Ribosome Life

From RNA duplication to polypeptide translation & beyond

Edward M. Hedgecock & Rui B. Proença

Department of Biology

Johns Hopkins University

Baltimore MD 21218

United States

PDF 2024 February 5

Summary

We first apprise readers of two earlier proposals for the origin of the ribosome and transfer RNAs (tRNAs) in the RNA world as means of RNA copying (Campbell 1991; Noller 2012). Like the modern ribosome, the ancient *duplisome* used anticodons of donor RNAs (dRNAs) to read templates from 5’ to 3’ in one codon steps, adding a matching *duplicon* from each dRNA to the nascent polynucleotide copy. Noticing the defects that vitiated these two schemes called *twist replication* and *indirect duplication*, respectively, we introduce a simpler scheme called *direct duplication* that avoids their most serious defects, yet retains the virtues of both. In twist replication, the duplicon is a trinucleotide 3’ extension of the dRNA, while in direct duplication, the duplicon is a dinucleotide 5’ extension of the dRNA.

Our choice of the size and location of duplicons has a number of major consequences, and raises novel questions. We conjecture that dRNAs correspond to the 5’ half (nucleotides 1-36) of modern tRNAs … adopted an extended DSL to pair the duplicon and anticodon during loading, and an open conformation to … during decoding …. We conjecture that dRNAs were loaded from random oligos by ribozyme P, the ancestral RNA of modern RNase P, working alternately as a ligase to charge free dRNAs, and an endonuclease to trim the excess leader. Thus, duplicon-anticodon pairing explains the fidelity of loading, and a two-step discrimation and explains a kinetic proof-reading of mischarged and misloaded dRNAs.

Unexpected conclusion that directional elongation was driven by the temperature difference between day and night decoding in the warming day and translocation in ….

Duplisome life explains two outstanding paradoxes of the RNA world: First, why there is no clear vestige of the ligase or polymerase ribozyme that assisted templated replication before protein life. Second, the origin of ribosomes and tRNAs in the RNA World before protein translation. Thus it has roles in decoding and phosphoryl transfer for the ancestral SSU and LSU rRNA,

1. The RNA world hypothesis

The *RNA world* conjectures that life on Earth began as self-replicating communities of RNA molecules, augmented later with coded proteins as more versatile gene products, and still later by duplex DNAs as more durable and reparable genes (Rich 1962; Woese 1967; Crick 1968; Orgel 1968; Pace & Marsh 1985; Gilbert 1986). Navigating a treacherous strait of natural philosophy, this hypothesis avoids *dualism* insomuch as these first genes were not materially different from prebiotic chemicals, nor so improbable in sequence as to be miraculous. It also avoids *reductionism* insomuch as the kinetic principles of RNA life realize the nested searches of living systems that explore their surroundings for no higher end than propagating those discoveries. In one early formulation of polymer life, selfish ends and communal means were realized as genes and proteins, respectively, coupled together in nested cycles of survival and reproduction (Eigen 1971). The new insight was that polyribonucleotides could play both roles: When unfolded, they were a peculiar kind of autocatalyst called *templates* of self-replication. When folded, they were common catalysts called *ribozymes* in analogy to catalytic proteins called *enzymes* (Kruger et al 1982; Guerrier-Takada et al 1983).[1]

Pari passu with the two distinct catalytic roles of RNA molecules, there were two distinct levels of evolutionary selection in the RNA world. As a selfish autocatalysts, each RNA molecule competed for common resources of copying, say a pool of precursors, and replicase ribozymes. As common catalysts, each molecule cooperated in promoting the survival and reproduction of the entire RNA community, competing as a whole with similar communities. Thus, at the dawn of life, evolutionary dynamics divided broadly into *vertical gene transmission*, that is, ways of promoting fairness, while limiting selfishness in reproduction of the whole community, and *horizontal gene transmission*, that is, ways of exploring opportunities, while mitigating dangers, from

[1] Here we reserve *ribozyme* and *enzyme* for RNA- or protein-based catalysts, respectively. Thus, it is inconsistent to say proteinaceous ribozyme, and needless to say proteinaceous enzyme. Names with the suffix *-ase* that describe a type of biochemical reaction, such as nuclease, polymer*ase*, etc., are applicable to catalysts of either composition. Thus, we say polymerase ribozyme, or polymerase enzyme where the type of catalyst matters.

mixis of entire communities, or more limited movements of particular RNAs from one community to another.

In any scheme of RNA life, compartments of some form, mediated the coopetition within and between RNA communities contained within aqueous droplets, or upon hydrated surfaces (Mizuuchi & Ichihashi 2021). Proposed physical compartments include naked aerosols, interstices of icy brines, and pores of rocks (Kanavarioti et al 2001; Attwater et al 2013; Mutscher et al 2015; Zhang et al 2022). According to whether their bulk fluid was air, water or oil, proposed organic compartments include encapsulated aerosols, aqueous emulsions in oil, micelle suspensions in water, polyelectrolyte coacervates, as well as true vesicles with lipid or polypeptide membranes (Ianeselli et al 2023).

Beyond conventional compartments with spatial boundaries, sequence-based compartments with kinetic barriers intrinsic to the polyribonucleotides themselves allowed genome segregation, a sophisticated form of chemical disproportionation, within otherwise mixed communities. These barriers include genome tags to identify members of the RNA community for selective aggregation and preferential replication, and endonuclease ribozymes to target foreign RNAs for preferential degradation (Weiner & Maizels 1987; Joyce & Orgel 1993). Like regular updating of IDs and passwords to limit counterfeiting, RNA tag and target sequences provided running definitions of self for replication, and non-self for destruction, respectively. Doubtless the first molecular protection rackets of infection and immunity, wherein parasites are contracted for host defense against similar invaders, as in modern toxin/antitoxin, or restriction/modification systems, trace to sequence-based compartments in the RNA world (Koonin et al 2020).

There are four principal constraints on exhuming the RNA world: (1) physical inferences about geology and planetology of the Hadean eon, (2) laboratory exploration of prebiotic chemistry in far-from-equilibrium environments, (3) rational design and *in vitro* selection of artificial ribozymes, and most clearly, (4) the molecular biology of cells and viruses today. There is some debate whether our planet attained a habitable crust and hydrosphere soon after the great impact that formed the Earth and the Moon, or only later, after a continued period of heavy meteorite bombardment (Benner et al,

2018; Pearce et al 2018). Thus, prebiotic organic chemistry may have commenced anywhere from 4.5 Ga to 3.9 Ga. Regardless of just when stable preconditions of life were finally attained, there is strong evidence from stromatolites, microfossils, and isotope ratios that cellular life had evolved by about 3.7 Ga (Nutman et al 2016; Javaux 2019).

Whether the period from a habitable planet to cellular life was nearly 800 million years, or barely 200 million, several major evolutionary transitions occurred within this interval. The major intermediate stages are characterized by (1) whether RNAs, proteins and DNAs were present, and (2) just how each polymer was made. In one popular scenario, life began with the fitful or steady reproduction of RNA communities by spontaneous copying, leading to faster and more accurate ribozyme-assisted replication (Table 1-1). After the invention of protein translation, these conjectural RNA-directed RNA polymerase ribozymes were retired, replaced by extant polymerase enzymes. Coupling RNA-directed DNA polymerase and DNA-directed RNA polymerase enzymes, duplex DNA stores immediately provided a much more stable backbone, as well as allowed evolution of error-free repair pathways that exploit the informational redundancy of complementary strands. The invention of DNA-directed DNA polymerase enzymes, allowing large chromosomes with many genes, consolidated the genome takeover of cellular life with the system popularized as the *central dogma* of molecular biology (Crick 1955/1958, 1956, 1970; Watson et al MBG). At this stage, the Janus-faced roles of polyribonucleotides in the RNA world, at once evolving genes and their present activities, were largely delegated to DNAs and proteins, respectively.

Ga	Era	RNA	polypeptides	DNA
4.5	prebiotic	random	random	none
	early RNA life	spontaneous copying	random	none
	late RNA life	RdRP ribozyme	random	none
	polypeptide life	RdRP ribozyme	ribosome	none
	protein life	RdRP enzyme	ribosome	none
	early DNA life	DdRP enzyme	ribosome	RdDP enzyme
3.7	late DNA life	DdRP enzyme	ribosome	DdDP enzyme

TABLE 1-1. MAJOR ERAS OF POLYMER LIFE DURING THE FIRST BILLION YEARS OF THE EARTH

2. Dawn of RNA life from prebiotic clutter

A bold defeasible conjecture about the origin of life on Earth, the RNA world has been used to triage the sundry, open-ended questions of astrobiology and geochemistry (Popper 1959). Its first three problems are to demonstrate (1) a prebiotic source of random oligonucleotides, (2) a spontaneous process of RNA copying to multiply chance sequences, and (3) a compatible compartmentation of RNA communities (Joyce & Orgel 1993; Szostak 2012, 2017). Its last three problems are to explain (4) the (now extinct) ribozymatic processes of RNA copying, (5) the breakout of polypeptide translation, and lastly (6) the handover of longterm genetic storage to DNA. Spanning both sets of problems, origin of life studies must explain the domestication of energy and metabolism, including (7) the sources of amino acids and random polypeptides along the way to coded polypeptides, and (8) the sources of chemical free energy along the way to the NTP currency of modern protein synthesis and polymerase enzymes.

There is no consensus on which prebiotic reactions were significant on the Hadean Earth, and many gaps must yet be filled, but confidence is high that such pathways can be fleshed out, given the compelling evidence of life itself (Orgel quote). Plausible prebiotic reactions have yielded scores of aliphatic amino acids, including ten proteinogenic amino acids, amidst a clutter of other amino and hydroxy acids (Miller 1953; Ring et al 1972; Wolman et al 1972; Cronin 1989). Decreasing in rough order G A D E V S I L P T, the yields of these prebiotic, or *primary amino acids* correlate to their chemical free energy of formation (Higgs & Pudritz 2009). More complex proteinogenic amino acids, K F R H N Q C Y M W, are rare or absent in abiotic syntheses, as well as carbonaceous chondrites such as the Murchison meteorite, indicating their synthesis likely required biotic catalysis (Pizzarello 2006; Cobb & Pudritz 2014; Koga & Naraoka 2017). It is unknown, however, which were *secondary amino acids*, first made in significant amounts by now extinct ribozymes, and which were *tertiary amino acids*, first made by extant enzymes.

Compared to the primary amino acids, prebiotic syntheses of nucleotides are more challenging (Joyce & Orgel 1993). Nonetheless, several known abiotic reactions can

generate sugars, nucleobases, and even nucleosides, amidst a clutter of other products (Butlerow 1861; Breslow 1959; Oro 1961; Yadav et al 2020).

In cellular metabolism, polynucleotides are cleaved in their middle, or shortened at one end, by nucleases that consume one water molecule to increase the number of polymer fragments (oligo- or mononucleotides) by one at the expense of one polymer link (phosphoester bond). The chemical free energy greatly favors the forward reaction (hydrolysis) over the reverse (condensation). The enthalpic term reflects the relative stability of the free ends after ionization of the terminal phosphate. The entropic term reflects the greater concentration of bulk water, nominally 55 M, than free ends. Conversely, polynucleotides are made by processive addition of activated mononucleotides to their 3' end, catalyzed by polymerase enzymes. Thus, the unfavorable free energy of polymer condensation, or "water problem", is solved by coupling condensation to a high energy leaving group, *viz*. inorganic pyrophosphate of (d)NTPs.

Like polynucleotides, polypeptides are cleaved in their middle, or shortened at one end, by peptidases that consume one water molecule to increase the number of polymer fragments (peptides or amino acids) by one at the expense of one polymer link (peptide bond). Again, the chemical free energy greatly favors the forward reaction (hydrolysis) over the reverse (condensation). The enthalpic term reflects the relative stability of the free ends after ionization of the terminal amine and carboxyl groups (Martin 1998). The entropic term reflects the greater concentration of bulk water, nominally 55 M, than free ends. Conversely, polypeptides are made by processive addition of aminoacyl-tRNAs to the C-terminus of the nascent polypeptide activated by its tRNA carrier, catalyzed by ribosomes. Thus, the unfavorable free energy of polymer condensation, or "water problem", is solved by coupling condensation to a high energy leaving group, *viz*. inorganic pyrophosphate of (d)NTPs. … It is solved for polypeptide formation in the ribosome by a rather exotic … exclusion of water and enormous alcohol, the tRNA as leaving group.

Prebiotic feedstocks of random polynucleotides and polypeptides cannot be inferred directly from the pathways and intermediates of cellular metabolism today. Beyond inorganic and organic polyphosphates, sundry less obvious molecules have

been proposed as primary stores of chemical free energy on the primitive Earth (Lohrmann & Orgel 1973). A great variety of leaving groups have been considered for prebiotic activation of nucleotides and amino acids. An ideal candidate must be energetically sufficient for spontaneous condensation, kinetically sluggish so that condensation can be directed catalytically, and of course, have plausible abiotic synthesis.

It is likely that the "water problem" of polymer condensation was first solved by physical processes that concentrated monomers and oligomers to increase their chemical activity, or excluded water to decrease its chemical activity (Rodriguez-Garcia et al 2015; Erastova et al 2017; Holden et al 2022). Among likely cyclic evaporites for dry-down condensation are dews, tidal froths and rain water, heated to dryness daily or seasonally by sunlight. With possible participation of deliquescent salts, mineral catalysts, clays and air-water interface ….completed the cycle of alternating between concentration and condensation of substrates and dilution and remixture of products completing the cycle of random polymer elongation (Campbell et al 2019).

Finally, the conventional distinction for polymer condensation between physical concentration and heating of monomers or oligomers, and their chemical activation and catalyzed condensation, is likely too sharp. For polypeptide synthesis, the principle of dry-down condensation can be combined with chemical activation of amino acids and peptides by esterification to say simple alcohols as leaving groups (Griffith & Vaida 2012; Forsythe et al 2015). Upon phosphorylation and activation, monomers can form random oligomers in the absence of any template, perhaps catalyzed by clay surfaces or metal ions (Gibard et al 2018; Liu et al 2020; Pasek 2020).

ability of deliquescent minerals to regulate …evaporate to dryness at high temperatues and spontaneous reacquire water vapor to form aqueous solutions at low temperatures .. diurnal hot-cold cycle or other fluctuations .. rather than unpredictable rains and massive unregulated flooding and dilution .. sea spray aerosols … wet phase difusion and mixing, dry phase condensation of neibhoring monomers and polymers … sprays and return are good for mixis, deliquescent microcompartments … compartments needed for Darwinian evolution but may or may not for abiotic chemistry of random polymers

[wet-dry cycles for polymer condensation] The prebiotic condensation reactions driven by regular diurnal oscillations in temperature and humidity, possibly in tides and surf sprays, not at the mercy of less regular rainstorms

[MOVE DOWN section 14]

NTPs polymerase enzymes

NDPs polynucleotide phosphorylase run backward in vitro

Lohrmann (1975) nucleoside 5' tetraphosphates from NMP and trimetaphosphate

As attractive as it seems, no one has demonstrated a one-pot synthesis for anything like a complete set of feedstocks for RNA life (Anastasi et al 2006; Powner et al 2009; Becker et al 2016, 2019; Sutherland 2016). Parsimony notwithstanding, there is no reason why sundry prebiotic reactions need to have occurred at once, or in the same place, nor to have used common chemical feedstocks and energy fluxes (Benner et al 2012). Indeed some organic compounds may have formed at great removes in time and space in extraterrestrial environments (Chyba et al 1990; Oba et al 2022). Thus, there were likely several material sources, as well as far-from-equilibrium environments, operating under steady conditions, regular cycles, or unpredictable fluctuations (Stueken et al 2013; Ianeselli et al, 2023).

Geological diversity and multi-pot prebiotic syntheses may solve problems of chemical incompatibility, but any such heterogeneity required matching means to stabilize, concentrate and purify oligonucleotides or their precursors, made in modest yield among a clutter of side products (Benner et al 2018; Sasselov et al 2020). Likely geophysical processes to concentrate and enrich these substrates from aqueous solutions include freezing-thawing, evaporation-rehydration, and mineral adsorption-elution (Bernal 1951). Whatever these mechanisms, some niches became in effect commissaries of life, where all essential ingredients from the rough-and-tumble of abiotic syntheses were available at once (Wu & Sutherland 2019).

Once provisioned with monomers and random oligomers in sufficient concentration, if not purity, the second problem of RNA life was a more-or-less faithful copying of oligoribonucleotides, allowing Darwinian selection of the first ribozymes. Considering

the genetical implications of the newly discovered structure of DNA, James Watson and Francis Crick first conceived of *template-directed copying* of polynucleotides (Watson & Crick 1953). Extrapolating phenomena of nucleation and growth from 3-dimensional crystals to linear polymers, they suggested wherever free monomers stacked together along an existing chain, adjacent monomers were positioned to polymerize just in case their nucleobases paired correctly to nucleobases of this template. In this way, each polymer could direct the synthesis of its own unique reverse complement. Watson and Crick were agnostic whether the specific template sufficed for replication, or a general replicase was required as well. In either case, a second round of polymerization could recreate the original template sequence.

As conjectured, template-paired monomers and oligomers can spontaneously polymerize faster than mispaired ones (Sulston et al 1968; Inoue & Orgel 1983; Wu & Orgel 1992; Zhou et al 2019). The kinetics of template-directed copying combines convergence toward fixed-points, familiar from purification through repeated crystallization, with divergence from branch-points seen here as autocatalytic amplification of sequence variants. Because canonical features of RNA, *viz*. D-ribose sugar, nucleobase alphabet, and 3',5' backbone linkages have modest kinetic advantages over competing clutter, chemical purity, homochirality, and regioselectivity all can increase, up to a point, through repeated rounds of copying (Giurgiu et al 2017; Kim et al 2021; tailwinds ref).

All three kinetic principles, perservation of polymer type, nearly faithful copying of polymer sequence, and fair copying of rare sequence variants, defined the genome, or

molecular store for hereditary knowledge, of the RNA community.[2] Here spontaneous copying performed purifying selection, a form of default repair, for generic features of RNA structure. Meanwhile, occasional changes in nucleobase sequence could be inherited, and importantly, selected upon. A life cycle of spontaneous copying, however fitful and error-prone, and folding of these products, however short and unstable, marked the dawn of RNA life. In their open-ended exploration of the affordances of this sequence space, communities discovered and exploited ribozymes that facilitated their survival and propagation in the RNA world. In this breakout biotic environment, first likened to a warm little pond, or later a rich organic soup, the comparatively delicate physiology of RNA life played out amongst molecules solvated and folded, at least from time to time, inside aqueous droplets not too different in from the intracellular milieu of modern life.

Beyond prebiotic clutter and intractable tars, RNA life faced novel thermodynamic sinks and kinetic traps. Strand annealing and tertiary folding are slow processes favored by cold, high salts and neutral pH, while unfolding and melting are favored by warmth, low salts, and acidic pH (Tinoco & Bustamante 1999). Fluctuations of one or more of these factors can shift the balance from annealing and folding, toward unfolding and melting (Ianeselli et al 2023). In thermodynamic sinks, long duplexes with high melting temperatures (T_m) required large environmental fluctuation to melt their secondary structure. In kinetic tars, RNA strands became trapped in useless folds, or entangled with one another, when cooled quickly or shifted abruptly to higher salts. All

[2] Pure RNA polymers may have emerged early, or there may have been a long *conviviencia* of RNA-like polymers with significant fractions of 2',5' linkages, deoxyribose, non-canonical nucleobases (*viz*. 2,6-diaminopurine and hypoxanthine), or modified nucleosides (Fialho et al 2020). Rather than admixtures of familiar elements, life may have begun with another linear copolymer entirely, perhaps a xeno-nucleic acid (XNA) with no counterpart in modern life, whose backbone and nucleobases were favored by prebiotic chemistry (Cairns-Smith & Davies 1977; Nelsesteun 1980; Schwartz & Orgel 1985; Weber 1989; Nielsen et al 1991; Joyce & Orgel 1993; Eschenmoser 1999, 2004, 2005). Finally, life may have used lattice imperfections on mineral surfaces as 2-dimensional templates (Cairns-Smith 1982). Beyond the problem of faithful copying, xeno-genomes introduce two difficult new problems. First, for Darwinian selection, xeno-genes must somehow act on their local environment. Whereas folded XNAs might have acted as xeno-zymes, possible chemical affordances of clay or organopyrite genes are less obvious, and necessarily, more inventive (Cairns-Smith & Hartman 1986; Wachtershauser 1988). Second, any xeno-genome must eventually be translated from the original mineral or XNA medium into familiar nucleic acids (Hud et al 2013). *Genetic metamorphosis* was a bold idea when first proposed for a mineral proto-genome (Cairns-Smith 1965). Since then, the wildly successful takeover of RNA genomes by duplex DNA has been reconstructed from extant reverse transcriptases and ancillary enzymes (section 14).

in all, selection favored polynucleotides with melting temperatures only slightly above the ambient high, poised to fold readily without becoming trapped. Other factors, notably metal cations and small organic molecules, including non-coded polypeptides, could modulate this folding.

Like the porridge of Goldilocks, our concept of physiological conditions allows a narrow range of optimal values for temperature, salinity, pH, redox potential, etc. Larger swings of these values can arrest growth, if not destroy life, and are at best tolerable, never obligate parts of the life cycles of modern organisms.[3] Absent the domesticated sources of free energy and ATP/GTP currency of modern cells, the first RNA communities relied more directly on environmental fluctuations in their life cycle. Moderate changes of temperature, salts and pH were needed to alternate between unfolding and melting for template copying, and annealing and folding of useful ribozymes. Still greater fluctuations were likely needed for concentration by adsorption, evaporation or freezing, as well as abiotic activation and condensation. Whatever the nature, magnitude, and cause of these fluctuations, they were not so lasting, that they degraded the RNA community, or destroyed its compartmentalization, nor so sudden that they trapped RNA in useless intermediates.

[3] The discovery of extremophile life in archaea and bacteria, as well as extremotolerance and cryptobiosis in eukarya such as lichen, nematodes, rotifers, and tardigrades, has probed earlier assumptions about the physicochemical requirements of life (Gade et al 2020). Although they are proof of principle that life can survive, or even thrive, in harsh environments, most or all of the extant adaptations for extremes of temperature, moisture, osmolatity, pH, etc. are likely derived traits, and not vestiges of primitive abilities of LUCA, or earlier polymer life.

3. Biological processes implemented in RNA

NB (Wu & Orgel 1992; Joyce & Orgel 1993) the most favored form of primer extension is hairpin self priming

One great temptation of the RNA world conjecture is to look to the mechanistic simplicity of polymer life for a deeper understanding of evolution. Indeed the same mistakes are made in our understanding of biological evolution as our historiography of biological sciences. In reductionist histories of biology, the pretense that biological processes have been discovered and understood from the bottom-up through their molecular and cellular mechanisms, the success of biological sciences comes from back-and-forth of higher-level process and lower-level realization seen in molecular biology as the the relation between Mendels characterization of genetic elements and cellular and molecular biology of chromosomes and then nucleic acids and proteins. revisionist histories strong prejudices and endless posturing … the success of biological sciences comes from … Mendel and chromosomes and then nucleic acids and polymers … back-and-forth] bottom up from RNA structure to catalytic functions … downward from the concept of process .. analysis of biological functions as formal processes that are realized or implemented in molecular, cellular networks … The most remarkable, and least obvious principle, is that the logic of life and evolution is least clear at the beginning, and becomes clearer and the specialization of … more distinct … thus the distinction between store and expression in polymer life, the difference between specifies instinct and individual experience in neuroscience, and the clearest distinction of search theory in computer science.

Imperfect hairpins, the shortest sequences that quickly find a stable fold under physiological conditions, were the low-hanging fruit of RNA life. Folding of longer RNA molecules is hierarchical and sequential, that is, early intermediates are dominated by favorable secondary structure, while final folds are determined by tertiary interactions, preserving some early secondary elements and rearranging others (Tinoco & Bustamante 1999). Thus, larger, potentially more useful, RNA folds are nucleated by hairpins, rearranged through toe-hold strand displacements or branch migrations, and stabilized by tertiary elements (Vicens & Kieft 2022). Rivaling the complexity of protein

folds, simple topological variations on Watson-Crick duplexes, *viz*. coaxial helices, kissing hairpins, multi-helix junctions, pseudoknots,, unpaired leaders or trailers, and a score of tertiary elements, *viz*. A-minor motifs, G-quadruplexes, kink turns, interdigitating T-loops, loop E motifs, ribose zippers, tetraloops/receptors, T-loops/ receptors, etc. have been described.

Along with intrinsically disordered polypeptides, ribozymes have probed and broadened our notions of folding and catalysis derived from seminal studies of globular proteins. Self-folding protein domains are sequences of about 25-200 residues that fold in one cooperative transition to a unique thermodynamic minimum, often with assistance of chaperones (Anfisen). Any of some twenty proteinogenic amino acids are exquisitely placed at specific positions for protein folding, as well as substrate recognition and cooperative catalysis in the folded protein. Unlike RNA 'breathing spaces' found in natural ribozymes and RNPs, these compact protein folds have few unnecessary cavities. Compared to the sequences of folding intermediates including alternative pathways found in ribozymes, the concerted folding of globular proteins is monotonous, unidirectional and terminal under physiological conditions.

Like concerted folding *simpliciter*, concerted reactions with one principal transition are rare amongst natural ribozymes and RNPs, while sequential reactions that alternate conformational and catalytic steps are common. Compensating for a comparative poverty of functional groups, ribozymes exploit partial unfolding, and alternative refolding, to catalyze sequential reactions. Through sequences of folding, unfolding, and refolding, substrates and intermediates are brought to and from a common active site. To accomplish this, large, more-or-less rigid parts pivot or rotate relative to another at flexible hinges, often with changes in base-pairing, strand displacement, junction slippage, as well as making or breaking of long-range tertiary interactions. Sometimes the rotation of a single conserved nucleobase, changing its stacking and pairing interactions, distinguishes one conformation from the next.

Sequences of conformational states and covalent intermediates reached their zenith in RNA splicing and protein translation whose splicosome and ribosome are conveniently described as *macromolecular machines*. Indeed a mechanistic view of the ribosome, with internal movements conveying substrates and intermediates from one

tRNA site to the next for mRNA decoding and peptidyl transfer, predates both the ribozyme concept and atomic-level descriptions of the ribosome (Bretscher 1968; Spirin 1968; ref).

Unlike most enzymes, the demarcation between substrates and catalyst is blurred for many ribozymes and RNPs. For instance, self-cleaving ribozymes and self-splicing introns, which undergo single turnover reactions upon themselves, fail the kinetic definition of catalysts as reactants restored unchanged at the end of the process.[4] In the back-and-forth of ribozyme engineering, and likely evolution itself, a single turnover reaction, where the substrate is a covalent extension of one or both ends of the ribozyme, can be turned into a multiple turnover reaction with encounter of free substrates and departure of free products, and *vice versa*. Substrates and catalyst are further confounded in substrate-assisted reactions, where a conventional stoichiometric reactant provides an essential element of the active site, complementing the catalyst, and in substrate-induced fit, where the substrate selects and positions the catalyst as much as the converse.

While biochemists reserved the suffix *-ase* for enzymes that make or break covalent bonds, molecular and cell biologists extended the kinetic concept of catalysis to all events characterized as transitions between two defined states. Many such events employ RNAs as catalysts, variously called adaptors, carriers, guides, messengers, scaffolds, switches or templates. In decoding, for example, mRNA, ribosome, and tRNAs selectively accelerate key non-covalent or conformational events, and are restored to their initial states at the end of an elongation cycle. Although they increase rates through the binding and positioning of other reactants, biochemists saw such components as ancillary to enzyme or ribozyme centers that provide say general acid-general base catalysis, or metal coordination to lower the activation energy of covalent events.

By mid 20th century, computer science had abstracted the kinetic concepts of reaction pathways and catalysis from 19th century chemistry as a theory of processes and communication (Shannon 1936, 1948; Turing 1936; Kleene; Petri 1962). Biologists

[4] To be sure, there are many examples of self-reactions in proteins …examples of auto-phosphorylation, auto-proteolysis CAMKII, inteins

imported these new ideas of information, memory and control, piecemeal as needed in their *lingua franca* of regulatory mechanisms. Like artificial processes, a typical biological process comprises a sequence of intermediate states with options of pausing, or even reversing, as well as choices between forward alternatives. These natural regulatory mechanisms were concurrent programs in all but name, effective flow-graphs of reusable instructions, both tests and actions, as well as memory elements, controlling sequential processes and their interactions.

The most kinetically curious, and evolutionarily profound, difference between informational polymers, the polynucleotides and coded polypeptides, and mere regular polymers (Crick). Regular (memoryless) enzymatic processes could output complex oligomers like polypeptide antibiotics, heparan sulphates, or gangliosides through sequential additions with alternative (Kleene). But without a means to modify the program itself … but the inclusion of reprogrammable memory elements ….

On physiological timescales, duplexes formed by Watson-Crick pairing provided a molecular means for storing and copying sequence information. On evolutionary timescales, but a dynamic data store on evolutionary, if not physiological timescale, allowing not just regular (memoryless) processes but molecular processes with true memory. [concept of lock-and-key complementarity. Emil Fischer; but each lock and its key was a peculiar one-off relationship, not part of a systematic relationship between a set of locks and a set of keys] Thus, systematic changes in helical segments caused systematic changes in substrate-product relations. …. or On the other hand, they select one substrate from among several and direct one product from among several similar one … through sequence-dependent interactions, usually Watson-Crick and related base-pairing rules. [change the sequence of the carrier or guide and change the sequence relations between substrates and products] direct and accelerate select, position and orient other substrates, and are restored at the end of the sequence, but are ancillary to the principal ribozyme activity..

…. when one substrate is a family of sequences, programmable machine

guide RNA, messenger RNA, siRNA … some change on physiological time scale, some on immunological, some on more or less evolutionary time scale … Third, [guide

RNAs] [FOOTNOTE] The closest in proteins is polypeptide-guided immunity involving proteolysis of proteins, selecting cognate polypeptides on MHC for activation of cognate TCRs. A remarkably roundabout way of approximating the mechanism of lock-and-key complementarity inherent in the base-pairing phenomenon of nucleic acids.

generalizes the Erhlich concept of fixed lock-and-key relation ….

What covalent reactions did the first ribozymes catalyze? Nearly all natural ribozymes known today act on RNA substrates to make and break phosphodiester linkages by one or two *transesterifications*. In a concerted (S_N2) mechanism, the ribose 2', 3' or 5' OH attacks a phosphodiester bond, with the inline 5' or 3' alcohol leaving. The rates of spontaneous transesterification (as well as hydrolysis) are generally low owing to electrostatic repulsion of oxyanion nucleophiles from the shared negative charge of the non-bridging oxygens (Westheimer 1987; Kamerlin et al 2013). Absent significant differences between the substrates and products in secondary or tertiary structure, transesterification, *viz*. substitution of one alcohol by another, is nearly isoergonic, driven primarily by mass action.

Acting on themselves, natural ribozymes catalyze single-turnover reactions with just one transesterification, *viz*. self-cleaving ribozymes, or two consecutive transesterifications, *viz*. self-splicing introns, and their splicosomal descendents (Kruger et al 1982; Chillon & Marcia 2021; Garside et al 2021; Wilson & Lilly 2021). In the small, self-cleaving ribozymes, both forward (*cleavage*) and reverse (*ligation*) transesterifications occur at appreciable rates. Assisted by general acid A and general base B, the 2' OH attacks the vicinal phosphodiester bond, making the 2',3' cyclic phosphate with the 5' alcohol leaving (Figure 3-1). So long as these cleavage products stay associated, the reverse reaction can relieve the strained cyclic phosphate to restore the original phosphodiester bond. Comparisons of self-cleaving ribozymes suggest that the loss of entropy upon ligation is greater for the flexible *hammerhead ribozyme* than the rigid *hairpin ribozyme* (Nesbitt et al 1999). Indeed, under high salts and low temperature, the hairpin ribozyme actually favors ligation over cleavage. Once the cleavage products dissociate, however, the strained 2',3' cyclic phosphate eventually opens by hydrolysis.

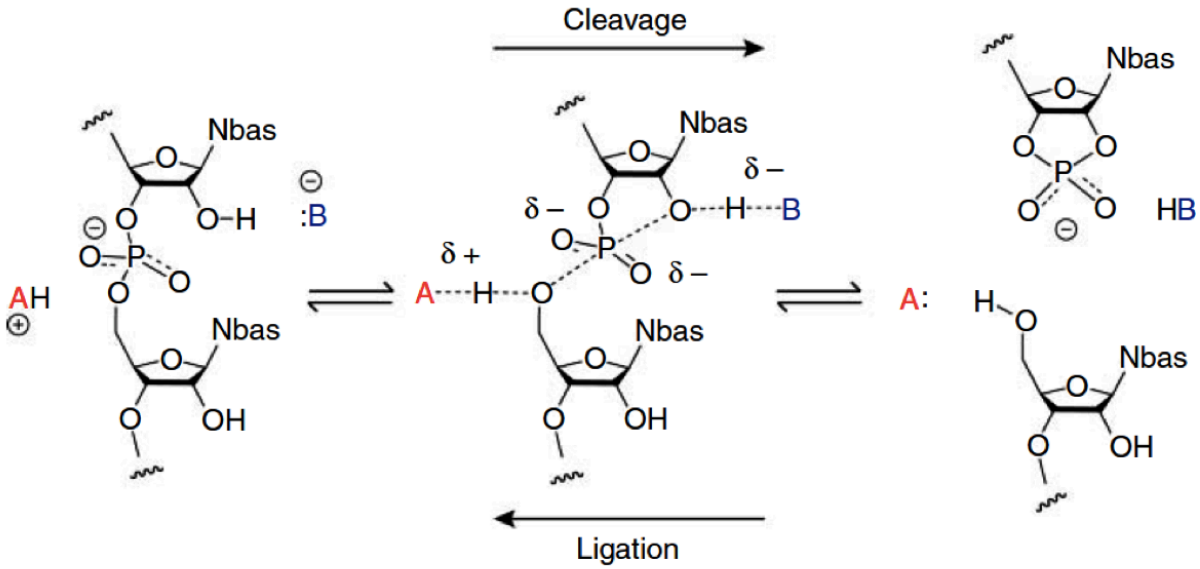


FIGURE 3-1. SELF-CLEAVING RIBOZYMES (WILSON & LILLEY 2021)

guanosine and its nucleotides GMP, GDP, GTP (Cech et al. 1981)

Self-splicing introns perform consecutive transesterifications at two distinct phosphodiester bonds called the *splice donor* and *acceptor* sites, respectively. Group I and group II introns differ in the oxyanion nucleophile of the first trans-esterification. In group I introns the 3' OH of a free guanosine Gα attacks the phosphodiester bond, with the 3' alcohol leaving. In group II introns the 2' OH of a downstream adenosine attacks the donor site, with the 3' alcohol leaving. As a result, the upstream cleavage products end with 3' OH, while the downstream products begin with the guanosine, or a lariat branched at the adenosine. In the second transesterification, the 3' OH of the upstream exon attacks the splice acceptor, with 3' alcohol leaving. After both transesterifications, upstream and downstream exons are joined by a phosphodiester bond based on the splice acceptor phosphorus. The second product is a linear intron, beginning with the Gα and ending with the internal terminal Gω, or a lariat-shaped intron ending with …, respectively. thermodynamics … a number of intermediates but critically the substrate in docking reaction redocking reaction undocking … retain the upstream exon at the active site, displace the intron part with the the acceptor splice site

Compared to proteins, simple RNAs have few functional groups for catalysis. Allowing regular hydrogen bonding and hydrophobic stacking, none of the four nucleobases is ionized at neutral pH. For general base catalysis, adenine and cytosine For general acid base catalysis, adenine (pK_a 3.5) and cytosine (pK_a 4.2) are protonated below pH 4; for general acid catalysis, guanine (pK_a 9.4) and uracil (pK_a 9.3) are deprotonated above pH 9 (ref). Alternatively, magnesium cations can activate oxyanion nucleophile of water or alcohol, or stabilize the the developing negative charge on oxyanion leaving group. In general, small self-cleaving ribozymes rely on general acid general base catalysis for transesterification yielding 2',3' cyclophosphate and 5' OH products, while self-splicing introns, RNase P, and the ???ribosome peptidyl transferase center …. ???

hydrolysis : metal-dependent protein RNases 3' OH and 5' phosphate

transesterification : metal-independent protein RNases 2',3' cyclo phosphate and 5' OH

RNase P RNA and the ribosome peptidyl transfer center extend the catalytic repertoire of ribozymes beyond *cis* transesterifications found in self-cleaving ribozymes and self-splicing introns. Both of these ribozymes act in *trans*, that is, on substrates that are not covalent extensions of the catalytic RNA, and mediate multiple turnovers with dissociation of products and association of new substrates. … downhill or exergonic …

RNase P RNA hydrolyzes the phosphodiester backbone of pre-tRNAs and other RNAs (Guerrier-Takada et al 1983). Hydrolysis differs from transesterification in that the oxyanion nucleophile comes from water, not the polyalcohol ribose. Enthalpically, hydrogen is more electropositive than carbon, so that the water is more polarizable than the alcohols. Entropically, the orientation of water is less constrained than ribose oxyanions, but its effective concentration is higher. Due to higher effective concentration of nucleophile than leaving group, mass action favors the forward (hydrolysis) over the reverse (condensation) reaction with *provisos* that solvent freely enters and products freely leave the active site.

The ribosome peptidyl transfer center catalyzes two distinct reactions, both exergonic or downhill reactions. During polypeptide elongation it catalyzes peptide

bond formation, an aminolysis of the acyl-ester that transfers from acyl-ester to amide. During polypeptide release, it catalyze hydrolysis of the same acyl-ester. The two reactions, hydrolysis or aminolysis of acyl-esters, are unique among natural ribozymes (Noller et al, 1992; Ban et al 2000; Nissen et al 2000). The hydrolysis is favored by mass action of water, while the aminolysis is favored by the greater stability of amide than acylester bond. Like the catalytic site of RNase P RNA, the chemical potential of water in the ribosome transfer center is far below bulk water; nominal concentration of water in free solution 56 M.

Two observations suggest that the repertoire of covalent catalysis was greater for ancient ribozymes than the surviving examples. First, through a back-and-forth of design and selection, artificial ribozymes have been successfully engineered for a wider variety of biochemical and general organic reactions (Wilson & Szostak 1999). Second, the sundry nucleotide-derived cofactors of modern enzymes are likely vestiges of ribozymes that predate protein life (White 1976; Jadjav & Yarus 2002; Goldman & Kacar 2021; Kirschning 2021). A few examples of these cofactors are shown in Figure 3-2. As a caveat, it is generally unknown whether the substrates of artificial ribozymes were available in the RNA world. This applies not just to exotic prebiotic feedstocks, but mundane metabolic intermediates. In particular, it is unclear whether NTPs, the energy currency of modern metabolism, were available as precursors in the RNA world, or came only later, in protein life (sections 4 & 14).

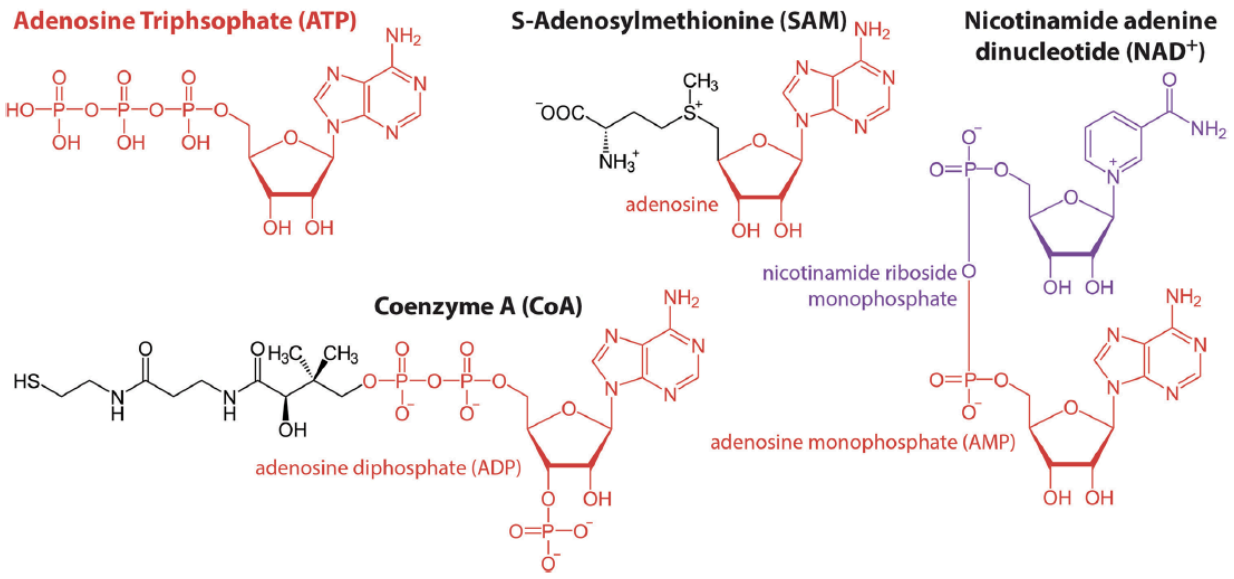


FIGURE 3-2. NUCLEOTIDE-DERIVED COFACTORS (GOLDMAN & KACAR 2021)

Generalizing the kinetic concept of catalysis to molecular conformations and cellular events beyond formation and dissolution of covalent bonds, a rich trove of natural RNAs, many likely ancient, have been found to regulate transcription, protein translation and export, and other cellular functions. There is a panoply of *riboswitches* that monitor everything from temperature, to inorganic ions, amino acids and nucleotide-related signaling molecules (cAMP, ppGpp, ZTP, c-di-GMP, ci-di-AMP), to the charge status of tRNAs, or the presence of hydrophobic signal sequences in the ribosome nascent polypeptide exit tunnel (Nelson & Breaker 2017). Changing conformation by unfolding and refolding, such riboswitches regulate options to pause or continue, as well as choices of two alternatives, within multi-step molecular processes.

Jash B, Tremmel P, Jovanovic D & Richert C (2021). Single nucleotide translation without ribosomes. Nat Chem

Muller F, Escobar L …. et al 2022. A prebiotically plausible scenario of an RNA-peptide world. Nature 605, 279-284.

4. RNA copying | ligase & polymerase ribozymes

Four types of ribozymes were prized discoveries at the dawn of RNA life: First, activities that improved the yields of mononucleotides and random oligonucleotides from prebiotic feedstocks, including salvage of hitherto dead-end molecules as metabolic intermediates. Second, activities that improved the generality, fidelity or speed of RNA copying, including protection or repair of existing molecules. Third, activities that improved genetic or metabolic compartmentalization, including selection of self RNA, rejection of nonself RNA, concentration of feedstocks, or diffusion of wastes. Finally, no one RNA community could acquire all of these useful innovations through in-house discovery and vertical gene transmission. Beginning with the simplest processes of horizontal gene transmission, assortment and recombination, RNA communities joined an ever-accelerating race to garner ribozymes from the pangenome that fit together into useful regulatory networks.

We use the term *copying* generically for processes of polymer reproduction, irrespective of substrates and intermediates, as well as any catalysts beyond the template polymers themselves. Today we associate RNA copying with long duplexes of two strands, each the *reverse complement* of the other. Nearly all schemes of spontaneous copying at the dawn of RNA life create long duplexes as intermediates. Moreover, the reverse complement is the key intermediate of enzymatic copying in modern cells and viruses. Polymerase enzymes may be non-templated () or template-directed. The later may be limited to fixed templates (telomerase) or be promiscuous. Promiscuous template-directed polymerases may copy short regions (repair polymerases) or entire chromosomes (replicative polymerases). Here we reserve the term *replication* for the common mechanism of all polymerase enzymes which read templates from 3’ to 5’ in one nucleotide steps, while making the reverse complement from 5’ to 3’ by one nucleotide additions.

We need some additional terminology to describe and compare a larger universe of copying schemes. First, there is a continuum of copying schemes between the most distributed and nondirectional (e.g., templated ligation of large oligomers) to the most processive and directional (e.g. primer extension by monomer addition). One key

mechanistic question is how the catalysts move from place to place along the template: associate and dissociate, freely sliding along the template (scanning) in one or both directions, processive and directional movements (translocation) in fixed-size steps. including possible helicase or other activities to bypass or remove or obstacles. Mechanistically catalysts can associate and disassociate or scan/slide … polymerase adds defined short oligonucleotides or mononucleotides … processive and directional . translocation. For processive copying we use the terms *5'-to-3*' and *3'-to-5'* to indicate the absolute direction of extending the product. Because we consider copying schemes that achieve in one elongation cycle what polymerase enzymes achieve in two rounds of copying …. We call the product a forward or reverse copy (relative to the template) if its backbone has the same or opposite polarity, respectively. More novel, we call the product a *duplicate* or *complement* (relative to the template) if its nucleobases are the same as the template or their Watson-Crick complements, respectively. We refer to copying schemes that over-specify the template, or under-specify the product, as having *restricted templates* or *degenerate products*, respectively. Unless so stated, all of the other schemes discussed here are proposed to be fully *general* and *faithful*.

as raw material rather than activated mononucleotides such as NTPs (Sharp 1985; Orgel 1986; Doudna & Szostak 1989)

[activated mononucleotides Pace & Marsh (1985)]

Ribozymatic RNA copying no doubt produced longer and more accurate copies than spontaneous copying. Over time ribozymatic copying overcame restrictions on templates and degeneracy or other errors in products, more universal, or less sequence-dependent in gene copying. Pari passu, became more selective in template recognition, or preferential replication of self genes, and. [functional distinction between plus strand (ribozyme) and minus strand (copying intermediate)] Whereas distributed spontaneous copying (simple templated ligation) made no genetic distinction between the complementary strands, in more processive copying a clear distinction emerged between the full-length template strand and the nascent copy strand. Although both are self RNAs, mechanisms to treat plus and minus strands differently.

Discovering that cells and viruses use enzymes to replicate and repair their DNA, Arthur Kornberg and colleagues characterized their DNA-directed DNA polymerases (see Kornberg 1969). Given all four dNTPs, these enzymes can extend a primer on any unique template, reading it from 3' to 5' in one nucleotide steps, while making a complementary copy from 5' to 3' by one nucleotide additions. In an inline S_N2 reaction, the 3' OH of the nascent polynucleotide attacks the dNTP α-phosphate with the γ-β pyrophosphate leaving. Completing the elongation cycle, the polymerase moves one nucleotide along the template before the next read-add step. The copy is formally the *reverse complement* of the template with reverse polarity of the backbone and Watson-Crick complement of each nucleobase. Another round of copying this intermediate, or minus strand, recreates the sequence of the original template, or plus strand.

The new paradigm of processive copying of nucleic acids was quickly cemented by discoveries of other template-directed polymerase enzymes (Watson et al MBG). DNA-directed RNA polymerases transcribed duplex DNA into mRNA and non-coding RNAs given all four NTPs (Burma et al 1961; Geiduschek et al 1961; Stevens 1961; Chamberlin & Berg 1962; Furth et al 1962). RNA-directed RNA polymerases replicated the genomes of RNA viruses (Haruna et al 1963). And lastly, RNA-directed DNA polymerases in retroviruses and telomerase reverse transcribe the genomes of retroviruses from RNA, the guide RNA of telomerase into duplex DNA, and retrotransposons (Baltimore 1970; Temin & Mizutani 1970; Greider & Blackburn 1989; Collins & Greider 1993).

