Surprising fidelity of template-directed chemical ligation of oligonucleotides

Kenneth D James and Andrew D Ellington

Background: Nucleic acid replication via oligonucleotide ligation has been shown to be extremely prone to errors. If this is the case, it is difficult to envision how the assembly and replication of short oligonucleotides could have contributed to the origin of life and to the evolution of a putative RNA world. In order to assess the fidelity of oligonucleotide replication more accurately, chemical ligation reactions were performed with constant-sequence DNA templates and random-sequence DNA pools as substrates.

Results: In keeping with earlier results, constant-sequence hairpin templates were not faithfully copied by random-sequence substrates. Linear templates, however, showed exceptional replication fidelity, particularly when random hexamers were ligated at 25°C. Surprisingly, at low temperatures the formation of G·A base pairs was common and sometimes occurred even more readily than the formation of the corresponding Watson–Crick A–T and G–C base pairs.

Conclusions: The fidelity of ligation reactions increases with temperature and decreases with the length of the random-sequence substrates. Oligonucleotides with a defined sequence can be copied faithfully in the absence of enzymes. Thus, to the extent that short oligonucleotides could readily have been generated by prebiotic mechanisms, it is possible that the earliest self-replicators arose via oligonucleotide ligation.

Address: Department of Chemistry, Indiana University, Bloomington, IN 47405, USA.

Correspondence: Andrew D Ellington
E-mail: adelling@indiana.edu

Key words: chemical ligation, G·A mismatch, *in vitro* selection, origins of life, self-replication

Received: **23 May 1997**
Revisions requested: **11 June 1997**
Revisions received: **2 July 1997**
Accepted: **3 July 1997**

Chemistry & Biology August 1997, **4**:595–605
http://biomednet.com/elecref/1074552100400595

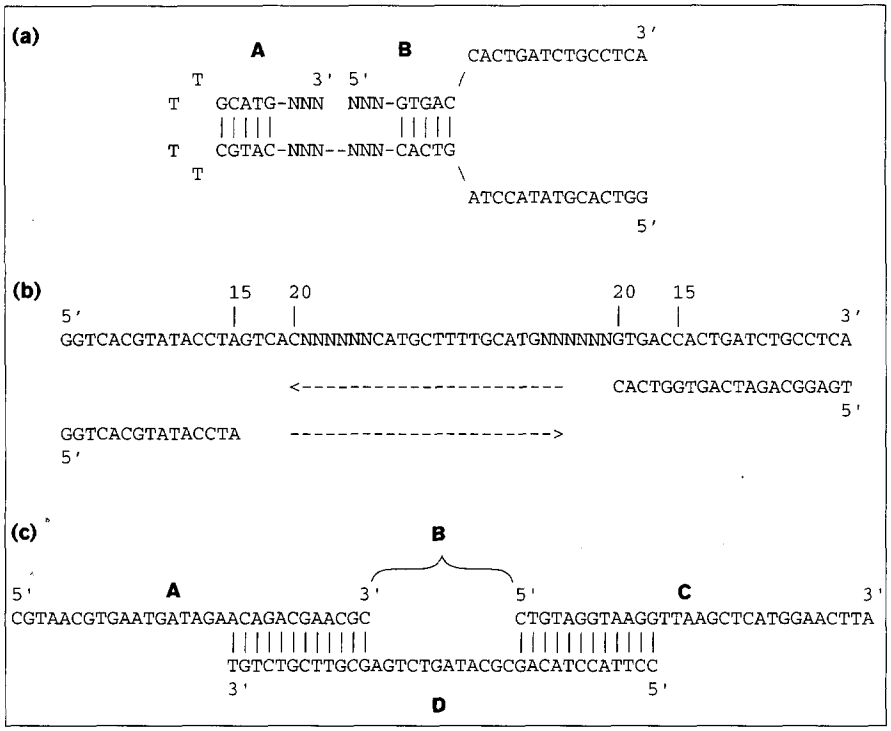
Introduction

It is widely believed that modern metabolism was preceded by an 'RNA world' in which catalysts were nucleic acids rather than proteins. According to this view, the ubiquitous distribution of nucleotide-based cofactors such as ATP is due to their conception in a metabolism in which nucleotide biochemistry played a central role. Likewise, the use of ribosomal RNA as a scaffold and catalyst for protein biosynthesis could be taken to imply that proteins were invented by RNA [1–3]. In support of this idea, several classes of reactions that were previously suspected to be catalyzed solely by protein enzymes have now been shown to be catalyzed by artificially evolved nucleic acid enzymes [4]. While demonstrations of the catalytic ability of nucleic acids lend credence to the possibility of a living world in the absence of proteins, they do not explain how functional nucleic acids may have arisen.

Self-replicating nucleic acids have been touted as the primogenitors of a more complex RNA world. Although the replication of natural nucleic acids is now enzyme catalyzed, replication via chemical catalysis might have been possible at one time. A host of experiments have been performed in which defined sequence templates have been shown to direct the polymerization of nucleotide monomers [5]. The most successful examples have been limited to the polymerization of activated monomers on short poly-cytodine (poly-C) templates [6] or on templates that were almost exclusively poly-C [7–11]. Template-directed ligation of short oligonucleotides has also been shown to be a viable mechanism for both copying and self-replication. In 1986, von Kiedrowski [12] showed that chemical ligation could support the autocatalytic accumulation of a hexamer from two trimers. Later, the oligonucleotide self-replicator was found to exhibit a parabolic growth curve [13] and to effect cross-catalysis as well as autocatalysis [14]. While these experiments illustrated potential prebiotic mechanisms for self-replication, the templates and substrates that were used in elongation or ligation reactions spanned a limited range of sequences. In contrast, it is likely that any oligonucleotides that were generated prebiotically would have spanned a wide variety of sequences and chemistries.

Studies that have addressed the fidelity of putative prebiotic replication mechanisms have generally shown that replication — other than that involving enzyme catalysis — is unfaithful. For example, nucleotide monomers could be polymerized best on templates that were C-rich [5], and not as well on other templates. Moreover, when both enantiomers of activated monomers were used, the insertion of a nucleoside of the opposite handedness effectively poisoned polymerization [15]. It might be expected that the fidelity of template-directed ligation would be much higher

Figure 1



Templates and substrates for ligation reactions. **(a)** The hairpin template (**A**) can pair with a single-stranded substrate (**B**) to form a ligation junction that spans six randomized residues (represented by N). In most experiments, the **B** oligonucleotide was 5′ phosphorylated and, in some instances, contained a radiolabeled phosphate. **(b)** The ligated hairpin oligonucleotide can potentially be unfolded to allow amplification by the polymerase chain reaction (PCR). The 3′ primer (complementary to the **B** oligonucleotide) extended into the paired stem region in order to destabilize the stem and facilitate amplification. **(c)** The linear template (**D**) can pair with the flanking oligonucleotides **A** and **C** to form a gapped hemiduplex. Each constant-sequence overlap was 12 residues in length, as was the gapped region. Several different **B** oligonucleotides were synthesized: a completely complementary dodecamer (12 W-C), a random dodecamer pool (R12), a random hexamer pool (R6), and a random trimer pool (R3). The **C** oligonucleotide was 5′ phosphorylated and, in some cases, contained a radiolabeled phosphate. All **B** oligonucleotides were 5′ phosphorylated.

than the fidelity of template-directed polymerization, both because the binding energy of oligonucleotides would be greater than that of mononucleotides and because fewer reactions would be required for the construction of polymers of equivalent length. Harada and Orgel [16] have shown, however, that, as with monomer addition, oligonucleotide polymerization does not appear to be faithful when random regions are present in the strands to be ligated. Taken together, these results have led those who research the origins of life to doubt whether nucleic acids were in fact the original self-replicators, and the results have prompted experiments with other prebiotic polymers [17–19].

