Microsatellites

Evolution and Applications

Edited by

David B. Goldstein
Department of Zoology,
University of Oxford

and

Christian Schlötterer
Institut fur Tierzucht und Genetik
Veterinämedizinsche Universität
Wien

OXFORD
UNIVERSITY PRESS

4 Mechanistic basis for microsatellite instability

Jonathan A. Eisen

Chapter contents

Abstract

The inherent instability of microsatellite loci makes them exceptionally useful for evolutionary and genetic studies. This instability is predominantly due to changes in the number of copies of the microsatellite repeat. Most copy number changes at microsatellites are caused by slip-strand mispairing errors during DNA replication. Some of these errors are corrected by exonucleolytic proofreading and mismatch repair, but many escape repair and become mutations. Thus microsatellite instability can be considered to be a balance between the generation of replication errors by slip-strand mispairing and the correction of some of these errors by exonucleolytic proofreading and mismatch repair. The factors that cause this process to occur much more frequently in microsatellites that in non-repeat containing DNA are discussed. However, not all microsatellites are equally unstable because not all are equally prone to this mutation process. The mechanisms by which a variety of factors cause this variation in stability among microsatellites are discussed.

4.1 Introduction

The characteristic that makes loci that contain microsatellite repeats particularly useful for evolutionary and genetic studies is their inherent instability. The mutation rates at most microsatellite loci are usually orders of magnitude higher than mutation rates at other loci within the same genome. Although many types of mutations occur at microsatellite loci, the elevated mutation rate is primarily caused by an elevated rate of one particular class of mutations: changes in the length of the repeat tract. Thus the term 'microsatellite instability' is frequently used to refer specifically to these tract-length changes. Since most of these tract-length changes result from changes in the integral number of copies of the repeat, they are also frequently referred to as copy number changes.

Ever since it was recognized that microsatellites are so prone to changes in tract length, researchers have been trying to determine why. A variety of approaches have been useful for this purpose. Evolutionary and population genetic comparisons have been used to document the patterns of tract-length variation at microsatellites, and to test the robustness of different mutation models when averaged over long time-scales. Biochemical experiments with purified proteins or cell extracts have been used to characterize each step in the mutation process, and to determine the factors that control that step. Genetic studies have given insight into the genes that control microsatellite stability, and have allowed the accurate quantification of the stability of different microsatellites in controlled genetic backgrounds. Only by combining the results of these different types of studies has the mechanism of the mutation process become well characterized. Since evolutionary studies of the mutation mechanism are described in detail elsewhere in this book, I focus here on the biochemical and genetic studies.

To have a complete understanding of the mechanism of microsatellite instability one must also explain why stability varies both within and between species. Clues to the cause of this variation have come from the identification of factors that correlate with the level of microsatellite stability. Such factors include size of the repeat unit, number of copies of the repeat, presence of variant repeats, and amount of transcription in the region of DNA containing the repeat. Many studies that use data on microsatellite variation use models of the mutation process to enhance the analysis being done. Such studies should be improved by a better understanding of the mechanism underlying microsatellite instability as well as the causes of differences in stability among microsatellites. In this chapter, I summarize what is known about the mechanism underlying microsatellite instability and discuss some of the factors that cause variation in stability within and between species.

4.2 Microsatellite mutation models

The central debate about the mechanism of microsatellite instability has focused on two competing but not necessarily mutually exclusive models. One model proposes that microsatellite instability is caused by an elevated rate of unequal crossing-over (UCO) within microsatellite repeats. Unequal crossing-over is the result of recombination between homologous chromosomes that are imperfectly aligned. The UCO microsatellite instability model suggests that UCO occurs at an elevated rate in microsatellites because the presence of repeats increases the likelihood of misalignment between homologues. A similar proposal has been made to explain the high rates of copy number changes observed in tandemly repeated genes (Smith 1973). The alternative model proposes that microsatellite instability is caused by an elevated rate of slip-strand mispairing (SSM) errors during DNA replication. The SSM process, which was first proposed to explain frameshift mutations in any type of DNA (Fresco and Alberts 1960), begins with the DNA polymerase 'slipping' during replication, causing the template and newly replicated strands to become temporarily unaligned. For replication to continue, the strands must realign. Mutations will be generated if this realignment is imperfect. The SSM microsatellite instability model proposes that SSM occurs at an elevated rate in microsatellites because the presence of repeats increases the likelihood of misalignment after slippage (since repeats can easily be looped out of the DNA double-helix) (Streisinger *et al.* 1966).

The results of many studies indicate that an elevated rate of SSM is the main cause of microsatellite instability. The key evidence that supports the SSM model against the UCO model is summarized below (see Sia *et al.* 1997*a*, for a review):

- Microsatellite stability is unaffected by defects in genes with major roles in recombination, such as *recA* in *Escherichia coli* (Levinson and Gutman 1987*b*), and *rad52* in *Saccharomyces cerevisiae* (Henderson and Petes 1992). This suggests against the UCO model since mutations are dependent on recombination in this model.
- In humans, copy number changes at microsatellites can be generated without exchange of flanking genetic markers (and thus probably without recombination) (Morral *et al.* 1991).
- In *S. cerevisiae*, microsatellite stability is similar in mitotic and meiotic cells (Strand *et al.* 1993). Since recombination occurs more frequently in meiosis than mitosis, if the UCO model were correct, microsatellites should be more unstable during meiosis.
- Microsatellite stability is reduced by defects in genes involved in DNA replication error correction pathways. This is consistent with the SSM model since this model requires DNA replication to occur. In addition, genetic and biochemical experiments show that these error-correction pathways can recognize and repair the types of DNA loops that would be created by SSM (Bishop *et al.* 1989; Parker and Marinus 1992).

