Epilogue

The contents of *Mysterium Tremendum* (MT)were written as part of an experiment in scientific writing. Instead of using the normal route to publication, involving referees or reviewers and editors, I wanted to see what would happen if I were free from all such editorial constraints and impediments and instead made public my thoughts as they evolved, more or less in real time. The hope was that being able to readily express tentative ideas would free the mind to be more creative and allow ideas to emerge naturally. Long delays and the negative feedback often associated with the process of publication in journals or books would be avoided. The internet makes this possible. At its start, the goals for MT were more pessimistic than turned out to be warranted. Rapid progress was made and a series of ideas emerged that are at the core of Earth-bound biological mechanisms.

Human beings are said to contain about 60 trillion cells. 60,000,000,000,000 cells !!! There may be 10^{11} Sun-like stars in our Galaxy, and 10^{11} Galaxies altogether. Everywhere the atoms, or elements, are the same, based on the same nucleosynthesis mechanism. Alpha-multiples show local abundance peaks in the periodic table of elements [Part 1, Elements of life]. These special elements, along with a few others such as N and P, are the basis for the chemistry of life and of cells. The "atom" of biology is the cell. The mechanisms at work inside a cell are molecular. Since the physics and the chemistry of the cell appear to be universal, applicable everywhere in the Universe, the biology should also be universal. Life as we know it is probably how life is everywhere in the Universe. Therefore the detailed mechanisms are also universal. At the core of the contemporary mechanism, elucidated by hundreds of thousands of researchers over very many years, is the so-called MT [Part 1, Part 8]:

How did the tRNA's and aaRS's emerge from the basic underlying processes?
How were amino acid residues and their "locked in" codon assignments first made?

Initially, I wasn't very optimistic about the outcome of this experiment because I did not yet know the ideas that were to subsequently emerge. Ideas did emerge, even jumping out at certain times. A proposal for a complete conceptual model did evolve, after all. I believe the reader will find MT to be comprehensive and penetrating in its analysis. In it is a detailed proposal about how the genetic code, and its associated protein biosynthesis machinery, emerged and evolved. A two base ancestor code and a two base alphabet are invoked based on a physico-chemical connection that sets the code, i.e. associates the specific aa's with their cognate tRNA's [Part 10] (In this model, all codons are three bases long, whether for an alphabet based on two bases, a "two base alphabet" with three base codons, or for an alphabet based on four bases, a "four base alphabet" with three base codons.) This is the physical-chemical part of the code and the mechanism of polypeptide synthesis associated with it is called the *primitive RNA translator* [Part 5]. With it, the molecular processes give rise to a purely genetic extension that attends the

transition to a four base code. The rest of the code evolves by genetics. This is like a "frozen accident" to a degree, but really a "frozen instance" of *felicity* [Part 5, Part 11]. The two base code is a CG code [Part 9]. Why cysteine is the first amino acid added to the core system when the codon numbers increase as the code becomes based on four bases instead of two is explained. Why the nonpolar amino acids are so important during codon number increases so that adequate amounts of hydrophobic polypeptides are made and incorporated into the porous membranes of the ur-cell [Part 3] is also explained. How the evolution of codons leads to three disjoint sets of codons when "complementarity" is invoked [Part 8, Part 11]is an unexpected byproduct of this approach. The transition to a four base code also involves the evolution of ur-tRNA's and ur-aaRS's [Part 11].

Starting with [Part 5] the presentation is given in the form of a Platonic Dialogue between Uranya and Reynard. This vehicle for expressing the ideas was chosen because it allows a frankly skeptical reading of the text. Many of the ideas are indeed highly speculative and the purpose of MT is not so much an attempt to identify what actually happened during evolution on the early Earth but to show that a comprehensive and consistent set of ideas, a complete model, can be established by reasoning tempered by the laws of physics, chemistry and biology. Right away, Uranya and Reynard discuss the origin of *chirality* [Part 5] in a manner that makes manifest its importance mechanistically, and putatively explains its origin. This is just one of the examples in which an answer unexpectedly jumps out at one from the general context. The reader may indeed remain skeptical or try to take part in the dialogue.

Another essential topic is *compartmentalization* [Part 3]. It is argued that initially there were racemic proteinoid microspheres capable of replication. No polynucleotides are needed nor present at this stage of development. These ideas are based on many experiments involving proteinoids and the microspheres that self-assemble from them. Several misconceptions and prejudices about proteinoids are dispelled. Indeed, their branched chain racemic structures prove to be a benefit, not a liability. Replication of the membranous ur-cells in on a par with the replicability of RNA as an essential mechanism in making a genetic system. These porous compartments are ideal for allowing the primordial soup in and out while trapping any polymers that might form inside. The microsphere membranes are made from racemic hydrophobic proteinoids. Other hydrophobic polypeptides made by other means, such as by the *primitive RNA translator*[Part 5], rather than by the thermal synthesis that produces proteinoids, can be self-assembled into the ur-cell membranes, causing the ur-cells to grow and divide. The combination of growing and replicating ur-cells and replicating RNA's leads to the possibility of ur-genetics [Part 9]. Considerations of thermodynamic and kinetic barriers to polymer formation are central issues [Part 3]. Initially, only short polymers can be formed because of these constraints. However, this allows the evolving processes *to try out all real physical possibilities* and thereby avoid the combinatorial limits forced on us by long polymers. Thus, some dedication to "counting" possibilities is required. Darwinian evolution at the molecular level takes place.

