Messenger RNA sequence determinants of protein production rate

Juraj Szavits-Nossan

School of Physics and Astronomy, University of Edinburgh

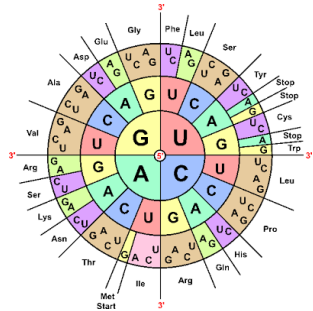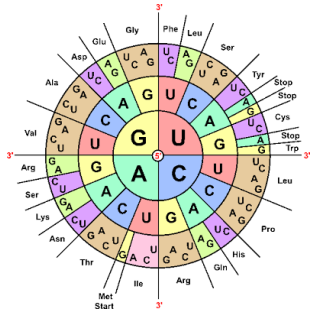May 17, 2018

Introduction

- 61 codons encode 20 amino acids

Introduction

- 61 codons encode 20 amino acids
- degeneracy: 18 of 20 amino acids encoded by more than one synonymous codon

Introduction

- 61 codons encode 20 amino acids

- degeneracy: 18 of 20 amino acids encoded by more than one synonymous codon

Example

a protein of 300 amino acids $\approx 10^{100}$ codon combinations

Introduction

- 61 codons encode 20 amino acids

- degeneracy: 18 of 20 amino acids encoded by more than one synonymous codon

Example

a protein of 300 amino acids $\approx 10^{100}$ codon combinations

Codon optimization

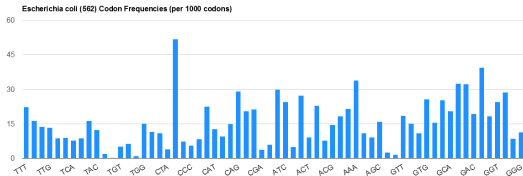Can we fine-tune expression levels by "silent" mutations?

Codon (usage) bias

codon bias = non-random distribution synonymous of codons

origin:
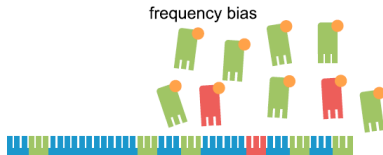
- mutation
- selection

E. coli K-12, Athley et al. (2017)

- codon adaptation index
  Sharp and Li (1986)

$$\mathrm{CAI} = \left(\prod_{k=1}^{L} w_k\right)^{1/L} \qquad w_k = \frac{f_k}{\max(f_k)}$$

- tRNA adaptation index
  dos Reis et al (2003)

frequency bias

Review: Quax et al (2015)
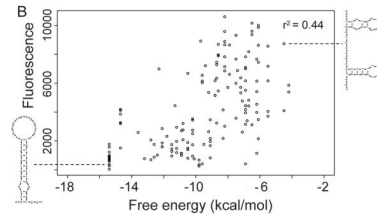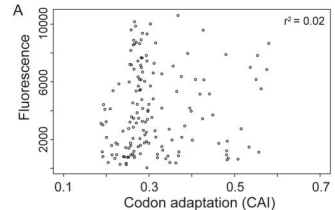
Translational efficiency

two definitions:

- protein production rate/mRNA
- ribosome density/mRNA Ingolia (2014)

rationale:

- faster codons → faster translation
- more bounded codons → more proteins

conflicting results in literature

154 gene variants of GFP in *E. coli*
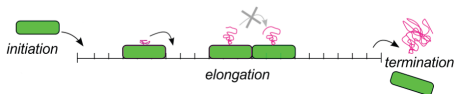

Kudla et al (2009)

Can physics help?
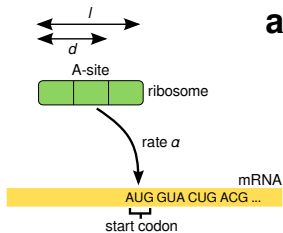
Biophysical modelling

MacDonald, Gibbs, Pipkin (1968)
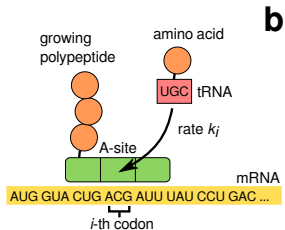

ribosomal current $J =$ mRNA translation rate

Totally Asymmetric Simple Exclusion Process (TASEP):

- more than one ribosome can simultaneously translate the same mRNA molecule
- unidirectional movement of ribosomes from 5' to 3' end
- excluded volume interactions that prevent ribosomes to pass each
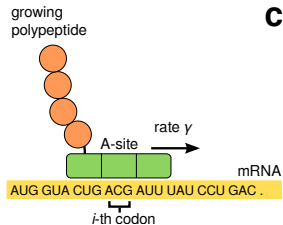- stochastic nature of the process Blanchard et al (2004)
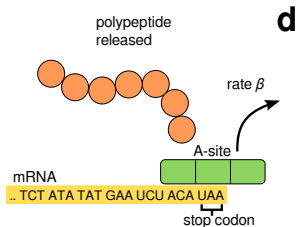
Kinetic rates

initiation, rate α

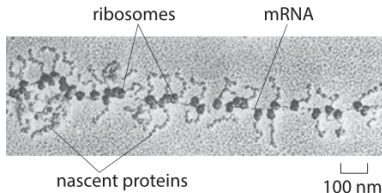aa delivery, rate $k_i \propto [\mathrm{tRNA}]_i$

translocation, rate γ

termination, rate β

Mathematical complications

ribosomal interference
(many-body problem)

codon-dependent
elongation rates
("disorder")

energy-driven process
(dissipative)

$$P(\text{steady state}) \neq \frac{1}{Z} e^{-\frac{E}{kT}}$$

Mathematical complications
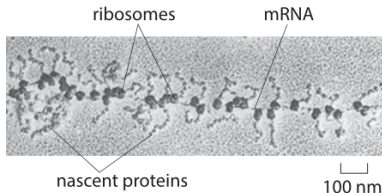
ribosomal interference
(many-body problem)

codon-dependent
elongation rates
("disorder")

energy-driven process
(dissipative)

$$P(\text{steady state}) \neq \frac{1}{Z}e^{-\frac{E}{kT}}$$

No analytical results

numerical simulations not helpful (too many parameters for optimization)

Parameter estimation for yeast (5836 genes)

Ciandrini et al (2013)

aa delivery rate
$k_i \propto [\text{tRNA}]_i = 1 - 35 \text{ codons/s}$

translocation rate
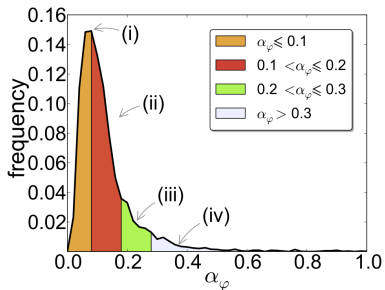$\gamma \approx 35 \text{ codons/s}$

termination rate
$\beta \approx \gamma$

initiation rate
$\alpha = 0.005 - 4.2 \text{ s}^{-1}$

$$J_{\text{simulations}}(\alpha) = \rho_{\text{experiments}}$$

Arava et al (2005)

translation initiation is a rate-limiting step

Series expansion in α

ribosomal current J (mRNA translation rate)

$$J(\alpha, \beta, \gamma, k_1, \ldots, k_L) = \alpha - \boxed{\sum_{i=1}^{10} \left(\frac{1}{\gamma} + \frac{1}{k_i} \right) \alpha^2} + O(\alpha^3)$$
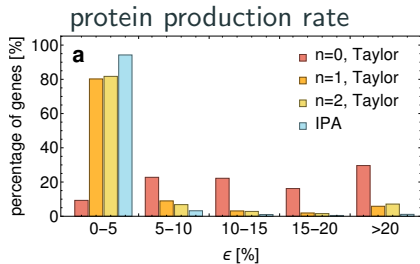
$\rightarrow$ first 10 codons most important

ribosome density

$$\rho(\alpha, \beta, \gamma, k_1, \ldots, k_L) = \frac{1}{L} \sum_{i=1}^{L} \left(\frac{1}{\gamma} + \frac{1}{k_i} \right) \alpha + O(\alpha^2)$$

$\rightarrow$ all codons important

Szavits-Nossan, Ciandrini, Romano (2018)

Comparison with numerical simulations
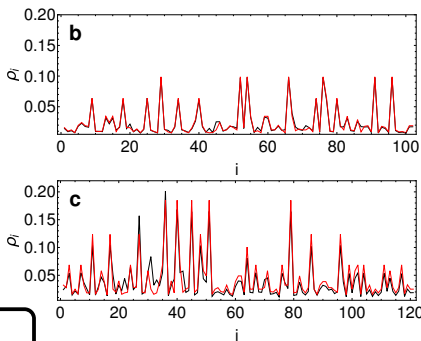

protein production rate

percentage error

$$\epsilon = \frac{|J_{\text{simulations}} - J_{\text{series}}|}{J_{\text{simulations}}}$$
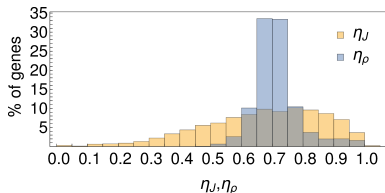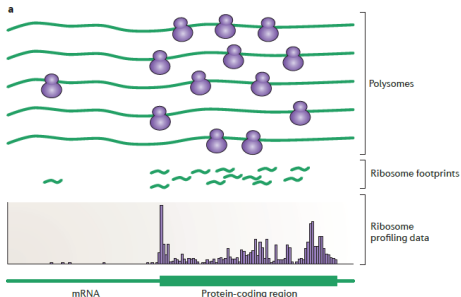
$J \approx \alpha$ accounts for 10% of genes

ribosome profiles for two genes

Conclusions

- codon adaptation metrics (CAI, tAI) unlikely to resolve codon bias puzzle

- dynamics essential

- translation can be monitored by ribosome profiling Ingolia et al (2008)

- prediction: only first 10 codons determine the translation rate

- optimal codons maximize number of free ribosomes?

Collaborators

Luca Ciandrini
Université de Montpellier, France

Mamen Romano
University of Aberdeen, UK

Thank you for your attention!