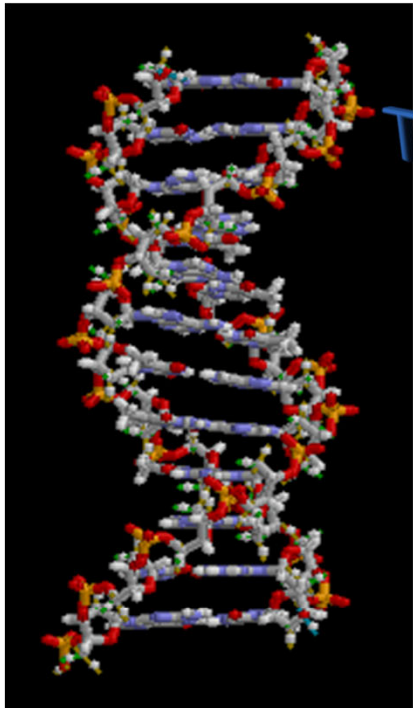
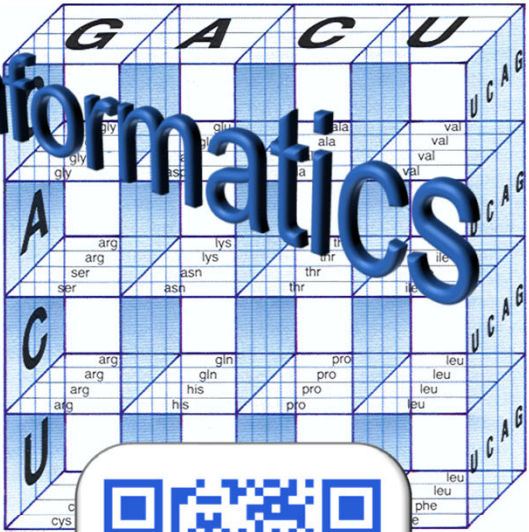


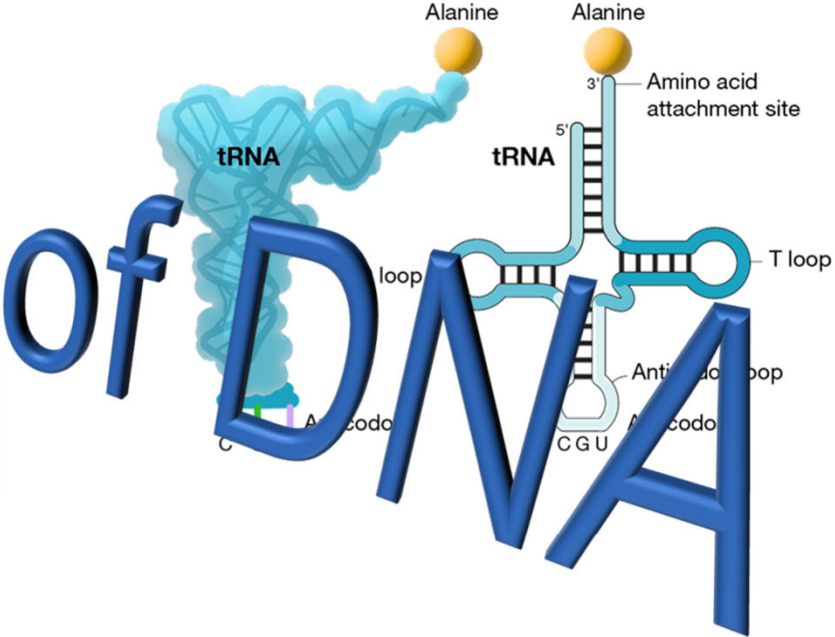
lecture 7: the informatics of DNA



The informatics of DNA



Common ways of depicting transfer RNA (tRNA)



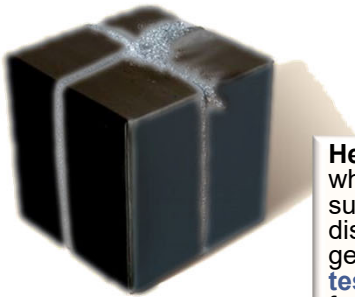
evaluation

- **Participation: 20%.**
 - class discussion, everybody reads and discusses every paper
 - engagement in class
- **Paper Presentation and Discussion: 20%**
 - **SSIE501** students are assigned to papers individually or as group lead presenters and discussants
 - all students are supposed to read and participate in discussion of every paper.
 - **Presenter prepares short summary of assigned paper (15 minutes)**
 - no formal presentations or PowerPoint unless figures are indispensable.
 - **Summary should:**
 - 1) Identify the key goals of the paper (not go in detail over every section)
 - 2) What discussant liked and did not like
 - 3) What authors achieved and did not
 - 4) Any other relevant connections to other class readings and beyond.
 - **ISE440** students chose one of the presented papers to participate as lead discussant
 - not to present the paper, but to comment on points 2-3) above
 - **Class discussion is opened to all**
 - lead discussant ensures we important paper contributions and failures are addressed
- **Black Box: 60%**
 - **Group Project (2 parts)**
 - Assignment I (25%) and Assignment II (35%)



bit.ly/atBIC

The Black Box: Due: **October 9th, 2024**



Herbert Simon: Law discovery means only finding **pattern** in the data; whether the pattern will continue to hold for new data that are observed subsequently will be decided in the course of **testing the law**, not discovering it. The **discovery process** runs from particular facts to general laws that are somehow induced from them; the **process of testing** discovers runs from the laws to predictions of particular facts from them [...] To explain why the patterns we extract from observations frequently lead to correct predictions (when they do) requires us to face again the problem of **induction**, and perhaps to make some hypothesis about the uniformity of nature. But that hypothesis is neither required for, nor relevant to, the theory of discovery processes. [...] By separating the question of pattern detection from the question of prediction, we can construct a **true normative theory of discovery**-a logic of discovery.

What is it???

500

Next n Step Revert Reset

Current step: 501

next readings (check brightspace)

■ Paper Presentation: 20%

- Present (501) and lead (501&440) the discussion of an article related to the class materials
- *section 01* presents in class, *section 20* (Enginet) posts videos on Brightspace (exceptions possible)

■ Module 2: Systems Science

- Discussion Set 3 (Group 3): September 19th or 24th
 - Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapters 1 and 2.
 - Optional:
 - Rosen, R. [1986]. "Some comments on systems and system theory". *Int. J. of General Systems*, **13**: 1-3. Available in: Klir, G.J. [2001]. *Facets of systems Science*. Springer. pp: 241-243.
 - Wigner, E.P. [1960], "The unreasonable effectiveness of mathematics in the natural sciences". Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Comm. Pure Appl. Math*, **13**: 1-14.
 - Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 3.

■ Future Modules

- See brightspace

more upcoming readings (check brightspace)

■ Paper Presentation: 20%

- Present (501) and lead (501&440) the discussion of an article related to the class materials
 - Enginet students post/send video or join by Zoom synchronously

■ Module 2: Systems Science

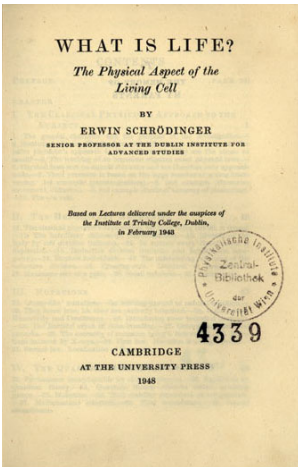
● Discussion Set 4 (Group 4): October 8th

- Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 8.
 - Optional: Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 11
- Schuster, P. (2016). The end of Moore's law: Living without an exponential increase in the efficiency of computational facilities. *Complexity*. **21**(S1): 6-9. DOI 10.1002/cplx.21824.
- Von Foerster, H., P. M. Mora and L. W. Amiot [1960]. "Doomsday: Friday, November 13, AD 2026." *Science* **132**(3436):1291-5.

