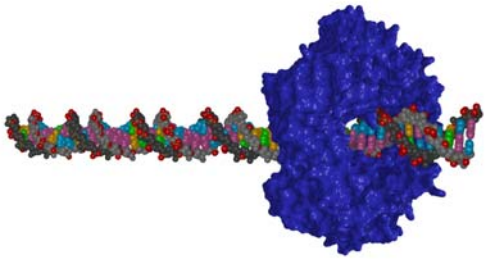


Daniel Barsky

Life on Earth: Instructions in Three Billion (tiny) Letters or Less

March 14, 2006

UCRL-PRES-220166

I. How things work—which axioms (“laws”), what level of detail?

How does that light work?

- To a theatre manager
- To an electrician
- To an electrical engineer
- To a physicist

In answering **how** and **why,** we are reducing the unfamiliar or complex to the familiar and simple, consistent with what we already know.

“Make everything as simple as possible but no simpler”

—A. Einstein

I. How things work—Conservation as a guide to mechanism

Even if we can't see how something happens, we can learn a lot by examining <u>what is conserved</u> as a guide…

What's needed to change a tire?
What about to go to the moon?
What's needed to live? (elements, organs, sleep)
What's needed for life to appear?

Abstract sequence comparisons…

```
SLINECWV
SLINENWV
SLINEDWV
SLINEQWV
```

```
WVVTAAHCEVGH
DATALAHCECLE
ATAMIAHCESYI
MKPASAHCEYRS
```

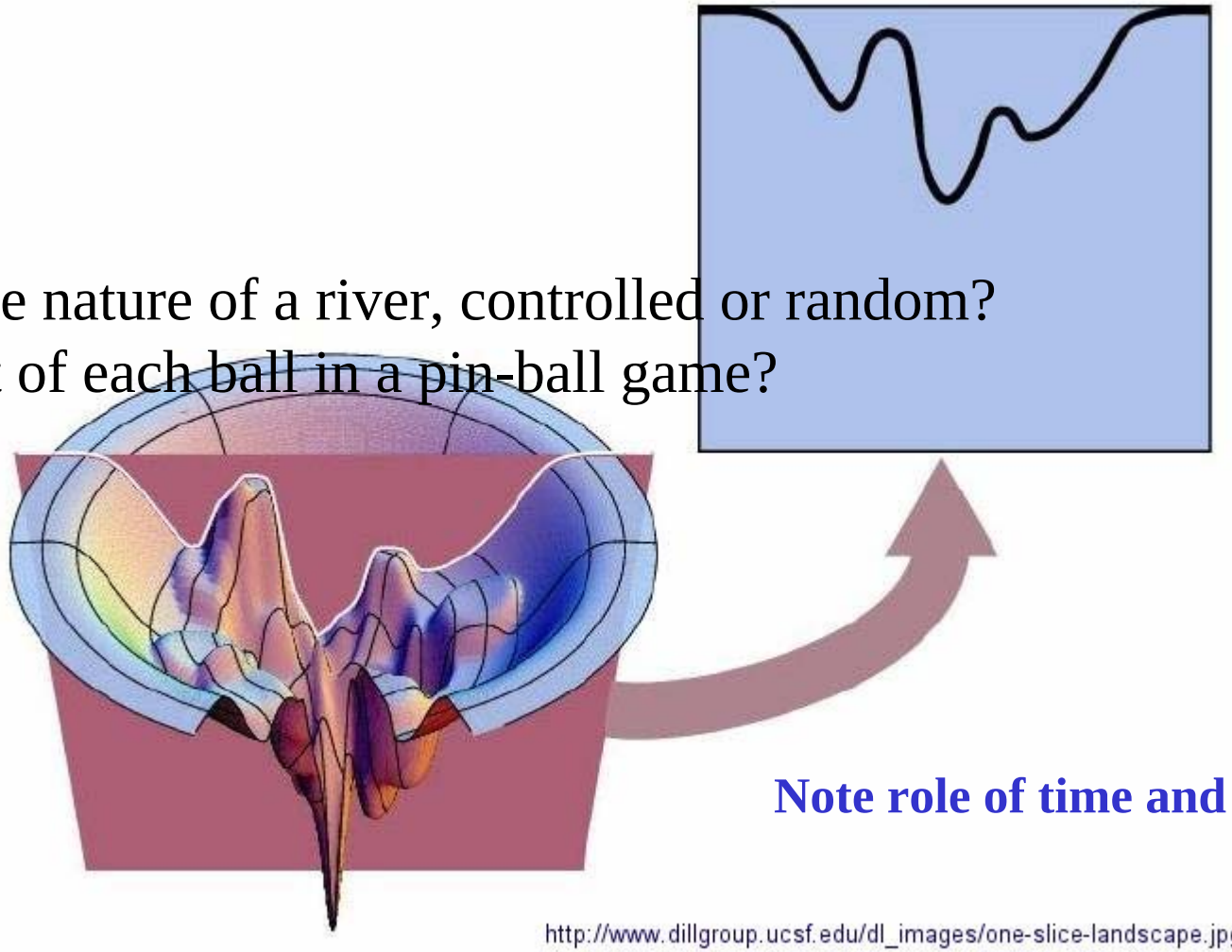
I. How things work—Control versus randomness

A. Biking to the store (strategies to reach goal despite randomness, try bumper car!)

B. Cellular function (“strategies:” goals and mechanisms)

C. Protein folding (“free energy:” goals and mechansims)

What is the nature of a river, controlled or random?
The plight of each ball in a pin-ball game?



Note role of time and size scale

http://www.dillgroup.ucsf.edu/dl_images/one-slice-landscape.jpg

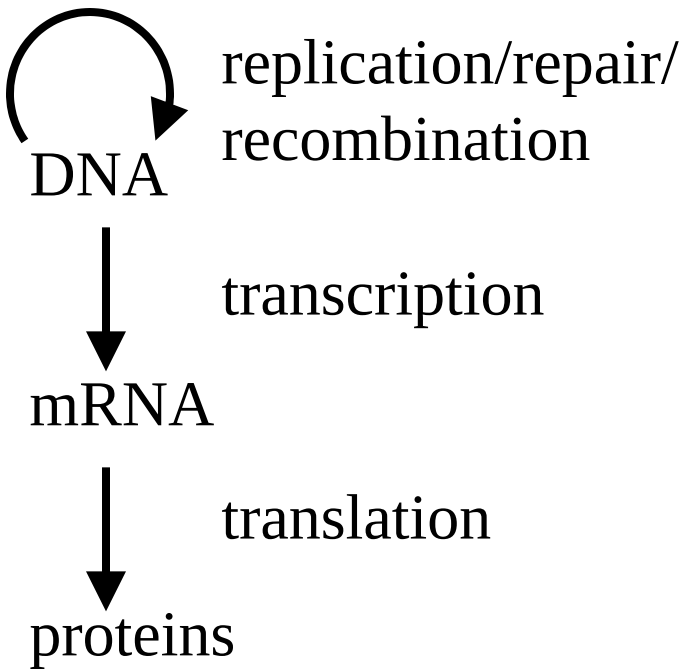
II. Central Dogma of Biology: DNA → mRNA → proteins

<u>metaphors</u>

master set of instructions

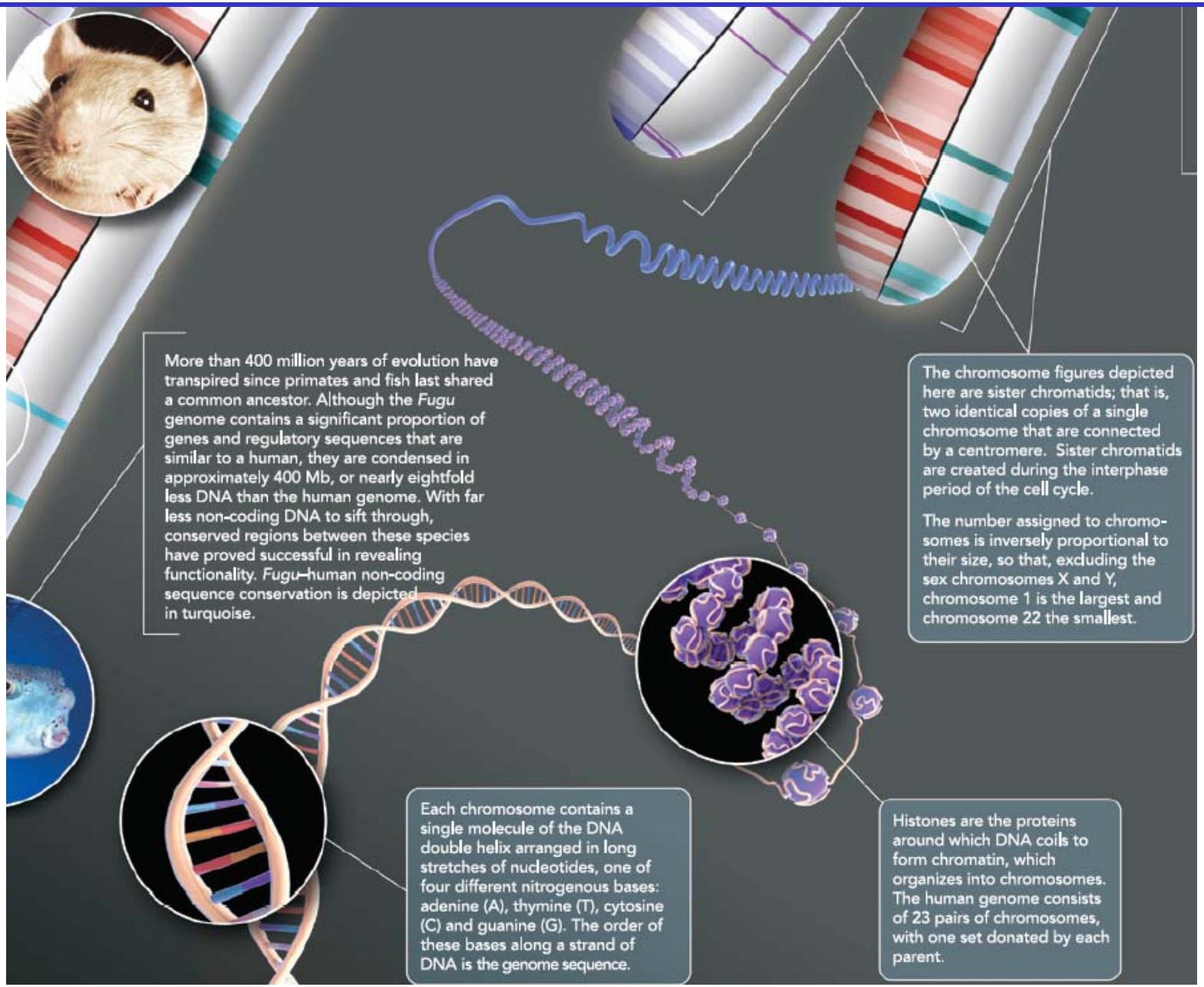
single page of instructions

devices, materials, machines



For the rest of the time we will be discussing the structure and function of DNA and proteins

