Self-Replication of Hexadeoxynucleotide Analogues: Autocatalysis versus Cross-Catalysis

Dirk Sievers and Günter von Kiedrowski*

Abstract: In the presence of a water-soluble carbodiimide, 5′-protected trideoxynucleotide 3′-phosphates and 3′-protected 5′-aminotrideoxynucleotides reacted to give hexadeoxynucleotides with a central 3′–5′-phosphoramidate linkage. The trimer sequences were CCG and CGG and were used as both the 5′-amine and the 3′-phosphate derivatives. The kinetics of the reactions were monitored by HPLC and analyzed with the *SimFit* program. In the first set of experiments, the formation of a single hexamer was studied in the absence of any hexadeoxynucleotide and in the presence of one of four hexadeoxynucleotides with the sequence of the reaction products. Here, self-complementary hexamers were always found to be formed faster than non-self-complementary hexamers because of autocatalytic self-replication. In the second set of experiments, the simultaneous formation of all four reaction products allowed both autocatalysis and cross-catalysis to occur. In this case all hexamers were synthesized with similar rates. In all experiments template effects were only observed for those sequences which were complementary to the reaction product. Autocatalysis as well as cross-catalysis was always accompanied by product inhibition leading to parabolic growth characteristics.

Keywords: kinetics · oligonucleotides · reaction mechanisms · self-replication · template-directed reactions

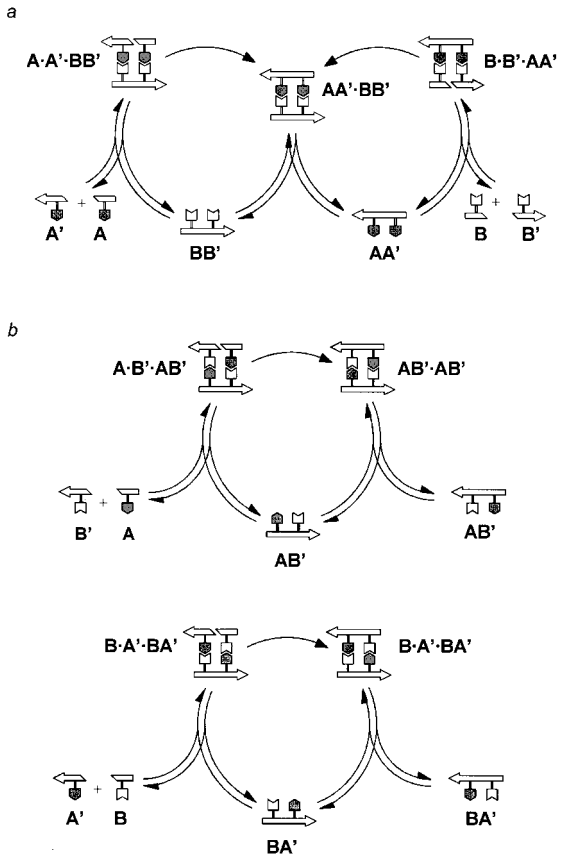
Introduction

The present design of artificial self-replicating systems is based on the principle of self-complementary templates.[1] Typically, a self-complementary template molecule is synthesized autocatalytically from two complementary template fragments; several examples for this type of self-replication have been published since 1986.[2] In contrast, natural nucleic acid replication utilizes two complementary strands rather than one self-complementary strand. We recently reported a minimal implementation of this type of replication based on cross-catalytic rather than autocatalytic reactions.[3] The general principle of a cross-catalytic self-replicating system is illustrated in Scheme 1a.[4] The templates AA′ and BB′ pick up their complementary fragments A, A′, B, and B′ to reversibly form the termolecular complexes $B \cdot B' \cdot AA'$ and $A \cdot A' \cdot BB'$, respectively. In these complexes the reactive ends of the respective fragments are in close spatial proximity, enabling ligation of the fragments by a covalent bond. Note that two pathways exist for the irreversible and rate-determining formation of the template duplex $AA' \cdot BB'$. Each template acts as a catalyst for the formation of its complementary copy, which in the parallel pathway acts as a catalyst for the formation of the original template (cross-catalysis). In other words, each couple of ligation steps is expected to double the total number of template molecules, if all template duplexes dissociate into single-stranded template molecules. As long as both templates are formed by the same type of chemistry, conditions that enable the cross-catalytic formation of complementary products AA′ and BB′ will also allow the parallel autocatalytic synthesis of self-complementary products AB′ and BA′ (Scheme 1b). Therefore a cross-catalytic self-replication cycle should be favoured if all four template-directed reactions proceed with similar rates.

The question of nonenzymatic cross-catalytic self-replication of oligonucleotides was first addressed by L. E. Orgel and colleagues.[5] Experiments on the template-directed polycondensation of nucleoside 5′-phosphoimidazolides revealed the general result that pyrimidine-rich oligo- and polynucleotides are operative as templates, whereas purine-rich strands are not.[5b] Since each pyrimidine-rich template generates a purine-rich product which in turn is inoperative as a template, it was concluded that nonenzymatic self-replication is unlikely to be achieved using activated mononucleotides as building material.[5c] The reported examples of self-replicating hexa-

[*] Prof. Dr. G. von Kiedrowski,[+] Dr. D. Sievers
Lehrstuhl für Organische Chemie I, Bioorganische Chemie
Ruhr-Universität Bochum, Universitätsstrasse 150
D-44780 Bochum (Germany)
Fax: (+ 49) 234-700-3218
E-mail: kiedro@ernie.orch.ruhr-uni-bochum.de
[+] Current address: Collegium Budapest
Institute for Advanced Studies
Szentháromság utca 2, H-1014 Budapest (Hungary)

  0947-6539/98/0404-0629 $ 17.50+.25/0



Scheme 1. a) Cross-catalytic self-replication of complementary templates; b) autocatalytic self-replication of self-complementary templates.

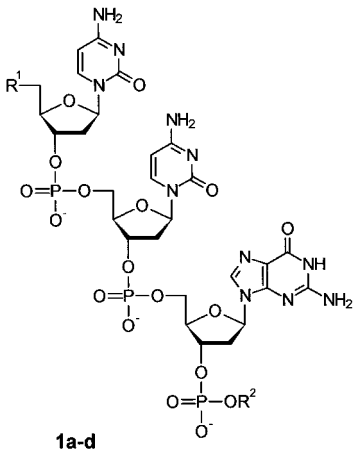
deoxynucleotides[2a,c,e] from trimeric building blocks were considered as special cases, because these examples were not based on arbitrary sequences but on self-complementary sequences.[1j] It is therefore appropriate to think about a generalization of replication schemes and to ask whether or not purine-rich templates can self-replicate in the absence of enzymes. Experiments on template-directed synthesis of hexadeoxynucleotides from trimeric fragments revealed that the condensation reactions are predominantly controlled by the stacking of the nucleotide bases flanking the reaction site.[6] The rate of condensations of trimers $B^1B^2B^3p$ and H_2N-$B^4B^5B^6$ (where B^1 to B^6 = C or G) decreased in the order B^3B^4: GG > GC > CG > CC,[6] a reactivity order which was also studied in chemical ligations of longer oligonucleotides.[7] From the above observations we concluded that a cross-catalytic self-replicating system may be operative if both the pyrimidine-rich and the purine-rich strand bear the same subsequence B^3B^4, since this would allow for similar condensation rates. For the case of hexamers containing only C and G as bases, the subsequence B^3B^4 may be either CG or GC, where GC was considered as the better choice for the above reasons. If B^3B^4 is chosen to be GC, B^1B^6 must be CG, since in the alternative combination of trimers B^3B^4 is determined by B^6B^1 of the original combination. The remaining bases of the trimer set, B^2B^5,were chosen such that one of the self-complementary hexamers would have the same sequence as in a system studied previously.[1e]

Another approach towards the implementation of a cross-catalytic replication system based on reciprocal template effects was recently described by Pieters et al.[8] In this system a concave template bearing two artificial receptors for adenine stimulated the synthesis of a convex bis-adenosine derivative which in turn acted as a template for the synthesis of the concave molecule. Although reciprocal template effects could be confirmed by initial kinetics, evidence for a full cross-catalytic replication cycle, which demands the simultaneous formation of both templates, was not given. On the basis of general conclusions from our present study, we are tempted to predict that the existence of a cross-catalytic replication cycle would be difficult to prove in this system, because the rates of formation of the complementary templates differ significantly. Moreover, since the self-complementary products are formed substantially faster than the complementary products, it seems most likely that the self-complementary products will win the race in a competition situation where all four products are synthesized simultaneously.

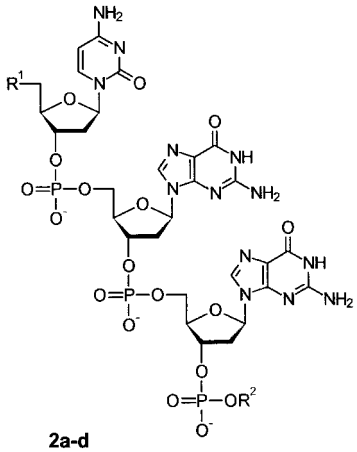
Results and Discussion

1. Experiments on the formation of single hexamers

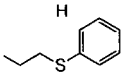
Basic reactions: Self-replication experiments were performed using a set of monoterminal protected trideoxynucleotides (Set I). Generally, the nucleotide sequence CCG is abbreviated with A, whereas the sequence CGG is referred to as B; symbols n and p always denote a 5′-amino- and a 3′-phosphate group and are referred to as amino and phosphate compounds, respectively. Compounds were synthesized by phosphotriester chemistry.[9, 10] In analogy to our earlier systems



