Research Proposal for Ph.D. Advisory Committee

Enrolment date (by conversion of MSc) 1 August 2005, expected completion July 2009

Simulation Models of Prebiotic Evolution of Genetic Coding

Sidney Markowitz
sidney@sidney.com

Supervisors: Alexei Drummond (CompSci) and Peter R. Wills (Physics)

Adviser: Allen Rodrigo (SBS)

Department of Computer Science

Abstract

Common to all life on Earth are the mechanisms of genetic encoding, in which specific trinucleotide sequences in DNA and RNA map to specific amino acids in synthesized proteins. The proteins which catalyse the coding are themselves products of that synthesis, posing the question of how the system bootstrapped itself and how it maintains stability. Mechanistic models based on chemical associations between groups of nucleotides and amino acids, or stochastic theories of a random “frozen accident” have not fully explained the emergence of the genetic code. The proposed thesis project considers the genetic code and protein synthesis as a molecular information processing system with mutually catalytic information storage and functional components, using simulation to investigate feasible models of the emergence of genetic encoding from an initially random population of proteins. The focus of the research is on developing an abstract framework that does not depend on explicit molecular details while demonstrating a plausible mechanism of self-organisation. It is hoped that an understanding of the conditions under which such a system self-organises will make it possible to build a complete model of the stages of emergence of the genetic code from the prebiotic environment.

Introduction

All life on Earth has in common the use of proteins made of long sequences of amino acids for structural and chemical purposes, and the use of replicating genes of DNA made of sequences of nucleotides as templates for the synthesis of the proteins. All life, with occasional minor variations, uses the same four nucleotides to make DNA, the same four to make RNA, the same set of 20 amino acids, and the same unambiguous mapping from trinucleotide sequences (codons) in DNA and RNA to amino acids. The pattern in DNA is replicated by messenger RNA (mRNA), whose

codons are then attached to molecules of transfer RNA (tRNA) that have been charged with the correct amino acids by special protein catalysts called aminoacyl-tRNA-synthetases. This universal mechanism must have had its origins in the very earliest beginnings of biological existence.

While the mechanism of protein synthesis from DNA templates is well understood, its complexity, the dependence of protein synthesis on protein catalysts that are themselves products of the synthesis, and the near universality of the standard Genetic Code collectively pose the question of how it bootstrapped itself into existence and how it maintains its stability in the face of errors and mutations.

Taking a molecular biological perspective in attempting to explain the origin of the genetic code, researchers have hypothesised stereochemical associations between nucleotide sequences and amino acids (i.e., affinities based on molecular structures that fit together), and coevolutionary paths that resulted from the biosynthesis of certain amino acids from certain earlier ones. Some researchers claim that Crick's Frozen Accident theory (Crick, 1968) was the dominant hypothesis for the two decades after it was proposed (Knight et al., 2001). The following Background section describes these various theories in more detail.

These theories have not been sufficient to fully describe the development of the genetic code starting from a prebiotic environment in which there was no coded assignment.

My proposed research looks at the genetic code and protein synthesis as a molecular information processing system, following the work of Bedian (1982) and Wills (1993). I am developing a simulation that models a replicating information storage, analogous to genes, that patterns the synthesis of functional components which in turn catalyse the replication and synthesis processes. The purpose is to investigate what is required for such a model to self-organise into a stable autocatalytic coding system.

Ultimately, I hope to have models of the emergence of the genetic code that come closer to completeness by combining the constraints I find are imposed by self-organisation with the latest results from the stereochemical and coevolution research.

Background

Nothing was known about the genetic code when Watson and Crick determined the molecular structure of DNA in 1953 (Watson and Crick, 1953). They showed that a DNA molecule is a double helix formed by two strands of four types of nucleotides, like two strings of four colours of beads twisted together. The exact sequence of nucleotides forms a code that specifies the order of 20 types of amino acids that string together to form a protein molecule. It was at first supposed that there is a direct chemical connection between sequences of nucleotide bases in DNA and the amino acids that they code for. Gamow (1954) proposed the Diamond Code, in which a portion of an amino acid fits into a space made by a group of four nucleotides, like a lock and key.