- The orientation of a microsatellite relative to the leading and lagging strands of replication influences its stability (Freudenreich *et al.* 1997). This is not expected by UCO model but is consistent with the SSM model since the leading and lagging strands have somewhat different mechanisms of replication.

These and other results show that SSM is an integral component of the mutation process leading to microsatellite instability. However, SSM alone does not provide a full picture of this mutation process. As suggested above, not all SSM errors become mutations—some are 'repaired' by error-correction mechanisms. The two error-correction pathways that have been shown to be important in repairing SSM errors are exonucleolytic proofreading and post-replication mismatch repair. Thus a complete description of the mutation process must include both the generation of replication errors by SSM and the correction of some of these errors by mismatch repair and proofreading (see Fig. 4.1). In the following sections I discuss each of the steps in the microsatellite instability mutation process, providing some details about the mechanism of each step and the methods used to study those mechanisms. In addition, I discuss how variability in each step contributes to variation in microsatellite stability within and between species (see Table 4.1).

4.3 Slip-strand mispairing replication errors

To study the mechanism of the SSM process, one must functionally isolate SSM from the downstream error-correction steps. One approach to achieve such functional separation is to study the replication of DNA *in vitro* (Kunkel 1986, 1990; Schlötterer and Tautz 1992). *In vitro* studies allow straightforward comparisons of replication errors by different polymerases, as well as comparisons of errors by the same polymerase using different templates. However, *in vitro* studies are limited because they may not accurately reflect what occurs during intracellular replication conditions. To study SSM errors *in vivo*, researchers have used strains with defects in either exonucleolytic proofreading or mismatch repair or both. In such strains, since SSM errors are not corrected, SSM error rates and patterns can be inferred directly from observed mutations (e.g., Wierdl *et al.* 1997). Results from many such *in vitro* and *in vivo* studies show that the SSM process can be subdivided into three distinct steps: slippage of the DNA polymerase during replication, mis-realignment of the template and newly replicated DNA strands, and continuation of replication from a misaligned template (see Fig. 4.1).

These studies confirm the prediction of the SSM model that SSM errors are more likely to occur in microsatellite repeats than in 'normal' DNA. However it has not been determined which step of SSM is most affected by the presence of repeats: slippage, misalignment, or extension. It is almost certain that misalignment is more common in repeat regions than in 'normal' DNA. Loops generated by misalignment will be more stable in microsatellites than in non-repeat regions since base pairing is not significantly changed when one or more copies of a

