

# SMALL HAIRPIN RNAs AND THE ORIGIN OF PROTEIN SYNTHESIS

Chris Allen Broka  
(chris.broka@gmail.com)

## Abstract

A model of the early RNA world is proposed. Nearly self-complementary sequences that could adopt double-stranded, small hairpin-like (shRNA), structures would be selected for due to their greater hydrolytic stability. These would be phosphorylated at their 5' ends. We suppose that dehydrating conditions arise (perhaps intermittently) in the early environment allowing amino acids to condense with these RNA molecules. The resulting phosphate-amino acid anhydrides would play the role of early, charged, tRNAs. A crude genetic code could emerge owing to the greater resistance of some amino acid-shRNA pairings to hydrolysis relative to others. Early on there is no division of labor between mRNAs and tRNAs; the same molecules perform both functions. But the first systems would have encoded little in the way of protein sequence information. Rather they would have served as catalysts for the random polymerization of amino acids. It is speculated that the selective advantage inhering in such systems lay in their ability to supply raw materials for the formation of coacervates within which the various molecules essential to proto-life could be concentrated. This would greatly facilitate the necessary chemistries. The evolution of homochiral protein and RNA populations is discussed. An appealing feature of this model is its ability to explain the transition from phosphorylated amino acids to the 3' ester-linked aminoacyl-tRNAs employed by modern life.

**Keywords:** Prebiotic Evolution, Protein Synthesis, Hairpin RNA, Genetic Code, Phosphate Anhydride.

## Introduction.

How did populations of well-defined, self-replicating, RNA sequences originate? And how did these come to figure in the synthesis of proteins? The first question we answer by appealing to the enhanced hydrolytic stability of double-stranded RNA compared to the other nucleic acid species that would also be present (1). These molecules would be gradually accumulated at the expense of other possibilities. The other question is more vexing. If early protein synthesis at all resembled its contemporary version there would have to have been populations of both mRNA-like molecules and tRNA-like ones. Such a situation would have to be considered unnaturally fortuitous, at best. We, therefore, conclude that things did not look this way. There must have been a far simpler process going on in which the same RNA molecules performed both functions. It is hard to imagine how such a primitive system could encode much in the way of protein sequence information. We, therefore, conclude that it did not. Rather, it only served to catalyze the random polymerization of amino acids present in the environment resulting in polypeptides/proteins. What advantage would the synthesis of such oligomers provide for a proto-organism? We are reminded of the very old, but very clever, suggestion of Oparin and Haldane. Proteins, in the presence of polysaccharides, as well as hydrophobic polypeptides by themselves (2) can, under the right circumstances, give rise to coacervate-like structures as well as proteinoid microspheres (3). These, we propose, served to isolate and concentrate the RNA species of interest.

## The Early RNA World.

We picture the prebiotic environment as a complex broth of amino acids, sugars, oligosaccharides, and nucleotides of various sorts, all produced through abiogenic chemistry. Within this world two countervailing processes seem to have operated. The first was 'constructive' and involved the condensation of these monomers

through phosphorylation and phosphate activation (the 'P-process'). We cannot propose any specific chemical mechanism here (but Montmorillonite clay has been suggested as a catalyst (4)) – only that it operated and contributed to the development of structure within the early world. The second process was simple hydrolysis (the 'H-process') which must also have been occurring and which acted to destroy the emerging ordered structures.

Through the agency of the P-process the various nucleotides present would be randomly polymerized into oligonucleotides of various types. Most of these would be incapable of forming double helices and would be subject to rapid hydrolytic degradation. Double-stranded (ds)RNA enjoys enhanced hydrolytic stability due, in part, to the conformational constraints resulting from its structure which prevent the 2'-hydroxy group from assisting in the cleavage of its adjacent phosphodiester linkage (1). We also speculate that, should hydrolysis break such a linkage, the broken strand would be held together by its partner long enough for the P-process to repair the broken bond. In this way dsRNA would come to predominate whereas other species would tend to remain very minor players. We also imagine that conditions were such that these dsRNAs existed in equilibrium with some corresponding single chains. (Perhaps the dsRNA melts somewhat during the heat of the day?) These would function as templates upon which the P-process could act thus resulting in self-replication and further increasing the concentration of these dsRNAs.