1a-d



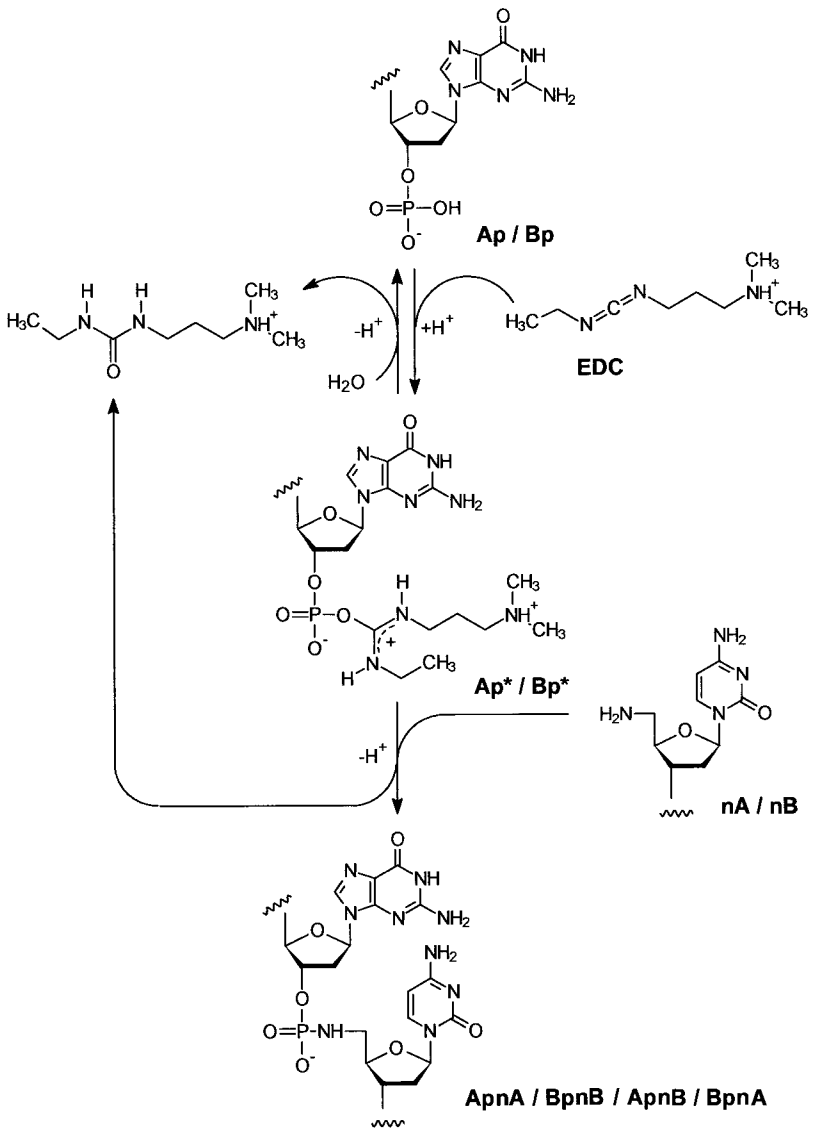
2a-d

1,2	R^1	R^2
a	N_3	H
b	CH_3SCH_2O	H
c	NH_2	
d	NH_2	

Set	Ap	Bp	nA	nB
I	1a	2a	1c	2c
II	1a	2b	1d	2d

0947-6539/98/0404-0630 $ 17.50+.50/0 *Chem. Eur. J.* **1998**, *4*, No. 4

tpl=-4mm>the water-soluble carbodiimide 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC) was employed as a condensing agent to activate the 3′-phosphate compounds, which then react with the 5′-aminotrideoxynucleotides to form hexadeoxynucleotides bearing a central 3′–5′-phosphoramidate linkage (Scheme 2).



Scheme 2. Reactions of Ap, Bp, nA and nB in the presence of water-soluble carbodiimide EDC.

Preliminary studies: In the first series of 20 individual experiments, one single 5′-protected trideoxynucleotide 3′-phosphate, ^{N_3}CCGp (Ap, N_3 is 5′-azido) or ^{N_3}CGGp (Bp), was reacted with one single 5′-aminotrideoxynucleotide 3′-(2-phenylthioethyl)-phosphate, nCCGp^{PTE} (nA) or nCGGp^{PTE} (nB), both in the absence of any and in the presence of one of the four hexadeoxynucleotide templates CCGCCG (ApA), CCGCGG (ApB), CGGCCG (BpA) and CGGCGG (BpB). All four templates bear a central 3′–5′-phosphodiester bond instead of a phosphoramidate linkage. The lack of any lipophilic protective groups at their terminal ends caused shorter retention times in reversed-phase chromatography. The template effects, however, are comparable to the respective products as follows from these and earlier results. The idea was to find out whether one or more templates are able to catalyze the formation of one or more of the following hexameric 3′-5′-phosphoramidates ApnB, BpnA, ApnA and BpnB. Product formation was monitored by HPLC. After a reaction time of 1 h the reactions are far from completion. Therefore, the yields of the hexamer reaction products in each single experiment provide a crude measure of the respective initial reaction rates. From the data in Table 1 three observations can be made: Firstly, only one of the four hexameric templates was able to catalyze the formation of one of the four 3′–5′-phosphoramidates. For each condensation reaction three templates did not at all influence the yields, which were found to be identical to those obtained in the absence of any template molecule. Secondly, the active template is always complementary to the phosphoramidate formed by ligation of the trideoxynucleotides. Thirdly, our experiments revealed similar yields for the pair of self-complementary phosphoramidates ApnB and BpnA as well as for the pair of complementary phosphoramidates ApnA and BpnB; however, yields of complementary products were significantly smaller than those of self-complementary templates.

Table 1. A comparision of 20 single experiments using set I of trideoxynucleotides: yields (%) of hexameric 3′–5′-phosphoramidates ApnB, BpnA, ApnA and BpnB in the absence (–) and in the presence of one single hexadeoxynucleotide template (ApA, BpB, ApB and BpA) after a reaction time of 60 min. Reaction conditions and initial concentrations (indicated by subscript zero): $T = 30\,^{\circ}C$, $[Ap]_o = [Bp]_o = [nA]_o = [nB]_o = 1$ mM, $[template]_o = 0/0.08$ mM, $[EDC]_o = 0.2$ M, [HEPES (Na^+/H^+)] = 0.1 M, pH 7.55.

	–	+ ApA	+ BpB	+ ApB	+ BpA
ApnB	23 %	23 %	23 %	32 %	23 %
BpnA	21 %	21 %	21 %	21 %	29 %
ApnA	4 %	4 %	15 %	4 %	4 %
BpnB	4 %	15 %	4 %	4 %	4 %

Our findings suggest template-directed reactions are always initiated by the formation of a termolecular complex. The observed template effects reflect the expected stabilities of these complexes, which in turn depend on the number of complementary base pairs to be formed between the trideoxynucleotides and the respective template molecule (Figure 1). In the presence of a fully complementary active template the termolecular complex is stabilized by a maximum of six cytidine–guanosine pairs. Even one single mismatch between two pyrimidines or two purines hinders the formation of the

termolecular complex and therefore prevents any template-directed reaction. Any hexameric 3′–5′-phosphoramidates observed in these cases must result from the existence of template-independent pathways. The difference in yields observed between self-complementary and non-self-complementary products is expected if template effects are operative. Only a self-complementary condensation product can auto-

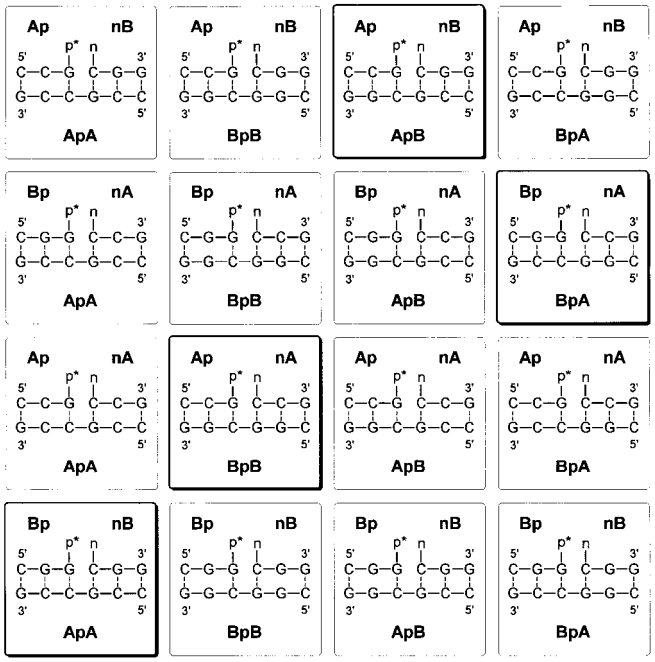
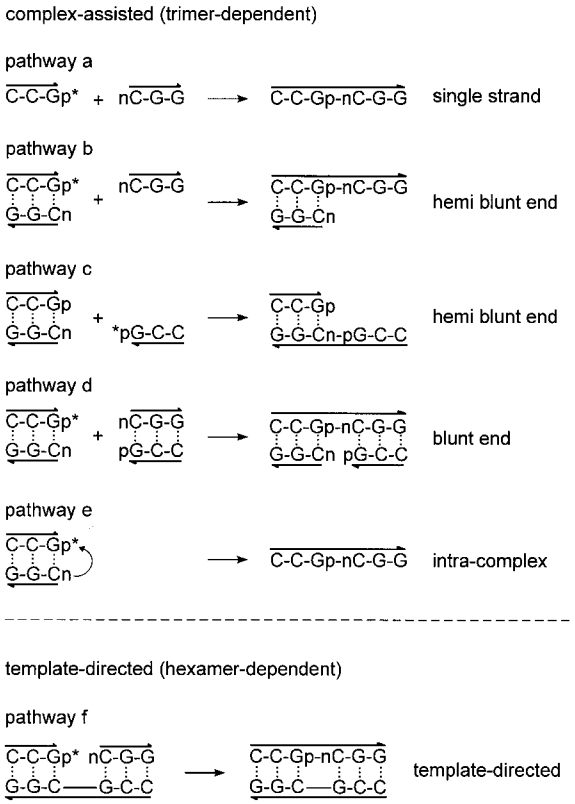


Figure 1. Possible termolecular complexes as theoretically expected for the separate formation of a single hexamer. Bold frames mark those complexes that contain a maximum of six complementary C/G base pairs. Terminal protective groups are omitted.

catalytically promote its own synthesis and hence accelerate the reaction pathway leading to template formation. In contrast a non-self-complementary product cannot catalyze its own synthesis and therefore its yield is expected to be significantly lower. The effect of external addition of a self-complementary template relative to the external addition of a non-self-complementary template is found to be less pronounced. This is again expected, since in the former case most of the product formation is catalyzed by the product itself (internal template) and not by the external template. After 1 h, for example, in the presence of 8% of external template BpA, the yield of internally synthesized BpnA increases from 21% to 29%. However, the assumption of template effects might be wrong and the higher yields of the self-complementary products might be due to the influence of complementarity between the trimers used. Self-complementary phosphoramidates are formed from complementary pairs of trideoxynucleotides, whereas non-self-complementary templates are formed from non-complementary pairs. It seemed conceivable that the formation of self-complementary products could involve complexes between complementary trideoxynucleotides (Scheme 3b–e). In order to distinguish between template-caused (template-directed) and trimer-



Scheme 3. Possible pathways for the formation of a self-complementary hexadeoxynucleotide: a) single-strand ligation, b/c) hemi-blunt-end ligation, d) blunt-end ligation, e) intracomplex ligation, f) template-directed ligation.

caused (complex-assisted) catalysis, the kinetics of the formation of single hexamers had to be carefully scrutinized.

