Proc. Nat. Acad. Sci. USA
Vol. 73, No. 2, pp. 415–419, February 1976
Biochemistry

Studies on nucleic acid reassociation kinetics: Empirical equations describing DNA reassociation*

(renaturation/nucleation inhibition/single strand ends)

ROY J. BRITTEN† AND ERIC H. DAVIDSON

Division of Biology, California Institute of Technology, Pasadena, Calif. 91125

Contributed by Roy J. Britten, November 26, 1975

ABSTRACT **The rate of appearance of duplex DNA renaturation, measured with single strand specific nuclease, deviates significantly from a second order reaction. Measurements reported in paper I of this series indicate an inhibition in the rate of reassociation of single strand tails on partially reassociated molecules by a factor of at least two. Equations are derived that describe the observed form of reassociation kinetics as measured with hydroxyapatite and with single strand specific nuclease. The free parameter that describes the extent of inhibition of nucleation with single strand tails in these equations has been evaluated by least squares methods and agrees with the experimentally measured value.**

The reassociation of DNA is now an important tool for the examination of the organization of the genome and for studying its evolution. The kinetics of the reassociation process are needed for the evaluation of the frequency of occurrence of repeated sequences and for the design of measurements comparing DNA sequences from different sources. However, this reaction is not fully understood, and the actual form of the kinetics has only been approximately evaluated (1, 2). This paper examines the effect of random shearing of DNA fragments on the kinetics of formation of strand pairs that contain duplex regions [assayed by hydroxyapatite (HAP)] and on the fraction of total nucleotides (NT) that are paired (assayed by single strand specific nuclease, or approximately by optical hyperchromicity). If all DNA fragments terminate at the same points in the sequence and have the same length and if there are no internal repetitions, then each effective nucleation leads to complete pairing of the fragments. Reassociation then takes the form of a second order reaction however it is assayed:

$$C/C_0 = (1 + KC_0t)^{-1} \quad [1]$$

Where C_0 is the DNA concentration (mol of NT/liter) and C is concentration in fragments without duplex regions. However, in most measurements the DNA has been sheared and each fragment terminates at a random place in the sequence. As a result, most nucleations lead to partial pairing of the fragments and single strand regions remain unpaired. Britten and Kohne (1) showed nevertheless that the kinetics of reassociation assayed by HAP follow Eq. 1 to a very close approximation.

In an earlier analysis (1, 3), a simple approximation was used to account for the reduction in yield of duplex in later collisions. It was assumed that effective nucleation rate is proportional to the square of the concentration of unpaired NT and that the yield of duplex per nucleation is reduced as the reaction proceeds in proportion to the fraction of unpaired NT remaining. The resulting approximate analysis yields Eq. 1 for the fraction of fragments without duplex regions (HAP assay) and yields the following equation for the fraction of NT unpaired:

$$S/C_0 = (1 + KC_0t)^{-n} \quad [2]$$

where S is the concentration of NT remaining single stranded and the other symbols are as in Eq. 1. K has the same numerical value as in Eq. 1, and the value of n resulting from this derivation was 0.5.

Morrow (4) and Smith *et al.* (5), in paper I of this series, measured the kinetics with which the fraction of NT remaining single stranded (S/C_0) changes in randomly sheared DNA, using the single strand specific nuclease S1. Their results fit closely the form of Eq. 2. Morrow's data yielded a best value of $n = 0.44$, while the best value for the data of Smith *et al.* was $n = 0.45$. It seems clear from the agreement of these results that n has been determined with some accuracy. Smith *et al.* also measured the kinetics with which the fraction of totally single stranded fragments (C/C_0) changes, by HAP assay, using portions of the same samples studied by nuclease. Their measurements showed that the best value of K in Eq. 1 is equal to the value of K in Eq. 2 for the S1 nuclease measurements.

Thus we have two simple equations that adequately express the kinetics of DNA reassociation and are suitable for the evaluation and interpretation of measurements. However these expressions provide little insight into the mechanisms responsible. In the current work we have taken into account the variation in length of the free single strands and of the single strands remaining on fragments that already contain duplex regions. These we term "particles", since, as the result of successive collisions, such partially duplexed renaturation products may grow to very large size. The variation in single strand length during the reaction was estimated using a computer program that models the reassociation reaction by a "Monte Carlo" method. These length changes, of course, retard the reassociation rate at later times. We have shown earlier that, probably due to some form of steric interference, the per NT rate of reaction of the particle single strands is inhibited. Here we show that this inhibition is required to explain quantitatively the observed forms of both HAP and S1 nuclease kinetics.

