The Specificity of Five DNAases as Studied by the Analysis of 5'-Terminal Doublets

Alberto BERNARDI, Claire GAILLARD, and Giorgio BERNARDI

Laboratoire de Génétique Moléculaire, Institut de Biologie Moléculaire, Paris

The 5'-terminal dinucleotides released by five deoxyribonucleases (spleen acid DNase, snail acid DNase, pancreatic DNase, *Escherichia coli* endonuclease I and crab DNase) have been determined on *E. coli* DNA (51% dG + dC) digests having different average sizes ($\bar{P}_n$) in the range 50 to 10. It has been shown that the composition of the 5'-terminal dinucleotide (a) is independent upon the degradation level, at least in the range explored; (b) is strongly different from the composition of *E. coli* DNA doublets, these differences being characteristic for each enzyme; (c) is very significantly different from the statistical composition of 5'-terminal dinucleotides, as calculated from the composition of 5'-terminal and penultimate nucleotides. A calculation of the statistical composition of the trinucleotides split by each enzyme, using the 3'-terminal nucleotide data in conjunction with the 5'-terminal dinucleotide results provided a qualitative "specificity spectrum" for each enzyme.

Recent work from this laboratory has shown that deoxyribonucleases (DNAases) recognize and split specific sets of short oligonucleotide sequences in DNAs. The basis for such conclusion is given by the analysis of the nucleotides next to the breaks introduced by the enzymes. In fact, the base compositions of the 3' and 5'-terminal and penultimate nucleotides released by DNAases not only differ from the overall base compositions of the degraded DNAs, but also do not show, in general, any nearest-neighbor relationships with each other, that is to say that the base composition of the nucleotides present in a particular position (3' or 5'-terminal or penultimate) is not that predicted, on the basis of nearest-neighbor analysis data, from the composition of its neighbors. A brief review of our work on DNAase specificity and its use in assessing the frequency of the sequences recognized and split by DNAases has been published elsewhere [1].

Very recently, a rapid and sensitive method for the analysis of the 5'-terminal doublets has been set up in our laboratory [2], enabling us to investigate in more detail the specificity of the five DNAases currently available to us: acid DNAase from hog spleen [3], acid DNAase from the hepatopancreas of *Helix aspersa* (Müll.) [4], DNAase from bovine pancreas, endonuclease I from *Escherichia coli* [5] and the neutral DNAase from the testes of *Cancer pagurus* [6]. For brevity we will refer to these enzymes as spleen, snail, pancreatic, *E. coli* and crab DNAases.

In the present work, we have studied the kinetics of liberation of the 5'-terminal dinucleotides from *E. coli* DNA (the choice of this DNA is justified in the Discussion). We have compared the experimental data with doublet-frequency results and with the statistical expectations. In addition, we have calculated the composition of the trinucleotides formed by each enzyme, using the 3'-terminal nucleotide data in conjunction with the 5'-terminal dinucleotide compositions. The results obtained in these investigations not only confirm and extend our previous conclusion that DNAases split specific sets of short nucleotide sequences, but also open the way to a finer assessment of oligonucleotide frequencies in different DNAs.

MATERIALS AND METHODS

E. coli DNA was prepared as described elsewhere [8]. Spleen DNAase was obtained as described by

Table 1. *5'-Terminal dinucleotides released by five different DNases*
The 5'-terminal dinucleotides are written with the 5'-terminal nucleotide on the left

