J Mol Evol (1990) 30:116–124

Journal of Molecular Evolution

The Biological Equilibrium of Base Pairs

Peter Strazewski*

Institut für organische Chemie der Universität, St. Johanns-Ring 19, CH-4056 Basel, Switzerland

Summary. An inherent feature of double-stranded DNA is the possible replacement of any base pair by another one upon replication. A replication-dependent substitution mutation of a matched base pair requires the temporary formation of a mismatched base pair (mispair). A functionally complementary pair of mispairs is ascribed to each of the four types of substitution mutations. Provided that all types of mispairs can be formed, a dynamic biological equilibrium between the four matched base pairs must exist in all DNA, which is directly related to the formation and stability of the corresponding eight mispairs in vivo. Each nucleotide position in a genome can therefore be described as a system of six dynamic equilibria between the four matched base pairs. After a sufficient number of replications, these equilibrium states will express an overall mutation-selection balance for each individual base pair. In a thermodynamic context, the mispairs represent intermediate states on the transformation pathways between the matched base pairs. Catalysts change the stability and probability of formation of intermediate states. Mutagenic proteins are proposed as hypothetical substitution mutation catalysts in vivo. Functionally, they would be capable of recognizing a particular DNA sequence, tautomerizing a nucleotide base thereof, and hence efficiently inducing a specific misincorporation. Phenomenologically such catalysts would accelerate the rates of substitution mutations and provide pathways for directional mutation pressure.

Key words: Thermodynamic model of DNA base pairs — Mutation-selection equilibrium — Directional mutation pressure — Optimons — Rate of mutation — Substitution mutations — Mismatches — Rare tautomers — Tautomerizing proteins

Introduction

There are sufficient results available from in vivo mutagenesis experiments to allow the conclusion that a significant fraction of spontaneous point mutations in DNA are replication-dependent substitution mutations. They are induced during replication by the formation of all the possible mismatched base pairs (mispairs) that are stable and inconspicuous enough to the repair systems to have a reasonable probability of surviving until the next round of replication. Moreover, results accumulated from in vitro mutagenesis experiments indicate that, apart from misincorporations due to sterically unfavorable rare base pair combinations, another more DNA sequence-specific and controllable mechanism of mispair formation may be involved in spontaneous substitution mutations (Strazewski 1988; Strazewski and Tamm 1990). According to this hypothesis, a mutagen could alter, under certain circumstances, the hydrogen substitution pattern of a template base, thus turning it temporarily into a functional analogue of another base and therefore causing a misincorporation. Mispairs containing unusual tautomeric forms are very similar in shape to normal Watson–Crick base pairs. Replication errors of this sort would efficiently introduce specific substitution mutations. Although the involvement of rare tautomers of normal DNA bases is not proven for in vivo conditions, the theory of substitution mutations presented below implies this possibility.

* *Present address until March 31, 1990:* University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, Great Britain
Offprint requests to: P. Strazewski

Pathways for Substitution Mutations

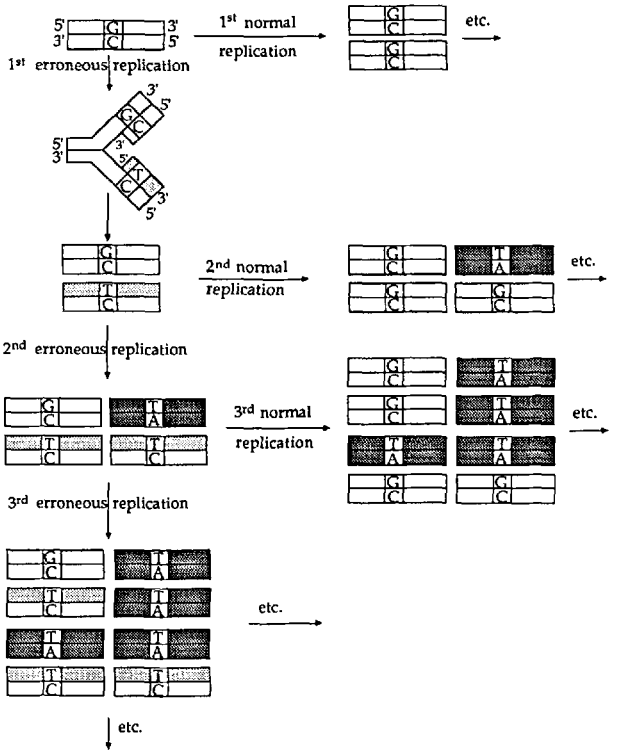
It is assumed that during replication misincorporation rates vary with the differences in the base-pairing strengths and polymerization rates between matched and mismatched pairs in polymerase–primer–template–nucleoside triphosphate complexes (Goodman and Branscomb 1986; Kuchta et al. 1988). Additionally, small transitory differences in the template DNA strand during replication, namely reversible changes in the hydrogen substitution pattern of the template nucleotide bases, are taken into account, which could be induced by proton-donating or -accepting mutagens.

A restriction to the model of substitution mutations shall be introduced. It is assumed that during replication a substitution mutation will be induced by a change in one base of the mother base pair, i.e., only one daughter strand will bear an error. This restriction expresses the fact that the probability for the formation of two mispairs per mother base pair is small enough to be neglected.

