# KINETIC EQUATION FOR DNA-DNA HYBRIDIZATION

SANTANU DUTTA
Biophysics Laboratory, Physics Department
University College of Science, 92, A.P.C. Road
Calcutta 700 009, India

THE kinetic equation describing DNA-DNA hybridization, a versatile tool to unravel genetic relatedness among various species, is deduced from a semi-empirical basis.

## *Introduction*

Though the study on the kinetics of DNA reassociation[1] has been carried out for more than a decade, there has been very little progress in the elucidation of DNA sequence homology among related species. This has been mainly due to the fact that the kinetics of DNA DNA hybridization has surprisingly remained a virgin field In this paper the kinetic equation describing DNA-reassociation is first derived in a slightly modified format from the one derived by Britten *et al.*[2] and then it is extended to deduce the more general case of DNA-DNA hybridization.

## *Theory*

### (A) *Reassociation*

Before presenting the arithmetic of reassociation, the symbols that will be used in the following section is clarified below.

$C_0$ = Total concentration of DNA.
$P$ = Length of sheared pieces of DNA.
$G$ = Genome length (haploid DNA content per cell).
$K$ = Reaction rate constant
$N$ = Number of single strand pieces of DNA.

Other symbols will be defined wherever necessary. The units of the symbols are suitably chosen so as to maintain the validity of the equation concerned. This is because we will be dealing mainly with dimensionless numbers.

Meaningful study of reassociation kinetics can only be done by fragmenting the DNA into suitable lengths. Fragmentation, though not mandatory in case of prokaryotes, is a must in case of eukaryotes. This is due to the presence of repeated sequences in the latter. Generally the length of fragmented pieces is of the order of the length of the repeating block.

Fragmentation leads to the separation of the pieces into two distinct classes : (1) Unique, (2) Repeated. To begin with we have $C_0/P$ fragments and its distribution is given by

$$\frac{C_0}{P} = \{N_u + (N_1 + N_2 + \cdots N_N)\}\frac{C_0}{G} \qquad (1)$$

where $N_u$ gives the number of unique fragments and $N_1, N_2, \ldots$, gives the frequencies of repeated sequences per genome.

Now reassociation takes place due to bimolecular collision of the two complementary strands,

$$\text{thus } \frac{dN}{dt} = -KN^2. \qquad (2)$$

The initial concentration of each unique fragment is $N_0$ where $N_0 = C_0/G$.

$$\therefore \int_{N_0}^{N_{u1}} \frac{dN}{N^2} = -\int_0^t K dt \qquad (3)$$

or,

$$N_{u1} = \frac{N_0}{1 + KN_0 t} \qquad (4)$$

where $N_{u1}$ is the number of single strand pairs of this unique class which has not reassociated in time $t$. Since we have $N_u$ unique fragments we should have $N_u$ equations identical to equation (4).

Adding these we have

$$N_u \cdot N_{u1} = \frac{N_u \cdot N_0}{1 + KN_0 t} \qquad (5)$$

For the repeated sequence of frequency $N_1$, the initial concentration is given by $N_{01}$

where

$$N_{01} = N_1 \cdot \frac{C_0}{G}.$$

Thus in a manner analogous to equation (4), we have

$$N_{11} = \frac{N_{01}}{1 + KN_{01} t} \qquad (6)$$

where $N_{11}$ is the number of single strand pairs of this repeated class which has not reassociated in time $t$. We should have similar equations for $N_2$, $N_3$, .... Thus adding equations (5), (6), ... we have

$$(N_u \cdot N_{u1} + N_{11} + N_{22} + \ldots N_{NN}) = \frac{N_u N_0}{1 + KN_0 t} + \frac{N_{01}}{1 + KN_{01} t} + \cdots \frac{N_{0N}}{1 + KN_{0N} t} \qquad (7)$$

Now L.H.S. $= C/P =$ Total number of single strand fragments remaining unreassociated in time $t$. Also substituting the values of $N_0$, $N_{01}$, $N_{02}$ ... and rearranging equation )7) we have

$$\frac{C}{C_0} = \frac{PN_u}{G + KC_0 t} + \sum_{i=1}^{N} \frac{PN_i}{G + KN_i C_0 t} \qquad (8)$$

This is the kinetic equation that describes the reassociation of eukaryotic DNA.

(B) *Hybridization*

DNA-DNA hybridization though a very powerful tool to unreveal genetic relatedness among species, find its limited use in the study of eukaryotic DNA. This is primarily because hybridization in eukaryotes is complicated by the presence of repeated sequences. So far to the best of my knowledge there has been no theoretical basis for the kinetics of interspecies reassociation (hybridization) in higher organisms. In this section, utilising the methodology developed in the preceding section, the kinetic equation that will adequately describe DNA-DNA hybridization is proposed.

