ROMA TRE
UNIVERSITÀ DEGLI STUDI

SCUOLA DOTTORALE IN BIOLOGIA

XXIV CICLO

Esplorazione funzionale e strutturale dello spazio di sequenze dell'RNA. Implicazioni per l'origine della Vita e le biotecnologie

Structural and functional exploration of the RNA sequence space. Implications for the origin of Life and Biotechnology

Fabrizio Maria Anella

Tutor: Prof. Fabio Polticelli

Co-tutors: Prof. Pier Luigi Luisi
Dr. Cristiano Chiarabelli

Coordinatore: Prof. Paolo Mariottini

“The important
is not the question
but the lesson
learned in time”

Contents

Riassunto

La teoria del *RNA World*, che parte dal presupposto che un mondo a RNA abbia preceduto il nostro mondo contemporaneo a DNA/ RNA/proteine, è diventata sempre più diffusa nel campo d'indagine dell'origine della Vita (Joyce, 2002; Orgel, 2003). Nonostante i recenti notevoli progressi compiuti in questo campo, alcune delle domande di base rimangono però ancora in attesa di una risposta: può l'RNA aver catalizzato le fondamentali reazioni di auto-replicazione sulla Terra primitiva? Può la vita basata sul RNA aver raggiunto la complessità metabolica necessaria per dare alla luce il mondo a DNA/proteine?

Per far fronte a queste domande una serie di lavori teorici e sperimentali (Szostack et al, 2001;. Muller, 2006) sono stati condotti, con l'obiettivo finale di ricreare un mondo a RNA in laboratorio. In questo quadro si inserisce il progetto "**Never Born Biopolymers**" (**NBBs**) (Luisi et al., 2006) e in particolare il progetto "**Never Born RNAs**" (**NB-RNAs**), che ha come obiettivo l'esplorazione dello spazio delle sequenze di RNA alla ricerca di sequenze con proprietà biologiche.

Questo progetto muove dalla constatazione che la collezione di molecole di RNA esistente è solo una piccola parte di tutte le teoricamente possibili sequenze di RNA (Luisi, 2003).

Sulla base di questa osservazione, la questione se la stabilità è una caratteristica comune o un raro risultato della selezione naturale è della massima importanza per far luce sul ruolo dell'RNA nell'origine della Vita e in modo da sfruttarne appieno il suo potenziale biologico.

Conoscendo la stretta relazione che intercorre tra struttura e funzione nelle molecole biologiche, abbiamo deciso di precedere l'esplorazione funzionale dello spazio delle sequenze dell'RNA con alcuni studi strutturali utilizzando l'**RNA Foster** (**RNA Folding Stability Test**) (De Lucrezia et al., 2006a). Questo test impiega la nucleasi S1, enzima specifico per filamenti singoli di DNA e RNA (Vogt, 1973) in grado di mantenere la propria attività in un ampia gamma di temperature, e permette il monitoraggio della presenza di domini a doppio filamento e indirettamente ogni eventuale struttura

terziaria. Infatti, gli RNA con un *fold* stabile sono più resistenti alla nucleasi S1 rispetto a quelli non strutturati, in quanto questi ultimi sono degradati più velocemente degli altri. Inoltre, sfruttando la capacità della nucleasi S1 di lavorare ad una vasta gamma di temperature, abbiamo verificato la stabilità dei domini secondari degli RNA in condizioni diverse. Infatti, un incremento della temperatura destabilizza il fold degli RNA, inducendo uno "svolgimento" globale o locale. Di conseguenza l'RNA diventa suscettibile all'attacco della nucleasi ed è facilmente degradato. In poche parole, le sequenze più stabili alle alte temperature saranno quelle con una struttura secondaria ed eventualmente terziaria più stabile.

Ad oggi il risultato più generale dei nostri studi risiede nella dimostrazione che gli RNA hanno la capacità di ripiegarsi in compatte strutture secondarie, anche in assenza di pressione selettiva (Anella et al., 2011). Questo conferma la nostra ipotesi che le molecole coinvolte attualmente nel fenomeno Vita non abbiano caratteristiche esclusive, almeno per quanto riguarda la capacità di adottare un *fold* stabile.

Nel dettaglio siamo riusciti a selezionare una sequenze, tra quelle analizzate, con una stabilità superiore ad un tRNA naturale a 70°C, che presenta una temperatura di *melting* superiore ad 80 °C. Alla luce di questi risultati si può postulare che il *fold* sia una proprietà intrinseca delle molecole di RNA e non il risultato di una lenta evoluzione.

I risultati saranno utilizzati per fornire indicazioni e suggerimenti per ulteriori studi riguardanti le proprietà funzionali del RNA in uno scenario prebiotico.

Abstract

The **RNA World** hypothesis, which assumes that an RNA World preceded our contemporary DNA/RNA/protein World, has become more and more popular in the field of the origin of life (Joyce, 2002; Orgel, 2003). Despite the recent progresses made in this field, some basic questions remain unanswered: Can the RNA catalyze the reactions needed for self-replication on the early Earth? Can the RNA-based life achieve the metabolic sophistication needed to give birth to the protein-nucleic acid World?
To tackle these questions a number of theoretical and experimental works (Szostack et al., 2001; Muller, 2006) have been carried out with the ultimate goal to re-create an RNA World in laboratory. Within this framework lies the "**Never Born Biopolymers** (**NBBs**)" project (Luisi et al., 2006) and in particular the "**Never Born RNAs**" (**NBRs**), project which has the goal to explore the RNA sequence space for catalytic functions. This project moves from the observation that the extant collection of RNA molecules is only a minor part of the all theoretical possible RNA sequences (Luisi, 2003).
On the basis of this observation, the question whether a functionality is a common feature or a rare result of natural selection is of the utmost importance to elucidate the role of RNA in the origin of Life and to fully exploit its biological potential.
Knowing the firm relation between structure and function of biological molecules we decided to precede the RNAs functional exploration with some structural studies using the **RNA Foster** (RNA Folding Stability Test) (De Lucrezia et al., 2006a). This assay employs S1 nuclease, a specific nuclease (Vogt, 1973) to cleave at different temperatures single-strand RNA sequences, monitoring the presence of double-stranded domains and indirectly any possible structure. In fact, folded RNAs are more resistant to S1 nuclease than unfolded ones, namely the latter are degraded faster than the former. In addition, we exploited the capability of nuclease S1 to work over a broad range of temperatures to probe RNA secondary domain stability at different conditions. In fact, an increase in temperature destabilizes the RNA fold, inducing either global or local unfolding. Consequently, the RNA becomes susceptible to nuclease attack and is readily degraded. In few words the most stable sequences at high temperature will be those with a more stable secondary and possibly tertiary structure.

Until now the most general result of our studies lies in the demonstration that RNAs have the capacity to fold into compact secondary structures, even in absence of selective pressure (Anella et al., 2011). This confirm our hypothesis that molecules involved in nowadays life don't have exclusive features at far as the ability to adopt a stable fold is concerned.
In detail one of the sequence analyzed has a stability higher than the tRNA at 70°C , with an approximately melting temperature higher than 80°C.
The results will be used to provide directions and suggestions for further studies concerning the functional properties of RNAs in the early evolution scenario.

1 Introduction

"It must be admitted from the beginning
that we do not know how life began"

"The origin of life on Earth"
Stanley Miller and Leslie Orgel

Biology is the science studying Life in all its forms and aspects. According to such definition, found in many dictionaries available on the market, the primary object of biological studies is Life; Life is such a special phenomenon that after its beginning, it followed a path 3.8 billion of years long, full of ups and downs, through trials and errors, successes and tragedies has also led to the appearance of this peculiar species to which we belong, the *Homo sapiens*.
Considering that Life is the focus of a scientific discipline, it can be surprising that there is a lack of a generally accepted definition by the scientific community. As a matter of fact, any attempt of giving a definition to this incredible phenomenon is incomplete, or elusive.
A famous evolutionary biologist, Jay M. Savage, noticed in the early 60s of the last century that:

"Life is very difficult to define because it is a dynamic process, constantly changing and extraordinarily complex".

Therefore, not being able to describe Life[1] exhaustively, we try to force the game by highlighting the features that differentiate it from its alter-ego, the non-living matter; this is all there would be if almost 4 billion years ago non-living matter didn't start a process that managed to harness the properties of matter towards the achievement of a goal, the acquisition of an identity, escaping from the tight control of thermodynamics and becoming the exception that proves the rule. As a matter of fact, living matter doesn't disobey to any law of physics, however, it is an unprecedented phenomenon in the physical universe: a desperate attempt, with an expiration date, to escape the second law of thermodynamics.

One way to get a characterization of the phenomenon of Life is to arm ourselves with the principles of physics and chemistry and explain the basis for the possibility of the occurrence of material systems with living organism features. This means to explain how, among the many possible physical systems *a priori*, there are also some with the properties and skills of life.

This is the field of investigation of the origin of Life.

1.1 The origin of Life

What is Life? How did it form? These existential questions have always intrigued the human mind.

According to Greek philosophers, Life was inherent in matter and it emerged spontaneously from it when the conditions were favorable. In the third century B.C., Aristotle picked up the ideas about spontaneous generation shaped by philosophers who lived before him and he summarized them into a theory whose effects were felt until very recently. For the great Greek philosopher, living organisms are born, generally, from other organisms like them, although they can sometimes arise from inert matter.

The theory of "spontaneous generation" has passed unharmed through the Middle Ages and the Renaissance and was supported by prominent thinkers such as Newton, Descartes and Bacon. In the sixteenth century, there was still someone willing to believe that geese were born from some trees that were in contact with the waters of the ocean, as travelers returning from the East were telling.

---

[1] But we must give justice to the numerous publications about the definition of Life. Look for example Luisi, 1998 and Margulis and Sagan, 1995.

In the seventeenth century the era of legends ended and the first trials to support the theory of spontaneous generation began. The Flemish physician Jean Baptiste van Helmont claimed to have conducted an experiment bringing together grains of wheat and a dirty shirt, from which some mice were born after 21 days! It's obvious that Van Helmont's experiment was a bad one, but the right path, the experimental verification of ideas, was open.
The scientific community deals with its issues using a mathematical logic system, that allows to develop a universal law by considering different aspects of a phenomenon in order to be able to interpret and place it in a theoretical framework.
In this light the origin of Life is a difficult event to define and to represent through a scientifically valid model. Moreover, science is not based on intangible dogmas, but on assumptions and theories always evolving. Over the years, various hypotheses have been formulated on the origin of Life, trying to limit, from time to time, the scope of the case and defining the probable.
A significant impact on these studies was provided by the work of one of the greatest naturalists of all times: Charles Darwin. His theory of evolution, formulated in the XIX century, argues that all the present living organisms may have originated from a few or perhaps even from a single common ancestor. According to Darwin (Darwin, 1859), the process of natural selection created new organisms to fit a constantly changing environment. With the consolidation of such theory and also thanks to experimental work carried out by Pascal and Spallanzani, the theory of abiogenesis (Life originates from non-living substances) was abandoned in favor of the theory of biogenesis (life comes from life). However, at the same time the door to new questions was open, first of all: who created the first living being if to create a living being another living being is necessary? In fact, one of the fundamental "*cruxes*" of the history of Life on Earth is, of course, its appearance. This is such a complex "*crux*" that Darwin himself preferred to leave it unresolved, circumventing the problem more or less brilliantly. Nevertheless, dissolving the "*crux*" of the origin of Life is essential for the whole construction of Darwinian evolutionism: if the first living being did not "evolve" from matter due to purely mechanical causes, why assign the appearance of all plant and animal species to the mechanisms of "small random variations" and "natural selection"?
Despite the latest achievements in Biology, Life, however, remains the great mystery of the Universe. Its origin is still under more or less daring hypotheses, on the border between science and religion. This incredible phenomenon, which occurred about 4 billion years ago under very different conditions from today and is therefore difficult to study various theories have been built, aiming to unravel the mysteries.
I will discuss them in the next few paragraphs.

1.2 The beginnings

How long back in time can we trace the formation of Earth? From dating systems studies using the calculation of the decay of the nucleus of atoms that exist today (such as ^{129}I, ^{53}Mn, ^{107}Pd), was obtained that the aggregation of small bodies in the solar nebula occurred about 10 million years after the birth of the Solar System (Lugmair and Shukolyukow, 1998) (Figure 1.1).

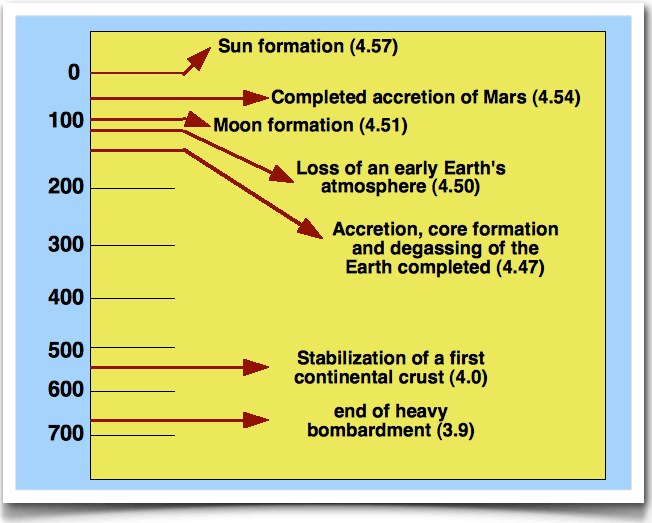

Figure 1.1: Chronology of the Solar system formation in billions of years

Between 4.51 and 4.45 billion years ago, the Earth reached its present mass, with a metal heart and a primitive atmosphere (Halliday, 2000). In these early stages, the Earth had an ocean of magma, due to the continuous bombardment of celestial bodies and to the effect of coverage provided by a warm and dense atmosphere, situations that left little space for a possible development of Life.
Much of this atmosphere would be amended because of the scattering of the solar nebula and the ongoing planetary collisions, causing a rapid cooling of the Earth which lead to the solidification of the outer portion, so to form a primitive crust. When can we trace the formation of the oceans, or better the first water in the liquid state? Recent data obtained by the analysis of very ancient zircon crystals, the oldest kind of terrestrial solids known, suggest that liquid water may have made its appearance around to 4.4 billion years

ago (Wilde et al., 2001; Mojzsis et al., 2001), finally making possible the development of life, even if the constant bombardment of extraterrestrial bodies may have induced a fusion of the surface rocks and the evaporation of the hydrosphere.
It is also possible that biogenesis has been repeated several times before the conditions became less prohibitive, and would have allowed the establishment the evolutionary process.
Between 4.0 and 3.9 billion years ago, the continents and oceans would have been stabilized and the first finds of microfossils in the sediments recovered in Apex Australia (Schopf, 1993) would dated about 3.5 billion years ago.

