Two RNA Worlds: Toward the Origin of Replication, Genes, Recombination, and Repair

CHAPTER · DECEMBER 2010

DOI: 10.1007/978-3-642-21625-1_11

1 AUTHOR:

Dirk-Henner Lankenau

Universität Heidelberg

42 PUBLICATIONS **638** CITATIONS

SEE PROFILE

Chapter 11
Two RNA Worlds: Toward the Origin of Replication, Genes, Recombination, and Repair

Dirk-Henner Lankenau

Abstract All modern organisms depend on genomes that encode a diversity of RNA molecules functioning in a plethora of physiological, regulatory, and fundamental functions. Processes like gene transcription into mRNA, ribozyme catalyzed translation in the heart of ribosomes, RNA interference (RNAi), reverse transcription and defense of transposons, retroelements, homing mobility of introns, and many other characteristics of life represent the smoking gun of primordial RNA based, complementary base pairing driven search processes crucial for the origin of life and all successive life. This chapter overviews the necessary steps of RNA dependent evolutionary processes that lead to the emergence of replication, genes, recombination, and repair. The top to bottom journey into the past starts with the concept of Popper's deductive cycle. It leads us through established knowledge accompanied by the conserved mechanism of homology search engines as a key feature of the modern RNA world to the structure based ribozyme function of the catalytic center of the peptidyltransferase in modern ribosomes. We then jump over to the ancient RNA world in a bottom to top approach, a tactic taken by other chapters in this book as well. We use theoretical evidence for a ~55 nt protoribosome as the endpoint of this bottom to top journey and join again with the current RNA world by pointing out the statistical detection of an ancient RNY code in modern genes.

D.-H. Lankenau (✉)
Hinterer Rindweg 21, 68526 Ladenburg, Germany
e-mail: d.lankenau@t-online.de

R. Egel et al. (eds.), *Origins of Life: The Primal Self-Organization*,
DOI 10.1007/978-3-642-21625-1_11,

Abbreviations

BIR	break induced replication (DNA repair)
dsRNA	double-stranded RNA
ESS	evolutionary stable strategy
FRET	fluorescence resonance energy transfer
Ga	1 giga year (anus) = 10^9 years
IDA	initial Darwinian ancestor
LTR	long terminal repeat
LUCA(S)	last universal common ancestor(al) state
Ma	million years (ani) = 10^6 years
NAD	Nicotinamide Adenine Dinucleotide
ncRNA	non-coding RNA
NHEJ	non-homologous end joining (DNA repair)
nt	nucleotide
ORF	open reading frame
QED	quantum electrodynamics
RdRP	RNA dependent RNA polymerase
RNAi	RNA interference
RNP(C)	ribo nucleo protein (complex)
RNY	pu**r**ine – any **n**ucleotide – p**y**rimidine = hypothetical ancestor of the genetic code
RRR	replication, repair, recombination
RT	reverse transcriptase
SDSA	synthesis-dependent strand annealing (DNA repair) = mechanistic ancestor of RRR
SSA	single strand annealing (DNA repair)
ssRNA	single-stranded RNA

The origin of life was the chemical event or series of events whereby the vital conditions for natural selection first came about. The major ingredient was heredity.... The origin of life only had to happen once. We therefore can allow it to have been an extremely improbable event many orders of magnitude more improbable than most people realize... And the beauty of the anthropic principle[1] *is that it tells us against all intuition that a chemical model need only predict that life will arise on one planet in a billion to give us a good and entirely satisfying explanation for the presence of life here. (But) I do not for a moment believe the origin of life was anywhere near so improbable in practice.*

(Dawkins 2006)

The word emergence is in a sense the opposite of reductionism where reductionism is the view that any phenomenon can be explained by understanding the parts of that system.

(Hazen 2005)

11.1 Introduction: Popper's Deductive Cycle

The emergence of life is a complex labyrinth, a historical puzzle that cannot be approached directly. However, indirect methodologies are possible but where is Ariadne's thread? Figure 11.1 shows four interlocked contemporary attempts to crack the origin of life question and to frame a theory of life escorted by a deductive concept that embraces a multidisciplinary, unifying synthesis. One approach (Fig. 11.1b) searches the cosmos for alien or strange traces of recent or fossil life analogs hoping to define constraints that sharpen models for the emergence of terrestrial life (Baross et al. 2007). Another effort (Fig. 11.1d) tries to create artificial Darwinian systems in the laboratory, an approach called synthetic biology (Benner and Sismour 2005; Gibson et al. 2010). Most chapters in this book deal with a third methodology called prebiotic (geo-) chemistry (Chaps. 2, 4–7). They are rooted in Stanley Miller's famous discharge experiments (Miller and Urey 1959). Miller's approach addresses the primal self organization of life, starting from abiotic processes, meteorological, astro-geophysical, and geochemical in their nature. Chemistry, is known as the most innovative of the sciences with chemists able to create and design from scratch new organic and inorganic compounds, unique in the entire universe. Many of them serve the well being of humanity. Despite hidden, life threatening dangers such as toxicity, explosiveness, ozone depletion etc. caused by novel chemicals, our civilization would not exist as is, if food chemistry, material grades like plastics or medicals had not been invented and produced by chemistry. Often, the new compounds synthesized by chemists make up tools of all kinds in industry and technology, the colors we paint our homes with and the medicine we so often depend on. Physics, the fundament of science, is likewise responsible for raw materials (steel, silicon chips, semiconductors) as applied in electrical engineering, civil-, and mechanical engineering. Further, physics is known to drive our understanding of nature deeply into the essence of matter itself, discovering fundamental laws, new elementary particles, and presently the existence of dark energy and dark

[1]The anthropic principle was named by the mathematician Brandon Carter in 1974 (Barrow and Tipler 1986).

matter that both only seem to interact with our baryonic matter via gravitation and the weak nuclear force. Murray Gell-Mann once said: *You don't need something more to explain something more*.[2] What he meant was that all the elementary building blocks and laws that chemistry and physics discovered in history plus a bunch of accidents are sufficient to explain the emergence of something more, i.e., *life*. Most chapters in this book take the bottom to top approach to experimentally and theoretically explore and reconstruct chemical means and conditions that could have led to the emergence of life (Fig. 11.1). An impressive example of non-Darwinian, purely geo-chemical evolution has been put forward recently by Robert Hazen exploring the evolution of minerals during the formation of our eight planets (Hazen 2010). This analysis shows that the early and present Earth is unique in its mineralogical composition compared to all other planets in this solar system, meaning that the emerging life, as we know it, had a different likelihood to self organize on Earth compared to other extraterrestrial bodies. The chapters of the book try to create a feeling for those geo-chemical processes and settings from which proto-life might have chosen and from which the first entities of life developed that coincide with a most basic definition of life. As of now, we do not know the exact sequence of events life took to emerge. The "golden spike" of Fig. 11.1 has not been bridged yet but a hope evoking concept is on the rise. There is a plethora of chemical compounds from where life could have started. In fact, the Beilstein and Gmelin databases which Stanley Miller assumed as a starting assumption of what he might encounter in the Miller-Urey discharge experiment – include all known organic, metalloorganic and inorganic compounds, and list more than 10 million structures and ten million reactions with 37 million factual datasets. Like in cosmology using quantum physics to elaborate the history of the universe, it is difficult to build up a history of the prebiotic world from bottom-up (Fig. 11.1). The bottom-up approach in cosmology is impracticable because in the quantum world there would be a nearly unlimited number of possible histories to follow. Therefore, to understand the origin of the universe, cosmology today uses the top-down approach with a high certainty-amplitude following the anthropic principle as a starting point (Dawkins 2004; Hawking and Mlodinow 2010).[3] For this reason, it seems unlikely that we would ever be able to choose the right track toward life based on the Beilstein/Gmelin database alone (Fig. 11.1, bottom wedges B, C, D). Luckily, Stanley Miller found in his spark experiment that some compounds, particularly amino acids, were far more abundant than predicted compared to other compounds such as ribonucleotides (Miller and Urey 1959). Richard Egel points out in the introduction to this book, that some predictable patterns may underlie those chemical reactions relevant for prebiotic synthesis (Egel, this book). Fortunately, the biologist's anthropic approach strongly assists the equally relevant, Miller-bottom-

[2] http://www.youtube.com/watch?v=ONiWmzrmfuY.

[3] Here, I underly a most foundational element of Western philosophy as a firm initial seed of our anthropic analysis: René Descarte's matured dictum: "*Sum res cogitans*" (I am a thinking substance").

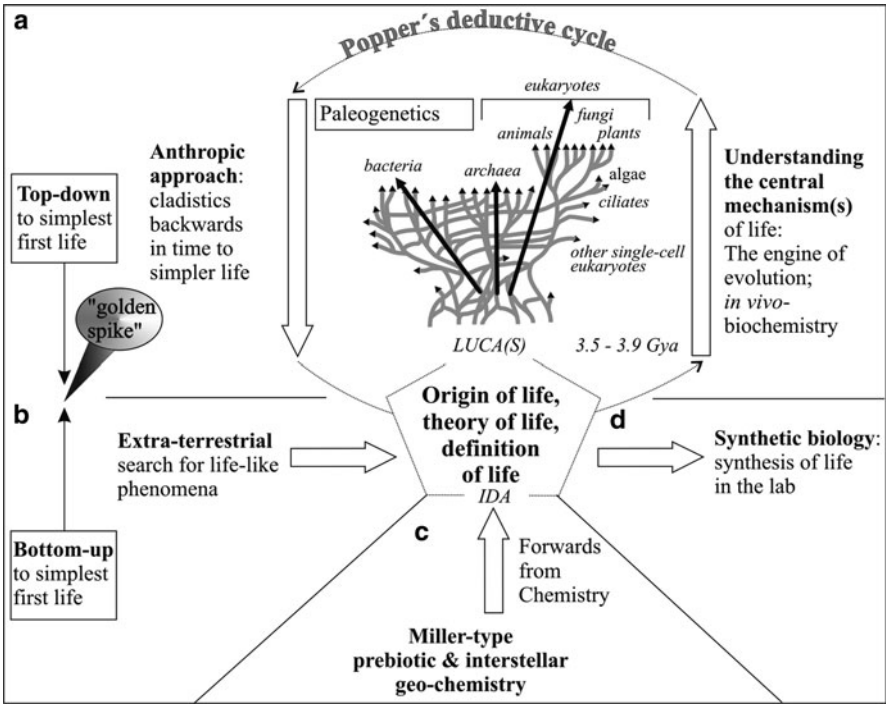

Fig. 11.1 The origin of life and the theory of life escorted by the deductive method – toward a multidisciplinary unifying synthesis. (**a**) Darwinian evolution and the origin of life conceptionalized by following the Popperian deductive cycle (Wächtershäuser 1997). It assumes that science moves from the current reality to the past particulars (*top to bottom*) and back to the general (same track *bottom to top*) – a *circling*, reductionistic, holistic process without end. The path backward in time is chosen by cladistic means following the Darwinian tree of life. As described in the text, first, Belozersky, Crick, Orgel, Woese, Britten & Davidson and Gilbert concluded that RNA was a primordial molecule of an RNA world. LUCAS was the Last Universal Common Ancestral State of all organisms but not a real individual entity as horizontal material, i.e., gene exchange, was common. LUCAS defines the ancient key molecules and metabolic processes present ubiquitously in contemporary life. Following these key processes from root level back-up the tree promises insight into the mechanisms of evolution and the historic and coincidental realities. Wedges (**b, c, d**) show non-Life Science approaches toward a theory of the origin of life. They can be seen as Popperian deductive mini-cycles that have the potential to fuse with any other deductive cycle. Stanley Miller's discharge experiments (**c**) represent the initial root of this bottom-up approach. Here, scientists look at chemical reactions that formed primeval, biologically relevant molecules. The goal is to analyze the variables leading to the initial Darwinian ancestor (IDA). The relevance of such reactions for the origin of life on the protoearth is then tested experimentally in the laboratory. Geologic and interstellar findings contribute further data of relevance. Between panel (**a**) on the one hand, and wedges (**b**, **c**, **d**) on the other hand resides the "golden spike": The "golden spike" is a paraphrase used by Wills and Bada to describe the bottom-up versus the top-down approach (for explanations see footnote 3). The *open pentagon* at the center represents the hub of interest for all parties. Benner recently noted that any definition

up approach (Fig. 11.1c). Biologists use the "backwards-in-time-to-ancient-life" method (Fig. 11.1a) in two ways: first, by simply reconstructing phylogenetic trees, and second, by resurrecting old, currently non-functional sequences and putting them into work in the cell in a Jurassic Park-like scenario after they lost or altered function millions of years ago. The latter approach is called paleogenetics (Dettai and Volff 2006; Eigen et al. 1985; Gaucher et al. 2003; Ivics et al. 1997; Liberles 2007; Noonan et al. 2006; Walisko et al. 2008). This biologist's approach in fact goes back to Charles Darwin. Since Darwin, biologists used the top to bottom approach to take advantage of the strong anthropic principle. They started their analysis from the most complex levels and varieties of contemporary life, that is the human brain on the one hand and modern species diversity on the other. Then, they observed, catalogued and systemized, and subsequently followed down the Darwinian tree of life using all available morphological traits of living and fossil organisms and from ubiquituously conserved molecular building blocks as a means to grasp the emergence of life and its evolving complexity. To this end, the phylogenetic systematics approach is the major quantitative means to reconstruct ancient beginnings (Dawkins 2004; Hennig 1950; Koonin et al. 2000). Chemists and crystallographers continued this anthropic path and by the same concept reached down and explored the most ancient of the molecular living-fossils. This "*Latimeria* of the molecular living fossils" is the ribosome, which has now been crystallized and its structure was revealed. The work was honored in 2009 with the Nobel prize in chemistry awarded to Venkatraman Ramakrishnan, Thomas A. Steitz, and Ada E. Yonath. Most important for our RNA world concept, however, was Ada Yonath's

Fig. 11.1 (continued) of life[4] must incorporate a "theory of life" (Benner et al. 2011). Vice versa, any theory of life must address the origin of life. A working-definition (Lankenau 2006), influenced by all wedges shown, that is compatible with Benner's requirements is taken as a basis. This biologist's, anthropic definition (for those who prefer a different) should at least help initially to focus efforts at understanding: *To comprehend the beginnings of life requires that we explain the origin of replication as well as of metabolism synergistically* (Maynard Smith and Szathmary 1997). *The genetic aspect of the modern definition of life was first proposed by Muller in 1966: "It is to define as alive any entities that have the properties of multiplication, variation and heredity"* (Muller 1966). *While metabolism supplies the monomers from which the replicators (i.e. genes) are made, replicators alter the kind of chemical reactions occurring in metabolism. Only then can natural selection, acting on replicators, power the evolution of metabolism.* The central idea incorporates the need of "instructive genetic information" that can replicate complements. Panel A further depicts the complex and simple three domains of life as discovered by studies on 16S rRNA (Woese and Fox 1977; Woese et al. 1990); Darwin's tree of life is rooted in the Last Universal Common ancestor (LUCA) that initially existed not as an entity but as the Last Universal Common Ancestral State (LUCAS) (Koonin 2009). Figure in part based on (Benner et al. 2011). The reticulated tree is based in part on (Doolittle 1999)

[4]The distinguished Asperger's Savant Daniel Tammet notes: "*Perhaps the most important logical errors to avoid are those caused by not being clear with definitions that we use. This is because careful and effective reasoning depends on precise definitions . . .*" (Tammet 2009, p. 251).

discovery of the perhaps 4 billion year old ancient catalytic pocket. This is an RNA molecule, i.e., a ribozyme, that carries out peptide bond linkages of tRNA-bound, activated amino acids. Ada Yonath and co-workers called it the protoribosome (Davidovich et al. 2009; Yonath 2009a, b, c). The protoribosome is definitely the most ancient of the ubiquitous living molecular fossils. It is real, not just hot air, and with its crystal structure and its catalytic activity we know a whole lot about it. And, what makes the case even stronger: *all* ribosomes in *all* organisms on earth use direct descendents of the same ancient ~55 nucleotide long RNA fragment to synthesize peptide bonds in the typical textbook peptidyltransferase reaction. For many scientists this fossil appears to be the closest we can get to the "golden spike"[5] using the top to bottom approach (Fig. 11.1). However, stringent theoretical considerations and experimental simulations allow us to push the top-down lead yet a little bit further beyond the golden spike (Fig. 11.1). Actually, we must conclude that there had to be an even older machine in evolution that secured the existence of the translating protoribosome itself. This was a prototype of the modern replication/repair/recombination (RRR) factories that enabled the protoribosome to either self-duplicate or to co-participate in hereditary processes together with other ribozyme species. Only nucleic acids have the property of forming complementarily paired double-stranded strings, a property immediately recognized in Watson and Crick's original paper: "*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for genetic material*" (Watson and Crick 1953b), and shortly later: "*The hypothesis we are suggesting is that the template is the pattern of bases formed by one chain of the deoxyribonucleic acid (in the RNA world it is RNA) and that the gene contains a complementary pair of such templates*" (Watson and Crick 1953a). Their hypothesis was confirmed by Meselson and Stahl experimentally showing that DNA is duplicated semi conservatively (Meselson and Stahl 1958a, b). I think, that from these ancient times onward to the present time (Fig. 11.1a right upward arrow), the RRR-factory, by its very nature experienced enormous functional, dynamic changes over 4 billion years of evolutionary history. This RRR-factory in all its facets, i.e., adaptations such as meiotic recombination, DNA-repair, transposition, mating-type

[5]"The Golden Spike": Fig. 11.1 shows the golden spike. While the bottom-up method to the origin of life is focused on the creation of some kind of self-replicating entity in the laboratory, starting with simple substances under some approximation of prebiotic conditions and ending up with a structure that has at least some properties of life, the top-down (anthropic) approach will succeed when it is possible to dissect modern life to its essentials, demonstrating the steps by which the elaborate machinery of living cells first appeared. Wills' & Bada's metaphor is based on a North American historical event. *The building of the transcontinental railroad, the Union and Central Pacific, was one of the great industrial achievements of the nineteenth century. Workers starting out from Omaha, Nebraska, in the east and Sacramento, California in the west met in triumph at Promontory Summit, northwest of Ogden, Utah, on May 10, 1869. There they drove in a golden spike to mark the railroad's completion. Like those gangs of railroad workers, science is working in two directions toward the origin of life. But rather starting from the east and the west, they are working from the top downward and the bottom upward* (Wills and Bada 2000).

switching, etc. is the core driving force of life's diversity and it represents the central interest of evolutionary biology attempting to penetrate the causes and mechanisms of evolutionary change (Lankenau 2006; Mayr 1963). This chapter sketches a reasonable, synthetic story of emerging RNA replicators (whose synonymous nature with transposons, viruses, and genes was recognized only later) synthesized from abiotic geo-metallo-molecular ancestors. Figure 11.1. shows that Popper's deductive cycle has taught us the advantages of starting out from the general (i.e., our human self, and our human complex brains), then moving to the particulars backward in time and back to the general – a circle without end, an anthropic approach – and hopefully, not just hot air about evolution[6] that happily will incorporate astro-geo-physico-chemical knowledge whenever it is meaningful.

11.2 The Current RNA World as Hint for an Ancient RNA World

11.2.1 The Modern RNA World

The concept of RNA as a primordial molecule preceding DNA was hypothesized first by A. N. Belozersky followed by Francis Crick, Leslie Orgel, and Carl Woese (Belozersky 1957, 1959; Crick 1968; Orgel 1968; Woese 1967) (see also Spirin 2005). In the early 1980s, catalytically active RNAs, called ribozymes were discovered (Guerrier-Takada et al. 1983; Kruger et al. 1982). Early on, it was recognized that RNA molecules may play a fundamental role in regulating eukaryotic gene expression (Britten and Davidson 1969) leading to a current paradigm shift in biology (Jordan and Miller 2008), and hooking RNA up with epigenetics. In search of the evolution of how catalytic RNA gave rise to the intron-exon structure of genes, Walter Gilbert coined the term "RNA world" (Gilbert 1986). First, Herbert Jäckle and colleagues showed experimentally that antisense RNA injection into *Drosophila* embryos produced mutant effects (i.e., "knockdowns") of the Krüppel gene (Rosenberg et al. 1985). In the mid 1980s, studies on the structure and function of Y-chromosomal lampbrushloops of *Drosophila hydei* lead to the insight that non-coding RNA (ncRNA) may be a transient recruiting tool for proteins functionally essential for fertility of the male sex (Hennig et al. 1989). The Y-loops also encompassed batteries of proviral retrotransposons that expressed an endogenous antisense RNA complementary to part of reverse transcriptase and the full RNase H (Huijser et al. 1988; Lankenau et al. 1988, 1994) (Fig. 11.2). This finding indicated that not only protein storage for the sake of maintaining fertility played a role but that a balanced parasitic/symbiotic immunity control of proviruses may be achieved by specific, endogenous antisense RNAs – a mechanism now

[6]Paraphrase on a paraphrase used for a title by Ramakrishnan: The Ribosome: Some Hard Facts about Its Structure and Hot Air about its Evolution (Ramakrishnan 2011).

