Blue Marble Space

The Center for Chemical Evolution

Research • Education • Outreach

The Origin of Life: What We Do and Don’t Know

Jim Cleaves

Blue Marble Space Institute for Science

Image courtesy of the American Museum of Natural History

How Do We Think Life Began on Earth? A View from 38 Years Ago…

"It must be admitted from the beginning that we do not know how life began. It is generally believed that a variety of processes led to the formation of simple organic compounds on the primitive Earth. These compounds combined together to give more and more complex structures until one was formed that could be called living.

No one should be satisfied with an explanation as general as this."

S.L. Miller and L.E. Orgel, The Origins of Life on Earth, 1974

NSCORT Exobiology

Modern Historical Background

Oparin postulated the self-organization of environmentally supplied compounds produced in an environment different from the modern one (*e.g.* with a different atmosphere); this was an outgrowth of 19^{th} c. thinking on the problem.

A.I. Oparin
1894-1980

H.C. Urey
1893-1981

S.L. Miller
1930-2007

Urey extended Oparin's model to ideas corroborated by mid-20^{th} c. conceptions of the origin of the solar system, backed up discoveries in isotope geochemistry.

Miller's electric discharge experiment provided an experimental validation of the possible link between the two concepts.

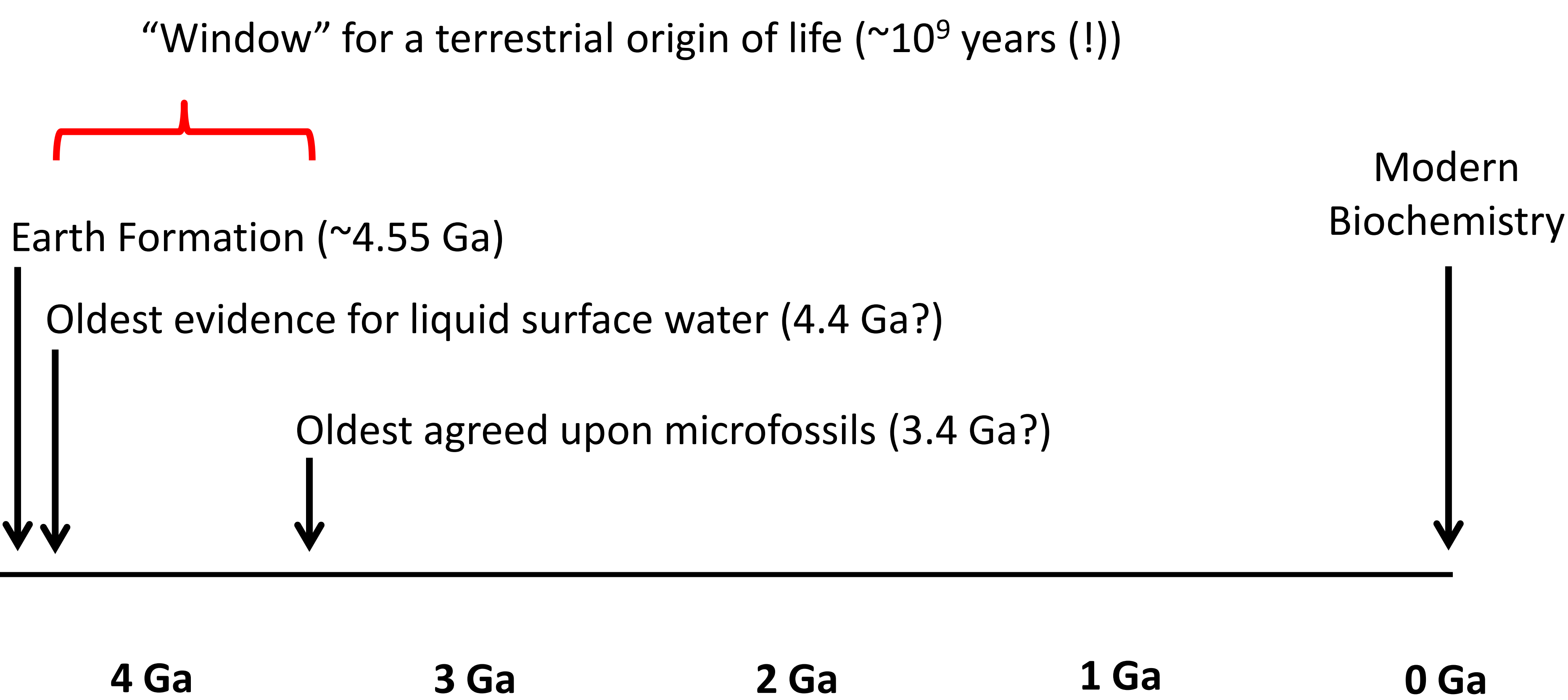
The Origin of Biochemistry (as a Scientific Field)

First amino acid discovered in **1806** (asparagine, interestingly threonine not until 1936!).
Proteins were not identified as such until **1838**; the **role of amino acids in forming them** was not understood until the **1860's**.
Role of proteins as catalysts was not clear until **1926** (urease), **first sequence** solved in **1955**, **first 3-D structure** solved in **1958**.

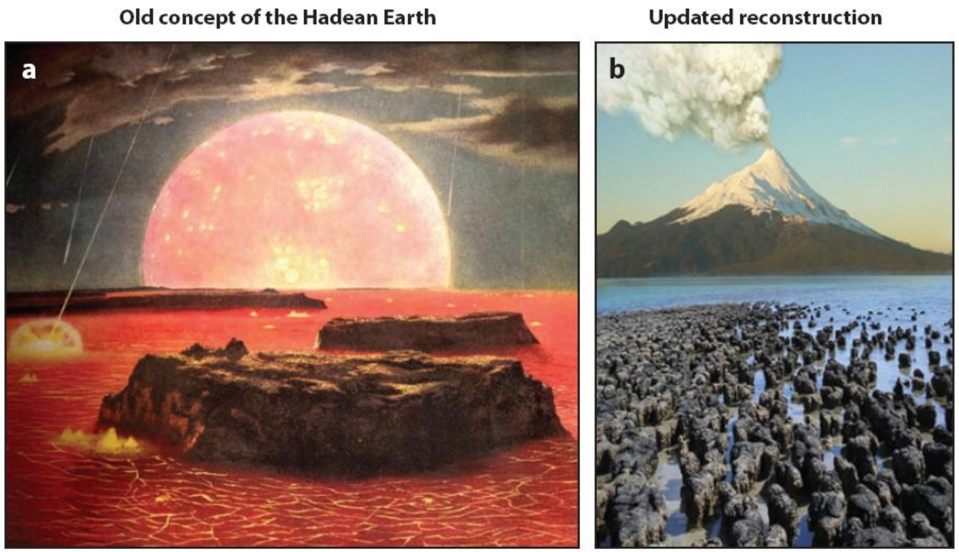
DNA was not discovered until **1869** (isolation) **1878** (purification).
The **nucleobases were isolated** from DNA in this period (*e.g.* G **1862**, A **1885**, C **1894**)
DNA structure was not solved until **1953**, the **genetic code** was not elucidated until **1961-1966**.

Yet we *still* do not completely understand how a cell functions,
much less how it came to be.

When Do We Think Life Began on Earth?



Changing Geological Constraints?



~3.43 Ga "Fossil" Stromatolite

~Modern Stromatitic Microbial Colonies Shark's Bay, Australia

The Earth <u>*may*</u> have cooled to "habitable temperatures" earlier than previously thought.