Propagation of Replication-Dependent Substitution Mutations

The products of a first erroneous replication of a given DNA sequence are one correctly replicated DNA duplex and one duplex bearing a mispair—a mismatched hybrid DNA. If we allow the mismatched hybrid to escape the postreplicative repair system, a second normal (i.e., faultless) replication will form three correct DNA duplexes and one that is mutated. This 3:1 ratio remains during all following normal replications. An example of this pathway is depicted in one part of Scheme 1 (first erroneous replication, second normal replication, etc.) using a G·C → T·A-transversion via the formation of a T·C pair. We shall call this type of pathway an unamplified mutation. Thus, an erroneous replication causes a labile change in the DNA sequence, which can be corrected by the mismatch repair system. Only a second replication or a mis-repair will convert the error into a mutation that is not correctable any more. The spread of a mutation in the offspring of the initial DNA will be more or less strongly inhibited, dependent on the type of mispair, the neighboring sequence, and the efficiency of the mismatch repair system. The 3:1 ratio therefore represents an upper limit for the production of the mutated DNA. For the purpose of this discussion, however, I shall concentrate on the highest possible mutation rates and ignore the effect of postreplicative mismatch correction.

It is widely assumed that spontaneous substitution mutations initially occur at random, and their frequency at a particular locus is well described by



Scheme 1. Propagation of a G·C → T·A-transversion via T·C.

the Poisson distribution (e.g., Kimura 1983). In consequence, repeated substitutions at one locus, given the enormous number of base pairs in a naturally occurring genome, are very rare, and the occurrence of consecutive substitutions at one locus must reach astronomically low probabilities. If, however, biological mechanisms exist that allow certain substitution mutations to be favored at a specific locus, the following propagation rule would apply to every possible substitution mutation (Scheme 1). Every consecutive formation of the same mismatched hybrid shifts the ratio rapidly in favor of the mutated product. Depending on which strand is chosen for observation, the ratio for normal to mutated DNA shifts either from 3:1 to 4:4, 5:11, 6:26, etc. (in the lower strand in Scheme 1), or from 1:3 to 1:7, 1:15, 1:31, etc., respectively (in the upper strand in Scheme 1). If the error were not propagated in every DNA copy, or would not last during a whole replication cycle, the shift of ratios would proceed at a slower rate. Therefore the mismatched hybrid operates as a mutation amplifier.

Rates of Amplified Substitution Mutations

An amplified mutation opens a possibility for a specific substitution mutation to spread through a population much more rapidly than under selection pressure, by random drift (Kimura 1983), or by the dissemination of a substitution mutation initially

$$\frac{[T{\cdot}A]}{[G{\cdot}C]+[T{\cdot}C]+[T{\cdot}A]} = \frac{2^n - n - 1}{2^n}$$

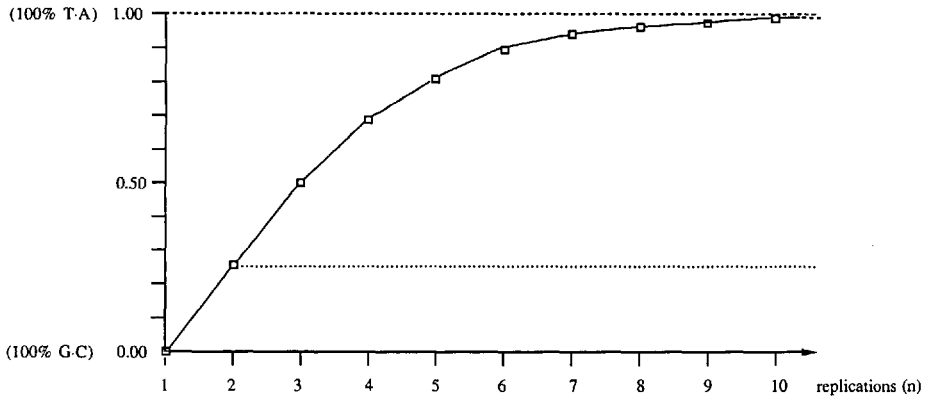
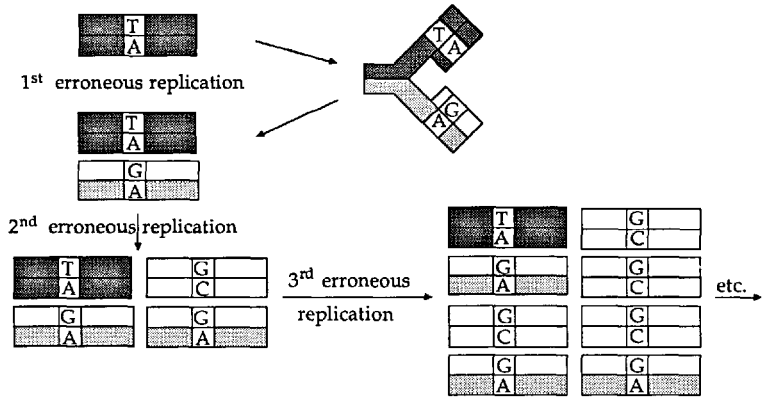


Fig. 1. Rate function for unamplified (·····) and amplified (——) substitution mutations, exemplified by a G·C → T·A-transversion via T·C.



