Formation of RNA Phosphodiester Bond by Histidine-Containing Dipeptides

Rafał Wieczorek, Mark Dörr, Agata Chotera, Pier Luigi Luisi, and Pierre-Alain Monnard

A new scenario for prebiotic formation of nucleic acid oligomers is presented. Peptide catalysis is applied to achieve condensation of activated RNA monomers into short RNA chains. As catalysts, L-dipeptides containing a histidine residue, primarily Ser-His, were used. Reactions were carried out in self-organised environment, a water-ice eutectic phase, with low concentrations of reactants. Incubation periods up to 30 days resulted in the formation of short oligomers of RNA. During the oligomerisation, an active intermediate (dipeptide–mononucleotide) is produced, which is the reactive species. Details of the mechanism and kinetics, which were elucidated with a set of control experiments, further establish that the imidazole side chain of a histidine at the carboxyl end of the dipeptide plays a crucial role in the catalysis. These results suggest that this oligomerisation catalysis occurs by a transamination mechanism. Because peptides are much more likely products of spontaneous condensation than nucleotide chains, their potential as catalysts for the formation of RNA is interesting from the origin-of-life perspective. Finally, the formation of the dipeptide–mononucleotide intermediate and its significance for catalysis might also be viewed as the tell-tale signs of a new example of organocatalysis.

Introduction

The abiotic emergence of catalytic polymers is thought to have been a fundamental event in the origin of life. The RNA world hypothesis postulates that modern life—based on DNA, RNA and proteins—was preceded by a primitive form of life in which RNA substitutes for DNA and proteins as, respectively, genetic and catalytic polymers. Its prebiotic synthesis is therefore one of the most important steps in chemical evolution.

A fundamental requirement of the RNA world hypothesis is a plausible nonenzymatic polymerisation of ribonucleotides that could occur in the prebiotic environment, but the nature of this process is still an open issue. In aqueous solution, condensation of ribonucleotides is thermodynamically disfavoured because hydrolytic reactions dominate. Microenvironments such as mineral surfaces or water-ice eutectic phases have previously been shown to promote condensation reactions by removal of water; but many hurdles remain, primarily related to catalyst nature, reaction regioselectivity and strand replication.

In present-day cells, polymerisation is carried out by enzymes with high efficiency and specificity. Enzymes are genetically encoded polymers requiring a complex, protein-based synthetic machinery that evolved over several billion years. We assume that proto-enzymes were much shorter, noncoded peptides that had low efficiency and specificity. Peptides, even as short as dimers, have been reported to have some catalytic properties, so could have participated in polymerisation reactions thereby leading to the formation of RNA polymers. Such a hypothesis is further supported by recent work that demonstrated that even amino acids could have mediated the enantiopure synthesis of ribonucleotides.

All chemical reactions are in principle reversible, and a catalyst of hydrolysis, for instance, should be able to catalyse the corresponding condensation reaction. The smallest known proteolytic and nucleolytic peptide, H-Ser-L-His-OH (Ser-His), has been shown to catalyse the condensation of amino acids in bulk aqueous solution but, to date, not that of ribonucleotides. Ser-His seems to be a good candidate catalyst for phosphodiester-bond formation in prebiotic chemistry and potentially in early molecular evolution.

Eutectic water-ice mixtures are a plausible prebiotic environment for the promotion of condensation reactions. Indeed, the formation of ice deposits on the early Earth might have been induced by rapid removal of greenhouse gases (principally CO₂) from the atmosphere, thereby leading to a situation (large ice deposits or an ice-covered Earth) similar to that of certain moons in the solar system today. Water-ice systems are formed when aqueous solutions are cooled to below their freezing point but above the eutectic point. The solid phase is mainly composed of pure-water ice, while the remaining liquid phase contains most of the solutes present in the original aqueous phase, but now at very high concentrations. This special environment has three major advantages for...
polymerisation reactions: 1) concentration of solutes and promotion of their self-assembly into organised structures, which are required for efficient polymerisation, 2) reduction of water activity as a nucleophile, and 3) reduction of the hydrolysis rate of the activated monomers and longer RNA products (because of the low temperatures). 

