

Appendix: Analysis of Nucleic Acid Association Reactions

Kinetic analysis of nucleic acid association by base pairing is a valuable technique for determining the complexity and relatedness of sets of nucleotide sequences. Various applications of this technique provide information on the genetic content, sequence organization, and expression of eucaryotic genomes. Similar methods can be used to study DNA-DNA, DNA-RNA, and RNA-RNA association reactions. This appendix considers the basic theory and some applications of nucleic acid association kinetics.

Concepts

A.1 Separated complementary strands of DNA can reassociate to form base-paired duplex structures.

A. The rate at which two complementary sequences reassociate depends on four parameters:

1. The concentration of cations, which decrease the intermolecular repulsion of negatively charged DNA strands
2. The incubation temperature, which is optimal for reassociation at about 25°C below the melting temperature (T_m) of the original duplex DNA (Concept 17.3)
3. DNA concentration, which determines the frequency of intermolecular collisions
4. The size of the DNA fragments

B. Reassociation studies generally are carried out as follows:

1. Purified DNA is sheared into small fragments by passage through a small orifice at high pressure, by high-speed blending, or by sonication. Uniform fragment sizes of 3×10^2 to 10^5 nucleotide pairs can be produced by controlled shearing.
2. The fragments are denatured by heating briefly to 100°C and then cooling rapidly. This procedure separates complementary strands without breaking a significant number of internucleotide bonds.
3. The dissociated fragments are incubated in a standard phosphate buffer (generally 0.18N monovalent cation) at a DNA concentration and temperature that will give an appropriate rate of reassociation (annealing conditions; Concept 17.3).
4. The extent of reassociation as a function of time is measured by one of the three following techniques, which distinguish between double- and single-stranded DNA:
 - a. Nuclease digestion. Using a nuclease that catalyzes hydrolysis of single-stranded but not double-stranded DNA (such as S1 nuclease from the mold

Aspergillus), samples can be digested to completion and the amount of nuclease-resistant material determined.

b. **Hypochromicity.** Because double-stranded DNA absorbs less ultraviolet light than single-stranded DNA at the same concentration, the degree of reassociation is directly proportional to the decrease in absorbance of 260 nm light (hypochromicity) as the reaction proceeds.

c. **Hydroxyapatite chromatography.** Because double-stranded DNA adsorbs to hydroxyapatite under conditions where single-stranded DNA does not, reassociation can be followed by passing samples through a column and determining the fraction of the DNA that adsorbs. Hydroxyapatite chromatography is useful also as a preparative technique for separating rapidly reassociating from slowly reassociating fragments.

A.2 Reassociation follows bimolecular reaction kinetics.

DNA reassociation is a bimolecular second-order reaction. The rate of disappearance of single strands is given by the expression

$$-\frac{dC}{dt} = kC^2 \quad (\text{A.1})$$

in which C is the concentration of single-stranded DNA in moles of nucleotide per liter, t is time in seconds, and k (liters per mole second) is a second-order rate constant. The value of k depends on cation concentration, temperature, fragment size, and the sequence complexity of the DNA population (Concept A.3). Equation A.1 can be rearranged to

$$-\frac{dC}{C^2} = k dt$$

and integrated from initial conditions $t = 0$ and $C = C_0$ to yield

$$\frac{1}{C} - \frac{1}{C_0} = kt$$

or

$$\frac{C}{C_0} = \frac{1}{1 + kC_0t} \quad (\text{A.2})$$

Because at $t = 0$ all the DNA is single-stranded, C_0 is equal to the total DNA concentration.

Equation A.2 shows that the fraction of single-stranded DNA remaining in a reassociation reaction (C/C_0) is a function of C_0t , the product of the initial concentration and the elapsed time. Therefore, the time course of a reassociation reaction is plotted conveniently as shown in Figure A.1, to give a so-called "Cot" curve.

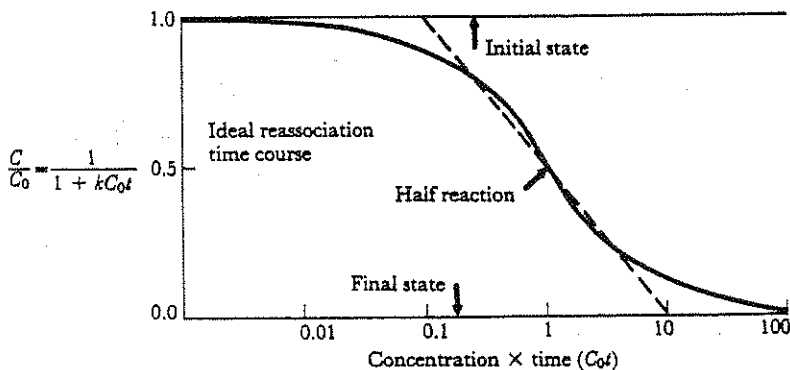


Figure A.1
 C_0t curve of an ideal second-order reassociation reaction. Note that 80% of the reaction occurs over a two-log interval of C_0t . [Redrawn from R. J. Britten and D. Kohne, *Science*, 161, 529 (1968).]

