Zeno Földes-Papp[1]
Gerd Baumann[1]
Eckhard Birch-Hirschfeld[2]
Holger Eickhoff[3]
Karl Otto Greulich[3]
Albrecht K. Kleinschmidt[4]
Hartmut Seliger[5]

[1] *Department of Mathematical Physics, University of Ulm, D-89069 Ulm, Germany*

[2] *Institute of Virology, University Jena, D-07745 Jena, Germany*

[3] *Department of Single-Cell and Single-Molecule Techniques, Institute of Molecular Biotechnology, D-07745 Jena, Germany*

[4] *University of Ulm, D-89069 Ulm, Germany*

[5] *Section for Polymers, University of Ulm, D-89069 Ulm, Germany*

The Analysis of Oligonucleotide Preparations by Fractal Measures

Received 24 February 1997; accepted 15 October 1997

Abstract: *In this paper we put forward improved mathematical methods for detecting synthesis parameters in connection with analyzing crude products of chemically synthesized oligonucleotides. The crude products experimentally sampled are separated by high-performance capillary electrophoresis and ion-exchange high-performance liquid chromatography. The measured separation profiles of experimental syntheses can be expressed as target and nontarget yields; they are characterized by a few parameters. These parameters account for nonlinear synthesis equations that are solvable by employing iteration procedures. We provide here a theoretical as well as computational analysis based upon specific models for stepwise chain growth.*

Under nonconstant (nonuniform) conditions we use here an exponential form of growth, with different expressions for calculating the fractal dimension of the biochemical process under study. Step lengths of parameter variations in an interval of finite length have to be adjusted properly to find convergent solutions in a mathematical, regularly four-dimensional

Correspondence to: Zeno Földes-Papp, Ph.D., Ph.D., at present address—Karolinska Institute, Department of Medical Biophysics, MBB, S-171 77 Stockholm, Sweden; e-mail: zeno@mango.mef.ki.se

Contract grant sponsor: Foundation for the Advancement of Molecular Biology

Biopolymers, Vol. 45, 361–379 (1998)

CCC 0006-3525/98/050361-19

parameter space. It is conceivable to have most, if not all, of the calculating and plotting carefully done by a computer.

This analysis represents the experimental situation up to 65-mer target oligonucleotides analyzed so far. We thus obtain the dynamics of the polymerization process limited in number by fractal models. The advantage, calculating these new methods as compared to qualitatively judged experimental methods, lies in the satisfactory evaluation of crude products, also of large amounts, of syntheses of these biopolymers. Biopoly 45: 361–379, 1998

Keywords: *fractal measures; high-performance capillary electrophoresis; ion-exchange high-performance liquid chromatography; high-performance liquid chromatography; oligodeoxyribonucleotides; oligoribonucleotides; oligonucleotides; polymerization processes; chemical synthesis; dynamics*

INTRODUCTION

Synthetic single strands of DNA are being used for a variety of applications, e.g., as gene fragments, primers, and more recently, as antisense and antigene oligonucleotides for potential therapeutic uses.[1,2] Nucleotide building blocks are chemically joined together in vitro to form usually short oligonucleotides that have been designed to recognize and bind to specific genes. Coupled with the solid-phase strategy for oligonucleotide synthesis[3] introduced through the work of Merrifield[4] and Letsinger,[3] the chemical methods are now the basis for technologies to produce such drugs. Presently, the synthesis of oligonucleotides and their analogues is performed in automated, microprocessor-controlled DNA synthesizers using mainly the phosphoramidite[5] or H-phosphonate[6] chemistry. For preclinical and clinical studies there is a need for scaling up oligonucleotide synthesis from the milligram to kilogram range.[7] Such clinical use opens not only the availability of large quantities of therapeutic oligonucleotides, but also the necessity of analyzing the sequence composition of such preparations. It is known that solid-phase methods produce the target oligonucleotide in a mixture with truncated and failure sequences called error sequences. In the context of oligonucleotide production, it is essential to minimize error sequences (coding variabilities).

In particular, the quantitative analysis of synthetic oligonucleotides and DNA preparations has already been a demand for modern molecular biology[8] applying rational or irrational drug design to therapeutic oligonucleotides, but the determination of the number and weight fractions of individual truncated or erroneous by-products with different sequence failures has, so far, not been reached. The quantitation methods for synthesis predominantly in use, although well studied, have limitations in many respects. A linear, additive formulation of models that underlies the usual analysis has inadequacies that, once revealed, point the way to better models.[9] Therefore, we attempted a fractal interpretation of multicyclic synthesis.[10,11]

We have developed a new approach which involves improved but useful nonlinear mathematical methods for detecting synthesis parameters by high-performance capillary electrophoresis, ion-exchange high-performance liquid chromatography, and highly resolved separation profiles from gel electrophoresis. The calculation of the characteristic propagation probability function d and the characteristic termination probability function p is central to the modeling of the growth process itself from experimentally determined separation profiles. Although chemical oligonucleotide synthesis leads to a large number of nontarget or intermediate products, the growth pattern can be described by power laws.[9,12,13] The growth process here is demonstrated in a few examples, expressed in fractional powers of nucleotide length N of the largest optimal, desirable component (target) in the nucleotide region of interest. The resulting fractal requires only the single growth parameter function d. Our methods lead to satisfying models of quantification of multicyclic syntheses of biopolymers, also in large amounts. Chemical oligonucleotide stepwise varied preparations of crude products become useful simulations, in terms of experimental synthesis parameters, and vice versa.

EXPERIMENTAL

Chemical Solid-Phase Oligonucleotide Synthesis

Chemical oligonucleotide syntheses on solid support were done by phosphoramidite chemistry using the automated DNA synthesizer Pharmacia Gene Assembler Plus according to standard manufacturer's recommendation and

the Applied Biosystems Model 380B and 394 DNA synthesizers according to special protocols. Reagents, polymer supports (polystyrene-grafted polytetrafluoroethylene supports, polystyrene primer support), and special protocols used are described in Refs. 14–16. The macroreticular polyvinylacetate support used is described in Ref. 17.

Ion-Exchange High-Performance Liquid Chromatography (HPIEC)

The high-performance liquid chromatography (HPLC) column Superformance 50-10 LiChrospher 4000 DMAE (5 μm) used throughout this study was a gift from E. Merck, Darmstadt, Germany. The anion-exchange HPLC column is made of wide-porosity, fine-particulate, spherical, silicate-based material whose surface is very inert.[18] The weak-basic dimethylaminoethyl (DMAE) groups are bound not directly to the surface of the core material. DMAE groups are attached on linear polymers grafted on the core. Thus, flexibility (so-called tentacle effect) leads to shortening the diffusion path of the analyte, higher throughput, but also results in more acidic groups being reached in or on the oligodeoxynucleotides. Greater selectivity is achieved.[18] For example, detecting the difference between 50 and 51 formal charges on the analyte, i.e., nucleotide units in oligodeoxynucleotides, was shown with such exchanger-analyte interactions.[18] Fractionating according to size in a salt gradient is obtained with good resolution, but if more demanding separations are required, e.g., complete component elution, the use of 20% acetonitrile in the eluent is recommended. In a preparative run up to 9.2 OD_{260} can be applied to the HPLC column.[18] Despite increasing the load, there were no signs of overload. Here we applied 50 μL containing 2.5 OD_{260}, 1.7 OD_{260}, and 2.3 OD_{260} (see caption to Figure 3; OD_{260} = optical density at a wavelength of 260 μm) in analytical runs. HPIEC analyses were carried out with Applied Biosystems Model 152A. Detection, uv 254 nm; flow rate, 1.5 mL/min; room temperature; eluent A: 20 m*M* sodium acetate, pH 6.5, 20% acetonitrile; eluent B: 20 m*M* sodium acetate, pH 6.5, 20% acetonitrile, 1*M* LiCl. Linear gradients were applied as indicated. Complete component elution occurred in order of increasing chain length; however, the retention time of the target sequence N was very close to that of erroneous sequences of chain length $N - 1$. The occurrence of minor peaks that seem to correspond to products longer than the target sequence has been occasionally observed (e.g., Ref. 19), but explanations are not at hand (see Results: Theory and Performance). The peaks of single nucleotides and dinucleotides were checked by standards. The products of protecting group cleavage were also eluated with the first and second peak. The detector capability to measure quantitatively both the smallest and largest peak did not exceed the absorbance detecting range set to appropriate attenuation. The computer output of the ABI 1783 A Absorbance Detector Controller (1 volt per absorbance unit as fixed detector sensitivity) was used with the Hitachi Model D-2000 Computing Integrator. The proper parameters for peak detection and baseline correction were set and, if necessary, changed before each measurement. Data processing was performed with the built-in chromatography software.

High-Performance Capillary Electrophoresis (HPCE) Procedure

In all capillary-electrophoretic separations a Beckman P/ACE System 5510 (Beckman Instruments, Palo Alto, CA, USA), equipped with an uv detector was used. The instrument allows to control temperature of probe and capillary. The separation of oligomers with a length of 30–65 nucleotides was achieved with original Beckman capillaries (Beckman ECAP 100, inner diameter 100 μm), containing a 12% linear polyacrylamide gel. All capillaries had an effective length of 47 cm with 40 cm distance from the capillary inlet to the detection window. The temperature of capillary, buffer, and probe was kept at 30 $\pm$ 1°C. This temperature was maintained during the whole separation procedure including equilibration of the capillary, probe injection, and final separation with base-pair resolution. The buffer system of choice was 1 $\times$ TBE buffer (Tris-borate EDTA) with 7% urea, adjusted to a pH of 8.4 (Beckman reagent kit for oligonucleotide separations). Gel-filled capillaries were mounted in the Beckman 5510 cartridge and the capillary ends were placed immediately in buffer reservoirs. Before each run the capillaries were equilibrated with the running buffer through application of an increasing voltage gradient (0–250 V/cm in 15 min). The polarity was reversed, ensuring injection of the samples at the cathode. Samples were injected electrokinetically with a voltage of 10 kV. For injection the dried oligomers (2.5 OD_{260}) were dissolved in 100 μL of bidistilled water. After a short heating to 95°C for 5 min, the probes were kept on ice until they were placed in the instrument. Electrokinetic injection introduces the charged single-stranded DNA (ssDNA) molecules in the sample onto the capillary by the application of a potential difference to the sample reservoir. More sample material can be put onto the capillary just by increasing the potential or the time applied. Although electrokinetic injection is less reproducible than hydrodynamic injection, electrokinetic injection was chosen because it offers better resolution. This is due to the fact that samples are introduced onto the column into a narrow sample zone. The total amount of capillary-introduced DNA can be calculated with the equation described in Ref. 20: $Q = [\pi \cdot r^2 \cdot c_s \cdot (\mu_{ep} + \mu_{eo}) \cdot E \cdot t \cdot \lambda_b]/\lambda_s$, with Q the amount of sample material injected, r the radius of the capillary, c_s the sample concentration, E the electric field applied, t the injection time, λ_s the sample conductivity, λ_b the conductivity of buffer electrolyte, μ_{ep} the mobility of sample molecules, and μ_{eo} the electroosmotic mobility.

