Evolution of Resistance During Clonal Expansion

Yoh Iwasa,*,[1] Martin A. Nowak[†] and Franziska Michor[†,‡]

*Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka 812-8581, Japan, [†]Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138 and [‡]Harvard Society of Fellows, Cambridge, Massachusetts 02138

ABSTRACT

Acquired drug resistance is a major limitation for cancer therapy. Often, one genetic alteration suffices to confer resistance to an otherwise successful therapy. However, little is known about the dynamics of the emergence of resistant tumor cells. In this article, we consider an exponentially growing population starting from one cancer cell that is sensitive to therapy. Sensitive cancer cells can mutate into resistant ones, which have relative fitness α prior to therapy. In the special case of no cell death, our model converges to the one investigated by Luria and Delbrück. We calculate the probability of resistance and the mean number of resistant cells once the cancer has reached detection size M. The probability of resistance is an increasing function of the detection size M times the mutation rate u. If $Mu \ll 1$, then the expected number of resistant cells in cancers with resistance is independent of the mutation rate u and increases with M in proportion to $M^{1-1/\alpha}$ for advantageous mutants with relative fitness $\alpha > 1$, to $\ln M$ for neutral mutants ($\alpha = 1$), but converges to an upper limit for deleterious mutants ($\alpha < 1$). Further, the probability of resistance and the average number of resistant cells increase with the number of cell divisions in the history of the tumor. Hence a tumor subject to high rates of apoptosis will show a higher incidence of resistance than expected on its detection size only.

CANCER therapy often fails because acquired resistance enables cancer cells to grow despite the continuous administration of therapy. In some cases, a single genetic alteration is sufficient to cause resistance to cancer treatment. For example, treatment of chronic myeloid leukemia with the chemotherapeutic agent imatinib fails due to resistance acquired through a single point mutation in the kinase domain of BCR-ABL, the oncogene driving this disease (GORRE et al. 2001). This also holds true for mutations in the Bcl-2 (SCHMITT et al. 2000), PI3K and RAS (GUPTA et al. 2001), or NF-κB (BENTIRES-ALJ et al. 2003) genes, respectively.

Consider the clonal expansion initiated by a single cancer cell that is sensitive to cancer therapy. The offspring of this cancer cell increase exponentially in abundance until the tumor is detected and cancer therapy is initiated. During clonal expansion, mutant cells can be generated that are resistant to therapy. In this article, we address the following questions: What is the probability that a tumor of a certain size contains resistant mutants? And what is the distribution of these mutants in the tumor?

These questions arise in many different contexts. Resistance mutations emerging during the exponential growth phase of a bacterial infection cause failure of

antibiotics treatment (LENSKI 1998; LEVIN et al. 2000; LEVY 2001). Viral infections are treated with antiviral drugs only after an initial phase of exponential growth during which resistance mutations arise (NOWAK et al. 1991; BONHOEFFER and NOWAK 1997; RIBEIRO et al. 1998; NOWAK and MAY 2000; RICHMAN 2001). Inherited childhood retinoblastoma ensues if retinal cells receive the second hit in the retinoblastoma tumor suppressor gene during the development of the retina (KNUDSON 1971). Chronic myeloid leukemia is diagnosed only after an extensive period of clonal growth and resistance is often detected at diagnosis (MICHOR et al. 2005). Of those chronic myeloid leukemia patients commencing imatinib (Gleevec) in early and late chronic phase, 6 and 14% develop resistance within the first year of therapy (BRANFORD et al. 2003).

In this article, we develop a mathematical model to calculate the probability of resistance of an exponentially growing cell population that starts from one sensitive cell and reaches detection size M. We determine the probability distribution of resistant cells in the tumor of size M. Our model is related to the question asked by LURIA and DELBRÜCK (1943) concerning the mutations of bacteria that confer resistance to phages. The distribution of the number of mutants in an exponentially growing population is called Luria–Delbrück distribution. This distribution has been studied for more than half a century (TLSTY et al. 1989; ZHENG 1999; FRANK 2003). Several different models have been

[1]Corresponding author: Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka 812-8581, Japan.
E-mail: yiwasscb@mbox.nc.kyushu-u.ac.jp

proposed that are all based on pure birth processes and do not include the possibility of cell death (ZHENG 1999). However, in most situations of cancer growth, bacterial and viral infections, death of the cancer cells or infecting agents cannot be neglected. In this article, we consider the case in which each cancer cell, virus, or bacterium experiences random birth or death. The total population size is growing exponentially with random fluctuations. Furthermore, most theoretical studies of the Luria–Delbrück distribution consider the situation in which the time period of exponential expansion is known. Conversely, in many medical situations the time between the initial start of growth and the detection of a cancer or infection is unknown, whereas the population size of the cells or infectious agents can be estimated. Therefore, we are interested in the probability of resistance until a clonally expanding population has reached a certain size M. We introduce a birth-and-death process model and calculate the number of resistant mutants in the population once it has reached size M. The probability distribution depends on the growth rate of the exponentially expanding agents, the mutation rate per replication, the relative fitness of mutants, and the detection size M. We derive formulas for the probability of resistance, for the mean number of mutants, and for the distribution of the number of mutants. These results are confirmed by exact computer simulations of the stochastic process.

THE MODEL

Consider an exponentially expanding cancer cell population starting from one cell that is sensitive to therapy. The cell population follows a continuous-time branching process. In a short time interval of length Δt, a sensitive cancer cell divides with probability $r\Delta t$ and dies with probability $d\Delta t$. A resistant cancer cell arises by mutation with probability u per cell division. Resistant cancer cells also follow a continuous-time branching process. In a time interval of length Δt, a resistant cancer cell divides with probability $a\Delta t$ and dies with probability $b\Delta t$. The relative fitness of resistant cancer cells as compared to sensitive cancer cells is given by $\alpha = (a - b)/(r - d)$. Resistant cancer cells can be advantageous ($\alpha > 1$), neutral ($\alpha = 1$), or deleterious ($\alpha < 1$). Both sensitive and resistant cancer cells grow on average exponentially, which means $r > d$ and $a > b$. Since the cancer cell population is initiated by a single cell, it might go extinct due to random events. Otherwise, the cancer cell population reaches detection size M. For the sake of mathematical simplicity, we assume that clinical detection occurs when the population size of sensitive cancer cells reaches M.

