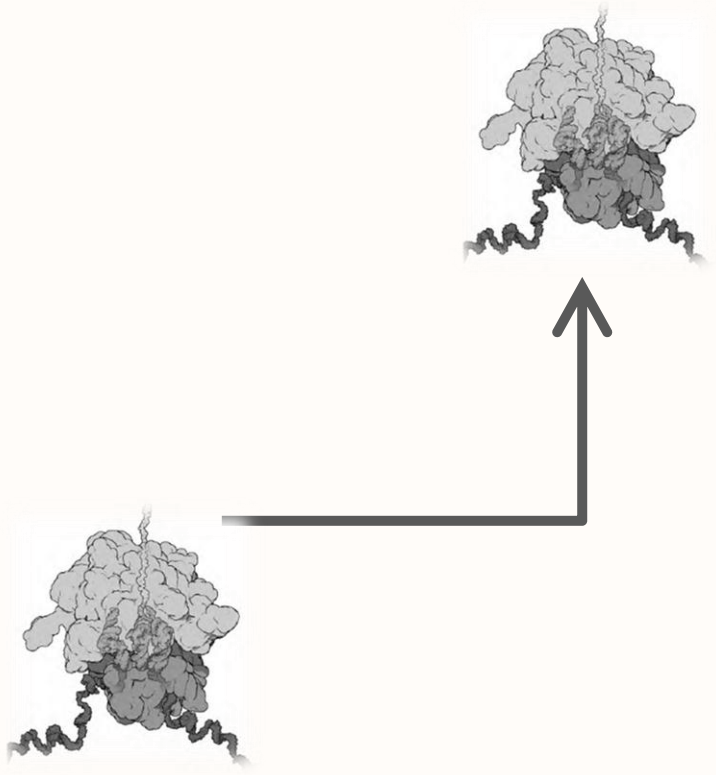
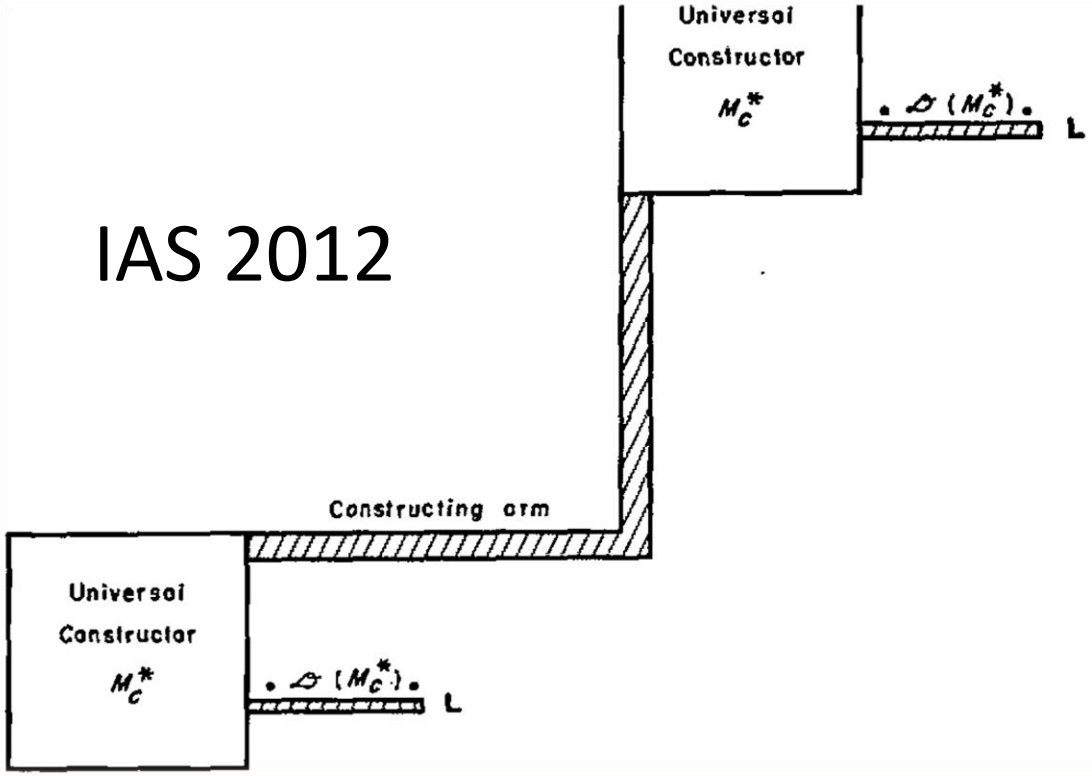
RIBOSOME: THE ENGINE OF THE LIVING
VON NEUMANN'S CONSTRUCTOR



IAS 2012



Von Neumann's universal constructor

Self-reproducing machine: constructor + tape (1948/9).

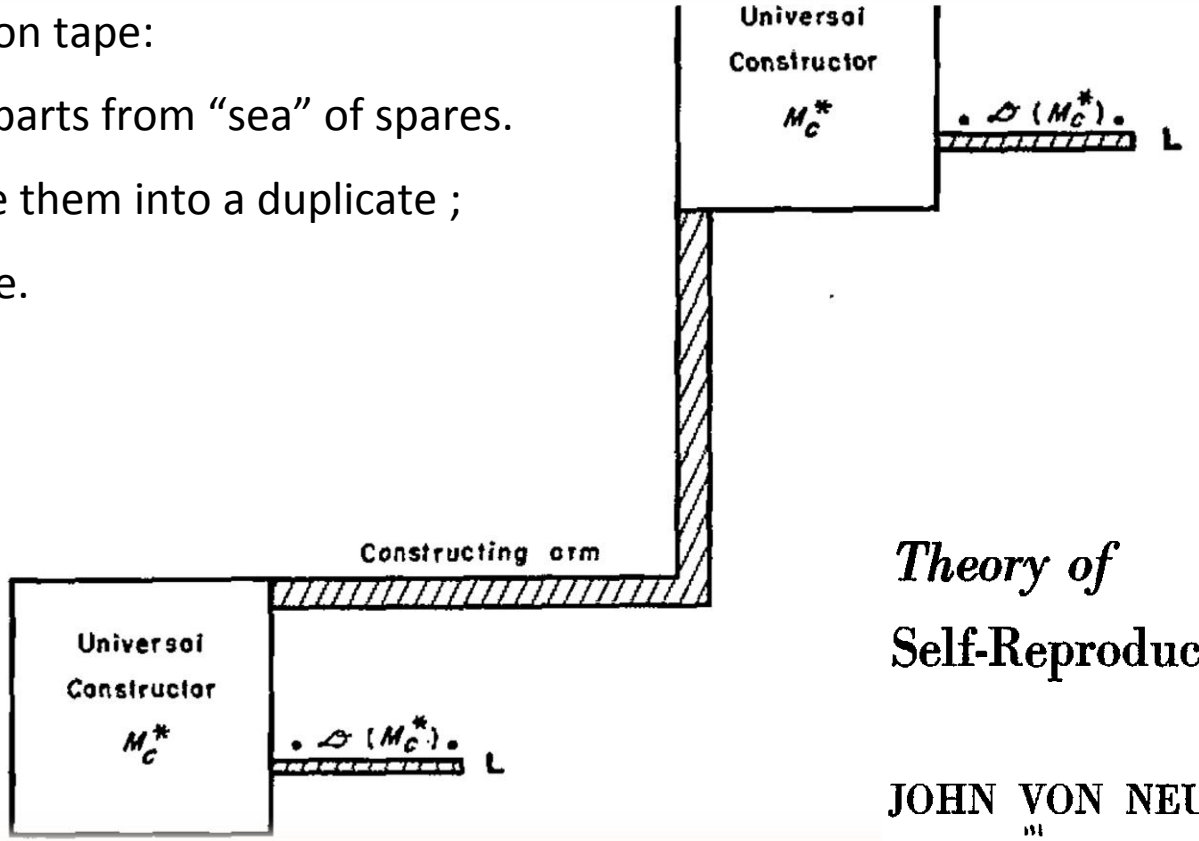
- Program on tape:

(i) retrieve parts from "sea" of spares.

(ii) assemble them into a duplicate ;

(iii) copy tape.

.