Kinetics of the individual formation of single hexamers: A series of experiments were carried out in which the formation of a single 3′–5′-phosphoramidate ApnB, BpnA, ApnA and BpnB (systems 1–4, respectively) was monitored as a function of the time in the presence of various concentrations of the corresponding complementary template. For each phosphoramidate 19 experiments were performed, which were organized into six groups for collective kinetic analysis. The composition of each group with respect to the template concentration is shown in Table 2. The HPLC elution profiles allowed the separation of all individual peaks in all cases (Figure 2). Time courses of product formation for all four

Table 2. Composition of each of the six groups with respect to template concentrations. Note that in group 6 template concentration is 0.12 mM for non-self-complementary sequences and 0.32 mM for self-complementary sequences (see Table 1 for other conditions).

Group	0 mM	0.04 mM	0.08 mM	0.12 mM/0.32 mM	0.16 mM
1	+	–	+	–	–
2	+	+	+	–	–
3	+	–	+	–	+
4	+	+	–	–	+
5	+	+	+	–	+
6	+	–	+	+	+

0947-6539/98/0404-0632 $ 17.50+.50/0 *Chem. Eur. J.* **1998**, *4*, No. 4

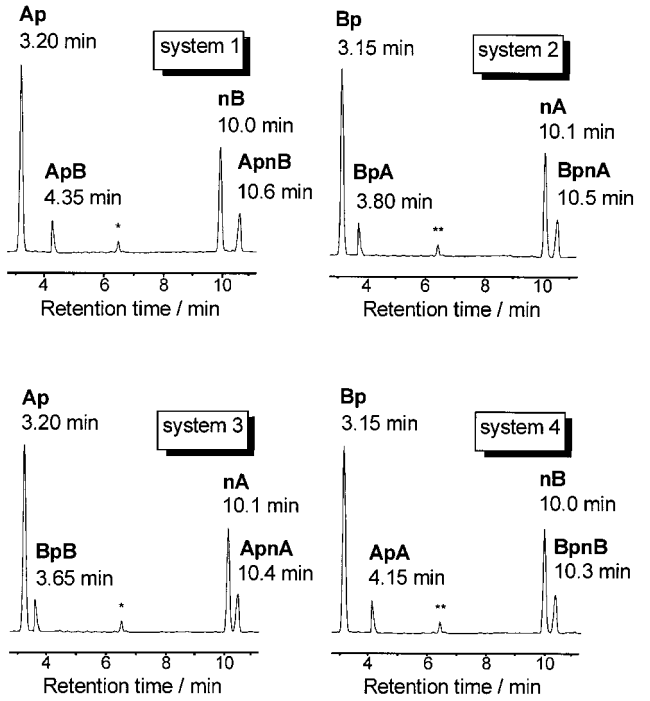


Figure 2. HPLC elution profiles for the separate formation of a single hexamer ApnB (system 1), BpnA (system 2), ApnA (system 3) and BpnB (system 4) from one single phosphate compound Ap or Bp and one single amino compound nA or nB of Set I after a reaction time of 90 min. ^{N_3}CCG and ^{N_3}CGG were used as standards and are shown as * and ** on the elution profiles.

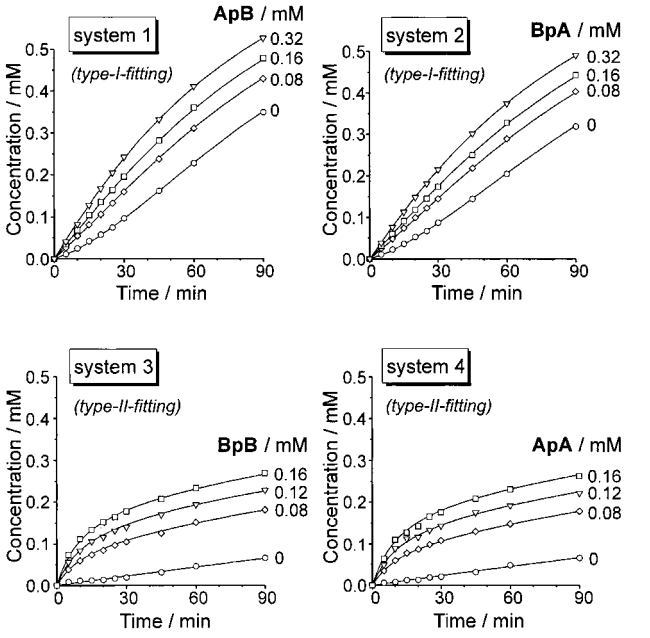
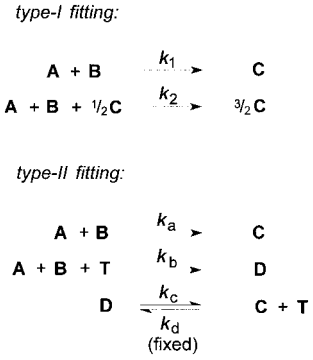


Figure 3. Experimental points and theoretical curves (always group 6) describing the time courses of formation of ApnB (system 1), BpnA (system 2), ApnA (system 3) and BpnB (system 4) from the respective trimers of Set I in the absence and the presence of variable concentrations of hexadeoxynucleotide templates ApB, BpA, ApA and BpB. In the calculations using the program *SimFit* the initial concentrations of the model species were calculated from the respective HPLC peak areas by calibration data for each compound. Nonlinear optimization was based on least-squares approximations employing the Newton–Raphson algorithm. The resulting rate constants, error estimates and covariances are presented in Tables 3 and 4.

systems are shown for groups 6 (Figure 3). Experimental data were rationalized using our *SimFit* program.[11] Calculations were based on two different simplified reaction models (Scheme 4). It was assumed that the template properties of externally added and internally synthesized templates are comparable. The validity of this assumption has been explicitly demonstrated in earlier systems, in which templates with a central 3′–5′-phosphoramidate bond and with a central 3′–5′-phosphodiester linkage were found to be barely distinguishable in their kinetic properties even if different terminal groups were involved.[1e, 6]



Scheme 4. Type-I and type-II fitting. Both simplified reaction models were employed to analyze the separate formation of single hexamers (systems 1–4). A, B: trimers; C: hexamer reaction product; T: template; D: duplex (between two hexamer strands). The reaction order of ½ (type-I fitting) appears as a consequence of the square-root law of autocatalysis. k_d was fixed to $10^6\,M^{-1}s^{-1}$ for all hexamers in our calculations.

Type-I fitting (Scheme 4) was employed to analyze both self-complementary systems (systems 1 and 2). The reaction model, based on the square-root law of autocatalysis, simply considers one template-independent pathway (k_1) and one template-directed pathway (k_2); the latter is characterized by a reaction order of ½ for the template. The rationale for this simplification is given in detail in our theory of minimal replicators.[12] Alternatively, a kinetic analysis of such reaction systems may be based on the full reaction model including all conceivable elementary steps.[2o] Although *SimFit* is programmed to deal with such situations, the experimental data do not necessarily support full modelling. This is because several rate and equilibrium constants are empirically linked and thus cannot be determined independently from each other. For example, the autocatalytic rate constant k_2 comprises not only the first-order rate constant for the transition of the termolecular complex to the template duplex but also the association constants for the latter complexes. Reliable thermodynamic data for the formation of termolecular complexes from self-complementary templates cannot be determined by current methods, and even for relatively stable termolecular complexes formed from non-self-complementary templates reliable thermodynamic data are difficult to ascertain.[13a] Consequently, full modelling requires the estimation of such formation constants from individual interactions[2o] and therefore a level of arbitrariness is introduced into the determination of other rate constants. In contrast, minimal modelling results in composite (apparent) rate constants whose values are more reliable, and therefore comparison of different systems is possible. The advantage of the computer program *SimFit* is that it can help to decide which and how many rate constants can be derived from the experimental data with sufficient independence. This is done by computing

the covariance matrix of the rate constants resulting from a given model.

Type-II fitting (Scheme 4) was employed to analyze both non-self-complementary systems (systems 3 and 4). In this case the rate constant k_a describes the template-independent synthesis of the phosphoramidates. In contrast, k_b results from the reversible formation of the termolecular complex and the irreversible template-directed reaction which yields the respective hexamer duplex. k_c deals with the reversible dissociation of the hexamer duplexes to give single-stranded hexadeoxynucleotides. The rate constant for the duplex association (k_d) can be chosen arbitrarily, so long as the irreversible reactions remain rate-determining (internal equilibrium). In our calculations we used a value of $10^6\,\text{M}^{-1}\,\text{s}^{-1}$ for all hexamers. Note that the equilibrium constants of the template–product complexes are not explicitly involved as parameters of the *SimFit* computation but can be easily derived from k_c and k_d. Deriving rate instead of equilibrium constants for reversible reactions offers the advantage that slow equilibria can be treated as well. This is in contrast to those fitting procedures where fast equilibria are computed separately prior to an integration step.[2o]

Activation of the 3′-phosphates by EDC and the hydrolysis of the activated species are not explicitly considered in either model. These processes are known to proceed as fast quasireversible reactions establishing the stationary concentration of activated phosphates within some minutes.[13b] Furthermore, as EDC is present in 100-fold excess over the 3′-phosphates, the degree of activation of the latter is assumed to be independent of time over the first 90 min. Thus the rate constants k_a and k_b are apparent rate constants, which depend on the concentration of the carbodiimide.

