Amino Acid-Specific, Ribonucleotide-Promoted Peptide Formation in the Absence of Enzymes

Helmut Griesser*, Maren Bechthold*, Peter Tremmel, Eric Kervio, and Clemens Richert*

In memory of Thomas Büchner

Abstract: Nucleic acids and polypeptides are at the heart of life. It is interesting to ask whether the monomers of these biopolymers possess intrinsic reactivity that favors oligomerization in the absence of enzymes. We have recently observed that covalently linked peptido RNA chains form when mixtures of monomers react in salt-rich condensation buffer. Here, we report the results of a screen of the 20 proteinogenic amino acids and four ribonucleotides. None of the amino acids prevent phosphodiester formation, so all of them are compatible with genetic encoding through RNA chain growth. A reactivity landscape was found, in which peptide formation strongly depends on the structure of the amino acid, but less on the nucleobase. For example, proline gives ribonucleotide-bound peptides most readily, tyrosine favors pyrophosphate and phosphodiester formation, and histidine gives phosphorimidazolides as dominant products. When proline and aspartic acid were allowed to compete for incorporation, only proline was found at the N-terminus of peptido chains. The reactivity described here links two fundamental classes of biomolecules through reactions that occur without enzymes, but with amino acid specificity.

Polypeptides are the best known biocatalysts of the cell. They are encoded in polynucleotide sequences that are translated by a highly evolved biochemical machinery. Therefore, it is not easy to say how the earliest encoded proteins were synthesized.[1] Phylogenetic analyses do not reach back far enough, and experiments trying to re-enact molecular evolution[2–4] have not yet provided a definite answer.[5] Because much evidence points to RNA as the earliest biopolymer that was able to encode genetic information and to act as biocatalyst in what is often called the “RNA world”,[6] it is likely that RNAs captured amino acids.[7] Both amino acids[8] and nucleobases[9] have been found to form under potentially prebiotic conditions. Pathways for the abiotic formation of nucleotides have also been described.[10,11] Furthermore, phosphorylation could have produced nucleoside monophosphates in the absence of enzymes.[12] Thus, it is reasonable to assume that both ribonucleotides and amino acids could have come in contact with each other[13] to start a very simple form of RNA-promoted peptide synthesis, most probably by reacting with each other.

Most studies on prebiotic RNA-induced peptide formation have focused on aminoacyl adenylates as starting materials, that is, the very labile intermediates of aminoacylation of tRNAs of present-day biochemistry,[14–17] which have to be prepared in a separate synthetic step. Synthetic 2'- or 3'-aminoacyl esters of phosphate-blocked nucleotides[18,19] or oligonucleotides[17,20] have been employed as mimics of the charged tRNAs of translation. The aminolysis of the blocked nucleotides has produced free dipeptides,[21] but not peptidyl RNAs. Short peptidyl RNAs were detected when an RNA strand was reacted with an aminoacyl adenylate in the presence of a small ribozyme, but not without this catalyst.[17]

Other forms of RNA-promoted peptide formation may have existed that started from free amino acids and free ribonucleotides and did not require prior synthetic steps or a ribozyme.

We recently found that a concentrated aqueous buffer with a water-soluble carbodiimide, ethylimidazole, and magnesium chloride drives RNA chain formation and genetic copying, starting from unactivated ribonucleotides.[22] At near-neutral pH, this combination of salts and organics activates both ribonucleotides and amino acids and induces reactions without potentially biasing chemical pre-activation and without a need for intervention during assays. In our assays, N-linked peptido RNAs (Figure 1) with chain lengths of up to 14 amino acid residues and up to five ribonucleotide residues were found.[23]
The low temperature and high salt concentration of our assays is reminiscent of eutectic phases, and is favorable for folding of biomacromolecules. Peptide formation in N-linked form also accelerates chain growth over background oligomerization, as detailed in the accompanying paper, so no enzymes are needed to start it. Because key molecules of encoding (RNA), catalysis (peptides), and metabolism (cofactors) are formed simultaneously, this reactivity system may have been relevant for the earliest forms of life. Thus far, only glycine and phenylalanine had been tested, and it was unclear whether amino acids with functional groups in their side chains would be tolerated. Herein, we report on the reactivity of all 20 proteinogenic amino acids in the presence of any of the four canonical ribonucleotides. The reactivity of the amino acids was found to favor the formation of some peptide sequences over others in this pristine form of RNA-promoted peptide synthesis.

