Lectures -5+6

# THE microbial genetics

**Dr. Yasir Alabdali**

# Mutation and Variation

**The cornerstone of bacterial genetics has been, until recently, the isolation of specific mutants, i.e. strains in which the gene concerned is altered (usually in a deleterious fashion). This alteration shows up as a change in the corresponding characteristics of the organism. It is this change in the observable properties of the organism (the phenotype) that is used to follow the transmission of the gene. The genetic nature of the organism (the genotype) is inferred from the observable characteristics.**

Any population of bacteria is far from homogeneous. A culture of E. Coli (under optimal conditions) will grow from a single cell to its maximum cell density (commonly about $10^9$cells $ml^{-1}$) in about 10–15 h, having passed through 30 generations. Within that culture, there will be some variation from one cell to another, which can be due to physiological effects or to genetic changes. Physiological variation means that, **due to differences in the environment, growth history or stage of growth at any one moment**, the cells may respond differently to some external influence. The key difference between physiological and genetic variation is whether the altered characteristic can be inherited.

# Types of mutations

## Point mutations

**Point mutation** is the sequence of the DNA has been altered at a single position. Where this change consists of replacing one nucleotide by another, it is known as a **base substitution**. The consequence of such a change depends both on the nature of the change and its location. If the change is within the coding region of a gene (i.e. the region which ultimately is translated into protein), it may cause an alteration of the amino acid sequence which may affect the function of the protein. The alteration may of course have little or no effect, either because the changed triplet still codes for the same amino acid or because the new amino acid is sufficiently similar to the original one for the function of the protein to remain unaffected.

**For example,** the triplet **UUA** codes for leucine; a single base change in the DNA can give rise to one of nine other codons as shown in Figure 2.3.

Two of the possible changes (**UUG, CUA**) are completely **silent,** as the resulting codons still code for **leucine**. These are known as synonymous codons.

Two further changes (**AUA and GUA)** may well have **little effect** on the protein since the substituted amino acids (**isoleucine and valine** respectively) are similar to the original leucine (they are all hydrophobic amino acids).

**Phenylalanine** (**UUU or UUC** codons) is also hydrophobic but is more likely to cause a **significant change** in the structure of the protein at that point.

The significance of the change to **UCA**, resulting in the substitution of **serine** (which is considerably different) for **leucine will depend on the role played by that amino acid** (and its neighbours) in the overall function or conformation of the protein.

The final two changes (**UAA, UGA**) are referred to as **stop** or termination codons (as is a third codon, UAG), since they result in termination of translation. The occurrence of such a mutation (also known as a **nonsense mutation**) will result in the **production of a truncated protein**; such a protein may or may not be functional, depending on the degree of shortening.

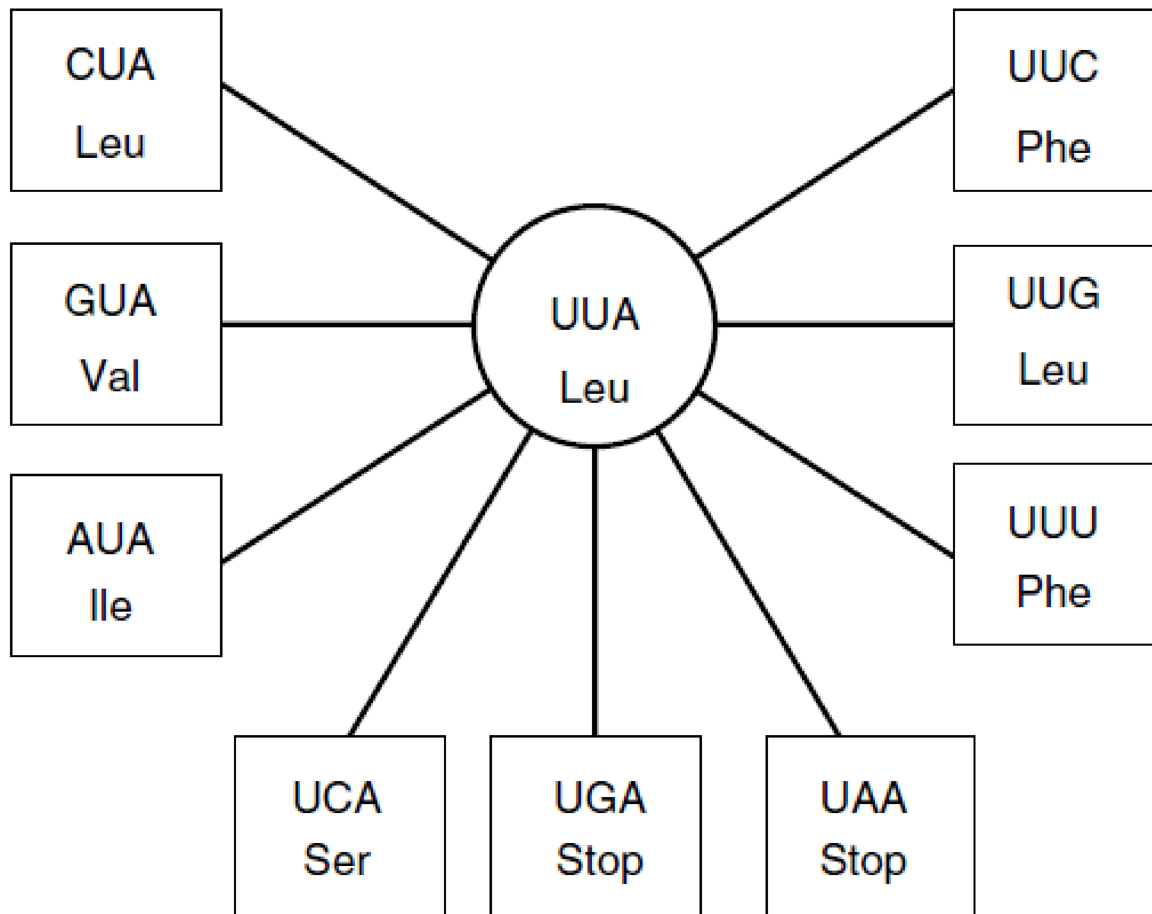


**Figure 2.3** Codons arising by single base substitutions from UUA

**Frameshift mutation**

The deletion or addition of a single nucleotide (or of any number other than a multiple of three). This is known as a **frameshift mutation,** since it results in the reading frame being altered for the remainder of the gene. Since the message is read in triplets, with no punctuation marks (the reading frame being determined solely by the translation start codon), an alteration in the reading frame will result in the synthesis of a totally different protein from that point on. Figure 2.4.

In fact, protein synthesis is likely to be terminated quite soon after the position of the deletion. For most genes the two alternative reading frames are blocked by termination codons, which serves to prevent the production of aberrant proteins by mistakes in translation.

(a) Original sequence

| AUG | CUA | GCU | AGC | UUA | CCU | AUU | CGA | UUC | UAC | CUG | AGC | U — — — |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Met | Leu | Ala | Ser | Leu | Pro | Ile | Arg | Phe | Tyr | Leu | Ser | — — — — |

(b) Sequence after deletion of a single base

Position of deletion

| AUG | CUA | CUA | GCU | UAC | CUA | UUC | GAU | UCU | ACC | UGA | GCU | — — — |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Met | Leu | Leu | Ala | Tyr | Leu | Phe | Asp | Ser | Thr | Stop | | |

Altered amino acid sequence

(c) Suppression of frameshift

Base added

| AUG | CUA | CUA | GUC | UUA | CCU | AUU | CGA | UUC | UAC | CUG | AGC | U — — — |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Met | Leu | Leu | Val | Leu | Pro | Ile | Arg | Phe | Tyr | Leu | Ser | — — — — |

Altered sequence

**Figure 2.4** Frameshift mutation and suppression. (a) Initial (mRNA) sequence and translated product. (b) Deletion of a single base alters the subsequent reading frame producing a different amino acid sequence and encountering a stop codon. (c) Addition of a base at a different position restores the original reading frame and may suppress the mutation

## 2.4 Restoration of phenotype

### *2.4.1 Reversion and suppression*

Since a point mutation arises by a change at a single point in the DNA, an event which can occur at random, it follows that there is a good chance of a second mutation occurring which will restore the original DNA sequence (or a sequence that is for practical purposes indistinguishable). In the example shown in Figure 2.3, a mutant strain with the UUU codon (phenylalanine) may undergo a further mutation which restores the UUA codon (a true *back mutation*) or one that substitutes UUG or CUU (both of which are also leucine codons). This strain will now have the same properties as the original one and is said to have *reverted*.

The effect of a mutation can also be negated by a second, unrelated mutation; this effect is known as *suppression*. This can take a wide variety of forms, most of which are specific to the particular gene involved and occur when alteration of a second gene can counteract the deleterious effects arising from the loss of the first gene function. There are two types of suppression that are of more general importance.

The first occurs with frameshift mutations. These may revert by the restoration of a deleted base for example, but can also be suppressed by the addition or deletion of further bases (not necessarily at the same place as the original mutation) so that the total number of bases added or lost is a multiple of three (see Figure 2.4). In this way the original reading frame is restored leaving a limited number of altered codons. Whether this altered product has sufficient biological function to result in observable suppression of the original mutation will of course depend on the size and nature of this altered sequence and its effect on the function of the protein.

Another important type of suppression occurs with 'nonsense' mutations, where a stop codon has been created within the coding sequence. These result in termination of translation largely because there is no corresponding tRNA to recognize them. However, tRNA molecules are themselves coded for by genes, which are of course susceptible to mutation. It is therefore possible for an existing tRNA gene to be changed in such a way that the tRNA it codes for will now recognize one of the stop codons rather than (or as well as) the codon it normally recognizes. For example, in Figure 2.5 the original mutation changes a glutamine codon (CAG) to a stop codon (UAG). Suppression of this mutation can occur by alteration of the glutamine tRNA gene so that its anticodon now pairs with the amber codon. Glutamine will therefore be inserted into the growing peptide chain and the final product will be identical with the wild-type protein. Since there is more than one glutamine tRNA gene, the cell does not lose the ability to recognize genuine glutamine codons.

