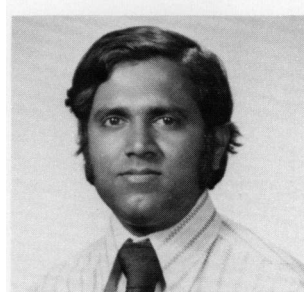
Modeling the shearing and rehybridization process of DNA

S. Sitharama Iyengar

SIMULATION 1981 36: 173
DOI: 10.1177/003754978103600505

Keywords: DNA, modeling, molecular biology, rehybridization, simulation

Modeling the shearing and rehybridization process of DNA



by

S. Sitharama Iyengar
Department of Computer Science
Louisiana State University
Baton Rouge, Louisiana 70803

S. SITHARAMA IYENGAR is an assistant professor of computer science at Louisiana State University in Baton Route. His research and teaching interests are centered around the design of algorithms, information system design, and modeling and simulation of systems. He is an invited editor for a book on *Computer Modeling of Complex Biosystems* for CRC Press, Inc., and he has published more than 20 papers in international journals and proceedings.

ABSTRACT

The physiological complexity of biological systems makes it very difficult to formulate hypotheses to explain their behavior and to test such hypotheses. Simulation models are precise statements of hypotheses; hence comparing the behavior of a model with that of the corresponding real-world system is a way of testing the adequacy of the hypothesis represented by the model.

In this paper we use a simulation model to assess a hypothesis about the hydrodynamic shearing of DNA. A second model uses the output of the first one as an input. This latter model is being used to study the rehybridization process, which involves several unknown stochastic distributions.

INTRODUCTION

Recent rapid advances in biology, medicine, and computer science have made it practical to model various aspects of biological systems. These systems often exhibit counterintuitive behavior which is difficult to explain, and their complexity is likely to cause problems in checking hypotheses by direct experimentation on the biological system. In this situation computer models, which give the researcher direct access to all the variables and to the stochastic distributions, can be of great help.

The DNA molecule contains an immense amount of information in coded form. A major problem in molecular biology is how this coded information controls the growth, development, and major characteristics of biological systems.[3] The first level at which control can be exercised is called transcription. At this level the DNA molecule can be regarded as a linear string of base sequences of four sets of atoms identified as A, T, G, and C. These sets are the four types of bridges that bind together the two long filaments that form the outside of the DNA double spiral. In the process of transcription, the DNA sequences of A, T, G, and C become new linear sets (codes) called the MRNA (messenger ribonucleic acid). The transcribed sets become new linear sequences of sets of atoms identified by the letters A, U, G, and C in MRNA. In the transcription U replaces A, A replaces T, G replaces C, and C replaces G. (For more information on this process, refer to any textbook in biology.) Transcription converts the DNA code into units called genes.

There are numerous hypotheses concerning the mechanisms that control the process of transcription. Some, perhaps all, of the hypothesized mechanisms may be operating, but none of the hypotheses satisfactorily explains all observed phenomena. Many of the recently proposed control mechanisms involve the organization of genes along the DNA strand.[1,2,3,4] Experimental evidence indicates that when very long strands of DNA are rehybridized (discussed below), networks of code sequences of the A, T, G, and C elements form as would be expected, but some of them are insoluble—that is, the two filaments of the DNA molecule cannot be separated, making further division and reproduction of the affected section of the DNA molecules impossible and preventing the formation of the corresponding genes by transcription. The insoluble DNA forms a precipant, making it very difficult to measure reaction rates achieved in a laboratory experiment.

REHYBRIDIZATION

The process

The very long double strands of the DNA molecule can be broken into shorter double strands by hydrodynamic shearing, which results from stirring the solution in which the strands are suspended. These shorter double strands can be separated into twice as many single strands by heat. When the heat is lowered, the single strands rehybridize by colliding with other single strands of the same length.

KINETICS OF REHYBRIDIZATION

The kinetics of the process of shearing can be explained by a second-order rate equation as follows:

$$\frac{d(M)}{dt} = K(N)^2 \qquad (1)$$

where

M represents the formation of double-stranded DNA

N represents the concentration of single-stranded DNA

Equation 1 can be written as follows:

$$\frac{d(\text{dsDNA})}{dt} = K_i \left[(\text{ssDNA}^+)_i * (\text{ssDNA}^-)_i \right] \qquad (2)$$

where

(dsDNA) represents the formation of double-stranded DNA

(ssDNA^+) and (ssDNA^-) represent the concentration of single-stranded DNA sequence i and its complement.