Now, in addition to the H-process, another factor serves to militate against the development of structure in our primitive world. This, of course, is entropy. For an RNA strand to benefit from stabilization as a double helix it must find and stick to its complementary partner. Since conditions in the early world may well have been rather dilute this would be difficult. This problem would be alleviated were the RNA sequence almost self-complementary and able to form a hairpin-like structure (fig. 1). Since only homochiral oligonucleotides form strong duplexes the early small hairpin RNAs (shRNAs) would have been composed of either all-D or all-L ribonucleotides. But populations of both may have coexisted.

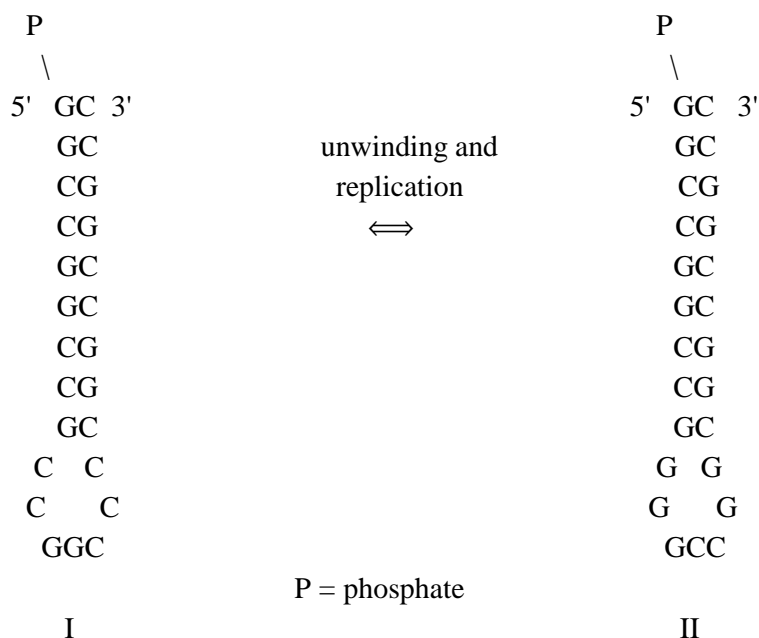


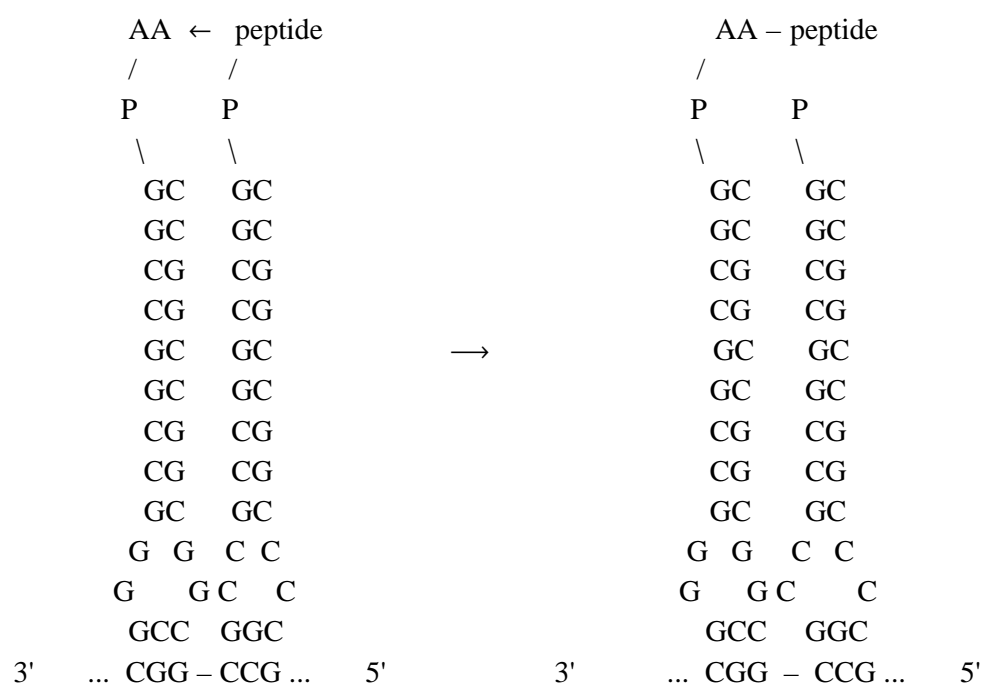
fig. 1

The above-suggested sequence is somewhat arbitrary and, at this point, serves only as an illustration. C and G are employed simply because they form tighter base pairs. The exact length of the "stem" is unknown. We propose the connecting loop to be seven bases in length by analogy to modern tRNA. This connector carries an early three-nucleotide anticodon (*vide infra*). In the modern world these are three bases in length and it is difficult to see how Nature could have transitioned to a three letter code from something else. It will be observed that replication, in our example, will lead to two closely similar populations (the I-strands and the II-strands). If such a system composed, for instance, of D-ribonucleotides came to dominate the early world the L-ribonucleotides and other species, enjoying less protection against hydrolysis, would be gradually eliminated.

## A Glimpse of Early Protein Synthesis?