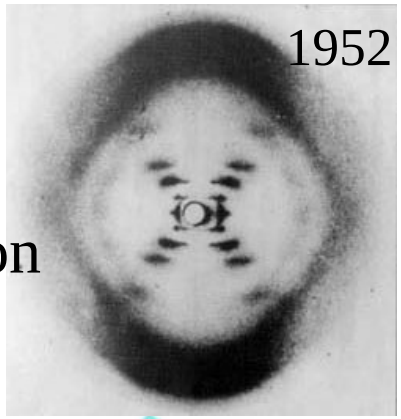
DNA: Instructions for an organism

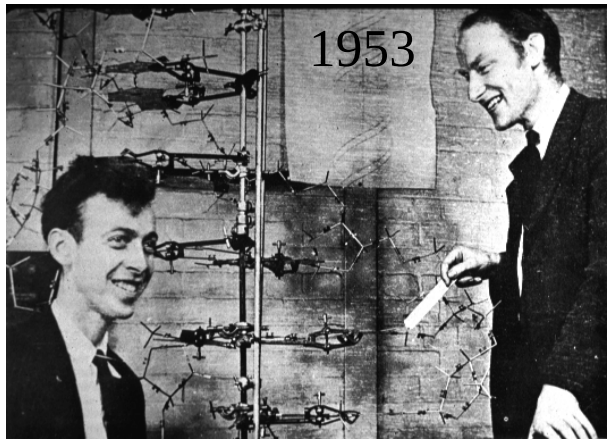


The structure of DNA: revealed by a back and forth between experiment and theory, creating new layers of understanding.