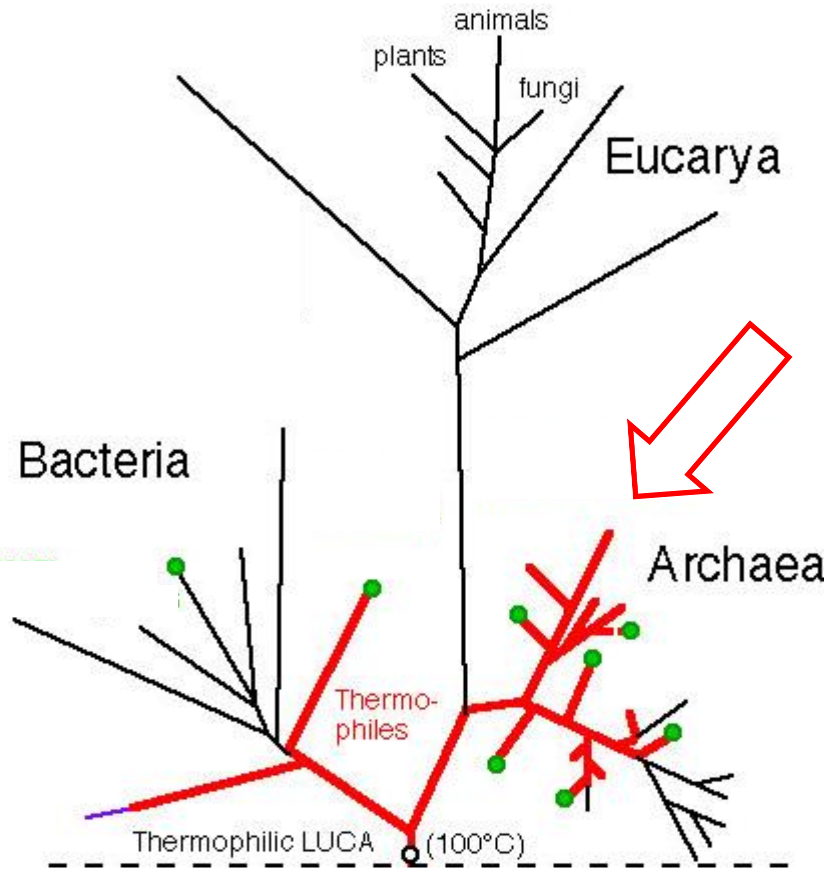
From Arndt & Nisbet (2012) *Annu. Rev. Earth Planet. Sci.* **40**:521–49

"…the Strelley Pool Formation suggests that microbial mat communities ***probably*** existed 3.43 Ga…inhabited by diverse microbial communities, ***possibly*** *including photoautotrophs*." Allwood *et. al.* (2009) ***PNAS*** **106**:9548-9555

It is likely that these were **not the first cells**, but are merely **the oldest surviving fossil evidence**.

LUCA and the Tree of Life

All extant organisms can be placed on a phylogenetic tree according to the degree of sequence similarity of their 16S rRNA:



Thermophilic prokaryotes cluster near the presumed base of this tree, which may suggest a thermophilic LUCA, and *perhaps* a thermophilic OOL.

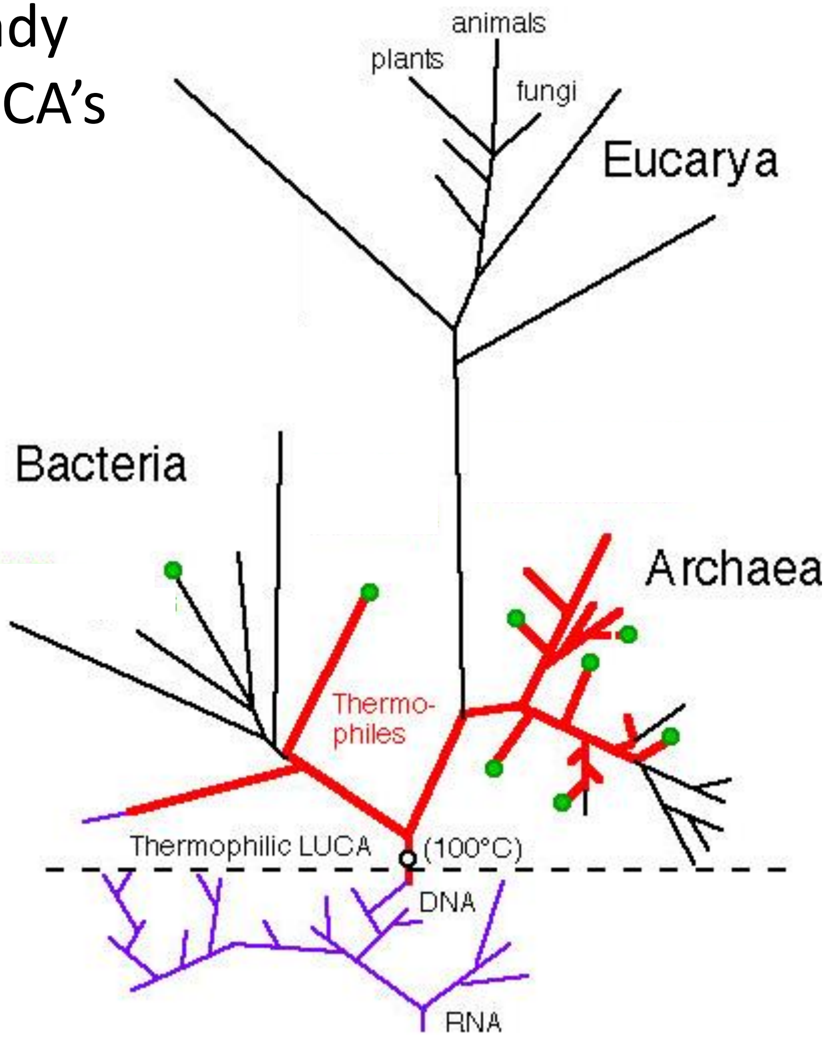
The Nature of LUCA: the “Last Universal Common Ancestor”

- Water-based cytoplasm surrounded and enclosed by a **lipid bilayer membrane**.
- (Na^+) lower, (K^+) higher, inside than outside. Gradients maintained by **protein** ion pumps.
- Cells multiplied by duplicating contents followed by cell division.
- **Genetic code based on DNA**
  - three-nucleotide codons, using essentially the **modern genetic code** and compliment of 20 **L-amino acids**.
  - **dsDNA copied by a protein template-dependent DNA polymerase**.
  - **DNA** integrity maintained by **protein** enzymes such as DNA topoisomerase, DNA ligase and others.
  - **DNA protected by DNA-binding proteins such as histones**.
- **RNA** produced by a **protein DNA**-dependent RNA polymerase.
- **Used essentially modern translation apparatus for protein synthesis.**
- **ATP** was used as an energy intermediate.
- **Several hundred protein enzymes** catalyzed chemical reactions.
- (Penny & Poole (1999) *Current Opinion in Genetics & Development* **9**:672–677)

- **Essentially a “Modern” Prokaryote!**

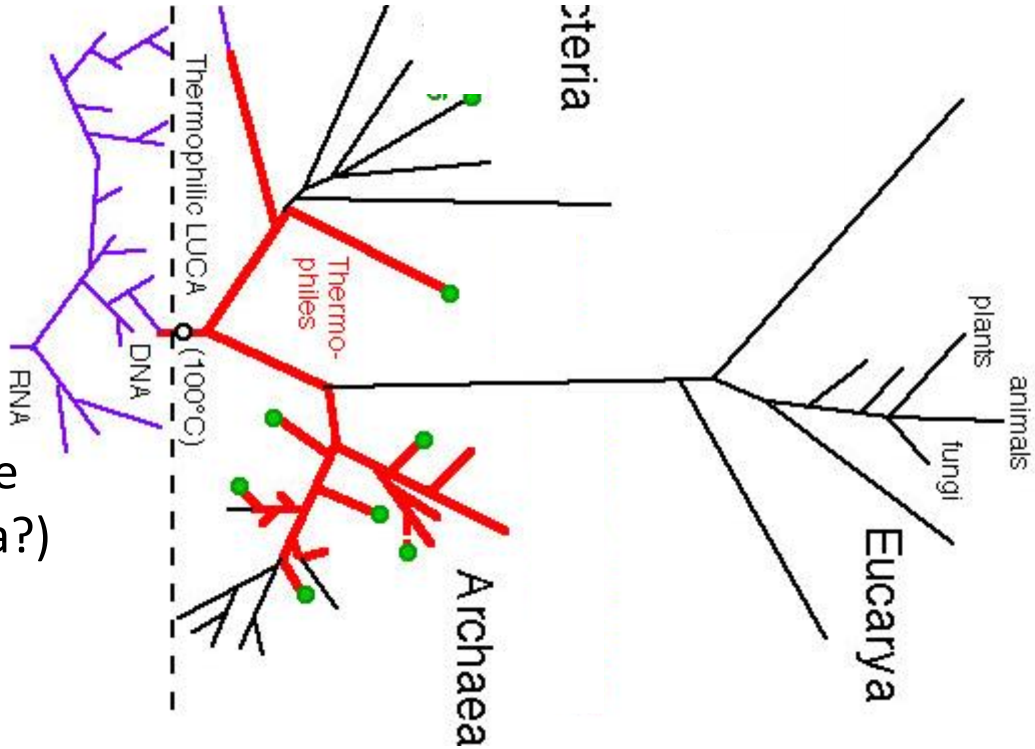
The Tree of Life

If LUCA was already a rather sophisticated and highly evolved "organism", this suggests a considerable amount of biochemical evolution had already occurred before LUCA's appearance.