In order to provide more evidence that might rule out or include oligonucleotide ligation as a potential mechanism for the evolution of self-replicating nucleic acids, we re-examined the question of the fidelity of template-directed, chemical ligation of oligonucleotides.

Results and discussion

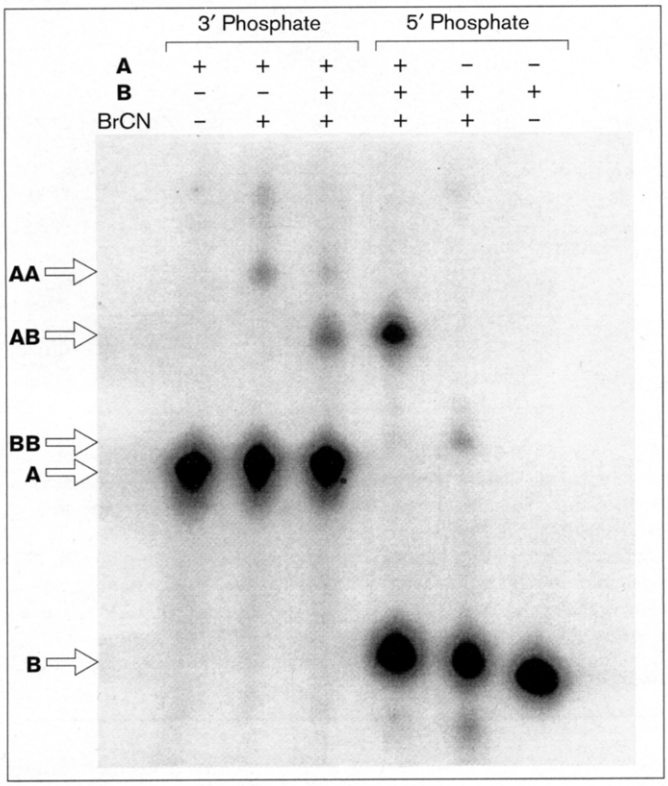
Substrates and conditions for ligation reactions

We attempted to mimic potential prebiotic copying mechanisms by examining the fidelity of chemically catalyzed oligonucleotide ligation using random-sequence substrates. The random-sequence oligonucleotides could either add to one another (Figure 1a) or fill gaps in constant-sequence hemiduplex templates (Figure 1c).

The use of random-sequence substrates for ligation reactions is a significant advance towards understanding how self-replication may have arisen in a prebiotic environment. We and others [20–22] have argued previously that oligonucleotide ligation may provide a means for the purification of particular replicators from complex mixtures, and that such replicators may have provided the fodder for the RNA world. In order for this scheme to have been successful, however, it would have been necessary for the replication process itself to have been faithful. Although the experiments described herein use DNA (rather than RNA) oligonucleotides as substrates, they nonetheless directly address the key issue of replication fidelity. Thus, results from these experiments may also provide insights into the dynamics of other replicating biopolymers, including non-natural nucleic acid analogs [17,19] or even replicating peptides similar to those observed recently by Ghadiri and colleagues [23].

Cyanogen bromide was used as the chemical coupling reagent. Phosphates that have been activated by cyanogen bromide are extremely reactive towards hydroxyl moieties, but are highly dependent on templates for fostering oligonucleotide ligation [24]. Reaction conditions were chosen based on model reactions. We synthesized constant-sequence oligonucleotides and examined their cyanogen bromide-mediated ligation in several different reaction buffers under a variety of coupling conditions that have been published previously [25–30]. The method originally used by Shabarova *et al.* [25] was found to give the best yields and was used for the selection experiments described below. Surprisingly, we found that

Figure 2



Comparison of the reactivity of 5′ and 3′ phosphates using a hairpin template and a single-stranded substrate (as shown in Figure 1a). The three lanes on the left represent *p**A**p (*p: radiolabeled phosphate group) coupling with **B**. The three lanes on the right represent **A** coupling with *p**B**. Reactions were allowed to proceed for 1 min prior to quenching. The expected sizes of the **A**, **B**, and **AB** oligonucleotides are indicated. Small amounts of other ligation products were also observed; the expected sizes of **AA** and **BB** dimers are indicated.

5′ phosphates and 3′ hydroxyls gave superior ligation yields when compared with the more commonly used combination of 3′ phosphates and 5′ hydroxyls. The preference for 5′ phosphate-mediated ligation was evident for both the hairpin and linear constructs (for example, see Figure 2). Consequently, the oligonucleotide pools for selection experiments were synthesized with 5′ phosphates at the ligation junctions.

Searching for base-pairing preferences at a ligation junction

The fact that there are differences in the hydrogen bonding and stacking energies of base pairs [31] suggests that some sequences should be more conducive to template-directed oligonucleotide ligation than others. In order to examine the fidelity of base pairing at a ligation junction and to have the potential to find optimal sequences for ligation, we designed and synthesized a hairpin oligonucleotide template and a substrate that could pair with one another to form a ligation junction. Six random-sequence positions were present at the junction; the randomized positions were held in place by flanking, complementary constant sequences (Figure 1a). Our divided hairpin was reminiscent of one used by Harada and Orgel [16] to select chemically ligated oligonucleotides from random-sequence pools. But, whereas Harada and Orgel randomized only the template strand, we randomized both the template and the substrate. Unpaired regions at the termini of the hairpin served as primer-binding sites and allowed amplification by the polymerase chain reaction (PCR; see Figure 1b).

Ligation time courses revealed that ligation products of the expected length were present even after 1 s (data not shown). Thus, to maximize selectivity, reactions were quenched immediately after the addition of cyanogen bromide (~1 s). The ligation products were isolated on a denaturing polyacrylamide gel and then amplified by PCR. In order to ensure that the amplified sequences arose from template-directed ligation, mock reactions were carried out in the absence of either substrate or template oligonucleotides. DNA molecules that might have co-migrated with a product of the correct size were isolated from these mock reactions and used to seed a PCR; in these cases, no amplification products were observed.

The amplification products from template-directed ligation reactions were cloned and the sequences of the products were determined (Figure 3). Surprisingly, the sequences of the ligated products seemed to reveal a bias against base pairing in the ligation reaction. Very few canonical pairs were formed between the residues in the template strand and the residues in the substrate strand. It was possible that a different ligation complex had formed that was different from the one originally intended, but no apparent pairing between residues in the randomized regions and residues in the constant regions was observed. Instead, the distribution of residues in the originally randomized positions appeared to remain random following ligation, with a small bias towards guanosine.

