The Molecular Biology of Genetic Transposition

Jerry Bergman*

Abstract

This paper reviews the functions, mechanisms, and types of movable genes called transposons. It is concluded that they play an important role in all living organisms, including antibiotic resistance and increased genetic diversity. This field has exploded in recent years, and indications exist that movable genetic elements are a significant part of the genome both in terms of size and importance. Research has also indicated that all transposons are controlled by a complex genetic mechanism that functions to reduce the likelihood of inappropriate transposon placement. The evidence now indicates transposons are part of a complex designed system that serves the function of producing morphological variety. Furthermore, the transposition mechanism cannot be explained by a step-by-step evolutionary process but must exist as a complete unit in order to function.

Introduction

Transposition is a complex genetic process involving organized movement of genetic elements such as *transposable elements*, or *transposons*, to new locations on the DNA. Often called jumping genes, transposition literally means "a change in position." Gene movement can be either to a position somewhere else on the same chromosome, or to another chromosome altogether. Transposons have been called the "smallest living creatures" because they consist of replicating units that seem to behave autonomously (Plasterk, 1993). It is now known that movable genetic elements play an important role in many living organisms, including as part of the antibiotic resistance mechanism and to increase genetic diversity. Evidence now also exists that they play a role in cancer development.

The Discovery of Transposons

The first transposable elements to be recognized include *controlling elements* (now called transposons), *activators* (now called regulators), and *disassociation elements* (now called autonomous elements). They were first identified as responsible for color alterations in *Zea Mays* corn by Barbara McClintock in the 1940s. McClintock presented her first paper on her findings at a 1951 genetics symposium at Cold Spring Harbor Laboratory on Long Island, N.Y. (McClintock, 1951). Her conclusion that genes actually move from one place to another on the genome in an organized manner was at first greeted incredulously (Dean, 1984). Her work on transposons had been an outgrowth of her earlier research confirming that genetic crossing-over during meiosis does in fact exchange genes (Landman, 1994).

A major observation which led her to postulate the transposon concept was the discovery that in plant offspring one sister cell (cells that are descended from a cell after differentiation, thus are the same kind of cell) *gained* what the other *lost* in a regular way. For example, McClintock noted that whenever one sector of a variegated chlorophyll pattern plant leaf showed a greatly increased number of green streaks on the white background, its twin showed a *reduced* number of green streaks (Dean, 1984). These plants led McClintock to conclude that one cell had gained some component that the sister cell had lost during a mitotic cycle (McClintock, 1987).

The specific factor McClintock found to be responsible for governing color was a *suppressor* gene which she called a *disassociater* or Ds element. When the disassociater lies near the gene for purple, for example, this color is repressed, but when it moves to another spot *away* from the purple gene, the color purple is expressed. The suppressor genes are in turn controlled by regulator genes that are located nearby the suppressor genes on the same chromosome.

Even after McClintock had published numerous well documented research reports in leading genetic journals, the scientific community resisted her conclusions. The response to her findings was in her words:

> . . . in some instances, hostility. A third attempt to support the thesis of the origin of mutable loci in maize appeared in 1953 in the widely read journal

*Northwest State College, Archbold, OH 43502
Received 18 March 1996; Revised 23 December 1999

> *Genetics*. It was titled "The Induction of Instability at Selected Loci in Maize." This article appeared before copying machines practically eliminated requests for reprints... In this instance I received a total of only three requests for this reprint! By then I had already concluded that no amount of published evidence would be effective...There were many vocal skeptics. Therefore, the method I had chosen to record data and conclusions from them was continued into the early 1960s.... In retrospect, it appears that the difficulties in presenting the evidence and arguments for transposable elements in eukaryotic organisms were attributable to conflicts with accepted genetic concepts. That genetic elements could move to new locations in the genome had no precedent and no place in these concepts (McClintock, 1987, p. x).

A major reason for the skepticism common at this time was because McClintock's findings were so radical and the phenomena she described "could not be explained in terms of the then-current framework of genetics. This mind-set began to change with the discovery of movable DNA systems in bacteria" (Landman, 1994, p. 70). McClintock spent decades collecting reams of data and eventually her conclusions could no longer be disputed (Keller, 1983). The transposon process was eventually so well empirically confirmed that she was awarded the Nobel Prize in 1983.

The transposition field has exploded in recent years, and indications now exist that movable genetic elements are a significant part of the genome, both in terms of size and importance (Holzman, 1991; Nitasaka, Yamazaki, and Green, 1995). Their contemporary importance is indicated by the fact that a *Medline* journal search of peer reviewed research revealed 2,376 published articles on transposons from 1995 to 1999 alone. Research has also indicated that transposons are controlled by a complex genetic mechanism that is used to reduce the likelihood of inappropriate placement. Some evidence now indicates transposons are part of a complex system that serves several functions including producing morphological variety (Craig, 1997; Russell, 1996).

Two types of transposons exist, a so-called simple system and a complex system. The term *insertion sequence* (IS) refers to the simplest transposition system, and is usually used to describe transposable segments that are less than 2 kilobases long (about 2,000 bases, therefore the three base code sequence codes 667 amino acids). Conversely, the term *transposons* (Tn) describes the more complex movable genetic elements that are usually *longer* than two kilobases. The smallest insertion sequence consists of only one gene flanked by short repeated DNA sequences (Plasterk, 1993; Green, Ellington and Szostak, 1990). A large variety of mechanisms are used to achieve the complex process of organized gene movement. Mechanisms that control movement are divided into two types, *non-composite*, meaning they are not flanked with the control elements called *insertion sequences*, and *composite* which describes those that are flanked with control sequences (Xiao et al., 1995; Pearlman and Butow, 1989).

Transposable elements have now been found in both prokaryotic (cells lacking a nucleus and organelles) and eukaryotic (nucleated cells) organisms. They are presumably universal to all living things including humans (Dombroski, Scott, and Hazazian, 1993). Prokaryotic transposable elements have so far been studied most extensively in *Escherichia coli*, but much research has been completed on vertebrate transposons. This research has found that transposable elements are similar in both prokaryotes and eukaryotes as well as viruses (Russell, 1996).

The discovery of transposable elements has revolutionized genetics, proving that the gene is "fluid and mobile, changing constantly in quality and quantity, and replete with hierarchical systems of regulation and control" (Gould, 1985, p. 712). The number of transposable elements in most eukaryotic organisms is evidently large. In *Drosophila melanogaster* over 30 different families of transposons are now known, accounting for close to 10% of the total DNA (Pimpinelli et al., 1995). On the average transposons account for about 0.3 percent of the bacterial genome. In humans, evidence exists that as much as 35 percent of our genome may be transposons if the Alu family of transposons is included (Bestor, 1998, Schwartz, 1995).

