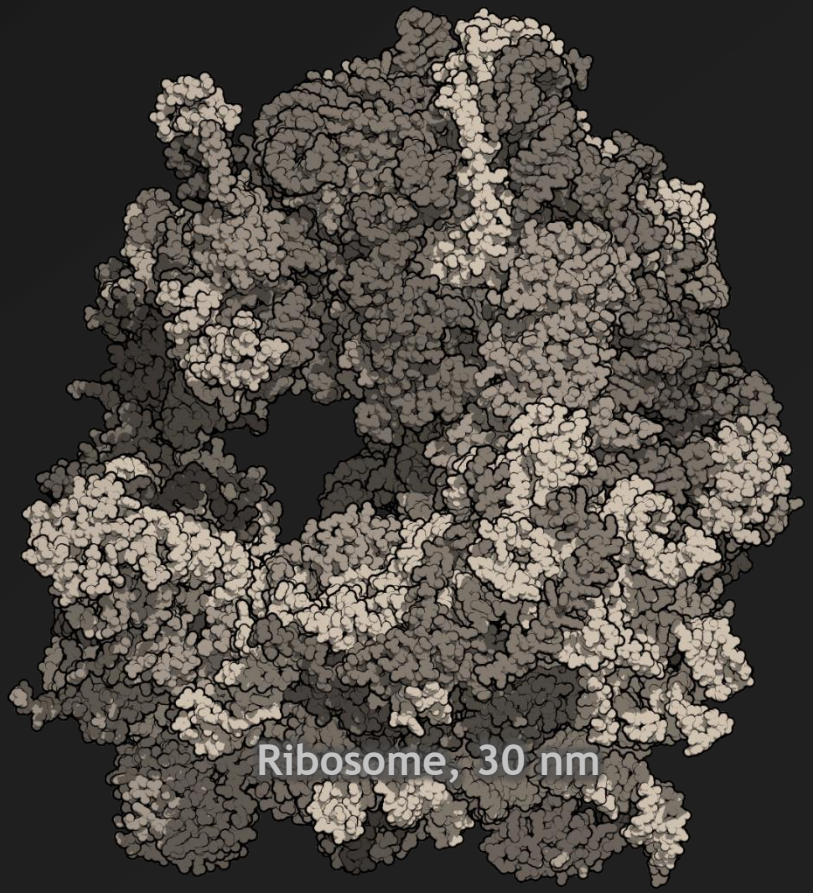
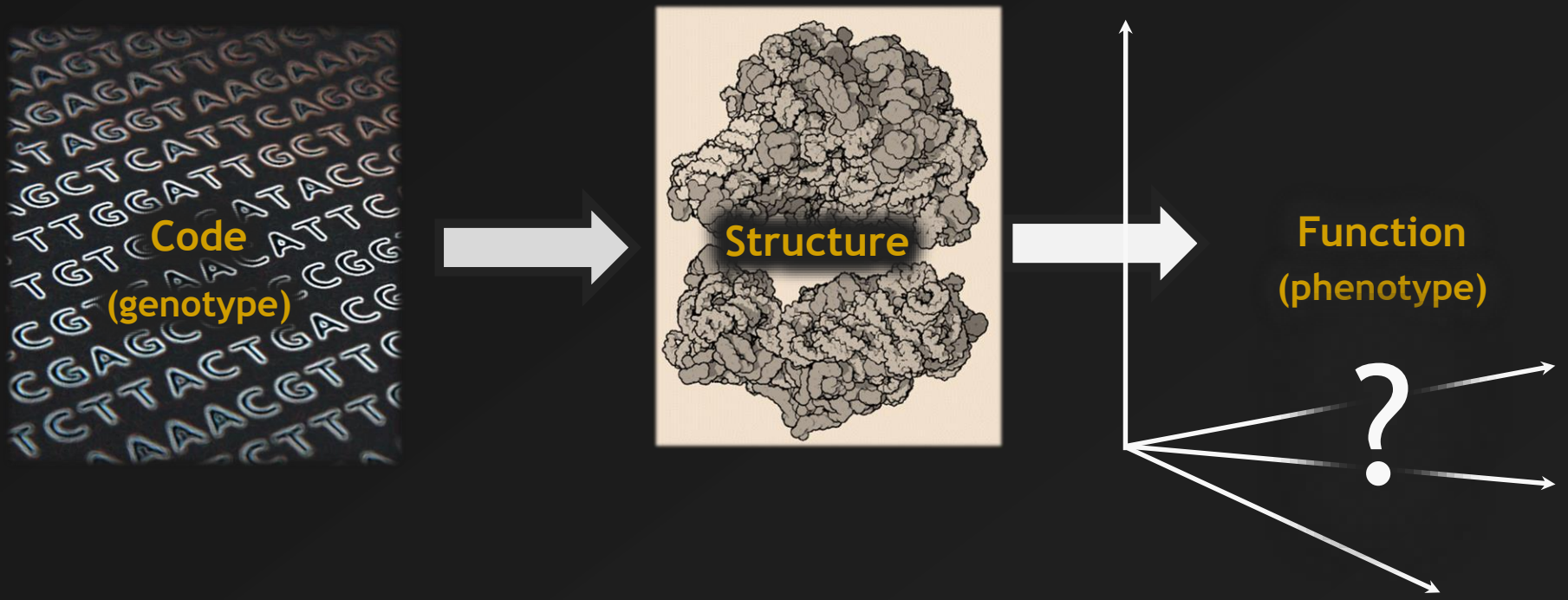
What is the dimension of molecular recognition?



Yoni Savir & TT

WAM seminar, Harvard 2015



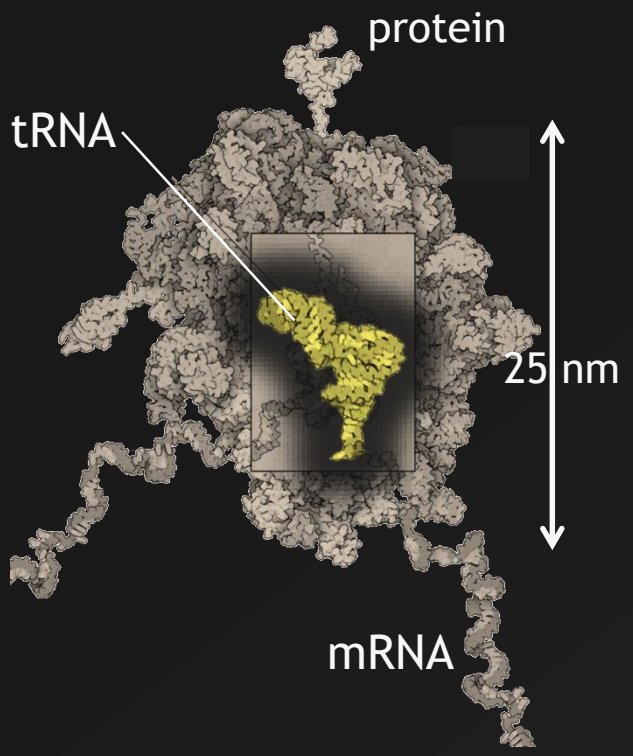
What are the relevant degrees-of-freedom?

Outline: are molecular phenotype spaces “flat”?

$$\dim_{\text{geno}} \gg \dim_{\text{pheno}}$$

- Molecular recognition: “fitness”? conformational changes?
- Generic principles: conformational tuning (induced fit).
- Physical basis? …

Ribosomes are engines of self-reproduction whose performance is critical to fitness



