CHROM. 19 420

STUDIES ON THE NUCLEOTIDE ARRANGEMENT IN DNA

ESTIMATION OF SEQUENCE ISOMERS IN PYRIMIDINE DEOXYRIBO-OLIGONUCLEOTIDES

HERMAN S. SHAPIRO

University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Department of Biochemistry, Newark, NJ 07103 (U.S.A.)

(First received August 8th, 1986; revised manuscript received January 12th, 1987)

SUMMARY

The population of pyrimidine deoxyribo-oligonucleotides obtained from DNA may be separated, first, into fractions comprising identical lengths (isotichs) and, subsequently, into the individual components of an isostich fraction varying in the cytidylic and thymidylic acid composition. At a third level of analysis the proportion of components exhibiting the same composition but different sequence may be determined. Selective removal of cytosine and the subsequent analysis for characteristic thymidine derivatives released during an acid-catalyzed β-elimination reaction allowed estimation of the relative proportion of sequence isomers for pyrimidine isostichs of length two to five.

INTRODUCTION

Hydroxylamine has proven to be a useful reagent in the study of the structure of ribonucleic acid[1] by virtue of its ability to effect a selective degradation of the uracil chromophore at pH 10. The application of this reaction to DNA, following preliminary deamination of the cytosine component to uracil was the subject of previous communications[2,3]. This differential degradation of pyrimidine structures allowed estimation of contiguous thymidylic acid units, T-runs, and was the basis of our conclusion that the longest T-run in T7 bacteriophage DNA has a length of fourteen and occurs, uniquely, once in the molecule[3]. Another application of our technique was demonstrated by Cox and Yanofsky[4] to elucidate distortion of the frequencies of T-runs in the DNA of *Escherichia coli*, mut T1; a strain capable of introducing and accumulating unidirectional transversions (A/T → C/G) into the bacterial genome. These present procedures have purpose in surveying sequence characteristics in DNA where unique-sequence technology[5,6] is not appropriate, and for the analysis of oligonucleotides containing methylcytidylic acid.

Chromatographic procedures[7] and enzymatic approaches[8], are applicable to estimate the proportion of sequence isomers in simple mixtures and have been effec-

tive for identification of longer sequences[9]. A reasonable and unique separation of sequence isomers for pyrimidine oligonucleotides up to length five by reversed-phase high-performance liquid chromatography (HPLC) has been described by Dizdaroglu and co-workers[10-12] and by Schott and co-workers[13,14].

The selective removal of erstwhile cytosine units from DNA by hydroxylaminolysis can, however, allow subsequent analysis of the proportion of sequence isomers in long pyrimidine tracts. The feasibility of estimating the molar proportions of sequence isomers by this method has previously been demonstrated[2]. In this communication our attention was primarily directed to estimation of the proportion of sequence isomers within pyrimidine nucleotide runs of length 2, 3, 4 and 5 of calf thymus DNA, and to establish the adequacy of the technique for application to oligonucleotides of more complex composition tabulated for plant DNA[15,16].

MATERIALS AND METHODS

Preparation of model pyrimidine deoxyribo-oligonucleotides

Freshly frozen calf thymus was the source of DNA used in this study. The DNA was isolated by the procedure of Kay *et al.*[17]. One gram of purified calf thymus DNA was suspended in 900 ml of 66% formic acid (v/v) containing 2% diphenylamine (w/v) and incubated for 18 h at 30°C, during which time the DNA became completely solubilized and the solution developed the typical purple-blue color for deoxyribose[18,19]. The solution was diluted with two parts water and subjected to continuous liquid–liquid extraction with ether. The clear aqueous phase was evaporated to dryness under reduced pressure at 37°C. The residue was dissolved in 2 l of 0.01 M lithium acetate, pH 5.5, and percolated through a column of DEAE-cellulose, Whatman DE-23, prepared from 50 g of cycled exchanger equilibrated with 0.01 M lithium acetate buffer. The separation of pyrimidine nucleotide isostich groups by a linear concentration gradient of lithium chloride in 0.01 M lithium acetate buffer has been described by Spencer and Chargaff[20]. The total column influent was 5.5 l. Ten-ml fractions were collected.

