

## BINF630/BIOL580 Bioinformatics Methods

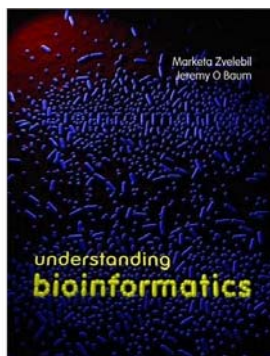
Iosif Vaisman

Department of Bioinformatics and Computational Biology

Email: [ivaisman@gmu.edu](mailto:ivaisman@gmu.edu)

Fall 2009

## Recommended book



Marketa J Zvelebil,  
Jeremy O Baum

**UNDERSTANDING  
BIOINFORMATICS**

New York: Garland Science, 2008.

## Bioinformatics

Bioinformatics is a field that deals with biological information, data, and knowledge, and their storage, retrieval, management, and optimal use for problem solving and decision making.

## NIH working definition of bioinformatics and computational biology (July 2000)

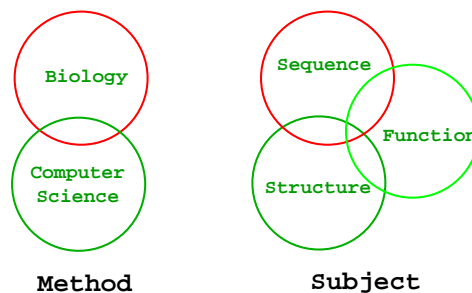
The NIH Biomedical Information Science and Technology Initiative Consortium agreed on the following definitions of bioinformatics and computational biology recognizing that no definition could completely eliminate overlap with other activities or preclude variations in interpretation by different individuals and organizations.

**Bioinformatics:** Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

**Computational Biology:** The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.

COMPUTATIONAL BIOLOGY  
COMPUTATIONAL STRUCTURAL BIOLOGY  
COMPUTATIONAL MOLECULAR BIOLOGY  
BIOINFORMATICS  
GENOMICS  
STRUCTURAL GENOMICS  
PROTEOMICS  
...  
...

## Bioinformatics



## Informatics

in•for•mat•ics (in'fər mat'iks) *n.* (used with a sing. v.) the study of information processing; computer science. [trans. of Russ informátika (1966); see INFORMATION, -ICS]

Random House Unabridged Dictionary

## Information

General

knowledge or intelligence communicated, received or gained

Information theory

indication of the number of possible choices

Th\_ qui\_k br\_wn \_ox ju\_ps ov\_\_ th\_ laz\_ d\_g  
Ae\_h uz\_ ko\_ wm so\_g oqr\_it ypu\_vn tr\_e oj\_

## Information

Th\_ qui\_k br\_wn \_ox ju\_ps ov\_\_ th\_ laz\_ d\_g  
Ae\_h uz\_ ko\_ wm so\_g oqr\_it ypu\_vn tr\_e oj\_

The quick brown fox jumps over the lazy dog  
Aedh uzh kox wm sobg oqrfit ypulvn tree ojc

## Information and uncertainty

Information is a decrease in uncertainty

$$\log_2(M) = -\log_2(M^{-1}) = -\log_2(P)$$

Shannon's formula for uncertainty

$$H = -\sum_{i=1}^M P_i \log_2 P_i$$

only infrmatn esentil to understandn mst b tranmitd

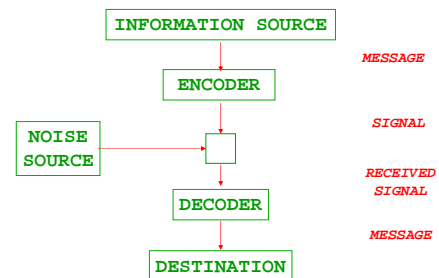
## Communication

Fundamental problem of communication:

reproducing at one point either exactly or approximately a message selected at another point

*The Mathematical Theory of Communication*  
Claude Shannon and Warren Weaver

## Communication system



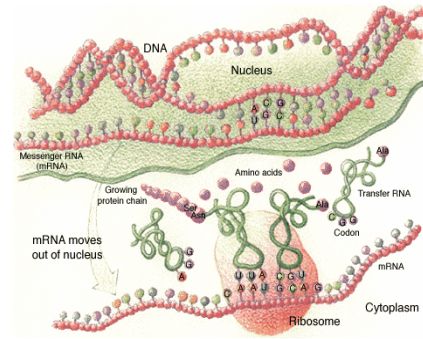
Adopted from C.E. Shannon,  
*The Mathematical Theory of Communication*, 1949

## Communication system duality

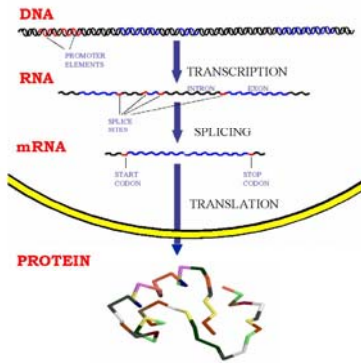
“This duality can be pursued further and is related to the duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past but cannot control it; we may control the future but have no knowledge of it.”

C. E. Shannon (1959)

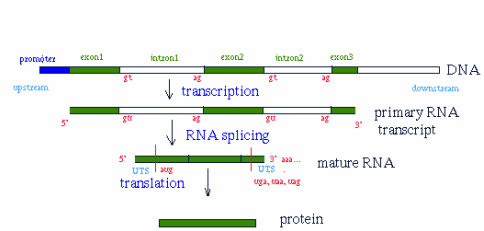
## Cell Informatics



## Cell Informatics



## Cell Informatics



## Error correcting codes

	a	b	c	d	e
a					
b					
c					
d					
e					

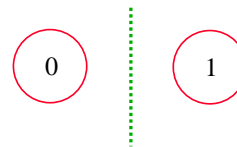
Code words ac, ba, be, db, ed in the permutation space of [a..e]x[a..e]

### Hamming metric

The sum of bit changes necessary to move from one point in the permutation space to another point in the permutation space

0000 and 0111 are separated by Hamming distance of 3:  
0000 - 0001 - 0011 - 0111

## Information Theory



1 bit