Experiment — Modeling/simulation

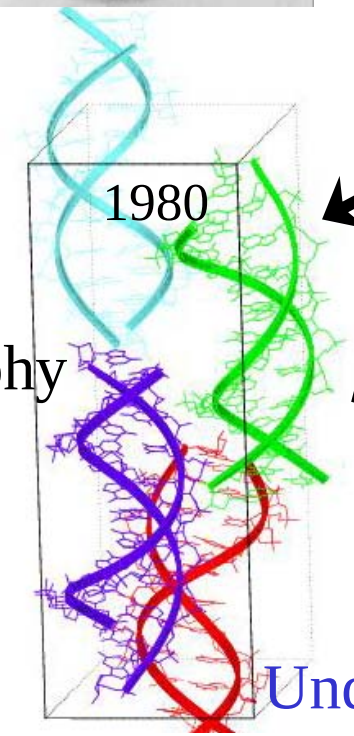
X-ray diffraction

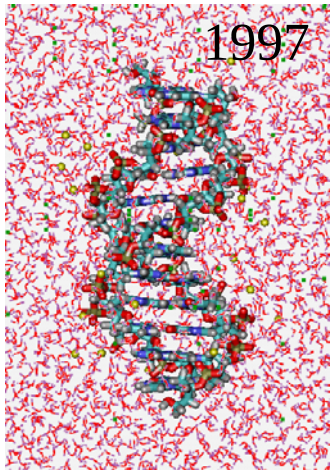




Molecular modeling

X-ray crystalography

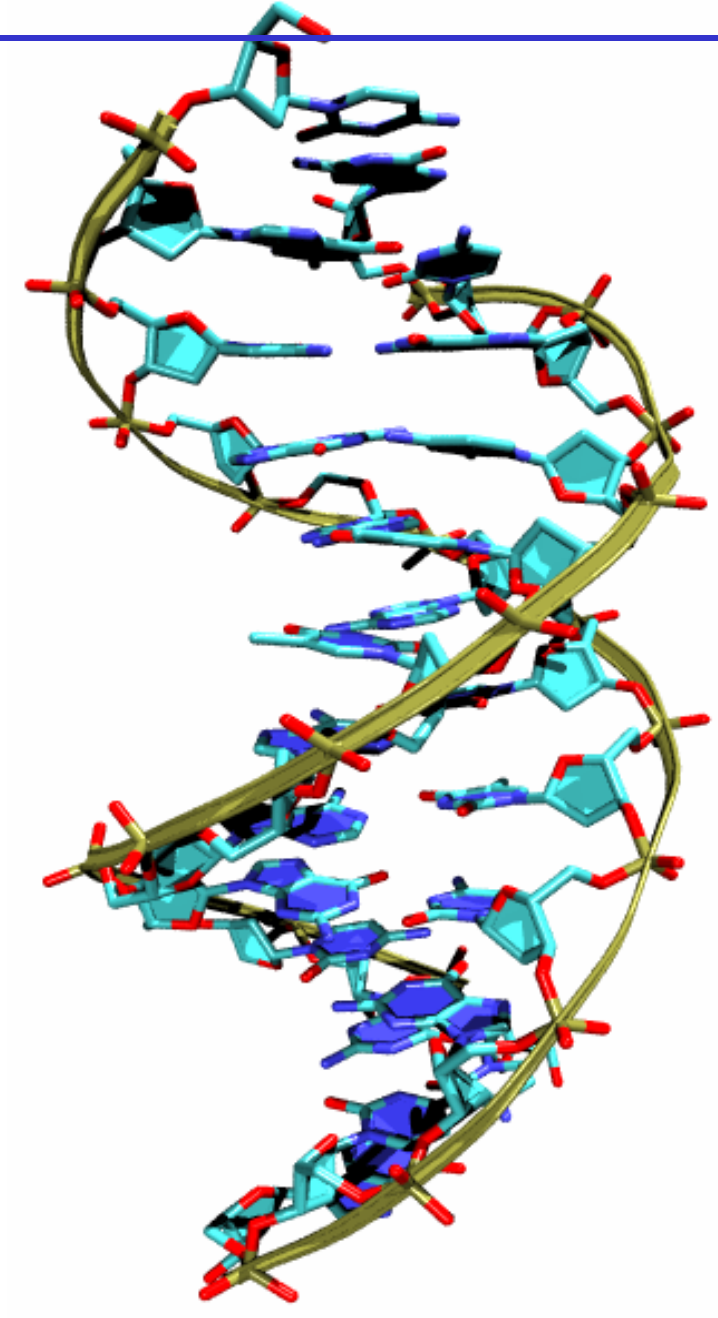


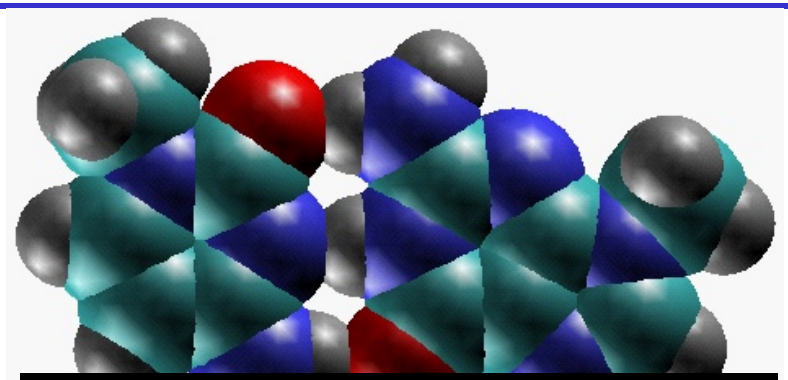


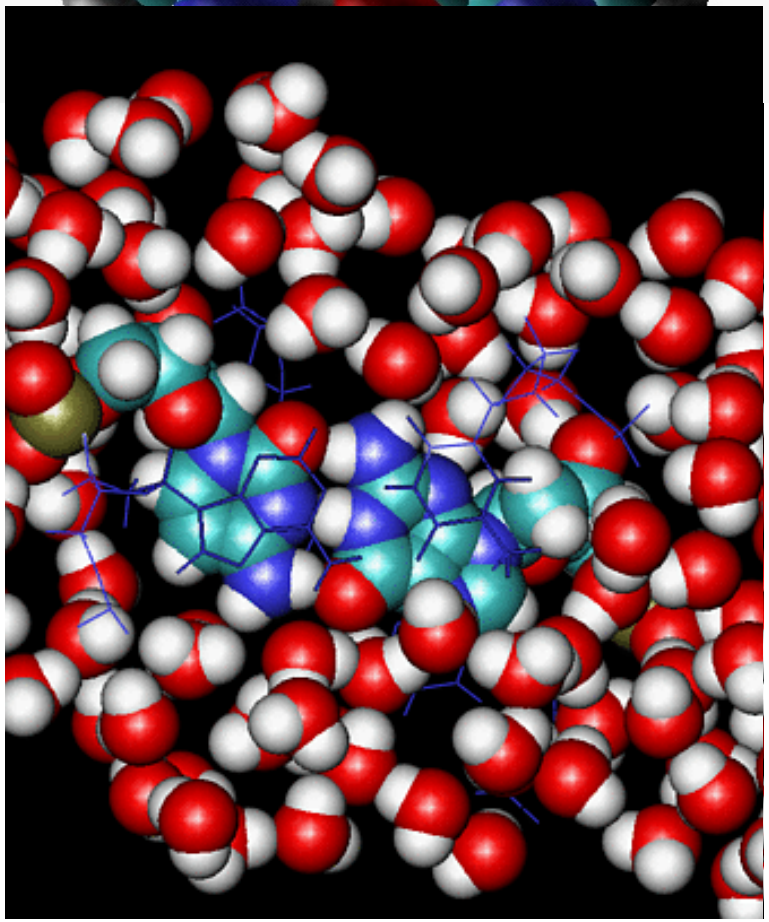
Molecular dynamics simulation

Understanding

Structure of DNA: Cell's Instructions

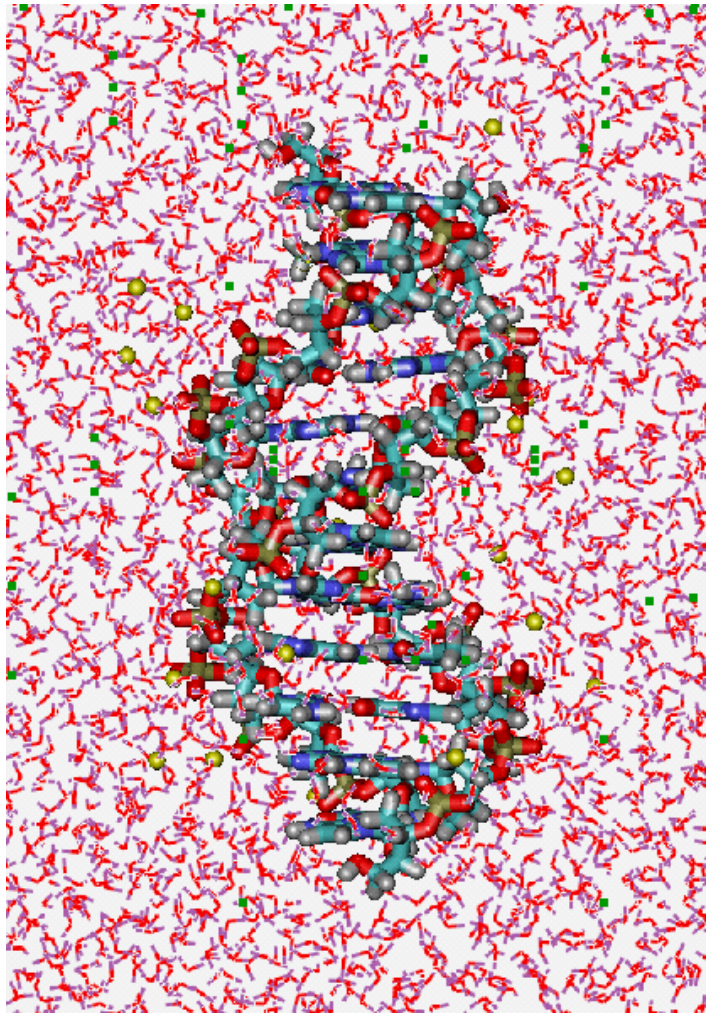


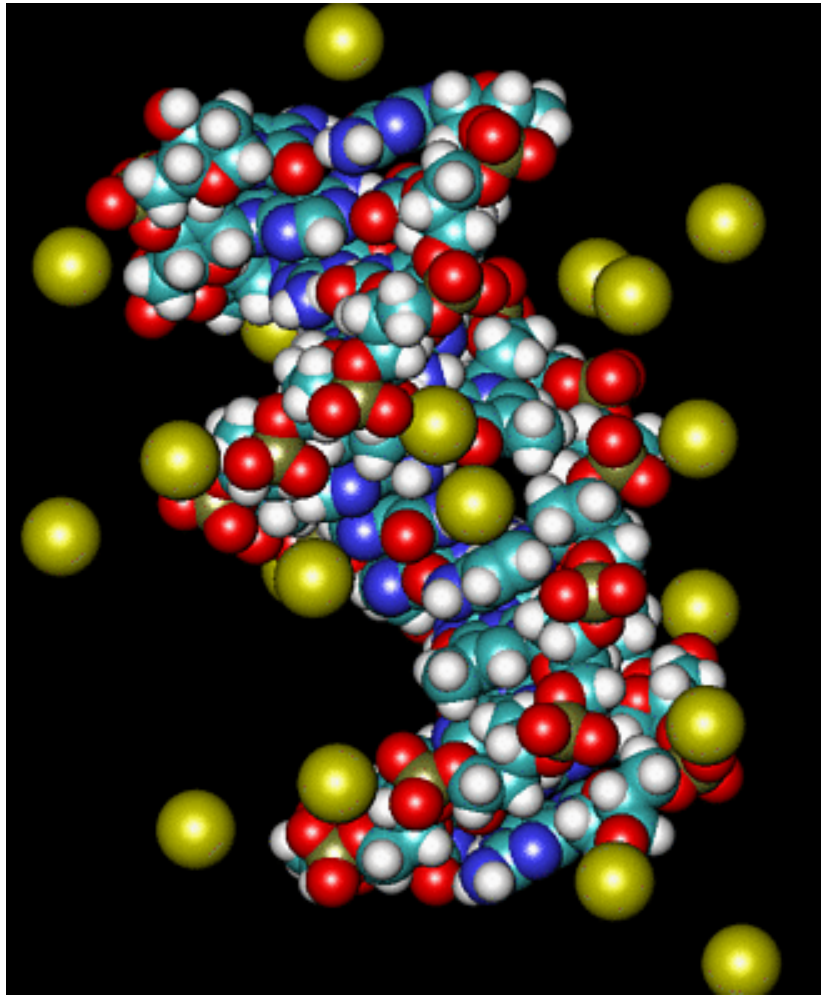




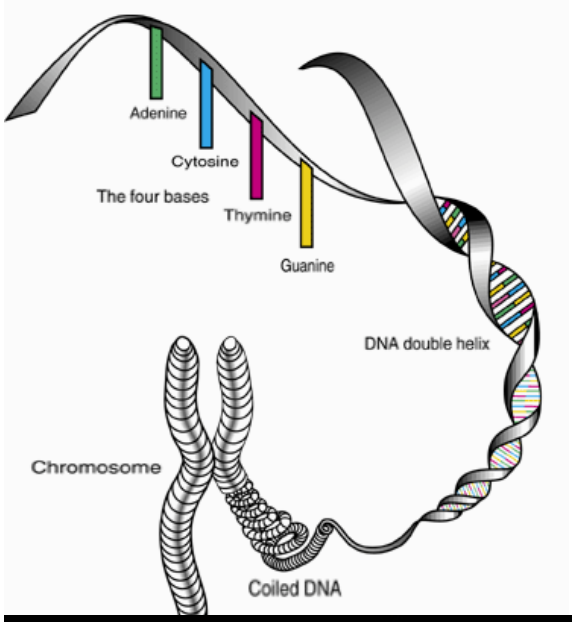
Understanding DNA damage by molecular dynamics simulation.

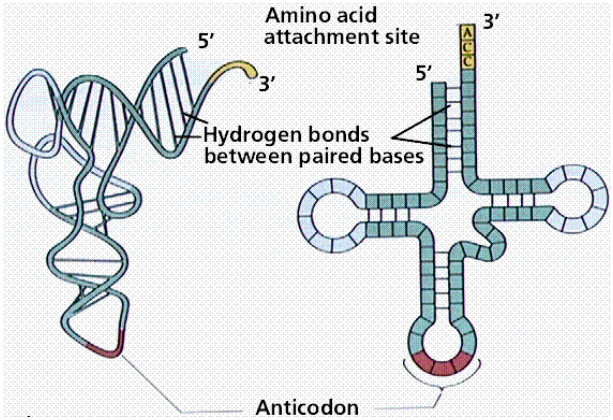
40,000 DNA bases lost per cell everyday!!





Proteins: from the chromosome (DNA), through mRNA and tRNA to the **ribosomes**, the protein factories.





DNA

↓ transcription

mRNA

+

tRNA

↓

proteins

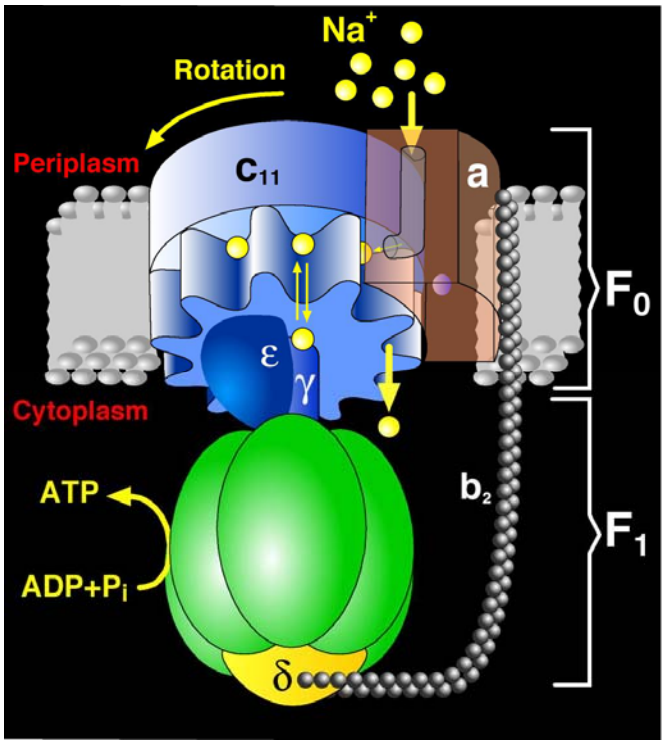
translation by the ribosome

X-ray crystallographic picture of a complete ribosome. The bacterial ribosome is composed of three different RNA molecules and more than 50 different to form the complete ribosome.

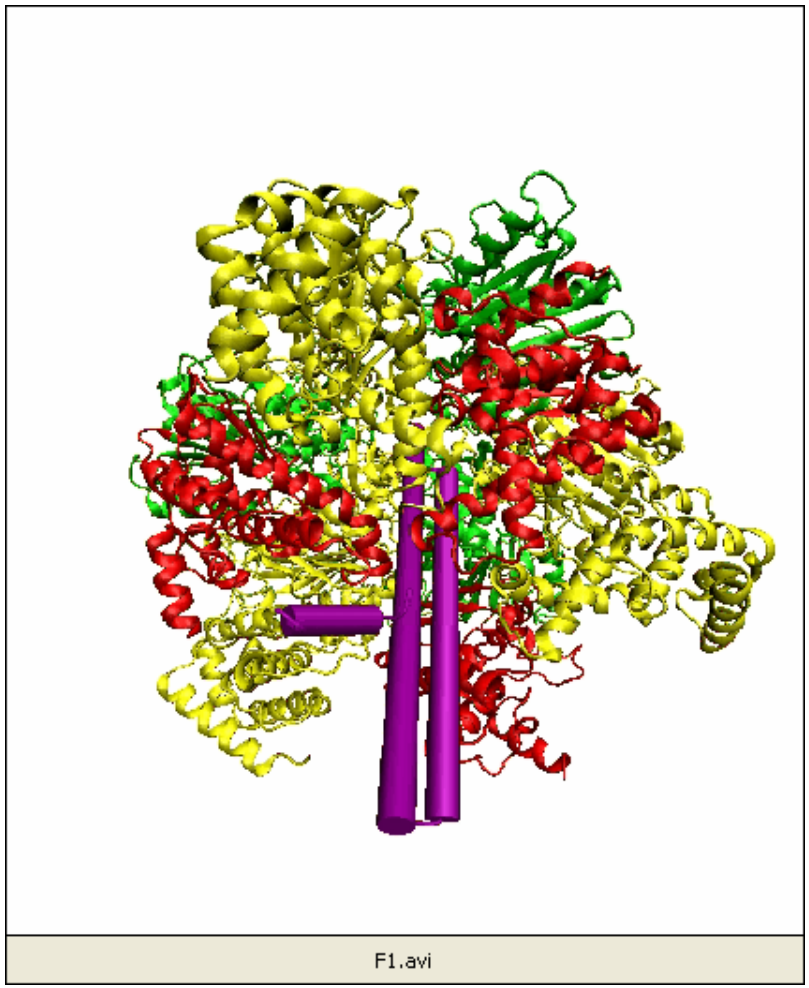
In this image, the large subunit is gray, the small subunit is violet (16S rRNA + 21 proteins), and the three transfer RNAs are green, blue, and red. [Cate et al. *Science* 1999.]

