

RESEARCH ARTICLE

Open Access

Synthesis of information-carrying polymers of mixed sequences from double stranded short deoxynucleotides

Olga Taran*, Oliver Thoennessen, Karin Achilles, Günter von Kiedrowski

Abstract

Background: The "RNA World" hypothesis suggests that an early form of life on Earth was based on nucleic acid strands able to store genetic information and catalyze a wide range of reactions including those which lead to self-replication. For this hypothesis to be true there must exist an efficient process for creating RNA or RNA-like polymers of mixed sequences from short precursors, where these polymers have to be long enough to fold into catalytically active structures (at least 40 bases).

Results: We report on the polymerization of dimeric to hexameric 5'-amino- oligodeoxynucleotides 3'-phosphates in the presence of the water-soluble carbodiimide EDC. Non-complementary single stranded nucleotides fail to polymerize and yield di- to hexameric cyclooligomers or capped EDC-adducts unable to undergo further 3'-5'-phosphoramidate formation. Complementary building blocks polymerize with a conversion close to 100% when starting from a concentration of typically 20 mM. The reactions proceed within a few hours yielding strands of mixed pyrimidine-purine sequences up to 300 bases long. The maximum length of the products depends on the type of the starting oligonucleotides. Copolymerization of a dimer alphabet consisting of equimolar quantities of all four sequences d(nYRp), where Y are pyrimidines and R are purines, generates a mixed-sequence library of 50-70 mers.

Conclusions: Libraries of long oligonucleotides with potentially catalytic activity are formed from short precursors within hours. Reactions occur via blunt end ligation of the double strands, and the reaction rates correlate with stacking interactions at the ligation sites. Circular dichroism measurements, polarized light microscopy and fluorescence microscopy suggest the formation of supramolecular aggregates during chain growth. These aggregates accelerate the reactions by increasing the local concentration of the reactants in a non-sequence-specific templating mode. Aggregation of the double strands into higher order "compartmented" structures might have been the key for the formation of the first inhabitants of the "RNA World".

Background "RNA World"

The first form of life on early Earth probably was based on groups of long aperiodic polymers capable of information transfer and catalysis. The discovery of catalytic RNA molecules led to the postulation of the "RNA World" hypothesis by Gilbert in 1986 [1]. Over the last two decades, strong experimental evidence has appeared in support of the theory. Ribonucleotides can be formed

from small organic molecules under conditions resembling those of the primitive Earth [2,3]; RNA and DNA molecules can form simple reaction networks by catalyzing replication of their complementary strands[4,5]; RNA random pools can be evolved *in vitro* into ribozymes with various catalytic functions. The dynamics of the formation of new functions in such pools and their subsequent evolution have been extensively explored, both experimentally [6] and theoretically [7]. To perform catalysis, RNA molecules need a minimal length to be able to fold into an functional tertiary structure. The normal minimal length of aptamers, ribozymes and

* Correspondence: olga@oc1.rub.de
Department of Organic Chemistry I, Ruhr University Bochum,
Universitätsstraße 150, 44780 Bochum, Germany
Full list of author information is available at the end of the article

deoxyribozymes is 30 bases [6], but small catalytic RNA molecules as long as 5 bp are also known [8].

S. Kauffman proposed, that in a mixture of catalytically active polymers able to catalyze different types of reactions, a reaction network will appear in which all the species are interconnected by the number of reactions where they function as mutual catalysts. He called these interconnected groups "autocatalytic sets" [9]. When natural selection leading to Darwinian Evolution is applied to the mixture, the groups have a significant evolutionary advantage for the propagation of their genetic information, compared to single self-replicating molecules. The appearance of autocatalytic sets on the early Earth required a method for the formation of the mixtures of polymers able to perform catalysis.

The idea of the "autocatalytic set" suggests that the basis of the first living organisms could have been not only RNA, but rather any kind of functional polymers able to catalyze reactions and to transfer information by templating. This suggestion is partially supported by experiments. Systems based on self-replicating peptides [10] and small molecules [11] have been reported in the last two decades.

Formation of information-carrying polymers under potentially prebiotic conditions

It is still unclear how catalytically active polymers can be formed spontaneously in solution from shorter precursors. Stepwise stochastic addition of mononucleotides to the growing strand is ineffective because of slow reaction rates, low conversion, numerous side products [12] and the hydrolysis of the already formed strands.

The low reactivity of the phosphate ester bond can be enhanced by the use of activated precursors. Common chemical modifications are amine or imidazole groups at the reactive ends [13] and the use of cyclic phosphate esters as precursors [12,14]. Heating can provide the necessary energy for the RNA bond formation, but strands formed this way will undergo rapid hydrolysis [14,15].

Catalysis on surface is one of the solutions to the problem. Conditions on the primitive Earth suggest that different types of surfaces such as clays [16], ice [17] or lipids [15] could promote the polymerization of the nucleic acids.

Ferris et al. have shown that montmorillonite clay catalyzes a ligation of phosphoimidazolid-activated ribonucleotide monomers, and oligomers 55 bases long can be formed on its surface when fresh supply of the mononucleotides is provided [16].

The phosphoimidazolid-activated ribonucleotide monomers in eutectic ice-water phase in presence of Pb^{2+} or Mg^{2+} can polymerize to give 17-mers, as shown by Monnard et al. [17]. The incorporation of all four

types of building blocks was observed in this case. Rajamani et al. have shown that heating and cooling cycles in presence of lipid membranes promote the oligomerization of not activated mononucleotides to oligomers up to 100 bases long [15]. The ordered surfaces formed by lipids instruct RNA polymerization, which is inefficient in case of the same tidal condition changes in solution of pure mononucleotides.

The reaction in bulk solution provides positive results when starting materials are changed. Heating of cyclic purine monoribonucleotides in water at 80°C allows their polymerization and the formation of the strands up to 120 bp long [14]. The reaction leads to correct regioselectivity (due to hydrolysis of the more labile 2'-3'- bond under the reaction conditions), but it does not occur with cyclic pyrimidines.

The use of double-stranded precursors can help to take the advantage of template properties of nucleic acids. Mononucleotides cannot be organized into double helices. Longer building blocks are needed.

The first system of this type was suggested by W. Ziehlinsky and L.E. Orgel in 1986 [18]. They studied the self-complementary 3'-amino-5'-phosphate dinucleotides pGCn and pCGn (here and below the "n" symbol corresponds to the amino group while the "p" symbol represents the phosphate group). Water soluble 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) was used as a condensing agent to increase ligation rates. Side reactions leading to cyclic products were partially reduced, compared to mononucleotide polycondensation, but not completely eliminated. Oligomers up to 20 bp long were observed. The polymerization of pGCn yielded approximately 20 bp long strands, while pCGn predominantly produced a cyclic dimer. The authors explained the results in terms of the internal stability of the mini-helices formed by the dinucleotides.

Bolli et al. have shown that pyranosyl-RNA tetramers with partially overlapping sequences polymerize in presence of EDC to give 36 bases long products after 6 weeks of reaction at 4°C [19]. Pyranosyl-RNA is the constitutional isomer of RNA with ribose units in their pyranose (instead of furanose) form and with phosphodiester bridges between the position 4' and 2' instead of 5' and 3' of the ribose ring. It has higher templating properties (both in strength and selectivity) than normal RNA. This system was able to discriminate mismatches in sequences and sugars of different chirality presented in the reaction mixture.

Recently, Horowitz et al. showed that, in the presence of intercalators e.g. ethidium bromide, that stabilize double strands, oligonucleotides up to 100 bp are formed, starting from concatenated tetrameric building blocks [20]. These authors demonstrated that a simple limitation of the cyclization pathway through the

formation of relatively rigid double strands is enough to displace the equilibrium toward the formation of the long oligonucleotides.

Nakata et al. have shown that, at high concentrations, 6 bp long double stranded oligonucleotides can form liquid crystalline phases, where the duplexes are merged together by “physical polymerization” based on stacking interactions [21]. The authors suggested that, in the presence of an efficient condensing agent, the molecules in the stacks could react and form long strands. The use of supramolecular self-organization to complete the polycondensation is a highly promising approach to solve this 30-year old problem.

Design of the model system

Our initial interest was based on the formation of a prototype of an “autocatalytic set”: a library of information containing polymers of different composition with a sufficient length to form catalytically active structures. To understand the basic principles of the library’s formation, we decided to work with a model system not directly related to the “RNA World”. We used chemically modified DNA-based oligonucleotides that are more stable in solution than RNA, easier to synthesize in large quantities and that do not form 2'-3'-bonds as a side product during polymerization. The target length of the polymers remained arbitrarily fixed to 30 bases, corresponding to the length set as a lower limit for the new ribozymes in RNA-based systems. Our approach was to design and study a model system, supposing that once basic principles governing the reaction are understood, these can be extended to other nucleic acid based polymers.

Analyzing the work done in this area and discussed above, we realized that it is easier to work with short oligonucleotides as a starting material than with mononucleotides.

The shortest known DNA-based self-replicating system consists of the CGCG tetranucleotide formed by ligation of two dinucleotides [22]. We chose dinucleotides as minimal building blocks that allow rate enhancement due to templating. Considering the differences in the efficiency of incorporation of purine and pyrimidine nucleotides [23], we used dimers with fixed 5'-pyrimidine-3'-purine composition, assuring that both types of the nucleotides are presented in the strand. The purine at the 3'-end and pyrimidine at the 5'-end provides stable purine-pyrimidine stacking between two reacting molecules, offering advantages for the system in the equilibrium step prior to polycondensation. Other types of base-pair stacking are less stable [24].