The mean values for the total amount of injected oligonucleotides, including target and error sequences, was

0.13 pmoles in the performed separations. These quantities lead to Gaussian-shaped peaks that do not show leading or tailing. Under the chosen separation conditions, especially in the region of short error sequences, the Gaussian-shaped peaks indicate uniform electrophoretic mobilities of the investigated probe compounds during electrokinetic injection. Elution of products with increasing length from dC_{30} to dC_{65} was obtained at virtually identical times by adjusting the voltage gradient. For detection of the ssDNA fragments the uv detector was set to a data rate of 5 Hz at 254 nm, with bandwidth of 2 nm. The sensitivity for the uv detector was set to 0.1 absorbance unit per volt. The quantitation of target sequences and by-products was achieved by using the System Gold 8.0 software package (Beckman Instruments, Forster City, CA, USA) (see Results). The correctness of the integration parameters was checked by the method of manual cutting off and weighing of peak areas.

Extended Software Package

The software package used in this paper is written in Mathematica (Wolfram Research, Inc., 100 Trade Center Drive, Champaign, Illinois 61820-7237, USA). The software is based on Eqs. (7)–(17), and on the Mathematica program in Ref. 21. The present package needs the relative areas of highly resolved chromatograms or electropherograms as experimental input data and allows us to calculate automatically the model parameters for constant as well as variable propagation and termination probabilities according to Eqs. (7)–(17). The analysis of potential sources of calculation errors showed that miscalculation depends on the precision of peak resolution. Therefore optimization of the chromatographic and electrophoretic separation conditions, respectively, is mandatory. Satisfactory separation conditions resulted here in a resolution of peaks with sufficient precision. Peak quantitation of overlapping peaks that can be fully identified as Gaussian peaks suffices in most cases. The program also produces a graphical representation of the input data and the results. The outlines of the algorithms as well as their implementations will be described elsewhere.

RESULTS: THEORY AND PERFORMANCE

Staggered Flux of Nucleotide Sequences

During synchronous growth of chains target sequence and truncated error sequences are formed at discrete periods of time i. We focus on error nucleotide sequences l that are truncated versions of the target sequence N. A *dynamical system* is considered whose state at time i is characterized by the value of a *stochastic variable.* After each cycle i this variable represents a realization of the *nucleotide length* l. In turn the stochastic process is connected with a probability P counting the relative frequency of any error and/or target sequences. We use this probability to define *moments* M (Ref. 22) measuring here the nucleotide length in connection with the target sequence N. The moment of probability M was derived from the total number of sequences given by a number of elongation cycles i and from the proportion of target to error sequences l (see Table I). The detailed theoretical analysis is given in Ref. 13.

Let $M(l, N)$ denote the local probability moment that after i steps (cycles) the walker reaches site l. Suppose d and p are a set of probabilities of making a step (cycle) i in a way that d_i is given for chain growth and p_i is given for termination of growth. We assume that d_i and p_i are independent from each other. The longer the walker goes along the nucleotide sequence of the target chain, the more the nucleotide length of error sequences increases. An outcome is shown in Figure 1 and Table I. The cyclic repetition of propagation (oxidation, detritylation, and coupling) and termination (capping) reactions[23] during the growth process of nucleotide polymerization is a *staggered flux of symbolic nucleotide sequences.* Nucleotide deletion may be present in any error nucleotide sequences and can arise in any cycle i, even though a capping reaction terminates growth.

Experimental results of chromatographic separations of crude oligonucleotide products from nonenzymatic polymerization obtained so far are consistent with the theoretical analysis (see Application to Experiments). At this point we need a plain argument to connect the synthesis parameter d to a fractal dimension D. The fractal models to be discussed are motivated by one-dimensional Cantor dusts and their global Hausdorff–Besicovitch dimension (a fractal dimension[10]).

Mapping Functions to Cantor Sets

We know that the error sequences undergo a bifurcation sequence resulting in a large number of similar products (see Figure 1). The final step of such a bifurcation tree reminds us of a geometric structure known as a *Cantor set.*[24] We use this analogy here to introduce the connection between fractals and the analysis of oligonucleotide preparations. Though there exist many control methods of oligonucleotide synthesis, we propose this analogy to promote profitable future techniques for generating large and small amounts of oligonucleotides. The property of *self-similarity* or *scaling,* as examplified here by *one-dimensional Cantor sets,* is one of the central concepts of

Table I Exact Numbers of Distinct Arrangements of Target and Error (Truncated) Sequences (up to Heptamers) for a Driven Multicycle Synthesis on Fixed Starting Sites[a]

l	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$
$l = 7$ Heptamers						$1 \cdot d_0^6$
$l = 6$ Hexamers					$1 \cdot d_0^5$	$6 \cdot a_0 b_0 d_0^5$ $1 \cdot a_0 p_0 d_0^5$
$l = 5$ Pentamers				$1 \cdot d_0^4$	$5 \cdot a_0 b_0 d_0^4$ $1 \cdot a_0 p_0 d_0^4$	$15 \cdot a_0^2 b_0^2 d_0^4$ $5 \cdot a_0^2 b_0 p_0 d_0^4$ $1 \cdot a_0 p_0 d_0^4$
$l = 4$ Tetramers			$1 \cdot d_0^3$	$4 \cdot a_0 b_0 d_0^3$ $1 \cdot a_0 p_0 d_0^3$	$10 \cdot a_0^2 b_0^2 d_0^3$ $4 \cdot a_0^2 b_0 p_0 d_0^3$ $1 \cdot a_0 p_0 d_0^3$	$20 \cdot a_0^3 b_0^3 d_0^3$ $10 \cdot a_0^3 b_0^2 p_0 d_0^3$ $4 \cdot a_0^2 b_0 p_0 d_0^3$ $1 \cdot a_0 p_0 d_0^3$
$l = 3$ Trimers		$1 \cdot d_0^2$	$3 \cdot a_0 b_0 d_0^2$ $1 \cdot a_0 p_0 d_0^2$	$6 \cdot a_0^2 b_0^2 d_0^2$ $3 \cdot a_0^2 b_0 p_0 d_0^2$ $1 \cdot a_0 p_0 d_0^2$	$10 \cdot a_0^3 b_0^3 d_0^2$ $6 \cdot a_0^3 b_0^2 p_0 d_0^2$ $3 \cdot a_0^2 b_0 p_0 d_0^2$ $1 \cdot a_0 p_0 d_0^2$	$15 \cdot a_0^4 b_0^4 d_0^2$ $10 \cdot a_0^4 b_0^3 p_0 d_0^2$ $6 \cdot a_0^3 b_0^2 p_0 d_0^2$ $3 \cdot a_0^2 b_0 p_0 d_0^2$ $1 \cdot a_0 p_0 d_0^2$
$l = 2$ Dimers	$1 \cdot d_0$	$2 \cdot a_0 b_0 d_0$ $1 \cdot a_0 p_0 d_0$	$3 \cdot a_0^2 b_0^2 d_0$ $2 \cdot a_0^2 b_0 p_0 d_0$ $1 \cdot a_0 p_0 d_0$	$4 \cdot a_0^3 b_0^3 d_0$ $3 \cdot a_0^3 b_0^2 p_0 d_0$ $2 \cdot a_0^2 b_0 p_0 d_0$ $1 \cdot a_0 p_0 d_0$	$5 \cdot a_0^4 b_0^4 d_0$ $4 \cdot a_0^4 b_0^3 p_0 d_0$ $3 \cdot a_0^3 b_0^2 p_0 d_0$ $2 \cdot a_0^2 b_0 p_0 d_0$ $1 \cdot a_0 p_0 d_0$	$6 \cdot a_0^5 b_0^5 d_0$ $5 \cdot a_0^5 b_0^4 p_0 d_0$ $4 \cdot a_0^4 b_0^3 p_0 d_0$ $3 \cdot a_0^3 b_0^2 p_0 d_0$ $2 \cdot a_0^2 b_0 p_0 d_0$ $1 \cdot a_0 p_0 d_0$
$l = 1$ Monomers	$1 \cdot a_0 b_0$ $1 \cdot a_0 p_0$	$1 \cdot a_0^2 b_0^2$ $1 \cdot a_0^2 b_0 p_0$ $1 \cdot a_0 p_0$	$1 \cdot a_0^3 b_0^3$ $1 \cdot a_0^3 b_0^2 p_0$ $1 \cdot a_0^2 b_0 p_0$ $1 \cdot a_0 p_0$	$1 \cdot a_0^4 b_0^4$ $1 \cdot a_0^4 b_0^3 p_0$ $1 \cdot a_0^3 b_0^2 p_0$ $1 \cdot a_0^2 b_0 p_0$ $1 \cdot a_0 p_0$	$1 \cdot a_0^5 b_0^5$ $1 \cdot a_0^5 b_0^4 p_0$ $1 \cdot a_0^4 b_0^3 p_0$ $1 \cdot a_0^3 b_0^2 p_0$ $1 \cdot a_0^2 b_0 p_0$ $1 \cdot a_0 p_0$	$1 \cdot a_0^6 b_0^6$ $1 \cdot a_0^6 b_0^5 p_0$ $1 \cdot a_0^5 b_0^4 p_0$ $1 \cdot a_0^4 b_0^3 p_0$ $1 \cdot a_0^3 b_0^2 p_0$ $1 \cdot a_0^2 b_0 p_0$ $1 \cdot a_0 p_0$

[a] For calculation of coefficients, see Ref. 13. N is the number of nucleotides of the target; l denotes the grown number of nucleotides of target or error sequences; d_0 = constant (average) propagation (detritylation, coupling and oxidation) probability; p_0 = constant (average) termination (capping) probability; $a_0 = 1 - d_0$; $b_0 = 1 - p_0$. The total number of sequences $a_{N,S}$ (sum of error and target sequences, S), for example, was calculated by $a_{N,S} = \sum_{l=1}^{N} \binom{N}{l}$.

fractal geometry. It is directly connected with the generalization to fractal dimension as illustrated below. For example, an object is said to be self-similar if it is formed by parts that are similar to the whole. For a full understanding, the iterative construction of the middle-third Cantor set L is shown in Figure 2. The generation of the Cantor set[10] may be started with a straight line of unity L_o. Then the initiator L_o is partitioned into three equal parts and the middle third is removed. This generates L_1 counting of $l = 2$ parts of the original set. The process is repeated on each of the two intervals, which increases the number of elements to four. In Figure 2, the length of the remaining intervals are $(1/r)^{k-1}$ with $r = 3$ in specific iteration steps $i = k$. In the $(k - 1)$th step (cycle) we observe l^{k-1} intervals.