~ 10^4 bases

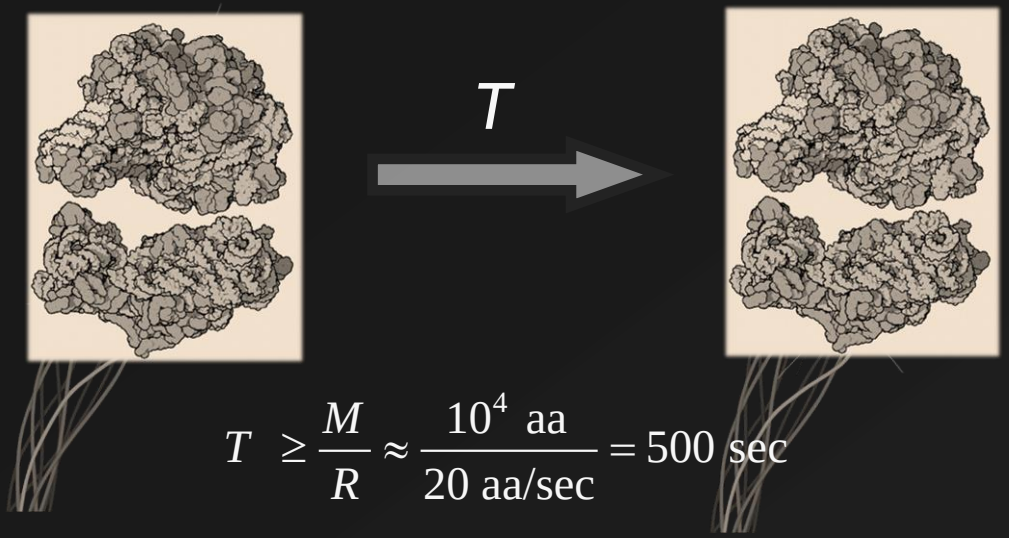
~ 10^4 aa

~ $3 \cdot 10^6$ Da

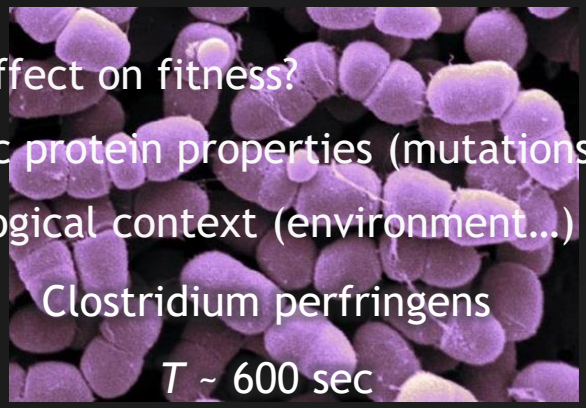
- Fast + accurate + small molecular decoder ?
- Relevant degrees of freedom ?

(Figure: D. Goodsell, PDB)

Ribosome limits self-replication rate

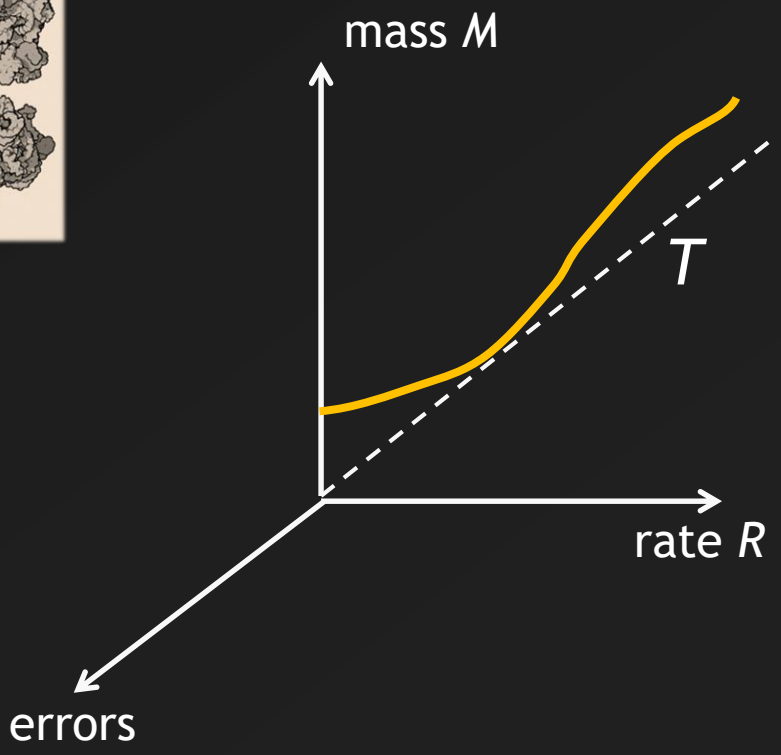


$$T \geq \frac{M}{R} \approx \frac{10^4 \text{ aa}}{20 \text{ aa/sec}} = 500 \text{ sec}$$

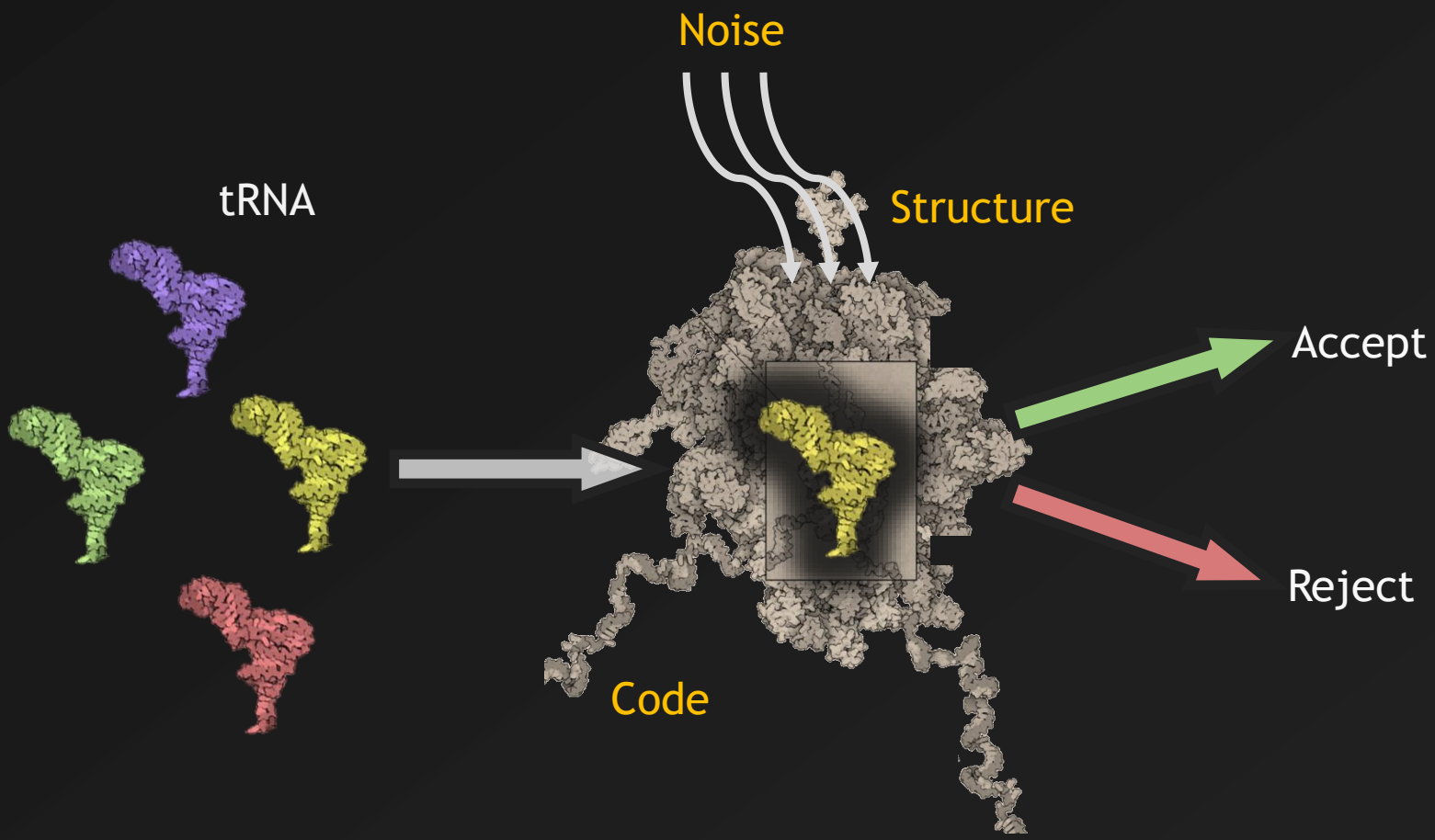


Errors effect on fitness?

- Basic protein properties (mutations…)
- Biological context (environment…)



