10 October 2019

MONOMERS, POLYMERS, THE LONG MOLECULES OF LIFE, & FROM POLYMERS TO LIFE

Homework #4 on WeBWorK – due 10/21 (after Fall Break) TI surveys!

Astronomy 106 | Fall 2019



10 October 2019

Monomers, polymers, the long molecules of life, & from polymers to life

1

Prebiotic molecules: delivered from the protoplanetary nebula or formed from the basics on infant planets? Origin of long, organic polymers, the predecessors of biomolecules that encode information Wet chemistry of nucelobases Nuances of polymerization Replication or mass production of nuceleic acids Transcription Codons Translation: the current fashion in protein manufacture The chicken-egg problem • Protein-based primitive life? • RNA world Emergence of the genetic code How long does all this take? The importance of baby steps and the heroism of time.

Is delivery necessary? Formation of prebiotic molecules *in situ*

It was long assumed that any prebiotic molecule would need to be formed in situ (in place, so on Earth) from simpler, inorganic molecules.

In the 1920s and 1930s, Oparin (USSR) and Haldane (UK) proposed that the early Earth would have had no prebiotic molecules and that a **reducing** atmosphere similar to the giant planets would give rise to them.

- Reducing = components of gas easily give up electrons. Examples include hydrogen, ammonia, methane.
- Opposite of oxidizing.

Miller and Urey (USA) reasoned that lightning and solar UV would provide the energy necessary for the neutral-neutral reactions and simulated these conditions in the lab.

10 October 2019

Astronomy 106 | Fall 2019

The original Miller-Urey experiment

A classic molecular biology experiment

- Ingredients: H₂, ammonia (NH₃), methane (CH₄) gas in 5-liter flask, connected by condenser to the "ocean," consisting of water in the 500 cc flask.
- A high-voltage discharge (a continuous spark) simulated lightning and provided ultraviolet light as well.
- Ran like this for one week. The experiment was repeated several times with new glassware and electrodes.

Result: After one week, 10-15% of the carbon had become incorporated into a wide variety of **prebiotic molecules** including several amino acids, sugars, lipids, and other organic acids.



10 October 2019

Compound	Yield	Compound	Yield
Glycine	270	Iminoacetic-propionic acid	6
Sarcosine	21	Lactic acid	133
Alanine	145	Formic acid	1000
N-methylalanine	4	Acetic acid	64
Beta-alanine	64	Propionic acid	56
Alpha-amino-n-butyric acid	21	Alpha-hydroxybutyric acid	21
Alpha-aminoisobutyric acid	0.4	Succinic acid	17
Aspartic acid	2	Urea	8
Glutamic acid	2	N-methyl urea	6
Iminodiacetic acid	66		

Miller 1953, Miller & Urey 1959

10 October 2019

Astronomy 106 | Fall 2019

Results of the original Miller-Urey experiment

Lots of formaldehyde (H_2CO) , hydrogen cyanide (HCN), cyanopolyyne (HC_3N) , and urea (NH_2CONH_2) formed in the gas and condensed into the "ocean."

Note that these are also seen in the ISM and/or meteorites.

The other molecules formed in the water with "wet chemical" reactions, among the formaldehyde, hydrogen cyanide, cyanopolyyne, and urea. A good example is the "Strecker synthesis" of the amino acids.

Modern "Miller-Urey" results

Alas, it is not so simple: it seems impossible that Earth's early atmosphere was *this* reducing.

- Gravity is insufficient to retain much hydrogen.
- Volcanism, heat, and solar UV light seem sure to have made the atmosphere oxidizing (opposite of reducing), producing lots of CO₂ and N₂, and vaporizing water. The atmosphere would have been more like that of Saturn's moon Titan.
- $\circ\,$ Formaldehyde, hydrogen cyanide, cyanopolyyne, and urea are not produced by reactions in a CO_2/N_2/H_2O atmosphere.
- This has moved attention to volcanic vents, which can produce locally high concentrations of reducing gases like methane (CH₄) and ammonia (NH₃), and where lightning naturally occurs.

As long as the precursors formaldehyde, hydrogen cyanide, cyanopolyyne, and urea are supplied, interesting molecules get formed in the "ocean."

10 October 2019

Modern Miller-Urey results

Hydrogen cyanide and ammonia have been shown to react with water to produce adenine, one of the **nucleobase** components of DNA and RNA (Oro 1961).

