Updated December 10th, 2015

CONTRIBUTIONS TO THE ORIGINS OF LIFE

JACQUES NINIO

(included in the web site http://www.lps.ens.fr/~ ninio)

TOPICS DISCUSSED HERE:

Biographical notes
Origin of the genetic code
Evolution of transfer RNA structure
Prebiotic replication
Catalysis by peptides
RNA catalysis
Against the 'RNA world' hypothesis
Criticism of Eigen's work
Criticism of chemical predestination
References
Appendix: Project submitted to EMBL on catalytic activities in peptides

BIOGRAPHICAL NOTES

For a long time, the origin of life appeared to me as a frontier in biology, the one to which I wished most to contribute. In this domain I had sound ideas and many advanced projects, but achieved little, in comparison to some other domains, being handicapped by my extremely poor understanding of chemistry. Yet, I made a number of valid publications, from which I have nothing to retract. In the theoretical domain, I published ideas on the origin of the genetic code, and on the evolution of transfer RNA structure [1,2]. Similar ideas were published later under other signatures. Experimentally, I did early work on prebiotic replication with oligonucleotides instead of monomers [3] and work on RNA catalysis with nucleotide analogues [4]. I was well accepted in the origins of life scientific community, and knew personally most of the people who were doing important work in the 1970's and the 1980's.

I started to speculate on the origin of the code while doing my thesis on transfer RNA structure. During the Christmas 1965 vacations, I had a wrong illumination about stereochemical complementarity between amino acids and anticodons. The main argument was that big amino acids (such as tryptophan and methionine) had non-degenerate codons, and small ones such as glycine or alanine had fourfold degenerate codons. This was taken as an indication that the big ones could interact with three nucleotides simultaneously, while small ones could make contact with only two

nucleotides. From this view of stereochemical relationships subtending the genetic code, I adhered to the idea of a reversibility in primitive translation, i.e., in the possibility of also translating from proteins to nucleic acids, in contradiction with the Central Dogma. These ideas were received with well-deserved scepticism, and Jacques Monod was seriously considering putting an end to my fellowship. I met at that time Mirko Beljansky, who was studying what he thought was non-ribosomal synthesis of peptides (it turned out to be a terminal acylation of viral RNA). In December 1966, there was a Meeting of the Biophysical Society in London, in which Francis Crick launched his ideas on the origin of the code (see updated accounts in [5, 6]). I was in the audience. According to Luisa Hirshbein's recollections, she did everything to dissuade me from taking part in the debate. Nevertheless, I did ask a question, after Crick's talk, possibly on what he thought of reverse translation, and I remember that he answered without aggressiveness. On the previous day (I think), I had paid a visit to Pelc and Welton, in London, who had published in Nature, models of stereochemical interactions between amino acids and codons [5], that turned out to be erroneous [6]. So, I was sinking into marginality.

In May 1967, as recounted in other sections, I had an intuition about codon-anticodon recognition, and from this, about recognition processes in general, which ultimately led to some of my most lasting contributions to molecular biology. At once, I abandoned most of my interest in stereochemical interactions between amino acids and nucleic acids. I maintained though a prejudice in favour of reverse translation, which is reflected in a sentence, Section 8, page 73 of the "missing triplet hypothesis" [9]. After the publications of my ideas on codon-anticodon recognition, I had a period of very intense thinking on the genetic code and its origin. At that time, I completely abandoned my ideas on reverse translation, but notice that other authors took reverse translation seriously, e.g. [10]. I had realized that the reverse of the peptide synthesis reaction was not reverse translation, but the hydrolysis of the terminal amino acid, coupled to a backward motion of the peptide chain ! I was in fact starting to think in terms of the kinetics of polypeptide synthesis on the ribosome.

Having rejected the "lock-and-key" very static imagery, I no longer thought it important to understand why a particular amino acid corresponded to a particular codon. The "dictionary" aspect of the code was played down, and I started thinking how the code came into being as a dynamical process. I had a theoretical idea about how non-coded peptide synthesis could evolve into coded peptide synthesis (see Section on the origin of the genetic code below). I put these ideas in a manuscript, which was submitted to the nascent "Journal of Molecular Evolution". It was rejected with unfair, insulting reports. Later, I incorporated my scheme in a review on codon-recognition [1], then in my book on molecular evolution [2]. There is nothing there that I might wish to retract.

In the summer 1971, I had in my hands a preprint of Manfred Eigen's hoax "Self-organization of matter and the evolution of biological macromolecules" [11]. I studied it carefully, and saw that it was scientifically vacuous. (See section below on the criticism of Eigen's work). In June 72, I gave a (critical) bibliographical seminar at the Pasteur Institute on this work. It took place in the conference room of the Molecular Biology building. It was attended by a large audience, François Jacob included.

In October 71, I joined the laboratory of François Chapeville at Institut de Biologie Moléculaire, Jussieu, Paris, in which I attempted to initiate biochemical work on

transfer RNA. I was also involved in writing the review on codon-anticodon recognition for "Progress in Nucleic Acid Research". I included there my ideas on the evolution of transfer RNA 3d structure. I do not know when exactly I got this idea. In any event, I consider that the idea is still valid [1, 2] (see section below on the evolution of tRNA 3d structure).

It is apparent, from my records, that I was following many threads at the same time. I was aware of the early results on prebiotic replication [12-14]. In June 71, I attended a prestigious conference "From Theoretical Physics to Biology" in Versailles (see the web chapter on stereoscopic vision). There, I met Leslie Orgel, he invited me for a post-doctoral period at The Salk Institute, and I prepared myself for the event. In my report to the CNRS, I read this detail which I had forgotten: it was envisioned that I would work on the possibilities of recognition of amino acids by oligonucleotides: which is the simplest nucleic acid structure capable of recognizing an amino acid? (i.e., tRNA ancestors).

I also devised an epistemological criterion for judging the validity of some proposals on the origin of life, which I called "l'axiome du choix permis" (the allowed choice axiom - see the section on "criticism of biochemical predestination").

