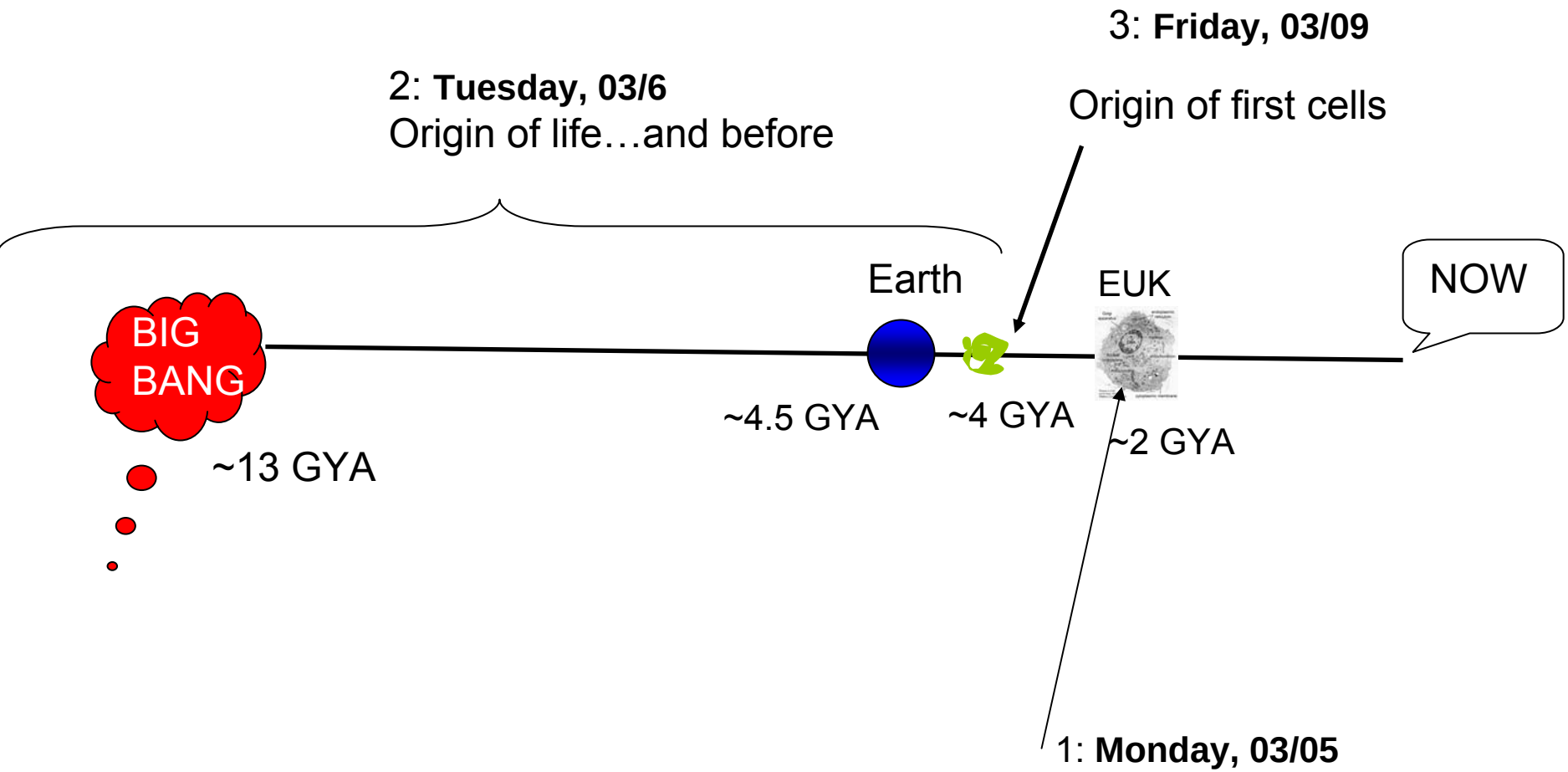
A three-part origin saga



The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life

Eugene V. Koonin

National Center for Biotechnology Information,
National Institutes of Health,
Bethesda, MD

KITP, UC Santa Barbara, March 6, 2007

“The Big Bang and the genetic code are probably the two scientific ideas that have most radically altered our view of the world in the twentieth century. The Big Bang seeks to explain how the Universe was created and how the primordial constituents came to be formed. The genetic code lays down the pattern for the formation of living material and the transmission of inherited characteristics.

Interestingly enough, the first serious theories of both these key ideas were formulated by the same man, a colourful, irresistibly playful physicist named **George Gamow”**.

Gino Segrè, **The Big Bang and the genetic code.** ***Nature*** **404, 437 (30 March 2000)**

Even Gamow apparently did not think that the Big Bang and the genetic code are linked directly…however, they just might be…

The message in a nutshell

•Origin of life is a chicken and egg problem: in order for biological evolution to take off, efficient systems for replication and translation are required, but these systems themselves appear to be products of extensive selection

•The standard solution is an ***RNA World*** without proteins in which replication is catalyzed by ribozymes and from which the translation system somehow emerges. However, the feasibility of the RNA world is dubious at best, and the path to translation is obscure

•The ***Many Worlds in One*** version of the cosmological model of eternal inflation (Vilenkin-Garriga) might offer a viable alternative

-all macroscopic histories permitted by conservation laws are repeated an ***infinite number of times in the infinite universe***

-this dramatically expands the range of possibilities for the transition from chance to biological evolution in the history of life

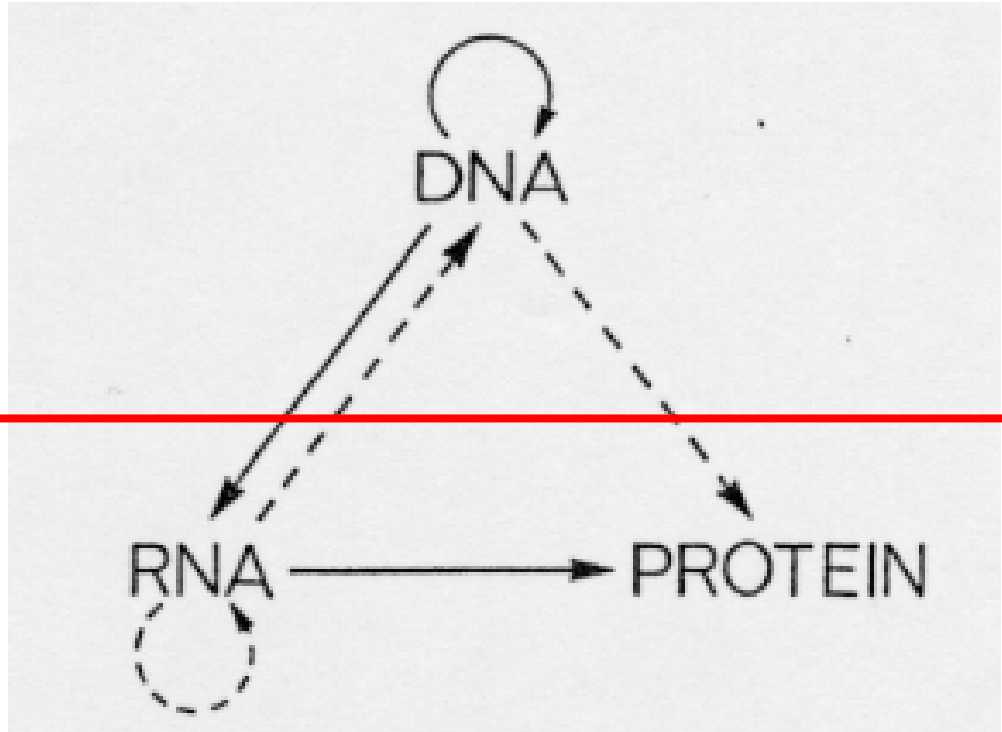
•complex systems ***inevitably*** emerge by chance

•biological evolution ***could*** start with a coupled system of replication and translation, hence ***NO*** bona fide RNA world (albeit key role of RNA in the origin of life)

•However improbable (***rare*** – in the infinite universe) the emergence of such a complex system by chance might be, its occurrence in our universe could be made inevitable by ***anthropic selection***

The “irreducible complexity” of biological information transfer and the mystery of its origin

Francis Crick's classification of information transfer pathways, 1970



Transfers relevant for origin of life

A general transfer is one which can occur in all cells. The obvious cases are
DNA --> DNA
DNA --> RNA
RNA --> Protein

A special transfer is one which does not occur in most cells, but may occur in special circumstances.

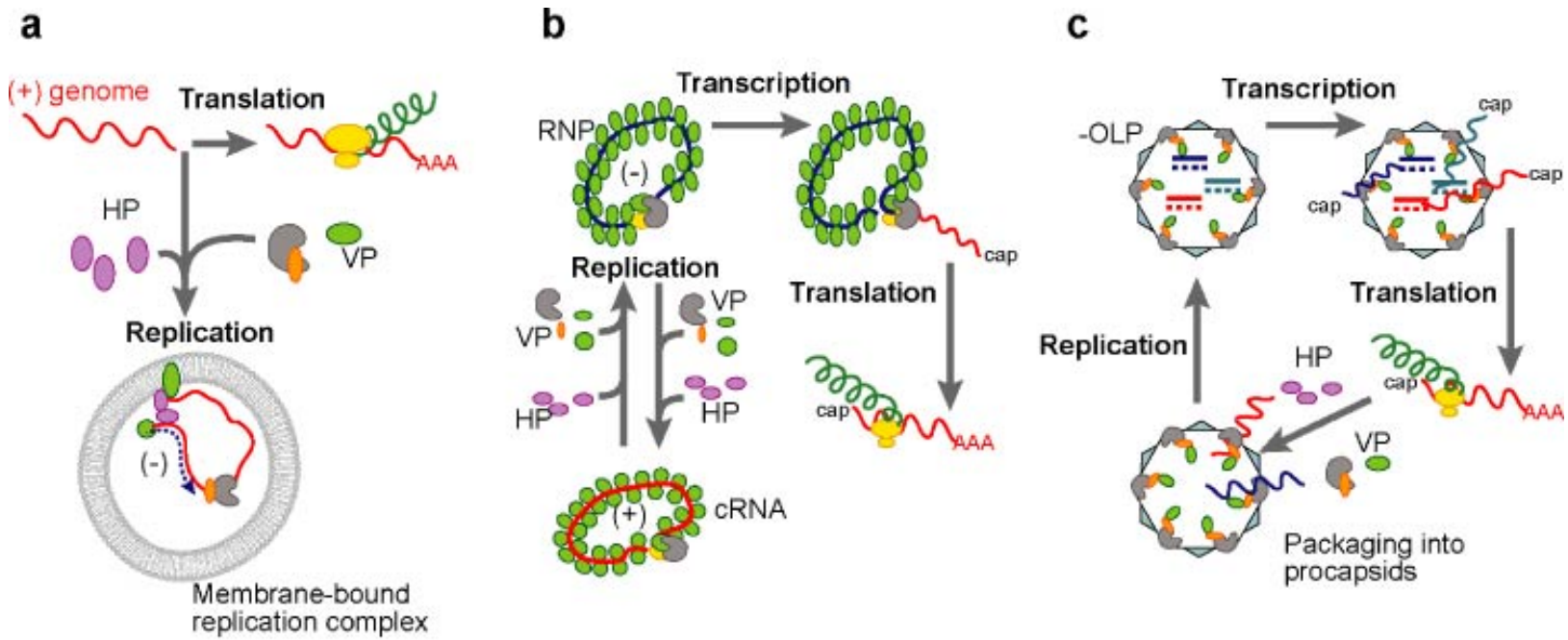
RNA --> RNA
RNA --> DNA
DNA --> Protein

It [classification] was intended to apply only to present-day organisms, and not to events in the remote past, such as the origin of life or the origin of the code

Crick FHC, in Symp. Soc. Exp. Biol. The Biological Replication of Macromolecules, XII, 138 (1958).

Crick FHC, Central Dogma of Molecular Biology, Nature, vol. 227, pp. 561-563 (August 8, 1970)

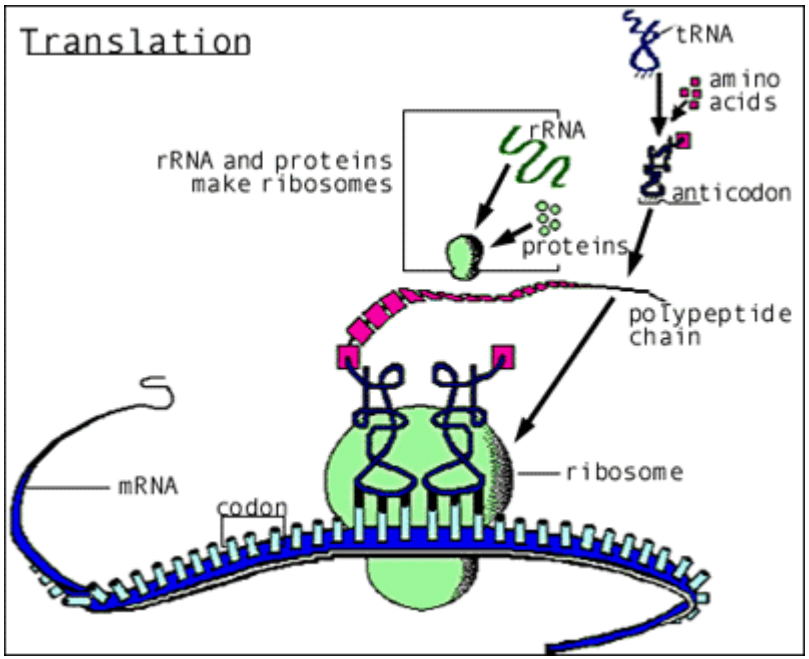
Even the simplest replication systems require specific RNA structures and several proteins that are produced by the staggeringly complex translation system