Equation 2 ignores the mismatching that occurs in the real-world experiment, but it should give a good approximation, since experimental data indicate that the error thus introduced is not great.

There is a general hypothesis that each gene found in the DNA strands of an organism's cells is unique. Moreover, each cell is thought to have one strand per cell or one strand per chromosome, such a strand being a chromatid. Therefore, the concentration of any particular code sequence i is N_i, the number of double strands of sequence i in shearing (per unit volume, of course). If the average strand length after shearing is 3 700, then there will be approximately 3 million fragments from each strand and each will have a sequence i. Therefore there will be $3N$ million double-strand fragments with sequence i. After shearing, the total number of single strand of fragments of chromatids in solution will be $6N$ million.

THE HYDRODYNAMIC SHEARING MODEL

The need

Because there is no way to determine experimentally how DNA breaks up during shearing, a computer simulation model is one of the best methods for obtaining evidence of this distribution. We shall therefore discuss the evolution and validation of a computer model which is capable of giving break distributions which have been validated by comparing the simulation results with those of real-world experiments.

The model was designed to be general in order to provide an opportunity for increasing or decreasing the completeness of system representation, a distinct advantage in dealing with a complex biological system.

Formulation

Iyengar and Quave[5] recently proposed a model of a distribution function for hydrodynamic shearing of DNA. Their conceptual model for DNA undergoing hydrodynamic shearing is that of blocks connected by stiff but flexible elements which can be broken, but the blocks themselves cannot be broken. Strands of these blocks with their connections are viewed as suspended in aqueous solution which is undergoing brownian motion and mass motion due to stirring.

The basic physical model is a flexible rod undergoing stress at some interval along the rod. The stress is supplied by shear planes, vortices, etc. The orientation of DNA molecules in dilute aqueous solution is believed to be uniformly random. However, the distribution of the turbulent shear forces is geometrically quite complex and not easily described.

As an approximation, the model uses a discrete Beta distribution (choosing an interval between two extremes) for choosing the interval where stress will occur. The probability that a breakdown will occur in a strand (nucleotide) undergoing an average shear force is exponentially dependent on its length. It can be expressed mathematically as follows:

$$P\ (\text{break at J will occur}) = 1 - \left[\exp\left(\frac{-\text{nucleotide length}}{\lambda} \right) \right]$$

Please note that whether or not a break occurs is a binomial process. To simulate the above process, a uniform random length R generated by the linear congruential method will be compared to P (breakdown at J). If $R \leq P(J)$, then the strand breaks at J.

STOCHASTIC BEHAVIOR

When forces are being applied to the connecting elements of the model, the blocks undergo finite but complex vibrational motion and have complex patterns of vibration. As a consequence, the connecting elements may be stretched or compressed. Therefore one cannot predict exactly what force will be required to break any given connector, but there will be an expected value and a distribution function that together describe the force that will break connectors. The distribution of strands in solution is assumed to be uniform random. The strength of the connector elements to which the shear force is applied is a stochastic variable whose distribution function is unknown but which has a strong central tendency. It is also assumed that breakage of a strand seldom if ever occurs at a distance greater than one-tenth of the strand length from the center of the strand.[8]

The following variables are believed to control the stochastic behavior of the process of breaking connectors:

(1) Velocity gradient in the solution (a function of unknown distribution)

(2) The position of the connectors on the strand (an unknown density function)

(3) The strand length which will result from breaking connectors (modeled as a stochastic difference equation)

(4) The position and orientation of the DNA chain in the solution (a uniform random variable).

One can see that the number of fragments that break out of N whole strands of DNA is a binomial process, but the probability of a break is a complex interaction of a function of several stochastic variables, some of which have unknown distribution functions. In this situation a Gaussian distribution function was chosen to locate the stressed connector which may break. The following function describes the properties of Gaussian function.

The random variable is normally distributed if its density function is a Gaussian curve. In this case,

$$f(y) = A\, e^{-\beta y^2} \qquad \beta > 0$$

Since

$$\left[\int_{-\infty}^{\infty} e^{-\beta y^2}\, dy = \sqrt{\pi/\beta} \right] \qquad (3)$$

For more information on the corresponding distribution function, see Reference 9.