When it was discovered that in eukaryotes (cells with nuclei that encapsulate the DNA) protein is synthesized outside of the nucleus (Hoagland et al., 1958), such direct theories became untenable. Nirenberg and Matthaei (1961) discovered the role of messenger RNA (mRNA) as the transport of information from DNA, demonstrating that a sequence of three U nucleotides forms a codon that maps to the amino acid Phenylalanine. Between 1961 and 1967 they found the remaining codon to

amino acid assignments, and it was shown that the same coding assignments are used by all organisms (Marshall et al., 1967).

By the time Crick wrote his influential review of the state of the research (Crick, 1968), the genetic code had been revealed to be an elegantly efficient coding scheme, remarkably good at minimising the effects of mutational and translational errors, and showing patterns that hinted at underlying chemical causes (Table 1). At the same time, the mechanism of protein synthesis proved to be so indirect and complex that researchers were left with no compelling theory as to how it could have emerged. One theory that Crick proposed in that paper said that the genetic code is a “frozen accident”, that once formed could not be changed because any mutation that changed the code would have drastic and fatal effects on an organism. While it has been claimed that the Frozen Accident theory was so influential that for two decades it inhibited research into the origins of the genetic code (Knight, 2001), a careful reading shows that Crick (1968) only used it to explain the near universality of the code, and did not consider Frozen Accident to have sufficient explanatory or predictive power to be a good theory of the emergence of the genetic code.

The various theories that have been proposed to explain the emergence of the genetic code fall into several classes, three of them described in similar but not identical ways in the review papers of Crick (1968), Knight et al., (1999), and Di Giulio (2005).

Adaptive theories say that the genetic code evolved via selection for an optimisation, for example, selecting for a code that reduces the impact of single errors by having amino acids with similar physicochemical properties have similar codons. These theories are sometimes called “physicochemical”, but their focus is more on code optimisation than on chemistry. The adaptive theories are supported by demonstrations that the standard genetic code is more optimal by some measure than randomly generated codes (Ardell, 1998). A difficulty with such theories is that they require selection between competing codes. Where there are no alternate codes, there could not be adaptation (Freeland et al., 2003; Osawa et al., 1992). Ardell (2001; 2002) used a population genetics model simulation to support theories of adaptation. His model is unusual in the work on adaptive theories in that it begins with random codon to amino acid assignments and does not assume chemical affinities.

Stereochemical theories say that amino acids and their codons or anticodons have a stereochemical affinity. In other words, a codon (or anticodon) presents a molecular structure in which a portion of the corresponding amino acid can fit like a lock and key. Such affinities, if they exist would provide a simple explanation for the code. But as Crick (1968) pointed out, such affinities don’t seem to exist now. Ongoing work such as by Knight et al., (1999) is testing the theory by using a technique called SELEX (Tuerk and Gold, 1990), an in vitro evolution method that isolates and amplifies RNA with selected characteristics, to look for RNA aptamers (short nucleotide sequences) that bind to amino acids, then determine if the aptamers are related to the codons that map to those amino acids.

The third class of theories are called “historical” (Knight et al., 1999) and “coevolutionary” (Di Giulio, 2005). These theories propose that an early code used fewer codons and amino acids, then underwent code expansion to evolve the current code. The stepwise evolution while maintaining viability of organisms could have resulted in amino acids with similar codons being related to each other, either by being physicochemically similar or as precursor/product in a biosynthetic pathway (Wong, 1975; Wong, 2005). Thus, coevolution could explain aspects of the

Table 1: The "Universal" Genetic Code

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	TER	UGA	TER
	UUG	Leu	UCG	Ser	UAG	TER	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Key

Saturation reflects molecular volume (Grantham 1974). Colourful = bigger.
Brightness reflects polar requirement (Woese et al. 1966). Lighter = hydrophobic.
Hue reflects side-chain composition. 0° (red) = acid; 30° (orange) = amide;
60° (yellow) = sulphur; 120° (green) = alcohol; 180° (cyan) = aromatic;
240° (blue) = basic; 270° (purple) = hydrophobic.