$C_0t_{1/2}$ is defined as the value of C_0t at which the reaction has proceeded to half-completion ($C/C_0 = \frac{1}{2}$). Substitution into equation A.2 shows that $C_0t_{1/2}$, which can be determined experimentally as shown in Figure A.1, is the reciprocal of the second-order rate constant:

$$C_0t_{1/2} = \frac{1}{k}$$

A.3 $C_0t_{1/2}$ is proportional to the sequence complexity of a DNA preparation.

The complexity (X)* of a sheared DNA preparation is the length in nucleotide pairs of the longest nonrepeating sequence that could be produced by splicing together fragments in the population. For example, a preparation of repeating dAT copolymer (. . . ATATAT. . .) has a complexity of 2; a preparation of a repeating tetramer of the form (ATGC)_n has a complexity of 4; and a preparation made from identical but internally nonrepeating DNA molecules 10⁵ nucleotide pairs in length has a complexity of 10⁵.

$C_0t_{1/2}$ is related directly to complexity. Consider the reassociation of two populations of denatured DNA fragments, one derived from identical nonrepeating molecules 10⁵ nucleotide pairs in length such as the DNA of a bacteriophage ($X = 10^5$), and the other derived from identical nonrepeating molecules 5 × 10⁶ nucleotide pairs in length such as the DNA of a bacterium ($X = 5 \times 10^6$). If the fragment size is 500 nucleotides, then in the first population the fragments representing any particular sequence and its complement make up 0.005% of the total DNA. In the second population, each 500-nucleotide sequence represents only 0.0001% of the total DNA. Therefore, if both populations are adjusted to the same *total DNA concentration*, the concentration of each 500-nucleotide sequence will be 50 times greater in the first population than in the second. Therefore $t_{1/2}$ for the first population will be $\frac{1}{50}$ of $t_{1/2}$ for the second. Or, if the DNA concentrations are unequal, $C_0t_{1/2}$ for the first population will be $\frac{1}{50}$ of $C_0t_{1/2}$ for the second.

The proportionality of $C_0t_{1/2}$ to complexity means that

$$X = KC_0t_{1/2} \quad (\text{A.3})$$

in which the proportionality constant K will depend on reaction conditions (cation concentration, fragment size, and so on). Under the standard conditions generally employed (0.18*N* cation concentration, 400-nucleotide fragment size), $K \approx 5 \times 10^5$ (liters × nucleotide pairs)/(moles of nucleotides × seconds), or

$$X = (5 \times 10^5)(C_0t_{1/2}) \quad (\text{A.4})$$

nucleotide pairs.

For organisms whose DNA contains no repeated sequences, X is simply the genome size (N) in nucleotide pairs, and N can be determined directly from measurements of $C_0t_{1/2}$ using equation A.3 or A.4. All procaryotic genomes examined so far fall into this category. The reassociation of sheared DNA from these organisms goes to completion over about a two-log interval of C_0t (see Figure A.1), indicating that the concentration of all sequences in the population is the same.

A.4 C_0t curves reveal the presence of repeated nucleotide sequences in eucaryotic genomes.

Most eucaryotic DNA preparations reassociate over a C_0t range that is much greater than two logs, indicating that different sequences in the population are present at

*In the literature complexity often is denoted by C . In this treatment X is used, to avoid confusion with the symbol C for DNA concentration.

different frequencies, and therefore reassociate at different rates. Consider a hypothetical example of an organism whose genome consists of 50% repeated sequences of total complexity 10^5 nucleotide pairs, each repeated 10^4 times per genome (component a), and 50% nonrepeated (unique) sequences of total complexity 10^9 nucleotide pairs (component b). N , the haploid genome size for this organism, is 2×10^9 . If the DNA fragments representing these two components could be separated and reassociated independently, they would give the C_0t curves shown by dashed lines in Figure A.2, with $C_0t_{1/2}$ values that differ by a factor of 10^4 . When reassociated together, the two components behave *independently* to give the biphasic C_0t curve shown by the solid line in Figure A.2. The following information can be obtained from such a curve:

1. **Genome fraction of each component (f).** The fraction of the genome represented by each component can be determined by extrapolation from the end point of each component of the reassociation curve to the ordinate. In the example, $f_a = f_b = 0.5$.