Using an assay relying on in situ $^{31}$P-NMR monitoring established recently, we screened the proteinogenic amino acids for their ability to form peptide RNA. The phosphorous signals provide a measure of what reaction channels a given ribonucleotide enters. Figure 2 gives an overview of the product classes formed, representative NMR spectra, and the distribution of products after 7 d at 0°C in condensation buffer containing 0.5 M HEPES, 80 mM MgCl$_2$, 0.15 M 1-ethylimidazole, and 0.8 M N-ethyl-N'-(3-dimethylamino-propyl) carbodiimide (EDC) at 0°C. Besides those for AMP, four reactivity maps each are shown for the remaining three canonical nucleotides (CMP, GMP, and UMP). The $^1$H,$^{31}$P-HMBC spectra for the 20 proteinogenic amino acids and AMP that confirmed the assignments are shown in Figures S1–S20 of the Supporting Information, together with the $^{31}$P-NMR spectra for all 80 combinations of amino acids and ribonucleotides (Figures S22–S31), and five assays performed with cyanoacetic acid as activating agent (Figures S33, S34).

The yields of the six main reaction channels shown in the bar graph “reactivity fingerprints” of Figure 2c are those of the initial phosphoramidate capture product (PA), peptide products of different lengths (PE), phosphodiester formed through nucleotide incorporation in RNA chains (PD), pyrophosphates (PP), and all the other phosphate products (PX). Their intensities vary with the number of nucleotides in the respective product class. The peaks for the miscellaneous products (PX) were more intense in the assays with GMP, which were run at much lower concentration (0.01 M instead of 0.2 M) due to low solubility. Here, the peak for the imidazo-
lium nucleotide, a reactive amide formed with the organo-catalyst, was more persistent, and so were the peaks for remaining free phosphate due to the lower probability of encountering a reaction partner.

The intensity of the peaks for peptido nucleotides are given as red bars in Figure 2c. Among all of the proteinogenic amino acids, proline consistently gave the most intense peaks for peptido phosphates (up to 62% of products). This is interesting because proline is well known for its organo-catalytic properties and its oligomers adopt a rigid conformation. On the other hand, cysteine, with its strongly nucleophilic side chain thiol, did not give significant amounts of peptido nucleotides with any of the four ribonucleotides (A/C/G/U). Minimal amounts were also found for tyrosine, aspartic acid, and glutamic acid, with free phosphate as the most dominant species.

Arginine, which is frequently found in ribosomal proteins and features a stably protonated guanidinium side chain, behaved like the well-soluble aliphatic or hydroxy-functionalized amino acids, giving very substantial amounts of peptido nucleotides. Asparagine and glutamine were poorer in their peptido-forming abilities, similar to the strongly hydrophobic amino acids Trp, Leu, and Phe, for which precipitates were common. For histidine and lysine, chain-linked products dominated after 7 d, either in the form of phosphoramidates involving the ε-amino group or as imidazolides (for the latter). Side-chain imidazolides of peptides can be substrates for enzyme-free phosphodiester formation.

None of the 20 proteinogenic amino acids suppressed chain growth in the phosphodiester-forming “RNA channel” (Figure 2, orange bars). This suggests that despite the range of functionalities found in the different amino acids, neither possesses a reactivity that prevents a simultaneous evolution of peptides and RNAs. Also, when comparing the reactivity landscapes of the different nucleotides, it is noteworthy that AMP consistently gives the highest yield of pyrophosphates. This is the very ribonucleotide found in pyrophosphate-containing cofactors such as NAD⁺, FAD, and acetyl CoA.

To test for sequence-selectivity, we then performed exploratory experiments with mixtures of two different amino acids. Figure 3 shows results from the combination of proline, aspartic acid, and AMP. While the general product distribution was not far from an average of the distributions found for the individual amino acids, analysis of the peptido RNA fraction indicated selectivity. Figures 3c and 3d shows an HMBC spectrum and a mass spectrum of the two main peptido RNA fractions from a five-step cartridge purification. The NMR spectrum of the fraction containing longer peptido chains shows cross peaks between the phosphate and proline residues, but not aspartic acid residues. The mass spectrum shown is for the more polar fraction containing shorter peptido chains, down to the single amino acid capture product. Here, the proline-bearing species (Pro-A₃) is observed, whereas the corresponding Asp-A₃ phosphoramidate is not (broken line in Figure 3d), confirming that there is selectivity in the incorporation at the N-terminus.

