Available online at www.sciencedirect.com
ScienceDirect
BioSystems 90 (2007) 870–880
BioSystems
www.elsevier.com/locate/biosystems

The stochastic model of non-equilibrium mutagen-induced alterations of DNA: Implication to genetic instability in cancer

Darapond Triampo [a], Wannapong Triampo [b,*], I.-Ming Tang [b], Yongwimon Lenbury [c]

[a] *Department of Chemistry and Center for Vectors and Vector-Borne Diseases, Faculty of Science, and Institute for Innovation and Development of Learning Process, Mahidol University, Rama 6 Rd., Rajchataywee, Bangkok 10400, Thailand*
[b] *R&D Group of Biological and Environmental Physics (BIOPHYSICS), Department of Physics, and Center of Excellence for Vectors and Vector-Borne Diseases, Faculty of Science, Mahidol University, Bangkok, Thailand*
[c] *Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, Thailand*

Received 6 January 2007; received in revised form 14 May 2007; accepted 18 May 2007

Abstract

Genetic alterations such as point mutations, chromosomal rearrangements, modification of DNA methylation and chromosome aberrations accumulate during the lifetime of an organism. They can be caused by intrinsic errors in the DNA replication and repair as well as by external factors such as exposure to mutagenic substances or radiation. The main purpose of the present work is to begin an exploration of the stochastic nature of non-equilibrium DNA alteration caused by events such as tautomeric shifts. This is done by modeling the genetic DNA code chain as a sequence of DNA-bit values ('1' for normal bases and '−1' for abnormal bases). We observe the number of DNA-bit changes resulting from the random point mutation process which, in the model, is being induced by a stochastic Brownian mutagen (BM) as it diffuses through the DNA-bit systems. Using both an analytical and Monte Carlo (MC) simulation techniques, we observe the local and global number of DNA-bit changes. It is found that in 1D, the local DNA-bit density behaves like $1/\sqrt{t}$, the global total number of the switched (abnormal) DNA-bit increases as $\sqrt{t}$. The probability distribution $P(b, 0, t)$ of $b(0, t)$ is log–normal. It is also found that when the number of mutagens is increased, the number of the total abnormal DNA-bits does not grow linearly with the number of mutagens. All analytic results are in good agreement with the simulation results.

PACS: 87.15.Aa

Keywords: DNA alteration; Genetic instability; Stochastic process; Mathematical modeling

1. Introduction

Genetic alterations such as point mutations, chromosomal rearrangements, unequal crossing over, loss of heterozygosity, modification of DNA methylation and chromosome aberrations accumulate during the lifetime of the organism. They are caused by intrinsic errors in the

* Corresponding author.
E-mail addresses: scwtr@mahidol.ac.th, wtriampo@yahoo.com (W. Triampo).

0303-2647/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2007.05.004

DNA replication and repair as well as by external factors such as exposure to mutagenic substances or radiation. Since the discovery that the configuration of a DNA or RNA molecule is a double helix (Watson et al., 1988), molecular biologists and geneticists have been studying the crucial role of DNA in the genome organization. Once it was recognized that DNA is the informational active chemical component of essentially all genetic materials, it was assumed that this macromolecule must be extraordinarily stable in order to maintain the degree of fidelity required of a master blueprint.

It was something of a surprise to learn that the primary structure of DNA is quite dynamic and subject to constant changes. For example, gene transposition is a well-established phenomenon in prokaryotic and eukaryotic cells (Finnegan, 1990; Kleckner, 1981). In addition, DNA is subject to alteration in the chemistry or sequence of individual nucleotides (Lindahl, 1993; Roberts, 1978; Singer and Kusmierek, 1982). Many of these changes arise as a consequence of errors introduced during replication, recombination and repairing itself. Other basic alterations arise from the inherent instability of specific chemical bonds that constitute the normal chemistry of nucleotides under physiological conditions of temperature and pH. Finally, the DNA of living cells reacts to a variety of chemical compounds and a smaller number of physical agents, many of which are present in the environment. Each of these modifications of the molecular structure of genetic material is appropriately considered to be a *DNA damage*. DNA damages can be classified into two major classes, spontaneous and environmental. However, in some cases the actual chemical changes in DNA that occur "spontaneously" are indistinguishable from those brought about through interaction with certain environmental agents. The term "spontaneous" may merely imply that we have not identified a particular environmental culprit. Changes in the DNA sequence may result from processes such as insertion, deletion, transversion and transition. For example, the genetic instability characteristic of cancer cells may be due, in part, to mutations in genes whose products normally function to ensure DNA integrity. DNA replication in normal human cells is an extremely accurate process. During the cell division cycle, copy errors occur with probabilities less than 10^{-9} to 10^{-10} per nucleotide. In contrast, the malignant cells that constitute cancer tissues are markedly heterogeneous and exhibit alterations in nucleotide sequence of DNA.

As initially proposed by Delbruck et al. (1935) and Watson and Crick (1953), spontaneous mutations are initiated by quantum jump events such as tautomeric shifts in single protons of DNA bases. Even what may be the most common of spontaneous mutations involves a chemical mechanism which must involve quantum uncertainty, since it occurs when individual electrons shift their positions to produce "tautomers".

Specifically, nucleotide transitions can be induced by exposure to endogeneous and exogeneous mutagens (agents causing genetic changes) such as chemical carcinogens. However, not all mutagens are carcinogenic. The nucleotide transitions are the interchange of bases of the same shape, e.g., the purine bases transition, C(cytosine) ↔ T(thymine) or the pyrimidine bases transitions, A(adenine) ↔ G(guanine). One of the mechanisms that can cause the transition is the shift of the positions of the electrons for the bases to become a transient form (known in organic chemistry as a *tautomeric shift*).