The speed and accuracy of copying … free energy of PP hydrolysis, or other helicase, to drive strand separation of duplex template, and kinetic proof-reading (ref). Beyond polymerase enzymes *tout court*, a handful of steps to complete the replication cycle engaged molecular biologists for years to come. Any list of these problems would include template selection, replication starts including primer synthesis and annealing, replication termination including end modifications, thermodynamics and kinetics of unwinding double-strand templates and separating template and copy strands,

topology of supercoils and linking, and bidirectional replication of leading and lagging strands.

With the central dogma of DNA replication, RNA transcription, and protein translation in place, molecular biologists asked how these informational (genetic) polymers and their polymerases evolved. Two contrasting approaches emerged in the search for tractable models of the evolution of polymer life, one narrow and the other eclectic. In the narrow approach, investigators sought a single selfish molecule capable of catalyzing its own replication. In the eclectic approach, they allowed communities of informational polymers of various types with interdependent cycles of survival and reproduction (Eigen 1971; Kaufmann).[5]

Studying *in vitro* replication of RNA virus genomes by their native polymerase, Solomon Spiegelman and colleagues showed that these enzymes favor their native template over other RNA or DNA templates, faithfully replicate rare sequence variants, and make copies of copies *ad nauseum* (Mills et al 1967; Kacian et al 1972). With Qβ polymerase, the only viral protein required for *in vitro* replication, supplied by the investigators, virus RNA was merely a template for copying, not an mRNA encoding useful proteins. Absent natural selection for the full-length functional products necessary for producing infective particles, virus RNA tolerated any mutations that did not disrupt the steps of polymerase recognition, priming and replication. In serial transfer experiments, selecting for faster replication, shorter sequences with possibly enhanced recognition, dubbed 'little monsters', soon took over the population.

Extrapolating the concept of Mendelian species with lawful matches between alleles of the same locus to a Hobbesian war of all genes against all, Richard Dawkins popularized the selfish behavior of gene within the genome and the viewpoint of replicator dynamics with evolution of lower level replicators nested within higher ones (their vehicles) (Dawkins 1976).

Richard Dawkins popularized the new idea of replicator dynamics, sequences that compete better for replication within the genome, but make little or no positive

[5] Finally, CNO nucleosynthesis, metabolic networks, lack the data storage principle that separates transmission of knowledge from its expression [metabolism-first]]

contribution to fitness of the whole, as selfish DNA (Dawkins 1976). Without mechanisms of enforcing the genome compact, this lawless was not an aberration but default of Hobbesian war of all against all. In successful organisms, bacteria and eukarya, the curious distribution of mobile genetic elements and repetitive sequences dispersed throughout the genome, and phyletic distribution could be explained … (cf. Doolittle & Sapienza 1980; Orgel & Crick 1980).

explained the widespread occurrence of mobile genetic elements as selfish genes that succeed through intragenomic competition without direct regard

[realized that repeated sequences dispersed through the genome, and genomes of distant species, were mobile genetic elements that … evolution of useless or parasitic RNA.

Dawkins emphasis on intracommunity competition; success within comes at a cost to the competition between communities

short-run advantage exploit known affordances, long run-advantage explore unknown; defense against like parasites

Manfred Eigen noticed two simple constraints on genome replication under purifying selection, one on genome size *tout court*, and the other on its sequence information (Eigen 1971). If the genome was not to degenerate in absolute length, the rate of replication must exceed the rate of decomposition. If it was not to degenerate in sequence information, replication must make at least one faithful, or error-free copy each generation. A simple argument shows that the effective size in nucleotides sustainable under purifying selection is about 1/k where k is the error-rate per nucleotide per replication. As Maynard-Smith put it : chicken-egg problem …. accurate replication of nucleic acids required a polymerase enzyme, but the gene for this enzyme required accurate replication.

Walter Gilbert identified the major evolutionary transition from early RNA life, based on feedstock ribozymes and spontaneous copying, to late RNA life, as the emergence of a single, conceivably large RNA molecule capable of “self-replication, mutation and hence evolution toward ever more efficient self-replication” (Gilbert 1986). Beyond prebiotic feedstocks of random oligonucleotides and spontaneous RNA copying, the

third Holy Grail for reincarnating the RNA world was to discover, design, or select a polymerase ribozyme whose overall reaction, if not mechanistic steps, modeled, more or less closely, natural RNA polymerase enzymes.

When Gilbert coined the *RNA world*, there were three main arguments for an (extinct) polymerase ribozyme: First, there were clear hints that RNA had greater catalytic versatility than the two known natural ribozymes, self-spicing introns and RNase P. Second, there seemed no clear alternative to a template-dependent ligase, if not a true polymerase, for RNA life moving beyond spontaneous copying. Finally, there were already suggests that template ligation of self-splicing introns … might be the basis for template-directed addition of ????

As a starting point in the quest for template-directed ligase or polymerase ribozymes, investigators turned to the group I self-splicing intron found in the large subunit rRNA of *Tetrahymena thermophilia*. Cech and colleagues modeled RNA copying with an isoergonic transesterifcation of the terminal nucleotide of a feedstock oligonucleotide to the 3' OH of the nascent polynucleotide.

pCCCCC or longer

PCCCC leaves

nascent polynucleotide grows 5' to 3'

products have 3' OH

leaving group is oligo not pyrophosphate

[transesterification]… a model of template addition of the terminal nucleotide of source to the end of polynucleotide acceptor ….(Zaug & Cech 1986a, 1986b; Cech 1986; Been & Cech 1988)

templated elongation by transesterification of terminal nucleotide of source to end of polynucleotide acceptor N5+ Nk > N4 + Nk+1

how to make recursive?

(Sharp 1985; Doudna & Szostak 1989)

[engineered GISSI] template-directed ligation of short oligonucleotides

allow single nucleotide additions or ligation of up to four RNA oligos on a complementary strand (Doudna & Szostak 1989)

improved by in vitro selection (Green & Szostak 1992)

Rather than start with group I self-splicing intron, or another natural ribozyme, investigators selected RNA ligase ribozyme *ab initio*. that uses oligonucleotide 5' triphosphates as substrates and forms 3',5' phosphodiester linkages (Bartel & Szostak 1993;)

[Ekland, Szostak & Bartel 1995]

required template tethered to ribozyme

low processivity

low speed

no strand displacement

specific sites of initiation and termination?

(Johnston et al 2001) primer extension up to 14 nt on general template

Johnston WK, Unrau PJ, Lawrence MS, Glasner ME & Bartel DP (2001). Science 292, 1319-1325.

(Joyce & Orgel 1993) doubt can be very processive

(Joyce & Orgel 1993) The quest for the nucleotide polymerase ribozyme for templated replication (Wachowius & Holliger 2021?). The problem of one RNA that folded ribozyme was general catalyst of its own self-replication as well as other members of its RNA community.

Variants of the original RNA polymerase ribozyme today (size?) can recognize a general primer-template helix in *trans*, adding hundreds of nucleotides in 3'-5' linkages with high speed and good fidelity. There are still obstacles of quantity or kind to robust .. including stability of the polymerase itself, template unwinding. Novel environments or triphosphate trinucleotide additions. The three decades of research on the class I ligase have shown something of what a RNA polymerase ribozyme can do. It would be ill-advised to draw firm conclusions from this about what a polymerase ribozyme can never do. Whereas the general problem of a robust RNA replication cycle

is difficult, the powers of RNA catalysis are considerable and the search spaces in RNA sequences is large, or scientific concepts is yet unfolded are large in evolution or lab is large, all of this invites a healthy optimism about its ultimate solution …, and agnosticism about its eventual form.

In the search for the missing link of late RNA life, ribozymatic RNA copying, templated ligation has been recognized as an attractive intermediate, or even a final alternative for the missing polymerase ribozyme (Levy & Ellington 2001).

Starting from the ligase activity of the *Azoarchus* group I intron [cooperative networks of RNAs that catalyzed reproduction of network; only limited variety of sequences satisfying strong sequence and folding constraints … inflexible genome tag - staring from the ligase activity of group I intron (Vaidya et al 2012)]

In four decades of exhuming or reinventing the extinct ligase or polymerase ribozyme for RNA copying, the argument for this missing-link ribozyme from RNA versatility has gotten considerably stronger with discoveries of natural ribozymes and riboswitches, including RNA-guided mechanism of genome recombination, defense, and gene regulation, and with selection of artificial ribozymes that can perform RNA-guided reactions from RNA copying to peptide synthesis. However, the argument for this missing-link ribozyme from wont of any alternative for processive RNA copying has gotten weaker. In particular, the conceptual monopoly of the RNA polymerase ribozyme has been challenged by novel alternatives based, on what was known, or imputed to the ribosome in protein translation (sections 5-8).

5. RNA copying | permuted reverse duplicate

The RNA world poses two great mysteries for comparative genomics. One is a *mystery of absence*, *viz*. something needed yet not clearly found, and the other a *mystery of presence*, *viz*. something found yet not clearly needed. Notwithstanding sundry RNA-directed enzymes of modern cells and viruses for RNA or DNA replication, modification, cleavage and repair, nor self-splicing introns and nuclear splicosomes for RNA recombination, there is no clear vestige of the RNA-directed ligase or polymerase ribozymes used for RNA copying in the RNA world. It is hardly tenable that this central process of RNA life vanished without a trace. Meanwhile, the machinery of protein translation, common to all cellular life, is based on rRNAs and tRNAs that seemingly sprang from nowhere. It is hardly tenable that these central molecules of protein life had no functional progenitors in the RNA world.

To get around the need for coded proteins in protein translation today, molecular biologists considered the possibility of ribosome-free polypeptide translation by direct interactions peptidyl- and aminoacyl-tRNAs brought together by their codon-anticodon pairing with the mRNA (Crick et al 1976; Eigen & Schuster 1979). Peptide bond formation was energetically favorable and only required placing the substrates in reasonable proximity for entropic catalysis. For processive translation, the free energy of peptide bond formation might drive translocation. Carl Woese had proposed a *reciprocating ratchet model* for translocation, soon modified as the RRY code based on the anticodon loops of the form 3’ UGYYRUU 5’, and later the RNY code (Woese 1970;). One unsolved problem with these mechanisms was there was no coupling of peptide bond formation to translocation that might drive processive elongation, that is, the ’ratchet’ was not a genuine unidirectional motor, but a reversible strand displacements (Woese 1980).

Mostly ending the quest for ribosome-free mechanisms of polypeptide translation, there is now ample evidence that the ribosome decoding and peptidyl transfer centers are themselves ribozymes formed from the small and large subunit rRNAs, respectively (Noller et al 1992; Moore & Steitz 2002). Moreover, there is now a fair case that bacterial T-box riboswitches are vestiges of extinct tRNA charging ribozymes (Ishida et

al 2020). But all of this evidence for translation mediated by primitive rRNAs, tRNAs and ribozymes *without any support from coded proteins* merely whets the appetite for an explanation of just what those several RNAs were doing *before the breakout of polypeptide translation*.

In “An RNA replisome as the ancestor of the ribosome” John H. Campbell at UCLA David Geffen School of Medicine dissolved both mysteries of RNA life, the disappearance of ribozymatic RNA copying without descendants, and the appearance of ribosomal polypeptide translation without ancestors (Campbell 1991). There was no ancient polymerase ribozyme to exhume, Campbell suggested, but an entirely different mechanism of processive RNA copying that read the template from 5’ to 3’ in *triplet codons*, while adding *triplet duplicons* to a nascent copy from 3’ to 5’. Unlike DNA replication, or RNA transcription, where one and the same (d)NTP that reads the next unpaired nucleobase of the template is concurrently added to the nascent copy, in this original form of processive RNA copying, and later in protein translation, adaptor RNAs separated template reading from product addition in both space and time.

Campbell’s proposal was the first of several schemes for RNA copying based on conjectural ancestors of the ribosome and its transfer RNAs (Campbell 1991; Weiss & Cherry 1993; Gordon 1995; Poole et al 1998; Yakhnin 2007; Noller 2012). Here we adopt a uniform terminology to describe those schemes, and to compare them to our own. To avoid confusion with the ribosome and tRNAs of protein translation, we refer to these conjectural ancestors as the *duplisome* and *donor RNAs (dRNAs)*, respectively. In all of these copying schemes, the template is read from 5’ to 3’ in one *codon* steps by matching to the dRNA *anticodon*, while the nascent product is lengthened by one *duplicon* additions.

The location and size of the duplicon varies in the different schemes of RNA copying, affecting all aspects of the elongation cycle from dRNA loading to decoding and nucleotidyl transfer. In a family of schemes that make the reverse complement as a replication intermediate, the same triplet acts first as anticodon and then as duplicon (Weiss & Cherry 1993; Gordon 1995; Poole et al 1998). Like the replication cycle of polymerase enzymes, the first round of copying creates a reverse complement of the template, while a second round of copying restores the forward polarity of the

backbone, as well as the original nucleobases (section 6). In another family of copying schemes that make a duplicate without an intermediate complement, both rounds of Watson-Crick basepairing occur *within the polynucleotide elongation cycle itself*, one in loading the dRNA with a cognate duplicon, and the other in decoding of the template RNA. In two such schemes the duplicon and its complementary anticodon are distinct parts of the dRNA (Campbell 1991; Hedgecock & Proenca present paper), while in the third scheme, the duplicon is a free triplet (Noller 2012; section 7). Finally, the *translocon* comprises the coaxial helices of template and dRNAs translocated within the duplisome; its distance moved is either three nucleotides as in protein translation, or just two nucleotides (sections 8, 11, 14).

Era	RNA	polypeptides
early RNA life	spontaneous copying	random
late RNA life	duplisome	random
polypeptide life	duplisome →	ribosome
protein life	RdRP enzyme	ribosome

TABLE 5-1. RIBOSOME LIFE

Like the discovery that birds are living descendants of dinosaurs hidden in plain sight, Campbell proposed that ribosomes are surviving descendants of a lost RNA replisome hidden in plain sight. Table 5-1 lists the major stages of ribosome life: First, in late RNA life, spontaneous RNA copying *without support of evolved ribozymes* was replaced by processive RNA copying mediated by the duplisome and its dRNAs. After the breakout of polypeptide translation, RNA duplication continued alongside translation. In the parlance of synthetic biology, the late duplisome and early ribosome were orthogonal polymerases that copied or translated template RNAs, respectively. Finally, in late protein life, the duplisome and its dRNAs were retired in favor of an RNA-

directed RNA polymerase enzyme. By then, the ribosome and its tRNAs were wildly successful exaptations ensconced in their roles of protein translation, and increasingly dependent on coded proteins, *viz*. ribosomal proteins, RNA modification enzymes and biogenesis factors, aminoacyl tRNA synthetases, translation factors, and signal recognition particle.

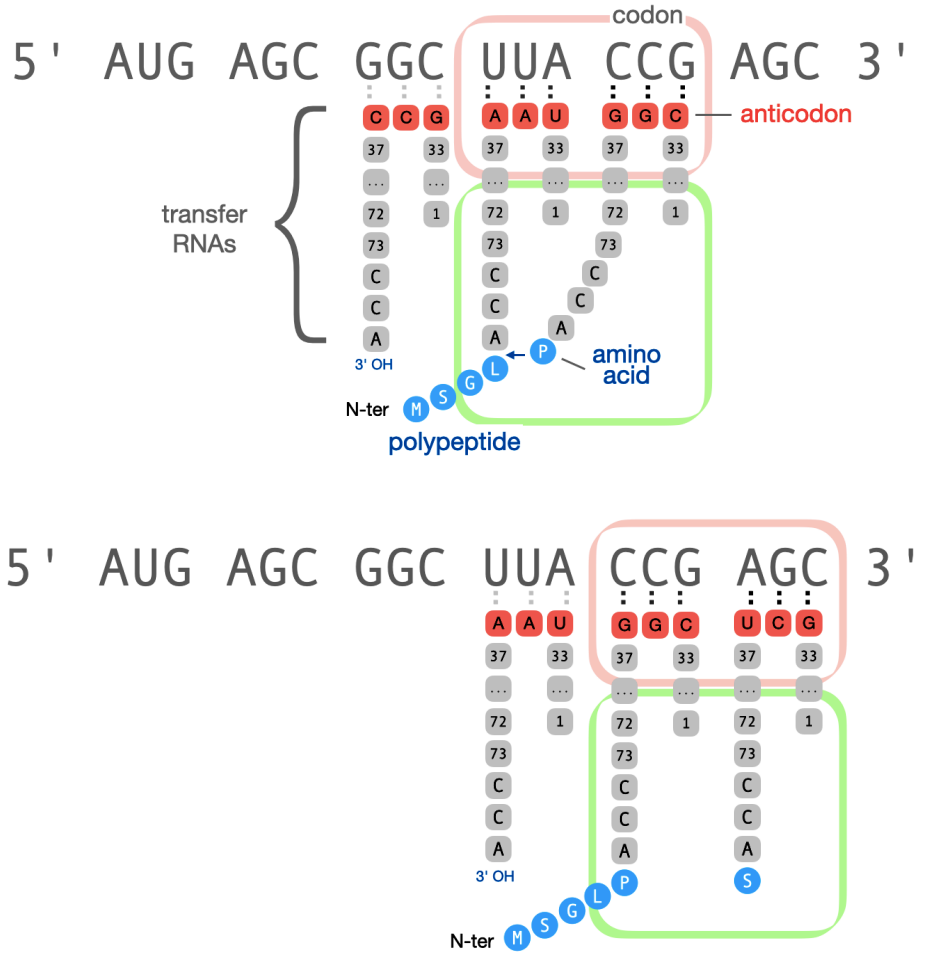


FIGURE 5-1. POLYPEPTIDE ELONGATION CYCLE

Figure 5-1 depicts one cycle of polypeptide elongation in the ribosome. In the upper panel, the α-amine of the aminoacyl-tRNA in the A-site attacks the acyl-ester of the peptidyl-tRNA in the P-site to transfer the nascent polypeptide chain. In the lower panel, the ribosome has moved along the mRNA to the next codon and a new aminoacyl-tRNA has entered the decoding center and been accommodated. Ghosts of the deacyl-tRNAs after leaving the E-site are shown for reference. Figure 5-2 depicts one cycle of polynucleotide elongation in Campbell's scheme of RNA copying. The template is read from 5' to 3' in codon triplets by the duplisome. The duplicon comprises the last three nucleotides of the dRNA, which Campbell suggested might be ancestral to the universal $_{74}$CCA$_{76}$ at the 3' end of mature tRNAs. For nucleotidyl

transfer, he implied that the 3’ OH of the duplicon-dRNA attacks the polynucleotidyl-dRNA, with the freed dRNA leaving. Thus, the product is made from 3’ to 5’ by addition of duplicon triplets.

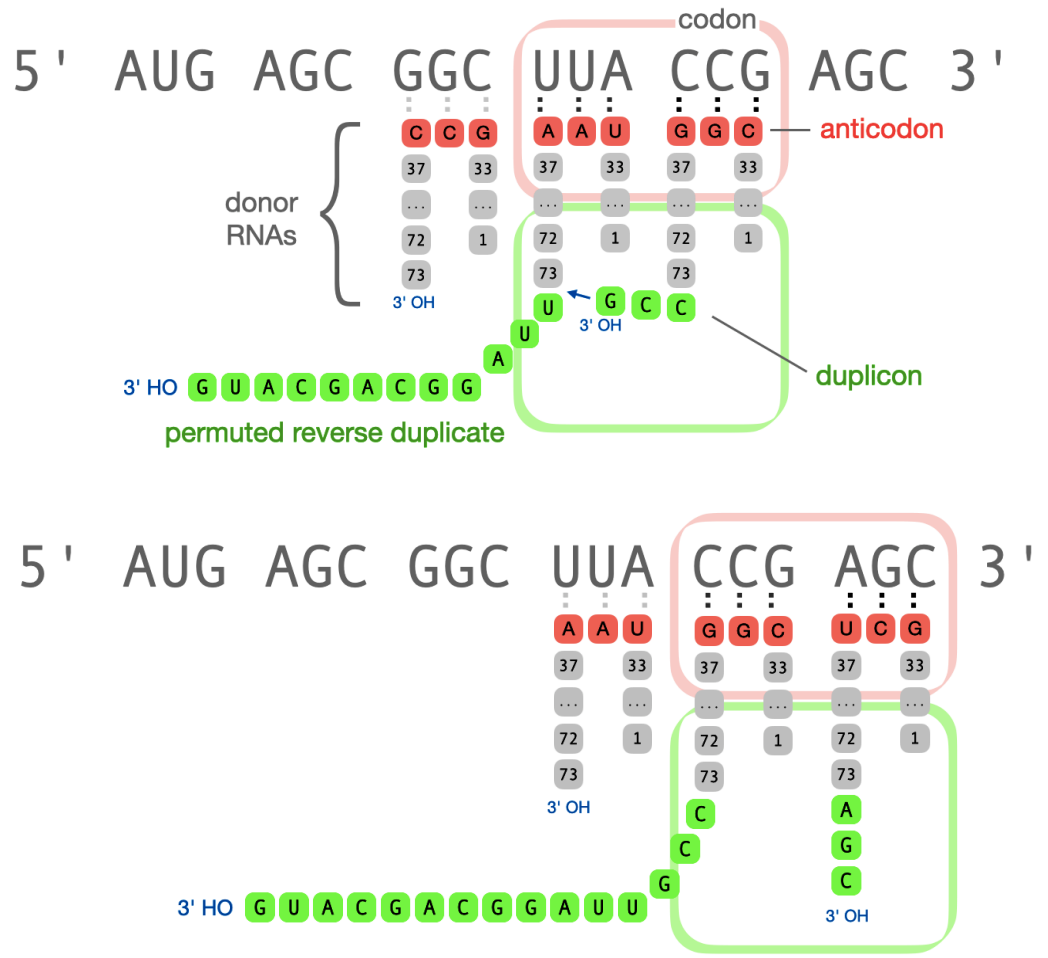


FIGURE 5-2. POLYNUCLEOTIDE ELONGATION CYCLE (CAMPBELL 1991)

An inspection of Campbell’s scheme of RNA duplication shows three serious problems. The first problem is the novel structure of the copy itself (Figure 5-3). This *permuted reverse duplicate* is neither a forward duplicate identical to the template, nor the familiar reverse complement of polymerase enzymes, but a reverse duplicate with the first and third nucleotides of each codon transposed.[6] One virtue, Campbell noticed, is that templates do not anneal to such copies, and hence, the long duplexes

[6] Although we strongly favor a scheme of RNA copying (sections 8-12) that produces forward duplicates, one might ask whether any imprimatur of Campbell’s permuted reverse duplicates is still be detectable by computational genomics.

formed in spontaneous RNA copying do not arise during processive copying, nor subsequent annealing of product and template. Like the familiar reverse complement, however, the permuted reverse duplicate still needs a second round of copying to recreate its original template.

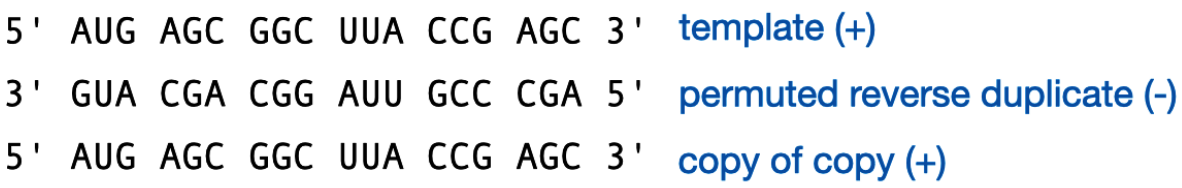


FIGURE 5-3. PERMUTED REVERSE DUPLICATE

The second difficulty of Campbell's scheme is that templates require as many as 64 different dRNA isoacceptors for faithful copying, one for each codon triplet. This is substantially more than the number of tRNA isoacceptors in modern genetic codes. The final problem is that the cycle still needs some means to repeatedly and accurately reload dRNAs with duplicons consumed in polynucleotide elongation. Campbell himself only hinted that the duplicon is somehow self-loaded onto the dRNA from its anticodon. Anticodon-directed dRNA loading presents two distinct subproblems, *viz*. placing the anticodon in spatial approximation to the duplicon, and the covalent chemistry of duplicon loading. Allowing that the proximity problem can be solved, to avoid a regression of nested Matryoshka dolls any plausible chemistry of anticodon-directed dRNA loading must be strictly simpler than our overall copying scheme. That is, we cannot evade the problem by invoking some vague, yet general mechanism of template-directed ligation or primer extension, *as this is the whole point of the duplisome and its dRNAs in RNA copying*.

Of those papers that have cited Campbell (1991) over the years, only two have pursued his scheme of RNA copying. In "A model for the origin of protein synthesis as coreplicational scanning of nascent RNA", Alexander Yakhnin …. depicts the riboreplisome as extending the nascent polynucleotide from 3' to 5' by triplet additions? , the reverse product.

whatever the virtues of scanning… the scheme of concurrent RNA copying and protein translation appears internally inconsistent, notably Campbell's adaptor RNA

carries a cuplicon …. overloaded adaptor RNA … carries aminoacyl and duplicon at 3’ end

In “Possible ancestral functions of the genetic and RNA operational precodes and the origin of the genetic system” ### Martínez-Giménez and ### Tabarés-Seisdedos (2021) modified an earlier proposal …

Like tRNA charging, dRNA loading poses a molecular ’recognition at a distance’ problem, *viz*. how to match a duplicon to the anticodon at another end of the molecule. In the genetic code, some 20 aminoacyl tRNA synthetase enzymes recognize one or more tRNA isoacceptors, charging the 3’ end of their acceptor arms with the cognate amino acid. To recognize and charge their tRNA substrates, these enzymes must span the distance from the 3’ end of the acceptor arm to determinants of tRNA identity found at various sites along the acceptor stem, the elbow and variable arm, to the anticodon arm and anticodon itself, about 75 angstroms away (ref).

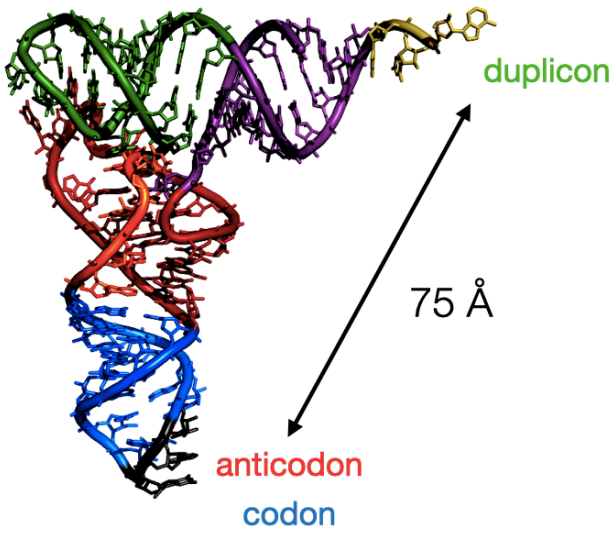


FIGURE 5-4. SPATIAL SEPARATION BETWEEN ANTICODON AND DUPLICON

Like polypeptide elongation in the ribosome, polynucleotide elongation in the duplisome entails a large spatial separation of between the anticodon in the small subunit decoding and the duplicon in the large subunit nucleotidyl transfer center

(Figure 5-4). Unlike the baroque distribution of tRNA determinants used for enzymatic aminoacylation today, dRNA loading required approximation of anticodon and duplicon.

Martínez-Giménez and ### Tabarés-Seisdedos (2021) have proposed a solution to this recognition at a distance problem based on transient homodimers of identical dRNAs, bringing the anticodon of one into approximation with the duplicon of the other for loading (Figure 5-5). This mating of dRNAs multiplies the problem of maintaining all 64 dRNA monomers, to forming all 64 homodimers without inhibition or other interference from the 2016 = (64 x 63) / 2 possible heterodimers.

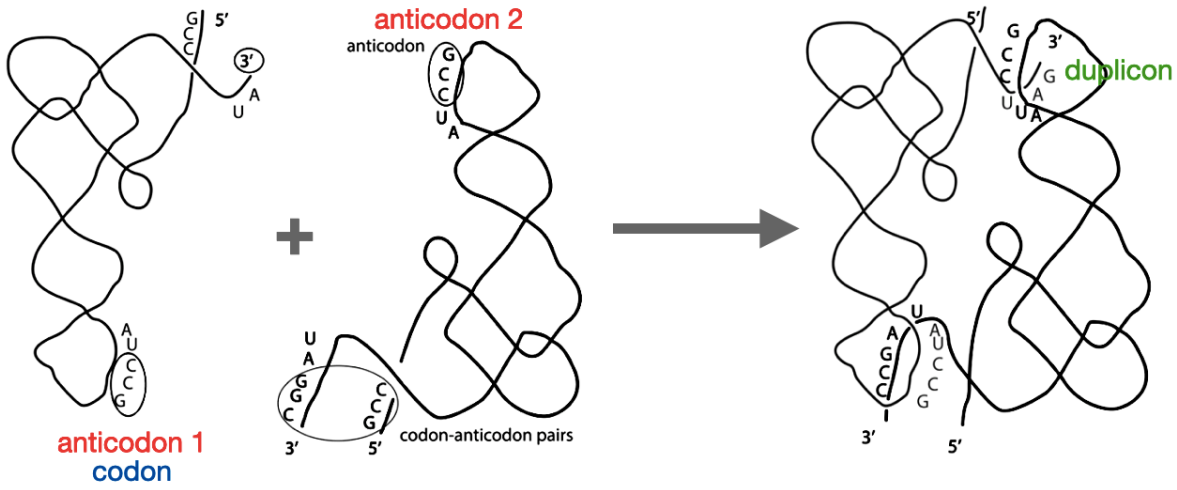


FIGURE 5-5. ANTICODON-DIRECTED dRNA LOADING FROM TRANSIENT HOMODIMERS (MARTÍNEZ-GIMÉNEZ & TABARÉS-SEISDEDOS 2021)

6. RNA copying | reverse complement

Unaware of Campbell (1991), Robert Weiss and Joshua Cherry shortly mooted a very different scheme of processive RNA copying inspired by what was then known of the ribosome (Weiss & Cherry 1993). Like Campbell, their duplisome reads the template from 5’ to 3’ in codon triplets and adds duplicon triplets to the nascent product. Whereas Campbell modeled his donor RNAs after tRNAs with central anticodon arm and terminal acceptor arm, Weiss and Cherry allowed donor oligoribonucleotides of random sequence and variable length. Their only requirement was a central triplet that functioned first as the anticodon to read the codon, and then as the duplicon added to the nascent polynucleotide. Because there is no dRNA loading step, codon reading is the only step of Watson-Crick pairing in the Weiss-Cherry elongation cycle, their scheme of RNA copying creates complements, not duplicates, of the template.

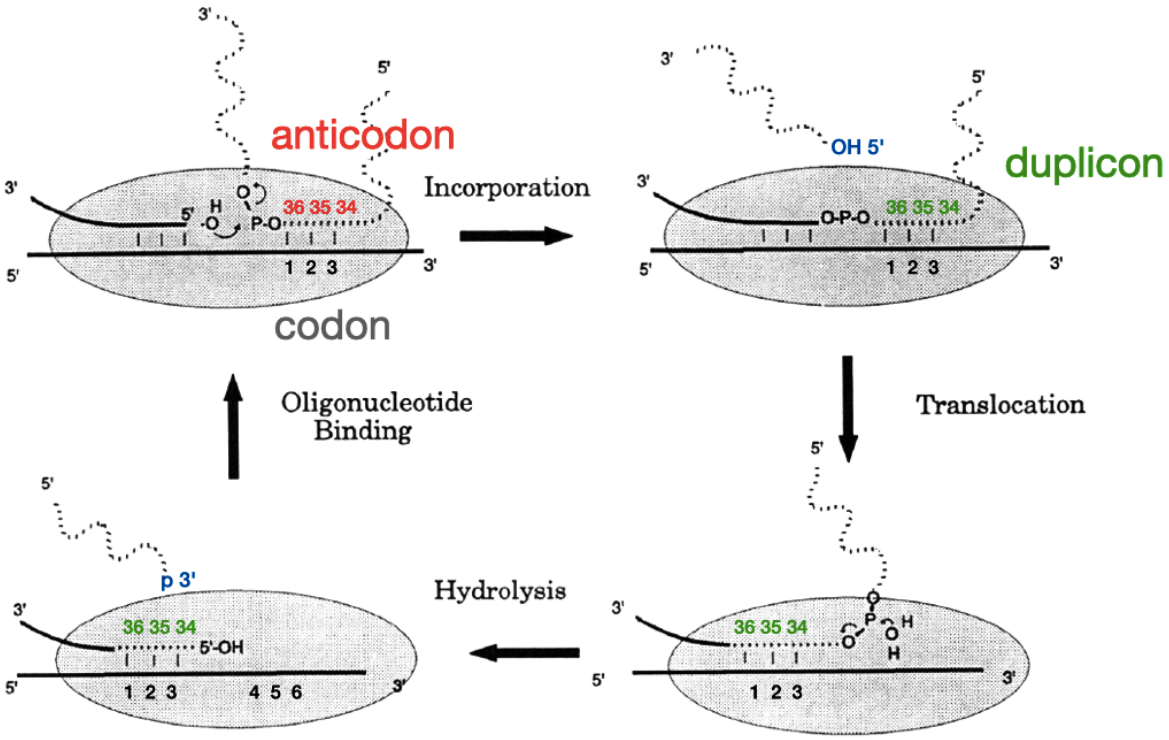


FIGURE 6-1. POLYNUCLEOTIDE ELONGATION CYCLE (WEISS & CHERRY 1993)

For ease of description as well as comparison with the other schemes of RNA copying, we refer to the anticodon (duplicon) triplet of the donor oligoribonucleotides as N34 N35 N36 although the actual length of the substrate and . is indeterminate.

donor oligonucleotide …32 33 34 35 36 37 38 …
incorporation …32 33 34 35 36 34 35 36 …

feedstock/tRNA = 1-33 p 3’ + 5’HO 34 35 36 p 3’ + 5’ HO 37-76

, constrained only by the requirement they have a triplet in the middle of the molecule that matches the codon. The same triplet functions first as the anticodon, reading the codon in the duplisome A-site, and then as the duplicon excised from the substrate oligonucleotide and added to the 5’ end of the nascent polynucleonucleotide.

…. two-step reaction that added the triplet from 3’ to 5’
transesterification with 5’ oxyanion attack on the phosphorus between N37 N36 with 5’ OH N37 N38 … leaving. [translocation] hydrolysis N33 N34 with

Although their primary object was to explain the origin of codon framing and translocation, they in fact proposed a general scheme of RNA copying from a pool of random oligonucleotides.

an oligonucleotide-consuming RNA replicase capable of processive template-directed elongation, including unwinding of template secondary structure ….

processive elongation along a template using decoding, sequence-specific, sequential binding of a substrate oligonucleotide, nucleotidyl transfer, and translocation

There were three apparent virtues of the Weiss-Cherry scheme of RNA copying: First, it required no special donor RNAs but used random oligonucleotides to read the codon from the template and donate the duplicon to the nascent polynucleotide. In particular, there is no requirement for dRNA reloading beyond the feedstock processes that nucleate and lengthen oligonucleotides by monomer addition, or oligomer ligation. Second, their scheme of RNA copying produces the familiar reverse complement of replicative polymerase enzymes (albeit made from 3' to 5' by triplet additions). The idea that duplexed RNA was a copying intermediate suggested evolutionary continuity, if not actual parsimony, from spontaneous through ribozymatic to enzymatic RNA copying.

Third, the Weiss-Cherry scheme had an immediate source of free energy within the elongation cycle itself to drive processive RNA copying. Thus, two phosphoester bonds were consumed to produce one new phosphoester bond, or net hydrolysis of one bond. In particular, they suggested that the duplisome tapped this free energy in the process of translocation to unwind duplexes or unfold other secondary structure in the templates, or separate template and product strands. Without this (or another) immediate source, they were skeptical that the initial inputs and final outputs of their scheme would afford a robust general process of RNA copying. Thus, reliance on mass action from a vast excess of random oligonucleotides would be vulnerable to deficiencies in the substrate pools. Similarly, reliance on extensive secondary and tertiary structure of substrates and products implied a difficult balance of template unfolding and refolding along with separation and folding of the nascent product all on the duplisome.

suggest that variants of this replicase, defective in one step of the polymerization cycle, were progenitors of the ribosome small subunit

[energetics] break two make one + acyl-ester to amide bond (Gordon 1995)

energetics: downhill hydrolysis followed by level transesterification

Weiss and Cherry (1991) inspired at least other schemes of RNA copying that made the reverse complement from 3' to 5' by excision and ligation of anticodon triplets to the nascent polynucleotide (Gordon 1995; Poole et al 1998; Yakhnin 2007). In "Were RNA replication and translation directly coupled in the RNA (+protein?) world, Karl Gordon altered the Weiss-Cherry scheme in three directions at once: First, he modeled the donor oligonucleotides as primitive tRNAs. Second, he modeled the nucleotidyl transfer reaction on group I self-splicing introns. And third, he allowed simultaneous peptidyl transfer.

() decoding center evolved as a decoding, nucleotidyl transfer center from intron

() not a general RNA copying but restricted to ….

() unknown ligase ribozyme to recreating the tRNA anticodon

() unknown ribozyme for aminoacylation of the 3' end of tRNA

() peptidyl transfer center of riboduplisome large subunitseparate and unrelated to the decoding/ nucleotidyl transfer center of small subunit

(1) rather than matching an internal triplet to the codon random oligomers matched to the codon … proposed that the donor RNAs resemble tRNAs with a definite length and position of anticodon …. the definite structure makes …. unlike the Weiss & Cherry scheme which made no assumptions beyond a feedstoack of random oligonucleotides, ribozymatic pathway to regenerate dRNAs … no constraint on the triplet anticodon …

feedstock/tRNA = 1-33 p 3' + 5'HO 34 35 36 p 3' + 5' HO 37-76

(2) model the chemistry of transesterification on group I self-splicing intron … propose homology

(3) proposed the dRNAs are charged by aa at 3’ end, thus tRNAs … non-specific charging by unspecified ribozyme …. random polypeptide synthesize pari passu with nascent polynucleotide … transpeptidation gives direction to tranesterification

simultaneous RNA replication and translation
decoding center was ligase for anticodon triplets excised from tRNAs
decoding center was ancestor of group I self-spicing introns
reversible transesterification driven by downhill transpeptidation

QUESTION how are oligomers made? how are dRNAs refilled with anticond/ duplicon in Gordon version

[Poole et al 1998; Jeffares et al 1998]

In “The path from the RNA world”, David Penny and colleagues presented a highly simplified amalgam of the Weiss-Cherry and Gordon schemes of RNA copying as an important predecessor of ribosomal protein translation. Their abstract promised “By focusing on the function of the protoribosome we develop a plausible model for the evolution of a protein-synthesizing ribosome from a high-fidelity RNA polymerase that incorporated triplets of oligonucleotides” (Poole et al 1998 p1). Their Figure 2 titled “An ancient RNA replicase as the precursor of the ribosome,” depicted the common features of both schemes, reading the template from 5’ to 3’ in codon triplets, and adding duplicon nee anticodon triplets to the nascent product from 3’ to 5’ so that the finished product is the reverse complement of the template. Cartooning the dRNA as a tRNA-like cloverleaf, they are agnostic whether duplicon transfer entails the Weiss-Cherry or Gordon reactions, or some altogether different mechanism of cleavage and ligation. Their one original suggestion is that a mechanism for aminoacylation and deacylation of the 3’ CCA of the acceptor arm provided a source of free energy to drive the polynucleotide elongation cycle both nucleotidyl transfer within the duplisome and the unknown anticodon reloading outside the duplisome … later co-opted to drive peptide bond formation by peptidyl transfer in the ribosome.

The paper of Poole et al (1998) appears to have anchored the duplisome of RNA copying in the RNA world in the literature.

ribo-organisms = RNA life

breakthrough organism = breakout of polypeptide translation

protoribosome = duplisome

[SECTION 14] the first coded polypeptides would by non-specific chaperone-like polypeptides rather than catalytic proteins … RNP (Poole et al 1998)

Yakhnin 2007

7. RNA copying | degenerate forward duplicate

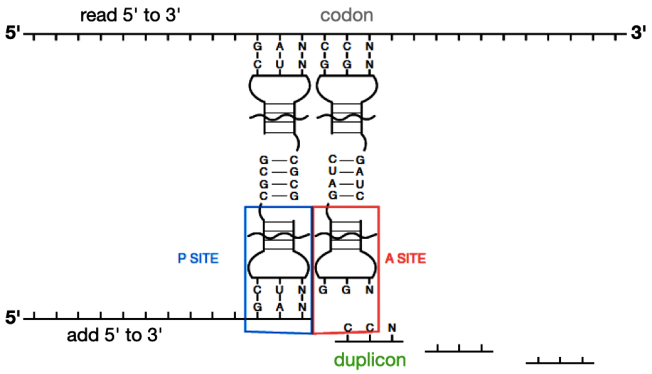


FIGURE 7-1. POLYNUCLEOTIDE ELONGATION CYCLE (NOLLER 2012)

In a remarkable decade, ribosomes in various stages of protein translation were characterized at atomic resolution by X-ray crystallography, and then cryo-electron microscopy (Ban et al 2000). In light of this more detailed picture of decoding, peptidyl transfer and translocation, Harry Noller at the UCSC Sinsheimer Laboratories pondered what functions of the ribosome and its tRNAs that may have predated protein translation. In his talk on the “Evolution of protein synthesis from an RNA world”, he conjectured that the decoding center of the small subunit rRNA evolved for RNA copying, and was later exapted for polypeptide translation (Noller 2012). In his scheme, the RNA template is read from 5’ to 3’ in codon triplets and its forward duplicate is made from 5’ to 3’ by addition of free duplicon triplets by an unknown ligase ribozyme (Figure 7-1). In lieu of dRNA reloading *outside of the duplisome* (Campbell 1991), or two entirely separate rounds of RNA copying (Weiss & Cherry 1993; Gordon 1995), Noller invoked two similar or identical decoding centers within the duplisome, one for decoding the template, and another for loading the duplicon. Finally, the forward duplicate avoids troublesome intermediates of other copying schemes, *viz*. there is no useless minus strand to copy again, much less a long duplex to unwind (Table 7-1).

	long duplex	minus strand
forward duplicate	NO	NO
reverse duplicate	NO	YES
reverse complement	YES	YES

TABLE 7-1. INTERMEDIATES IN RNA COPYING

In his scheme of RNA copying, Noller required that dRNAs form stable homodimers, so that the anticodon of one subunit reads the codon, while the identical anticodon of the other subunit loads the cognate duplicon for ligation to the nascent product. He noticed that with third position superwobble, just 16 dRNA isoacceptors could read all 64 codon triplets, and similarly, there are exactly 16 self-complementary quadruplets, or palindromes of four nucleotides. Without suggesting any particular relation between anticodons and palindromes, Noller proposed that dRNAs were held together by such quadruplets at their 3' end (Figure 7-2). While stable matings of 16 homodimers by such self-complementary tails is more energetically and kinetically plausible than the transient matings of 64 homodimers (*pace* Martínez-Giménez & Tabarés-Seisdedos 2002, 2021), anticodon and palindrome codes would likely drift apart without strong selection. But our criticism of this Rube Goldberg contraption just buries the lede: *the conjecture that every third nucleobase of the product is unspecified, or degenerate, vitiates its usefulness as a process of RNA copying.*

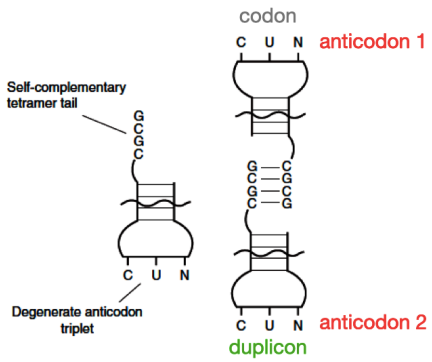


FIGURE 7-2. 16 STABLE dRNA HOMODIMERS (NOLLER 2012)

8. RNA copying | faithful forward duplicate

Perhaps because the permuted reverse duplicate is unprecedented, if not frankly bizarre, Campbell never wrote again about the possible origin of the ribosome and tRNAs as a means of RNA copying, and indeed, his paper was quickly lost in the literature.[7] The Weiss-Cherry and Gordon schemes of RNA copying briefly attracted some interest because their product is the familiar reverse complement of replicative polymerase enzymes (albeit made from 3' to 5' by triplet additions). As more was learned of protein translation, however, these schemes seemed less and less likely as ancestors of the ribosome and tRNAs. Finally, no one followed up on Noller's scheme for degenerate RNA copying, and indeed, it is unclear whether Noller himself regarded it as more than a fanciful notion for a pending talk.

Coming upon the newly published Noller (2012), then tracking down Weiss & Cherry (1993), and finally exhuming Campbell (1991), we were struck by all three explanations of the origin of protein translation from a lost process of RNA copying. Framing the problem, these authors proposed clear solutions that seem to fail not for obscure reasons, but patent defects. Here we present an equally simple scheme of RNA copying that retains virtues of all three without those defects. *The key difference of our proposal from Campbell (1991) is that the first two nucleotides of loaded dRNAs, not their final three nucleotides, comprise the duplicon.* Two obvious virtues of this revised dRNA structure are that the product of RNA copying is a faithful forward duplicate made 5' to 3' (Figure 8-1), and just 16 dRNAs are needed to copy any template, not Campbell's 64 isoacceptors nor Noller's 16 degenerate homodimers.

[7] Among authors speculating about RNA copying mediated by an ancestor of the ribosome, or its small subunit, who failed to notice Campbell (1991) are Weiss & Cherry (1993), Gordon (1995), Poole et al (1998), Bernhardt (2012), Noller (2012).

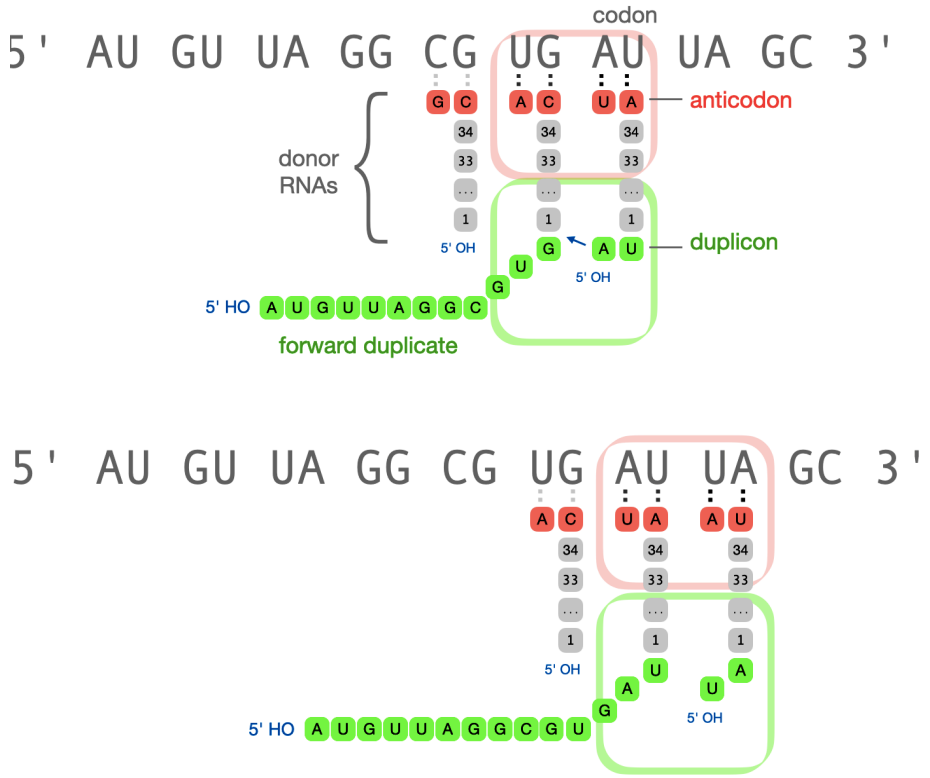


FIGURE 8-1. POLYNUCLEOTIDE ELONGATION CYCLE (PROVISIONAL)

The provisional scheme depicted in Figure 8-1 is just the start of our investigation, not the conclusion. Like a row of dominos standing upon end, the seemingly minor change in the size and location of the duplicon triggers a cascade of problems for the polynucleotide elongation cycle, and raises unexpected questions about the evolution of ribosome life. In the next three sections we discuss the structure of dRNAs and their ribozymatic loading from random oligonucleotides (section 9), the elongation and termination reactions of the duplisome nucleotidyl transfer center (section 9), the energetics and kinetics of the elongation cycle (section 10), and the mechanism of decoding (section 11). In the final three sections we discuss the origin of the duplisome (section 12), the origin of tRNAs (section 13), and the breakout of polypeptide translation (section 14).