Two more nucleobases, cytosine and uracil, are made in water by reacting cyanopolyyne and urea (Robertson & Miller 1995).

The sugar ribose can be formed in high concentrations of formaldehyde: five formaldehyde molecules make one ribose (<u>Reid & Orgel 1967</u>). This works better in soil or clay than in water, though.

From these, entire **nucleotides** – combinations of nucleobases, ribose and phosphate from which RNA is built – form, all in the same "ocean" (<u>Powner et al. 2009</u>).



10 October 2019

M-U formation of nucleotides (McCollum 2013) Astronomy 106 | Fall 2019

Summary: synthesis of prebiotic molecules

Either by the slight reducing properties of the early atmosphere **or** by the reducing atmosphere near volcanic vents **or** by delivery via IDPs and meteorites of molecules formed in the reducing atmosphere of the ISM, it seems that the early Earth had supplies of formaldehyde, hydrogen cyanide, cyanopolyyne, and urea.

With the liquid water supplied by meteorites serving as the solvent, the primary prebiotic molecules – the amino acids and the nucleoside bases – can be made without much trouble or imagination.

The next step is the polymerization of these monomers: the generation of long, much more complex molecules that can encode information.

In water, important modifications are made to amino acids and nucleoside bases, owing to the **propensity of water to grab positive charges**.

When any acid dissolves in water, loosely-bound hydrogen nuclei attach themselves to water and leave their electrons behind. Example of hydrochloric acid:

 $HCI + H_2O \rightarrow CI^- + H_3O^+$ (in liquid water)

This would happen just as well in liquid ammonia, which shares this propensity with water:

 $HCI + NH_3 \rightarrow CI^- + NH_4^+$

(in liquid ammonia)

10 October 2019

Astronomy 106 | Fall 2019

Wet chemistry of amino acids

Amino acids express these tendencies on both of their ends: "amino" refers to the NH_2 group they all have at one end, and the H in the COOH organic-acid group they have at the other end is the one that comes off in the solution.

Example: the simplest amino acid, glycine – NH₂CH₂COOH



Note the color code, which we will use from here on.

The cool thing about them is that the NH₂ group on the end has the same propensity to grab positive ions as ammonia does in liquid.

10 October 2019

Astronomy 106 | Fall 2019



In an aqueous solution, the H in the COOH group is swiped by a water molecule...



This is a general and key property of amino acids in solution: there is a positive electrical charge localized on one end (the amino group) and a negative charge on the other end, with the molecule being electrically neutral.

These molecules are called **zwitterions** by biochemists. All amino acids are zwitterions when dissolved in water.

Zwitterions can thus undergo ion-molecule chemical reactions with each other, plus-end to minus-end. This is the same class of reactions as in molecular synthesis at low T in the interstellar medium.



10 October 2019

Astronomy 106 | Fall 2019

What is special about ion-molecule chemical reactions as compared to neutral-neutral (n-n) reactions?

Question

14

- A. They have no energy barrier, and the n-n reactions do.
- B. They are much rarer under most conditions because high degrees of ionization are rare.
- C. They heat their surroundings much less than n-n reactions do.
- D. They only work in the interstellar medium, and n-n reactions only work elsewhere.

10 October 2019

That the electrical charges are localized on these groups rather than shared among the atoms of the molecule as, for example, the electrons are among a benzene ring, means that the amino ends of amino acids will be electrically attracted to the acid ends of other molecules.



It is as if each amino acid has hook and eye connectors on opposite ends.

10 October 2019

Astronomy 106 | Fall 2019

Wet chemistry of amino acids

The electrostatic attraction between hook and eye can lead to chemical bonds, creating a longer zwitterion and a spare water molecule. Example of two glycines: $2[(NH_3)^+CH_2COO^-] \rightarrow (NH_3)^+CH_2CONHCH_2COO^- + H_2O$



10 October 2019

Astronomy 106 | Fall 2019

18

19

Wet chemistry of amino acids

Or, using our hook and eye shorthand, and noting that a spare water molecule is generated every time a hook and eye connect:



Obviously, this can be repeated ad nauseam, and it does not need to be the same amino acid every time. A long-chain polymer has been created from amino-acid monomers.



A polymer of amino acids is generally called a protein (or a polypeptide).

