ur-tRNA and ur-aaRS

Reynard: In the context we have developed here, the origin of *tRNA* is intimately tied to the origin of *aaRS*'s. What good would the aaRS's be if there were no tRNA's for them to service? So, tRNA's emerge first.? Propagating racemic proteinoid microsphere populations exhibit internal polymerizations of RNA and of pure chiral linear polypeptides. The spheres, *ur-cells*, are porous to the monomers but not to polymers of sufficient length. However, polymerization is uphill thermodynamically and kinetically constrained. These barriers are overcome by robust energy ur-metabolism culminating in strong, pyrophosphate driven polymerizations. These trapped polymers constitute the ur-genome and ur-proteinome. RNA strand propagation, catalyzed by an ur-polymerase that makes a complement of a complement that is equal to a replica, must keep pace with the ur-cell replication rate. Not all steps are uphill thermodynamically or are driven by a high P~P concentration. The self-assembly of microspheres from proteinoid is a case in point, i.e. a *spontaneous* process. If RNA replication is fast enough, robust Brownian motion fairly distributes the RNA replicas between the daughter ur-cells. Since the RNA genome propagates with the ur-cells, a population of identical genome containing spheres grows, and can do genetics. So tell me your vision of the tRNA origin.

Uranya: First, Let me review the kinetic limits. The redox-thioester world in which lots of pyrophosphate is made can support production of RNA strands and polypeptides. In [Part 3] it was argued that the lengths of the polymeric products is distributed according to the formula for the number of polymers of length i, N_i, given M_0,the initial number of monomers:

$$N_i = \frac{M_0}{e}\frac{1}{(i-2)!\,i}$$

This means that cM (*centimolar*) amounts of mononucleotides initially ($M_0 = 3 \times 10^5$) can generate about 20 octomers in a one cubic micron microsphere. The numbers of hexamers and heptamers respectively is about 750 and about 140. In the two base, CG, system, there are only 64 different hexameric sequences. Many copies of each are possible. Magnesium ions, Mg^{2+}, *may be* the original ur-polymerase.

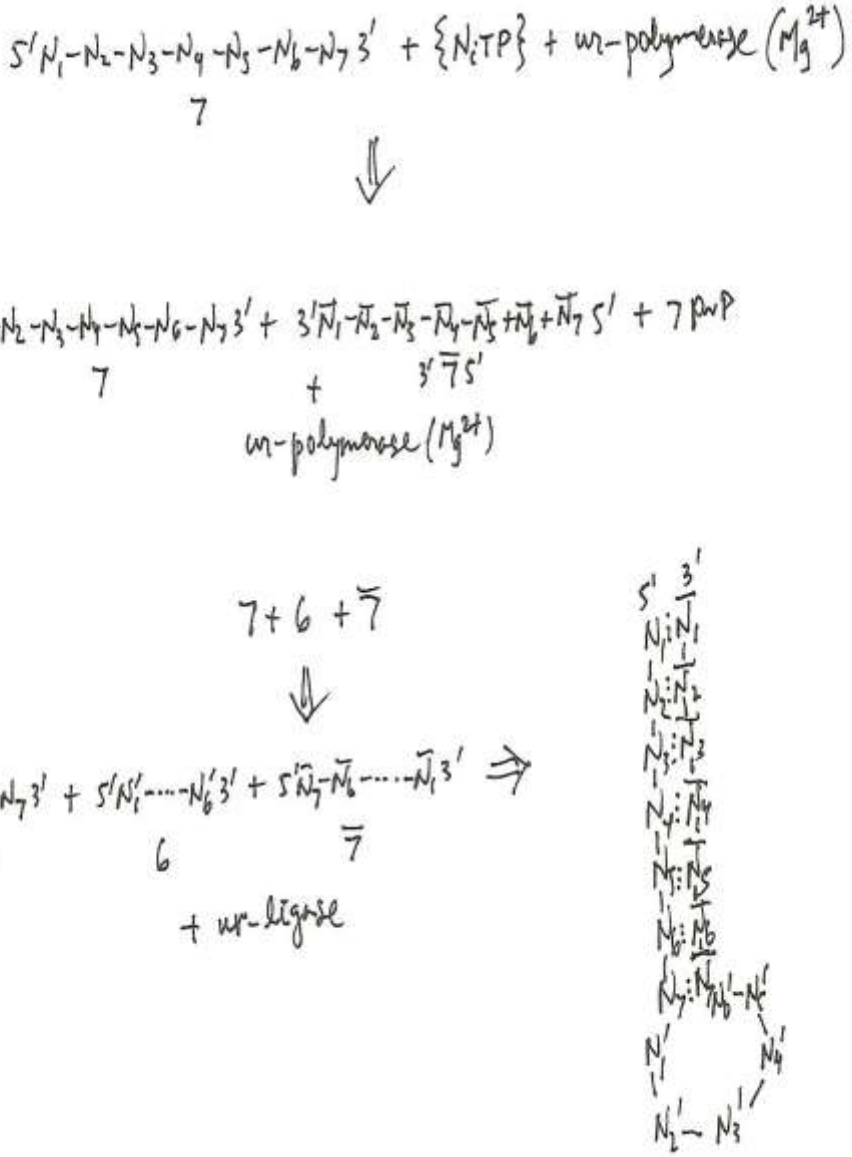
R: *Maybe* !