Basic Mechanism of Transposition

The gene transposition process is initiated by one or more enzymes which excise the movable genetic element from one location on a chromosome then integrate it into another chromosome location. Next, the transposition enzymes repair the DNA insertion junction to enable the transplanted genes to function properly in their new location. The DNA normally is cleaved at a specific location, and the specific break site is called a *dissociation point*. *Transposase*, the primary enzyme that catalyzes the transposition process, consists of at least one polypeptide, but most other transposases are oligomers consisting of two to four polypeptides. The active unit of the Mu bacteriophage which integrates DNA in the bacteria genome is a tetramer, and all four subunits are evidently required for strand transfer to occur (Schwartz, 1995). In transcellular transposition, after a transposon is inserted the precise repair of adjacent DNA is completed by the host's DNA replication and repair system.

Replicative and Non Replicative Transposition

Two major transposition mechanisms are now known to exist, those involving *non-replicative* and those involving *replicative* transposons (Figure 1). Most organisms use either one or the other (although some systems can transpose using both methods). A few organisms, such as the bacteriophage Mu, use both types, though its replicative transposition does not use an RNA intermediate (Craig, 1995). *Insertion sequences* (IS) are named acording to when they were discovered and the "1" refers to the first one discovered. This family of insertion sequence transposons alone now includes over 300 types.

Replicative transposition involves a two-step process: 1) making an mRNA copy of the transposon, and 2) moving *only* the copy to another location, leaving the original gene intact. The mRNA is then transcribed back into DNA by reverse transcriptase. This transposition method is typified by the *Tn*3 bacteria transposon (*Tn* means transposon, 3 the third one discovered). The first step of integration involves a transposon forerunner unit which integrates at the DNA target site, then the transposon is replicated. The second step utilizes the co-integration process, a procedure called *resolving* that is catalyzed by *resolvase*, an enzyme coded by the transposon. The result of this process is that a single copy of Tn3 is integrated at the target site (Moran et al., 1994). This process is common because evidently the genomes of most animals contain multiple members of the Tcl family of transposable elements (Avancini, Walden and Robertson, 1996).

Replicative transposition requires an mRNA intermediate called a *retrotransposon* because the process of replication it uses is similar to how retroviruses function. Examples include some yeast transposons which almost always integrate in areas *outside* of open reading frames. Their most salient trait is the fact that they encode polypeptides which have reverse transcriptase activity.

Two known types of transposons that use the retrovirus system are: 1) retrotransposons that use *long terminal repeats* (LTRs) and 2) those retrovirus systems that lack these terminal repeats called *retroposons*. Both long terminal repeats and retroposon transposon types contain sequences similar to the *gag* and *pol* retrovirus genes, and not unexpectedly both utilize mechanisms similar to those used by retroviruses.

The advantage of the reverse transcription replicative process is the fact that it does not require excision of the gene element from the donor site. The long terminal repeat transposon type resembles retrovirus proviral DNA which also transposes by means of reverse transcription of mRNA but through a mechanism that differs in many ways from the retrovirus mechanism. One difference is the retroposons use an Adenine-rich sequence at their 3' ends in place of terminal repeats (Pimpinelli et al., 1995).

The second major transposon mechanism is called *non-replicative transposition* because it does *not* involve replication of the transposon. This method is by far the most common, and an example is the 9,300 base pair transposon *Tnl0* that contains a tetracycline resistance gene. This transposition type involves excision of the strand from its original site which requires coding for the transposases that cut both DNA strands at both ends of the transposing element. Subsequently, both DNA strands are excised, leaving behind four single-stranded staggered cuts at the target site. The transposon is then inserted into the new site, bonded, and the DNA sugar phosphate backbone molecules are properly ligated together.

Transposition also has two subtypes: 1) those with both short inverted repeats at their termini that encode a transposase which is required for their transposition and 2) those which use heterogeneous central portions flanked by long inverted termini called *foldback elements*. The first class is also called *composite* or *compound* transposons because they contain insertion sequence elements at each end (usually the insertion sequence element is inverted). This type requires specific transposases which are coded by the terminal insertion sequence elements. The second class of transposons utilize terminal inverted repeats that are about 30 base pairs long which do *not* contain insertion sequences (Kendrew and Lawrence, 1994, p. 1098).

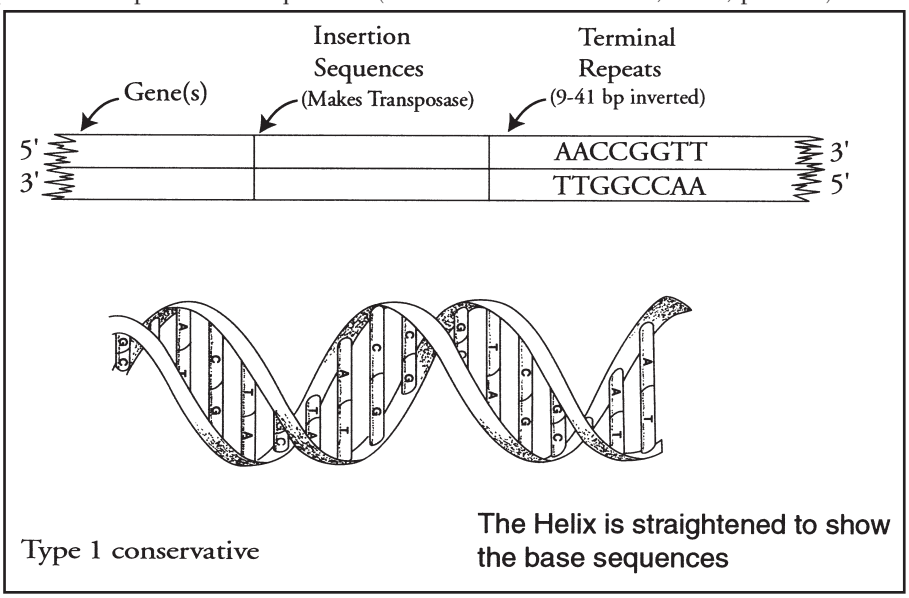

Figure 1. Non-replicative or conservative transposition which shows the process this transposon type uses to insert genes in a DNA strand. Drawings by Richard Geer.