We can imagine these hairpin-like RNAs coming to dominate the prebiotic world owing to their resistance to hydrolytic degradation. But, thus far, they perform no function other than to exist. We propose that their first useful function was to catalyze the random polymerization of amino acids present in the environment. Modern protein synthesis proceeds by the transfer of growing peptide chains from the 3'-hydroxyl of one tRNA to the neighboring aminoacyl-tRNA. This process is catalyzed by the ribosome itself (which would, obviously, not have existed at this early time). In the absence of this catalysis such a process would have been anything but efficient. Esters are very poor acylating agents. We cannot think things happened this way early on.

We suppose that, in addition to polymerizing nucleotides leading to RNA, the P-process also joined amino acids and peptides to the 5'-ends of our hairpin-like RNAs resulting in aminoacyl and peptide phosphate anhydrides ( $\text{RCOO-PO}_2\text{-oligonucleoside}$ ) (5, 6). These would be highly activated acylating species and function as substrates for template-directed protein synthesis (fig. 2).



AA = amino acid

fig. 2

We now realize that the 'arbitrary' hairpin RNA structure given above is, actually, not so arbitrary after all. We see that it is ideally suited to the synthesis of peptide bonds. We suppose it is in equilibrium with its melted form. The latter would provide a catalytic template whose function would be to bring into close proximity the growing peptide chain and the next amino acid to be added. In this way our hairpin RNA acts to catalyze the polymerization of amino acids.

We imagine this polymerization, initially, to be a completely random process; the shRNA encodes no

particular sequence information. It merely allows for the generation of larger polypeptides from smaller ones and amino acids. As alluded to above, we think the advantage of this lies in its ability to produce coacervate-like structures within which the components of our system can be concentrated and confined. The efficiency of our processes would be greatly enhanced thereby. This would result in selective pressure that would favor species such as our I-II system at the expense of an RNA system like III-IV (fig. 3) which could not catalyze protein synthesis.

