Pyranosyl-RNA: chiroselective self-assembly of base sequences by ligative oligomerization of tetranucleotide-2′,3′-cyclophosphates (with a commentary concerning the origin of biomolecular homochirality)

Martin Bolli[1,2], Ronald Micura[1] and Albert Eschenmoser[1,2]

Background: Why did Nature choose furanosyl-RNA and not pyranosyl-RNA as her molecular genetic system? An experimental approach to this problem is the systematic comparison of the two isomeric oligonucleotide systems with respect to the chemical properties that are fundamental to the biological role of RNA, such as base pairing and nonenzymic replication. Pyranosyl-RNA has been found to be not only a stronger, but also a more selective pairing system than natural RNA; both form hairpin structures with comparable ease. Base sequences of pyranosyl-RNA can be copied by template-controlled replicative ligation of short activated oligomers (e.g. tetramer-2′,3′-cyclophosphates) under mild and potentially natural conditions. The copying proceeds with high regioselectivity as well as chiroselectivity: homochiral template sequences mediate the formation of the correct (4′→2′)-phosphodiester junction between homochiral tetramer units provided they have the same sense of chirality as the template. How could homochiral template sequences assemble themselves in the first place?

Results: Higher oligomers of pyranosyl-RNA can self-assemble in dilute solutions under mild conditions by ligative oligomerization of tetramer-2′,3′-cyclophosphates containing hemi self-complementary base sequences. The only side reaction that effectively competes with ligation is hydrolytic deactivation of 2′,3′-cyclophosphate end groups. The ligation reaction is highly chiroselective; it is slower by at least two orders of magnitude when one of the (D)-ribopyranosyl units of a homochiral (D)-tetramer-2′,3′-cyclophosphate is replaced by a corresponding (L)-unit, except when the (L)-unit is at the 4′ end of the tetramer and carries a purine, when the oligomerization rate can be ~10% of that shown for a homochiral isomer. The oligomerization of homochiral tetramers is not, or only weakly, inhibited by the presence of the non-oligomerizing diastereomers.

Conclusions: Available data on the chiroselective self-directed oligomerization of tetramer-2′,3′-cyclophosphates allow us to extrapolate that sets of tetramers with different but mutually fitting base sequences can be expected to co-oligomerize stochastically and generate sequence libraries consisting of predominantly homochiral (D)- and (L)-oligomers, starting from the racemic mixture of tetramers containing all possible diastereomers. Such a capability of an oligonucleotide system deserves special attention in the context of the problem of the origin of biomolecular homochirality: breaking molecular mirror symmetry by de-racemization is an intrinsic property of such a system whenever the constitutional complexity of the products of co-oligomerization exceeds a critical level.

Addresses: [1]Laboratory of Organic Chemistry, Swiss Federal Institute of Technology, Universitätstrasse 16, CH-8092 Zürich, Switzerland and [2]The Skaggs Institute for Chemical Biology at the Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA-92037, USA.

Correspondence: Albert Eschenmoser

Key words: biomolecular homochirality, chiroselectivity, oligonucleotides, pyranosyl-RNA, self-assembly

Received: **12 March 1997**
Accepted: **3 April 1997**

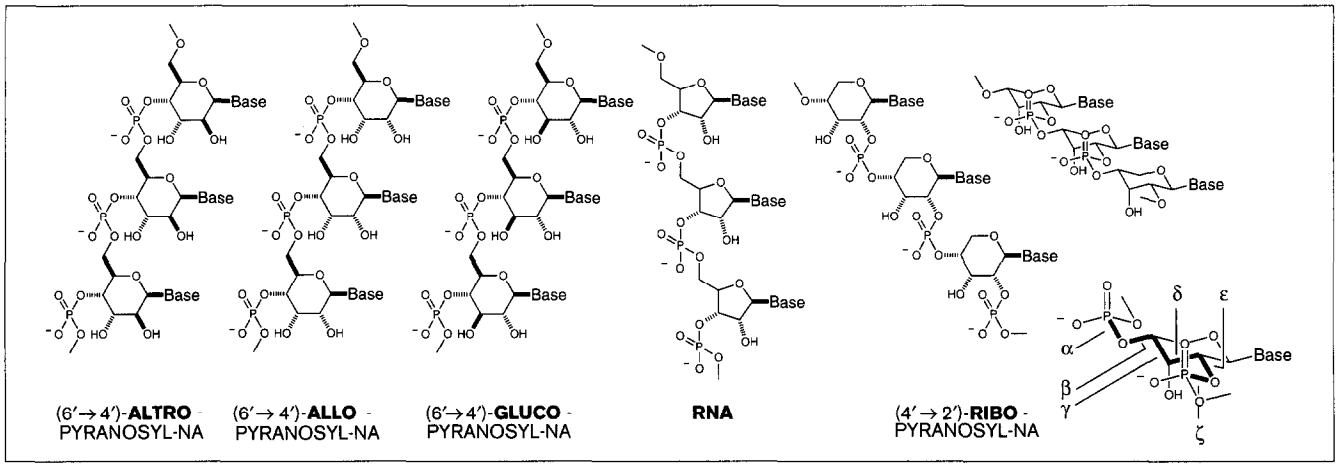
Chemistry & Biology April 1997, **4**:309–320
http://biomednet.com/elecref/1074552100400309

Introduction

Pyranosyl-RNA (p-RNA; [1–6]; M.B., R.M., S. Pitsch and A.E., unpublished results) is the constitutional isomer of RNA with ribose units in their pyranose (instead of furanose) form and with phosphodiester bridges spanning the ribose units between the position 4′ and 2′ instead of 5′ and 3′ (Fig. 1). Its chemical properties are being systematically studied as part of a project which is oriented towards a chemical etiology of nucleic acid structure [7]. According to the chemical criteria which direct this project, p-RNA is a potentially natural nucleic acid alternative which Nature could have selected, but either had not or had only transiently, as her molecular genetic system.

Potential nucleic acid alternatives derived from the close constitutional neighborhood of RNA would be expected

Figure 1

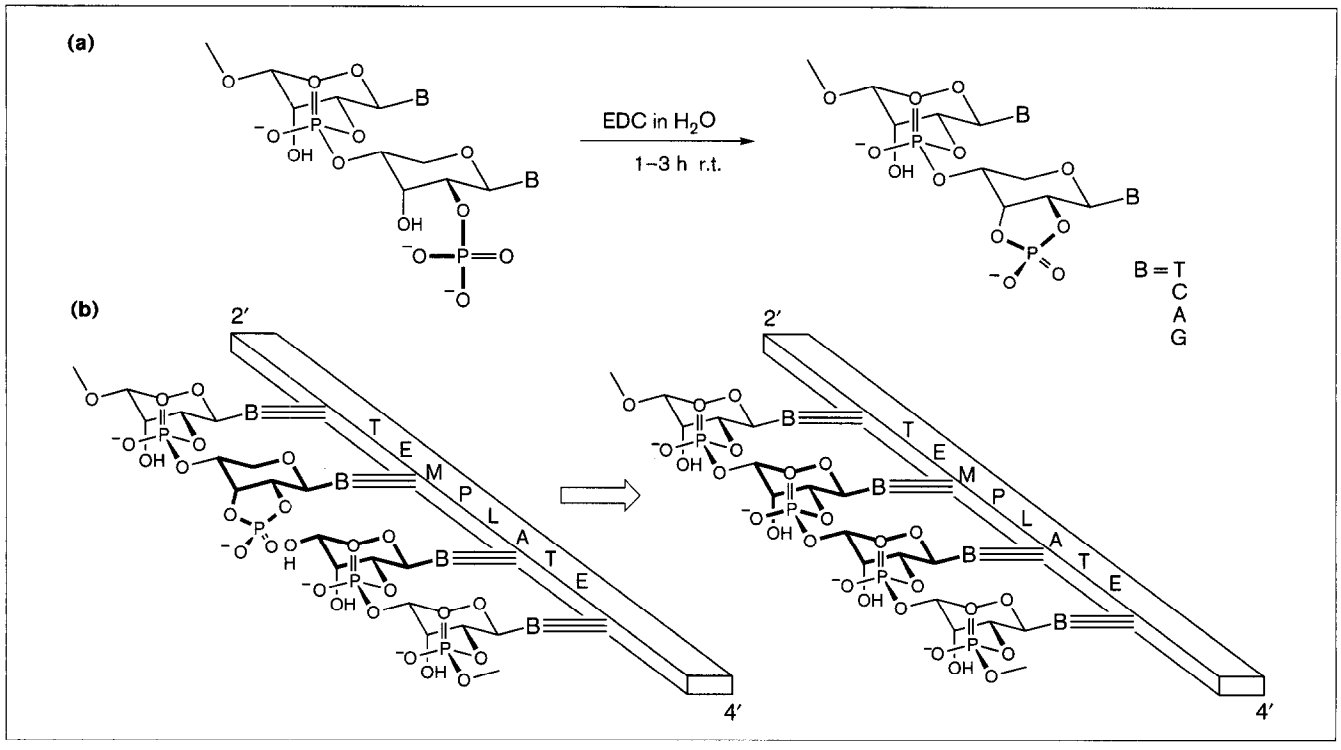


Hexopyranosyl- and pentopyranosyl-oligonucleotide systems studied so far in the context of a chemical etiology of nucleic acid structure. Upper right: pairing conformation of p-RNA strand (idealized) [2,5].