U: It is clear that RNA's of this length, serving as mRNA's using a three base code, can only make di- and tri-peptides. That is why one of the earliest acquisitions had to be an ur-ligase. As we have discussed before, the C and G bases can produce an arginine rich polypeptide by means of the *primitive RNA translator* mechanism that is based on a physical-chemical interaction. However, as we just said this mechanism won't produce long enough polypeptides to do the jobs needed. Initially, the ur-ligase may have to be present among the racemic proteinoids. If it is (*felicity*) then there is a way for short RNA polymers, such as hexamers to octomers, to become ligased into longer polymers. The key observation is that initially RNA's up to lengths of 6-8 monomers are all that we can expect. If these can become ligased then longer chains are possible but the kinetic limits still apply. That is, the hexamers, say, are now the *monomers* for the ligase and kinetics says that on average we should expect not quite 3 ($e < 3$) such *monomers* to become ligased, or an RNA of length ~19 (say 6+7+6). Note that when the original polymerization of nucleotide triphosphates occurs, the 5' end remains charged with triphosphates while the 3' end is available for nucleophilic attack on the α-phosphate of a 5' triphosphate anywhere in the ur-cell. So there is already enough energy available to form the phosphodiester bonds needed for ligation.

R: What is the effect of having both an ur-ligase and an ur-polymerase functioning together?

U: That is the really interesting case. If there is already a functioning ur-polymerase, probably based on Mg^{2+}, then once some hexameric to octomeric RNA has formed, lots of replicas are possible. Moreover, since replication is really complementation of the complement, there are also lots of complements around inside the ur-cell. Therefore, when the ur-ligase acts it is likely to produce *trimers* perhaps of the form 5'(7 + 6 + $\bar{7}$)3' where $\bar{7}$ denotes the complement of the initial heptamer, denoted by 7. Now think about the conformations of these segments. The $\bar{7}$ is made from the 7 with polarity 3' to 5' by antiparallel complementarity. When it is added to the 5'(7 + 6)3' it must be reversed so that it too is 5' to 3'. Therefore, ignoring the middle 6, we have a *palindrome* and a hairpin will readily form with the 6 becoming the loop of the hairpin. Let me draw this on the blackboard. In the drawing $\bar{N}$ is the base complementary to N. Because the 6 is drawn from the same pool as are 7 and $\bar{7}$ it is also likely that the sequence of bases in 6 is either part of that for 7 or for $\bar{7}$. The 3' end of the hairpin is a site for amino acid esterification. The loop end is a site for the anticodon. Thus the

anticodon sequence is also present in the stem, and by complementarity, so is the codon sequence !



