High resolution thermal denaturation of mammalian DNAs

T.Guttmann, A.Vítek and L.Pivec

Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences,
166 10 Prague 6, Flemingovo nám. 2, Czechoslovakia

Received 4 October 1976

ABSTRACT

High resolution melting profiles of different mammalian DNAs are presented. Melting curves of various mammalian DNAs were compared with respect to the degree of asymmetry, first moment, transition breadth and T_{mi} of individual subtransitions. Quantitative comparison of the shape of all melting curves was made. Correlation between phylogenetical relations among mammals and shape of the melting profiles of their DNAs was demonstrated. The difference between multicomponent heterogeneity of mammalian DNAs found by optical melting analysis and sedimentation in CsCl-netropsin density gradient is also discussed.

INTRODUCTION

Recent theoretical^{1,2} and experimental investigations have shown that thermal resolution data may provide a detailed qualitative and quantitative information on compositional heterogeneity of DNAs as to the level of molecular segments representing individual genes³⁻⁵. These investigations are in accordance with our earlier denaturation experiments which indicated plurimodal base composition heterogeneity in DNAs of various origin⁶⁻⁸. Because of the high complexity of naturally occurring DNAs we have developed a computerized method for the analysis of the pluriphase absorption melting profiles⁹. Our analysis is based on the use of the statistical criteria to distinguish between artificial and physically significant derivative subtransitions. This analysis was used in this paper for the study of the mutual relations in the multicomponent base composition heterogeneity among DNAs of various mammalian species.

MATERIAL AND METHODS

Liver nuclei were prepared from the frozen tissue as previously described¹⁰. DNAs from mammalian cell nuclei were isolated according to Marmur¹¹. Molecular weights of all mammalian DNAs were in range of $7-19 \cdot 10^6$ and were presented in our previous paper¹².

Absorption melting curves of DNAs were measured in $0.1 \times$ SSC ($0.15M$ NaCl + $0.015M$ trisodium citrate) + $10^{-4}M$ EDTA as described previously⁹. The concentration of DNA was $15 \mu g/ml$. Absorbancy values were not corrected for changes of the solution volume due to heating. Measurements of the melting curves of all DNAs were repeated 2-4 times. Experimental points were fitted by a cubic spline function and derivated as previously described⁹. Quantitative comparison of the melting curves of different DNAs was carried out at the confidence level of smoothing $\phi(k) = 0.68$ (ref.⁹). At this level, the reproducibility of dH/dT values (expressed as maximum difference between two measurements) is equal to $0.6 k^{-1}$ and is designated in Fig. 1 as the error bar. (For the lower probability level $\phi(k) = 0.50$ see ref.⁸).

For the estimation of the weighting factor used for calculation of residual sum of squares in the fitting process the standard deviation of the absorbance measurement σ_A and temperature reading σ_T was 0.0005 and $0.04^\circ C$, respectively. However, these values were obtained from the dispersion of both these quantities in the stationary measurements. Therefore an additional test of random distribution of positive and negative deviations of experimental points from the spline function was performed for different values of σ_A and σ_T . As a criterion the number of iterations (i.e. number of point, where differences change their sign) was used. The expected number of iterations for most random distribution¹³ is equal to

$$E(i) = 2mn/(m+n) + 1, \quad (1)$$

where m and n are the number of positive and negative differences, respectively. The most random distribution of $\pm$ deviations was in each case very close to the above mentioned values of σ_A and σ_T . This result justifies the adoption of

these values also for our non-stationary experimental conditions (the rate of temperature increase was 0.1 to 0.2°/min).

From the absorption melting profiles of DNAs were determined the T_m , $\bar{\epsilon}_1$ and $\bar{\epsilon}_2$ values where $T_m = T_{0.5}$, $\bar{\epsilon}_2 = T_{0.75} - T_{0.50}$ and $\bar{\epsilon}_1 = T_{0.5} - T_{0.25}$. The symbol T with index x denotes the temperature at which the normalized hyperchromicity equals x . First moment ($\bar{T}$) was used for the determination of the mean G+C content of DNAs^{6,29}. The values of G+C content were corrected with respect to the difference of the contribution to hyperchromicity from (dA.dT) base pairs and (dG.dC) base pairs using data of Felsenfeld et al.¹⁴.