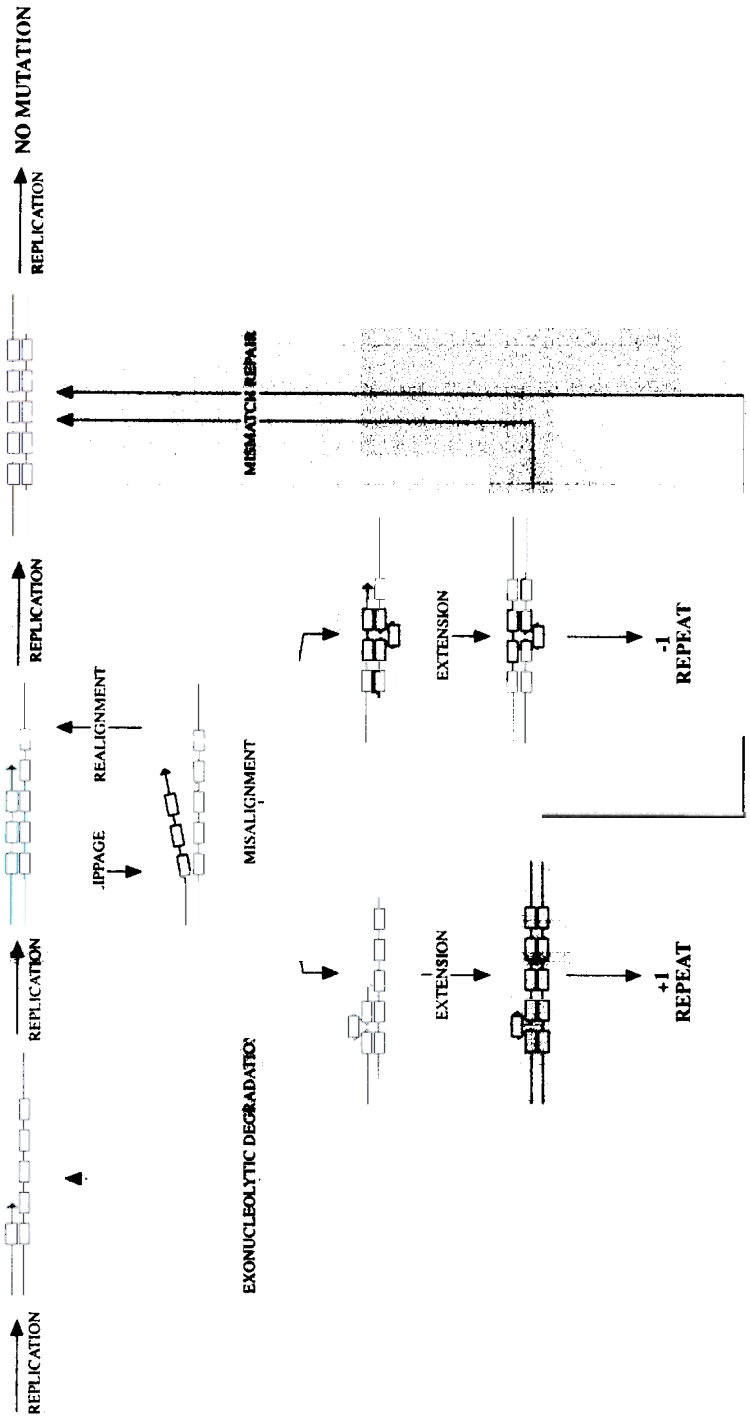

Fig. 4.1 Model of the mutation process at microsatellite loci. Cartoons of double-stranded DNA containing a microsatellite repeat are shown at different stages of the replication and mutation process. In the cartoons, DNA strands are represented by thin lines, microsatellite repeats by small boxes, and ongoing replication by small arrows. Flow arrows point down for steps that lead to mutations, up for steps that prevent mutations from occurring, and to the right for steps in the DNA replication process. The exonuclease step is shown with a dashed line since it has only a limited role in regulating microsatellite mutations. Details about each step are provided in the main text.

repeat are in a loop (see Fig. 4.1). However, there is also reason to believe that slippage occurs more frequently in microsatellite repeats than in normal DNA. *In vitro* studies show that DNA containing microsatellite repeats is particularly prone to the formation of unusual DNA structures. Such structures likely interfere with the replication process, which could lead to slippage by the polymerase (Kang *et al.* 1995; Samadashwily *et al.* 1997). Thus, the elevated SSM rates at microsatellites relative to normal DNA may be caused by an increased likelihood of both slippage and misalignment.

4.3.1 SSM variation: effects of the nature of the microsatellite

Although in general SSM errors are more frequent in microsatellite-containing regions than in other regions of the genome, the rate and type of such errors are not equal for all microsatellites. The nature of the microsatellite itself has a large impact on SSM. For example, the likelihood of SSM for a particular microsatellite is correlated with the number of copies of the repeat. The most detailed study of this copy number effect is that of Wierdl *et al.* (1997), in which the stability of five microsatellites with different numbers of copies of a GT repeat was analysed. The mutation *rate* was found to increase with more repeats (as is expected since there are more places to slip and misalign) but the increase was greater than expected (more than two orders of magnitude between loci with 7.5 and 52.5 repeats). The *types* of mutations also differed between the microsatellites with different numbers of the repeat. The long tracts (those with more repeats) were more likely to have large, multi-repeat deletions than short tracts. In addition, the mutations that resulted in single-repeat changes (plus or minus one repeat) were different between long and short tracts. The single copy changes in long tracts were mostly additions while those in short tracts included roughly equal numbers of additions and deletions. Wierdl *et al.* showed that these copy number effects were not due to biases in mismatch repair, since the effects were seen in mismatch repair mutants. Therefore, they concluded that the copy number effects were probably caused by differences in SSM between microsatellites with different numbers of repeats. However, they were not able to determine the step of SSM that was influenced by copy number. One possibility is that the unusual DNA structures, discussed above as a potential cause of increased slippage in microsatellite repeats, may be even more likely to occur as the number of repeats increases. Regardless of the exact mechanism, the details of the effects of copy number on SSM (and thus on microsatellite stability) help to explain why the number of repeats at a particular microsatellite is somewhat stable over evolutionary time. Long tracts may be biased towards getting shorter (due to the large deletions) and short tracts may be biased towards getting longer (because of a slight bias in additions over deletions). An effect of copy number may also explain why certain microsatellites (e.g. those associated with some human diseases) become particularly unstable after they cross a threshold number of copies of the repeat (see chapter by Rubinsztein, this volume).

'able 4.1

Step in mutation process affected by factor	Nature of the microsatellite				
	Repeat unit size	Number of repeats	Type of repeat	Variant repeats	Replication orientation
SSM (any step)	±				
Replication slippage	±				
Misalignment	±	±			
Extension[5]		?			
Exonuclease					
Mismatch repair					

Another aspect of the microsatellite that influences the likelihood of SSM is the presence of variant repeats. Evolutionary and genetic studies have shown that the presence of variant repeats is correlated with the stability of a microsatellite (e.g. Goldstein and Clark 1995). Petes *et al.* (1997) have studied this effect in controlled laboratory conditions in *S. cerevisiae* to try to determine the underlying mechanism. This study showed that the presence of variant repeats leads to an approximately fivefold stabilization of GT repeats. Since this stabilizing effect was also seen in mismatch-repair mutants, the authors suggested that the variant repeats exerted their effect by reducing the likelihood of SSM errors. However, as with the copy-number effect described above, it has not been possible to determine what step of SSM was most 'stabilized' by variant repeats.

4.3.2 SSM variation: effects of external factors

There are many reasons to believe that external factors (i.e. factors other than characteristics of the microsatellite) can influence SSM error rates and patterns. For example, base misincorporation error rates and patterns are influenced by many external factors. Since base misincorporation and SSM are both forms of polymerase error, it is likely that these factors will also influence the SSM process. External factors that influence misincorporation errors include local DNA sequence (e.g. the GC content or the ability to form secondary structures), genome position (e.g. proximity to replication origins or chromosome ends), and even the chromosome in which a sequence is found (e.g. nuclear, organellar, plasmid) (Wolfe *et al.* 1987, 1989; Kunkel 1992; Hess *et al.* 1994). In addition, misincorporation error rates are dependent on many conditional factors including methylation state, amount of chromosome packaging, temperature, phase of the cell cycle during which a particular section of DNA is replicated, and amount of DNA damage and repair prior to replication. Future studies of microsatellite mutation mechanisms would benefit by examining whether some of these factors influence SSM errors.

