Statistical Fluctuations in Autocatalytic Reactions

Max Delbrück

Statistical Fluctuations in Autocatalytic Reactions

MAX DELBRÜCK[1]

William G. Kerckhoff Laboratories of the Biological Sciences, California Institute of Technology, Pasadena, California

(Received August 29, 1939)

The differential equations describing the statistical fluctuations of a simple autocatalytic reaction mechanism are set up and solved completely. The fluctuations are found to approach a constant limiting value when the amount of reaction product is large compared to the amount which initiates the reaction. The relative fluctuations which arise from the reaction mechanism are then equal to the reciprocal of the square root of the initial number of particles. Conversely the fluctuation in the time required to attain a threshold amount of reaction product also approaches a limiting value for high thresholds. This limiting value is greater than one time unit if the reaction is initiated by one particle, and becomes equal to one divided by the square root of the number of initiating particles if this number is large.
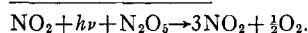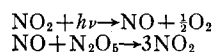
R ECENT work by Kunitz and Northrop, and by Herriott[2] on the proteolytic enzymes, trypsin and pepsin, has shown that these substances can be produced from precursors by an autocatalytic process, the rate of production of the enzyme being proportional both to the concentration of precursor and to the concentration of the enzyme itself. Such a type of reaction can function like an amplifier, inasmuch as a very small amount of enzyme is sufficient to initiate the reaction and to let the amount of enzyme grow exponentially to macroscopic size. In fact one molecule of pepsin should be sufficient under favorable conditions to convert in a few hours any weighable amount of pepsin-precursor. In experiments designed to test this one must be prepared to encounter very large statistical fluctuations in the amount of reaction taking place in a given time, or vice versa in the time required to effect a given amount of reaction. Since in all other cases of chemical kinetics the statistical fluctuations are negligibly small, these reactions offer a unique opportunity for the study of such fluctuations.[3] A closer theoretical study of their finer details seems therefore desirable as an aid for the design of experiments and to prepare the way for a discussion of the possible importance of such fluctuations for cell physiology. From such an aim we are however still far removed, because for none of the biological *S* curves, in which enthusiastic biological speculators have seen the manifestation of autocatalytic reactions has such a mechanism ever been established.

For the sake of simplicity of calculation it will be assumed that the amount of precursor throughout the course of the reaction is large compared to the amount of enzyme and that the change in concentration of precursor is negligible, i.e., we will only consider the initial stages of the reaction.

We further assume that the rate of the reaction is always strictly proportional to the concentration of enzyme. Since we want to consider the case of few molecules we have to explain a little further what we mean by "rate." Let us consider an experiment where we set up at time zero a very large number of samples containing each one molecule of enzyme and a large amount of precursor. We want to assume then that the fraction of samples which form their second

[1] International Fellow of the Rockefeller Foundation.
[2] For a summary see J. H. Northrop, *Crystalline Enzymes, The Chemistry of Pepsin, Trypsin and Bacteriophage* (Columbia University Press, New York, 1939).
[3] While we do not feel qualified to review adequately the chemical literature on autocatalysis it should be pointed out that none of the well-known autocatalytic reactions lend themselves to a study of these fluctuations. Those reactions where the catalyst is the H^+ ion are unsuitable, because the H^+ ion concentration cannot be made small enough. The photochemical decomposition of N_2O_5 into NO_2 and O_2 is autocatalyzed by the light-

absorbing NO_2

$$NO_2 + h\nu \rightarrow NO + \tfrac{1}{2}O_2$$
$$NO + N_2O_5 \rightarrow 3NO_2$$

$$\overline{NO_2 + h\nu + N_2O_5 \rightarrow 3NO_2 + \tfrac{1}{2}O_2.}$$

But here the spontaneous thermal decomposition of N_2O_5 is too fast even at 0°, so that this reaction initially will furnish more NO_2 than the autocatalytic reaction.

enzyme-molecule in a given time-element is always proportional to the fraction of samples which have not yet done this at the beginning of the time-element, and independent of the time which has elapsed since the samples were set up. This means that we rule out a reaction mechanism in which the reaction consists of separable steps each requiring the completion of its forerunner before it can occur, and each requiring a finite average time for its completion. In actual fact it is quite probable that most autocatalytic reactions will be separable into such steps, but our simplified model will still apply to such cases if one step of the reaction is much slower than all the others, for instance if the average time between collisions of an enzyme-molecule with any precursor-molecule is sufficiently long, and this can always be secured by working at low concentrations.