2. **$C_0t_{1/2(\text{pure})}$ for each component.** The $C_0t_{1/2}$ value for each component in the mixture can be determined by extrapolating from the midpoint of each component of the reassociation curve to the abscissa. In the example, $C_0t_{1/2(\text{mixture})}$ for a = 2×10^{-1} mol sec/L and $C_0t_{1/2(\text{mixture})}$ for b = 2×10^3 mol sec/L. Note that these values are different from those determined from the C_0t curves of each component alone (Figure A.2, dashed lines). This discrepancy arises because for the mixture C_0 is taken to be the *total DNA concentration*, whereas for an isolated component C_0 is taken to be the *concentration of the component*. To determine the complexity of a component, it is necessary to know the $C_0t_{1/2}$ value that the component would exhibit alone ($C_0t_{1/2(\text{pure})}$). This value can be obtained by multiplying the observed $C_0t_{1/2(\text{mixture})}$ by the fractional contribution of the component to the genome:

$$C_0t_{1/2(\text{pure})} = fC_0t_{1/2(\text{mixture})} \quad (\text{A.5})$$

In the example, the $C_0t_{1/2(\text{pure})}$ values for components a and b are 10^{-1} and 10^3 mol sec/L, respectively.

3. **Repetition number (R).** The *relative* frequencies of sequences in the various components of a mixture are inversely proportional to their $C_0t_{1/2(\text{mixture})}$ values. If the most slowly reassociating class is known to be unique ($R = 1$), or if the genome size, N , is known, then the absolute frequency or repetition number of the sequences in a component can be calculated. For any component, i , R_i is related to f_i , N , and X_i by the expression

$$R_i = \frac{f_i N}{X_i} \quad (\text{A.6})$$

This relationship also can be used to calculate the complexity X_i of a component if R_i , f_i , and N are known.

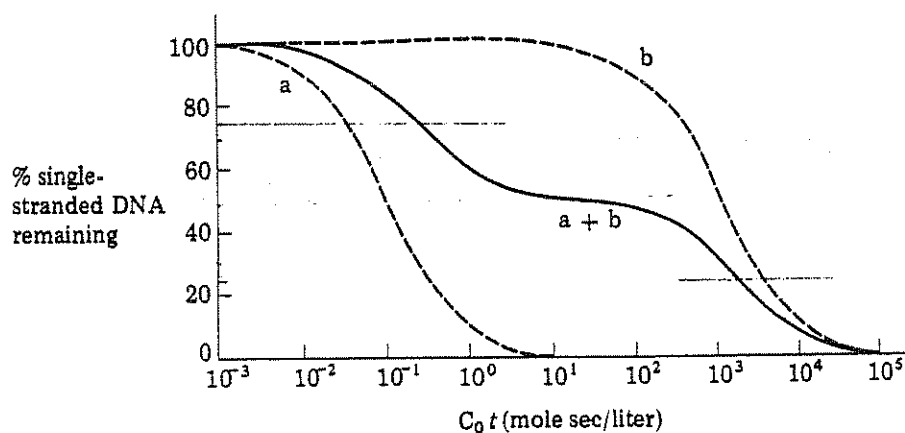


Figure A.2
 C_0t curves for whole DNA (solid line) and isolated frequency components (dashed lines) from a hypothetical eucaryotic organism.

A.5 Simple C_0t curves provide no information on arrangement of sequences or degree of complementary matching.

- A.** The analysis described here provides no information on the lengths of unique and repeated sequences as functional units in the genome, or on the organization of these units. The question of organization and an experimental approach to it are considered in Problem A.10.
- B.** Complications in the analysis of C_0t curves arise when the members of a family of repeated DNA sequences are similar but not identical. The extent of mismatching in a component isolated after reassociation (using hydroxyapatite chromatography) can be determined from its absorbance-temperature profile in a thermal denaturation experiment (Concept 17.3). DNA duplexes that contain mismatches have a lower thermal stability than perfectly complementary duplexes of the same nucleotide composition. Generally, a drop of 1°C in the melting temperature (T_m) of a DNA duplex indicates about 1% mismatching. Some examples of mismatching and the complications that arise from it are considered in Problems A.6 through A.8.

References

- R. J. Britten, D. E. Graham, and B. R. Neufeld, "Analysis of Repeating DNA Sequences by Reassociation," in L. Grossman and K. Moldave (Eds.), *Methods in Enzymology*, Vol. 29, Academic Press, New York, 1974, p. 363.
- R. J. Britten and D. Kohne, "Repeated Sequences in DNA," *Science*, 161, 529 (1968).
- G. A. Galau, R. J. Britten, and E. H. Davidson, "Measurement of the Sequence Complexity of Polysomal Messenger RNA in Sea Urchin Embryos," *Cell*, 2, 9 (1974).
- G. A. Galau, R. J. Britten, and E. H. Davidson, "Studies on Nucleic Acid Reassociation Kinetics: Rate of Hybridization of Excess RNA with DNA, Compared to the Rate of DNA Renaturation," *Proceedings of the National Academy of Sciences (U.S.)*, 74, 1020 (1977).
- G. A. Galau, M. J. Smith, R. J. Britten, and E. H. Davidson, "Studies on Nucleic Acid Reassociation Kinetics: Retarded Rate of Hybridization of RNA with Excess DNA," *Proceedings of the National Academy of Sciences (U.S.)*, 74, 2306 (1977).
- J. Wetmur and N. Davidson, "Kinetics of Renaturation of DNA," *Journal of Molecular Biology*, 31, 349 (1968).