10 October 2019

Astronomy 106 | Fall 2019

Other monomers: wet chemistry of ribose

Ribose $(C_5H_{10}O_5)$ is a sugar that is made in the "ocean," or even better on dry land, by formaldehyde either delivered or made by Miller-Urey means.

Some of the dangling OH groups are good at making bonds in much the same way as the dangling OH in glycine. A shorthand schematic:

Other monomers: phosphoric acid

Phosphoric acid (H₃PO₄) can also be made in electrical discharges in a reducing atmosphere if there is any phosphorous around. (There is never much.)

In solution, $H_2O + H_3PO_4 \rightarrow H_3O^+ + H_2PO_4^-$

The two dangling OH groups of $H_2PO_4^-$ (phosphate ion) are good at making bonds with the "eyelet" groups on sugars in much the same way that the "hook" groups do in amino acids. The phosphate ion hook plus sugar eye lead to a bond that, like a peptide bond, releases a water molecule.

10 October 2019

What do zwitterions have to do with the polymerization of amino acids (AAs)? Select all that apply.

- A. AAs are zwitterions in the interstellar medium.
- B. AAs are zwitterions in solution.
- C. Electrostatic attraction guarantees that NH₃ groups of two AAs will match up.
- D. Electrostatic attraction guarantees that NH₃ groups will match up with OH groups.
- E. Peptide bonds form between two AAs when an NH₃ group encounters an OH group.
- F. Peptide bonds form between two AAs when their NH₃ groups encounter each other.

10 October 2019

Astronomy 106 | Fall 2019

Question

Other monomers: wet chemistry of nucleobases

A particularly interesting family of molecules made in the Miller-Urey ocean are the **nucleobases**, which divide into two categories: **purines** and **pyrimidines**. Five of these are of particular importance.

- Purines: adenine (C_5H_5N_5) and guanine (C_5H_5N_5O), planar molecules each containing two CN rings

These two purines are essential in all Earth life forms. Another purine, caffeine, is essential to
most of the life forms in this room. ☺

10 October 2019

Astronomy 106 | Fall 2019

Wet chemistry of nucleobases: pyrimidines

Pyrimidines: cytosine (C₄H₅N₃O), thymine (C₅H₆N₂O₂), and uracil (C₄H₄N₂O₂), also planar molecules. Their structure, in order (C-T-U):

- These are also essential to Earthly life forms: thymine in DNA, uracil in RNA, cytosine in both.
- The sizes of these five molecules match each other in a special way that promote **hydrogen bonding**.

10 October 2019

Astronomy 106 | Fall 2019

Wet chemistry of nucleobases: hydrogen bonds

Hydrogen bonds are weaker than the usual (covalent) chemical bonds: they are the manifestation of the attraction of an atom (like O and N) that is good at stealing hydrogen, for the hydrogen atoms are covalently bound to other atoms.

The essential purines and pyrimidines match up in pairs with particularly strong hydrogen bonds: adenine with thymine or uracil... Hydrogen bonds

10 October 2019

Astronomy 106 | Fall 2019

Wet chemistry of nucleobases: H bonds

...and guanine with cytosine

It will be useful to have a schematic shorthand for these monomers, too, to reflect their key-like ability to bond with each other:

Note that A-T/U and G-C pairs are precisely the same length, 1.09 nm between the NH groups. All other abundant purine-pyrimidine pairs have quite different lengths.

10 October 2019

Astronomy 106 | Fall 2019

Wet chemistry of nucleobases

Each nucleobase has an NH group that is good at combining with a dangling OH group on ribose, making a bond rather like a peptide bond and releasing yet another water molecule. The combination is called a **nucleoside**.

Hanging a phosphate group on the sugar (releasing a water molecule, of course) produces an interesting monomer called a **nucleotide**. Note that they can be chained together, hook to eye, making a sugarphosphate backbone with nucleobases sticking out.

10 October 2019

Astronomy 106 | Fall 2019

Nucleotide polymers

Nucleotides can polymerize into long chains, like amino acids can polymerize into proteins.

Owing to the use of ribose and nucleobases, and the hook-and-eye ends, this particular kind of polymer is called a ribonucleic acid (RNA).

10 October 2019

Astronomy 106 | Fall 2019

27

Nature "makes" choices

There are several chemically equivalent ways to polymerize amino acids and nucleotides, but only one for each prevailed in Earthly life.