The results described above are similar to those reported by Harada and Orgel [16], who examined the requirements for the ligation of constant-sequence substrates to a random-sequence template and found that base pairing was necessary in the two positions that were directly proximal to the ligation junction, but base pairing was not necessary in positions that were three or more residues away from the junction. We found that when the template and the substrates were randomized, oligonucleotide ligation reactions appeared even less discriminating. The simplest explanation for the observed differences in fidelity is that more possible junctions were able to form in our experiments. But the differences between the results obtained by Harada and Orgel and those reported here may also be due to differences in experimental design. In their study, a different coupling reagent, ethyl diaminopropyl carbodiimide, was used and the stem of the hairpin was longer

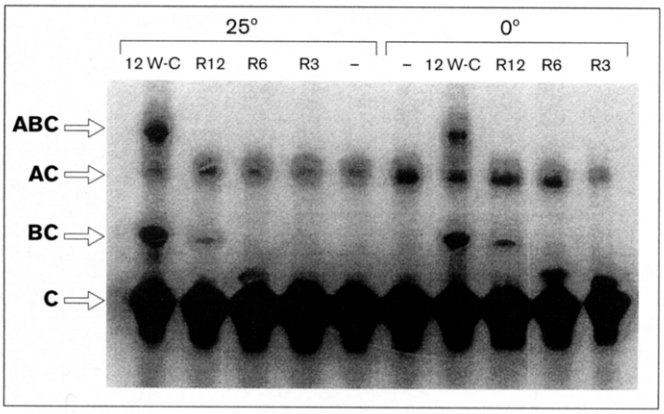
Figure 3

3′ Phosphate	Template region	Loop region	Ligation junction
TATACCTAGTCAC	NNNNNN	CATGCTTTTGCATG	NNNNNN GTGACCACTGATC
TATACCTAGTCAC	CCATGT	CATGCTTTTGCATG	CAGCGT GTG-CCACTGATC
TATACCTAGTCAC	CAGTAA	CATGCTTTTGCATG	GCGCTT GTGACCACTGATC
TATACCTAGTCAC	TGAGCG	CATG-TTTTGCATG	AGCGAG GTGACCACTGATC
TATACCTA----C	ATAGGG	CATGCTTTTGCATG	G--TGG GTGACCACTGATC
TATACCTAGTCAC	TTTTTG	CATGCTTTTGCATG	---ATG GTCACGTATACCT
TATACCTAGTCAC	GGTTCG	CATGCTTTTGCATG	GTGTTG GTGACCACTGATC
TATACCTAGTCAC	TAGAGA	CATGCTTTTGCATG	AGGTCG GTGACCACTGATC
TATACCTAGTCAC	GGGACG	CATGCTTTTGCATG	GGAGTG GTGACCACTGATC
TATACCTAGTCAC	AGGCCG	CATGCTTTTGCATG	TTG--- GTGACCACTGATC
TATACCTAGTCAC	TTAGGA	CATGCTTTTGCATG	CAGCTG GTGACCACTGATC
TATACCTAGTCAC	GCGGAC	CATGCTTTTGCTCG	------ GTGACCACTGATC
TATACCTAGTCAC	TAGATA	CATGCTTTTGCATG	AAGCAA GTAAGCCGAATTC
TATACCTAGTCAC	GTGGGA	CATGCTTTTGCATG	AAAATT GTGACCACTGATC
TATACCTAGTCAC	CAACAA	CATGCTTTTGCATG	GACGAT GTGACCACTGATC
TATACCTAGTCAC	TTGTGT	CATGCTTT-GCATG	AGGGGG GTGACCACTGATC
TATACCTAGTCAC	GCTGAT	CATGCTTTTG-ATG	TGGACTGTA GTGACCACTGATC
TATACCTAGTCAC	CTTTGT	CATGCTTT-GCATG	ATCACTGA GTGACCACTGATC
TATACCTAGT$_G$C$_G$AC	TAATGG	CATGCTTTTGCATG	TTCT-- GTGACCACTGATC

5′ Phosphate	Template region	Loop region	Ligation junction
TATACCTAGTCAC	NNNNNN	CATGCTTTTGCATG	NNNNNN GTGACCACTGATC
TATACCTAGTCAC	ATGGGT	CATGCTTTT-CATG	CGGGGT GTGACCACTGATC
TATACCTAGTCAC	GGCCGT	CATGCTTTTGCATG	TGGTGA GTGACCACTGATC
TATACCTAGTCAC	AGGGCT	CATGCTTTTGCA-G	----CT GTGACCACTGA--
TATACCTAGTCAC	TGAGAT	CATGCTTTTGCATG	CGGATT GTGACCACTGATC
TATACCTAGTCAC	GGGTAT	CATGCTTTT-CATG	GGTAGT GTGACCACTGATC
TATACCTAGTCAC	GATCAA	CATGCTTTTGCATG	GCCTGT GTGACCACTGATC
TATACCTAGTCAC	TTATCC	CATGCTTTTGCATG	CGC--A GTGACCACTGATC
TATACCTAGTCAC	TGTCAC	CATGCTTTT-CATG	AATCGT GTGACCACTGATC
TATACCTAGTCAC	GGGAAG	CATGCTTTTGCATG	AGTTTT GTGACCACTGATC
TATACCTAGTCAC	TAGTGT	CATGCTTTTGCATG	GGCCGT GTGACCACTGATC
TATACCTAGTCAC	AGCTGG	CATG-TTTTGCATG	AATTGT GTGACCACTGATC
TATACCTAGTCAC	AAGGGG	CATGCTTTTGCATG	GGGCGA GTGACCACTGATC
TATACCTAGTCAC	GGGGGG	CATGCTTTTGCATG	AGCAAG GTGACCACTGATC
TATACCTAGTCAC	ACGAGG	CATGCTTTTGCATG	GCATCA GTGACCACTGATC
TATACCTAGTCAC	GGTTAA	CATGCTTTTGCATG	GGTAGG GTGACCACTGATC

Sequences of PCR products derived from the ligation reaction with a hairpin template and a single-stranded substrate. Constant sequences are shown in black; the randomized regions are colored; potential base pairs within the randomized regions are shown in red; mismatches are shown in blue; and deletions are indicated by dashes. The sequences corresponding to the randomized template are on the left and the sequences corresponding to the randomized ligation junction are on the right. Some sequence substitutions and insertions were noted in constant regions; these are shown in italics and subscript italics, respectively. It is unclear whether these substitutions arose during amplification or cloning of the ligation products.

Figure 4



A comparison of the ligation of linear templates and substrates (Figure 1c) at different temperatures. Each lane is labeled according to the **B** oligonucleotide that was used. Only oligomer *p**C** contained a radioactive tag. The expected sizes of the **C**, **BC**, **AC**, and **ABC** oligonucleotides with dodecamer substrates are indicated; smaller **BC** ligation products can be seen in the R6 and R3 lanes at both temperatures. A fully complementary **B** oligonucleotide (12 W-C) was the only one that gave a substantial yield of **ABC** ligation product at both temperatures. The **AC** ligation product was more prevalent at 0°C than at 25°C.

(28 residues rather than 16). More importantly, because the oligonucleotides used by Harada and Orgel terminated in constant sequences rather than in random sequences, any templates that were selected could be regenerated by a nested PCR and the ligation reaction could be carried out again. Optimal templates could therefore be selected over multiple cycles of selection and amplification, and the improved fidelity observed by Harada and Orgel may have been the result of the selection of their ligation junctions over four cycles of selection and amplification.

