
The Molecular Evolution of Complexity at the Origin of Life

Nisha Dhar

A thesis submitted to the Faculty of Health Sciences, University of the
Witwatersrand, in fulfilment of the requirements for the degree of

Doctor of Philosophy

Johannesburg, 2016

DECLARATION

I, Nisha Dhar declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

.....

...10th...day of.....June....., 2016

PUBLICATIONS AND PRESENTATIONS RELATED TO THESIS

Publications

1. Dhar N., Weinberg M., Michod R., Durand P.M. Ribozyme plasticity and molecular trade-offs can account for complexity at the origin of life (Manuscript in preparation).
2. Dhar N., Weinberg M., Michod R., Durand P.M. A conceptual model for network stability at the origin of life (Manuscript in preparation).
3. Durand P.M. and Dhar N. The evolutionary ecology of the first living system. In: “The evolutionary origins of life and death” (Durand P.M.). University of Chicago press. In preparation (accepted).

Conference Proceedings

1. Dhar N., Weinberg M., Michod R.E., Durand P.M. Ribozyme plasticity and molecular trade-offs can account for increasing complexity and network stability at the origin of life, Origins of Life (Gordon Research Conference), Jan 17-22, 2016, Galveston, Texas, USA. **Awarded GRC Carl Storm International Diversity Award.**
2. Dhar N., Weinberg M., Michod R.E., Durand P.M. Ribozyme plasticity and molecular trade-offs can account for increasing complexity and network stability at the origin of life, Evolution 2014, June 20-24, 2014, Raleigh, North Carolina, USA (Presented by Dr. P.M. Durand)

3. Dhar N., Weinberg M., Michod R.E., Durand P.M. Evolution of Complexity at the origin of life, Cooperation and Major Evolutionary Transitions, The Kavli Institute for Theoretical Physics, Feb 4-8, 2013, Santa Barbara, California, USA.

4. Dhar N., Weinberg M., Michod R.E., Durand P.M. Cooperation and Conflict at the Origin of Life, SASBMB/FASBMB Congress 2012, Jan 29-Feb 1, 2012, Drakensberg, South Africa. **Awarded second best poster prize.**

Other Presentations

Dhar N., Weinberg M., Michod R.E., Durand P.M. Ribozyme plasticity and molecular trade-offs can account for increasing complexity and network stability at the origin of life, Wits Health Sciences Research Day 2014, 17 September 2014, University of the Witwatersrand, South Africa (Published in South African Journal of Infectious Diseases. Website link: <http://www.sajei.co.za>).

Dhar N., Weinberg M., Michod R.E., Durand P.M. Cooperation and Conflict at the Origin of Life, Wits Health Sciences Research Day 2012, 19 September 2012, University of the Witwatersrand, South Africa.

ABSTRACT

A widely accepted hypothesis for the origin of life is that it was based on catalytic RNA or ribozymes (the RNA world hypothesis). In this paradigm, one of the earliest and essential functions for an RNA based life to emerge was polymerisation. Although polymerisation activity has been demonstrated in extant and engineered ribozymes, these molecules are large and too complex to have formed spontaneously in the prebiotic world. Furthermore, the evolution and stability of RNA based life would have required a pool of diverse complex ribozymes. An understanding of the basic mechanistic processes implicated in the emergence of a minimal polymerase and diverse complex molecules from small oligomers remains a major gap. This project examined the ligation activity of a polymerase and its smaller derivatives with random oligonucleotide substrates and revealed how the molecular dynamics of ligation would have affected the evolution of complexity in the early stages of an RNA world.

The size and structural complexity of a minimal polymerase (called R18 polymerase ribozyme) was reduced in a stepwise fashion. All RNA constructs were examined for self-ligation function with 24 random oligonucleotide substrates (each 35 nucleotides long) in the absence of experimentally designed base pairing. The smallest element (40 nucleotides long) was able to non-specifically ligate substrates to its own end, however, with low efficiency. A gradual increase in specificity for the substrates and overall functional efficiency was observed with an increase in structural complexity of the ribozymes. The most complex R18 polymerase ligated only selected substrate variants to itself, although with much greater efficiency than the smaller constructs. These findings suggest that the complexity in a primitive molecular system increased in a modular fashion via ligases. Furthermore, general compatibility of the ligases with the substrates was a

mechanism for increase in the molecular complexity and functionality. The inverse correlation between functional flexibility and efficiency with increase in structural complexity of the catalysts points to a molecular trade-off. In the ecology of the RNA world, this molecular trade-off would have been central to ribozyme population stability and for the development of functional specialisation. The findings in this project point to a form of hypercycle composed of a complementary set of processes stabilised by inherent molecular trade-offs. Such a hypercycle is suggested to facilitate the emergence of a stable molecular network and a replicative unit essential for life to begin.

ACKNOWLEDGEMENTS

1. I would like to thank my supervisor, Dr. Pierre Durand, for his valuable support and guidance during the project. I would also like to appreciate his faith and optimism which provided me strength to cope with difficult times in my research work.
2. I would like to thank my co-supervisor, Prof. Marco Weinberg for his support and guidance.
3. I would like to thank Dr. Abdullah Ely for his kind assistance on some practical aspects of my project. I would also like to thank him for his encouragement during the research work.
4. I would like to thank Prof. Theresa Coetzer and Prof. Patrick Arbuthnot for generously providing some of their lab facilities.
5. I would like to thank all my lab members and friends for their support. Special thanks to Dr. Kubendran Naidoo and Dr. Sonja Lauterbach.
6. Finally, I would like to thank the following funding bodies for financial support, Wits University Postgraduate Merit Award, Wits University local merit scholarship, Wits University Medical Research Endowment Fund, National Research Foundation, and Medical Research Council.

TABLE OF CONTENTS

DECLARATION	ii
PUBLICATIONS AND PRESENTATIONS RELATED TO THESIS	iii
ABSTRACT	v
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES	xii
LIST OF TABLES	xvii
LIST OF ABBREVIATIONS	xix
LIST OF SYMBOLS	xxi
TERMINOLOGY	xxii
1. INTRODUCTION	1
1.1 Ribozymes and the origin of life.....	2
1.1.1 Origin of RNA building blocks	5
1.1.1.1 Formation of Nucleobases	5
1.1.1.2 Formation of Ribose sugar.....	6
1.1.1.3 Formation of Nucleosides	7
1.1.1.4 Formation of Nucleotides	8
1.1.1.5 RNA as a product of the pre-RNA world	9
1.1.2 Abiotic synthesis of RNA oligomers.....	9
1.1.3 Non-enzymatic replication of RNA oligomers.....	11
1.1.4 Replication and stability of catalytic RNA polymers.....	12
1.2 Emergence of complexity at the origin of life	15
1.2.1 Hypercycles and the emergence of complexity in the RNA world	16

Table of Contents

1.2.2	The history of engineered RNA polymerases.....	21
1.2.3	Question 1: How could a minimal polymerase emerge from short molecules? 24	
1.2.4	Question 2: How could complex diverse molecules emerge from short molecules?.....	29
1.2.5	The proposed argument	30
1.2.5.1	The possible role of small ribozymes	31
1.2.5.2	The possible role of unconstrained ligation reactions.....	32
2.	MATERIALS AND METHODS	36
2.1	Preparation of Round 18 Polymerase ribozyme and truncated derivatives	36
2.1.1	Preparation of template DNA for <i>in vitro</i> transcription of RNAs	36
2.1.2	In vitro transcription and purification of RNA products	37
2.2	Design of the Oligonucleotide substrates	38
2.3	Ribozyme self-ligation assay	42
2.4	Sequence analysis of the ligation product.....	46
2.5	Analysis of the substrates with respect to the ligation activity of RNA	46
2.6	Structures and complexity predictions.....	47
2.7	Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of ribozyme self-ligation activity.....	47
2.7.1	Design of Probe	48
2.7.2	Preparation of samples for standard curves.....	48
2.7.3	Real-time quantitative PCR of samples and generation of standard curves..	49
2.7.4	Determination of the rate of ribozyme self-ligation activity	50
3.	RESULTS.....	55
3.1	Overview.....	55

Table of Contents

3.2	RNA preparation for Self-ligation assays	55
3.3	Self-Ligation activity of RNAs	57
3.3.1.	Self-Ligation activity of R18 RNA	59
3.3.2.	Self-Ligation activity of R18-T1 RNA.....	59
3.3.3.	Self-Ligation activity of R18-T2 RNA.....	59
3.3.4.	Self-Ligation activity of R18-T3 RNA.....	60
3.3.5.	Self-Ligation activity of R18-T4 RNA.....	60
3.4	Summary of RNA activity with the oligonucleotide substrates	66
3.5	Analysis of sequence patterns in the substrate sequences	69
3.5.1	Substrates sequence analysis for R18-T4 and R18-T3 ribozyme activity.....	71
3.5.2	Substrates sequence analysis for R18-T2 ribozyme activity.....	74
3.5.3	Substrates sequence analysis for R18-T1 and R18 ribozyme activity.....	74
3.6	Rate of self-ligation activity of the ribozymes.....	79
3.6.1	Comparison of the rates of ribozyme self-ligation activity with different substrates	80
3.6.2	Dynamics of the rates of ribozyme self-ligation activity with increase in their complexity.....	81
3.7	Correlation between biochemical traits with increase in ribozyme complexity ...	95
4.	DISCUSSION	102
4.1	Self-ligation function in polymerase and its smaller components	104
4.2	Self-ligation reactions under no experimentally designed pairing	105
4.3	Flexibility of self-ligation function in early RNA molecules	106
4.4	Analysis of the substrates for the self-ligation activity of the ribozymes.....	108
4.5	Rate of self-ligation activity	110

Table of Contents

4.6	Correlations and trade-offs between molecular traits and the implications for the origin of life	112
4.7	Conceptual model for network stability at origin of life.....	115
5.	CONCLUSION.....	124
6.	FUTURE DIRECTIONS	125
7.	APPENDIX A	126
A.1	Recipe for 10x TBE solution	126
A.2	Recipe for preparing 10% polyacrylamide gel	127
A.3	Recipe for preparing 8% polyacrylamide-8M urea gel.....	127
A.4	Protocol for <i>in vitro</i> transcription of RNA using the Megashortscript T7 transcription kit.....	128
A.5	Preparation of RNA samples before loading on gel	128
A.6	Gel electrophoresis	128
8.	APPENDIX B.....	130
B.1	Electro-elution protocol for purification of RNAs.....	130
9.	APPENDIX C	132
C.1	Confirmation of self-ligation activity of R18 RNAs by sequence analysis.....	132
C.2	Confirmation of self-ligation activity of R18-T1 RNAs by sequence analysis..	136
C.3	Confirmation of self-ligation activity of R18-T2 RNAs by sequence analysis..	140
C.4	Confirmation of self-ligation activity of R18-T3 RNAs by sequence analysis..	144
C.5	Confirmation of self-ligation activity of R18-T4 RNAs by sequence analysis..	149
10.	APPENDIX D	153
11.	APPENDIX E.....	159
12.	ETHICS WAIVER CERTIFICATE	164
13.	REFERENCES	165

LIST OF FIGURES

Figure 1.1: Time line of events for the evolution of an RNA world.	5
Figure 1.2: The various scenarios of Hypercycles.	20
Figure 1.3: Structures of the engineered RNA polymerases.	23
Figure 1.4: A model proposed for the evolution of a RNA polymerase.	25
Figure 1.5: A model based on a stepwise process for the emergence of RNA polymerase.	28
Figure 2.1: Synthesis of the template DNA sequence for R18 polymerase.	39
Figure 2.2: Synthesis of R18 polymerase and its truncated RNA sequences.	40
Figure 2.3: Schematic representation of the ribozyme self-ligation assay.	43
Figure 2.4: Schematic representation of the assay performed for quantitative analysis of the ribozyme's ligation activity.	53
Figure 2.5: Comparison of the fluorescence from ribozyme self-ligation assay and the corresponding standard sample.	54
Figure 3.1: Purified RNAs used for self-ligation assays.	56
Figure 3.2: Self-ligation activity of R18 RNA.	61
Figure 3.3: Self-ligation activity of R18-T1 RNA.	62
Figure 3.4: Self-ligation activity of R18-T2 RNA.	63
Figure 3.5: Self-ligation activity of R18-T3 RNA.	64
Figure 3.6: Self-ligation activity of R18-T4 RNA.	65
Figure 3.7: Sequence logo of all the substrates used for MEME analysis.	70
Figure 3.8: Time course analysis of R18-T4 ribozyme self-ligation activity.	82
Figure 3.9: Time course analysis of R18-T3 ribozyme self-ligation activity.	84
Figure 3.10: Time course analysis of R18-T2 ribozyme self-ligation activity.	86
Figure 3.11: Time course analysis of R18-T1 ribozyme self-ligation activity.	88

List of Figures

Figure 3.12: Time course analysis of R18 ribozyme self-ligation activity.	90
Figure 3.13: Consistency of the ribozymes with respect to their rate of self-ligation reaction with the substrates.	92
Figure 3.14: Correlation of complexity of the ribozymes with respect to their rate of self-ligation activity with the substrates.	93
Figure 3.15: Correlation between size, structural stability and functional flexibility of ribozymes.	97
Figure 3.16: Correlation between size, structural stability and the rate of self-ligation activity of ribozymes.	98
Figure 3.17: Trend between rate of self-ligation activity and functional flexibility.	99
Figure 3.18: Correlation between size, functional flexibility and the rate of self-ligation activity of ribozymes.	100
Figure 3.19: Correlation between structural stability, functional flexibility and the rate of self-ligation activity of ribozymes.	101
Figure 4.1: Truncated constructs of the R18 polymerase	103
Figure 4.2: Schematic representation of the conceptual model for network stability at the origin of life.	123
Figure B.1: Electro elution set up for elution of RNA from gel pieces.....	131
Figure C.1.1: Sequence alignment of R18 with oligonucleotide substrate 1	132
Figure C.1.2: Sequence alignment of R18 with oligonucleotide substrate 2	133
Figure C.1.3: Sequence alignment of R18 with oligonucleotide substrate 3	133
Figure C.1.4: Sequence alignment of R18 with oligonucleotide substrate 6	134
Figure C.1.5: Sequence alignment of R18 with oligonucleotide substrate 6a.....	134
Figure C.1.6: Sequence alignment of R18 with oligonucleotide substrate 6b	135
Figure C.1.7: Sequence alignment of R18 with oligonucleotide substrate 7A.....	135

List of Figures

Figure C.2.1: Sequence alignment of R18-T1 with oligonucleotide substrate 1.....	136
Figure C.2.2: Sequence alignment of R18-T1 with oligonucleotide substrate 2.....	136
Figure C.2.3: Sequence alignment of R18-T1 with oligonucleotide substrate 3.....	137
Figure C.2.4: Sequence alignment of R18-T1 with oligonucleotide substrate 4.....	137
Figure C.2.5: Sequence alignment of R18-T1 with oligonucleotide substrate 6.....	138
Figure C.2.6: Sequence alignment of R18-T1 with oligonucleotide substrate 6a.....	138
Figure C.2.7: Sequence alignment of R18-T1 with oligonucleotide substrate 6b.....	138
Figure C.2.8: Sequence alignment of R18-T1 with oligonucleotide substrate 7A.....	139
Figure C.2.9: Sequence alignment of R18-T1 with oligonucleotide substrate 7.....	139
Figure C.2.10: Sequence alignment of R18-T1 with oligonucleotide substrate 8B.....	139
Figure C.3.1: Sequence alignment of R18-T2 with oligonucleotide substrate 1.....	140
Figure C.3.2: Sequence alignment of R18-T2 with oligonucleotide substrate 2.....	140
Figure C.3.3: Sequence alignment of R18-T2 with oligonucleotide substrate 3.....	141
Figure C.3.4: Sequence alignment of R18-T2 with oligonucleotide substrate 4.....	141
Figure C.3.5: Sequence alignment of R18-T2 with oligonucleotide substrate 5.....	141
Figure C.3.6: Sequence alignment of R18-T2 with oligonucleotide substrate 6a.....	142
Figure C.3.7: Sequence alignment of R18-T2 with oligonucleotide substrate 6b.....	142
Figure C.3.8: Sequence alignment of R18-T2 with oligonucleotide substrate 7A.....	142
Figure C.3.9: Sequence alignment of R18-T2 with oligonucleotide substrate 7.....	143
Figure C.3.10: Sequence alignment of R18-T2 with oligonucleotide substrate 7B.....	143
Figure C.3.11: Sequence alignment of R18-T2 with oligonucleotide substrate 8A.....	143
Figure C.3.12: Sequence alignment of R18-T2 with oligonucleotide substrate 8.....	144
Figure C.4.1: Sequence alignment of R18-T3 with oligonucleotide substrate 1.....	144
Figure C.4.2: Sequence alignment of R18-T3 with oligonucleotide substrate 2.....	145
Figure C.4.3: Sequence alignment of R18-T3 with oligonucleotide substrate 3.....	145

List of Figures

Figure C.4.4: Sequence alignment of R18-T3 with oligonucleotide substrate 4.....	145
Figure C.4.5: Sequence alignment of R18-T3 with oligonucleotide substrate 5.....	146
Figure C.4.6: Sequence alignment of R18-T3 with oligonucleotide substrate 6.....	146
Figure C.4.7: Sequence alignment of R18-T3 with oligonucleotide substrate 6A.....	146
Figure C.4.8: Sequence alignment of R18-T3 with oligonucleotide substrate 6b.....	147
Figure C.4.9: Sequence alignment of R18-T3 with oligonucleotide substrate 7A.....	147
Figure C.4.10: Sequence alignment of R18-T3 with oligonucleotide substrate 7.....	147
Figure C.4.11: Sequence alignment of R18-T3 with oligonucleotide substrate 7B	148
Figure C.4.12: Sequence alignment of R18-T3 with oligonucleotide substrate 8A.....	148
Figure C.4.13: Sequence alignment of R18-T3 with oligonucleotide substrate 8B	148
Figure C.5.1: Sequence alignment of R18-T4 with oligonucleotide substrate 1.....	149
Figure C.5.2: Sequence alignment of R18-T4 with oligonucleotide substrate 2.....	149
Figure C.5.3: Sequence alignment of R18-T4 with oligonucleotide substrate 3.....	149
Figure C.5.4: Sequence alignment of R18-T4 with oligonucleotide substrate 4.....	150
Figure C.5.5: Sequence alignment of R18-T4 with oligonucleotide substrate 5.....	150
Figure C.5.6: Sequence alignment of R18-T4 with oligonucleotide substrate 6.....	150
Figure C.5.7: Sequence alignment of R18-T4 with oligonucleotide substrate 6a.....	150
Figure C.5.8: Sequence alignment of R18-T4 with oligonucleotide substrate 6b.....	151
Figure C.5.9: Sequence alignment of R18-T4 with oligonucleotide substrate 7A.....	151
Figure C.5.10: Sequence alignment of R18-T4 with oligonucleotide substrate 7.....	151
Figure C.5.11: Sequence alignment of R18-T4 with oligonucleotide substrate 7B	151
Figure C.5.12: Sequence alignment of R18-T4 with oligonucleotide substrate 8A.....	152
Figure C.5.13: Sequence alignment of R18-T4 with oligonucleotide substrate 8B	152
Figure E.1: Standard curve for quantification of R18 ribozyme self-ligation activity.....	159
Figure E.2: Standard curve for quantification of R18-T1 ribozyme self-ligation activity.	160

List of Figures

Figure E.3: Standard curve for quantification of R18-T2 ribozyme self-ligation activity.161

Figure E.4: Standard curve for quantification of R18-T3 ribozyme self-ligation activity.162

Figure E.5: Standard curve for quantification of R18-T4 ribozyme self-ligation activity.163

LIST OF TABLES

Table 2.1: Sequences of the oligonucleotide substrates.	41
Table 2.2: Sequences of the primers used for detection of self-ligation activity of the ribozymes	45
Table 2.3: The copies of DNA prepared for generating the standard curves.	50
Table 3.1: Summary of the self-ligation activity of the RNAs with the substrates.	68
Table 3.2: Analysis of the substrates sequence pattern with respect to R18-T4 activity. ...	72
Table 3.3: Analysis of the substrates sequence pattern with respect to R18-T3 activity. ...	73
Table 3.4: Analysis of the substrate sequence pattern with respect to R18-T2 activity.	75
Table 3.5: Analysis of the substrates sequence pattern with respect to R18-T1 activity. ...	76
Table 3.6: Analysis of the substrates sequence pattern with respect to R18 activity.	77
Table 3.7: Predicted secondary structures of the substrates by mfold.	78
Table 3.8: Rates of reaction of R18-T4 ribozyme.	83
Table 3.9: Rates of reaction of R18-T3 ribozyme.	85
Table 3.10: Rates of reaction of R18-T2 ribozyme.	87
Table 3.11: Rates of reaction of R18-T1 ribozyme.	89
Table 3.12: Rates of reaction of R18 ribozyme.	91
Table 3.13: Comparison of the ribozymes rates of reaction.	94
Table 3.14: Summary of correlation between the biochemical traits of the ribozymes with increase in their complexity.	96
Table D.1: Probability matrix of all the substrates used in MEME analysis.	153
Table D.2: Probability matrix of the substrates for the analysis of R18-T4 activity.	154
Table D.3: Probability matrix of the substrates for the analysis of R18-T3 activity.	155
Table D.4: Probability matrix of the substrates for the analysis of R18-T2 activity.	156

List of Tables

Table D.5: Probability matrix of the substrates for the analysis of R18-T1 activity.....	157
Table D.6: Probability matrix of the substrates for the analysis of R18 activity.	158
Table E.1: Crossing point (CP) values of DNA copies of R18 with substrate 1.....	159
Table E.2: Crossing point (CP) values of DNA copies of R18-T1 with substrate 1.....	160
Table E.3: Crossing point (CP) values of DNA copies of R18-T2 with substrate 1.....	161
Table E.4: Crossing point (CP) values of DNA copies of R18-T3 with substrate 1.....	162
Table E.5: Crossing point (CP) values of DNA copies of R18-T4 with substrate 1.....	163

LIST OF ABBREVIATIONS

A	-	Adenine
APS	-	Ammonium persulfate
ATP	-	Adenosine triphosphate
bp	-	Base pairs
C	-	Cytosine
cDNA	-	complementary DNA synthesized from RNA
Cl	-	Chloride
CoA	-	Coenzyme A
CP	-	Crossing Point
DNA	-	Deoxyribonucleic Acid
dNTP	-	Deoxy-nucleotide triphosphate
dsDNA	-	double stranded DNA
DTT	-	Dithiothreitol
EDTA	-	Ethylenediaminetetraacetic acid
EPSP	-	4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid
FAM	-	6-carboxyfluorescein
G	-	Guanosine
GMP	-	Guanosine monophosphate
GNA	-	Glycol Nucleic Acid
ΔG	-	Gibbs free energy
HCN	-	Hydrogen Cyanide
HDV	-	Hepatitis Delta Virus
KCl	-	Potassium Chloride

List of Abbreviations

KH ₂ PO ₄	-	Potassium di-hydrogen phosphate
M	-	Molar
MEME	-	Multiple Em for Motif Elicitation
MGB	-	Minor Groove Binder
MgCl ₂	-	Magnesium Chloride
MLS	-	Multilevel Selection
MLST	-	Multilevel Selection Theory
NADH	-	Nicotinamide adenine dinucleotide (reduced form)
NASA	-	National Aeronautics and Space Administration
NFQ	-	Non-fluorescent quencher
nt	-	Nucleotides
NTPs	-	Nucleoside triphosphates
PCR	-	Polymerase chain reaction
PNA	-	Peptide Nucleic Acid
R18	-	Round 18
RNA	-	Ribonucleic Acid
SELEX	-	Systematic evolution of ligands by exponential enrichment
TBE	-	Tris-Borate-EDTA
TEMED	-	Tetramethylethylenediamine
TNA	-	Threose Nucleic Acid
tRNA	-	Transfer RNA
U	-	Uracil
UV	-	Ultra-violet
VS	-	Varkud Satellite

LIST OF SYMBOLS

Δ	-	delta
μ	-	micro
$^\circ$	-	degree
α	-	alpha
β	-	beta

TERMINOLOGY

Evolutionary terms are often interpreted or given slightly different meanings by researchers in different fields (e.g. molecular biologists and ecologists) especially as applied in the origin of life literature. A glossary of terms as applied in this thesis is, therefore, provided.

Abiotic process: A process characterised by the absence of life

Activated nucleotides: Nucleotides that are primed with a high-energy bond to facilitate their condensation with other nucleotides (Higgs and Lehman, 2015).

Altruism: A behaviour which is costly to the actor and beneficial to the recipient. Cost and benefit are defined on the basis of the lifetime direct fitness consequences of a behaviour (West et al., 2007).

Autocatalytic set: A collection of molecules that mutually cooperate in the sense that none of them can replicate without all the others. The reactions that form the components of the set are catalysed by other components of the set (Higgs and Lehman, 2015).

Cooperation: A behaviour that provides a benefit to another individual (recipient) without an associated cost to the actor.

Error threshold: The theoretical maximum mutation rate that can sustain information genetic polymers of a particular length (Higgs and Lehman, 2015).

Evolutionary Constraint: The bias or limitation in phenotypic variation that a biological system produces (Wagner, 2011).

Functional complexity: The kind of function performed by the RNA molecule, the functional flexibility, and the functional efficiency of a molecule.

Functional flexibility: The ability of a RNA molecule to ligate different kinds of oligonucleotides to their own end.

Functional plasticity: Flexibility of a molecule in its interactions with other molecules and execution of a function.

Hypercycles: Cooperative replicative sets of molecules in which hyperbolic growth is possible (Higgs and Lehman, 2015).

Inclusive fitness: ‘the effect of one individual’s actions on everybody’s numbers of offspring ... weighted by the relatedness’ (Grafen, 1984); the sum of direct and indirect fitness; the quantity maximized by Darwinian individuals (West et al., 2007). Direct fitness is the component of fitness gained through the impact of an individual’s behaviour on the production of offspring. Indirect fitness is defined as the component of fitness gained from aiding the reproduction of related individuals (West et al., 2007).

Level of selection: A hierarchical level at which Darwinian principles (heritable variation in fitness) apply (Nedelcu et al., 2011).

Natural selection: It is the process by which fitness is maximized.

Prebiotic: Existing before the emergence of the first living entities.

Quasispecies: A well-defined distribution of mutants that is generated by a mutation-selection process (Nowak, 1992).

Structural complexity: The RNA secondary structure that contributes to the minimum free energy (measured by the predicted thermodynamic stability).

Structural stability: The predicted thermodynamic stability of a molecule measured by the Gibbs free energy.

Trade-off: An increased investment in one component causes a reduced investment in another component.

Unit of selection: The unit at which natural selection operates e.g. group of molecules, cell etc.

1. INTRODUCTION

The origin and evolution of life remains one of the most mysterious and challenging questions in biology. Scientists have long struggled to define “life” in a way that is broad enough to encompass forms not yet discovered. Physicist Erwin Schrodinger suggested that a defining property of living systems is that they self-assemble against nature’s tendency towards disorder or entropy (Erwin, 1944). Chemist Gerald Joyce’s “working definition” adopted by NASA, is that life is a self-sustaining chemical system capable of Darwinian evolution (Joyce et al., 1994). In the “cybernetic definition” by Bernard Korzeniewski, life is a network of feedback mechanisms (Korzeniewski, 2001). A satisfactory definition for life and its origin that is acceptable to everyone seems unlikely (Luisi, 1998a). However, for investigating the question of life’s evolutionary origin, outlining the necessary components of living systems is at least helpful. These include (i) an energy source (and energy gradient), (ii) basic biochemistry (small molecules and reactions driven by the energy gradients), (iii) organisation (membranes, compartmentalisation and separation from the external environment), and (iv) reproducibility (genetic heritability, information transfer, evolvability) (Penny, 2005). Living organisms are dependent on the highly synergistic molecular cooperation of polymers and multi-molecular assemblies (i.e. nucleic acids, proteins, polysaccharides, lipid membranes). Disagreements have persisted over the first molecules giving rise to life, however, a balance of opinion and empirical evidence support a view of early life based on RNA (Joyce and Orgel, 1999, Kruger et al., 1982, Gilbert, 1986, Joyce, 2004, Muller and Bartel, 2008). This stage of life was termed the “RNA world”. The comparisons between metabolism-first and replication-first scenarios indicate that non-metabolic replication-associated molecules were likely to be the first steps in the RNA world (Wagner et al.,

2010). The conceptual framework for the origin of life from RNA-based replicating polymers, the “RNA world hypothesis” (Gilbert, 1986) is adopted here. Simply stated, this model describes heritable information encoded by RNA polymers and biocatalysis performed by the structure of the folded molecule. These RNA catalysts (ribozymes) served both as genotype and phenotype. While the RNA world hypothesis serves to reduce the number of molecules for sustaining life, there is no reason to believe that other kinds of molecules did not simultaneously emerge along with nucleic acids. This notwithstanding, the origin of RNA and the origin of functional biopolymers in general, is of central importance to understanding the origin of life.

1.1 Ribozymes and the origin of life

The foremost stage in the evolution of life based on catalytic RNA has been supported by the present day role of RNA in cell biology. Some of the most fundamental and highly conserved cellular processes present prominent examples. These include the self-splicing group I and group II introns (Kruger et al., 1982, Peebles et al., 1986), the RNA component of RNase P, which cleaves precursor tRNAs to generate mature tRNAs (Guerrier-Takada et al., 1983) and the various self-cleaving RNAs including the “hammerhead”, “hairpin”, HDV (hepatitis delta virus), and VS (*Neurospora Varkud* satellite) motifs (Prody et al., 1986, Buzayan et al., 1986, Sharmeen et al., 1988, Saville and Collins, 1990). Furthermore, strong evidence that the early biosphere on the earth relied on RNA before the emergence of encoded proteins comes from the structural studies of contemporary catalytic proteins, the ribosomes (Ban et al., 2000, Wimberly et al., 2000, Yusupov et al., 2001). The active site for peptide-bond formation by ribosomes lies deep within a central core of RNA, whereas proteins decorate the outside of this RNA core. It was concluded that the ribosome is a ribozyme (Steitz and Moore, 2003). This supports the

suggestions made long ago that the primitive ribosome could have been made entirely of RNA (Crick, 1968, Orgel, 1968). In fact, all chemical group transfers, and even the information transfers required for coded protein biosynthesis, have precedents within the chemical repertoire of pure small RNAs (Yarus, 1991, Yarus et al., 2009). Individual ribonucleotides (ATP) serve as important signalling molecules and their coenzyme derivatives (acetyl-CoA, NADH, biotin etc.) participate in many of the central metabolism reactions across all domains of life. It was suggested that these molecules are remnants of an earlier RNA based metabolism (White, 1976, Visser and Kellogg, 1978, White, 1982). Looking at the “fossil” evidence of a primordial RNA world in present day cells, extensive discussions were initiated on the role of RNA in the origins of life (Sharp, 1985, Pace and Marsh, 1985, Lewin, 1986).

Scientists have provided additional support for the RNA world by the artificial selection of RNA activities essential for survival of RNA based life. Such activities have been discovered using SELEX (Systematic Evolution of Ligands by Exponential Enrichment) from large combinatorial libraries of RNA sequences (Tuerk and Gold, 1990). It has been observed that RNA can catalyse a broad range of chemical transformations by virtue of its tertiary structure (Chen et al., 2007, Ellington et al., 2009, Joyce, 2004). These include nucleotide synthesis (Unrau and Bartel, 1998), RNA polymerisation (Johnston et al., 2001, Zaher and Unrau, 2007, Cheng and Unrau, 2010), aminoacylation of transfer RNA (Lee et al., 2000), and peptide bond formation (Zhang and Cech, 1997). The other functions include binding small metabolites (such as guanine, S-adenosylmethionine, and lysine), switching from one RNA structure to another, the riboswitches (Breaker, 2012, Garst et al., 2011), self-cleavage activity (Ferré-D'Amaré and Scott, 2010), and RNA splicing reactions (Lambowitz and Zimmerly, 2011). Ribozymes were also obtained for

catalysing chemical reactions such as acyl transfer (Lohse and Szostak, 1996, Jenne and Famulok, 1998), N- and S-alkylation (Wilson and Szostak, 1995, Wecker et al., 1996), carbon-carbon bond formation (Tarasow et al., 1997, Seelig and Jaschke, 1999), amide bond formation (Wiegand et al., 1997), and Michael addition (Sengle et al., 2001). The functional versatility of RNA which is analogous to present day structured proteins supports the capability of RNA to form early life. Based on the biochemical properties of ribozymes, a RNA based organism (ribo-organism) that carried out complex metabolism was conceived (Benner et al., 1989). It was speculated that this RNA based life must have entailed some form of encapsulation (Luisi, 1998b, Szostak et al., 2001). Such compartmentalisation would be advantageous for retaining the fruits of RNA based metabolism for the benefit of the system that produced them and also protection of genomic RNA from degradation. The concept of a ribo-organism is also supported by the evidence that shows the ability of RNA to change the permeability of a membrane (Khvorova et al., 1999), ribozymes capable of acting as membrane transporters (Janas et al., 2004), and redox ribozymes capable of producing energy (Tsukiji et al., 2004). Based on evidence from geophysics, geology, paleobiology, and molecular biology, the widely accepted order of events for the evolution of an RNA world and from the RNA world to contemporary biology is summarised in Figure 1.1. In addition to the catalytic role of RNA in nature and in the laboratory, the origin of life from RNA is also supported by evidence from prebiotic chemistry (Neveu et al., 2013). Components of RNA have been synthesised under plausible prebiotic conditions.

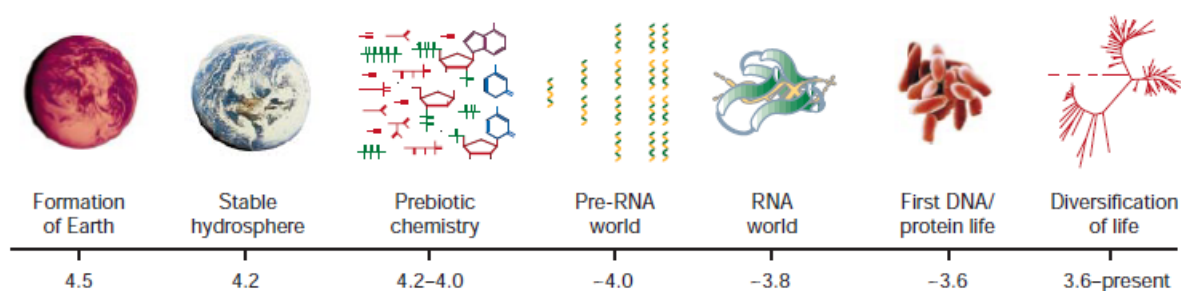


Figure 1.1: Time line of events for the evolution of an RNA world.

The time scale is shown in billions of years (approximately) pertaining to early history of life on earth before the present. Taken from (Joyce, 2002a).

1.1.1 Origin of RNA building blocks

RNA is composed of three substructures: the nucleobases (adenine, guanine, cytosine and uracil), a sugar (ribose), and an inorganic molecule (phosphate group). The three substructures of RNA are chemically bonded to each other via condensation-dehydration reactions. The modular structure of RNA has led chemists to hypothesise that the formation of a RNA polymer started with distinct formation of the substructures followed by sequential formation of nucleosides (nucleobase bonded with ribose sugar), then nucleotides (nucleoside bonded with a phosphate group) and finally a RNA polymer. Assuming this scenario, considerable progress has been made in the chemical synthesis of RNA substructures.

1.1.1.1 Formation of Nucleobases

In this area, the pioneering work was done by Juan Oró and his co-workers who showed that an appreciable yield of adenine can be produced by refluxing a solution of ammonium cyanide (Oró, 1960, Oró and Kimball, 1961, Oró and Kimball, 1962). Adenine

and small amounts of guanine were also detected among products of HCN polymerisation (Miyakawa et al., 2002b, Miyakawa et al., 2002a). These reactions have, however, been complex involving multiple steps. Purines were also synthesised by a mechanism independent of HCN (Miyakawa et al., 2000). In this method, guanine, uracil and cytosine were synthesised by quenching a 90%N₂–10%CO–H₂O high temperature complex organic product (plasma). High temperature plasma used in the study was reported to form by lightning and meteor impacts. Based on the presence of substantial amounts of adenine in carbonaceous chondrites, it was suggested that purines were deposited on earth in meteorites from elsewhere in the solar system (Oró, 1961, Chyba and Sagan, 1992).

The synthesis of pyrimidines was promising from reactions between cyanoacetylene or cyanoacetaldehyde and cyanate ions, cyanogen or urea. Cyanoacetylene was the major product formed on passing an electric discharge through a mixture of nitrogen and methane. Hydrolysis of cyanoacetylene forms cyanoacetaldehyde (Orgel, 2002). Cytosine was obtained in appreciably high yields by incubation of cyanoacetaldehyde with a saturated solution of urea (Robertson and Miller, 1995a, Robertson and Miller, 1995b). Modest yields of cytosine were also obtained from reaction of cyanoacetylene with cyanate (Ferris et al., 1968). It was suggested that these reactions could have proceeded in parallel with synthesis of adenine from HCN (Orgel, 2004a). Simple hydrolysis of cytosine forms uracil (Robertson and Miller, 1995a). This was proposed as a plausible prebiotic formation of uracil.

1.1.1.2 Formation of Ribose sugar

One of the cornerstones in the synthesis of sugars is Butlerow's synthesis (Butlerow, 1861). Bulterow reported that the polymerisation of formaldehyde in the

presence of simple mineral catalysts can form a mixture of sugars (Formose reaction). The reaction, however, produced ribose sugar as a minor product. Eschenmoser and his co-workers demonstrated that the pattern of products formed in formose reaction could be greatly simplified using monophosphate derivatives of glycolaldehyde and glyceraldehyde under alkaline conditions (Mueller et al., 1990). They showed that ribose-2-4-diphosphate was a major product using this approach. Eschenmoser's synthesis has been considered as a promising reaction for the synthesis of ribose sugar if ribose 2-4-diphosphate could be converted to a 5-phosphate or a 1-5-diphosphate (Orgel, 2004b). Another synthesis based on Zn-proline catalysed aldolisation of glycolaldehyde and *rac*-glyceraldehyde formed sugars with 20% of ribose (Kofoed et al., 2005). Ribose has also formed from methanol and water at 70K in simulated interstellar ice grains (Meinert et al., 2016).

1.1.1.3 Formation of Nucleosides

Orgel was the first chemist to initiate the research for producing nucleosides by drying and heating free nucleobases with ribose sugar. The reactions produced ribose nucleosides of adenine (β -adenosine) in modest yields (Fuller et al., 1972a, Fuller et al., 1972b). However, analogous reactions of other nucleobases (guanine, cytosine, uracil and thymine) have not formed their nucleosides in detectable yields (Orgel, 2004b). Sutherland and co-workers demonstrated a multicomponent reaction in water as a possible alternative route to canonical purine nucleosides (Powner et al., 2011). Another possibility of nucleoside formation is from the reaction of ribose sugar with alternative forms of nucleobases. Miller, and later Hud and co-workers, demonstrated that urazole (resembling uracil) and 2-Pyrimidinone (resembling uracil and cytosine) form ribose nucleosides in good yields (Kolb et al., 1994, Bean et al., 2007).

1.1.1.4 Formation of Nucleotides

There was a historical focus on phosphorylating nucleosides with phosphates. Orgel and co-workers demonstrated efficient phosphorylation by drying and heating nucleosides with acidic ammonium phosphate minerals in presence of urea as a catalyst (Lohrmann and Orgel, 1971). This reaction produced a complex mixture of phosphorylated products. Attempts to direct this reaction to the synthesis of particularly nucleoside-5'-phosphate or 5'-triphosphate met with some success (Handschuh et al., 1973, Österberg et al., 1973, Reimann and Zubay, 1999). Nucleoside phosphorylation in good yields was also shown using calcium phosphate minerals (hydroxylapatite) (Lohrmann and Orgel, 1971). In addition to using phosphates, chemists also investigated more reactive forms of phosphorous such as inorganic phosphate, anhydrides of phosphate (e.g. pyrophosphate, metaphosphate), and reduced forms of phosphorous (e.g. phosphite). These phosphorous species were shown to result from the interaction of iron-rich meteorites with water (Pasek et al., 2013). A reaction of nucleosides with tri-metaphosphates in strongly alkaline conditions formed nucleoside 2'-3'-cyclic phosphates that hydrolyse to a mixture of nucleoside 2'- and 3'-phosphates (Saffhill, 1970, Schwartz, 1969). An alternative route to nucleotide synthesis which bypasses problematic piece-wise assembly from ribose sugar and nucleobases was demonstrated (Powner et al., 2009). In this approach, cytosine nucleotides formed from small molecules through a mostly water-based multistep synthesis. Subsequent UV irradiation of cytosine nucleotides formed uracil nucleotides. In another study, glycosidic bond formation in water and nucleotide assembly was shown (Cafferty et al., 2016)

1.1.1.5 RNA as a product of the pre-RNA world

Some of the current challenges faced by prebiotic chemists in synthesising the substructures of RNA *de novo* have led them to hypothesise that RNA is a product of a pre-RNA world (Orgel, 2004b, Hud et al., 2013, Robertson and Joyce, 2012). According to this proposal, some simple and synthetically easily accessible nucleic acid analogues such as Threose Nucleic Acids (TNAs) (Schoning et al., 2000, Yu et al., 2012), Peptide Nucleic Acids (PNAs) (Miller, 1997) and Glycol Nucleic Acids (GNAs) (Zhang et al., 2005) were the ancient molecules from which RNA could possibly have formed. TNAs were shown to form stable Watson crick structures with themselves and with RNA (Ebert et al., 2008, Yang et al., 2007b, Nelson et al., 2000). Although, there is a possibility of a pre-RNA life, a transition to RNA took place at some stage in evolution. The chemical optimality of RNA as an informational and functional molecule has led biochemists to provide experimental evidence for the assembly of RNA oligomers from ribonucleotide monomers.

1.1.2 Abiotic synthesis of RNA oligomers

The polymerisation of nucleotides in aqueous solution is an uphill reaction and does not occur spontaneously to any significant extent. Therefore, activated derivatives of nucleotides such as nucleoside 5'-polyphosphates or nucleoside 5'-phosphorimidazolides have been used for the experimental synthesis of oligomers. The three principal nucleophilic groups in activated nucleotides, which participate in the polymerisation reaction are: the 5'-phosphate, the 2'-hydroxyl and the 3'-hydroxyl group, in order of decreasing reactivity. Therefore, the reaction of a nucleotide or oligonucleotide with an activated nucleotide normally yields 5', 5'-pyrophosphate, 2', 5'- phosphodiester, and 3',5'-phosphodiester-linked adducts in order of decreasing abundance (Sulston et al., 1968). Coupling of more than 15 activated adenine and uridine- monophosphates were

obtained using lead (II) as a catalyst under eutectic conditions (Kanavarioti et al., 2001, Monnard et al., 2003). The product, however, contained a large proportion of 2'-5' linkages.

The most appealing strategy for catalysing highly regiospecific reactions was adsorption to a mineral surface. Surface enhanced oligomerisation of nucleoside 5'-phosphorimidazolides and related activated nucleotides was extensively studied on clay mineral montmorillonite (Ferris and Ertem, 1993, Kawamura and Ferris, 1994, Miyakawa and Ferris, 2003, Ferris, 2006). Using this approach, the formation of 40-50 nucleotide long oligomers was shown, wherein, adenine oligomers were mainly 3',5'- linked and pyrimidine oligomers were predominantly 2',5'- linked (Ferris, 2002, Ferris et al., 1996). Some other studies showed the formation of 40 nucleotide long oligomers on montmorillonite using 1-methyl-adenine activated adenine and uridine-monophosphates (Prabakar and Ferris, 1997, Huang and Ferris, 2003). The adenine and uridine oligomers consisted of ~75% 3',5'- linkages and ~60% 3',5'- linkages, respectively (Huang and Ferris, 2006). A detailed analysis of catalysis by montmorillonite has suggested that oligomerisation occurs at a limited number of structurally specific active sites within the interlayers of the clay platelets (Wang and Ferris, 2001, Joshi et al., 2009). These studies have represented one of the most impressive prebiotically plausible examples of oligonucleotides synthesis. The findings have reinforced the suggestion that life may have started on mineral surfaces, perhaps in clay—rich muds at the bottom of the water pools formed by hot springs (Hazen, 2001). An alternative approach has shown dry state oligomerisation of non-chemically activated nucleotides. Acidic form of cyclic 3',5'- GMP (cGMP) polymerises to form 40 nucleotide long oligomers if dried at elevated temperatures (Morasch et al., 2014). Once short oligomers were established, the next stage

would have been their replication so that a process equivalent to natural selection could begin.

