Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

Chaos and Order in Spontaneous Mutation

John W. Drake[1]

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709-2233

THIS story starts with one of those unanticipated turning points that mark the lives of most of us. Around 1960 I was working at the University of Illinois at Urbana using Newcastle disease virus (NDV), a tractable animal virus that infects chickens rather than humans. (Working previously with polioviruses, I had acquired extraordinarily high serum titers against all three strains.) The NDV plaque assay required that chicken embryos of the correct age be harvested and dissociated into single cells, which were then used to produce monolayers that became confluent in a few days. Virus samples were adsorbed to the monolayers, which were then incubated for a few more days while plaques formed. The regimen was cumbersome and contained day-long gaps that, as it turned out, were incompletely filled with the planning and analysis of experiments, the preparation and delivery of lectures, and the miscellaneous duties of a bottom-rung academic.

As a graduate student I had followed the antics of the Caltech phage group and had also become aware of the mutagenic base analog 5-bromouracil. Over the next few years I was increasingly impressed by the early work on mutation by Ernst Freese and Seymour Benzer using phage T4. Ernst in particular was dissecting out mutational pathways with mutagens that seemed to be highly specific in their actions, for instance, the base analogs 5-bromouracil and 2-aminopurine that induced transitions in both directions and hydroxylamine that induced only G·C → A·T transitions. I often had occasion to UV irradiate NDV, and I wondered if the mutagenic specificity of UV irradiation could be ascertained by performing reversion tests on UV-induced mutations with base analogs, hydroxylamine, and proflavin, the specificity of the last having been demonstrated by Benzer and Sydney Brenner. Starting with a protocol developed by Raymond Latarjet for phage T4, I began to intercalate phage experiments into the free days provided by the NDV protocols. I quickly found that phages produced interesting results at least an order of magnitude more frequently than did NDV.

Phages and antimutators: There followed a burst of investigations into the mutation process in phage T4. A key component of these investigations was reversion analysis of *rII* mutations, which tended to reveal that the mutants I was making—not only with UV, but also arising spontaneously in free phages and induced by photosensitizers—contained several different kinds of mutations. Here, however, a barrier arose. Transitions of both types could be recognized, as well as many frameshift mutations (which could be reverted by proflavin), but no agent was known to induce only transversions. The answer, I hoped, might lie in an altogether different kind of mutagen. Joe Speyer (1965) had just reported that two temperature-sensitivity mutations in T4 gene *43* (which encodes the viral DNA polymerase) were also strong mutator mutations. Perhaps some such mutators would make transversions specifically. To that end, I acquired the entire Caltech collection of gene-*43* mutants. Elizabeth "B. J." Allen hunted among these for mutators and found many. Unexpectedly, she found others that seemed to be antimutators.[2] These were much more interesting, so we quickly changed course.

Our first report on antimutators was at a Cold Spring Harbor Symposium (Drake and Allen 1968) and contained a formal description of the antimutators plus some general observations based on the literature. One of the latter was that the phages T4 and λ and the bacteria *Escherichia coli* and *Salmonella typhimurium* shared similar genomic mutation rates of roughly 0.2% per chromosome replication. In a huge leap, I wrote that "This result suggests that mutation rates are usually

[1]*Author e-mail:* drake@niehs.nih.gov

[2]I write "seemed to be" because most of the initial antimutator candidates were selected because of deleterious effects of the gene-*43* mutations that rendered the mutants that Allen was scoring difficult to detect, a frequent artifact in the history of mutation research.

about as large as can be tolerated" (p. 339). It was only several years later that I discovered an important article by A. H. Sturtevant (1937) and contemporary (1960s) work by Motoo Kimura, addressing both the issue of downward pressures on mutation rates and the limits of such reductions.

When the formal antimutator story was published soon thereafter, it was divided (by the journal editor) into two parts: one an article about the T4 work (Drake *et al.* 1969) and the other a short note (Drake 1969) on comparative genomic mutation rates in diverse organisms. The note presented the phage/bacteria rate concordances but also cited a rate about tenfold lower for *Neurospora crassa* (which eventually turned out to be an underestimate) and a value for *Drosophila melanogaster* that must have confused many readers because it was a rate (of about 0.8) per sexual generation rather than per germ-line cell division. The same table and its legend allowed astute readers to estimate that the rate per germline cell division was about 1.4%. Although this number was based on several components now known to be inaccurate, the result was close to a later calculation based on somewhat better information. Not yet having discovered Kimura's articles on the subject, I wrote that "Mutation rates may be adjusted by natural selection to achieve a balance among the mostly deleterious effects of mutation, the need for variation, and the cost of suppressing mutation" (p. 1132). Well, two out of three was perhaps not too bad for a beginner; Kimura (1960, 1967) had already pointed out that selection for downmodifiers of mutation rates would be balanced by the cost of further reducing rates and that selection for upmodifiers of mutation rates hitchhiking with rare advantageous mutations was rapidly obviated by recombination.

