ELSEVIER

Journal of Chromatography A, 706 (1995) 405–419

JOURNAL OF CHROMATOGRAPHY A

Efficiency of chemical oligonucleotide synthesis evaluated by ion-exchange high-performance liquid chromatography

Zeno Földes-Papp[a,*], Eckhard Birch-Hirschfeld[b], Rudi Rösch[a], Manfred Hartmann[b], Albrecht K. Kleinschmidt[c], Hartmut Seliger[a]

[a]*Sektion Polymere, Universität Ulm, D-89069 Ulm, Germany*
[b]*Institut für Biotechnologie e.V., D-07745 Jena, Germany*
[c]*Universität Ulm, D-89069 Ulm, Germany*

Abstract

Ion-exchange high-performance liquid chromatography (HPIEC) is a method of choice for the separation of crude solid-phase synthesized oligonucleotides. It shows peaks for target and error sequences and it is used for quantification of oligonucleotide synthesis. This paper presents a homeodynamic model of polymer-supported oligonucleotide synthesis and a numerical solution, which accounts for elution profiles obtained by HPIEC. The elution profiles are consistent with the model. The chain-length distributions for 30mer syntheses on different support materials were also studied and the probability moment analysis was applied as a quantitative measure of the differential relationship of fractal dimension. The support systems for chemical synthesis of short oligonucleotides are classified according to this relationship.

1. Introduction

The preparation of high-purity oligonucleotides of extended length, large number or quantity is becoming increasingly important in genetic engineering [1–3], molecular biology and molecular medicine [4–6]. This is accomplished by chemical solid-phase synthesis on surface-reactive polymers as support systems [7].

Ion-exchange chromatography (IEC) is a method of choice for the purification and analysis of chemically synthesized oligonucleotides [8–11]. The separation of crude products of solid-phase synthesis is mainly based on the interaction of the negatively charged phosphate groups of the oligonucleotide backbone with the cations of the stationary phase. The elution profiles show peaks for target and error sequences. Error sequences are truncated or failure sequences formed in the course of oligonucleotide synthesis. No oligonucleotide synthesis provides 100% of the target sequence.

We are working on a theory for analysing chain-length distributions obtained in non-enzymatic and enzymatic nucleotide polymerization processes and at applying it to real systems. The modelling of HPIEC elution profiles is of practical relevance to oligonucleotide chemistry. The knowledge of the composition of the product provides information for various applications of synthetic nucleic acids fragments and their structurally modified analogues. Here we used the simulated HPIEC elution profiles to provide a basis for judging support systems of chemical

* Corresponding author.

0021-9673/95/$09.50
SSDI 0021-9673(94)01221-0

synthesis of short oligonucleotides by the fractal dimension, D_a.

2. Theory

Today, chemical oligonucleotide syntheses on solid supports are most frequently performed by phosphoramidite chemistry [12–14]. The synthesis is done by a cyclic repetition i of four reactions: detritylation, coupling (elongation), capping (termination) and oxidation in organic solvents. The chemical method of phosphoramidite synthesis used in this study exemplifies the general principle of multi-step condensation and capping reactions arranged in cyclic format. The same cyclic reaction principle of growth and termination of growth is put into practice in other chemical methods of oligonucleotide synthesis, e.g., when using H-phosphonate synthons.

In the condensation reaction the 3′ end of the incoming nucleotide building block binds covalently to the 5′ end of the immobilized chain. Chain propagation proceeds in the 3′ to 5′ direction (5′ to 3′ reactions are equally possible). In the capping reaction, the former unreacted chains are chemically capped, to some extent, to prevent them from reaction with the next incoming nucleotide building block. Nucleotide chains grow randomly with the coupling (elongation) probability d. The former unreacted chains (error sequences) stop growing with the capping (termination) probability p; d and p are independent of each other.

We have modelled analytically the dynamics of a non-enzymatic polymerization process by a new fractal formalism. Our fractal approach comes from fractal methods [15–22]. Here we elaborate on our theoretical concept in more details and compare it with experimental results obtained through IEC.

2.1. Growth model of chemical solid-supported oligonucleotide synthesis

We have recently formulated a model which generates variations of error sequences [23]. Error sequences l are truncated versions of a target sequence N. A sequence is based on entries of {0, 1, 2} [24]; 1 = nucleotide (arbitrary A, T, C, G); 0 = no nucleotide at position referring to the target sequence; 2 = stopping of growth. The sequence l of real numbers is a map $l: \mathbb{N} \rightarrow \mathbb{R}$ where the non-empty set of its members $L(\mathbb{N}) = \{l_j : j \in \mathbb{N}\}$. $\mathbb{N}$ and $\mathbb{R}$ denote the sets of positive integers and real numbers, respectively. The sum of all sequences might be generated by an iterative process (variation generator). The results of a probabilistic experiment such as chemical solid-supported oligonucleotide synthesis confirm this model. For the set L we assume that $N = \sup_{l \in L} l = \max\ L$ and $1 = \inf_{l \in L} l = \min\ L$. The model function

$$M(l, N) = M(l, N-1) + [M^*(l-l, N-1) - M^*(l, N-1)] \cdot d(N-1) \quad (1)$$

describes the relationships between these generated discrete sequences in the mathematical sequence space. Thus, M is the probability density of this homeodynamic system. M^* is the probability density of non-terminated error sequences:

$$M^*(l, N) = M^*(l, N-1) \cdot [1 - d(N-1) - p(N-1) + d(N-1) \cdot p(N-1)] + M^*(l-1, N-1) \cdot d(N-1) \quad (2)$$

The probabilities d and p are explicit, real functions of the iteration step i satisfying $0 \leqslant d, p < 1$; d and p can be properly fitted in iteration steps i. Each particular solution is obtained from this general solution. This model depends on the nucleotide length N of a target sequence, the stepped-up probability of growth d and the stepped-up probability of termination of growth p. The experimental data are not sufficient to refine the model to highest resolution.

Theoretical analysis of properties of the model function (e.g., driving nucleotide polymerization) in the mathematical sequence space [25] shows, in principle, the validity of this function as applied to branched pseudorandom walks [26,27]. Operations of modelling were made for

arbitrary nucleotide sequences. We found that this kind of pseudo-random walks can be associated with frequency distributions M between the moments of probability of error sequences $M(l, N)$ and the moment of probability of the target sequence $M(N, N)$ with $l = \{N\}$.