Comparison of observed and computer simulated reaction kinetics

A computer simulation of the reassociation reaction has been used to calculate the expected or ideal reaction kinetics. By summing up the fraction of simulated fragments that contain "duplex" regions, the ideal rate of appearance of HAP binding is calculated. A similar calculation in which the

Abbreviations: HAP, hydroxyapatite; NT, nucleotides; C_0t, mol of nucleotides per liter × sec.

* This is paper II in a series. Paper I is ref. 5.

† Also Staff Member, Carnegie Institution of Washington.

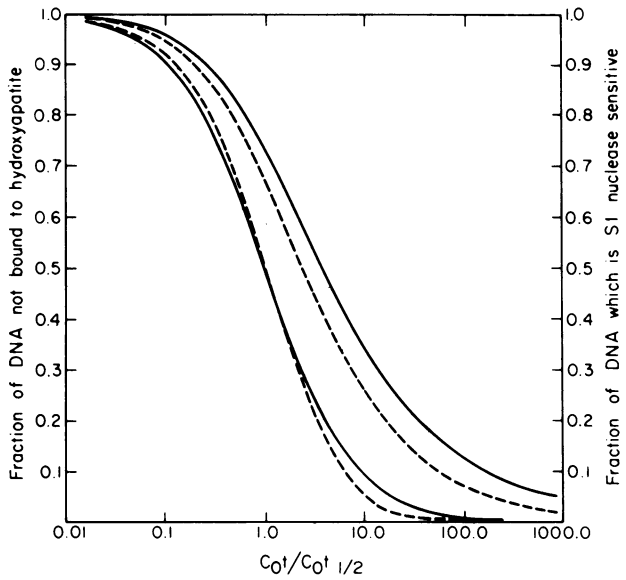

FIG. 1. Comparison of calculated and observed reassociation kinetics. The solid lines describing the reassociation of *Escherichia coli* DNA, as measured by S1 nuclease resistance and by HAP binding, are reproduced from Fig. 1 of Smith *et al.* (5). The calculated curves are shown by dashed lines. These are the HAP binding curve (inner curve) and the calculated S1 nuclease resistance curve (outer curve). The calculation was carried out by a computer program which simulates the reassociation reaction using a realistic fragment size distribution as described in the *text*. The computer calculations presented are the average of eight runs of the simulation program, all closely agreeing. The rate of the calculated curves has been chosen so that the calculated and measured hydroxyapatite curves cross at half-reaction. When Eq. 2 is used to fit the computer generated curves, the calculated HAP binding curve has n = 1.85 and the calculated S1 nuclease curve has n = 0.55.

"duplex" content is summed provides the ideal rate of appearance of S1 nuclease resistance. The calculated kinetics are compared to the observed kinetics in Fig. 1.

The reassociation reaction was modeled in the following way. Two arrays of 5000 elements were established representing the complementary strands of DNA. The array is divided into fragments using a random number generator according to a Gaussian distribution of lengths with the average and standard deviation specified for each run. For the results utilized in this paper, the average was 20 elements long and the standard deviation was 10. The distribution was truncated so that no lengths below 2 or above 40 were included. This distribution matched fairly closely the length distribution of sheared fragments determined by electron microscopy (6). The array of 5000 elements was assumed to repeat every 31 elements, in order to increase the effective rate of "nucleation" to a practical value. The ratio of the repeat length to the average fragment length was chosen so that the fragments terminated randomly within the repeat length. During a simulation run two elements are selected at random, and if either is paired, this is taken as an unproductive collision and the try is repeated. When two unpaired elements are selected and the two single stranded regions are "zippered" as far as their unpaired regions overlap; that is, each of the included elements is marked as paired. The number of elements paired is scored for "duplex formation," and the total length of the fragments is scored for "HAP binding" if they did not previously have paired regions. Where inhibition of particle single strand ends is to be included, a chosen fraction of the "nucleations" in such regions is not scored. Each try is considered a collision and the number of trys is the measure of the "C_0t." When the simulated rate is calculated so that the probability of "nucleation" is proportional to the square root of the "fragment" length rather than to the "fragment" length itself, there is little change in the result.