The introduction of the general number of remaining elements l and the factor of reduction r for a specific iteration step k allows us to generalize the Cantor set. To quantitatively characterize such sets the cluster structure is theoretically covered by a layer of balls with volume V of radius ϵ. It is known from Refs. 10 and 25 that the volume increases as

$$V(\epsilon) \propto \epsilon^{-D_o}, \tag{1}$$

where D_o is the fractal dimension:

$$D_o = \lim_{\epsilon \to 0} \frac{\ln V}{\ln\{1/\epsilon\}}. \tag{2}$$

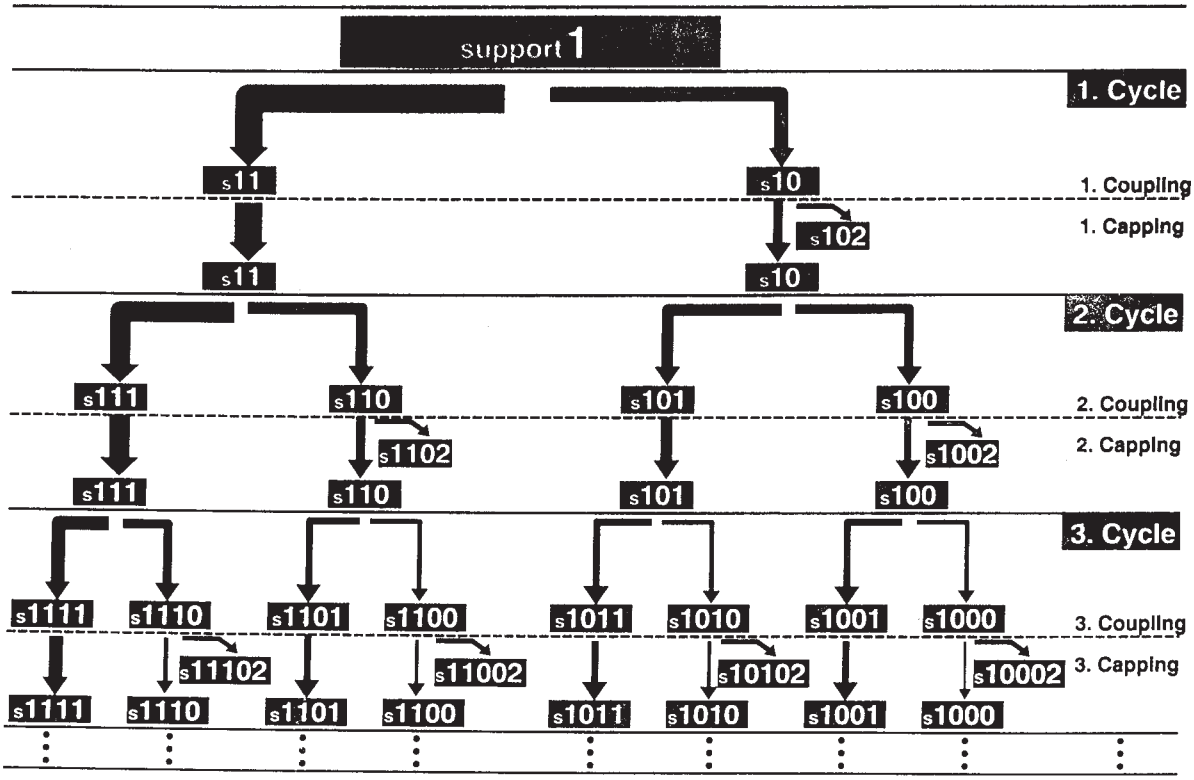

FIGURE 1 Idealized staggered (nonuniform, weighted) flux of arbitrary nucleotide sequences. The sequence space is based on entries of {0, 1, 2}. 1 = nucleotide (arbitrary A, T, G, C). 0 = no nucleotide at position referring to the target sequence. 2 = stop of growth. We define forward and backward reactions of propagation and termination by following notation: $1 \rightarrow 11$, $1 \rightarrow 10$ and $10 \rightarrow 102$, $10 \rightarrow 10$. See text for explanation.

We choose in our discussion $\epsilon = (1/r)^{k-1}$, by scaling down the number of particle V on the structure, $V = l^{k-1}$. Inserting these expressions for the Cantor set into Eq. (2), we get

$$D_o = \frac{\ln l}{\ln r}. \tag{3}$$

So far we derived a formula connecting the elements l of the set with the reduction factor r.

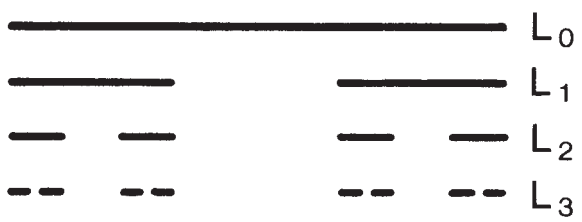

FIGURE 2 The first three stages of the triadic one-dimensional Cantor set L generated on the unit interval for $l = 2$ and $r = 3$. See text for explanation.

Since the experimentally observable is the propagation probability d, we need to modify the formulas above in such a way that they are useful in connection with our experiments. Hence, we replace r by $\chi \cdot 1/d_o$, where χ is a scaling factor and d_o is the average (constant) propagation probability. Using Eq. (3), D_o is given by the fractal dimension

$$D_o = \frac{\ln l}{\ln\{\chi \cdot 1/d_o\}}, \tag{4}$$

where d_o satisfies $0 < d_o < 1$. To include dynamic properties of growth systems of multicyclic synthesis, we use the fractal dimension $D(N) = \lim_{d_o \rightarrow 1} D_{a,\text{measurable}}(N, d_o)$.[15] Inserting this expression into Eq. (4) and rearranging terms, we end up with

$$\chi = d_o \cdot \exp\left\{\frac{\ln l}{2 - [N/(N-1)]}\right\}. \tag{5}$$

This formula connects the error sequences l, the propagation probability d_o, and the nucleotide sequence N in one expression. In the limit $N \gg 2$ a nontrivial transformation is shown here from the embedding dimension 2 (Ref. 13) to the embedding dimension 1.

Let us now consider the case where the reduction factor r is not identical for all of the set elements l. However, we assume that the overall reduction factor is represented by $r = \sum_{k=1}^{l} r_k$. The construction is again designed to give a fractal dimension between 0 and 1:

$$D_o = \frac{\ln l}{\ln\{\sum_{k=1}^{l} \chi_k\} - \ln d_o}, \qquad (6)$$

where $\chi = \sum_{k=1}^{l} \chi_k$, $l = \{2, 3, 4, \ldots, N\}$, and $0 < d_o < 1$. This would be true if the iterations were extended to higher orders, i.e., to large N. Equation (6) gives us the relationship between the geometrical fractal Cantor set and quantities of the experiments.

The fractal concept sketched above can be applied to *cumulative distributions of sequence variabilities* $\sum_{l=1}^{N-1} (M(l, N))$. In this way, we gain here valid access to the overall efficiency[26] of a support system if respective experiments are of good similar performance. A quantitative description of this overall efficiency in terms of fractal mathematics (D_o) may assist in the choice of appropriate experimental synthesis D_o methods.

Application to Experiments

We prefer *dynamics of sequence variabilities* to generalize some fractal aspects about the synthesis of oligonucleotides. When time i for the cycle proceeds, the specific models we consider account for chemical synthesis as approaching forward and also possible backward reactions of nucleotide growth in synchronous mode. In synchronous mode the sequences are always generated in weighted arrangements starting from monomer up to the intended nucleotide length at the time period i. The resulted target length is uniform. These stepped-up walks can be associated with the formal generating function $M(l, N)$. After suitable formalistic changes[21] we get a model approach to total yield of synthesis regardless of what average optimum and standard conditions of synthesis and preparation are chosen:

$$d_o^{N-1} = \frac{A^{(\mathrm{expl})}(N, N) \cdot \sum_{l=1}^{N} (l \cdot M(l, N))}{N}, \qquad (7)$$

$$p_o = \frac{1}{(1/d_o) - 1} \times \frac{N \cdot A^{(\mathrm{expl})}(N - 1, N)}{(2 - N) \cdot (N - 1) \cdot A^{(\mathrm{expl})}(N, N)} + \frac{N - 1}{N - 2}, \qquad (8)$$

where $A^{(\mathrm{expl})}$ is the relative area of the Nth and the $(N - 1)$th peak in chromatograms or electropherograms. Equations (7) and (8) are a set of iterated functions. As a consequence, the four chemical reactions of detritylation, coupling (elongation), capping, and oxidation[23] are completely covered by our Eqs. (7) and (8) as well as by our Eqs. (9) and (10) [for Eqs. (9) and (10) see below].

To compute the extent of *nonconstant growth conditions*, we further use the theory described above (see also Refs. 15, 21, and 27). This method requires numerical solution of the coupled nonlinear equations

$$\prod_{i=1}^{N-1} d_i = \frac{A^{(\mathrm{expl})}(N, N) \cdot \sum_{l=1}^{N} (l \cdot M(l, N))}{N} \qquad (9)$$

and

$$\frac{A^{(\mathrm{expl})}(l, N)}{A^{(\mathrm{expl})}(l - 1, N)} = \frac{l \cdot \{M(l, N - 1) + [M^{\bullet}(l - 1, N - 1) - M^{\bullet}(l, N - 1)] \cdot d_{N-1}\}}{(l - 1) \cdot \{M(l - 1, N - 1) + [M^{\bullet}(l - 2, N - 1) - M^{\bullet}(l - 1, N - 1)] \cdot d_{N-1}\}}. \qquad (10)$$

The special nonlinear, recursive functions M and $M^{\bullet}$ (first moments) were found by projecting $M(l, N)$ and $M(l - 1, N)$ onto $M(l, N + 1)$.[15] Using this procedure, several coefficients d_i and p_i are superimposed.