Denote by R_x the expected number of resistant cancer cells produced by a tumor with exactly x sensitive cancer cells. We assume that the number of resistant cancer cells follows a Poisson distribution with mean R_x.

Consider the clones each initiated by one resistant cancer cell. Their population sizes follow continuous-time branching processes. Let Z be the number of resistant cancer cells at the time of diagnosis. The generating function of Z is given by $G(\xi) = E[\xi^Z]$. The generating function is useful for the calculation of the probability of resistance, of the average number of resistant cells, and also of higher moments of Z. The total number of resistant cells, Z, is the sum of the number of resistant cancer cells in each clone. The generating function for the number of resistant cancer cells at diagnosis in the clone whose initiating resistant cell was produced when there were x sensitive cancer cells is denoted by $g_x(\xi)$. The generating function of the total number of resistant cells, Z, is given by

$$G(\xi) = \exp\left[-\sum_{x=1}^{M-1} R_x(1 - g_x(\xi))\right]. \quad (1)$$

See APPENDIX A for derivation of Equation 1.

Let us consider the Markov chain of the number of sensitive cancer cells, x. Their number follows a birth-and-death process (FELLER 1971). The cell division rate is denoted by r, the cell death rate is denoted by d, and resistance mutants arise with probability u per cell division. We consider the process starting from $x = 1$; it can end either with extinction, $x = 0$, or with exponential increase until reaching the detection size, $x = M$. The transition rate from x to $x + 1$ is given by $rx(1 - u)$, which can be approximated as rx because $u \ll 1$. The transition rate from x to $x - 1$ is given by dx. Both 0 and M are absorbing states of the process. Mutation does not change the state.

Let $f_x(t)$ be the probability that there are x sensitive cells at time t. Then we have

$$\frac{df_x}{dt} = r(x - 1)f_{x-1} + d(x + 1)f_{x+1} - (r + d)xf_x, \quad (2a)$$

where $x = 2, 3, \ldots, M - 2$. Furthermore, we have

$$\frac{df_1}{dt} = 2df_2 - (r + d)f_1, \quad (2b)$$

$$\frac{df_{M-1}}{dt} = r(M - 2)f_{M-2} - (r + d)(M - 1)f_{M-1}. \quad (2c)$$

The initial condition is $f_1(0) = 1$ and $f_x(0) = 0$ for $x = 2, 3, \ldots, M - 1$. The production rate of resistant cancer cells is given by the product of the cell division rate, r, the mutation rate per cell division, u, and the number of sensitive cancer cells, x. The expected time until the number of sensitive cancer cells is x is given by $\int_0^\infty f_x(t)dt$. Then the expected number of resistant cancer cells produced when there are x sensitive cancer cells is given by

$$R_x = \frac{rux}{1 - d/r}\int_0^\infty f_x(t)dt. \quad (3)$$

The factor $1 - d/r$ arises because we consider the conditional mean number of resistant cancer cells and exclude evolutionary trajectories in which sensitive cancer cells go extinct.

In APPENDIX B, we can derive $R_x \approx u/(1 - d/r)$. Therefore, equal numbers of resistant cancer cells are produced by any number of sensitive cancer cells, $x = 1$, $2, \ldots, M - 1$. Then the generating function, Equation 1, becomes

$$G(\xi) = \exp\left[-\frac{u}{1 - d/r} \sum_{x=1}^{M-1}(1 - g_x(\xi))\right]. \qquad (4)$$

After a single resistant cancer cell has been created at time t, the number of resistant cancer cells follows a continuous-time branching process with the generating function

$$\bar{g}(\xi, t) = \frac{(\xi - 1)(b/a)e^{(a-b)t} - (\xi - b/a)}{(\xi - 1)e^{(a-b)t} - (\xi - b/a)}. \qquad (5)$$

The derivation is explained in APPENDIX C. Equation 5 is the generating function for the number of resistant cancer cells at time t after the occurrence of the resistance mutation. Here we are interested in the number of resistant cancer cells once the number of sensitive cancer cells has reached detection size M. The time period during which the number of sensitive cancer cells increases from x to M is not a fixed constant, but rather a stochastic variable. Assuming deterministic population growth of the sensitive cancer cell population once a resistant cancer cell has been produced, we have $M = x\exp[(r - d)t]$. With $e^t = (M/x)^{1/(r-d)}$ and $\alpha = (a - b)/(r - d)$, the generating function becomes

$$g_x(\xi) = \frac{(\xi - 1)(b/a)(M/x)^\alpha - (\xi - b/a)}{(\xi - 1)(M/x)^\alpha - (\xi - b/a)}. \qquad (6)$$

THE PROBABILITY OF RESISTANCE

Let us first calculate the probability of resistance. This probability is given by 1 minus the probability of extinction of resistant cancer cells (Equation 6 with $\xi = 0$). The probability of nonextinction of a clone of resistant cancer cells is given by

$$1 - g_x(0) = \left(1 - \frac{b}{a}\right) \bigg/ \left[1 - \frac{b}{a}\left(\frac{x}{M}\right)^\alpha\right].$$

Then the probability of resistance is given by

$$P = 1 - G(0) = 1 - \exp\left[-\frac{MuF}{1 - d/r}\right]. \qquad (7)$$

Here $F = \int_0^1 (1 - b/a)/(1 - (b/a)y^\alpha)\,dy$ and $\alpha = (a - b)/(r - d)$. The sum over x was replaced by an integral because M is very large. Equation 7 can be used for any fitness of resistant cancer cells as compared to sensitive

cancer cells. Resistant cancer cells can be advantageous ($\alpha > 1$), neutral ($\alpha = 1$), or deleterious ($\alpha < 1$). If $Mu \ll 1$, then we can approximate Equation 7 as $P = MuF/(1 - d/r)$.

When resistant cancer cells are neutral as compared to sensitive cells, we have $a = r$ and $b = d$. Then Equation 7 can be simplified as

$$P = 1 - \exp\left[-Mu\frac{r}{d}\ln\frac{r}{r - d}\right]. \qquad (8)$$

The traditional Luria–Delbrück distribution assumes that there is no cell death, $b = d = 0$. In that case, Equation 7 becomes

$$P = 1 - \exp[-Mu]. \qquad (9)$$

This result can also be derived from the fact that it takes $M - 1$ cell divisions until one cell has generated a population of M cells; if resistant mutants are produced at rate u per cell division and the produced mutants survive until detection time, then the probability of resistance is given by Equation 9.