Fig. 1 shows that the measured kinetics of the reaction are of a different form from the calculated kinetics. The rate of the calculated curves has been chosen so that the calculated and measured hydroxyapatite curves cross at half-reaction. Given the same initial rate, the calculated reaction would proceed more rapidly than the observed second order form. Thus when Eq. 2 is used to fit the simulated HAP binding curve, the value obtained for n is 1.85 rather than 1.0, as in Eq. 1. Similarly, when Eq. 2 is used to fit the simulated S1 nuclease curve, the value obtained for n is 0.55, rather than the measured 0.45. We now consider the implications of the disagreement between the ideal and observed kinetic calculations.

As pointed out above, the rate of disappearance of free single strands depends not only on the reaction between free single strands, but also on the reaction between free single strands and the single stranded regions on particles. A correct description of this process would take the form

$$dC/dt = -KC^2 - KC(S - C) \qquad [3]$$

where the meaning of K and C are as above and $(S\text{-}C)$ is the concentration of single stranded NT on particles. As a first approximation the same rate constant, K, is used for both terms of the expression. Eq. 3 is of course a non-second order form, and it predicts that measurements of single stranded DNA concentration made with HAP should display faster than second order kinetics, given the same initial rates. The effect of fragment length variation is to blur this distinction somewhat by slowing down the reaction as it approaches termination. This is because longer fragments tend to react first so that the mean free fragment length, and hence the reaction rate, decreases. However a realistic fragment length distribution is already included in the simulated reactions illustrated in Fig. 1. The conclusion is that the reaction as measured with HAP has a rate which in the second half of the reaction is slower than the prediction even when the effect of fragment length distribution is included. This result strongly implies the existence of some form of inhibition that retards the later phase of the reaction. Such an inhibition has also been directly measured by Smith *et al.* (5).

Similar considerations apply to the comparison of calculated and observed S1 nuclease kinetics. Here again it is evident that some inhibition exists which has the effect of retarding the actual appearance of S1 nuclease resistance. The calculated uninhibited case begins at the expected rate, with the n of Eq. 2 at 0.55. This is equal to our best estimate of the extent of overlap (α) of two random sheared single strands as discussed in Smith *et al.* (5). The simulated reaction continues throughout in accordance with Eq. 2 with n = 0.55. The observed reaction begins at this rate since the measured value of α is 0.55. Soon, however, the overall rate declines, due to some form of inhibition, and the best fit value of n for the whole reaction is 0.45. We conclude that an inhibition affecting the latter part of the reassociation reaction is required in order to explain the experimental measurements of both HAP and S1 nuclease kinetics. Neither the apparent second order form described by Eq. 1 for HAP kinetics nor the use of Eq. 2 with n = 0.45 for S1 nuclease

measurements can be quantitatively understood without considering such an inhibition.

Equations describing reassociation in more detail

We now describe differential equations for the disappearance of free single strands (HAP kinetics) and the disappearance of single stranded NT (S1 nuclease kinetics). These equations include appropriate terms for the reaction of free single strands with particle single strands and for the reaction of particle single strands with each other. They permit a more detailed analysis of the nature of the apparent inhibition than has previously been possible.

The parameters used in the following expressions are the same as those defined earlier in this paper. In addition, there are terms for fragment length, particle inhibition, and nucleation rate constant. The symbols applied are: C_0, the total DNA NT concentration; L, the average free single strand fragment length at any time in the reaction; L_R, the average length of single strands on particles at any time in the reaction; K_n, the nucleation rate for each NT, not varying with fragment length (but varying according to the complexity); E, a particle inhibition factor expressed as a fraction of the uninhibited rate; α, the average overlap between two free single strands of average length L or between two particle single strands of average length L_R, expressed as fraction of single strand length; and α_R, the average overlap between free single strands of average length L and particle single strands of average length L_R expressed as a fraction of L_R. In constructing these equations we use the following assumptions:

(*i*) The observed rate of nucleation varies with fragment length as $K_n L^{-1/2}$ (2).

