Biol 213 Genetics: Friday, November 17, 2000
Mutagenesis

Genetics Bulletin Board

Found: One calculator, found after last exam. Brad has it.

Old exams available: Some of you have not picked up the second exam, passed out last week. One of you hasn't picked up the first exam. Jeff has them.

Helen Fillmore to speak: Is neural dysfunction your thing? If so, be sure to come to next Monday's seminar at 12:30 in CB-01. Helen Fillmore of the Medical College of Virginia will talk about treatment of Parkinson's and other neurodegenerative diseases.

Outline (topics in italics to covered on Wednesday)

I. Introduction

II. The clonal nature of mutation (pp.463-466)

III. Spectrum of spontaneous mutations

IV. Types of base substitutions
- *A. Spontaneous base substitutions (pp.455-456)*
- B. Environmentally-induced base substitutions (pp. 469-471)

V. Duplication/deletion
- A. Duplication/deletion at the lacI hotspot
- B. Duplication as a cause of human disease (pp.459-460)

VI. Repair of mutation (pp.472-478)
- A. Bypass replication
- B. Excision repair
- C. Mismatch repair

IV. Types of base substitutions (continued)

IV.B. Environmentally-induced base substitution (pp.469-471)

It is not enough that bases change on their own. We are bombarded by chemicals that increase the frequency of base alterations. Despite the variety of compounds that affect DNA, their mechanisms of action can be put into four general categories:

Base modification: Certain chemicals may react with bases to change their pairing properties.

Partial mimicry: Certain chemicals may be incorporated into DNA in place of a base. Often the chemical is capable of pairing with more than one base, leading to mutation.

Intercalation: Certain chemicals may insert themselves between bases stacked in double-stranded DNA. During replication, this can lead to single base insertions or deletions.

DNA damage: Certain chemicals and physical insults can alter DNA to the extent that it can no longer be replicated. This is not directly mutagenic, since the damage is not inherited. However, cellular efforts to rescue the damaged DNA may lead to heritable mutations.

IV.B.1. Example of base modification: nitrous acid

Fig. 17-8 shows nitrous acid (HNO_2) chemically altering cytosine and adenine. (What is nitrous acid? Read food labels!). One result of exposure to nitrous acid is that some cytosines are converted to uracil, giving rise to a CG → TA transition (Fig. 1). Note that not all cytosines are affected. In fact, the reaction is very rare. Nitrous acid also reacts with adenine, converting it to hypoxanthine, which is capable of basepairing with cytosine. This leads to an AT → GC transition. The chemical has no effect on thymine (because it has no amine group) nor on guanine (because the chemical alteration doesn't happen to produce different base-pairing properties).

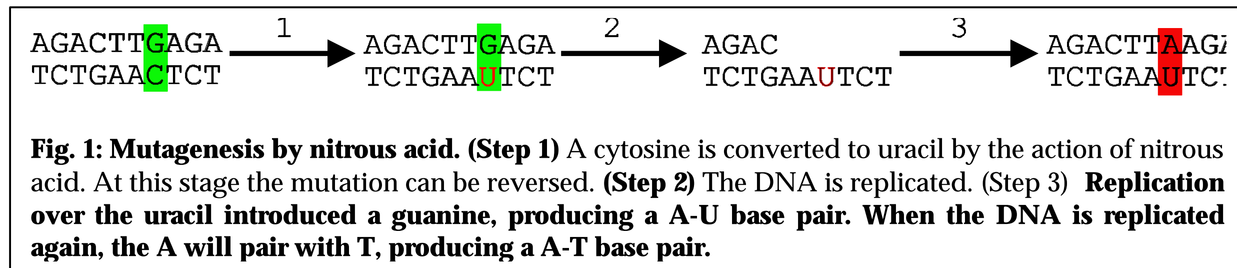

Fig. 1: Mutagenesis by nitrous acid. (Step 1) A cytosine is converted to uracil by the action of nitrous acid. At this stage the mutation can be reversed. **(Step 2)** The DNA is replicated. (Step 3) **Replication over the uracil introduced a guanine, producing a A-U base pair. When the DNA is replicated again, the A will pair with T, producing a A-T base pair.**