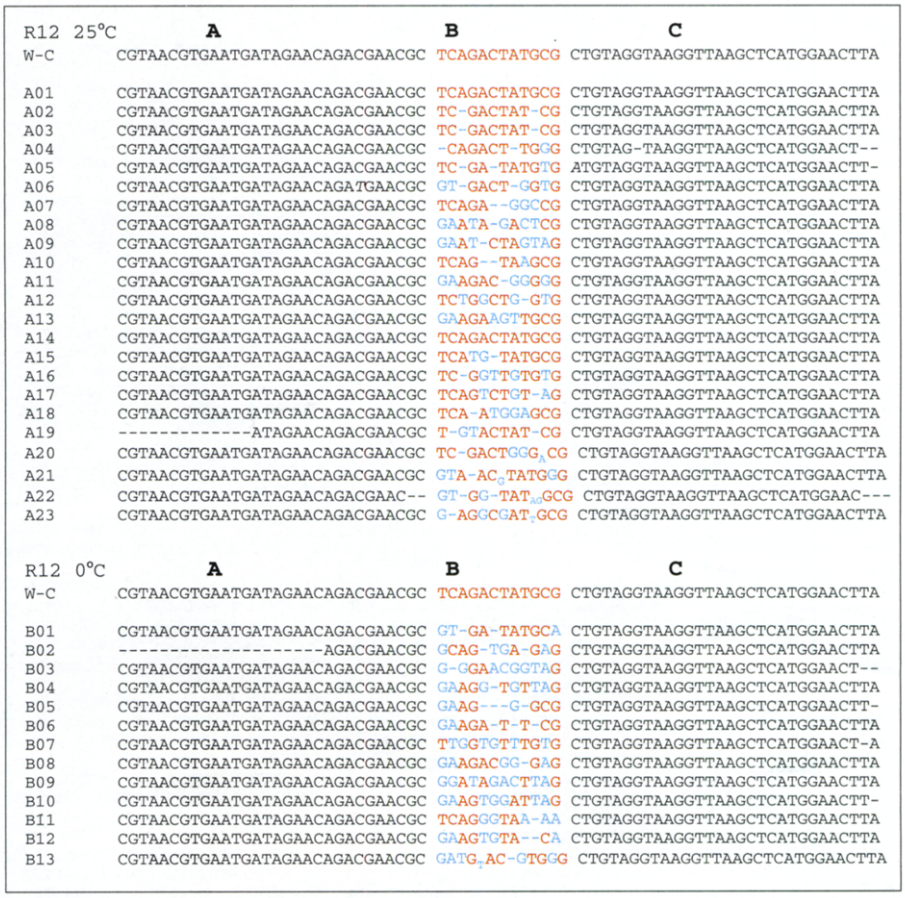
Nonetheless, the overall lack of base pairing that was observed in both studies seemed extremely odd, especially given that in each case a template was present that could have been used. Other studies have shown that single-stranded linear DNA templates, when present, dramatically increase the rate and extent of chemically catalyzed oligonucleotide ligation [29]. The apparent discrepancy between these studies and our results may be due to differences between ligation reactions carried out with constant-sequence rather than random-sequence oligonucleotides, or it may be the result of differences between ligation reactions carried out with linear rather than hairpin oligonucleotides. In the latter case the poor apparent fidelity may be purely artefactual. The oligonucleotides that were cloned and sequenced did not come directly from the ligation reaction, but were first amplified by PCR. Given that the efficiency of template-directed nucleic acid polymerization is sensitive to the secondary structure of the template, it is possible that less stable hairpin structures in the ligated population would have unfolded and been amplified more readily than perfectly paired hairpins. If so, a small subset of the ligated products may have been amplified preferentially and the amplification products that were eventually cloned and sequenced may have been the result of an unintended selection for replicability rather than the intended selection for ligation efficiency. This explanation was supported by experiments in which we attempted to copy (rather than amplify) the single-stranded hairpin template. Following the annealing of the 3′ PCR primer to the hairpin template, neither Taq polymerase, T4 DNA polymerase, nor the Klenow fragment was found to elongate this primer efficiently (data not shown).

Assessing the fidelity of linear ligations

In order to assess better the fidelity of chemically catalyzed oligonucleotide ligation, we switched to a linear

Figure 5

Sequences of PCR products derived from the ligation reaction with linear templates and substrates (randomized dodecamers, R12). Constant sequences are shown in black; randomized regions are colored; residues that can form canonical base pairs with the template are in red; those that are likely mismatches are in blue; and deletions are indicated by dashes. Residues that we suspect are bulged are shown as subscripts and in blue. We attempted to keep predicted bulged residues to a minimum; a residue is presented as bulged only when two or more canonical pairs are aligned as a result. The fully complementary strand (W-C) is presented at the head of each list.



'gap-filling' system (Figure 1c). Although in this configuration the template is constant rather than random, the differential amplification of ligated oligonucleotides should be largely obviated. Random sequence pools containing oligonucleotides of twelve, six or three residues were synthesized as potential gap-filling substrates (via two, three, or five chemical-ligation reactions, respectively). The efficiency of gap-filling reactions with these substrates was monitored relative to a complementary dodecamer (Figure 4). The fully complementary oligonucleotide was clearly orders of magnitude more efficient than any of the random-sequence pools, suggesting that, in this system, base pairing plays a significant role in the efficiency of ligation. This observation held at both 0°C and 25°C.

Although very little ligation product was observed following gel electrophoresis, oligonucleotides of the appropriate length could nonetheless be gel-isolated, amplified, cloned, and sequenced. Several precautions were taken to avoid isolating, amplifying, or cloning artefactual products. First, primer-binding sites were localized within single-stranded overhangs to ensure that no template strands would inadvertently be amplified. Second, the full-length ligation product was isolated prior to amplification to avoid the artefactual generation of a full-length, fully complementary product by polymerase elongation of template and substrate oligonucleotides. Finally, ligation products that were eventually sequenced, and which were derived from random oligonucleotide substrates, were never present on the same gel as control ligation products derived from constant-sequence oligonucleotide substrates.

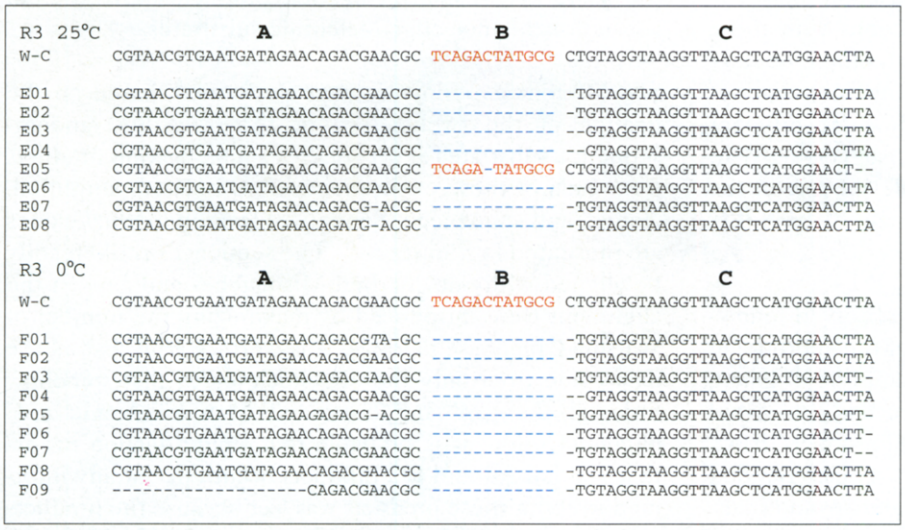
Six different experiments were performed, based on the ligation of the twelve, six, or three residue random-sequence substrates at 0°C or 25°C. The ligation products from each of these experiments were isolated, amplified, cloned, and sequenced. Sequences from individual clones are presented in Figures 5–7. The sequences could readily be aligned with one another. Mismatches and gaps in the alignment were introduced by maximizing the number of Watson–Crick base pairings. It is worth noting that all the data come from only the oligonucleotides that somehow 'bridged the gap' and contained both an **A** and a **C** oligonucleotide. It is probable that single ligation events that yielded **AB** and **BC** products occurred, but any information about these products was lost because the products were not amplifiable.