In standard complementary pairing, G pairs with C and A with T. Keto-enol tautomeric shift leads to non-standard form of G: G ↔ G^* resulting in G^* pairing with T. Amino-imino tautomeric shift leads to a non-standard form of A: A ↔ A^* resulting in A^* pairing with C. Non-standard bases alter the pairing specificity, i.e., modified purine pairs with the *wrong* pyrimidine and modified pyrimidine pairs with the *wrong* purine. Fig. 1 shows an example of the keto-enol tautomeric shift that results in a transition mutation of the complementary strand. Consider the pairing of ATGC with TACG: Let G in the first strand undergo a tautomeric shift to G^*. The complementary strand that is generated would be TATG, not TACG. This would be a transition from C ↔ T. To complete the process of producing a mutation, a tautomeric shift must take place during replication, either in the template chain, or in the deoxyribonucleotide being added by the DNA polymerase. Since the shifted form retains its rare mis-matching structure for only a brief period, the next replication cycle will most likely find itself reverted

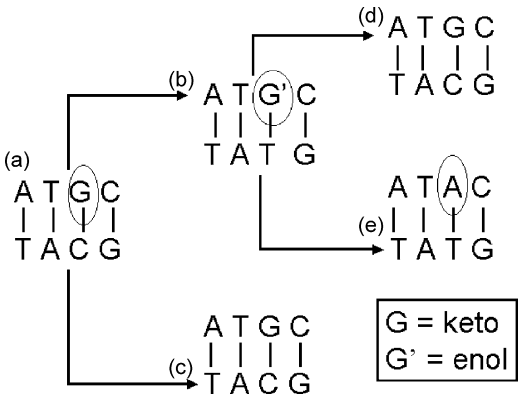

Fig. 1. A diagram of the keto-enol tautomeric shift that results in a transition mutation of the complementary strand.

to its normal form, and the polymerase will pair it with its normal mate. Thus, in two cycles of replication, an A–T pair is changed to a G–C pair, or vice versa. This, in turn, can often result in a change in a triplet code, leading to an amino acid substitution in a protein, and a modification of some visible phenotypic property of the organism.

Although it has never been demonstrated experimentally that rare tautomers are responsible for spontaneous mutations, subsequent experimental and theoretical investigations (Leszczynski, 1999; Radchenko et al., 1983) seem to confirm the essential correctness of this postulate. It should be remarked that Neo–Darwinian evolutionary theory is founded on the principle that mutations occur randomly, and the direction of evolutionary change is provided by selection for advantageous mutations. However, the central tenet, that mutations occur randomly, has recently been challenged by the finding of the phenomenon termed adaptive or directed mutation.

There have been a few approaches used to investigate this mutation complex process ranging from wet lab research to highly complicated computational calculations. Theoretical models fall into two very broad classes: deterministic and stochastic models. Deterministic models attempt to model or predict the average behavior of systems according to some precise rules. In contrast, stochastic models describe the probability of very specific behaviors of individuals rather than average behavior of the population. Stochasticity has been recognized in the biology field of research and modeling as the description of life systems (Kurakin, 2006). It had appeared as general principles underlying the dynamics and organization of biological systems at all scales: gene expression (Kurakin, 2005), enzymes (Xie and Lu, 1999), self-organization of macromolecular complexes mediating transcription (Dundr et al., 2002; Kimura et al., 2002), and DNA repair (Essers et al., 2002; Hoogstraten et al., 2002).

Because a gene or DNA is a molecule, the statistical fluctuations of atomic or molecular scale cannot be avoided. Mathematical modeling of genetic instability has led to considerable insight into human tumorigenesis. One study of the mutational spectrum gave the type, location and frequency of DNA changes in a particular gene (Hussain and Harris, 1999). Claytong and Robertson (1955) proposed a random walk mutation model as a model for genetic analysis. It was later proposed explicitly by Crowj and Kimura (1964), by Kimura (1965), and subsequently popularized by Lander (1975). Zeng and Cockerham (1993) proposed a more general mutation model, called the regression mutation model. This model regards the regression coefficient of the effect of an allele after mutation on the effect of the allele before mutation as a parameter.

In 1989, Nowak and Schuster (1989) investigated error thresholds in finite populations. They determined that, at error rates above the critical value, the quasispecies ceases to be localized in sequence space and start to drift randomly. Sole' and Deisboeck (2004) used a quasispecies model to investigate the error threshold in cancer cells. They demonstrated that, once the threshold is reached, the highly unstable cancer cells become unable to maintain their genetic information, leading to a decrease in the velocity of tumor growth. The original quasispecies model assumes that genomes replicate conservatively, i.e., each single-stranded genome replicates by producing a new, possibly error-prone, single stranded copy without affecting the original. In this form, the quasispecies model predicts the existence of an error catastrophe or "error threshold", a threshold mutation rate above which no viable species can exist. This threshold depends on the replication rate of the fittest sequence, the master sequence (Komarova et al., 2002) utilize a stochastic model to evaluate the rate of formation of dysplastic crypts by chromosomal instability (CIN) and microsatellite instability (MIN) mechanisms in sporadic colon cancer to obtain broad qualitative agreement with the relative importance of CIN and MIN and the number of polyps generated under these conditions.

The main purpose of the present work is to begin an exploration of the stochastic nature of non-equilibrium DNA alteration caused by events such as tautomeric shifts in a theoretical DNA-bit alteration model This is done by modeling the genetic DNA (or RNA) code chain as a sequence of DNA-bit values ('1' for normal bases and '−1' for abnormal bases). This is similar to what is used in computers or electronics. We observe the number of DNA-bit changes resulting from the random point mutation process (to mimic tautomeric shifts) which is being induced by a stochastic Brownian mutagen (BM) as it diffuses through the DNA-bit systems. We will make analytic predictions and simulate the non-equilibrium process using the Monte Carlo (MC) method. To the best of our knowledge, there has not been a stochastic approach to investigate the non-equilibrium stochastic kinetics of DNA-alteration. This work therefore represents a new avenue for studying non-equilibrium mutation.

2. Theoretical model and analytic predictions

As mentioned, earlier theoretical models fall into two very broad classes: deterministic and stochastic models.

Stochastic models evaluate the entire probability distribution of random individual events. This kind of model is potentially more informative in that it considers rare events, not just average properties. Typically one defines a variety of discrete states, and the rates or probabilities of transition between the states. Often the different states of a phenomenon of interest can be represented as a Markov process. In a Markov process, the system passes through the defined states in discrete steps with a given set of transition probabilities. The possibilities for where the system will go next, and the chance it will "select" a particular option, depend only on where the system is at the moment (i.e., its present state) rather than on how it got there (its history). This type of analysis can in principle give the chance that the system is in a given state as a function of time or other key variables. However, utilizing this approach often requires a detailed understanding of individual states and transitions, which is not always available. As the system complexity increases, the definition of all the relevant states and the mathematical analysis of all the transitions between them can become daunting. The results will