Nevertheless, I was also interested in the non-enzymatic replication experiments. Most internucleotide linkages that were obtained at that time were of the 2'-5' type instead of the biological 3'-5' type or, even worse they were often of the 5'-5' type, that did not allow an extension of the chain beyond the dinucleotide level. I thought that the problem would be solved by using nucleotide oligomers as polymerizing units, instead of using nucleotide monomers. I wrote a little note, based upon model-building with space-filling components, explaining why in a universe of nucleic acids containing both RNAs with 2'-5' bonds, and RNAs with 3'-5' bonds, the latter should replicate more faithfully. The argument was astute, but with an unjustified hidden assumption. The reviewer pointed out this flaw, and I did not insist because, in any event I had the hope of settling the problem experimentally in the near future.

I also spent two months in Alan Michelson's laboratory , perhaps in 1972, during which I acquired some experience with oligonucleotides. In particular, I did experiments on oligomer condensation with carbodiimide.

In September 1972, I joined Orgel's laboratory at the Salk Institute in La Jolla (Southern California). I immigrated with my wife, two children, and a baby-sitter from Brittany. I came alone first, and was hosted by Daniel Blangy. Within a few days I got a car: A gorgeous 1951 pink Cadillac (a vintage coupé Sedan model), acquired from Rolf Lohrmann at the symbolic price of 75 dollars, and I rented a superb house in Del Mar, with two levels, two bathrooms and a big terrace right on the beach. At high tide, the waves were breaking on the terrace. We were coming from France with low salaries, and this represented an almost unbelievable increase in living standards. Very luckily, it was the case that the currency exchange rate between French Francs and US dollars was at that time extremely favourable to the French Francs. I had also a modest complement from the Salk Institute. But essentially we were living on the French salaries.

This was an exceptional period in my life. In Orgel's laboratory, I was pursuing two lines of research in parallel. One was the prebiotic replication of nucleic acids, in which I linked oligomers instead of monomers (see section on prebiotic replication below). The other was all that turned around the accuracy of molecular processes. I worked very

hard experimentally on the first subject, and had daily discussions with Orgel on the second subject (see the web chapter on the kinetic theory of accuracy). By March 74, I had made several theoretical breakthroughs on accuracy [15, 16], and had also nearly completed a massive work on prebiotic replication, which I completed during a subsequent period of two months at the Salk Institute (August 25th - October 25th, 1975). There was a multidisciplinary library at the Salk Institute, and I gained there contact with many important books on disciplines outside molecular biology (d'Arcy Thompson, Karl von Frisch, Bela Julesz). I also had the occasion to meet many important people who visited the Salk, or just gave a seminar.

In November 73, Leslie Orgel organized an informal meeting with top scientists such as Richard Feynmann, Alexander Rich, Murray Goodman, Stanley Miller, Don Glaser, etc. Manfred Eigen was also invited. Orgel wished to discuss the possibility of developing an advanced experimental strategy for in vitro selection of peptides with a given catalytic activity. There would be a combinatorial synthesis of peptides, and a mechanism to amplify the synthesis of the peptides having the desired catalytic activity. The meeting is mentioned in my book ([2], Princeton edition, page 88. I added:

"Half-way between the sifting method inspired by biochemistry and genetics which I propose, and the peptide-selection machine which Orgel envisaged, are various semi-selective methods. For example, a means of obtaining a peptide catalyst capable of acting on DNA would be to prepare a mixture of peptides, and filter them through a resin containing DNA which would retain peptides having affinity for it. After this selection, the peptides could be subjected to a sifting procedure to detect either destabilizing cutting peptides, or stabilizing ones."

Upon my return in France, in Chapeville's laboratory, I made several attempts to develop an experimental activity on the origins of life. My strategy, more modest than Orgel's was to prepare random peptide mixtures, and make a screening for the peptides having the highest catalytic activity (see section below on peptide catalysis). Françoise Bernardi started to work on this theme, had preliminary encouraging results, but we had difficulties identifying the most active peptides.

In March 1976, I wrote a letter to John Kendrew, then director of the European Molecular Biology Laboratory, in which I proposed to launch a team of about six individuals (three researchers, three technicians) on experimental molecular evolution. More specifically, I proposed the theme of the "search for peptides with catalytic activity by selective methods" (see Appendix). Before consulting the Advisory Committee, Kendrew wrote "I had always had the impression that your own interests were primarily in doing theoretical work, and only secondarily in experimental work. Was I wrong ? And if we came to the conclusion that a group on the scale you suggest could not be fitted into the laboratory, would you also be interested in a smaller operation with a more theoretical slant, or even a completely theoretical one ?". To this, I answered in a two-pages letter, to which I would not change a line to-day. In particular, I wrote : « When I was working at the Salk Institute it appeared in the course of many discussions with Leslie Orgel that some experiments of fundamental importance were never carried out. The reason, according to him was that theoreticians are not good enough experimentalists to do them themselves, and experimentalists are not good enough theoreticians to grasp the importance of such experiments ». The project was not retained by the Advisory Committee. Kendrew wrote that "everybody

thought it a most interesting and worthwhile one, and we had some comments from Francis Crick in the same direction". However, the committee doubted whether the proposed work "would interact much with the work of other groups already planned".

So, this line of work came to an end. I started many things, and in 1978, I found myself at the head of a small team named "biochimie de l'évolution". Two researchers were doing experimental work on DNA polymerase kinetics (see the web chapter on the enzymology of DNA and RNA polymerases), a technician was doing computer work on the prediction of RNA secondary structures (see section on bio-informatics), and there was a young student, Philippe Marlière who had both an interest in bio-informatics, and an interest in the origins of life. He had a solid background in chemistry, and started a project of his own, on analogues of nucleic acids, with simpler backbones. Most of the work was done in external laboratories, having better competence in chemistry. Marlière first constructed models of nucleic acids with a peptide backbone. He saw that "some simple polymers to which nucleotide bases where hooked, could well form good helices when the repeating backbone unit contained an even number of atom links" (as reported in [2], Princeton edition, page 62). I do not remember precisely what he did after that along this line (he moved to the Pasteur Institute).

I maintained a theoretical interest for the subject, which is reflected in my book, which I started to write in 1977 [2]. After the discovery of RNA catalysis in living cells, I considered making with RNA what I hat attempted to do with peptides: look for catalytic activities in randomly synthesized RNAs. I took in my group a laboratory technician to initiate the work (June 1982), and made, in this choice, one of the biggest mistakes in my scientific career. Although we had at that time good competence in all the molecular biology techniques we needed, there was always something wrong in his experiments. Ultimately, I had the revelation that he was, probably, a drug addict. Fortunately, he left in October 84 for a laboratory on the Mediterranean coast. In the mean time he had not produced a single valid experiment.