To design the model system, we used amino-modified DNA in presence of EDC as a condensing agent, to ensure fast and efficient ligation. 3'-amino-2', 3'-

dideoxynucleotides have been used as model systems by the groups of *Orgel* [18], *Richert* [25] and *Szostak* [26]. After analyzing *Orgel's* work, we concluded that non-templated cyclization of the single strands can be reduced if the rate of ligation to elongate the double strands is enhanced. It has been previously reported by *Shabarova et al.* [27] that EDC-driven ligation of a phosphate group with amine at the 3'-end is 75 times faster than the ligation of the phosphate group with the hydroxyl, while an amino group at the 5'-end accelerates this reaction 100 times. The study of self-replicating molecules by *von Kiedrowski et al.* [28] had shown that the difference between 3'-amino and 5'-amino was increased almost 20 times in presence of template molecules. Our model system is based on 5'-amino-3'-phosphate terminated short oligodeoxynucleotides.

The polymerization of four possible 5'-aminopyrimidine-3'-purine phosphate dinucleotides was studied: nCGp, nTAp, nCAp and nTGp. The general reaction describing the system is shown in the Figure 1. Several experiments with tetra and hexanucleotides as a starting material have been additionally performed in order to understand the reaction mechanism.

Oligonucleotides longer than 100 bases are produced within hours when the building blocks are complementary or self-complementary. The goal of the study is to understand the reaction mechanism and explore the efficiency of the reaction for nucleotides of different length and composition. Qualitative characterization of reaction provides a better understanding of the reaction mechanism.

The solution of the oligomerization problem allows us to model the origin of the “RNA World” in the laboratory and can bring us one step further in understanding of the mechanisms of the transition from simple organic molecules to living organisms.

Results and Discussion

Polycondensation of dinucleotides

5'-amino-3'-phosphate dinucleotides, tetranucleotides and hexanucleotides have been used as a starting material for the polycondensation reaction. Reactions have been performed at 2°C in HEPES buffer pH 7.5 in the presence of EDC as a condensing agent. Table 1 summarizes the sequences of all the compounds used in the experiments. The course of the reactions was followed by HPLC, MALDI and denaturing polyacrylamide gel electrophoresis (PAGE).

Initially, the polycondensation of non self-complementary building blocks was studied. nCAp and nTGp were left to react in the presence of 0.4 M EDC for 72-100 h (Figure 2A-B). Both dinucleotides reacted similarly. At 2°C and 20 mM dinucleotide concentration, several peaks were observed by HPLC analysis after 24 h of

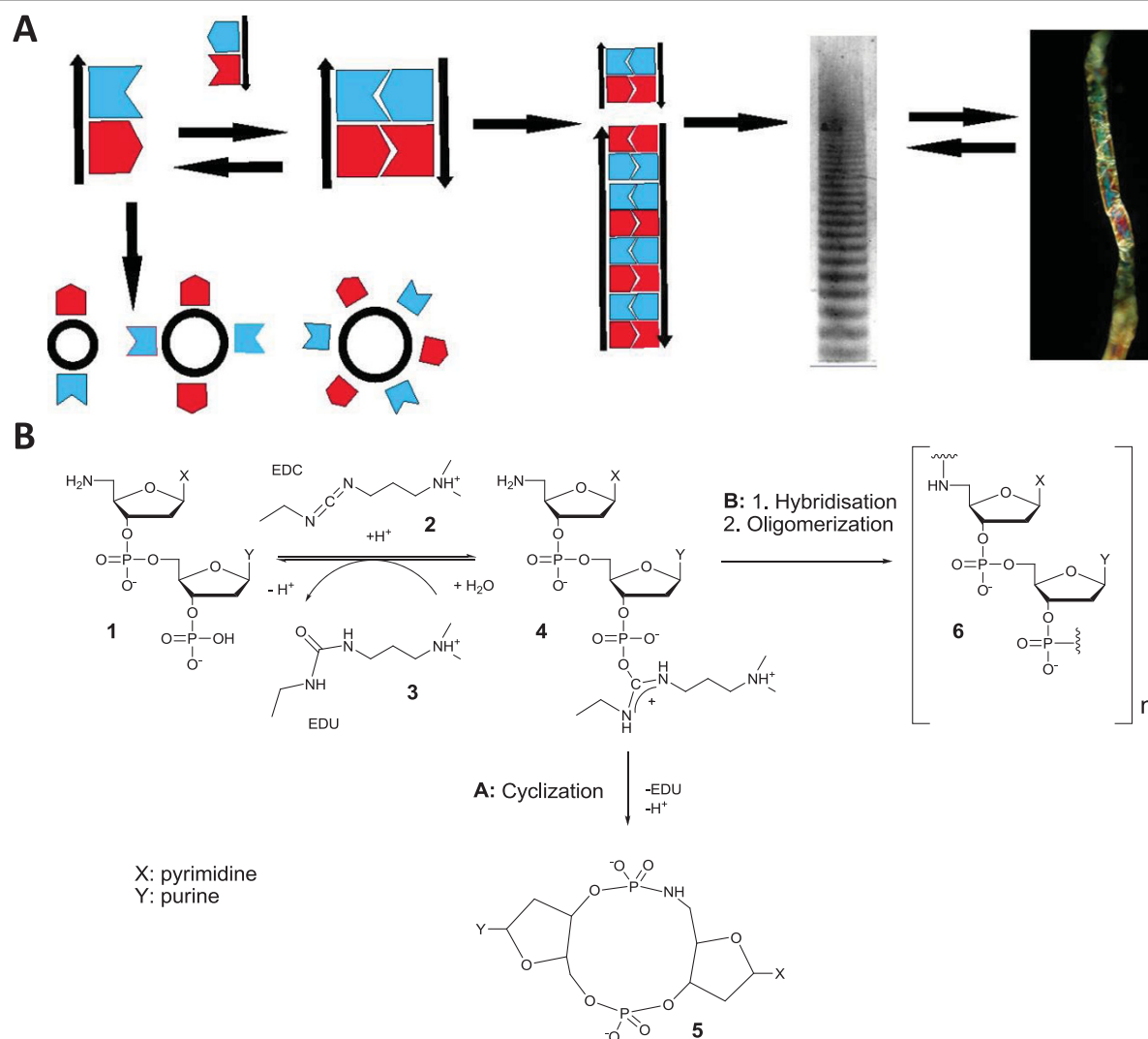


Figure 1 Reaction mechanism. (A) Schematic representation of the general reaction mechanism. (B) Detailed reaction mechanism. 5'-amino-3'-phosphate pyrimidine-purine dinucleotide (**1**) is activated by EDC (**2**) binding to phosphate group with uptake of one proton. EDC (**2**) in water is partially hydrolyzed with release of the proton and EDU (**3**) formation. Single stranded activated molecules (**4**) undergo pathway A, forming an inactive cyclic side product (**5**) and releasing EDU and proton. At high concentrations and low temperatures where duplex formation is possible, polycondensation and chain growth via pathway B takes place, leading to the product (**6**).

reaction and the conversion was complete while no further change in the spectrum was detected after 72 h or 100 h of reaction. The MALDI analysis of the reaction mixture showed that all the starting dinucleotides were consumed to yield several products not suitable for further polymerization. Among them the non reactive

EDC adduct (identified by single MALDI peak with a mass $M+155$, where M is a mass of a dinucleotide, reported in the Table S1 [Additional file 1]), cyclic 4-6-mers (identified by the $M-18$ peaks) and cyclic guanine EDC-adducts (identified by the $M+133$ peaks). Similar compounds were observed by R  thlingsh  fer and Richert in their study of the reaction of 3'-aminonucleotides with EDC [25]. The complete tables of registered peaks are presented in the Supplementary Information [Additional file 1]. The results are similar to those obtained by other investigators [12,25].

Reactions with self-complementary building blocks were studied next. At 2  C and the 20 mM nCGp

Table 1 Compounds used in the study.