DNA-based microbes: I soon became attached to the possibility of a universal genomic mutation rate. As reports appeared offering numerical orts that could reasonably be converted into genomic mutation rates, my table was adjusted and expanded and published serially in several chapters and reviews in the 1970s and 1980s. The data of best quality were those for DNA viruses, bacteria, and fungi, and by the early 1990s I was able to produce a compilation that showed that all the DNA-based microbes for which data were available shared a genomic rate of spontaneous mutation of close to 0.003 (mean 0.0033; range 0.0019–0.0046, excluding two formally defined outliers) (Drake 1991). This result fascinated me because it implied that the action of powerful evolutionary forces was capable of finely tuning the rate of spontaneous mutation. Both Kimura and I had imagined a mutation rate balanced between the deleterious effects of most mutations and what Kimura called the "physiological" cost of further reducing mutation rates. However, the defining microbes are extraordinarily diverse in both their genomics and their life histories: a tiny phage with single-stranded DNA, a genetically complex and completely lytic phage pair, a lysogenic phage (whose rate, however, was based on its lytic cycle), a bacterium, a yeast, and a filamentous fungus. Why should such diversity not lead to different optimal rates in such different microbes? I concluded that the balancing forces must be very deep, but they remain mysterious to this day.

In an effort to expand the range of DNA microbes exhibiting this standard rate, more examples were slowly added to the table. The most recent was for a DNA virus that is also a human pathogen, herpes simplex virus (Drake and Hwang 2005). The most bizarre was for an archaeon adapted to growth in hot acid, *Sulfolobus acidocaldarius* (Grogan *et al.* 2001). Both organisms display typical genomic rates, but the Sulfolobus rate of 0.0018 may have been slightly lower than usual for reasons to be discussed later.

Rates per gene and per base pair: Half a century of work in many laboratories has revealed such a plethora of mechanisms that it is now abundantly clear that mutagenesis reflects the sum of many ways in which a complex consortium of fidelity processes can go wrong. This being the case, it is unlikely that any single mutation could lower the genomic mutation rate very much. I argued thusly against the existence of general antimutator mutations of any considerable strength and free of strongly deleterious pleiotropic effects (Drake 1993a) and supported the argument with the experimental demonstration that all of the nine antimutators in phage T4 gene *43* and one in gene *45* exhibit mutation rates the same as or slightly greater than that of the wild type when assayed using reporter genes containing roughly 3600 bp; clearly, while mutation rates at some sites were decreased in these antimutators, rates at other sites were increased. This result had been modestly anticipated by some of the values posted by Drake *et al.* (1969) and strongly anticipated by Lynn Ripley when she developed the first T4 system for scoring transversions (Ripley and Drake 1972; Ripley 1975). In fact, the T4 antimutators all seem to do the same thing: reducing the rate of A·T → G·C transitions while modestly increasing rates of transversions and small indels.[3] In an extensive series of investigations, Roel Schaaper has had a similar experience: antimutators of modest effect can be selected in *E. coli* but they are again rather pathway specific. The only exception that one would expect would be an organism that already harbors a mutator mutation, perhaps not known to the investigator; reversion or suppression of that specific defect could then produce a strong antimutator effect, but would simply lower the rate toward, but not below, the standard value.

[3] Curiously, the term "indel" (meaning insertions and deletions, frequently without regard to size), although in widespread use by evolutionists, is still often unfamiliar to students of the mutation process. It is, of course, superior to "frameshift mutations" because it includes indels of a size that is a multiple of three nucleotides or resident in sequences that do not encode proteins.

Until recently, tabulations of mutation rates in textbooks tended to present a chaotic picture. The rates were rarely genomic, so the overall regularities were obscured; microbial genome sizes vary by over 6000-fold, with the result that genic and average base-pair rates vary by at least that much. Genic rates also vary markedly because of large variations in gene size and irregular content of rapidly mutating base pairs. At the level of individual base pairs, huge variation reigns. This became obvious from the very first T4 *rII* spectrum published by Seymour Benzer (1955) and reflects not only factors such as indel-prone local sequence repeats (Streisinger *et al.* 1966) but also the sequence in which a mutating base resides. My favorite example of the latter also comes from Benzer's T4 *rII* system, where the mutation rate corresponding to ochre → opal conversions (UAA → UGA) varies by more than 1000-fold across the *rII* locus (Salts and Ronen 1971). This result shows not only that base-pair substitution (BPS) rates vary extremely depending on context, but also that the context extends beyond the two neighboring base pairs.

