Biophysical Chemistry 18 (1983) 11–13
Elsevier

BPC 00776

THE EVALUATION OF AN AVERAGE NUCLEOTIDE COMPOSITION FROM MELTING CURVES

R.R. VARDAPETJAN [a], K. LETNANSKY [a], P.O. VARDEVANJAN [b] and G.H. PANOSJAN [b]

[a] *Institute for Cancer Research, University of Vienna, Vienna, Austria, and* [b] *Department of Biophysics, Faculty of Biology, State University of Erevan, Erevan, U.S.S.R.*

Received 20th October 1982
Revised manuscript received 7th February 1983
Accepted 9th March 1983

Key words: Thermal denaturation; G + C content; DNA

A new formula has been derived for the calculation of the average G+C content $\bar{X}$ of DNAs from different origins using thermal melting data. As compared to existing formulas the new method gives highly accurate results, although being much easier to use than similar equations.

1. Introduction

The average G + C content ($\bar{x}$) of the DNA of a given organism is one of the most important taxonomic features which is determined by means of chemical analyses of hydrolysates [1].

Essential physical methods of $\bar{x}$ identification differ in precision and simplicity of performance. One of the commonly used methods of calculating $\bar{x}$ is a procedure based on the determination of the melting temperature (T_m) of DNA [2]:

$$\bar{x} = \frac{T_m - T_{AT}}{T_{GC} - T_{AT}} \quad (1)$$

where T_{AT} and T_{GC} are the melting temperatures of poly(AT) and poly(GC). T_m is the temperature characteristic for 50% helicity; it is defined by $\theta(T_m) = 0.5$ for a DNA with a quasi-random sequence of nucleotides, where $\theta(T)$ is the degree of helicity of the DNA. If this method of determining T_m for DNAs of different origin is applied, the value of $\bar{x}$ defined by eq. 1 gives different results compared to the chemical analyses of hydrolysates for DNA with block heterogeneity: All natural DNAs may be devided into two types, quasi-random (DNAs of phage T2, T7, SD, T4, etc.) and DNAs with block heterogeneity (DNAs from bacteria, plants and higher animals). DNA with block heterogeneity contains alternating regions, within which the sequence of base-pairs is quasi-random. The observed melting curve for DNAs with block heterogeneity is the sum of the melting curves of particular blocks. Direct conformation of the block structure of DNA is given in refs. [3–7]. This leads to a discrepancy between the values of $\bar{x}$ obtained by the chemical method and by eq. 1.

Taking into consideration the block structure of DNA, a special method of defining $\bar{x}$ was worked out:

$$\bar{x} = \int_0^1 x\rho(x)\,dx \quad (2)$$

where $\rho(x)$ is the density function of distribution of a DNA in which blocks with different G + C composition are linked [8].

For making use of eq. 2 it is necessary to know the analytical expression of the function $\rho(x)$, the calculation of which requires the use of special methematical procedures.

In this communication a simple method of defining $\bar{x}$ is suggested, irrespective of the type of distribution of nucleotides, based on calculations

of the area under the melting curve of DNA. We believe that the application of this method results in more accurate values as compared to those obtained with the simple formula, eq. 1, although complicated calculations, like those based on eq. 2, can be avoided.

2. Methods

DNA melting curves were performed by thermal denaturation in a Unicam SP8–100 spectrophotometer at concentrations of 20–30 mg/ml in 15 mM NaCl, 1.5 mM sodium citrate using 1 cm quartz cells with Teflon stoppers. Heating was applied in a SP876 2 temperature programmer at a linear rate of 0.25°C/min. Absorbance at 260 nm was registered simultaneously on an HP-97S programmable calculator; the frequency of absorbance registration was 6 readings/min. The calculation and construction of curves were carried out by an EC 1033 computer (U.S.S.R). In 15 mM NaCl, 1.5 mM sodium citrate, $T_{GC} - T_{AT} = 42.38$ [10]. The protein content of DNA preparations never exceeded 0.5%.