1.3 The primordial atmosphere

There is no general agreement about the composition of the primitive atmosphere, with theories ranging from a reductive composition ($CH_4 + N_2$, $NH_3 + H_2O + H_2 + CO_2$ or N_2) to a neutral one ($CO_2 + H_2O + N_2$). The only aspect in which both theories agree is the absence of free O_2 (Lazcano and Miller, 1996). On the other hand, reductive conditions are absolutely required for the synthesis of amino acids, purines, pyrimidines and sugars (Stribling and Miller, 1997). Alternatively, we should postulate an extraterrestrial origin for the first organic compounds (Anders, 1989). However, it has been argued that their amount was not in all cases sufficient to trigger the early biosynthetic processes (Chyba et al., 1990).
Another highly debated issue is related to the temperature of the primitive Earth. It is believed that the whole planet has remained to the molten state for several hundred million years after its formation (approximately 4.6 x 10^9 years ago) (Wetherill, 1990). Some models foresee a high partial pressure of CO_2 or other greenhouse gases, resulting in increased temperatures, that would have compensated for a reduction of the 30% in the perceived solar luminosity, so avoid in a complete freezing of the oceans (Kasting, 1993). The high temperatures, since they favor all exothermic reactions, would have also played in favor of the transitions from abiotic reactions to the first self-replicating entities. Moreover, the most recent phylogenetic reconstructions have suggested that the common ancestors of today's living organisms have been hyperthermophile microorganisms (Pace, 1997, Di Giulio, 2003) (Figure 1.2).

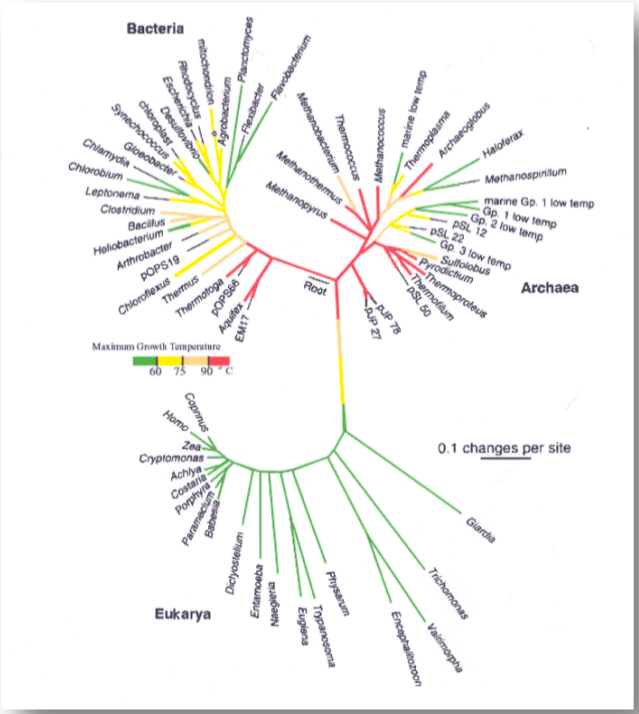

Figure 1.2: Phylogenetic tree based on rRNA sequences

1.4 Three models to explain the origin of Life

To answer the question concerning the origin of Life, three models were mainly considered with interest by the scientific community to explain how, starting from basic compounds, it has come to the formation of biopolymers that underlie the phenomenon of Life.

1.4.1 The autotrophic origin

The autotrophic theory is based on a model of primitive atmosphere rich in CO_2 (Kasting, 1993). A high level of CO_2 and a high pressure (10-100 atm), however, imply the absence of reducing conditions and then the synthesis of organic compounds, thus making it necessary to postulate that small amounts of these were brought by comets, meteorites or star dust. In 1988,

Wächtershäuser developed an evolved autotrophic model which could prove the biosynthesis and polymerization of simple molecules on the surface of metal sulphide FeS and FeS_2. The reaction:

$$FeS + H_2S \quad ==> \quad FeS_2 + H_2$$

is particularly thermodynamically favored (AG ° = -9.23 kcal / mol, E ° = - 620mV at pH 7 and 25 °C) and the combination FeS / H_2S is a strong reducing agent (Wächtershäuser, 1988). This combination can be used to reduce double bonds, α-ketoglutarate to glutamic acid, hydrocarbon thiols, and many other organic molecules (Hafenbradl et al., 1995). However, this system cannot reduce CO_2 to amino acids, purines or pyrimidines although there would be more of the free energy needed to drive these reactions (Keefe et al., 1995).

1.4.2 The heterotrophic origin

The first formulations of the heterotrophic theory were advanced at the end of the twenties of the last century by the Anglo-Indian biologist Haldane (Haldane, 1928). He started from the observation that the early Earth should have very different characteristics from today. If today some organic material spontaneously forms - he pointed out - this would be immediately devoured by some living organism, while the organic matter on the early Earth, not decomposed by bacteria or other microorganisms, could have continued and, left quiet, would had enough time to develop and evolve.
The same concept was formulated long ago by Darwin, whom in a letter to a friend expressed the following consideration: "*It is often said that all the conditions for the first production of a living organism are present, which could ever have been present. But if (and Oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., present, that a protein compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed*" (Darwin, 1888).
In 1924, the Soviet scientist Oparin published a book (later translated into English in 1938) (Oparin, 1938) where he assumed a strongly reducing primitive atmosphere composed by hydrogen, water vapor, methane, nitrogen and ammonia. In such atmosphere, which existence is not supported by experimental proof, the electrical discharges of lightning and ultraviolet solar radiation would have caused the synthesis of simple organic compounds, including amino acids, purines and pyrimidines that dispersed into the oceans, would have formed the so-called "*primordial soup*", where

for subsequent chemical reactions, would have randomly formed nucleic acids and proteins.
The Oparin-Haldane's ideas were the starting point for one of the key experiments of modern research on the origin of Life. In 1953 (year really fruitful for biology...), an American researcher, Stanley L. Miller, devised an experiment that confirmed the hypothesis of the "*primordial soup*" (Miller, 1953).
The necessary equipment (Figure 1.3), arranged for this purpose, was quite basic and consisted of two glass vials filled respectively with water kept at high temperature (which mimicked the ancestral ocean) and with a mixture of hydrogen (H_2), ammonia (NH_3) and methane (CH_4) (gasses that, along with water vapor, were thought to be the main constituents of the gaseous envelope that more than four billion years ago surrounded the Earth).

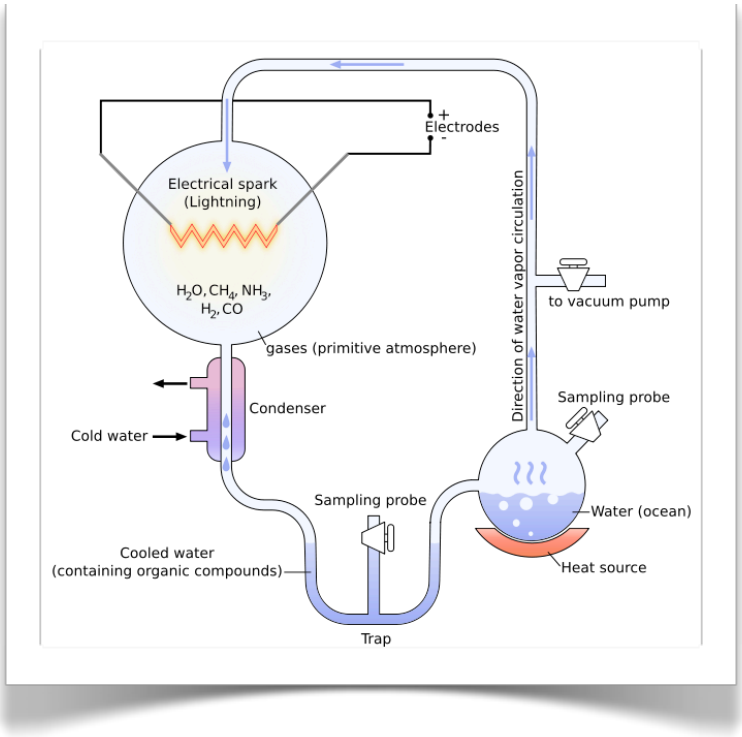

Figure 1.3 : The Miller's experiment

The vapor produced from water was pumped through a tube into the vessel containing the gases, in which electric shocks were generated trying to reproduce the presumably frequent and intense storms of the early history of our planet. The analysis of the compounds mixture formed after one week of treatment showed, with surprise, the presence of amino acids, the basic building blocks of today's proteins.
The initial experiment, was followed by others with different starting mixtures and with various sources of energy (ultraviolet, X-rays, electrons flow, high temperature) from whose were obtained the formation of carbohydrates, lipids, amino acids and nucleotides (Miller, 1955).

To date, three main scenarios of the heterotrophic theory are considered by the scientific community:

a) **RNA-World:** it assumes that RNA was the first macromolecular species, from which DNA and proteins have originated. I will return to this issue, central to my thesis, in section 1.5;

b) **Compartmental Approach:** it is based on the necessity of a spherical closed boundary at the beginning of the business. All life on Earth, in fact, is based on closed compartments and the interaction between the inner and outer worlds. This idea is reinforced by the fact that cell-like compartments formed spontaneously from prebiotic molecules (Deamer, 1985; Deamer, 1998; Bachmann et al., 1992; Walde et al., 1994). The starting framework of such approach is that these shells formed spontaneously and would have encapsulated some simple peptides, which, together with other molecules, started to catalyzed reactions, triggering a simple primitive metabolism.

c) **Prebiotic metabolism:** it based on the possibility of establishing a prebiotic metabolism before the advent of the enzymes. It is divided into two strands (Luisi, 2006):

   i) *Universal metabolism:* it examines the chemistry of model systems of C, H and O. Starting from CO_2 and reducing agents, using the power generated by these as a form of energy, it would have been feasible the set up of a chemical system (Morowitz, 1992).

   ii) *Metabolism on clays:* this scenario born from the pioneering work of Cairn-Smith, who developed an idea proposed by Bernal (Bernal, 1949), suggesting a model where monomers of

biological macromolecules present very diluted in the "*primordial soup*" could have been concentrated by adsorption on the mineral surfaces (Cairns-Smith, 1986; Ferris 1993). Moreover the clays could have had the ability to create a stabilizing microenvironment (Franchi et al., 1999), which preserved the monomers from the hydrolysis, the main reaction in an environment dominated by the water as a means (Ferris et al., 2003). This stabilizing environment, together with the clay characteristic to dispose certain molecules in specific positions, thanks to the presence of pores or charged regions on their surfaces, would allowed the polymerization of these monomer units (Kebbekus, 1988; Ferris and Ertem, 1993; Kawamura and Ferris, 1994).

1.4.3 Extraterrestrial origin

As described in paragraph 1.4.1, the autotrophic model implies an extraterrestrial origin for some prebiotic compounds (Anders, 1989). On the other hand, the theory of panspermia was already advanced in the early twentieth century by Swedish chemist Arrhenius, who suggested migration through space and the colonization of new planets by simple forms of life (Arrhenius, 1907).
The subsequent discovery of numerous organic molecules in meteorites, comets and stellar dust would support the hypothesis of a strong organic chemistry in the Universe and a not insignificant contribution of the extraterrestrial material to the formation of life on our planet. The total amount of prebiotic compounds in an interstellar cloud is quantifiable on the order of a few tons/year (Ehrenfreund et al., 2002), although only a small percentage of these compounds survived the impact with the Earth. About a qualitative analysis, the Murchison meteorite (discovered in the 1970), for example, contains approximately 3% of organic carbon (in the form of formaldehyde, hydrogen cyanide, formamide, acetaldehyde) and 0.1 g of amino acids for kg of meteorite (Dworkin et al., 2002). The purines adenine, guanine, xanthine and hypoxanthine and the pyrimidine uracil were found in concentrations from 200 to 500 parts per million (Van der Velden and Schwartz, 1977).