■ Future Modules

- See brightspace

Erwin Schrödinger(1943-1944)



- puzzled by the persistence of living structures
 - Call to understand how life stores and perpetuates order
 - “[...] **chromosomes**[...] contain in some kind of **code-script** the entire pattern of the individual’s future development.”
 - “complete (double) copy of the code-script.”
- **aperiodic crystals as structures that can replicate themselves**
 - “We believe a gene—or perhaps the whole chromosome **fiber**—to be an aperiodic solid.”
 - “structure without predictable repetition”
 - DNA is entirely regular
 - Instead of “aperiodicity” we have **encoded information**: separated **description/construction**

“Turing invented the stored-program computer, and von Neumann showed that the description is separate from the universal constructor. This is not trivial. Physicist Erwin Schrödinger confused the program and the constructor in his 1944 book *What is Life?*, in which he saw chromosomes as “*architect’s plan and builder’s craft in one*”. This is wrong. The code script contains only a **description** of the executive function, not the **function** itself.” (Sydney Brenner)

Brenner, Sydney. [2012]. “Life’s code script.” *Nature* **482** (7386): 461-461.



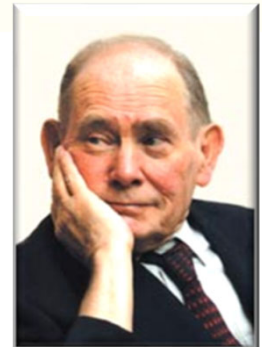
Schrodinger vs. Von Neumann

self-replication vs. decoupled, encoded information



Brenner, Sydney. [2012]. "Life's code script." *Nature* **482** (7386): 461-461.

"Turing invented the stored-program computer, and von Neumann showed that the description is separate from the universal constructor. This is not trivial. Physicist Erwin Schrödinger confused the program and the constructor in his 1944 book *What is Life?*, in which he saw chromosomes as "*architect's plan and builder's craft in one*". This is wrong. The code script contains only a **description** of the executive function, not the **function** itself." (Sydney Brenner)

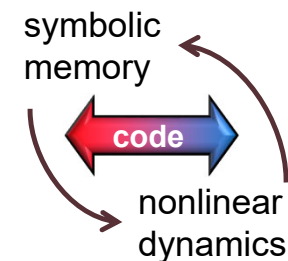


Von Neumann, J. [1949]. "Theory and organization of complicated automata."
5 lectures at University of Illinois

two roles of information

data/program (Turing)
passive/active (Von Neumann)
description/construction-function (Pattee)
genotype/phenotype (Biology)

semiotic closure (semiotic coupling)



Howard Pattee

fundamental principle of *organized complexity*
Leads to **open-ended evolution**
General principle that includes *Natural Selection*
Von Neumann described this scheme **before**
structure of DNA molecule was identified in
1953 by Watson & Crick

Rocha, L.M. & W. Hordijk [2005] *Artificial Life* **11**:189 - 214.

Rocha, L.M. [2001] *Biosystems* **60**: 95-121.

Rocha, L.M. [1996] *Systems Research* **13**: 371-384.

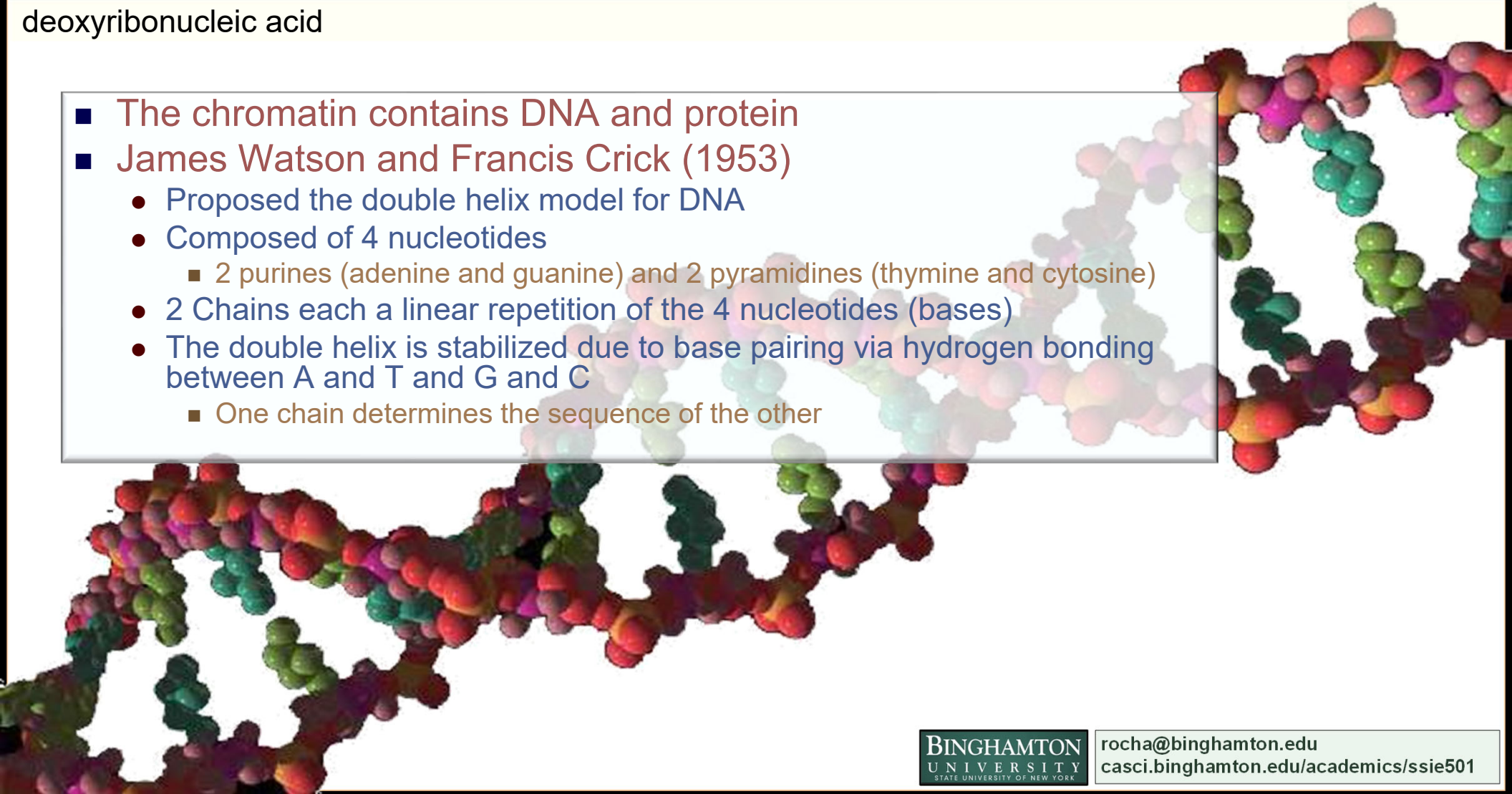
Pattee, HH [2001] *Biosystems* **60** (1):5-21



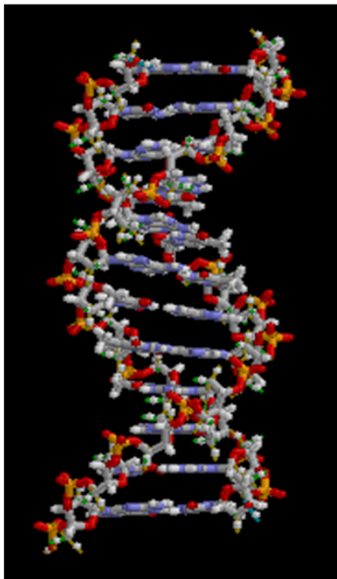
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

deoxyribonucleic acid

- The chromatin contains DNA and protein
- James Watson and Francis Crick (1953)
 - Proposed the double helix model for DNA
 - Composed of 4 nucleotides
 - 2 purines (adenine and guanine) and 2 pyrimidines (thymine and cytosine)
 - 2 Chains each a linear repetition of the 4 nucleotides (bases)
 - The double helix is stabilized due to base pairing via hydrogen bonding between A and T and G and C
 - One chain determines the sequence of the other



a molecular language system: nucleotide “bases” (the genotype “tape”)

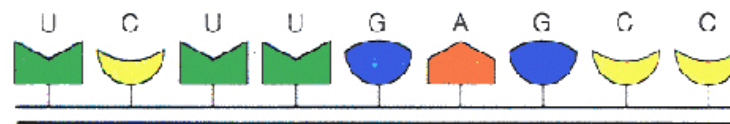


Purine (R) → Adenine (A)
 → Guanine (G)
 Nucleotides
 Pyrimidine (Y) → Cytosine (C)
 → Thymine (T)
 → Uracil (U)