Table 1: **The Standard Genetic Code.** The genetic code and its patterns, as illustrated in Knight (2001)

genetic code that have been cited as supporting stereochemical and adaptive theories. Di Giulio (1998) extended the coevolution theory to say that anticodons and biosynthetic pathways of amino acids evolved together.

Both Knight et al., (1999) and Di Giulio (2005) find ways to make the three classes of theories compatible. However, Di Giulio somewhat cryptically concludes, “Finally, although this model (Di Giulio 1998) makes compatible the theories of the origin of the genetic code, nevertheless I believe that the coevolution theory and the stereochemical theory are incompatible since these are based on a different determinism: historical the former, physicochemical the later.” (Di Giulio, 2005)

Knight et al., (1999) proposes that stereochemistry was the major force behind the origination of the early genetic code, then came code expansion under control of coevolutionary forces. Once there were sufficient numbers of codons and amino acids in the code and assignments were mediated by adaptor molecules, code adaptation for optimisation became dominant. He suggests that the evolutionary forces could be complementary, giving an example that current codon assignments could assign biosynthetically related amino acids to similar codons, which would meet both stereochemical and adaptive criteria (Caporaso et al., 2005).

None of the three theories have been developed enough to completely describe the emergence of the genetic code, even when combined in a staged form as in Fig. 1 (from Knight et al., (1999) Figure 5). Open questions include: What was the prebiotic environment in which the genetic code developed? Did the current genetic code develop early or late in the development of macromolecular metabolism? What was the origin of metabolism, an understanding of which would provide contexts for theories of origins of the genetic code? Which came first, RNA metabolism adding proteins (the so-called RNA world of Gilbert (1986)), or a protein metabolism adding RNA? In support of stereochemical theories, how many aptamers are yet to be found linking various anticodons to amino acids?

SELEX is currently being used to test hypotheses related to coevolutionary as well as stereochemical theories of the origin of the genetic code. (Vergne et al., 2006; Yarus et al., 2005)

There also continues to be speculation regarding the nature of the specific chemical nature of the prebiotic environment and possible mechanisms for early metabolism, as referenced in the introduction to Vergne et al., (2006). (Di Giulio, 2003; Lazcano and Miller, 1996; Szathmary, 1997)

My research area

My proposed research is part of a fourth theoretical approach to explaining the emergence of the genetic code, which I call “self-organisation”. It began with Eigen’s (1971) investigations into the fundamental requirements that allow a molecular system to self-organise. He concluded that self-organisation requires a catalytic coupling with cyclical feedback between replicating information carriers and the production of functional catalysts. Nucleic acids and proteins provide the proper prerequisites. In contrast to the research described in the previous sections, Eigen did not attempt to describe details of the emergence of the genetic code, saying, “We should not pretend to explain the historical path of evolution. All we can try to do is to state the minimum prerequisites and obtain some insight into/the physical principles of the evolutionary process.” He began with the question, “How can such a system, represented by an ensemble of nucleic acids and