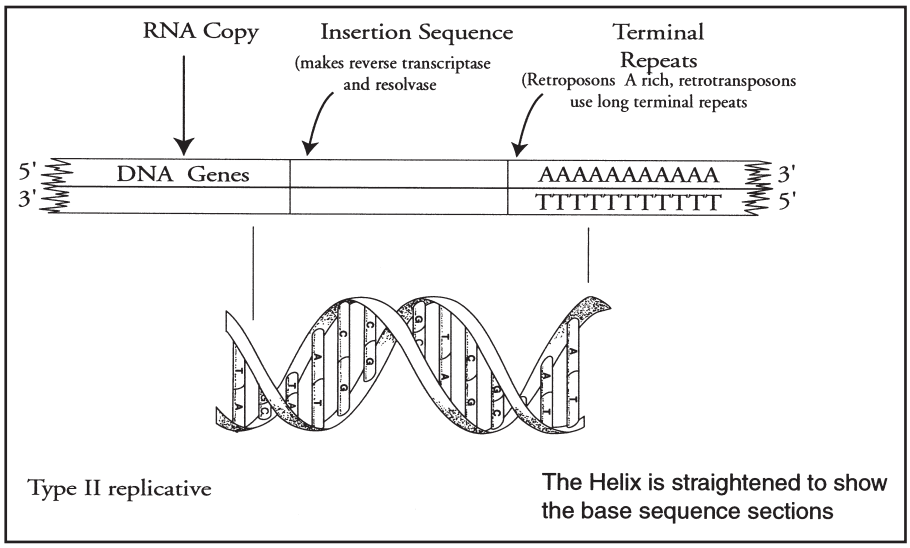

Figure 2. Replicative transposition which uses reverse transcriptase and resolvase in order to produce a gene copy.

Transposons are also now known to use mechanisms similar to those used in other biological gene exchange systems including: 1) the acquisition of bacterial genes for antibiotic resistance, 2) the replication of certain bacteriophages, 3) the integration of retroviruses, and 4) the intracellular movement of retroviral like elements.

New evidence points to unexpected parallels among these many transposition events: They all occur by similar DNA breakage and joining reactions. Moreover, the structures of transposition proteins from very different biological sources have remarkable overall structural similarity, even though they lack extensive primary sequence homology (Craig, 1995, p. 253).

Transposons can be *autonomous* (largely able to direct their own transposition), and *nonautonomous* (those that lack transposition genes and therefore must be activated by an autonomous element elsewhere in the genome) (Russell, 1996). The latter are referred to as *stable* because they cannot transpose by themselves; the former are called *unstable* because they can excise themselves to a new location. Most autonomous transposons can become stable nonautonomous elements by internal deletions or complex sequence rearrangements. This control system introduces some stability and order to the genome.

Target Choice

The criteria used to select the target choice is not yet fully understood, but some criteria likely exist because many sites attracted by transposons are largely predictable. Russell concludes that "transposons integrate at a target site by a precise mechanism so that the trans-integrated elements are flanked at the insertion site by a short duplication of target site DNA" (1996, p.629). Craig (1997, p. 437) adds that some transposition elements "display considerable target site selectivity and others display little obvious selectivity, although none appears to be truly random."

Generalizations in understanding target site are problematic because a variety of mechanisms are used. The hot spots are not easy to identify because identifying the obvious criteria, including primary sequences or morphological traits such as bends and kinks in the DNA, requires a greater understanding of the factors that influence DNA topography. A mechanism must normally exist, though, because in many cases strong hot spots are only a "few base pairs from sites that are not hit at all" (Plasterk, 1993, p. 782). Accessory proteins that communicate with the target DNA recombinase and direct interactions between target DNA and recombinase are all used to determine target sites (Craig, 1997).

A common target regulation method is the histone coiling-wrapping system which helps to determine where the mobile genetic element can be inserted (Wolfe, 1995). Yet another control method involves placing of transposable elements into *heterochromatin*, a specialized chromatin material which often remains tightly coiled in the nondividing nucleus and can be seen easily in the microscope because it stains darkly during interphase. The rest of the chromatin, called *euchromatin*, stains with low intensity and uncoils during interphase, usually condensing only during mitosis. Evidently the heterochromatin condition also prevents, or at least impedes, transposable-element movement. Conversely, the euchromatin position facilitates such movement. For this reason mobile genetic elements cannot normally move to another area of the genome.

Pimpinelli et al. (1995) determined the distribution of eleven different transposable elements on *Drosophila melanogaster* mitotic chromosomes. They found that nine of the eleven transposable elements were preferentially clustered onto one or more discrete heterochromatic regions in the chromosomes of the stock they tested (the offspring of samples collected near Baria, Italy). An analysis of the location of heterochromatic transposable element clusters found that they were also very similar in geographically distant strains. Researchers also concluded that these transposable elements assumed the role of major structural components in *Drosophila* heterochromatin.

Although the integration sites for "most mammalian and drosophila retroelements" at this early stage in the research may appear "to be distributed *more* or less ran-

domly," this could be true mostly because we do not yet understand the site selection criteria. As research in this area has progressed, increased knowledge has revealed a tendency to transpose to a specific area (Labrador and Corces, 1997, p. 392). Transposase is evidently a requirement for transposition, and Labrador and Corces (1997 p. 383) concluded that transposase is normally tightly regulated at several levels including transcription, differential splicing, translation and protein-protein interactions. One regulation system involves a repressor that inhibits transposition by binding to certain sites in the genome.

A totally or largely random integration site would eventually wreck havoc to the genome even if only a small proportion of random integrations occurred in the larger mammal and vertebrate genome. This would be true even in animals that have large amounts of extra DNA compared to lower life forms such as yeast. Yeast genome contains wall to wall genes and "most or all yeast retroelements [in yeast] show target site" selectivity and random integration would commonly damage genes (Labrador and Corces, 1997, p. 393).

Furthermore, all mobile elements share fundamental biochemical and structural similarities, and for this reason all mobile elements would be expected to behave in ways similar to those in yeast and manifest tight control over the target site (Craig, 1995, p. 254). Nevertheless, the mechanism occasionally does not work properly, and a gene is damaged by a transposition insertion, causing that gene to malfunction. If the gene is a tumor suppressor gene, oncogene, or another gene implicated in cancer, tumor progression will result. Also, much of the damage caused by transposition may be blocked by methylation.

All transposable genetic elements utilize at least one transposase, and several use more than one. Transposases in bacteria are usually encoded by the insertion sequence elements located on each side of all transposons. The two known roles for transposase are, first, the cleavage of the correct DNA sequence that is to be moved, and, second, the accurate recognition of the two transposon ends. This mechanism evidently usually insures that only the proper set of DNA bases is excised and the movable genetic elements are spliced into the new correct location of the gene. The synthesis of transposases is in turn under the direct control of a *repressor* which rigidly controls transposition frequency. This tight control is necessary because transposases could otherwise damage the genome, causing cancer (McKee and McKee, 1996).