Theory of
Self-Reproducing Automata

(1966)

JOHN VON NEUMANN

edited and completed by Arthur W. Burks

Von Neumann's design allows open-ended evolution

Motivated by biological self-replication:

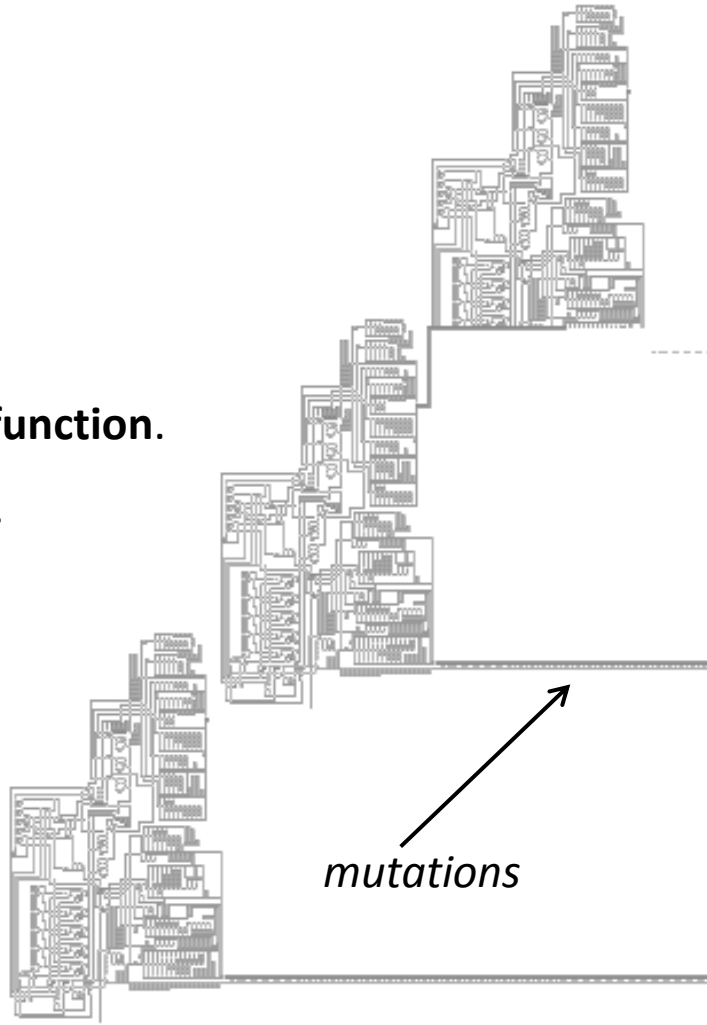
- Construction universality.
- Evolvability.

Key insight **(**before DNA) **separation of information and function**.

- Tape is read twice: for construction and when copied.

How to design

fast/accurate/compact constructor?



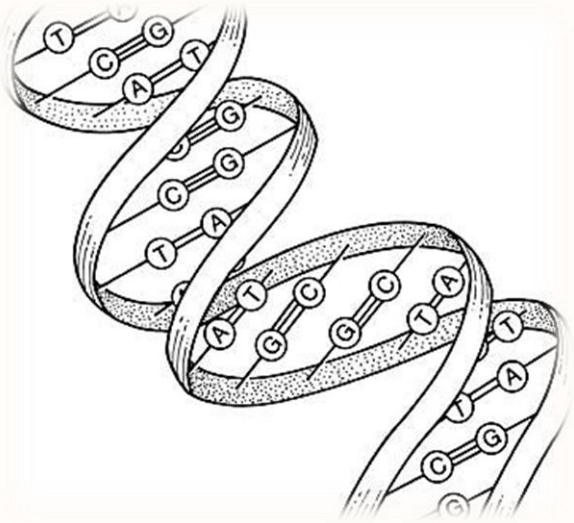
Implementation by Nobili & Pesavento (1995)

Dual spaces of DNA and proteins

DNA

- Building blocks :

  4 nucleic bases = {A, T, G, C}.

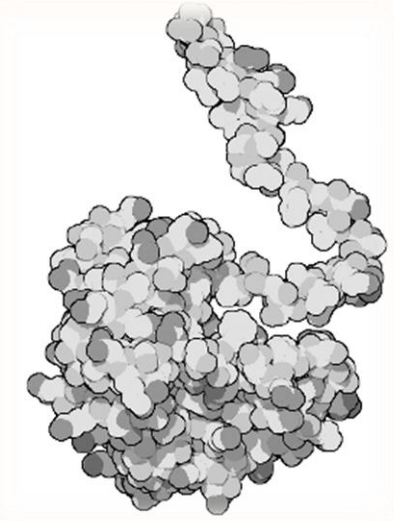


- Polymer: DNA double-helix.
- Inert information storage (“tape”)

protein

- Building blocks:

  20 amino acids.



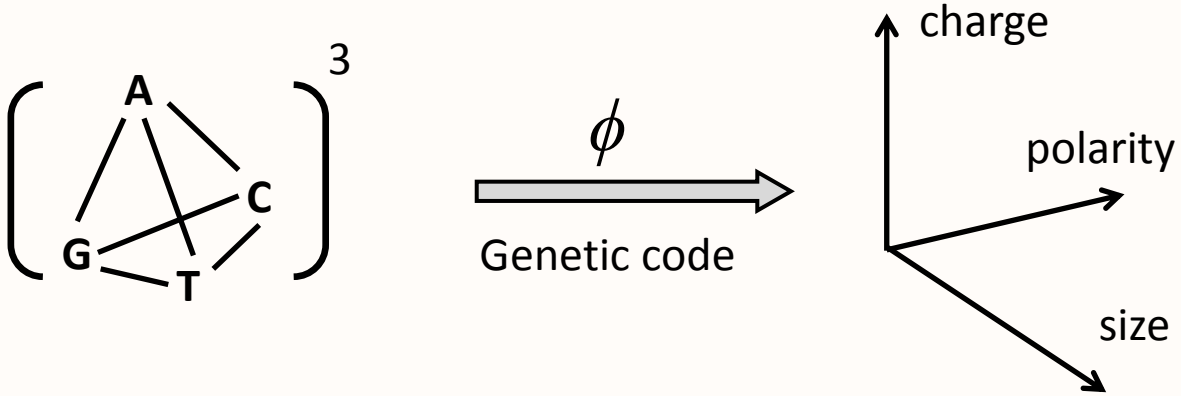
- Polymer = protein.
- Functional molecules (“constructor”)

The genetic code maps DNA to protein

- **Genetic code**: maps 3-letter words in 4-letter DNA language ($4^3 = 64$ codons) to protein language of 20 amino acids.

$$\text{codon} = b_1 b_2 b_3,\ b_i \in \{\text{A, T, G, C}\}.$$
$$\phi(\text{codon}) \rightarrow \text{amino-acid}.$$

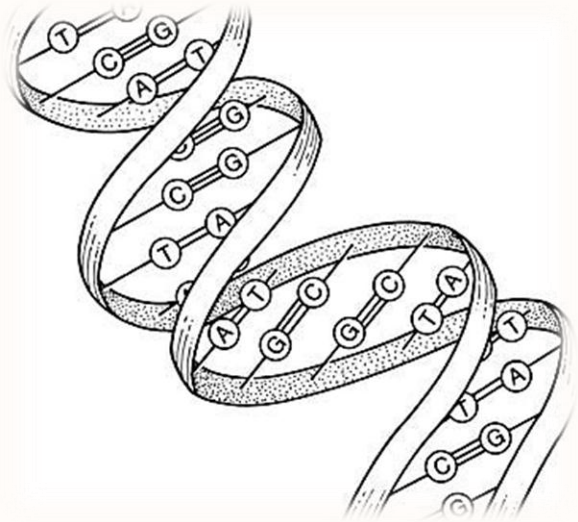
- Genetic code embeds the codon-graph (Hamming graph) into space of amino-acids.



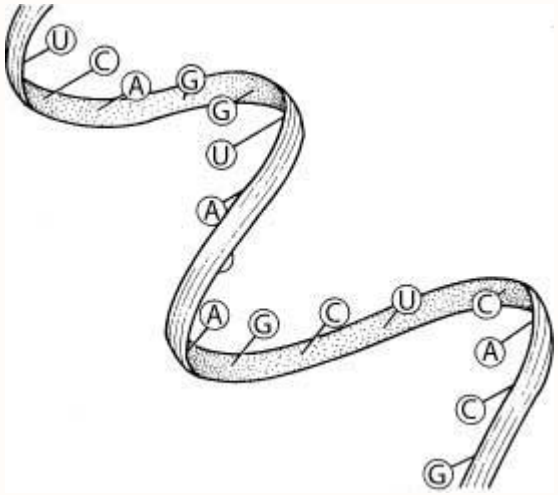
- **Translation machinery**, whose main component is the **ribosome**, facilitates the map.