Each pooled isostich fraction, corresponding to pyrimidine nucleotide runs 1–5 was diluted four-fold with water and re-absorbed into columns prepared from 10 g of DEAE-cellulose. Separation of the isostich components on the basis of composition was achieved by a sodium chloride concentration gradient (0 to 0.4 M) in 0.1 M formic acid[21,22]. The volume of column influent was 900 ml. Five-ml fractions were collected during each chromatographic separation.

Each fraction described by the general formula $p(dC_n,dT_m)p$ with $n + m = 2, 3, 4$ and 5 was freed of salt by differential adsorption on acid-activated charcoal[23]. The nucleotide material was recovered from the charcoal by elution with 50% ethanol–conc. ammonium hydroxide (99/1, v/v), and evaporated to dryness.

Dephosphorylation and deamination of pyrimidine oligonucleotides

Each oligonucleotide, $p(dC_n,dT_m)p$, was dissolved in 5 ml of 5 mM Tris [tris(hydroxymethyl)aminomethane]–HCl, pH 8.0, and dephosphorylated by incubation with 50 μg *E. coli* alkaline phosphatase for 18 h at 37°C. One-fifth of each fraction was removed for subsequent use as chromatographic standards. The remaining material was treated with one volume of 1.3 M nitrous acid, pH 3.35, for

TABLE I

SOLVENT PARTITION CHROMATOGRAPHY OF DEAMINATED OLIGONUCLEOTIDES AND THYMIDINE DERIVATIVES

Compound	Solvent system[*]		
	n-Propanol–2 N hydrochloric acid	Ammonium isobutyrate, pH 3.6	70% Isopropanol in NH_3 atmosphere
(UT)p	0.63	0.95	1.75
$(U_2T)p_2$	0.45	0.68	1.21
$(UT_2)p_2$	0.54	0.77	1.40
$(U_3T)p_3$	0.35	0.43	0.58
$(U_2T_2)p_3$	0.37	0.46	0.68
$(UT_3)p_3$	0.40	0.53	0.88
$(U_4T)p_4$	–	0.20	0.23
$(U_3T_3)p_4$	–	0.27	0.36
$(U_2T_3)p_4$	–	0.34	0.40
$(UT_4)p_4$	–	0.36	0.49
pT	1.00	1.00	1.00
Tp	1.13	1.07	1.03
pTp	1.06	0.52	0.55
pTpTp	0.85	0.42	0.41
pTpTpTp	0.63	0.34	0.29
pTpTpTpTp	0.43	0.26	0.22
pTpT	0.76	0.74	0.88
pTpTpT	0.63	0.53	0.83
TpT	0.76	1.09	1.98
TpTpT	0.63	0.85	1.47

[*] Mobility of components is relative to thymidine 5'-phosphate.

72 h at 37°C in sealed vials. The deaminated, dephosphorylated, pyrimidine components were finally obtained salt-free by differential ammonium bicarbonate elution on small DEAE-cellulose columns[24]. Removal of the ammonium bicarbonate in solution was done under vacuum at 40°C in a rotating evaporator. The salt-free solutions of oligonucleotides were freeze-dried and the residue stored at −20°C. From our previous experience, the recovery of dephosphorylated, deaminated material is approximately 90%[2,19]. Some spectrophotometric and chromatographic characteristics of these deaminated compounds and relevant degradation products are included in Tables I and II. Identification of components relied on these data and upon composition; $T/U/PO_4$.

Hydroxylamine digestion of dephosphorylated, deaminated pyrimidine oligonucleotides

Salt-free hydroxylamine was prepared from hydroxylamine–HCl by the procedure of Verwoerd et al.[1]. The hydroxylamine was recovered by vacuum distillation and assayed by titration with sulfuric acid. Hydroxylamine (10 N) was stable for over six months when stored at −20°C.