R: Don't you mean to have only 6 P~P ? And doesn't the ligase make a random coil RNA that self-assembles into the hairpin conformation because of the palindromic stem ? There would also be *trimers* among the natural products that are not palindromes and that serve as ur-genes and ur-mRNA's for the translator.

U: Yes, of course, 6 P~P. Every little detail needs to be right. So we see that hairpin precursors to the now much larger tRNA' (~76 bases) are a natural product of propagating ur-cells that contain pyrophosphate driven polymerizations *and* ur-

polymerases *and* ur-ligases. These functions are coded in ur-genes that are RNA molecules that can be translated into pure chirality, linear polypeptides. The *primintive RNA translator* does the translation. The proteinoid ur-ligase could be replaced by an ur-ligase coded in an RNA ur-gene. The ur-tRNA's provide an opportunity for the emergence and formation of a *faster* and more *faithful* translator. How tRNA's became 4 times larger remains for later exploration. Some comtemporary tRNA's still possess the signature of their anticodon in there 3' acceptor stems (in modern aaRS's, recognition of tRNA's does not involve the anticodon loop segment at all for all of Class II). For the smaller ur-tRNA's there remain other matters, such as the CCA 3' terminus found in all modern tRNA's. Is it necessary at this simpler level of mechanism? If so, what basic steps does it's presence allow?

Also notice that if the ur-tRNA in the drawing forms, it is very likely that its exact complement also forms. Therefore, we would expect to see ur-tRNA anticodons emerging as complementary pairs. This is consistent with codon complementarity. Is this the origin of the symmetry seen by the Rodin's ?

R: Go back to your ur-tRNA structures and exchange the 6's and the 7's. This means 5'(7 + 6 + $\bar{7}$)3' is really 5'(6 + 7 + $\bar{6}$)3'. Total length of 19 monomer units. I like 19 better than 20 since 76 = 4 x 19.

Let's get into the details of P~P and aa binding polypeptides. The material of length 19 can also exist without palindrome symmetry. These lengths can serve as ur-mRNA for the *primitive RNA translator*. Other lengths can as well. Thus, pure chirality, linear polypeptides (hexamers, heptamers, octomers) can arise naturally and serve many functions.

Binding P~P, and then binding P~P *and* specific aa's, are evolutonary steps needing details. If the function of a specific sequence is *felicitous*, then it can be "locked in" by an ur-cell population in which RNA replication keeps pace with the ur-cell division rate. Within this population there will be many virtually identical ur-cells in terms of the many RNA sequences they contain. An ur-gene is any RNA that is preserved by an ur-polymerase and translates into a polypeptide catalyst that *increases the overall rate of ur-cell replication*. We have already seen how polymers rich in arg could serve as ur-polymerases (plus Mg^{2+}, Z^{2+}). Their ur-genes, something like N(CGN)n, are preserved by replication. So preserved too, are all other sequences, since the ur-polymerase replicates all sequences irrespective of their specific sequence. Adding such ur-genes to the ur-genome is the result of more rapid ur-cell population growth using these ur-gene

products, compared to the growth rates for other genomes. Basically any ur-gene having a *felicitous* translation product will be added to the genome so long as a population of ur-cells with this genome remains viable. Note that there is no value judgement made, simply, faster is better than slower in sustaining a growing population of this cell type. The population dynamics is what it is. Thus, to argue that a sequence is actually a gene, the gene product must function *felicitously*. The ur-polymerase is special in that its gene product is itself ! It functions as an RNA [Uroboros].