Theoretically the rate constants k_2 (type-I fitting), and k_b and k_c (type-II fitting), which describe the template-dependent product formation, are related by a simple approximation [Eq. (1)].[12] An apparent rate constant k_2 results from an

$$k_2 \approx k_b \times (k_d/k_c)^{-1/2} \quad (1)$$

arbitrary number of k_b/k_c pairs. As long as no other conditions specify this pair of rate constants, a splitting of k_2 into its component rate constants [Eq. (1)] inevitably surcharges the input data. This limitation is true for both self-complementary systems, where the time courses of product formation for different k_b/k_c pairs are qualitatively equivalent in shape. Therefore, type-II fitting failed to analyze the formation of self-complementary hexamers, but could be applied to the formation of non-self-complementary hexadeoxynucleotides, for which the ratio between template–trimer complexes and template–hexamer complexes depends strongly on the reaction time. At the beginning of the reaction the concentration of single-stranded templates available to form termolecular complexes is almost identical to the entire concentration of externally added templates; thus, initially no product inhibition can take place. At late reaction times, the concentration of hexameric reaction products exceeds the concentration of externally added templates and a significant number of complementary template molecules are now deactivated by duplex formation. Therefore different k_b/k_c pairs (k_2 fixed) cause different patterns of product formation. For example a large k_b (and small k_c) describes a fast template-directed reaction at early reaction times that slows down significantly later in the reaction owing to the formation of stable duplexes from hexameric phosphoramidates and the corresponding templates. Therefore the number of single-stranded templates, which are essential for the formation of termolecular complexes, decreases significantly. In contrast, a small k_b (and large k_c) describes a slow template-directed reaction which, due to low duplex stability, slows down only slightly as it proceeds. The obtained experimental data of both complementary systems, that is, the pattern of the time courses of product formation, could be precisely correlated with one single k_b/k_c pair.

The apparent rate constant of the template-independent formation of non-self-complementary hexamers (k_a) could be determined independently. For this purpose, experiments were performed in the absence of any template molecule. No sequence with inherent template properties was formed in these experiments and no acceleration of product formation took place. This is again in sharp contrast to the formation of self-complementary hexadeoxynucleotides where, even in the absence of any externally added templates, the reaction products were able to autocatalytically promote their own formation. The results obtained from numerical fitting of the experimental data within each of the six groups of each of the four systems are summarized in Tables 3 and 4. From Table 4 the average dissociation constants for the template–product complexes can be calculated to be $K_{ass} = 7.61 \pm 0.07\,\mu\text{M}$ for system 3 and $K_{ass} = 7.07 \pm 0.04\,\mu\text{M}$ for system 4.

Template-independent product formation: Comparison of the kinetic data showed very similar rate constants for self-complementary systems and also for non-self-complementary systems in template-independent reactions. Obviously reactions proceed faster in the former case, where trimeric template fragments are complementary to each other. These findings can only be explained if one or more complex-assisted pathways in the template-independent formation of self-complementary products occurs. No such pathways can

Table 3. Results obtained from numerical fitting (type I) of the experimental data of both self-complementary systems (systems 1 and 2). The rate constants are listed together with their error estimates. Cov indicates the respective covariances; R is the percentage of the mean absolute difference between experimental and theoretical concentrations with respect to the mean of experimental concentrations.

System/group	k_1 [$10^{-2}\,\text{M}^{-1}\,\text{s}^{-1}$]	k_2 [$\text{M}^{-3/2}\,\text{s}^{-1}$]	Cov (k_1, k_2)	R [%]
1/1	1.77 ± 0.04	7.24 ± 0.05	-0.923	$+1.0229$
1/2	1.76 ± 0.03	7.30 ± 0.03	-0.939	$+0.8698$
1/3	1.77 ± 0.03	7.24 ± 0.03	-0.912	$+0.7597$
1/4	1.78 ± 0.04	7.25 ± 0.03	-0.920	$+1.0225$
1/5	1.74 ± 0.05	7.31 ± 0.04	-0.928	$+1.2695$
1/6	1.79 ± 0.04	7.24 ± 0.03	-0.896	$+0.9341$
2/1	1.73 ± 0.02	6.14 ± 0.02	-0.924	$+0.5555$
2/2	1.78 ± 0.03	6.16 ± 0.03	-0.943	$+0.9519$
2/3	1.76 ± 0.03	6.16 ± 0.03	-0.915	$+0.7343$
2/4	1.75 ± 0.02	6.19 ± 0.02	-0.920	$+0.6226$
2/5	1.71 ± 0.04	6.21 ± 0.03	-0.931	$+1.2605$
2/6	1.73 ± 0.03	6.18 ± 0.02	-0.899	$+1.0106$

0947-6539/98/0404-0634 $ 17.50+.50/0 *Chem. Eur. J.* **1998**, *4*, No. 4

Table 4. Results obtained from numerical fitting (type II) of the experimental data of both non-self-complementary systems (systems 3 and 4). See Table 3 for the quantities listed.

System	k_a [10^{-2} M^{-1}s^{-1}]	k_b [10^{3} M^{-2}s^{-1}]	k_c [s^{-1}]	Cov (k_a,k_b)	Cov (k_a,k_c)	Cov (k_b,k_c)	R [%]
3/1	1.33 ± 0.01	1.76 ± 0.02	7.60 ± 0.24	-0.014	-0.226	-0.867	$+1.138$
3/2	1.33 ± 0.01	1.82 ± 0.02	7.55 ± 0.18	$+0.021$	-0.335	-0.867	$+0.955$
3/3	1.27 ± 0.01	1.77 ± 0.02	7.51 ± 0.19	-0.016	-0.315	-0.800	$+1.033$
3/4	1.30 ± 0.01	1.78 ± 0.01	7.67 ± 0.15	$+0.018$	-0.344	-0.798	$+0.842$
3/5	1.31 ± 0.01	1.77 ± 0.01	7.67 ± 0.15	$+0.013$	-0.404	-0.788	$+0.911$
3/6	1.29 ± 0.03	1.79 ± 0.03	7.66 ± 0.39	-0.015	-0.373	-0.794	$+2.367$
4/1	1.31 ± 0.01	1.68 ± 0.02	7.01 ± 0.19	-0.013	-0.234	-0.862	$+0.876$
4/2	1.31 ± 0.01	1.68 ± 0.01	7.03 ± 0.15	$+0.008$	-0.340	-0.850	$+0.770$
4/3	1.31 ± 0.01	1.65 ± 0.01	7.13 ± 0.12	-0.020	-0.333	-0.789	$+0.612$
4/4	1.31 ± 0.01	1.69 ± 0.01	7.09 ± 0.16	$+0.017$	-0.353	-0.783	$+0.896$
4/5	1.31 ± 0.01	1.66 ± 0.01	7.06 ± 0.12	-0.005	-0.416	-0.771	$+0.766$
4/6	1.29 ± 0.03	1.63 ± 0.03	7.09 ± 0.39	-0.036	-0.386	-0.767	$+2.213$

exist for the template-independent formation of non-self-complementary strands. The difference in rate, however, is only about 30% under the conditions of the experiments. This means that most of the hexadeoxynucleotides are synthesized via single-stranded trimers (Scheme 3a).

Our data are unable to distinguish between the possible complex-assisted pathways (Scheme 4b–e). The populations of pathways b–d, which describe hemi-blunt-end and blunt-end ligations, are expected to be influenced by the stacking of the nucleobases at the reaction site. We speculate that in particular pathway d benefits from an increased stacking of the nucleobases at the reaction site of the complex. Earlier kinetic data did not reveal sufficient evidence for complex-assisted pathways under the given conditions. However, because of an increase in the accuracy of our data we now find that there is some, albeit not very significant, involvement of trimer–trimer complexes in our reactions. This is kinetically evident, although the equilibrium concentration of these species appears to be rather small when thermodynamic data are taken into account.[14a] Molecular modelling suggests that the distance between the reactive ends in the complex formed by pathway e (Scheme 3) is too large to allow for covalent bond formation.

Template-directed product formation: Rate constants for template-directed reactions differ only slightly. The major influence on determining k_2 for the self-complementary systems, and k_b and k_c for the non-self-complementary systems, is expected to arise from the stacking of the nucleotide bases flanking the nick in the termolecular complex. As well as the effects caused by this central stack the nearby heterobases also exert a notable influence on the ligation. Our findings are in excellent agreement with data from previous experiments, where the kinetic properties of the self-complementary sequence CCGCGG were analyzed.[2e] Rate constants k_b and k_c are comparable to k_2 [Eq. (1)], so long as one considers that the formation of self-complementary duplexes is entropically disfavoured by a factor of 2 compared to the formation of complementary duplexes.[14b] Thus k_2 for a non-self-complementary sequence, calculated from Equation 1, must be multiplied by $\sqrt{2}$ for comparison with k_2 for a self-complementary sequence. When this symmetry factor is taken into account the rate constants of the individual reactions are considered similiar enough to be taken as identical in the elucidation of the parallel formation of all four sequences simultaneously (Table 5).

Table 5. Mean values of the rate constants k_1 (template-independent reactions) and k_2 (template-directed reactions) for all four systems describing the separate formation of single hexadeoxynucleotides.

System/fitting	k_1 [10^{-2}M^{-1}s^{-1}]	k_2 [M$^{-3/2}$s^{-1}]
1/I	1.78	7.26
2/I	1.74	6.17
3/II	1.31	4.91[a]
4/II	1.31	4.43[a]

[a] Calculated from k_b and k_c by Equation (1).

2. Experiments on combinatorial hexamer formation

Kinetics of combinatorial synthesis: The following experiments were performed with a new set of trimers (Set II, cf. Scheme 1a) because the hexamers from Set I could not be sufficiently resolved by HPLC when all four products were synthesized simultaneously. The four trimers of set II were reacted in the presence of a water-soluble carbodiimide EDC. The formation of the four 3′–5′-phosphoramidate-linked hexamers ApnA, BpnB, ApnB and BpnA from combinatorial synthesis was again monitored by HPLC. Typical elution profiles showing the composition of the reaction mixtures after 0, 30, 60 and 90 min reaction time are depicted in Figure 4. Figure 5 depicts the changes in the HPLC profile (90 min) caused by seeding the reaction mixtures with the respective templates. Figure 6a reflects the time course of product formation in the absence of any template. All four hexameric products are formed with similar rates, a result in sharp contrast to those from experiments on the individual formation of single hexamers (Figure 3), in which the yields of the self-complementary products (after 1 h) were five times higher than those of the non-self-complementary products (Table 1). The observation of similar reaction rates in combinatorial synthesis provides evidence for cross-catalysis between the complementary hexamers (for a general scheme see Scheme 1a). Cross-catalysis can occur if both complementary products are formed simultaneously. This is the case only during combinatorial synthesis but not during the individual formation of single hexamers.