2.2. Experimental validation of the growth model of chemical solid-supported oligonucleotide synthesis

For experimental validation of the model as described above, we consider here P as a probability which designates the relative frequency of the stochastic variable $\mathscr{L}(i)$. $\mathscr{L}(i)$ is completely characterized by the cumulative probability distribution $\Phi(l, N) = P(\mathscr{L} \leq l)$ with

$$\Phi(l, N) = \begin{cases} \sum_{l_j \leq 1} M(l_j, N) \text{ for discrete } \mathscr{L}: \\ \quad \begin{pmatrix} l_1 & l_2 & \cdots & l_n \\ M(l_1, N) & M(l_2, N) & \cdots & M(l_n, N) \end{pmatrix} \\ \quad \text{with} \begin{cases} M(l_j, N) = P(\mathscr{L} = l_j) \\ \sum_{l_j=1}^{N} M(l_j, N) = 1 \end{cases} \\ \int_1^l M(l, N)\, \mathrm{d}l \text{ for continuous } \mathscr{L}: \\ \quad \text{probability density } M(l, N) \\ \quad \text{with} \begin{cases} M(l, N) \geq 0 \\ \int_1^N M(l, N)\, \mathrm{d}l = 1 \end{cases} \end{cases} \quad (3)$$

The definition of a continuous stochastic variable $\mathscr{L}$ makes sense, because a suitable meaning for this property is, for example, the obtained resolution of the experimental technique used for separation of $\mathscr{L}$. Further, these formulations permit the quantitative evaluation of the synthesis parameters N, d and p from the HPLC experiments.

2.3. Fractal dimension D_a of a linear non-enzymatic nucleotide polymerization process

The dynamics of chemical solid-phase oligonucleotide synthesis is naturally expressed in terms of the exponent $D_a(N)$. The general solution of $D_a(N)$ for discrete $\mathscr{L}$ indicates that whenever N is scaling, $D_a(N)$ has also a geometric interpretation over a D_a-dimensional observing set. In this case, $D_a(N)$ is of geometric fractal dimensionality [22]. This is the content of Eq. (5) with "embedding" dimension 2.

If the generating function $M(l, N)$ satisfies the condition of Eq. 1 and also Eq. 2, and if

$$\Phi(l, N) = \sum_{l_j \leq l} M(l_j, N) \quad (4)$$

then

$$D_a(N) = 2 - N\left(\frac{\partial\{\ln[\Phi(N-1, N) - \Phi(1, N)]\}}{\partial N}\right) \quad (5)$$

With the additional restriction of $d(i) = d_0 =$ constant (or $\bar{d}$, average), our formalism implies that

$$D_a(N) = 2 - N \cdot \frac{d_0^{N-1}}{1 - d_0^{N-1}} \cdot \ln\left(\frac{1}{d_0}\right) \quad (6)$$

In practice, the additional restriction of $d(i) = d_0 = \bar{d} =$ constant often applies to short target oligonucleotides (in case of routine chemical oligonucleotide synthesis). We found here that for our models

$$\lim_{d_0 \to 0} D_a(N) = 2 \quad (7)$$

$$\lim_{d_0 \to 1} D_a(N) = 2 - \frac{N}{N-1} \quad (8)$$

$$\lim_{d_0 \to 1} \lim_{N \to \infty} D_a(N) = 1 \quad (9)$$

There is another interesting property of $D_a(N)$ that is of practical relevance. Let $I = \{i \in \mathbb{R}: 1 \leq i < N\}$, then the non-empty set I is obviously bound above and below. If $\mathscr{L}(i)$ is either a discrete or a continuous stochastic variable and there is a real function $d(i)$ so that for every $i \in I$

$$\prod_{i \subset \{1, N\}} d(i) = M(N, N) \quad (10)$$

with $0 \leq d(i) < 1$, and if

$$1 - M(N, N) = P(1 \leq \mathscr{L} < N) \quad (11)$$

then

$$D_a(N) = 2 - N\left(\frac{\partial\{\ln[1 - M(N, N)]\}}{\partial N}\right) \qquad (12)$$

Consequently, mappings for $\Sigma_{l_j \leq l} M(l_j, N)$, $\int_1^l M(l, N)\, dl$ and $\Pi_{i \in [1,N)} d(i)$ are equivalent in terms of their dynamics $D_a(N)$. For example, computing the analytical function for all truncated or failure sequences measured by a separation technique might be a hopeless task, but $D_a(N)$ guarantees that these sequences are considered and, moreover, $D_a(N)$ gives a measure of all error sequences. Eq. 12 shows the power of symbolic dynamics [24] used.

For constant values of d and p (d_0, p_0), the arguments for Eqs. 5 and 12 were given earlier. We show that if

$$\sum_{l=1}^{N-1} M(l, N) = 1 - M(N, N) = C\left(\frac{1}{N}\right)^{D_a(N)-2} \qquad (13)$$

then, by changing the mathematical values of $N \to \infty$, we go to Eqs. 5 and 12. C is a constant of proportionality. The real, physical functions $d(i)$ and $p(i)$ are similar to those shown in Ref. [23], where for short oligonucleotides (e.g., $N \leq 50$) constant parameters are used.

3. Experimental

3.1. Reagents

5′-O-Dimethoxytrityldeoxynucleoside-3′-O-(2-cyanoethyl)-N,N-diisopropylaminophosphanes were purchased from Millipore, MWG Biotech, Roth and Applied Biosystems. The standard solutions for oligonucleotide synthesis (activation, capping, oxidation, detritylation) were also obtained from these companies. The phosphoramidites were dissolved in acetonitrile with water content ≤10 ppm (HPLC-grade acetonitrile was obtained from Merck) at final concentrations of 0.1 *M*. Acetonitrile (impurities ≤30 ppm) from Roth was used for routine syntheses. Other reagents were of analytical-reagent grade.

3.2. Chemical solid-phase oligonucleotide synthesis

Syntheses were performed on Applied Biosystems Model 380B and 394 DNA synthesizers, and on Pharmacia–LKB Gene Assembler Special/4 Primers with integrated software packages according to 0.2 μmol standard cycles (small scale cyanoethyl cycle 103a of version 2.0 or 2.01 for ABI 380B synthesizer or 0.2 μmol cyanoethyl cycle for ABI 394 synthesizer of Model 392/394T system software ROM version 2.0). Using the styrene-grafted polytetrafluoroethylene supports P_{29} and P_{19}, coupling and washing steps during the initial three elongations were run on the ABI Model 380B synthesizer for double the time (60 s) compared with the 0.2 μmol standard synthesis cycles. Polymer-supported oligonucleotides were cleaved from supports and deprotected by treatment with 28% aqueous ammonia solution for 5–12 h at 55°C. The following 30mer oligonucleotide sequences were synthesized: (i) dC_{30}; and (ii) 5′ d GAA CTG ACT GGT CAA CGT CTG CGT GAA GGT. The heterooligonucleotide target sequence is the initial part of the codogenous strand of the human γ-lipotropin gene [28].

3.3. Polymer supports

Polystyrene-grafted polytetrafluoroethylene

The procedures and results of the grafting reactions of styrene on to polytetrafluoroethylene are described elsewhere [29–31]. A coarse-grained polytetrafluoroethylene powder (Polychrom I) with an average particle diameter of 500–1000 nm was chosen as starting material. The grafting of styrene on to this material was accomplished by ^{60}Co irradiation in an eddy flow reactor in which a styrene-saturated nitrogen stream flowed through the polytetrafluoroethylene gravel. This procedure yielded products highly homogeneous in degree of grafting and avoided the formation of large amounts of styrene homopolymers. The degree of grafting (% styrene) was determined from the ratio of intensities of IR bands at 1560 and 1610 cm^{-1}, compared with a calibration graph or from

corrected values of carbon content in elemental analysis. The improved pathway of support functionalization with a long alkylamine spacer is briefly described in Ref. [32]. Characteristics of supports P_{19} and P_{29} are reported in Ref. [32]. P_{19} has a degree of grafting of 5–7% styrene, corresponding to 42–48 μmol nucleoside/g support, whereas the degree of grafting of P_{29} is 2–3%, corresponding to 11–18 μmol nucleoside/g support.