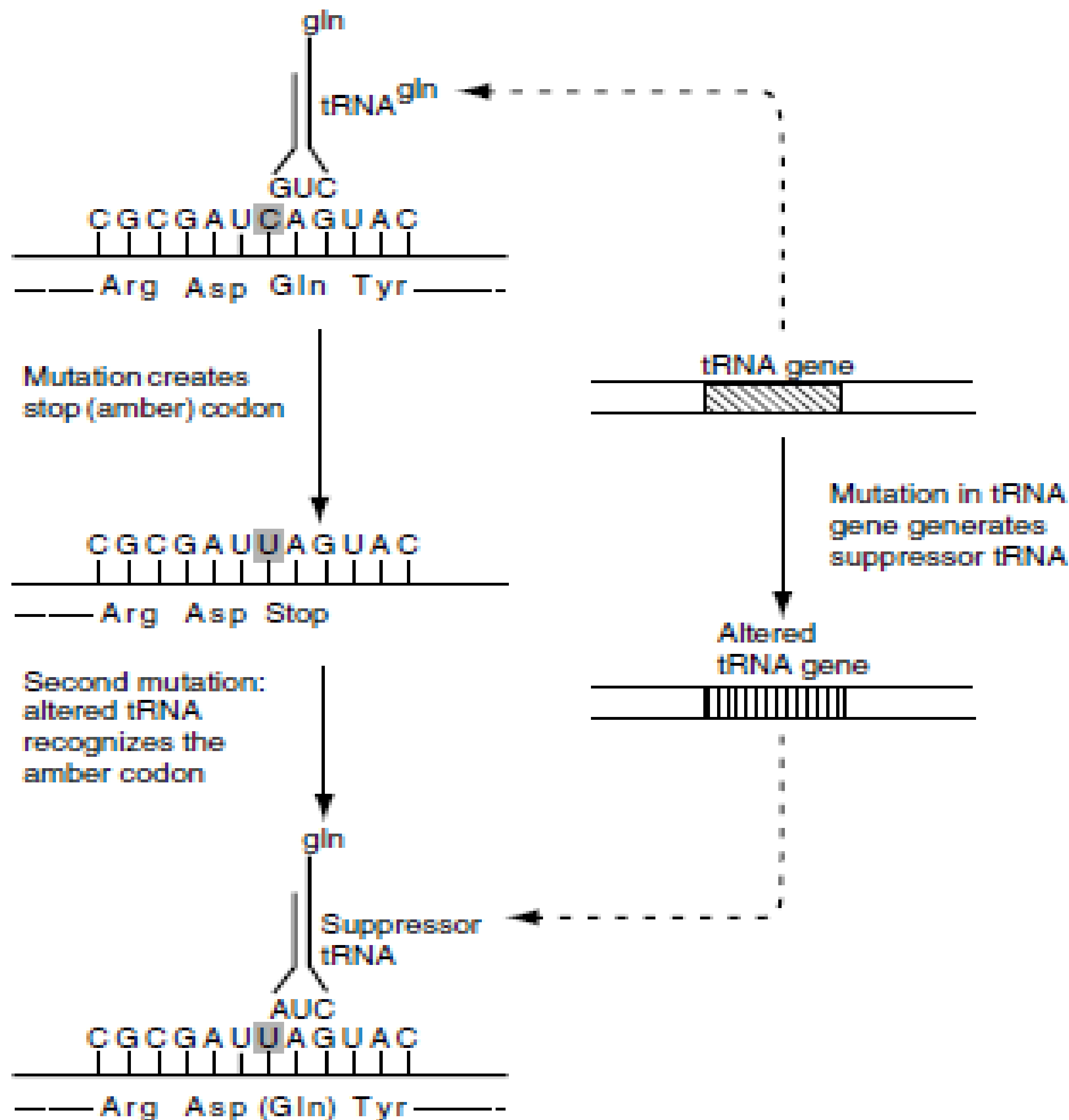


**Figure 2.5** Suppression of a nonsense mutation. A base substitution changes CAG to the stop codon UAG, causing premature termination of translation. This can be suppressed by a separate mutation in a tRNA gene, giving rise to a tRNA that can recognize the UAG codon

It might be expected that this suppressor tRNA would now prevent normal chain termination at the end of the translated region which would clearly be very damaging to the cell. However, translational termination is reinforced by the action of release factors which means that suppression is far from absolute (it may range from 10–50 per cent for an amber suppressor). In addition, at the genuine termination site, there are often multiple termination codons which will lead to efficient termination even in the presence of a suppressor tRNA.

The importance of this type of suppression is that the tRNA mutation is able to suppress any corresponding mutation, not just the original one it was selected for. Of course, each of the stop codons can arise by mutation of a number of different



codons (and there are also a number of possible suppressor tRNAs with different amino acid specificities) which means that in many cases the activity of the suppressor tRNA will result in the insertion of an amino acid other than the correct one. So, although a full-length product will be obtained, it may not be

### 2.4.2 Complementation

Another way in which a mutant phenotype can be converted back to the wild type is by acquisition of a plasmid that carries a functional version of the affected gene. For example, a $Lac^-$ strain of *E. coli* will become able to use lactose again following introduction of a plasmid carrying the relevant genes. In this case, we say that the plasmid has *complemented* the chromosomal defect. This only works if the functional version is effective in the presence of the mutated gene, i.e. the mutation is recessive.

Traditionally, genetics has followed the route described above: isolating variants on the basis of the altered phenotype and then attempting to identify the nature of the genetic change responsible. The advent of gene cloning technology, and especially of genome sequencing, has opened up a different route. Using these techniques, we commonly know the sequence of many, or all, of the genes of an organism, but without knowing their functions. If we have some idea of the possible function of a specific gene (for example by comparing its sequence with that of known genes from other sources), we can test this hypothesis by examining the ability of the cloned DNA fragment to complement a characterized mutant.

## 2.5 Recombination

Recombination, in the sense of re-assorting the observable characteristics in the progeny of a cross, has been a fundamental feature of genetics since long before its inception as a formal discipline. The term 'recombination' can be used in an analogous fashion in bacterial genetics, but is also used to refer to the physical breaking and joining of DNA molecules.

At the simplest level, we can consider two linear DNA molecules: breaking both molecules at a single point, crossing them over and rejoining them will produce two recombinant DNA molecules, both of which have a part of each of the parental molecules (Figure 2.6). This general concept applies to a variety of recombinational mechanisms, of which the principal one is known as *general* or *homologous* recombination; this requires a substantial degree of homology between the sequences to be recombined but will work with any two pieces of homologous DNA. In contrast, *site-specific* recombinational mechanisms require little or no homology, but (as the name implies) operate only within specific sequences. The RecA protein is required for homologous recombination, but not for site-specific processes. Recombination mechanisms are considered further in Chapter 6.

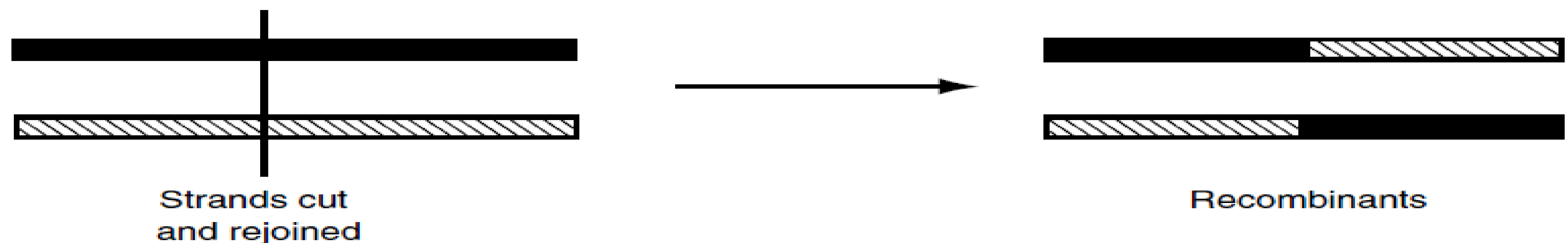


**Figure 2.6** Recombination between two linear DNA molecules

## 2.6 Mechanisms of mutation

### 2.6.1 Spontaneous mutation

Spontaneous mutation occurs through errors in the replication of DNA. However, the model presented for the structure and replication of DNA (Chapter 1) does not leave room for the existence of errors so it is necessary to consider how this may happen. Within the normal structure of the double helix of the DNA molecule, the only base-pairing combinations that are allowed are A-T and G-C; any other combinations would result in a distortion of the helix, and such distortions will be removed enzymically and repaired (see Chapter 1).

This is the basic dogma. However, it makes assumptions about the structure of the bases. If they exist only in the form in which they are usually drawn, then the statement is true. However, each of the bases can exist in alternative tautomeric forms with different hydrogen bonding capabilities. Two examples of tautomerism, the change from an amino to an imino form and a keto-enol tautomerism, are shown in Figure 2.7. (Other forms of tautomerism also exist, but will not be discussed here). The consequences for base pairing are illustrated in Figure 2.8. The normal pairing of adenine and thymine (Figure 2.8a) occurs with adenine in the amino form and the keto form of thymine. The imino form of adenine (Figure 2.8b) however, will base pair with cytosine rather than thymine, while the enol form of thymine will hydrogen bond to guanine (Figure 2.8c). These alternative forms of the bases are thermodynamically unfavoured; if considered as an equilibrium between the two states, then about 1 in $10^4 - 10^5$ molecules will be in the

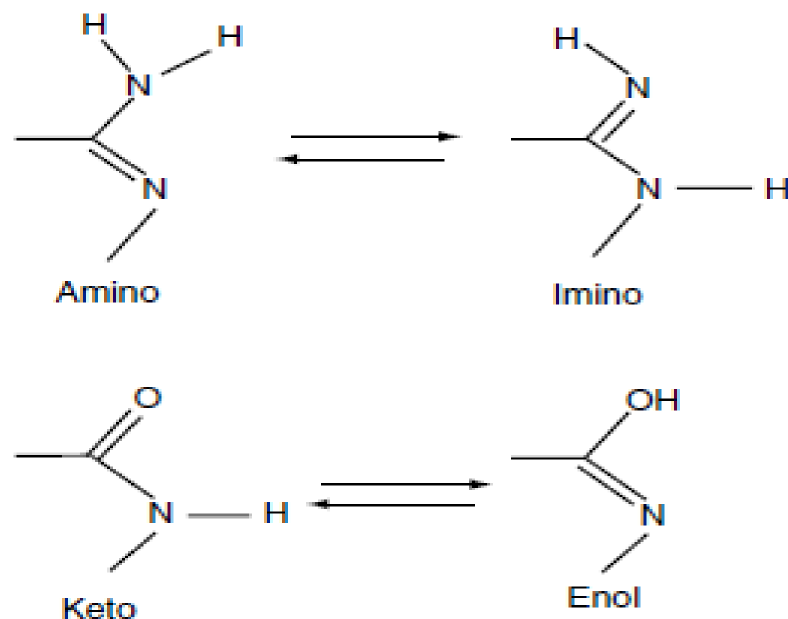


**Figure 2.7** Examples of tautomerism

alternative state at any one point in time. Incorporation of one of these into the DNA will result in a mutation at that point, which would thus be expected to occur with a frequency of the order of $10^{-4} - 10^{-5}$ per base per generation.

The problem with this explanation is that the observed frequency of mutation is in fact many orders of magnitude lower than this (with *E. coli* at least). The discrepancy is accounted for by the proof-reading activity of the DNA replication machinery, arising from the 3′–5′ exonuclease function of the DNA polymerase, as described in Chapter 1, which enables it to remove incorrectly paired bases at the 3′ end of the growing strand. The abnormal tautomer can be considered as reverting rapidly to the more energetically favoured form, in which state it will be unable to pair correctly with the opposing base in the original strand. It will therefore be removed by the proof-reading mechanism. Spontaneous mutations will then only occur when this mechanism fails.