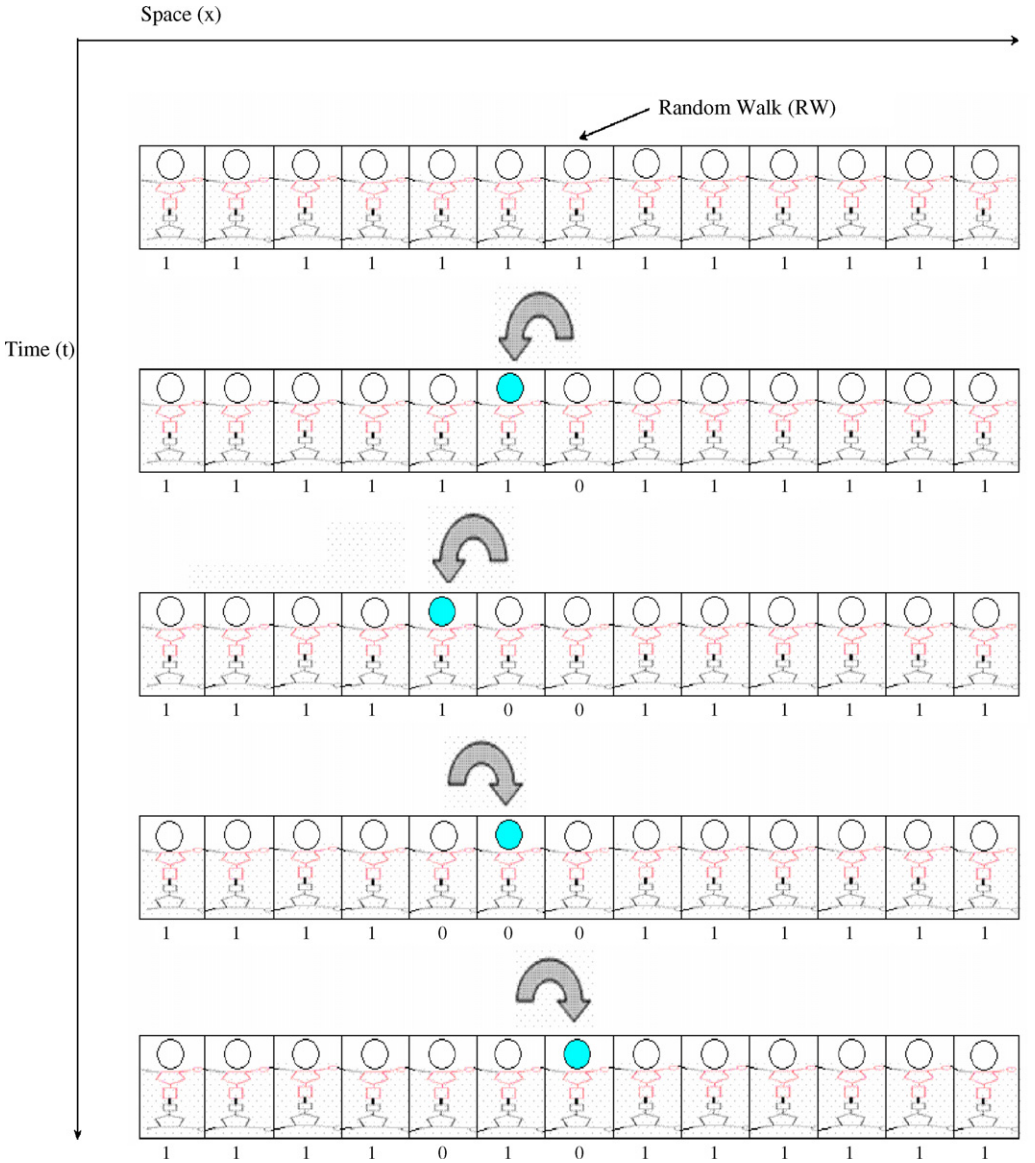

Fig. 2. Illustration of the discrete system incorporated with DNA-bit switching processes for 1D and switching rate $q = 1$. The initial 100% normal DNA-bit state is shown on the top, with the Brownian mutagen (BM) represented by the filled circle. From the top, we show a possible configurations after 3 step-moves. The BM switches a DNA-bit each visit, so those DNA-bits visited an even number of times are restored to their original value.

be probabilities of discrete events, rather than average properties.

We now look at a system of one-dimensional (1D) chain of the DNA-bits (Fig. 2). The position of each DNA-bit is labeled by "x" and that of a mutagen by a lattice vector $R(t)$. The DNA-bits are described by the variables σ_x which may take the values "1" or "−1". The bit variables encode the information about the status of the nucleotide sequences for the transition creation process. As we stated above, normal DNA-bits are denoted by "1" and abnormal DNA-bits by "−1". The mutagen has a probability p for moving to one of its two nearest neighbor sites in a time step δt. After making such a jump, there is a probability q that the DNA-bit on the site departed from is switched. As known, parameters are the variables which, based on the theoretical analysis, are expected to influence the outcome of interest. In some cases, values of the parameters are known from prior experiments, and therefore these values can be fixed. In other cases, the values of the parameters are unknown or could reasonably be expected to vary over a known range. In this case, the parameters are adjustable. The greater the excess of independent experimental data points over adjustable parameters, the more valid the experimental confirmation of the theory. Starting with the initial condition that all bits are "1", we monitor the time evolution of the numbers of the abnormal "−1" and normal bits "1" for the different situations which shall be specified later.

To mathematically model the stochastic mutagenesis, we write the temporal probability distribution $P(R, \{\sigma_x\}, t)$ which is the probability that at time t, the mutagen is at position $R(t)$ and the DNA-bits have values given by the set, $\{\sigma_x\}$. This distribution evolves according to a master equation (Gardiner, 1985) of the form

$$\begin{aligned} &P(R, \{\sigma_x\}, t+\delta t) \\ &=(1-p)P(R, \{\sigma_x\}, t)+\frac{p(1-q)}{2d}\sum_l P(R+l, \{\sigma_x\}, t) \\ &+\frac{pq}{2d}\sum_l P(R+l, \ldots, -\sigma_{R+l}, ..., t) \end{aligned} \tag{1}$$

where l represents the two orthogonal lattice vectors or go-left and go-right vector (which have magnitude l). In principle one can solve this system by the use of linear difference equations. This approach would suffer from having a too large of a number of degrees of freedom. For the case of Brownian mutagen, we focus on the specific case where $p = 1$ and $q = 1$.

An alternative continuum description was obtained by viewing the process as a stochastic cellular automaton (SCA). The process is then defined in terms of the position $R(t)$ of the BM. Each time step the agent makes a random jump to one of its nearest neighbors and in which the bit at the site it leaves behind definitely switches. This corresponds to setting $p = q = 1$. Let us denote a randomly chosen unit lattice vector by $l(t)$, and the time-dependent value of the spin at site x by $\sigma_x(t)$. Then we have

$$R(t+\delta t) = R(t) + l(t), \tag{2}$$

and

$$\sigma_x(t+\delta t) = \sigma_x(t)(1-2\delta_{x,R(t)}) \tag{3}$$

We are interested in a continuum limit of these two equations. In this limit, the first equation becomes the Langevin stochastic equation for the random walk,

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \xi(t), \tag{4}$$

where $\xi(t)$ is an uncorrelated Gaussian random variable with zero mean (i.e., $\xi(t)$ is a white noise process). The correlator of $\xi(t)$ is given by

$$\langle \xi(t)\xi(t')\rangle = D\delta(t-t'), \tag{5}$$

where δ is the Dirac delta function and D is the diffusion constant. $\langle \cdots \rangle$ indicates an average over the noise (or equivalently the paths of the agent). The agent is chosen to reside initially at the origin.