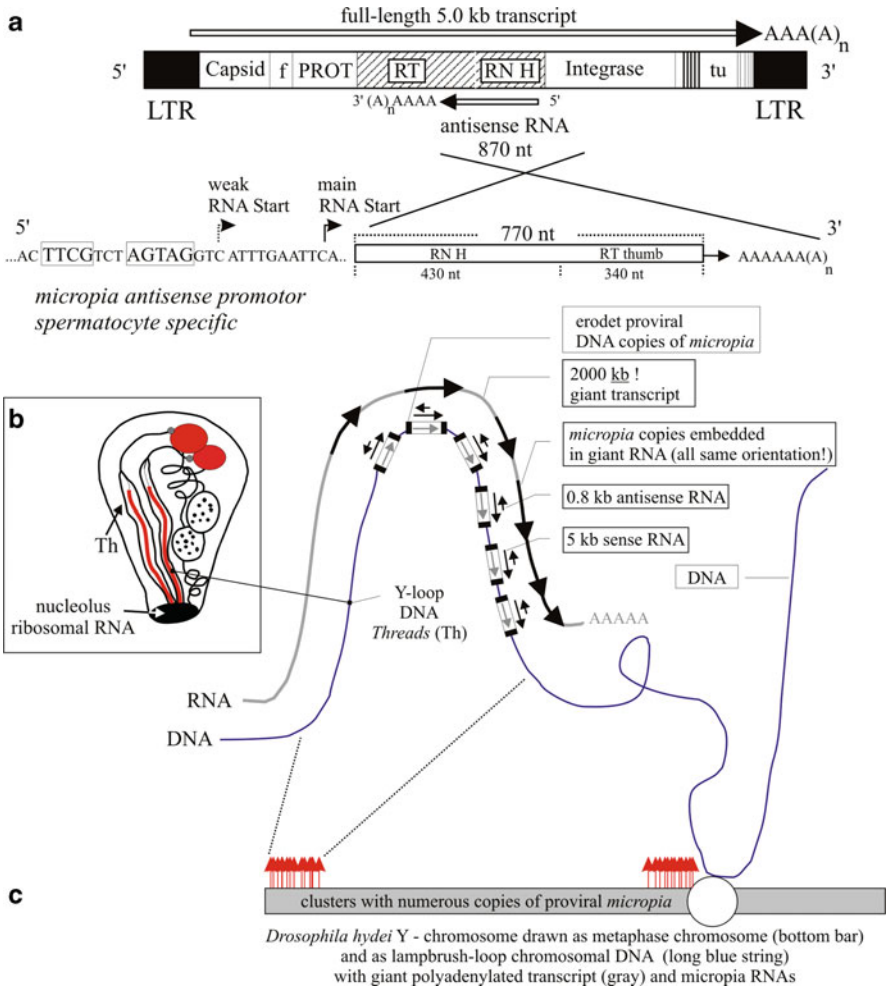

Fig. 11.2 Endogenous expression of the *micropia*-retrotransposon on Y-chromosome loops in *Drosophila hydei*. (**a**) *Micropia* encoded antisense RNA complementary to reverse transcriptase (RT) and RNase H. The antisense RNA is meiotic-germline specific in primary spermatocytes. (**b**) Schematic nucleus of primary spermatocyte with giant Y-loop transcripts. Th, giant RNP-transcripts termed *Threads*. (**c**) Expression and genomic distribution of *micropia* retrotransposon clusters on Y-chromosome loops pointed toward a molecular immune system (Huijser et al. 1988; Lankenau 1999; Lankenau et al. 1988, 1994). Transposon silencing pathways were subsequently confirmed with Argonaut subfamily Piwi proteins as the central players of such a molecular immune system controlling endogenous transposable elements of all shades (Aravin et al. 2007; Brennecke et al. 2007; Sarot et al. 2004). *Red filling* and *red arrows* indicate location of *micropia* clusters

known as RNA interference (RNAi) (Joshua-Tor and Hannon 2011). RNAi injection into living organisms as a knockdown tool (i.e., gene-silencing) for studying gene function has been used since then.

Andy Fire then showed experimentally that antisense RNA injected into the nematode *Caenorhabditis* interfered quantitatively best with complementary mRNA when injected as double-stranded RNA (Fire 2006; Fire et al. 1998). Similar evidence for a recent RNA world had accumulated from gene silencing phenomena in plants. Here, it was known that infection by viruses was combated by a mechanism that involved destruction of viral RNA. Earlier it had been shown that in plants genes could be switched off by copies of homologous genes in the same cell. These two observations were linked by infecting plant cells with a virus that carried a copy of an endogenous plant gene. This endogenous plant gene was subsequently silenced by the homologous transgene within the virus (reviewed in (Matzke and Matzke 2004). All the phenomena are now known to be mechanistically linked. Most importantly, a genetic screen for RNAi-resistant mutants in *Caenorhabditis* identified the gene *rde-1* (RNAi-defective-1) as being essential for knocking down gene function in response to exogenously introduced-double stranded RNA (Tabara et al. 1999). The *rde-1* gene turned out to be a homolog of the gene *ago1* in the plant *Arabidopsis* that had been noticed because of developmental mutant flower phenotypes reminiscent of the pelagic octopus *Argonauta argo* (Bohmert et al. 1998). Sequence alignments of Argonaut (Ago) proteins revealed a wide spread distribution of this family present in archaea (e.g., *Pyrococcus furiosus*), bacteria (e.g., *Aquifex aeolicus*) and eukaryotes, the latter split into three protein-clades (reviewed in Tolia and Joshua-Tor 2007). This actually opened a window into a distinct new world – a modern RNA world with reflections into the past. For their discovery of RNAi and gene silencing by double-stranded RNA, Andrew Fire and Craig Mello received the Nobel Prize in Physiology or Medicine in 2006. Argonaute proteins plus 23 nucleotides long dsRNA fragments called small interfering RNAs (siRNAs) are the signature components of an RNA-induced silencing complex called RISC (Song et al. 2004). **RISC** itself appears to be a programmable sequence-*homology search engine* that engages in various genomic targeting activities (Fig. 11.3). How does it work? To explain the collaborating key mechanisms of RNA silencing another key factor must be considered: Dicer. Dicer is an RNaseIII-like enzyme that recognizes and digests long dsRNAs. The genomes of dsRNA viruses are good substrates as they are double stranded from the beginning. Messenger RNAs or transcripts like the 770 nt *micropia* antisense RNA (Fig. 11.2a) naturally fold into low energy secondary structures giving rise to dsRNA stretches. In some organisms, ssRNAs are made dsRNAs through RNA-dependent RNA polymerase (RdRP) before further processing. Such long RNAs are the substrate of Dicer, chopping the RNA into double stranded fragments of 21–26 nt length. These small RNAs (siRNA) are then incorporated into multiprotein *silencing-effector complexes*. Using the sequence information of the dsRNA fragment, these complexes are then guided by some sort of homology search to complementary nucleic acid targets (Fig. 11.3). As the complex is not necessarily bound to other cell constituents, it is likely that it functions as a diffusible, trans-acting homology signal. The protein composition of the complex varies and depends on the nature of the target sequence. RISC is just one example of different effector-complexes, with different types of silencing possible (Matzke and Birchler

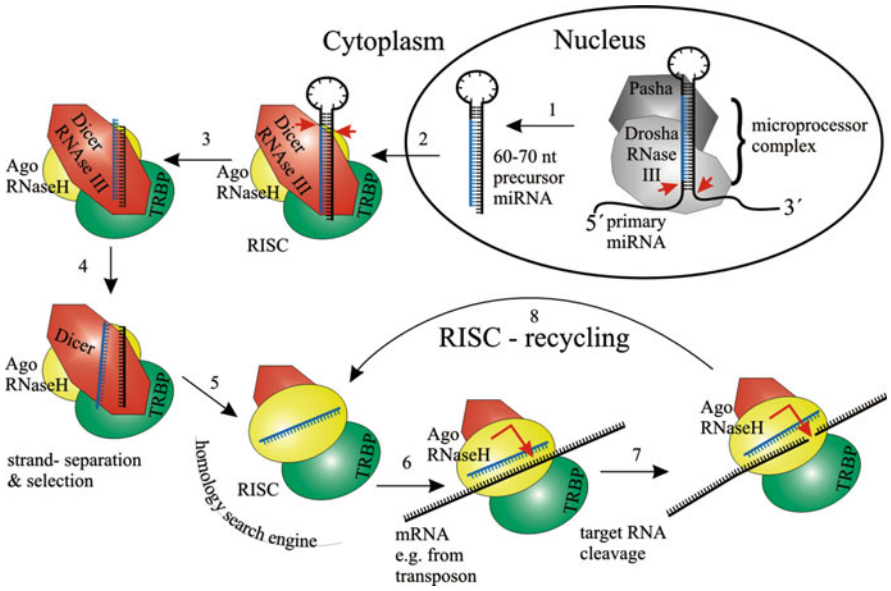

Fig. 11.3 The diffusible homology search engine and silencing-effector complex RISC, assembly and function (*Homo sapiens*). Modified after Gregory et al. (2005). Successive steps indicated by numbers 1 through 8. (*1*) The microprocessor protein complex comprised of the dsRNA binding domain (dsRBD) protein (Pasha = DGCR8) and the RNAse III protein Drosha recognizes a primary miRNA transcript. *Red arrows*, microprocessor cleaves specifically the base of stem loop RNAs. A precursor-micro RNA of ~60–70 nt is released with 2 nt single stranded 3′ overhang. (*2*) Transport through nuclear membrane to cytoplasm. Recognition by RISC including Dicer-RNaseIII, Ago-RNase H and double-stranded RNA binding protein TRBP. (*3*) Dicer cleaves ~22 nt by 2 nt staggered cuts with ~22 nt duplex miRNA remaining bound to the RISC RNP complex. (Note the similarity to the nuclease reactions of retroelement integrases & RNasesH). (*4*) Guide strand identification in dsRNA for homology search and strand separation. (*5*) Homology search engine, where the guide strand is kept bound within the RISC. (*6*) Target recognition. The guide miRNA fragment searches cytoplasmic mRNAs for homology and directs RISC to a complementary sequence. (*7*) Crooked *red arrow*, Ago-RNase H endonucleolytically cleaves the mRNA, destroying it partially for posttranscriptional silencing. (*8*) Cleaved mRNA is released. The RISC homology search engine can engage in a new cycle of mRNA hunting and destruction. According to Gregory et al. (2005), no energy is needed for any of the ssRNA release steps

2005). In the late 1980s and early to mid 1990s, my wife and I worked on the Y-chromosome lampbrush loop of *Drosophila hydei*. The Y-chromosome is heterochromatic in all tissues and developmental germline stages except during the prophase of meiosis in the male germline. During the primary spermatocyte stage, it forms so-called lampbrush loops (Y-loops) that express giant transcripts of a size of 2,000 kb (Hennig et al. 1989). In addition to the *micropia* antisense RNA (Fig. 11.2a) mentioned above, we discovered perplexing expression patterns of various transcripts cross hybridizing with single stranded probes detecting *micropia* sequence-similarity (Lankenau 1996). Figure 11.2b, c shows a graphic summary of the results. Based on early data from prokaryotes and eukaryotes, we

speculated that the giant transcripts and *micropia*-embedded transcripts might be subject to RNA degradation representing a kind of molecular immune system against transposable elements (Lankenau 1996). Unfortunately, these Y-loops and the antisense transcript were specific of *Drosophila hydei*, which is not a good genetic model system, such that we were unable to pursue further functional analysis. Today, we know that RNAi can mediate heterochromatin assembly in the presence of transposon clusters. Volpe and colleagues discovered the involvement of the RNAi machinery in histone methylation and heterochromatin formation at centromeres of fission yeast (Volpe et al. 2002). In yeast, tandem repeats or multiple copies of transposable elements in heterochromatin generate dsRNAs or ssRNAs that can be transcribed to dsRNA through the activity of RdRP. The dsRNA is then cleaved by Dicer producing small interfering RNAs (siRNAs). These siRNAs associate with histone methyltransferases (HMTs) that together represent a *silencing-effector complex* termed SHREC. In their function to guide the complex to a target RNA or DNA sequence, the siRNAs guide the HMTs to the centromeric chromatin to methylate histone H3 on lysine 9 (H3K9). The modified H3 is then bound by Swi6 or heterochromatin protein HP1, which also associate with methyltransferases, to maintain a silenced state of heterochromatin (for review, see Matzke and Birchler 2005). This RNAi silencing is mediated by SHREC which appears to regulate nucleosome positioning to assemble higher-order chromatin structures critical for heterochromatin functions (Sugiyama et al. 2007). This example demonstrates how deeply the modern RNA world is embedded in the physiology and metabolism (chromatin controls gene expression -> gene expression controls the quantity and content of enzymes etc., -> this controls the making of new metabolites) controlling and penetrating the endogenous integrity of all modern organisms. For example, genome integrity is also modified by RNAi-mediated pathways. In the ciliate *Tetrahymena thermophila* during the process of germline-soma differentiation, RNAi mediates DNA elimination of a chromosomal fragment called internal eliminated segment (IES) (Mochizuki and Gorovsky 2004). In other diploid species, RNAi dependent silencing of unpaired genomic regions takes place during meiosis (Shiu et al. 2001). In *Neurospora crassa,* this mechanism requires RdRP and an Argonaut like protein distinct from “normal” RNAi in this organism (for review, see Matzke and Birchler 2005). During meiosis in the filamentous fungus *Ascobolus immersus* DNA methylation can be transferred from one allele to another in a pattern typical of gene conversion, indicating the involvement of DNA-DNA pairing (Colot et al. 1996).

In summary, RNA plays a crucial role in epigenome regulation and epigenetic inheritance. Today, the prominent role for small RNAs, RNAi, and the multiprotein silencing-effector complexes as diffusible homology search engines in mediating sequence specific silencing at genome levels has been broadly established. The origin of these ubiquitous mechanisms, however, remains to be worked out as it may provide a window into an even more ancient RNA world directly linked to the geochemical origin of replicators, genes, and life itself. As a PhD student, I ran into this modern RNA world by discovering the *Drosophila* genus-specific *micropia* retrotransposon family and their perplexing transcripts (Lankenau et al.

1988). Knowing that the genomes of most eukaryotes consist of significant amounts of transposable elements and other non-coding sequences (Davidson 1986), where the human genome consists of 50% transposable elements and 98% of non-protein-coding DNA (Venter et al. 2001), the idea rose that *micropia* must be involved in intragenomic molecular immune responses. In addition, the unidirectional tandem repeat organization of *micropia* (Fig. 11.2c) also pointed to homologous recombination and other DNA repair mechanisms responsible for amplifying a transposon in a way that established the unidirectional tandem patterns (i.e., Synthesis Dependent Strand Annealing, SDSA (Fig. 11.8b) (Lankenau 2006; Paques et al. 1998)). In this context, I am thankful for a note by Matzke and Birchler (Matzke and Birchler 2005) recognizing the intermingled, ancient symbiotic/parasitic correlation of transposons where I cannot better do but to quote: "*Consistent with the role of RNAi in defense against invasive sequences (Matzke et al. 2000*; Vastenhouw and Plasterk 2004), *transposable elements and related repeats are preferred natural targets of the RNAi-mediated silencing pathways in the nucleus. Although often considered solely as molecular parasites, transposon sequences that function as foci for RNAi-based chromatin modifications benefit the host through these mechanisms by contributing to gene regulation and to chromosome structure and function*." Recent, more general theoretical work further underscores this statement (Branciamore et al. 2009) and its spirit leads us directly to address the question of the origin of transposons and to follow the Popperian deduction further top-down (Fig. 11.1).

Another modern, prominent example of the RNAi world relates to the human X-chromosome. One of the two human female X chromosomes is transcriptionally silenced in every cell by ncRNAs. X-inactivation is controlled by Xist and Tsix, two non-coding genes of antagonistic function. The Xist gene produces a 17-kb-long non-coding RNA that localizes along one of the two female X chromosomes and triggers chromosome-wide silencing X-inactivation by coating the inactive X. Tsix RNA is transcribed in antisense to Xist, and is critical for the labeling of the active X-chromosome through cis-expression of Xist RNA accumulation. It has been proposed that LINE-1 retrotransposons (L1) serve as DNA signals and mediate the spreading of the X inactivation signal along the chromosome. Apparently, a subset of L1 elements in the X chromosome is enriched that were active less than 100 Ma ago (Bailey et al. 2000; Lyon 2000). Thus, the X-inactivation center (Xic) represents an idiosyncratic RNA world of gene-dosage compensation that evolved about 50–200 Ma ago in Eutherian mammals (Chaumeil et al. 2006; Chow et al. 2005; Lee 2011; Navarro and Avner 2010; Ng et al. 2007). From the human perspective, it appears to be very old. However, in terms of geologic deep time, the X-inactivation center is only a relative recent evolutionary novelty. It may encompass primordial mechanisms but for a first anthropic top-down causality in the Popperian deductive cycle, it would be necessary to find evidence for an ancient RNA world. The X inactivation world is not yet far enough explored to tell us anything. What could a

relict from the primordial beginnings of an old RNA world within Xic or any other RNAi related process[7] look like? What shall we call young and what is old?

Let us briefly address the latter key question. Since a long time, we know about one component of the modern RNA world that is actually very old and dates back to an old if not very distant ancient RNA world – a true molecular living fossil that likely is conserved within all life for more than 3.5 Ga. Such a molecular living fossil in fact exists. It is a constituent of all life's genomes: the ribosome mentioned in the introduction. As long as none of the modern RNA world constituents – and with modern I mean younger than 3.5 Ga (Giga years = billion years) – can be causally connected to the times the very old ribosome emerged (~ $\geq$ 3.5 Ga) this modern transposon/RNAi RNA system world like the Xist/Tsix world, mentioned above, could in theory have evolved de novo from scratch. 4.3 Ga ago, the earth had cooled enough such that water had precipitated. We know this from zircon crystals whose age could be determined using radioisotopic methods. $ZrSiO_4$ only builds crystals in the presence of water. Zircon crystals from the Murchison district in West Australia were dated using the ^{18}O-isotope and provided evidence for liquid water at the Earth's surface 4.3 Ga ago. Around this time, the astro-geophysical parameters were mild enough such that primeval synthesis of life could have started. We then have to ask, what is ancient from the modern RNA world that could still exist and is detectable in modern cells and genomes *in addition* to the ribosome? Fortunately, there are numerous links that connect RNAi systems such as Xic (via LINE elements) to the primordial world. First, Argonaut proteins belong to an ancient family! They are related to the RNase H clade of enzymes typical of retroviruses and LTR-retrotransposons. At their core, Argonaut proteins have the typical RNase H fold and the two conserved aspartates invariably present on flanking β-strands of retroelements (Joshua-Tor and Hannon 2011). Above, we learned about RISC as an example of a *silencing-effector complex* and its role as a homology search engine. Among other functions, the most conserved function of Argonaut proteins is endonucleolytic cleavage of a target sequence (Song et al. 2004) similar to the activity of RNase H during reverse transcription-replication of retrotransposons, retroviruses, and other retroelements (Kohlstaedt et al. 1992). The Argonaut-RNase H activity creates a 5′ product 3′OH and a 3′ product carrying a 5′phosphate, where the target DNA strand is generally replaced by a guide RNA. The RNase H family is ubiquitously conserved in all domains of the tree of life consisting of well-characterized enzymes such as the integrases and transposases of DNA-transposons and retroelements (Nowotny 2009). While *Escherichia coli* RNase H1 catalyzes a single reaction resulting in substrate cleavage, integrases, and transposases catalyze two consecutive reactions such as during DNA transposon integration events resulting in strand transfer (Rio 2002). The nucleophile in these reactions is a water molecule or the 3′-OH of a nucleotide. Involved is a two-metal ion catalysis mechanism, with one metal activating the nucleophile and the second

[7]Ubiquity of proteins or genes or ncRNAs in all organisms means that they are likely of primordial universality.

stabilizing the intermediate (Nowotny 2009). Crystal structures of substrate-bound complexes of the transposon Tn5 transposase and of human and *Bacillus halodurans* RNase H1 revealed the involvement of two metal ions likewise (Lovell et al. 2002; Nowotny et al. 2005, 2007). So far, the participation of one metal ion has been confirmed for Argonaut protein structures as well as the two invariant aspartates and a conserved histidine critical for catalysis (Joshua-Tor and Hannon 2011). Also Piwi clade proteins possess conserved catalytic activity, and are supposedly active, based on their conserved mechanism of transposon recognition and silencing (Faehnle and Joshua-Tor 2007).

Argonaut-RNase H active proteins also participate in processing of human immuno deficiency virus (HIV) transcripts (Fig. 11.3). HIV is the most well studied retrovirus and we can learn a lot from its cellular functions and its association with ancient RNA world processes. A *silencing-effector complex* containing Dicer, Argonaut and HIV-tar-RNA binding protein (TRBP) is able to guide-strand loading and multiple rounds of target cleavage, stimulated by nucleotides (Gregory et al. 2005). This example together with Julius Brenneke's ping-pong model (Brennecke et al. 2007), where transposon transcript abundance is kept under control involving Ago3 complexes, piwi-interacting RNA (piRNA) cluster transcripts, and transcripts of active transposons, may turn out as the explanation for what is going on with the *micropia* Y-loop transcripts in primary spermatocytes (Fig. 11.2), and may further give the explanation for why and how these unusual Y-loop structures evolved after all. Probably their existence is a sort of exaptive molecular drive involving the RISC-machine and invasion-control of the *micropia* family.

RNA world derived defense strategies against mobile genetic elements have now also been characterized in prokaryotes and compared to RNAi in eukaryotes (Jore et al. 2011; Shah and Garrett 2011).

11.2.2 Ribonucleases: RNaseH and Integrase Meet Reverse Transcriptase

The evidence is overwhelming for a contemporary RNA world being a central element of living cells and cell differentiation. The major signatures that underscore the importance of RNA in modern cells are: (1) transcription into mRNAs from DNA genes; (2) protein translation from mRNAs mediated by tRNAs in a ribosome with 23S or 28S rRNA as catalytic ribozyme; and (3) we now know that gene and chromosome regulatory control mechanisms involve a full panoply of RNA associated activities that even tightly pervade metazoan germline processes (Lankenau 2007, Box 1). Now, why do we think RNA is relevant for the origin of life? Would protein or DNA not suffice? Did DNA or proteins precede RNA in the primal synthesis? The answer seems "no" for the following, selected, rational arguments:

Anthropic arguments:

1. RNA is both a biocatalyst and an informational molecule. Thus both, genotype and phenotype are a feature of RNA. Protein on the other hand has extremely limited ability to transmit information (as in prions).
2. DNA has an excellent ability to transmit information but it lacks the biocatalyst property in modern cells.
3. RNA is capable of replicating itself and can perform the chemistry needed for RNA replication (Cech 1986).
4. DNA replication in modern cells always needs the 3′OH end of an initial, independently synthesized RNA primer.
5. The ribosome ubiquitously uses the catalytic activity of RNA to perform peptide synthesis.
6. RNA preceded DNA, because in modern organisms the biosynthesis of deoxyribonucleotides is by reduction of previously synthesized ribonucleotides. Two additional enzymes are needed to make DNA from ribonucleotide precursors: ribonucleotide reductase and thymidylate synthase.

Bottom up arguments:

7. Prebiotic synthesis of ribose appears simpler by alkaline aldolcondensation than to think of the synthesis of deoxyribose from scratch.
8. The nucleophilicity of the 2′, 3′OH-groups is higher than of the 3′OH group of deoxyribose (Lohrmann and Orgel 1977).
9. It is more parsimonious to think of a single type molecule (i.e., RNA) replicating itself than to imagine that two different molecules (e.g., random peptide plus nucleic acid) were synthesized by random chemical reactions at the same time in the same place – and that repeatedly for only by redoing the trick over and over again a new individual entity would emerge.[8]

In summation of our current deductive top down analysis, we can so far ask if there is a preliminary unifying theme from the modern RNA world that might unite all three domains of life (i.e., bacteria, archaea, eukaryota) with LUCA at its root? As RNA polymerization is the basic requirement for any cellular RNA world, it is not surprising that two proteins are ubiquitous in all living organisms: DNA-directed RNA polymerase, subunits α, β, β′ responsible for the transcription of genes and transcription antiterminator NusG (Charlebois and Doolittle 2004; Koonin 2003) (not further elaborated here). The second conserved topic is that of two key players of RISC, i.e., two families of RNases (Fig. 11.3).

