Nucleosides, Nucleotides, and Nucleic Acids, 25:705–717, 2006

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257770600725903

Taylor & Francis
Taylor & Francis Group

SYNTHESIS OF POLY d(G-C) OLIGONUCLEOTIDES

John P. Hachmann and Michal Lebl □ *Illumina, Inc., San Diego, California, USA*

□ *Model sequences for evaluation of the GC dimer sequence repetition on synthesis success were prepared and analyzed by HPLC. Contiguous d(G-C) or d(C-G) sequences have a deleterious effect on DNA oligonucleotide synthesis. The critical number seems to be about 6 GCs in a row. If the GCs are separated by other nucleotides, the effect is not as severe.*

Keywords Solid phase synthesis; Difficult sequences; GC motif

INTRODUCTION

Difficulties are sometimes encountered in the synthesis of oligonucleotides. Some of these can be traced to problems with starting reagents, instrument failure, or occasionally, human error. Sometimes, however, even when the reagents are tested and found to meet specifications, the instruments validated, and procedures carefully followed, a poor synthesis might occur. In the early days of oligonucleotide synthesis,[1,2] when the manual syntheses were laborious and time consuming, it was extremely difficult to conduct extensive systematic studies as to what might be causing the problems. From time to time, reports emerged based largely on anecdotal evidence that some sequences were more difficult to synthesize than others. It was claimed, for example that purine-purine couplings gave poorer yields than pyrimidine-pyrimidine or purine-pyrimidine reactions. In particular, it was believed that runs of deoxyguanosine had a deleterious effect on synthesis.

Now with the advent of high-throughput parallel synthesis using the solid-phase phosphoramidite approach,[3,4] it is possible to carry out such structure reactivity studies in a quick and efficient manner. This article describes one such study of the effect of sequence on

Received 8 August 2005; accepted 29 December 2005.
Address correspondence to Michal Lebl, Illumina, Inc., 9885 Towne Centre Dr., San Diego, CA 92130, USA. E-mail: mlebl@illumina.com

oligodeoxyribonucleotide synthesis. In particular, the effects of stretches of $d(G\text{-}C)_n$ are discussed.

This study was initiated after an attempt was made to synthesize oligonucleotide standards for an electrophoresis study. It was decided to synthesize a number of oligos with differing chain lengths, base composition, and degree of secondary structure. In one parallel synthesis, 24 oligomers were prepared. Satisfactory results were obtained for 23 of them, but the synthesis of $d(G\text{-}C)_{16}$ resulted in poor yield and analysis by HPLC showed many shorter failure sequences. When a repeat of this experiment gave the same result, we decided to examine the effect of $d(G\text{-}C)_n$ on oligonucleotide synthesis.

EXPERIMENTAL

Materials and Methods

All DNA synthesis reagents were of standard formulation and were obtained from commercial sources. Phosphoramidites, supports, capping reagents, and oxidizer solution were purchased from Glen Research, Sterling, VA. The activator solution was 4,5-dicyanoimidazole in acetonitrile and was also purchased from Glen. Acetonitrile was purchased from Burdick and Jackson, Muskegon, MI. Anhydrous acetonitrile for dissolving phosphoramidites, dichloroacetic acid, and ammonia solution were purchased from Sigma-Aldrich, Milwaukee, WI. Dichloromethane was purchased from VWR International, West Chester, PA.

HPLC analyses were carried out on Agilent (Palo Alto, CA) 1100 HPLC systems. Eluents were monitored at 260 nm. Measurements of OD_{260} were done with a Tecan (US, Durham, NC) SPECTRAfluor Plus microtiter plate reader.

Oligonucleotide Synthesis

The synthesis of the oligonucleotides was carried out in our Pet Oligator[46] using commercial reagents and our own protocols. This instrument has a 24-well rotor, which allows 24 syntheses to be carried out in parallel. For this study, oligodeoxyribonucleotides 24, 32, 40, and 70 bases in length were prepared.

Preparation of Reagents

Cap A solution (tetrahydrofuran (THF)/acetic anhydride (9:1), Cap B solution (10% N-methylimidazole in THF/pyridine (9:1), oxidizer solution (0.02 M iodine in THF/pyridine/H_2O (89.6:0.04:10), and activator solution (4,5-dicyanoimidazole in acetonitrile) were purchased and the vendor's bottles were attached directly to the instrument. Low water content (<30 ppm)

acetonitrile was purchased in 4-L bottles. To each bottle, a large Trap Pak™ (ABI) was added and the bottle was attached to the machine. Deblock solution was prepared by addition of 100-mL dichloroacetic acid (DCA) to 4 L dichloromethane, and this bottle was also attached to the instrument.

The phosphoramidites, d-(Bz)A-CEP, d-(Ac)C-CEP, d-(iBu)G-CEP, and d-T-CEP (Bz = N^6-benzoyl, Ac = N^4-acetyl, iBu = N^2-isobutyryl, CEP = 3′-O-cyanoethyl-N,N-diisopropylaminophosphine) were purchased in serum bottles and were dissolved in anhydrous acetonitrile transferred by syringe from an Aldrich Sure-Seal™ bottle. When dissolved, they were transferred to an oven-dried 60-mL serum bottle, to which a small Trap-Pak™ (ABI) had been added before sealing. Again using a syringe, an additional anhydrous acetonitrile was added to make the final volume. The amidite bottles were then attached to the instrument.