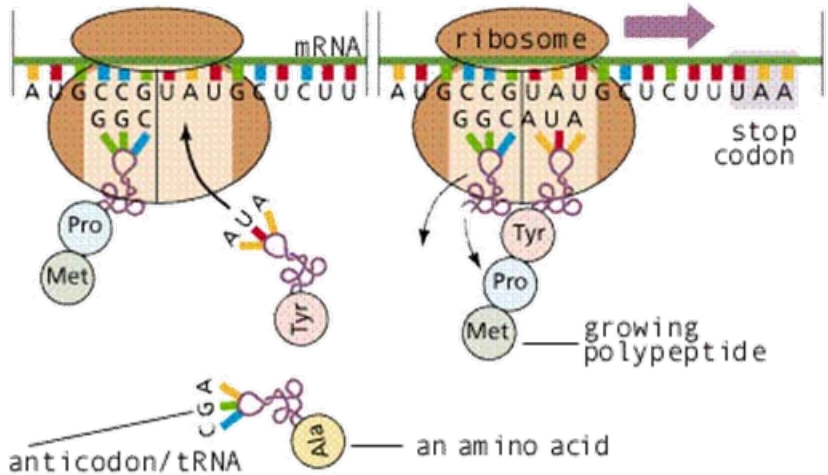


Ortín J, Parra F. 2006.
Annu. Rev. Microbiol. 60:305–26

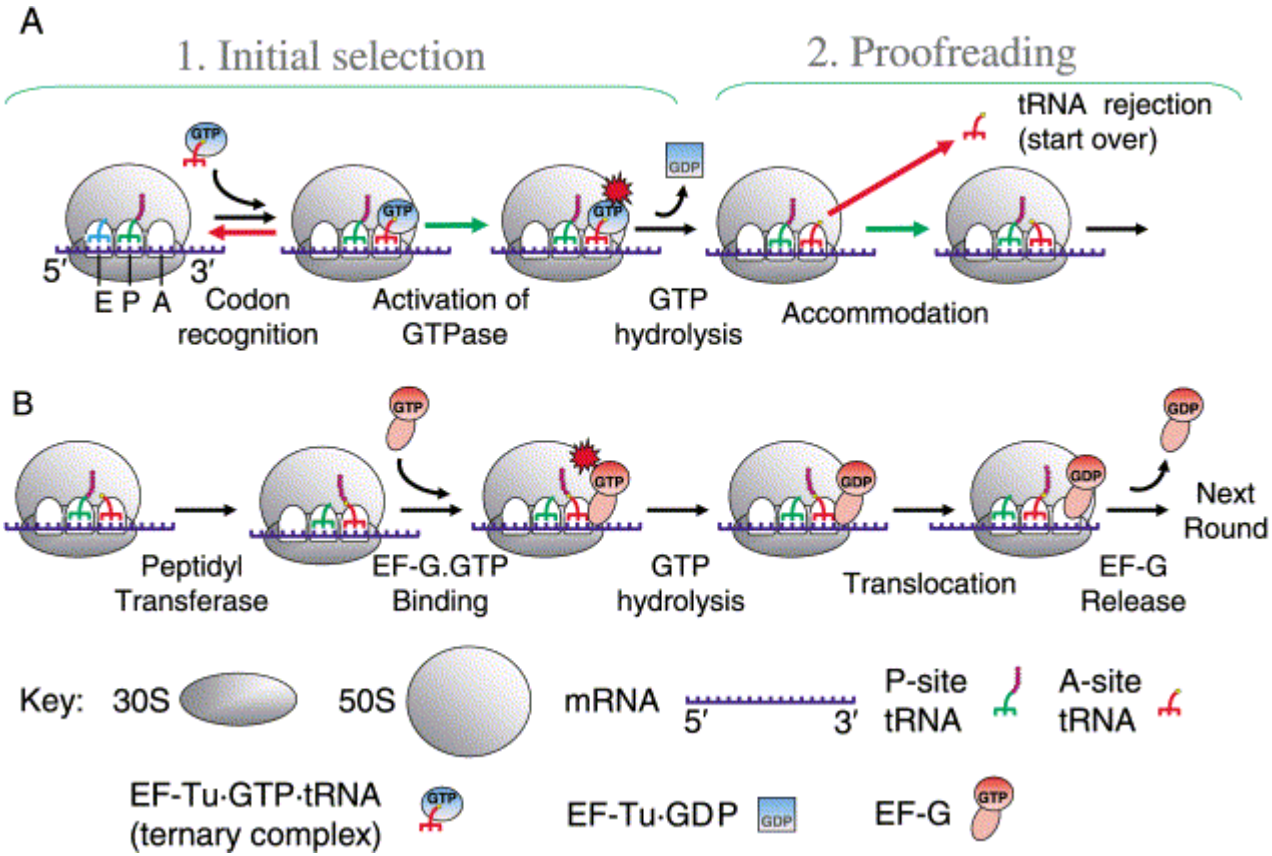
Replication strategies of RNA viruses

The dramatic complexity of the translation system

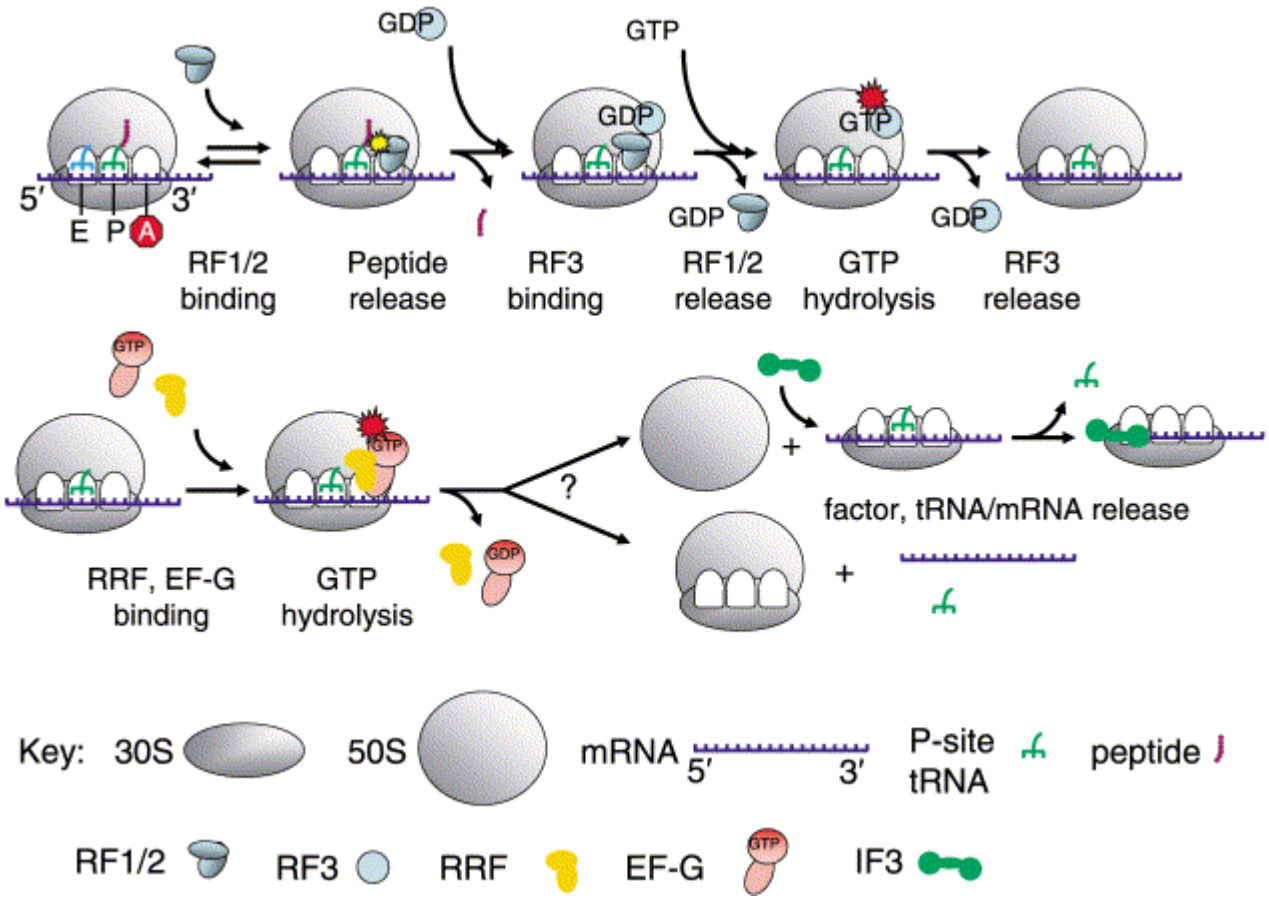




Abridged list of components:
3 ribosomal RNAs
~50 ribosomal proteins
~30 tRNAs
~20 aminoacyl-tRNA synthetases
~10 translation factors
~20 RNA modification enzymes