4 Letter Alphabet
 DNA: A, G, C, T
 RNA: A, G, C, U

Form sequences that can store information

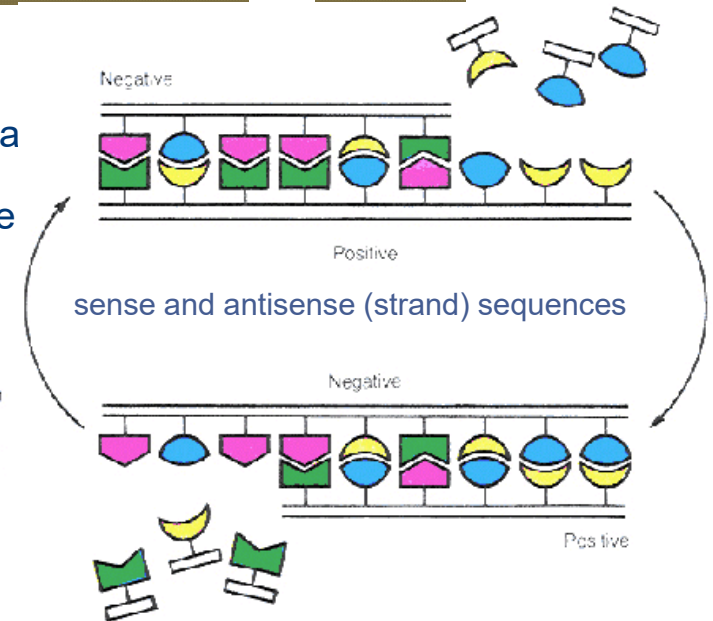
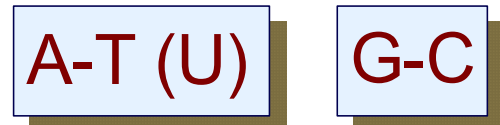
Linear molecules with a phosphate-sugar backbone (deoxyribose and ribose)



Figures from Eigen [1992]. *Steps Towards Life.*

Complementary base pairing

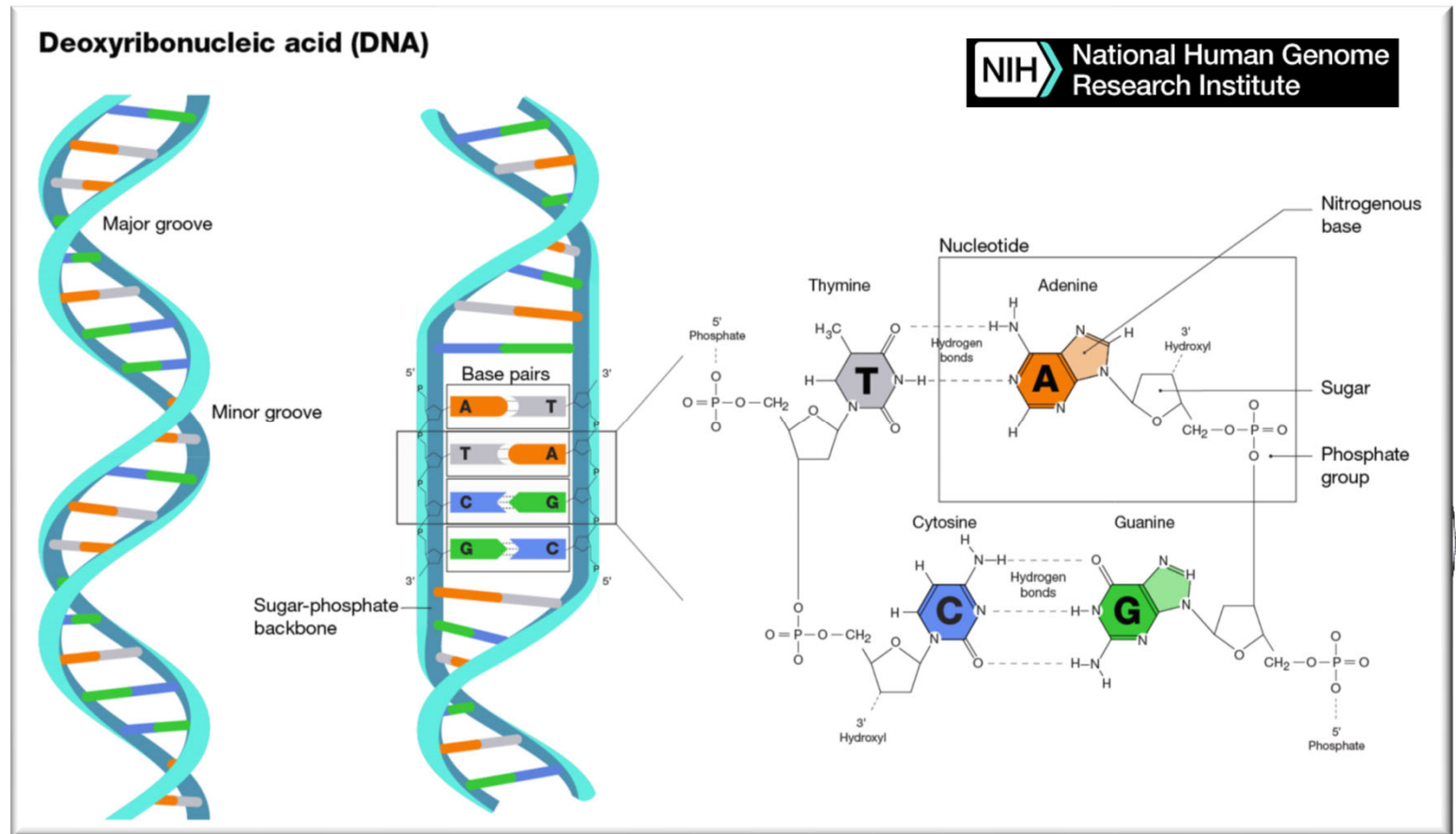
(Hydrogen-bonding between purines and pyrimidines)



Requirements for structural information

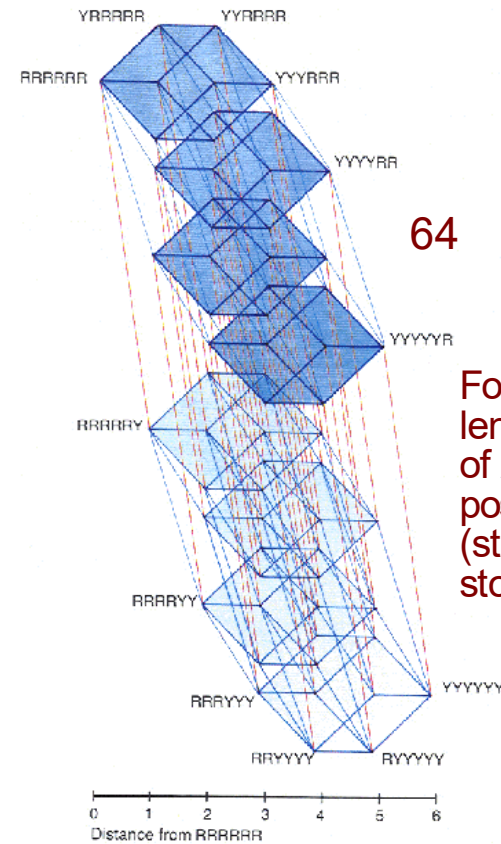
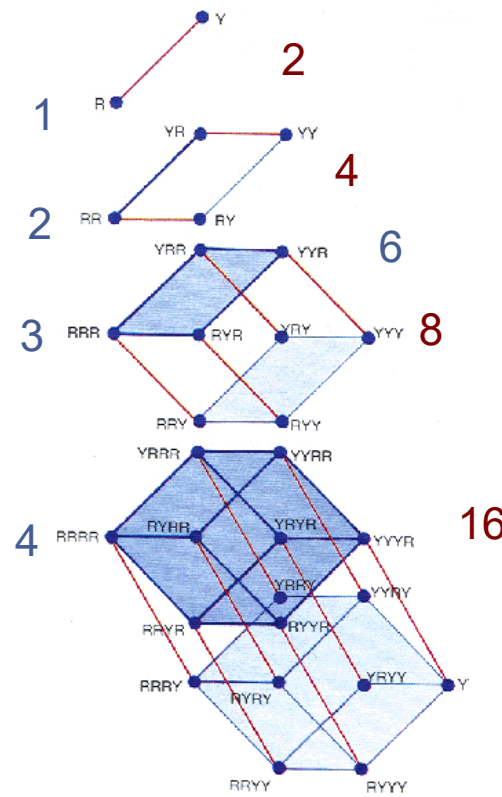
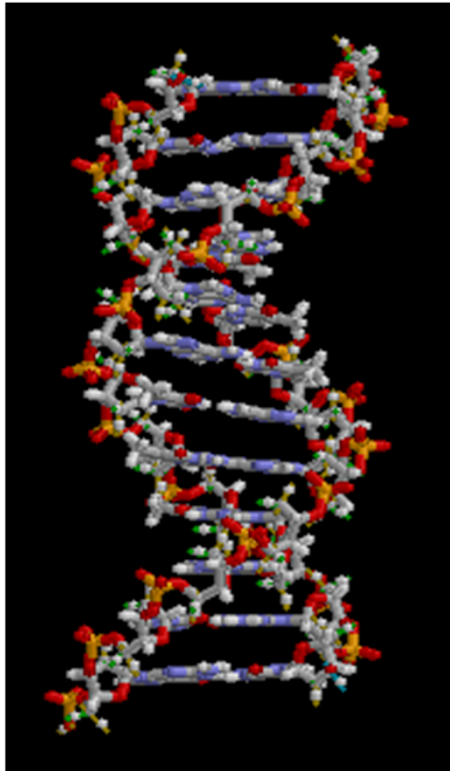
Possibility of repeated copying

