

Efficiency of carcinogenesis with and without a mutator mutation

Robert A. Beckman*[†] and Lawrence A. Loeb[‡]

*Department of Clinical Research and Development, Hematology/Oncology, Centocor, Inc., Malvern, PA 19355-1307; and [‡]Department of Pathology, University of Washington School of Medicine, Mail Stop SM-30, Seattle, WA 98195-7470

Communicated by Manfred Eigen, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, July 25, 2006 (received for review April 7, 2006)

Carcinogenesis involves the acquisition of multiple genetic changes altering various cellular phenotypes. These changes occur within the fixed time period of a human lifespan, and mechanisms that accelerate this process are more likely to result in clinical cancers. Mutator mutations decrease genome stability and, hence, accelerate the accumulation of random mutations, including those in oncogenes and tumor suppressor genes. However, if the mutator mutation is not in itself oncogenic, acquiring that mutation would add an extra, potentially time-consuming step in carcinogenesis. We present a deterministic mathematical model that allows quantitative prediction of the efficiency of carcinogenesis with and without a mutator mutation occurring at any time point in the process. By focusing on the ratio of probabilities of pathways with and without mutator mutations within cell lineages, we can define the frequency or importance of mutator mutations in populations independently of absolute rates and circumvent the question of whether mutator mutations are “necessary” for cancers to evolve within a human lifetime. We analyze key parameters that predict the relative contribution of mutator mutants in carcinogenesis. Mechanisms of carcinogenesis involving mutator mutations are more likely if they occur early. Involvement of mutator mutations in carcinogenesis is favored by an increased initial mutation rate, by greater fold-increase in mutation rate due to the mutator mutation, by increased required steps in carcinogenesis, and by increased number of cell generations to the development of cancer.

cancer | mathematical model | mutator hypothesis

The development of a cancer requires multiple steps as suggested by epidemiologic data and animal models (1–5). At least some of these steps are believed to involve oncogenic mutations or other genetic changes.

The mutator hypothesis originally stated that mutations in DNA polymerases and DNA repair enzymes would play a critical role in carcinogenesis by accelerating the acquisition of oncogenic mutations (6). This concept has been broadened to include a variety of mutator mutations that create genetic instability by mechanisms including microsatellite instability, chromosomal instability, and aberrations of checkpoint control (7–10).

An alternative hypothesis is that enhanced rates of genetic instability are not necessary for carcinogenesis and that cancer arises from mutations occurring at the normal rate followed by multiple rounds of lineage selection and expansion (11–14).

These two hypotheses are not mutually exclusive. Indeed it is likely that carcinogenic pathways corresponding to both hypotheses contribute concurrently. For example, carcinogenesis could involve a mutator mutation and subsequent mutations augmented by lineage selection and expansion.

The common outcome of a malignant lineage is likely approachable by numerous pathways, and we postulate that these pathways will be seen in clinical cancers in proportion to their relative efficiencies, irrespective of whether a given pathway occurs at an absolute rate that could potentially account for observed cancer incidence. Thus, the question of whether mutator mutations are “necessary” can be circumvented.

In this paper, we calculate probabilities of cancer arising at a fixed time within a single cell lineage by pathways with and without a mutator mutation. By focusing on the ratio of probabilities of pathways with and without mutator mutations within cell lineages, we can define the frequency or importance of mutator mutations in populations independent of absolute carcinogenesis rates.

Mutator mutations accelerate the acquisition of oncogenic mutations but need not, in themselves, be oncogenic. In cases in which mutator mutations are not oncogenic, they impose an additional delay required for their acquisition. We evaluate this tradeoff in this paper.

Mutator lineages also are at greater risk of accumulating deleterious mutations that reduce their fitness and could lead to extinction [negative clonal selection (NCS)]. This effect has been studied mathematically (15) and found to be relatively small in most situations.

Simplifying Assumptions

We have used a number of simplifying assumptions to establish a mathematical model for analyzing the role of mutator mutations in carcinogenesis. These assumptions and their limitations are discussed below.

1. We assume that the mutation rate, or probability per unit time in oncogenic or mutator loci, is similar to the average mutation rate throughout the genome and that it is relatively uniform except as affected by mutator mutations. It is nonetheless known that mutation hot spots do indeed occur at specific loci throughout the genome. However, hot spots are not known to be characteristic of oncogenic or mutator loci.
2. It is assumed that a critical number of mutation steps are required for cancer development. Mutations in recessive oncogenes are counted as two steps. This critical number of steps may vary among individual tumor types, as classified either by organ site and histology or by modern molecular classification methods.
3. It is assumed that the critical number of mutations may be acquired in any order and in any combination. In fact, for colon cancer and melanoma there is likely a specific order of acquisition of critical mutations (16, 17), and in animal models only certain combinations of oncogenic mutations lead to malignant transformation (4–5). These assumptions do not affect the key calculation of this paper, the relative efficiency of mutator and nonmutator mechanisms, although they do affect the calculation of absolute rates.
4. The model assumes that the sets of oncogenic and mutator mutation loci are mutually exclusive, which is known not to be true in that the retinoblastoma and p53 gene products are both mutator genes and recessive oncogenes (18). This

Conflict of interest statement: No conflicts declared.