The Molecular Biology of the Process of Transposition

The enzyme catalyzing transposition is an "impressively" long rod-like structure two micrometers in length (Plasterk, 1993, p.783). The two transposon ends must be integrated into the target DNA strand by a rotationally symmetric reaction. The transposase then binds to the end recognition sequences. A second enzyme family called *integrases* functions as a transposase for both retrotransposons and retroviruses which are located in eukaryotes (Craig, 1997). Cointegrate resolution of Tn3 phage DNA also requires the integrase named *resolvase*. It has now been confirmed that these enzymes are all structurally similar, even though surprisingly little sequence homology exists between them:

> That a retroviral integrase and a bacterial transposase are indeed fundamentally related has now been spectacularly demonstrated by the structures of the catalytic domains of these proteins at the atomic level. The overall topology of the HIV integrase and MuA transposase structures is similar, although there is little extended primary sequence homology between them. That these proteins might be structurally related was not entirely unanticipated; these and other recombinases contain a signature array of conserved acidic amino acids, the D, D(35)E motif, so-called because of the usual 35-amino acid spacing between the last two residues. These conserved amino acids are critical for the 3' end processing reactions, suggesting that these amino acids are part of (or at least closely related to) the active sites of the enzymes. Moreover, in both the integrase and MuA transposase structures, these amino acids are close together, forming a plausible binding site for a metal ion cofactor essential to recombination (Craig, 1995, p. 254).

All insertion sequences and transposons that have been sequenced so far are flanked with perfect or close to perfect *terminal repeats* that are between 9 and 41 base pairs long. The repeats are also found *inverted* with respect to each other: the same or a very similar set of sequences is found at each end but in the opposite orientation (Russell, 1996). The number of inverted repeat sequences at the transposon termini area ranges from a few to several hundred bases. The terminal repeats are essential for the transposition process because only specific sequences are recognized by transposase, and these are essential to successfully catalyze the DNA cuts and initiate transposition.

Short DNA sequences are duplicated during transposition in order to repair the insertion ends at the *target site*. The target site cut is often staggered, resulting in an overhang of from 4 to 13 bases. The last required step for both insertion sequences and transposons is for DNA polymerase and DNA ligase to repair the ladder frame by filling the 4 to 13 base gap—a process called *target site duplications*. Repair of the break in the DNA is often template dependent, usually relying on the homologous chromatids, the sister chromatid, and even *ectopic* templates (nonhomologous gene segments).

Virtually all transposable elements utilize this system to repair their insertion site DNA regardless of the transposition mechanism they use. The number of bases duplicated is either constant or may vary with each strand position or event. Why this difference exists and the exact role of these inverted repeat sequences are not understood, but it is known that they are necessary in order for transposition to occur.

Transposon sequences of one species are often more similar to those in closely related species than they are to less closely related species. Exceptions to this tendency suggest the possibility that transposable elements may occasionally be transmitted horizontally, that is from one species to another morphologically similar species. Some transposons integrate primarily in one position, others in several, and they do not normally integrate randomly. One possible reason for a variety of integration sites may be the result of mutations or structural defects in chromosomes which cause a transposon to integrate at the wrong site in the DNA.

One method of transposon control which illustrates what can go wrong is located in transposable factors called *P elements*. The 3,000 bp (base pairs) long P elements are prevented from jumping around on the somatic chromosomes by some unknown mechanism. They are normally able to jump only in gametes due to three introns that break up the transposase gene. These introns serve as a control element that must be removed from the mRNA before the transposase protein can be produced (Green, Ellington, and Szostak, 1990).

Normal splicing removes only two of these introns from the nuclear DNA of the somatic cells, and for this reason it cannot make transposase. Conversely, *all three introns* are evidently removed from the germ cell line (Wills, 1989, p. 266). When the third intron was removed from the cloned element and the modified P element was reintroduced in *Drosophila melanogaster*, it transposed "everywhere," not just in the germ line. As a result, small patches of mutant tissue appeared in many locations on the fly. Control must normally exist because most transposons show only occasional shifts, and rarely do they transpose everywhere as indicated in the research cited by Wills above. Other possible causes of transposon movement includes environmental circumstances, or an optimal number of transposable elements may exist which will result in their blockage if this number is exceeded.

Horizontal Spread of Transposons

The P elements of *Drosophila* have been studied most extensively to understand the mechanism's spread of transposons within an animal type. Research on captive populations of these flies indicates that the transposon was transmitted from *Drosophila willistoni* into a *Drosophila melanogaster* fly about 70 years ago. Today all living *Drosophila melanogaster* flies evidently contain the P element except those isolated in laboratories (Plasterk, 1993). The P elements can be accurately detected because the repressor gene which normally regulates the P elements' activities does not exist in *Drosophila melanogaster*.

Both Tcl and the mariner superfamily of transposons exist in the genome of most insects, platyhelminthes and even mites (Plasterk, 1993). The Tcl transposon has been found in nematodes, fungi, ciliated protozoa, and several kinds of vertebrates. This is explained by evolutionists as a result of horizontal spread distribution, but the fact that they are "absent in whole parts of evolutionary trees and then show up with 95% nucleotide conservation in species thought to have diverged 200 million years ago" indicates either they were created as part of the genome in diverse forms of life or for the purpose of horizontal spread (Plasterk, 1993, p.781). Avancini et al. (1996, p. 131) found that this family of transposons is "extremely common in animal genomes with multiple representatives in most genomes." Their commonality indicates transposons are important and implies that they have a central function in many living things. If their commonality is due to horizontal spread, this indicates that a mechanism exists that permits horizontal spread and they serve a purpose in the genome.

Looped Transposons

Another type of transposon control involves the disassociation of an entire transposon, which then forms a DNA ring similar to a plasmid. For this reason the DNA is usually unable to rejoin the chromosome. Consequently, all of the genetic information the ring contains is evidently lost to the genome. Many transposons are evidently repressors, and when removed by translocation all of the genes it normally represses can then be expressed. As these control sequences travel around the genome, the genes are turned on and off (Moran et al., 1994). Allowing repressors and other control sequences to randomly move around the genome surely would eventually wreck havoc to the genome, arguing for some type of control system.

