Quick Bio Review

CMSC 423

E. coli

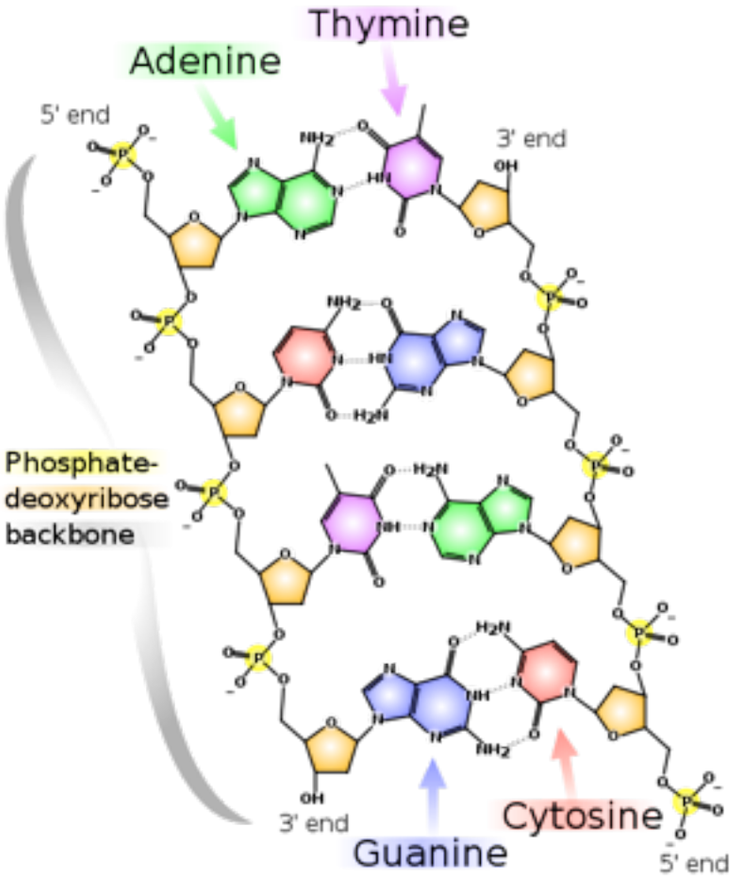
- E. coli is an example of a bacterium.

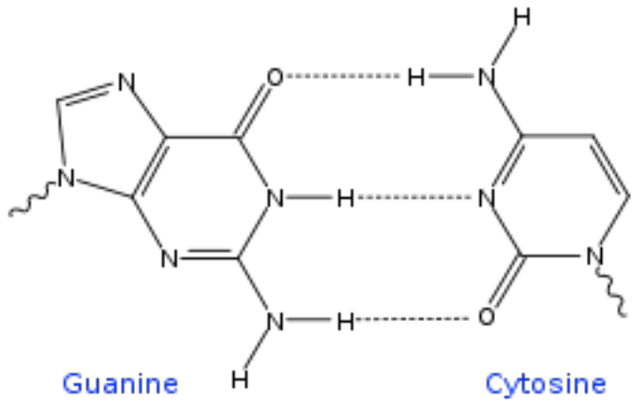
	% total dry weight
DNA	3.1
RNA	20.5
Protein	55.0
Lipid	9.1

} Algorithms are used to understand these important components.

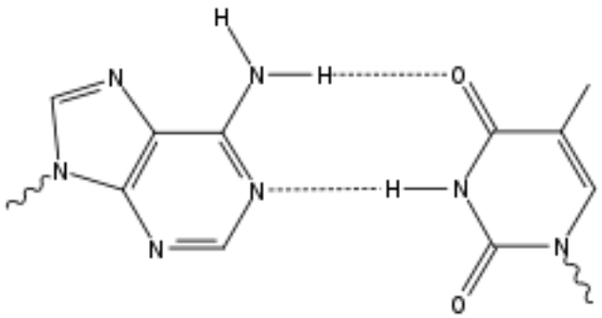
Christos Savva (Microscopy & Imaging Center) and Thomas Wood (Dept. of Chemical Engineering) at Texas A&M University.