Why? The leading explanation is that those were the ones easiest to replicate, and **natural selection** therefore favored them: they were made consistently, and thus outnumbered the ones that were randomly made.

How? That is debated, but several experiments and theories provide explanations.

10 October 2019

Isomers & chirality

Isomers: Molecules with identical chemical contents but different atomic arrangements

Isomers can be classified as either lefthanded (levo, L) or right-handed (dextro, D), depending on the specific arrangement of the atoms; this is known as chirality.

Discovered by Biot, who also discovered that meteorites fall from the sky.

10 October 2019

Astronomy 106 | Fall 2019

Isomers & chirality

All amino acids in Earthly life are L-type. (Except glycine, which is achiral.) We are not sure why.

- Chemical reactions (like the Miller-Urey experiment) produce an equal amount of L and D (called a racemic mixture).
 - The ISM probably does as well.
- There is an excess of L-type amino acids found in meteorites.
 - Certain minerals in meteorites have a helical form.
 - UV circularly polarized light preferentially destroys either L- or D-type molecules, depending on its polarization.

10 October 2019

Astronomy 106 | Fall 2019

30

31

Chirality & the alpha helix

A protein of all L or all D amino acids forms a spiral structure called an **alpha helix**, with the alpha Cs (the ones between the NH₂ and COOH groups) and Ns on the inside, the side groups on the outside, and hydrogen bonds between the C=Os and the NH groups four amino acids down the chain.

Ls make a right-hand spiral; Ds would make a left-hand spiral.

A racemic mix of amino acids could not make such a spiral.

СООН

(H)

NH₂

COOH

NH₂

CI

Alanine is a chiral amino acid.

10 October 2019

Nucleotides – binding to ribose

Binding to ribose: Carbons 1', 2', 3', and 5' look identical, but a long straight polymer chain is only possible if a nucleobase bonds to 1' and phosphates to 3' and 5'.

Not sure how it originally effected, but it was found by <u>Leslie Orgel</u> that, when in a solution, double-oxidation-state ions like Zn⁺⁺ catalyze polymerization of long chains (as long as 50) in the correct alignment.

10 October 2019

Sugars are also isomers

The five important nucleobases and phosphoric acid do not exhibit isomerization, but sugars do.

Sugars used in Earthly life are all D-type: not just the ribose that forms the backbone of RNA, but also the other sugars like glucose that play roles in energy and metabolism of organisms.

10 October 2019

33

Nucleotide polymerization

Astronomy 106 | Fall 2019

10 October 2019

10 October 2019

 DNA
 FNA

 • CHa
 •

DNA

When deoxyribose is the sugar, thymine is substituted for uracil; otherwise, the same bases are used.

The polymer that uses deoxyribose is called a deoxyribonucleic acid (DNA).

Both DNA and RNA have helical structures, dictated by the twists in their D-sugar backbones.

35

Geometry of nucleobase pairs

Bonds to the sugars are made by the NH groups on opposite sides of the hydrogen-bonding sites.

In adenine-thymine or guanine-cytosine pairs, these sugar-bonding groups are precisely 1.09 nm apart.

The five essential nucleobases are the only ones this size. We cannot substitute others, or else DNA and RNA would not replicate...

10 October 2019

Astronomy 106 | Fall 2019

Replication of nucleic acid

Requirement (#3) of life is replication. Can proteins and nucleic acids replicate, or at least mass-produce?

 If they can, this could explain the emergence of classes of proteins and nucleic acids as components, the assembly of which could serve more complex, life-like roles.

Perhaps the easiest to envision is for short strands of RNA. Suppose a short RNA lies in a solution containing nucleobases, ribose, and phosphoric acid.

- A complementary polymer can form by hydrogen-bonding nucleotides onto the RNA's nucleobases, hooking up the phosphates and sugars, and then severing the hydrogen bonds.
- The complementary polymer can do the same, thus replicating the original.
- Hydrogen bonds are much weaker than covalent bonds. The "copies" can be stripped off without harming the original chain.
- Other polymers are capable of "unzipping" the H bonds in DNA. Once unzipped, both sides can bond to nucleotides in the solution just like RNA, and thus the DNA replicates: bases in same order as original.