1.5 RNA World

"…the problem of the origin of life is the problem of the origin of RNA-World"
Leslie Orgel

If we start from the undoubtedly strong statement of Leslie Eleazer Orgel (Orgel, 2004) and then we consider the RNA World, the first biochemical organized system on the primitive Earth, the starting point is to shed light on why RNA holds this key role.
First of all the ribonucleic acid is a very old molecule, very versatile and involved in all processes related to the perpetuation of cell life. When in 1982 the molecular biologists Thomas R. Cech (Kruger et al., 1982) and Sidney Altman (Guerrier-Takada et al., 1983) independently discovered that RNA molecules can act not only as messengers and custodians of information, but even as enzymes catalysing chemical reactions, a hypothesis already advanced in the mid-'60s by Carl R. Woese (Woese, 1965), the assumption that the biological catalysis belonged only to the realm of the protein enzyme was shattered.
By the time the list of RNA catalysts is lengthened by the discovery of new ribozymes (Peebles et al. 1986; Prody et al., 1986; Buzayan et al., 1986; Sharmeen et al., 1988; Saville and Collins, 1990). Moreover It was also clarified the key role of the ribonucleic acid in the complex of the spliceosome and in the ribosome (Sharp, 1991; Noller et al., 1992).
The theoretical framework of the RNA-World began to be outline, which would have preceded the world we know today, where the RNA would have carried out many functions currently performed by other molecules, including DNA and proteins.
The RNA-World hypothesis is supported by numerous observations on the present forms of life. First, DNA replication requires the catalysis of proteins and at the same time the protein synthesis requires that the amino acid sequence is encoded by the DNA. The RNA World explains how this interdependent situation, comparable to the classic puzzle game of "*who came first, the chicken or the egg?*", would have been derived from a more simple one: a world in which RNA acted both as conservative of the genetic information and as catalyst, then passing respectively the two roles to DNA and proteins (Gilbert, 1986).
Second, the hypothesis explains a number of biochemical observations: the ribosome and RNase P are ribozymes (Steitz and Moore, 2003; Guerrier-Takada et al., 1983), the nucleotide cofactors are conserved ubiquitously in biology (White, 1976; Benner et al., 1989), the DNA replication requires RNA primers (Itoh and Tomizawa, 1980) and the DNA is synthesized from

RNA precursors (Lazcano et al., 1988). All these subjects are considered as biochemical molecular fossils (White, 1976) of a primordial RNA World.
So, although the RNA seems the ideal candidate to be awarded with the endowment of key molecule for the origin of life, some doubts persist.
From a purely chemical point of view, it is difficult to imagine how RNA molecules can be formed in non-enzymatic ways.
First of all, its precursors, the ribonucleotides, are difficult to form by means of synthesis with a prebiotic plausibility. Several synthesis, in compatible conditions with a prebiotic scenario, have been set up: from the synthesis of adenine from hydrogen cyanide and ammonia (Orò, 1960), to that of cytosine and uracil in the presence of urea from cyano aldehyde (Robertson and Miller, 1995), and finally to the various reinterpretations of the formose cycle of Butlerow (Butlerow, 1861) that produces a variety of treosi sugars, pentoses and hexoses (Socha et al., 1980; Reid and Orgel, 1967).
For the next steps too, the formation of nucleosides and nucleotides, several reactions have been investigated, such as the production of purine nucleosides from D-ribose, of hypoxanthine in the presence of divalent ions (Fuller et al., 1972) and of the nucleotides in the presence of inorganic phosphates and polyphosphates (Schwartz, 1997; Peyser and Ferris, 2001).
However, to date, these syntheses do not provide large quantities of products and tend to form complex mixtures, including the racemic (Shapiro, 1999). In addition, the synthesis of pyrimidine nucleosides doesn't start from any prebiotic molecule, such as the arabinose-3-phosphate, while the phosphorylations of them lead to complex mixtures of isomeric mono- and polyphophates (Orgel, 2004). In addition, the formation of an RNA molecule requires a long series of 3'-5' phosphodiester bonds in the context of a series of competing reactions such as hydrolysis and the formation of 2'-5' bonds. Anyway, even if all the problems that affect the synthesis of RNA polymers under prebiotic conditions, to date, are unresolved, the difficulties concerning the necessity to satisfy the request of a large number of copies of the same molecule to ensure that the RNA-World could be a feasible chemical system and not only a fascinating hypothesis, can be overcome if one takes into account the possibility that different sequences can be able to perform the same function.

1.6 The sequences space

"The real is not as rich as the possible"
Ilya Prigogine

As introduced in the previous paragraph, a prerequisite for the initiation of the RNA-World, in order to establish a chemical feasible system, is the request for a high number of copies of the same RNA sequence.
This constraint is currently one of the most serious limitations to the emergence of the RNA-World. In fact, a single RNA sequence is only a point in a 4^l space of all the possible sequences of a given length l (Luisi, 2006; Schuster et al., 1994). This dimension 4^l comes from the fact that for a given polymer of length l, 4^l is the number of all the possible generable sequences, being each position of the sequence occupied by one of the four nucleobases.
If RNA molecules emerge from prebiotic chemistry, the ability to synthesize repeatedly a certain sequence is near to impossibility, due to the random nature of all synthesis routes in a plausible prebiotic scenario. However, this request is true only if one assumes that a metabolic function is performed only by a single RNA sequence, so to require its replication repeatedly with a high degree of fidelity. This is not the case of the RNA, because many theoretical studies have stressed that there is a high redundancy in the sequence-structure map of RNA, so many different and unrelated sequences can easily take the same structure and could probably play the same function (Schuster, 1995).
Several independent investigations have shown that for a fixed-length polynucleotide, the number of different generable sequences far exceeds the number of the possible structures. Schuster and colleagues (Schuster et al., 1994), using an inverse folding algorithm to calculate the number of different sequences with the same secondary structure, found that the sequence-space structures is the type of many-to-one (Figure 1.4) and that the different sequences that take the same shape are spread evenly in the space of sequences, showing a small or none homology degree (Schuster, 1995).
These theoretical indications are experimentally confirmed by the observation that only seven nucleotides are strictly conserved in the intron self-splicing group I (Lisacek et al., 1994), although the secondary structure of the ribozyme and its function are preserved. This case clearly illustrates the fact that sequences with little or no homology can achieve the same structure and perform the same function in spite of the diversity of their sequences.

The reasons behind the redundancy of the sequence/structure space of RNA can be found in the modular organization of the secondary structure of the RNA. Knight and colleagues have highlighted how RNA molecules are organized into independent functional domains connected by flexible spacers, whose sequences are irrelevant for the secondary and tertiary structures (Knight and Yarus, 2003). In this work they developed a method to calculate the abundance of these functional domains, in a pool of random RNA sequences. The results showed that the modularity of RNA sequences further increases the redundancy of the sequence/structure space, since the number of ways that can be achieved a specific fold increases markedly due to the modularity of the domains of RNA.
In addition, the RNA secondary motifs, crucial for ribozymic activity, are due to the annealing of complementary bases that follow the Watson and Crick rules (A:U and C:G), increasing even more the probability of finding these domains in different regions of the sequences space and inside sequences no related to each other (Knight et al., 2005).
The simple consequence of these results is that the RNA sequence space is rich in folded RNAs. In addition, the RNA sequence/structure space seems to be characterized by a high redundancy of related sequences that share the same structure and probably the same function too.

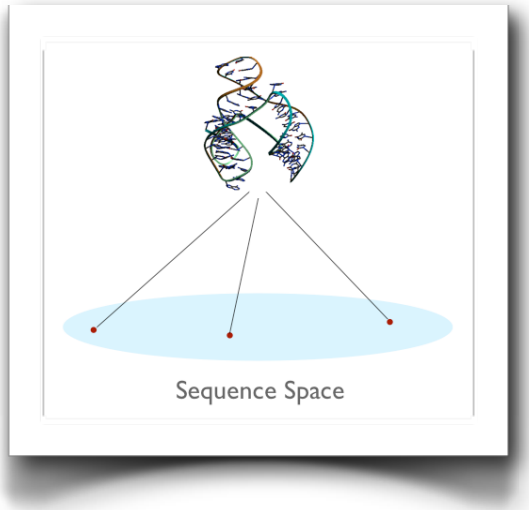

Figure 1.4: Structural redundancy of the sequences space. The gray flat represents the potential diversity of molecular species, i.e. all the different sequences can be generated for a given length l. From several studies, both computational and experimental, it was found that different sequences of RNA, with little or no homology, fold getting the same structure. This indicates that sequence/structure space is many-to-one.

Taken together, these results may suggest that in principle there is no need for identical copies of the same RNA sequence for the initiation of an RNA-World, making this scenario more plausible than it would have been if the sequences would have demanded fidelity of RNA replication beyond the ability of the prebiotic standards of synthesis.

1.7 Our "hunting" for folded RNA

Starting from the considerations on the structural redundancy of the sequence space, i.e. sequences with low homology may take the same structure, in our research group a comprehensive research has been undertaken to study the structural properties of random biopolymers, that is, the draft of the *Never Born Proteins* (*NBPs*) and *Never Born RNAs* (*NB-RNAs*) (Luisi, 2006).
This research starts from the consideration that the number of biopolymers currently present in living organisms is only a small fraction of those theoretically possible (Xia and Levitt, 2004). For example, the number of possible different peptides of 50 amino acid residues synthesized using the 20 natural amino acids is equal to 20^{50}, ie 10^{65}.
This number is only an order of magnitude less than the number of atoms in our galaxy. The number of possible different proteins becomes even greater if you take into account the living organisms, where the average length of proteins is much greater (Lipman et al., 2002).
The difference between the number of possible proteins and the number of those actually present in living organisms is comparable to the difference that exists between a grain of sand and the entire Sahara desert (Luisi, 2006). In the same way this considerations can be broadened to the RNAs.
On this basis it was decided to explore the RNA sequence space in search of random RNA molecules that have a stable folding at high temperatures, by an assay set up in our lab: the *RNA Foster* (RNA Folding Stability Test). Basically, the RNA Foster assay allows to determine the presence and thermal stability of secondary domains in RNA molecules by coupling enzymatic digestion with temperature denaturation (Figure 1.5). The assay is capable to quantitatively determine the fraction of folded RNAs (f_{fold}) as a function of the temperature (De Lucrezia et al., 2006a).
The RNA Foster assay employs S1 nuclease, a single-strand-specific nuclease, to cleave single-stranded regions, thus monitoring the presence of double-stranded domains and indirectly any possible tertiary structure. Folded RNAs are more resistant to S1 nuclease than unfolded ones, as the latter are degraded faster than the former. The capability of nuclease S1 to work over a broad range of temperatures can be used to probe RNA secondary domain stability at different temperatures. An increase in

temperature destabilizes the RNA fold, inducing either global or local unfolding. The fraction of folded RNA f_{fold} at each experimental temperature is determined by measuring the amount of RNA remaining after S1 digestion, using electrophoresis and a suitable staining method. The f_{fold} at different temperatures is used to assess the Tm of the RNA molecules.
The Foster assay has several advantages: the simplicity and rapidity of execution, the requirement of simple instrumentation, and in particular the minimal RNA amount required. In general, the most notable advantage of this methodology is that it is a time-saving technique and is therefore, suitable for the screening of large libraries and could be easily adapted for high-throughput studies, like *in vitro selection* experiments (Wilson and Szostak, 1999).

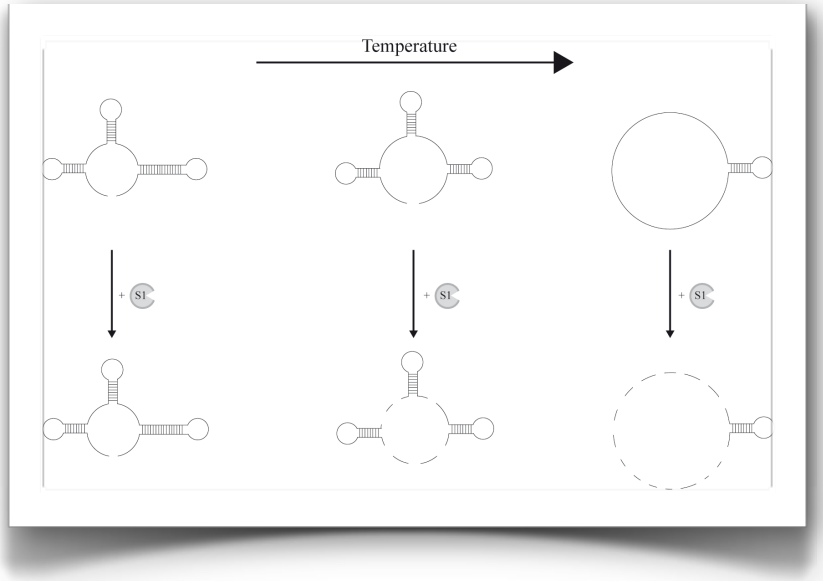

Figure 1.5: Schematic representation of the RNA Foster assay. It combines the thermal incubation with nucleasic activity of nuclease S1, which is acted on single strand RNAs. In fewer words, the most stable sequences at high temperature will be those with a more stable secondary and eventually tertiary structure.

In particular, we have prepared a library of totally random RNAs containing 60 residues (plus a pair of constant flanked regions for operational reasons), asking the question of stability: whether or not, and to what extent these totally random RNA sequences are able to exhibit a stable compact fold.
In addition to the goal of isolating the folded RNAs, this work is guided by three main intentions.
First, it would provide an experimental test of contingency, namely the interpretation of the history of life as opposed to the deterministic point of view (Luisi, 2003).
According to this theoretical approach the whole phenomenon of life we see today is the result of a long series of events, that occurred one after one, each of which is closely linked to the conditions in which it occurred, even one different variable would have lead to a completely different result, but even it with a sense .
This way of describing the history of Life is perfectly summarized by a quote from one of his biggest supporters, the evolutionary biologist Stephen Jay Gould (Gould, 1989):

"If you could rewind the tape of life and operate it again would result in a completely different result that would also make sense"

This is the nature of contingency in the history of Life. All this does not mean that everything is random and inexplicable, but it just says that what happens is unrepeatable and so complex, so dependent on each of the thousands of previous stages, each of which could have been carried out in a different way, that it is not possible to predict the result based on the immutable laws of Nature.
In addition,it has to be considered that during the evolutionary process not all possibilities have been explored. This statement is based on some reasoning (De Duve, 2002, Luisi et al., 2006). Besides the fact that to obtain a copy of all possible polynucleotide sequences of a given length, for example, of 100 residues, all atoms of the universe would not be enough (Luisi, 2006), the time required for the synthesis of all sequences of a given length presents a paradox. In fact, estimating a rate of synthesis of a polynucleotide chain to second, the time required for the synthesis of all sequences of a given length, in this case sequences of 50 residues, is even more than the time estimated from *BigBang* to today. All these observations suggest that there is a whole universe of possibilities unexplored where lie undiscovered new biopolymers (Figure 1.6).