1.1.3 Non-enzymatic replication of RNA oligomers

The reaction thought by many to be central to replication of nucleic acids is template-directed synthesis, that is, the synthesis of a complementary oligonucleotide under the direction of a pre-existing oligonucleotide. The basic principle is the formation of a double-stranded complex if a polynucleotide is incubated with an appropriate mixture of complementary mononucleotides or short oligonucleotides. Researchers have tried understanding the replication scheme using monomers activated with 2-methylimidazole or 5'-phosphorimidazole (Inoue and Orgel, 1981, Inoue and Orgel, 1982). Efficient and highly regiospecific formation of poly (G) up to 50-mer over poly (C) templates was demonstrated. Reactions with random co-polymers with an excess of C residues have given complementary products containing G residues with a mean chain length of 6-10 (Inoue and Orgel, 1983). The fidelity of the reactions was extensively studied (Inoue et al., 1984, Acevedo and Orgel, 1987, Wu and Orgel, 1992b, Hill et al., 1993, Wu and Orgel, 1992a). Incorporation of G opposite C in the template was most efficient, while incorporation of U opposite A was least efficient. Incorporation of A opposite U or of C opposite G were of intermediate efficiency. The fidelity was usually very good, except for the misincorporation of G on some RNA templates because of GU wobble pairing. Successful primer extension across A rich sequences such as, AA, AU and AG was demonstrated within ice eutectics containing lead (II) and magnesium (II) ions (Monnard and Szostak, 2008, Loffler et al., 2013). In another approach, cold temperature and immobilisation of template-primer strands with periodic replenishment of activated

ribonucleotides has led to the successful copying of all the four nucleobases (Deck et al., 2011).

A viable alternative scheme of chemical self-replication system was devised based on template directed ligation of activated short oligonucleotides (James and Ellington, 1999, Rohatgi et al., 1996a, Rohatgi et al., 1996b). It was found that 3', 5'- linked oligonucleotides were superior to mononucleotides as substrates with respect to regiospecificity and the temperature ranges at which the reactions occurred. The fidelity of the reactions was, however, compromised since single base mismatched oligomers hybridised as efficiently as fully complementary oligonucleotides. The non-enzymatic RNA polymerisation has commonly resulted in heterogeneous backbone linkages. A mixture of 2', 5' linkage and 3', 5' linkage have formed rather than only 3', 5' linkage as found in the contemporary RNA. RNAs containing remarkably high proportions of 2', 5' linkages have, however, formed functional nucleic acids and templates for replication (Engelhart et al., 2013, Prakash et al., 1997). Non-heritable backbone heterogeneity has, therefore, been considered one of the essential features for the emergence of RNA. Assuming that a library of replicating oligomers would have formed by chemical processes, some of them would have performed functions essential for the survival and stability of RNA based life.

1.1.4 Replication and stability of catalytic RNA polymers

One of the essential functions for the survival and stability of an RNA based life is self-replication. Researchers have made extensive efforts in finding a true replicase, i.e. a polymerase ribozyme that replicates itself. Such a molecule must act on itself to produce complementary RNAs, and act on the complementary RNAs to produce additional copies

of itself. The development of such a molecule has, however, been a challenge. In the efforts to study how a replicase could possibly function, researchers explored template directed polymerisation of mononucleotides and oligonucleotides by ribozymes (Joyce, 2007). This function involved the chemical joining of ribonucleotide monomers or oligonucleotides complementary to a template with the help of a ribozyme. The first success in this regard was the development of a “Class I” RNA ligase ribozyme that catalysed joining of two template bound oligonucleotides (Bartel and Szostak, 1993). Condensation occurred between the 3' hydroxyl of one oligonucleotide and the 5' triphosphate of another, forming a 3'-5' phosphodiester linkage and releasing inorganic pyrophosphate. It is classified as a ligation reaction because of the nature of the oligonucleotide substrates, however, it involves the same chemistry by which a RNA polymerase functions. Furthermore, efforts were made to develop polymerase ribozymes which can operate on a separate RNA template and polymerise mononucleotides. A *bona-fide* RNA polymerase (Round 18 ribozyme) was developed from Class I ligase using *in vitro* evolution experiments (Johnston et al., 2001). Round 18 ribozyme could add 14 nucleoside triphosphates (NTPs) in 24 hrs to a primer-template complex with an average fidelity of 0.967. A few mutations in this polymerase ribozyme further improved the fidelity, length of extension and rate of polymerisation for an external favourable template (Wochner et al., 2011, Attwater et al., 2013). The improved ribozyme polymerases; although far from self-replication, were able to synthesise RNAs up to 95 nucleotides in length (average fidelity; 0.991) and around 200 nucleotides (average fidelity; 0.974) from a RNA template.

Other kinds of self-replicating systems have been developed are based on RNA molecules that mutually assemble each other. These systems typically entailed a template

that directs the ligation of two substrates to form a product that is identical to the template. Autocatalysis occurs when the product can also function as a template, directing the formation of additional products. The first demonstration of this kind was an autocatalytic system of short oligomers such as hexa-deoxynucleotide and tetra-deoxynucleotides (Sievers and von Kiedrowski, 1994, Zielinski and Orgel, 1987). Following this, an autocatalytic system of a ligase ribozyme was devised (Paul and Joyce, 2002). Exponential growth, however, was limited in these systems because of slow dissociation of the template-product complex. This problem was resolved when a cross catalytic system of ligase ribozymes was developed (Lincoln and Joyce, 2009, Kim and Joyce, 2004). In this system, two species could act as templates for each other without being self-complementary. Exponential growth was observed and the system was self-sustainable. Apart from ligation based autocatalytic systems, recombination based self-sustained systems have also been developed. Cross catalytic recombination has proved to be successful for the autocatalytic assembly of group I self-splicing ribozyme and Azoarcus ribozyme from their constituent fragments (Doudna et al., 1991, Vaidya et al., 2012).

Considerable progress has been made in understanding the origin of life from RNA. Biochemists have provided evidence for the prebiotic formation of small RNA molecules (40-50 nucleotides long). Molecular biologists have developed sophisticated molecules with functions essential for survival and stability of RNA based life. A gap in the evolution of RNA based life, which remains unclear, are the processes by which complex molecules and systems emerged from small, less complex RNAs. Researchers have tried to address this problem theoretically, computationally and to some extent empirically. Understanding these processes in more detail is the focus of this project.

1.2 Emergence of complexity at the origin of life

A vital function for an RNA based life to emerge and evolve is replication. The replicative potential of RNA systems has been described in the previous section. The success of such systems, however, requires RNA polymers that are far too long to accumulate spontaneously under plausible prebiotic conditions. Furthermore, in the early world, the high mutation rate limited sequence length that could be accurately replicated (the Eigen limit). There is a well-established theoretical framework for assessing the effect of genome size, replication rate and replication fidelity on the ability to maintain genetic information (Eigen, 1971). Generally, if the number of replicable error copies of an advantageous molecule exceeds the number of accurate copies, then the fittest molecules cannot be enriched by selection. This is Eigen's error catastrophe (Eigen 1971, Eigen and Schuster 1977), and a critical value of replication accuracy was established known as Eigen's error threshold. As a rough guide, the error rate of replication per nucleotide must be no more than about the inverse of genome length. This corresponds to a 99% fidelity for the replication of a 100 mer and 97.5% fidelity for replication of a 40 mer. A high mutation rate, therefore, severely constrained the amount of genetic information that could be stored and reliably transmitted to subsequent generations. A limitation in genetic information could have further limited the functional capabilities of the evolving molecules. In addition, the knowledge that only protein-based replication system can achieve the required high fidelity, led to the framing of Eigen's paradox: "no genomes without enzymes and no enzymes without genomes".

1.2.1 Hypercycles and the emergence of complexity in the RNA world

One of the seminal works providing an adaptive explanation for the emergence of genomes is Eigen's theory of hypercycles (Eigen and Schuster, 1977, Eigen and Schuster, 1979). According to this theory, individual catalytic information containing molecules cooperated to form a functionally or replicatively integrated network and subsequently a primitive genome. The theory addressed the error catastrophe by invoking the stability of molecules (oligonucleotides and their encoded proteins) in the form of various-membered hypercycles. Based on Eigen's theory of hypercycles, mathematicians tried to address the problem and worked out the formation of autocatalytic sets for the stabilisation of molecules. The first proposed model was that all molecules in the set can be synthesised by reactions that are catalysed by other molecules in the set (Kauffman, 1993). The set as a whole is mutually autocatalytic, even though none of the molecules are individually autocatalytic. This model was well studied and applied to chemical reaction systems and RNA replicator systems (Hordijk and Steel, 2004, Hordijk and Steel, 2013, Smith et al., 2014). In addition to addressing the origin of replicator networks, hypercycle theory also overcomes the problem the low copying fidelity of early enzymes. Shorter, catalytically inferior ribozymes could become functionally connected, facilitating the evolution of cooperation and the origin of complex networks. The stability and evolution of groups of interacting ribozymes have been investigated using a multilevel selection theory (MLST) approach (Takeuchi and Hogeweg, 2009, Takeuchi and Hogeweg, 2012). It was concluded that natural selection can act at multiple levels of organisation based on spatial clustering and compartmentalisation of catalysts. In the case of primitive gene networks, natural selection may act on the individual ribozymes in the group as well as the group itself (Durand and Michod, 2010, Michod, 1983). Various scenarios of hypercycles have been

proposed (Meyer et al., 2012, Higgs and Lehman, 2015) for the sustainability of a self-replicating system of ribozymes: (A) Complex emergence, (B) Cooperative emergence and (C) Cooperative replication hypercycle (illustrated in Figure 1.2).

- (i) In the first scenario, a complex catalyst capable of full replication emerges from either mononucleotides or oligonucleotides. Such a self replicase or polymerase could sustain an autocatalytic cycle by alternating copying of itself and its complementary sequence (Figure 1.2 A). This kind of replication can be summarised as:



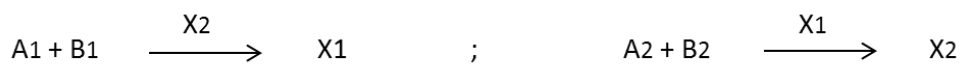
In this, P is a polymerase and P' is its complementary sequence (not a catalyst). According to this scenario, P would have to act as a catalyst as well as a template. The molecule would, therefore, have to exist in two alternate conformations: the folded state where the secondary structure is necessary for catalysis and the unfolded state so that it can act as a template. The two states are mutually exclusive, which could be an evolutionary constraint. This kind of replication mechanism has not been empirically demonstrated so far, however, general polymerase ribozymes which can copy an external template have been developed (Attwater et al., 2013, Johnston et al., 2001, Wochner et al., 2011). Such a polymerase is an altruistic cooperator because it replicates other sequences, however, it would only be replicated when another polymerase uses it as a template.

- (ii) In the second scenario, replication occurs by a hypercycle composed of two or more components where each component catalyses the next one in a circular arrangement. Such an autocatalytic set can be summarised as:



In this, X1 and X2 replicate themselves with the aid of catalysis provided by the other sequence. For example, a minimal polymerase and a minimal recombinase emerge from random oligonucleotides. This kind of replication could involve cooperation between a recombinase and a polymerase (Figure 1.2 B).

- (iii) In the third scenario, replication could occur by mutual cooperation between different molecules using precursors in the pool. Each co-operator in the set needs the other components in the set to replicate. Such an autocatalytic set can be summarised as:



In this, X1 and X2 are catalysts that catalyse each other's formation from precursors (As and Bs). For example, a replication hypercycle can be sustained through interconnected polymerisation and recombination cycles. In one cycle, polymerisation of the short RNA fragments comprising the polymerase and recombinase occurs. In the other cycle, the reconstituted recombinase stitches the RNA fragments (Figure 1.2 C). This kind of cooperation may be composed of catalysts with different functions and each contributes to the replication of the system as a whole. An example of this kind of cooperation which has been

empirically demonstrated is the autocatalytic system of ligases (Lincoln and Joyce, 2009) and recombinases (Vaidya et al., 2012, Vaidya et al., 2013).

While the theoretical models address the stability of complex systems by cooperation, empirical observation has revealed sets of cooperative molecules that are far too complex to be explained by our current understanding of the prebiotic world. How did these complex molecules emerge? The missing link, which is yet not clearly understood, is the process implicated in the increase in complexity from short oligonucleotides to large sophisticated catalysts. Specifically, laboratory investigations of both the ultimate (“why” this increase in complexity arose) and proximate (“how” this happened - the mechanistic framework) questions are lacking. Furthermore, the disconnect between theoretical models, which have concentrated on ultimate reasoning without fully accounting for empirical observation, and biochemical mechanisms has severely hampered progress. The need for experimental evaluation of theoretical concepts is a scientific imperative (West et al., 2007).

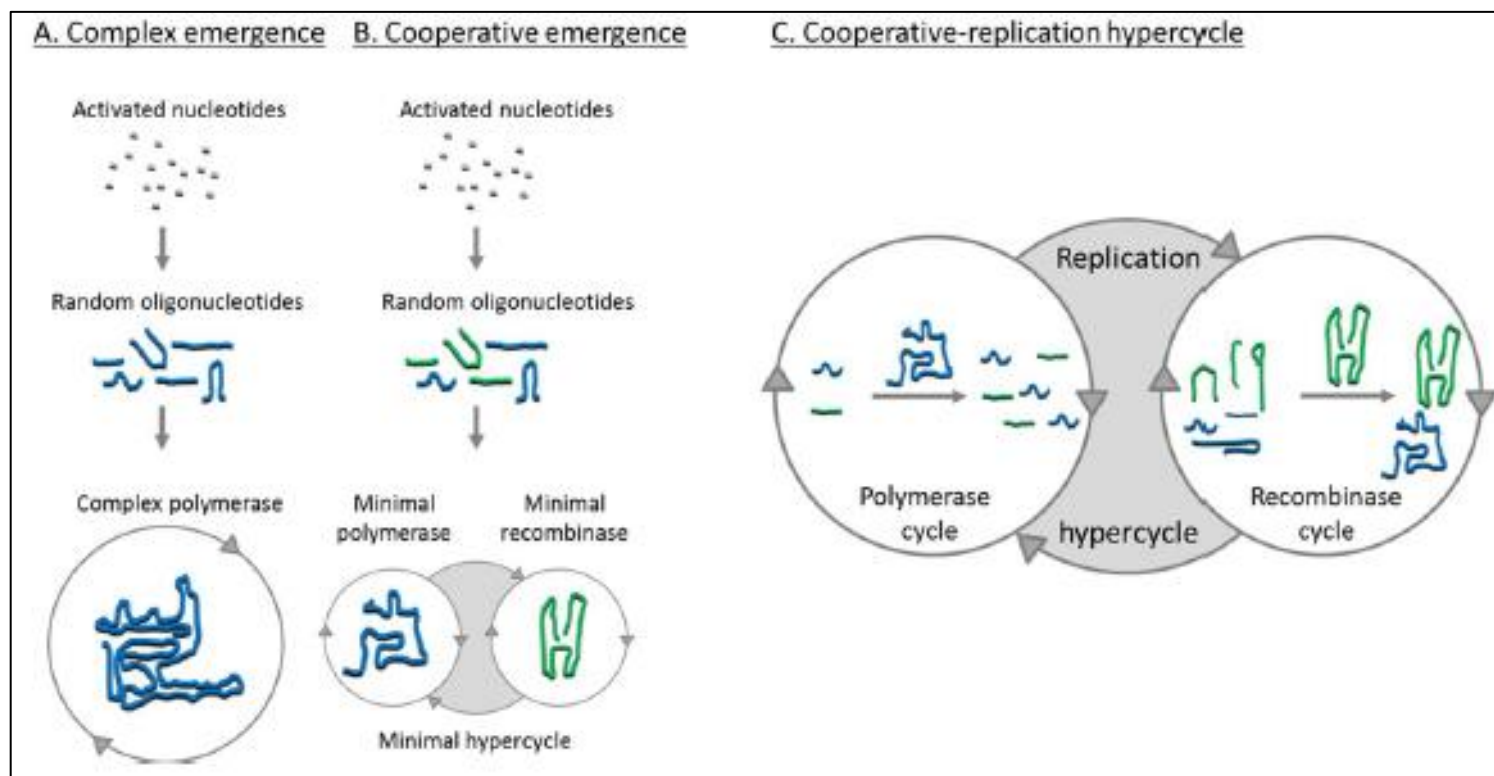


Figure 1.2: The various scenarios of Hypercycles.

In Complex emergence, a self-replicase emerged from mono- or oligonucleotides capable of replicating itself. In cooperative emergence, a minimal polymerase and a recombinase emerged from mono- or oligonucleotides. The two molecules cooperated in a cycle to sustain each other. In cooperative-replication hypercycle, a minimal polymerase and a recombinase emerged. The two molecules cooperated using their constituent fragments. Taken from (Meyer et al., 2012).

One of the processes which could assist in the assembly of complex catalysts from short precursor strands in the pool was recombination. Such a reaction system has been empirically demonstrated for the formation of group I self-splicing ribozyme (Doudna et al., 1991, Hayden and Lehman, 2006, Hayden et al., 2008). The assembly of the recombinase ribozyme results from secondary structure formation from fragments directed by complementarity between sequences. A prebiotic environment, however, was more likely to be abundant in random fragments and could be limited in required complementary fragments in the zone of catalysis. It is less clear how such complex functional molecules might have emerged in a heterogeneous environment. Such reaction sets are, however, more likely to emerge in the presence of a polymerase which partially polymerised templates and generated complementary fragments. A minimal polymerase, therefore, seems essential for the stability of recombination based reaction sets. The theoretical co-participation of a polymerase and a recombinase in a hypercycle for emergence of complexity has also been proposed (Figure 1.2 C). A minimal polymerase would also be one of the prerequisites for the emergence of the other complex autocatalytic systems which are functionally based on complementarity between sequences.

1.2.2 The history of engineered RNA polymerases

The efforts towards engineering of RNA polymerases started with the *in vitro* selection of RNA ligases which could efficiently join two oligonucleotides in a template-directed manner (Bartel and Szostak, 1993, Eklund et al., 1995, Eklund and Bartel, 1995). The reaction used specific substrates partially complementary to the ligase ribozyme at the ligation junction. It was anticipated that such ligases could be further coaxed to accept NTPs as substrates and to add multiple NTPs in succession in order to create an RNA system with true auto-replicative potential. Some of the selected ligases when tested for

polymerisation function showed limited activity (Ekland and Bartel, 1996). An efficient ligase ribozyme (98 nt) selected from a random pool was able to add only six mononucleotides in 144 hrs (average fidelity of 0.88) to a primer-template complex (Figure 1.3). The polymerisation efficiency of the ligase was improved by appending an additional domain of 76 nucleotides to the ligase catalytic core with some minor nucleotide mutations (Johnston et al., 2001). This developed ribozyme polymerase (Round 18 ribozyme; 189nt) was able to extend a general primer-template by 14 nucleotides in 24 hrs (average fidelity of 0.967-0.985; approaches the fidelity of a yeast polymerase “pol η ” needed for accurate replication of UV damaged DNA). It was found from the study that the auxiliary domain conferred an ability to effectively recognise the primer template complex and improved the polymerisation efficiency. The processivity and fidelity of the Round 18 ribozyme was further enhanced by a few mutations and selection by *in vitro* compartmentalisation (Zaher and Unrau, 2007). The variant B6.61 polymerase incorporated 20 nucleotides onto a primer template complex. Some of the other variants of R18 polymerase with minor mutations showed an increased efficiency of replicating a favourable external template. One of the variants was able to polymerise RNAs up to 95 nucleotides in length (average fidelity; 0.991) from an RNA template (Wochner et al., 2011). Another minor mutational variant of R18 polymerase when selected at cold temperatures polymerised RNAs up to 206 nucleotides in length (average fidelity; 0.974) (Attwater et al., 2013).

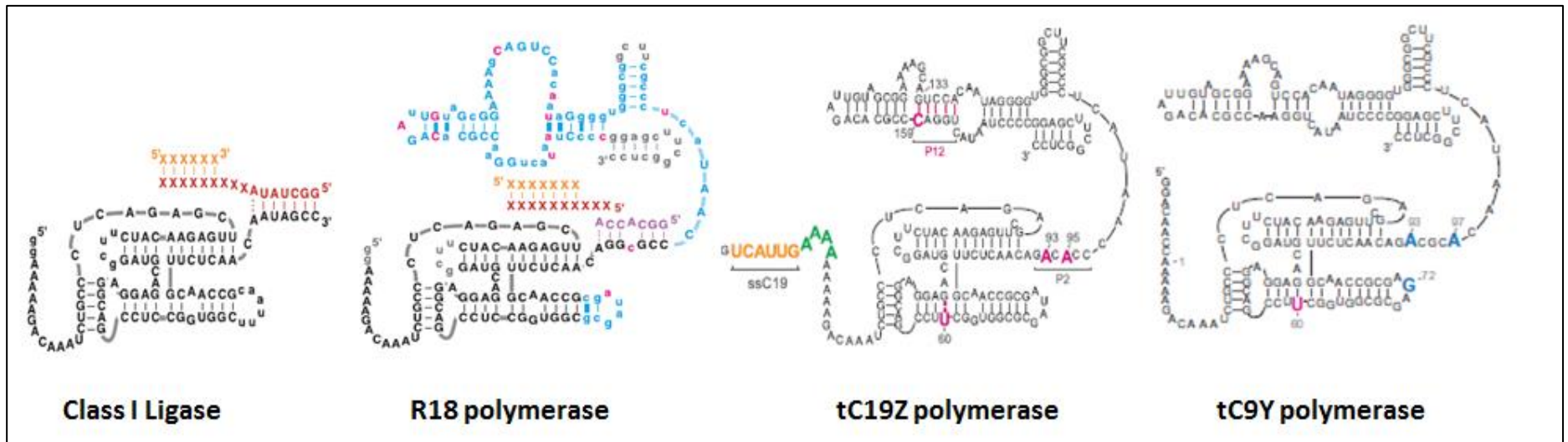


Figure 1.3: Structures of the engineered RNA polymerases.

R18 polymerase, developed from Class I ligase, replicates a template up to 14 nucleotides. The structures are taken from (Johnston et al., 2001).

Polymerases tC19Z and tC9Y (minor mutational variants of the R18 polymerase) replicate a template up to 95 nucleotides and 206 nucleotides, respectively. The structures are taken from (Wochner et al., 2011, Attwater et al., 2013).

The *in vitro* evolution experiments have evidently shown that 1) the size and the structural complexity of RNA molecules was an important determinant in the development of sophisticated metabolic functions like polymerisation. The small size of the ligase molecule constrained its ability to perform polymerisation. And essentially, the first breakthrough in this activity was achieved only when the length of the RNA catalyst was increased to ~200 nucleotides; and 2) only after polymerases of this size formed, an active replication process of molecules could have begun, which was essential for the stability of RNA based life. The engineered RNA polymerases are, however, far too complex to have assembled by passive chemical processes without catalysts. **This leads to the question: Which basic processes were implicated in the emergence of a minimal RNA polymerase from much shorter oligonucleotides?**

1.2.3 Question 1: How could a minimal polymerase emerge from short molecules?

A model for the evolution of a polymerase function from early self replicators or short RNA polymers based on template directed oligonucleotide ligation reactions was proposed (James and Ellington, 1999). In this scenario, it was envisioned that longer oligomers could have formed via interaction between different self-replicators (Figure 1.4). The potential role of ligation of oligonucleotides in building a replication system has also been suggested by others (Bartel, 1999, Szostak, 2011, Rohatgi et al., 1996b).

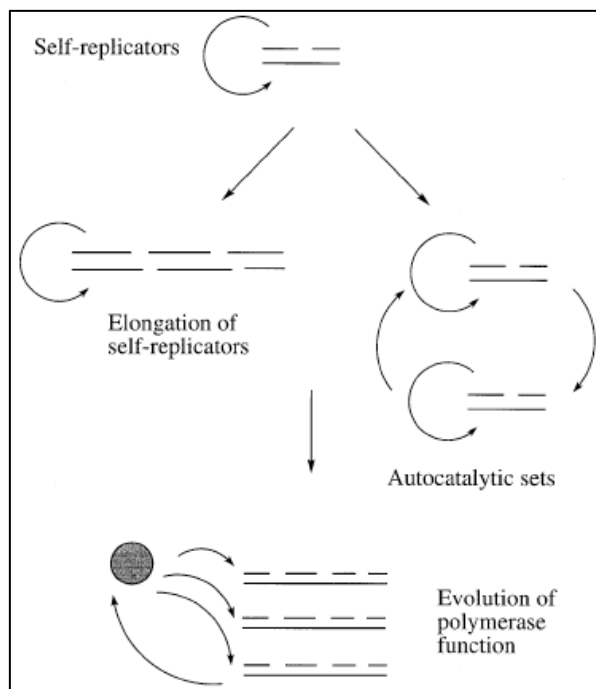


Figure 1.4: A model proposed for the evolution of a RNA polymerase.

Self-replicating short oligomers arose from non-enzymatic process. This process entailed a template that directed joining of two substrates to form a product that is identical to the template. The self-replicators elongated further by template directed ligation mechanisms acquiring additional, catalytic information and functionality Adapted from (James and Ellington, 1999).

The limitation of the model in Figure 1.4 is that it requires the presence of specific substrates which are at least partially complementary to the template. Such reactions could be impaired in a heterogeneous prebiotic microenvironment. The lack of specific substrates is a constraining feature which would have led to a collapse of the emerging complexity. The model reactions, however, are more likely to be plausible and sustainable in the presence of a RNA polymerase that generates complementary substrates. Furthermore, such reactions are more likely to utilise substrates of heterogeneous lengths. This might result in intermediate longer products separated by gaps of one or more nucleotides. The gapped products could be filled by RNA polymerases. Such gap filling activity, however, has only been reported in molecules which are themselves very complex (McGinness et al., 2002). This presents a conundrum for the emergence of a minimal RNA polymerase: A polymerase seems inevitable for the evolution of a polymerase.

In an alternative model, computational methods have demonstrated ligation reactions between independently evolved RNA molecules as a potential pathway for the increase in molecular length and functional complexity (Manrubia and Briones, 2007). A combination of different simple RNA modules has been proposed as a much more likely mechanism for emergence of a RNA replicase (Briones et al., 2009). In this scenario, large repertoires of short, genetically different molecules would likely have been produced and folded into secondary or tertiary structures (illustrated in Figure 1.5). A fraction of these molecules could have ligase activity. Ligation of different oligomeric RNA structures could have progressively given rise to more complex molecules. Modular evolution has been proposed to shorten adaptation times and allow formation of complex structures that could not otherwise be directly selected. This kind of process could have overcome Eigen's mutational constraint in the evolution of complexity. **Based on this model, this research**

seeks to empirically understand the processes which led to an increase in complexity and the emergence of a minimal polymerase from short RNA oligomers.

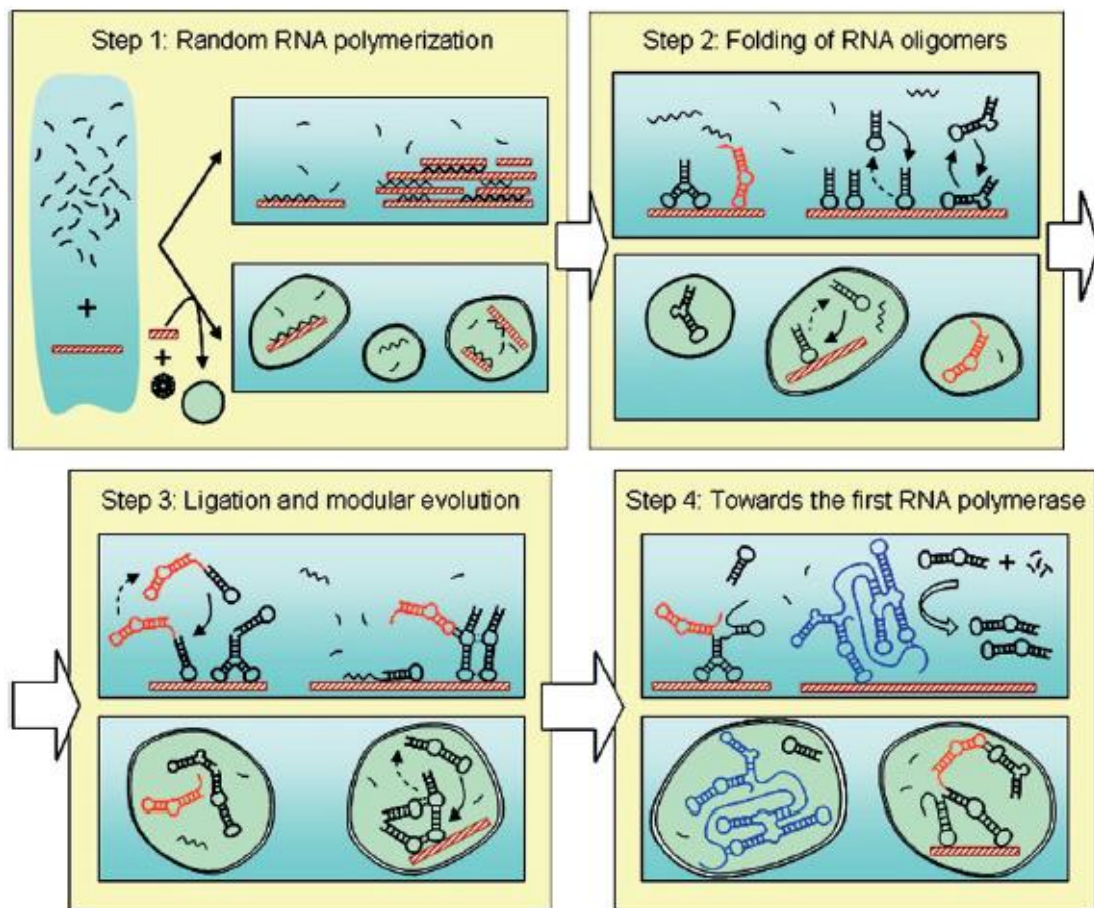


Figure 1.5: A model based on a stepwise process for the emergence of RNA polymerase.

In every step two possible and compatible scenarios are depicted: evolution on mineral surfaces (shown as brown rectangles) in bulk solution, as well as evolution inside vesicles that could also encapsulate mineral particles. Functional hairpin structures (with ligase activity) are shown in red. Solid and dotted arrows stand for the surface-bound to in-solution equilibria. The RNA polymerase emerging from this process is depicted in blue. Taken from (Briones et al., 2009).

Theoretically, the stability of complex systems has been hypothesised to occur in the context of hypercycles where RNA catalysts are functionally or replicatively integrated (Eigen and Schuster, 1977, Eigen and Schuster, 1979, Meyer et al., 2012, Higgs and Lehman, 2015). Such interactive or mutually connected networks would be more likely to emerge from a pool of structurally and functionally diverse molecules. A prebiotic pool which would possibly be limited in the kinds of RNA molecules formed could have poor evolutionary potential. **This leads to the second question: which basic evolutionary processes allowed the progressive emergence of diversity for the stability of RNA networks?**

1.2.4 Question 2: How could complex diverse molecules emerge from short molecules?

Mutations inherent to replication errors in the prebiotic scenario could be one of the processes that increased molecular diversity (quasispecies). The resulting heterogeneous pool could give rise to molecules with optimal and sub-optimal activities. The structural and functional diversity generated by mutagenic processes would, however, be limited to the available sequence space. It has been demonstrated *in vitro* that generally a pool complexity of 10^{14} - 10^{16} randomly mutated molecules is needed for the evolution of an efficient functional molecule (Ellington and Szostak, 1990, Bartel and Szostak, 1993, Joyce, 2004, Joyce, 2007). An early pool of molecules which would likely be limited in its sequence length would, therefore, be limited in the number of newer phenotypes and functions generated from random mutations. For the evolution of RNA networks, the

enrichment of the pool with diverse informational, structural and functional complexities for network stability must have come from mechanisms besides mutations.

The sequence length is an important determinant of the functional complexity of molecules (Gevertz et al., 2005, Sabeti et al., 1997, Joyce, 2002b). It has been noted in several studies that variations based on the addition or exchange of nucleotide domains significantly change the structural properties of molecules. A classic example is the *in vitro* development of the R18 polymerase ribozyme itself (discussed earlier); in which the domain addition to a ligase was necessary for engineering the polymerisation function (Johnston et al., 2001). Other examples are the selection of new functional RNAs engineered through ligation and recombination of structural domains (Burke and Willis, 1998, Joyce, 2004). Additionally, bi-functional enzymes endowed with RNA ligase and cleavage functionalities have been produced by joining catalytic RNA motifs (Kumar and Joyce, 2003, Landweber and Pokrovskaya, 1999). Also, allosteric ribozymes and effector-activated ribozymes have been designed by combining different functional RNA domains (Tang and Breaker, 1997, Komatsu et al., 2002, Robertson and Scott, 2007). New catalytic RNA functions have evolved by appending random sequence segments to a pre-existing functional domain of natural ribozyme (Jaeger et al., 1999). **Based on the above studies, this research seeks to empirically understand the key mechanisms which led to a diversity of complex molecules from a limited pool of prebiotically generated RNA polymers.**

1.2.5 The proposed argument

Could ligation of simple RNA oligomers possibly explain the process of increase in complexity and emergence of a minimal polymerase?

1.2.5.1 The possible role of small ribozymes

In the primitive world, the role of small functional RNA molecules would be particularly relevant before any of the complex systems emerged. Small RNA structures have been found to be evolutionary significant in the early world due to their genetic or mutational robustness (Manrubia and Briones, 2007). Robustness is a feature of different RNA sequences folding into the same secondary structure. With robust traits different sequences in a quasispecies population preserve their phenotype in the face of genetic perturbations (de Visser et al., 2003). This characteristic has been linked to long term increased adaptive potential (Masel and Trotter, 2010). Mutational robustness has been empirically detected, among others, in the genome of RNA viruses (Wagner and Stadler, 1999) and in the structure of natural miRNAs (Borenstein and Ruppin, 2006). Robustness of small RNA modules could lead to selection and preservation of a functional phenotype over generations (Knight and Yarus, 2003, Wang and Unrau, 2005). A combination of such selected modules could result in the evolution of complex molecules and functionalities (Manrubia and Briones, 2007, Briones et al., 2009). This has been explored from the analysis of structural motifs and their catalytic function in extant RNA molecules (Fontana et al., 1993, Knight and Yarus, 2003, Gevertz et al., 2005, Bourdeau et al., 1999, Hendrix et al., 2005). Topologically, simple RNA modules like stem loops and hairpin structures have been obtained in abundance (Stich et al., 2008). Hairpin structures of different sizes are common in current viral and cellular RNAs, being involved in RNA–RNA and RNA–protein interactions that guide RNA folding, ribozyme function, RNP structure, and gene expression regulation (Svoboda and Di Cara, 2006). Certain hairpin-like structures have been endowed with the ability to catalyse RNA cleavage/ligase reactions in a reversible way (Ivanov et al., 2005). Naturally occurring hairpin ribozymes

have been found to be functionally diverse (Puerta-Fernández et al., 2003). Interestingly, even the truncated and fragmented derivatives of a hairpin ribozyme showed ligase activity (Vlassov et al., 2004). Mapping experiments of aptamers selected for interaction with organic dye molecules have revealed the binding sites comprised of only 20-40 nucleotides (Ellington, 1994a, Ellington, 1994b). Selected sequences that bind vitamin B12 have been found to be short helical structures (Lorsch and Szostak, 1994). Other selection experiments have also revealed minimal active RNA motifs (Lozupone et al., 2003, Wang and Unrau, 2005). All the above studies emphasise the fact that simple nucleic acids are quite adept at forming active sites for catalysis. **Stable short motifs with ligase function could possibly join oligomers to their own end (self-ligation) forming larger and phenotypically diverse molecules. This process of self-ligation needs to be experimentally examined in short RNA oligomers and in the larger oligomers.**

1.2.5.2 The possible role of unconstrained ligation reactions

The potential of ligation reactions by ribozymes has been explored primarily to develop a replicating system, which was essential for life to begin (Joyce, 2004, Joyce, 2007, Johnston et al., 2001, Wochner et al., 2011, Attwater et al., 2013, Paul and Joyce, 2002, Kim and Joyce, 2004, Lincoln and Joyce, 2009). These replication systems have been based on the interaction of ribozymes with specific substrates. An early heterogeneous prebiotic pool could have become an evolutionary constraint for the emergence of such systems. In such a case, processes that utilise heterogeneous pool of molecules could have built complex replicating systems (Szostak, 2011).

The potential of ligation reactions by ribozymes in a system of unrelated RNA oligomers has not been explored in detail. Such reactions could play a role in increasing

the prebiotic complexity of early molecules and in the emergence of essential functions. The catalytic reactions could utilise differential pairing according to the stereochemistry between the molecules. Forms of functional plasticity and flexibility in catalytic nucleic acid systems have unexpectedly been observed in various *in vitro* studies. In one of the studies, selected deoxyribozymes ligated multiple substrates via a common ligation junction (Levy and Ellington, 2002). The deoxyribozyme eschewed the substrate binding site provided for them in favour of their own unique substrate binding site. In another study, a small ligase ribozyme utilised non Watson-crick base pairs at its ligation junction and not the designed base pairing for its activity (Robertson et al., 2001). It was proposed that a non-Watson crick stack may have formed a secondary structure that was particularly conducive to ligation. Another study that was designed for selection of ligase ribozymes to yield template directed 3'-5' linkage reported emergence of ribozymes that accelerated an unexpected 5'-5' linkage (Chapman and Szostak, 1995). Functional flexibility of RNA systems was also observed in a cross chiral polymerase system (Sczepanski and Joyce, 2014). Opposite handed versions of a polymerase ribozyme, the d-and the l-enantiomer efficiently catalysed their respective joining reactions in a mixture containing both d- and l- versions of the substrates and templates. The opposing enantiomers of RNA molecules which are unable to form consecutive Watson–Crick base pairs with each other recognised each other based on tertiary interactions. All these studies suggest that catalytic molecules can overcome sequence restrictions by adapting alternative mechanisms beyond mere templating. **In an environment with limited complementary substrates, such processes might be potentially significant for the emergence of complexity. Relaxation of pairing constraints between molecules, therefore, might prove beneficial in tackling the apparently more difficult problem of the emergence of complex catalysts. Ligation**

activity of RNA oligomers with oligonucleotides in the absence of designed base pairing needs to be examined experimentally.

The focus of this project was to investigate the possible evolutionary processes underpinning the formation of complex diverse molecules and a minimal polymerase from short RNA oligomers.

Hypothesis:

- Small functional ribozymes with ligase activity in a heterogeneous pool of sequences may have an important role in the progressive (step by step) emergence of complex structures from short oligomers.
- Self-ligation reactions that were not directed by any defined sequence or template could possibly explain the mechanistic process implicated in the emergence of a polymerase.

Aim:

The overall aim was to understand the steps by which a collection of biomolecules (catalytic RNA) are able to form higher levels of complexity that display some of the biochemical and evolutionary properties that we associate with life.

Objectives:

To study the processes that may have been important for the emergence of diverse and complex molecules such as a minimal polymerase ribozyme from short RNA oligomers.

- Decrease the size and structural complexity of a minimal polymerase ribozyme (R18 polymerase) to short RNA oligomers
- Investigate the ability of the polymerase and its truncated derivatives to ligate oligonucleotide substrates to their own end (self-ligation function).
- Study the dynamics of self-ligation function with nucleotide variations in the substrates.
- Correlate the molecular traits of the ribozymes with increase in their size and structural complexity.
- Integrate these data in understanding the process of building functional complexity in the RNA world and the evolutionary ecology of molecules at the origin of life.

2. MATERIALS AND METHODS

2.1 Preparation of Round 18 Polymerase ribozyme and truncated derivatives

2.1.1 Preparation of template DNA for *in vitro* transcription of RNAs

The template DNA for *in vitro* transcription of Round 18 Polymerase ribozyme (R18) was designed to include a T7 promoter followed by the sequence for the ribozyme. The sequence of the ribozyme was selected from (Johnston et al., 2001). The double stranded template DNA for R18 was generated by primer extension of overlapping oligonucleotides (R18-F and R18-R) as shown in Figure 2.1. Oligonucleotides R18-F and R18-R were synthesised by standard phosphoramidite chemistry (Integrated DNA Technologies, USA). Briefly, the primer extension reaction was performed as a PCR using *Pfu* DNA polymerase with 0.5 pmoles of each R18-F and R18-R as forward and reverse primers in a total volume of 50 μ l. PCR was carried out with the following thermocycling conditions: Initial denaturation at 95 °C for 5 mins, 30 cycles of [denaturation at 95 °C for 1 min, annealing at 68 °C for 45 sec, extension at 72 °C for 45 sec], and final extension at 72 °C for 10 mins. The amplicon (206 nt) was purified to homogeneity on a 2.5% agarose gel (Lonza, USA) using the NucleoSpin Extract II kit (Macherey-Nagel, Germany). A single adenine was added to the 3' end of the eluted product using Taq DNA polymerase (Thermo Scientific, USA) and purified using the NucleoSpin Extract II kit (Macherey-Nagel, Germany). The purified DNA was ligated into pTZ57R/T vector using the InsTAclone PCR cloning kit (Thermo Scientific, USA) according to the manufacturer's instructions. The ligated product was transformed into competent DH5 α cells and analysed

for white colonies. Plasmid DNA was extracted from single bacterial colony using the NucleoSpin Plasmid extraction kit (Macherey-Nagel, Germany). The putative clone was verified by restriction digestion using *Pvu II* restriction enzyme (Thermo Scientific, USA). The plasmid construct containing the insert of expected size was sent for sequence analysis to Inqaba Biotechnology, South Africa. The sequenced plasmid construct was further used to amplify templates for *in vitro* transcription of R18 and its 3' truncated derivatives using the primer sets as shown in the Figure 2.2. The primers were synthesised by standard phosphoramidite chemistry (Integrated DNA Technologies, USA). Briefly, the amplification was carried out as a PCR using *Pfu* DNA polymerase with 25 pmoles of each forward and reverse primers in a total volume of 50 μ l. The following thermocycling conditions were used: Initial denaturation at 95 °C for 5 mins, 30 cycles of [denaturation at 95 °C for 1 min, annealing at 58 °C for 30 sec, extension at 72 °C for 30 sec], and final extension at 72 °C for 10 mins. All the amplified templates were purified to homogeneity on a 2.5% agarose gel (Lonza, USA) using the NucleoSpin Extract II kit (Macherey-Nagel, Germany).

2.1.2 In vitro transcription and purification of RNA products

The purified templates were used for *in vitro* transcription of Round 18 polymerase ribozyme (R18) and its 3' truncated RNA molecules (R18-T1, R18-T2, R18-T3, and R18-T4). The sequence details of all the RNA molecules are given in Figure 2.2. RNA was transcribed using the Megashortscript T7 transcription kit (Ambion, USA) according to the manufacturer's instructions (described in Appendix A). The transcribed RNA was electrophoresed through an 8% polyacrylamide-8M urea gel (described in Appendix A). The full length transcripts were excised from the gel and purified to homogeneity using the electro-elution protocol (described in Appendix B). The integrity of the purified RNA was

verified on 8% polyacrylamide-8M urea gel. The concentration of RNA was measured by the absorbance at 260 nm using a Nanodrop spectrophotometer ND-1000 (Thermo Scientific, USA). The purified RNA molecules were then used for self-ligation assays.

2.2 Design of the Oligonucleotide substrates

Oligonucleotide substrates (35 nt in length) were synthesised by standard phosphoramidite chemistry (Integrated DNA Technologies, USA). Sequences were DNA-RNA chimaeras with four ribonucleotides at the 3' end. The sequences of the designed substrates are given in Table 2.1. The nucleotides at the 3' end of the substrates were mutated (from position 19-34). The last ribonucleotide at the 3' end involved in the ligation junction remained constant. The 5' end of the substrates (from position 1 to 18) was also kept constant in all the substrates (except four substrates; 2, 3, 4 and 5) since it was used as the generic primer binding region for the detection of catalytic activity of the RNAs by PCR. The substrates had no experimentally designed region complementary to the RNAs used in the study. The synthesised substrates were dissolved in nuclease free water (Sigma-Aldrich, USA) to a stock concentration of 100 μ M.

