

Life's Interpreter

COMS W4115

Prof. Stephen A. Edwards
Spring 2002

Columbia University
Department of Computer Science

Life

We're built from complex molecules called proteins.

Each protein is built from a string of amino acids.

A little interpreter called a *ribosome* builds these.

DNA

Every known living organism uses DNA to record its construction program.

Double helix assembled from sequences of base pairs.



The Twenty Amino Acids

A	Alanine	Ala	M	Methionine	Met
C	Cysteine	Cys	N	Asparagine	Asn
D	Aspartic Acid	Asp	P	Proline	Pro
E	Glutamic Acid	Glu	Q	Glutamine	Gln
F	Phenylalanine	Phe	R	Arginine	Arg
G	Glycine	Gly	S	Serine	Ser
H	Histidine	His	T	Threonine	Thr
I	Isoleucine	Ile	V	Valine	Val
K	Lysine	Lys	W	Tryptophan	Trp
L	Leucine	Leu	Y	Tyrosine	Tyr

The Genetic Code

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC " TAA STOP TAG STOP	TGT Cys (C) TGC " TGA STOP TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " ATG Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

Building Proteins

ATGGAATTCTCGCTCTAA	DNA
TACCTTAAGAGCGAGATT	DNA'
AUGGAAUUCUCGCUCUAA	RNA
Met-Glu-Phe-Ser-Leu	Protein

Other Genetic Information

Terribly simplified model.

Much more information in DNA:

Start sequences ("start copying RNA here")

Sequences for rate regulation ("Make this protein faster")

More than 100k proteins in humans.

Human DNA: 23 chromosomes comprising 3 billion base pairs.

Human genome project: enumerate these 3 billion pairs (done).