To characterize the difference between the melting curves of any two DNAs p_{ij} parameter was calculated:

$$p_{ij} = \int_{60}^{85} |H_i(T) - H_j(T)| dT \quad (2)$$

where $H_i(T)$ is normalized hyperchromicity-temperature dependence defined by i -th melting curve. The integral in equation (2) is equal to area confined by i -th and j -th melting curves. For two measurements of the same DNA the parameter p_{ii} should be equal to zero. However, the experimental error in the determination of $H_i(T)$ causes that p_{ii} ranges from 0.05 to 0.15. These values may be used as an estimate of uncertainty in the determination of the p_{ij} parameter which ranges from 0.25 - 1.51.

RESULTS

Derivative absorption melting curves of different mammalian DNAs are shown in Fig. 1. In all DNAs investigated the presence of 7-10 maxima and shoulders on each melting curve indicate a high degree of compositional heterogeneity. The T_m values and the magnitude of subtransitions characterize DNA of each species. All derivative melting curves are asymmetric and are elongated toward higher temperatures. This asymmetry is quantitatively expressed by the difference in $\bar{T} - T_m$ values and by the ratio $\bar{\epsilon}_2/\bar{\epsilon}_1$ (Tab. 1).

Two full lines in Fig. 2a represent the area in which the melting curves of all DNAs investigated, except the DNA

Fig.1 - First derivatives of absorption melting curves of different mammalian DNAs. The parallel derivative curves are shifted vertically from each other by 4.5 units of the scale on the ordinate (see melting curve on the top). Short vertical bars on each melting curve represent the T_m value of resolved bands and shoulders.

of guinea pig, are situated (dotted line). The shapes of the melting curves of different mammalian DNAs are more similar to each other in the first half of the temperature interval of the melting and markedly differ at temperatures higher than T_m (70°C). This difference reaches maximum at approximately 74°C , which corresponds to G+C content of 48%. The DNA of guinea pig differs from the group of other mammalian DNAs by a relatively high percentage of lower melting subbands. These differences are more illustrative in the plot $H(T)_{\max} - H(T)_{\min}$ against temperature (Fig. 2b); $H(T)_{\max}$ and $H(T)_{\min}$ represent the highest and lowest values of hyperchromicity $H(T)$, respectively, among melting curves of all DNAs at temperature T .

Table 1. Thermal denaturation characteristics of mammalian DNA

Source of DNA	T_m	T	$G+C^a$ (%)	σ_1	σ_2	σ_2/σ_1
Mouse	69.7	70.0	40.3	1.8	2.6	1.4
Rat	69.4	69.6	39.4	2.1	2.5	1.2
Guinea pig	69.0	69.5	39.5	2.4	3.2	1.3
Rabbit	70.0	70.8	42.3	3.0	4.2	1.4
Hedgehog	70.1	70.3	41.0	2.4	2.5	1.0
Human	70.0	70.6	41.6	2.5	3.1	1.2
Makakus rhesus	69.7	70.3	41.1	2.2	3.2	1.5
Dog	69.8	70.6	41.8	2.4	3.4	1.4
Cat	69.8	70.6	41.9	2.5	3.8	1.5
Roe deer	69.7	70.4	40.9	2.3	3.2	1.4
Calf	69.9	70.5	41.5	3.0	3.8	1.3
Pig	69.8	70.6	41.7	2.8	3.6	1.3

^a Calculated from the $\bar{T}$ values²⁹.

Fig.2 - a) The plot of extreme courses of the integral melting curves of mammalian DNAs. The solid lines represent limits of lowest temperature and highest temperature parts of the melting curves using all mammalian DNAs investigated except DNA of guinea pig (dotted line). b) The plot $H(T)_{max} - H(T)_{min}$ against temperature. Dashed line represents the plot of $H(T)_{max}$ of guinea pig (dotted) - $H(T)_{min}$ of other DNAs.

Melting curves of different DNAs considered as functions $H(T)$ were quantitatively compared using p_{ij} parameter (see Material and methods). The p_{ij} parameter, characterizing degree of dissimilarity between i -th and j -th curves, was calculated for every two curves. All combinations among 12 melting curves represent 66 values of p_{ij} (Tab.2). These relations implied in the data of Tab.2 become more evident if the data are presented graphically as a phylogenetic tree. The statistically optimal phylogenetic tree (Fig.3) was calculated according to Fitch and Margoliash¹⁵ using programable calculator Hewlett-Packard 9830A. The sum of the number values of individual tree branches between any pair of mammalian DNAs is equal to the parameter $\bar{p}_{ij}$ (Tab.2) which reflects their degree of phylogenetic relation. The main three branches (from the top) indicate phylogenetic separation of representatives of the order rodentia, second group of orders carnivora, ungulates and primates and third group lagomorpha.

Fig.3 - Phylogenetic tree calculated from the data in Tab. 2.
 $\bar{p}_{ij}$ - mean value of the parameter p_{ij} .

Table 2. The parameter p_{ij} (upper triangle) and recalculated¹⁵ mean parameter $\bar{p}_{ij}$ (lower triangle) of mammalian DNAs. Both p_{ij} and $\bar{p}_{ij}$ values $\times 10^2$

No.	Source of DNA	1	2	3	4	5	6	7	8	9	10	11	12
1	Mouse	-	62	80	151	43	77	84	87	109	92	131	116
2	Rat	85	-	38	137	76	95	54	90	106	97	104	93
3	Guinea pig	88	37	-	137	81	98	58	92	108	99	87	79
4	Rabbit	146	131	134	-	112	77	86	71	47	70	65	62
5	Hedgehog	42	61	82	122	-	44	61	60	76	60	101	85
6	Human	77	82	85	87	73	-	51	28	37	37	63	55
7	Makakus rhesus	92	77	80	82	68	49	-	42	58	48	63	51
8	Dog	93	77	81	83	69	28	45	-	28	32	54	41
9	Cat	100	85	88	90	76	35	52	27	-	37	37	34
10	Roe deer	99	84	87	89	75	36	51	32	39	-	59	46
11	Calf	111	96	99	98	87	62	57	58	65	64	-	25
12	Pig	100	85	88	87	76	51	46	47	54	53	25	-

Mammalian DNAs dealt with in this paper were analysed in CsCl-netropsin density gradients (recently published¹²). This type of density gradient is very sensitive to detection of minor components, about 2.5 times more than simple CsCl density gradient. The CsCl-netropsin density gradients and melting profiles of DNAs represent the G+C distribution of these DNAs when G+C content is assigned to temperature or buoyant density of sedimenting and melting segments of DNA respectively. The width of G+C distribution calculated from melting data was greater than that calculated from CsCl-netropsin density gradient profiles. This was found in all DNAs concerned and is shown in two representatives of mammalian DNAs in Fig.4. However, a significant portion of the melting subbands in the G+C rich region of the distribution represent denaturation of the heavy satellite components of mammalian DNAs. This follows from the comparison of the melting profile of calf total DNA and composed melting profile of the same DNA calculated as a superposition of the individual melting curves of calf satellites and the main band DNA component previously published by Filipaki et al.¹⁶ (Fig.4b). For calculation of the relative percentage of each satellite, width and T_m value of corresponding melting curve were used. There is a good accordance between these two melting curves.

Fig.4 - The distribution of the G+C content of DNAs calculated from absorption melting curves (a,b; ref.²⁹) and sedimentation in CsCl-netropsin in density gradient (a',b'; ref.¹²). a, a' - Makakus thesis DNA; b, b' - calf thymus DNA. Dashed line in Fig.4b represents composed melting profile calculated according to data of Filipski et al.¹⁶ (see text). The vertical dashed lines represent the G+C content of individual melting and buoyant density components (designated by the number values).

In all mammalian DNAs investigated, except those of mouse and man, there are no satellite components detected at lower densities than the main band density area. However, the thermal denaturation reveals the presence of more components melting in the low temperature region of the melting curves. The histogram in Fig. 5 represents the G+C content of discrete components of mammalian DNAs found in CsCl-netropsin profiles¹²

Fig.5 - The distribution of the G+C contents of melting components of mammalian DNAs obtained from Fig.1, a), and components resolved on CsCl-netropsin density gradients recently published¹² b). The dashed lines represent the highest components. G+C values calculated from Tm1, see ref.²⁹.

(Fig.5b) or absorption melting curves (Fig.5a). The G+C values of the highest melting components of different mammalian DNAs are in the range between 35-41% G+C which corresponds to the range of the G+C content of the main band components in CsCl-netropsin density gradients. It seems probable that these components are not distributed randomly in the mentioned interval of the G+C content, i.e. the components having similar G+C values are observed in different mammalian DNAs.