Polystyrene primer support

This support was obtained from Pharmacia–LKB Biotechnology (Uppsala, Sweden). Nucleoside loading was 20.8 μmol/g support material. Derivatized polystyrene particles were optimized for oligonucleotide synthesis up to 50 bases in length [33].

Controlled-pore glass (CPG)

In the case of CPG materials, the accessibility of the growing oligonucleotide chains is connected with the pore size. Small-pore CPG allows higher loading. Long chains can be made only with low-capacity supports [34]. For syntheses of 30mers, CPG 500 Å (small-pore CPG) was routinely used with 27 μmol nucleoside/g support or with 43 μmol nucleoside/g support. CPG 500 Å was purchased from Eppendorf-Biotronic and Roth. CPG 1000 Å was obtained from Millipore. PGL 1000 Å (pore glass) was purchased from Schuller in unmodified form and functionalized with a long alkylamine spacer in our laboratory. The nucleoside loading was 28 μmol/g support.

3.4. Estimation of repetitive trityl yields

Experimental estimates of relative average coupling yields were obtained by relating the 495-nm absorption of the solution of the individual detritylation steps to that of the detritylation of the support-bound nucleoside. In order to obtain correct yields, the solutions resulting from cleavage of the dimethoxytrityl protecting group were separately collected and absorptions at 495 nm were determined manually after acidification by addition of 0.1 *M* *p*-toluenesulfonic acid in acetonitrile [34].

3.5. Determination of oligonucleotide concentrations

Oligonucleotide concentrations in crude products were determined directly after cleavage from polymer support. Absorbance (*A*) was measured at 260 nm on a Pharmacia–LKB Ultrospec Plus spectrophotometer or on a Beckman DU 7500 spectrophotometer with integrated software.

3.6. HPIEC analysis

HPIEC analyses, as represented in Fig. 1, were carried out with an Applied Biosystems Model 152A system. The HPLC-column, Superformance 50–10 LiChrospher 4000 DMAE (5 μm), was a gift from Merck (Darmstadt, Germany). The following conditions were used: detection, UV at 254 nm; flow-rate, 1.5 ml/min; room temperature; eluent, buffer A, 20 m*M* sodium acetate (pH 6.5)–acetonitrile (80:20), buffer B, 20 m*M* sodium acetate (pH 6.5)–acetonitrile (80:20)–1 *M* LiCl. A linear gradient was applied as indicated in Fig. 1.

HPIEC analyses reported in Fig. 2 and Tables 1 and 2 were carried out on a Bio-Rad Model 2700 system (Software Series 800 HRLC System, Version 2.30.1a) equipped with a column oven and with an AS-100T HRLC automatic sampling system. A Bio-Rad UV-1806 UV–Vis detector was used. A Mono Q HR 5/5 anion-exchange column was obtained from Pharmacia–LKB Biotechnology. The following conditions were used: detection, UV at 260 nm; flow-rate, 1 ml/min; temperature, 60°C; eluent, buffer A, 10 m*M* NaOH–0.1 *M* NaCl (pH 11), buffer B, 10 m*M* NaOH–1 *M* NaCl (pH 11). Linear gradients from 0 to 100% buffer B were applied over 30 and 40 min, respectively. Water used for mobile phases was purified with a Milli-Q Plus water-purification system (Millipore, Bedford, MA, USA). Sample solutions were filtered through a Millipore filter (pore size 0.45 μm).