to be inferior to RNA with respect to chemical properties which, in the case of RNA, are fundamental to biological function. To observe such inferiority on a chemical level would constitute support for a hypothesis that Nature's evolutionary choice of RNA was made from a diversity of constitutionally related alternatives on the basis of functional criteria. Functional inferiority on a chemical level was, in fact, found for hexopyranosyl-(6′→4′) systems synthesized and systematically studied in our laboratory [1,7–11]. Representative (D)-β-allo-, altro- and glucopyranosyl-(6′→4′)-oligonucleotides (Fig. 1) turned out either to not be pairing systems at all (glucopyranosyl), or to show a base-pairing behavior that was far inferior in strength, regularity and selectivity to that of corresponding DNA (or RNA) oligonucleotides. The reason for this is intrastrand steric hindrance in the pairing conformation of these hexopyranosyl oligonucleotide strands (c.f. the pairing properties of 'homo-DNA', the model system that lacks the two hydroxyl groups at positions 2′ and 3′ [12,13]). Studies pursued more recently on pyranosyl-RNA, the pentopyranosyl isomer of RNA (Fig. 1), revealed chemical properties that are in sharp contrast to those observed for the hexopyranosyl systems described above. Not only is p-RNA a stronger pairing system than RNA (and DNA), its pairing is also more selective in the sense that exclusively Watson–Crick purine–pyrimidine pairing in strictly antiparallel strand orientation and neither Hoogsteen, nor reverse-Hoogsteen purine–purine self-pairing is observed [2,3]. As a consequence of the specific structural features responsible for this unique selectivity, p-RNA base sequences can be reliably copied under a variety of conditions by template-controlled replicative ligation of short oligonucleotide units ([3]; M.B., R.M., S. Pitsch and A.E., unpublished results). The phosphate activation required for the formation of phosphodiester bonds can take the simplest possible path, namely that of an intramolecular and, therefore, easily achieved dehydration converting 2′-phosphates into 2′,3′-cyclophosphates (Fig. 2). Intermolecular 2′,3′-cyclophosphate→4′,2′-phosphodiester transesterification proceeds smoothly under the conditions used in the presence of a template, but not in its absence, and produces exclusively (4′→2′)-phosphodiester junctions (Fig. 2; [3]; M.B., R.M., S. Pitsch and A.E., unpublished results). Sequence copying is regioselective, as well as chiroselective and transcribes guanine-rich sequences with comparable efficiency to complementary cytosine-rich sequences ([3]; M.B., R.M., S. Pitsch and A.E., unpublished results). Although replication conditions under which product inhibition is overcome remain to be demonstrated and template-turnover is achieved without an experimentalist's interference, p-RNA can be stated to represent an oligonucleotide system for which a capacity for informational replication [14] under nonenzymic, robust and potentially natural conditions has been demonstrated.

One of the prerequisites for sequence-copying in p-RNA by replicative ligation is the homochirality of both template and ligand oligonucleotides involved. Where could the homochirality of the template sequences of an oligonucleotide system be derived from? Here we report that ligative oligomerization of tetramer-2′,3′-cyclophosphates proceeding within growing duplexes via mutual overlap of overhanging base sequences can result in constitutional self-assembly of predominantly homochiral p-RNA sequences by a remarkably efficient chiroselection from mixtures of heterochiral tetramer-2′,3′-cyclophosphate diastereomers.

Figure 2



Template-controlled ligation of short p-RNA oligonucleotide 2′,3′-cyclophosphates. **(a)** Activation of phosphate groups at the 2′ end of p-RNA tetramers by conversion to 2′,3′-cyclophosphates (EDC = 1-ethyl-3(3′-dimethylaminopropyl) carbodiimide; [3]; M.B., R.M., S. Pitsch and A.E., unpublished results); **(b)** template-directed ligation of p-RNA oligomer-2′3′-cyclophosphates leads regioselectively to (2′→4′)-phosphodiester junctions by an in-line reaction path favored by the constellation of reaction centers which is caused by the strong inclination of backbone and base-pair axes in the p-RNA pairing conformation (see [3,5]; M.B., R.M., S. Pitsch and A.E., unpublished results). When not stated otherwise, p-RNA sequences mentioned in this paper are homochiral and belong to the (D)-ribose series.