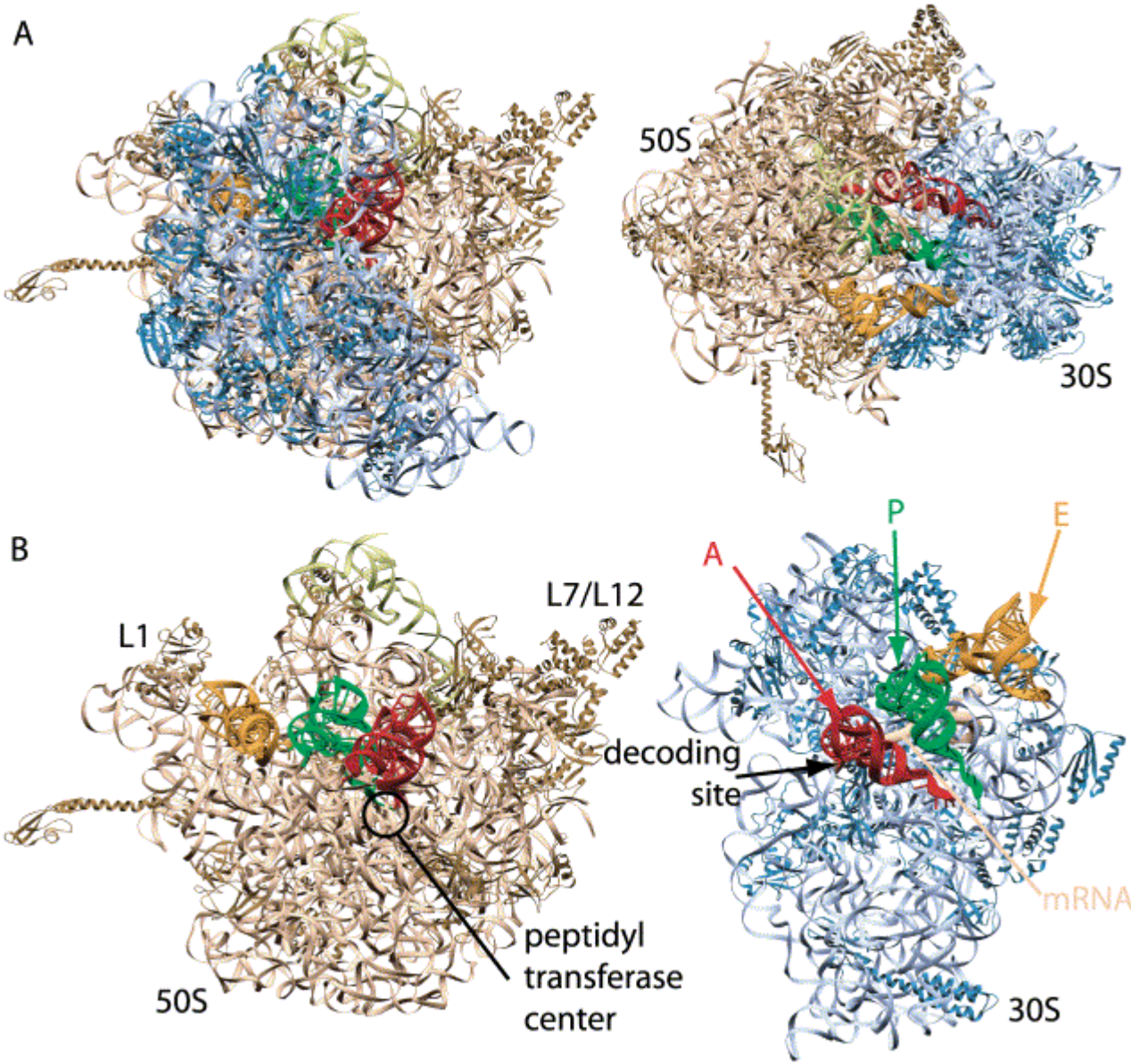


Elongation of translation

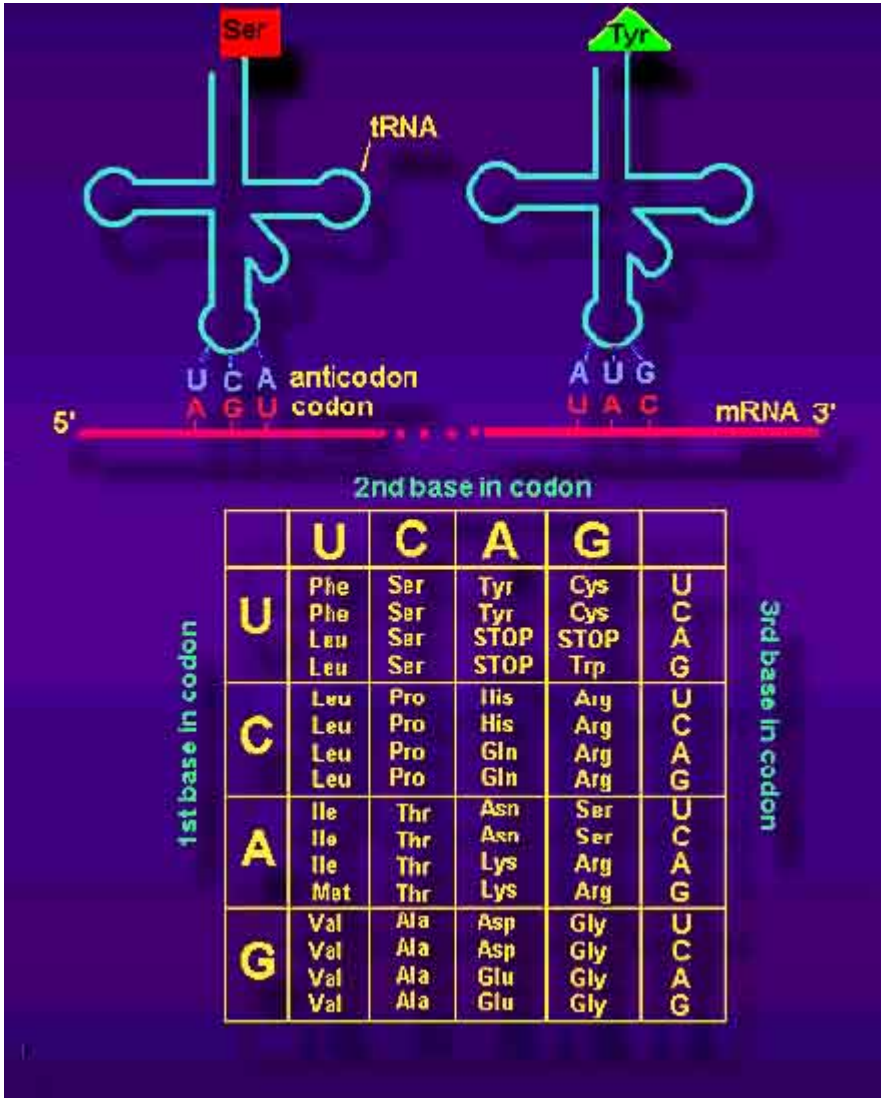


Termination of translation

Structures of the ribosome subunits



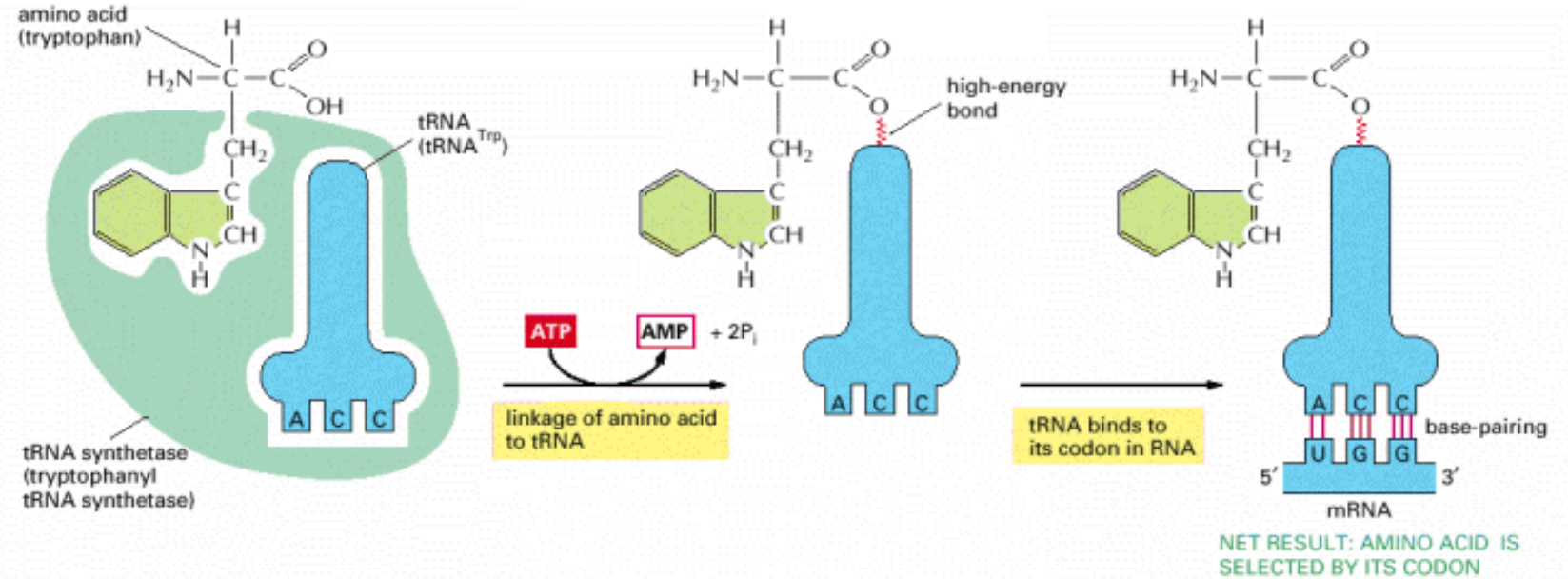
The genetic code



www.bioscience.org/atlases/genecode/genecode.htm

The code is finely structured: the 3rd letter is (partially) redundant and highly robust to mutational and translational errors: Freeland SJ, Hurst LD, The genetic code is one in a million. J Mol Evol. 1998 Sep;47(3):238-48

What makes the code work?



The genetic code is translated by means of two adaptors that act one after another.
The first adaptor is the aminoacyl-tRNA synthetase, which couples a particular amino acid to its corresponding tRNA; the second adaptor is the tRNA molecule itself, whose *anticodon* forms base pairs with the appropriate *codon* on the mRNA.
An error in either step would cause the wrong amino acid to be incorporated into a protein chain.
In the sequence of events shown, the amino acid tryptophan (Trp) is selected by the codon UGG on the mRNA.
Alberts et al. 2002. Molecular Bology of the Cell

Thus, the code specificity is mediated by a set of ~20 complex, highly evolved protein enzymes, the aminoacyl-tRNA synthetases

Origin of replication and translation (OORT): the mother of all problems in biology

Origin of complex systems: the quintessential Darwinian problem

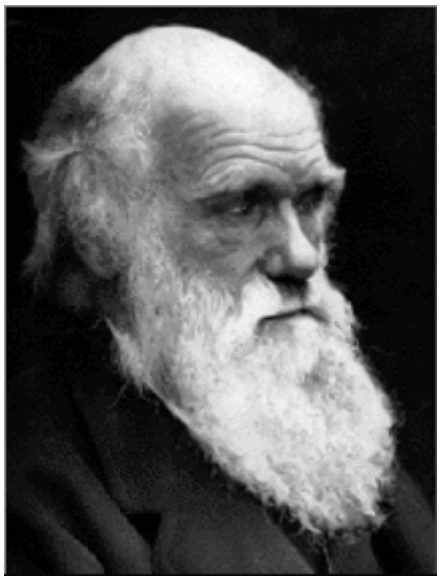
•Generally solved by Darwin – **natural selection** of successive slight improvements:
"If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down. But I can find no such case. No doubt many organs exist of which we do not know the transitional grades…"
Ch. Darwin, The Origin of Species, Chapter VI

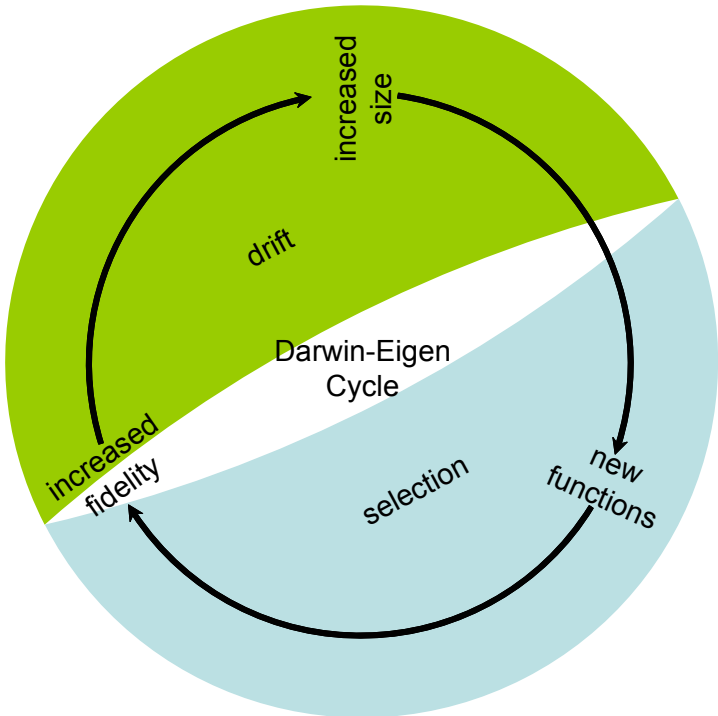
•Amended by Gould/Lewontin – **exaptation** of spandrels, by-products of evolution, for new functions

However, neither natural selection nor exaptation applies to OORT because neither process was established prior to OORT

The Darwin-Eigen Cycle

(increase of complexity via mutation and selection)





The problem is: How to start the cycle?

The paradoxes of OORT (Origin Of Replication and Translation)

1. To attain the minimal complexity required for a biological system to start on the path of biological evolution, a system of a far greater complexity, i.e., a highly evolved, coupled system of replication and translation, appears to be required

2. High translation fidelity is unachievable without a complex, highly evolved set of RNAs and proteins but an elaborate protein machinery could not evolve without an advanced translation system

Thus, it is impossible to develop a scenario of OORT without postulating evolutionary predecessors ***qualitatively different*** from the modern systems

The popular (partial) solution: The RNA World – a diverse population of RNA molecules with numerous RNA enzymes (ribozymes) including replicases capable of replicating other RNAs with an accuracy >Eigen limit (*crucial point: not just many ribozymes but effective, protein-less, RNA-catalyzed replication*)

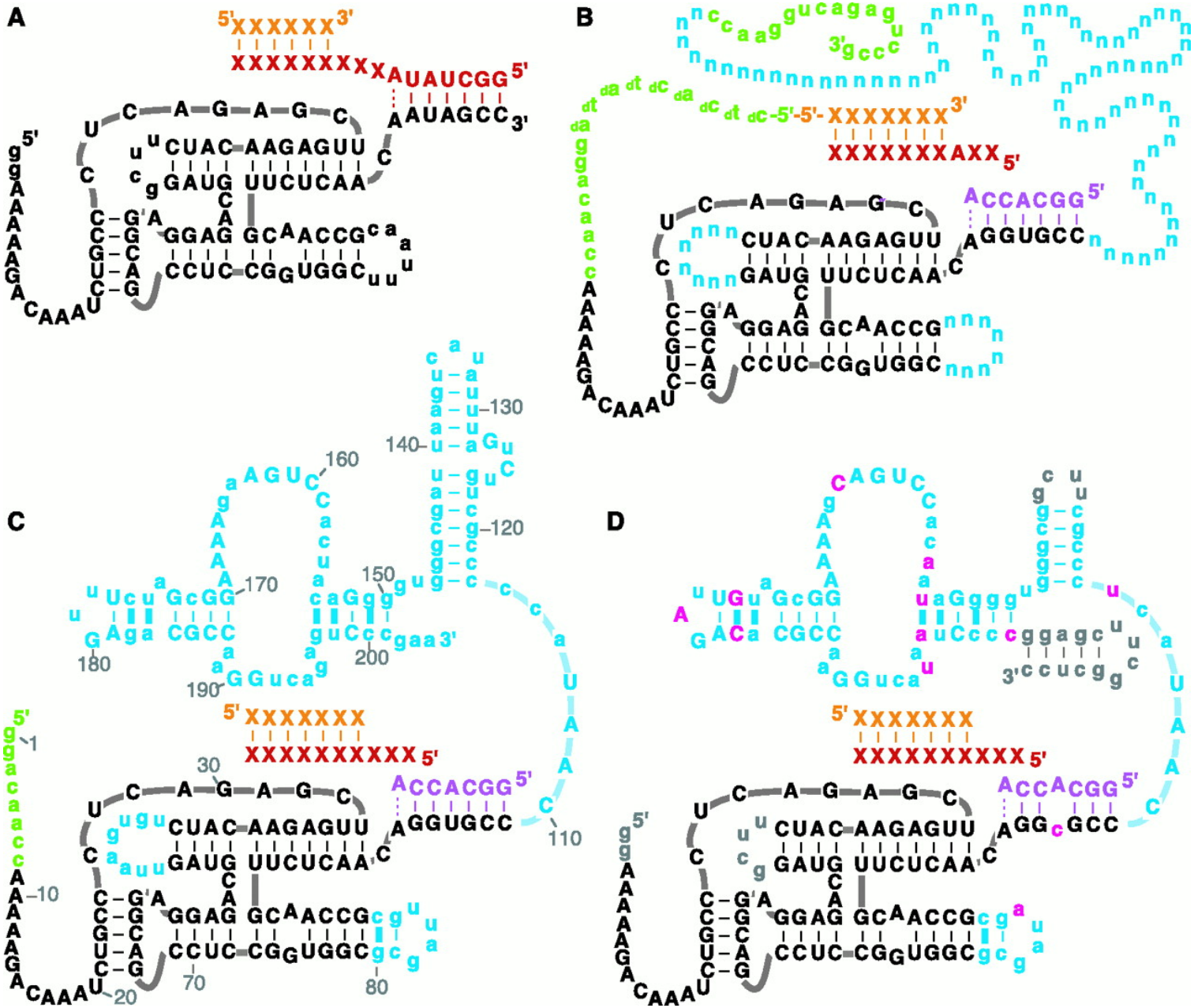
Ribozyme polymerases

Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP.

RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension.
Science. 2001 May 18;292(5520):1319-25

The RNA world hypothesis regarding the early evolution of life relies on the premise that some RNA sequences can catalyze RNA replication. In support of this conjecture, we describe here an RNA molecule that catalyzes the type of polymerization needed for RNA replication. The ribozyme uses nucleoside triphosphates and the coding information of an RNA template to extend an RNA primer by the successive addition of up to 14 nucleotides - more than a complete turn of an RNA helix. Its polymerization activity is general in terms of the sequence and the length of the primer and template RNAs, provided that the 3' terminus of the primer pairs with the template. Its polymerization is also quite accurate: when primers extended by 11 nucleotides were cloned and sequenced, 1088 of 1100 sequenced nucleotides matched the template.

Ribozyme polymerases: adding up to 14 nt



The paradoxes and problems of the RNA World

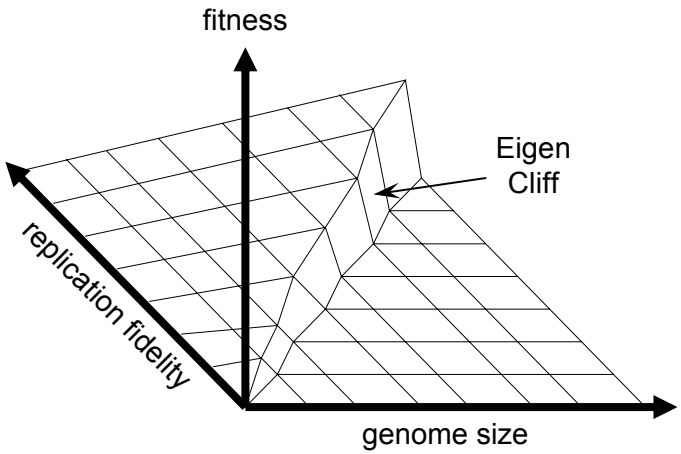
•The crux of the matter is that a sophisticated replication machinery is needed to maintain a sizeable genome, but a sizeable genome is needed in the first place to encode a sophisticated replication machinery

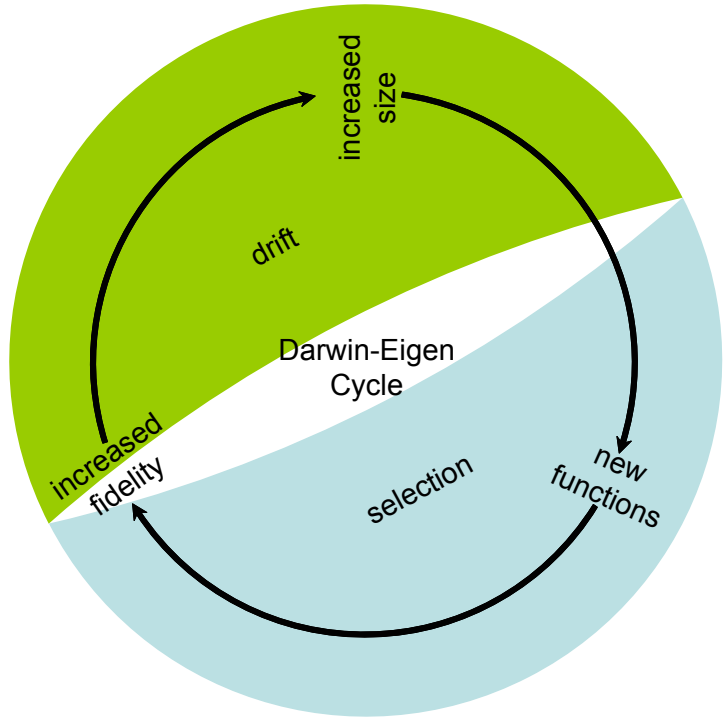
Gerald F. Joyce, Molecular evolution: Booting up life. *Nature* 2002, **420**, 278-279

•A translation system appears useless before it can produce proteins (at least, peptides) - what would be the selective forces for the evolution of translation in the RNA World?

