

# Understanding Cells and Systems

---

*Or...The meaning of life, the universe,  
and everything*

Lecture 6

Heidi J. Sofia, Computational BioSciences

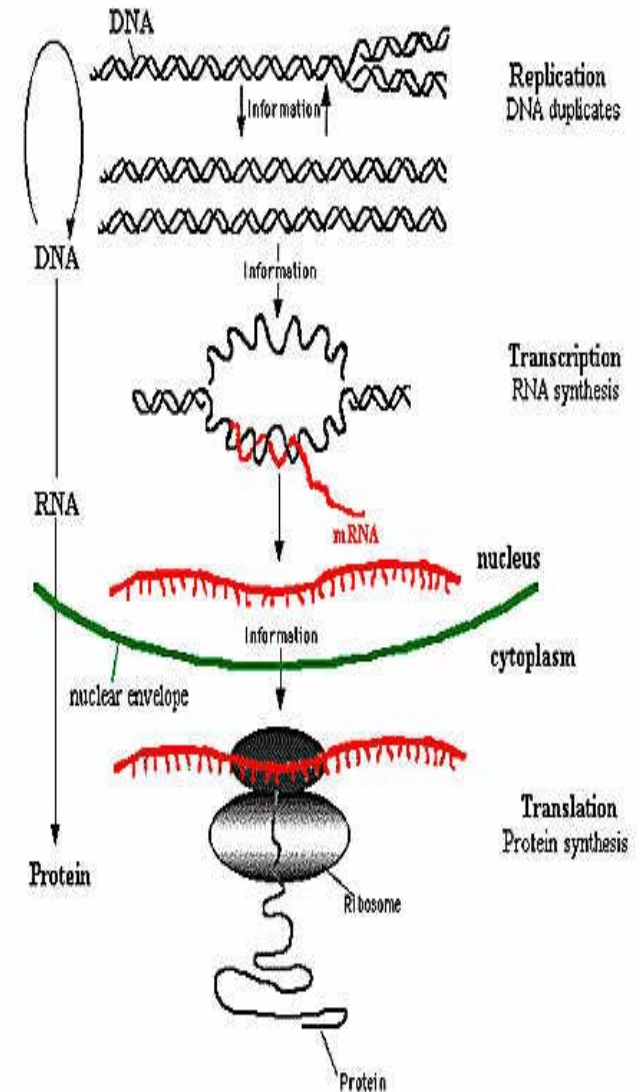
# *A Brief Review*

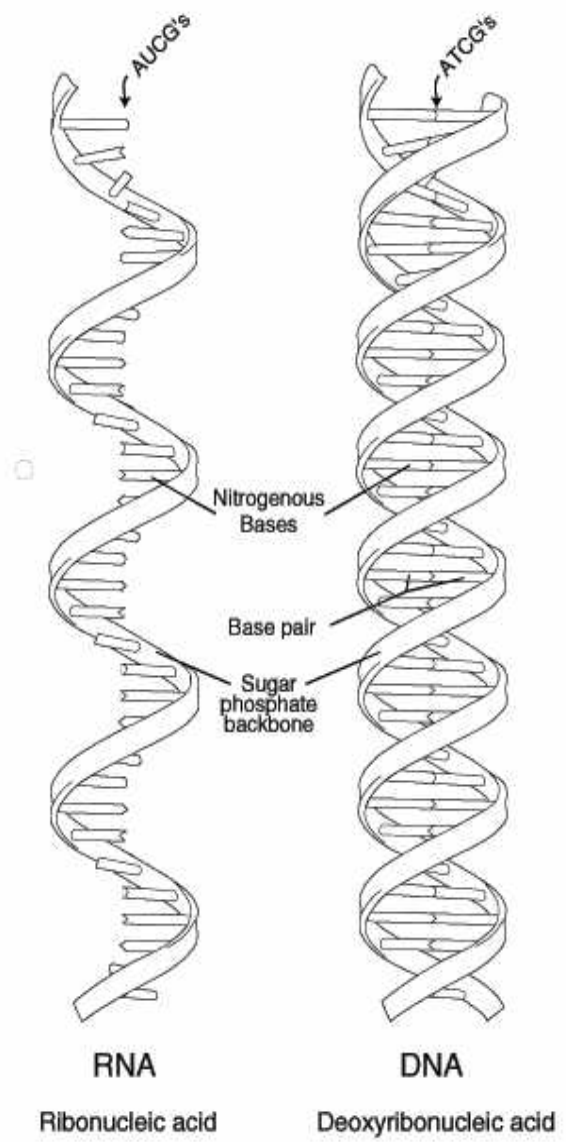
Central dogma for biology:

**DNA → RNA → Protein**

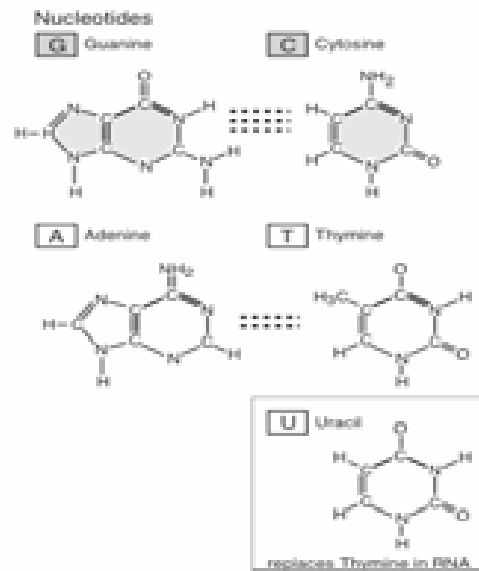
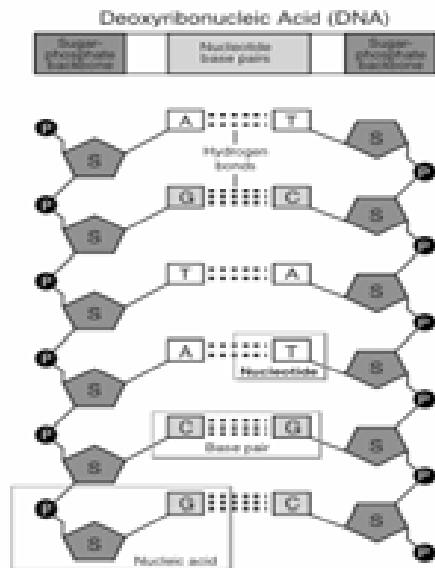
## *Analogies*

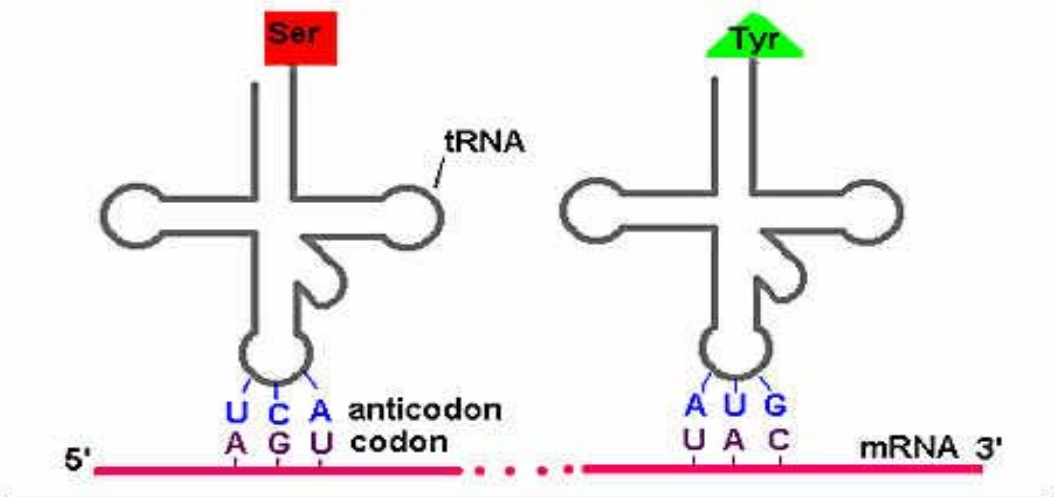
Blueprint → Xerox → Machinery  
Source code → Ftp copy → Executable  
Cookbook → Recipe → Cookie





## Nucleotide





2nd base in codon

		U	C	A	G		
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G	3rd base in codon
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G	
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G	
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G	

# The Genetic Code

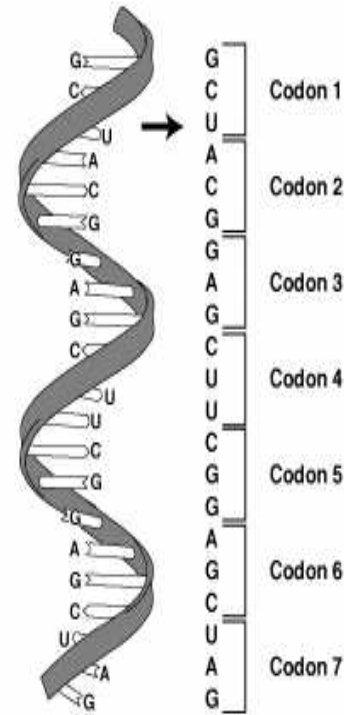
# What is DNA?