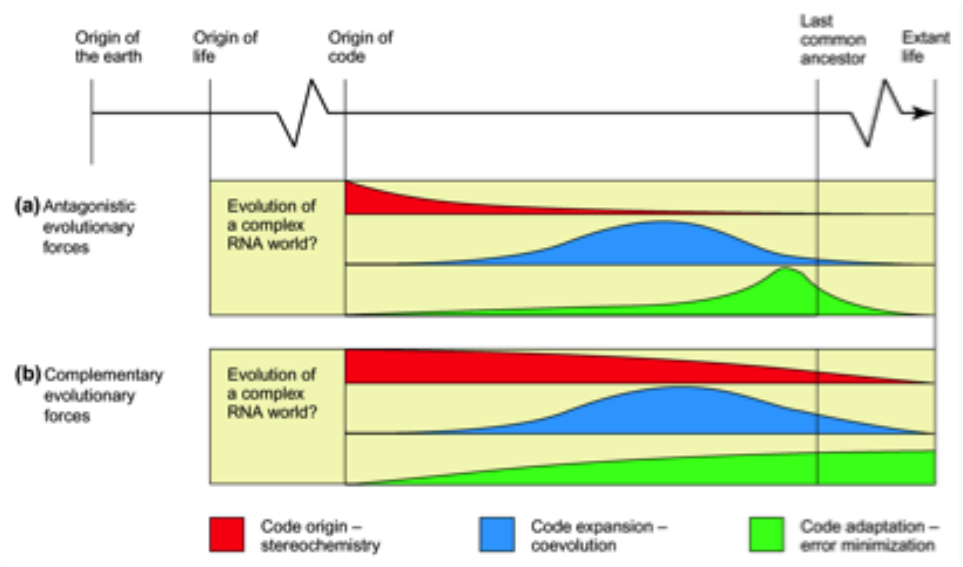

Figure 1: **Three facets of code evolution.** Taken from (Knight et al., 1999) figure 5, this illustrates Knight's concept of how the three classes of theories of the evolution of the genetic code can be combined as stages in the overall evolution of the code

proteins, organize itself into a stable selfreproducing, further evolving unit?" One of his results was to elucidate the notion of an error threshold in replication as a solution to the error catastrophe problem raised by Orgel (1963; 1970).

(Hoffmann, 1974) applied a similar threshold to the process of translation of codons to amino acids, showing that translation does not have to proceed perfectly. Bedian (1982; 2001) demonstrated that such a system can be tolerant of coding ambiguity, defined as the existence of codons that code for multiple amino acids, and derived detailed equations describing systems with binary codes. Wills (1993) and Nieselt-Struwe and Wills (1997)) developed a model of protein translation which allows the derivation of formal constraints on genetic coding systems. Wills (1994) showed that in a chemically homogeneous self-organised coding system mutation during gene replication inevitably leads to collapse.

The initial conditions in these models (Bedian, 1982) (Wills, 1993) (Nieselt-Struwe and Wills, 1997) are similar to the early environment proposed by Woese as in his TE model of evolution of the genetic code (Woese, 1965), in which there is a primitive cell with random, ambiguous codon assignments and very error-prone translation. Woese made qualitative arguments regarding the nature of evolution of a genetic code starting from such a state. Bedian, Wills, and Nieselt-Struwe looked for formal constraints and equations that describe how self-organisation can take place in such a model.

Füchslin and McCaskill (2001) added an error-prone replication process to a simplified version of the error-prone translation process model to show that reaction-diffusion coupling can allow the two processes to self-organize into a stable system. They described what they call a Gene Replicase Translatase (GRT) system that includes gene replication with mutation catalysed by a "replicase" protein, protein synthesis with coding catalysed by a "translatase" and two alternate

translatase proteins that implement non-target codings.

My work (Markowitz et al., 2006) has extended research in this area by modelling a GRT system that is more complex and more realistic than that of Füchslin and McCaskill (2001). Separate translatases for each codon-to-amino acid assignment are embedded as catalytic centres in the protein sequence. Our reaction employs a separate translatase for each amino acid (like the real process) whereas Füchslin and McCaskill employed a single translatase for the entire synthesis of a new protein. Their model was embedded on a three dimensional lattice with one molecule per node and diffusion between nodes. The model I use also demonstrates stabilization through reaction-diffusion coupling, but uses a one-dimensional lattice of "well-mixed" compartments containing variable numbers of molecules, with diffusion between compartments. In this, the results can be compared to research that has shown that a combination of chaotic flows and spatial constraints is required for evolutionary improvement of replication and translation efficiency (Scheuring et al., 2003).

Ongoing work, goals, and objectives

Eigen (1971) formulated an abstract and general model that provides insights that are useful when attempting to design more realistic models of self-organizing processes. The subsequent cited work added features that bring the model closer to natural genetic systems, while elucidating constraints on self-organization. My goal is to eventually model processes that can describe self-organization starting from a random initial state through simple autocatalytic systems all the way to the modern systems of genetic coding. To that effect, in my most recent work I have developed a model that is intermediate, in complexity and realism, with respect to the models of prior work and the systems of natural biology.