The only thinkable solution: translation evolved by exaptation – the original selection was for some other function

Only limited complexity (at best) is attainable in an RNA World





Eigen threshold:
L = genome length (nt)
f = replication fidelity (errors/nt/replication cycle)

Lf < 1 – reproduction sustainable
Lf > 1 – error catastrophe

$L_{max} \sim 1/f$

RNA replicases: $f \sim 10^{-2}$

$L_{max} \sim 100$ nt

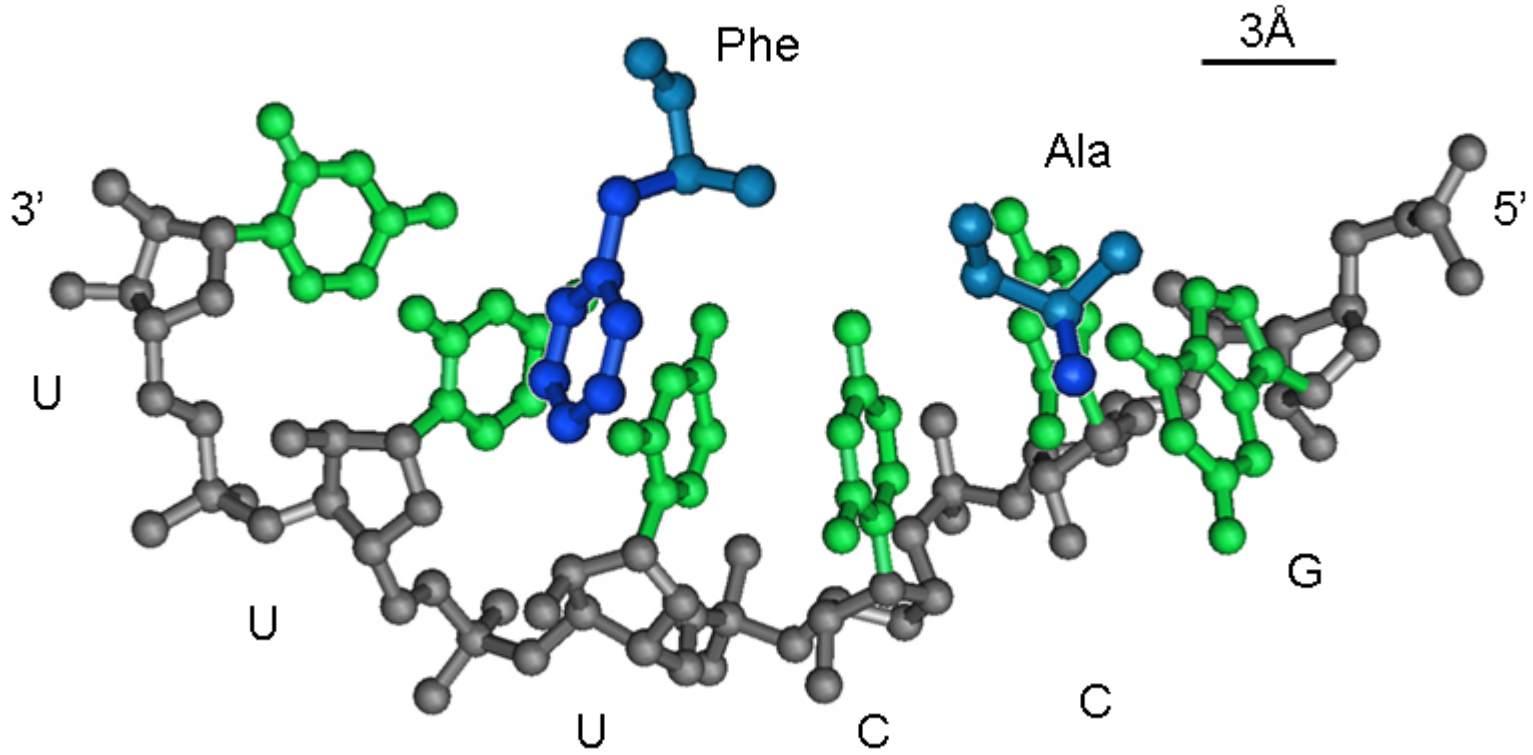
Eigen paradox:
- High fidelity unattainable without new functions (proofreading etc)
- New functions unattainable without large genomes= high fidelity

Origin of translation and coding by biological evolution: Our earnest best effort on conceptual modeling

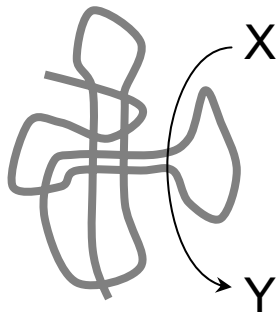
Key assumption: a full-fledged RNA world - RNA-catalyzed replication of a diverse RNA population

(Yuri Wolf and EVK, in preparation)

Direct translation of RNA is sterically impossible:
Adaptors must have been involved from the start



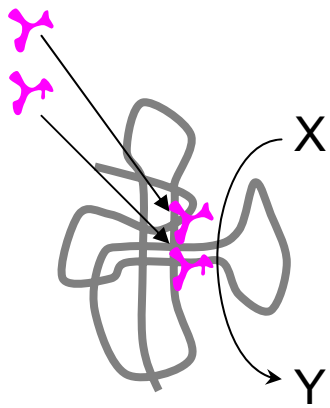
Key assumption: a full-fledged RNA world - RNA-catalyzed replication of a diverse RNA population



Ribozyme R performing an X→Y function. Could be anything…

- ***The idea of the scenario to follow: translation evolved as a collateral of stimulation of ribozymes by peptides***

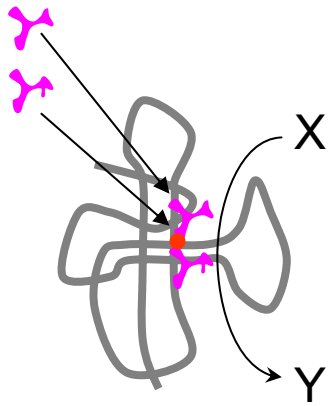
(also: Noller, The driving force for molecular evolution of translation. RNA. 2004 Dec;10(12):1833-7)



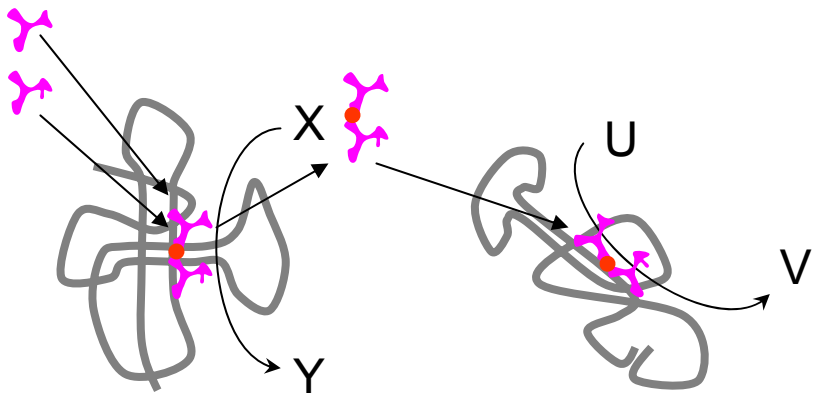
Ribozyme R with amino acid cofactors in the active site. Abiogenic amino acids; binding to an *ad hoc* site within R. Enhance the X→Y reaction. Some experimental evidence does exist.

Szathmary E. The origin of the genetic code: amino acids as cofactors in an RNA world.

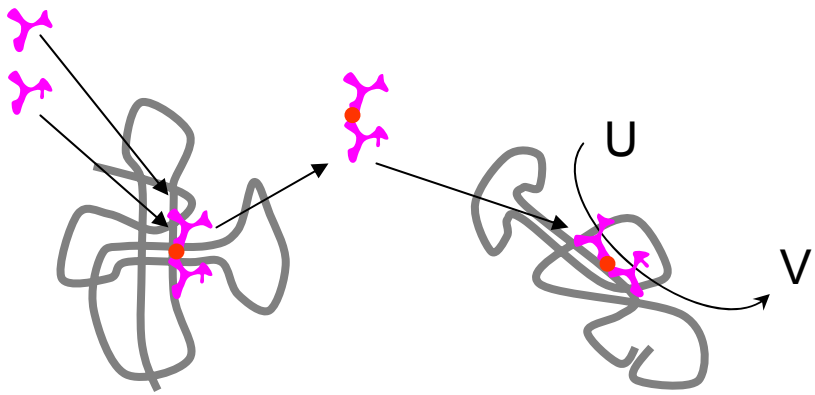
Trends Genet. 1999 Jun;15(6):223-9



Ribozyme R with amino acid cofactors linked by peptide bonds due to an additional amino acid ligase activity of R. Further boost to the ribozyme efficiency. Ribozyme ligases exist.

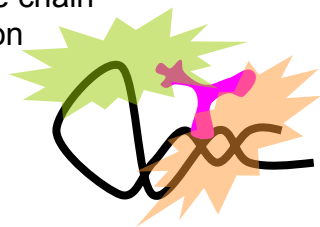


Spontaneous disassembly (decay) of R would release the peptide P which is recruited by another ribozyme E, utilizing peptide's generic catalytic properties (e.g. DD?). Advantage to the U→V reaction; co-adjustment between E and P (R indirectly).

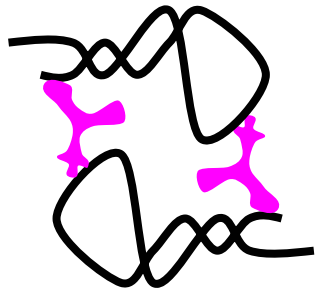


Subfunctionalization of R: R_0 retains the X→Y function; R loses the original catalytic activity; shifts to peptide production within a selfish cooperator ensemble. Boost to the X→Y and U→V reactions outweighs the extra replication costs. **Future LSU**

recognition of an
amino acid side chain
by an anti-codon

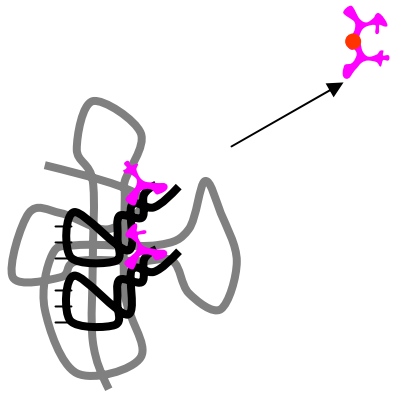


recognition of an
amino acid backbone
chain by an *ad hoc* site

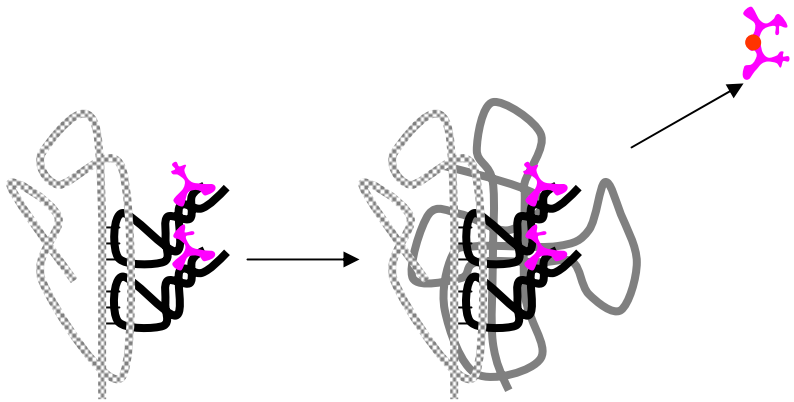


Amino acids become a commodity. Small, polar, easily diffusible molecule. Amino-acid-binding RNA species T evolves for amino acid hoarding (storage for in-house use; resource denial for competitors). Amino acid backbone chains are strongly bound by a non-specific binding site; this process is modulated by a specific recognition of a side chain (anticodon-based stereochemical proto-code) by a different part of the molecule. The side chain-specific component of the binding is of secondary importance and could have evolved later to increase the preferential storage of the more valuable amino acids. Cooperative binding by dimers can be invoked to ease the possible conformational conflicts. Duplication of T and mutation in the proto-anticodon loop allowed to fine-tune the binding for several different amino acids.

Autocatalytic aminoacylation provides a stronger, but still reversible (due to the high energy the aminoacyl-RNA bond) association between the RNA T and the amino acid, improving the hoarding/storage function.

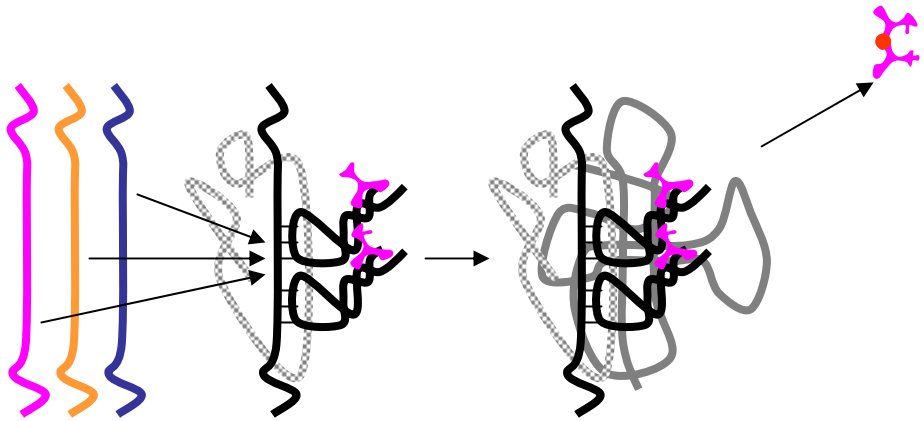


Amynoacyl-T complex provides a better way to deliver and bind amino acids into the active site of R. Dramatic improvement of stability and spatial precision of binding. The activity of R changes from peptidyl-ligation to trans-peptidylation (better thermodynamics).

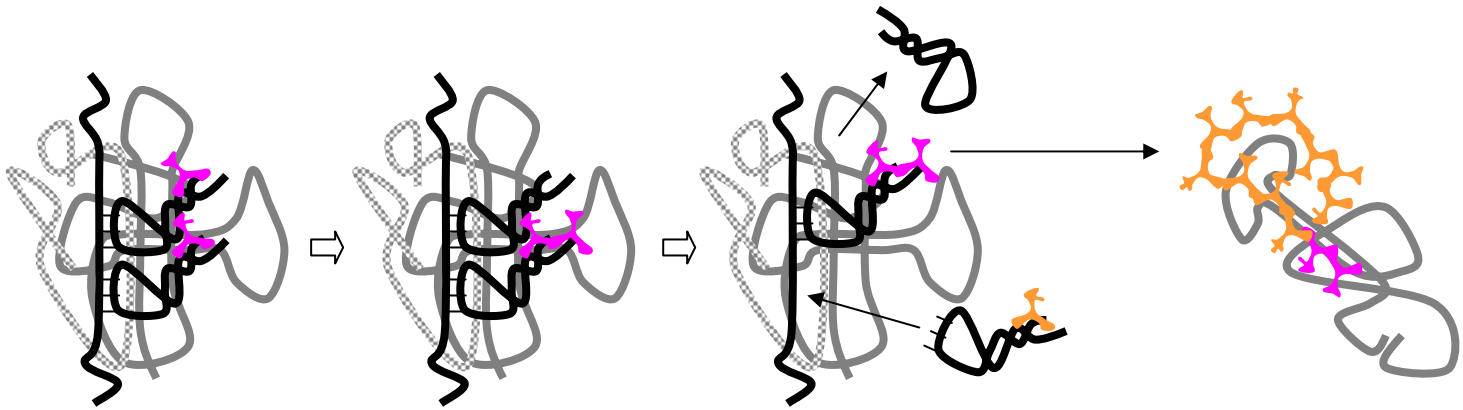


An accessory RNA subunit R_S evolves to facilitate binding and positioning of aa-T complex. The burden of specific recognition shifts from an *ad hoc* basis (Van der Vaals and hydrogen-bonds) to a more regular complementary binding between the anticodon loop of T and an extended RNA strand of R_S. Trans-peptidylation activity remains the function of R_L. **Future SSU.**

Mutant forms of R_S bind different species of aa-T, increasing the variety of peptides, produced by R_SR_L (note that mutations in R_S do not affect the catalytic performance of R_L).