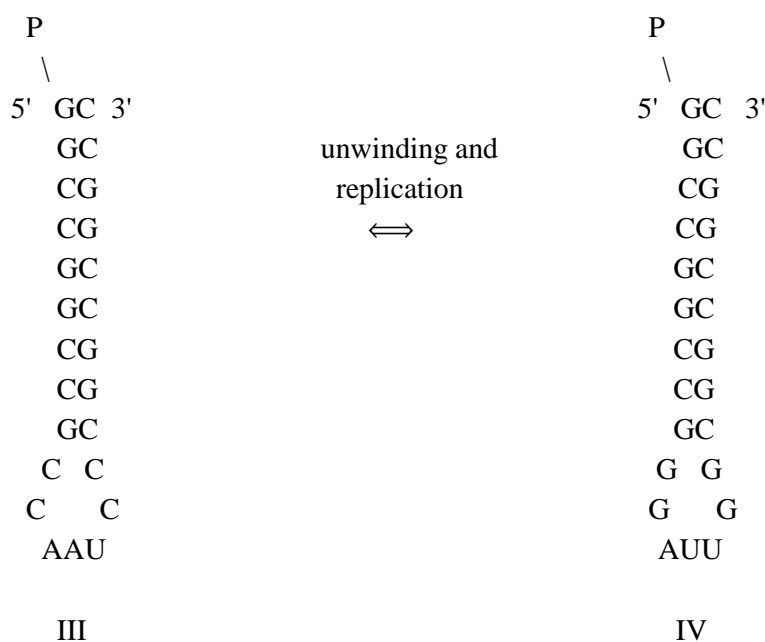


fig. 3

Note that it does not matter where along the single-stranded RNA the binding occurs as long as it is at two compatible neighboring sites. Protein synthesis could proceed in either the  $3' \rightarrow 5'$  or  $5' \rightarrow 3'$  direction or both. It might be objected that rather special conditions would have to obtain for our hairpin-like RNA to coexist with its, melted, single-stranded form as in fig. 2. This is true. But we can also imagine the proto-mRNA to be just a smaller, broken fragment of the shRNA molecule. Certainly, many such fragments would be present owing to the H-process. (A different mechanism for hairpin RNA-based protein synthesis has been given by Di Giulio (7). Unlike the present model it proposes the early aminoacyl-tRNAs to be ester-linked.)

Toward a Genetic Code.

Suppose that, in addition to the hairpin RNA discussed above, our coacervate was also populated with species V and VI (fig. 4). These could also catalyze the polymerization of amino acids in the same random way as I and II. So far none of these RNAs encode any information whatsoever. The evolution of more specific systems is proposed to involve an idea suggested by Hopfield (8). (In contrast to the present model, Hopfield pictures the early aminoacyl-tRNAs as ester-linked.) Imagine that there were two amino acids,  $AA_1$  and  $AA_2$ , present in the early environment. The P-process would lead to the formation of phosphate anhydrides having  $AA_1$  and  $AA_2$  randomly linked to I, II, V, and VI. All of these active acylating species would be subject to hydrolysis, of course. Suppose that species having  $AA_2$  linked to I or II, and those having  $AA_1$  linked to V or VI, were relatively more hydrolytically labile than the other possibilities. These differences in stability would result from purely chemical factors (e.g. favorable or unfavorable noncovalent interactions between the amino acid side-chain and the nucleotides constituting the 5' end of the shRNA). If such were the case we would end up with a coacervate enriched in  $AA_1$ -(I,II) and  $AA_2$ -(V,VI) phosphate anhydrides. Thus the (I, II) RNA system would lead, mostly, to the synthesis of poly- $AA_1$  and the (V, VI) system poly- $AA_2$ . Protein synthesis would no longer be a completely random process. We need only imagine that there was some selective advantage in having this be the case; maybe the two different proteins cooperate to form better coacervates than a random mixture of proteins with no defined structure. Coacervates containing both the (I, II) and (V, VI) RNA systems would enjoy an advantage and be selected for. Since modern life employs L-amino acids only we would suppose that L- $AA_1$ -(D-I, II) would be more stable than, say, D- $AA_1$ -(D-I, II). But, conceivably, the  $AA_2$ -(V, VI) system could be of the opposite chirality.