5'-Terminal dinucleotides	Doublet frequency	Spleen DNase	Snail DNase	Pancreatic DNase	E. coli DNase	Crab DNase
	%					
dA-dA	7.1	9.9 ± 0.17	4.8 ± 0.18	4.1 ± 0.32	3.2 ± 0.18	3.1 ± 0.09
dA-dC	5.4	2.0 ± 0.04	3.4 ± 0.16	4.5 ± 0.13	1.9 ± 0.07	2.8 ± 0.13
dA-dG	5.5	2.7 ± 0.27	2.0 ± 0.09	3.8 ± 0.21	3.5 ± 0.23	3.9 ± 0.45
dA-dT	6.8	2.1 ± 0.17	6.4 ± 0.38	2.5 ± 0.11	6.2 ± 0.55	14.2 ± 0.35
dC-dA	7.1	14.8 ± 0.50	5.3 ± 0.13	7.6 ± 0.12	6.2 ± 0.20	6.6 ± 0.46
dC-dC	5.6	5.3 ± 0.11	8.3 ± 0.20	7.2 ± 0.22	3.6 ± 0.34	5.9 ± 0.67
dC-dG	6.7	10.4 ± 0.45	4.9 ± 0.30	4.5 ± 0.29	6.2 ± 0.50	7.4 ± 0.72
dC-dT	5.5	5.2 ± 0.42	9.6 ± 0.26	2.3 ± 0.14	5.0 ± 0.31	16.7 ± 0.73
dG-dA	5.5	13.4 ± 0.81	5.1 ± 0.55	5.9 ± 0.51	8.9 ± 0.44	4.8 ± 0.72
dG-dC	8.3	5.9 ± 0.25	10.0 ± 0.21	10.0 ± 0.31	6.2 ± 0.32	6.0 ± 0.33
dG-dG	5.6	8.8 ± 0.30	9.2 ± 0.50	8.1 ± 0.38	9.9 ± 0.39	7.4 ± 0.55
dG-dT	5.5	5.5 ± 0.27	11.7 ± 0.30	3.0 ± 0.20	12.6 ± 0.65	17.4 ± 0.81
dT-dA	5.1	5.1 ± 0.25	4.4 ± 0.40	10.9 ± 0.31	8.0 ± 0.43	0.5 ± 0.06
dT-dC	5.6	2.2 ± 0.12	4.5 ± 0.09	8.3 ± 0.46	3.2 ± 0.46	0.7 ± 0.10
dT-dG	7.1	3.9 ± 0.17	3.6 ± 0.17	14.1 ± 0.35	11.7 ± 0.51	1.0 ± 0.10
dT-dT	7.6	2.8 ± 0.21	6.8 ± 0.32	3.2 ± 0.32	3.8 ± 0.49	0.7 ± 0.10
5'-Terminal nucleotides						
dA		16.7 ± 0.35	16.6 ± 0.70	14.8 ± 0.28	14.7 ± 0.74	24.0 ± 0.68
dC		35.7 ± 0.42	28.0 ± 0.59	21.6 ± 0.19	21.0 ± 0.79	36.6 ± 1.30
dG		33.6 ± 0.72	36.0 ± 0.31	27.1 ± 0.56	37.6 ± 0.80	35.7 ± 1.69
dT		14.0 ± 0.49	19.4 ± 0.70	36.5 ± 0.51	26.7 ± 0.96	3.2 ± 0.34[a]
5'-Penultimate nucleotides						
dA		43.2 ± 1.09	19.6 ± 0.93	28.5 ± 0.77	26.3 ± 0.97	15.0 ± 0.66
dC		15.4 ± 0.27	26.2 ± 0.29	30.0 ± 0.48	14.8 ± 0.38	15.5 ± 0.41
dG		25.8 ± 0.53	19.7 ± 0.77	30.5 ± 0.67	31.6 ± 1.04	19.6 ± 1.06
dT		15.6 ± 0.71	34.5 ± 0.52	11.0 ± 0.50	27.5 ± 1.34	48.3 ± 1.06
Number of determinations	15	17	10	11	7	

[a] Direct analysis of 5'-terminal nucleotides indicates that these values are overestimated, the total amount of terminal dT being about 1% (see [7]).