1. The Fundamental Equation

In ordinary reaction-kinetics we are only concerned with the average number of molecules of each reactant as function of time, since the deviations from these average values in individual cases are negligibly small. For our model we would then have

$$dn/dt = kn, \quad n = n_0 e^{kt}, \tag{1}$$

where n is the number of enzyme-molecules present at time t, and n_0 is the number of enzyme molecules present at time zero. k is the reaction rate constant, which will depend on the concentration of precursor, but will be constant during the course of the reaction according to our assumption. This equation will hold also for our case if we substitute for n the *average* number of molecules $\bar{n}$, averaged over a very large number of samples starting from the same initial conditions. We must therefore write

$$d\bar{n}/dt = k\bar{n}, \quad \bar{n} = \bar{n}_0 e^{kt}. \tag{2}$$

However we want to inquire beyond this information about the average number $\bar{n}$, into the statistical fluctuations of n as function of time and of n_0. We therefore do not deal with n as function of time but with the probability of having n particles at time t. This probability, i.e., the fraction of a very large number of similar samples, which contains n particles, we

will call $p_n(t)$. Instead of the one function $\bar{n}(t)$ we therefore will have to handle the infinite series of functions $p_n(t)$.

Let us find the differential equation determining p_n as function of time. The probability that of n particles any one will double in the small interval dt is equal to $nkdt$. The samples in which this happens will leave the class p_n and enter the class p_{n+1}. The probability that more than one particle will double in this interval is small of a higher order in dt and can therefore be neglected. The quantity p_n will change in the time interval dt for two reasons. Firstly it will increase by $(n-1)kp_{n-1}dt$ because so many of the samples containing $n-1$ particles will produce one additional particle and will thereby be promoted to the class p_n. Secondly it will decrease by nkp_ndt because so many of the samples containing n particles will advance to the class p_{n+1}.

We therefore have as our differential equation

$$dp_n/dt = (n-1)kp_{n-1} - nkp_n. \tag{3}$$

These equations, together with the initial conditions $p_n(0)$ will determine all the functions p_n for all times.

2. The Standard Deviations

Before we go on to establish and to discuss the general solutions of these equations we will consider what we can derive about the fluctuations without explicit knowledge of the p_n. It will be found that we can follow the change in the mean square deviation of the number of particles from their mean value without knowing the actual distribution of the samples over the classes p_n at any time.

The standard or root mean square deviation s is defined by

$$s^2 = \langle (n-\bar{n})^2 \rangle_{Av} = \langle n^2 \rangle_{Av} - \bar{n}^2.$$

Let us consider the change in time of the mean square $\langle n^2 \rangle_{Av}$.

We have, by definition,

$$\frac{d}{dt}\langle n^2 \rangle_{Av} = \frac{d}{dt} \sum_{n=0}^{n=\infty} n^2 p_n.$$

The right-hand side we are now going to differentiate term by term and to substitute for the

derivative of p_n its value from the fundamental Eq. (3)

$$= \sum n^2 [(n-1)kp_{n-1} - nkp_n].$$

Now, since we have to sum over all values of n, we can always combine the $(n+1)$ term of the first member with the n term of the second member, and obtain in this way

$$= \sum [(n+1)^2 n - n^3]kp_n$$
$$= \sum (2n^2 + n)kp_n$$
$$= 2k\langle n^2 \rangle_{\text{Av}} + k\bar{n}.$$

Now we know already $\bar{n}$ as a function of time, Eq. (2); therefore this is a differential-equation for $\langle n^2 \rangle_{\text{Av}}$ alone. Its solution is

$$\langle n^2 \rangle_{\text{Av}} = Ae^{2kt} - \bar{n}_0 e^{kt} = A(n/\bar{n}_0)^2 - n. \quad (4)$$

The mean square deviation is obtained from this by simply subtracting $\bar{n}^2$. The constant A will be determined by the standard deviation at time zero, which we will call s_0^2. Thus we obtain for the relative standard deviation at an arbitrary time

$$s^2/\bar{n}^2 = s_0^2/\bar{n}_0^2 + 1/\bar{n}_0 - 1/\bar{n}. \quad (5)$$

Here the last term is negligible compared to $1/n_0$ at any late time when $\bar{n} \gg n_0$. The relative standard deviation is then the sum of its initial value and the constant $1/n_0$. If the initiating amount is obtained by sampling from a large reservoir, the first term will also be $1/\bar{n}_0$ and the total relative standard deviation will be twice this value.

3. A COMPLETE SOLUTION OF THE FUNDAMENTAL EQUATIONS FOR THE CASE, IN WHICH THE INITIAL NUMBER OF PARTICLES IS KNOWN

So far we have derived information about the average number of particles and about the standard deviation from this average as functions of time. This would be sufficient for all practical purposes if we could be assured that the probability distribution was always normal. However, since we have already seen that in cases where the initial number of particles is very small the standard deviation is of the same order of magnitude as the mean, the distribution is not likely to be normal and we must try to obtain

the complete distribution. This is fortunately possible.