The analytic formula for the probability of resistance is confirmed by the exact computer simulation of the stochastic process. We define two integer variables for the numbers of sensitive and resistant cancer cells, $x \in \{0, 1, \ldots, M\}$ and $y \in \{0, 1, \ldots, M\}$, subject to the constraint $x + y \le M$. Each process is initiated with one sensitive cancer cell, $x = 1$. Let $\Gamma = (r + d)x + (a + b)y$. The transition probabilities between states are given by

$$\Pr[(x, y) \to (x + 1, y)] = (rx(1 - u))/\Gamma$$

$$\Pr[(x, y) \to (x - 1, y)] = dx/\Gamma$$

$$\Pr[(x, y) \to (x, y + 1)] = (rxu + ay)/\Gamma$$

$$\Pr[(x, y) \to (x, y - 1)] = by/\Gamma.$$

Once the total population size reaches the detection size, $x + y = M$, the process is stopped.

We perform many independent runs to account for random fluctuations. Then we determine the fraction of runs in which resistant cells were produced. Figure 1 shows the very good agreement between the analytic formula for the probability of resistance, P, given by Equations 7 and 8 (lines) and the exact computer simulation of the stochastic process for the probability of resistance (circles).

Figure 1 also shows that the probability of resistance, P, depends on the product Mu rather than on M and u separately. If $Mu \ll 1$, then P is proportional to Mu (Figure 1a). If the relative fitness of resistant cells, α, is constant, then P decreases with increasing growth rate of sensitive cells, r (Figure 1b). Therefore, a slowly growing cancer is more likely to develop resistance than a quickly growing one, because the former undergoes a larger number of cell divisions than the latter until reaching a certain size. If the growth rate of sensitive cancer cells, r, is constant, then P increases with the
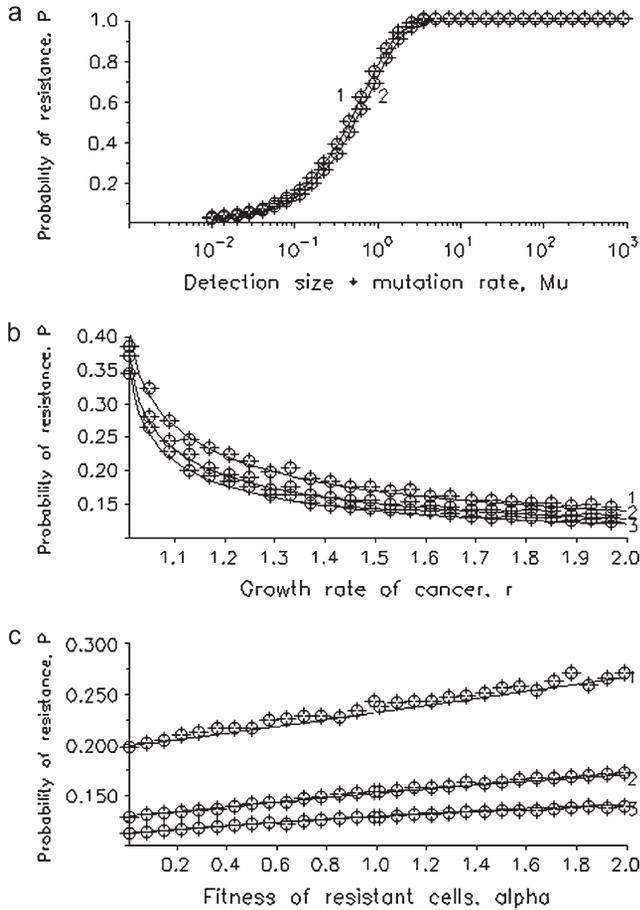
FIGURE 1.—The probability of resistance. We show the probability of resistance in dependence of (a) the detection size times the mutation rate, Mu, (b) the growth rate of sensitive cancer cells, r, and (c) the fitness of resistant cancer cells, α. Equations 7 and 8 are shown as lines and direct computer simulations as circles. Parameter values are $M = 10^4$, $d = b = 1$, and (a) $r = 2$, $\alpha = 2$ (line 1), and $\alpha = 0.5$ (line 2); (b) $u = 10^{-5}$, $\alpha = 2$ (line 1), $\alpha = 1$ (line 2), and $\alpha = 0.5$ (line 3); and (c) $u = 10^{-5}$, $r = 1.1$ (line 1), $r = 1.5$ (line 2), and $r = 2$ (line 3). The results of the stochastic computer simulation are averaged over 100,000 runs.

relative fitness of resistant cells, α (Figure 1c). This implies that advantageous mutations are more dangerous than deleterious mutations under the assumption of constant mutation rates.

THE AVERAGE NUMBER OF RESISTANT CELLS

Let us now calculate the average number of resistant cancer cells from the generating function, Equation 4. We consider the conditional average, *i.e.*, the average number of resistant cells in cancers that have developed resistance before reaching size M. This average number is related to the time period during which cancer therapy is effective. The conditional average number of resistant cancer cells is given by $\bar{Y} = G'(1)/P$. Note that $\partial \bar{g}/\partial \xi(1, t) = e^{(a-b)t} \approx (M/x)^{\alpha}$. Then we have

$$\bar{Y} = \frac{G'(1)}{P} = \frac{u}{P(1 - d/r)} \sum_{x=1}^{M-1} \left(\frac{M}{x}\right)^{\alpha}$$
$$\approx \frac{Mu\left(1 - 1/M^{1-\alpha}\right)}{P(1 - d/r)(1 - \alpha)}. \tag{10}$$

Here $P = 1 - \exp[-MuF/(1 - d/r)]$ and $F = \int_0^1 (1 - b/a)/(1 - y^{\alpha}b/a)\,dy$.