Codon-bearing chain of R_S physically disassociates from the subunit. R_S becomes able to accommodate different RNA molecules M. This further releases the template from evolutionary constraints on the structure and catalytic activity of R_S and R_L.



Complex molecular movements involved in binding and release of RNA T form a mechanical ratchet that drags RNA M by 3 nucleotides through the interface of R_S and R_L. Longer (and longer and longer) peptides can be synthesized. Protein breakthrough is achieved.

A skeptical summary of the attempts (including ours) to model OORT as a case of biological evolution (experimentally, computationally, “biologically”):

there are many interesting findings but little real progress…
MANY unknown activities are postulated

…hence alternatives – if any - have to be taken seriously

A biologist's introduction to eternal inflation

- the key global features of our universe:

-spatial flatness (Ω=1)
-uniformity, in particular, homogeneity of CMB (the horizon problem)
-***amplitude of CMB variation***
-deficit of magnetic monopoles

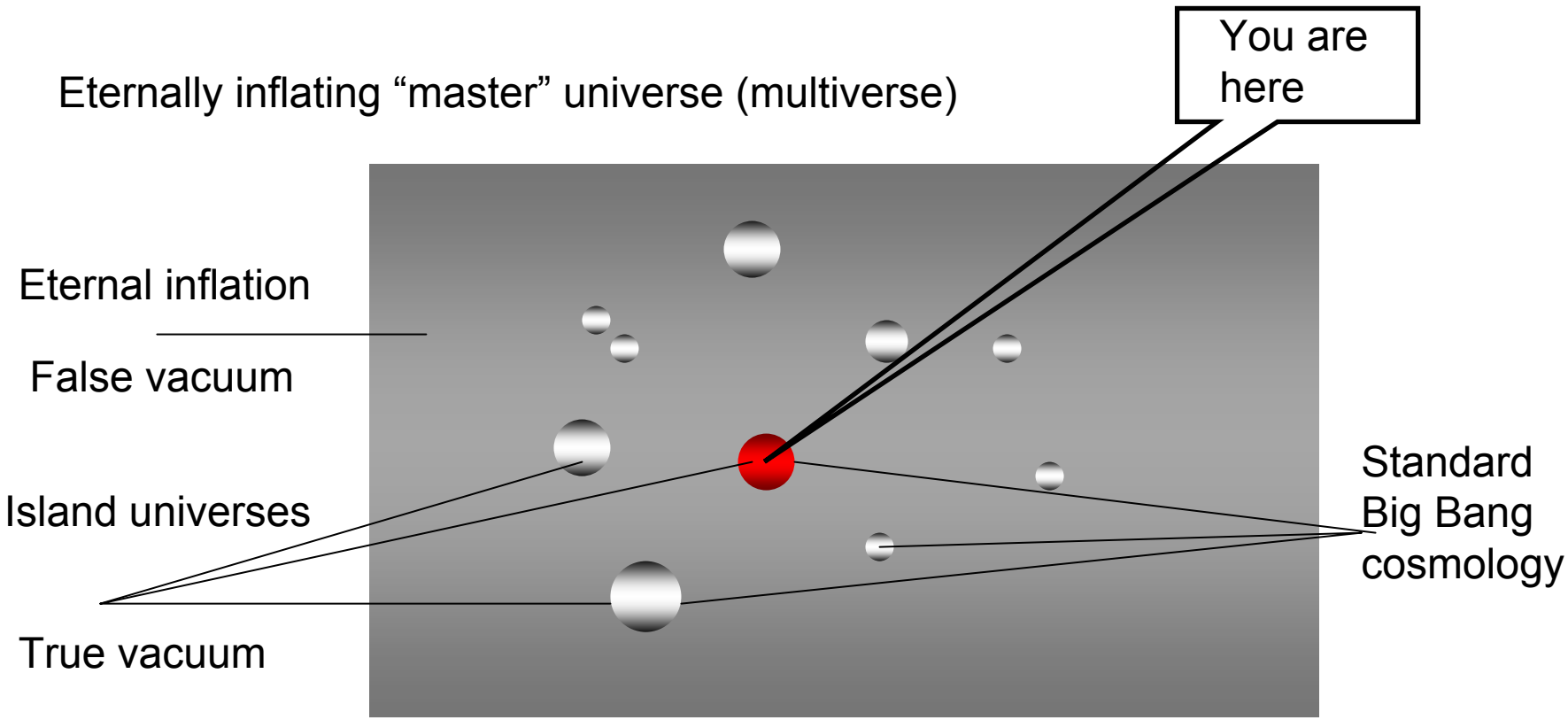
appear enigmatic in the standard Big Bang cosmology but are readily explained by ***inflation***, a period of initial, exponentially fast expansion caused by the repulsive gravity of false (high energy) vacuum

A. H. Guth, The inflationary universe: A possible solution to the horizon and flatness problems, Phys. Rev. D23, 347 (1981)

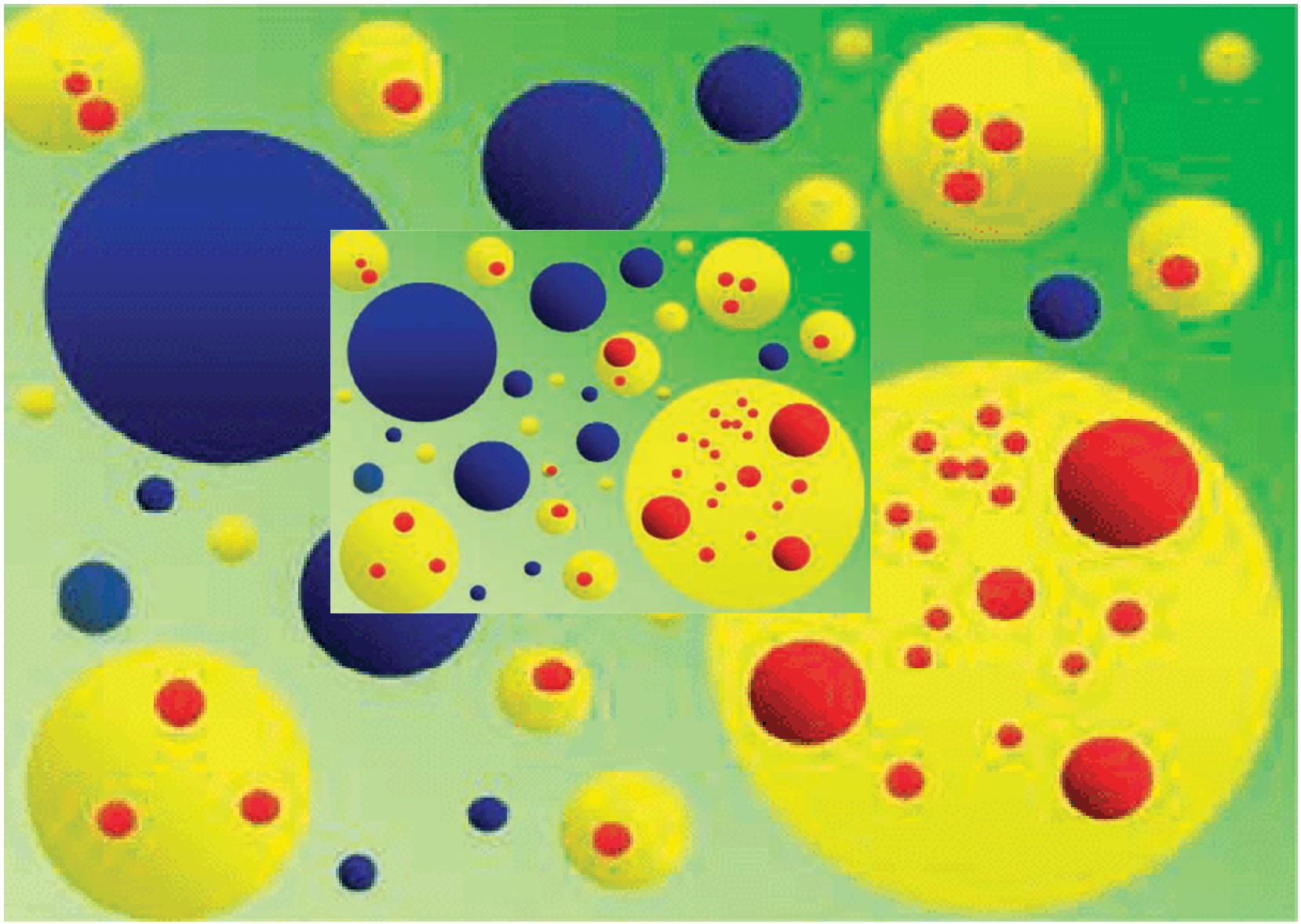
•False vacuum decay signifies the end of inflation and the beginning of "normal" cosmological evolution (standard Big Bang cosmology)

Graceful exit problem - *eternal inflation*:

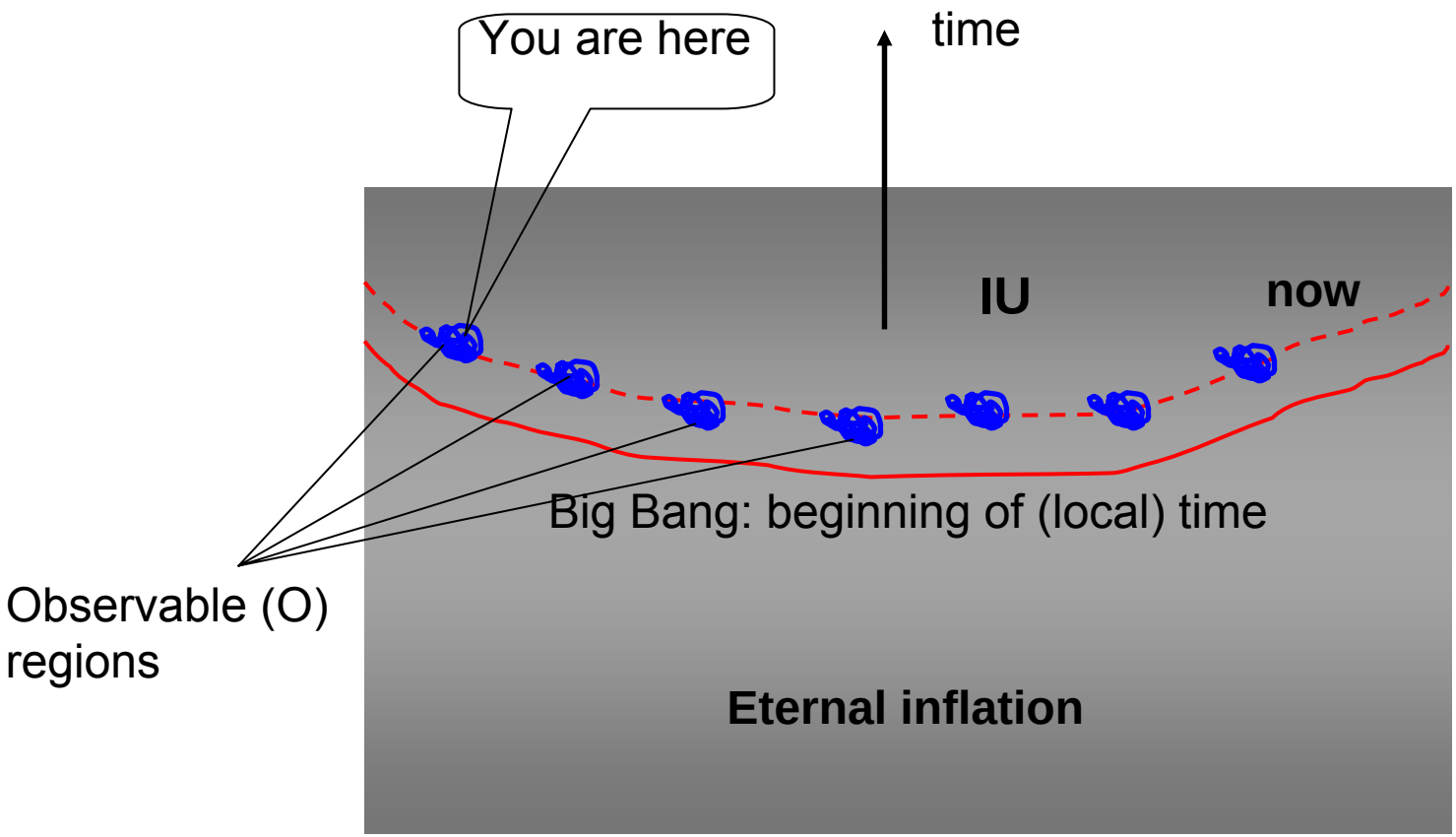
"…false vacuum regions multiply much faster than they decay. This means that inflation never ends in the entire universe and the volume of inflating regions keeps growing without bound!"
Alex Vilenkin, Many Worlds in One, chapter 8

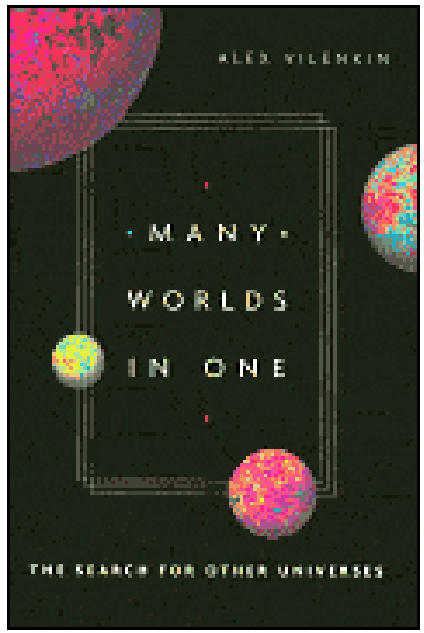


“Bubbles with different properties nucleate and expand in the inflating high-energy background. We live in one of the bubbles and can observe only a tiny part of it.”
Alex Vilenkin, *Science* 26 May 2006



Each island universe is infinite from the point of view of its inhabitants





Alex Vilenkin and the ‘Many Worlds in One’ worldview

Many Worlds in One (MWO)

- The number of all possible macroscopic histories that can be realized in an O-region is ***finite*** even if vast -

  ~$10^{10^{150}}$ – a corollary of the Heisenberg principle

N - number of times a history is repeated in an island universe

H – number of histories (finite)

O – number of O-regions (infinite)

- N=O/H = ***infinite***

Each history permitted by laws of physics is repeated an infinite number of times in any (infinite) island universe