Problems

- A.1 Answer the following questions with reference to Figure A.7.
- How many of these DNA preparations contain more than one frequency class of sequences? Explain your answer.
 - If the genome size of *E. coli* is taken to be 4.5×10^6 nucleotide pairs, what is the genome size of T4?
 - What is the complexity of mouse satellite DNA?
 - Mouse satellite DNA represents 10% of the mouse genome. What is the repetition number for mouse satellite sequences, given that the haploid genome size is 3.2×10^9 nucleotide pairs?
 - The calf genome is the same size as the mouse genome. What fraction of the calf genome is composed of unique sequences?
 - The reassociation curve for *E. coli* DNA, shown in Figure A.7, shows a $C_0t_{1/2}$ value of 8 mol sec/L. How will this value change if *E. coli* DNA is allowed to reassociate in the presence of a 1000-fold excess of calf thymus DNA?
- A.2 The reassociation curves for calf thymus DNA and an internal standard, *E. coli* DNA, are shown in Figure A.8.

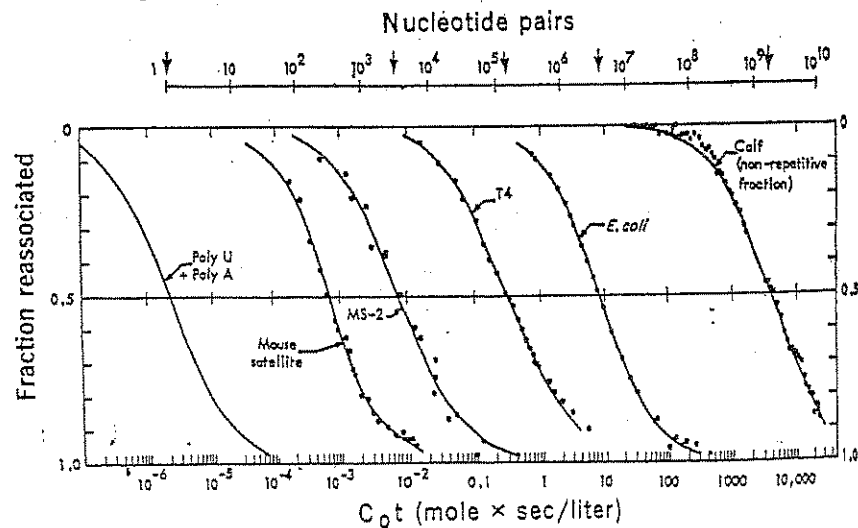
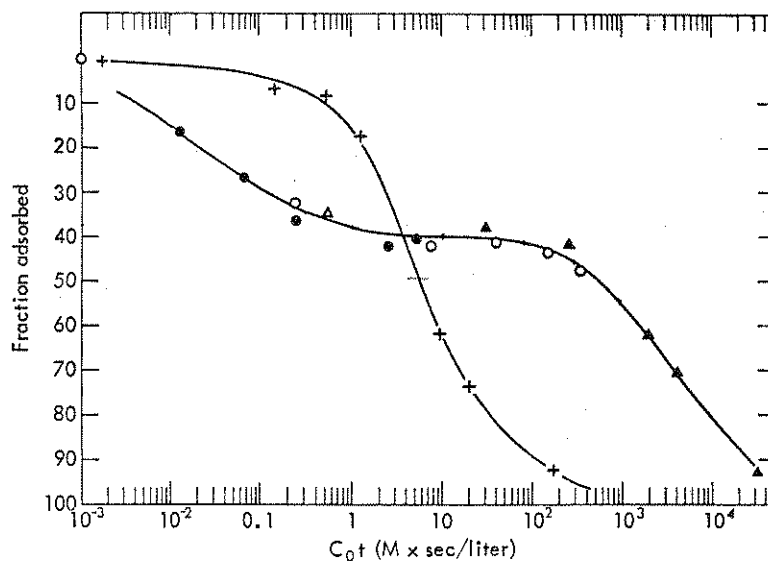


Figure A.7
Reassociation of nucleic acids, sheared to 500-nucleotide fragments, from various sources (Problem A.1). [From R. J. Britten and D. Kohne, *Science*, 161, 529 (1968).]

APPENDIX
Analysis of Nucleic
Acid Association
Reactions

Figure A.8
Reassociation curves for calf thymus DNA (\circ , Δ , \bullet , and \blacktriangle) and *E. coli* DNA ($+$) (Problem A.2). [From R. J. Britten and D. Kohne, *Science*, 161, 529 (1968).]