In 1983, I think, Marie-Christine Maurel, then a college teacher, came to my laboratory, saying that she was very interested in the origins of life, and was prepared to all sacrifices to work on the subject. I asked her to acquire first some laboratory practice, and she spent two years preparing a diploma in Alain Favre's laboratory. Then she started working in my lab on catalysis by RNA molecules. We transposed to RNA the strategy which Françoise Bernardi had applied to peptides around 1975. Nothing worked as well as one of the controls — adenine. I then remembered a result by Fuller, Sanchez and Orgel on "prebiotic adenosine" [17, 18], and worked out the implications. So, we published in 1987 one of the earliest papers ever published on artificial RNA catalysis [4] (see section on RNA catalysis below). Subsequently, she worked mostly on her own. I considered that I did not have enough merit in her production to put my name on her articles.

In 1988 I became more and more involved in cognitive sciences, and at the end of 1991, I resigned from my position of leader of the "biochimie de l'évolution" group. I joined the physicists at Ecole Normale Supérieure. In 2000, I started writing an updated version of my book on evolution [2], and completed a first draft, invited by Luna Han of Wiley and Sons. On this occasion I sharpened a number of ideas, in the light of the last 20 years of progress, and these were turned later into publishable articles [19, 20]. However, the book was not published, possibly due to the xenophobic craze that

followed the September 11th, 2001 events in the United States.

ORIGIN OF THE GENETIC CODE

My ideas on the origin of the code emerged from a gedanken experiment of "evolutionary regression". I was wondering how a primitive bacterium, in which proteins were synthesized at a low accuracy level could escape from the error-catastrophe. I thought that part of the answer was in the possibility that primitive bacteria used much shorter proteins than they use to-day. The point had been made in the MS that had been submitted to the Journal of Molecular Evolution, and rejected. Soon afterwards, Ycas published in the Journal of Theoretical Biology his own gedanken experiment on evolutionary regression [21]. To my knowledge, experiments in which one would follow the adaptive evolution of bacteria growing under ever decreasing accuracy levels, in Richard Lenski's spirit [22] have yet to be done.

From there, I started thinking about an evolutionary path from non-coded peptide synthesis to coded translation. Here is the theory, reproduced from [2], Princeton edition, page 86. I have rewritten here each sentence as a separate paragraph.

"I am supposing that we have primitive synthesis, catalysed by a crystal or whatever, of a defined dipeptide, for example methionine-tyrosine.

"Later, this synthesis becomes complicated by one step; a tripeptide is formed: methionine-tyrosine-valine, the bond between the second and third amino acids being facilitated by the presence of an oligonucleotide, let us say AGCG.

"There is no coding relationship, just a coupling between two events: binding of a cofactor AGCG and addition of valine to the dipeptide methionine-tyrosine.

"At a third stage a tetrapeptide methionine-tyrosine-valine-histidine, or the same with glutamine as the fourth amino acid, is synthesized.

"As for the tripeptide, synthesis would depend on the presence of an oligonucleotide, the nature of which would determine which of the two tetrapeptides were made.

"We still have no genetic code, but peptide synthesis with 'options' on the fourth position.

"We now see the distance which separates this synthesis from one using a code: the system with options has to become repetitive so that, from the fifth amino acid on for example, the same regular process is reiterated allowing the next amino acids to be put into place.

"Later, the regular iteration can be made to start right from the beginning of the chain.

"To sum up, what is fundamental in the genetic code, from my point of view, is not linear correspondence between a messenger RNA and a protein but the existence of an 'elongation cycle'; a reiterative process which causes the (n)th amino acid and the (n+1)th to be added on in exactly the same manner."

For reasons unknown to me, it seems that this hypothesis on the origin of genetic code has never been quoted, and not even plagiarized. Yet, I think that what I did, which was to de-emphasize the code as a linear correspondence, and reinterpret it as an iterative process is still worth serious consideration.

In [2], I noted briefly (in chapter 11) that Woese « very clearly affirmed the primacy of the process over the dictionary ». My theory can be viewed as a particular implementation of Woese's doctrine [23].

THE EVOLUTION OF TRANSFER RNA 3d STRUCTURE.

Here again, I quote the two paragraphs in [2], Princeton edition, pages 86-87 in which the hypothesis is given (see also Figure 16, page 88 in [2], and Figure 5, page 326 in [1]).

"We can imagine a primitive tRNA made of two pieces; a short oligonucleotide to which an amino acid is attached non-specifically and a hairpin - that is, the end of a nucleic chain folded on itself and forming a loop in its middle:

(short figure provided, not reproduced here)

"We know that quite often a nucleic acid double helix can be associated with a supplementary filament which come to lie in the wide groove of the double helix, forming a three filament structure.

"The primitive tRNA would have been formed from the association of the hairpin with the oligomer carrying the amino acid, energy for the interaction being derived principally from formation of the triple helix between nucleic chains.

"The amino acid, attached to the oligomer, would come into contact with the loop, which would influence its position in space for various reasons (attractions due to opposite charges, repulsions, steric hindrance).

"This primitive tRNA, without an anticodon, combines a general principle of attraction without specificity (between nucleotide chains) with a principle of specific positioning not requiring attraction (interaction between the amino acid and the loop).

"From here, it was possible to conceive successive enlargements of the molecule, according to the order of events shown in Fig. 16."

The last sentence may appear very elliptical here, but there lies some of the strength of the proposal. The triple helix configuration is in fact at the core of tRNA 3d structure (the DiHU stem + the extra loop), a field in which I was particularly competent (see the Section on RNA structure). The model was developed before the elucidation of tRNA 3d structure, and it remains essentially unchanged with the modern structure.

Similar proposals were made later by several other authors, in prestigious journals, without giving credit to my earlier work.