If $Mu \ll 1$, we can further simplify as

$$\bar{Y} = \frac{1 - M^{\alpha-1}}{(1 - \alpha)F}. \tag{11}$$

If resistant cells are neutral, $\alpha = 1$ because $a = r$ and $b = d$, Equation 10 becomes

$$\bar{Y} = \frac{1}{(r/d - 1)\ln(r/(r - d))} \sum_{x=1}^{M-1} \frac{1}{x}$$
$$\approx \frac{\ln M}{(r/d - 1)\ln(r/(r - d))}. \tag{12}$$

Figure 2 shows the agreement between the analytic formula for the conditional average number of resistant cancer cells given by Equations 10 and 12 (line) and the exact computer simulation of the stochastic process (circles). If resistant cancer cells are deleterious or neutral, then there is very good agreement between the analytic prediction and the exact computer simulation. If resistant cells are advantageous, however, the equations above greatly overestimate the number of resistant cells. Therefore, we consider a different approach for calculating the average number of advantageous resistant cancer cells in the next section.

If resistant cells are deleterious or neutral, then the dependence of the conditional average number of resistant cancer cells, $\bar{Y}$, on the mutation rate, u, is weak. If Mu is small, $\bar{Y}$ is even independent of the mutation rate. Note that $\bar{Y}$ is the conditional average of resistant cells—the probability of resistance, however, increases with the mutation rate, as discussed above. The dependence of $\bar{Y}$ on the detection size, M, is less than linear. If resistant cancer cells are neutral, then $\bar{Y}$ increases in proportion to $\ln M$. If resistant cancer cells are deleterious, then $\bar{Y}$ does not increase strongly with M and converges to an upper limit (Figure 2a). Furthermore, $\bar{Y}$ decreases with increasing growth rate of sensitive cancer cells, r, but the decrease is faster for deleterious than for neutral resistant cancer cells (Figure 2b). Finally, $\bar{Y}$ increases with the relative fitness of resistant cancer cells, α (Figure 2c).

HEURISTIC FORMULA FOR THE NUMBER OF ADVANTAGEOUS RESISTANT CANCER CELLS

An unexpected result of our analysis is that the number of resistant cancer cells, R_x, produced by x sensitive cancer cells is almost independent of x. If we classify new resistant cancer cells according to the number of
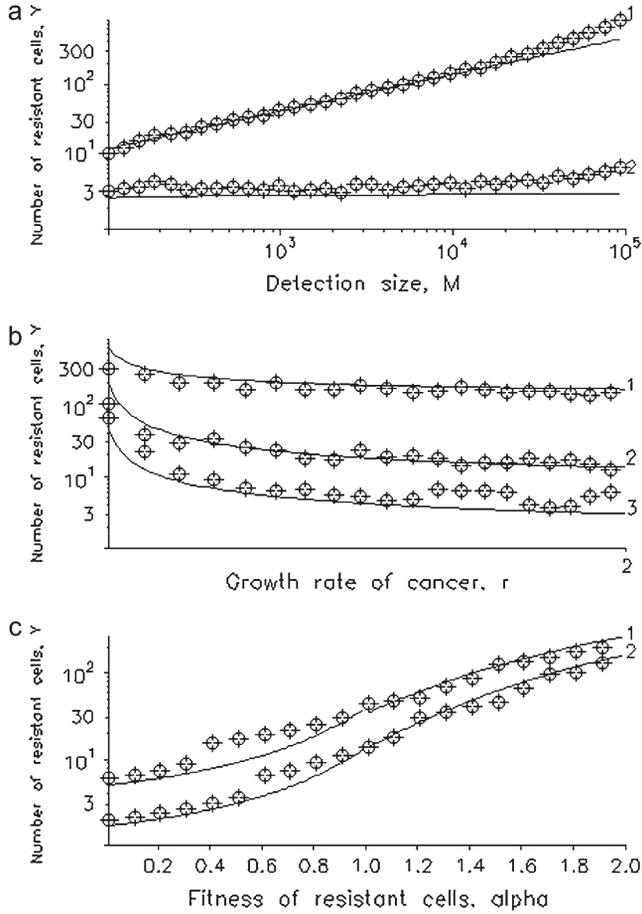
FIGURE 2.—The average number of resistant cells. We show the expected average number of resistant cells in dependence of (a) the detection size, M, (b) the growth rate of sensitive cancer cells, r, and (c) the fitness of resistant cancer cells, α. Equations 11–13 are shown as lines and direct computer simulations as circles. Parameter values are $d = b = 1$ and (a) $r = 2$, $\alpha = 2$ (line 1), and $\alpha = 0.5$ (line 2); (b) $M = 10^4$, $u = 10^{-5}$, $\alpha = 2$ (line 1), $\alpha = 1$ (line 2), and $\alpha = 0.5$ (line 3); and (c) $M = 10^4$, $u = 10^{-5}$, $r = 1.1$ (line 1), and $r = 2$ (line 2). The results of the stochastic computer simulation are averaged over 100,000 runs.

sensitive cancer cells present at the time of the production of resistant cancer cells, then the latter are evenly distributed from 1 to $M - 1$. Once one resistant cancer cell has been produced, its offspring increase in number until reaching $y = e^{(a-b)t} \approx (M/x)^\alpha$. This is the expected number of resistant cancer cells once the sensitive cancer cells, x, reach detection size M, and it is the average over exponentially growing lineages and extinct lineages. The mean number of resistant cancer cells conditional to nonextinction is given by $(M/x)^\alpha/F$, where $F = \int_0^1 (1 - b/a)/(1 - y^\alpha b/a)dy$. Equation 10 gives the average number of resistant cancer cells when x is distributed uniformly from 1 to $M - 1$. However, the conditional number of resistant cancer cells, $(M/x)^\alpha/F$, can increase beyond M. Hence Equation 10 or 11 cannot be used for advantageous resistant cells, $\alpha > 1$.