The equation governing the evolution of the DNA-bit density denoted by ϕ is described by

$$\partial_t \phi(x,t) = -\lambda\phi(x,t)\delta(x-R(t)), \tag{6}$$

where λ is a phenomenological parameter which describes how strongly the DNA-bit density is coupled to the BM. It is the coarse-grained version of σ. Taking an initial condition $\phi(x, 0) = 1$ for all x and straightforward integration of Eq. (6) gives the explicit functional solution

$$\phi(x,t) = \exp[-\lambda \int_0^t \mathrm{d}t'\delta(x-R(t')). \tag{7}$$

We will now use the stochastic solution of local DNA-bit density to calculate several interesting quantities. The simplest quantity to consider is the mean local DNA-bit density given by

$$b(x,t) = \langle \phi(x,t)\rangle = \sum_{n=0}^{\infty}(-\lambda)^n \chi_n(x,t), \tag{8}$$

where $\chi_0(x, t) = 1$ and for $n > 0$,

$$\chi_n(x,t) = \frac{1}{n!}\left\langle \left[\int_0^t \mathrm{d}\tau\, \delta(x-R(\tau))\right]^n \right\rangle. \tag{9}$$

It can be shown that

$$\chi_n(x,t) = \int_0^t \mathrm{d}\tau_1 \int_0^{\tau_1} \mathrm{d}\tau_2 \cdots \int_0^{\tau_{n-1}} \mathrm{d}\tau_n \, g(0,\tau_1-\tau_2) \times \cdots \times g(0,\tau_{n-1}-\tau_n) g(x,\tau_n), \quad (10)$$

where $g(x,t) = (2\pi Dt)^{-1/2}\exp(-x^2/2Dt)$ is the probability density of the random walk. Eq. (10) is an n-fold convolution. Therefore, if we apply the temporal Laplace transform, we get (for $n>0$)

$$\tilde{\chi}_n(x,s) \equiv \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-st}\chi_n(x,t) = \frac{1}{s}\tilde{g}(0,s)^{n-1}\tilde{g}(x,s), \quad (11)$$

where

$$\tilde{g}(x,s) = \frac{1}{(2DS)^{1/2}}\exp\left[-\left(\frac{2S}{D}\right)^{1/2}|x|\right]. \quad (12)$$

where $\tilde{g}(x,s)$ is the Laplace transform of the diffusion equation Green function.

Summing over these function as given in Eq. (6) we find

$$\tilde{b}(x,s) = \frac{1}{s}\left[1 - \frac{\lambda\tilde{g}(x,s)}{1+\lambda\tilde{g}(0,s)}\right]. \quad (13)$$

This exact result allows one to extract a great deal of statistical information about the process. First, one can simply invert the Laplace transform to find the average local DNA-bit density (or average density of switching relative to 1/2) as a function of x and t. The explicit forms are given as

$$b(x,t) = \mathrm{erf}\left[\frac{|x|}{(2Dt)^{1/2}}\right] + \exp\left(\frac{\lambda|x|}{D} + \frac{\lambda^2 t}{2D}\right) \times \mathrm{erfc}\left[\lambda\left(\frac{t}{2D}\right)^{1/2} + \frac{|x|}{(2Dt)^{1/2}}\right] \quad (14)$$

where $\mathrm{erf}(z)$ and $\mathrm{erfc}(z)$ are the error function (Abramowitz and Stegun, 1972). Considering the long time behavior of the above expression, we find that the average local DNA-bit density at the origin ($x=0$) decays asymptotically as

$$b(0,t) = \left(\frac{2D}{\pi\lambda^2 t}\right)^{1/2}\left[1 + O\left[\frac{D}{\lambda^2 t}\right]\right]. \quad (15)$$

We note here that the continuum solution has the important property that $\langle\phi(x,t;\lambda)^n\rangle = \langle\phi(x,t;nt)\rangle$. This allows us to utilize the exact solution to reconstruct the probability density for the local DNA-bit density. Theoretically, it can be proved that the average DNA-bit density by N agents $\equiv b_N(0,t)$ is proportional to $b(0,t)^N$, i.e.,

$$b^{(N)}(0,t) = [b^{(1)}(0,t)]^N = t^{N/2}, \quad \text{as } t \gg 1. \quad (16)$$

Another interesting quantity which may be extracted from $b(0,t)$ is the global switched DNA-bits, $B(t)$, defined as

$$B(t) = \int \mathrm{d}x[\langle b(x,0)\rangle - \langle b(x,t)\rangle].$$

This quantity obeys the exact relation

$$\frac{\mathrm{d}B(t)}{\mathrm{d}t} = \lambda b(0,t). \quad (18)$$

We find that asymptotically $B(t) \approx (t)^{1/2}$, independent of the coupling. In other words, the total amount of disorder created by a single BM on average increases as $t^{1/2}$ that is rather independent of the coupling between the BM and the DNA-bits for large time.

We now consider the probability distribution function $P(b,x,t)$ of the local corrupted bit density. This P function will provide the information about the time evolution of the probability distribution that describes the local corruption behavior. Obviously, at the very early times, the peak of the distribution is supposed to occur in the vicinity of the origin. The complete analytic structure of $b(x,t;\lambda)$ is needed to reconstruct the distribution function P. This suggests that by knowing the first moment of the corruption density, we can generate the higher moment. Therefore, we can reconstruct the probability density function. We define P via

$$P(b,x,t) = \langle\delta(b - b_R(x,t)\rangle \quad (19)$$

where $b_R(x,t)$ is the stochastic field solution given in Eq. (15). We can express the δ function using a frequency integral, and then expand it in powers of the field as follows:

$$\begin{aligned} P(b,x,t) &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\,\mathrm{e}^{-\mathrm{i}\omega b}\langle\mathrm{e}^{\mathrm{i}\omega b_R(x,t)}\rangle \\ &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\,\mathrm{e}^{-\mathrm{i}\omega b}\sum_{n=0}^{\infty}\frac{(\mathrm{i}\omega)^n}{n!}\langle b_R(x,t)^n\rangle \\ &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\,\mathrm{e}^{-\mathrm{i}\omega b}\sum_{n=0}^{\infty}\frac{(\mathrm{i}\omega)^n}{n!}\langle b(x,t;n\lambda)\rangle. \end{aligned} \quad (20)$$