(*ii*) In reactions between roughly equal concentrations of longer and shorter fragments the observed rate of nucleation varies with the length of the shorter participant (7). While this assumption may not be correct, it is in accord with the best current measurements on which we rely in the following. Thus the observed nucleation rate in collisions between free single strands and particle single strands is considered to vary as $K_n L_R^{-1/2}$ since generally $L_R < L$.

(*iii*) The yield per free single strand in a successful collision between free single strands is L NT withdrawn from C, and is αL NT withdrawn from S. Similarly, in nucleations between a free single strand and a particle single strand the amount of NT removed from S per strand is $\alpha_R L_R$. The yield for a particle-particle collision is αL_R, and so forth.

For the disappearance of free single strands (HAP kinetics) we write

$$\frac{dC}{dt} = -\frac{K_n}{L^{1/2}}C^2L - \frac{EK_n}{L_R^{1/2}}C(S - C)L \qquad [4]$$

This is equivalent to Eq. 3, except that the effect of length on rate and yield are here explicit, and a factor is provided for particle inhibition. In contrast to K_n of Eq. 4, the Ks of Eqs. 1, 2, and 3 all include the effects of yield and fragment length. In Eq. 4 the rate of disappearance of free single strands is taken to be proportional to $L^{1/2}$ (2).

Similarly, for the disappearance of free single strands (S1 nuclease kinetics):

$$\frac{dS}{dt} = -\frac{K_n}{L^{1/2}}C^2\alpha L - \frac{EK_n}{L_R^{1/2}}C(S - C)\alpha_R L_R - \frac{EK_n}{L_R^{1/2}}(S - C)^2\alpha L_R \qquad [5]$$

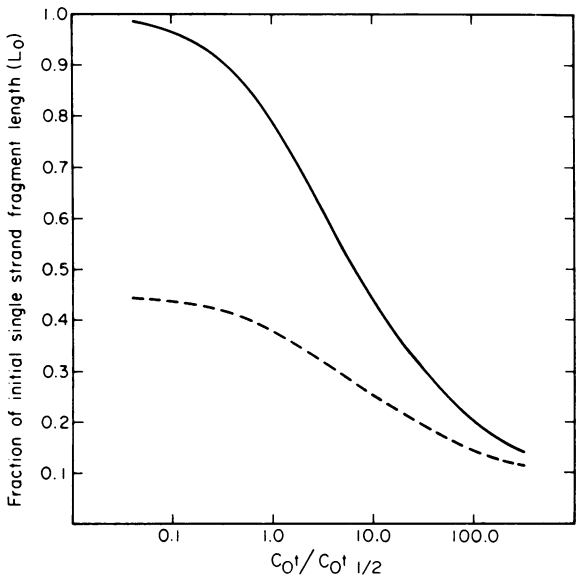

FIG. 2. Calculated change in length of free single strands (L/L_0) and single strand regions in particles (L/L_R) with C_0t. The ordinate is expressed as the single strand lengths relative to the starting fragment length, L_0. The calculations were carried out with the computer simulation of the reassociation reaction described in the *text*. The abscissa is expressed in terms of C_0t normalized to a simulated HAP binding rate of 1.0 liter mol^{-1} sec^{-1}. The functions used to generate the lines plotted are Eq. 6 for L_R/L_0 (dashed line) and Eq. 7 for L/L_0 (solid line), least squares fitted to the results of the computer simulation.

Eqs. 4 and 5 state the expected strand length effects on nucleation rate, and the expected per collision yield for each portion of the reaction. We assume that the value of E is the same for particle–particle as for particle-free reactions and that E is independent of L. The integration of these differential equations would provide solutions suitable for the quantitative examination of the rate of duplex formation and HAP binding. In order to integrate these equations, we must evaluate L and L_R as a function of t or C_0t. We follow an approximation method which, though it may lack elegance, is sufficiently accurate for the study of DNA reassociation kinetics. On the right-hand side of the differential equations we express C and S approximately in terms of C_0t using Eqs. 1 and 2. In the next section we describe the evaluation of the variation of L and L_R with C_0t and the method for the solution of Eqs. 4 and 5.

Approximate solution of the differential equations

Fig. 2 shows a calculation of particle single strand length (L_R) as a function of C_0t obtained with the "Monte Carlo" reassociation program and the distribution of fragment lengths described above. The graph shows the average length of single strand regions in particles (L_R) as a fraction of L_0, the starting single strand length. Thus, in Fig. 2 L_R/L_0 is plotted against the "C_0t" of the simulated reaction.