Why and how would a polypeptide that binds P~P be the product of an ur-gene? One argument is that a P~P binding peptide limits the available P~P to those molecules that are attracked to the bound P~P. Amino acids may be more interactive than carbohydrates or hydrocarbons, i.e. sugars and hydrophobes. Thus, aa activation would dominate over other uses of P~P. This leads to more polypeptide production, some of which feeds the cell membrane growth and some of which enhances RNA transcription and replication. So a gene for a peptide that binds P~P is a good addition to the genome.

A cell with copies of the gene for P~P binding can afford to let at least one copy mutate as an unfaithful transcript. If the gene product of this mutant is *felicitous*, this mutated gene will be selected for preservation in the cell population. It may be that ligasing the genes of the genome into one strand helps guarantee replication of the *entire* genome, at a faster rate than for cells having fragmented genomes. The joining of genes allows smooth replication of one gene and then another. Again recall that arg rich peptides can serve as ur-ligases in addition to already serving as part of the ur-polymerase.

The gene for P~P binding is the progenitor for genes for P~P *and* aa binding. The first step could be ligation of the P~P binding gene to a gene for generic aa binding. *Fusion* of two genes is a great way to couple functions. The second step is a mutant of the P~P binding *and* aa binding fused gene that still binds P~P but is dominantly specific for a particular aa residue. Several of these with differing specificities could evolve in turn. The aa's that would be used would have to be those that are *available*, some from abiotic syntheses like gly, ala, pro,… and some from ur-metabolism like arg (the urea ur-cycle)…. The end result is a set of genes that code for a set of polypeptides, each member of which binds P~P and also a specific aa. This set has evolved one gene at a time from a fused progenitor. When viewed as a "set" from the outset, without recourse to evolution, ones sees many specific gene products all at the same time. How could all of this specificity

emerge at once? Is this an apparent case of [IC]? Not after one sees how to do it *one polymer at a time* !

Another early polypeptide binds trinucleotides. The cell interior is rich in hydrolysis products of the longer RNA's, that include trimers, and is rich in trimers based solely on kinetics for a P~P driven system. The triplet binding peptide is rich in arg, and the arg residue interaction with the ribophosphate RNA backbone is non-specific for particular triplets. Thus one triplet binding peptide makes 64 bound triplet states. Each of these is quite specific for binding a particular ur-tRNA loop by complementary, antiparallel base pairing. By itself this binding specificity doesn't confer any special properties on the cell. Bring in the ur-tRNA's and we can picture the origin of ur-aaRS's.

An ur-aaRS must be able to attach a *specific* aa to a *specific* ur-tRNA^{aa}. Let the triplet binding peptides with bound triplets serve as the ur-aaRS's. Their specifity for aa is the natrual one already discussed that utilizes the small set {gly, ala, pro, arg} and a CG code. Binding specificity is good but not perfect. The aa forms a ribo-carboxyl ester at the 2' –OH of the ribose attached to the second base in the triplet, just like in the *primitive RNA translator*. The intermediate that reacts to form the ester is an aa-carboxyl phosphate made from P~P. The appropriate ur-tRNA^{aa} is recognized by the ur-aaRS through the base pairing of the codon (bound triplet) and the anticodon loop. Thus, the ur-aaRS is a ribozyme when it comes to ur-tRNA recognition. The triplet RNA component carries the recognition function. Today's aaRS's are pure proteins containing no ribozyme activities. Continuing with the model, how does the aa transfer from the ribose to the tail end of the hairpin stem? Does CCA3' arise as a small ligased gene fragment that is selected from many ur-tRNA's that also have 3' extentions. Imagine having the extra CCA extention. Once the ur-tRNA binds the ur-aaRS by base pairing, a conformation change in the shape of the hairpin could occur in which the hairpin bends its 3' end down to close to the bound aa on the triplet (the helical stem is involved and is particularly well suited to coupling a twist and a bend). The base 5' inside the CCA end has been called the *discriminator* base. In this model, it would be nice (*felicitous*) if this base were the complement of the triplet's second base. I have looked at recent data (B. Mallick et al., *DNA Research* **12**, 235-246 (2005)) about the base at position 73 (counting the 76 bases begins at the 5' end and ends at the 3' end).