The Eqs. (3) are linear. The general solution consists therefore of the linear combinations of any linearly independent complete set of solutions. As such we can choose those for which we have at time zero a definite number, say n_0, particles, i.e.,

$$p_{n_0} = 1 \quad (6)$$

and $p_s = 0$ for all values of s unequal to n_0. This particular set of solutions, which conforms to the initial conditions just mentioned we will designate by $p_n^{(n_0)}$. We find

$$p_n^{(n_0)} = \frac{(n-1)!}{(n_0-1)!(n-n_0)!} \frac{n_0^{n_0}}{\bar{n}^{n_0}} \left(1 - \frac{n_0}{\bar{n}}\right)^{n-n_0}. \quad (7)$$

These are functions of the initial number of particles, n_0, of the time (through $\bar{n}$, which is a function of n_0 and of the time, see Eq. (2)), and of n, the index number of the function, designating the class of samples the frequency of which is given by the formula. It is easily verified that these functions satisfy both the initial conditions and the differential-equations. To prove that the initial conditions are satisfied, one simply has to put $\bar{n} = n_0$. To prove, that the differential equations are satisfied one has to substitute our solution in both sides of these equations. It is then found that both sides are equal to

$$kn_0 \cdot p_n^{(n_0)} \cdot (n - \bar{n})/(\bar{n} - n_0). \quad (8)$$

4. REACTIONS INITIATED BY ONE PARTICLE

We will use the solution (7) first to consider the case where we start with one particle at time zero, $n_0 = 1$. We then have

$$p_n^{(1)} = \frac{1}{\bar{n}} \left(1 - \frac{1}{\bar{n}}\right)^{n-1}. \quad (9)$$

We are mainly interested in the probability distribution at a late time, when $\bar{n}$ represents a chemically or enzymatically measurable amount, i.e., when $\bar{n}$ is large compared to one. We can then replace the second factor by an exponential function and obtain a very simple expression for our distribution, viz.

$$p_n^{(1)} = \frac{1}{\bar{n}} e^{-n/\bar{n}}. \quad (10)$$

We see then that the distribution is by no means normal. At all times the probability of finding one particle is greater than the probability of finding any other specified number of particles. In particular, the probability of finding the average number of particles is $1/e$ times the probability of finding one particle. As function of n the probability falls off with a "decay-constant" $1/\bar{n}$. It is easily verified, that the standard deviation is in this case equal to the mean in agreement with the general formula (5).*

5. REACTIONS INITIATED BY n_0 PARTICLES

Returning to the general solution (7) we will also here restrict our discussion to a phase of the reaction when the number of particles is already large compared to the initial number n_0.

We will first find out for which value of n the distribution has its maximum. To do this we observe that p_{n+1} and p_n stand in the simple relation

$$p_{n+1} = \frac{n}{n+1-n_0}(1-n_0/n) \cdot p_n. \qquad (11)$$

p_{n+1} arises from p_n by multiplication with a factor which is at first greater than one and continuously decreases. p_n will have its maximum where this factor is equal to one. This occurs for

$$n = \bar{n}(1-1/n_0). \qquad (12)$$

At this point the distribution has therefore its maximum. It is interesting to note that the maximum does not occur at the average value $\bar{n}$, but a definite distance towards smaller values of n away from it.

We now want to know what the distribution looks like in the neighborhood of this maximum. We introduce a new variable f to replace n. f shall measure the distance from the maximum

---

* In general experimental conditions will be such that also initially we will have a statistical distribution rather than having precisely one particle in each sample. We can however approximate the above condition by making the *average* initial number of particles, $\bar{r}$, small compared to unity. Then $e^{-\bar{r}}$ samples will have no particles, $\bar{r}e^{-\bar{r}}$ will have one particle and a much smaller fraction $(\bar{r}^2/2)e^{-\bar{r}}$ will have two particles. These will cause an alteration in the final distribution, which works out to be

$$f(n) = Cp_n^{(1)}\left[1+\frac{\bar{r}}{2}\frac{n-1}{\bar{n}}\right]. \qquad (10')$$

The change becomes large for $n/\bar{n} > 1/\bar{r}$.