- deoxyribonucleic acid
- an autoreplicating helical double-stranded linear polymer of sugar, phosphate, and bases including adenine (A), thymine (T), cytosine (C), and guanine (G)
- the molecule that encodes information about all life forms on the third planet from the sun.

➤ *For example, some dinosaur DNA:*

5' GCGTTGCTGGCGTTTTTCCATAGGCTCCG 3'

Level of abstraction: **DNA is just A's, C's, T's, G's**



RNA

Ribonucleic acid

*Life is simple:*  
just A's, T's, C's, and G's

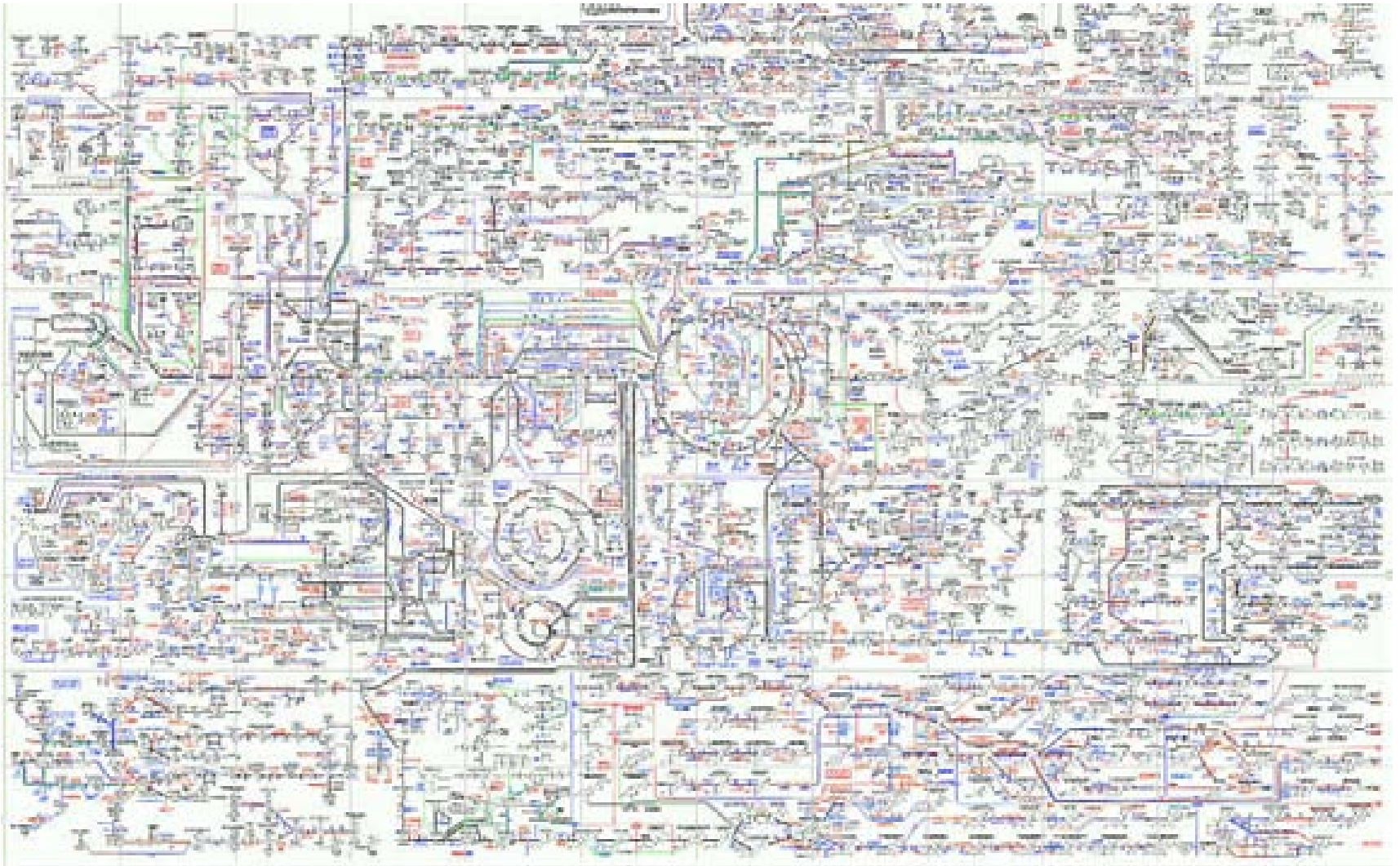
*Life is complex:*  
biology is hard because of complexity

Evolutionary theory is the foundation of modern biology.