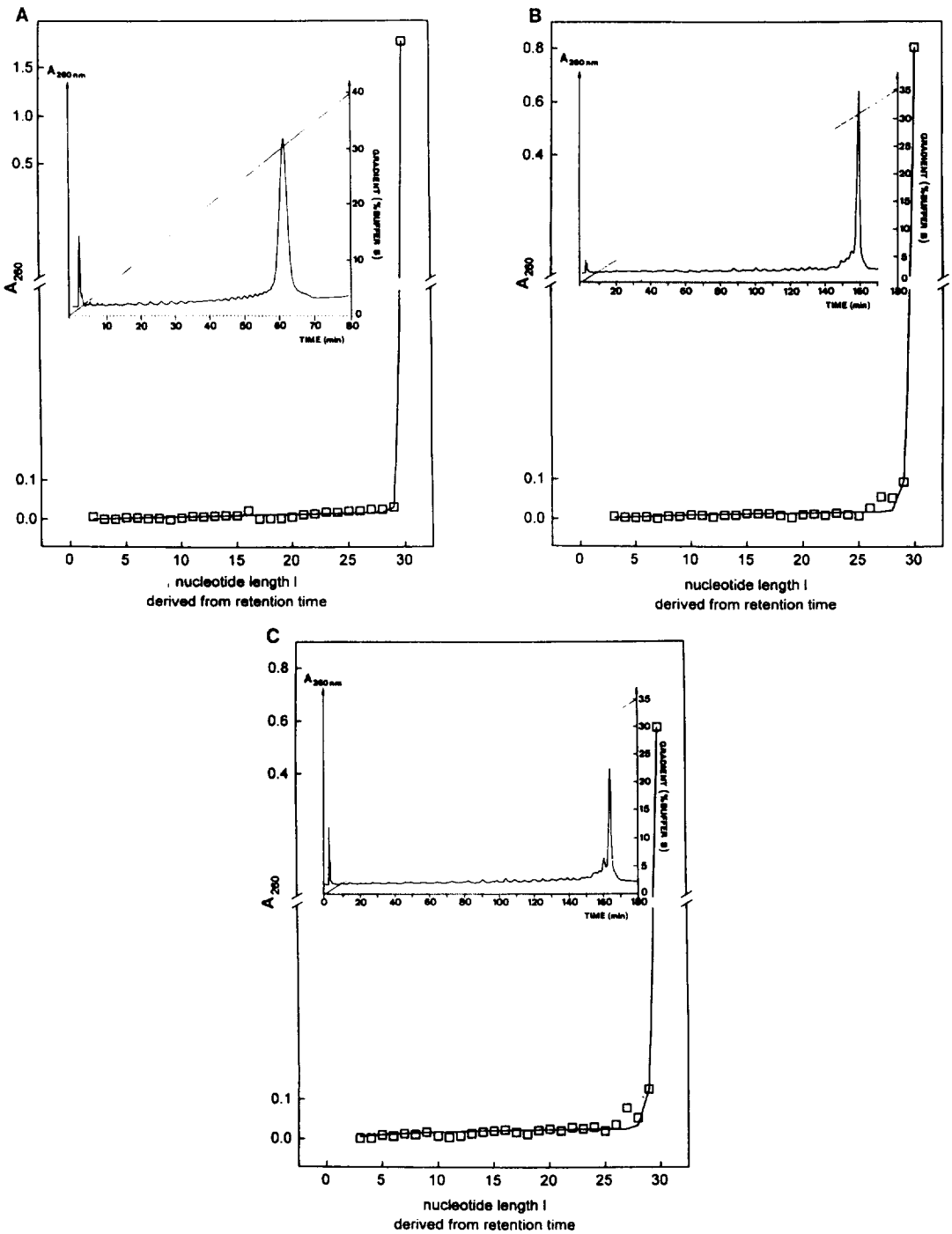


Fig. 1. Experimental anion-exchange HPLC elution profiles (insets) of crude products of (A) dC_{30} synthesis on CPG 500 Å support (Eppendorf-Biotronic), (B) 30mer heterooligonucleotide synthesis on P_{29} support and (C) 30mer heterooligonucleotide synthesis on P_{19} support. Experimental values (□) of chromatograms are compared with theoretical values (——) obtained by proper setting of synthesis parameters d_0 (constant coupling efficiency) and p_0 (constant capping efficiency). 1.25 AU of crude products were applied in (B) and (C).

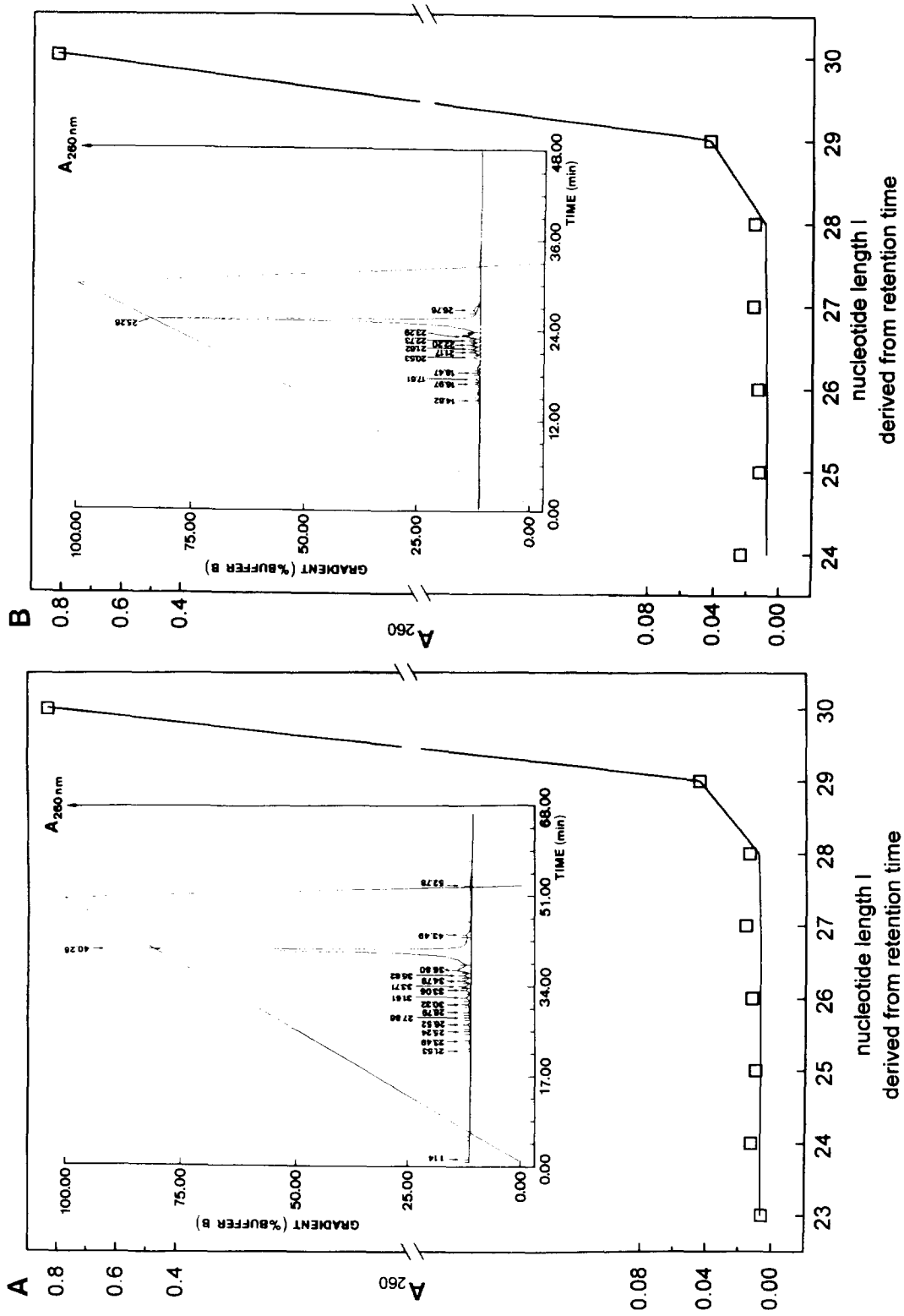
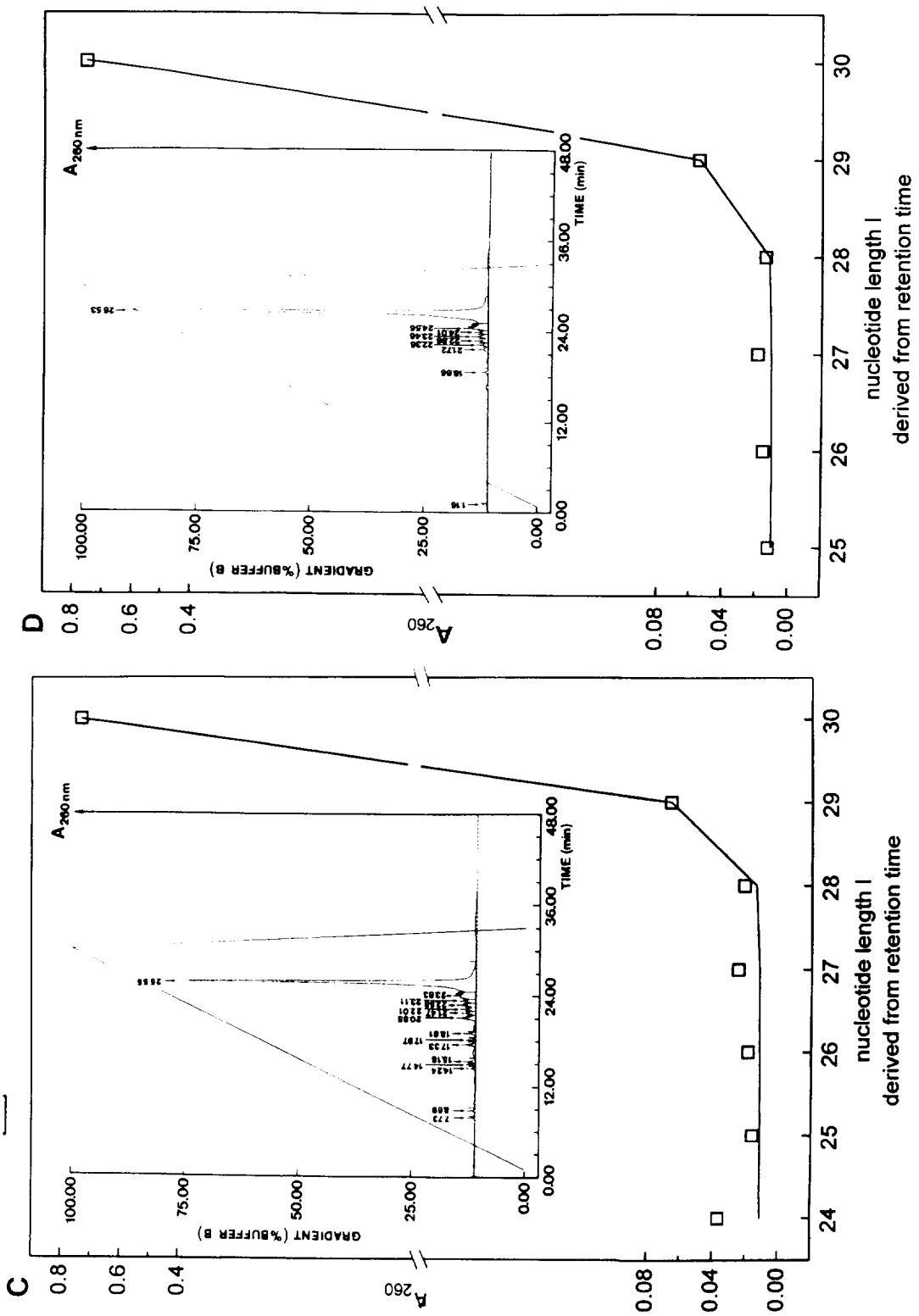