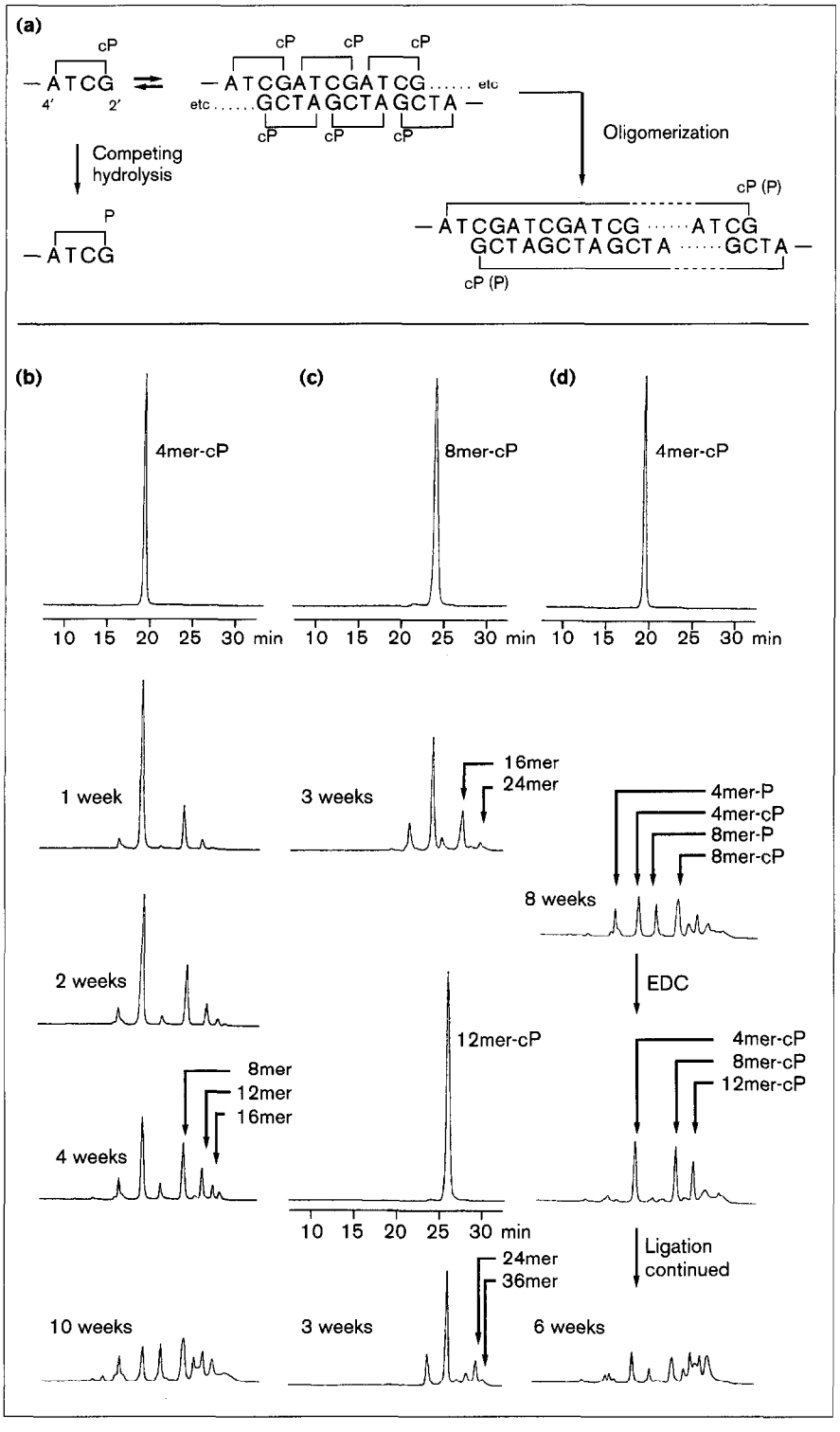
Results and discussion

Self-directing oligomerization of homochiral p-RNA tetramers

Under reaction conditions in which the rate ratio of ligation to hydrolytic deactivation had been found to be in favor of the former (2.5 mM tetramer 2′,3′-cyclophosphate (cP) in aqueous 1.5 M LiCl, pH 8.5, 4°C), the tetramer pr(ATCG)-cP (pr = pyranosyl-ribo) smoothly oligomerizes within weeks to produce a mixture of the cPs of higher oligomers accompanied by their respective (2′,3′-cP→2′ and 3′-phosphate (P)) hydrolysis products (Fig. 3a). Progress of reaction was followed by high performance liquid chromatography (HPLC), under conditions (pH 11.5) where the strands of the product duplexes run as single strands (Fig. 3b). Products were identified on the basis of HPLC analysis of the octamer- and dodecamer-cPs by coinjection with authentic samples, by comparison of the HPLC traces of the product mixture with those obtained in analogous experiments starting from the (authentic) octamer pr(ATCGATCG)-cP or dodecamer pr(ATCGATCGATCG)-cP (Fig. 3c) and, finally, by comparison of the HPLC traces obtained before and after treatment of the oligomerization product mixtures with water-soluble carbodiimide under conditions which convert product 2′-phosphates into the corresponding cPs (Fig. 3d). The partial hydrolysis of the reactands' cyclophosphate end groups to the corresponding 2′ and 3′ phosphate groups is, in essence, the only process that efficiently competes with ligative oligomerization; it restricts both oligomerization yields and oligomer chain lengths. The distribution of chain length can be pushed to higher values by treatment of reaction mixtures with water-soluble carbodiimide, causing reactivation of the 2′- and 3′-phosphate end groups by converting them back to cP groups (Fig. 3d).

Auto-oligomerization of p-RNA tetramer-cPs is, in principle, possible with any tetramer of the type PU–py–py–PU, py–PU–PU–py, PU–py–PU–py or py–PU–py–PU in which the base sequence shows dual hemi-complementarity in antiparallel strand orientation. Representative examples for each sequence motive were tested experimentally under standard conditions; they are listed in Table 1 (Nos 1–10) which shows the outcome in terms of the yields of higher oligomers, qualitatively reflecting the dependence of the ratio between cyclophosphate ligation and cyclophosphate hydrolysis on the tetramer's base sequence.

Figure 3



(a) Reaction scheme for the auto-oligomerization experiment starting with pr(ATCG)-2'3'-cyclophosphate [pr(ATCG)-cP]. **(b)** HPLC of product mixtures of experiment (a) after reaction times of up to 10 weeks. For reaction conditions, see the caption for Table 1. HPLC: Mono-Q-HR 5/5 Pharmacia, denaturing conditions (10 mM Na-phosphate buffer, pH 11.5), linear NaCl gradient (0→1.0 M within 30 min), detection 270 nm. **(c)** HPLC of product mixtures of oligomerization experiments starting with pr(ATCGATCG)-cP (above) and pr(ATCGATCGATCG)-cP (below), conditions as in experiment (a). **(d)** HPLC of product mixtures of the auto-oligomerization experiment starting with pr(ATCG)-cP before (above) and after (middle) treatment with EDC (excess EDC, in H_2O, r.t., 2.5 h). Below: HPLC trace of reaction mixture treated with EDC and then exposed to ligation conditions again for 6 weeks. (Synthesis of cPs of p-RNA sequences is given in [2,3] and M.B., R.M., S. Pitsch and A.E., unpublished results.)

Estimated rate differences clearly point to a sequence dependence of ligation that parallels a sequence dependence of the thermodynamic stability of the pre-ligation complex, superimposed with a kinetic sequence dependence of interstrand base-stacking at the ligation site (Fig. 4). Two pyrimidines in interstrand-stacking relationship at these sites result in slower ligation relative to the rate observed when these stackings are purine–purine or purine–pyrimidine.

Table 1

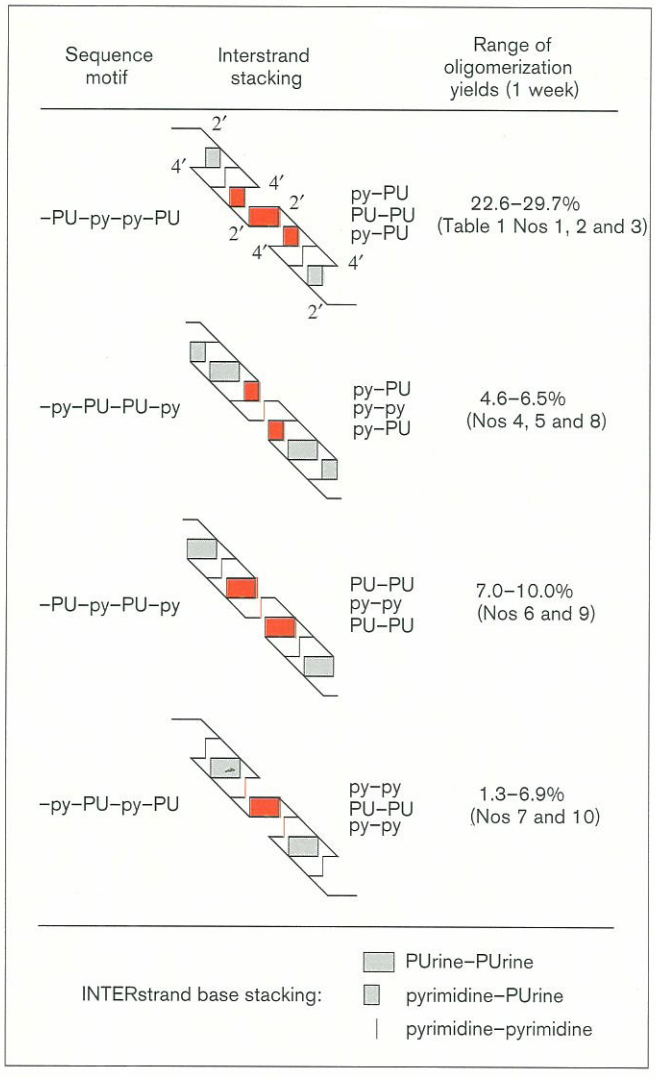
Oligomerization experiments referred to in this paper.

	Tetramer sequence	Yield in % after ... weeks 1	2	8
1	ATCGcP	22.6	38.4	71.2
	ATC-((L)-G)cP	0	0.04	0.2
	AT-((L)-C)-GcP	0.03	0.3	1.1
	A-((L)-T)CGcP	0	-	0.3
	A-((L)-TCG)cP	2.5	3.2	5.8
	AT-((L)-CG)cP	0	0.1	0.2
	A-((L)-TC)-GcP	0	0	0
	A-((L)-T)-C-((L)-G)cP	0	0	0
2	DTCGcP	26.4	45.2	76.2
	((L)-D)TCGcP	3.1	5.2	12.2
3	GCCGcP	29.7	43.2	73.7
	GCC-((L)-G)cP	0	0	0
	GC-((L)-C)-GcP	0	0	0
	G-((L)-C)-CGcP	0	0	0
	((L)-G)-CCGcP	2.8	5.6	12.1
4	CGGCcP	6.5	11.8	30.6
	CGG-((L)-C)cP	0	0	0
	CG-((L)-G)-CcP	0	0	0.13
	C-((L)-G)-GCcP	0	0	0
	((L)-C)-GGCcP	0	0.20	0.60
5	TAGCcP	4.6	10.7	32.0
	((L)-T)-AGCcP	0	0	0
6	ATGCcP	7.0	14.2	41.5
	A-((L)-TGC)cP	0.85	1.7	5.2
7	TACGcP	1.3	2.7	7.0
	TAC-((L)-G)cP	0	0	0
	((L)-T)-ACGcP	0	0	0.53
8	TDGCcP	5.1	10.2	30.5
9	DTGCcP	10.0	20.2	51.8
10	TDCGcP	6.9	14.9	45.6
11	GTCGcP	0	0	0
12	GCTGcP	0	0.2	0.5
13	GCCGcP + DTCGcP	23.0	36.1	70.5
14	ACGGcP + GTCCcP	2.6	5.4	26.1
15	GCGGcP + GCCCcP	11.4	21.1	50.7
	GCGGcP + GCC-((L)-C)cP	0.1	0.3	0.3
	GCGGcP + GC-((L)-C)-CcP	0	0	0.3
	GCGGcP + G-((L)-C)-CCcP	0.6	1.1	5.4
	GCGGcP + ((L)-G)-CCCcP	4.1	7.8	28.0
16	GCGCcP	0.37	0.32	1.1
17	CGCGcP	0	0.2	0.8
18	GCCGcP + CGGCcP 1:1	0.3	0.6	1.3
19	GCCGcP + CGGCcP 2:1	18.8	34.5	69.3
20	GCCGcP + CGGCcP 1:2	2.9	6.5	18.4