DNA

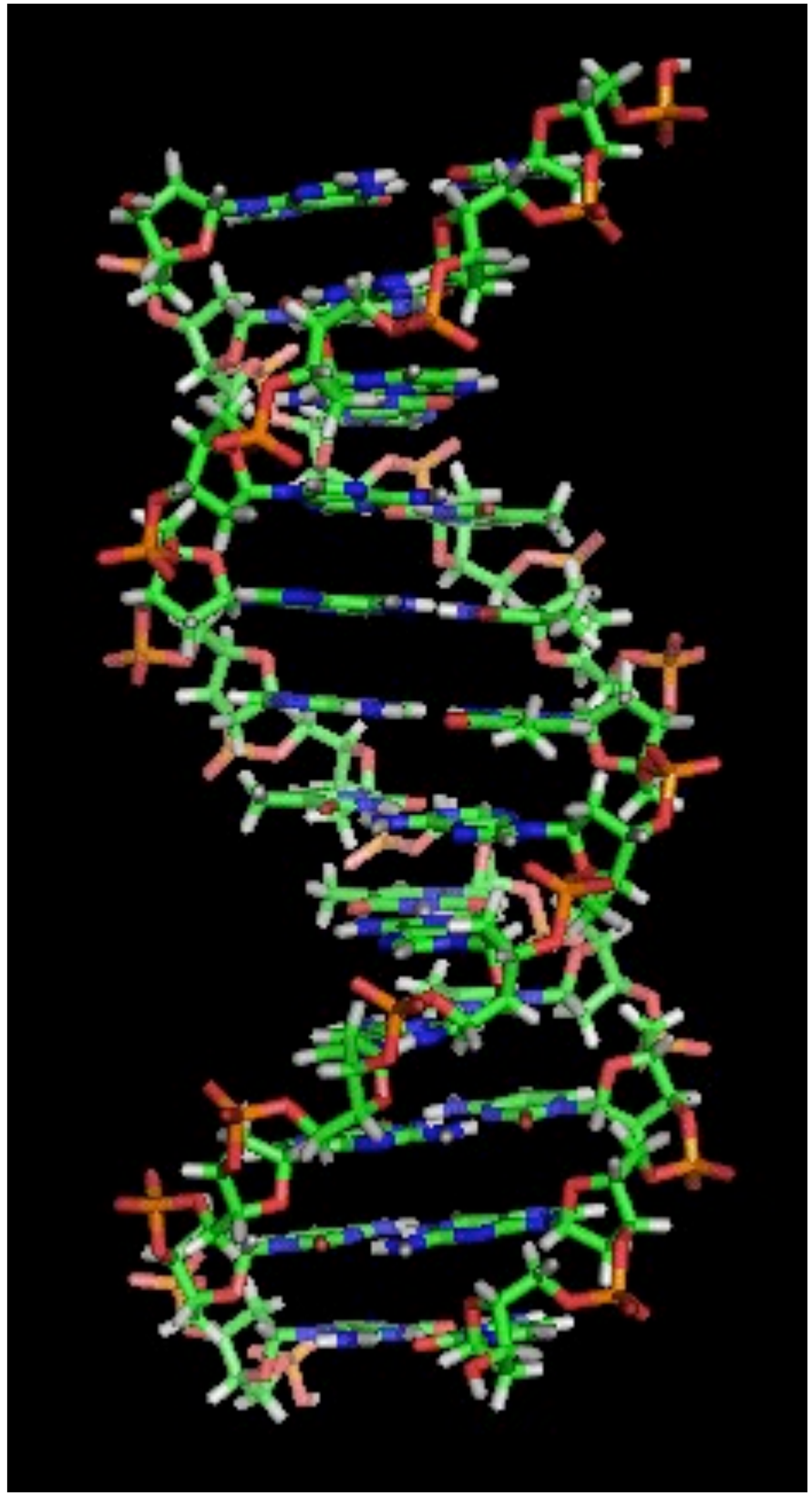

Thymine
Adenine
5' end
3' end
Phosphate-
deoxyribose
backbone
3' end
Cytosine
Guanine
5' end