What came before LUCA? An RNA-based organism? Something else?

Calibrating Phylogenetic Trees with the Geologic Record



Liquid surface water (4.4 Ga?)

Modern Biochemistry

Oldest microfossils (3.4 Ga?)

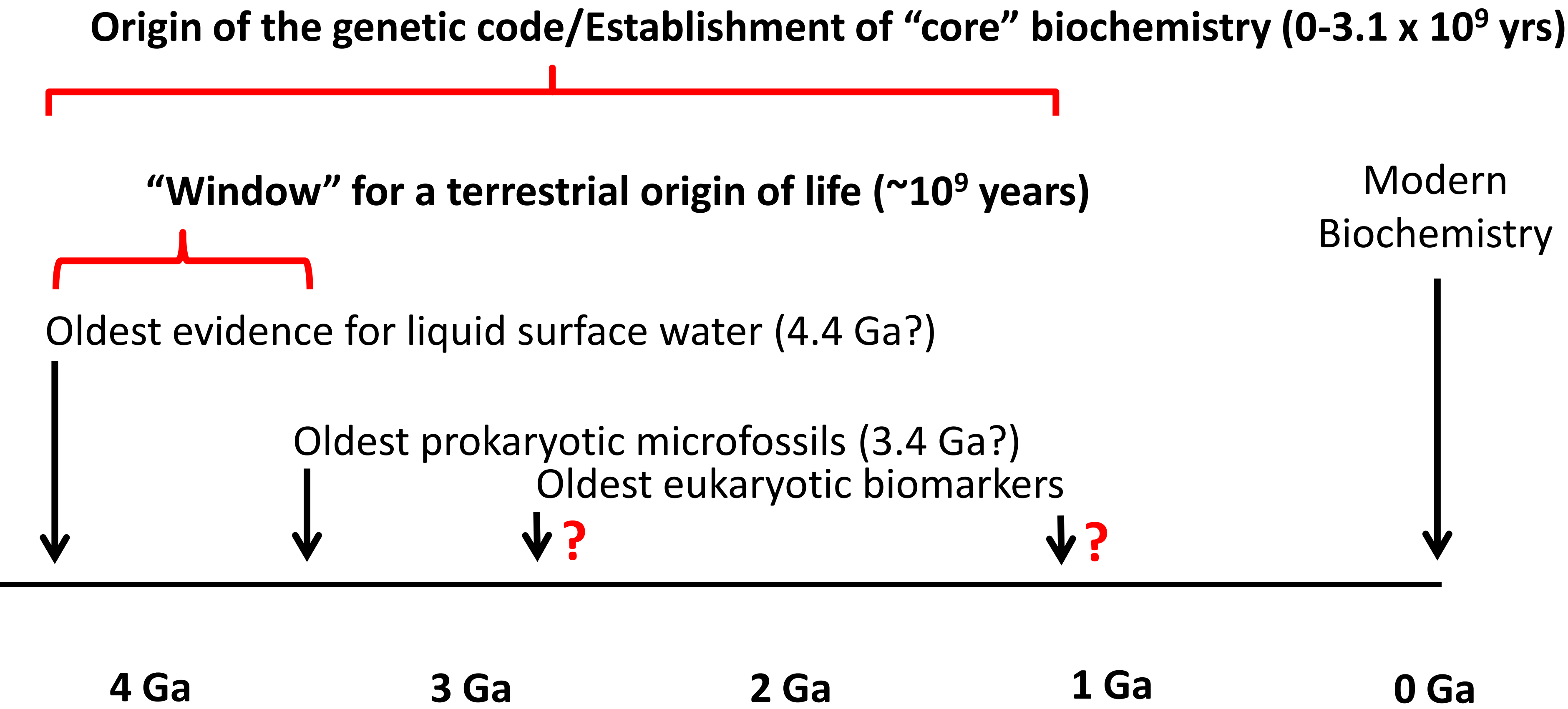
Oldest eukaryotic molecular biomarkers

? ?

4 Ga **3 Ga** **2 Ga** **1 Ga** **0 Ga**

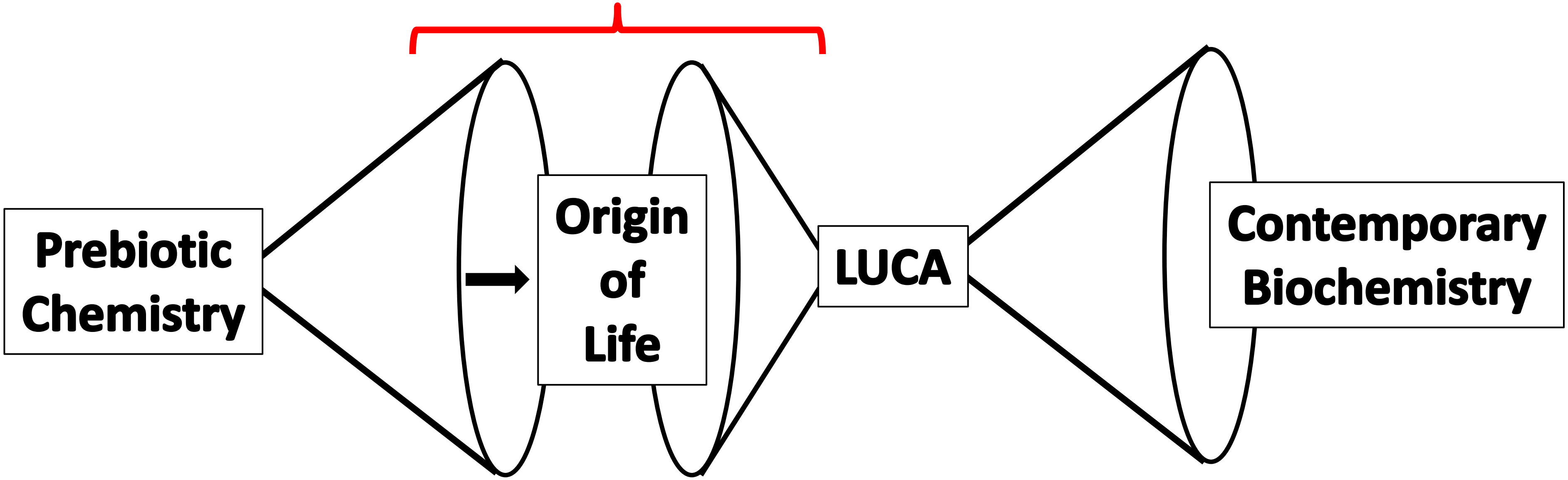
The Very Large Temporal Window for OOL and pre-LUCA Evolution

The genetic code is one of 1.5×10^{84} possible triplet codes: it is extremely unlikely that it sprang into existence without undergoing some degree of natural selection.



General Scheme for the Origin and Evolution of Life

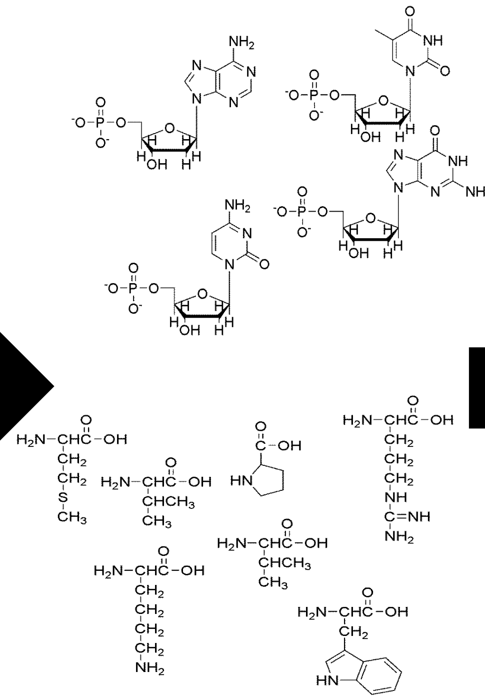
Many unknowns, but the most baffling ones are in this region:



Which "prebiotic" compounds were available/crucial?
Are some environments necessary for "complexifying" the available organics through key selective pressures?
How did energy transduction and genetic inheritance become established?
Were earlier states of living matter culled into LUCA-like ones?

The Dominant Paradigm

Reduce biological systems to their components → mix these together and watch them come alive. If this is unproductive, it is because it is a rare combinatorial event, and more trials need to be conducted.



This idea largely ignores the considerable amount of <u>biochemical evolution</u> which likely occurred between the OOL and LUCA.