a molecular language system: nucleotide “bases” (the genotype “tape”)



Possibility of repeated copying

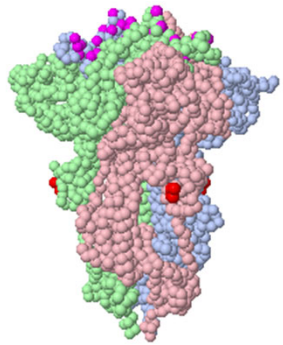
the genotype “tape” encodes an enormous amount of information



For a sequence of length n , composed of m -ary symbols, m^n possible values (structures) can be stored

Figures from Eigen [1992] . *Steps Towards Life*.

functional products that build up (self-organize) the phenotype



Jmol

Polypeptide chains of aminoacids
Primary Structure



Folding

3-dimensional structure
Secondary and tertiary bonds

- In proteins, it is the 3-dimensional structure that dictates function
 - ▶ The specificity of enzymes to recognize and react on substrates
- The functioning of the cell is mostly performed by proteins
 - ▶ Though there are also ribozymes

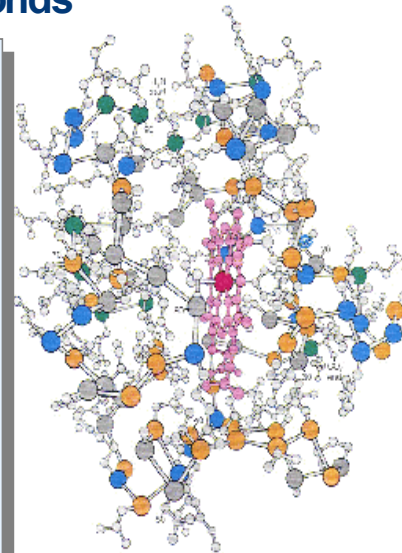
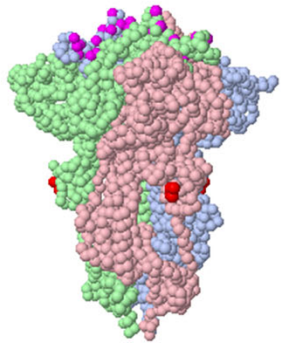


Table 1.4. Amino acid codes

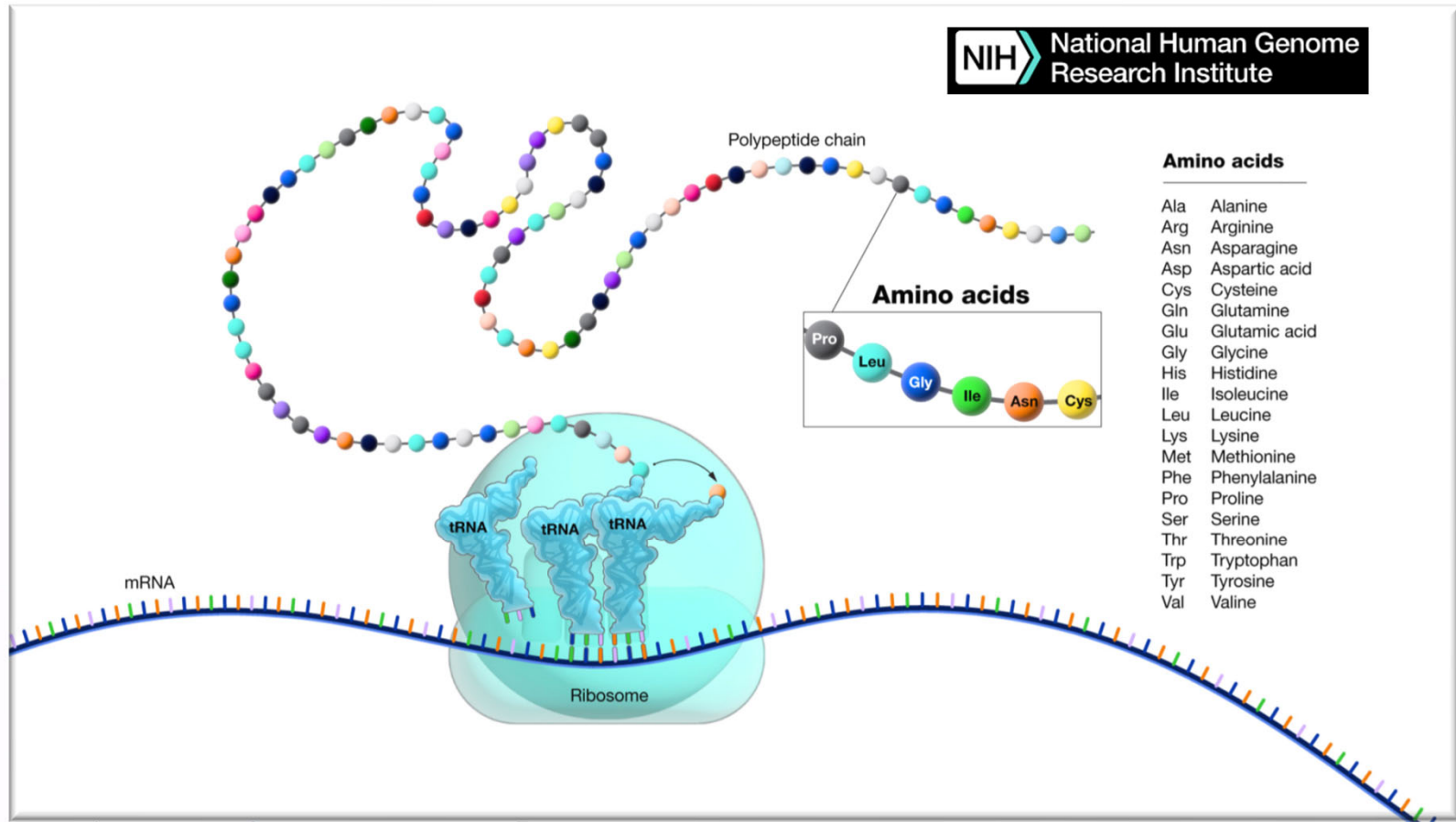
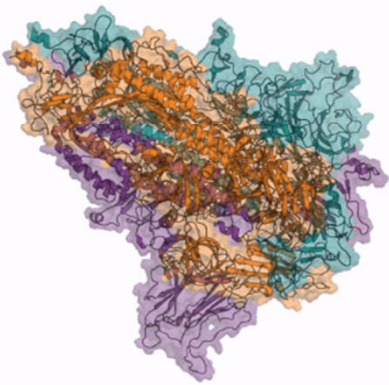
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine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

Figures from Eigen [1992] . *Steps Towards Life*.