Product yields (oligomers higher than tetramers in the form of 2′,3′-cyclophosphates and 2′ (and 3′) phosphates) as determined by HPLC are given. 0 = no product detectable (< 0.03%). All experiments were carried out under the same conditions: 2.5 mM tetramer-cP, in H_2O, 1.5 M LiCl, pH 8.5; 0.1 M HEPES-buffer; 4°C; within weeks (M.B., R.M., S. Pitsch and A.E., unpublished results). D = 2,6-diaminopurine. Sequences are formulated in the 4′→2′ direction.

Auto-oligomerization is highly sensitive to the notorious guanine–uracil mismatch. Substituting one of the cytosine bases in readily oligomerizing pr(GCCG) with thymine to produce pr(GTCG) and pr(GCTG) stops the reaction (Table 1, Nos 11,12; [2,3]; M.B., R.M., S. Pitsch and A.E., unpublished results). The two experiments show, as do others discussed below, that auto-oligomerization of a tetramer-cP does not occur under the conditions used when

Figure 4



Parallel trends in sequence dependence of auto-oligomerization rates and interstrand base-stacking at ligation sites. For the structural aspects of interstrand stacking in p-RNA, see [5].

the tetramer lacks the dual hemi self-complementarity of its base sequence.

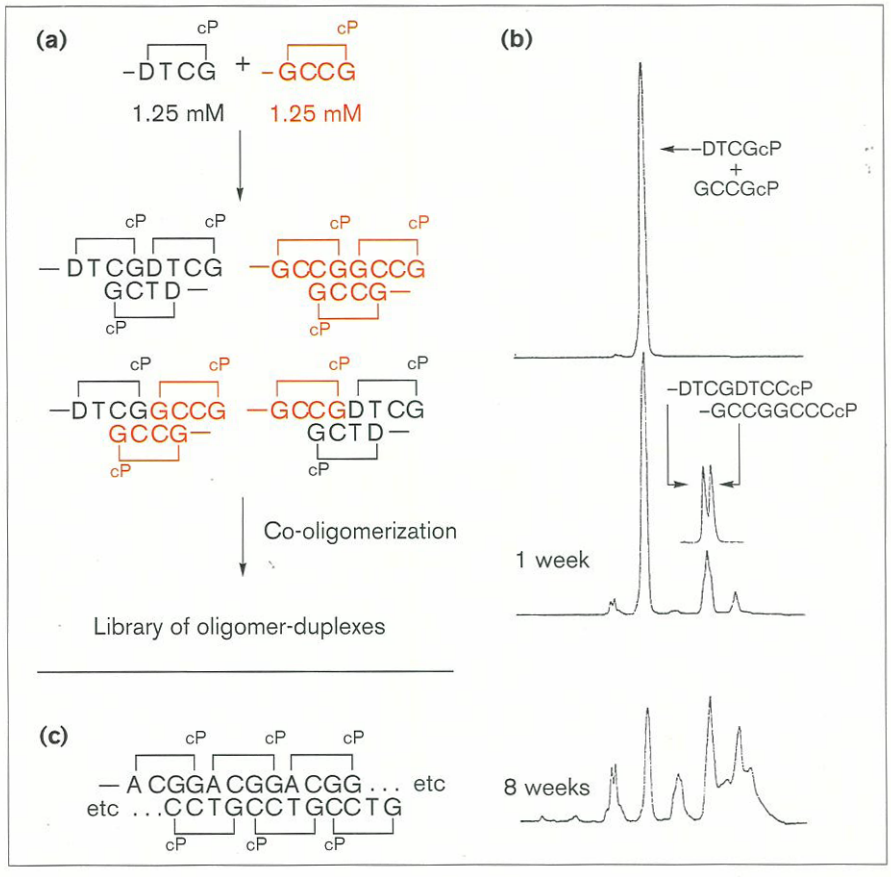
We have exemplified two variants of co-oligomerization in which two different tetramer sequences collaborate with each other in building the strands of a growing duplex. In the first, the two tetramer units can — besides oligomerizing by themselves — stochastically co-oligomerize in both strands, giving rise to a library of oligomer duplexes, which contain the two tetramer sequences in both strands in an essentially random distribution. Equimolar amounts of the sequences pr(DTCG) and pr(GCCG) (Table 1, No. 13), both of which had been observed to auto-oligomerize with comparable speed (Nos 2,3), led to a co-oligomerization-product mixture for which HPLC analysis of the octamer fraction indicated the presence of not two but of what are probably four different octamer-cPs in about equal amounts (Fig. 5a,b). This points to comparable rates of homoligation and heteroligation of the two tetramer sequences and, by extrapolation, to an essentially stochastic formation of sequence libraries, which contain $2^{n/4}$ different sequences of the two tetramer units within sublibraries of oligomers containing n monomer units. Another example of this type of co-oligomerization are the two sequences pr(GCGG)-cP and pr(GCCC)-cP (Table 1, No. 15).

In a second variant of co-oligomerization, the base sequences of two tetramer-cPs are chosen in such a way that a duplex can grow by oligomerization of each tetramer in its own strand assisted by mutual templating. Complementary strands with base sequences consisting of homotetramer repeats will result (Fig. 5c). The reaction starting from a mixture (1:1) of the cPs of pr(ACGG) and pr(GTCC) proceeds smoothly, although more slowly than one might have expected (Table 1, No. 14). HPLC of the product mixture reveals the presence of the expected two different oligomer sets in the ratio of about 2:1, with oligomers of pr(GTCC)-cP as the main components. The two sets are identified by comparison with product HPLC traces of oligomerization experiments in which one of the two tetramers is not a 2′,3′-cyclophosphate, but the 2′-phosphate.

Co-oligomerizations that involve more than two different tetramer sequences have not been tested, mostly because of foreseeable difficulties in product analysis. Any base sequence containing 4n bases could, in principle, be formed by self-directing ligative co-oligomerization, given enough time and a library of requisite tetramer-cPs as starting material. Such a process, however, would require that the necessary two-base-overlap combinations (there are eight in total) were all prone to mediate ligation and, moreover, that those tetramer-cPs which have fully complementary sequences would not block each other under the reaction conditions by mutual pairing and withdraw themselves from participating in co-oligomerization. Although the first of the requirements may well be fulfilled, the second is more critical, and relevant observations are more scarce. The cyclophosphates containing the self-complementary sequences pr(CGCG) and pr(GCGC) auto-oligomerize under standard conditions (4°C) only very slowly (Table 1, Nos 16,17). Interstrand-stacking analysis (c.f. Fig. 4) predicts a self-complementary (4′→2′)-(py–PU–py–PU) sequence to form a more stable tetramer duplex than the isomeric (4′→2′)-(PU–py–PU–py) sequence [6] and the duplex of the latter to tolerate the overlap frame-shift required for auto-oligomerization more readily. Such shifts seem to occur readily in homopurinic and homopyrimidinic sequences as we had observed in ligative replication of the sequence pr(GGGGGGGG) [3]. A mixture (1:1) of pr(GCCG)-cP and pr(CGGC)-cP shows

Figure 5

Co-oligomerization of two different tetramer sequences. **(a)** Example with stochastic mixing of the two sequences in both strands of growing duplex (D=2,6-diaminopurine). **(b)** HPLC of product mixtures of experiment (a). The octamer-cP peak (after 1 week) consists of three fractions in an approximate ratio of 1:2:1, the shoulders coinciding with the peaks of the authentic homo-octamer-cPs (synthesized for comparison). **(c)** Example of a co-oligomerization proceeding without mixing of participating tetramer sequences.