Fig. 2. (Continued on page 412)


C
D
A260
A260nm
GRADIENT (%BUFFER B)
TIME (min)
nucleotide length l
derived from retention time

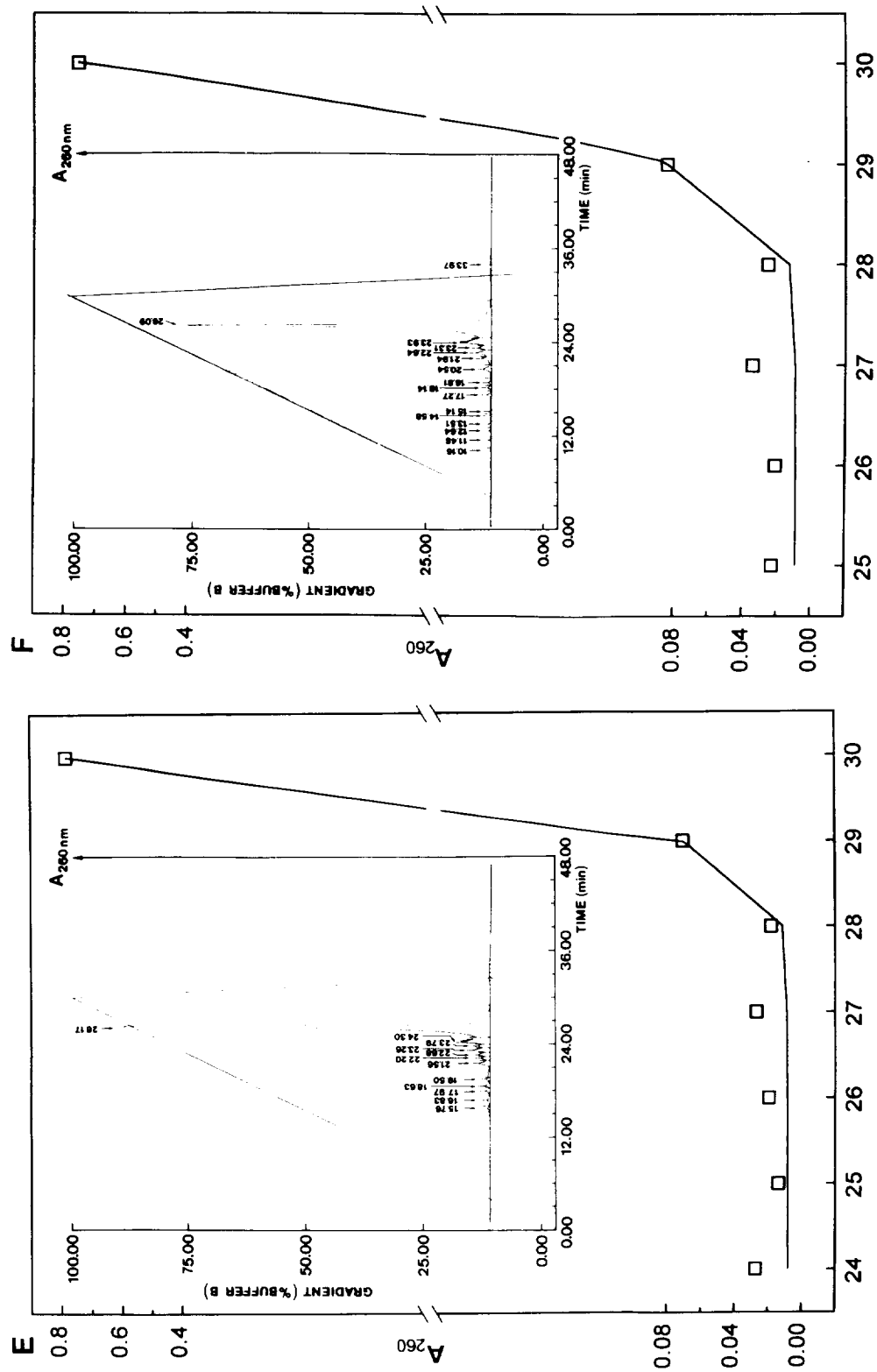


Fig. 2. Determination of parameters d_0 (constant coupling efficiency) and p_0 (constant capping efficiency) characterizing the course of chemical oligonucleotide synthesis. Theoretical curves (——) were fitted to experimental values (□) taken from anion-exchange HPLC elution profiles (insets) of crude products of 30mer heterooligonucleotide. 1 AU of crude products were analysed. (A) CPG 500 Å support obtained from Roth, $d_0 = 0.98897$, $p_0 = 0.85950$; (B) CPG 1000 Å support obtained from Millipore, $d_0 = 0.98789$, $p_0 = 0.87755$; (C) PGL 1000 Å support, $d_0 = 0.98149$, $p_0 = 0.85850$; (D) polystyrene primer support obtained from Pharmacia–LKB Biotechnology, $d_0 = 0.98373$, $p_0 = 0.87600$; (E) P_{29} support [32], $d_0 = 0.98613$, $p_0 = 0.80011$; (F) P_{19} support [32], $d_0 = 0.98455$, $p_0 = 0.77500$.