Abbreviation: NCS, negative clonal selection.

[†]To whom correspondence should be addressed. E-mail: eniac1@snip.net.

© 2006 by The National Academy of Sciences of the USA

simplifying assumption will lead to a slight underestimate of the importance of mutator mutations in carcinogenesis, because it does not consider a subclass of mutator mutations, such as the retinoblastoma gene product.

5. We assume that cell proliferation and death are balanced for each cell lineage until the critical number of oncogenic mutations are attained, at which point the lineage is highly selected and expands. Expansion of lineages with only a subset of critical mutations (“pre-malignant” lineages) is not an explicit part of the model. Mutations affecting cellular fitness may be cooperative in their effects such that the most dramatic increase in cellular fitness occurs only at the point of malignant transformation. The degree to which this assumption is true is unknown.
6. We do not need to assume that the probability of mutation in any given nucleotide locus during the lifetime of a single cell and its progeny is considerably less than 1. This assumption is clearly true for a single cell and its progeny in the absence of lineage expansion as for any given cell most nucleotide loci are unmutated. It would not be true in reference to an overall sufficiently large population of N cells, where N is equal to or greater than the reciprocal of the per nucleotide mutation rate per cell generation times the number of cell generations (19), such as would occur in the case of lineage expansion. However, in that instance the model can be reformulated to address the question of which mechanism will produce the most malignant cells. Relative efficiency would be defined as the “expectation value,” i.e., the mean number of expected malignant cells, of one mechanism compared with another. This definition of relative efficiency may be relevant *in vivo*, where not every malignant cell necessarily leads to a clinical cancer. Therefore, mechanisms that produce more malignant cells are more likely to lead to clinical cancer. In this regard, it is feasible to broaden this type of model to include pre-malignant lineage expansion.
7. NCS, i.e., selective extinction of mutator lineages due to accumulation of mutations reducing their fitness, is not considered. An analysis of this effect suggests that it is not quantitatively significant over a broad range of parameter values (15).
8. Only dominant mutator mutations are considered. It is assumed that recessive mutator mutations will be less quantitatively significant because they require two hits to occur. This assumption is different from that for oncogenic mutations, for which we assume that some of the critical oncogenic mutations will be in tumor suppressor genes, therefore requiring two mutations to alter the cellular phenotype.
9. Reversions are expected to be quantitatively significant only at very high mutation frequencies and are not considered.
10. Once the critical number of oncogenic mutations are acquired, the lineage is assumed committed to cancer and will produce a clinical cancer with a constant rate and probability. Whether or not the mutation steps required for commitment are rate limiting, the ratio of clinical cancers occurring by different mechanisms is determined by their ratio at the point of commitment, which is evaluated at fixed time points chosen to correspond to the typical times of onset of most cancers.

Key Parameters and Their Values

N_{OL} is the number of oncogene loci, the mutation of which could lead to cancer, on a nucleotide basis. N_{OL} is assumed to be 10^2 , of the approximate order of the number of published oncogenes (4). For dominant oncogenes, we assume that mutations in only a small number of loci result in a gain of function. Although absolute probabilities depend on this number, relative probabilities of mutator and nonmutator pathways do not.

C is the number of oncogene mutations required to produce a cancer and is varied in this paper from 2, as shown in the case of retinoblastoma (20), to 12, as is suggested by epidemiologic data in the case of prostate cancer (21). Based on studies of metabolic pathways involved in cellular transformation, as well as the epidemiologic data, a number of ≈ 6 is a reasonable expectation (4–5, 22).

k_{mut} is the spontaneous mutation rate or probability per cell division of a mutation in any one nucleotide locus, estimated at 10^{-11} for stem cells, which are believed to give rise to cancers (23), to 10^{-9} in somatic cells (24). k_{mut} is varied within that range in this paper.

T is the number of cell generations at which the relative efficiency of different mechanisms is evaluated, based on typical times to onset of cancer and cell cycle kinetics. T is varied in this paper from 170 for tissues with a low-cycling fraction (15, 25, 26) to 5,000 for highly proliferative tissues, such as colonic epithelium (27).

p_{mut} is the spontaneous mutation frequency or probability in the lifetime of the cell lineage that any given critical nucleotide locus in a gene is mutated in the absence of a mutator mutation. (If $k_{mut}T \ll 1$, then $p_{mut} \approx k_{mut}T$; see Appendix).

