PHYSICS

1960

SovietRxiv

Abstract

Full Text

PHYSICS

M. V. VOL' KENSHTEIN and Al. M. ELYASHEVICH

THE INFLUENCE OF THE SOLVENT AND OF THE SPECIFICITY OF NUCLEOTIDE ADDITION ON THE REDUPLICATION OF DEOXYRIBONUCLEIC ACID

(Presented by Academician I. V. Obreimov, November 5, 1959)

In a paper by one of us (^{1}), a model statistical theory of the reduplication of deoxyribonucleic acid (DNA) was proposed. In the present paper we consider the influence on the reduplication process of important entropic and energetic factors: the interaction of DNA with the solvent (water), the specificity of nucleotide addition, and also the flexibility of the free DNA chain.

We shall make use of the previously described model (^{1}). The double helix of DNA consists of two polymer chains with N units in each. The surrounding solution contains $n - M$ water molecules and M free nucleotides of four types, containing thymine, cytosine, guanine, and adenine (^{2}). For simplicity we take the quantities of all four nucleotides in the solution to be equal, $M/4$. The energy released upon polymerization of a pair of nucleotides is $-\varepsilon_1$; the energy of the hydrogen bond connecting two nucleotides in the double helix is $-\varepsilon_2$. We continue to assume that rupture of hydrogen bonds occurs at one of the ends of the double helix. At the moment of reduplication there are $2r$ units of separated chains, to which, with an energy gain ε_2, any nucleotides from the solution may attach. If two "suitable" nucleotides have attached side by side, then they polymerize with an energy gain ε_1. A water molecule may also attach to a given free unit, with an energy gain ε_3 (^{3}). We shall characterize the state of an r-membered released chain by a set of quantities σ_j $(j = 1, 2, \dots, r)$, where σ_j refers to the j-th unit and is equal to 0 if a "suitable" nucleotide has attached, $+1$ if an "unsuitable" nucleotide has attached, and -1 if the unit has remained free. For a state with the given set $\{\sigma_j\}$, the number of "suitable" nucleotides attached to the two released chains is equal to

$$2l = 2\left(r - \sum_{j=1}^{r} \sigma_j^2\right). \tag{1}$$

The total number of attached nucleotides is equal to

$$2m = 2\left(r - \frac{1}{2}\sum_{j=1}^{r}\sigma_j^2 + \frac{1}{2}\sum_{j=1}^{r}\sigma_j\right). \tag{2}$$

The number of free units is equal to $2(r-m)$. The statistical sum is written in the form

$$Q = \sum_{r=0}^{N}\sum_{\{\sigma_j\}} Q_{12}Q_3Q_4, \tag{3}$$

where

$$Q_{12} = q^{2N}\exp\frac{2(N-1)\varepsilon_1 + (N-r)\varepsilon_2}{kT} Z^{2(r-m)} \tag{4}$$

refers to the nucleotides of two partially separated chains. Here q is the internal statistical sum of such a nucleotide; Z is the number of additional conformations of a unit to which a new nucleotide has not attached. Obviously, Z determines the flexibility of the DNA chain.

$$Q_3 = q^{2m}\exp\left(\frac{2\sum_{j=1}^{r}V_{j,j+1}}{kT}\right)3^{2(m-l)} \tag{5}$$

refers to $2m$ newly attached nucleotides, where

$$V_{j,j+1} = -\varepsilon_1 - \varepsilon_2 + \frac{2\varepsilon_1+\varepsilon_2}{2}\sigma_j^2 + \varepsilon_1\sigma_{j+1}^2 - \frac{1}{2}\varepsilon_2\sigma_j - \varepsilon_1\sigma_j^2\sigma_{j+1}^2. \tag{6}$$