We introduce a truncation as follows. The stochastic process stops when the sum of sensitive and resistant cancer cells reaches the detection size, $x + y = M$. Let us approximate this process in the model by stopping when the number of sensitive cancer cells reaches half of the threshold size, $M/2$. We define a threshold value of x as $(M/x_c)^\alpha/F = M/2$, leading to $x_c = M^{1-1/\alpha} \cdot (2/F)^{1/\alpha}$. Then we replace $(M/x)^\alpha/F$ in Equation 11 by $M/2$ for $x \le x_c$. Hence, if $Mu \ll 1$, we have

$$
\bar{Y} = \frac{1}{M}\left(\frac{M}{2}x_c + \sum_{x=x_c}^{M-1}(M/x)^\alpha/F\right)
$$
$$
\approx \frac{\alpha(M/2)^{1-1/\alpha}}{(\alpha-1)F^{1/\alpha}} - \frac{1}{(\alpha-1)F}. \tag{13}
$$

Figure 2c shows the agreement of Equation 13 (lines) with the direct computer simulation (circles). Equation 13 shows the parameter dependence of the conditional average number of resistant cancer cells, $\bar{Y}$, very clearly. First, $\bar{Y}$ is independent of the mutation rate, u, and of the growth rate of sensitive cancer cells, r. Note that only the conditional mean number of resistant cancer cells is independent of u and r; the probability of resistance, P, increases with u and decreases with r. Second, $\bar{Y}$ increases as a power function of M. Third, $\bar{Y}$ increases with increasing fitness of resistant cells, α.

When there is no cell death, $b = d = 0$, and when $Mu \ll 1$, we have

$$
\bar{Y} \approx \frac{1 - M^{\alpha-1}}{1-\alpha} \qquad \text{if } \alpha < 1, \tag{14a}
$$

$$
\bar{Y} \approx \ln M \qquad \text{if } \alpha = 1, \tag{14b}
$$

$$
\bar{Y} \approx \frac{\alpha}{\alpha-1}(M/2)^{1-1/\alpha} - \frac{1}{\alpha-1} \quad \text{if } \alpha > 1. \tag{14c}
$$

This result corresponds to the traditional Luria–Delbrück distribution (ZHENG 1999).

THE DISTRIBUTION OF THE NUMBER OF RESISTANT CANCER CELLS

On the basis of the approximation for the deterministic growth of resistant cancer cells, we can derive the distribution of the number of resistant cancer cells at detection time. We use the result of the uniform distribution of the number of sensitive cancer cells when a resistant cancer cell arises. The number of resistant cancer cells is given by $(M/x)^\alpha/F$, where $\alpha = (a - b)/(r - d)$ and $F = \int_0^1 (1 - b/a)/(1 - y^\alpha b/a)dy$. Let $1 < y_1 < y_2 < M$. The probability that the number of resistant cancer cells is greater than y_1 is given by

$$
P(Y \ge y_1) = \Pr[(M/x)^\alpha/F \ge y_1]
$$
$$
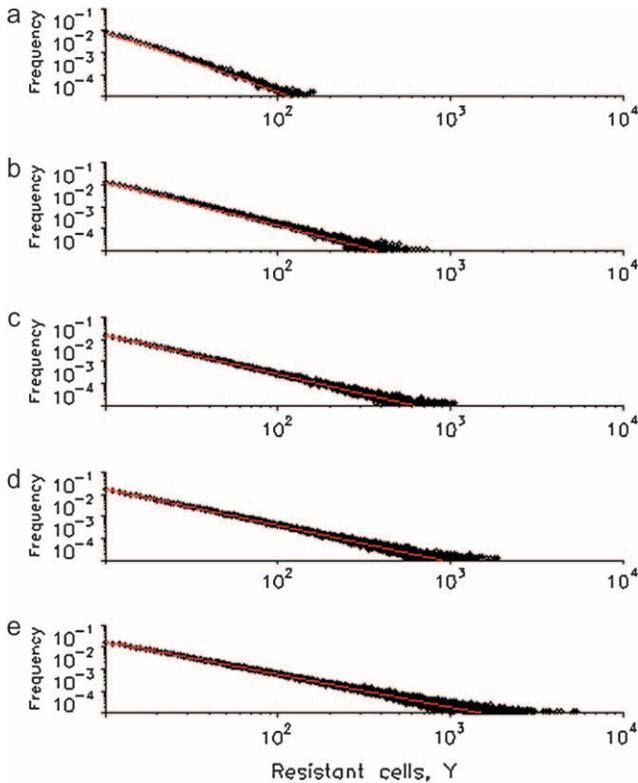= \Pr[x \le M/(Fy_1)^{1/\alpha}] = 1/(Fy_1)^{1/\alpha}. \tag{15a}
$$

FIGURE 3.—The distribution of resistant cells. We show the probability distribution of resistant cells for different fitness values of resistant cells, α. Equations 15b and 16 are shown as red lines and direct computer simulations as circles. Parameter values are $M = 10^4$, $u = 10^{-5}$, $r = 2$, and $d = b = 1$ and (a) $\alpha = 0.5$, (b) $\alpha = 1$, (c) $\alpha = 1.25$, (d) $\alpha = 1.5$, and (e) $\alpha = 2$. The results of the stochastic computer simulation are averaged over 10^7 runs.

Note that this and all the following expressions give the probabilities conditional to the existence of some resistant cancer cells at diagnosis. These formulas apply if y is clearly smaller than M. Similarly, the probability that the number of resistant cells is between y_1 and y_2 is given by

$$P(y_1 < Y \leq y_2) = 1/(Fy_1)^{1/\alpha} - 1/(Fy_2)^{1/\alpha}. \quad (15b)$$

Figure 3 shows the agreement between the computer simulation and Equation 15. If $y_1 \leq 10\%$ of the threshold M, then Equation 15 fits very well to the results of the computer simulation.

The same argument applies to the case of deleterious or neutral resistant cancer cells ($\alpha \leq 1$). Here the conditional average number of resistant cancer cells reaches less than the maximum, M, but Equation 15 holds for $1 < y_1 < y_2 < M^\alpha/F < M$. Figure 3 illustrates the results of the computer simulations and Equation 15.

Comparing these results with direct computer simulations, we find that the formulas above are very accurate for advantageous resistant cancer cells with $y < 10\%$ of M. However, they are not accurate for deleterious resistant cancer cells, especially for small y, because the

logic of Equations 15 relies on the exponential growth of the resistant cancer cell lineages. Many deleterious resistant cancer cells eventually go extinct, but might not yet have disappeared at the time of detection. This effect causes an underestimation for small y. In APPENDIX D, we derive a formula for the number of resistant cancer cells that follows from the generating function (Equations 1 and 6) by Taylor expansion with respect to ξ. When $Mu \ll 1$, we have

$$\Pr[Y = y] = \frac{(1 - b/a)^2}{\alpha F} \int_0^1 \frac{z^{1/\alpha}(1 - z)^{y-1}}{(1 - (b/a)z)^{y+1}}dz,$$
$$y = 1, 2, 3, \ldots. \quad (16)$$

Equation 16 is especially useful for small y and is accurate for deleterious resistant cancer cells. If there is no cell death, we have explicit solutions (see APPENDIX D). Equation 15 is accurate for advantageous and neutral resistant cells, and Equation 16 is accurate for deleterious resistant cells (Figure 3).