Bernardi *et al.* [9]; snail DNAase was also obtained essentially according to this reference [9] with our unpublished modifications; pancreatic DNAase (code D) was purchased from Worthington (Freehold, N. J.): E. coli DNAase was prepared by C. Cordonnier and G. Bernardi according to a procedure to be published elsewhere; crab DNAase was obtained as described by Sabeur *et al.* [7]. DNA digestions, evaluations of average sizes of digests, dialysis and dephosphorylation of the digest (when necessary), were done as already described [10 – 13, 7]. 5'-Terminal dinucleotides were determined according to Bernardi [2]. The statistical evaluation of data was done using an Olivetti (Ivrea, Italy) Programma 102 desk computer (program 1 A).

RESULTS

Kinetics of Liberation of 5'-Terminal Dinucleotides

5'-Terminal dinucleotides were determined on E. coli DNA degraded by spleen, snail, E. coli and pancreatic DNAases to oligonucleotides having five different average sizes ($\bar{P}_n$) ranging from 50 to 10. For each enzyme, the results obtained on digests having different average sizes were not significantly different and could therefore be averaged out (see below). In the case of the crab enzyme, three hydrolysates having an average size of 25 were analyzed; kinetic results obtained on other DNAs, showed, however, that in this case, too, the composition of the 5'-terminal

dinucleotides was not dependent upon the size of the digest, at least in the $50-10\ \bar{P}_n$ range [7].

Composition of 5'-Terminal Dinucleotides

All the results obtained are presented in Table 1. Also shown in the table are the doublet frequencies for *E. coli* DNA [14], which have been used in a further evaluation of the data (see below), and the composition of the 5'-terminal and penultimate nucleotides, as calculated from the composition of the 5'-terminal dinucleotides. Since all results obtained were used in calculating the average values and the standard deviations, occasionally aberrant values were included. Elimination of such values would have, however, only reduced the standard deviation without affecting significantly the average values. The ratios of standard deviation to average value for 5'-terminal dinucleotides were in the large majority of cases lower than 5% and never higher than 10% (except for dinucleotides dA-dG, dG-dA, dT-dA, dT-dC, dT-dG, dT-dT as released by the crab enzyme; it should be noted that values for the thymidine-starting dinucleotides were very low and overestimated by the present method; see Table 1). The same ratios were always lower than 5% for 5'-terminal and penultimate nucleotides.

The 5'-terminal and penultimate nucleotides as calculated from the 5'-terminal doublets were in good agreement with values previously obtained by a completely different procedure, for the spleen enzyme [8] and, to a lesser extent, for the snail enzyme [15]. For the other enzymes, data on the 5'-terminal and penultimate nucleotides released from *E. coli* DNA were not available for comparison.

Fig. 1−3 display the 5'-terminal doublets released by the five DNAases in the form of histograms; these show very evident differences in the specificities of the enzymes. Fig. 4 shows the histograms of the 5'-terminal and penultimate nucleotides (Table 1).

A better comparison of the enzymes specificities than those of Fig. 1−3 is given by histograms plotting the differences between the percentage of each 5'-terminal dinucleotide and that of the same doublet in *E. coli* DNA (Fig. 5−7).

Fig. 8 shows the ratio of the percentage of each 5'-terminal dinucleotide, as experimentally determined, to the percentage of the same 5'-terminal dinucleotide, as calculated by multiplying the percentage of 5'-terminal and 5'-penultimate nucleotides given by Table 1. Fig. 8 shows very significant deviations, showing that the statistical values cannot replace the experimental one. The calculated, or statistical, 5'-terminal doublets still reflect, however, the real situation. The distribution of the trinucleotides split by the enzymes was therefore calculated by multiplying
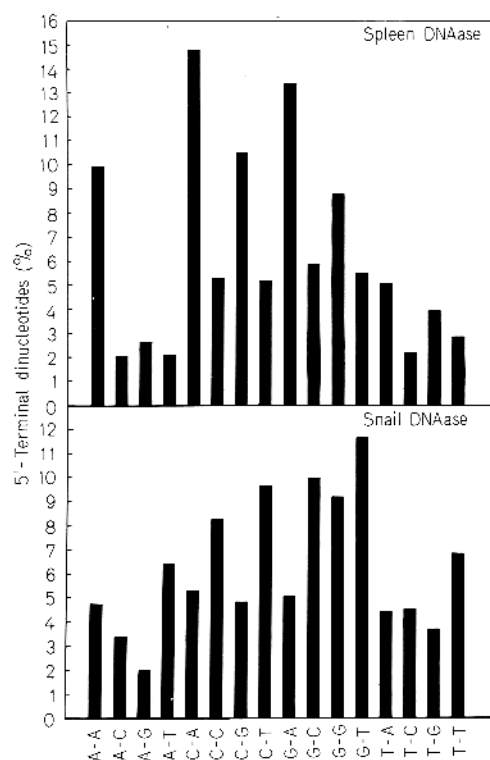