9. From dRNA loading to polynucleotide transfer

To turn our scheme of RNA duplication into a definite elongation cycle subject to experimental tests, we must specify how the nascent polynucleotide was transferred within the duplisome, and how the freed dRNA was reloaded afterward. Moreover, as any plausible mechanism of dRNA loading must be strictly simpler than our overall copying scheme, we cannot invoke some vague form of templated ligation or primer extension. After pondering the biochemistry of polynucleotide transfer and dRNA loading separately, we have converged on one parsimonious solution to both problems. Here we explain how dRNAs are loaded from the pool of random oligomers in a condensation-hydrolysis sequence catalyzed by ribozyme P, progenitor of the catalytic RNA of RNase P. Conversely, polynucleotides are transferred in a hydrolysis-condensation sequence, catalyzed by the duplisome nucleotidyl transfer center, progenitor of the ribosome peptidyl transfer center.

The two key ideas in polynucleotide transfer and dRNA loading are: (1) tight control of water at the active site allows the downhill hydrolysis to respectively push or pull the uphill condensation, and (2) the sites of hydrolysis and condensation are staggered by one duplicon between the P-site and A-site dRNAs in transfer, or along the single dRNA in loading. This solution for the cycle of RNA elongation solves the *water problem* of polymer condensation in bulk solvent, and frames new questions about the energetics of duplisome life (section 10). Finally, considering basic problems of recombination and repair in the RNA world, we trace the duplisome nucleotidyl transfer center back to a primordial ligase ribozyme, while we trace ribozyme P, and perhaps the decoding center itself, back to a complementary repair ribozyme (section 12).

Rather than solving the matching at a distance problem with Rube Goldberg homodimers, we propose each dRNA had two conformations, *closed* and *open*. Anticodon and duplicon come together in the closed conformation for dRNA loading, then come apart in the open conformation for decoding and polynucleotide transfer, respectively, at spatially separated sites. To be definite, we model the ancestral dRNA as nucleotides 1-36 of a modern tRNA, with a 15 base-pair stem in the closed conformation, capped by an 8 nucleotide D-loop (14-21), for a total of 38 nucleotides,

including the duplicon (Figure 9-1). Using this numbering, the duplicon N2 N1 pairs with the anticodon N35 N36, N1 pairs with N34, and so forth. Our choice of dRNA structure, motivated by the need to pair duplicon and anticodon for loading, imposes remarkable constraints on the entire scheme of RNA duplication (sections 8-11). From this, we reconstruct the rich history of polymer life, from the origins of RNA duplication, to the breakout of polypeptide translation, and onward, to the retirement of the duplisome (sections 12-14).

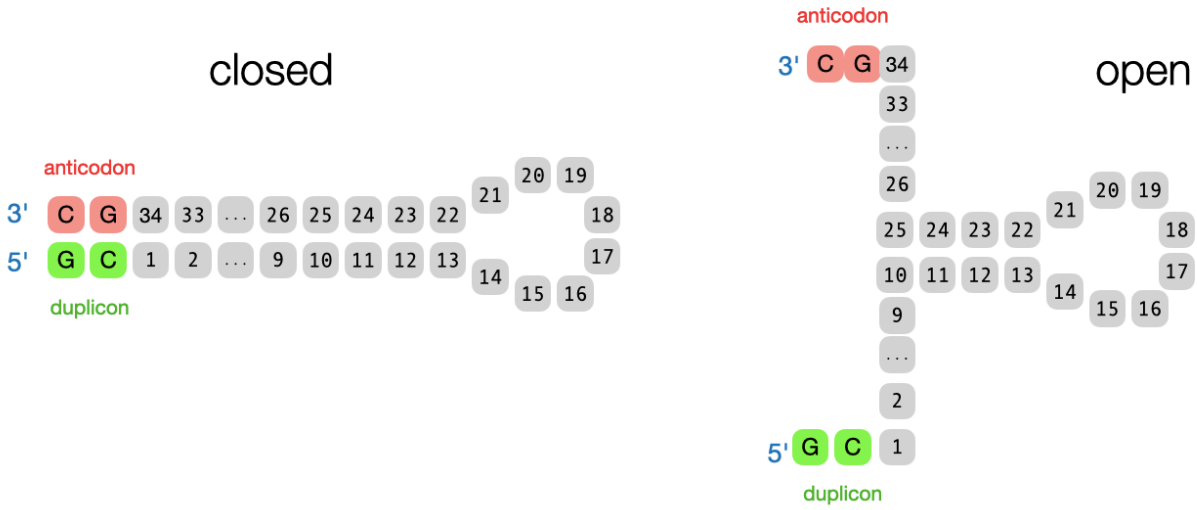


FIGURE 9-1. dRNA CLOSED AND OPEN CONFORMATIONS

- multi-turnover
- substrate specificity and promiscuity

Found in all kingdoms of cellular life, the ribozymatic subunit of RNase P removes the 5’ leader of pre-tRNAs and eclectic other substrates (Guerrier-Takada et al 1983; Gößringer et al 2021; Phan et al 2021). In bacteria, RNase P cleaves pre-tRNAs, pre-tmRNA, pre-SRP RNA, pre-rRNA and certain mRNAs. In eukaryotes, gene duplication of the catalytic RNA gave rise to two very similar complexes called RNase P and RNase MRP that share most of their accessory proteins (Welting et al 2006; Coughlin et al 2008). The former cleaves pre-tRNAs and certain pre-snoRNAs, while the latter cleaves pre-rRNA and mitochondrial primers, as well as certain lncRNAs and mRNAs.

In both natural and artificial substrates, an unpaired 5' leader is cleaved from a paired stem, e.g., a simple helix and loop for pre-SRP RNA, or two coaxial helices and loop for pre-tRNAs, leaving hydroxyl and phosphate at the 3' and 5' ends, respectively (McClain et al 1987; Kirsebom & Trobro 2009). Studies on pre-tRNAs and their mimics have identified key features in substrate recognition, positioning and hydrolysis. Two features at some remove from the scissile bond assist binding of pre-tRNAs to bacterial RNase P. First, in what has been dubbed a *molecular ruler*, the pre-tRNA elbow binds the interdigitating T-loop motif of the specificity domain, placing the topological junction of leader/stem at about the right position in the catalytic domain for trimming (Chan et al 2013; Lehmann et al 2013; Zhang & Ferré-D'Amaré 2016a,b). Second, sequence-independent interactions of the distal leader with the accessory protein increase the affinity of RNase P for pre-tRNAs relative to mature tRNAs, allowing product release, and preventing product inhibition.

Once interactions with the pre-tRNA elbow and distal leader form the substrate encounter complex with RNase P, additional interactions dock the scissile bond at the active site in the catalytic conformation (Lan et al. 2018; Zhu et al 2022). The preferred cleavage site is not specified by the nucleotide sequence, but by the topological junction between the single-strand leader and double-strand stem. The crucial steps are unwinding any fortuitous pairing between leader and trailer preceding the mature stem, and rotating nucleotide U51 of helix P4.[8] Unwinding entails stacking A248 of junction J5/15 on base-pair N1:N72, thereby unstacking N73 from the acceptor stem and exposing the Hoogsteen edge of A248 to pair with N1. In bacteria, where the terminal 3' C74 C75 A76 of the mature tRNA are encoded in the primary transcript, G292 G293 U294 in the internal loop of P15[9] pair with these nucleotides, holding the proximal trailer flayed away from the 5' leader of the pre-tRNA (Kirsebom & Svard 1994). With the substrate unwound, U51 rotates into the active site, coordinating one Mg^{++} that generates the hydroxide nucleophile, while a second Mg^{++} assists the leaving

[8] Numbers of the P RNA sequence are from *Escherichia coli*.

[9] Internal loop L15 is absent in the P RNA of eukarya and many archaea?, consistent with the post-transcriptional addition of the terminal 3' CCA after pre-tRNA processing by RNase P.

group departure. After S_N2 hydrolysis, the leader quickly dissociates, while a slower return from catalytic to encounter conformation releases the mature tRNA (Tallsjo & Kirsebom 1993). Under physiological conditions, the holoenzyme completes about … turnovers per second.

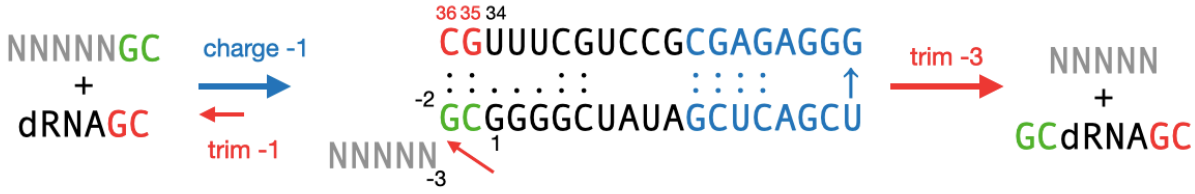


FIGURE 9-2. dRNA LOADING FROM COGNATE OLIGOMER

We propose that dRNAs were reloaded from the pool of random oligomers in a coupled condensation-hydrolysis sequence catalyzed by an ancestor of P RNA called *ribozyme P* for sake of discussion. We propose this ribozyme loaded the free dRNA from a cognate oligomer via the sequence of docking-condensation-redocking-hydrolysis-undocking. For convenience, we refer to the covalent steps of condensation and hydrolysis as *charging* and *trimming*, respectively, and the complete sequence as *loading* (Figure 9-2). Complexed with ribozyme P, the dRNA anticodon samples the pool of oligomers for a match. At the end of the loading sequence, the input oligomer is shortened by two nucleotides at its 3' end, while the free dRNA is now loaded with these same two nucleotides at its 5' end. There is no net change in the number of RNA molecules or phosphoester bonds, but free energy of base-pairing between duplicon and anticodon (- 4-6 kcal /mol) favors charging cognate oligomers over non-cognate ones, and drives the overall reaction toward loading.

Our model of dRNA loading conjectures that (1) the tightly coupled, condensation-hydrolysis sequence was a quasi-reversible process, and that (2) the active site shifted between the two catalytic steps. The irreversible maturation of pre-tRNAs by RNase P, *viz*. docking-hydrolysis-undocking, is a vestige of dRNA trimming; there is no similar vestige of dRNA charging. If our reaction sequence is thermodynamically novel, pulling

the first step (condensation) uphill via a second step (hydrolysis) downhill, self-splicing introns provide examples *par excellence* of staggering reaction sites by redocking the intermediate (refs).

In dRNA loading, the scissile bond rewinds from the site of condensation N1↑N1 to the site of hydrolysis N3↓N2. Thus, A248 stacks on N1:N34 during charging, and re-stacks on N2:N36 during trimming. Other interactions used for pre-tRNA positioning on RNase P were different or absent in dRNA loading. The length of the closed dRNA stem, say 13 or 15 basepairs for charging and trimming, respectively, was comparable to the 12 or 13 basepairs in the coaxial helices of acceptor- and T-arms in pre-tRNA trimming today. But only a primitive D-loop, not the modern D/T elbow was available to position the scissile bond by a ruler mechanism. Whereas pre-tRNAs have a sizable 3' trailer, beginning N73 C74 C75 A76 in bacteria, dRNAs had only the unpaired anticodon N35 N36 during charging, and no trailer at all during trimming. Finally, ribozyme P had no coded accessory protein in the RNA world.

There is indirect evidence from comparative genomics, and direct evidence from mutant pre-tRNAs, that the preferred site of RNase P cleavage can be wound in either direction by matches that lengthen the paired stem, or mis-matches that shorten it. In some cases both a decrease in cleavage at the canonical site and an increase in off-site cleavage have been observed. A natural -1 shift in the cleavage site due to pairing with the unusual discriminator C73 creates the 8 base-pair acceptor stem of bacterial $tRNA^{His}$, compared to the 7 base-pair stem of most tRNAs (Orellana et al 1986). In rat nuclear pre-$tRNA^{Lys}$, base substitutions that leave N1:N72 paired are well-tolerated, while substitutions that mis-pair these nucleotides reduce cleavage at the canonical site without an apparent increase in +1 cleavage (Paisley & Van Tuyle 1994). In yeast nuclear pre-tRNAs a mismatch at -1/73 appears important to prevent leader-trailer pairing …. (Lee et al 1997).

Kinetic and structural studies on cleavage site selection in pre-tRNAs are hard to interpret because remote features, not just the topology and sequence of the cleavage site, affect substrate docking and catalysis. Substrates lacking the specific elbow and trailer interactions of pre-tRNAs may better illustrate topological docking of the scissile

bond. In *E. coli* pre-SRP RNA, the 5'-leader is cleaved at the same site U24↓G25 whether from the intermediate hairpin H1 formed after transcription of the first 36 nucleotides, or from the mature hairpin formed after transcription of all 138 = 24 + 114 nucleotides (Fukuda et al 2020). In living cells, the leader is likely removed co-transcriptionally once the H1 hairpin forms.

Minimal substrates retaining few of the features of pre-tRNAs …. Thus, bacterial P RNA can efficiently cleave the (length?) 5' leader from a model substrate with a 12 basepair stem and 7 nucleotide loop, and 5 nucleotide 3' trailer NCCAN (McClain et al 1987). A partial shift to -1 cleavage was seen in an artificial hairpin-loop substrate with ?leader? 4 bp stem, GAAA tetraloop, and 3' trailer CCAC (Brännvall et al 2007).

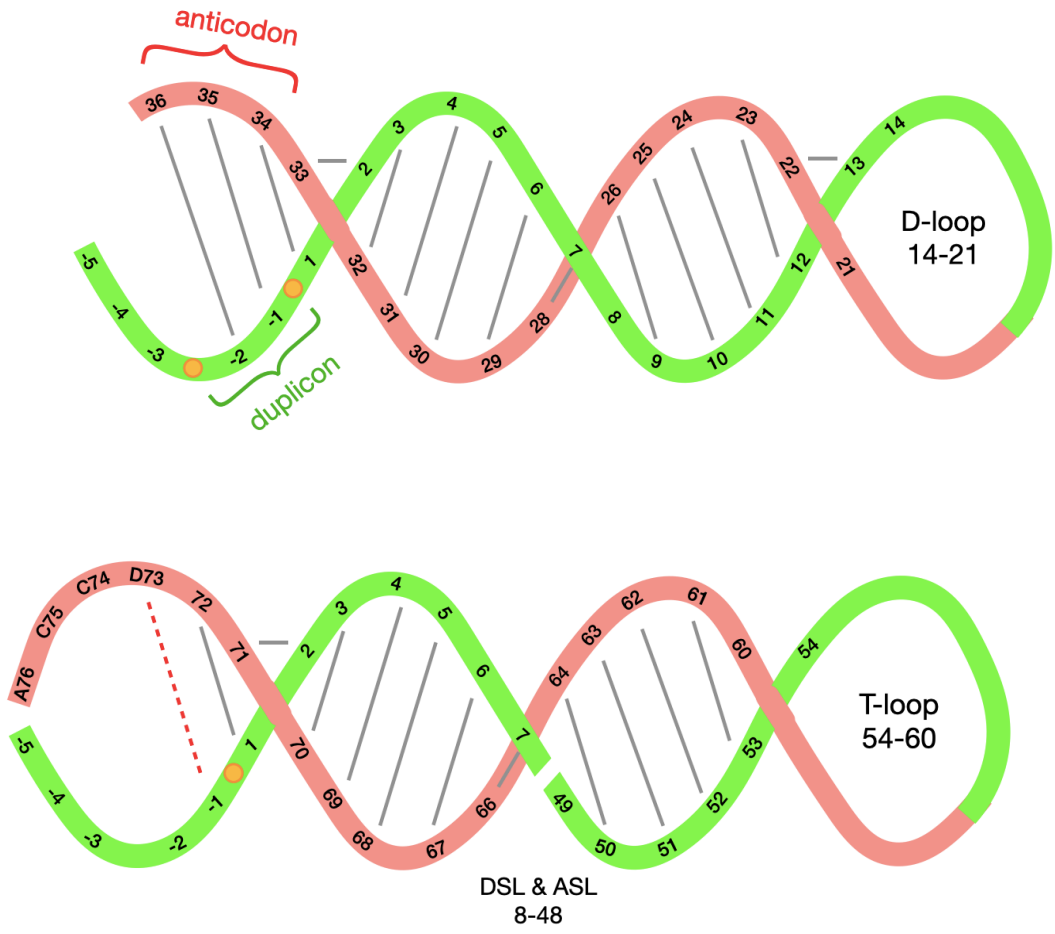


FIGURE 9-3. dRNA LOADING (TOP) & pre-tRNA MATURATION (BOTTOM)

In Figure 9-3 we compare dRNA loading by ribozyme P and pre-tRNA maturation by RNase P. Scissile phosphates are shown as brown circles. The closed dRNA is modeled as an RNA duplex topped by D-loop (Dickerson et al 1982; Saenger 1984). It has been charged at N1↑N1 with a cognate oligomer, and awaits trimming at N3↓N2. dRNAs end with N36, the third position of the anticodon. In the pre-tRNA, the acceptor and T arms coaxial helices topped by the T-loop, where a gap marks the D and anticodon arms. The pre-tRNA awaits trimming at N1↓N1. The critical mismatch or bulge between N1 and N73 is shown by a dashed red line.

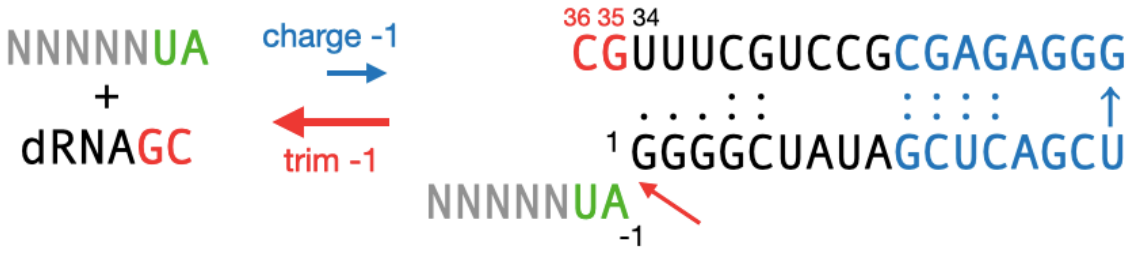


FIGURE 9-4. dRNA CHARGING WITH NON-COGNATE OLIGOMER

Until the excess leader is trimmed away, an overloaded dRNA likely cannot enter the duplisome decoding center, much less accommodate the nucleotidyl transfer center. But misloaded dRNAs whose duplicon mismatches the anticodon at one, or rarely both positions, were a likely source of substitution errors, and might even cause misreading of codons (section 11). For non-cognate oligomers, the ratios of forward and reverse reactions are shifted in both charging and trimming (Figure 9-4). Thus, *mispaired* dRNAs are more likely to dissociate rather than mischarge, and *mischarged* dRNAs are more likely to hydrolyze at N1↓N1 without winding to N3↓N2 for trimming. Interestingly, *misloaded* dRNAs still might be *retrimmed* via a second hydrolytic cleavage at N1↓N1 (Figure 9-5). As the loss of one phosphoester bond separates the trimming at N3↓2 from the retrimming at N1↓N1, this rejection pathway for *misloaded* dRNAs is a simple form of kinetic proof-reading (Hopfield 1974; Ninio 1975).

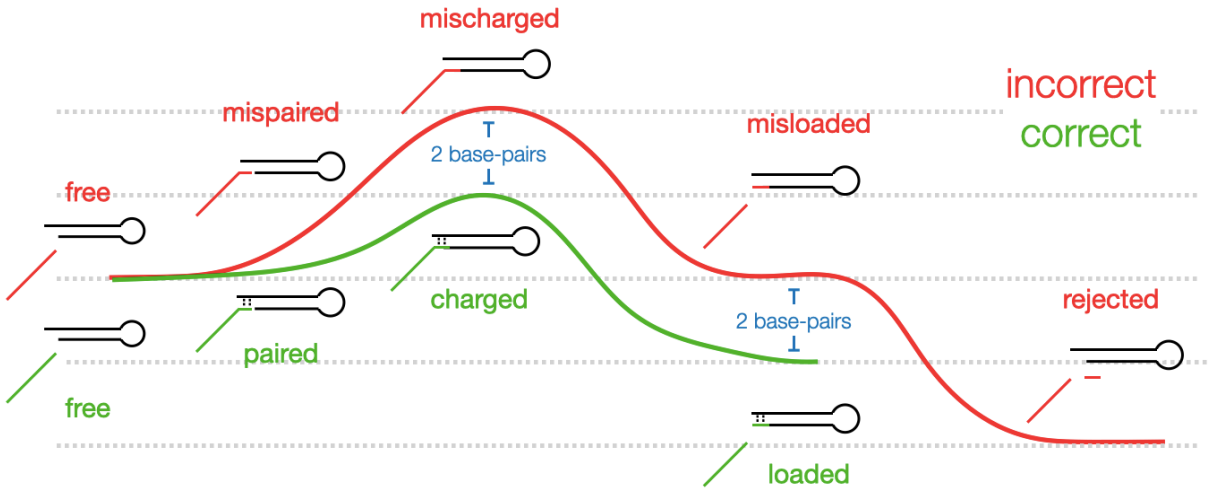


FIGURE 9-5. REJECTION OF MISLOADED dRNA

Our scheme of dRNA loading catalyzed by ribozyme P requires a phosphate at the 5’ end of free dRNAs, and leaves one at the 5’ end of loaded dRNAs. One plausible mechanism of polynucleotide transfer uses a 5’ hydroxyl of loaded dRNAs for nucleophilic attack, either directly on the polynucleotidyl-dRNA phosphodiester bond, or on the polynucleotidyl 2’,3’ cyclic phosphate intermediate formed by strand scission as in self-cleaving ribozymes. But to use this 5’ oxyanion as nucleophile, we must somehow remove the 5’ phosphate from the loaded dRNA *before it enters the A-site*, and restore the 5’ phosphate to the free dRNA *after it leaves the E-site*. Thus, loading and transfer reactions could be coupled together via a *dRNA 5’-phosphotransferase* that shuttles phosphate between loaded and free dRNAs (Figure 9-6). Or, rather than this isoergonic phosphate shuttle, a *duplicon-dRNA 5’ phosphatase* and *free-dRNA 5’ kinase* could work in tandem to drive the elongation cycle from some high-energy phosphate donor.

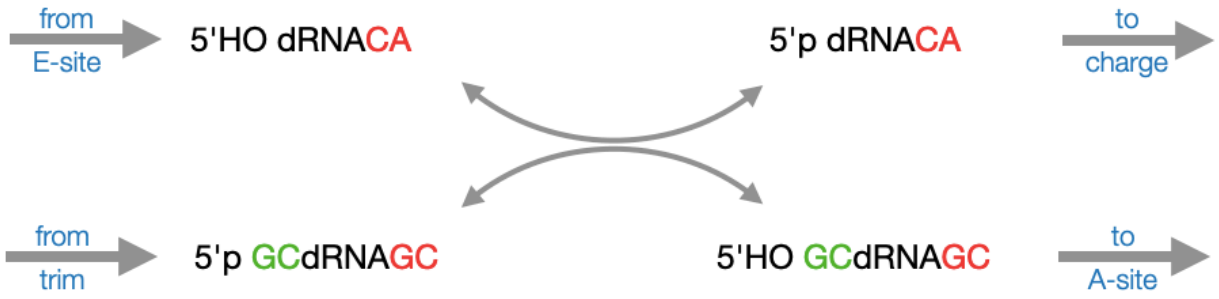


FIGURE 9-6. dRNA 5' PHOSPHOTRANSFERASE

Figure 9-7 depicts the substrates and products of nucleotidyl transfer by concerted trans-esterification where the 5' oxyanions of the duplicon-dRNA and free dRNA are the inline nucleophile and leaving group, respectively. The reaction passes through a trigonal bipyramidal (sp^3d) transition state, or a more stable phosphorane intermediate. The electron movements are complemented by a series of proton transfers, or *proton shuttle*, mediated by the substrates, transfer center, and water. After transfer, the P-site dRNA begins with 5' OH N1, while the A-site dRNA carries the nascent polynucleotide chain elongated by one duplicon. Whereas peptidyl transfer in the ribosome is exergonic because of the greater stability of amide than acylester, polynucleotide transfer is nearly isoergonic, with no net change in the number or kinds of covalent bonds.

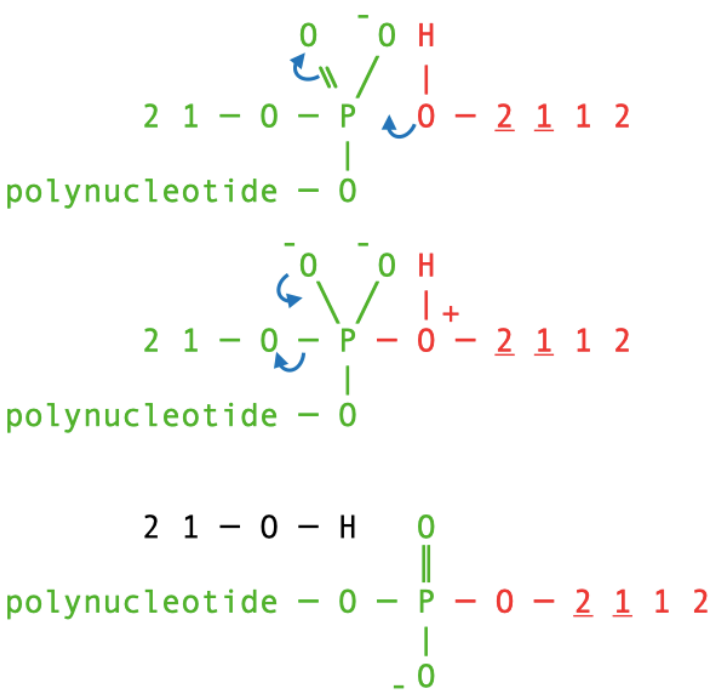


FIGURE 9-7. NUCLEOTIDYL TRANSFER BY CONCERTED TRANS-ESTERIFICATION

Reversible strand scission in self-cleaving ribozymes (section 3) suggest a two-step pathway for nucleotidyl transfer mediated through a polynucleotidyl 2',3' cyclic phosphate intermediate (Figure 9-8). Converting the abortive cycle of self-cleaving ribozymes into a productive sequence, polynucleotidyl-dRNA in the P-site undergoes strand scission by the vicinal 2' OH to free the 5' OH dRNA and retain the nascent polynucleotide with 2',3' cyclic phosphate. In the second step, the 5' OH duplicon-dRNA in the A-site attacks the strained cyclic phosphodiester bond, recreating a polynucleotidyl-dRNA, now elongated by the new duplicon and its cognate dRNA.

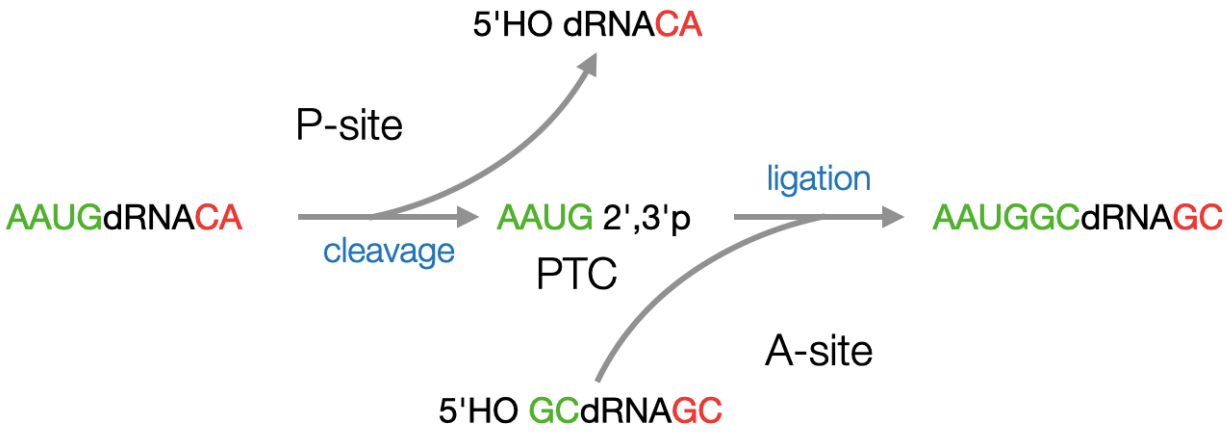


FIGURE 9-8. NUCLEOTIDYL TRANSFER BY SEQUENTIAL TRANS-ESTERIFICATION

The elongation cycles for RNA duplication outlined above use ribozyme P for dRNA loading and the 5’ oxyanion of duplicon-dRNA as nucleophile for nucleotidyl transfer. A serious, perhaps fatal drawback of these schemes to couple dRNA loading and polynucleotide transfer is that they require one (*phosphotransferase*) or even two (*phosphatase*, *kinase*) extinct ribozymes for manipulating dRNA 5’ ends, *in addition to ribozyme P known through extant RNase P*. What if instead we simply retain the 5’ phosphate of the newly loaded dRNA as it enters the duplisome A-site? This alternative nucleotidyl transfer reaction, or *trans-phosphorylation* is formulated in Figure 9-9. Like trans-esterification, this reaction is isoergonic, but unlike the former, it is sterically impossible as a concerted mechanism. Below we propose a sequential mechanism of trans-phosphorylation without precedent in natural or artificial ribozymes.

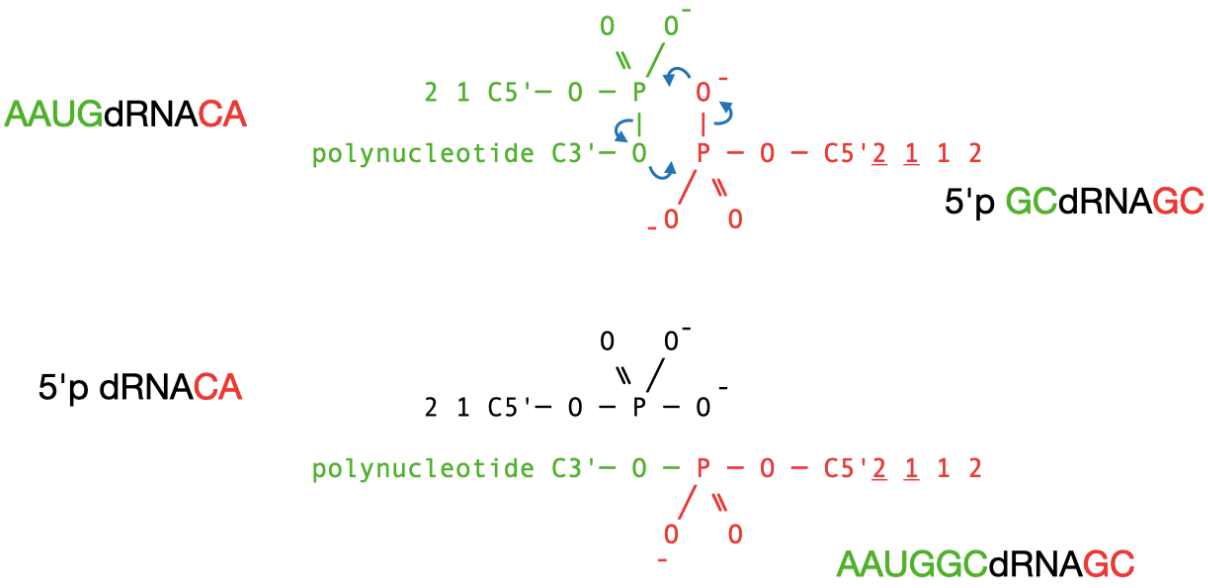


FIGURE 9-9. CONCERTED TRANS-PHOSPHORYLATION

Inspired by the conjectured condensation-hydrolysis sequence in dRNA loading catalyzed by ribozyme P, we now consider a hitherto obscure possibility, *viz*. the 3' OH of the nascent polynucleotide attacks the 5' phosphate of the duplicon-dRNA as the *second step* of a hydrolysis-condensation sequence, confined from bulk solvent within the duplisome nucleotidyl transfer center. So long as free water is kept from the reaction center, there seems no obstacle to this tightly coupled hydrolysis-condensation sequence with a suitable proton shuttle (Figure 9-10).

Here we propose that by carefully controlling solvation, the duplisome transfer center catalyzed polynucleotide elongation via a nucleotidyl hydrolysis-condensation sequence, dehydration improbable in free aqueous solution, made probable in ptc … coupled by limiting the free entry of water and directed by P-site closing, the converse of oasis … dry island in a sea of water [?proton shuttle]

IDEA reread Weiss & Cherry 1992 on hydrolyze/condense sequence

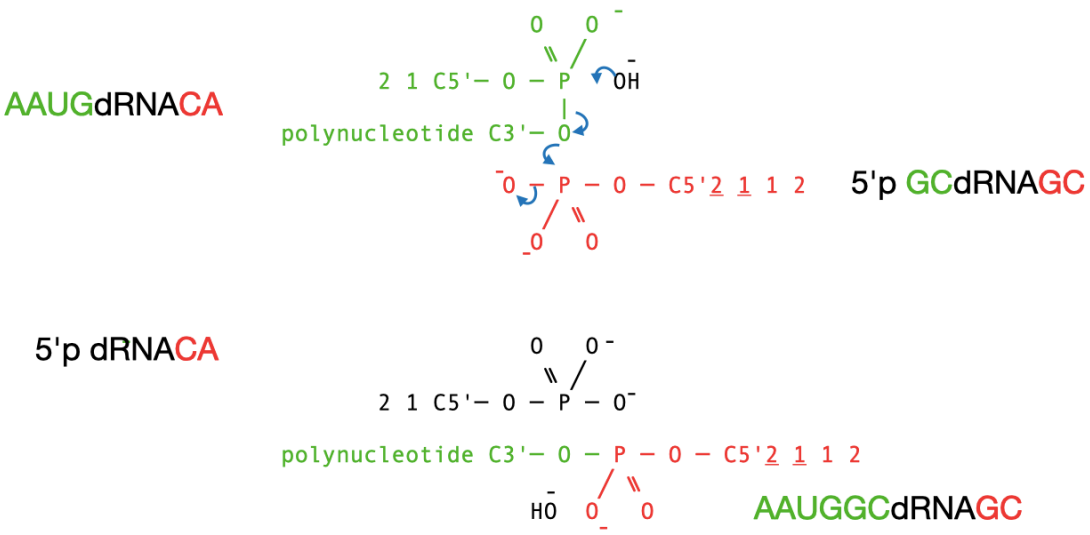


FIGURE 9-10. SEQUENTIAL HYDROLYSIS AND CONDENSATION

In Figure 9-11, we depict the catalytic events in (left) dRNA loading by ribozyme P and (right) polynucleotide transfer by the duplisome center. Small blue arrows connect each nucleophile (arrow-tail) to its electrophile (arrow-head). P-site and A-site reactants in the duplisome are highlighted in orange and blue, respectively. In dRNA loading condensation precedes hydrolysis, while in polynucleotide transfer hydrolysis precedes condensation. If hydrolysis is not followed immediately by condensation, say when free water enters via an empty A-site, elongation terminates in release of the nascent polynucleotide through the P-site exit tunnel.

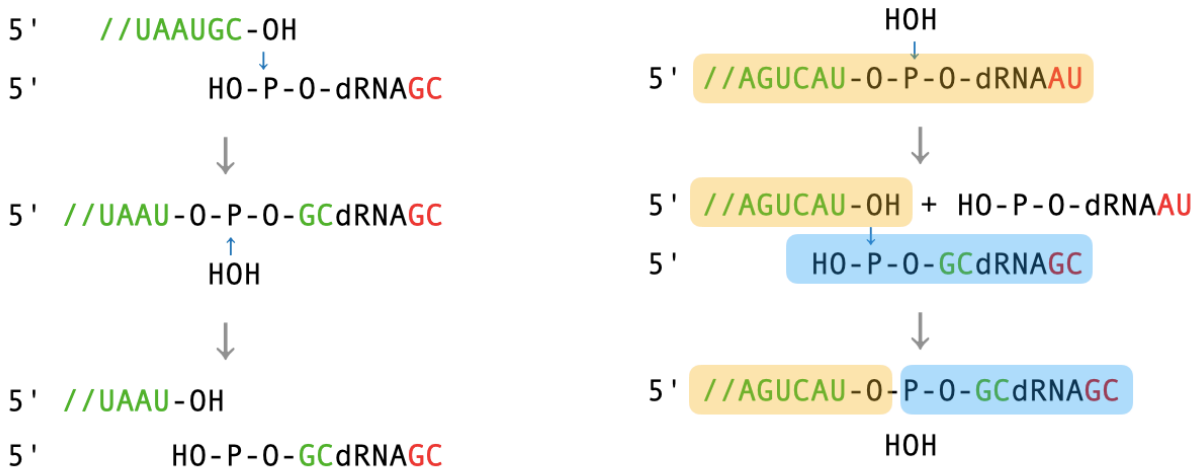


FIGURE 9-11. SUMMARY OF RNA ELONGATION CYCLE

The peptidyl transfer center at the heart of the ribosome large subunit is a quasi-symmetrical pocket formed from the P- and A-regions of the central loop of domain V, about 180 nucleotides overall that likely arose by a tandem duplication (Krupkin et al 2011). Possibly the oldest part of the large subunit rRNA, this center likely originated as a free-standing ribozyme in the RNA world (Bokov & Steinberg 2009; Petrov et al 2014). One suggestion is that this *protoribosome* played a role in random polypeptide synthesis, catalyzing the same ester-to-amide transfer from peptidyl- to aminoacyl-RNA substrates as modern ribosomes (Agmon et al 2009; Tamura 2015). If an earlier ancestor of this *peptidyl transferase ribozyme* acted on purely RNA substrates, that function was lost before the protoribosome began accretion of RNA extensions and functions along its way to becoming the large subunit rRNA.

Two observations have been taken as evidence for a free-standing peptidyl transferase ribozyme as progenitor of the ribosome large subunit. First, artificial ribozymes can catalyze peptide bond formation by positioning model substrates relative to one another (?Weber & Orgel 1980; Lohse & Szostak 1996; Zhang & Cech 1997; Tamura & Schimmel 2003). In another model reaction, the α-amine of phenylalanine, tethered to the ribozyme 5’ end can attack the carbonyl carbon of aminoacyl-esters on the 3’ OH of adenosine monophosphate, albeit with 20,000-fold

slower turnover than the ribosome elongation cycle (Zhang & Cech 1997). Compared to water or alcohols, the non-protonated primary amine is a strong nucleophile with a pK_a about 8. The tetrahedral transition state has a relatively low activation energy, and the products are energetically favored. [aa-AMP models]

Second, like the peptidyl transferase ribozymes selected *ab initio*, fragments of the P- and A-regions of the ribosome peptidyl transfer center itself can catalyze peptide bond formation from aminoacyl-tRNA minihelices or even simpler substrate mimics (Bose et al 2022; Kawabata et al 2022). Heterodimers of the P- and A-region cores, or homodimers of the P-region core alone, are active. Having striped away the later additions to the ribosome and its tRNAs, the activities of these core fragments of the peptidyl transfer center, and the tRNA acceptor arms, are proposed to exhume the ancient peptidyl transferase ribozyme and it substrates in a lost process of random polypeptide formation. “Although doubts and caveats remain, Yonath’s and Tamura’s work seems to recapitulate a milestone on the road from primordial organic molecules to the ribosome used by the last common ancestor or all living things” (Dance 2023).

We have now proposed that the progenitor of the ribosome peptidyl transfer center was a duplisome nucleotidyl transfer center that catalyzed sequential hydrolysis-condensation of polynucleotides. *If so, the backward extrapolation from the ribosome peptidyl transfer center to a protoribosome that made random polypeptides is mistaken.* In section 12 we trace the proposed duplisome nucleotidyl transfer center backward to a primordial ligase ribozyme, the workhorse of sequence-independent recombination and repair in early RNA life. In sections 13 and 14 we trace the duplisome nucleotidyl transfer center and its dRNAs forward to the ribosome peptidyl transfer centers and its tRNAs through two new reactions, polynucleotide tagging and polypeptide elongation (Table 9-1).

Reaction		P-site	A-site	ΔG kcal / mol
polynucleotide	ligation	5' polynucleotide	3' polynucleotide	net -1 bond
polynucleotide	elongation	polynucleotidyl-dRNA	duplicon-dRNA	0
polynucleotide	release	polynucleotidyl-dRNA	empty	net -1 bond
polynucleotide	tagging	polynucleotidyl-dRNA	aminoacyl-tRNA	0
polypeptide	elongation	polypeptidyl-tRNA	aminoacyl-tRNA	- ?
polypeptide	release	polypeptidyl-tRNA	empty	net -1 bond

TABLE 9-1. EVOLUTION OF THE POLYMER TRANSFER CENTER

Our proposed origin and evolution of the polymer transfer center helps explain several features of the ribosome peptidyl transfer center including its catalytic mechanism, natural substrate promiscuity, and artificial substrate plasticity. Most enzymes are enthlapic catalysts that lower the free energy of the transition state through non-covalent bonds to its protons (general acid/general base), monopoles (ionic bonds), dipoles (electrostatic bonds), or electrophiles/nucleophiles (hydrogen bonds). The ribosome peptidyl transfer center is a predominantly entropic catalyst that raise the free energy of the pre- and post-states through geometric constraint of bound reactants (Sievers et al 2004). The thermodynamic distinction is the temperature-dependence of the reaction as say Arrhenius plot of log k versus 1/T (ref).

By positioning the electrophile and nucleophile, and lubricating exit of the nascent polypeptide, the ribosome peptidyl transfer center and its exit tunnel accelerate peptide bond formation about 10^5-10^7 fold (Sievers et al 2004). The transfer center comprises the P- H74 H89 and A-sites H90 H93 of the large subunit rRNA. The central loop of the peptidyl transfer center is closed by helix H73, while the region between H73 and H74 forms the entrance of the nascent polypeptide exit tunnel. The folded

peptidyl transfer center has an inner core of some 120 nucleotides, and an outer shell of 60 nucleotides that includes the P- and A-loops.

To achieve favorable geometry of substrates, allowing their great variety of amino acid sidechains and tRNA acceptor stems, the ribosome specifically interacts with the unpaired universal nucleotides C74 C75 A76 forming the 3' end of the tRNA acceptor arm (Samaha et al 1995; Kim & Green 1999; Nissan et al 2000), and the nascent polypeptide backbone (Syroegin et al 2023). Thus, the nucleophile α-amine nitrogen is positioned by basepairing of aminoacyl-tRNA tC75 with rG2553 of the A-loop atop H92, as well as A-minor interaction of tA76 with rU2506-G2583. Similarly, the electrophile carbonyl carbon is positioned by basepairing of peptidyl-tRNA tC74 C75 with rG2251 G2252 of the P-loop atop H80, and A-minor interaction of tA76 with rA2450-C2501.

The main pushing, or propulsive force on nascent polymers during elongation results from the transfer and translocation reactions that replace the P-site nucleotidyl-dRNA or peptidyl-tRNA by equivalent ones, now lengthened by two nucleotides, or one amino acid, respectively. Tugging on the nascent polypeptide, a network of conserved hydrogen bonds between the backbone of the three proximal residues and rRNA nucleotides G2061 A2062 U2506 ensures that it maintains an extended β-strand conformation needed for peptide bond formation and tunnel passage (Syroegin et al 2023). Additional pulling, or tensile forces result from the configurational freedom of segments emerging from the cramped tunnel, the free energy of protein folding, and co-translational secretion (Leininger et al 2019). Whether pushing or pulling, these forces are transmitted along the polypeptide backbone with little attenuation by the tunnel wall.

Desolvation, or excluding bulk water from the active site, averts side reactions that compete with polymer elongation. In polynucleotide ligation (section 12). In nucleotidyl transfer, one activated water is needed for hydrolysis, and in the absence of condensation, results in polynucleotide termination and release. In polynucleotide termination by amino acid tagging, ….(section 13). In peptide bond formation, aminolysis has replaces hydrolysis, but one activated water is still needed for polypeptide termination and release (section 14).

Entropic trapping is the predominant, but not sole contribution of the ribosome to peptide bond formation. High resolution structures and kinetic studies , rRNA mutations, and substrate mimics in search for enthalpic interactions lower the free energy of the tetrahedral transition state by activation of the nucleophile or stabilization of the leaving group. [zwitterion, anion or neutral] ribosome and **P-site tRNA** together activate water molecule that deprotonate the nucleophile amine

[proton relay MOVE DOWN section 14 peptide bond formation]

is the original proton wire paper a stack or a queue

this back and forth relay of protons through hydrogen bonds has been called a proton shuttle or proton wire but its forth and back movements are most clearly described as the push and pop operations of a stack of hydrogen bonds. The first operation pushes a proton from the nucleophile amine onto the [2'-OH A76 | 2' OH A2451 | W1] to form the stack [proton |]. This is proposed to be the rate limiting event in forming the tetrahedral transition state. The second operation pops this proton onto the leaving group 3' OH A76 and restores the original stack.

push then pop a proton onto a stack of hydrogen bonds

nucleophile amine *pushes* proton onto 2' OH A76 onto 2' OH A2451 onto W1

leaving group 3' OH A76 *pops* proton from 2' OH A76 from A2451 from W1

A minimum of three proton abstractions and additions occur: (1) deprotonation of α-NH_3^+ to prepare the nucleophile, (2) deprotonation of this α-NH_2 nucleophile during or after attack, and (3) protonation of the leaving group 3' OH of tA76 in the P-site. The rate-limiting step is a concerted deprotonation of the α-amine as it attacks the carbonyl carbon (Polikanov et al 2014). Several proposed networks of hydrogen bonds called proton shuttles or wires couple the abstraction and addition of single protons into longer sequences. The proposed sequences range from …. to … include tRNA, rRNA and structured water molecules.

[proton wire] rate-limiting step of transpeptidation is

concerted attack of alpha-amine of aa-tRNA and deprotonation of nucleophile

proton wire : α-amine → 2' OH P-tRNA A76 → 2' OH A2451 → W1 H2O

W1 H2O → {N6 A2602, O of P between A76/C75 A-tRNA, a-amine L27 Ala2}

2’ OH A76 in P-site needed for peptide bond formation and release (Zaher et al 2011)

deprotonation of the charged primary amine …

network of hydrogen bonds allowing concerted proton transfer in formation of the rate-limiting tetrahedral intermediate … shuttle or wire involving … ? general acid/base … ?metals … one or more structured water …. and substrates themselves …. (Polikanov et al 2014)

ASIDE 2’ OH of A76 in proton shuttle/wire but why A?