functional products that build up (self-organize) the phenotype

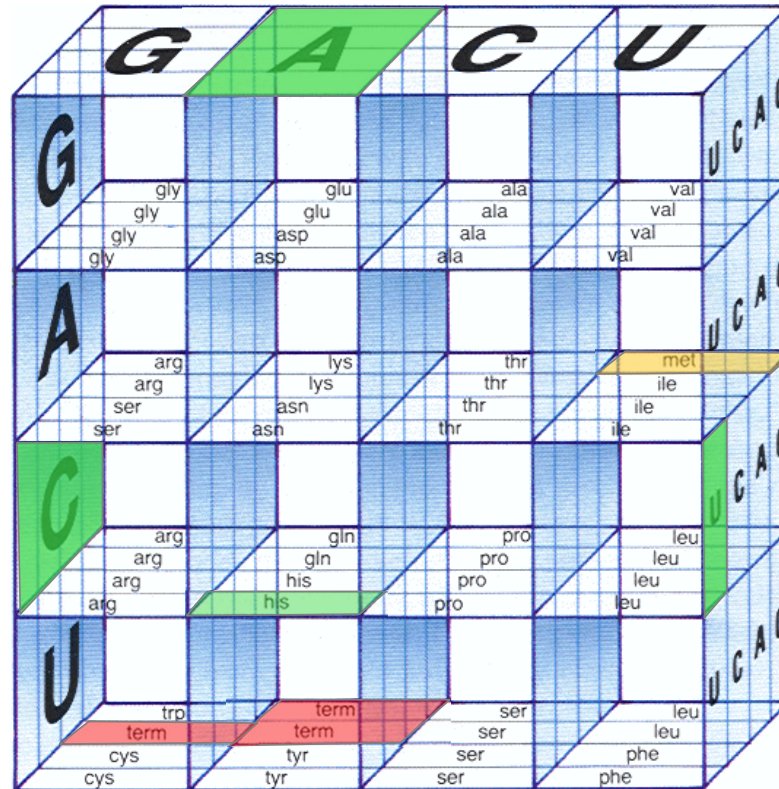


Jmol



between **genotype** and **phenotype**

Triplets of 3 Nucleotides can define 64 possible codons, but only 20 amino acids are used (redundancy)



- The genetic code maps information stored in the genome into functional proteins
 - Triplet combinations of nucleotides into amino acids

Purine (R) → Adenine (A)
 → Guanine (G)
 Nucleotides
 Pyrimidine (Y) → Cytosine (C)
 → Thymine (T)
 → Uracil (U)

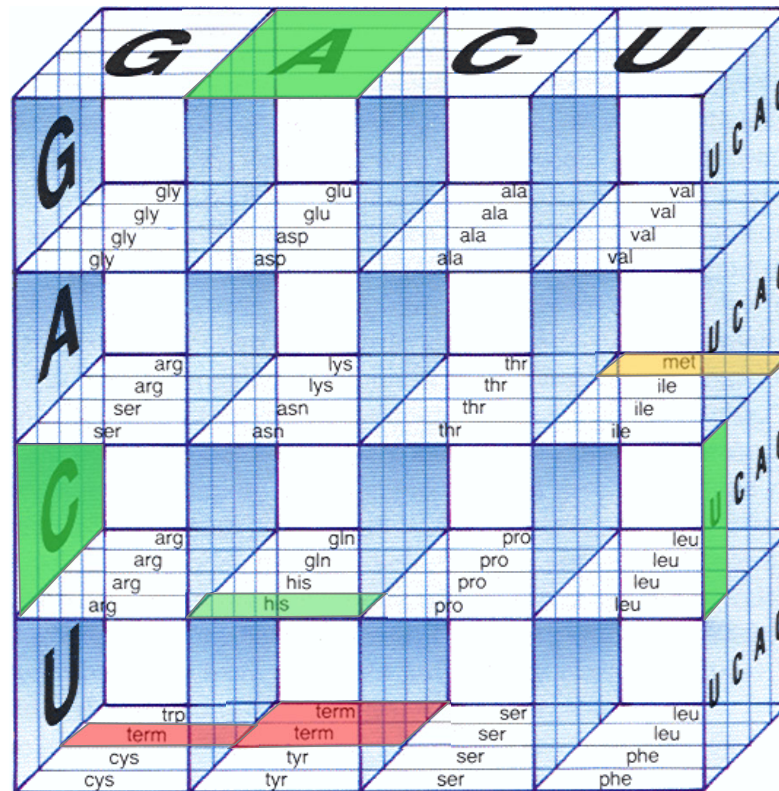
4 Letter Alphabet
DNA: A, G, C, T
RNA: A, G, C, U

Linear molecules w phosphate-sugar

Figures from Eigen [1992] . *Steps Towards Life*.

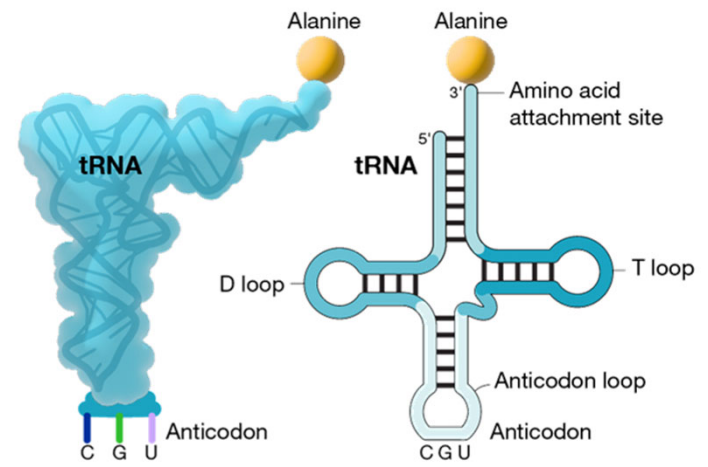
between **genotype** and **phenotype**

Triplets of 3 Nucleotides can define 64 possible codons, but only 20 amino acids are used (redundancy)



- The genetic code maps information stored in the genome into functional proteins
 - Triplet combinations of nucleotides into amino acids

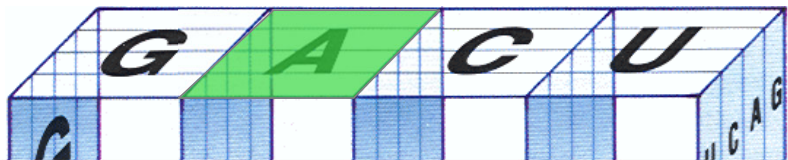
Common ways of depicting transfer RNA (tRNA)



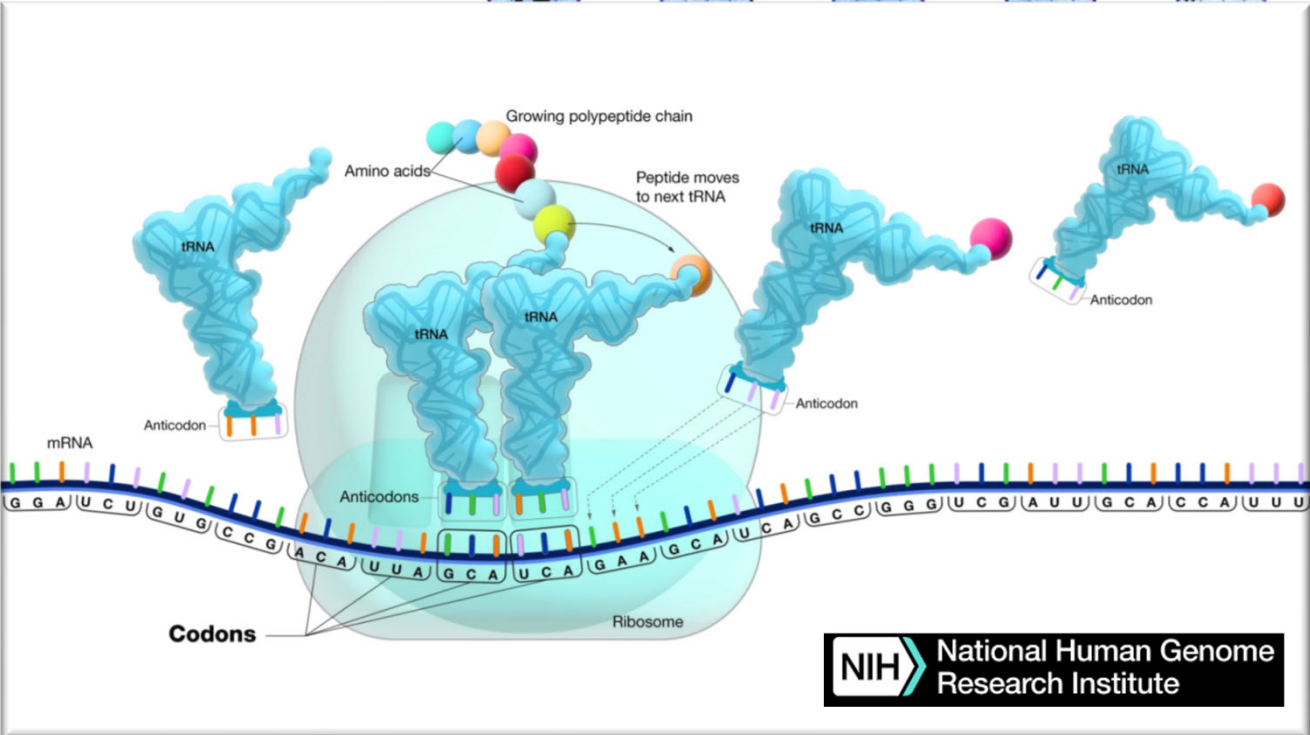
Figures from Eigen [1992] . *Steps Towards Life*.