Fig. 1. *Frequencies of 5'-terminal doublets released from* E. coli *DNA by spleen and snail DNAases*
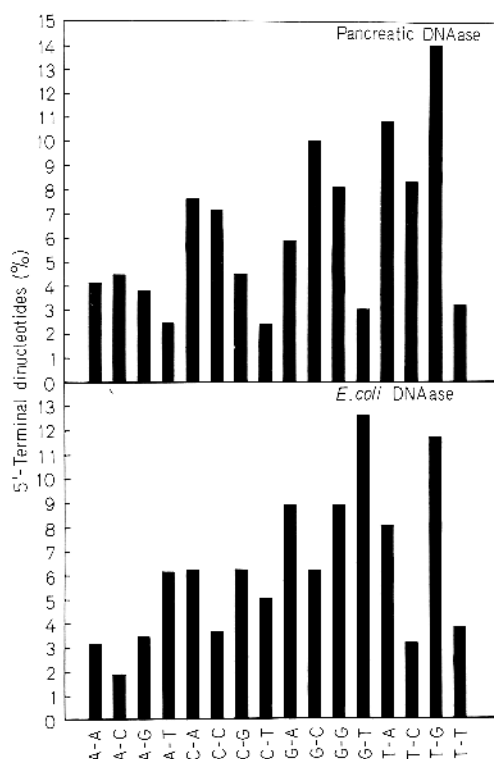

Fig. 2. *Frequencies of 5'-terminal doublets released from* E. coli *DNA by pancreatic and* E. coli *DNAases*
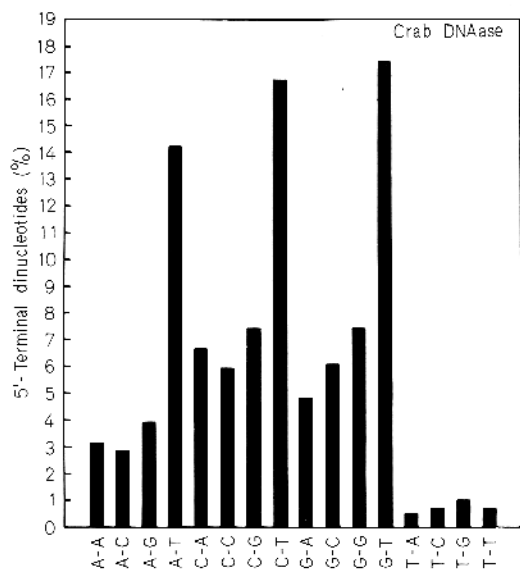
Fig. 3. *Frequencies of 5'-terminal doublets released from* E. coli *DNA by crab DNAase [7]*
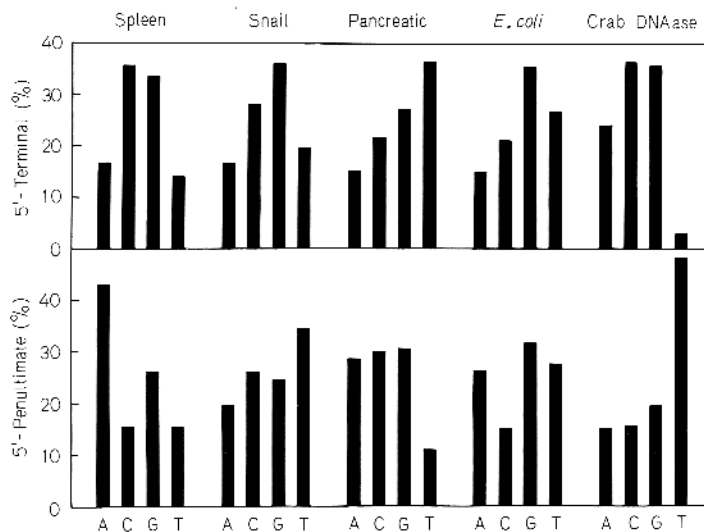