Central to the entire discussion in MT is the role of energy [Part 2]. UV driven oxidation of ferrous iron supports redox reactions on the early Earth. Together with thermal energy, redox energy generates thioesters. The thioesters in turn drive many processes similar to modern metabolism, and indeed modern metabolism contains many thioester mechanisms, perhaps relics of this earlier stage [Energy metabolism]. Finally, thioester energy can produce polyphosphate energy, especially in the form of pyrophosphate, the precursor to ATP. Phosphates serve as activators for polymerizations. In the case of amino acids, that are naturally abundant products of abiotic synthesis, the two distinct functional groups, the carboxyl and the amino, are ideal for both activation by pyrophosphate and polymerization into polypeptides by nucleophilic attack of the activated carboxyl phosphates by the amino groups. These circumstances answer the question: why were amino acids selected out of all the many molecular products of abiotic synthesis for polymerization? It is clear from the analysis in MT that the structure of biomolecular mechanisms is a result of the types of energy that flow through biomolecular matter: light, heat, redox, thioester and polyphosphate [Part 2]. Elsewhere [*Energy and the Evolution of Life*, W. H. Freeman and Sons, New York, 1988] I have expanded on this theme with regard to human evolution by considering the roles in societal evolution played by: sunlight, fire, gravity, metallurgy, water wheels, norias, harnessed animals, irrigation, steam, electricity, petroleum and nuclear fission and fusion. Electricity in particular has transformed human existence in so many profound ways. Hydroelectric generators (either gravitational or steam) are analogs of energy coupling between redox energy and thioesters or between thioesters and polyphosphates. Would our society be what it is today without these types of energy? Is not cellular life what it is today as a consequence of the types of energy (UV light, heat, redox, thioester, polyphosphate) that are natural on the Earth, or anywhere else in the Universe suitable for life?

Let me go back to the important issue of complementarity in the evolution of the genetic code [Part 11]. There, the argument leads to the table below showing how complementarity sets are evoked by the mapping made by expressing the code as aa's for all the complements of a given codon. Closed groups of aa's are formed. As originally proposed by the Rodin's, there is a tight connection to the aaRS classes I and II, given by green and red, respectively.

arg	CGN	→	$\bar{N}$CG	pro	ala	thr	ser
gly	GGN	→	$\bar{N}$CC	pro	ala	thr	ser
pro	CCN	→	$\bar{N}$GG	arg	gly	arg	trp
ala	GCN	→	$\bar{N}$GC	arg	gly	ser	cys
thr	ACN	→	$\bar{N}$GU	arg	gly	ser	cys
ser	UCN	→	$\bar{N}$GA	arg	gly	arg	stop
ser	AGY	→	RCU		ala	thr	
arg	AGR	→	YCU	pro			ser
cys	UGY	→	RCA		ala	thr	
trp	UGG	→	CCA	pro			
stop	UGA	→	UCA				ser
glu	GAR	→	YUC	leu			phe
gln	CAR	→	YUG	leu			leu
leu	CUN	→	$\bar{N}$AG	gln	glu	lys	stop
leu	UUR	→	YAA	gln			stop
phe	UUY	→	RAA		glu	lys	
lys	AAR	→	YUU	leu			phe
stop	UAR	→	YUA	leu			leu
asp	GAY	→	RUC		val	ile	
asn	AAY	→	RUU		val	ile	
his	CAY	→	RUG		val	met	
tyr	UAY	→	RUA		val	ile	
val	GUN	→	$\bar{N}$AC	his	asp	asn	tyr
ile	AUN'	→	$\bar{N}'$AU		asp	asn	tyr
met	AUG	→	CAU	his			

The arg/cys block is discussed in detail in [Part 11]. The glu/asp blocks are also discussed. Could it be that in some of the many other places in the Universe where there is life that the glu is coded by GAY and asp is coded by GAR instead. If these connections are made by the proposed mechanism, a genetic mechanism, then it seems natural for this variation to be realized. Variations regarding which hydrophobic residues are associated with glu or with asp are also possible. In all cases the general structure is a CG two base code core using the *primitive RNA translator* mechanism [Part 5], followed by the addition of the cysteine group of aa's. Which hydrophobics come along with cysteine is not decided. That there are the two groups based on glu and asp is also universal, but exactly as in the table above is unlikely. Hydrophobics are important because it is necessary to have enough hydrophobic polypeptides, like the ur-collagen [Part 9], for self-assembly of the ur-cell membranes. After all, there can be no genetics if there isn't both ur-cell replication and RNA replication inside the ur-cells. In [Part 10] the role of Mg^{2+}

as the ur-polymerase for RNA's is elucidated. So is the emergence and evolution of ur-tRNA's and ur-aaRS's [Part 11].