Table 4.1

Neighbouring DNA[1]		Cellular conditions				Organismal differences	
GC content	Sequence content	Transcription	Methylation state	Cell cycle stage	Pathway used[2]	Pathway presence[3]	Pathway biases[4]
+	+	±		+	+	?	±
±	±	±		±	+	?	±
±	±			±		?	
±	±			±	+	?	
+	+			±	+	+	
+	+			+	+	+	

4.3.3 SSM variation: differences between individuals or species

Although the SSM mechanism and its role in causing microsatellite instability is conserved between species, it is likely that the specific rates and patterns of SSM differ greatly between species. For example, polymerases from different species have significantly different base misincorporation error rates (Kunkel 1992) and thus likely also have different SSM rates at microsatellites. In addition, many of the factors described above as influencing SSM errors within a species differ greatly between species (e.g. GC content, temperature, methylation). Thus it remains to be seen whether all species are affected by copy number and variant repeats in the same ways as described above.

4.4 Correcting SSM errors I: exonucleolytic proofreading

Exonucleolytic proofreading is a process in which DNA that has been recently synthesized is examined for errors made by the DNA polymerase. If errors are found, the exonuclease will degrade the newly replicated strand, the DNA polymerase will back up, and the strand will be recopied. Thus many errors made by the DNA polymerase will not become mutations because they will be 'erased' by proofreading. Proofreading was originally characterized for its role in limiting mutations due to base misincorporation errors. The role of proofreading in regulating microsatellite stability has been determined by methods that are similar to those used to study SSM. *In vitro* studies have been used to compare the

error rates and patterns of polymerases with and without associated exonucleases, and to determine the types of substrates that the proofreading exonucleases will degrade. *In vivo* studies have allowed the determination of errors with and without exonucleases under realistic cellular conditions. In such *in vivo* studies, it has been helpful to use strains with defects in mismatch repair, so that the role of the proofreading step is clear.

Studies such as the ones described above have shown that proofreading is involved in regulating the stability of microsatellites, but the extent of this role is limited in two ways. First, proofreading only significantly influences the stability of a subset of microsatellites: those with both small unit size (mostly mono- and di-nucleotide repeats) (Kroutil *et al.* 1996; Sia *et al.* 1997*a*; Strauss *et al.* 1997) and few copies of the repeat (Streisinger and Owen 1985; Kroutil *et al.* 1996; Tran *et al.* 1997). In addition, even for this subset of microsatellites, the impact of proofreading is limited—the stability of such microsatellites only decreases by about five- to tenfold in exonuclease mutants.

The details of the mechanism of proofreading help to explain why this process has only a limited role in regulating microsatellite stability (for reviews see Echols and Goodman 1991; Kunkel 1992). Proofreading exonucleases detect errors by monitoring the DNA that has just been replicated to determine whether it forms normal double-helical DNA structures with the template strand. Abnormal DNA structures trigger the exonuclease activity. This is how proofreading prevents many base misincorporation errors from becoming mutations. A base misincorporation error will lead to a base : base mismatch between the newly replicated and template DNA strands and many such mismatches will be recognized by proofreading exonucleases. However, proofreading exonucleases are only able to monitor the DNA within a few bases of the active site of the polymerase. This proximity effect explains why proofreading has at most a small impact on microsatellite stability. Most loops generated by SSM will be too far from the replication fork to be recognized by proofreading exonucleases. The lack of a role of exonucleases in repairing most SSM errors at microsatellites helps to explain the high rate of microsatellite copy-number changes relative to point mutation rates.

4.4.1 Proofreading variation

While the impact of proofreading on microsatellite stability is limited, variation in proofreading can account for some of the variation in stability of microsatellites. As with SSM, the nature of the microsatellite has a profound impact on proofreading. The best example of this was described above—proofreading only works on microsatellites that are short and in which the repeat unit size is small. The mechanism of both of these biases is directly related to the proximity effect described above. As the number of copies of a repeat increases, the impact of proofreading decreases because those loops that are generated by SSM will be even more likely to be far from the replication fork. In addition, in microsatellites with repeats of large unit size (e.g. 5 bp repeats), a loop just one repeat away from the replication fork may be too far away to be proofread (the base pairing of one repeat may be enough to stabilize the DNA structure at the fork). Proofreading is also likely to be affected by many external factors. For example, the efficiency of some exonucleases is affected by both GC content and sequence context (Kunkel 1992). Thus the sequence around a mononucleotide repeat may influence its mutation rate by altering the efficiency of proofreading. Finally, the impact of proofreading on microsatellite stability is also likely to vary greatly between species. For example, some species do not even have proofreading exonucleases associated with their DNA polymerases. Microsatellites with short mono- and dinucleotide tracts should be more unstable in species without proofreading than in species with proofreading.