The plasticity of entropy trapping … using basepairing between substrates and the ribosome transfer center … the most likely substrates were RNAs themselves undergoing nucleotidyl transfer …. at their 3’ or 5’ ends ….

Electrostatic shield of developing charges …

binding peptidyl-tRNA to P-site A2451 U2506 U2585 along with G2252

accommodation gate C2573 with U2492 C2556 facilitates placement of aa-tRNA at the A-site

least mutationallly flexible

base triple C2501 (A2450.C2063)

stacks with triple G2447-A2451-G2061

P-site from A2058 to C2501 of central loop

H74 H75 H80 H89

A-site from G2502 to A2610 of central loop

H90 H91 H92 H93

H90 C2507 2’ OH hydrogen bonds with 2’ OH and O2 of C75 in aa-tRNA

both A-region and P-region of PTC have stem-elbow-stem motif (also found in tRNAs and ribozymes ref 13)

The ribosome transfer center and its tRNAs are tuned to form amide bonds between any two proteinogenic α-L-amino acids. The natural promiscuity of the ribosome peptidyl transfer center … formation of peptide bonds between any of some 400 pairs of proteinogenic L-α-amino acids at comparable rates. [slower reaction of peptide bonds when one or both residues are the imido acid α-L-proline][EF-P]… proof-reading or discrimination Green

Doerfel LK et al (2013). EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science 339, 85-88.

Early evidence established the independence of peptide bond formation from decoding tout court … … nonsense suppressor tRNAs that change anticodon and decoding, not tRNA charging or peptidyl transfer, misacylation (von Ehrenstein experiment), misacylation (chemical oxidation of phe-tRNA to alpha hydroxy acid) … rate of LSU alone

The evolutionary plasticity of the ribosomal peptidyl transfer center has been demonstrated by design and selection of rRNA variants that catalyze …. (Dedkova & Hecht 2019)

- IDEA accommodates a large set of tRNA adaptors as well as diverse amino acids on the aminoacyl-tRNA in the A-site …

- α-D-amino acids are excluded from proteins by aaRSs, proof-reading deacylases, EF-Tu, and reduced reactivity in peptidyl transfer … as well as ribosome arrest of the incorporated peptide

peptide bond formation with α-D-aminoacyl-tRNA in the A-site is about 10^3 slower than with the natural α-D-aminoacyl-tRNA (Englander et al 2015). The reason is that the sidechain must be accommodated in the A-site cleft between C2452 and U2506. … which rotates the H not the amine N of Calpha for inline attack ….. (Melnikov et al 1019).

the amine is too far from the 2’ OH tA76 in P site for hydrogen bonding and nucleophile activation

three factors (1) unfavorable geometry for inline attack (2) failure of hydrogen bond to 2’ OH tA76 in P-site that activates nucleophile (3) failure of protonate the LG 3’ OH of tA76 in P-site

α-hydroxy acids … polyesters

cyclic analogs
N-alkylated amino acids
oligomeric
foldamers
The backbone extended by lengthening the distance between nucleophile and acyl-ester to tA76 …. backbone-extended β-, γ- and δ-amino acids … longer additions by repositioning the tRNA-rRNA interactions.
and dipeptides (Maini et al 2015)

IDEA While substrate positioning ….. kinetics, the energetics of polymer formation … compare the intrinsic stability of the polymer bond to the monomer bond to its RNA carrier … varies isoergonic polyesters … exergonic polypeptides … isoergonic polynucleotides ….
linkage itself … peptide/amide, thioamide, acyl-ester, phosphoester

The idea that polymer transfer center …. entropic control, solvent exclusion, proton shuttles or wires …. Molecular modeling of dRNAs in the modern ribosome transfer center may clarify … … Because 3’ extensions of tRNAs run anti-parallel to their 5’ ends, the aminoacyl-tRNA and peptidyl-tRNA inherit from the proper relative orientation for inline attack from the duplicon-dRNA and polynucleotidyl-dRNA.
[IDEA] The speed of entropic catalysis by transfer center is more than adequate because the rate limiting steps are faithful decoding and precise translocation ….

[MOVE DOWN section 14] Alongside the polymer transfer center, the nascent polymer exit tunnel was co-opted from polynucleotides to polypeptides, changing the mechanical forces of elongation. The exit tunnel at the breakout of polypeptide translation was likely shorter and wider than the tunnel of modern ribosomes, and had to accommodate polypeptides of simple sequences and compositions (Fritch et al 2018; Dao Duc et al 2019).

80-100 A in length, average 15 A in diameter; somewhat wider folding vestibule last 30 A … [90 A length by 6 A width (Dao Duc et al 2019)] co-translational folding within the vestibule and just outside the ribosome.

NB PTC remains rigid during transpeptidation

[MOVE up section 9/10] processive RNA folding … One important continuity in RNA copying from duplisome to RdRP and DdRP enzymes is that RNAs are made from 5' to 3', albeit by different mechanism and step sizes. This means that vectorial folding of the nascent polyribonucleotide procedes in the same direction and likely similar sequence of folding intermediates … this would not be true of an RNA copying scheme that read from 5' to 3' but made a reverse product (*pace* Campbell 1991; Weiss & Cherry 1993; Gordon 1995). [idea] elongation was so much slower!

[MOVE ?] For purposes of probing the substrate promiscuity of the ribosome, as well as applications to expand the genetic code, have been probed by experimental misacylation of tRNAs, either *in vitro* (cell-free) chemical or ribozyme (flexizyme) catalyzed charging, or *in vivo* (cell-based) enzyme catalyzed charging using an orthogonal tRNA/aaRS pair.

Maini R, Dedkova LM, Paul R, Madathil MM, Chowdhury SR, Chen S & Hecht SM (2015). Ribosome-mediated incorporation of dipeptides and dipeptide analogues into proteins in vitro. J Am Chem Soc 137, 11206-11209.

Dedkova LM & Hecht SM (2019). Expanding the scope of protein synthesis using modified ribosomes. *J Am Chem Soc* 141, 6430-6447.

Kofman C, Lee J & Jewett MC (2021). Engineering molecular translation systems. Cell Systems 12, 593-607.

10. What drove RNA elongation?

For the dawn of RNA life, we stipulated a general feedstock of random oligonucleotides and compartmented processes of spontaneous copying, but were otherwise agnostic about any abiotic materials, energy sources, or reaction mechanisms (sections 1-3). In sections 8 and 9, we proposed that spontaneous copying was supplanted by RNA duplication catalyzed by the duplisome and its dRNAs, which tapped the existing feedstock of random oligomers. The covalent reactions of this elongation cycle, polynucleotidyl transfer and dRNA loading, were each analyzed at two steps separated by ribozyme-bound intermediates, *viz*. an uphill condensation coupled to a downhill hydrolysis occurring immediately before or immediately after, respectively. With no net consumption or production of high-energy bonds, this polymer elongation cycle driven by oligomer shortening was thermodynamically isoergonic as well as kinetically reversible (Ross & Deamer 2016).

As dRNA loading and polynucleotide transfer *tout court* cannot explain the processivity of RNA duplication, we must look elsewhere for a source of free energy to drive the polymer elongation cycle forward. First, we consider (and rule out) the possibility that duplication was pulled forward by favorable free energy of product folding. Next we consider how mass action of the loaded / free dRNA ratios might drive elongation forward. Although the oligonucleotide feedstock could likely shift these ratios to favor elongation over de-elongation, this mechanism is vulnerable to pausing or even reversing with the vagaries of oligonucleotide supplies. We consider (and rule out) other ways to concentrate/activate these substrates, or remove/inactivate their by-products, beyond the proposed reaction of dRNA loading *tout court* from section 9. Finally, having said all we can about the initial inputs and final outputs of RNA duplication, we ask whether there was some proximate source of free energy, comparable to ribosome translation factors EF-Tu and EF-G, that might operate within the RNA elongation cycle itself?

Eschewing additional chemical intermediates, here we suggest the duplisome and its dRNAs formed a simple heat engine, harnessing the diurnal warming and cooling of the Earth's surface to drive RNA elongation. Our key idea is that useful work was

extracted from a thermal cycle of opening dRNA hairpins during the day and closing them during the night. Remarkably, the slow, but steady addition of one duplicon per day afforded sufficient generations for natural selection of RNA life. With duplication times of 1 to 10 years, the covalent stability of RNAs placed strong constraints on habitable environments, and provided strong selection for mechanisms of damage prevention, recombination and repair (section 12).

In Figure 10-1 we show the reactants for RNA duplication from a pool of loaded dRNAs. The duplisome is a common catalyst restored to its initial state at the termination of copying, having cycled through various intermediate states each elongation step. Similarly, the template is unfolded for copying, but ultimately refolded to its original structure. Formally, it is an autocatalyst in that the major product, or duplicate is identical to the template, excepting any errors in copying, and possibly the ends.[10] Loaded and free dRNAs are stoichiometric reactants, *viz*. substrates and by-products, respectively. The net reaction does not change the total number of RNA molecules (or equivalently, the number of phosphoester bonds) but redistributes their lengths and sequence. Hence, the chemical free energy available to drive duplication comes from (1) the reactions that reload dRNAs, and (2) differences of secondary or tertiary structure between these substrates (duplicon-dRNAs), their by-products (freed dRNAs), and the major product (duplicate RNA). Gains of base pairing and stacking in the folded duplicate are countered by losses of pairing and stacking between duplicon and anticodon in loaded dRNAs. Notwithstanding extremely stable RNAs with favorable stacking, base triples, ion binding etc., it seems unlikely that vagaries of product folding could drive a robust general mechanism of RNA copying.

[10] In section 12 we consider problems of end replication and repair, including how the nucleotidyl transfer center, or ribozyme P might release the final dRNA from the 3' end of the duplicate.

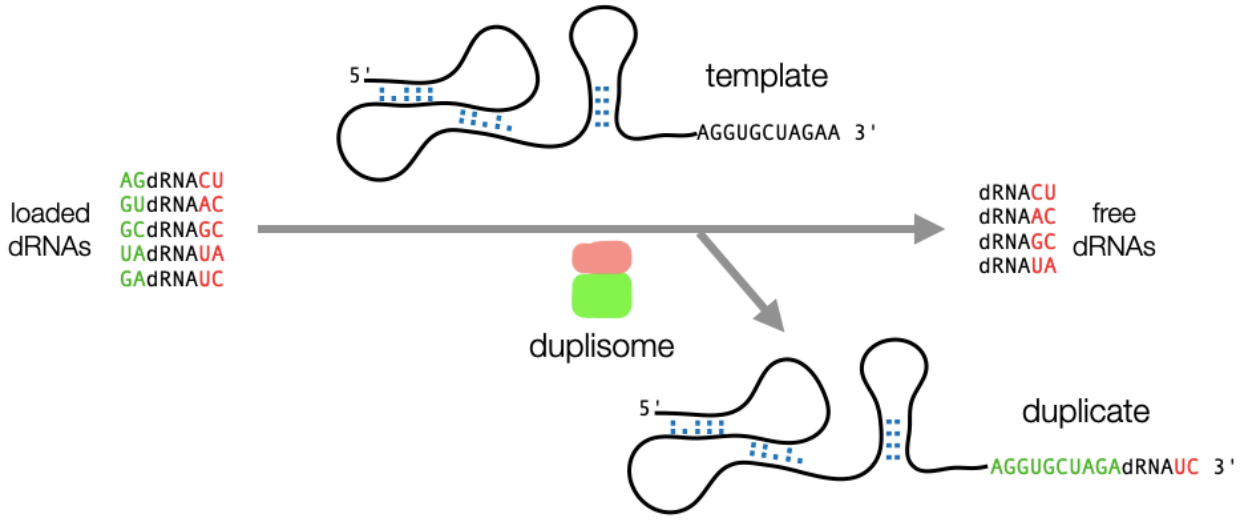


FIGURE 10-1. MASS ACTION OF LOADED AND FREE dRNAs

Could mass action of loaded dRNAs drive elongation? The reaction in Figure 10-1 is thermodynamically isoergonic, and perhaps kinetically reversible, with an equilibrium constant shown in Equation 10-1. The indices σ_i counts how many times each of the 16 different duplicons occurs in the copy. The reaction is driven toward duplicon addition when loaded dRNAs are in excess, and toward subtraction when free dRNAs are in excess. Taxing our intuitions, we imagine codon-directed subtraction from senescent polynucleotides as the reverse of codon-directed addition to nascent polynucleotides. In this spooky exercise of microscopic reversibility, we picture de-exit of closed dRNA into the E-site, opening this free dRNA for reverse translocation along with the polynucleotidyl-dRNA, reverse nucleotidyl transfer, and finally, closing the duplicon-dRNA with duplicon displacement of the codon from its anticodon, followed by de-entry from the A-site.

$$K_{eq} = \frac{k_+}{k_-} = \frac{[template + duplicate]}{[template]} \cdot \prod_{i=1}^{16} \left(\frac{[free]}{[loaded]} \right)^{\sigma_i}$$

EQUATION 10-1. MASS ACTION OF LOADED / FREE dRNAs

In Figure 10-2 we show the reactants in RNA duplication from a pool of random oligonucleotides. Once again, the template is an autocatalyst, while the duplisome, ribozyme P and free dRNAs are common catalysts. Of the stoichiometric reactants, oligomers consumed in dRNA loading are returned, shortened by two nucleotides at their 3' end, as by-products. Despite the rearrangements of primary structure, there is no net change in the number of RNA molecules or phosphoester bonds. Unspecified reactions that produce longer oligonucleotides and consume shorter ones complete the elongation cycle. [??????] One simple way such feedstock reactions might elevate the loaded/free dRNA ratios to drive elongation is to produce longer oligomers (say 6 nucleotides or more) that are preferred substrates for dRNA loading, and consume shorter oligomers that cause product inhibition in dRNA loading.

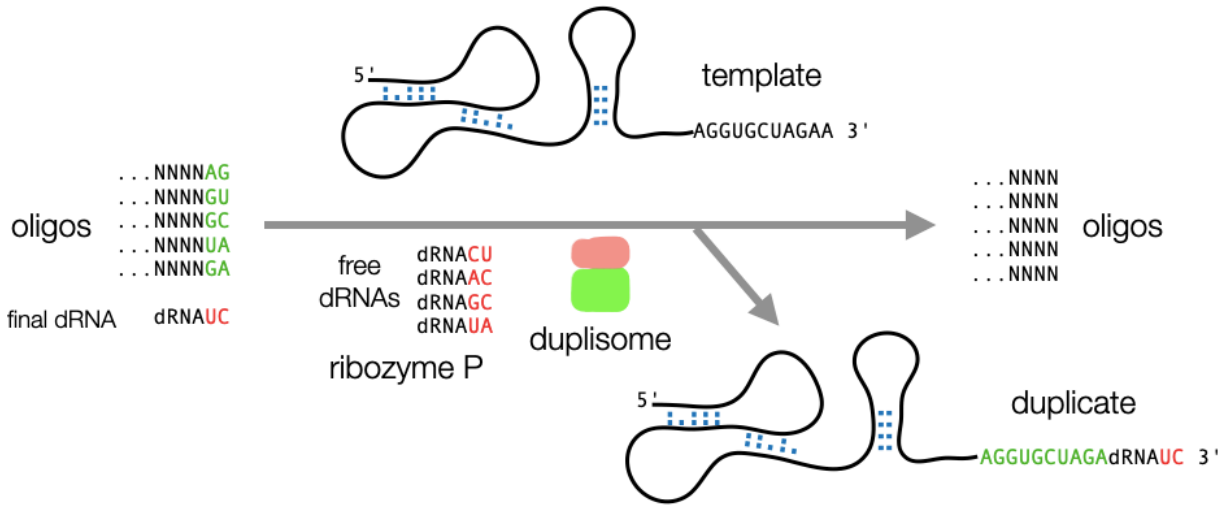


FIGURE 10-2. INPUT AND OUTPUT OLIGOMERS

Tracing back to LUCA, isoergonic and endergonic reactions are metabolically coupled, more or less directly, to the hydrolysis of ATP or GTP. Beyond the favorable free energy of the peptidyl transfer reaction (Table 9-1), driven indirectly by the tRNA charging reactions catalyzed by ATPase enzymes known as aminoacyl-tRNA synthetases, polypeptide elongation is coupled directly to translation factors (GTPases). Was there any comparable source of chemical free energy for RNA elongation, and a means of tapping it? In section 9, we noticed that the nucleotidyl transfer center might use the 5' OH in the A-site, removing the 5' phosphate of loaded

dRNAs before duplisome entry, and restoring this phosphate to free dRNAs after duplisome exit. If so, a pair of ribozymes, *viz*. duplicon-dRNA 5' phosphatase and dRNA 5' kinase, might tap a high-energy phosphate donor to effectively shift the substrate / by-product ratios to drive elongation. There is great interest and uncertainty about prebiotic chemical activation and the first ribozymes to tap these high-energy donors. It is unclear which, if any components of the modern NTP currency preceded, rather than followed, the evolution of ribozymatic RNA copying, or indeed protein translation itself (section 14). Beyond the duplicon itself, we pursue additional reaction intermediates no further here, preferring our parsimonious proposal that the polynucleotide 3' OH made by hydrolysis in the P-site condensed immediately with the 5' phosphorus of the duplicon-dRNA in the A-site (section 9).

Beyond fortuitous product folding, or mass action of loaded / free dRNA ratios, was there any proximate means to drive RNA elongation without invoking additional reaction intermediates? To this end, we briefly survey likely sources of free energy on the Hadean Earth (Deamer & Weber 2010). There are three broad classifications of energy sources based on their distal origin, geological distribution, and temporal patterning, respectively. First, the ultimate origin of this energy may be solar, terrestrial, or extraterrestrial. Thus, solar radiation drove chemical reactions directly, surface warming less directly, and atmospheric (wind, rain, lightning) or oceanic (currents) phenomena least directly. Meanwhile, geological and geochemical processes, as well as extraterrestrial impactors, provided fluxes of both energy and materials.

Second, few if any energy sources were everywhere available: Some only operated at defined planetary latitudes, atmospheric heights, or oceanic depths. Some only operated at definite interfaces of air and water, water and land, or land and air. Some only operated in isolated, if not frankly exotic locations *viz*. volcanos, hydrothermal vents, salt flats. Of these interfaces and locals, some where determined by extrinsic discontinuities, while others arose by disproportionation of initially homogeneous environments (nuclear geysers).

Third, some energy sources at the Earth's surface, including geochemical reactions and radioactive decays, were more or less steady, others were entrained and driven by the cycles of planetary motion, and some fluctuated erratically (Gordon & Mikailowsky

2021). The most regular cycles included once-a-day warming and twice-a-day tides, as well as seasonal atmospheric, terrestrial and oceanic phenomena. Electrical storms, and perhaps volcanic eruptions, occurred more erratically, but on comparable time scales of days or years. Finally, some events occurred quite sporadically, even just once (Moon formation), in the planets history.

The exotic energy sources of the Hadean Earth, both steady and episodic, are challenging to comprehend. One exception is the sunlight so familiar to us today that provides a large flux of energy to the Earth's surface in daily and yearly cycles. Granting this streetlight effect, here we consider whether radiant energy from the Hadean Sun drove RNA duplication. Rather than coupling the elongation cycle to sunlight directly, we suggest that the opening and closing of dRNAs within the duplisome was entrained to the daily heating and cooling of the Earth's surface, driving regular addition of one duplicon per day.

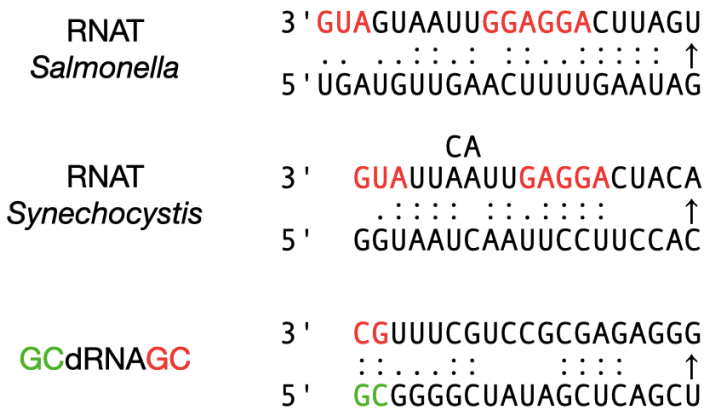


FIGURE 10-3. RNA THERMOMETERS & LOADED dRNA

RNA hairpins are used in bacteria as thermal riboswitches to control mRNA translation in response to temperature (Kortmann & Narberhaus 2012). In a change of perspective, here we regard RNA hairpins as thermal ribomotors that perform mechanochemical work. In typical RNA thermometers found in the 5' UTR of an mRNA, a critical hairpin masks the Shine-Dalgarno (SD) box, preventing protein translation in the cold. Upon warming, melting of this hairpin unmasks the SD box and initial AUG for translation. In Figure 10-3 we compare the critical hairpins of two RNA thermometers

to GCdRNAGC modeled on bacterial $tRNA^{Ala}$. The SD box and AUG start codon in these bacterial RNA thermometers are shown in red.

$$W = \Delta T / T_m \cdot \Delta H_{open}$$

EQUATION 10-2. WORK AVAILABLE FROM HOT OPENING AND COLD CLOSING OF HAIRPIN RNA

The principle of our heat engine is that RNA hairpins are harder to pry open when cold, than when warm. The free energy available from this cycle can be calculated from the melting temperature (T_m) and opening enthalpy (ΔH_{open}) of the hairpin, along with the temperature difference (ΔT) between the hot and cold baths (Privalov 2012). If the melting temperature is say 350 K, and these baths are 35 K apart, up to 10% of the enthalpy of melting can be extracted as useful work, assuming the enthalpy and entropy of opening are insensitive to temperature (Equation 10-2). By opening the duplicon-dRNA in the A-site during the day and then closing the freed dRNA in the P-site during the night, the duplisome extracts as much as 10 kcal / mol useful work, comparable to hydrolysis of 1 GTP to GDP in a modern cell. In Figure 10-4 we depict a cycle of opening GCdRNAGC in the warm bath and then closing dRNAGC in the cold bath. Using the RNAfold Vienna program to estimate the free energies of opening and closing these hairpins at 310 K and 273 K, respectively, we obtained a net free energy of nearly -9 kcal / mol / cycle (Gruber et al 2008; http://rna.tbi.univie.ac.at/).

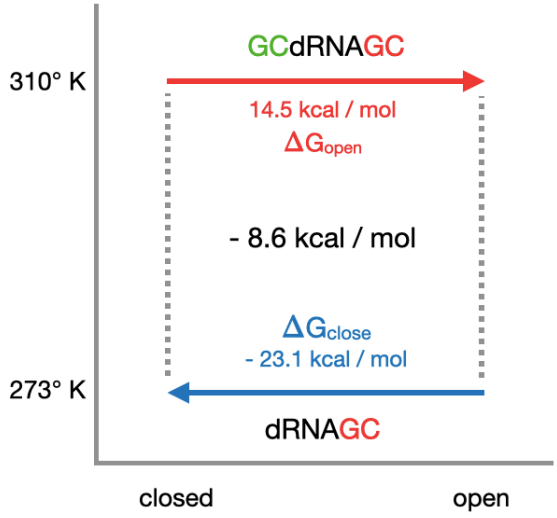


FIGURE 10-4. HEAT CYCLE OF dRNA OPENING AND CLOSING

In section 9 we proposed that dRNAs close for duplicon loading, and open for nucleotidyl transfer, respectively. Here we suggest that the loaded dRNA opens, and the freed dRNA closes *within the duplisome itself*, to drive decoding and translocation, respectively. Before discussing this elongation cycle, we remark on the secondary and tertiary structure of dRNAs. If closed dRNAs were more-or-less regular hairpin extensions of the DSL of modern tRNAs (*cf*. Figure 9-3), what might open dRNAs look like? In Figure 10-5 we model an open dRNA on the tertiary structure of yeast $tRNA^{Phe}$ (Kim et al 1974; Robertus et al 1974). None of the interactions with the 3' half (37-76) would have been present, but any hydrogen bonding and nucleobase stacking contained strictly within the 5' half of the modern tRNA (1-36) may have been present in the dRNA ancestor. These include conserved base triples U8-A14-A21 and A9-U12-A23 that pin the **duplicon-leader** back against the **D arm**. Conceivably, this augmented D-helix held the leader, **hinged** back between A9 and G10, open to accommodate the nucleotidyl transfer center. Conversely, the **trailer** from G26 to the anticodon A35 A36, has no interactions with the D-arm, suggesting it swung freely to pair with the template in the decoding center.

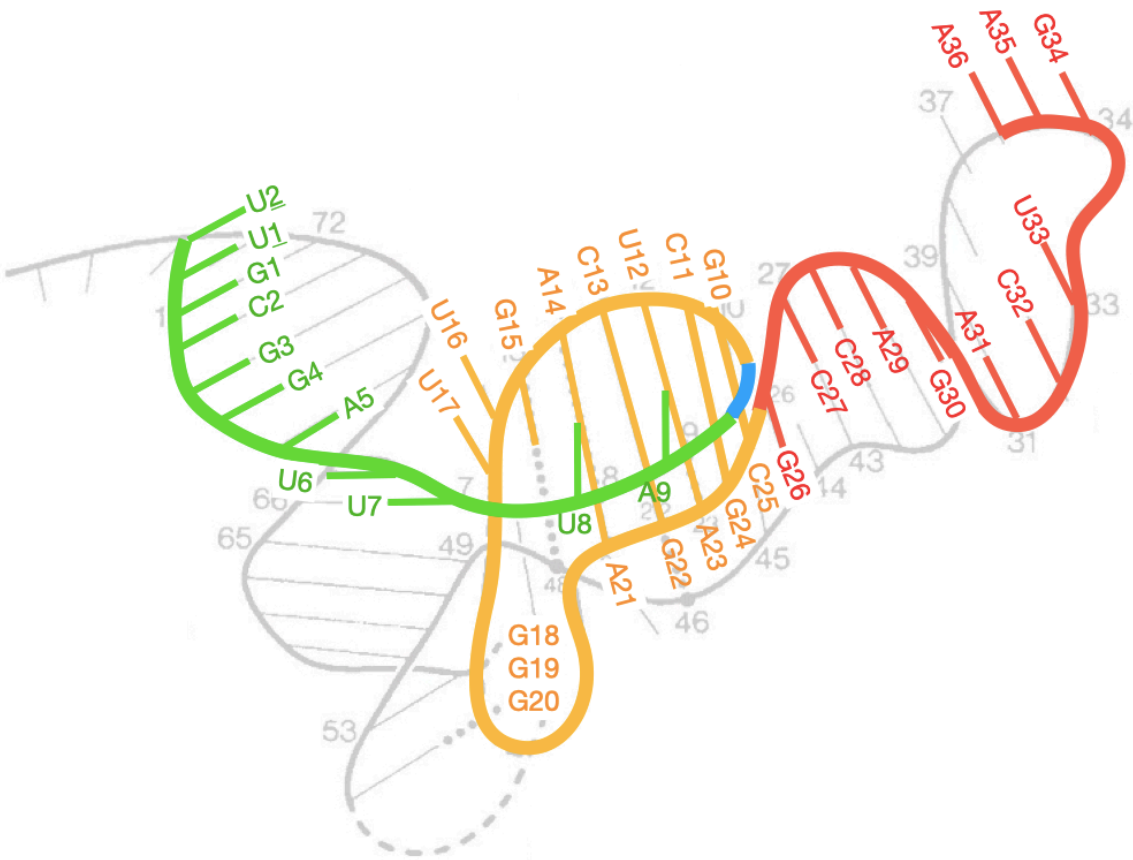


FIGURE 10-5. OPEN dRNA MODELED ON tRNA (AFTER ROBERTUS ET AL 1974)

In section 11 we examine how the duplicon-dRNA is pried open by codon-anticodon pairing to displace the duplicon, freeing the entire leader to fold back onto the D arm. Large deformations, as well as rigid rotations, of aminoacyl-tRNA during accommodation may be vestiges of this dRNA opening (ref). Indeed some releasing factors, perhaps protein mimics of “lost” nonsense tRNAs, show analogous movements of accommodation (ref). After nucleotidyl transfer, the freed dRNA closes to exit from the duplisome. We suggest this duplisome-bound opening of the duplicon-dRNA during the hot day, and closing of the freed dRNA during the cold night may

have provided the free energy to drive an otherwise, nearly isoergonic elongation cycle for RNA duplication.[11]

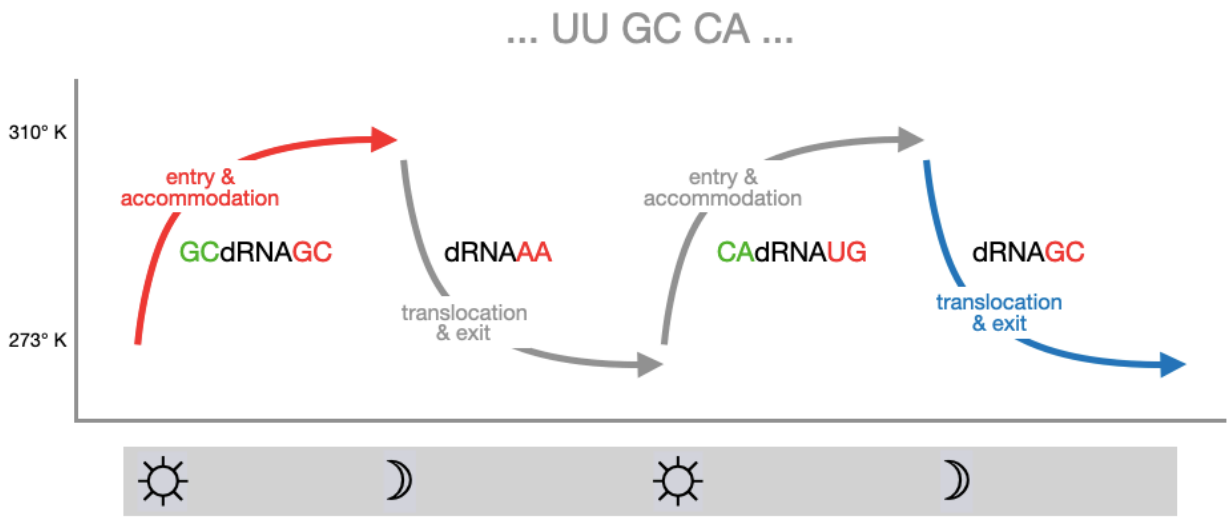


FIGURE 10-6. THERMAL ELONGATION CYCLE

In Figure 10-6 above we illustrate the thermal elongation cycle where heating drives decoding and cooling drives translocation with dRNAs acting as thermal ribomotors.[12] On the first morning, closed GCdRNAGC enters the A-site where its anticodon gains a toehold with the cognate codon GC, prying the dRNA open to accommodate. Near midday the nascent duplicate-dRNA and duplicon-dRNA under isoergonic exchange at the nucleotidyl transfer center. On the first evening the free dRNAAA closes, driving translocation and exit. On the second morning the closed CAdRNAUG enters the A-site for another decoding. Thus, the original GCdRNAGC exits as a freed dRNAGC on the second evening. In this way, one new duplicon is added each day, yet any

[11] Although dRNAs combine their motoring with loading and transfer functions in a uniquely elegant way, thermal drive is not limited to hairpin structures. Thus, regions of the rRNAs themselves might also undergo reversible changes in secondary or tertiary structure to drive elongation. Indeed it has been conjectured that the SRP RNA Alu domain, which occupies the factor binding site at the ribosome subunit interface to slow or arrest elongation, was an elongation factor before translational GTPases (Ahl et al 2015).

[12] The sequential reactions of dRNA loading may also have been entrained to the diurnal temperature cycle, charging at night when the entropic cost of substrate docking was low, and trimming in day when the entropic benefit of product release was high.

particular dRNA takes two full days from entry to exit. A full reaction cycle of two successive days is shown in Figure 9-7 below.

...UUdRNAAA + GCdRNAGC ⇔ dRNAAA + …UUGCdRNAGC

…UUGCdRNAGC + CAdRNAUG ⇔ dRNAGC + …UUGCCAdRNAUG

FIGURE 10-7. COVALENT REACTIONS OF ELONGATION CYCLE

Could duplisome life have evolved adding (at most) one duplicon per template per day? There are two concerns: Could this rate of synthesis have outpaced decomposition? And would there be enough rounds of Darwinian selection for robust evolution? Stipulating the first answer is yes, the second is clearly yes as well. If we assume one duplicon was added per day over a 6 month growing season, we could add some 360 nucleotides per year. Under these conditions 3600 nucleotides would be duplicated in 10 years, sufficient for modern rRNAs. The generation time could, of course, be shorter if ancestral rRNAs were duplicated in smaller fragments.[13] The population-doubling time could be shorter if polyduplisomes shared one template just as polyribosomes share one mRNA. Finally, the Earth's rotation has slowed considerably since the Hadean when there were likely twice as many days per year (Gordon & Mikhailovsky 2021).

A doubling time, or even a generation time, of just 1 year is conceivable for duplisome life. But even if a generation required 10 years, and duplisome life lasted only 200 million years, there would have been 20 million generations, a figure comparable say to the mammalian radiation of recent evolution. Major events in this era, from amino acid tagging and the breakout of polypeptide translation to the retirement of RNA duplication, are discussed in sections 13 and 14.

But what of the chemical stability of the RNA backbone and the nucleobases? Protected from environmental insults, *viz*. UV irradiation, exogenous nucleophiles,

[13] Besides the universal partition of SSU and LSU rRNAs, there are three precedents for further partition: In eukaryotes, cytoplasmic LSU rRNA is fragmented into 5.8S and 28S rRNAs. In protosomes, 28S rRNA is fragmented into nearly equal halves called 28Sα and 28Sβ (Natsidis et al 2019). Finally, in *Euglena gracillis*, 28S rRNA is fragmented into multiple pieces (Matzov et al 2020). ?insect 5.8S cleavage?

biotic nucleases, RNA in neutral solution is still prone to spontaneous scission of the backbone, as well as damage or loss of nucleobases. At 20-25 °C, the half-life of one phosphodiester linkage in water at pH 6-7 is estimated between 5-1000 years (Eigner et al 1961; Li & Breaker 1999; Wolfenden 2011; Ross & Deamer 2016; Lonnberg 2023).[14] The stability of any phoshodiester linkage is significantly greater in paired regions. Similarly, the rate of spontaneous deamination of cytosine decreases over 100-fold in paired regions.

It is admittedly unclear whether a copying time of 180 days for an RNA of 360 nucleotides is fast enough to stay apace its half-life from spontaneous decomposition, estimated anywhere from 5-1000 days. Clearly, the common forms of covalent chemical damage placed strong constraints on diurnal elongation, or any comparably slow process of RNA replication: First, habitable environments were constrained by the need to minimize RNA decomposition, favoring conjectures say that ribozymes functioned in dark and icy brines. Second, there was strong selection for shorter, or less damage-prone ribozyme sequences and folds. Finally, there was strong selection for ways of protecting critical sequences, as well as mechanisms of recombination and repair. We return to these problems in section 12.

[14] By comparison, phosphodiester bonds in duplex DNA have an estimated half-life of 31 million years at 25 °C in neutral water (Wolfenden 2011).

11. The RNA code

The universal genetic code maps codons to amino acids, and by iteration, maps mRNAs to polypeptides. This code is the product of two simpler relations involving tRNA adaptors. Complexed with its cognate aminoacyl tRNA synthetase (aaRS), each tRNA is charged at its 3' end with the correct amino acid, and then, complexed with the ribosome decoding center, matched at its anticodon to the mRNA codon. The composite function is many-to-one, *i.e*. *synonymous codons* assign the same amino acid, and only partially defined, *i.e*. *nonsense codons* assign no amino acid. Iterating from one codon to an entire mRNA, deacyl- and peptidyl-tRNAs are shifted to the E- and P-sites, respectively, tugging the next codon into the A-site.

The RNA code maps codons to duplicons, and by iteration, maps templates to duplicates. Like the amino acid code, this older code is the product of two relations involving dRNA adaptors. Complexed with ribozyme P, each dRNA is charged at its 5' end with the correct duplicon, and then, complexed with the duplisome decoding center, matched at its anticodon to the template codon. The composite function is one-to-one, and formally, the identify. Iterating from one codon to an entire template RNA, freed- and polynucleotidyl-dRNAs are shift to the E- and P-sites, respectively, tugging the next codon into the A-site.

The polymer mappings from template RNA to duplicate RNA, or from mRNA to polypeptide, are remarkable but imperfect. As in protein translation, the fidelity of RNA duplication depends on the accuracy of various steps including (1) dRNA loading by ribozyme P, (2) duplicon-dRNA selection by the duplisome decoding center, (3) polynucleotide transfer in the duplisome nucleotidyl transfer center, and (4) translocation of dRNAs and template. Errors in each step produced characteristic defects in the duplicate, i.e., nucleobase substitutions from errors in dRNA loading and decoding, premature polynucleotide release from errors in nucleotidyl transfer, and indels from frameshifts during translocation and decoding. Many mechanisms have evolved to suppress these errors, and improve the fidelity of polymer mapping. We introduce one improvement here … may have been the original mechanism. Some of

these mechanisms arose in RNA duplication and others pertain more specifically to amino acid tagging (section 13) and polypeptide translation (section 14).

One obvious way of improving the accuracy of RNA duplication is to lengthen the codon-anticodon helix from two to three basepairs (Grosjean & Westhof 2016). Naively, this comes at the immense cost of increasing the number of dRNAs from 16 to 64 (*pace* Campbell 1991; ###), or an entirely unacceptable, degeneracy of the copies (*pace* Noller 2012). Here we suggest a simple solution to attain the accuracy of triplet decoding with a set of only 16 dRNA isoacceptors. For purpose of discussion we introduce this mechanism called *triplet decoding with doublet addition* as an improvement on *doublet decoding with doublet addition* outlined in section 8. In fact we are agnostic whether this triplet decoding replaced the older doublet decoding, or was itself the original mechanism of decoding in RNA duplication. Before we refine our concept codon size in RNA duplication, we apprise readers of five intertwined concepts of step size in protein translation (Table 10-1). To wit, reading matches *one codon* with *one anticodon*, transfer adds *one residue* to the nascent polypeptide, translocation shifts tRNAs *one site* in the ribosome, tugging the mRNA *one translocon*.

	match codon & anticodon	add duplicon	shift dRNA	tug translocon
duplication simple doublet	2 nt	2 nt	1 site	2 nt
duplication superwobble	2 ❇ nt	2 nt	1 site	2 nt
translation superwobble	2 ❇ nt	1 aa	1 site	3 nt
translation complete triplet	3 nt	1 aa	1 site	3 nt

TABLE 11-1. FIVE CONCEPTS OF STEP SIZE

The codon and anticodon, *viz*. stretches of mRNA and tRNA matched at the decoding center, define the size of one another, and proved to be triplets (Nirenberg et al 1963; Holley et al 1965; Jones & Nirenberg 1966). However, the simple idea of *complementary triplets* proved inadequate because the rules and underlying mechanism of nucleobase matching vary by nucleotide position. The first two positions are restricted to Watson-Crick basepairs, but the third codon position is more flexible. In his *wobble hypothesis*, Crick noticed that an anticodon U34 could read either A or G in the third codon position, reducing the number of tRNAs needed for the protein code (Crick 1966). In this, and other cases, the map from codon to anticodon is not invertible as the identity of the third codon position is partly lost in decoding. In mitochondria and plastids, as well as reduced bacterial genomes (*Mycoplasma*), a single tRNA with unmodified U34 can read a four-codon box, e.g., $tRNA^{Ala}$ (UGC) reads all four alanine codons GCN (Bonitz et al 1980; Heckman et al 1980; Andachi et al 1989; Rogalski et al 2008; Alkatib et al 2012). Known as *superwobble*, here the adaptor and decoder merely confirm the presence of the third nucleotide, but ignore its identity.

The translocon proved to be a triplet as well, so that the ribosome translates a "comma-free" succession of codons with no gaps or overlaps. In deciphering the protein code, molecular biologists had considered other formal possibilities including a triplet codon with doublet translocon, or a doublet codon with triplet translocon (Crick 1968). The former in effect reads the identity of odd-numbered nucleotides twice, first in the third codon position, and then again in the first codon position, while the latter ignores the identity of every third nucleotide, viz. what is now known as superwobble.

The two formal requirements for faithful RNA duplication are that the duplicon matches the template, and that the translocon is the same length as the duplicon, not one nucleotide more, nor one less. Thus, there is no reason that the decoding center cannot match triplets, so long as the duplicon and translocon are both doublets or both triplets (*pace* Campbell 1991; Noller 2012). Here we modify the scheme of doublet decoding with doublet addition from section 8, to incorporate superwobble, *viz*. monitoring the presence of the third position nucleotide, but ignoring its identity. Our new scheme of triplet decoding with doublet addition has obvious parsimony with decoding in protein translation. There are two further reasons for triplet decoding with

doubt addition that pertain to RNA duplication *per se*. First, as discussed below, the third basepair improves decoding accuracy by increasing the stability of the codon-anticodon helix (Grosjean & Westhof 2016). Second, as discussed in section 12, we suggest that the wobble position monitors template termination.

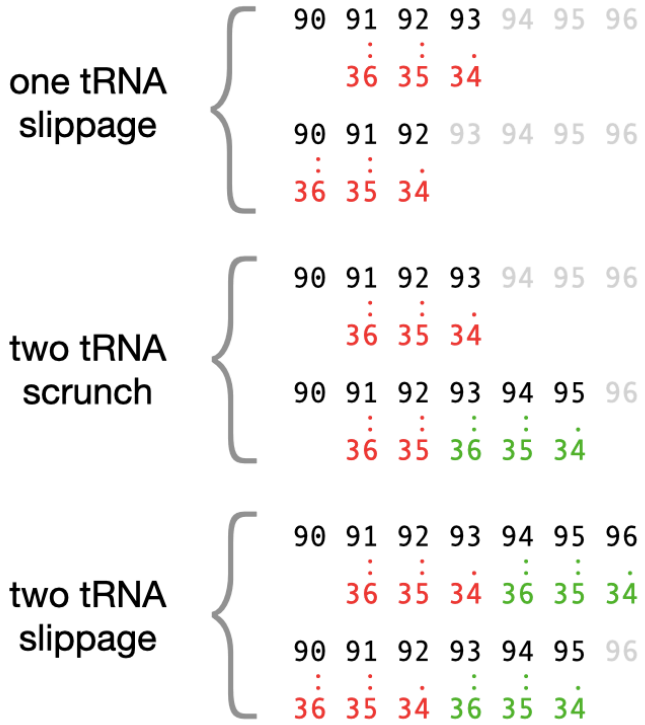


FIGURE 11-1. THREE RIBOSOMAL MECHANISMS OF -1 FRAMESHIFT

Early theories of ribosomal frameshift errors explained reading frame iteration as an active mechanism of translocation, followed by passive entry of aminoacyl-tRNA into an empty A-site. In fact, the division of labor between translocation and tRNA selection is more fluid, so that the incoming tRNA helps to determine, not simply respect, the current frame. The most common errors are shifts of the reading frame upstream on the mRNA by 1 nucleotide called -1 frameshifts (Figure 11-1). Frameshift can occur by *one tRNA slippage* in the POST translocation ribosome with an empty A-site. Here the peptidyl-tRNA slips -1 nucleotide, facilitated by a slippery mRNA sequence X XXY, as well as depletion of the cognate aminoacyl-tRNA. Frameshift can also occur after peptidyl transfer by *two tRNA slippage* in the PRE translocation ribosome. Here the deacyl-tRNA and peptidyl-tRNA together slip by -1 nucleotide, facilitated by slippery

mRNA sequence X XXY YYZ, and programmed by downstream hairpins or pseudoknots. Frameshift can perhaps also occur during decoding itself by what we call *two tRNA scrunch* where nucleobase N34 of the peptidyl-tRNA flips out of the anticodon stack, allowing nucleobase N36 of the incoming aminoacyl-tRNA to read the -1 nucleotide (Licznar et al 2002; Atkins & Bjork 2009).[15] One tRNA slippage, two tRNA scrunch, and two tRNA slippage, may be distinguished by their intermediate states and regulation, but the nascent polypeptide is identical for all three pathways.

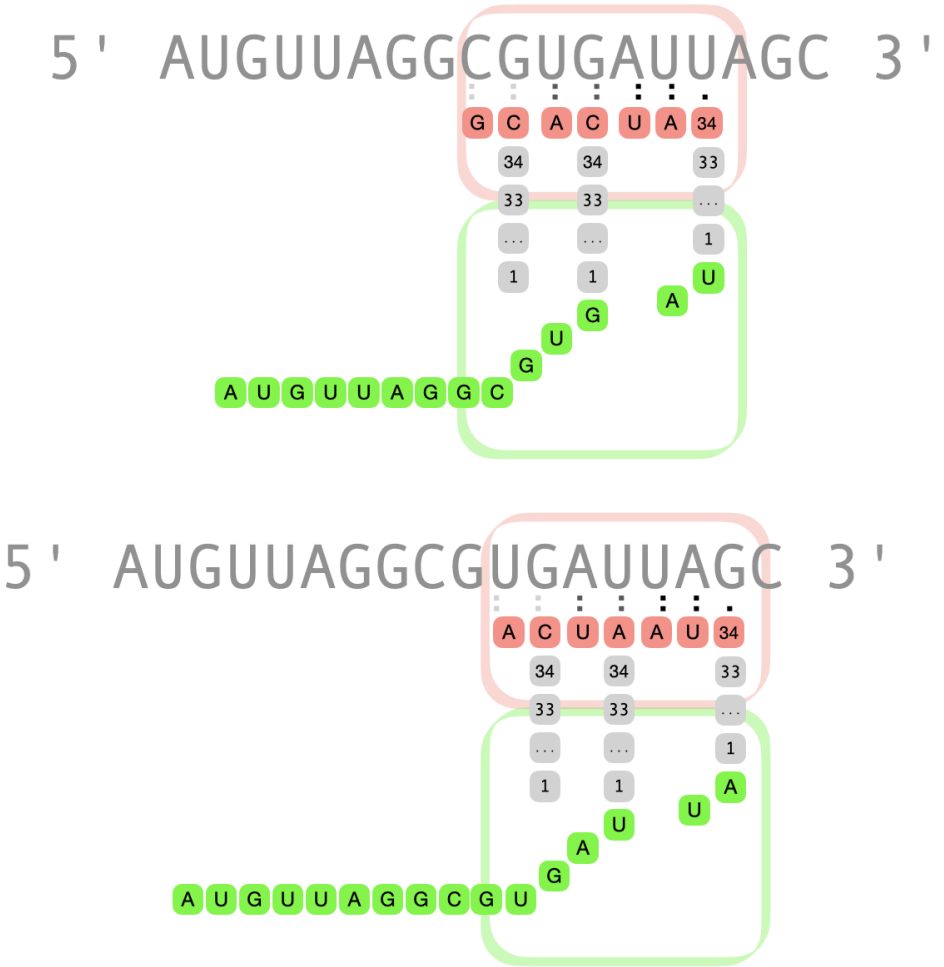


FIGURE 11-2. TRIPLET DECODING WITH DOUBLET ADDITION

[15] Analogous mechanisms of one and two tRNA slippage have been proposed for +1 frameshifts … called spread for +1 frameshifting where the peptidyl-tRNA in the P-site not only defends its wobble position, but *spreads* out to the next nucleotide downstream, forcing the incoming aminoacyl-tRNA to read the +1 codon. May explain how tRNAs with enlarged anticodon loops pair normally in the A-site yet cause +1 shifts in the P-site.