Type of nucleotide	Sequence
Dimer	nCGp, nCAp, nTGp, nTAp
Tetramer	nTGCAp, nCATGp, nGTGTp
Hexamer	nTGTACAp

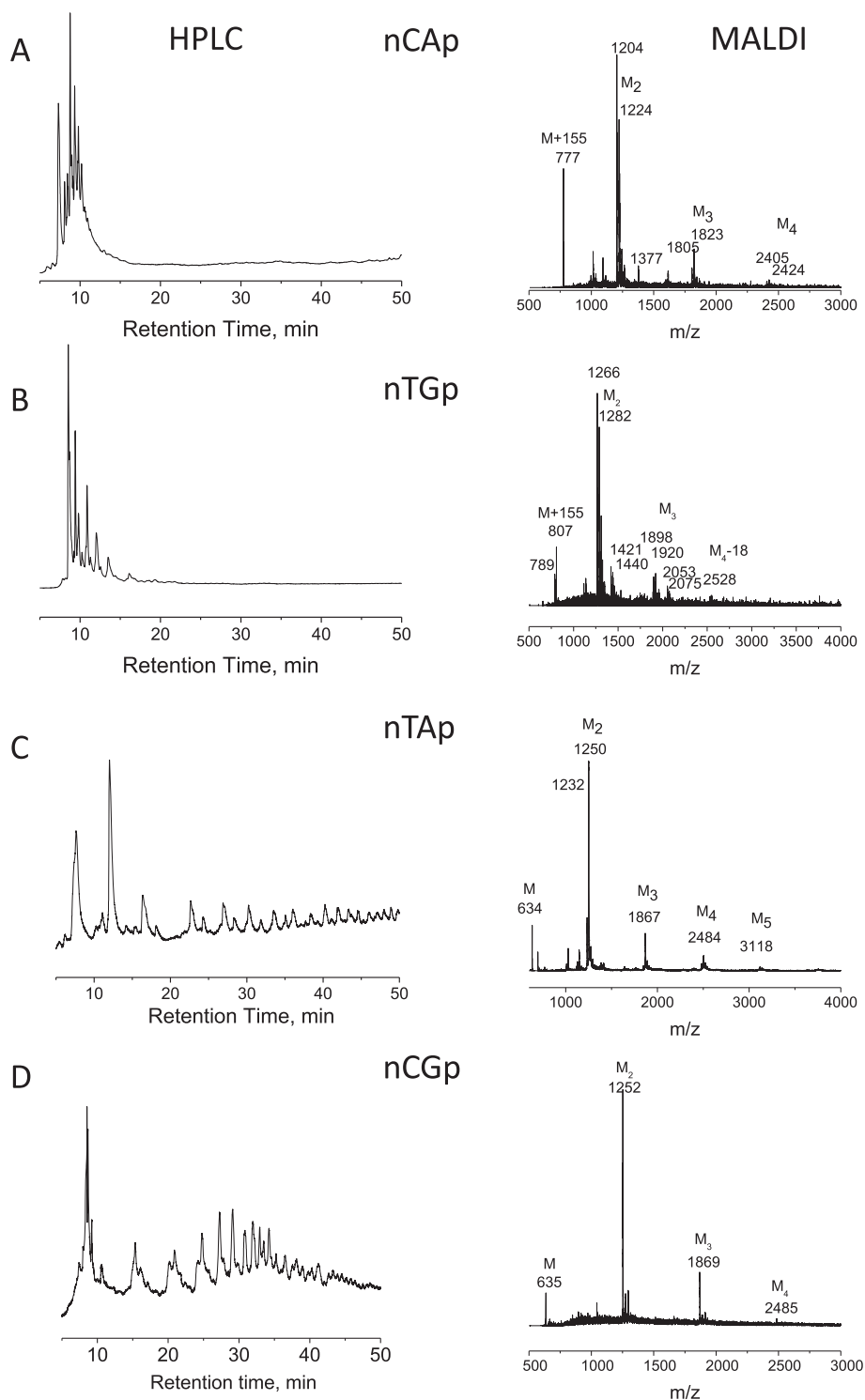


Figure 2 Polymerization of dinucleotides. IE HPLC (left) and MALDI (right) plots. In a MALDI plot, masses of dimers, tetramers, etc. are labeled as M , M_2 , M_3 etc. initial conditions in all the cases are: 20 mM dinucleotide, 0.4 M EDC, 0.1 M HEPES, pH 7.55, 2°C. (A) nCAp polycondensation after 24 h. EDC adducts and cyclic products are observed. (B) nTGp polycondensation after 24 h. EDC adducts and cyclic products are observed. (C) nTAp polycondensation after 24 h. The amounts of EDC adducts and cyclic products are reduced and longer oligomers are observed. (D) nCGp polycondensation after 4 h. Apart from the oligomers M_n no other side products are detected.

concentration, rapid polymerization was observed. Products up to 20 bases were registered by HPLC (Figure 2C). Longer products formed already after 4 h of reaction were observed by gel electrophoresis (Figure 3A). After 48 h of reaction, practically all the dimers were consumed, strands 140 bases long were formed and the chain stopped growing (Figure 3A and Figure SI1 in

[Additional file 1]). The addition of fresh EDC or dinucleotides did not change the final length of the chain. MALDI spectrum showed the absence of the EDC adducts and cyclic products, and a formation of M_n oligonucleotides (Figure 2D).

nTAp also polymerized. Peaks with high elution times were observed by HPLC and identified as oligomers by

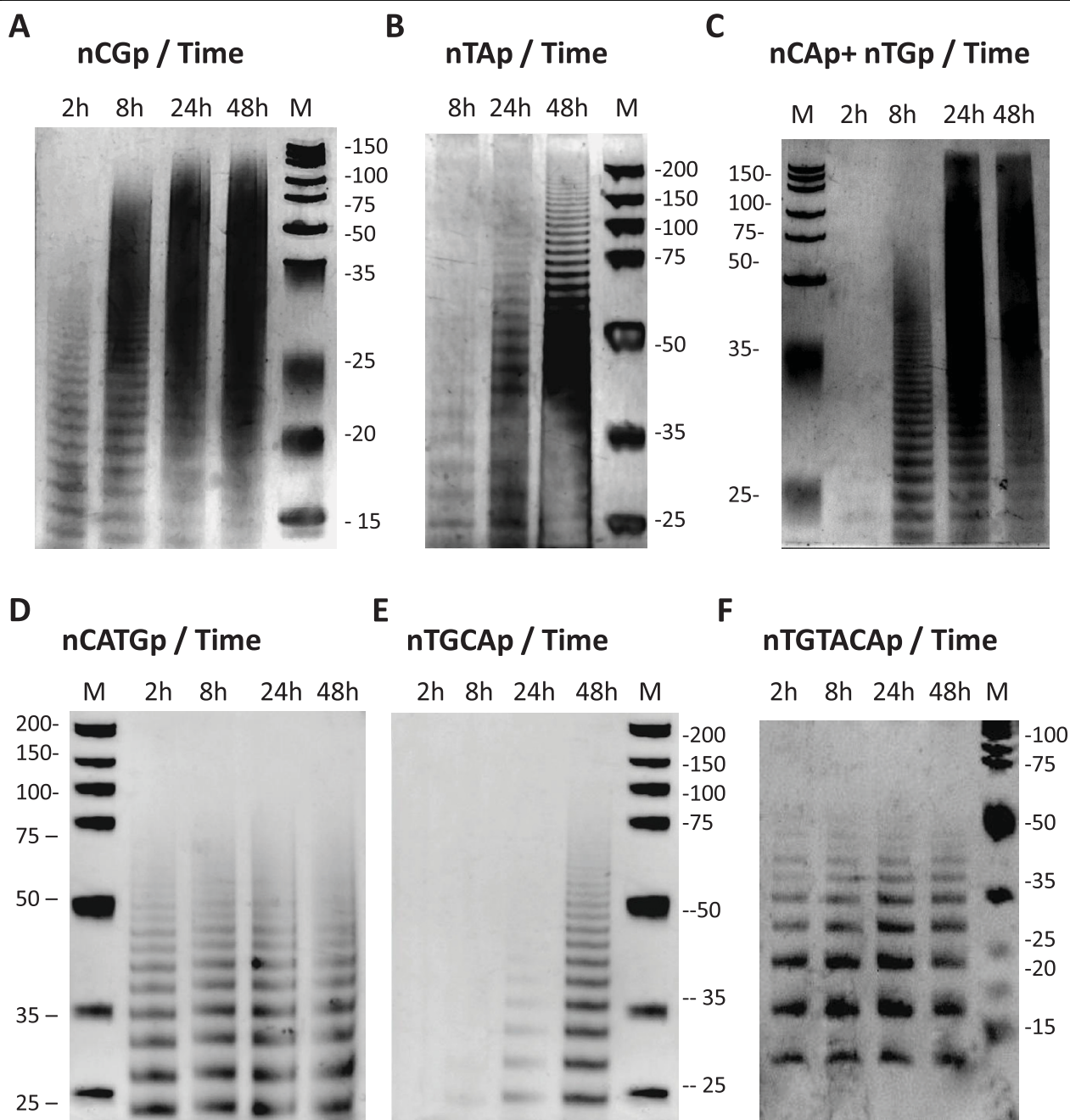


Figure 3 20% PAGE analysis of polycondensation products. Initial conditions in all the cases are: 20 mM dinucleotide, 0.4 M EDC, 0.1 M HEPES, pH 7.55, 2°C. All gels are stained with SybrGold® gel stain for 20 min. Line M is a marker line. (A) nCGp, (B) nTAp, (C) nTAp and nCap copolymerization, (D) nTGCAp, (E) nCATGp, (F) nTGTAcap.

MALDI. TA duplexes can form only two hydrogen bonds for each base pair, instead of the three bonds formed by the CG dimer. Apparently, under the reaction conditions, only a fraction of the nTAp dinucleotides was presented as a duplex, and is able to undergo polycondensation. Due to the lower duplex stability part of the nTAp dimers remained single stranded and cyclization and EDC adduct formation were detected (Figure 2C). The observed reaction rates of nTAp polymerization were much lower than in the case of nCGp. As registered by PAGE, strands longer than 100 nucleotides were formed only after 48 h of reaction (Figure 3B).