Twenty five years later, I was invited by Piet Herdewijn to contribute to a volume of essays in the honour of Leslie Orgel. On this occasion, I updated my ideas, taking into account the recent work on catalytic RNAs, and in particular some publications on self-aminoacylating RNAs [24, 25]. I thus wrote in [19] that I was tempted to make a few amendments to my initial model : *« 1) The amino acid could be chemically linked to the second oligomer after binding, but not before binding ; both options are worth keeping in mind. 2) The recognition site for the amino acid on the first RNA need not be a hairpin loop, but might as well be an internal loop or any structural feature on the RNA complex. 3) Similar complexes with one RNA playing the role of the scaffold, and one or two short RNA oligomers carrying amino acids, might turn out to be efficient*

catalysts ».

PREBIOTIC REPLICATION

Soon after my arrival in Leslie Orgel's laboratory, it was decided that I would work on non-enzymatic replication of nucleic acids, by condensing oligonucleotides on a polymer matrix, instead of condensing monomers. The idea that oligonucleotide ligation was the primitive form of replication was natural to me. Prior to my work, most prebiotic replication experiments had been done with monomers, using homopolymeric templates (Poly(A), Poly(U), Poly(C)). Some success had been achieved in linking purine residues (GMP and AMP) opposite the complementary templates [13, 14]. However, the incorporation of pyrimidines was problematic. UMP could be incorporated, but in a triple helix configuration. I expected that U could be incorporated as part of an oligonucleotide, in a double helix configuration. Naylor and Gilham [12] had been able, in 1966 to condense two hexamers of T on a dodecamer of dA.

There was also the problem that very little 3'-5' bonds were formed in the prebiotic replication experiments with monomers. (Later, Usher and McHale, in a much praised 1976 article « proved » that this was as expected, chemically, but that in the long run, 3'-5' would outcompete the others, being more resistant to hydrolysis [26]).

I expected that if oligonucleotides were used instead of monomers, the geometric constraints at the site of condensation might be different, and might favour the 3'-5' bond. Orgel was, on the other hand, mostly concerned with the search for the perfect condensing agent. At that time, he had developed with Rolf Lohrmann a strategy of derivatization of the nucleotide units, in which an imidazole group was hooked onto the 5' phosphate of the nucleotide, and they were studying a still more powerful derivatization, with methyl-imidazole. The preparation of these active monomers required a large number of horrifying procedures in organic chemistry, in which the compounds were dissolved into anhydrous organic solvents.

For my work on oligomer condensation, I needed to prepare a number of different oligomers and polymers, and I did it using some molecular biology techniques, such as enzyme purification and radioactive labelling of a terminal phosphorus. There was no expertise for this in the lab, except at the beginning, thanks to the presence of a German post-doc, the late Christoph Biebricher, who wrote numerous articles with Manfred Eigen on RNA replication, e.g. [27]. I went through all the molecular biology preparations honourably (some of my oligomer preparations were even used, subsequently, by Hiroaki Sawai in the same lab). Then, I had to derivatize all the oligomers, using the established organic chemistry procedures. Then I had to combine oligomers and polymers, and study many condensation reactions. Then I had to analyse the products of the reactions. The work was a massive one, and on several occasions, I stayed in the lab until midnight. On the whole, the results were quite good, and in the direction of my expectations.

They are summarized as follows [3]:

"We have studied a number of condensation reactions involving ImpU, ImpT, ImpC, ImpA, ImpG, ImpUpG and ImpCpA as activated nucleotide donors and a variety of homo- and hetero-polynucleotides as templates. We did not obtain any evidence of a template effect with ImpU and ImpT, but observed some condensations of ImpC with

GpG on appropriate templates. ImpA and ImpG take part in a number of more or less efficient template-directed reactions, as do ImpUpG and ImpCpA.

"Our results suggest that, on the primitive Earth, pyrimidine nucleotides could most easily have been incorporated into polymers as constituents of short oligomers, which contained one or more purine nucleotide. The linkage of the product depends strongly on the nature of the substrates; the percentage of the natural 3'-5' linkage was, in some cases, less than 10% and, in others, as high as 70%. Wobble-pairing was often very effective in promoting condensations, suggesting that transition mutations would have been very frequent in prebiotic polynucleotide replication".

I feel that this article stands very honourably in the line of Orgel's production. It was published in 1978, three years after the work was fully completed. It seems that Orgel felt very insecure with my results that showed, in many cases, a high proportion of 3'-5' linkages, but he became progressively convinced, from further experiments performed in his laboratory.

There is a printing mistake in this paper, which was made by the publisher after receiving the corrected proofs. In many cases, I had done experiments and controls in which the incubation times were identical. In [3], Table 1, page 96 all incubation durations of 32, 33 or 36 days are in fact durations of 2, 3 or 6 days respectively. I thought that Leslie would request a correction, but he did not, and I left things as they were, expecting that the readers would correct by themselves (which they did not).

In line with the current 'RNA world' theories of the origins of life, there has been a burst of papers on artificial RNA polymerases, obtained by selection from enormous pools of randomized sequences (e.g., [27-30]). However, most of the results, in this area are in fact not related to monomer by monomer additions, but to ligation of preexisting large oligomers.

CATALYSIS BY PEPTIDES.

At the end of 1974, Françoise Bernardi initiated in my embryonic laboratory a program of search for catalytic activities in small peptides. After some unsuccessfull attempts with peptides generated by protein hydrolysis, the peptides were prepared from several separate amino acid mixtures. Each mixture was formed by random condensation of 2 or 3 different amino acids, chosen among the 20 canonical amino acids of the genetic code. The peptide mixtures were spotted on chromatographic paper, and partially separated by paper chromatography. They were revealed, I think, by the ninhydrin reaction. The tested catalytic activity was the hydrolysis of para-nitro-phenyl-acetate, a colourless reagent, well known in studies of artificial catalysis, that turned brown when it was hydrolysed .

Very rapidly, we saw that peptide mixtures that contained serine, histidine and methionine gave the best results. One of the fractions, eluted from paper chromatography, was (slightly) more active, per mole of amino acid, than the most active model peptide studied so far. However, we were not able to purify this activity, or learn more about it. This type of difficulty was one of the main reasons for abandonning this line of work.

RNA CATALYSIS

When it became clear, from the work of Tom Cech, Sidney Altman and co-workers that some RNA catalysts were used in living cells, I thought of transposing to RNA the strategy I had initiated with peptides.