The oligonucleotides (dU,dT), (dU_2,dT), (dU,dT_2), (dU_3, dT), (dU_2,dT_2), (dU,dT_3), (dU_4,dT), (dU_3,dT_2), (dU_2,dT_3) and (dU,dT_4) were individually treated with salt-free hydroxylamine. A shorter period of digestion was used in the present

TABLE II

ABSORBANCE RATIOS OF DEAMINATED PYRIMIDINE OLIGONUCLEOTIDES

Compound	O.D. 250 nm / O.D. 260 nm	O.D. 280 nm / O.D. 260 nm	O.D. 290 nm / O.D. 260 nm
pT	0.64	0.73	0.24
pU	0.74	0.38	0.03
(UT)p	0.69	0.51	0.13
$(U_2T)p_2$	0.81	0.54	0.18
$(UT_2)p_2$	0.74	0.64	0.21
$(U_3T)p_3$	0.77	0.47	0.08
$(U_2T_2)p_3$	0.74	0.52	0.12
$(UT_3)p_3$	0.76	0.68	0.26
$(U_4T)p_4$	0.71	0.47	0.07
$(U_3T_2)p_4$	0.69	0.53	0.11
$(U_2T_3)p_4$	0.68	0.60	0.16
$(UT_4)p_4$	0.66	0.66	0.20

work than was reported previously[2]. Approximately 25 absorbance units (260 nm) of each oligonuleotide were dissolved in 0.5 ml of 7.5 N hydroxylamine adjusted to pH 10 with dilute ammonium hydroxide. The reaction tubes were not sealed during 2 h of gentle mixing at 37°C. The contents of each tube were freeze-dried. The residues were solubilized with small portions of water and evaporated to dryness *in vacuo*. This procedure was repeated until all hydroxylamine was removed and a vitreous residue remained. The residues were dissolved in 0.5 ml water and shaken with several drops of benzaldehyde for several hours[25]. Excess benzaldehyde and benzaldoxime were removed by ether extraction. The aqueous phase was evaporated to dryness.

The hydroxylaminolysis degradation products of the (dU_ndT_m) compounds were now hydrolyzed with 0.25 ml of 0.1 M sulfuric acid at 100°C for 35 min to catalyze the β-elimination of thymidine derivates[26,27]

Chromatographic analysis for thymidine derivatives

Two dimensional descending solvent partition chromatography on sheets of Whatman 1 filter paper adequately separated the products derived from the series of reactions which we have just described. Preliminary washing of the filter paper (46 cm × 57 cm) sequentially with 1 N hydrochloric acid, distilled water, and 95% ethanol was done prior to their use. Portions of a neutralized sulfuric acid digest corresponding to 30–50 μg nucleotide phosphorus were applied to the filter paper and developed with n-propanol–2 N HCl (3:1) for 18 h. The papers were dried in air, rotated 90° and developed for 18 h with a buffered isobutyrate solvent mixture, pH 3.6, isobutyric acid–0.5 M ammonium hydroxide (5:3). Areas of ultraviolet absorption were excised, extracted with standard volumes of 0.01 N hydrochloric acid (3–5 ml) and quantitatively estimated spectrophotometrically. (Our results have been calculated using the following molar extinction data: deoxyuridylic acid, $\Delta\varepsilon_{260} - \varepsilon_{290} = 9700$; uracil, $\Delta\varepsilon_{260} - \varepsilon_{290} = 7410$; thymidylic acid, $\Delta\varepsilon_{260} - \varepsilon_{300} = 8620$; thymine, $\Delta\varepsilon_{260} - \varepsilon_{290} = 6730$.)