Reliable data can be calculated with minimal time course studies by applying the following criteria, serving for the best quality estimation ($\delta z_{\max}$) between experimental and theoretical results, as

$$\Delta z_{\max} = \sum_{j=1}^{N-l_{\min}} \left| \frac{\overline{\kappa_j} - \kappa_j}{\kappa_j} \right|, \qquad (11)$$

where $l_{\min}$ is the smallest peak resolved in the measured chromatogram/electropherogram. The j counted to resolved number of peaks. For a quality assessment we used in our calculation the ratio

$$\delta z_{\max} = \frac{\Delta z_{\max}}{z} \qquad (12)$$

with $z = \sum_{j=1}^{N-l_{\min}} \overline{\kappa_j}$.[27,21] If the conditions $\kappa_j = \frac{f_j(l, N)}{f_j(l-1, N)}$ *and* $\overline{\kappa_j} = \frac{g_j(l, N)}{g_j(l-1, N)}$ *with* $l = \{N, N-1, \ldots, N-l_{\min}+1\}$, $j = \{1, 2, \ldots, N-l_{\min}\}$ are fulfilled, and max $A^{(\text{expl})}(N, N)$, min $A^{(\text{expl})}(N, N)$, max $A^{(\text{expl})}(N-1, N)$, min $A^{(\text{expl})}(N-1, N)$ are given, we obtained parameter values for nonconstant as well as constant growth. The functions $f_j(l, N)$ and $g_j(l, N)$ describe the theoretical, and experimental values of the relative peak areas, respectively.

Parameter Values* d *from Experiments. The following question arises: how do we get parameters d and p for short time series? Relative quantitation by referring to the integrated absorbance of the peaks (absorbance/mm^2) is always performed in order to obtain the experimental input data for the theoretical analysis.

Effort has been made to minimize the overlaps of peaks by running appropriate separation conditions. Differences in relative peak areas have effects on the parameter values calculated.

At the *integrator terminal* a correct sampling period was set before separation analyses. If this setting was wrong, consequently noise or baseline drift could be detected as a broad peak. For detecting single peaks, peak top points were automatically looked for and, after collecting data, peak start and end points were found by the built-in chromatographic and electrophoretic software packages. A peak was detected, when the product of noise value measured at test runs and sensitivity is less than variations of sampling data at consecutive points.

Measured chromatograms and electropherograms include solvent and instrumental background. In order to obtain correct values of sample components, the background needs to be subtracted by baseline corrections. Therefore baseline measurements were always carried out before running analytical separation in the same conditions as in the sample measurements and automatically stored on computer memory. These measurements were read out to integrator terminal by specifying the memory numbers. Variation of baseline from start to stop of the chromatogram and electropherogram, respectively, are removed before calculating peak areas. Factors causing variations of the baseline are, for example, status change of moving and stationary phases, drift by temperature rise, and drift by changing over the solvent.

On Synthesis Details Related to Simulation. After automatic peak identification and baseline correction, the simple percentage method and the normalized percentage method with scale factor, respectively, were used as quantitative methods for determination of true peak areas. Further, the proper estimation of the baseline was visually controlled for all chromatograms and electropherograms. The gradual increase in absorbance background in HPIEC (e.g., Figure 3A) or HPCE (e.g., Figure 4C) was constant. It did not arise from the lag in elution of reaction products.

We found by our experiments (see Figure 3) that average (constant) conditions of synthesis and preparation can be assumed for the short target oligonucleotides on CPG-500 Å material as well as on a macroreticular polyvinylacetate support.[17] As seen from Figure 3, the experimental and theoretical curves do correspond quite well. The theoretical curves were iterated to the best confidence with experimental data. This iteration procedure gives a full fit with all measured OD values of the experimental curves. The numerical accuracy (tolerance) θ was chosen to be 10^{-5}, for d and p. The appearance of the $N+1$ peak is occasionally found. To our knowledge, this phenomenon was not subject to detailed investigations.[19] The experimentally founded question comes up that the acidity of tetrazole causes a trace amount of detritylation that may lead to multiple incorporation of a given nucleotide. Normally, one suppresses this phenomenon by the synthesis protocol, but in general the theory presented accounts also for $N = N+1, N+2, \cdots$ peaks.

Suppose the data obtained from an experiment cannot be analyzed by average (constant) growth conditions. In this situation the expressions $\prod_{i\in[1,N)} d_i$ of Eq. (9) as well as $\prod_{i\in[1,N)} p_i$ can be weighted by variable growth conditions. We apply here the following models allowing systematic variation of d and p. The exponential growth is given by $d_i = \alpha \cdot \exp\{-\beta \cdot (i-1)\}$ and p_i

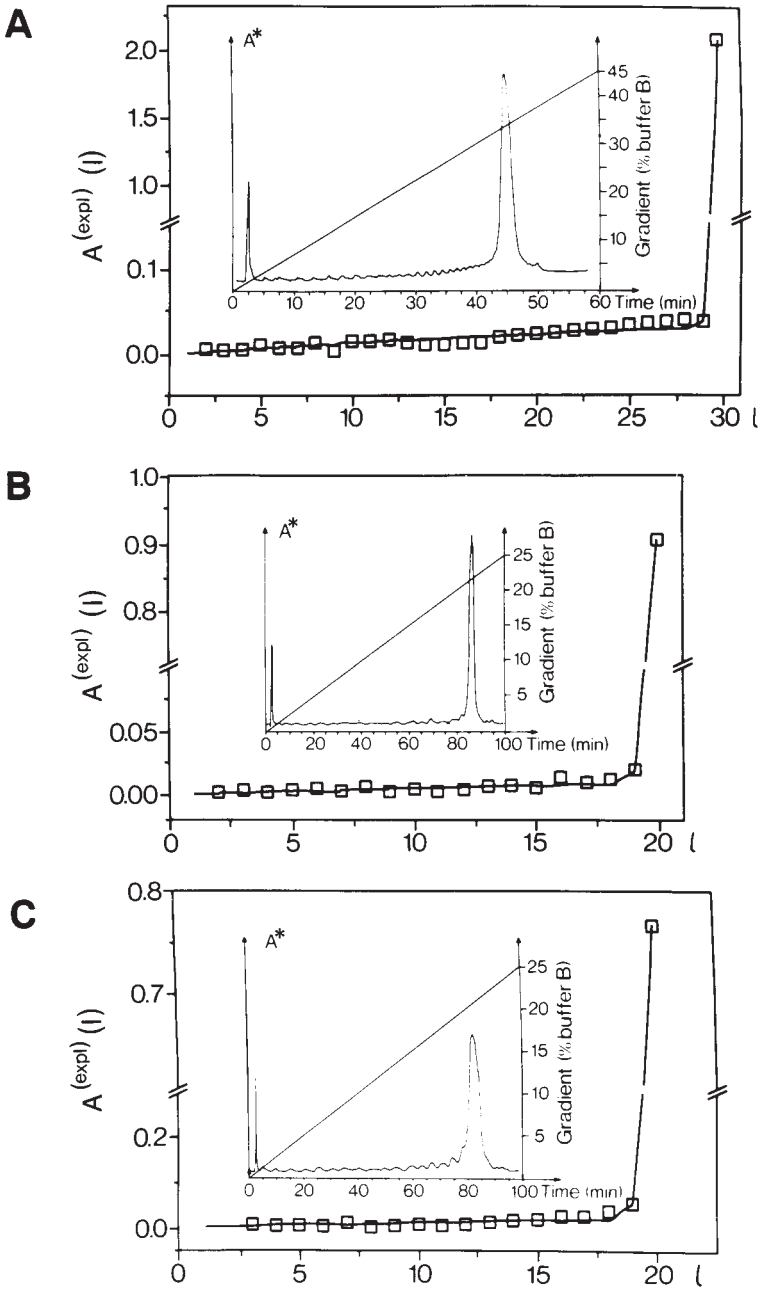

FIGURE 3 Comparison of experimental and calculated distribution of chain length for oligonucleotide preparations on a CPG-500 Å support (A) and a macroreticular polyvinylacetate support (B, C). Abscissa: nucleotide length (l) derived from retention time. Ordinate: $A^{(\mathrm{expl})}(l, N) \equiv A^{(\mathrm{expl})}(l)$. Experimental values (□ □ □) taken from anion exchange HPLC (insets). Theoretical values (——) calculated by the constant (average) propagation (d_o) and termination (p_o) efficiencies. The DNA synthesizer Pharmacia Gene Assembler Plus was used according to standard manufactures's recommendation. See Table II for parameter comparison. (A) 30-mer, $(dC)_{30}$. Synthesis scale: 0.2 μmol. Analyzed amount: 2.5 OD_{260}. $d_o = 0.9870$, $p_o = 0.9900$. (B) 20-mer, 5′ *d* CGG ATT CAT AGC TGA GTC AT. Synthesis scale: 1 μmol. Analyzed amount: 1.7 OD_{260}. $d_o = 0.9905$, $p_o = 0.9335$. (C) 20-mer, same sequence. Synthesis scale: 10 μmol. Analyzed amount: 2.3 OD_{260}. $d_o = 0.9755$, $p_o = 0.8960$.