The instrument was connected to a pressure line containing dry nitrogen. The bottles containing the reagents were pressurized to 3.0 ± 0.1 psi. During the run, dry nitrogen was allowed to flow through the reaction chamber at 5 liters per minute.

The solid support used was the Universal Support from Glen Research (Sterling, VA). For oligomers of 40 bases in length or less, 500 Å pore size CPG (controlled pore glass) was used. For longer sequences we used 1000 Å pore size support. The CPG support was plated onto a metal ring, which had holes drilled in it calibrated to contain approximately 4 mg of the support. The 500 Å CPG with typical loadings (approx. 50 μmol/g) resulted in a 200 nmol synthesis per well, and 1000 Å support (approx. 33 μmol/g) gave approximately 133 nmol per well.

The rotor was inverted and placed over the ring. The assembly was inverted again so that the support was deposited into the rotor wells. The rotor was then placed in the synthesizer.

The PET Oligator is computer controlled and its standard phosphoramidite synthesis protocol was used. The instrument allowed the cycle to be paused during the deblock step so that the dimethoxytrityl (DMT) color could be inspected visually. The synthesis cycle was repeated until the complete oligonucleotide was synthesized and then stopped. The final detritylation step was conducted by using a separate program. This allowed the operator to observe the final deblock color. The rotor was removed from the instrument and used in the cleavage and deprotection step described below.

Cleavage and Deprotection

After synthesis, the oligonucleotides were cleaved from CPG and the protecting groups removed by treatment in either concentrated ammonium hydroxide or 40% aqueous methylamine. (The acetyl group was used to protect the N^4 position of cytosine to avoid a methylation side

TABLE 1 Gradients Used for HPLC Analysis of Model Oligonucleotides

Gradient	Length and flow rate	Time (min.)	% A	% B
I	32 and 40 mers	0	90	10
	Flow = 2.3 mL/min	3.3	75	25
		6.5	65	35
		8.5	60	40
		9.5	90	10
II	24 mers	0	95	5
	Flow = 1.5 mL/min	1	95	95
		15	55	45
		16	95	5
III	70 mers	0	90	10
	Flow = 2.3 mL/min	6.5	75	25
		9.8	60	40
		11	90	10

reaction.)[5] In the syntheses in this article, only concentrated ammonia was used.

Analysis

The rotor wells were extracted with 4 × 150 μL water using a Packard Multiprobe™ (Perkin Elmer Life and Analytical Sciences, Boston, MA) robotic dilutor. The cleaved oligonucleotide solutions were transferred to a 1-mL × 96-well microtiter plate. The oligonucleotide solutions were analyzed without additional purification. For HPLC analysis, 50 μL of the extract was transferred to an Evergreen (Evergreen Scientific, Los Angeles, CA) 96-well v-bottom microtiter plate and was diluted by addition of 150 μL water.

A 20 μL aliquot of each well was taken and diluted to 200 μL in a BD Falcon Microtest™ 96-well plate (VWR International, West Chester, PA) with a UV transparent film bottom. The plates were read at 260 nm in a Tecan SpectraFluor Plus plate reader.

HPLC analysis was carried out with an Agilent 1100 HPLC system using a Dionex (Sunnyvale, CA) DNAPac™ PA-100 4 × 250 mm anion exchange column. The runs were carried out at room temperature. Buffer A consisted of 0.02 M NaOH, pH 12.0 ± 0.1, and buffer B was 0.02 M NaOH, pH 12.0 ± 0.1, 2.0 M NaCl. The flow rate is indicated in Table 1 and the absorbance was monitored at 260 nm.

RESULTS

Figure 1 shows the HPLC trace for one of our first attempts to make $d(G\text{-}C)_{16}$. This result proved to be very reproducible over several runs, and therefore it was concluded that it was not a result of the reagents used

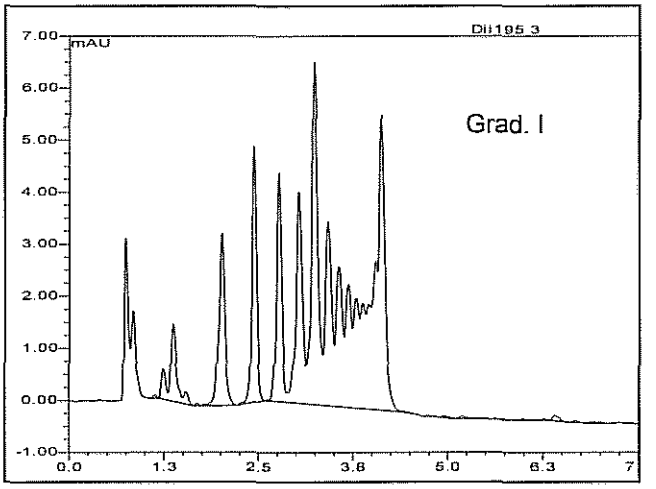

FIGURE 1 HPLC trace of $d(G\text{-}C)_{16}$.

or the instrument. Other sequences synthesized in parallel gave normal chromatograms. We designed a series of experiments to test whether or not these results were due to the sequence. Reversing the order of the bases to make $d(C\text{-}G)_{16}$ gave essentially the same result shown in Figure 2.