How can such complexity simply evolve?



# Complexity of Biochemical Pathways in the Cell

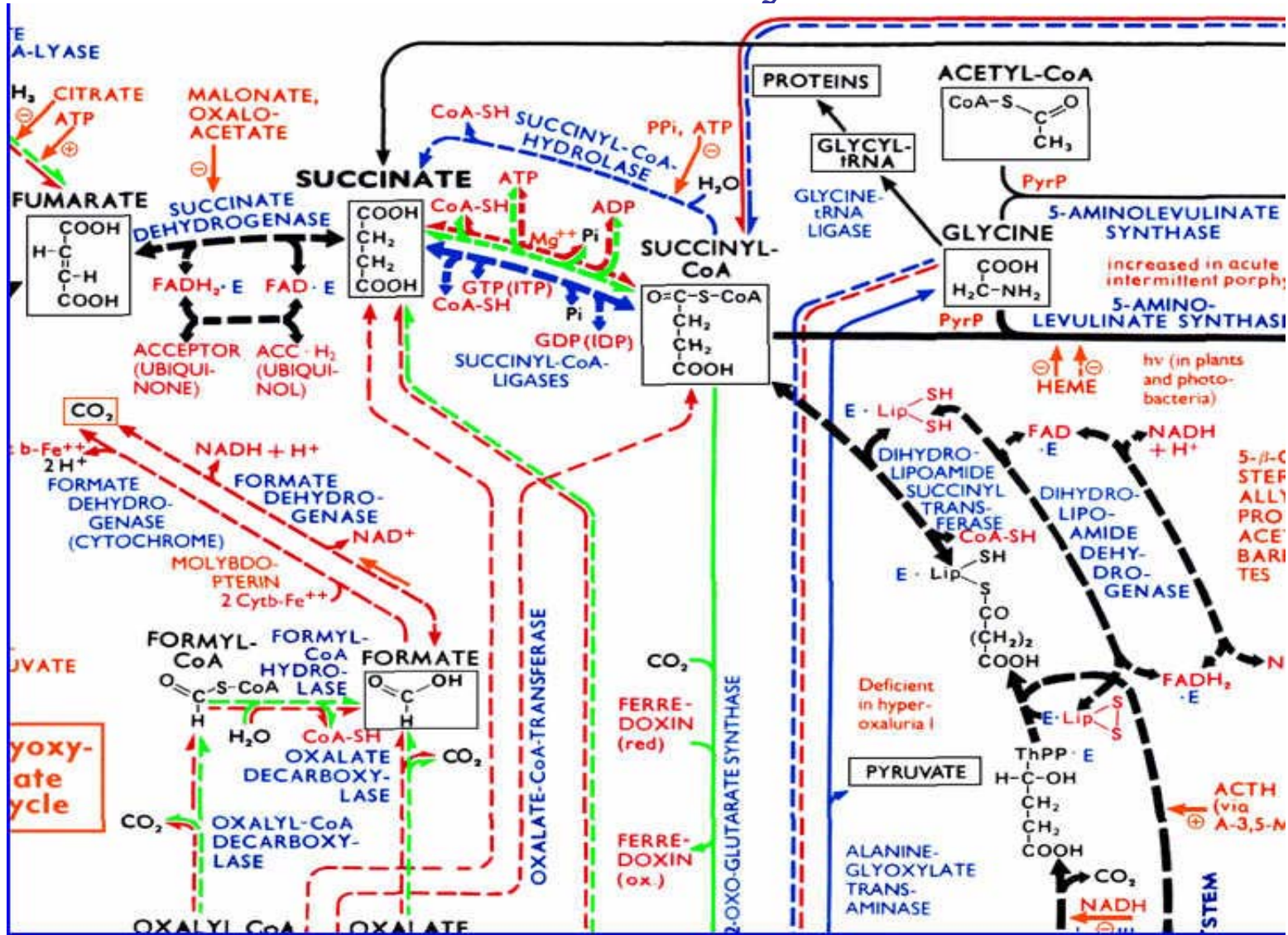


**Pacific Northwest  
National Laboratory**

Operated by Battelle for the  
U.S. Department of Energy



# Krebs Cycle



# Evolutionary Theory is the Foundation of Modern Biology and Biotechnology

➤ Evolution is genetic change over time through natural selection.