Guanine
Cytosine

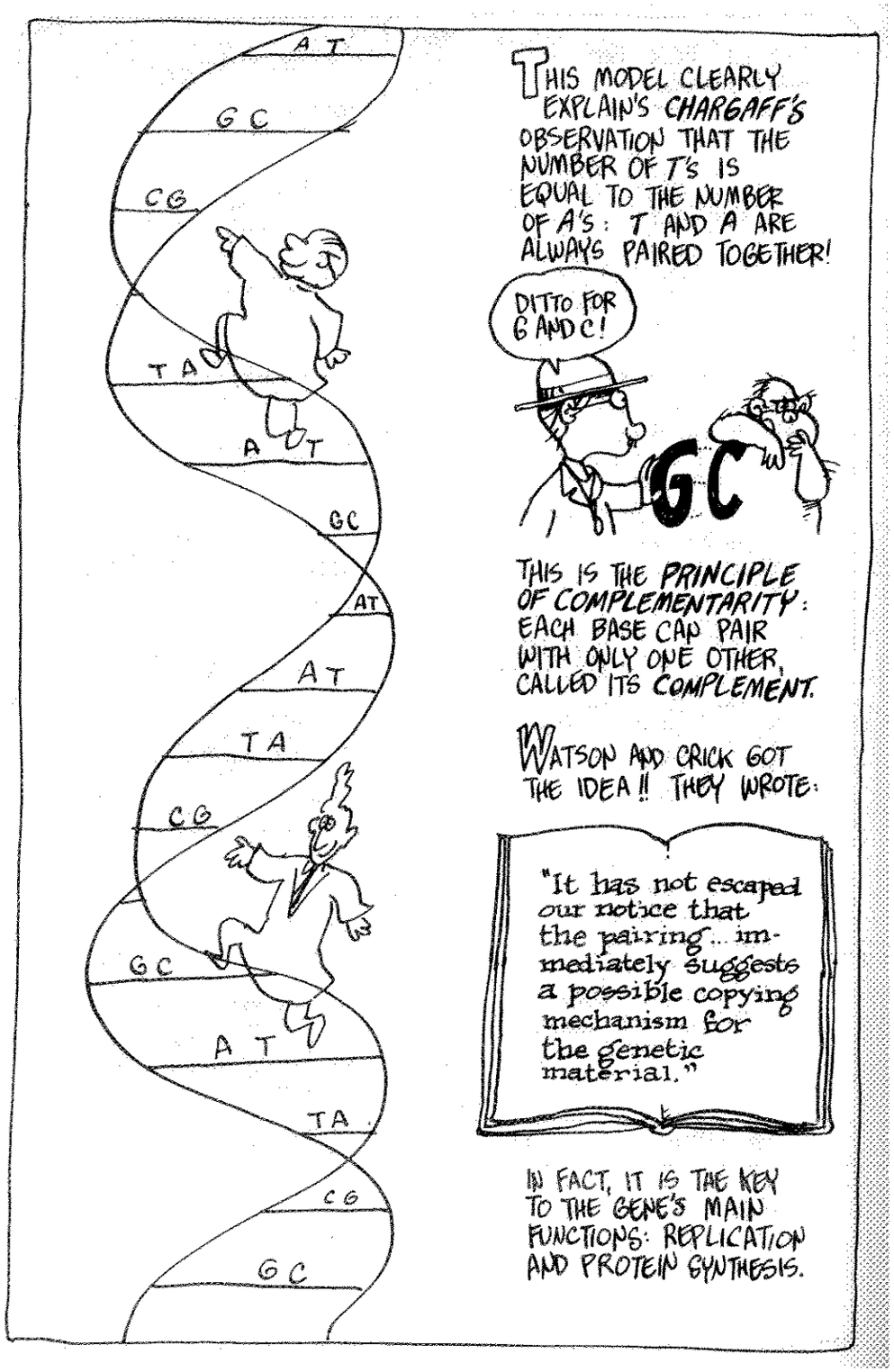
G C

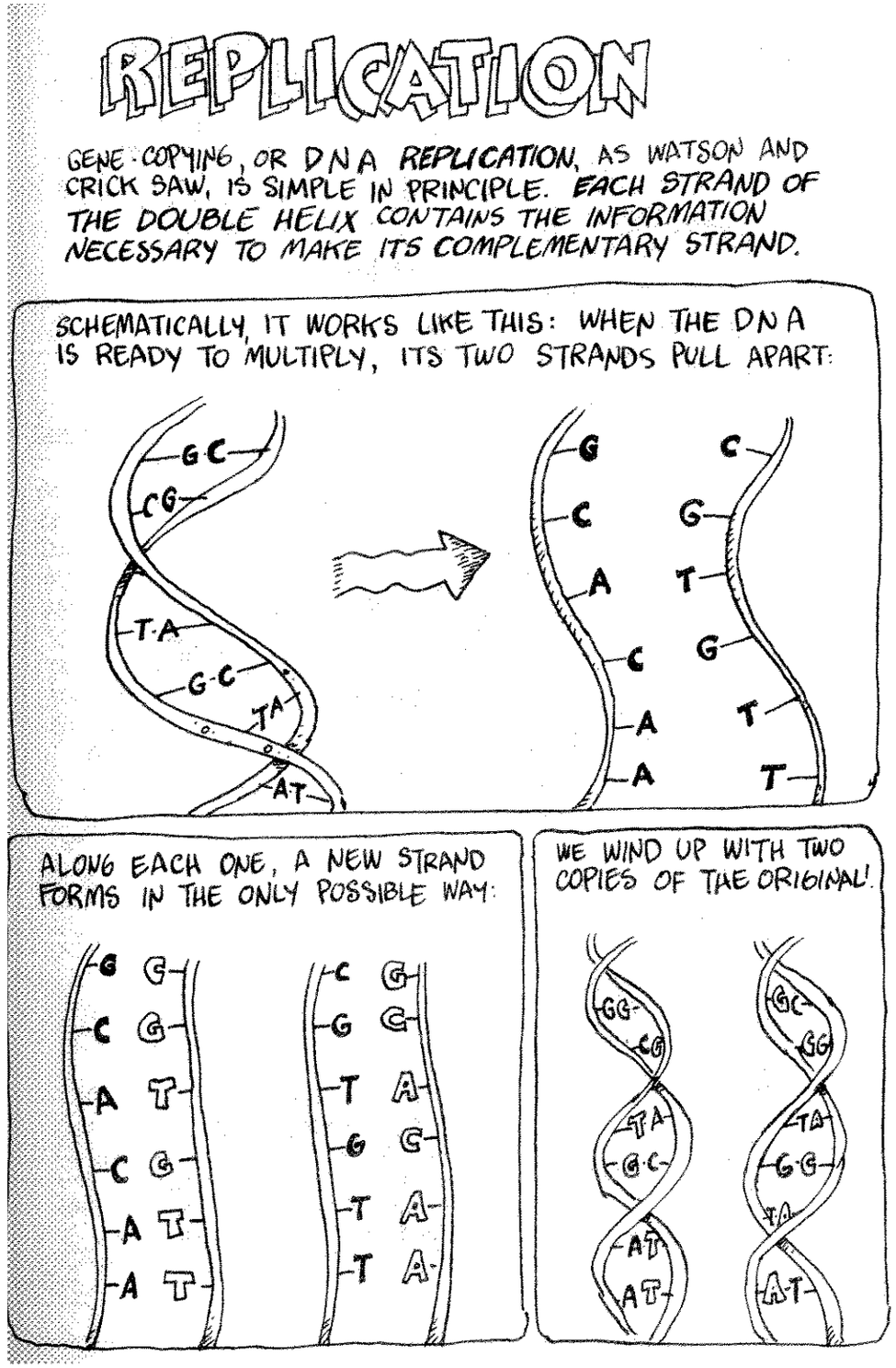

Adenine
Thymine

A T



DNA Replication



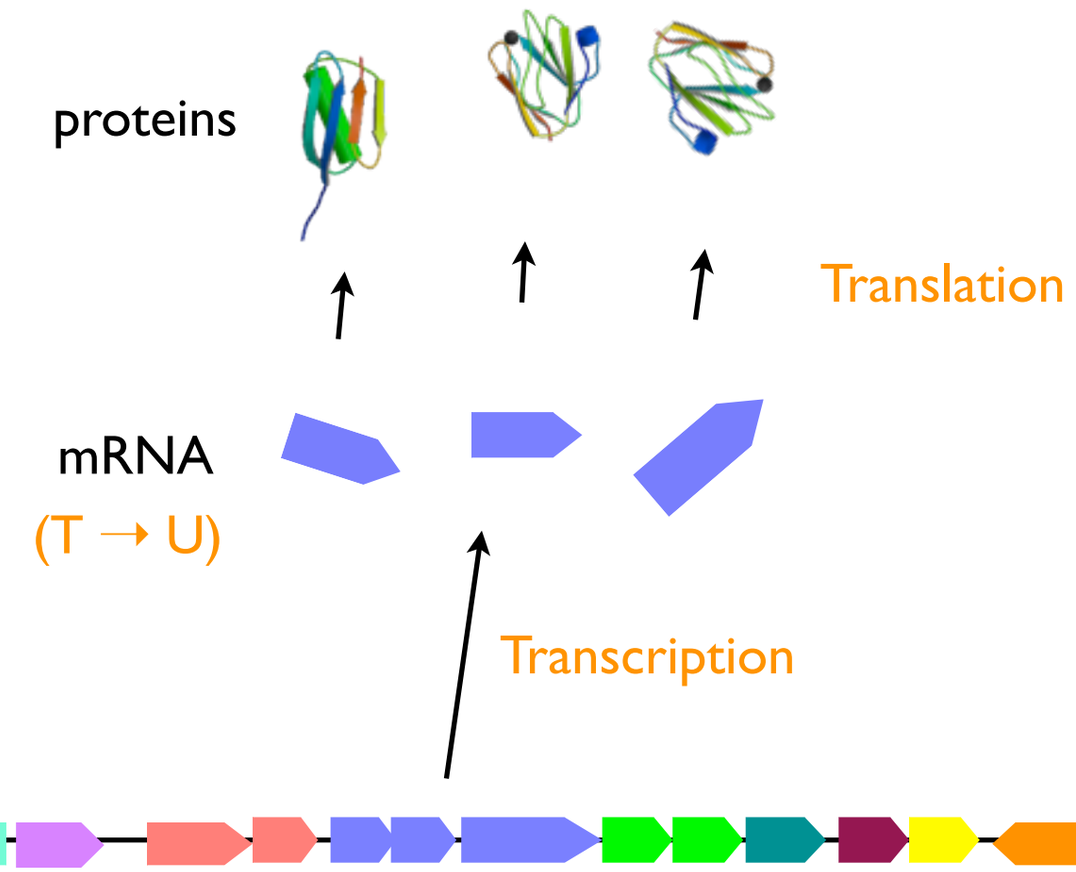


The Cartoon Guide to Genetics
Larry Gonick & Mark Wheelis, 1983

Most Genes Encode For Proteins

- Make up structural components of the cell.