The time, in cell generations, when a mutator mutation occurred if it contributed to the pathogenesis of the cancer is t , where $0 < t < T$.

N_{ML} is the number of nucleotides corresponding to critical mutator loci in which a mutator mutation can occur. This parameter affects the calculations of relative efficiencies. Studies in yeast have identified ≈ 30 potential mutator genes (28), and we assume ≈ 100 for humans. We vary the number of critical nucleotide loci per mutator gene from 1 to 10. The latter value is still a very conservative estimate given data on loss-of-function mutations (29); therefore, N_{ML} varies from 100 to 1,000 in this work. If a less conservative, i.e., higher number, were assumed, the estimated relative importance of mutator mechanisms would be increased.

α is the multiplicative factor by which a mutator mutation increases p_{mut} and k_{mut} , where $\alpha > 1$. Mutations in base selection and proofreading generally increase mutation rates 10- to 100-fold (30, 31) but can increase them up to 50,000-fold (31). Thus, after a mutator mutation the spontaneous mutation rate within a lineage would be αk_{mut} .

The number of the mutation step at which the mutator mutation occurs in the pathogenesis of the tumor is termed k , $1 \leq k \leq C$. That is, the mutator mutation can occur as the first step in the pathogenesis of the tumor ($k = 1$), as the second step after one oncogene mutation, ($k = 2$), all the way up to the second to last step after $C - 1$ oncogene mutations ($k = C$).

Results

We see that the relative importance of mutator mutations increases with increasing fold-increase α in mutation rate, increasing number of nucleotide loci available for a mutator mutation N_{ML} , increasing number of cell generations T , increasing initial mutation rate k_{mut} , and increasing number of steps C required before commitment to cancer. Depending on the other factors, especially C , a significant increase α in the mutation rate could be required for increased carcinogenic efficiency to be realized from a mutator mutation.

Table 1 shows the value of $\alpha_{50\%}$, or the minimum fold increase in error rate at which a mutator mutation will contribute to at least 50% of the observed cancers, for different values of T and C assuming the initial mutation rate $k_{mut} = 10^{-11}$ and the number of nucleotide loci available for a mutator mutation $N_{ML} = 100$ or 1,000. Table 1 also shows the same information assuming $k_{mut} = 10^{-9}$. All values of $\alpha_{50\%}$ are within the same orders of magnitude observed for known mutator mutations.

Table 1. $\alpha_{50\%}$ for various values of T and C

T at various N_{ML} and k_{mut}	C							
	2	3	4	5	6	8	10	12
$N_{ML} = 100/k_{mut} = 10^{-11}$								
170	4,200	287	74	32	19	9.2	6.0	4.5
5,000	775	93	32	16	11	6.1	4.3	3.4
$N_{ML} = 1,000/k_{mut} = 10^{-11}$								
170	1,330	133	41	20	13	6.9	4.8	3.7
5,000	245	43	18	10	7.2	4.5	3.4	2.8
$N_{ML} = 100/k_{mut} = 10^{-9}$								
170	420	62	23	13	8.6	5.2	3.8	3.1
5,000	77	20	10	6.5	4.9	3.4	2.7	2.3
$N_{ML} = 1,000/k_{mut} = 10^{-9}$								
170	133	72	13	8.1	5.9	3.9	3.0	2.6
5,000	24	9.3	5.6	4.1	3.3	2.5	2.2	1.9

The results were calculated by using Eq. 9. The calculations assume that $N_{ML} = 100$ or $1,000$, as indicated. It is further assumed that $k_{mut} = 10^{-11}$ or 10^{-9} , as indicated.

We see as the number of steps required in carcinogenesis decreases, the threshold for α at which a mutator mutation is important in carcinogenesis increases dramatically. Thus, mutator mutations are more likely to play a role as more steps are required in carcinogenesis. Table 1 also shows that the effect of increased T and increased wild-type mutation rate in favoring mutator mechanisms is greater when the number of mutations required for a cancer is small. For example, we see that, for $C = 2$, $\alpha_{50\%}$ decreases 5- to 6-fold, from 4,200 to 775, when T increases from 170 to 5,000. However, for $C = 12$, the same change in T leads to only a 24% decrease in $\alpha_{50\%}$, from 4.5 to 3.4.