Structure of proteins: a huge variety of structures and functions

$$\text{ATP} \underset{\text{Synthesis ATP}}{\overset{\text{Pump ion}}{\rightleftharpoons}} \text{ADP} + \text{Pi}$$



THE F_1F_o ATPase

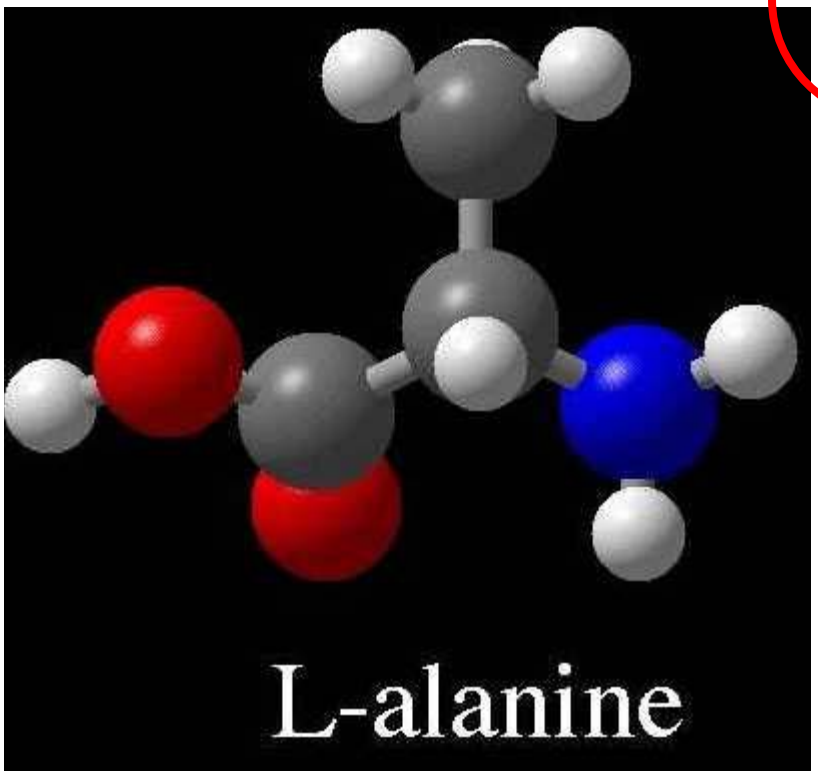


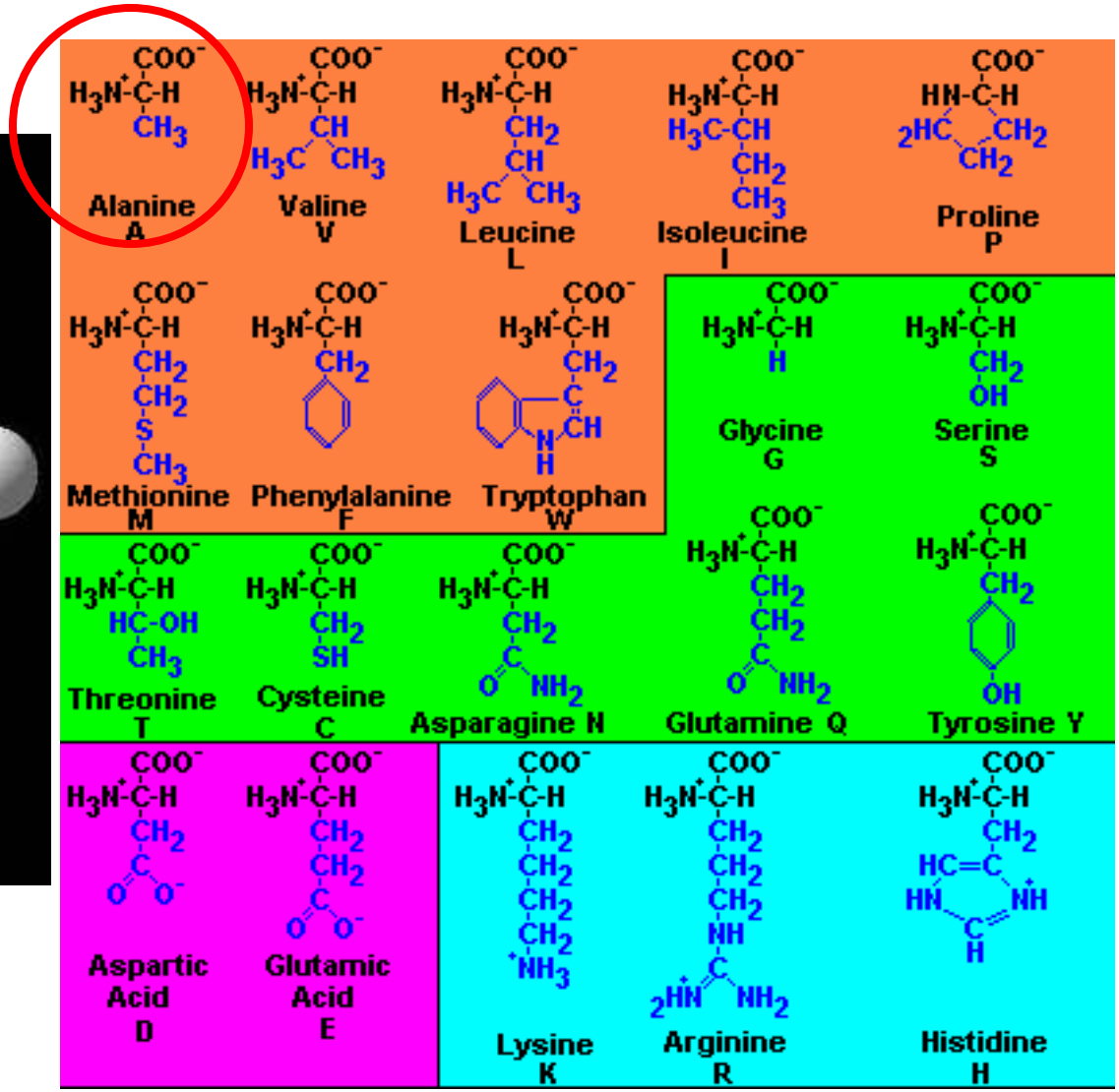
THE F_1 PART

nature.berkeley.edu/~xing

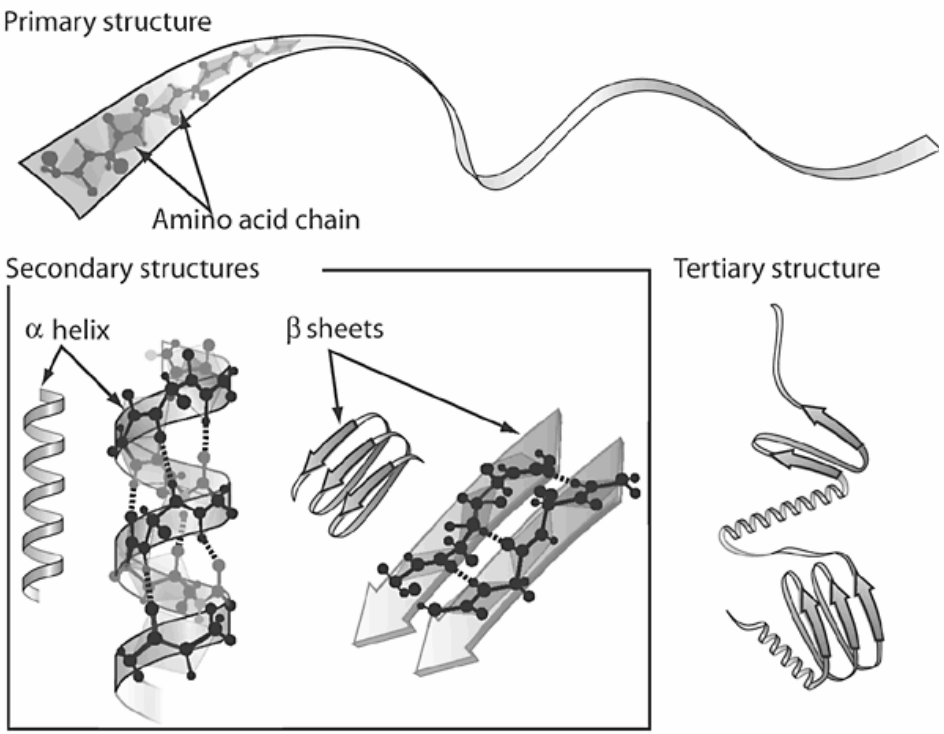
Check out the molecules of the month at **www.pdb.org**