- Pass signals from environment to the cell and between locations within the cell.
- Act as *enzymes:* catalyze reactions.
- Work as molecular motors
- Many other functions...

"Central Dogma" of Biology



DNA =

- double-stranded, linear molecule
- each strand is string over {A,C,G,T}
- strands are complements of each other (A ↔ T; C ↔ G)
- substrings encode for genes most of which encode for proteins

Proteins

mRNA
$\Sigma = \{A,C,G,U\}$

↓

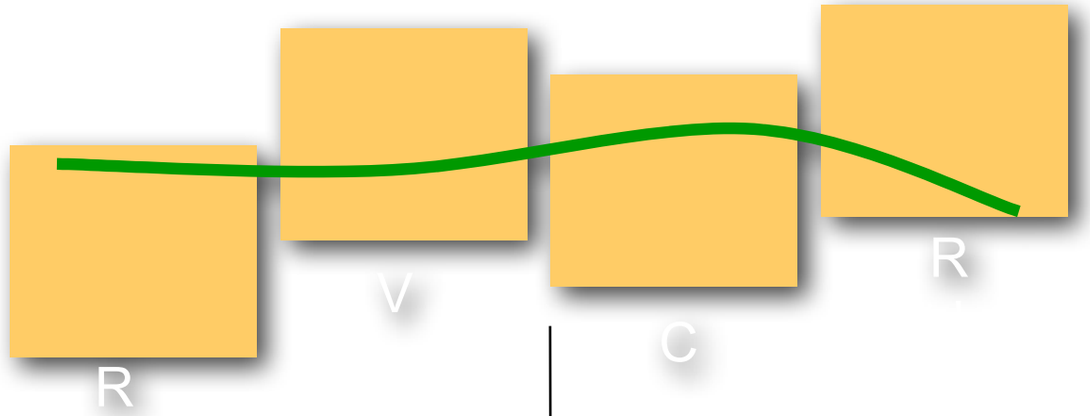
protein
$|\Sigma| = 20$ amino acids

AGG → R

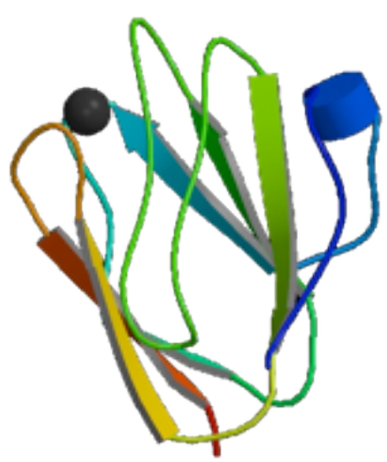
GUC → V

UGU → C

CGA → R



Amino acids with flexible side chains strung together on a backbone



Proteins are the Building Blocks of Life

Their shape is instrumental in determining their function.