The equimolar mixture of nCap and nTGp (complementary to each other) gave strands longer than 300 bases long after 24 h of the reaction (Figure 3C). Even at early reaction stages a complex mixture of the products was formed in this case and for this reason the course of the reaction was analyzed only by PAGE. Again, as in the case of nCGp polycondensation, once achieved, the maximal product length remained constant.

The polycondensation reaction was observed only when the building blocks were complementary and able to form double strands. However, the dimers are too short and cannot form double strands in a cooperative way, as longer oligonucleotides do. The double strand formation of the dinucleotides is in simple equilibrium with single strands in solution. Duplex formation is favored at high concentrations of the dinucleotides and low temperatures. The consumption of the double strands by polycondensation reaction displaces the equilibrium towards the duplex formation.

Polycondensation of tetranucleotides and hexanucleotides

Our first approach toward understanding the reaction mechanism suggested that the Watson-Crick base pair recognition may play an important role. In this case, short concatenated single stranded fragments could appear at the growing chain ends and then react with each other by overlapping "sticky ends". To prove the relevance of the proposed mechanism, we switched from dimers to tetramers and hexamers. The fraction of the double stranded complexes increases with the growth of the strand, which will change the observed reaction behavior and can allow estimating the relationship between the double strand and single strand ligation mechanisms.

For the tetramer polymerization three different strands were used: non-self-complementary nTGTGp and self-complementary nTGCAp and nCATGp. Polymerization of nTGTGp, monitored by PAGE, showed only traces of short oligonucleotides (less than 20 bases long) after several days of reaction. Polymerization of

self-complementary building blocks of nTGCAp or nCATGp led to the formation of ~60 bases long products (Figure 3D-E). The two tetramers have the same CG-content, but different bases at the ends, resulting in different energies of stacking between reacting building blocks. The reaction rate of nCATGp polymerization (where at 5'-3' direction the G from one building block comes in contact with C from another) is much faster than the reaction of nTGCAp (where A and T from the ends of two building blocks are interacting).

The self-complementary hexamer nTGTACAp reacts faster than the tetramers (Figure 3F). If one compares the reaction rates among dimer nTAp, tetramer nTGCAp and hexamer nTGTACAp that have the same base pairs at the ends, it can be noticed that the polycondensation reaction rate increases with the size of the building block. One of the reasons for this change may be a larger fraction of the reactive duplexes in solution related to the increase of the length of the starting oligonucleotides.

Copolymerization of the complementary dimers with corresponding tetramers showed little template effect from tetramers. When nTGTGp was copolymerized with nCap, the chain grew with a step $2 \times n + 2$, indicating that there is no significant difference between the ligation rates of two double strands and the ligation of two dimers on the tetrameric template (Figure 4, Figure SI2 [Additional file 1] and Table S6 [Additional file 1] for MALDI analysis).

Main trends in polycondensation reactions

Final product lengths, estimated from gel electrophoresis for different combinations of nucleotides used in the study, are reported in Table 2.

The rates of the polycondensation reactions in all studied systems varied in the following order:

nCGp > nCap + nTGp > nTAp >> nTGp \approx nCap, for dimers

nCATGp > nTGTGp + nCap > nTGCAp >> nTGTGp, for tetramers

In both series, the results correlate with the stacking energies between the building blocks [24] and not with the stability of the double strands (which, for example, is expected to be lower for the ternary complex formed by the tetramer and two dimers than for the duplex formed by two tetramers). The results suggest that the stacking interactions are the main driving force of the reaction and the Watson-Crick base pairing seems to play a minor role in the model system.

The results can be understood better if one considers the work done by the *Frank-Kamenetskii's* group, who found that stacking, and not base pairing, makes a major contribution to the stability of the double strand

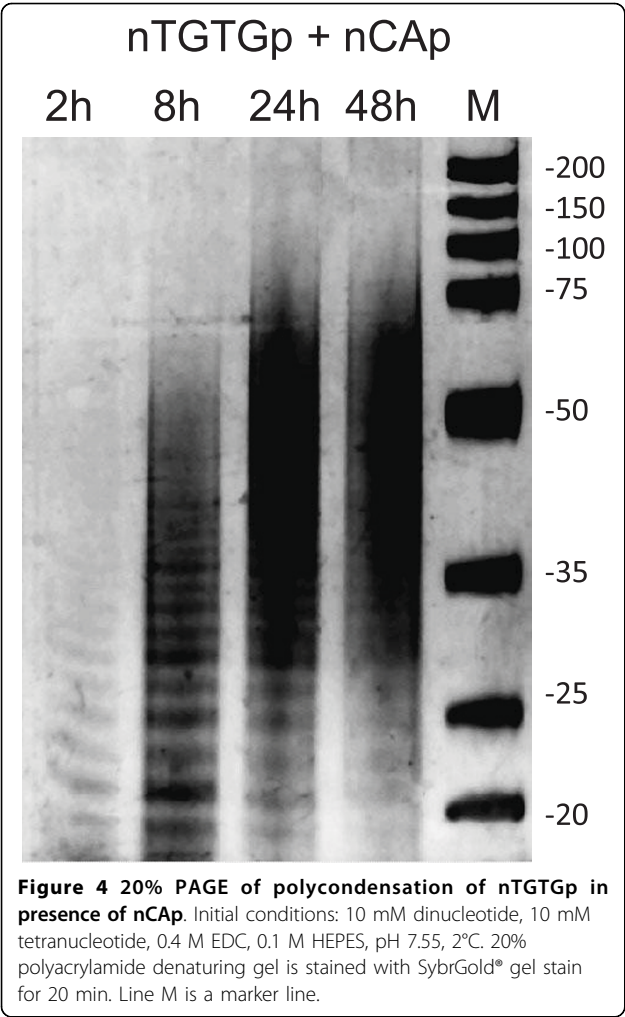


Table 2 Estimated length of the reaction products.

Nucleotide combination	Estimated length, bp
nCGp	80 - 150
nTAp	200
nCGp + nCAp (1:1)	50
nCGp + nTGp (1:1)	50
nCGp + nTAp (1:1)	60
nCAp + nTGp (1:1)	> 300
nCAp + nTGp + nTAp (1:1:1)	70
nCAp + nTGp + nCGp (1:1:1)	> 150
nCAp + nTGp + nTAp + nCGp (1:1:1:1)	50
nCATGp	60
nTGCAp	60
nTGTGp + nCAp (1:2)	70
nTGCAp + nCAp + nTGp (1:1:1)	150
nGTACAp	54

in solution [29]. The ligation between blunt-ended duplexes is the main reaction mechanism. This may explain why the reaction is, in principle, possible, but it does not explain the tendency towards long strand formation.

Criteria for dimer polycondensation

Other condensing agents have been used but EDC resulted to be the only efficient one (some shorter products were formed in case of EDC with 1-methylimidazole additive and no products were observed in cases of BrCN or 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT MM) condensing agents). It is not clear if EDC has a special function or it just acts under optimal reaction conditions (neutral pH, medium ionic force, low temperature, longer stability in solution as compared to other condensing agents used in the study). EDC variation has been performed and it was observed that while the reaction rate varied with the EDC concentration, after 100 h of the reaction time all the products achieved the same final length. We would suggest that EDC is specific to chosen model system but only due to its maximum reactivity under optimal reaction conditions and not because of its involvement in the later reaction stages.

Under prebiotic conditions, pure solutions where 1:1 pairing could have occurred were highly improbable. A variety of contaminating sequences and small organic molecules were present, together with the short nucleic acid fragments that could polymerize. To evaluate how robust the system is, polymerization in the presence of non-complementary building blocks was studied.

Either nCAp or nTGp was added to nCGp, thus keeping the total dinucleotide concentration constant and equal to 20 mM. In both cases, identical behavior was observed. Polymerization proceeded efficiently and the length of the products remained almost constant for up to 50% of the “contaminating” sequences present in solution. At higher concentration of non-complementary dimers, the main product started to become shorter. The growth stopped completely when the nTGp (or nCAp) concentration was higher than 80% of the total nucleotide concentration (Figure 5A-B).

Reactions in mixtures with unbalanced stoichiometry have shown that the polycondensation observed in the model system is tolerant to side reactions. As previously noticed in the study of one-component systems, the final length of the product depends on the composition of the starting material and thereof of the strands that are formed. The constant length maintained by the products over the wide range of concentrations suggests that the polymerization between double strands excludes side reactions, where single strands are spontaneously ligated to double strands.

