

REMARKS ON SEQUENCE CHARACTERISTICS OF THE DNA AND TRANSFER RNA OF YEAST*

BY ERWIN CHARGAFF AND HERMAN S. SHAPIRO

CELL CHEMISTRY LABORATORY, DEPARTMENT OF BIOCHEMISTRY, COLUMBIA UNIVERSITY

Communicated November 13, 1967

To what extent does a short section of a DNA chain reflect, in its sequence characteristics, the entire molecule? With this question in mind, we have examined the available evidence on the frequencies of purine and pyrimidine isostichs in several species of transfer RNA of yeast and compared them with the corresponding data on yeast DNA.

This comparison makes use of the fact that a nucleic acid chain can be regarded as an array of successive tracts of purine and pyrimidine nucleotides of different length and composition.¹ A size group comprising all tracts of the same length has been termed an isostich.² In a fully base-paired, double-stranded DNA molecule we deal thus with a series composed of an equal number of purine and pyrimidine isostichs of varying length whose relative frequencies can be predicted if the arrangement is random.

The DNA of baker's yeast³ belongs to the AT-type⁴ and contains, as mole per cent, A, 31.5; G, 18.5; C, 17.3; T, 32.7.^{5, 6} The distribution of pyrimidine isostichs of length $n = 1$ to 6 in this nucleic acid has been determined recently.⁷

The species of yeast transfer RNA ready for inspection are those for which the complete nucleotide sequence has been proved or suggested, namely, the nucleic acids specific for the transfer of alanine,⁸ phenylalanine,⁹ serine,¹⁰ tyrosine,¹¹ and valine.¹² These five molecular species, which will be regarded together, as if they were aligned in one row, contain a total of 201 purines and 197 pyrimidines, apportioned among isostichs of length 1 to 7 as shown in Table 1. The double-stranded segment of yeast DNA, one strand of which could act as the template for the specification of the combined tRNA sequences, will consist of one DNA chain repeating the RNA sequence and one complementary to it. This constructed portion of yeast DNA would hence comprise 398 base pairs. It would correspond to only 0.0018 per cent of the yeast genome, if the figure

TABLE 1. *Distribution of purine and pyrimidine isostichs in yeast transfer RNA.**

Isostich (<i>n</i>)	Purines					Pyrimidines				
	Ala	Phe	Ser	Tyr	Val	Ala	Phe	Ser	Tyr	Val
1	10	8	7	5	10	11	5	6	5	10
2	7	3	5	4	8	2	4	6	5	3
3	2	2	4	3	—	3	6	3	3	4
4	—	1	1	3	3	1	1	2	1	1
5	—	1	1	—	—	1	—	—	2	1
6	—	1	1	1	—	—	—	1	—	1
7	1	1	—	—	—	1	—	—	—	—
Total nucleotides	37	42	44	40	38	40	35	41	38	43

* Based on published data.⁸⁻¹² The two varieties of serine tRNA described¹⁰ have been treated as one, since they do not differ in isostich distribution. The unidentified component *Y* in Phe tRNA⁹ has been arbitrarily listed as a purine.

TABLE 2. *Isostich frequencies in tRNA samples (summation of values in Table 1); the composition of corresponding segment of DNA and of total yeast DNA.*

Isostich (n)	tRNA, No. of Isostichs		No. of Py isostichs in DNA segment	Moles Py Isostich/100 Moles Py in:	
	Purines	Pyrimidines		DNA segment	Total DNA*
1	40	37	77	19.3	26.9
2	27	20	47	11.8	13.3
3	11	19	30	7.54	5.98
4	8	6	14	3.52	2.90
5	2	4	6	1.51	1.47
6	3	2	5	1.26	0.65
7	2	1	3	0.75	

* Based on published values.⁷

of 2.2×10^7 nucleotide pairs per haploid nucleus of *S. cerevisiae* is accepted.¹³

The distribution of purine and pyrimidine isostichs in the five varieties of yeast tRNA is shown in Table 1. The isostich frequencies resulting from the summation of the findings for these five samples and the expected composition of the reconstructed segment of double-stranded DNA that could specify these sequences are tabulated in Table 2, together with the experimentally found⁷ distribution of the pyrimidine isostichs of length 1 to 6 in the total DNA of yeast.