Figure 6

Sequences of PCR products derived from the ligation reaction with linear templates and substrates (randomized hexamers, R6). Colors and the assignment of bulged residues are as in Figure 5.

Even a cursory examination of the sequence data reveals that, as expected, the linear template was copied much more faithfully than the hairpin template. A more complete analysis of these data is shown in Figure 8. Watson–Crick pairings predominated in all the experiments, but longer substrates and lower temperatures led to less faithful

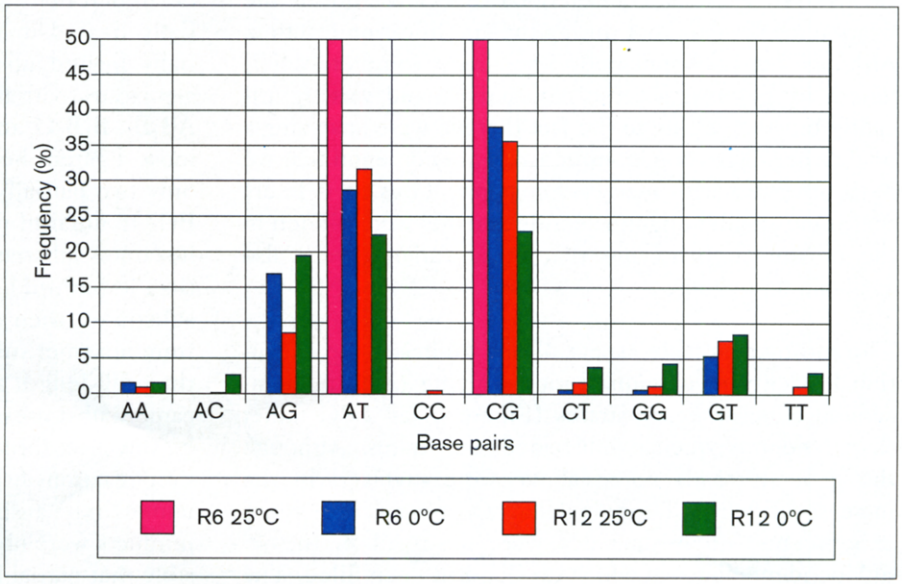
Figure 7



Sequences of PCR products derived from the ligation reaction with linear templates and substrates (randomized trimers, R3). Colors and the assignment of bulged residues are as in Figure 5.

Figure 8

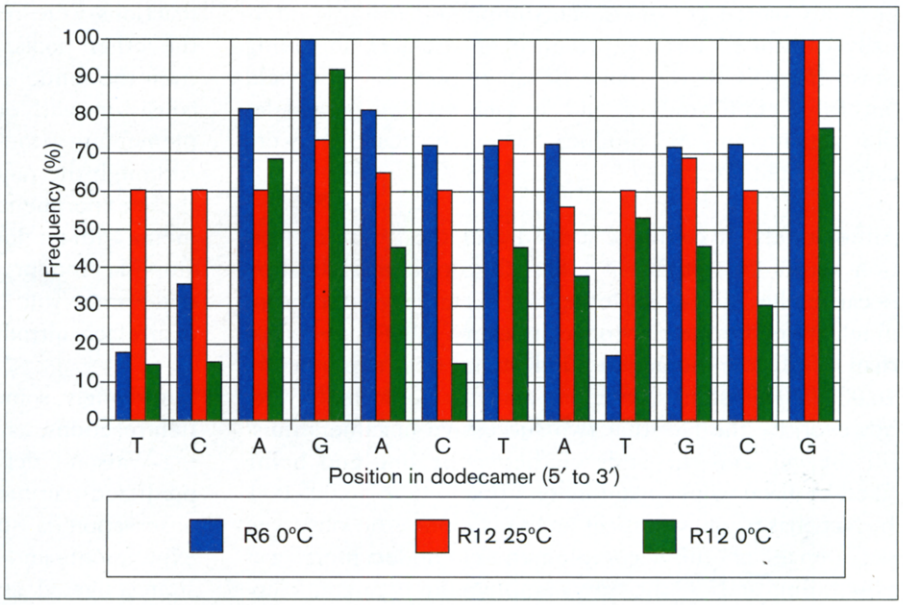
The relative frequency of predicted base pairings in ligation products derived from random hexamer (R6) and dodecamer (R12) substrates. The number of Watson–Crick pairings and mismatches were calculated on the results and alignments shown in Figures 5 and 6; ligation products that contained no insert or that contained only one of two hexamers were not counted in these results. The tallies are expressed as a proportion of total possible pairings (%). Deletions and bulges are not included in this figure, so the bars do not add up to 100%. Because the template contained an equal amount of each nucleotide, perfect pairing would be expected to result in 50% of the pairs being G·C and 50% being A·T; these were in fact the values observed with the random hexamers at 25°C (shown in pink).



copying of the constant-sequence template. Both of these trends were expected: because at higher temperatures (25°C) mismatches were unfavorable, more accurate base pairing was required to bring about ligation than at 0°C. Similarly, the hexamer oligonucleotides were predicted to have lower melting temperatures (~25°C each, disregarding coaxial stacking; 32.5°C and 46.5°C, including coaxial stacking) than the dodecamer oligonucleotides (~62°C, disregarding coaxial stacking, ~71°C including coaxial stacking; predictions were made at 1.0 M NaCl for 20 μM oligonucleotide concentration; John SantaLucia, personal communication). More accurate base pairing would therefore have been required to hold and align the ligation junctions of the hexamers than the dodecamers. When Watson–Crick base pairing was examined as a function of position in the hemiduplex, the two trends were confirmed further (Figure 9). It also appears that there was a more stringent requirement for Watson–Crick complementarity at the 3′ end of the substrate than at the 5′ end, especially at 0°C. In addition, many of the amplification products derived from dodecamer substrates contained one or more deletions. The deletions appear to be localized to the dodecamer substrate and do not extend into the flanking constant sequences, even though some of the constant

Figure 9

The relative frequency of predicted base pairings as a function of position. The results were calculated as in Figure 8, but only Watson–Crick pairings are shown as a function of position in the hemiduplex gap. For simplicity, the random hexamers at 25°C are not shown because they exhibited 100% fidelity at each position. Several of the sites that appear to show little preference for Watson–Crick pairings are instead predicted to form A·G pairings.



regions were not covered by primers. This observation seems to suggest that the deletion events arose during the ligation reaction and not during the amplification reaction. Although the pool appeared to be almost completely full-length by gel analysis, the radioactive band was significantly broadened due to the fact that we were analyzing a pool rather than a single sequence, a phenomenon that we have observed with many other pools. Thus, it is likely that some smaller oligonucleotide variants were hidden in the broadened band and that these variants were also available for ligation.