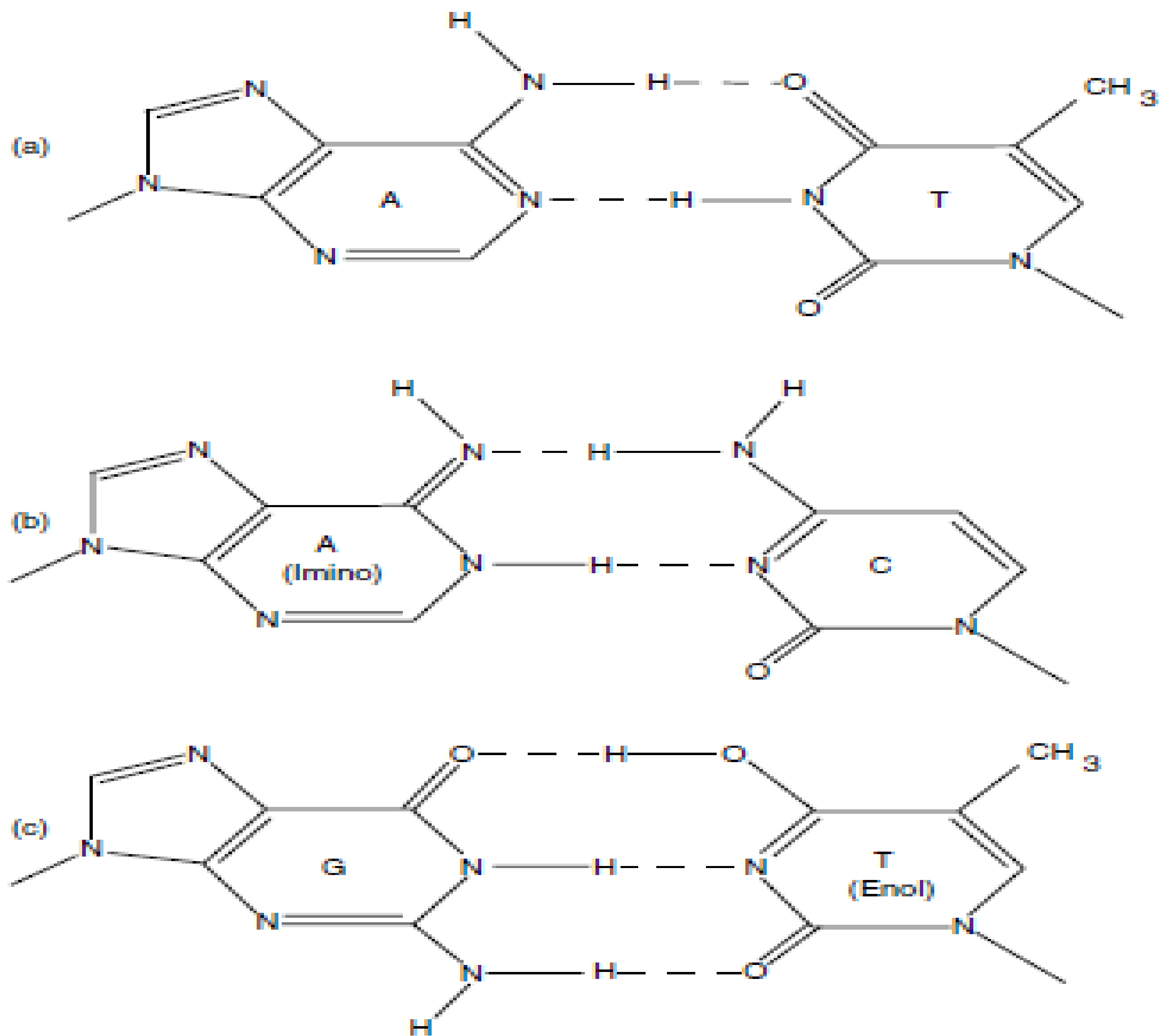


**Figure 2.8** Tautomerism of bases leads to mispairing. (a) Normal pairing of adenine (amino form) and thymine (keto form). (b) Adenine in the imino form pairs with cytosine. (c) Thymine in the enol form pairs with guanine

## 2.6.2 Chemical mutagens

The natural rate of spontaneous mutation is much too low for convenient isolation of most types of mutants (apart from a handful of easily selected mutations such as antibiotic resistance). Ways must be found of enhancing that frequency. It is often possible to use *in vitro* mutagenesis or transposon mutagenesis (see

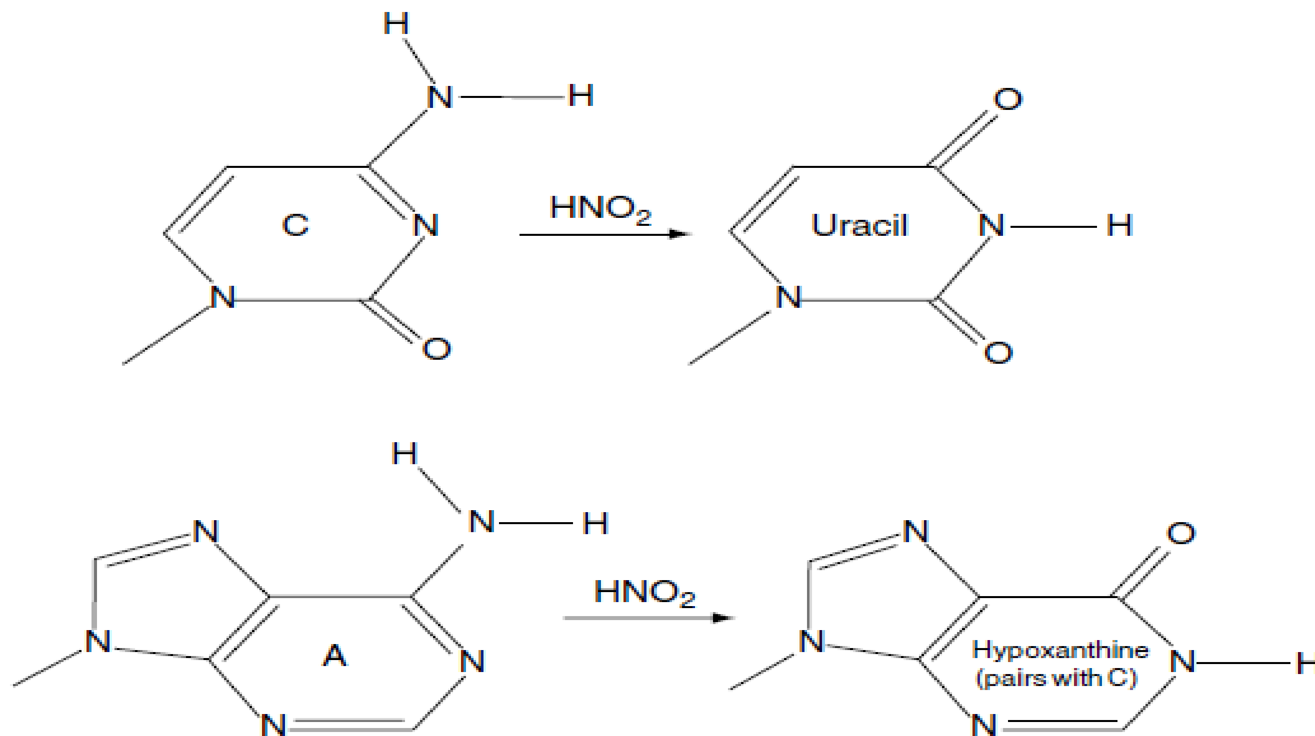


**Figure 2.9** Nitrous acid causes oxidative deamination of bases

Chapter 10), but there are still many situations where chemical or physical procedures are preferred or essential.

Many different chemical agents interact with DNA or the replication machinery so as to produce alterations in the DNA sequence. Of these, the simplest to understand are those agents that act by chemically modifying a base on the DNA so that it resembles a different base. For example, nitrous acid causes an oxidative deamination in which amino groups are converted to keto groups and thus cytosine residues for example will be converted to uracil (Figure 2.9). Uracil is not a normal base in DNA and the cell contains enzymes that will remove it. However, if it persists through to replication it will be capable of pairing with adenine, thus causing a change from a C-G pair to U-A and ultimately T-A. Similarly deamination of adenine creates the base hypoxanthine which will base-pair with cytosine.

Nitrous acid can react directly with isolated DNA, although to produce mutations the DNA must of course be reintroduced into a bacterial cell. Treatments of this sort are especially useful for producing alterations of bacteriophages or plasmids where isolation and reintroduction of the DNA is readily achieved and where it is desirable to mutate the phage or plasmid without exposing the bacterial cell itself to the mutagenic agent. This ensures that any mutations selected are a consequence of changes to the DNA of the bacteriophage or plasmid, rather than being due to alterations in a chromosomal gene.

Some types of chemical agents act against the DNA within cells, rather than against isolated DNA. Alkylating agents such as ethyl methane sulphonate (EMS) and 1-methyl-3-nitro-1-nitroso-guanidine (MNNG) are extremely

powerful mutagens – so much so that the latter in particular is extremely hazardous to use. They act by introducing alkyl groups onto the nucleotides at various positions, especially the $O_6$ position of guanine, and tend to cause multiple closely linked mutations in the vicinity of the replication fork.

The intercalating agents, such as acridine orange and ethidium bromide, have a different mechanism of action. These molecules contain a flat ring structure (Figure 2.10) which is capable of inserting (intercalating) into the core of the double helix between adjacent bases. The consequences of this are the addition (or sometimes deletion) of a single base when the DNA is replicated, giving rise to a frameshift mutation. These dyes (ethidium bromide in particular) are also much used in molecular biology in the detection of DNA, since the complex formed with DNA is fluorescent.

Another type of agent that acts only against growing cells (but with a very different mechanism) consists of the base analogues such as 5-bromouracil (Figure 2.11). Despite its name, this is an analogue of thymine in which the methyl group is replaced by a bromine atom which is a similar size. 5BU can be incorporated into DNA in place of thymine, since it will form base pairs with adenine residues on the template strand. However, the tautomerism referred to above is much more pronounced with 5BU. Therefore, in subsequent rounds of replication, it may pair with guanine rather than with adenine, thus giving rise to an A-T to G-C mutation.

### 2.6.3 Ultraviolet irradiation

Any agent that damages DNA can in principle lead either to the death of that organism or, amongst the survivors, to mutation. This is true of irradiation as