Since the synthesis of 32 base long oligonucleotides consisting of only GC dimers is not likely to be routine, the effects of shorter stretches of GC were next studied. Table 2 shows the sequences which were synthesized for this study.

Sequence 1 is a control that, although it contains dG and dC, does not contain d(G-C) dimers. In the sequences 2–5, d(G-C) dimers were

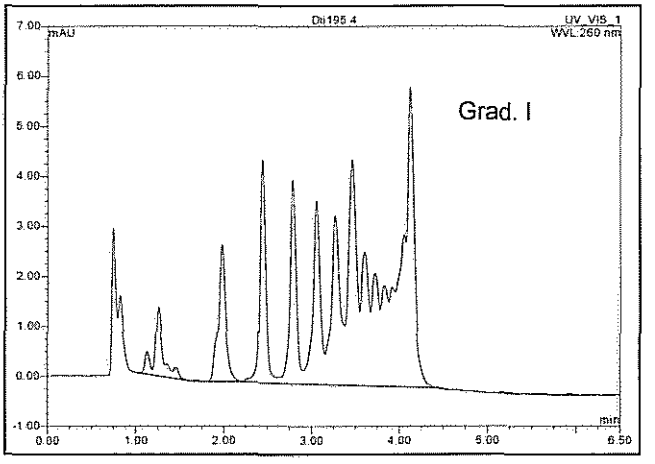

FIGURE 2 HPLC trace of $d(C\text{-}G)_{16}$.

TABLE 2 Oligonucleotides Synthesized to Study the Effect of GC Dimer Cluster Localization in the Sequence

No.	Sequence	$d(G\text{-}C)_n$
1	ATCTAGTTGATGTCCCATTCCCCATGAGTTCT	0
2	ATCTAGTTGATGTCCCATTCCCCATGAG*GCGC*	2
3	ATCTAGTTGATGTCCCATTCCCCA*GCGCGCGC*	4
4	ATCTAGTTGATGTCCCATTC*GCGCGCGCGCGC*	6
5	ATCTAGTTGATGTCC*GCGCGCGCGCGCGCGC*	8
6	*GCGC*AGTTGATGTCCCATTCCCCATGAGTTCT	2
7	*GCGCGCGC*GATGTCCCATTCCCCATGAGTTCT	4
8	*GCGCGCGCGCGC*TCCCATTCCCCATGAGTTCT	6
9	*GCGCGCGCGCGCGCGC*ATTCCCCATGAGTTCT	8
10	ATCTAGTTGATGTC*GCGC*TCCCCATGAGTTCT	2
11	ATCTAGTTGATG*GCGCGCGC*CCCCATGAGTTCT	4
12	ATCTAGTTGA*GCGCGCGCGCGC*CATGAGTTCT	6
13	ATCTAGTT*GCGCGCGCGCGCGCGC*TGAGTTCT	8

substituted for the bases at the 3′ end of the oligonucleotide (Fig. 3 left), in 6–9 at the 5′ end (Figure 3 middle), and 10–13 in the middle (Figure 3 right).

Addition of two or four d(G-C) dimers does not seem to affect synthesis to an appreciable degree. With the addition of six or eight, the effect is more pronounced as shown in the figures below. In Figure 3 left, it should