RNA intermediates can be both tapes and machines

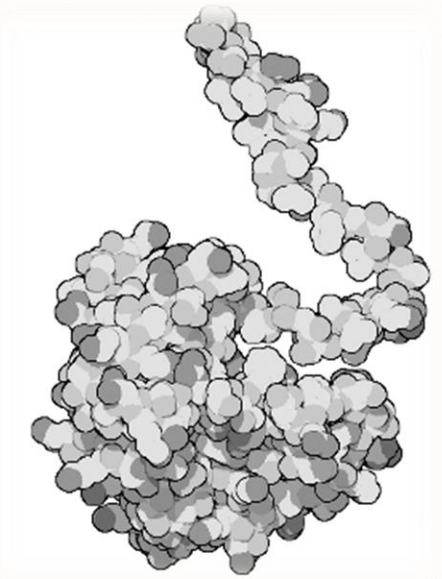
DNA



RNA



protein

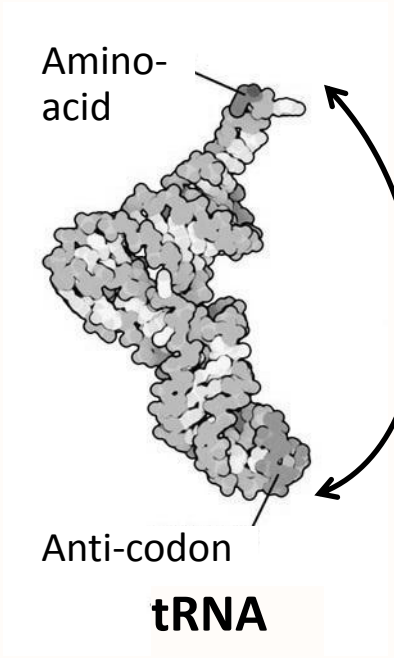


- Primordial “RNA world” :

RNA molecules are both information carriers (DNA) and executers (proteins).

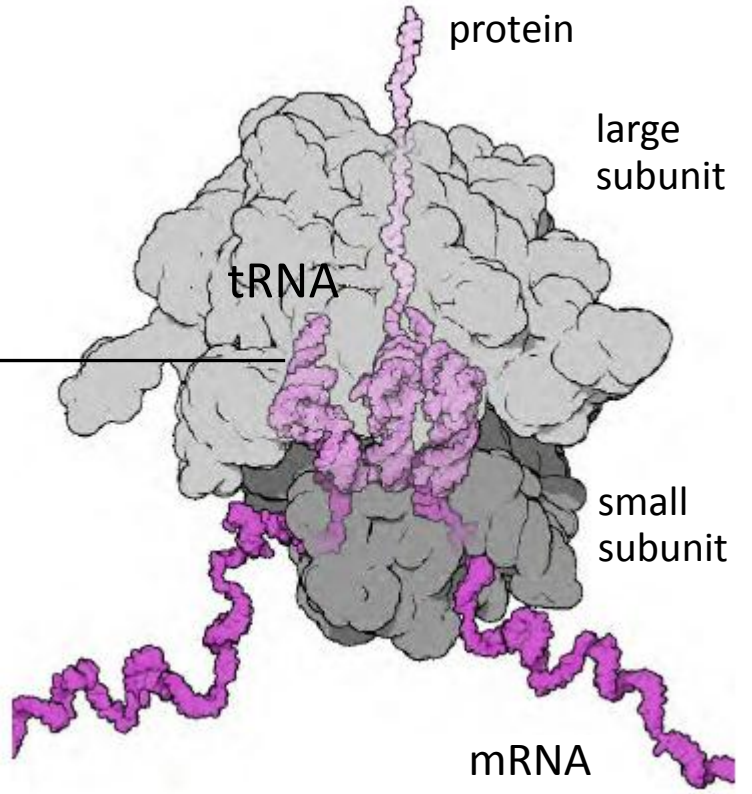
Ribosomes translate nucleic bases to amino acids

- **Ribosomes** are *large* molecular machines that synthesize proteins with **mRNA** blueprint and **tRNA**s that carry the genetic code.



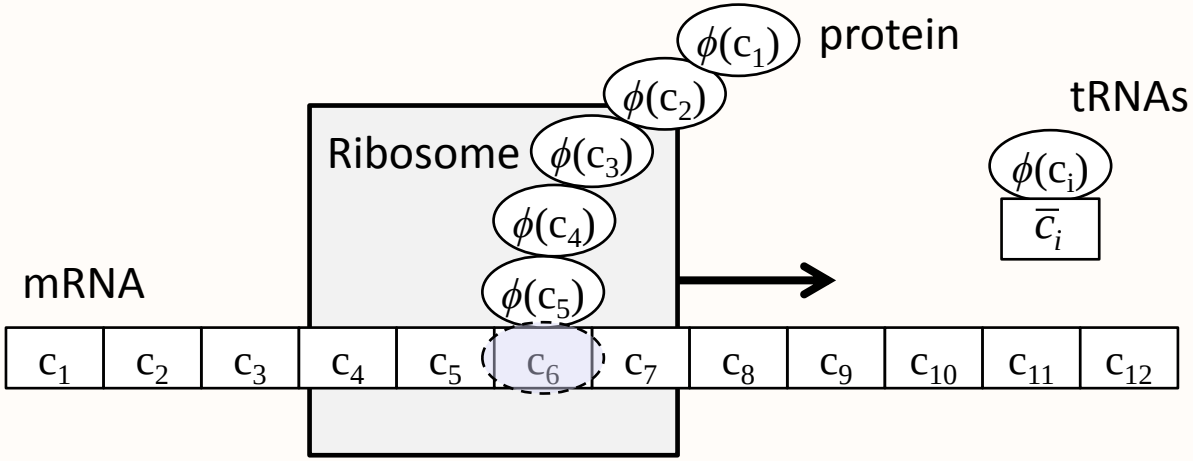
genetic code

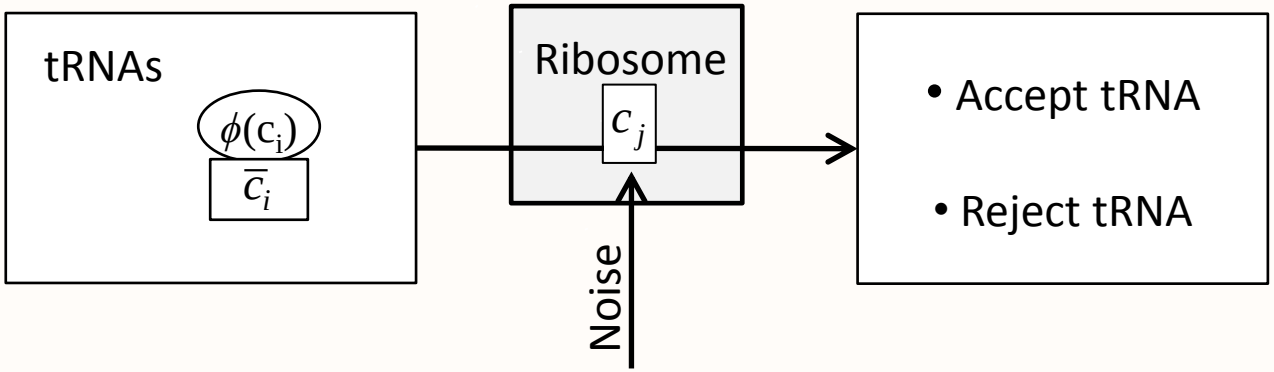
$$\text{amino-acid} = \phi(\text{codon})$$



Goodsell, *The Machinery of Life*

Ribosome needs to recognize the correct tRNA



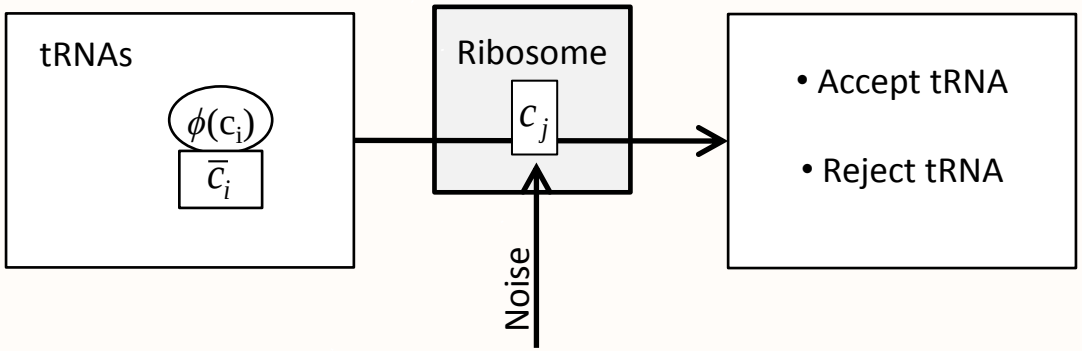


(i) binding wrong tRNAs: $\text{amino-acid} \neq \phi(\text{codon})$

(ii) unbinding correct tRNAs: $\text{amino-acid} = \phi(\text{codon})$

How to construct fast\accurate\small ***molecular*** decoder ?

Decoding at the ribosome is a molecular recognition problem



- Central problem in biology and chemistry:

How to evolve molecules that recognize in a noisy environment?

(crowded, thermally fluctuating, weak interactions).

- How to estimate recognition performance ("fitness")?
- What are the relevant degrees-of-freedom? **Dimension**? **Scaling**?
- What is the role of conformational changes?