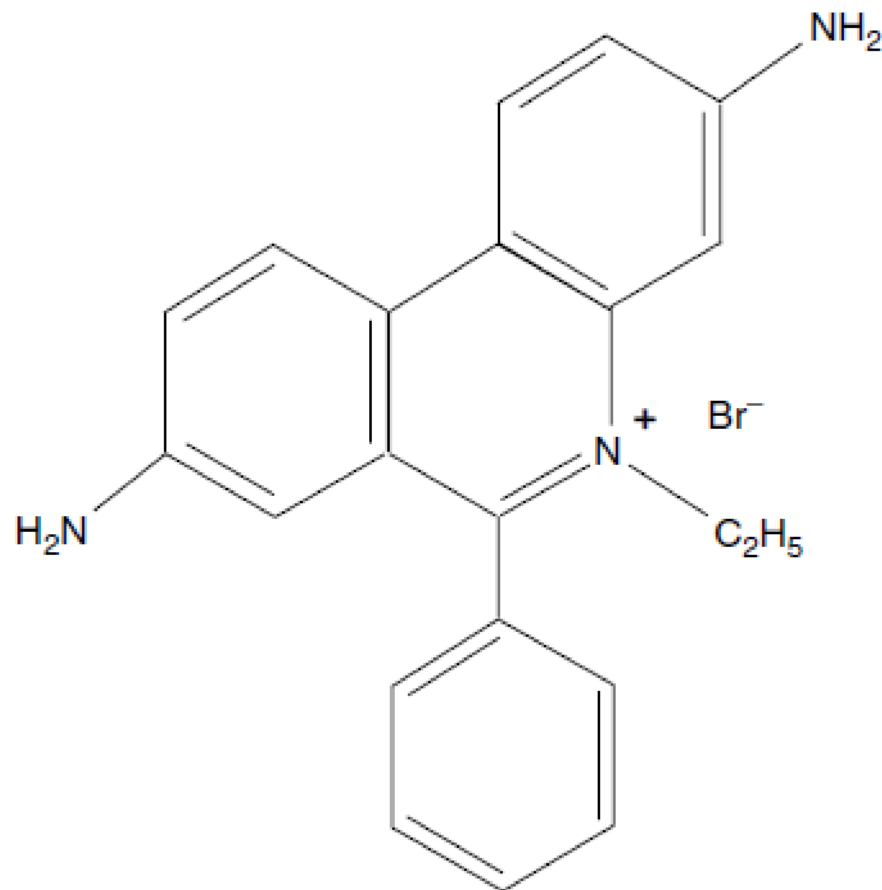


**Figure 2.10** Ethidium bromide

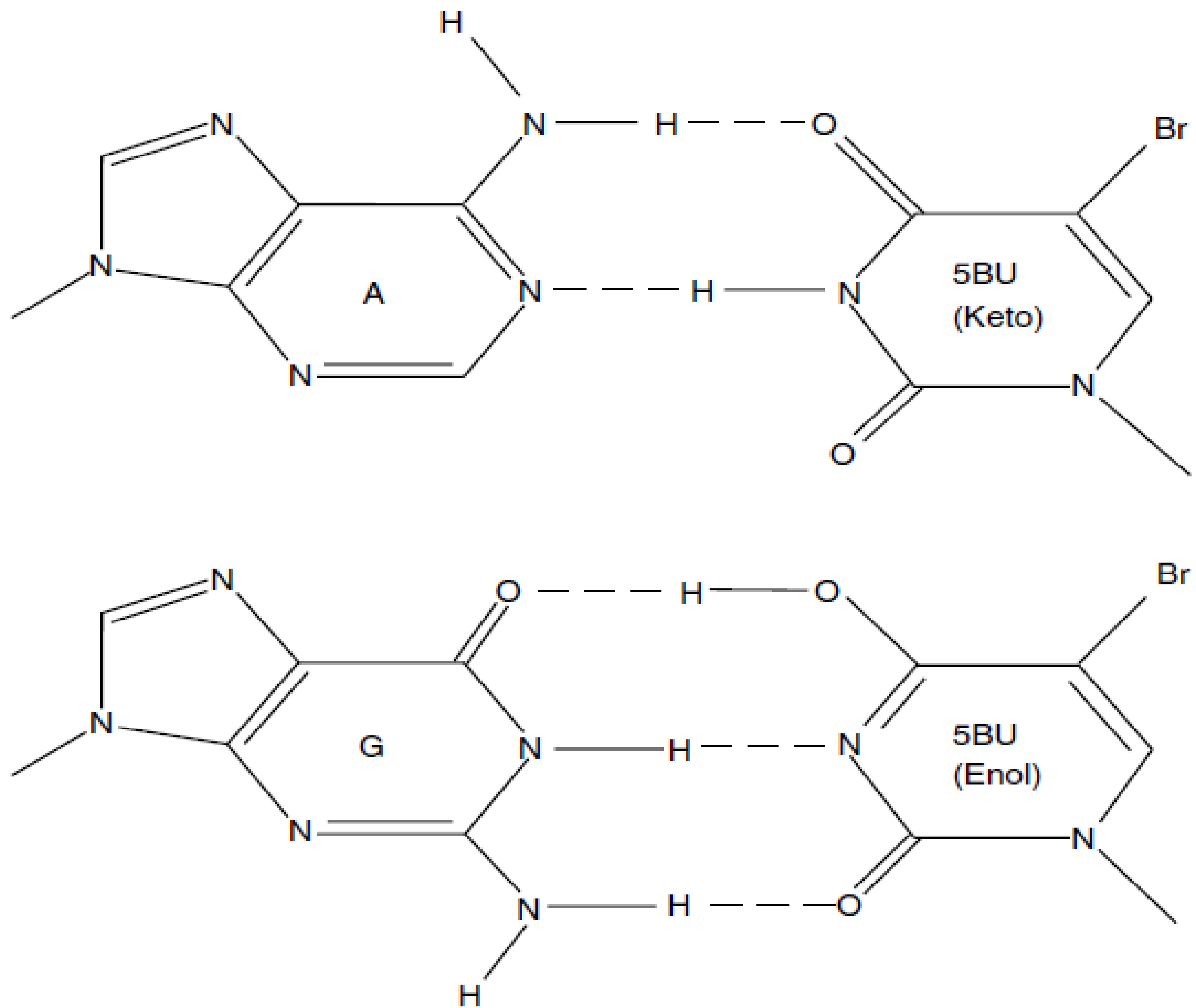


**Figure 2.11** Enhanced tautomerism by the base analogue 5-bromouracil

well as of chemical agents. Many types of irradiation have been used to generate mutations. The higher energy rays such as X-rays and gamma rays however require expensive apparatus and safety equipment and are not really suitable for routine use in a microbiology laboratory. In addition, they produce an excessive amount of gross chromosomal damage that is not easily repaired by the micro-organism. Ultraviolet irradiation on the other hand is easily controlled (although eye and skin protection is necessary) and requires only comparatively inexpensive equipment.

The principal effect of UV irradiation with which we are concerned is the production of *pyrimidine dimers* (commonly referred to as thymine dimers, although the effect can also occur with cytosine). Where two pyrimidine residues are adjacent on the same DNA strand (Figure 2.12) the result of UV irradiation is the creation of covalent links between them. These pyrimidine dimers cannot be replicated and are therefore lethal to the cell unless it is able to repair the damage.

It is the attempts to repair the damage caused by ultraviolet irradiation that can lead to mutagenic effects. Although most repair mechanisms are reasonably accurate (error-free repair), in the event of these mechanisms being unable to cope with the damage an additional defence comes into play. This error-prone

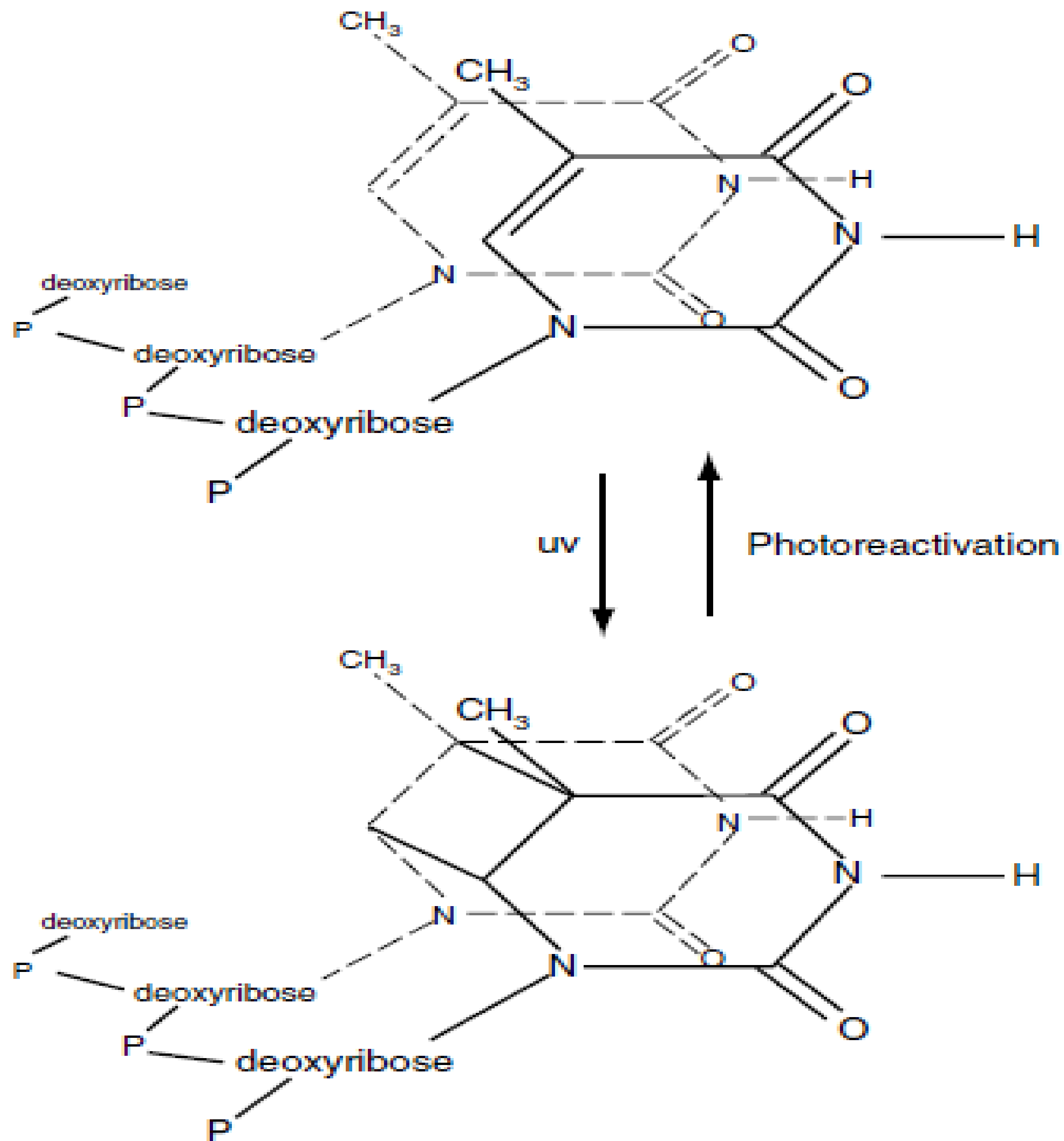


**Figure 2.12** Structure of thymine dimers

### *Photoreactivation*

The best defence that is mounted against damage by UV irradiation is known as *photoreactivation*. This is catalysed by an enzyme (photolyase) within the cells that in the presence of visible light can break the covalent bonds linking the two pyrimidine residues, thus re-establishing the original nature of the base sequence at that point. This method is very efficient and clearly does not lead to the establishment of mutations. For this reason, when UV mutagenesis is being performed in the laboratory, it is necessary to exclude light from the cultures



(e.g. by wrapping the bottles in foil) while the cells are recovering from the UV treatment.

*SOS repair*

Photoreactivation is an error-free process, and therefore does not lead to mutation. If light is excluded, and photoreactivation prevented, the cell still has recourse to other error-free repair processes such as excision repair and recombination repair (post-replication repair), as described in Chapter 1. It is only when these processes are overwhelmed (or if we use mutant strains that lack these error-free repair mechanisms) that significant numbers of mutations result. This is due to yet another repair mechanism that is error-prone which is part of the so-called SOS response. (Although we are considering this effect in relation to UV irradiation, it is also involved in repair of other types of DNA damage and is related to the more general stress response systems).

In the presence of damaged DNA, the expression of a number of genes involved in DNA repair is induced; these genes include the excision repair genes *uvrA* and *uvrB*, the *recA* gene and the genes involved in error-prone repair. This inducibility of the repair pathways can be demonstrated by irradiating lambda bacteriophage and testing its ability to form plaques on irradiated and non-irradiated host cells. More plaques are obtained if the host cells are also subjected to a low dose of UV irradiation before infection by the phage due to the induction of the repair enzymes in the host by the pre-existing DNA damage.

The mechanism of induction involves the products of two genes: *recA* and *lexA*. The LexA protein acts as a repressor of the genes of the SOS response, including both *recA* and *lexA* itself. The LexA protein also has the ability to cleave itself but only after the RecA protein binds to it. The RecA protein has a co-protease function in stimulating the self-proteolysis of LexA. This activity of RecA arises after it binds to single-stranded DNA which arises as a consequence of DNA damage. This causes a conformational change in the protein that enables it to bind to LexA, resulting in cleavage of LexA and expression of the SOS genes.

Two of the SOS genes in particular, *umuC* and *umuD*, are involved in mutagenesis, since strains that are defective in these genes not only have increased UV sensitivity but also are not susceptible to UV-induced mutagenesis. Certain plasmids carry analogous genes (*mucA* and *mucB*); the presence of these genes increases resistance to UV and also results in an increased level of mutagenesis by UV and many other mutagenic agents. The UmuCD complex is able to act as a DNA polymerase, taking over from the normal polymerase (DNA polymerase III) when that enzyme stalls due to the presence of DNA damage. The low specificity of the UmuCD polymerase enables it to continue synthesis of DNA past the lesion, but at the expense of producing errors in the new DNA strand.

## Mutation Sources

Mutations arise from a variety of sources, including:

## Tautomeric shift of bases

Each of the bases in DNA can appear in one of several forms, called tautomers, which are isomers that in the positions of their atoms and in the bonds between the atoms. The keto(C-O) and amino($C-NH_2$) form of each base is normally present in DNA, whereas the imino(C=NH) and enol(C-OH) forms of the bases are rare (Figure 1).Conversion of keto group in thymine & guanine to enol form and changing amino group in adenine & cytosine to imino form, are examples of tautomerization. Naturally, A in amino form pairs with T in keto form, whereas A in its imino form pairs with C and T in enol form match with G [1,6,8] (Figure 2).