3. Results and discussion

As for the blocks in DNA, the condition $\theta(T_m) = 0.5$ results in considerable distortion for determining T_m. In general, it is necessary to use the first moment of the function $\theta'(T)$

$$T_m = \int_{T_{AT}}^{T_{GC}} -\frac{\partial\theta}{\partial T} T\,dT \qquad (3)$$

where the DNA with quasi-random distribution coincides with the meaning of T_m defined from the condition $\theta(T_m) = 0.5$.

Let us take an example on the basis of which the definition of T_m according to eq. 3 results in a certain average temperature characterizing the process of melting of DNA in such a manner that it does not depend on the type of distribution of nucleotides along the DNA molecules. If we represent the function $\theta'(T)$ as the sum of Gaussian components and extend the integration from $-\infty$ up to $+\infty$ (which does not influence the result as the integral equals 0 beyond the limits T_{AT} and T_{GC}), we obtain

$$T_m = -\int_{-\infty}^{\infty} T\theta'(T)\,dT = -\int_{-\infty}^{\infty} T\sum_{l=1}^{n}\theta'_{l\tau}(T)\,dT$$

$$= -\sum_{l=1}^{n}\int_{-\infty}^{\infty} T\theta_{l\tau}(T)\,dT$$

$$= \sum_{l=1}^{n}\int_{-\infty}^{\infty} T\frac{S_l}{\Delta\tau_l}\exp\left[-\frac{\pi(T-T_l)^2}{\Delta\tau_l^2}\right]dT$$

where S_l, $\Delta\tau_l$ and T_l are area, width of the temperature interval of transition and position of the constituent on the temperature scale, respectively. Taking into consideration that

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \pi^{1/2}$$

we get

$$T_m = \sum_l T_l S_l$$

If nucleotides are distributed quasi-randomly we obtain $l = 1$ and we are able to pass on to the definition of T_m, according to the condition $\theta(T_m) = 0.5$

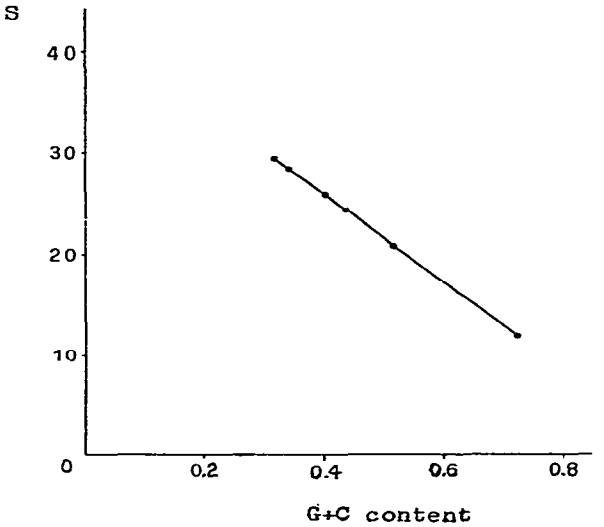

Fig. 1. Dependence of S (ordinate) in eq. 7 from the G+C content $\bar{x}$ (abscissa) of different DNA species as listed in table 1.

Table 1

Comparison of the G+C content of the DNA from different sources as obtained by different methods

Values are expressed as per cent of total nucleotide content.

Source of DNA	% G+C from melting curve and:			%G+C from chemical analysis [11–14]
	eq. 1	eq. 2	eq. 6	
Clostridium perfringens	31.0	31.5	31.4	31.3±1.0
Phage T2	32.4	34.2	34.5	34.4±0.4
Rat liver	38.7	39.5	39.9	40.0±0.4
Calf thymus	42.0	43.4	43.3	43.3±0.5
Escherichia coli	52.5	51.5	51.3	51.3±0.6
Micrococcus lysodeikticus	72.0	72.5	72.2	72.3±0.6