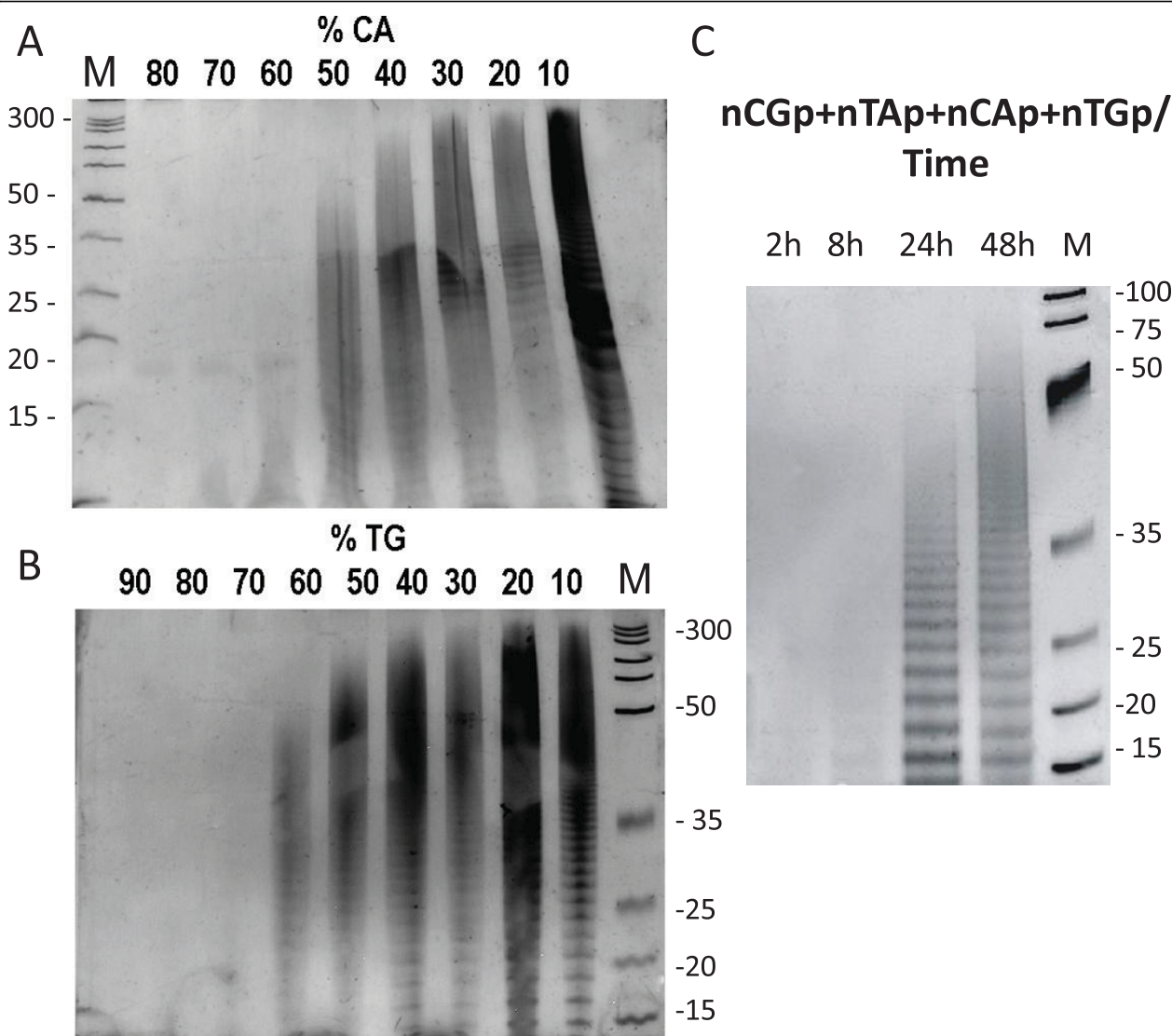


Figure 5 Polycondensation in mixtures. Total nucleotide concentration is 20 mM. 20% polyacrylamide denaturing gel is stained with SybrGold® gel stain for 20 min. Line M is a marker line. (A) nCGP polymerization in presence of nCAp. Ratio of the two dinucleotides varies as shown in the picture. Reaction is analyzed after 48 h of reaction. (B) nCGP polymerization in presence of nTGp. Ratio of the two dinucleotides varies as shown in the picture. Reaction is analyzed after 48 h of reaction. (C) nCGp + nCAp + nTG + nTAp at 1:1:1:1 ratio gives strands approximately 50 bases long. All four dinucleotides are incorporated.

When all four types of dimers were left to polymerize together in 1:1:1:1 ratio, the complementarity conditions were fulfilled for every type of building block and oligonucleotides of at least 50 bases long were formed (Figure 5C). All building blocks were consumed, indicating that the product was the library of mixed sequences. Further study to reveal the catalytic potential of these spontaneously formed libraries is needed.

The temperature dependence for reaction of nCAp with nTGp showed that the rate is almost the same at 2°C and 10°C. At 20°C, only traces of long products

were observed and at 30°C no oligomers were detected (Figure 6A). At high temperatures, all building blocks existed as monomeric species and only cyclization occurred, as a more detailed study of nCGp oligomerization at 30°C has shown (Figure SI3 [Additional file 1]).

Ionic strength is known to stabilize double strand formation by neutralizing the negative charges on the phosphate backbone. The reaction proceeded at high ionic strength, tolerating Na⁺ concentration up to 0.5 M. At 0.15 M ionic strength, a maximum of the reaction rate was observed. At 1-3 M NaCl, no product

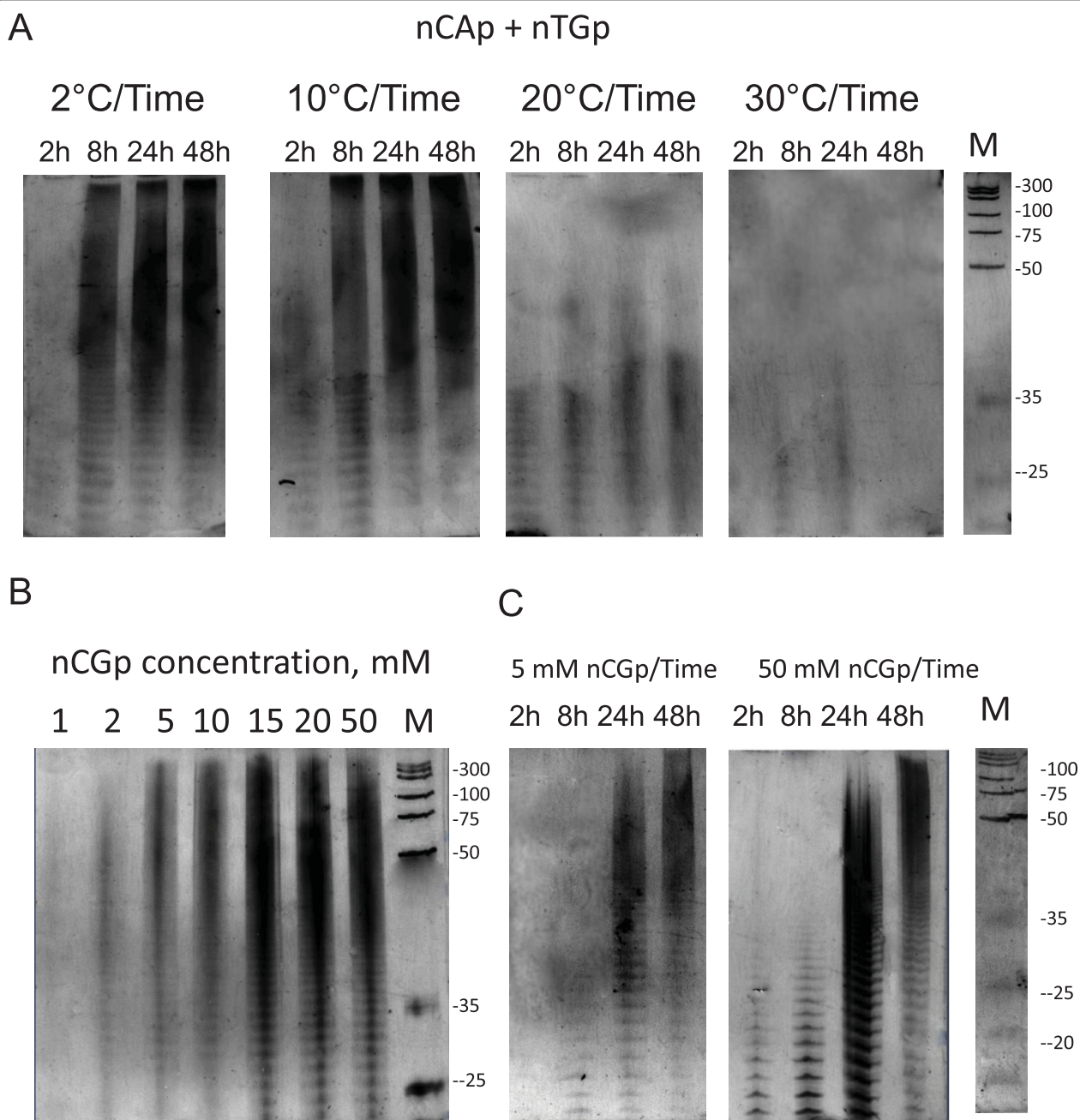


Figure 6 Temperature and concentration effect. 20% polyacrylamide denaturing gel is stained with SybrGold® gel stain for 20 min. Line M is a marker line. (A) Effect of the temperature on polycondensation efficiency of nTGp with nCAp in 1:1 ratio. Total nucleotide concentration is 20 mM. (B) Concentration effect on nCGp polymerization. Reaction products are visualized after 48 h of reaction. (C) Comparison of the reaction rates at different concentrations of the dinucleotide. 5 mM nCGp and 50 mM nCGp show similar reaction rates and size of the final product.

formation was detected (Figure SI4 [Additional file 1]). It has been previously observed that high ionic strength inhibits EDC-driven ligation. Probably this is the reason for the low efficiency of the reaction at high ionic strength, which, in theory, should stabilize the supramolecular DNA structures. Dications and transition metal ions, known to accelerate RNA ligation in solution, also

interact with EDC chemistry and inhibit the reaction (results not shown).