Scheme 2. Propagation of a G·C ← T·A-reversion via G·A.

occurring only once through repeated duplications of the mutation-containing sequence by a DNA turnover mechanism, the molecular drive (Dover 1987). In the absence of selective pressures and without any postreplicative repair of the mutation amplifier, the occurrence of an amplified mutation—expressed as the amount of mutated DNA (mut) divided by the sum of the original (orig), the mismatched (mism), and the mutated DNA after n replications (percentage of a particular substitution after n replications)—will follow the function

$$\begin{aligned}\%\mathrm{mut}(n) &= 100\% \cdot [\mathrm{mut}(n)] \\ &\quad \div ([\mathrm{orig}(n)] + [\mathrm{mism}(n)] \\ &\quad + [\mathrm{mut}(n)]) \\ &= (2^n - n - 1)/2^n \qquad \text{(cf. Fig. 1)} \quad (1)\end{aligned}$$

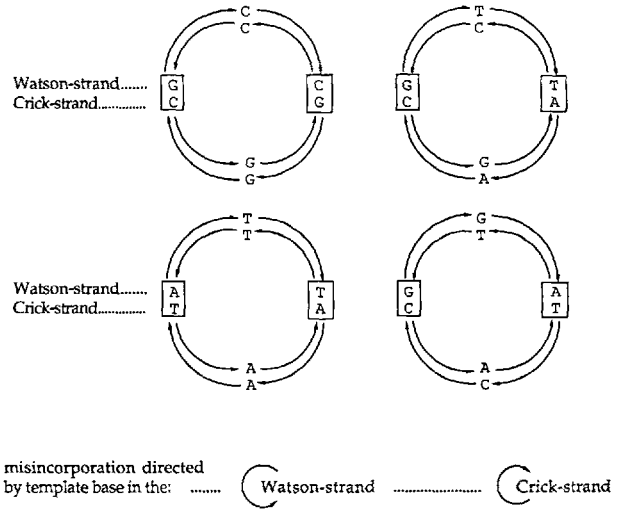
This enables a mutated DNA to occur in 99% of all offspring of the initial DNA after only 10 replications. This maximum will hardly ever be reached, but one can still imagine that even a mismatched hybrid that is efficiently corrected can have a remarkable effect on the mutation rate, if the corresponding mutation amplifier is produced at a comparably high rate.

Reversions

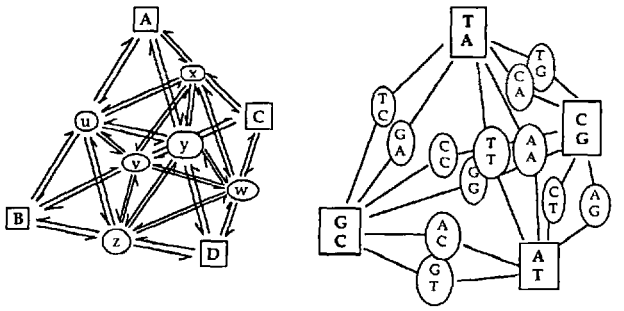
Because this pathway is valid for all kinds of substitution mutations, every transition and transversion caused by the formation of a mispair can be reversed through the formation of a functionally complementary mispair (or reciprocal, Dohet et al. 1985), under the same mechanism for the propagation of mutations. In Scheme 2 the G·C ← T·A-reversion is followed on the most amplified mutational pathway where the mutation amplifier is the G·A pair, the complementary mispair to T·C.

The Biological Equilibirum of Base Pairs

The mechanism described is reminiscent of a chemical reaction system in a thermodynamic equilibrium involving intermediate states in each transformation pathway. In vivo, within a sufficient number of replications, essentially every matched base pair in a DNA duplex is in biological equilibrium with the other possible matched base pairs. Every base pair can be transformed into another



Scheme 3. Pathways for substitution mutations.



Scheme 4. **Left:** Hypothetic system of thermodynamic equilibria between reactants A–D through intermediate states u–z. **Right:** System of biological equilibria between the matched base pairs through the corresponding mispairs. The lines symbolize transformation pathways. The connections between the mispairs are omitted for clarity.

one more or less readily, depending on its mutability, through the formation of a mispair, and can be reversed through the complementary mispair. In Scheme 3 the mutational pathways of each of the four types of substitution mutations are depicted. Additionally, all four types of substitution mutations are connected through transformation pathways between mismatched base pairs (not shown), as a mispair could in principle be erroneously replicated to form another one (e.g., C·T → C·C).

This biologically equilibrated system resembles a system of true thermodynamic equilibria. It contains four reactants, A–D, that are all interconvertible on six transformation pathways leading through the intermediate states u–z (Scheme 4, left system). Moreover, every intermediate state is also convertible into four others through 12 additional transformation pathways. In the biological counterpart the matched base pairs represent the reactants A–D, and each intermediate state consists of two complementary mispairs (Scheme 4, right system).