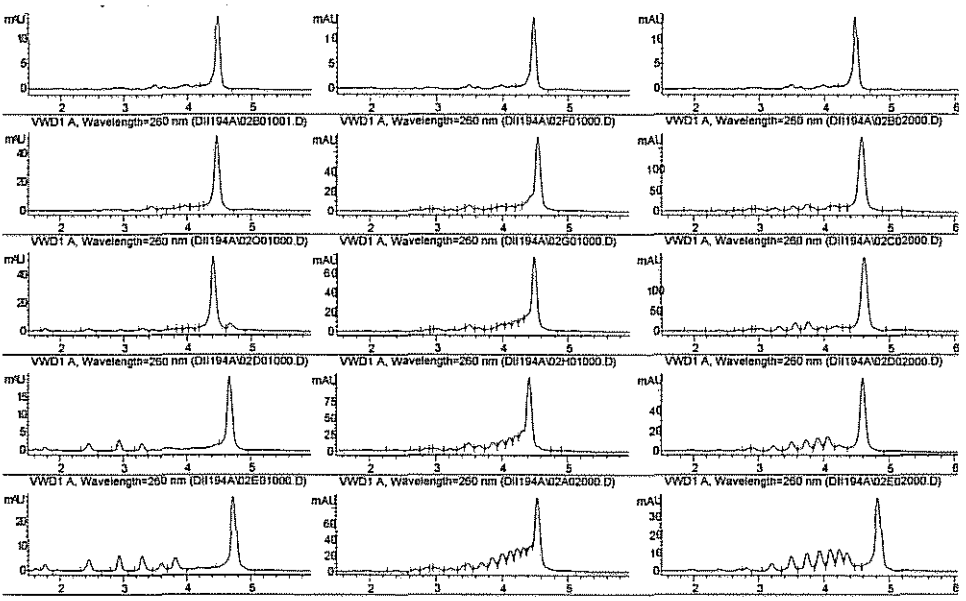

FIGURE 3 HPLC traces of sequences 2–5 with GC clusters at 3′ end (left traces), sequences 6–9 with GC clusters in the middle (right traces), and sequences 10–13 with GC clusters at the 5 end. Top trace in each column is a trace of a control sequence 1.

be noted that the failure sequences elute fairly early in the chromatogram which would indicate that they are of shorter length as one would expect if they were formed early in the synthesis. In Figure 3 middle, they appear later, and in Figure 3 right, they are intermediate. These results indicate that the difficulties appear in the regions where the d(G-C) dimer runs are introduced. Once the synthesis proceeds past that d(G-C) rich region, the synthesis appears to continue normally.

The results in Figure 3 also argue against hydrolysis during cleavage and deprotection steps since no longer sequences are seen in Figure 3 left and no shorter sequences are seen in Fig. 3 middle. Since the oligomers are not purified, all hydrolysis products should remain in the deprotection mix. The conclusion is that the difficulty must arise during synthesis.

To determine whether or not the d(G-C) dimers had to be contiguous for the effect to take place, sequence 14 (A*GCTGCAGCTGCAGCT GCAGCTGCAGCTGCA*T) was synthesized. The oligonucleotide contained 10 d(G-C) dimers, but each was separated by a d-A or d-T. The chromatogram is shown in Figure 4 and clearly demonstrates that the d(G-C) dimers need to be contiguous to have an adverse effect on synthesis.

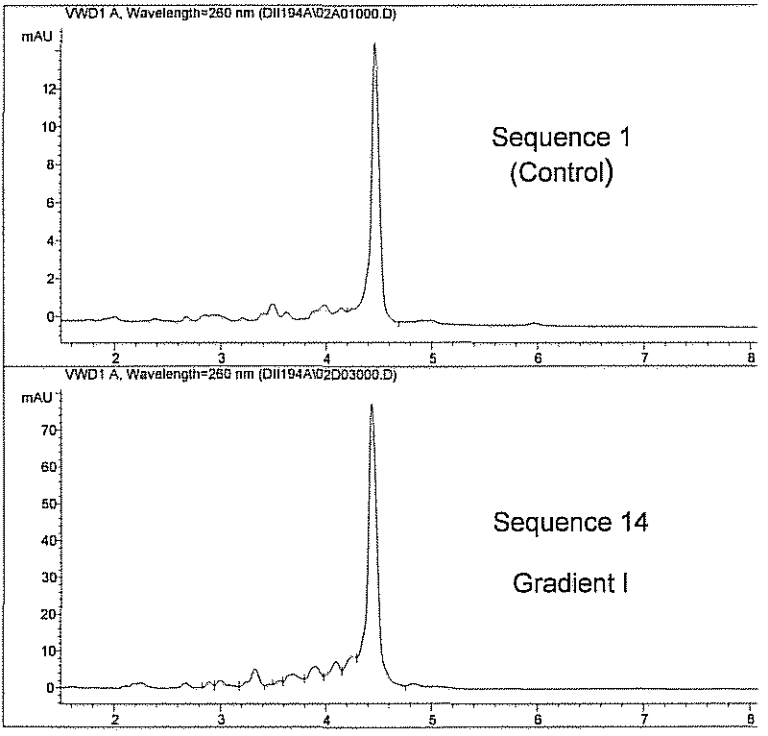

FIGURE 4 HPLC of sequence 1 and 14.

TABLE 3 Sequences with the Same Content of G and C and Variable Content of GC Dimers

No.	Sequence	$d(G\text{-}C)_n$
15	CCCCCCCCCCCCCCCCGGGGGGGGGGGGGGGG	0
16	*GCCCCCCCCCCCCCCCGGGGGGGGGGGGGGGC*	2
17	*GCCCCCCCGCCCCCGCGCGGGGGGGCGGGGGC*	6
18	*GCGCCCCCGCGCCCGCGCGCGGGGGCGCGGGC*	10
19	*GCGCCCCCGCGCGCGCGCGCGGGGGCGCGCGC*	12
20	*GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC*	16