$tRNA^{aa}$	identity element	codon
ala	A73	GCN
arg	A/G73	CGN/AGR
asn	G73	AAY
asp	G73	GAY
cys	U73	UGY
gln	A73	CAR
glu	-	GAR
gly	A73	GGN
his	C73	CAY
ile	A73	AU(Y,A)
leu	A73	CUN
lys	G73	AAR
met	U72	AUG
phe	A73	UUY
pro	A73	CCN
ser	G73	UCN/AGY
thr	U73	ACN
trp	A73	UGG
tyr	G72	UAY
val	A73	GUN

At the present time only five cases of such complementarity (val, ser, phe, leu and ile) are extant. Fully half of the cases are A73. However, if the ur-tRNA can bend down it can take away the aa as a ribo-carboxyl ester on its 3' CCA end. The added tail gives the tRNA a *scorpion-tail* mechanism. Removal of the aa from the triplets triggers a triplet readjustment that disrupts the stability of the base pairing interaction and the ur-tRNA leaves the ur-aaRS with its aa in tow. Cognate aa residues and codons become connected. None of this is *felicitous* unless the aa-charged ur-$tRNA^{aa}$'s can get together so that polypeptides form. An ur-mRNA that is stretched out linearly, perhaps by binding to a particular polypeptide or RNA strand (part of the ur-ribosome), could be read by the ur-tRNA's by antiparallel base pairing rules. The polypeptide product is the same as that made from this ur-mRNA by the *primtive RNA translator* mechanism. Another binding polymer (polypeptide or RNA) that binds simultaneously two aa-charged ur-tRNA's so that peptide bond formation is favored would itself be favored. Part of the recognition could be for the CCA3' end of the ur-tRNA that is general and non-specific. RNA fragments, e.g. triplets, are well suited for this task. Thus positioning of aa-charged ur-tRNA's on the ur-ribosome may use a ribozyme mechanism just as we have postulated for the peptidyl transferase activitiy in [Part 7].

What if the aa/P~P binding polypeptide gene and the RNA triplet binding polypeptide gene fused by ligation? By now you must be thoroughly bewildered?

U: Indeed, your argument sounds fragmented. There are so many parts. Are you going to argue that after the initial CG system emerged, coding for gly, ala, pro and arg, more codons and aa's are added by natural variations, mutations, in the genes followed by selection based on function? Whatever the intervening details, the structure of the evolutionary sequence of events can be characterized as an early stage based on physico-chemical intereactions that use gly, ala, pro and arg, followed by a stage of genetic acquisition of more monomer types and their codons. In [Part 10] you sounded ambivalent, even ambiguous, about whether (cys, thr and ser) deserve similar status to (gly, ala, pro and arg), i.e. having roots in a physico-chemical interaction, or is instead the first case of genetic variation as the mechanism of recruitment. That is probably what you meant by "locking in" in our earlier discussion. A convincing scenario for "locking in" is what you are starting to describe, the genetic versus the physico-chemical.

R: Let me try to clarify a few points. Since the second base of the codons for asp and glu is A, these aa's enter the scene only after the tight C/G code is broadened to include U/A as well. So I do see that asp and glu came after the C/G set. Similarly, I see cys coming in after the initial C/G set (gly, ala, pro and arg) because its first base is U. So the C in CGN for arg mutates to U and must be read by an ur-tRNA with the properly modified anti-codon. The cognate ur-tRNA needs to be esterified with cys. How does this connection get made? It cannot depend on the scorpion-tail mechanism because the cys is not already esterified to a bound triplet. The physico-chemical basis for recognition only applies to the pure C/G process. Cys gets associated with UGY because it is available and makes a bigger difference to functionality than perhaps any other candidate. All associations are tried. The combinatorics does not overwhelm the possibility of trying all real physical manifestations and noting which outlasts the rest. So cys apparently won this contest on the young Earth, perhaps because

1) it can form disulfide bonds that greatly enhance structural potentialities,
2) it can be the active site in a polypeptide that speeds up many thioester mediated reactions, and
3) it can chelate metal ions such as Fe, Mg and Z.