The factor $3^{2(m-l)}$ accounts for the possibility of attachment, to each unit, of any of the three "unsuitable" nucleotides. Q_4 is the statistical sum of free nucleotides and water molecules—free and adsorbed on DNA,

$$Q_4 = {q'}^{M-2m}\sum_{p=0}^{2(r-m)} Q'_p, \tag{7}$$

where q' is the internal statistical sum of a free nucleotide, and Q'_p is the statistical sum of water molecules in the state corresponding to the attachment of p water molecules to the free units of the two DNA chains. We have

$$Q'_p = q_0^p {q'_0}^{n-M-p}\frac{(2r-2m)!}{p!(2r-2m-p)!}\exp\frac{p\varepsilon_3}{kT}\cdot R_p, \tag{8}$$

where q_0 is the internal statistical sum of an attached water molecule, and q_0' that of a free water molecule, while R_p characterizes the mixing of water with nucleotides,

$$R_p = \frac{(n-2m-p)!}{(n-M-p)!\prod_{i=1}^{4}\left(\frac{M}{4}-m_i\right)!}; \tag{9}$$

m_i is the number of attached nucleotides of type i, $m_1+m_2+m_3+m_4=2m$.

If $n \gg M \gg m_1, m_2, m_3, m_4$, then R_p practically does not depend on m_i, but only on their sum. Using Stirling' s formula, we obtain

$$R_p \simeq \left(\frac{n}{n-M}\right)^n \left(\frac{4(n-M)}{M}\right)^M \left(\frac{M}{4n}\right)^{2m} \left(\frac{n-M}{n}\right)^p. \tag{10}$$

By the binomial formula we obtain

$$Q_4 = q'^{M-2m} q_0'^{n-M} \left(1+\frac{q_0}{q_0'}\frac{n-M}{n}\exp\frac{\varepsilon_3}{kT}\right)^{2(r-m)} \left(\frac{n}{n-M}\right)^n \left(\frac{4(n-M)}{M}\right)^M \left(\frac{M}{4n}\right)^{2m}. \tag{11}$$

Finally we obtain

$$Q = \sum_{r=0}^{N} K e^{Dr} J, \tag{12}$$

where

$$K = \left(\frac{n}{n-M}\right)\left(\frac{4(n-M)}{M}\right)^M q_0^{n-m} q'^M q^{2N} \exp\frac{2\varepsilon_1(N-1)+N\varepsilon_2}{kT}, \tag{13}$$

$$J = \sum_{\{\sigma_i\}} \exp\left[\sum_{j=1}^{r}\left(A\sigma + B\sigma_j^2\sigma_{j+1}^2 + C\sigma_j^2\right)\right], \tag{14}$$

$$D = \frac{2\varepsilon_1+\varepsilon_2}{kT} + 2\ln\left(\frac{q}{q'}\frac{M}{4n}\right), \tag{15}$$

$$A = \frac{\varepsilon_2}{kT} + \ln\left(\frac{q}{q'}\frac{M}{4n}\right) + \ln 3 - \ln\left(1+\frac{q_0}{q_0'}\frac{n-M}{n}\exp\frac{\varepsilon_3}{kT}\right) - \ln Z, \tag{16}$$

$$B = \frac{2\varepsilon_1}{kT}, \tag{17}$$

$$C = -\frac{4\varepsilon_1 + \varepsilon_2}{kT} - \ln\left(\frac{q}{q'}\frac{M}{4n}\right) + \ln 3 + \ln\left(1 + \frac{q_0}{q_0'}\frac{n-M}{n}\exp\frac{\varepsilon_3}{kT}\right) + \ln Z, \tag{18}$$

J is computed by the Ising method (4). J can be represented as the trace of a certain matrix Λ^r. The elements of the matrix Λ in our case have the form

$$\Lambda_{\sigma\sigma'} = \exp\left(A\sigma' + B\sigma^2\sigma{a'}^2 + C\sigma{a'}^2\right), \tag{19}$$

where $\sigma, \sigma' = 0, +1, -1$. The trace of the matrix Λ^r is equal to the r-th power of the largest root $\lambda_{\max}$ of the characteristic equation

$$|\Lambda - I\lambda| = 0, \tag{20}$$

I is the identity matrix. Equation (20) has the form

$$\lambda^3 - \lambda^2(1 + 2e^{B+C}\operatorname{ch} A) + 2\lambda\operatorname{ch} A\,(e^{B+C} - e^C) = 0. \tag{21}$$