Another means of transposon spread is by bacteriophages that replicate the transposon each time they reproduce. Some plasmids called *episomes* are capable of splicing their way into bacteria DNA. An example is *E. coli* fertility factor F which consists of 94,500 bp of DNA and codes for a variety of proteins (Russell, 1996). Nonmobile transposable element clusters may regulate both the activity and the copy number of mobile transposable elements.

Functions of Transposons

Many evolutionary theorists view most transposons as parasitic DNA which often randomly move around the genome. The result is a major source of random genetic variations that natural selection can select from, increasing the likelihood of producing genetic changes that convey a survival advantage for the animal from its mixed bag of largely destructive, useless, or otherwise non-functional DNA.

Although once felt to be only "rogue DNA's that seemed to land anywhere in the genome" many transpositions are now recognized as having "profound effects on the expression of genes" (Pennisi, 1998, p. 1134). While most of the genome is not yet understood, many current findings, some of which are reviewed in this section, indicate that what is known about transposition does not seem to support this hypothesis. Rather, it suggests that the transposition system is a complex organized system of gene movement and "each system has its own highly specific requirements for how these reactions must be temporally and spatially coordinated ... [by] an intricate nucleoprotein architecture" that orchestrates transposition (Craigie, 1996).

The list of known functions for transposons is growing yearly, and transposons are now believed to play an important role in the genome of most organisms (Craig, 1997). A primary role is to produce color and trait variety, but transposase enzymes also are critical for antibiotic resistance (Pennisi, 1998). Travis (1992) concluded that transposons are a "constructive driving force" which generates much genetic variety. Most of the antibiotic resistance transposons that have been studied in detail use an adjunct transfer vector to distribute their antibiotic resistance genes. An example is the R plasmids which are located in the pathogenic bacteria *Shigella*. Some transposable elements even function as introns which are removed from the mRNA by RNA splicing (Craig, 1997).

Except for bacteria, the transposition mechanism evidently usually does not involve genes that code for trait variations that influence the survival of the animal or plant. Aside from the variety they produce related to defense against antimicrobial compounds in bacteria, transposition usually seems to be limited to influence traits such as color variety. As noted, without an organized and designed control system the transposition machinery would also soon cause deleterious mutations and changes in the genome:

> Insertions and inversions can disable the gene into which they enter by disrupting the reading frame. In this way, they act as switches to turn a gene OFF. The insertion can also be precisely removed, and the inversion precisely reversed, to permit the gene once more to function. The insertion sequences (IS) have special indicators on their ends that identify them for removal and transposition. If it were not for the precision with which they act, they would be turning genes OFF at random, wreaking havoc in the genome. Moreover, if not for the special indicators on the ends of the IS, once a gene is turned OFF in this manner there would be little chance the IS could be precisely removed to turn the gene ON again. The chance that a random deletion will precisely take out a previous insertion is very small.
>
> The chance is also small for a random inversion to reverse a previous inversion. The chance that a transposition will occur in the genome of a bacterium is about one in a million generations... If they are random, then they would be equally likely to occur anywhere in the genome... (Spetner, 1997, p 89–90).

Spetner notes that the genome of a bacterium has about a million nucleotides. Therefore the chance that an inversion or a deletion will occur and one of its endpoints will land on a particular nucleotide would be about one per trillion replications. Such events can occur in bacteria, but cannot be important in higher animals because their reproduction and mutation levels are far less frequent. An example he gives is during the entire putative 65 million year long horse evolution, only about a trillion replications have occurred.

Research on the mechanisms of transposition reveals that this system contains a far greater level of organization and complexity than previously understood. The extremely complex transposition mechanism could not exist without the simultaneous presence of all its many parts, supporting the concept of irreducible complexity. Even the existing complex transposition system is imperfect and occasionally results in a gene being placed in an area of the genome or in another gene that causes genetic damage. Spetner concludes that

> there are good reasons to believe that these genetic rearrangements are not random. Insertions and inversions are complex rearrangements of the gene. Inversions occur when two sequences recombine in just the right way. Inversions seem to have important roles to play in both cells and organisms, but we don't yet know what those roles are. We do know, however, that they are not just genetic mistakes. They are controlled by a set of special enzymes... Some of the enzymes they need are encoded in the transposon itself, and some are encoded in other parts of the cell's genome. The rearrangements seem to be deliberate acts performed on behalf of the cell (or the organism). They do not seem to be the random stuff that the NDT (NeoDarwinism) says propels evolution (1997, p. 89)

If further research supports the thesis of this review, transposition will be one of more than a score of systems

which eloquently support the conclusion that the amount of information necessary to control the body's genetic systems is phenomenal. This view is supported by the growing information about the complexity of the human genome and the enormous amount of information it contains.

Although the role of many transposable elements is still unknown, it would be rash to assume they have no function. Past experience of human anatomists in labeling 180 structures as vestigial organs, all which are now known to serve a function, provides an excellent illustration of the fallacy of labeling a structure useless. Transposons in eukaryotes have been studied so far mostly in yeast. Other studies have thus far been largely confined to *Drosophila*, corn, and humans. Many of these transposons aside from those that code for the transposition mechanism, also carry genes, and many of these genes have as of yet unknown functions (Russell, 1996). Nonetheless, the still widely held selfish DNA theory (Osawa, 1995; Black, 1999; Dawkins, 1976), which holds that the only role of most transposons is to reproduce themselves and spread throughout the genome is becoming less tenable as more is learned about transposable elements (Travis, 1992).

Labrador and Corces (1997, p. 389) concluded that although the functional versus selfish DNA debate has not been settled yet, recent findings support the functional view. They note that the finding that retrotransposons can serve as repair systems such as "healing *Drosophila* telomeres opens a new scenario" on this question (1997, p. 399). Furthermore, the new evidence that indicates that transposable elements can "contribute to the maintenance of heterochromatin, and consequently to chromosome structure ... opens new and exciting avenues for experimentation ..." (1997, p. 399).

In humans, the SINEs family is a 300 base long sequence that is repeated up to 500,000 times and amounts to 3 percent of the total genome. SINEs are a member of the large Alu family of genes which are all hypothesized to be transposable elements. Likewise the LINEs family member called L1 is also believed to have transposed via a retrotransposon method. The fact that as many as 100,000 copies—fully 5 percent of the human genome—consists of L1 elements, also indicates that transposons may involve a significant fraction of the human genome (Hadler, Devadas and Mahalingham, 1998; Russell, 1996).