As mentioned above, at low nucleotide concentrations and elevated temperatures, cyclization rather than polymerization takes place. The transition has been studied systematically. The total concentration of the building blocks was varied between 1 - 50 mM. The polymerization

took place in the relatively concentrated solutions, but the final length of the products did not depend on the concentration of the monomers. The polycondensation is a second order reaction; however no significant change in the reaction rate has been observed when the concentration of the dinucleotides was varied from 2 to 50 mM. In the case of the nTGp with nCap mixture, 2 mM total dinucleotide concentration is a minimum necessary to start the growth; in the case of nCGp, the minimum is 1 mM (Figure 6B).

Phase transition during the course of reaction

In the presence of a "crowding agent" such as 30% polyethylene glycol PEG 6000 solution (which does not participate in the reaction but increases the local concentration of the nucleotides in solution), the formation of long strands is faster and, after several hours, only high molecular mass products were observed with gel electrophoresis.

PEG is known to induce liquid crystalline aggregation in DNA molecules. Droplets with ordered structure were observed under polarized light microscopy (Figure 7A). Similar structures in smaller amount appeared in reaction mixtures in the absence of PEG at late stages of reaction, where long strands were formed. Two types of structures could be identified: small droplets with high birefringence and thin ribbon-like crystals (Figure 7B-C). The growth of the crystals correlated in time with the formation of large oligonucleotide strands (Figure SI 5 [Additional file 1]). To assure that they contain DNA, Hoechst 33258 stacking dye was added to the reaction mixture 72 h after the reaction had started (final concentration of the dye in solution was 0.5 $\mu\text{g/mL}$). The dye stained ribbon-like structures and the droplets (Figure SI6 [Additional file 1]), proving that the structures were formed by double stranded DNA molecules and cannot be attributed to any side reactions or impurities.

Supramolecular aggregates of DNA in water are known for a long time. They can be characterized by circular dichroism (CD). If a liquid crystalline phase is formed, the spectrum becomes asymmetric, with intense bands between 250-300 nm [30]. To understand the structural changes responsible for the reaction, we analyzed the solution of dinucleotides before the reaction started and at the stage where long products had been already formed. Dinucleotides in absence of EDC showed a CD spectrum with bands between 250 and 300 nm at the concentration range used in the reactions (10-20 mM) and 2°C. This suggests that even short strands can form aggregates in solution.

Single non-complementary dinucleotides lost their internal structure after several days of reaction with EDC, when cyclic side products and EDC adducts were formed (Figure 8). The spectrum of a reaction mixture of complementary or self-complementary dinucleotides

after 72 h showed an intense band at 255 nm, similar to those, observed when liquid crystalline DNA domains are dispersed in solution (Figure 8) [30].

nCGp polymerization kinetics was monitored online because it polymerizes relatively fast, forming products of different length and with only one type of sequence. The appearance of a non-conservative spectrum with the minimum at 250 nm has been registered. The intensity of the negative band is almost 4 times greater than the positive part of the spectrum (Figure 8C-D). No lag period was observed during the reaction.

The solution of unreacted dimers lost the intensity of the CD spectrum peaks during heating to 30°C (Figure SI 7 [Additional file 1]). Non cooperative gradual melting was observed. Diluted solutions of reaction products (the 200 times dilution was necessary to reduce the EDC concentration that could damage the formed strands at high temperatures), showed a 2 stage cooperative melting curve with a mean melting temperature around 70°C (Figure SI8 [Additional file 1]).

In living cells, DNA is packed in dense and highly ordered structures. Self-organization of DNA in solution into ordered supramolecular structures beyond the double helix has been known for more than 50 years. At high concentrations, DNA molecules form a cholesteric liquid crystalline phase, where the molecules are organized horizontally into pseudoplanes that are twisted with respect to each other. At lower concentrations, and in the presence of condensing agents, such as polyamides, DNA forms small hard spheres known as spherulites. Experiments on DNA aggregates have mostly been performed with DNA molecules of ~150 bp, which is approximately the percolation length for double stranded DNA. Recently, it was shown that 6-20 bp long double strands with identical sequences can form liquid crystalline domains, in which the short fragments are held together by stacking forces. This phenomenon has been called "physical polymerization" and has been described for different types of double stranded DNA and RNA molecules [21]. In the cited work the authors suggested that, in the presence of an efficient condensing agent, this pre-aggregation should lead to long strand formation. It seems to be the case in our system.

Aggregation of DNA strands with identical sequences into separate liquid crystalline domains [31] has several interesting features that could be relevant for the prebiotic chemistry. First, with the growth of the chain, the number of charges on the formed polyelectrolyte increases, thus increasing the attraction between the strands. This can shift the reaction equilibrium toward the formation of long strands, once a minimal threshold length is achieved. Second, the recognition of homolog double strands by unique ion distribution in the major groove of the helix [32] suggests that, under conditions

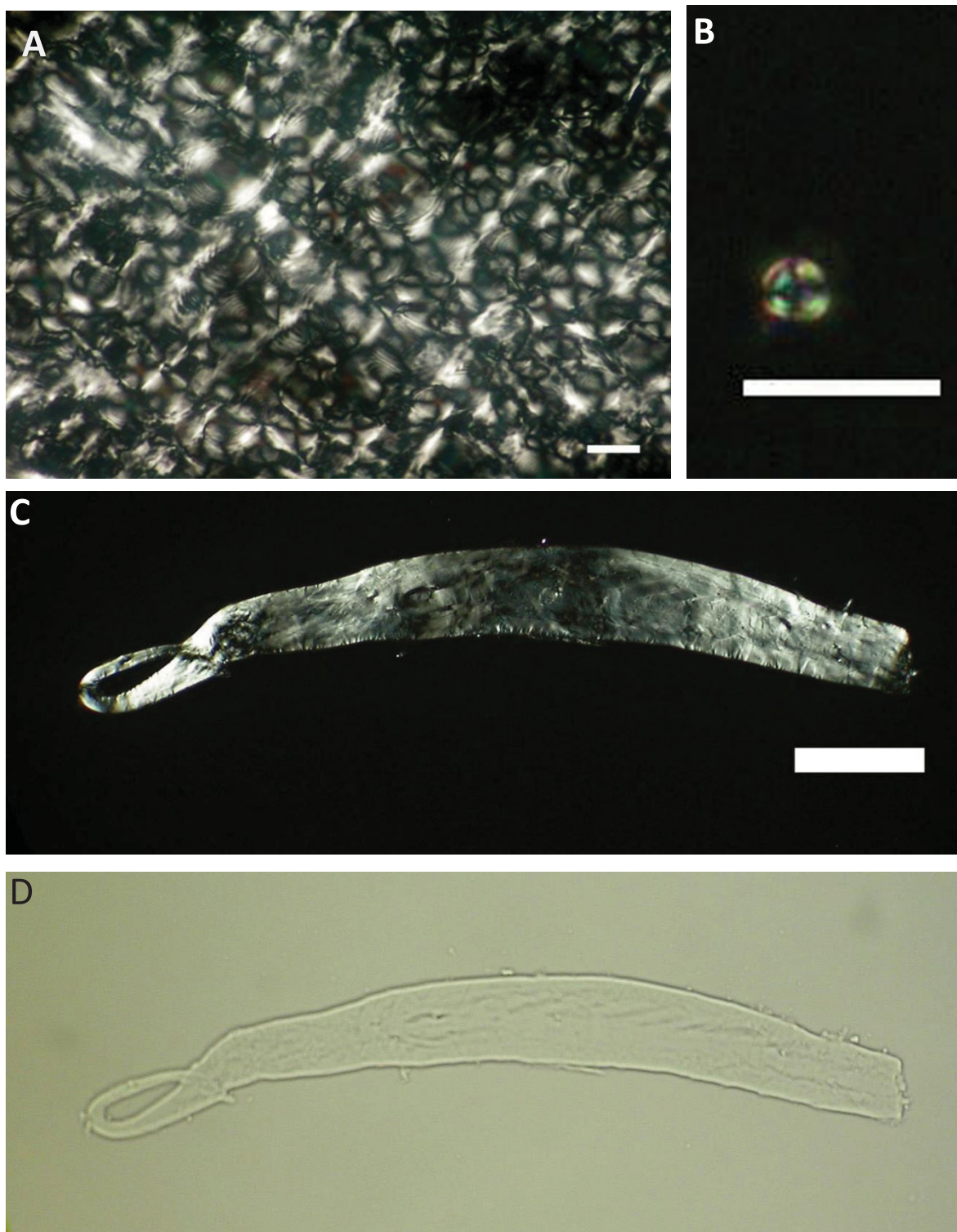
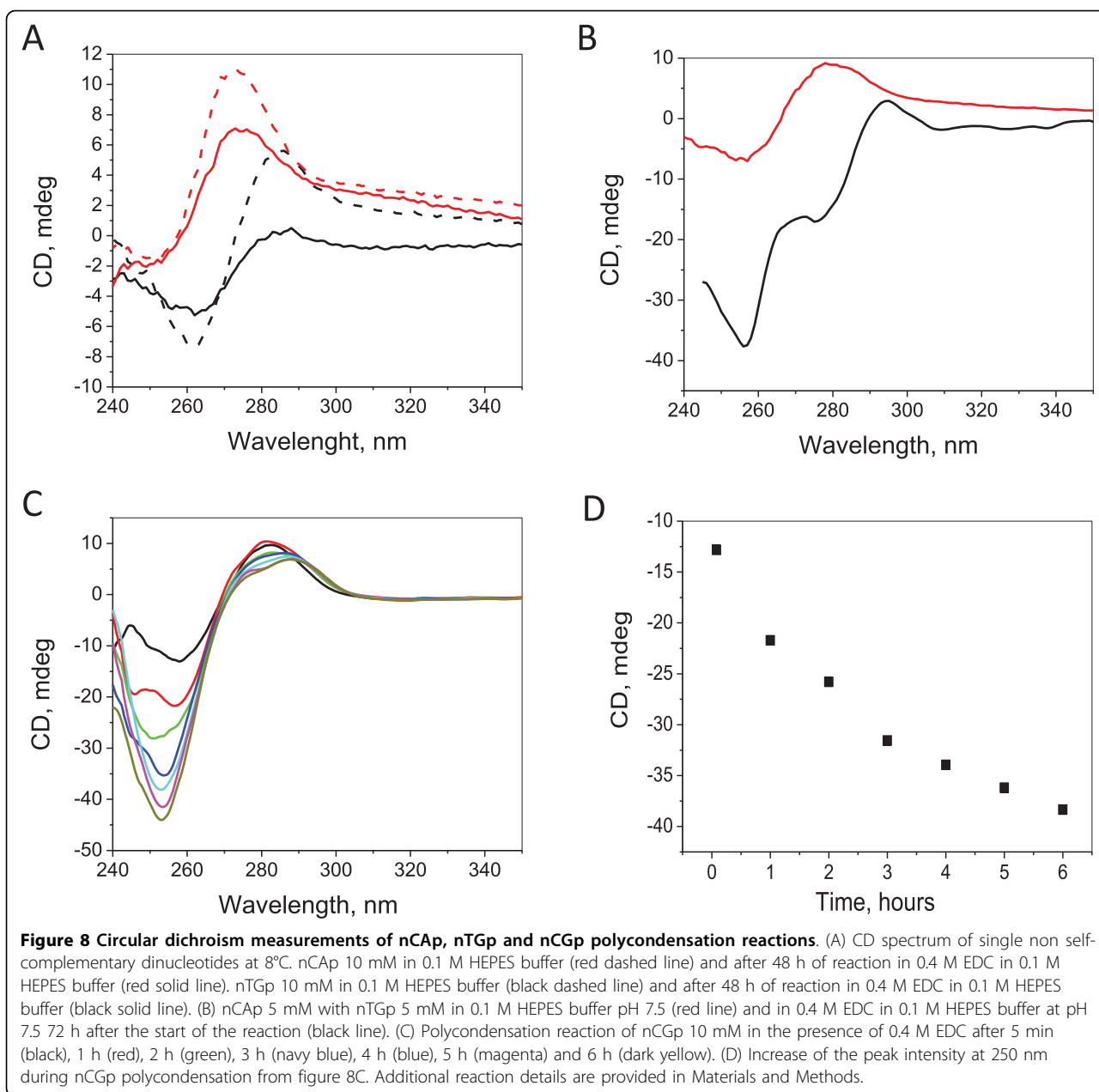