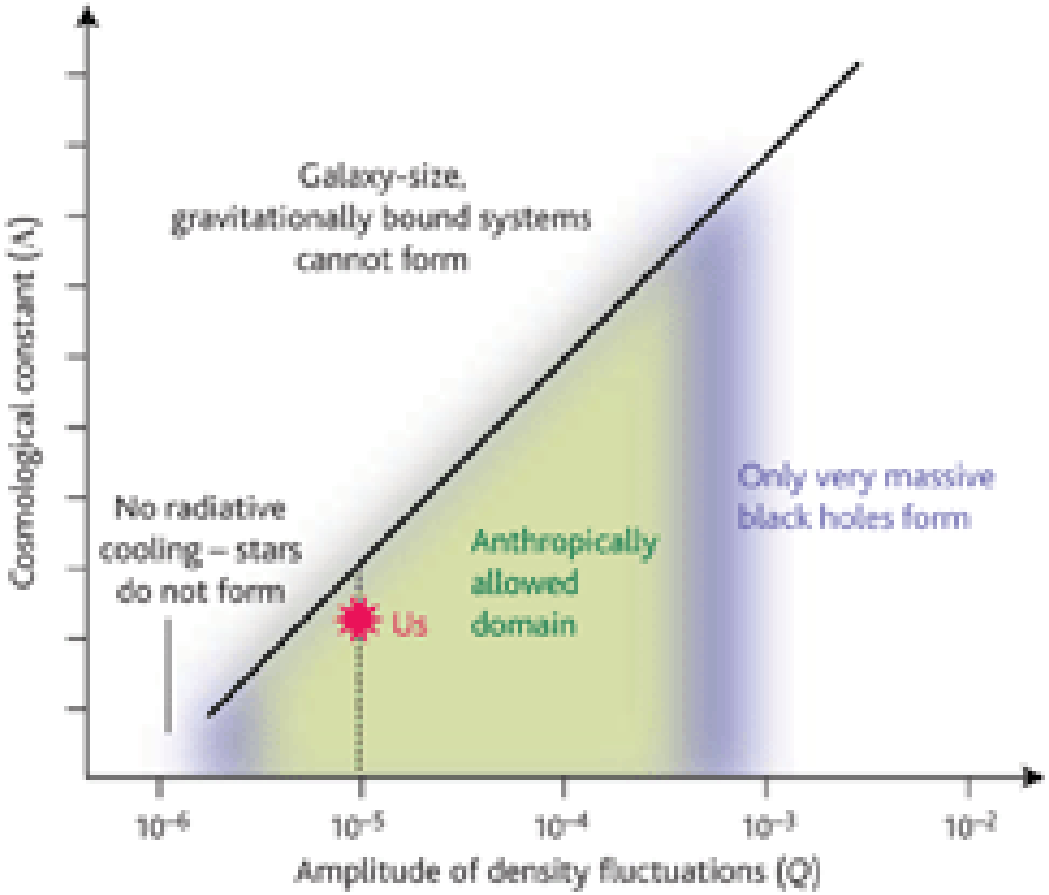
J. Garriga, A. Vilenkin, Many worlds in one, Phys. Rev. D64, 043511 (2001)

The MWO is closely linked to the ***anthropic principle (anthropic selection***):

According to the anthropic principle, the only "reason" our *O*-region has its specific parameters is that, otherwise, there would be no observers to peer into the universe.

In the MWO model, anthropic selection has a straightforward interpretation: the parameters of our *O*-region are selected among the vast number of parameter sets existing in the universe (in an infinite number of copies each) by virtue of being conducive to the emergence and sustenance of life.

Anthropic selection, the “biophilic domain” and the emergence of complexity



Livio, M., Rees, M.J. *Science* 12 August 2005

However counterintuitive, MWO is ***not a piece of science fiction or a philosophical conjecture but a straightforward implication of eternal inflation***

"…when we say that every possible history is realized in infinitely many regions, we are making a straightforward physical claim about regions of our universe"

Knobe, J., **Olum**, K. & **Vilenkin**, A. (2006). Philosophical Implications of Inflationary Cosmology. British Journal for the Philosophy of Science, 57, 47-67

MWO changes the very notions of chance and probability - under MWO, ***probability = frequency -*** any event that is allowed by the laws of physics will happen in the infinite universe with *p=1*, even though the probability that it happens in a given *O*-region might be infinitesimally small

Put another way: the 2nd law of thermodynamics is a ***statistical law*** that can be violated, just very rarely – in the MWO multiverse, it is severely violated an infinite number of times

MWO/anthropic selection have profound implication for the history of any phenomenon in the universe, and origin of life cannot be an exception…

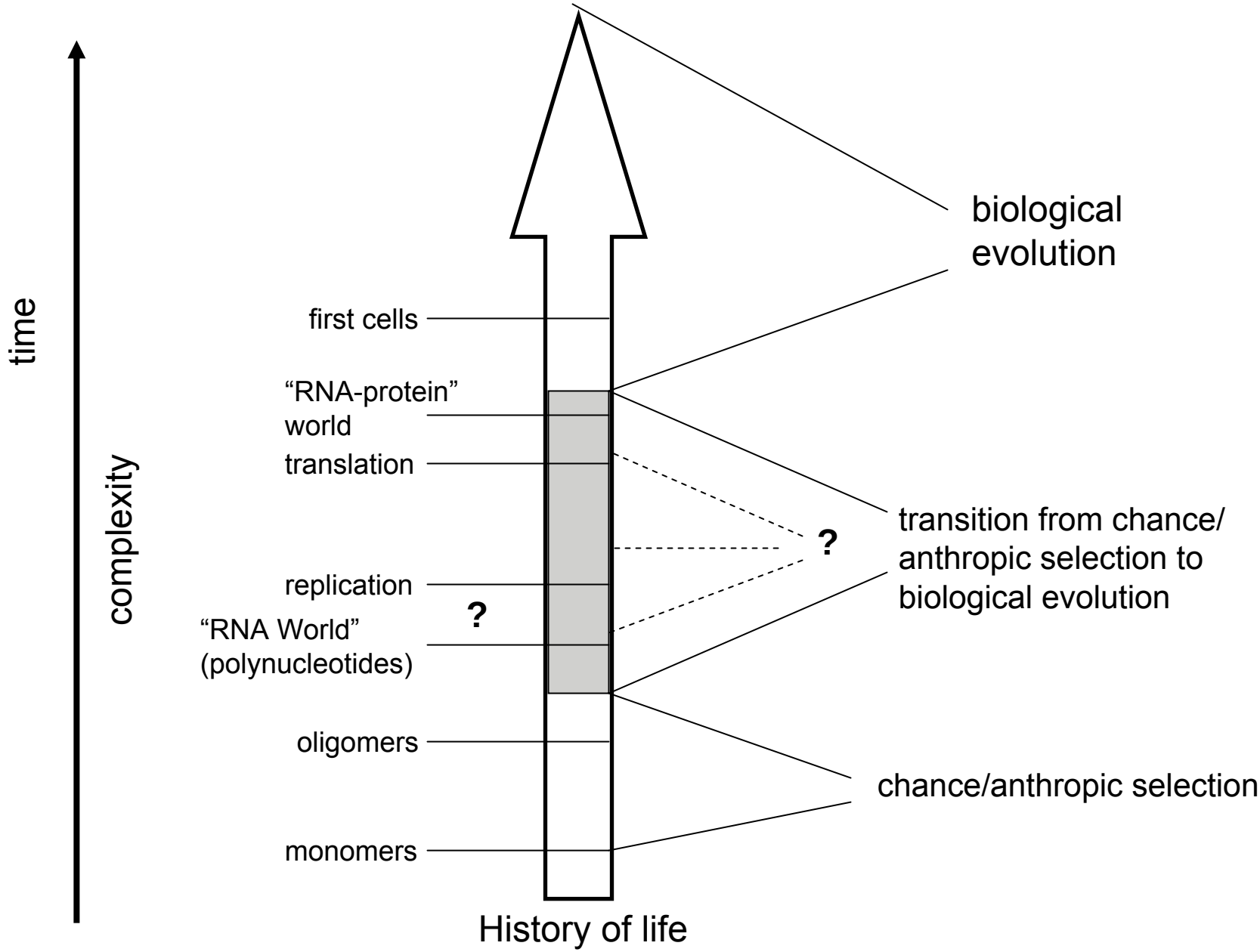
A radical alternative to Darwinian evolution of replication and translation inspired by MWO

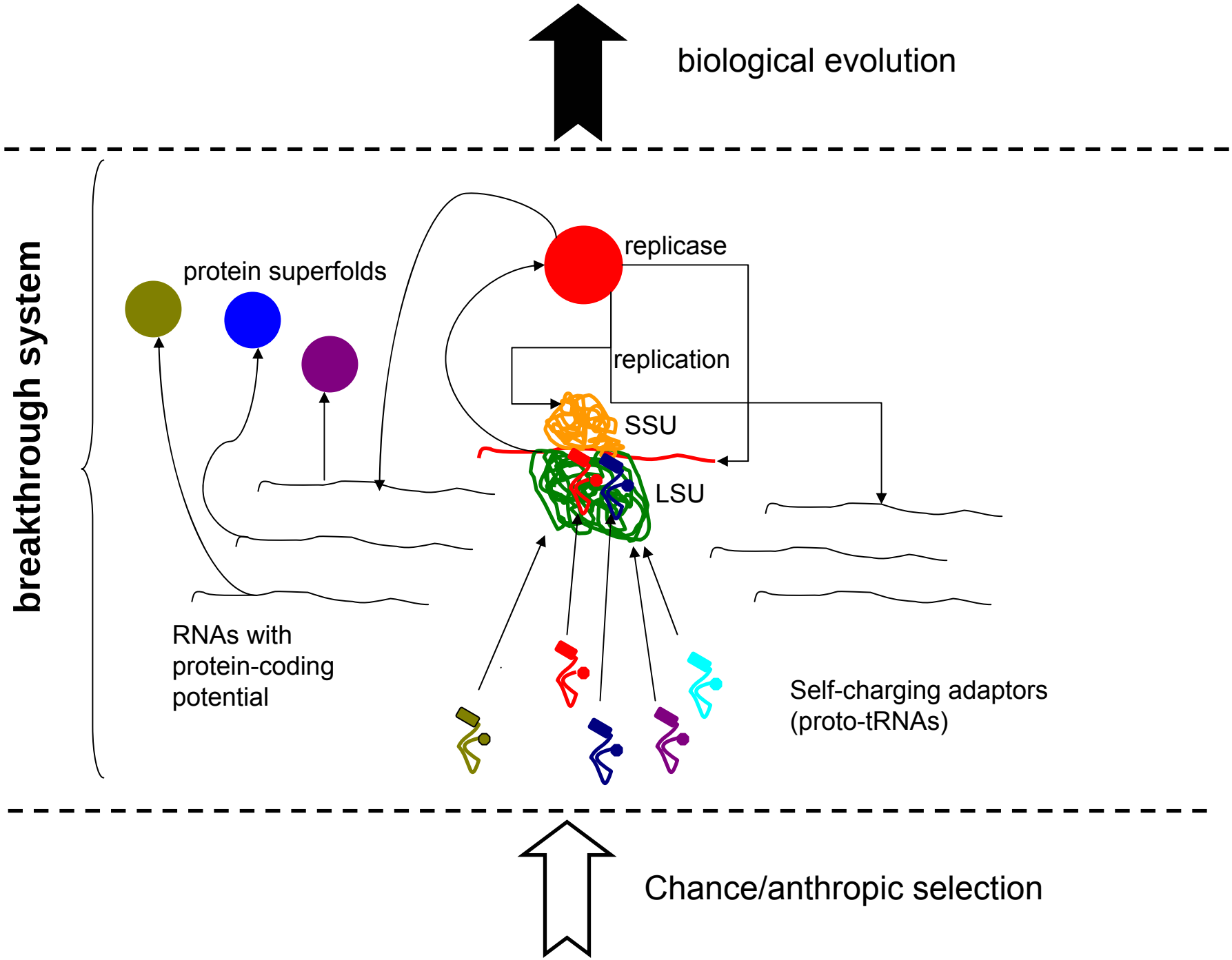
The universe is infinite
+
The number of macroscopic histories is finite

=Each physically admissible history is repeated an infinite number of times

=Emergence of an infinite number of copies of a vast variety of complex systems by chance is not just possible but inevitable

Transition from chance/anthropic selection to biological evolution



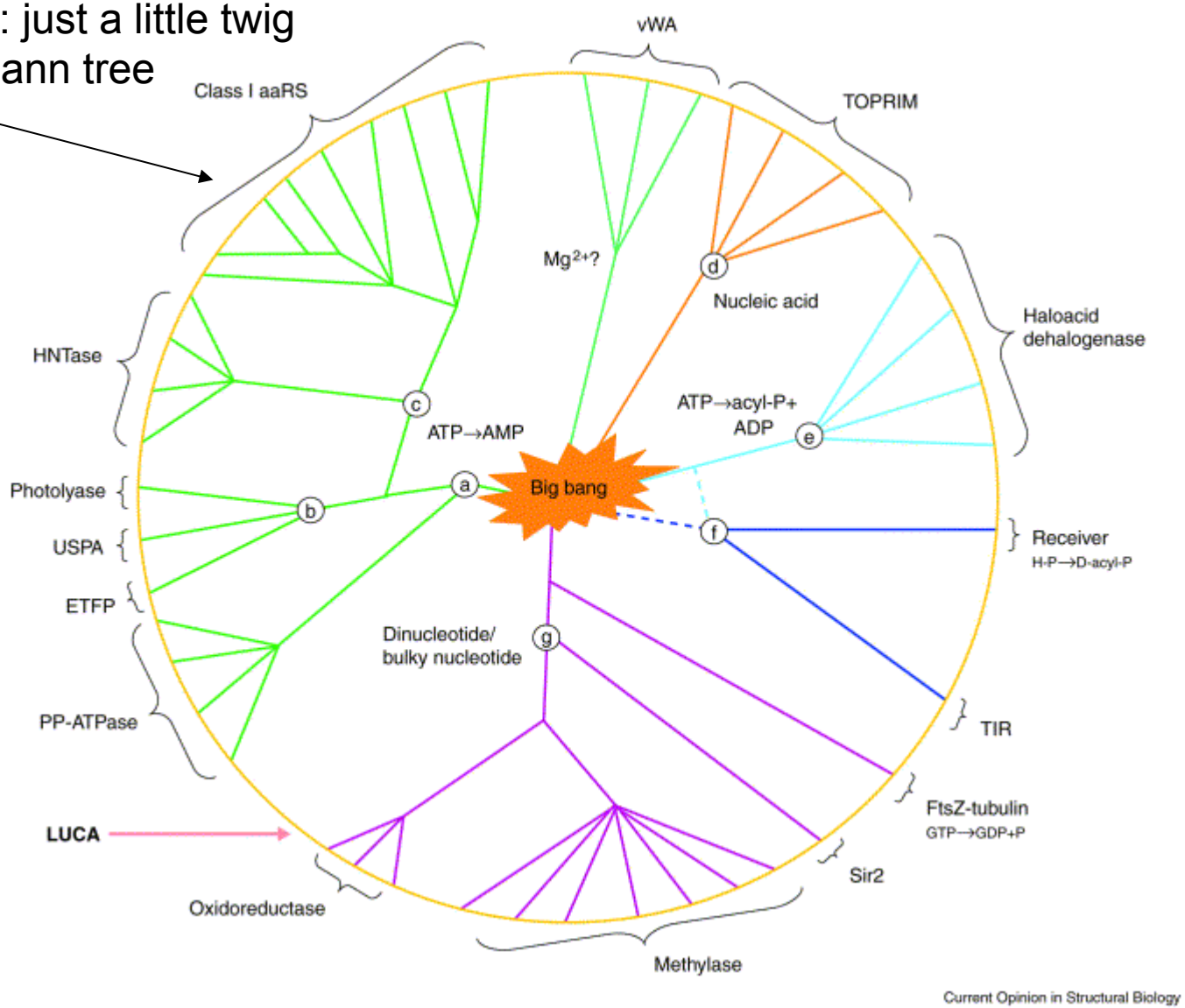

biological evolution
breakthrough system
protein superfolds
replicase
replication
SSU
LSU
RNAs with
protein-coding
potential
Self-charging adaptors
(proto-tRNAs)
Chance/anthropic selection