What other aa offers so many diverse advantages all at once? The versatility of cys is most *felicitous*.! The connection between UGY and cys is not a "frozen accident" because it is *no accident*, rather it is a "frozen instance" of *felicity*.

Making the connection is the key issue. This is a physical, molecular connection. The peptide that binds RNA triplets can become attached to the peptides that bind P~P and a specific aa by fusion (ligation) of their respective genes. But this produces a longer polypeptide that at one end binds P~P and a specific aa, and at the other end binds a RNA triplet. Because the RNA triplet binding is non-specific, all possible triplets are allowed and are connected to the specific aa binding component.

This doesn't work to create a *specific-specific* recognition !

Remember how all the P~P *and* aa binding polypeptides emerged, as diversifications of a single progenitor P~P and generic aa binding polypeptide. Imagine the same sort of thing happening to the RNA triplet-binding-polypeptide. Genes for polypeptides that can mimic the binding of anti-codons by the codons, but only for specific anti-codons, could evolve. If the triplet involved is one of the important triplets needed so far for (gly, ala, pro and arg) then the new polypeptide provides an alternative site for ur-tRNA binding. Now let this gene fuse with the gene for a P~P and aa binding polypeptide. All the cases can be tried and those that agree with the earlier connections are selected by population dynamics. Thus the ur-aaRS has evolved from a very limited one that was a polypeptide bound RNA triplet, to one that carries two fused regions, one for *specific* anti-codon recognition and one for P~P and *specific* aa recognition and binding. Now the ur-mRNA can be read linearly by ur-tRNA's reading codons and carrying specific cognate aa's. By adding cys to the pool created by a single mutation (C to U in the first position of the codon) of the arg ur-tRNA, one also has the opportunity to add the complementary codon(s) having just one U/A base. Two important structural variations have taken place.

1) The recognition of anti-codon by the ur-aaRS no longer utilizes the ribozymic RNAtriplet but is done case by case by a specific polypeptide.

2) The activation of the aa as an aa~P hasn't already happened and is instead done on the P~P *and* aa binding component, once the 3' tail of the ur-tRNA becomes situated next to the aa and P~P binding region.

The end result is cognate esterification of the aa to the ur-tRNA^{aa} at the expense of P~P. One could say that the linear reading of an ur-mRNA increases greatly the repertoire of aa's because it allows a more indirect reading mechanism than the phyico-chemical mechanism used by the *primitive RNA translaton*, and that

permits genetic construction of much more functionally diverse polypeptides. The sequence of aa acquisitions by this system tells a story about what functionalities needed to be added and when, if the genome was to survive population dynamics. Clearly letting one U/A change in the first position lets cys in and letting G change to A in the second base lets in asp and glu. These two amino acids are the only two with carboxyl groups, the negative charge on which is highly versatile in polypeptide functions. After cys, these two were apparently the next most versatile aa's for making *felicitous* polypeptides. The availability of asp and glu as products of abiotic aa synthesis is high, *and* as by-products of the citric acid ur-cycle [Energy metabolism].

U: I think you've done it. I now understand "locking in."
In answer to the question:

what is the origin of the genetic code?

you say first there was the tight C/G code and (gly, ala, pro and arg) using the mechanism of the *primitive RNA translator* to translate RNA genes into pure chirality, linear polypeptides, some of which enter the membrane phase and some of which become catalysts. Second, once hairpin RNA's of size 19 emerge in the highly energized molecular soup, it is possible for ur-tRNA's to exist, maybe with CCA3' additions already. These give rise to a new type of translation wherein ur-mRNA in association with various ur-ribosomal factors is read linearly by the ur-tRNA's charged with cognate aa. This system can do exactly what the *primitive RNA transpator* does. Its ur-aaRS is just a RNA triplet binding polypeptide rich in arg. The attachment of aa to cognate ur-tRNA is achieved by the scorpion-tail mechanism. Eventually the mechanism evolves into the fusion of variants of two simple genes, one for P~P and specific aa binding, and one for pure polypeptide recognition of anti-codons. These primitive ur-aaRS's associate cognate aa's and codon's, and then attach the aa's to the cognate ur-tRNA's.