We next take the Laplace transform of $b(x,t;n\lambda)$. From Eq. (12) we have

$$b(x,s;n\lambda) = \frac{1}{s}\left[1 - \frac{n\lambda\tilde{g}(x,s)}{1+n\lambda\tilde{g}(0,s)}\right] = \frac{\tilde{g}(0,s)-\tilde{g}(x,s)}{s\tilde{g}(0,s)} + \frac{\tilde{g}(x,s)}{s\tilde{g}(0,s)[1+n\lambda\tilde{g}(0,s)]}. \quad (21)$$

The first term is handled as it is independent of n. Thus, the sum over n for this yields a factor $\mathrm{e}^{\mathrm{i}\omega}$ which leads to

the factor $\delta(1-b)$ when integrated over ω. The details of how to perform the sum over n for the second term, we refer to a reference by Newman and Triampo (1999). The final result for $\hat{P}(b,x,s)$ is

$$\hat{P}(b,x,s) = \frac{\hat{g}(0,s)-\hat{g}(x,s)}{s\hat{g}(0,s)}\delta(1-b) + \frac{\hat{g}(x,s)}{\hat{g}(0,s)^2}\frac{1}{s\lambda b} \times \exp\left[-\frac{1}{\lambda\hat{g}(0,s)}\ln\left(\frac{1}{b}\right)\right]. \quad (22)$$

To this end, we need the explicit form for $\hat{g}(x,s)$ which is given by Eq. (11). Inserting this into Eq. (22) and inverting the Laplace transform, we have our final result

$$P(b,x,t) = \delta(1-b)\mathrm{erf}\left[\frac{|x|}{(2Dt)^{1/2}}\right] + \frac{1}{(\pi t)^{1/2}}\frac{1}{\tilde{\lambda}b} \times \exp\left\{-\left[\frac{|x|}{(2Dt)^{1/2}} - \frac{\ln b}{2\tilde{\lambda}t^{1/2}}\right]^2\right\} \quad (23)$$

where $\mathrm{erf}(z)$ is the error function where $\tilde{\lambda} = \lambda/(2D)^{1/2}$. In particular, the probability distribution for the average bit corruption density at the origin takes the form

$$P(b,0,t) = \frac{1}{(\pi t)^{1/2}}\frac{1}{\tilde{\lambda}b} \times \exp\left\{\left[-\frac{\ln(b)^2}{4\tilde{\lambda}^2 t}\right]\right\} \quad (24)$$

which is a log–normal distribution and where we have defined $\tilde{\lambda} = \lambda/\sqrt{2}$. This indicates the extreme nature of the fluctuations in the system. For instance, the typical value of the magnetization density can be found from the above expression to decay exponentially.

For the asymptotic behavior of $b(0,t)$ as $b(0,t) \approx 1/\sqrt{t}$, $P(b,0,t)$ in Eq. (24) becomes

$$\begin{aligned} P(b,0,t) &= \frac{1}{\sqrt{\pi}\tilde{\lambda}} \times \exp\left\{-\frac{1}{4\tilde{\lambda}^2 t}\left[\ln\frac{1}{\sqrt{t}}\right]^2\right\} \\ &\approx \exp\left\{-\frac{(\ln t)^2}{t}\right\} = \exp\left\{-\frac{1}{144t} - \frac{1}{6t^2} \right. \\ &\left. -\frac{3}{4t^3} + \frac{26}{9t^4} + O\left(\frac{1}{t^5}\right)\right\} \approx \exp\left\{-\frac{1}{t}\right\} \\ &= \exp\left\{-\left(\frac{1}{\sqrt{t}}\right)^2\right\} = \exp\{-b^2\} \end{aligned} \quad (25)$$

which is a normal distribution and where $O(1/t^5)$ is the correction to the order of $1/t^5$. We now claim that log–normal distribution approaches normality when t is infinitely large. Finally, we analyze the effects of many BM's within the system. We assume the BM's to be non-interacting, i.e., they are unaware of each other's immediate presence. The non-trivial statistics reside in the fact that the mutating effects of the BM's statistically interact via the overlapping of the BM histories. As we have already seen, a single BM interferes with the previous switched DNA-bit it has created, such that the amount of mutating does not simply increase linearly in time. This effect is more severe when more than one BM is present, as each BM can disturb the mutation that another BM has previously created. We measure the strength of this interference via a quantity called the "mutation efficacy" of the mutagents, defined as

$$\sigma_N \equiv \lim_{t\to\infty}\frac{B^N(t)}{B(t)}$$

where $B^N(t)$ is the average global mutation created by N mutagents. If the BM's were truly independent (in terms of the mutation they create), then we would expect $\sigma_N \propto N$.

3. Monte Carlo numerical results and discussion

Our aim in this section is to show the validity of our predicted results obtained in the previous section. To do so, we have performed the Monte Carlo simulations of the discrete model defined in Section 2. All results are obtained for a 1D chain of DNA-bits which at each site can take either the value 1 (normal) or -1 (abnormal). The chain length is considered negligible, as long as one ensures that the BM never touches the system boundaries in any of its realizations up to the latest time at which data is extracted. Thus, the system is infinitely large. We performed an average over realizations (or runs) with each run starting with the same initial configuration; namely all DNA-bits are normal. The DNA-bit at the starting point has value = 1 as shown in Fig. 2, then the BM are introduced to the starting point (origin). At each Monte Carlo step, the BM randomly walks to either one of its two neighboring sites and switches the DNA-bit of visited site before leaving. We let the BM mutate the system independently with the consequence that multiple occupancies are allowed.

We have focused on the local DNA-bit density at the origin so we measured the average altered DNA-bits density at the origin where the BM started switching the system denoted by $b(0,t)$. To investigate the accumulated DNA alteration, we measure the total number of abnormal DNA-bits $B(t)$ versus time. Then we defined a coarse-grained bit corruption over a patch containing 20 bits representing the bit at the origin to measure the probability distribution of the local DNA-bit density, $P(b,0,t)$. The 20 bit patch size is chosen because, computationally, this is primarily as a consequence of

the optimization of the simulation technique. This is technically to compromise the length and time scale of BM versus the DNA global alteration time scale. If the patch size is too small (it has been tested), it would not quite allow us to obtain the data for the reasonably good enough histogram data resulting in the good quality probability distribution. In contrast, if the patch size is too large, it could lead to the scenario where the BM would be spending to long time in just one site in unreasonable large frequency to alter each DNA coarse-grained bit. This situation could lead to the local change as unrealistic "over-express" of a spontaneous or one time-step of a single DNA in relation with global change that is not in the reasonable time scales. Biologically, this optimal bit patch size might imply some biological counterparts, i.e., how could each BM be able to induce DNA-alteration. In addition, this optimal size might reflect the BM capability or efficiency to alter DNA. Moreover, it is known that several factors like the sensitivity of each genetic site to BM, specificity of BM/DNA matching, the fluctuation of the response due to either "on or off" genes or the inhomogeneity of DNA array landscape, etc.