Subsequent work on the proteins conducting DNA synthesis and error avoidance or correction revealed large differences in their impacts on mutagenesis. In extensive studies of polymerase fidelity *in vitro*, Tom Kunkel showed that different DNA polymerases produce strikingly different patterns of differential BPS rates, thus implicating the polymerase, with its many contacts to DNA atoms in the vicinity of a replicating base, as a major determinant of error rates. The proofreading 3′-exonucleases also turned out to exhibit strong site specificities. DNA mismatch repair introduces further variability, although phage T4 happens not to have mismatch repair. The most recently discovered determinant of site-specific mutability is the array of polymerase accessory proteins, including at least the processivity clamp and the protein that binds and coats single-stranded DNA (Bebenek *et al.* 2002, 2005). Why does natural selection allow the persistence of high site-specific mutability, by which some genes experience a majority of their deleterious mutations at a single site? This is really two questions. The first is simply why some sites are so mutable. The answer, I suspect, is that the replication apparatus confronts a vast number of different short sequences and simply cannot evolve to minimize error rates at all of these, particularly when other needs such as velocity and coordination must be addressed. The second question is why selection against hypermutable sequences does not remove mutational hotspots, especially given the high level of degeneracy in the genetic code. Here the answer is less clear. Sometimes there may be constraints on synonymous replacements, but more often the answer may be that selection against high mutability at single sites is simply ineffective.

As Benzer published ever more dense T4 *rII* mutational spectra, it became clear that most sites were never detected by BPSs. Was this because their mutation rates were far below those of the detected sites? Or were the missing sites those at which BPSs simply failed to produce a mutant phenotype? This question was addressed experimentally by Bob Koch (Koch and Drake 1970), a brilliant graduate student who was confined to a wheelchair. Starting with two temperature-sensitive *rII* mutations that had only a slight mutant phenotype at 30°, Bob treated the corresponding mutants with nitrous acid to produce diverse *rII* transition mutations. He then screened the treated population for strong *r* mutants at 30° and constructed mutational spectra. In addition to the characteristic nitrous-acid spectrum, each mutant produced a number of mutations at previously undetected sites scattered over the *rII* locus. When backcrossed away from the starting *rIIts* mutations, the newly induced mutations had so little mutant phenotype that they would almost always have escaped detection. So many new sites were discovered that it was arguable that most or all undetected BPS sites were simply the result of very weak phenotypes.

One day, Bob asked me if he could do a project on the side while working on another that we had started together, but he said nothing about the nature of the project. "Of course," I said. A few months later a draft of an article appeared on my desk. It was amazing: using 2-aminopurine to constrain the mutations to A·T → G·C transitions, he had carried off the first experimental demonstration that BPS mutation rates depend on their sequence context, in this case their neighboring base pairs (Koch 1971). A little later, he discovered an interaction between two *rII* alleles in which one allele modulated both spontaneous and 2-aminopurine-induced mutation rates at the other allele, while recombination tests suggested that the interacting alleles were at least several base pairs apart. Bob died in 1977 from muscular dystrophy, but this work was completed by Mark Conkling (Conkling *et al.* 1980), and Akio Sugino later determined that the alleles were a dozen nucleotides apart (Sugino and Drake 1984). These studies showed that site-specific mutation rates could be affected not only by adjacent but also by somewhat more distant base pairs, which further helped to explain the huge variation in site-specific mutabilities.

BPSs *vs.* indels: Genomic mutation rates are of course the sum of diverse BPSs, small indels, and various large changes. When we characterized the Sulfolobus mutations, we noticed that most of them were small indels. My impression had been that most spectra were dominated by BPSs. A survey of a considerable collection of spectral data confirmed that, most of the time, roughly two-thirds of detected mutations are BPSs, whereas only about one-third were for Sulfolobus (Grogan *et al.* 2001; supplemental Table 4 at http://www.pnas.org/cgi/content/full/141113098/DC1). We had applied both relaxed and stringent selection for our mutants, but the fraction of BPSs in Sulfolobus was indistinguishable in

the two collections, which argued against an unusually high fraction of the BPSs being leaky mutants that escaped detection. It occurred to me that the average missense mutation might be more deleterious at a high temperature than in a mesophile, a notion that structurally oriented colleagues found reasonable but was untested. If so, then an extra mutational load would accrue to BPSs in a thermophile, a cost perhaps justifying the countercost of accumulating additional downmodifiers of BPS mutagenesis. One further result would be to reduce the ratio of missense mutations to silent mutations during the course of molecular evolution. This hypothesis was confirmed by Friedman *et al.* (2004), who found that the ratio of nonsynonymous to synonymous substitutions in diverged pairs of prokaryotes was about 1.7-fold lower in thermophiles than in mesophiles, regardless of whether they were archaeons or eubacteria. An unanticipated bonus from this work was the conclusion, based on an update of previous work from another group (Kollman and Doolittle 2000), that rates of molecular evolution are indistinguishable among eubacteria, archaeon, and eukaryotes.

