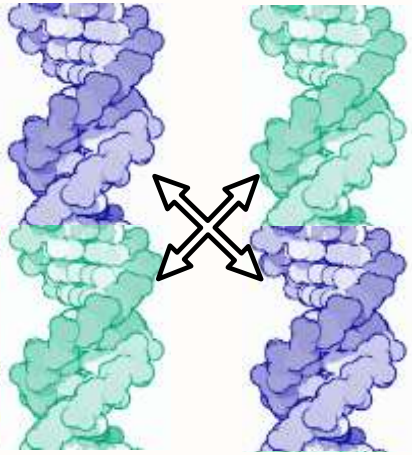
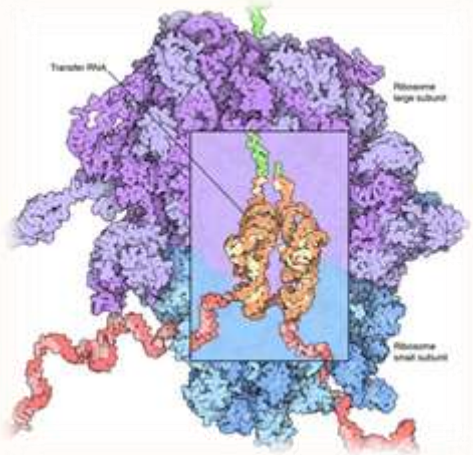
# SEQUENCE RECOGNITION BY CONFORMATIONAL CHANGES:

# HOMOLOGOUS RECOMBINATION AND THE RIBOSOME



YONI SAVIR AND T.T.

WEIZMANN INSTITUTE

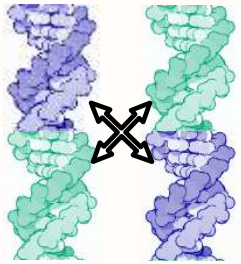


**Conformational proofreading,** a general mechanism of molecular recognition: *Structural mismatch* or *energy barrier* between molecular recognizer and target enhance quality and specificity of detection.

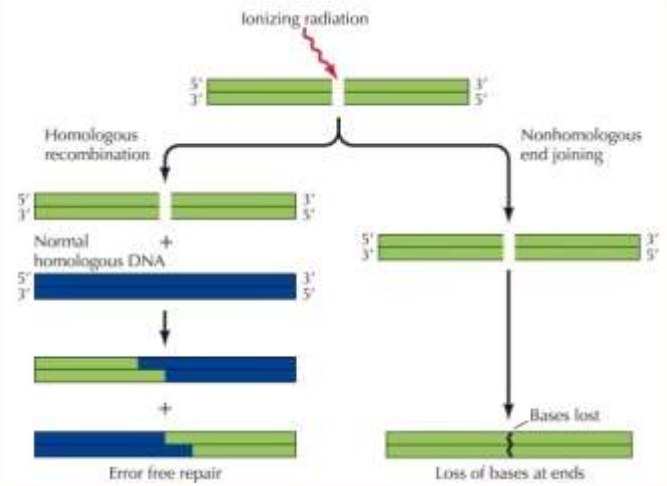
(Savir & TT, *PLoS1* 2007, *IEEE J 20*08, *Molecular Cell 20*10)

# Recombination machinery recognizes homologous DNA and avoids false targets

- Exchange between two homologous DNAs**.**



D Goodsell




Essential roles:

Genetic **diversity** (crossover, sex).

← Genome **integrity** (repair machinery).

- **Task**: Detect correct, homologous DNA targets

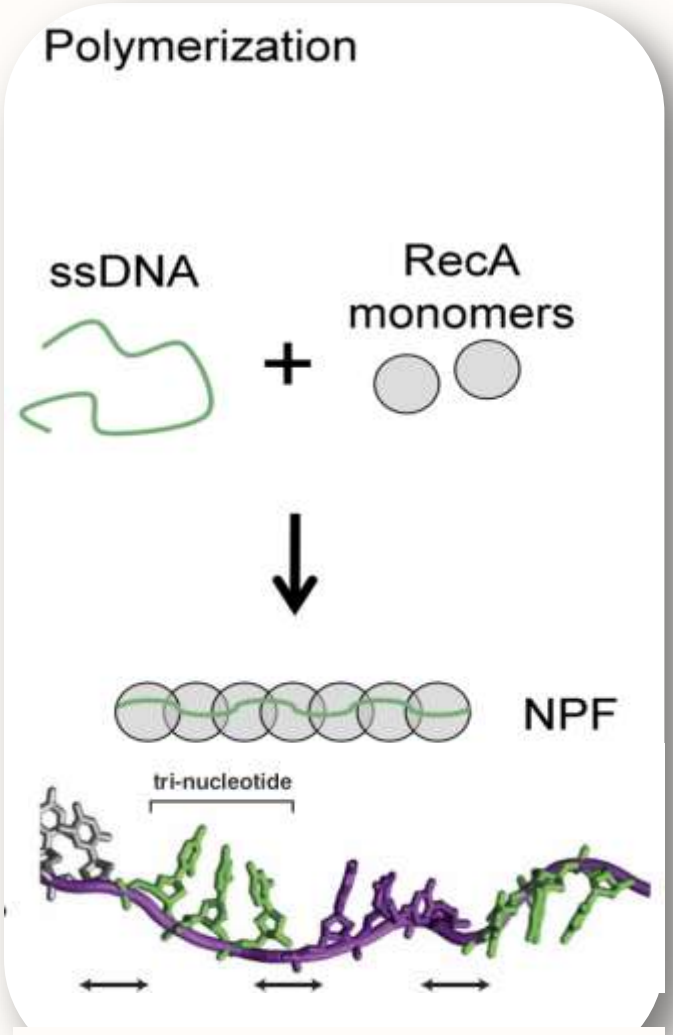
  among incorrect lookalikes in a noisy environment.