Table 2 shows the ratio $P_{rel, mut:no mut}$ of cancers arising with a mutator mutation in their pathogenesis to those without a mutator mutation for various values of T , C , and α , assuming the initial mutation rate $k_{mut} = 10^{-11}$ and the number of nucleotide loci available for a mutator mutation $N_{ML} = 100$ or

$N_{ML} = 1,000$. Also shown is the same information assuming $k_{mut} = 10^{-9}$. A ratio of >1 indicates a predominance of cancers arising with a mutator mutation, and a ratio of <1 indicates a predominance of cancers arising without it. The fraction of cancers arising with a mutator mutation in their pathogenesis is given by $P_{rel, mut:no mut}/(1 + P_{rel, mut:no mut})$.

A selection of these results is also represented graphically in Fig. 1, in which the common logarithm of $P_{rel, mut:no mut}$ is plotted on the y axis vs. the common logarithm of α on the x axis for values of $C = 2, 6, \text{ or } 12$ and for $T = 170$ or $5,000$, assuming $k_{mut} = 10^{-11}$ and $N_{ML} = 100$ (Fig. 1A) or $k_{mut} = 10^{-9}$ and $N_{ML} = 1,000$ (Fig. 1B). As can be seen by taking the logarithm of Eq. 8, these curves represent a family of lines with slope C . We note that the lines in Fig. 1B are parallel to those in Fig. 1A, and indeed any 10-fold increase in either k_{mut} or N_{ML} moves the curve up by one logarithm unit without changing its slope. Thus, the

Table 2. $P_{rel, mut:no mut}$ at various values for T , C , and α

α at various N_{ML} and k_{mut}	$T = 170$			$T = 5,000$		
	$C = 12$	$C = 6$	$C = 2$	$C = 12$	$C = 6$	$C = 2$
$N_{ML} = 100/k_{mut} = 10^{-11}$						
10	1.5×10^4	2.7×10^{-3}	6.3×10^{-6}	4.3×10^5	0.79	1.9×10^{-4}
20	5.6×10^7	1.6	2.4×10^{-5}	1.7×10^9	48	7.0×10^{-4}
100	1.3×10^{16}	2.5×10^4	5.7×10^{-4}	3.9×10^{17}	7.2×10^5	1.7×10^{-2}
1,000	1.3×10^{28}	2.4×10^{10}	5.7×10^{-2}	3.9×10^{29}	7.2×10^{11}	1.7
$N_{ML} = 1,000$ and $k_{mut} = 10^{-11}$						
10	1.5×10^5	2.7×10^{-2}	6.3×10^{-5}	4.3×10^6	7.9	1.9×10^{-3}
20	5.6×10^8	16	2.4×10^{-4}	1.7×10^{10}	480	7.0×10^{-3}
100	1.3×10^{17}	2.5×10^5	5.7×10^{-3}	3.9×10^{18}	7.2×10^6	0.17
1,000	1.3×10^{29}	2.4×10^{11}	0.57	3.9×10^{30}	7.2×10^{12}	17
$N_{ML} = 100$ and $k_{mut} = 10^{-9}$						
10	1.5×10^6	0.27	6.3×10^{-4}	4.3×10^7	79	1.9×10^{-2}
20	5.6×10^9	160	2.4×10^{-3}	1.7×10^{11}	4.8×10^3	7.0×10^{-2}
100	1.3×10^{18}	2.5×10^6	5.7×10^{-2}	3.9×10^{19}	7.2×10^7	1.7
1,000	1.3×10^{30}	2.4×10^{12}	5.7	3.9×10^{31}	7.2×10^{13}	170
$N_{ML} = 1,000$ and $k_{mut} = 10^{-9}$						
10	1.5×10^7	2.7	6.3×10^{-3}	4.3×10^8	790	0.19
20	5.6×10^{10}	1.6×10^3	2.4×10^{-2}	1.7×10^{12}	4.8×10^4	0.7
100	1.3×10^{19}	2.5×10^7	0.57	3.9×10^{20}	7.2×10^8	17
1,000	1.3×10^{31}	2.4×10^{13}	57	3.9×10^{32}	7.2×10^{14}	1.7×10^3

The results were calculated by using Eq. 8. The calculations assume $N_{ML} = 100$ or $1,000$, as indicated. It is further assumed that the initial mutation rate, $k_{mut} = 10^{-11}$ or 10^{-9} , as indicated. Values of $P_{rel, mut:no mut}$ at >1 indicate more than half the cancers arising with a mutator mutation somewhere in their pathogenesis. The fraction of cancers arising via a mutator mutation is given by $P_{rel, mut:no mut}/(1 + P_{rel, mut:no mut})$.