The curve shown in Fig. 2 can be fit reasonably well by an expression of the form

$$(L_R/L_0) = (1 - \alpha)(1 + KC_0t)^{-y} \qquad [6]$$

where y is about 0.24.

The value of L, the free single strand length, also changes as the reaction proceeds, due to the fact that longer fragments tend to react more rapidly, always leaving a distribution of free single strands of shorter mean length in the yet

unreacted class. An analysis similar to that in Fig. 2 shows that a reasonable form to describe the change in L as a function of C_0t is

$$(L/L_0) = (1 + KC_0t)^{-y} \quad [7]$$

where y is evaluated at about 0.34 for the fragment length distribution we have studied.

The observed disappearance of free single strands is fit almost perfectly by the simple second order expression, Eq. 1. Except for the first few percent of the reaction, the observed S1 nuclease kinetics are described adequately by Eq. 2. In making these substitutions we insert approximate solutions for C and S. When integrated we obtain more accurate functions for C and S which furthermore include explicit length corrections. Thus all the variables in Eqs. 4 and 5 can be related to C_0t by expressions of the form $(1 + KC_0t)^{-y}$. For C, $y = 1.0$ (Eq. 1); for S, $y = 0.45$ (Eq. 2); for L, $y = 0.34$ (Eq. 7); and for L_R, $y = 0.24$ (Eq. 6). The only remaining problem is the evaluation of α_R. As will be described elsewhere (Britten and Davidson, unpublished), α_R/α can be shown analytically to be equal to $[1 + (\alpha/2)]$ for all cases where L is much longer than L_R. Using these approximations, Eqs. 4 and 5 may now be integrated either algebraically or by numerical procedures. When integrated, Eqs. 4 and 5 yield the following expressions, where $m = y/2$ in Eq. 6 or about 0.12, and $x = y/2$ in Eq. 7 or about 0.17.
For HAP reassociation kinetics

$$\frac{C}{C_0} = \exp\left[\frac{V^{-x} - 1}{x} - \frac{E[V^{(1+m-n-2x)} - 1]}{(1 + m - n - 2x)(1 - \alpha)^{1/2}} + \frac{E[V^{(m-2x)} - 1]}{(m - 2x)(1 - \alpha)^{1/2}}\right] \quad [8]$$

where $V = 1 + KC_0t$. Here, as earlier, K is the same as in Eq. 1, i.e., the observed second order rate constant for the HAP reaction. For S1 nuclease reassociation kinetics

$$\frac{S}{C_0} = \exp\left[\frac{(V^{(n-x-1)} - 1)}{(n - x - 1)} + \frac{E\left(\frac{\alpha}{2} - 1\right)(1 - \alpha)^{1/2}(1 - V^{-m})}{m} - \frac{E(1 - \alpha)^{1/2}(V^{1-n-m} - 1)}{(1 - n - m)} + \frac{E\frac{\alpha}{2}(1 - \alpha)^{1/2}(V^{(n-m-1)} - 1)}{(n - m - 1)}\right] \quad [9]$$

We note that all of the constants in Eqs. 8 and 9 except E, the particle inhibition, have been evaluated from other evidence or calculations and are not free.

Evaluation of the particle inhibition, E

Eqs. 8 and 9 were evaluated at a series of C_0t's to determine the best value of E. As a check, Eqs. 4 and 5 were numerically integrated as well, with almost identical results. In Fig. 3 the data of Fig. 1 of Smith *et al.* (5) are replotted with the best least squares fits of Eq. 9. The form of Eq. 9 can be seen to fit the data excellently. The value of α was taken as 0.55, and K is normalized to 1.0 for ease of calculation. The best value of E is then 0.6. That is, slightly less than a 2-fold particle inhibition factor is required. The main significant feature of the calculation illustrated in Fig. 3 is that it shows that the "length effect" built into Eqs. 8 and 9 combined with the "particle inhibition" suffices to explain the observed S1 nuclease kinetics. It is possible that the length effect is slightly different from that we have calculated. New evidence shows that the dimensions of the longer strand may have to be taken into account in considering reactions between longer and shorter strands. As noted above we have followed Wetmur (7) in calculating the nucleation rate as a function of the shorter strand length. The net effect would be small, however, resulting in a minor increase in the particle inhibition. That is, there would be a slight reduction in the value of E if the assumption that rate depends solely on the length of the smaller reacting participant is wrong. We find that a slightly larger inhibition is also needed to explain the observed HAP kinetics, with a best value of E around 0.2, but less discrimination is possible than with the S1 nuclease kinetics. Furthermore the variation in root mean square error in fitting the calculated curves to the data with change in E is gradual. The most conservative conclusion is that an inhibition of 1.8- to 4-fold (i.e., E of 0.25–0.6) is required in order to provide a quantitative interpretation of the observed kinetics.