Here we suggest how the duplisome and dRNAs combined triplet decoding with doublet addition in their normal cycle of RNA elongation (Figure 11-2). In this best of both worlds, the triplet codon-anticodon helix provided stability for accurate decoding, while the doublet duplicon (translocon) provided true duplication with just 16 dRNA isoacceptors. To be clear, two dRNA scrunch was the normal mechanism of duplisome decoding whereas two tRNA scrunch, if it occurs at all in the ribosome, is a form of -1 frameshift error. Finally, although we suggest dRNA N34 superwobble was the perfected mechanism of triplet decoding in RNA duplication, analogous cases of tRNA N34 superwobble in protein translation are clearly derived, not primitive characters. Irrespective of the duplisome hypothesis, superwobble at the third codon position has been invented independently a number of times.

Here we suggest that the fidelity of decoding in RNA duplication was determined by direct kinetic competition between codon and duplicon for anticodon affection. To wit, the duplisome sampled prospective duplicon-dRNAs that entered the A-site in their closed conformation. If non-cognate or near cognate, any duplicon-dRNA soon left the A-site, allowing the duplisome to sample another. If it sampled a set of 16 isoacceptors without replacement, it would need 8 1/2 trials *on average* to encounter the cognate duplicon-dRNA. More realistically, if it sampled these isoacceptors with replacement, it would need 16 trials *on average*. Biases in the frequencies of codons and isoacceptors could increase or decrease the mean sample size in decoding. Once the duplisome encounters the cognate duplicon-dRNA at the A-site, the codon can displace the duplicon from the anticodon, retaining the dRNA through this toehold until it fully opens as the day warms and the duplicon arm accommodates the nucleotidyl transfer center. An example of toehold competition and duplicon-dRNA opening is shown in Figure 11-3. In the parlance of polynucleotide strand exchange, the *invader* (template codon $_{1}$GCG$_{3}$) displaces the *incumbant* (dRNA duplicon $_{-2}$GCA$_{1}$) from the *substrate* (dRNA anticodon $_{34}$UGC$_{36}$).

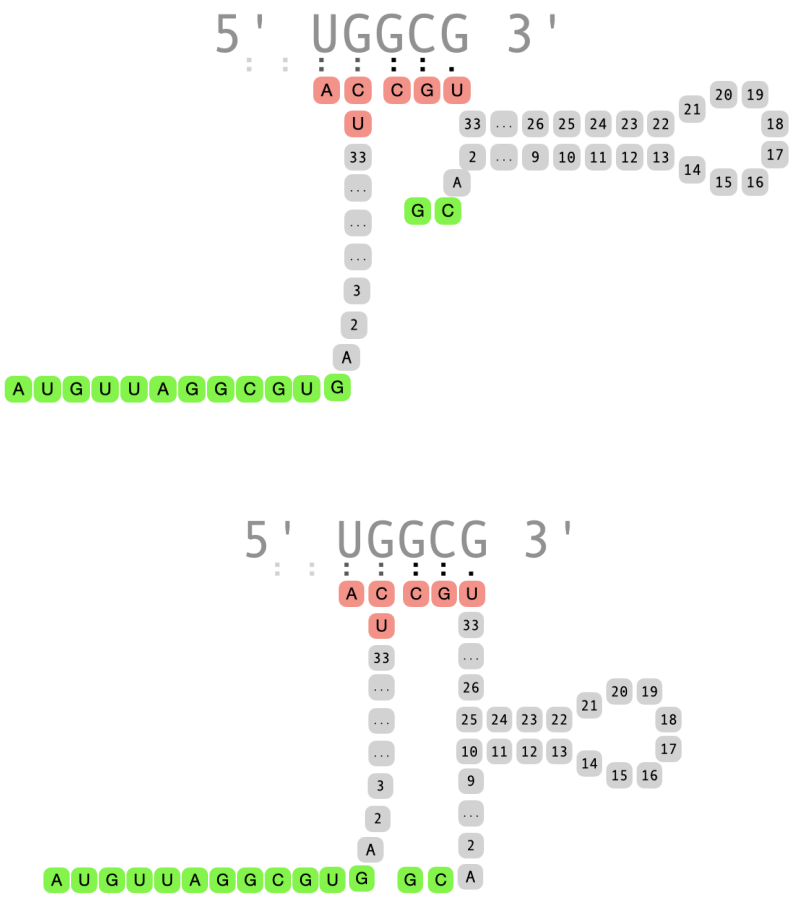


FIGURE 11-3. DECODING BY DUPLICON DISPLACEMENT

1st \ 2nd	U	C	A	G
U	UUAdRNAUAA	UCAdRNAUGA	UAAdRNAUUA	UGAdRNAUCA
C	CUAdRNAUAG	CCAdRNAUGG	CAAdRNAUUG	CGAdRNAUCG
A	AUAdRNAUAU	ACAdRNAUGU	AAAdRNAUUU	AGAdRNAUCU
G	GUAdRNAUAC	GCAdRNAUGC	GAAdRNAUUC	GGAdRNAUCC

TABLE 11-2. 16 dRNAs OF TRIPLET DECODING WITH DOUBLET ADDITION

Table 11-2 shows the 16 dRNAs of triplet decoding with doublet addition based on scrunch (destacking) of polynucleotide-dRNA N34 at P-site, and superwobble of duplicon-dRNA N34 at the A-site. To be definite, during dRNA loading the universal nucleobase U34 pairs with the universal nucleobase A1. During decoding, U34 in the P-site yields to the incoming duplicon-dRNA N36 at the A-site, while U34 in the duplicon-dRNA U34 in the A-site pairs with any nucleobase in the third codon position. In sections 13 and 14 on amino acid tagging and the breakout of polypeptide translation, we discuss the origin of the triplet translocon in polypeptide translation, and the splitting of 4-codon boxes in the amino acid code evolved from the duplisome and its dRNAs. [both during aminoacyl-tRNA selection and translocation]

12. RNA recombination and repair

[Poole & Logan 2005]

RNA viruses TMV (Fraenkel-Conrat 1956; Gierer & Schramm 1956)

plant viroids (Diener 1971; Flores et al 2004)

[arrival versus survival of the fittest]

- attempts to explain evolution by eliminative without creative [Marshall vs global fixed point … Schaupenauer, connectionism, population genetics] hit apparent ceilings , error-catastrophes …. attempts to capture difference by features such as complexity or non-linear also miss the point … nesting of firms within markets within sectors …

[Darwin and Wallace macroevolutionary competition of sibling species, innovation fixed within a species, competition with sister species] The idea of population genetics was that alternative alleles at each locus competed with one another by their contribution to phenotype and fitness of the individual organism. [genome compact Mendel] competition between alternative alleles at any locus, occasional interactions between alleles at two loci, but generally neutral and additive independent, cooperation …. a gene succeeded only if its genome succeeded… mechanisms of sexual reproduction segregation and recombination

on closer inspection many forms of selfish behavior of individual genes or small cliques gaming the rules ….

… interaction of system and surrounding ….

In the argot of biological philosophy for the relationship of system and surround, we say “any embodied and situated process that explores open-ended affordances of its surroundings”.

With an apology to the philosopher C. D Broad, “life was the glory of nature and scandal of natural science” insomuch that biologists, chemists and physicists struggled to agree on its essential phenomena, much less one principle that distinguished living systems from their lifeless surroundings. A bold conjecture about polymer chemistry at the dawn of life on Earth, the RNA world promised to describe

the phenomena of life, and to clarify its principles, straightforwardly realized in polymer kinetics. As in other attempts to model learning and evolution *ab initio*, although the mechanisms were especially simple in RNA life, the fundamental logic of discovery with its evolutionary principle of adaptation became more clear and distinct in later worlds with greater division of functions.[16]

In the parlance of goal-directed search, our general theory of discovery, plans to follow familiar paths are nested within plans to explore unfamiliar ones. Known in search theory as *exploitation* and *exploration*, biologists refer to these nested plans as physiology and evolution, respectively.

The great insight of Charles Darwin and Alfred Russell Wallace was that living species explore immediate paths of their surroundings, and invent mediate ones, for no higher purpose than preserving and propagating these maps. For many, both before and after them, the idea that evolutionary search has no given purpose or *teleos*, but maps new paths to preserve the existing map and plans, the conventional survival and reproduction of Darwinism, seemed paradoxical.[17] Because of the changing surroundings, including the species and other biota, the knowledge acquired is not itself final. life was not a fixed Eigenstate, but a search process with embedded records of its past paths and planned future paths. [memoryless automata]

Today the structure, function, and transmission of the genome and its epigenome provide an example *par excellence* of adaptive systems that map their surroundings to attain, or maintain given goals. But in their theory of evolution by means of natural selection, Darwin and Wallace skirted about the unsolved problems of heredity. Answering these questions has since occupied biologists for over 150 years. First, what maps are inherited from past experiences of the species? Second, how are they

---

[16] replicator dynamics, connectionism or parallel distributed processing

[17] Biological species are the most famous, but remarkably, not the sole example of living systems that invent and test new affordances for no higher purpose than continuity of knowledge and plans. Human social organizations, ranging from churches to nation states, have institutional memories, laws or constitutions, as well as compiled plans of self-preservation, *genomes in all but name*, that exceed the individual powers and memories of human agents contracted to advance the interests of their principals. These organizations explore and exploit their affordances on the timescale of human history not human lifespans. Lower-level organizations, like colleges and business firms, have less autonomy from … and greater … with their human agents. Some cultural institutions, notable the basic sciences are abstracted means …

compiled as goal-directed plans and executed to regulate or mediate the varied processes of cell biology, development and behavior? To routinely exploit the past, organisms must inherit the records of past searches in a useful form, *viz*. as a more or less compiled (aka habitual, or executable) plan of present behavior, or compile or translate them from the language or format of the map to the language or format of the plan. In general, the same map can be used for navigational plans for changing goals at hand. [biologists emphasize this aspect of plans as regulationFinally, how are these maps transmitted from generation to generation? [nesting of continuity within spontaneous or planned variation]

and preserve for the next generation. [plan of present behavior]… transcription of mRNAs and translation of proteins, gene networks of regulation and action … [exploit]

The problem of faithful transmission of maps of past experience and plans of present behavior, preservation and replication of the genome. Easily understood in population genetics … steady-states and purifying selection … and the new molecular genetics … the main challenge to replicate faster than degradation, protect and repair, including error-free repair … running in place … faithful vertical transmission, suppress selfish and useless … recombination in repair

One litmus test for any process of genome replication is that it can reproduce its common catalysts with sufficient speed and fidelity (Eigen 1971). For RNA duplication these catalysts would include the duplisome, ribozyme P, and the set of dRNAs. Speculations about the RNA world have rightly focused on the speed and fidelity of copying larger RNAs, as well as their stability. Like medieval manuscripts before the invention of printing, large RNA molecules were premium goods in RNA life, copied slowly and accurately, and folded with care. There was strong selection to prevent or repair common forms of RNA damage from spontaneous or biotic insults.

Processivity doubtless had priority over fidelity of RNA duplication. If the cognate duplicon-dRNA were depleted, the decoding center likely accommodated a near cognate dRNA as the day warmed, hazarding a nucleobase substitution error to avert an indefinite delay, or premature termination. If the template had damaged or missing nucleobases, the duplisome likely bypassed these stretches to reinitiate at a downstream codon, producing an internal deletion, but no break in the RNA copy.

Finally, as the duplisome unfolded and copied large templates it likely continued past frank breaks or gaps by rethreading of the free 5' end. The copy might suffer a point mutation, but a full-length RNA would be made. Meanwhile, the template could refold after copying, but was not itself covalently repaired by this means.

Because terminal units differ chemically from internal units, synthesizing and maintaining the beginnings and ends of polymers are fraught with problems, and replete with opportunities. Whether a large ribozyme was assembled from several polynucleotide chains, or folded as a single chain, it was important to copy each RNA as close to both ends as possible. There was also strong selection to protect these ends from spontaneous erosion, as well as to restore ends degraded by wear, or shortened in copying.

There were three special problems associated with copying the dRNAs themselves: First, their loading, decoding and transfer reactions required chemical definition of both 5' (duplicon) and 3' (anticodon) ends. Thus, if the 5' phosphate were lost from the free dRNA or duplicon-dRNA, the adaptor would no longer be capable of oligomer charging and nucleotidyl transfer, respectively. Damaged dRNAs might be salvaged as feedstock oligonucleotides, but there would also be strong selection for a repair process. Second, and more novel, the set of dRNAs must be complete and balanced, that is all 16 anticodons must be represented, and in comparable numbers. A Noah's arc set of dRNAs, one of each isoacceptor, would comprise nearly 600 nucleotides (576 nt = 16 x 36 nt). If some isoacceptors went missing from the set, there was strong selection for some means of restoring them. Finally, the set of dRNAs might drift apart in sequence, perhaps improving the performance of the particular duplicon-anticodon helix, but challenging the co-evolution of ribozyme P, as well as the duplisome decoding and transfer centers, that must work with the entire family of dRNAs.

Interestingly, the problems of 3' end duplication, and maintaining a complete and balanced dRNA set, are closely connected. What happens when we duplicate any template RNA to the very end likely depended on whether the final codon has all three or just two nucleotides. In Figure 12-1 we show an odd length template ending GCAUG 3' with the final AUdRNAUAU at the A-site after nucleotidyl transfer when the duplicon A3 U2 matching the codon A3 U2 G1 has been added. The final template nucleotide

G1 is missing from the copy, but is too short alone to form the first position of do not form a complete codon for another elongation. The common fate is IDEA translocation occurs so free dRNA is in E-site polynucleotidyl-dRNA is in the P-site. Insufficient template for another dRNA in A-site. PTC hydrolysis and release of the polynucleotide-OH 3’ and free dRNA

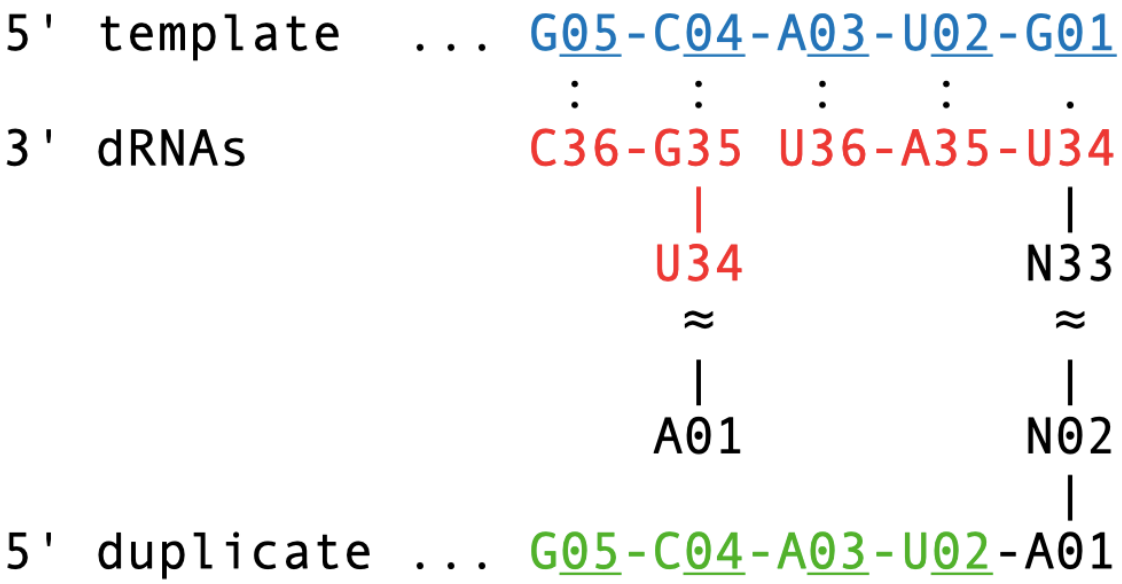


FIGURE 12-1. TERMINATION OF ODD LENGTH TEMPLATE WITH END EROSION

Here we suggest that the final two nucleotides of an even length template were read with somewhat less fidelity than the final three nucleotides of an odd length template. The result was a full-length duplicate without end erosion but with possibly nucleobase substitutions in one, or more rarely, both of the final two positions. For RNAs generally, this error-prone decoding, or *end wobble*, meant that the final two nucleotides were more variable than internal positions. This may have been important for primitive mechanisms of recombination and repair. For dRNAs particularly, end wobble ensured any vacancies in the set of isoacceptors were soon filled, and balance restored. It also allowed the dRNA core sequence (nucleotides 2-33) to evolve independently of the duplicon-anticodon helix so that the optimal distribution of dRNA

length and sequence in the core stem and loop could adapt to changes in the duplisome, ribosome P, and the physical environment.

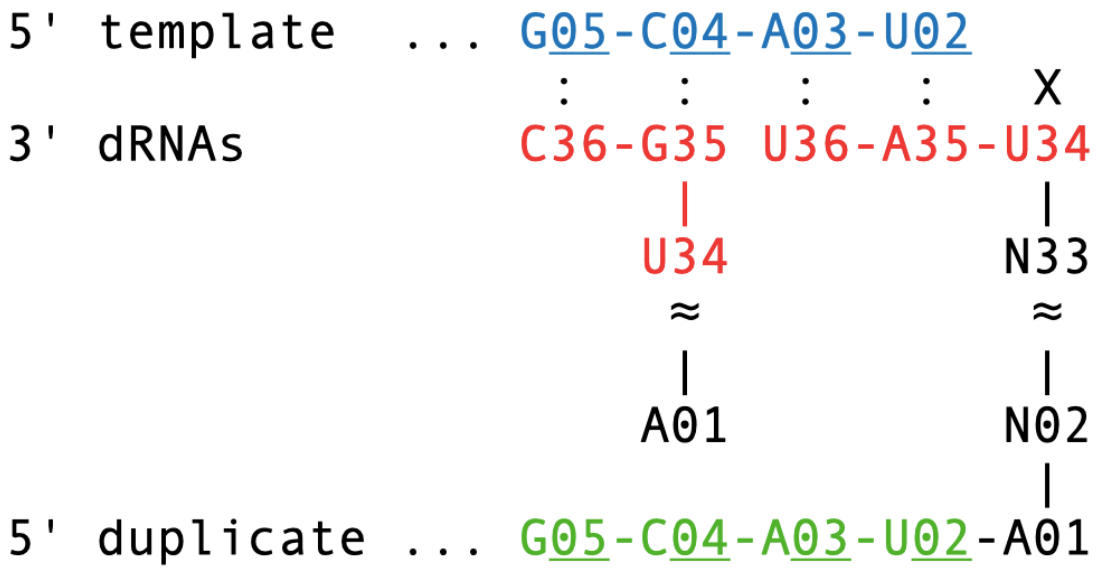


FIGURE 12-2. TERMINATION OF EVEN LENGTH TEMPLATE WITH END WOBBLE

In Figure 12-3 we depict the final covalent intermediate in the duplication of an unfolded dRNA template (without the 5' duplicon). In this example, the covalent intermediate is the tandem dRNACAdRNAUG of length 36 + 36 = 72 nucleotides. After translocation to the P-site for hydrolysis (small blue arrow) and release, the product and by-product are a duplicate of the template dRNACA in the nascent polynucleotide exit tunnel, and the freed dRNAUG that translocates to the E-site. In the absence of end wobble, these two dRNAs load complementary duplicons. In section 13 we use these tandem dRNAs that arise as physiological intermediates in dRNA duplication as ancestors of the tRNA.

FIGURE 12-3. TANDEM dRNA INTERMEDIATE IN DUPLICATION OF dRNA TEMPLATE

The example above of duplication to the 3' end followed by hyrolysis and release from the P-site in the absence of duplicon-dRNA in the A-site. The other striking feature / question is termination and release.

[MOVE UP to section 11 polynucleotide hydrolysis NOT = release]

Where and how was RNA duplication terminated? … at the end of the template … In section 8 we proposed that the nascent RNA was released from the nucleotidyl transfer center if the hydrolysis was not shortly followed by condensation with the duplicon-dRNA in the A-site. One simple suggestion is that in the absence of a duplicon-dRNA in the A-site for condensation, the nascent polynucleotide was simply released through its exit tunnel at the peak warmth, and the free dRNA from the E-site

in the cooling phase. [distinguish between absent dRNA and absent duplicon only] The possibility of constitutive termination codons with no matching dRNA.

[MOVE UP] In RNA duplication, hydrolysis is the obligated first step of nucleotidyl transfer, followed by condensation in ordinary elongation, release, or amino acid tagging (section 13). hydrolysis-condensation hydrolysis-release hydrolysis-tagging-release

What happens when we attempt to duplicate an individual dRNA from end to end?

can dRNA duplication repair some forms of damage? restore balanced pool?

it may still be a oligomer substrate for dRNA loading or a template for RNA copying; seems like we may want a way to repair 5' end of dRNA, a 5' kinase ribozyme? or transferase? the things we rejected in section 8

Our concept of vertical gene transmission, including faithful replication, as well as genome protection and repair, falls under the search theory rubric of preserving and exploiting our maps of the environment. Here evolution is viewed as the inheritance and preservation of knowledge, without acquisition of new knowledge through discovery and invention. This incomplete perspective exaggerates certain limits and perils of evolution captured in such phrases as genetic load, error catastrophy, and mutational meltdown. Long before 20th century population genetics formulated the problem of purifying selection, 19th century critics of Darwinism mocked the theory of natural selection as … survival of the fittest but not arrival of the fittest. … notion of small steps of variation in all directions, not large steps … monsters and sports … twentieth-century population genetics could fixation of favorable alleles and haplotypes was the obverse of elimination of unfavorable ones … the creative aspect of sexual reproduction was reduced to … linkage disequilium … Muller's ratchet

Our concept of horizontal gene transmission, including mixis and recombination, fails under the rubric of exploration, discovery and invention in search theory. remarkable unforeseen and unforeseeable opportunities, breakouts … immediate advantages and emergent opportunities

[exploration]The second problem is discovery …. not as Darwin supposed small variation in all directions … most common mutations have been exhaustively tried and

selection to reduce their frequency …. without violating teleos, explore directions believed promising based on recent experience, abductive bias …. the mechanisms of recombination underlying gene duplication, regulation of euchromatin and heterochromatin, to maximize R & D without compromising the faithful transmission of well-tested or tenured genes .. mixis, horizontal, risk selfish and useless at cost of discovery …. recombination in regular mixis, recombination in radical exploration

sexual species, core and pangenomes

idea population genetics is not creative; survival of fittest, arrival of fittest

survival of the fitest, not arrival of the fitest

The fundamental error in 20th century population genetics as a theory of creative evolution was formulating the invention and selection of favorable and unfavorable mutations as complementary problems, so that discovery was not a creative problem but mirrored purifying selection [create/degenerate]. [mutation was invention] The complexity of search … that exploration is a higher-level of search complexity than exploitation … this is remarkably obvious in the evolution of polymer life were the space of alternative sequences of length m has size 4m or 20m, while the space of recombinant sequences 4m+n or 20m+n. The exponential increase means that sequences that have exhausted their searches are only small seeds within the mixed space. Nonetheless, mixis of RNAs, tandem duplications, and end ligations form the … for exploration of genome space in the RNA world.

[creative, degenerative

[initiation] [termination] [release] [recycling]

Here we discuss several problems of RNA copying peculiar to the duplisome, as well as relations of RNA duplication to generic problems of recombination and repair in the RNA world.

In this section we discuss …

In protein translation, the initial codon not only determines the start of translation, but more crucially, the reading frame. In RNA duplication, although there are two frames, which utilize distinct sets of loaded dRNAs, their polynucleotide products are

identical apart from the terminal duplicons. Early duplisomes likely had few requirements for template recognition, and no preferred dRNA or start codon for initiation. Like translation of leaderless mRNAs today, intact duplisomes with a loaded dRNA in the P-site may have accommodated the 5’ end of the template to commence duplication on the first recognized codon (Leiva & Katz 2022).

Finally, when the end of the template was reached, or no more codons could be matched, termination of elongation, release of the nascent polynucleotide and recycling the duplisome.

cis template swapping to readthrough breaks or small gaps in the template, rejoining the correct ends albeit with short indel or base substituions …. FOOTNOTE NHEJ of dsDNA in modern cells ….

ASIDE unlike templated ligation and primer extension, direct duplication is extremely short-sight … their is no duplex longer than the codon and method for joining sequences based on longer regions of sequence identity …. conversely allows blunt-end? recombination … scanning and slippage …. the helicase action of template unfolding and co-duplicational refolding of template …

Beyond the reparative role of template rethreading, or *cis template swapping*, duplisomes likely played a creative role in recombining sequences by *trans template swapping*. There are two obvious ways that duplisomes might have mediated RNA recombination through *template swapping*. Stalled partway along the template, the duplisome might disassemble, allowing the large subunit with nascent nucleotidyl-dRNA in the P-site to reassemble with a different small subunit, and resume duplication on a new template (Figure). The product is ….

Alternatively, stalled at the end of the template, the duplisome might remain assembled as a *donor-template RNA* entered at the A-site, induced hydrolysis-condensation with the P-site nucelotidyl-dRNA in the P-site, swapping its 3' template-like domain into the empty template channel. The product is … We conjecture that this mechanism of template-swapping without disassembly gave rise to bacterial trans-translation based on the *transfer-messenger RNA* that rescues ribosomes stalled at the ends of nonstop mRNAs.

IDEA what precisely is the structure of the dtRNA? duplicon-dRNA-anticodon-template

?is there an end stalled problem at all; idea tmRNA evolved from dtRNA for recombination not ribosome rescue?

IDEA the duplisome-mediated template swapping for repair and recombination of copy but here we speculate on likely simpler older mechanism in RNA life for break repair and recombination of templates themselves … the first is a primordial ligase as ancestor of the nucleotidyl transfer center of the duplisome …

(I) forms nucleotide bonds - forms peptide bonds - forms coded polypeptides

(II) forms nucleotide bonds - forms templated polynucleotides - forms coded polypeptides

Here we propose the duplisome nucleotidyl transfer center originated from a primordial RNA ligase ribozyme that spliced together any two polynucleotides via a homologous chemical reaction, *viz*. sequential hydrolysis-transesterification.

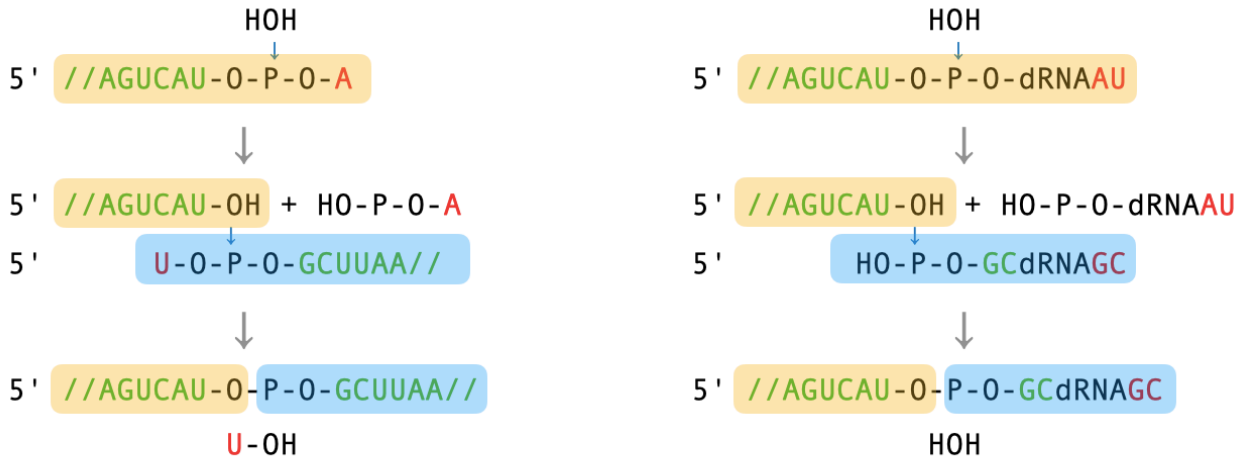


FIGURE 12-. PRIMORDIAL LIGASE RIBOZYME (LEFT) & DUPLISOME NUCLEOTIDYL TRANSFER CENTER (RIGHT)

In Figure 12- we depict sequential hydrolysis of the P-site RNA, creating a 3' OH and 5' phosphate, followed by transesterification with the A-site RNA.[18] The transesterification step of the ligase ribozyme is chemically homologous to the condensation step of the polynucleotidyl transferase center, but thermodynamically more favorable. Complementing the unknown, prebiotic processes of templated RNA ligation, the primordial ligase ribozyme catalyzed non-templated RNA ligation, making it one of the principal workhorses of recombination and repair in early RNA life. Thus,

[18] For simplicity, the non-bridging oxygens are not shown. The small blue arrows denote the movement of electron pairs from nucleophile oxygen to electrophile phosphorus. For simplicity the electron movements from the electrophile to leaving group oxygen are not shown.

any two RNAs could be joined near their ends in an energetically favorable reaction with a minimal sacrifice of the terminal nucleotide of each. Evolved for terminal joins, the P-region tunnel of the ligase ribozyme allowed RNA substrates with indefinite 5' extensions, while its A-region tunnel allowed RNA substrates with indefinite 3' extensions.

We suggest that the primordial ligase solved one grave threat to the elongation cycle of duplisome life, damage to the 5' end of dRNAs including the loss of the 5' phosphate.

IDEA the ancestor of ribozyme P that adds two nucleotides to 5' end of any phosphorylated RNA, the primordial ligase that preys on first and last nucleotides of A and P site RNAs, or just hydrolyzes the P site or ??????????

We suggest two ribozymes, ligase and elongase, were the principal ribozymes of RNA life before ribozymatic copying, allowing both quick and dirty repair of broken strands, and occasionally joining unrelated RNAs.

Whereas duplisome-mediated template swapping could create new RNA combinations in the course of duplication, without consuming either template, the primordial ligase ribozyme created new combinations at the expense of both. A drawback of the ligase ribozyme for break repair was the product of each round of repair was strictly shorter than the original RNA. We suggest that a second workhorse repair ribozyme, ancestral to ribozyme P, extended the 5' end of RNAs by say two nucleotides. Working together, the primordial ligase and this 5' elongase allowed repeated repair of broken strands without net loss of nucleotides, or change in length.

[details of 5' elongase, Figure]

[ribozyme P/Q, adds 2 nucleotide leader using guide sequence][DCC]

Unlike the primordial ligase ribozyme, the 5' elongase uses a guide sequence …

Is this ribozyme ancestor of DCC which has lost its ability to condense and hydrolyze, …. and co-opted the primordial ligase to hydrolyze and condense. If so, RNase P RNA and the ribosome decoding center have a common ancestor.

[MOVE section 11] group I introns are MGE, excise by splicing and integrate at homologous target sites defined by pairing with internal guide sequence by reverse-splicing

Self-splicing introns, the first mechanism of site-specific recombination, are believed to have originated in the RNA world, and then continued, via reverse transcription, into DNA life. Interestingly, there are not one but two, ancient families of self-splicing introns. Viewed as mobile genetic elements, these introns insert themselves into a host RNA by reverse splicing, where they need contribute little to share whatever prospects the host confers, so long as they can remove themselves precisely by splicing, allowing an uninterrupted host fold, and available for new insertions. In this abortive cycle of reverse then forward and splicing at the same site, introns did no great harm, nor much good, nominally mobile genetic elements but with really no where to go. But once there are two copies of the element at the same site in homolgous RNAs, or at distinct sites in non-homolgous RNAs, … roles in recombination, a productive form of *trans* splicing.

In RNA life …. (1) maintainance of large RNAs by avoiding the mutational meltdown and Muller-ratchet. (2) concerted evolution of families [idea these are not deleterious mutations per se but the drift apart of multicopy genes is problematic] including mix-and-match families, concerted evolution {goal is tRNA families}, (3) creative evolution …. More productive variations of the life cycle to introduce novel elements at novel sites, or create novel sequence combinations. internal guide sequence Two introns of the same type can catalyze recombination …. unlike the irreversible random joins of the ligase ribozyme, introns provided dedicated sites for regular exchange …. mixis by assortment of unlinked RNAs in compartments and mixis by recombination and exchange …

(Zaug & Cech 1985) intermolecular ligation to form oligomeric IVS

heating-cooling cycle of denaturation and renaturation

Zaug AJ & Cech TR (1985). Oligomerization of intervening sequence RNA molecules in the absence of proteins. Science 229, 1060-1061.

(Weiner & Maizels 1987). tRNAs at the 3’ ends of RNAs acted as telomers against end erosion; ribozyme P evolved to remove these tRNA-like tags; substrates for CCA addition and aminoacylation … ancient untagging activity

Watson (1972) nature end erosion problem

Watson, James D. "Origin of concatemeric T7DNA." *Nature New Biology* 239 (1972): 197-201.

Reanney (1979) Nature

Darnell Doolittle (1986) PNAS

[RNA modification]

Besides the 5’ cap and 3’ tail structures that mark or protect RNA ends, a remarkable variety of internal RNA modifications (1) stabilize the folded structure and (2) protect critical sequences from spontaneous damage, biotic nucleases, or antibiotics. These modifications have expanded their functions to (3) balance interaction with multiple partners, such as anticodon with alternative codons, (4) checkpoints in biogenesis, markers of quality of completed ribosome (5) regulation (6) physiological toggle between alternative functions (7) evolutionary silencing of vestigial functions. Many RNA modifications trace back to LUCA, and perhaps to protein life, or even earlier .. in ribosomal RNAs and tRNAs …strong first impression that concentrated. at import sites of substrate interactions and catalysis

IDEA RNA modification enzymes using sRNA guides to target modification may have been ribozymes complexed with sRNA guides in RNA world to protect critical sites

Hints that some RNA modifications trace to RNA life … guide RNAs and ribozymes …

ribozymatic RNA modification guide RNAs like snoRNAs of archaea and eukaryotes today

2’ OH methylation protects against strand scission

[guide RNAs for 2’-O-methyl] C/D box sRNAs, or less often H/ACA box sRNAs

snoRNA C/D box family : methylation

snoRNA H/ACA box family : pseudouridylation

snoRNAs U8, U3 endonucleolytic cleavage pre-rRNA [**MOVE kinetic compartments**]

13. From programmed polynucleotide termination to amino acid tagging

Beyond exhuming prebiotic and ribozymatic processes of RNA copying, the origin of polypeptide translation is the greatest challenge to tracing the history of life back to an RNA world. Our universal genetic code maps codons to amino acids, and by iteration, mRNAs to polypeptides. This map is many-to-one, that is any codon assigns (at most) one amino acid, while any amino acid may have six (L R S), four (A G P T V), three (I), two (C D E F H K N Q Y), one (M W), or no codons (non-proteinogenic amino acids) at all. The function is partially defined as three codons (UAA UAG UGA) map to no amino acid.[19] The brilliant mechanistic insight and experimental breakthrough was that this amino acid code is the product of two simpler relations involving small adaptor RNAs, known now as transfer RNAs (Crick 1955/1958; Crick et al 1957; Hoagland 1959; Zamecnik 1960; Fry 2022). Each tRNA, complexed with its cognate aminoacyl-tRNA synthetase (aaRS), is charged at its 3' end with an amino acid, and then, complexed with the ribosomal decoding center, is matched via its anti-codon to an mRNA codon. These charging and matching relations together specify the amino acid code.

Untangling the origins of protein coding, we analyze polypeptide translation *before support from coded proteins*. At this breakout stage of protein life, anything that today requires coded proteins, e.g., RNA modification, tRNA charging, elongation factors, and protein secretion, was catalyzed by (possibly extinct) ribozymes and non-coded polypeptides, happened spontaneously, or did not happen at all. Under this assumption, we compare two scenarios for the origins of translation: First, ancestors of rRNA and tRNAs had no recognized functions in the late RNA world, or second, these molecules functioned in RNA duplication more-or-less as sketched in sections 8-12. The first hypothesis places no constraints on primitive ribosomes and tRNAs beyond parsimony with modern ones, while the second hypothesis requires parsimony with ancient duplisomes and dRNAs, their conjectured progenitors, as well as with modern ribosomes and tRNAs, their extant progeny.

[19] Curious exceptions in recoding of stops codons for selenocysteine and pyrrolysine (ref).

Era	RNA	polypeptides
early RNA life	spontaneous copying	random
late RNA life	duplisome	random
polypeptide life	duplisome →	ribosome
protein life	RdRP enzyme	ribosome

TABLE 13-1. FROM INVENTING THE RIBOSOME TO RETIRING THE DUPLISOME

Along any evolutionary path from RNA duplication to polypeptide translation there were key changes in (1) the structure and charging of adaptor RNAs, (2) the mechanism of decoding, (3) the chemistry of polymer transfer, (4) the process of translocation, and (5) the energetics of elongation. Meanwhile, duplisomes remained the principal mechanism of RNA copying right up until the invention of RdRP enzymes. Hence, remodeling them for protein translation was akin refitting a ship at open sea, not in the shipyard. There were, we suggest, two key points of regulation, (6) one focused on the adaptor RNAs, and (7) another on the ribosome née duplisome, that allowed concurrent RNA duplication and polypeptide translation without costly redundancy or mutual interference. We identify the RNA changes associated with each step, and suggest their chronology along an adaptive path from late RNA life to early protein life (Table 13-1). In apologetics of our just-so story of ribosome evolution, we provide explanations for several hitherto inexplicable, features of cellular life, venture a few testable predictions, and raise some unexpected questions.

In this section, we place two, defeasible constraints on any possible evolutionary path from RNA duplication to polypeptide translation that make the breakout problem tractable, and our solution unique: Our first conjecture, the presence of a brisk market in amino acids and random polypeptides in the late RNA world, is clearly prerequisite for the breakout of polypeptide translation. Thus, despite the moniker “RNA world”, we assume that simple amino acids and random polypeptides were abundant and useful to late RNA life (Cech 2009; Frenkel-Pinter et al 2020). These polypeptides likely included intrinsically disordered sequences that co-assembled with lipids to form membrane pores, or with polynucleotides to form mixed polymer coacervates and discrete ribonucleoproteins (ref). During the long *conviviencia* of templated nucleic acids and random polypeptides in the RNA world, there was strong selection for ribozymes that improved the syntheses of amino acids and useful polypeptides. To focus on the breakout of polypeptide coding, we simply stipulate several aspects of polypeptide synthesis that do not prejudice the mechanism of RNA copying.

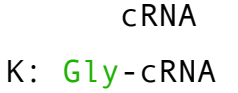
Our second conjecture is far bolder, namely, that before any iterative translation of polypeptides, the duplisome acquired the means of tagging nascent polynucleotides with one terminal amino acid as a modification of polynucleotide release. Thus, the first tRNAs and their charging ribozymes were invented to improve polynucleotides, not to make polypeptides. In our chronology, the changes for template-programmed polynucleotide termination (this section), leading to the mature amino acid tagging code, precede the changes for mRNA-programmed polypeptide initiation, elongation and release (section 14).

To frame the breakout of polypeptide translation, we briefly discuss likely sources of amino acids and random peptides in the late RNA world. Beyond the abiotic sources of amino acids and spontaneous condensation of polypeptides (section 2), ribozymes had no doubt evolved for activation of amino acid and peptides, and formation of peptide bonds in the heyday RNA world (ref). There are various proposals for the initial high-energy substrates handed to ribozymes, as well as the downstream intermediates in the catalyzed pathway to polypeptides (Liu et al 2020). Extrapolating from modern cells, the most likely activated carriers of amino acids and polypeptides in late RNA life were nucleosides or polynucleotides, notably their 5’ acyl-phosphate mixed

anhydrides, and 2’ or 3’ acyl-esters. The free energy of hydrolysis for aminoacyl-phosphates is ~ 4 kcal / mol higher than aminoacyl-esters, which in turn, is ~ 8 kcal / mol higher than peptide bonds (Carpenter 1960). Thus, aminoacyl-phosphate anhydrides of RNA carriers were likely abiotic inputs, or early intermediates in ribozymatic peptide synthesis (Leman et al 2006).

A handful of proposals for ribozymatic formation of random polypeptides (ref). Without prejudicing the question of RNA duplication, we adapt the simple scheme of Koji Tamura and Paul Schimmel (Tamura & Schimmel 2003). To wit, amino acids are charged at the 5’-phosphate of the *source RNA*, and then transferred to the growing polypeptidyl-ester at the 3’ OH of the *target RNA*. In model reactions on tRNA mimics, aminoacyl-5’-phosphate of a source RNA is positioned near 3’ OH of a target RNA through direct base-pairing of source and target RNAs, or via a bridging RNA intermediate (Wu et al 2021). With suitable reactants the product can be extended from the aminoacyl-ester to peptidyl- and dipeptidyl-esters. One curious result is that the D-ribose of modern RNA favors charging with the L-amino acid of modern proteins (Tamura & Schimmel 2006). Those authors suggested that as RNA communities settled on D-ribose, they put the imprimatur of nucleotide chirality onto amino acids and peptides through this charging reaction.

The gist of the ribozymatic synthesis of random polypeptides is shown in Figure 13-1. For sake of discussion, we name these conjectural RNA substrates and catalysts: *Ribozyme K* amino-acylates the 5’-phosphate of source RNA from the (abiotically activated) amino acid. There may have been several variants of this charging ribozyme to accommodate diverse amino acids, or to regulate polypeptide composition. *Ribozyme L* and *ribozyme M* catalyze the first acyl transfer (ester formation) and subsequent transfers (amide formation), respectively. Conceivably, one and the same ribozyme might catalyze both ester and amide formation. Finally, *ribozyme N* hydrolyzes the peptidyl-RNA ester, although spontaneous hydrolysis might suffice. Thus, the rate of spontaneous hydrolysis of peptidyl-tRNA is about 1 per 14 hours outside the ribosome.



```
L:      cRNA-Gly
K: Asp-cRNA-Gly
M:      cRNA-GlyAsp
K: Val-cRNA-GlyAsp
M:      cRNA-GlyAspVal
N:      cRNA + GlyAspVal
```

FIGURE 13-1. RANDOM POLYPEPTIDE FORMATION (AFTER TAMURA & SCHIMMEL 2003)

The ribozymatic activities in the Tamura-Schimmel pathway of polypeptide formation might be associated with carrier or bridging RNAs *tout court*. In any event, these RNAs contribute to catalysis through the proximity and orientation of reactants. For simplicity, we show no bridging intermediate, and use one carrier RNA (cRNA) as both source and target in the elongation cycle, providing the 5'-phosphate for amino acid charging and the 3'-OH for peptide growth, respectively. One striking difference from ribosomal protein translation on tRNAs, or nonribosomal protein synthesis on thioester carriers, is that the amino acid is transferred to the N-terminus of the polypeptide, rather than the polypeptide to the N-terminus of the amino acid (Lipmann 1971; Finking & Marahiel 2004). As a consequence, polypeptides in the Tamura-Schimmel pathway are made from C-terminal to N-terminal. … free energy of hydrolysis of kcal / mol compared to kcal / mol for aminoacyl-esters.

We suggest that polynucleotide termination by amino acid tagging was exapted from the hitherto independent pathways for random polypeptide synthesis and RNA duplication, respectively. It likely began as a simple tag-and-release option used to control the length of nascent polynucleotides, protect their 3' ends from recombination or degradation, or target them to specific macromolecular compartments. One primitive tRNA and charging ribozyme sufficed for simple tagging. This charging complex no doubt accepted a variety of amino acids, including ones no longer found in proteins, according to their affinity and availability. It likely favored diversity over specificity, so that tagged polynucleotides were not so much targeted to compartments and complexes, as selected by their affinity for them.

First we explain the origins of tRNAs and tRNA charging ribozymes: To wit, a primitive tRNA arose from tandem dRNAs made in the ordinary course of dRNA duplication (section 12). This first tRNA was charged by existing ribozymes KL of random polypeptide formation. Second, we explain how the primitive aminoacyl tRNA interacted with the duplisome decoding and nucleotidyl transfer centers. If the origins of this tRNA and its charging ribozymes are plausible enough, it seems miraculous that the duplisome A-site evolved for the duplicon-dRNA could accommodate an aminoacyl-tRNA, a substrate mimic fully twice its length. [compare the folding and accommodation of both adaptors][kinetics and energetics of non-covalent interactions in the decoding center, and covalent reaction of the nucleotidyl transfer center] We put the case that an awkward squeeze sufficed for simple tag-and-release, noting that selection for greater fidelity, and then speed, came with alternative tagging, and polypeptide translation, respectively.

Nonetheless, we notice that the structure of the tRNA anticodon arm had two profound consequences for the subsequent evolution of ribosomal decoding and translocation. First, steric hinderance from N37 of the A-site tRNA prevented dRNA tRNA scrunch on the template, that is, reading the -1 frame. This in turn, favored a triplet translocon at the breakout of polypeptide translation, and selection for the elaborate mechanisms of reading frame defense found in protein translation today (section 14). Second, absent pairing of the duplicon to anticodon in closed dRNAs, and duplicon displacement by the codon in dRNA opening (section 11), there was an empty cast at the decoding center, as well as selection for better tRNA decoding.

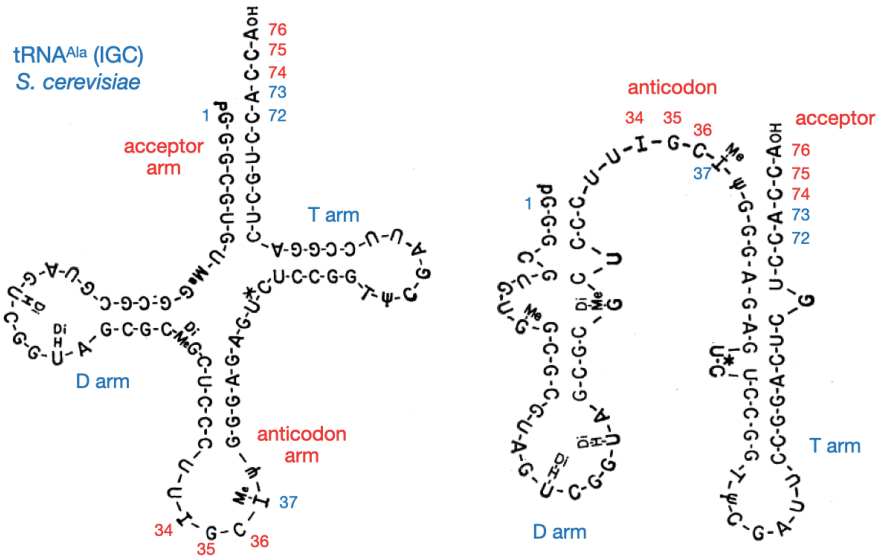


FIGURE 13-2. CLOVERLEAF & DOUBLE HAIRPIN STRUCTURES (HOLLEY ET AL 1965)

From the sequence of yeast tRNA^{Ala} (IGC), Robert Holley and colleagues proposed the four-way junction, or *cloverleaf* model of tRNA secondary structure (Figure 13-2). One alternative was an imperfect, double hairpin that extended the D- and T-stems at the expense of acceptor and anticodon stems. In various *tandem hairpin models* of tRNA evolution, the fusion of two hairpin sequences gave rise to the primitive cloverleaf tRNA (Eigen & Winkler-Oswatitsch 1981a,b; Di Giulio 1992, 2004; Dick & Schamel 1995; Schimmel & Ribas de Pouplana 1995; Nagaswamy & Fox 2003; Widmann et al 2005).[20] Besides the lengths of hairpin stems and loops, these models vary whether the hairpins were nearly perfect, weakly paired, or bulged, and whether the two tandem hairpins were nearly identical, highly diverged, or entirely unrelated. More importantly, these models vary in proposed functions of the ancestral hairpins in RNA replication or polypeptide synthesis, in when, how and why hairpin fusion occurred, and in when, how and why the family of tRNAs radiated.