IV.B.2. Example of base mimicry: 5-bromouracil (5-BU)

Chemicals that incorporate into DNA in place of G, A, T, or C would do little harm if they were good enough mimics. The problem arises when their mimicry fails. Fig. 2 provides the example of 5-bromouracil (5-BU), which behaves similarly to the structurally similar thymine (5-methyluracil). This analog, however, is much more prone than thymine to experience the imido tautomerization shown in Fig. 2D, which leads to a significantly higher rate of mispairing with guanine. Like all tautomerizations, the change is transient and leads to heritable mutation only when made permanent by DNA replication (Fig. 17-9b).

IV.B.3. Intercalation example: proflavin

We spent considerable time earlier in the semester understanding how proflavin produced either single-base insertions or single-base deletions, owing to its ability to fit between the bases of double stranded DNA while the molecule was being replicated.

IV.B.4. Example of DNA damage: ultraviolet (UV) radiation

Some chemical agents and physical treatments go beyond the ambiguity of the previously described mutagens. UV radiation, for example, damages DNA to such an extent that replication is no longer possible without repair. It does so because thymine absorbs UV to create a highly reactive species capable of reacting covalently with an adjacent thymine (Fig. 17-10). The result is the cross-linking of the two pyrimidines,

known as a thymidine dimer. Cytosine can be similarly induced to crosslink, so “pyrimidine dimer” is the more accurate term. Such dimers block DNA replication and must be excised for replication to continue. This kind of damage occurs only when two pyrimidines are next to each other (T-T, C-T, or T-T). Different environmental insults produce characteristic signatures of mutations, making it possible sometimes to proceed backwards, deducing the mutagen from the type of mutations it produces.

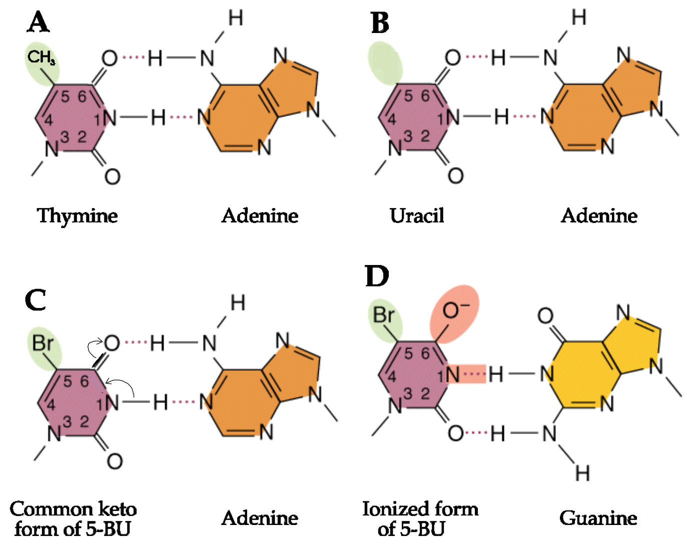

Fig. 2: Molecular mechanism underlying mutagenesis by 5-bromouracil (5-BU). Green oval emphasizes the differences amongst thymine, uracil, and 5-BU. **(A)** Normal thymine-adenine pairing. **(B)** normal uracil-adenine pairing. **(C)** Pairing between adenine and 5-BU in its usual form. Arrows show transition from keto form to ionized form. **(D)** Pairing between guanine and 5-BU in its less common form. (From Griffiths et al (2000), *An Introduction to Genetic Analysis*, 7th edition. WH Freeman, New York]

SQ1. Which of the mutagens that were discussed above would be expected to produce:

- **a. A mutation from the codon TGG (tryptophan) to CGG (arginine)**
- **b. A mutation from GCA (alanine) to GTA (valine)**
- **c. A mutation from GGA (arginine) to GAA (glutamate)**
- **d. A frameshift mutation**
- **e. Cell death arising from damage at a ATT (isoleucine) codon**