10 October 2019

Astronomy 106 | Fall 2019

Replication of RNA

10 October 2019

Astronomy 106 | Fall 2019

38

Transcription

Long chains of DNA can be partially unzipped, capture a complementary chain of nucleotides, have this new chain of RNA zip off, and have the DNA zip back up. This form of partial replication is called **transcription**.

Unzipping does not frequently happen spontaneously, except when the solution temperature experiences small increases.

- Certain polymers with the right molecules in the right spacing can do it, though.
- Among Earth life forms, the unzipping and rezipping of DNA is done with a protein we call RNA polymerase.
- This is an example of an **enzyme**: proteins that catalyze chemical reactions which means that they do not get chemically changed in the process, like the dust grains that catalyze the formation of molecular hydrogen.
- Replications of DNA and RNA, and transcriptions, require in the current age a special class of proteins.

10 October 2019

Astronomy 106 | Fall 2019

Animation of transcription

10 October 2019

Astronomy 106 | Fall 2019

40

Codons

Three-nucleotide sequences have a special significance, and thus their own name: codon.

Because there are four different nucleobases used in either nucleic acid, there are

 $4 \times 4 \times 4 = 64$

different codons.

As we will see, this has significance for the encoding of information used by modern organisms to build proteins.

- For example, there are 20 amino acids* used in human proteins, and the base sequences that indicate the beginnings and ends of codon sequences. 20 + 2 = 22 < 64, so three-base codons work.
- *Note: There are really 21 amino acids used in human proteins, but the 21st (selenocysteine) is relatively rare in proteins and uses a special signal in addition to the Stop codon.

10 October 2019

Astronomy 106 | Fall 2019

Codons – the genetic code

Astronomy 106 | Fall 2019

10 October 2019

Instead of 20, suppose Earthly organisms used only 10 different amino acids. How many nucleotides could be used in a codon?

Question

A. 1

- B. 2
- C. 3
- D. 4
- E. Any number greater than 1

Replication of proteins

This has been envisaged in several different ways...

Long chains of amino acids can polymerize in solution without much help.

Some configurations might be **self-organizing**: certain sequences have higher binding energies than others and will preferentially form.

There will be lots of copies of these for this reason.

Alternative mass production or replication: perhaps formation on a sequenced structure, like a crystal or another molecule, that preferentially attracts the monomers in a certain order?

10 October 2019

Astronomy 106 | Fall 2019

44

Translation

In modern lifeforms, this is quite complicated and intimately involves nucleic acids following a transcription.

Section of DNA or RNA transcribed onto a short length of RNA, called **messenger RNA** (mRNA).

mRNA is transported to a ribosome, a collection of proteins and nucleic acids that will hydrogen-bond onto a certain codon of the mRNA.

The ribosome has shorter stretches of RNA around, called **transfer RNA (tRNA)**, that have exposed codons and, bonded elsewhere, an amino acid captured from the solution that is unique to the exposed codon.

10 October 2019

Protein replication

The tRNA complementary to the first codon on the mRNA H-bonds to it, and the AA is the first element of the protein. The next element is brought in with the next tRNA, and so on.

The ribosome's (and tRNA's) job is called translation.

https://youtu.be/2BwWavExcFI?t= <u>2m48s</u>

†RNA

10 October 2019

The original chicken and egg problem

In modern Earthly life, the replication or mass production of nucleic acids and protein are interdependent: each needs certain of the others in order to be produced in large quantities.

Which came first? Was the most primitive life originally just a protein or nucleic acid?

How did they come to be interdependent?

How did useful proteins and nucleic acids develop?

10 October 2019

Astronomy 106 | Fall 2019

Protein-based primitive life?

There is no experimental evidence that protein-only life could exist, but there are theories and candidate substances.

Closest to modern life: peptide nucleic acids, which are proteins but which have nucleobases that allow replication and information storage à la RNA.

So far, peptide nucleic acids are a lab curiosity; no evidence in substructure of cells or chemistry of primitive monomers that it is a viable life form.

• They can hydrogen-bind to DNA, though.

10 October 2019

Astronomy 106 | Fall 2019

Protein-based primitive life?

Prions do not count: they reproduce by folding existing proteins into weird shapes, and in general seem to be things that developed after life was already on its way with both nucleic acids and proteins.

