CMSC423: Bioinformatic Algorithms, Databases and Tools
Lecture 1

Instructor: Mihai Pop

MW 3:30-4:15 CSIC 3120

INTRODUCTIONS

- Instructor: Mihai Pop (mpop at umiacs.umd.edu)
  Office hours: Tuesdays 1-2pm, 3120F Bio. Sci. Bldg.
- TA: Behjat Siddiquie (behjat at cs.umd.edu)
  Office hours:
- You

- Class webpage:
  http://www.cbcb.umd.edu/confcour/CMSC423.shtml-

What is bioinformatics?

- Biology can be viewed as an information science (e.g. DNA is just a string of letters)
- Computers are essential in generating, managing, and analyzing biological data
- “Bioinformatics” or “Computational Biology” encompasses all applications of computers to the analysis of biological data

Why study bioinformatics?

- Exciting field! Help biologists figure out what life is all about.
- Work with people different from you – bio-geeks
- Many programmer/software engineer jobs in biotech industry currently filled by biologists – great need for people with CS backgrounds

Overview of course

- No knowledge of biology required
- Will cover areas of interest in CURRENT bioinformatics research
- Overall flow: data management (databases), data generation (sequencing), data analysis (extracting meaning)
- Examples based on real data (note: instructor spent 5 years in a biotech research institute)

Policies

- Attendance - follow University policy
  - you must claim excused absences in writing
  - written documentation of illness is required (from Dr. not yourselves)
  - if possible inform me prior to the class you will skip
- Disabilities
  - must inform me during the first 2 weeks of the semester if special accommodations necessary
  - request letter from Office of Disability Support Services
- General – communication is key
  - talk to me about any issues whether covered or not by University policies

Grading & workload

- Homework (10%)
- Goal: 5-10 assignments
  - exercises from textbook
  - small programming assignments
  - “discovery” exercises (find something in public databases or using public software)
- Programming projects (15% + 15%)
  - Project 1 – assigned by instructor
  - Project 2 – chosen by student
- In-class midterm (25%) & final (35%)
- Late policy: 1 day late – 10 points off; 2 days late – 20 points off; 3 days late – 0 points

Academic Honesty

http://www.studenthonorcouncil.umd.edu/code.html

- No cheating on homeworks/projects/exams
- No making up data/results
- No copying of other people's code
- You can work together on homeworks/projects but WRITE THE ANSWER BY YOURSELF

I pledge on my honor that I have not given or received any unauthorized assistance on this examination.