The first is represented by Drosha and Dicer which belong to the RNase III enzyme family. RNase III enzymes specifically bind to and cleave dsRNA. There

[8]This statement must not be confused with the assumption that the first RNA replicators emerged in sterile environments. There always must have been organic and anorganic compounds coexisting with any informational RNA molecule. Peptide replicators are possible (Ashkenasy et al. 2004). They possibly have coexisted and facilitated RNA relicators from begin on. For an emergence scenario encompassing systemic properties, see Dyson (1999); Kauffman (1993).

are three classes of RNaseIII: Class 1 process precursors of ribosomal RNA and in fungi small nuclear RNAs (snRNAs). Class 2 comprises Drosha with rRNA and miRNA processing functions. Class 3 includes the Dicer family involved in RNAi. Especially interesting is the processing aspect of 12S pre-rRNA pointing toward an ancient pathway (Fukuda et al. 2007; Redko et al. 2008).

The second ubiquitous RNase enzyme activity in the modern RISC RNA world however is that of the Argonaut proteins, which are related to the RNase H clade. As mentioned above, Argonaut proteins have the typical RNase H fold and the two conserved aspartates invariably present on flanking β-strands like RNase H encoded by LTR retrotransposons (Joshua-Tor and Hannon 2011). Even more prominent is the structural similarity to integrases and DNA transposon transposases, and the Double Holliday Junction processing protein RuvC, the latter providing a link to DNA repair (Yang and Steitz 1995). However, there appears yet another even more crucial link to that stage of LUCA, where RNA started to be transcribed into DNA. Modern retrotransposons and retroviruses represent this transition of an RNA world into the DNA world. Retroviral RNaseH activity in retroviral replication is tightly connected to reverse transcriptase (RT) activity (Kohlstaedt et al. 1992). Let the HIV RT protein serve here as the role model for other retroviruses and retrotransposons. RT is carved out from the pol gene product as a 66 kD polypeptide. A 66 kD peptide includes both RT and RNaseH functions. Two 66 kD monomers aggregate to form a dimer, where both monomers have slightly different conformations. One is susceptible to proteolytic digestion by the retroviral protease such that a RT/RNaseH heterodimer is released with a 66 kD unit containing RT/RNaseH and a 51 kD unit with a second RT subunit (Arnold et al. 1992; Huang et al. 1998). The p66-p51 heterodimer has one RT polymerase active site, one RNaseH active site, and one tRNA binding site. This intricate interrelationship between RT and RNaseH reflects an intriguing replication mechanism that likely dates back to the times of LUCAS. The most interesting part linking retroids[9] to RISC are the intramolecular and intermolecular strand exchange reactions during reverse transcription (details see below). As the RNaseH is always an integral part of homology-dependent strand transfer reactions, we may speculate that the RISC homology search machine including Argonaut-RNaseH proteins was evolutionarily derived directly from the RT/RNaseH driven mechanism (Fig. 11.5).

[9]All elements and sequences encoding a reverse transcriptase: retroviruses, retrotransposons, group II introns, LINEs, plus-strand RNA viruses, pararetroviruses, retrons.

11.2.3 Reverse Transcription: An Ancient Mechanism of Replication

Retrotransposons and retroviruses are rather autonomous folks that personify idiosyncratic toolboxes of replication. How do retroids replicate? The proviral structure of *micropia* representing a typical modern RNA world LTR-retrotransposon is shown in Fig. 11.4 (Lankenau et al. 1988). Initially, an RNA polymerase, guided by upstream enhancer bound transcription factors, binds to a promoter such as the classical TATA box. Without the need for the 3′OH end of a primer, it starts transcription about 28 nucleotides upstream. Typically, the retrotransposon transcript terminates after about 5 kb at a AA(G)TAAA termination signal – still inside the boundaries of the proviral DNA sequence entity. As transcription started from a transposon internal promoter, and terminated at the stop signal well before the 5′end of the transposon this transcript by no means represents the full-length boundaries of the transposable element. And yet, this transcript – an incomplete copy of the transposon – is the only source to restore a complete, new copy of the proviral DNA fragment. How is that possible? How can an incomplete RNA fragment restore a complete DNA provirus?

The trick is that the mRNA[10] encompasses all sequence information necessary to restore the full length proviral DNA entity which sooner or later may integrate at a different site in the "host" genome. It also contains a sequence fragment called R duplicated at its opposite ends. Retroviral replication is complex. Figure 11.5 shows the mechanism of reverse transcription as combined from several sources (for review, see Telesnitsky and Goff (1997), Varmus and Brown (1989), Voytas and Boeke (2002)). Both Figs. 11.4 and 11.5 represent structural and mechanistic details of the replication steps mediated by RT/RNaseH and, because these two proteins can be seen as a homology search engine like the RISC complex, they represent a key to the modern RNA world as well as to an ancient RNA world. Figure 11.5 is simplified. In vitro experiments suggest that an intact virion core is required for efficient elongation and cis or trans strand transfers. There is evidence of limited DNA synthesis in HIV-1 virions prior to the entry of the virion into a target cell. But reverse transcription generally appears to be activated by entry of the viral core into the cytoplasm of the target cell (Telesnitsky and Goff 1997). As endogenous LTR retrotransposons or endogenous retroviruses such as the insect *gypsy* retrotransposon form virions or virus like particles (VLPs), there are many variations in detail in the timing and location of reverse transcription processes (Adams et al. 1987; Flavell and Ish-Horowicz 1983; Garfinkel et al. 1985; Kikuchi et al. 1986; Mellor et al. 1985; Shiba and Saigo 1983; Song et al. 1994).

[10]The long mRNAs of retrotransposons are better called full length RNAs. There may be epigenetic differences between the two.

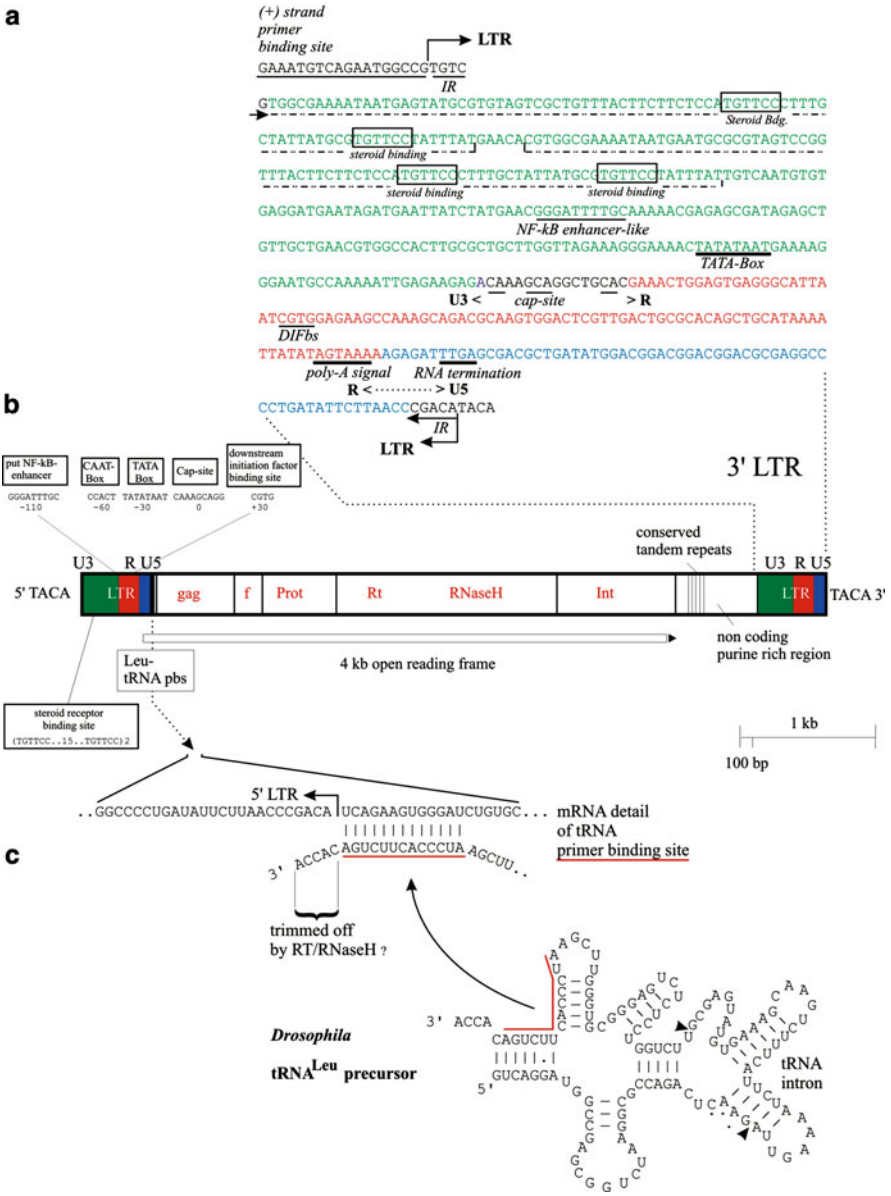

Fig. 11.4 Structure and priming model for *micropia* Dm11 of *Drosophila melanogaster*. (**a**) Nucleotide sequence of 3′ LTR. The putative order of U3, R, and U5 sequences are colored *green*, *red,* and *blue* respectively. Putative upstream elements relative to TATA box are indicated. (**b**) Genome organization with encoded genes. LTR regions colored as in A. (**c**) tRNA priming model, data based on (Lankenau et al. 1988)

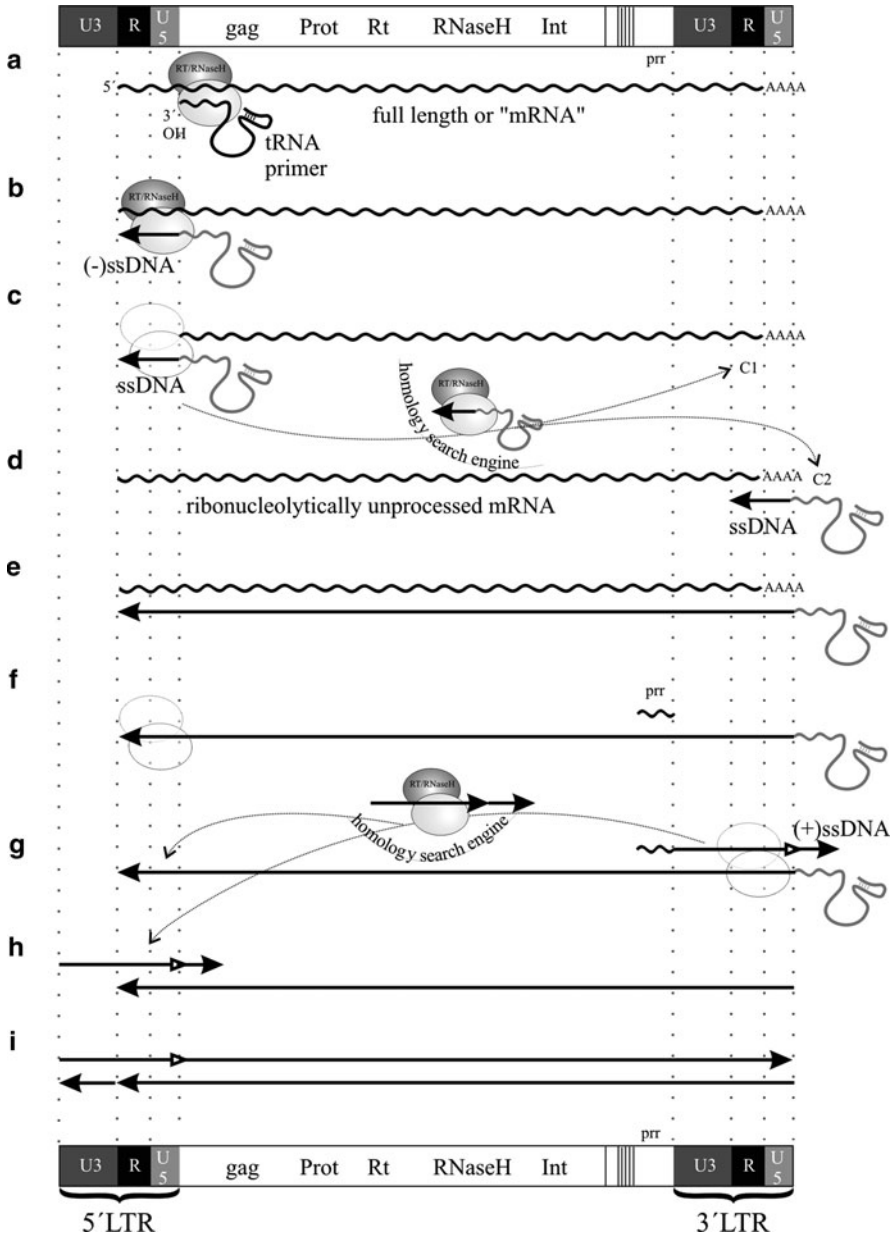

Fig. 11.5 Mechanism of reverse transcription and synthesis of proviral DNA. The processes are drawn as linear cis and trans primer strand-transfer reactions; circular cis-reactions are possible as well. The *top* and *bottom* drawings represent the proviral, genomic DNA-integrates of a LTR-retrotransposon with gag, protease, RT, RNaseH, and integrase genes – before and after a reverse transcription driven replication cycle. Between the *top* and *bottom* proviral schemes there are numerous cell physiological steps not shown, including formation/unpacking of virus like particles (VLPs or virions) and the cytoplasmic processes where replication takes place. (**a, b**) Two

Once a full length "mRNA" has been produced,[11] a host-endogenous tRNA is binding by complementary base pairing to the 5′ tRNA primer binding site (Fig. 11.5a). This initiates the DNA polymerization steps of RT and the RNA degradation steps committed by RNaseH reminding us of the RISC-RNAi shredder mill. The most significant aspect of the replication of a complete copy of a provirus from an incomplete retroid mRNA however is the use of redundancies and unannealing, strand transfer and annealing reactions. In the modern RNAi world, this is performed by the RISC "search engine" (Fig. 11.3), for retroid replication it is performed by RT/RNaseH (Fig. 11.5c, d and g, h). Interestingly, the modern DNA world encompasses a similar genome wide homology search program as part of homologous recombinational DNA repair-reactions using the synthesis dependent strand annealing pathway (Engels et al. 1994; Lankenau 1995). The SDSA DNA repair homology search of modern organisms however faces the far more byzantine world of modern chromatin factors. SDSA will only sporadically be mentioned here (see section "Group I Introns" and Fig. 11.8b) (Lankenau et al. 1996, 2000). The HIV RT/RNaseH search engine serves as a prime model system here as well. Recently, with the aid of smFRET techniques it was possible to understand the fundamental molecular-dynamics of the homology-search at single-molecule resolution (Abbondanzieri et al. 2008; Liu et al. 2008, 2010; Tinoco et al. 2011). This work shows that on a DNA template, RT/RNaseH binds to the template with a DNA or RNA primer in two opposite orientations. Either the DNA polymerase domain or the RNaseH domain are located close to the 3′end of the primer. This is what would be expected from the activity of RT/RNaseH functioning in

Fig. 11.5 (continued) molecules of full length, poly adenylated mRNA are present within or released from the VLP or virion. The 3′OH end of a host tRNA anneals to one mRNA molecule and primes minus-strand DNA synthesis. RT synthesized until the 5′ end of the full length transcript generating a short DNA fragment on the order of 100–150 nucleotides length. This fragment is termed the minus strand strong stop DNA ((−)ssDNA). (**c**) Accompanying DNA synthesis the RNaseH activity of RT/RNaseH degrades the mRNA fragment already copied into ssDNA. (**c**, **d**) The ssDNA contains the repeated (*R*) sequence that is occurring at both 5′ and 3′ ends of the mRNA. The first strand transfer (here hypothetically shown as mediated by RT/RNaseH working as a homology search engine) leads to the annealing of (−)ssDNA with *R* either on the same mRNA molecule in cis or on the second mRNA molecule in trans. (**d, e**) RT/RNaseH synthesize the full mRNA and simultaneously degrade all mRNA except a short purine rich fragment (prr) (**f**). (**g**) The 3′OH end of the prr RNA primer serves RT/RNaseH to start plus strand DNA synthesis resulting in plus strand strong stop DNA ((+)ssDNA). In a second homology search process the (+)ssDNA fragment translocates and anneals to the opposite DNA end on redundant LTR sequences (**h**). (*i*) Plus strand synthesis completes production of a full double-stranded LTR retrotransposon ready to be integrated into the genome with the aid of the self-encoded integrase. Abbreviations: ssDNA, minus- and plus-strand strong stop DNA; prr, purine rich region; cDNA copy DNA; *Open arrows* and stippled lines: Intra- or intermolecular translocation of complementary cDNA. (Diagram based on: (Peliska and Benkovic 1992; Telesnitsky and Goff 1997; Temin 1993; Varmus and Brown 1989; Voytas and Boeke 2002))

[11]The many splice variants and RNA editing is not further elaborated here.

DNA-directed DNA synthesis (e.g., Fig. 11.5h, i) or DNA directed RNA hydrolysis (Fig. 11.5e, f). On chimeric DNA/RNA primers, RT was observed to occupy both the DNA polymerase competent and RNaseH-competent orientations. The measured rate of primer extension correlated with the fraction of time for which the RT enzyme bound in the polymerase competent orientation. Obviously, RT/RNaseH distinguishes between different substrates. The HIV genome is ~10 kb long. Therefore, RT, with its low processivity in DNA synthesis (only a few to a few hundred nucleotides on a ~10 kb HIV genome) must accomplish location to its target site very efficiently using a highly capable searching mechanism crucial for RT function. With smFRET, it was demonstrated that RT/RNaseH slides thermally driven between two ends of an experimental substrate even if longer than ~20 nt. Once RT/RNaseH snapped into place at the catalytic target site with atomic precision, it can flip its binding orientation to orient the correct functional domain (DNA polymerization or RNA hydrolysis) close to the target substrate. Thus, the homology search engine combines trembling sliding motions and flipping (Abbondanzieri et al. 2008; Liu et al. 2008, 2010). This key mechanism likely helps to explain the differences of cis- versus trans- genome wide homology searches accompanying recombinational SDSA repair (Engels et al. 1994), but as said before, homology search engine processes and the byzantine organization of chromatin must be viewed on an equal footing (Lankenau 1995; Lankenau et al. 1996).

11.2.4 The Retroelement Ancestor Hypothesis

11.2.4.1 "Linné's Revenge" – or the Attempt to Classify Transposable Elements

What is Linné's Revenge? Attempting to classify transposable elements of all shades confronts us with similar problems today as Carolus Linnaeus was confronted with when writing his famous Systema Naturae of 1758. Even though Linné was not aware of our current understanding and the logics of phylogenetic systematics, his invention of the binomial nomenclature of species nevertheless reflected most aspects of today's accepted tree of life. While it is the goal of modern systematics to define strictly circumscribed monophyletic groups (Hennig 1950, 1966), Linné, unknowingly, established polyphyletic and paraphyletic groups that did not necessarily reflect monophyletic common decent. Today, the same holds true for molecular entities. For example, standard classifications of viruses into groups such as Herpes, HepaDNA, Adeno-, Retro-, Onco-, Lentiviridae, etc. do not reflect phylogenetic relationships but rather represent a grouping based on chemical, physical, physiological, or medical similarities. Such classifications are termed paraphyletic and polyphyletic of typological classification systems (Hennig 1966; Lankenau et al. 1988). The problem grows when approaching more and more ancient taxa. Popper's deductive cycle in Fig. 11.1a circles around the phylogenetic tree of life reflecting many horizontal gene exchanges between different clade-

branches, especially at the base near LUCA (Doolittle 1999, 2000). With regard to retroelements, it has become clear that between classifications of transposable elements based on the RNaseH or on the RT or on capsid proteins or on the integrase or on protease, there are large inconsistencies (Capy et al. 1998). Here, I have no space to elaborate on the primordial phylogenetic processes and I like to refer to the excellent analyses of Eugene Koonin and colleagues (Jordan et al. 2005; Koonin 2003, 2006a, 2006b, 2007, 2009; Koonin et al. 1980, 2000, 2006; Koonin and Martin 2005; Koonin and Novozhilov 2009; Koonin and Wolf 2008; Leipe et al. 1999). Thus, because the ancient "evolutionary temperature" (Woese 1998), as reflected by frequent horizontal gene exchanges between replicating entities, was very high in the early beginnings, any attempt to root the trunks of the three domains of life, i.e., bacteria, archaea, and eukaryota, appears to be futile. This matter becomes even more tricky if we accept the promiscuity of many proteins acting in historically totally unrelated biological functions and phenomena (Khersonsky et al. 2006). Stephen Jay Gould coined the concise term *exaptation*, that is, recruitment of an existing structure or function for a new function by unrelated selection forces (Gould 2002). The transposon/virus-like stage in life's early evolution belongs to the same kind of solutions and might be the most plausible if not the only way to avoid what Koonin termed "irreducible complexity trap" associated with the origin of cellular organization itself and subsequently of primordial entities representing LUCAS (Koonin 2009). To this end, let us get a feeling for the RNA world of retroids and other transposable elements. Many phylogenetic trees have been published on retroid and DNA transposon clades. Neither here can I elaborate fully on this issue and I like to refer to some excellent books for reference (Berg and Howe 1989; Capy et al. 1998; Coffin et al. 1997; Cooper et al. 1995; Craig et al. 2002; Fedoroff and Botstein 1992; Sherratt 1995; Skalka and Goff 1993). Figure 11.6 shows a phylogenetic tree of retroids representing all domains of life based on their RT amino acid sequences. The tree is outlined at two scales. It reflects the pseudo-phylogenetic relationship between RT of the major groups of retroids. It also represents an analysis of the sequence distance of the RT of the primordial *Penelope* retrotransposon to all major retroids as an outgroup. *Penelope* represents an average similarity distance to RT of all other retroids of about 14%. The analysis indicates that even though all RTs may share a common ancestor – probably at the time of LUCAS – the evolutionary change of sequences was so significant and sequence stability was compromised by horizontal transfer and xenologous recombination[12] (because evolutionary temperature was hot) (McClure 1991, 1993), that we cannot really identify a defined common ancestor among retroid elements. LUCAS therefore is more likely the stem group of the three domains of life with no need to identify an individual ancestor. With this insight, we are getting close to the golden spike of Fig. 11.1 in the top-down approach. But before jumping to the bottom-up path, one of the major retroids

[12]Replacement of a resident gene by a homologous foreign gene.