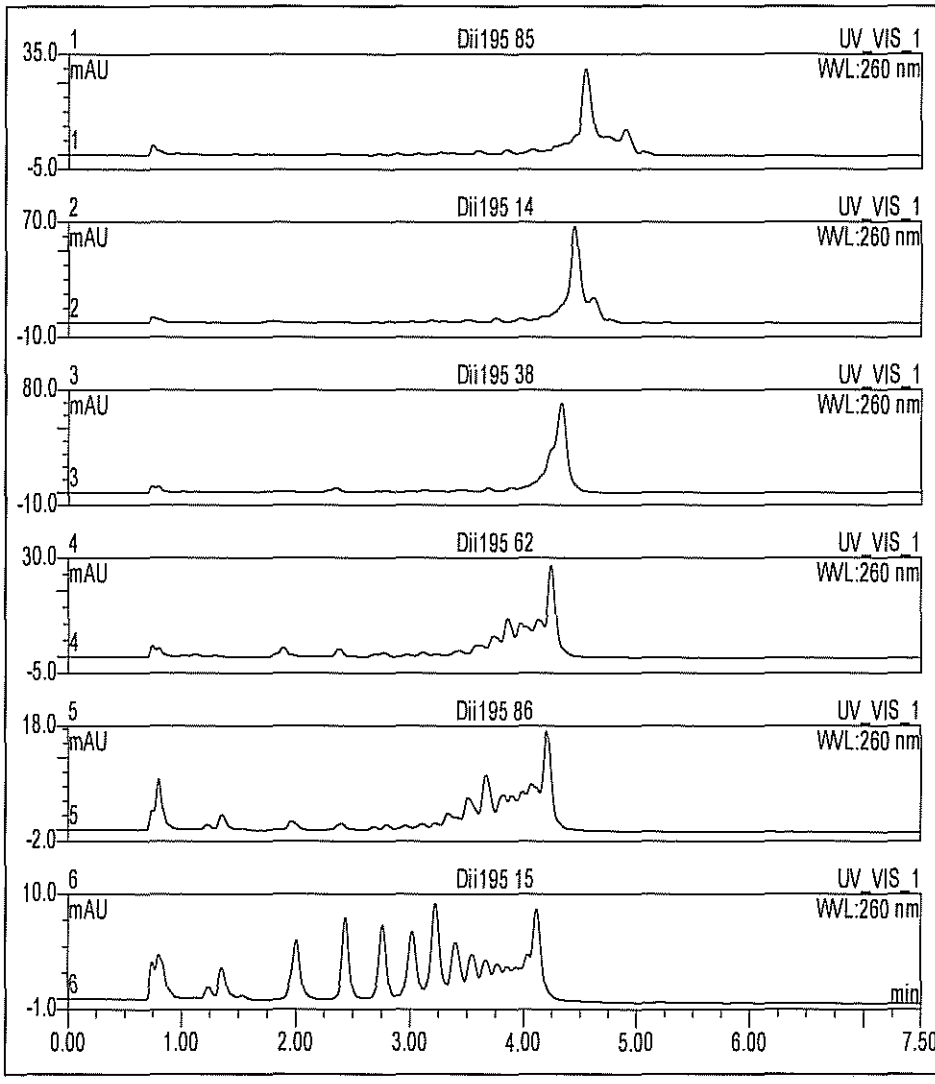

FIGURE 5 HPLC traces of sequences 15–20 (Table 3) with increasing numbers of GC dimer sequences (top to bottom).

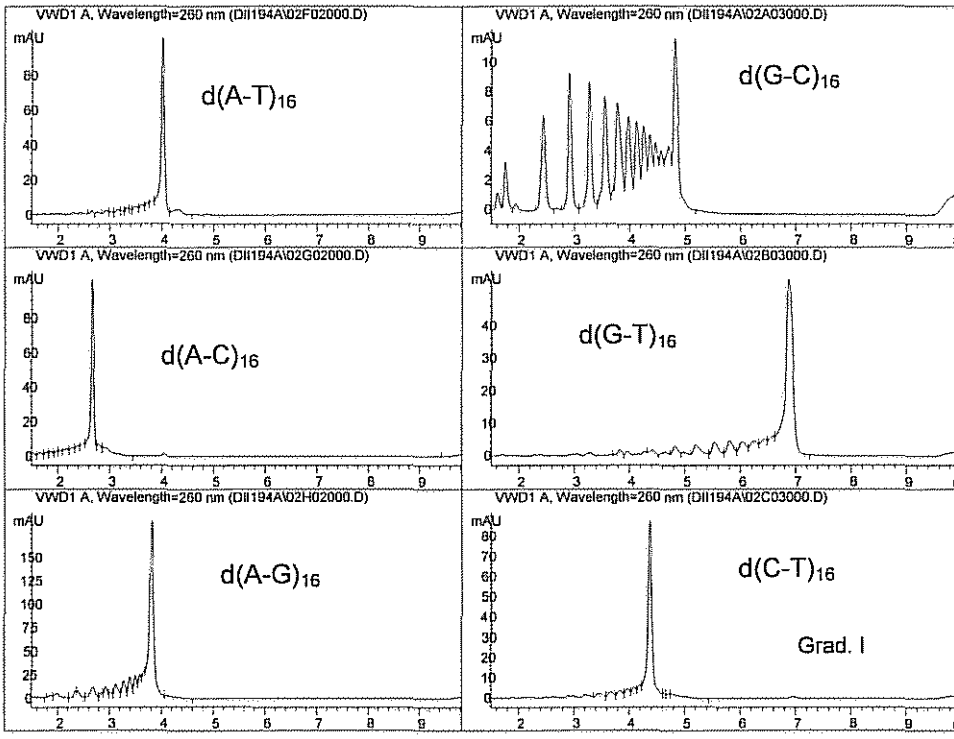

FIGURE 6 HPLC of all possible poly-dimers.