We illustrate some common themes in tandem hairpin models of tRNA origin with one example. Massimo Di Giulio conjectured that two identical RNA hairpins gave rise to the D-arm, and the T-arm, respectively, with both halves contributing equally to the

[20] We leave several non-hairpin models of the origin of cloverleaf tRNAs from simple helices or repeating sequences to interested readers (see Agmon 2022).

acceptor and anticodon stems (Di Giulio 1992, 2004). For sake of discussion, he proposed an ancestral hairpin of 38 nucleotides, comprising a perfect stem of 12 basepairs, a loop of 10 nucleotides, and a 3' trailer of 4 nucleotides (Figure 13-3). In his model, the ancestral hairpins were already a family of sequences with different triplet anticodons (ANT), while the 3' trailer comprised the discriminant (D73) nucleotide and universal CCA of modern tRNAs. This model was agnostic about the mechanism of fusion, but suggested it introduced the variable arm, and that the anticodon and discriminant nucleotides of the 3' hairpin formed the amino acid identifier (ID) sequence of the tRNA acceptor arm.

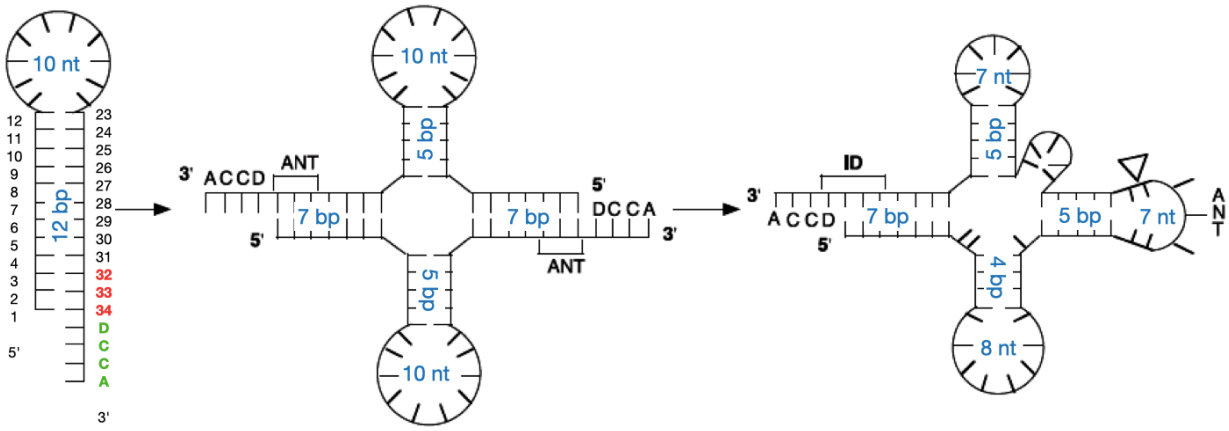


FIGURE 13-3. TANDEM HAIRPIN ORIGIN OF TRANSFER RNA (AFTER DI GIULIO 1992, 2004)

Several aspects of the chemistry and biology of modern tRNAs have been taken as evidence for or against a tandem hairpin origin. Double hairpins have not been reported as folding intermediates in tRNA biogenesis (?is this true?), nor in heating mature tRNAs *in vitro* where, after melting the tertiary elbow structure, D-stems are generally the first, and T-stems the last to melt under various salt conditions (see Privalov 2012). Nonetheless, some tRNAs can adopt a double hairpin as a minor conformation *in vivo* or *in vitro*. For example, mature tRNAiMet from *Drosophila*, forms a double hairpin as a minor conformation in vitro as evidence by cleavage into two fragments by RNase P RNA from *Escherichia coli* (Kikuchi et al 1990). At least two other *Drosophila* tRNAs can be cleaved after the anticodon *in vitro* by the bacterial RNase P ribozyme (Hori et al 2000; Tanaka & Kikuchi 2001).

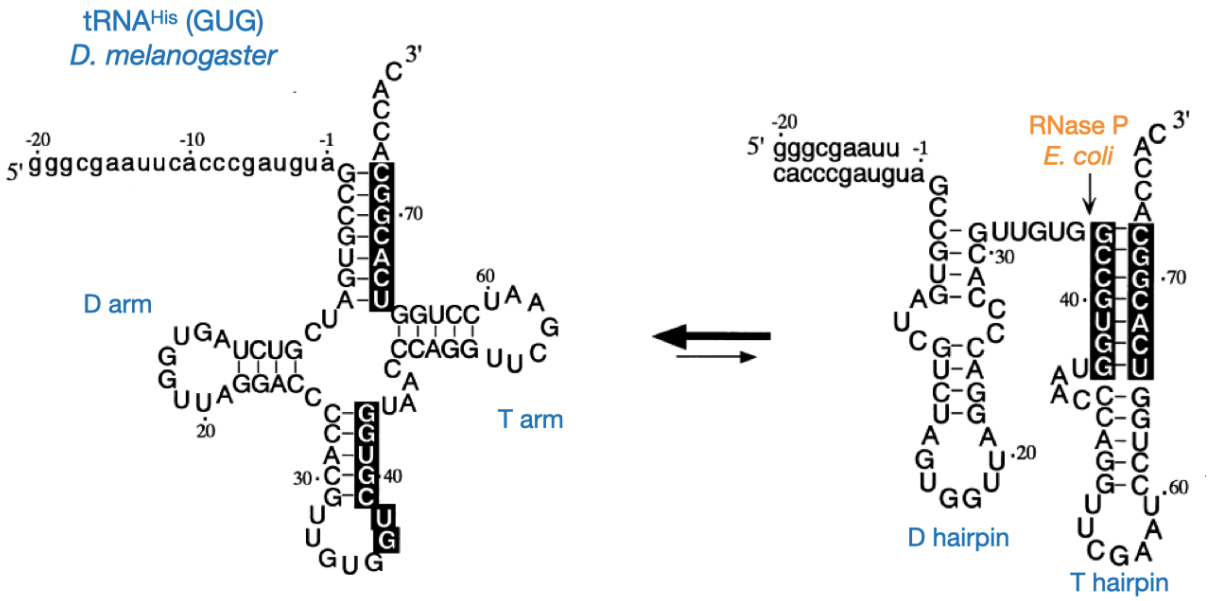


FIGURE 13-4. INTERNAL CLEAVAGE OF tRNA BY RNAse P RNA (TANAKA & KIKUCHI 2001)

The *in vitro* cleavage of certain insect tRNAs into two halves by bacterial RNase P, or hyperprocessing, provided evidence of double hairpin conformations as the likely source of . but the physiological role of double hairpins. The motivation behind these experiments was that the 39 nucleotide 5' fragment of tRNAiMet, and not the 3' end of the intact tRNA, is the primer for minus-strand reverse transcription of some insect copia/Ty1 retrotransposons (Kikuchi et al 1990).

In vivo the cleavage of tRNAiMet at nucleotides 39/40 produces the 39 nucleotide 5' tRNA fragment which ….has a physiological and evolutionary significance in

This cleavage likely uses the retrotransposon RNase H *in vivo* where the 39 nucleotide 5' fragment, not the 3' end of the intact tRNAiMet, is used as primer for minus-strand reverse transcription of some copia/Ty1 elements (ref). Although the physiological relevance of the double hairpin *in vitro* hyperprocessing of tRNAs by bacterial RNase P RNA is unclear, and retrotransposon replication seemed the sole example of in vivo processing the existence of double hairpins in pre-tRNAs or tRNAs … of double hairpin conformations in pre-tRNAs, or their hyperprocessing by RNase P, is unclear (Figure 13-4). More recently, a large literature has emerged on the processing

of pre-tRNAs and mature tRNAs into tRNA fragments (tRFs) with functions in regulation (section 14).

The splicing of tRNAs with intron ….[tRNA splicing]

IDEA we want dRNA N34 N35 N36 to be exactly tRNA N34 N35 N36

tRNA = 36 + 36 =72 and group I intron NOT between 37/38 and need to gain 1 nt after initial anticodon

What happens when group I intron inserts into de novo site? free guanosine?

MODEL before/after RdRp enzyme is available group I intron inserts into double dRNA / joins not contiguous dRNAs creating full tRNA that can be copied

[MOVE section 12]

dRNA GISSI template yield donor-trailer RNAs

template GISSI dRNA yield leader-donor RNAs

QUESTION are group I and group II introns different in their insertion site relative to the host target sequence (that is, more 5’ versus more 3’?)

Our proposed origin of tRNAs from tandem dRNAs differs from other tandem hairpin models in three important ways: First, whereas those theories have either (1) no particular function in mind for the ancestral hairpin, (2) a role in tagging RNA templates for recognition by the replicase ribozyme, (3) a role in priming RNA replication, (4) an acceptor function in random polypeptide synthesis, or (5) an anticodon function in primitive decoding, we propose that (6) dRNAs were central players in duplisome-mediated RNA copying. Second, the functional requirements of duplicon loading, template decoding, duplicon transfer, and translocation, constrain dRNA structure, so our primitive tRNA must be parsimonious with this progenitor structure, not just the structure of modern tRNAs. Finally, whereas most tandem hairpin models are agnostic about how and when hairpin joining occurred, tandem dRNAs were likely physiological intermediates in the duplication of dRNAs (section 12) before their exaptation for amino acid tagging.

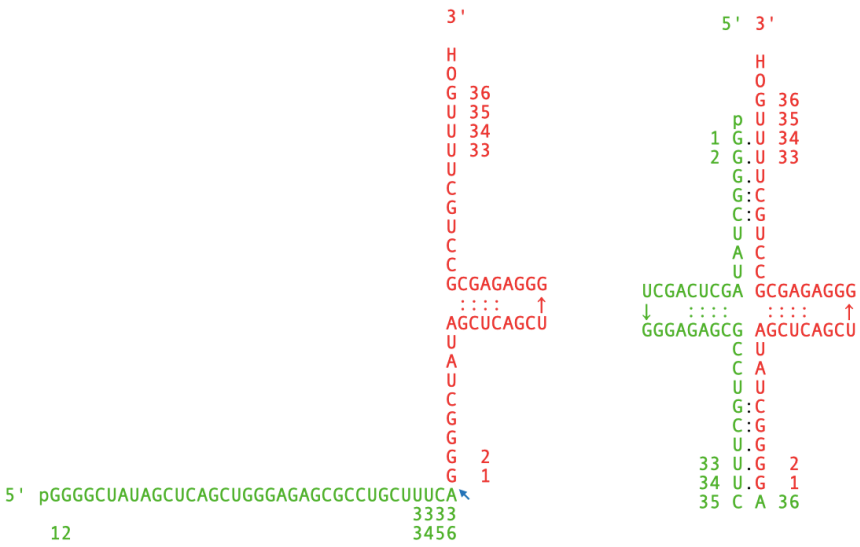


FIGURE 13-5. TANDEM dRNAS FOLDED AS PRIMITIVE CLOVERLEAF tRNA

For at least some tandem dRNAs, the cloverleaf was an alternative to the double hairpin fold. Figure 13-5 (left) shows the tandem dRNA taken from Figure 12-. If it were released intact from the duplisome, *viz*. without hydrolysis (small blue arrow) in the nucleotidyl transfer center, it might fold as a double hairpin for hydrolysis ("hyperprocessing") by ribozyme P either before or after the internal anticodon. The only necessary structural relation between the component dRNAs is that the anticodon of the 5' dRNA is one-and-the-same dinucleotide as the duplicon of the 3' dRNA. These components are drawn from the common pool of dRNAs, the 5' dRNA as the duplicated template and the 3' dRNA as the final substrate. If as shown in Figure 13-5, they were identical dRNAs (excepting their complementary anticodons), the anticodon and acceptor stems would correspond in length and sequence to the distal stem of the closed dRNA. *More generally*, the two halves of the cloverleaf would be similar, not identical in sequence, and might pair together more perfectly, or less perfectly, than either of closed progenitor hairpin. Thus all dRNAs would form tandem hairpin intermediates during RNA duplication, but only rare combinations could also form stable cloverleafs. This is opposite to the situation in modern tRNAs where all sequences form stable cloverleafs, but only rare tRNAs can also form double hairpins.

In modern tRNAs, the characteristic elbow interactions between the D- and T-loops stabilize the orthogonal (L-shaped) arrangement of the coaxial helices formed by the D-stem plus anticodon stem, and the T-stem plus acceptor stem, respectively (Zhang & Ferre-D'Amare 2016). We suggest that the evolution of this elbow, the *sine qua non* of dedicated tRNAs compared with tandem dRNAs, commenced with the invention of amino tagging and continued past the breakout of polypeptide translation. By suppressing the double hairpin conformation, the elbow prevented internal cleavage (aka *hyperprocessing*) of tRNAs by ribozyme P, the ordinary fate of tandem dRNAs released from the duplisome. Other advantages that emerged later include better recognition by the charging ribozymes and accommodation by the ribosome.

Evolution of the tRNA elbow from tandem dRNAs began with sequence specialization of the D-loop née 5' dRNA loop and the T-loop née 3' dRNA loop, and continued with structural refinements by RNA modification (ref). As discussed in section 9, the original hairpin loop formed the closed dRNA for duplicon loading, and likely stabilized the open dRNA for decoding and accommodation. The two halves of the tRNA mimicked a pair of open dRNAs, paired from head to tail like synchronous skaters, except for their loops. Tertiary interactions between the D-loop sequence D16 D17 G18 G19 G20 and the T-loop sequence A58 G57 C56 Ψ55 T54 form the elbow. … modifications dihydrouridines D16 D17 and ribothymidine T54 pseudouridine Ψ55 ….At the center of the elbow the Watson-Crick pair between G19:C56 presents the flat face of these nucleobases to the solution.

Changes in three interacting RNAs followed the evolution of the tRNA elbow, either more or less immediately. While ribozyme P remained the means of dRNA loading by oligonucleotide charging and duplicon trimming, its role for tRNAs was limited to clearing any leader that might interfere with aminoacylation and transfer at the 3' end. The tRNA elbow suppressed internal cleavage of the double hairpin, and drove evolution of the specificity domain S of ribozyme P and its the molecular ruler mechanism to prevent adventitious duplicon loading at the 5' end (ref). … interdigitated double T-loop motif (IDTM)

Finally, interactions with the large subunit rRNA of the tagging duplisome, and breakout ribosome, at the A-site and E-site.

The interaction of tRNA elbow with the large subunit rRNA at helix H38 called the A-site finger ….. Remarkably, the interactions of tRNA elbows with the large subunit rRNA L1 stalk, with ribozyme P .., and with ribozyme T …. IDTMs …convergent evolution

interacts with the small subunit rRNA at both the A-site, with helix H38 called the A-site finger, and the E-site with an interdigitated double T-loop motif (IDTM) in the L1 stalk. Remarkably, the IDTM has evolved in two other RNAs, RNase P RNA and the T-box riboswitch recognition domain, for recognition of tRNAs by their elbows.

finger in small subunit rRNA where it helps ….

The tRNA elbow also interacts with the P-site

E-site L1 stalk IDTM

RNase P RNA J11-12 J12-11 IDTM

T-box riboswitch recognition domain stem I IDTM

In ribozyme T, the tRNA charging ribozyme (discussed below), the elbow recognition ….

The success of programmed polynucleotide termination by the first amino acid code with just one primitive tRNA and charging ribozyme, created the opportunity for pairing two (or more) tRNAs with cognate charging ribozymes. While the tRNA elbow and its interaction with the new S domain of ribozyme P thwarted the undesirable reactions of tRNA cleavage into halves, and duplicon loading at its 5' end, the necessary reaction, amino acid charging of the tRNA 3' end, required a new family of ribozymes. Few things about the origin of protein coding have been more puzzling than tRNA charging, catalyzed by twenty-odd aminoacyl-tRNA synthetase (aaRS) enzymes in modern cells, one for each proteinogenic amino acid, with some curious exceptions (Rubio-Gomez & Ibba 2020). Each aaRS enzyme recognizes one or more tRNAs by positive and negative identifier nucleotides in the acceptor arm, the anticodon, and elsewhere in the mature tRNA ().

To avert a chicken-or-egg paradox of needing a family of coded proteins (aaRSs) to make any coded protein, early molecular biologists looked for direct affinities of amino acids for tRNAs, or just their anticodons, to explain selective tRNA charging *before support of coded enzymes*. Fitting individual amino acids into the crevices of duplex DNA between adjacent basepairs, George Gamov launched the quest for

stereochemical fits between amino acids and code words of nucleic acids (Gamov 1954). Besides the combinatorial degeneracy and non-polarity of Gamov's code, Crick doubted that van der Waals interactions with the hydrophobic faces of nucleobases could discriminate between amino acid sidechains (Crick 1955). After elucidation of the triplet codons and anticodons of the tRNA-based genetic code, many looked for direct binding of amino acids and tRNAs, or spontaneous tRNA charging from aminoacyl-AMPs, or other activated amino acids. These searches failed to uncover any clear stereochemical fit between tRNAs and amino acids underlying the modern genetic code (Schimmel & Ribas de Pouplana 1995).

Under the RNA world hypothesis, the search for a primitive mechanism of selective tRNA charging shifted, from spontaneous reactions exploiting direct fits between tRNAs and amino acids, to reactions mediated by ribozymes. We can infer some likely features of those extinct ribozymes for tRNA recognition and charging from artificial ribozymes and extant riboswitches. A variety of artificial ribozymes can catalyze RNA aminoacylation. For experimental convenience, as well as their research objectives, these models of tRNA charging depart from the enzymatic reactions in one or both substrates. Supplying the electrophile (carbonyl carbon), chemically activated amino acids include AMP anhydrides (Illangasekare et al 1995, 1999), CoA thioesters (Li & Huang 2005), cyanomethyl and 3,5-dinitrophenol esters (Murakami et al 2006; Ohuchi et al 2007), and oxalones (Pressman et al 2019; Liu et al 2020; Janzen et al 2022). Supplying the nucleophile (alcohol oxygen), polynucleotide substrates include full-size tRNAs, minihelices, NCCA, or 2' OH of an internal ribose. In some models, the same RNA acts as catalyst and substrate, allowing direct selection of catalytically active sequences via self-aminoacylation (Pressman et al 2019).

Seeking a practical reagent to charge tRNAs with non-cognate, or even non-proteinogenic amino acids, Hiroaki Suga and colleagues perfected a short (46 nt) artificial ribozyme that acylates virtually any natural tRNA using 3,5-dintrophenol esters of α-amino- or α-hydroxy-acids (Lee et al 2000; Murakami et al 2006). This reagent dubbed the *flexizyme* recognizes only the terminal 3' NCCA of the acceptor arm, not the L-shaped fold, much less discriminates among tRNAs. Designed and selected for substrate promiscuity, the flexizyme says little about the likely substrates or

mechanism of extinct charging ribozymes for primitive tRNAs or other polynucleotides. More recently, ribozymes have been selected for self-aminoacylation at internal 2' OH using biotinyl-tyrosyl-oxazolone (Pressman et al 2019). Starting from three of these ribozymes as seeds, a spectrum of ribozymes have been selected for biotinyl-aminoacyl oxazolones of F I L M V W (Janzen et al 2022). In general, there is a positive correlation between specificity and activity.

One attractive conjecture is that T-box riboswitches, found in 5' UTRs of bacterial mRNAs, are vestiges of extinct tRNA charging ribozymes (Grundy & Henkin 1993; Suddala & Zhang 2019; Ishida et al 2020; Zhang 2020). These riboswitches recognize individual tRNAs, monitoring their aminoacylation status to regulate transcription and translation of enzymes involved in amino acid biosynthesis and transport, as well as aaRS enzymes. T-box riboswitches have been found in Gram-positive bacteria for isoacceptors of all twenty amino acids. The riboswitch 5' domain has an IDTM to recognize the tRNA elbow and a triplet *codon* to recognize the tRNA anticodon (Zhang & Ferre-D'Amare 2016). Together these two interactions ensure that the riboswitch has bound a tRNA of specified isoacceptor type. The riboswitch 3' domain pairs with the 3' NCCA of the uncharged tRNA, assessing whether or not an amino acid, or other small ester, is present, not whether it is the correct one for that tRNA. Different T-box riboswitches read-out the absence of an amino acid on tA76 in either of two ways: unmasking the anti-terminator to allow transcription to continue, or unmasking the Shine-Dalgarno box to allow translation to commence.

We suggest that the recognition domain of T-box riboswitches including the elbow IDTM and the anticodon specifier, descend from ancestral tRNA charging ribozymes. In combination with say ribozymes KL of the random polypeptide formation, this ancient RNA, or RNA complex complex could recognize the tRNA elbow and anticodon, and selectively charge the acceptor arm. For sake of discussion we refer here to this combination as *ribozyme TKL*. … tRNA specificity from ribozyme T and amino acid specificity from ribozymes KL … purely conventional, no stereochemical relation between codon/anticodon pairing and amino acid aptamer … the co-evolution and phylogenies of tRNA and charging ribozymes …

In proof of concept that T-box riboswitches can mediate selective tRNA charging, Suga and colleagues have recently engineered the T-box riboswitch from the *Bacillus subtilis* glyQS gene to charge its cognate $tRNA^{Gly}$ (GCC) using phenylalanyl-cyanomethyl ester (Ishida et al 2020). Whether or not these riboswitches descend from extinct tRNA charging ribozymes, the chimeric ribozyme demonstrates the feasibility of an amino acid code based purely on RNA catalysts and carriers. As these modular RNAs inspect well-separated features of the L-shaped tRNA from elbow to anticodon to acceptor NCCA, it is easy to see that the matching of anticodons and amino acids in the early code may have been purely conventional with no subtle stereochemical predispositions. That is, the modular recognition elements for tRNAs and amino acids were entirely distinct, and their combinations in the amino acid code were determined by accidents of RNA pairing or ligation.

IDEA the two steps of charging KL have specificity for amino acid from K, but perhaps no specificity for aminoacyl-phosphate-carrier cRNACCA.

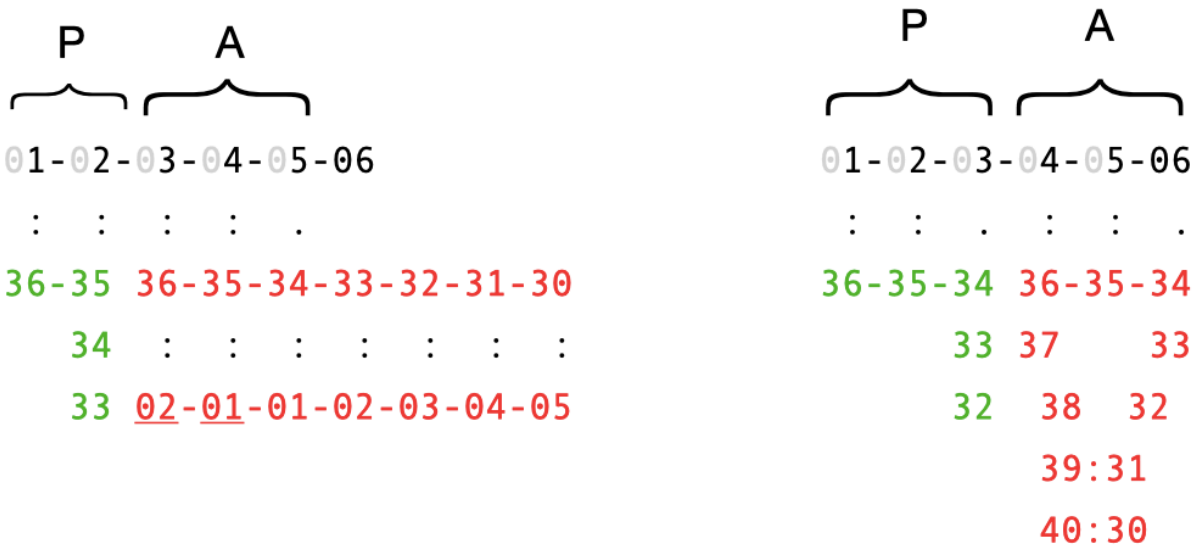


FIGURE 13-6. dRNA dRNA VERSUS dRNA tRNA DECODING

At the outset of tagging, the duplisome A-site had to accommodate either a duplicon-dRNA, or a primitive aminoacyl-tRNA, for polynucleotide elongation and termination, respectively. The ribosome A-site, of course, decodes and accommodates modern aminoacyl-tRNAs, but is this plausible for a duplisome evolved to decode

duplicon-dRNAs? The primitive tRNA was a partial and imperfect mimic of an open dRNA. At the duplisome decoding center, its anticodon arm could pair with the A-site codon, allowing the transfer center to accommodate its acceptor arm with an amino acid esterifed to the 2' or 3' OH. With no duplicon to displace, the codon engaged the anticodon loop directly, if indiscriminately. Relying on the thermodynamic stability of the codon-anticodon helix, tRNA decoding was decidedly inferior to dRNA decoding based on kinetic competition of codon and duplicon for the anticodon. With no hairpin to open, the primitive tRNA acceptor arm accommodated quickly, so that any near match, dominated by the second codon position, sufficed for amino acid tagging. One immediate difference in decoding the primitive tRNA, unimportant at the time, but consequential thereafter, was that the anticodon trailer, notably nucleotide 37, resisted *dRNA tRNA scrunch* on the template (Figure 13-6). The precise choice of codon frame (scrunched or unscrunched) was no doubt sloppy, and likely unimportant, in the original tagging duplisome.

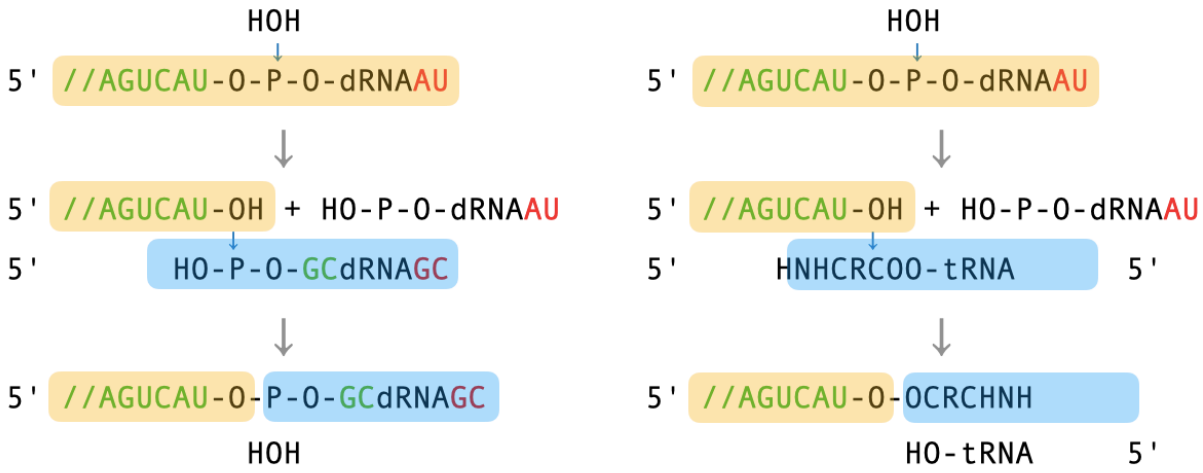


FIGURE 13-7. POLYNUCLEOTIDE ELONGATION VERSUS TERMINATION BY AMINO ACID TAGGING

Once the tRNA anticodon arm had paired with the A-site codon, the charged acceptor arm was accommodated in the nucleotidyl transfer center for amino acid tagging. As in polynucleotidyl transfer or hydrolytic release, the first step in amino acid

tagging is hydrolysis of the polynucleotidyl-dRNA in the P-site. But now the A-site is occupied by an aminoacyl-tRNA, not duplicon-dRNA, so that the second step, nucleophilic attack by the 3' OH, results in aminoacylation, or terminal tagging (Figure 13-7). The 3' tagged polynucleotide exits through the nascent chain tunnel, while the side-products, free dRNA and free tRNA in the P-site and A-site, respectively, are released upon duplisome recycling. [energetics of tagging?] There are precedents for ribosome catalyzed trans-acylation. Thus, ribosomes synthesize a polyester-tRNA in the P-site through iterative trans-acylation to α-hydroxy acyl-tRNAs entering the A-site (Fahnestock & Rich 1971). Similarly, ethanol and other small organic alcohols can attack peptidyl-tRNA in the P-site to release the nascent polypeptidyl-ester (Caskey et al 1971).

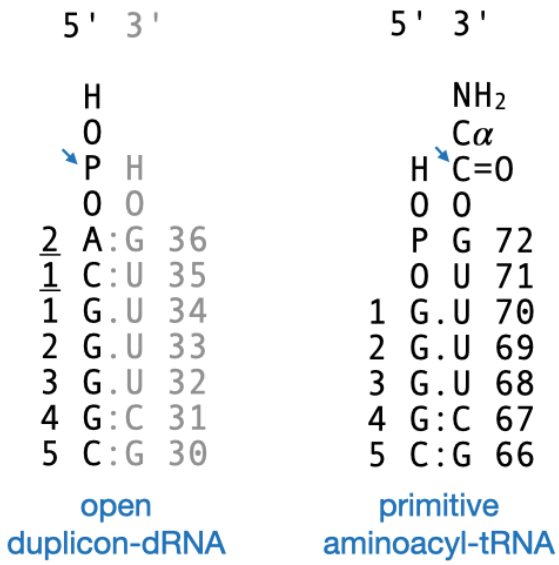


FIGURE 13-8. DUPLICON-dRNA VS PRIMITIVE AMINOACYL-tRNA ACCEPTOR ARMS IN THE A-SITE OF THE DUPLISOME TRANSFER CENTER

In Figure 13-8 small blue arrows point to the electrophile atoms, phosphoryl phosphorus and carbonyl carbon, in polynucleotide elongation and amino acid tagging, respectively. These are on opposite strands of the acceptor stems, but otherwise are remarkably close in position, judged by the homologous 5' ends of both adaptors. Molecular modeling might clarify the relative position and alignment of nucleophile 3' OH in the transfer center for these two alternative substrates of the tagging duplisome. The cast of the 3' trailer ending with the anticodon (N34 N35 N36), which has opened

away during decoding, is shown (light grey) for comparison to the 3' end of the tRNA. The nucleotides of the primitive tRNA has been renumbered 1-72. For simplicity, only the bridging oxygens are shown on the phosphate.

Once there were two or more tRNAs and charging ribozymes, favoring distinct sets of amino acids, the uses of template-programmed tagging multiplied. For sake of discussion, we conjecture that the mature tagging code at the breakout of polypeptide translation had anywhere from two to four tRNAs, and the same number of cognate charging ribozymes. From the first tag-and-release code with just one tRNA, to the mature tagging code with several tRNAs, and all subsequent refinements of the protein code, the evolving binary relation from codons to amino acids was determined by the tRNA charging complexes and the decoding center of the ribosome *née* duplisome. Comparing the universal protein code today to early amino acids suggests one intermediate code had just four tRNAs whose pairing was dominated by the second codon position, charged perhaps with small (NGN), medium (NCN), large apolar (NUN) and large acidic residues (NAN), respectively (Higgs 2009).[21] This four tRNA code was likely anticipated by an even simpler two tRNA code, discriminating say purines and pyrimidines at the second position, and accepting say polar and apolar amino acids, respectively.

To exploit amino acid tagging, and avert conflicts with RNA duplication, both tRNAs and duplisome centers meanwhile underwent a number of changes in no particular order. Any tight squeeze in aminoacyl-tRNA accommodation was relaxed, without compromising accommodation of duplicon-dRNAs. Additions of more tRNAs and charging ribozymes to the original one tRNA tag-and-release code, selected for improvements in tRNA decoding, without compromising the fidelity of dRNA decoding. Thus, by the breakout of polypeptide translation, the tagging duplisome was well adapted to accept either duplicon-dRNA or aminoacyl-tRNA in the A-site, for polynucleotide elongation and termination, respectively.

---

[21] Some authors identify those coded amino acids as G A V D, respectively, although the first charging enzymes, more so the ribozymes that preceded them, were no doubt promiscuous (Weber & Miller 1981; Eigen & ### 1982; Trifonov 2004).

If this sounds like a mouthful, we suggest simple defaults in RNA duplication appear baroque if we use the evolved features of the modern ribosome as our starting point. Thus, features that apparently defend the reading frame of protein translation against scrunch were rudimentary or entirely absent in RNA duplication. There was no enzymatic RNA modification, in particular nucleotide 34 that might defend the P-site wobble pair, nor modifications of rRNA forming the wobble seal were likely absent. There was no nucleotide 37 on the duplicon-dRNA for cross-strand stacking, nor indeed the 3' half of the modern tRNA. Finally, there were no proteins, in particular no translocational GTPase, … no ribosomal proteins (S9 gripe of peptidyl-tRNA in P-site). Similarly, the distribution of tRNA superwobbling found today in protein translation of organelles and mycoplasmal bacteria are derived characters. Here we suggest that dRNA superwobbling was the primitive default in RNA duplication. Finally, we suggest a translocon of 2 nucleotides conforms to the default module of scrunched dRNAs and template.

The duplisome hypothesis explains the curious maturation of pre-tRNAs by RNase P as a vestige of dRNA loading. Unlike their crucial 3' end, hitherto there seemed nothing special about tRNA 5' ends to warrant precise processing by an ancient ribozyme. Indeed as first reported for the initiator tRNA-fMet of the archaea *H volcanii*, the mature tRNA is triphosphorylated at its 5' end, indicating it is the start of transcription of a leaderless pre-tRNA and not a product of RNase P cleavage (Gupta 1984). Presumably, it has been easier to retain RNase P for processing other pre-tRNAs than to move all their starts of transcription so precisely. Since the radiation of LUCA, RNA P has acquired new protein partners and new RNA substrates, and been replaced in some lineages by non-orthologous, all protein RNase P.

Whereas RNA duplication used 16 different dRNAs, and amino acid tagging began with just one tRNA, we conjecture that polypeptide translation began with a mature tagging code of at least two, but likely no more than four tRNAs. The thermodynamics of codon-anticodon pairing, dominated by the second position, was likely adequate for an amino acid code with four columns.

In the ribosome decoding center … h44 …

Changes in the decoding center, notably h44 allowed faithful base-pairing at the first two codon positions, amino acid code of 16 quartets, rivaling the fidelity of the RNA code. As discussed in section 10, h44 may trace to dRNA decoding, or only after the breakout of polypeptide translation. IDEA h44/h18 refinement before EG-Tu and proofreading … not clear whether 4 codon boxes are split down yet

[decoding crisis]

principal interaction between codon and anticodon

monitored by rRNA DCC 'initial codon selection'

if 36 35 are WC and 34 is GU wobble or modified 34 pair then latch

IDEA raw breakout based on energetics of codon-anticodon pair; has lost kinetic filter of duplicon; second position more important than first (read 2 or read 4)

ADD rRNA helix 44 latch and up the difficulty of accommodation (recover read 16)

E-site

mRNA -3 -2 -1

tRNA 36 35 34

rRNA G963 uS7 G926

P-site

mRNA 1 2 3

tRNA 36 35 34

rRNA A790 C1400

A-site

mRNA 4 5 6

tRNA 36 35 34

rRNA A1913 A1493 A1492 G530 C1054

A1492 A1493 h44

G530 h18

codon	1	2	3
h44/h18	93	92	
anti	36	35	34
duplicon	2	1	1

JARGON the duplicon left a molecular cast filled by the decoding center?

wrong side? wrong orientation?

We suggest that the trimolecular decoding in protein translation (codon-anticodon-decoder) arose from the simpler bimolecular decoding in RNA duplication (codon-anticodon-duplicon). **Molecular modeling might reveal** whether both decoding mechanisms were compatible in the same decoding center, or only a dedicated ribosomal decoding center. the decoding helices h44/h18 were compatible with continued toehold decoding of duplicon-dRNAs, or occupy the molecular cast of the duplicon?

ribosome decoding center

first codon-anticodon basepair monitored & locked by A1493

second codon-anticodon basepair monitored & locked by A1492 & G530

all three SSU rRNA nucleobases are universally conserved

[Khade et al 2013] all of the hydrogen bonds from rRNA decoding center to codon-anticodon duplex are sequence independent/balanced

obvious for 2'-OH groups

true for contacts with nucleobases of mRNA

2'-OH of A1493 bonds O2 if pyrimidine and N3 if purine

() A-minor interactions in the decoding center to make first and second positions reliable again , recover the reading fidelity sacrificed

Thus, filling the spatial and functional void of the duplicon, the oppositely-oriented strand of SSU rRNA (A1492 A1493) monitors the minor groove of the codon-anticodon helix for correct pairing (Ogle).

1st \ 2nd	U	A
U	UUA tRNAUAA apolar	UAA tRNAUUA empty
A	AUA tRNAUAU blocked	AAA tRNAUUU polar

14. Breakout of polypeptide translation

The duplisome hypothesis divides polymer life into seven eras punctuated by major evolutionary transitions: (1) An era of spontaneous RNA copying leading to the first ribozymes for RNA recombination and repair. (2) An era of processive RNA copying mediated by the duplisome and its dRNAs. (3) An era of programmed polynucleotide termination and amino acid tagging using primitive tRNAs and charging ribozymes. (4) An era of polypeptide translation with little or no assistance from coded polypeptides. (5) An era of protein translation incorporating coded proteins and enzymes in all aspects of bioenergetics and metabolism, including RNA and protein synthesis. Protein life and its NTP currency opened up a new era of polymerase enzymes that heralded (6) the retirement of the duplisome, and (7) the takeover of DNA life.

At the breakout of polypeptide translation in the late RNA world, there was already a brisk market for random polypeptides of simple compositions, including intrinsically disordered sequences that assembled into polynucleotide-polypeptide coercevates, discrete RNPS, or lipid membrane pores. Useful polypeptides were subject to depletion, creating pent-up demand for any mechanism that better supplied them. Coming at the heyday of amino acid tagging and exploiting random polypeptides, the first coded polypeptides did not need any great variety of residues, nor precise control of sequence, to be useful. Thus, a rudimentary code of just two or three classes of amino acids could supply current demands, as well as hock new wares. Programmed polypeptides of different size, composition and sequence, were made concurrently from the common pool of amino acids using different mRNAs, or one mRNA with different starts and reading frames. Among their virtues, the first coded polypeptides likely had greater, more reproducible length, regular repeats, and distinct domains marked by abrupt transitions from one composition or sequence to another.

Given the tagging duplisome with 16 dRNAs, one empty tRNA for simple termination, and two or more charged tRNAs for tagged termination, we propose that the polynucleotide elongation cycle was co-opted for polypeptide elongation. The mechanisms of tRNA charging and decoding carried over unchanged from amino acid tagging for polynucleotide termination, *viz*. the polynucleotide tagging code became

the first polypeptide translation code. Contrasting the continuity of tRNA decoding, polypeptide translation required an abrupt, if catalytically modest, change in the polymer transfer reaction and energetics of elongation (Table 9-). In Figure 14-1 we compare polymer transfer for amino acid tagging and peptide bond formation. Both reactions require an aminoacyl-tRNA in the A-site, but amino acid tagging requires a nucleotidyl-dRNA in the P-site, while peptide bond formation requires a peptidyl-tRNA in that site. Only small changes in substrate positioning were needed in the center evolved for nucleotidyl transfer and polynucleotide release, to catalyze peptidyl transfer and polypeptide release.

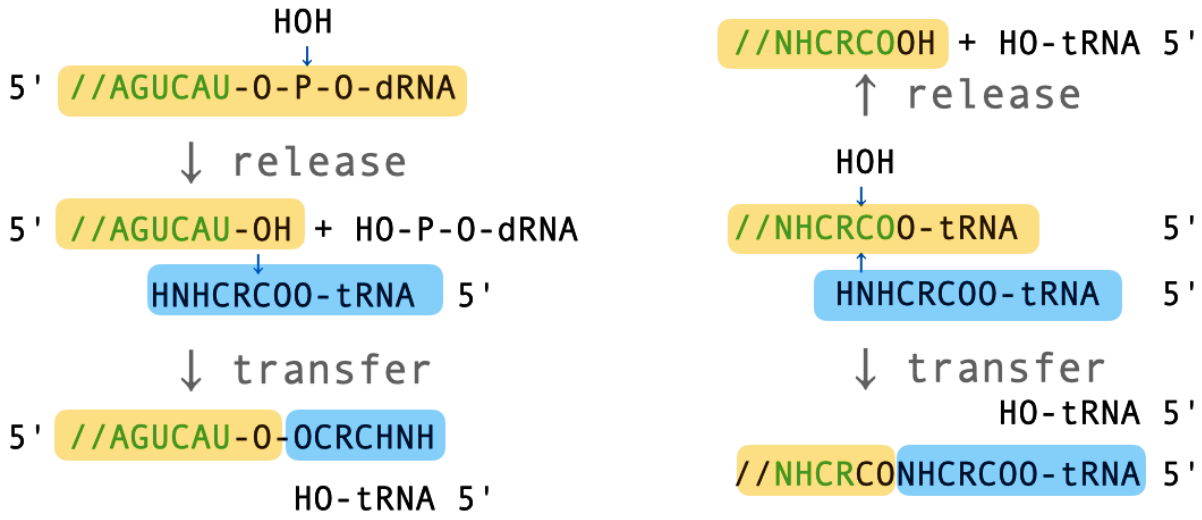


FIGURE 14-1. POLYNUCLEOTIDE RELEASE WITH AMINO ACID TAGGING (LEFT) POLYPEPTIDE RELEASE OR PEPTIDE BOND FORMATION (RIGHT)

Perhaps the simplest model for initiation of polypeptide translation comes from abortive tagging, *viz*. occasional failure of the second step in polynucleotide termination. To wit, the polynucleotidyl-dRNA in the P-site undergoes hydrolysis, the usual first step in termination, but simply releases the polynucleotide 3' OH through the polymer exit tunnel without attacking the aminoacyl-tRNA in the A-site. *Indeed this was the normal mode of polynucleotide termination and release before the invention of tagging.*

After an abortive tagging, translocation driven by closing and exit of the freed dRNA, brings the aminoacyl-tRNA into the P-site, allowing a new aminoacyl-tRNA to enter the A-site. Now the stage is set for the first peptide bond with an aminoacyl-tRNA in the P- as well as A-site, allowing the nucleophile (α-amine nitrogen) of one to attack the electrophile (carbonyl carbon) of the other (Table 9-). In amino acid tagging the A-site aminoacyl-tRNA supplies the electrophile, while in peptide bond formation it supplies the nucleophile. Conceivably, some subtlety of relative position, or else, frank modification of the initiating residue, analogous to formylation of methionine in the initiator tRNA in the P-site in bacteria and mitochondria, ensured that the A-site α -amine formed the first peptide bond. Whatever caused of the role reversal of nucleophile and electrophile in forming the initial peptide bond, once substrate symmetry was broken, so that the P-site tRNA carried say a mono- or dipeptide in the proximal exit tunnel, the direction of transfer was sterically constrained for all further elongation.

We conjectured that translocation in polynucleotide elongation was driven by dRNAs acting as thermal motors, particularly cold closing of freed dRNA in the P-site after nucleotidyl transfer. Locked in the open conformation, L-shaped tRNAs were weak thermal motors at best. Although work from hairpin closing was not available from deacyl tRNAs in the P-site after peptidyl transfer, a new source of chemical free energy was available to drive polypeptide elongation, *viz*. the greater stability of the peptidyl-amide than peptidyl-ester bond (Table 9-; Krayevsky & Kukhanova 1979; Leung et al 2011). Decoding and accommodation of open tRNAs, as well as peptidyl transfer itself, were relatively fast, so once a polypeptide was initiated, multiple peptide bonds could be made isothermally in a short while in comparison to the diurnal hot-cold cycle of polynucleotide elongation. If so, the breakout ribosome née tagging duplisome had two modes of polymer elongation, slow and careful RNA duplication using duplicon-dRNAs, or quick and dirty polypeptide translation using aminoacyl-tRNAs.

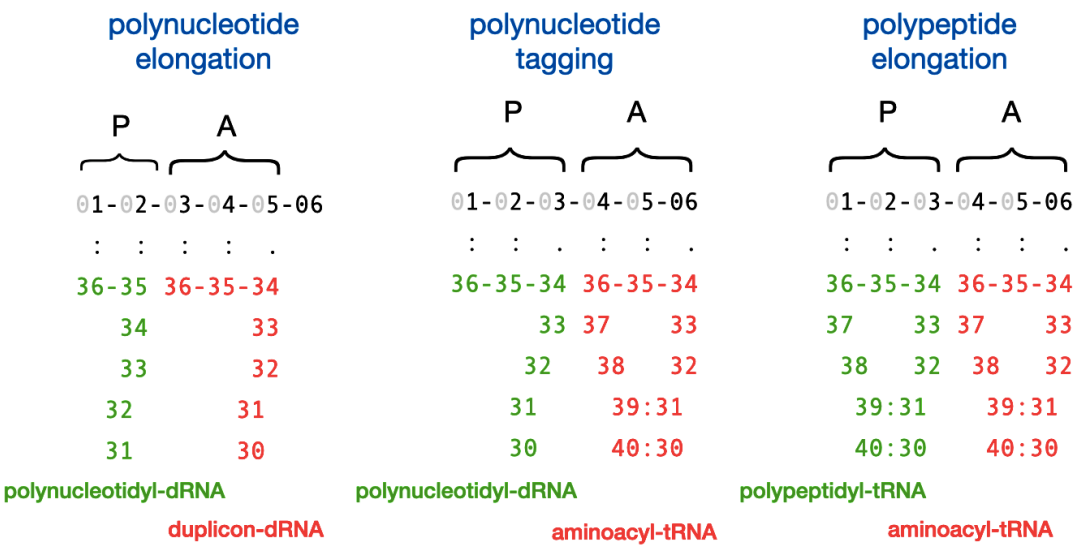


FIGURE 14-2. FROM RNA DUPLICATION TO AMINO ACID TAGGING TO PROTEIN TRANSLATION

Beyond the polymer transfer reaction itself, the most striking change from polynucleotide to polypeptide elongation is the increased net movement of the template RNA, from two to three nucleotides in each elongation cycle.[22] We suggest the larger translocon size arose from +1 shift in the A-site reading frame when the elongator dRNA was replaced with the terminator tRNA (Figure 14-2). Much as the non-scrunch reading frame for programmed polynucleotide termination was likely sloppy at first, the translocon size was likely sloppy at the breakout of polypeptide translation. The advantages on non-overlapping triplet codons for splitting quartet boxes emerged only later, after improvements in reading frame defense during decoding and translocation.

[22] With co-transcriptional translation and polyribosomes on large mRNAs, we may go back and forth between regarding the ribosome as motoring along a stationary template, or a stationary ribosome pumping a template through its channel. MOVE WHERE?

In the parlance of synthetic biology, RNA duplication and protein translation were orthogonal functions throughout the long *conviviencia* of the duplisome and its dRNAs, with the ribosome and its tRNAs. The first tasks of protein life were to improve the speed, accuracy, and versatility of polypeptide translation without interference with the well-oiled machinery of RNA duplication, or costly redundancy. A small number of changes, in no necessary order, could improve polypeptide translation on the breakout ribosome: (1) Changes in the P-site of the large subunit rRNA to better accommodate tRNAs. (2) Changes within the polymer transferase center to position and orient the A-site tRNA 3' , the P-site tRNA 3', or both for peptidyl transfer. (3)Changes in the nascent polymer exit tunnel to better accommodate polypeptides, and facilitate peptide bond formation, without compromising the exit of nascent polynucleotides during RNA duplication.