Table 1
Comparison of a typical result for solid-supported oligonucleotide syntheses performed on Applied Biosystems Model 380B and 394 DNA synthesizers in the optimum reaction mode

Synthesizer	N	DMTr[a] (%)	d_0	p_0	D_a	$\frac{2-\frac{N}{N-1}}{2-N\cdot\left(\frac{d_0^{N-1}}{1-d_0^{N-1}}\cdot\ln\frac{1}{d_0}\right)}$ [b]
Model 380B	30	66.92	0.98908 ±0.00090 ($n=3$)	0.86450 ±0.00999 ($n=3$)	1.12141 ±0.01215 ($n=3$)	0.86105 ±0.00935 ($n=3$)
Model 394	30	85.80	0.98846 ±0.00101 ($n=3$)	0.85471 ±0.01664 ($n=3$)	1.12983 ±0.01353 ($n=3$)	0.85465 ±0.01019 ($n=3$)

The support material was CPG 500 Å obtained from Roth. The 30mer heterooligonucleotide target was synthesized. 1 AU of crude products were analysed by anion-exchange HPLC.
Randomization of different syntheses and of different HPLC runs. Estimates were calculated from a sample using the following equations for the mean ($\bar{x}$) and for the standard deviation (s) of the mean from a normally distributed population: $\bar{x}=\frac{1}{n}\cdot\Sigma_{j=1}^{n}x_j$, $s=\sqrt{\frac{1}{n-1}\cdot\Sigma_{j=1}^{n}(x_j-\bar{x})^2}$.
[a] Total trityl yield after $N-1$ cycles (see Experimental).
[b] The relation $[2-N/(N-1)]/D_{a,\text{measurable}}$ is a universal measure for multi-cycle synthesis conditions (for explanation, see text).

4. Results and discussion

The crude product of synthesis of a dC_{30} obtained after cleavage from the support and deprotection was chosen as the sample species (Fig. 1A). We first studied a homooligonucleotide synthesis in order to avoid different elution behaviours of truncated or failure sequences with increasing chain length. The separation of the dC_{30} was examined using the Superformance 50–10 LiChrospher 4000 DMAE (5 μm) column. As shown in the inset in Fig. 1A, a component elution was obtained. The peaks of single nucleotides and of dinucleotides were checked by comparison with standards. The products of protecting group cleavage were also eluted with the first and second peaks. However, the retention time of the target was very close to that of error sequences of length $N-1$. The open squares are the data from the HPLC experiment. The solid line without symbols is the result obtained by our theory. It is worth noting that A values of a clean separation of N and $N-1$ sequences are most important for the computer simulation of anion-exchange HPLC profiles of crude products of short oligonucleotides.

The next series of experiments together with computer simulation results are shown in Fig. 1B and C. Samples of syntheses of a heterooligonucleotide target sequence (see Experimental) were used. Although selected from systematic investigations, the mobile phase for complete component elution was unable to give an adequate resolution of $N-1$ peaks. The slight discrepancies observed between the calculated parameters can be explained by this experimental factor.

The insets in Fig. 2A–F show HPLC analyses with the Mono Q HR 5/5 anion-exchange column. This system is optimized for separation of N and $N-1$ peaks. Further, we checked the "purity" of the N and $N-1$ peaks by re-chromatography. The occurrence of minor peaks which seem to correspond to products longer than the target sequence has been reported previously (see, for instance, the examples shown in Ref. [35]), but explanations are not at hand. These results indicated that, in each instance, the experimental values are within acceptable limits.

Table 2
Influence of different support materials on synthesis result

Support	N	DMTr[a] (%)	d_0	p_0	D_a	$\frac{2-\frac{N}{N-1}}{2-N\cdot\left(\frac{d_0^{N-1}}{1-d_0^{N-1}}\cdot\ln\frac{1}{d_0}\right)}$ [b]
CPG 500 Å (Roth)	30	77.36 ($n=2$)	0.98877 ±0.00092 ($n=6$)	0.85961 ±0.01339 ($n=6$)	1.12562 ±0.01239 ($n=6$)	0.85785 ±0.00942 ($n=6$)
CPG 1000 Å (Millipore)	30	69.47 ($n=1$)	0.98655 ±0.00122 ($n=3$)	0.88653 ±0.00910 ($n=3$)	1.15527 ±0.01621 ($n=3$)	0.83586 ±0.01178 ($n=3$)
PGL 1000 Å	30	58.90 ($n=1$)	0.97989 ±0.00164 ($n=3$)	0.87974 ±0.01878 ($n=3$)	1.24033 ±0.02018 ($n=3$)	0.77857 ±0.01265 ($n=3$
Polystyrene primer support (Pharmacia)	30	86.33 ($n=3$)	0.98699 ±0.00488 ($n=7$)	0.83528 ±0.05336 ($n=7$)	1.14820 ±0.06309 ($n=7$)	0.84296 ±0.04387 ($n=7$)
P_{29}	30	70.94 ($n=3$)	0.98516 ±0.00421 ($n=9$)	0.76151 ±0.07125 ($n=9$)	1.17254 ±0.05550 ($n=9$)	0.82514 ±0.04025 ($n=9$)
P_{19}	30	61.40 ($n=2$)	0.98591 ±0.00494 ($n=7$)	0.75857 ±0.09647 ($n=7$)	1.16242 ±0.06413 ($n=7$)	0.83275 ±0.04525 ($n=7$)

The 30mer heterooligonucleotide target was synthesized. 1 AU of crude products were analyzed by anion-exchange HPLC. Randomization of different syntheses and of different HPLC runs. Estimates were calculated from a sample using following equations for the mean ($\bar{x}$) and for the standard deviation (s) of the mean from a normally distributed population: $\bar{x}=\frac{1}{n}\cdot\Sigma_{j=1}^{n}x_j$, $s=\sqrt{\frac{1}{n-1}\cdot\Sigma_{j=1}^{n}(x_j-\bar{x})^2}$.

[a] Total trityl yield after $N-1$ cycles of different syntheses (see Experimental).

[b] See Table 1, footnote b.

By assuming given values of d_0 and p_0, theoretical curves were calculated and iterated to best fit the experimental data. By means of our theory we thus generated parameters d_0 *and* p_0 from the experimental HPLC elution profiles obtained; d_0 is the average (constant) coupling (elongation) probability of each synthesis cycle and p_0 is the average (constant) capping (termination) probability of each synthesis cycle. These parameters depend on the number of reaction cycles i and on internal or external conditions, which all are taken into account by proper scaling of d and p. The d_0 values obtained from experiments were used for calculating the fractal dimension D_a of the growth process.