Before the genetic code was traced to DNA, the originator of the idea behind the Miller-Urey experiment, Alexander Oparin (1924), noted the tendency of amino acids to combine and fatty acids (lipids) to form in solution, and postulated that this would lead to spherical **coacervate** droplets of these compounds in the "primordial pond."

He thought that this would lead to cells: membranes made from lipids, most of the rest proteins.

Similarly, Fox later proposed **proteinoid** droplets (protein-like molecules which form abiotically from amino acids).

10 October 2019

Astronomy 106 | Fall 2019

Protein-based primitive life? – Dyson

Freeman Dyson (<u>1982</u>, <u>1999</u>), a famous physicist turned evolutionary biologist, picked up on Oparin's scheme and conjured a model in which metabolism and proteins were primary, with nucleic acids and the genetic code developing much later.

- "Cells" first, enzymes next, nucleic acids much later.
- Cells provide sites for absorption of monomers.
 - He had acervates/proteinoids and amino acids in mind, but noted that the reasoning would work on crystals and nucleotides as well.
- Repeatable assembly of complex polymers through "order-disorder" transition of absorbed molecules.
- Dyson's model predicts that polymers with 100-200 monomers can be made repeatedly.
 - About the minimum required for a reproducing system, according to experiments on viruses (and thus RNA, not proteins).
- Works best if there are 8-10 different monomers; does not work at all if there are 3-4, so the mechanism prefers amino acids to nucleotides.
- Also achieves an error rate of about 1 in 4 if the total number of monomers present is of order 10,000 (fewer errors if more monomers).
- Hard to extend to nucleic acids, though, so this model has not found much favor.

10 October 2019

Astronomy 106 | Fall 2019

Nucleic-acid based primitive life?

RNA can

- Replicate, as we have seen (transcription)
- Catalyze the formation of proteins (tRNA + RNA) and nucleic acids.
- Store information: the genetic code, so that specific proteins can be mass-produced.
- Transmit information: mRNA copies of short sections of a master RNA (translation).

This is the most consistent, and therefore popular, story of emergence of biology from chemistry: an RNA-dominated primitive form of life that preceded proteins and DNA.

This stage of evolution is generally called the RNA world.

10 October 2019

Astronomy 106 | Fall 2019

52

RNA world

Scenario, due mainly to Eigen, Orgel, and Cech:

- In the beginning, there were short strands of bare, self-replicating RNA (Eigen et al. 1981). Eigen calls the varieties of the strands quasispecies.
- Most numerous and robust quasispecies predominate (appearance of natural selection).
- Quasispecies mutate and therefore diversify as time goes on.
- Truncated stretches of these RNAs catalyze reactions of other RNAs (Cech & Bass 1986).
- Example: 192-base RNA that catalyzes the replication of an RNA as long as itself with 98.9% accuracy.
- D-sugars "chosen" by now.
- These RNA catalysts, called ribozymes, presage the function of modern ribosomes. (1989 Nobel Prize to Cech and Altman for discovering ribozymes.)
- Soon, the ribozymes begin to catalyze protein formation. L-amino acids "chosen" at this point.
- In turn, some of the proteins thus formed begin to influence RNA cleaving and hydrolysis; these
 are the first enzymes.
- This interchange of RNA and protein catalysts is termed a hypercycle by Eigen.

10 October 2019

RNA world

Simulations and theories show that proteins made from other quasispecies would be likely to perform novel catalytic functions on a given quasispecies.

RNA can hydrogen-bond to complementary DNA. Eventually, DNA took over as the more robust and stable master copy of the base sequences that encode the formation of proteins, but RNA retained its role in transcription and translation.

10 October 2019

Ribozymes in action on RNA (<u>Cech 2004</u>)

Astronomy 106 | Fall 2019

The genetic code

One of the most striking adaptations of ribozymes is the development of **tRNA**: rather long molecules, each with a three-base sequence on one end which uniquely matches to an amino acid on the other.

As we have mentioned, life on Earth has 20° amino acids in use, which could not be specified by two bases (only 16 possible combinations) but can be with three (64 possible combinations).

This code must have been established very early in the development of life, as even cell-organelle DNA and RNA differ very little from nuclear DNA in the code they use.