Fig. 4. *Frequencies of 5'-terminal and penultimate nucleotides released from* E. coli *DNA by five different DNAases*
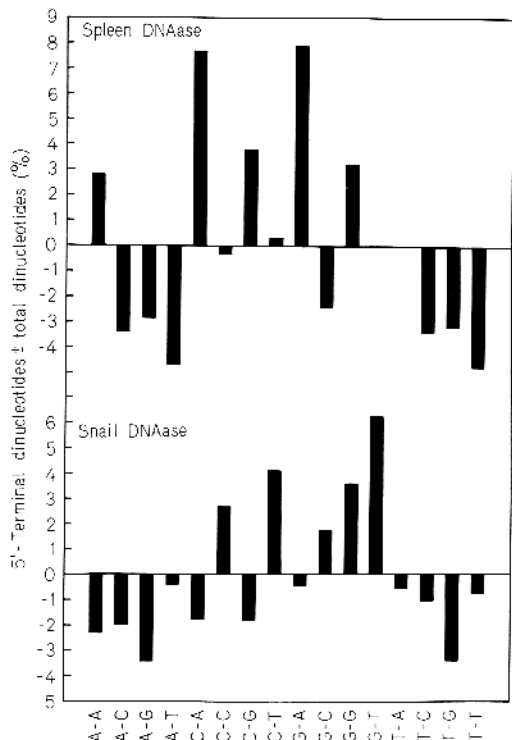

Fig. 5. *Difference histograms of 5'-terminal doublets and total doublets of* E. coli *DNA. Data for spleen and snail DNAases*
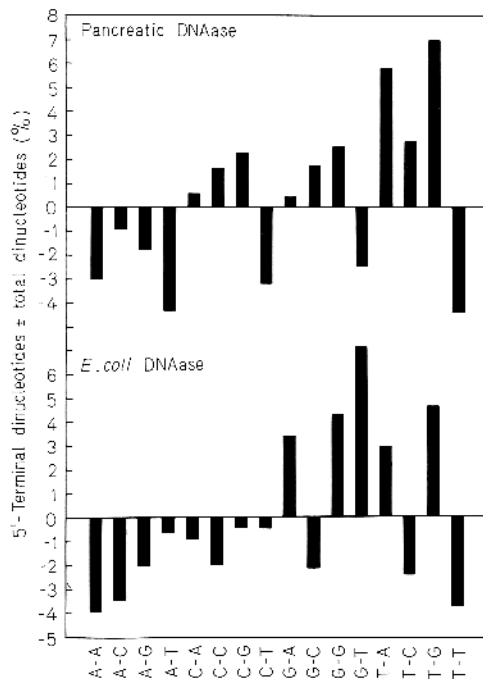

Fig. 6. *Difference histograms of 5'-terminal doublets and total doublets of* E. coli *DNA. Data for pancreatic and* E. coli *DNAases*

the experimental percentages of the 3'-terminal nucleotides, previously determined [7,8,12,13,15] by those of the 5'-terminal dinucleotides (Table 1). The results are shown in Fig. 9 and are taken to represent, in a qualitative way, the trinucleotide sequence split by the enzyme investigated. As expected, the trinucleotide "spectra" strongly differ for different enzymes.