Materials and Methods

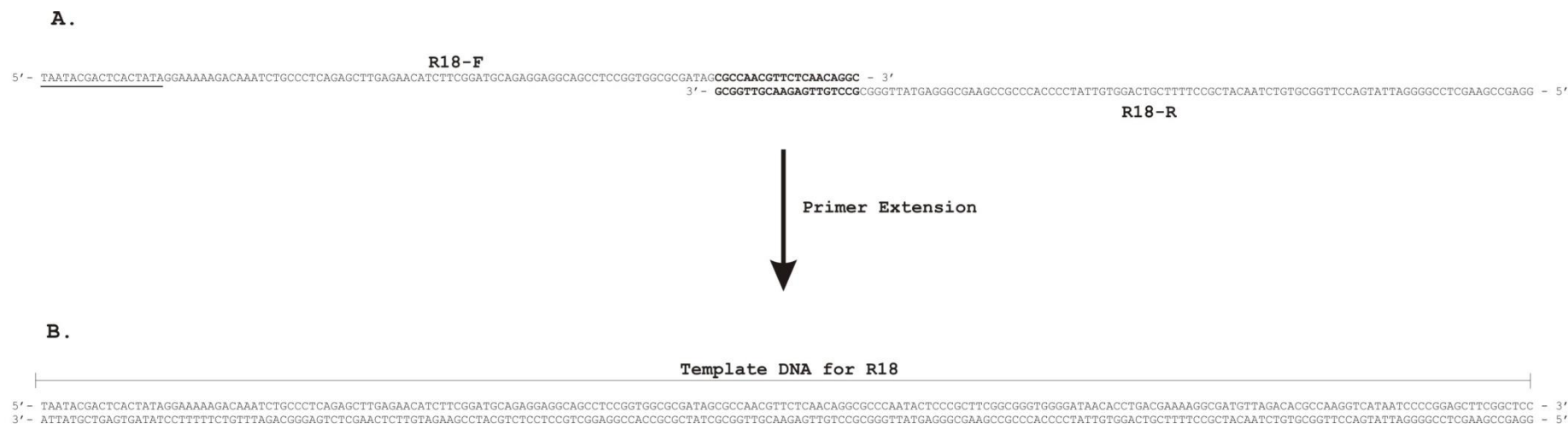


Figure 2.1: Synthesis of the template DNA sequence for R18 polymerase.

A. Oligonucleotides R18-F and R18-R were designed such that they were complementary in their 3' end (shown in bold) and R18-F included a T7 promoter sequence at the 5' end (underlined). **B.** The full length template DNA sequence for in vitro transcription of Round 18 polymerase ribozyme (R18) was generated by primer extension of the two overlapping oligonucleotides.

Materials and Methods

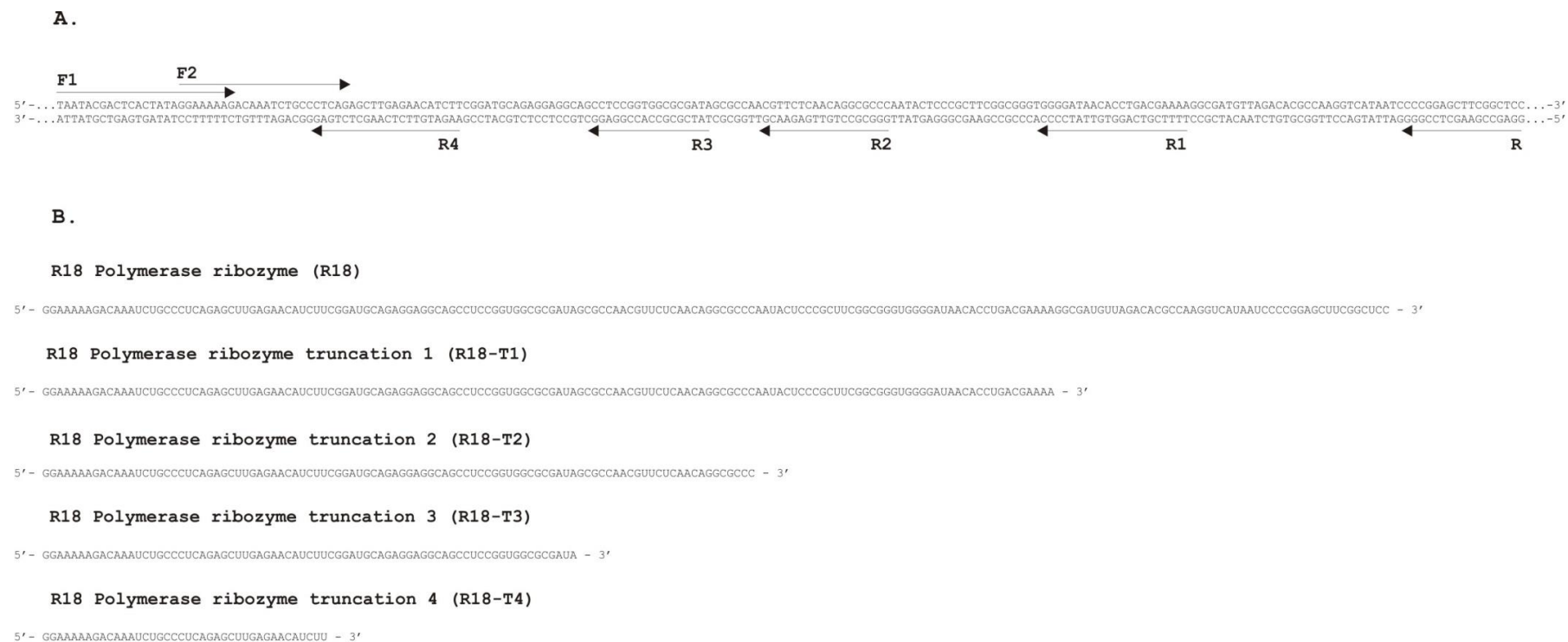


Figure 2.2: Synthesis of R18 polymerase and its truncated RNA sequences.

A. Template DNA for Round 18 polymerase ribozyme (R18) cloned and sequenced in pTZ57R/T vector (vector shown with dotted lines). The plasmid was used for amplification of template DNA sequences for *in vitro* transcription of R18 and its truncations. Primers F1 and R were used for amplification of template for R18, primers F1 and R1 were used for amplification of template for R18-T1, primers F1 and R2 were used for amplification of template for R18-T2, primers F1 and R3 were used for amplification of template for R18-T3, primers F1 and R4 were used for amplification of template for R18-T4. **B.** RNA sequences of R18 and its truncations (shown in 5' to 3' direction) after *in vitro* transcription.

Substrates	Sequence in 5' to 3' direction
Substrate 1	CTC GAC GTC AGC CTG GAC TAA TAC GAC TCA CUA UA
Substrate 2	GTC AAC TTC CGC ATG AAC GAA TAC TAC GCA CUA AA
Substrate 3	CAC GAC GAC AAC CTG GTC TAA TAC GCC TCA CGA UA
Substrate 4	CTG GAT GTA AGT CTT GAA TAT ATG GAA TCG CUC GA
Substrate 5	TAA TAC TCA TAA CGA CTA CAT GGA CCT CGC CUC AA
Substrate 6	CTC GAC GTC AGC CTG GAC TAA TAC TAA AAA CUA UA
Substrate 7	CTC GAC GTC AGC CTG GAC TAT ATG GAA TCG CUC GA
Substrate 8	CTC GAC GTC AGC CTG GAC CAT GGA CCT CGC CUC AA
Substrate 6a	CTC GAC GTC AGC CTG GAC TAA TAC TAT TTA CUA UA
Substrate 6b	CTC GAC GTC AGC CTG GAC TAA TAC TAG GGA CUA UA
Substrate 6c	CTC GAC GTC AGC CTG GAC GGG GGC TAG GGA CUA UA
Substrate 7a	CTC GAC GTC AGC CTG GAC TAT ATG GAC TCA CUA UA
Substrate 7b	CTC GAC GTC AGC CTG GAC TAA TAC GAA TCG CUC GA
Substrate 8a	CTC GAC GTC AGC CTG GAC CAT GGA GAC TCA CUA UA
Substrate 8b	CTC GAC GTC AGC CTG GAC TAA TAC CCT CGC CUC AA
Substrate G1	CTC GAC GTC AGC CTG GAC CCC CGC GAC TCC CUC CA
Substrate G2	CTC GAC GTC AGC CTG GAC CCC CGG CTG AGC CUC CA
Substrate G3	CTC GAC GTC AGC CTG GAC GGG GCC GAC TCC CUC UA
Substrate G4	CTC GAC GTC AGC CTG GAC GGG GCG CTG AGC CUC UA
Substrate G5	CTC GAC GTC AGC CTG GAC CGG CGC GAC TCT CUU UA
Substrate G6	CTC GAC GTC AGC CTG GAC CGG CGG CTG AGT GUU UA
Substrate G7	CTC GAC GTC AGC CTG GAC GCG CCT ATA AGG GUG CA
Substrate 6d	CTC GAC GTC AGC CTG GAC GGG GGC GAC TCA CUA UA
Substrate 6e	CTC GAC GTC AGC CTG GAC CCC CCC TAC CCA CUA UA

Table 2.1: Sequences of the oligonucleotide substrates.

Substrate sequences are shown in 5' to 3' direction. All the sequences were DNA-RNA chimaeras with four ribonucleotides at the 3' end (shown in bold).

2.3 Ribozyme self-ligation assay

The purified RNAs were assayed for self-ligation activity with each of the designed oligonucleotide substrate. The activity was assayed in a reaction buffer composed of 25 mM MgCl₂ (Merck, Germany), 50 mM KCl (Merck, South Africa), 4 mM DTT (Calbiochem, USA), 50 mM EPPS pH 8.2 (Sigma-Aldrich, USA) in nuclease free water (Sigma-Aldrich, USA). RNA (2 μM final concentration) was incubated in nuclease free water (Sigma-Aldrich, USA) at 80°C for 1 min, then cooled to 37°C for 5 mins. This was followed by simultaneous addition of reaction buffer and oligonucleotide substrate (5 μM final concentration) and the reaction was incubated in a total volume of 20 μL at 37°C for 40 mins. At the end of incubation period, 25 pmol of the primer complementary to the 3' end of the RNA (given in Table 2.2), 0.4 mM of each dNTP (Thermo Scientific, USA), 200 units Superscript III reverse transcriptase (Invitrogen, USA) were added to the above reaction mixture and incubated in a total volume of 25 μl at 55°C for 30 mins. After incubation, 5 μl was removed from the reaction and amplified by PCR using primers complementary to the 5' end of the oligonucleotide substrate and 3' end of the RNA template (given in Table 2.2). Briefly, the PCR was carried out using *Pfu* DNA polymerase with 25 pmoles of each forward and reverse primers in a total volume of 50 μl. The following thermocycling conditions were used: Initial denaturation at 95 °C for 5 mins, 30 cycles of [denaturation at 95 °C for 1 min, annealing at 58 °C for 30 sec, extension at 72 °C for 30 sec], and final extension at 72 °C for 10 mins. The schematic representation of the assay is shown in Figure 2.3. Negative controls for the ribozyme assay were set up by incubating the RNA without the addition of oligonucleotide substrate in the reaction buffer at 37 °C for 40 mins and then reverse transcribed at 55 °C for 30 mins. The control reactions were PCR amplified with primers used for detection of self-ligation activity of ribozymes (given in Table 2.2).

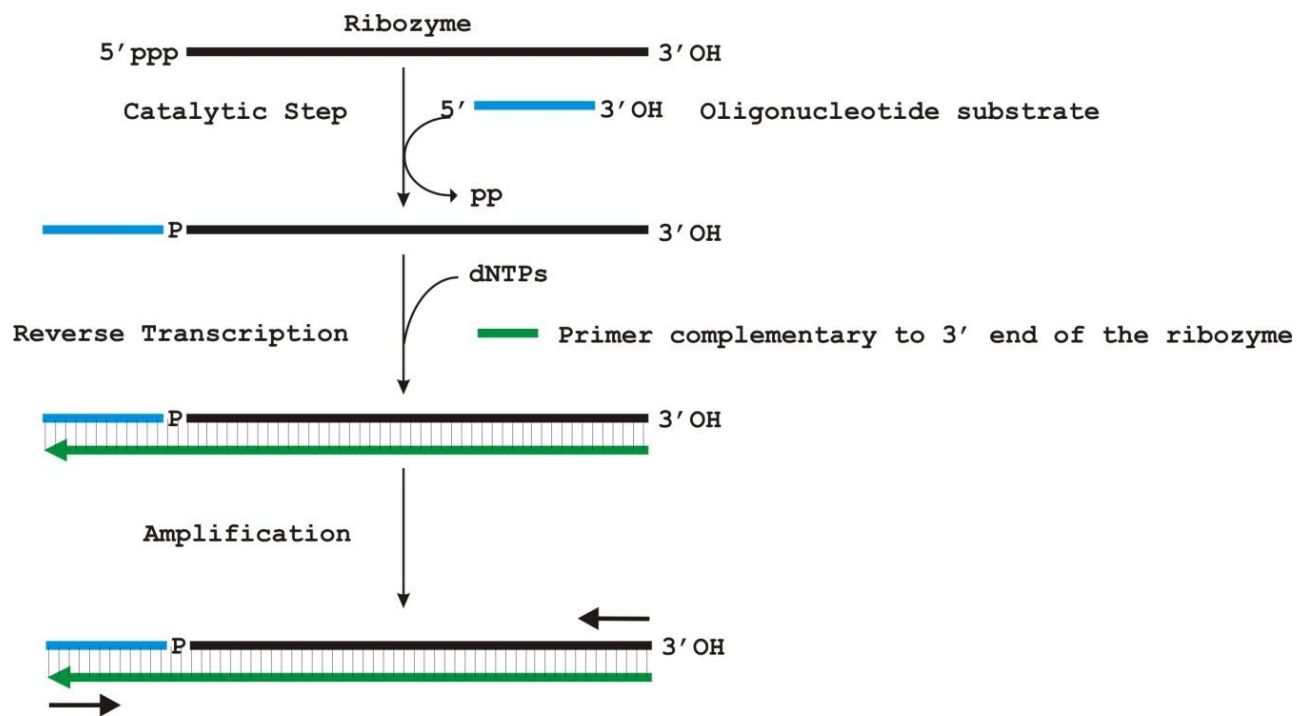


Figure 2.3: Schematic representation of the ribozyme self-ligation assay.

The black solid line represents the RNA (shown in 5'-3' direction). The blue solid line represents the oligonucleotide substrate used in the assay (shown in 5' - 3' direction). The green solid arrow line represents the cDNA formed from the reverse transcription of the ligated product using the primer complementary to the 3' end of the ribozyme. The primers used to amplify the reverse transcribed product are shown as black solid arrows.

Additional negative controls were set up in which a purified random RNA (ORF region of pET 15b vector) was incubated with the substrates in the reaction buffer conditions used for the ribozyme assay. After the incubation, the reactions were reverse transcribed using the primer complementary to the 3' end of the random RNA and then PCR amplified using the primer complementary to the 5' end of the substrates and the 3' end of the RNA.

The PCR products were electrophoresed through a 2.5 % agarose gel (Lonza, USA) at 100 V. To assess the self-ligation activity of the RNA with the oligonucleotide substrate, a reference DNA amplicon was run on the same gel. The reference was set up by PCR amplification of the DNA sequence corresponding to the sequence of the RNA (as shown in Figure 2.2). PCR products amplified using primers F2 and R, primers F2 and R1, primers F2 and R2, primers F2 and R3, and primers F2 and R4 corresponding to the sequence of R18 RNA, R18-T1 RNA, R18-T2 RNA, R18-T3 RNA and R18-T4 RNA, respectively, were used as reference DNAs. The self-ligation activity of the RNA was assessed by comparing the size of the amplicon after the assay to the respective reference DNA. A 35bp difference in size was indicative of a ligation reaction. The PCR products indicating positive activity were sequenced for confirmation.

RNA	Primer complementary to 5' end of substrate (Forward primer)	Primer complementary to 3' end of the RNA template (Reverse Primer)
R18	CTCGACGTCAGCCTGGAC	GGAGCCGAAGCTCCGGG
R18-T1	CTCGACGTCAGCCTGGAC	TTTTCGTCAGGTGTTATCCCC
R18-T2	CTCGACGTCAGCCTGGAC	GGGCGCCTGTTGAGAACG
R18-T3	CTCGACGTCAGCCTGGAC	TATCGCGCCACCGGAGG
R18-T4	CTCGACGTCAGCCTGGAC	AAGATGTTCTCAAGCTCTGAG

Table 2.2: Sequences of the primers used for detection of self-ligation activity of the ribozymes

Primers (shown in 5' to 3' direction) used for detection of self-ligation activity of R18 ribozyme and its truncated RNAs with the substrates (except substrates 2, 3, 4, and 5). The self-ligation activity of RNAs with the substrates 2, 3, 4 and 5, was detected by using forward primers 5'-GTCAACTTCCGCATGAAC-3' (complementary to 5'end of substrate 2), 5'-CACGACGACAACCTGGTC-3' (complementary to 5'end of substrate 3), 5'-CTGGATGTAAGTCTTGAA-3' (complementary to 5'end of substrate 4), 5'-TAATACTCATAACGACTA-3' (complementary to 5'end of substrate 5) and reverse primers for respective RNAs as given in the table.

2.4 Sequence analysis of the ligation product

For sequence analysis, the PCR product indicating positive activity was purified to homogeneity on a 2.5% agarose gel (Lonza, USA) using the NucleoSpin Extract II kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. A single adenine was added to the 3' end of the eluted product using Taq DNA polymerase (Thermo Scientific, USA) and purified using the NucleoSpin Extract II kit (Macherey-Nagel, Germany). The purified DNA was ligated into pTZ57R/T vector using the InsTAclone PCR cloning kit (Thermo Scientific, USA) according to the manufacturer's instructions. The ligated product was transformed into competent DH5 α cells and analysed for white colonies. Plasmid DNA was extracted from single bacterial colony using the NucleoSpin Plasmid extraction kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. The putative clone was verified by restriction digestion using fast digest *Pvu II* restriction enzyme (Thermo Scientific, USA). The plasmid construct containing the insert of expected size was sent for sequence analysis to Inqaba Biotechnology, South Africa. The self-ligation reaction was confirmed by sequence alignment using EMBOSS-needle algorithm. An alignment of the cloned sequence was performed with the sequence expected for a positive result.

2.5 Analysis of the substrates with respect to the ligation activity of RNA

Comparative analysis of the substrates with respect to the ligation activity of RNA was performed using MEME – Multiple Em for Motif Elicitation Suite 4.10.0 (Bailey et al., 2009). Substrate sequences were grouped into two categories; substrates ligated by the RNA and substrates not ligated by the RNA. The grouping for MEME was done using all

the substrate sequences except substrates 2, 3, 4 and 5. Only 17 positions which were variable at the 3' end were used for analysis since all the substrates had the same sequence in the first 18 positions. The group of substrate sequences that were ligated by the RNA were aligned in the 5' to 3' direction and analysed for sequence patterns. The sequence pattern which was present in the majority of the grouped sequences was chosen. Similarly, the group of substrates which were not ligated by the RNA were analysed and a sequence pattern was chosen. Based on the probability matrix, the two sequence patterns were compared for nucleotides that occurred at a probability of $\geq 50\%$. The probability matrix is given in Appendix D. The above procedure was done for all the RNA constructs.

2.6 Structures and complexity predictions

The structures of the substrates were predicted using Mfold RNA folding form (Zuker, 2003). The structures and the stability of the RNA molecules were predicted using RNA fold (Zuker and Stiegler, 1981). Structural complexity of the RNA molecules was determined by the minimum free energy structure and the predicted thermodynamic stability (Gibbs free energy).

2.7 Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of ribozyme self-ligation activity

Reverse transcription-quantitative real time PCR (RT-qPCR) has been widely adopted for highly specific detection and quantification of RNAs. Specificity is conferred at three levels: via two PCR primers and a probe. The present study employed the technique for the analysis of ribozyme self-ligation activity (described in Figure 2.4). The following procedure was followed:

2.7.1 Design of Probe

A TaqMan probe was synthesised (Life technologies, USA) specific to a region common in the cDNA sequences of all the ribozymes; R18, R18-T1, R18-T2, R18-T3 and R18-T4. The sequence of the designed probe was 5'-GGAAAAGACAAATCTGCCC-3'. The probe sequence included a fluorescent dye 6-carboxyfluorescein (FAM) on the 5' end. The 3' end consisted of non-fluorescent quencher (NFQ) conjugated to major groove binder (MGB) moiety. The signals from FAM are quenched by NFQ. The MGB moiety increases the T_m of the probe and stabilises probe–target hybrids.

2.7.2 Preparation of samples for standard curves

Sequenced plasmid construct cloned with ligation product of each ribozyme with substrate 1 was PCR amplified using the respective forward and reverse primers as given in Table 2.2. Briefly, PCR was carried out using *Pfu* DNA polymerase with 25 pmoles of each forward and reverse primers in a total volume of 50 μ l with the following thermocycling conditions: Initial denaturation at 95 °C for 5 mins, 30 cycles of [denaturation at 95 °C for 1 min, annealing at 58 °C for 30 sec, extension at 72 °C for 30 sec], and final extension at 72 °C for 10 mins. The amplified product was run on a 2.5% agarose gel (Lonza, USA) and was purified to homogeneity using the NucleoSpin Extract II kit (Macherey-Nagel, Germany). It was sequenced for confirmation. The concentration of the gel purified and sequenced PCR product was determined using the Qubit dsDNA HS Assay kit according to the manufacturer's instructions (Life technologies, USA). Serial dilutions of the DNA were prepared in nuclease free water to obtain copy numbers

corresponding to a final concentration as given in Table 2.3. The amounts of DNA needed for obtaining X number of DNA copies was calculated as follows:

Amount of DNA (in gms) = Moles (X number of Copies / 6.023×10^{23}) x Molecular weight of DNA (gms/mol)

A real time quantitative PCR of the known DNA copies of the ribozymes R18, R18-T1, R18-T2, R18-T3 and R18-T4 ligated to substrate 1 was performed and the corresponding standard curves were generated.

2.7.3 Real-time quantitative PCR of samples and generation of standard curves

Real-time quantitative PCR was performed as follows: One microlitre from each dilution sample was added to 19 μ L PCR set up. The set up consisted of 12.5 pmoles of forward primer (specific to the 5' end of substrate), 12.5 pmoles of reverse primer specific to a region common in all the ribozymes template DNA sequences (i.e. primer complementary to the 3' end of R18-T4 ribozyme template), and 5 pmoles of the designed probe. The sequences of the primers are given in Table 2.2. A control reaction with no DNA was also set up. The reactions were amplified using 10 μ l TaqMan gene expression master mix (Applied Biosystems, USA) on the Roche Lightcycler v.2 (Roche Diagnostics, GmbH, Germany) with following thermal cycling conditions; Step 1: 50 °C for 2 mins, Step 2: 95 °C for 10 mins, Step 3: 95 °C for 15 sec, Step 4: 60 °C for 1 min. Step 3 and 4 were repeated for 40 cycles. Fluorescence was measured at 530 nm. The CP values of known copies of DNA were obtained. The CP value corresponds to the cycle number at which there is first detectable increase in fluorescence as a result of cleavage of probe during polymerisation reaction (described in Figure 2.4). A standard curve was generated using base 10 log of initial target copy number versus corresponding CP value (given in

Appendix E). The standard curve was used to quantify unknown cDNA copies produced from the ribozyme reaction product at different time intervals as described in the next section.

Sample Name	No. of copies per μl
E	1×10^7
F	1×10^6
G	1×10^5
H	1×10^4
I	1×10^3
J	1×10^2
K	1×10^1

Table 2.3: The copies of DNA prepared for generating the standard curves.

Samples E-K are the dilutions of DNA prepared with the corresponding number of copies per μl

2.7.4 Determination of the rate of ribozyme self-ligation activity

The rate of reaction of ribozymes was determined using reverse transcription-quantitative real time PCR (RT-qPCR). A time course analysis for the reactions of ribozymes R18, R18-T1, R18-T2, R18-T3, and R18-T4 that were positive with the substrates 1, 6, 7, 8, 6a, 6b, 6c, 7a, 7b, 8a, 8b was done. The rates of the ribozyme activity were not studied with substrates 2, 3, 4 and 5 since the primer sequences for detection of the ligation product were different to the ones used for generation of the standard curve.

Self-ligation reactions were performed in a reaction buffer composed of 25 mM MgCl_2 (Merck, Germany), 50 mM KCl (Merck, South Africa), 4 mM DTT (Calbiochem,

USA), 50 mM EPPS pH 8.2 (Sigma-Aldrich, USA) in nuclease free water (Sigma-Aldrich, USA). The reactions were set up by incubating purified ribozyme (1 μ M final concentration) in nuclease free water (Sigma-Aldrich, USA) at 80°C for 1 min, then cooled to 37°C for 5 mins. This was followed by simultaneous addition of reaction buffer and oligonucleotide substrate (100nM final concentration) in a total volume of 20 μ L. Incubation was performed at 37°C with eight time points ranging from 5 mins to 40 mins that were set up in different reaction tubes. After the incubation time, the reactions were stopped by snap freezing them in liquid nitrogen. After completion of all the time points, the reaction tubes were simultaneously transferred on ice. A mixture containing 25 pmoles of the primer complementary to the 3' end of the RNA (given in Table 2.2), 0.4 mM of each dNTP (Thermo Scientific, USA), and 200 units Superscript III reverse transcriptase (Invitrogen, USA) were simultaneously added to the above reaction mixture and incubated in a total volume of 25 μ l at 55°C for 30 mins. In the whole procedure, two negative controls were set up; one of the controls was set up by incubating the ribozyme without the addition of oligonucleotide substrate in the reaction buffer at 37°C for 40 mins and reverse transcribed at 55°C for 30 mins. The other control was set up by incubating the ribozyme with the addition of oligonucleotide substrate in the reaction buffer at 37°C for 40 mins and not reverse transcribed. After incubation, one microlitre was removed from each reaction was added to 19 μ L PCR setup and amplified as described previously in section 2.7.3. All the PCR assays included a control reaction with no DNA. A sample of standard curve with known DNA copy number was also included for calibration of each run. The fluorescence was measured at 530 nm. The CP value of the reaction at each of the eight time points was obtained. The number of cDNA copies in the reaction was quantified using the standard curve. This was done as follows: the CP value in the standard curves is based on fluorescence from double stranded DNA in the first cycle. The CP value obtained from the

ribozyme reaction was based on fluorescence from cDNA (single stranded) in the first cycle. Thus, the fluorescence obtained from double stranded DNA after x number of cycles will be double than the fluorescence obtained from single stranded DNA (compared in Figure 2.5). Therefore, for an equivalent CP value obtained, the number of cDNA copies present in the ribozyme reaction was quantified to be one half of the copies calculated from the standard curve. A graph of cDNA copies quantified in the reaction was plotted versus duration of incubation using Microsoft Excel. The rate of reaction was determined from the slope of the curve and was given as the copies of ligated product cDNA formed per minute.

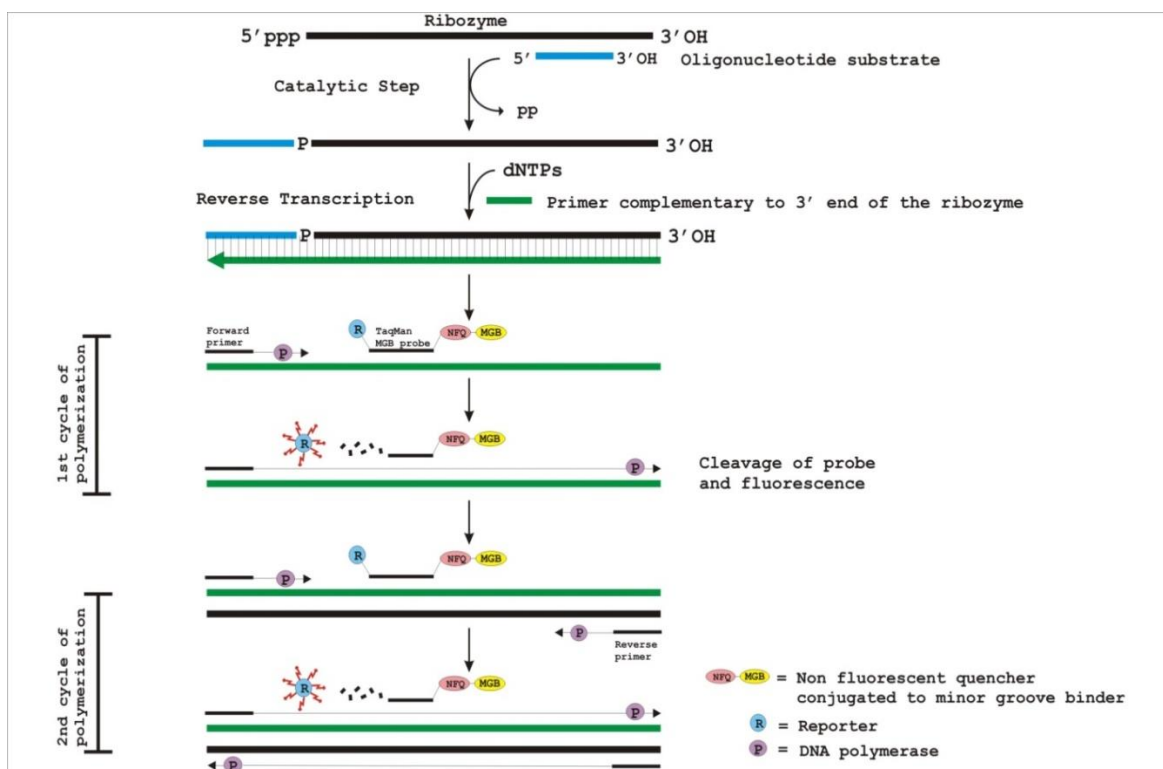


Figure 2.4: Schematic representation of the assay performed for quantitative analysis of the ribozyme’s ligation activity.

The ribozymes reaction products were reverse transcribed and PCR amplified using a probe designed to bind a specific region common in the reverse transcribed sequences of the ribozyme ligation product. When the probe is intact, signals from reporter dye FAM are quenched by NFQ. During the PCR reaction, the 5’ nuclease activity of DNA polymerase cleaves the probe separating FAM and NFQ, resulting in fluorescence of FAM.

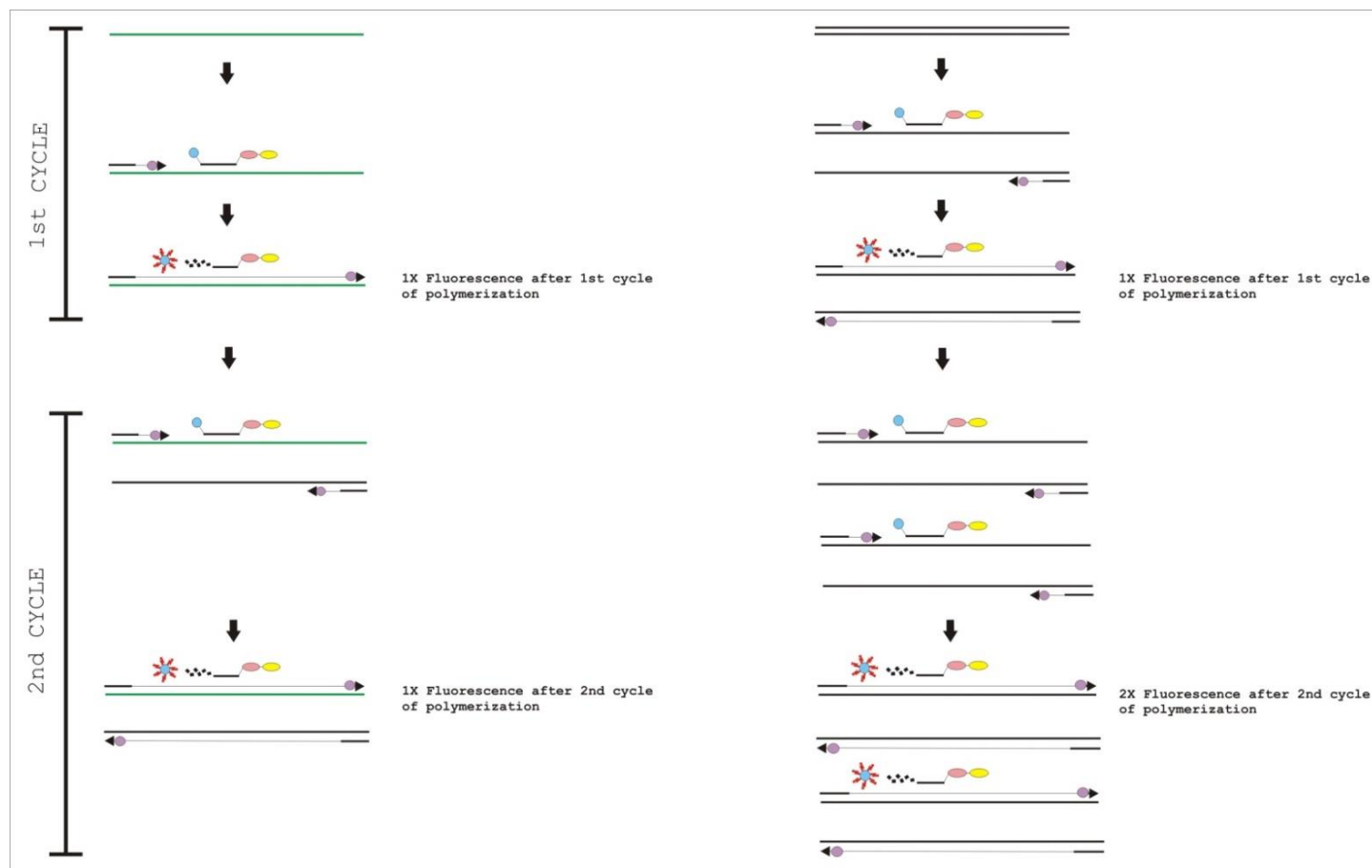


Figure 2.5: Comparison of the fluorescence from ribozyme self-ligation assay and the corresponding standard sample.

Quantitative comparison of the fluorescence after 2 cycles of polymerisation using real time PCR from cDNA of ribozyme's self-ligation reaction (left) and the corresponding dsDNA used as standard (right).

3. RESULTS

3.1 Overview

The main questions that have been posed in this study concern the processes that may have been important for the emergence of a minimal polymerase ribozyme and diverse complex molecules from short RNA oligomers. A polymerase ribozyme (R18 polymerase) was selected as the model for the study. This polymerase is composed of a Class I ligase core (100 nucleotides; developed by Bartel and group) and an additional domain appended to the 3' end of the core. The first objective was to determine a possible route by which the polymerase and its ligase core emerged. To study this process, the polymerase and its smaller components were prepared and examined for their ability to ligate 35 nucleotides long substrates to their own end (self-ligation function). The substrates were variable and not specifically designed for any specific base pairing with the RNA constructs. The self-ligation function of the molecules was analysed in terms of their flexibility in ligating different substrates and the efficiency. In addition, the sequence patterns in substrates were also analysed using MEME and their relation to the ligation function. The correlations between traits like size, functional flexibility and efficiency of all the RNA constructs were studied to understand their role in the emergence of a polymerase and stability of RNA based life.

3.2 RNA preparation for Self-ligation assays

The structure reported by (Johnston et al., 2001) was used as a reference for selecting the size of the smaller components. Since the catalytic core in the polymerase essential for

Results

ligation function is at the 5' end, the polymerase was truncated from the 3' end. The truncations were based on reduction of the stem-loops and the structural complexity of the polymerase in a step-wise fashion. The study employed use of five RNA constructs (R18, R18-T1, R18-T2, R18-T3 and R18-T4). The structures of these constructs (only for illustration purpose) are given in Figure 4.1. The smallest construct was the 5' region of the active ligase core; R18-T4 RNA (a 40 nucleotide long molecule with a simple stem-loop structure as predicted by RNA fold, Table 3.14).

Purified DNA sequences for R18 and its truncated derivatives were prepared and used for *in vitro* transcription of R18 RNA, R18-T1 RNA, R18-T2 RNA, R18-T3 RNA, and R18-T4 RNA. The full length transcripts of size 189 nt (R18 RNA), 142 nt (R18-T1 RNA), 100 nt (R18-T2 RNA), 75 nt (R18-T3 RNA), 40 nt (R18-T4 RNA) were purified to homogeneity. Figure 3.1 shows the purified RNAs used for self-ligation assays.

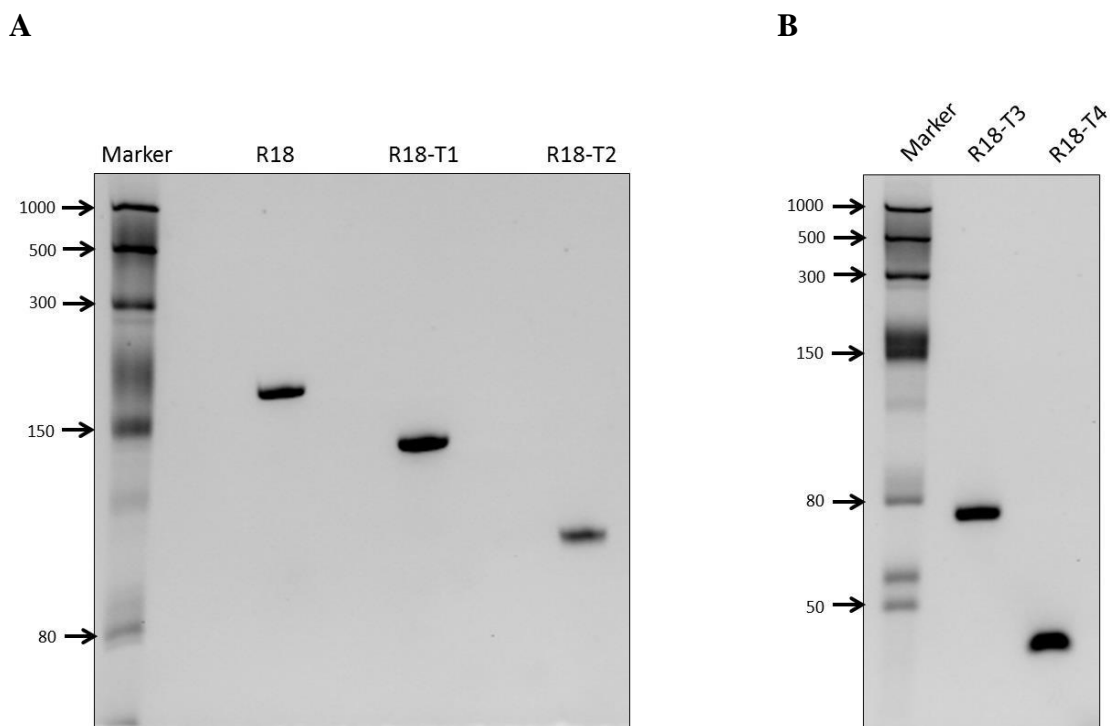


Figure 3.1: Purified RNAs used for self-ligation assays.

A: Purified R18 RNA, R18-T1 RNA and R18-T2 RNA on 8% polyacrylamide-8M urea gel. B: Purified R18-T3 RNA, R18-T4 RNA on 8% polyacrylamide-8M urea gel.

3.3 Self-Ligation activity of RNAs

The self-ligation assay of the RNA constructs was initiated with the DNA-RNA chimaera substrate 1 (5'- CTC GAC GTC AGC CTG GAC TAA TAC GAC TCA CUA UA - 3'). This substrate was chosen as it was used in some of the previous studies for the continuous *in vitro* evolution of more efficient ligases from the Class I ligase core (Wright and Joyce, 1997). In these studies, the catalytic core included a random segment at the 5' end for pairing with this substrate. The R18 polymerase is also composed of the Class I ligase core; hence, this substrate was used to initiate the study. The ligase core of the polymerase, however, lacks the 5' random segment (Johnston et al., 2001). The assays in this study examined the self-ligation activity of the polymerase and the smaller components in the absence of an explicitly designed base pairing with the substrate. The activity of the ribozymes constructs were also analysed with nucleotides variations at the 3' end of the substrate (given in Table 2.1). In a few substrates, the 5' end of the substrate was also varied for analysis. The rationale behind this approach was that the early stages of the RNA world were likely to have heterogeneous substrates. Assuming that the evolution proceeded from such an environment, the activity of the ribozymes might have been based on mechanisms that were independent of specific complementarity with the substrates. The reaction systems in this study investigated the self-ligation activity of the ribozymes with the substrates in the absence of a guided template.

Each purified RNA molecule was assayed for self-ligation activity with each of the designed oligonucleotide substrates. In the given conditions, the reaction products of the P^{32} labeled RNA with the substrates were not detected when directly analysed using phosphorimaging. The use of radio-isotopic labeling is a standard method for ribozyme assays and is sensitive for detection of low levels of activity (< 1% based on literature). Since, the products were not visible using this method, it was anticipated that they could be present in even lower levels due to low efficiency of the reactions. The RT-PCR based method offers detection sensitivity of femtograms (fg) of the transcripts and amplification sensitivity for up to 10 copies in the samples (Appendix E). Therefore, for greater sensitivity, this method was chosen to study the activity of the molecules. It involved the detection using a two-step process: reverse transcription of the reaction and then PCR amplification. The PCR products were electrophoresed through a 2.5% agarose gel. Positive activity was assessed by comparing the size of the amplicon after the assay to the respective reference DNA run on the same gel. A 35 bp difference in size was indicative of a ligation reaction. The PCR product indicating positive activity of the RNA molecule was gel purified, cloned into pTZ57R/T vector and sequenced. Self-ligation activity was confirmed by alignments of the cloned sequence (top sequence) with the sequence expected in case of a self-ligation activity of the RNA with the oligonucleotide substrate (bottom sequence). Sequence alignments are provided in Appendix C. In the control reactions, the RNA construct was incubated without any substrate; reverse transcribed and PCR amplified with the primer sets that were used to detect its activity with the substrates. All the RNA constructs failed to show any amplification in the control reactions (Figure 3.2 - 3.6). The additional negative controls that were set up using pET 15b RNA also failed to show any amplification and the gels were blank.

Each RNA construct self-ligated only some kinds of substrates, the results of which are given in the sections below (Figure 3.2 - 3.6). The reaction with the other substrates did not show any amplification and the gels were blank.

3.3.1. Self-Ligation activity of R18 RNA

R18 RNA showed positive activity with oligonucleotide substrates 1, 6, 6A, 6B, 7A, 2 and 3 only (Figure 3.2). An amplicon of size 224 bp as compared to the size of reference DNA (189bp) was indicative of self-ligation activity of R18 RNA. The amplicons were sequenced for confirmation of the ligation activity (given in Appendix C; C.1).

3.3.2. Self-Ligation activity of R18-T1 RNA

R18-T1 RNA showed positive activity with oligonucleotide substrates 1, 2, 3, 4, 6, 6A, 6B, 7, 7A and 8B only (Figure 3.3). An amplicon of size 177 bp as compared to the size of reference DNA (142 bp) was indicative of self-ligation activity of R18-T1 RNA. The amplicons were sequenced for confirmation of the ligation activity (given in Appendix C; C.2).

3.3.3. Self-Ligation activity of R18-T2 RNA

R18-T2 RNA showed positive activity with oligonucleotide substrates 1, 2, 3, 6A, 6B, 7, 7A, 7B, 8, 8A, 4 and 5 only (Figure 3.4). An amplicon of size 135 bp as compared to the size of reference DNA (100 bp) was indicative of self-ligation activity of R18-T2 RNA. The amplicons were sequenced for confirmation of the ligation activity (given in Appendix C; C.3).

3.3.4. Self-Ligation activity of R18-T3 RNA

R18-T3 RNA showed positive activity with oligonucleotide substrates 1, 2, 3, 6, 6A, 6B, 7, 7A, 7B, 8A, 8B, 4 and 5 only (Figure 3.5). An amplicon of size 110 bp as compared to the size of reference DNA (75 bp) was indicative of self-ligation activity of R18-T3 RNA. The amplicons were sequenced for confirmation of the ligation activity (given in Appendix C; C.4).

3.3.5. Self-Ligation activity of R18-T4 RNA

R18-T4 RNA showed positive activity with oligonucleotide substrates 1, 6, 6A, 6B, 7, 2, 3, 7A, 4, 5 7B, 8A, 8B only (Figure 3.6). An amplicon of size 75 bp as compared to the size of reference DNA (40 bp) was indicative of self-ligation activity of R18-T4 RNA. The amplicons were sequenced for confirmation of the ligation activity (given in Appendix C; C.5).

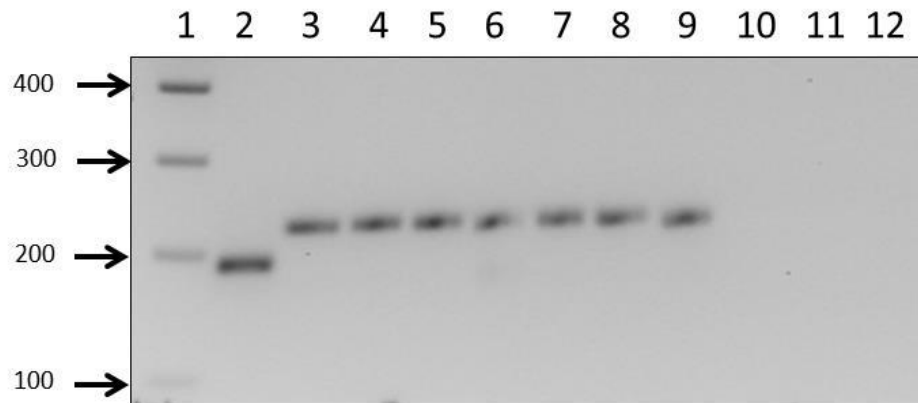


Figure 3.2: Self-ligation activity of R18 RNA.