Advice: how to do well in the class

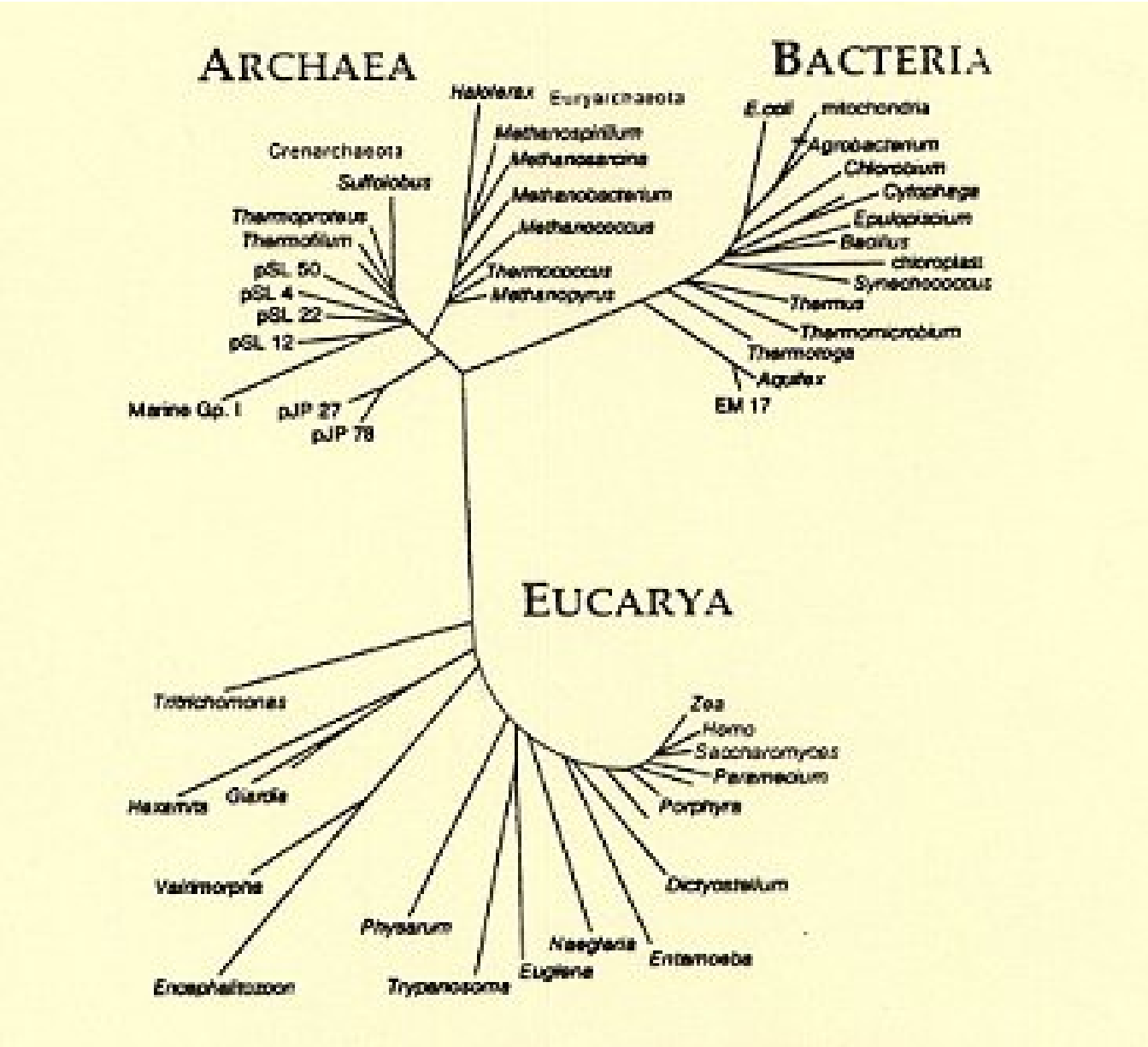
- Start early on assignments – at least read the assignment after class
- Ask questions – during class, exams, office hours, using email (I’m available most time by email)
- Be inquisitive – follow up on topics discussed in class: Google, Wikipedia