between **genotype** and **phenotype**

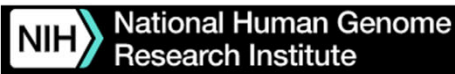
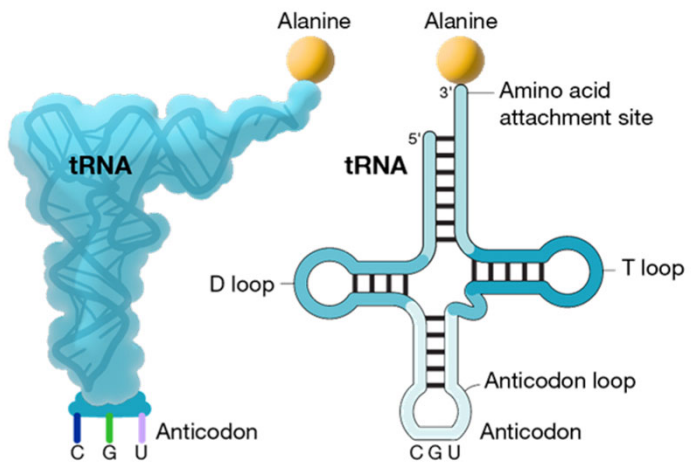
Triplets of 3 Nucleotides can define 64 possible codons, but only 20 amino acids are used (redundancy)



- The genetic code maps information stored in the genome into functional proteins
 - Triplet combinations of nucleotides into amino acids



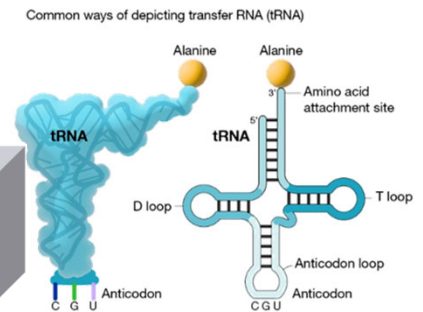
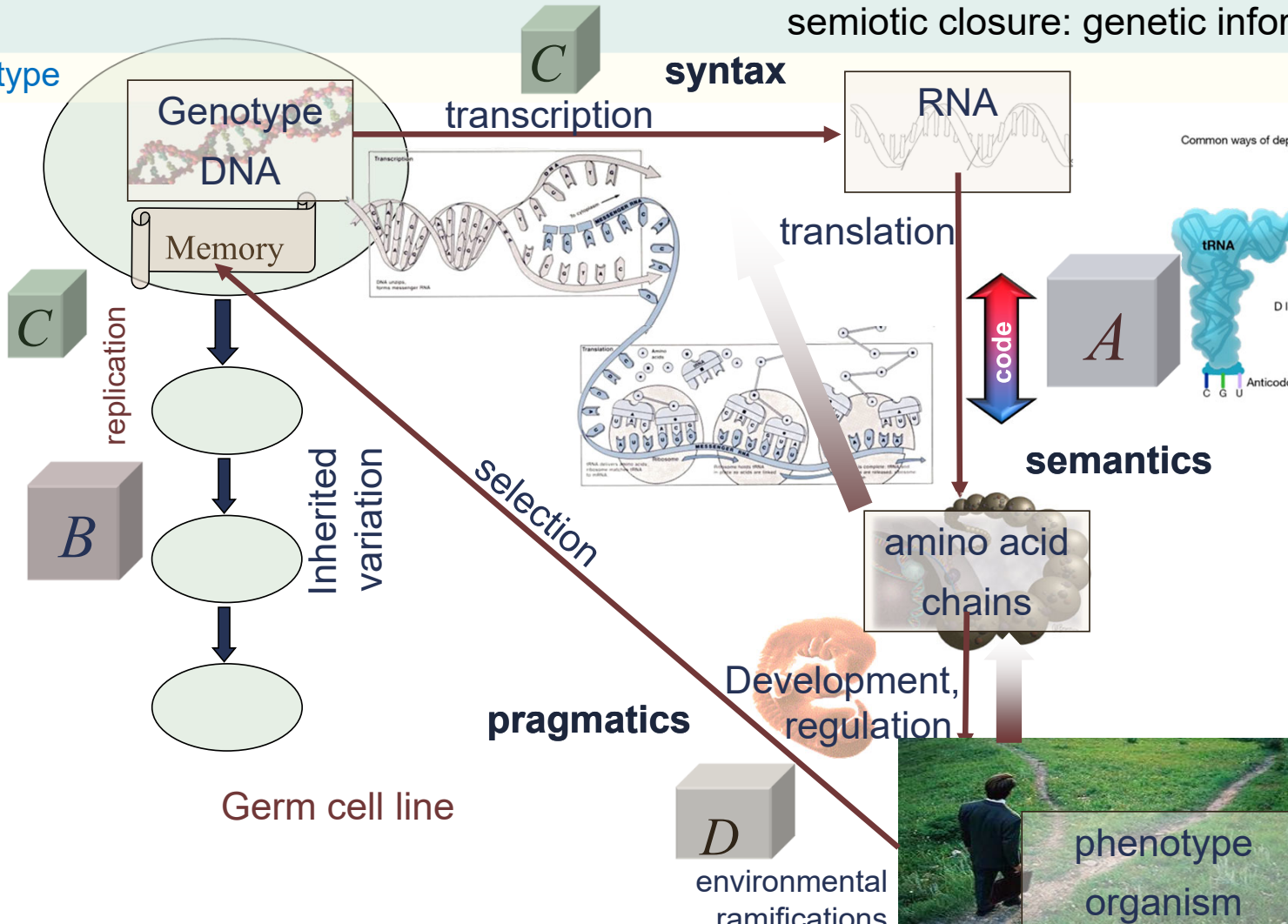
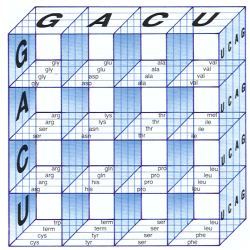
Common ways of depicting transfer RNA (tRNA)



rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

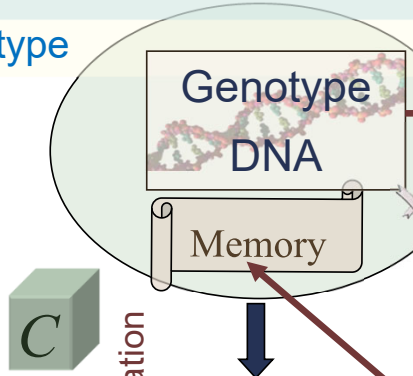
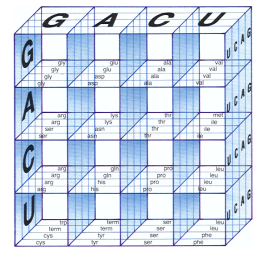
semiotic closure: genetic information at work