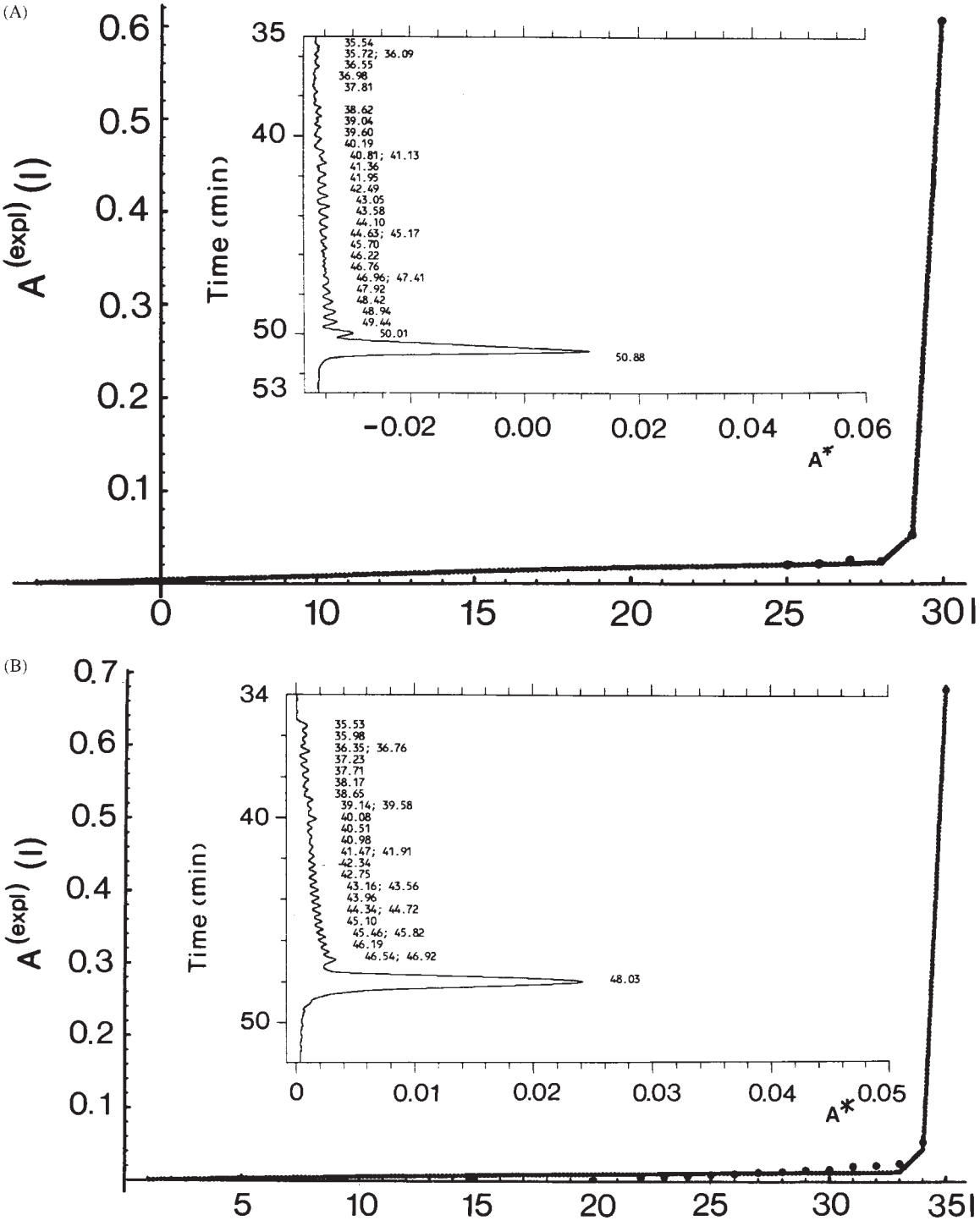

FIGURE 4 Experimental electropherograms (high-performance capillary electrophoreses, insets) of crude products of (A) dC_{30}, (B) dC_{35}, (C) dC_{40}, (D) dC_{50}, (E) dC_{60}, (F) dC_{65}. (A), (D)–(F) were prepared on P_{29} polystyrene-grafted (2–3%) polytetrafluoroethylene support, whereas (B) and (C) were prepared on polystyrene primer support (Pharmacia). The Applied Biosystems Model 380B (A) and 394 DNA synthesizers (B–F) were used according to special protocols (see Experimental). Experimental values (● ● ●) taken from HPCE (insets) for calculated distributions of chain length (——). The minimalization of δz_{max} for all parameters was the criterion for data fitting. Abscissa: nucleotide length (l) derived from retention time. Ordinate: $A^{(expl)}(l, N) \equiv A^{(expl)}(l)$. See Table II for parameter comparison.

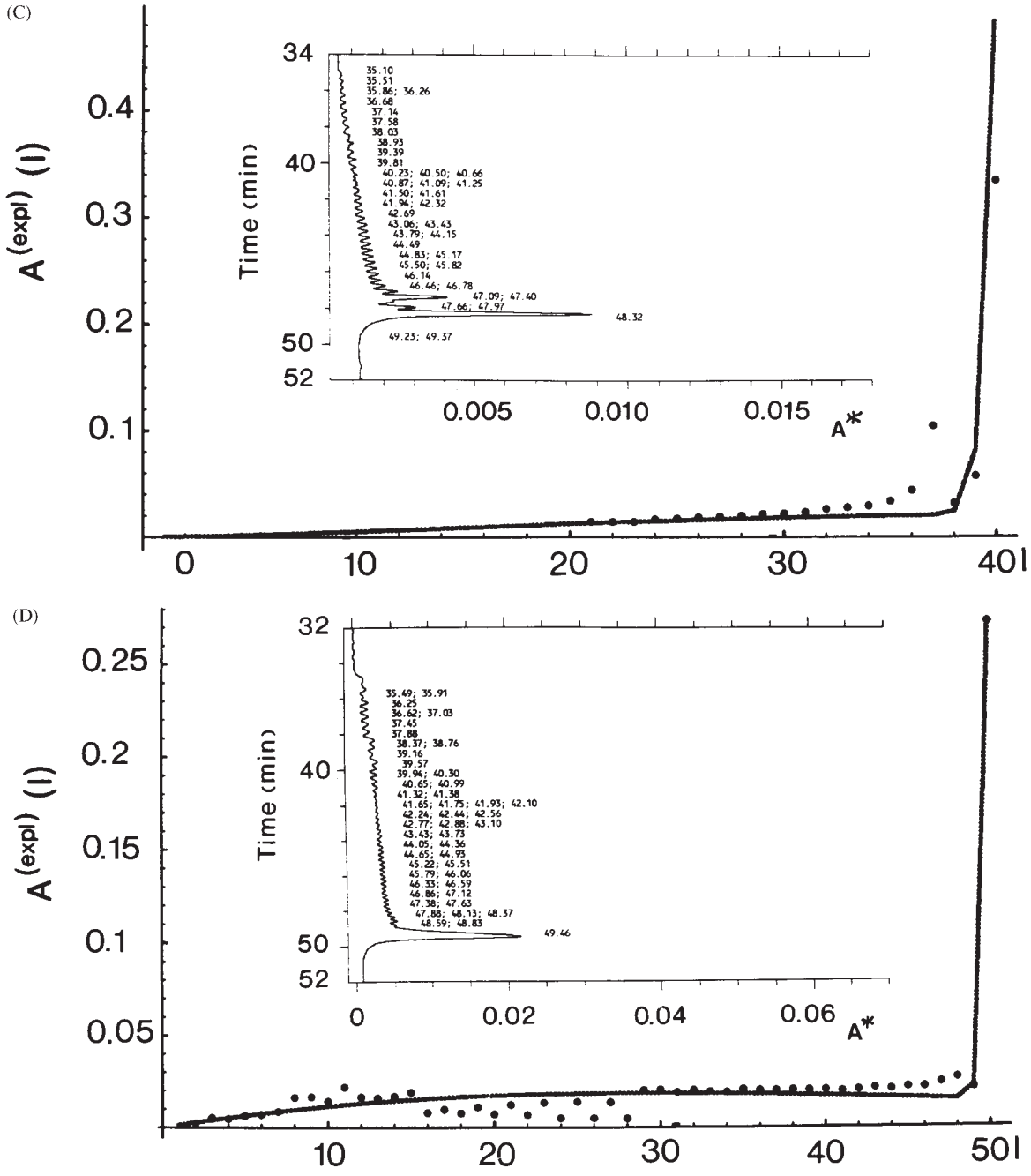

FIGURE 4 (*Continued from the previous page*)

$= \gamma \cdot \exp\{-\delta \cdot (i-1)\}$.[21] The constant parameter α is interpreted in terms of the average (constant) propagation probability d_o, whereas the constant parameter γ is given by the average (constant) termination probability p_o. This result is confirmed here by numerical and analytical methods. The exponents in β and δ are new parameters describing a deviation from average (constant) growth. The parameter values α, β, γ, and δ with $0 < \alpha, \beta, \gamma, \delta < 1$, of course, depend on the particular details of the experiments.

Experimental vs Theoretical Approach. In Figure 4A–F, the separations of the crude products by high-performance capillary electrophoresis are shown for several target nucleotide length N using

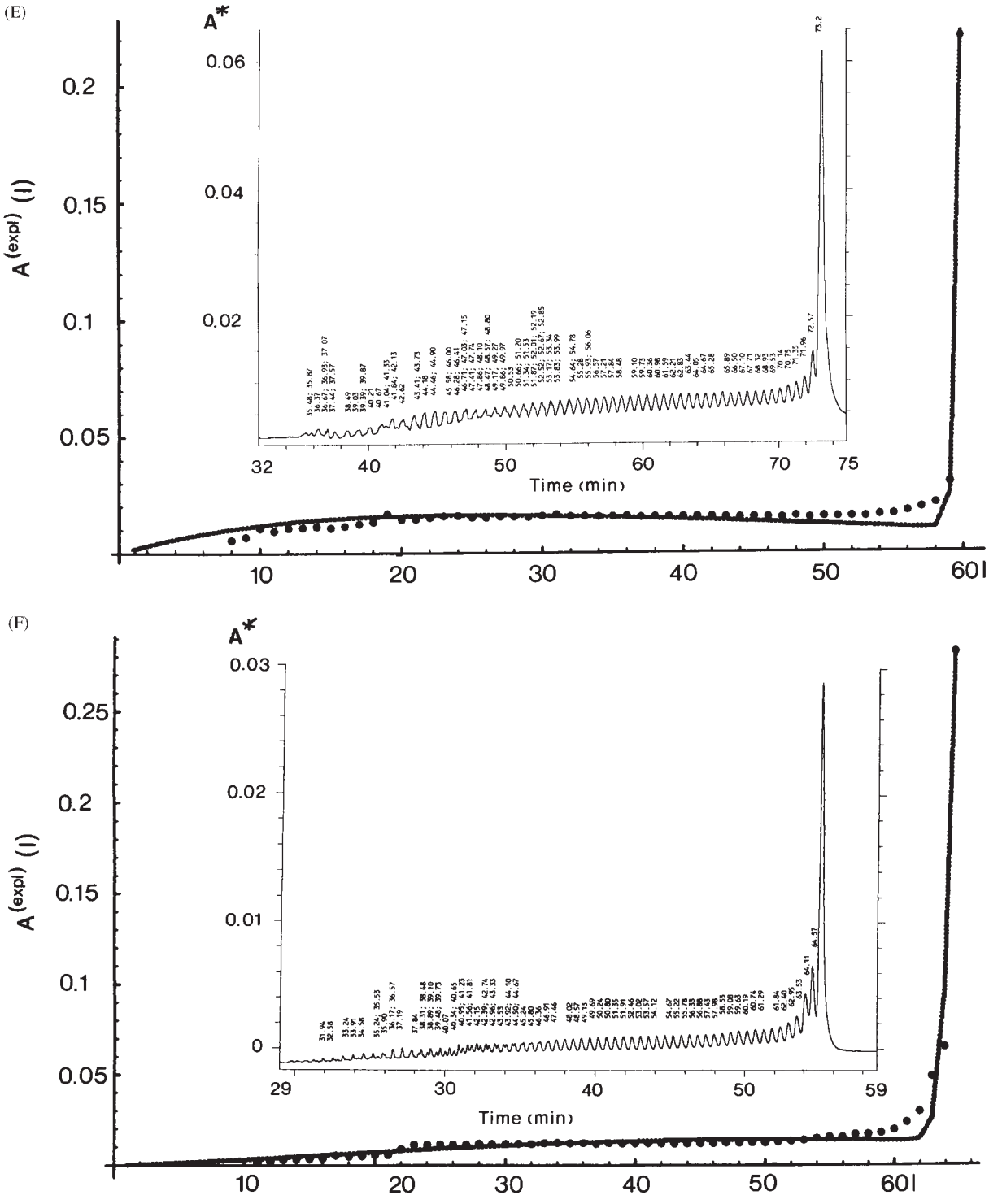

FIGURE 4 (*Continued from the previous page*)