Another question arose as to whether or not it was the d(G-C) dimers responsible for the difficulties in synthesis or was it the dG and dC content alone. To test this, the oligomers in Table 3 were synthesized.

Again, in the absence of runs of d(G-C) the synthesis appears to proceed normally. As the dCs and dGs are "scrambled" the failure sequences become more pronounced.

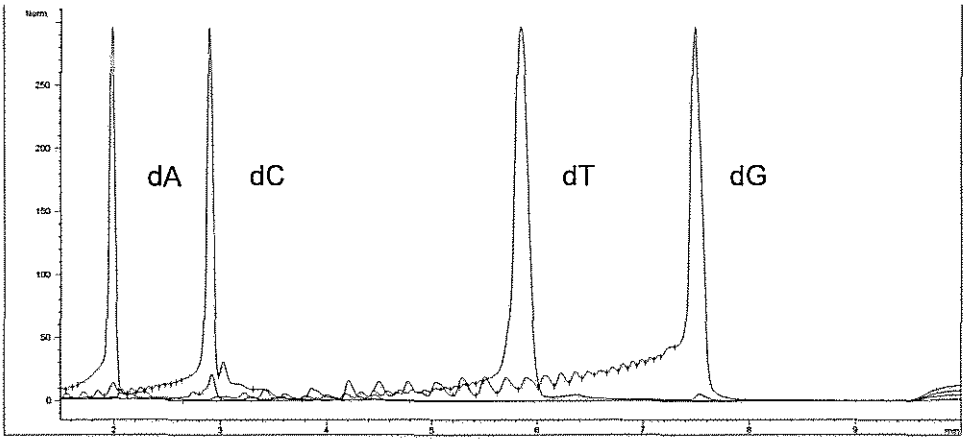

FIGURE 7 HPLC of homooligomers.