- Be social – get to know some biologists – learn what they do, what they are interested in
- Get to know your colleagues

Summer internships

- Venter Institute
  http://www.venterinstitute.org/education/internship.php

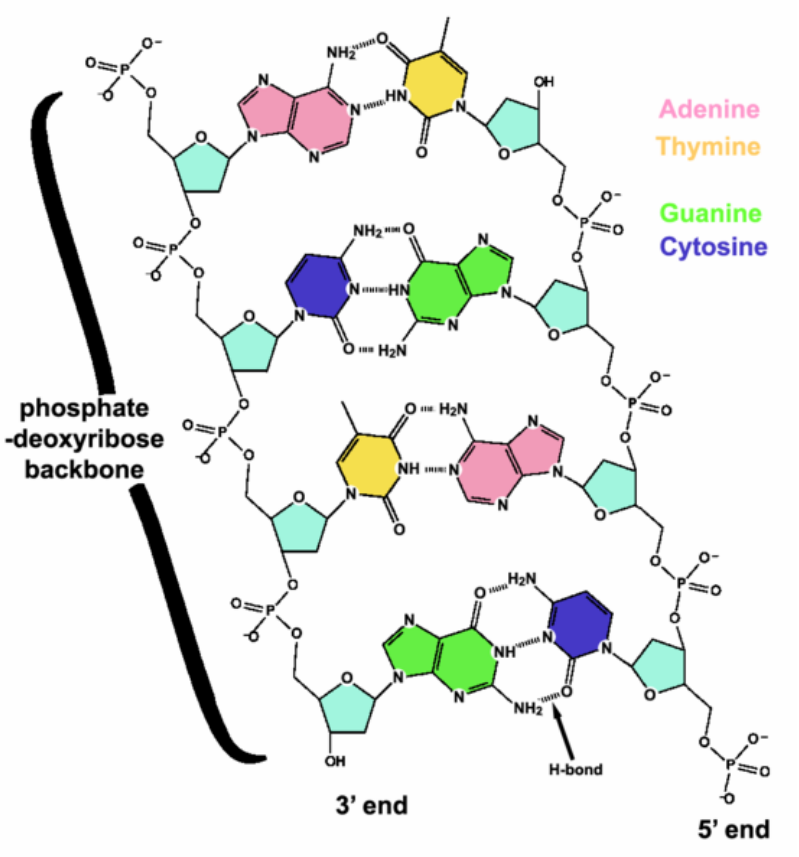
- Center for Bioinformatics and Computational Biology
  http://www.cbcb.umd.edu