One difference between a system of thermodynamic equilibria and the system of biological equilibria is the reversibility of the transformation pathways. Although in thermodynamic equilibria the formation and decomposition of all intermediate states must be reversible elementary processes (double arrows in Scheme 4, left system), the connection between the four matched base pairs is accomplished by 12 transformation pathways (single lines in Scheme 4, right system), and the connection between the 12 mispairs consists of 24 transformation pathways (not shown). The formation and decomposition of the intermediate states (mispairs) are not necessarily based on actual reversible elementary processes, but proceed under given, constant (steady-state) conditions always in one direction. The reversibility of the transformations originates from the presence of similar complementary transformation pathways: For example, the simplest way to revert a G·C → C·G-transversion that proceeded through C·C, i.e., through an erroneous replication in the Crick-strand (Scheme 3), is either the formation of the same mispair due to an erroneous replication in the complementary Watson-strand, or the formation of the complementary mispair due to the erroneous replication in the Crick-strand.

A weight is attributed to each transformation, representing its probability or relative frequency, similar to kinetic velocity constants in the thermodynamic counterpart. The replication process takes into account only polymerization and proofreading. Postreplicative influences derived from the internal environment such as base methylation or base hydrolysis leading to additional intermediate states are not treated separately in the described system of equilibria. The probability of formation of such modified intermediate states can be included in the probability of formation of one or several of the unmodified mispairs.

In spite of this complication, a DNA molecule functioning throughout multiple generations may be regarded as a chain system of m_i-dimensional hyper-tetrahedrons, where m_i is the number of normal replications of a matched base pair i = 1–4 (Fig. 2). Every hyper-tetrahedron has an m_k-dimensional hyper-octahedral fine structure, where m_k is the number of replications via a particular mispair k = 1–12, expressing the degree of mutation amplification. Every tetrahedral unit is anchored on one corner, onto a stable base pair, and bears certain probabilities to flip along a given transformation pathway onto another corner. All transformation probabilities are influenced by neighboring tetrahedral units and by the environmental conditions. This can be compared to linked, interdependent thermodynamic systems, each containing four stable states of variable energies and different activation barriers.

There are two distinct mutation pressures, in-

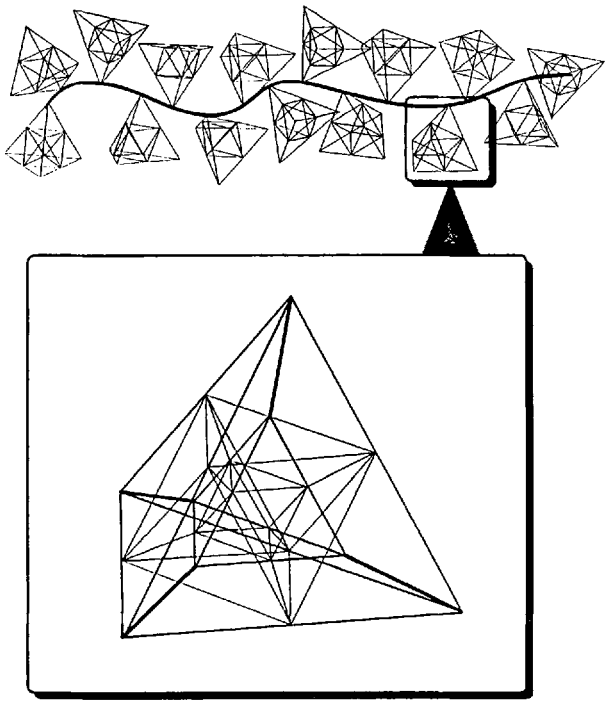


Fig. 2. A 15-base pair fragment of a genome shown as a sequence of biological equilibrium systems. The *ith* system is depicted as an m_i-dimensional hyper-tetrahedron with $m_i = 1$ and an inserted m_k-dimensional hyper-octahedron with $m_k = 1$ (i for matched, k for mismatched base pair). Bold lines from the inner tetrahedron outward are the usually most probable pathways for normal replications of matched base pairs.

dependent of external environment. The first can be expressed as the product of the probability of mispairing on replication and the probability of escaping proofreading. The second results from internal environmental factors such as the presence of DNA repair enzymes. Additionally, there are selection pressures due to the external environment, which effect mutation rates and patterns. We can deduce the following relations for base pairs in vivo. The biological equilibirum constant $K^{biol}(i/j)$ between two stable (matched) base pairs i and j in the same position of the genome shall be defined as the quotient of the probabilities W_j that one of the three other matched base pairs results and W_i that the original matched pair remains, both after an infinity of replications. $K^{biol}(i/j)$ is determined by the DNA sequence-dependent probabilities of formation w of the appropriate mispairs k and l and their DNA sequence-dependent lifetime τ:

$$K^{biol}(i/j) \equiv W_j/W_i = f(w_k, \tau_k, w_l, \tau_l) \qquad i, j = 1, \ldots 4 \qquad i \neq j \qquad (2)$$

This equilibrium rigorously implies that control over the formation and stabilization of mispairs means control over the occurrence of stable base pairs in a given position.