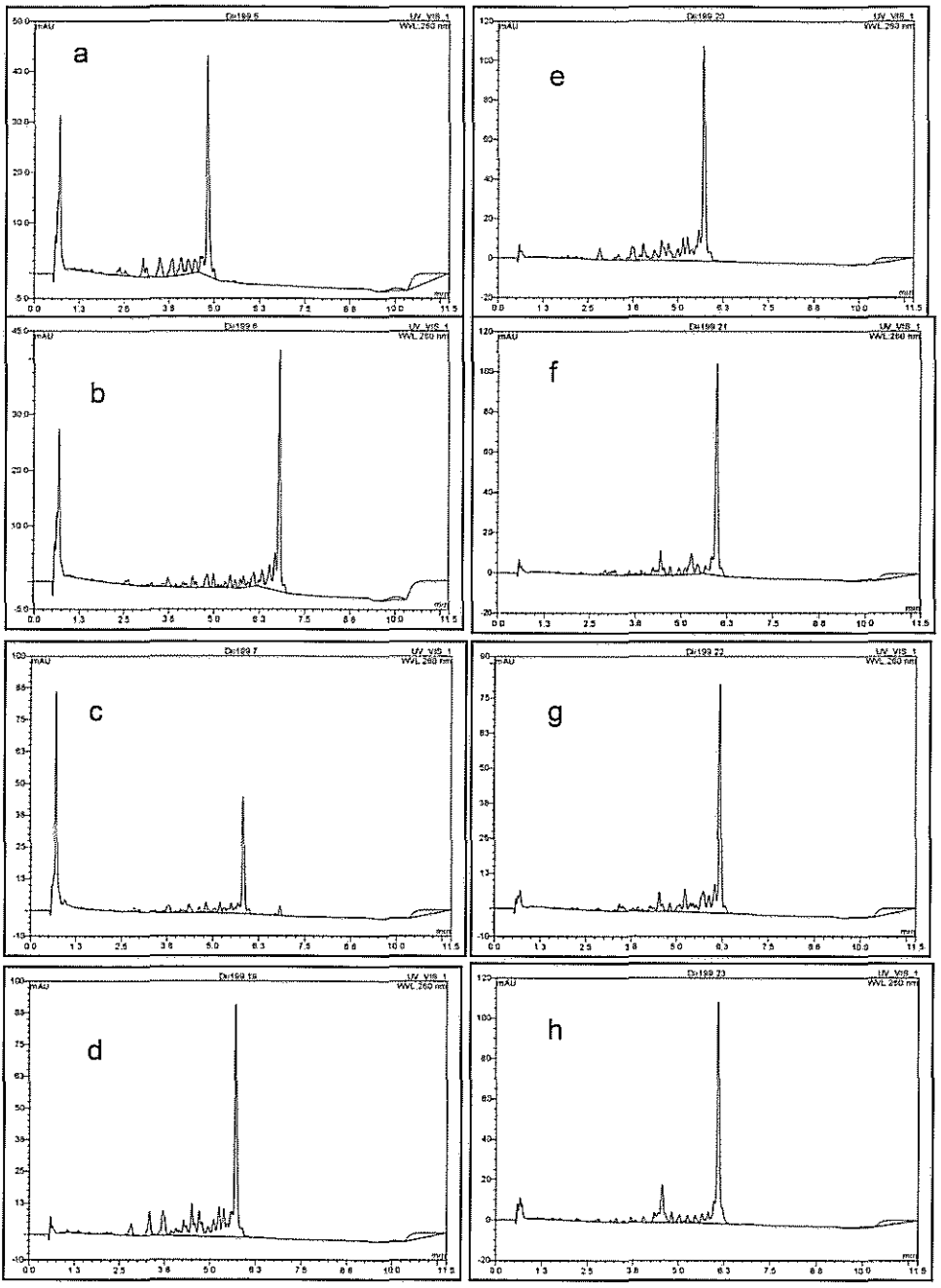

FIGURE 8 HPLC of oligonucleotides 21–28 (see Table 4).

Figure 6 shows the effect of other dimers on synthesis. Again, the most noticeable effect is with $d(G\text{-}C)_n$. Figure 7 shows the chromatograms of the homooligomers $d(A)_{32}$, $d(C)_{32}$, $d(G)_{32}$, and $d(T)_{32}$.

TABLE 4 Sequences of Oligonucleotides with Different G and C Containing Motifs

Number	Trace	Sequence	Motif
21	a	$d(G\text{-}C\text{-}C)_{10}\text{-}G\text{-}C$	GCC Trimer
22	b	$d(G\text{-}G\text{-}C)_{10}\text{-}G\text{-}G$	GGC Trimer
23	c	$d(G\text{-}G\text{-}C\text{-}C)_8$	GGCC Tetramer
24	d	$d(G\text{-}G\text{-}C\text{-}C\text{-}G\text{-}C\text{-}G\text{-}C)_4$	GGCCGCGC Octamer
25	e	$d(G\text{-}C\text{-}C\text{-}G\text{-}G\text{-}C)_5\text{-}G\text{-}C$	GCCGGC Hexamer
26	f	$d(G\text{-}G\text{-}G\text{-}G\text{-}C\text{-}C\text{-}C\text{-}C)_4$	GGGGCCCC Octamer
27	g	$d(G\text{-}G\text{-}G\text{-}C\text{-}C\text{-}C)_5\text{-}G\text{-}G$	GGGCCC Hexamer
28	h	$d(G\text{-}G\text{-}G\text{-}G\text{-}G\text{-}G\text{-}G\text{-}G\text{-}C\text{-}C\text{-}C\text{-}C\text{-}C\text{-}C\text{-}C\text{-}C)_2$	GGGGGGGGCCCCCCCC Hexadecamer

In Figure 7, the retention times of the dG and dT oligomers are longer than those for dA and dC presumably due to the ring dissociations at the N-1 of guanine and the N-3 of thymine in the pH 12 buffer system. The dG coupling does not proceed as well as the others, but it is still more efficient than in the case of $d(G\text{-}C)_n$.

Figure 8 shows the effects of other GC motifs on synthesis. The sequences of the oligonucleotides chromatographed in Figure 8 are given in Table 4.

Although some failure sequences are observed, none give results comparable to $d(G\text{-}C)_{16}$. All of the results described up to now have been with 32 base long oligonucleotides. We also wished to investigate the effect of d(G-C) runs with other chain lengths. Figure 9 shows the results with $d(G\text{-}C)_{12}$, $d(G\text{-}C)_{20}$, and $d(G\text{-}C)_{35}$.

DISCUSSION

In this study, it was found that a series of contiguous d(G-C) or d(C-G) sequences can have a deleterious effect on DNA oligonucleotide synthesis. The critical number seems to be about 6 GCs in a row. If the GCs are separated by other nucleotides, the effect is not as severe. The problem is specific to d(G-C-G-C-G-C-etc). Repeating trimers, such as $d(G\text{-}C\text{-}C)_n$ or $d(G\text{-}G\text{-}C)_n$, do not show the effect. Runs of other dimers or monomers do not seem to have the same effect. Rudimentary molecular modeling (MM2, Chem3D, CambridgeSoft) suggests that the AT sequence is nearly linear, but the GC sequence is more convoluted. We may speculate that the accessibility of the 5′ end of the molecule plays a large role in reactivity during the coupling reaction. Alternative explanation that GC rich sequences are degraded during deprotection was ruled out by experiment in which products were exposed to repeated deprotection. The profile of impurities did not change significantly in these experiments.