By finding the ranges of parameters in which self-organization is demonstrated, we can then further refine the model in the direction of added complexity and realism. My recent paper (2006) had the modest goal of demonstrating that there are values of parameters of the model in which self-organization occurs. My Ph.D. research will explore the range of parameter space and more precisely characterize the elements that are necessary for self-organization.

To date, no theories of the emergence of the genetic code have been complete in the sense of describing plausible pathways from a state of prebiotic disorganisation all the way to the extant biological system. Models of self-organisation may provide the missing pieces to the theories. In addition, the insight from studying biological models may in the long term help in the design artificial self-organising information processing systems, for applications such as artificial life, artificial intelligence, or nanoscale manufacturing.

Methods and approach

My recent paper (Markowitz et al., 2006) describes a discrete event step simulation of a GRT system in one dimension that demonstrates an apparent attractor state for some range of parameters. The software is a purpose-built C program that models genes and protein as 24-bit sequences contained in a closed one-dimensional loop of 1500 cells. Each cell contains a maximum of two five-gene "genomes" and 200 proteins. There are four types of codons and four

types of amino acids. Simulations have been carried out to on the order of 10^{10} event steps, which takes several days running on a 3.4 GHz dual Xeon processor with 1Gb RAM. While a networked cluster was used for the simulation, each computer ran its own copy of the simulation with different parameters. The current software does not contain any parallel or distributed processing capability.

I would like to continue work on the model in the following directions:

- Continue working with the discrete event step simulation, extending the dimensionality of the lattice to two and three dimensions.
- Explore the parameter space, looking for regions that lead to stability, and attempt to formulate general principles regarding stability in self-organising systems of this type.
- Investigate ways to simulate larger systems and for more event steps on clusters of computers, possibly by using optimistic asynchronous distributed discrete event simulation techniques (Lin and Fishwick, 1996).
- Derive systems of differential equations that characterise the model and apply numerical modelling methods to perform simulations, if possible using a software package such as XMDS (Collecutt and Drummond, 2001).
- Extend the simulation models to investigate models in which simple codes self-organise into increasing complexity, following the work of Wills (1994).

As an ongoing matter, while working on self-organisation theories of the emergence of the genetic code, I will stay aware of ongoing research on the stereochemical and coevolutionary theories, as well as theories of early emergence of metabolism, looking for opportunities to incorporate new results into better combined models.

Four year time line

First year:

- Model evolution of genetic code as a discrete event step simulation of a simplified (compared to biological) system that includes genes, replication, mutation, synthesis of protein catalysts, spatial heterogeneity, and reaction-diffusion coupling. Explore the parameter space that demonstrates stability. determine minimum values for sequence length, number of codons, number of amino acids, etc. for which the model works. Use a one-dimensional model, then extend it to two and three dimensions.
- Begin derivation of differential equations describing the model and development of simulations based on them using XMDS or similar software. Compare the results with those from the discrete event step simulations. Look for basins of attractors, regions of stability, and so on.
- Produce first year work products, including final research proposal, results of literature search, published conference paper, submitted journal article, introductory chapter of thesis, preliminary chapter of thesis based on the conference paper and journal article, and presentation with written report on state of my research.

Second year:

- Continue to work with differential equations to describe the model.
- Model self-organisation of more complex codes from simpler ones, following Wills (2004). Characterise the constraints on the parameters that allow self-organisation of increased complexity. Relate the results to coevolutionary theories.

Second to third year:

- Derive predictions from the model. Look for them in biological systems. Perhaps work with wetware labs to test predictions of the model using manufactured RNA or virus.

Second year to completion:

- Write chapters of thesis as work is completed, submitting material for publication when possible. Increasingly focus on writing leading up to finishing the dissertation by end of the fourth year.

Possibilities in various years, depending on results:

- See if a model of evolutionary development from a small number of codons and amino acids through increasing complexity to the extant biological coding system could be reflected in evidence for a phylogenetic tree of aminoacyl-tRNA-synthetase structure.
- Look for similar possibilities of connection with work being done in stereochemical and physiochemical aspects of the genetic code.