genotype/phenotype

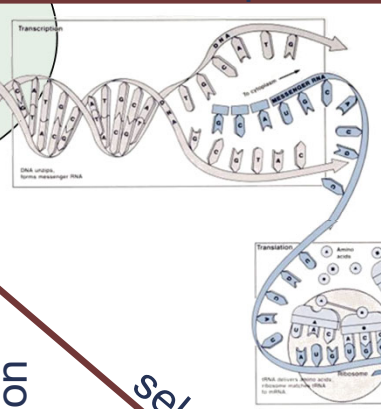


semiotic closure: genetic information at work

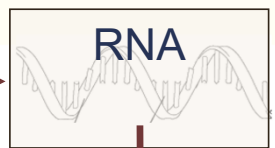
genotype/phenotype



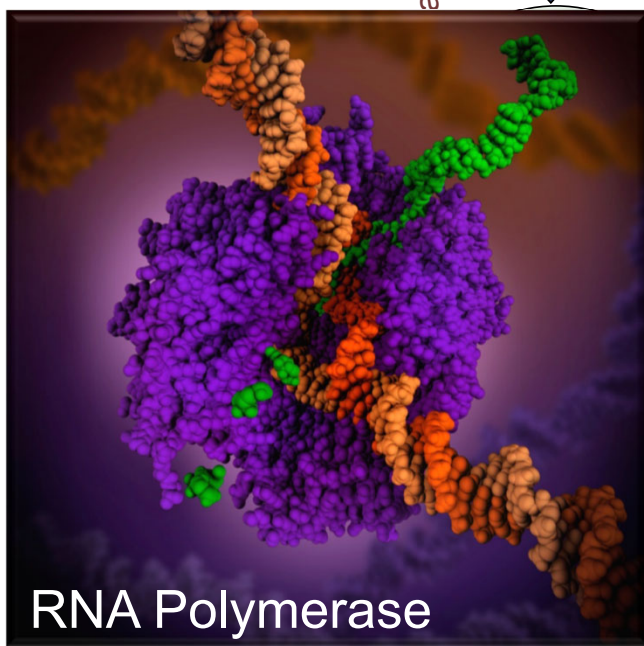
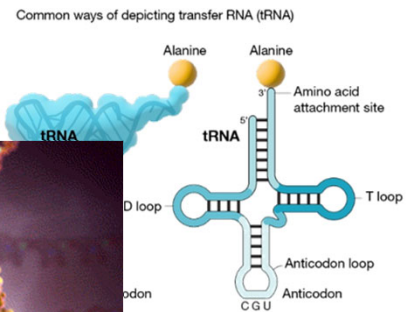
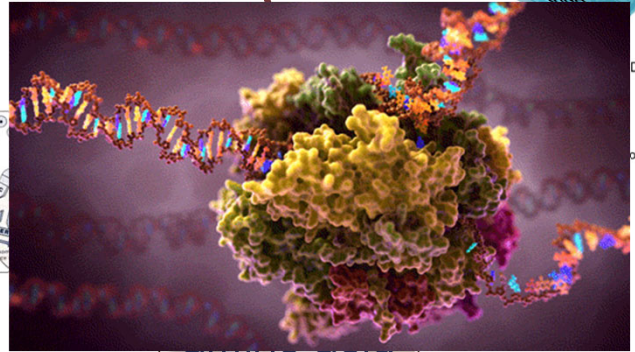
transcription



syntax



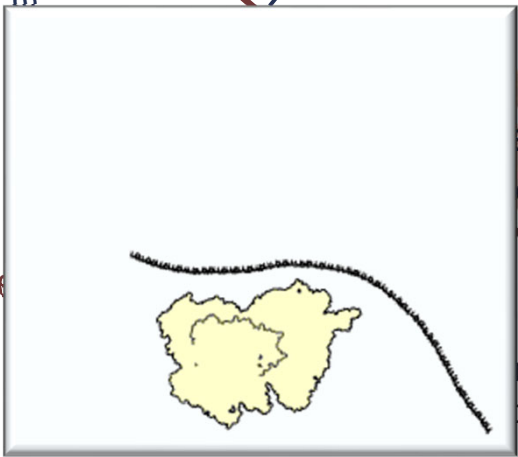
translation



RNA Polymerase

Inherited variation

selection



chains

developmental regulation

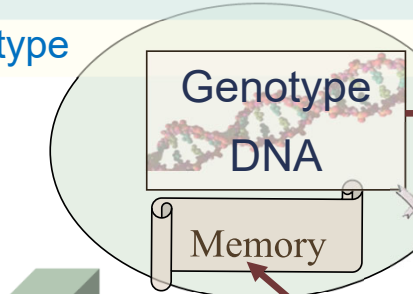
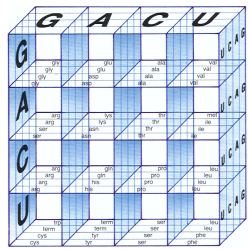
experimental applications



Ribosome and tRNA

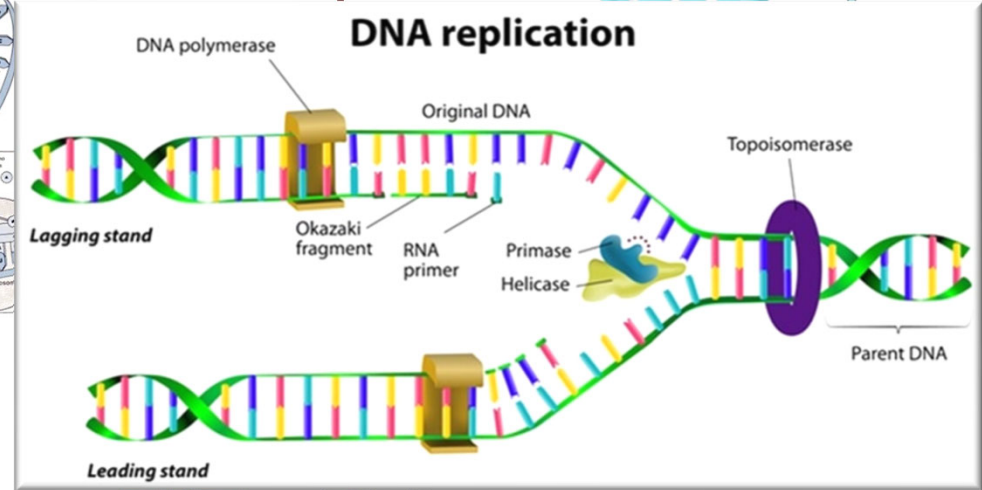
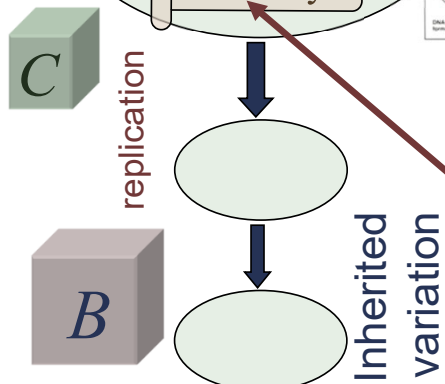
semiotic closure: genetic information at work