DISCUSSION

The choice of *E. coli* DNA as the substrate for testing the specificity of DNAases was due to the fact that this DNA (a) does not contain short repetitive sequences, which may affect the results [1,16]; (b) has a known nearest-neighbor analysis; (c) has a very
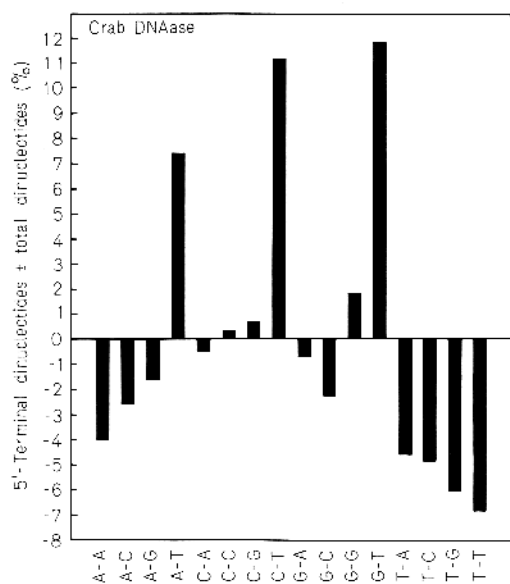
Fig. 7. *Difference histograms of 5'-terminal doublets and total doublets of* E. coli *DNA. Data for crab DNAase*
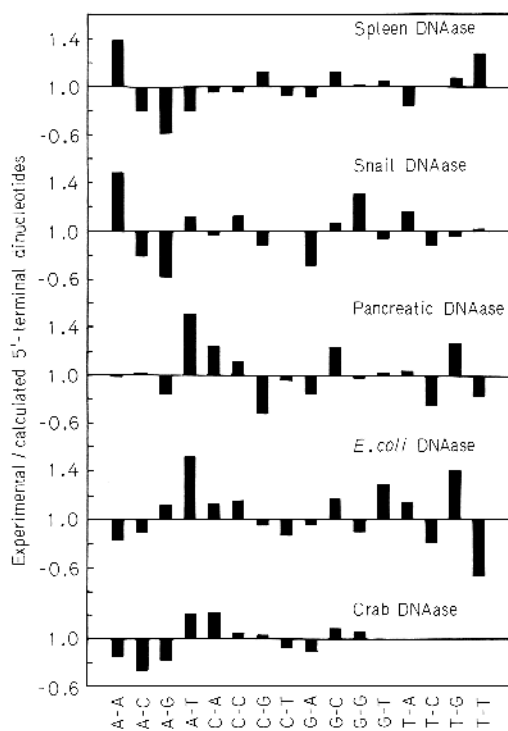

Fig. 8. *Ratio of experimentally found 5'-terminal doublets to "calculated" 5'-terminal doublets.* Calculated values were obtained by multiplying the experimentally found 5'-terminal and 5'-penultimate nucleotides. The last four values for crab DNAases were not plotted, the ratios being very uncertain
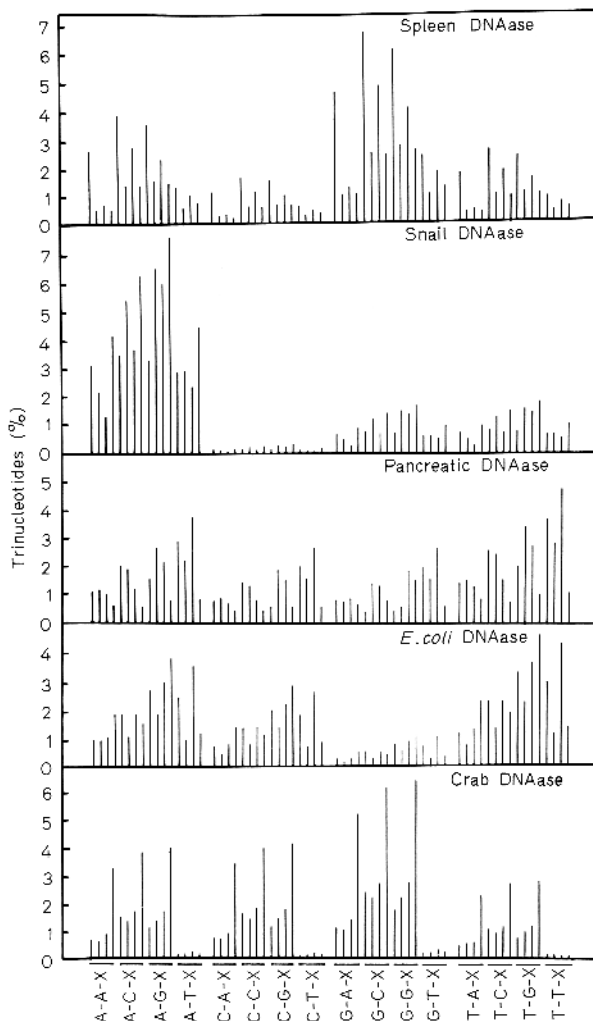