Ribosome sets physical limit on self-reproduction rate

Large fraction of cell mass is ribosomes.

- For self-reproduction each ribosome should self-reproduce.
- Sets lower bound on self-reproduction rate .

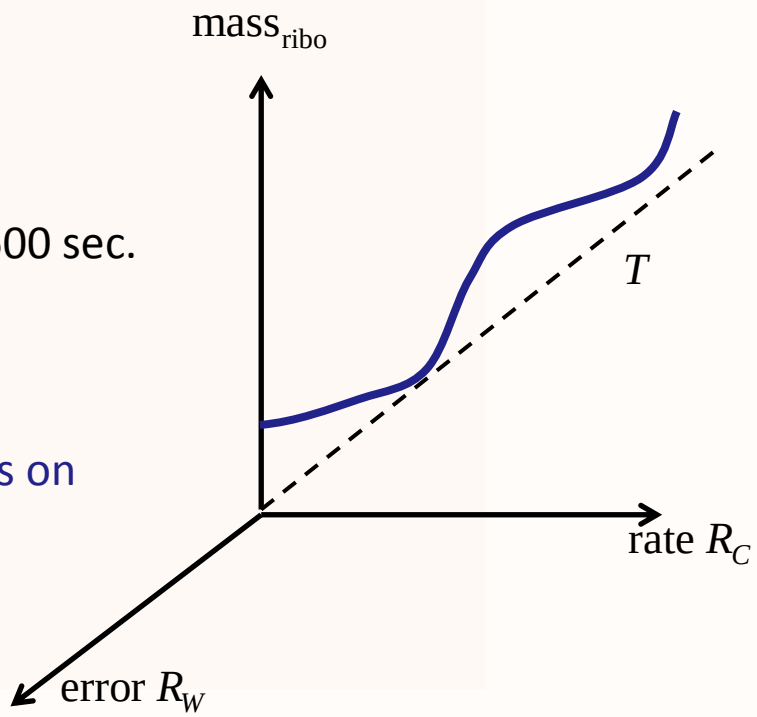
$$T \geq \frac{\text{mass}_{\text{ribo}}}{R_C} \approx \frac{10^4 \text{ amino-acids}}{20 \text{ amino-acids/sec}} = 500 \text{ sec}$$

- Fastest growing bacteria (*Clostridium perfringens*): T ~ 600 sec.

Problem: how ribosome accuracy affects fitness depends on

(i) Basic protein properties (mutations).

(ii) Biological context (environment etc.).



Ribosomes are complicated machines with many d.o.f.

Ribosomes are made of proteins and RNAs:

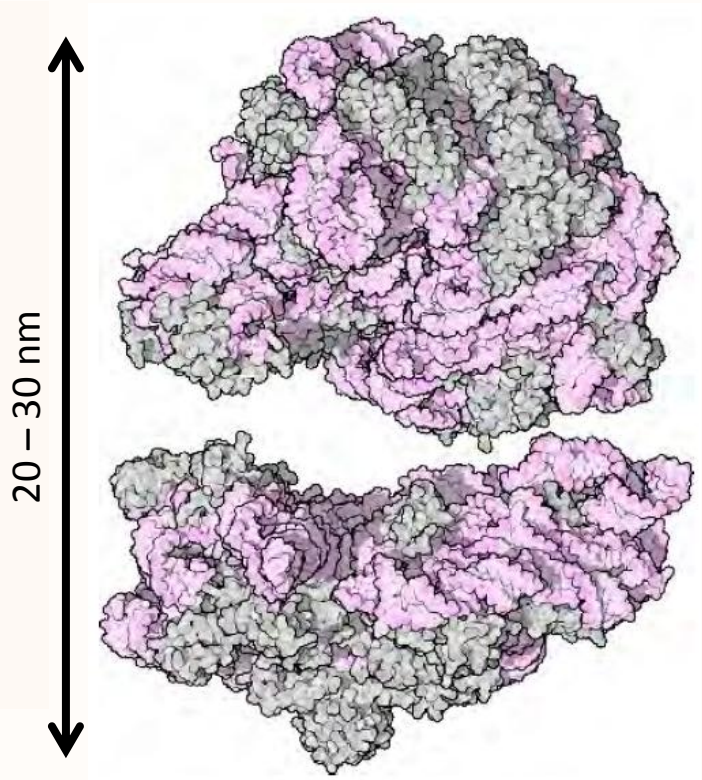
- ~ 10^4 nucleic bases in RNA.
- ~ 10^4 amino-acids in proteins.
- Total mass : ~ $3 \cdot 10^6$ a.u.

- High-res structure is known (Yonath et al.).

Within this known complexity:

- **What are the relevant degrees-of-freedom?**

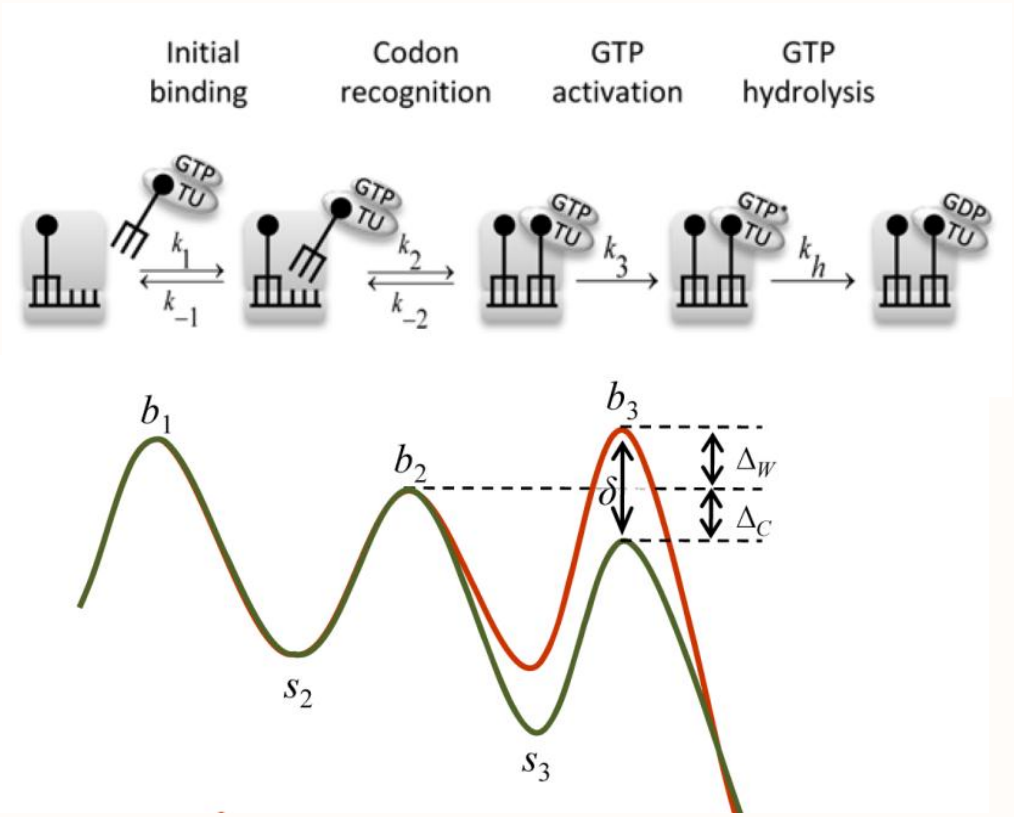
- **How does this machine operate?**



(magenta – RNA, grey – protein, from Goodsell, *Nanotechnology*)

Decoding is determined by energy landscapes of correct and wrong tRNAs

- Decoding is multi-stage process.
- Kinetics involves *large* conformational changes.



Steady-state decoding rates

(Arrhenius law, $k \propto e^{-\Delta G}$)

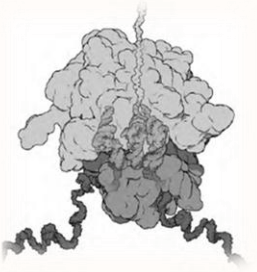
$$R_C \sim \frac{1}{e^{b_1} + e^{b_2} + e^{b_3}}$$

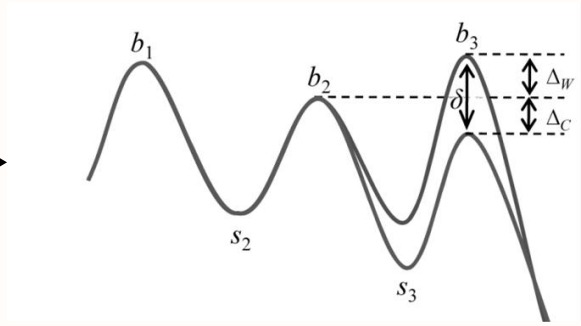
$$R_W \sim \frac{1}{e^{b_1} + e^{b_2} + e^{\boldsymbol{\delta}+b_3}}$$

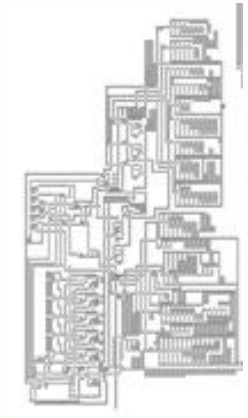
Ribosome kinetics exhibits large dimensionality reduction

- Effective dimension decreases by at least 3 orders of magnitude:

~ 10^4 structural parameters → ~ 10 kinetic parameters (energy landscape).