There was no essential difference between the tagging duplisome and the breakout ribosome, only between dRNAs and primitive tRNAs. From the breakout of polypeptide translation through to the invention of the RdRP enzyme, cells performed RNA duplication and protein translation side-by-side without mutual interference. Improvements in the speed and fidelity of polypeptide translation required significant changes in large and small subunit rRNAs, all without compromising RNA duplication. Ability to discriminate between duplicon-dRNA and aminoacyl-tRNA at the decoding center, catalyze the correct transfer at the polymer transfer center, and … nascent chain through the exit tunnel. Presumably all of this could be accomplished by duplication of the rRNAs and specialization apart.

There is a continuum of possibilities for the orthogonal evolution of RNA duplication and polypeptide translation. At one extreme, a dedicated ribosome with novel large and small subunits might perform polypeptide translation exclusively. At the other extreme, some novel ribosome component, or RNA modification might toggle the core duplisome from its default mode of RNA duplication to the new mode of polypeptide translation. This might make an irreversible commitment to translation during ribosome biogenesis, or bind or modify the core duplisome reversibly to toggle the elongation cycle from RNA duplication to polypeptide translation, and back. Obvious possibilities for regulating ribosome biogenesis or the elongation cycle are a free-standing

riboswitch, or perhaps a ribosomal protein. There are now several examples from synthetic biology of managing conflicts between orthogonal translation systems using tethered or stapled ribosomes (Orelle et al 2015; Schmied et al 2018; Carlson et al 2019; Aleksashin et al 2020).

To encourage discussion, we conjecture that 5S rRNA (~120 nt) evolved in early protein life to toggle biogenesis of the polymer transfer center and exit tunnel from polynucleotide to polypeptide elongation. In the mature ribosome large subunit, the 5S RNP forms part of the central protuberance just above the peptidyl transfer center, between the L1 stalk and factor binding site. Two observations suggest that 5S rRNA has no catalytic role in the mature ribosome. First, antibiotics that bind large subunit domains D2 and D5 rescue peptide bond formation in the absence of 5S rRNA (Khaitovich & Mankin 1999). To work, these drugs must be present during biogenesis, but can then be washed out with ethanol without abolishing activity of the folded subunit. Second, 5S rRNA has been lost entirely in evolution of mitochondrial ribosomes (Koripella et al 2020).

Two other observations suggest that 5S rRNA has a regulatory role in ribosome biogenesis. First, 5S RNP docks midway during large subunit biogenesis and acts as a wrench-like chaperone for domain D5 (Zhou et al 2019; Micic et al 2020). At the nucleolar stress response checkpoint in eukarya, accumulation of 5S RNP due to problems in large subunit maturation inhibits E3 ubiquitin ligase MDM2 allowing p53 mediated cell cycle arrest and apoptosis (Bohnsack & Bohnsack 2019). Second, in engineered ribosomes with circularly permuted 5S rRNA fused to 23S rRNA, the large subunit adopts either a normal fold that supports protein translation, or an alternative fold that shifts H89 and trailer (nucleotides 2490-2505) as much as 30 A (Huang et al 2020). In this latter fold, which does not support translation, pairing H80 P-loop rG2251 G2252 G2253 with C2498 C2499 U2500, and thus precluding its normal pairing with the terminal 3’ CCA of P-site tRNA to position the carbonyl carbon for peptidyl transfer.

WHERE??? QUESTION dtRNA versus tmRNA

Refinement of the protein code

Both the immediate advantages of coded polypeptides, and emergent advantages of self-folding protein domains and their catalytic centers, selected for a greater variety of amino acids and greater control of mRNA translation. At the heart of these changes in protein translation and refinement of the genetic code, the radiation and specialization of tRNAs and their charging complexes, while maintaining interactions with universal machinery of translation. … On one hand, specialization of the tRNA acceptor arms and charging complexes enlarged the palette of proteingenic amino acids even as tRNA charging ribozymes/enzymes became more selective about their set of interchangeable amino acids. On the other hand, changes in tRNA anticodon arms and the ribosome decoding center. *Pari passu* with new tRNAs, modifications of their anticodon arm, and changes in the ribosomal decoding center, restored quartet decoding, sacrificed in amino acid tagging, and later, split quartet codon boxes into duets, and even singlets. Finally, changes in the tRNAs and ribosome increased the stability of the tRNA tRNA :: mRNA minihelix to defend the reading frame during translocation.

Splitting quartets into duets required 5-methylation or other modification of U34 to suppress superwobble. In modern cells, these RNA modifications require enzymes. …. reading frame defense ….

New tRNAs

Pari passu with the changes in tRNA anticodon arms and ribosome decoding center to split codon boxes from 16 to 8 to 4 to 2 to 1, as well as improve the fidelity of decoding, … split codons to specify hitherto interchangeable amino acids, as well as accommodate entirely new ones. () at the same refinement in tRNA charging culled out some non-proteinogenic aas

WILD Question was ribozyme P related to the DCC?

PTC nearly blunt RNA ligase

DCC/PTC 2 nt 3’ overhang RNA ligase

IDEA DCC lost ligase activity, taken over by PTC

more tRNAs and cognate charging ribozymes, or enzymes

refinement of codon blocks

- parcellation between existing proteogenic amino acids
    - needs to refine the charging and the matching
- squeeze current aa to admit a new one
- w/o changing the matching, refine the charging or reprogram the tRNA

reassignment of existing amino acids

mitochondria UGA stop > W

trash releasing factor, change anticodon from CCA > UCA

CCA reads only UGG by WC

UCA reads UGG by wobble as well as UGA by WC

mitochondria? AUA I > M

standard anticodon K2CAU pairs with A only hence AUA

UAU pairs WC AUA and wobble AUG

() greater stability of reading frame for longer, unique folds

The effect of frame shift during elongation is very different for RNA duplication and polypeptide translation. The impact of -1 or +1 frameshift on nascent polynucleotides is strictly local. Although different sets of loaded dRNAs are utilized in the original (even) and shifted (odd) frames, the RNA duplicate produced is identical to the template in sequence, excepting the 1 nucleotide indel at the site of shifting. The impact of -1 or +1 frameshift on protein translation is generally catastrophic, as the downstream sequences are unrelated to that coded in the 0 frame, and usually terminate prematurely.

For simple polypeptide repeats, it is possible that frameshift products were useful repeats themselves, or that random shifting between frames produced a spectrum of polypeptides of useful compositions and sequence. The full advantages of faithful maintenance of reading frame appeared *pari passu* with refinement of tRNA charging and decoding. With a half dozen or more choices, and ORFs of a couple dozen or more residues, the space of polypeptide sequences to explore … (ref ancient alphabet) Predict that piecemeal refinement with discrimination between hitherto interchangeable residues, culling of previously used residues, addition of new feedstocks. As the lengths got longer, and the positional constraints on sidechains at key positions became greater, the rewards drove biosynthesis and charging, refinement and fidelity of decoding, and defense of reading frame.

A frameshift error during polypeptide translation might spoil the nascent protein, but a frameshift error during RNA copying spoiled any protein made from that mRNA. Thus, the improvements in protein translation, and the increasing cost of ribosome frameshift errors, selected for improved frame preservation in RNA copying, and later, DNA replication. RNA templated RNA polymerase faithful in nucleotide and in frame.

UNSCRUNCH IDEA the shift from 2 nucleotide to 3 nucleotide translocation predates all unique protein folds extant today …. The immediate advantage was likely the greater stability of 3 + 3 module in translocation … other changes for better defense of reading frame… … Any advantages from splitting hitherto interchangeable residues or adding entirely new ones were realized more gradually. It is unknown whether any extant protein folds evolved from a pre-modern genetic code with fewer tRNAs and perhaps more promiscuous aaRSs, but it seems unlikely that any unique folds survived the transition from scrunch to rigid.

mutations in tRNA C74 act as frameshift suppressors (Green et al 1998) .. C74 pairs with LSU rRNA G2252 which also causes frameshifts (Gregory et al 1994)

QUESTION does LSU play a role in frameshift?

IDEA the reading frame first defined during initiation must be preserved during elongation, first during decoding at the A-site and then translocation

RNA-based reading frame defenses

Can we explain this rudimentary protein code requiring the abrupt appearance of tRNAs and their (now extinct) charging ribozymes? And, what of its immediate crisises in the mechanism of decoding, the chemistry of polymer transfer, the mechanism of translocation, and the energetics of elongation? This major evolutionary transistion is implausible if we assume that the breakout ribosome was all round as good, if not better, than the well-oiled duplisome. Rather, as discussed below, polypeptide translation was quick and dirty from the start compared to RNA duplication, with a great sacrifice of the fidelity of decoding, and the processivity of elongation. It had one brilliant new trick, programmable control over polypeptide sequence. Over time, the advantages of improved fidelity, control of initiation and preservation of reading frame, refinement of the genetic code to split codon boxes, the charging complexes split apart hitherto interchangeable amino acids, introduce useful new amino acids, and cull hitherto acceptable ones. Finally, no matter how useful the new technology of coded polypeptides, early protein life relied on continued RNA duplication.

what was the order of improvements of decoding? EF-Tu before RNA polymerase?

much less anticipate the fruits of exploring protein folds based on precise sequence and novel catalytic side-chains.

(3) amino acids important as metabolic intermediates, cofactors , or … histidine, cysteine, methionine … implicated in liquid-liquid phase separation compartment R W amino acids or polypeptides attached to RNA

The biogenesis of the ribosome large subunit in eukaryotes is conventionally staged by intracellular location: The solvent-exposed outer shell and nascent polypeptide exit tunnel fold within the nucleolus. Next the central protuberance and initial inter-subunit surface fold within the nucleoplasm. And finally, the PTC and final inter-subunit surface fold within the cytoplasm.

IDEA PTC is not toggled until after CP formed

L1 stalk = H75 H76 H77 H78

The question of when any two evolutionary innovations arose in a lineage has a clear and simple answers in phylogenies of strictly vertical gene transmission. But real species undergo mixis, or horizontal gene transmission, as well, with frequent introductions of genes from near relatives (population genetics, fixation, sweeps) and less frequent introductions from distant ones. Besides introducing new genes, may replace existing ones in NOGDs. outside the immediate lineage, in that makes sense in strictly VGT … invention/fixation. The radiation of LUCA into cellular kingdoms of bacteria and archaea began with many rival innovations unsorted, and continued with robust HGT across the nascent cellular kingdoms. In particular, the idea that the universal protein code was settled before any radiation of LUCA has been questioned. In the short run, there was strong selection for divergence of codes, or at least codon frequencies, to defend against the dangers of MGEs. In the long run, there was strong selection for convergence of codes to explore the opportunities of the advantages of universal protein code as the *lingua franca* of HGT .. successful those lineages than could interpret the universal genetic code … dangers of HGT s … selection for divergence of code to defend against MGE, advantageous of HGE converged on the universal protein code despite considerable different strategies of adding new anticodons or new modifications …. (Grosjean et al 2010)

tRNA biogenesis & evolution

The primitive tRNA underwent a number of structural changes along the evolutionary path from programmed polynucleotide termination and amino acid tagging in late RNA life to protein translation. Some were one-off heritable changes, while others were biochemical modifications mediated by ribozymes or enzymes of tRNA biogenesis and repair. *Pari passu* with the changes in tRNA structure, there were corresponding changes in the ribozymes/enzymes of tRNA charging, as well as ribosomal centers of decoding and peptidyl transfer. Meanwhile, the Noah's arc of

tRNAs radiated from just one terminator tRNA to over twenty charged isotypes and forty isoacceptors. For each innovation, we seek a parsimonious explanation of when, how, and why that change arose that is mutually consistent with our explanations of other features. Here we sketch a likely partial ordering of four key events: (1) As suggested in section 13, specialization of the D and T loops to form the elbow that stabilizes the constitutively open (L-shaped) tertiary structure likely occurred in the first tRNA for programmed polynucleotide termination. (2) The formation of the variable loop and the splicing of tRNA halves likely occurred during tRNA radiation. (3) Various RNA modifications to stabilize the fold, split codons, or defend reading frame … (4) Finally, the addition of the universal 3' CCA tail likely occurred in late protein life.

(Kuhsel et al 1990; Xu et al 1990) cyanobacteria and chloroplasts have group I intron in same position of leucyl tRNA-UAA anticodon back to back in Science 250 … after N34?

Reinhold-Hurek B & Shub DA (1992). Self-spicing introns in tRNA genes of widely divergent bacteria. Nature 357, 173-176.

Agrobacterium tumefaceins tRNA Arg-CCU (intron after N36)

Azoarcus sp. tRNA Ile-CAU (intron after N36)

[MOVE DOWN section 14] A Noah's arc of bacterial tRNAs, one of each of the 43 isoacceptors would comprise over 3000 nucleotides. In either case, as much or more RNA mass in the large and small subunit RNAs of the duplisome and ribosome, respectively. Might expect that multiple sets of adaptor RNAs per duplisome or ribosome, but surprisingly, …

E coli isotypes [20?] isoacceptors [43] isodecoders [15]

I'm guessing they mean 43 + 15 = 58 tRNA genes, and don't count isodecoder classes with just one representative as isodecoders.

The first step of tRNA maturation, after removing any introns, is trimming the 5' leader between -1/+1 by RNase P, followed by trimming the 3' trailer between 73/74 by

tRNase Z, and then addition of the universal nucleotides C74 C75 A76 by the CTP/ATP nucleotidyltransferase.

Before retiring **RNA direct duplication**, some other means was needed for forming the tandem dRNAs that make tRNAs. This new mechanism of joining dRNAs was likely the self-splicing group I intron found today in some bacterial and organellar tRNA genes between nucleotides 37 and 38. Likely vestiges of this group I intron, non-autonomous introns found in this same position in archyaeal and nuclear tRNA genes are removed by tRNA splicing endonuclease (TSEN) followed by exon ligation.

Just when tRNA splicing emerged in protein life is uncertain, only that it was needed before the retirement of RNA direct duplication. [MOVE section 20 all 16 dRNAs form eight types of ribbons] If group I self-splicing introns afforded a means not only to preserve the supply of tandem dRNAs, but to create and control new combinations of halves. Whether ribozymatic or enzymatic charging, possible advantages for elaboration of the genetic code of combining tRNA halves not just tandem dRNAs. [splitting of the 4-box codons]. contribution to early protein life.

may be the ancestral function of group I introns …

polyphyletic or mix-and-match evolution of tRNAs from tRNA halves … split genes in archaea today … redundacy of the genetics in that isoacceptor family has multiple isodecoders ….

The human genome still harbor remnants of the original tRNA halves (~140) that lead to the origin of tRNAs (Zuo et al 2013). peak length of about 38 nt…correspond to 5’ or 3’ half …. tRNA processed fragments, tRNA splicing …. regulate translation ….

[from breakout to LUCA]

We focus our discussion of late protein life, and then DNA life, on a few curious problems that perhaps depend on whether late RNA life used the duplisome, or an extinct polymerase ribozyme for replication, or more pointedly, whether ribosome and tRNAs are exaptations of duplisome and dRNAs, or *deus ex machina*. To frame these curiosities, we mention in passing some major events that occur along any possible path from early protein life to late DNA life.

Doubtless the first mRNAs encoded proteins and enzymes to improve the speed and fidelity of translation, enlarge the set of proteinogenic amino acids, and direct the export of proteins.

[ribosomal proteins][RNA modification enzymes]

Remarkable modifications of rRNAs and especially tRNAs and their anticodon arm … to refine and perfect … suggestion that in some cases these modifications may suppress or mask once desirable functions …not just side-reactions …

Polypeptide initiation

The initiation and termination factors of protein translation tuned after bacteria/archaea split 3 billion years ago.

initiation is rate-limiting requires several seconds in bacteria, elongation 20 per sec

[internal termination]

[internal initiation]

IDEA MOVE UP the polynucleotidyl-aminoacyl ester is not released through the exit tunnel and a new aminoacyl-tRNA enters the A-site. amine attacks the ester bond at the P-site, forming a peptide bond and freeing the polynucleotide … rather than tagged product, freed product and continued polypeptide … switch from dRNA elongation to tRNA elongation … latter true initiation by loading an aminoacyl-tRNA in the P-site …

Like RNA duplication, there were special problems, as well as opportunities for gene regulation, associated with the initiation and termination of polypeptide translation. Unlike simple polypeptide repeats of indefinite length, for self-folding protein domains, it is crucial that the entire sequence is present, and oftentimes, that there is no excess N-terminal leader or C-terminal trailer. Thus, there was strong selection for precise control of polypeptide initiation and termination, as well as maintenance of the reading frame, to encode longer unique sequences, culminating in the self-folding domains with catalytic centers of modern enzymes.

In modern cells, initiation of protein translation is a key event in the regulation of gene expression. The initiation codon sets the reading frame as well as the polypeptide start.

At the breakout of polypeptide translation, the ribosome likely had little or no control over the site of initiation and termination. On the breakout ribosome née tagging duplisome, the question of whether to translate or to duplicate likely had precedence over just where to start translation. Moreover, the reading frame of polypeptide translation was more important than the position of the initiation site along the mRNA. the evolution of dedicated initiator tRNAs and start codons allowed programmed initiation and reinitiation at internal sites in the mRNA

?problem/opportunity if the RBS is folded …stand-by model (de Smit van Duin 2003)

IRES

There are striking departures from factor-assisted initiation sketched above, revealed by viral mRNAs or cellular stress. Found in all cellular kingdoms, as well as mitochondria, leaderless mRNAs require only the start codon and initiator tRNA, but not upstream elements such as the Shine-Dalgarno box (prokarya), or the 5' cap (eukarya). In some cases, leaderless mRNAs can initiate on undissociated ribosomes in absence of protein initiation factors. Although this mechanism is formally simpler than factor-dependent initiation, it is unknown whether it is a primitive mechanism for initiation, or a later adaptation for cellular stress.

We suggest that leaderless initiation carried over from RNA duplication to polypeptide translation *before any support of coded initiation factors*. Likely any codon could initiate polypeptide translation from its cognate aminoacyl-tRNA at the P-site. ceribus paribus the three nucleotides of the mRNA would be used as in leaderless mRNAs today…

After protein termination on polycistronic mRNAs, the 70S ribosome can scan the mRNA for the next Shine-Dalgarno box with downstream start codon *without subunit dissociation*. [how does the new initiator tRNA enter P-site?] The ability of intact ribosomes to initiate on leaderless mRNAs, and to reinitiate on a downstream ORF in polycistronic mRNAs, raises the question of whether the early ribosome, and its duplisome ancestor, ever needed to dissociate into large and small subunits. Today … subunit assembly and disassembly allows swapping mRNAs with long 5' or 3' UTRs in and out of the small subunit without scanning, much less translating, from the beginning, or to the end. A mobile interface between the two subunits appears central to the mechanics of translocation, but must they dissociate completely between mRNAs?

LOOKUP gene order, ITSs and processing sequence

QUESTION could the 3' end of small subunit rRNA have once been continuous with 5' end of large subunit rRNA

[MOVE TO DUMP] bacteria leaderless mRNAs that begin precisely at AUG? (thus lack A/U-rich sequence for S1 ribosomal protein and SD consensus AGGAGG about 7-10 nt upstream of start codon … SD pairs with anti-SD at 3’ end of SSU rRNA

Polypeptide Termination & Release

[TERMINATION] Ribosomes seldom translate mRNAs to the end, but terminate when they encounter the first stop codon, releasing the nascent protein, and then are recycled to initiate translation on a new mRNA. Release factors in the A-site recognize stop codons, open and trigger hydrolysis of the peptidyl-tRNA in the P-site. Ribosomes stalled on aberrant mRNAs, or reaching the end without encountering a stop codon, are recognized and rescued by diverse quality control mechanisms. … remarkable similarity between translational termination and ribosome rescue mechanisms (Korostelev 2021).

Unlike sense codons read by aminoacyl-tRNAs, stop codons are recognized by release factors. Release factors are bifunctional proteins: they recognize stop codons and trigger the hydrolysis of the ester bond in peptidyl-tRNA. like dRNAs, release factors sample the A site in a compact form, opening into an extended form upon recognition of the stop codon.

RF recognition of a stop codon results in stacking the third nucleobase of the stop codon on G530, and rearranges A1492 and A1493 into a termination-specific conformation. Packing the switch loop W319 in RF2 against A1492 and A1493 directs domain 3 into PTC for release.

Elongation in the early duplisome, polynucleotidyl-dRNA in the P-site and duplicon-dRNA in the A-site, and eventual release if no dRNA in the A-site. In the tagging duplisome, amino acid tagging with rapid release if aminoacyl-tRNA in the A-site. In the breakout ribosome, peptide bond formation if peptidyl-tRNA in the P-site and aminoacyl-tRNA in the A-site, and eventual release if no tRNA in the A-site.

dedicated initiator codon and tRNA

dedicated termination codons …. ribozymatic with uncharged tRNAs? …enzymatic with coded releasing factors ….

One question for early ribosome … aminoacyl-tRNA in the P-site initiates polypeptide translation. Is a peptidyl-tRNA in the P-site able to block entry or accommodation of duplicon-dRNA in the A-site? [ASIDE the curious influence of

multiple mismatches] QUESTION can one dRNA in the A-site flush the polypeptidyl-tRNA by peptide hydrolysis NO!! ???? can dRNA tRNA scrunch I think no!

LOGIC OF TERMINATION tagging tRNA takes advantage of hungry ribosome to tag-and-release … regulation of elongation versus termination on individual template or all templates … mechanism to prevent premature tagging? tagging/termination codon requirement

Here we suggest that ribosome hydrolysis of peptidyl-tRNA for protein release was exapted from duplisome hydrolysis of nucleotidyl-dRNA for RNA elongation, release or tagging. Normal termination of protein translation requires a nonsense codon in the decoding center A-site, confirmed by a class I release factor that undergoes an opening movement, or accommodation in the large subunit. In bacteria there are two class I release factors, RF1 and RF2, that read UAR and URA codons, respectively. Accommodation places the conserved GGQ motif, which has a methylated amide on the glutamine side chain, into the peptidyl transferase center. It is unknown whether this motif activates hydrolysis directly, or acts indirectly by allowing solvent access. Under non-physiological conditions (e.g., 30% acetone), ribosomes can efficiently hydrolyze the peptidyl-tRNA in the P-site with the class I release factor replaced by deacyl-tRNA or just the mimic CCA in the A-site (Caskey et al 1971; Bao et al 2022). The GTPase class II release factor RF3 ensures the fidelity of termination … .

IDEA Before the evolution of coded release factors, any mechanism of polypeptide release was spontaneous, or mediated by RNA and noncoded polypeptides. Conceivable slow release from the absence of any tRNA in the A-site. Alternatively, nonsense tRNAs that pair with stop codons but cannot be charged uncharged tRNA release factors not available to early ribosome … uncharged nonsense tRNA? … just absence of anything in A-site … maybe no designated stop codons, but slow release when no tRNA or no charged tRNA in A-site … [compare polypeptide and polynucleotide release] IDEA hydrolysis is an obligate first step in polynucleotide transfer center … if only it happens because no duplicon-dRNA in A-site then this is polynucleotide release ; in transpeptidation the tetrahedral intermediate SN2 … don’t hydrolyze ester bond then condense amide bond … but in peptide release still hydrolyze ester bond

How was polypeptide translation terminated before the evolution of coded release factors? One suggestion is that the peptidyl-tRNA underwent slow hydrolysis whenever elongation was stalled by prolonged absence of the cognate aminoacyl-tRNA in the A-site. This might occur in unfed, or hungry ribosomes with nothing in the A-site, as well as misfed ribosomes with a deacyl- or similarly damaged tRNA accommodated in the A-site. There was likely a trade-off between hasty termination and endless delay. Besides slow hydrolysis as a default for hungry ribosomes, deacyl-tRNAs might trigger a fast affirmative decision to terminate. The possibility of programmed and regulated termination at designated termination codons in the presence of a terminator tRNA that was currently deacylated or constitutively unacylated for want of a cognate charging ribozyme. [Crick? Brenner?] [nonsense codons]

The regulation of protein termination after the … protein translation. [The rapid termination of polypeptides when uncharged tRNA accommodated in the A-site; selected for better means of keeping them out] EF-Tu and monitoring acylation status …() occurs spontaneously if there is no faster competing reaction from the aminoacyl-tRNA in the A-site. () uncharged A-site tRNA () catalyzed by releasing factors in the modern ribosome that read nonsense codons () conjectured that uncharged tRNAs, either accidentally uncharged or dedicated nonsense tRNAs with no active ribozyme T

[conformation switch ONE] between empty and occupied A-site

[conformation switch TWO] between aa-tRNA and deacyl-tRNA/RF occupied A-site

There have been many attempts to separate the ribosomal contributions to peptide bond formation, and peptide release, respectively, and to distinguish direct roles in catalysis, from possible conformational switching in the peptidyl transfer center. In bacteria and eukarya, respectively, fungal antibiotics lincosamide/lincomycin and streptogramin A/anisomycin inhibit peptide bond formation and stimulate peptide release (Caskey et al 1971; Polacek et al 2003).

Several conserved nucleotides in the inner shell of the peptidyl transferase center are essential for protein release, but not peptide bond formation (ref). the highly mobile universal nucleotide H93 A2602 is crucial for release but not for transpeptidation (?also U2585) …. “inner shell” peptide release (not transfer) A2451 U2506 **U2585 A2602**

Amort et al 2007 .. ribose of A2602

the two reactions, hydrolysis and condensation, nucleotide H93 A2602 requires for release in ribosome, conformational switch between hydrolysis and peptide bond formation, suggest the switch between hydrolysis and condensation;

Focused on the universal nucleobases in the inner shell of the polypeptidyl transferase center, notably A2451 U2506 U2585 A2602, each within 5 A of the carbonyl carbon of the peptidyl-tRNA. Nearly all substitutions of these four nucleotides are dominant lethals; generally rate of polypeptide bond formation reduced 30- to 9000-fold with A-site puromycin (Youngman et al 2004). Even so, more important for polypeptide release than peptide bond formation.

U2506 U2585 A2602 crucial for peptide-bond formation and release; switch from uninduced to induced state from accommodation of the A-site by aa-tRNA or class I release factors to allow attack of activated nucleophiles

induced U2506.G2583 pair

conserved nucleobases within 5 A of tetrahedral carbon

C2063 A2451 U2585 A2602

A2451 2'OH important, but not nucleobase

C2063.A2450 wobble basepair

none of the four nucleobases affect peptidyl transfer too much

delete A2602 or A2602C abolishes all detectable hydrolysis

A2058 A2059 entrance NPET

C2063 universal

G2251 mutations affect transfer

G2252 mutations affect release only

A2451 2' OH transfer wire, nucleobase unimportant

C2501 (A2450.C2063) base triple

C2501//G2502 P-site//A-site

G2553 mutations affect transfer

A2572 release only; sensor of conformation of aa-tRNA

C2573 release only

C2585 universal

H93 A2602

Whereas nucleotidyl transfer involves two sequential reactions, hydrolysis and condensation, peptidyl transfer rapid/concerted aminolysis … peptide bond formation and peptide release .. (Polacek & Mankin 2005)

15-20 peptide bonds per second in bacteria (Katunun et al 2002).

RF catalyzed peptide release 0.5-1.5 per second (Zavialov et al, 2002)

E coli A2602 (yeast A2971) in H93 of domain V important in coordination of bond formation & release

Protein families that use (d)NTPs

Following the breakout of polypeptide translation, immediate changes in the rRNAs, tRNAs, and other ribozymes and riboswitches to improve the speed and fidelity of translation. More gradually, and more profoundly, coded polypeptides and folded proteins were incorporation into metabolism and energetics generally, and protein synthesis and translation. In an extended feedback of coded polypeptides and folded proteins on the evolution of translation, many of the later improvements in decoding and translocation entailed protein biogenesis factors including RNA modification enzymes, stably bound ribosomal proteins, and cyclical translation factors.

[kinase] ROH attacks gamma phosphate NTP => Rp + NDP leaving
ROH attacks beta phosphate NTP => Rpp + NMP leaving
ROH attacks alpha phosphate NTP => RONMP + pp leaving

[nucleotidyl transferase superfamily] [alpha phosphate electrophile | (d)NTP + ROH => RO(d)NMP + pp]

[P-loop NTP hydrolases]
    G-proteins: small GTPases, EF-Tu, G alpha, FtsY, Ffh
    kinases: thymidylate kinase
    motors: myosin, kinesin
    RecA: RecA, ATP-dependent helicases, ABC transporters
[Ras-fold, P-loop related]
    tubulin
    FtsZ
[P-loop, but no Ras-fold]
    PEP-carboxykinase

motor proteins from here to there

myosin, kinesin,

Whatever the sources of mechanochemical free energy in RNA life, metabolism in protein life consolidated around nucleotide triphosphates as high-energy intermediates. A panoply of new enzymes, *viz*. nucleotidyl transferases, kinases, phosphatases, etc., used this currency to drive reactions uphill, impose direction on otherwise reversible reactions, or choose between alternative conformations or products. As a rule, enzymes used ATP for covalent modification of small molecule, and GTP for non-covalent complexes with macromolecules. That is, ATPases mediated covalent events and GTPases regulated conformational changes. Biochemists and cell biologists discovered that hydrolysis of GTP was used to choose between … select between two otherwise choices, or to repeat a choice (kinetic proof-reading) so that a final error occurred only if both choices were wrong. [enzymes of NTP synthesis]

Notwithstanding nucleotide-derived cofactors of modern enzymes, or artificial polymerase ribozymes, there is no clear evidence that NTPs, not just oligonucleotides, trace to the RNA world, much less were used in polynucleotide synthesis. By late protein life, at least three enzyme families had evolved to transfer NTPs to alcohols such as oligonucleotide 3' OH with the release of inorganic pyrophosphate. We conjecture that a promiscuous terminal transferase enzyme took over the provisioning of random oligonucleotides, bypassing and replacing older ribozymatic or spontaneous feedstock reactions. That is, a nucleotidyl transferase enzyme made the random oligonucleotides for dRNA charging in the RNA duplication cycle *before any RdRP enzymes* of RNA. Likely a founding member of the Polβ superfamily whose descendants include non-templated addition 3' tails …. poly(A) polymerases of mRNA polyadenylation, tRNA CCA nucleotidyltransferase, and deoxyribonucleotide terminal transferase. single nucleotide additions, substrate specificity …

Despite their ancient radiation and continued divergence, the common structure of tRNAs is constrained by their interactions … Besides the universal 5' processing of pre-tRNA leaders by RNase P, the universal CCA trailer at the 3' end of all tRNAs. Like any evolutionary innovation, we seek a parsimonious explanation of when, how, and why the CCA trailer arose, consistent with our explanations of other features. The

obvious function of the universal CCA trailer today is to provide a common handle on the Noah's arc of tRNAs for generic aspects of amino acid charging and peptidyl transfer. In archaea and eukarya, as well as some bacteria, it is added late in tRNA biogenesis to the 3' OH of N73 by a nucleotidyl transferase. [IDEA] with a family of tRNAs … need unique features of anticodon and recognition by cognate aaRS ribozymes or enzymes … need generic features like the elbow for interaction with ribosome …. **the problem of drift in length and sequence** of the 3' trailer used for substrate positioning at the A- and P-sites….

In some bacteria, the terminal CCA is transcribed from the gene and processed by the endonuclease enzyme tRNase Z or. In either case, the same tRNA CCA nucleotidyl transferase repairs tRNA 3' ends that have lost one or more nucleotides. The parsimonious explanation is that enzymatic addition of CCA to tRNA N73 was the primitive state of LUCA and the genomic encoding and processing found is some bacteria is the derived state.

The pairing of the terminal CCA to several RNAs of ancient origin raises the possibility that this tRNA sequence might trace to early protein life, if not to the RNA world itself. These RNA-RNA interactions include: (1) domain of bacterial RNase P RNA, (2) T-box riboswitches, and (3) A-loop and P-loop of the peptidyl transfer center in the large subunit rRNA. The interaction with bacterial RNA P …derived after the encoding of CCA in DNA life. Although the charging ribozymes of primitive tRNAs likely gave rise to the 5' domain of T-box riboswitches, the 3' domain of these riboswitches that regulate transcription and translation may be much younger than translation itself. In particular, there is no clear argument for the presence of the CCA trailer in primitive tRNAs of amino acid tagging or the breakout of translation.

The strongest case for the antiquity of the tRNA CCA trailer comes from its interactions with the ribosome peptidyl transfer center. In relaxed accounts of ribosome evolution, a parsimonious explanation for pairing of tRNA 3' CCA ends within the peptidyl transferase center is that the A- and P-loops are homologous elements dating back to formation of the primordial transferase center by tandem duplication. But our stringent account, neither loops pairing with 3' CCA ends, nor hydrogen bonds to the polypeptide backbone in the exit tunnel have necessary ancestors in the duplisome.

IDEA The the A-loop and P-loop interactions with the 3' CCA extensions evolved piecemeal, first to perfect tagging by the A-site carbonyl carbon, and then after the breakout to target the P-site carbonyl with the A-site amine or water. We propose that these changes in the outer shell of the transferase center and the proximal exit tunnel occurred in breakout ribosome nee tagging duplisome to improve the speed and fidelity of polypeptide elongation. In particular, the short A- and P-loops are not structural homologs, but functional analogs from convergent evolution.[23] A76 in proton shuttle.

If the universal 3' CCA extension of tRNAs indeed arose before the nucleotidyl transferse enzyme, our only recourse would be to invoke an extinct ribozyme, or even less likely, its addition to founding tRNA and perservation against drift in length and sequence by strong selection and frequent recombination. As none of its interactions are proof of such antiquity, its most parsimonious origin is just enzymatic addition by a nucleotidyl tranferase enzyme ancestral to the modern CCA …. A variety of modern nucleotidyl transferase …This primordial nucleotidyltransferase enzyme may have repaired tRNAs shorted by lost of terminal nucleotides. The original reaction may have added one or more nucleotides without strict control of sequence or length, perhaps a single C. [polyadenylation] A back and forth … piecemeal ….precise control of sequence and length of tRNA trailer along with A-loop and P-loop for PTC, and bacterial RNase P.

aaRS enzymes ….

arose only after the breakout of polypeptide translation, and that NCCA recognition elements in the ribosome peptidyl transfer center, bacterial RNase P RNA, and T-box riboswitch discriminator domain, as well as aaRS enzymes all postdate this addition.

The electrophile (carbonyl carbon) of the aminoacyl-tRNA was better positioned for tagging by pairing with the novel A-loop, without compromising the transfer reaction of RNA duplication.

today final step in 3' end maturation made by nucleotidyltransferase enzyme using CTP and ATP substrates on tRNA with N73 the determinant overhang … even in

[23] Helix H80 has a stem of 3 WC pairs topped by a 7 nt loop; H92 has a stem of 5 WC pairs topped by a 5 nt loop.

bacteria with CCA transcribed from gene then tRNA Z ??? the NT is used for repair … also moonlights in regulation?

In the invention of tRNAs from dRNAs, both ends of these molecule were changed, reflecting their new roles in loading and transfer. In maturation of pre-tRNA transcripts, any 3' trailer is trimmed between nucleotides 73/74 by tRNase Z, followed by addition of the universal sequence 3’ CCA by tRNA nucleotidyltransferase. Even in bacteria where the 3’ CCA is generally transcribed from the gene, this nucleotidyltransferase is required to repair erosion of the mature 3’ end. The unpaired 3’ trailer NCCA of the acceptor stem is critical to amino acid charging and peptidyl transfer.

Two changes in ribozyme P likely prevented interference of primitive tRNAs with the dRNA loading cycle, and perhaps inadvertent charging of tRNAs with oligomers at their 5’ end. Threading the universal 3’ NCCA tailor through the new P15/16/17/6 round arch to pair with G292 G293 U294, reinforced by RNA ruler monitoring of the tRNA elbow, locked the substrate at 1/1 alignment in the catalytic site at the, and prevented the reversible two step reaction of staggered condensation and hydrolysis of duplicon loading. [product release is rate-limiting step today … avoid product inhibition, back reaction]

In bacteria, the universal 3’ trailer NCCA of tRNAs interacts with no fewer than four different RNA segments: ribosome peptidyl transfer center P- and A-loops, RNase P RNA, T-box riboswitch discriminator] These RNA RNA interactions have suggested to some that the NCCA extension dates to the RNA world itself, and was perhaps present on the original hairpin that formed the 3’ half, or both halves of the primitive cloverleaf tRNA. After pondering this argument, we are not convinced that the NCCA trailer predates protein life. The interaction between bacterial RNase P RNA and the NCCA is likely a derived, and not a primitive feature of the RNase P RNA. In particular, it is absent in archaea and eukarya RNase P which acquire CCA by NT. If so, it is quite plausible that the other short basepairing interactions between ribosomal RNA and the tRNA trailer and between T-box riboswitches and the tRNA trailer are derived at well.

From this viewpoint, that the several basepairing interactions do not preclude the comparatively late acquisition of the NCCA trailer, the most parsimonious origin of the

3’ extension is simply a nucleotidyl transferase enzyme, ancestral to the modern CCA adding enzyme. If so, the tRNA trailer arose in protein life and used NTPs as substrate. This conjecture of when, and how, only sharpens the question of why.

We are agnostic whether addition of the 3’ tail to the primitive tRNA was part of more general mechanism to protect and repair RNA ends, or more specific mechanism to reduce interference between RNA duplication and polypeptide translation, to improve the rate of peptide bond formation by substrate position, or interactions with the new charging enzymes.

To encourage discussion, we conjecture that the addition of the universal 3’ CCA by terminal nucleotidyltransferase enzyme demarked tRNAs for enzymatic charging from those for ribozymatic charging … pari passu … the original TENT may have added oligo(A) and been more tolerant of sequence and perhaps length as well

PROBLEM? T-box riboswitches inspect CCA motif?

IDEA the complementary changes in the polypeptidyl transferase center … accretion of the A-loop and the P-loop.

IDEA changes in bacterial RNase P RNA for transcribed CCANNNN

[distance] duplicon extends 2 nucleotides past N1, if N1 paired with N72 in tRNA then N76 is 4 nucleotides plus amino acid [angstroms?]

target is 4 nt past N72, aa is 4 nt + 1 aa past N72

target is 0 nt past N1, duplicon is 2 nt past N1

The remarkable variety, speed and fidelity of protein synthesis, as well as its enormous burden on the cellular economy, trace to GTPase enzymes of translation and protein secretion, as well as ATPase enzymes of amino acid anabolism and tRNA charging.

After the early expansion of the amino acid code by duplications of charging ribozymes, these were supplemented and replaced piecemeal by charging enzymes, probably because proteins could discriminate between hitherto interchangeable amino acids. That is, the conjectured charging ribozymes T were better suited to reading one anticodon, than picking out just one amino acid. Moreover, one enzyme could charge

several tRNA isoacceptors with different anticodons. For example, *Escherichia coli* LeuRS charges five isoacceptors with anticodons GAG, CAG, UAG, CAA and UAA, reading codons CUY, CUG, CUR, UUG and UUR, respectively. Even after all ribozymes T were replaced by charging enzymes, there were likely further duplications of the enzymes and their tRNAs. Twenty-three different aaRS enzymes, comprising two distinct catalytic folds, are found in cellular life today (Rubio Gomez & Ibba 2020). [as many as ## isoacceptors][FOOTNOTE isotype, isoacceptor, isodecoder]

Whereas the stereochemical fit theory of the amino acid code proved unsupported, and unnecessary, other early explanations have fared better. …. [MOVE UP to section 13; seem compatible with a modular charging ribozyme like the conjectural ribozyme TKL (ref).] In the error minimization theory, chemically similar amino acids were assigned near cognate codons to reduce the harm of occasional errors in charging or decoding. [Janzen et al 2022] We suggest a simple process … duplication and specialization of tRNAs and their charging enzymes partitioned blocks of codons x amino acids that had been interchangeble into … During protein translation, the distinction between common near-cognate mistakes, and rare non-cognate mistakes may explain at least some of the assignments in the genetic code (Koonin).

The ideas of error minimization and piecemeal accretion of new amino acids, and splitting of codon boxes … produce the genetic code today which indeed not so much as a frozen accident as a frozen habit that satificed ….

more anabolically complex amino acids added later; co-evolution of ribozymatic and enzymatic amino acid biosynthetic pathways, and tRNA loading and code …. frozen accident …

Amidst the refinement of the genetic code, splitting up ….some amino acids were culled from of Conversely, other amino acids with the right characteristics for ribozymatic tRNA charging may have been used that were later culled from the set of proteinogenic amino acids.

Pari passu with the **perfection of reading frame**

() modifications of tRNAs, rRNAs to defend frame, refine or relax decoding

() from 2 nucleotide to 3 nucleotide shift of mRNA, defend against scrunch, opens up the code to discriminate third position

- improve frame (no frame at all in tagging, short polypeptides, statistical products)

-improve speed

N37 modifications help maintain reading frame

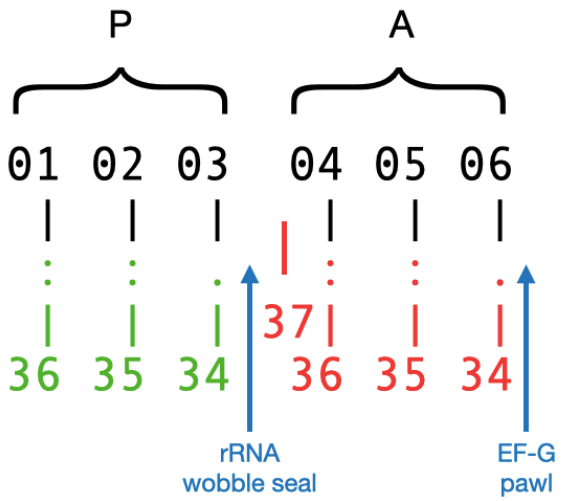


FIGURE 14-. DEFENDING THE READING FRAME

Nearly one-third of the proteome of modern cells is targeted for membrane insertion or secretion. Already LUCA had a phospholipid cell membrane, as well as a mechanism for co-translational secretion based on the signal recognition particle (SRP) bound to the ribosome. SRP scans emerging polypeptides for a hydrophobic segment, either an N-terminal signal sequence, or a presumptive first transmembrane α-helix, both of which cytoplasmic proteins lack (Voorhees & Hedge 2015). Once it engages this first hydrophobic segment, SRP pauses ribosomal elongation until it docks with the SRP receptor (SR) on the cell or ER membrane. There it hands the ribosome-nascent chain complex over to the membrane translocon (prokarya SecY, eukarya ER Sec61) to resume translation, coupled now with secretion.

In all cellular kingdoms, SRP comprises SRP RNA and a conserved secretory GTPase (prokarya Ffh, eukarya SRP54), while its membrane receptor comprises an

homologous GTPase (prokarya FtsY, eukarya SRα). SRP RNA itself has two domains called S and Alu that recognize the hydrophobic segment for membrane docking, and arrest elongation, respectively. The S domain binds Ffh/SRP54 via conserved bulges in H8, presenting it to the nascent polypeptide exit tunnel where this protein recognizes emergent hydrophobic sequences via its methionine-rich C-terminus. N-terminal helix of SRP54 interacts near exit tunnel with 5.8S rRNA, uL29, uL23. The SRP/SR GTPases act as mutual GTPase activating proteins, but SRP RNA accelerlates their binding and dissoaciation by over 100 fold . tetraloop GNRA in SRP RNA catalyses the productive interaction between fh/FtsY GTPases. reciprocal activation of GTPase NG domains of SRP54 & SRα …The GTPases of the SRP and its receptor SR … reciprocal activation in docking ..

The Alu domain of SRP RNA docks at the interface of the two ribosome subunits, where it blocks the binding site for translational GTPases (pka translation factors), and slows or pauses elongation. In *Bacillis subtilis* this domain interacts directly with elements of the large subunit rRNA, *viz*. the stalk-base helices H43 H44 and the α -sarcin-ricin loop on H95 (Beckert et al 2015). Elongation arrest is more important in eukaryotic cells where ribosome-SRP complexes need considerable time to encounter SR receptors on the ER, and hand over to the translocon, than smaller, bacterial cells where these complexes encounter SR receptors on the cell membrane more quickly (Wild et al 2020). In *Escherichia coli*, for example, the Alu domain of SRP RNA has been lost entirely, leaving only the S domain.

To encourage discussion, we conjecture that SRP RNA originated in the RNA world as a riboswitch to regulate the elongation of polynucleotides, and later, the elongation of polypeptides, *before any support of coded proteins*. In particular, the ribosome factor binding site for translational GTPases was originally a duplisome binding site for Alu-riboswitches that enabled or disabled elongation (Ahl et al 2015). An early form of gene regulation, there were several likely roles for Alu-switches with sensor domains: (1) pausing polymer elongation to regulate or synchronize growth, forcing polymer release, and diverting resources from completion to initiation; (2) blocking elongation of

polymers with specific identifier sequences; and (3) blocking polymers with foreign, pathogenic, or misfolded segments.

In late protein life, the ribosome factor binding site was co-opted by translational GTPases, and the only vestige of ancient Alu-riboswitches in LUCA was SRP RNA. Today SRP RNA works with signal recognition GTPases for membrane targeting, and works against translational GTPases for elongation arrest.[24]

Like pre-tRNAs, pre-SRP RNA is an ancient substrate of RNase P, hinting at possible coordinations of RNA duplication and polypeptide translation.

Q. E. coli lacks Alu domain, check ribozyme P sites of full-length SRP RNAs

Coupling initiation, elongation, and termination to the free energy of GTP hydrolysis, translational GTPases lower the energy barrier between ribosome states, and ratchet these transitions in one direction (Table 14-).[25] The high activation barrier of modern ribosomes for transitions between PRE and POST states (20-25 kcal / mol) likely evolved *pari passu* with an elongation factor as the means of overcoming it (Schilling-Bartetzko et al 1992). Gene duplication gave rise to two elongation factors, one GTPase specialized for POST to PRE transitions (bacteria EF-Tu / eukarya EF1), and the other for PRE to POST transitions (bacteria EF-G / eukarya EF2). By separating decoding into two step, initial matching and final accommodation, EF-Tu allows codon/anticodon pairing with the minor groove monitoring to be used twice, for a nearly multiplicative improvement. As a result of this kinetic proofreading, substitution errors of the modern ribosome is only ##. … bacteria 20 residues per second polypeptide elongation.

[move EF-Tu proof-reading here]

latched lets EF-Tu hydrolyze GTP and dissociated

now stay latch long enough for accommodation or unlatch and release aa-tRNA

IDEA non-cognate tRNAs leave with intact ternary complex

near-cognate tRNAs pass the GDP test but unlatch before accommodation

---

[24] Quite recently, in primates the SRP Alu-riboswitch has been co-opted by the SINE retrotransposon to help steal the nascent RdRP enzyme of LINE mRNA (ref).

[25] reverse direction GTPase EF4

ADD EF-Tu and use latch discrimination twice in kinetic proof-reading

EF-Tu but not EF-G mutations cause frameshifts???

EF-Tu proof-reading

[polypeptides/proteins in decoding]

S12 serine46 (E. coli) also invovled with second base-pair

EF-Tu

[move EF-G reading frame here]

[factor assisted initiation]

In factor-assisted initiation, the ribosome small subunit encounters an unfolded region of the 5' UTR and scans the downstream mRNA for the start codon by its interaction with the anticodon of the charged initiator tRNA in the P-site. Once the GTPase initiation factor (IF2) validates the tripartite complex of small subunit rRNA, mRNA, and initiator tRNA, it assembles with the large subunit, allowing the aminoacyl-tRNA complexed with the GTPase elongation factor (EF-Tu) to enter the A-site for decoding and peptide bond formation. In prokaryotes, an RNA duplex between the Shine-Dalgarno box, and the anti-SD sequence at the 3' end of the small subunit rRNA, roughly positions the mRNA. and finally, matching the anticodon of the initiator tRNA at the P-site. initiation factors recruit fMet-tRNA^{fMet} then locates initiator tRNA with AUG in P-site and prevent association with the large subunit until correct assembly of small subunit mRNA and initiator tRNA … then acquires large subunit then aa-tRNA in A-site …. The correct conformation recognized by IF2 GTPase and assembly with the large subunit for the elongation cycle. In eukaryotes, encounters the capped 5' end of the mRNA, then scans the 5' UTR with the initiator tRNA for the first start codon.