(Savir &TT, *Molecular Cell* 2010)

# Homologous recombination is a multistage process

- HR mediated by **RecA** (or members of superfamily)

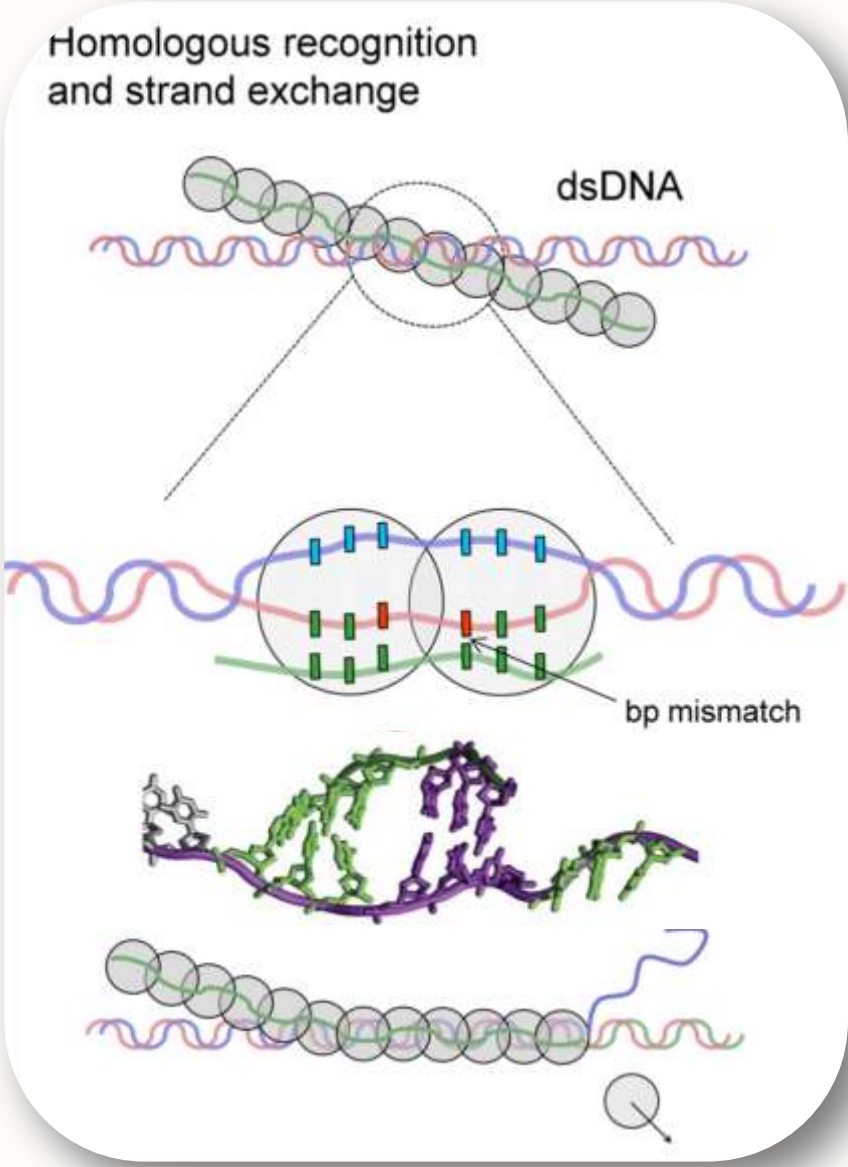
- **Non-uniform** ssDNA deformation :
  - Each triplet ~ B-DNA structure ~ 4 Å
  - Axial rise between triplets ~ **8 Å** .




PDB (Chen,Yang,Pavletich, Nature 2008)

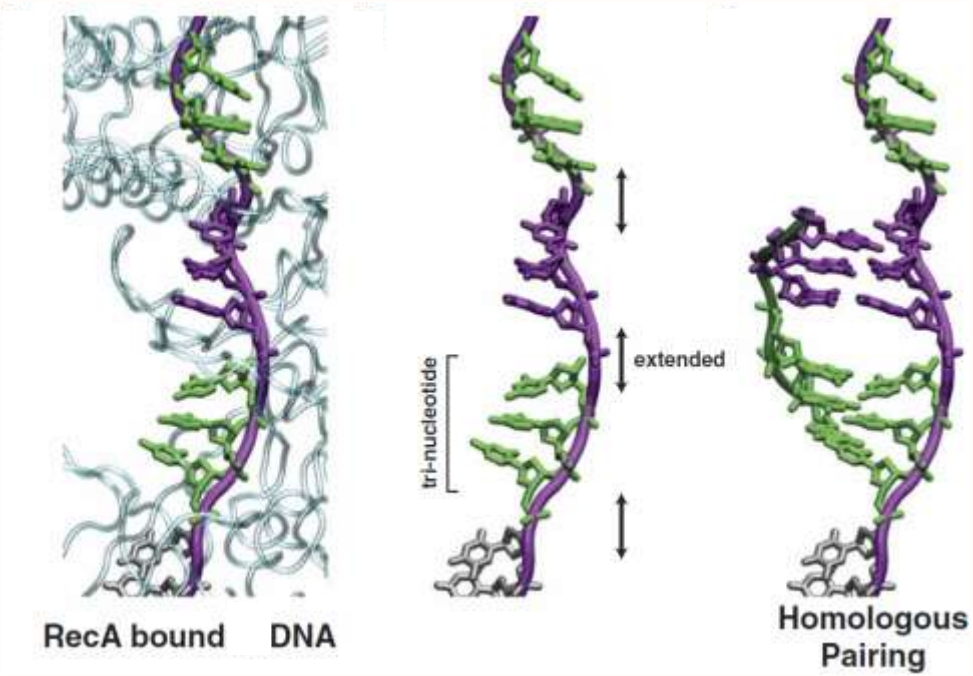
# Homology check requires deformation of dsDNA

- Homology check involves **steps** of a few bp ~3.
- dsDNA **deforms** to allow recognition via Watson-Crick pairing.
- Once bases are aligned, ‘decision’ is made: proceed or abort recombination.
- ATP hydrolysis is *not* required for check.
- Global pairing problem → Avi Minsky’s talk.



# DNA is extended by a factor of two – a seeming paradox because of the energy cost

- **RecA-induced** extension is large
  (1.5 on average, up to 2.4 )
- Costs 3-4 $k_BT$/bp.
- Mechanism to detect Right DNA within many similar Wrong targets.