Analyses of reaction products

Control compounds, UpU and TpT were processed through the hydroxylamine and acid treatment which has been described. Changes in the proportion of organic phosphorus and recovery of chromophores were estimated. The only measurable change for TpT was a 0.8% conversion of organic P to inorganic P. For UpU a 60% decrease in the ultraviolet absorption maximum of deoxyuridine coincided with a 47% conversion of organic P to inorganic P. These values reflect the extent of destruction of the uracil chromophore by hydroxylamine and the subsequent acid-catalyzed β-elimination process which produces inorganic P^2. Longer periods of exposure to hydroxylamine (3–5 h) cause complete destruction of the uracil chromophore. While the refractory nature of the thymine chromophore is still evident under similar treatment, unacceptable levels of phosphate ester hydrolysis occurs, *i.e.* 10%. Therefore, in the experiments reported here the shorter 2-h exposure to hydroxylamine has been employed. We have also assumed that the hypochromism in pyrimidine oligonucleotides is negligible[8].

RESULTS AND DISCUSSION

The results of our analyses are included in Table III which lists the yields of thymidine derivatives expected from the acid catalyzed β-elimination degradation of the chemically modified (dC_n,dT_m) oligonucleotides. An estimation of the proportion of sequence isomers from each set of data is not difficult (Table IV). For example: The original pyrimidine oligonucleotide (dC_2,dT_2) is composed of six isomers; 1,

TABLE III

THYMIDINE DERIVATIVES RELEASED FROM SEQUENCE ISOMERS

Serial treatment by nitrous acid, hydroxylamine and sulfuric acid are described in Materials and methods.

Gross composition of initial components	Molar recovery of thymidine derivatives[*]								Others
	pT	Tp	pTp	pTpT	TpTp	pTpTp	pTpTpT +	TpTpTp	
(dC,dT)	61	39	—[**]	—	—	—	—	—	
(dC₂,dT)	68	19	12	—	—	—	—	—	
(dC,dT₂)	33	29	—	29	10	—	—	—	
(dC₃,dT)	46	22	32	—	—	—	—	—	
(dC₂,dT₂)	31	17	11	21	12	8	—	—	
(dC,dT₃)	27	16	—	14	23	—	20		
(dC₄,dT)	41	36	23	—	—	—	—	—	
(dC₃,dT₂)	36	21	16		22§	9	—	—	
(dC₂,dT₃)	22	19	9		20§	9	21		pTpTpTp[***]
(dC,dT₄)	22	12	—	13	11	—	30		pTpTpTpT + TpTpTpTp 13

[*] Corrected to 100%.
[**] Indicates not expected and not observed.
[***] pTpTpTp expected but not observed.
§ Indicates summation of pTpT and TpTp.

TABLE IV

ESTIMATED PROPORTIONS OF SEQUENCE ISOMERS

Gross composition of initial component	Percent contribution of each isomer unit
(dC,dT)	CpT(61), TpC(39)
(dC$_2$,dT)	CpCpT(68), CpTpC(12), TpCpC(19)
(dC,dT$_2$)	CpTpT(41), TpCpT(44), TpTpC(14)
(dC$_3$,dT)	CpCpCpT(46), CpCpTpC + CpTpCpC(32), TpCpCpC(22)
(dC$_2$,dT$_2$)	CpCpTpT(30), CpTpCpT(16), TpCpCpT(26), TpCpTpC(10), TpTpCpCp(17), CpTpTpC(11)
(dC,dT$_3$)	CpTpTpT + TpTpTpC(60), TpCpTpTp(15), TpTpCpT(25)
(dC$_4$,dT)	CpCpCpCpT(41), CpCpCpTpC + CpCpTpCpC + CpTpCpCpC(23), TpCpCpCp(36)
(dC$_3$,dT$_2$)	CpCpCpTpT(22), CpCpTpTpC + CpTpTpCpC(13), TpTpCpCpC(10), CpCpTpCpT + CpTpCpCpT(22), TpCpCpCpT(30), TpCpCpTpC + TpCpTpCpC(2), CpTpCpTpC (0)
(dC$_2$,dT$_3$)	CpCpTpTpTp + TpTpTpCpC(37), CpTpCpTpT + TpTpCpTpC(4), CpTpTpCpT + TpCpTpTpC(16), TpCpTpCpT(12), TpCpCpTpT + TpTpCpCpT(31), CpTpTpTpC(0)
(dC,dT$_4$)	CpTpTpTpT + TpTpTpTpC(22), TpCpTpTpT(20), TpTpTpCpT(37), TpTpCpTpT(20)