Riboviruses: As a Caltech graduate student in Renato Dulbecco's animal-virus group, I had become aware from the work of colleagues that polioviruses tended to be highly mutable: the few mutants that could be obtained exhibited high revertant frequencies. The few mutants that were stable were later found to contain multiple mutations. By the 1980s, the high mutability of RNA viruses had become widely recognized (Holland *et al.* 1982) but quantitation was based on rates of evolution and on mutation frequencies rather than on mutation rates.

Bolstered by the robustness of the standard genomic rate for DNA microbes, I decided to see what the riboviruses had to say. Two difficulties quickly became apparent. The first was the lack of a well-defined geometry of viral RNA replication: Was it exponential with a persistently double-stranded intermediate, linear as in the simple iterative copying of Salvador Luria's "stamping machine" (Luria and Delbrück 1943), or a mixture such as rolling-circle replication? To minimize this difficulty, I converted frequencies into rates using the algebra for both exponential and linear replication and took the means of the two values, which usually differed by only a few-fold. The second difficulty was the data themselves. Controls for selection were infrequent, there was apt to be imperfect correspondence between genotype and phenotype, and, most important of all, the reporter sequences were usually tiny, such as a single base-substitution pathway at a single site, and thus subject to the large variability in site-specific mutabilities already well described for genes made of DNA. The result (Drake 1993b) was a list of eight rates from four riboviruses: a coliphage, a poliovirus, an influenza virus, and vesicular stomatitis virus (VSV). The median value of the genomic rates was 2 but the range was a discouraging 0.1–17, so that this analysis showed only that riboviral rates could be estimated and were indeed very high. However, because most of the values seemed too large to sustain viable populations, I suspected an ascertainment bias: perhaps only relatively high mutation frequencies tended to be measured.

This situation did not improve until later in the decade after further explorations of the literature convinced me that riboviruses replicate mainly by the stamping machine mode, the initial copy serving to template numerous complements that in turn each template numerous progeny genomes. However, the simplicity of the resulting algebra for converting mutant frequencies into mutation rates became obvious to me only in 1999 after the first bottle of wine on a flight from Raleigh–Durham to London. I turned for help to Jack Holland, who helped to assemble a set of nine reliable riboviral mutation frequencies from poliovirus, measles virus, rhinovirus, and VSV. The median genomic rate per chromosome replication was 0.76 and the range was 0.13–1.15 (Drake and Holland 1999), still wide but at least narrower than in 1993. The experimental uncertainties precluded any estimates of an average per-base rate among these four viruses. The rate still seemed high to me, because it meant that the average mutation frequency after a single cycle of infection was about 1.5, so that only about a fifth of progeny viruses would be free of newly arisen mutations. Older experiments had shown that ribovirus populations were extinguished by increasing the mutation frequency by about threefold (Holland *et al.* 1990), which would seem to mean that the average mutation frequency would have to be as high as 4.5 to achieve error catastrophe, despite the highly compact nature of riboviral genomes and the involvement of many otherwise codon-degenerate sites in important secondary structures.[4] However, as we shall see below, the average mutation frequency of a ribovirus may conceal important differences among lineages within the population.

The next improvement in understanding riboviral mutation came from tobacco mosaic virus (TMV) (Malpica *et al.* 2002). A strain of tobacco carrying an entire TMV transgene made possible the first ribovirus mutational spectrum wherein most or all deleterious mutations could be recovered and maintained by complementation in the transgenic host. The genomic mutation rate from this system was about 0.1, the lower end of the distribution for animal viruses. The mutational spectrum contained the usual mixture of base substitutions and indels. However, as with Sulfolobus, the base

[4]Retroelements, including retroviruses, seem to have genomic mutation rates several-fold smaller than do riboviruses (Drake *et al.* 1998) and are slightly more resistant to artificial mutagenesis (Pathak and Temin 1992). The retroviral rate is the average of three very different kinds of replication (transcription of a provirus, then RNA → DNA, and then DNA → DNA), each with unknown individual contributions to the average rate.

substitutions were in a minority (11/35), which might reflect a higher average deleterious impact of substitutions in these genomes because of the major role of RNA secondary structure in ribovirus life histories, often rendering even synonymous mutations deleterious. The spectrum also contained several examples of a class of mutations that I believe had not been recorded previously, poly(A) insertions of 4–84 residues. Given the mutation frequency, a surprising number of the mutants contained two or three mutations, an anomaly to be considered later.