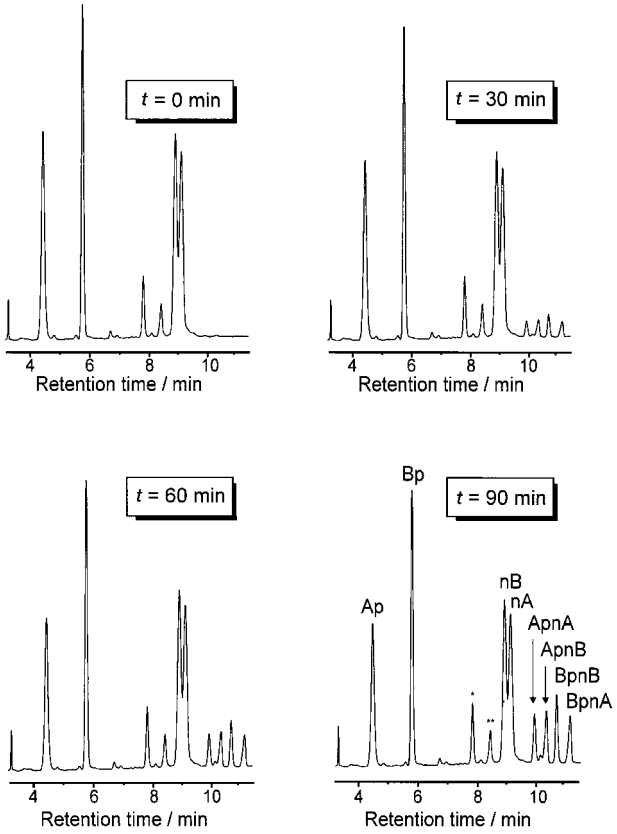


Figure 4. HPLC elution profile from the reaction between Ap, Bp, nA, nB (Set II) and EDC in the absence of any hexadeoxynucleotide template after reaction times of 0, 30, 60 and 90 min; ^{N_3}CCG and ^{MTM}CGG were used as standards, and are shown as * and ** on the elution profile. Reaction conditions and initial concentrations (indicated by subscript zero) as follows: $T = 30\,^{\circ}C$, $[Ap]_o = [Bp]_o = [nA]_o = [nB]_o = 1\,mM$, $[EDC]_o = 0.2\,M$, $[HEPES\ (Na^+/H^+)] = 0.1\,M$, pH 7.55.

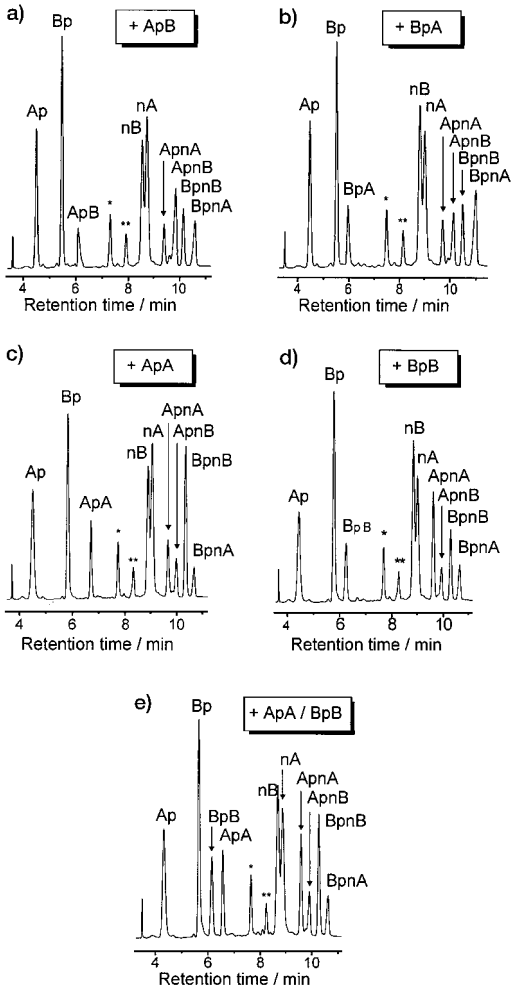


Figure 5. HPLC elution profiles from the reactions between Ap, Bp, nA, nB (Set II) and EDC in the presence of approximately 0.16 mM concentrations of hexadeoxynucleotide templates ApB (a), BpA (b), ApA (c), BpB (d), and ApA+BpB (e) after a reaction time of 90 min; ^{N_3}CCG and ^{MTM}CGG were used as standards, and are shown as * and ** on the elution profile. For reaction conditions and initial concentrations see Figure 4.

The reactions were then followed in the presence of one of the hexadeoxynucleotide templates ApA, BpB, ApB and BpA. Elution profiles of the reaction mixtures after a reaction time of 90 min are depicted in Figure 5a – e. Figure 6b – e provides evidence that each hexamer template stimulates the formation of the one and only reaction product whose sequence is complementary to the respective template sequence. Again, the effect of a single non-self-complementary template, ApA or BpB, is more pronounced than the effect of a single self-complementary template. When both ApA and BpB are present (Figure 5e and 6f), the effects are comparable to those of the self-complementary templates. Remarkably, the effects of ApA and BpB are similar, although the base compositions are different.

Kinetic modelling: Experimental data from Figure 6a – f were analyzed by using type-II fitting as a simplified reaction model. AA, BB, AB and BA are general species representing both externally added and internally synthesized hexameric templates. Equations (2 – 12) were employed to calculate the theoretical time courses of the entire experiments by means of the computer program *SimFit*. The experimental time courses of all experiments were evaluated simultaneously to generate a single set of three apparent rate constants k_a, k_b and k_c from

$$Ap + nA \rightarrow AA \quad (2)$$

$$Bp + nB \rightarrow BB \quad (3)$$

$$Ap + nB \rightarrow AB \quad (4)$$

$$Bp + nA \rightarrow BA \quad (5)$$

$$Ap + nA + BB \rightarrow AA \cdot BB \quad (6)$$

$$Bp + nB + AA \rightarrow AA \cdot BB \quad (7)$$

$$Ap + nB + AB \rightarrow AB \cdot AB \quad (8)$$

$$Bp + nA + BA \rightarrow BA \cdot BA \quad (9)$$

$$AA + BB \rightleftharpoons AA \cdot BB \quad (10)$$

$$2\,AB \rightleftharpoons AB \cdot AB \quad (11)$$

$$2\,BA \rightleftharpoons BA \cdot BA \quad (12)$$

0947-6539/98/0404-0636 $ 17.50+.50/0 *Chem. Eur. J.* **1998**, *4*, No. 4

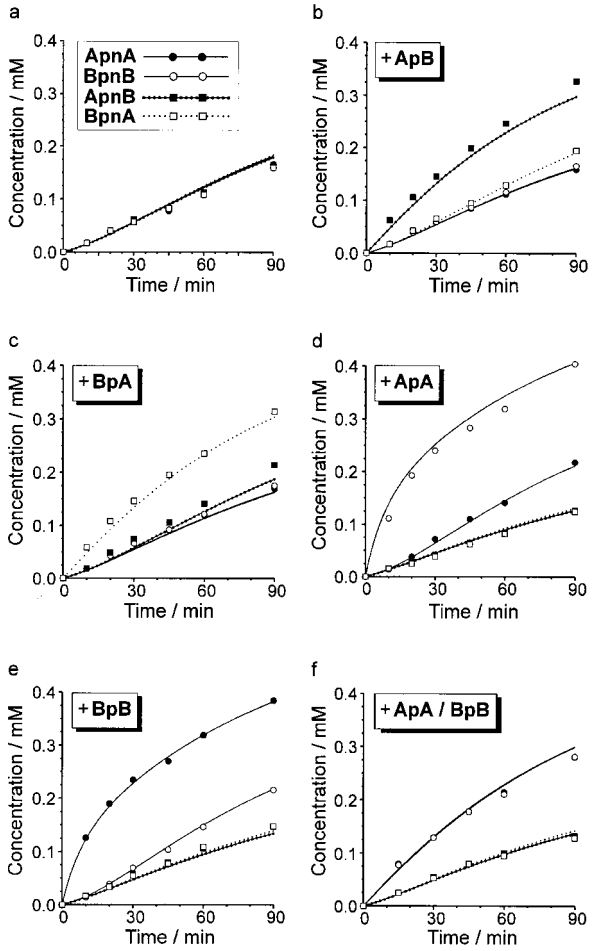


Figure 6. Experimental points and theoretical curves describing the time courses of the formation of ApnA, BpnB, ApnB, and BpnA (from Set II) in the absence (a), and in the presence of approximately 0.16 mM of hexadeoxynucleotide templates ApB (b), BpA (c), ApA (d), BpB (e), and ApA+BpB (f). In the *SimFit* computations, the initial concentrations of the model species were calculated from the respective HPLC peak areas with calibration data for each component. $[AA]_o$, $[BB]_o$, $[AB]_o$, and $[BA]_o$ were taken from the concentrations of ApA, BpB, ApB, and BpA, respectively. All duplex concentrations were set to zero. Theoretical species concentrations were computed by stiff integration of the rate equations which *SimFit* derived from the model. *SimFit* was instructed to compute theoretical values for the observable concentrations using the expressions: $[ApnA]_{theor} = [AA] + [AA \cdot BB] - [AA]_o$, $[BpnB]_{theor} = [BB] + [AA \cdot BB] - [BB]_o$, $[ApnB]_{theor} = [AB] + 2[AB \cdot AB] - [AB]_o$, $[BpnA]_{theor} = [BA] + 2[BA \cdot BA] - [BA]_o$. Nonlinear optimization was based on least-squares approximations employing the Newton–Raphson algorithm. The resulting rate constants and error estimates were $k_a = (2.07 \pm 0.04) \times 10^{-2}\,M^{-1}s^{-1}$, $k_b = (1.68 \pm 0.07) \times 10^{3}\,M^{-2}s^{-1}$, $k_c = (8.98 \pm 0.81)\,s^{-1}$. The off-diagonal elements of the covariance matrix were: $cov(k_a, k_b) = -0.368$, $cov(k_a, k_c) = +0.219$, $cov(k_b, k_c) = -0.970$.

which the theoretical time courses for all experiments were derived. We assumed equally efficient template-independent reactions for Equations (2–5) (k_a), equally efficient template-directed reactions for Equations (6–9) (k_b), and equally efficient duplex dissociations. The latter were characterized by the backward rate constants of Equations (10–12) (k_c) providing a symmetry factor of 2 for the self-complementary hexamers.[14b] The forward rate constants of reactions (10–12) were fixed to a value of $10^6\,M^{-1}s^{-1}$ for all hexamers. From the resulting value of k_c it follows that the dissociation constant of the complexes is around 10 μM at $T = 30\,^{\circ}C$, in good agreement with thermodynamic data on similar hexadeoxynucleotides.[6] The numerical results of our calculations are summarized in Table 6.

Table 6. Rate constants of the separate and combinatorial formation of hexadeoxynucleotides.