The PCR products from the self-ligation reactions of the R18 RNA with the substrates are shown on a 2.5% agarose gel. Lane 1 shows the DNA marker. Lane 2 shows the reference DNA for assessing R18 RNA self-ligation activity. Lanes 3-9 show the R18 RNA self-ligation activity with substrate 1, substrate 6, substrate 6A, substrate 6B, substrate 7A, substrate 2 and substrate 3, respectively. Lanes 10-12 show the control reactions.

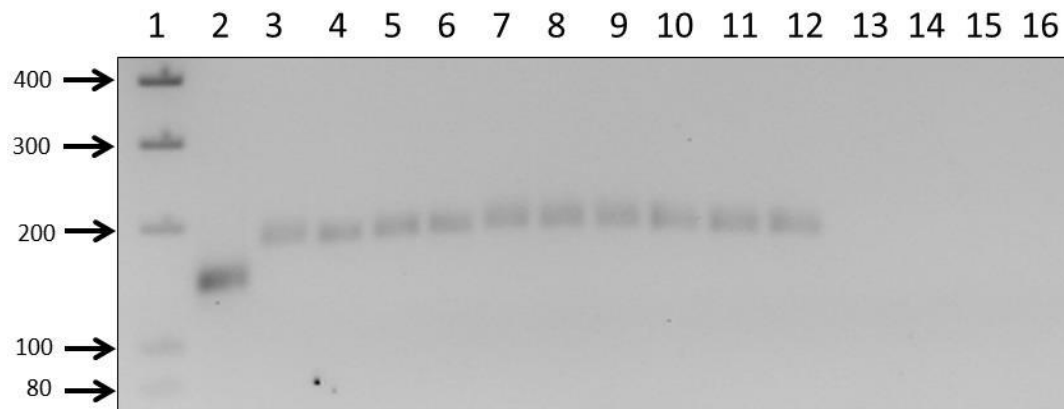


Figure 3.3: Self-ligation activity of R18-T1 RNA.

The PCR products from the self-ligation reactions of the R18-T1 RNA with the substrates are shown on a 2.5% agarose gel. Lane 1 shows the DNA marker. Lane 2 shows the reference DNA for assessing R18-T1 RNA self-ligation activity. Lanes 3-12 show the R18-T1 self-ligation activity with substrate 1, substrate 2, substrate 3, substrate 4, substrate 6, substrate 6A, substrate 6B, substrate 7, substrate 7A, and substrate 8B, respectively. Lanes 13-16 show the control reactions.

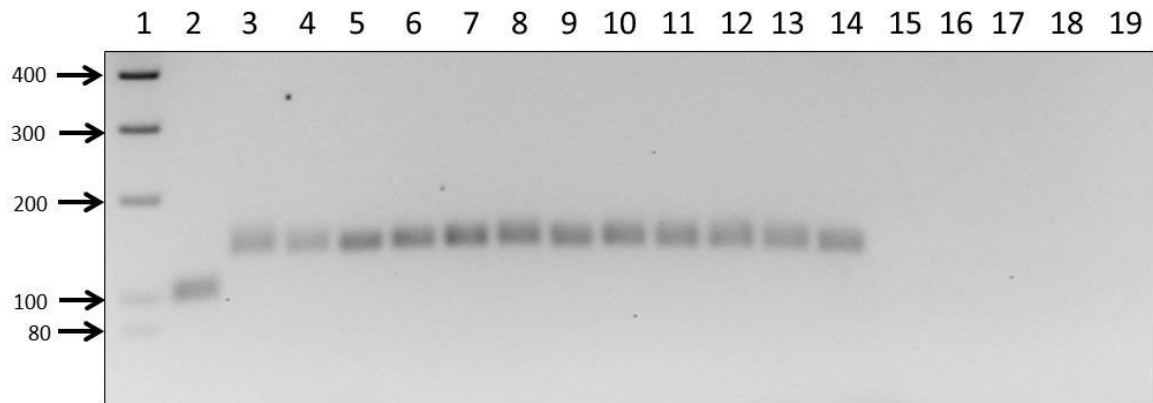


Figure 3.4: Self-ligation activity of R18-T2 RNA.

The PCR products from the self-ligation reactions of the R18-T2 RNA with the substrates are shown on a 2.5% agarose gel. Lane 1 shows the DNA marker. Lane 2 shows the reference DNA for assessing R18-T2 RNA self-ligation activity. Lanes 3-14 show the R18-T2 RNA self-ligation activity with substrate 1, substrate 2, substrate 3, substrate 6A, substrate 6B, substrate 7, substrate 7B, substrate 7A, substrate 8, substrate 8A, substrate 4, and substrate 5, respectively. Lanes 15-19 show the control reactions.

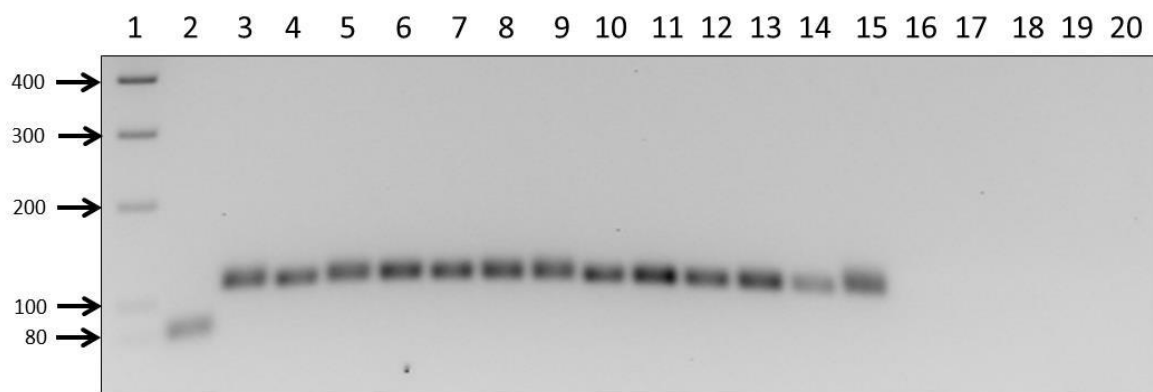


Figure 3.5: Self-ligation activity of R18-T3 RNA.

The PCR products from the self-ligation reactions of the R18-T3 RNA with the substrates are shown on a 2.5% agarose gel. Lane 1 shows the DNA marker. Lane 2 shows the reference DNA for assessing R18-T3 RNA self-ligation activity. Lanes 3-15 show R18-T3 RNA self-ligation activity with substrate 1, substrate 2, substrate 3, substrate 6, substrate 6A, substrate 6B, substrate 7, substrate 7A, substrate 7B, substrate 8A, substrate 8B, substrate 4, and substrate 5, respectively. Lanes 16-20 show the control reactions.



Figure 3.6: Self-ligation activity of R18-T4 RNA.

The PCR products from the self-ligation reactions of the R18-T4 RNA with the substrates are shown on a 2.5% agarose gel. Lane 1 shows the DNA marker. Lane 2 shows the reference DNA for assessing R18-T4 RNA self-ligation activity. Lanes 3-15 show R18-T4 RNA self-ligation activity with substrate 1, substrate 6, substrate 6A, substrate 6B, substrate 7, substrate 2, substrate 3, substrate 7A, substrate 4, and substrate 5, substrate 7B, substrate 8A, and substrate 8B, respectively. Lanes 16-20 show the control reactions.

3.4 Summary of RNA activity with the oligonucleotide substrates

- R18 RNA, R18-T1 RNA, R18-T2 RNA, R18-T3 RNA, and R18-T4 RNA were analysed for self-ligation activity with 24 oligonucleotide substrates. The RNAs demonstrated catalytic activity by ligating oligonucleotide substrates to their own 5' end. This was confirmed by analysing the reaction PCR products on an agarose gel (Figure 3.2 - 3.6), purification of the product and sequence analysis (Appendix C).
- There was no explicit experimentally designed base pairing between the ribozymes and the substrates. The catalytic activity observed was, therefore, based on interactions independent of experimentally defined base pairing between the molecules.
- The smallest functional truncation of R18 RNA which demonstrated self-ligation activity was R18-T4 RNA (40 nucleotides in size).
- The smallest ribozyme R18-T4 was general in its activity and was able to ligate 13 out of 24 different kinds of substrates to its own end. However, as the size and structural complexity of the R18-T4 ribozyme increased, there was a gradual decrease in the kinds of substrates ligated. The R18 ribozyme was more specific in its activity and ligated 7 out of 24 different kinds of substrates to its own end.
- With the set of substrates used in this study, the following three patterns were noteworthy (summarized in Table 3.1)
 - a. In the case of catalytic reactions of R18 RNA, R18-T1 RNA, R18-T2 RNA, R18-T3 RNA, and R18-T4 RNA with oligonucleotide substrates G1, G2, G3, G4, G5, G6, G7, 8, 6c, 6d and 6e, no self-ligation activity was observed (except substrate 8 was self-ligated by R18-T2 RNA). The region in the Table 3.1 highlighted in pink indicates this trend.

- b. In the case of catalytic reactions of R18 RNA, R18-T1 RNA, R18-T2 RNA, R18-T3 RNA, and R18-T4 RNA with oligonucleotide substrates 1, 2, 3, 6a, 6b, 7a, and 6, self-ligation activity was observed at all the complexity levels of RNA (except substrate 6 was not self-ligated by R18-T2 RNA). The region in the Table 3.1 highlighted in green indicates this trend.
- c. In the case of catalytic reactions of R18 RNA, R18-T1 RNA, R18-T2 RNA, R18-T3 RNA, and R18-T4 RNA with oligonucleotide substrates 4, 7, 5, 7b, 8a, and 8b, a trend in functional activity was observed. R18-T4 RNA demonstrated self-ligation activity with all the substrates. However, there was a gradual decrease in the kinds of substrates self-ligated with increase in size and structural complexity of R18-T4 RNA. The R18 RNA did not show self-ligation activity with any of these substrates. The region in the Table 3.1 highlighted in yellow indicates this trend.

Results






		1	2	3	6a	6b	7a	6	4	7	5	7b	8a	8b	8	G1	G2	G3	G4	G5	G6	G7	6c	6d	6e
R18		+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R18-T1		+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
R18-T2		+	+	+	+	+	+	-	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	-
R18-T3		+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
R18-T4		+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-

Table 3.1: Summary of the self-ligation activity of the RNAs with the substrates.

The RNA constructs (represented by rows) were assayed with the 24 substrates (represented by columns). The (+) sign denotes the presence of ligation and the (-) the absence. There were three different patterns observed from the ribozyme assays. The patterns are highlighted in green, yellow and pink and the details are discussed in the text.

3.5 Analysis of sequence patterns in the substrate sequences

The reactions of the ribozymes with the substrates were not experimentally designed for any specific base pairing, however, some substrates were ligated and others were not. The substrate sequences were analysed for nucleotide patterns that might have determined the ligation activity of the ribozymes. This was done using MEME, a tool which identifies motifs (sequence pattern) in a group of related DNA sequences. It represents motifs as position-dependent letter-probability matrices, which describe the probability of each possible letter at each position in the pattern. Based on the probability matrices, MEME generates a sequence logo which depicts the probability of nucleotides at each position on the aligned sequences. A sequence logo consists of a stack of letters at each position. The relative sizes of the letters indicate their frequency in the sequences. The total height of the letters depicts the information content of the position in bits.

All the substrate sequences that were used in the study (except substrates 2, 3, 4, and 5) were first verified for their variability with respect to each other. The variable region of all the substrates were aligned and analysed for a sequence pattern using MEME. Substrates 2, 3, 4 and 5 were not used in this analysis since their 5' end was also variable unlike the other substrates. In case of inclusion of these substrates for the analysis, MEME may not find one or more consensus sequences and/or generate a dysfunctional consensus sequence. It was found from the sequence logo (Figure 3.7) that the analysed region was not enriched in any one particular nucleotide (except position 32 and 35). This confirmed that mostly all the positions in the region were designed to be randomly variable (Figure 3.7). Position 35 was designed to be constant as it was involved in the ligation junction.

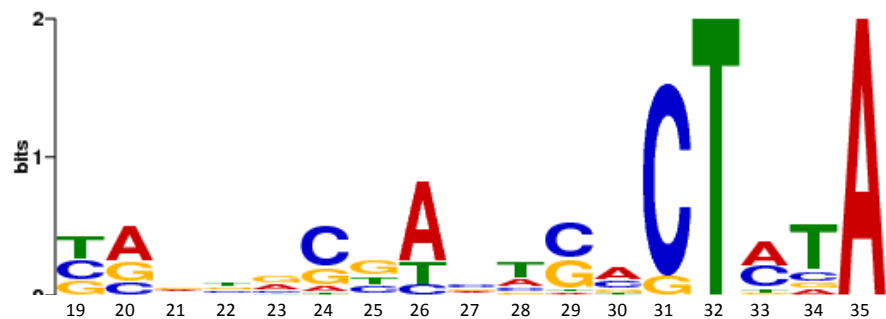


Figure 3.7: Sequence logo of all the substrates used for MEME analysis.

Positions 19 to 35 on the 3' end of all the substrate sequences were aligned in 5' to 3' direction. The logo depicts the probability of the nucleotides present at each position on the substrates used in the study.

For each ribozyme, the substrates that were ligated were aligned and analysed for a sequence pattern. Similarly, the substrates that were not ligated were analysed. Based on the probability matrix (Appendix D), the two sequence patterns were compared for the nucleotides that occurred at a probability of $\geq 50\%$.

3.5.1 Substrates sequence analysis for R18-T4 and R18-T3 ribozyme activity

In the case of R18-T4 and R18-T3 ribozyme, it was found that nucleotide sequence A, A, T and A occurred at a probability of $\geq 60\%$ at the positions 20, 21, 22, and 23, respectively in the substrates that were ligated (Table 3.2 and 3.3). While at the same probability, nucleotide sequence G, G, C and G occurred at those positions in the substrates not ligated. The differences of nucleotides indicated that these positions on the substrates might be important for determining the self-ligation activity of these ribozymes. What is also significant is that, based on the Mfold structure predictions (given in Table 3.7), these positions are in the unfolded region in most cases of the ligated substrates; and in the folded region in most cases of the substrates not ligated. This suggests that the activity of the ribozymes could be determined by the availability of this region on the substrates for pairing.

Results

Probability of the nucleotides in the sequence pattern	SUBSTRATES LIGATED BY R18-T4 (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES LIGATED BY R18-T4 RIBOZYME	SUBSTRATES NOT LIGATED BY R18-T4 (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES NOT LIGATED BY R18-T4 RIBOZYME
=1	20A, 31C, 32T, 35A			
≥0.8	19T, 20A, 26A, 31C, 32T, 35A			
≥0.7	19T, 20A, 26A, 31C, 32T, 35A			
≥0.6	19T, 20A, 21A, 22T, 23A, 24C, 26A, 28T, 30A, 31C, 32T, 33A, 34T, 35A			
≥0.5	19T, 20A, 21A, 22T, 23A, 24C, 25G, 26A, 28T, 29C, 30A, 31C, 32T, 33A, 34T, 35A			

Table 3.2: Analysis of the substrates sequence pattern with respect to R18-T4 activity.

The sequence pattern for the substrates ligated (left) and not ligated (right) by R18-T4 ribozyme were generated by MEME. Based on the probability matrix, the two sequence patterns were compared for the nucleotides that occurred at a probability of ≥ 1 , ≥ 0.8 , ≥ 0.7 , ≥ 0.6 , and ≥ 0.5 . The boxes outline the difference in the two sequence patterns at a probability of ≥ 0.6 .

Results

Probability of the nucleotides in the sequence pattern	SUBSTRATES LIGATED BY R18-T3 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES LIGATED BY R18-T3 RIBOZYME	SUBSTRATES NOT LIGATED BY R18-T3 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES NOT LIGATED BY R18-T3 RIBOZYME
=1	20A, 31C, 32T, 35A		35A	
≥0.8	19T, 20A, 26A, 31C, 32T, 35A		31C, 35A,	
≥0.7	19T, 20A, 26A, 31C, 32T, 35A		31C, 35A, 21G, 34T	
≥0.6	19T, 20A, 21A, 22T, 23A, 24C, 26A, 28T, 30A, 31C, 32T, 33A, 34T, 35A		20G, 21G, 22C, 23G, 24C, 26A, 31C, 34T, 35A	
≥0.5	19T, 20A, 21A, 22T, 23A, 24C, 25G, 26A, 28T, 29C, 30A, 31C, 32T, 33A, 34T, 35A		19C, 19G, 20G, 21G, 22C, 23G, 24C, 26A, 27C, 29C, 29G, 31C, 34T, 35A	

Table 3.3: Analysis of the substrates sequence pattern with respect to R18-T3 activity.

The sequence pattern for the substrates ligated (left) and not ligated (right) by R18-T3 ribozyme were generated by MEME. Based on the probability matrix, the two sequence patterns were compared for the nucleotides that occurred at a probability of ≥ 1 , ≥ 0.8 , ≥ 0.7 , ≥ 0.6 , and ≥ 0.5 . The boxes outline the difference in the two sequence patterns at a probability of ≥ 0.6 .

3.5.2 Substrates sequence analysis for R18-T2 ribozyme activity

On comparison of the sequence patterns for the R18-T2 ribozyme activity, no significant difference was found at a probability of $\geq 60\%$. However, at a probability of $\geq 50\%$, nucleotide sequence T, A, A, T, A occurred at the positions 19, 20, 21, 22, and 23, respectively in the substrates that were ligated; while, C/G, G, G, C, G occurred in the substrates not ligated (Table 3.4). This suggested that this region might be important in determining the ligation activity of R18-T2 ribozyme as well.

3.5.3 Substrates sequence analysis for R18-T1 and R18 ribozyme activity

The comparison of the substrates sequence patterns for R18-T1 and R18 ribozyme activity did not show any significant differences in the nucleotides at a probability of $\geq 60\%$ or $\geq 50\%$ (Table 3.5 and 3.6). The sequence patterns were neither similar. The results did not indicate any specific region important for determining the activity of the ribozyme.

Results

Probability of the nucleotides in the sequence pattern	SUBSTRATES LIGATED BY R18-T2 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES LIGATED BY R18-T2 RIBOZYME	SUBSTRATES NOT LIGATED BY R18-T2 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES NOT LIGATED BY R18-T2 RIBOZYME
=1	20A, 26A, 31C, 32T, 35A		32T, 35A	
≥ 0.8	19T, 20A, 26A, 28T, 31C, 32T, 35A		31C, 32T, 35A	
≥ 0.7	19T, 20A, 25G, 26A, 28T, 29C, 30A, 31C, 32T, 33A, 34T, 35A		20G, 21G, 22C, 23G, 24C, 26A, 31C, 32T, 34T, 35A	
≥ 0.6	20A		20G	
≥ 0.5	19T, 20A, 21A, 22T, 23A, 24C, 25G, 26A, 27T, 29C, 30A, 31C, 32T, 33A, 34T, 35A		19C, 19G, 20G, 21G, 22C, 23G, 24C, 26A, 27C, 29C, 29G, 31C, 32T, 34T, 35A	

Table 3.4: Analysis of the substrate sequence pattern with respect to R18-T2 activity.

The sequence pattern for the substrates ligated (left) and not ligated (right) by R18-T2 ribozyme were generated by MEME. Based on the probability matrix, the two sequence patterns were compared for the nucleotides that occurred at a probability of ≥ 1 , ≥ 0.8 , ≥ 0.7 , ≥ 0.6 , and ≥ 0.5 . The boxes outline the difference in the two sequence patterns at a probability of ≥ 0.5 .

Results

Probability of the nucleotides in the sequence pattern	SUBSTRATES LIGATED BY R18-T1 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES LIGATED BY R18-T1 RIBOZYME	SUBSTRATES NOT LIGATED BY R18-T1 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES NOT LIGATED BY R18-T1 RIBOZYME
=1	19T, 20A, 31C, 32T, 35A		32T, 35A	
≥0.8	19T, 20A, 26A, 31C, 32T, 35A		31C, 32T, 35A	
≥0.7	19T, 20A, 21A, 22T, 23A, 24C, 26A, 30A, 31C, 32T, 33A, 34T, 35A		31C, 32T, 34T, 35A	
≥0.6	--		---	
≥0.5	19T, 20A, 21A, 22T, 23A, 24C, 26A, 28T, 30A, 31C, 32T, 33A, 34T, 35A		29C, 31C, 32T, 34T, 35A	

Table 3.5: Analysis of the substrates sequence pattern with respect to R18-T1 activity.

The sequence pattern for the substrates ligated (left) and not ligated (right) by R18-T1 ribozyme were generated by MEME. Based on the probability matrix, the two sequence patterns were compared for the nucleotides that occurred at a probability of ≥ 1 , ≥ 0.8 , ≥ 0.7 , ≥ 0.6 , and ≥ 0.5 . There were no differences or similarities found in the two sequence patterns at any of the probabilities.

Results

Probability of the nucleotides in the sequence pattern	SUBSTRATES LIGATED BY R18 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES LIGATED BY R18 RIBOZYME	SUBSTRATES NOT LIGATED BY R18 RIBOZYME (Type of Nucleotide and the position in the sequence pattern based on probability)	SEQUENCE PATTERN FOR SUBSTRATES NOT LIGATED BY R18 RIBOZYME
=1	19T, 20A, 26A, 30A, 31C, 32T, 33A, 34T, 35A		32T, 35A	
≥0.8	19T, 20A, 21A, 22T, 23A, 24C, 26A, 30A, 31C, 32T, 33A, 34T, 35A		31C, 32T, 35A	
≥0.7	19T, 20A, 21A, 22T, 23A, 24C, 25T, 26A, 28T, 30A, 31C, 32T, 33A, 34T, 35A		26A, 31C, 32T, 35A	
≥0.6	----		---	
≥0.5	19T, 20A, 21A, 22T, 23A, 24C, 25T, 26A, 28T, 30A, 31C, 32T, 33A, 34T, 35A		26A, 29C, 31C, 32T, 33C, 34T, 35A	

Table 3.6: Analysis of the substrates sequence pattern with respect to R18 activity.

The sequence pattern for the substrates ligated (left) and not ligated (right) by R18 ribozyme were generated by MEME. Based on the probability matrix, the two sequence patterns were compared for the nucleotides that occurred at a probability of ≥ 1 , ≥ 0.8 , ≥ 0.7 , ≥ 0.6 , and ≥ 0.5 . There were no differences or similarities found in the two sequence patterns at any of the probabilities.

Results

Substrate (5' to 3')	Predicted Mfold structure	Substrate (5' to 3')	Predicted Mfold structure	Substrate (5' to 3')	Predicted Mfold structure	Substrate (5' to 3')	Predicted Mfold structure
1		8A		G6		8B	
6		G1		G7		7B	
6A		G2		8		7A	
6B		G3		6C		6E	
7		G4		6D		G5	

Table 3.7: Predicted secondary structures of the substrates by Mfold.

The putative region (positions 20, 21, 22 and 23) on the substrates which might be critical for ligation activity of R18-T4, R18-T3, and R18-T2 ribozymes is shown with a line (curved or straight).

3.6 Rate of self-ligation activity of the ribozymes

The self-ligation reactions were also studied in terms of efficiencies of ribozymes with the substrates. A highly sensitive, probe based quantitative reverse transcription real time PCR technology was used for this purpose. The ribozyme activity was determined using two additional enzymatic steps; reverse transcription of the ribozyme reactions and then PCR amplification. The reverse transcription reactions was carried out for more than the recommended time given by the manufacturer instructions of the reverse transcriptase to ensure that all the ligated product copies are fully reverse transcribed. The efficiencies of the ribozymes were reported by quantifying the copies of reverse transcribed product formed at different time points for a maximum time period of 40 mins. This time period was selected such that the activity of all the ribozymes could be compared in the linear range of product formation before saturation. The aim of the experiments was to compare the efficiencies of the ribozymes with their size and structural complexity. The correlation was confirmed by analysing the activity of the ribozymes with a set of different substrates.

A real time quantitative PCR of the known DNA copies of the ribozymes R18, R18-T1, R18-T2, R18-T3 and R18-T4 ligated to substrate 1 was performed and the corresponding standard curves were generated (Appendix E). A common set of probe and primer sets were used for the amplification of all the samples. The CP values for equivalent copy numbers in all the standard curves of different ribozymes were similar. This showed the reproducibility of the assay and excluded any possibility of error induced due to biasedness of PCR. All the standard curves displayed sensitivity for detection up to 10 copies in the samples. These standard curves were used for quantification of ribozyme reactions. Each ribozyme reaction was incubated from 0 min to 40 mins, reverse transcribed at 8 time points. One microlitre of the reverse transcribed reaction was PCR

amplified using the set of the primers and the probe that were used for the generation of the standard curves. The copies of ligated product cDNA formed at the different time points were determined using the standard curve prepared for the respective ribozyme (Appendix E). The quantified cDNA copies were plotted against time. The rate of reaction (quantified as copies of ligated product cDNA formed per minute) was determined from the slope of the curve.

The standard direct ribozyme assays entail incubation of the ribozyme with the substrate (one of reactants is radiolabeled) and separation of the products on a polyacrylamide gel. The rate of reactions is determined as the fraction of the radio-labeled reactant converted to product over a period of time. The limitation of the RT-qPCR method employed in this study is that it is an indirect method of estimation of the reaction products after a given period of time. Since this method involved two additional enzymatic steps, the fraction calculated will not be an accurate estimation and, therefore, the “Rate of reaction” by standard definitions was not applied. However, for the purpose of comparison of the activities of different R18 polymerase truncations, the term “rate” or “efficiency” has been used as a surrogate for the amount of ligated product formed over a period of time.

3.6.1 Comparison of the rates of ribozyme self-ligation activity with different substrates

In all the reactions of the ribozyme with the substrates, the copies of ligated product cDNA increased linearly with increase in incubation time (Figure 3.8 – 3.12). The rate of self-ligation activity of each ribozyme with different substrates was compared. The smallest ribozyme R18-T4 showed the rate of activity in the range of 7.2 - 15.5 (copies of

ligated product cDNA formed per minute) - Table 3.8. The rate of activity of the largest ribozyme was in the range of 44.18 - 54.70 (copies of ligated product cDNA formed per minute) - Table 3.12. Ribozymes R18-T3, R18-T2, and R18-T1 exhibited rates of activity in the range of 27.18 - 47.04, 18.70 - 38.24, and 15.75 - 17.33 (copies of ligated product cDNA formed per minute), respectively (Table 3.9 - 3.11). Generally, the rates of activity of each ribozyme with different substrates were in a narrow range (Figure 3.13).

3.6.2 Dynamics of the rates of ribozyme self-ligation activity with increase in their complexity

In addition, the ribozymes were compared for the rate of their self-ligation activity with an increase in their size and structural complexity. For this analysis, a set of substrates which were ligated by all the ribozymes were selected i.e. substrates 1, 6, 6a, 6b, 7a. The rate of reactions of each ribozyme with these substrates was plotted (Figure 3.14, Table 3.13). Overall, with an increase in size and structural complexity of the ribozymes, their rate of self-ligation activity increased linearly in a similar fashion with each of the substrates (Figure 3.14). Specifically, the following trend in the efficiency of the ribozymes was observed; $R18-T4 < R18-T1 < R18-T2 \leq R18-T3 < R18$. The smallest ribozyme R18-T4 exhibited the lowest efficiency; while, the largest ribozyme R18 exhibited the highest efficiency. The rates of ribozymes R18-T3 and R18-T2 were of intermediate efficiencies and were in an overlapping range. An exception to the trend of increase in efficiency with size was the activity of R18-T1 ribozyme. Although, larger than ribozymes R18-T2 and R18-T3, it demonstrated relatively lower activity.

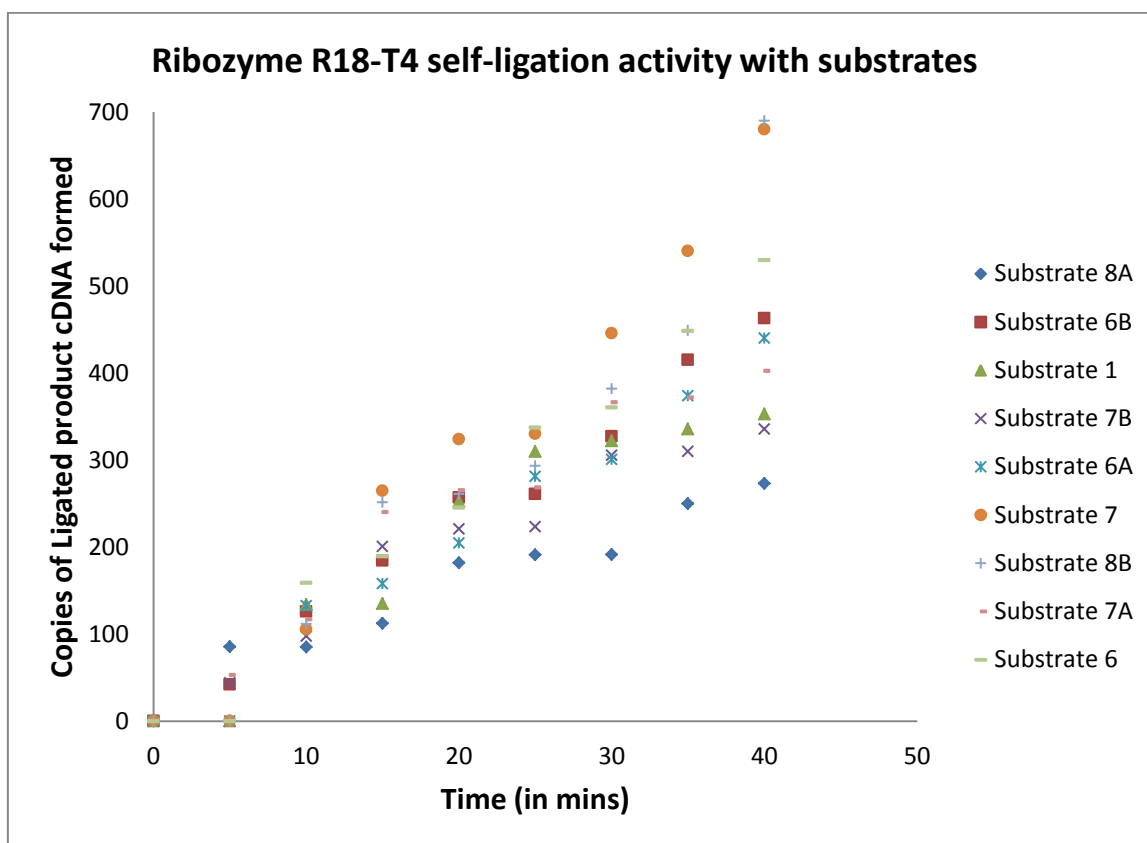


Figure 3.8: Time course analysis of R18-T4 ribozyme self-ligation activity.

The X axis shows the incubation time (in minutes). The Y axis shows the copies of cDNA formed in the reaction. The symbols represent the activity of the R18-T4 ribozyme with the substrates in the course of time.

Results

R18-T4 ribozyme reaction with substrates	Copies of ligated product cDNA formed per minute
Substrate 1	10.13
Substrate 6	12.84
Substrate 7	15.53
Substrate 6a	10.68
Substrate 6b	11.52
Substrate 7a	11.21
Substrate 7b	9.34
Substrate 8a	7.21
Substrate 8b	14.16

Table 3.8: Rates of reaction of R18-T4 ribozyme.

The rates of reaction are given as the copies of ligated product cDNA formed per minute in the assay of R18-T4 ribozyme with the substrates (determined from the slopes of the curves in Figure 3.8).

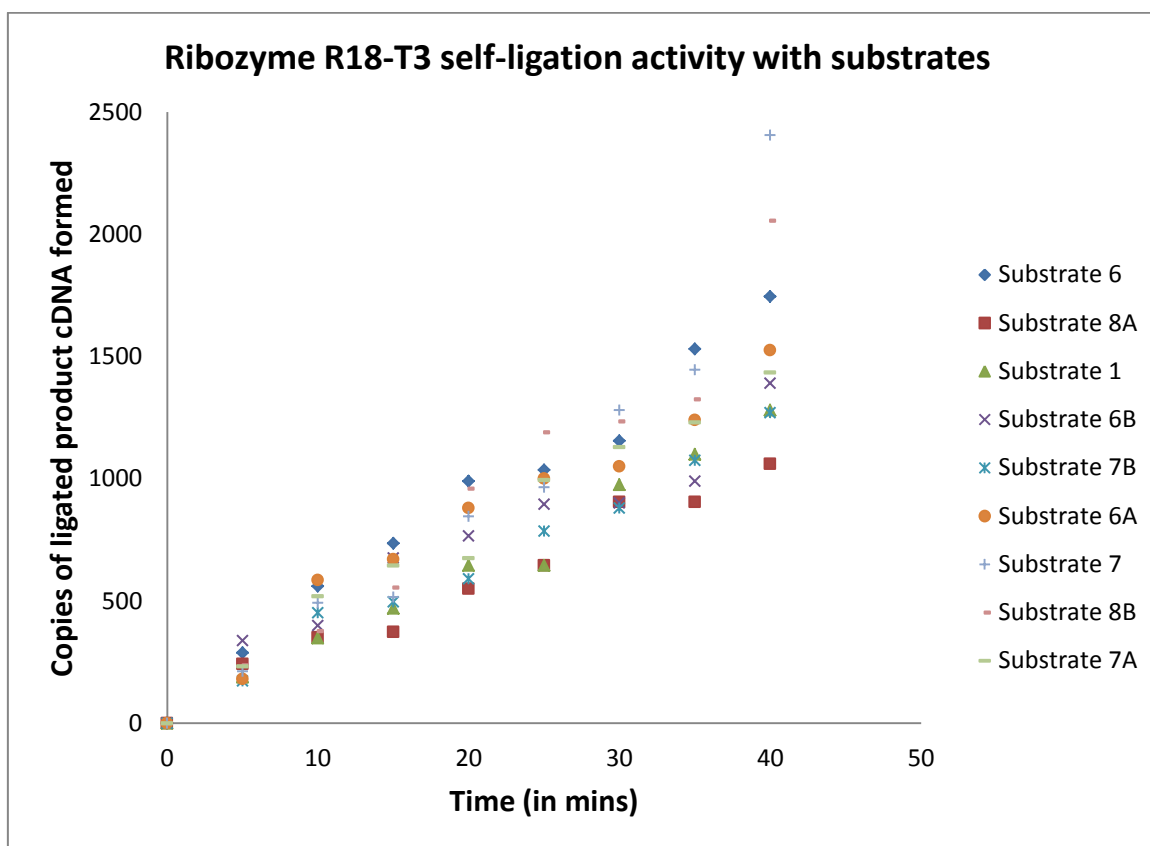


Figure 3.9: Time course analysis of R18-T3 ribozyme self-ligation activity.

The X axis shows the incubation time (in minutes). The Y axis shows the copies of cDNA formed in the reaction. The symbols represent the activity of the R18-T3 ribozyme with the substrates in the course of time.

R18-T3 ribozyme reaction with substrates	Copies of ligated product cDNA formed per minute
Substrate 1	31.26
Substrate 6	43.47
Substrate 7	47.04
Substrate 6a	38.29
Substrate 6b	33.47
Substrate 7a	37.01
Substrate 7b	31.18
Substrate 8a	27.18
Substrate 8b	44.67

Table 3.9: Rates of reaction of R18-T3 ribozyme.

The rates of reaction are given as the copies of ligated product cDNA formed per minute in the assay of R18-T3 ribozyme with the substrates (determined from the slopes of the curves in Figure 3.9).

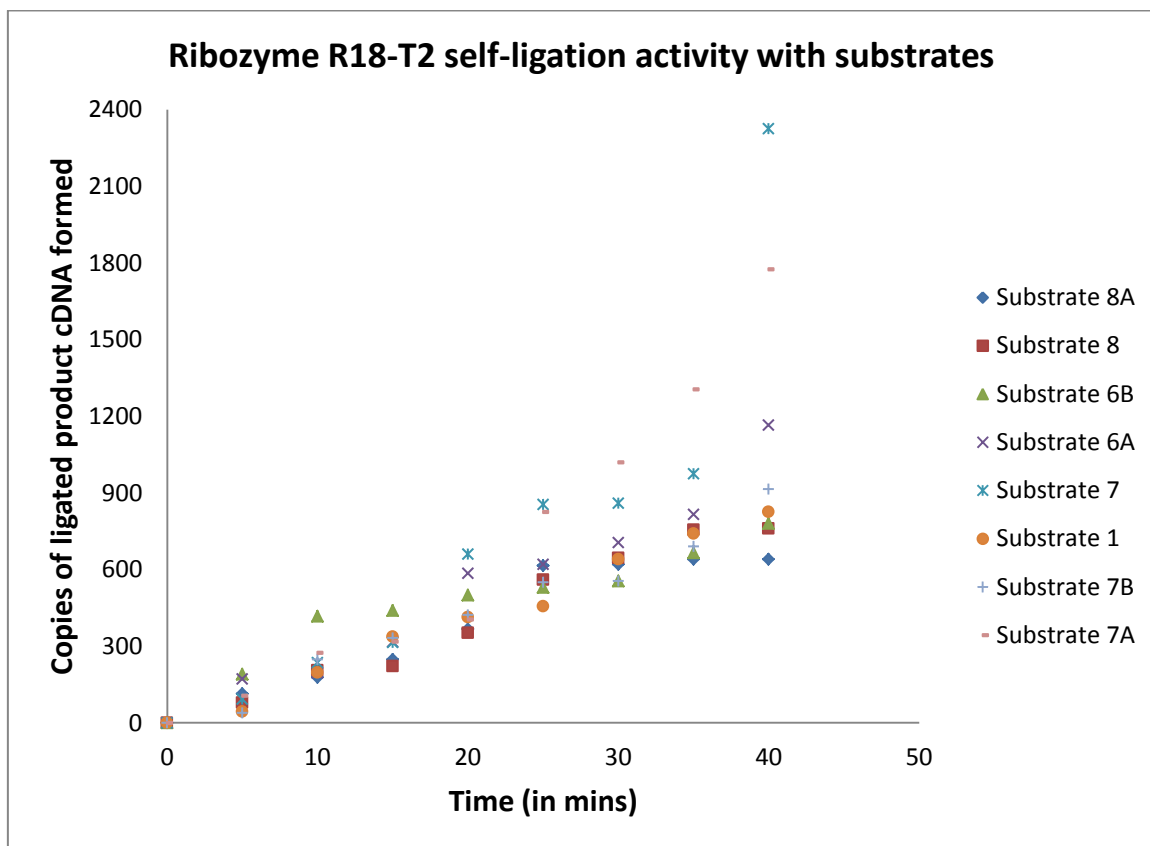


Figure 3.10: Time course analysis of R18-T2 ribozyme self-ligation activity.

The X axis shows the incubation time (in minutes). The Y axis shows the copies of cDNA formed in the reaction. The symbols represent the activity of the R18-T2 ribozyme with the substrates in the course of time.

R18-T2 ribozyme reaction with substrates	Copies of ligated product cDNA formed per minute
Substrate 1	20.59
Substrate 7	38.24
Substrate 8	20.19
Substrate 6a	25.69
Substrate 6b	20.80
Substrate 7a	36.08
Substrate 7b	21.02
Substrate 8a	18.70

Table 3.10: Rates of reaction of R18-T2 ribozyme.

The rates of reaction are given as the copies of ligated product cDNA formed per minute in the assay of R18-T2 ribozyme with the substrates (determined from the slopes of the curves in Figure 3.10).

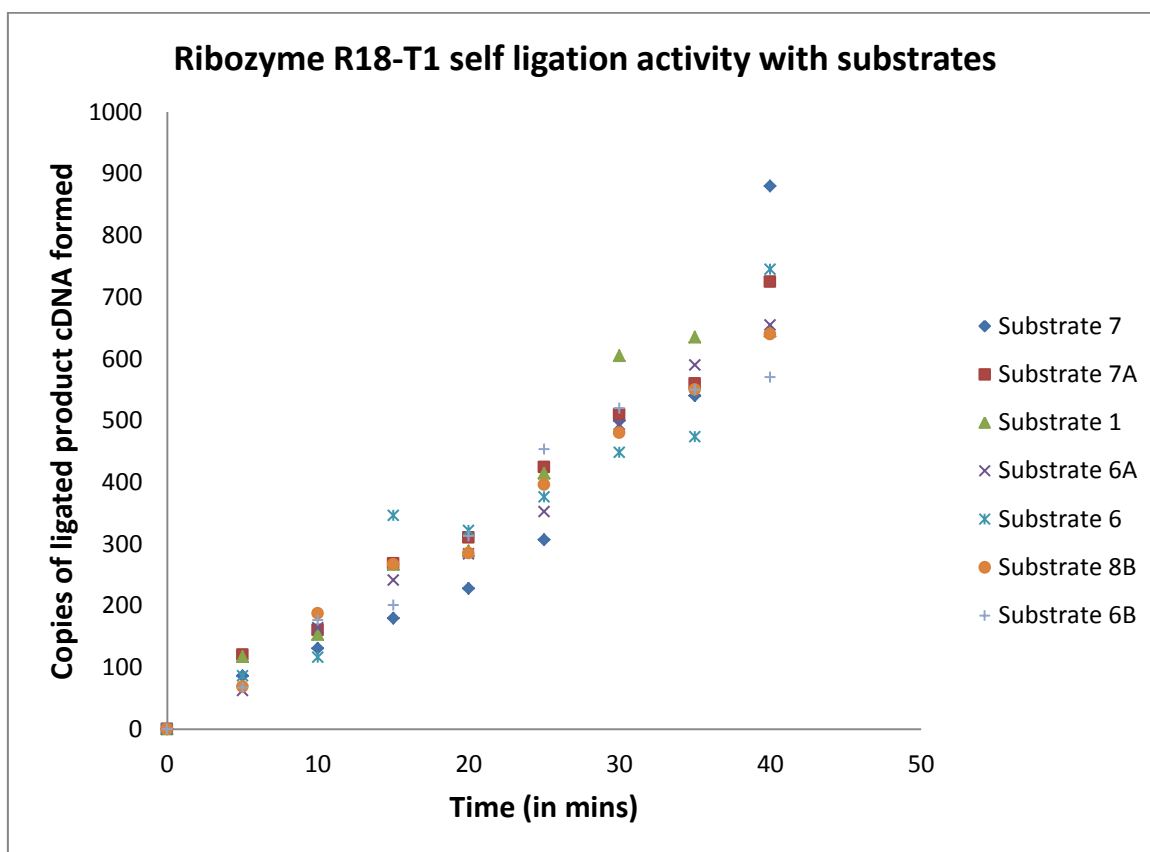


Figure 3.11: Time course analysis of R18-T1 ribozyme self-ligation activity.

The X axis shows the incubation time (in minutes). The Y axis shows the copies of cDNA formed in the reaction. The symbols represent the activity of the R18-T1 ribozyme with the substrates in the course of time.

R18-T1 ribozyme reaction with substrates	Copies of ligated product cDNA formed per minute
Substrate 1	17.33
Substrate 6	16.17
Substrate 7	16.81
Substrate 6a	16.02
Substrate 6b	15.75
Substrate 7a	17.05
Substrate 8b	15.89

Table 3.11: Rates of reaction of R18-T1 ribozyme.

The rates of reaction are given as the copies of ligated product cDNA formed per minute in the assay of R18-T1 ribozyme with the substrates (determined from the slopes of the curves in Figure 3.11).