The tree of life



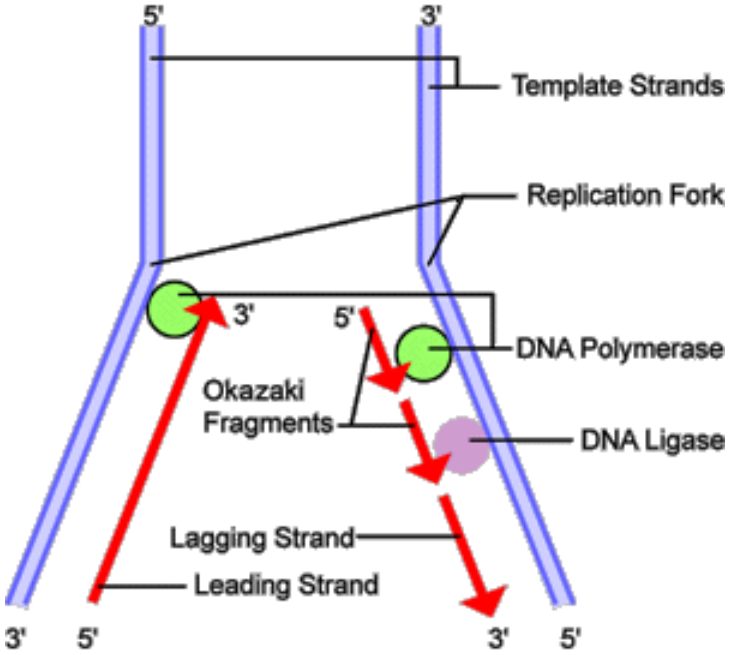
http://www.fossilmuseum.net/Tree_of_Life/Domains_Archaea_Bacteria/

DNA – the code of life



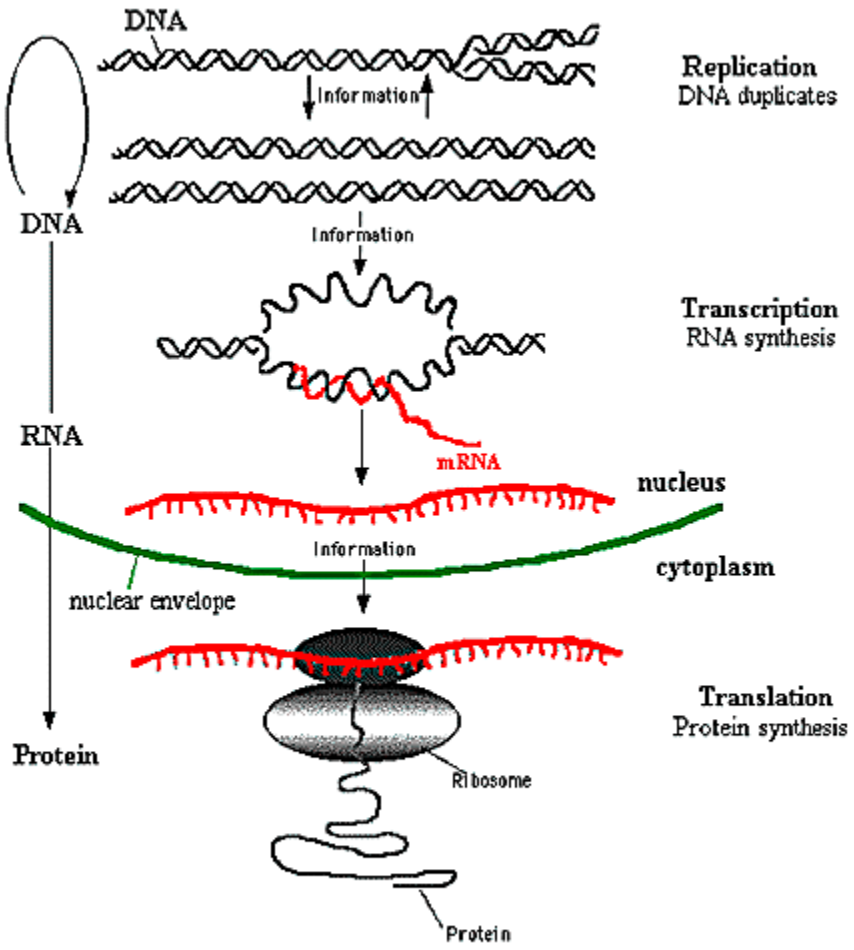
- Purines A, G, caffeine
- Pyrimidines C, T
- Sugar backbone (ticker tape)