The most remarkable aspect of our studies is the exceptional fidelity of oligonucleotide copying using random-sequence hexamer substrates (Figure 6) at 25°C rather than random-sequence dodecamer substrates. Although the general trends noted above were involved in the increased fidelity of copying that was observed in these experiments, no mismatches were observed in the 19 independent clones in which the gap had been filled. The degree of selectivity is impressive, especially for a non-enzymatic reaction from a pool of oligonucleotides. The observed selectivity may have a mechanistic rationale: in order to bridge the gap in the hemiduplex by ligation, only the termini of oligonucleotide substrates had to have been positioned properly. Assuming that some base pairing was necessary to position the termini, residues in the interior need not have been paired, an explanation that has previously been championed by Harada and Orgel [16]. Thus, fewer residues could have varied in the shorter ligation products than in the longer ones. As oligonucleotide substrates become shorter and shorter, it would be predicted that there should come a point at which no mismatches can be tolerated. The fact that this point was apparently reached using the hexamer substrates correlates nicely with the predicted melting temperatures (~25°C) of the perfectly paired hexamers. This important result is internally controlled for the presence of artefacts, including those from PCR: the same hexamer pool did not yield faithful ligation products and, in fact, yielded both deletion variants and **AC** products, when the reactions were carried out at 0°C.

Alternatively, it could be argued that the hexamer ligation shows greater fidelity than the dodecamer ligation because the fully complementary hexamer was present much more frequently (once in every 4,096 substrates) than a fully complementary dodecamer (less than once in 16,000,000 substrates). Such an analysis would be flawed, however, by the fact that two successive hexamers must fill the gap and the odds of their annealing and being fixed in place would again be less than one in 16,000,000. In fact, based on statistical analyses alone, the observed prevalence of fully complementary dodecamers was astounding: at 25°C the fully complementary oligonucleotide occurred in 2 out of 23 clones (1 out of every 12 compared to 1 out of every 16,000,000). Similarly, it could be argued that there are more possible ways to make partially ligated unamplifiable hexamer products than partially ligated unamplifiable dodecamer products. There are four partially ligated hexamer products — $\mathbf{AB}_1$, $\mathbf{AB}_1\mathbf{B}_2$, $\mathbf{B}_1\mathbf{B}_2\mathbf{C}$, and $\mathbf{B}_2\mathbf{C}$ (where $\mathbf{B}_1$ and $\mathbf{B}_2$ are the 5′ and 3′ most ligated hexamer, respectively) — compared with only two partially ligated dodecamer products — **AB** and **BC**. If unfaithful partial ligation products derived from hexamer substrates are also more difficult to complete than unfaithful ligation products derived from dodecamer substrates, then full-length amplifiable products derived from hexamer substrates might appear to be more faithful than full-length amplifiable products derived from dodecamer substrates. Again, though, this does not appear to be the case: three of the 17 sequences derived from ligation reactions with the hexamer substrate at 0°C contained only a single hexamer in the gap and no such products were observed at 25°C. Thus, although it is possible that partial ligation products may lead to an overestimate of the fidelity of ligation for hexamer substrates, it is apparent that this effect is relatively minor.

On the basis of the analysis outlined above, oligonucleotides of a particular length might be required for high-fidelity template copying, but oligonucleotides below this length would not be expected to promote efficient copying. In support of this interpretation, the random trimers clearly did not function as well as the longer oligonucleotides in the gap-filling experiments (Figure 7). Even though the regions of the gel that were excised corresponded to the position where the full-length product should have been, the amplification products were almost exclusively ligation products of the **A** and **C** oligonucleotides (**AC** products). The reduced frequency of gap filling may also have been due to the fact that five ligation reactions were required, rather than the two or three for the other pools. The one full-length product obtained from the random trimers, E05, contained no mismatches, consistent with our observation of the importance of complementarity for gap filling by shorter oligonucleotides. Although the random trimers did not excel at gap filling, these experiments are an excellent control for the experiments with random hexamers and random dodecamers: the amplification products in the experiments with random dodecamers and hexamers were not likely to be the result of artefacts introduced by amplification or other means.

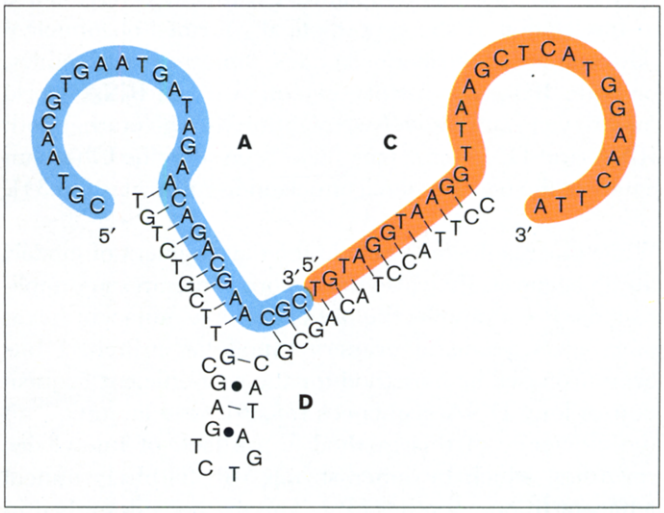
Surprisingly, some mismatches were clearly favored in ligation reactions, especially those carried out at 0°C. Excluding Watson-Crick interactions, the most populous class of predicted pairings were G·A interactions (Figure 8). This observation is supported by examinations of individual experiments and clones. At the 5′ end of dodecamer substrates ligated at 0°C, guanosine (which would be predicted to form an G·A pair) was selected in preference to

thymidine (which would have formed the canonical T·A pair) in 11 of the 13 clones (B01–B13). A possible explanation is that a guanosine at the 5′ terminus of oligonucleotide **B** gave a favorable stacking interaction with the cytidine at the 3′ terminus of oligonucleotide **A**. While this energetic argument may be true, it does not account for other sites at which G·A pairing was observed. In the same sequences (B01–B13), the penultimate residue at both the 3′ and 5′ ends of oligonucleotide **B** would have been expected to be cytidine, which would form a C·G pair. Instead, it was adenosine forming an A·G pair in 7 of the 13 clones. The formation of G·A pairings also explains the relatively small number of T·A pairings seen at the ninth position in the copied DNA strand (Figure 9).