on the n axis and its unit shall be this n_{max}. We define f by the equation

$$n = n_{max}(1+f)$$
$$= \bar{n}(1-1/n_0)(1+f). \qquad (13)$$

Substituting this value for n into our solution (7) we obtain (for n large compared to n_0)

$$p_n = C[(1+f)(1-n_0/\bar{n})^{f\bar{n}/n_0}]^{n_0-1}, \qquad (14)$$

where C is independent on f. Here we can replace

$$(1-n_0/\bar{n})^{f\bar{n}/n_0} \quad \text{by} \quad e^{-f}.$$

If $\log p_n$ is then developed into a power series of f it is found that the first-order term vanishes, the second-order term has the coefficient $-\frac{1}{2}(n_0-1)$ and the terms of higher order in t are small in the region where p_n is appreciable. We obtain therefore

$$p_n = C \cdot e^{-\frac{1}{2}(n_0-1)f^2}. \qquad (16)$$

This is indeed a normal distribution with a relative standard deviation equal to one divided by the square root of the initial number of particles, as given in (5). Under ordinary experimental conditions already the initiating sample n_0 would exhibit a standard deviation from n_0 by this amount. The square of the total final relative standard deviation of n would then be the sum of these two (5), i.e. it would be $2/\bar{n}_0$.

6. THE DISTRIBUTION OF THE TIME AT WHICH A GIVEN NUMBER OF PARTICLES IS ATTAINED

Up till now we have been concerned with the probability distribution of the number of particles at a given time. This corresponds to a type of experiment, where we set up at time zero a large number of samples with each containing the same number of initiating particles, and where we test at a certain later time the number of particles produced in the interval in each sample. We can, however, look at the problem of the fluctuations from a different standpoint. Suppose that a certain threshold-value n is required to produce a given effect. We must then expect great fluctuations in the time required by different samples starting with the same number of particles to attain this threshold-value. Let us call the fraction of samples which
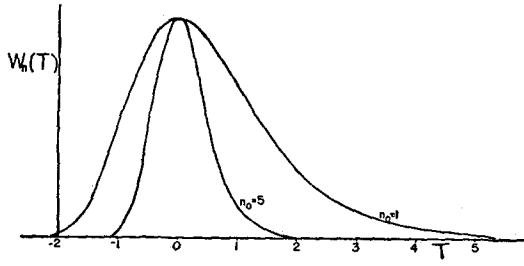
FIG. 1. The distribution in time of the arrival at a high threshold value of samples, starting from one or five particles.

attain this threshold-value in the time interval between t and $t+dt$ $w_n(t)dt$. This fraction is given by

$$w_n(t)dt = (n-1)kp_{n-1}dt, \qquad (17)$$

because this is the rate at which the class $n-1$ goes over into the class n, multiplied by the abundance of the class $n-1$. By virtue of Eq. (3) it follows from this that

$$-w_n(t) = \frac{d}{dt}(p_1 + p_2 + \cdots p_{n-1}). \qquad (18)$$

That means that the flux into the class n is equal to the net loss of all the lower classes.

Substituting our general solution (7) into (17) we obtain

$$w_n(t) = C\frac{1}{\bar{n}^{n_0}}\left(1 - \frac{n_0}{\bar{n}}\right)^{n-1-n_0}. \qquad (19)$$

For late times, when $\bar{n}$ is large compared to n_0, this can again be approximated by

$$w_n(t) = C\frac{1}{\bar{n}^{n_0}} \cdot e^{-(n/\bar{n})n_0}, \quad \bar{n} = n_0 \cdot e^{kt}. \qquad (20)$$

Here C stands for all the factors which are not dependent on the time, and $\bar{n}$ is given by (2) as a function of t and of n_0.

Let us now take a closer look at these functions $w_n(t)$, which give the rate at which samples starting with an equal number of particles reach their threshold-value of n particles. For n greater than n_0+1 they all start with zero at time zero (when $n=n_0$), then rise to a maximum and fall again to zero for large times. The maximum is easily found to occur at

$$t_{\text{max}} = \frac{1}{k}\log\frac{n-1}{n_0} \quad \text{when} \quad \bar{n} = n-1. \qquad (21)$$

We will introduce a new time-variable T, taking t_{max} as the origin and $1/k$ as the time unit. This time unit is the span in which the average number of particles increases by a factor e,

$$\bar{n} = (n-1)e^T. \qquad (22)$$

Substituting this in (20) we obtain

$$w_n(T) = C\exp\left[-n_0(T + e^{-T})\right]. \qquad (23)$$

This function we have plotted for $n_0 = 1$ and for $n_0 = 5$ in Fig. 1. It will be seen that the curve for $n_0 = 1$ is rather asymmetrical, it rises in two time units from 0.005 to its maximum and takes five time units to fall back to the same value. The spread in time naturally is smaller when we start with more than one particle. For large values of n_0 we obtain a very sharp maximum, we can then expand the second term in the exponent on the right-hand side of (23) and obtain a simple normal distribution curve

$$w_n(T) = C\exp\left(-\frac{n_0}{2}T^2\right). \qquad (24)$$

The standard deviation of this distribution is again equal to one divided by the square root of the initial number of particles:

$$\langle \Delta T^2 \rangle_{\text{Av}} = 1/n_0. \qquad (24a)$$