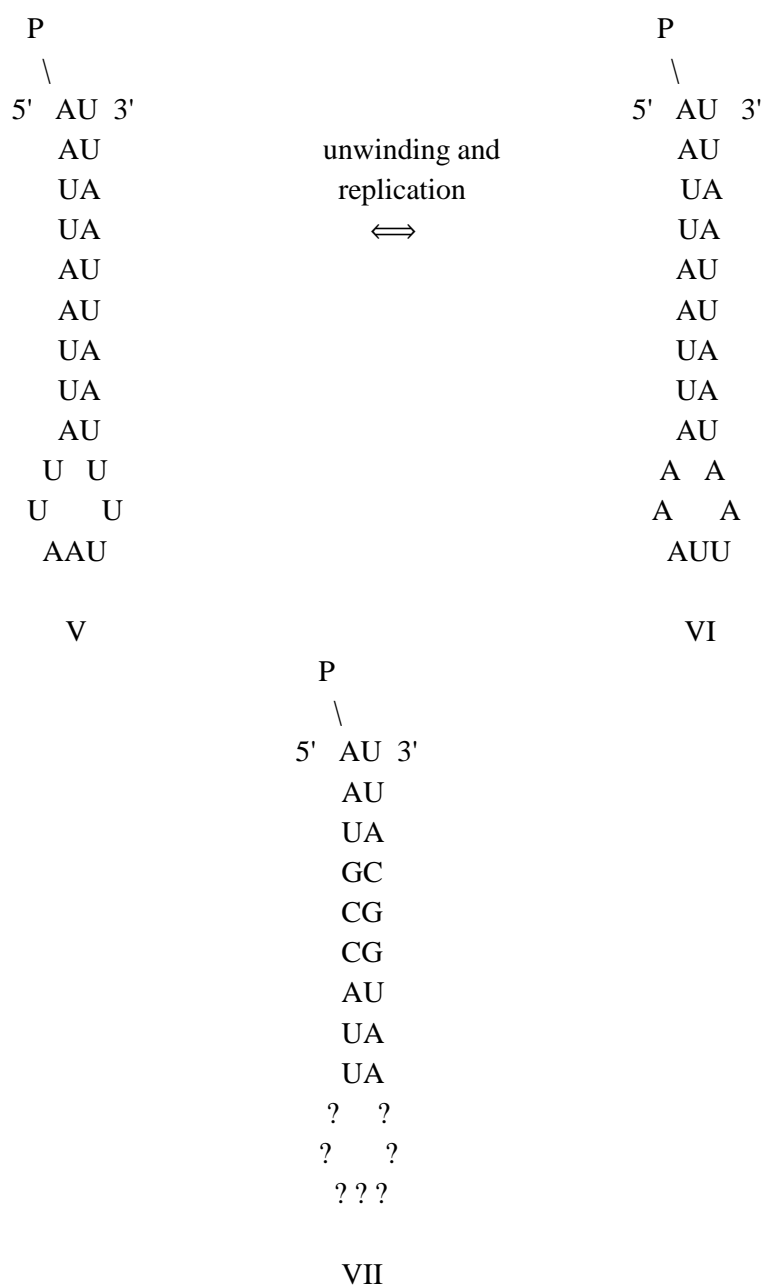


fig. 4

Imagine that such a coacervate developed to also contain VII. This would enable the production of  $(AA_1 - AA_2)_n$  proteins. If there were some selective advantage to this such coacervates would come to dominate the early prebiotic world. (Note also that it is at point that strict homochirality would be imposed upon the pre-living system.)

Over time we suppose that the coacervate evolves to contain pools of several, differentially acylated, shRNAs. The amino acid mostly attached to each species is, again, determined by the hydrolytic stability of its phosphate anhydride linkage. We can begin to see an emerging distinction between what will become mRNAs

and tRNAs. Indeed, we can imagine the coacervate accumulating a pool of non-shRNAs if these encoded a protein sequence that was beneficial to it. This encoding process would, undoubtedly, have been miserably sloppy. The number of amino acids so coded for would probably have been far fewer than the 20 life now utilizes. But it would be a start. Besides producing more stable coacervates, what benefits might a more-or-less defined protein sequence confer upon a proto-organism? If the protein were able to perform some catalytic function leading to more efficient protein synthesis it would certainly be selected for very strongly. This catalytic function might have consisted of stabilizing the unwound form of the pre-mRNAs making them more available to act as templates. Or the protein might have facilitated, somehow, the acylation process necessary for peptide growth. We see what begins to look like a primitive ribosome.

### From 5' to 3'.

A peculiar feature of this model is its invocation of 5'-phosphate amino acid anhydrides as early tRNAs. In the modern world amino acids are generally linked to their appropriate tRNAs by 3'-ester linkages. We cannot think that such poor acylating agents would have allowed for much protein synthesis absent some catalysis. It may seem unlikely that life could have transitioned over from one protein synthesis mechanism to the other without the complete disruption of its nascent genetic code. But this is not the case. Consider the process illustrated in fig. 5. It would be a very chemically normal thing for an active acylating species to react with a nucleophile (the 3'-OH group) that was brought into its close proximity by the structure of the (somewhat flexible) shRNA.