DISCUSSION

In this article, we have analyzed the dynamics of resistance mutations emerging from an exponentially growing population of cancer cells. Resistant cancer cells can be advantageous, neutral, or deleterious as compared to sensitive cancer cells in the absence of therapy. We use a continuous-time branching process to calculate the probability of resistance once the tumor has reached a certain size. We also calculate the probability distribution of resistant cancer cells conditional to the emergence of resistance in the tumor until clinical diagnosis. The accuracy of our model is confirmed by exact computer simulations of the underlying stochastic process. Tables 1 and 2 show the formulas for the probability of resistance and for the average number of resistant cancer cells with and without cell death, respectively.

Consider the number of resistant cancer cells that arise when the number of sensitive cancer cells is x. One might expect that more resistant cancer cells are produced when x is large, because more sensitive cancer cells divide more often. On the other hand, if x is large, events changing x (i.e., cell death and cell division) occur faster, and the number of sensitive cancer cells, x, stays at any given value for a shorter time. Our model shows that exactly the same number of resistant cancer cells is produced by different numbers of sensitive cancer cells, x. This uniform distribution allows us to derive a formula for the probability that the number of resistant cells at diagnosis exceeds a certain number.

The probability of resistance, P, is given by Equation 7. This equation is accurate for both deleterious and advantageous resistant cancer cells. If resistant cancer cells are neutral as compared to sensitive cancer cells, then the probability of resistance is given by an even

TABLE 1

The probability of resistance and the average number of resistant cells with cell death

	Deleterious resistant cells ($\alpha < 1$)	Neutral resistant cells ($\alpha = 1$)	Advantageous resistant cells ($\alpha > 1$)
Probability of resistance	$P = 1 - e^{-MuF/(1-d/r)}$	$P = 1 - e^{-Mu(r/d)\ln(r/(r-d))}$	$P = 1 - e^{-MuF/(1-d/r)}$
Expected average no. of resistant cells	$\bar{Y} = \dfrac{1 - M^{\alpha-1}}{(1-\alpha)F}$	$\bar{Y} = \dfrac{\ln M}{(r/d - 1)\ln(r/(r-d))}$	$\bar{Y} = \dfrac{1}{\alpha - 1}\left[\dfrac{\alpha}{F^{1/\alpha}}\left(\dfrac{M}{2}\right)^{1-(1/\alpha)} - \dfrac{1}{F}\right]$

The formulas are shown for the probability of resistance, P, and the expected average number of resistant cells, $\bar{Y}$, for deleterious ($\alpha < 1$), neutral ($\alpha = 1$), and advantageous ($\alpha > 1$) resistant cells. Parameter values are M, detection size; u, mutation rate per cell division; r and d, growth and death rates of sensitive cancer cells; and a and b, growth and death rates of resistant cancer cells. $F = \int_0^1 (1 - b/a)/(1 - (b/a)y^\alpha)dy$ and $\alpha = (a - b)/(r - d)$.

simpler formula (Equation 8). The probability of resistance, P, increases with the detection size, M, the mutation rate, u, and the relative fitness of resistant cells, α. Conversely, P decreases with increasing growth rate of sensitive cancer cells, r. This implies that the probability of resistance increases with the number of cell divisions that occurred until clinical detection of the cancer. The number of cell divisions increases if the growth rate is small or the death rate is large. Therefore, cancer cell clones with genetic instability (LENGAUER *et al.* 1998; LOEB 2001; NOWAK *et al.* 2002) might have a larger probability of generating resistant cells not only because of increased mutation rates, but also due to increased apoptosis rates, as such cell clones are more likely to receive deleterious or lethal mutations.

We derive equations for the mean number of resistant cancer cells, $\bar{Y}$, conditional that the cancer has produced at least one resistant cell until diagnosis. Interestingly, $\bar{Y}$ is independent of the mutation rate, u. However, $\bar{Y}$ increases with the relative fitness of resistant cells, α. The dependence of $\bar{Y}$ on the detection size M differs among neutral, deleterious, and advantageous resistant cancer cells: $\bar{Y}$ increases as $M^{1-1/\alpha}$ if they are advantageous and increases as $\ln M$ if they are neutral, but it increases and saturates to an upper limit for large M if resistant cells are deleterious. Note that this result holds for the number of resistant cancer cells averaged over tumors with one or more resistant cells and only when $Mu \ll 1$.

One intriguing result is that the conditional average number of resistant cancer cells, $\bar{Y}$, does not increase with M if M is large and if resistant cancer cells are deleterious (see Equations 11 and 14a). The conditional number of deleterious resistant cancer cells increases exponentially, but their growth rate is slower than the growth rate of sensitive cancer cells. However, the number of tumors with newly emerging resistance increases. Therefore, the number of resistant cancer cells averaged over tumors with preexisting resistance and tumors with newly emerging resistance is constant for very large M.

The unconditional mean number of resistant cancer cells at diagnosis is given by $P\bar{Y}$. This number always increases with the detection size, M. It is proportional to M for deleterious resistant cells, but is faster than linear for advantageous resistant cells. Therefore, a decrease in the detection size reduces the risk of resistance, especially if resistance is caused by advantageous mutations. In this case, however, resistance caused by deleterious mutations remains unchanged and the relative importance of deleterious mutations increases.