W_j and W_i represent kinetic reaction velocity constants in the thermodynamic equilibrium conterpart. Every base pair position in a genome has its own set of six K^{biol}'s (C·G ↔ G·C; C·G ↔ A·T; G·C ↔ T·A; A·T ↔ T·A; C·G ↔ T·A; G·C ↔ A·T) under constant environmental conditions. The sum of all sets in a genome contains the selected species and all phenotypically neutral or nearly neutral mutants (Kimura 1983) that maintain the genetic flexibility of the species, i.e., all sets of K^{biol} define a quasi species (Eigen and Schuster 1977, 1978; Eigen 1978).

The probability for correct symbol reproduction q_i of each base pair i in DNA, which, assuming that the probability to find a mispair approaches 0, is determined by the quotient W_i divided through the sum of probabilities of all pairs $(W_i + \sum_{j=1}^{3} W_j)$ [analogous eq. (IV-5) for mRNA–tRNA interactions in Eigen 1971] and can be calculated as $q_i = 1/[1 + \sum_{j=1}^{3} K^{biol}(i/j)]$. All q_i of a gene determine its quality factor Q [eq. (IV-2) in Eigen 1971 and eq. (25) in Eigen and Schuster 1978] or $\mathfrak{L}$ [eq. (II-17) in Eigen 1971], i.e., its overall mutability.

Discussion

Direct Influence of the Phenotype on the Genotype

As in the thermodynamic treatment of molecular evolution at the amino acid level (Vogel and Zuckerkandl 1971), the biological equilibrium of base pairs is balanced by two forces, the mutation pressure imposed by the DNA structure itself and by the internal environmental conditions (as mentioned above) and the natural selection pressure imposed by the external environment on the phenotype. Experimental in vitro and in vivo results from molecular biology suggest that the phenotype, here defined in a restricted sense as being identical with the internal environment of the genome, might act on DNA and in principle provide replication-dependent substitution mutations (cf. extensive literature about protein–DNA interactions). Every phenotype, irrespective of whether it contains mutagenic molecules or not, can be altered or influenced by the external environment to a certain extent and is therefore exposed to natural selection pressures. Some selective constraints, the ones that act on phenotype molecules that are (potentially) mutagenic, could be at this stage indistinguishable from mutation pressure. Or vice versa, one can think of natural pressures directly selecting base pairs at specific loci through the influence of the phenotype on the

genotype. Because most cellular DNA in a mature organism hardly ever replicates (mortal DNA), this direct action of the phenotype on the genotype is in most cases expected to be far out-weighed by the Darwinian indirect action of nature, which selects through the extinction of maladapted with respect to better-adapted alleles. Yet, there are special cases where the direct action could provide clues to a variety of observed and insufficiently explained phenomena.

Differential Mutability

In spite of in vitro and in vivo evidence for the DNA sequence-dependence of replication fidelity (Fowler et al. 1986; Schaaper et al. 1986; Schaaper and Dunn 1987; Kunkel and Bebenek 1988), population geneticists often ignored differential mutability of base pairs without the influence of natural selection. Only very recently has this opinion been questioned by the population geneticists themselves. Wolfe et al. (1989) interpret the observed DNA region-dependent mutation patterns in neutral or nearly neutral positions of the mammalian genome as being based on varying nucleoside triphosphate pools and the coherent timing of replication. They also imply a sequence-dependent efficiency of DNA repair systems.

In the following, I propose a replication-dependent mechanism of differential mutability that parallels the mutagenic effect of varying nucleoside triphosphate availabilities and assumes aimed at and effective misincorporations. If any sequence-dependent mutagenic activity derived from the internal environment of the genome exists during replication, this would lead to amplified mutations. Such amplification would increase not only the rates of substitution mutations (as shown in Fig. 1), but also the variability of the rates with time, i.e., the dynamics of substitution rates.

Some Testable Hypotheses: Tauterogens

Being well aware of the heresy against the central dogma of biology and the experimental results by Cairns et al. (1988), which prompted some researchers to reconsider the principles of Lamarckism (see Discussions about Cairns et al. 1988/1989), I shall start from another viewpoint, that of a chemist.

Genetic information is carried by the four natural heterocyclic nucleotide bases guanine, cytosine, thymine, and adenine. Each bears a distinctive hydrogen substitution pattern located on the reading part of the molecule. In a strict sense, it is this part that carries the actual genetic information, which is passed on during replication and transcription. Protolytic reactions are known to occur quite readily in many chemical systems and are also known to be effectively and reversibly influenced by catalysts of basic or acidic character. What prevents us then from assuming that in vivo these heterocyclic bases could be influenced by basic or acidic catalysts, which would alter the normal hydrogen substitution pattern into a rare one? Such protolytic rearrangements (protonation, deprotonation, tautomerism) are reversible and will be short-lived. But perhaps the influence of a catalyst could occur precisely during replication.[1] Hence, a mispairing would be provoked by presenting toward the environment, the nucleoside triphosphates in a polymerase, a false hydrogen substitution pattern, mimicking one of the three other bases (Topal and Fresco 1976; Strazewski and Tamm 1990). Because rare tautomeric forms of nucleotide bases are kinetically labile, it has been impossible so far to visualize them in an active site of a DNA polymerase. Yet results from various mutagenesis experiments suggest their existence and show that they can direct misincorporations upon replication in vitro (reviewed in Strazewski 1988). After that process, especially after the proofreading, the absence of the catalyst allows the base to rearrange into its natural form, and the mispair is at the mercy of the repair system.