Technical challenges from a Computer Science point of view that might arise:

- How to distribute the simulations?
- Can the event step simulation be done using optimistic asynchronous distributed discrete event simulation? (Lin and Fishwick, 1996)
- Can the differential equations be specified sufficiently for numerical modelling, if so can I use XMDS (Collecutt and Drummond, 2001), and if so can the model be parallelised using the MPI facilities of XMDS on a Beowulf cluster (Sterling et al., 1995)?

Speculation that could lead to topics to pursue if there is time:

- Will the simulation models provide insight into the rapid evolution of RNA viruses in a host?
- What directions for further research into evolution of the genetic code will result from adding the self-organization approach?
- Will there be applications for the simulation methods?
- Will the models match any from physics? What insight might similar systems in physics give to biological systems?

References

Ardell, D. H.: 1998, On error minimization in a sequential origin of the standard genetic code, *Journal of Molecular Evolution* **47(1)**, 1

Ardell, D. H. and Sella, G.: 2001, On the evolution of redundancy in genetic codes, *Journal of Molecular Evolution* **53(4-5)**, 269

Ardell, D. H. and Sella, G.: 2002, No accident: genetic codes freeze in error-correcting patterns of the standard genetic code, *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **357(1427)**, 1625

Bedian, V.: 1982, The possible role of assignment catalysts in the origin of the genetic code, *Orig Life* **12(2)**, 181

Bedian, V.: 2001, Self-description and the origin of the genetic code, *Biosystems* **60(1-3)**, 39

Caporaso, J. G., Yarus, M., and Knight, R.: 2005, Error minimization and coding triplet/binding site associations are independent features of the canonical genetic code, *Journal of Molecular Evolution* **61(5)**, 597

Collecutt, G. and Drummond, P. D.: 2001, Xmds: extensible multi-dimensional simulator, *Computer Physics Communications* **142(1-3)**, 219

Crick, F. H. C.: 1968, Origin of genetic code, *Journal of Molecular Biology* **38(3)**, 367

Di Giulio, M.: 1998, Reflections on the origin of the genetic code: a hypothesis, *J Theor Biol* **191(2)**, 191

Di Giulio, M.: 2003, The early phases of genetic code origin: Conjectures on the evolution of coded catalysis, *Origins of Life and Evolution of Biospheres (Formerly Origins of Life and Evolution of the Biosphere)* **33(4 - 5)**, 479

Di Giulio, M.: 2005, The origin of the genetic code: theories and their relationships, a review, *Biosystems* **80(2)**, 175

Eigen, M.: 1971, Selforganization of matter and evolution of biological macromolecules, *Naturwissenschaften* **58(10)**, 465

Freeland, S. J., Wu, T., and Keulmann, N.: 2003, The case for an error minimizing standard genetic code, *Origins of Life and Evolution of the Biosphere* **33(4-5)**, 457

Füchslin, R. M. and McCaskill, J. S.: 2001, Evolutionary self-organization of cell-free genetic coding, *Proceedings of the National Academy of Sciences of the United States of America* **98(16)**, 9185

Gamow, G.: 1954, Possible relation between deoxyribonucleic acid and protein structures, *Nature* **173(4398)**, 318

Gilbert, W.: 1986, Origin of life - the rna world, *Nature* **319(6055)**, 618

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., and Zamecnik, P. C.: 1958, A soluble ribonucleic acid intermediate in protein synthesis, *J Biol Chem* **231(1)**, 241

Hoffmann, G. W.: 1974, On the origin of the genetic code and the stability of the translation apparatus, *J Mol Biol* **86(2)**, 349

Knight, R. D.: 2001, *The origin and evolution of the genetic code: statistical and experimental investigations, Ph.D. thesis*, Princeton University

Knight, R. D., Freeland, S. J., and Landweber, L. F.: 1999, Selection, history and chemistry: the three faces of the genetic code, *Trends in Biochemical Sciences* **24(6)**, 241

Knight, R. D., Landweber, L. F., and Yarus, M.: 2001, How mitochondria redefine the code, *Journal of Molecular Evolution* **53(4-5)**, 299

Lazcano, A. and Miller, S. L.: 1996, The origin and early evolution of life: Prebiotic chemistry, the pre-rna world, and time, *Cell* **85(6)**, 793