The total risk of resistance, $P\bar{Y}$, increases with the mutation rate, u, and decreases with the growth rate of sensitive cancer cells, r. Therefore, cancer cell clones

TABLE 2

The probability of resistance and the average number of resistant cells with no cell death

	Deleterious resistant cells ($\alpha < 1$)	Neutral resistant cells ($\alpha = 1$)	Advantageous resistant cells ($\alpha > 1$)
Probability of resistance	$P = 1 - e^{-Mu}$	$P = 1 - e^{-Mu}$	$P = 1 - e^{-Mu}$
Expected average no. of resistant cells	$\bar{Y} = \dfrac{1 - M^{\alpha-1}}{1-\alpha}$	$\bar{Y} = \ln M$	$\bar{Y} = \dfrac{\alpha(M/2)^{1-(1/\alpha)} - 1}{\alpha - 1}$

The formulas are shown for the probability of resistance, P, and the expected average number of resistant cells, $\bar{Y}$, for deleterious ($\alpha < 1$), neutral ($\alpha = 1$), and advantageous ($\alpha > 1$) resistant cells. Parameter values are M, detection size; u, mutation rate per cell division; r and $d = 0$, growth and death rates of sensitive cancer cells; and a and $b = 0$, growth and death rates of resistant cancer cells.

with genetic instability have a much greater risk of resistance than cells with normal mutation rates.

Here we assumed constant rates of cell division and cell death. However, the tumor growth rate is likely to slow down once its size becomes very large due to interference between cancer cells. Our results suggest that the reduction of $r - d$ would enhance the risk of resistance. A detailed analysis of the effect of cell–cell interactions is an important subject of future theoretical investigations.

LITERATURE CITED

Bentires-Alj, M., V. Barbu, M. Fillet, A. Chariot, B. Relic et al., 2003 Nf-kb transcription factor induces drug resistance through mdr1 expression in cancer cells. Oncogene **22:** 90–97.

Bonhoeffer, S., and M. A. Nowak, 1997 Pre-existence and emergence of drug resistance in HIV-1 infection. Proc. Biol. Sci. **264:** 631–637.

Branford, S., Z. Rudzki, A. Grigg, J. F. Seymour, K. Taylor et al., 2003 The incidence of BCR-ABL kinase mutations in chronic myeloid leukemia patients is as high in the second year of imatinib therapy as the first but survival after mutation detection is significantly longer for patients with mutations detected in the second year of therapy. Blood **102:** 414.

Feller, W., 1971 *Introduction to Probability Theory and Its Applications*, Vol. II. John Wiley & Sons, New York.

Frank, S. A., 2003 Somatic mosaicism and cancer: inference based on a conditional Luria-Delbrück distribution. J. Theor. Biol. **223:** 405–412.

Gorre, M. E., M. Mohammed, K. Ellwood, N. Hsu, R. Paquette et al., 2001 Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science **293:** 876–880.

Gupta, A. K., V. J. Bakanauskas, G. J. Cerniglia, Y. Chen, E. J. Bernhard et al., 2001 The ras radiation resistance pathway. Cancer Res. **61:** 4278–4282.

Iwasa, Y., F. Michor and M. A. Nowak, 2004 Evolutionary dynamics of invasion and escape. J. Theor. Biol. **226:** 205–214.

Knudson, A. G., 1971 Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA **68:** 820–823.

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

Lenski, R. E., 1998 Bacterial evolution and the cost of antibiotic resistance. Int. Microbiol. **1:** 265–270.

Levin, B. R., V. Perrot and N. Walker, 2000 Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics **154:** 985–997.

Levy, S. B., 2001 Antibiotic resistance: consequences of inaction. Clin. Infect. Dis. **33:** 124–129.

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

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

Michor, F., T. P. Hughes, Y. Iwasa, S. Branford, N. P. Shah et al., 2005 Dynamics of chronic myeloid leukemia. Nature **435:** 1267–1270.

Nowak, M. A., and R. M. May, 2000 *Virus Dynamics*. Oxford University Press, Oxford.

Nowak, M. A., R. M. Anderson, A. R. McLean, T. Wolfs, J. Goudsmith et al., 1991 Antigenic diversity thresholds and the development of AIDS. Science **254:** 963–969.

Nowak, M. A., N. L. Komarova, A. Sengupta, P. V. Jallepalli, I. M. Shih et al., 2002 The role of chromosomal instability in tumor initiation. Proc. Natl. Acad. Sci. USA **99:** 16226–16231.

Ribeiro, R. M., S. Bonhoeffer and M. A. Nowak, 1998 The frequency of resistant mutant virus before antiviral therapy. AIDS **12:** 461–465.

Richman, D. D., 2001 HIV chemotherapy. Nature **410:** 995–1001.

Schmitt, C. A., C. T. Rosenthal and S. W. Lowe, 2000 Genetic analysis of chemoresistance in primary murine lymphomas. Nat. Med. **6:** 1029–1035.

Tlsty, T. D., B. H. Margolin and K. Lum, 1989 Difference in the rates of gene amplification in nontumorigenic and tumorigenic cell-lines as measured by Luria-Delbrück fluctuation analysis. Proc. Natl. Acad. Sci. USA **86:** 9441–9445.

Zheng, Q., 1999 Progress of a half century in the study of the Luria-Delbrück distribution. Math. Biosci. **162:** 1–32.

APPENDIX A: DERIVATION OF EQUATION 1

Let k_x be the number of resistant cancer cells produced when there are $x = 1, 2, \ldots, M - 1$ sensitive cancer cells. The clone produced by each resistant cancer cell contributes to the number of resistant cells at diagnosis. The number of resistant cells at diagnosis has the generating function $g_x(\xi)$. From the assumption that $k_1, k_2, \ldots, k_{M-1}$ follow a Poisson distribution with mean $R_1, R_2, \ldots, R_{M-1}$ and that they are independent of each other, their probability distribution is given by $\prod_{x=1}^{M-1} \left((R_x^{k_x}/k_x!)e^{-R_x} \right)$. The generating function of the total number of resistant cancer cells is given by

$$G(\xi) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{M-1}=0}^{\infty} \prod_{x=1}^{M-1} \left(\frac{R_x^{k_x}}{k_x!} e^{-R_x} g_x(\xi)^{k_x} \right)$$

$$= \prod_{x=1}^{M-1} \exp(R_x g_x(\xi) - R_x)$$

$$= \exp\left(-\sum_{x=1}^{M-1} R_x(1 - g_x(\xi)) \right). \tag{A1}$$

This is Equation 1.