the polystyrene-grafted polytetrafluoroethylene support P_{29} and the polystyrene primer support (Pharmacia) in the syntheses. Syntheses and separations of target nucleotide lengths N up to 65 were performed, i.e., N covers a length scale up to 65, in the study. The resolution of defined product peaks by HPCE is not complete, but in most cases the uv-electropherogram separated peaks from N to N

− 15. From the insets of Figure 4 it is evident that at shorter target length N of oligonucleotide syntheses (Figure 4A–C) the peak quantitation of $N - 1$ and $N - 2$ sequences is more difficult due to the excess of target products. As one can see in Figure 4D to F, under the electrophoretic conditions used (voltage gradient conditions, mobile phase, pH, temperature, etc.), a component separation was obtained up to $N - 48$, $N - 52$, and $N - 54$, respectively.

The experimental uncertainties do not seriously compromize the theoretical analysis because local experimental data points cannot be expected to give a global picture of a nonlinear growth system. Thus, the criterion δz_{max} [Eq. (12)] is applied in order to obtain good overall fits to measured experimental curves with a numerical accuracy (tolerance) θ of 10^{-5} (same as in Figure 3). The calculations of the growth parameters were carried out from the separation profiles according to Eqs. (7)–(12). The confidence limit of numbers of experimental *integrated absorbance* (absorbance/mm^2) depends on the special models, for example, an average or exponential stepwise synthetic process used for theoretical analysis.

Computer simulations of chemical oligonucleotide syntheses[15] indicated that values of N and $N - 1$ sequences are most important for determination of d_o and p_o parameters of average (constant) growth. Considering the nonlinear equations of the IPLCS (Inverse Power Law of Driven Multicycle Synthesis) for average growth [e.g., Eqs. (7) and (8)], it is reasonable to fit the curves to N and $N - 1$ values; all other values are on (or very close to) the obtained curves. For the specific, exponential growth model more input data than the N and $N - 1$ values have to be collected.

The optimal model fits shown in Figure 4A to Figure 4C are obviously nonlinear for panels D–F. It is surprising that these calculated curves are nonuniform. We devised a system of equations (IPLCS) that takes into account different situations characterized by d and p (see, for example, theoretical curves of the IPLCS in Ref. 28, therein Figures 1–4, 7, and 8). The theoretical basis for this phenomenon is the nonlinearity of a stepwise synthetic process, e.g., chemical oligonucleotide synthesis. Due to the nonlinearity, the approximation is well suitable for realistic patterns of synthesis. However, it appears that the large amount of full-length product would cause increases in absorbances for the $N - 1$ and slightly shorter products, that might contribute to a loss-of-fit. As shown in Table II the experimental data indicate a change from constant growth to variable growth. In this case the exponential model was used to get an optimal model-to-reality correspondence. The residuals δz_{max} between the observed data and the calculated model data varies randomly. Constant growth was observed for the 60-mer syntheses and for one of the 45-mer syntheses. Inherently, the best fit of all experimental data may be to the exponential growth model including longer synthesis length and higher coupling efficiencies. At shorter synthesis length ($N \leq 30$) and/or with synthesis supports with steric restrictions (PA, CPG), this is experimentally indistinguishable from, or limited to, constant growth. When in individual syntheses coupling efficiencies are low (below 0.9825) due to inherent support access limitations (only seen for P_{29} resin) or external and variable limitations in other factors like chemicals, the constant growth model becomes the better fit.

Fractal Measures for the Overall Efficiency. We proposed that nonenzymatic (and enzymatic) multicyclic growth processes are statistically fractal structures whose dimensions can be described in mathematical terms.[28] The mathematical definition of our fractal dimension D for multicyclic nucleotide processes producing sequence variabilities in the two-dimensional embedding space is given in Ref. 9 for the first time according to Mandelbrot.[12] The very generality of this result holds for all fractal sets of cumulative sequence variabilities, i.e., truncated, deleted, *and* point-mutated sequence variabilities. The minimal requirements sufficient to produce, or to degrade, a target DNA sequence N are multistep additions, recombinations, or degradations of some monomeric or oligomeric building blocks with the propagation probability function $d = d_i$, at cycle i. Hence, the general expression of the fractal dimension $D_{a,\text{measurable}}$ given in Ref. 13 can be written in the following form:

$$D_{a,\text{measurable}} = 2 - N\left(\frac{(-\partial(\prod_{i\in[1,N)} d_i))/\partial N}{1 - \prod_{i\in[1,N)} d_i}\right). \quad (13)$$

Equation (14) describes the response to the average (constant) growth[15]:

$$D_{a,\text{measurable}}(N, d_o) = 2 - N\frac{d_o^{N-1}}{1 - d_o^{N-1}}\ln\frac{1}{d_o}\ . \quad (14)$$

For exponential growth we obtain from Eq. (13)[21]:

Table II Table of Parameter Comparison[a]

N	Support	Separation Method	Figure	*d*		*p*	
20	PA	HPIEC	3C	0.9755		0.8960	
20	PA	HPIEC	3B	0.9905		0.9335	
30	CPG-500Å	HPIEC	3A	0.9870		0.9900	
				α	$\beta \cdot 10^{-4}$	γ	$\delta \cdot 10^{-3}$
30	P_{19}	HPCE	—, 6	0.9829	9.21	0.9662	0.30
30	P_{29}	HPCE	4A, 6	0.9847	8.75	0.9854	3.30
35	PS	HPCE	4B, 6	0.9885	4.37	0.9288	1.25
40	PS	HPCE	—, 6	0.9825	9.06	0.9892	4.13
40	PS	HPCE	4C, 6	0.9906	9.37	0.9645	4.24
45	P_{29}	HPCE	—, 6	0.9364	0.00	0.9589	0.00
45	P_{29}	HPCE	—, 6	0.9713	8.75	0.9858	3.30
45	P_{29}	HPCE	—, 6	0.9887	4.37	0.9092	4.37
50	P_{29}	HPCE	4D, 6	0.9671	3.75	0.9847	0.00
60	P_{29}	HPCE	4E, 6	0.9591	0.00	0.9714	0.00
60	P_{29}	HPCE	—, 6	0.9591	0.00	0.9712	0.00
60	P_{29}	HPCE	—, 6	0.9695	0.00	0.9559	0.00
65	P_{29}	HPCE	4F, 6	0.9904	5.78	0.9745	4.37

[a] PA: macroreticular polyvinylacetate support; CPG-500Å: controlled pore glass; PS polystyrene primer support; P_{19}: polystyrene grafted (5–7%) polytetrafluoroethylene support; P_{29}: polystyrene grafted (2–3%) polytetrafluoroethylene support. See text for definition of the parameters α, β, γ, δ.

$$D_{a,\text{measurable}}(N, \alpha, \beta) = 2 - N\left(\frac{\alpha^{N-1}\cdot(-3\beta + 2\beta N - 2\ln\alpha)\cdot\exp\left\{\frac{-\beta\cdot(1-N)\cdot(2-N)}{2}\right\}}{2\left(1-\alpha^{N-1}\cdot\exp\left\{\frac{-\beta\cdot(1-N)\cdot(2-N)}{2}\right\}\right)}\right). \quad (15)$$

Generality of Fractal Dimensions $D_{a,measurable}$, D_a, $D(N)$.

The usefulness of constructing the functions $D_{a,\text{measurable}}$ arises from the theorem [Eqs. (10)–(12) in Ref. 15] that as d varies between 0 and 1, $D_{a,\text{measurable}}$ tends to 2 (upper limit) for all possible non-negative densities $M(N, N)$. Thus we may determine the dynamics $D_{a,\text{measurable}}$ using Eq. (13) from d_i. The upper and lower bounds to $D_{a,\text{measurable}}$ can be calculated from d_o and α, β. For $d_o \to 0$ in Eq. (14) and $\beta \to 0$, $\alpha \to 0$ in Eq. (15) we obtain the upper limit 2. For $d_o \to 1$ in Eq. (14) and $\beta \to 0$, $\alpha \to 1$ in Eq. (15), we find the special lower limit:

$$D(N) = 2 - \frac{N}{N-1} \quad (16)$$

Equation (16) quantifies the influence of N on the dynamics.[15] Then, we can write

$$D_a = \frac{2 - (N/(N-1))}{D_{a,\text{measurable}}}. \quad (17)$$

Equation (17) now allows to quantify the influence of d on the dynamics.[15,21] The argument of lower bounds for average and exponential development can be used to apply D_a to a universal measure of the two classes of multicyclic process that is being made of multistep additions/recombinations of some building blocks resulting in a largest nucleotide sequence N of interest (universality classes). However, in exponential growth the symbolic expressions of $D_{a,\text{measurable}}$ are more complicated [see

e.g., Eq. (15)], inasmuch as the functions $\prod_{i\in[1,N)} d_i$ have to be used. The problem breaks up into the following:

It is true that there exists situations corresponding to $D_{a,\text{measurable}} \leq 2 - (N/(N-1))$. This result indicates the interference of Eq. (17) because we obtain values that cannot all be expressed in terms of D_a. In order to describe the influence of d on the dynamics, we must have $D_{a,\text{measurable}} > 2 - (N/(N-1))$, meaning $0 \leq D_a < 1$. The values of D_a, N, and stepped-up probabilities result in surfaces that are plotted in Figure 5. It is, in fact, a most interesting demonstration how much of the d parameter is not related to different target length N in constant (Figure 5A) as well as exponential (Figure 5B) growth models. If values of the universal measure D_a are equal, the stepwise synthetic processes are of same performances (efficiencies), which now are independent of the target length N for producing high yields of the target sequence. The parameter d (in the chemical literature used as ''coupling efficiency'') or individual yields (e.g., trityl yields) do not allow the quantitative characterization of chemical oligonucleotide synthesis at different target lengths N. The d parameter and individual yields reflect the quantitative influence on any reaction step only. It is evident from Figure 5A and B that the bold points, which are extracted from measurements, do not fit in the much flatter surface, as indicated by changes in color presentation. Chemical oligonucleotide syntheses, which are not optimized to give high target length yields, make no sense. In Figure 5B some of the bold points are positioned outside the surface. The reason for that pattern is as follows: In order to obtain a three-dimensional representation we fitted the experimental data for Figure 5B with an assumed arbitrary α value of 0.9887 for all syntheses, so exponential oligonucleotide syntheses at longer target length N and lower or higher α values (see Table II) could be visualized in one plot.