In addition, the BM is initially allocated evenly on boundary of the patch to avoid the internal decimation by the transient motion of the BM. Lastly, we consider the situation of more than one BMs. We have measured the asymptotic long time value of the ratio between the number of the abnormal DNA-bit when many agents are present and that when only one agent is present. It is denoted by σ_N.

In Fig. 3, the plot between the local DNA-bit density and time is shown. $b(0, t)$ can be viewed as being the frequency of the local changes of DNA-bit caused by the random or stochastic induction of the mutagen(s). It is found that $b(0, t) \approx \sqrt{t}$ for one BM and $b^{(N)}(0, t) = [b^{(1)}(0, t)]^N = t^{N/2}$ for $N > 1$ which are in good agreement with the analytic prediction. This indicates that $b(0, t)$ depends sensitively on the number of BMs in terms of the overlap of the paths of different walkers and how often the BMs have visited the origin. The decay of $b(0, t)$ due to N BAs is not linearly proportional to $b(0, t)$ due to 1 BM but instead it varies as the power N of $b(0, t)$. For $N = 2$, it gives $b(0, t) \approx t$. This result is consistent with observations that the between-population genetic variance (Roychoudhury and Nei, 1988; Lynch and Hill, 1986), and that the cumulative selection response from mutation (Hill, 1982) asymptotically increase linearly with time. It should be pointed out that the origin is strongly altered by the multiple BMs since all of the independent BMs always return to the origin and switched its DNA-bits. In the process of N agents which are non-interacting, they will interfere strongly with each other. In other words, the overlap of their histories is found. When time is infinitely large, $b(0, t)$ approaches 0. It implies 50% chance of finding the site to be normal or abnormal. This agrees with the time limit of $b(0, t)$. The fluctuation at this equilibrium is relatively large compared to that in the scaling regime. This results from thermal fluctuation (Burgess, 1969).

Fig. 4 shows the probability distribution $P(b, 0, t)$ of $b(0, t)$. In the early time regime, the simulated probability distribution is log–normal. The distribution curve has a robust tail for larger value of average DNA-bit density. This reveals the extreme fluctuation at the origin and the high probability that the origin will only be slightly mutated. The fact that BM on 1D lattice always returns to the origin (Hughes, 1995) is critical for this event. At large time the distribution has completely changed from

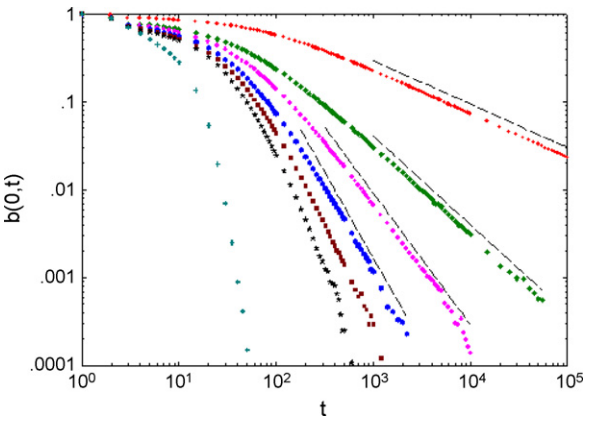

Fig. 3. The log–log plot of local DNA-bit density at the origin $b(0, t)$ vs. time, $d = 1$, $p = 1$ and $q = 1$ due to 1, 2, 3, 4, 5, 6, 16 and 32 BMs. The arrow directs the increment of number of BMs. The dash lines have slope −0.4992, −1.0104, −1.4999 and −1.9487, respectively, along the arrow direction and show the range of time in which the exponent is extracted.

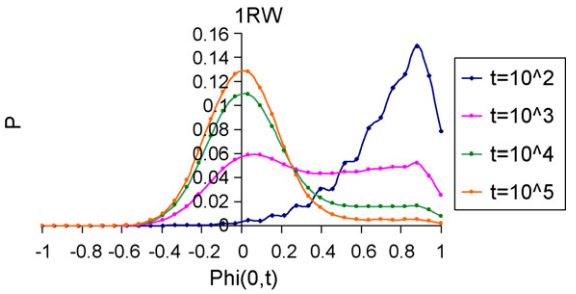

Fig. 4. The simulated probability distribution of a local DNA-bit density at the origin $b(0, t)$, $d = 1$, $p = 1$ and $q = 1$ due to 1 BM. In the early time regime when the simulated probability distribution is log–normal (Eq. (24)) and at large time the distribution has completely changed from log–normal to normal.

log–normal to normal. The peak occurs with a probability of 0.1275. An important feature that this distribution unambiguously points out is that the realizations where the DNA-bits at the origin will be half abnormal and half normal will have the highest probability of occurrence. The distribution approaches normality as time proceeds with the highest probability occurring at $b(0, t) = 0$. This means that the origin is steadily altered. It is expected that the distribution approaches a normal distribution quicker when there are more than one BM. The peak of the normal distributions remains at the same place. In contrast to $b(0, t)$, the characteristics of the normal distributions are independent of N and time.

In Fig. 5, we present the results of the numerical simulations which gives the values of σ_N. We have performed numerical simulations of the many mutagens system in order to test the prediction result. The microscopic rule we use is that there is no hard-core exclusion between the mutagens, and that for each time step the N mutagens are in turn moved to a randomly chosen nearest neighbor site. A DNA-bit which is occupied by two mutagens, say, will thus (for $q = 1$) be switched twice in that time step. We observe the evolution of the ratio of the average global DNA-bit switching for N agents as compared to one agent for $d = 1$. Asymptotically, this ratio is the mutating efficacy by definition. Results are shown for $N = 2$, 3, and 4. The curves are asymptotic to constants as expected. As we see, σ_N does not increase linearly as the number of mutagens is increased. From the stochastic point of view, this implies that there is a degree of interference between the mutagens. In the process of N BMs which are non-interacting, they will interfere strongly with each other. In other words, the overlap of their histories is found.

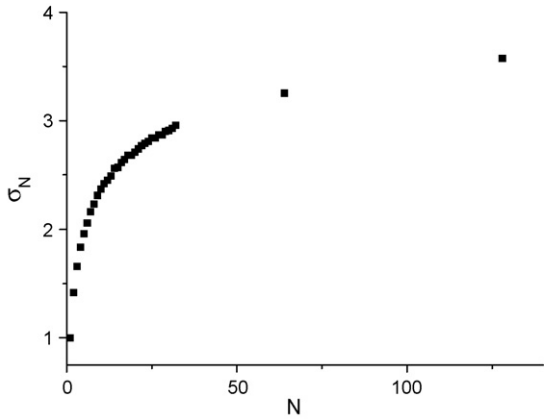

Fig. 5. Plot of σ_N vs. number of BMs. It shows that σ_N does not increase linearly as the number of mutagens is increased and, from the stochastic point of view, this implies that there is a degree of interference between the mutagens.