Marie-Christine Maurel did all the experimental work. She used the same model-reaction (the hydrolysis of para-nitrophenyl acetate), and studied the catalytic activity, in this reaction, of ribosomal RNA or transfer RNA fragments. The results were negative. Then she used random polynucleotides of various nucleotide composition, and the results were again negative. As a matter of fact, the most active compound was the control, pure adenine! I then remembered a few things I had learnt in Orgel's laboratory, and put them together (i) imidazole is a powerful catalyst. It is the active principle in histidine, which is often found in the catalytic centres of proteins (ii) adenine is composed of a 5-membered and a six-membered ring. The 5-membered ring has the imidazole structure. (iii) in standard RNA structures, the imidazole activity of adenine is masked, because it is through the imidazole group that adenine is linked to ribose (iv) however, in prebiotic condensations of adenine with ribose, the ribose is hooked to the N6 of adenine, leaving the imidazole group free [17, 18]. From this, one could expect that the "prebiotic" adenosine (i.e., N6-ribosyl adenine) could be a good catalyst. It turned out to be half as good as histidine [3]. At the end of the note describing the work we wrote:

"It is difficult however to conceive a precise prebiotic replication of polymers containing four canonical nucleotides and, in addition, modified nucleotides with good catalytic potential. One might envisage instead the existence of two classes of compounds. On one side, there would be regular nucleic acids, of the size of tRNA molecules, providing some kind of scaffold. On the other, there would be small oligonucleotides with a nucleotide catalyst at their end. The oligonucleotides would base-pair to the scaffold. If two or three such oligonucleotides become fixed, they might form the equivalent of the catalytic site of an enzyme".

Maurel pursued the work on catalysis by RNA plus adenine. For instance, she and her co-workers isolated by in vitro selection [31] two adenine-dependent hairpin ribozyme [31], capable of self-cleavage according to different pathways.

Recently [19], I made an alternative suggestion: That there were replicable RNA sequences, catalytically inactive in their standard form, but active when, by mistake, some active nucleotide analogue was incorporated at some specific position :

« Finally, I note that in most writings on ribozymes in the context of the RNA world, the same RNA molecules are assumed to be replicated faithfully, and assumed to function as ribozymes. Here also, alternatives should be considered. For instance, a nucleic acid could be used as a template for its own replication, and as a template for the synthesis of a different kind of nucleic acid, used for another purpose. (…). Furthermore, it is conceivable that occasionally, during the replication of RNA templates, a non-standard nucleotide with interesting catalytic activity could be incorporated at a crucial position, instead of the standard one. This would be in effect an alternative implementation of dual replicating and catalytic functions derived from a same RNA molecule. RNA polymers containing all kinds of modified nucleotides may be synthesized by transcription of DNA molecules with cellular polymerases (e.g., [32]), and this could lead into a new subfield of ribozyme studies, in which a large pool of

RNA molecules containing a few modified nucleotides in various positions would be prepared by replication of a single RNA or DNA sequence »

AGAINST THE RNA WORLD HYPOTHESIS

Adapted from my 2007 review in [19]

There is an argument which is, in my opinion, rather damaging to the RNA world hypothesis. The argument is derived from experimental results and theoretical estimates by Carothers et al. [33]. These authors studied how the efficiency of GTP-binding RNA aptamers varied in relation to the sizes of the pools from which they were selected. They found that about « 10 bits of additional information are required to specify RNA structures with 10-fold better binding to GTP, over a range of 3 orders of magnitude of binding affinity ». From there, I deduced that nucleotide substitutions or additions produce roughly affinity increases or decreases by steps of 0.2 kcal, merely one fifth the value of an additional hydrogen bond. Similarly, Paegel and Joyce [28] found that 11 mutations were needed to obtain a 90 fold improvement in affinity, corresponding, roughly, to a global 2 kcal improvement.

If we accept this interpretation, then even if the chances of finding an interesting ribozyme within a pool are high enough, very little improvement can be gained by changing just one or two nucleotides in the sequence. The improvements being small, changes will not produce substantial selective advantages that could be used in a darwinian mode of evolution. Indeed we know that drift prevails in a replicating population when the selective advantage is less than the reciprocal of the size of the population [34]. The evolutionary problem with catalytic RNA (in a darwinian context) is not that beneficial mutations are rare. It is that gains acquired by mutations are very small, so they have little chance of spreading into the population.

We can now better appreciate the advantage of proteins from this perspective. The amino acid side chains offer a wide spectrum of chemical reactivities, and they have substantial differences in polarity and hydrophobicity. So, starting with a catalytically efficient polypeptide, although beneficial amino acid replacements may be rare, when they occur, they have good chances of generating substantial selective advantages. It must also be kept in mind that what matters, in a biological context, is not just binding affinity for a cognate ligand, but also specificity, which is the capacity to discriminate between cognate and non cognate ligands (see, e.g. [15, 16]). Let us consider, as before that binding energy gains or losses go by steps of 0.2 kcal. A loss of 0.2 kcal for the incorrect substrate may, in favourable kinetic cases correspond to a 40% decrease in error-rate [15, 16]. This can lead to a substantial selective advantage, if the incorrect product is frankly noxious. This is precisely the situation with cellular ribozymes, most of them being involved in RNA splicing. They need not work very fast, but they must carry out their excision activities at very precise places.

I expressed previously strong reservations about naked genes theories, and my adhesion to a view in which darwinian evolution took place at a rather advanced stage in cellular evolution [2]. Thus, I do acknowledge the importance of the work on primitive metabolism, energetics, and primitive cellular organization. Maintaining the focus on proteins and nucleic acids I also do acknowledge the value of the effort invested by so many authors in understanding how the two classes of molecules came to be tied into

complex relationships which led to bootstrapping and the emergence of the genetic code. I will discuss here at a more elementary level some issues on primitive replication and catalysis in a way which associates the protein and the RNA worlds.

Before the advent of naked genes theories, whole cell duplications could be attributed to the conjunction of two processes : growth and fragmentation. Growth by aggregation of compounds with similar properties (for instance, hydrophobic material in an aqueous medium), followed by fragmentation above a certain size due to physical constraints is conceivable, both in the organic and the mineral worlds, and such processes have often been invoked in relation to prebiotic organization. In this spirit, can we conceive that « pure » residue by residue replication of nucleic acids was preceded by alternative forms of self-propagation ?