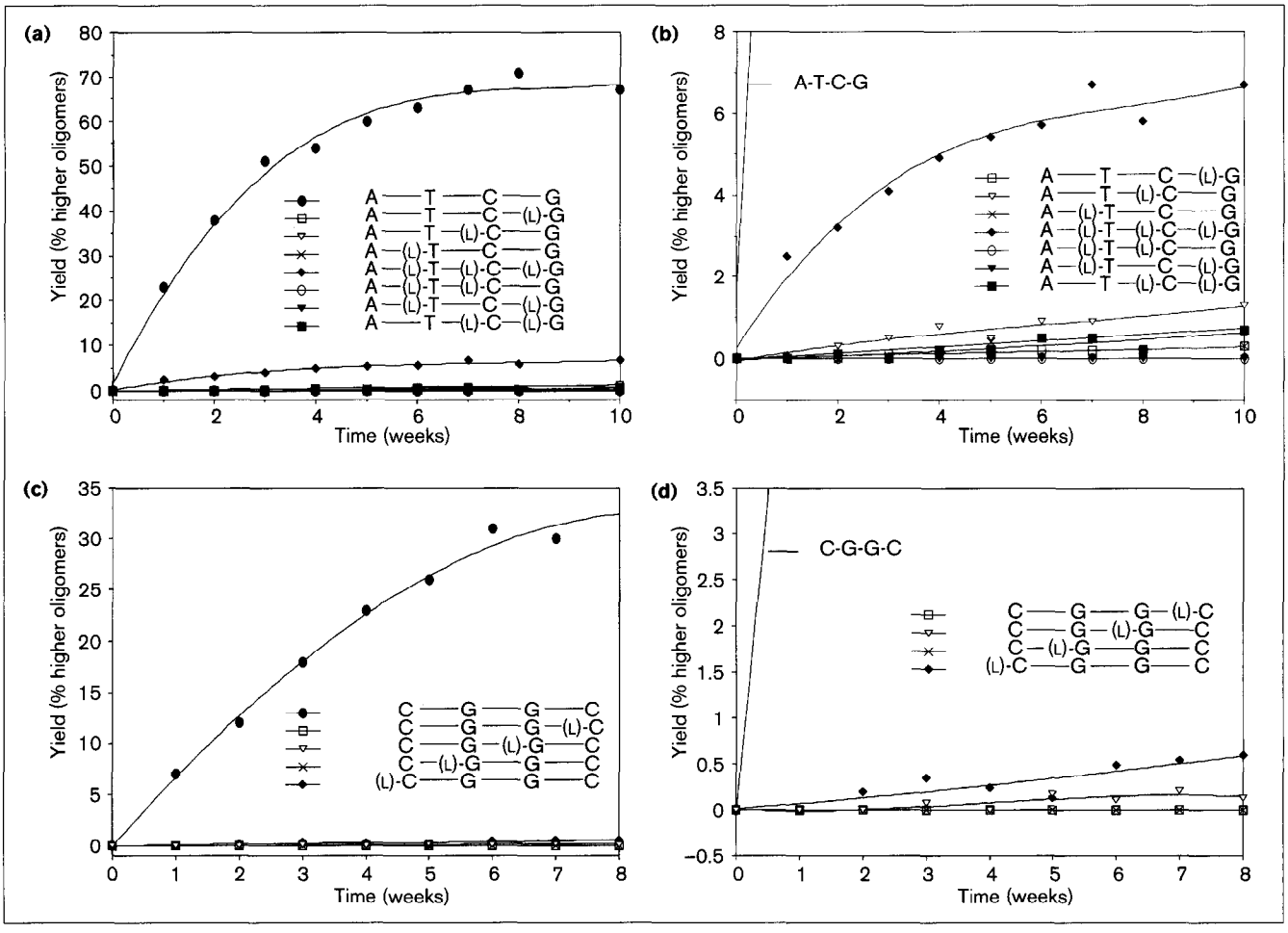
concomitant, but again very slow auto-oligomerization of both sequences; in 2:1 and 1:2 mixtures of the two, on the other hand, the excess component oligomerizes in about the expected rate (Table 1, Nos 19,20).

Chiroselectivity of oligomerizations

Ligative oligomerizations of p-RNA tetramers-cPs are highly chiroselective (Fig. 6). The selectivity is comparable to the one observed in replication of p-RNA base sequences by template-controlled ligation of tetramer-cPs (M.B., R.M., S. Pitsch and A.E., unpublished results). Replacement of the (D)-ribopyranosyl-guanine unit of homochiral (D)-pr(ATCG)-cP by the corresponding (L)-unit reduces the rate of oligomerization (as judged from the absence of the HPLC trace at the octamer-cP position after the first week) by a factor of two to three orders of magnitude (Fig. 6a, Table 1). Analogous mono-replacements involving the cytosine-bearing and thymine-bearing sugar units induce a comparable inhibition. On the other hand, substitution of the adenine-bearing sugar unit at the tetramer's 4′ end is less effective in suppressing oligomerization; a residual oligomer production amounting to about 10% of the one of the homochiral tetramer is observed (Fig. 6b). With the tetramer sequence pr(GCCG)-cP, the same chiroselectivity pattern is found as with pr(ATCG)-cP (Table 1, Fig. 6c). The isomeric tetramer sequence pr(CGGC)-cP shows a chiroselectivity pattern in which again no reaction can be observed after (D→L) substitutions in either one of the two central positions or the one at the 2′ end. Importantly, replacement of the unit at the 4′ end, which now carries a cytosine and not a guanine, leads to a rate which is much less than before, namely about two orders of magnitude less than the rate of the homochiral tetramer (Fig. 6d). Thus, oligomer production from tetramers bearing a chirality mismatch at their 4′ end is much lower when the involved nucleobase is a pyrimidine rather than a purine; this was observed in all relevant examples investigated so far (Table 1, Nos 1–7).

The remarkable difference between purines and pyrimidines at the 4′ end in their effect on ligation may be related to a difference in efficiency by which interstrand base-stacking (in presumably the syn conformation of the 4′-purine) [15] achieves conformational fixation of the 4′ end in the pre-ligation complex. That, on the other hand, (D→L) substitution at the 2′ end, irrespective of whether the base is a purine or a pyrimidine, blocks oligomerization completely, is not necessarily surprising,

Figure 6



Chiroselectivity of auto-oligomerization of the sequences **(a)** and **(b)** pr(ATCG)-cP and **(c)** and **(d)** pr(CGGC)-cP under standard conditions (c.f. Fig. 3a; Table 1, Nos 1,4). Curves (•): observations for homochiral all-(D) tetramers; other curves: observations for chirodiastereomers of indicated configuration; (D→L) substitution of adenine unit is made in the enantio-series. Yield scales in (b) and (d) are enlarged by a factor of 10 to compare the effect of 4′-end substitution with that of other substitutions. Note the difference between (b) purines and (d) pyrimidines in (D→L) substitutions at the 4′ end.

because ligation can be expected to show a special stereoelectronic sensitivity towards a change in the conformational positioning of the cyclophosphate group (note also the differences in interstrand stacking). The sensitivity of ligation towards (D→L) substitution of one of the two central units can be qualitatively understood by analyzing the type of reorganization the pairing conformation of a homochiral p-RNA strand would need to undergo in order to allow an (L)-unit to participate in base-pairing in an environment of (D)-units. Such an analysis is given in Figure 7 on the qualitative level of idealized conformations.

To what extent does chiroselectivity of ligation persist in co-oligomerization when (D→L) substitutions are made in only one of two co-oligomerizing sequences? Co-oligomerization of pr(GCCC)-cP with pr(GCGG)-cP (Table 1, No. 15) stops when, in the former sequence, either the unit at the 2′ end or its neighbor is substituted by an (L)-unit. Reaction to a minor extent (with product consisting exclusively of oligomers of the homochiral partner) occurs when the third cytosine unit is made (L). Finally, when (D→L) substitution refers to the guanine-bearing unit at the 4′ end, oligomerization of the hetereochiral pr-(GCCC) sequence is found to be inhibited more strongly than that of the homochiral pr(GCGG) sequence.