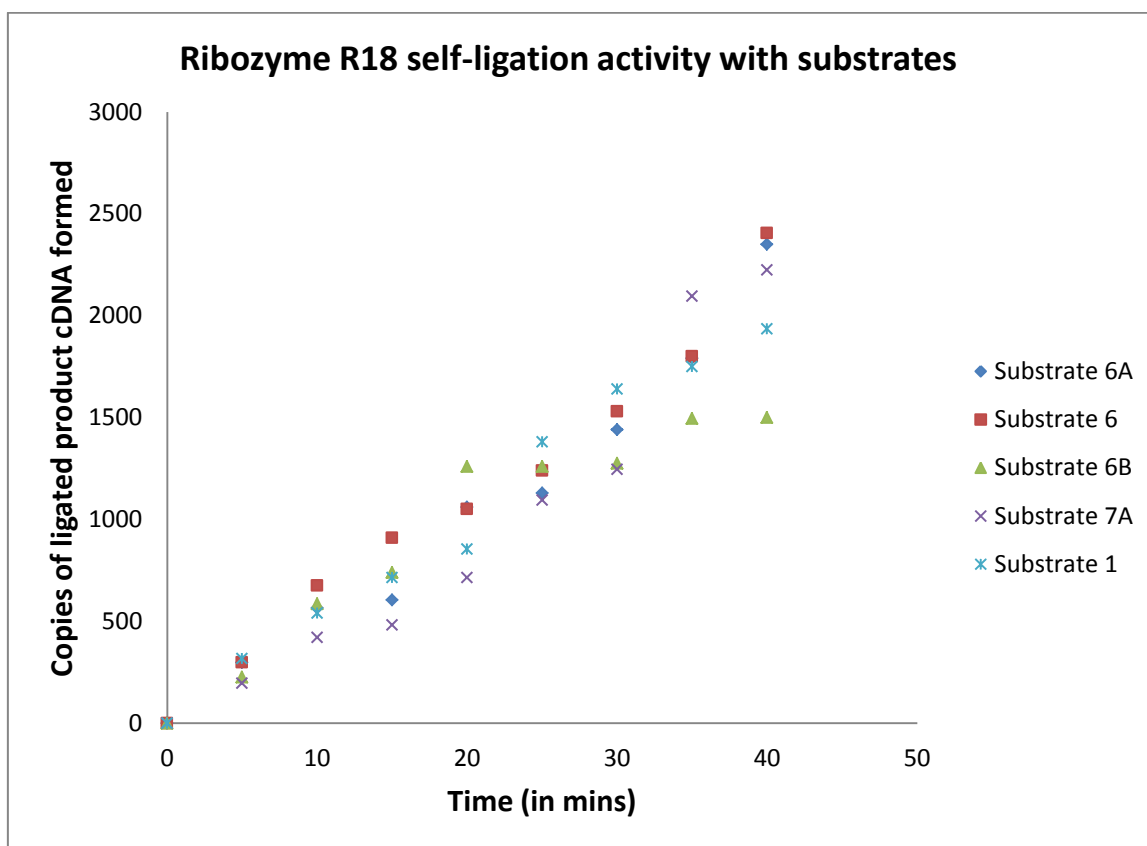


Figure 3.12: Time course analysis of R18 ribozyme self-ligation activity.

The X axis shows the incubation time (in minutes). The Y axis shows the copies of cDNA formed in the reaction. The symbols represent the activity of the R18 ribozyme with the substrates in the course of time.

R18 ribozyme reaction with substrates	Copies of ligated product cDNA formed per minute
Substrate 1	50.42
Substrate 6	54.70
Substrate 6a	52.06
Substrate 6b	44.18
Substrate 7a	49.75

Table 3.12: Rates of reaction of R18 ribozyme.

The rates of reaction are given as the copies of ligated product cDNA formed per minute in the assay of R18 ribozyme with the substrates (determined from the slopes of the curves in Figure 3.12).

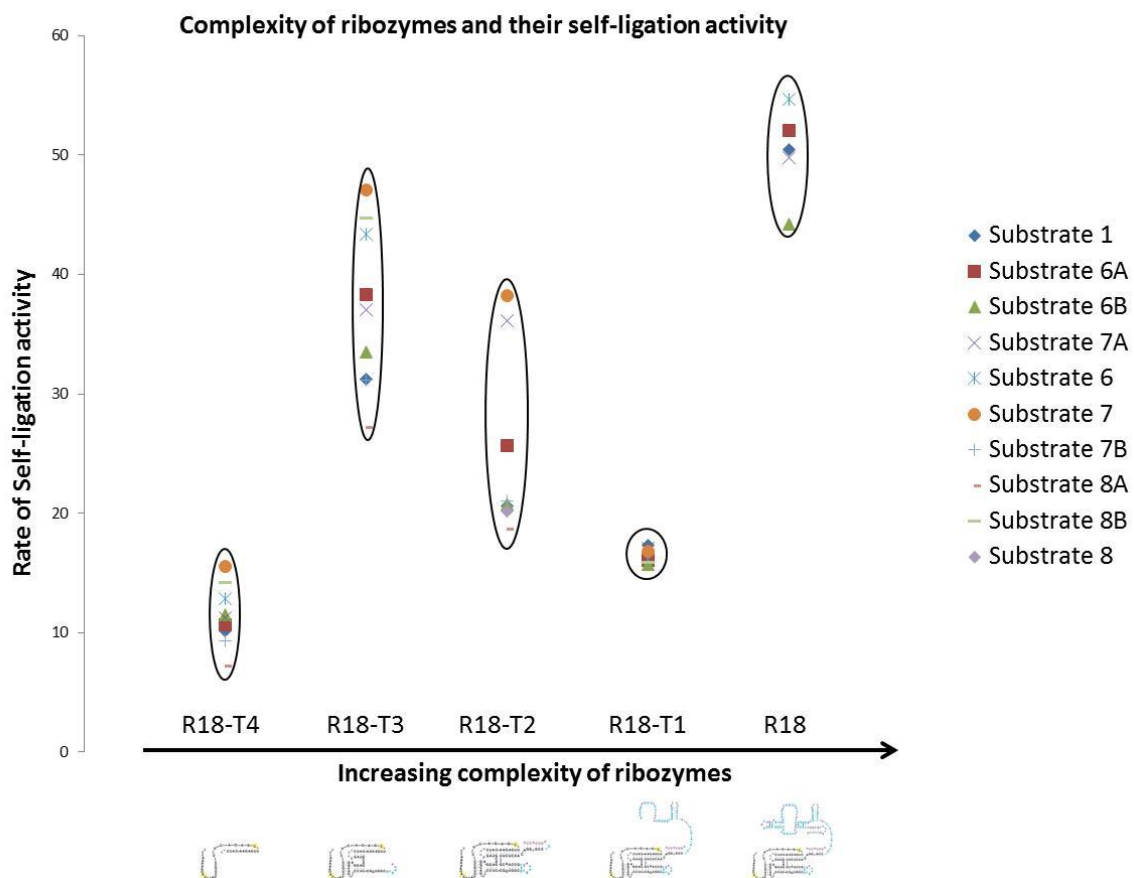


Figure 3.13: Consistency of the ribozymes with respect to their rate of self-ligation reaction with the substrates.

The X axis represents the increase in size and structural complexity of ribozymes. The Y axis represents the copies of the ligated product cDNA formed per minute (given in Table 3.8-3.12). The circles denote the consistency of rate of ribozyme activity with different substrates.

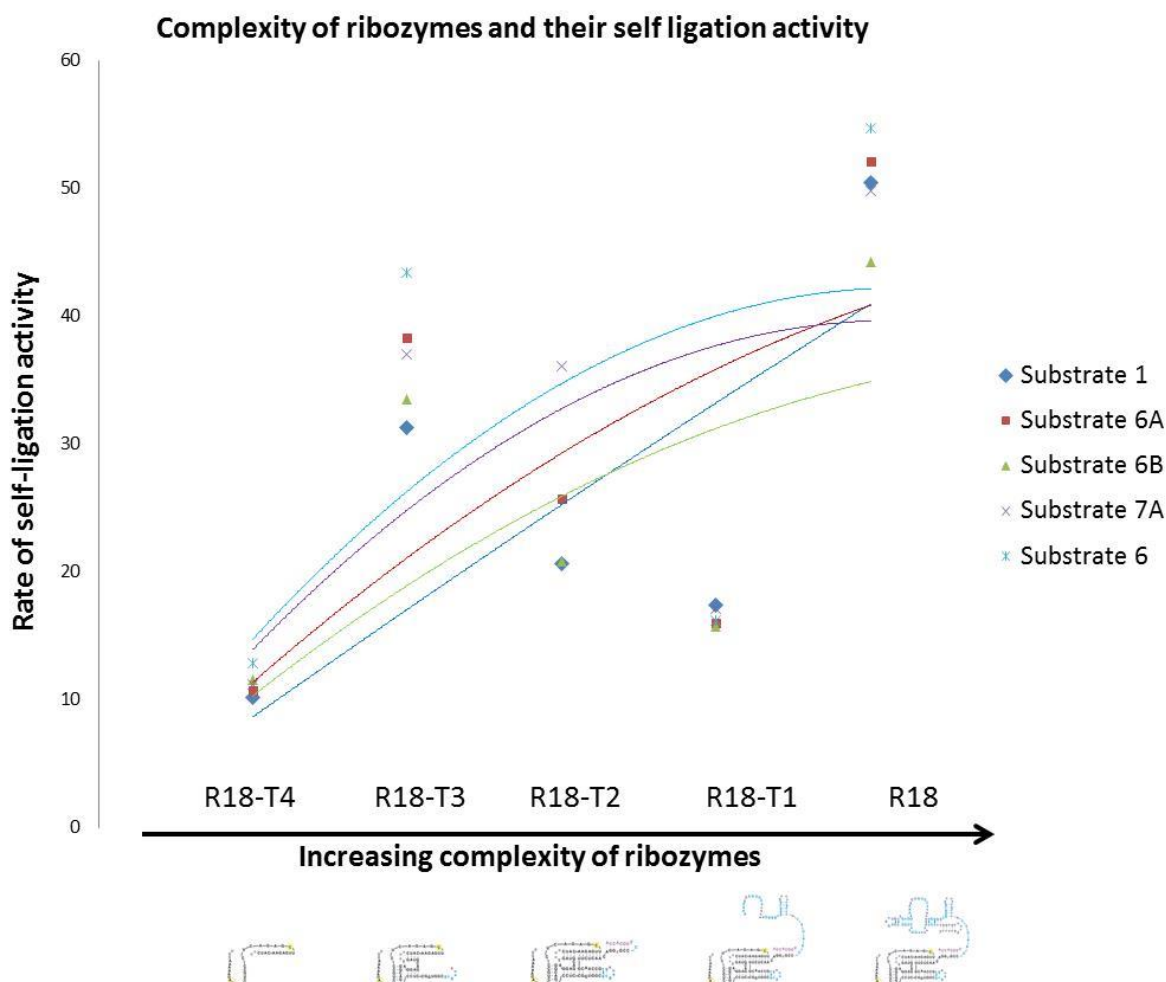


Figure 3.14: Correlation of complexity of the ribozymes with respect to their rate of self-ligation activity with the substrates.

The X axis represents the increase in size and structural complexity of ribozymes. The Y axis represents the copies of ligated product cDNA formed per minute (given in Table 3.13). The curves represent a linear transition in the efficiency of the ribozyme with increase in their size and structural complexity.

	Ribozyme R18-T4	Ribozyme R18-T3	Ribozyme R18-T2	Ribozyme R18-T1	Ribozyme R18
Substrate 1	10.13	31.26	20.59	17.33	50.42
Substrate 6A	10.68	38.29	25.69	16.02	52.06
Substrate 6B	11.52	33.47	20.8	15.75	44.18
Substrate 7A	11.21	37.01	36.08	17.05	49.75
Substrate 6	12.84	43.4	No activity	16.17	54.7
Average rate	11.27	36.6	25.79	16.46	50.22

Table 3.13: Comparison of the ribozymes rates of reaction.

The columns represent the ribozymes and the rows represent the substrates used in the assay. The rates of reaction are given as the copies of ligated product cDNA formed per minute.

3.7 Correlation between biochemical traits with increase in ribozyme complexity

The relationships between various biochemical and molecular traits of the ribozymes were analysed, including size, predicted structural stability, functional flexibility and the rate of self-ligation activity (Summarized in Table 3.14). The objective was to gain an insight into the dynamics of the traits and their role in the evolution of complexity in the RNA world. It was observed that;

- 1) With an increase in size of the ribozymes, the structural stability of the molecules increased, however, the functional flexibility with different kinds of substrates decreased (Figure 3.15).
- 2) With an increase in size of the ribozymes, both the structural stability of the molecules and their rate of self-ligation activity with specific substrates increased. (Figure 3.16).
- 3) With an increase in size and structural stability of ribozymes, their rate of self-ligation activity with specific substrates increased, however, the functional flexibility with different kinds of substrates decreased (Figure 3.18 and 3.19). The functional flexibility and efficiency of molecules were inversely correlated (Figure 3.17).

The correlations point towards molecular trade-offs. The implications of these trade-offs in the evolutionary ecology of RNA world is elaborated in the discussion section. Based on the results, a conceptual model for origin of life is proposed.


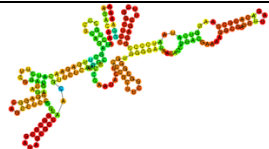



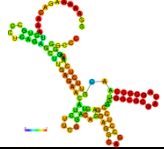

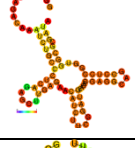

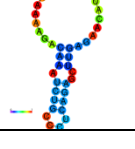
Length of RNA	Structural Complexity		Functional Complexity		
	Secondary structure	Thermodynamic stability (Gibbs free energy; ΔG)	Type of function	Functional flexibility (Kinds of substrates self-ligated)	Rate of self-ligated activity (Copies of cDNA of ligated product formed per minute)
5'  3' (189 nt)		-67.30 kcal/mol	Polymerization Self-Ligation	Self-ligation with 7 out of 24 substrates	50.22 (average activity with 5 substrates); Range: 44.18 – 54.70 50.22 (average activity with substrates 1, 6, 6a, 6b, 7a)
5'  3' (142 nt)		-46.7 kcal/mol	Self-Ligation	Self-ligation with 10 out of 24 substrates	16.43 (average activity with 7 substrates); Range 15.75 - 17.33 16.46 (average activity with substrates 1, 6, 6a, 6b, 7a)
5'  3' (100 nt)		-31.20 kcal/mol	Limited Polymerization Self-Ligation	Self-ligation with 12 out of 24 substrates	25.16 (average activity with 8 substrates); Range: 18.70 - 38.24 25.79 (average activity with substrates 1, 6a, 6b, 7a)
5'  3' (75 nt)		-18.50 kcal/mol	Self-Ligation	Self-ligation with 13 out of 24 substrates	37.05 (average activity with 9 substrates); Range: 27.18 - 47.04 36.6 (average activity with substrates 1, 6, 6a, 6b, 7a)
5'  3' (40 nt)		-3 kcal/mol	Self-Ligation	Self-ligation with 13 out of 24 substrates	11.40 (average activity with 9 substrates); Range: 7.2 - 15.5 11.27 (average activity with substrates 1, 6, 6a, 6b, 7a)

Table 3.14: Summary of correlation between the biochemical traits of the ribozymes with increase in their complexity.

The columns represent the various biochemical traits and the rows represent analysis of the traits at different complexities of the ribozymes.

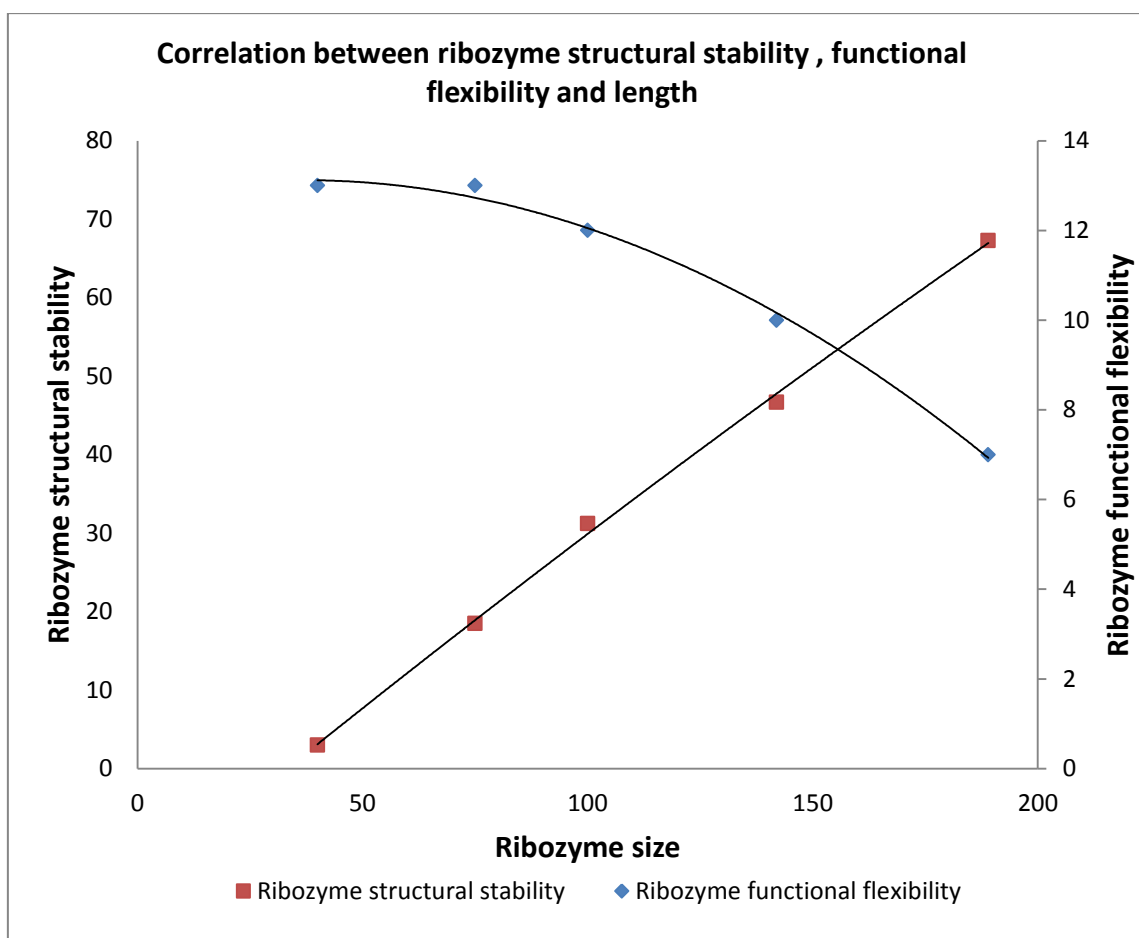


Figure 3.15: Correlation between size, structural stability and functional flexibility of ribozymes.

The X axis represents the size measured by number of nucleotides in the ribozyme. The primary Y axis represents the structural stability measured by predicted Gibbs free energy (ΔG) using RNAfold (the values of ΔG are in negative). The secondary Y axis represents the functional flexibility measured by the number of different oligonucleotide substrates which the ribozyme could ligate to its own end.

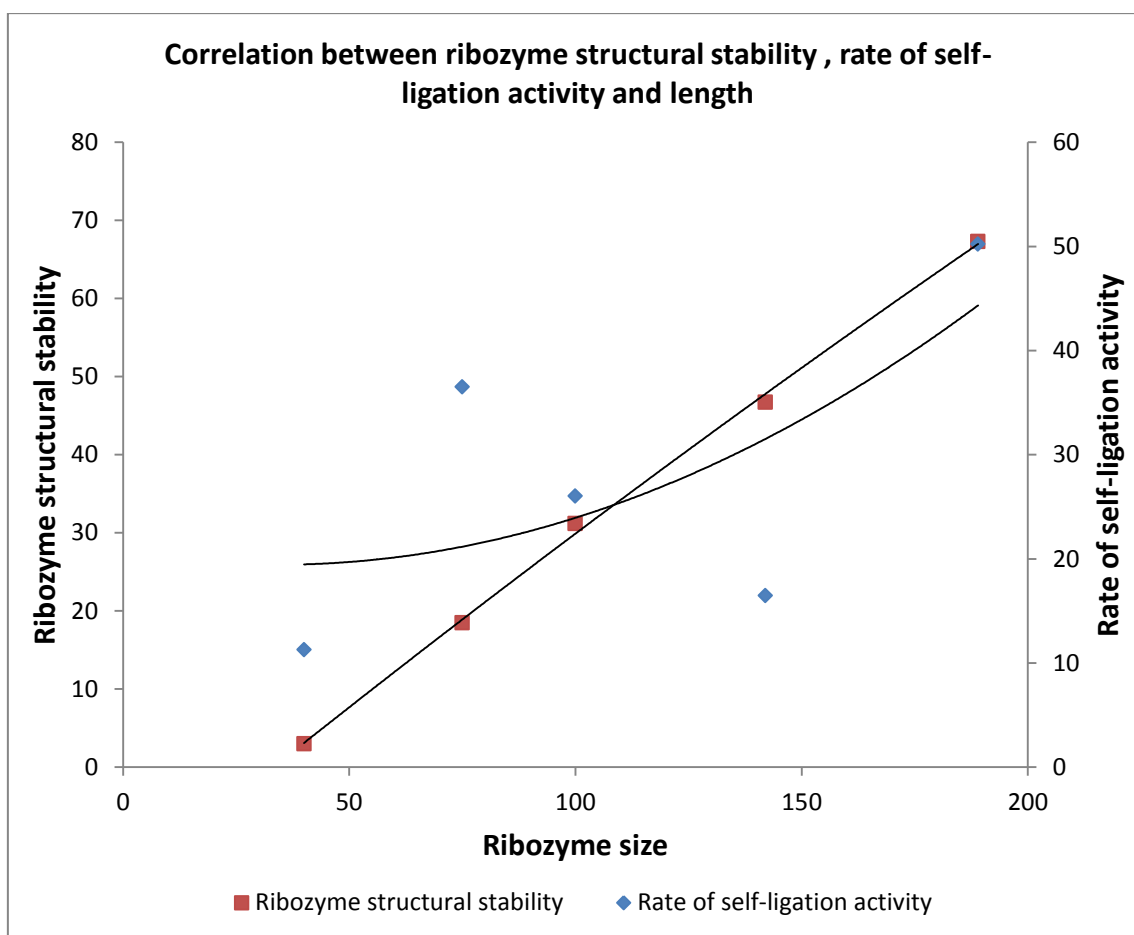


Figure 3.16: Correlation between size, structural stability and the rate of self-ligation activity of ribozymes.

The X axis represents the size measured by number of nucleotides in the ribozyme. The primary Y axis represents the structural stability measured by predicted Gibbs free energy (ΔG) using RNAfold (the values of ΔG are in negative). The secondary Y axis represents the rate of self-ligation activity measured by the copies of ligated product cDNA formed per minute. The represented value of rate is the average rate at which a ribozyme self-ligated 5 substrates (1, 6A, 6B, 7A, 6) - as given in Table 3.13.

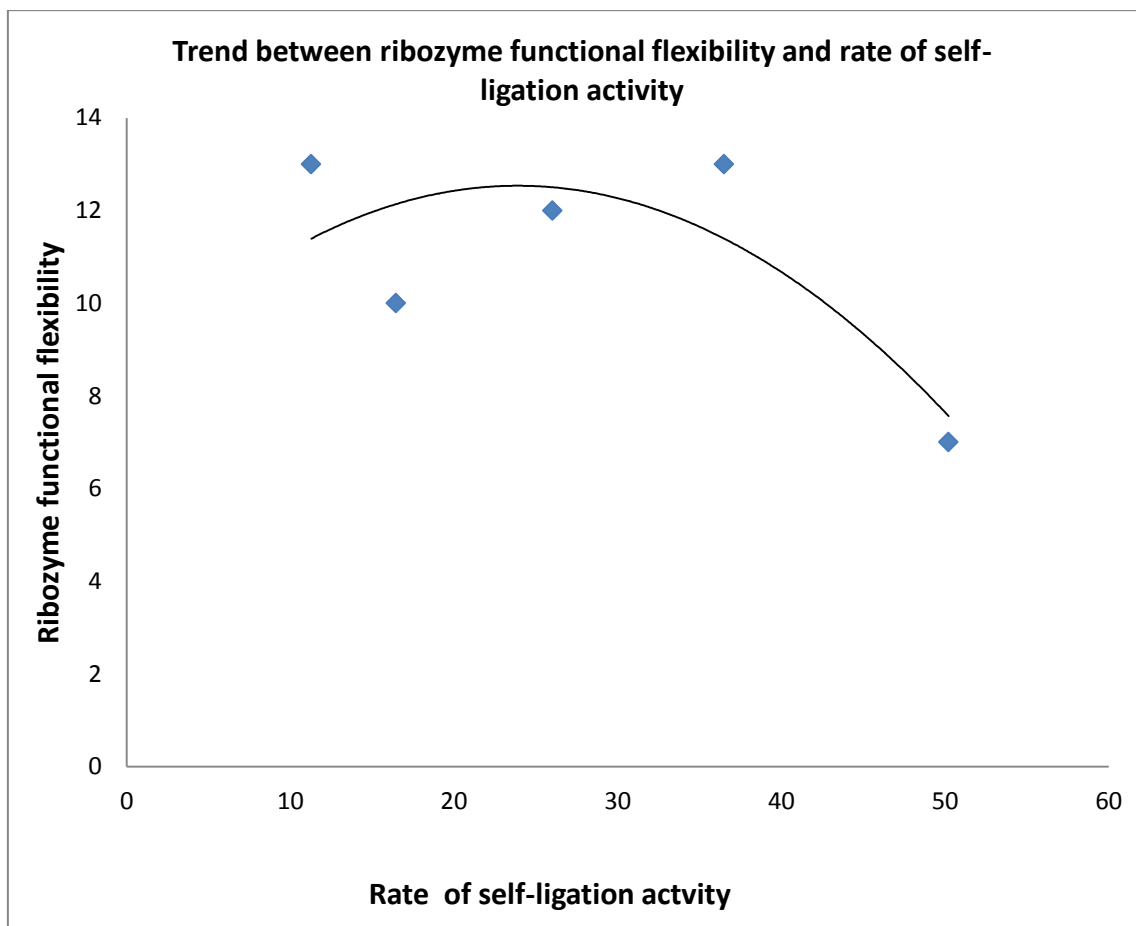


Figure 3.17: Trend between rate of self-ligation activity and functional flexibility.

The X axis represents the rate of self-ligation activity measured by the copies of ligated product cDNA formed per minute. The represented values of rate is the average rate at which a ribozyme self-ligated 5 substrates (1, 6A, 6B, 7A, 6) - as given in Table 3.13. The Y axis represents the functional flexibility measured by the number of different oligonucleotide substrates which the ribozyme could ligate to its own end.

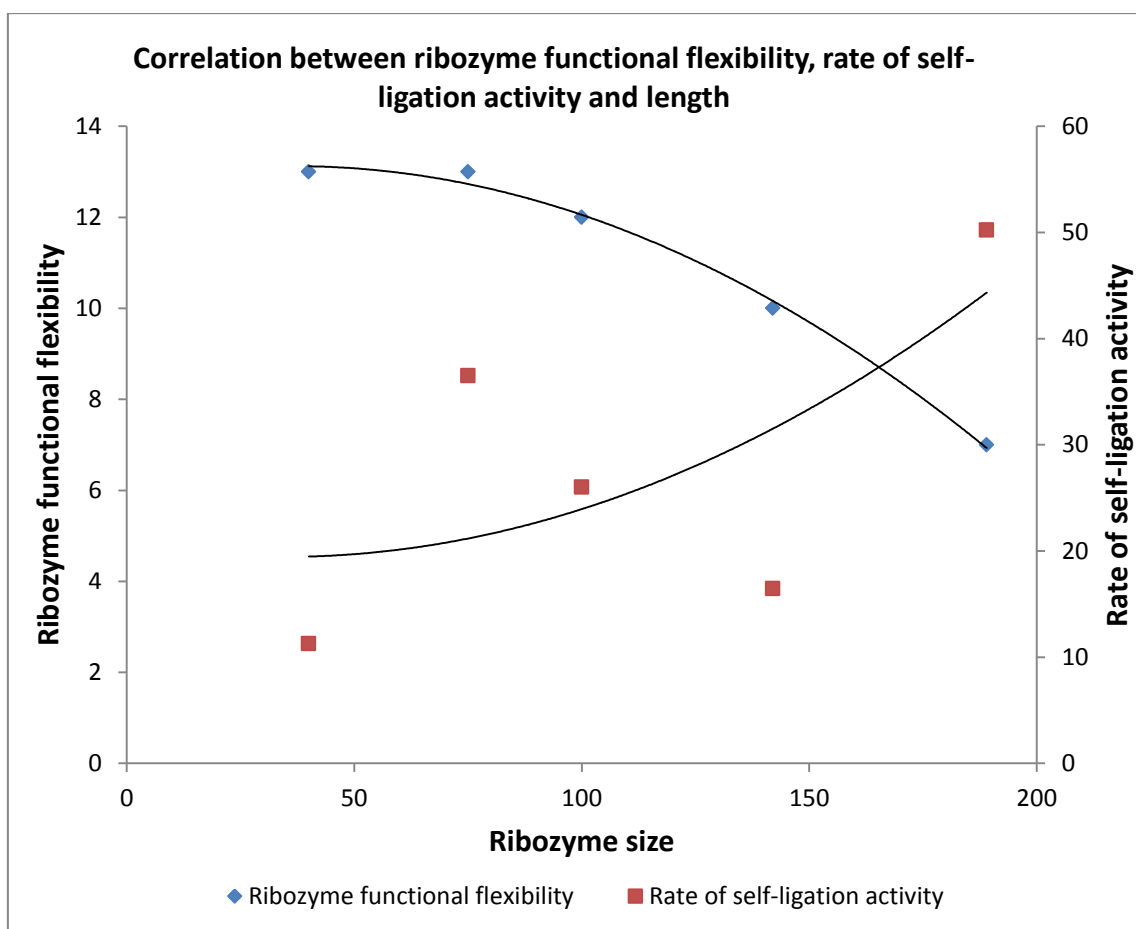


Figure 3.18: Correlation between size, functional flexibility and the rate of self-ligation activity of ribozymes.

The X axis represents the size measured by number of nucleotides in the ribozyme. The primary Y axis represents the functional flexibility measured by the number of different oligonucleotide substrates which the ribozyme could ligate to its own end. The secondary Y axis rate of self-ligation activity measured by the copies of ligated product cDNA formed per minute. The represented value of rate is the average rate at which a ribozyme self-ligated 5 substrates (1, 6A, 6B, 7A, 6) - as given in Table 3.13.

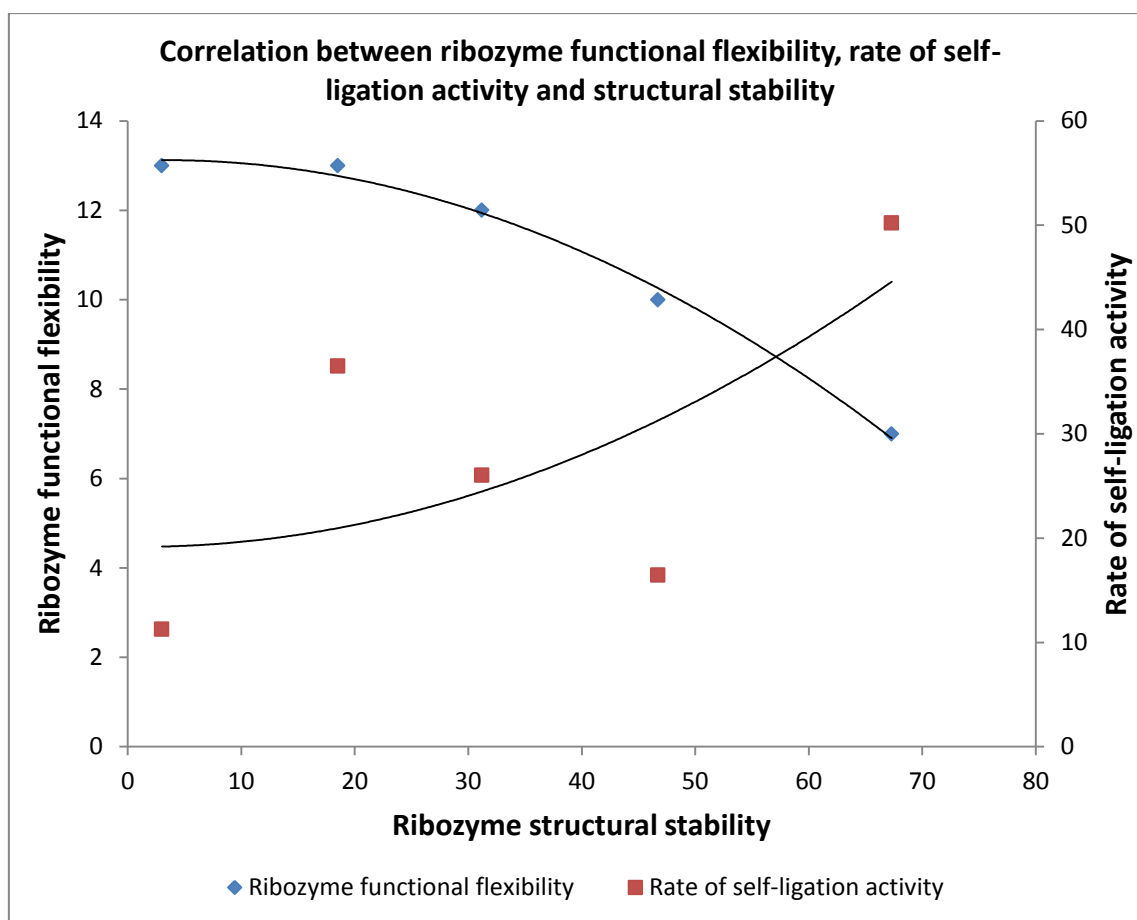


Figure 3.19: Correlation between structural stability, functional flexibility and the rate of self-ligation activity of ribozymes.

The X axis represents the structural stability measured by predicted Gibbs free energy (ΔG) using RNAfold (the values of ΔG are in negative). The primary Y axis represents the functional flexibility measured by the number of different oligonucleotide substrates which the ribozyme could ligate to its own end. The secondary Y axis rate of self-ligation activity measured by the copies of ligated product cDNA formed per minute. The represented value of rate is the average rate at which a ribozyme self-ligated 5 substrates (1, 6A, 6B, 7A, 6) - as given in Table 3.13.

4. DISCUSSION

For a feasible RNA world to be the source of life, the evolution of a minimal RNA polymerase was imperative (James and Ellington, 1999, Joyce, 2007). The emergence of an RNA polymerase would have been vital for (a) the replication of functional RNA molecules formed by passive chemical processes, (b) the formation of a stable network of functional RNA molecules. The engineered polymerases indicate a threshold size which was essential for the function i.e. around 200 nucleotides. (Eklund and Bartel, 1996, Johnston et al., 2001, Wochner et al., 2011, Zaher and Unrau, 2007, Attwater et al., 2013). Although, 40-50 nucleotides long strands are able to form by passive processes without catalysts on montmorillonite clay, the formation of larger molecules has not been possible (Huang and Ferris, 2006, Huang and Ferris, 2003, Ferris, 2002, Ferris, 2006). In this study, the basic evolutionary processes which could account for increase in complexity from short RNA molecules before a polymerase emerged were examined.

The model used in this study was a minimal polymerase, the R18 RNA polymerase, capable of polymerising a given primer-template by 14 nucleotides in 24 hrs. (Johnston et al., 2001). This project was a retrospective study of how this polymerase might have emerged by investigating the ability of the polymerase and its smaller components to ligate oligonucleotides to their own end (self-ligation). The components (R18-T1, R18-T2, R18-T3 and R18-T4 RNA) were synthesised by reducing the length of the polymerase from the 3' end. A set of 24 different substrates (35 nucleotides long) were used to determine the ligation ability of the RNA molecules without any experimentally designed pairing.

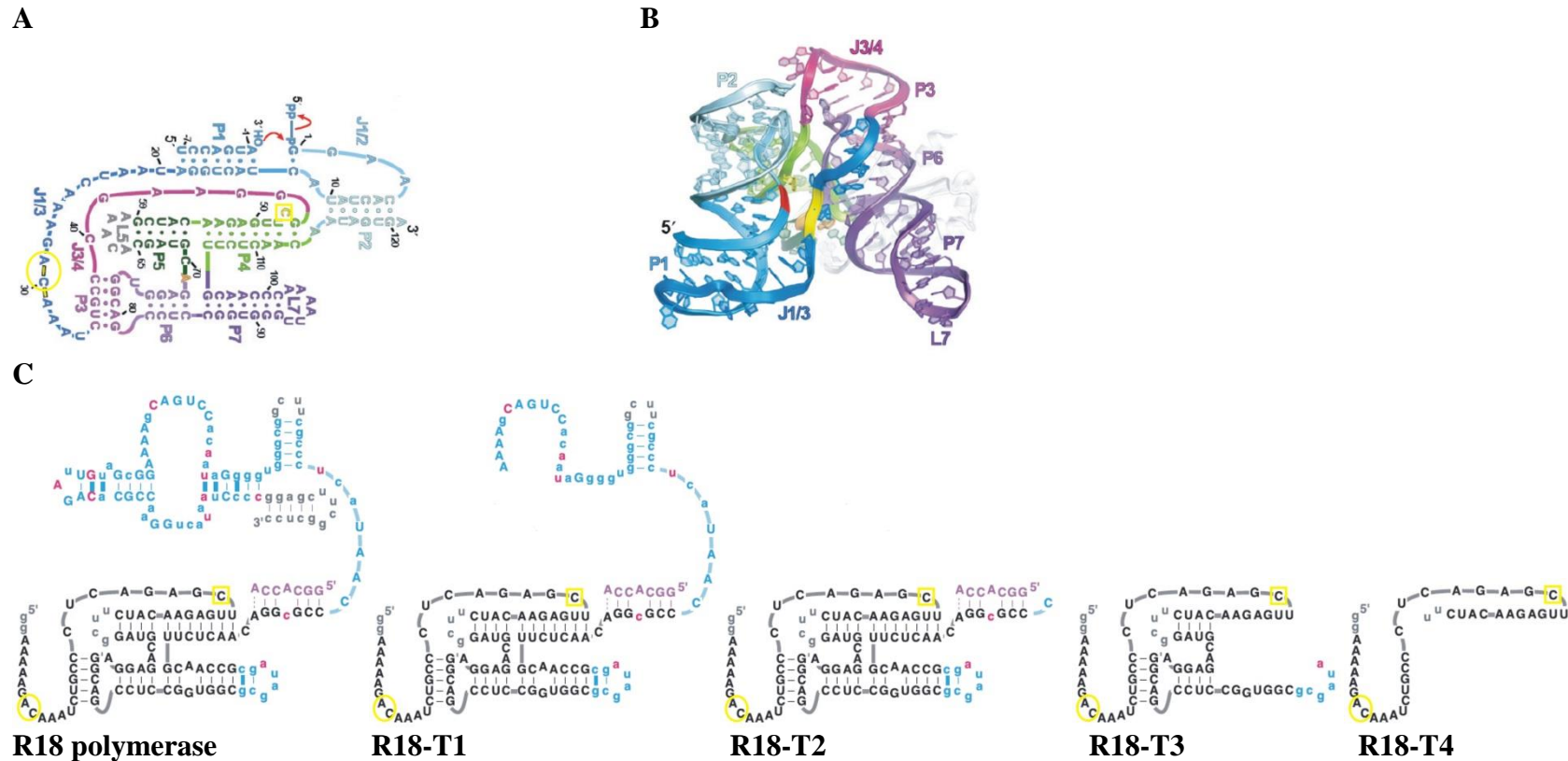


Figure 4.1: Truncated constructs of the R18 polymerase

A. Secondary structure of an improved variant of parental Class I ligase and B. Revised secondary structure of the variant based on crystallization studies (taken from (Shechner et al., 2009)). The nucleotide C (marked with yellow box) and nucleotides A and C (marked with yellow circle) formed the active site for ligation activity. C. Secondary structure of the R18 polymerase (taken from (Johnston et al., 2001)). The truncated constructs of R18 polymerase used in the study are shown for illustration only and do not depict their secondary structures. The residues essential for ligation activity are marked in yellow.

4.1 Self-ligation function in polymerase and its smaller components

It was observed that the polymerase as well as its smaller components ligated substrates to their own 5' end. This indicates that the self-ligation function was preserved in the smaller derivatives of the polymerase. Ligation of small oligomers could have been the early steps for the emergence of complex molecules like a polymerase. It has also been alluded to by others that a polymerase or a replication system may have assembled based on oligonucleotide ligations (Rohatgi et al., 1996b, Bartel, 1999). However, a mechanistic framework for explaining the details of this process remains largely unexplored. In this study, the smallest element of the polymerase that exhibited self-ligation was R18-T4 RNA; a 40 nucleotide molecule with a hairpin structure as predicted by RNAfold software (Table 3.14). Such a molecule is a model system of small molecules, which could “elongate” substantially by virtue of their structure and function using random oligonucleotides. Small RNA ligases are capable of joining template directed oligonucleotides (Vlassov et al., 2004, Robertson et al., 2001, Landweber and Pokrovskaya, 1999). The structural and functional properties of naturally occurring hairpin ribozymes suggest that they are quite adept at forming active catalytic sites (Puerta-Fernández et al., 2003, Svoboda and Di Cara, 2006). Hairpin ligases could, therefore, have played an important role in building complexity. The R18-T4 ligase derived in this project is an exemplar of a molecule that may have existed at the origin of an RNA world.

Based on the global structure features of the improved Class I ligase variant (with minor mutations) derived from the crystallisation studies, it was proposed that C47, as

positioned by C30, and the backbone phosphates of A29 and C30 comprise the ligase active site (Shechner et al., 2009, Bagby et al., 2009). The C47 was predicted to have a more direct role in the catalysis. The R18 polymerase is composed of the minor mutational variant of the active Class I ligase core (excluding the first random 24 bases at the 5' end of the original catalytic core). All the constructs that were examined in this study (including the 40 nucleotide construct; R18-T4 RNA) consist of the 3 bases that were found to be essential for ligation activity (Figure 4.1). Their position numbers are explained in the figure. The R18-T4 RNA forms a minimal functional motif for ligation activity. The precise mechanism of catalysis; however, cannot be predicted and needs further studies. Although, the prediction of structures by RNAfold has limitations, the predicted structure of R18-T4 RNA (Table 3.14); shows the partial reconstruction of a stem similar to the P4 stem in the crystallised core and extrusion of C47 (C27 in the constructs) from the stem. The interaction of the extruded C47 residue with the A29 and A30 was found to be essential in the formation of the active ligase site.

4.2 Self-ligation reactions under no experimentally designed pairing

In the early stages of an RNA world, besides functionality in small molecules, the micro environment would critically determine the success of molecular processes. Biochemists have tried to develop self-sustaining molecular systems based on specific base pairing between the ribozymes and substrate molecules (Joyce, 2004, Joyce, 2007, Paul and Joyce, 2002, Lincoln and Joyce, 2009, Kim and Joyce, 2004). In the early microenvironment, different kinds of molecules in terms of their sequence composition and

structures would have formed by chemical processes. The molecular heterogeneity would have further increased due to an inherently poor replication process. A heterogeneous pool of substrates, therefore, could have constrained the catalytic processes which are functionally based on complementarity between molecules. As a result, would have limited the evolution of molecular activity dependent on the presence of specific substrates. Could there be an alternative process which drove complexity?

This study revealed that the RNA molecules could perform the self-ligation function in the absence of a designed base pairing with the substrates, indicating that ligation reactions based on interactions beyond mere template binding could occur in the early molecular systems. Such molecular interactions formed larger molecules and sustained complexity. At the same time, they assisted in the formation of diverse phenotypes of larger RNA population and in building a foundation for the evolution of an integrated metabolic network and stability of an RNA based life (Eigen and Schuster, 1977, Eigen and Schuster, 1978). The processes which could utilise heterogeneous pool of molecules have been suggested to potentially build complex replicating systems in a prebiotic pool (Szostak, 2011).

4.3 Flexibility of self-ligation function in early RNA molecules

The RNA molecules were analysed for their functional flexibility i.e. their ability to ligate different kinds of substrates to their own end. The smallest component of the R18 RNA polymerase, R18-T4 RNA was more general in its function and self-ligated 13 out of

24 different kinds of substrates (Table 3.1). However, with an increase in size of the ribozyme, there was a gradual restriction in the kind of substrates preferred for self-ligation. Ribozymes R18-T3, R18-T2, R18-T1 self-ligated 13, 12, 10 different kinds of substrates, respectively. The R18 RNA polymerase was more specific in its function and self-ligated only 7 out of 24 kinds of substrates. Specifically, 3 types of patterns were found (Table 3.1); **Pattern 1** showed that the substrates that were self-ligated by R18-T4 RNA were not ligated with an increase in the size of the ribozyme catalyst. This demonstrates a constraint in self-ligation activity with increasing size in ribozymes. The generality of ligation in the R18-T4 ribozyme could be due to its less folded nature which allowed interaction with different kinds of molecules. With increase in the size of the ribozymes, their folding increased (Table 3.14), which limited their interaction with different kinds of molecules. The highly folded nature of R18 polymerase provided specificity to the ribozyme. Thus, in the early stages of RNA world, the structural complexity of the catalysts could critically influence their functional flexibility. **Pattern 2** showed that a subset of the substrates was self-ligated by all the types of ribozymes investigated in the study. This suggested that, although the self-ligation was gradually restricted with an increase in size of the catalysts, the function was still preserved before a polymerase emerged. **Pattern 3** showed that a subset of substrates was not self-ligated by any of the ribozymes irrespective of their size or structural complexity. This suggests that although self-ligation was a basic function in the smaller components of the polymerase, it could be completely absent in some substrate pools. The three patterns provided an insight into the dynamics of the structural nature and functional flexibility of the ribozymes. Furthermore, the nature of the substrates and their effect on the self-ligation reaction by the ribozymes was studied.