Going backwards from the contemporary mode of replication, it seems natural to introduce oligonucleotide ligation reactions in the replication process. It is usually objected that replication by ligation is very inaccurate, and my own results with Orgel [3] point in this direction. However, if we believe in early ligation reactions, we can as well believe in early excision reactions which would extract the poorly paired regions in structured single-stranded RNAs or in imperfect double-stranded RNAs. We are used to think of primitive replication as though it were occurring in a chemical reactor containing an initial supply of templates, regularly fed with nucleotides or oligonucleotides to copy the templates. If energy was abundant, under prebiotic conditions, then polymers rather than monomers may have been the raw material. So, our reactor was perhaps fed with polymers, and the first catalytic reactions to consider might have been extractions of oligonucleotides (or peptides) from the polymers.

Under the joint actions of excision and ligation, RNAs with random sequences would have had a tendency to evolve towards a folded state with increasingly long complementary sections. For instance, « At first, nucleic acid chains of any sequence whatever fold up or associate in twos or more, matching any complementary sequences as best as they can. Afterwards, the badly paired segments are excised and finally the holes are refilled. Thus any sequence evolves towards a complementary double chain. The chains separate, fold up on themselves and the process starts again. » [2].

The intermediate stage, in which a good deal of an RNA sequence contains double-stranded sections, while some small sections are still forming loops is interesting, because aptamers or ribozymes derived from pools of random sequences obey this pattern, including amino acid binding aptamers [35, 36], self-aminoacylating RNAs [24, 25] and ribozymes that mediate peptide bond formation [37, 39].

Thus, I do believe that ribozyme studies should pay more attention to the advantages of combining different RNAs to form a catalytically active complex. The interest of « recombining » covalently short RNA pieces has been acknowledged [40-42], but the idea has not been pushed to the point of considering non covalent associations between separate RNA molecules.

It is a pleasure to note that many authors take now their distances with the RNA world, and insist on alternatives that involve primitive metabolism, peptide networks, or proto-cells [43 – 46].

CRITICISM OF EIGEN's WORK

Manfred Eigen has launched what looks like a theory of prebiotic replication of nucleic acids [11]. It is in reality a work on the population genetics of bacteria: autonomous organisms that self-replicate accurately and occasionally mutate. The property he attributes to "RNA species" are just those of bacterial sub-populations. This work is, in my opinion, one of the biggest scientific mystifications of the twentieth century. What is tragic there, is that so many uninformed scientists have lost so much time refining Eigen's calculations, with the sincere belief that they would contribute to the origins of life. This is discussed in [2], Princeton edition, page 56, in which I wrote:

"Recently some physicists attempted a take-over bid of the study of evolution, assuming that, with their superior science of differential equations, they would be able to reveal the essential truths of the phenomena. What came out of this was the application of mathematical treatments to the simplest test case of population genetics: bacterial competition in the chemostat. Or rather, after having changed the words, these physicists described competition between DNA molecules in the prebiotic soup, attributing to the molecules the same reproductive properties as to bacteria".

See also [2], Princeton edition, pages 89-91, in relation to the origin of the code.

Of course, Manfred Eigen was aware of my book, he read it , and he took much inspiration from it, although he never quoted it. In 1987, I was invited to the 52nd Cold Spring Harbor Symposium on Quantitative Biology. In the leaflet announcing the program of the meeting, a talk by Eigen was scheduled with a title about sequence space, a new concept in molecular evolution. I kept quiet during the talk, but towards the end of the meeting, I informed Alan Weiner (who was in charge of the concluding talk) of my earlier chapter on the sequence space in [2]. Or perhaps I informed him after his talk. In any event, Alan Weiner was perfectly honest, he quoted Maynard Smith and my book, writing in his published summary [47], p.938: "(The notion of amino acid sequence space was originally invented by Maynard Smith [see Ninio 1983], and Eigen introduced sequence space early in the symposium …)". Furthermore, Eigen had to change the title of his published text, it became "new concepts for dealing with the evolution of nucleic acids" [48].

This happened nearly thirty years ago. However, Eigen is still highly praised by new generations of biologists who did not have the opportunity to inquire about the history of the subject. He gained recently a prestigious advocate, Eugene Koonin, a prolific biologists, author of seminal papers on the minimal set of genes required to make a bacterium work and on horizontal gene transfer. In his book "The logic of chance" [49], Eugene Koonin goes as far as equating Darwin with Eigen, writing about the "Darwin-Eigen" theory of evolution and about the "Darwin-Eigen" transition from pre-darwinian life to life that would follow a mechanism of replication and selection.

CRITICISM OF CHEMICAL PREDESTINATION

In theories of the origins of life, one often encounters "why" questions of the kind: "why nucleic acids use the 4 canonical nucleotides" or "why the proteins use this set of 20 amino acids" or "why particular codons correspond to particular amino acids" ?

There are answers such as:

- the canonical nucleotides, or the canonical amino acids were those obtained most

easily under prebiotic conditions

- the canonical nucleotides, or the canonical nucleotides are used because they are the best suited for their functions.

What I felt was that both answers could not be correct simultaneously, that it would be too much of a coincidence if the products generated most abundantly at one step in evolution were exactly the most suited products for the next step.

So, I thought one could use quantitative arguments to remove the "overdeterminacies" in the theories. I wrote an article on this, and sent it to François Jacob, with the hope he would submit it to PNAS. But he said that there would be problems with such philosophical articles. This line of thinking is somewhat reflected in [2], Princeton edition, bottom of page 77.

MISCELLANEOUS

Discussions of the evolution or the origine of the genetic code, see [1, 2, 50-53].

Popular science articles on the origin of life [54-57].

Reviews of books on the origin of life and evolution [58]

REFERENCES

[1] Ninio, J. (1973) Recognition in nucleic acids and the anticodon families. Progress in Nucleic Acids Research and Molecular Biology 13, 301-337.

[2] Ninio, J. (1979) *Approches moléculaires de l'évolution*. Masson, Paris. (1982) *Molecular approaches to evolution.* Pitman, London. (1983) updated edition, Princeton University Press, Princeton. (1984) Japanese translation. Kinokuniya Press, Tokyo.

[3] Ninio, J. and Orgel, E. (1978) Heteropolynucleotides as templates for non-enzymatic polymerizations. Journal of Molecular Evolution 12, 91-99.