So How Did Life Start?

Metabolism? e.g. Wächtershäuser’s surface metabolism model

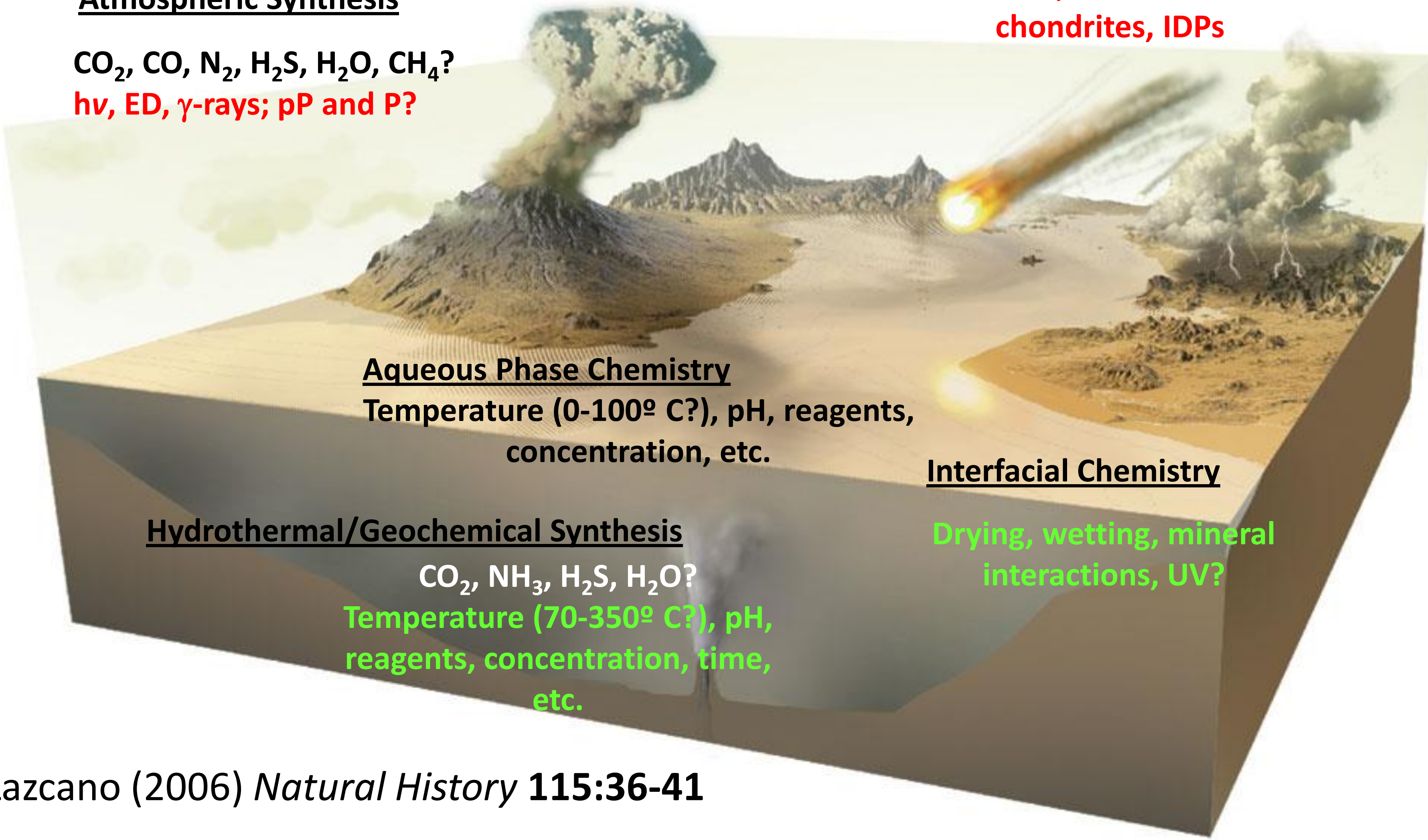
Genetics? e.g. Crick and Orgel’s RNA World model, Joyce *et al.*’s pre-RNA World Model

Encapsulation? e.g. Lancet’s GARD model

All three phenomena may be required for a minimal living system.

- There were likely a great variety of organic compounds available on the primitive Earth.
- These included some compounds which are important in modern biochemistry, ***along with many more which are not***.

Possible Sources of Prebiotic Compounds



Lazcano (2006) *Natural History* **115:36-41**

Carbonaceous Chondrite Meteorites

Hundreds of indigenous organic compounds identified in CCs, **some** are important in modern biochemistry:

Class	Concentration (ppm)	# Compounds Identified
Aliphatic Hydrocarbons	>35	140
Aromatic Hydrocarbons	22	87
Polar Hydrocarbons	<120	10
Carboxylic Acids	>300	48
Amino Acids	60	74
Hydroxy Acids	15	38
Dicarboxylic Acids	>30	44
Dicarboximides	>50	2
Pyridine Carboxylic Acids	>7	7
Sulfonic Acids	67	4
Phosphonic Acids	2	4
N-Heterocycles	7	31
Amines	13	20
Amides	n.d.	27
Polyols	30	19
Imino Acids	n.d.	10

(Adapted from Pizzarello et al. (2006) Meteorites and the Early Solar System (UofA Press) pp. 625–651)

Murchison CM2 CC

2-3 % Organic Carbon:

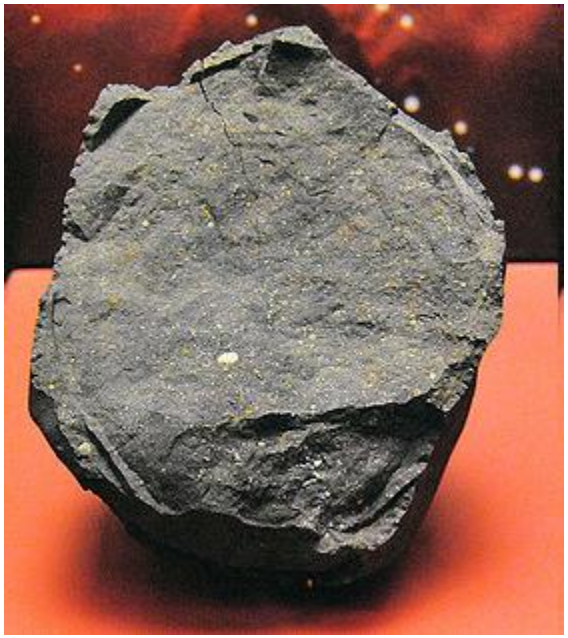
~25 mg/g

30 % Soluble OC: 7.5 mg

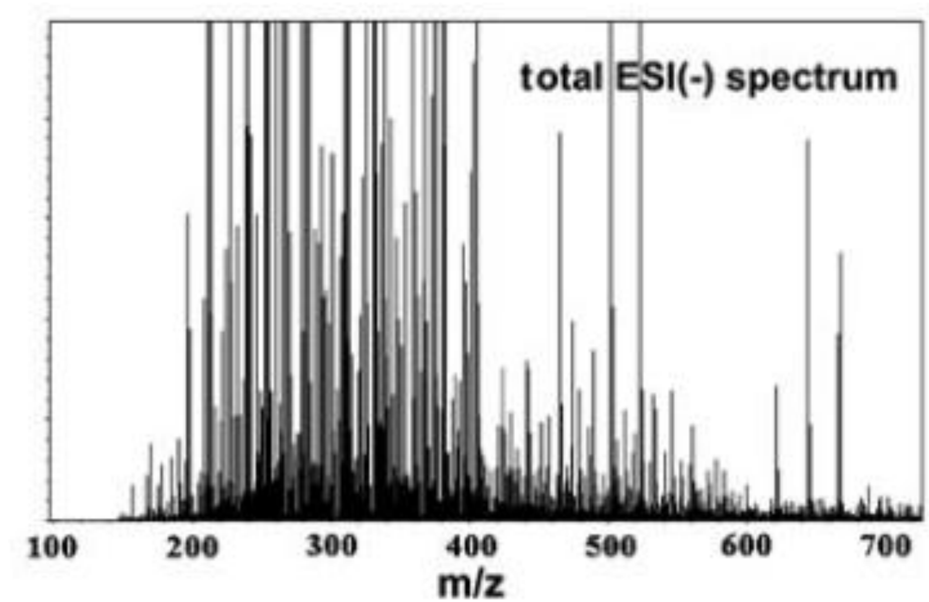
Σ compounds identified = 0.76 mg

~ 90 % of the soluble fraction is unidentified small molecules.