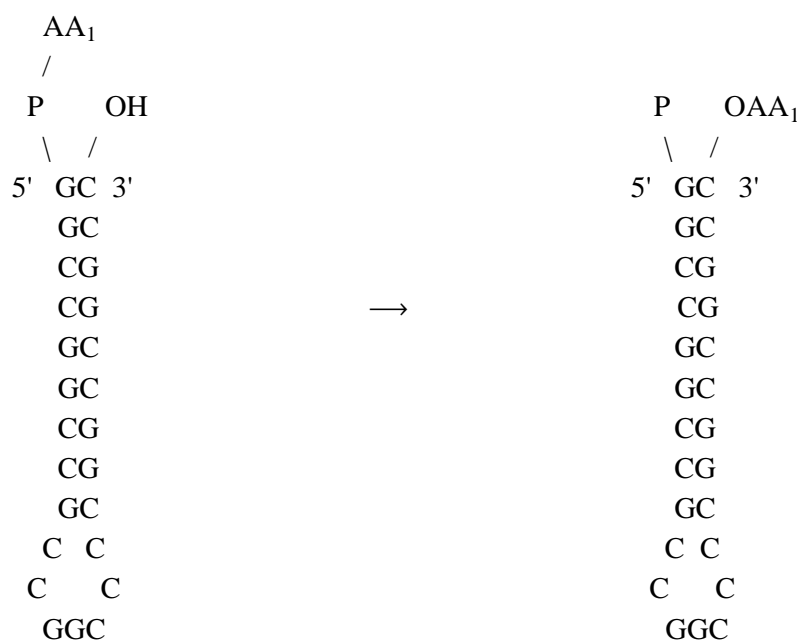


fig. 5

Early on this migration of the amino acid from 5' to 3' would have been nothing more than an unhelpful side-reaction; it would deplete the amino acids needed for protein synthesis and result in an inactive species. If, however, a sort-of sequence-definite protein were to arise that could catalyze what is, normally, a poor and



sluggish acylation process these 3'-aminoacyl tRNAs would offer a great advantage to the coacervate – they are stable and could accumulate to high concentrations easily. As soon as some ribosome-like catalytic activity began to emerge, the switchover from phosphate anhydrides to the (previously useless) 3'-aminoacyl versions thereof would be favored strongly. And we observe that the preferential pairing of a given amino acid with its corresponding tRNA sequence would be preserved. The primitive genetic code would carry on intact.

As the early organism becomes more adept at protein synthesis we can envision the development of simple aminoacyl-tRNA synthetases. These would obviate the phosphate linkages altogether. And evolution could fine-tune these synthetases so as to effect a much more accurate pairing of each amino acid with its tRNA than could Hopfield's mechanism. Since complementary anticodons do not, nowadays, generally correspond to the same (or similar) amino acids we have to conclude that most of our modern genetic code evolved well after the emergence of aminoacyl-tRNA synthetases. All or most evidence for our much simpler early code seems to have been erased over time.

## References.

- 1) Elliot, David; Lodomery, Michael (2011). *Molecular Biology of RNA* (1 ed.). New York: Oxford University Press. 34– 64.
- 2) Sharpe, Simon; Keeley, Fred W.; Muiznieks, Lisa D.; Reichheld, Sean E. (2017). Direct observation of structure and dynamics during phase separation of an elastomeric protein. *Proc. natl. Acad. Sci.* **114** (22): E4408– E4415. doi:10.1073/pnas.1701877114.
- 3) Fox, Sidney W. (1965). *The origins of prebiological systems and of their molecular matrices*. New York: Acad. Pr.
- 4) Ferris, James P. (2006). Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361** (1474): 1777– 1786. doi: 10.1098/rstb.2006.1903.
- 5) Liu, Ziwei; Beaufils, Damien; Rossi, Jean-Christophe; Pascal, Robert (2014). Evolutionary importance of the intramolecular pathways of hydrolysis of phosphate ester mixed anhydrides with amino acids and peptides. *Scientific Reports*. **4**: Article number: 7440.
- 6) Tamura, Koji; Schimmel, Paul (2003). Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proc. natl. Acad. Sci.* **100** (15): 8666-8669. doi.org/10.1073/pnas.1432909100
- 7) Di Giulio, Massimo. (1994). On the origin of protein synthesis: A speculative model based on hairpin RNA structures. *Journal of Theoretical Biology*. **171** (3), 303-308. doi.org/10.1006/jtbi.1994.1233.
- 8) Hopfield, John J. (1978). Origin of the genetic code: A testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proc. natl. Acad. Sci.* **75** (9): 4334-4338.