[4] Maurel, M-C. and Ninio, J. (1987) Catalysis by a prebiotic nucleotide analog of histidine. Biochimie 69, 551-553.

[5] Crick, F. H. C. (1967) Origin of the genetic code. Nature 213, 119.

[6] Orgel, L. E., and Crick, F. H. C. (1993) Anticipating an RNA world. Some past speculations on the origin of life: where are they today? FASEB Journal 7, 238-239.

[7] Pelc, S. R., and Welton, M. G. E. (1966) Stereochemical relationship between coding triplets and amino-acids. Nature 209, 868-870

[8] Crick, F. H. C. (1967) An error in model building. Nature 213, 798.

[9] Ninio, J. (1971) Codon-anticodon recognition : the missing triplet hypothesis. Journal of Molecular Biology 56, 63-82.

[10] Cook, N. D. (1977) The case for reverse translation. Journal of Theoretical Biology 64, 113-135.

[11] Eigen, M. (1971) Selforganization of matter and the evolution of biological

macromolecules. Naturwissenshaften 58, 465-523.

[12] Naylor, R., and Gilham, P. T. (1966) Studies on some interactions and reactions of oligonucleotides in aqueous solutions. Biochemistry 8, 2722-2728.

[13] Sulston, J., Lohrmann, R., Orgel, L.E., and Miles, T. H. (1968) Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. Proceedings of the National Academy of Sciences of the USA 59, 726-733.

[14] Sulston, J., Lohrmann, R., Orgel, L.E., Schneider-Bernloehr, H., Weiman, B. J., and Miles, T. H. (1969) Non-enzymic oligonucleotide synthesis on a polycitydylate template. Journal of Molecular Biology 40, 227-234.

[15] Ninio, J. (1974) A semi-quantitative treatment of missense and nonsense suppression in the strA and ram ribosomal mutants of *Escherichia coli* . Evaluation of some molecular parameters of translation *in vivo*. Journal of Molecular Biology 84, 297-313.

[16] Ninio, J. (1975) Kinetic amplification of enzyme discrimination. Biochimie 57, 587-595.

[17] Fuller, W. D., Sanchez, R. A., and Orgel, L. E. (1972) Studies in prebiotic synthesis: VI. Synthesis of purine nucleosides. Journal of Molecular Biology 67, 25-33

[18] Fuller, W. D., Sanchez, R. A., and Orgel, L. E. (1972) Studies in prebiotic synthesis. VII. Solid-state synthesis of purine nucleosides. Journal of Molecular Evolution 1, 249-257.

[19] Ninio, J. (2007) Errors and alternatives in prebiotic replication and catalysis. Chemistry & Biodiversity 4, 622-632.

[20] Ninio, J. (2010) Frail hypotheses in evolutionary biology. PloS Genetics 6 issue 9, e1001067.

[21] Ycas, M. (1974) On earlier states of the biochemical system. Journal of Theoretical Biology 44, 145-160.

[22] Cooper, T. F., Rozen, D.E., and Lenski, R. E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proceedings of the National Academy of Sciences of the USA 100, 1072-1077.

[23] Woese, C.R. (1967) The genetic code. Harper, New-York.

[24] Illangasekare, M., Kovalchuke, O., and Yarus, M. (1997) Essential structures of a self-aminoacylating RNA. Journal of Molecular Biology 274, 519-529.

[25] Ramaswamy, K., Wei, K., and Suga, H. (2002) Minihelix-loop RNAS: minimal structures for aminoacylation catalysts. Nucleic Acids Research 30, 2162-2171.

[26] Usher, D. A., and McHale, A. H. (1976) Hydrolytic stability of helical RNA: a selective advantage for the natural 3',5'-bond. Proceedings of the National Academy of Sciences of the USA 73, 1149-1153.

[26a] Biebricher, C. K., Eigen, M., Gardiner Jr, W. C. (1985) Kinetics of RNA replication: competition and selection among self-replicating RNA species. Biochemistry 24, 6550-6560.

[27] Green, R., and Szostak, J. W. (1992) Selection of a ribozyme that functions as a

superior template in a self-copying reaction. Science 258, 1910-1915.

[28] Paegel, B. P., and Joyce, G. F. (2008) Darwinian evolution on a chip. PloS Biology 6, article 900

[29] Lincoln, T. A., and Joyce, G. F. (2009) Self-sustained replication of an RNA enzyme. Science 323, 1229-1232.

[30] Sczepanski, J.T., and Joyce G.F. (2014) A cross-chiral RNA polymerase ribozyme. Nature 515, 440-442.

[31] Meli, M., Vergne, J., and Maurel, M.-C. (2003) In vitro selection of adenine-dependent hairpin ribozymes.

[32] Kozlov, I. A., Zielinski, M., Allart, B., Kerremans, L., Van Aerschot, A., Busson, R., Herdewijn, P., and Orgel, L. E. (2000) Nonenzymatic template-directed reactions on altriol oligomers, preorganized analogues of oligonucleotides. Chemistry-A European Journal 6, 151-155.

[33] Carothers, J. M., Oestreich, S. C., Davis, J. H., and Szostak, J. W. (2004) Informational complexity and functional activity of RNA structures. Journal of American Chemical Society 126, 5130-5137

[34] Kimura, M. (1983) The neutral theory of molecular evolution. Cambridge University Press, London.

[35] Connell, G. J., Illangaserkare, M., and Yarus, M. (1993) Three small ribooligonucleotides with specific arginine sites. Biochemistry 32, 5497-5502.

[36] Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A., and Famulok, M. (1996) RNA aptamers that bind L-argine with sub-micromolar dissociation constants and high enantioselectivity. Nucleic Acids Research 24, 1029-1036

[37] Zhang, B., and Cech, T. R. (1997) Peptide bond formation by *in vitro* selected ribozymes. Nature 390, 96-100.