Lastly, we suggest that our theoretical results can be tested, at least in principle, directly by experiments. Recently, it was shown that CIN and MIN can be introduced in cancerous cell lines through specific mutagenesis (Bardelli et al., 2001). In addition, Greenman et al. (2007) and Haber and Settleman (2007) have shown large scale analysis of DNA mutations across cancer arrays. While an in-depth study of dynamics above the error threshold along with careful consideration of the enzymatic interactions, both subjects of future research, would be necessary to rigorously quantify this statement particularly to non-equilibrium aspects. This is one example of a quantifiable and testable hypothesis that can be used to experimentally test our theoretical work. Theoretical modeling frequently uses simplifying assumptions. Simplifying assumptions eliminate complexities which may be peripheral to the issue under consideration, allowing a focus on key features of a complex biological system. For example, most models of genetic instability assume that the rate of genetic change is constant at any location in the genome, even though there is evidence of mutation "hot spots" which violate this assumption (Schaaper and Dunn, 1991). In determining whether this simplifying assumption impacts the results when modeling genetic change in carcinogenesis, one would need to know whether mutation "hot spots" exist at key loci within cancer-associated genes.

4. Implication to genetic instability in cancer and conclusion

In this work, we have modeled the stochastic kinetics of the spontaneous mutation induced by nucleotide transition as a problem of a mutagen. The dynamics in the model is to mimic the mutagenesis due to the tautomeric shift which may occur when a mutagen interacts with one of the bases in the DNA chain. The "tautomers" are created when the interaction causes some of the electrons in the base to shift their positions. To understand how this model may feature the real world phenomena, we have used both analytical model and computer simulation techniques. Analytically, we have set up the master equation and solved for local DNA-bit density, global abnormal DNA-bits, and the probability distribution function to describe the non-equilibrium nature of mutagenesis. To confirm the theoretical findings, we have performed computer simulations by applying some stochastic cellular automata rules to a DNA-bit system. Evidently, the model is non-trivial since the values of DNA-bits depend very sensitively on the path of the

BM, i.e., how often the BM has visited and switched the DNA-bits. We find that the local DNA-bit density, $b(0, t) \approx 1/\sqrt{t}$ the global DNA-bit $B(t) \approx \sqrt{t}$, probability density function $P(b, 0, t)$ is log–normal, and also for the case of many mutagens, σ_N does not increase linearly as the number of the BMs increases. Instead, it increases more slowly due to interference effects occurring along the path of the mutagens. Our work may relate to genetic instability in cancer.

Genetic instability is a hallmark of human cancers (Lengauer et al., 1998; Loeb, 2001). Genetic changes which are required in carcinogenesis are divided into two very broad classes: those which are dominant, requiring alteration of only one gene copy to contribute to a premalignant or malignant phenotype, and those which are recessive, requiring alteration of both gene copies to contribute to a premalignant or malignant phenotype. Genetic alterations can happen in stem cells and differentiated cells. If those genetic alterations affect genes involved in cellular proliferation, cell-cycle regulation or apoptosis, then neoplastic growth might be initiated (Levine, 1993; Mitelman et al., 1994; Kinzler and Vogelstein, 1998; Lengauer et al., 1998; Knudson, 2001; Hahn and Weinberg, 2002). The alteration of one gene, however, does not suffice to give rise to full-blown cancer. For progression towards malignancy and invasion, further mutational hits are necessary (Knudson, 2001). Hence the risk of cancer development does not only depend on mutations initiating tumourigenesis, but also on subsequent mutations driving tumor progression.

One point that we wish to make is that biological processes such as mutagenesis can be modeled by a simple model with reasonable assumptions. Although for very complex biological system, perhaps such simple model may not be valid or can only be partially modeled. To date, the number of theoretical investigation of the kinetics of mutagenesis is scant, which is one of the reasons we have modeled this problem. This work gives an example of how an interaction between a living system and its environment can be described as a stochastic process. This work can also be viewed as a problem in non-equilibrium disordering. Here, we started with an initially ordered configuration and by applying local update rules (the dynamics), we can tract the time evolution of the degree of disordering. A more detailed model would require, for example, a complete description of the tautomeric shifts, in order to understand the kinetics of mutagenesis more fully. This may involve quantum mechanics theory since one would need to know position of the localization of the electrons in the bases as they shift from a nitrogen ion to a hydrogen ion. It should be pointed out that we have not addressed a very important issue, the survivability of the mutation. At what degree of mutagenesis is the DNA sequence not able to replicate itself. In order to correct the errors which occur during the DNA synthesis, DNA polymerase checks the newly-synthesized DNA strand and corrects most of the incorrect bases (Kornberg, 1974; Watson, 1970). It was shown experimentally that this "proof-reading" step reduces the number of mutations by a factor of 10^2 to 10^3. Such significant reduction should be also considered while comparing the calculated and observed frequencies of the mutations. Therefore, the frequency of the spontaneous GC → AT before the checking step should be in the approximate range of 10^{-6} to 10^{-8}.

Also we would like to note that the values of the predicted non-equilibrium quantities are sensitive to the level of calculations (level of theory and the basis set), which suggests that a higher-level calculations should be also performed. So far, computational calculations of the nucleotide sequences done within the framework of the "human genome" have proven to be useful since they provide deeper insight into the principle of genome organization and function. Much more work has to be done to close the gap between the complexities of real biological entities and grossly oversimplified mathematical (modeling) descriptions used to study biological and medical systems. The increase appreciation of stochasticity in biological research is observed in all scales of biological systems (Kurakin, 2005, 2006; Xie and Lu, 1999; Dundr et al., 2002; Kimura et al., 2002; Essers et al., 2002; Sirakoulis, 2004; Hoogstraten et al., 2002). We believe our model will complement detailed stochastic modeling by providing a set of powerful mathematical tools and concepts to visualize DNA alteration.

Acknowledgements

This research work was financially supported in part by Mahidol University, The Thailand Research Fund (TRF), The National Center for Engineering and Biotechnology, Thailand (BIOTEC), and the Third World Academy of Sciences (TWAS 04-446 RG/PHYS/AS). The authors thank the reviewer of the manuscript for all the valuable recommendations and comments.

References

Abramowitz, M., Stegun, I.A., 1972. Handbook of Mathematical Functions, 10th ed. Dover, New York.