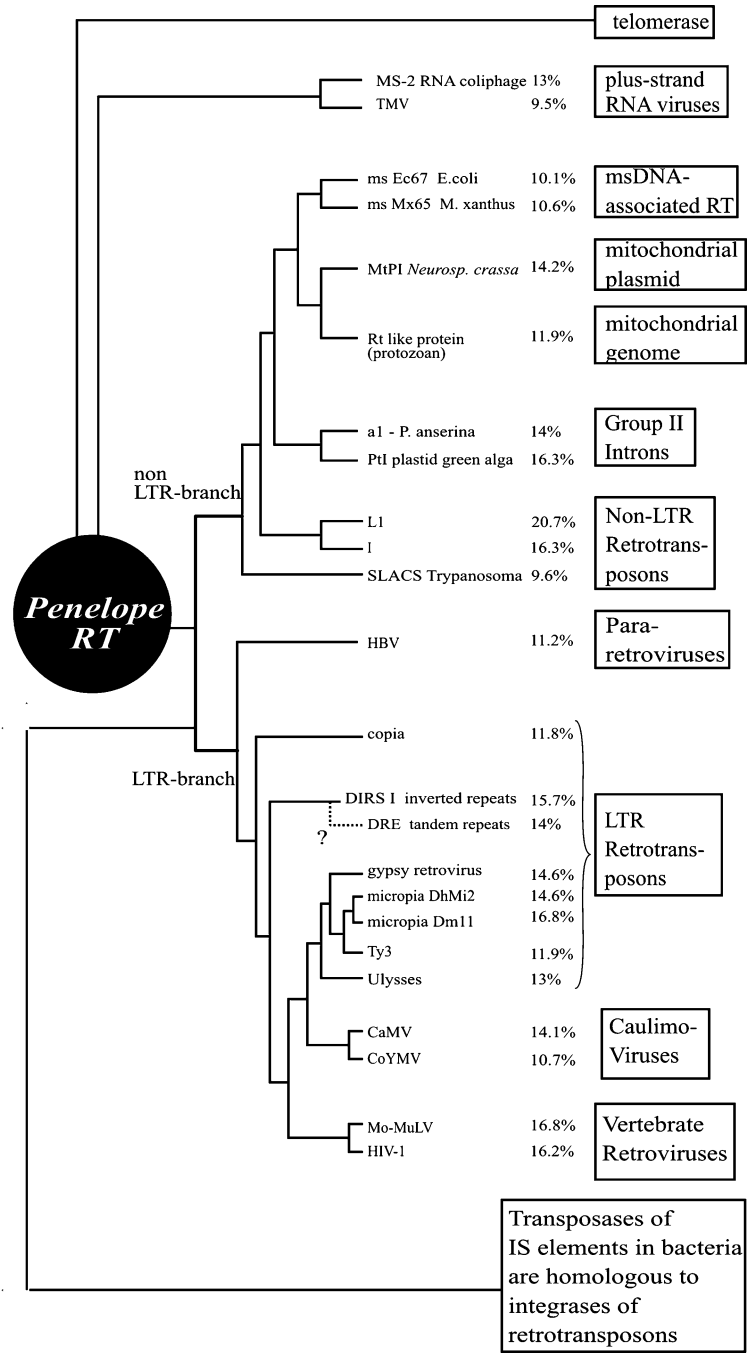

Fig. 11.6 Sequence relationship of RTs et al. between retroid elements. The pseudo-phylogenetic tree is based on the analysis of Xiong and Eickbush and own data. The tree uses the RT of the

seems to give us a lead anyway hinting toward mechanisms that circumscribe the molecular conditions at the origin of modern RNA replicators (retroids) and subsequent DNA genes. *Penelope* already is a very unusual retroid element with links to all retroid groups, but the Group II introns possess ribozyme catalytic properties that strongly help understanding the molecular events at the LUCAS stage.

11.2.4.2 Mobile Introns

For many years, the established term "gene" was dominated by Beadle's and Tatum's "*one gene-one enzyme*" concept (Beadle 1958). The discovery of split genes, i.e., genes that contain introns being removed from primary gene transcripts, then came as a surprise (Sharp 1985). It was even more surprising that RNA transcripts triggered their own excision of introns. Now, the splicing mechanism points toward an ancient origin deep within the realm of LUCAS. Introns are ribozymes that splice by three fundamentally different pathways. Group I introns splice by the guanosine-initiated pathway. Group II introns and the related group III introns splice by the lariat pathway. Introns found in rRNA and tRNA genes of *Archaea* use the nuclease-ligase pathway (Belfort et al. 2002).

Group I Introns

Group I introns are common in *eukaryota* and *bacteria* but do not occur in *archaea*. Intron mobility refers to two mechanistic types of mobility. The first is intron homing that actually is a DNA associated gene conversion process making use of a fundamental recombinational double strand break (DSB) repair reaction called synthesis dependent strand annealing (SDSA) (Gloor and Lankenau 1998; Haber 2008; Lankenau 1995, 2006; Lankenau and Gloor 1998; Nassif et al. 1994). A homing endonuclease is encoded within the group I intron of a host gene. After transcription of that gene with the intron still present in the primary RNA transcript, the 3′-OH group of an exogenous guanosine cofactor triggers a cascade of transesterification reactions splicing out the intron and self-ligating it to a full-length intron circular RNA molecule (Nielsen et al. 2003). Homing endonucleases like the *I-CreI* homing mega-endonuclease of *Chlamydomonas* are often used in

Fig. 11.6 (continued) ancient *Penelope* retrotransposon of *Drosophila virilis* as an outgroup for comparison (Evgen'ev et al. 1997). Amino acid sequence identities between the RT of *Penelope* and other major retroids are shown. Sequence similarities of the putative *Penelope* RT to all other known categories of RT sequences is, on an average, about 14%. The alignment of RT sequences therefore is based upon groups of conserved amino acid residues that can be identified in published RT-like sequences available (Xiong and Eickbush 1988, 1990). Alignments were carried out with the program MultAlin (Corpet 1988). The groups of conserved amino acid sequences are as defined (Xiong and Eickbush 1990). Sequences of *Ulysses*, and *micropia* were taken from published data (Evgen'ev et al. 1992; Huijser et al. 1988; Lankenau et al. 1988)

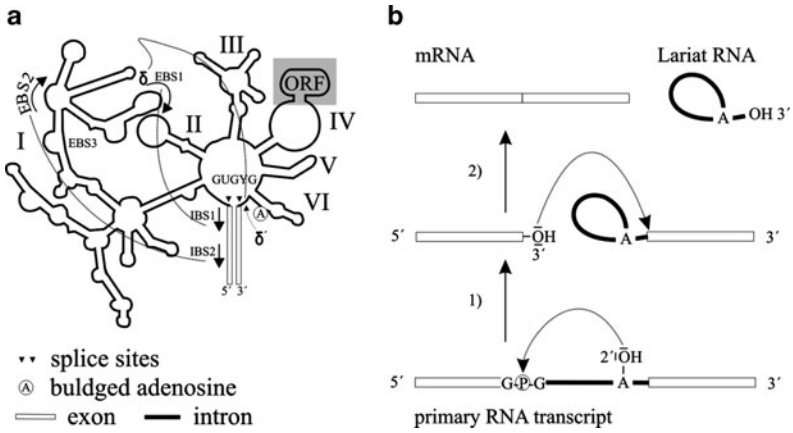

Fig. 11.7 Structure of group II intron RNA and the splicing mechanism. (**a**) Secondary structure of a group II intron (Perlman and Podar 1996; Schmidt et al. 1996). *EBS* exon binding site, *IBS* intron binding site; *dotted curved lines* link complementary binding sites. (**b**) Transesterification and self splicing mechanism producing lariat RNA. See Fig. 11.8c for group II intron retro homing and SDSA of the lariat molecule

experimental gene targeting systems (Lankenau 2006; Lankenau et al. 2003). The circular RNA molecule can be translated into the mega nuclease (Perriman and Ares 1998). The endonuclease then specifically produces a staggered double strand break within an intronless copy of the host gene. Using the intact, complementary host gene sequences as template, the DSB then is repaired via SDSA gene conversion with donor DNA that still contains the original group I intron DNA (Belfort et al. 2002). The group I introns therefore represent the very dynamic phase of the LUCAS employing recombinational mechanisms that mediate between the ancient RNA world and the modern DNA world.

Group II Introns

The second mechanistic type of intron mobility is represented by group II introns. They occur in 25% of the genomes of eubacteria, in mitochondrial and chloroplast genomes of fungi and plants but not in the nuclear genomes of eukaryotes nor in *archaea*. Like group I introns they are ribozymes as well. However, group II introns do not self-splice without co-factors such as intron-encoded and/or host-encoded splicing factors. These include reverse transcriptase, maturase,[13] RNaseH, and endonuclease. The open reading frame (ORF) of these genes are always inserted in domain IV of the group II intron (Fig. 11.7a). Sometimes the ORF extends 5′ to

[13]Many self-splicing introns code for matureases that help with the splicing process, generally only the splicing of the intron that encodes it.

form a continuous ORF with the upstream exon. The length of the group II introns are as long as 3,000 nucleotides.

The splicing mechanism is shown in Fig. 11.7b. Like group I introns, group II introns possess no conserved sequences at base level. Conserved sequences occur at the intron boundaries, with GUGYG and AY[14] as consensus sequence at the 5′ and 3′ends respectively. Domain double helix VI contains a buldged adenosine ribonucleotide which defines the branch site of the lariat structure. The free electron pair of the 2′-OH of the buldged A initiates a nucleophilic attack on the 5′splice site, producing a 2′–5′ linkage and the lariat RNA structure still remaining attached to the 3′exon. The free pair of electrons of a 3′-OH group at the 3′end of the free 5′exon attacks the 3′splice site resulting in exon ligation and lariat RNA release. The exact recognition of the splice sites is mediated by defined complementary base pairing interactions between intron binding sites (IBSs) and exon binding sites (EBSs) and between δ and δ' (Fig. 11.7a).

Figure 11.6 shows that group II intron RT is closely related to non-LTR retrotransposons – of which one representative, L1, makes up 5% of the human genome. The mechanism of transposition of LINEs within DNA genomes is related to that of group II intron retro homing (see color plate 47 in Craig et al. 2002; Belfort et al. 2002; Luan et al. 1993). The mechanism of retrohoming involves retrotransposition of the intron lariat RNA into genomic DNA. Details of the mechanism were explored in yeast and in bacteria (Belfort et al. 2002). The protein encoded within the ORF (Fig. 11.7a) is translated after transcription of the pre-mRNA and before splicing sets in. It then binds to the intron and induces RNA splicing with the aid of its maturase activity. The active RNP-lariat complex is in fact a homology search engine like the RISC complex in RNAi mediated pathways (Fig. 11.3) and in the RT/RNaseH strand transfer reactions during retroviral replication (Fig. 11.5). The lariat-RNA contains the EBS1 and EBS2 sequences of 14–16 nucleotide length. These, together with the active RNP complex perform the homology search and recognize the complementary IBS1 and IBS2 sequences in the DNA target site (Guo et al. 1997; Matsuura et al. 1997; Mohr et al. 2000; Saldanha et al. 1999). Other nucleotides in the homing site are recognized by the intron encoded protein in a restriction enzyme like fashion (Yang et al. 1998). The first cut is made by the lariat RNA reverse splicing into the sense DNA strand (Fig. 11.8a). The endonuclease domain of the protein cleaves the complementary DNA strand creating a staggered DSB. Figure 11.8b, c shows a comparison between a fundamental mechanism of recombinational DSB repair called SDSA and intron retrohoming.

From Fig. 11.6, it appears that the RT of group II introns is not the most ancient of the RTs of other retroids. However, like the multicellular group of algae, *Volvocinae*, which serves as model of a line of evolution toward multicellularity but in fact is too young to be the real ancestor of metazoans (Lankenau 2007),

[14]Y designates pyrimidine.

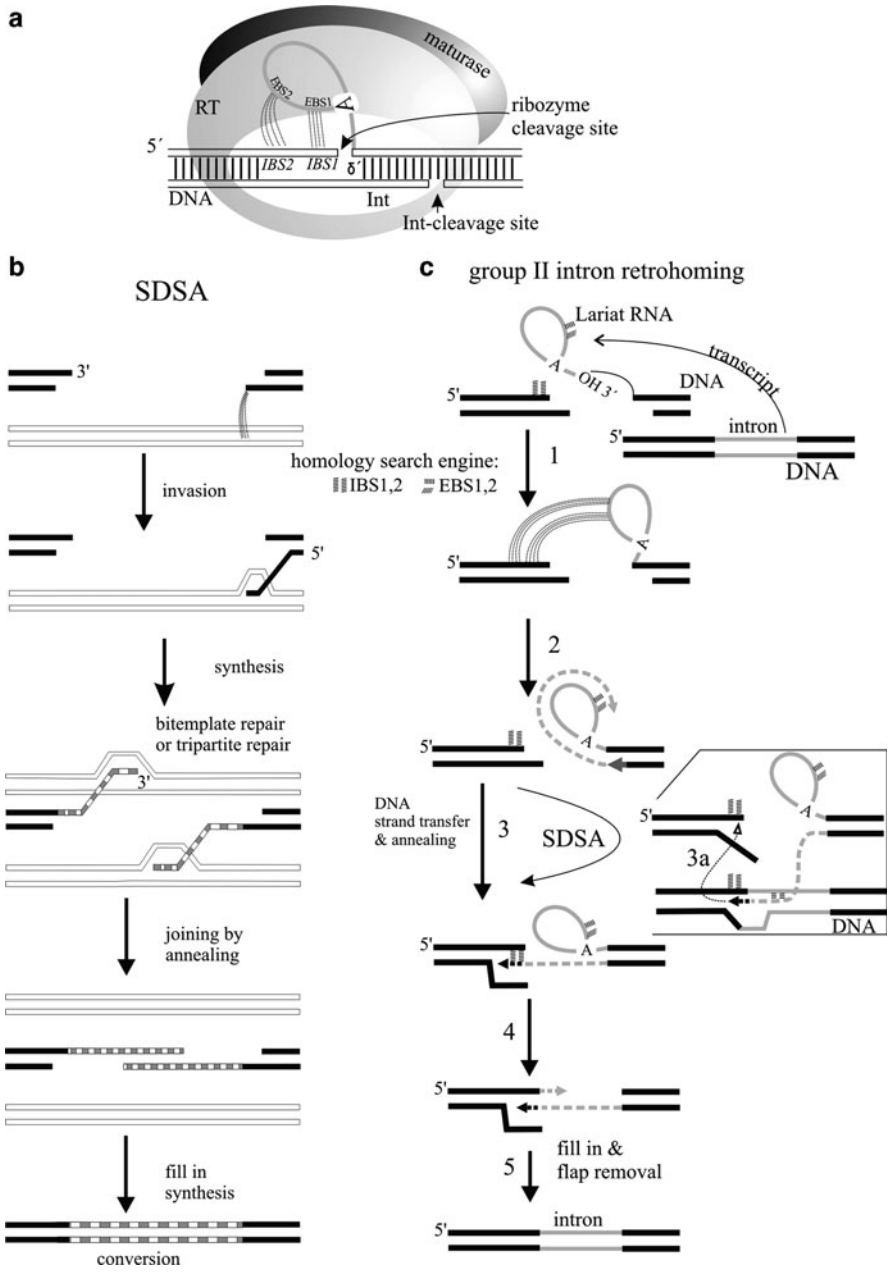

Fig. 11.8 SDSA and group II intron retrohoming. (**a**) Landed homology search engine RNPC initiating group II intron retrohoming on DNA target. The ovals symbolize the intron-domain IV encoded protein consisting of the following catalytic domains: *RT*, integrase (=endonuclease), maturase, DNA-binding. Integrase is an Mg^{2+} dependent endonuclease. Stippled rays indicate Watson-Crick base pairing interactions between intron and exon binding sites 1 and 2. Interactions

within the framework of our deductive cycle group II introns serve as a good model for an ancestor of the eukaryotic spliceosomal introns as well as the eukaryote non-LTR retrotransposons (i.e., LINEs) (Lambowitz and Zimmerly 2011). The relationship between group II introns and non-LTR retrotransposons is evident from similarities in DNA sequence of their RTs (Xiong and Eickbush 1988, 1990). The target-primed reverse transcription mechanism (TPRT) used by non-LTR retrotransposons to integrate in DNA is very similar to the retrohoming mechanism of group II introns as well (Fig. 11.8) (Luan et al. 1993). Based on phylogenetic studies, the LTR-retrotransposons and retroviruses then subsequently evolved from non-LTR retroelements. Group II introns are rare in archaea and are probably derived from horizontal events. The assumption that group II introns originated in bacteria of the archaic LUCA age as retroelements is supported by the observation that bacterial group II introns include all known lineages and generally behave as functional retroelements whereas organellar introns belong only to two clades that are frequently eroded. The ancestral eubacterial retroelements might be a descendent of a self-splicing ribozyme with or without RT dating back to the LUCA age. Group IIC introns appear to be the earliest branching clade perhaps dating back into the LUCA age, but as with the tree in Fig. 11.6 statistical significance is weak and in agreement with the stem group character and evolutionary temperature of evolution in the LUCA age (Simon et al. 2009). Post-LUCA, when the eukaryote nucleus had evolved, group II introns are thought to have invaded the nucleus and proliferated to many genomic sites as is typical of hybrid dysgenesis causing invasions of other transposable elements (Engels 1997; Evgen'ev et al. 1997; Kidwell et al. 1977). As with transposons, over several generations, the ribozyme structure degenerated producing non autonomous copies that became fragmented into snRNAs that now function in trans in a common splicing gadget, the modern spliceosome (Sharp 1991; Will and Lührmann 2011). A recent hypothesis posits that at the root of the three domains of life the introduction of group II introns by bacterial endosymbionts was the trigger for a fundamental step in the evolution of eukaryotes. This step was the formation of the nuclear membrane separating transcription from translation aiding to prevent translation of incompletely spliced RNAs (Martin and Koonin 2006). In any case, the separation of transcription and translation into different cellular compartments prevents direct

Fig. 11.8 (continued) shown here and in Fig. 11.7 have been confirmed by crystal structure analysis. (**b**) Mechanism of synthesis dependent strand annealing (SDSA) (Gloor and Lankenau 1998; Lankenau 1995; Lankenau and Gloor 1998; Nassif et al. 1994) compared to (**c**) group II intron retrohoming. In (**c**) only one example involving the intron lariat RNA is shown. (**c**) Step 1, the group II intron is transcribed and the lariat forms as in Fig. 11.7b, the lariat RNA EBS sequences associate with complementary IBS sequences in DNA target. Step 2, RT synthesizes along lariat. Step 3, in order to anneal with target DNA and proceed replicating the entire intron information, the reverse transcript DNA performs strand transfer and annealing reactions as in bitemplate SDSA repair (3a). Step 4, fill in reactions on both DNA strands and removal of lariat RNA by RNaseH. Step 5, fill in and flap removal as in standard SDSA DSB repair (see also color plate 47 in Craig et al. 2002; Belfort et al. 2002; Lambowitz and Zimmerly 2011)

access of the intron's domain IV ORF-encoded protein to the intron RNA. This triggered the evolution of splicing factors functioning in trans (Lambowitz and Zimmerly 2011). Eukaryotic genomes contain numerous introns. In an adaptationist setting, they evolved snRNAs derived from group II intron domains into a general RNA-based catalytic machinery replacing that in individual introns. Modern snRNAs still recognize modern introns via conserved 5′- and 3′sequences and a branch-point nucleotide similar to those of group II introns. Splicing then occurs by the same transesterification reactions. This intricate RNA-based spliceosomal machinery of modern eukaryotes seems to be a persisting living molecular fossil like the ribosome. It is actually the strongest evidence that the eukaryotic splicing engine evolved from mechanisms represented by group II introns with probably lots of related molecules replicating in LUCA settings (Lambowitz and Zimmerly 2011). Further, the homology search engines performing genome wide homology searches today also trigger in modern organisms a presumably ancient DNA repair mechanism, i.e., SDSA (Fig. 11.8b).

11.3 The Ancient RNA World

As we have seen above, a thriving modern RNA world is ubiquitously present within all contemporary organisms (see also Wang et al. 2011). It follows that there must have been an ancient RNA world within their ancestors during the LUCA(S) age as well. Indeed, since long it was suspected that an RNA world was historical reality. In 1976, White noted that RNA fragments attached to various cofactors are widely distributed in modern terran life (White 1976). These RNA-cofactors today serve ubiquitous metabolic key functions: adenosine triphosphate (ATP) transfers phosphate transfer energy, S-adenosylmethionine performs one carbon transfers, flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide (NAD) serve in redox electron transfer chains, and coenzyme A (CoA) assists carbon-carbon bond formation. Therefore, these cofactors were likely present in the LUCA(S) and indicate that they facilitated the historical RNA replicators from the beginning. Other molecules such as random peptides, lipid molecules, and inorganic atoms must have coexisted and facilitated RNA strand formation and replication as cofactors as well.

Entering this primordial RNA world, we now, more often have to rely on reconstructions and simulations in our rhetorical disquisition, but a trajectorial concept is visible. We shall accept that about every strategy available to draw conclusions about the primal synthesis leading toward an ancient RNA world is welcome. Therefore, physico-chemistry based bottom up-approaches as well as geological and astronomical data and experimental approaches likewise are most welcome to nourish ideas about the LUCA(S) age and to feed concepts for better comprehension into Popper's cycle (Fig. 11.1).

11.3.1 The Protoribosome, Bridging the Current and the Ancient RNA Worlds

Group I and II introns opened a window into the past toward understanding the LUCA world at the root of the three domains of life, i.e., archaea, bacteria, and eukaryotes. Group I and Group II introns have secondary structures conserved within each clade but distinct from each other between the clades. The common feature is that splicing is mediated by intron RNA with ribozyme activity that can proceed in the absence of proteins and both intron groups involve SDSA-like features including complementarity search and strand transfer reactions (Saldanha et al. 1993) (Fig. 11.8c). The sequence structure of mobile introns however is not conserved in evolution and that makes it difficult to move further back in time with mobile introns.