- How many frequency classes of DNA are present in the cow?
 - What fraction of the DNA does each class represent?
 - What is the $C_0t_{1/2(\text{mixture})}$ of each class?
 - How many times are the most rapidly reassociating sequences repeated with respect to the most slowly reassociating sequences (S fraction)?
 - Given that the *E. coli* genome size is 4.5×10^6 nucleotide pairs, what is the complexity of the S fraction?
- A.3 A small five-legged creature called a cloewg, which resembles the terrestrial cow, is brought back from an early morning expedition to Disneyland. The reassociation curve of its DNA is shown in Figure A.9.
- How many frequency classes of sequences are present in this DNA?
 - What fraction of the total genome does each class represent?
 - Explain how you might isolate the fastest (F) and the slowest (S) classes of DNA.
 - If you were to rerun C_0t curves on these isolated classes, what $C_0t_{1/2(\text{pure})}$ values would you obtain?
 - What is the complexity (X) of the S class of DNA, assuming that the reassociation was carried out under standard conditions (as defined in Concept A.3)?
 - Given that the haploid genome size of the cloewg is 2.5×10^9 base pairs, how many copies of each sequence are present in the S class (i.e., what is the repetition number R)?
 - What is the repetition number for the F class?

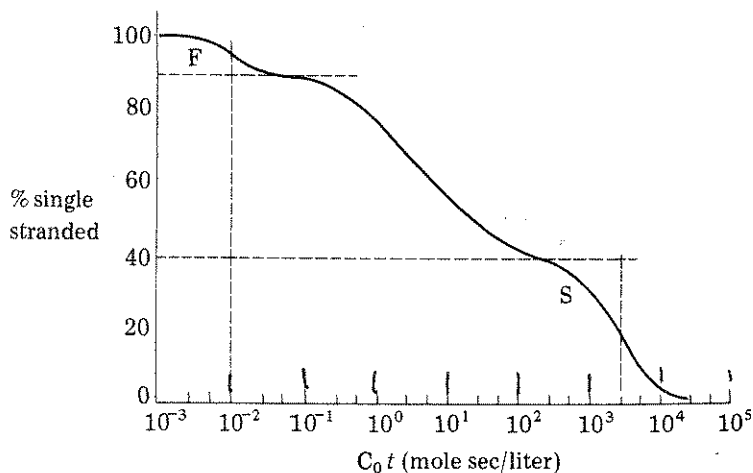
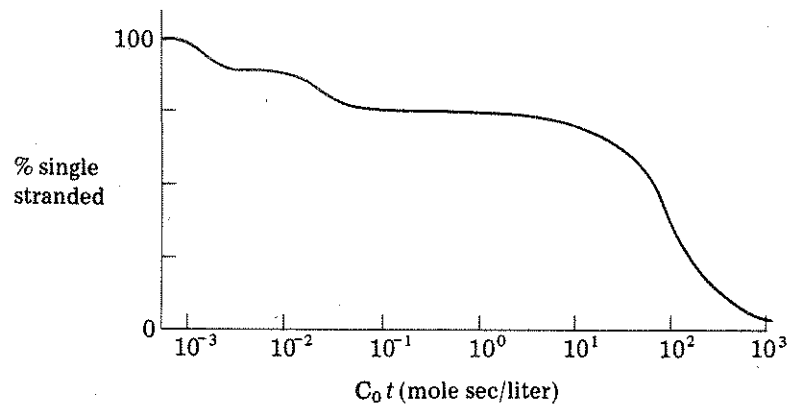
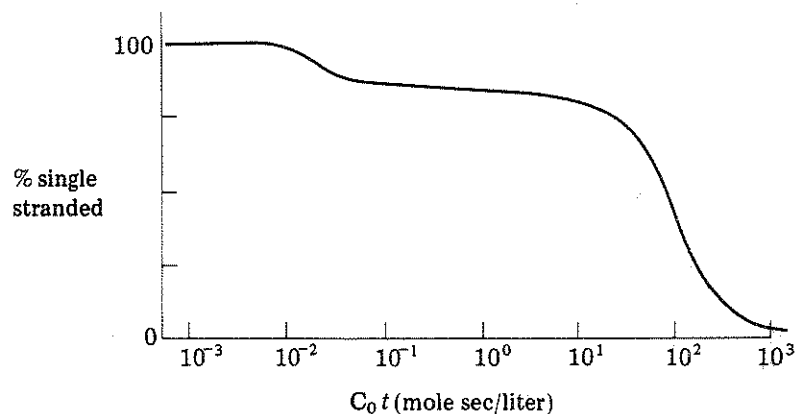


Figure A.9
The reassociation curve of DNA fragments from the cloewg (Problem A.3) (F = fastest, S = slowest).