In their pioneering work in the RNA series, Joyce *et al.* [15] have shown that poly(C)-directed oligomerization of (D)-guanosine-5′-phospho-2-methylimidazole to oligomers of (D)-guanosine shows marked chiroselection. In addition, they discovered the reaction to be inhibited by the presence of the activated monomer of opposite sense of chirality, a phenomenon ('enantiomeric cross inhibition') recognized by the authors to be of importance to theories

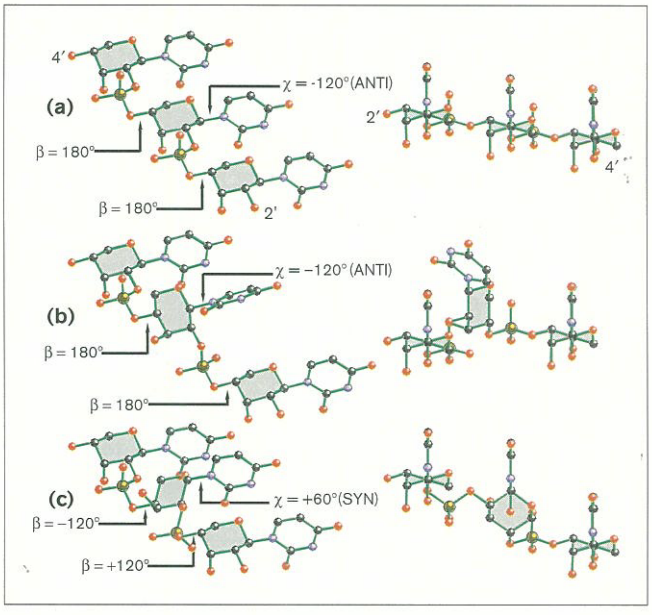
of biogenesis. In template-directed ligation of short oligomers, inhibition of this type is expected to be less important than when a template directs the oligomerization of monomer units. In p-RNA, do chiro-diastereomeric tetramer-cPs interfere with the auto-oligomerization of their homochiral isomers? Experiments with pr(ATCG)-cP, pr(GCCG)-cP and pr(CGGC)-cP and their respective chiro-diastereomers show that they do not, or only weakly so. Figure 8a illustrates an example: oligomerization starting from the entire mixture of the homochiral and the four heterochiral diastereomers of pr(GCCG)-cPs (2.5 mM each) proceeds only slightly slower than the reaction with 2.5 mM of the pure homochiral tetramer under otherwise similar conditions (Fig. 8b).

Extrapolation of the available data predicts that a mixture of the racemic pairs of enantiomers of the complete set of eight diastereomers of pr(GCCG)-cPs (Fig. 8c) should auto-oligomerize to a mixture of oligomers containing predominantly, if not exclusively, homochiral (D) and (L) sequences. Furthermore, we would expect co-oligomerization of tetramers with different but mutually fitting base sequences to proceed with comparable chiroselectivity. In summary, our observations point to the remarkable potential of p-RNA to generate sequence libraries consisting of predominantly homochiral (D) and (L) oligomers, respectively, through chiroselective co-oligomerizations starting from the racemic mixtures of activated tetramers containing all possible chiro-diastereomers of the participating tetramers. The term predominantly homochiral means that not all, but most of the ribose units of a given sequence have the same sense of chirality. So far, p-RNA is the only oligonucleotide system for which the nonenzymic reactivity as well as the chiroselectivity required for such a potential have been demonstrated experimentally. For relevant studies on the chemistry of nucleotide-2′,3′-cyclophosphates in the RNA series see the work of Orgel [16], Usher [17,18] and coworkers, as well as recent unpublished results by M.B., R.M., S. Pitsch and A.E.

Concerning the problem of the origin of biomolecular homochirality

The potential of oligonucleotides to self-assemble by chiroselective co-oligomerization of short activated oligomers calls for an analysis of its relevance to one of the standing problems of an etiology of life, the origin of biomolecular homochirality (see the review of Bonner [19]). In fact, by considerations of the following kind, one can convince oneself at a qualitative level that the capacity for breaking molecular mirror symmetry must be an intrinsic property of an oligonucleotide system. If a mixture consisting of the racemic enantiomer pairs of all eight diastereomers of all 4^4 possible oligonucleotide tetramers were to co-oligomerize stochastically by chirospecific ligation to (D) and (L) libraries consisting each of higher oligomers, say 100-mers, these two libraries would consist of

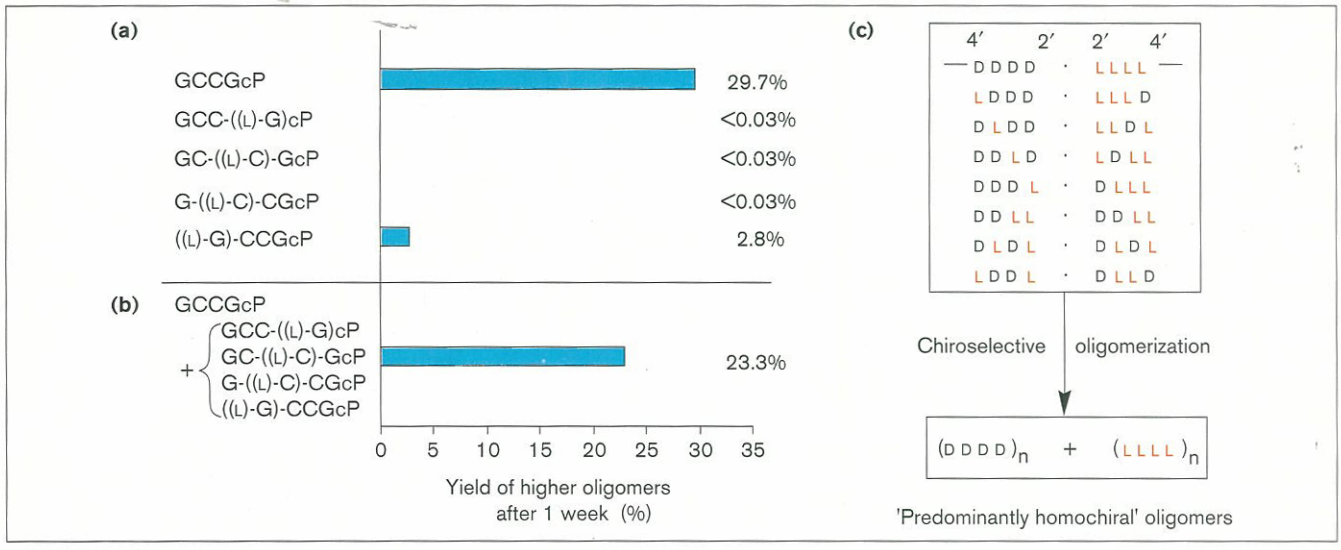
Figure 7



Idealized pairing conformation of: **(a)** excerpt of a homochiral all-(D) p-RNA strand ($\alpha = +60°$, $\beta = 180°$, $\gamma = 180°$, $\delta = 180°$, $\varepsilon = -60°$ $\zeta = 180°$); **(b)** as for (a), except with middle ribopyranosyl unit in an inverted sense of chirality and connected to neighboring (D)-units in such a way that all scalar conformational relationships between phosphodiester and sugar centers remain the same (torsion angle ε at (L)-unit +60° and α of same phosphodiester group −60°); **(c)** same, but with the two torsion angles β flanking the (L)-unit changed to −120° and +120°, respectively, and the nucleobase of the (L)-unit in the (idealized) syn-conformation. Note that in the conformation in (c), the nucleobase of the (L)-unit is in approximate stacking and pairing position; this is, however, only at the cost of two eclipsed bonds (angles β) flanking the (L)-unit and causing steric repulsion between phosphodiester oxygens and (C–5′) methylene groups at the (L) and the adjacent (D) sugar unit. In (D→L) substitutions at the 4′ end, this constraint is one-sided only.

equal amounts of homochiral all-(D) and all-(L) oligomers. The two libraries taken together would no longer constitute a racemic mixture, however. Breaking symmetry by de-racemization would be an inevitable property of the system because, in the formation of both the (D) library and the (L) library, the number of possible sequences with growing oligomer length would soon rise far beyond the number of actually forming sequences. The number of possible sequences in each library would be $(4^4)^{25} = 4^{100}$ ($\sim 10^{60}$), and the (maximum) number of sequences which possibly could be formed would be $\sim 10^{24}$ if co-oligomerization started from a mixture containing 1 mole of each tetramer species (= 4096 moles, corresponding to 1.3×10^3 kg material; also see Fig. 9). Such being the case, and provided that co-oligomerization proceeded stochastically, as supposed, any given sequence in both libraries would eventually occur only once. Thus, the constitutional composition (sequence composition) of the (D) library and the (L) library would not be identical and, therefore, the two

Figure 8



(a) Product yields (after 1 week) referring to auto-oligomerization of homochiral pr(GCCG)-cP and of chiro-diastereomers; **(b)** product yield after oligomerization of the mixture of the five chiro-diastereomers indicated (2.5 mM each); **(c)** extrapolation of observation (b) to a mixture of racemic pairs of enantiomers of the complete set of eight chiro-diastereomers of a tetramer-cP sequence. For the non-reactivity of doubly inverted tetramers in auto-oligomerization see experiments with the sequence pr(ATCG)-cP (Table 1, No. 1). ('Chiro-diastereomerism': diastereomerism with reference to the sense of chirality of nucleotide units only.)

libraries taken together could not constitute a racemic mixture. Two libraries composed of molecular components that have a different constitutional distribution have, in principle, different chemical properties. Specifically, their behavior in amplification by chirospecific sequence replication — again, in principle — would be different, and if the system were to evolve eventually against internal and external selection pressures, emerging winner sequences would be homochiral and belong to either the (D) series or the (L) series.