4.5 Correcting SSM errors II: mismatch repair

Mismatch repair was named based on its role in recognizing and repairing base : base mismatches that arise due to base misincorporation errors. It is now clear that the same process can repair DNA containing loops such as those generated by SSM at a microsatellite (see Fig. 4.1). Mismatch repair has a much more significant impact on microsatellite stability than proofreading. Defects in mismatch repair can cause microsatellite instability to increase by many orders of magnitude (see below for more details). Since mismatch repair plays such a key role in regulating microsatellite stability, differences in the repair of loops by mismatch repair could account for a great deal of the variation in microsatellite stability within and between species.

Before discussing the specifics of loop repair and how it varies within and between species, it is useful to review some details about the general mechanism of mismatch repair. Mismatch repair has been found in a variety of species from bacteria to humans. It has been characterized in the most detail in *E. coli*. In the other species in which it has been characterized, the overall scheme of mismatch repair works in much the same way as in *E. coli*. Thus the *E. coli* system has served as a useful model for mismatch repair in all species. The first critical step in mismatch repair in *E. coli* is the recognition of mismatched DNA by the MutS protein (see Modrich 1991 for a review). Specifically, a dimer of MutS (two MutS proteins bound together) binds to the site of a mismatch in double-stranded DNA. Subsequently, through an interaction between the MutS dimer, a dimer of the MutL protein, and a single MutH protein, a section of one of the DNA strands at that location is targeted for removal. Other proteins complete the repair process: the section of DNA that has been targeted is removed and degraded, a patch is synthesized using the complementary strand as a template, and the patch is ligated into place resulting in a repaired section of double-stranded DNA without mismatches.

The evidence that mismatch repair is involved in repairing SSM errors at microsatellites comes from three types of studies. First, defects in mismatch

repair cause decreases in microsatellite stability (anywhere from 10- to 5000-fold, depending on the species and the microsatellite). In addition, when DNA containing loops is transformed into cells, the loops can be repaired, but only if the cells have functional mismatch repair (Dohet *et al.* 1986; Bishop *et al.* 1989; Parker and Marinus 1992). Finally, *in vitro* studies have shown that repair of loops can be carried out by purified mismatch repair proteins (Learn and Grafstrom 1989; Parker and Marinus 1992). Each of these results has been found in a variety of species, showing that the role of mismatch repair in repairing loops at microsatellites is highly conserved. Incidentally, this is what led to the discovery that mismatch repair genes are defective in hereditary non-polyposis colon cancer in humans—cells from patients with this disease showed high levels of microsatellite instability. In summary, these studies show that the repair of loops is very similar to the repair of mismatches.

4.5.1 Mismatch repair variation: effects of the nature of the microsatellites

Perhaps the most important cause of variation in mismatch repair is the nature of the microsatellite. Loops are not all recognized equally by mismatch repair system, and this specificity varies between species. One factor that is very important to the recognition step is the size of loop. For example, in *E. coli*, transformation studies have shown that loops of 1–3 bases are repaired well, those of 4 bases are repaired poorly, and those greater than 4 bases are not repaired at all. *In vitro* studies of purified mismatch repair proteins show that this is due to inability of MutS to recognize loops larger than 4 bases in size (Learn and Grafstrom 1989; Parker and Marinus 1992). Thus in *E. coli*, microsatellites in which the repeat unit size is 4 bp or greater have especially high rates of instability, since SSM errors in such regions are not repaired well. Mismatch recognition is also biased by loop size in many other species, although the specific size preferences are not completely conserved. For example, the yeast mismatch repair system appears to be able to recognize and repair loops up to 6 bp well (and possibly even up to 14 bp, although this has not been confirmed). More details about the mechanism causing the different size preference are given in the section on variation in mismatch repair between species. For the purposes of the discussion here, all that is important is that in many species the size limits of loop recognition help to explain why microsatellites with different repeat unit sizes have different mutation rates.

The size specificity of loop recognition also helps to explain variation in mutation patterns between microsatellites with different sized repeats. For example, in *S. cerevisiae*, the majority of mutations in mononucleotide repeats are additions or deletions of one repeat (i.e. plus or minus 1 bp). However, the majority of mutations at microsatellites with 5 bp repeats are additions or deletions of two or more repeats (Sia *et al.* 1997*b*). To understand this phenomenon, it is important to recognize that the mutation rate and pattern for a microsatellite is determined by a combination of the rate and type of SSM errors and how well these errors are repaired. Thus a particular mutation may occur at a high rate either because it is a common SSM error or because it is repaired poorly. For the mononucleotide repeat described above, most SSM errors are repaired about equally well (errors involving even five repeats at a time can be repaired by mismatch repair). Thus the most common mutations are those that are the most common SSM errors. In contrast, for the microsatellite with the 5 bp repeat, mismatch repair will only repair single repeat changes. Thus, although SSM errors involving two or more repeats are not very frequent, most of the mutations are changes in two or more repeats because many of the single repeat changes are repaired. The size dependence of mismatch repair also explains why 20 bp repeats are so unstable in *S. cerevisiae* (Sia *et al.* 1997*b*); mismatch repair will not recognize any SSM error involving such a large repeat. Since both the number of repeats and the size of the repeat influence microsatellite stability, it is important to compare repeats of the same unit size when studying copy-number effects, and repeats with the same number of copies when studying unit size effects.

One aspect of loop repair that has been poorly studied is the role of the type of microsatellite (e.g. GT vs. GA repeats). Since base : base mismatch repair is not uniform for all mismatches (e.g., C : C mismatches are not repaired well in many species), it is likely that loop repair will also not be uniform. Since most of the studies of microsatellite mutation mechanisms have been done on limited types of microsatellites, it will be important to determine if the results of these studies are universal to all types of repeats.