It has been common practice to attribute the high mutation rates of riboviruses to their need to escape host defenses. To me, this always seemed unlikely because the RNA coliphage Qβ (genome ≈ 4200 bases) has a characteristic RNA mutation rate (Drake 1993b) while the DNA coliphage M13 (genome ≈ 6500 bases) has a characteristically DNA mutation rate about 100-fold lower (Drake 1991). Very recently, Furió *et al.* (2005) used a VSV system to conclude that the magnitudes of mutation rates did not correlate with rates of adaptation. They suggested that viral adaptation is not limited by mutant frequencies but rather that high replication rates result in low replication fidelity. However, another report (Vignuzzi *et al.* 2006) reached a very different conclusion, namely that poliovirus adaptation can be limited both by the mutation rate and by the quasi-species dynamics of the population. Thus, the adaptive value of high riboviral mutation rates remains an open question.

It is also common practice to attribute the high mutation rates of riboviruses to the lack of a proofreading 3′-exonuclease and, of course, to the lack of mismatch repair, which is less often mentioned. In contrast, it has long seemed more probable to me that riboviral life histories drive their high mutation rates, the particular mechanisms simply reflecting how the preferred values are achieved. There is a good example from DNA-based organisms. Phage T4 and *E. coli* differ in average mutation rate per base pair by about 30-fold. Mechanistically, this difference reflects the absence of mismatch repair in T4 plus a modestly higher rate of proofreading than in *E. coli*, but the driving force behind the higher T4 rate is surely not lack of mismatch repair but rather the deeper evolutionary forces that favor the standard genomic microbial rate. If lower riboviral mutation rates were adaptive, that is, if their costs were more than countered by their benefits, they could probably be achieved quite easily by increasing the fidelity of base selection and/or by adding proofreading. Indeed, a kind of proofreading has been observed during transcription (Erie *et al.* 1993) and might be acquired inexpensively by a ribovirus. What, then, is the component of the riboviral life history that drives the observed high mutation rates? I suspect a specific chemical hazard: RNA is highly susceptible to backbone hydrolysis at even moderate temperatures, and this instability is enhanced by traces of common metal ions such as magnesium, calcium, iron, and zinc (Eigner *et al.* 1961; Lindahl 1967, 1993; Larralde *et al.* 1995). Indeed, it is notoriously difficult to recover and purify intact riboviral genomes. As a result, high replication rates and large numbers of progeny will be required and, concomitantly, high mutations rates may become inevitable.

Higher eukaryotes: Sequence-based mutational spectra are needed for estimating genomic mutation rates because only such spectra can provide confidence that the reporter gene is mutating in an unexceptional manner. Unfortunately, good germ-line spectra simply do not yet exist for higher eukaryotes. In 1998 I undertook a collaboration with Jim Crow and Brian and Deborah Charlesworth to survey mutation rates in all possible groups of organisms (Drake *et al.* 1998). (I originally proposed to entitle this article "Rates of Spontaneous Mutation from Microbes to Man" until I was convinced by my co-authors that this rate was immeasurably small.) Relying on very incomplete information, Jim and I guesstimated haploid rates for four animals: *Caenorhabditis elegans*, *D. melanogaster*, *Mus musculus*, and *Homo sapiens*. Realizing that selection acting on mutation rates would be indifferent to mutations occurring in DNAs lacking mutation-sensitive functions, we proposed to use an "effective" genome size consisting of those base pairs in which most mutations mattered, that is, were deleterious.[5] Genomic rates estimated for total genomes, or for effective genomes per sexual generation, varied considerably among these animals, but rates per effective genome per germline cell division showed surprisingly less variation (values of 0.004, 0.005, 0.014,[6] and 0.004, respectively). However, the similarity of three of these values to the rate for DNA-base microbes is probably misleading, because our values were not adjusted for mutations of small impact and are probably substantially underestimated. If there is such a standard rate for higher eukaryotes, it will probably not be defined soon. For instance, there is increasing evidence that purifying selection acts on large blocks of noncoding DNA, which tends to increase estimates of the size of the effective genome.

Why should mutation rates be higher rather than lower in these animals compared to DNA microbes, especially when effective genome mutation rates summed over an entire sexual generation often generate values in the neighborhood of 1? The most common response

[5]As far as I know, the effective genome size in microbes is generally fairly close to 1 (*e.g.*, ≥0.8). However, the proportion of synonymous changes that is deleterious rises not only in thermophiles, but also particularly in riboviruses where extensive secondary structure is important to gene regulation and to resistance to cellular nucleases.