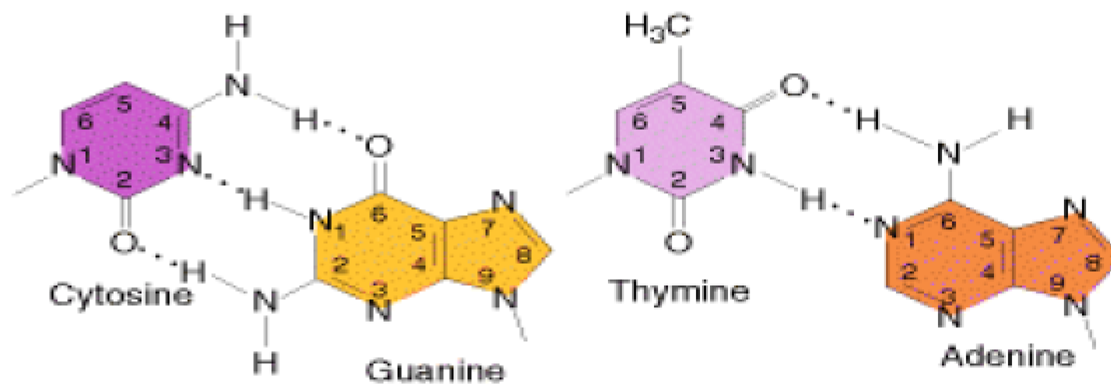


**Figure 1 :** Pairing between the normal (keto) forms of the bases.

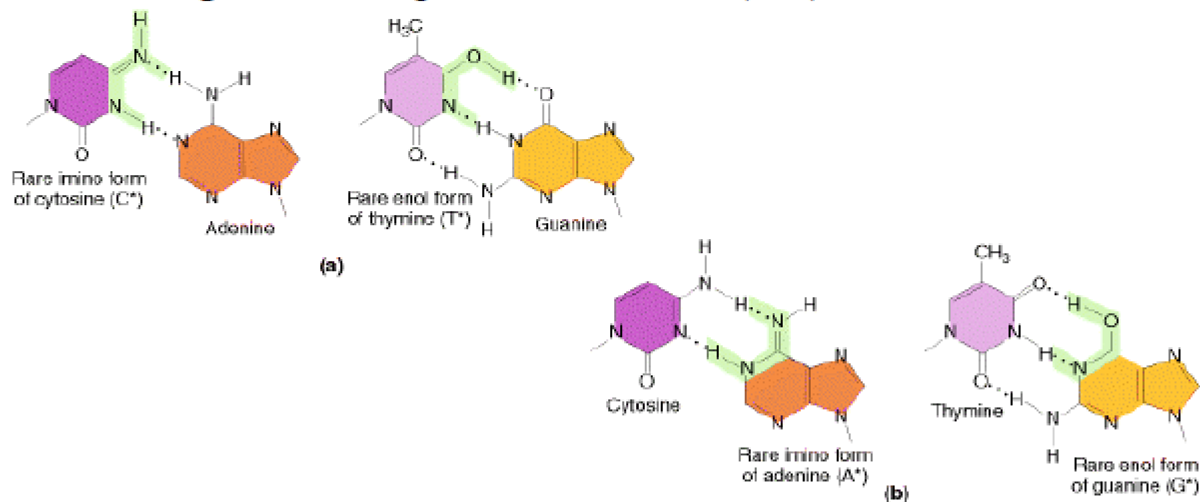


**Figure 2 :** Mismatched bases. (a) Mispairs resulting from rare tautomeric forms of the pyrimidines; (b) mispairs resulting from rare tautomeric forms of the purines.

## Depurination

Depurination consists of the interruption of the glycosidic bond between the base and deoxyribose and the subsequent loss of a guanine or an adenine residue from the DNA. in replication, the resulting apurinic sites cannot specify a base complementary to the original purine. So, efficient repair systems remove apurinic sites. Under this certain conditions, a base can be inserted across from an apurinic site; this insertion will frequently result in a mutation [6] (Figure 3) .

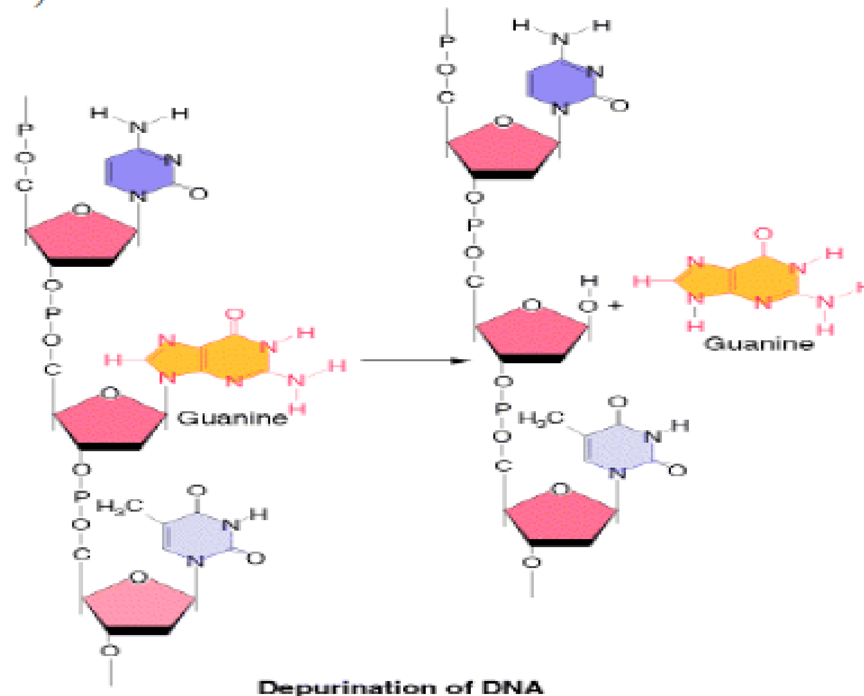


**Figure 3 :** The loss of a purine residue (guanine) from a single strand of DNA. The sugar-phosphate backbone is left intact.

### Deamination

The deamination of cytosine yields uracil. Unrepaired uracil residues will pair with adenine in replication, resulting in the conversion of a G–C pair into an A–T pair (a GC → AT transition). Another example for deamination, is conversion of 5-methylcytosine to thymine [1,2,6] (Figure 4) .

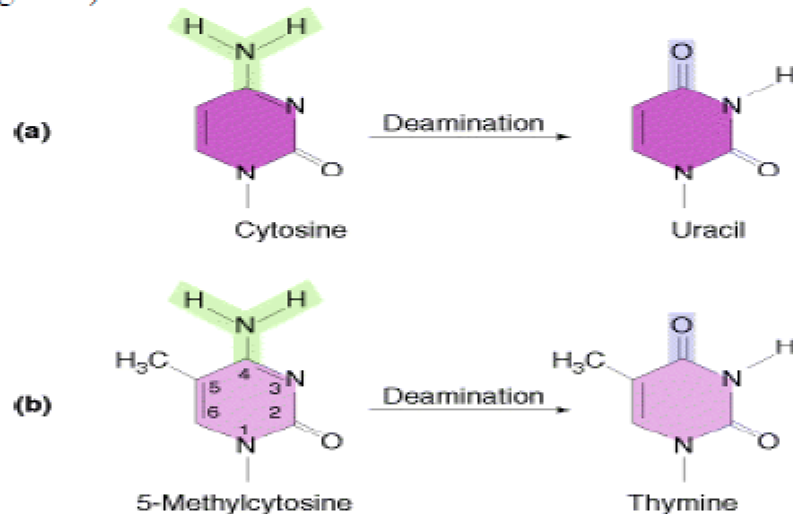


**Figure 4** : Deamination of (a) cytosine and (b) 5-methylcytosine.

### Oxidatively damaged bases

Active oxygen species, such as superoxide radicals ($O_2\cdot$), hydrogen peroxide ($H_2O_2$), and hydroxyl radicals ($OH\cdot$), are produced as by-products of normal aerobic metabolism. They can cause oxidative damage to DNA. For example, The 8-oxo-7-hydrodeoxyguanosine (8-oxodG, or GO) product frequently mispairs with A, resulting in a high level of G → T transversions. Thymidine glycol which induced of hydroxyl radical attack to thymine , blocks

DNA replication if unrepaired but has not yet been implicated in mutagenesis [6,10] (Figure 5) .

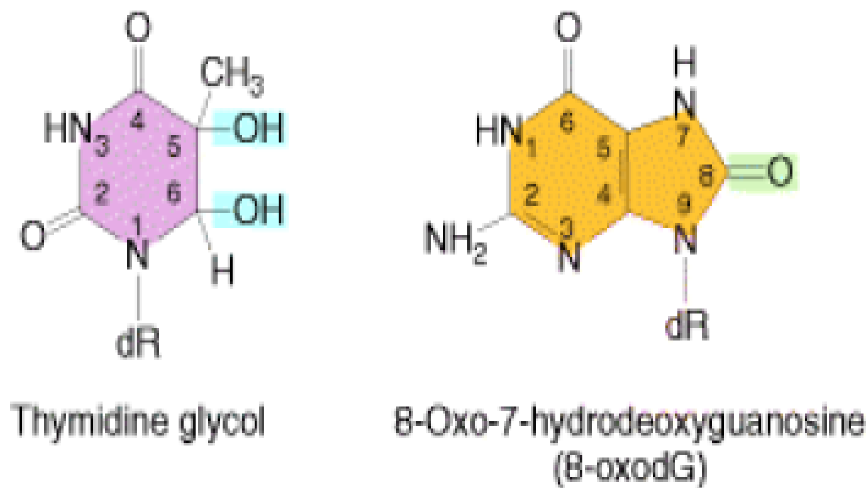


**Figure 5 :** DNA damage products formed after attack by oxygen radicals.

**Ultraviolet irradiation**

Ultraviolet light (opt 250 nm) causes bonds to form between adjacent pyrimidine residues (commonly referred to as thymine dimers, although the effect can also occur with cytosine) in the same polynucleotide strand. These are called pyrimidine dimers. UV irradiation can result in the formation of covalent links between pyrimidine dimers (Figure 6) .These bonds distort the DNA conformation and inhibit DNA replication and transcription [1,2,6,7,11].