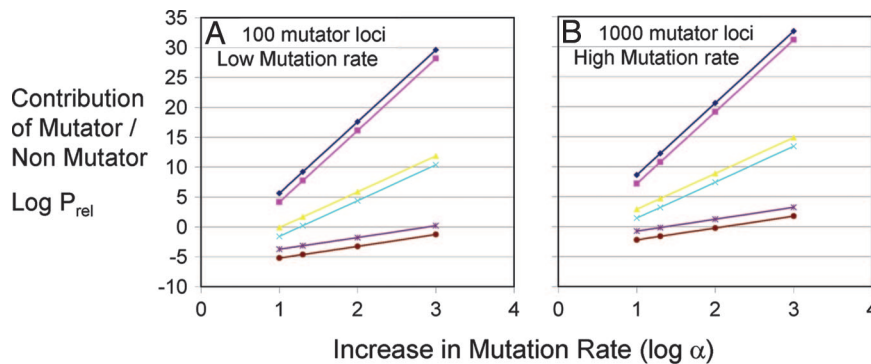


Fig. 1. $\log(P_{\text{rel, mut:no mut}})$, the common logarithm of the relative efficiency of carcinogenesis with a mutator mutation compared to that without a mutator mutation, plotted vs. $\log \alpha$, the common logarithm of the fold increase in mutation rate due to a mutator mutation. The graphs were calculated by using Eq. 8 and by assuming that $N_{\text{ML}} = 100$ (A) or 1,000 (B) and that the initial mutation rate $k_{\text{mut}} = 10^{-11}$ (A) or 10^{-9} (B). $T = 170$ (brown, light blue, and magenta) or 5,000 (purple, yellow, and dark blue). $C = 2$ (lower lines), 6 (middle lines), or 12 (upper lines).

slopes and locations of the curves corresponding to data in Table 2 not shown in Fig. 1 can easily be inferred. A positive value of $\log(P_{\text{rel, mut:no mut}})$ corresponds to >50% of the cancers arising by a pathway including a mutator mutation, whereas a negative value of $\log(P_{\text{rel, mut:no mut}})$ corresponds to <50% of the cancers arising by such a pathway. We observe that under most (but not all) conditions, mutator mutations are favored, in some cases by >30 orders of magnitude.

We see again that when only two steps are required for carcinogenesis ($C = 2$) mutator mutations are not likely to be important unless the fold increase in error rate caused by them is large (although still within ranges that have been seen). Indeed, the effect of changes in C is dramatic. Overall, we see that the relative importance of mutator mutations can vary widely depending on the parameter values.

Discussion

Cancer is a common endpoint that can likely occur by multiple mechanisms. We have used a direct deterministic mathematical model to assess the importance of mutator mutations in carcinogenesis as a function of a variety of conditions.

By using an approach in which we calculate the ratio of efficiency of carcinogenic mechanisms with a mutator mutation to those without a mutator mutation, we circumvent the debate about whether or not mutator mutations are necessary to produce cancers within a human lifetime. We calculate relative efficiencies rather than absolute rates. We implicitly assume that not every malignant cell develops into a clinical cancer and that the mechanism that produces the greatest number of malignant cells is most likely to be clinically significant. Thus, the focus shifts from evaluating whether steps are rate-limiting to evaluating relative efficiencies of different carcinogenic pathways.

We find that mutator mutations are important in most of the carcinogenic scenarios we analyzed. Mutator mutations are of greatest importance when they occur as an initial step in carcinogenesis.

Our results indicate that the importance of mutator mutations is greatest when more oncogenic mutations are required before commitment to cancer. When only two oncogenic mutation steps are required, mutator mutations are not favored as a pathway to cancer, because they are unlikely to accelerate such a small number of steps sufficiently to offset the time required for acquisition of the mutator mutation itself. As seen in Table 1, however, at as few as three steps, mutator mutations begin to be favored at reasonable fold increases (α) in mutation rates, and, for four or more steps, mutator mutations are clearly favored except for very small α . We note that mutation of both copies of a recessive oncogene would count as two separate events. In the

case of embryonal carcinomas, where mutation of both copies of a single recessive oncogene are the only two events required (32), mutator mutations are not likely to play a significant role. Thus, in hereditary, nonpolyposis colon cancer, for which there is a single-copy mutation of a mutator gene and a second mutation is required to lead to a mutator phenotype, there is no increased risk of embryonal cancers despite an increased risk of colon cancer. The case of retinoblastoma (20) is an exception in that the critical oncogene mutation also causes genetic instability; therefore, no extra step is required to acquire the mutator mutation.