Fortunately, the universal sequence conservation of another molecule, i.e., the ribosome, grants moving back even further in deep time. The three essential functions of the ribosome are: (1) decoding of the genetic code, where tRNA anti-codons pair with a complementary triplet on a mRNA strand; (2) peptidyl transferase, where two amino acids are covalently linked to a di- or polypeptide; and (3) the translocation reaction, where an aminoacyl tRNA at site A moves into a new sterical position in the peptidyl site P within the ribosome. In contemporary bacteria, the 23S rRNA component of the ribosome contains the peptidyl transferase activity, central to translation, and acts as a ribozyme. Proof that none of the ribosomal protein moieties plays a role in the catalytically active site was revealed by the atomic crystal structure of ribosomes (Nissen et al. 2000). No protein moieties were found within 17Å of the active site. Therefore, peptide bond formation is indeed catalyzed by RNA with no part of any protein component close enough to play a direct (chemical) role in the reaction (Moore and Steitz 2011; Ramakrishnan 2011). Because of its universal distribution in all life, the catalytic center of the ribosome is the most ancient living molecular fossil on earth. The subcomponents of the ribosome are shown in Table 11.1. The 23S rRNA of prokaryotes and the 28S rRNA molecules

Table 11.1 Ribosome composition

Domain	Sedimentation coefficient	rRNA	Proteins	Catalytic activity
Prokaryotic	70S			
	Small subunit 30S	16S = 1,500 nt	21 proteins S1–S21	
	Large subunit 50S	5S = 120 nt *23 S = 2,900 nt*	31 proteins L1–L31	Peptidyltransferase ribozyme
Eukaryotic	80S			
	Small subunit 40S	18S = 1,900 nt	~33 proteins S1–S33	
	Large subunit 60S	5S = 120 nt 5.8S = 156 nt *28S = 3,400-4,700 nt*	~50 proteins L1–L50	Peptidyltransferase ribozyme

of eukaryotes are 2,900 nt and 3,400–4,700 nt long respectively. In all modern cells, this translation activity catalyzed by the ribozyme must be facilitated by a large number of co-factor proteins and other rRNAs (Table 11.1) that, figuratively speaking, were added successively in evolutionary history like the shells around the core of an onion (Bokov and Steinberg 2009). The long standing hypothesis, that the first ribosome-like entity, which probably appeared between 3.5 and 4.0 Ga ago, was made entirely of RNA, today seems more likely than ever. Nevertheless, it would be foolish to assume that the biological world in which that first ribosome appeared was one in which RNA was the only relevant polymer (random or informational) (Moore and Steitz 2011). It likely needed co-factors from start on (this book).

Everything we know about the ribozymal catalytic site in the ribosome is based on *modern* ribosome crystal structures, but we cannot go back in time 4 Ga to check out the primordial ribosome directly. However, based on structural data, Ada Yonath and colleagues suggest that there is evidence for a ~55 nt long RNA molecule which they termed the protoribosome (Davidovich et al. 2009; Yonath 2009c). It would be a member of the ancient RNA world from which our present RNA world evolved by Darwinian evolution. According to their work, the ancient translation apparatus may have survived selection pressures, and its vestiges may indeed be embedded within all modern ribosomes (Davidovich et al. 2009). The secondary structure of the protoribosome comprises an RNA-homodimer of *stem-elbow-stem* elements. From the basic ~55 nt RNA, the ribosome evolved successively over billions of years into the modern ribosome, which includes many protein components and other rRNAs (Bokov and Steinberg 2009). The protoribosome originated from gene (!) fusion and gene (!) duplication – a process that reminds us of the SDSA-like processes mentioned in Fig. 11.8 (Yonath 2009a). Independently from the sequence (RNA and protein) the three-dimensional structure of the two halves of the ribosomes that carry out the peptidyl transferase reaction, has been conserved stereo chemically until today. According to this, the ancestral ribosome possessed a central pocket built of two chains of the same ~55 nt RNA forming a dimer. These two RNA molecules were sufficient, according to Yonath's hypothesis, to carry out peptide synthesis.

The ~55 nt short protoribosomal RNA is an extrapolation from atomic ribosome structures (Davidovich et al. 2009; Yonath 2009c). Nevertheless, it is hypothetical, and we have now reached the other side, down under the *golden spike*, and outside Popper's cycle (Fig. 11.1). At this point, there is only one theoretical extrapolation from the preceding RNA world possible where we are still on firm ground: The very first protoribosome-like entity that existed must have done so in multiple copies and not just in one. If it had been only one molecule, it would not have existed as an entity. Reproduction was mandatory. Any further, reproduction needed to be reproducibly stable reproducing similar molecules over many generations. Thus, because there had to be a multi-copy population of protoribosome molecules in order to be called an entity, something (mechanistic or structural) must have been responsible for making copies of the protoribosome. This something would be called "template." The protoribosome already encompassed a distinct secondary structure (*stem-elbow-stem*) that was enforced by some sort of primary RNA sequence and

base-pairing constraints. Thus, the protoribosome likely was in state to self-prime fragments of its own sequence and possibly could have served as template and ribozyme in mobile intron-like priming and retro-replication activities (Fig. 11.8c).

If we would proceed further from here, the amount of speculation likely became too much for the good. Therefore, we will now briefly leave the RNA world and make a jump to a new, bottom up hypothesis based on the simultaneous consideration of many bioenergetic, physical, and geological constraints. This hypothesis was recently developed by Armen Mulkidjanian (Mulkidjanian 2009; Mulkidjanian and Galperin 2009). It leads us back into a distant, post-hadean age, bare of life and filled with hardening minerals, consolidating igneous rocks, and sunlight triggered photo polymerization resulting in the emergence of the first replicators. The hypothesis is so attractive because it unifies seemingly opposing ideas such as "metabolism first" versus "replication first" scenarios. According to the Zn world hypothesis, life emerged on earth as a "proto metabolism-driven replication." From a sun light penetrated, primordial hydrothermal field environment, RNA molecules started to replicate at some point, facilitated by many coexisting and interacting protometabolites (Chap. 9) possibly already involving simple proto-peptide replicators (Ashkenasy et al. 2004). From there, we shall arrive at the protoribosome again as the earliest trace record of a replicator bridging the ancient and modern RNA worlds.

11.3.2 The Zn World Hypothesis and First RNA Replicators

The Zn world hypothesis represents a perfectly reasonable deviation from a beaten track[15] of alternative hypotheses. The accompanying chapter by Mulkidjanian and Belozersky (this book) unfolds the details and handles the matter with utmost expertise. Therefore, let me only give a brief, overlapping account in order to link the *two* RNA worlds above and below this section. In two seminal papers, Mulkidjanian proposes that the Zn world involved sub-aerial, sun exposed ZnS (sphalerite) precipitates supporting a mechanism of continuous abiogenic photosynthesis of prebiotic metabolites and their further conversion by ZnS-confined replicating entities (Mulkidjanian 2009; Mulkidjanian and Galperin 2009) (see Fig. 1 in Mulkidjanian & Belozersky, this book) Two basic assumptions are critical for the hypothesis:

1. Availability of enough UV-irradiation from the Sun
2. The requirement of high atmospheric pressure

According to the "faint young Sun puzzle" 4–3.5 Ga ago, the Sun was about 30% less bright than today (Nisbet and Sleep 2001). If such a sun shone on the current

[15]In paraphrase to Richard Feynman's letters collected in the book: "Perfectly Reasonable Deviations from the Beaten Track."

Earth, there would be global glaciation. However, in contrast to visible light, X-ray and UV luminosity emitted from the early Sun was likely higher by analogy with other Sun-like stars that generally show decrease in X-ray and UV luminosity with age (Cnossen et al. 2007). Therefore, the Zn world hypothesis starts from the premise that without the ozone shield solar UV-radiation reaching Earth's surface land or water was 10–1,000 times higher than today. Further, the atmosphere was dominated by CO_2, with smaller amounts of CH_4, N_2, and condensing H_2O vapor. The surface pressure was 10–100 times higher than today changing the critical points for phase transitions of precipitating minerals according to their specific phase diagram. Currently, it is thought that CO_2 was the main material from which primal organic polymers were formed. CO_2 reduction occurred close to sub-aeral hydrothermal vents that were common in costal and shallow waters as well as on land like nowadays in Yellowstone, or in the Afar Danakil depression. The energy for forming organic compounds came from UV light and CO_2 embodied the building blocks for photo polymerization. The ZnS hypothesis is based on Wächershäuser's insight that an abundant mineral on earth could have provided the energy for carbon fixation in a redox process converting ferrous ions and hydrogen sulfide of basaltic origin into pyrite (FeS_2) (Wächtershäuser 1988). The detailed reasons for why ZnS or MnS are better candidates for this process is explored in the chapter of Mulkidjanian and Belozersky (this book).

Indeed, the ZnS and MnS precipitates close to sub-aerial or coastal hydrothermal vents appear ideal as a primal model system to focus efforts at understanding the emergence of life from abiotic settings. Pumice-like textured ZnS precipitates formed honeycomb micro compartments of <1–100 μm in diameter, such as in fossil and modern vents (Branciamore et al. 2009; Kelley et al. 2005; Koonin and Martin 2005; Martin and Russell 2003). We now can imagine that many, if not all, requirements existed within ZnS compartments for the production of many organic, pre-metabolic molecules that accompanied or even facilitated the polymerization of the first RNA polymers (see Chap. 1; Mulkidjanian 2009; Mulkidjanian and Galperin 2009). Next to a number of pre-metabolic compounds, the most significant coexisting polymers were abiotic, random peptides or proteinoids, lipids, and RNA precursors – all in a heterochiral mixture. Polymerization of monomers (both nucleotides and amino acids) in aqueous solution under experimental conditions produced only short, let it be, heterochiral oligomers (≤10 mer). High photostability

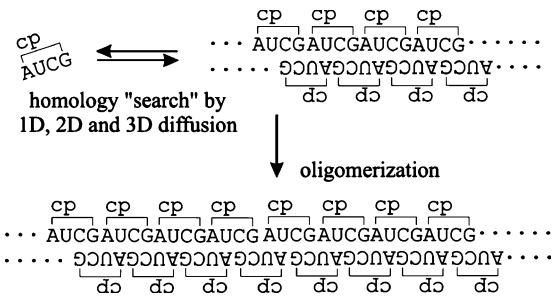

Fig. 11.9 Simple oligomerization of tetra nucleotides involving a homology search and primordial, pre-SDSA-like annealing; cp = 2′, 3′ cyclophosphate; (Szathmary 1997)

of the monomers and polymers was a side-effect as the ZnS compartments catalyzed photopolymerization under UV illuminated exposure exerting high selection pressures on photo-instable monomers (Chap. 1). In Chap. 1, Mulkidjanian and Belozersky point out the UV-protection mechanism for the evolution of primordial RNA-like polymers. According to this mechanism, it is likely that small double stranded snippets of RNA, composed of single stranded fragments that were dynamically synthesized and degraded and synthesized again, may have been annealed to random complementary sequences, forming Watson-Crick base pairs, that provided an enhanced photostability and a selective advantage for survival. It was a process very similar to and predating the universal SDSA-mechanism (Lankenau 2006). Surface-bound template polymerization at ZnS surfaces was possible as well. It has been shown that presence of montmorillonite surfaces induces formation of oligomers up to ~55 monomers long, as is the size of the predicted protoribosome (Davidovich et al. 2009; Ferris et al. 1996). Simple ribozyme activity was therefore feasible on mineral surfaces as experiments by the Bartel lab have shown (Bartel and Szostak 1993; Ekland and Bartel 1995, 1996; Ekland et al. 1995; Johnston et al. 2001). These experiments demonstrate the practicability of autocatalytic simple ligations up to multiple, successive ligations of several nucleotides. Multiple ligations in fact represent nothing else but the origin of continuous replication. However, not every sequence is replicated yet in these experiments. An important feature in the evolution of RNA polymers was probably the involvement of complementary annealing. Tetranucleotide-2′, 3′- cyclophosphates were shown to assemble into oligomers of up to 36 nucleotides in dilute solution, with oligomerization proceeding highly chiroselectively (Bolli et al. 1997) (Fig. 11.9).

Homochiral templates catalyze the formation of the correct phosphodiester junction between homochiral tetramers that have the same chirality as the template. The origin of homochirality in biomolecules is a significant issue for the evolution of replicators and here is one facet of this subject. Using purine ribonucleoside 5'-phosphoimidazolides as activated monomers and oligomeric or polymeric ribonucleotides as templates allows for no enzymatic, template-directed syntheses of oligonucleotides (Naylor and Gilham 1966). Even though it is only a historically unrealistic simulation, von Kiedrowski showed that these imidazolides in vitro grow as artificial replicators without replication enzymes or ribozymes (Sievers and von Kiedrowski 1994; von Kiedrowski 1986). Thus, it appears imaginable that in ZnS micro compartments, homochiral templates of D- and L-enantiomers evolved with the degree of oligomerization growing in time. Each chiral template then gives rise to its own kind. When photochemically synthesized raw-monomer supplies became limiting, the replicator sequences started to compete for raw materials and an imbalance of chiral kinds built up. Once the symmetry of the molecular population was broken only one kind of chirality survived. Of course, there were many hydrothermal vents all over the earth and feeding only one particular kind of chirality into LUCA(S) was likely a long-term Darwinian process. Random peptide replicators could have emerged as well (Ashkenasy et al. 2004; Paul and Joyce 2004). In coexistence with RNA replicators, they might have been selected for coexistence with RNA replicators from the beginning. The most stable and most

abundant peptides may have had developed binding affinities stabilizing replicators and further photo protecting RNA from destruction. Thus, the first RNP replicator network could have been emerged and continued the Darwinian path toward the protoribosome (Cech 2009). Is all this still to be called pre-Darwinian evolution? At which point of replicators does Darwinian evolution start?

11.3.3 Quasispecies and the Error Threshold

11.3.3.1 The Initial Darwinian Ancestor (IDA)

In a bottom-up journey, we cannot explore LUCA(S) as this is the stemgroup of the three domains of life caught in Koonin's "irreducible complexity trap" (Koonin 2009) (to be published elsewhere). LUCA(S) remains causally attached to present life forms as viewed from the anthropic perspective (Fig. 11.1). Based on Bayes' Theorem, i.e., a quantitative version of the anthropic argument, Michael Yarus suggests that there was a substantially older IDA where nicotinamide adenine dinucleotide (NAD) or its congeners were the earliest cofactors to enter the biochemical inventory of all living cells (Yarus 2011) (Chap. 9). According to this hypothesis, in step 1, activated nucleotides and coexisting, interacting molecules oligomerized arbitrarily. In step 2, initially random replicators such as AMP-containing cofactors became abundant with simple templating and minimal catalysis. In step 3, dinucleotide 5′–5′replicators such as NAD participated and were selected for "usefulness" (increasing fitness) in metabolism. Then, in step 4, the ancestral RNA world emerged (or RNP world, as random peptides may have coexisted in the ZnS/MnS compartments (Cech 2009)), where 5′–5′ cofactor initiation and reactivity may or may not have participated. Yarus states that the 5′–3′ RNA replicase replicator created the first RNA world at 4 Ga ago. If it took place in the ZnS world, the window would be rather between 3.5 and 3.9 Ga according to the Zn world hypothesis (Mulkidjanian 2009; Mulkidjanian and Galperin 2009). In step 5, translation and coded peptides were invented employing Ada Yonath's et al. protoribosome (Davidovich et al. 2009). 5′–5′ cofactors like NAD, FAD, NADP were adopted by peptide catalysts. They were readily available in spark experiments and several synthetic routes employing plausible prebiotic chemicals (Cleaves and Miller 2001; Friedmann et al. 1971). The stability of the ribose moiety was increased substantially in presence of borate minerals where thermal springs and hydrothermal fluids are an important source of B to continental evaporites (Grew et al. 2011). This points to a sub-aerial relevance as with ZnS. Within sphalerite (ZnS) compartments, photo-excited electrons could further be transferred from the surface of photo-excited ZnS particles to NAD^+/NADH and co-electron acceptors in order to evolve energy-managing redox systems (Mulkidjanian and Galperin 2009; Mulkidjanian et al. 2009). It has not been shown yet that single activated nucleotides polymerize on dinucleotide templates, especially on 5′–5′dinucleotides – perhaps a challenge to show for photo excited ZnS

compartments? Indeed, it would be a great step forward if it could be shown that a random replicator such as a dinucleotide- IDA evolved into longer chain replicators under any conditions. In 1975, Sumper and Luce demonstrated that a mixture containing no RNA chain at all but only activated RNA base monomers plus the enzyme Qβ-replicase spontaneously generated long, self-replicating RNA chains (Sumper and Luce 1975). Such a simulation does not prove that random prebiotic peptides could do the same job, but it appears just a matter of time for the initial evolution, e.g., in ZnS compartments to imagine the evolution of effective replicators. As a next step, only a population of RNA sequences representing a single consensus sequence stands for such an effective replicator and can be called an individual genetic entity from where Darwinian evolution can start (Robertson and Joyce 2011). A population of such sequences operating at the boundary between chaos and order was called *quasispecies* (Eigen 1987, 1992) which we will learn about next.

11.3.3.2 In Vitro Evolution and the Spiegelman Monster

Sol Spiegelman is known for his recognition that for a certain fragment of DNA, only one of the two DNA strands (the Watson or the Crick strand) encodes the genetic information. However, for our purpose of the ancient RNA world, it is Spiegelman's evolution experiments that are of relevance. The Spiegelman Monster is a specific RNA molecule of ~218 nt length that is able to reproduce and evolve in vitro in the presence of an RNA polymerase, i.e., the Qβ replicase (Kacian et al. 1972).

In the 1960s, Spiegelman and colleagues employed the plus-strand RNA virus, Qβ, which infects the bacterium *Escherichia coli*. The RNA genome of Qβ has a length of 4,500 nt and encodes its own replicase and capsid proteins for packaging the virion. Infection starts with the docking of a Qβ encoded adsorption protein to the sex-pili on the outside of a cell. After infection, the plus-strand RNA serves simultaneously as genomic and as mRNA inside the infected *E. coli* cell. When Qβ encoded replicase is translated, it rapidly catalyzes replication. Qβ-replicase consists of four subunits and recognizes both, plus strand RNA as well as minus strand RNA. It binds to the 3′ends of both RNA strands and both strands act as templates reciprocally to one-another (Domingo et al. 1976). This sets up a replication cycle similar to what we expect might have happened in ZnS compartments or other locations on the primordial Earth. A stable internal secondary structure prevents the RNA from forming an RNA helix during replication. After about ten to one hundred thousand Qβ molecules have been produced, the plus strand RNA molecules are packaged into capsid protein again, and a Qβ encoded lysis factor ministers lysis of the host cell and release of free Qβ virions.

A population of 100,000 Qβ virions resembles populations of some animal and plant species. However, higher organisms, and even *E. coli* cells possess the proofreading activity of modern DNA polymerases, making replication of their genomes highly reliable and safeguarded against too many mutations (Kornberg 1989; Kornberg and Baker 1992). The Qβ replicase does not possess proofreading

and the quantity of errors is high, i.e., 3/10,000 = 3×10^{-4}. Because of this high mutation rate, only a fraction of the released Qβ virions are infectious. The whole population of virions - or better the population of RNA genomes - including its mutant spectrum, was termed *quasispecies* by Manfred Eigen and colleagues (Biebricher and Eigen 2006; Eigen et al. 1988; Epstein and Eigen 1979). In *natural populations* of the *E. coli* bacterium, the Qβ virus coexists together with a non-infectious Qβ variant (called minivariant) that acts as a superparasite on Qβ (Orgel 1979). The length of this superparasite is ~220 nt and it evolved from the 4,500 nt long Qβ RNA. Since the minivariant does not encode the Qβ replicase anymore, replication and the infection potential completely depend on the intact Qβ replication and infection machinery. To overcome the negative effects of its dependence on the original Qβ virus, the minivariant RNA evolved a size and a secondary structure that increased replication speed by Qβ replicase (Eigen 1992). Such a minivariant of ~220 nt length is the naturally occurring equivalent of the experimental Spiegelman Monster.

Superimposed systems of parasites are ubiquitous for life and well known from the transposable element world. For example, SINEs (short interspersed nuclear elements) are a subclass of transposable elements that do not encode any proteins themselves but instead have evolved to parasitize the LINE retrotransposition machinery. Many SINES are also chimeric with a 5′ region derived from tRNA and a 3′ tail derived from the 3′ end of a LINE. Most famous are the 280 nt long *Alu* elements with 1.5 million copies, i.e., 11% of the haploid human genome. They are sequence related to 7SL RNA.[16] Despite their parasitic nature, both LINES and SINES are now known (as are many examples of endogenous retroviruses) for being incorporated into novel genes, so as to evolve new functionality.

The Qβ minivariant, therefore, serves as an in vitro model system simulating the ancient conditions of mass-effects of small RNA replicators as they likely existed in primordial environments such as ZnS compartments 4 Ga ago. At the same time, this gives us information about the dynamics at the genome level in modern organisms. Thus, the question we pose is: under which conditions did minivariants evolve initially? What were the conditions for gliding from the abiotic, non-Darwinian side with "random replicators" as depicted by Yarus (2011) to the side of informational *quasispecies* employing Darwinian selection, and driving evolution to be discharged into the worlds of adaptations, exaptations, punctuated equilibria, and molecular and organismic drivers under environmental constraint (Burt and Trivers 2006; Gould 1997, 2002).

Based on the mechanisms of Fig. 11.5, a new experimental system was set up termed self-sustained sequence replication (3SR) (Gebinoga and Oehlenschläger 1995). In 1997, Eigen and Oehlenschläger showed that the Spiegelman Monster

[16]The signal recognition particle RNA, also known as 7SL, 6S, or 4.5S RNA, is the RNA component of the signal recognition particle (SRP) ribonucleoprotein complex. SRP is a universally conserved ribonucleoprotein that directs the traffic of proteins in the cell and allows them to be secreted.

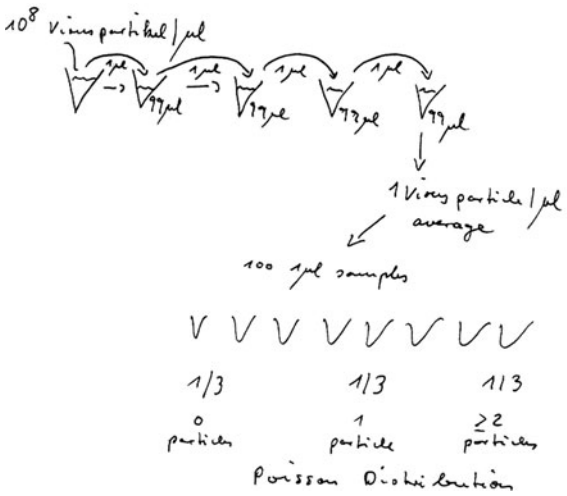

Fig. 11.10 Serial dilution of viral suspension with established titer (Drawing from the author's Baltimore/ Heidelberg lectures)

eventually becomes even shorter, containing only 48 or 54 nucleotides, which are simply the binding sites for the reproducing RNA polymerase (Oehlenschläger and Eigen 1997). This size is effectively the size of the protoribosome as reconstructed by Ada Yonath and colleagues (Davidovich et al. 2009).