V. Duplication/Deletion

V.A. Duplication/Deletion at the lacI hotspot (pp.587-588)

We've been worrying about the small stuff as far as *lacI* is concerned. Go back to Table 1 in Wednesday’s notes and you’ll see that the most frequent cause of *lacI* mutations, 54%, is insertion at a certain hotspot and that the second most frequent cause, 13%, is deletion at a certain hotspot. And it's the same hotspot! That's more than two-thirds of the mutations occurring at the same place in the gene (pUR3 coordinate 2776, for those following along in the lab manual). It clearly behooves us to find out what's going on.

Fig. 3 shows the sequence of the hotspot, the insertion mutant, and the deletion mutant. The hotspot consists of a repeat of four bases: TGGC. The insertion mutant has one too many, and the deletion mutant has one too few. Either way, the gene is in trouble, because the insertion or deletion causes a frameshift, changing all amino acids after the deletion.

wild type *lacI*	Leu Arg Leu Ala Gly His Trp Lys Tyr CTG CGT CTG GCT GGC TGG CAT AAA TAT . . .
Insertion mutant	Leu Arg Leu Ala Gly His *Leu Ala* ***STOP*** CTG CGT CTG GCT GGC TGG CTG GCA TAA
Deletion mutant	Leu Arg Leu Ala Gly *Ile Asn Ile* CTG CGT CTG GCT GGC ATA AAT ATC . . .

Fig. 3. Sequence of *lacI* mutational hotspot. Wild-type *lacI* has three iterations of the tetranucleotide TGGC at position 2778. The most common spontaneous mutants of *lacI* show either one too many tetranucleotide (insertion mutant) or one too few (deletion mutant). In either case, the reading frame is thrown off and the resulting LacI protein has no activity as a repressor.

SQ2. You might think that *lacI* insertion mutants with more than four tetranucleotide repeats would be even deader than those with four, if that were possible. However, it may well be that a mutant *lacI* with <u>six</u> tetranucleotide repeats encodes a functional repressor! Why is that?

SQ3. Fig 3 shows the repeat as a TGGC tetranucleotide. In Wednesday's notes it is shown as a CTGG tetranucleotide. Which is correct?

How do the deletions and insertions come about? It turns out that repeated sequences are intrinsically unstable in this way, some more than others. The problem comes during replication. If the replicating strand slips and mispairs with the template, continuing replication will duplicate one of the repeats (Fig. 4). If, on the other hand, the <u>template</u> strand slips and mispairs, then continuing replication will <u>delete</u> one of the repeats (not shown).

SQ4. Draw a diagram showing how mispairing of the template strand leads to <u>deletion</u> of a repeat in the replicating strand.