[38] Illangasekare, M., and Yarus, M. (1999) A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. RNA 5, 1482-1489

[39] Sun, L., Cui, Z., Gottlieb, R. L., and Zhang, B. (2002) A selected ribozyme catalyzing diverse dipeptide synthesis. Chemical Biology 9, 619-628

[40] Burke, D. H., and Willis, J. H. (1998) Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. RNA 4, 1165-1175

[41] Wang, Q. S., and Unrau, P. J. (2005) Ribozyme motif structure mapped using random recombination and selection. RNA 11, 404-411

[42] Striggles, J.C., Martin, M. B., and Schmidt, F. J. (2006) Frequency of RNA-RNA interaction in a model of the RNA world. RNA 12, 353-359

[43] Dyson, F. J. (1985) Origins of life. Cambridge University Press, London

[44] Lifson, S. (1997) On the crucial stages in the origin of animate matter. Journal of Molecular Evolution 44, 1-8.

[45] Carny, H., and Gazit, E. (2005) A model for the role of self-assembled peptides in the very early stages of the origin of life. FASEB Journal 19, 1051-1055.

[46] Segré, D., and Lancet, D. (2000) Composing life. EMBO Reports 1, 217-222.

[47] Weiner, A. M. (1987) Summary. Cold Spring Harbor Symposia on Quantitative Biology 52, 933-941.

[48] Eigen, M. (1987) New concepts for dealing with the evolution of nucleic acids. Cold Spring Harbor Symposia on Quantitative Biology 52, 307-320.

[49] Koonin, E. V. (2011) The logic of chance. The nature and origin of biological evolution. FT Press Science, Upper Saddle River, New Jersey.

[50] Ninio, J. (1975) La précision dans la traduction génétique. Dand *Ecole de Roscoff 1974. L'évolution des macromolécules biologiques*. (Sadron, C., ed.) pp. 51-68. CNRS, Paris.

[51] Ninio, J. (1975) Considerations on the problem of the joint evolution of two different translation apparatuses within the same cell. In *Molecular biology of nucleocytoplasmic relationships* (Puiseux-Dao, C., ed.) pp. 31-39, Elsevier, Amsterdam.

[52] Ninio, J. (1990) The revised genetic code. Origins of Life and Evolution of the Biosphere 20, 167-171.

[53] Ninio, J. (1986) Divergence in the genetic code. Biochemical Systematics and Ecology 14, 455-457.

[54] Ninio, J. (1976) Evolutionnisme et origines de la vie. La Recherche 7, 325-324.

[55] Ninio, J. (1976) Les origines de la vie. Echos Elf-Aquitaine 38, 37-44.

[56] Ninio, J. (1993) Algunos artificios de la vida/ Alguns artificis de la vida/ Some of life’s artifices. In *Artificial Life*, pp. 44 - 51, catalog of the exhibit Art Furtura 93, Barcelone.

[57] Debate: (1992) L'évolution du vivant (B. Brun, J.-F. Gérard, P.-H. Gouyon, J. Ninio, P.-Y. Quenette, P. Vernier). Agone 8/9, 9-59.

[58] BOOK REVIEWS:

(1980) The origin of life. A warm little pond. (C.E. Folsome) FEBS Letters 112, 116-117.

(1980) The evolution of adaptation by natural selection. (J.M. Smith and R.E. Holliday, eds). FEBS Letters 120, 306.

(1981) The hypercycle. (M. Eigen and P. Schuster) Biochimie 63, N°6, pp. XXIV-XXV.

(1981) The origins of life and evolution. (H.O. Halvorson and K.E. van Holde, Eds). FEBS Letters 136, 186.

(1982) Evolution now. A century after Darwin (J.M. Smith, ed.) FEBS Letters 150, 263.

(1983) The physicochemical factors of biological evolution (S.E. Shnol). Biochimie 65 N°3, XI-XII.

(1983) Genetic Takeover and the mineral origins of life (Cairns-Smith). FEBS Letters 154, 219-220.

(1983) Molecular theory of evolution (B.-O. Küppers). Biochimie 65 N° 8-9, XVII.

(1985) A stir in the primeval soup: Seven clues to the origin of life (A.G. Cairns-Smith). Nature 318, 119-120.

(1987) The semantic theory of evolution (M. Barbieri). FEBS Letters 222, 220.

(1990) La vie plus têtue que les étoiles (F. Trémollières) Alliage 6, 115.

(1991) Une aurore de pierres. Aux origines de la vie (A. Danchin) Pour la Science 167. 106.

APPENDIX :

Quotation from a letter of application to the European Molecular Biology Laboratory (EMBL) in 1976:
Search for peptides with catalytic activities by selective methods. The initial motivation of this work is related to the problem of the origin of the genetic code. Theories of the evolution of the translation apparatus can be distinguished according to the hypothesized class of products that the apparatus was making: “statistical” proteins, repeating polypeptides, or small peptides. In every case, it is assumed that if a member of the class is chosen at random, the chances that it will possess an interesting catalytic activity will be small, but non negligible. Thus, I propose as an experimental approach to evolutionary problems, an exploration of the “protein space”, with the aim to obtain quantitative informations: with what probability is a given catalytic activity found within a class, what is the probability distribution of enhancement factors, what specificity can be expected?

A number of research groups, working on enzyme catalysis have attempted to construct simple peptides or polypeptide catalysts by chemical design, with some positive but generally disappointing results. I suggest the use of screening and selective methods instead.

Screening methods are inspired by a classical approach in molecular biology: one looks for a given enzymatic activity within a crude cell extract, or after fractionation. If it is found, one pushes the purification further, with the hope to obtain the pure enzyme and analyze it thoroughly. Similarly, one would start with mixtures of peptides. For instance, one would synthesize N mixtures of random peptides each mixture containing only three different amino acids. The N mixtures would be tested alone or in conjunction with each of p metal ions for the presence of various catalytic activities.

If some interesting activity is found, one identifies the peptide responsible for it and attempts variations on its structural theme. The tested activities will fall into three classes : a) those which are traditionally studied when model-catalysts are designed ; b) the biologically important reactions ; c) reactions of prebiotic chemistry.

Selective methods. The ideal would be to find methods which enable the product of a given catalyzed reaction to amplify the synthesis of the polymer responsible for its production. This possibility appears far-fetched at the moment but when the screening methods will be mastered, possible pathways towards selective methods may become

practicable.

Such ideas were also developed in my book [2], in particular in pages 87-89 of the Princeton edition. Later, Stuart Kaufman told me that he had issued patents on variants of these ideas, and it seems that when « combinatorial chemistry » became a fashion and a business, my book was used in court to argue about the legitimacy of some patents.