High Molecular Diversity of Extraterrestrial Organic Matter in Murchison Meteorite Revealed 40 Years After its Fall. Schmitt-Kopplin et al. (2010) *PNAS* **107:2763-2768**



Murchison Meteorite

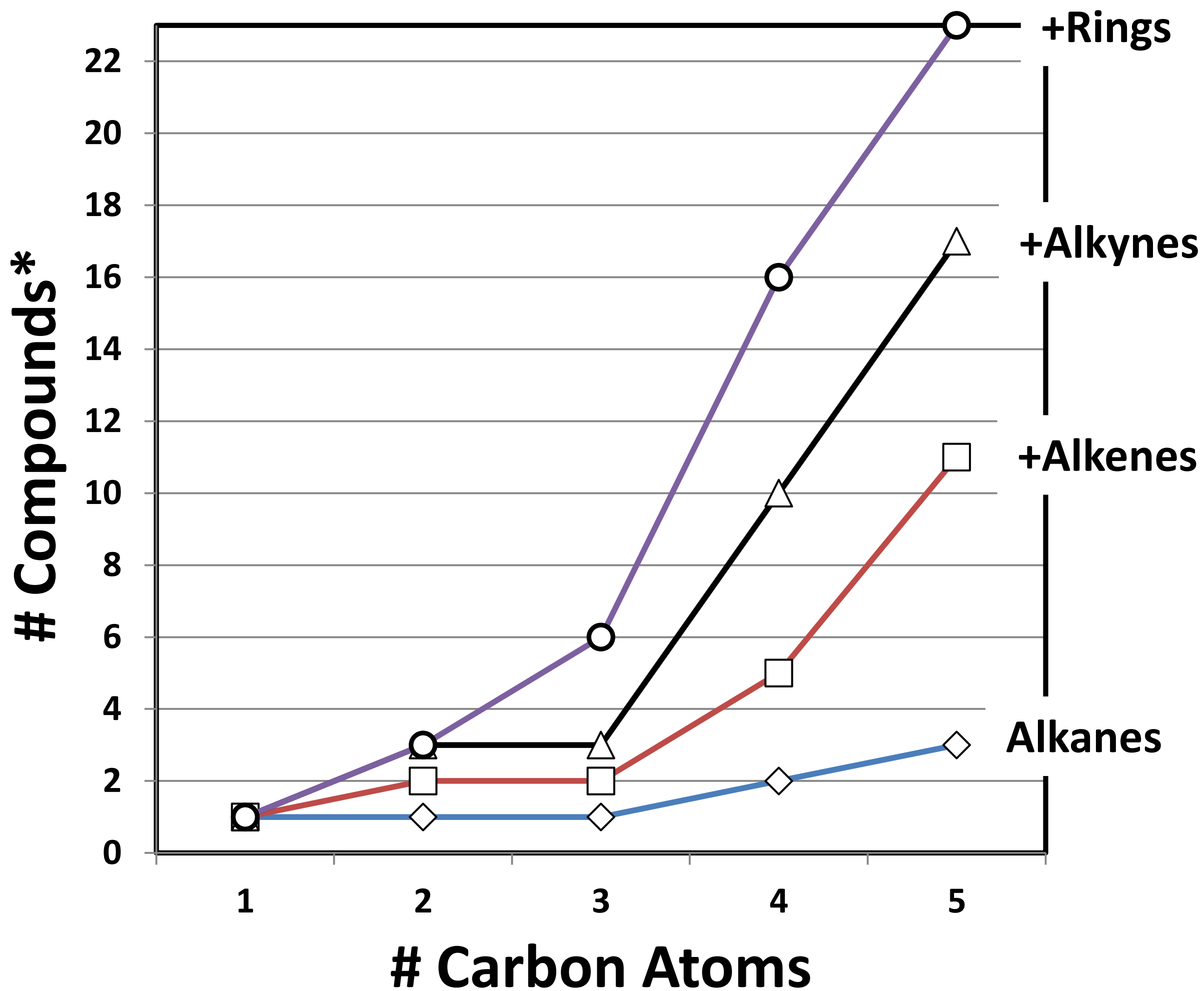


High resolution MS analysis shows **14,197** distinct elemental formulas. If typical structural and stereoisomerism is considered, there could be **50,000 to several million** distinct molecule types containing CHONS and P present.

~ 600 discrete compounds have been identified so far in CCs.
This would then be between **0.0005 – 1 %** of the total number possibly present.

What Makes Organic Chemistry Complex?

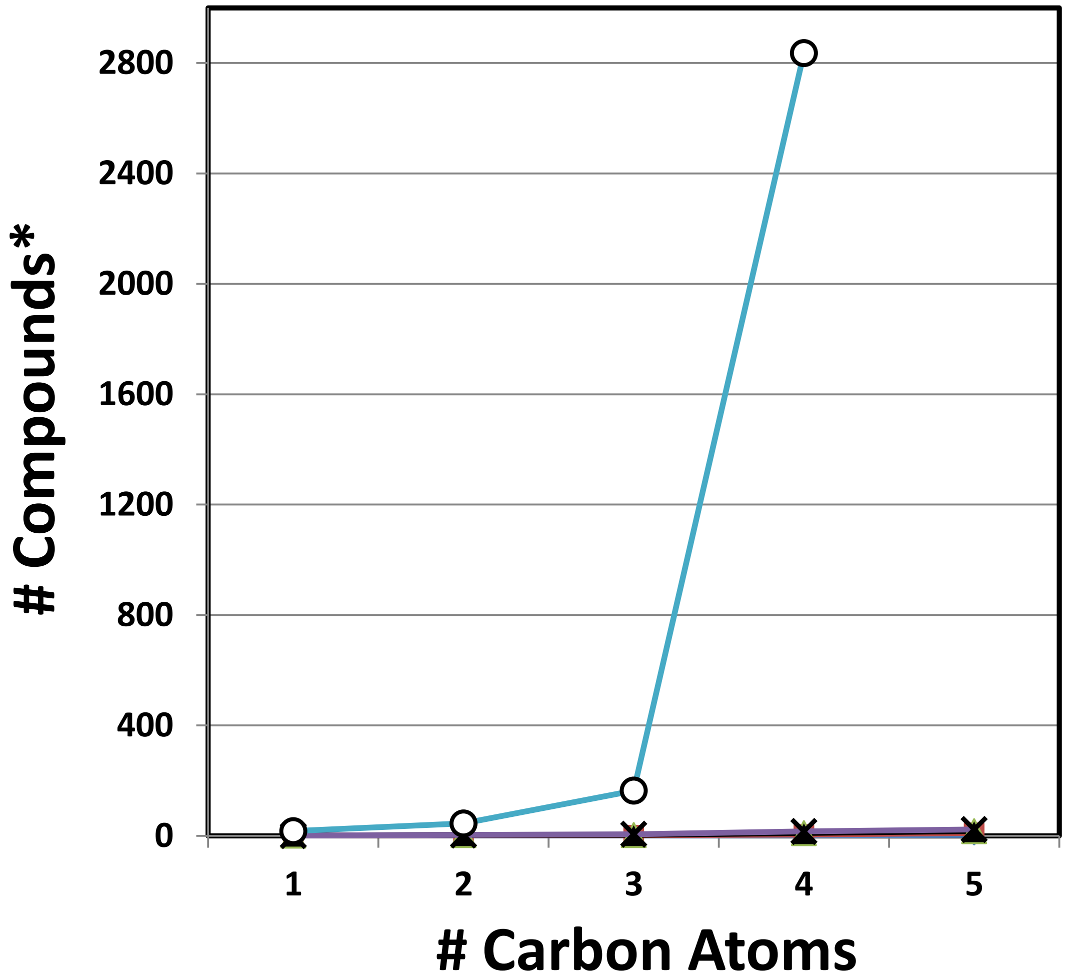
Compounds containing only C and H

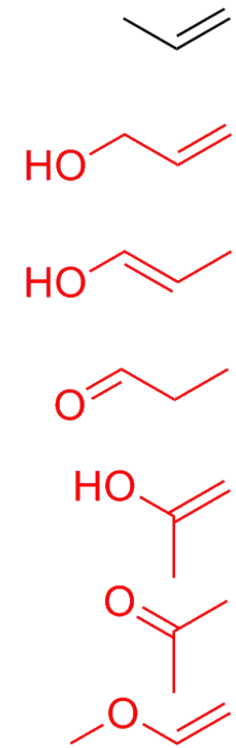


*Ignoring stereoisomerism

Answer: Heteroatoms (S,N,O…)