4.5.2 Mismatch repair variation: effects of external factors

As with SSM and proofreading, many factors in addition to the nature of the microsatellite itself can influence the effectiveness of mismatch repair. For example, the location of the mismatch within the genome is important. In *S. cerevisiae*, loop recognition appears to be biased between loops on the template versus nascent strand of replication. For loops including a single repeat, mismatch repair appears to repair preferentially those that are on the template strand, resulting in a bias towards single repeat additions. The exact mechanism of this strand bias is not known although some of the genes involved have been identified (Sia *et al.* 1997*b*). Another effect of location is whether the mismatch is in nuclear or organellar DNA. Although organellar mismatch repair has not been characterized in detail, it is likely to be quite different from nuclear mismatch repair. The surrounding DNA also influences mismatch repair. For example, studies of base : base mismatches have shown that mismatch recognition is affected by sequence context (Cheng *et al.* 1992), and by GC content (Jones *et al.* 1987). It is likely that the recognition of loops will also be affected by these factors. Finally, mismatch repair can also be influenced by conditional factors including the presence of strand-recognition signals, methylation state, and level of transcription.

4.5.3 Mismatch repair variation: differences within a species

Differences in mismatch repair among individuals of a particular species have been well documented. For example, many strains of *E. coli* in the 'wild' are defective in mismatch repair (LeClerc *et al.* 1996; Matic *et al.* 1997). Since there are adaptive benefits to having modest increases in mutation rates in certain circumstances (Taddei *et al.* 1997*a*,*b*), and since one way to alter mutation rates is by altering mismatch repair, many strains may be found to have defects in mismatch repair. Also, since mismatch recognition is involved in other cellular processes such as the regulation of interspecies recombination, there may be other selective pressures that lead to variation in mismatch repair capabilities within a species. Finally, since organisms appear to be able to turn mismatch repair on and off in certain situations (Harris *et al.* 1997; Macintyre *et al.* 1997; Torkelson *et al.* 1997), environmental conditions may play a major role in determining mismatch repair capabilities.

4.5.4 Mismatch repair variation: differences between species

Although mismatch repair is a highly conserved process, there are many ways in which it varies between species. For example, the mismatch recognition process is not completely conserved between bacteria and eukaryotes. The best characterized eukaryotic mismatch repair system is that of *S. cerevisiae*. As suggested above, the general mechanism of *S. cerevisiae* mismatch repair is very similar to that of *E. coli* (see Kolodner 1996 for a review). In particular, the role of the MutS and MutL proteins is highly conserved—*S. cerevisiae* uses homologues of these proteins in essentially the same way that they are used in *E. coli*. Even the use of the proteins as dimers is conserved. However, unlike *E. coli*, *S. cerevisiae* uses multiple homologues of both MutS and MutL for mismatch repair. These multiple homologues are used to make separate mismatch repair complexes with unique and distinct functions. The specificity of each of these complexes is determined almost entirely by its particular combination of MutS homologues (which are referred to as MSH proteins for *MutS H*omologue). For mismatch repair of nuclear DNA there are two recognition complexes: an MSH2–MSH6 heterodimer for recognizing and repairing base : base mismatches and loops of 1–2 bases, and an MSH2–MSH3 heterodimer for recognizing and repairing loops of 1–6 bases (and possibly even up to 14 bases—see Sia *et al.* 1997*b*). Genetic studies suggest that there may also be a mitochondrial-specific mismatch repair complex. Defects in another MutS homologue, MSH1, cause increases in the mutation rates in mitochondrial DNA. However, the details of mitochondrial mismatch repair are not well understood. In particular, it is not known what role mismatch repair plays in microsatellite stability in mitochondrial DNA. Interestingly, *S. cerevisiae* encodes two additional MutS homologues (MSH4 and MSH5) that do not function in mismatch repair, but instead appear to use mismatch recognition to regulate meiotic crossing-over and chromosome segregation. The mismatch recognition process of other eukaryotes is highly similar to that of *S. cerevisiae* (Fishel and Wilson 1997). One of the results of the differences in mismatch repair between eukaryotes and *E. coli* is that eukaryotes can repair loops of larger sizes than *E. coli*. This explains why microsatellites with these larger sized repeats are more stable in eukaryotes than in *E. coli*.

Another major difference in mismatch repair between species is in the mechanism used to determine which strand is the recently replicated strand (and thus is the strand that contains the error). In *E. coli* the 'incorrect' strand is determined by its methylation state—the newly replicated strand is unmethylated and thus can be distinguished from the template strand. In some other species, strand recognition is thought to be based on the presence of nicks, which are more likely to occur on the newly replicated strand. In such species, there may be differences in mismatch repair efficiency between the leading and lagging strands, since nicks are more common on lagging strand.

Although the process of mismatch repair is highly conserved, some species may not have the process at all. For example, analysis of complete genome sequences shows that some bacterial and Archaeal species do not encode any likely MutS or MutL homologues (Eisen *et al.* 1997; Eisen 1998). It is likely that these species do not have any mismatch repair, since functional MutS and MutL homologues are absolutely essential to the mismatch repair process. Any species without mismatch repair should have significantly elevated levels of microsatellite instability. In addition, differences between species could arise from the number and types of MutS and MutL homologues that are present.