What makes this type of protolytic catalyst rather special is that it could be DNA sequence-specific. Proteins are often both acidic and basic catalysts and, in the case of many DNA-binding proteins, sequence-specific. Moreover, they are the direct products of the genetic information carriers themselves. In these terms, a genetic self-control is imaginable where the mutability of base pairs in response to the mutagens present would constitute a self-regulating driving force for directional mutation pressure. We could call this mutagenic class of tautomerases tauterogens.

GC/AT Pressure

GC or AT pressures are well documented (Jukes and Bhushan 1986; Sueoka 1988) albeit their source is unknown. They could be a consequence of such protein–DNA interactions leading to substitution mutations. In Scheme 3, the G·C ↔ T·A-transversion and the G·C ↔ A·T-transition provide pathways for either a GC or an AT pressure, respectively. Both pathways are likely because although the transversion-mispairs, C·T and G·A, once formed, are in general poorly recognized by repair systems in vivo, the transition-mispairs, G·T and A·C, are generally formed more frequently, yet their repair

[1] The x-ray structure of the large fragment of *Escherichia coli* DNA polymerase I suggests that molecules like single-stranded DNA-binding proteins could have close contact with the template strand shortly before or during replication (Ollis et al. 1985)

will be quite efficient. DNA regions that have a high GC content might produce DNA-binding proteins that, after having bound to an A·T site, tend to generate G·T or C·T mispairs during replication. If there is no selective constraint on this site, an AT → GC drift will result. If natural selection counteracts the GC pressure, the mispair-generating proteins will maintain a mutation-selection equilibrium. Such an equilibrium was recently measured by Sueoka (1988) in various organisms and indicated the existence of a biological equilibrium of base pairs.

Codon Pressure

If we assume that some mutagenic proteins may be highly sequence-specific, another efficient directional mutation pressure will result. The most frequent opinion about the observed nonrandom usage of synonymous codons is that certain isoaccepting tRNA molecules happened to be less available than others and that this bias was fixed for some reason and gives rise to a biased use of modern synonymous codons (Zuckerkandl 1965; Zuckerkandl and Pauling 1965; Ikemura 1985). The hypothesis that biased tRNA usage is correlated with gene expression and that there are constraints at the RNA level that preserve this translational regulation mechanism (e.g., Gutman and Hatfield 1989) does not necessarily imply that it was this selection pressure that led to biased codon utilization. It may well be that biased codon utilization due to a DNA-inherent differential mutability caused the translation system to react with the above selective constraints in order to ensure a translational control.

If a self-regulating directional mutation pressure is present that, for example, maintains a certain GC content or favors certain dinucleotides and disfavors certain others (Ohno 1988b; Beutler et al. 1989), the observed biased usage of synonymous codons could emerge automatically. The variation in the preference for alternative codons in different organisms, different parts of an organism, or even in different DNA regions could be explained by different mutagenic conditions that act on the DNA. It should be noted that such codon pressure would force a biased tRNA usage irrespective of whether the tRNAs are isoacceptors or not. It would also result in unused tRNA genes mutating out of function in the absence of selective constraints.

According to this perspective, organisms have adapted to biased codon availability, which leads to biased tRNA usage, rather than the other way around. The reversed direction of causes emphasizes a basic character of DNA, its differential mutability, that may have originated in the absence of selection pressures. But it does not exclude selection pressures being responsible for the conservation of the resulting biases in synonymous codons, as both directional mutation pressure imposed by tauterogens and nonrandom tRNA abundancies feeding back on biased codon usage are certainly thought to be selected.

Lately, the socalled codon preference has been recognized as a natural consequence of the construction principle of coding sequences that were derived from ancestral oligomeric DNA sequences. These sequences have been conserved throughout their existence in the genomes examined (Ohno 1988a). This analysis, which focuses not on our modern codon triplets but on longer sequences, suggests subtle mechanisms for the action of mutagenic proteins. The modern codon preferences seem to be derived from very old DNA relics. Because each ancient DNA sequence presently encodes different peptide sequences depending on the position of the codon triplets and therefore possibly encodes very different functions, it is not clear how selective pressures could preserve these DNA sequences against everpresent sequence-randomizing forces throughout the ages. Proteins capable of recognizing such ancestral oligomeric DNA sequences could be as old as their operons. Such proteins would help to stabilize their sequences, if they could cause slightly mutated sequences to revert upon replication. Naturally, one thinks of a DNA sequence-recognizing protein as a folded structure, which, in addition to sequence-neutral DNA-binding contacts, will form distinctive hydrogen bonds to the nucleotide bases, possibly mediated by ordered water molecules. If the sequence that is recognized is long enough, then the binding would be sufficiently strong to tolerate a false base in the sequence. This very base might be forced to tautomerize, thus improving the binding to the protein and provoking, in turn, mispairing upon replication. Experiments will be needed to determine whether such mispairing leads to the original sequence.

Sudden Mutational Outbreaks

At present, it is unclear how genomes or cells can suddenly accelerate the rate of somatic substitution mutations. Examples are secondary mutations in transformed cells, periodic mutations in surface protein genes of certain microorganisms and viruses, or the production of as many as a thousand substitution mutations per cell per generation in the variable parts of antibody genes belonging to activated immune B cells during the stage of somatic hypermutation (Tonegawa 1988; French et al. 1989).