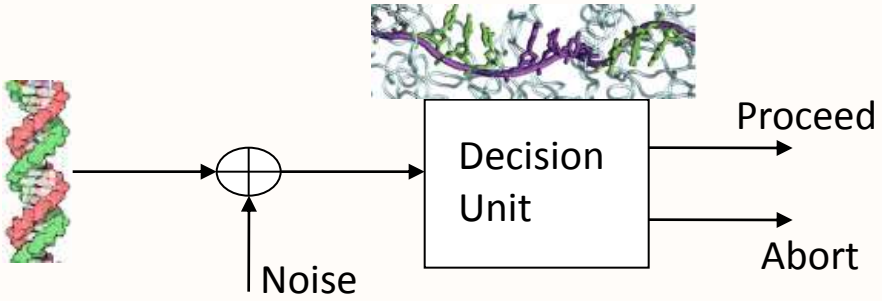



General mechanism: conformational proofreading.

(Savir & TT, Mol Cell 2010)

# Molecular recognition is a decision problem

- Each 3-base step is **decision-making**: 2x2 channel.




| Output \ Input | Homologous | Non-hom. |
|---|---|---|
| Proceed HR | True Positive | False Positive |
| Abort HR | False Negative | True Negative |

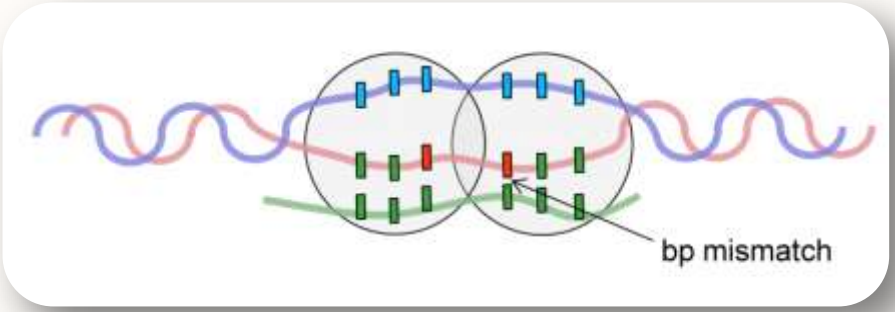
- Each event bears cost/benefit, $f_{\text{event}}$ .
- **"Fitness" :** $F = \sum_{\text{event}} f_{\text{event}} \times P_{\text{event}}.$ (**Distortion** in info theory)
- ***F*** depends on **structural** parameters and can be optimized.

(Savir &TT, Plos 1 2007, Mol Cell 2010)

# Binding probability depends on extension and mismatches

- Binding probability

$$p = \frac{1}{1 + e^{-N \cdot G_t + m \cdot G_{mis}}}$$




$N$: Step size

$m$: mismatches

$G_t$: total free energy/bp

$G_{mis}$: loss by mismatch

- The total binding energy depends on binding gain and extension cost :

$$G_t = G_b - G_{ext}$$

# Optimal binding energy maximizes “fitness”

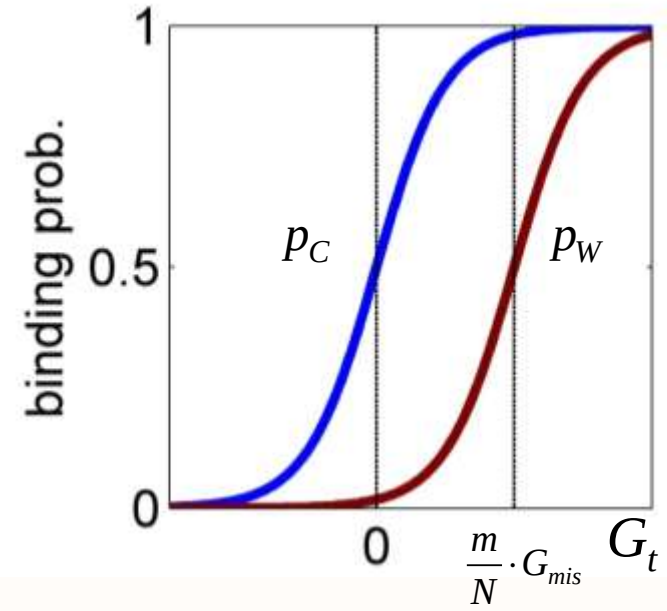
$$F = t \cdot p_C - p_W$$

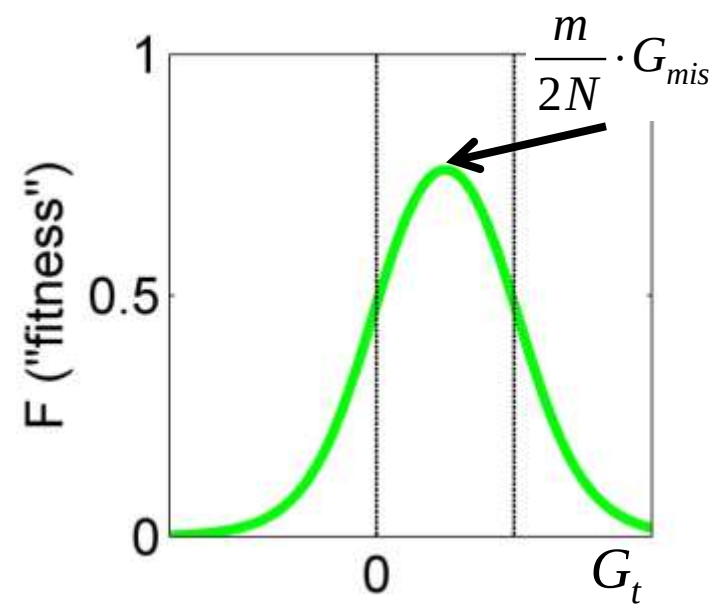
- *t* is the ***tolerance*** of the system: accounts for occurrence, functionality and cost\benefit.
- Reducing $G_b$ decreases binding probability of **both** correct and incorrect sequences .
- reduction of incorrect sequences is **larger**

  → overall detection improves.

$$\overline{G_t} = \frac{m}{2N} \cdot G_{mis}$$

- Corresponds to significant, non-zero deformation.