Crucial to all experimental systems is the serial dilution and transfer principle. It was first applied in 1938, when Max Delbrück started to work with bacterial viruses (Ellis and Delbrück 1939). For example, when a bacterial culture is infected with phage lambda, optical density clears after a couple of hours because all replicating bacteria have been lysed and killed. Plating bacteria on a solid agarose plate and pouring phage solution on top will clear the plate off bacteria as well. However, when the phage suspension is diluted such that only very few phages remain, individual phage particles produced by a single ancestor phage produce singular clear plaques in the bacterial lawn. Delbrück realized that when diluting the phage suspension such that, on average, *no* phage particle per vial was the most common dilution event, he could apply the Poisson formula and calculate future serial transfers (Poisson 1837) as shown in Fig. 11.10. Application of the method initiated molecular biology and experiments on mutant distributions as explicated next.

11.3.3.3 Mutant Distributions of Replicators

Sol Spiegelman, Charles Weissmann, and Manfred Eigen set up experimental systems in which sequence evolution relative to the reproductive speed (fitness) of RNA molecules could be exactly monitored and controlled (for review and references, see Eigen 1983, 1992). Qβ-replicase is able to replicate any RNA molecule in the presence of the four NTPs (ATP, GTP, UTP, CTP). An important experimental strategy is to clone each individual sequence mutant from large

quasispecies populations. Based on a known Qβ virus titer, a serial dilution must be carried out (Fig. 11.10). Sequencing from over diluted vials allows for sequencing any number of single molecules. An experiment starts by adding the four NTPs and Qβ replicase under defined conditions. The reaction takes place in a so called "evolution reactor" (Eigen 1992). It is initiated and stopped by increasing/decreasing the temperature from 0°C to 37°C within a second. Growth of the increasing number of replicons is monitored by a laser fluorimeter with glass fiber optic and the reaction is stopped in the exponential growth phase. A fraction of the RNA molecules will be reverse transcribed, cloned, and sequenced. Today, next generation sequencing can be applied. The experiment can go on by starting a new round of replication from a single RNA molecule. This can be repeated over many generations under defined experimental conditions.

The most spectacular molecules, replicating under presence of increasing amounts of ethidiumbromide or RNaseT1, evolved resistance against such detrimental agents. As mentioned above, the Spiegelman Monster itself evolved from a ~4,500 nt long Qβ virus RNA as starting material. Finally, a short ~220 nt long, fast replicating RNA molecule emerged which very much resembled the Qβ minivariant superparasite of natural *E. coli* populations (Orgel 1979). After few generations of independent evolution, individual subpopulations possessed the same consensus sequence. But, after many generations of multiplication, each independently evolved RNA population started to establish its own characteristic mutant spectrum represented by a characteristic consensus sequence.[17]

A combination of quantitative experimental and theoretical simulation assisted research lead to fundamental insights into the very matter of the origin of life. Let me draw the picture by starting out with a population of say 10^8 RNA replicators each 50 nt long. This population dwells in one of the micro-compartments of Mulkidjanian's Zn world.[18] Here, it is irrelevant by which mechanism the RNAs or RNPs emerged – e.g., either, via random peptide facilitated short RNA chains that became more abundant by von Kiedrowski/Yarus like processes (Sievers and von Kiedrowski 1994; Terford and Von Kiedrowski 1992; von Kiedrowski 1986; Yarus 2011) or by Qβ-type de-novo synthesis facilitated by some random peptide (or an RNP) instead of Qβ replicase as reported by Eigen's group (Eigen 1983).

Figure 11.11 summarizes Manfred Eigen's group's results. Let me recapitulate the example (Eigen 1988, 1992; Eigen et al. 1988): We start in Fig. 11.11a with a theoretical population of RNA replicators, each 50 nt long, like the protoribosome. Only purines (R) and pyrimidines (Y) are considered. Thus, only two types of mutations are possible: Y - > R and R -> Y. The coordinate system depicts the replication error rates ($1 - q$; abscissa) versus the relative population number (X_d; ordinate). Initially, we assume that all molecules are identical and the replication

[17] *Consensus sequence* refers to the most common nucleotide at a particular position after multiple sequences are aligned. In case of the mutant spectrum of a *quasispecies*, the consensus sequence is identical with the master sequence.

[18] Other locations such as beaches or brine ice may do the job as well.

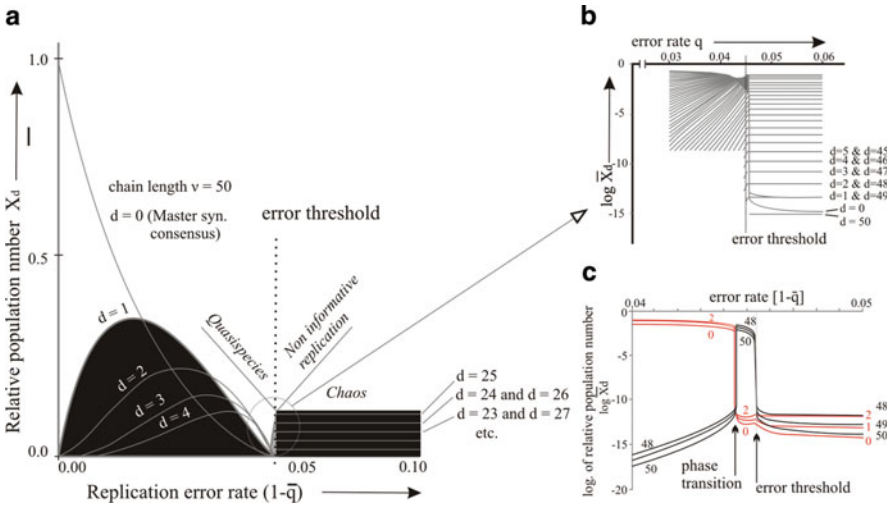

Fig. 11.11 Conditions for a replicating entity of protoribosome size, i.e., a *quasispecies* of 50 nt chain length (ν) at different error rates. For detailed explanations see main text (Combined, redrawn, and modified after Eigen (1992); Eigen et al. (1988)))

error rate $(1 - q)$ is zero. Therefore, the population number X_d is 1.0, that is all molecules (100%) are identical. The index d is the distance (in numbers of nucleotides) of a mutant sequence to the best adapted master sequence, called the Hamming distance. As all sequences are identical and the Hamming distance of all RNA replicators is 0, $X_d = X_0$. Therefore, X_0 designates the master sequence of the population. X_d is the ratio between the amounts of individual molecules with the distance d and the total amount of all existing individuals. The X values with indices $d = 1$ to $d = 50$, represent the sum of all mutants within each error class 1–50. For example, X_1 encompasses all 1-error mutants independently to the position within the 50 nt sequence. The X-values in the ordinate are given as relatives between 0.0 and 1.0 (0–100%), summing up to 1.0. The replication error rate $(1 - q)$ relates to the probability in which a position mutates. An average of each 1/100 base wrongly incorporated gives a mean error rate of 0.01. At an error rate zero, there is a strong "all or nothing" judgment. Either the sequence is identical with the master and lives or it will die after the first mutation (*survival of the fittest*). With a growing rate of mutations, the population number of the master sequence (X_0) decreases rapidly to low numbers. Quickly, there are (in the sum) more $X_1, X_2, X_3, \ldots$ error mutants than X_0 molecules representing the master. Experimentally, such an individual master-molecule could not be identified by sequencing, as it will become too rare. It is still mastering the population for its superior fitness through positive selection, but it may become nearly extinct, while its mutants are still symmetrically distributed around the master.

Nevertheless, the master sequence can be reconstructed from aligning many mutant sequences such that their consensus sequence identifies the master sequence. In other words, with increasing error rate the master molecules drop in

copy number to nihility, but the master's information still determines the mutant distribution. Finally, when the replication error rate increases even further, a threshold value is reached (Fig. 11.11a, b at $1 - q = 0.046$) and the *information* represented by the master sequence instantaneously disintegrates. The population number of the master abruptly collapses to 15^{-15} (Fig. 11.11b). The master sequence then is just one of $2^{50} \times 10^{15}$ possible variants (remember, we are dealing with binary R, Y sequences). The formerly selected sequence information has now been lost completely, chaos is ruling sequence space. Eigen and colleagues termed this informational break down a *phase transition* in analogy to thermodynamics, e.g., when a liquid becomes gas upon heating to the boiling point, resulting in abrupt change in the state of aggregation. The information evaporates.

Figure 11.11c depicts details at the phase transition point and the error threshold for two similar fit sequences. Just around the error threshold, Eigen identified the most favorable conditions for evolutionary change. In his example (Eigen 1988, Abb. 3 p.127), he examines a binary R/Y sequence of 50 nt length again (i.e., protoribosome size) and explores the consequences of minute fitness differences. The sequence with the highest selective value (I) (master) is arbitrarily set to 1.0 with I_0 (zero errors). Its antipode, a sequence that possesses the complement in all 50 positions (i.e., 50-error mutant) is called I_{50}. This mutant has a similar high selective value of 0.9. Therefore, I_0 and I_{50} represent two similar fitness peaks with only slight differences. All other mutants get a low selective value of 0.1, except the 50 possible one-error-mutants of I_{50} receiving the selective value 0.5. The latter class of mutants are at the same time the 49-error-mutants of I_0 and are, therefore, called I_{49}. In this distribution, the best adapted singular sequence I_0 competes with the slightly less selectively favored singular antipode I_{50}, which is surrounded by multiple, better adapted I_{49} mutants.

Figure 11.11c details the results of this competition around the error threshold: Shown are again the logarithmic values of the relative population numbers $\log X_d$ for all complete error classes I_0 to I_{50} (sum of all individual sequences within the error class) as a function of the error rate $(1 - q)$. With two "*races or breeds*" of sequences, we observe two phase transitions; with a relative low error rate, I_0 wins, because it is the best adapted individual sequence. This is what would be expected from Darwin's principle, *survival of the fittest*. At an error rate of 0.0445, i.e., the phase transition point the information of I_0 "evaporates" (Fig. 11.11c *red lines:* distributions with 0 and 2 mutant molecules are shown). At this point, however, the sequence I_{50} (a race of its antipode distribution in population biological language) is still stable because it has a better value topography of its neighbors (the I_{50} mutant with a selection value of 0.9 and the I_{49} mutant with a selection value of 0.5, whereas I_0 with selection value 1.0 is only surrounded by selection values of 0.1). In this way, I_{50} with its idiosyncratic clan of mutants (fifty possible one-error-mutants I_{49}) now becomes the target of selection (Fig. 11.11c black line distributions). Only at the error threshold with an error rate beyond 0.045, the information finally becomes victim of an error catastrophe.

In this example, we see that the target of the selective evaluation is not the individual sequence type (I_0 or I_{50}) but the whole mutant spectrum. This is the

quasispecies, and in my opinion this is a plausible starting-framework of conditions for the IDA as well. Typical of the phase transition are the dramatic changes in population numbers of several orders of magnitude. As mentioned above, a binary sequence with 50 positions has $2^{50} \gg 10^{15}$ possible variants. When the threshold is passed, the population numbers of individual sequences – visible for I_0 and I_{50} (Fig. 11.12c) – fall to statistically insignificant numbers of 10^{-15}. The error classes I_1 to I_{49} of course include different individual alternatives. Therefore, the statistically most abundant sequences I_{25} dominate (Fig. 11.12a).

For our analysis of the random RNA molecules in some kind of hydrothermal vent compartment, the experiments and simulations above have tremendous consequences. If we read each plot of Fig. 11.11 from the right to the left, ZnS chambers might have produced RNPs of the type mentioned by Cech (Cech 2009). However, sequence information as in a *quasispecies* could not have existed a priori. It must have been a chance event when the first single RNP[19] molecule arose and produced a replica with a low enough error replication rate and high enough fitness value to overcome the error threshold and make a phase transition into a stable *quasispecies*.

In natural systems, however, the error rate topic can be much more complicated. For example, the Eigen threshold made us believe for long that primitive RNA sequences could not evolve into significantly longer chains exceeding about 50 nt. This created the Eigen's paradox: "*no protein without longer gene, no gene without protein*." A recent paper however argues that replication seizes and slows down in speed at a mismatched point mutation using the complementary sequence as template. In that view, it becomes anticlimactic that master sequences are preferably replicated and longer chains can evolve (Rajamani et al. 2009). An experimental indication that long chains indeed can evolve using imidazolid-replicators was demonstrated for sea ice micro chambers (Trinks et al. 2005). Further, an individual ZnS micro chamber in which a *quasispecies* had emerged, could then have started to export molecules into the environment infecting other micro chambers, neighboring vents, entire hydrothermal systems but also beaches and ice etc. Vice versa, ZnS chambers could have absorbed solvents from a primordial soup generated elsewhere. Relevant is that at some point, replicators evolved that exchanged RNA fragments interdependently through homology search processes and annealing steps. Synthesis dependent strand annealing (SDSA) as in group II intron retrohoming evolved as a novel mechanism, now ubiquitously present in all life. Simultaneously, the protoribosome-like replicators evolved and learned to translate information into protein. This gave rise to Beadle-Tatum-type genes (*one gene-one enzyme concept*) (Beadle 1958).

[19]With "P" in RNP, I refer to random, non informational, non encoded, abiotic peptides associated with random abiotic ribonucleotides.

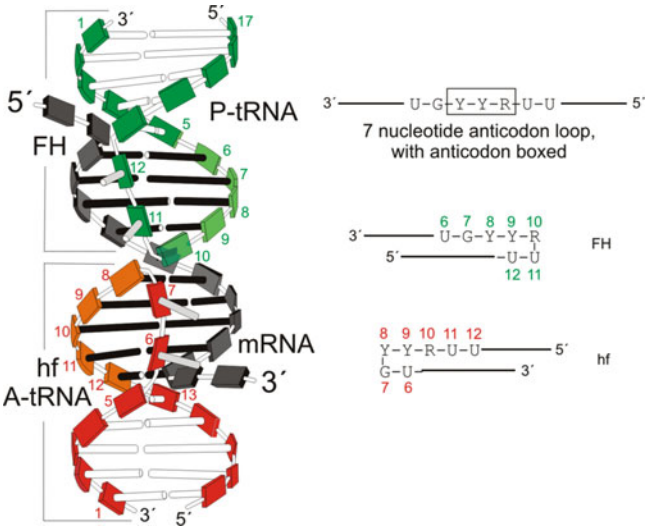

Fig. 11.12 Sterochemical model of a pre-protoribosomal replicator interaction complex based on Woese and Crick (Crick 1968; Crick et al. 1976; Woese 1970). The complex consists of two tRNA anticodon loops in the two possible conformations: FH for the peptidyl tRNA (P-tRNA) shown in green color and hf for the aminoacyl tRNA (A-tRNA) shown in *red color*. Both tRNA anticodon loops are in antagonistic conformations where the bases 6 through 10 (*green*) continue the double helical turn from the tRNA stalk of P-tRNA while bases 8 through 12 (*red*) continue the double helical turn from the tRNA stalk of A-tRNA. Two nucleotides within the anticodon loop are energetically, unfavorably kinked falling out of the double helices (i.e., *red bases* 6, 7 and *green bases* 11, 12). This is compensated for by the total base stacking energy of the 10 bp double helix formed between mRNA (*gray* and *black symbols*) and the two anticodon loops stabilizing the complex. Solid bars represent two arrays of pentaplet Watson-Crick base pairs between mRNA and tRNA anticodon loop. Graphic symbols and numbering of bases are that used by Woese (Woese 1970). FH and hf designates Fuller& Hodgson and hodgson & fuller conformations of the anticodon loops. This structure is the sterical prerequisite for Crick's et al. and Schuster & Eigen's RNY-code hypothesis (Eigen and Schuster 1978b). The energetically unfavorable kinks make sure that the structure is highly unstable and dynamic – switching between pentaplet and triplet Watson-Crick base pairs (for further detail see Woese 1970). The flexibility in addition to the much longer, flexible cloverleaf structure of the tRNA (not shown here) still leaves enough potential for the peptidyl transferase interactions. Most relevant however, the structure allows multiple tRNA like molecules to "creep" along the mRNA strand in a caterpillar like movement (see Fig. 11.13a–d)

11.3.4 Positive Pondering on the Origin of RNY[20]

The origin of protein synthesis is a notoriously difficult problem. We do not mean by this the formation of random polypeptides but the origin of the synthesis of polypeptides directed,

[20]RNY = purine-any nucleotide-pyrimidine "code."

> *however crudely, by a nucleic acid template and of such a nature that it could evolve by steps into the present genetic code, the expression of which now requires the elaborate machinery of activating enzymes, transfer RNAs, ribosomes, factors, etc. (Crick et al. 1976).*

In the previous sections, we learned about the central importance of the homology search engines such as RISC, the RT/RNaseH strand transfer mechanism working in trembling, flipping, sliding motions, and the SDSA-step in group II intron homing. SDSA-like templating, annealing, and priming appears to be the universal ingredient of replicating systems taking advantage of perhaps initially random complementarity in RNA sequences, annealing different molecules to each other, serving as informational backups of each other, and continuing in the production of either random RNA chain assemblies (Fig. 11.11a, b right half of abscissa) or of Darwinian, master-governed mutant assemblies, termed *quasispecies* and called to be alive (Fig. 11.11a, b left half of abscissa).

The specific properties of any kind of such RNA(P) molecule assemblies further remain a mystery. However, game theory teaches that some sort of evolutionary stable strategy (ESS) coevally considering Nash-equilibria must have been reached once or at multiple occasions (Maynard Smith and Price 1973; Nash 1994). The RNA(P) assembly of molecules would be but one example of such equilibrated systems. Any system of this kind, replicating and encoding intrinsic genetic information[21] that was stable for enough generations to serve as a unit of selection would have been alife and simultaneously fitted the anthropic, *deep definition* of the term *gene*.[22] Game theory provides a conceptional bridge and is relevant for these initial beginnings of life to the highest complexities, e.g., as in the *prisoners' dilemma* or as for the life strategies of social superorganisms such as the honeybee.

Thus, game theory represents a well tied Ariadne's thread throughout life sciences. But finding a concrete, uninterrupted, contiguous path from a first ribozyme-activity of a replicator-*quasispecies* to the (theoretical) protoribosome (Davidovich et al. 2009; Yonath 2009b, 2009c) calls for bridging substantial leaps of another kind. Only few theoretical and fossil threads remain to conceptualize such a reasonable bridge. The only hint for such a link was motivated by Crick's et al. recognition that the anticodon loop of present tRNAs encompasses a consensus sequence motive, which reads 3'NRαβμUY, where αβμ is what we call the modern anticodon (Crick et al. 1976). One possible explanation for the universal

[21] i.e., informational replicators as defined (Zachar and Szathmary 2010) with the slight difference that the most basic replicator representing life (i.e., the *gene* as defined in footnote 22) itself manifests a phenotype (secondary structure) as well.

[22] *A gene is defined as any portion of chromosomal material that potentially lasts for enough generations to serve as a unit of selection.* G.C. Williams cited in Dawkins (1976) – where a unit of selection is any ESS in Nash-equilibrium. This gene definition is more general and it embraces the classic, technical Beadle-Tatum definition of gene (one gene, one enzyme hypothesis) used in everyday laboratory practice. The gene as a physical reality was recognized by physicists only after Muller's demonstration that radiation triggers mutations in DNA visible as a changed phenotype (Muller 1941).

conservation of this motive is the stereo chemical theory in contrast to the frozen accident theory (Crick 1968). From the motive, these and other authors deduced an ancient RNY code on what would become mRNA later (see below). Today, these inferences, still provide the best hypothetical link between the ancient and the current RNA world. At present, the most complete study to address the path from random to living replicator-quasispecies (e.g., in ZnS micro compartments) to the protoribosome still is the analysis by Eigen and colleagues (Eigen and Schuster 1977, 1978a, b). As the first replicating RNA(P) entities did certainly not exist alone in the vastness of primordial earth's geological environments, *quasispecies* likely exchanged molecular mutants within or between entitary compartments (ZnS/MnS or other analogous ones) and those molecules must have interacted with each other by random complementarities. Therefore, the authors developed the concept of interacting and interdependent hypercycles, representing ESSs as mentioned above. There is no space here to go deep into this matter but the following remark of Friedrich Cramer hits the bull's-eye best:

> *For the understanding of the mechanism of evolution, the theory of hypercycles is comparably the same as what is the quantum mechanics* (and I would rather add quantum electrodynamics (QED)[23]) *for the physics of elementary processes* (transl. from Cramer 1989).

As the theoretical protoribosome had to be replicated under the rules of the *quasispecies* concept, it could have been a part of some kind of hypercycle itself. Several interacting RNA(P) molecules took part as parasites or symbionts in replication cycles in analogy to Spiegelman's monster or functionally as ribozyme-proto-chaperones analogous to modern peptide chaperones (Karbstein 2010). Thus, molecular drivers (Dover 1986; Dover et al. 1982; Flegr 2002; Strachan et al. 1985) were there from the beginning of life with "transposons" as their modern descendants (Lankenau and Volff 2009).