System	Fitting	k_a [$10^{-2}\,M^{-1}s^{-1}$]	k_b [$10^{3}\,M^{-2}s^{-1}$]	k_c [s^{-1}]
1	I	1.78	–	–
2	I	1.74	–	–
3	II	1.31	1.78	7.61
4	II	1.31	1.67	7.07
comb.	II	2.07	1.68	8.98

A comparison of these rate constants with the ones obtained from the analysis of the individual formation of single hexamers reveals the following results: Firstly, the rate constants for the template-directed processes (k_b, k_c) are in good agreement if the simplification of combined analysis is taken into account. Secondly the rate constants k_a describing the template-independent hexamer formation exhibit differences that cannot be explained by assuming similar reaction pathways. Instead, the increase in rate found in the combinatorial syntheses must be due to processes that cannot occur in separate syntheses (systems 1–4). In particular for trimer–trimer complexes, fully complementary complexes are twice as likely in the combinatorial experiments compared to the separate experiments involving self-complementary hexamers (systems 1 and 2). This is due to the fact that a trideoxynucleotide 3′-phosphate can not only pair with a complementary 5′-aminotrideoxynucleotide 3′-phosphate, but also with a complementary trideoxynucleotide 3′-phosphate. Furthermore, a trimeric amino compound can also pair with a complementary amino compound instead of a complementary phosphate compound. Assuming that complex-assisted pathways involving such complexes are comparable in efficiency with the pathways observed for the individual formation of self-complementary hexamers, one can expect a value of about 2.20×10^{-2} for the apparent rate constant k_a $[1.31 + 2(1.76 - 1.31)]$. Indeed, the experimentally observed value of 2.07×10^{-2} is higher than 1.76×10^{-2}. This shows that the complex-assisted pathways are indeed operative, although they may be not as productive in the combinatorial experiments as in the separate formation of self-complementary hexamers. Again, these pathways supporting the template-independent formation of hexamers exist only as a minor source of production.

Conclusion

Cross-catalytic self-replication can be observed in template-directed reaction systems if the reciprocal template effects are similar and if the reaction system is not dominated by autocatalytic syntheses of self-complementary products which occur as parallel reactions. In a minimal implementation of cross-catalytic self-replication each template is synthesized by

a single condensation step. For systems based on oligonucleotide templates the template effects depend mainly on the stacking of nucleobases flanking the condensation site. A purine-rich original has a similar template effect as its pyrimidine-rich copy if: 1) short oligomers instead of monomers are used as building material, and 2) both complementary strands bear the same base sub-sequence at the newly formed internucleotide link. The competing self-complementary autocatalysts then also bear the above base sub-sequence. The latter results in a kinetically balanced situation in which an autocatalytic product does not outcompete the cross-catalytic products. Such a situation may also be found in a similarly designed system based on the reaction between pRNA tetramers,[15a] whose kinetic analysis is eagerly awaited.[15b] Although neither our system nor the one outlined above is based on RNA oligomers, we believe that the study of such systems can give basic insights into the required replicative pathways that enabled RNA, or a similar class of molecules, to create its own world. This confidence is based on the fact that the present menagerie of chemical replicators, ranging from nucleic acids through artificial templates and even to peptides,[2p] is diverse from a structural point of view yet very similar with respect to the dynamics involved. So far every nonenzymatic chemical system exhibiting template feedback shows parabolic growth characteristics,[12] regardless of whether autocatalytic or cross-catalytic systems are involved. Theoretical biology tells us that the ability to evolve is not a consequence of a particular template structure but primarily a consequence of the order of autocatalytic feedback expressed by the whole system.[16] Molecular evolution of the Darwinian type simply necessitates exponential growth of the competing replicators.[17] The latter may be based on DNA, RNA or peptides as well as on artificial templates. Our research program will thus continue to search for other implementations of template replication that might enable exponential amplification under enzyme-free conditions. Screening the structural diversity of templates is an important but, in our program, secondary goal. The ultimate aim is a chemical algorithm enforcing exponential amplification and general enough to give room to a whole variety of template structures. Whether this goal can be achieved by Kauffman sets, that is collectively closed autocatalytic networks,[18] minimal replicases[2q] or by enforcing replication on surfaces[19] remains to be investigated.

Experimental Section

Synthesis of trideoxynucleotides (prepared as triethylammonium salts) and hexadeoxynucleotides is described elsewhere.[9] EDC (Merck), HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid, Aldrich), acetonitrile (Roth) and NH_4HCO_3 of p.a. or synthesis quality were used. Water was doubly distilled with a WM Muldestor SE apparatus (Wagner & Munz). Lyophilization was carried out by a SpeedVac concentrator (Savant). 1 μL and 2 μL precision capillaries (minicaps end to end, nonheparinized, accuracy $R \leq 0.5\,\%$, precision $CV \leq 1\,\%$) were from Hirschmann; pipettes (0.5–10, 10–100, 50–250 and 200–1000 μL) were from Eppendorf.

Self-replication experiments: A set of parent solutions of single oligonucleotides were prepared in Eppendorf tubes (1.5 mL) by pipetting doubly distilled water (1 mL) to the lyophilized trideoxynucleotides (40–60 A_{254}) and hexadeoxynucleotides (20–30 A_{254}). After vortexing and centrifuging the tubes, the concentration was determined by UV spectroscopy at $\lambda =$ 254 nm from an aliquot (5 μL) diluted with doubly distilled water to a total volume of 5 mL. The respective extinction coefficient ε_{254} ($L\,mol^{-1}\,cm^{-1}$) was calculated from the following increments: CMP: 6541, GMP: 13679, PTE: 8104 and *o*-ClPh: 200. Experiments required identical initial concentrations of starting material in each single experiment of a series. In order to increase the reproducibility of the experiments, multicompound parent solutions of two or four trideoxynucleotides were prepared by mixing defined aliquots (about 100 μL) of each trimer parent solution in Eppendorf tubes (1.5 mL). The composition of these multicompound parent solutions was monitored by HPLC and, if necessary, corrected to the nominal value. For self-replication experiments, aliquots of the respective multicompound parent solution (usually 8–22 nmol in each trideoxynucleotide) were lyophilized in Eppendorf tubes (0.7 mL) for 1 h using a SpeedVac evaporator. For experiments in the presence of a template, aliquots were colyophilized with the hexadeoxynucleotide (usually 0.3–3.5 nmol). A solution of the carbodiimide EDC (0.2 M) in HEPES buffer (0.1 M, pH 7.55) was freshly prepared under inert gas by dissolving the appropriate amount of EDC. The experiments were started by pipetting an aliquot of the carbodiimide solution (kept at 30 °C) to the lyophilized oligodeoxynucleotides. The mixtures were immediately vortexed, centrifuged and drawn into a series of 2 μL (separate experiments) or 1 μL precision capillaries (combinatorial experiments) provided with a silicon rubber ring. Usually ten or seven capillaries were prepared to allow for the monitoring of the reaction mixtures after ten or seven different reaction times in the separate or combinatorial experiments respectively. Drops of reaction solution were wiped off the outer walls of the capillaries, which were then placed in a plastic rack and stored in small airtight glass containers under an atmosphere saturated with water (bath thermostat at 30 °C). After the given time interval one capillary was removed and its contents transferred into a 250-fold volume of HPLC initial buffer (IB). The quenching procedure decreased the reaction rate by a factor of at least 62500, which was sufficient because HPLC analysis was done immediately after the reaction.

HPLC analysis: Analysis was performed on the following equipment (Kontron): two pumps 420, autosampler 460, ultraviolet detector 430, data acquisition on system D450; Nucleosil 120–7 C_{18} column (Macherey and Nagel). The initial buffer (IB) consisted of aqueous NH_4HCO_3 (0.1 M), the final buffer (FB) of acetonitrile/water 3:7 v/v). The following gradient program was employed for experiments on the separate formation of single hexamers: 20 % to 55 % FB within 10 min, 55 % to 75 % FB within 3 min, 75 % to 20 % FB within 4 min and final reequilibration for 10 min. The experiments on the combinatorial hexamer formation needed a slightly modified program: 15 % to 55 % FB within 10 min, 55 % to 75 % FB within 3 min, 75 % to 15 % FB within 4 min and final reequilibration for 10 min. Flow rate was 1 mL min^{-1}. Injection volume was 150 μL. HPLC peaks were detected by the simultaneous monitoring of the UV absorbances at $\lambda =$ 254 and 273 nm.

Baseline separation of HPLC peaks: The increased number of oligonucleotides involved in the combinatorial experiments (up to ten different compounds) hampered baseline separation of HPLC peaks when the trimers of set I were used. Exhaustive variations of the standard HPLC gradient did not result in a significant improvement. From earlier experiments it was known that the retention times of trideoxynucleotides are significantly influenced by the lipophilicity of their terminal protective groups. Therefore, two 5′-protected trideoxynucleotide 3′-phosphates, ^{MTM}CCGp (**1b**) and ^{MTM}CGGp (**2b**) were synthesized in which the 5′-azido group was *o*-chlorophenyl group. Furthermore two 5′-aminotrideoxynucleotide 3′-phosphate esters, nCCGp^{ClPh} (**1d**) and nCGGp^{ClPh} (**2d**) were synthesized replacing the 2-phenylthioethyl group with the less lipophilic *o*-chlorophenyl (ClPh) group. HPLC separation (Figure 7) was optimized by substitution of one single trimer of set I (5′-azido- and 3′-phenylthioethyl-protected) by the respective trideoxynucleotide of the new trimers (5′-MTM- and 3′-chlorophenyl-protected). No sufficient baseline separation for the hexamer peaks was observed (Figure 7a–d). However, in one case (Figure 7b) one missing hexamer peak seemed to be hidden under the partly separated peaks of the amino compounds. This remaining hexamer peak was observed by choosing less lipophilic 3′-*o*-chlorophenyl-protected aminotrideoxynucleotides instead of 3′-phenylthioethyl-protected compounds. This replacement resulted in the proper combination of trimers (Set II) enabling complete baseline separation of all four hexamer peaks