Evidence that one function of transposons is to produce phenotypic variety is indicated by research on the P elements. Loss of a P element may cause the organism to lose "some potential for generating genetic variability" (Wills, 1989, p.268). Certain domestic animals may have lost some of their transposons as indicated by the fact that isolated laboratory flies do not have P elements. Wills (1989, p. 268) speculates that this may also have happened to humans, noting that it would be intriguing to determine if groups such as the Australian Aboriginals who have not yet acculturated in the cities "have more mobile elements in their genomes than the effete Western man."

Transposition Can Produce an Increase in Gene Number

Replicative transposition noted above can function to increase the gene number as a result of each transposition. Transposons can also increase gene number if their removal results in a chromosomal break that can be repaired by using the sister chromatid (the homologous chromosome) as a template for DNA synthesis to effectively make a new copy of the now-departed transposon. If the donor site is restored with its transposable control gene, which is likely the case if the sister chromatid was used as the template to make the repair, the transposon command would again be initiated. This process may occur repeatedly, producing many copies of a gene (Fedoroff, 1991).

Mechanisms involved in producing these repeats are not fully understood. However, certain excessive repeats cause instability of the organism, and consequently reversion back to the wild type is common. This usually occurs as a result of a complete or partial excision of certain transposon elements. The reason they revert back to the wild type is evidently due to another repair process which determines that uncontrolled replication of the transposon has occurred and corrects the problem. Some results of this multiple transpositions process in plants include producing a striking physical appearance in the plants leaves, petals, seeds, and other tissues (Plasterk, 1993).

When the System Fails: Transposition as Intragenomic Parasites

Although some evolutionists argue that most transposons are intragenomic parasites, the so called parasitic DNA theory (Yoder, Walsh, and Bestor, 1997), much research contradicts this view, including evidence that transposons are evidently usually tightly regulated by the host and usually insert into a specific site which does not adversely affect the host (Craig, 1997). It is known that the host can determine the transposition rate of the Tc1 transposon, and this may also be true for many other transposons (Labrador and Corces, 1997).

Furthermore, the fact that most transposable elements normally avoid certain sites and/or regions of the genome indicates that a negative feedback system exists to control transposition rates and sites (Craig, 1997, p. 438). Creationists argue that transposition is a complex system designed for a constructive purpose including to produce genetic variety, and the most important and best documented function is to transpose antibiotic resistance genes.

Some movable genetic elements *seem* to act like parasites, existing in DNA and reproducing as the cell reproduces but serving no useful purpose. To explain the existence of these transposons, some researchers assume that many of the parasitic DNA elements were mutated and could have been selected for because they benefited the host (McDonald, 1990). Intragenomic parasite transposons are evidently usually blocked by 5-methylcytosine. Evidence exists that most of the 5-methylcytosine in mammalian DNA resides in the specialized intragenomic parasite transposons (Yoder et al., 1997).

Yoder et al. (1997) also claim recent evidence indicates that reversible promoter methylation is not commonly involved in developmental gene control as once thought. Rather the primary function of cytosine methylation is evidently the suppression of parasitic or unneeded sequence elements, and the secondary role of cytosine methylation is control of allele-specific gene expression including both X inactivation and genomic imprinting.

Aside from the subset of genes subject to inactivation in females to achieve dosage compensation and genomic imprinting, no other cellular gene in non-expressing tissue has been proven to be methylated in a pattern that prevents transcription. The transposon promoters are inactivated when methylated, and the C→T transition mutations which occur at methylated sites usually destroy the parasitic transposons. Consequently, these mutated movable genetic elements are removed from the genome. Therefore, parasitic theory cannot explain most of the permanent transposons. They also conclude that

> It has become increasingly difficult to hold that reversible promoter methylation is commonly involved in developmental gene control; instead, suppression of parasitic sequence elements appears to be the primary function of cytosine methylation, with crucial secondary roles in allele-specific gene expression as seen in X inactivation and genomic imprinting (Yoder et al., 1997, p. 335).

Use of Transposons for Molecular Biology Research

Implications of transposons for molecular biology is enormous. Due to various difficulties in using viruses for vectors, researchers have experimented with using transposons as gene carriers. Since many transposons insert themselves at specific sites in the gene, the researcher can clone the desired genes into the transposon by recombinant DNA techniques. Transposons have potential as a vector to transfer genes not only from one organism into another, but because transposons are evidently "not restricted in their host range," they may have the flexibility necessary to function as effective yet versatile vectors (Plasterk, 1993, p.785). The transposon elements of *Drosophila melanogaster* have already been used as vectors.

Transposable elements are now also important experimental tools that can be used not only as vectors but also as probes. A process called *transposon tagging* uses a transposon probe to identify recombinant DNAs in order to determine if they have successfully inserted themselves into the genome and also to determine if they have carried the gene of concern into the correct area. Many genes responsible for murine cancer have been identified by retroviral tagging.

Transposon Mutations

Occasionally the transposon mechanism does not function properly and produces a spontaneous mutation. If a transposon is inserted inside a gene, it can inactivate either necessary or optional genes which may explain the origin of some pseudogenes. Incorrect insertion can also cause gross chromosomal rearrangements, both due directly to the results of their transposase cutting an incorrect site or as a result of a transposon being spliced into an area of the genome where it interfears with normal gene function. Insertion inside a reading frame can destroy the gene, causing a *null mutation*, and insertion into a genetic control area such as a CpG island or enhancers can cause changes in genetic expression (Akopyants et al., 1998).

If the transposable element is imprecisely excised, the polypeptide that the gene produces may be altered. If the bases added or deleted are a multiple of three, the polypeptide may be functional but usually only if the changes occur at the genes' ends which are less critical in producing functional proteins. If a number of bases other than three is lost or gained, a string of nonsense genes will likely result, and totally nonfunctional protein will be produced.

Since mutations caused by transposons are relatively uncommon in humans and many other animals compared to those caused by mutagens, clastogens and transcription errors, mechanisms likely exist to reduce the likelihood of their placement between a gene or other inappropriate locations in the genome. Although it is assumed that a large proportion of spontaneous mutations in *Drosophila melanogaster* that are of laboratory origin are associated with mobile DNA elements, Nitasaka, Yamazaki and Green found that none of the mutants they examined are associated with an transposition insertion. They concluded that:

> Clearly much more information, both of terms of numbers of genes studied and numbers of alleles per gene analyzed, is needed before it is possible to answer with any confidence the question of whether the spectrum of spontaneous laboratory mutations is predictive for what takes place in the wild (1995, p. 168).