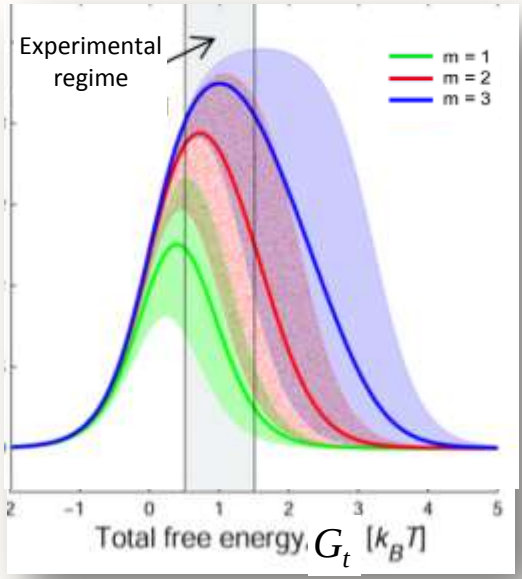



# Observed energies are nearly optimal

- Prediction: energetics optimize recognition.

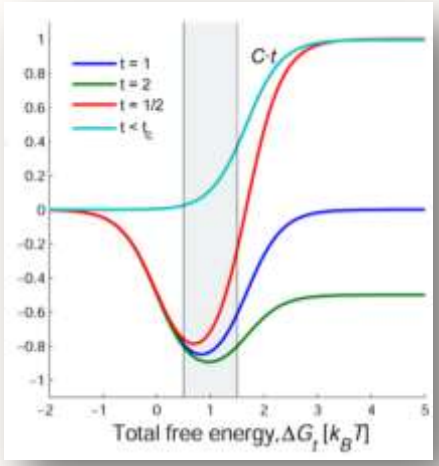
**Experimental energy ~ optimal.**

➔ fits measured $G_t$ ~ 3-5 $k_BT$/triplet.



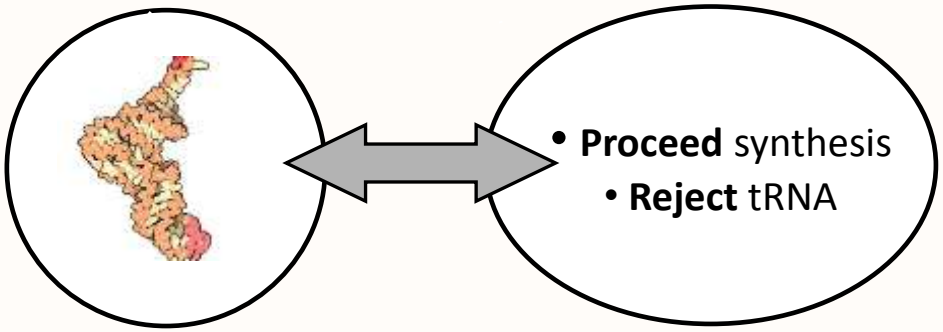

- Optimality is robust to tolerance parameter *t*.

- General mechanism: **Conformational Proofreading**
  - Structural *mismatch or energy barrier* .
  - Reduce Right, *but* reduces Wrong even more.

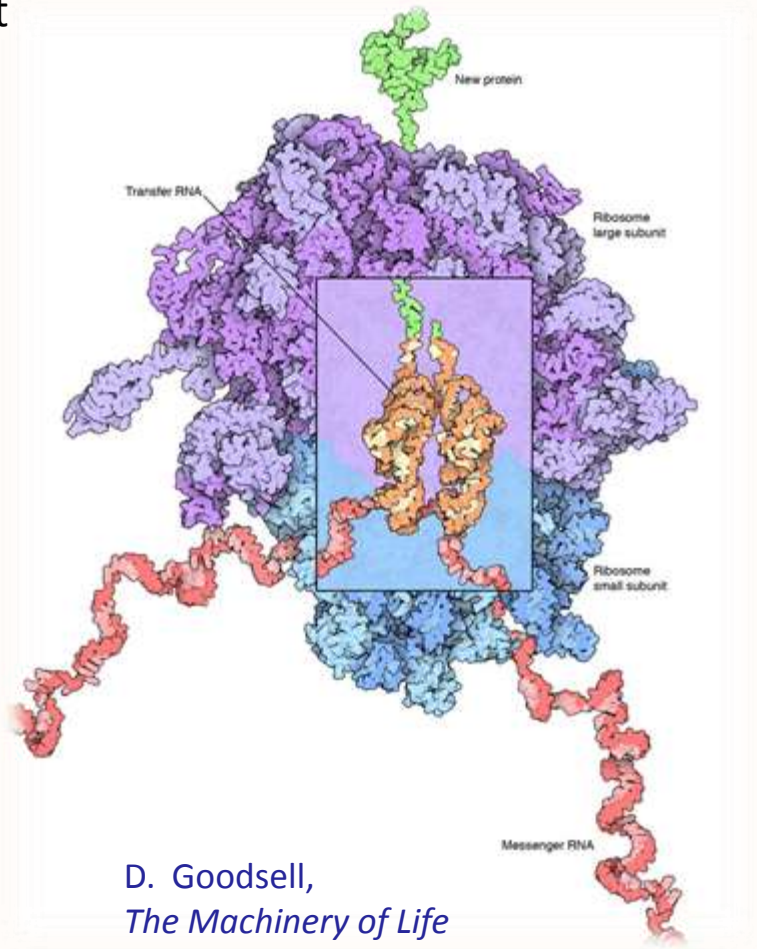


# Ribosomes translate nucleic bases to amino acids

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




- Ribosome needs to recognize and select the correct tRNA among many competitors.




D. Goodsell,
*The Machinery of Life*

How to decode ***fast*** and ***accurately***?