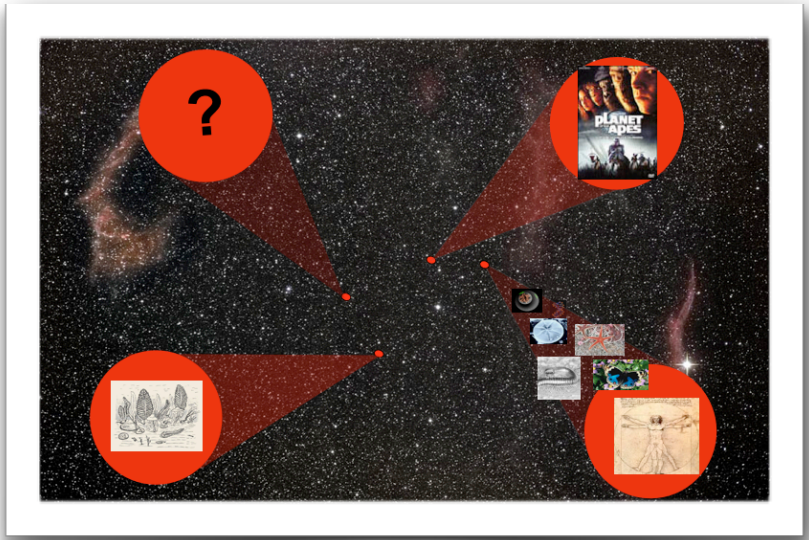

Figure 1.6: And if you would have started from another point of the space of sequences?

Secondly, our choice of starting from a population of random RNA sequences may represent a model of the early stages of RNA evolution, during which the prebiotic catalysis (i.e. non-enzymatic) lacked of specificity essentially generating populations of random RNA. Our choice fell on a population of RNA with a length of 60 nucleotides to enhance even further this comparison, as this is more or less the length of tRNA and of the ribozymes core today known, because the synthesis of sequences of 60 residues length, although as yet unrealized under plausible prebiotic conditions, retains a kind of reasonableness, because could have been the result of condensation reactions between shorter nucleic acids.

Finally, all these topics have a particular relevance for the origin of Life.

In fact assuming a hot-start origin for Life (see paragraph 1.3) and assuming a key role to the RNA in a prebiotic scenario (see paragraph 1.5), a requirement that this molecule has to satisfy is the stability at high temperatures.

To date there is little evidence supporting this feature of RNAs (Bada and Lazcano 2002, Moulton at al, 2000). Therefore, the search for stable sequences has the purpose to reinforce the candidature of the RNA World as the first organized system in the evolution of Life.

2 Aim of the project

The importance of RNA is determined by its role in biological systems as intermediary in the translation of information from genes to proteins, gene expression regulation and development timing. Moreover, RNA might have had a crucial role in the origin of life (Orgel, 2003) and it could be of great relevance in the development of biotechnologies (Scanlon, 2004).
RNA exerts its role by mean of its tree-dimensional structure, which is the result of a well-defined and stable spatial arrangement of its chemical groups. Indeed, an essential requirement to perform any biological function is a well-defined and stable three-dimensional structure. Besides, the observation that the extant collection of RNA molecules is only a minor part of the all possible RNA sequences (Luisi, 2006), rises the question about the structural properties and potential activity of all the RNA molecules not existing in nature.
On the basis of this observations, the question whether a stable fold is a common feature or a rare result of natural selection is of the utmost importance to elucidate the role of RNA in the origin of Life and to fully exploit its biotechnological potential.
One of the aims of this work was to tackle this question and to this aim the spontaneous occurrence of folded RNAs in a random library was investigated. Furthermore, structural properties of randomly synthesized RNAs, such as thermal stability and main unfolding temperature were studied.
A random DNA library of 60 residues (plus a pair of constant flanked regions for operational reasons) was designed and produced to perform *in vitro* transcription and the resulting RNAs were screened to evaluate their structural properties. Because no convenient method was known to investigate structural properties of large RNA library, a new assay was developed. The screening methodology is based on the observation that RNA tertiary and secondary structures are unavoidably grounded on intermolecular double helices formation (Draper, 1996). Therefore the screening methodology, named RNA folding stability test (RNA Foster) (De Lucrezia et al., 2006a), was used. The S1 nuclease, a single-strand specific nuclease (Vogt, 1973) was employed, to test the presence of double helices and indirectly any possible secondary and tertiary structure. In fact, folded RNAs are expected to be more resistant to S1 nuclease than unfolded ones, namely the latter should be degraded faster than the former. The method was validated with extensive control experiments.
The results were used to provide indications and suggestions for further studies concerning the stability and structural properties of RNAs in the early evolution scenario.

3 Experimental results

To investigate the possibility of isolating, from a broad population of RNA, sequences with a stable fold at high temperature, a library of random DNA sequences was designed and produced (section 3.1). At the same time was designed and validated an enzymatic assay with the aim to individuate folded RNA within a random population (section 3.2). This assay was used first to evaluate the rate of folded sequences in a population of RNAs 60 nucleotides long, transcripted from the corresponding plasmidic library (section 3.3), then to select a random population for thermal stability (section 3.4).

3.1 Construction of the plasmidic library

The DNA library was designed to possess specific features (Figure 3.1). An oligonucleotide containing the T7 RNA polymerase promoter was cloned into the plasmid pIIIdmy/HisTag-A upstream the 60-nucleotide-long random DNA region. The vector, which offers better performances in the cloning and transformation steps, is a derived plasmid produced from the commercial vector pHEN1.
The library was obtained by annealing two oligonucleotides, one carrying the T7 promoter (RT-PrimerFw-wT7) and the second one the random sequence (LibRNA$(NNN)_{20}$). The random region was designed according to the randomization NNN scheme (where N is an equimolar representation of all four bases). Two short constant regions (RT-Fw and RT-Rv) were placed upstream and downstream the random DNA region to allow the annealing of two primers for an eventual RNA reverse transcription. At the two ends there are unique restriction sites (NcoI and NotI at the 5', XbaI and BamHI at the 3' end) that were used both to prepare the insert for the cloning and to linearize the plasmid for the RNA transcription.
The single-stranded regions were converted into double strands by fill-in reaction using the Klenow's polymerase fragment after the annealing of the oligos LibRNA$(NNN)_{20}$ and RT-PrimerFw-wT7. The success of the reaction was verified by a run on 2% agarose gel. The main band of the reaction was at the correct height in comparison to the marker corresponding to a length of approximately 150 bp (Figure 3.2), quite comparable to that predicted by *in silico* PCR using the program *pDRAW32 1.0*, a product of 148 bp.

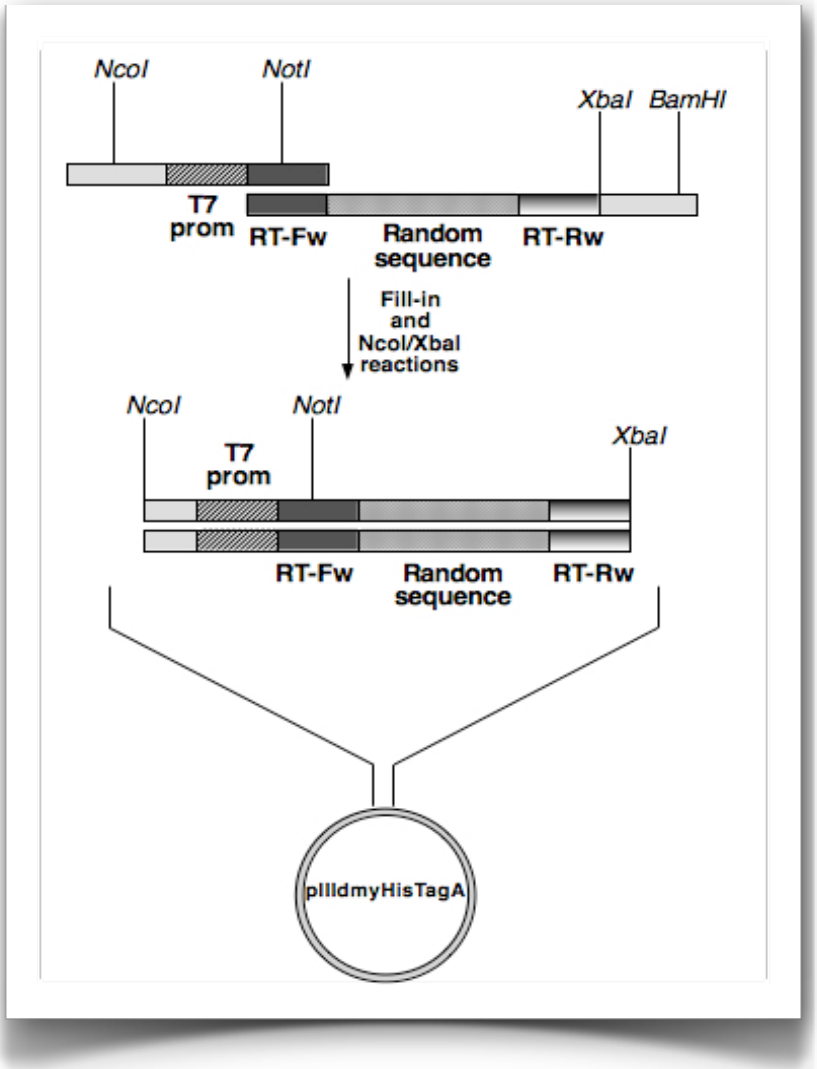

Figure 3.1: Structure of the DNA library codifying for 97-nucleotide-long partially random RNAs. The library insert was obtained annealing two oligomers, one carrying the T7 promoter sequence (RT- PrimerFw-wT7) and the other with the 60 random nucleotides (LibRNA(NNN) 20). Single-stranded oligomers were converted to double-strand by fill-in reaction employing the Klenow's fragment. The insert was cloned in the pIIIdmy/HisTag-A vector between NcoI and XbaI restriction sites. The 60-mer random sequence is flanked by two constant regions (RT-Fw and RT-Rv) useful for an eventual RNAs retro-transcription. At the end of the RT-Rv region, an XbaI unique site useful to linearize the vector for a run-off transcription was placed.

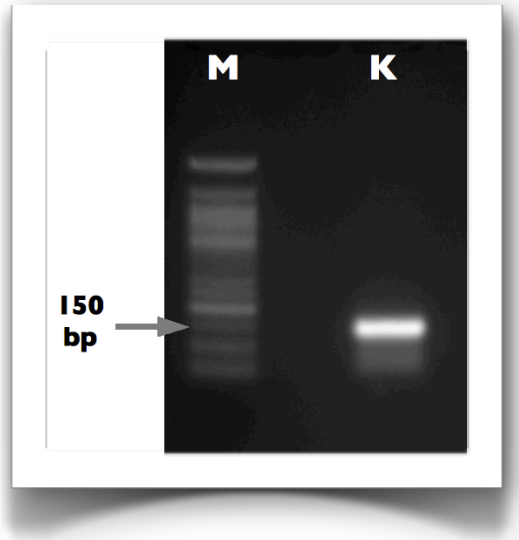

Figure 3.2 : 2% Agarose gel of the klenow reaction for the conversion to double-stranded of the DNA library. ***M*** *: 50 bp marker from NEB;* ***K*** : klenow reaction.

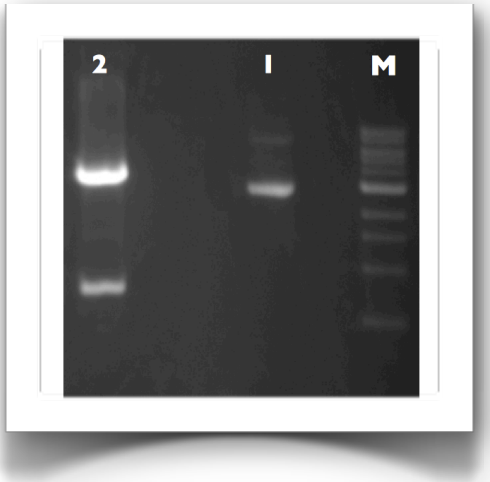

Figure 3.3: 1% Agarose gel of the double digestion reaction of pIIIdmy/HisTag-A. ***M****: 1 Kb marker from NEB;* ***1****: pIIIdmy/HisTag-A;* ***2****: pIIIdmy/HisTag-A double-digested by NcoI and XbaI.*

Subsequently, the dsDNA after being purified, was prepared for the cloning by a double digestion with restriction enzymes *NcoI* and *XbaI*. At the same time the vector *pIIIdmy/HisTag-A* was subjected to double digestion with the same enzymes as well (Figure 3.3), to create complementary ends to those generated in the insert.
Performed and controlled the ligation reaction (Figure 3.4), the resulting plasmid library was inserted into HIT-Blue cells, a modified XL1-Blue strain resulting in greater transformation efficiency (up to 5 x 10^8 transformants for µg of plasmid), by electroporation. The complexity of the library was estimated as 1.5 x 10^9 different sequences on the basis of serial dilutions on LB agar in the presence of ampicillin and tetracycline.