Lin, Y. B. and Fishwick, P. A.: 1996, Asynchronous parallel discrete event simulation, *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans* **26(4)**, 397

Markowitz, S., Drummond, A., Nieselt, K., and Wills, P. R.: 2006, Simulation model of prebiotic evolution of genetic coding, in L. M. Rocha, M. Bedau, D. Floreano, R. Goldstone, A. Vespignani, and L. Yaeger (eds.), *ALIFE X: 10th International Conference on the Simulation and Synthesis of Living Systems*, p. in press, Bloomington, Indiana, MIT Press

Marshall, R. E., Caskey, C. T., and Nirenberg, M.: 1967, Fine structure of rna codewords recognized by bacterial amphibian and mammalian transfer rna, *Science* **155(3764)**, 820

Nieselt-Struwe, K. and Wills, P. R.: 1997, The emergence of genetic coding in physical systems, *Journal of Theoretical Biology* **187(1)**, 1

Nirenberg, M. and Matthaei, J. H.: 1961, Dependence of cell-free protein synthesis in e coli upon naturally occurring or synthetic polyribonucleotides, *Proceedings of the National Academy of Sciences of the United States of America* **47(10)**, 1588

Orgel, L. E.: 1963, Maintenance of accuracy of protein synthesis and tis relevance to ageing, *Proceedings of the National Academy of Sciences of the United States of America* **49(4)**, 517

Orgel, L. E.: 1970, The maintenance of the accuracy of protein synthesis and its relevance to ageing: a correction, *Proc Natl Acad Sci U S A* **67(3)**, 1476

Osawa, S., Jukes, T. H., Watanabe, K., and Muto, A.: 1992, Recent-evidence for evolution of the genetic-code, *Microbiological Reviews* **56(1)**, 229

Scheuring, I., Czárán, T., Szabó, P., Károlyi, G., and Toroczkai, Z.: 2003, Spatial models of prebiotic evolution: Soup before pizza?, *Origins of Life and Evolution of Biospheres (Formerly Origins of Life and Evolution of the Biosphere)* **33(4 - 5)**, 319

Sterling, T. L., Savarese, D., Becker, D. J., Dorband, J. E., Ranawake, U. A., and Packer, C. V.: 1995, Beowulf: A parallel workstation for scientific computation, in P. Banerjee (ed.), *Proceedings of the 1995 International Conference on Parallel Processing*, Vol. Volume I: Architecture, pp 11–14, Urbana-Champain, Illinois, USA, CRC Press

Szathmary, E.: 1997, Origins of life - the first two billion years, *Nature* **387(6634)**, 662

Tuerk, C. and Gold, L.: 1990, Systematic evolution of ligands by exponential enrichment - rna ligands to bacteriophage-t4 dna-polymerase, *Science* **249(4968)**, 505

Vergne, J., Cognet, J. A., Szathmary, E., and Maurel, M. C.: 2006, In vitro selection of halo-thermophilic rna reveals two families of resistant rna, *Gene*

Watson, J. D. and Crick, F. H. C.: 1953, Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid, *Nature* **171(4356)**, 737

Wills, P. R.: 1993, Self-organization of genetic coding, *Journal of Theoretical Biology* **162(3)**, 267

Wills, P. R.: 1994, Does information acquire meaning naturally, *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics* **98(9)**, 1129

Wills, P. R.: 2004, Stepwise evolution of molecular biological coding, in J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. A. Watson (eds.), *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pp 51–56, Boston, MA, MIT Press

Woese, C. R.: 1965, On evolution of genetic code, *Proceedings of the National Academy of Sciences of the United States of America* **54(6)**, 1546

Wong, J. T. F.: 1975, Co-evolution theory of genetic code, *Proceedings of the National Academy of Sciences of the United States of America* **72(5)**, 1909

Wong, J. T. F.: 2005, Coevolution theory of the genetic code at age thirty, *Bioessays* **27(4)**, 416

Yarus, M., Caporaso, J. G., and Knight, R.: 2005, Origins of the genetic code: The escaped triplet theory, *Annual Review of Biochemistry* **74**, 179