Transforming eq. 3, we take into consideration the fact that after integration we have the following expression

$$\frac{\partial\theta}{\partial T}\mathrm{d}T=\mathrm{d}\theta,\quad T_{\mathrm{m}}=T_{\mathrm{GC}}-\int_{T_{\mathrm{AT}}}^{T_{\mathrm{GC}}}(1-\theta)\mathrm{d}T \tag{4}$$

(The integral will be replaced by S below).
We may substitute the expression T_{m} from eq. 4 into eq. 1 and get

$$\bar{x}-1-\frac{S}{T_{\mathrm{GC}}-T_{\mathrm{AT}}} \tag{5}$$

S is numerically equal to the area formed by the DNA melting curve with the temperature axis and the vertical line at T_{GC}. To sum up:

$$\bar{x}=1-\frac{S}{T_{\mathrm{GC}}-T_{\mathrm{AT}}}\quad S=\int_{T_{\mathrm{AT}}}^{T_{\mathrm{GC}}}(1-\theta)\mathrm{d}T \tag{6}$$

It is noteworthy that the parameter S is independent of the shape of the DNA melting curve.

Experimental melting data of DNAs with different G + C composition also indicate that there is a linear dependence between $\bar{x}$ and S [9] (fig. 1)

$$\bar{x}=\alpha S+\beta \tag{7}$$

where again S is the area under the melting curve.

Comparison of eqs. 7 and 5 shows that

$$\alpha=-\frac{1}{T_{\mathrm{GC}}-T_{\mathrm{AT}}};\quad \beta=1$$

Proceeding from eq. 6, values of $\bar{x}$ were calculated for DNAs with block and quasi-random distributions of nucleotides (G + C content from 0.21 to 0.72) and are shown in table 1. This table further shows values of $\bar{x}$, obtained by chemical analyses of hydrolysates and a comparison of these data with those obtained by melting experiments and calculation with eqs. 1, 2, and 6, respectively. In these experiments the standard deviation never exceeded ±0.2. It is demonstrated that values calculated with eq. 6 correspond equally well to the chemical analyses as compared to those obtained with eq. 2. As to eq. 1, there is an even better correspondence.

References

1 B.F. Vanjushin, S.G. Tkacheva and A.N Belozersky, Nature 225 (1970) 948.
2 J. Marmur and P. Doty, J. Mol. Biol. 5 (1962) 109.
3 A.V. Shugalii, M.D. Frank Kamenetskii and Yu.S. Lazurkin, Mol. Biol. 4 (1970) 275.
4 A.V. Shugalii, M.D. Frank Kamenetskii and Yu.S. Lazurkin, Mol. Biol. 5 (1971) 766.
5 A. Skalka, E. Burgi and A.D. Hersbez, J. Mol. Biol. 34 (1968) 1.
6 T. Huguet and L. Jouanin, Biochim. Biophys. Acta 262 (1972) 431.
7 Yu.L. Lyubchenko, M.D. Frank Kamenetskii, A.V. Vologodskii, Yu.S. Lazurkin and G.G. Gause, Biopolymers 15 (1976) 1019.
8 A.A. Achrem, V.M. Aslanjan, Yu.S. Babayan and D.J. Lando, (1980) Dokl. Akad. Nauk B.S.S.R. 24, n 3, 264.
9 R.R. Vardapetjan, P.O Vardevanjan, G.H. Panosjan and A.T. Karapetjan, Stud. Biophys., 91, (1982) 237.
10 B.N. Belinzev, A.V. Vologodskii and M.D. Frank Kamenetskii, Mol. Biol. (U.S.S.R.) 4, (1976) 764.
11 A.S. Antonov, The role of DNA Structure in the classification of organisms (Nauka, Moscow, 1972).
12 C.L. Schildkraut, J. Marmur and P. Doty, J. Mol. Biol. 4 (1962) 430.
13 R.L. Sinsheimer, in: The Nucleic Acids, vol. 3, eds. E. Chargaff and J.N. Davidson (Academic Press, New York, 1960), p. 187
14 W. Szybalski, R.I. Erikson, G.A. Gentry, L.G. Gafford and C.C. Randall, Virology 19 (1963) 586.