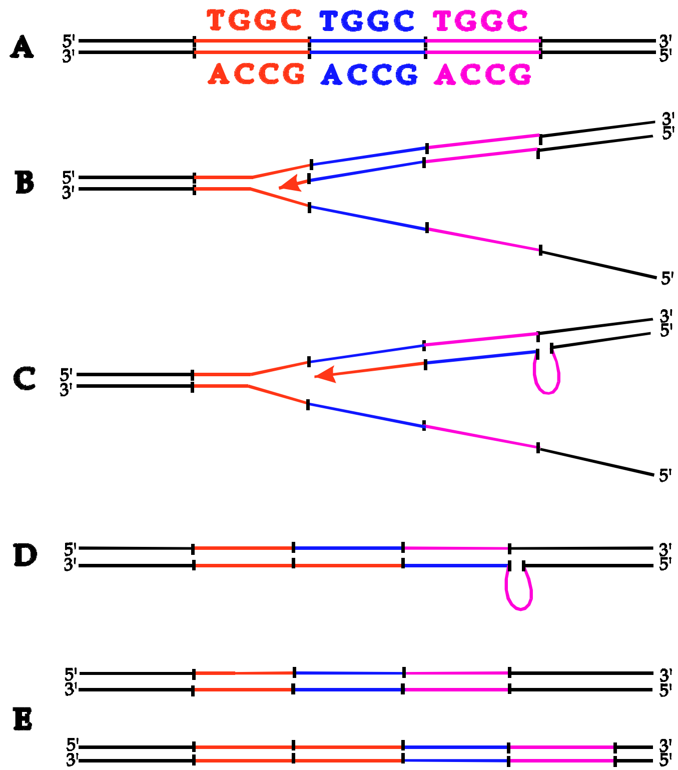

Fig. 4: Insertion mutation generated by slipped strand mispairing. (**A**) Wildtype *lacI*, with three tetranucleotide repeats. (**B**) Part way into replicating the repeated region. (**C**) The replicating strand slips and mispairs, taking advantage of the duplication so that one repeat pairs with the adjacent repeat. (**D**) Continuing replication generates a fourth repeat. This process, slipped strand mispairing is also diagrammed in Fig. 19-4.

V.B. Insertion as a cause of human disease (pp.459-460)

In the last decade, several human genetic diseases have been shown to result from insertions within long repeats (some are shown in Table 17-3). The role of slipped strand mispairing (Fig. 4) in the inheritance of this disease solves a mystery of long standing. It has long been noted that there is an excess of males among mentally retarded people. While this may not come as a surprise to roughly half of you, who knew it all along, it nonetheless is worthy of explanation. The answer, we can now appreciate, is that there are many X-linked syndromes associated with mental retardation, fragile X syndrome being the most common. The name "fragile X" refers to the appearance of the tendency of the X chromosome to break under certain artificial conditions. Fragile X, though certainly inherited, has decidedly weird genetic properties, which we'll now consider in depth.

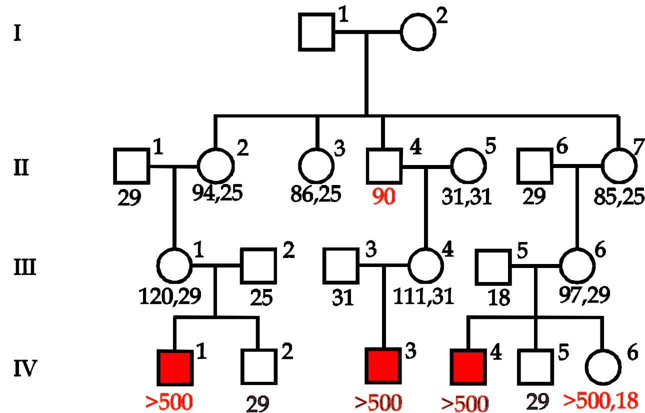

Fig. 5: Pedigree of family with high incidence of fragile X syndrome. Affected members have red-filled symbols. Numbers below symbols represent number of repeated sequences in *FMR1* gene.

Our story begins with a type of inheritance that defies Mendelian analysis. Fig. 5 shows a pedigree of a family with fragile X syndrome (ignore the numbers below the boxes for the moment).

SQ5. From the pedigree, would you describe the trait as dominant or recessive?

Try it as you like, the pedigree does not fit with any of the models we've come up with so far. Even low penetrance doesn't have much appeal, since you would have to postulate that it is mere coincidence that all of the affected individuals are in the same generation. On the contrary, what you see is the norm for fragile X syndrome: clusters of affected individuals, related to each other through a common grandparent!

The gene determining fragile X syndrome, *FMR1*, was cloned earlier this decade, and the sequence offered a surprise:

The region just upstream from the gene contained several copies of a three-basepair unit. In people from families with no history of the syndrome, the number of repeats varied from 6 to 52. In contrast, individuals affected by the disease had hundreds, even thousands of copies of the repeated trinucleotide. The numbers in the pedigree shown in Fig. 5 shows for each individual how many repeats they have.

SQ6. Why do females have two number of repeats and males only one?

SQ7. Can you deduce from which parent each fragile X allele was donated?