Several important reactions related to an RNA-world scenario have been achieved in water-ice environments. These include the synthesis of nucleobases from simple prebiotic chemicals (e.g. NH₄CN), the metal-ion-catalysed polymerisation of RNA monomers into RNA oligomers, and the preservation of ribozyme activity. However, polymerisation reactions under these conditions yield mainly RNA analogues with predominantly 2'--5' phosphodiester linkages; additionally, metal ions lack evolvability.

For this study, we therefore selected water-ice mixtures to test whether Ser-His can catalyse phosphodiester-bond formation between ribonucleotides. In agreement with earlier reports of nonenzymatic oligomerisation, the formation of phosphodiester bonds requires imidazole-activated monomers. This activation method is a widely used model in nonenzymatic oligomerisation, but its prebiotic plausibility is controversial. Furthermore, we explored the chemical mechanism of the oligomerisation by using several derivatives of Ser-His in control experiments, and carried out a thorough analysis of reaction intermediates. From these results, we propose a likely mechanism for the formation of phosphodiester linkages between ribonucleotide monomers.

**Results and Discussion**

The typical reaction mixtures used in this study contained guanosyl imidazolate (ImpG, 1, 5 mM), MES buffer (5 mM, pH 6.5) and Ser-His (8 mM). Reactions were in most cases incubated at −18.4 °C for up to 30 days. A temperature of −18.4 °C allows formation of the eutectic phase in water-ice close to the eutectic point of the mixture.

In the presence of Ser-His, samples contained products that eluted during standard HPLC analysis at retention times consistent with ribonucleotide dimers and longer products (Figure 1; see also Figure S4 in the Supporting Information) as well as monomeric compounds. The dimeric elution region of the chromatogram (22–30 min; Figure 1) contained three dimeric compounds besides GppG (see also detailed view in Figure S3). Two of these eluted as 2'-- and 3'--linked G dimer standards. The latter—labelled SH-pGpG (9)—eluted later and was found by MALDI-TOF analysis and NMR to be a dipeptide linked to a G dimer (see the Supporting Information).

In the monomer elution region (10–18 min in Figure 1), two peaks—SH-pG (3) and pG-SH (8)—specific to the peptide-catalysed reaction, were recorded. These peaks do not exist in the absence of Ser-His or in metal-ion-catalysed polymerisation.

These monomeric products represented a large fraction of the total initial ImpG and were determined to have a mass consistent with a conjugate of a dipeptide and GMP. Moreover, the formation of this conjugate occurred rapidly (Figure S1): up to 50% of the initial ImpG in the sample was converted within the first few hours of reaction. During the subsequent incubation this compound was slowly consumed, whereas G accumulated at a constant and slow rate over the total incubation period (Figure S1). By contrast, in incubation without dipeptide, the two main products were pG (from hydrolysed ImpG) and the phosphosphate dimer GppG. No other dimers (or longer products) were observed. These findings are consistent with results reported in the literature.

Although the maximum yield for oligomer formation (10% under standard conditions, excluding pyrophosphate dimer; Figure 1) was obtained at equimolar concentrations of 2 and 1, the dipeptide catalytic activity could already be detected at 8 μM Ser-His (a 3125-fold substrate molar excess; Figure 2). Ser-His catalysis was also observed with other activated monomers (Chart S3).

The analysis of the reaction regioselectivity hints at preferential formation of one isomer under optimal conditions. However, direct determination was impaired by the lack of activity of...
were tested: Ser-His(1-Me), acetyl-Ser-His, Ser-His-NH₂, as well as mixtures of serine and histidine as free amino acids. None showed any oligomerisation activity. This implies that the complete dipeptide is important for the catalysis, not just a specific functional group(s). Ala-His, a dipeptide analogue without a hydroxyl side chain, catalysed oligomerisation, but at a significantly slower rate, thus showing the importance of the serinyl hydroxyl group for the reaction kinetics. Trimmers of RNA could be detected within 14 days with Ser-His as catalyst (Figure S2), and with Ala-His as catalyst after longer incubation time (Figure S4).