The importance of mutator mutations also increases as α increases. The value of α must be greater than a certain minimum for 50% of the cancers to occur via a mechanism that includes a mutator mutation. In this model, the number of nucleotide loci that result in a mutator mutation when mutated is kept fixed. However, in reality, the fraction of mutator loci that result in α above a certain cutoff will be smaller the higher the cutoff is. Mutations of proofreading, base selection, DNA repair, and chromosomal instability can produce values of α within the ranges described in Table 1 (9, 30, 31, 33). Mutations in *Escherichia coli* DNA polymerase I exhibit up to a 50,000-fold increase in mutation rate (31). Catalytic site mutations of DNA polymerase δ , a major DNA-replicating polymerase, exhibit a 50-fold increase in mutation rate in yeast when substituted for the wild-type gene (33). Moreover mutations that abolish the proofreading activity of DNA polymerase δ resulted in cancers in mice harboring and expressing the mutant gene (33).

The importance of mutator mutations increases with a higher wild-type k_{mut} and with T in this model, despite acquisition of oncogenic mutations being less “rate limiting” with increasing k_{mut} and T . This surprising conclusion is a direct result of the focus on relative efficiencies in contrast to calculation of absolute rates and comparison to incidence. The dependence on T suggests a possible importance of mutator mutations in tumors of rapidly proliferating tissues, such as hematologic malignancies or colon cancer, and the effect of the number of oncogenic steps suggests that mutator mutations may play a very significant role in prostate cancer, the development of which may require up to 12 steps (21). Given the relatively late onset of prostate cancer, it is identified as potentially highly susceptible to prevention by agents that inhibit the generation of mutations.

The effect of premalignant lineage expansion on our results will likely depend on its extent and timing relative to the acquisition of oncogenic mutations. The effect of lineage expansion on absolute carcinogenesis rates may also be distinct from that on relative rates. Although lineage expansion affects the absolute rates of developing cancer, preliminary results

indicate that the effect on the relative efficiency of mutator and nonmutator mechanisms is less significant.

NCS is not explicitly considered in this model but was considered previously in isolation (15). This work demonstrated that NCS was not quantitatively significant over a broad range of parameters, but it should be noted that the current model may be valid only within that parameter range. In particular, although the results in this article suggest the increasing importance of mutator mutations with increasing k_{mut} , α , and T , in the presence of NCS these parameters cannot continue to increase indefinitely without the lineages eventually being limited by this effect.

Mutator mutations and lineage expansion of selected pre-malignant lineages are not mutually exclusive hypotheses, and both likely play significant roles in carcinogenesis, depending on the conditions. Mutator mutations will play a larger role when more oncogenic mutation steps are required for carcinogenesis and when the mutator mutations occur early in carcinogenesis.

Within the limits imposed by NCS, mutations may occur in multiple mutator genes, resulting in an ever-increasing mutation rate and accumulation during tumor progression. Tumors may be considered as a quasispecies, and the evolution of a tumor may depend on multiple mutant cells within the tumor (34) rather than on the emergence of single clones with a proliferative advantage. The genetic variability of tumors produced by mutator mutations increases the likelihood that a clinical tumor will contain one or more cells resistant to any particular therapy, consistent with the observed utility of therapeutic combinations (22).

Methods

The probability of a single cell or its progeny developing cancer in its lifetime without a mutator mutation occurring is

$$P_{\text{cancer}, 0} = p_{\text{mut}}^C [N_{\text{OL}}! / \{(N_{\text{OL}} - C)! C!\}] \quad [1]$$

or, equivalently, because $p_{\text{mut}} \approx k_{\text{mut}}T$,

$$P_{\text{cancer}, 0} = k_{\text{mut}}^C T^C [N_{\text{OL}}! / \{(N_{\text{OL}} - C)! C!\}]. \quad [2]$$

The factorial terms are the number of ways C mutations can be selected from a set of N_{OL} sites without regard to the order of selection, whereas p_{mut} is the probability of one of these sites being mutated, and p_{mut}^C is the probability of C sites independently being mutated.

The probability $P_{\text{cancer}, k}$ of a single cell or its progeny developing into a cancer in its lifetime with a mutator mutation occurring as the k th step in carcinogenesis at any generation t , $0 < t < T$, is given by the integral (from generation 0 to T) of the product of the probability $P_{k-1, t}$ of $k-1$ oncogene mutations occurring by generation t , the instantaneous probability of a mutator mutation occurring at generation t ($N_{\text{ML}} k_{\text{mut}} dt$), and the probability $P_{C, k-1, t}$ of the remaining $C - (k - 1)$ mutations occurring between generation t and T after a mutator mutation at generation t increases the mutation rate α -fold:

$$P_{\text{cancer}, k} = \int_0^T P_{k-1, t} P_{C, k-1, t} N_{\text{ML}} k_{\text{mut}} dt \quad [3]$$

where

$$P_{k-1, t} = (k_{\text{mut}}t)^{k-1} [N_{\text{OL}}! / \{(N_{\text{OL}} - k + 1)! (k - 1)!\}] \quad [4]$$

$$P_{C, k-1, t} = \{\alpha k_{\text{mut}}(T - t)\}^{C-k+1} [(N_{\text{OL}} - k + 1)! / \{(N_{\text{OL}} - C)! (C - k + 1)!\}]. \quad [5]$$