Building blocks of proteins

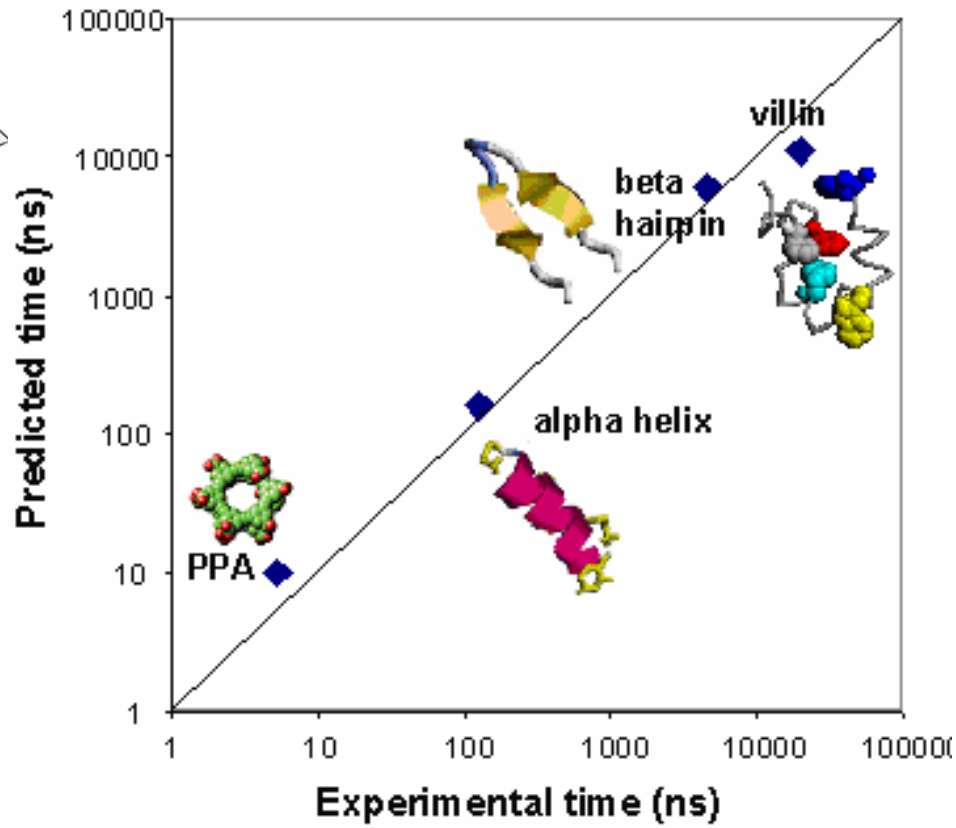




Structure of proteins 1o, 2o, 3o, 4o

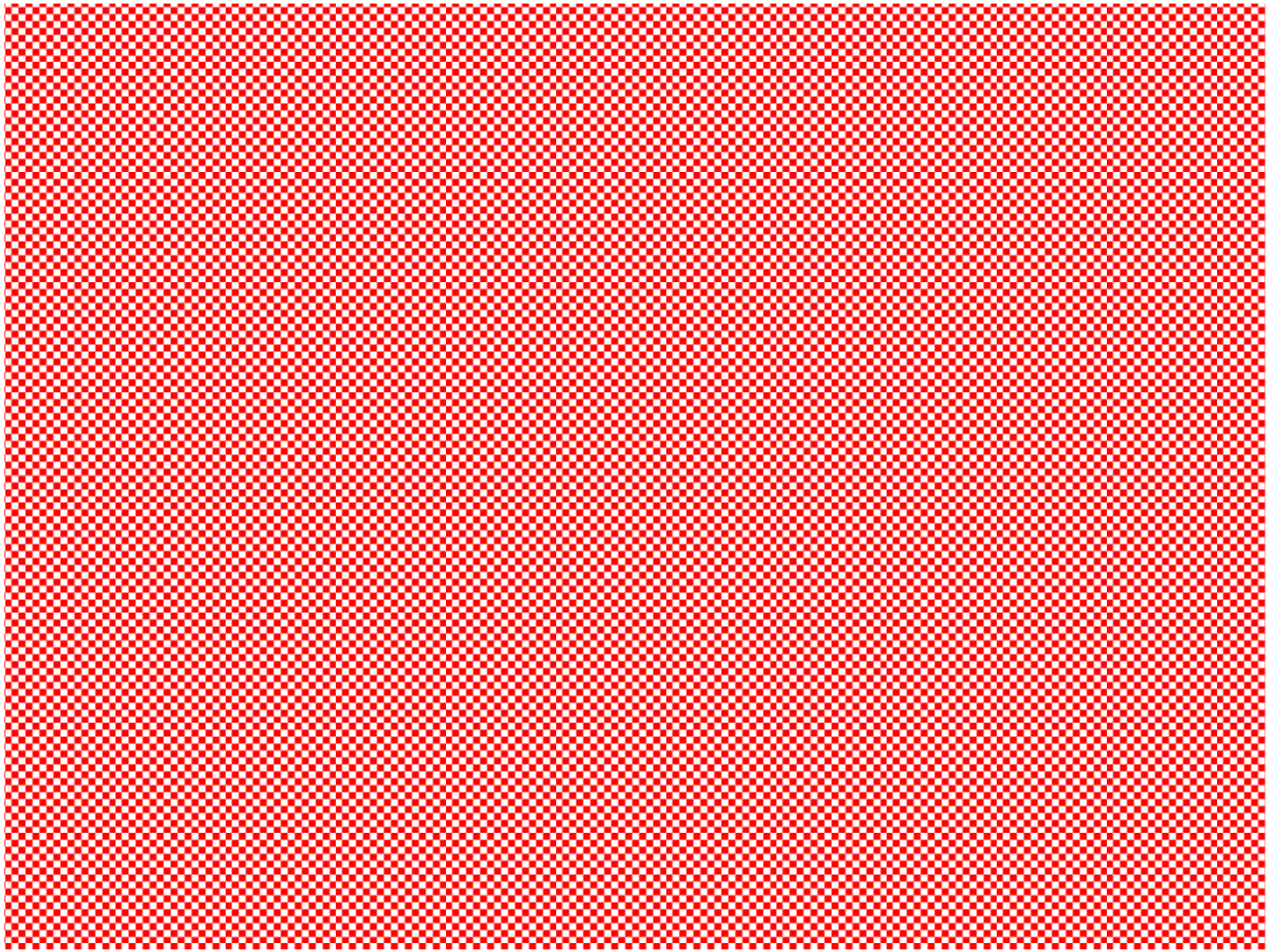


"folding@home"



folding.stanford.edu

What do proteins see? Water everywhere!



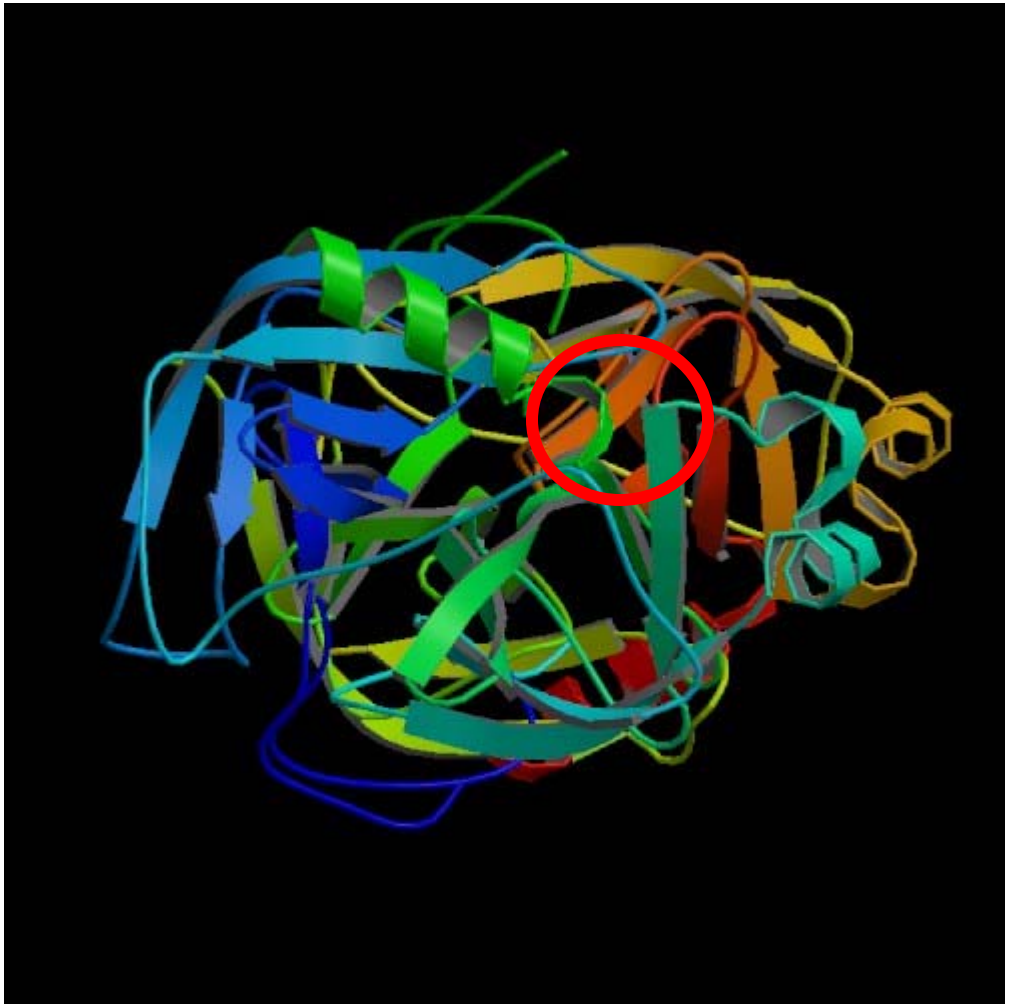
Proteins cannot “see” at all.

IV. ACTIVITY! Making a human protein: a digestive enzyme

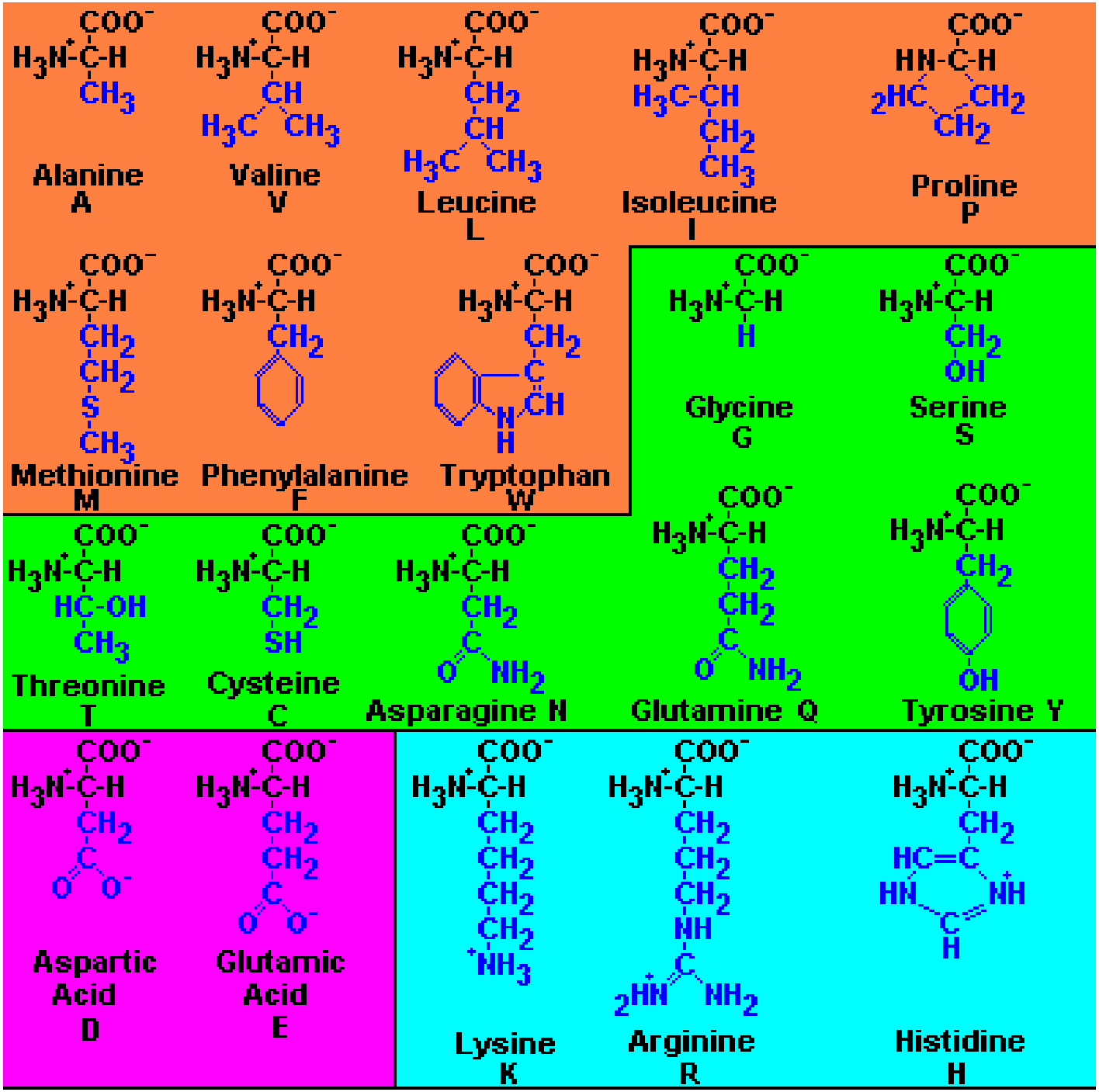
The catalytic triad:
Asp 102 – His 57 – Ser 195

```
CGVPAIQPVLSGLXXIVNGE
EAVPGSWPWQVSLQDKTGFH
FCGGSLINENWVVTAAHCGV
TTSDVVVAGEFDQGSSSEKI
QKLKIAKVFKNSKYNSLTIN
NDITLLKLSTAASFSQTVSA
VCLPSASDDFAAGTTCVTTG
WGLTRYXXANTPDRLQQASL
PLLSNTNCKKYWGTKIKDAM
ICAGASGVSSCMGDSGGPLV
CKKNGAWTLVGIVSWGSSTC
STSTPGVYARVTALVNWVQQ
TLAAN
```