Figure 7 Liquid crystalline phase and crystals formed during the polycondensation of nCAp with nTGp. (A) Liquid crystalline domain formed by polycondensation of 5 mM nCAp with 5 mM nTGp in the presence of 30% PEG 6000 after 48 h of reaction. Scale bar is 10 μm . (B) Spherulites observed during nCAp polycondensation after 48 h of reaction. Scale bar is 10 μm . (C) Ribbon-like crystals observed in the reaction mixture at late reaction stages (here 72 h) in polycondensation of 10 mM nCAp with 10 mM nTGp under polarized light (the length of the scale bar is 50 μm). (D) The same crystal under the white light.



that allow the formation of such aggregates, marked similarity in the sequences of synthesized strands may be present. If, upon the formation of long oligonucleotides, the strands of the same or similar sequences can aggregate, the reaction rate can accelerate due to the increased local concentration. Such increase in the production of the strands with the same sequence can be considered as rudimentary selection in prebiotic systems. Currently, there is insufficient experimental evidence to demonstrate conclusively that this pathway is truly autocatalytic.

Discussion

Presumably, different mechanisms operate at different reaction stages and the reaction products are the results of a combination of polymerization by double strands and self-organization. When dimers and short nucleotides are present in solution, polycondensation due to reduced cyclization is the major reaction pathway. When longer strands are formed, supramolecular interactions among them increase. Here, the DNA aggregates can be seen, abstractly, as the surface used to increase the local concentration of activated building blocks.

This can displace the equilibrium towards the preferential formation of longer products.

The model system correlates with the work done by the groups of *Orgel* [18] and *Hud* [20]. In their systems, the decrease of the cyclization, resulting from the formation of double strands, leads to polymers formed by ligation of 10-20 building blocks. The results can be explained in terms of stepwise polymerization.

Orgel's system is closely related to our experiments, in spite of the different final results. The reaction conditions were the same in both cases: 2°C and 20-50 mM total dinucleotide concentration; the same type of condensation chemistry in presence of EDC was applied; the same base sequences were used. The main difference between the experiments leading to the different final results was the use of 5'-phosphates in case of *Orgel* and 3'-phosphates in the present work. In the work by *Zielinsky and Orgel* pCGn gave cyclic dimers as a main reaction product while in our case the reaction proceeded to the formation of long strands.

The formation of the cyclic product is probably more favored in the case of pCGn than for nCGp, due to the better bond rotation of the end phosphate group at 5'-position. The use of 3'-phosphate groups in the experiment does not contradict conditions expected to exist on the prebiotic Earth. In nucleoside phosphorylation experiments, both 3'- and 5'- phosphate nucleotides have been formed [12]. 3'-phosphates are the main reaction products in prebiotic purine nucleotide synthesis, as shown by *Powner et al.* in 2009 [2], though the real significance of this coincidence is hard to judge at the current stage of research.

The main difference between our and previously reported systems emerges when relatively long stands are already formed in solution. In our experiments, the concentration of nucleotides is below the reported value at which short strands of natural DNA start to condense [21]. Probably this effect is induced by the use of low temperatures and amino modified nucleotides, which possess lower solubility than non modified DNA (with 5'-amino dinucleotides the observed solubility limit is at a concentration of 50 mg/mL). *Gryaznov et al.* have compared structural and physicochemical properties of 3'-amino, 5'-amino modified and natural DNA oligonucleotides [33]. They observed that, at the single strand level, all the chains are isoelectronic, and have similar conformation, bond length and angle. An amino group is more hydrophilic than an ester bond in natural DNA. It slightly increases solvation of chemically modified DNA at the single strand level. When double strands are formed, the 3'-amino group is more exposed to water than the 5'-amino group. The reduced solvation reduces the general solubility of the 5'-amino modified double strands. Another factor that may facilitate the

aggregation of oligonucleotides at relatively low concentrations is the use of EDC. EDC can be seen as a polyamine condensing agent that neutralizes one of the negative charges on the phosphate backbone.

Conclusions

The formation of oligonucleotides from small precursors is a long-standing problem in prebiotic chemistry. We have shown that at low temperatures and high concentrations short complementary oligonucleotides can polymerize and give long products within hours. Different building blocks can be incorporated into the strand, promoting the formation of combinatorial libraries of oligonucleotides long enough to be folded into specific catalytically active structures and to potentially form initial autocatalytic sets. Formation of supramolecular aggregates during the polymerization reaction triggers the long strand formation and leads to high yields.

Environmental conditions possibly provided numerous situations where concentrated solutions of nucleotides and small organic molecules could have existed. It is commonly believed that long RNA strands may have required millions of years for their formation. Probably this is not true. If long strands can be formed in a fast and efficient way, we can suggest that the origin of life is not an isolated accident under very specific conditions, but rather a common form of self-organization of complex organic molecules.

Experimental

Materials

The synthesis of 5'-amino- 3'-phosphate dinucleotides is reported elsewhere [5]. Tetra- and hexanucleotides were produced by machine-assisted synthesis; dinucleotides were synthesized in solution. All the building blocks were purified by preparative HPLC, lyophilized, diluted with double distilled water, filtered and stored at -20°C.

HPLC

Reverse phase and ion exchange chromatography were performed with Kontron *HPLC equipment*. Detection of the peaks was done at 254 nm. RP HPLC was carried out with 4 × 250 mm Nucleosil 120-5 C18 and VP 250/10 Nucleodur 100-5 C18 ec columns with 0.1 M NH₄HCO₃ in H₂O/CH₃CN gradient (see SI for the details of the gradient). IE HPLC was performed with a column Mini Q anion exchanger from Amersham Pharmacia; NaClO₄ gradient from 0 to 1 M at pH 12.5 was run.

UV

Determination of the concentration of the stock solutions was carried out with a *UV Spectrophotometer Cary 1E* from *Varian* using a 1 cm cell at 254 nm.