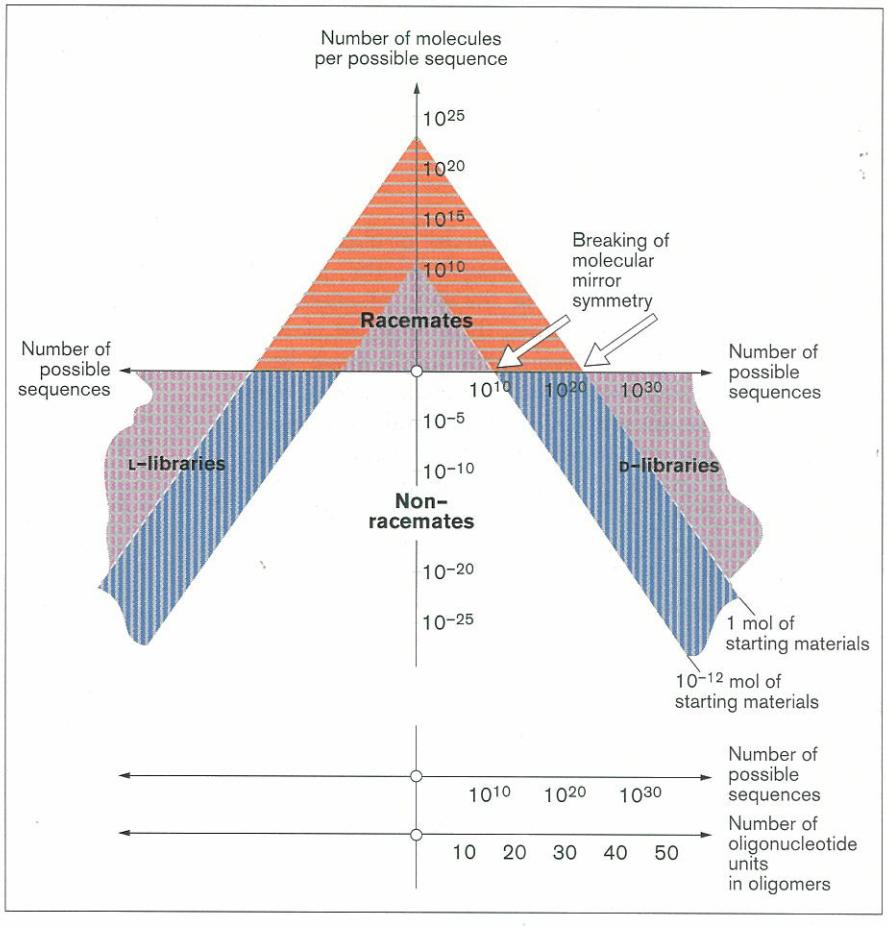
In an etiological context, chiroselective co-oligomerization of activated oligonucleotide tetramers is an exemplary process in the sense that it makes us recognize in a qualitative, yet compelling way the intrinsic capacity of an informational pairing system to break molecular mirror symmetry through its chemistry. The process provides a clear example of how a molecular system is bound to break mirror symmetry in moving, by stochastic chemical changes, from a level of low constitutional complexity (small molecules, low constitutional diversity) to one of high complexity (large molecules, large constitutional diversity), the critical requirement being that it thereby reaches a level of complexity on which the number of possible constitutions far exceeds the number of actually formed constitutions.

Although the stringency of the thought-experiment proposed is independent of whether the suppositions made for the sake of the argument's transparency on a qualitative level could correspond to reality, it is of interest to check these suppositions against what we might expect in real experiments on the basis of what we know about the nonenzymic chemical properties of oligonucleotide systems, such as p-RNA or RNA. Whether the type of oligomerization chemistry of activated tetramers demonstrated with p-RNA also exists in the RNA series remains to be seen. For work which is relevant in this context see [16–18,20,21].

The most serious difficulty for an experimental demonstration of the type of symmetry breaking referred to above is the fact that the efficiency of nonenzymic informational replication required for amplifying sequences for the purpose of documenting such a phenomenon is, at present, out of reach ([15]; M.B., R.M., S. Pitsch and A.E., unpublished results). This may change in the future, however. Other gaps between suppositions and reality refer to the chiroselectivity as well as the kinetic sequence dependence (as compared to a stochastic nature) of co-oligomerization. (D) Libraries and (L) libraries of higher oligomers would not be composed of homochiral sequences, but rather of heterochiral sequences that would be 'predominantly homochiral', however, in the sense that most, rather than all, of the ribose units of a given sequence would have the same sense of chirality. With regard to the sequence dependence of co-oligomerization rate, it would seem that, apart from the possibility of having non-stochiometric ratios of co-oligomerization partners, a rather large allowance for non-participation of tetramer sequences among the total of 4^4 possible tetramers could be made without losing the

Figure 9

Symmetry breaking by stochastic chiroselective co-oligomerization in oligonucleotide systems. The level of sequence diversity on which a stochastic and chirospecific co-oligomerization starting from four racemic mononucleotides would be expected to break molecular mirror symmetry depends on the amount of starting material involved. With the (drastically) simplifying assumption that only oligomers of the same length are formed, symmetry breaking occurs at the level of 37-mers when starting with one mole of each racemic unit, and at the level of 18-mers with a picomole of starting units.



potential for de-racemization. The same holds for the constraints due to self-inhibition of self-complementary, or complementary pairs of tetramer sequences in co-oligomerization. In the following example of a closed co-oligomerization set of 16 tetramer sequences: -ACCA, -GTTG, -ACTG, -GTCA, -CCCA, -CCTG, -ACCC, -GTCC, -GGCA, -GGTG, -ACGG, -GTGG, -GGCC, -CCGG, -GGGG, -CCCC, two sequences only are self-complementary and two form a complementary pair. Stochastic co-oligomerization involving the entire set of the 16 tetramers could potentially produce a library of 16^{25} ($\sim 10^{30}$) 100-mer sequences, and one within a closed subset of only four tetramers (e.g. -ACCA, -GTTG, -ACTG, -GTCA) a library of 10^{15} 100-mers.

Breaking molecular mirror symmetry as an intrinsic consequence of chiroselective combinatorial synthesis of oligonucleotide libraries starting from racemic precursors and being driven beyond a critical level of constitutional complexity seems not to have been — as far as we are aware — an explicit part of the discussions devoted in the literature to the problem of the origin of biomolecular homochirality [19]. The idea of arriving (at least in principle) at molecular homochirality on a macroscopic level through amplification of the handedness of a single molecule, however, goes back, in a way, as far as 1898 ([22], cited in [19]) and is, when referring to RNA, a crucial aspect of Kuhn and Wasser's [23,24] etiological model. The stereochemical consequences of rising configurational complexity of polymers to a level of structural uniqueness of single molecules were pointed out some years ago by Green and Garetz [25] in the context of polymer chemistry and they are also implicit in a recent article [26] that was devoted to the present topic. Finally, chiroselective autocatalysis in self-replication of molecular systems as the means for the amplification of molecular handedness after initial symmetry breaking has been a subject of comprehensive theoretical analyses in the literature (for a review see [19]) and the central role it is supposed to have played in the emergence of Darwinian evolution seems undisputed. Matters of continuing debate refer to the nature and the timing of the initial symmetry breaking combined with the question of whether the sign of today's biomolecular homochirality is accidental or not

[19,26–28]. The widespread conjecture that it is accidental (e.g. [23,29,30]) may be considered to be strengthened by the results presented in this paper, since it is pointing, both conceptually and experimentally, to a robust chemical pathway for oligonucleotide systems to break molecular mirror symmetry before replication (see the recent report in [31], however).