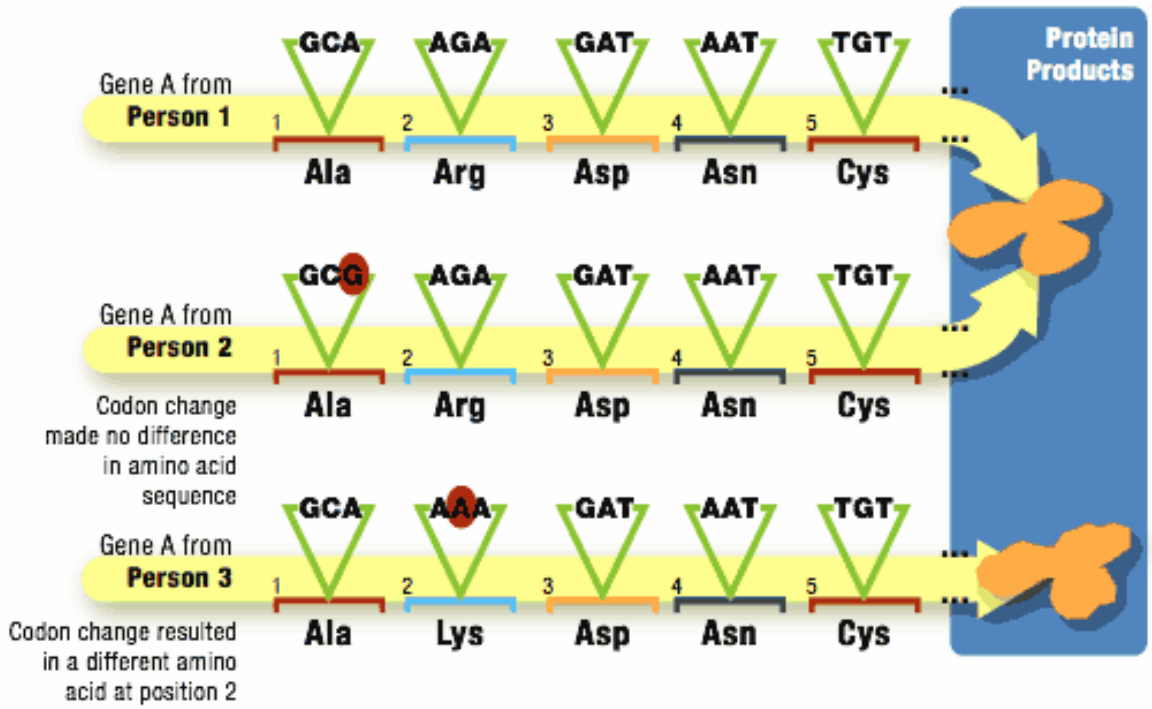
α-chymotrypsin



http://www.pdb.org/ 3VGC


Alanine
A
Valine
V
Leucine
L
Isoleucine
I
Proline
P
Methionine
M
Phenylalanine
F
Tryptophan
W
Glycine
G
Serine
S
Threonine
T
Cysteine
C
Asparagine N
Glutamine Q
Tyrosine Y
Aspartic
Acid
D
Glutamic
Acid
E
Lysine
K
Arginine
R
Histidine
H

The Human Genome: Mutations in DNA may or may not change proteins



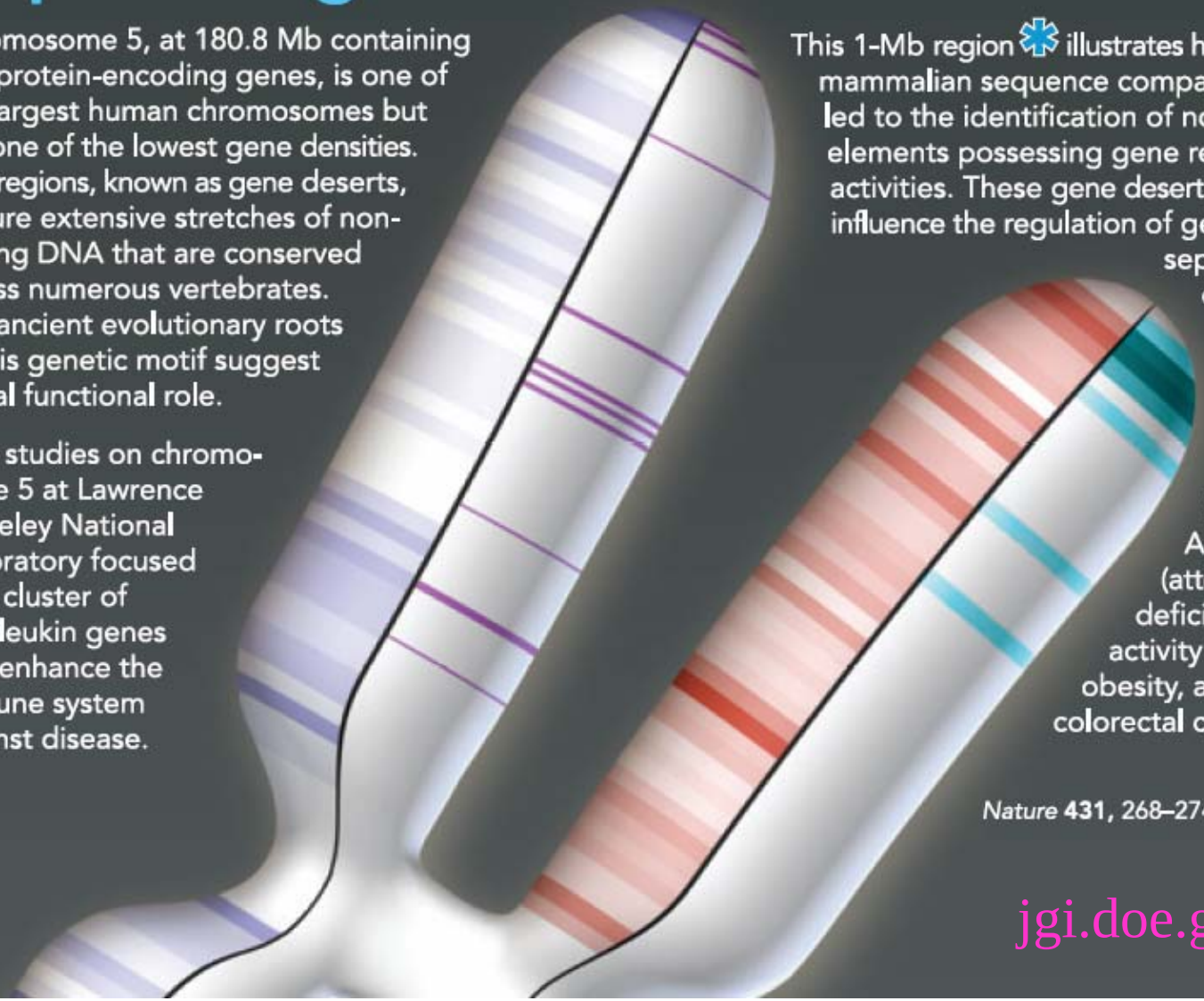
When change in DNA changes protein sequence, a “mutant” form of protein is produced. This may or may not affect the protein function, and it may help or hurt the survival of the organism.

CHROMOSOME 5

Expansive gene 'desert'

Chromosome 5, at 180.8 Mb containing 923 protein-encoding genes, is one of the largest human chromosomes but has one of the lowest gene densities. Vast regions, known as gene deserts, feature extensive stretches of non-coding DNA that are conserved across numerous vertebrates. The ancient evolutionary roots of this genetic motif suggest a vital functional role.

Pilot studies on chromosome 5 at Lawrence Berkeley National Laboratory focused on a cluster of interleukin genes that enhance the immune system against disease.



This 1-Mb region * illustrates how multi-mammalian sequence comparisons have led to the identification of non-coding elements possessing gene regulatory activities. These gene deserts seem to influence the regulation of genes separated by distances of as much as 1 Mb or more. Genes of interest include ADHD (attention-deficit/hyper-activity disorder), obesity, asthma and colorectal cancer.