4.6 Additional factors that influence microsatellite stability

Although the studies of microsatellite mutation mechanisms have been extensive, there are still many factors that have been found to influence microsatellite stability but for which the mechanism of the effect is unknown. For example, Wierdl *et al.* (1996), following up previous studies (Datta and Jinks-Robertson 1995), showed that transcription leads to a 4–9-fold destabilization of a poly(GT) repeat. One explanation for this is that transcription will increase the likelihood of repair by the process of transcription-coupled repair and this process is mutagenic in some conditions (Wang *et al.* 1996). Alternatively, transcription could interfere with either mismatch repair or replication. Another unexplained observation is that microsatellites in the chromosome are usually more stable than identical microsatellites on a plasmid (Henderson and Petes 1992). Finally, many studies have shown that microsatellite stability is dependent on the orientation of the microsatellite within the DNA (Kang *et al.* 1995; Maurer *et al.* 1996; Freudenreich *et al.* 1997). For example, Freudenreich *et al.* showed that a microsatellite with 130 CTG repeats was more unstable when the CTG was on the lagging strand (Freudenreich *et al.* 1997). They suggested that this could be due to differences in the likelihood of slippage on the leading vs. lagging strand of replication. However, they could not rule out differences in mismatch repair, transcription,

proofreading or other factors between the strands as the explanation. An alternative explanation for the orientation effect is that loops may be better recognized on the nascent strand than on the template strand (Sia *et al.* 1997*b*). More detailed studies of microsatellite mutation mechanisms will likely sort out how these and factors influence microsatellite stability.

4.7 Conclusions and summary

The mutation process at microsatellites can be considered to be a balance between the generation of replication errors by slip-strand mispairing and the correction of some of these errors by exonucleolytic proofreading and mismatch repair. The mutation rate and pattern for a particular microsatellite will be determined by the rate and type of SSM errors, as well as by how well these errors are recognized and repaired by exonucleases and mismatch repair. The details of the mutation mechanism explain why microsatellites are so unstable. First, SSM occurs much more frequently in microsatellites than in normal DNA. In addition, exonucleolytic proofreading, which prevents a large proportion of base misincorporation errors from becoming mutations, has only a limited role in preventing SSM errors from becoming mutations. The details of the mutation mechanism also help to understand why microsatellite stability varies within and between species. For example, the high mutation rate of microsatellites with repeats of large unit size can be explained by the inability of mismatch repair to recognize SSM errors in such large repeats. In addition, the positive correlation between number of copies of a repeat and stability can be explained by an increased likelihood of SSM errors in microsatellites with more repeats. The details not only help to understand the mutation process causing microsatellite instability, but they can be used to improve models of microsatellite evolution. Just as better models of nucleotide substitution processes have improved the analysis of DNA sequence variation, better models of microsatellite instability should improve the analysis of copy number variation at microsatellite loci.

References

Bishop, D. K., J. Andersen, and R. D. Kolodner. 1989. Specificity of mismatch repair following transformation of *Saccharomyces cerevisiae* with heteroduplex plasmid DNA. *Proceedings of the National Academy of Sciences U.S.A.* **86**:3713-3717.

Cheng, J. W., S. H. Chou, and B. R. Reid. 1992. Base pairing geometry in GA mismatches depends entirely on the neighboring sequence. *Journal of Molecular Biology* **228**:1037-1041.

Datta, A., and S. Jinks-Robertson. 1995. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**:1616-1619.

Dohet, C., R. Wagner, and M. Radman. 1986. Methyl-directed repair of frameshift mutations in heteroduplex DNA. *Proceedings of the National Academy of Science U.S.A.* **83**:3395-3397.

Echols, H., and M. F. Goodman. 1991. Fidelity mechanisms in DNA replication. *Annual Review of Biochemistry* **60**:477-511.

Eisen, J. A. 1998. Evolution of the MutS family of proteins: gene duplication, loss, and functional divergence. *Submitted*

Eisen, J. A., D. Kaiser, and R. M. Myers. 1997. Gastrogenomic delights: a movable feast. *Nature (Medicine)* **3**:1076-1078.

Fishel, R., and T. Wilson. 1997. MutS homologs in mammalian cells. *Current Opinion in genetics and Development* **7**:105-113.

Fresco, J. R., and B. M. Alberts. 1960. The accomodation of noncomplemetary bases in helical polyribonucleotides and deoxyribonucleic acids. *Proceedings of the National Academy of Sciences U.S.A.* **85**:311-321.

Freudenreich, C. H., J. B. Stavenhagen, and V. A. Zakian. 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Molecular and Cellular Biology* **17**:2090-2098.

Goldstein, D. B., and A. G. Clark. 1995. Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Research* **23**:3882-3886.

Harris, R. S., G. Feng, K. J. Ross, R. Sidhu, C. Thulin, S. Longerich, S. K. Szigety, M. E. Winkler, and S. M. Rosenberg. 1997. Mismatch repair protein MutL becomes limiting during stationary-phase mutation. *Genes and Development* **11**:2426-2437.

Henderson, S. T., and T. D. Petes. 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **12**:2749-2757.

Hess, S. T., J. D. Blake, and R. D. Blake. 1994. Wide variation in neighbor-dependent substitution rates. *Journal of Molecular Biology* **236**:1022-1033.

Jones, M., R. Wagner, and M. Radman. 1987. Mismatch repair and recombination in *Escherichia coli*. *Cell* **50**:621-626.

Kang, S., A. Jaworski, K. Ohshima, and R. D. Wells. 1995. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genetics* **10**:213-218.

Kolodner, R. 1996. Biochemistry and genetics of eukaryotic mismatch repair. Genes and Development **10**:1433-1442.