Transposons and Cancer

Transposons also play a major role in the development of tumors and cancer (Cao, Brown and Sognier, 1997; Miki et al., 1996; Ramel et al., 1996). The best example is retroviruses which can function as transposons to transport oncogenes into a genome such as the papillomas virus. Understanding transposons will help researchers better understand the gene carrier initiated changes which cause cancer. Many, likely most, heterochromatin genes "have organizations that resist inactivation by insertional mutagenesis," consequently protecting them (Pimpinelli et al., 1995, p. 3808). For this reason euchromatin may be the major mutational concern.

Transposons for Cancer Therapy

The implications of transposons for curing disease is enormous. Since they insert themselves at specific places in the gene, the researcher simply clones the desired genes into the transposon by recombinant DNA techniques. Once this is achieved, one goal is to cause transposons to carry genes into cells which can cause cancer cells to react a certain way so the tumor can be effectively and accurately detected or to cause the cancer cells to revert to normal or self-destruct.

Significance of Transposons for Creationism

Evolutionists have postulated that the genome has been evolving for billions of years and is a hodge podge of not only useful genes, some which are still evolving and many others that are all but useless, but also currently non-functional DNA and genes such as pseudogenes (Travis, 1992; McDonald, 1990). Some genes were once useful but are now no longer needed due to evolutionary changes, and yet others have mixed or no clear function (Smit, 1996). Transposons have even been postulated to contribute to the "explosion of new organisms, many...bizarre in appearance [that] appeared suddenly in the Cambrian fossil record" (Travis, 1992, p. 884). The discovery of transposition at first seemed to support this evolutionary view.

Transposons were assumed to randomly move around the genome, often causing damage, and often they were selected against but would occasionally cause beneficial mutations. Some evolutionists believe they are a major source of genetic variation because they can cause random variations that selection can select from, increasing the likelihood of mutations that convey a survival advantage for the animal from its mixed bag of largely destructive, useless or otherwise non-functional DNA.

While the function of most of the genome is not yet understood, many current findings, some of which are reviewed in this paper, indicate that transposition does not support this hypothesis. Rather, it suggests an intelligent design origin because the transposition system is a complex highly organized system of gene movement and "each system has its own highly specific requirements for how these reactions must be temporally and spatially coordinated ... [by] an intricate nucleoprotein architecture" that orchestrates transposition (Craigie, 1996).

Evolutionary naturalism postulates that the genome accumulated as a result of the selection of random mutations which has occurred throughout evolutionary history. Research on the genome such as the mechanisms of transposition indicates that a far greater level of complexity exists than previously imagined. The extremely complex transposition mechanism can not function without the simultaneous presence of *all* its many parts and thus can not be explained by natural selection. The mechanism would not be selected for *unless* and *until* it confers a survival advantage to the organism. Furthermore, in the case of transposition, many of its parts such as the enzymes and DNA would damage the genome unless the control that results from the whole system existed.

Variation is critical for evolution, but a mechanism which produces variation would not be selected until it functioned properly, and until then many of its parts would be deleterious. Except for bacteria, the transposition mechanism is not known to involve trait variations that favorably influence survival. Except for the variety they produce related to defense in bacteria, the function of transposition seems to be limited to influencing traits such as color variety which usually doesn't have survival significance. Without an organized and designed control system the transposition machinery would also soon cause the genome to become chaotic.

Even the current highly perfected system occasionally results in a transposition being placed in the wrong area of the genome, causing a mutation or other damage. Furthermore, the entire system must be in place as a unit for it to function properly. If most or possibly all of its parts are not present, the system will not function or will malfunction, causing major genetic damage. In a literature search, the author did not locate a single article which even tried to hypothesize how this system could have evolved by natural selection selecting mutations. It is a prime example of Behe's concept of irreducible complexity (Behe, 1996).

Transposons are especially important for creationism because the biological mechanism it requires reveals a complexity at the molecular biology level which is far beyond the complexity that is known to exist at the gross anatomy level. The estimated 150,000 proteins that exist in humans and many times this in all life involve hundreds of enormously complex systems. Other discoveries

of the last decade include the signal transduction system, the cell cycle control mechanism, and the large cytokine and kinase families. The DNA transcription and translation control and regulatory systems alone reveal a level of complexity unsuspected only a few years ago (Quinn, 1975, 1976).

If further research supports the thesis of this paper, transposition will be one of more than a score of systems which eloquently supports the intelligence design model and worldview. The amount of information existing in the genome necessary to control these systems is phenomenal. Intelligent design is also supported by growing understanding about the complexity of the human genome and the enormous amount of information it contains which defies naturalistic explanations.

Summary

Rather than a printed book, the genome may be more accurately likened to a manuscript stored on a word processing program. The information in the manuscript is not set in stone, but is changeable in ways that biologists are just beginning to understand (Henningfeld and Hecht, 1995; Holzman, 1991; Fedoroff, 1991; Borst and Graves, 1987). The current state of research on transposons is best summarized by McClintock:

> Only now, more than forty years after the discovery of transposable elements, are we beginning to understand enough about the ways that they can affect genes to decipher some intriguing new aspects of gene control from their study (1987, p. xi).

We are also finding that a designed order exists in the many systems which were once assumed to be random or disordered:

> The biological and conceptual connections between many different mobile elements have long been recognized. It is now clear that biologically diverse [transposon like] systems also share fundamental biochemical and structural similarities (Craig, 1995, p. 254).

Acknowledgement

I wish to thank John Woodmorappe for his comments on an earlier draft of this manuscript.

References

CRSQ: Creation Research Society Quarterly.

Akopyants, N.S., S.W. Clifton, D. Kersulyte, J.E. Crabtree, B.E. Youree, C.A. Reece, N.O. Bukanov, E.S. Drazek, B.A. Roe, and D.E. Berg. 1998. Analyses of the cag pathogenicity island of Helicobacter pylori. *Molecular Microbiology* 28(1):37–53.

Avancini, Rita, Kimberly Walden and Hugh Robertson. 1996. The genomes of most animals have multiple members of the Tc1 family of transposable elements. *Genetica* 98:131–140.

Behe, Michael. 1996. *Darwin's black box*. The Free Press, New York.

Bestor, T.H. 1998. The host defense function of genomic methylation patterns. *Novartis Foundation Symposium* 214:187–195, discussion 195–199, 228–232.

Black, Jacquelyn G. 1999. *Microbiology: principles and applications*, fourth edition. Prentice-Hall, Upper Saddle River, NJ.

Borst, P. and D. R. Graves. 1987. Programmed gene rearrangements altering gene expression. *Science* 235:658–667.

