

BINF 731

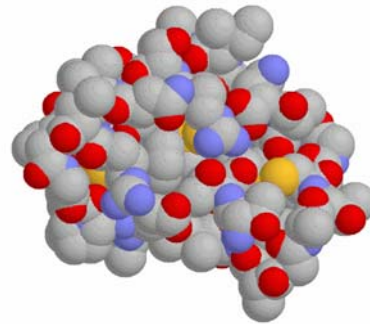
# Protein Structure Analysis

<http://binf.gmu.edu/vaisman/binf731/>

Iosif Vaisman

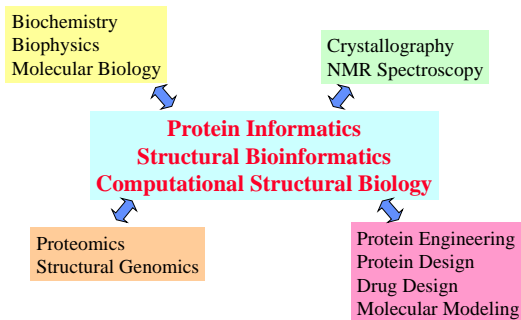
2009

## Proteins



TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIPGATCPGDYAN

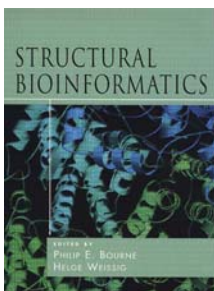
## Protein Science



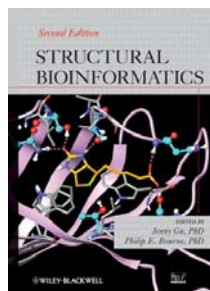
## Protein Structure and...

Business  
 Law  
 Ethics  
 Medicine  
 ...

## Recommended book



Philip Bourne, Helge Weissig (Eds)  
**Structural bioinformatics**  
Hoboken, N.J. : Wiley-Liss, 2003.



Jenny Gu, Philip Bourne (Eds)  
**Structural bioinformatics**  
Hoboken, N.J. : Wiley-Liss, 2009.

## 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

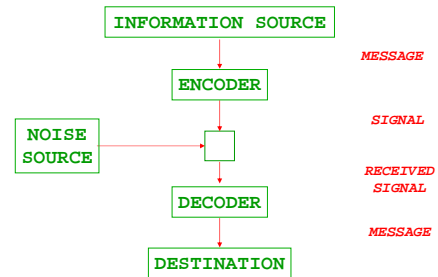
## Communication

Fundamental problem of communication:

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

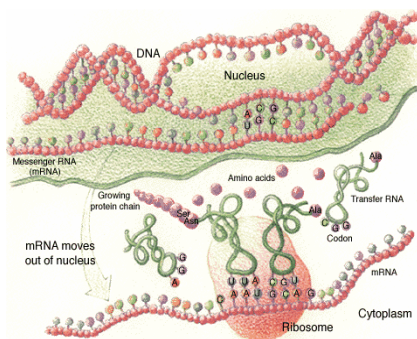
*The Mathematical Theory of Communication*  
 Claude Shannon and Warren Weaver

## Communication system

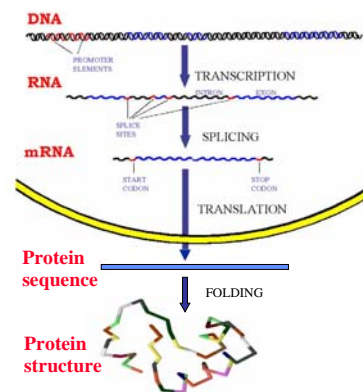


Adopted from C.E. Shannon,  
*The Mathematical Theory of  
 Communication*, 1949

## Cell Informatics



## Cell Informatics

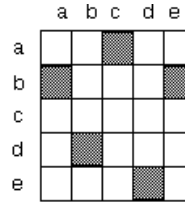


## Communication system duality

“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)

## Error correcting codes



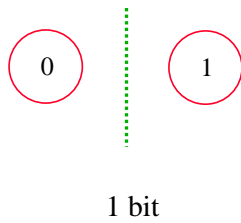
Code words ac, ba, be, db, ed in the permutation space of [a..e]x[a..e]

### Hamming metric

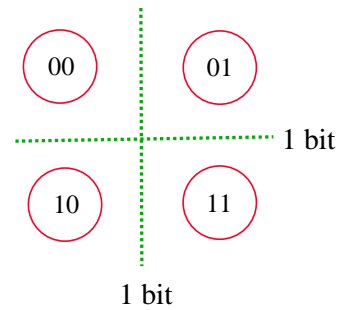
The sum of bit changes necessary to move from one point in the permutation space to another point in the permutation space

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

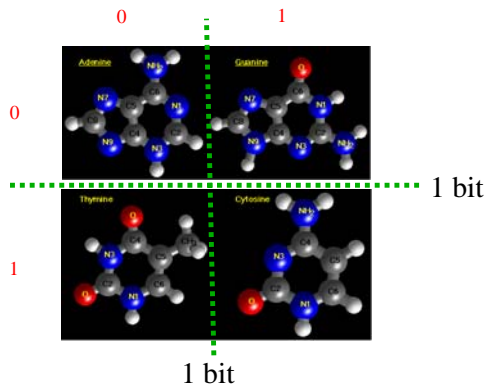
## Information Theory



## Information Theory



## Nucleotide permutation space



## Standard genetic code

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