The reaction was observed over a pH range of 5.0 to 8.2; yield decreased significantly below pH 6. This indicates that the dipeptide catalytic activity is optimal when the unprotonated dipeptidyl histidine is present.

Figure 3 A shows the dependence of SH(-pG) formation on the initial concentration of ImpG, for two different concentrations of Ser-His. The maximum velocity doubled with double the catalyst concentration. Formation of the intermediate 3 could be reversed with the addition of imidazole, thereby leading to reformation of 1 (Figure 3B). Subsequent addition of excess Ser-His could then shift the equilibrium back towards the formation of 3. These results underline the catalytic nature of the reaction mechanism.

It was previously proposed that Ser-His functions during hydrolysis reactions in a manner analogous to that of serine proteases, whose mechanisms includes attack of a peptide bond at the carbonyl carbon by the hydroxyl group of serine—the nucleophile. The acyl-enzyme complex is attacked by water or another nucleophile. If the Ser-His mechanism resembles this during phosphodiester bond synthesis, one can expect that the structure of the Ser-His–nucleotide intermediate complex is a phosphoserine analogue. However, 31P NMR studies indicated that the intermediate complex 3 exhibits a linkage at the imidazole ring instead: the chemical shift of the isolated intermediate was –10.36, very similar to the chemical shift of the imidazole substrate (–10.39), but radically different to phosphoserine amino acid (–0.03; Table S1). This is supported by the fact that the reaction does not proceed at the pH at which histidine residues become protonated. Another pathway could be a rapid conversion of a serine complex to a more stable histidine form: an O-to-N acyl transfer reaction. However, such an intermediate was not detected, and the reactivity of Ala-His does not seem to support this explanation as its intermediate formed as rapidly as that of Ser-His 3 (Figure S4). We therefore suggest a transamination reaction followed by condensation between two intermediates.

Scheme 1 illustrates and summarises the proposed mechanism for the formation of RNA oligomers. The polymerisation of imidazole-activated nucleotides 1 by Ser-His involves the creation of an intermediate 3 (a dipeptide covalently bound to a mononucleotide) that can react with another SH(-pG) molecule, thereby leading to the formation of the SH-capped dimer 9. We propose that Ser-His 2 displaces the activating group (imidazole) and covalently bonds to a nucleotide mono-phosphate (Scheme 1). This compound is subsequently attacked by 3'- and 2'-hydroxyl nucleophiles from another nucleotide, for example 1 or 3, to form a dimer or SH-capped dimer, respectively. Side products were also observed (Scheme 1, bottom), which were the results of nucleophilic attack on 3 by other nucleophiles, such as water, that produce deactivated nucleotide pG (5), the 5’-phosphate of 5, which creates the pyrophosphate dimer 7, and the N terminus of the dipeptide. This last reaction depletes the total amount of catalyst available by creating an inactive compound, 8; but this accounted for only 20% of converted monomers after 30 days.

Compound 3 is the active species, and even alone it can form RNA dimers (Figure 4A and 8). After dimer formation the dipeptide can be “reloaded” with 1 (Figure 4C). Thus, each Ser-His can in principle catalyse multiple synthetic steps. This demonstrates that the peptide works in the formation of phosphodiester bonds like a true catalyst (performs multiple turnover cycles).
Are His-containing peptides plausible in an origin-of-life context? Serine and alanine are among the amino acids present in carbonaceous meteorites and in Miller–Urey-type reactions.\(^\text{20}\) Although histidine does not form as easily, its synthesis in simulated prebiotic conditions has been reported.\(^\text{21}\) The fact that several His-containing dipeptides catalyse phosphodiester-bond formation to varying extents is also advantageous in the context of abiotic synthesis of peptides, where we expect a large mixtures of dipeptides, each likely at low concentration.\(^\text{26}\)

From the prebiotic viewpoint, the suggested mechanism (transamination before oligomerisation) might also be advantageous in abiotic synthesis. If activated nucleotides were produced, it is unlikely that they would all possess the same activation group. The peptide catalysis described here can take advantage of various phosphor-imidazolide activation groups, provided that they contain a secondary amine directly linked to the phosphorus (Chart S2).