0947-6539/98/0404-0638 $ 17.50+.50/0 *Chem. Eur. J.* **1998**, *4*, No. 4

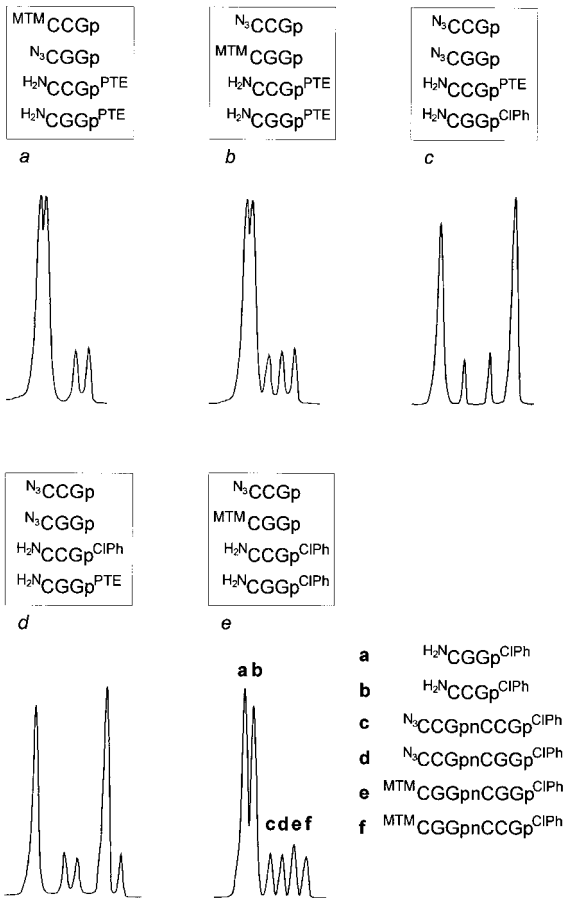


Figure 7. Optimization of baseline separation. a–d) Formal replacement of one single trimer of the old set, 5′-azido- and 3′-phenylthioethyl-protected, by the respective trideoxynucleotide of the new set, 5′-MTM- and 3′-chlorophenyl-protected. e) Separation of remaining hexamer peak (see b) by use of less lipophilic 3′-*o*-chlorophenyl-protected aminotrideoxynucleotides instead of 3′-phenylthioethyl-protected compounds. For assignment of each of the four peaks to a given hexamer see text.

Table 7. Calibration data [factors a and b, Eq. (2)] to calculate concentrations from the respective HPLC peak areas.

Set	Compound	ε [L mol^{-1} cm^{-1}]	a [10^{-5}]	b [10^{-5}]
I, II	Ap	26761	1.67	1.58
I, II	Bp	33899	1.41	1.31
I	nA	34865	1.39	1.29
I	nB	42003	1.22	1.11
II	nA	26961	1.66	1.57
II	nB	34099	1.40	1.30
I, II	ApB	60660	1.14	1.94
I, II	BpA	60660	1.14	1.94
I, II	ApA	53522	1.22	2.11
I, II	BpB	67798	1.05	1.77
I	ApnB	68764	1.04	1.75
I	BpnA	68764	1.04	1.75
I	ApnA	61626	1.12	1.91
I	BpnB	75902	0.95	1.60
II	ApnB	60860	1.14	1.94
II	BpnA	60860	1.14	1.94
II	ApnA	53722	1.22	2.11
II	BpnB	67998	1.05	1.77

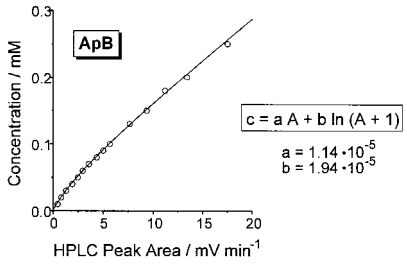


Figure 8. Calibration curve to calculate concentrations of hexadeoxynucleotide template CCGCGG (ApB) from the respective HPLC peak areas. The data were derived from HPLC analysis of samples of known concentration. Because of decreasing recovery rates at low concentrations, the calibration curves diverge significantly from a linear relationship between concentration and peak areas. Empirically, we found that Equation (13) gives a very good approximation of the calibration measurements.

(Figure 7e). The assignment of each of the four peaks to a given hexamer was based on two considerations. Firstly, dual channel monitoring of the UV absorbances at $\lambda = 254$ nm and 273 nm gave UV-absorbance ratios that depended heavily on the base composition. Thus, non-self-complementary hexamers CCGCCG and CGGCGG could be assigned nonarbitrarily, whereas self-complementary sequences CCGCGG and CGGCCG, bearing the same base composition, were indistinguishable. Secondly, an assignment of the latter sequences was possible based on the lipophilicity of the terminal protective groups. Thus, the earlier eluting self-complementary hexamer was assigned to the 5′-azido group whereas the later eluting hexamer was assigned to the 5′-MTM group. Only the latter assignment was consistant with the kinetic data.

Calibration of HPLC peak areas: Concentrations of oligodeoxynucleotides were calculated from the respective HPLC peak areas by means of calibration data for each compound (Table 7). These data were derived from HPLC analysis of samples of known concentration. Decreasing recovery rates at low concentrations meant that the calibration curves diverged significantly from a linear relationship between concentration and peak areas (Figure 8). Empirically, we found that Equation (13) approximates the relationship between concentrations c and peak areas A of the calibration measurements. From Equation (13), appropriate calibration factors a and b could be derived by nonlinear regression using least-squares techniques.

$$c = aA + b\ln(A+1) \qquad (13)$$

Data analysis using *SimFit*: The computer program *SimFit* (nonlinear fitting by dynamic simulation)[11] was employed to derive the rate constants together with their errors and covariances as shown in Tables 3, 4 and 6. Generally, the program supports the kinetic evaluation of experimental time courses of concentrations based on any given reaction model. As many as 16 different experiments with varying initial concentrations can be evaluated simultaneously in a single run. Input of reaction equations is supported at run-time (i.e., no compiling and linking prior to program start). The program then generates the corresponding set of differential equations and its concentration derivatives (Jacobi matrix). The user may choose either the Runge–Kutta algorithm or, for the case of stiff differential equations, a modified Gear algorithm as the method of integration. Nonlinear curve fitting by least squares may be achieved either with the Simplex algorithm or with the Newton–Raphson algorithm. The output data consist of the approximated concentration time profiles, the determined rate constants and associated variances and covariances. The program supports the input of symbolic assignments between the concentrations of mechanistic species and the observable concentrations (expressions with sums and differences). Thus, it is applicable also in those kinetic studies where only the total concentrations can be monitored (e.g. in the case of intermediate complexes which form and dissociate rapidly).

The course of data evaluation using *Simfit* is illustrated as follows. Firstly, the experimental input data were organized as tables of HPLC integrals observed at the given reaction times and for the given experiments. These tables were written as text files (*.txt) consisting of a header with the names

of observable compounds and the data itself. A typical example showing the original data for system 4, group 6 (Figure 3) is listed in Table 8. Secondly, so-called command files (*.cmd) were prepared, which consisted

Table 8. Sample input data comprising 4 experiments. Note that for time >0, an entry of zero means no consideration of that variable.

No.	Exp No.	Time	Bp	nB	ApA	BpnB
1	1	0	68.0	78.1	0	0
1	2	0	69.4	79.7	3.85	0
1	3	0	70.2	80.6	6.50	0
1	4	0	70.4	80.8	9.25	0
2	1	5	0	0	0	0.25
2	2	5	0	0	0	1.90
2	3	5	0	0	0	2.90
2	4	5	0	0	0	4.00
3	1	10	0	0	0	0.35
3	2	10	0	0	0	3.65
3	3	10	0	0	0	6.00
3	4	10	0	0	0	7.85
4	1	15	0	0	0	0.55
4	2	15	0	0	0	5.05
4	3	15	0	0	0	8.10
4	4	15	0	0	0	9.40
5	1	20	0	0	0	0.60
5	2	20	0	0	0	5.85
5	3	20	0	0	0	8.40
5	4	20	0	0	0	10.8
6	1	25	0	0	0	0.90
6	2	25	0	0	0	6.90
6	3	25	0	0	0	9.95
6	4	25	0	0	0	12.9
7	1	30	0	0	0	1.00
7	2	30	0	0	0	7.70
7	3	30	0	0	0	10.9
7	4	30	0	0	0	13.9
8	1	45	0	0	0	1.65
8	2	45	0	0	0	9.65
8	3	45	0	0	0	13.8
8	4	45	0	0	0	17.0
9	1	60	0	0	0	2.80
9	2	60	0	0	0	11.3
9	3	60	0	0	0	15.4
9	4	60	0	0	0	19.2
10	1	90	0	0	0	4.15
10	2	90	0	0	0	14.1
10	3	90	0	0	0	18.3
10	4	90	0	0	0	22.3

of instructions to the program on how to analyze the input data. The command file for the evaluation of the above input data is listed in Table 9. The number of rate constants to be iterated was entered first (DIM). Next, the names of compounds for which observable HPLC data had been

Table 9. Sample command file to evaluate the data.

```
* Specify number of variable rate constants:
DIM (3)
* Define and scale observable data to be read
* from input data file S4F.TXT;
* format of commands:
* DEFINE (running number,
* name of observable as in headline,
* type (e, p, i or c for educt, product,
* intermediate or catalyst, respective
* ly), screen color during plot)
* SCALE (number of scale function,
* calibration data):
DEFINE (1,Bp,e,0) SCALE (2, 1.41e5, 1.31e5)
DEFINE (2,nB,e,0) SCALE (2, 1.22e5, 1.11e5)
DEFINE (3,ApA,c,0) SCALE (2, 1.22e5, 1.11e5)
DEFINE (4,BpnB,p,12) SCALE (2, 0.95e5, 1.60e5)
* Read the above input data file,
* filename S4F.TXT:
READ (S4F)
* Select observable/s for screen plots:
SELECT (BpnB)
* Define unit of time for screen plots:
TIME (min)
* Define window in concentration-time space;
* format of command:
* WIN (min. time, max. time,
* time-axis interval,lenght of conc. tic,
* min. conc., max. conc.,
* conc.-axis interval, lenght of time tic.)
WIN (0, 90, 15, 1, 0, 5e4, 1e4, 1e6)
* Input of reaction equations and input of
* estimates for the rate constants of the above
* reactions. Rate constants may be variable,
* fixed or coupled. Format of commands:
* REACTION (expression)
* CONSTANT (number of reaction equation,
* numerical value of rate constant,
* number of variable
* (zero means fixed),
* coupling factor, minmax-factor):
REACTION (Bp + nB > BpnB)
CONSTANT (1, 1.75e2, 1, 1, 50)
REACTION (Bp + nB + ApnA > C)
CONSTANT (2, 2.50e + 2, 2, 1, 1000)
REACTION (ApnA + BpnB = = > C)
CONSTANT (3, 1.00e6, 0)
CONSTANT (4, 10.0, 3, 1, 1000)
* Encode rate equations:
REACTION (compile)
* Print rate equations and Jacobian matrix onto
* the screen:
REACTION (show)
* Wait for a keypress:
KEY (> > > > Press any key to continue < < < < )
* Input of assignments
* ASSIGN (obs, X = Y) means:
* the theoretical concentration of the observa-
* ble X is taken from the species Y. Symbolic
* expressions as ASSIGN (obs, X = A + B) are
* possible.
* ASSIGN (spec, X = Y) means:
* the initial concentration of species X is
* taken from the observable Y:
ASSIGN (obs, Bp = Bp)
ASSIGN (obs, nB = nB)
ASSIGN (obs, BpnB = C + BpnB, prod)
ASSIGN (spec, Bp = Bp)
ASSIGN (spec, nB = nB)
ASSIGN (spec, ApnA = ApA)
ASSIGN (spec, BpnB = BpnB)
* Wait for a keypress:
KEY (> > > > Press any key to continue < < < < )
* Clear the screen:
CLS
* Select experiment/s in the input data set:
CHOOSE (expall)
* Plot observable points and theoretical curves
* onto the screen:
PLOT
GRAPH
* Open the document file S4F.DOC:
DOCUMENT (S4F)
* Write reaction and rate equations as well as
* rate constants into the documentfile S4F.DOC:
REACTION (doc)
CONSTANT (doc)
* Input of integration and regression parame-
```