[6]Even though these values are not expected to be particularly accurate, this value, for the mouse, might reflect a true difference because the rate of synonymous substitutions in the mouse over recent evolutionary time appears to be roughly 3.6-fold higher than in other major groups of animals (*e.g.*, Bulmer *et al.* 1991).

is that the impact of mutations should be lower in a diploid than in a haploid, thus relaxing selection for downmodifiers of mutation rates. Certainly most of the haploid phases in these organisms enjoy sheltered lives compared to, for instance, yeast. Thus, a role for the sheltering nature of diploidy is a reasonable conjecture, but one deserving of experimental exploration.

Students of mutation have long hoped for a more objective measure of mutation frequency, one based directly on changes in DNA sequence rather than on phenotypes. However, a recent such report (Denver *et al.* 2004) was disconcerting. It described a genomic mutation rate in *C. elegans* corresponding to 0.24 per germ-line cell division, about 13-fold higher than the corresponding rate estimated by Drake *et al.* (1998). In addition, the mutations themselves exhibited differences from general trends. First, BPSs were a 13/30 minority, different from the usual 2/3 majority and hugely different from the 12/13 majority expected after adjusting for a roughly fivefold excess of undetected BPSs in the more conventional spectra (Grogan *et al.* 2001; supplemental Table 3 at http://www.pnas.org/cgi/content/full/141113098/DC1). Second, among the smaller indels there was a 12/14 majority of additions instead of the more common majority of deletions. Third, the pattern of molecular evolution in *C. elegans* is also discordant with some of their results. Thus, it will be important to determine whether these substantial differences reflect the true rate and pattern of mutation in *C. elegans*, or whether the method is not yet the long-sought paradigm.

When reading early articles about mutation in cultured mammalian somatic cells, I used to worry about whether the culture conditions (hypoxia, mutagenic fluorescent lighting, and various medium components) might themselves be mutagenic, and I am still unsure how well this question has been resolved. In any case, that question led to the more interesting one of whether mutation rates might be lower in germ cells than in somatic cells, a reasonable evolutionary conjecture. When mutation frequencies began to become available from studies using mice carrying *E. coli lacI* mutation-reporter transgenes, it soon became clear that testicular samples rich in germ cells averaged about threefold fewer mutations than were scored in somatic cells from diverse organs. At a mid-1990s meeting in Japan where some such data reemerged, I asked Barry Glickman if he knew the fraction of germ cells in the samples called "testes" and he responded that it might be about 80%. A quick calculation suggested that pure male germ cells would then have mutation frequencies roughly sixfold lower than somatic cells. An article soon appeared with the answer: a tenfold difference (Walter *et al.* 1998). This appeared to be a dramatic fine tuning of mutation rates, but certain details were even more provocative. The average mutation frequency for somatic cells was 4.8×10^{-5} and for spermatocytes plus spermatids plus spermatozoa it was 0.48×10^{-5}, but for type A spermatogonia the frequency was an intermediate 2.2×10^{-5}. How can a mutation frequency be *reduced* by a factor of five? There is a wave of apoptosis between type A and type B spermatogonia, but how could apoptosis selectively kill *lacI*-mutant cells? Or could there be some other kind of selection against such mutant cells, as occurs (as selection for rather than against) in the case of human male germ cells carrying *FGFR2* mutations (Goriely *et al.* 2003)? This is clearly a fundamental question ripe for investigation.

Yet another parameter along which mutation rates vary is developmental stage. More than 4 decades ago, *Saccharomyces cerevisiae* was observed to undergo a burst of hypermutation during meiosis (Magni and von Borstel 1962; Magni 1963). Other than associating these mutations with nearby recombinational exchanges, this phase of transient hypermutation remains undissected. Much more recently, many years of scoring spontaneous mutations in the mouse also revealed a burst of hypermutation approximately during meiosis (Russell and Russell 1996; Russell 1999). If the mutation rate per sexual generation receives a large contribution during such a narrow window, then the calculations for animals presented in Drake *et al.* (1998) would be unbalanced. However, if this bias were limited to the mouse, then only that value would be enhanced, and the mouse value does in fact appear to be considerably higher than the other values.