- h. What is the complexity of the F class?
- i. How do the frequency classes of the Disneyland cloewg differ from those of the ordinary cow?
- j. From the data presented, what can you determine about the organization of these sequences?
- A.4 **Polytene chromosomes** are found in certain tissues of several organisms. These chromosomes arise from the chromosomes of diploid nuclei by successive duplications that differ from those of the normal mitotic cycle. Following replication, daughter chromosomes do not segregate but remain paired with each other. The nuclear membrane remains intact following replication, and there is no cell division. Nine to ten cycles of such replication result in a nucleus containing giant polytene chromosomes, each composed of 1000–2000 identical extended DNA molecules in a multistranded cable. The duplication process for polytenization appears to be complex, in that different portions of the genome are duplicated to different extents. The presence of polytene chromosomes is referred to as **polyteny**.
- a. A C_0t curve for DNA from a fly embryo is shown in Figure A.10a. Would you expect a C_0t curve of polytene chromosomal DNA from this insect to be identical? Why?
- b. Suppose that you obtain the C_0t curve shown in Figure A.10b for the polytene DNA. How does this C_0t curve differ from the preceding one, and how would you interpret this difference?
- A.5 The most slowly reassociating fraction of the DNA from *Amphiuma*, an amphibian, gives a $C_0t_{1/2(\text{mixture})}$ value of 1.6×10^5 mol sec/L. Assuming that the reassociation was carried out under standard conditions and that $R = 1$ for the most slowly reassociating fraction, calculate the genome size of *Amphiuma*.



(a)



(b)

Figure A.10
Hypothetical C_0t curves for
(a) fly embryo DNA and
(b) DNA from polytene nuclei
(Problem A.4).

APPENDIX
Analysis of Nucleic
Acid Association
Reactions

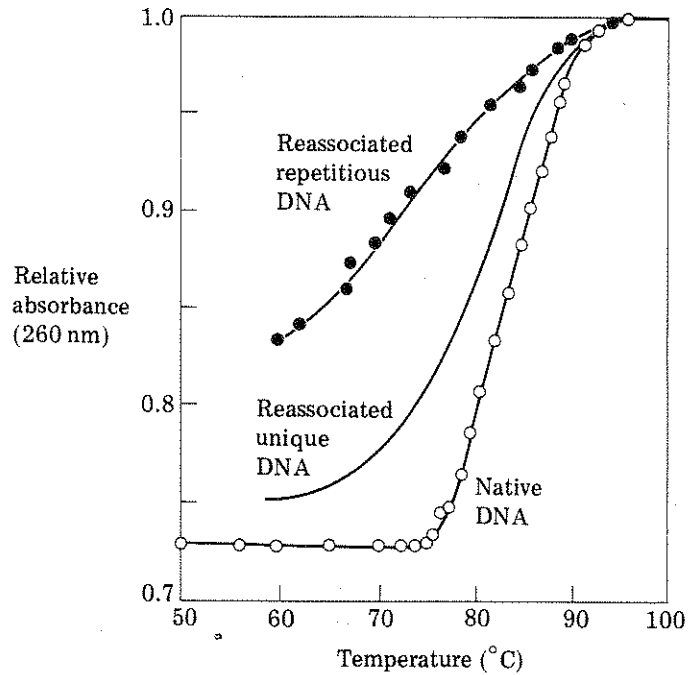


Figure A.11
Absorbance-temperature profiles of various calf DNAs. Native DNA is high-molecular-weight (Problem A.6). [Adapted from R. J. Britten and D. Kohne, *Science*, 161, 529 (1968).]

- A.6 Figure A.11 shows absorbance-temperature profiles for various calf DNAs.
- How would you explain the large difference in the melting profiles of the reassociated repetitive and unique DNAs?
 - What is the significance of the difference between the profiles of native and reassociated unique DNA?
- A.7 Thermal denaturation experiments can yield valuable information about the evolutionary relatedness of various frequency classes from different animals. In one such experiment, a mixture of denatured ^3H -mouse DNA (290 $\mu\text{g}/\text{mL}$) and ^{14}C -rat DNA (0.086 $\mu\text{g}/\text{mL}$) are incubated together to a C_0t value sufficient for reassociation of the repeated sequences but not the unique sequences. Because of the difference in concentration of mouse and rat DNA there is essentially no self-association of the rat sequences. The mixture then is passed over a hydroxyapatite column. About 55% of the added DNA is bound to the column. The bound material is eluted by passing a phosphate buffer through the column while increasing its temperature linearly with time. The fractions then are assayed for acid-precipitable radioactivity. The results are shown in Figure A.12.
- Which column fractions correspond to the reassociated sequences with the least mismatch?