Let the two types of DNA have concentration $C_{01}$ and $C_{02}$ respectively. Hence the total concentration $C_0 = C_{01} + C_{02}$. Now as in equation (1) we have

$$\frac{C_{01}}{P} = \left(N_u + \sum_{i=1}^{N} N_i\right) \frac{C_{01}}{G_1} \tag{9}$$

$$\frac{C_{02}}{P} = \left(M_u + \sum_{i=1}^{N} M_i\right) \frac{C_{01}}{G_2} \tag{10}$$

where $G_1$, $G_2$ are the genome lengths, $N_i$, $M_i$ are the frequencies and N, M are the classes of repeated sequences of the two species respectively. Now homology may be in unique as well as in the repeated sequences.

Let $H_u$ be the number of unique fragment common to both. Thus as in equation (5), we have for non-homologous unique fragments,

$$(N_u - H_u)\, N_{u1} = \frac{(N_u - H_u)\, N_0}{1 + KN_0 t} \tag{11}$$

$$(M_u - H_u)\, M_{u1} = \frac{(M_u - H_u)\, M_0}{1 + KM_0 t} \tag{12}$$

Again each homologous unique class has $(N_0 + M_0)$ fragments, so for homologus unique fargments we have

$$H_u \cdot H_{u1} = \frac{H_u (N_0 + M_0)}{1 + K (N_0 + M_0)\, t} \tag{13}$$

In case of repeated sequences, say there is $c$ common class which are homologous,

*i.e.*, $N_1 \ldots N_c$ and $M_1 \ldots M_c$ are common

*or*,

$$R_{01} = N_1 \cdot \frac{C_{01}}{G_1} + M_1 \cdot \frac{C_{02}}{G_2} \tag{14}$$

where $R_{01}$ is the total number of fragments of the first common class.

*Hence*

$$R_{11} = \frac{R_{01}}{1 + KR_{01} t} \tag{15}$$

where $R_{11}$ is the number of single strands of this repeated class which has not reassociated in time $t$. Similar expressions are for the rest of the homologous repeated class. Adding these expressions we should have,

$$(R_{11} + R_{22} + \cdots R_{cc}) = \sum_{i=1}^{c} \frac{R_{0i}}{1 + KR_{0i} t} \tag{16}$$

For non-homologous repeated sequence we should have

$$N_{jj} = \frac{N_{0j}}{1 + KN_{0j} t} \tag{17}$$

and

$$M_{jj} = \frac{M_{0j}}{1 + KM_{0j} t} \tag{18}$$

where $j = c + 1$.

Similar expression should be for the rest of the non-homologous repeated classes,

*i.e.*,

$$\sum_{j=c+1}^{N} N_{jj} = \sum_{j=c+1}^{N} \frac{N_{0j}}{1 + KN_{0j} t} \tag{19}$$

and

$$\sum_{j=c+1}^{M} M_{jj} = \sum_{j=c+1}^{M} \frac{M_{0j}}{1 + KM_{0j} t} \tag{20}$$

Thus summing up equations (11), (12), (13), (16), (19), (20), we have

$$\frac{C}{P} = \frac{(N_u - H_u)\, N_0}{1 + KN_0 t} + \frac{(M_u - H_u)\, M_0}{1 + KM_0 t} + \frac{H_u (M_0 + N_0)}{1 + K (N_0 + M_0)\, t} + \sum_{i=1}^{c} \frac{R_{0i}}{1 + KR_{0i} t} + \sum_{i=c+1}^{N} \frac{N_{0i}}{1 + KN_{0i} t} + \sum_{i=c+1}^{M} \frac{M_{0i}}{1 + KM_{0i} t} \tag{21}$$

If the two DNA type is mixed in a ratio such that

$$\frac{C_{01}}{G_1} = r \cdot \frac{C_{02}}{G_2}$$

then we have

$$R_{0i} = (r \cdot N_i + M_i) \frac{C_{02}}{G_2}$$

and

$$N_{0i} = r \cdot N_i \cdot \frac{C_{02}}{G_2}$$

where $r$ = ratio of genome concentration.