4.4 Analysis of the substrates for the self-ligation activity of the ribozymes

The nature of the substrates for the self-ligation activity of the ribozymes was analysed using MEME. For each ribozyme, the sequence pattern of the substrates that were ligated was compared to those that were not ligated. In the case of R18-T4 ribozyme, nucleotide sequence A, A, T and A occurred at a probability of $\geq 60\%$ at the positions 20, 21, 22, and 23, respectively in the substrates that were ligated (Table 3.2). While at the same probability, nucleotide sequence G, G, C and G occurred at those positions in the substrates not ligated indicating that this region was important for determining the activity of the R18-T4 ribozyme. It further suggested that, although there was no experimentally designed base pairing for the reactions, R18-T4 ribozyme preferred substrates composed of AATA over GGCG in that region. The structures of the substrates were also analysed using Mfold (Table 3.7). It was found that a change of nucleotides from AATA to GGCG modified the secondary structure of the substrates such that it allowed self-base pairing. This may have rendered the substrate inaccessible to the ribozyme. Comparatively, substrates with nucleotides AATA were more open and less folded structures that may have provided more unpaired regions for the ribozymes to bind. The analysis suggested that in a pool of random molecules, the smallest component of the polymerase was more likely to self-ligate substrates which were less folded. For the ribozymes, R18-T3 and R18-T2, the sequence patterns of the substrates were different in the same region i.e. positions 20, 21, 22, and 23 (Table 3.3 and 3.4). This region on the substrates, therefore, remained important for determining the activity of these larger components of the polymerase. For the ribozymes R18-T1 and R18 polymerase, the sequence patterns of the substrates were neither significantly different nor similar in any region (Table 3.5 and 3.6). This result

could not be used to determine the interactions. However, since the MEME tool is capable of identifying un-gapped motifs based on the sequence pattern, it has limitations in identifying structural motifs. It is possible that the activity of these ribozymes with the substrates were based more on tertiary interactions. The highly folded catalytic molecules provided fewer unpaired regions for binding and, hence, only specific kinds of substrates compatible with the stereochemistry of the ribozymes were ligated.

The nature of the ribozymes and the substrates demonstrate that in a mixed pool of oligomers, general ligation reactions via differential pairing based on unpaired regions in molecules formed larger molecules. The smaller components of the polymerase utilised their own tag sequence in different substrates and performed self-ligation functions. In the early stages of life, this would have assisted with building complexity. The flexible interaction of functional nucleic acids has been unexpectedly observed in reactions designed for specific base pairing between molecules (Levy and Ellington, 2002, Robertson et al., 2001, Chapman and Szostak, 1995). Catalytic reactions between opposing enantiomers have occurred simply based on tertiary interactions (Sczepanski and Joyce, 2014). The degree of secondary structures formed by nucleic acids, however, influenced the molecular interactions. Highly folded structures were less available for pairing with different nucleic acid sequences as compared to less folded structures. This further determined the ligation reactions between the molecules. The molecular ecology, thus, played an important role in the evolution of ligation function and primitive molecular processes. The success of such processes would also be governed by the efficiency of the ligases.

4.5 Rate of self-ligation activity

The efficiency of each type of ribozyme was studied by determining the rate of self-ligation activity. This was done by quantifying the cDNA copies of the ligated product formed relative to time. All the ribozyme-substrate reactions showed a linear increase of the product with increase in incubation time (Figures 3.8 - 3.12). This suggested the capability of the systems to increase in complexity with time. The efficiencies of each ribozyme with different substrates were in a narrow range (Figure 3.13). This showed the consistency of different ribozymes in their respective functional efficiencies. The activity of the ribozymes indicated their robustness for the reaction with different substrates.

Furthermore, the catalytic efficiencies of the ribozymes were compared to analyse the dynamics of the self-ligation process with increase in molecular size. This analysis was performed with a subset of five substrates. With each of the substrates, the efficiencies increased similarly in a linear fashion with increase in the size of the ribozymes (Figure 3.14). The R18-T4 ribozyme had the lowest catalytic rate. However, with increase in the size of the ribozymes, an overall increase in the catalytic rate was observed. The R18 polymerase showed the highest self-ligation rate. The low turnover of product formation in case of R18-T4 ribozyme could be due to its small structure, which may result in weak binding with the substrates and susceptibility to dissociation before completion of the reaction. The turnover of product formation increased with increase in molecular size of the catalysts. This is likely due to a stronger binding of the catalysts with the substrates conferred by tertiary interactions resulting in increased stability of the ribozyme-substrate complex for completion of the reaction. An exception to this trend was the activity of R18-T1 ribozyme. Although, larger in size than ribozymes R18-T2 and R18-T3, it demonstrated

lower efficiency. An intermediate increase in complexity from R18-T3 to R18 polymerase could have made unfavourable contacts with the catalytic domain for the activity. A general trend of increase in specificity and efficiency with increase in structural complexity of RNA has been found in the engineered and naturally occurring ribozymes (Carothers et al., 2004, Scott, 2007, Ekland et al., 1995) (Lai et al., 2010). These data suggested that before a polymerase emerged in the RNA world, its smallest component; the R18-T4 ribozyme was adept at generally ligating different kinds of substrates, however, with lower efficiency. The increase in size and structural complexity of the ribozymes by self-ligation function was essential for developing their specificity and efficiency. In the early stages of life, the progressive increase in self-ligation efficiency of larger molecules would have fuelled the formation of more complex molecules. There could, however, be a stage of molecular complexity (R18-T1 ribozyme) which was less efficient in self-ligation function. Such a stage could become a cost to the increasing complexity of the overall system.

Ligases have been explored primarily with the aim of developing a self-replicase with the use of a substrate that is base paired with the ribozyme (Joyce, 2007). The self-replication of the molecules was a very important aspect in the RNA world. A pool of molecules would; however, have been required for such sophisticated catalysts to emerge and for the stability of RNA based life. The prebiotic chemistry on the other hand would have generated a limited set of molecules. The current study examined the role of ligases in increasing molecular complexity in a setting where the substrates might not necessarily be complementary to aid the favourability of the reactions. In the absence of a designed base pairing, the affinity of the molecules for the substrates could become a limiting factor. The efficiency of such ligation reactions was, thereby low (to be detected by direct ribozyme assays). In the RNA world, this low catalytic efficiency of each molecule might not have

sufficed their ability to replicate on their own; however, the reactions would have contributed as a whole in building the early molecular pool for a replicating system to emerge. These molecules would have replenished based on the conceptual model provided in Section 4.7. The study revealed various correlations between molecular traits of catalytic RNA such as size, functional flexibility and catalytic efficiency. The probability of evolution of molecular complexity in the RNA world would lay on the structural complexity of the catalytic RNA molecules as well as the kinds of substrates around them are evident.

4.6 Correlations and trade-offs between molecular traits and the implications for the origin of life

The R18 polymerase and its smaller components were examined for correlation between traits such as structural stability, functional flexibility and catalytic efficiency. With an increase in size and the structural stability of the ribozyme the flexibility in self-ligating different kinds of substrates decreased (Figure 3.15). At the same time, the rate of self-ligation activity with specific substrates increased (Figure 3.16). This revealed an inverse correlation between the functional flexibility and the catalytic efficiency of the molecules with increase in their size and structural stability (Figure 3.17, 3.18 and 3.19). The correlation suggested that molecular trade-offs are at play in this system.

The concept of how trade-offs and constraints can shape the evolution of complex functions and life history traits has been applied to organismal populations to study their

evolutionary dynamics. Classical life history traits are directly related to two major components of fitness, i.e., survival and reproduction (Flatt and Heyland, 2011, Stearns, 1992). In addition to these traits, morphological, physiological, or behavioural traits which may contribute to fitness and have major effects on reproduction and survival have been called life history traits (Roff, 2007). A life-history trade-off occurs when an increased investment in one fitness component causes a reduced investment in another fitness component i.e. fitness benefit in one trait exacts a fitness cost in another. Examples include survival versus reproduction, number versus size of offspring. According to life history theory, trade-offs and constraints within the organism limited life history traits and could have promoted evolutionary transitions (Roff, 2002, Stearns, 1992).

While, the fitness of organisms can be defined based on life history traits, it is less clear how to apply the term at the level of RNA molecules at the origin of life. Researchers have tried to extend the concept of fitness theoretically and empirically to RNA populations (Takeuchi and Hogeweg, 2012, Joyce, 2004, Athavale et al., 2014, Lincoln and Joyce, 2009). Fitness in RNA molecules has been measured in terms of their efficiency of self-replication. Evolution of the catalytic RNA molecules has been studied by *in vitro* continuous system that mimics evolution of organisms in nature (Arenas and Lehman, 2010, Ellington et al., 2009). In such a system, repeated rounds of random mutations are introduced to maintain variation in the population and catalytically efficient molecules are selected. The principle of fitness in such an evolving system is based on selection of RNA molecules in their ability to catalyse a reaction with their cognate substrates (Diaz Arenas and Lehman, 2013, Joyce, 2009). The primitive environment which may not be necessary rich in the related kind of substrates may critically influence the emergence and selection

of molecules. The role of ecological dynamics in varying the selection pressure occurring during the evolution of a population has been seen at the organismal level (Conner and Hartl, 2004). The change in selection pressure can affect the mean fitness value of a phenotypic trait (Falconer, 1983). This demands the question that before life history traits like replication evolved in molecules, could there be characteristics of catalytic molecules which indirectly determined the fitness of the molecular pool?

This study found that molecular traits like functional flexibility, specificity, and efficiency could determine an increase in complexity from small RNAs to a minimal polymerase. The trade-offs in these traits associated with the molecular complexity could collectively shape the fitness of the molecular pool. This could happen as follows:

- (a) The small size of RNA molecules allowed them to be flexible in self-ligation function. The functional flexibility increased the molecular diversity in the pool, essential for evolution of new functions. The efficiency of the self-ligation function was low, however, was a consistent function due to low specificity.
- (b) In the process, the small RNA ligases formed larger molecules with increased structural complexity and thermodynamic stability. The increased structural stability made the molecules less prone to thermal degradation and thus enhanced the survivability of the molecular pool.
- (c) The large size and increased structural stability of RNA molecules constrained their self-ligation function, however, at the same time increased their catalytic efficiency. The limited self-ligation function reduced the phenotypic variability in molecules. The latter was essential for preservation of precious structural folds and maintenance of functional integrity of the molecular pool. The increase in the self-

ligation efficiency of the molecules developed the processivity of the molecular pool.

The fitness of a group of molecules is manifested through their phenotype (Orr, 2009). The above characteristics of RNA ligases would have conferred variation, stabilisation and evolution of phenotypes to the molecular pool. This would have consequently determined the fitness of the pool until a polymerase emerged. Based on the observations in this study, a conceptual model for the origin of a replicative unit in the RNA world is proposed.

4.7 Conceptual model for network stability at origin of life

The processes essential for the emergence of an RNA based life, besides replication, would be 1) elongation of small replicators for emergence of larger catalysts, 2) generation of molecular diversity, which was important for network stability, and 3) structural and functional stability of molecules. The study employed R18 RNA polymerase as the model system and its smaller components to understand these processes. It revealed the following:

- (a) The RNA polymerase and its reduced complexity levels exhibited self-ligation function. This indicated that such a function was preserved in the polymerase as well as its smaller components.
- (b) All the ribozymes studied performed the self-ligation function and importantly this was occurred without any experimentally designed base pairing to the substrate sequence. This indicated that in the absence of a favourable substrate, the catalytic molecules employed differential base pairing with the substrates conducive to the

stereochemistry. This could be one of the primitive functions in the early stages of the RNA world.

- (c) The smallest component of the polymerase was more flexible in ligating different kinds of substrates. This indicated that it would have been more adept in elongation and capable in giving rise to larger and diverse molecules.
- (d) There was a gradual decrease in functional flexibility with increase in size of the catalytic molecules. The most complex RNA polymerase was functionally least flexible in self-ligating different kinds of substrates. This indicated development of elongation restriction and specificity with increase in molecular size.
- (e) The smallest component of R18 RNA polymerase demonstrated the lowest catalytic efficiency. However, an overall increase in efficiency of self-ligation was observed with increase in the size of the catalytic molecules. The polymerase demonstrated highest catalytic efficiency. This indicated development of molecular processivity with increase in molecular size.
- (f) A basic trade-off was observed between functional flexibility and catalytic efficiency with increase in size and structural complexity of molecules.

Bases on these observations, a conceptual model for increase in complexity and network stability at the origin of life is proposed. The proposed model accounts for the emergence of a polymerase and a replicative unit based on a complementary set of processes and molecular trade-offs (illustrated in Figure 4.2).

Process 1: Elongation of small replicators and generation of molecular diversity

In the first step, a set of molecules (40-50 mers) formed prebiotically (Huang and Ferris, 2006, Huang and Ferris, 2003, Ferris, 2002, Ferris, 2006). Replication of molecules occurred by passive template directed mechanisms (Wu and Orgel, 1992b, Wu and Orgel, 1992a, Deck et al., 2011, Rohatgi et al., 1996b, Rohatgi et al., 1996a). By virtue of simple Watson crick pairing possibilities within the sequences, the molecules assumed a compact secondary and tertiary structure. A random pool of 25-40 nt in length was abundant in structures such as stem loops and hairpins (Stich et al., 2008, Gevertz et al., 2005, Knight and Yarus, 2003, Hendrix et al., 2005). It is evident that RNA molecules have minimum structural requirements to function as ligases (Vlassov et al., 2004, Robertson et al., 2001, Landweber and Pokrovskaya, 1999). The R18-T4 ligase (a 40 nucleotides long hairpin molecule) and several similar ligases, therefore, would have commonly formed.

In the second step, active small RNAs (like R18-T4 ligase) generally ligated other oligomers (35mers) in the pool to their own end via their own defined base-pairing. By this process, small RNA molecules gave rise to a diverse pool of larger and structurally more complex molecules. The increase in structural complexity of the molecules imparted thermodynamic stability to them. This was essential to maintain the structural and functional integrity of the molecular pool before it succumbed to environmental degradation. It is known that phylogenetically variable auxiliary structural elements do not directly participate in catalysis, however, they do have an essential role in the stability and structure of catalytic RNAs. Examples include ribonuclease P ribozyme (Darr et al., 1992,

Westhof and Massire, 2004), trans-activation of the Tetrahymena group I intron ribozyme (Ikawa et al., 1999), and natural hammerhead ribozymes (Penedo et al., 2004).

In the third step, a pool of diverse and longer molecules substantially allowed the appearance of phenotypic variants of RNA for further selection under different microenvironments. Informational complexity has been found to be an important determinant in accessing functional RNAs from a random pool (Carothers et al., 2004, Sabeti et al., 1997, Joyce, 2002b). The increase in phenotypic diversity increased the functional repertoire of the pool. Furthermore, the diverse larger molecules formed were embedded with small ligase motifs. This would have allowed the emergence of molecules like the R18-T3 ligase (75mer) and similar ligases. Such catalysts generally ligated oligomers (35mers) to their own end and, further, increased the pool diversity of larger and structurally more stable molecules. Likewise, in successive steps, self-ligation function guided by common ligase motifs increased the pool complexity in leaps shortening the evolutionary time for the emergence of larger functional catalysts. Such a step-wise process would have led to the formation of ligases like R18-T2 (100mer), R18-T1 (142mer) and eventually the R18 polymerase in the pool. The ability of all the smaller substructures of R18 polymerase (both as individual modules and in combined structures) to form progressively longer molecules by ligating oligonucleotides evidently suggested that R18 polymerase evolved in a modular fashion. Computational and theoretical methods have predicted that evolution of molecules essential for an RNA based life could have resulted from functional short RNA molecules which could increase in structural complexity by virtue of their ligation function (Manrubia and Briones, 2007, Briones et al.,

2009). This kind of modular assembly has also been proposed for evolution of the contemporary ribosomal RNA (Bokov and Steinberg, 2009).

Process 2: Structural and functional stability of molecules

The generality of ligation in the small molecules assisted in their elongation, thus, giving rise to larger molecules. Additionally, this increased the diversity of structurally more complex molecules in the pool. The formation of diverse catalytic molecules from common structural elements is evident in contemporary ribozymes. Conserved structural elements similar to those found in the bacterial RNase P RNAs have been identified in all archaeal counterparts (Evans et al., 2006, Brown, 1999), suggesting a common evolutionary origin for the RNA subunit and a shared catalytic core in all RNase P RNAs. The diversification process based on ligation of structural motifs could have led to progressive building of structural and functional complexity. With increasing length of the molecules, however, the base pairing probabilities and tertiary contacts within the molecule would have also increased. Unrestricted elongation events could have increased the phenotypic instability and loss of essential functional folds. A general elongation mechanism could, therefore, have led to an elongation catastrophe (Kun et al., 2015, Fernando et al., 2007). How was this catastrophe avoided?

In the model studied, with an increase in size and structural complexity of RNA ligases, the flexibility in ligating different kinds of substrates decreased. The R18 polymerase exhibited limited elongation potential. In the early stages of RNA world, this restricted addition of arbitrary structural elements by larger ligases minimized folding

alterations and reduced the phenotypic perturbations. This assisted in the preservation of the native secondary structure of the molecules essential for the stabilisation of ligation and polymerisation function. The conservation of structural arrangement and development of functional specialization is evident in the extant functional RNA molecules such as Group I introns, tRNAs (Bailor et al., 2011, Mustoe et al., 2015b, Mustoe et al., 2014, Michel and Westhof, 1990, Mustoe et al., 2015a).

An increase in molecular length and structural complexity via ligation was accompanied by a parallel process of development of self-ligation specificities. This was essential for the phenotypic and functional stability of the catalysts. Such a process could account for selection of RNA polymerase and its smaller component ligase modules from a random pool. The relationship between reduction of phenotypic plasticity by natural selection and origin of modularity has been studied computationally (Ancel and Fontana, 2000). Molecular trade-offs, thus, played an essential role in the emergence and stability of sophisticated functions.

Process 3

The process of diversification and stabilisation of molecules would also be determined by the efficiency of the catalytic molecules. By self-ligation reactions, small RNA ligases formed components of larger molecules with increased structural stability. The low catalytic efficiency and poor thermodynamic stability of small RNAs (like R18-T4 ribozyme) could have, however, depleted such ligases from the early pool. In the early environment, the larger molecules formed could also be subject to a partial fragmentation

process with the environmental fluctuations. Such natural fragmentation processes would have compensated the population of small RNA ligases. The fragmented RNAs would have further participated in ligation reactions in a cycle. Thus, self-ligation reactions assisted with fragmentation processes would have formed interconnected cycles. Such multiple and consistent events would have sustained the process of increase in complexity. The process of fragmentation has been a part of all natural systems and has played a vital role in the evolution of the group II introns tRNA (Zimmerly and Semper, 2015, Mami and Pallet, 2015). Thus, in parallel to other processes, fragmentation could have constituted the early molecular systems.

In the model studied, the smaller component ligases of the R18 polymerase would have been recycled by fragmentation process. These ligases, further, formed the larger component ligases in a cycle. A positive correlation between the size and efficiency of the ligases in this system suggested that the population of larger RNA molecules once established were catalytically more efficient. In the early stages of RNA world, all the coupled processes, thus, occurred in an accelerated fashion until components of a polymerase and eventually a polymerase itself emerged. Once a minimal polymerase like R18 polymerase emerged, polymerisation of simple oligomers was possible. The polymerisation cycle comprising interconnected self-ligation and fragmentation cycles could be envisaged as forming a replicative kind of unit. This form of hypercycle could have sustained the whole molecular system.

It is proposed that a polymerase emerged by a sequence of steps which might not have occurred contemporaneously, however, they were the first steps towards the development of a self-sustained system. The self-ligation reactions were one of the processes underpinning stability and the emergence of larger catalysts and a polymerase. Basic trade-offs and constraints in molecules shaped the development of specialized molecules and a replicative unit. Such a unit was subject to Darwinian selection. With the evolution of such a system, there could be an associated cost. The increase in structural stability in the molecular system could, however, render complex catalysts unable to unfold and be replicated. This could be overcome by the development of a cooperative cycle with other molecules as proposed by Eigen (Eigen and Schuster, 1978, Eigen and Schuster, 1979). A scenario can be envisaged in which RNA molecules with helicase like properties emerged from self-ligation activity of small RNAs. Bifunctional ribozymes and new functional ribozymes have been evolved by combining catalytic RNA motifs (Kumar and Joyce, 2003, Komatsu et al., 2002, Joyce, 2004). A helicase unwound the polymerase and its components. This kind of cooperation could have assisted complex catalysts like polymerases to partially overcome their stable self-structure and be passively replicated in different sequence registers. The helicase was in turn replicated by the polymerase. The role of RNA helicases have been essential for most processes of modern day RNA metabolism such as ribosome biogenesis, pre-mRNA splicing, and translation initiation (Jankowsky, 2011, Yang et al., 2007a). In this way, we can envisage building up a system of many ribozymes that control a metabolic function, all of which are copied by the same polymerase.

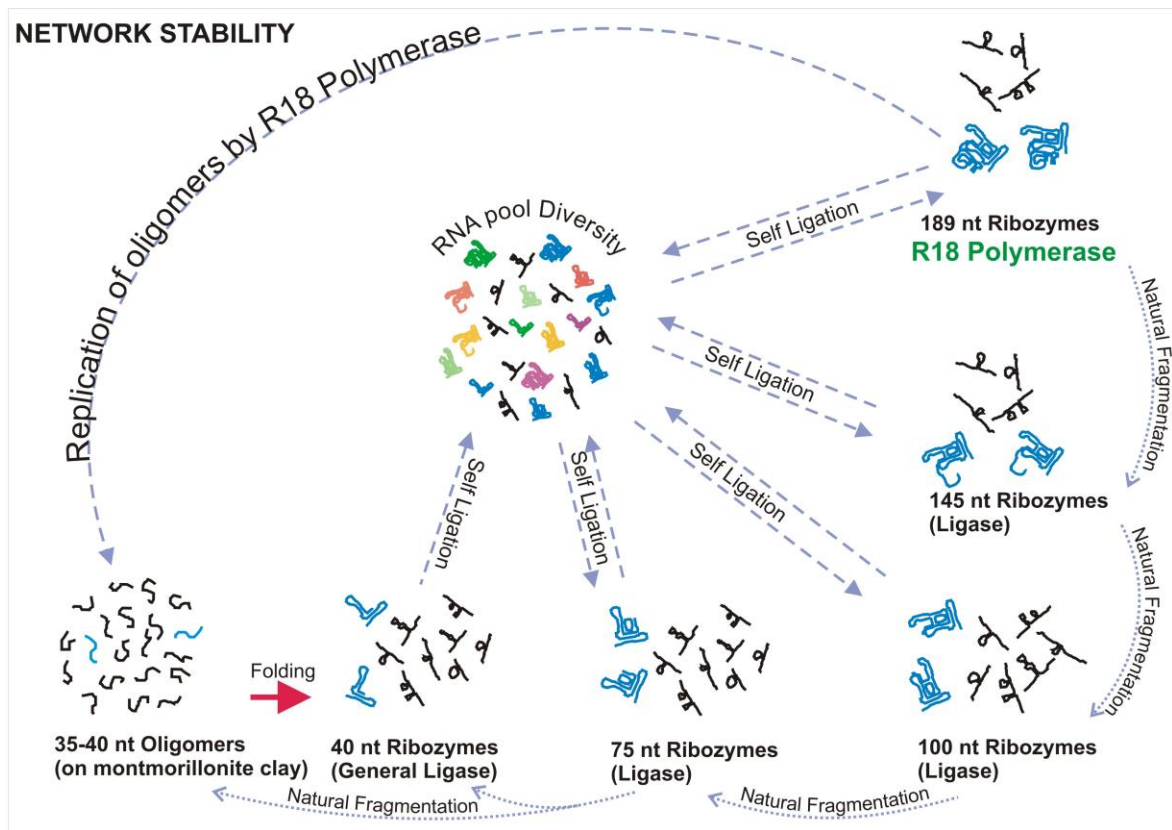


Figure 4.2: Schematic representation of the conceptual model for network stability at the origin of life.

The R18 polymerase ribozyme and its smaller component ligases are depicted in blue and the random substrates in black. The RNA pool diversity depicts the diversity of longer and structurally complex molecules generated from the self-ligation activity of the ribozymes. The variation of phenotypes in the RNA pool is shown in different colours. The number of different random substrates self-ligated by the ribozymes is illustrated around the ribozymes. Arrows represent the processes towards emergence of the R18 polymerase and network stability. Further details are given in the main text.

5. CONCLUSION

In the last four decades an extensive body of work has emerged to support the theory of early life based on an RNA world scenario. The functional versatility and efficiency of RNA molecules has provided a feasible model system to examine the evolution of molecular systems that, to some degree at least are associated with life. The emergence and stability of a complex RNA world from early much simpler oligomers, however, was poorly understood. This work highlights the ecological dynamics of self-ligation processes in shaping the evolution of a minimal polymerase from its reduced elements. It points out that the early evolution of molecular complexity could be a product of catalytic processes based on general structural compatibility. These molecular processes likely played a significant role in the early stages before catalytic molecules with replicative potential emerged.

6. FUTURE DIRECTIONS

Ecology plays a crucial role in the selection of traits at the level of organisms. The ecological dynamics of the catalytic molecular systems for origin of life has been largely unexplored. A complete evolutionary ecology of the early molecular life could be developed using the other RNA catalytic systems. Such research is anticipated to give further insight into the evolutionary origins of complex molecular systems and a replicative unit.

7. APPENDIX A

Materials

- Megashortscript T7 transcription kit (Ambion, USA)
- Urea (Sigma-Aldrich, USA)
- Tris base (Roche Diagnostics, GmbH, Germany)
- Boric Acid (Merck, Germany)
- 0.5 M EDTA solution (pH 8): For 100 ml solution, 18.61 gms of EDTA, disodium salt, dihydrate (Calbiochem, USA) was dissolved in 80 ml of nuclease free water. The pH of the solution was adjusted to 8 using NaOH and the total volume was made up to 100 ml.
- 40% Acrylamide solution (19:1): For a 100 ml solution, 38 gms of Acrylamide (Sigma-Aldrich, USA) and 2 gms of N, N'-Methylenebisacrylamide (Sigma-Aldrich, USA) was dissolved in a total volume of 100 ml nuclease free water. The solution was filtered using 0.45 micron filter.
- 30% w/v Ammonium persulfate solution (Sigma-Aldrich, USA): For 1 ml of solution, 0.3 gms of Ammonium persulfate was dissolved in a total volume of 1 ml of nuclease free water. The solution was filtered using 0.45 micron filter.
- Gel loading Buffer II (supplied with the Megashortscript T7 transcription kit from Ambion, USA)
- Nuclease free water (Sigma-Aldrich, USA)

A.1 Recipe for 10x TBE solution

The following components were dissolved with stirring in about 850 mL nuclease-free water. 108 gms of Tris base, 55 gms of Boric acid, 40 ml of 0.5 M EDTA solution (pH 8).

The final volume was adjusted to 1 litre. For preparing 1x TBE buffer, the above solution was diluted 10 times.

A.2 Recipe for preparing 10% polyacrylamide gel

For 10 ml gel solution, the following components were added

Component	Amount
40% Acrylamide solution	2.5 ml
10x TBE	1 ml
30% (w/v) Ammonium persulfate (APS)	33 μ l
TEMED	2.5 μ l
Nuclease free water	to 10 ml

Note: APS and TEMED were added after dissolving urea and other components. The gel solution was filtered using 0.45 micron filter before pouring.

A.3 Recipe for preparing 8% polyacrylamide-8M urea gel

For 15 ml gel solution, the following components were added:

Component	Amount
8M Urea	7.2 gms
40% Acrylamide	3 ml
10x TBE	1.5 ml
30% (w/v) Ammonium persulfate (APS)	49.5 μ l
TEMED	3.75 μ l
Nuclease free water	to 15 ml

Note: APS and TEMED were added after dissolving urea and other components. The gel solution was filtered using 0.45 micron filter before pouring.

A.4 Protocol for *in vitro* transcription of RNA using the Megashortscript T7 transcription kit

For a 20 μ l transcription reaction, the following components were added

Component	Amount
Template DNA	300-500 ng
T7 10X reaction buffer	2 μ l
T7 ATP Solution (75 mM)	2 μ l
T7 CTP Solution (75 mM)	2 μ l
T7 GTP Solution (75 mM)	2 μ l
T7 UTP Solution (75 mM)	2 μ l
T7 Enzyme Mix	2 μ l
Nuclease free water	to a final volume of 20 μ l

The reaction was incubated at 37 °C for 3 hrs. After incubation, 1 μ l of TURBO DNase was added to the reaction and incubated for 37 °C for 15 mins.

A.5 Preparation of RNA samples before loading on gel

The RNA sample was mixed with one volume of Gel Loading Buffer II. The samples were heated for 3 mins at 95 °C and immediately chilled on ice for 2 mins.

A.6 Gel electrophoresis

Prior to loading of the samples, the gel was pre-run for 30 mins using 1x TBE buffer at a constant current of 25mA. After the pre-run of the gel, the wells were rinsed with 1x TBE

Appendix

buffer using a syringe needle immediately before loading. The RNA samples (prepared as described above) were loaded and the gel was run at a constant current of 25mA.

8. APPENDIX B

Materials

- 10% polyacrylamide gel and 1x TBE Buffer (recipe given in Appendix A)
- Acid-Phenol: Chloroform pH 4.5 (Ambion, USA)
- Ethanol (Merck, South Africa)
- 3M Sodium Acetate (pH 5.2): For 100 ml solution, 24.6 gms of anhydrous Sodium Acetate (Merck, South Africa) was dissolved in 80 ml of nuclease free water. The pH adjusted to 5.2 with glacial acetic acid and the total volume was adjusted to 100 ml.
- Nuclease free water (Sigma-Aldrich, USA)

B.1 Electro-elution protocol for purification of RNAs

A mix of 10% polyacrylamide gel (recipe given in Appendix A) was prepared and 20 μ l of the gel mix was added to 0.5 ml microcentrifuge tubes. The mix was allowed to set for 45 mins. Carefully, without dislodging the polyacrylamide plug using a scalpel, the bottom tip of the tube was cut off. An appropriate volume of 1x TBE buffer (recipe given in Appendix A) was added to the 0.5 ml tube. This tube was placed inside a 1.5 ml microcentrifuge tube with an appropriate amount of the same buffer such that the tip of 0.5 ml tube is just immersed. The excised gel slice was cut into around 2 mm pieces and added to the 0.5 ml tube. A platinum wire connected to anode was inserted into the 1.5 ml microcentrifuge tube and a platinum wire connected to cathode was inserted into the 0.5 ml microcentrifuge tube. The setup (illustrated in Figure B.1) was run at a constant current of 5 mA for 30 mins. The elution in the 1.5 ml microcentrifuge tube was collected. An equal volume of Acid phenol: chloroform, premixed with isoamyl alcohol was added to the

elution and mixed well. It was followed by centrifugation at 13000 rpm for 10 mins. The aqueous part containing RNA was collected in a microcentrifuge tube. Further, precipitation of RNA was done by adding 2.5 volume of ethanol and 1/10th volume of 3M sodium acetate (pH 5.2). The mixture was chilled at -70 °C for 2 hours and immediately centrifuged at 13000 rpm for 1 hour. The pellet was washed with 70% ethanol and centrifuged at 13000 rpm for 15 mins. The washed pellet was dissolved in 20 µl of nuclease free water.

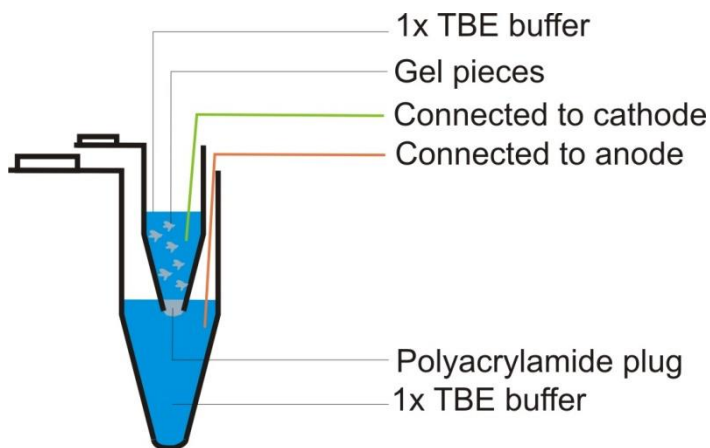


Figure B.1: Electro elution set up for elution of RNA from gel pieces

9. APPENDIX C

C.1 Confirmation of self-ligation activity of R18 RNAs by sequence analysis

The amplicons (224 bp in size) indicating self-ligation activity of R18 RNA with substrates 1, 2, 3, 6, 6a, 6b, 7a were sequenced and aligned with the expected sequence. Sequence alignments confirmed that R18 RNA ligated the substrates to its own 5'end. The alignments are given below.

```

EMBOSS_001      51 CGGGCCCGGGATCCGATTCTCGACGTCAGCCTGGACTAATACGACTCACT      100
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACGACTCACT      32
EMBOSS_001     101 ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC      150
EMBOSS_001      33 ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC      82
EMBOSS_001     151 AGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCG      200
EMBOSS_001      83 AGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCG      132
EMBOSS_001     201 CCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGA      250
EMBOSS_001     133 CCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGA      182
EMBOSS_001     251 TGTTAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCCAATCTAGA      300
EMBOSS_001     183 TGTTAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCC-----      224

```

Figure C.1.1: Sequence alignment of R18 with oligonucleotide substrate 1

```

EMBOSS_001      51  ACGGGCCCCGGGATCCGATTGYCMACTTCCGCATGAACGAATACTACGCAC      100
                    |.|.||||||||||||||||||||||||||||||||
EMBOSS_001      1  -----GTCAACTTCCGCATGAACGAATACTACGCAC      31

EMBOSS_001     101  TAAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      150
                    |||
EMBOSS_001      32  TAAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      81

EMBOSS_001     151  CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      200
                    |||
EMBOSS_001      82  CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      131

EMBOSS_001     201  GCCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCG      250
                    |||
EMBOSS_001     132  GCCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCG      181

EMBOSS_001     251  ATGTTAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCCAATCTAG      300
                    |||
EMBOSS_001     182  ATGTTAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCC-----      224

```

Figure C.1.2: Sequence alignment of R18 with oligonucleotide substrate 2

```

EMBOSS_001      51  GGGCCCCGGGATCCGATTACCSACGACAACCTGGTCTAATACGCCTCACGA      100
                    |||.||||||||||||||||||||||||||||
EMBOSS_001      1  -----CACGACGACAACCTGGTCTAATACGCCTCACGA      33

EMBOSS_001     101  TAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCA      150
                    |||
EMBOSS_001      34  TAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCA      83

EMBOSS_001     151  GAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGC      200
                    |||
EMBOSS_001      84  GAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGC      133

EMBOSS_001     201  CCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGAT      250
                    |||
EMBOSS_001     134  CCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGAT      183

EMBOSS_001     251  GTTAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCCAATCTAGAT      300
                    |||
EMBOSS_001     184  GTTAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCC-----      224

```

Figure C.1.3: Sequence alignment of R18 with oligonucleotide substrate 3

```

EMBOSS_001      851 CGAATGCATCTAGATTCGACGTCAGCCTGGACTAATACTAAAACTATAG      900
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTAAAACTATAG      36
EMBOSS_001      901 GAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAG      950
EMBOSS_001      37 GAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAG      86
EMBOSS_001      951 GAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCA      1000
EMBOSS_001      87 GAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCA      136
EMBOSS_001     1001 ATACTCCCCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGTT      1050
EMBOSS_001     137 ATACTCCCCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGTT      186
EMBOSS_001     1051 AGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCCATCGGATCCCGG      1100
EMBOSS_001     187 AGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCC-----      224

```

Figure C.1.4: Sequence alignment of R18 with oligonucleotide substrate 6

```

EMBOSS_001      851 GCGAATGCATCTAGATTCGACGTCAGCCTGGACTAATACTATTTACTATA      900
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTATTTACTATA      35
EMBOSS_001      901 GGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGA      950
EMBOSS_001      36 GGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGA      85
EMBOSS_001      951 GGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCC      1000
EMBOSS_001      86 GGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCC      135
EMBOSS_001     1001 AATACTCCCCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGT      1050
EMBOSS_001     136 AATACTCCCCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGT      185
EMBOSS_001     1051 TAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCCATCGGATCCCGG      1100
EMBOSS_001     186 TAGACACGCCAAGGTCATAATCCCCGGAGCTTCGGCTCC-----      224

```

Figure C.1.5: Sequence alignment of R18 with oligonucleotide substrate 6a

```

EMBOSS_001      801 GACGGCCAGTGAATTCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTC      850
                                |||
EMBOSS_001      1  -----CTC      3
EMBOSS_001      851 GACGTCAGCCTGGACTAATACTAGGGACTATAGGAAAAGACAAATCTGC      900
|||
EMBOSS_001      4  GACGTCAGCCTGGACTAATACTAGGGACTATAGGAAAAGACAAATCTGC      53
|||
EMBOSS_001      901 CCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGG      950
|||
EMBOSS_001      54 CCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGG      103
|||
EMBOSS_001      951 CGCGATAGCGCCAACGTTCTCAACAGGCGCCAATACTCCCGCTTCGGCG      1000
|||
EMBOSS_001      104 CGCGATAGCGCCAACGTTCTCAACAGGCGCCAATACTCCCGCTTCGGCG      153
|||
EMBOSS_001     1001 GGTGGGGATAACACCTGACGAAAAGGCGATGTTAGACACGCCAAGGTCAT      1050
|||
EMBOSS_001     154 GGTGGGGATAACACCTGACGAAAAGGCGATGTTAGACACGCCAAGGTCAT      203
|||
EMBOSS_001     1051 AATCCCGGAGCTTCGGCTATCGGATCCCGGGCCCGTCGACWGCRGAGGC      1100
|||
EMBOSS_001     204 AATCCCGGAGCTTCGGC-----TCC-----      224

```

Figure C.1.6: Sequence alignment of R18 with oligonucleotide substrate 6b

```

EMBOSS_001      51 CGACGGGCCCGGGATCCGATTCGACGTCAGCCTGGACTATATGGACTCAC      100
                                .|||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGGACTCAC      31
EMBOSS_001     101 TATAGGAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      150
|||
EMBOSS_001     32 TATAGGAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      81
|||
EMBOSS_001     151 CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      200
|||
EMBOSS_001     82 CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      131
|||
EMBOSS_001     201 GCCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCG      250
|||
EMBOSS_001     132 GCCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCG      181
|||
EMBOSS_001     251 ATGTTAGACACGCCAAGGTCATAATCCCGGAGCTTCGGCCATCTAGATG      300
|||
EMBOSS_001     182 ATGTTAGACACGCCAAGGTCATAATCCCGGAGCTTCGGC--TCC-----      224

```

Figure C.1.7: Sequence alignment of R18 with oligonucleotide substrate 7A

C.2 Confirmation of self-ligation activity of R18-T1 RNAs by sequence analysis

The amplicons (177 bp in size) indicating self-ligation activity of R18-T1 RNA with substrates 1, 2, 3, 4, 6, 7, 6a, 6b, 7a, 8b were sequenced and aligned with the expected sequence. Sequence alignments confirmed that R18-T1 RNA ligated the substrates to its own 5' end. The alignments are given below.

```

EMBOSS_001      901 CGGCCAGTGAATTCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTCGA      950
                                     |||||
EMBOSS_001       1  -----CTCGA      5
EMBOSS_001      951 CGTCAGCCTGGACTAATACGACTCACTATAGGAAAAGACAAATCTGCC      1000
                                     |||||
EMBOSS_001       6  CGTCAGCCTGGACTAATACGACTCACTATAGGAAAAGACAAATCTGCC      55
EMBOSS_001     1001 TCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCG      1050
                                     |||||
EMBOSS_001      56  TCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCG      105
EMBOSS_001     1051 CGATAGCGCCAACGTTCTCAACAGGCGCCAATACTCCCGCTTCGGCGGG      1100
                                     |||||
EMBOSS_001     106  CGATAGCGCCAACGTTCTCAACAGGCGCCAATACTCCCGCTTCGGCGGG      155
EMBOSS_001     1101 TGGGATAACACCTGRCGAATCGGATCCCGGGCCCGTCGACTGCRGRGGC      1150
                                     |||||.|||.
EMBOSS_001     156 TGGGATAACACCTGACGAAAA-----      177

```

Figure C.2.1: Sequence alignment of R18-T1 with oligonucleotide substrate 1

```

EMBOSS_001      901 GCATCTAGATTGTCAACTTCCGCATGAACGAATACTACGCACTAAAGGAA      950
                                     |||||
EMBOSS_001       1  -----GTCAACTTCCGCATGAACGAATACTACGCACTAAAGGAA      39
EMBOSS_001      951 AAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAG      1000
                                     |||||
EMBOSS_001      40  AAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAG      89
EMBOSS_001     1001 GCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCAATA      1050
                                     |||||
EMBOSS_001      90  GCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCAATA      139
EMBOSS_001     1051 CTCCCGCTTCGGCGGGTGGGGATAACACCTGACRAAAAAATCGGATCCCG      1100
                                     |||||.|||.
EMBOSS_001     140 CTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAA-----      177

```

Figure C.2.2: Sequence alignment of R18-T1 with oligonucleotide substrate 2

```

EMBOSS_001      51  ACGGGCCCGGGATCCGATTCACGACGACAACCTGGTCTAATACGCCTCAC      100
                   |||
EMBOSS_001      1  -----CACGACGACAACCTGGTCTAATACGCCTCAC      31
                   |||
EMBOSS_001     101  GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      150
                   |||
EMBOSS_001      32  GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      81
                   |||
EMBOSS_001     151  CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      200
                   |||
EMBOSS_001      82  CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      131
                   |||
EMBOSS_001     201  GCCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAAATC      250
                   |||
EMBOSS_001     132  GCCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAA----      177

```

Figure C.2.3: Sequence alignment of R18-T1 with oligonucleotide substrate 3

```

EMBOSS_001     851  GGTACCTCGCGAATGCATCTAGATTCTGGATGTAAGTCTTGAATATATGG      900
                   |||
EMBOSS_001      1  -----CTGGATGTAAGTCTTGAATATATGG      25
                   |||
EMBOSS_001     901  AATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT      950
                   |||
EMBOSS_001      26  AATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT      75
                   |||
EMBOSS_001     951  CGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCA      1000
                   |||
EMBOSS_001      76  CGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCA      125
                   |||
EMBOSS_001    1001  ACAGGCGCCAATACTCCCGCTTCGGCGGGTGGGGATAACMCCTGACGAA      1050
                   |||
EMBOSS_001     126  ACAGGCGCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAA      175
                   |||
EMBOSS_001    1051  AAAATCGGATCCCGGGCCCGTTCGACTGCAGAGGCCTGCATGCAAGCTTCC      1100
                   ||
EMBOSS_001     176  AA-----      177

```

Figure C.2.4: Sequence alignment of R18-T1 with oligonucleotide substrate 4

Appendix

```
EMBOSS_001      851 CCAGTGAATTCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTCGACGT      900
EMBOSS_001      1  -----CTCGACGT      8
EMBOSS_001      901 CAGCCTGGACTAATACTAAAACTATAGGAAAAGACAAATCTGCCCTCA      950
EMBOSS_001      9  CAGCCTGGACTAATACTAAAACTATAGGAAAAGACAAATCTGCCCTCA      58
EMBOSS_001      951 GAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGA      1000
EMBOSS_001      59  GAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGA      108
EMBOSS_001     1001 TAGCGCCAACGTTCTCAACAGGCGCCAATACTCCCGCTTCGGCGGGTGG      1050
EMBOSS_001     109  TAGCGCCAACGTTCTCAACAGGCGCCAATACTCCCGCTTCGGCGGGTGG      158
EMBOSS_001     1051 GGATAACACCTGACGAAAAATCGGATCCCGGGCCCGTCGACTGCRGAGGC      1100
EMBOSS_001     159  GGATAACACCTGACGAAAA-----CTCGACGT      177
```