Compounds containing C, H, and N, O or S





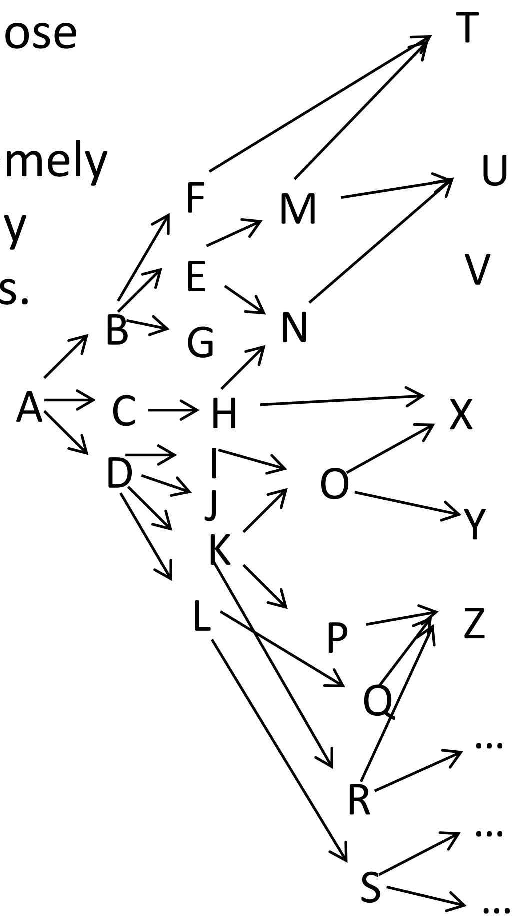
*Ignoring stereoisomerism

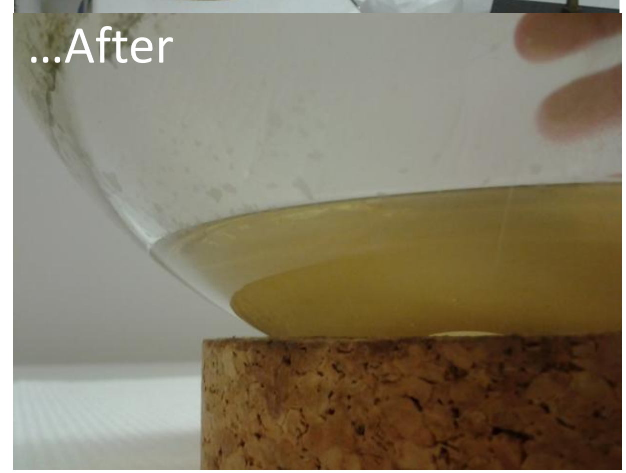
Lab Analogues Show Similar Complexity to CCs

These systems are characterized by complex kinetic multi-reactant chemistry.

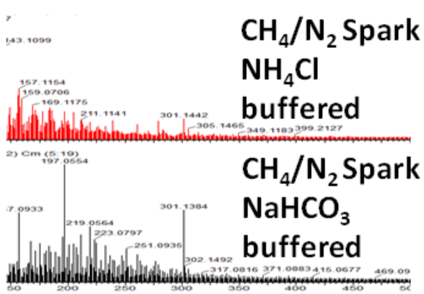
Many of these systems are remarkably reproducible despite their complexity.

Reactions like MU, the formose (HCHO) reaction and HCN polymerization can be extremely complex, despite having only one or a few initial reactants.

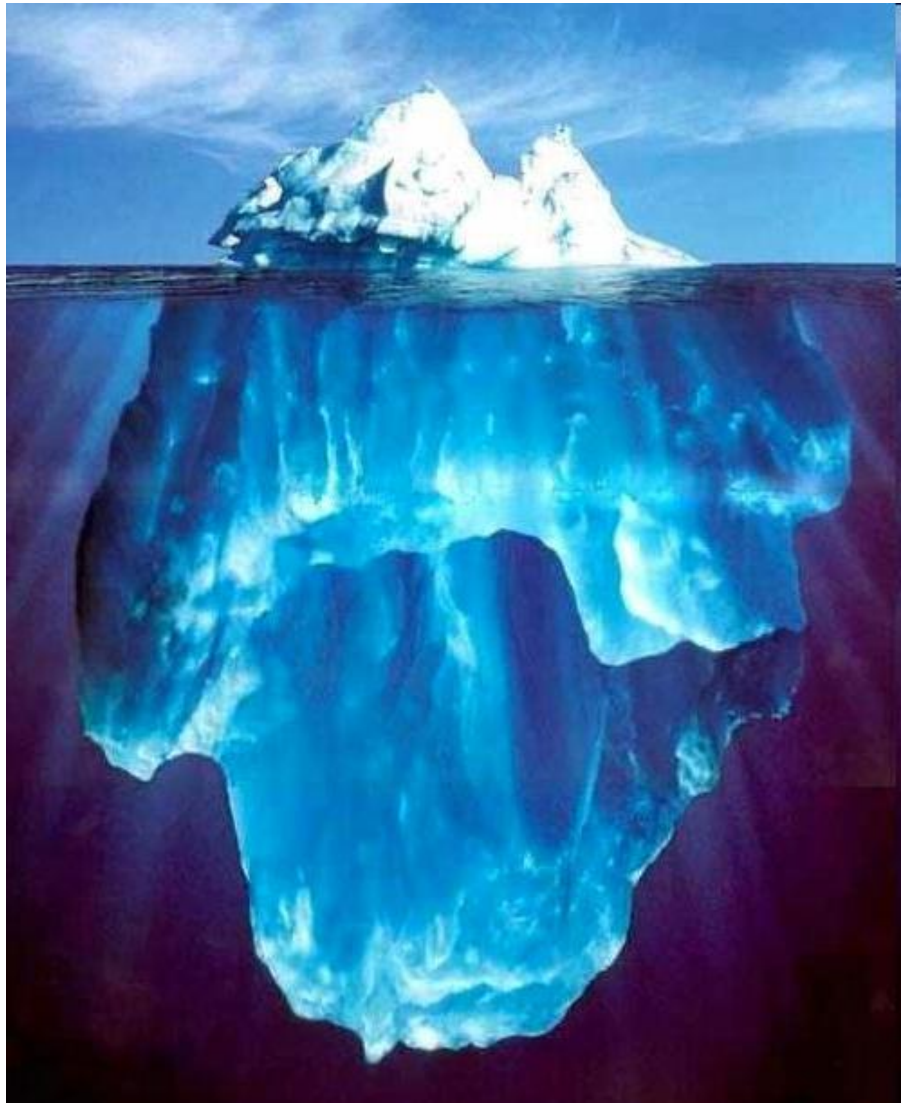




Mass Spectral Analysis

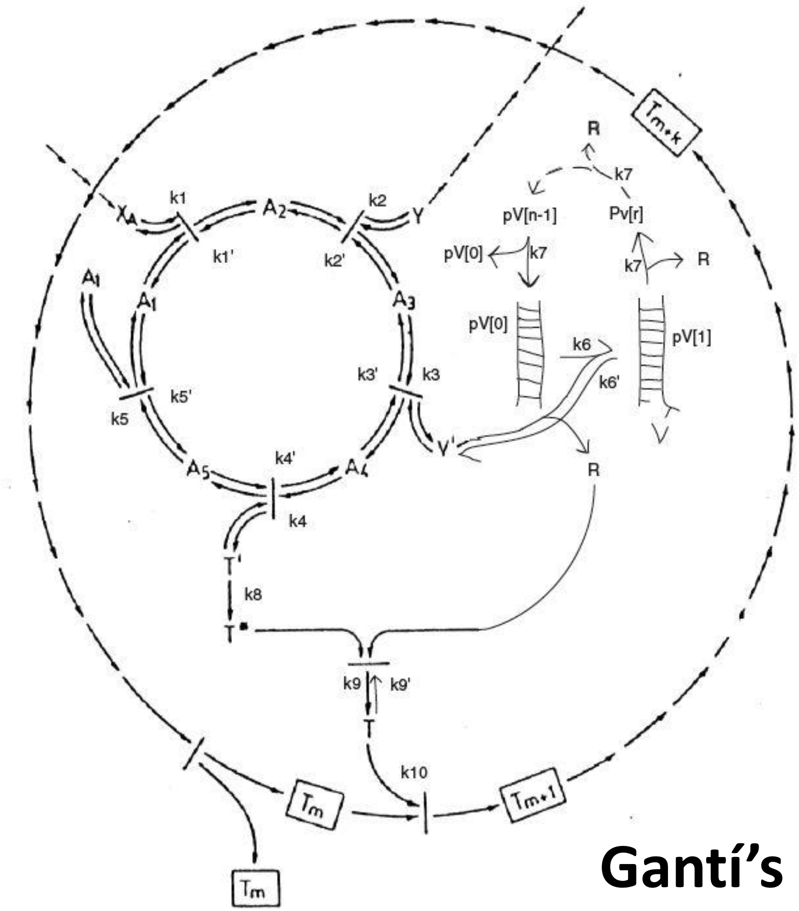


We have tended to study molecules common in modern biochemistry so far in our pursuit of understanding the origin of life.