- Generic phenomenon in biomolecules: many catalytic molecules (enzymes) can be described by a few kinetic parameters (transition state landscape).

What is the origin of dimensionality reduction?

- Hints:

- Protein function mainly involve the lowest modes of their vibrational spectra (hinges).

- Sectors: “Normal modes” of sequence evolution (Leibler & Ranganthan).

Transition states reduce the dimensionality of effective parameter space

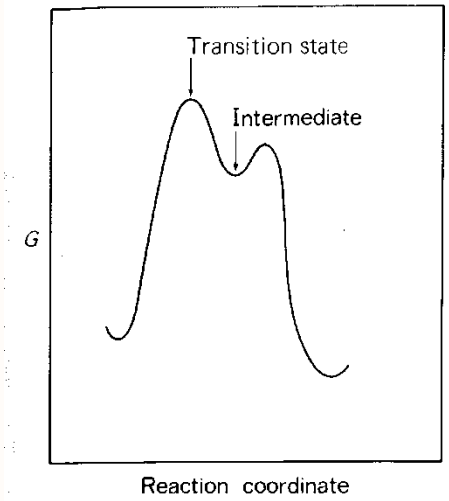


Figure 2.1 Transition states occur at the peaks of the energy profile of a reaction, and intermediates occupy the troughs.

$$k_1 = \left(\frac{kT}{h}\right) \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right)$$

FEBRUARY, 1935 JOURNAL OF CHEMICAL PHYSICS VOLUME 3

The Activated Complex in Chemical Reactions

HENRY EYRING, *Frick Chemical Laboratory, Princeton University*

(Received November 8, 1934)

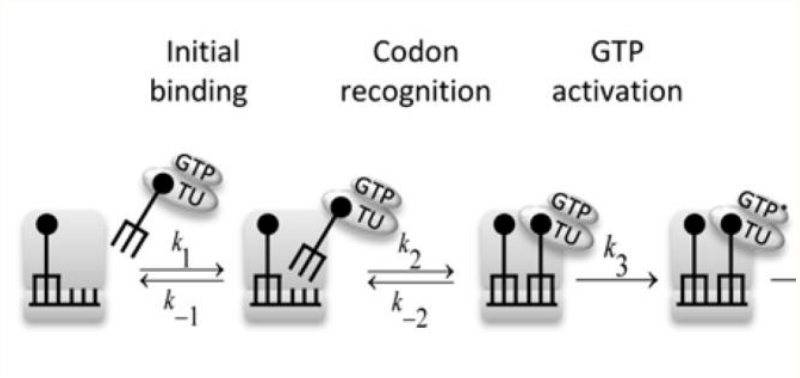
leave no doubt of the proposed method of procedure in a particular case. We may write for the specific reaction rate constant for a reaction of any order

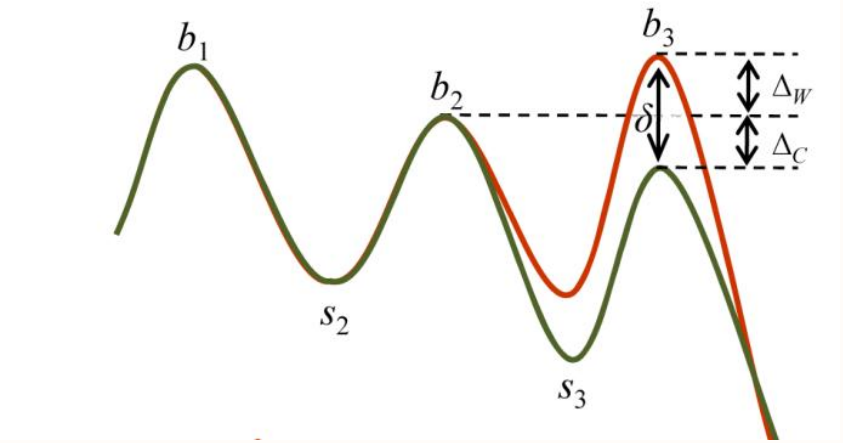
$$k_i = c(F_a/F_n)(\bar{p}/m^*) = c(F_a'/F_n)(kT/h)e^{-E_0/kT} \qquad (10)$$

where F_a is simply the partition function (or Zustandsumme) for the activated state and F_n is the same quantity for the normal state. F_a' is the partition function for the activated complex for all the normal coordinates except the one in which decomposition is occurring. The partition function for this normal coordinate is included in the factor $(kT/h)e^{-E_0'/kT}$. The other quantities have been defined.

Theory can be tested with measured rates

- The codon-specific stages are Codon recognition and GTP activation.





	(UUU) Cognate	(CUC) Non-cognate
k_1	100-140 1/(μM·s)	100-140 1/(μM·s)
k_{-1}	80-100 1/s	80-100 1/s
k_2	190 1/s	190 1/s
k_{-2}	0.23 1/s	100 1/s
k_3	260 1/s	0.6 1/s

(Rodnina's lab, Gottingen)

$$R_C \sim \frac{1}{e^{b_1} + e^{b_2} + e^{b_3}}$$

$$R_W \sim \frac{1}{e^{b_1} + e^{b_2} + e^{\boldsymbol{\delta}+b_3}}$$

How to estimate recognition performance (“fitness”) ?

What is the actual dimension of the problem ?

Recognition fitness has generic features

$$R_C \sim \frac{1}{e^B + e^{b_3}}$$

$$R_W \sim \frac{1}{e^B + e^{\delta + b_3}}$$

$$(e^B = e^{b_1} + e^{b_2})$$

- "Fitness" *F* is often obscure and context-dependent.:

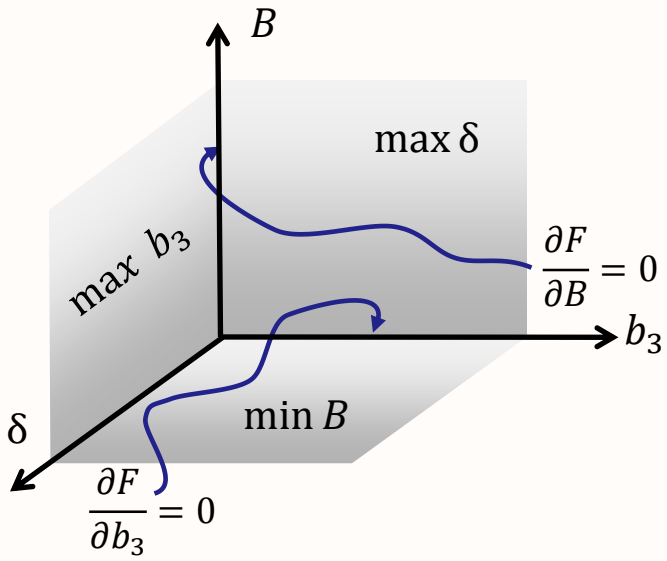
  → look for generic properties of $F(R_C, R_W) = F(B, \delta, b_3)$.

- Only requirement: "biologically reasonable", $\frac{\partial F}{\partial R_C} \geq 0$, $\frac{\partial F}{\partial R_W} \leq 0$.

- Searching for optimum in (B, δ, b_3) space:

(i) $\frac{\partial F}{\partial \delta} \geq 0$: $\boldsymbol{\delta}$ **approaches biophysical limit**.

(ii) $\left\{\frac{\partial F}{\partial B} = 0 \;\&\; \frac{\partial F}{\partial b_3} \geq 0\right\}$ or $\left\{\frac{\partial F}{\partial B} \leq 0 \;\&\; \frac{\partial F}{\partial b_3} = 0\right\}$:



Optimization is essentially 1D (2 other parameters approach limit).