We have

$$Q = K\sum_{r=0}^{N}\left(e^D\lambda_{\max}\right)^r = K\frac{\left(e^D\lambda_{\max}\right)^{N+1} - 1}{e^D\lambda_{\max} - 1}. \tag{22}$$

For fixed r, the mean number of attached "suitable" nucleotides is equal to

$$\bar{l} = r - r\frac{\partial\ln\lambda_{\max}}{\partial C}. \tag{23}$$

The mean value of r

$$\bar{r} = \frac{\partial\ln Q}{\partial D} = \frac{Ne^{D(N+2)}\lambda_{\max}^{N+2} - (N-1)e^{D(N+1)}\lambda_{\max}^{N+1} + e^D\lambda_{\max}}{e^{D(N+2)}\lambda_{\max}^{N+2} - e^{D(N+1)}\lambda_{\max}^{N+1} - e^D\lambda_{\max} + 1}. \tag{24}$$

Since ε_1 is of the order of 10 kcal/mole, $e^B \gg 1$, and ε_2 is of the order of 4 kcal/mole. Then the roots of equation (21) have the form

$$\lambda_1 = 0, \quad \lambda_2 \simeq 1, \quad \lambda_3 \simeq e^{B+C+A} + e^{B+C-A},$$

where $B + C + A < 0$, $|B + C + A| \gg 1$.

Let us consider the following cases:

1) $B+C-A<0$. Then $\lambda_{\text{max}}=\lambda_2$ and $\bar{l}=r$. If $D<0$, then $\bar{r}\simeq 0$; if $D>0$, then $\bar{r}\simeq N$. Thus, when D passes through zero, reduplication occurs;

2) $B+C-A>0$. Then $\lambda_{\text{max}}=\lambda_3$, $\bar{l}=0$ and $\bar{m}=0$. If $B+C-A+D<0$, then $\bar{r}\simeq 0$ and the double helix is preserved; but if $B+C-A+D>0$, then $\bar{r}\simeq N$, and the double helix separates into two solvated chains that coil into globules (factor $\ln Z$), i.e., denaturation occurs. Denaturation occurs at $B+C-A+D=0$.

3) $B+C-A=0$. In this case $\lambda_2\simeq\lambda_3$, and both roots must be used. We have

$$\bar{l}=r-\frac{\partial\ln(\lambda_2^r+\lambda_3^r)}{\partial C}=\frac{r}{2}. \tag{25}$$

If $D<0$, then $\bar{r}\simeq 0$; if $D>0$, then $\bar{r}\simeq N$, $\bar{l}=N/2$, i.e., half the links of the chain have adsorbed nucleotides and half have adsorbed water molecules. For $D>0$ a sufficiently small deviation of $B+C-A$ from 0 is enough for reduplication or denaturation to occur.

Both transitions are the sharper, the larger N is (1). We see that the concentration of nucleotides necessary for reduplication is 4 times greater than that found without taking into account the specificity of nucleotide attachment (1). The fate of DNA depends on the flexibility of the chain and on the difference between the solvation and polymerization energies. At large Z and ε_3, not reduplication but denaturation is possible. In the future, it will be necessary in the calculations also to take into account electrostatic interactions within the chain and of DNA with dissolved ions.

Institute of High-Molecular Compounds
Academy of Sciences of the USSR

Received
5 XI 1959

CITED LITERATURE

1. M. V. Vol' kenshtein, DAN, **130**, No. 3 (1960).
2. M. V. Vol' kenshtein, *Configurational Statistics of Polymer Chains*, Publishing House of the Academy of Sciences of the USSR, Moscow—Leningrad, 1959, § 31.
3. J. H. Gibbs, E. A. Di Marzio, *J. Chem. Phys.*, **30**, 271 (1959).
4. G. Newell, E. Montroll, *Rev. Mod. Phys.*, **25**, 353 (1953).

Note: Figure translations are in progress. See original paper for figures.