Amazingly enough, mutation in this region is frequent, <u>very</u> frequent, quite unlike the one in a million mutation rate we've come expect. In fact, the mutation rate approaches 100% per generation for alleles with more than 60 repeats. Individuals with 60 to 200 repeats are said to have a premutation: no symptoms, but an unstable allele that usually expands through passage to the next generation. The number of repeats often explodes in progeny of mothers with greater than 90 repeats.

SQ8. What molecular mechanism can you suggest to explain the frequent increase in the number of repeating units?

VI. Repair of Mutation (pp.472-478)

Base substitutions, insertions, deletions... the world is a dangerous place! How can our DNA deal with the ill effects of mutation? The best defense is to avoid mutation in the first place, and the cell is pretty good at doing this. DNA polymerase is accurate to start with, but much of this low error rate in *E. coli* is attributable to the ability of the primary DNA polymerase (polymerase III) to sense mistakes and back up to correct them (described in the text on pp. 300-301).

But errors do occur, what then? If you consider that perpetuating its DNA intact is arguably the prime reason for being of an organism, then it comes as no surprise that evolution has produced a bewildering repertoire of methods by which organisms deal with the prospect of mutation. We'll look at a few examples.

VI.A. <u>Direct repair of damaged bases</u>

The best solution way to fix an altered base is to change it back to what it was before. Fig. 17-12 shows how the enzyme DNA photolyase recognizes the pyrimidine dimers created by UV radiation and reversed the crosslinking. Other enzymes exist that can recognize the most common modifications to bases. Unfortunately, sometimes the damage is irreversible, and in any event there are too many possible ways bases can be altered to have an enzyme for each one. Other repair mechanisms are necessary.

VI.B. <u>Bypass Replication</u> (or *Better Read than Dead)*

For a single-cell organism, the worst case is that the mutation goes undetected by the organism, and a base change takes place, but in the case of mutagens like UV, this does not occur. The lesion is so severe that it cannot help but be noticed. This is because the error correcting DNA polymerase that replicates the genome cannot make sense out of the pyrimidine dimers crosslinked by UV or the bulky nucleotide derivatives produced by many other mutagens. But stopping DNA replication is a recipe for death, since a broken chromosome is as good as none at all, and the organism must fix the chromosome or suffer genetic death. The solution is to use an alternative DNA polymerase that is less fastidious about faithful replication (Fig. 6). When DNA Polymerase I encounters the incomprehensible bases, it puts something in and proceeds. The normal DNA polymerase would keep backing up trying vainly to correct the error, forever.

Pyrimidine dimer (*TT*) stops replication by DNA Polymerase III

```
5'-AGATTGACCTAGAT
3'-TCTAACTGGATCTATTGTAACAATGACTGGCGTATAG
```

Lesion is bypassed by error-prone DNA Polymerase I

```
5'-AGATTGACCTAGATAACATTGTTACTGACCGCATATC
3'-TCTAACTGGATCTATTGTAACAATGACTGGCGTATAG
```

Process may lead to error

```
5'-AGATTGACCTAGATAACATTGTTACTGACCGCATATC
3'-TCTAACTGGATCTATCGTAACAATGACTGGCGTATAG
```

Fig. 6: Bypass response to lethal DNA lesions. Ultraviolet light chemically alters two adjacent thymidine residues, linking them together. The resulting thymidine dimer on the template strand blocks replication by DNA polymerase III. Replication is resumed by DNA polymerase I, which usually guesses to place adenosines across lesions. If the lesion were a thymidine linked to a cytosine, the practice of blindly inserting adenosines would lead to mutation.

Will this process lead to mutation? Sometimes yes, sometimes no. What base would you insert if forced to read over a lesion that stops conventional DNA polymerase? You couldn't do much better than what actually occurs. Since UV often affects thymines, the error-prone DNA polymerase makes the best guess it can, generally putting adenines where it can't read. That's great when the lesion was a T-T dimer, not so good when it isn't. Bypass replication, therefore, sometimes leads to mutation and sometimes doesn't, but either way, the results to a single cell organism are preferable to death. As we'll see later, death may be the preferable fate for a cell in a multicellular organism

SQ9. Bypass replication across pyrimidine dimers generally produces which kind of base substitution: transition or transversion?