Kroutil, L. C., K. Register, K. Bebenek, and T. A. Kunkel. 1996. Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry* **35**:1046-1053.

Kunkel, T. 1992. Biological asymmetries and the fidelity of eukaryotic DNA replication. *Bioessays* **14**:303-308.

Kunkel, T. A. 1986. Frameshift mutagenesis by eucaryotic DNA polymerases in vitro. *Journal of Biological Chemistry* **261**:13581-13587.

Kunkel, T. A. 1990. Misalignment-mediated DNA synthesis errors. Biochemistry **29**:8004-8011.

Kunkel, T. A. 1992. DNA replication fidelity. *Journal of Biological Chemistry* **267**:18251-18254.

Learn, B. A., and R. H. Grafstrom. 1989. Methyl-directed repair of frameshift heteroduplexes in cell extracts from *Escherichia coli*. *Journal of Bacteriology* **171**:6473-6481.

LeClerc, J. E., B. Li, W. L. Payne, and T. A. Cebula. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**:1208-1211.

Levinson, G., and G. A. Gutman. 1987. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Research* **15**:5323-5338.

Macintyre, G., K. M. Doiron, and C. G. Cupples. 1997. The Vsr endonuclease of *Escherichia coli*: an efficient DNA repair enzyme and a potent mutagen. *Journal of Bacteriology* **179**:6048-6052.

Matic, I., M. Radman, F. Taddei, B. Picard, C. Doit, E. Bingen, E. Denamur, and J. Elion. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* **277**:1833-1834.

Maurer, D. J., B. L. O'Callaghan, and D. M. Livingston. 1996. Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **16**:6617-6622.

Modrich, P. 1991. Mechanisms and biological effects of mismatch repair. *Annual Review of Genetics* **25**:229-253.

Morral, N., V. Nunes, T. Casals, and X. Estivill. 1991. CA/GT microsatellite alleles within the cystic fibrosis transmembrane conductance regulator (CFTR) gene are not generated by unequal crossingover. *Genomics* **10**:692-698.

Parker, B. O., and M. G. Marinus. 1992. Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proceedings of the National Academy of Sciences U.S.A.* **89**:1730-1734.

Petes, T. D., P. W. Greenwell, and M. Dominska. 1997. Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. **Genetics** 146:491-498.

Samadashwily, G. M., G. Raca, and S. M. Mirkin. 1997. Trinucleotide repeats affect DNA replication in vivo. *Nature Genetics* **17**:298-304.

Schlötterer, C., and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Research **20**:211-215.

Sia, E. A., S. Jinks-Robertson, and T. D. Petes. 1997. Genetic control of microsatellite stability. *Mutation Research* **383**:61-70.

Sia, E. A., R. J. Kokoska, M. Dominska, P. Greenwell, and T. D. Petes. 1997. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Molecular and Cellular Biology* **17**:2851-2858.

Smith, G. P. 1973. Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symposium of Quantitative Biology* **38**:507-513.

Strand, M., T. A. Prolla, R. M. Liskay, and T. D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**:274-276.

Strauss, B. S., D. Sagher, and S. Acharya. 1997. Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA. *Nucleic Acids Research* **25**:806-813.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. *Cold Spring Harbor Symposium of Quantitative Biology* **31**:77-84.

Streisinger, G., and J. Owen. 1985. Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics* **109**:633-659.

Taddei, F., I. Matic, B. Godelle, and M. Radman. 1997. To be a mutator, or how pathogenic and commensal bacteria can evolve rapidly. *Trends in Microbiology* **5**:427-428.

Taddei, F., M. Radman, J. Maynard-Smith, B. Toupance, P. H. Gouyon, and B. Godelle. 1997. Role of mutator alleles in adaptive evolution. *Nature* **387**:700-702.

Tishkoff, D. X., A. L. Boerger, P. Bertrand, N. Filosi, G. M. Gaida, M. F. Kane, and R. D. Kolodner. 1997. Identification and characterization of *Saccharomyces cerevisiae* EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proceedings of the National Academy of Sciences U.S.A.* **94**:7487-7492.

Tishkoff, D. X., N. Filosi, G. M. Gaida, and R. D. Kolodner. 1997. A novel mutation avoidance mechanism dependent on *S. cerevisiae* RAD27 is distinct from DNA mismatch repair. *Cell* **88**:253-263.

Torkelson, J., R. S. Jarris, M.-J. Lombardo, J. Nagendran, C. Thulin, and S. M. Rosenberg. 1997. Genome-wide hypermutation in a subpopulation of stationary-phase cells underlies recombination-dependent adaptive mutation. *EMBO Journal* **16**:3303-3311.

Tran, H. T., J. D. Keen, M. Kricker, M. A. Resnick, and D. A. Gordenin. 1997. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Molecular and Cellular Biology* **17**:2859-2865.

Wang, G., M. M. Seidman, and P. M. Glazer. 1996. Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science* **271**:802-805.

Wierdl, M., M. Dominska, and T. D. Petes. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**:769-779.

Wierdl, M., C. N. Greene, A. Datta, S. Jinks-Robertson, and T. D. Petes. 1996. Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics* **143**:713-721.

Wolfe, K., W.-H. Li, and P. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Sciences U.S.A.* **84**:9054-9058.

Wolfe, K. H., P. M. Sharp, and W. H. Li. 1989. Mutation rates differ among regions of the mammalian genome. *Nature* **337**:283-285.