Figure C.2.5: Sequence alignment of R18-T1 with oligonucleotide substrate 6

```
EMBOSS_001      51 GACGGGCCCGGATCCGATTCSACGTCAGCCTGGACTAATACTATTACT      100
EMBOSS_001      1  -----CTCGACGT      32
EMBOSS_001     101 ATAGGAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC      150
EMBOSS_001     33  ATAGGAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC      82
EMBOSS_001     151 AGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCG      200
EMBOSS_001     83  AGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCG      132
EMBOSS_001     201 CCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAACTAG      250
EMBOSS_001     133 CCCAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAA-----      177
```

Figure C.2.6: Sequence alignment of R18-T1 with oligonucleotide substrate 6a

```
EMBOSS_001      51 AGATTCTCGACGTCAGCCTGGACTAATACTAGGGACTATAGGAAAAGAC      100
EMBOSS_001      1  -----CTCGACGT      45
EMBOSS_001     101 AAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCC      150
EMBOSS_001     46  AAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCC      95
EMBOSS_001     151 TCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATACTCCCG      200
EMBOSS_001     96  TCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATACTCCCG      145
EMBOSS_001     201 CTTCCGGCGGGTGGGGATAACACCTGACGAAAAAATCGGATCCCGGGCCCG      250
EMBOSS_001     146 CTTCCGGCGGGTGGGGATAACACCTGACGAAAA-----      177
```

Figure C.2.7: Sequence alignment of R18-T1 with oligonucleotide substrate 6b

```

EMBOSS_001      901 TAGATTCTCGACGTCAGCCTGGACTATATGGACTCACTATAGGAAAAAGA      950
      |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGGACTCACTATAGGAAAAAGA      44
EMBOSS_001      951 CAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGC      1000
      |||
EMBOSS_001      45  CAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGC      94
EMBOSS_001     1001 CTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATACTCCC      1050
      |||
EMBOSS_001      95  CTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATACTCCC      144
EMBOSS_001     1051 GCTTCGGCGGGTGGGGATAACACCTRACGAAAAAATCGGATCCCGGGCCC      1100
      |||
EMBOSS_001     145  GCTTCGGCGGGTGGGGATAACACCTGACGAAAA-----              177

```

Figure C.2.8: Sequence alignment of R18-T1 with oligonucleotide substrate 7A

```

EMBOSS_001      901 CGAATGCATCTAGATTCTCGACGTCAGCCTGGACTATATGGAATCGCTCG      950
      |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGGAATCGCTCG      34
EMBOSS_001      951 AGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAG      1000
      |||
EMBOSS_001      35  AGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAG      84
EMBOSS_001     1001 AGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC      1050
      |||
EMBOSS_001      85  AGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC      134
EMBOSS_001     1051 CAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAAATCGGA      1100
      |||
EMBOSS_001     135  CAATACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAA-----      177

```

Figure C.2.9: Sequence alignment of R18-T1 with oligonucleotide substrate 7

```

EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACCCTCGCCTCAAGGAAAAAGA      44
      |||
EMBOSS_001     901 TAGATTCTCGACGTCAGCCTGGACTAATACCCTCGCCTCAAGGAAAAAGA      950
EMBOSS_001      45  CAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGC      94
      |||
EMBOSS_001     951 CAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGC      1000
EMBOSS_001      95  CTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATACTCCC      144
      |||
EMBOSS_001     1001 CTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATACTCCC      1050
EMBOSS_001     145  GCTTCGGCGGGTGGGGATAACACCTGACGAAAA-----              177
      |||
EMBOSS_001     1051 GCTTCGGCGGGTGGGGATAACWCCTGACGAAAAAATCGGATCCCGGGCCC      1100

```

Figure C.2.10: Sequence alignment of R18-T1 with oligonucleotide substrate 8B

C.3 Confirmation of self-ligation activity of R18-T2 RNAs by sequence analysis

The amplicons (135 bp in size) indicating self-ligation activity of R18-T2 RNA with substrates 1, 2, 3, 4, 5, 7, 8, 6a, 6b, 7a, 7b, 8a were sequenced and aligned with the expected sequence. Sequence alignments confirmed that R18-T2 RNA ligated the substrates to its own 5' end. The alignments are given below.

```
EMBOSS_001      951  ATTCTCGACGTCAGCCTGGACTAATACGACTCACTATAGGAAAAAGACAA 1000
      |||
EMBOSS_001      1    ---CTCGACGTCAGCCTGGACTAATACGACTCACTATAGGAAAAAGACAA  47
EMBOSS_001     1001  ATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTC 1050
      |||
EMBOSS_001      48  ATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTC  97
EMBOSS_001     1051  CGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGSCCAATCGGATCCCG 1100
      |||
EMBOSS_001      98  CGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC----- 135
```

Figure C.3.1: Sequence alignment of R18-T2 with oligonucleotide substrate 1

```
EMBOSS_001      951  TCTAGATTGTCAACTTCCGCATGAACGAATACTACGCACTAAAGGAAAAA 1000
      |||
EMBOSS_001      1    -----GTCAACTTCCGCATGAACGAATACTACGCACTAAAGGAAAAA  42
EMBOSS_001     1001  GACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCA 1050
      |||
EMBOSS_001      43  GACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCA  92
EMBOSS_001     1051  GCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCSAATCGGA 1100
      |||
EMBOSS_001      93  GCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC----- 135
```

Figure C.3.2: Sequence alignment of R18-T2 with oligonucleotide substrate 2

```

EMBOSS_001      51  ACGGGCCCCGGGATCCGATTCACGACGACAACTGGTCTAATACGCCTCAC      100
                    |||
EMBOSS_001      1  -----CACGACGACAACTGGTCTAATACGCCTCAC      31

EMBOSS_001     101  GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      150
                    |||
EMBOSS_001      32  GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      81

EMBOSS_001     151  CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      200
                    |||
EMBOSS_001      82  CAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGC      131

EMBOSS_001     201  GCCCAATCTAGATGCATTCGCGAGGTACCGAGCTCGAATTCACTGGCCGT      250
                    |||
EMBOSS_001     132  GCCC-----                                  135
    
```

Figure C.3.3: Sequence alignment of R18-T2 with oligonucleotide substrate 3

```

EMBOSS_001     951  TCGGTACCTCGCGAATGCATCTAGATTCTGGATGTAAGTCTTGAATATAT      1000
                    |||
EMBOSS_001      1  -----CTGGATGTAAGTCTTGAATATAT      23

EMBOSS_001    1001  GGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATC      1050
                    |||
EMBOSS_001      24  GGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATC      73

EMBOSS_001    1051  TTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCT      1100
                    |||
EMBOSS_001      74  TTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCT      123

EMBOSS_001    1101  CWACAGGCGCCCAATCGGATCCCGGGCCCGTCGACTGCAGAGGCCTGCAT      1150
                    |.|||
EMBOSS_001     124  CAACAGGCGCCC-----                                  135
    
```

Figure C.3.4: Sequence alignment of R18-T2 with oligonucleotide substrate 4

```

EMBOSS_001      51  GACGGGCCCGGATCCGATTTAATACTCATAACGACTACATGGACCTCGC      100
                    |||
EMBOSS_001      1  -----TAATACTCATAACGACTACATGGACCTCGC      30

EMBOSS_001     101  CTCAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      150
                    |||
EMBOSS_001      31  CTCAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      80

EMBOSS_001     151  GCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGG      200
                    |||
EMBOSS_001      81  GCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGG      130

EMBOSS_001     201  CGCCCAATCTAGATGCATTCGCGAGGTACCGAGCTCGAATTCACTGGCCG      250
                    |||
EMBOSS_001     131  CGCC-----                                  135
    
```

Figure C.3.5: Sequence alignment of R18-T2 with oligonucleotide substrate 5

Appendix

```
EMBOSS_001      951 CTCGACGTCAGCCTGGACTAATACTATTTACTATAGGAAAAAGACAAATC 1000
      |||
EMBOSS_001      1  CTCGACGTCAGCCTGGACTAATACTATTTACTATAGGAAAAAGACAAATC 50
      |||
EMBOSS_001     1001 TGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGG 1050
      |||
EMBOSS_001      51  TGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGG 100
      |||
EMBOSS_001     1051 TGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCAATCGGATCCCGGGCC 1100
      |||
EMBOSS_001      101 TGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC----- 135
```

Figure C.3.6: Sequence alignment of R18-T2 with oligonucleotide substrate 6a

```
EMBOSS_001      51  AGATTCTCGACGTCAGCCTGGACTAATACTAGGGACTATAGGAAAAAGAC 100
      |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTAGGGACTATAGGAAAAAGAC 45
      |||
EMBOSS_001     101  AAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCC 150
      |||
EMBOSS_001      46  AAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCC 95
      |||
EMBOSS_001     151  TCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCAATCGGATCC 200
      |||
EMBOSS_001      96  TCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC----- 135
```

Figure C.3.7: Sequence alignment of R18-T2 with oligonucleotide substrate 6b

```
EMBOSS_001      951 ATGCATCTAGATTCTCGACGTCAGCCTGGACTATATGGACTCACTATAGG 1000
      |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGGACTCACTATAGG 37
      |||
EMBOSS_001     1001 AAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGG 1050
      |||
EMBOSS_001      38  AAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGG 87
      |||
EMBOSS_001     1051 AGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGSGCCCAA 1100
      |||
EMBOSS_001      88  AGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCC-- 135
```

Figure C.3.8: Sequence alignment of R18-T2 with oligonucleotide substrate 7A

Appendix

```
EMBOSS_001      951 TACCTCGCGAATGCATCTAGATTCTCGACGTCAGCCTGGACTATATGGAA 1000
                    |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGGAA 27
EMBOSS_001     1001 TCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCG 1050
                    |||
EMBOSS_001      28 TCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCG 77
EMBOSS_001     1051 GATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAAC 1100
                    |||
EMBOSS_001      78 GATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAAC 127
EMBOSS_001     1101 AGGCGCCCAATCGGATCCCGGGCCCGTCGACTGCRGAGGCCTGCATGCAA 1150
                    |||
EMBOSS_001     128 AGGCGCCC----- 135
```

Figure C.3.9: Sequence alignment of R18-T2 with oligonucleotide substrate 7

```
EMBOSS_001      951 CTAGATTCTCGACGTCAGCCTGGACTAATACGAATCGCTCGAGGAAAAAG 1000
                    |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACGAATCGCTCGAGGAAAAAG 43
EMBOSS_001     1001 ACAAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAG 1050
                    |||
EMBOSS_001      44 ACAAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAG 93
EMBOSS_001     1051 CCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCCAATCGGAT 1100
                    |||
EMBOSS_001      94 CCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCGCCC----- 135
```

Figure C.3.10: Sequence alignment of R18-T2 with oligonucleotide substrate 7B

```
EMBOSS_001      51 CGGGCCCGGATCCGATTTCYCSACKTCAGCCTGGACCATGGAGACTCACT 100
                    |.|.|||.|||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACCATGGAGACTCACT 32
EMBOSS_001     101 ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC 150
                    |||
EMBOSS_001     33 ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC 82
EMBOSS_001     151 AGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCG 200
                    |||
EMBOSS_001     83 AGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCAACGTTCTCAACAGGCG 132
EMBOSS_001     201 CCCAATCTAGATGCATTTCGCGAGGTACCGAGCTCGAATTCAGTGGCCGTC 250
                    |||
EMBOSS_001     133 CCC----- 135
```

Figure C.3.11: Sequence alignment of R18-T2 with oligonucleotide substrate 8A

```

EMBOSS_001      951 TTCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTCGACGTCAGCCTGG      1000
                               |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGG      16
EMBOSS_001     1001 ACCATGGACCTCGCCTCAAGGAAAAGACAAATCTGCCCTCAGAGCTTGA      1050
                               |||
EMBOSS_001      17  ACCATGGACCTCGCCTCAAGGAAAAGACAAATCTGCCCTCAGAGCTTGA      66
EMBOSS_001     1051 GAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCA      1100
                               |||
EMBOSS_001      67  GAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAGCGCCA      116
EMBOSS_001     1101 ACGTTCTCAACAGGSGCCCAATCGGATCCCGGGCCCGTGCAGTGCAGAG      1150
                               |||
EMBOSS_001     117  ACGTTCTCAACAGGCGCC-----      135

```

Figure C.3.12: Sequence alignment of R18-T2 with oligonucleotide substrate 8

C.4 Confirmation of self-ligation activity of R18-T3 RNAs by sequence analysis

The amplicons (135 bp in size) indicating self-ligation activity of R18-T3 RNA with substrates 1, 2, 3, 4, 5, 6, 7, 6a, 6b, 7a, 7b, 8a, 8b were sequenced and aligned with the expected sequence. Sequence alignments confirmed that R18-T3 RNA ligated the substrates to its own 5' end. The alignments are given below.

```

EMBOSS_001      51 GACGGGCCCGGATCCGATTCYCGACKTCAGCCTGGACTAATACGACTCA      100
                               |.||||.||||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACGACTCA      30
EMBOSS_001     101 CTATAGGAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      150
                               |||
EMBOSS_001     31  CTATAGGAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      80
EMBOSS_001     151 GCAGAGGAGGCAGCCTCCGGTGGCGCGATAATCTAGATGCATTCGCGAGG      200
                               |||
EMBOSS_001     81  GCAGAGGAGGCAGCCTCCGGTGGCGCGATA-----      110

```

Figure C.4.1: Sequence alignment of R18-T3 with oligonucleotide substrate 1

```

EMBOSS_001      951 CGAGCTCGGTACCTCGCGAATGCATCTAGATTGTCAACTTCCGCATGAAC 1000
                    |||
EMBOSS_001      1  -----GTCAACTTCCGCATGAAC 18
EMBOSS_001     1001 GAATACTACGCACTAAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGA 1050
                    |||
EMBOSS_001      19  GAATACTACGCACTAAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGA 68
EMBOSS_001     1051 ACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGSGATAAATCGGAT 1100
                    |||
EMBOSS_001      69  ACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATA----- 110

```

Figure C.4.2: Sequence alignment of R18-T3 with oligonucleotide substrate 2

```

EMBOSS_001      51 ACGGGCCCGGGATCCGATTCACGACGACAACCTGGTCTAATACGCCTCAC 100
                    |||
EMBOSS_001      1  -----CACGACGACAACCTGGTCTAATACGCCTCAC 31
EMBOSS_001     101 GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG 150
                    |||
EMBOSS_001      32  GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG 81
EMBOSS_001     151 CAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCTAGATGCATTGCGGAG 200
                    |||
EMBOSS_001      82  CAGAGGAGGCAGCCTCCGGTGGCGCGATA----- 110

```

Figure C.4.3: Sequence alignment of R18-T3 with oligonucleotide substrate 3

```

EMBOSS_001      51 GACGGCCCGGGATCCGATTCCTGGATGTAAGTCTTGAATATATGGAATCG 100
                    |||
EMBOSS_001      1  -----CTGGATGTAAGTCTTGAATATATGGAATCG 30
EMBOSS_001     101 CTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT 150
                    |||
EMBOSS_001      31  CTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT 80
EMBOSS_001     151 GCAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCTAGATGCATTGCGGAG 200
                    |||
EMBOSS_001      81  GCAGAGGAGGCAGCCTCCGGTGGCGCGATA----- 110

```

Figure C.4.4: Sequence alignment of R18-T3 with oligonucleotide substrate 4


```

EMBOSS_001      51  ACGGGCCCCGGGATCCGATTTAATACTCATAACGACTACATGGACCTCGCC      100
                    |||
EMBOSS_001      1   -----TAATACTCATAACGACTACATGGACCTCGCC      31

EMBOSS_001     101  TCAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      150
                    |||
EMBOSS_001     32  TCAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      81

EMBOSS_001     151  CAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCTAGATGCATTTCGCGAGG      200
                    |||
EMBOSS_001     82  CAGAGGAGGCAGCCTCCGGTGGCGCGATA-----      110

```

Figure C.4.5: Sequence alignment of R18-T3 with oligonucleotide substrate 5

```

EMBOSS_001      51  GACGGGCCCGGGATCCGATTCYCGACGTCAGCCTGGACTAATACTAAAAA      100
                    |.
EMBOSS_001      1   -----CTCGACGTCAGCCTGGACTAATACTAAAAA      30

EMBOSS_001     101  CTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      150
                    |||
EMBOSS_001     31  CTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      80

EMBOSS_001     151  GCAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCTAGATGCATTTCGCGAGG      200
                    |||
EMBOSS_001     81  GCAGAGGAGGCAGCCTCCGGTGGCGCGATA-----      110

```

Figure C.4.6: Sequence alignment of R18-T3 with oligonucleotide substrate 6

```

EMBOSS_001      51  GACGGGCCCGGGATCCGATTCYCGACGTCAGCCTGGACTAATACTATTTA      100
                    |||
EMBOSS_001      1   -----CTCGACGTCAGCCTGGACTAATACTATTTA      30

EMBOSS_001     101  CTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      150
                    |||
EMBOSS_001     31  CTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGAT      80

EMBOSS_001     151  GCAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCTAGATGCATTTCGCGAGG      200
                    |||
EMBOSS_001     81  GCAGAGGAGGCAGCCTCCGGTGGCGCGATA-----      110

```

Figure C.4.7: Sequence alignment of R18-T3 with oligonucleotide substrate 6A

```

EMBOSS_001      51  CGGGCCCGGGATCCGATTCTCGACGTCAGCCTGGACTAATACTAGGGACT   100
                   |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTAGGGACT   32
                   |||
EMBOSS_001     101  ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC   150
                   |||
EMBOSS_001      33  ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGC   82
                   |||
EMBOSS_001     151  AGAGGAGGCAGCCTCCGGTGGCGCGATAATCTAGATGCATTCGCGAGGTA   200
                   |||
EMBOSS_001      83  AGAGGAGGCAGCCTCCGGTGGCGCGATA-----   110

```

Figure C.4.8: Sequence alignment of R18-T3 with oligonucleotide substrate 6b

```

EMBOSS_001     951  GGTACCTCGCGAATGCATCTAGATTCTCGACGTCAGCCTGGACTATATGG   1000
                   |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGG   25
                   |||
EMBOSS_001    1001  ACTCACTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT   1050
                   |||
EMBOSS_001      26  ACTCACTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT   75
                   |||
EMBOSS_001    1051  CGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCGGATCCCGGGC   1100
                   |||
EMBOSS_001      76  CGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATA-----   110

```

Figure C.4.9: Sequence alignment of R18-T3 with oligonucleotide substrate 7A

```

EMBOSS_001     951  ATTCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTCGACGTCAGCCTG   1000
                   |||
EMBOSS_001      1  -----CTCGACGTCAGCCTG   15
                   |||
EMBOSS_001    1001  GACTATATGGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTG   1050
                   |||
EMBOSS_001      16  GACTATATGGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTG   65
                   |||
EMBOSS_001    1051  AGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATAAATCG   1100
                   |||
EMBOSS_001      66  AGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGGTGGCGCGATA-----   110

```

Figure C.4.10: Sequence alignment of R18-T3 with oligonucleotide substrate 7

Appendix

```
EMBOSS_001      51 ACGGGCCCGGGATCCGATTTCYCGACGTCAGCCTGGACTAATACGAATCGC      100
                   |.|||||||||||||||||||||||||||||||||||||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACGAATCGC      31

EMBOSS_001     101 TCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      150
                   |||||||||||||||||||||||||||||||||||||||
EMBOSS_001     32 TCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATG      81

EMBOSS_001     151 CAGAGGAGGCAGCCTCCGGTGGCGCGATAATCTAGATGCATTCGCGAGGT      200
                   |||||||||||||||||||||||
EMBOSS_001     82 CAGAGGAGGCAGCCTCCGGTGGCGCGATA-----      110
```

Figure C.4.11: Sequence alignment of R18-T3 with oligonucleotide substrate 7B

```
EMBOSS_001     901 CTCGACGTCAGCCTGGACCATGGAGACTCACTATAGGAAAAAGACAAATC      950
                   |||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1  CTCGACGTCAGCCTGGACCATGGAGACTCACTATAGGAAAAAGACAAATC      50

EMBOSS_001     951 TGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGK      1000
                   |||||||||||||||||||||||||||||||||||.
EMBOSS_001      51 TGCCCTCAGAGCTTGAGAACATCTTCGGATGCAGAGGAGGCAGCCTCCGG      100

EMBOSS_001    1001 TTKCGCGATAAATCGGATCCCGGGCCCGTCGACTGSAGAGGCCTGCATKC      1050
                   |..|||||
EMBOSS_001    101 TGCGCGATA-----      110
```

Figure C.4.12: Sequence alignment of R18-T3 with oligonucleotide substrate 8A

```
EMBOSS_001      51 GGGCCCGGGATCCGATTCTCGACGTCAGCCTGGACTAATACCCTCGCCTC      100
                   |||||||||||||||||||||||||||||||||||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACCCTCGCCTC      33

EMBOSS_001     101 AAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCA      150
                   |||||||||||||||||||||||||||||||||||
EMBOSS_001     34 AAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTCGGATGCA      83

EMBOSS_001     151 GAGGAGGCAGCCTCCGGTGGCGCGATAAATCTAGATGCATTCGCGAGGTA      200
                   |||||||||||||||||||
EMBOSS_001     84 GAGGAGGCAGCCTCCGGTGGCGCGATA-----      110
```

Figure C.4.13: Sequence alignment of R18-T3 with oligonucleotide substrate 8B

C.5 Confirmation of self-ligation activity of R18-T4 RNAs by sequence analysis

The amplicons (135 bp in size) indicating self-ligation activity of R18-T4 RNA with substrates 1, 2, 3, 4, 5, 6, 7, 6a, 6b, 7a, 7b, 8a, 8b were sequenced and aligned with the expected sequence. Sequence alignments confirmed that R18-T4 RNA ligated the substrates to its own 5' end. The alignments are given below.

```
EMBOSS_001      51 GACGGGCCCGGGATCCGATTCYCGACGTCAGCCTGGACTAATACGACTCA      100
                  |.|||||||||||||||||||||||||||||||||||||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACGACTCA      30

EMBOSS_001     101 CTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCT      150
                  |||||||||||||||||||||||||||||||||||
EMBOSS_001     31 CTATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
```

Figure C.5.1: Sequence alignment of R18-T4 with oligonucleotide substrate 1

```
EMBOSS_001     1001 TCGCGAATGCATCTAGATTGTCAACTTCCGCATGAACGAATACTACGCAC      1050
                  |||||||||||||||||||||||||||||||||||
EMBOSS_001      1  -----GTCAACTTCCGCATGAACGAATACTACGCAC      31

EMBOSS_001     1051 TAAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCGG      1100
                  |||||||||||||||||||||||||||||||||||
EMBOSS_001     32 TAAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
```

Figure C.5.2: Sequence alignment of R18-T4 with oligonucleotide substrate 2

```
EMBOSS_001      51 ACGGGCCCGGGATCCGATTCACGACGACAACCTGGTCTAATACGCCTCAC      100
                  |||||||||||||||||||||||||||||||||||
EMBOSS_001      1  -----CACGACGACAACCTGGTCTAATACGCCTCAC      31

EMBOSS_001     101 GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCTA      150
                  |||||||||||||||||||||||||||||||||||
EMBOSS_001     32 GATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
```

Figure C.5.3: Sequence alignment of R18-T4 with oligonucleotide substrate 3

```

EMBOSS_001      51  ACGGGCCCCGGGATCCGATCTGGATGTAAGTCTTGAATATATGGAATCGC      100
                |||
EMBOSS_001      1  -----CTGGATGTAAGTCTTGAATATATGGAATCGC      31
EMBOSS_001     101  TCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCTA      150
                |||
EMBOSS_001      32  TCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
    
```

Figure C.5.4: Sequence alignment of R18-T4 with oligonucleotide substrate 4

```

EMBOSS_001      51  ACGGGCCCCGGGATCCGATTTAATACTCATAACGACTACATGGACCTCGCC      100
                |||
EMBOSS_001      1  -----TAATACTCATAACGACTACATGGACCTCGCC      31
EMBOSS_001     101  TCAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCTA      150
                |||
EMBOSS_001      32  TCAAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
    
```

Figure C.5.5: Sequence alignment of R18-T4 with oligonucleotide substrate 5

```

EMBOSS_001     1001  GCATCTAGATTCTCGACGTCAGCCTGGACTAATACTAAAACTATAGGAA      1050
                |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTAAAACTATAGGAA      39
EMBOSS_001     1051  AAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCGGATCCCGGG      1100
                |||
EMBOSS_001      40  AAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
    
```

Figure C.5.6: Sequence alignment of R18-T4 with oligonucleotide substrate 6

```

EMBOSS_001     1001  TGCATCTAGATTCTCGACGTCAGCCTGGACTAATACTATTTACTATAGGAAA      1050
                .|||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTATTTACTATAGGAAA      40
EMBOSS_001     1051  AAGACAAATCTGCCCTCAGAGCTTGAGAACA KCTTATCGGATCCCGGGCC      1100
                |||
EMBOSS_001      41  AAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
    
```

Figure C.5.7: Sequence alignment of R18-T4 with oligonucleotide substrate 6a

```

EMBOSS_001      51  CGGGCCCGGGATCCGATTCTCGACGTCAGCCTGGACTAATACTAGGGACT      100
                               |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTAATACTAGGGACT      32
EMBOSS_001     101  ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAATCTAG      150
                               |||
EMBOSS_001     33  ATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT-----      75
    
```

Figure C.5.8: Sequence alignment of R18-T4 with oligonucleotide substrate 6b

```

EMBOSS_001      51  KTCGACGGGCCCGGGATCCGATTCTCGACGTCAGCCTGGACTATATGGAC      100
                               |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGACTATATGGAC      27
EMBOSS_001     101  TCACATATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTTAA      150
                               |||
EMBOSS_001     28  TCACATATAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCTT--      75
    
```

Figure C.5.9: Sequence alignment of R18-T4 with oligonucleotide substrate 7A

```

EMBOSS_001     1001  ATTCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTCGACGTCAGCCTG      1050
                               |||
EMBOSS_001      1  -----CTCGACGTCAGCCTG      15
EMBOSS_001     1051  GACTATATGGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTG      1100
                               |||
EMBOSS_001     16  GACTATATGGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTG      65
EMBOSS_001     1101  AGAACATCTTAATCGGATCCCGGGCCGTCGACTGCRGAGGCCTGCATGC      1150
                               |||
EMBOSS_001     66  AGAACATCTT-----      75
    
```

Figure C.5.10: Sequence alignment of R18-T4 with oligonucleotide substrate 7

```

EMBOSS_001     1001  TCGAGCTCGGTACCTCGCGAATGCATCTAGATTCTCGACGTCAGCCTGGA      1050
                               |||
EMBOSS_001      1  -----CTCGACGTCAGCCTGGA      17
EMBOSS_001     1051  CTAATACGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAG      1100
                               |||
EMBOSS_001     18  CTAATACGAATCGCTCGAGGAAAAAGACAAATCTGCCCTCAGAGCTTGAG      67
EMBOSS_001     1101  AACATCTTAATCGGATCCCGGGCCGTCGACTGCRGAGGCCTGCATGCAA      1150
                               |||
EMBOSS_001     68  AACATCTT-----      75
    
```

Figure C.5.11: Sequence alignment of R18-T4 with oligonucleotide substrate 7B

10. APPENDIX D

Motif Letter-Probability matrices

The letter probability matrix is a table of probabilities where the rows are positions in the sequence pattern and the columns are four nucleotides A, C, G, and T. Probability matrix gives the probability of each nucleotide at each position in the sequence pattern.

Position (5' to 3')	A	C	G	T
19	0	0.35	0.25	0.4
20	0.5	0.2	0.3	0
21	0.3	0.15	0.35	0.2
22	0.1	0.3	0.3	0.3
23	0.3	0.2	0.4	0.1
24	0.1	0.6	0.25	0.05
25	0.05	0.25	0.45	0.25
26	0.7	0.1	0	0.2
27	0.2	0.4	0.25	0.15
28	0.25	0.15	0.1	0.5
29	0.05	0.5	0.4	0.05
30	0.45	0.3	0.15	0.1
31	0	0.9	0.1	0
32	0	0	0	1
33	0.45	0.4	0.05	0.1
34	0.1	0.15	0.1	0.65
35	1	0	0	0

Table D.1: Probability matrix of all the substrates used in MEME analysis

Position (5' to 3')	A	C	G	T
19	0	0.1111	0	0.8888
20	1	0	0	0
21	0.6666	0	0	0.3333
22	0.2222	0	0.1111	0.6666
23	0.6666	0	0.1111	0.2222
24	0.1111	0.6666	0.2222	0
25	0	0.1111	0.5555	0.3333
26	0.8888	0.1111	0	0
27	0.3333	0.3333	0.1111	0.2222
28	0.1111	0.1111	0.1111	0.6666
29	0.1111	0.5555	0.2222	0.1111
30	0.6666	0.1111	0.2222	0
31	0	1	0	0
32	0	0	0	1
33	0.6666	0.3333	0	0
34	0.1111	0	0.2222	0.6666
35	1	0	0	0

Position (5' to 3')	A	C	G	T
19	0	0.5	0.5	0
20	0	0.4	0.6	0
21	0	0.3	0.7	0
22	0	0.6	0.4	0
23	0	0.4	0.6	0
24	0	0.6	0.3	0.1
25	0.1	0.3	0.4	0.2
26	0.6	0	0	0.4
27	0.1	0.5	0.4	0
28	0.4	0.1	0.1	0.4
29	0	0.5	0.5	0
30	0.3	0.4	0.1	0.2
31	0	0.8	0.2	0
32	0	0	0	1
33	0.3	0.4	0.1	0.2
34	0	0.3	0	0.7
35	1	0	0	0

Table D.2: Probability matrix of the substrates for the analysis of R18-T4 activity.

Probability matrix of the substrates ligated (left) and not ligated by R18-T4 ribozyme (right).

Position (5' to 3')	A	C	G	T	Position (5' to 3')	A	C	G	T
19	0	0.1111	0	0.8888	19	0	0.5	0.5	0
20	1	0	0	0	20	0	0.4	0.6	0
21	0.6666	0	0	0.3333	21	0	0.3	0.7	0
22	0.2222	0	0.1111	0.6666	22	0	0.6	0.4	0
23	0.6666	0	0.1111	0.2222	23	0	0.4	0.6	0
24	0.1111	0.6666	0.2222	0	24	0	0.6	0.3	0.1
25	0	0.1111	0.5555	0.3333	25	0.1	0.3	0.4	0.2
26	0.8888	0.1111	0	0	26	0.6	0	0	0.4
27	0.3333	0.3333	0.1111	0.2222	27	0.1	0.5	0.4	0
28	0.1111	0.1111	0.1111	0.6666	28	0.4	0.1	0.1	0.4
29	0.1111	0.5555	0.2222	0.1111	29	0	0.5	0.5	0
30	0.6666	0.1111	0.2222	0	30	0.3	0.4	0.1	0.2
31	0	1	0	0	31	0	0.8	0.2	0
32	0	0	0	1	32	0	0	0	1
33	0.6666	0.3333	0	0	33	0.3	0.4	0.1	0.2
34	0.1111	0	0.2222	0.6666	34	0	0.3	0	0.7
35	1	0	0	0	35	1	0	0	0

Table D.3: Probability matrix of the substrates for the analysis of R18-T3 activity.

Probability matrix of the substrates ligated (left) and not ligated by R18-T3 ribozyme (right).

Position (5' to 3')	A	C	G	T
19	0	0.1428	0	0.8571
20	1	0	0	0
21	0.5714	0	0	0.4285
22	0.2857	0	0.1428	0.5714
23	0.5714	0	0.1428	0.2857
24	0.1428	0.5714	0.2857	0
25	0	0	0.7142	0.2857
26	1	0	0	0
27	0.2857	0.4285	0.1428	0.1428
28	0	0	0.1428	0.8571
29	0	0.7142	0.1428	0.1428
30	0.7142	0	0.2857	0
31	0	1	0	0
32	0	0	0	1
33	0.7142	0.2857	0	0
34	0	0	0.2857	0.7142
35	1	0	0	0

Position (5' to 3')	A	C	G	T
19	0	0.5	0.5	0
20	0	0.4	0.6	0
21	0	0.3	0.7	0
22	0	0.6	0.4	0
23	0	0.4	0.6	0
24	0	0.6	0.3	0.1
25	0.1	0.3	0.4	0.2
26	0.6	0	0	0.4
27	0.1	0.5	0.4	0
28	0.4	0.1	0.1	0.4
29	0	0.5	0.5	0
30	0.3	0.4	0.1	0.2
31	0	0.8	0.2	0
32	0	0	0	1
33	0.3	0.4	0.1	0.2
34	0	0.3	0	0.7
35	1	0	0	0

Table D.4: Probability matrix of the substrates for the analysis of R18-T2 activity.

Probability matrix of the substrates ligated (left) and not ligated by R18-T2 ribozyme (right).

Position (5' to 3')	A	C	G	T	Position (5' to 3')	A	C	G	T
19	0	0	0	1	19	0	0	0	0
20	1	0	0	0	20	0	0	0	0
21	0.7142	0	0	0.2857	21	0	0	0	0
22	0.2857	0	0	0.7142	22	0	0	0	0
23	0.7142	0	0	0.2857	23	0	0	0	0
24	0	0.7142	0.2857	0	24	0	0	0	0
25	0	0.1428	0.4285	0.4285	25	0	0	0	0
26	0.8571	0.1428	0	0	26	0	0	0	0
27	0.2857	0.2857	0.1428	0.2857	27	0	0	0	0
28	0.1428	0.1428	0.1428	0.5714	28	0	0	0	0
29	0.1428	0.4285	0.2857	0.1428	29	0	0.5384	0.4615	0
30	0.7142	0.1428	0.1428	0	30	0.3076	0.3846	0.1538	0.1538
31	0	1	0	0	31	0	0.8461	0.1538	0
32	0	0	0	1	32	0	0	0	1
33	0.7142	0.2857	0	0	33	0.3076	0.4615	0.0769	0.1538
34	0.1428	0	0.1428	0.7142	34	0.0769	0.2307	0.0769	0.6153
35	1	0	0	0	35	1	0	0	0

Table D.5: Probability matrix of the substrates for the analysis of R18-T1 activity.

Probability matrix of the substrates ligated (left) and not ligated by R18-T1 ribozyme (right).

Position (5' to 3')	A	C	G	T	Position (5' to 3')	A	C	G	T
19	0	0	0	1	19	0	0	0	0
20	1	0	0	0	20	0	0	0	0
21	0.8	0	0	0.2	21	0	0	0	0
22	0.2	0	0	0.8	22	0	0	0	0
23	0.8	0	0	0.2	23	0	0	0	0
24	0	0.8	0.2	0	24	0	0	0	0
25	0	0	0.4	0.6	25	0.0666	0.3333	0.4666	0.1333
26	1	0	0	0	26	0.6	0.1333	0	0.2666
27	0.2	0.4	0.2	0.2	27	0.2	0.4	0.2666	0.1333
28	0.2	0	0.2	0.6	28	0.2666	0.2	0.0666	0.4666
29	0.2	0.4	0.2	0.2	29	0	0.5333	0.4666	0
30	1	0	0	0	30	0.2666	0.4	0.2	0.1333
31	0	1	0	0	31	0	0.8666	0.1333	0
32	0	0	0	1	32	0	0	0	1
33	1	0	0	0	33	0.2666	0.5333	0.0666	0.1333
34	0	0	0	1	34	0.1333	0.2	0.1333	0.5333
35	1	0	0	0	35	1	0	0	0

Table D.6: Probability matrix of the substrates for the analysis of R18 activity.

Probability matrix of the substrates ligated (left) and not ligated by R18 ribozyme (right).

11. APPENDIX E

Standard Curves used for quantification of ribozymes ligation activity

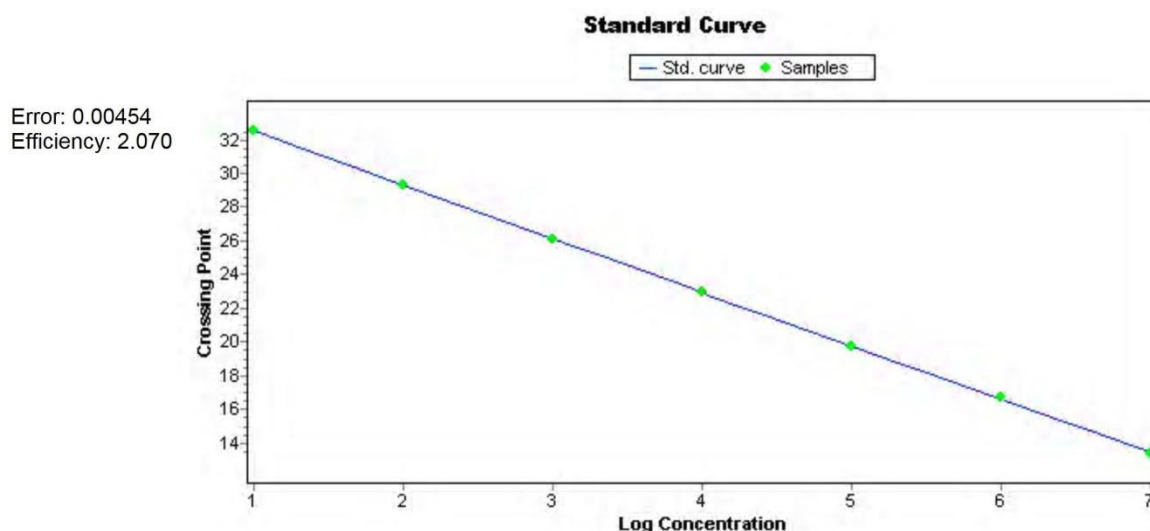


Figure E.1: Standard curve for quantification of R18 ribozyme self-ligation activity.

The curve was generated using log base 10 of initial target copy number versus corresponding crossing point (CP) values obtained (given in the table below).

Inc	Pos	Name	Type	CP	Concentration	Standard
<input checked="" type="checkbox"/>	1	E	Standard	13.37	1.04E7	1.00E7
<input checked="" type="checkbox"/>	2	F	Standard	16.66	9.51E5	1.00E6
<input checked="" type="checkbox"/>	3	G	Standard	19.77	9.94E4	1.00E5
<input checked="" type="checkbox"/>	4	H	Standard	22.94	9.87E3	1.00E4
<input checked="" type="checkbox"/>	5	I	Standard	26.06	1.03E3	1.00E3
<input checked="" type="checkbox"/>	6	J	Standard	29.30	9.95E1	1.00E2
<input checked="" type="checkbox"/>	7	K	Standard	32.55	1.00E1	1.00E1
<input checked="" type="checkbox"/>	8	No DNA	Unknown	>35.00	[<1.83E0]	

Table E.1: Crossing point (CP) values of DNA copies of R18 with substrate 1.

Samples E, F, G, H, I, J and K were the prepared dilutions corresponding to 10^7 , 10^6 , 10^5 , 10^4 , 10^3 , 10^2 , 10 copies of DNA per μl respectively. A sample with no DNA was also prepared.

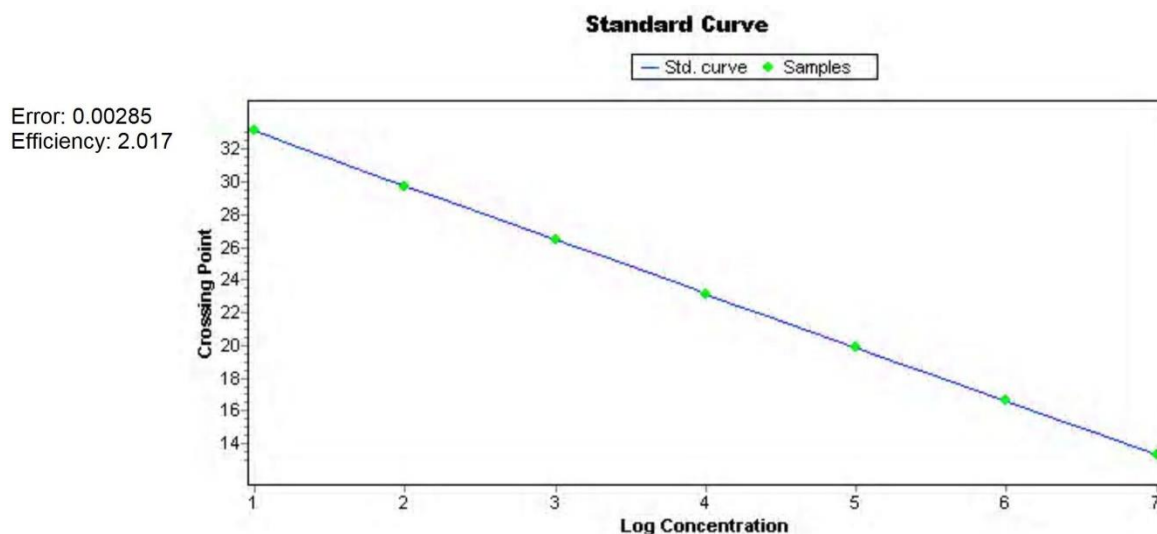


Figure E.2: Standard curve for quantification of R18-T1 ribozyme self-ligation activity.

The curve was generated using log base 10 of initial target copy number versus corresponding crossing point (CP) values obtained (given in the table below).

Inc	Pos	Name	Type	CP	Concentratio	Standard
<input checked="" type="checkbox"/>	1	E	Standard	13.32	1.01E7	1.00E7
<input checked="" type="checkbox"/>	2	F	Standard	16.64	9.77E5	1.00E6
<input checked="" type="checkbox"/>	3	G	Standard	19.85	1.03E5	1.00E5
<input checked="" type="checkbox"/>	4	H	Standard	23.16	1.01E4	1.00E4
<input checked="" type="checkbox"/>	5	I	Standard	26.49	9.75E2	1.00E3
<input checked="" type="checkbox"/>	6	J	Standard	29.72	1.01E2	1.00E2
<input checked="" type="checkbox"/>	7	K	Standard	33.14	1.00E1	1.00E1
<input checked="" type="checkbox"/>	8	No DNA	Unknown	[>35.00]	[<3.07E0]	

Table E.2: Crossing point (CP) values of DNA copies of R18-T1 with substrate 1.

Samples E, F, G, H, I, J and K were the prepared dilutions corresponding to 10^7 , 10^6 , 10^5 , 10^4 , 10^3 , 10^2 , 10 copies of DNA per μl respectively. A sample with no DNA was also prepared.

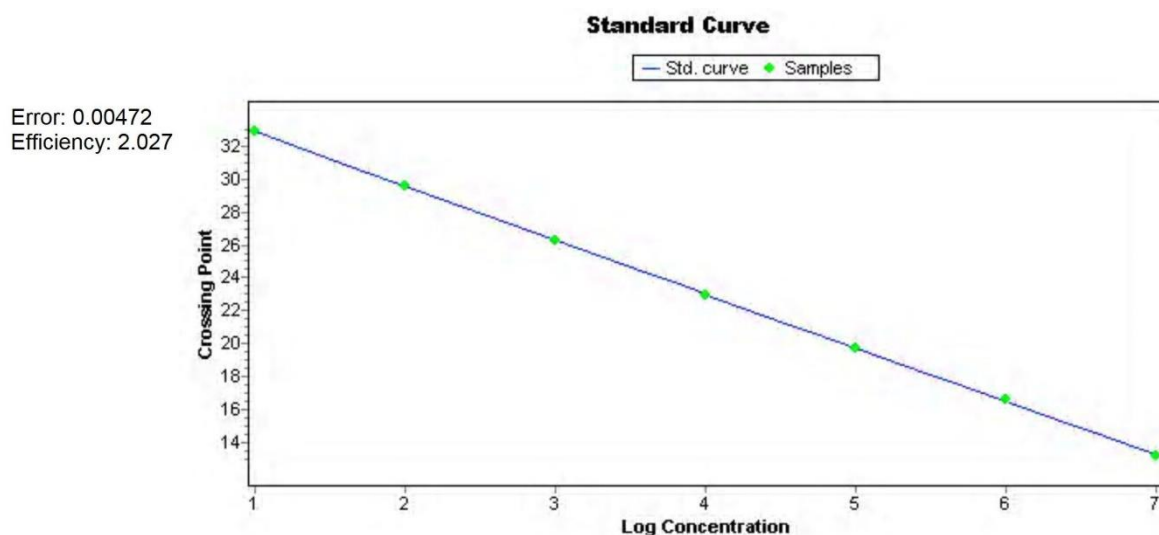


Figure E.3: Standard curve for quantification of R18-T2 ribozyme self-ligation activity.

The curve was generated using log base 10 of initial target copy number versus corresponding crossing point (CP) values obtained (given in the table below).