Fig. 9. *Histograms of "calculated" trinucleotides split by five DNAases.* Experimental values for 3'-terminal nucleotides were multiplied by experimental values for 5'-terminal dinucleotides. X corresponds to A, C, G, T, in this order

equilibrated base composition (dG + dC = 51%). An investigation on the effects of differences in the dG + dC contents and of the presence of short repetitive sequences in other DNAs is in progress at the present time. These are very important points as far as the practical use of DNAase specificity in the assessment of oligonucleotide frequencies is concerned, and hopefully should lead to an improved resolution over methods involving the analysis of terminal and penultimate nucleotides independently from each other [1].

The lack of variation in the composition of the 5'-terminal dinucleotides, as released by the different enzymes investigated here, with average size of the DNA hydrolysates (at least in the range under consideration) confirms a conclusion already drawn in

previous work on the composition of the 5'-terminal and penultimate nucleotides [8,11—13]. The very slight variations observed in the composition of the 5'-terminal nucleotides released by the spleen enzyme at different levels of degradation [8] are, therefore, of uncertain significance.

The frequencies of the 5'-terminal doublets (Table 1, Fig. 1—3) show patterns which are characteristic for each enzyme, an expected finding in view of the fact that DNAases recognize and split specific sets of short nucleotide sequences [1].

It is very useful to compare our experimental results with a number of other data. First of all, a comparison of the frequencies of the 5'-terminal doublets (Fig. 1—3) should be done with the frequencies of the 5'-terminal and penultimate nucleotides (Fig. 4), since such a comparison makes very evident the much greater wealth of information obtained by the doublet analysis. This result is quite expected in view of the averaging process which is implicit in the separate assessment of terminal and penultimate nucleotides.

A second comparison which can be made is between the experimental frequencies of 5'-terminal doublets and the statistical value of 6.25%, which corresponds to an equal number of doublets for a DNA having a dG + dC content of 50%. Since E. coli DNA has a doublet pattern slightly differing from that expected for a statistical distribution of nucleotides, a more meaningful comparison can be done with the experimentally determined [14] frequencies of all E. coli DNA doublets. Such a comparison, which obviously is more valuable as the doublet frequencies of the DNAs under consideration deviate more from statistical values, is presented in the form of difference histograms (Fig. 5—7) for E. coli DNA digests.

It is clear that, in this representation, the larger the deviations, the higher the enzyme specificity. On the basis of Fig. 5—7 one should conclude that the crab and, to a lesser extent, the spleen enzyme outweigh the other three DNAases as far as specificity is concerned. In fact, a proper comparison of specificities should take into account the entire recognized sequences. This point is particularly evident if one considers the case of the snail enzyme: this DNAase does not recognize the 5'-penultimate nucleotide [15], but is very highly specific for the 3'-terminal nucleotide. Failure to take these facts into consideration may lead to wrong conclusions.