Perhaps we need to consider the rest of the prebiotic “iceberg”?

Tibor Gantí's Chemoton Model (1971)



For Gantí, a living <u>system</u> (a "chemoton"):

- must be an individual unit: it cannot be subdivided without losing its properties
- has to perform metabolism ("work")
- must be inherently stable (homeostasis and excitability).
- must have a subsystem carrying information which is useful for the function and continuation of the whole system

Gantí's model is significant in that it <u>does not depend on the molecular identities of the chemoton's components</u>.

Are the Biological Amino Acids Special?

There are *millions (or billions)* of possible C2-C11 α-amino acids. Terrestrial biology almost exclusively encodes <u>20</u>.

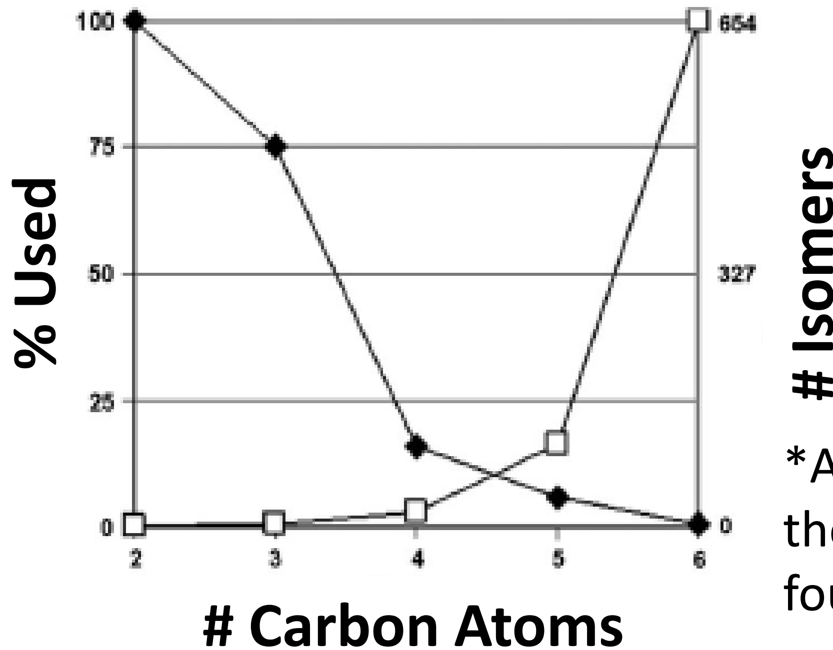
Why does biology use this particular set of 20 α-amino acids?
Exactly how many are possible?

Has terrestrial biochemistry found the *universally optimal* solution, to the degree that we would have a hard time identifying a novel biochemistry if we found one?

If not, which amino acids would be good signatures of a novel biochemistry?

Coded Amino Acids in Amino Acid Space

For small amino acids (C≤ 4), biology uses a good sampling of the possible “chemical space”.
For larger amino acids, this relationship falls off rapidly.



*AAs only containing the functional groups found in the coded AAs.

Why is this so?

Cleaves, HJ (2010) *J. Theor. Biol.* **263:490-498**

Computationally Exploring Molecular Complexity

- Software developed at Bayreuth University, Germany.
- An indirect descendant of **Lederberg's 1960's DENDRAL** project developed for use in exobiology as a computing system to help study alien organic compounds.
- Allows the rapid computation of <u>ALL</u> molecular structures which obey Lewis electron-pairing rules for a given molecular formula.

Molecular Formulas →

Enumerated Structures →

Filtered Structures →

Calculated Physical Properties →

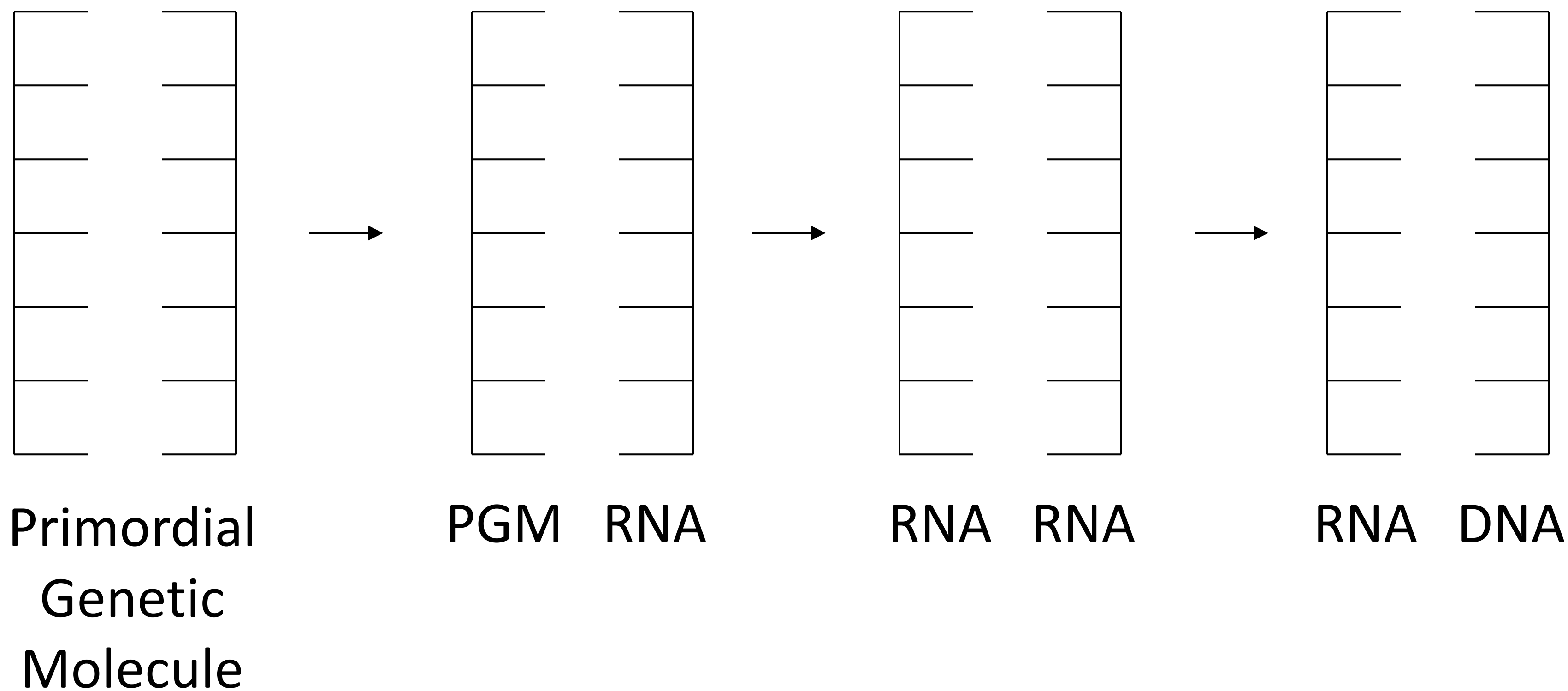
Comparison with Coded AAs

Genetic Takeover

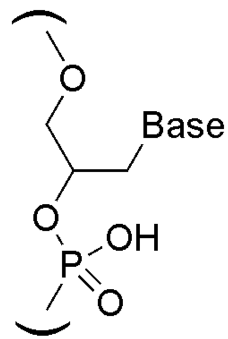
Life may have evolved considerably since its origin.

This seems very likely for amino acids, how about for nucleic acids?