The problem of comparison in growth patterns such as chemical oligonucleotide synthesis is, in essence, a matter of converting the experimentally accessible information into otherwise inaccessible information about the real process. By the differential relation D_a we produce a geometric image of error sequences formed by the course of chemical oligonucleotide synthesis. This physical growth pattern is directly related to the efficiency of error sequences dynamics. Different syntheses of oligonucleotides and single-stranded DNA sequences can now be directly compared. If their values of D_a are equal, they are of the same performance in all steps of oligonucleotide growth. Further, the local fractal

dimension D_a quantifies the influence on nucleotide length N of different target sequences. The measurable values of the exponent D_a can be normalized to an idealized synthesis of constant growth without error production. In the case of a synthesis of constant growth without error production, D_a becomes $D_a = 2 - N/(N-1)$ (Eq. 8). Hence the relation $[2 - N/(N-1)]/D_{a,measurable}$ is a universal quantitative measure for multi-cycle synthesis conditions. This measure is revealed by two-dimensional graphs of functions M. The function M specifies the growth process under study. For example, M expresses chemical solid-supported oligonucleotide synthesis. It is clear from Eq. 12 that the relation $[2 - N/(N-1)]/D_{a,measurable}$ can also be a quantitative measure for chemical oligonucleotide synthesis carried out in solution. Obviously, the coupling efficiency (d) or individual yields (e.g., repetitive trityl yields) do not allow the quantitative characterization of chemical oligonucleotide synthesis of different target length. They reflect the quantitative influence on any reaction step only.

The coupling efficiency d of chemical oligonucleotide synthesis is not a constant value in the proper sense. It is convenient to use constant values for solid-phase synthesis of short oligonucleotides. In this case, our universal quantitative measure becomes

$$\frac{2 - \dfrac{N}{N-1}}{2 - N \cdot \dfrac{d_0^{N-1}}{1 - d_0^{N-1}} \cdot \ln\left(\dfrac{1}{d_0}\right)} \qquad (14)$$

The parameters N and d_0 of the relationship can be combined with experiments. The relationship can be extended to average values of d ($\bar{d}$). Hence more complicated growth systems are described with the same simplicity.

In Table 1 we compare a typical result of chemical solid-phase oligonucleotide syntheses performed on the Applied Biosystems Model 380B and 394 DNA synthesizers in the optimum reaction mode. In our study the parameters d_0, p_0 and the measures D_a and $[2 - N/(N-1)]/D_{a,measurable}$ did not reveal any differences related to the apparatus used for these syntheses. One of the major interests has been to explore new polymer support systems, because the good accessibility of a reactive polymer support is responsible for the overall efficiency of a chemical oligonucleotide synthesis.

Table 2 shows the influence of different support materials on the synthesis results. First we considered the most widely used support materials such as controlled-pore glass and polystyrene support (Primer Support from Pharmacia). These materials show similar efficiencies for producing high yields of the target sequence {$[2 - N/(N-1)]/D_{a,measurable}$ values in column 7 of Table 2}. It is noteworthy that the capping probability p_0 is much lower for syntheses with styrene supports (polystyrene primer support) than for syntheses carried out with controlled-pore glass. The decrease in p_0 values may be due to free functional groups of styrene supports which did not react during the further functionalization of support materials or to a slight steric hindrance of chains resulting in decreased accessibility of some growing ends. Differences in capping efficiency may shift the composition of the truncated and failure sequences toward longer chains. A practical consequence of this finding might be that the isolation of the target sequence N from error sequences $N-1$, $N-2$, $N-3$, $N-4$ in crude products can be more easily handled in the case of CPG synthesis. Such differences of composition may be of importance for application of solid-phase synthesized oligonucleotides, especially in the biomedical field. A zero-capping step prior to synthesis is to be recommended to block unreacted COOH and NH_2 groups on styrene supports. Recently, we have described a versatile support material based on tetrafluoroethylene powder grafted with polystyrene. A feature of this support material is that the oligonucleotide reactions proceed only in the polystyrene surface coat, the thickness of which can be easily regulated by the grafting procedure [32]. We have measured the efficiency of this support material using the same criteria as applied above for CPG and polystyrene supports. The results summarized in Table 2 suggest that this new material is comparable in its performance to those supports which are established and most widely used in oligo-

nucleotide chemistry. In principle, the physical model applied to D_a is sensitive enough to distinguish various support materials. However, for short oligonucleotides the theoretical result is involved in an optimum experimental resolution between N and $N-1$ peaks obtained in the chromatogram.

A phenomenon observed in the experiments and proved by simulations of anion-exchange HPLC elution profiles is that A values of $N-1$ error sequences must decrease with the increase in d values. Similar considerations hold for the influence of p_0 values. Usually the yield of the target sequence is determined only by measured absorbance at 260 nm. For the calculation of d values in the literature this may lead to an overestimation. Monitoring the trityl yield (column 3 of Table 2) also provides some obstacles [36] which may result in an overestimation of d values. In our homoeodynamic model of solid-supported oligonucleotide synthesis, the yields of all truncated or failure sequences and the yield of the target are considered for obtaining d and p values in each reaction cycle.

Automation and optimization of chemical reactions on solid support materials allow syntheses of chains of extended nucleotide length and the preparation of oligonucleotides in large amounts [32,37–39]. As an example, we previously analysed experimentally and theoretically the attempted synthesis of an unusually long 238mer sequence. Neither HPLC, capillary electrophoresis nor polymerase chain reaction techniques in the crude product of chemical synthesis followed by sequencing led to the isolation of any amount of the 238mer target sequence. Nevertheless, the calculation of d and p could be done by simulation of the overall pattern of experimental nucleotide length distribution of error sequences obtained by quantitative agarose gel electrophoresis [25]. In a parallel approach, this sequence has been prepared in the meantime by splint ligation of two 119mer oligonucleotides with a 40 nucleotide splint. From the resulting 238 nucleotide single-stranded DNA, the double-stranded fragment was amplified via PCR using special primers carrying, in addition, unique restriction sites for subsequent cloning in the BamHI and HindIII region of the PTZ18/19 plasmid [3]. Sequencing of clones proved the success of the preparation.

5. Conclusions

In the literature there are no investigations dealing with non-enzymatic and enzymatic nucleotide growth processes under the dynamic aspect of error sequences production. Here we analysed the dynamics of chemical oligonucleotide synthesis by a new fractal formalism. We are currently elaborating on our theoretical concept in the direction of both non-enzymatic and enzymatic nucleotide polymerization processes. In this paper, theoretical results are compared with the experimental separation of the products of solid-supported oligonucleotide synthesis by HPIEC. Fig. 2 shows that there is a good correlation between the results of HPIEC and theoretical calculations by assuming given values of coupling and capping efficiencies. One of the novel aspects of this work is that the distribution of solid-phase products which is based on the efficiencies of coupling and capping is now accessible to calculation. On this basis, we have presented first data that allow the evaluation of support systems for chemical synthesis of short oligonucleotides by our differential relationship of fractal dimension. From these results slight differences can be seen between the performances of different support materials (Table 2, column 7). It is possible that with more experimental material at hand, quantitative differences concerning the performance of the support systems will more clearly emerge. Our approach has the advantage that model parameters and measures derived from the theory possess physical and chemical significance for the multi-step process which we see through HPIEC.

Acknowledgements

We thank Dr. Wei-Guo Peng of Softlab (Munich, Germany) for writing the computer program used. The computer code was written in C algorithmic language on PC under DOS/Windows and VAX under VMS, compiled with

Turbo C^{++} from Borland and VAX C from Digital Equipment, respectively. We thank Ms. Anita Willitzer for skilful technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft and the Stiftung zur Förderung der molekularbiologischen Forschung Universität Ulm.