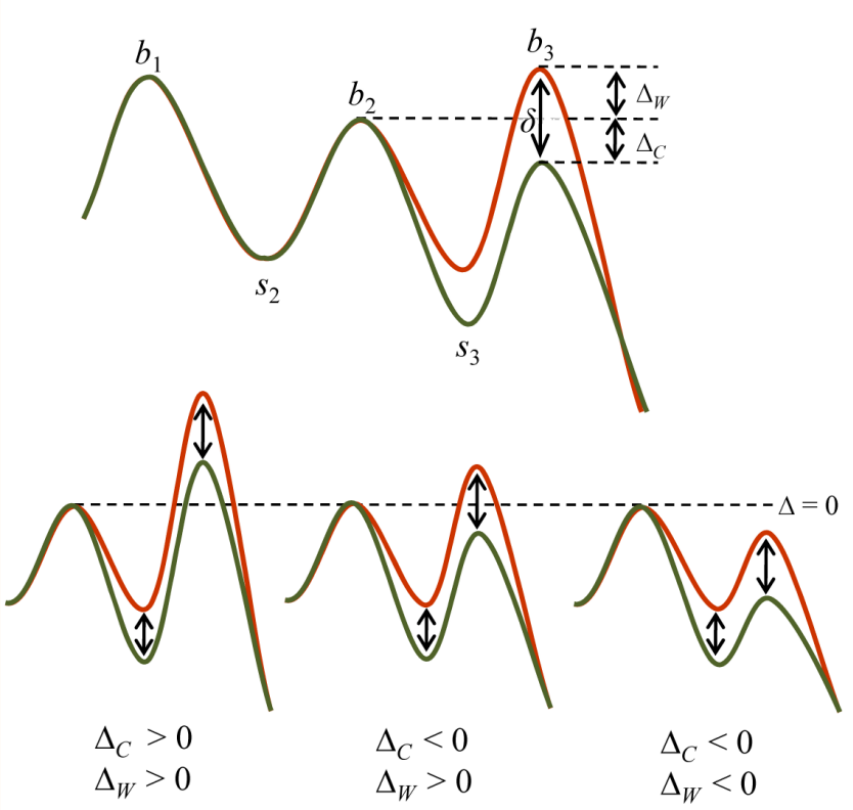
What is the optimal energy landscape of the ribosome?

- For example, distortion fitness from engineering (weight d is context-dependent) :

$$F = R_C - d \cdot R_W \propto \frac{1}{e^B + e^{b_3}} - \frac{d}{e^B + e^{b_3+\delta}}$$

- 1D problem: optimum is along b_3.

(measured: $\Delta = b_3 - b_2$)



- **What is the optimal b_3 (or Δ)?**
- **Is the ribosome optimal ?**
- **Role of conformational changes ?**

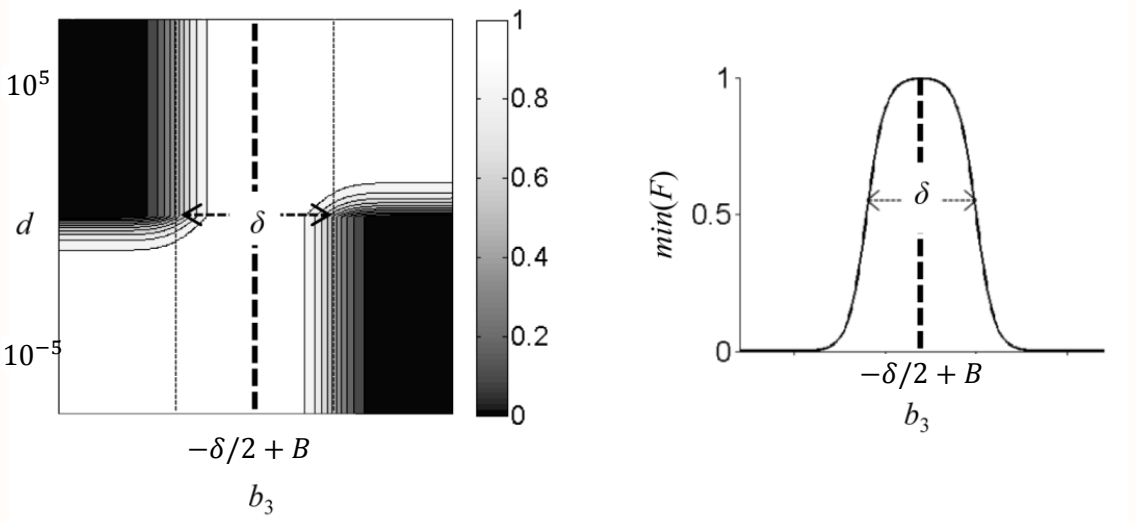
Optimal design is a Max-Min strategy

- Weight d can vary.

(i) For each d normalize F.

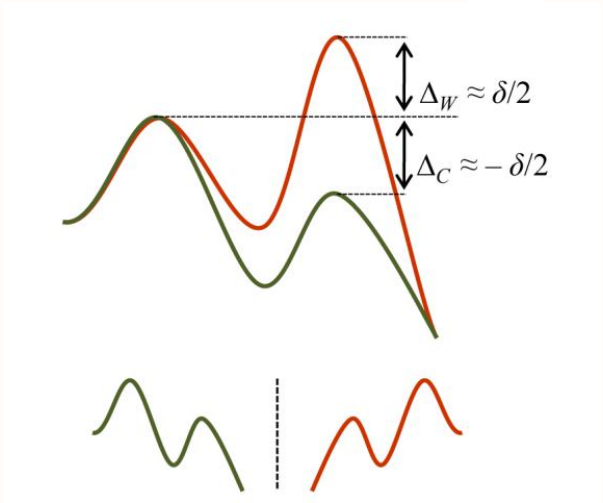
(ii) "Worst case scenario":

$$\max\big(\min(F)\big).$$



- Max-Min solution is "symmetric":

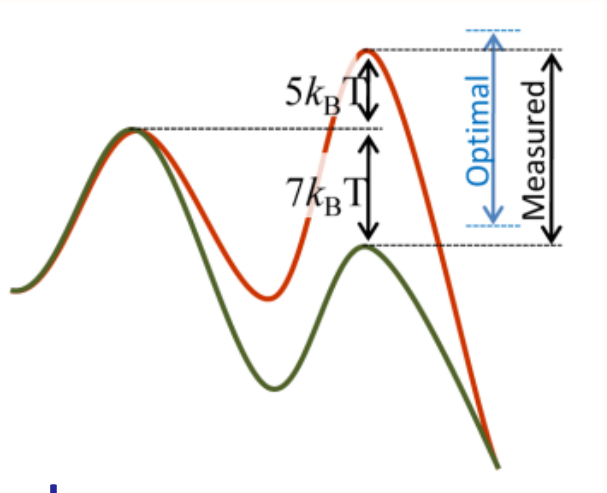
$$b_3 = -\frac{1}{2}\delta + B$$

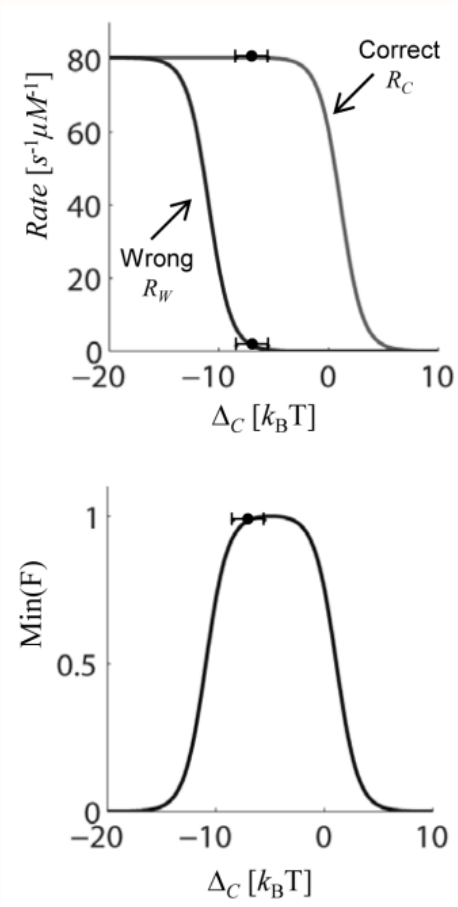


Ribosome shows an energy barrier which is nearly optimal

- Measurements: $\Delta_C \approx -7\ k_BT,\ \delta \approx 12k_BT\ , \bar{B} = B - b_2 \approx 1k_BT.$
- Prediction: the optimal regime is **symmetric**, $\Delta_C < 0,\ \Delta_W > 0.$