- Double-stranded – allows replication



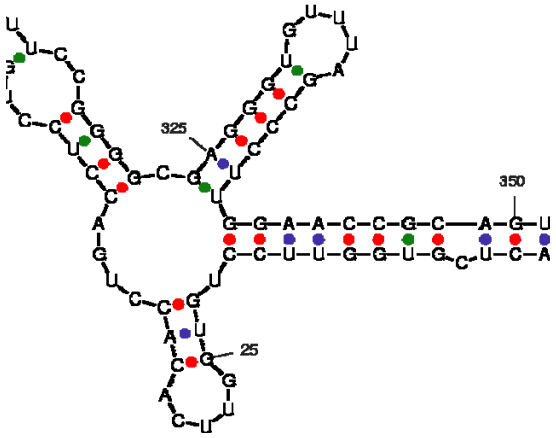
pictures from wikipedia

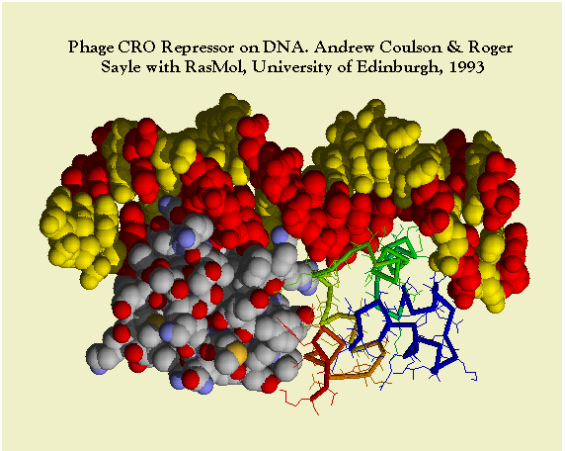
Central dogma



The Central Dogma of Molecular Biology

AGGTACGCGTACCTGACAGG





http://www.accessexcellence.org/RC/VL/GG/central.html

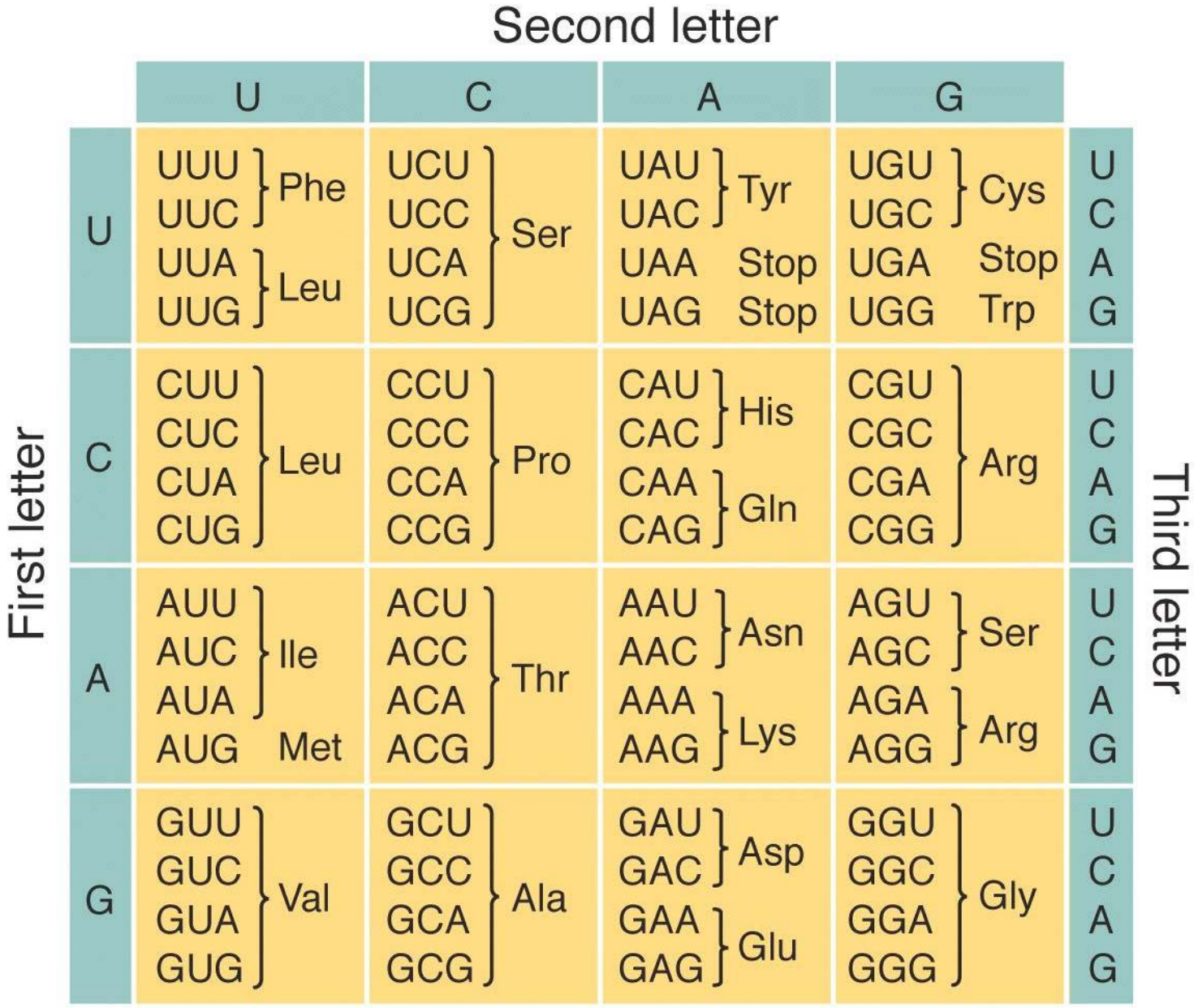
Genes, transcription, translation

- DNA – RNA - Thymine replaced by Uracil (T-U)
- The transcribed segments are called genes