Significance

Our observations on the chiroselectivity of oligomerization of tetramer-2′,3′-cyclophosphates in the p-RNA series influence our thinking in the field of etiological nucleic acid chemistry as follows. The standard view of how long oligonucleotide sequences might have been able to assemble themselves by stochastic oligomerization of activated mononucleotide units should be complemented by consideration of the 'short-oligonucleotide oligomerization pathway'. This pathway is the self-assembly of large and predominantly homochiral sequence libraries by stochastic co-oligomerization of libraries of activated, short oligomers of critical length. The critical length is the length that allows base pairing in overlapping regions of oligomerizing units to be just strong enough for template-directed oligomerization to proceed and, at the same time, weak enough for template control to be able to discriminate sharply between chiro-diastereomeric oligomer units. Predominantly homochiral (D) libraries and (L) libraries would assemble out of a mixture of all possible diastereomers of the polymerizing oligomer units, without stereochemical control by an external catalyst. A type of phosphate activation that is weak enough for oligomerization not to occur except under template control is essential. In the case of p-RNA, activation of the oligomer units as 2′,3′-cyclophosphate seems to provide a nearly ideal solution for that purpose.

The delineation of the existence of a simple and robust chemical process among oligonucleotides that, in principle, is bound to break molecular mirror symmetry whenever the products of the process transcend a critical level of constitutional complexity brings a new element into the ongoing discussion on the origin of biomolecular homochirality.

Acknowledgements

This paper is communication No. 7 in the series 'Pyranosyl-RNA'; for No. 6 see [6]. The work has been supported by Ciba-Geigy AG in Basel, and by the Schulleitung of ETH. R.M. thanks the Fonds zur Förderung der Wissenschaftlichen Forschung, Wien (Austria) for a Schrödinger fellowship. A.E. thanks K. Mislow for informing him of [25] and J. Siegel for making available before publication the manuscript of his forthcoming paper 'The homochiral imperatove of molecular evolution' to appear in the journal 'Chirality'.

References

1. Eschenmoser, A. (1993). Hexose nucleic acids. *Pure Appl. Chem.* **65**, 1179–1188.
2. Pitsch, S., Wendeborn, S., Jaun, B. & Eschenmoser, A. (1993). Why pentose- and not hexose-nucleic acids? Pyranosyl-RNA ('p-RNA'). *Helv. Chim. Acta* **76**, 2161–2183.
3. Pitsch, S., *et al.*, & Eschenmoser, A. (1995). Pyranosyl-RNA ('p-RNA'): base-pairing selectivity and potential to replicate. *Helv. Chim. Acta* **78**, 1621–1635.
4. Krishnamurthy, R., Pitsch, S., Minton, M., Miculka, C., Windhab, N. & Eschenmoser, A. (1996). Pyranosyl-RNA: base pairing between homochiral oligonucleotide strands of opposite sense of chirality. *Angew. Chem. Int. Ed. Engl.* **35**, 1537–1541.
5. Schlönvogt, I., Pitsch, S., Lesueur, C., Eschenmoser, A., Jaun, B. & Wolf, R.M. (1996). Pyranosyl-RNA ('p-RNA'): NMR and molecular-dynamics study of the duplex formed by self-pairing of ribopyranosyl-(C-G-A-A-T-T-C-G). *Helv. Chim. Acta* **79**, 2316–2345.
6. Micura, R., Bolli, M., Windhab, N. & Eschenmoser, A. (1997). Pyranosyl-RNA also forms hairpin structures. *Angew. Chem. Int. Ed. Engl.*, in press.
7. Eschenmoser, A. (1993). Toward a chemical etiology of the natural nucleic acids' structure. In *Proceedings of The Robert A. Welch Foundation 37th Conference on Chemical Research, '40 Years of the DNA Double Helix'*, pp. 201–235, The Robert A. Welch Foundation, Housten, Texas, USA.
8. Fischer, R.W. (1992). Allopyranosyl-Nukleinsäure: synthese, paarungseigenschaften und struktur von adenin-/uracil-haltigen oligonukleotiden. [Allopyranosyl nucleic acid: synthesis, pairing properties and and structure of adenine-/uracil-containing oligonucleotides]. *Thesis No. 9971*, ETH Zürich.
9. Helg, A.G. (1994). Allopyranosyl-nukleinsäure: synthese, paarungseigenschaften und struktur von guanin-/cytosin-enthaltenden oligonukleotiden. [Allopyranosyl nucleic acid: synthesis, pairing properties and structure of guanine-/cytosine-containing oligonucleotides]. *Thesis No. 10464*, ETH Zürich.
10. Diederichsen, U. (1993). Hypoxanthin-basenpaarungen in homo-DNA-oligonucleotiden; zur frage des paarungsverhaltens von glucopyranosyl-oligonucleotiden. [Basepairs involving hypoxanthine in homo-DNA oligonucleotides; concerning the basepairing in glucopyranosyl oligonucleotides]. *Thesis No. 10122*, ETH Zürich.
11. Groebke, K. (1993). Über purin–purin-paarungen bei hexopyranose-nukleinsäuren. [Purine–purine pairing in hexopyranose nucleic acids]. *Thesis No. 10149*, ETH Zürich.
12. Eschenmoser, A. (1991). Warum pentose- und nicht hexose-nucleinsäuren? [Why pentose and not hexose nucleic acids?]. *Nachr. Chem. Tech. Lab.* **39**, 795–806.
13. Hunziker, J., *et al.*, & Eschenmoser, A. (1993). Warum pentose- und nicht hexose-nucleinsäuren? Teil III. Oligo(2′,3′-dideoxy-β-D-glucopyranosyl)nucleotide ('homo-DNS'): paarungseigenschaften. [Why pentose and not hexose nucleic acids? Part III. Oligo(2′,3′-dideoxy-β-D-glucopyranosyl)nucleotides ('homo-DNA'): pairing properties.] *Helv. Chim. Acta* **76**, 259–352.
14. Orgel, L.E. (1992). Molecular replication. *Nature* **358**, 203–209.
15. Joyce, G.F., Visser, G.M., Van Boeckel, C.A.A., van Boom, J.H., Orgel, L.E. & van Westrenen, J. (1984). Chiral selection in poly(C)-directed synthesis of oligo(G). *Nature* **310**, 602–604.
16. Renz, M., Lohrmann, R. & Orgel, L.E. (1971). Catalysts for the polymerization of adenosine cyclic 2′,3′-phosphate on a poly (U) template. *Biochim. Biophys. Acta* **240**, 463–471.
17. Usher, D.A. (1972). RNA double helix and the evolution of the 3′,5′ linkage. *Nature (New Biology)* **235**, 207–208.
18. Usher, D.A. & McHale, A.H. (1976). Hydrolytic stability of helical RNA: a selective advantage for the natural 3′,5′-bond. *Proc. Natl Acad. Sci. USA* **73**, 1149–1153.
19. Bonner, W.A. (1991). The origin and amplification of biomolecular chirality. *Orig. Life Evol. Biosphere* **21**, 59–111.
20. Ertem, G. & Ferris, J.P. (1996). Synthesis of RNA oligomers on heterogeneous templates. *Nature* (London, UK) **379**, 238–240.
21. Ferris, J.P., Hill Jr., A.R., Liu, R. & Orgel, L.E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61.
22. Strong, W.M. (1898). (Letter to the Editor). *Nature* **59**, 53–54.
23. Kuhn, H. (1972). Self-organization of molecular systems and evolution of the genetic apparatus. *Angew. Chem. Int. Ed. Engl.* **11**, 798–820.
24. Kuhn, H. & Waser, J. (1981). Molecular self-organization and the origin of life. *Angew. Chem. Int. Ed. Engl.* **20**, 500–520.
25. Green, M.M. & Garetz, B.A. (1984). The configurational stereochemistry of atactic vinyl homopolymers. *Tetrahedron Lett.* **25**, 2831–2834.
26. Avetisov, V. & Goldanskii, V. (1996). Mirror symmetry breaking at the molecular level. *Proc. Natl. Acad. Sci. USA* **93**, 11435–11442.
27. Mason, S.F. & Tranter, G.E. (1985). The electroweak origin of biomolecular handedness. *Proc. R. Soc. Lond. A* **397**, 45–65.
28. Mason, S.F. (1984). Origins of biomolecular handedness. *Nature* **311**, 19–23.
29. Eigen, M. & Winkler, R. (1975). *Das spiel*, p. 145, R. Piper & Co. Verlag, München.
30. Orgel, L.E. (1973). The origins of life. In *Molecules and Natural Selection*. J. p. 167, Wiley & Sons, N.Y., USA.
31. Cronin, J.R. & Pizzarello, S. (1997). Enantiomeric excesses in meteoritic amino acids. *Science* **275**, 951–955.