1st base	2nd base: U		2nd base: C		2nd base: A		2nd base: G	
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine
	UUC	(Phe/F) Phenylalanine	UCC	(Ser/S) Serine	UAC	(Tyr/Y) Tyrosine	UGC	(Cys/C) Cysteine
	UUA	(Leu/L) Leucine	UCA	(Ser/S) Serine	UAA	Ochre *Stop*	UGA	Opal *Stop*
	UUG	(Leu/L) Leucine	UCG	(Ser/S) Serine	UAG	Amber *Stop*	UGG	(Trp/W) Tryptophan
C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine
	CUC	(Leu/L) Leucine	CCC	(Pro/P) Proline	CAC	(His/H) Histidine	CGC	(Arg/R) Arginine
	CUA	(Leu/L) Leucine	CCA	(Pro/P) Proline	CAA	(Gln/Q) Glutamine	CGA	(Arg/R) Arginine
	CUG	(Leu/L) Leucine	CCG	(Pro/P) Proline	CAG	(Gln/Q) Glutamine	CGG	(Arg/R) Arginine
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine
	AUC	(Ile/I) Isoleucine	ACC	(Thr/T) Threonine	AAC	(Asn/N) Asparagine	AGC	(Ser/S) Serine
	AUA	(Ile/I) Isoleucine	ACA	(Thr/T) Threonine	AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine
	AUG [A]	(Met/M) Methionine	ACG	(Thr/T) Threonine	AAG	(Lys/K) Lysine	AGG	(Arg/R) Arginine
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine
	GUC	(Val/V) Valine	GCC	(Ala/A) Alanine	GAC	(Asp/D) Aspartic acid	GGC	(Gly/G) Glycine
	GUA	(Val/V) Valine	GCA	(Ala/A) Alanine	GAA	(Glu/E) Glutamic acid	GGA	(Gly/G) Glycine
	GUG	(Val/V) Valine	GCG	(Ala/A) Alanine	GAG	(Glu/E) Glutamic acid	GGG	(Gly/G) Glycine

Another View of the Genetic Code

- There are 20 different amino acids & 64 different codons.
- Lots of different ways to encode for each amino acid.
- The 3rd base is typically less important for determining the amino acid
- Three different "stop" codons that signal the end of the gene
- Start codons differ depending on the organisms, but AUG is often used.

Databases of Biological Data

- General Repositories
  - GenBank - USA
  - EMBL - Europe: http://www.ebi.ac.uk/embl/

- Specialized by data type
  - NCBI Trace Archive - raw reads from sequencing machines
  - SwissProt - curated protein information: http://www.expasy.org
  - KEGG - metabolic pathways: http://www.genome.jp/kegg/
  - Gene Expression Omnibus (GEO) - gene expression
  - PDB - protein structures

- Specialized by organism
  - ZFIN - zebrafish
  - SGD - yeast
  - WormBase - c. elegans
  - FlyBase - fruit fly

NCBI

GenBank: central repository for genome sequences

Genome Home

http://www.ncbi.nlm.nih.gov/sites/genome

Google

NCBI

Entrez Genome

My NCBI [Sign In] [Register]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search Genome for Go Clear