➤ Evolutionary theory “works” (evidence: medical advances, new drugs and treatments)

➤ Direction of flow of information:

DNA → Organism

*not*

Organism → DNA (“Lamarckian”)

# Life Itself: Its Origin and Nature (1981)

Francis Crick

- proposed the scenario that life on earth originated when an unmanned rocket carrying primitive spores was sent to the earth by a higher civilization that had developed billions of years ago. [Theory of Directed Panspermia]

# A Larger Perspective

## Stars

hundred billion stars in our galaxy ( $10^{11}$ )

hundred billion galaxies in visible universe ( $10^{11}$ )

## Earth

earth is 4.55 billion years old

sun is halfway through its lifecycle

## Life

Life on earth is between 3.5 to 4 billion years old.

*Predicted demise of life on earth*: 10 minutes to 10 million years from now, due to supernova of nearby star *Betelgeuse* (pronounced “Beetlejuice”)

# What is Life?

This is a challenging question.

Life forms store information, transmit information through generations, and have activity and complex structures.

<u>Entity</u>	<u>Structure</u>	<u>Characteristics</u>
Plant/Animal	Higher organism	Information + <b>Activity</b>
Eukaryote	Complex cell	Information + <b>Activity</b>
Prokaryote (bacteria)	Simple cell	Information + <b>Activity</b>
Virus	Parasitic cell	Information
Transposon	Rogue gene	Information
Prion	Rogue protein	Information

# Current Paradigm

**DNA**

Information

**RNA**

Information

**Protein**

Machinery

## Primordial World

**RNA** both encoded information and catalyzed chemical reactions.

The modern paradigm has erased all but traces of this earlier world.

# The Primordial Soup

- no oxygen (O<sub>2</sub>) on the primitive earth
- electrical discharge in a primitive atmosphere of H<sub>2</sub>O, H<sub>2</sub>, NH<sub>3</sub>, and CH<sub>4</sub> (water, hydrogen, ammonia, methane)
- with energy from lightening strikes and UV light, organic compounds can be synthesized abiotically
- rich “primordial soup” of precursor molecules
- some molecules became self-replicating

“...some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity ...”