The scaling of the target nucleotide length N within the region of size defines the fractal dimension $D_{a,\text{measurable}}$. Figure 6 shows the log–log linear relationship for the data from Table II. We provide here error bars (standard deviations from the mean values). It is evident that the dynamics of sequence variability of the chemical oligonucleotide syntheses can be modeled by a scale-invariant power law. The linear fit corresponds to $D_{a,\text{measurable}} = 1.26$. This value is consistent with the theoretical mean value of the fractal dimension which is calculated by Eqs. (14) and (15).

DISCUSSION

With the progress in using oligonucleotide sequences in analyses for basic research in molecular biology and molecular medicine, for diagnostic application and for gene therapy in medicine, it has become obvious that experimental quantitative control and constancy in oligonucleotide sequence production are desirable for research and development in biomedicine. Several shortcomings in testing oligonucleotides indicate the urge for control improvements. The minimizing of error oligonucleotide length and the maximizing of intending target lengths in small or even large yields is mandatory for experiments (also in the case of syntheses of oligonucleotide libraries). Under the chemical nonenzymatic synthesis of oligonucleotides,[5,6,23] any intrinsic deviation within target sequences may be rare, if not totally excluded, in the experiments we presented here. Thus, only length determination—and not nucleotide sequence determination—in the target sequence applying chromatographic and electrophoretic methods,[15] can gain desirable effectiveness superior to summative spectroscopic determination on an approximative average length scale during synthesis.[8] The crude yields as described above must be determined in ion-exchange high-performance liquid chromatography, high-performance capillary electrophoresis, or gel electrophoresis. Computer simulations of elution profiles demonstrate the influence of nucleotide length l *and* capping reaction p on the peaks $A^{(\text{expl})}(l, N)$ in chromatograms or electropherograms.[15] The existing chromatographic and chemical parameters [e.g., retention time, trityl yield, average (constant) yield, etc.] are not suitable to characterize quantitatively the composition of the crude product as well as the time course of chemical oligonucleotide synthesis. We show here that the immeasurability of details of the dynamic process[29] of chemical synthesis demands an adequate simulation, or fitting model procedures of those nucleotide growth processes.

In the presented analyses of oligonucleotide structures, fractality at different spatial, geometric scales, or dynamic processes at different time scales are related to each other.[29] By dynamic models we extend the suggested simulations to a best fit of useful form. Further, the prominent function d_i can be used in Eq. (13) for an extended fractal dimension $D_{a,\text{measurable}}$ in terms of stretched exponential and polynomial development.[27] This is a new set of parameters that augment the formulative capacity by at least two by two terms. However, the experimental results are not sensitive and constant enough

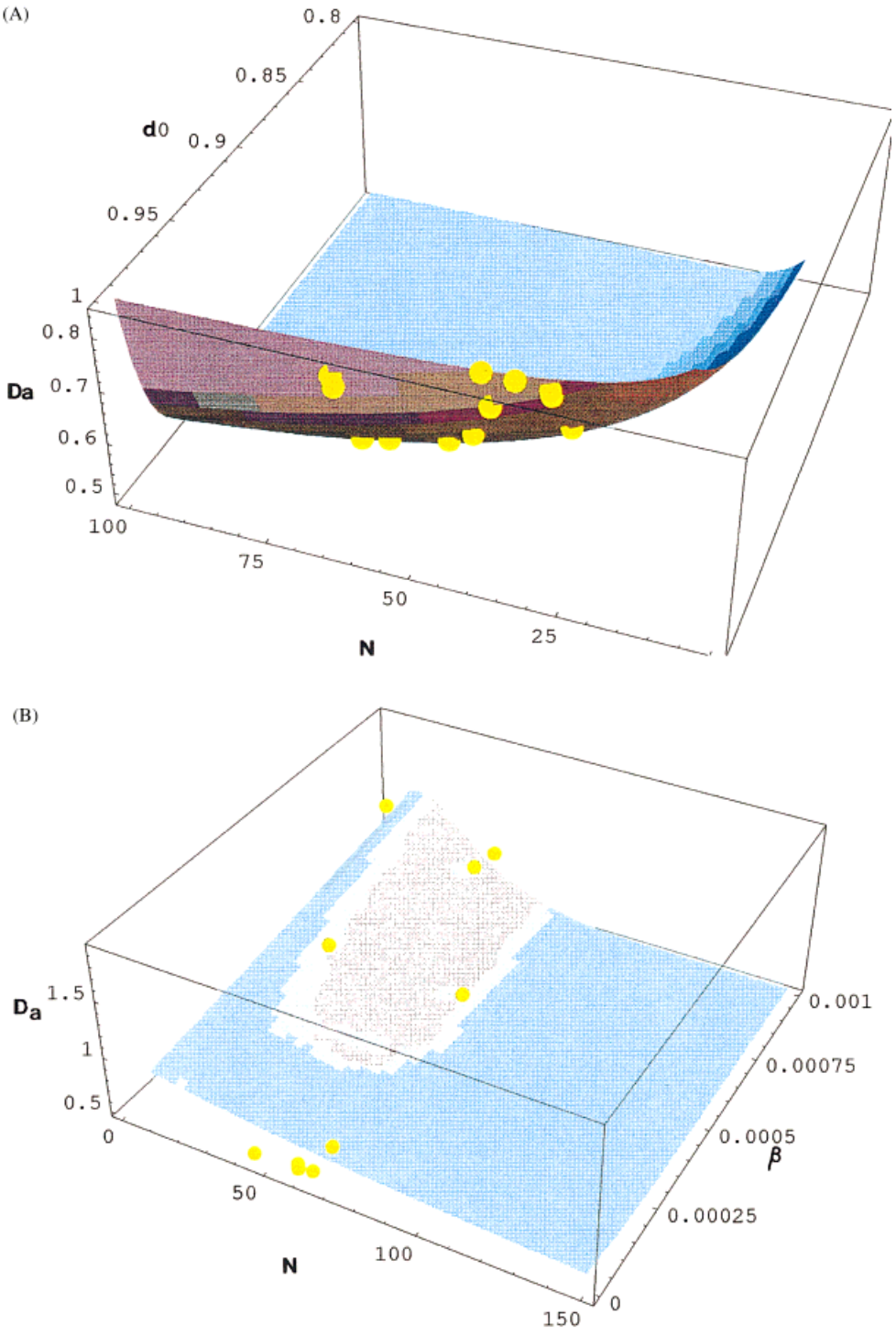

FIGURE 5 Three-dimensional surfaces are formed by $D_a = \frac{2 - (N/(N-1))}{D_{a,\text{measurable}}}$, N, and stepped-up probabilities. Numerical results were taken for the specific growth model of oligonucleotide synthesis. (A) constant (average) growth. (B) exponential growth with fixed $\alpha = 0.9887$ for all data. The bold points indicate the calculated values extracted from measurements. See text for explanation.

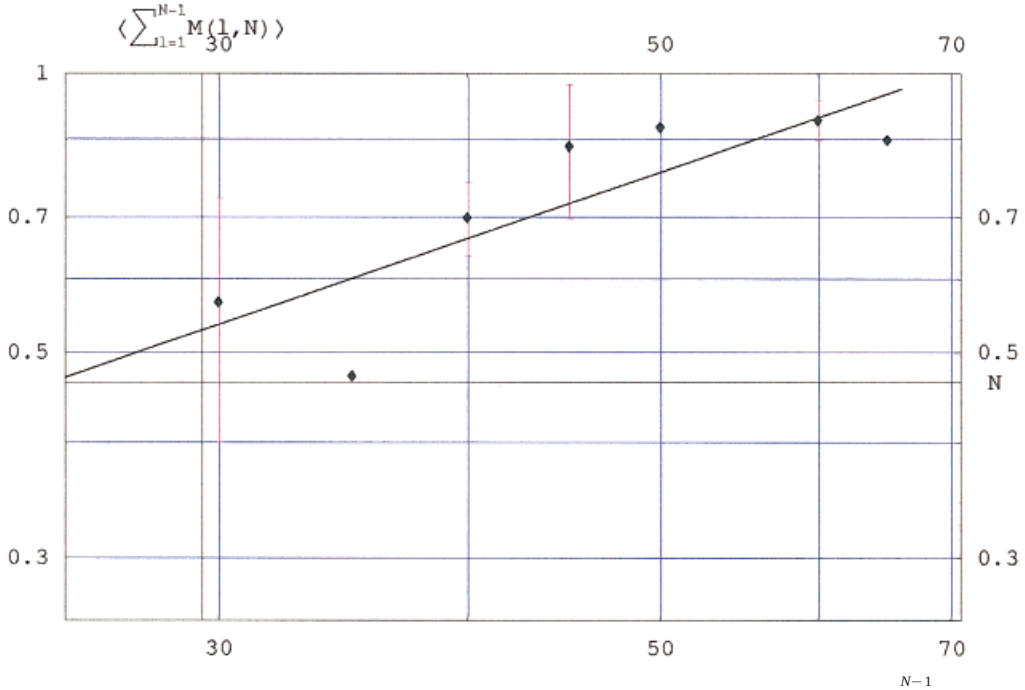

FIGURE 6 Log–log plot of the sequence variability distribution $\Phi(N-1, N) = \sum_{l=1}^{N-1} M(l, N)$ vs N, obtained from experimentally measured elution profiles of Table II. A power law is observed with $D_{a,\text{measurable}} = 1.26$. See text for explanation.

to correlate the developed expressions of D functions[27] to such more complex dynamic approach.