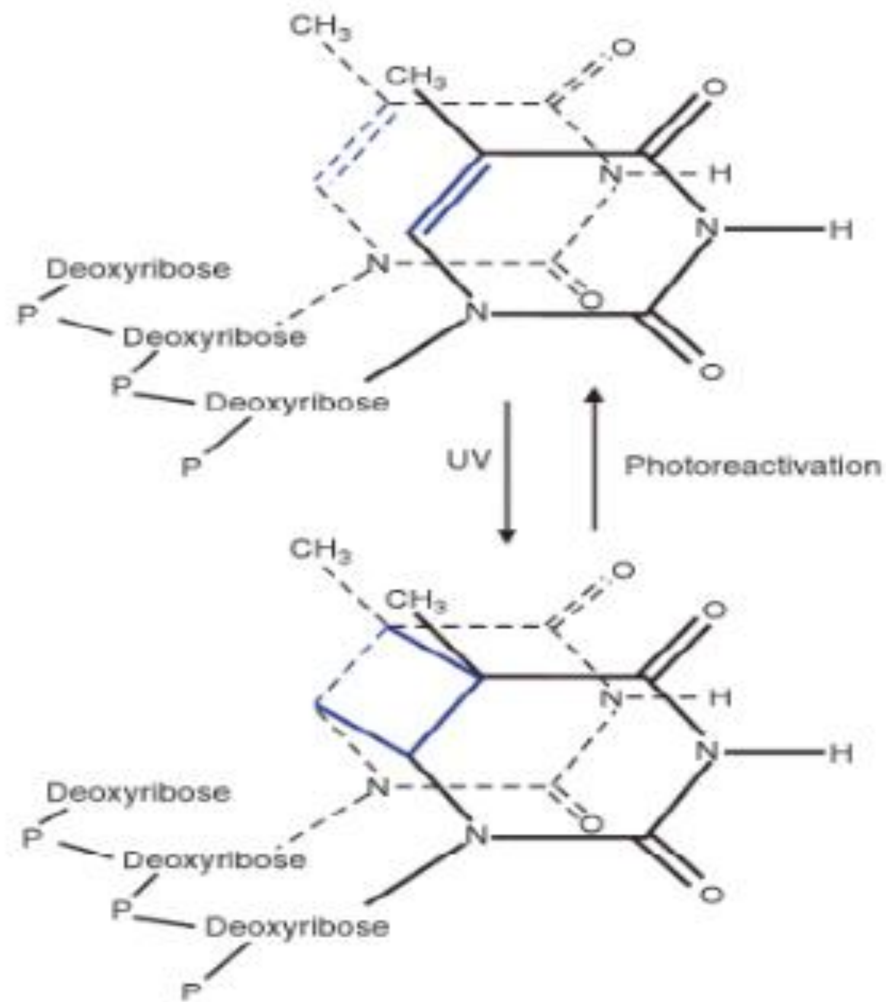


**Figure 6 :** Structure of thymine dimers.

# Insertion sequences

## 7.1 Insertion sequences

Insertion of a DNA fragment into a gene will usually result in the inactivation of that gene, and it is by the loss of that function that such events were initially recognized. A number of genetic elements, including some phages and plasmids (see earlier chapters), can be inserted into the bacterial chromosome. However, this chapter is concerned with elements that do not usually have any independent existence but are only found as a part of some other DNA molecule. The simplest of these genetic elements are known as Insertion Sequences (IS). As will be shown later, many of the other elements that participate in genetic rearrangements share key features with IS elements.

### *7.1.1 Structure of insertion sequences*

There are many IS elements known. They differ in size and other details, but the overall structure of most such elements is similar. One example (IS*1*) is shown in

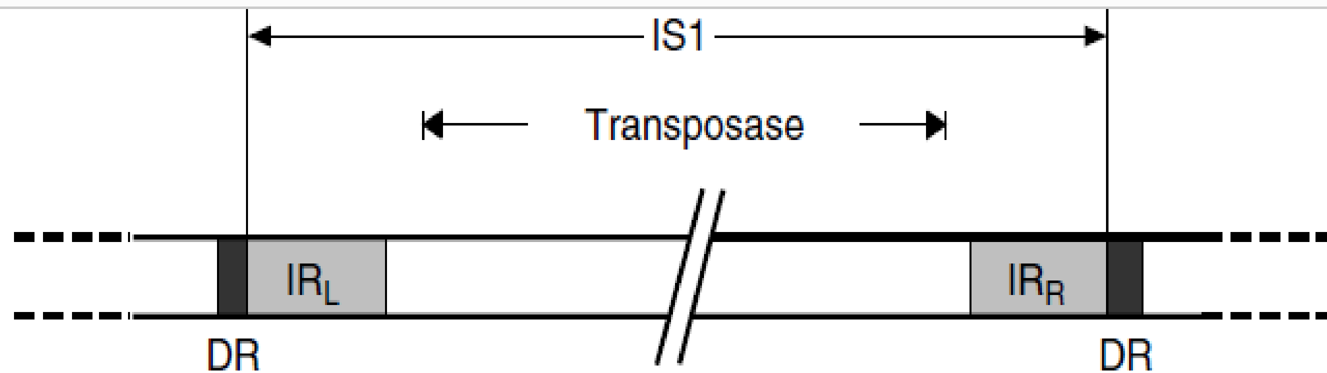


**Figure 7.1** Structure of the insertion sequence IS*1*. DR, direct repeat (duplicated target sequence); IR, inverted repeats

Figure 7.1; IS*1* is 768 bases long but many other IS elements are longer (usually 1300–1500 bases). The central region of an IS element codes for a protein (known as a *transposase*) which is necessary for the movement of the element from one site to another. At the ends of the insertion sequence are almost perfect inverted repeat (IR) sequences, which in IS*1* consist of 23 nucleotides. A minority of elements, such as IS*900* from *Mycobacterium paratuberculosis* do not have inverted repeat ends.

It must be stressed that reference to an inverted repeat of a DNA sequence does NOT mean that the sequence on an individual strand is repeated backwards, but that the sequence from left to right on the 'top' strand is repeated from right to left on the 'bottom' strand so that reading either copy of the IR in the 5′ to 3′ direction will result in the same sequence of bases. Since DNA sequences are often presented as just one of the two strands, an inverted repeat of the sequence CAT will appear as ATG.

In addition to the inverted repeats, inspection of a DNA region containing an insertion sequence usually shows a further short sequence that is duplicated – but this sequence is repeated in the same orientation and is therefore referred to as a *direct* repeat (DR). This is not part of the IS, but arises from duplication of the DNA at the insertion site (Figure 7.2) and therefore different copies of IS*1* will have different target sequence repeats depending on the point of insertion. Transposition of IS*1* generates rather long direct repeats (nine base pairs). With other insertion sequences, the direct repeats are commonly as short as two to three base pairs. The presence of these direct repeats is linked to the mechanism of transposition which is considered later in this chapter.

## 7.1.2 Occurrence of insertion sequences

Insertion sequences have been identified in most bacterial genera, although the presence and the number of copies of any one element often varies from strain to strain. A typical laboratory strain of *E. coli* for example might contain six copies of IS*1* as well as a number of copies of other insertion sequences.

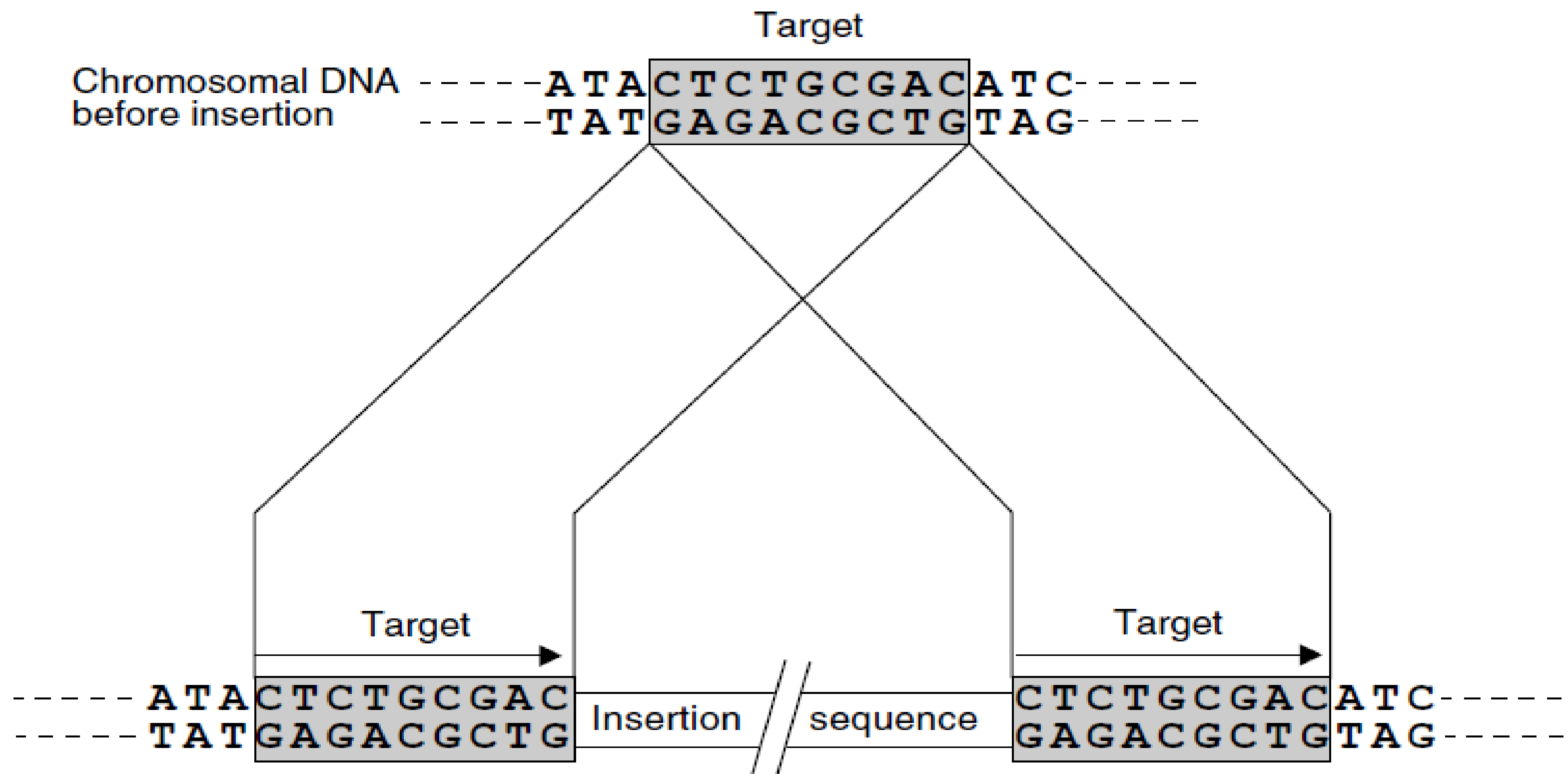


**Figure 7.2** Target duplication following insertion of IS*1*. Flanking IS*1*, there are direct repeats of a 9-bp sequence, derived from duplication of the target region. The top part of the diagram shows the target structure before insertion of IS*l*, while the lower part shows the insertion of IS*l* and the duplication of the target sequence

Hybridization of a Southern blot (see Chapter 2, Box 2.4) with a probe for a specific IS will produce a banding pattern which depends on the number of copies of the element and their position on the chromosome. Since both the copy number and the chromosomal distribution of an IS may vary from one strain to another, a different pattern may be seen for different isolates of the same species. This is exploited in a technique known as *restriction fragment length polymorphism* (or RFLP) for typing bacterial strains which is described more fully in Chapter 9.

IS elements are also commonly found on bacterial plasmids. For example, in Chapter 5 it was shown that the antibiotic resistance plasmid R-100 (Figure 5.5) carries two copies of IS*1* and two copies of a different insertion sequence, IS*10*. The presence of IS elements can be a major cause of plasmid instability since recombination between two insertion sequences on the same plasmid will lead to inversion or deletion of the intervening region. Insertion of the plasmid into the chromosome or recombination between two plasmids can also arise by homologous recombination between IS elements present on each DNA molecule. Similarly, the presence of two copies of an IS element in the chromosome can result in inversion or deletion of the region between them due to homologous recombination. IS elements can therefore play a significant role in the variation of genomic structure between one strain and another.

Although IS elements can affect the phenotype by inactivation or deletion of genes, they do not carry any genetic information other than the transposase needed for transposition. They are therefore of little or no direct benefit to the

bacterium. Why then does the cell tolerate their presence? The Darwinian notion of the 'survival of the fittest' would suggest that evolutionary pressure will tend to eliminate such elements that are not beneficial since their presence will constitute a metabolic drain on the cell (even if small). The simplest answer is that these sequences are essentially parasitic and have devised strategies that prevent their elimination.

For some IS elements, insertion into another DNA site is a replicative event: the original copy remains and a duplicate is inserted at the new site. The number of copies carried by the cell will therefore tend to increase. When there are a certain number of copies present in the cell, this process is repressed, which prevents the cell from being overwhelmed and dying (thus eliminating the 'parasite'). However, if the cell does lose one or two copies, the number is now lower than that needed for repression, leading to a renewed round of replication and insertion. Thus, like any well-adapted parasite, an insertion sequence 'colonizes' its host while refraining from doing sufficient damage to seriously weaken it.