$$\Delta_C = -\frac{1}{2}\delta + \bar{B}$$

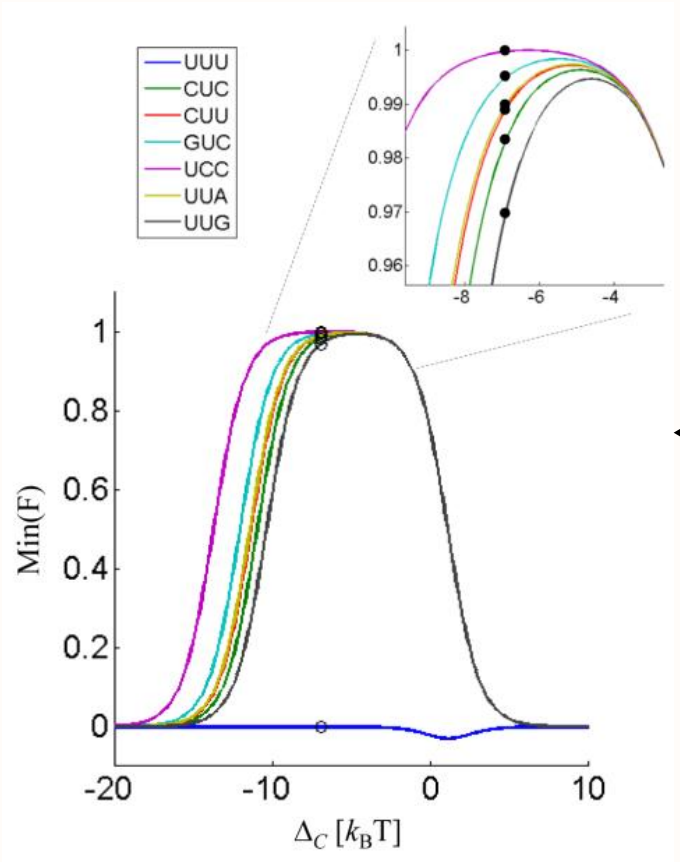


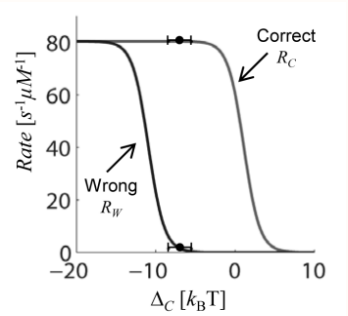


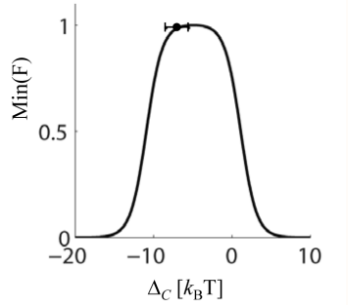
- **The ribosome is nearly optimal**

  (according to Max-Min prediction).

Decoding is optimal for all six measured tRNAs







- Except for UUC which encodes the same amino-acid

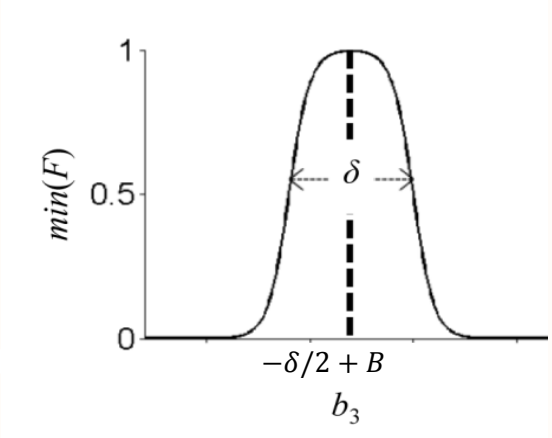
$$\phi(\mathrm{UUC}) = \phi(\mathrm{UUU}) = \text{phenylalanine.}$$

Optimality is valid for wide range of fitness functions

- Ribosome optimal in wide region:

$$5\cdot 10^{-6} = e^{-\delta} \le d \le e^{\delta} = 2\cdot 10^{5}.$$

$F = R_C - d \cdot R_W$

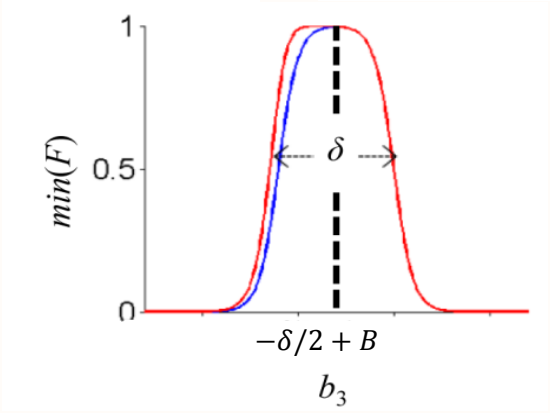


- **General feature**: any fitness function $F(R_C, R_W)$

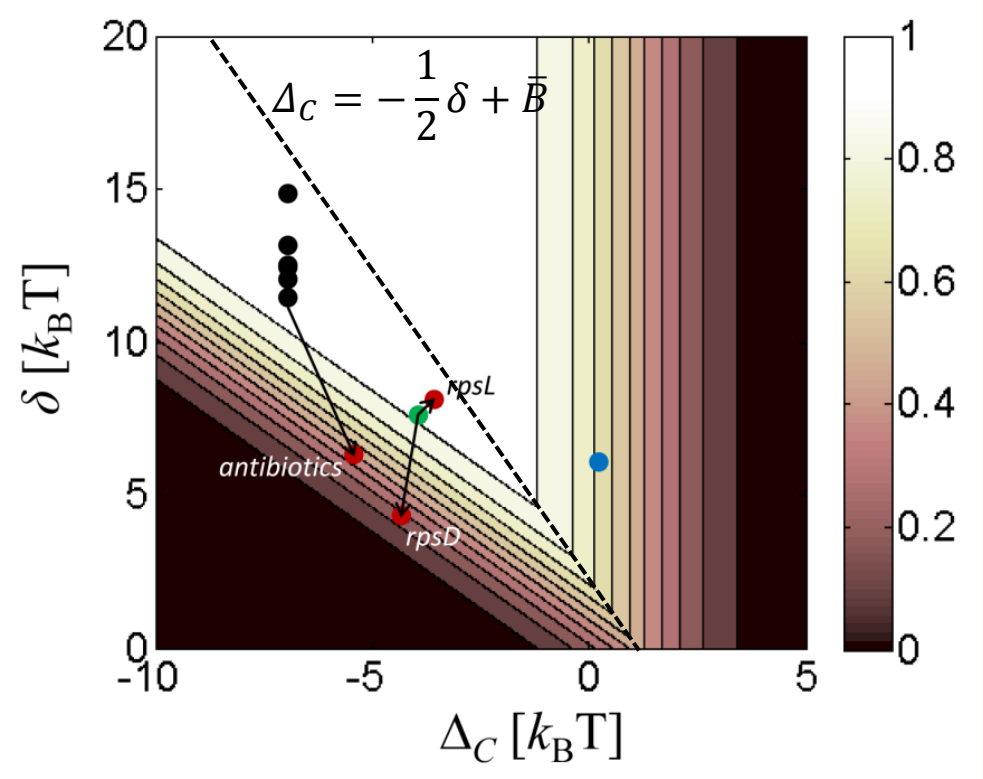
exhibits optimum as long as both rates are "relevant".

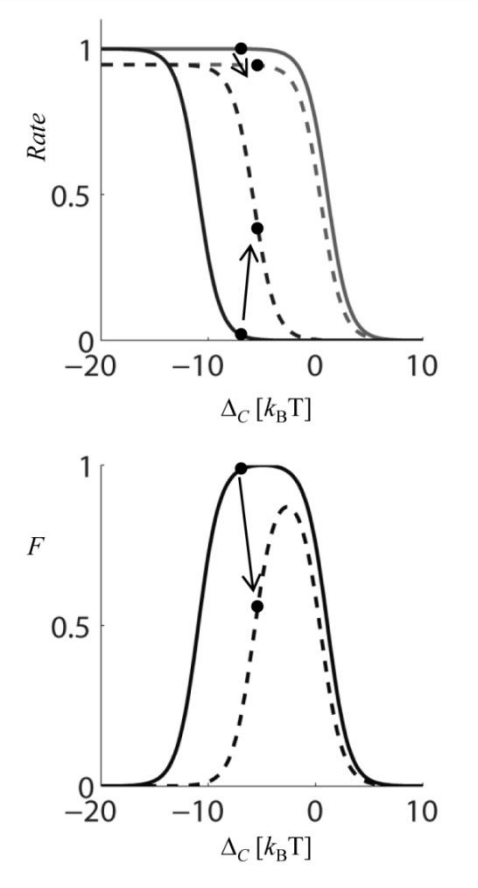
$$e^{-\delta} < \left| \frac{\partial F}{\partial R_C} \Big/ \frac{\partial F}{\partial R_W} \right| < e^{\delta}$$

$R_C - d \cdot R_W{}^2$ $\quad$ $R_C - d \cdot (R_W/R_C)$



Theory predicts optimal regime of ribosomes for all organisms



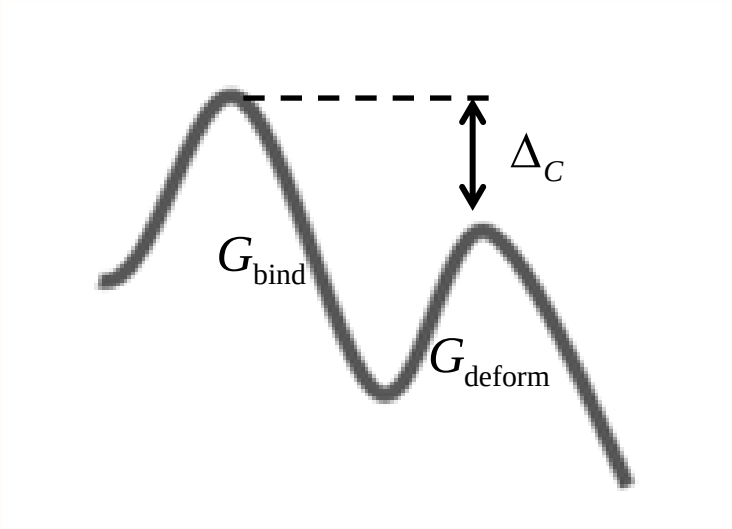


- Optimal region in the space of all possible landscapes, $-\delta \leq \Delta_C \leq 0$.
- Mutations and antibiotics tend to push away of optimality.