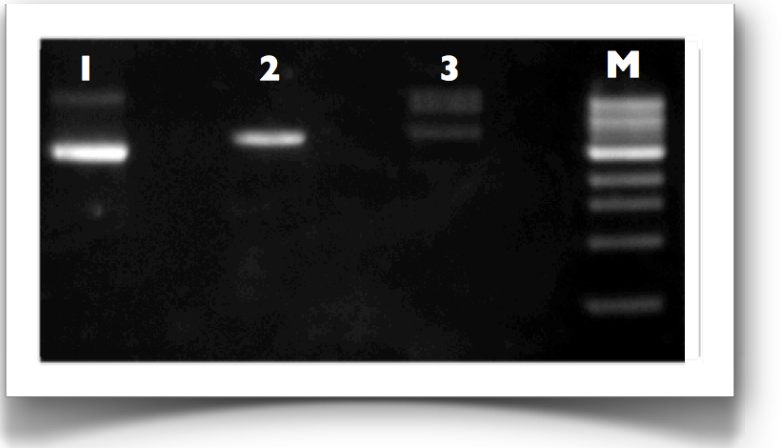

Figure 3.4: 1% Agarose gel of the ligation reaction. ***M*** *: 1 Kb marker from NEB;* ***1****: pIIIdmyHisTag-A;* ***2****: pIIIdmy/HisTag-A double digested by NcoI and XbaI;* ***3****: ligation reaction of the library insert and the pIIIdmy/HisTag-A double digested by NcoI and XbaI.*

A number of colonies were individually grown, and plasmids were extracted to determine the efficiency of the cloning step by plasmids sequencing (Figure 3.5). At the end of such controls the final complexity of the library was estimated to be approximately 0.85 x 10^9, still a good value for this kind of investigations.

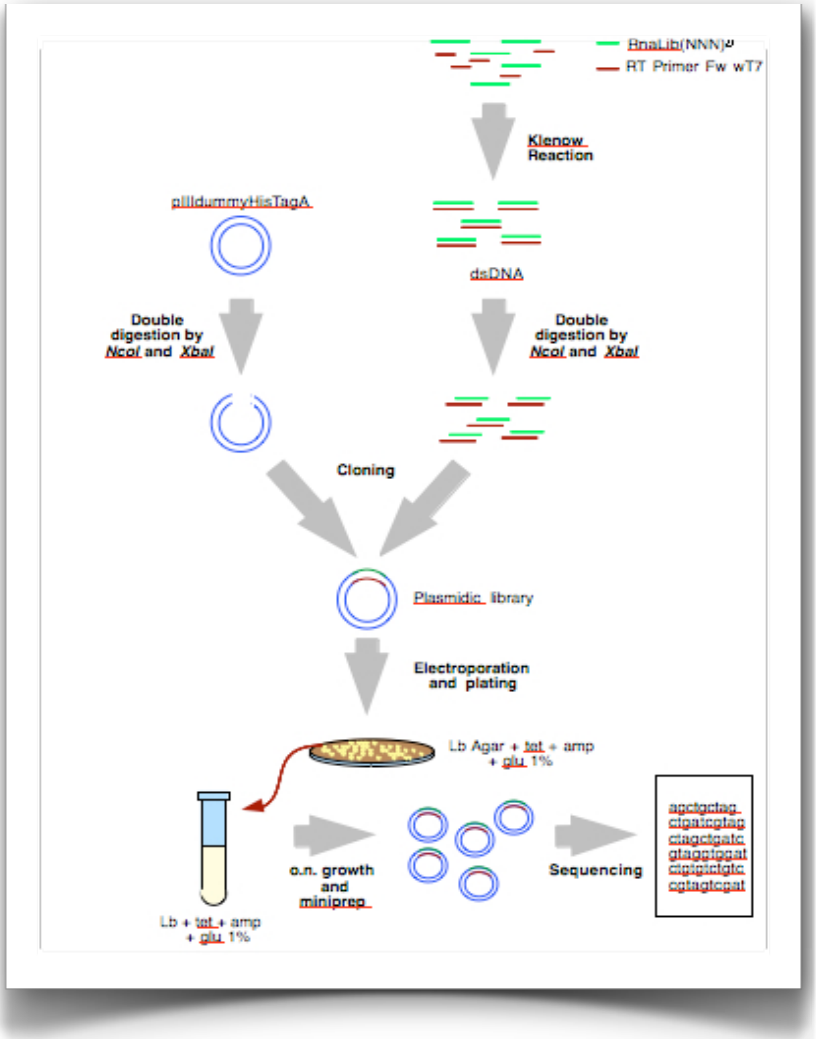

Figure 3.5 : Schematic summary of the procedure for the production of RNA library. For details of the different steps see text.

3.2 Optimization of the RNA Foster Assay

The RNA Foster assay allows to determine the presence and the thermal stability of the secondary domains in RNAs. The RNA Foster assay uses S1 nuclease, an endonuclease active against single stranded of RNA and DNA molecules, to test the presence of double helices and, indirectly, any possible structure. Moreover, RNA Foster assay allowed us to test the thermal stability and calculate the melting temperature (Tm) of random RNA molecules. RNA Foster assay is a powerful and simple approach, as it couples enzyme activity with thermal denaturation of the RNA molecules.
First of all, the incubation time and the S1 nuclease concentration conditions were optimized to be neither too harsh nor too mild. To this aim, I chose the *E. coli* tRNA-Trp as standard, because, in terms of length and structure, it is comparable to the random RNA population and can be used as control for the natural sequences.
For the optimization of the incubation time four samples of tRNA-Trp were incubated with different final concentrations of S1 nuclease (1, 0.1, 0.01, and 0.001 U/µl) for different time periods (15, 30, 45 and 60 min). Each sample was incubated for 15 min at 70ºC and then, for each test time, in presence of S1 nuclease at 70ºC as before. The optimal incubation time was estimated in 45 minutes (Figure 3.6).

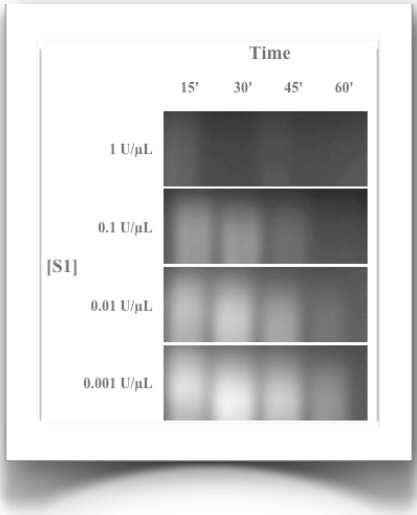

Figure 3.6: Optimization of the incubation time. 2% Agarose gels. The same quantity of tRNA-Trp was incubated with different final concentration os S1 nuclease (y-axis) for different times (x-axis). The fittest condition was 45'.

For the optimization of S1 nuclease concentration, the tRNA-Trp was incubated with different final concentrations of S1 nuclease (10, 1 and 0.1 U/µl). Each sample was incubated for 15 min at the test temperatures without S1 nuclease (30º, 50º, 70º and 90ºC) and then 45 min in presence of S1 nuclease at the same temperature. The optimal final S1 concentration was estimated in 0.1 U/µl (Figure 3.7).

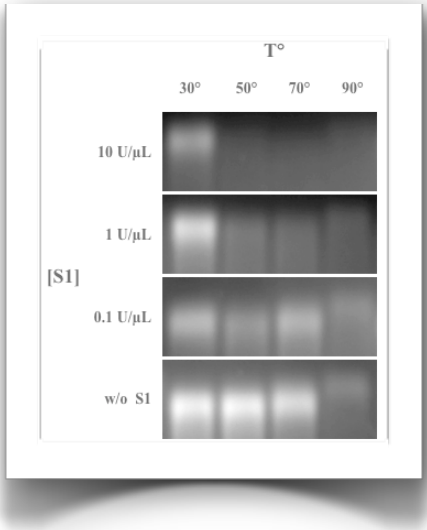

Figure 3.7: Optimization of S1 nuclease concentration. 2% agarose gel. The same quantity of tRNA-Trp was incubated with different final concentration os S1 nuclease (y-axis) for different temperatures (x-axis). The fittest condition was a final concentration of S1 equal to 0.1 U/µl.

3.3 Stability studies on single Never Born RNA

I selected 18 different RNA sequences in a completely random fashion (Table 3.1), namely without any kind of selection bias, in order to investigate their structural properties by means of the RNA Foster assay.

Name	Sequence
1.4	CGGAAAGCAGACCCCAAUGUAUCUGCCCCGGACCACAUUUCGAUACGAAGAGAAUAACUC
1.5	UGACACGCCCAGGUCUGGACUAAGUCCGCCGUCCGGGGUCGGAGCACCCACCUCACCCGG
3.1	AAGCAAGACUAUAAACCAGUCCAGUCCUACGAACUGAAUCCCUACAAACCCCCAUCACAC
3.2	CCUGAGACCCGAGUAUGACCUCAAAAUCAAUAUCUACCUUGAAUCACCUGCACCCAUUGA
3.3	AGCGAACGGCACCUCUUGACGAAACCUGAUCCGUCGAAUUACGCACUACAAGGCGACUAA
3.4	UGCGCGCACCCCAAUGGCAGAUUAUGCAUGAGCCGACGGUUACAAGAUGGCAUCAGCCCC
3.5	CCGGCUGUUGCACCCACUCCUUCCGUGUGUCAGACCGUACCCCGCUAACAAACGCAGGAC
5.1	CACACGGAAUACAGAUAAAGCCACUACAGCGGCAACCAGAUUAUACGGCGAUCGCACAUA
5.2	UACACACGUAUCAACAUAAUCUUCAUACGCCCAAACAAACCCAUUGCGAAGGCCAACACU
5.3	CCGGGUCUUCGACGGCCGCAAGUCCCAAUACCCAUCGCCCAACCUUCGCAGGCAAGGCUC
5.4	CUUUAAGAGACGGCGCUGAACUCCCUGACACAAACUUAAAAUAUAUACGAUCAGACCAGG
5.5	CAAAUUGUCAAAUUUGACCGAGUGAAAUCGGAAUGAAUUCAUACUCCGCACUGCCAUGAC
13	UUCGACGCGCCCAGGUGCGCCACGCACAUCCUGAUGAUACCUUUGCCCACAAUCAGCAAU
16	ACCCAUUCGCACCGAGACAUUUCUACUGCCUGUUCUCCUACCACCACGCCCUUAAGCUCU
18	UUAUACUUCCAUCGAUUCGAAAGGCACUACGCGGCCUCUGAACACACGAAUGAUAUACAG
20	CCUGCCUGAACAAACCGCGCAAAAAGCCAUACUGGGCCACGACUCAGACCAUCAACCCAA
21	GAGUUAAGCCAACCACCAUAAAAUCGCUAAACCUAGUUUAAUUGGCUCACGCUCAAUCGU
22	AUACAAGUGAAACAAAUCCUUUGGUACUCUCUUAUGCUAUAAACUGCUGUCUCCCCCAAU

Table 3.1: List of the 18 Random RNA sequences analyzed

After having selected the experimental conditions, the chosen RNAs were individually tested by the RNA Foster assay.
The transcription product of each linearized plasmid carrying a *NB-RNA* sequence was incubated for 15' at the test temperatures (30, 50, 70 and 90°C) and then 45' in presence of 0.1 U/µL of nuclease S1 at the same

temperatures. A control reaction was set incubating the sequences for 60' at 30°. The fraction of folded RNA for each temperature was estimated by image analysis of the band intensity, which corresponds to the amount of un-cleaved RNA after S1 incubation. All the amounts were normalized respect to control one at 30ºC.
The results obtained were used to determine the RNAs folding properties and then compared to those for the tRNA standard and population average (Figure 3.8). The criterion is such that the most stable sequences at high temperature are those with a more stable secondary and eventually tertiary structure.

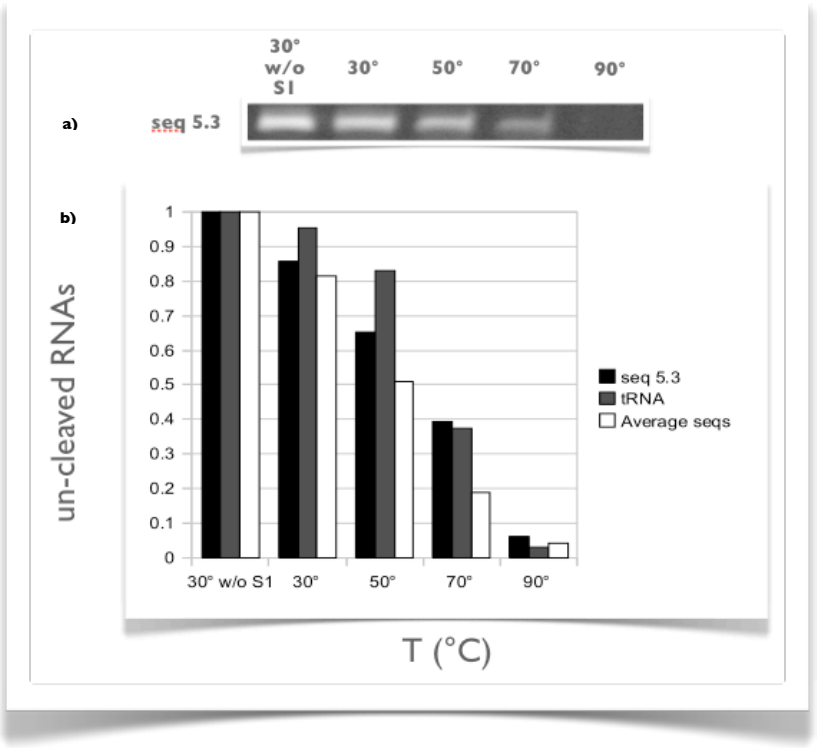

Figure 3.8 : RNA Foster on a single Never Born RNA. a) 12% Page urea gels of the residual amount of RNAs after each test temperatures (30, 50, 70 and 90°C). A control reaction was set incubating the sequence for 60' at 30°C. b) Plot showing the residual RNA amount (in black) at the different temperatures. The fraction of folded RNA for each temperature was estimated by image analysis of the band intensity, which corresponds to the amount of un-cleaved RNA after S1 incubation. All the amounts were normalized respect to control one at 30ºC. The sequence stability was compared with the one of the natural tRNA (in grey) and the population average (in white).