Too many mutants with multiple mutations: During the course of several rather large experiments exploring the structural basis of DNA polymerase fidelity, we noticed that some mutants produced *in vivo* and many produced *in vitro* contained two or more mutations. It was almost immediately evident that the multiple mutants ("multiples") were usually more frequent than expected from the mutant frequency. (If the mutant frequency is F and the mutations are randomly dispersed according to a Poisson distribution, then the average number of mutations per mutational target is $a = -\ln(1-F)$. If M mutants are sequenced, the expected number of doubles is $E_2 = Ma^2e^{-a}/2F$, the expected number of triples is $E_3 = E_2a/3$, and so on.) When I explored numerous published spectra, it soon became clear that most contained too many multiples (Drake *et al.* 2005). This is seen with diverse DNA polymerases *in vitro*, with numerous viral and cellular microbial systems, with cultured mammalian cells, and with mouse and human tissues. There were strong indications that the multiples were not produced by a background of mutator mutants, but by transient phenotypic hypermutation.

Populations with too many multiples can be modeled as the sum of two or more subpopulations: a majority with a lower mutation frequency and one or more minorities with much higher mutation frequencies. This skewed distribution may impact important processes. For instance, some prospective paths of adaptation will

contain mutations that are neutral or deleterious individually but beneficial when combined, and these paths are much more efficiently traversed by mutational clusters than by sequential single mutations. Human cancers require more mutations than can be produced by the typical mutation rate. Many cancers seem to acquire a mutator mutation early in their lineage, but others do not and may be initiated by bouts of transient hypermutation. In most organisms, the contribution of transient hypermutation to the average mutation rate is small, but in TMV it appears to be large, so that the majority of the population has a mutation frequency substantially lower than the average. As a result, ribovirus populations may be somewhat more genetically stable than estimated previously.

Mechanisms of transient hypermutation have been explored in stationary-phase bacterial cells but not elsewhere. They are likely to involve diverse accidents of genome metabolism, such as unstable mutator polymerases caused by errors of transcription, translation, or folding. As I approach the start of a fourth quarter-century of life (hopefully linearly rather than asymptotically), I look forward to exploring these mechanisms.

I thank Marilyn Diaz and Mike Resnick for their insightful comments on the manuscript. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences.

LITERATURE CITED

Bebenek, A., G. T. Carver, H. K. Dressman, F. A. Kadyrov, J. K. Haseman *et al.*, 2002 Dissecting the fidelity of bacteriophage RB69 DNA polymerase: site-specific modulation of fidelity by polymerase accessory proteins. Genetics **162:** 1003–1018.

Bebenek, A., G. T. Carver, F. A. Kadyrov, G. E. Kissling and J. W. Drake, 2005 Processivity clamp gp45 and ssDNA-binding-protein gp32 modulate the fidelity of bacteriophage RB69 DNA polymerase in a sequence-specific manner. Genetics **169:** 1815–1824.

Benzer, S., 1955 Fine structure of a genetic region in bacteriophage. Proc. Natl. Acad. Sci. USA **41:** 344–354.

Bulmer, M., K. H. Wolfe and P. M. Sharp, 1991 Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. Proc. Natl. Acad. Sci. USA **88:** 5974–5978.

Conkling, M. A., R. E. Koch and J. W. Drake, 1980 Determination of mutation rates in bacteriophage T4 by unneighborly base pairs: genetic analysis. J. Mol. Biol. **143:** 303–315.

Denver, D. R., K. Morris, M. Lynch and T. W. Kelley, 2004 High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature **430:** 679–682.

Drake, J. W., 1969 Comparative rates of spontaneous mutation. Nature **221:** 1132.

Drake, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

Drake, J. W., 1993a General antimutators are improbable. J. Mol. Biol. **229:** 8–13.

Drake, J. W., 1993b Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. USA **90:** 4171–4175.

Drake, J. W., and E. F. Allen, 1968 Antimutagenic DNA polymerases of bacteriophage T4. Cold Spring Harbor Symp. Quant. Biol. **33:** 339–344.

Drake, J. W., and J. J. Holland, 1999 Mutation rates among lytic RNA viruses. Proc. Natl. Acad. Sci. USA **96:** 13910–13913.

Drake, J. W., and C. B. C. Hwang, 2005 On the mutation rate of herpes simplex virus type 1. Genetics **170:** 969–970.

Drake, J. W., E. F. Allen, S. A. Forsberg, R.-M. Preparata and E. O. Greening, 1969 Genetic control of mutation rates in bacteriophage T4. Nature **221:** 1128–1132.

Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics **148:** 1667–1686.

Drake, J. W., A. Bebenek, G. E. Kissling and S. Peddada, 2005 Clusters of mutations from transient hypermutability. Proc. Natl. Acad. Sci. USA **102:** 12849–12854.

Eigner, J., H. Boedtker and G. Michaels, 1961 The thermal degradation of nucleic acids. Biochim. Biophys. Acta **51:** 165–168.