Decoding at the ribosome is a molecular decision problem



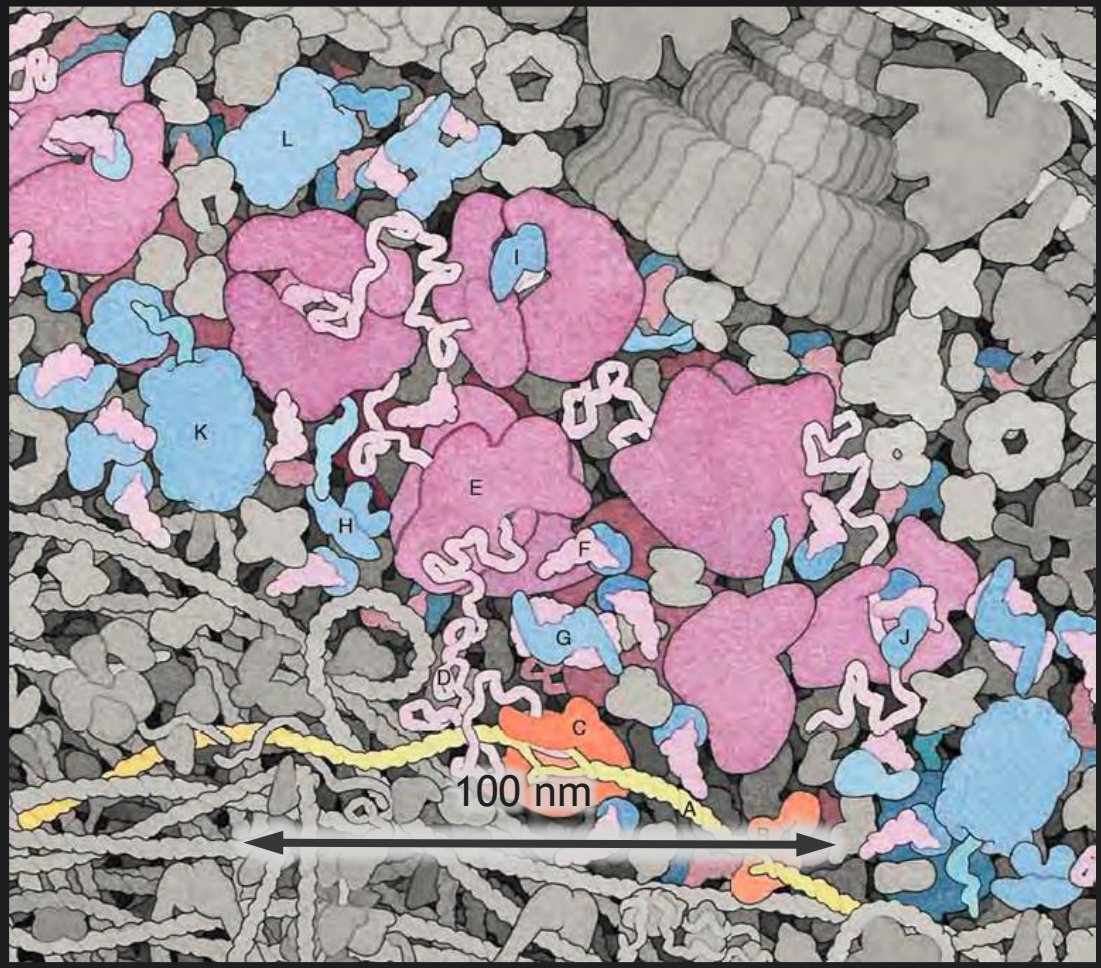
- How to evolve molecules that recognize in a noisy environment?

Savir et. al, *Cell* 2013

Recognizing molecular targets is a formidable task

- tRNA ~ 10^6
- Ribosomes ~ 10^5
- tRNA species = 41
- D_{tRNA} ~ 10^2 μm^2/sec
- tRNA-ribosome ~ 10^4/sec

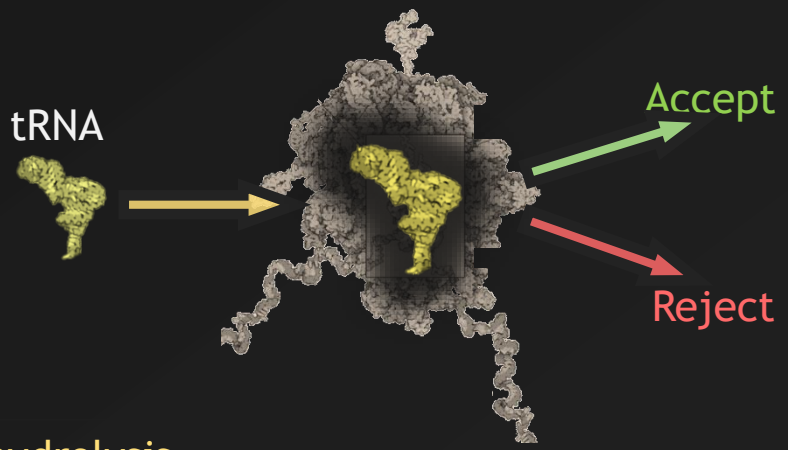
- Only ~1/10 tRNAs is correct.



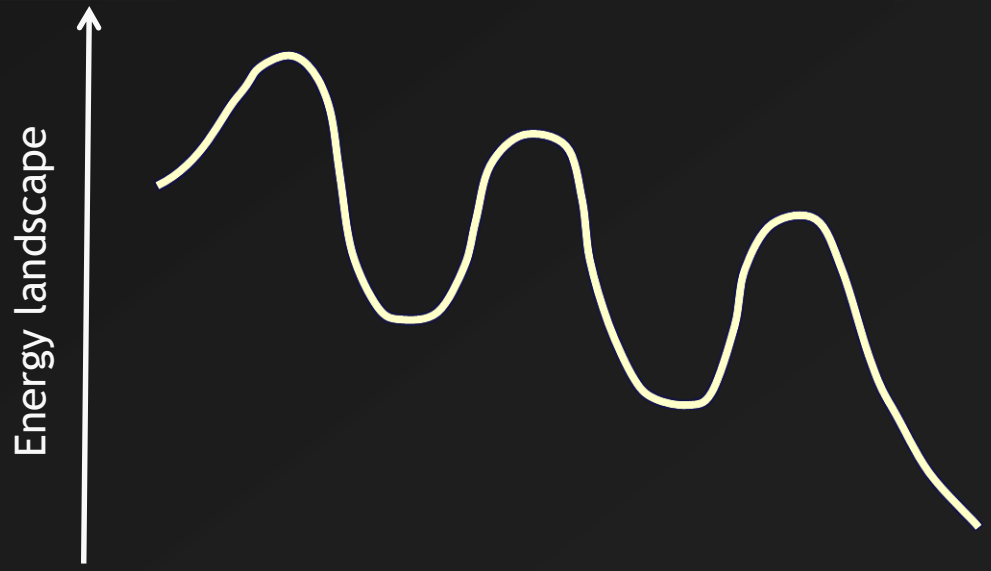
(Figure: D. Goodsell, PDB)

Decoding governed by energy landscapes of tRNA recognition

- Measured transition sequence.
- Large conformational changes.



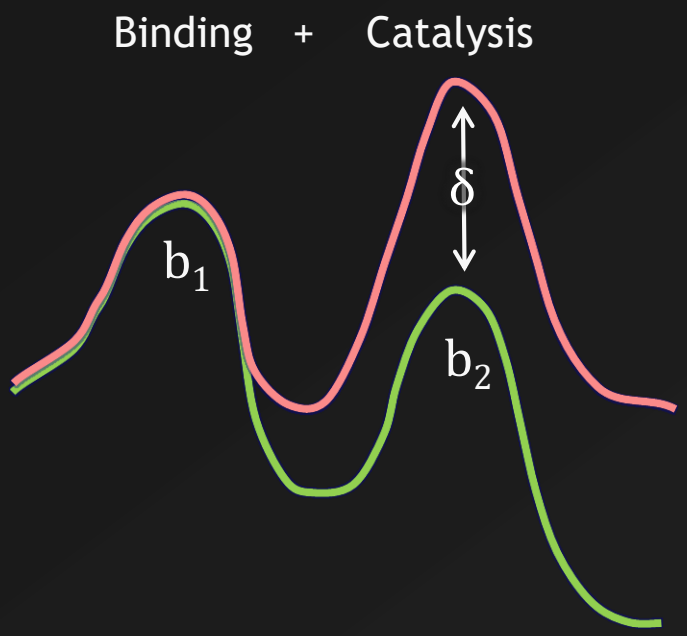
Reject ← Initial binding ↔ Specific binding ↔ GTP activation → hydrolysis → Accept



$$\text{Rates} \sim e^{-\Delta G/k_B T}$$ (Arrhenius)

Ribosome decoding problem is an example of recognition in the presence of competition

- Generic enzyme kinetics (MM):



$$R_B \sim \frac{1}{e^{b_1} + e^{b_2+\delta}}$$

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

- The enzyme dilemma: rate R_G vs. specificity R_G/R_B ?

Given the phenomenology of tRNA decoding:

- Recognition performance (“fitness”) ?
- Effective dimension?
- Role of conformational changes?