Except for the sequence 21 (Table 3.1), all the RNAs were resistant to S1 nuclease digestion at 30ºC (Figure 3.9, panel 30ºC), indicating that the RNA molecules could posses a stable secondary structure at that temperature under the conditions used during the assay.
As the temperature increases, the number of RNA molecules sensitive to S1 nuclease increases as well. Assuming a degradation rate lower than 50%, as criterion of stability, we observed that 60% of the sequences analyzed retains a stable secondary structure at up to 50ºC (Figure 3.9, panel 50ºC). Furthermore, an interesting result, also in view of the restricted number of analyzed sequences and their random fashion, was the comparable stability observed for two random sequences (number 13 and 16) and the natural tRNA at 50ºC.
Instead, at 70ºC all the sequences, including the tRNA standard, unfolded and were degraded by S1 nuclease to less than 50% (Figure 3.9, panel 70ºC).
Finally, at 90ºC all the sequences were readily degraded by S1 (Figure 3.9, panel 90ºC).
These results can be explained assuming that most of the RNA nucleotides were engaged in some kind of secondary structures up to the temperature of 70ºC, which would prevent the activity of S1 nuclease. Surprisingly, the sequence 13 presents a stability at 70ºC close to 50%, showing that its melting temperature was around 70ºC (Figure 3.9, panel 70ºC). Note that both sequence numbers 13 and 5.3 are more stable than the standard tRNA at the same temperature (70ºC). This supports the assumption that the folding is not a unique property of biologically extant RNAs (which have undergone a lengthy process of evolution); and, by inference, this means that the sequence space of the NB-RNAs could be within reach of folded stable structures. It is noteworthy that the tested RNAs were randomly chosen, and no selection was exerted on them prior to the analysis, so that finding a sequence nearly stable at 70ºC comes unexpectedly and support the hypothesis that thermostable RNAs may not be so rare in the RNA sequence space (Figure 3.9).

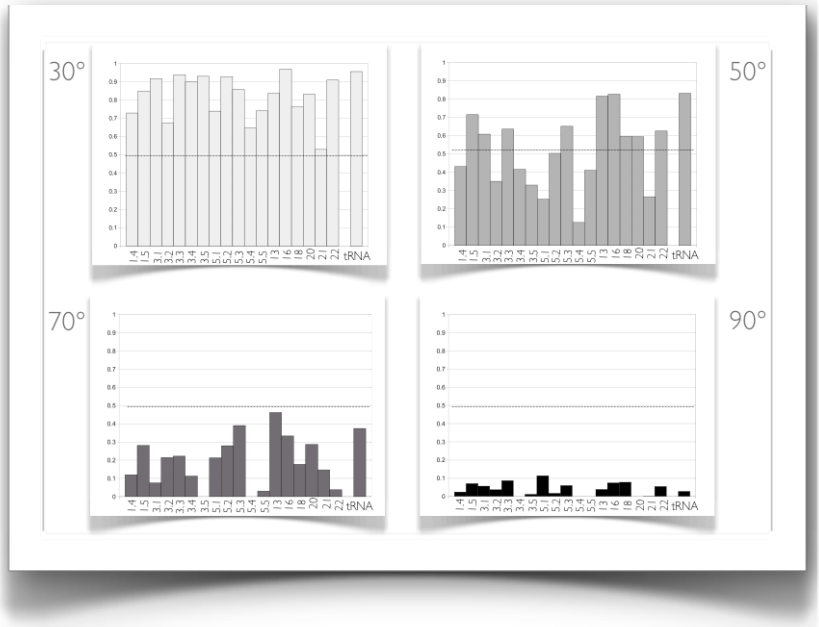

Figure 3.9: RNA Foster Assay summary for the eighteen analyzed sequences. The image shows the plots at each experimental temperature (30, 50, 70 and 90°C) for each random RNA sequence (x-axis). In y-axis the amount of un-cleaved RNA.

The investigation of the relationship between GC content and thermal stability showed no significant correlation as shown in Figure 3.10. This result corroborates previous findings and suggests that random RNA thermal stability is not univocally determined by sequence composition. This lack of correlation can be explained by assuming that the tertiary structures and long-range interactions play a crucial role in stabilizing the RNAs, coherently with the notion that a stable three-dimensional structure increases secondary structure stability too, shifting the Tm of the individual secondary domains towards higher values.

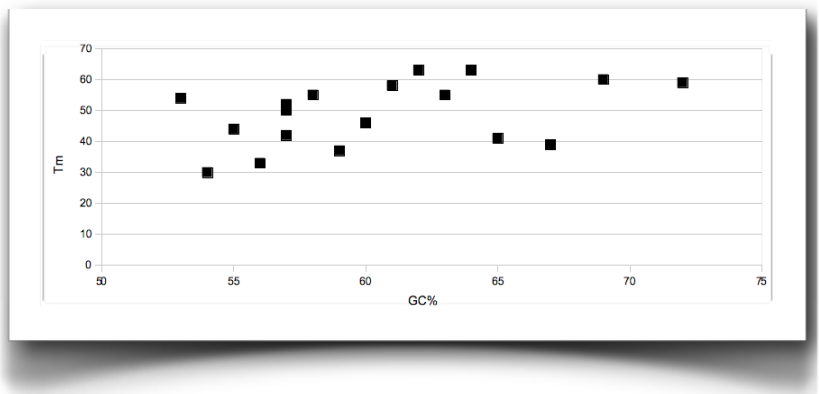

Figure 3.10: Correlation between thermal stability and GC content. Plot showing the RNAs melting temperatures (T_m) plotted versus their GC content. No significant correlation could be found.

3.4 In vitro selection for thermal stability of the Never Born RNA library

In light of the results obtained from the studies on the stability of single Never Born RNAs by RNA Foster, we decided to extend our analysis to a larger number of RNAs, in order to investigate if sequences stable at high temperatures could be found in a random RNA population.
With this aim we used the *in vitro selection* (Klug and Famulok, 1994), a technique based on the fundamental principles of Darwinian theory, which in general allows the exploration of functional properties of DNAs, RNAs and proteins in a quick, systematic and direct way, to select the random RNA library. Beginning with a broad population of RNAs we applied a selective pressure, then we recovered, reverse-transcripted and amplified once again those molecules which satisfy the established condition (Figure 3.11).
Then the population can be submitted to a new in vitro selection cycle. By doing this we isolated sequences that satisfy a well determined request.

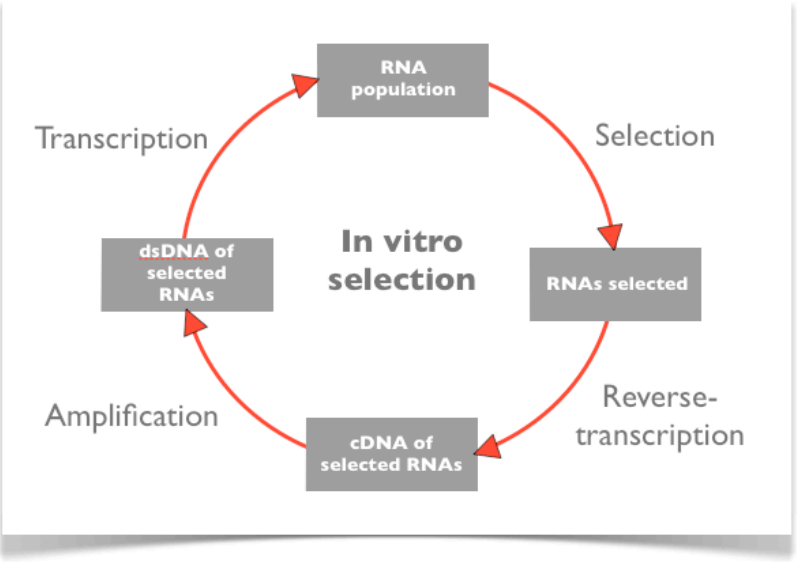

Figure 3.11: Summary of the in vitro selection. Once produced the RNA population, it is selected for a established property, then the molecules which satisfy such condition are recovered, reverse-transcipted and amplified. Then the population can be submitted to a new in vitro selection cycle. By doing this we isolated sequences that satisfy a well determined request.

In detail, the first step was the DNA library transcription to produce the corresponding RNA population. About 300 ng of DNA plasmid library, linearized by *XbaI* digestion, were used as template for the transcription by T7 polymerase of the RNAs, and later on they were submitted to a Foster assay at the test temperatures of 30°C.
After the test the un-cleaved RNAs were recovered by reverse transcription using the RT-PrimerRv11, which has an annealing site on the constant region downstream the random sequence, and the relative cDNAs were produced. At this point, by PCR reaction using the primers RT-Primer-RV11 and RT-Primer-Fw11, the double-stranded DNAs were obtained (Figure 3.12).
Such DNAs were double digested by *NotI* / *XbaI* enzymes (Figure 3.13), in order to be readily cloned into a vector (double digested by *NotI* / *XbaI* as well).

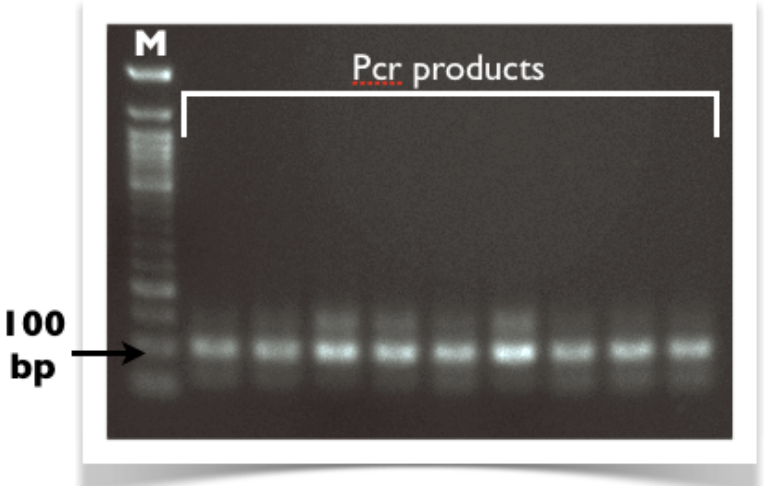

Figure 3.12: 2% Agarose gel. PCR reactions on the cDNA synthesized from the un-cleaved RNAs after 30° selection cycle. The aspected length is 102 bp, completely comparable with the amplification profiles obtained. ***M*** *: 50 bp Marker from NEB.*

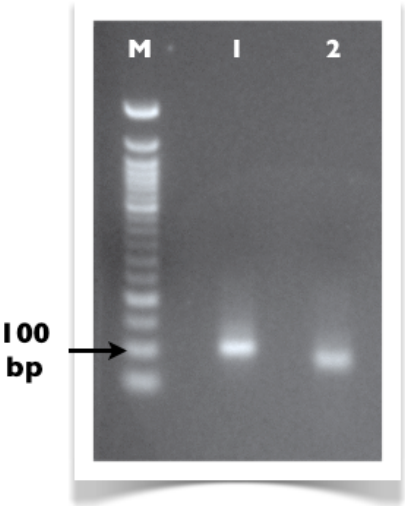

Figure 3.13: 2% Agarose gel. Double digestion check of the insert. ***M****: 50bp marker from NEB;* ***1****: dsDNA produced from the cDNA of the recovered RNAs;* ***2****: sample 1 double digested by NotI and XbaI enzymes.*

The ligated product (see Materials and Methods) was inserted into *E. coli* cells (Electro Ten-Blue strain) by electro-transformation and plated on agar to obtained the selected RNA sub-library.
The RNA population was transcripted again and a new cycle of selection performed, increasing the test temperature up to 50°C end the whole cycle was repeated once again. After the two cycles of selection a representative number of clones were recovered and sequenced (Table 3.2).

Selected sequences after 30°C cycle	Selected sequences after 50°C cycle
	AGTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTC
	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
TGTAGATCAATTCCGCCTCTCTGCCCCCCCGTGCACCGACCAGGACCCAACACCCCCCAG	ACCGTCTTTACTCAATACGCCATCCCCAAAGTAAATATACGGTGCCCGTACTCCGCCAAC
ACCTCTCTCAAATTCCACATGGGCATGACAACCCCTCAACTTCCACCTACCTGCCCGGTA	AGATGGTACAGCTGAGCAACCCACCATGATTACCCCCAATACTATCCACGCGCCCTCCTG
GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA	AGTTCCTAAGATTGGCAAGCCAGGGCCGAGCAAAAGCACAGCTGCGTAATCCGACCACAT
TTATCCCCCCCCCCACCTTACACAGTGCCCCTTCGACGATTGACAAGAAGCGGTCCGCCA	GAGCAACCTAGTAAACCCCGTGATCATTACATCTCACAATCCCCAGGAGATTCGAGTGGT
GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
GCGGAAGCCAACACAAAACACGCAAGTAAGAAAGTCCAGACCGACGTGAATAGTGATTAT	GGCCCACTAGACACACACGATCAGGCCCAGCGCGTTATNNNGTACNCACACACATACCAN
GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA	TGATACCCCCCGAGAACAAACGTTGAATAGCATTCAGCTGCTAACAGGCGTCAGCACTCT
CTCCCTGGCTATACCACCAAACAACACCAAAGTCCACTTCGCTCGGCAAAAGAACAGTTT	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA	TGCTCCTCCCAGTGTCTAGCCTACCCCGCTCCGTTTGCTTCCAAGTATCCAGCTCCCGAT
TTACACTACCCACACCCTACTGGAACCCACTAAGCAATGATGGGTGCATGTGCACCCGGT	TCCCCGAACAAAGAAAAGAGAATACTGACTTCATACCCCCGCAAAGTGACACAGCTCCCN
GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA	TGCTCCTCCCAGTGTCTAGCCTACCCCGCTCCGTTTGCTTCCAAGTATCCAGCTCCCGAT
AGTTCCTAAGATTGACAAGCCAGGGCCGAGCAAAAGCACAGCTGCGTAATCCGACCACAT	AGTTCCTAAGATTGACAAGCCAGGGCCGAGCAAAAGCACAGCTGCGTAATCCGACCACAT
CGAAGCCCTCAACCAACTTTATAACTACGACCGCTATTCATCGGACAACCGTGCCCCTGC	TAACACGCCTCCGAAATGGAGTCGACAAAAGTTAACGGTAAGCTACCTATCCCTCCCCTT
	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
	GTCCTGGAACAGGCCGGTAAGCTCCCAAGACAAATCGTACGCCCTCGAACTGAAAAGTCA
	TTATCCCCCCCCCCACCTTACACAGTGCCCCTTCGACGATTGACAAGAAGCGGTCCGCCA

Table 3.2: Selected sequences after 30°C and 50°C selection cycles. With the same color the sequences present in multi-copy.