So the ur-tRNA system starts with only the C/G four codons (third base degeneracy). Now the repertoire can increase through mutations of the primal quartet. These mutants may incorporate U/A bases into the first and second positions of the codons. Cys is added, along with complementarity aa's, ser and thr. It is the unique versatility of cys as structure factor, as catalyst, and as chelator, that makes it the first aa to be added. No physico-chemical interaction is making UGY the codon for cys the way CGN is the codon for arg. Cys, thr and ser have been added by the outcome of microsphere population dynamics. Because of Mg^{2+} an ur-polymerase can function so that RNA strand replication

can keep pace with the membrane growth and ur-cell division. Hence a genome exists and is manifested by a population of ur-cells carrying that genome. Ur-genes for *felicitous* polypeptides are"locked in" by the population dynamics. For this new system to work, ur-aaRS's had to evolve as well as the ur-tRNA's. You have argued for variation in and fusion of two ur-genes so that many combinations are tried and from which the most *felicitous* are selected. The "first", CGN goes to UGY, codes for cys because cys is the most versatile aa available. The "next", GCN goes to GAN, codes for (asp, glu). The point is: subsequent acquisitions are genetic in origin rather than physico-chemical.

I also see now why you answered all the questions in [Universality] affirmatively, especially question 14.

R: Thank you for understanding.

As you see, the problem has transformed into a detailed accounting of the sequence of incorporation of the aa's into the genetic code. At the CG level there are (gly, ala, pro and arg) . These are readily available, arg perhaps coming from a urea ur-cycle. There are one positively charged residue and three hydrophobic residues. This is enough diversity to make polypeptides either rich in arg or rich in aa's having hydrophobic residues. Thus both the membrane needs and the catalysis needs can be met by polymer synthesis. Among the catalysts are the ur-ligase and the ur-polymerase. Both membrane growth and RNA transcription/replication are strongly supported by an underlying energy driven soup and ur-cells using the CG code only. Adding (cys, thr and ser) has been argued for, at least for the addition of highly versatile cys. Why thr and ser at this point? I think there are several reasons. As other functional residues are added to the code, adding more hydrophobic residues is necessary so that enough hydrophobic membrane constituents are still made in sufficient amounts to promote membrane growth and ur-cell division.Having many hydrophobic residues, in essence degeneracy of this residue type, increases the chances of making long chains of hydrophobic residues. These are needed for the membrane. Pure hydrocarbon residues, such as ile, leu, and valine, or polar neutral residues containing hydroxyl groups, such as ser and thr, are the choices. The proposal that cys comes with thr and ser can be read to mean that on the primitive Earth, population competitions between ur-cells containing genomes ended up favoring the choice of thr and ser, over ile, leu, and valine, and reflects the added versatility of the hydroxyl group over pure hydrocarbon while maintaining hydrophobicity. I am also struck by the fact that thr and ser incorporate the OH group while cys utilizes the SH group and S is directly below O in the periodic

table. Does (met, val and his) come next or does (ile, leu, asp and glu)? Again we see a mixture of functional groups and hydrocarbon residues in each case. Surely gln and asn came later, first as a two step process

1) glu + glnRS becomes $glnRS^{glu}$.

2) $glnRS^{glu}$ + NH_3 becomes $glnRS^{gln}$ catalyzed by an ur-amidotransferase.