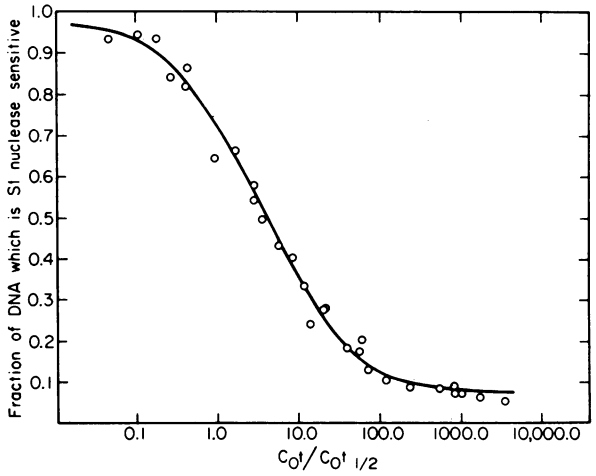

FIG. 3. S1 nuclease reassociation data fit with Eq. 9. Data are transcribed directly from Fig. 1 of Smith *et al.* (5). The solid line is calculated with Eq. 9. By least squares analysis the best fit value of E was 0.6. Other parameters were $K = 1.0$, $n = 0.45$, $m = 0.12$, and $x = 0.17$, as described in the *text*.

We now summarize the evidence for the existence of the particle inhibition factor. If the value of E is set at 1.0 (i.e., no inhibition) a less satisfactory fit of Eq. 9 to the data is obtained, and the root mean square error rises to about 1.7 times its value when $E = 0.6$. In addition experiments of Smith *et al.* (5) indicate a particle inhibition factor in the range 0.5–0.6, in excellent agreement with the quantitative treatment just described. However, both the numerical interpretation of these experiments and the calculation shown in Fig. 3 rely in part on the length reduction function illustrated in Fig. 2 and formalized in Eq. 6. If for some reason length reduction for particle single strands follows a much different course than the computer simulation predicts, the value of the particle inhibition factor would change. However we see no reason why the length reduction calculation should not be realistic. Our belief is that the particle inhibition probably accounts for about half of the total observed retardation. That is, if a calculation similar to those resulting in Eqs. 8 and 9 and Fig. 3 is carried out without any length

reduction terms, the value found for E is about 0.25, a little less than one-half that observed when the expected length reduction is taken into account. The general conclusion is that whatever its exact causes, a progressive reduction of the particle-single strand nucleation rate occurs. This rate reduction is the underlying reason that kinetics follow the familiar forms described by Eqs. **1** and **2.**

We thank Drs. Norman Davidson, Michael Smith, and Lynn Klotz for their critical reviews of a draft of this paper. This work was supported by NIH Grants HD-05753 and GM-20927 and by NSF Grant GB-33441.

1. Britten, R. J. & Kohne, D. E. (1967) *Carnegie Inst. Wash. Year Book* **65,** 78–106.
2. Wetmur, J. G. & Davidson, N. (1968) *J. Mol. Biol.* **31,** 349–370.
3. Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974) in *Methods in Enzymology,* eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 29, part E, pp. 363–418.
4. Morrow, J. (1974) Ph.D. Thesis, Stanford University.
5. Smith, M. J., Britten, R. J. & Davidson, E. H. (1975) *Proc. Nat. Acad. Sci. USA.* **72,** 4805–4809.
6. Chamberlin, M. E., Britten, R. J. & Davidson, E. H. (1975) *J. Mol. Biol.* **96,** 317–333.
7. Wetmur, J. G. (1971) *Biopolymers* **10,** 601–613.