Figure 3. Formation of peptido RNA from a mixture of proline, aspartic acid, and AMP leads to selective incorporation of proline at the N-terminus. Conditions: condensation buffer, 200 mM AMP, and 100 mM of each amino acid, 7 d at 0°C. a) Reaction scheme. b) Product distribution in 31P-spectra for control assays of aspartic acid alone or proline alone (at 200 mM), and for the mixture of Asp and Pro (100 mM each), with color code according to Figure 2a; y-axis maxima are 60% fraction of products. c) Phosphoramidate region of the 1H,31P-HMBC spectrum of the main peptido RNA fraction obtained by reversed-phase cartridge purification. Cross peaks from the phosphate to protons of the proline spin system are detected, but no cross peaks to aspartic acid residues. d) MALDI-TOF mass spectrum of an early fraction of the cartridge separation showing trinucleotide species with short peptide chains. Longer peptide chains contain aspartic acid residues, but the initial capture product is observed for proline only. The missing peak for Asp-A₃ is labeled with a broken line.
Taken together, our results led to the mechanistic proposal shown in Figure 4. Several molecular species are likely to form when amino acids react with the 5′-phosphate of a nucleotide. After activation of the phosphate or the carboxylic acid, a mixed anhydride can form. This is the species implied in the aminoacylation of tRNA.[29] But mixed anhydrides are very reactive,[30,31] perhaps explaining why they do not reach a steady state concentration observable by NMR. Other structures that may form independently or through intramolecular substitution are the phosphoramidate, and, for side chain-functionalized amino acids, a side-chain-linked conjugate (SCC). Upon renewed activation, mixed anhydride or phosphoramidate may then cyclize to give the 1,3,2-oxazaphospholidin-2,5-dione called cyclic aminoacyl phosphoramidate (CAPA) [31,32] which has been cited as a species whose reactivity explains why only α-amino acids are encoded genetically.[33] Some side chain-functionalized amino acids, like histidine and lysine, form sufficiently stable SCCs to be detectable by NMR; others appear to form SCCs too labile to remain covalently linked to the nucleotide, being hydrolyzed instead.

The attack of the side-chain nucleophile may also occur on the level of the CAPA, an uncharged species that does not repel nucleophiles electrostatically. A rapid conversion of a CAPA to a less productive side-chain-linked intermediate may prevent peptide coupling.[25] Together with stereoelectronic and conformational factors, such as the rigidity of proline, intrinsic reactivity may thus provide the structural basis for whether peptido RNAs form efficiently or not. The actual extent of peptide production will vary not only with structure, but also with the relative abundance of amino acids, so that the reaction network may turn out different quantities of peptido species without higher-level control.

It is interesting to ask how a system producing peptido RNAs can evolve into one that produces peptides of a specific length and sequence, and how it can later be turned into a genetically encoded system with 3′/5′-linked intermediates. It is not easy to follow this experimentally, but it is noteworthy that transfer reactions from the phosphate-anchored form to 2′/3′-esters are known.[34,35] Furthermore, the confinements of a pre-ribosomal apparatus may increasingly limit the length of peptido chains, reducing polydispersity. The carboxy terminus of the peptido chains is free to be captured through esterification to the 2′,3′-dial of RNA strands, opening a covalent pathway for such a process. Subsequent transacylations are likely because esters are not long-term stable in condensation buffer containing free amino acids, as detected when glycine methyl ester was used as starting material for assays with a 5′-phosphate-terminated RNA strand (Figure S35).

In conclusion, RNA-promoted peptide formation can produce different peptide sequences, starting from free amino acids and ribonucleotides. Chain growth is strongly affected by the reactivity of the first amino acid residue captured by the priming ribonucleotide. The apparent absence of peptido RNAs in present-day cells, where peptidyl tRNAs dominate as the intermediates of RNA-directed polypeptide synthesis, may mean that they were transitional species, but ribonucleotide-promoted peptide growth leaves free the carboxy terminus of the peptide so that a transition to 2′,3′-ester chemistry is conceivable. Like a molecular Archaeopteryx, the peptido RNAs may have filled the gap between a pure RNA world and a molecular world with ribosomal machinery that ensures a formally encoded, RNA-controlled growth of polypeptides. Independent of such speculations, our results show transformations of key biomolecules, interacting in un instructed fashion, driven by intrinsic reactivity in cold aqueous solution. A rich systems chemistry of some of nature’s most pivotal molecules remains to be fully explored.
Acknowledgements

The authors thank S. Motsch, M. Sosson, and S. Kaspari, for sharing results, Drs. D. Garnier and B. Claesen for NMR spectra, D. Gühringer, and T. Krahl for expert technical assistance. This study is supported by grants from DFG (RI 1063/8-2 and RI 1063/16-1) and EU COST action CM1304.

Conflict of interest

The authors declare no conflict of interest.

Keywords: amino acids · nucleotides · peptides · prebiotic chemistry · RNA

How to cite: Angew. Chem. Int. Ed. 2017, 56, 1224–1228

Angew. Chem. 2017, 129, 1244–1248


Manuscript received: October 31, 2016
Revised: December 1, 2016
Final Article published: December 21, 2016