10 October 2019

Astronomy 106 | Fall 2019

	Nucleotide 2				
		U	С	Α	G
ľ		UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine
	п	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine
	U	UUA Leucine	UCA Serine	UAA Ochre (Stop)	UGA Opal (Stop)
		UUG Leucine	UCG Serine	UAG Amber (Stop)	UGG Tryptophan
le 1 (5')		CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine
	c	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine
	C	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine
€		CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine
le		AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine
Š	•	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine
_	A AUA Isol	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine
		AUG Methionine	ACG Threonine	AAG Lysine	AGG Arginine
		GUU Valine	GCU Alanine	GAU Aspartic acid	GGU Glycine
¢	C	GUC Valine	GCC Alanine	GAC Aspartic acid	GGC Glycine
	G	GUA Valine	GCA Alanine	GAA Glutamic acid	GGA Glycine
		GUG Valine	GCG Alanine	GAG Glutamic acid	GGG Glycine

The genetic code

56

Amino acid	Codons		
Alanine	GCU, GCC, GCA, GCG		
Arginine	CGU, CGC, CGA, CGG, AGA, AGG		
Asparagine	AAU, AAC		
Aspartic acid	GAU, GAC		
Cysteine	UGU, UGC		
Glutamine	CAA, GAG		
Glutamic acid	GAA, GAG		
Glycine	GGU, GGC, GGA, GGG		
Histidine	CAU, CAC		
Isoleucine	AUU, AUC, AUA		
Start	AUG		
Leucine	UUA, UUG, CUU, CUC, CUA, CUG		
Lysine	AAA, AAG		
Methionine	AUG		
Phenylalanine	υυυ, υυς		
Proline	CCU, CCC, CCA, CCG		
Serine	UCU, UCC, UCA, UCG, AGU, AGC		
Threonine	ACU, ACC, ACA, ACG		
Tryptophan	UGG		
Tyrosine	UAU, UAC		
Valine	GUU, GUC, GUA, GUG Also used to code	;	
Stop	UAA, UGA, UAG seleno-cysteine		

The genetic code

Baby steps

Every step in the development of life has involved the development, not just of molecules of a certain size and class, but of molecules that *do* something specific. That is a big restriction.

For example, take the minimum-length RNA-replication ribozyme mentioned earlier: 192 nucleotides.

There are four different nucleotides, so the number of combinations of 192 nucleotides is $n = 4 \times 4 \times 4 \times K \times 4 = 4^{192} \cong 4 \times 10^{115}$

192 factors of 4

If they form randomly, we may have to try 4×10^{115} times for a good chance of getting the correct one. (Why? Well...)

10 October 2019

Astronomy 106 | Fall 2019

58

Question! Quick education in probability...

Flip a coin. What is the probability of it coming up heads?

- A. 1/2
- B. 1/20
- C. 1
- D. 2/3
- E. Cannot tell until you flip the coin.

What is the probability of the coin turning up heads in ten consecutive tosses?

Question!

A.
$$\frac{1}{2} = 0.5$$

B. $\frac{1}{10} = 0.1$
C. $\left(\frac{1}{2}\right)^{10} = 9.8 \times 10^{-4}$
D. $\left(\frac{1}{2}\right)^{\frac{1}{10}} = 0.933$
E. $\left(\frac{1}{10}\right)^2 = 0.01$
F. $\left(\frac{1}{10}\right)^{\frac{1}{2}} = 0.316$

10 October 2019

So how many trials of ten coin tosses would you think you would have to make to have a good chance of ten consecutive "heads" at some point?

Question!

A. 2 B. 10 C. 100 D. $2^{10} = 1024$ E. $2^{100} = 1.3 \times 10^{30}$

10 October 2019

Astronomy 106 | Fall 2019

In total, how many coin tosses would that be?

Question!

- A. 20
- B. **40**
- C. 10240
- D. 1.3x10³
- E. Cannot estimate from the given information.

10 October 2019

Astronomy 106 | Fall 2019

Time to get the right RNA randomly (years)

Baby steps

If that many attempts were needed to have a reasonable chance of getting the correct molecule, **making them out** of single nucleotides one by one, it would take *much* longer than the age of the Universe.

But it takes much less time to make shorter RNAs. And that, plus natural selection, is the key: make short RNAs first, then make longer RNAs out of these.

10 October 2019

Baby steps

For example, consider a two-step process.

Start with water at T = 300 K that has the Earth-crust abundance of carbon, 10% of it in nucleotides.