From the sequencing data we obtained some interesting results (Figure 3.14).
At 30°C, on the 14 sequences with the right frame analyzed, ten are present in single copy and only one is repeated four times. The population retains a 79% of diversity.
At 50°C, on 21 sequences with the right frame, 9 sequences are present only one time while three sequences are present in multi-copy. Within the last group one sequence is repeated 8 times and it is the same sequence found in four copies in the 30°C subpopulation, while the other two sequences are represented twice. At this step the population retains a diversity of 57%.

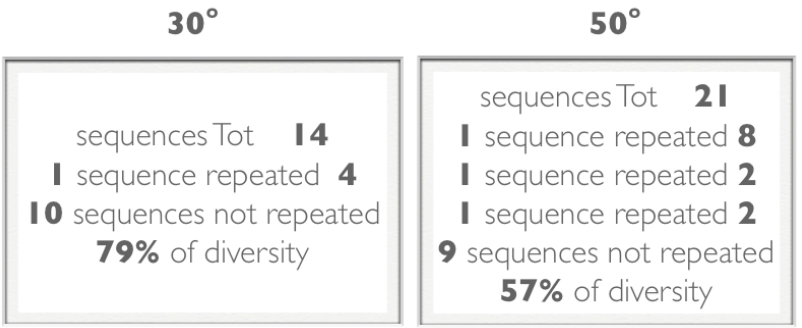

Figure 3.14: summary of the sequence distribution after 30°C and 50°C selection cycles

Furthermore, we found that, according to our predictions, the diversity of the RNA population tends to decrease as the selection keeps on, indicating a skimming working on it. In addition we found a sequence repeated 8 times at the end of the 50°C selection cycle, data which suggests a possible specific stability for such sequence. To verify this hypothesis the sequence number 81 was submitted to a Foster Assay at different temperatures in order to evaluate its stability (Figure 3.15).

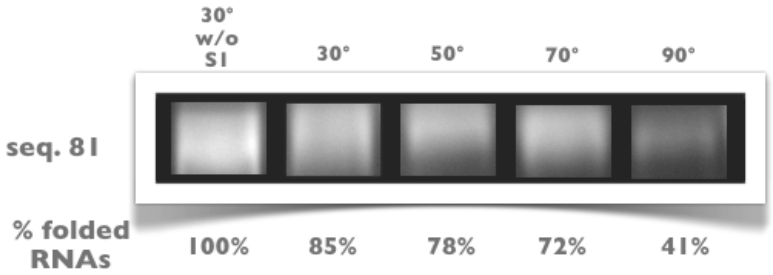

Figure 3.15: Foster Assay on the sequence 81, the most represented sequence after the 50°C selection cycle. On the top the temperatures test, on the bottom the fraction of folded RNAs

The sequence number 81 shows a stability up to 70°C, with an approximate melting temperature higher then 80°C.
The comparison between the stability of the selected sequence, the tRNA-Trp and the resistance average from the population composed by the eighteen individually analyzed sequences (see paragraph 3.3), shows how the sequence number 81 is more stable at 70° and 90°C (Figure 3.16).
In detail, at 90°C the selected sequence retains a fraction of folded RNA of 41% suggesting that at the next step, the selection at 70°C of the sub-population at 50°C, could be possible find sequences stable up to 90°C.

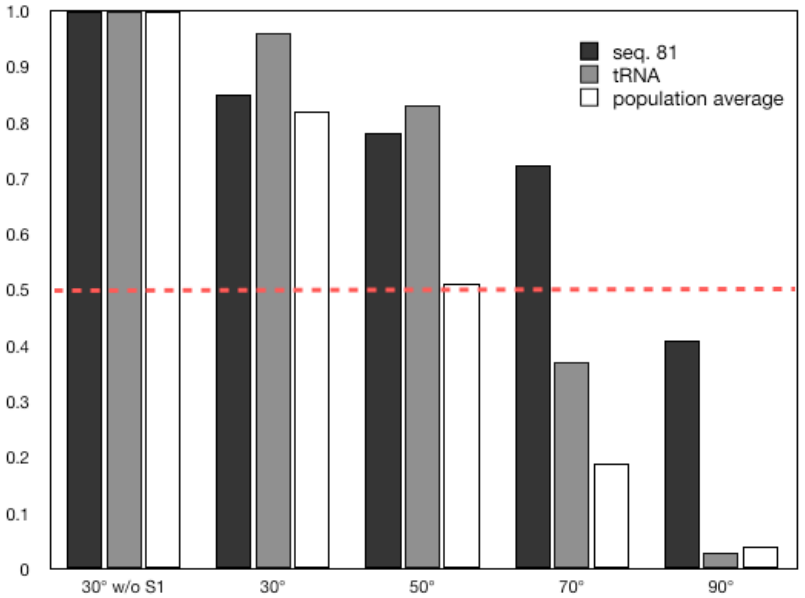

Figure 3.16: Plot representing the fraction of folded RNAs (%) at each experimental temperature. The results for the sequence 81 (in black) are compared with the stability of the tRNA-Trp (in grey) and the population average (in white) of the eighteen individually analyzed sequences.

The result confirms in a more strong way that stable RNA sequences could be found easily in a random population, supporting the idea that folding properties are intrinsic in the RNA molecule and are not the product of a lengthy evolution.

4 Discussion

The main aim of this research project was to estimate the frequency of the spontaneous RNA folding in a random library, to determine whether a stable folding is a common feature of RNAs or the fine result of natural selection.
I will briefly summarize the answers given by my work to those questions, and discuss the single aspects one by one.
The results will be discussed in terms of RNA random library construction (see paragraph 4.1), preliminary analysis of random RNA structural properties of (paragraph 4.2) and possible outlooks (paragraph 4.3).

4.1 Construction of a RNA random library

To evaluate the occurrence of spontaneously folded RNAs, a random DNA library was designed and produced. The library codifies for 0.85×10^9 RNA different sequences 97 residues long.
The library construction is a remarkable achievement because of its complexity and the completely RNA randomness according to the nature of the DNA sequences used to produce the library.
The construction of RNA random libraries is a common molecular biology procedure nowadays, especially in the attempt to discover potentially active RNA molecule. However, RNA random libraries reported in literature are actually designed as random regions engrafted on a constant scaffold inspired to known natural functional domains.
In contrast, this library is designed in a complete random fashion. The only constrain imposed is the presence at the edges of the random region of two constant sequences required for the primers annealing and cloning reactions. The two constant regions codifies for 37 nucleotides out of 97 residues reducing the randomness of the RNA sequence of only the 38%. In comparison, some RNA libraries known from literature are constituted by constant sequences, which represent up to the 50% of the "random sequence" (Davis and Szostak, 2002). Moreover, the two constant regions present in this library do not share any homology with extant functional domains.
About the actual complexity of the RNA random library (0.85×10^9), it results to be some order of magnitude smaller than the primary DNA library it derives from. The low yield is due to the low efficiency of the cloning step and was limited by the technical bottleneck of plasmid transformation. The lower efficiency of the transformation step is an unavoidable limit in this kind of procedures. However, 0.85×10^9 different RNA sequences are still a remarkable collection of random RNA molecules.

4.2 About the structural properties of a random RNA library

In the project, the RNA Foster was successfully used to evaluate the structural properties of 18 random RNAs. RNA molecules were tested individually at 4 different temperatures, in order to evaluate their ability to form internal double helices and, thus, to acquire secondary and tertiary structures. The results were used to assess the temperature of the main unfolding transition of individual RNAs. The experimental data show that almost all screened RNAs possess folded structures at 30°C. Furthermore, half of the sampled RNAs maintain the fold up to 50°C. Even though the limited number of sequences explored is not statistically significant, collected data suggest that the folding capability is a relatively common property, almost an intrinsic feature, of RNAs. Because a well-defined folding is essential to acquire any biological function, one may claim that the possibility of a spontaneous emergence of a potentially functional RNA is a highly probable event. This is an important issue from the biotechnological point of view; if folded RNAs are common in random RNA libraries the probability of selecting potentially functional RNAs by mean of large RNA pool screening becomes high.
The most general result of our studies lies in the demonstration that RNAs have the capacity to fold into compact structures, even in the absence of selective pressure. In other words, the stable fold appears to be an intrinsic tendency, as already mentioned in our previous work for longer RNAs (De Lucrezia et al., 2006b) . In addition, Schuster demonstrated that the sequence-to-structure space is of the kind many-to-one, where the RNA sequences share common shapes even if they have little or no primary structure homology (Schuster, 1995). In this way, different RNA sequences can have the same tertiary structure and possibly the same activity. Such RNA tendency, coupled to the highly redundancy of RNA sequence-to-structure map, could suggest a possible answer to the problem of a critical RNA concentration for a workable chemical system in the RNA World scenario. In other words, if a critical concentration is necessary, only zeptomoles of RNA molecules, as proposed by Knight and Yarus (Knight and Yarus, 2003), for the beginning of an RNA metabolism, such a problem in a prebiotic scenario could have been overpassed by the presence of different RNA sequences presenting the same active structure (Luisi, 2006; De Lucrezia et al., 2007). Of course, the missing link is still the prebiotic synthesis (enzyme-free) of long RNA chains in many identical copies, but some hints in the right direction appear to come from the condensation reactions of shorter oligonucleotides under alleged prebiotic conditions (Rehatgi et al., 1996; Pino et al., 2008).

All these considerations are important for the Origin of Life debate. In fact, it appears possible to suggest that the emergence of de novo stable RNAs is a probable event. In this regard, it is worth noticing that the experimental length of our NB-RNAs (97 nucleotides), as already mentioned, corresponds to the length of ribozymes and, therefore, one can see the relevance for the possible RNA World scenario (Cech, 1993; Muller, 2009).
In addition from the in vitro selection of the RNA population we got very interesting results: the in vitro selection reduced the population diversity and the most represented sequence after the selection at 50 degrees shows a very high stability, namely over 70 degrees.
Those preliminary studies support in a stronger way our idea that the RNA molecules have an intrinsic fold stability, even at high temperature, not exclusive of the natural RNAs.
All those results implying a support to the RNA world theory and providing a model for the early stage of the RNA evolution.
Finally such outstanding results could be a further support to the contingency theory, because could demonstrated that the real is not as rich as the possible!

4.3 Outlooks

The large library of completely random RNAs that were produced may represent a tank of never-investigated molecules. Thus, we plan to screen this library in an attempt to select RNAs with potential biotechnological applications.
The RNA Foster was used during this project to test the structural properties of RNA molecules. We plan to use the RNA Foster as pre-screening methodology to improve the efficiency of random RNA libraries screening.
Furthermore, the preliminary analysis of the random RNAs structural properties gave interesting and unexpected results that may have strong implication for the evolution of Life on Earth studies. A well-defined fold is essential to acquire any biological function and it does seem a relatively common feature of randomly synthesized RNA. Thus, one might claim that if the fraction of foldable RNAs among all random RNA sequences is large, then the emergence of potentially functional RNAs might have been a common event during the first steps of Life's evolution on Earth. In other terms, if this observation will be confirmed by further investigations, RNA molecules with unique structure-function features could have evolved easily in the prebiotic scenario. However, it must be stressed that in the present study the correlation between foldability and activity has not been investigated.

Thus, the suggestion that functional RNAs might have been common during the early evolution of life remains only a speculation, although intriguing. In this respect, on the basis of the obtained results, a further development could be the extensive study of the random RNA structural properties with the aim to elucidate their role in the early evolution scenario.