By and large, our observation that G·A pairs appear to be favored at lower temperatures is congruent with thermodynamic and structural observations derived from nuclear magnetic resonance studies. Patel and coworkers [32] analyzed double-stranded DNA decamers containing two G·A base pairs that were each surrounded by canonical pairings. They found that the imino proton of guanosine in a G·A base pair was quite resistant to exchange with solvent at low temperatures. In addition, base pairs adjacent to the G·A pair were not disrupted. In contrast, at higher temperatures (25–35°C) the imino protons in an A·G base pair were readily exchanged with the solvent and adjacent base pairs were disrupted. It should be noted that this analysis also confirms that the composition of the original random-sequence populations did not distort our results. Although some positions showed a greatly reduced tendency to form Watson–Crick base pairings, in many cases this was due to the formation of alternative pairings. Overall, each residue was present in sufficient abundance, so canonical pairings could have readily formed if they had been favored.

Curiously, a cytidine was omitted at the ligation junction of every **AC** ligation product that was derived from experiments with the trimer or other substrates (Figure 7; see also sequences C11, C13, C18, D03, and D06). But this cytidine was never omitted from the ligation junction of full-length or other products. The deleted ligation products were most likely to have been derived from foreshortened **C** oligonucleotides that arose during chemical synthesis. The preferential inclusion of such failure sequences in the **AC** ligation products may have been the result of the formation of an alternative structure for ligation. If the terminal residue in the **C** oligonucleotide was missing prior to ligation, a structure could have formed in which the 3′ end was held, by three canonical pairings, adjacent to the 5′ terminus of the **C** oligonucleotide (Figure 10). This alternative structure could not have been formed readily by **C** oligonucleotides that contained the 5′ terminal cytidine. Formation of this alternative structure may have been mediated by a short hairpin hinge that is predicted to contain two A·G base pairs. The stability of these A·G pairings at 0°C may explain why the **AC** product was far more abundant at 0°C than at 25°C (see Figure 4).

Figure 10



Proposed intermediate in the formation of **AC** ligation products. When the cytidine at the 5′ terminus of oligonucleotide **C** (shaded in red) was missing (as shown in this figure), the 3′ terminus of oligonucleotide **A** (shaded in blue) could form three adjacent base pairings. Chemical ligation would link the 5′ phosphate of oligonucleotide **C** with the 3′ hydroxyl of oligonucleotide **A**, generating an **AC** ligation product with a missing cytidine residue. This ternary complex may be stabilized by a short stem–loop structure in oligonucleotide **D** that contains two A·G pairings.

Significance

Our results bear on hypotheses regarding the origin and evolution of life. Given that the prebiotic synthesis of short (~3–6 residue) oligonucleotides can be demonstrated experimentally and that long (> 50 residue) ribozymes could have been part of a putative 'RNA world', a viable scenario for the evolution of self-replicating oligonucleotide 'missing links' must be developed. Until now, no mechanism has been demonstrated that could have led to the efficient copying of mixed sequences longer than hexamers. Moreover, because faithful copying of templates by random-sequence substrates mimics the boundary conditions in the primordial soup, it is noteworthy that this copying scenario has not been observed previously. In this paper, we show that an oligonucleotide with a defined sequence can be synthesized from a random pool of short oligonucleotides at ambient temperatures in the absence of enzymes. This observation suggests that prebiotic oligonucleotide replication might also have occurred with a high degree of fidelity. Specificity increases with temperature but decreases with the length of the oligonucleotides to be ligated, suggesting that different replicators might have been favored in different environments. If chemically catalyzed oligonucleotide ligation can now be considered as a mechanism for

the faithful replication of DNA oligonucleotides, then perhaps comparable templating effects could have aided in the ligation of short, prebiotically formed oligonucleotides composed of ribonucleotides, unnatural nucleotides, or both. If so, the further evolution of an RNA world may have been relatively straightforward given that both RNA and DNA enzymes have been identified that can catalyze the ligation of oligonucleotide substrates [33,34].

While our results may help in the development of models for abiogenesis, they also have more contemporary applications. For example, the use of multiple, adjacent hexamers as sequencing primers ('modular priming') has been proposed as a method for the cost-efficient acquisition of long DNA sequences [35,36]. One limitation on the efficiency of this method is the lack of knowledge regarding which hexamer strings will faithfully anneal adjacent to one another [35]. The methods described in this paper can provide such insights. Similarly, the fidelity of oligonucleotide hybridization and ligation is critical to the success of several schemes that have been proposed for DNA computing [37–39]. Again, experiments similar to those described in this paper might be used to determine the error rate for such algorithms.

Materials and methods

Materials and reagents

All oligonucleotides were synthesized on an ABI PCR Mate DNA synthesizer. Reagents were purchased from Glen Research (Sterling, VA, USA); columns were purchased from Cruachem (Dulles, VA, USA). Terminal 5′ and 3′ phosphates were added to oligonucleotides during automated synthesis. Oligonucleotides longer than 15 residues were purified by gel electrophoresis. T4 polynucleotide kinase (New England Biolabs, Beverly, VA, USA) and [γ-^{32}P]ATP (NEN Research Products, Boston, VA, USA) were used to end-label oligonucleotides. Taq polymerase was purchased from Promega (Madison, WT, USA); Sequenase was purchased from United States Biochemical (Cleveland Oh, USA). Cyanogen bromide was obtained from Aldrich (Milwaukee, WT, USA).

Chemical ligations

Oligonucleotides were combined in a thin-walled PCR tube at a concentration of 20 μM each in a final volume of 4.5 μl of 0.25 M MES buffer (pH 7.5; adjusted with Et_3N) and 20 mM $MgCl_2$. The strands were denatured at 95°C for 3 min and cooled to room temperature over 15 min. The solutions were then placed at the desired temperature (0° or 25°C) and cyanogen bromide (0.5 μl of a 5 M solution in acetonitrile) was added. The reactions were quenched at times ranging from 1 s to 1 min by the addition of absolute ethanol (100 μl). After centrifugation at 4°C for 30 min, the supernatant was removed from the precipitated DNA. The pellets were further dried in a vacuum centrifuge and then resuspended in 5 μl water.

Gel isolation and PCR amplification

Ligation products were separated from nonligated oligonucleotides by gel electrophoresis (12% polyacrylamide; 19:1 monoacrylamide : bisacrylamide). Radioactive bands of appropriate length were cut from the gels and eluted overnight at 37°C into 0.3 M NaCl. The eluted oligonucleotides were precipitated by addition of three volumes of ethanol followed by centrifugation (4°C). The pellets were washed with 70% ethanol, air-dried, and resuspended in 5 μl water. The oligonucleotides were amplified by the PCR. For the hairpin ligation reactions, 2 μl of the isolated ligation product was added to a reaction mixture that contained 5% acetamide, 0.05% NP-40, 200 μM dNTPs, 500 nM of each primer (Figure 1b), 2.0 mM $MgCl_2$, 15 mM $(NH_4)_2SO_4$, 60 mM Tris-HCl, pH 8.5. For the linear ligation reactions, 2 μl of the isolated ligation product was added to a reaction mixture that contained 5% acetamide, 0.05% NP-40, 200 μM dNTPs, 500 nM for each primer, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl, pH 8.3. The differences in reaction buffers were not found to affect the relative efficiency of amplification of hairpin versus linear templates: in both cases hairpin templates amplified poorly. The second set of buffer conditions, however, proved to be more efficient for the amplification of the linear templates. In each thermal cycle, the temperature was held at 94°C for 30 s, 45°C for 30 s, and 72°C for 30 s.