DISCUSSION

Results of the analysis of the absorption melting curves show that all nuclear mammalian DNAs investigated consist of minimally eight melting subbands representing the molecular segments of DNA differing in their base composition. The number of melting components would probably further increase if more sensitive methods of measurement of absorption melting curve were used. The range of the G+C content of melting components of mammalian DNAs represents (practically) the whole range of the G+C content which was found in naturally occurring DNAs. These melting components may be divided into

three main groups. The first group includes the biggest components near the T_m value ($67-70^{\circ}\text{C}$, i.e. 35-41% G+C), the second and third groups include the components melting at low and high temperatures, respectively. The variation of the base composition heterogeneity among different mammalian DNAs is not the same in the three categories of DNA components mentioned.

The similarity in the physical properties of the melting components in the first two categories among various DNAs indicates that these components represent probably the most conservative sequences of mammalian genomes. This set of DNA sequences form relatively short intramolecular segments of DNA molecules and reflect the compositional complexity of the non-satellite "main band" DNA recently described in mammalian DNAs^{17,18}. It may be suggested that this region contains an interspersion of short repetitive and unique DNA sequences. On the other hand, the highly repeated sequences which form typical satellites appear in these two categories of DNA sequences relatively rarely and were detected only in mouse and man DNAs.

The third category of the components melting at high temperatures represents denaturation of the G+C rich short intramolecular segments of DNA and predominantly the CsCl heavy satellite components of mammalian DNAs. It was previously found in several mammalian DNAs that some of these satellites consist of repeated sequences^{19,20}. The sequence complexity of the most newly detected components is yet unknown. However, in several ungulate DNAs increases this complexity with the decrease of buoyant density of the satellite components towards the density of the main band^{21,22}. The DNA components appearing in the range of the maximal compositional variation among different DNAs contain probably highly repeated sequences, which are not generally transcribed for structural proteins. Low natural selection pressure on these sequences can explain their mentioned compositional variability²³ and possible regulatory functions on the level of the chromosome topology²⁴.

The G+C distribution calculated from the melting curves

and our recently published buoyant density data in CsCl-netropsin gradients have in both cases a multicomponent character. Partial differences between these distributions may be explained by several effects. First: The higher number of resolved components obtained from melting data is due to higher sensitivity of the analysis of absorption melting curves to distinguish between base composition of two molecular segments of DNA. Second: The greater widths of the "melting distribution" are affected by much shorter size of the cooperatively melting molecular segments of DNA (600 base pairs¹) than have DNA molecules forming CsCl-netropsin density gradient components. In our group of mammalian DNAs all intramolecular melting segments of DNA (appearing in the first two categories of the components) are shorter than about 10^4 nucleotide pairs (ref.¹², Tab.1). Third: Some DNAs and especially satellite components do not fit exactly the empirical equations between temperature of melting or buoyant density on the one side and the mean G+C content on the other side. The deviations from these equations are supposed to be sequence specific and are very probably responsible for the differences in G+C content calculated from the mentioned physical methods used in our papers and G+C content obtained by direct chemical analysis¹⁶.

The quantitative determination of the degree of relationship between two organisms may be expressed as a sum of the similar nucleotide sequences of their DNAs²⁵. A comparison of the p_{ij} parameter among DNAs of different mammals shows that the representatives of a certain order are approximately situated in the same limited area of the phylogenetic tree. Mutual distances between these areas roughly reflect a phylogenetic relation among mammalian orders^{26,27}. However, there exist p_{ij} values widely spread among representatives of one order (mouse-guinea pig). It may be suggested that this phenomena have been related to the age of the considered order and to the number of its surviving suborders. The highest distance between two mammals in Fig. 3 is that between the representatives of rodents and rabbit. The rabbit has been classified as a rodent before paleontology found evidence

of the phylogenetic difference between rabbit and Rodentia. On the basis of these facts a new order, Lagomorpha, was established²⁶. Our results also strongly support this new differentiation.