0947-6539/98/0404-0640 $ 17.50+.50/0 *Chem. Eur. J.* **1998**, *4*, No. 4

```
* ters. The system of coupled differential
* equations is stiff and needs the respective
* integration algorithm. The initial integra-
* tion step size * is 10^-7 s. A minimum of 16
* steps between any adjacent observation times
* is chosen. Any species concentration is not
* allowed to change for more than 1% if the
* step size is doubled. Plots of theoretical
* curves have 50 steps. The maximal number of
* iteration cycles is 10:
INTEG (stiff, 1e7, 16, 0.01, 50, 10)
* Start the optimization using the Simplex
* algorithm (first 10 cycles) followed by the
* Newton - Raphson algorithm. The results are
* written into the document file S4F.DOC:
SIMPLEX (doc)
NEWTON (doc)
CONSTANT (doc)
* Write experimental points and theoretical
* curves into output-data files S4FP.DAT and
* S4FC.DAT:
PLOT (file)
GRAPH (file)
```

collected were defined (DEFINE). Calibration factors for each of the defined compounds were entered in order to derive the corresponding concentrations from the HPLC peak areas (SCALE). The input data themselves were then read (READ) and selected for plotting purposes onto the screen (SELECT). The latter also required definition of an appropriate window in concentration–time space (WIN). Then the reaction equations for type-II fitting were entered (REACTION) and estimates for the rate constants that varied in the course of the optimization were provided (CONSTANT). The observable concentrations derived from HPLC integrals had to be assigned to concentrations of species as given by the reaction equations (ASSIGN). For example, the observable concentration named with BpnB consisted of the concentration of single-stranded species BpnB plus the concentration of the complex C (Scheme 1). Furthermore, assignments were necessary to specify the initial concentration of each species in terms of the respective observable initial concentrations. Ten rounds (defined by INTEG) of simplex optimization (SIMPLEX) followed by ten rounds of Newton–Raphson optimization (NEWTON) proved to be sufficient to find the minimum of the sum of squares with respect to the rate constants. The calculation in the given example took 6.15 minutes on a 486 processor running at 66 Mhz and 2.45 minutes on a pentium processor running at 90 MHz. Finally, the experimental points and the theoretical time courses were exported to other programs (PLOT, GRAPH).

Acknowledgments: This work was supported by Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, NATO, and GIF.

Received: August 1, 1996
Publication delayed at authors' request [F428]

[1] Reviews: a) B. G. Bag, G. von Kiedrowski, *Pure Appl. Chem.* **1996**, *68*, 2145–2152; b) G. von Kiedrowski, J. Helbing, B. Wlotzka, S. Jordan, M. Matzen, T. Achilles, D. Sievers, A. Terfort, B. C. Kahrs, *Nachr. Chem. Tech. Lab.* **1992**, *40*, 578–588; c) D. Sievers, T. Achilles, J. Burmeister, S. Jordan, A. Terfort, G. von Kiedrowski in *Self-Production of Supramolecular Structures—From Synthetic Structures to Models of Minimal Living Systems* (Eds: G. Fleischaker, S. Colonna, P. L. Luisi), Kluwer, Dordrecht, **1994**, pp. 45–64; d) J. Rebek, *ibid.* 75–88; e) E. A. Wintner, M. M. Conn, J. Rebek, *Acc. Chem. Res.* **1994**, *27*, 198–203; f) L. E. Orgel, *ibid.* **1995**, *28*, 109–118; g) S. Hoffmann, *Angew. Chem.* **1992**, *104*, 1032–1035; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1013–1016; h) A. Eschenmoser, E. Loewenthal, Chem. Soc. Rev. **1992**, *21*, 1–16; i) R. Hoss, F. Vögtle, *Angew. Chem.* **1994**, *106*, 389–398; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 375–384; j) A. Kanavarioti, *J. Theor. Biol.* **1992**, *158*, 207–219.

[2] a) G. von Kiedrowski, *Angew. Chem.* **1986**, *98*, 932–934; *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 932–935; b) W. S. Zielinski, L. E. Orgel, *Nature* **1987**, *327*, 346–347; c) G. von Kiedrowski, B. Wlotzka, J. Helbing, *Angew. Chem.* **1989**, *101*, 1259–1261; *Angew. Chem. Int. Ed. Engl.* **1989**, *28*, 1235–1237; d) T. Tjivikua, P. Ballester, J. Rebek, *J. Am. Chem. Soc.* **1990**, *112*, 1249–1250; e) G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen, S. Jordan, *Angew. Chem.* **1991**, *103*, 456–459 and 1066; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 423–426 and 892; f) J. S. Nowick, Q. Feng, T. Tjivikua, P. Ballester, J. Rebek, *J. Am. Chem. Soc.* **1991**, *113*, 8831–8839; g) A. Terfort, G. von Kiedrowski, *Angew. Chem.* **1992**, *104*, 626–628; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 654–656; h) J.-I. Hong, Q. Feng, V. Rotello, J. Rebek, *Science* **1992**, *255*, 848–850; i) Q. Feng, T. K. Park, J. Rebek, *Science* **1992**, *256*, 1179–1180; j) C. Böhler, W. Bannwarth, P. L. Luisi, *Helv. Chim. Acta* **1993**, *76*, 2313–2320; k) T. Achilles, G. von Kiedrowski, *Angew. Chem.* **1993**, *105*, 1225–1228; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1198–1201; l) F. M. Menger, A. V. Eliseev, N. A. Khanjin, *J. Am. Chem. Soc.* **1994**, *116*, 3613–3614; m) M. M. Conn, E. A. Wintner, J. Rebek, *J. Am. Chem. Soc.* **1994**, *116*, 8823–8824; n) F. M. Menger, A. V. Eliseev, N. A. Khanjin, M. J. Sherrod, *ibid.* **1995**, 2870–2878; o) D. N. Reinhoudt, D. M. Rudkevich, F. de Jong, *ibid.* **1996**, *118*, 6880–6889; p) D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadiri, *Nature* **1996**, *382*, 525–527; q) J. Burmeister, G. von Kiedrowski, A. Ellington, *Angew. Chem.* **1997**, *109*, 1321–1381; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1321–1324.

[3] a) D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221–224;see also b) J. Ferris, *ibid* **1994**, *369*, 184–185; c) J. Zubay, *Chemtracts: Org. Chem.* **1995**, 368–370.

[4] For another scheme see: T. Li, K. C. Nicolaou, *Nature* **1994**, *369*, 218–221.

[5] a) L. E. Orgel, R. Lohrmann, *Acc. Chem. Res.* **1974**, *7*, 368–379; b) T. Inoue, L. E. Orgel, *Science* **1983**, *219*, 859–862; c) G. F. Joyce, L. E. Orgel, *J. Mol. Biol.* **1986**, *188*, 433–441.

[6] B. Wlotzka, thesis, Universität Göttingen, **1992**.

[7] I. N. Merenkova, T. S. Oretskaya, N. I. Sokolova, Z. A. Shabarova, *Bioorg. Khim.* **1993**, *19*, 1205–1214.

[8] a) R. J. Pieters, I. Huc, J. Rebek, Jr., *Angew. Chem.* **1994**, *106*, 1667–1669; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1579–1581; b) R. J. Pieters, I. Huc, J. Rebek Jr., *Tetrahedron* **1995**, *51*, 485–498.

[9] D. Sievers, *Selbstreplikationsexperimente mit Oligodeoxynucleotiden: Autokatalyse versus Kreuzkatalyse*, Cuviller, Göttingen, **1994**.

[10] a) S. Jordan, thesis, Universität Göttingen, **1993**; b) J. Helbing, thesis, Universität Göttingen, **1990**.

[11] Available from the author (GvK) on request.

[12] G. von Kiedrowski, *Bioorg. Chem. Front.* **1993**, *3*, 113–146.

[13] a) N. G. Dolinnaya, N. I. Sokolova, O. I. Gryaznova, Z. A. Shabarova, *Nucleic Acids Res.* **1988**, *16*, 3720–3728; b) N. G. Dolinnaya, A. V. Tsytovich, V. N. Sergeev, T. S. Oretskaya, Z. A. Shabarova, *ibid.* **1991**, *19*, 3073–3080.

[14] a) R. Breslauer, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3746–3750; b) C. R. Cantor, P. R. Schimmel, *Biophysical Chemistry, Part III: The behavior of biological macromolecules*, Freeman, New York, **1980**, pp. 1197–1198.

[15] a) S. Pitsch, R. Krishnamurthy, M. Bolli, S. Wendeborn, A. Holzner, M. Minton, C. Lesueur, I. Schloenvogt, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1995**, *78*, 1621–1635; b) M. Bolli, R. Micura, A. Eschenmoser, *Chem. Biol.* **1997**, *4*, 309–320.

[16] J. Maynard-Smith, E. Szathmáry, *The Major Transitions in Evolution*, Freeman, Oxford, **1995**.

[17] E. Szathmáry, *Trends Ecol. Evol.* **1991**, *6*, 366–370.

[18] S. A. Kauffman: *The Origins of Order*, Oxford University Press, New York, **1993**.

[19] J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59–61; see also G. von Kiedrowski, *ibid.* 20–21.