*Charles Darwin on the origins of life in tidal pools*

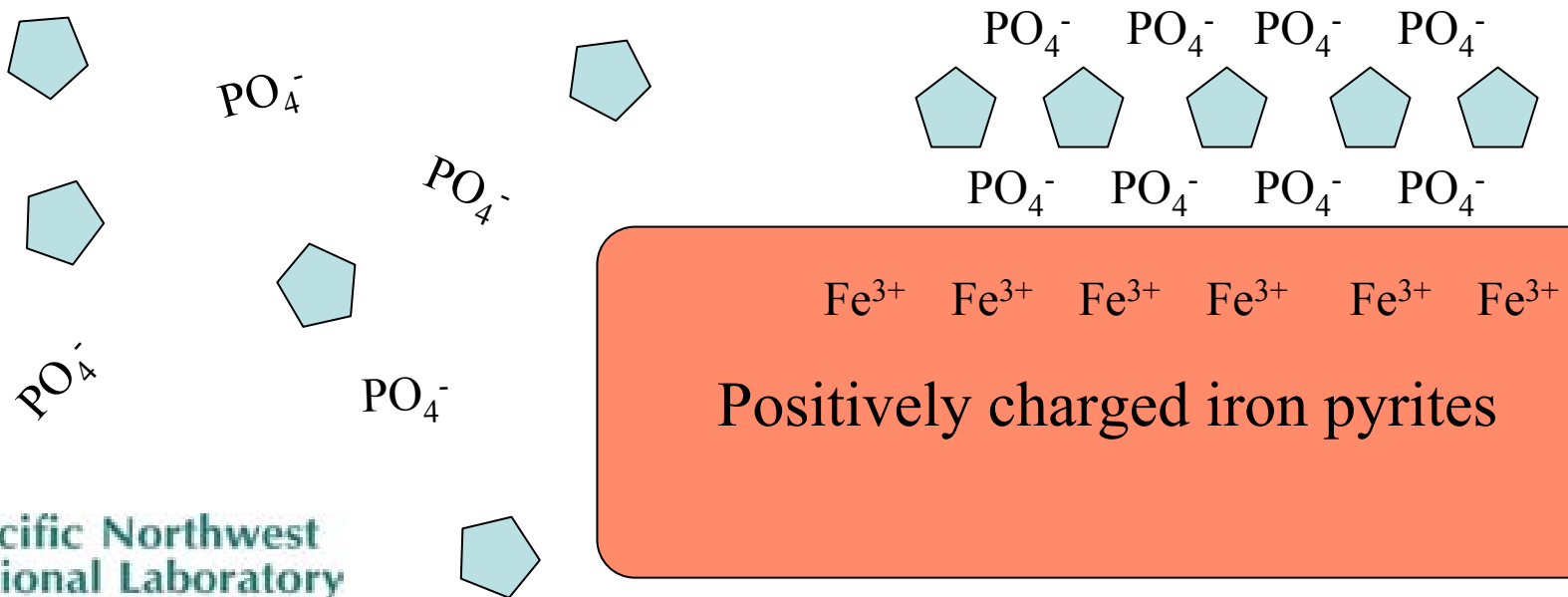
# Polymerization

Polymerization reactions can occur naturally.

(For example, heating and drying a mixture of amino acids will produce amino acid chains.)

A charged mineral surface can support orderly polymerization

Huber, C. and Wächterhäuser, G., 1998. Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science* **281**(5377):670–672.

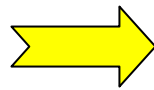
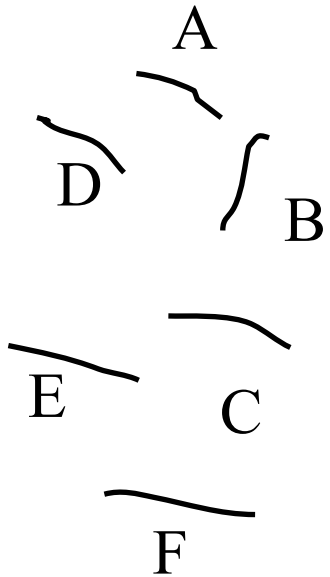




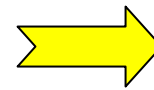
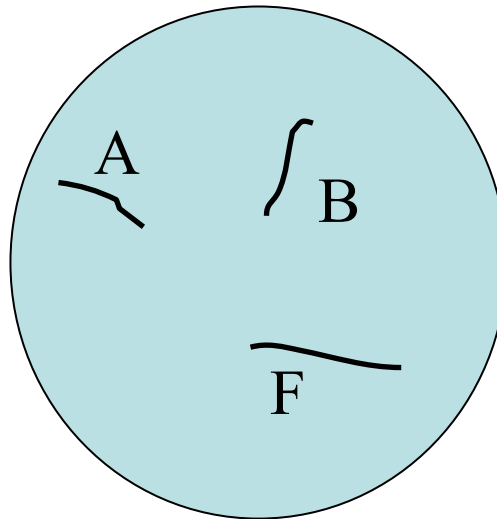
# Competition vs. Cooperation in Genetic Elements

Linkage between elements shifts evolutionary drivers for selection from competition toward cooperation

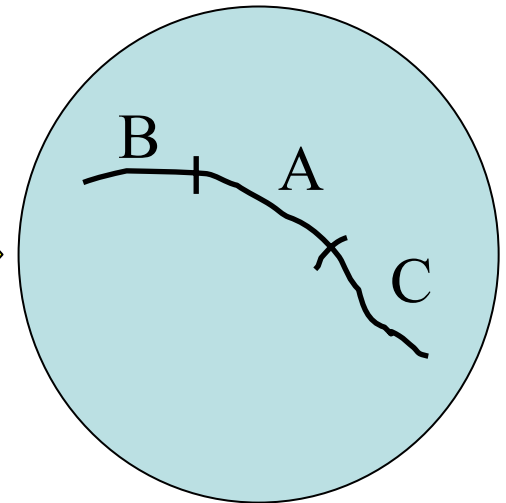
Independent units compete for same resources