The comparison of these data, which we make in Figure 1, is based on previous studies from this laboratory.^{1, 14} In this work, it was pointed out that the relative frequencies of pyrimidine isostichs (as moles per 100 moles of DNA pyrimidines) present in the DNA of several microbial species and in the DNA of *E. coli* phage T3, plotted as the logarithms against the length n of the isostichs, were defined by a straight line (calculated by means of the method of least squares) whose slope did not deviate significantly from that of the line describing a random distribution of isostichs. The equations defining these lines were: for random distribution, $\log F = -0.301 n + \log 50$; for microbial DNA, $\log F = -0.282 n$

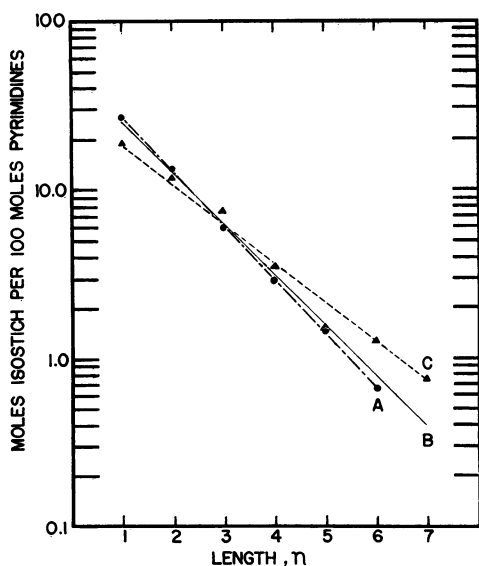


FIG. 1.—Frequencies of pyrimidine isostichs of length n in total yeast DNA (curve A; circles) and in the reconstructed segment of yeast DNA capable of specifying the five yeast tRNA varieties discussed in text (curve C; triangles). Curve B corresponds to the frequencies expected from a random distribution of isostichs. Compare text for equations defining the lines.

+ log 42.5. (F refers to the frequency, i.e., the moles of isostich per 100 moles of pyrimidines in DNA; n to isostich length.) A different slope, most probably deviating from randomness, was found in a similar graph constructed from the rather scarce material on mammalian DNA. Here, the resulting equation was: $\log F = -0.238 n + \log 30.1$. For details and for our conclusions on the possible mechanisms of the evolutionary diversification of DNA the previous paper should be consulted.¹ It is, in fact, not impossible that DNA molecules from different taxonomic groups will be found to be distinguishable by a characteristic distribution of isostich frequencies.

The curves assembled in Figure 1 show that the frequency distribution of isostichs in total yeast DNA (curve A) again deviates little from conditions of randomness (curve B), in this respect reminiscent of bacterial DNA, whereas the distribution in the postulated DNA segment specifying the five tRNA samples (curve C) shows a marked deviation from randomness. The equations are: total yeast DNA, $\log F = -0.322 n + \log 57.0$; postulated DNA segment, $\log F = -0.231 n + \log 31.2$. That a relatively short DNA section comprising only 400 base pairs would exhibit a considerable divergence from randomness, even if the entire DNA is constructed randomly, was to be expected. It will be interesting to see to what extent the present conclusions will be modified as more tRNA sequences become available.

* Supported by a research grant from the National Institutes of Health, U.S. Public Health Service.

¹ Shapiro, H. S., R. Rudner, K. Miura, and E. Chargaff, *Nature*, **205**, 1068 (1965).

² Shapiro, H. S., and E. Chargaff, *Biochim. Biophys. Acta*, **91**, 262 (1964).

³ Chargaff, E., and S. Zamenhof, *J. Biol. Chem.*, **173**, 327 (1948).

⁴ Abbreviations: A, adenine; G, guanine; C, cytosine; T, thymine; Py, pyrimidine; tRNA, transfer RNA.

⁵ Vischer, E., S. Zamenhof, and E. Chargaff, *J. Biol. Chem.*, **177**, 429 (1949).

⁶ Zamenhof, S., and E. Chargaff, *J. Biol. Chem.*, **187**, 1 (1950).

⁷ Symons, R. H., and B. W. Ellery, *Biochim. Biophys. Acta*, **145**, 368 (1967).

⁸ Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir, *Science*, **147**, 1462 (1965).

⁹ Raj Bhandary, U. L., A. Stuart, R. D. Faulkner, S. H. Chang, and H. G. Khorana, in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 31 (1966), p. 425.

¹⁰ Zachau, H. G., D. Dütting, and H. Feldmann, *Z. Physiol. Chem.*, **347**, 212 (1966).

¹¹ Madison, J. T., G. A. Everett, and H. Kung, *Science*, **153**, 531 (1966).

¹² Bayev, A. A., T. V. Venkstern, A. D. Mirzabekov, A. I. Krutilina, V. A. Axelrod, L. Li, and V. A. Engelhardt, in *Genetic Elements, Properties and Function*, ed. D. Shugar (London and New York: Academic Press, 1967), p. 287.

¹³ Esser, K., and R. Kuenen, *Genetik der Pilze* (Berlin, Heidelberg, New York: Springer-Verlag, 1965), p. 228.

¹⁴ Rudner, R., H. S. Shapiro, and E. Chargaff, *Biochim. Biophys. Acta*, **129**, 85 (1966).