## 7.2 Transposons

When resistance plasmids were first discovered, there was much speculation as to how a single element could have evolved to carry a number of different antibiotic resistance genes and in particular how apparently related plasmids could have different combinations of such genes (or, conversely, how otherwise dissimilar plasmids could carry related resistance genes). It was assumed that a basic

plasmid, having the ability to replicate independently but not carrying any other information, had somehow picked up a resistance gene from the chromosome of a resistant host strain. Transfer of this plasmid to an otherwise sensitive strain then produces a selective advantage for that strain, and therefore indirectly a selective advantage for this 'new' plasmid. As the plasmid moves from one organism to another it has the opportunity to acquire additional resistance genes, thus giving rise to a family of plasmids containing different combinations of resistance genes.

Since this model implies that unrelated plasmids could pick up the same gene independently, this would explain the widespread distribution of certain resistance genes, notably a type of β-lactamase (the enzyme that destroys penicillin and hence confers resistance to penicillins). This particular enzyme, the TEM β-lactamase, is the commonest type amongst plasmids in the Enterobacteriaceae and is also present in many members of the genus *Pseudomonas*. The same gene has also been found in connection with plasmid-mediated penicillin resistance in species as diverse as *Haemophilus influenzae* and *Neisseria gonorrhoeae*.

The reason behind the ubiquity of the TEM β-lactamase became apparent from the discovery that this gene could move (transpose) from one plasmid to another. This is exemplified by the conjugation experiment shown diagrammatically in

Figure 7.3. A strain of *E. coli* containing two different plasmids, one with an ampicillin resistance gene and one conferring resistance to kanamycin, is used as a donor. The recipient strain is sensitive to both drugs (but resistant to nalidixic acid). Plating the mixed culture on a medium containing nalidixic acid, ampicillin and kanamycin will therefore select for recipient cells that have received both resistance genes from the donor. It was found that resistance to both antibiotics was transferred at a rate much higher than would be predicted from the rate of independent transfer of the two plasmids. It was also found that the recipients which were resistant to both drugs contained a single plasmid carrying both resistance genes.

This effect was not due to ordinary recombination between the two plasmids since it occurred equally well in recombination deficient (*recA*) strains. From additional evidence, it was deduced that the ampicillin resistance gene had moved (transposed) from one plasmid to the other. The term *transposon* was coined to signify an element that was capable of such behaviour, i.e. a mobile genetic element containing additional genes unrelated to transposition functions.

This movement of resistance genes can occur not only between two plasmids but also from plasmid to chromosome and vice versa. It therefore provides part of the explanation for the observed rapid evolution of resistance plasmids and also of plasmids that carry genes other than antibiotic resistance. Although resistance

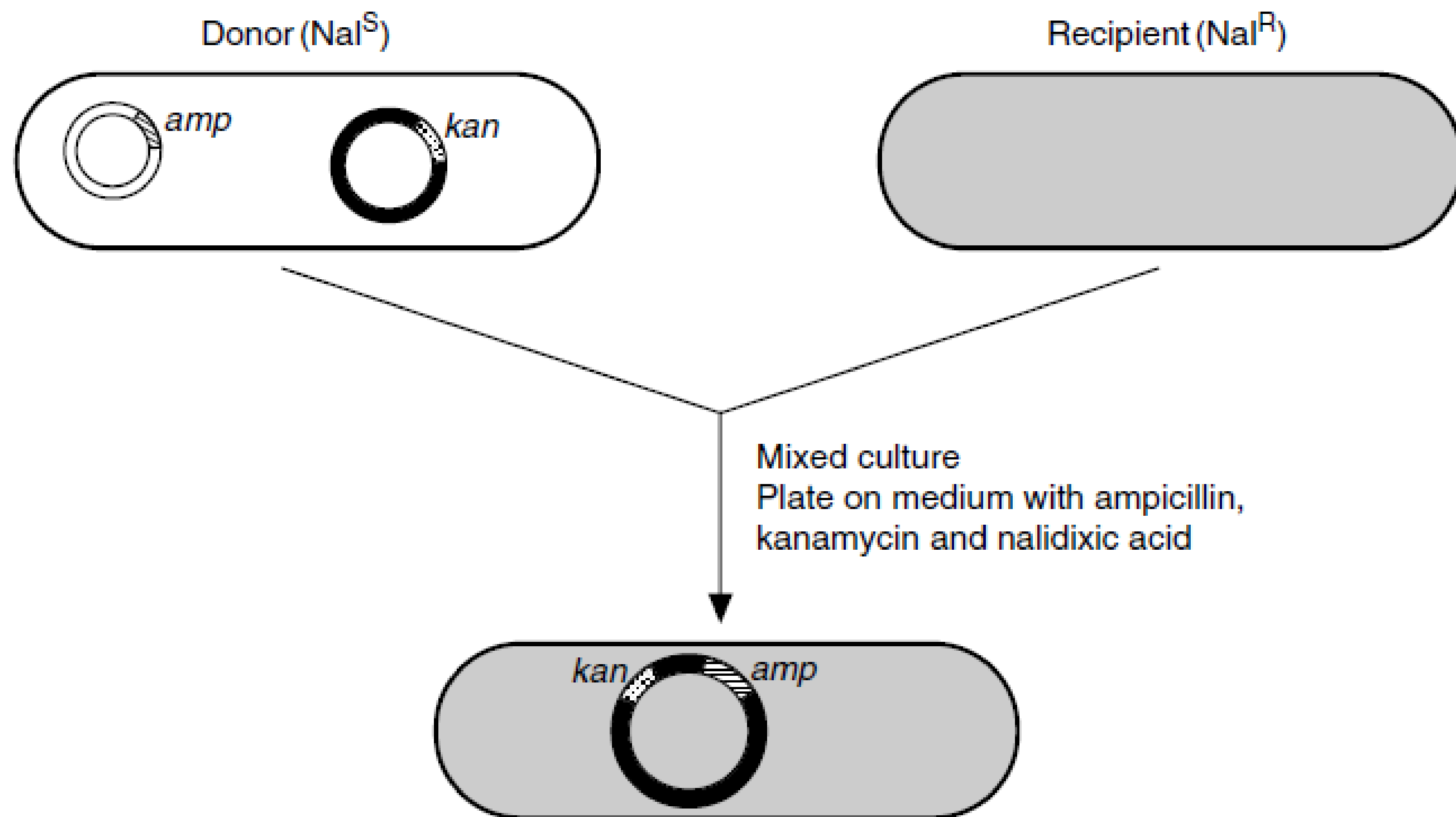


**Figure 7.3** Transposition of a resistance gene between plasmids. The donor strain has two plasmids, carrying ampicillin- and kanamycin-resistance genes (*amp* and *kan*) respectively. Conjugation with a nalidixic acid-resistant recipient, using a selective medium containing all three antibiotics, leads to colonies with a single plasmid carrying both *amp* and *kan*, due to transposition of *amp* to the second plasmid

### 7.2.1 Structure of transposons

The structure of a simple transposon, Tn*3*, is shown in Figure 7.4; it consists of about 5000 base pairs and has a short (38 bp) inverted repeat sequence at each end. It is therefore analogous to an insertion sequence, the distinction being that a transposon carries an identifiable genetic marker – in this case the ampicillin resistance gene (*bla*, β-lactamase). Tn*3* codes for two other proteins as well: a transposase (TnpA), and TnpR, a bifunctional protein that acts as a repressor and is also responsible for one stage of transposition known as resolution (this is explained more fully later on). As with the insertion sequences, there is a short direct repeat at either end of the transposon (five base pairs in the case of Tn*3*).

Some transposable elements have a more complex structure than Tn*3*. These *composite transposons* consist of two copies of an insertion sequence on either side of a set of resistance genes. For example the tetracycline resistance transposon Tn*10*, which is about 9300 bp in length, consists of a central region carrying the resistance determinants flanked by two copies of the IS*10* insertion sequence in opposite orientations (Figure 7.5; see also Figure 5.5). IS*10* itself is about 1300 bp long with 23-bp inverted repeat ends and contains a transposase gene.

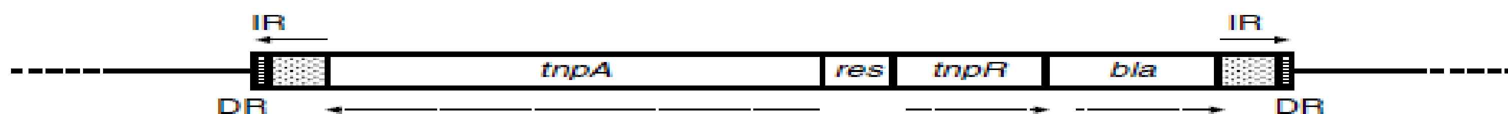


**Figure 7.4** Structure of the transposon Tn*3*. DR, five-base pair direct repeat (target duplication); IR, 38-base pair inverted repeats; *res*, resolution site; *tnpA*, transposase; *tnpR*, resolvase; *bla*, β-lactamase (ampicillin resistance)

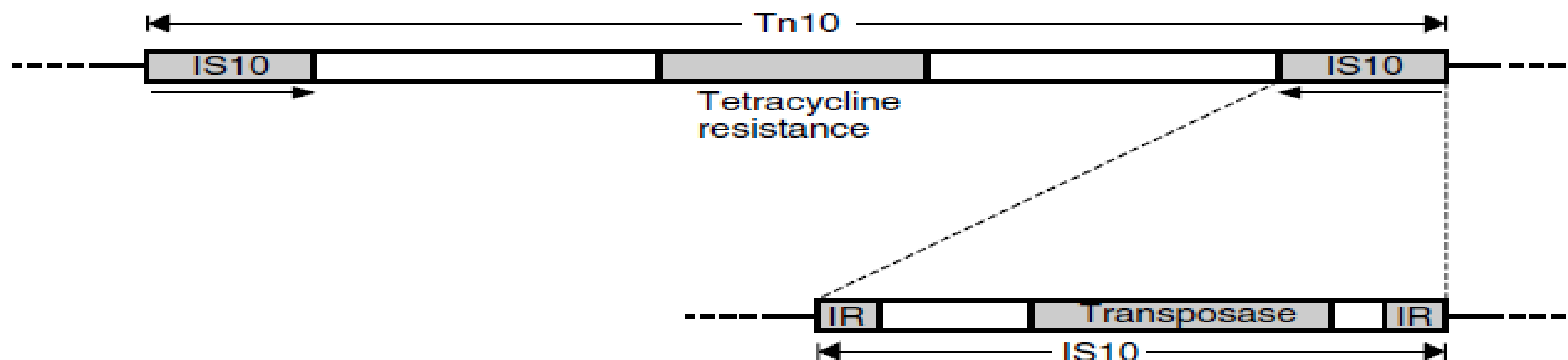


**Figure 7.5** Structure of a composite transposon Tn10

### 7.2.2 Integrons

Extremely complex and large transposons can also be built up by insertion of additional genes within an existing transposon. Many large transposons have been identified which are related to Tn*21* which has a structure analogous to class II transposons such as Tn3: it has inverted repeats (38 bp) at each end and carries genes for transposition functions (Figure 7.7). Tn*21* may have developed from a smaller transposon (such as Tn*2613*) by acquisition of additional genes. Tn*2603* and Tn*1696* (and a family of other transposons) are also very similar to Tn*21* but contain additional resistance genes.

It is now known that the transposons in the Tn*21* family have acquired resistance genes by a specific mechanism. Each individual gene has been inserted separately, as a *gene cassette* which contains a single gene and a recombination site. Tn*21* contains a site known as an *integron* into which such gene cassettes can be inserted by site-specific recombination. The integron region in Tn*21* also contains a gene coding for an integrase which is responsible for the site-specific recombination (and is related to the bacteriophage λ integrase; see Chapter 4). After insertion of a gene cassette into the integron, the recombination site remains available for insertion of a further gene cassette, enabling the build-up of an array of several cassettes within the integron. A further twist to the story is that the gene cassettes do not normally contain a promoter. However there is a promoter region within the integron itself, upstream from the insertion site, so each of the gene cassettes is transcribed from the integron promoter. Integrons are thus a naturally-occurring analogy to the expression vectors that will be discussed in Chapter 8, for obtaining expression of foreign genes by inserting them into a vector adjacent to a promoter.