# Ribosomal conformation changes place large energy barriers – that have no compelling explanation

- Significant changes occur in all participating molecules:

  small\large subunits, EF-Tu and tRNA.

tRNA samples codon-anticodon pairing

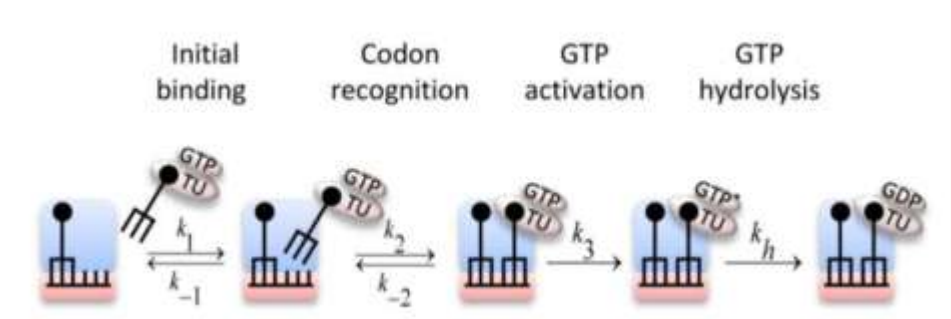
Decoding center probes minor groove

Ramakrishnan's lab

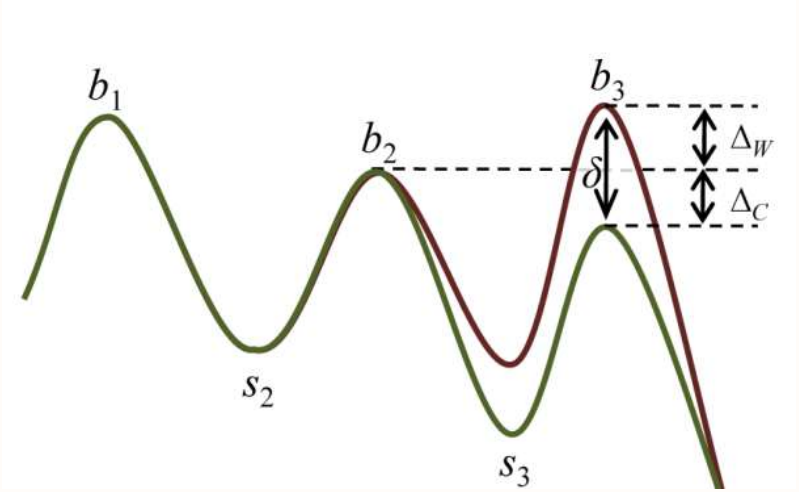

## Do conformational changes enhance tRNA recognition?

# Recognition is determined by energy landscapes of correct and wrong tRNAs

- Theory is constrained by the measured kinetic rates.




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



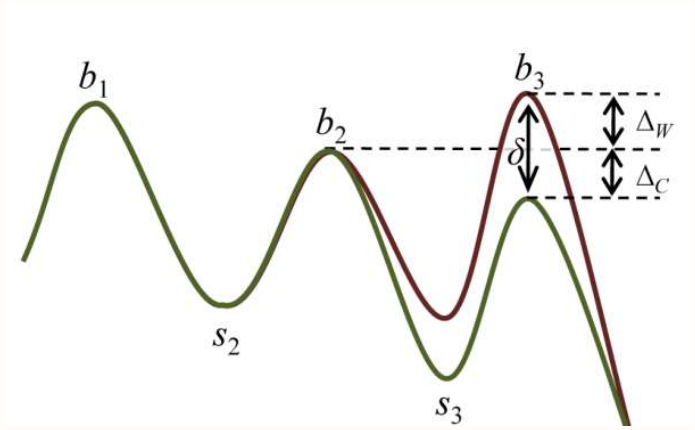

| | Cognate | Non-cognate |
|---|---|---|
| $k_1$ | 100-140 1/(μM·s) | 100-140 1/(μM·s) |
| $k_{-1}$ | 80-100 1/s | 80-100 1/s |
| $k_2$ | 190 1/s | 190 1/s |
| $k_{-2}$ | 0.23 1/s | 100 1/s |
| $k_3$ | 260 1/s | 0.6 1/s |

Correct = UUU
Wrong = CUC

(Rodnina lab)

# Conformations change the total energy barrier, Δ

- Steady-state **decoding rate** (per ribosome) is:

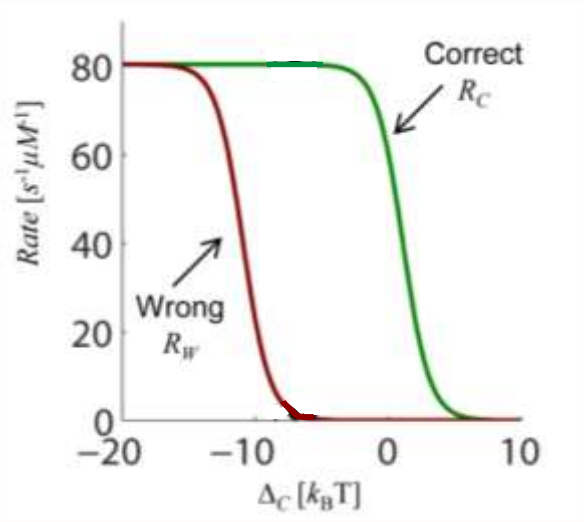



$$R_C \propto \frac{[t_W]}{p^{-1} + e^{\Delta_C}}$$

$$R_W \propto \frac{[t_C]}{p^{-1} + e^{\Delta_C + \delta}}$$

where $p = \frac{k_{-1}}{k_2 + k_{-1}} \approx 0.3, \quad \delta \approx 12\ k_B T$

- In terms of energy landscape, Right and Wrong decoding rates are:

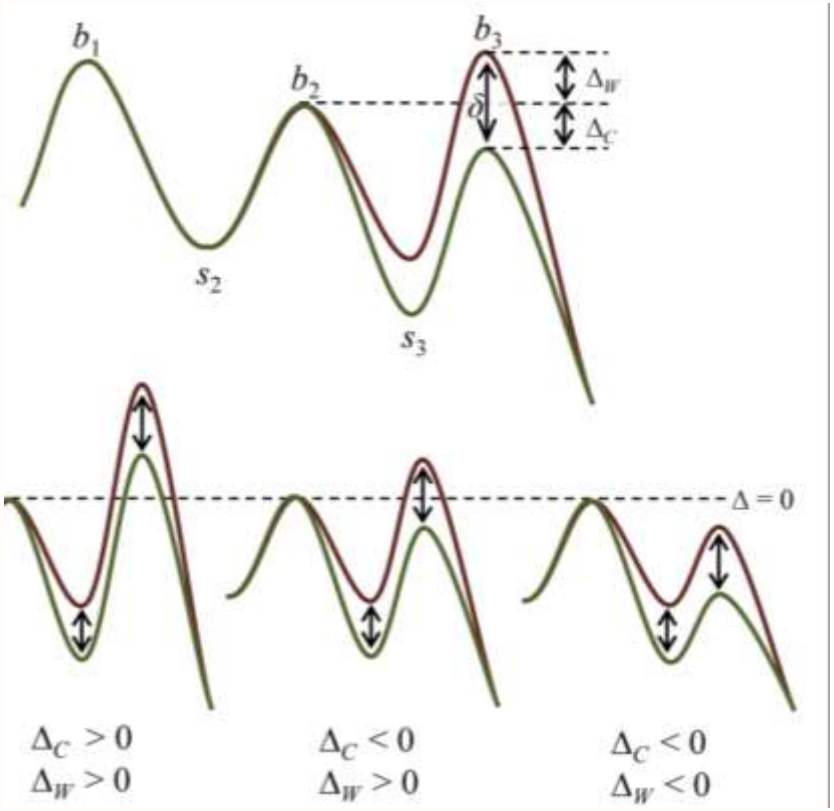


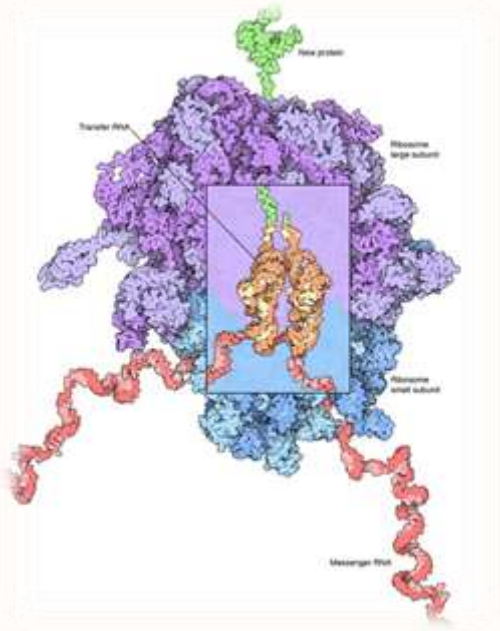
# What is the optimal landscape?

- Given *p* and δ, what is the optimal $\Delta_C$ ?
- Three possible regimes:

$$R_C \propto \frac{1}{p^{-1} + e^{\Delta_C}}$$

$$R_W \propto \frac{1}{p^{-1} + e^{\Delta_C + \delta}}$$



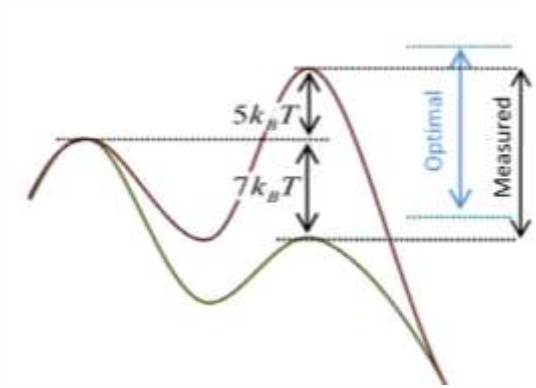
# Ribosome shows an energy barrier Δ which is nearly optimal

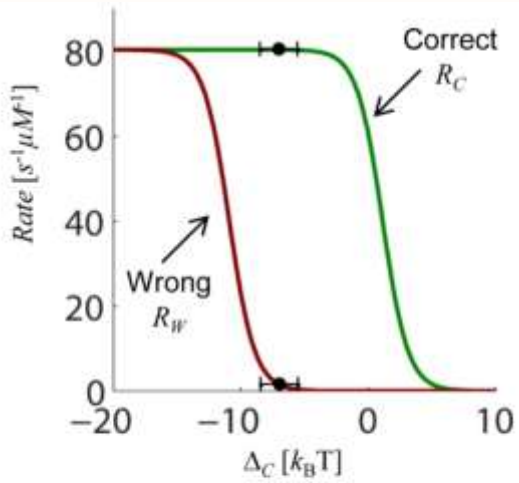
- For example, simple performance measure ("fitness"):

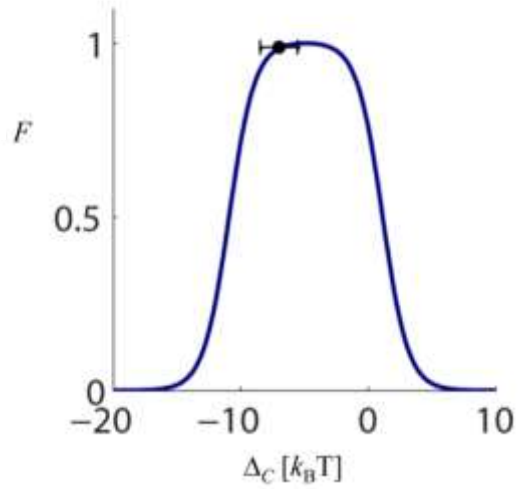
$$F = R_C - R_W = \frac{1}{p^{-1} + e^{\Delta_C}} - \frac{1}{p^{-1} + e^{\Delta_C + \delta}}$$

- Prediction: the optimal regime is the **symmetric**, $-\delta < \Delta_C < 0$.