Inc	Pos	Name	Type	CP	Concentration	Standard
<input checked="" type="checkbox"/>	1	E	Standard	13.16	1.05E7	1.00E7
<input checked="" type="checkbox"/>	2	F	Standard	16.55	9.48E5	1.00E6
<input checked="" type="checkbox"/>	3	G	Standard	19.76	9.82E4	1.00E5
<input checked="" type="checkbox"/>	4	H	Standard	22.97	1.02E4	1.00E4
<input checked="" type="checkbox"/>	5	I	Standard	26.24	1.01E3	1.00E3
<input checked="" type="checkbox"/>	6	J	Standard	29.57	9.84E1	1.00E2
<input checked="" type="checkbox"/>	7	K	Standard	32.92	1.00E1	1.00E1
<input checked="" type="checkbox"/>	8	No DNA	Unknown	>35.00	[<2.51E0]	

Table E.3: Crossing point (CP) values of DNA copies of R18-T2 with substrate 1.

Samples E, F, G, H, I, J and K were the prepared dilutions corresponding to 10^7 , 10^6 , 10^5 , 10^4 , 10^3 , 10^2 , 10 copies of DNA per μl respectively. A sample with no DNA was also prepared.

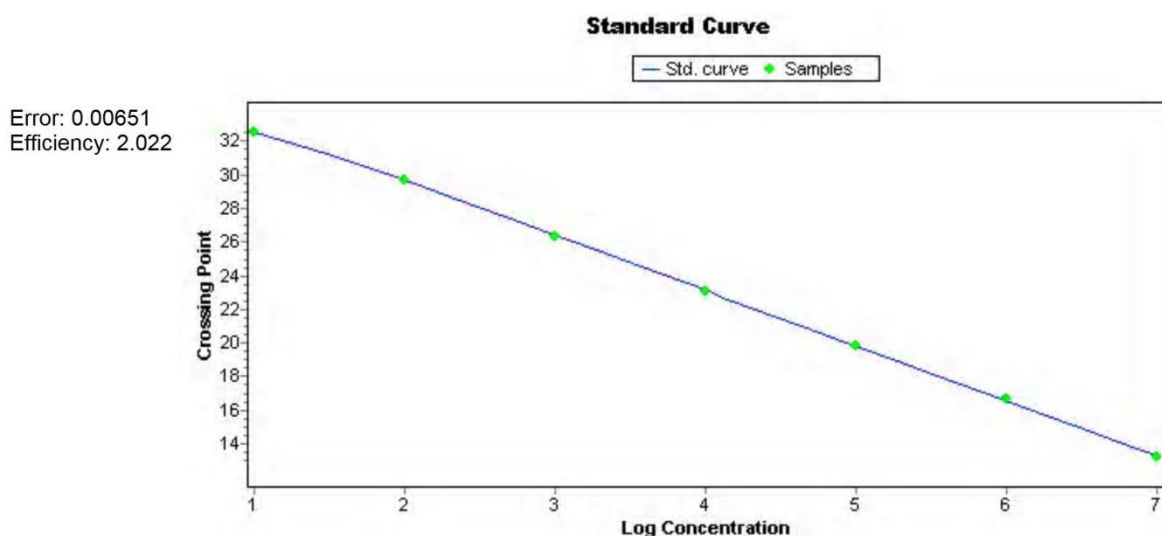


Figure E.4: Standard curve for quantification of R18-T3 ribozyme self-ligation activity.

The curve was generated using log base 10 of initial target copy number versus corresponding crossing point (CP) values obtained (given in the table below).

Inc	Pos	Name	Type	CP	Concentratio	Standard
<input checked="" type="checkbox"/>	1	E	Standard	13.22	1.06E7	1.00E7
<input checked="" type="checkbox"/>	2	F	Standard	16.69	9.21E5	1.00E6
<input checked="" type="checkbox"/>	3	G	Standard	19.84	1.00E5	1.00E5
<input checked="" type="checkbox"/>	4	H	Standard	23.10	1.01E4	1.00E4
<input checked="" type="checkbox"/>	5	I	Standard	26.33	1.03E3	1.00E3
<input checked="" type="checkbox"/>	6	J	Standard	29.67	9.87E1	1.00E2
<input checked="" type="checkbox"/>	7	K	Standard	32.52	1.00E1	1.00E1
<input checked="" type="checkbox"/>	8	No DNA	Unknown	>35.00	[<8.59E-1]	

Table E.4: Crossing point (CP) values of DNA copies of R18-T3 with substrate 1.

Samples E, F, G, H, I, J and K were the prepared dilutions corresponding to 10^7 , 10^6 , 10^5 , 10^4 , 10^3 , 10^2 , 10 copies of DNA per μl respectively. A sample with no DNA was also prepared.

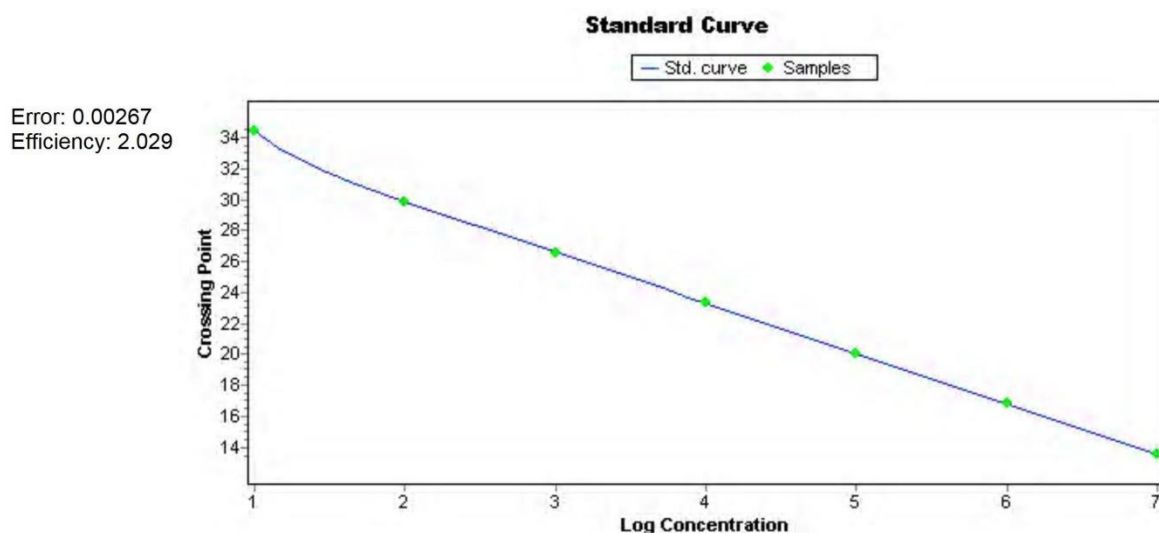


Figure E.5: Standard curve for quantification of R18-T4 ribozyme self-ligation activity.

The curve was generated using log base 10 of initial target copy number versus corresponding crossing point (CP) values obtained (given in the table below).

Inc	Pos	Name	Type	CP	Concentratio	Standard
<input checked="" type="checkbox"/>	1	E	Standard	13.56	9.96E6	1.00E7
<input checked="" type="checkbox"/>	2	F	Standard	16.84	9.77E5	1.00E6
<input checked="" type="checkbox"/>	3	G	Standard	20.02	1.03E5	1.00E5
<input checked="" type="checkbox"/>	4	H	Standard	23.30	1.01E4	1.00E4
<input checked="" type="checkbox"/>	5	I	Standard	26.57	9.97E2	1.00E3
<input checked="" type="checkbox"/>	6	J	Standard	29.84	9.88E1	1.00E2
<input checked="" type="checkbox"/>	7	K	Standard	34.42	1.00E1	1.00E1
<input checked="" type="checkbox"/>	8	No DNA	Unknown	>35.00	[<8.57E0]	

Table E.5: Crossing point (CP) values of DNA copies of R18-T4 with substrate 1.

Samples E, F, G, H, I, J and K were the prepared dilutions corresponding to 10^7 , 10^6 , 10^5 , 10^4 , 10^3 , 10^2 , 10 copies of DNA per μl respectively. A sample with no DNA was also prepared.

12. ETHICS WAIVER CERTIFICATE



UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
7 York Road, Parktown, 2193 South Africa * E-mail physiology@health.wits.ac.za * Telephone (011)717-2363 * Fax (011) 643-2765

29th October 2015

To: Whom it may concern,

Re: **Waiver from Animal Ethics Screening Committee**

This letter is to confirm that Nisha Dhar currently enrolled for postgraduate studies (PhD) in the department of molecular medicine does not require clearance from the Animal Ethics committee for the study titled "Ribozyme evolution and molecular trade-offs at the origin of life".

The study supervised by Dr Pierre Durand utilises artificially synthesised DNA or RNA molecules. No cells, tissues or animals are used in this project. The project is based entirely on synthetic molecules.

Yours sincerely,

A handwritten signature in black ink, appearing to read 'K. Erlwanger'.

Kennedy Erlwanger
(Chairman: Animal Ethics Screening Committee, University of the Witwatersrand)

Reference: Dhar Waiver 29.10.15-O

Assoc Prof Kennedy H. Erlwanger
School of Physiology
Faculty of Health Sciences, University of the Witwatersrand
7 York Road, Parktown, 2193
SOUTH AFRICA

Private bag 3, Wits, 2050, South Africa.
Tel: +27 (0)11 717 2454
Fax: +27 (0)11 643 2765
Email: Kennedy.Erlwanger@wits.ac.za

13. REFERENCES

- Acevedo, O. L. & Orgel, L. E. 1987. Non-enzymatic transcription of an oligodeoxynucleotide 14 residues long. *J Mol Biol*, 197, 187-93.
- Ancel, L. W. & Fontana, W. 2000. Plasticity, evolvability, and modularity in RNA. *J Exp Zool*, 288, 242-83.
- Arenas, C. D. & Lehman, N. 2010. The continuous evolution in vitro technique. *Curr Protoc Nucleic Acid Chem*, Chapter 9, Unit 9 7 1-17.
- Athavale, S. S., Spicer, B. & Chen, I. A. 2014. Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Curr Opin Chem Biol*, 22, 35-9.
- Attwater, J., Wochner, A. & Holliger, P. 2013. In-ice evolution of RNA polymerase ribozyme activity. *Nat Chem*, 5, 1011-8.
- Bagby, S. C., Bergman, N. H., Shechner, D. M., Yen, C. & Bartel, D. P. 2009. A class I ligase ribozyme with reduced Mg²⁺ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA*, 15, 2129-46.
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. & Noble, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, 37, W202-8.
- Bailor, M. H., Mustoe, A. M., Brooks, C. L., 3rd & Al-Hashimi, H. M. 2011. Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation. *Curr Opin Struct Biol*, 21, 296-305.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, 289, 905-20.

References

- Bartel, D. P. 1999. 5 Re-creating an RNA Replicase. *Cold Spring Harbor Monograph Archive*, 37, 143-162.
- Bartel, D. P. & Szostak, J. W. 1993. Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science*, 261, 1411-8.
- Bean, H. D., Sheng, Y., Collins, J. P., Anet, F. A., Leszczynski, J. & Hud, N. V. 2007. Formation of a beta-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. *J Am Chem Soc*, 129, 9556-7.
- Benner, S. A., Ellington, A. D. & Tauer, A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci U S A*, 86, 7054-8.
- Bokov, K. & Steinberg, S. V. 2009. A hierarchical model for evolution of 23S ribosomal RNA. *Nature*, 457, 977-80.
- Borenstein, E. & Ruppin, E. 2006. Direct evolution of genetic robustness in microRNA. *Proc Natl Acad Sci U S A*, 103, 6593-8.
- Bourdeau, V., Ferbeyre, G., Pageau, M., Paquin, B. & Cedergren, R. 1999. The distribution of RNA motifs in natural sequences. *Nucleic Acids Res*, 27, 4457-67.
- Breaker, R. R. 2012. Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology*, 4, a003566.
- Briones, C., Stich, M. & Manrubia, S. C. 2009. The dawn of the RNA World: toward functional complexity through ligation of random RNA oligomers. *RNA*, 15, 743-9.
- Brown, J. W. 1999. The Ribonuclease P Database. *Nucleic Acids Res*, 27, 314.
- Burke, D. H. & Willis, J. H. 1998. Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. *RNA*, 4, 1165-75.
- Butlerow, A. 1861. Bildung einer zuckerartigen Substanz durch Synthese. *Justus Liebigs Annalen der Chemie*, 120, 295-298.
-

References

- Buzayan, J. M., Gerlach, W. L. & Bruening, G. 1986. Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA.
- Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. 2016. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nat Commun*, 7, 11328.
- Carothers, J. M., Oestreich, S. C., Davis, J. H. & Szostak, J. W. 2004. Informational complexity and functional activity of RNA structures. *J Am Chem Soc*, 126, 5130-7.
- Chapman, K. B. & Szostak, J. W. 1995. Isolation of a ribozyme with 5'-5' ligase activity. *Chem Biol*, 2, 325-33.
- Chen, X., Li, N. & Ellington, A. D. 2007. Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers*, 4, 633-55.
- Cheng, L. K. & Unrau, P. J. 2010. Closing the circle: replicating RNA with RNA. *Cold Spring Harb Perspect Biol*, 2, a002204.
- Chyba, C. & Sagan, C. 1992. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature*, 355, 125-32.
- Conner, J. K. & Hartl, D. L. 2004. *A primer of ecological genetics*, Sinauer Associates Incorporated.
- Crick, F. H. 1968. The origin of the genetic code. *J Mol Biol*, 38, 367-79.
- Darr, S. C., Zito, K., Smith, D. & Pace, N. R. 1992. Contributions of phylogenetically variable structural elements to the function of the ribozyme ribonuclease P. *Biochemistry*, 31, 328-33.
- De Visser, J. A., Hermisson, J., Wagner, G. P., Ance Meyers, L., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., Gibson, G.,
-

- Hansen, T. F., Krakauer, D., Lewontin, R. C., Ofria, C., Rice, S. H., Von Dassow, G., Wagner, A. & Whitlock, M. C. 2003. Perspective: Evolution and detection of genetic robustness. *Evolution*, 57, 1959-72.
- Deck, C., Jauker, M. & Richert, C. 2011. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat Chem*, 3, 603-8.
- Diaz Arenas, C. & Lehman, N. 2013. Partitioning the fitness components of RNA populations evolving in vitro. *PLoS One*, 8, e84454.
- Doudna, J. A., Couture, S. & Szostak, J. W. 1991. A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science*, 251, 1605-8.
- Durand, P. M. & Michod, R. E. 2010. Genomics in the light of evolutionary transitions. *Evolution*, 64, 1533-40.
- Ebert, M. O., Mang, C., Krishnamurthy, R., Eschenmoser, A. & Jaun, B. 2008. The structure of a TNA-TNA complex in solution: NMR study of the octamer duplex derived from alpha-(L)-threofuranosyl-(3'-2')-CGAATTCG. *J Am Chem Soc*, 130, 15105-15.
- Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58, 465-523.
- Eigen, M. & Schuster, P. 1977. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, 64, 541-65.
- Eigen, M. & Schuster, P. 1978. The hypercycle. *Naturwissenschaften*, 65, 7-41.
- Eigen, M. & Schuster, P. 1979. The hypercycle. *New York*.
- Ekland, E. H. & Bartel, D. P. 1995. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res*, 23, 3231-8.
-

- Ekland, E. H. & Bartel, D. P. 1996. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature*, 382, 373-6.
- Ekland, E. H., Szostak, J. W. & Bartel, D. P. 1995. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, 269, 364-70.
- Ellington, A. D. 1994a. RNA selection. Aptamers achieve the desired recognition. *Curr Biol*, 4, 427-9.
- Ellington, A. D. 1994b. The RNA world. Empirical explorations of sequence space: Host-guest chemistry in the RNA world. *Berichte der Bunsengesellschaft für physikalische Chemie*, 98, 1115-1121.
- Ellington, A. D., Chen, X., Robertson, M. & Syrett, A. 2009. Evolutionary origins and directed evolution of RNA. *Int J Biochem Cell Biol*, 41, 254-65.
- Ellington, A. D. & Szostak, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346, 818-22.
- Engelhart, A. E., Powner, M. W. & Szostak, J. W. 2013. Functional RNAs exhibit tolerance for non-heritable 2'-5' versus 3'-5' backbone heterogeneity. *Nat Chem*, 5, 390-4.
- Erwin, S. 1944. What is life? The physical aspect of the living cell. *Cambridge (niem.: Was ist Leben)*.
- Evans, D., Marquez, S. M. & Pace, N. R. 2006. RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci*, 31, 333-41.
- Falconer, D. S. 1983. *Problems on quantitative genetics*, Longman.
- Fernando, C., Von Kiedrowski, G. & Szathmary, E. 2007. A stochastic model of nonenzymatic nucleic acid replication: "elongators" sequester replicators. *J Mol Evol*, 64, 572-85.
-

- Ferré-D'amaré, A. R. & Scott, W. G. 2010. Small self-cleaving ribozymes. *Cold Spring Harbor perspectives in biology*, 2, a003574.
- Ferris, J. P. 2002. Montmorillonite catalysis of 30-50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world. *Orig Life Evol Biosph*, 32, 311-32.
- Ferris, J. P. 2006. Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Philos Trans R Soc Lond B Biol Sci*, 361, 1777-86; discussion 1786.
- Ferris, J. P. & Ertem, G. 1993. Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA. *J Am Chem Soc*, 115, 12270-5.
- Ferris, J. P., Hill, A. R., Jr., Liu, R. & Orgel, L. E. 1996. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381, 59-61.
- Ferris, J. P., Sanchez, R. A. & Orgel, L. E. 1968. Studies in prebiotic synthesis. 3. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J Mol Biol*, 33, 693-704.
- Flatt, T. & Heyland, A. 2011. *Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-offs*, Oxford University Press.
- Fontana, W., Konings, D. A., Stadler, P. F. & Schuster, P. 1993. Statistics of RNA secondary structures. *Biopolymers*, 33, 1389-404.
- Fuller, W. D., Sanchez, R. A. & Orgel, L. E. 1972a. Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides. *J Mol Biol*, 67, 25-33.
- Fuller, W. D., Sanchez, R. A. & Orgel, L. E. 1972b. Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides. *J Mol Evol*, 1, 249-57.
-

References

- Garst, A. D., Edwards, A. L. & Batey, R. T. 2011. Riboswitches: structures and mechanisms. *Cold Spring Harbor perspectives in biology*, 3, a003533.
- Gevertz, J., Gan, H. H. & Schlick, T. 2005. In vitro RNA random pools are not structurally diverse: a computational analysis. *RNA*, 11, 853-63.
- Gilbert, W. 1986. Origin of life: The RNA world. *Nature*, 319.
- Grafen, A. 1984. Natural selection, kin selection and group selection. *Behavioural ecology: an evolutionary approach*, 2.
- 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, 849-57.
- Handsuh, G. J., Lohrmann, R. & Orgel, L. E. 1973. The effect of Mg²⁺ and Ca²⁺ on urea-catalyzed phosphorylation reactions. *J Mol Evol*, 2, 251-62.
- Hayden, E. J. & Lehman, N. 2006. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol*, 13, 909-18.
- Hayden, E. J., Von Kiedrowski, G. & Lehman, N. 2008. Systems chemistry on ribozyme self-construction: evidence for anabolic autocatalysis in a recombination network. *Angew Chem Int Ed Engl*, 47, 8424-8.
- Hazen, R. M. 2001. Life's rocky start. *Sci Am*, 284, 76-85.
- Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. 2005. RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys*, 38, 221-43.
- Higgs, P. G. & Lehman, N. 2015. The RNA World: molecular cooperation at the origins of life. *Nat Rev Genet*, 16, 7-17.
- Hill, A. R., Jr., Orgel, L. E. & Wu, T. 1993. The limits of template-directed synthesis with nucleoside-5'-phosphoro(2-methyl)imidazolides. *Orig Life Evol Biosph*, 23, 285-90.
-

- Hordijk, W. & Steel, M. 2004. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J Theor Biol*, 227, 451-61.
- Hordijk, W. & Steel, M. 2013. A formal model of autocatalytic sets emerging in an RNA replicator system. *J Syst Chem*, 4.
- Huang, W. & Ferris, J. P. 2003. Synthesis of 35-40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem Commun (Camb)*, 1458-9.
- Huang, W. & Ferris, J. P. 2006. One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis. *J Am Chem Soc*, 128, 8914-9.
- Hud, N. V., Cafferty, B. J., Krishnamurthy, R. & Williams, L. D. 2013. The origin of RNA and "my grandfather's axe". *Chem Biol*, 20, 466-74.
- Ikawa, Y., Shiraishi, H. & Inoue, T. 1999. Trans-activation of the Tetrahymena group I intron ribozyme via a non-native RNA-RNA interaction. *Nucleic Acids Res*, 27, 1650-5.
- Inoue, T., Joyce, G. F., Grzeskowiak, K., Orgel, L. E., Brown, J. M. & Reese, C. B. 1984. Template-directed synthesis on the pentanucleotide CpCpGpCpC. *J Mol Biol*, 178, 669-76.
- Inoue, T. & Orgel, L. E. 1981. Substituent control of the poly (C)-directed oligomerization of guanosine 5'-phosphorimidazolide. *Journal of the American Chemical Society*, 103, 7666-7667.
- Inoue, T. & Orgel, L. E. 1982. Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(C). An RNA polymerase model. *J Mol Biol*, 162, 201-17.
- Inoue, T. & Orgel, L. E. 1983. A nonenzymatic RNA polymerase model. *Science*, 219, 859-62.
-

References

- Ivanov, S. A., Vauleon, S. & Muller, S. 2005. Efficient RNA ligation by reverse-joined hairpin ribozymes and engineering of twin ribozymes consisting of conventional and reverse-joined hairpin ribozyme units. *FEBS J*, 272, 4464-74.
- Jaeger, L., Wright, M. C. & Joyce, G. F. 1999. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc Natl Acad Sci U S A*, 96, 14712-7.
- James, K. D. & Ellington, A. D. 1999. The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Orig Life Evol Biosph*, 29, 375-90.
- Janas, T., Janas, T. & Yarus, M. 2004. A membrane transporter for tryptophan composed of RNA. *RNA*, 10, 1541-9.
- Jankowsky, E. 2011. RNA helicases at work: binding and rearranging. *Trends Biochem Sci*, 36, 19-29.
- Jenne, A. & Famulok, M. 1998. A novel ribozyme with ester transferase activity. *Chem Biol*, 5, 23-34.
- Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. 2001. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, 292, 1319-25.
- Joshi, P. C., Aldersley, M. F., Delano, J. W. & Ferris, J. P. 2009. Mechanism of montmorillonite catalysis in the formation of RNA oligomers. *J Am Chem Soc*, 131, 13369-74.
- Joyce, G., Deamer, D. & Fleischaker, G. 1994. Origins of life: the central concepts. Deamer, DW.
- Joyce, G. F. 2002a. The antiquity of RNA-based evolution. *Nature*, 418, 214-21.
- Joyce, G. F. 2002b. Molecular evolution: booting up life. *Nature*, 420, 278-9.
-

References

- Joyce, G. F. 2004. Directed evolution of nucleic acid enzymes. *Annu Rev Biochem*, 73, 791-836.
- Joyce, G. F. 2007. Forty years of in vitro evolution. *Angew Chem Int Ed Engl*, 46, 6420-36.
- Joyce, G. F. 2009. Evolution in an RNA world. *Cold Spring Harb Symp Quant Biol*, 74, 17-23.
- Joyce, G. F. & Orgel, L. E. 1999. 2 Prospects for Understanding the Origin of the RNA World. *Cold Spring Harbor Monograph Archive*, 37, 49-77.
- Kanavarioti, A., Monnard, P. A. & Deamer, D. W. 2001. Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology*, 1, 271-81.
- Kauffman, S. A. 1993. *The origins of order: Self-organization and selection in evolution*, Oxford university press.
- Kawamura, K. & Ferris, J. P. 1994. Kinetic and mechanistic analysis of dinucleotide and oligonucleotide formation from the 5'-phosphorimidazolide of adenosine on Na(+)-montmorillonite. *J Am Chem Soc*, 116, 7564-72.
- Khvorova, A., Kwak, Y. G., Tamkun, M., Majerfeld, I. & Yarus, M. 1999. RNAs that bind and change the permeability of phospholipid membranes. *Proc Natl Acad Sci U S A*, 96, 10649-54.
- Kim, D. E. & Joyce, G. F. 2004. Cross-catalytic replication of an RNA ligase ribozyme. *Chem Biol*, 11, 1505-12.
- Knight, R. & Yarus, M. 2003. Finding specific RNA motifs: function in a zeptomole world? *RNA*, 9, 218-30.
- Kofoed, J., Reymond, J. L. & Darbre, T. 2005. Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of alpha-hydroxy aldehydes and ketones. *Org Biomol Chem*, 3, 1850-5.
-

- Kolb, V. M., Dworkin, J. P. & Miller, S. L. 1994. Alternative bases in the RNA world: the prebiotic synthesis of urazole and its ribosides. *J Mol Evol*, 38, 549-57.
- Komatsu, Y., Nobuoka, K., Karino-Abe, N., Matsuda, A. & Ohtsuka, E. 2002. In vitro selection of hairpin ribozymes activated with short oligonucleotides. *Biochemistry*, 41, 9090-8.
- Korzeniewski, B. 2001. Cybernetic formulation of the definition of life. *J Theor Biol*, 209, 275-86.
- Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31, 147-57.
- Kumar, R. M. & Joyce, G. F. 2003. A modular, bifunctional RNA that integrates itself into a target RNA. *Proc Natl Acad Sci U S A*, 100, 9738-43.
- Kun, A., Szilagyi, A., Konnyu, B., Boza, G., Zachar, I. & Szathmary, E. 2015. The dynamics of the RNA world: insights and challenges. *Ann N Y Acad Sci*, 1341, 75-95.
- Lai, L. B., Chan, P. P., Cozen, A. E., Bernick, D. L., Brown, J. W., Gopalan, V. & Lowe, T. M. 2010. Discovery of a minimal form of RNase P in *Pyrobaculum*. *Proc Natl Acad Sci U S A*, 107, 22493-8.
- Lambowitz, A. M. & Zimmerly, S. 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor perspectives in biology*, 3, a003616.
- Landweber, L. F. & Pokrovskaya, I. D. 1999. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *Proc Natl Acad Sci U S A*, 96, 173-8.
- Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. 2000. Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol*, 7, 28-33.
-

References

- Levy, M. & Ellington, A. D. 2002. In vitro selection of a deoxyribozyme that can utilize multiple substrates. *J Mol Evol*, 54, 180-90.
- Lewin, R. 1986. RNA Catalysis Gives Fresh Perspective on the Origin of Life: The old chicken-and-egg problem of the origin of life is illuminated in unexpected ways by recent results on the splicing of RNA precursors. *Science*, 231, 545-6.
- Lincoln, T. A. & Joyce, G. F. 2009. Self-sustained replication of an RNA enzyme. *Science*, 323, 1229-32.
- Loffler, P. M., Groen, J., Dorr, M. & Monnard, P. A. 2013. Sliding over the blocks in enzyme-free RNA copying--one-pot primer extension in ice. *PLoS One*, 8, e75617.
- Lohrmann, R. & Orgel, L. E. 1971. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science*, 171, 490-4.
- Lohse, P. A. & Szostak, J. W. 1996. Ribozyme-catalysed amino-acid transfer reactions. *Nature*, 381, 442-4.
- Lorsch, J. R. & Szostak, J. W. 1994. In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry*, 33, 973-82.
- Lozupone, C., Changayil, S., Majerfeld, I. & Yarus, M. 2003. Selection of the simplest RNA that binds isoleucine. *RNA*, 9, 1315-22.
- Luisi, P. L. 1998a. About various definitions of life. *Orig Life Evol Biosph*, 28, 613-22.
- Luisi, P. L. 1998b. About various definitions of life. *Origins of Life and Evolution of the Biosphere*, 28, 613-622.
- Mami, I. & Pallet, N. 2015. Transfer RNA fragmentation and protein translation dynamics in the course of kidney injury. *RNA Biol*, 0.
- Manrubia, S. C. & Briones, C. 2007. Modular evolution and increase of functional complexity in replicating RNA molecules. *RNA*, 13, 97-107.
- Masel, J. & Trotter, M. V. 2010. Robustness and evolvability. *Trends Genet*, 26, 406-14.
-

- Mcginness, K. E., Wright, M. C. & Joyce, G. F. 2002. Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem Biol*, 9, 585-96.
- Meinert, C., Myrgorodska, I., De Marcellus, P., Buhse, T., Nahon, L., Hoffmann, S. V., D'hendecourt Lle, S. & Meierhenrich, U. J. 2016. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science*, 352, 208-12.
- Meyer, A. J., Ellefson, J. W. & Ellington, A. D. 2012. Abiotic self-replication. *Acc Chem Res*, 45, 2097-105.
- Michel, F. & Westhof, E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol*, 216, 585-610.
- Michod, R. E. 1983. Population biology of the first replicators: on the origin of the genotype, phenotype and organism. *American Zoologist*, 23, 5-14.
- Miller, S. L. 1997. Peptide nucleic acids and prebiotic chemistry. *Nat Struct Biol*, 4, 167-9.
- Miyakawa, S., Cleaves, H. J. & Miller, S. L. 2002a. The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig Life Evol Biosph*, 32, 195-208.
- Miyakawa, S., Cleaves, H. J. & Miller, S. L. 2002b. The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions. *Orig Life Evol Biosph*, 32, 209-18.
- Miyakawa, S. & Ferris, J. P. 2003. Sequence- and regioselectivity in the montmorillonite-catalyzed synthesis of RNA. *J Am Chem Soc*, 125, 8202-8.
- Miyakawa, S., Murasawa, K., Kobayashi, K. & Sawaoka, A. B. 2000. Abiotic synthesis of guanine with high-temperature plasma. *Orig Life Evol Biosph*, 30, 557-66.

- Monnard, P. A., Kanavarioti, A. & Deamer, D. W. 2003. Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J Am Chem Soc*, 125, 13734-40.
- Monnard, P. A. & Szostak, J. W. 2008. Metal-ion catalyzed polymerization in the eutectic phase in water-ice: a possible approach to template-directed RNA polymerization. *J Inorg Biochem*, 102, 1104-11.
- Morasch, M., Mast, C. B., Langer, J. K., Schilcher, P. & Braun, D. 2014. Dry polymerization of 3',5'-cyclic GMP to long strands of RNA. *ChemBiochem*, 15, 879-83.
- Mueller, D., Pitsch, S., Kittaka, A., Wagner, E., Wintner, C. E. & Eschenmoser, A. 1990. Chemistry of α -aminonitriles. Aldomerization of glycolaldehyde phosphate to rac-hexose 2, 4, 6-triphosphates and (in presence of formaldehyde) rac-pentose 2, 4-diphosphates: rac-allose 2, 4, 6-triphosphate and rac-ribose 2, 4-diphosphate are the main reaction products. *Helvetica Chimica Acta*, 73, 1410-68.
- Muller, U. F. & Bartel, D. P. 2008. Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA*, 14, 552-62.
- Mustoe, A. M., Al-Hashimi, H. M. & Brooks, C. L., 3rd 2015a. Secondary structure encodes a cooperative tertiary folding funnel in the *Azoarcus* ribozyme. *Nucleic Acids Res*.
- Mustoe, A. M., Brooks, C. L., 3rd & Al-Hashimi, H. M. 2014. Topological constraints are major determinants of tRNA tertiary structure and dynamics and provide basis for tertiary folding cooperativity. *Nucleic Acids Res*, 42, 11792-804.
- Mustoe, A. M., Liu, X., Lin, P. J., Al-Hashimi, H. M., Fierke, C. A. & Brooks, C. L., 3rd 2015b. Noncanonical secondary structure stabilizes mitochondrial
-

- tRNA(Ser(UCN)) by reducing the entropic cost of tertiary folding. *J Am Chem Soc*, 137, 3592-9.
- Nedelcu, A. M., Driscoll, W. W., Durand, P. M., Herron, M. D. & Rashidi, A. 2011. On the paradigm of altruistic suicide in the unicellular world. *Evolution*, 65, 3-20.
- Nelson, K. E., Levy, M. & Miller, S. L. 2000. Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc Natl Acad Sci U S A*, 97, 3868-71.
- Neveu, M., Kim, H. J. & Benner, S. A. 2013. The "strong" RNA world hypothesis: fifty years old. *Astrobiology*, 13, 391-403.
- Nowak, M. A. 1992. What is a quasispecies? *Trends Ecol Evol*, 7, 118-21.
- Orgel, L. E. 1968. Evolution of the genetic apparatus. *J Mol Biol*, 38, 381-93.
- Orgel, L. E. 2002. Is cyanoacetylene prebiotic? *Orig Life Evol Biosph*, 32, 279-81.
- Orgel, L. E. 2004a. Prebiotic adenine revisited: eutectics and photochemistry. *Orig Life Evol Biosph*, 34, 361-9.
- Orgel, L. E. 2004b. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol*, 39, 99-123.
- Oró, J. 1960. Synthesis of adenine from ammonium cyanide. *Biochemical and Biophysical Research Communications*, 2, 407-412.
- Oró, J. 1961. Comets and the formation of biochemical compounds on the primitive Earth.
- Oró, J. & Kimball, A. 1961. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Archives of biochemistry and biophysics*, 94, 217-227.
- Oró, J. & Kimball, A. 1962. Synthesis of purines under possible primitive earth conditions: II. Purine intermediates from hydrogen cyanide. *Archives of biochemistry and biophysics*, 96, 293-313.
- Orr, H. A. 2009. Fitness and its role in evolutionary genetics. *Nat Rev Genet*, 10, 531-9.
-

References

- Österberg, R., Orgel, L. & Lohrmann, R. 1973. Further studies of urea-catalyzed phosphorylation reactions. *Journal of molecular evolution*, 2, 231-234.
- Pace, N. R. & Marsh, T. L. 1985. RNA catalysis and the origin of life. *Orig Life Evol Biosph*, 16, 97-116.
- Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. 2013. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc Natl Acad Sci U S A*, 110, 10089-94.
- Paul, N. & Joyce, G. F. 2002. A self-replicating ligase ribozyme. *Proc Natl Acad Sci U S A*, 99, 12733-40.
- Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. & Cheng, H. L. 1986. A self-splicing RNA excises an intron lariat. *Cell*, 44, 213-23.
- Penedo, J. C., Wilson, T. J., Jayasena, S. D., Khvorova, A. & Lilley, D. M. 2004. Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA*, 10, 880-8.
- Penny, D. 2005. An interpretive review of the origin of life research. *Biology and Philosophy*, 20, 633-671.
- Powner, M. W., Gerland, B. & Sutherland, J. D. 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459, 239-42.
- Powner, M. W., Sutherland, J. D. & Szostak, J. W. 2011. Chemoselective multicomponent one-pot assembly of purine precursors in water. *Journal of the American Chemical Society*, 133, 4149-4150.
- Prabakar, K. J. & Ferris, J. P. 1997. Adenine derivatives as phosphate-activating groups for the regioselective formation of 3',5'-linked oligoadenylates on montmorillonite:

- possible phosphate-activating groups for the prebiotic synthesis of RNA. *J Am Chem Soc*, 119, 4330-7.
- Prakash, T. P., Roberts, C. & Switzer, C. 1997. Activity of 2', 5'-Linked RNA in the Template-Directed Oligomerization of Mononucleotides. *Angewandte Chemie International Edition in English*, 36, 1522-1523.
- Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. 1986. Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231, 1577-80.
- Puerta-Fernández, E., Romero-López, C., Barroso-Deljesus, A. & Berzal-Herranz, A. 2003. Ribozymes: recent advances in the development of RNA tools. *FEMS microbiology reviews*, 27, 75-97.
- Reimann, R. & Zubay, G. 1999. Nucleoside phosphorylation: a feasible step in the prebiotic pathway to RNA. *Orig Life Evol Biosph*, 29, 229-47.
- Robertson, M. P., Hesselberth, J. R. & Ellington, A. D. 2001. Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA*, 7, 513-23.
- Robertson, M. P. & Joyce, G. F. 2012. The origins of the RNA world. *Cold Spring Harb Perspect Biol*, 4.
- Robertson, M. P. & Miller, S. L. 1995a. An efficient prebiotic synthesis of cytosine and uracil. *Nature*, 375, 772-4.
- Robertson, M. P. & Miller, S. L. 1995b. Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. *Science*, 268, 702-5.
- Robertson, M. P. & Scott, W. G. 2007. The structural basis of ribozyme-catalyzed RNA assembly. *Science*, 315, 1549-53.
- Roff, D. A. 2002. *Life history evolution*, Sinauer Associates Sunderland.
-

- Roff, D. A. 2007. Contributions of genomics to life-history theory. *Nat Rev Genet*, 8, 116-25.
- Rohatgi, R., Bartel, D. P. & Szostak, J. W. 1996a. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J Am Chem Soc*, 118, 3332-9.
- Rohatgi, R., Bartel, D. P. & Szostak, J. W. 1996b. Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3'-5' phosphodiester bonds. *J Am Chem Soc*, 118, 3340-4.
- Sabeti, P. C., Unrau, P. J. & Bartel, D. P. 1997. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem Biol*, 4, 767-74.
- Saffhill, R. 1970. Selective phosphorylation of the cis-2',3'-diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution. *J Org Chem*, 35, 2881-3.
- Saville, B. J. & Collins, R. A. 1990. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell*, 61, 685-96.
- Schoning, K., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. & Eschenmoser, A. 2000. Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3'-->2') oligonucleotide system. *Science*, 290, 1347-51.
- Schwartz, A. 1969. Specific phosphorylation of the 2'-and 3'-positions in ribonucleosides. *Journal of the Chemical Society D: Chemical Communications*, 1393a-1393a.
- Scott, W. G. 2007. Morphing the minimal and full-length hammerhead ribozymes: implications for the cleavage mechanism. *Biol Chem*, 388, 727-35.
- Szczepanski, J. T. & Joyce, G. F. 2014. A cross-chiral RNA polymerase ribozyme. *Nature*, 515, 440-2.
-

References

- Seelig, B. & Jaschke, A. 1999. A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol*, 6, 167-76.
- Sengle, G., Eisenfuhr, A., Arora, P. S., Nowick, J. S. & Famulok, M. 2001. Novel RNA catalysts for the Michael reaction. *Chem Biol*, 8, 459-73.
- Sharmeen, L., Kuo, M., Dinter-Gottlieb, G. & Taylor, J. 1988. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *Journal of Virology*, 62, 2674-2679.
- Sharp, P. A. 1985. On the origin of RNA splicing and introns. *Cell*, 42, 397-400.
- Shechner, D. M., Grant, R. A., Bagby, S. C., Koldobskaya, Y., Piccirilli, J. A. & Bartel, D. P. 2009. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science*, 326, 1271-5.
- Sievers, D. & Von Kiedrowski, G. 1994. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369, 221-4.
- Smith, J. I., Steel, M. & Hordijk, W. 2014. Autocatalytic sets in a partitioned biochemical network. *J Syst Chem*, 5, 2.
- Stearns, S. C. 1992. *The evolution of life histories*, Oxford University Press Oxford.
- Steitz, T. A. & Moore, P. B. 2003. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci*, 28, 411-8.
- Stich, M., Briones, C. & Manrubia, S. C. 2008. On the structural repertoire of pools of short, random RNA sequences. *J Theor Biol*, 252, 750-63.
- Sulston, J., Lohrmann, R., Orgel, L. E. & Miles, H. T. 1968. Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc Natl Acad Sci U S A*, 59, 726-33.
- Svoboda, P. & Di Cara, A. 2006. Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci*, 63, 901-8.
-

- Szostak, J. W. 2011. An optimal degree of physical and chemical heterogeneity for the origin of life? *Philos Trans R Soc Lond B Biol Sci*, 366, 2894-901.
- Szostak, J. W., Bartel, D. P. & Luisi, P. L. 2001. Synthesizing life. *Nature*, 409, 387-90.
- Takeuchi, N. & Hogeweg, P. 2009. Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization. *PLoS Comput Biol*, 5, e1000542.
- Takeuchi, N. & Hogeweg, P. 2012. Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. *Phys Life Rev*, 9, 219-63.
- Tang, J. & Breaker, R. R. 1997. Rational design of allosteric ribozymes. *Chem Biol*, 4, 453-9.
- Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. 1997. RNA-catalysed carbon-carbon bond formation. *Nature*, 389, 54-7.
- Tsukiji, S., Pattnaik, S. B. & Suga, H. 2004. Reduction of an aldehyde by a NADH/Zn²⁺ - dependent redox active ribozyme. *J Am Chem Soc*, 126, 5044-5.
- Tuerk, C. & Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249, 505-10.
- Unrau, P. J. & Bartel, D. P. 1998. RNA-catalysed nucleotide synthesis. *Nature*, 395, 260-3.
- Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J. & Lehman, N. 2012. Spontaneous network formation among cooperative RNA replicators. *Nature*, 491, 72-7.
- Vaidya, N., Walker, S. I. & Lehman, N. 2013. Recycling of informational units leads to selection of replicators in a prebiotic soup. *Chem Biol*, 20, 241-52.
- Visser, C. M. & Kellogg, R. M. 1978. Biotin. Its place in evolution. *J Mol Evol*, 11, 171-87.
-

References

- Vlassov, A. V., Johnston, B. H., Landweber, L. F. & Kazakov, S. A. 2004. Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world. *Nucleic Acids Res*, 32, 2966-74.
- Wagner, A. 2011. Genotype networks shed light on evolutionary constraints. *Trends in Ecology & Evolution*, 26, 577-584.
- Wagner, A. & Stadler, P. F. 1999. Viral RNA and evolved mutational robustness. *J Exp Zool*, 285, 119-27.
- Wagner, N., Pross, A. & Tannenbaum, E. 2010. Selection advantage of metabolic over non-metabolic replicators: a kinetic analysis. *Biosystems*, 99, 126-9.
- Wang, K. J. & Ferris, J. P. 2001. Effect of inhibitors on the montmorillonite clay-catalyzed formation of RNA: studies on the reaction pathway. *Orig Life Evol Biosph*, 31, 381-402.
- Wang, Q. S. & Unrau, P. J. 2005. Ribozyme motif structure mapped using random recombination and selection. *RNA*, 11, 404-11.
- Wecker, M., Smith, D. & Gold, L. 1996. In vitro selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA*, 2, 982-94.
- West, S. A., Griffin, A. S. & Gardner, A. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J Evol Biol*, 20, 415-32.
- Westhof, E. & Massire, C. 2004. Structural biology. Evolution of RNA architecture. *Science*, 306, 62-3.
- White, H. 1982. Biosynthetic and salvage pathways of pyridine nucleotide coenzymes. *The pyridine nucleotide coenzymes*, 225-248.
- White, H. B., 3rd 1976. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol*, 7, 101-4.
-

References

- Wiegand, T. W., Janssen, R. C. & Eaton, B. E. 1997. Selection of RNA amide synthases. *Chem Biol*, 4, 675-83.
- Wilson, C. & Szostak, J. W. 1995. In vitro evolution of a self-alkylating ribozyme. *Nature*, 374, 777-82.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T. & Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature*, 407, 327-39.
- Wochner, A., Attwater, J., Coulson, A. & Holliger, P. 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science*, 332, 209-12.
- Wright, M. C. & Joyce, G. F. 1997. Continuous in vitro evolution of catalytic function. *Science*, 276, 614-7.
- Wu, T. & Orgel, L. E. 1992a. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J Am Chem Soc*, 114, 5496-501.
- Wu, T. & Orgel, L. E. 1992b. Nonenzymatic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J Am Chem Soc*, 114, 317-22.
- Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. 2007a. DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell*, 28, 253-63.
- Yang, Y. W., Zhang, S., Mccullum, E. O. & Chaput, J. C. 2007b. Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA. *J Mol Evol*, 65, 289-95.
- Yarus, M. 1991. An RNA-amino acid complex and the origin of the genetic code. *New Biol*, 3, 183-9.
-

References

- Yarus, M., Widmann, J. J. & Knight, R. 2009. RNA-amino acid binding: a stereochemical era for the genetic code. *J Mol Evol*, 69, 406-29.
- Yu, H., Zhang, S. & Chaput, J. C. 2012. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat Chem*, 4, 183-7.
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science*, 292, 883-96.
- Zaher, H. S. & Unrau, P. J. 2007. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA*, 13, 1017-26.
- Zhang, B. & Cech, T. R. 1997. Peptide bond formation by in vitro selected ribozymes. *Nature*, 390, 96-100.
- Zhang, L., Peritz, A. & Meggers, E. 2005. A simple glycol nucleic acid. *J Am Chem Soc*, 127, 4174-5.
- Zielinski, W. S. & Orgel, L. E. 1987. Autocatalytic synthesis of a tetranucleotide analogue. *Nature*, 327, 346-7.
- Zimmerly, S. & Semper, C. 2015. Evolution of group II introns. *Mob DNA*, 6, 7.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31, 3406-15.
- Zuker, M. & Stiegler, P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9, 133-48.