Later a glnRS evolved that uses gln directly. The same is surely true for asp and asn. When was lys incorporated? It is natural to suppose that (phe, tyr and trp) came last. Perhaps some of the highly redundant ser codons came with (phe, tyr and trp) as well. Complimentarity connects lys to leu one way and lys to phe another way. The sequence of transition could have been early leu, lys and phe late.

I think this outlines the details of how ur-tRNA's, ur-aaRS's and the set of coded aa's evolved. The genetic code does not appear all at once and all of a sudden, creating [IC] concerns. Rather polypepetides are added *one at a time* and aa's are added in small codon complementarity based groups determined by availability and *felicity* of function.

U: I think this is a very appealing way to look at the problem. Using complementariy of codons it is possible to give structure to the contemporary code. Consider the construction of a table of complementarity aa cousins by the following means. Take an aa and write down all of its codons. Construct the complements of these codons and write down the aa's to which they correspond. The surprise is that this leads to three *disjoint* sets of aa's and codons. In the table below I use the color coding from [Part 8], and if the arg codons AGR are colored as if they were S/G as was explained in [Part 8], then we get (on the right the sequence of aa's is listed according to first base in the order C, G, A and U; N is any base, N' is C, A or U, R is G or A and Y is C or U)):

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 first grouping (arg, pro, ala, gly, thr, ser, cys, trp and stop) contains the primitive physico-chemical quartet (arg, pro, ala and gly) as well as the genetic threesome (cys, thr and ser). Why the other two groupings, (glu, gln, leu, lys, phe and stop) and (asp, asn, val, ile, his, tyr and met), are divided by asp and glu is not obvious. As noted earlier the second base mutation for the ala codon, GCN to GAN, allows asp (GAY) and glu (GAR) to become coded because of their availability and rich functionality. Clearly, this change and complementarity

would eventually yield all of the contemporary code. It is still consistent to expect late arrivals for (asn, gln, tyr, trp and phe). These are distributed among the three sets.

R: If I make use of the color scheme above, the genetic code table given on page 6 of [Part 10], becomes a colored map of aaRS class I and aaRS class II distribution. The Rodin symmetry is manifest.

1rst base	C	G	A	U	3rd base
C	pro	arg	his	leu	C
C	pro	arg	his	leu	U
C	pro	arg	gln	leu	G
C	pro	arg	gln	leu	A
G	ala	gly	asp	val	C
G	ala	gly	asp	val	U
G	ala	gly	glu	val	G
G	ala	gly	glu	val	A
U	ser	cys	tyr	phe	C
U	ser	cys	tyr	phe	U
U	ser	trp	stop	leu	G
U	ser	stop	stop	leu	A
A	thr	ser	asn	ile	C
A	thr	ser	asn	ile	U
A	thr	arg	lys	met	G
A	thr	arg	lys	ile	A

The first two columns allow C/G to U/A mutations in the first base position only. In these two columns are the entire physico-chemical arg quartet, the genetic csy triplet and a syntax stop. This sytem is very versatile and can support an ur-polymerase and a ur-tRNA based translation machinery. Once second base C/G to U/A mutations are allowed, asp and glu become possible, coded-for aa's by way of the GCN to GAN muation. Why these two aa's? As we said, they bring in the carboxyl group with its negative charge. So much versatility is gained by the addition of a carboxyl group that of all available aa residues, asp and glu are best for linking to an anti-codon recognizing subunit of an ur-aaRS. Thus GAN to GAY plus GAR, with asp linked to GAY, and glu linked to GAR, yields two disjoint sets of aa's by the complementarity mapping manifested in your table. Each new set of aa's includes an allotment of hydrophobic residues. These guarantee that the ur-cells will be able to makes lots of ur-collagen membrane and promote membrane division. All of this is made possible by a robust energy

metabolism that is based on UV-redox-thioester-P~P energy interconversions. The Iron-Thioester-P~P World is an ideal setting for these events.