Erie, D. A., O. Hajiseyedjavadi, M. C. Young and P. H. von Hippel, 1993 Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. Science **262:** 867–873.

Friedman, R., J. W. Drake and A. L. Hughes, 2004 Genome-wide patterns of nucleotide substitution reveal stringent functional constraints on the protein sequences of thermophiles. Genetics **167:** 1507–1512.

Furió, V., A. Moya and R. Sanjuán, 2005 The cost of replication fidelity in an RNA virus. Proc. Natl. Acad. Sci. USA **102:** 10233–10237.

Goriely, A., G. A. T. McVean, M. Röjmyr, B. Ingemarsson and A. O. M. Wilkie, 2003 Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. Science **301:** 643–646.

Grogan, D. W., G. T. Carver and J. W. Drake, 2001 Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. Proc. Natl. Acad. Sci. USA **98:** 7928–7933.

Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol *et al.*, 1982 Rapid evolution of RNA genomes. Science **215:** 1577–1585.

Holland, J. J., E. Domingo, J. C. de la Torre and D. A. Steinhauer, 1990 Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. J. Virol. **64:** 3960–3962.

Kimura, M., 1960 Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. J. Genet. **57:** 21–34.

Kimura, M., 1967 On the evolutionary adjustment of spontaneous mutation rates. Genet. Res. **9:** 23–34.

Koch, R. E., 1971 The influence of neighboring base pairs upon base-pair substitution mutation rates. Proc. Natl. Acad. Sci. USA **68:** 773–776.

Koch, R. E., and J. W. Drake, 1970 Cryptic mutants of bacteriophage T4. Genetics **65:** 379–390.

Kollman, J. M., and R. F. Doolittle, 2000 Determination of the relative rates of change for prokaryotic and eukaryotic proteins with anciently duplicated paralogs. J. Mol. Evol. **51:** 173–181.

Larralde, R., M. P. Robertson and S. L. Miller, 1995 Rates of decomposition of ribose and other sugars: implications for chemical evolution. Proc. Natl. Acad. Sci. USA **92:** 8158–8160.

Lindahl, T., 1967 Irreversible heat inactivation of transfer ribonucleic acids. J. Biol. Chem. **242:** 1970–1973.

Lindahl, T., 1993 Instability and decay of the primary structure of DNA. Nature **362:** 709–715.

Luria, S. E., and M. Delbrück, 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28:** 491–511.

Magni, G. E., 1963 The origin of spontaneous mutations during meiosis. Proc. Natl. Acad. Sci. USA **50:** 975–980.

Magni, G. E., and R. C. von Borstel, 1962 Different rates of spontaneous mutation during mitosis and meiosis in yeast. Genetics **47:** 1097–1108.

Malpica, J. M., A. Fraile, I. Moreno, C. I. Obies, J. W. Drake *et al.*, 2002 The rate and character of spontaneous mutation in an RNA virus. Genetics **162:** 1505–1511.

Pathak, V. K., and H. M. Temin, 1992 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate. J. Virol. **66:** 3093–3100.

Ripley, L. S., 1975 Transversion mutagenesis in bacteriophage T4. Mol. Gen. Genet. **141:** 23–40.

Ripley, L. S., and J. W. Drake, 1972 A genetic assay for transversion mutations in bacteriophage T4. Mol. Gen. Genet. **118:** 1–10.

Russell, L. B., 1999 Significance of the perigametic interval as a major source of spontaneous mutations that result in mosaics. Environ. Mol. Mutagen. **34:** 16–23.

Russell, L. B., and W. L. Russell, 1996 Spontaneous mutations recovered as mosaics in the mouse specific-locus test. Proc. Natl. Acad. Sci. USA **93:** 13072–13077.

Salts, Y., and A. Ronen, 1971 Neighbor effects in the mutation of *ochre* triplets in the T4*rII* gene. Mutat. Res. **13:** 109–113.

Speyer, J. F., 1965 Mutagenic DNA polymerase. Biochem. Biophys. Res. Commun. **21:** 6–8.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita *et al.*, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

Sturtevant, A. H., 1937 Essays on evolution. I. On the effects of selection on mutation rate. Q. Rev. Biol. **12:** 464–467.

Sugino, A., and J. W. Drake, 1984 Modulation of mutation rates in bacteriophage T4 by a base-pair change a dozen nucleotides removed. J. Mol. Biol. **176:** 239–249.

Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron and R. Andino, 2006 Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature **439:** 344–348.

Walter, C. A., G. W. Intano, J. R. McCarrey, C. A. McMahan and R. B. Walter, 1998 Mutation frequency declines during spermatogenesis in young mice but increases in old mice. Proc. Natl. Acad. Sci. USA **95:** 10015–10019.