5 References

Anders, E. (1989): Pre-biotic organic matter from comets and asteroids. *Nature* **342**, 255- 257

Anella, F., De Lucrezia, D., Chiarabelli, C. and Luisi, P.L. (2011): Stability studies on random folded RNAs (Never Born RNAs), implications for the RNA World. *Chem. Biodivers.* **8**, 1422-1432

Arrhenius, S. (1907): Panspermy: The transmission of life from star to star. *Scientific American*, **196**

Bachmann, P.A., Luisi, P.L. and Lang, J. (1992): Autocatalytic self-replicating micelles as model for prebiotic structures. *Nature* **357**, 57-59

Bada, J.L. and Lazcano, A. (2002): Some like it hot, but not the first biomolecules. *Science* **296**, 1982-1983

Benner, S.A., Ellington, A.D. and Tauer, A. (1989): Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. USA* **86**, 7054-7058

Bernal, J.D. (1949): The physical Basis of Life. *Proc. R. Soc. Lond*. **357A**, 537-558

Butlerow, A. (1861): Formation synthetique d'une substance sucree. *Compt. Rend. Acad. Sci.* **53**, 145-147

Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986): Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature* **323**, 349-353

Cairns Smith, G. (1986): Clay Minerals and the Origins of life. *Cambridge University Press*, Cambridge

Cech, T.R. (1993): The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **135**, 33-36

Chyba, C.F., Thomas, P.J., Brookshaw, L. and Sagan, C. (1990): Cometary delivery of organic molecules to the early Earth. *Science* **249**, 366-373

Darwin, C. (1859): On the origin of Species by Means of Natural selection. *John Murray*, London

Darwin, F. (1888): The life and letters of Charles Darwi. *John Murray*, London

Davis, J.H. and Szostak, J.W. (2002): Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci. USA* **3**, 11616-11621

De Duve, C. (2002): Life evolving: molecules, mind and meaning. *Oxford University Press*, Oxford, UK

De Lucrezia, D., Anella, F. and Chiarabelli, C. (2007): On the chemical reality of the RNA World. *Orig. Life Evol. Biosph.* **37**, 379-385

De Lucrezia, D., Franchi, M., Chiarabelli, C., Gallori, E and Luisi, P.L. (2006a): Investigation of de novo totally random biosequences, part III: RNA Foster: A novel assay to investigate RNA folding structural properties. *Chem. Biodivers.* **3**, 860–868

De Lucrezia, D., Franchi, M., Chiarabelli, C., Gallori, E. and Luisi, P.L. (2006b): Investigation of de novo totally random biosequences, part IV: folding properties of de novo, totally random RNAs. *Chem. Biodivers.* **3**, 869–877

Deamer, D.W. (1985): Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* **317**, 792-794

Deamer, D.W. (1998): Possible starts for primitive life. In *Molecular Origins of Life*, ed. A. Brack, Cambridge University Press

Di Giulio, M. (2003): The universal ancestor and the ancestor of bacteria were hyperthermophiles. *J. Mol. Evol.* **57**, 721-30

Draper, D.E. (1996): Strategies for RNA folding. *Trends Biochem Sci.* **21**,145-149

Dworkin, J.P., Lazcano, A. and Miller, S.L. (2002): The roads to and from the RNA world. *J. of Theoretical Biology* **222**, 127-134

Ehrenfreund, P., Irvine,W., Becker, L., Blank, J., Brucato, J.R., Colangeli, L., Derenne, S., Despois, D., Dutrey, A., Fraaije, H., Lazcano, A., Owen, T. and Robert, F. (2002): Astrophysical and astrochemical insights into the origin of life. *Prog. Phys.* **65**, 1427-1487

Ferris, J.P. (1993): Catalyses and prebiotic RNA synthesis. *Orig. Life Evol. Biosph.* **23**, 307-315

Ferris, J.P. and Ertem, G. (1993): Montmorillonite catalysis of RNA oligomer formation in acqueous solution. A model for the prebiotic formation of RNA. *J. Am. Chem. Soc.* **115**, 12270-12275

Ferris, J.P., Joshi, P.C., Wang, K.J., Miyakawa, S. and Huang, W. (2003): Catalysis in prebiotic chemistry application to the synthesis of RNA oligomers. *Advances in Space Research* **33**, 100-105

Franchi, M., Bramanti, E., Bonzi, L.M., Orioli, P.L., Vettori C. and Gallori, E. (1999): Clay- nucleic acid complexes: characteristics and implications for the preservation of genetic materials in primeval habitats. *Orig. Life Evol. Biosph.* **29**, 297-315

Fuller, W.D., Orgel, L.E. and Sanchez, R.A. (1972): Studies in prebiotic synthesis : VII. Solid- state synthesis of purine nucleosides. *J. Mol. Evol.* **1**, 249-257

Gilbert, W. (1986): Origin of life – the RNA world. *Nature* **319**, 618

Gould, S.J. (1989): Wonderful life. *Penguin Books*, London

Guerrier-Takada C., Gardiner K., Marsh T., Pace N. and Altman, S. (1983): The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857

Hafenbradl, D., Keller, M., Wächtershäuser, G. and Stetter, K.O. (1995): Primordial amino acids by reductive amination of a-oxo acids in conjunction with the oxidative formation of pyrite. *Tetrahedron Lett.* **36**, 5179-5182

Haldane, J. B. S. (1928): Possible Worlds. *Hugh & Bros*, New York

Halliday, A.N. (2000): Terrestrial accretion rates and the origin of the Moon. *Earth Planet Sci. Lett.* **176**, 17-30

Itoh, T. and Tomizawa, J. (1980): Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl. Acad. Sci. USA* **77**, 2450-2454

Joyce, G.F. (2002): The antiquity of RNA-based evolution. *Nature* **418**, 214-221

Kasting, J.F. (1993): Earth's early atmosphere. *Science.* **259**, 920-926

Kawamura, K. and Ferris, J.P. (1994): Kinetic and mechanistic analysis of dinucleotide and oligonucleotide formation from the 5'-phosphroimidazolide of adenosine on Na^+- montmorillonite, *J. Am. Chem. Soc*., **116,** 7564-7572

Kebbekus, P. (1988): Formation of RNA oligomers on montmorillonite clay under possible prebiotic conditions. *Chemistry Rensselaer Polytechnc Institute*, Troy NY

Keefe, A.D., Miller, S.L., McDonald, G. and Bada, J. (1995): Investigation of the prebiotic synthesis of amino acids and RNA bases from CO2 using FeS/H_2S as a reducing agent. *Proc. Natl. Acad. Sci. USA* **92**, 11904-11906

Klug, S.J. and Famulok M. (1994): All you wanted to know about SELEX. *Mol. Biol. Rep.* **20**, 97-107

Knight, R. and Yarus, M. (2003): Finding specific RNA motifs: function in a zeptomole world?. *RNA* **9**, 218–230

Knight, R., De Sterck, H., Markel, R., Smit, S., Oshmyansky, A. and Yarus M. (2005): Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. *Nucleic Acids Res.* **33**: 5924–5935

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982): Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147-157

Lazcano, A. and Miller, S.L. (1996): The origin and early evolution of life: prebiotic chemistry, the pre-RNA world and time. *Cell Press* **85**, 793-798

Lazcano, A., Guerrero, R., Margulis, L. and Orò, J. (1988): The evolutionary transition from RNA to DNA in early cells. *J. Mol. Evol.* **27**, 283-290

Lipman, D.J., Souvorov, A., Koonin, E.V., Panchenko, A.R. and Tatusova, T.A. (2002): The relationship of protein conservation and sequence length. *BMC Evolutionary Biology* **2**, 20- 30

Lisacek, F., Diaz, Y. and Michel, F. (1994): Automatic identification of group-I intron cores in genomic DNA sequences. *J. Mol. Biol.* **235**, 1206–1217

Lugmair, G.W. and Shukolyukow, A. (1998): Early solar system timescales accordino to ^{53}Mn-^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62**, 2863-2886

Luisi, P.L. (1998): About various definitions of life. *Orig. Life Evol. Biosph.* **28**, 613-622

Luisi, P.L. (2003): Contingency and determinism. *Phil. Trans. R. Soc. Lond. A* **361**, 1141- 1147

Luisi, P.L. (2006): The emergence of life. From chemical origins to synthetic biology. *Cambridge University Press*, Cambridge

Luisi, P.L., Chiarabelli, C. and Stano, P. (2006): From Never Born Proteins to Minimal Living Cells: Two Projects in Synthetic Biology. *Orig. Life Evol. Biosph.* **36**, 605-616

Margulis, L. and Sagan, D. (1995). What is life?. *University of California Press*, Berkeley and Los Angeles, California

Miller, S.L. (1953): A production of amino acids under possible primitive Earth conditions. *Science* **117**, 528

Miller, S.L. (1955): Production of some organic compounds under possible primitive Earth conditions. *J. Am. Chem. Soc.* **77**, 2351-2361

Mojzsis, S.J., Harrison,T.M. and Pidgeon, R.T. (2001): Oxygen isotope evidence from ancient zircons for liquid water at Earth's surface 4,400 Myr ago. *Nature* **409**, 178-181

Morowitz, H.J. (1992): Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis. *Yale University Press*, New Haven

Moulton, V., Gardner, P.P., Pointon, R.F., Creamer, L.K., Jameson, G.B. and Penny, D. (2000): RNA folding argues against a hot-start origin of life. *J. Mol. Evol.* **51**, 249-257

Muller, U.F. (2006): Re-creating an RNA world. *Cell. Mol. Life Sci.* **63**, 1278-1293

Muller, U.F. (2009): Evolution of ribozymes in an RNA world. *Chem. Biol.* **16**, 797-798

Noller, H.F., Hoffarth, V. and Zimnlak, L. (1992): Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416-1419

Oparin, A.L., (1938): The origin of life. *MacMillan,* New York

Orgel, L.E. (2003): Some consequences of the RNA world hypothesis. *Orig. Life Evol. Biosph.* **33**, 211-218

Orgel, L.E. (2004): Prebiotic Chemistry and the Origin of the RNA World, *Critical Reviews in Biochemistry and Molecular Biology* **39**, 99-123

Orò, J. (1960): Synthesis of adenina from ammonium cyanide. *Biochem. Biophys. Res. Comm*. **2**, 407-412

Pace, N.R. (1997): A molecular view of microbial diversity and the biosphere. *Science* **276**, 734-740

Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrel, K.A. and Cheng, H.L. (1986): A self-splicing RNA excise an intron lariat. *Cell* **44**, 213-223

Peyser, J.R. and Ferris, J.P. (2001): The rates of hydrolysis of thymidyl-3',5'-thymidine H phosphonate: the possible role of nucleic acids linked by diester of phosphorus acid in the origins of life. *Orig. Life Evol. Biosph.* **31**, 363-380

Pino, S., Ciciriello, F., Costanzo, G. and Di Mauro, E. (2008): Non enzymatic RNA ligation in water. *J. Biol. Chem.* **283**, 36494-36503

Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R. and Bruening, G. (1986): Autolytic Processing of Dimeric Plant Virus Satellite RNA. S*cience* **231**, 1577-1580

Rehatgi, R., Bartel, D.P. and Szostak, J.W. (1996): Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3'-5' phosphodiester bonds. *J. Am. Chem. Soc.* **118**, 3340-3344

Reid, C. and Orgel, L.E. (1967): Synthesis of sugars in potentially prebiotic conditions. *Nature* **216**, 455

Robertson, M.P. and Miller, S.L. (1995): An efficent prebiotic synthesis of cytosine and uracil. *Nature* **375**, 772-772

Saville, B.J. and Collins, R.A. (1990): A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **61**, 641-671

Scanlon, K.J. (2004): Anti-Genes: siRNA, Ribozymes and Antisense. *Current Pharmaceutical Biotechnology* **5**, 415-420

Schopf, J.W. (1993): Microfossils of the early Archaean Apex chest, new evidence for the antiquity of life. *Science* **260**, 640-646

Schuster, P. (1995): How to search for RNA structures theoretical concepts in evolutionary biotechnology. *J. Biotechnol.* **41**, 239–257

Schuster, P., Fontana, W., Stadler, P.F. and Hofacker, I.L. (1994): From sequences to shapes and back – a case-study in RNA secondary structures. *Proc. Natl. Acad. Sci. USA* **255**, 279–284

Schwartz, A.W. (1997): Prebiotic phosphorus chemistry reconsidered. *Orig. Life Evol. Biosph.* **27**, 505-512

Shapiro, R. (1999): Prebiotic cytosine synthesis: a critical analysis and implications for the origin of life. *Proc. Natl. Acad. Sci. USA* **96**, 4396–4401

Sharmeen, L., Kuo, M.Y.P., Dinter-Gottlieb, G. and Taylor, J. (1988): Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.* **62**, 2674-2679

Sharp, P. (1991): Five easy pieces. *Science* **253**, 663

Socha, R.F., Weiss, A.H. and Sakharov, M.M. (1980): Autocatalysis in the formose reaction. *React. Kinet. Catal. Lett.* **14**, 119-124

Steitz, T.A. and Moore P.B. (2003): RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* **28**, 411-418

Stribling, R. and Miller, S.L. (1997): Energy yields for hydrogen cyanide and formaldehyde synthesis : the HCN and amino acid concentrations in the primitive ocean. *Orig. Life Evol. Biosph.* **17**, 261-273

Szostak, J.W., Bartel, D.P. and Luisi, P.L. (2001): Synthesizing life. *Nature* **409**, 387-390

Van der Velden, W. and Schwartz, A.W. (1977): Search for purine and pyrimidines in the Murchison meteorite. *Geochim Cosmochim Acta* **41**, 961-968

Vogt, VM. (1973): Purification and further properties of single-strand-specific nuclease from Aspergillus oryzae. *Eur J Biochem.* **33**, 192-200

Wächtershäuser, G. (1988): Before enzymes and templates, theory of surface metabolism. *Microbiol. Rev.* **52**, 452-484

Walde, P., Wick, R., Fresta, M., Mangone, A. and Luisi, P.L. (1994): Autopoietic self- reproduction of fatty acid vescicles. *J. Am. Chem. Soc.* **116**, 11649-11654

Wetherill, G.W. (1990): Formation of the Earth. *Annu. Rev. Earth Planet Sci.* **18**, 205-256

White, H.B. (1976): Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101-104

Wilde, S.A., Valley, J.W., Peck, W.H. and Graham, C.M. (2001): Evidence from detrital zircons for the existence of continental crust and oceans on the Earth. *Nature* **409**, 175-178

Wilson, D.W. and Szostak, J.W. (1999): In vitro Selection of Functional Nucleic Acids. *Ann. Rev. Biochem.* **68**, 611-648

Woese, C.R. (1965): On the evolution of genetic code. *Proc. Natl. Acad. Sci. USA* **54**, 1546-1552

Xia Y. and Levitt M. (2004): Funnel-like organization in sequence space determines the distributions of protein stability and folding rate preferred by evolution. *Proteins: Structure, Function, Bioinformatics* **55**, 107-114

Acknowledgements