Limits | Preview/Index | History | Clipboard | Details

About Entrez

Entrez Genome
Help

Submiting
Genome Project
Genome sequence

Microbial
Genome Projects

Genomic BLAST
Microbial
Eukaryotic

Resources
Organelles
Viruses

Archaea
Chromosome
Plasmid
DraftAssembly

Bacteria
Chromosome
Plasmid
DraftAssembly

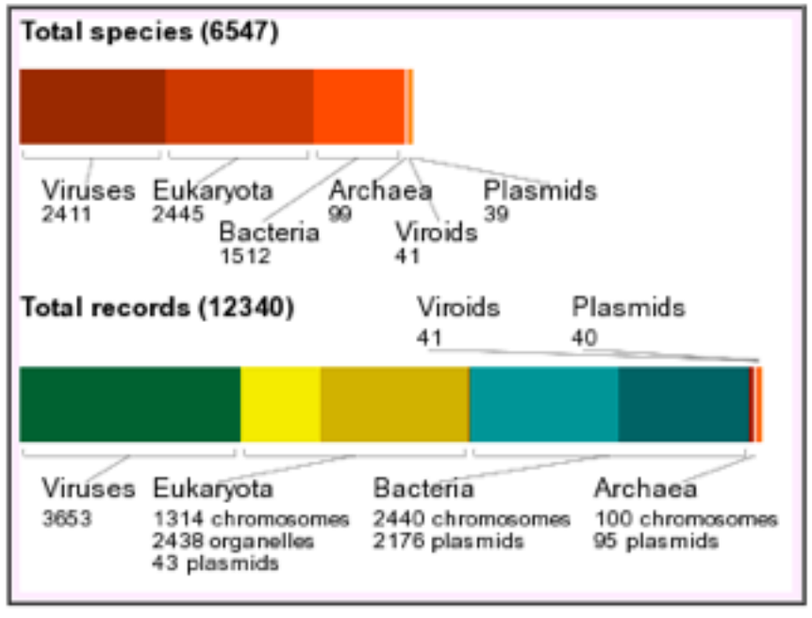
Eukaryota
Chromosome
Plasmid
Organelles

Viruses
Phages

Viroids

All Plasmids

The Genome database provides views for a variety of genomes, complete chromosomes, sequence maps with contigs, and integrated genetic and physical maps. The database is organized in six major organism groups: Archaea, Bacteria, Eukaryotae, Viruses, Viroids, and Plasmids and includes complete chromosomes, organelles and plasmids as well as draft genome assemblies.



Genome Sequencing Milestone Reached! There are now 1000 complete Prokaryotic Genomes available in Entrez Genome. See the full list of complete bacterial and archaeal genomes. Microbial Resources are available for search, retrieval, and analysis of all genomes.

Microbial Genomes Resources presents public data from prokaryotic genome sequencing projects. The sequence collection contains data from finished genomes as well as draft assemblies. The analytical tools include specialized BLAST with microbial genomes, newly developed Concise