genotype/phenotype



C
transcription

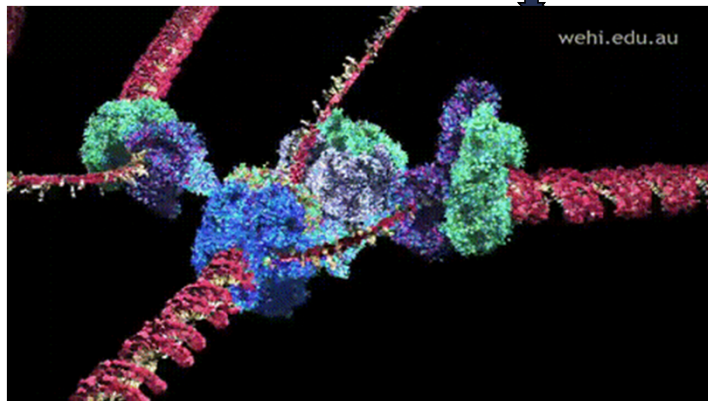
syntax



selection

pragmatics

Development, regulation



line

D
environmental ramifications



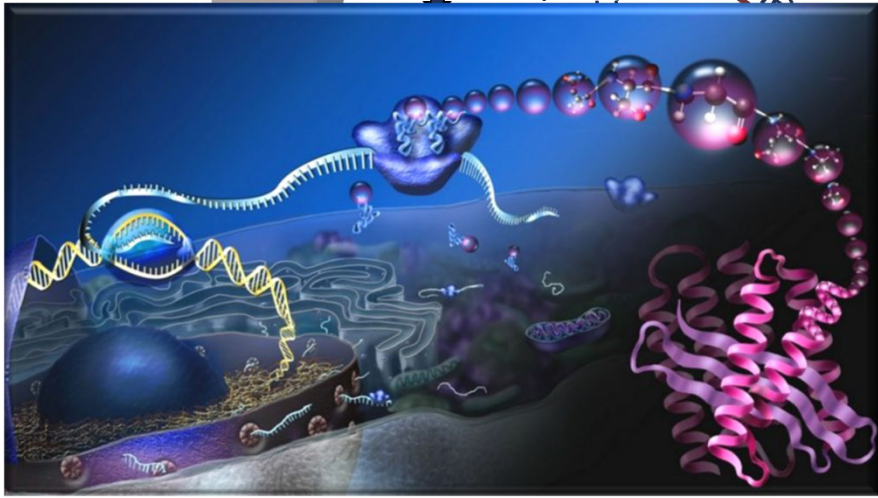
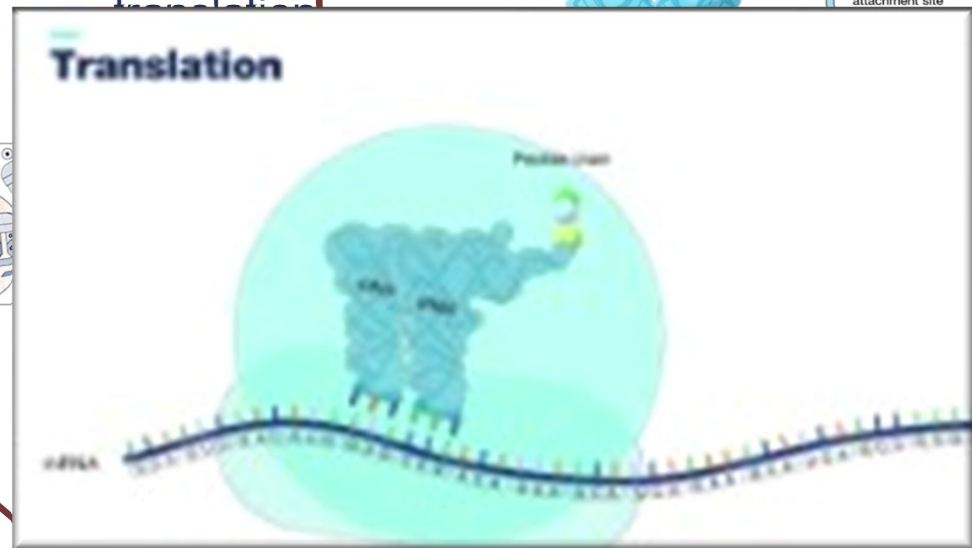
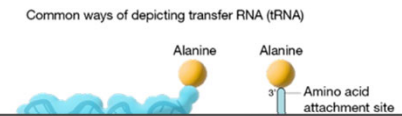
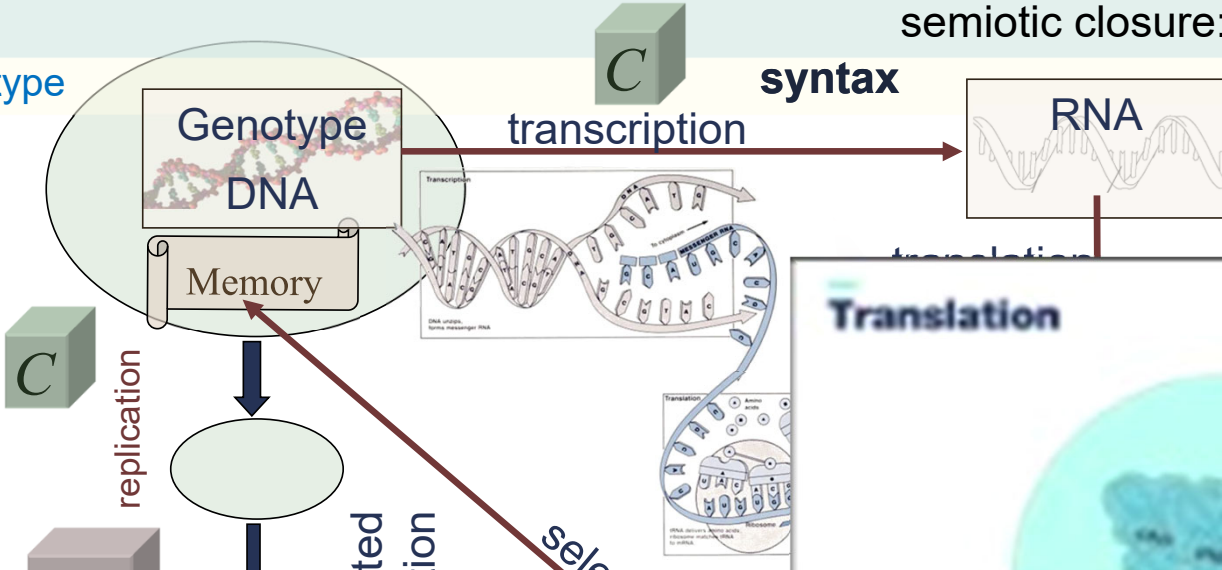
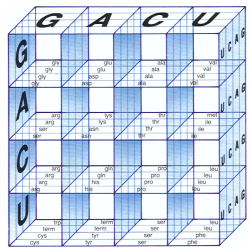
phenotype organism



rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

semiotic closure: genetic information at work

genotype/phenotype



atics

regulation ↓

D
environmental
ramifications



phenotype
organism



rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

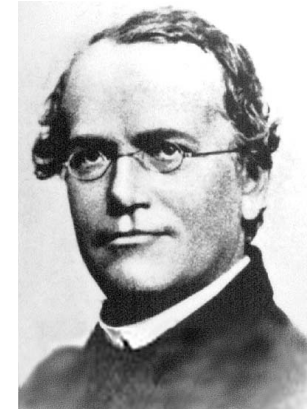
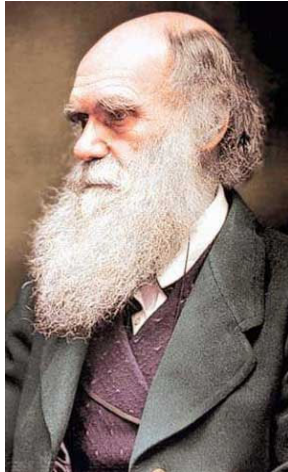
■ The “information turn”

- Unlike Schrödinger, Turing and Von Neumann had no direct effect on molecular biology
- But the “external tape” separated from the constructor (semiotic closure) has become an unavoidable **principle of organization of biocomplexity**
- A new synthesis?
 - In 1971 Brenner: “in the next twenty-five years we are going to have to teach biologists another language still, [...] where a science like physics works in terms of laws, or a science like molecular biology, to now, is stated in terms of mechanisms, maybe now what one has to begin to think of is algorithms. Recipes. Procedures.”



“The concept of the gene as a symbolic representation of the organism — a **code script** — is a fundamental feature of the living world and must form the kernel of biological theory. [...] at the core of everything are the tapes containing the descriptions to build these special Turing machines.” (Sydney Brenner)

fundamental principle of organisms as *cybernetic mechanisms*



readings

■ Class Book

- Klir, G.J. [2001]. *Facets of systems science*. Springer.

■ Papers and other materials

- Discussion Set 3 (Group 3): September 19th

- Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapters 1 and 2.

- Optional:

- Rosen, R. [1986]. "Some comments on systems and system theory". *Int. J. of General Systems*, **13**: 1-3. Available in: Klir, G.J. [2001]. *Facets of systems Science*. Springer. pp: 241-243.

- Wigner, E.P. [1960], "The unreasonable effectiveness of mathematics in the natural sciences". Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Comm. Pure Appl. Math*, **13**: 1-14.

- Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 3.