Nature **431**, 268–274 (2004)

jgi.doe.gov

The Human Genome… by the numbers

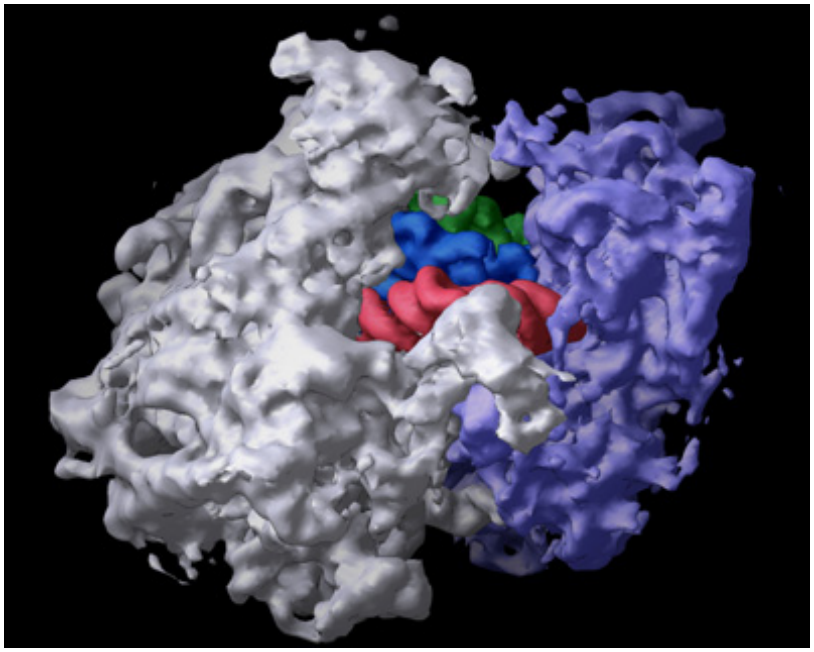
75-100 trillion . . . Cells in the human body
3.1 billion . . . Base pairs in each cell
2.4 million . . . Base pairs in the largest human gene (dystrophin)
28,000-35,000 . . . Genes in the human genome

SPECIES	CHROMOSOMES	GENES	BASE PAIRS
Human *(Homo sapiens)*	46 (23 pairs)	28-35,000	~3.1 billion
Mouse *(Mus musculus)*	40	22.5-30,000	~2.7 billion
Pufferfish *(Fugu rubripres)*	44	~31,000	~365 million
Malaria Mosquito *(Anopheles gambiae)*	6	~14,000	~289 million
Sea Squirt *(Ciona intestinalis)*	28	~16,000	~160 million
Fruit Fly *(Drosophila melanogaster)*	8	~14,000	~137 million
Roundworm *(C. elegans)*	12	19,000	~97 million
Bacterium *(E. coli)*	1*	~5,000	~4.1 million

Bacterial chromosomes are **chromonemes, not true chromosomes.*

Ribosomal RNA is the gold standard for phylogeny.

1. All organisms must have protein production machinery (ribosomes).
2. RNA is a direct translation of DNA and rRNA is non-redundant (unlike the codons of mRNA).
3. 16S rRNA (within the small subunit) is most commonly compared



X-ray crystallographic picture of a complete ribosome. The bacterial ribosome is composed of three different RNA molecules and more than 50 different to form the complete ribosome.

In this image, the large subunit is gray, the small subunit is violet (16S rRNA + 21 proteins), and the three transfer RNAs are green, blue, and red. [Cate et al. *Science* 1999.]

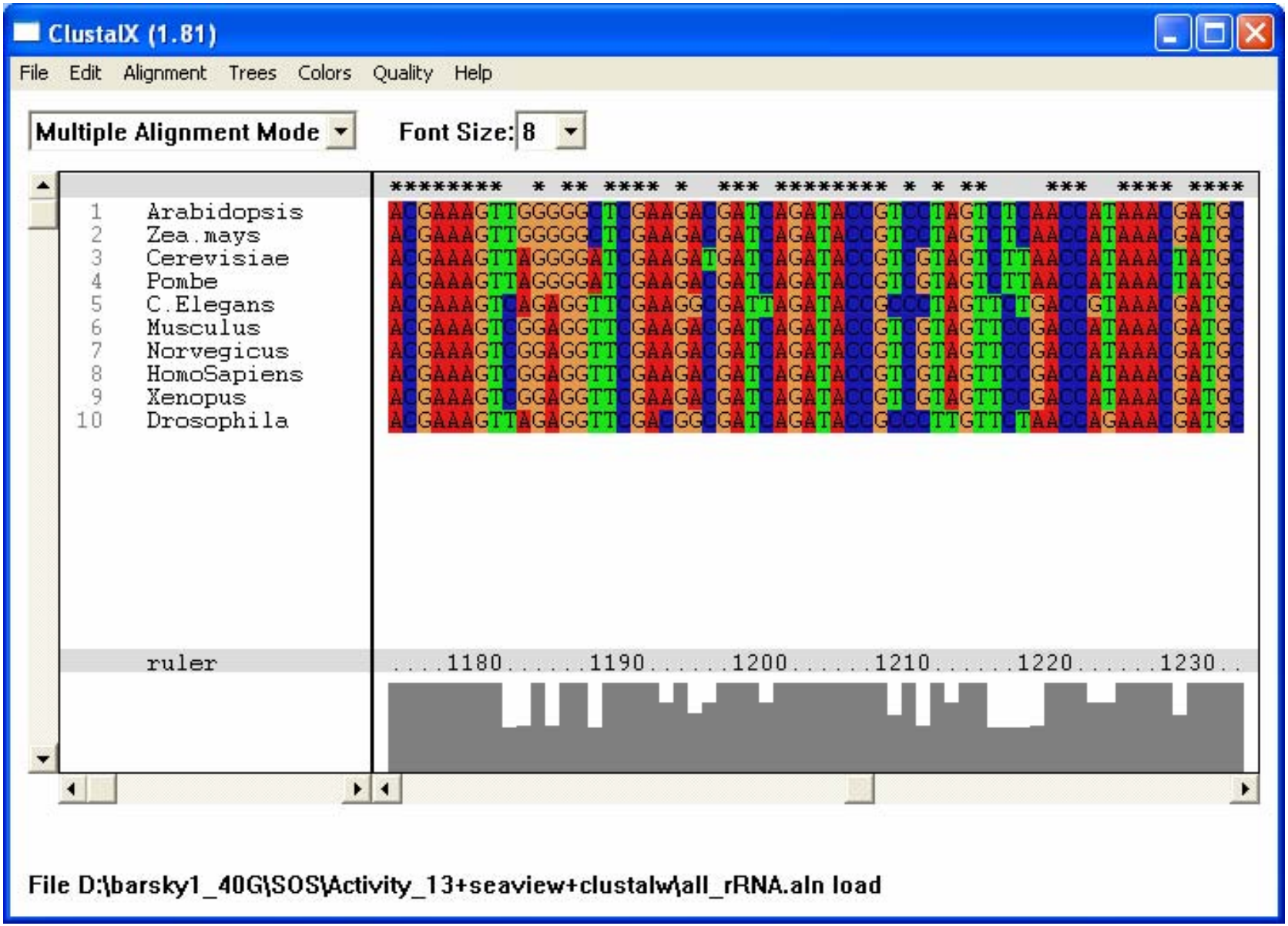
The genome comparisons: 16S rRNA from a wide range of species

>Drosophila
TCTAGATAACATGCAGATCGTATGGTCTTGTACCGACGACAGATCTTTCAAATGTCTGCCCTATCAACTT
TTGATGGTAGTATCTAGGACTACCATGGTTGCAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGG
GAGCCTGAGAAACGGCTACCACATCTAAGGAAGGCAGCAGGCGCGTAAATTACCCACTCCCAGCTCGGGG
AGGTAGTGACGAAAAATAACAATACAGGACTCATTTCCGAGGCCCTGTAATTGGAATGAGTACACTTTAA
ATCCTTTAACAAGGACCAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC
GTATATTAAAGTTGTTGCGGTTAAAACGTTCGTAGTTGAACTTGTGCTTCATACGGGTAGTACAACTTAC
AATTGTGGTTAGTACTATACCTTTATGTATGTAAGCGTATTACCGGTGGAGTTCTTATATGTGATTAAAT
ACTTGTATTTTTTCATATGTTCCTCCTATTTAAAAACCTGCATTAGTGCTCTTAAACGAGTGTTATTGTG
GGCCGGTACTATTACTTTGAACAAATTAGAGTGCTTAAAGCAGGCTTCAAATGCCTGAATATTCTGTGCA
TGGGATAATGAAATAAGACCTCTGTTCTGCTTTCATTGGTTTTCAGATCAAGAGGTAATTATTAATAGAA
GCAGTTTGAGGGCTTAAGACTAACTTAAGCGAAAGCATTTGCCAAAGATGTTTTCATTAATCAAGAACGA
AAGTTAGAGGTTCGACGGCGATCAGATACCGCCCTTGTTCTAACCAGAAACGATGCCAGCTAGCAATTGG
GTGTAGCTACTTTTATGGCTCTCTCAGTCGCTTCCCGGAAACCAAAGCTTTTGTGCTCCGGGGGAAGTAT
GGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTT
GACTCAACACGGGAAAACTTACCAGGTCCGAACATAAGTGTGTAAGACAGATTGATAGCTCTTTCTCGAA
TCTATGGGTGGTGGTGCATGGGCGTTCTTAGTTCGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAAC
GAGACTCAAATATATTAAATAGATATCTTCAGGATTATGGTGCTGAAGCTTATGTAGCCTTCATTCATGT
TGGCAGTAAAATGCTTATTGTGTTTGAATGTGTTTATGTAAGTGGAGCCGTACCTGTTGGTTTGTCCCAT
TATAAGGACACTAGCTTCTTAAATGGATAAATTGCGTCTTGCAATAATGAGATTGGACAATAACAGGTCT
GTGATGCACTTAGATGTCCTGGGCTGCGTGCAGGCTACAATGAAAGTATCAACGTGTACACGTTTCTAGA
CCGAGAGGTCCGGGTAAAACGCTGAACCACTTTCATGCTCGGAATTGTGAACTGAAACTGTTCACATGAA
CTTGGAATTTCCATTAACTCGCATTCATTACGTCCCAGCCCTGTGTACACACTGCCGTCGCTACTACCGA
TTGAATTATTTAGTGAGGCCTCCGGA