Related resources

Entrez Genome Project complete and incomplete large-scale sequencing projects

Entrez Protein Clusters a collection of related protein sequences

Eukaryotic genome projects and sequences

Genomes of Bacillus anthracis reference genome and related sequences

Influenza Virus Resource sequence database and analyses

Microbial Genomes reference sequences and resources

Organelle reference sequences and tools

Plant Genomes Central major plant genome projects

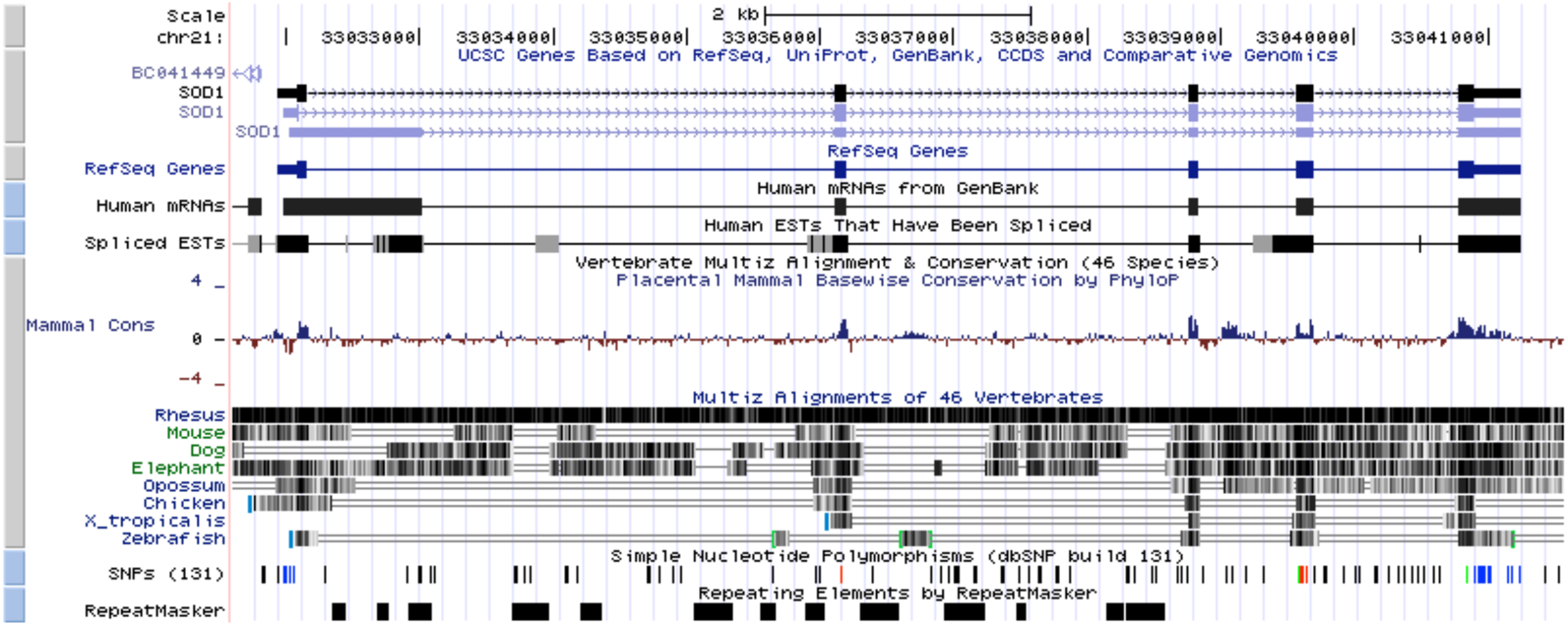
SARS Coronavirus Resource sequence data and analyses

Salmonella SNPs SNP data in two Salmonella enterica pathotypes

Viruses reference sequences and

Human Genome Browser

- http://genome.ucsc.edu/



Home Hide
News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
New Website Features

Deposition Hide
All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamline
Related Tools

Search Hide
Advanced Search
Latest Release
Latest Publications
Sequence Search
Chemical Components
Unreleased Entries
Browse Database
Histograms

Explorer:

Last Structure: 2KZ1

Results:

Query (18 hits):
- Query Results
- Query Details
- Query History
- Save Query to MyPDB

Tools Hide
File Downloads

Summary | Derived Data | Sequence | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Inter-molecular interactions in a 44 kDa interferon-receptor complex detected by asymmetric back-protonation and 2D NOESY

2KZ1

Display Files ▾
Download Files ▾
Print this Page
Share this Page

DOI:10.2210/pdb2kz1/pdb

ENTRY 2KZ1 SUPERSEDES 2KSX

Primary Citation

Intermolecular interactions in a 44 kDa interferon-receptor complex detected by asymmetric reverse-protonation and two-dimensional NOESY.

Nudelman, I., Akabayov, S.R., Schnur, E., Biron, Z., Levy, R., Xu, Y., Yang, D., Anglister, J.

Journal: (2010) Biochemistry **49:** 5117-5133

PubMed: 20496919
DOI: 10.1021/bi100041f
Search Related Articles in PubMed

PubMed Abstract:

Type I interferons (IFNs) make up a family of homologous helical cytokines initiating strong antiviral and antiproliferative activity. All type I IFNs bind to a common cell surface receptor consisting of two subunits, IFNAR1 and IFNAR2, associating upon binding of...
[Read More & Search PubMed Abstracts]

Molecular Description Hide

Classification: Antiviral Protein
Structure Weight: 43587.80

Molecule: Interferon alpha-2
Polymer: 1 **Type:** polypeptide(L) **Length:** 165
Chains: A

Molecule: Soluble IFN alpha/beta receptor

Structure Image

More Images...

View in Jmol
Other Viewers ▾
SimpleViewer
Protein Workshop
Ligand Explorer

Deposition Summary Hide

Authors: Nudelman, I., Akabayov, S.R., Schnur, E., Biron, Z., Levy, R., Xu, Y., Yang, D., Anglister, J.

Recap

- Central dogma of biology: DNA -> RNA -> Proteins
  - DNA encodes genes, most of which encode for proteins (via the genetic code)
  - Proteins perform much of the work of the cell.
  - RNA acts as an intermediate step (it also has other functions as well)
- Huge amount of data now available, need algorithms to make sense of it.
- Next up: sequence comparison using dynamic programming.