In 1976, Crick et al. published a paper on the origin of protein synthesis (Crick et al. 1976). The authors based their argument only on the assumption, that originally no ribosome at all was necessary and that the ordering of amino acids in protein synthesis was accomplished using only "mRNA" and a few primitive tRNAs. The idea helped to blaze the trail for the modern protoribosome idea (Davidovich et al. 2009). However, the deep key question raised by these classical think tanks was the question how initially do tRNA like molecules (as emerged from ZnS/quasispecies world origins) anneal to mRNA like molecules and then to move the mRNA further with melting and annealing dynamics of incoming and outgoing proto-tRNAs. May be or may not be that tRNA like molecules such as Noller's duplicator RNA (dRNA) monomers (Noller 2011) participated in the molecular mechanics of *quasispecies* RNA(P) replication. Important is that at some point, proto-tRNA anticodon loops did "learn" to anneal with "mRNA" monomers. (The causal reason does not matter primarily.) Protoribosomal

[23]Bracketed note, added by D.-H.L related to Feynman (1985).

ribozymes(P)[24], similar but smaller than contemporary 23S and 28S rRNAs, joined into these annealing mechanistics. Aminoacyl tRNAs, i.e., tRNAs that were associated with an amino acid – having some stereochemic stability property – joined as well and linked two different amino acids in a ribozyme-catalytic process for reasons we do not know. In modern ribosomes, two tRNAs bind with their anticodon loop to two adjacent triplets of the mRNA termed aminoacyl (A) and peptidyl (P) site and perform a translocation step. This A- to P-site tRNA translocation comprises two highly correlated motions, i.e., a sideways shift and a ribosomal navigated rotatory motion (Yonath 2009b). According to Crick's et al. argument early on, there was only a skeletal scaffold without all the proteins added during 4 Ga of evolution. The old pioneers came up with an alternative, much less sophisticated primordial mechanism for translocation along a comma-free read-off (Crick 1968; Crick et al. 1976; Eigen and Schuster 1978b; Woese 1970). Their idea was based on perfectly plausible, causal inferences: A primordial code, firmly rooted in an *quasispecies*-replicator world, must have had a frame structure, analogous to the modern triplet, otherwise any germinating message could not have been read off uninterruptedly. They proposed a particular base sequence to which all codons had to adhere. Crick et al. recognized a consensus sequence regularity found in the anticodon loop of present tRNAs, which reads 3'NRαβμUY, where αβμ is what we call the modern anticodon. N designates any nucleotide, R and Y stand for purine and pyrimidine. Another requirement of ribosome-free or protoribosome-led translation was the stability of the evolving machine. Until the "complete" message is translated, the peptidyl-t-RNA must not fall off before the translocation of the subsequent aminoacyl-tRNA is accomplished. The stereochemistry with the two t-RNA molecules (peptidyl and aminoacyl tRNAs) each binding to an mRNA via dynamic 7 bp interactions involving the ancient anticodon loop consensus is shown in Figs. 11.12 and 11.13. In this model (Fig. 11.12, slightly altered after Woese (Woese 1970)), the mRNA meanders through two sterically antagonistic anticodon loop conformations (FH and fh) of the two tRNAs binding not to three codon-nucleotides as a triplet but to five complementary nucleotides within each anticodon loop. Note that, according to the current structure of the modern ribosome, tRNA anticodon loops only complement with three nucleotides of the mRNA and the decoding translocation mechanics is highly dependent on higher order ribosomal structures (Ramakrishnan 2011).

Because of the crystallography data explaining the movement of the decoding machine, it seems futile to discuss Crick's et al. RNY code mechanism any further. The 5 nt tRNA/mRNA codon interactions are just hot air (Ramakrishnan 2011). Nevertheless, statistical analysis of hundreds of modern genes revealed that it is

[24]P designates the option for the association of a random peptide forming random, primordial RNP chromatin.

always the coding strand with the open reading frame (ORF) that requires the lowest number of computational back mutations in order to reinstall a comma free RNY pattern. Figure 11.13e, f show two examples of such an RNY rhythm analysis. No matter any criticism that such a pattern might be a statistical artifact or a statistical correlation based on codon bias phenomena, it is stunning that Crick et al. posited their model *before* the RNY pattern in modern ORFs was discovered (Lankenau 1990; Shepherd 1981a, b). Thus, even with our modern knowledge of a structure based ribosome function, the RNY code hypothesis is still reasonable and would call for being replaced by an alternative better one if that existed. Thus, from our present speculations about the ancient RNA world, the RNY code hypothesis still is the best we can derive in a bottom to top approach.

However, what role does the protoribosome (Davidovich et al. 2009; Yonath 2009a, 2009b, 2009c) play in this RNY-code world. Figure 11.13a–d merely explains the translocation by sterical conformation shifts (hf<=>FH) of tRNA molecules along an mRNA. Thereby, it consents tacitly that the in-brought tRNA-coupled amino acids are "somehow" connected to form a polypeptide chain. Nothing is said about the peptidyl transferase step. At this stage, however, the protoribosome enters. In this hypothesis, the protoribosome is an additional ribozyme *quasispecies* with polymerization capabilities. In association with the Crick-Brenner- Klug-Pieczenik mechanism (Crick et al. 1976) (Fig. 11.13a–d), the protoribosome was complexed with amino acid-coupled tRNAs (Joyce 2002) and mRNAs (Fig. 11.12) and acted as the primordial translation machine. Only if a combination of:

1. Navigated annealing (as in SDSA) and decoding,
2. Conformation-change based translocation, and
3. Peptidyltransfer

was accomplished in the same complex could translation have emerged.

The theoretical power of Eigen's and colleague's coexistent, interdependent *quasispecies* and *hypercycles*, perhaps thriving inside Mulkidjanian-ZnS compartments, truly reflects a fundamental theory similarly important for life sciences as what QED is for the physics of elementary processes. Joyce and Orgel's note that RNA is a prebiotic chemist's nightmare (Joyce and Orgel 1999) may be true, but the frameworks concerning two RNA worlds set up to explore the obstacles of the primal synthesis (Eigen 1971; Wächtershäuser 1997) (Fig. 11.1) stand and thrive firmly.

11.4 From Grassroots Level Back to and Beyond the Golden Spike

The majority of sequences comprising modern genomes express ncRNAs that are now recognized to act as sensors, integrators, catalysts, defenders in a universe of possible metabolic processes (Fig. 11.2) (Wang et al. 2011). The transposon world

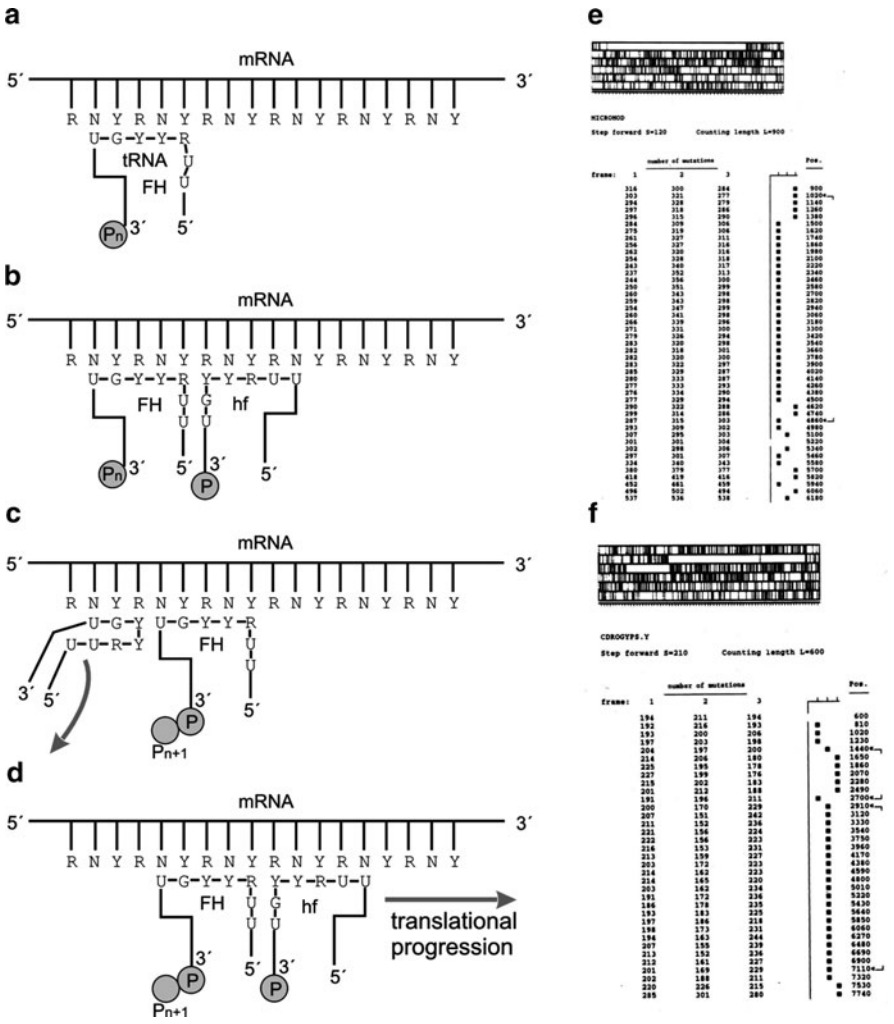

Fig. 11.13 The RNY code hypothesis. (**a**–**d**) Decoding and conformation driven, translational progression along a comma free RNY code in mRNA. The pre-protoribosomal mechanism requires annealing of the tRNA anticodon loop in the hf conformation (see Fig. 11.12). Movement along mRNA is driven by conformational flips of the tRNA loop (hf <=>FH). (**e**, **f**) Identification of the RNY code in modern genes – two examples. Results of only three of the six possible frames are shown. (**e**) RNY pattern analysis of the six reading frames of the *micropia* retrotransposon. Stop codons indicated as bars. RNY is only associated with the long open reading frame (ORF) encoding RT/RNaseH. (**f**) Positive correlation of RNY with the gag and pol ORFs of the proviral retrotransposon *gypsy*. This kind of correlation is ubiquitously conserved in all genes analyzed so far. Method as described in Lankenau (1990) and Shepherd (1981a, b)

is a driving part of it. This is a reflection of the primal beginnings 4 Ga ago when genetic information emerged. The ancient, short RNA monomers and first short RNA chains certainly interacted with many different prebiotic metabolites where amino acids or even short, random proto-peptides must have interacted with RNA. But only when translation – by whatever geo-environmental, chemical, physical, or mechanistic constraints – was established, the tight interaction of proteins and nucleic acids emerged together as chromatin, transporting the genetic information in an uninterrupted lineage (i.e., line of "clones" as in prokaryotes or germlines as in metazoans) to subsequent generations and protecting it by DNA repair.

The phylogenetic tree of life was reconstructed mainly based on ribosomal gene information. Here, the ribosomal 16S rRNA played a major role in assessing phylogenetic relationships (e.g., Woese et al. 1990; Yang et al. 1985). However, the major goal of evolutionary research is not tree-reconstruction but to elucidate the driving mechanisms of evolution. The ribosome is a true molecular living fossil. Charles Darwin did not like living fossils for deciphering evolutionary change, as they do not evolve significantly enough.[25] The same holds true for 16S rRNA and other ribosomal components symbolizing the "*Latimeria* of the molecular living-fossils." Even though for tree of life reconstructions rRNAs did a great job, they did not evolve speedily enough to enlighten us deeply to understand the major, cutting edge transitions (including promiscuous events) as rooted in particular basic causes and mechanisms of evolutionary change. For example, the evolution of HOX genes gave insight into major evolutionary transitions of metazoans (Akam et al. 1994; Gehring 1998; Holland and Garcia-Fernandez 1996; Valentine et al. 1996). However, mediating both, significant evolutionary changes and evolutionary driving forces throughout all kingdoms of life and spanning the entire deep history of life is the realm of RRR, especially of the homology search engine of SDSA. Therefore, as Ernst Mayr once said, it is these basic causes and mechanisms that are the central themes of evolutionary biology (Mayr 1963).

As shown here, (proto)ribosomes were a part of an ancient RNA world where other RRR-factories were as old as the protoribosome or older (Eigen's *quasispecies* and *hypercycles*). Thus, RRR-factory related phenomena were the earliest actors of darwinian evolution. As ingredients of their very own geochemical emergence these reproducing molecule assemblies used homology search processes conceptualized with techniques like FRET (Abd..et al. 2008) to form the anthropic basis the anthropic basis for research in moving the science from the general to the particulars, and back to the general, where moving back means: following Popper's deductive cycle from bottom to top, beyond the golden spike into the complex worlds of replication, and DNA-repair mechanisms (i.e., NHEJ, SDSA, SSA, BIR) (Lankenau 2006; Wächtershäuser 1997) Recently, Marcel Weber developed the thesis of "causal specificity" which further strengthens our concept here (Weber 2006).

[25]Darwin refers to living fossils as "*anomalous forms*" or "*wrecks of ancient life*" saved from competition and extermination (see also Fricke 2010, pp. 81–82).

To explore modern RNA world interactions with modern, chromatin embedded DNA repair will be a lead for forthcoming research. We now know (and re-experience) that attempts to understand chromatin and DNA repair had a difficult start as Francis Crick remarked: "*We totally missed the possible role of . . . repair. I came later to realize that DNA is so precious that probably many distinct repair mechanisms would exist. Nowadays one could hardly discuss mutation without considering repair at the same time*" (Crick 1974). Also Roger Kornberg reports on difficulties conceptualizing DNA-repair related processes: " '*I'll bet you a bottle of champagne' said Buzz Baldwin, a colleague in my department, 'that the nuclease activity in DNA polymerase is part of the enzyme.' The preparation of the replicating enzyme we had purified extensively could still degrade DNA chains. In the absence of nucleotide building blocks needed for synthesis, nucleotides were cleaved slowly and serially from DNA. I took Baldwin's bet because it made no sense to me at the time that DNA polymerase would degrade the very end of the chain it would normally be extending*" Kornberg (1989). So the RRR properties of life embedded inside molecular 3D networks of chromatin, membrane enveloped nuclei and cells, and even up to replication phenomena of superorganisms such as ants, termites, and honey bees as spearheads of ecology provide one contiguous trajectory throughout evolution, with enough change at hand for each step and enough depth of conserved traits to represent a current theory of evolutionary synthesis.

Acknowledgments I dedicate this chapter to father and son W. Hennig, i.e., Willi Hennig and Wolfgang Hennig – the latter celebrating his 70th birthday this year. I apologize to all colleagues whose key contributions were not cited due to space restriction and focus. I am grateful to Peter Vogt for pointing out to me the original RNY-code concept, to Manfred Eigen for discussing the topic and the quasispecies concept with me a long time ago, and to Carsten Lankenau for his engaged help in writing the RNY pattern analysis software during the 1980s. I also thank Victor Corces and Bill Engels for having me in their labs in Baltimore and Madison where the antisense story and the SDSA connections were established. Many thanks also go to Armen Mulkidjanian for accompanying me while learning about his ZnS theory, and to Susanne Lankenau and Richard Egel for engaged discussions and for comments on the manuscript.

References

Abbondanzieri EA, Bokinsky G, Rausch JW, Zhang JX, Le Grice SF, Zhuang X (2008) Dynamic binding orientations direct activity of HIV reverse transcriptase. Nature 453:184–189

Adams SE et al (1987) The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. Cell 49:111–119

Akam M, Averof M, Castelli-Gair J, Dawes R, Falciani F, Ferrier D (1994) The evolving role of Hox genes in arthropods. Development Suppl:209–215

Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318:761–764

Arnold E et al (1992) Structure of HIV-1 reverse transcriptase/DNA complex at 7A resolution showing active site locations. Nature 357:85–89

Ashkenasy G, Jagasia R, Yadav M, Ghadiri MR (2004) Design of a directed molecular network. Proc Natl Acad Sci USA 101:10872–10877
Bailey JA, Carrel L, Chakravarti A, Eichler EE (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc Natl Acad Sci USA 97:6634–6639
Baross J et al (2007) The limits of organic life in planetary systems. The National Academies Press, Washington DC
Barrow JD, Tipler FJ (1986) The anthropic cosmological principle. Oxford University Press, Oxford
Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. Science 261:1411–1418
Beadle GW (1958) Genes and chemical reactions in Neurospora. Nobel Foundation, Stockholm, pp 587–599
Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM (2002) Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, pp 761–783
Belozersky AN (1957) On species specificity of nucleic acids in bacteria. In: Oparin AIea (ed) Vozniknovenie zhizni na Zemle (The origin of life on Earth). Akad. Nauk. SSSR, Moscow, pp 198–205
Belozersky AN (1959) On species specificity of nucleic acids in bacteria. In: Oparin AI, Pasynskii AG, Braunshtein AE, Pavlovskaya TE, Clark F, Synge RLM (eds) The origin of life on the Earth. Pergamon, London, pp 322–331
Benner SA, Kim HJ, Yang Z (2011) Setting the stage: the history, chemistry, and geobiology behind RNA. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 7–19
Benner SA, Sismour AM (2005) Synthetic biology. Nat Rev Genet 6:533–543
Berg DE, Howe MM (eds) (1989) Mobile DNA I. American Society of Microbiology, Washington DC
Biebricher CK, Eigen M (2006) What is a quasispecies? In: Domingo E (ed) Quasispecies: concepts and implications for virology. Springer, Heidelberg/New York, pp 1–31
Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J 17:170–180
Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. Nature 457:977–980
Bolli M, Micura R, Eschenmoser A (1997) Pyranosyl-RNA: chiroselective self-assembly of base sequences by ligative oligomerization of tetranucleotide-2′,3′-cyclophosphates (with a commentary concerning the origin of biomolecular homochirality). Chem Biol 4:309–320
Branciamore S, Gallori E, Szathmary E, Czaran T (2009) The origin of life: chemical evolution of a metabolic system in a mineral honeycomb? J Mol Evol 69:458–469
Brennecke J et al (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128:1089–1103
Britten RJ, Davidson EH (1969) Gene regulation for higher cells: a theory. Science 165:349–357
Burt A, Trivers R (2006) Genes in conflict. The Belknap Press of Harvard University Press, Cambridge
Capy P, Bazin C, Higuet D, Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience and Chapman & Hall, Georgetown
Cech TR (1986) A model for the RNA-catalyzed replication of RNA. Proc Natl Acad Sci USA 83:4360–4363
Cech TR (2009) Evolution of biological catalysis: ribozyme to RNP enzyme. Cold Spring Harb Symp Quant Biol 74:11–16
Charlebois RL, Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res 14:2469–2477

Chaumeil J, Le Baccon P, Wutz A, Heard E (2006) A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev 20:2223–2237
Chow JC, Yen Z, Ziesche SM, Brown CJ (2005) Silencing of the mammalian X chromosome. Annu Rev Genomics Hum Genet 6:69–92
Cleaves HJ, Miller SL (2001) The nicotinamide biosynthetic pathway is a by-product of the RNA world. J Mol Evol 52:73–77
Cnossen I, Sanz-Forcada J, Favata F, Witasse O, Zegers T, Arnold NF (2007) Habitat of early life: solar X-ray and UV radiation at Earth's surface 4–3.5 billion years ago. J Geophys Res 112: E02008
Coffin JM, Hughes SH, Varmus HE (1997) Retroviruses. Cold Spring Harbor Laboratory Press, New York
Colot V, Maloisel L, Rossignol JL (1996) Interchromosomal transfer of epigenetic states in Ascobolus: transfer of DNA methylation is mechanistically related to homologous recombination. Cell 86:855–864
Cooper GM, Temin RG, Sugden B (eds) (1995) The DNA provirus: Howard Temin's scientific legacy. ASM Press, Washington DC
Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890
Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) (2002) Mobile DNA II. ASM Press, Washington
Cramer F (1989) Chaos und Ordnung, 2nd edn. Deutsche Verlags-Anstalt GmbH, Stuttgart
Crick F (1974) The double helix: a personal view. Nature 248:766–769
Crick FH (1968) The origin of the genetic code. J Mol Biol 38:367–379
Crick FH, Brenner S, Klug A, Pieczenik G (1976) A speculation on the origin of protein synthesis. Orig Life 7:389–397
Davidovich C, Belousoff M, Bashan A, Yonath AE (2009) The evolving ribosome: from non-coded peptide bond formation to sophisticated translation machinery. Res Microbiol 160:487–492
Davidson EH (1986) Gene activity in early development, 3rd edn. Academic, Orlando
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Dawkins R (2004) The ancestor's tale: a pilgrimage to the dawn of evolution. Houghton Mifflin, Boston
Dawkins R (2006) The God delusion. Transworld Publishers – Black Swan ed, London
Dettai A, Volff JN (2009) Morphological Characters from the Genome: SINE Insertion Polymorphism and Phylogenies. In: Lankenau D-H, Volff JN (eds) Transposable Elements and Genome Dynamics, vol 4. Springer, Heidelberg, pp 45–75
Domingo E, Flavell RA, Weissmann C (1976) In vitro site-directed mutagenesis: generation and properties of an infectious extracistronic mutant of bacteriophage Qbeta. Gene 1:3–25
Doolittle RF (2000) Searching for the common ancestor. Res Microbiol 151:85–89
Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2129
Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165
Dover GA, Brown S, Coen E, Dallas J, Strachan T, Trick M (1982) The dynamics of genome evolution and species differentiation. In: Dover GA, Flavell RB (eds) Genome evolution. Academic, London, pp 343–372
Dyson FJ (1999) Origins of life, Revth edn. Cambridge University Press, Cambridge/New York
Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58:465–523
Eigen M (1983) Self-replication and molecular evolution. In: Bendall DS (ed) Evolution from molecules to men. Cambridge University Press, Cambridge, pp 106–130
Eigen M (1987) Stufen zum Leben. Piper, Muenchen/Zuerich
Eigen M (1988) Perspektiven der Wissenschaft. DVA, Stuttgart

Eigen M (1992) Steps towards life: a perspective on evolution. Oxford University Press, Oxford
Eigen M, Lindemann B, Winkler-Oswatitsch R, Clarke CH (1985) Pattern analysis of 5s rRNA. Proc Natl Acad Sci 82:2437–2441
Eigen M, McCaskill J, Schuster P (1988) Molecular quasi-species. J Phys Chem 92:6881–6891
Eigen M, Schuster P (1977) The hypercycle. A principal of natural self organisation. Part A: emergence of the hypercycle. Naturwissenschaften 64:541–565
Eigen M, Schuster P (1978a) The hypercycle. A principal of natural self organisation. Part B: the abstract hypercycle. Naturwissenschaften 65:7–41
Eigen M, Schuster P (1978b) The hypercycle. A principle of natural self-organisation. Part C: the realistic hypercycle. Naturwissenschaften 65:341–369
Ekland EH, Bartel DP (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. Nucleic Acids Res 23:3231–3238
Ekland EH, Bartel DP (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature 382:373–376
Ekland EH, Szostak JW, Bartel DP (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. Science 269:364–370
Ellis EL, Delbrück M (1939) The growth of bacteriophage. J Gen Physiol 22:365–384
Engels WR (1997) Invasions of P elements. Genetics 145:11–15
Engels WR, Preston CR, Johnson-Schlitz DM (1994) Long-range cis preference in DNA homology search over the length of a Drosophila chromosome. Science 263:1623–1625
Epstein IR, Eigen M (1979) Selection and self-organization of self-reproducing macromolecules under the constraint of constant flux. Biophys Chem 10:153–160
Evgen'ev MB, Corces VG, Lankenau DH (1992) Ulysses transposable element of Drosophila shows high structural similarities to functional domains of retroviruses. J Mol Biol 225:917–924
Evgen'ev MB et al (1997) Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci USA 94:196–201
Faehnle CR, Joshua-Tor L (2007) Argonautes confront new small RNAs. Curr Opin Chem Biol 11:569–577
Fedoroff N, Botstein D (1992) The Dynamic genome: Barbara McClintock's ideas in the century of genetics. Cold Spring Harbor Laboratory Press, New York
Ferris JP, Hill AR Jr, Liu R, Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. Nature 381:59–61
Feynman RP (1985) QED: the strange theory of light and matter. Princeton University Press, Princeton
Fire A (2006) Gene silencing by double stranded RNA. In: Nobel lecture, Stockholm,The Nobel Prizes". Nobelprize.org. http://nobelprize.org/nobel_prizes/ pp 198–233,
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391:806–811
Flavell AJ, Ish-Horowicz D (1983) The origin of extrachromosomal circular copia elements. Cell 34:415–419
Flegr J (2002) Evolutionary mechanisms and processes. In: Encyclopedia of life support systems (EOLSS). Eolss Publishers, Oxford
Fricke H (2010) Der Fisch, der aus der Urzeit kam – Die Jagd nach dem Quastenflosser. Deutscher Taschenbuch Verlag, München
Friedmann N, Miller SL, Sanchez RA (1971) Primitive earth synthesis of nicotinic acid derivatives. Science 171:1026–1027
Fukuda T et al (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. Nat Cell Biol 9:604–611
Garfinkel DJ, Boeke JD, Fink GR (1985) Ty element transposition: reverse transcriptase and virus-like particles. Cell 42:507–517
Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425:285–288