APPENDIX B: SOLUTION OF EQUATION 3

We consider the integral of both sides of Equation 2 from $t = 0$–∞. We note $f_1(0) = 1$ and $f_x(0) = 0$ for $x = 2, 3, 4, \ldots, M - 1$. With $h_x = x \int_0^\infty f_x(t)\, dt$, we have

$$0 = rh_{x-1} + dh_{x+1} - (r + d)h_x, \quad \text{where } x = 2, 3, \ldots, M - 2, \tag{B1a}$$

$$-1 = dh_2 - (r + d)h_1, \tag{B1b}$$

$$0 = rh_{M-2} - (r + d)h_{M-1}. \tag{B1c}$$

From Equations 2a–2c, Equation B1b becomes $h_2 - h_1 = (r/d)h_1 - 1/d$, and then Equation B1a becomes $h_{x+1} - h_x = (r/d)^{x-1}((r/d)h_1 - 1/d)$. Hence $h_x = (1/(r/d - 1))\big[((r/d)^x - 1)h_1 - (1/d)((r/d)^{x-1} - 1)\big]$. From Equation B1c, we have

$$h_1 = \frac{(r/d)^{M-2} - (r/d)^{M-3} - (r/d)^{M-4} + 1}{d\big((r/d)^{M-1} - (r/d)^{M-2} - (r/d)^{M-3} + 1\big)} \approx \frac{1}{r}.$$

We further have

$$h_x = \frac{(r/d)^{x-1} + (r/d)^{M-4}\big((r/d)^2 - r/d - 1\big)}{d\big((r/d)^{M-1} - (r/d)^{M-2} - (r/d)^{M-3} + 1\big)} \approx \frac{1}{r}, \quad x = 2, 3, \ldots, M - 1 \tag{B2}$$

except for those x that are very close to M. Using this result, we have

$$R_x = \frac{ru}{1 - d/r} h_x \approx \frac{u}{1 - d/r}. \tag{B3}$$

APPENDIX C: DERIVATION OF THE BRANCHING PROCESS GENERATING FUNCTION

Let $\bar{g}(\xi, t) = E\big[\xi^{Z(t)} \mid Z(0) = 1\big]$. Considering the events happening in a short time interval of length Δt, we have

$$\bar{g}(\xi,\, t + \Delta t) = a\Delta t \bar{g}(\xi,\, t)^2 + b\Delta t \cdot 1 + (1 - (a + b)\Delta t)\bar{g}(\xi,\, t).$$

We have used $E\big[\xi^{Z(t)} \mid Z(0) = 2\big] = E\big[\xi^{Z(t)} \mid Z(0) = 1\big]^2 = \bar{g}(\xi,\, t)^2$, because two lineages starting from different cells at a given time behave independently (IWASA *et al.* 2004).

In the limit of $\Delta t \to 0$, this leads to

$$\partial\bar{g}/\partial t = (b - a\bar{g})(1 - \bar{g}). \tag{C1}$$

Using the initial condition $\bar{g}(\xi, 0) = \xi$, we have Equation 5 in the text. It satisfies $\bar{g}(1, t) = 1$ (because the sum of all probabilities is 1), $\partial\bar{g}/\partial\xi(1, t) = e^{(a-b)t}$, which is the mean number of mutants, and $\bar{g}(0, t) = (b/a)\big(e^{(a-b)t} - 1\big)/\big(e^{(a-b)t} - b/a\big)$, which is the extinction probability.

APPENDIX D: THE NUMBER OF RESISTANT CELLS

We can derive a formula for the number of resistant cancer cells, y, directly from the generating function (Equation 1). Considering the case $Mu \ll 1$, we have

$$G(\xi) \approx 1 - \frac{u}{1 - d/r} \sum_{x=1}^{M-1} (1 - g_x(\xi)). \tag{D1}$$

From Equation 6, we have

$$1 - g_x(\xi) = \frac{(\xi - 1)(1 - b/a)(M/x)^\alpha}{(\xi - 1)(M/x)^\alpha - (\xi - b/a)}$$

$$= \frac{1 - b/a}{1 - (b/a)y^\alpha}[1 + (L - 1)(\xi + L\xi^2 + L^2\xi^3 + L^3\xi^4 + \cdots)]. \tag{D2}$$

Here $y = x/M$ and $L = (1 - y^\alpha)/(1 - (b/a)y^\alpha)$. Hence we have

$$G(\xi) = 1 - \frac{Mu}{1 - d/r} \int_0^1 \frac{1 - b/a}{1 - (b/a)y^\alpha}[1 + (L - 1)(\xi + L\xi^2 + L^2\xi^3 + L^3\xi^4 + \cdots)]\,dy. \tag{D3}$$

The coefficient of ξ^k is

$$p_k = \frac{Mu}{1 - d/r} \int_0^1 \frac{(1 - b/a)^2 y^\alpha}{(1 - (b/a)y^\alpha)^2}\left(\frac{1 - y^\alpha}{1 - (b/a)y^\alpha}\right)^{k-1}dy \quad \text{for } k = 1, 2, 3, \ldots, M - 1. \tag{D4}$$

This is the unconditional probability that the number of resistant cancer cells is k. The mean number conditional to the existence of resistance is given by

$$\Pr[Y = k] \approx \frac{\int_0^1 \left[((1 - b/a)^2 y^\alpha)/(1 - (b/a)y^\alpha)^2\right]((1 - y^\alpha)/(1 - (b/a)y^\alpha))^{k-1}\,dy}{\int_0^1 (1 - b/a)/(1 - (b/a)y^\alpha)\,dy}$$

$$= \frac{(1 - b/a)^2}{\alpha F} \int_0^1 \frac{z^{1/\alpha}(1 - z)^{k-1}}{(1 - (b/a)z)^{k+1}}dz, \quad \text{where } K = 1, 2, 3, \ldots, M - 1.$$

This is Equation 16 in the text.

If there is no cell death, we have $d = b = 0$, and hence

$$\Pr[Y = k] = \frac{1}{\alpha} \int_0^1 z^{1/\alpha}(1 - z)^{k-1}\,dz$$

$$= \frac{(k - 1)!}{\alpha(k + 1/\alpha)(k - 1 + 1/\alpha) \ldots (1 + 1/\alpha)}, \quad \text{where } K = 1, 2, 3, \ldots, M - 1. \tag{D5}$$

If in addition the resistant cancer cells are neutral ($\alpha = 1$), we have

$$\Pr[Y = k] = \frac{1}{(k + 1)k} = \frac{1}{k} - \frac{1}{k + 1}, \quad \text{where } K = 1, 2, 3, \ldots, M - 1. \tag{D6}$$