Thus substituting the values of $N_0$, $M_0$, $R_{0t}$, $N_{0t}$, $M_t$, and rearranging equation (21) we have,

$$\frac{C}{C_{0t}} = P\left[\frac{(N_u - H_u)\,r}{G_2 + KrC_{0t}t} + \frac{(M_u - H_u)}{G_2 + KC_{0t}t} + \frac{(r+1)\,H_u}{G_2 + (r+1)\,KC_{0t}t} + \sum_{i=1}^{i=c} \frac{A_i}{G_2 + KA_iC_{0t}t} + \sum_{i=c+1}^{i=N} \frac{r \cdot N_i}{G_2 + KrN_iC_{0t}t} + \sum_{i=c+1}^{i=M} \frac{M_i}{G_2 + KM_iC_{0t}t}\right] \quad (22)$$

where

$$A_i = (r \cdot N_i + M_i).$$

This is the equation that describes hybridization, and it is easy to see that reassociation is a special case of hybridization, as equation (22) reduces to eqution (8) when the two DNA type is equal in all respects.

*Discussion*

The utility of these equations depends on whether the unknown parameters are computable or not. In case of equation (8), the unknowns (Repetition Class and Frequency) can be computed using non-linear least square fit method[3], or by "trial linearizing method" (unpublished result). Evaluation of the unknown parameters of equation (22) cannot however be easily done if we are working with two unknown species. However if on the other hand the repeated class and frequency of the two species are known then $H_u$ and $c$ remain only the two unknowns. For the evaluation of $c$ the simplest method would be to put $c = 0, 1, 2, \ldots N$ and see which of the values of $c$ fits best into equation (22) at various $C_0t$ values. This is facilitated by the fact that reassociation of unique sequences occur when the reassociation of repeated sequence is more or less complete. Thus while evaluating $c$ we can take the first three terms of equation (22) as constants and equal to $P(r \cdot N_u + M_u)/G_2$. For evaluation of $H_u$ we use late $C_0t$ values where the last three terms equal zero.

Utilizing these $C_0t$ values, evaluation of $H_u$ is simply straightforward. From the experimenters point of view, $r$ is a vital parameter. In fact this is the only parameter that can be judiciously varied to give different degrees of resolution in the analysis of sequence homology among various species. Optimum values of $r$ depends on the ratio of genome size of the two species. Detailed analysis of equation (22) with respect to $r$, corresponding to various degrees of homology among species, will be published elsewhere.

*August* 8, 1980.

1. Britten, R. J. and Kohne, D. E., *Science*, 1968, **161**, 529.
2. —, Graham, D. E. and Neufeld, B. R., In *Methods in Enzymology* (eds. L. Grossman and K. K. Moldave), Part E, Academic Press, New York, 1974, **29**, 363.
3. Pearson, W. R., Davidson, E. H. and Britten, R. J., *Nucleic Acids Res.*, 1977, **4**, 1727.

## ANTIMICROBIAL PROPERTIES OF THE ESSENTIAL OIL OF *VATERIA INDICA*

G. S. Grover and J. Tirumala Rao
Department of Chemistry
University of Saugar, Sagar (M.P.)

*Vateria indica* Linn (Syn. *V. malabarica* and *Chloroxylon dupada*) belongs to the natural order *Dipterocarpaceae*. It is distributed in the forests of Western Ghats of India from North Kanara to Tarvancore and Tirunelveli at elevations up to 4000'[1]. Various parts of the plant are used for curing different ailments[2,3]. The oleo-resin is used as an incense in paints, varnishes and in ointments. It is also used as stimulant, dressing for carbuncles and other ulcerations. In the present investigations the study of the antimicrobial activity of the essential oil extracted from the oleo-resin has been undertaken.

The oleo-resin when extracted by water and steam distillation has yielded a yellowish brown coloured essential oil in 1·5% yield. For the determination of antibacterial activity the "Oxoid Nutrient Broth" was used for making the inoculum and the media was prepared by adding 2% agar to the "Oxoid Nutrient Broth". For the determination of antifungal activity "Saboraud's Broth" was used for making the inoculum and the media was prepared by adding 2% agar to the "Saboraud's Broth". For determining the antimicrobial activity paper disc diffusion method of Maruzzella and Henry[4] was used.

The bacteria tested are—*Bacillus anthraces*, *Bacillus subtilis*, *Corynebacterium pyogenes*, *Escherichia coli*, *Haemophillus influenzae*, *Klebsiella pneumoniae*, *Pasteurella* sp., *Proteus vulgaris*, *Pseudomonas aeruginosa*, *Salmonella pullorum*, *Salmonella newport*, *Salmonella richmond*, *Salmonella stanley*, *Salmonella typhimurium*, *Staphylococcus aureus* and *Streptococcus agalactiae*.

The fungi tested are—*Aspergillus flavous*, *Aspergillus fumigatus*, *Aspergillus niger*, *Candida albicans*, *Fusarium* sp., *Penicillium digitatum* and *Rhizophus stolonifera*.