Units linked by membrane

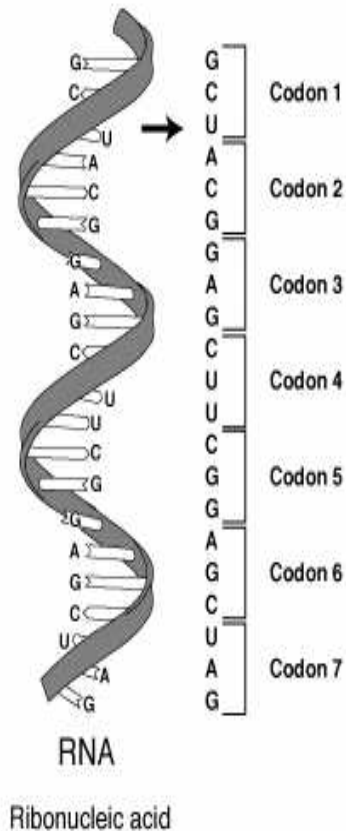


Units linked genetically



cooperate or die

# RNA World to Modern World



➤ In the primordial world RNA encoded information and catalyzed reactions.

➤ RNA chemistry may have been extended with amino acid “cofactors.”

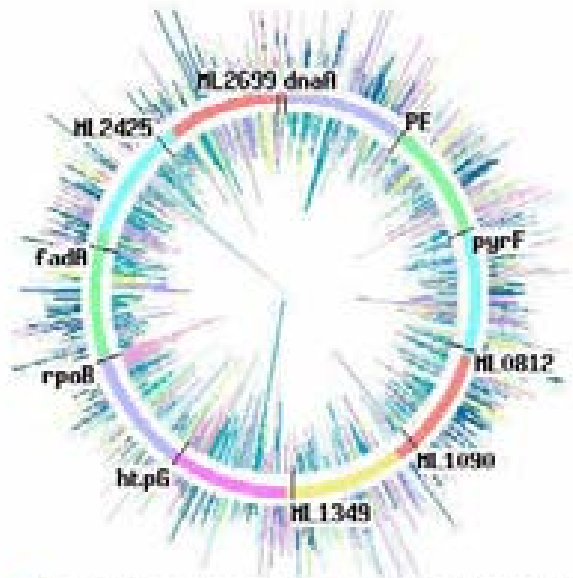
➤ Eventually, functions are separated into DNA, RNA, and proteins.

# Protein-Coding Genes

*Mycobacterium leprae*

*Genome size*

1.5 Megabases



- Translation, ribosomal structure and biogenesis
- Transcription
- DNA replication, recombination and repair
- Cell division and chromosome partitioning
- Posttranslational modification, protein turnover
- Cell envelope biogenesis, outer membrane
- Cell motility and secretion
- Inorganic ion transport and metabolism
- Signal transduction mechanisms
- Energy production and conversion
- Carbohydrate transport and metabolism
- Amino acid transport and metabolism
- Nucleotide transport and metabolism
- Coenzyme metabolism
- Lipid metabolism
- Secondary metabolites biosynthesis, transport and catabolism
- General function prediction only
- Function unknown
- No COG match

Pacific Northwest  
National Laboratory

Operated by Battelle for the  
U.S. Department of Energy

# Basic Local Alignment Search Tool (BLAST)

- **BLAST** is like a telescope looking backwards through evolutionary time.
- Use **BLAST** to **observe** sequence similarity, then **infer** homology
- **Homology**: a shared common ancestor  
*Examples of shared common ancestor:*
  - mitochondrial Eve
  - Y chromosome Adam
  - King George III

# National Center for Biotechnology Information

<http://www.ncbi.nlm.nih.gov>

## Genetic sequence databases and tools

<b>GenBank</b>	Sequence database
<b>BLAST</b>	Sequence comparison
<b>MedLine</b>	Biomedical literature
<b>Entrez</b>	Data integration

## Database Sequence

```
LOCUS      NP_799976                141 aa          linear  BCT 11-MAR-2003
DEFINITION universal stress protein A [Vibrio parahaemolyticus]
```

```

1 mkykhilval elsdestvli dravsmany1 dsdisfihid gthgeiyrel vdikenpdqr
61 plnehsmecl rtfSDYMDYP lkhffvgtgd ladklevtik egevdllicg hhqdfwskii
121 sysrhlinks pvdilvvp1h e
```

## FASTA format

```
>gi|28900321|ref|NP_797.1| universal stress protein A [Vibrio parahaemolyticus]
MKYKHILVALELSDESTVLIDRAVSMANYLDSDISFIHIDGTHGGEIYRELVDIKENPDQRPLNEHSMECL
RTFSDYMDYPLKHFFVGTGDLADKLEVTIKEQEVDLLICGHHQDFWSKIISYSRHLINKSPVDILVVP1HE
```

# Measuring Sequence Similarity

BLAST compares the query sequence to each database sequence and scores each pairwise alignment.