SHEAR-STRESS MODEL

An algorithm is basically a finite set of instructions to accomplish a particular task. The shear-stress model embodies the following algorithm:

Step 1: Choosing the stress point, the site of the event

Step 2: Determining the probability of occurrence of a break at the stress point; we refer to this process as the probability-of-event process

Step 3: Executing a Bernoulli trial (deciding whether a break occurs)

Step 4: Storing a temporary event vector (breakage distribution) for use in the time-slice processing

Step 5: Updating the event vector for time slice θ

Step 6: Collecting and printing break events and statistics for each strand

Step 7: Printing a summary for all strands.

REHYBRIDIZATION MODEL

The following algorithm describes the process for rehybridization:

Step 1: Using the shear model, find the number of fragments from N identical chromatids. (Single and double strands are stored within the shear-stress model.)

Step 2: Select from the pool two fragments of the same length, using a uniform probability distribution.

Step 3: Determine the length of an uninterrupted double-strand chain required for stability.

Step 4: If the fragments selected in Step 2 are long enough to form a stable chain, eliminate them from the sample pool and score a success.

Step 5: If the selected fragments are not stable, return them to the pool and score a failure.

In the above algorithmic process, the kinetics of formation of double-stranded DNA is calculated using weighted counters. The appropriate weights for the counters depend on unknown distribution functions. We shall attempt to find the weights by means of simulation runs in which only a single DNA molecule is used. Under these conditions perfect first-order kinetics should prevail.

SOFTWARE

The computer program for the shear-stress model is written in FORTRAN IV. Since ease of use was given the high priority in designing the computer program, it was written in structured form to facilitate modification. The program for the rehybridization is not complete at this time since there are some unsolved problems arising from the computational complexity of the system.

RESULTS

The shear-stress model produces fragment lengths with a Gaussian distribution. The average strand length was 50.78. The standard deviation was 9.98, and the coefficient of variation was 0.196. An F test on the polynomial regression gives a value of 81.7 with three degrees of freedom for the regression. Most of the total variance was the result of the regression procedure. The results obtained by this model are consistent with the real-world evidence, but the algorithm underlying the rehybridization model is not working very well, the obvious reason being the complexity and the unknowns of the system being modeled.

FINAL THOUGHTS ON THIS MODEL

Biological systems are very complex, which makes computer simulations of them difficult and challenging. The effort reported in this paper shows that models of biological systems can be structured for convenient use and to represent and test hypotheses concerning microbiological systems by means of computer simulation techniques. Future endeavors to further generalize the rehybridization model and to better define the relevant stochastic distributions would add significantly to its value. The algorithm for this model is more difficult to program than the shear-stress algorithm. It is not clear at this point how to develop a good simulator for the rehybridization process, the obvious reason being the complexity of the system.

ACKNOWLEDGEMENTS

The author is very grateful to the referees and the copy editors of the journal, who have improved the content and readability of the paper. The author also acknowledges the assistance of Stephen A. Quave in preparing this paper.

REFERENCES

1 BONNER, J. WU, J.-R.
A Proposal for the Structure of the Drosophils Genome
Proceedings of the National Academy of Sciences vol. 70 1973 pp. 535-537

2 CRICK, F.
General Model for the Chromosomes of Higher Organisms
Nature vol. 234 1971 pp. 25-27

3 DAVIDSON, E.H. BRITTEN, R.J.
Organization, Transcription, and Regulation in the Animal Genome
Quarterly Review of Biology vol. 48 1973 pp. 563-613

4 GEORGIEV, G.P.
On the Structured Organization of Operon and the Regulation of RNA Senthesis in Animal Cells
Journal of Theoretical Biology vol. 25 1969 pp. 473-490

5 IYENGAR, S.S. QUAVE, S.A.
A Computer Model for Hydrodynamic Shearing of DNA
Computer Programs in Biomedicine vol. 9 1979 pp. 160-168

6 IYENGAR, S.S. QUAVE, S.A.
A Computer Model for Hydrodynamic Shearing of DNA. Part II: Gibbs Phenomenon
Computer Programs in Biomedicine vol. 10 1979 pp. 133-135

7 IYENGAR, S.S.
A Computer Model for Hydrodynamic Shearing of DNA: Part III
Computer Programs in Biomedicine In press

8 LEVINTHAL, C. DAVISON, P.F.
Degradation of Deoxyribonucleic Acid under Shearing Forces
Journal of Molecular Biology vol. 3 1961 pp. 674-683

9 PAPOULIS, P.
Probability Random Variables and Stochastic Process
McGraw-Hill New York 1965