More on the breakthrough system

- Translation system shows clear signs of evolution by duplication/diversification
- Key proteins of the translation system (aaRS, translation factors) are late elaborations of the respective protein folds
- Hence the primordial translation system must have consisted, largely, of RNA
- The tRNAs also appear to have evolved by duplication, hence different, simpler adaptors in the primordial system

Phylogeny of Rossmann fold domains

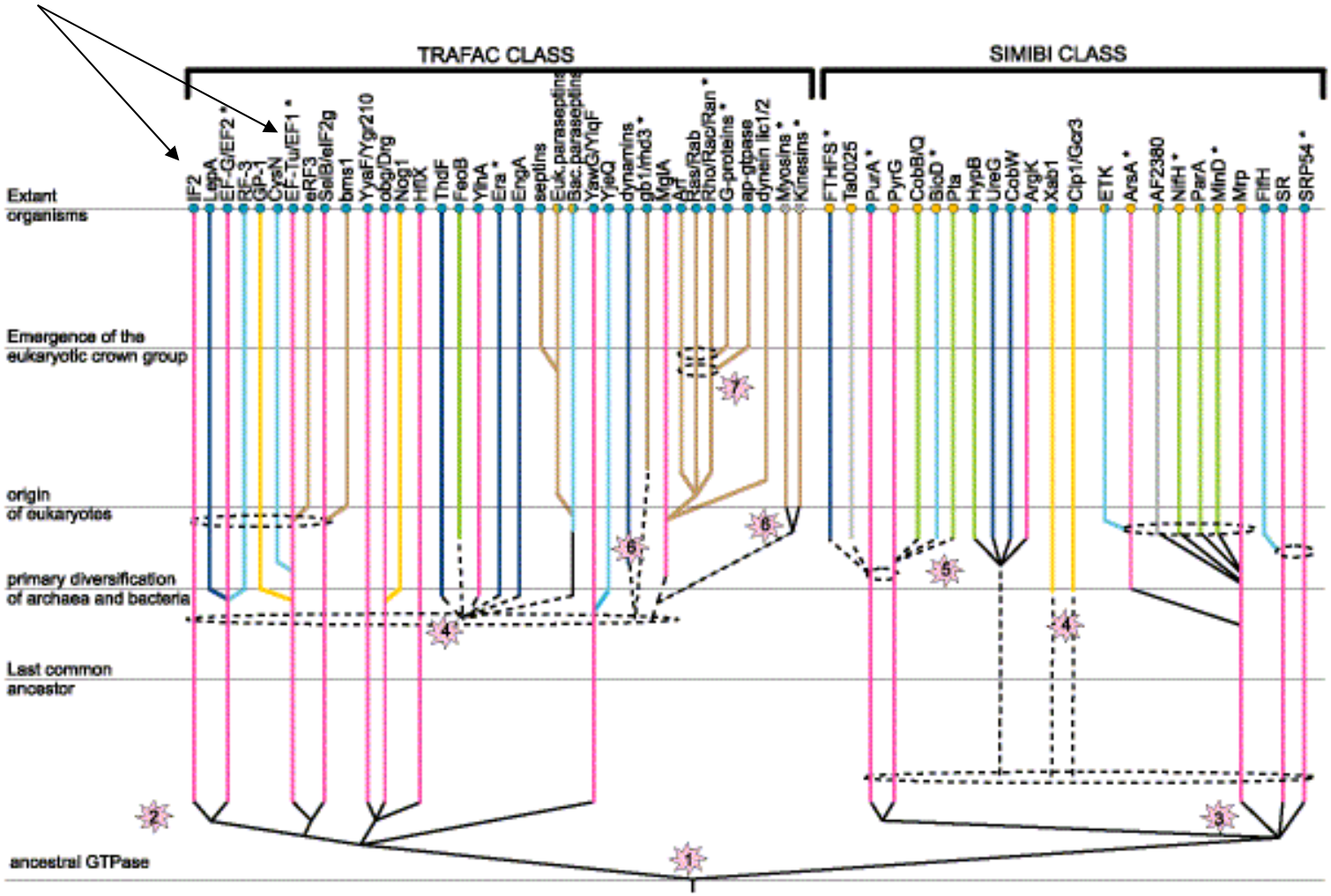
Class I aaRS: just a little twig of the Rossmann tree



Aravind L, Mazumder R, Vasudevan S, Koonin EV Curr Opin Struct Biol. 2002 Jun;12(3):392

Phylogeny of GTPases

Translation factors: scattered twigs



Leipe DD, Wolf YI, Koonin EV, Aravind L. J Mol Biol. 2002 Mar 15;317(1):41

The composition of the breakthrough system: *what* would have to emerge by chance/anthropic selection

- RNA-only proto-ribosomes
- ~20 primitive but specific adaptors activating amino acid in a ribozyme-catalyzed reaction – genetic code
- Translatable (m)RNAs for RNA-dependent RNA polymerase and prototypes of other major protein folds

The breakthrough system was much unlike the modern translation system:
A much greater role of RNA/ribozymes…though ***not*** an RNA World

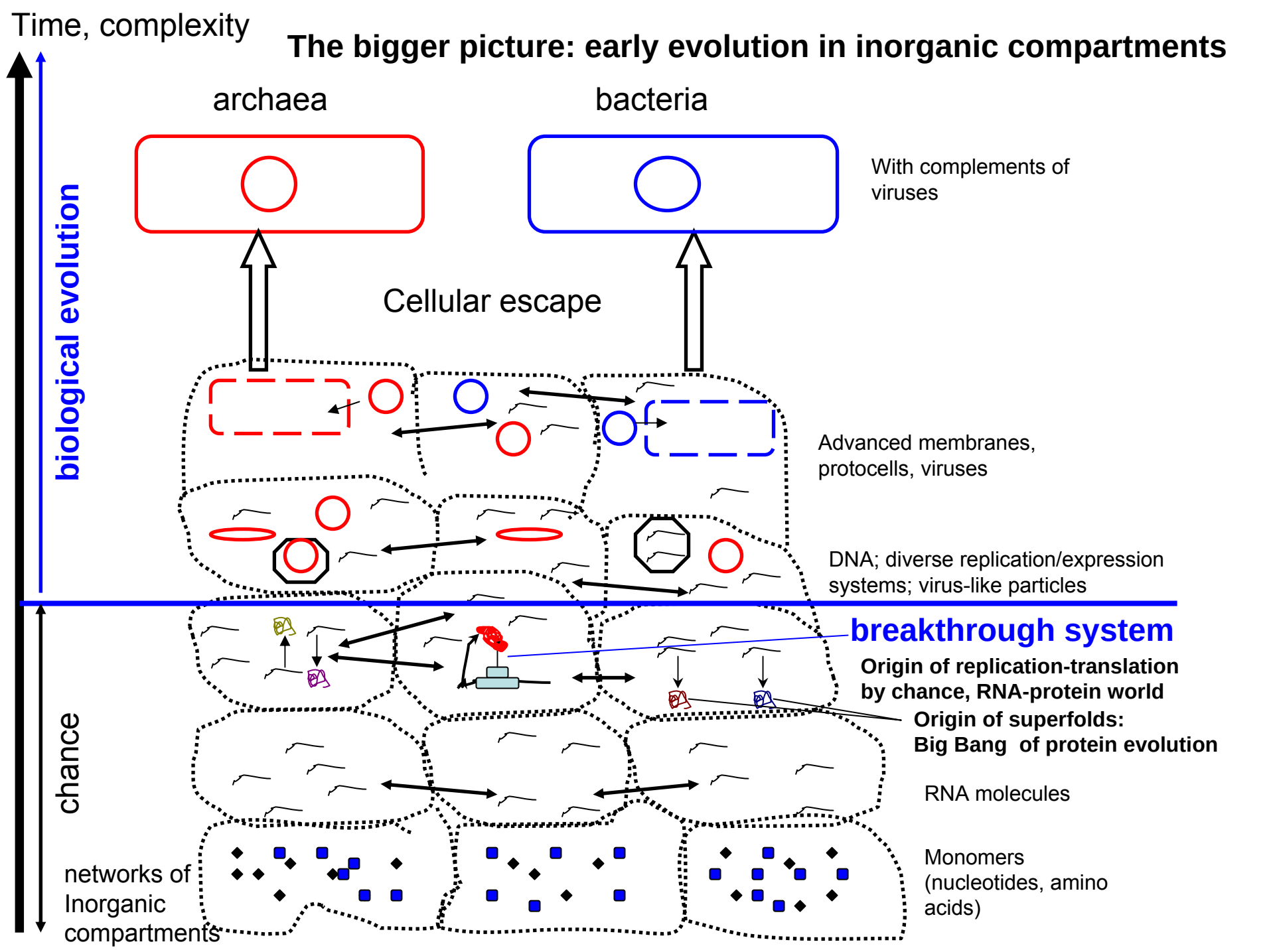
The origin and evolution of the genetic code

The code is highly robust/optimized and is usually thought to be a product of extensive evolution:

Freeland SJ, **Hurst LD**. The genetic code is one in a million.
J Mol Evol. 1998 Sep;47(3):238-48

We thus conclude not only that the natural genetic code is extremely efficient at minimizing the effects of errors, but also that its structure reflects biases in these errors, as might be expected were the code the product of selection.

However, an anthropic alternative is apparent:
If sufficient translation accuracy in the breakthrough system was unattainable with a code significantly less robust than the current one, this could have been selected anthropically, perhaps, with further evolutionary adjustment


Time, complexity
The bigger picture: early evolution in inorganic compartments
archaea
bacteria
With complements of viruses
biological evolution
Cellular escape
Advanced membranes, protocells, viruses
DNA; diverse replication/expression systems; virus-like particles
breakthrough system
Origin of replication-translation by chance, RNA-protein world
Origin of superfolds: Big Bang of protein evolution
chance
RNA molecules
Monomers (nucleotides, amino acids)
networks of Inorganic compartments

What about the RNA World alternative: emergence of a ribozyme replicase by chance/anthropic selection?

This cannot be ruled out – and might have happened in some universes

However:

No precedent: *bona fide* ribozyme replicase so far not known

More importantly: path to translation from the RNA world via biological evolution remains unclear

The similarities and the key distinction between the present hypothesis and (anthropically selected) RNA world:

- in both cases, anthropic selection yields complex RNA machinery
- however, under the present hypothesis, what is created is the machine and substrates for translation not for replication, hence no bona fide RNA World (replication is part of the definition)
- The OORT problem is solved in its entirety (unlike with RNA World)

A back of the envelope calculation…

Probabilities of the emergence, by chance, of different versions of the breakthrough system in an *O*-region: a naïve calculation of the upper bounds

General assumptions:

10^{22} starts, 10^{21} habitable, earth-size planets in an O-region
A 10 km thick habitable layer on a planet: $5x10^{24}$ cm3
RNA synthesis occurs in 1% of the volume – RNA-making "reactor" of $5x10^{22}$ cm3
RNA synthesis rate: 1 *n*-mer/cm^3/sec
Life time of an O-region and all planets: 10^{10} yrs = $3x10^{17}$ sec
Number of *n*-mers tried during the time after Big Bang: $S \approx 5x10^{22}x10^{21}x3x10^{17} \approx 1.5x10^{61}$
Number of unique *n*-mer sequences: $N=4^n=10^{0.6n}$
Expected number of instantiations of a unique *n*-mer in an O-region:
$E=S/N=1.5x10^{61}/10^{0.6n}$ **and** $n = \log(Ex1.5x10^{61})/0.6$

For *E*=1, *n*=102 (a gross over-estimate but highly robust) – notably, close to Eigen threshold for RNA World

Reasonable for a ribozyme replicase –
chance emergence in an O-region "imaginable"

Primitive coupled translation-replication system:

two rRNAs with a total size of at least 1000 nucleotides
~10 primitive adaptors of ~30 nucleotides each, in total, ~300 nucleotides
at least one RNA encoding a replicase, ~500 nucleotides (low bound)
is required.

n=1800, $E < 10^{-1018}$ ***Conceivable only under MWO or a similar model***

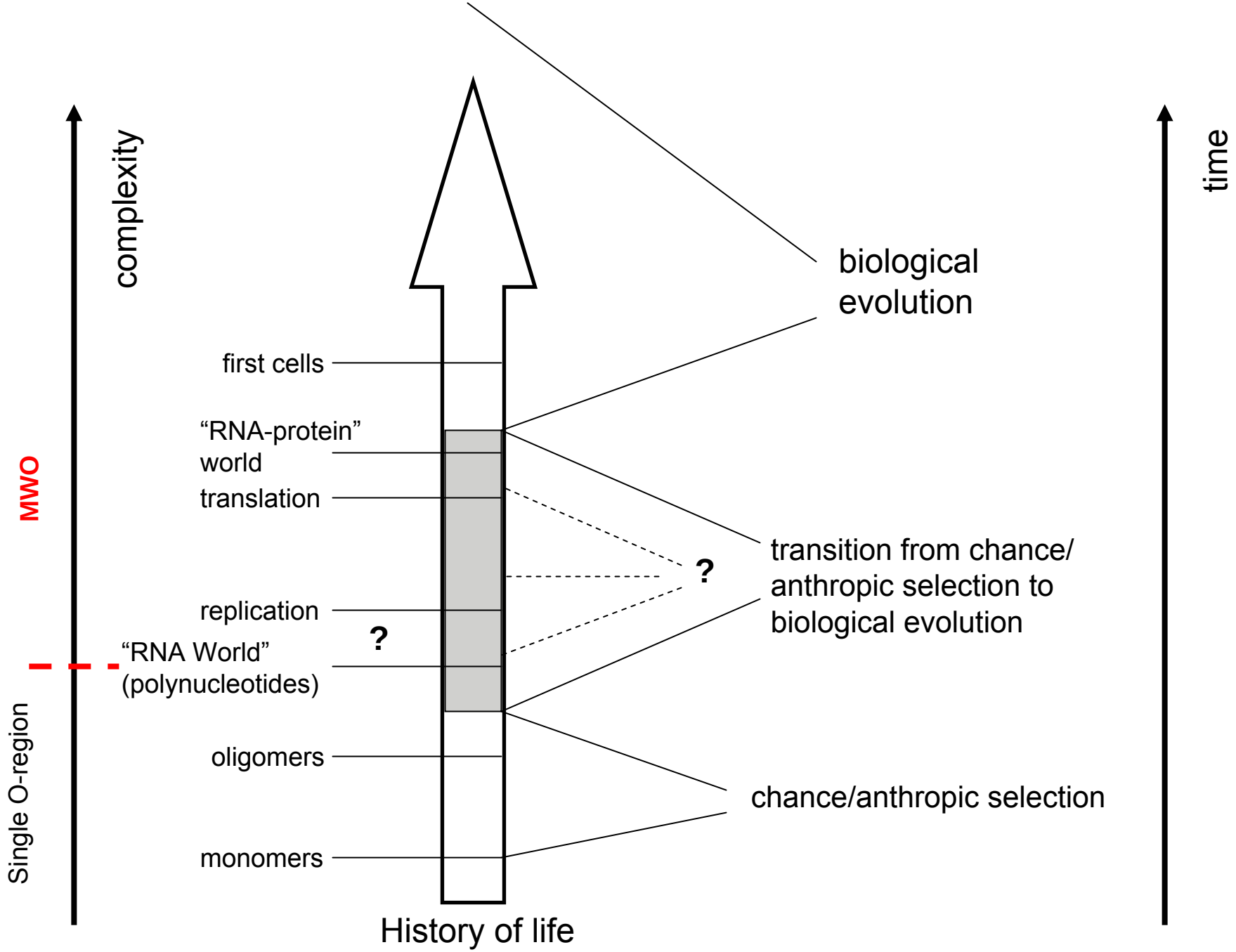

complexity
time
MWO
Single O-region
first cells
"RNA-protein" world
translation
replication
"RNA World" (polynucleotides)
oligomers
monomers
?
?
biological evolution
transition from chance/ anthropic selection to biological evolution
chance/anthropic selection
History of life