By comparing dipeptide catalysis with processes observed for more commonly used catalysts (metal ions\(^\text{23}\) or mineral surfaces),\(^\text{4, 16}\) it is clear that the reported catalysis is less efficient in terms of yield and length of product. But, peptide catalysis, even at low catalytic rate, is very promising, as peptides can in principle be transformed into better catalysts by the introduction of new residues or sequence arrangements, whereas metal ions or mineral surfaces have a finite improvement potential.

**Conclusions**

We have shown that the formation of the RNA phosphodiester bonds can be catalysed by simple \(l\)-dipeptides containing a histidine residue (Ser-His, Ala-His). During the catalytic process, an active intermediate product (dipeptide–mononucleotide) is produced. This intermediate forms rapidly and is in equilibrium with the mononucleotide substrate. The peptide itself can participate in multiple catalytic cycles. The results suggest that small peptides produced during prebiotic processes might have played a significant role in the emergence of an RNA world, by catalysing the oligomerisation of the first RNA chains.

The catalytic properties of small peptides are also of great importance for understanding the origin of enzymatic catalysis. The polymeric nature of these enzymes is thought to be essential to their function.\(^\text{22}\) However, serine hydrolases (the largest enzyme family, throughout all the domains of life) seem to have a functional equivalent in a dipeptide whose activity can be reversed rather easily (from hydrolysis to condensation). Thus, selection toward highly efficient catalytic peptides, which eventually resulted in present day enzymes, could have started at a very early stage of chemical evolution.

This catalysis reaction can also be viewed as an interesting case of organocatalysis. To date, most examples of small-mole-
cule organocatalysis are confined to enantioselective synthesis.[23] The nucleotide condensation reported in this article will broaden the scope of this rapidly emerging field.

**Experimental Section**

**Synthesis of guanosine-5′-imidazole monophosphate (ImpG):**

Procedure adapted from Kanavarioti et al.[24] Guanosine monophosphate (GMP) disodium salt (407.2 mg, 1 mmol, \(M_r = 407.19\)) was converted to its free acid form by cation-exchange chromatography (Dowex 50WX8, H⁺-form, slurry in MiliQ water; column: 25 cm × 1 cm) by elution with acetic acid (0.1 M). Sixteen fractions were collected for lyophilisation, thereby yielding dry free-acid GMP (255 mg, \(M_r = 352.22\), 72.4% yield). Imidazole and GMP were dried over P₂O₅ in an argon atmosphere overnight. All further treatments were performed under argon. GMP (182 mg, 0.5 mmol) and imidazole (340 mg, 5 mmol) were dissolved in dry DMSO (10 mL). Ph₃P (262 mg, 1 mmol) and Et₃N (279 μL, 2 mmol) were dissolved in dry DMF (10 mL). The solutions were mixed in a 50 mL round-bottomed flask. To initiate the reaction, dithiopyridine (441 mg, 2 mmol) was added to the reaction. The reaction was stirred at RT for 5 h and monitored by TLC [water (5 mL), 1-butanol (3 mL), acetic acid (2 mL)]. The product was precipitated as a sodium salt by adding the reaction mixture dropwise to a stirring solution of acetone (120 mL; dried over CaSO₄), dry Et₂O (75 mL), Et₃N (1.8 mL) and NaClO₄ (0.31 g); all kept on ice. A fine, light-yellow precipitate was observed. The solution was stirred for 10 min on ice. The precipitate was removed by filtration through a glass fibre filter under a flow of argon. The filter cake was washed with cold Et₂O/acetone (1:1, 100 mL) and then with cold Et₂O (100 mL). The material was then quickly placed under vacuum to prevent condensation of water on the hygroscopic material. The precipitate had a light-yellow colour. The product was transferred to a small vial and lyophilised overnight. This product was used without further purification to minimise hydrolysis (225 mg; including salts, \(M_r = 435.3\)). Reversed-phase HPLC showed no further side products.