What is the role of conformational changes?

- Energy barrier results from binding energy and deformation energy penalty:

$$\Delta_C = G_{\text{deform}} - G_{\text{bind}} = -\frac{1}{2}\delta + \overline{B}.$$



- Therefore

$$G_{\text{deform}} = G_{\text{bind}} - \frac{1}{2}\delta + \overline{B}.$$

- For any $G_{\text{bind}} \geq -\frac{1}{2}\delta + \overline{B} \approx 5\ \text{k}_\text{B}\text{T},$

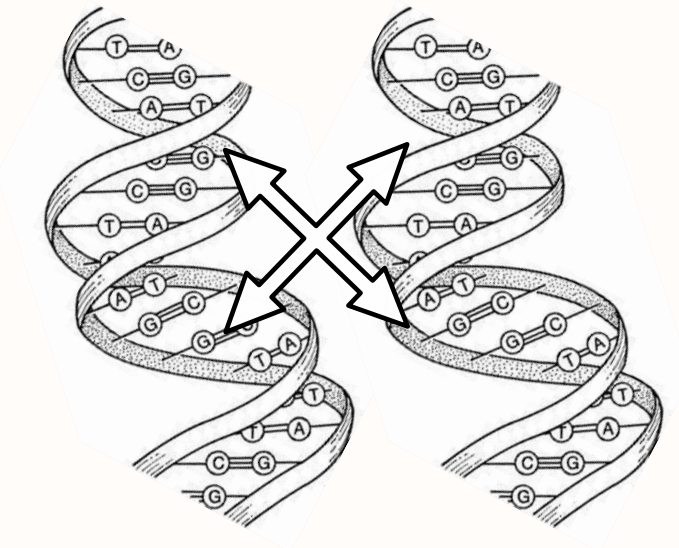
$G_{\text{deform}} \geq 0$ non-zero deformation is optimal for tRNA recognition.

Energy barrier that discerns the right target from competitors.

Yonatan Savir

Recombination machinery recognizes homologous DNA

- Exchange between two *homologous* DNAs.
- Essential for:
  - *Genome integrity* (repair machinery).
  - *Genetic diversity* (crossover, sex).

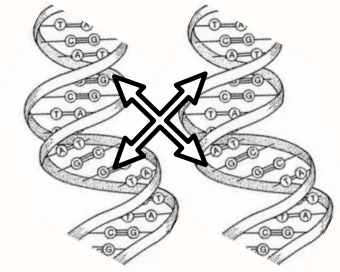


- Task: Detect correct, homologous DNA target among many incorrect lookalikes.

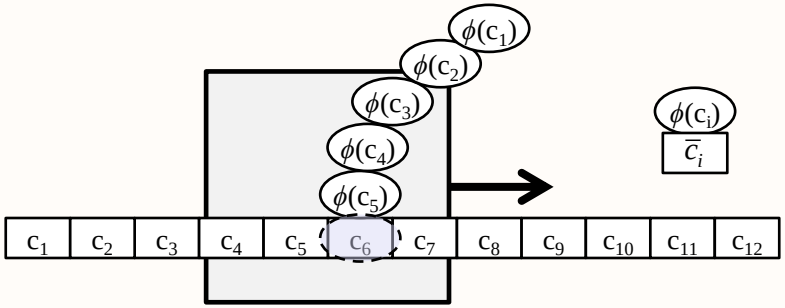
- DNA stretches during recombination:

  large deformation energy barrier.

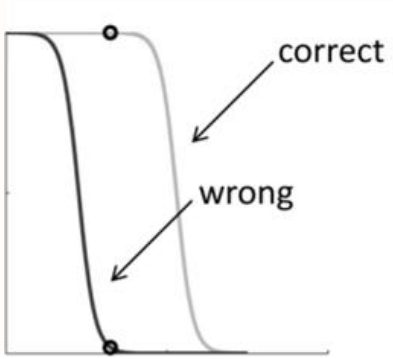
Energy barriers for optimal recognition may be a general design principle of recognition systems with competition



Recombination optimizes extension energy of dsDNA.

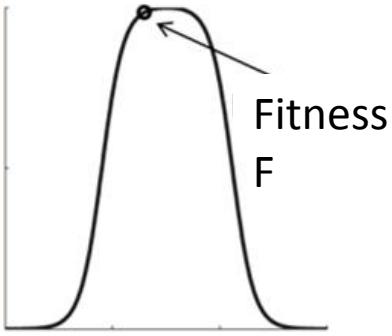


Ribosome optimizes energy barriers of decoding



Relevant energy

- **Conformational proofreading:** Design principle follows from optimization of information transfer function.
- May explain **induced fit** (Koshland 1958). Why molecules deform upon binding to target.



Relevant energy

- **Applies to any enzymatic kinetics in the presence of competition...**

Open questions, future directions…

Understanding evolvable matter:

- What are the degrees-of-freedom underlying dimensional reduction?
  (Rubisco and other enzymes)

- Basic logic of molecular information cahnnels
  (e.g. utilizing conformational changes, worst case scenario).

Translation machinery coevolved with proteins →

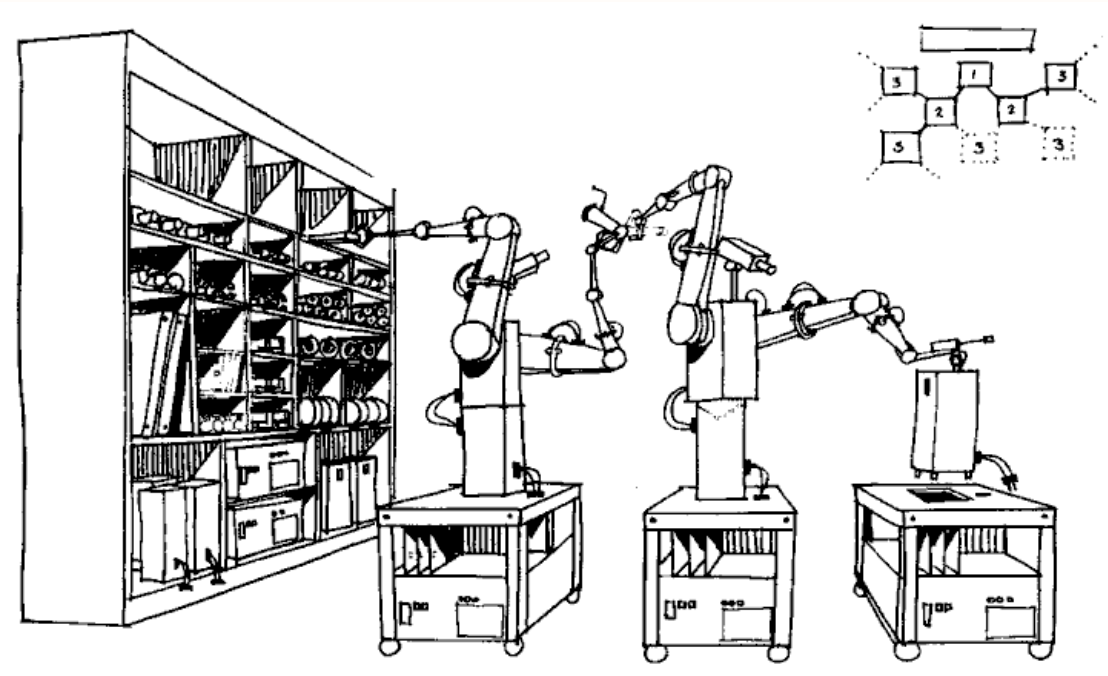
- Physics of the state of matter called “proteins”
  (evolvable, mapped from DNA space, glassy dynamics) .

THANKS

Yonatan Savir

more: www.weizmann.ac.il\complex\tlusty

Self-printing?



Proposed demonstration of simple robot self-replication,

from *advanced automation for space missions*, NASA conference 1980.