continuum of fragment sizes ranging from long (at the top of the gel) to short (at the bottom). In contrast, the *Hpa* II enzyme cuts only the un-methylated CCGG sequences . The sample on the left is therefore not cut into nearly so many pieces as the sample on the right; but the fuzzy series of bands on the lower left shows that some of the DNA fragments are short, indicating a tight clustering of the un-methylated cutting sites of HpaII into 'CG islands'. The lower diagram shows approximately the overall distribution of CG steps in the two parts of the genome. In (c) we show the treatment of human cells with a chemical 5-aza-cytosine, which inhibits enzymes that add methyl groups to cytosine, and thereby leads to loss of 5-methyl-cytosine across the entire genome; by studying the DNA extracted from such cells with *Hpa* II or *Msp* I, and the RNA extracted from such cells to look for messenger-RNA, scientists can determine whether loss of methylation might activate any particular gene.

The remaining (99%) of the genomic DNA had its CG sequences methylated; and so one could see a striking contrast in the gel for DNA digested with *Hpa* II: between the short pieces resulting from cutting *within* CG islands, *versus* the long uncut pieces of DNA



found *between* the islands. A striking contrast could also be seen between the high CG-dinucleotide content of the 'island' DNA – 10 times more than expected randomly – and a low CG-dinucleotide content of DNA found elsewhere in the same genome. That contrast will be explained below.

We know now that the DNA within any CG island often contains a promoter sequence for some nearby gene, and is usually protected from methylation by proteins that facilitate transcription; however, that was not known at the time. All of this shows that methylation of cytosine can indeed affect the activity of a gene – which may be related to cancer and other diseases, as we shall see below.

A third early technique was to treat cells in culture with a chemical known as '5-aza-cytosine', in order to *remove* methyl groups from many different cytosine bases across the whole genome: see Fig. 11.2(c). Then scientists could extract the total DNA from those cells, and digest it with *Hpa* II/*Msp* I to see whether any particular gene had lost its methyl groups. They could also extract total RNA, to see whether any particular gene had begun to make more messenger-RNA. The chemical 5-aza-cytosine is an analogue of normal