Cao, S., D. Brown, and M. Sognier. 1997. Alu sequences and the formation of double minutes/homogeneously staining regions in colon cancer cell lines. *Proceedings the Annual Meeting of the American Association of Cancer Research* 38:A2773.

Craig, Nancy L. 1995. Unity in transposition reactions. *Science* 270:253–254.

Craig, Nancy L. 1997. Target site selection in transposition. *Annual Review of Biochemistry* 66:437–474.

Craigie, Robert. 1996. Quality control in Mu DNA transposition. *Cell* 85:137–140.

Dawkins, Richard. 1976. *The selfish gene*. Oxford University Press. New York.

Dean, Donald. 1984. Barbara McClintock, pioneer. *The American Biology Teacher* 46:361–362

Dombroski, Beth A., Alan F. Scott, and Haig H. Hazazian. 1993. Two additional potential retrotransposons isolated from a human family Li subfamily that contains an active retrotransposable element. *Proceedings of the National Academy of Sciences USA* 90:6513–6517

Fedoroff, Nina V. 1991. The restless gene. *The Sciences* 31(1):22–28.

Gould, Stephen Jay. 1985. Jumping genes. *Chemtech* 15: 712–715.

Green, R., A.D. Ellington, and J.W. Szostak. 1990. *In vitro* genetic analysis of the *tetrahymena* self-splicing intron. *Nature* 347:406–408.

Hadler, H.I., K. Devadas, and R. Mahalingam. 1998. Selected nuclear LINE elements with mitochondrial-DNA-like inserts are more plentiful and mobile in tumor than in normal tissue of mouse and rat. *Journal of Cellular Biochemistry* 68(1):100–109.

Henningfeld, Kristine A. and Sidney M. Hecht. 1995. A model for topoisomerase I-mediated insertions and deletions with duplex DNA substrates containing

branches, nicks, and gaps. *Biochemistry* 34:6120–6129.

Holzman, David. 1991. A jumping gene caught in the act. *Science* 254:1728.

Jackers, P., N. Clausse, M. Fernandez, A. Berti, F. Princen, U. Wewer, M.E. Sobel, and V. Castronovo. 1996. Seventeen copies of the human 37 kDa laminin receptor precursor/p40 ribosome-associated protein gene are processed pseudogenes arisen from retropositional events. *Biochimica et Biophysica Acta* 1305(1–2):98–104.

Keller, Evelyn Fox. 1983. *A feeling for the organism: The life and work of Barbara McClintock*. Freeman, New York.

Kendrew, John and Eleanor Lawrence (editors). 1994. *The encyclopedia of molecular biology*. Blackwell Science, Oxford.

Labrador, Mariano and Victor Corces. 1997. Transposable elements-host interactions: regulation of insertion and excision. *Annual Review of Genetics* 31:381–404.

Landman, Otto. 1994. The uniqueness of Barbara McClintock. *The Quarterly Review of Biology* 69(1) 69–72.

Miki, Y., T. Katagiri, F. Kasumi, T. Yoshimoto, and Y. Nakamura. 1996. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nature Genetics* 13(2): 245–247.

McClintock, Barbara. 1951. Chromosome organization and genetic expression. *Cold Spring Harbor Symposia on Quantitative Biology* 16:13–47. 15. 1987.

———. 1987. *The discovery and characterization of transposable elements: The collected papers of Barbara McClintock*. John A. Moore (Editor). Garland, N. Y.

McDonald, John F. 1990. Microevolution and retroviral elements. *BioScience* 40:183–191

McKee, Trudy and James R. McKee. 1996. *Biochemistry*. William C. Brown, NY.

Moran, Laurence A., K. Gray Scrimgeour, H. Robert Horton, Raymond S. Ochs, and J. David Rawn. 1994. *Biochemistry*. Neil Patterson Publishers-Prentice Hall, NJ.

Nitasaka, Eiji, Tsuneyuki Yamazaki, and M. M. Green. 1995. The molecular analysis of brown eye color mutations isolated from geographically discrete populations of *Drosophila melanogaster*. Molecular and General Genetics 247:164–168.

Osawa, Syozo. 1995. *Evolution of the genetic code*. Oxford Univ. Press, New York.

Pennisi, Elizabeth. 1998. How the genome readies itself for evolution. *Science* 281:1131–1134.

Pearlman, Phillip S. and Ronald Butow. 1989. Mobile introns and intron encoded proteins. *Science* 246:1106–1109

Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Banaccorsi, E. Marchetti, R. Caizzi, C. Caggese, and M. Gatti. 1995. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proceedings of the National Academy of Sciences USA* 92:3804–3808.

Plasterk, Ronald A. 1993. Molecular mechanisms of transposition and its control. *Cell* 74:781–786

Quinn, L.Y. 1975. Evidence for the existence of an intelligible genetic code. *CRSQ* 11:188–198

———. 1976. Comments on the manipulation of genes. *CRSQ* 12:213.

Ramel, C., H. Cederberg, J. Magnusson, E. Vogel, A.T. Natarajan, L.H. Mullender, J.M. Nivard, A. Leyson, M.A. Comendador, L.M. Sierra, J.A. Ferreiro, and S. Consuegra. 1996. Somatic recombination, gene amplification and cancer. *Mutation Research* 353(1–2):85–107.

Russell, Peter J. 1996. *Genetics*. Harper Collins, New York.

Schwartz, M.D. 1995. Molecular medicine: jumping genes. *Molecular Medicine* 332:941–945.

Smit, A.F. 1996. The origin of interspaced repeats in the human genome. *Current Opinion in Genetics & Development* 6(6):743–748.

Spetner, Lee. 1997. *Not by chance*. Judaica Press, New York.

Travis, John. 1992. Possible evolutionary role explored for 'jumping genes.' *Science* 257:884–885.

Wills, Christopher. 1989. *The wisdom of the genes*. Basic Books, New York.

Wolfe, Alan P. 1995. Genetic effects of DNA packaging. *Science and Medicine* 2(6):68–77.

Xiao, Bing, Stephen J. Smerdon, David H. Jones, Guy G. Dodson, Yasmina Soneji, Alastair Aitken, and Steven J. Gamblin. 1995. Structure of a 14-3-3 protein and implications for coordination of multiple signaling pathways. *Nature* 376:188–190.

Yoder, Jeffery, Colum Walsh and Timothy Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* 13(8):335.

Erratum

CRSQ 35(4), June 2001, page 46. The following is the correct equation 2.

$$\Delta S = S_f - S_i = \int_{T_i}^{T_f} C_p \frac{dT}{T} - R \ln\left(\frac{P_f}{P_i}\right) \qquad (2)$$