First, we describe nucleotide polymerization by classic one-dimensional Cantor sets. The length scale is clustering in volume growth V in exponential form where the power law follows Eq. (1). Furthermore, the construction is such that D_o takes the form of Eq. (6). The approach is very common.[30] The power-law distribution is the only distribution independent of any length scale. Thus, it allows infinite iterations making the remaining line lengths (Figure 4) shorter and shorter (Cantor dust[12]) and leads to a theoretical global fractal dimension D_o as presented in Eq. (6).

Second, in nature there are scaling limits everywhere[31]; sets and distributions obey fractal scalings within limited scale invariance.[32,33] Therefore, driven, multicyclic syntheses of different nucleotide length N of the largest optimal component (target) in the nucleotide region of interest are fractal under certain constraints.[28]

We have been able to calculate a well-defined theoretical response of a fractal system by deterministic quantitation. Two types of dynamic behavior were confirmed by experiments done for this paper (Figures 3 and 4, Table II). These types depend on the growth coefficient d. The fundamental character of chemical oligonucleotide synthesis does not change, but the dynamic behavior of the reported syntheses is different. For example, for the polystyrene-grafted polytetrafluoroethylene supports or polystyrene primer support (Pharmacia; see Table II), the optimal average growth was found dependent on longer target oligonucleotide preparations. On the other hand, using the polyvinylalcohol support and the CPG-500 Å support we obtained here an optimum of average (constant) growth at target nucleotide length N of 20 as well as 30 nucleotides. This is the first time that it has become possible to give a complete description of the number of all individual components and their relative as well as molar ratio within a solid-phase product based on measurements of the product distribution of individual chain length. We obtain solution profiles of a nonlinear growth process, which allow us to treat them as fractal by fractal power law formulation.

APPENDIX: DEFINITIONS

Some quantities used in the text are defined here:

N: Nucleotide length of the largest optimal component (target) in the

nucleotide region of interest.

l: Nucleotide length of sequence variabilities.

l_{min}: Smallest peak resolved in the measured chromatogram/electropherogram.

i: Number of reaction cycle.

$M(l, N)$: Probability density (first moment) of all sequence variabilities including the target $N = l$.

$M^{\cdot}(l, N)$: Probability density (first moment) of nonterminated sequence variabilities.

$\Phi(l, N)$: Cumulative probability distribution of all sequence variabilities including the target $N = l$.

d_i: Propagation probability at i.

d_o: Average (constant) propagation probability at i.

p_i: Termination probability at i.

p_o: Average (constant) termination probability at i.

A^*: Measured absorbance, e.g., at 254 or 260 nm. In papers sometimes used as $A_{260\,nm}$.

$A^{(expl)}(l, N)$: Experimental or theoretically calculated relative area of the absorbance peak l in the chromatogram/electropherogram. In papers sometimes used as A_{260} (concentration) or OD_{260} (amount).

$f(l, N)$: Theoretically calculated $A^{(expl)}(l, N)$.

$g(l, N)$: Experimental $A^{(expl)}(l, N)$.

$\overline{\kappa_j}$: Measured sequence variability ratio.

κ_j: Calculated sequence variability ratio.

Δz_{max}: Absolute maximum miscalculation.

δz_{max}: Relative maximum miscalculation.

$D(N)$: Fractal dimension of the idealized multicyclic synthesis without error production, i.e., with synthesis of the target only.

$D_{a,measurable}$: Measurable fractal dimension of a multicyclic synthesis.

D_a: Fractal dimension of normalized data sets against the idealized multicyclic synthesis without error production.

We thank Referee One, acting for the editors, who gave detailed support in revising the manuscript. We thank Professor Dr. Rudolf Rigler, Department of Medical Biophysics, MBB, Karolinska Institute, Stockholm, Sweden, for discussion and support of a fellowship (Z.F.-P.) at the Karolinska Institute, Stockholm, from the Swedish Foundation Wenner-Gren. We gratefully acknowledge financial support by the Foundation for the Advancement of Molecular Biology at the University of Ulm, Germany. Dr.Dr. Zeno Földes-Papp thanks for a German Academic Exchange Fellowship and for technical support of this work from DIONEX Cooperation, Sunnyvale, CA, USA, and DIONEX GmbH, Idstein, Germany.

REFERENCES

1. Uhlmann, E. & Peyman, A. (1990) *Chem. Rev.* **90,** 545–584.
2. Gura, T. (1995) *Science* **270,** 575–577.
3. Letsinger, R. L. & Kornet, M. J. (1963) *J. Am. Chem. Soc.* **85,** 3045–3046.
4. Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85,** 2149–2154.
5. Beaucage, S. L. (1993) in *Protocols for Oligonucleotides and Analogs: Methods in Molecular Biology*, Vol. 20, Agrawal, S., Ed., Humana Press, Totowa, pp. 33–61.
6. Froehler, B. C. (1993) in *Protocols for Oligonucleotides and Analogs: Methods in Molecular Biology*, Vol. 20, Agrawal, S., Ed., Humana Press, Totowa, pp. 63–80.
7. Seliger, H. (1993) in *Protocols for Oligonucleotides and Analogs: Methods in Molecular Biology*, Vol. 20, Agrawal, S., Ed., Humana Press, Totowa, pp. 391–435.
8. Agrawal, S., Ed. (1993) *Protocols for Oligonucleotides and Analogs: Methods in Molecular Biology*, Vol. 20, Humana Press, Totowa.
9. Földes-Papp, Z., Angerer, B., Ankenbauer, W., Baumann, G., Birch-Hirschfeld, E., Björling, S., Conrad, S., Hinz, M., Rigler, R., Seliger, H., Thyberg, P. & Kleinschmidt, A. K. (1998) in *Fractals in Biology and Medicine*, Vol. 2, Nonnenmacher, T. F., Merlini, D., Losa, G. A. & Weibel, E. R., Eds., Birkhäuser, Boston, in press.
10. Mandelbrot, B. B. (1983) *The Fractal Geometry of Nature*, Freeman, New York, pp. 79–83.
11. Mandelbrot, B. B. (1975) *Les Objects Fractals: Forme, Hasard et Dimension*, Flammarion, Paris.
12. Mandelbrot, B. B. (1994) in *Fractals in Biology and Medicine*, Nonnenmacher, T. F., Losa, G. A., Weibel, E. R., Eds., Birkhäuser, Boston, pp. 8–21.
13. Földes-Papp, Z., Herold, A., Seliger, H. & Kleinschmidt, A. K. (1994) in *Fractals in Biology and Medicine*, Nonnenmacher, T. F., Losa, G. A. & Weibel, E. R., Eds., Birkhäuser, Boston, pp. 165–173.
14. Birch-Hirschfeld, E., Földes-Papp, Z., Gührs, K.-H. & Seliger, H. (1996) *Helv. Chim. Acta* **79,** 137–150.

15. Földes-Papp, Z., Birch-Hirschfeld, E., Rösch, R., Hartmann, M., Kleinschmidt, A. K. & Seliger, H. (1995) *J. Chromatogr. A* **706,** 405–419.
16. Birch-Hirschfeld, E., Eickhoff, H., Stelzner, A., Greulich, K. O., Földes-Papp, Z., Seliger, H. & Gührs, K.-H. (1996) *Collect. Czech. Chem. Commun.* **61,** 311–314.
17. Seliger, H. (1975) *Makromol. Chem.* **176,** 1611–1627.
18. Jansen, C. & Mueller, W. (1989) *Merck Spectrum* **2,** 1–5.
19. Warren, W. J. & Vella, G. (1993) *BioTechniques* **14,** 598–606.
20. Rose, D. J. & Jorgenson, J. W. (1988) *Anal. Chem.* **60,** 642–648.
21. Földes-Papp, Z., Birch-Hirschfeld, E., Eickhoff, H., Baumann, G., Peng, W.-G., Biber, T., Seydel, R., Kleinschmidt, A. K. & Seliger, H. (1996) *J. Chromatogr. A* **739,** 431–447.
22. Small, E. W. (1992) *Methods Enzymol.* **210,** 237–279.
23. Caruthers, M. H., Beaton, G., Wu, J. V. & Wiesler, W. (1992) *Methods Enzymol.* **211,** 3–20.
24. Vicsek, T. (1996) in *Fractal Geometry in Biological Systems,* Iannaccone, P. M., Khokha, M., Eds., CRC Press, Boca Raton, FL, p. 317.
25. Vicsek, T. (1989) *Fractal Growth Phenomena,* World Scientific, Singapore, pp. 9–46.
26. Seliger, H., Bader, R., Birch-Hirschfeld, E., Földes-Papp, Z., Gührs, K. H., Hinz, H., Rösch, R. & Scharpf, C. (1995) *React. Funct. Polym.* **26,** 119–126.
27. Földes-Papp, Z., Birch-Hirschfeld, E., Seliger, H. & Kleinschmidt, A. K. (1997) in *Spurenanalytische Bestimmung von Ionen: Ionenchromatographie und Kapillarelektrophorese,* Kettrup, A., Weiss, J. & Jensen, D., Eds., ecomed Verlagsgesellschaft, Landsberg, Germany, pp. 158–172.
28. Földes-Papp, Z., Peng, W.-G., Seliger, H. & Kleinschmidt, A. K. (1995) *J. Theor. Biol.* **174,** 391–408.
29. Bassingthwaighte, J. B., Liebovitch, L. S. & West, B. J. (1994) *Fractal Physiology,* Oxford University Press, New York.
30. Jannaccone, P. M. & Khotha, M., Eds. (1996) *Fractal Geometry in Biological Systems,* CRC Press, Boca Raton, FL.
31. Nonnenmacher, T. F., Losa, G. A. & Weibel, E. R., Eds. (1994) *Fractals in Biology and Medicine,* Birkhäuser, Boston.
32. Nonnenmacher, T. F. (1994) in *Fractals in Biology and Medicine,* Nonnenmacher, T. F., Losa, G. A. & Weibel, E. R., Eds., Birkhäuser, Boston, pp. 22–37.
33. Baumann, G., Barth, A. & Nonnenmacher, T. F. (1994) in *Fractals in Biology and Medicine,* Nonnenmacher, T. F., Losa, G. A. & Weibel, E. R., Eds., Birkhäuser, Boston, pp. 182–189.