Useful as it may be, the representations of Fig. 5—7 have two drawbacks, both of them typical of difference histograms: the first one is that they do not give any idea of the relative values of deviations (it is not ob-

vious that the frequencies of 5'-terminal dC-dT and dG-dT in the crab DNAase hydrolysate are over 300% the frequencies of the same doublets in the whole DNA); the second one is that 5'-terminal doublets which are almost absent (like dT-dA, dT-dC, dT-dG and dT-dT in the crab DNAase hydrolysates) have absolute values close to the frequencies of the same doublets in the whole DNA; in this way an important point is missed. In conclusion, deviation histograms of Fig. 5—7 should be used in connection with the simple histograms of Fig. 1—3.

A third comparison which is useful concerns the experimental *versus* the calculated frequencies of the 5'-terminal doublets, the latter being based on the assumption of a statistical distribution of the nucleotides which are experimentally found in the 5'-terminal and penultimate positions. The percentage difference histogram of Fig. 8 shows that the calculated values may easily be off by as much as 50% compared to the experimental values. It is interesting, but not surprising, that deviations are smaller for the more specific enzymes, the spleen DNAase and, more so, the crab DNAase. This can be understood if one thinks that in this case the values to be multiplied by each other are either very large or very small.

The results of Fig. 8 encouraged us to "calculate" the trinucleotides (3'-terminal nucleotide + 5'-terminal doublet) recognized and split by the five DNAases under consideration. The histograms obtained (Fig. 9) have only an indicative value, yet they stress in a graphical way the idea that different sets of short nucleotides are "seen" by different DNAases.

REFERENCES

1. Bernardi, G., Ehrlich, S. D. & Thiery, J. P. (1973) *Nat. New Biol. 246*, 36—40.
2. Bernardi, A. (1975) *Anal. Biochem.*, in press.
3. Bernardi, G. & Griffé, M. (1964) *Biochemistry, 3*, 1419—1426.
4. Laval, J. & Paoletti, C. (1972) *Biochemistry, 11*, 3596—3603.
5. Lehman, I. R., Roussos, G. C. & Pratt, E. A. (1962) *J. Biol. Chem. 237*, 819—828.
6. Sabeur, G., Sicard, P. & Aubel-Sadron, G. (1974) *Biochemistry, 13*, 3203—3207.
7. Sabeur, G., Bernardi, A., Aubel-Sadron, G. & Bernardi, G. (1975) *Eur. J. Biochem. 52*, 445—449.
8. Thiery, J. P., Ehrlich, S. D., Devillers-Thiery, A. & Bernardi, G. (1973) *Eur. J. Biochem. 38*, 434—442.
9. Bernardi, G., Bernardi, A. & Chersi, A. (1966) *Biochim. Biophys. Acta, 129*, 1—11.
10. Soave, C., Thiery, J. P., Ehrlich, S. D. & Bernardi, G. (1973) *Eur. J. Biochem. 38*, 423—433.

11. Laval, J., Thiery, J. P., Ehrlich, S. D., Paoletti, C. & Bernardi, G. (1973) *Eur. J. Biochem. 40*, 133−137.

12. Ehrlich, S. D., Bertazzoni, U. & Bernardi, G. (1973) *Eur. J. Biochem. 40*, 143−147.

13. Ehrlich, S. D., Bertazzoni, U. & Bernardi, G. (1973) *Eur. J. Biochem. 40*, 149−153.

14. Josse, J., Kaiser, A. D. & Kornberg, A. (1961) *J. Biol. Chem. 236*, 864−875.

15. Ehrlich, S. D., Devillers-Thiery, A. & Bernardi, G. (1973) *Eur. J. Biochem. 40*, 139−141.

16. Ehrlich, S. D., Thiery, J. P., Devillers-Thiery, A. & Bernardi, G. (1974) *Nucleic Acids Res. 1*, 87−96.

A. Bernardi, C. Gaillard, and G. Bernardi, Institut de Biologie Moléculaire, Université de Paris VII, 2, Place Jussieu, F-75005 Paris, France