Perhaps the enhanced expression of genes encoding mutagenic proteins would be the answer? In healthy cells sudden outbreaks of somatic substitution mutations have to be carefully regulated so

as to ensure a controlled mutation process. During the stage of somatic hypermutation in immune B cells, for instance, low affinity contacts between antigen and antibody stimulate the antibody-producing cells to replicate and to produce substitution mutations exclusively at certain loci. High affinity contacts at a later stage of the immune response, however, are bound to stimulate even more intensely the replication of such highly adapted B cells, and yet inhibit the mutations of their antibody genes. The expression of mutagenic proteins in B cells could be coupled to these changing signals received from the antibodies. A microselection of the most effective B cells would result. In cancerous cells this regulation may be disturbed, leading to uncontrolled production of somatic mutations.

It is worth investigating any mutagenic proteins in these systems for many reasons. Their activity is expected to be difficult to assay, as they would probably be present at very low concentrations and/or only during certain periods of cell life. In contrast to conventional mutagenic mechanisms, which change the structure of DNA substantially (alkylation, intercalation, complexation), tautomerization should make the activity of these mild mutagens invisible to both the DNA repair system and the researcher.

Conclusion

Only an Idea . . .

The proposed existence of DNA-binding proteins and their mutagenic interactions with replicating DNA is a speculation with, as far as I am aware, no experimental support. This proposal was triggered by the inconsistence between conclusions from experiments performed by molecular biologists and those by population geneticists. Although it was shown that replication fidelity in vitro and in vivo is highly dependent upon the DNA sequence, hence, that on a local DNA level a substantial differential mutability is measurable, random mutation patterns are assumed on a much wider scale of genome lengths, and randomness is utilized in boundary conditions for calculations. This assumption of randomly occurring mutations spreading through a population with the help of selection pressures or merely by random drift proved to be valid in most cases in vivo. Yet I suspect that the apparent random behavior may mask fine nonrandom mutational pressures operating on small DNA regions like codons. It seems important therefore to think of ways in which the phenotype might alter the genotype in an efficient and controlled manner.

The idea of mutagenic proteins possibly acting as tautomerases (Strazewski 1988) stems from mutagenesis results in vitro using a base analogue in the DNA template strand. This analogue, a mimicry of a rare tautomeric form of a natural nucleotide base, showed how easily an efficient specific misincorporation during replication could be generated, by just changing the positions of the hydrogen atoms on the base (Charczuk et al. 1986).

On first glance, it may look as if the assumption of a mutagenic phenotype ignores the central dogma of biology. According to the neo-Darwinian idea of adaptation, neither direct nor phenotype-mediated action of the external environment on the genome, but only environmental action on the phenotype, eventually produces genomic evolution. Nevertheless a direct action on the genome can be safely allowed. The reason why a selectively mutagenic environment would usually have virtually no effect on the evolution of species is that germ-line DNA, responsible for sexual reproduction and heredity, is apparently much better protected against external influence than somatic DNA, involving immune responses and cell growth. Germ-line DNA protects itself from external influences by being exceptionally well isolated from the extracellular medium, as in oocytes, and by avoiding the presence of any intracellular medium, as in spermatocytes.

. . . For a Way of Thinking

The view of DNA base pairs presented here as an in vivo dynamic biological equilibrium demonstrates the uniqueness of our genetic information carrying system. I doubt if there exists a true chemical, nonreplicating thermodynamic counterpart that can be described by a tetrahedral/octahedral system of equilibria.

Apart from this aesthetic aspect, the concept helps us to think of DNA base pairs as the smallest units upon which natural selection acts (Dawkins 1982). If a DNA sequence is regarded as having always evolved as a function of the mutagenic activity of its direct environment, silent positions in a coding region cannot be called silent any more. This was pointed out by Zuckerkandl (1987), who emphasized different senses of neutrality. Silent positions are parts of synonymous codons—synonymous with respect to the translated amino acid—but their mutability in response to the mutagenic environment will depend on the occupying base pair. This difference may vary when mutagens change. Even identical triplets located in different neighboring sequences could respond with different mutabilities to the mutagens present (Gutman and Hatfield 1989).

The existence of a dynamic biological equilibrium of base pairs also involves the relation between molecular evolutionary changes and time. In reality

environmental conditions are hardly ever perfectly steady, because of omnipresent fluctuations. Even nonfluctuating conditions can produce a rapidly changing (chaotic) behavior. Therefore the question arises, how and how fast can an equilibrium be achieved after a period of rapid changes? Does the analogy to the system of thermodynamic equilibria make sense in the region of nonequilibrated states? What is the difference between the evolutionary mutation patterns in a steady state and far from this state? Could it be that substitutions fixed by random drift are only favored during steady periods, whereas in times of rapid changes other pathways involving amplified substitutions are preferred?