CpCpTpT; 2, CpTpCpT; 3, TpCpCpT; 4, TpCpTpC; 5, TpTpCpC; and 6, CpTpTpC. Upon serial dephosphorylation (to remove terminal phosphate groups), deamination (to convert cytosine to uracil), hydroxylaminolysis (to destroy the uracil chromophore) and acid treatment (to catalyze the β-elimination reactions), the folowing thymidine derivatives would be produced: 1, pTpT; 2, pTp = pT; 3, Tp = pT; 4, Tp = pTp; 5, TpTp; and 6, pTpTp, respectively. Isomers 2, CpTpCpT; 3, TpCpCpT; and 4, TpCpTpC have derivatives in common and the proportion of isomers 2, 3 and 4 relies on algebraic solution of the following set of equations

$$31x + 17y = 1$$

$$31(1-x) = 17(1-y)$$

employing the data from table III. Where possible, similar calculations have been made and the proportion of isomers for each structure has been estimated.

For the unit (dC$_2$,dT$_3$), the thymidine derivative pTpTpTp would be expected on the basis of a random distribution of linear events. This compound was not observed. Similarly, for the unit (dC,dT$_4$), components pTpTpTpT and TpTpTpTp were expected but not observed. In several analyses the fragment pairs pTpT + TpTp and pTpTpT + TpTpTp were not chromatographically resolved. The experimental data, therefore, did not always allow a unique solution for the relative frequency by each isomer.

The elegant reversed-phase HPLC separation described by Schott and Eckstein[14] has allowed extensive tabulation of the relative proportion of sequence-isomeric pyrimidine oligonucleotides in herring sperm DNA. There had been a paucity

of data prior to those analyses. Some data, notably that of Peterson and co-workers[7,23] show a CpT to TpC ratio in calf thymus DNA ranging from 1.33 to 1.55. This compares to our ratio of 1.56. Statistically, the frequency ratio of these pyrimidine-sequence isomers should, of course, be unity. A set of data for TpC/CpT from five bacterial DNAs indicated no obvious correlation with the T/C composition of the DNA[23,28]. Another unexpected aspect of these analyses was that the TpC/CpT isomer proportion form calf thymus and herring sperm DNA; having identical nucleotide composition; were not similar. We prefer, therefore, not to compare the data from calf thymus DNA reported in this communication with the unique HPLC data from herring sperm DNA. Those HPLC analyses for units of length two to five demonstrated the presence of most of the pyrimidine sequence isomers in DNA. The pyrimidine run, (dC_4, dT) was not observed on HPLC although its occurrence has been documented by us and others for eukaryote and prokaryote DNA[16,29].

The introduction of 5-methyl deoxycytidylic acid as a third component in pyrimidine oligonucleotide tracts would show a complex HPLC elution profile whose components would be more difficult to fully separate and quantitate. Although animal DNA contains a low proportion of this minor component, plant DNA contains nearly ten nucleotide percent of methylcytidylic acid[30]. Based on our results and on work in progress we believe it is possible to distinguish between cytosine and methylcytosine loci within pyrimidine oligonucleotides by a modification of the procedure employed for chemical deamination of DNA. Under appropriate conditions, 3 M sodium bisulfite deaminates cytosine to uracil while methylcytosine, adenine and guanine are refractory to this reagent[31,32]. Incorporation of this slective procedure into the protocol described in this communication would allow an estimation of the proportion of sequence isomers of small pyrimidine oligonucleotides composed of thymine, cytosine and methylcytosine.

We have continued to explore the differential chemical stability of DNA in an effort to obtain maximum information of some features of nucleotide arrangement. Integration of HPLC separation techniques with the chemical manipulations discussed here would extend analytical capabilities for the estimation of sequence isomers of more complex pyrimidine oligonucleotides.

