Volume 226, number 2, 247-249 FEB 05450 January 1988

Vestiges of lost introns in the thermal stability map of DNA

Ryo Hanai, Akira Suyama* and Akiyoshi Wada

Department of Physics, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Received 7 September 1987; revised version received 21 October 1987

The absence of introns in prokaryotic genes has been explained by intron loss on various bases. Here we report another piece of evidence on intron loss, which was found in the thermal stability map of DNA. We calculated the local melting temperature of cDNA sequences and found that (i) gaps in thermal stability tend to occur near intron positions with a statistical significance, and (ii) one-third of the gaps far from intron positions can be assigned to lost introns. From these results we conclude that the gaps of thermal stability in protein coding regions are the vestiges of lost introns.

Intron; Exon; DNA; Double-helix stability; Thermal stability map; Gene homostabilizing propensity

1. INTRODUCTION

The thermal stability map of DNA is a plot of the local melting temperature of the DNA double-helix along its base sequence (fig.1). So far we have found that the map has an inclination to be flat in each protein coding region: the gene homostabilizing propensity [1-4]. Even so, gaps are often found in long coding regions. In this study we tried to elucidate the gaps in terms of lost introns. A direct study in this context is to compare the thermal stability maps of genes without introns to the exon structures of homologous genes. However, few such sequence data are available. Therefore, we examined thermal stability maps of cDNA sequences with known exon structures.

Correspondence address: R. Hanai, Department of Physics, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

* Present address: Department of Mechanical Engineering, Technological University of Nagaoka, Nagaoka, Niigata 940-21, Japan

2. RESULTS AND DISCUSSION

Fig.1 shows thermal stability maps calculated by the algorithm of Poland-Fixman-Freire [5,6], with parameters $T_{AT} = 53.1°C$, $T_{GC} = 94.1°C$, $\Delta S = -24.54$ e.u., $\sigma_l = 6 \times 10^{-6}(l+450)^{-1.75}$, and $\beta C_0 = 10^{-7}$. The gaps are located near intron positions indicated by arrows. Moreover, the gap indicated by the asterisk in fig.1b coincides with the lost intron suggested by Gō [7].

We tested the statistical significance of this gap-intron correlation by analysis of 32 thermal stability maps consisting of 31 726 bp and 227 exons (not including flanks). 114 gaps (larger than 0.3°C) were observed, and classified according to their size and distance from the nearest (present) introns (table 1). The results show that larger gaps are apparently near introns. We considered only the gaps larger than 0.8°C and defined the gaps within 20 bp from introns as 'near'. Of the 78 gaps, 26 are near introns and 10 are near coding region boundaries. Because the 10 gaps are not directly related to introns, we discuss only the other 68 gaps. If the gaps are scattered randomly, the probability for a gap to appear near an intron is 0.28, for 40 bp in each exon are 'near'. Thus only 19.0 gaps with a

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

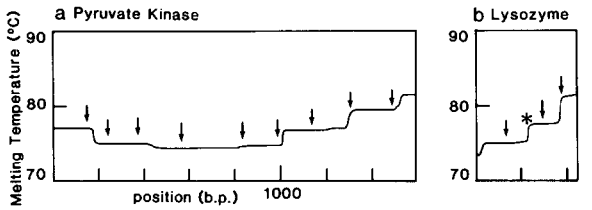

Fig.1. Thermal stability maps of cDNA sequences of (a) chicken pyruvate kinase and (b) chicken egg white lysozyme. Arrows indicate the positions of introns. The asterisk indicates the position of Gō's module junction [7].

standard deviation of 3.7 would be expected to appear near introns by binomial distribution. In fact, 26 are near introns; the deviation is statistically significant.

Though 42 remain unassigned to present introns, 15 gaps among 42 gaps were assigned to lost introns: 8 were near introns of homologous genes, 5 were near gaps in thermal stability maps of homologous genes, and 2 were near junctions of the modules of Gō (fig.1b, and personal comunication). Nevertheless, 27 gaps still remain, however, we believe that most of them should imply lost introns not yet identified.

Table 1

Classification of gaps of thermal stability in coding regions

Size of gaps (°C)	Distance to the nearest introns/ coding region boundaries (bp)				
	1-10	11-20	21-30	31-	Total
0.3-0.6	4	3	2	16	25
0.6-0.8	3	1	2	5	11
0.8-1.0	0	2	1	6	9
1.0-2.0	8 (1)	3 (1)	4	11	27 (2)
2.0-	12 (2)	11 (6)	4	16	43 (7)
Total	27 (3)	20 (7)	13	54	114 (10)

From the coding region of 32 thermal stability maps of cDNA sequences of various kinds of proteins. The size is the temperature difference of the flat regions forming a gap, and the distance is the number of base pairs from a gap to the nearest intron (or coding region boundary). Values in parentheses indicate the numbers of the gaps near coding region boundaries, for example 8(2) means that 2 of the 8 gaps are near coding region boundaries and 6 are near intron positions

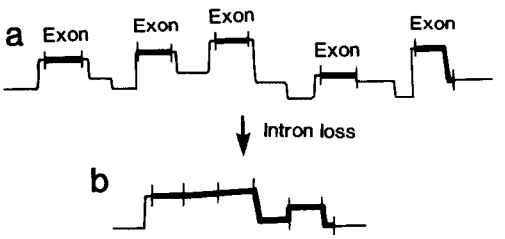

Fig.2. Model of the relationship of intron loss and gaps in thermal stability map of coding regions.

We also calculated maps of 4 yeast sequences originally lacking introns. Of the 15 gaps (larger than 0.8°C) found, 6 were near coding region boundaries, and 5 were assigned to lost introns.

From these results we conclude that the gaps in protein coding regions are the vestiges of lost introns. We explain this as follows. Our recent unpublished data show that exons have the third letter counterbalance tendency, a base-sequence characteristic of the homostabilizing propensity [3], and exons usually have higher G+C content and stability than introns in each gene [1,8]; the thermal stability map can be schematically depicted as in fig.2a. When introns are lost, gaps remain at the junctions of large stability differences (fig.2b). Conversely, gaps are not always found at lost intron positions.

Finally, we would like to point out that the present results cannot be obtained by direct analyses of DNA sequences. First, donor and acceptor consensus sequences [9] reduce to 5′-$\underset{G}{\underset{C}{}}$AG/G-3′ when introns are lost. This short sequence would occur too frequently to be a mark of lost introns. Second, the relationship between the base sequence and the thermal stability map is obscure [1,2].

ACKNOWLEDGEMENTS

We thank Dr Mitiko Gō for unpublished data and discussions. This work was partly supported by grants from the Ministry of Education, Science and Culture, Japan.

REFERENCES

[1] Wada, A. and Suyama, A. (1983) J. Phys. Soc. Jap. 52, 4417-4422.
[2] Wada, A. and Suyama, A. (1984) J. Biomol. Struct. Dyn. 2, 573-591.

[3] Wada, A. and Suyama, A. (1985) FEBS Lett. 188, 291–294.
[4] Wada, A. and Suyama, A. (1986) Prog. Biophys. Mol. Biol. 47, 113–157.
[5] Poland, D. (1974) Biopolymers 13, 1859–1871.
[6] Fixman, M. and Freire, J.J. (1977) Biopolymers 16, 2693–2704.
[7] Gō, M. (1983) Proc. Natl. Acad. Sci. USA 80, 1964–1968.
[8] Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Gerard, C., Meunier-Rotival, M. and Rodier, F. (1985) Science 228, 953–958.
[9] Mount, S.M. (1982) Nucleic Acids Res. 10, 459–472.