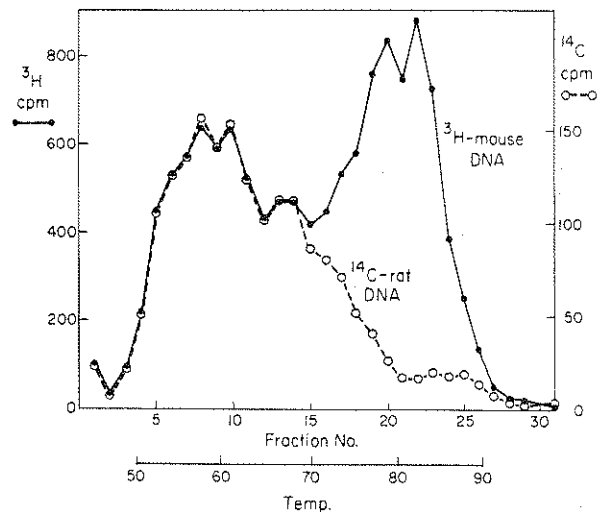


Figure A.12
Thermal denaturation of reassociated repeated sequences of mouse and rat DNA (Problem A.7). [From N. Rice in H. H. Smith (Ed.), *Evolution of Genetic Systems, Brookhaven Symposia in Biology*, Vol. 23, Gordon and Breach, New York, 1972.]

- b. The repeated sequences of the mouse can be divided into two general categories, middle-repetitive and highly repetitive. Which column fractions correspond to each class?
- c. Which of these frequency classes shows more homology with rat DNA?
- d. The rat and mouse evolutionary lines diverged about 10^7 years ago. From this experiment, what can you say about the relative times of evolution of the middle- and highly repetitive sequences of DNA in the mouse?

A.8 Consider the data in Table A.1.

Table A.1

Effect of Reassociation Temperature on Apparent Percentage of Repetitious DNA (Problem A.8)^a

DNA	Temperature of reassociation (°C)	Percentage of genome as repetitious DNA
Human	51	43
	60	34
	66	25
Mouse	51	53
	60	34

^aFrom D. Kohne, *Quarterly Review of Biophysics*, 33, 327 (1970).

- a. Why does a change in the reassociation temperature affect the percentage of the genome classified as repetitious DNA?
- b. What does this temperature effect indicate about mammalian DNA?
- c. How would you expect the melting curves of DNA reassociated at 60°C and 51°C to differ?

A.9 Differential polyteny has been suggested to account for large differences in DNA content per nucleus within certain taxonomic groups that do not show corresponding differences in chromosome number. According to this hypothesis, the larger genomes evolved from the smaller ones by lateral duplication of the entire genome to produce multistranded chromosomes.

Vicia faba contains five times as much DNA per nucleus as *Vicia sativa*, but both these species of the pea genus have a haploid chromosome number of 8. Because the two species are closely related, it has been argued that the additional DNA of *V. faba* is unlikely to represent new genetic loci not present in *V. sativa*. Moreover, studies of chromosome structure using trypsin digestion have indicated that *V. faba* has more lateral strands in its chromosomes than does *V. sativa*. Thus *V. sativa* and *V. faba* would appear to be likely candidates for differential polyteny.

The differential polyteny hypothesis can be tested using DNA reassociation kinetics. An increase in polyteny increases the frequency of each sequence by the same amount and thus leaves the concentration of each sequence per gram of DNA unchanged. Because the rate of DNA reassociation is concentration dependent, differential polyteny will not change the rate of DNA reassociation. However, if new sequences are responsible for the increase in DNA content, the concentrations of individual sequences will be lower, and their rate of reassociation will decrease.

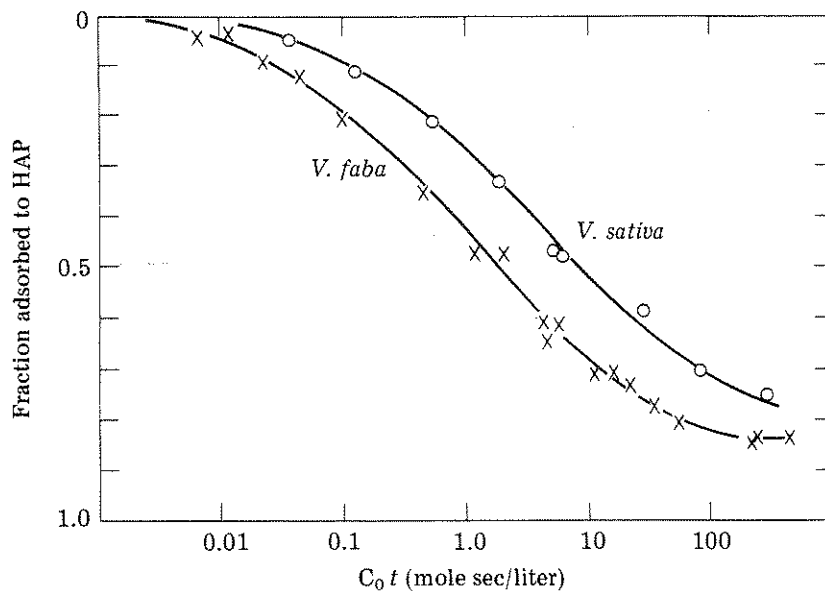
With these considerations in mind, the reassociation kinetics of *V. faba* and *V. sativa* DNA were studied, with the results shown in Figure A.13. Under the conditions used, sequences present once per haploid genome would be half-reassociated at C_0t values of 3000 and 15,000 mol sec/L for DNA from *V. sativa* and *V. faba*, respectively. Therefore Figure A.13 represents the reassociation curves of repeated DNA only. Do the results support the differential polyteny hypothesis or not? Explain.