Polyacrylamide gel electrophoresis

BIORAD® Mini-PROTEAN Tetra Electrophoresis System was used; the gels were viewed on BIORAD® Fluor-S MultiImager. 5 µL of the diluted probes were mixed with 5 µL of denaturing INVITROGEN® loading dye solution. Aliquots with 50-100 ng of the nucleotides were loaded into 20% denaturing polyacrylamide gel (0.5 mm thick), run for 50 min at 55 V/cm voltage and stained for 20 min with SYBR Gold® nucleic acid gel stain (10000X in DMSO, Molecular Probes®). To estimate the length of the product Ultra Low Range DNA ladder by Fermentas® was used as a marker with a strands of 300, 200, 150, 100, 75, 50 (the most intense band), 35, 25, 20 and 15 nt.

MALDI

Equipment: *Autoflex II*-Massenspektrometer from *Bruker* with laser: N2 (337 nm). Method: 10 µL of sample were treated with Dowex 50WX8-200 cation - exchange resin in ammonium form, mixed with 5 µL of matrix solution (0.3 M of 2,4,6-trihydroxyacetophenone in MeOH mixed with 0.1 M diammonium citrate in water in proportion 2:1), 1 µL aliquot of the mixture was taken and allowed to dry in a steel target plate.

Circular dichroism

Samples were measured at a JASCO J-710 spectropolarimeter with temperature control. 0.01 cm and 1 cm quartz cells were used. 0.1 M HEPES buffer pH 7.5 with 0.4 M EDC was used as a baseline. The spectra were taken at a speed of 100 nm/min with 5 spectra accumulation.

Microscopy

The images were produced from a LABORLUX 12 POL optical microscope with linearly polarized filter and a Nikon Coolpix 4500 camera. 1 µL of the reaction mixture was placed under the object glass, allowed to equilibrate for 30 min and visualized under polarized light. For fluorescent microscopy a Zeiss Microscope LSM 510 was used with excitation at 450-480 nm and emission at 500 nm. The stock solution of the Hoechst 33258 stacking dye was added to the reaction mixture to achieve a 0.5 µg/mL final concentration.

General Methods

A desired amount of the starting nucleotides from the stock solutions were mixed in the 0.5 mL Eppendorf tube and lyophilized. Reactions were initiated with 6-10 µL of ice- cold filtered 0.4 M EDC solution in HEPES buffer (0.1 M, pH 7.5, 0.05 M Na⁺) and stored at 2°C (or another temperature, depending on the experiment) without stirring for 48-72 h, 1 µL was withdrawn at

indicated times and the reaction was quenched by 100 times dilution with 0.01 M HEPES buffer.

Additional material

Additional file 1: Supplementary information. Supplementary figures and tables.

Abbreviations

EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide; PAGE: polyacrylamide gel electrophoresis; CD: circular dichroism; HPLC: high pressure liquid chromatography; MALDI: Matrix-assisted laser desorption/ionization; bp: base pair; nt: nucleotide.

Acknowledgements

OT was supported by a DAAD Scholarship. We thank M. Wüstefeld and Dr. O. Medenbach for technical assistance and Dr. N. Takeuchi for comments and suggestions on the manuscript.

Authors' contributions

GvK, OTh and OT designed the experiments. OT, OTh and KA performed the experiments. OT and GvK wrote the paper.

Competing interests

The authors declare that they have no competing interests.

Received: 1 May 2010 Accepted: 18 August 2010

Published: 18 August 2010

References

1. Gilbert W: The RNA World. *Nature* 1986, **319**:618.
2. Powner MW, Gerland B, Sutherland JD: Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 2009, **45**:239-242.
3. Fuller WD, Sanchez RA, Orgel LE: Studies in prebiotic synthesis VI. Synthesis of purine nucleosides. *J Mol Biol* 1972, **67**:25-33.
4. Lincoln TA, Joyce GF: Self-sustained replication of an RNA enzyme. *Science* 2009, **323**:1229-1232.
5. Sievers D, von Kiedrowski G: Self-replication of complementary nucleotide-based oligomers. *Nature* 1994, **369**:221-224.
6. Eklund EH, Szostak JW, Bartel DP: Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* 1995, **269**:364-370.
7. Takeuchi N, Hogeweg P: Evolution of complexity in RNA-like replicator systems. *Biology Direct* 2008, **3**:11.
8. Turk RM, Chumachenko NV, Yarus M: Multiple translational products from a five-nucleotide ribozyme. *Proc Natl Acad Sci USA* 2010, **107**:4585-4589.
9. Kauffman SA: Autocatalytic Sets of Proteins. *J Theor Biol* 1986, **119**:1-24.
10. Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR: A self-replicating peptide. *Nature* 1996, **382**:525-528.
11. Vidonne A, Philp D: Making molecules make themselves - the chemistry of Artificial Replicators. *Eur J Org Chem* 2009, **5**:593-610.
12. Orgel LE, Lohrmann R: Prebiotic chemistry and nucleic acid replication. *Acc Chem Res* 1974, **7**:368-377.
13. Sleeper HR, Lohrmann R, Orgel LE: Formation of the Imidazolides of dinucleotides under potentially prebiotic conditions. *J Mol Evol* 1978, **11**:87-93.
14. Constanzo G, Pino S, Ciciriello F, Di Mauro E: Generation of Long RNA Chains in Water. *J Biol Chem* 2009, **284**:33206-33216.
15. Rajamani S, Vlassov A, Benner S, Coombs A, Olasagasti F, Deamer D: Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig Life Evol Biosph* 2008, **38**:57-74.
16. Ferris JP, Hill AR Jr, Liu R, Orgel LE: Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 1996, **381**:59-61.
17. Monnard PA, Kanavarioti A, Deamer DW: Eutectic phase polymerization of activated ribonucleotide mixtures yields quasiequimolar incorporation of

- purine and pyrimidine nucleobases. *J Am Chem Soc* 2003, **125**:13734-13740.
18. Zielinski WS: Oligoaminonucleoside phosphoramidates. Oligomerization of dimers of 3'-amino-3'-deoxynucleotides (GC and CG) in aqueous solution. *Nucl Acids Res* 1987, **15**:1699-1715.
 19. Bolli M, Micura R, Eschenmoser A: 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). *Chemistry & Biology* 1997, **4**:309-320.
 20. Horowitz ED, Engelhart AE, Chen MC, Quarles KA, Smith MW, Lynn DG, Hud NV: Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. *Proc Natl Acad Sci* 2010, **107**:5288-5293.
 21. Nakata M, Zanchetta G, Chapman BD, Jones CD, Cross JO, Pindak R, Bellini T, Clark NA: End-to-end stacking and liquid crystal condensation of 6-to 20-base pair DNA duplexes. *Science* 2007, **318**:1276-1279.
 22. Zielinski WS, Orgel LE: Autocatalytic synthesis of a tetranucleotide analogue. *Nature* 1987, **327**:346-347.
 23. Lohrmann R, Orgel LE: Self-Condensation of activated dinucleotides on polynucleotide templates with alternating sequences. *J Mol Evol* 1979, **14**:243-250.
 24. Gotoh O, Tagashira Y: Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers* 1981, **20**:1033-1042.
 25. Röthlingshöfer M, Richert C: Chemical primer extension at submillimolar concentration of deoxynucleotides. *J Org Chem* 2010, **75**:3945-3952.
 26. Rajamani S, Ichida JK, Antal T, Treco DA, Leu K, Nowak MA, Szostak JW, Chen IA: Effect of stalling after Mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J Am Chem Soc* 2010, **132**:5880-5885.
 27. Shabarova ZA, Dolinnaya NG, Drutsa VL, Melinkova NP, Purmal AA: DNA-like duplexes with repetitions. III. Efficient template-guided chemical polymerization of d(TGGCCAAGCTp). *Nucl Acids Res* 1981, **9**:5747-5761.
 28. Von Kiedrowski G, Wlotzka B, Helbing J: Sequence dependence of template-directed synthesis of hexadenoxynucleotide derivatives with 3'-5' pyrophosphate linkage. *Angew Chem Int Ed Engl* 1989, **28**:1235-1237.
 29. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD: Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucl Acids Res* 2006, **34**:564-574.
 30. Ghirlando R, Wachtel EJ, Arad T, Minsky A: DNA packaging induced by micellar aggregates: a novel in vitro DNA condensation system. *Biochemistry* 1992, **31**:7110-7119.
 31. Kornyshev AA, Wynveen A: The homology recognition well as an innate property of DNA structure. *Proc Natl Acad Sci* 2009, **106**:4683-4688.
 32. Baldwin GS, Brooks NJ, Robson RE, Wynveen A, Goldar A, Leikin A, Seddon JM, Kornyshev AA: DNA double helices recognize mutual sequence homology in a protein free environment. *J Phys Chem B* 2008, **112**:1060-10648.
 33. Barsky D, Colvin ME, Zon G, Gryaznov SM: Hydration effects on the duplex stability of phosphoramidate DNA-RNA oligomers. *Nucl Acids Res* 1997, **25**:830-835.

doi:10.1186/1759-2208-1-9

Cite this article as: Taran et al.: Synthesis of information-carrying polymers of mixed sequences from double stranded short deoxynucleotides. *Journal of Systems Chemistry* 2010 **1**:9.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.chemistrycentral.com/manuscript/