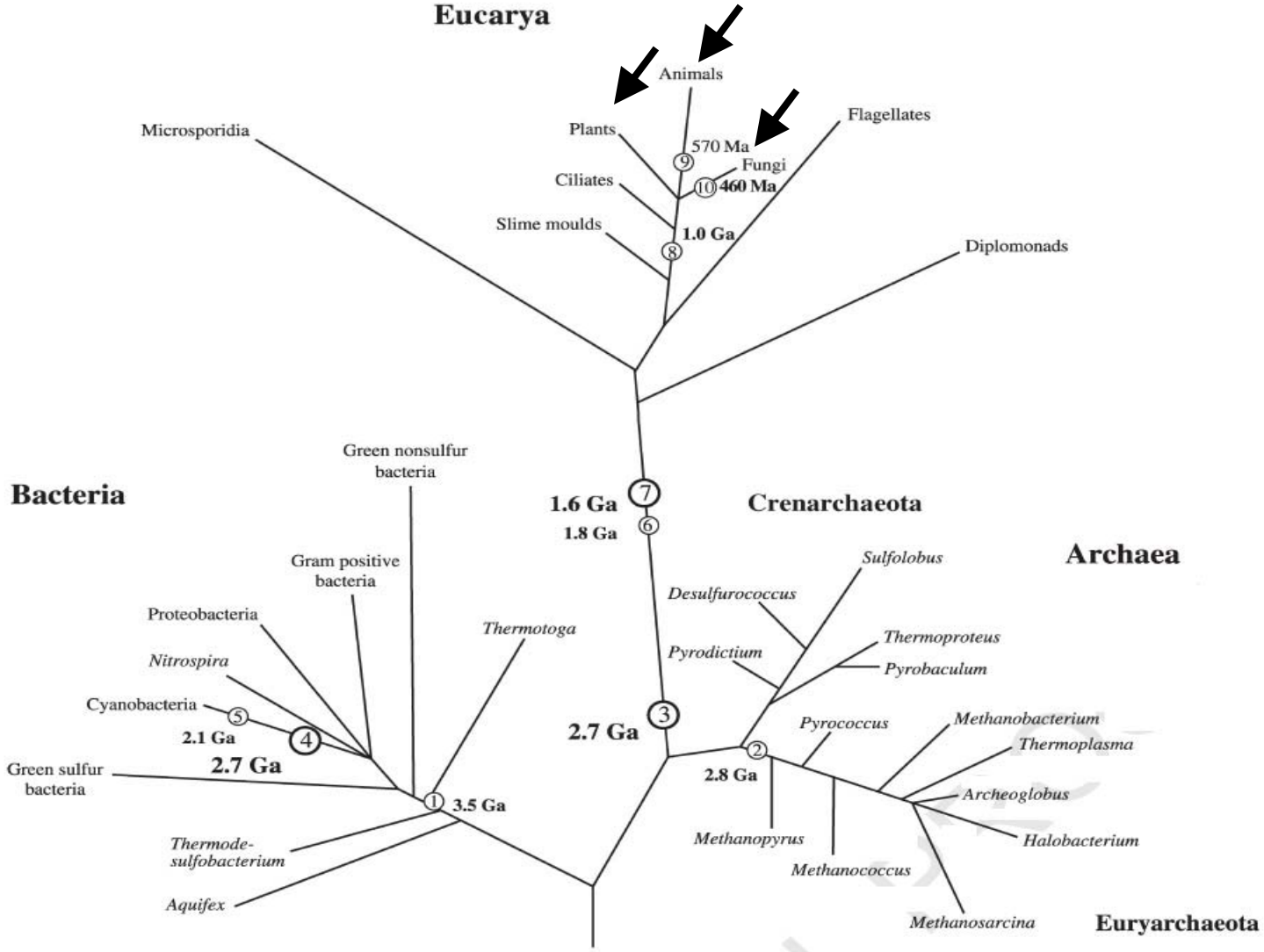
>HomoSapiens
CCGTCCGTCCGTCGTCCTCCTCGCTTGCGGGGCGCCGGGCCCGTCCTCGAGCCCCCNNNNNCCGTCCGGC
CGCGTCGGGGCCTCGCCGCGCTCTACCTACCTACCTGGTTGATCCTGCCAGTAGCATATGCTTGTCTCAA
AGATTAAGCCATGCATGTCTAAGTACGCACGGCCGGTACAGTGAAACTGCGAATGGCTCATTAAATCAGT
TATGGTTCCTTTGGTCGCTCGCTCCTCTCCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCC
GACGGGCGCTGACCCCCTTCGCGGGGGGGATGCGTGCATTTATCAGATCAAAACCAACCCGGTCAGCCCC
TCTCCGGCCCCGGCCGGGGGGCGGGCCGCGGCGGCTTTGGTGACTCTAGATAACCTCGGGCCGATCGCAC
GCCCCCCGTGGCGGCGACGACCCATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCCGTGCCTA
CCATGGTGACCACGGGTGACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCAC
ATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCACTCCCGACCCGGGGAGGTAGTGACGAAAAATAACAA
TACAGGACTCTTTCGAGGCCCTGTAATTGGAATGAGTCCACTTTAAATCCTTTAACGAGGATCCATTGGA
GGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGCTGCAGTTAA
AAAGCTCGTAGTTGGATCTTGGGAGCGGGCGGGCGGTCCGCCGCGAGGCGAGCCACCGCCCGTCCCCGCC
CCTTGCCTCTCGGCGCCCCCTCGATGCTCTTAGCTGAGTGTCCCGCGGGGCCCGAAGCGTTTACTTTGAA
AAAATTAGAGTGTTCAAAGCAGGCCCGAGCCGCCTGGATACCGCAGCTAGGAATAATGGAATAGGACCGC
GGTTCTATTTTGTTGGTTTTCGGAACTGAGGCCATGATTAAGAGGGACGGCCGGGGGCATTCGTATTGCG
CCGCTAGAGGTGAAATTCTTGGACCGGCGCAAGACGGACCAGAGCGAAAGCATTTGCCAAGAATGTTTTC
ATTAATCAAGAACGAAAGTCGGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCCGACCATAAACGATG
CCGACCGGCGATGCGGCGGCGTTATTCCCATGACCCGCCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGG
TTCCGGGGGGAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAG
CCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCCGGCCCGGACACGGACAGGATTGACAGATTGA
TAGCTCTTTCTCGATTCCGTGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTA
ATTCCGATAACGAACGAGACTCTGGCATGCTAACTAGTTACGCGACCCCCGAGCGGT
CTTCTTAGAGGGACAAGTGGCGTTCAGCCACCCGAGATTGAGCAATAACAGGTCTGT
GTCCGGGGCTGCACGCGCGCTACACTGACTGGCTCAGCGTGTGCCTACCCTACGCCG
ACCCGTTGAACCCCATTCGTGATGGGGATCGGGGATTGCAATTATTCCCCATGAACG
AAGTGCGGGTCATAAGCTTGCGTTGATTAAGTCCCTGCCCTTTGTACACACCGCCCGT
TTGGATGGTTTAGTGAGGCCCTCGGATCGGCCCCGCCGGGGTCGGCCCACGGCCTGG
AGACGGTCGAACTTGACTATCTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGC
GGATCATTA

>Musculus
TACCTGGTTGATCCTGCCAGTAGCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
GCCGGTACAGTGAAACTGCGAATGGCTCATTAAATCAGTTATGGTTCCTTTGGTCGCT
ACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCCGACGGGCGCTGACCCCCC
ATGCGTGCATTTATCAGATCAAAACCAACCCGGTGAGCTCCCTCCCGGCTCCGGCCG
GGCGGCTTGGTGACTCTAGATAACCTCGGGCCGATCGCACGCCCCCCGTGGCGGCGA
CGTCTGCCCTATCAACTTTCGATGGTAGTCGCCGTGCCTACCATGGTGACCACGGGTC
GGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGC
CCCACTCCCGACCCGGGGAGGTAGTGACGAAAAATAACAATACAGGACTCTTTCGAC
AATGAGTCCACTTTAAATCCTTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGCCA
TTCCAGCTCCAATAGCGTATATTAAAGTTGCTGCAGTTAAAAAGCTCGTAGTTGGATC
GGGCGGTCCGCCGCGAGGCGAGTCACCGCCCGTCCCCGCCCCTTGCCTCTCGGCGCC
TAGCTGAGTGTCCCGCGGGGCCCGAAGCGTTTACTTTGAAAAAATTAGAGTGTTCAA
CGCCTGGATACCGCAGCTAGGAATAATGGAATAGGACCGCGGTTCTATTTTGTTGGT
GCCATGATTAAGAGGGACGGCCGGGGGCATTCGTATTGCGCCGCTAGAGGTGAAATT
AAGACGGACCAGAGCGAAAGCATTTGCCAAGAATGTTTTCATTAATCAAGAACGAA
AGACGATCAGATACCGTCGTAGTTCCGACCATAAACGATGCCGACTGGCGATGCGGC
TGACCCGCCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGGTTCCGGGGGGAGTATGC
ACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGGCCTGCGGCTTAATTTGAC
CCTCACCCGGCCCGGACACGGACAGGATTGACAGATTGATAGCTCTTTCTCGATTCC
CATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAG
AACTAGTTACGCGACCCCCGAGCGGTCGGCGTCCCCCAACTTCTTAGAGGGACAAGT
CCGAGATTGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCTGCACGCGC
GCTCAGCGTGTGCCTACCCTGCGCCGGCAGGCGCGGGTAACCCGTTGAACCCCATTC
GGGATTGCAATTATTCCCCATGAACGAGGAATTCCCAGTAAGTGCGGGTCATAAGCT
TCCCTGCCCTTTGTACACACCGCCCGTCGCTACTACCGATTGGATGGTTTAGTGAGGC
CCCGCCGGGGTCGGCCCACGGCCCTGGCGGAGCGCTGAGAAGACGGTCGAACTTGAC
TAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTA

The Genome comparisons



To see the vastness of biological diversity, think small.



The Universal Phylogenetic Tree with branch lengths and branching order based on SSU rRNA (Adapted from Brocks et al. 2004 *Treatise on Geochemistry*)

The end—so many answers…so many more questions!

Molecules of the month:
www.pdb.org

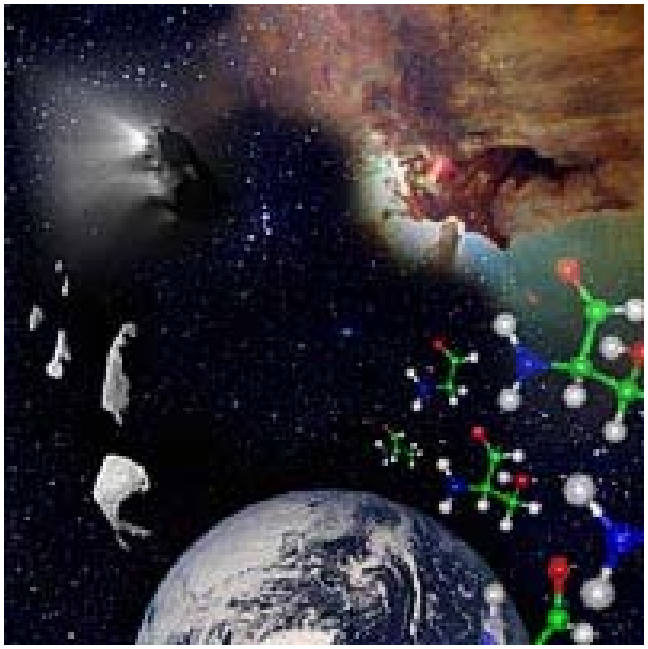
Genomes/bioinformatics:
www.ncbi.nih.gov
jgi.doe.org

me: compbio.llnl.gov/barsky

F1 ATPase:
nature.berkeley.edu/~xing

folding@home:
folding.standford.edu

Where did amino acids come from?



Contact me: **barsky@llnl.gov**

This presentation was created under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48. UCRL-PRES-220166