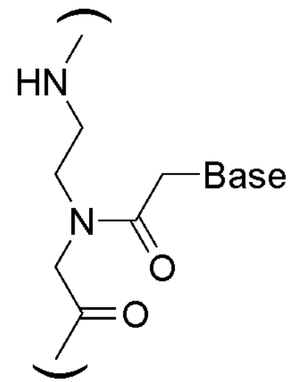
It has been postulated that RNA may have been preceded by another genetic material which was more easily produced prebiotically (Joyce *et al.* (1987) *PNAS* **84:4398–4402**)



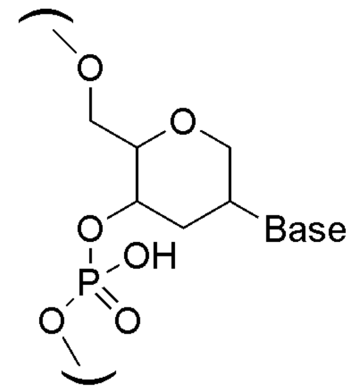
Some Characterized Alternative Nucleic Acids



Glycerol Nucleic Acid (GNA)



Peptide Nucleic Acid (PNA)



Hexitol Nucleic Acid (HNA)

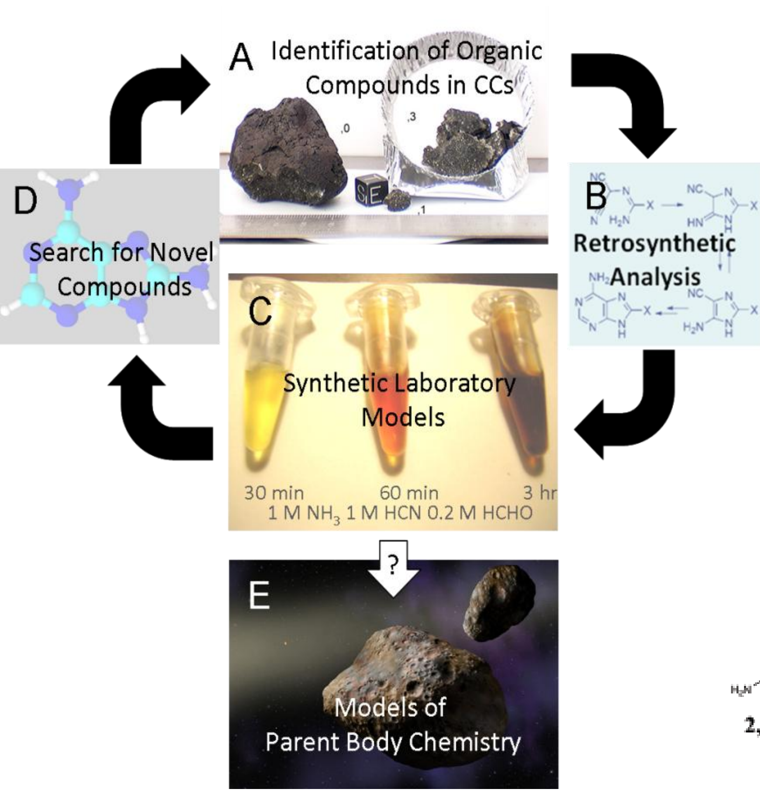
Many molecules appear to have the capacity to template and store information.

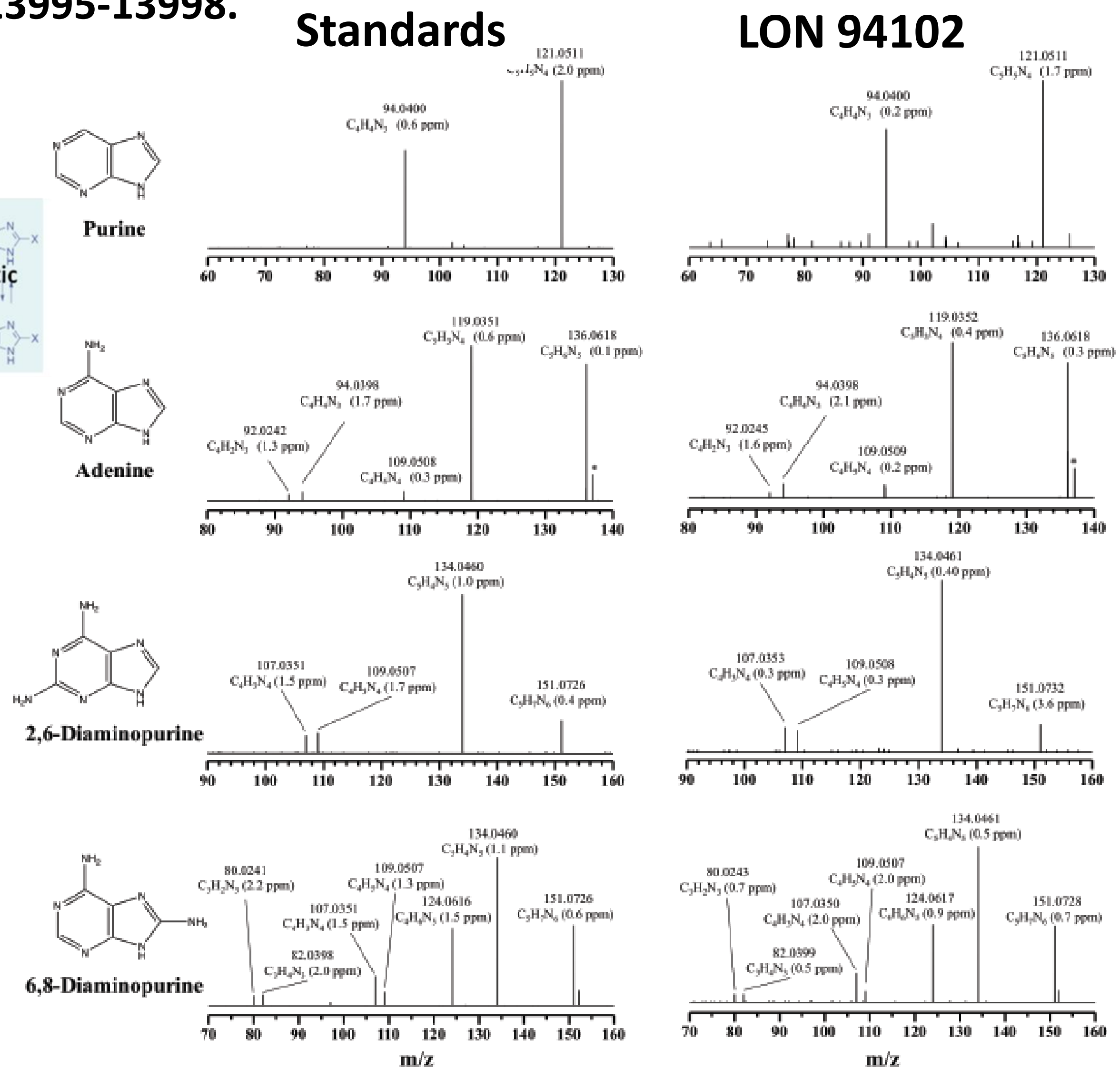
While a great deal of effort has gone into trying to make RNA prebiotically, little has gone into making or detecting alternatives.

What properties should they have? How would one look for them?

Search Method Applied to Purines in CCs

Callahan et al. (2011) *PNAS* 108: 13995-13998.





Since CC samples are relatively scarce, we can use lab analogues to search for novel compounds.

A Return to Biochemistry 101

- We have only begun to scratch the surface of the organic complexity present in CCs or generated in laboratory analogues: **many, many thousands of molecules remain to be discovered**.
- If Life was <u>ever</u> very different from how it is now, it was likely so near its origin.
- Novel self-replicating systems could be difficult to recognize with modern biological preconceptions and analytical techniques (PCR?).
- Should we perhaps try to know more about this complexity before formulating OOL models in *too* much detail?
- The most fundamental <u>analyzable</u> signature of a living system is likely chemical disequilibrium.
  - Disequilibrium requires knowing something about the *baseline* equilibrium signal.
- The combination of high resolution MS and computational structure searching is a powerful means of attacking some of these problems.
- **It took ~ 125 years from Wöhler's urea synthesis to Watson and Crick's DNA structure – how quickly could we decipher a novel biochemistry now?**

Acknowledgments

NAI/Exobiology

NSF/NASA Center for Chemical Evolution

KUL

Piet Herdewijn

Jef Rosenzky

Elizabetha Groaz

GA Tech

Irena Mamjanov

Facundo Fernandez

Manshui Zhou

Eric Parker

Emory U.

David Lynn

Jay Goodwin

Neil Anthony

NASA GSFC

Mike Callahan

Danny Glavin

Jason Dworkin

UHI

Steve Freeland

DLR

Markus Meringer

SIO

Jeffrey Bada

CIW

George Cody

Bob Hazen

Spelman College

Jean-Marie Dimandja