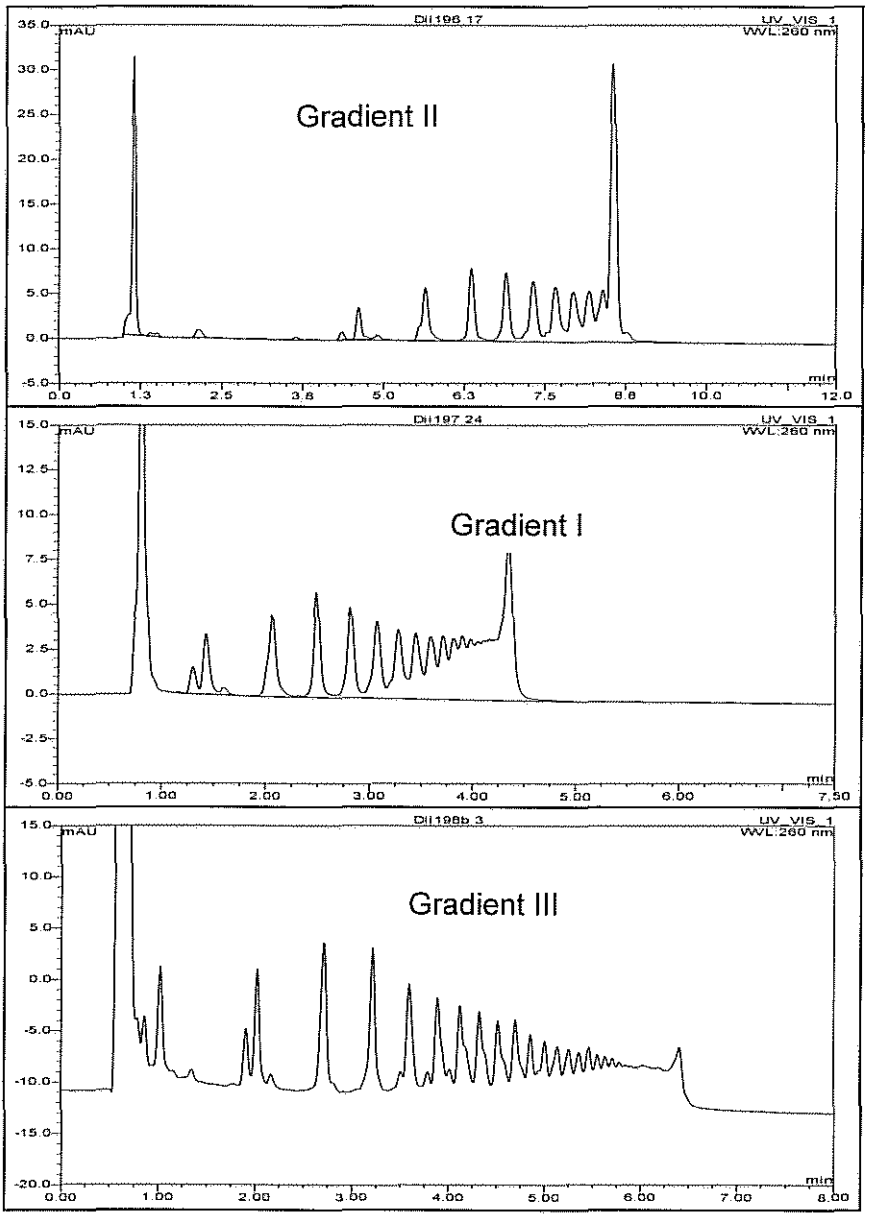

FIGURE 9 HPLC of various lengths GC repeat oligonucleotides. 24-mer $d(G\text{-}C)_{12}$, top, 40-mer $d(G\text{-}C)_{20}$, middle trace, and 70-mer $d(G\text{-}C)_{35}$, bottom. (Note different gradient for each trace.)

REFERENCES

1. Schaller, H.; Weimann, G.; Lerch, B.; Khorana, H.G. *Journal of the American Chemical Society* **1963**, 85, 3821.
2. (a) Letsinger, R.L.; Mahadevan, V. *Journal of the American Chemical Society*, **1965**, 87, 3526–3527. (b) Itakura, K.; Bahl, C.P.; Katagisi, N.; Michniewiez, J.J.; Wightman, R.H.; Narang, S.A. *Canadian Journal of Chemistry*, **1973**, 51, 3649–2651.

3. (a) Beaucage, S.L.; Caruthers, M.H. *Tetrahedron Letters*, **1981**, 22, 1859–1862. (b) Sinha, N.D.; Biernat, J.; Köster, *Nucleic Acids Research*, **1984**, 12, 4539–4557.
4. (a) Lebl, M.; Burger, C.; Ellman, B.; Heiner, Ibrahim, G.; Jones, A.; Nibbe, M.; Thompson, J.; Mudra, P.; Pokorny, V.; Poncar, P.; Zeníšek, K. *Collection of Czechoslovak Chemical Communication*, **2001**, 66, 1299–1314. (b) Lebl, M. *JALA*, **2003**, 8, 30–35.
5. Reddy, M.P.; Hanna, N.B.; Farooqui, F. *Tetrahedron Letters* **1994**, 35, 4311–4314.