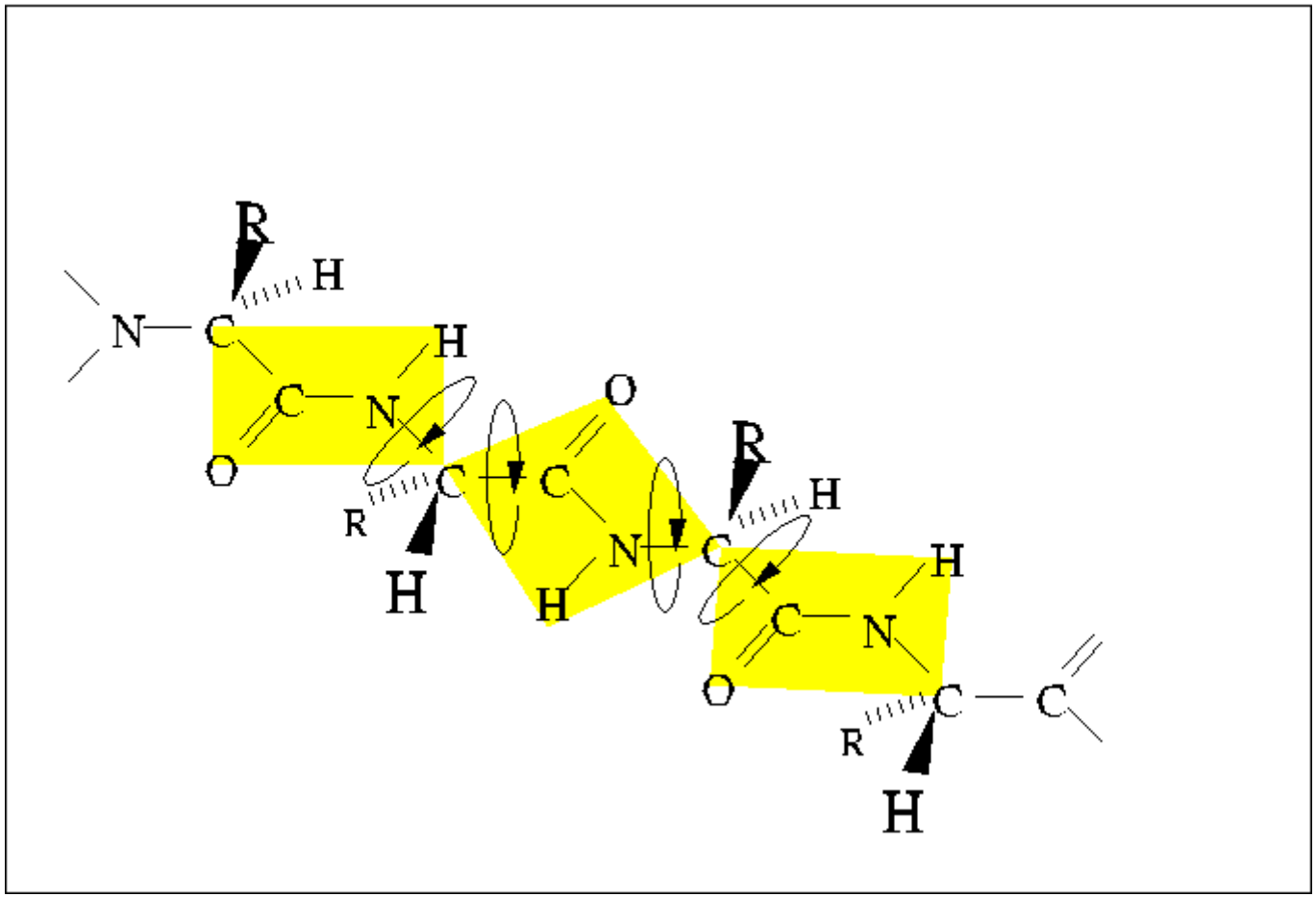
ACCGUACC**AUG**UUA...AUAGGC**UGA**GCA

- AUG – start codon (also amino-acid Methionine)
- UAA, UAG, UGA – stop codons
- Genes are read in sets of 3 nucleotides during translation – $4^3 = 64$ possible combinations
- Each combination codes for one of 20 amino-acids – the building blocks for proteins

Amino-acid translation table

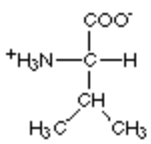


Protein structure

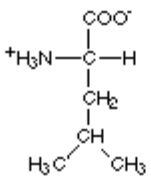


http://www.tulane.edu/~biochem/med/second.htm

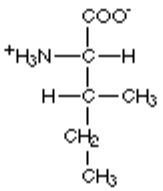
Amino acids with hydrophobic side groups



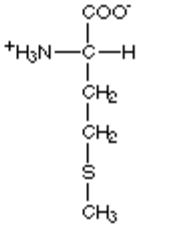
Valine
(val)



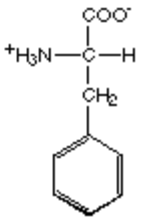
Leucine
(leu)