Recognition fitness has generic features

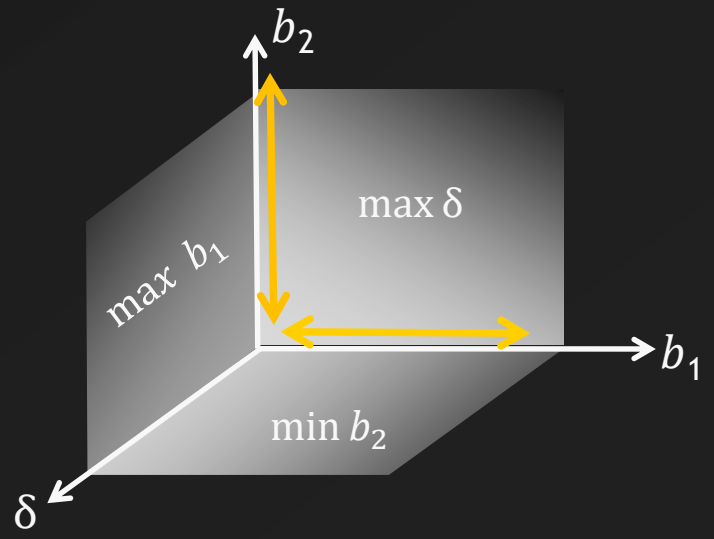
- "Fitness" $F(R_G, R_B)$ is context-dependent.
- "Biologically reasonable": $\frac{\partial F}{\partial R_G} \geq 0, \quad \frac{\partial F}{\partial R_B} \leq 0.$

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

$$R_B \sim \frac{1}{e^{b_1} + e^{b_2+\delta}}$$

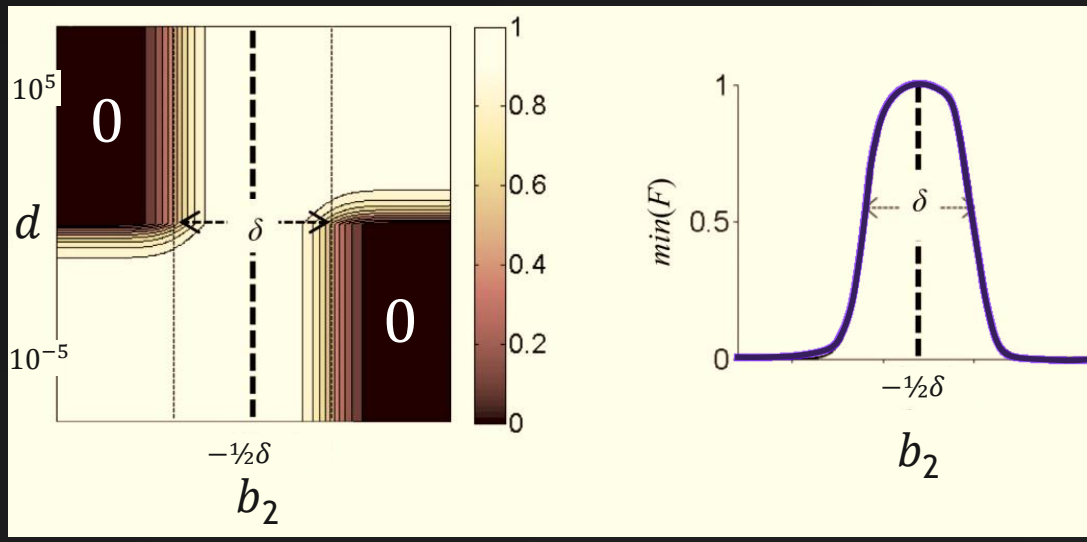
- Optima in landscape (b_1, b_2, δ) space:
  - $\delta \rightarrow$ limit. $\quad \frac{\partial F}{\partial \delta} \geq 0$
  - Optimization is effectively 1D.

$$\left\{\frac{\partial F}{\partial B} = 0 \;\&\; \frac{\partial F}{\partial b_3} \geq 0\right\} \text{ or } \left\{\frac{\partial F}{\partial B} \leq 0 \;\&\; \frac{\partial F}{\partial b_3} = 0\right\}$$



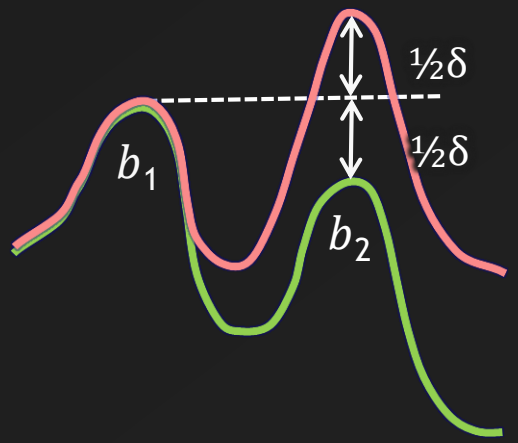
Adaptation to worst case yields symmetric design

- Example: $F = R_G - d \cdot R_B$.
- "Worst case scenario" $\max(\min(F))$.



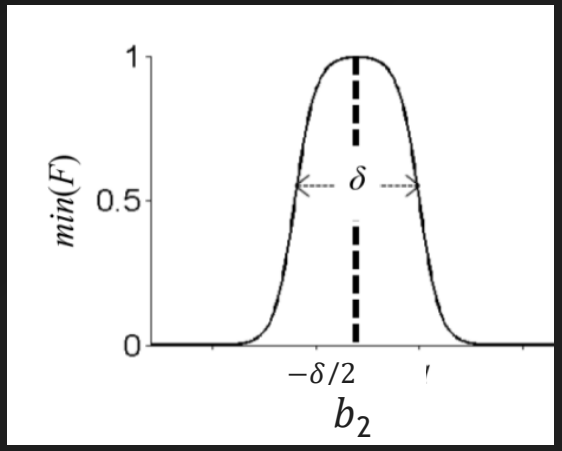
- Max-Min solution is symmetric:

$$\Delta = b_2 - b_1 \approx -½\delta$$



Solution is valid for any “reasonable” fitness function

$F = R_G - d \cdot R_B$



- Ribosome optimal in wide region:

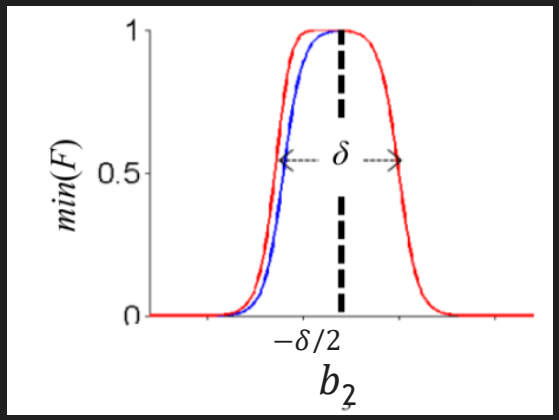
$$10^{-5} = e^{-\delta} \le d \le e^{\delta} = 10^{5}.$$

- Any $F(R_G, R_B)$ *exhibits maximum*

as long as both rates are “relevant”.

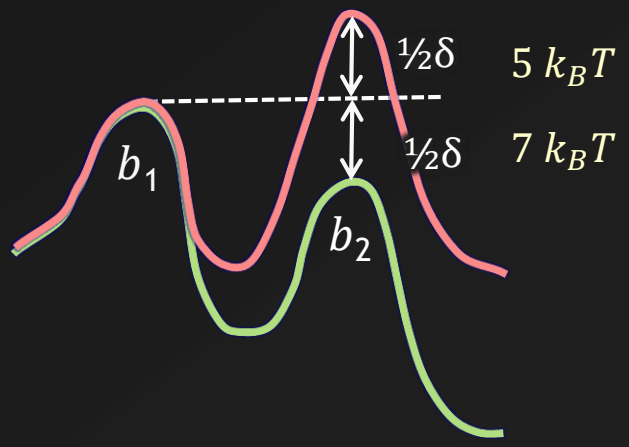
$$e^{-\delta} < \left| \frac{\partial F}{\partial R_B} \Big/ \frac{\partial F}{\partial R_G} \right| < e^{\delta}$$

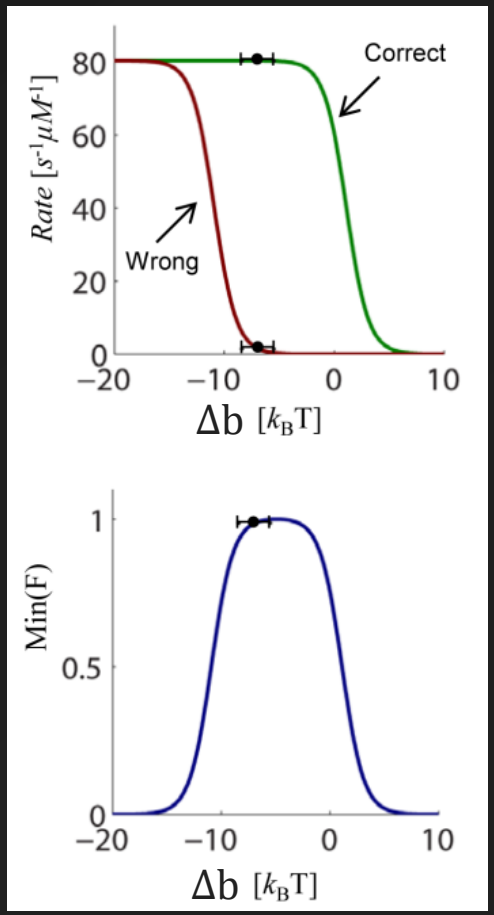
$R_G - d \cdot R_B^{\,2}$ $R_C - d \cdot (R_B/R_G)$