The heterogeneity of the main CsCl band was previously found in several mammalian DNAs¹⁶⁻¹⁸. Our results also suggest the multicomponent character of the G+C distribution in the range of the biggest melting components and CsCl-netropsin main band components. Considering that the G+C content of the highest component of DNAs (which is usually close to the mean G+C value) may be regarded as an important taxonomic value²⁸, general polymodality in these groups of components seems to be a factor which should be taken into consideration for classification of mammalian DNAs. The existence of three distinct G+C values (see Fig.5), around which the highest components or their contiguous satellites (or melting components) in DNAs of different mammalian species are variably distributed, indicates the presence of several new mammalian subclasses. Other criteria, mainly the sequence complexity and the repetitiveness, should also be taken into account in a more precise characterization of these components. Further investigations in this field probably could bring into proper correlation the taxonomic classification of mammals on the basis of analysis of the compositional heterogeneity of their DNAs.

ACKNOWLEDGEMENTS

We are grateful to all persons who helped us in providing mammalian livers, especially to Dr J. Jirásek, Dr L. Šimek and Dr J. Lilák. Our thanks are due to Dr V. Hanák for valuable discussions, as well as to Mrs. M. Vítěčková for excellent technical assistance.

REFERENCES

- 1 Crothers, D.M. (1968) Biopolymers 6, 1391-1404
- 2 Wartell, R.M. and Montroll, E.W. (1972) Advances in Chem. Phys. Vol. 22 (John Wiley & Sons)
- 3 Michel, F. (1974) J.Mol.Biol. 89, 305-326
- 4 Vizard, D.L. and Ansevin, A.T. (1976) Biochemistry 15, 741-750

-
- 5 Karrer, K.M. and Gall, J.G. (1976) *J.Mol.Biol.* 104, 421-453
 - 6 Pivec, L., Pivcová, H. and Šormová, Z. (1970) *Biochim. Biophys.Acta* 213, 343-351
 - 7 Pivec, L., Štokrová, J. and Šormová, Z. (1972) *Biochim. Biophys.Acta* 272, 179-190
 - 8 Pivec, L., Horská, K., Vítek, A. and Doskočil, J. (1974) *Biochim.Biophys.Acta* 340, 199-206
 - 9 Vítek, A., Reddy, C.R. and Pivec, L. (1974) *Biochim. Biophys.Acta* 353, 385-391
 - 10 Pivec, L. and Štokrová, J. (1971) *FEBS Lett.* 14, 157-160
 - 11 Marmur, J. (1961) *J.Mol.Biol.* 3, 208-218
 - 12 Guttman, T., Votavová, H. and Pivec, L. (1976) *Nucleic Acids Res.* 3, 835-845
 - 13 Swed, F.S., Eisenhart, C. (1943) *Ann.Math.Statistics* 14, 66-87
 - 14 Felsenfeld, G. and Hirschman, S.Z. (1965) *J.Mol.Biol.* 13, 407-427
 - 15 Fitch, W.M. and Margoliash, E. (1967) *Science* 155, 279-284
 - 16 Filipski, J., Thiery, J.P. and Bernardi, G. (1973) *J.Mol. Biol.* 80, 177-197
 - 17 Cech, T.R. and Hearst, J.E. (1976) *J.Mol.Biol.* 100, 227-256
 - 18 Marx, K.A., Allen, J.R. and Hearst, J.E. (1976) *Biochim. Biophys.Acta* 425, 129-147
 - 19 Flamm, W.G., Walker, P.M.B. (1969) *J.Mol.Biol.* 42, 447-455
 - 20 Curtain, C.C., Pascoe, G. and Hayman, R. (1973) *Biochem. Gen.* 10, 253-262
 - 21 Votavová, H. and Šponar, J. (1975) *Nucleic Acids Res.* 2, 185-196
 - 22 Pivec, L., Votavová, H., Forstová, J. and Guttman, T.: in preparation
 - 23 Smith, G.P. (1976) *Science* 191, 528-535
 - 24 Miklos, G.L.G. and Nankivell, R.N. (1976) *Chromosoma* 56, 143-167
 - 25 DeLey, J. (1969) *J.Theoret.Biol.* 22, 89-116
 - 26 Simpson, G.G. (1945) *Bull.Amer.Mus.Natl.Hist.* 85, 1-350
 - 27 Thenius, E. *Stammesgeschichte der Säugetiere*, Berlin 1969
 - 28 Sueoka, N. (1961) *J.Mol.Biol.* 3, 31-41
 - 29 Mandel, M., Igambi, L., Bergendahl, J., Dodson, M.L. and Scheltgen, E. (1970) *J.Bacteriol.* 101, 333-338