Gebinoga M, Oehlenschläger F (1995) Comparison of self-sustained sequence replication reaction systems. Eur J Biochem 235:256–261
Gehring WJ (1998) Master control genes in development and evolution. Yale University Press, New Haven
Gibson DG et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56
Gilbert W (1986) The RNA world. Nature 319:618
Gloor GB, Lankenau D-H (1998) Gene conversion in mitotically dividing cells: a view from Drosophila. Trends Genet 14:43–46
Gould SJ (1997) The exaptive excellence of spandrels as a term and prototype. Proc Natl Acad Sci USA 94:10750–10755
Gould SJ (2002) The structure of evolutionary theory. Belknap Press of Harvard University Press, Cambridge
Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell 123:631–640
Grew ES, Bada JL, Hazen RM (2011) Borate minerals and origin of the RNA world. Orig Life Evol Biosph, DOI 10.1007/s11084-010-9233-y
Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35:849–857
Guo H, Zimmerly S, Perlman PS, Lambowitz AM (1997) Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. EMBO J 16:6835–6848
Haber JE (2008) Evolution of models of homologous recombination. In: Egel R, Lankenau D-H (eds) Recombination and meiosis – models, means and evolution, vol 3. Springer, Berlin Heidelberg, pp 1–64
Hawking S, Mlodinow L (2010) The grand design, 1st edn. Bantam Books, New York
Hazen RM (2005) Genesis: the scientific quest for life's origins. Joseph Henry Press, Washington
Hazen RM (2010) Evolution of minerals. Sci Am 302:58–65
Hennig W (1950) Grundzüge einer Theorie der phylogenetischen Systematik. Deutscher Zentralverlag, Berlin
Hennig W (1966) Phylogenetic systematics. University of Illinois Press, Urbana/Chicago/London
Hennig W et al (1989) Y chromosomal fertility genes of Drosophila: a new type of eukaryotic genes. Genome 31:561–571
Holland PWH, Garcia-Fernandez J (1996) Hox genes and chordate evolution. Dev Biol 173:382–395
Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282:1669–1675
Huijser P, Kirchhoff C, Lankenau DH, Hennig W (1988) Retrotransposon-like sequences are expressed in Y chromosomal lampbrush loops of Drosophila hydei. J Mol Biol 203:689–697
Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510
Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science 292: 1319–1325
Jordan IK et al (2005) A universal trend of amino acid gain and loss in protein evolution. Nature 433:633–638
Jordan IK, Miller WJ (2008) Genome defense against transposable elements and the origins of regulatory RNA. In: Volff J-N, Lankenau D-H (eds) Transposons and genome dynamics, vol 4. Springer, Heidelberg, pp 77–94
Jore MM, Brouns SJJ, van der Oost J (2011) RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds:

from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 231–242
Joshua-Tor L, Hannon GJ (2011) Ancestral roles of small RNAs: an ago-centric perspective. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA world's: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 243–253
Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–221
Joyce GF, Orgel L (1999) Prospects for understanding the origin of the RNA world. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA world, 2nd edn. Cold Spring Harbor Laboratory Press, New York, pp 49–77
Kacian DL, Mills DR, Kramer FR, Spiegelman S (1972) A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. Proc Natl Acad Sci USA 69:3038–3042
Karbstein K (2010) Chaperoning ribosome assembly. J Cell Biol 189:11–12
Kauffman SA (1993) The origins of order: self organization and selection in evolution. Oxford University Press, New York
Kelley DS et al (2005) A serpentinite-hosted ecosystem: the lost city hydrothermal field. Science 307:1428–1434
Khersonsky O, Roodveldt C, Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. Curr Opin Chem Biol 10:498–508
Kidwell MG, Kidwell JF, Sved JA (1977) Hybrid dysgenesis in Drosophila melanogaster:a syndrome of aberant traits including mutation, sterility, and male recombination. Genetics 86:813–833
Kikuchi Y, Ando Y, Shiba T (1986) Unusual priming mechanism of RNA-directed DNA synthesis in copia retrovirus-like particles of Drosophila. Nature 323:824–826
Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256:1783–1790
Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1:127–136
Koonin EV (2006a) On the origin of cells and viruses: a comparative-genomic perspective. Isr J Ecol Evol 52:299–318
Koonin EV (2006b) Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases. Biol Direct 1:39
Koonin EV (2007) The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life. Biol Direct 2:15
Koonin EV (2009) On the origin of cells and viruses: primordial virus world scenario. Ann NY Acad Sci 1178:47–64
Koonin EV, Aravind L, Kondrashov AS (2000) The impact of comparative genomics on our understanding of evolution. Cell 101:573–576
Koonin EV, Chumakov KM, Agol VI (1980) A comparative study on the UV resistance of double-stranded and single-stranded encephalomyocarditis virus RNAs – evaluation of the possible contribution of host-mediated repair. J Gen Virol 49:437–441
Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. Trends Genet 21:647–654
Koonin EV, Novozhilov AS (2009) Origin and evolution of the genetic code: the universal enigma. IUBMB Life 61:99–111
Koonin EV, Senkevich TG, Dolja VV (2006) The ancient virus world and evolution of cells. Biol Direct 1:29
Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36:6688–6719
Kornberg A (1989) For the love of enzymes. Harvard University Press, Cambridge/London
Kornberg A, Baker TA (1992) DNA replication. W.H. Freeman & Co., New York

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell 31:147–157

Lambowitz AM, Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 103–121

Lankenau D-H (1990) Molecular structure and evolution of a retrotransposon family in Drosophila. Department of Human Genetics, Nijmegen, Nijmegen, pp 1–167

Lankenau D-H (1995) Genetics of genetics in Drosophila: P elements serving the study of homologous recombination, gene conversion and targeting. Chromosoma 103:659–668

Lankenau D-H (1999) DNA-Reparatur und Transposons in der Drosophila Keimbahn (Habilitation). In: DKFZ (ed) Universität Heidelberg, Ladenburg

Lankenau D-H (2006) Germline double-strand break repair and gene targeting in Drosophila: a trajectory system throughout evolution. In: Lankenau D-H (ed) Genome integrity: facets and perspectives, vol 1. Springer, Berlin Heidelberg, pp 153–197

Lankenau D-H (2007) The legacy of the germ line – maintaining sex and life in metazoans: cognitive roots of the concept of hierarchical selection. In: Egel R, Lankenau D-H (eds) Recombination and meiosis – models, means and evolution, vol 3. Springer, Berlin Heidelberg, pp 289–339

Lankenau D-H, Gloor GB (1998) In vivo gap repair in Drosophila: a one-way street with many destinations. Bioessays 20:317–327

Lankenau D-H, Peluso MV, Lankenau S (2000) The Su(Hw) chromatin insulator protein alters double strand-break repair frequencies in the Drosophila germ line. Chromosoma 109:148–160

Lankenau D-H, Volff J-N (2009) Transposons and the dynamic genome. Springer, Heidelberg

Lankenau DH, Corces VG, Engels WR (1996) Comparison of targeted-gene replacement frequencies in Drosophila melanogaster at the forked and white loci. Mol Cell Biol 16: 3535–3544

Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W (1988) Micropia: a retrotransposon of Drosophila combining structural features of DNA viruses, retroviruses and non-viral transposable elements. J Mol Biol 204:233–246

Lankenau S (1996) Expressionsanalyse des micropia Retrotransposons bei Drosophila. In: DKFZ, vol. Dr. rer. nat. Johns Hopkins University/University of Madison, Wisconsin/Universität Heidelberg, Heidelberg, p 110

Lankenau S, Barnickel T, Marhold J, Lyko F, Mechler BM, Lankenau DH (2003) Knockout targeting of the Drosophila Nap1 gene and examination of DNA repair tracts in the recombination products. Genetics 163:611–623

Lankenau S, Corces VG, Lankenau DH (1994) The Drosophila micropia retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. Mol Cell Biol 14:1764–1775

Lee JT (2011) The X as model for RNA's niche in epigenomic regulation. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA world's: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 267–278

Leipe DD, Aravind L, Koonin EV (1999) Did DNA replication evolve twice independently? Nucleic Acids Res 27:3389–3401

Liberles DA (ed) (2007) Ancestral sequence reconstruction. Oxford University Press, Oxford

Liu S, Abbondanzieri EA, Rausch JW, Le Grice SF, Zhuang X (2008) Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. Science 322:1092–1097

Liu S, Harada BT, Miller JT, Le Grice SF, Zhuang X (2010) Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. Nat Struct Mol Biol 17:1453–1460

Lohrmann R, Orgel LE (1977) Reactions of adenosine 5'phosphorimidazolide with adenosine analogs on a polyuridylic acid template. J Mol Biol 113:193–198

Lovell S, Goryshin IY, Reznikoff WR, Rayment I (2002) Two-metal active site binding of a Tn5 transposase synaptic complex. Nat Struct Biol 9:278–281
Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605
Lyon MF (2000) LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? Proc Natl Acad Sci USA 97:6248–6249
Martin W, Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. Nature 440:41–45
Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc Lond B Biol Sci 358:59–83; discussion 83–55
Matsuura M et al (1997) A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. Genes Dev 11:2910–2924
Matzke MA, Birchler JA (2005) RNAi-mediated pathways in the nucleus. Nat Rev Genet 6:24–35
Matzke MA, Matzke AJ (2004) Planting the seeds of a new paradigm. PLoS Biol 2:E133
Matzke MA, Mette MF, Matzke AJ (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Mol Biol 43:401–415
Maynard Smith J, Price GR (1973) The logic of animal conflict. Nature 246:15–18
Maynard Smith J, Szathmary E (1997) The major transitions in evolution. Oxford University Press, Oxford
Mayr E (1963) Animal species and evolution. Belknap Press of Harvard University Press, Cambridge
McClure MA (1991) Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol Biol Evol 8:835–856
McClure MA (1993) Evolutionary history of reverse transcriptase. In: Skalka AM, Goff S (eds) Reverse transcriptase. Cold Spring Harbor Laboratory Press, New York, pp 425–444
Mellor J et al (1985) Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. Nature 318:583–586
Meselson M, Stahl FW (1958a) The replication of DNA. Cold Spring Harb Symp Quant Biol 23:9–12
Meselson M, Stahl FW (1958b) The replication of DNA in Escherichia coli. Proc Natl Acad Sci USA 44:671–682
Miller SL, Urey HC (1959) Organic compound synthesis on the primitive earth. Science 130:245–251
Mochizuki K, Gorovsky MA (2004) Small RNAs in genome rearrangement in Tetrahymena. Curr Opin Genet Dev 14:181–187
Mohr G, Smith D, Belfort M, Lambowitz AM (2000) Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. Genes Dev 14:559–573
Moore PB, Steitz TA (2011) The roles of RNA in the synthesis of protein. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 123–139
Mulkidjanian AY (2009) On the origin of life in the zinc world: 1. Photosynthesizing, porous edifices built of hydrothermally precipitated zinc sulfide as cradles of life on Earth. Biol Direct 4:26
Mulkidjanian AY, Galperin MY (2009) On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on Earth. Biol Direct 4:27
Mulkidjanian AY, Galperin MY, Koonin EV (2009) Co-evolution of primordial membranes and membrane proteins. Trends Biochem Sci 34:206–215

Muller H (1941) Induced mutations in Drosophila. Cold Spring Harb Symp Quant Biol 9:151–167
Muller HJ (1966) The gene material as the initiator and organizing basis of life. Am Nat 100:493–517
Nash J (1994) The work of John Nash in game theory. In: Nobel Foundation (ed) Nobel seminar. Nobel Foundation, pp 160–190, Nobelprize.org. http://nobelprize.org/nobel_prizes/
Nassif N, Penney J, Pal S, Engels WR, Gloor GB (1994) Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol 14:1613–1625
Navarro P, Avner P (2010) An embryonic story: analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells. Bioessays 32:581–588
Naylor R, Gilham PT (1966) Studies on some interactions and reactions of oligonucleotides in aqueous solution. Biochemistry 5:2722–2728
Ng K, Pullirsch D, Leeb M, Wutz A (2007) Xist and the order of silencing. EMBO Rep 8:34–39
Nielsen H, Fiskaa T, Birgisdottir AB, Haugen P, Einvik C, Johansen S (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. RNA 9:1464–1475
Nisbet EG, Sleep NH (2001) The habitat and nature of early life. Nature 409:1083–1091
Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis [In Process Citation]. Science 289:920–930
Noller HF (2011) Evolution of protein synthesis from an RNA world. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 141–154
Noonan JP et al (2006) Sequencing and analysis of Neanderthal genomic DNA. Science 314:1113–1118
Nowotny M (2009) Retroviral integrase superfamily: the structural perspective. EMBO Rep 10:144–151
Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell 121:1005–1016
Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. Mol Cell 28:264–276
Oehlenschläger F, Eigen M (1997) 30 years later – a new approach to Sol Spiegelman's and Leslie Orgel's in vitro evolutionary studies. Dedicated to Leslie Orgel on the occasion of his 70th birthday. Orig Life Evol Biosph 27:437–457
Orgel LE (1968) Evolution of the genetic apparatus. J Mol Biol 38:381–393
Orgel LE (1979) Selection in vitro. Proc R Soc Lond B Biol Sci 205:435–442
Paques F, Leung WY, Haber JE (1998) Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol Cell Biol 18:2045–2054
Paul N, Joyce GF (2004) Minimal self-replicating systems. Curr Opin Chem Biol 8:634–639
Peliska JA, Benkovic SJ (1992) Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science 258:1112–1118
Perlman PS, Podar M (1996) Reactions catalyzed by group II introns in vitro. Methods Enzymol 264:66–86
Perriman R, Ares M Jr (1998) Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. RNA 4:1047–1054
Poisson SD (1837) Recherches sur la probalité des judgement en materière criminelles et en materière civile, précédées des règles générales du calcul des probalités. édition Bachelier, Paris, pp 1–415
Rajamani S et al (2009) Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. J Am Chem Soc 132:5880–5885
Ramakrishnan V (2011) The ribosome: some hard facts about its structure and hot air about its evolution. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 155–164
Redko Y, Bechhofer DH, Condon C (2008) Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in B. subtilis. Mol Microbiol 68:1096–1106

Rio DC (2002) P transposable elements in Drosophila melanogaster. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, pp 484–518
Robertson MP, Joyce GF (2011) The origins of the RNA world. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 21–42
Rosenberg UB, Preiss A, Seifert E, Jäckle H, Knipple DC (1985) Production of phenocopies by Krüppel antisense RNA injection into Drosophila embryos. Nature 313:703–706
Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM (1999) RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. Biochemistry 38:9069–9083
Saldanha R, Mohr G, Belfort M, Lambowitz AM (1993) Group I and group II introns. FASEB J 7:15–24
Sarot E, Payen-Groschene G, Bucheton A, Pelisson A (2004) Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics 166:1313–1321
Schmidt U, Podar M, Stahl U, Perlman PS (1996) Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing in vitro and in vivo: phenotypes and suppressor mutations. RNA 2:1161–1172
Shah SA, Garrett RA (2011) CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol 162:27–38
Sharp PA (1985) On the origin of RNA splicing and introns. Cell 42:397–400
Sharp PA (1991) Five easy pieces. Science 254:663
Shepherd JCW (1981a) Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc Natl Acad Sci USA 78:1596–1600
Shepherd JCW (1981b) Periodic correlations in DNA sequences and evidence suggesting their evolutionary origin in a comma-less genetic code. J Mol Evol 17:94–102
Sherratt DJ (1995) Mobile genetic elements. IRL Press at Oxford University Press, USA
Shiba T, Saigo K (1983) Retrovirus-like particles containing RNA homologous to the transposable element copia in Drosophila melanogaster. Nature 302:119–124
Shiu PK, Raju NB, Zickler D, Metzenberg RL (2001) Meiotic silencing by unpaired DNA. Cell 107:905–916
Sievers D, von Kiedrowski G (1994) Self-replication of complementary nucleotide-based oligomers. Nature 369:221–224
Simon DM, Kelchner SA, Zimmerly S (2009) A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. Mol Biol Evol 26:2795–2808
Skalka AM, Goff S (eds) (1993) Reverse transcriptase. Cold Spring Harbor Laboratory Press, New York
Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305:1434–1437
Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG (1994) An Env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev 8:2046–2057
Spirin AS (2005) The RNA world and its evolution. Mol Biol 39:550–556
Strachan T, Webb D, Dover G (1985) Transition stages of molecular drive in multiple-copy DNA families in Drosophila. EMBO J 4:1701–1708
Sugiyama T et al (2007) SHREC, an effector complex for heterochromatic transcriptional silencing. Cell 128:491–504
Sumper M, Luce R (1975) Evidence for de novo production of self-replicating and environmentally adapted RNA structures by bacteriophage Qbeta replicase. Proc Natl Acad Sci USA 72:162–166
Szathmary E (1997) Origins of life. The first two billion years. Nature 387:662–663
Tabara H et al (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell 99:123–132

Tammet D (2009) Embracing the wide sky. Free Press, New York/London/Toronto
Telesnitsky A, Goff SP (1997) Reverse transcriptase and the generation of retroviral DNA. In: Coffin JM, Hughes SH, Varmus HE (eds) Retroviruses. Cold Spring Harbor Laboratory Press, New York, pp 121–160
Temin HM (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc Natl Acad Sci 90:6900–6903
Terford A, Von Kiedrowski G (1992) Self-replication by condensation of 3-aminobenzamidines and 2-formylphenoxyacetic acids. Angew Chem Int Ed Engl 31:654–656
Tinoco I, Chen G, Qu X (2011) RNA reactions one molecule at a time. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 321–331
Tolia NH, Joshua-Tor L (2007) Slicer and the argonautes. Nat Chem Biol 3:36–43
Trinks H, Schröder W, Biebricher CK (2005) Ice and the origin of life. Orig Life Evol Biosph 35:429–445
Valentine JW, Erwin DH, Jablonski D (1996) Developmental evolution of metazoan bodyplans: the fossil evidence. Dev Biol 173:373–381
Varmus HE, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington DC, pp 53–108
Vastenhouw NL, Plasterk RH (2004) RNAi protects the Caenorhabditis elegans germline against transposition. Trends Genet 20:314–319
Venter JC et al (2001) The sequence of the human genome. Science 291:1304–1351
Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837
von Kiedrowski G (1986) A self-replicating hexadeoxynucleotide. Angew Chem Int Ed Engl 25:932–934
Voytas DF, Boeke JD (2002) Ty1 and Ty5 of *Saccharomyces cerevisiae*. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA 2. ASM Press, Washington DC, pp 631–662
Wächtershäuser G (1988) Pyrite formation, the first energy source for life: a hypothesis. Syst Appl Microbiol 10:207–210
Wächtershäuser G (1997) The origin of life and its methodological challenge. J Theor Biol 187:483–494
Walisko O, Jursch T, Izsvák Z, Ivics Z (2009) Transposon-Host Cell Interactions in the Regulation of Sleeping Beauty Transposition. In: Volff J-N, Lankenau D-H (eds) Transposons and Genome Dynamics, vol 4. Springer, Heidelberg, Berlin, pp 109–132
Wang X, Song X, Glass CK, Rosenfeld MG (2011) The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 279–292
Watson JD, Crick FH (1953a) Genetical implications of the structure of deoxyribonucleic acid. Nature 171:964–967
Watson JD, Crick FH (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–738
Weber M (2006) The Central Dogma as a Thesis of Causal Specificity. Hist. Phil. Life Sci. 28:595–610
White HB 3rd (1976) Coenzymes as fossils of an earlier metabolic state. J Mol Evol 7:101–104
Will CL, Lührmann R (2011) Spliceosome structure and function. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 181–203
Wills C, Bada J (2000) The spark of life: Darwin and the primeval soup. Perseus Publishing, Cambridge
Woese C (1970) Molecular mechanics of translation: a reciprocating ratchet mechanism. Nature 226:817–820

Woese C (1998) The universal ancestor. Proc Natl Acad Sci USA 95:6854–6859
Woese CR (1967) The genetic code: the molecular basis for genetic expression. Harper & Row, New York
Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090
Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proc Natl Acad Sci USA 87:4576–4579
Xiong Y, Eickbush TH (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol 5(6):675–690
Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9(10):3353–3362
Yang D, Oyaizu Y, Olsen GJ, Woese CR (1985) Mitochondrial origins. Proc Natl Acad Sci USA 82:4443–4447
Yang J, Mohr G, Perlman PS, Lambowitz AM (1998) Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity of aI1 and reverse splicing into DNA transposition sites in vitro. J Mol Biol 282:505–523
Yang W, Steitz TA (1995) Recombining the structures of HIV integrase, RuvC and RNase H. Structure 3:131–134
Yarus M (2011) Getting past the RNA world: the initial Darwinian ancestor. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York, pp 43–50
Yonath AE (2009) Ribosome: An Ancient Cellular Nano-Machine for Genetic Code Translation. In: Puglisi JD (ed) Biophysics and the Challenges of Emerging Threats. Springer Science Business Media B.V., pp 121–155
Yonath AE (2009a) Hibernating bears, antibiotics and the evolving ribosome. Nobel lecture. Nobelprize.org. http://nobelprize.org/nobel_prizes/
Yonath AE (2009b) Large facilities and the evolving ribosome, the cellular machine for genetic-code translation. J R Soc Interface 6(Suppl 5):S575–S585
Yonath AE (2009c) Ribosome: an ancient cellular nano-machine for genetic code translation. In: Puglisi JD (ed) Biophysics and the challenges of emerging threats. Springer, Netherlands, pp 121–155
Zachar I, Szathmary E (2010) A new replicator: a theoretical framework for analysing replication. BMC Biol 8:21