A.10 In an important series of experiments with *Xenopus laevis* DNA, hydroxyapatite chromatography was used to assay the reassociation of labeled fragments of various lengths with a 10,000-fold excess of unlabeled 450-nucleotide fragments. Figure A.14 shows the fraction of labeled fragments reassociated sufficiently to bind to the column as a function of labeled fragment length (L), at $C_0t = 50$ mol sec/L. Under these conditions, only repetitive sequences associate.

In a separate experiment, larger quantities of 450-nucleotide and 1400-nucleo-

APPENDIX
Analysis of Nucleic
Acid Association
Reactions

Figure A.13
Reassociation kinetics of isolated repetitive DNA from *V. faba* and *V. sativa* monitored by adsorption to hydroxyapatite (Problem A.9). [From N. Straus, *Carnegie Institution Year Book*, 71, 258 (1972).]



the fragments each were reassociated independently to $C_0t = 50$ mol sec/L. The reassociated fractions then were isolated by hydroxyapatite chromatography and analyzed by thermal denaturation. The hyperchromicities observed were 17% and 10%, respectively, where hyperchromicity is defined as the increase in optical density from 60°C to 98°C expressed as a percentage of the 98°C value. The hyperchromicity of native *Xenopus* DNA is 27%.

- When *Xenopus* DNA is sheared to an average chain length of 4000 nucleotides, what fraction of the fragments consists entirely of unique sequences, based on the data in Figure A.14?
 - From the data in Figure A.14, what fraction of the *Xenopus* genome is made up of repetitive sequences? Is your answer more, or less, reliable than the value for this fraction that would be obtained from a standard C_0t curve?
 - Propose an interpretation for the change in slope of the curve in Figure A.14 at $L \simeq 800$ nucleotides, and for the continued slow rise beyond this point.
 - Based on the hyperchromicity experiment, what are the average lengths of repetitive sequence in 450-nucleotide and 1400-nucleotide fragments, respectively? How would you explain the differences in these lengths?
 - Suggest an alternative and more direct experiment that would answer the first question in part d, using methods that you have encountered in this Appendix.
- A.11 Hybridization experiments using labeled 28S rRNA and unlabeled DNA from *Xenopus laevis* demonstrate that at saturation 0.070 μg of 28S rRNA (molecular weight 1.6×10^6) is bound for each 100 μg of DNA. *Xenopus* somatic cells contain 3.6×10^{12}

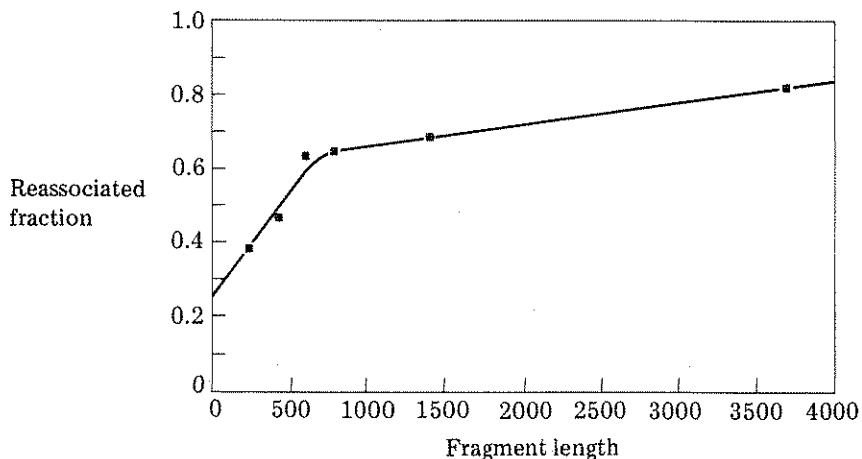


Figure A.14
Reassociation of *Xenopus laevis* DNA at $C_0t = 50$ as a function of fragment length (Problem A.10). [Adapted from E. H. Davidson and R. J. Britten, *Quarterly Review of Biology*, 48, 565 (1973).]