ID	Annotation	Species	Score	E value
gi 28900321 ref NP_799976.1	universal stress protein A [Vibrio ...		266	5e-71
gi 16124098 ref NP_407411.1	universal stress protein A [Yersini...		121	2e-27
gi 16766877 ref NP_462492.1	universal stress protein A [Salmon...		120	4e-27
gi 16272756 ref NP_438975.1	universal stress protein A [Haemoph...		110	3e-24
gi 15603151 ref NP_246223.1	UspA [Pasteurella multocida] >gi 12...		107	4e-23
gi 27905001 ref NP_778127.1	universal stress protein [Buchnera ...		86	1e-16
gi 26248161 ref NP_754201.1	Hypothetical protein yecG [Escheric...		64	4e-10
gi 23029259 ref ZP_00067612.1	hypothetical protein [Microbulbif...		48	3e-05
gi 15598213 ref NP_251707.1	conserved hypothetical protein [Pse...		41	0.004
gi 23102010 ref ZP_00088542.1	hypothetical protein [Azotobacter...		40	0.005
gi 26988856 ref NP_744281.1	universal stress protein family [Ps...		40	0.006
gi 28870681 ref NP_793300.1	universal stress protein family [Ps...		40	0.009
gi 23472231 ref ZP_00127558.1	hypothetical protein [Pseudomonas...		39	0.019
gi 3800760 gb AAC68863.1	ubiquitin specific protease 46 [Gallus...		37	0.053
gi 21673839 ref NP_661904.1	universal stress protein family [Ch...		35	0.18
gi 23058583 ref ZP_00083629.1	hypothetical protein [Pseudomonas...		35	0.29
gi 29655266 ref NP_820958.1	universal stress protein A [Coxiell...		34	0.36
gi 21356575 ref NP_648665.1	CG8745-PA [Drosophila melanogaster]...		33	0.54
gi 28210234 ref NP_781178.1	ABC transporter permease protein [C...		33	0.68
gi 27731453 ref XP_218577.1	similar to otogelin; twister [Mus m...		33	0.91
gi 19718764 ref NP_570854.1	ubiquitin protein ligase E3A isoform...		33	1.1
gi 12803511 gb AAH02582.1 AAH02582	ubiquitin protein ligase E3A ...		32	1.1

Y

?

N

# BLAST Pairwise Alignment View

**Query=** gi|28900321|ref|NP\_799976.1|universal stress protein A  
[Vibrio parahaemolyticus] (141 letters)

**Database:** All non-redundant GenBank CDS translations+PDB+SwissProt+PIR  
1,438,044 sequences; 462,300,935 total letters

>[gi|23029259|ref|ZP\\_00067612.1|](#) hypothetical protein [Microbulbifer degradans]  
Length = 149

Score = 48.1 bits (113), Expect = 3e-05

Identities = 32/141 (22%), Positives = 67/141 (47%), Gaps = 2/141 (1%)

```
Query: 3   YKHILVALELSDESTVLIDRAVSMANYLDSDISFIHIDGTHGEIYRELVDIKENPDQRPL 62
          Y H+LV L+LS++ST ++ +A+ +A ++ +S H+ Y V + Q +
Sbjct: 9   YSHLLVGLDLSEDSTAILSKAIKLAKACNAQVSLAHVVEPLAFAYGGDVPMDLTEAQVVM 68

Query: 63  NEHSMECLRTFSDYMDYPLKHFFVGTGDLADKLEVTIKEQEVDLLICGHH--QDFWSKII 120
          + + L +D ++ + + G+ A +L +E DL++ G H F +
Sbjct: 69  EQQAQARLDKIADELELTPANMAILVGNTAGELHRYAEEHSADLIVVGSHGRHGFALLLG 128

Query: 121 SYSRHLINKSPVDILVVPIHE 141
          S + +++ + D+L V + E
Sbjct: 129 STANGVLHGAKTDVLAVRVGE 149
```

# Multidimensional View of Large-Scale Sequence Data

## *Similarity box interactive visualization software*