[factor assisted termination]

IDEA the modern ribosome is remarkably faster than other ribozymes … 20 elongations per second versus one per day …10**6 faster than conjectural diurnal duplisome … ribosomal proteins & translational GTPases

The elongation cycle of the ribosome has about -30 kcal / mol from the hydrolysis of two GTPs and the favorable free energy of peptide bond formation, compared to only -12 kcal / mol for the conjectured duplisome driven by the opening and closing dRNA thermal motors (Table 14-). The free energy of folding for RNA and protein domains may be comparable, but the folding of RNA domains is immediately offset by the cost of unfolding the RNA template, while the free energy of protein domains ultimately derives from cellular metabolism that makes the amino acid pools.

	duplisome	ΔG kcal / mol	ribosome	ΔG kcal / mol
Initiation			IF2	-12
Elongation		- 12 / step		- 30 / step
Read	dRNA opening	12	EF-Tu	-12
Add	nucleotidyl transfer	~ 0	peptidyl transfer	-6
Move	dRNA closing	-24	EF-G	-12
Termination			RF3	-12
Recycling			EF-G	-12
Folding	RNA domain	-10	protein domain	-10

TABLE 14-#. ENERGETICS OF RNA DUPLICATION & PROTEIN TRANSLATION

The greater speed or fidelity of RNA-templated RNA polymerase (RdRP) enzymes caused the more-or-less abrupt retirement of ribozymatic copying in favor of enzymatic replication. The duplisome hypothesis suggests this selective sweep was driven mostly by increased speed, while the processivity and fidelity of modern polymerase enzymes from helicases, sliding clamps, proof-reading etc., evolved later. Thus, given simply

adequate fidelity, primer extension using NTPs would easily outpace cells that added only one duplicon per day. One serious drawback in comparison to RNA duplication, however, is that the RdRP replication cycle produced long duplexes that required strand separation, and invested every other round of copying in making minus strands that neither folded as ribozymes nor functioned as mRNAs. [FOOTNOTE] Similar limitations on spontaneous RNA copying gave a leg up to the duplisome in the RNA world.

[original replicative RdRP] Polymerase enzymes arose independently several times in protein life (Koonin et al 2020). Three lineally unrelated core catalytic domains called Polβ-like, RRM-Palm, and 2xDPBB are found in polymerases today, any, or all, of which may trace to RdRPs involved in RNA replication, recombination, or repair in protein life.[26] The Polb are found in … The RdRPs of RNA viruses have the RRM-Palm domain, while the RdRPs of eukaryotic RNA interference have two DPBB domains in a single subunit. When the duplisome and its dRNAs were retired, the ribosome and its tRNAs were already wildly successful exaptations, ensconced in the role of protein translation.

The breakout of DNA life required two polymerase enzymes: DdRP transcriptase (2xDPBB fold) and RdDP reverse transcriptase (RRM-Palm fold). Both of these enzymes were no doubt exapted from older RNA polymerases of the corresponding fold. Together these polymerases copied RNA into duplex DNA for longterm storage, and copied DNA into noncoding RNAs and mRNAs for gene expression and action. One parsimonious suggestion is that transcriptase arose from the replicative RdRP through a modest change in template preference from RNA to DNA. If so, the breakout of DNA life only required reverse transcriptase, plus a robust source of dNTPs. In modern cells dNTPs are made from NDPs by *ribonucleotide reductase* which converts them to the corresponding dNDPs, followed by *nucleoside diphosphate kinase* which converts (d)NDPs to (d)NTPs. This observation that deoxyribonucleotides are made from their ribonucleotides, and not from deoxyribose directly, was one of the original arguments for the RNA world (Long et al 2022).

---

[26] (Polβ-like) Polβ-like nucleotidyl transferase, (2xDPBB) two double-psi beta-barrel domains contained in either one subunit or two separate subunits, (RRM-Palm) RNA recognition motif or Palm domain,.

Of two chief advantages of DNA, greater chemical stability of the backbone was immediate, while various error-free repair pathways exploiting the informational redundancy of the regular duplex evolved over time. The ribose 2' OH needed for folding motifs and catalytic virtuosity of RNA was a liability for information storage. Without energetic incentives to find hydrogen bonding, reverse complements DNA strands formed long regular duplexes with the B form sugar pucker precluded for ribose. … most high complexity sequences … still easy to separate …

the RNA versatility in folding and catalysis, catalytic The the folding motifs [ribose zipper] and catalytic virtuosity of the ribose 2' OH not needed, the B form pucker, its contribution to strand scission via 2',3' cyclic phosphate became a major liability.

Without energetic incentives to find tertiary, DNA forms settles regular double-helix … redundant sequence information complementary strand paired and available not just to template copies but for error-free repair of damage confined to either strand. Now worthwhile to replace the problematic uracil, which might belong or be product of cytosine deamination, by thymine, exposing the remaining uracil in DNA as a damaged nucleobase for base-excision repair. protect, repair DNA ends .

[dUTP to dTTP] why isn't dUTP incorporated into DNA?

the replacement of uracil by 5-methyl uracil (aka thymidine) in DNA … in modern metabolism dUMP is converted to dTMP by *thymidylate synthetase*, and then to dTDP by *thymidylate phosphate kinase*, and finally dTTP by the common nucleoside diphosphate kinase. The dUMP itself derives from either dUTP or dCMP itself from dCDP???

DdDPs do not distinguish between major pool of dTTP and minor pool of dUTP … kept low by dUTPase … the U:A can be removed immediately, and U:G that forms post-replicatively from cytosine deamination can be removed … finally DdRPs can bypass uracil in transcription

In the final curtain call of the RNA world, DdDPs all but eliminated the need for RNA intermediates in DNA replication. Remarkably, some major components of DNA replication including primases and main helicases, as well as the replicative DdDPs themselves, are not homologous among bacteria, archaea, and eukarya (Koonin et al

2020). Other components of DNA replication are universally conserved including the sliding clamp (PCNA), clamp loader ATPase, and ssDNA-binding protein =?RecA. One parsimonious proposal is that the replicative DdDP of LUCA, like the universal DdRP, arose from a 2xDPBB-family RdRP. This PolD (DPBB-type) polymerase remains the main replicative DdDP in archaea, but has been replaced by PolC (Polβ-type) or PolB (RRM-type) DdDPs in bacteria and eukarya, respectively. Whatever the actual history of genome replication and repair in cellular and viral life, these fundamental processes were fluid, not fixed, with occasional switching of polymerases from RNA to DNA templates, and back, as well as non-orthologous gene displacements of polymerases and other protein components.

DNA genomes brilliantly satisfied the criteria of ideal storage from computer science. Shannon had formulated memory stores to WRITE and later READ information, or communication channels to SEND and remotely RECEIVE information as devices adopting any of several alternative configurations, or MESSAGES, often realized as a sequence of letters from some fixed alphabet (Shannon 1948). What made an ideal storage medium were: (1) Uniform storage of stores all sequences irrespective of actual content. (2) Stability in passive storage and active use (non-destructive or repeatable read). (3) One-shot copying (writing) from one store to another of the same (reading with one-shot writing or overwriting) (4) or translation into a different medium or format (5) Separation of well-formatted messages to allow error detection and correction.

Underpinning the success of DNA chromosomes, the imprimatur of the RNA world is still seen in the processes of vertical/faithful and horizontal/creative genome copying and transmission. In VGT, or the physiology of exploitation (physiology)] … vestiges in RNA primers of DdDp … what makes primers?? … IDEA transcription and translation is physiology of gene expression … [in HGT, the process of exploration (evolution)] more importantly, RNA viruses and transposons … reminder that novel genes in heterochromatin are not passive copied from DNA but creative RNA intermediates (Huttar et al 2000) … Meanwhile, RNA poses a means to mixis but a threat of hostile takeover, and selfish replication (Fire & Mello).

VGT and HGT, cellular and MGE/viral,…. reverse transcription is normal part of the life cycle for retrotransposons and retroviruses but considered a vestigial and exotic .. in fact, creative process of new genes, and epigenome process of ….

Ga	Era	RNA	polypeptides	DNA
4.5	prebiotic	random	random	none
	early RNA life	spontaneous copying	random	none
	late RNA life	duplisome	random	none
	early protein life	duplisome →	ribosome	none
	late protein life	RdRP enzyme	ribosome	none
	early DNA life	DdRP enzyme	ribosome	RdDP enzyme
3.7	late DNA life	DdRP enzyme	ribosome	DdDP enzyme

TABLE 14-2. FROM EARLY TO LATE DNA LIFE

Acknowledgements

This paper grew out of lectures on the RNA world to our human genetics class at Johns Hopkins University; we wish to thank the students for letting us explore these ideas. We also wish to acknowledge our mentors who introduced us to science: Maurice Bessman, Marshall Blaufarb, Glenn Flowers, Phil Hartman, Ed Lewis, Hill, Ray Owen, Dick Russell, Aaron Shatkin, John Sulston and John White. Finally, we thank our current colleagues …[Rachael, Vince, Joe Culotti, Any Fire, Bob Herman, Edward Kipreos, Carolyn Norris, Bruce Vogel, Haiqing Zhao] … for helpful feedback, and especially, encouragement to tackle something so inherently speculative, as well as outside our scientific comfort zone.

References

Agmon (2022).

Agmon I, Bashan A, Zarivach R & Yonath A (2005). Symmetry at the active site of the ribosome: structural and functional implications. Biol Chem 386, 833-844.

Agmon I, Davidovich C, Bashan A & Yonath A (2009). Identification of the prebiotic translation apparatus within the contemporary ribosome. Nat Prec …

Ahl et al (2015).

Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schottler MA, Ruf S & Bock R (2012). The contributions of wobbling & superwobbling to the reading of the genetic code. *PLOS* 8, 1-16.

Amort et al (2007).

Anastasi et al (2006).

Andachi Y, Yamao F, Muto A & Osawa S (1989). Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*. Resemblance to mitochondria. *J Mol Biol* 209, 37-54.

Atkins JF & Bjork GR (2009). A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol Mol Biol Rev* 73, 178-210.

Attwater J, Wocher A & Holliger P (2013). In-ice evolution of RNA polymerase ribozyme activity. *Nat Chem* 5, 1011-1018.

Baltimore D (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226, 1209-1211.

Ban et al (2000).

Bao et al (2022).

Becker S, Thoma I, Deutsch A, Gehrke T, Mayer P, Zipse H & Carell T (2016). A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* 352, 833-836.

Becker S, Feldmann J, Wiedmann S, Okamura H, Schneider C, Iwan K, Crisp A, Rossa M, Amatov T & Carell T (2019). Unified prebiotic syntheses of pyrimidine and purine RNA building blocks driven by wet-dry cycles. *Science* 366, 76-82.

Beckert B, Kedroy A, Sohmen D, Kempf G, Wild K, Sinning I, Stahlberg H, Wilson DN & Beckmann R (2015). Translational arrest by a prokaryotic signal recognition particle is mediated by RNA interactions. *Nature Structural Mol Biol* 22, 767-773.

Been MD & Cech TR (1988). RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* 239, 1412-1416.

Benner et al (2012).

Benner S, Kim HJ & Biondi E (2018). Mineral-organic interactions in prebiotic synthesis. The discontinuous synthesis model for the formation of RNA in naturally complex geological environments. In *Prebiotic Chemistry & Chemical Evolution of Nucleic Acids* (ed Menor-Salvan C) Springer International.

Bernal (1951).

Bohnsack & Bohnsack (2019).

Bokov K & Steinberg SV (2009). A hierarchical model for evolution of 23S ribosomal RNA. Nature 457, 977-980.

Bonitz SG, Berlani R, Coruzzi G, Li M, Macino G, Nobrega RG, Nobrega MP, Thalenfeld BE & Tzagoloff A (1980). Codon recognition rules in yeast mitochondria. *Proc Natl Acad Sci USA* 77, 3167-3170.

Bose T, Fridkin G, Davidovich C, Krupkin M, Dinger N, Falkovich AH, Peleg Y, Agmon I, Bashan A & Yonath A (2022). Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds. *Nucleic Acids Res* 50, 1815-1828.

Bousset L, Mary C, Brooks MA et al (2014). Crystal structure of a signal recognition particle Alu domain in the elongation arrest conformation. RNA 20, 1955-1962.

Brännvall M, Kikovska E, Wu S & Kirsebom LA (2007). Evidence for induced fit in bacterial RNase P RNA-mediated cleavage. *J Mol Biol* 372, 1149-1164.

Breslow (1959).

Bretscher (1968).

Butlerow (1861).

Cairns-Smith AG (1965). The origin of life and the nature of the primitive gene. *J Theoret Biol* 10, 53-88.

Cairns-Smith AG & Davies CJ (1977). The design of novel replicating polymers. In *Encyclopaedia of Ignorance* (editors Duncan R & Weston-Smith M). Pergamon Press, Oxford. pp 391-403.

Cairns-Smith AG (1982). *Genetic Takeover: and the Mineral Origins of Life*. Cambridge University Press, Cambridge UK.

Cairns-Smith AG & Hartman H (1986). *Clay Minerals and the Origin of Life*. Cambridge University Press, New York.

Campbell JH (1991). An RNA replisome as the ancestor of the ribosome. *J Mol Evol* 32, 3-5.

Campbell TD, Febrian R, McCarthy JT, Kleinschmidt HE, Forsythe JG & Bracher PJ (2019). Prebiotic condensation through wet-dry cycling regulated by deliquescence. Nature Comm 10, 1-7.

Carpenter FH (1960). The free energy change in hydrolytic reactions: the non-ionized compound convention. *J Am Chem Soc* 82, 1111-1122.

Caskey CT, Beaudet AL, Scolnick EM & Rosman M (1971). Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc Natl Acad Sci USA* 68, 3163-3167.

Cech TR (1986). A model for the RNA-catalyzed replication of RNA. Proc Natl Acad Sci 83, 4360-4363.

Cech (2009).

Chan et al (2013).

Chillon I & Marcia M (2021). Self-splicing group II introns. In Ribozymes … 143-167.

Chyba CF, Thomas PJ, Brookshaw L & Sagan C (1990). Comentary delivery of organic molecules to the early Earth. *Science* 249, 366-373.

Cobb & Pudritz (2014).

Collins & Greider (1993).

Coughlin DJ, Pleiss JA, Walker SC, Whitworth GB & Engelke DR (2008). Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. *Proc Natl Acad Sci USA* 105, 12218-12223.

Crick FHC (1955/1958). On degenerate templates and the adaptor hypothesis. Symp Soc Exp Biol 12, 138-163.

Crick FHC (1956). On protein synthesis. *Symp Soc Exp Biol* 12, 138-163.

Crick FHC (1966). Codon-anticodon pairing : the wobble hypothesis. *J Mol Biol* 19, 548-555.

Crick FHC (1968). The origin of the genetic code. *J Mol Biol* 38, 367-379.

Crick FHC (1970). Central dogma of molecular biology. *Nature* 227, 561-563.

Crick FHC, Griffith JS & Orgel LE (1957). Codes without commas. Proc Natl Acad Sci USA 43, 416-421.

Crick et al (1976).

Cronin (1989).

Dance (2023).

Dao Duc et al (2019).

Dawkins R (1976). *The Selfish Gene*. Oxford University Press.

Deamer & Weber (2010).

Dedkova & Hecht (2019).

Dick TP & Schamel WWA (1995). Molecular evolution of transfer RNA from two precursor hairpins: implications for the origin of protein synthesis. J Mol Evol 41, 1-9.

Dickerson et al (1982).

Di Giulio (1992). On the origin of the transfer RNA molecule. *J theor Biol* 159, 199-214.

Di Giulio M (2004). The origin of the tRNA molecule: implications for the origin of protein synthesis. *J theor Biol* 226, 89-93.

Doolittle WF & Sapienza C (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601-603.

Doudna JA & Szostak JW (1989). RNA-catalysed synthesis of complementary-strand RNA. *Nature* 339, 519-522.

Eigen M (1971). Self-organization of matter & the evolution of biological macromolecules. *Naturwissenschaften* 58, 465-523.

Eigen M & Schuster P (1979). The hypercycle: a principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64, 541-565.

Eigen M & Winkler-Oswatitsch R (1981a). Transfer-RNA: the early adaptor. Naturwissenschaften 68, 217-228.

Eigen M & Winkler-Oswatitsch R (1981b). Transfer-RNA, an early gene? Naturwissenschaften 68, 282-292.

Eigner J, Boedtker H & Michaels G (1961). The thermal degradation of nucleic acids. *Biochim Biophys Acta* 51, 165-168.

Eschenmoser A (1999). Chemical etiology of nucleic acid structure. *Science* 284, 2118-2124.

Eschenmoser A (2004). The TNA-family of nucleic acid systems: properties and prospects. *Orig Life Evol Biosph* 34, 277-306.

Fahnestock S & Rich A (1971). Ribosome-catalyzed polyester formation. *Science* 173, 340-343.

Fialho DM, Roche TP & Hud NV (2020). Prebiotic syntheses of noncanonical nucleosides and nucleotides. *Chem Rev* 120, 4806-4830.

Fry M (2022). Crick's adaptor hypothesis and the discovery of transfer RNA: experiment surpassing theoretical prediction. Philos Theor Pract Biol 14, 11-31.

Fujishima K, Sugahara, Tomita M & Kanai A (2008). Sequence evidence in the Rachael genomes that tRNAs emerged through the combination of ancestral genes. *PLoS ON*E 3, e1622.

Fujishima K, Sugahara J, Kikuta K, Hirano R, Sato A, Tomita M & Kanai A (2009). Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. *Proc Natl Acad Sci USA* 106, 2683-2687.

Gamov G (1954). Possible relation between deoxyribonucleic acid and protein structures. Nature 173, 318.

Gibard C, Bhowmik S, Karki M, Kim EK & Krishnamurthy R (). Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nature Chem* 10, 212-217.

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

Giurgiu C, Li L, O'Flaherty DK, Tam CP & Szostak JW (2017). A mechanistic explanation for the regioselectivity of nonenzymatic RNA primer extension. *JACS* 139, 16741-16747.

Goldman AD & Kacar B (2021). Cofactors are remnants of life's origin and early evolution. *J Mol Evol* 89, 127-133.

Gordon KHJ (1995). Were RNA replication & translation directly coupled in the RNA (+protein?) world? J theor Biol 173, 179-193.

Gößringer M, Schencking I & Hartmann RK (2021). The RNase P ribozyme. In *Ribozymes* 1, 227-279.

Grosjean H & Westhof E (2016). An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res* 44, 8020-8040.

Gruber AR, Lorenz R, Bernhart SH, Neubock R & Hofacker IL (2008). The Vienna RNA websuite. *Nucleic Acids Res* 36, W70-W74.

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.

Gupta R (1984). *Halobacterium volcanii* tRNAs. *J Biol Chem* 259, 9461-9471.

Haruna I, Nozu K, Ohtaka Y & Spiegelman S (1963). An RNA "Replicase" induced by and selective for a viral RNA: Isolation and properties. *Proc Natl Acad Sci USA* 50, 905-911.

Haugen P, Simon DM & Bhattacharya D (2005). The natural history of group I introns. *Trends Genet* 21, 111-119.

Higgs PG (2009). A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biol Direct* 4, 16-.

Higgs PG & Pudritz (2009). A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* 9, 483-490.

Hoagland M (1959). Biochemical activities of nucleic acids: the present status of the adaptor hypothesis. Brookhaven Symp Biol 12 (November), 40-46.

Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR & Zamire A (1965). Structure of a ribonucleic acid. Science 147, 1462-1465.

Hopfield JJ (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic proccesses requiring high specificity. Proc Natl Acad Sci 71, 4135-4139.

Huang S, Aleksashin NA, Loveland AB, Klepacki D, Reier K, Kefi A, Szal T, Remme J, Jaeger L, Vazquez-Laslop N, Korostelev AA & Mankin AS (2020). Ribosome engineering reveals the importance of 5S rRNA autonomy for ribosome assembly. Nature Communication 11, 1-13.

Hud NV, Cafferty BJ, Krishnamurthy R & Williams LD (2013). The origins of RNA and "my grandfather's axe." *Chem Biol* 20, 466-474.

Ianeselli A, Salditt A, Mast C, Ercolano B, Kufner CL, Scheu B & Braun D (2023). Physical non-equlibria for prebiotic nuclei acid chemistry. *Nature Rev Physics*

Inoue T & Orgel LE (1983). A nonenzymatic RNA polymerase model, *Science* 219, 859-862.

Ishida S, Terasaka N, Katoh T & Suga H (2020). An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nature Chemical Biology* 16, 702-709.

Jadhav VR & Yarus M (2002). Coenzymes as coribozymes. *Biochimie* 84, 877-888.

Javaux EJ (2019). Challenges in evidencing the earliest traces of life. *Nature* 572, 451-460.

Jeffares DC, Poole AM & Penny D (1998). Relics from the RNA world. J Mol Evol 46, 18-36.

Jencks WP (1975). Binding energy, specificity, and enzymic catalysis: the circe effect. Adv Enzymol Relat Areas Mol Biol 43, 219-410.

Joyce GF & Orgel LE (1993). Prospects for understanding the origin of the RNA world.

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.

Kamerlin SCL, Sharma PK, Prasad RB & Warshel A (2013). Why nature really chose phosphate. *Quart Rev Biophysics* 46, 1-132.

Kanavarioti A, Monnard PA & Deamer DW (2001). Eutectic phase in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiol* 1, 271-281.

Kawabata M, Kawashima K, Mutauro-Aoki H, Ando T, Umehara T & Tamura K (2022). Peptide bond formation between aminoacyl-minihelices by a scaffold derived from the peptidyl transferase center. *Life* 12, 573-583.

Khaitovich P & Mankin AS (1999). Effects of antibiotics on large ribosomal subunit assembly reveals possible function of 5 S rRNA. *J Mol Biol* 291, 1025-1034.

Kim SH, Sussman JL, Suddath Fl, Quigley GJ, McPherson A, Wang, AHJ, Seeman NC & Rich A (1974). The general structure of transfer RNA molecules. *Proc Nat Acad Sci USA* 71, 4970-4974.

Kim SC, O'Flaherty DK, Giurgiu C, Zhou L & Szostak JW (2021). The emergence of RNA from the heterogeneous products of prebiotic nucleotide synthesis. *J Am Chem Soc* 143, 3267-3279.

Kirsebom LA & Svard SG (1994). Base pairing between Escherichia RNase P RNA and its substrate. *EMBO J* 13, 4870-4876.

Kirsebom LA & Trobro S (2009). RNase P RNA-mediated cleavage. *Life* 61, 189-200.

Kirschning A (2021). Coenzymes and their role in the evolution of life. *Angew Chem Int Ed* 60, 6242-6269.

Koga T & Naraoka H (2017). A new family of extraterrestrial amino acids in the Murchison meterorite. *Scientific Reports* 7, 636-643.

Koonin EV, Mushegian AR & Bork P (1996). Non-orthologous gene displacement. Trends Geneti 12, 334-

Kornberg A (1969). Active center of DNA polymerase. *Science* 163, 1410-1418.

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.

Krupkin M, Matzov D, Tang H, Metz M, Kalaora R, Belousoff MJ, Zimmerman E, Bashan A & Yonath A (2011). A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. *Phil Trans R Soc* B 366, 2972-2978.

Leidig C, Thoms M, Holdermann I, Bradatsch B, Berninghausen O, Bange G, Sinning I, Hurt E & Beckmann R (2014). 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. Nature Comm 5, 3491.

Leininger SE, Narayan K, Deutsch C & O'Brien EP (2019). Mechanochemistry in translation. *Biochemistry* 58, 4657-4666.

Levy M & Ellington AD (2001). The descent of polymerization. *Nat Struct Mol Biol* 8, 580-582.

Li Y & Breaker R (1999). Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J Am Chem Soc* 121, 5364-5372.

Licznar P, Mejlhede N, Prere MF, Wills N, Gesteland RF, Atkins JF & Fayet O (2002). Programmed translational -1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *EMBO J* 22, 4770-4778.

Lipmann F (1971). Attempts to map a process evolution of peptide biosynthesis. *Science* 173, 875-884.

Liu Z, Wu LF, Xu J, Bonflo C, Russell DA & Sutherland JD (2020). Harnessing chemical energy for the activation and joining of prebiotic building blocks. *Nat Chem* 12, 1023-1028.

Martínez-Giménez JA & Tabarés-Seisdedos R (2021). Possible ancestral functions of the genetic and RNA operatonal precodes and the origin of the genetic system. *Origins Life Evol Biospheres* 51, 167-183.

Matzov D, Taoka M, Nobe Y, Yamauchi Y, Halfon Y, Asis N, Zimermann E, Rozenberg H, Bashan A, Bhushan S, Isobe T, Gray MW, Yonath A & Shalev-Benami M (2020). Cryo-EM structure of the highly atypical cytoplasmic ribosome of Euglena gracilis. *Nucleic Acids Res* 48, 11750-11761.

Maynard-Smith J (1983). Models of evolution. *Proc Roy Soc Lond* 219, 315-325.

McClain WH, Guerrier-Takada C & Altman S (1987). Model substrates for an RNA enzyme. *Science* 238, 527-530.

Meng K, Chung CZ, Soll D & Krahn N (2022). Unconventional genetic code systems in archaea. *Frontiers Microbiol* ##, 1-18.

Miller SL (1953). A production of amino acids under possible primitive earth conditions. *Science* 117, 528-529.

Mills DR, Peterson RL & Spiegelman S (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci USA* 58, 217-224.

Mizuuchi R & Ichihashi N (2021). Primitive compartmentalization for the sustainable replication of genetic molecules. *Life* 11, 1-17.

Moore PB & Steitz TA (2002). The invovlement of RNA in ribosome function. Nature 418, 229-235.

Murakami H, Ohta A, Ashigai H & Suga H (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* 3, 357-359.

Mutscher H, Wochner A & Holliger P (2015). Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nature Chem* 7, 502-508.

Nagaswamy U & Fox GF (2003). RNA ligation and the origin of tRNA. Origins Life Evol Biosphere 33, 199-209.

Natsidis P, Schiffer PH, Salvador-Martinez I & Telford MJ (2019). Computational discovery of hidden breaks in 28S ribosomal RNAs across eukaryotes and consequences for RNA Integrity Numbers. *Nature Scientific Reports* 9, 19477.

Nelsesteun GL (1980). Origin of life: consideration of alternatives to proteins and nucleic acids. J Mol Evol 15, 59-72.

Nelson JW & Breaker RR (2017). The lost language of the RNA World. *Sci Signal* 10, 1-10.

Nielsen PE, Egholm M, Berg RH & Buchardt O (1991). Sequence-selective recognition of DNA by strand displacement with a thymine-substitute polyamide. *Science* 254, 1497-1500.

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

Nissen P, Hansen J, Ban N, Moore PB & Steitz TA (2000). The structural basis of ribosome activity in peptide bond synthesis. Science 289, 920-930.

Nissen P, Ippolito JA, Ban N, Moore PB & Steitz TA (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proc Natl Acad Sci USA 98, 4899-4903.

Noller HF, Hoffarth V & Zimniak L (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256, 1416-1419.

Nutman AP, Bennett VC, Friend CRL, Van Kranendork MJ & Chivas AR (2016). Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* 537, 535-538.

Oba Y, Takano Y, Furukawa Y, Koga T, Glavin DP, Dworkin JP & Naraoka H (2022). Identifying the wide diversity of extraterrestrial purine and pyrimidine nucleobases in carbonaceous meteorites. *Nature Comm* 13, 1-10.

Orellana O, Cooley L & Soll D (1986). The additional guanylate at the 5' terminus of Escherichia coli $tRNA^{His}$ is the result of unusual processing by RNase P. *Mol Cell Biol* 6, 525-529.

Orgel LE (1968). Evolution of the genetic apparatus. *J Mol Biol* 38, 381-393.

Orgel LE & Crick FHC (1980). Selfish DNA : the ultimate parasite. *Nature* 284, 604-607.

Oro J (1961). Mechanism of synthesis of adenine from hydrogen cyanide under plausible primitive earth conditions. *Nature* 191, 1193-1194.

Pace NR & Marsh TL (1985). RNA catalysis and the origin of life. *Origins Life* 16, 97-116.

Pearce BKD, Tupper AS, Pudritz RE & Higgs PG (2018). Constraining the time interval for the origin of life on earth. *Astrobiol* 18, 343-364.

Phan HD, Lai LB, Zahurancik WJ & Gopalan V (2021). The many faces of RNA-based RNase P, an RNA-world relic. *Trends Biochem Sci* 46, 976-991.

Polacek N & Mankin AS (2005). The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Critical Reviews Biochem Mol Biol* 40, 285-311.

Polikanov YS, Steitz TA & Innis CA (2014). A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation of the ribosome. *Nature Structural Mol Biol* 21, 787-793.

Poole AM, Jeffares DC & Penny D (1998). The path from the RNA world. *J Mol Evol* 46, 1-17.

Powner MW, Gerland B & Sutherland JD (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459, 239-242.

Randau L & Soll D (2008). Transfer RNA genes in pieces. *EMBO Rep* 9, 623-628.

Rich A (1962). On the problems of evolution and biochemical information transfer. In *Horizons in Biochemistry* (ed Kasha M & Pullman B). New York: Academic Press. 103-126.

Ring D, Wolman Y, Friedmann N & Miller SN (1972). Prebiotic synthesis of hydrophobic and protein amino acids. *Proc Natl Acad Sci* 69, 765-768.

Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BFC & Klug A (1974). Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature* 250, 546-551.

Rogalski M, Karcher D & Bock R (2008). Superwobbling facilitates translation with reduced tRNA sets. *Nature Structural Mol Biol* 15, 192-198.

Rubio Gomez MA & Ibba M (2020). Aminoacyl-tRNA synthetases. *RNA* 26, 910-936.

Schimmel P & Ribas de Pouplana L (1995). Transfer RNA: from minihelix to genetic code. Cell 81, 983-986.

Schwartz AW & Orgel LE (1985). Template-directed synthesis of novel, nucleic acid-like structures. Science 228, 585-587.

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Balleannee C, Kucukural A, Serra RW, Sun F, et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 351, 391-396.

Sharp PA (1985). On the origin of RNA splicing and introns. *Cell* 42, 397-400.

Stueken EE, Anderson RE, Bowman JS, Brazelton WJ, Colangelo-Lillis J, Goldman AD, Som SM & Baross JA (2013). Did life originate from a global chemical reactor? *Geobiology* 11, 101-126.

Sulston J, Lohrmann R, Orgel LE & Miles HT (1968). Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc Nat Acad Sci USA* 59, 726-733.

Sutherland JD (2017). Studies on the origin of life - the end of the beginning. *Nature Rev Chem* 1, 1-7.

Syroegin EA, Aleksandrova EV & Polikanov YS (2023). Insights into the ribosome function from the structure of non-arrested ribosome-nascent chain complexes. *Nature Chem* 15, 143-153.

Szostak JW (2012). The eightfold path to non-enzymatic RNA replication. *J Syst Chem* 3, 1-14.

Szathmary E (1999). The origin of the genetic code: amino acids as cofactors in an RNA world. Trends Genet 15, 223-229.

Tallsjo A & Kirsebom LA (1993). Product release is a rate-limiting step during cleavage by the catalytic RNA subunit of *Escherichia coli* RNase P. *Nuc Acids Res* 21, 51-57.

Tamura K & Schimmel P (2003). Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proc Natl Acad Sci* 100, 8666-8669.

Tamura K & Schimmel P (2006). Chiral-selective aminoacylation of an RNA minihelix: mechanistic features and chiral suppression. *Proc Natl Acad Sci USA* 103, 13750-13752.

Tamura K (2015). Origins and early evolution of the tRNA molecule. *Life* 5, 1687-1699.

Temin HM & Mizutani S (1970). RNA-dependent DNA polymerase in virions of rous sarcoma virus. *Nature* 226, 1211-1213.

Tinoco I & Bustamante C (1999). How RNA folds. *J Mol Biol* 293, 271-281.

Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ & Lehman N (2012). Spontaneous network formation among cooperative RNA replicators. *Nature* 491, 72-77.

Vicens Q & Kieft JS (2022). Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad Sci USA* 119, 1-9.

Visser CM (1984). Evolution of biocatalysts 1. Possible pre-genetic-code RNA catalysts which are their own replicase. *Orig Life* 14, 291-300.

Voorhees RM & Hedge RS (2015). Structure of the scanning and engaged states of the mammalian SRP-ribosome complex. Elife 4, e07975.

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

Watson JD & Crick FHC (1953). Genetical implications of the structure of DNA. *Nature* ##, 964-967.

Weber AL (1989). Model of early self-replication based on covalent complmentarity for a copolymer of glycerate-3-phosphate and glycerol-3-phosphate. Origins Life 19, 179-186.

Weiner AM & Maizels N (1987). tRNA-like structures tag the 3' ends of genomic RNA molecules for replication : Implications for the origin of protein synthesis. *Proc Acad Sci USA* 84, 7383-7387.

Weiss R & Cherry J (1993). Speculations on the origin of ribosomal translocation. *Cold Spring Harbor Monograph Series* 24, 71-89.

Welting TJM, Kikkert BJ, Van Venrroij WJ & Pruijn GJM (2006). Differential association of protein subunits with human RNase MRP and RNase P complexes. *RNA* 12, 1373-1382.

Westheimer FH (1987). Why nature chose phosphate. *Science* 235, 1173-1178.

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

Widmann J, Di Giulio M, Yarus M & Knight R (2005). tRNA creation by hairpin duplication. J Mol Evol 61, 524-530.

Woese C (1967). *The Origins of the Genetic Code*. New York: Harper & Row.

Wolfenden R, Lu X & Young G (1998). Spontaneous hydrolysis of glycosides. *J Am Chem Soc* 120, 6814-1815.

Wolfenden R (2011). Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu Rev Biochem* 80, 645-667.

Wu T & Orgel LE (1992). Nonenzymatic template-directed synthesis on hairpin oligonucleotides. J Am Chem Soc 114, 5496-5501.

Wu LF & Sutherland JD (2019). Provisioning the origin and early evolution of life. *Emerging Topics in Life Sciences* 3, 459-468.

Yakhnin AV (2007). A model for the origin of protein synthesis as coreplicational scanning of nascent RNA. *Orig Life Evol Biosph* 37, 523-536.

Zamecnik PC (1960). Historical and current aspects of the problem of protein synthesis. *Harvey Lectures* 1958-1959, 250-281. New York: Academic Press.

Zaug AJ & Cech TR (1986). The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* 231, 470-475.

Zhang B & Cech TR (1997). Peptide bond formation by *in vitro* selected ribozymes. *Nature* 390, 96-100.

Zhang J & Ferre-DAmare AR (2016). Trying on tRNA for size: RNase P and the T-box riboswitch as molecular rulers. *Biomolecules* 6, 1-14.

Zhang SJ, Duzdevich D, Ding D & Szostak JW (2022). Freeze-thaw cycles enable a prebiotically plausible & continuous pathway from nucleotide activation to nonenzymatic RNA copying. *Proc Natl Acad Sci USA* 119, 1-7.

Zhou L, Kim SC, Ho KH, O'Flaherty DK, Giurgiu C, Wright TH & Szostak JW (2019). Non-enzymatic primer extension with strand displacement. *eLife* 8, 1-14.

Zhu J, Huang W, Zhao J, Huynh L, Taylor DJ & Harris ME (2022). Structural and mechanistic basis for recognition of alternative tRNA precursor substrates by bacterial ribonuclease P. *Nature Comm* 13, 1-13.

Zuo Z, Peng D, Yin X, Zhou X, Cheng H & Zhou R (2013). Genome-wide analysis reveals origin of transfer RNA genes from tRNA halves. *Mol Biol Evol* 30, 2087-2098.

THE DUMPING GROUND

Poole AM & Logan DT (2005). Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome? Mol Biol Evol 22, 1444-

Noller HF (2004). The driving force for molecular evolution of translation. RNA 10, 1833-

Doudna JA & Lorsch LR (2005). Ribozyme catalysis: not different, just worse. Nat Struct Mol Biol, 395-

Doudna JA & Szostak JW (1989). RNA-catalysed synthesis of complementary-strand RNA. Nature 339, 519-522.

Green R & Szostak JW (1992). Selection of a ribozyme that functions as a superior template in a self-copying reaction. Science 258, 1910-1915.

Bartel DP & Szostak JW (1993). Isolation of new ribozymes from a large pool of random sequences. Science 261, 1411-1418.

Ekland EH & Bartel DP (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature 382, 373-376.

Attwater J, Raguram A, Morgunov AS, Gianni E & Holliger P (2018). Ribozyme-catalysed RNA synthesis using triplet building blocks. eLife 7, 1-25.

Cojocaru R & Unrau PJ (2021). Processive RNA polymerization & promotor recognition in an RNA World. Science 371, 1225-1232.

Wachowius F & Holliger P (2021). RNA replication and the RNA polymerase ribozyme. In Ribozymes (edited Muller S, Masquida B & Winkler W) Wiley-VCH.

McRae EKS, Wan CJK, Kristoffersen EL, Hansen K, Gianni E, Gallego I, Curran JF, Attwater J, Holliger P & Andersen ES (2022). Cryo-EM structure and functional landscape of an RNA polymerase ribozyme. bioRxiv

Lorenz MR et al (2017). Proton gradients and pH oscillations emerge from heat flow at the microscale. Nat Commun 8, 1897.

Atkins JF, Herr AJ, Massire C, O'Connor M, Ivanov I & Gesteland RF (2000). Poking a hole in the sanctity of the triplet code: inferences for framing. In The Ribosomes: Structure, Function, Antibiotics, and Cellular Interactions (editors Garrett RA, Douthwaite SR, Liljas A, Matheson AT, Moore PB & Noller HF) 369-383.

Wilson DN & Nierhaus KH (2006). The E-site story: the importance of maintaining two tRNAs on the ribosome during protein synthesis. Cell Mol Life Sci 63, 2725-2737.Bjork GR, Wikstrom PM & Bystrom AS (1989). Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. Science 244, 986-989.

Carlson BA, Kwon SY, Chamorro M, Oroszlan S, Hatfied DL & Lee BJ (1999). Transfer RNA modification status influences retroviral ribosomal frameshifting. Virology 255, 2-8.

Demeshkina N, Jenner L, Yusupova G & Yusupov M (2010). Interactions of the ribosome with mRNA & tRNA. Current Opinion Struc Biol 20, 325-332.

Rozov A, Demeshkina N, Khusainov I, Westhof E, Yusupov M & Yusupova G (2016). Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. Nature Comm 7, 1-10.

Zhou J, Lancaster L, Donohue JP & Noller HF (2019). Spontaneous ribosomal translocation of mRNA and tRNAs into a chimeric hybrid state. Proc Natl Acad Sci USA 116, 7813-7818.

Djumagulov M, Demeshkina N, Jenner L, Rozov A, Yusupov M & Yusupova G (2021). Accuracy mechanism of eukaryotic ribosome translocation. Nature 600, 543-546.

Kisonaite M, Wild K, Lapouge K, Ruppert T & Sinning I (2022). High-resolution structures of a thermophilic eukaryotic 80S ribosome reveal atomistic details of translocation. Nature Comm 13, 1-12.

Poulis P (2022). Ribosome dynamics during spontaneous frameshifting. Doctoral Dissertation, Univ Gottingen.

Poulis P, Patel A, Rodnina MV & Adio S (2022). Altered tRNA dynamics during translocation on slippery mRNA as determinant of spontaneous ribosome frameshifting. Nature Comm 13, 4231-####.

Woese CR (1967). The Genetic Code. NY : Harper & Row.

Martínez-Giménez JA, Saez GT & Tabarés-Seisdedos R (1998). On the function of modified nucleosides in the RNA world. J theor Biol 194, 485-490.

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

Tanaka T & Kikuchi Y (2001). Origin of the cloverleaf shape of transfer RNA - the double-hairpin model : implication for the role of tRNA intron and the long extra loop. Viva Origino 29, 134-142.

Kanai A (2015). Disrupted tRNA genes and tRNA fragments : a perspective on tRNA gene evolution. Life 5, 321-331.

Maracci C & Rodnina MV (2016). Translational GTPases. Biopolymers 105, 463-475.

Koonin EV (2017). Frozen accident pushing 50: stereochemistry, expansion, and chance in the evolution of the genetic code. Life 7, 1-13.

Koonin EV, Makarova KS, Wolf YI & Krupovic M (2020). Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nature Rev* 21, 119-131.

Jacome R, Campillo-Balderas JA, Becerra A & Lazcano A (2022). Structural analysis of monomeric RNA-dependent polymerases revisited. J Mol Evol 90, 283-295.

Cojocaru R & Unrau PJ (2021). Phosphoryl transfer ribozymes. In Ribozymes … 331-358.

Zaug AJ & Cech TR (1986). The Tetrahymena intervening sequence ribonucleic acid enzyme is a phosphotransferase and an acid phosphatase. Biochemistry 25, 4478-4482.

Garside EL, Kent OA & MacMillan AM (2021). The spliceosome: an RNA-protein ribozyme derived from ancient mobile genetic elements. In Ribozymes … 169-191.

Chan CW, Chetnani B & Mondragon A (2013). Structure and function of the T-loop structural motif in noncoding RNAs. *WIREs RNA* 4, 507-522.

Lehmann J, Jossinet F & Gautheret D (2013). A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNAse P and T-box leaders. *Nucleic Acids Res* 41, 5495-5502.

Paisley TE & Van Tuyle GC (1994). The processing of wild type and mutant forms of rat nuclear pre-tRNALys by homologous RNase P. *Nucl Acids Res* 22, 3347-3353.

Pettersson BMF, Ardell DH & Kirsebom LA (2005). The length of the 5' leader of Escherichia coli tRNA precursors influences bacterial growth. J Mol Biol 351, 9-15.

Kerkhofs K, Garg J, Fafard-Couture, Elela SA, Scott MS, Pearlman RE & Bayfied MA (2022). Altered tRNA processing is linked to a distinct and unusual La protein in Tetrahymena thermophila. Nature Comm 13, 1-17.

Persson T, Cuzic S & Hartmann RK (2003). Catalysis by RNase P RNA: unique features and unprecedented active site plasticity. *J Biol Chem* 278, 43394-43401.

Sun L, Campbell FE, Zahler NH & Harris ME (2006). Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P. EMBO J 25, 3998-4007.

Wilusz JE (2016). Long noncoding RNAs: re-writing dogmas of RNA processing and stability. Biochim Biophys Acta 1859, 128-138.

most of anticodon loop bases form a 3’ stack

3 : 4 stacks 32 33 34 35 36 37 38 abnormal

2 : 5 stacks 32 33 34 35 36 37 38 normal

1 : 6 stacks 32 33 34 35 36 37 38 abnormal

3 : 4 stacks??? yields the wobble position in P-site????

inosine/hypoxanthine deamination of adenosine/adenine

I pairs best to C then others

dC to dU 200 / day in mammalian cels

dA to dI at 2-3% of this rate

deaminations in dsDNA are 0.5-0.7% those in ssDNA

[FOOTNOTE] purine only nucleic acids / code A.I pairing (Crick 1968)

[alarmones] In the bacterial stringent response to amino acid starvation and stress, RelA/Rel proteins detect uncharged tRNAs in ribosome A-site and make alarmones (p)ppGpp.

alarmones inhibit translational GTPases

bacterial (p)ppGpp alarmones bind SRP and SR GTPases and block docking (Czech et al 2022).

Cells generally have one aaRS for each of the canonical 20 amino acids, one SepRS for tRNA-Sec that assigns selenocysteine at UGA, and in many archaea and some bacteria, one PylRS for tRNA-Pyl that assigns pyrrolysine at UAG. Finally, there are two different LysRSs, one of each class, though few organisms have both LysRSs. …. tRNAGln and tRNASec are unusual in that they are loaded by ND GluRS and SepRS, respectively, and the last steps convert the amino acid occur on the esterified tRNA (Meng et al 2020).

B. subtilis one GluRS charges Glu-tRNA(Glu) and Glu-tRNA(Gln); specific amidotransferase converts Glu-tRNA(Gln) to Gln-tRNA(Gln) on the tRNA; similarly for Asp and Asn

SerRS charges tRNASec with serine … bacteria convert ser to sec in one step & then SelB delivers to ribosome

archaea/eukarya two steps from Ser to Sec

archaea PylRS related to PheRS class II

Schimmel et al (1993). operational RNA code

Arnez & Moras (1997). Structural and functional considerations of the aminoacylation reaction.

(Syroegin et al 2023) four hydrogen bonds between rRNA and three C-terminal residues of nascent peptide plus one hydrogen bond within the nascent peptide

-1 residue carbonyl - G2061

-1 residue amide - A2062

-2 residue carbonyl - N3 of U2506

-2 residue amide - O4 of U2506

amide -0 residue - carbonyl -2 residue

carbonyl -0 residue - water W2 - A2602

The first accessory proteins of RNase P apparently arose after the split of bacteria and archaea.

[Alu retrotransposon]

In primates Alu retrotransposons arose from the 7SL RNA through deletion of the central S domain, followed by tandem duplication of the Alu domain and acquisition of the polyA tail.

[5S RNA] uL18/uL5 = Rpl5/Rpl11 respectively - what a nightmare!

The common B-form of DNA duplex has two anti-parallel right-handed helices with 10 base pairs per turn, and a rise of 3.38 A per base pair.

DNA forms two different anti-parallel right-handed double helices, the common B-form and the rare A-form, depending on the pucker of the furanose ring (Dickerson et al 1982; Saenger 1984). The B-form has 10 base pairs per turn, with a rise of 3.38 A per base pair.

base pairs nearly perpendicular to the helix axis, *anti*-conformation for all glycosyl torison angles, C2'-*endo* sugar pucker. 20 A diameter, major and minor grooves equally deep, major wider than minor; helix axis passes straight through the base pairs

[stout and hollow] Because the 2' OH constrains the ribofuranose ring to C3'-*endo* pucker, RNA duplexes generally adopt the right-handed A-form double helix with 11 bp

per turn, a rise of 2.55 A per base pair, an inclination of the base pairs of ~19 degrees aginst the helix, and anti-conformation glycosyl torsion angles. A diameter of ~23 A, deep narrow major goove, pushing the base pairs away from the helix, and shallow minor groove.

wide + shallow minor groove, edge of nucleobase accessible

narrow + deep major groove, access hindered

Because the 2' OH constrains the ribofuranose ring to C3'-*endo* pucker, RNA duplexes generally adopt the right-handed A-form double helix… the 2' OHs line up in the minor groove

recycling?? EF-G and/or HflX

To gather translational GTPases near the factor binding site, proteins of the P-stalk.

SRL sarcin-ricin loop

[L1 stalk] H76 attached to a 3-way junction H75 H76 H79 … swings between open and closed to release tRNA from E-site

POL III small ncRNAs

() transciption 7SK indirect inhibitor of POL II

() RNA processing U6 RNA (active site of splicosome), RNaseP RNA, RMRP RNA

() RNA localization vault RNA nuclear pore complex

() Y RNA binds ring-shaped protein Ro60 that binds misfolded ncRNAs and pre-5S rRNA

() translation 7S RNA = SRP RNA

() primate BC200 RNA /rodent Bc1 RNA regulate translation in neuron dendrites