References

[1] J.W. Engels and E. Uhlmann, *Angew. Chem., Int. Ed. Engl.*, 28 (1989) 716–734.
[2] L.C. Klotz, R.W. Schatz, A.P. Kerr and C.R. Morris (Editors), *The Commercial Potential of Human Oligonucleotide Therapy*, Decision Resources, Burlington, 1992.
[3] Z. Földes-Papp, *Studies on Synthesis of Oligonucleotides and DNA Sequences: Dynamics of Error Sequences: Fractality and Experiments, Ph.D. Thesis*, University of Ulm, Ulm, 1994.
[4] E. Wickstrom (Editor), *Prospects for Antisense Nucleic Acid Therapy of Cancer and AIDS*, Wiley, New York, 1991.
[5] J.A.H. Murray (Editor), *Antisense RNA and DNA: a Comprehensive Guide to Successful Use of Antisense Nucleic Acids*, Wiley, New York, 1992.
[6] Z. Földes-Papp, *Gen. Physiol. Biophys.*, 11 (1992) 3–38.
[7] H. Seliger, R. Bader, E. Birch-Hirschfeld, Z. Földes-Papp, K.H. Gührs, M. Hinz, R. Rösch and C. Scharpf, *React. Polym.*, (1995) in press.
[8] H. Seliger, in S. Agrawal (Editor), *Methods in Molecular Biology, Vol. 20: Protocols for Oligonucleotides and Analogs: Synthesis and Properties*, Humana Press, Totowa, NJ, 1993, pp. 391–435.
[9] S. Agrawal (Editor), *Methods in Molecular Biology, Vol. 26: Protocols for Oligonucleotide Conjugates: Synthesis and Analytical Techniques*, Humana Press, Totawa, NJ, 1994.
[10] H. Kössel and H. Seliger, in W. Herz, H. Grisebach, G.W. Kirby (Editors), *Progress in the Chemistry of Organic Natural Products*, Vol. 32, Springer, Vienna, 1975, pp. 297–508.
[11] W.J. Warren and G. Vella, *BioTechniques*, 14 (1993) 598–606.
[12] R.L. Letsinger, J.L. Finnan, G.A. Heavner and W.B. Lunsford, *J. Am. Chem. Soc.*, 97 (1975) 3278–3279.
[13] S.L. Beaucage and M.H. Caruthers, *Tetrahedron Lett.*, 22 (1981) 1859–1862.
[14] M.H. Caruthers, G. Beaton, J.V. Wu and W. Wiesler, *Methods Enzymol.*, 211 (1992) 3–20.
[15] B. Mandelbrot, *Les Objets Fractals: Forme, Hasard et Dimension*, Flammarion, Paris, 1975.
[16] B.B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, New York, 1983.
[17] T.F. Nonnenmacher, in T.F. Nonnenmacher, G.A. Losa and E.R. Weibel (Editors), *Fractals in Biology and Medicine*, Birkhäuser, Basle, 1994, pp. 22–38.
[18] G. Baumann, A. Barth and T.F. Nonnenmacher, in T.F. Nonnenmacher, G.A. Losa and E.R. Weibel (Editors), *Fractals in Biology and Medicine*, Birkhäuser, Basle, 1994, pp. 182–189.
[19] A.B. Çambel, *Applied Chaos Theory: A Paradigm for Complexity*, Academic Press, Boston, 1993.
[20] A. Bunde and S. Havlin (Editors), *Fractals in Sciences*, Springer, Berlin, 1994.
[21] J.L. Casti, *Reality Rules: II, Picturing the World in Mathematics: The Frontier*, Wiley, New York, 1992.
[22] B.J. West, *Fractal Physiology and Chaos in Medicine*, World Scientific, Singapore, 1990.
[23] Z. Földes-Papp, A. Herold, H. Seliger and A.K. Kleinschmidt, in T.F. Nonnenmacher, G.A. Losa and E.R. Weibel (Editors), *Fractals in Biology and Medicine*, Birkhäuser, Basle, 1994, pp. 165–173.
[24] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings, Menlo Mark, CA, 1986.
[25] Z. Földes-Papp, W.-G. Peng, H. Seliger and A.K. Kleinschmidt, *J. Theor. Biol.*, (1995) in press.
[26] S. Redner and F. Leyvraz, in A. Bunde and S. Havlin (Editors), *Fractals in Sciences*, Springer, Berlin, 1994, pp. 212–217.
[27] G.H. Weiss, in A. Bunde and S. Havlin (Editors), *Fractals in Sciences*, Springer, Berlin, 1994, pp. 119–161.
[28] A.S.Y. Chang, M. Cochet and S.N. Cohen, in R. Håkanson and J. Thorell (Editors), *Biogenetics of Neurohormonal Peptides*, Academic Press, London, 1985, pp. 15–28.
[29] R. Weiss, E. Birch-Hirschfeld, W. Wittkowski and K. Friese, *Z. Chem.*, 26 (1986) 127–130.
[30] W. Witkowski, E. Birch-Hirschfeld, R. Weiss, V.F. Zarytova and V.V. Gorn, *J. Pract. Chem.*, 326 (1984) 320–328.
[31] E. Birch-Hirschfeld, Z. Földes-Papp, K.-H. Gührs, R. Weiss and H. Seliger, presented at the *3rd Swedish–German Workshop on Nucleic Acid Synthesis, Structure and Function, Uppsala, 1992*, abstracts, p. 35.
[32] E. Birch-Hirschfeld, Z. Földes-Papp, K.-H. Gührs and H. Seliger, *Nucleic Acids Res.*, 22 (1994) 1760–1761.
[33] *Product Bulletin Primer Support*, 56-1142-52, edition AA, Pharmacia, Uppsala.
[34] H. Seliger, A. Herold, U. Kotschi, J. Lyons and G. Schmidt, in K.S. Bruzik and W.J. Stec (Editors), *Biophosphates and Their Analogues—Synthesis, Structure, Metabolism and Activity*, Elsevier, Amsterdam, 1987, pp. 43–58.
[35] W.J. Warren and G. Vella, in S. Agrawal (Editor), *Methods in Molecular Biology, Vol. 26: Protocols for Oligonucleotide Conjugates: Synthesis and Analytical Techniques*, Humana Press, Totowa, NJ, 1994, pp. 233–264.

[36] *Applied Biosystems User Bulletin, 40 nanomole Polystyrene: New Highly Efficient DNA Synthesis Columns, No. 61*, ABI, 1991.
[37] R.W. Barnett and H. Erfle, *Nucleic Acids Res.*, 18 (1990) 3098.
[38] R.B. Ciccarelli, P. Gunyuzlu, J. Huang, C. Scott and F.T. Oakes, *Nucleic Acids Res.*, 19 (1991) 6007–6013.
[39] Z. Földes-Papp, R. Rösch, F. Ramalho Ortigao, M. Hinz, S. Conrad, R. Weiss, E. Birch-Hirschfeld and H. Seliger, presented at the *3rd Swedish–German Workshop on Nucleic Acid Synthesis, Structure and Function, Uppsala, 1992*, abstracts, p. 60.