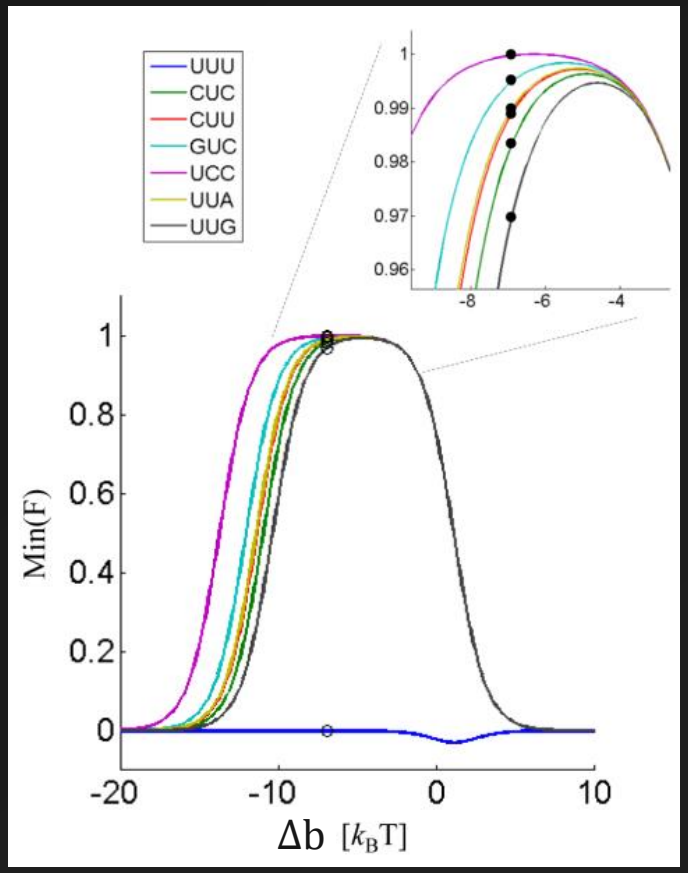
Ribosome energy barriers appear adapted

- Predicted symmetry: $\Delta b = b_2 - b_1 \approx -½\delta$.
- Measurements: $\delta \approx 12 k_B T$.



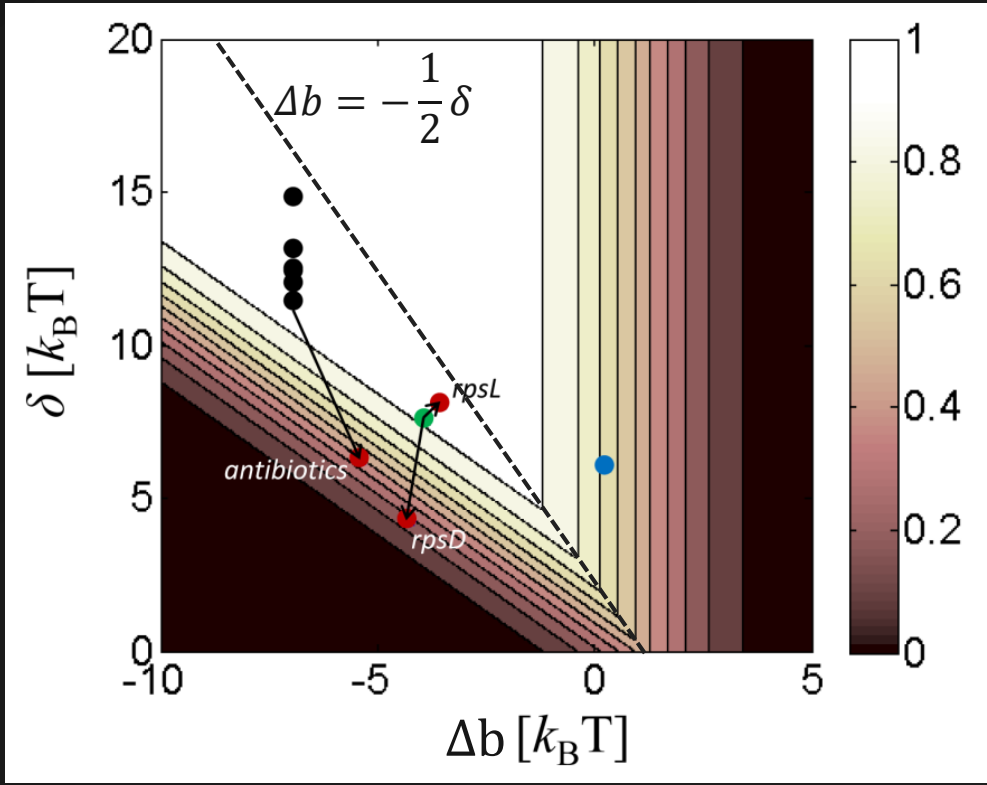


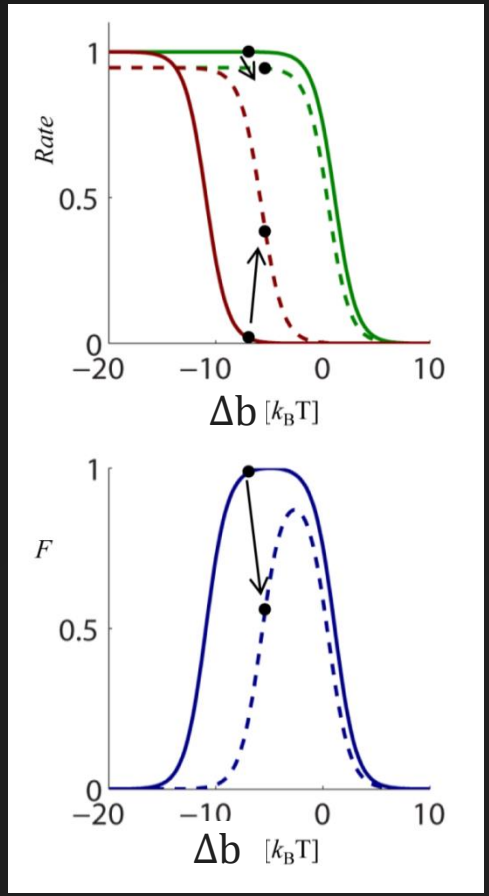
Decoding is adapted for all measured tRNAs



- Except for UUC: $\{\mathrm{UUC}, \mathrm{UUU}\} \rightarrow$ phenylalanine.

Predicted adaptation regime of other ribosomes





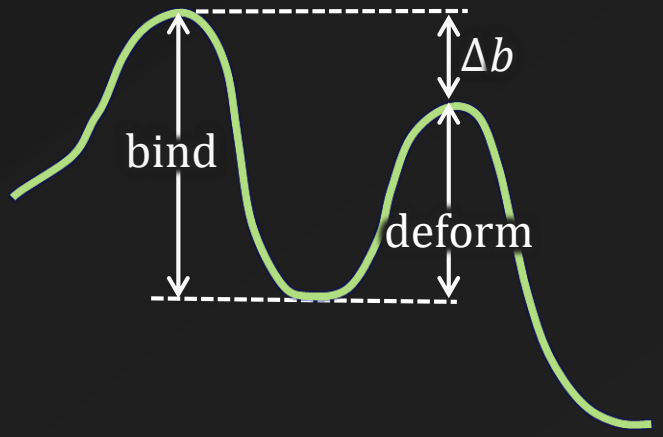
- Adaptation regime in ribosome phenotype space, $-\delta \leq \Delta b \leq 0$.
- Mutations\antibiotics reduce adaptation.

Savir et. al, *Cell* 2013

What is the role of conformational changes?

- Energy barrier $\Delta b = \text{deform} - \text{bind} = -\frac{1}{2}\delta \approx -6\ k_B T.$

- For $\text{bind} \geq 6\ k_B T,\ \ \text{deform} \geq 0.$



- Conformational proofreading: deformation improves tRNA recognition.

- Energy barrier that selects right targets among competitors.

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

extension energy of dsDNA.

[Dekker et al., *Mol Cell*, 2012]

energy barriers of decoding

- Conformational proofreading:

  optimization of information transfer function.

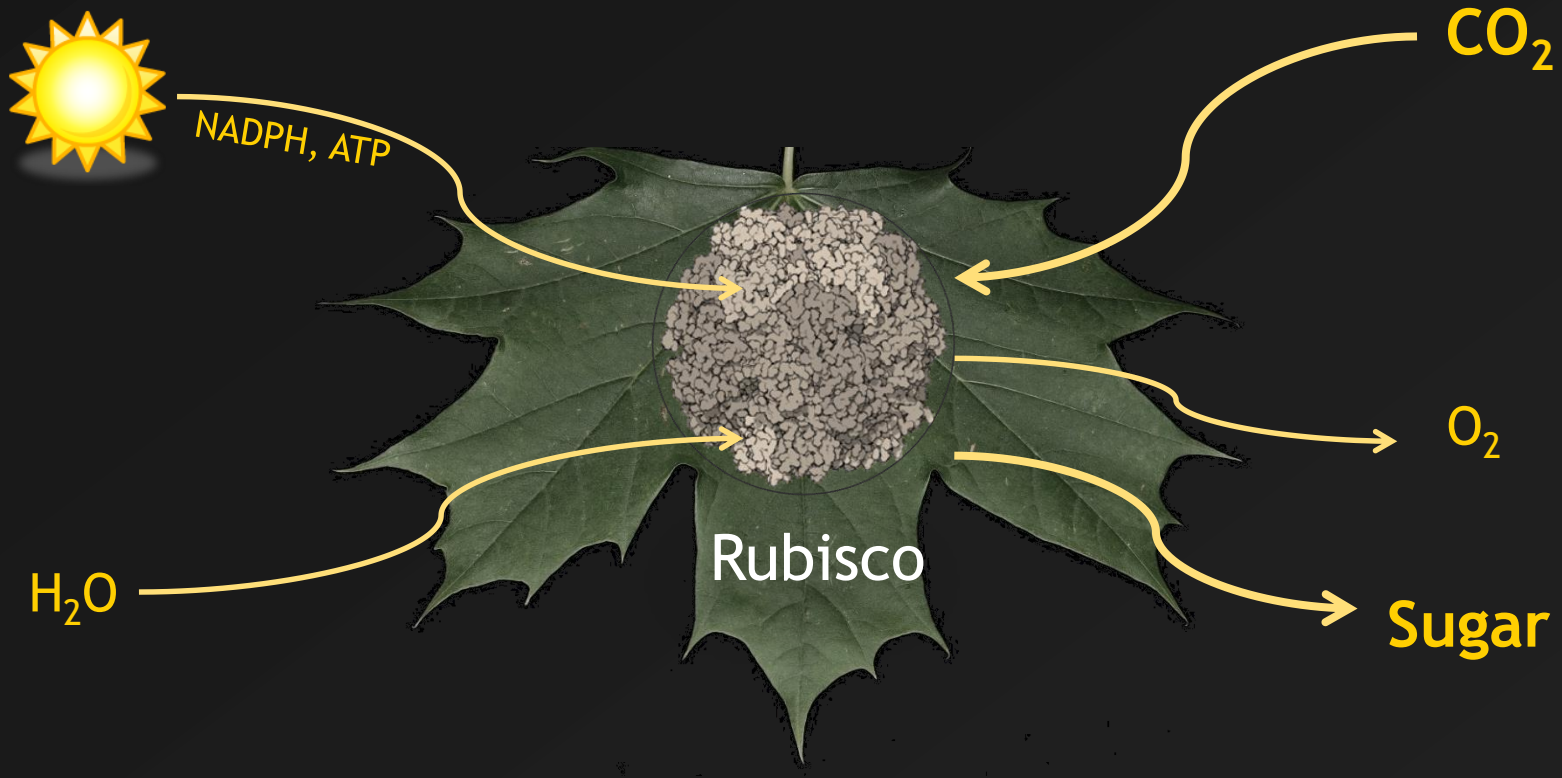
- Explains induced fit (Koshland 1958)?

- Applies to any enzymatic kinetics with competition...

Savir et al., *Mol Cell*, 2010

- Generic properties of adaptation in function space.
- Why is coarse-grained transition-state description effective?

Rubisco captures atmospheric CO_2 to make sugar



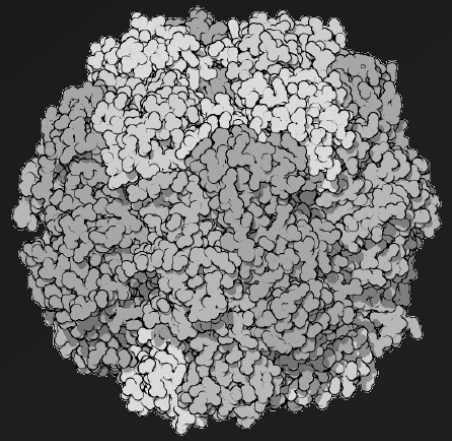
- Photosynthesis fixes C into sugar.
- Rubisco captures CO_2.
- 8 large + 8 small subunits = 540 kDa.

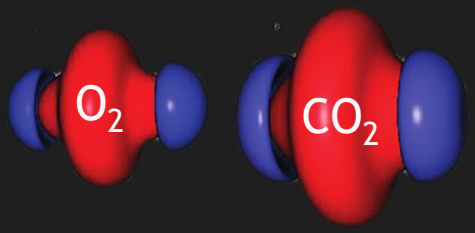
(Savir et al., PNAS 2010)

Rubisco is thought to be a slow and confused enzyme

- #1 protein: most C fixation.

- **Slow:** ~10 CO_2 /sec.
- **Low specificity:** O=C=O vs. O=O.





- Can Rubisco be improved ?
- Already adapted by evolution?
- Inefficiency due to biochemical constraints?

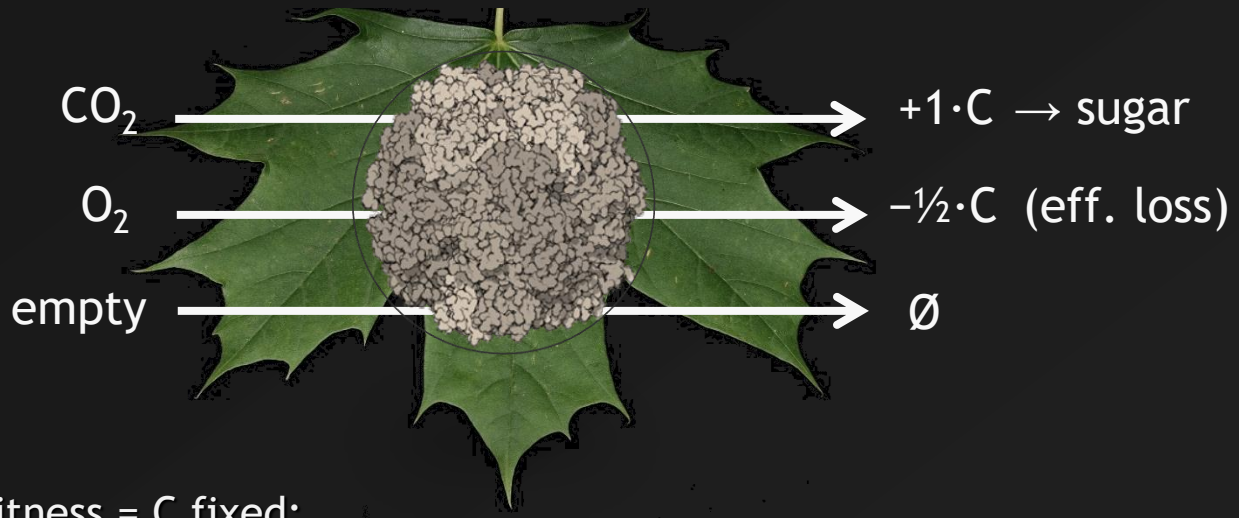
- Is Rubisco **optimal**, **constrained** or both?

(Rubisco figure: D. Goodsell, PDB)

Rubisco needs to recognize CO_2 above a high background of O_2

- 2 affinities to CO_2 and O_2 : $\{K_C \; K_O\}$
  + 2 catalysis rates : $\{\nu_C \; \nu_O\}$

Atmosphere:
20% O_2
0.04% CO_2



- Fitness = C fixed:

$$F = \nu_C P_C - \tfrac{1}{2}\nu_O P_O + 0 \cdot P_{empty}$$

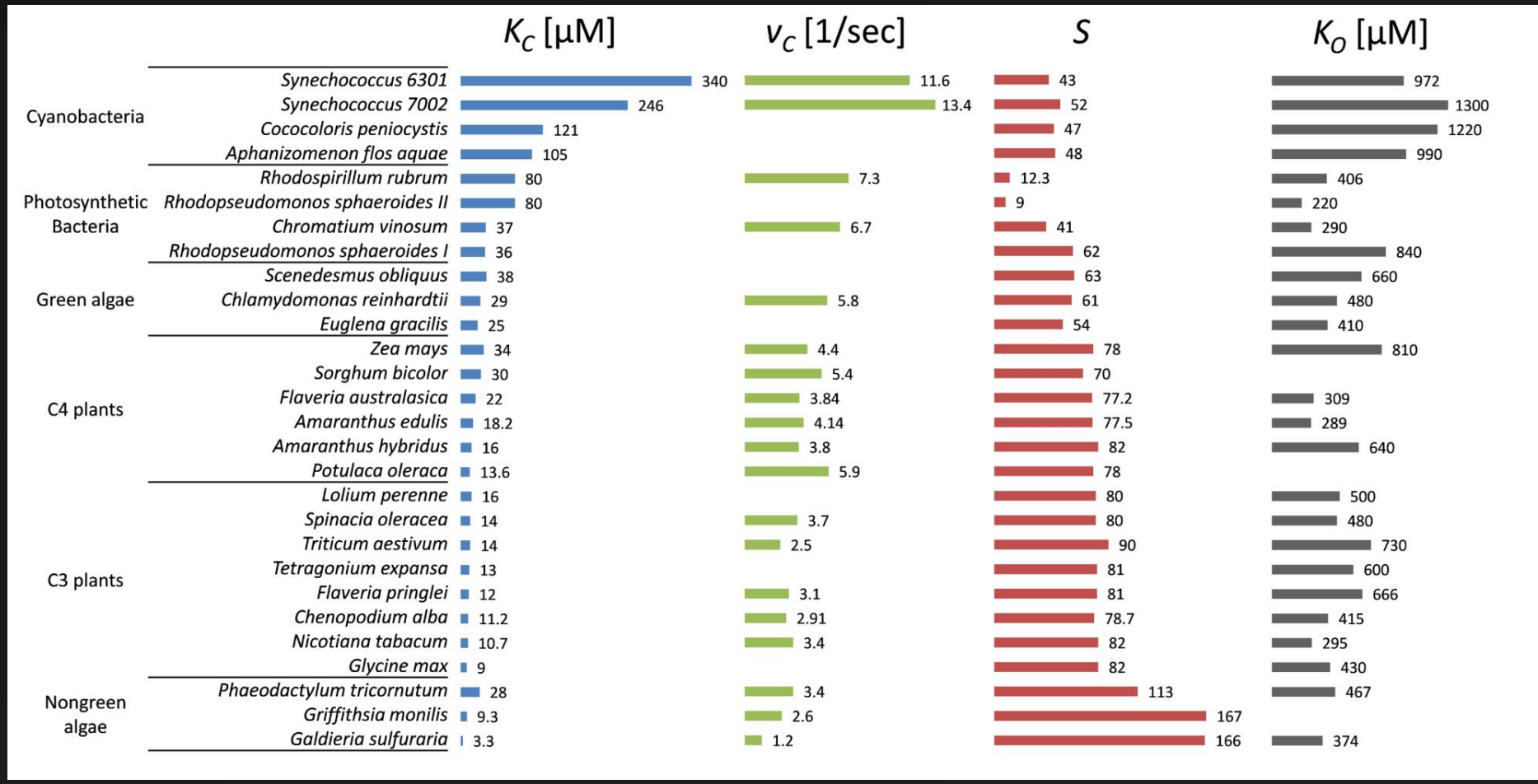
- Adapted Rubisco?

$$P_C = \frac{\frac{[CO_2]}{K_C}}{1 + \frac{[CO_2]}{K_C} + \frac{[O_2]}{K_O}},$$

$$P_O = \frac{\frac{[O_2]}{K_C}}{1 + \frac{[CO_2]}{K_C} + \frac{[O_2]}{K_O}}.$$

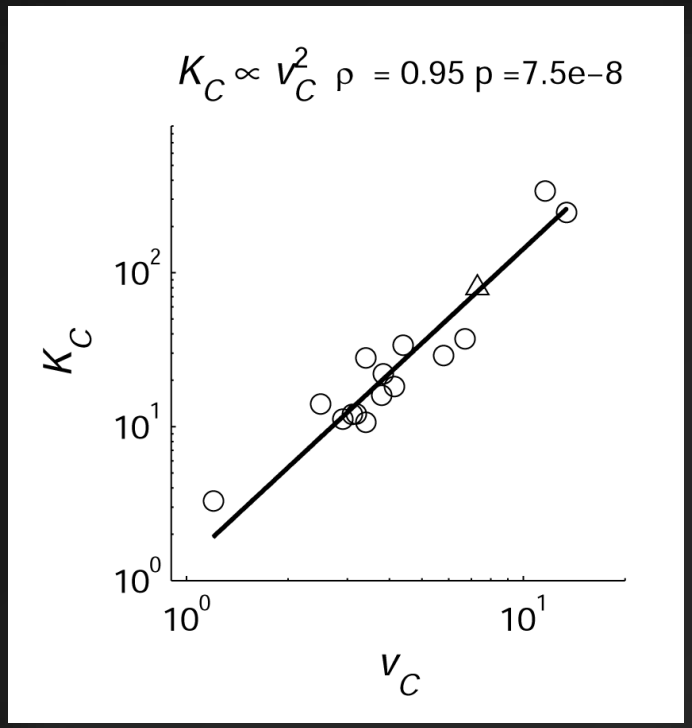
Rubisco kinetic parameters were measured in diverse organisms

- data from 28 eukaryotes and prokaryotes:



Rubisco's parameter space is effectively 1D

- **4D** data space (S, v_C , K_C , K_O).
- constrained to **1D power law.**



$$K_C \propto v_C^{2.0\pm0.2}$$

$$K_O \propto v_C^{0.5\pm0.1}$$

$$S \propto v_C^{-0.5\pm0.1}$$

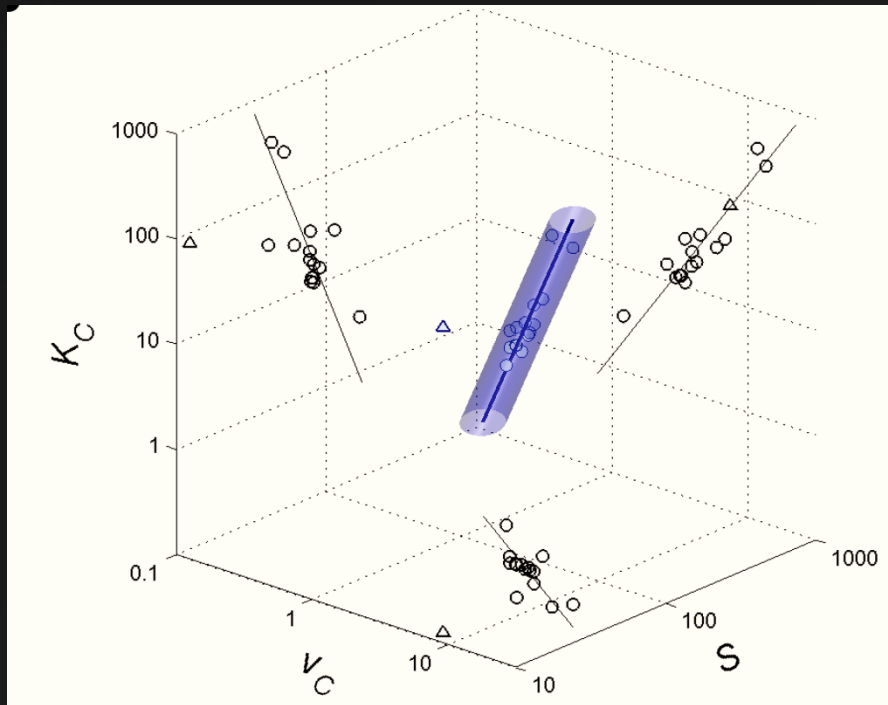
Dimensional reduction

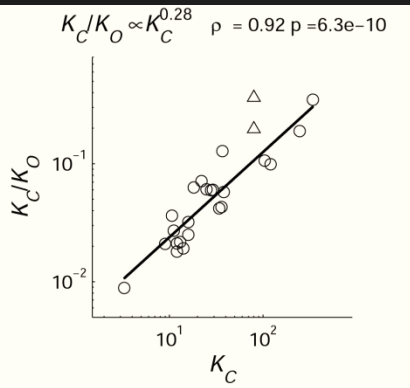
("Constrained Plasticity")

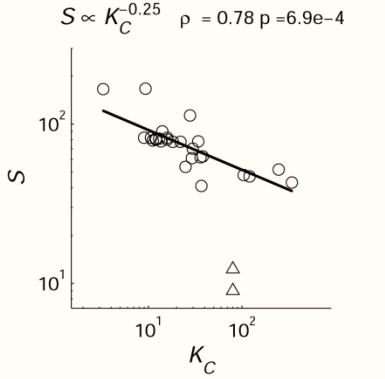
(Savir, Noor, Milo et al., PNAS 2010)

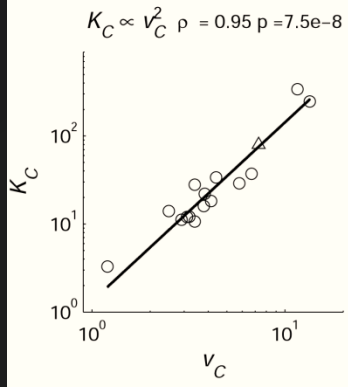
Effective 1D landscape for Rubisco evolution

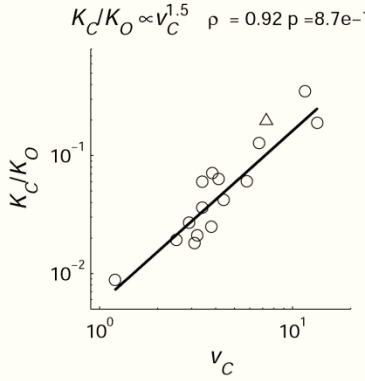
- All organisms scattered around a straight line in 4D parameter space.

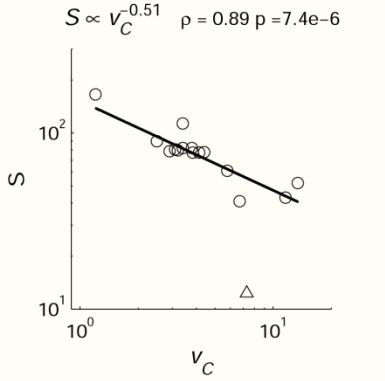






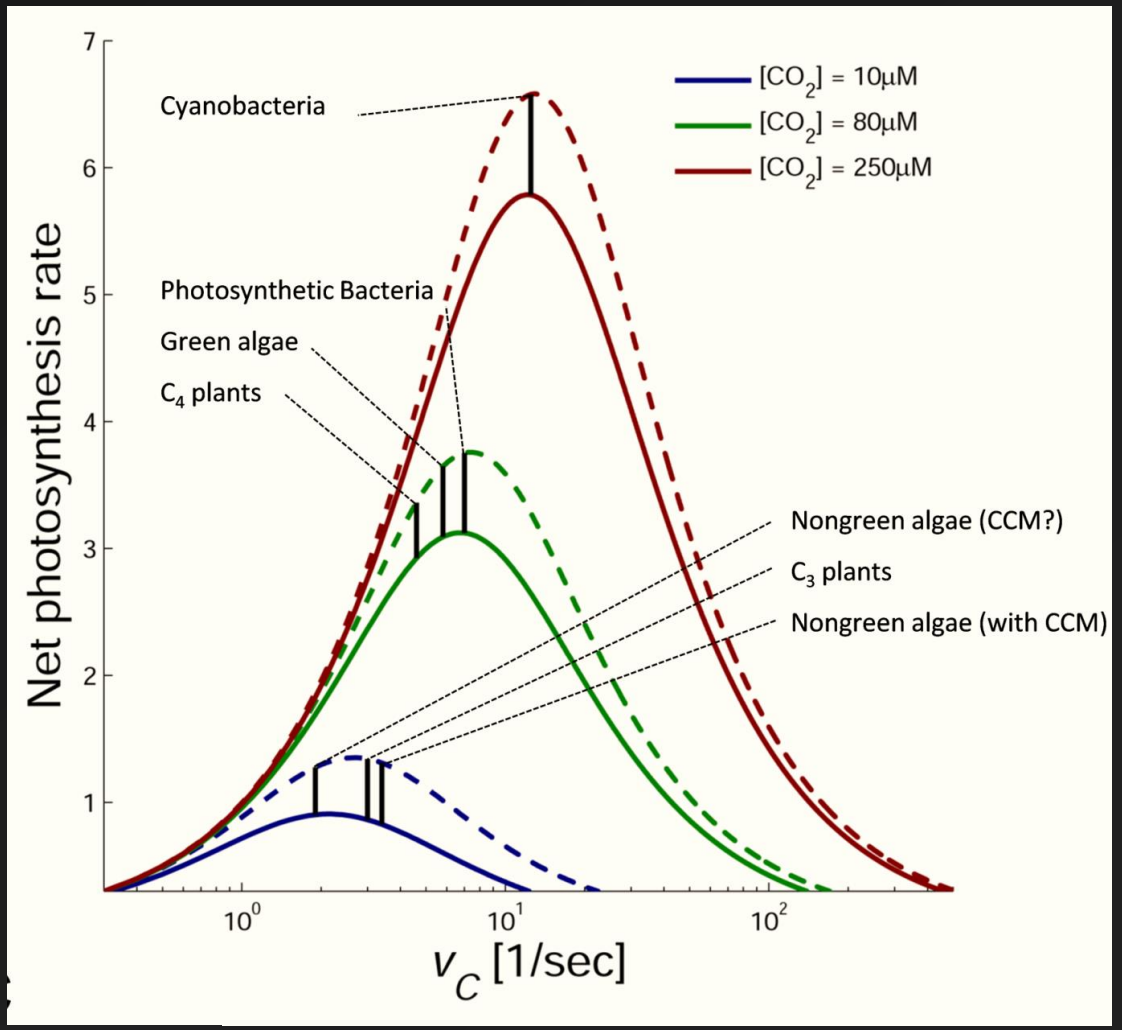






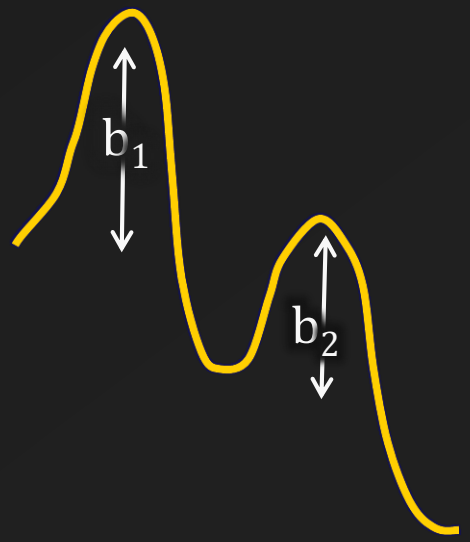
Rubisco energy barriers are nearly optimal for recognition of CO_2 in their habitats

- Max(F) along power-laws in given habitat (O_2 + CO_2).



$$F(v_C) = \frac{v_C}{1 + 1.3 \cdot v_C^2/[CO_2]} + f([O_2])$$

- Linear energy constraints?



$$K_C \sim v_C^2 \quad \leftrightarrow \quad b_1 + b_2 = \text{const.}$$

- Molecular phenotype spaces appear low-dimensional,
  ribosome, RecA, rubisco… – is it generic?

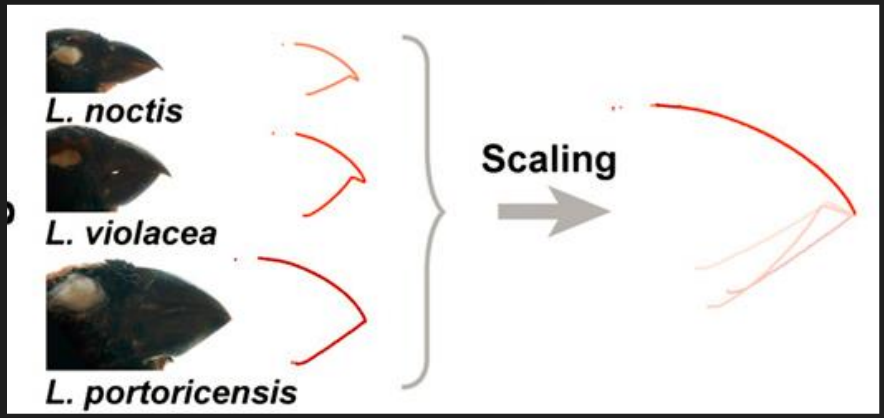
- Simple, coarse-grained thinking about function may explain basic features.

- Conformational changes enhance recognition.

- Dimensional reduction in other living systems (development, behavior, neural…)

- General feature of mapping

  genotype → phenotype?



Mallarino, Campàs, Fritz, Burns, Weeks, Brenner, Abzhanov (PNAS 2012)

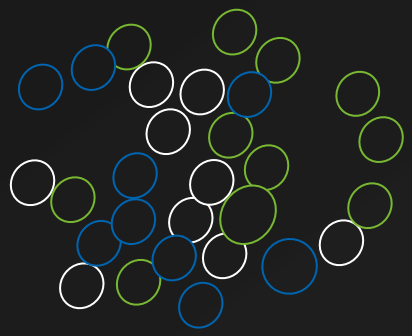
Savir et al.: *Cell* 2013, *Molecular Cell* 2010, *PNAS* 2010

What is the relation to physical properties of proteins as an evolvable state of matter?

- Slow manifolds of low dimension in various living systems.
- Hints: - transition states approximate kinetics.

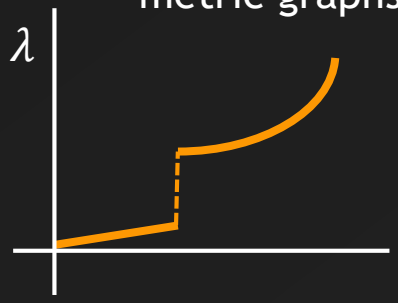
  - Low modes dominate protein function (spectral gap?)

- **What are the biophysical degrees-of-freedom of *evolvable, information-rich soft matter*.** (work in progress…)



Soft modes in amorphous solids

Soft $\omega \approx 0$ modes in metric graphs



Viscous response $\omega \leq 10^2$ Hz

Zocchi et al. (2012)

Thanks:

Yonatan Savir

Elad Noor, Ron Milo