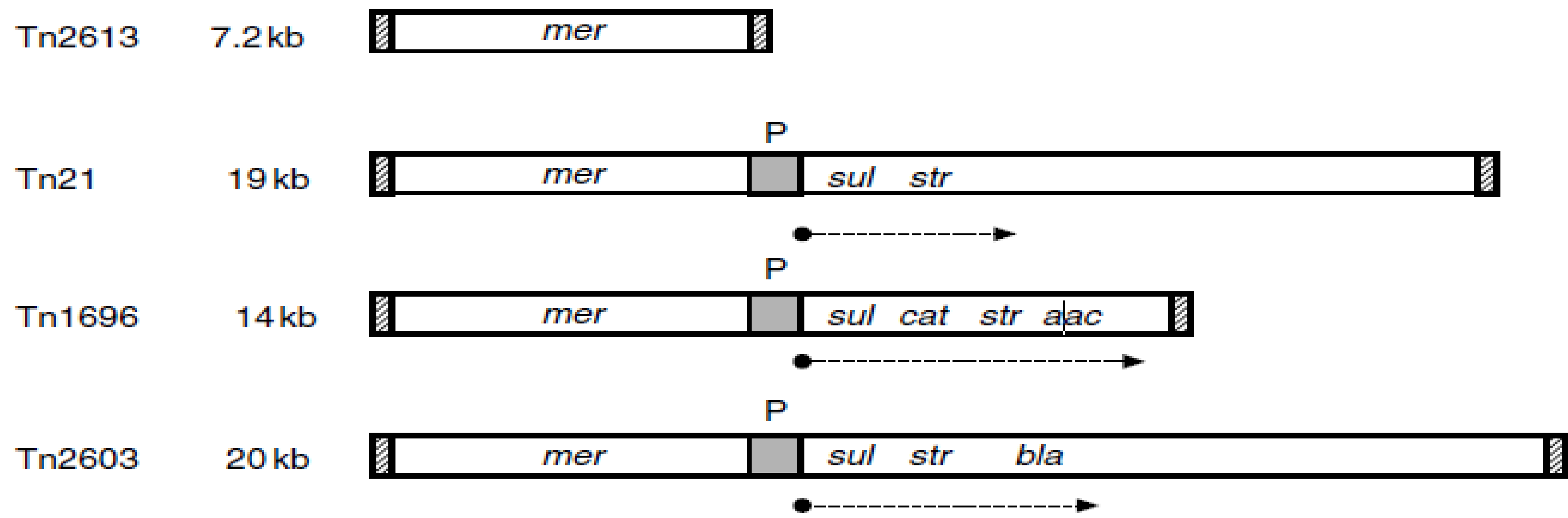


**Figure 7.7** Tn21 family of transposons: integrons. Gene cassettes are inserted to the right of the promoter (P), by means of a transposon-encoded integrase (position not shown). *aac*, aminoglycoside acetyltransferase (gentamicin resistance; see Figure 7.6 for the identity of the other genes)

## Methods for mutant selection

### Replica plating

In this procedure, the mutagenized culture is plated out to obtain single colonies on a nutrient medium on which mutants and parents will grow. After incubation, the colonies are replicated, using a sterile velvet pad, onto a minimal agar plate and then a similar plate to which the appropriate supplement (in this case, histidine, since we are looking specifically for histidine auxotrophs) has been added. Histidine requiring auxotrophs will be unable to grow on the first plate, but will grow on the second one. Thus, mutant colonies can be identified with localization of colonies that have failed to grow in second plate in comparison with first one [1,5,14] (Figur 12).

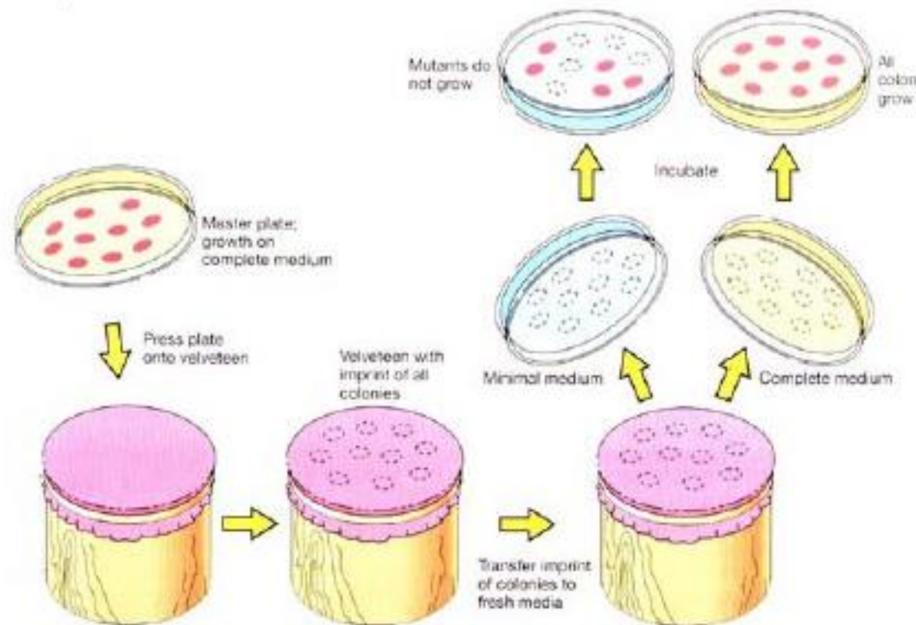


**Figure 12** : Replica plating to isolate auxotrophic mutants.

**Ames test**

The Ames test uses several strains of the bacterium *Salmonella typhimurium* that carry mutations in genes involved in histidine synthesis i.e. it is an auxotrophic mutant, so that they require histidine for growth. The method tests the capability of mutagen in creating mutations that can result in a reversion back to a non-auxotrophic state so that the cells can grow on a histidine-free medium. The bacteria are spread on an agar plate with a small amount of histidine. This small amount of histidine in the growth medium allows the bacteria to grow for an initial time and have the opportunity to mutate. When the histidine is depleted, only bacteria that have mutated to gain the ability to produce their own histidine will survive. The plate is incubated for 48 hours. The mutagenicity of a substance is proportional to the number of colonies observed [7,14 ,17,18] (Figur 14).

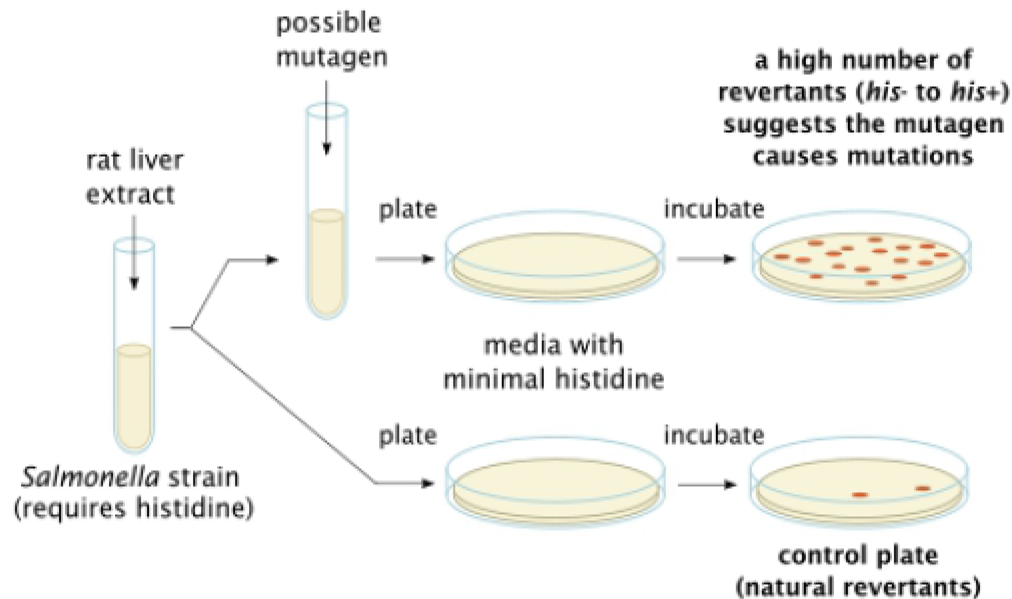


**Figure 14 :** Ames test procedure.

**Use of chromogenic substrate**

Mutants that are not able to utilize a particular carbon source (lactose, for example) can be isolated with use a chromogenic substrate that shows an easily detectable color change when acted on by the enzyme concerned. In this case the enzyme is *β*-galactosidase, which catalyses the hydrolysis of lactose into its constituent sugars glucose and galactose. A commonly used chromogenic substrate for *β*-galactosidase is 5-bromo-4-chloro-3-indolyl-*β*-D-galactoside, more popularly known as X-gal. This is a synthetic analogue of the natural substrate, containing a dye linked to galactose. X-gal itself is colorless; the color of the dye is only manifest when it is released by hydrolysis of the linkage by *β*-galactosidase. Lac+ colonies will be blue on a medium containing X-gal and colonies that do not produce *β*-galactosidase will be white[1,19].

**Polymerase chain reaction (PCR) and Gel electrophoresis**

In this method, target gen (mutant sequence in this case) can be increased exponentially. If the mutation has caused a significant change in the size of a specific gene (such as an insertion or a deletion) it will be detect by a change in the size of the PCR product, using gel electrophoresis [1,20,21,22,23].

**Gene probes and Southern blotting**

This method involves separating fragments of DNA by electrophoresis in an agarose gel and transferring them to a filter which can then be hybridized with the labeled probe. By using highly specific probes, detection of difference in the sequence is possible [1,14,21,23].

**DNA sequencing**

DNA sequencing template is amplified using PCR and oligonucleotide primers flanking the region of interest. The amplified fragment is directly cycle sequenced using fluorescent sequencing primers, Sanger dideoxy sequencing chemistry and an enzyme mixture of a mutant Taq DNA polymerase and thermostable pyrophosphatase. The sequence ladders produced are analyzed on a real-time, automated four-color sequencing system. The method produces sequence ladders from unpurified PCR fragments of sufficiently high quality such that heterozygotes can be reproducibly detected and identified by software that recognizes signal-strength patterns indicative of mixed-base positions [1,24,25,26].

**DNA microarray**

Base-pairing or complementarity is the principle behind this emerging technology. The potential applications of microarray technology are gene expression profiling and identification of gene sequences (including sequences that bear mutations). In this method, a large number of DNA fragments are placed on a glass slide. The fragments are allowed to complement or bind with the labeled DNA (probes), which hybridize with the DNA on the glass slide. The amount of hybridization is then analyzed in each spot on the slide. The genes are given a color where the hybridized ones are colored red and the genes that are hybridized least are colored green. It has been reported that microarray technology could be used in monitoring chromosome gains and losses, tumor classification, drug discovery and development [27,28,29].

# References

Molecular Genetics of Bacteria. By Jeremy W. Dale and Simon F. Park2004 , 4 edition, John Wiley & Sons Ltd.

p.:

http://www.google.iq/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0ahUKEwjFzsnz7rzSAhUC3iwKHRI9BhgQFggqMAM&url=http%3A%2F%2Fshahroodut.ac.ir%2Ffa%2Fdownload.php%3Fid%3D1111127349&usg=AFQjCNH3Q6IyrUwkBEqLYW_wTj1iyJj3QQ&bvm=bv.148747831,d.bGg