Eqs. 3–5 can be shown (see *Derivation of Eq. 6 in Appendix*) to give an intuitive result regarding the probability of a single cell

or its progeny developing into cancer in T generations with a mutator mutation occurring as the k th step:

$$P_{\text{cancer}, k} = \alpha^{C-k+1} (k_{\text{mut}}T)^{C+1} [N_{\text{ML}}(N_{\text{OL}})! / \{(N_{\text{OL}} - C)! (C + 1)!\}]. \quad [6]$$

We determine the relative probability $P_{\text{rel}, k, 0}$ of a carcinogenesis pathway with a mutator mutation at the k th step to one without a mutator mutation by dividing Eq. 6 by Eq. 2:

$$P_{\text{rel}, k, 0} = P_{\text{cancer}, k} / P_{\text{cancer}, 0} = \alpha^{C-k+1} k_{\text{mut}} T [N_{\text{ML}} / (C + 1)]. \quad [7]$$

The relative probability $P_{\text{rel}, \text{mut:non-mut}}$ of a carcinogenic pathway involving a mutator mutation at any step compared with one without a mutator mutation is the sum of Eq. 7 over all integral values of k from 1 to C . This sum can be computed as a geometric series and then approximated to within $1/(\alpha - 1)$ by the value of the term corresponding to the mutator mutation at the first step ($k = 1$) when $\alpha \gg 1$:

$$\begin{aligned} P_{\text{rel}, \text{mut: no mut}} &= \sum_{k=1, C} P_{\text{rel}, k, 0} \\ &= [(\alpha^{C+1} - \alpha) / (\alpha - 1)] (k_{\text{mut}}T) [N_{\text{ML}} / (C + 1)] \\ &\approx \alpha^C k_{\text{mut}} T [N_{\text{ML}} / (C + 1)]. \end{aligned} \quad [8]$$

It is apparent from Eqs. 7 and 8 that most of the cancers arising from pathways containing a mutator mutation have that mutation as the first step in their pathogenesis, because that pathway corresponds to the highest power of α in Eq. 7, and $\alpha \gg 1$ in most cases. It is clear from Eq. 8 that a mutator mutation must result in a sufficiently large fold increase α in mutation rate to make up for the extra time required to acquire the mutator mutation. For $\geq 50\%$ of cancers to have a mutator mutation in their pathway, $P_{\text{rel}, \text{mut: no mut}} \geq 1$, or, rearranging Eq. 8,

$$\alpha_{50\%} \approx \{[1 / (k_{\text{mut}}T)] \{(C + 1) / (N_{\text{ML}})\}\}^{1/C}, \quad [9]$$

where $\alpha_{50\%}$ is the minimum value of α that is sufficiently advantageous such that $\approx 50\%$ of the cancers will have a mutator mutation in their pathogenesis.

Appendix

Proof That $p_{\text{mut}} \approx k_{\text{mut}}T$ When $k_{\text{mut}}T \ll 1$. The rate of decrease in the fraction of unmutated loci is proportional to the mutation rate constant k_{mut} and the fraction of unmutated loci:

$$d[1 - p_{\text{mut}}(t)]/dt = -k_{\text{mut}}[1 - p_{\text{mut}}(t)]. \quad [10]$$

Solving Eq. 10 subject to the boundary condition that $p_{\text{mut}}[0] = 0$ (i.e., the lineage starts out as wild-type), we obtain

$$p_{\text{mut}}(t) = 1 - e^{-k_{\text{mut}}t}. \quad [11]$$

Substituting $t = T$, expanding the exponential in a Taylor series, and truncating the series after the linear term because $k_{\text{mut}}T \ll 1$, we obtain $p_{\text{mut}} \approx k_{\text{mut}}T$.

Derivation of Eq. 6. We start with Eqs. 3–5 for $k = 1$. Removing terms independent of t from under the integral, we obtain

$$P_{\text{cancer}, 1} = \alpha^C k_{\text{mut}}^{C+1} I_1(C, T) [N_{\text{ML}} N_{\text{OL}}! / \{(N_{\text{OL}} - C)! C!\}] \quad [12]$$

$$I_1(C, T) = \int_0^T (T - t)^C dt. \quad [13]$$

The integral is solved with the substitution $t' = T - t$.