$$\Delta_C^* = -\frac{1}{2}\delta - \ln p.$$



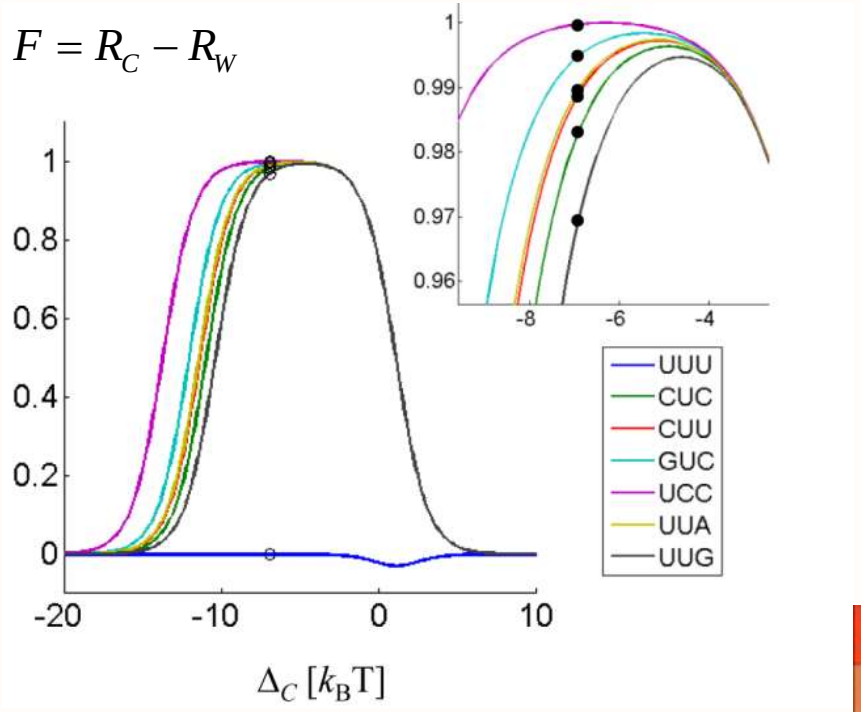



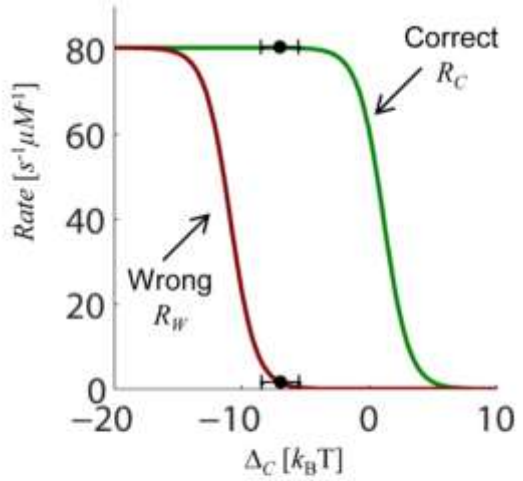



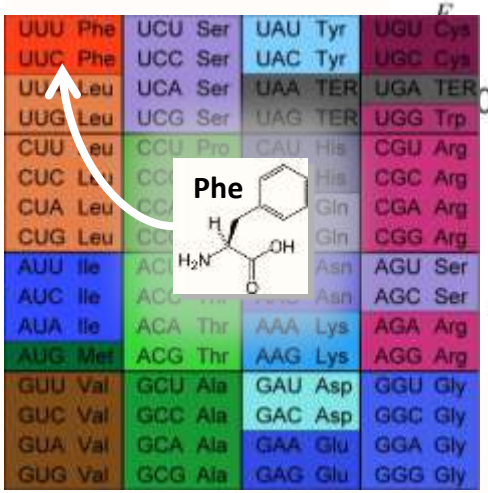

- **The ribosome is nearly optimal**

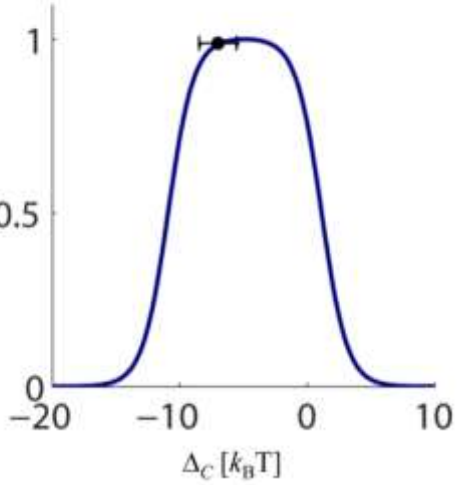
(according to info theory prediction).

# Decoding is optimal for all six measured tRNAs




- Except for UUC, the other codon of amino-acid phenylalanine (“wobble”).

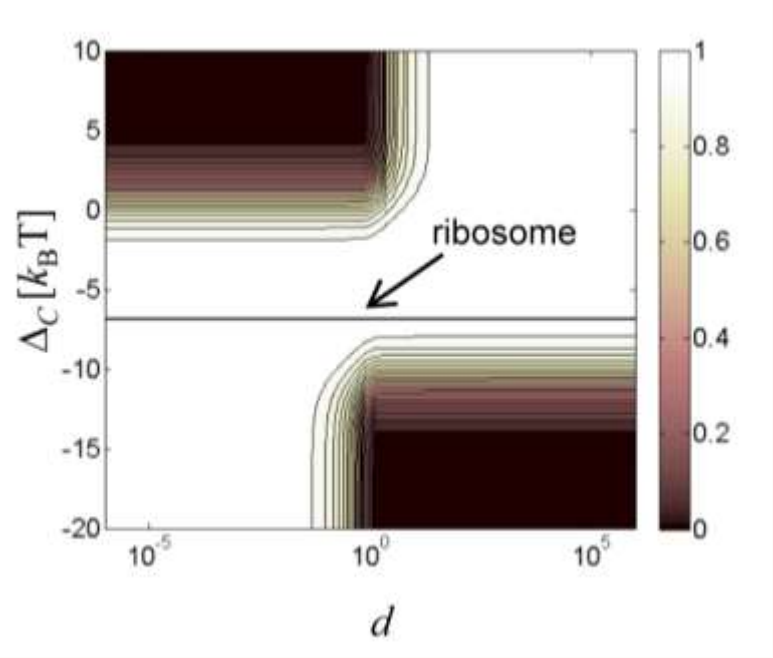
# Optimality is valid for wide range of fitness functions

$$F = R_C - d \cdot R_W$$

- Ribosome is optimal in wide region:

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

- Ribosome parameters achieve 95% of maximal normalized fitness.
- *Worst case scenario* (“max-min” design).




- **General feature**: any fitness function

$F(R_C, R_W)$ with $\partial F / \partial R_C > 0, \partial F / \partial R_W < 0$

exhibits optimum .