Fig. 1

Corollary: there might be a feedback from biology to cosmology

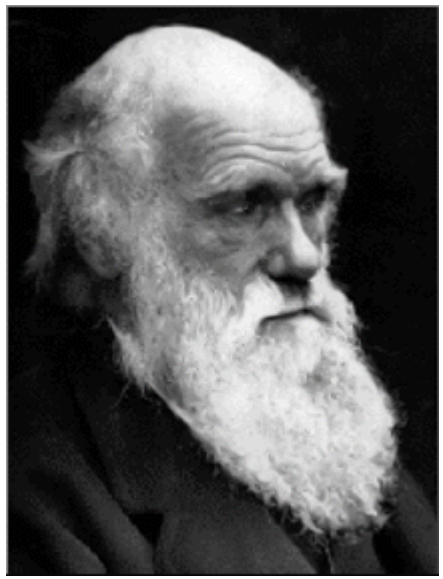
IF strong evidence is obtained in support of a highly complex breakthrough system…

Models of infinite universe like MWO get a boost through anthropic reasoning

A nightmarish thought…

Isn't ***anything*** (within laws of physics) possible under MWO?

If so, do we need Darwin at all?

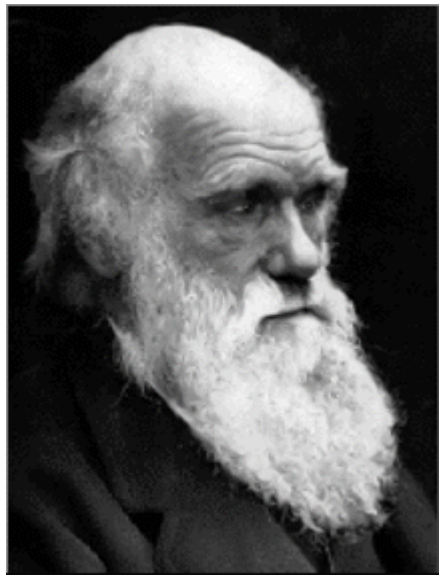


Pretty much anything does seem to be possible but this is beside the point…

…as soon as replication >Eigen threshold is achieved, Darwin rules

time

complexity



Darwinian selection

Threshold of biological evolution

Anthropic selection

Chance histories

Histories before and after the transition from chance to biological evolution

Another nightmare: the strong version of the present concept seems to propose chance emergence of a system designed for a specific function (translation)…

This seems to be a mere semantic trap: outside the biological context, the emerging translation system would ***not*** be functional, just complex, hence allowed by MWO and necessitated on earth by anthropic selection

Falsifiable predictions

- No way to evolve a coupled replication-translation system by biological evolution

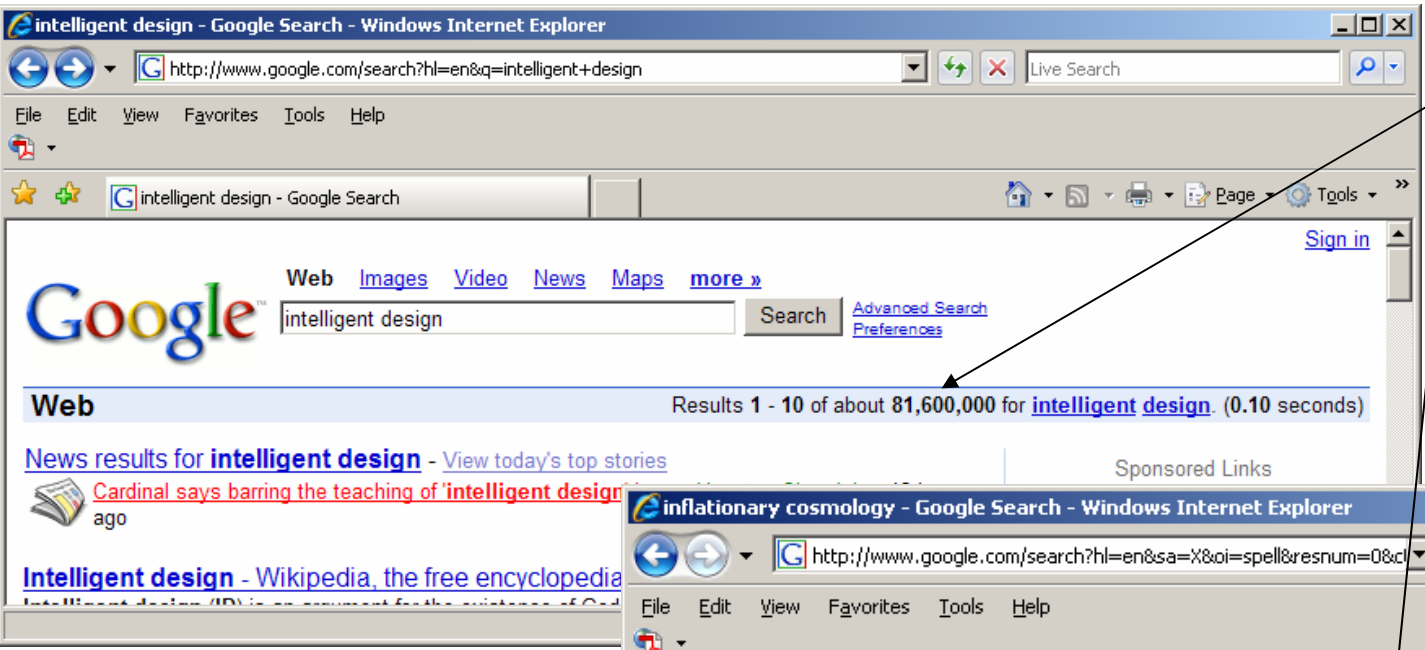
- *To falsify, develop a compelling (experimental or at least computational) evolutionary scenario*

- Life is **extremely** rare in the universe

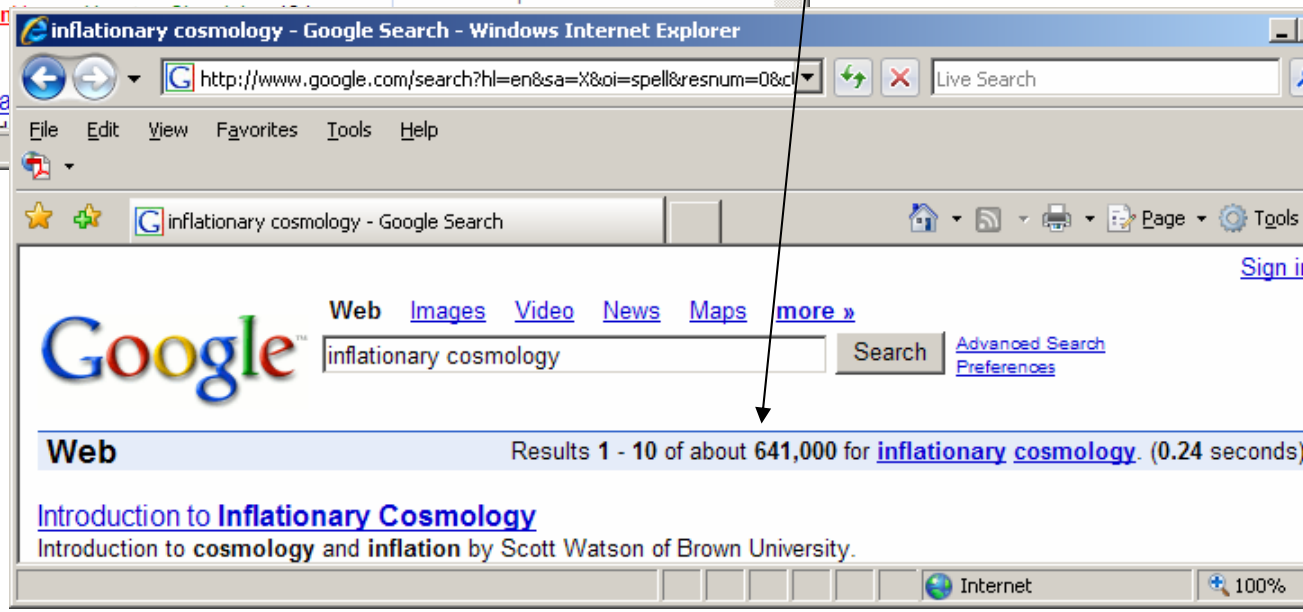
- *To falsify, discover independent life in 1-2 locations in our Galaxy*

- Converse prediction: any life forms that might be discovered on, say, Mars or Europa will have a common origin with Earth life

What about Intelligent Design?

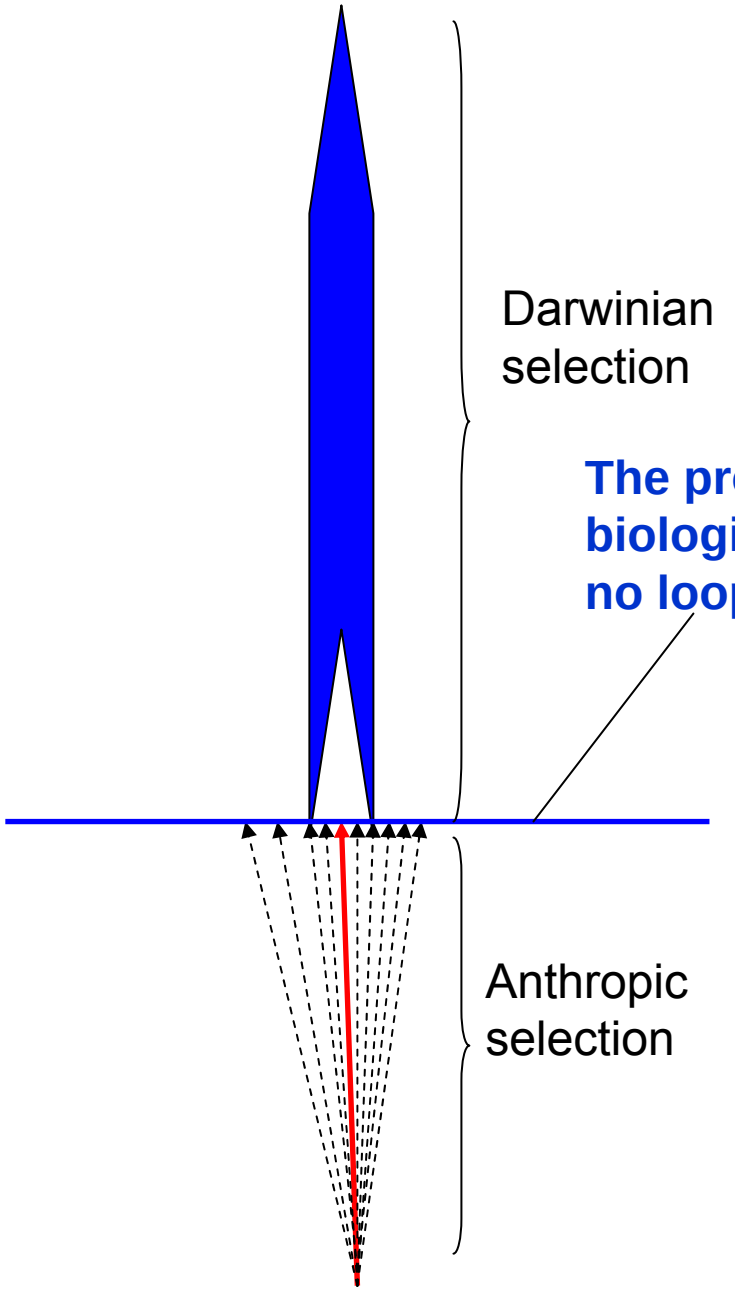


ID/IC >100



Science Declares Our Universe IS Intelligently Designed
Robert A. Herrmann Ph. D.
Professor of Mathematics
U. S. Naval Academy; 5 AUG 2000. Revised 19 JAN 2006
http://www.serve.com/herrmann/gidt.htm

What about Intelligent Design?



The proposed transition from anthropic selection to biological evolution at the breakthrough stage leaves no loophole for Intelligent Design

Conclusions

•The emergence of any phenomenon can be understood only within the framework of a more general, embedding system evolving in spacetime

•The Solar System might be sufficient as the framework for biological evolution

•However, for the origin of life, the framework is the evolution of the universe – cosmological models are directly relevant

•At least some of the inflationary cosmology models suggest that, in an infinite universe, origin of complex systems, such as a coupled replication-translation system, by chance/anthropic selection is possible

•This is distinct from trivial anthropic reasoning (the world is the way it is because, otherwise, we won't be here): under MWO, ***all histories are real***, the question is, which one is ours

•The strong version of this hypothesis is readily falsifiable. However, such falsification will not render eternal inflation/MWO irrelevant for the origin of life but just lower the threshold of biological evolution

Acknowledgments

Alex Vilenkin (Tufts) – cosmology reality check and discussion

Tania Senkevich (NIAID/NIH), Yuri Wolf, Valerian Dolja (Oregon State) – discussion

Yuri Wolf (NCBI, NIH) – “Conventional” modeling of the origin of translation

Bill Martin (Univ. Duesseldorf) – early evolution in inorganic compartments

L. Aravind (NCBI, NIH) – evolution of the translation system

Preliminary version: http://www.arxiv.org/abs/q-bio.PE/0701023