**Scheme 1.** Proposed reaction scheme. Top: The oligomerisation of imidazole-activated nucleotides 1 by Ser-His 2 involves the creation of a covalent linkage between amino acid and mononucleotide (3), thereby releasing imidazole (4). Compound 3 rapidly forms and is then slowly consumed. It can be attacked by 3′- or 2′-hydroxyl group of another nucleotide, thereby forming dimers 9 and 6 (only 3′–5′ bonds shown). Bottom: alternatively, it can be converted into side products: pyrophosphate dimer 7, ribonucleotide 5 and an inactive, stable complex 8 are formed upon nucleophilic attack by a 5′-phosphate of 5, water or the N terminus of dipeptide 2, respectively.
HPLC analysis: Buffer A: phosphate buffer (20 mM), TFA (0.2 %); buffer B: acetonitrile (30 % v/v), buffer A (70 %, v/v); flow rate 0.75 mL min\(^{-1}\); gradient: 0–5 min 100 % A, 5–15 min gradient to 30 % B, 15–20 min isocratic 30 % B, 20–30 min gradient to 60 % B, 30–31 min increased to 100 % B, 31–34 min 100 % B; 34–35 min to 0 % B, 35–48 min eluant maintained at 100 % B. The eluant absorption was monitored by UV/Vis absorption at 260 nm; the column thermostat was set at 20 °C. The identity of the product was confirmed by MS and NMR spectroscopy. \( ^1 \)H NMR (\( ^1 \)H, DMSO, 400 MHz): \( \delta = 7.92 \) (s, 1 H), 7.62 (d, \( J = 13.0 \) Hz, 1 H), 7.09 (d, \( J = 17.6 \) Hz, 1 H), 6.87 (d, \( J = 28.5 \) Hz, 1 H), 6.54 (s, 2 H), 5.65 (d, \( J = 6.6 \) Hz, 1 H), 5.36 (d, \( J = 6.4 \) Hz, 1 H), 5.18 (d, \( J = 4.1 \) Hz, 1 H), 4.53 (d, \( J = 5.6 \) Hz, 1 H), 3.91 (d, \( J = 30.7 \) Hz, 2 H), 3.71 ppm (d, \( J = 40.5 \) Hz, 2 H); MALDI MS: m/z 414 (30 % [M+H]\(^+\)), 436 (40 % [M+Na]\(^+\)).

Synthesis of other 5-imidazole monophosphates (ImpN): Other 5-imidazole monophosphates [ImpN (1)] and 5'-monophosphates of other imidazole-derived compounds (2-methylimidazole (2-MeIm), 2,4-dimethylimidazole (2,4-dMeIm), 2-aminobenzimidazole (2-NH\(_2\)BenzIm)) were synthesised analogously to ImpG.

Standard reaction procedure: Reactions were performed in 0.5 mL glass ampules. The following concentrations were used in most experiments: 6 mM Ser-His, 5 mM ImpG, 5 mM MES buffer. The reaction vials were incubated at 37 °C.

Preparation of reaction mixtures: Typically, reaction samples were prepared with various concentrations of Ser-His, ImpNs, different buffers (Tris-maze base, AcOH) and different substitutes for Ser-His: His-Ser, serine, histidine (Sigma–Aldrich), Ser-OEt, Ac-Ser-OMe (Bachem), Ser-His(1-Me) (a gift from Katarzyna Adamala, Department of Genetics, Harvard Medical School, MA). This work was supported by the Danish National Research Foundation, which supports the Center for Fundamental Living Technology (FLinT, with Steen Rasmussen as centre leader) and by funding from the EU FP7 programme under grant agreement 249032. R.W. was supported by the EU FP6 programme (043359, SYNTHCELLS).

Acknowledgements

We would like to thank Katarzyna Adamala for the Ser-His(1-Me), Gonen Ashkenazy, Stephen Douthwaite, David W. Deamer, Martin Hanczyc, John D. Sutherland and Günther von Kiedrowski for discussions, as well as Michael Pedersen for help with the \(^{31} \)P NMR measurements. This work was supported by the Danish National Research Foundation, which supports the Center for Fundamental Living Technology (FLinT, with Steen Rasmussen as centre leader) and by funding from the EU FP7 programme under grant agreement 249032. R.W. was supported by the EU FP6 programme (043359, SYNTHCELLS).

Keywords: catalysts · nucleotides · origin of life · peptides · RNA formation

References