Bardelli, A., Cahill, D.P., Lederer, G., Speicher, M.R., Kinzler, K.W., Vogelstein, B., Lengauer, C., 2001. Genetic instability and the quasispecies model. Proc. Natl. Aacd. Sci. 98, 5770.

Burgess, R.E., 1969. Fluctuation Phenomena in Solid. Academic Press, New York.

Claytong, A., Robertson, A., 1955. Mutation and quantitative variation. Am. Nat. 89, 151–158.

Crowj, F., Kimura, M., 1964. The theory of genetic loads. Proc. XI Int. Congr. Genet., 495–505.

Delbruck, M., Timofeeff, N.W., Zimmer, K.G., 1935. Nachr. Biol. Ges. Wiss. Gottingen 1, 189.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D., Misteli, T., 2002. A kinetic framework for a mammalian RNA polymerase in vivo. Science 298, 1623–1626.

Essers, J., Houtsmuller, A.B., van Veelen, L., Paulusma, C., Nigg, A.L., Pastink, A., Vermeulen, W., Hoeijmakers, J.H., Kanaar, R., 2002. Nuclear dynamics of RAD52 group homologous recombination proteins in response to DNA damage. EMBO J. 21, 2030–2037.

Finnegan, D.J., 1990. Transposable elements and DNA transposition in eukaryotes. Cuur. Opin. Cell Biol. 21, 471–477.

Gardiner, C.W., 1985. Handbook of Stochastic Processes, second ed. Springer-Verlag, Berlin.

Greenman, C., et al., 2007. Patterns of somatic mutation in human cancer genomes. Nature 446, 153–158.

Haber, D.A., Settleman, J., 2007. Drivers and passengers. Nature 446, 145–146.

Hahn, W.C., Weinberg, R.A., 2002. Rules for making human tumor cells. N. Engl. J. Med. 347, 1593–1603.

Hill, W., 1982. Rates of change in quantitative traits from fixation of new mutations. PNAS 79, 142–145.

Hoogstraten, D., Nigg, A.L., Heath, H., Mullenders, L.H., van Driel, R., Hoeijmakers, J.H., Vermeulen, W., Houtsmuller, A.B., 2002. Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. Mol. Cell 10, 1163–1174.

Hughes, B.H., 1995. Random Walks and Random Environments, vol. 1. Clarendon Press, Oxford.

Hussain, S.P., Harris, C.C., 1999. Mutation spectrum and load: the generation of hypotheses linking the exposure of endogeneous or exogeneous carcinogens to human cancer. Mutat. Res. 428, 33–39.

Kimura, M., 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. U. S. A. 54, 731–736.

Kimura, H., Sugaya, K., Cook, P.R., 2002. The transcription cycle of RNA polymerase II in living cells. J. Cell Biol. 159, 777–782.

Kinzler, K.W., Vogelstein, B., 1998. The Genetic Basis of Human Cancer. McGraw-Hill, Toronto, CA.

Kleckner, N., 1981. Transposable elements in prokaryotes. Annu. Rev. Genet. 15, 341–404.

Knudson, A.G., 2001. Two genetic hits to cancer. Nat. Rev. Cancer 1, 157–162.

Komarova, N.L., Lengauer, C., Vogelstein, B., Nowak, M.A., 2002. Dynamics of genetic instability in sporadic and familial colorectal cancer. Cancer Biol. Ther. 1, 685–692.

Kornberg, A., 1974. DNA Synthesis. W.H. Freeman & Co, San Francisco.

Kurakin, A., 2005. Self-organization versus watchmaker: stochastic gene expression and cell differentiation. Dev. Genes Evol. 215, 46–52.

Kurakin, A., 2006. Self-organization versus watchmaker: Molecular motors and protein translocation. Biosystems 84, 15–23.

Lander, R., 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. 26, 221–235.

Lengauer, C., Kinzler, K.W., Vogelstein, B., 1998. Genetic instabilities in human cancers. Nature 396, 643–649.

Leszczynski, J. (Ed.), 1999. Computational Molecular Biology of "Theoretical and Computational Chemistry Book Series, vol. 8. Elsevier.

Levine, A.J., 1993. The tumor suppressor genes. Annu. Rev. Biochem. 62, 623–651.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature (London) 362, 709–715.

Loeb, L.A., 2001. A mutator phenotype in cancer. Cancer Res. 61, 3230–3239.

Lynch, M., Hill, W.G., 1986. Phenotypic evolution by neutral mutation. Evolution 40, 915–935.

Mitelman, F., Johansson, B., Mertens, F., 1994. Catalog of Chromosome Aberrations in Cancer, vol. 2. Wiley–Liss, New York.

Newman, T.H., Triampo, W., 1999. Binary data corruption due to a Brownian agent. Phys. Rev. E 59 (5), 5172–5186.

Nowak, M., Schuster, P., 1989. Error tresholds of replication in finite populations, mutation frequencies and the onset of Muller's ratchet. J. Theor. Biol. 137, 375.

Radchenko, E.D., Plohotnichenko, A.M., Sheina, G.G., Blagoi, Yu.P., 1983. Ab initio study of the prototropic tautomerism of cytosine and guanine and their contribution to spontaneous point mutations. Biophysics (Engl. Ed.) 28, 559.

Roberts, J.J., 1978. The repair of DNA modified by cytotoxic, mutagenic and carcinogenic chemicals. Adv. Radiat. Biol. 7, 211–435.

Roychoudhury, A.K., Nei, M., 1988. Human polymorphic genes: world distribution. Oxford University Press, New York.

Schaaper, R.M., Dunn, R.L., 1991. Spontaneous mutation in the Escherichia coli lacI gene. Genetics 129, 317–326.

Singer, B., Kusmierek, J.T., 1982. Chemical mutagenesis. Annu. Rev. Biochem. 51, 655–693.

Sirakoulis, G.Ch., 2004. An algorithm for the study of DNA sequence evolution based on the genetic code. BioSystems 77, 11–23.

Sole, R.V., Deisboeck, T.S., 2004. An error catastrophe in cancer. J. Theor. Biol. 228, 47–54.

Watson, J.D., Crick, F.H.C., 1953. Genetic implications of the structure of deoxyribonucleic acid. Nature 171, 964–969.

Watson, J.D., 1970. Molecular Biology of Gene. W.A. Benjamin Inc., New York.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., Weiner, A.M., 1988. Molecular Biology of the Gene. The Benjamin/Cummings Publishing Company Inc., CA/Reading, MA.

Xie, X.S., Lu, H.P., 1999. Single-molecule enzymology. J. Biol. Chem. 274, 15967–15970.

Zeng, Z.B., Cockerham, C., 1993. Mutation models and quantitative genetic variation. Genetics 133, 729–736.