Darwin's bicentennial
Origin of the Species sesquicentennial

It is perhaps fitting that today, February 12, 2009, is Charles Darwin's 200th birthday. This year is also the sesquicentennial of *The Origin of the Species*. The model presented in MT functions by Darwinian selection at the polymer level. All possible variations have to be realized physically for this mechanism to work [Part 3]. That is why the model begins with sufficiently short polypeptides and sufficiently short RNA's. Darwin did not know anything about the molecular structure of the cell, must less about RNA, yet his penetrating analysis at the level of multicellular species can be transferred to the nanoscale world of molecules.

Feedback is most welcome. Established researchers may find this material suitable for exciting students to think about the issues addressed. There is a lot to read and about which to think.

Epilogue redux was posted on February 10, 2010

Epilogue redux

In MT, the origin of life on the young Earth is explained by the merging of three conceptual categories:

1) *energy*
2) *compartments*
3) *polymers*

1) The geophysically natural *energy* transductions develop as a series that includes heat, UV, Fe^{2+}/Fe^{3+}, thioesters and P~P in that temporal order. Dry heating is also productive of P~P. Eventually thermal oligomers are replaced by oligomers made from activated monomers. P~P, by way of Fe^{2+}/Fe^{3+} and thioesters, takes over from heat as the activating agency.

2) Natural *compartments* are microspheres self-assembled from thermal proteinoids, that are racemic and branched thermal polymers of amino acids. As these microspheres accrete proteinoid the membrane grows and divides or buds, producing a population of spheres. Much can still be learned about the biophysics of this ur-cell replication mechanism. These events are independent of polynucleotides and co-equal in importance when making a model.

3) The activation energy can drive the generation of *polymers* from activated monomers inside the microspheres. At some minimal length, perhaps only ~ 6-9 units, the polymers are effectively trapped inside the porous membranes while the primordial energized soup readily permeates and is shared by every sphere.

⇔

Trapped sequences of nucleotides inside microspheres constitute an ur-genome. If replication and transcription of the polynucleotides can take place in synchrony with the microsphere division (budding) then the genome can propagate, and the setting becomes ur-cellular genetics.

Replication of polynucleotides is essential for their function as information carrying macromolecules. This is where chirality comes in. When the activated monomers, the nucleotide tri-phosphates, are all of the same chirality and the template is also homo-chiral, replication or transcription can take place smoothly and relatively quickly. If there is mixed chirality in the template the replication or transcription processes will jam at a junction between opposite chiralities. Thus ribose chirality was selected when polynucleotides became active biologically (genetically). That D-ribose was selected on the Earth rather then L-ribose may be a simple accident. Homo-chirality for polypeptides may have arisen in order to make the *primitve RNA translator* run more smoothly. Given D-ribose, L-aa's may have been necessary, and preferrable to D-aa's, for translation of the ur-messenger RNA into homo-chiral, unbranched polypeptides.

In the present model the genetic code began with a three base spacing, but with only a two base alphabet, C and G. By physico-chemical attraction CGN codes for arg. This could be coulombic attraction between the negative phophates of the triplet nucleotides and the positive guanidinium group of the arg. Rather then having its origin in a typical amino acid abiosynthesis like the smaller amino acids, arg may be part of the nitrogen metabolism, particularly FeS-P~P-thioester-World precursors to the urea cycle and the metabolic processing of ammonia.

Using the *primitive RNA translator*, polyarginine could be an early product of the gene $(CGN)_n$. It could act as a ligase for polynucleotides, and with Mg^{2+} serve as an ur-polymerase. With the incorporation of some A and U, new amino acids are added to the code, still three base spacing but with a four base alphabet. Only because of the remarkable symmetry between CG and GC, and AU and UA, and!! CG and AU is it possible to enlarge the codon lexicon. However, with just C and G the mechanism supports arg, gly, ala and pro. These lead to polymers of ur-collagen. This precursor to modern collagen is strong and elastic and well suited for self-assembly into the growing membranes of microspheres. Perhaps ur-collagen based membranes replaced the thermal proteinoid membranes. The next amino acid to be incorporated after arg, gly

ala and pro will likely be cys. The versatility of csy, as a structural element (disulfide bonds), as a catalytic element (thio group) and as a metal chelator (Zn^{2+}) makes it the best candidate for the next coded aa. From here until the coding is finished, the rules are versatility, availability and benefit to the entire system (lots of hydrophobic residues for membrane constituents; aa's with carboxyl group residues) as well as symmetry rules. With cys and arg, Zn-finger motif could have emerged and proven helpful in promoting replication and transcription.

Several present day metabolic and biochemical features show signatures of the past. That Fe-S, P~P, coenzymes, chirality, ur-collagen and zinc-fingers are a few of these shows how rich the model is.