Acknowledgments. I am grateful to Prof. Christoph Tamm (Institut für organische Chemie der Universität Basel, Switzerland) who encouraged and enabled me to write this work. I thank Dr. Alfons Hädener (presently at the University Chemical Laboratory, Cambridge, Great Britain) and Dr. Petr Zimák (Physikalisches Institut der Universität Basel, Switzerland) for many invaluable discussions. I also thank Prof. Thomas Bickle (Abteilung Mikrobiologie, Biozentrum der Universität Basel, Switzerland), Dr. William Hunter, Dr. William Cruse, and Prof. Olga Kennard (University Chemical Laboratory, Cambridge, Great Britain), who made the manuscript readable. This investigation was supported by the Swiss National Science Foundation.

References

Beutler E, Gelbart T, Han J, Koziol JA, Beutler B (1989) Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. Proc Natl Acad Sci USA 86:192–196

Cairns J, Overbaugh J, Miller S (1988) The origin of mutants. Nature 335:142–145

Charczuk R, Tamm C, Suri B, Bickle T (1986) An unusual base pairing between pyrimidine and pyridine nucleotides. Nucleic Acids Res 14:9530

Dawkins R (1982) The extended phenotype. The gene as the unit of selection, ed 2 (1987). Oxford University Press, New York, pp 81–96

Discussions about Cairns et al. (1988/1989) in Nature (1988) 335:112–113, 336:21–22, 336:525–528, in Nature (1989) 337: 119–120, 337:123–124

Dohet C, Wagner R, Radman M (1985) Repair of defined single base-pair mismatches in *Escherichia coli.* Proc Natl Acad Sci USA 82:503–505

Dover GA (1987) DNA turnover and the molecular clock. J Mol Evol 26:47–58

Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 10:465–523

Eigen M (1978) The hypercycle. Part B: the abstract hypercycle. Naturwissenschaften 65:7–41

Eigen M, Schuster P (1977) The hypercycle. Part A: emergence of the hypercycle. Naturwissenschaften 11:541–565

Eigen M, Schuster P (1978) The hypercycle. Part C: the realistic hypercycle. Naturwissenschaften 7:341–369

Fowler RG, Schaaper RM, Glickman BW (1986) Characterization of mutational specificity within the lac-1 gene for a mutD5 mutator strain of *Escherichia coli* defective in 3'-5' exonuclease proofreading activity. J Bacteriol 167:130–137

French DL, Laskov R, Scharff MD (1989) The role of somatic hypermutation in the generation of antibody diversity. Science 244:1152–1157

Goodman MF, Branscomb EW (1986) DNA replication fidelity and base mispairing mutagenesis. In: Kirkwood TBL, Rosenberger RF, Galas DJ (eds) Accuracy in molecular processes. Chapman & Hall, London, New York, pp 191–232

Gutman GA, Hatfield GW (1989) Nonrandom utilization of codon pairs in *Escherichia coli.* Proc Natl Acad Sci USA 86: 3699–3703

Ikemura T (1985) Codon usage, tRNA content, and rate of synonymous substitution. In: Ohta T, Aoki K (eds) Population genetics and molecular evolution. Springer-Verlag, Berlin, pp 385–406

Jukes T, Bhushan V (1986) Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. J Mol Evol 24:39–44

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, London

Kuchta RD, Benkovic P, Benkovic SJ (1988) Kinetic mechanism whereby DNA polymerase I (Klenow) replicates with high fidelity. Biochemistry 27:6716–6725

Kunkel TA, Bebenek K (1988) Recent studies of the fidelity of DNA synthesis. Biochim Biophys Acta 951:1–15

Ohno S (1988a) Codon preference is but an illusion created by the construction principle of coding sequences. Proc Natl Acad Sci USA 85:4378–4382

Ohno S (1988b) Universal rule for coding sequence construction: TA/CG deficiency–TG/CT excess. Proc Natl Acad Sci USA 85:9630–9634

Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature 313:762–766

Schaaper RM, Danforth BN, Glickman BW (1986) Mechanisms of spontaneous mutagenesis. An analysis of the spectrum of spontaneous mutation in the *Escherichia coli* lac-1 gene. J Mol Biol 189:273–284

Schaaper RM, Dunn RL (1987) Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction. The nature of in-vivo DNA replication errors. Proc Natl Acad Sci USA 84:6220–6224

Strazewski P (1988) Mispair formation in DNA can involve rare tautomeric forms in the template. Nucleic Acids Res 16: 9377–9398

Strazewski P, Tamm C (1990) Replication experiments with nucleotide base analogues. Angew Chem 102 (in press)

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Tonegawa S (1988) How does an organism manage during its lifetime to respond to a huge number of different antigens. Angew Chem Int Ed Engl 27:1028–1039

Topal MD, Fresco JR (1976) Complementary base pairing and the origin of substitution mutations. Nature 263:285–289

Vogel H, Zuckerkandl E (1971) Randomness and "thermodynamics" of molecular evolution. In: Schoffeniels E (ed) Biochemical evolution and the origin of life. North-Holland, Amsterdam, pp 352–365

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

Zuckerkandl E (1965) Remarques sur l'évolution des polynucléotides comparée à celle des polypeptides. Bull Soc Chim Biol 47:1729–1730

Zuckerkandl E (1987) On the molecular evolutionary clock. J Mol Evol 26:34–46

Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. J Theoret Biol 8:357–366

Received February 1, 1989/Revised and accepted September 15, 1989