Isoleucine
(ile)



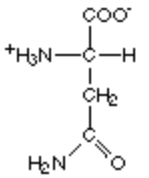
Methionine
(met)



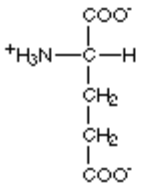
Phenylalanine
(phe)

hate water

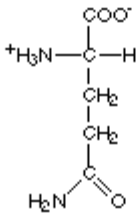
Amino acids with hydrophilic side groups



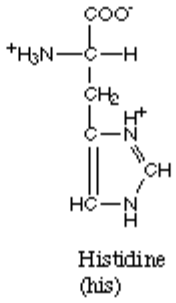
Asparagine
(asn)



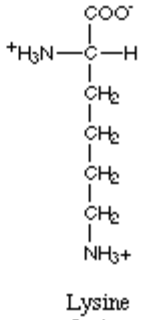
Glutamic acid
(glu)



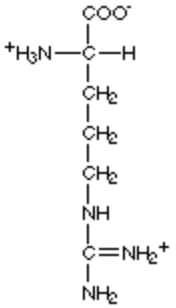
Glutamine
(gln)



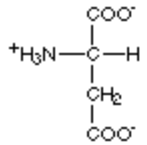
Histidine
(his)



Lysine
(lys)



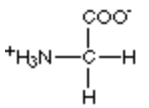
Arginine
(arg)



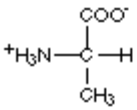
Aspartic acid
(asp)

like water

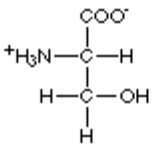
Amino acids that are in between



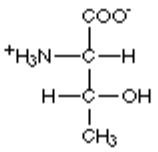
Glycine
(gly)



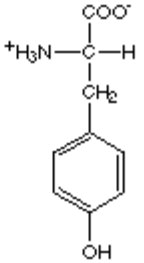
Alanine
(ala)



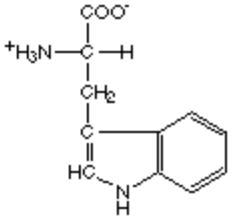
Serine
(ser)



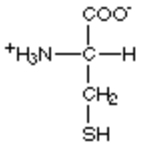
Threonine
(thr)



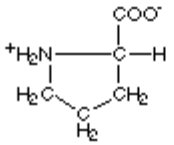
Tyrosine
(tyr)



Tryptophan
(trp)



Cysteine
(cys)

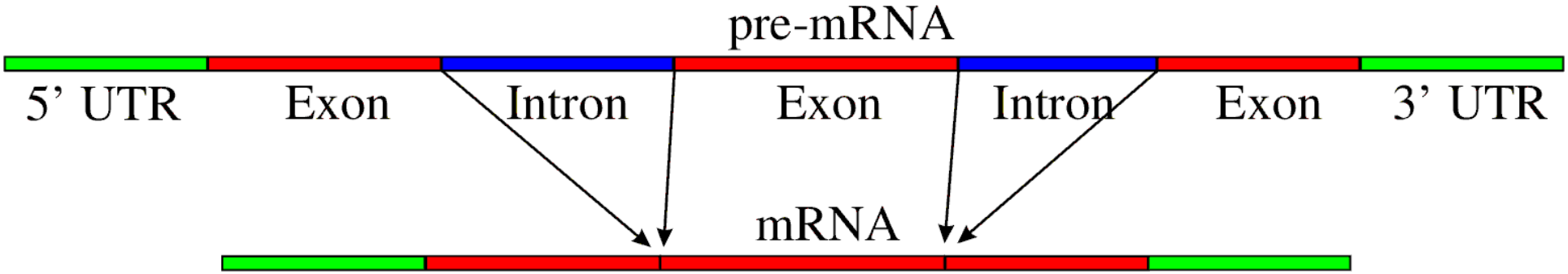


Proline
(pro)

can’t decide

http://web.mit.edu/esgbio/www/lm/proteins/aa/aminoacids.html

Translation – complications



Homework 1

- Reverse complement some sequences
- Translate some DNA to the corresponding protein sequence
- Look things up on the internet.