REFERENCES

1 D. W. Verwoerd, W. Zilling and H. Kohlhage, *Physiol. Chem.,* 332 (1963) 184–203.
2 H. S. Shapiro and E. Chargaff, *Biochemistry,* 5 (1966) 3012–3019.
3 H. S. Shapiro, *Fed. Proc., Fed. Am. Soc. Exp. Biol.,* 32 (1973) 664.
4 E. C. Cox and C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.,* 58 (1967) 1895–1902.
5 A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.,* 74 (1977) 560–564.
6 F. Sanger, S. Nicklen and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.,* 74 (1977) 5463–5467.
7 K. Burton and G. B. Petersen, *Biochem. J.* 75 (1960) 17–27.

8 H. S. Shapiro and E. Chargaff, *Biochim. Biophys. Acta*, 26 (1957) 596–608.
9 C.-P. D. Tu and R. Wu, *Methods Enzymol.*, 65 (1980) 620–638.
10 M. Dizdaroglu, W. Hermes, C. von Sonntag and H. Schott, *J. Chromatogr.*, 169 (1979) 429–435.
11 M. Dizdaroglu and W. Hermes, *J. Chromatogr.*, 171 (1979) 321–330.
12 M. Dizdaroglu, M. G. Simic and H. Schott, *J. Chromatogr.*, 188 (1980) 273–279.
13 H. Schott, H. D. Meyer and E. Bayer, *J. Chromatogr.*, 280 (1983) 297–311.
14 H. Schott and H. Eckstein, *J. Chromatogr.*, 296 (1984) 363–368.
15 H. S. Shapiro and E. Chargaff, *Biochim. Biophys. Acta*, 39 (1960) 68–82.
16 J. H. Spencer and E. Chargaff, *Biochim. Biophys. Acta*, 68 (1963) 18–27.
17 E. R. M. Kay, N. S. Simmons and A. L. Dounce, *J. Am. Chem. Soc.*, 74 (1952) 1724–1726.
18 K. Burton and G. B. Peterson, *Biochim. Biophys. Acta*, 26 (1957) 667–668.
19 H. S. Shapiro and E. Chargaff, *Biochim. Biophys. Acta*, 91 (1964) 262–270.
20 J. H. Spencer and E. Chargaff, *Biochim. Biophys. Acta*, 68 (1963) 9–17.
21 G. B. Petersen and J. M. Reeves, *Biochim. Biophys. Acta*, 129 (1966) 438–440.
22 R. Cerny, W. Musynski and J. H. Spencer, *Biochim. Biophys. Acta*, 169 (1968) 439–450.
23 G. B. Petersen, *Biochem. J.*, 87 (1963) 495–500.
24 G. W. Rushizky and H. A. Sober, *Biochim. Biophys. Acta*, 55 (1962) 217.
25 A. Temperli, H. Türler, P. Rüst, A. Danon and E. Chargaff, *Biochim. Biophys. Acta*, 91 (1964) 462–475.
26 H. S. Shapiro and E. Chargaff, *Biochim. Biophys. Acta*, 76 (1963) 1–8.
27 H. S. Shapiro, *Methods Enzymol.*, 12A (1967) 205–212.
28 T. Hudnik-Plevnik and L. A. Stocken, *Nature (London)*, 192 (1961) 554–555.
29 R. Rudner, H. S. Shapiro and E. Chargaff, *Biochim. Biophys. Acta*, 129 (1966) 85–103.
30 H. S. Shapiro, *Chemical Rubber Company Handbook Biochem. and Mol. Biol. (U.S.A.), 3rd ed., Nucleic Acids*, Vol. II (1975) 282–283.
31 R. Shapiro, B. Braverman, J. B. Louis and R. E. Servis, *J. Biol. Chem.*, 248 (1973) 4060–4064.
32 R. Y-H. Wang, C. W. Gehrke and M. Ehrlich, *Nucleic Acids Res.*, 8 (1980) 4777–4790.