SQ10. *E. coli* mutants lacking DNA Polymerase I is perfectly viable under normal growth conditions. Furthermore, the mutants have a much lower mutation rate than the wild type. If they're viable and don't mutate as much, then why does *E. coli* bother to have the enzyme in the first place?

VI.C. Excision Repair

The organism can do better than this if it can recognize the lesion in the DNA and cuts it out. Fig. 17-13 shows how the general excision repair pathway can recognize a gross defect in DNA and correct it. If the damage suffered by DNA produces a gross distortion, the affected region is targeted for excision. DNA on either side of the lesion is nicked, and the resulting gap is filled in with the repair DNA polymerase, DNA Polymerase I. The figure shows how this process can be used to replace DNA carrying a thymidine dimer with wholesome DNA in which the adjacent thymidines are not connected to each other.

This is a great mechanism, and cells have extended it to those situation where the lesion is not so obvious as to be recognized by the general excision repair system. Cells possess a large number of enzymes (DNA glycolases) specialized to detect specific kinds of damaged bases (even when they do not distort the double helix) and to remove them (Fig. 17-14). With the base gone, the DNA has a defect -- the absence of a purine or pyrimidine -- that the cell is adept in handling. The strategy is to funnel diverse lesions into a common AP pathway ("AP" stands for "apurinic" or "apyrimidinic", i.e., the absence of a purine or pyrimidine).

SQ11. Consider two mutations: one in Ura-DNA glycosylase, the enzyme that detects and removes uracils from DNA, and one in AP endonuclease, which detects and excises apurinic/apyrimidinic residues. If the two mutations are combined into the same strain, which phenotype is epistatic over the other (i.e., does the resulting strain have the same phenotype as the Ura-DNA glycosylase-minus mutant or the AP endonuclease mutant)?

VI.D. Mismatch Repair

When the lesion is a simple base substitution that occurs during replication, you can see that the cell has what seems to be an impossible task. The substituted base is G, A, T, or C, so there is no DNA glycolase to remove it. Furthermore, the general excision repair system has no way of determining which is the wrong base. With nothing to go on but the mismatch itself, how would you tell which base to excise?

Fig. 17-16 illustrates the problem and the solution. A thymidine has been put where a cytosine ought to go (perhaps by a transient tautomerization by guanidine on the other strand). The region containing the wrong base is excised, and the affected strand is resynthesized to give DNA polymerase a second chance. All well and good, but how did the system know which strand to work on? Isn't it just as likely that the G is in error rather than the T? You'd think that half the time the system would guess wrong and excise the good base, leaving the mutation intact.

Russian roulette with two chambers is not a good bet, and organisms have developed means to better the odds. What can organisms see in this situation that you can't? The trick, in *E. coli* at least, is for newly synthesized DNA to be a different color than DNA of the parental strand. Now the excision system knows which strand to chop: always the new strand.

What's the basis of this trick? *E. coli*, like almost all organisms, use the conventional nucleotides GTP, ATP, TTP, and CTP to make its DNA. However, after the DNA is made, it marks it by modifying some adenines with methyl groups. Fig. 7 shows how *E. coli* methylates certain adenine residues, those in the sequence GATC. This modification does not affect basepairing, but the excision repair system can use the methyl group to recognize the DNA. Recall that newly made DNA is unmethylated, since it is made with unmethylated ATP. It takes *E. coli* about 10 minutes for the new DNA to be methylated. During this 10 minute window, the newly synthesized DNA strand is different from the old template strand, because only the old strand is

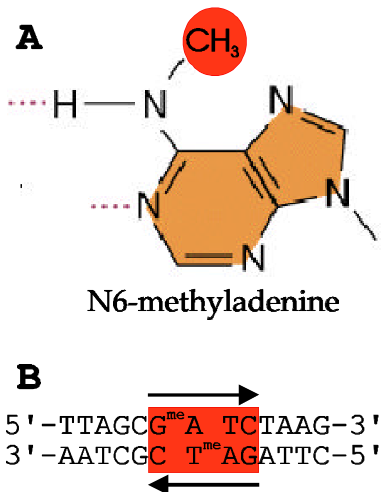

Fig. 7. Methylation of adenine within GATC. (A) Structure of methylated adenine, differing from adenine only in the presence of the methyl group (in red) attached to the amine. Note that the hydrogen bonding properties are not affected. **(B)** Sequence of DNA with GATC site. The adenine within GATC on both strands are methylated, but adenines elsewhere in the sequence are not affected.