- Then it takes about 0.04 seconds to form an RNA with 24 nucleotides and 0.32 seconds to form one with 192, adding nucleotides one by one.
- To have a good chance to get a specific RNA with 24 nucleotides by this means, it takes $4^{24} \times 0.04 \text{ s} = 3.6 \times 10^5$ years, while it takes $4^{192} \times 0.32 \text{ s} = 4 \times 10^{107}$ years to get a specific 192-long RNA.

Clearly, 24-nucleotide RNAs are going to form the slow way much faster than the 192s; also, at least some of them can **self-replicate**, as we have already discussed.

```
10 October 2019
```

Astronomy 106 | Fall 2019

Question

How many different four-nucleotide-long RNA molecules are there?

А.	4
Β.	8
C.	16
D.	32
E.	64
F.	128
G.	256

Astronomy 106 | Fall 2019

About how many times would you have to construct 4-base RNAs at random to have a good chance of getting a specific one?

Question

A.	4
Β.	8
C.	16
D.	32
E.	64
F.	128
G.	256

10 October 2019

Astronomy 106 | Fall 2019

6

In the worst case, you would have to do this serially (one RNA at a time). If it takes 0.05 s to synthesize a 4-base RNA, how long does it take to have a good chance of synthesizing a specific one?

Question

10 October 2019

Baby steps

Suppose that once certain 24-long RNAs formed, self-replication led quickly ($\ll 10^5$ years) to the incorporation of all the nucleotides around into chains of 24.

• It gives about 1.6 s to make a chain of 192 out of chains of 24, given the above conditions.

Furthermore, let us assume that there are 24 different ones that have risen to prominence by replication.

Then it will take $24^{\left(\frac{192}{24}\right)} \times 1.6 \text{ s} = 5000$ years to have a good chance of making a specific 192-long RNA like our "RNA replicase" ribozyme. The total time is therefore less than a million years, instead of greater than 10^{107} years.

• And thus much less than the age of the Solar System.

· Goes even faster with more, shorter steps, as we will see.

10 October 2019

Astronomy 106 | Fall 2019

69

Probability math simplified

Two pieces of information to figure out:

- Length of molecule (number of blanks)
- How many options you have for each blank

10 October 2019

Baby steps

It would be nice to start with primordial soup and see the evolution of ribozymes produced in the lab in this manner, but unfortunately there are few graduate students willing to wait 5000 years to get their PhDs.

• Existing ribozymes have been "intelligently designed" by the likes of Cech and Altman, rather than evolved from a mixture of prebiotic chemicals.

Meanwhile, there are **computer simulations** based on lab data:

 Ma et al. (2007, 2010a): starting with a prebiotic mixture and allowing the basic sequences to be 5-10 nucleotides long (instead of the 24 we used a minute ago), "RNA replicase" is generated in only 52 years.

Astronomy 106 | Fall 2019

70

cott Lab, UCSC

Baby steps

Also in the simulations: once the first kind of RNA replicase is created, more efficient, *mutated* replicases begin to arise.

In similar computer simulations, Ma et al. (2010b) have also shown that ribozymes provided with a protective cell-like membrane soon (within only a few years!) produce ribozymes which can replicate membrane-component molecules.

• Emergence of cell self-replication?

and the second s 2000 1500 0 r of re 1000 0 Jumber 500 0 (200000) 12 (a) Time (Monte Carlo Steps) x 10 and the second of the second o 400 of replicases 300 00 200 0 Number 100 \cap 0œ 2 3 4 5 6 7 8 9 10 11 12 Time (Monte Carlo Steps) x 10 (c)

Computer simulation of RNA-replicase evolution, using 10-nucleotide (upper) and 6-nucleotide (lower) basic sequences. A "Monte Carlo step" is approximately 0.38 hours (Ma et al. 2007).

10 October 2019

Astronomy 106 | Fall 2019

Time is still the hero

A sophisticated biochemical function like the "RNA replicase" ribozyme takes less than one million years to produce from scratch, via chemistry and biological **natural selection**.

Of course, there are lots of functions to evolve, but all the functions of the molecules of life did not need to develop serially.

Thousands of millions of years (i.e. Gyr), very roughly, seem reasonable for making many functions of this level of sophistication. Complex simulations bear this out.

Still a long time, but it is much shorter than the time available and is consistent with the fossil record of primitive life on Earth.

10 October 2019