# Theory predicts optimal regime of ribosomes of other organisms

- Given δ, optimal Δ resides in a triangle around

$$\Delta^* \approx -\delta / 2$$

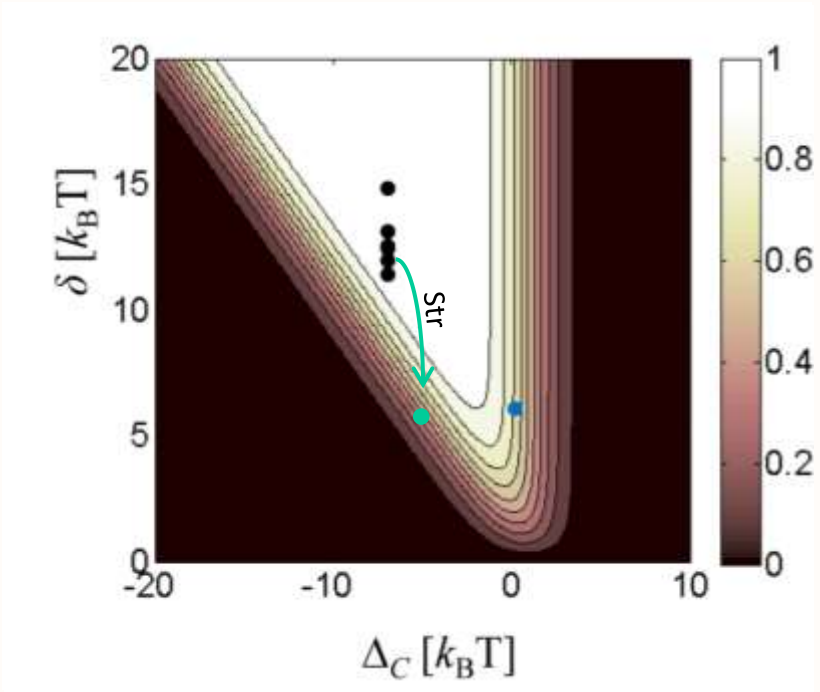
- Kinetic and FRET indicate:

  E. coli is in the triangle.

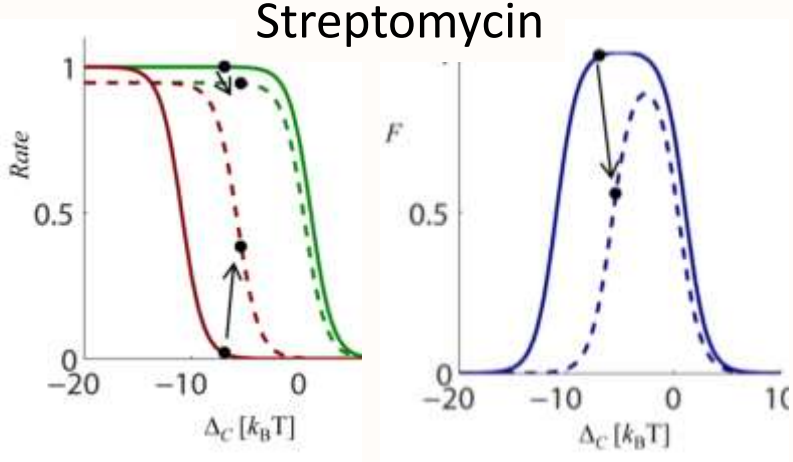
- **Hypothesis:**

  evolution drove ribosomes into triangle.

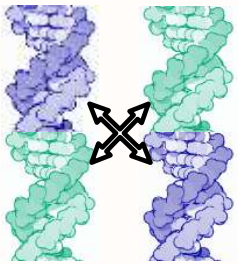
- **Tests:**
  - ribosomes of other organisms.
  - Mutations\antibiotics will be pushed to edges (even if faster\more accurate).



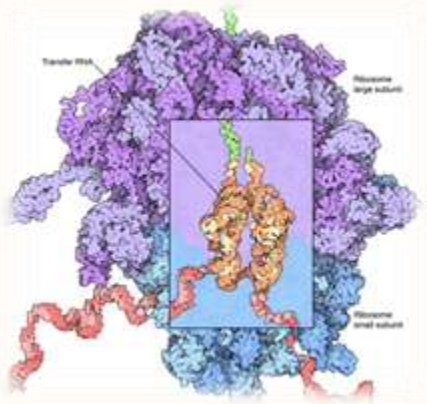

Streptomycin



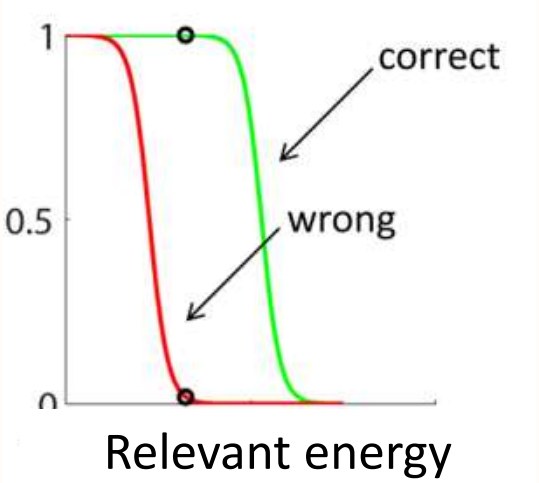
# Energy barriers for optimal recognition may be a general design principle of recognition systems with competition

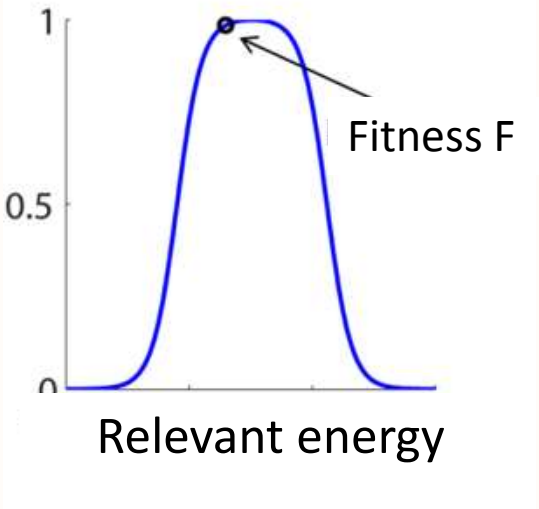


Recombination optimizes extension energy of dsDNA.



Ribosome optimizes energy barriers of decoding




- Design principle follows from optimization of information transfer function:
- **Conformational proofreading.**

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

# Evaluating the involved energies

- Total binding energy

  $G_t \approx 0.5 – 1.5\ k_BT$

- Destabilization free energy:

  $G_{mis} \approx 1.5\text{-}3.5\ k_BT$