A similar approach is applied to Eqs. 3–5 when $k > 1$, leading to

$$P_{\text{cancer}, k} = \alpha^{C-k+1} k_{\text{mut}}^{C+1} I_k(C, T) [N_{\text{ML}} N_{\text{OL}}! / \{(N_{\text{OL}} - C)!(C - k + 1)!(k - 1)!\}], \quad [14]$$

$$I_k(C, T) = \int_0^T t^{k-1} (T - t)^{C-k+1} dt. \quad [15]$$

Integration of Eq. 15 by parts leads to a recursion relation for I_{k+1} in terms of I_k :

$$I_{k+1}(C, T) = k I_k(C, T) / (C - k + 1). \quad [16]$$

Repeated application of this recursion relation from the known value of I_1 allows evaluation of I_k for all k , and substitution of those values into Eq. 14 leads to Eq. 6.

We thank Mike Schmitt for insightful comments. This work was supported by National Cancer Institute Grants CA102029, CA78885, and CA105704.

1. Nordling CO (1953) *Br J Cancer* 7:68–72.
2. Armitage P, Doll R (1954) *Br J Cancer* 8:1–12.
3. Cook PJ, Doll R, Fellingham SA (1969) *Int J Cancer* 4:93–112.
4. Hanahan D, Weinberg RA (2000) *Cell* 100:57–70.
5. Rangarajan A, Hong SJ, Gifford A, Weinberg RA (2004) *Cancer Cell* 6:171–183.
6. Loeb LA, Springgate CF, Battula N (1974) *Cancer Res* 34:2311–2321.
7. Ionov Y, Peinado MA, Malkhoysan S, Shibata S, Perucho M (1993) *Nature* 363:558–561.
8. Fishel R, Lescoe MK, Rao MRS, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R (1993) *Cell* 75:1027–1038.
9. Lengauer C, Kinzler KW, Vogelstein B (1998) *Nature* 396:643–649.
10. Paulovich AG, Toczyski DP, Hartwell LH (1997) *Cell* 88:315–321.
11. Nowell PC (1976) *Science* 194:23–28.
12. Tomlinson IPM, Novelli MR, Bodmer WF (1996) *Proc Natl Acad Sci USA* 93:14800–14803.
13. Fisher JC (1958) *Nature* 181:651–652.
14. Cairns J (1975) *Nature* 255:197–200.
15. Beckman RA, Loeb LA (2005) *Genetics* 171:2123–2131.
16. Fearon ER, Vogelstein B (1990) *Cell* 61:759–767.
17. Balaban GB, Herlyn M, Clark WH, Jr, Nowell PC (1986) *Cancer Genet Cytogenet* 19:112–122.
18. Knudson AG (1993) *Proc Natl Acad Sci USA* 90:10914–10921.
19. Nowak MA, Michor F, Komarova NL, Iwasa Y (2004) *Proc Natl Acad Sci USA* 101:10635–10638.
20. Knudson AG (1971) *Proc Natl Acad Sci USA* 101:820–823.
21. Cook PJ, Doll R, Fellingham SA (1969) *Int J Cancer* 4:93–112.
22. Beckman RA, Loeb LA (2005) *Semin Cancer Biol* 15:423–435.
23. Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ (2002) *Proc Natl Acad Sci USA* 99:3586–3590.
24. Albertini RJ, Nicklas JA, O'Neill JP, Robison SH (1990) *Annu Rev Genet* 24:305–326.
25. Baca OG, Tacheeni OS, Akporiaye ET, Deblasie R, Crissman HA (1985) *Infect Immun* 47:366–369.
26. Baker FL, Sanger LJ, Rodgers RW, Jabboury K, Mangini OR (1995) *Cell Prolif* 28:1–15.
27. Tomlinson I, Sasienu P, Bodmer W (2002) *Am J Pathol* 100:755–758.
28. Kolodner RD, Putnam CD, Myung K (2002) *Science* 297:552–557.
29. Guo HH, Choe J, Loeb LA (2004) *Proc Natl Acad Sci USA* 101:9205–9210.
30. Beckman RA, Loeb LA (1993) *Q Rev Biophys* 26:225–331.
31. Camps M, Naukkarinen J, Johnson BP, Loeb LA (2003) *Proc Natl Acad Sci USA* 100:9727–9732.
32. Hethcote HW, Knudson AG (1978) *Proc Natl Acad Sci USA* 75:2453–2457.
33. Venkatesan N, Hsu JJ, Lawrence NA, Preston BD, Loeb LA (2006) *J Biol Chem*, in press.
34. Eigen M (2002) *Proc Natl Acad Sci USA* 99:13374–13376.