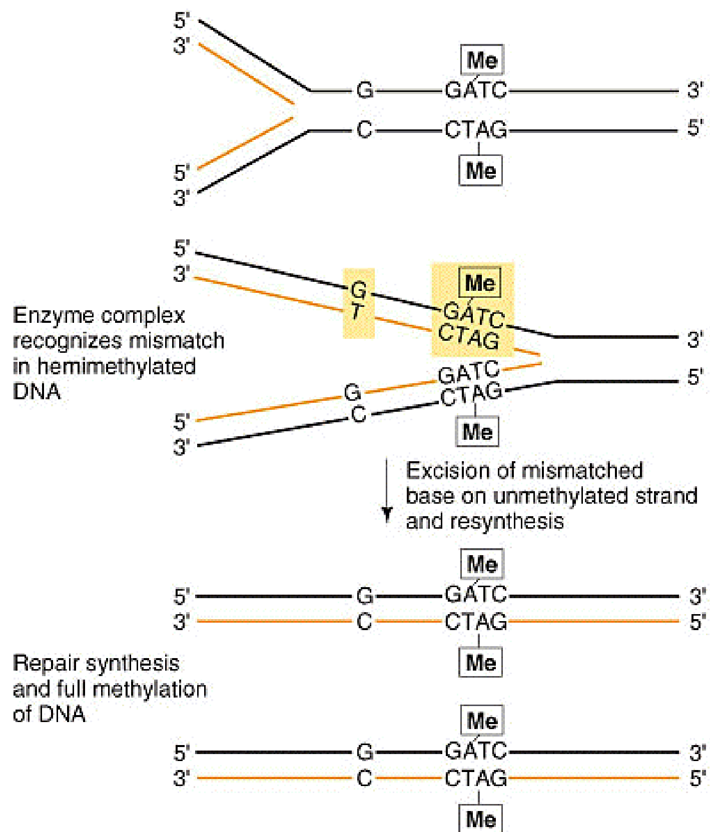

Fig. 8: Methylation-directed mismatch repair. **See notes for explanation.**(From Griffiths et al (1996), *An Introduction to Genetic Analysis*, 6th edition. WH Freeman, New York]

methylated. This difference enables the mismatch repair system to make the right decision

Fig. 8 shows how the difference in methylation between the two strands is used to direct the mismatch repair system to remove the appropriate base from a mismatch. Initially, both DNA strands are methylated, but as replication proceeds, newly replicated DNA is methylated only on the old strand. If the wrong base is put in, the mismatch is recognized by a slight bulge in the double helix, and the base on the NONmethylated strand is replaced. Eventually, the new strand is methylated, and replication can proceed again.

The process works by means of a multiprotein complex (Fig. 17-16). Protein MutS detects the mismatch that was introduced during replication but cannot by itself tell which base is correct and which is wrong. MutH recognizes and cuts a nearby unmethylated GATC site (only the unmethylated strand), and the nick in the DNA is used as a starting point to degrade DNA of the strand past the mismatch. The gap can now be filled in by DNA Polymerase I. Although humans don't have the same mismatch repair system as *E. coli*, human proteins similar to MutS and MutL are clearly

important in avoiding mutation. Defects in these proteins predispose the affected person to colon cancer.

SQ12. What would be the phenotype of *E. coli* that lacks the ability to methylate GATC sequences?

SQ13. What would be the phenotype of *E. coli* that overmethylates GATC sequences (i.e., has such a high activity of the methylating enzyme that the sequences are methylated almost immediately after DNA replication)?