Molecular biology

Double-stranded PCR products were cloned using TA Cloning and Zero Blunt-End Cloning kits by Invitrogen (Carlsbad, CA, USA). Minipreps of plasmids were performed by a common protocol [40]. Sequencing was performed using a standard dideoxy method.

Acknowledgements

We wish to thank J. SantaLucia for many helpful discussions. This work was supported by a Cottrell Scholar Award from the Research Corporation (A.D.E.) and a National Science Foundation Young Investigator Award (A.D.E.).

References

1. White III, H.B. (1976). Coenzymes of fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101-104.
2. Visser, C.M. & Kellogg, R.M. (1978). Bioorganic chemistry and the origin of life. *J. Mol. Evol.* **11**, 163-169.
3. Benner, S.A., Ellington, A.D. & Tauer, A. (1989). Modern metabolism as a palimpsest of the RNA world. *Proc. Natl Acad. Sci. USA* **86**, 7054-7058.
4. James, K.D. & Ellington, A.D. (1997). In *The Molecular Origins of Life* (Brack, A., ed.), University of Cambridge Press, Cambridge, UK, in press.
5. Orgel, L.E. (1992). Molecular replication. *Nature* **358**, 203-209.
6. Inoue, T. & Orgel, L.E. (1982). Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C). An RNA polymerase model. *J. Mol. Biol.* **162**, 201-217.
7. Inoue, T. & Orgel, L.E. (1983). A nonenzymatic RNA polymerase model. *Science* **219**, 859-862.
8. Inoue, T., *et al.*, & Reese, C.B. (1984). Template-directed synthesis on the pentanucleotide CpCpGpCpC. *J. Mol. Biol.* **178**, 669-676.
9. Joyce, G.F., Inoue, T. & Orgel, L.E. (1984). Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, U) templates. *J. Mol. Biol.* **176**, 279-306.
10. Joyce, G.F. & Orgel, L.E. (1986). Non-enzymic template-directed synthesis on RNA random copolymers. Poly(C, G) templates. *J. Mol. Biol.* **188**, 433-441.
11. Joyce, G.F. & Orgel, L.E. (1988). Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, A) templates. *J. Mol. Biol.* **202**, 677-681.
12. von Kiedrowski, G. (1986). A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed. Engl.* **25**, 932-935.
13. von Kiedrowski, G., Wlotzka, B., Helbing, J., Matzen, M. & Jordan, S. (1991). Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3′-5′-phosphoamidate linkage. *Angew. Chem. Int. Ed. Engl.* **30**, 423-426.
14. Sievers, D. & von Kiedrowski, G. (1994). Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221-224.
15. Joyce, G.F., *et al.*, & van Westrenen, J. (1984). Chiral selection in poly(C)-directed synthesis of oligo(G). *Nature* **310**, 602-604.
16. Harada, K. & Orgel, L.E. (1994). *In vitro* selection of optimal DNA substrates for ligation by a water-soluble carbodiimide. *J. Mol. Evol.* **38**, 558-560.
17. Visscher, J. & Schwartz, A.W. (1988). Template-directed synthesis of acyclic oligonucleotide analogues. *J. Mol. Evol.* **28**, 3-6.
18. Eschenmoser, A. & Dobler, M. (1992). Why pentose- and not hexose-nucleic acids? Introduction to the problem, conformational analysis of oligonucleotide single strands containing 2′, 3′-dideoxyglucopyranosyl building blocks ('homo-DNA'), and reflections on the conformation of A- and B-DNA. *Helv. Chim. Acta* **75**, 218-259.
19. Bohler, C., Nielsen, P.E. & Orgel, L.E. (1995). Template switching between PNA and RNA oligonucleotides. *Nature* **376**, 578-581.
20. Ellington, A.D. (1993). Experimental testing of theories of an early RNA world. *Methods Enzymol.* **224**, 646-664.

21. James, K.D. & Ellington, A.D. (1995). The search for missing links between self-replicating nucleic acids and the RNA world. *Orig. Life Evol. Biosph.* **25**, 515-530.
22. Bolli, M., Micura, R. & Eschenmoser, A. (1997). 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). *Chem. Biol.* **4**, 309-320.
23. Lee, D.H., Granja, J.R., Martinez, J.A., Severin, K. & Ghadiri, M.R. (1996). A self-replicating peptide. *Nature* **382**, 525-528.
24. Dolinnaya, N.G., Sokolova, N.I., Ashirbekova, D.T. & Shabarova, Z.A. (1991). The use of BrCN for assembling modified DNA duplexes and DNA-RNA hybrids; comparison with water-soluble carbodiimide. *Nucleic Acids Res.* **19**, 3067-3072.
25. Shabarova, Z.A., *et al.*, & Bogdanov, A.A. (1991). Chemical ligation of DNA: the first non-enzymatic assembly of a biologically active gene. *Nucleic Acids Res.* **19**, 4247-4251.
26. Dolinnaya, N.G., *et al.*, & Shabarova, Z.A. (1993). Oligonucleotide circularization by template-directed chemical ligation. *Nucleic Acids Res.* **21**, 5403-5407.
27. Kanaya, E. & Yanagawa, H. (1986). Template-directed polymerization of oligoadenylates using cyanogen bromide. *Biochemistry* **25**, 7423-7430.
28. Luebke, K.J. & Dervan, P.B. (1989). Nonenzymatic ligation of oligodeoxyribonucleotides on a duplex DNA template by triple-helix formation. *J. Am. Chem. Soc.* **111**, 8733-8735.
29. Li, T. & Nicolaou, K.C. (1994). Chemical self-replication of palindromic duplex DNA. *Nature* **369**, 218-221.
30. Rubin, E., Rumney, S.T., Wang, S. & Kool, E.T. (1995). Convergent DNA synthesis: a non-enzymatic dimerization approach to circular oligodeoxynucleotides. *Nucleic Acids Res.* **23**, 3547-3553.
31. SantaLucia, J., Allawi, H.T. & Seneviratne, P.A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555-3562.
32. Patel, D.J., Kozlowski, S.A., Ikuta, S. & Itakura, K. (1984). Deoxyguanosine-deoxyadenosine pairing in the d(C-G-A-G-A-A-T-T-C-G-C-G) duplex: conformation and dynamics at and adjacent to the dG:dA mismatch site. *Biochemistry* **23**, 3207-3217.
33. Green, R. & Szostak, J.W. (1992). Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* **258**, 1910-1915.
34. Cuenoud, B. & Szostak, J.W. (1995). A DNA metalloenzyme with DNA ligase activity. *Nature* **375**, 611-614.
35. Kieleczawa, J., Dunn, J.J. & Studier, F.W. (1992). DNA sequencing by primer walking with strings of contiguous hexamers. *Science* **258**, 1787-1791.
36. Kotler, L.E., Zevin-Sonkin, D., Sobolev, I.A., Beskin, A.D. & Ulanovsky, L.E. (1993). DNA sequencing: modular primers assembled from a library of hexamers or pentamers. *Proc. Natl Acad. Sci. USA* **90**, 4241-4245.
37. Adleman, L.M. (1994). Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021-1024.
38. Lipton, R.J. (1995). DNA solution of hard computational problems. *Science* **268**, 542-545.
39. Rozen, D.E., McGrew, S. & Ellington, A.D. (1996). Molecular computing: does DNA compute? *Curr. Biol.* **6**, 254-257.
40. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). In *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA.