

Mathematical models for gene expression data

Juraj Szavits-Nossan

School of Physics and Astronomy, University of Edinburgh

Growth and division in mathematics and medicine

5 November 2019, UCL



The Leverhulme Trust

Background

nonequilibrium statistical physics

stochastic processes

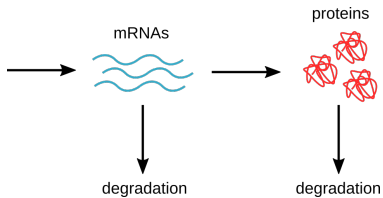
large deviations theory

Interests

mathematical models of transcription and translation

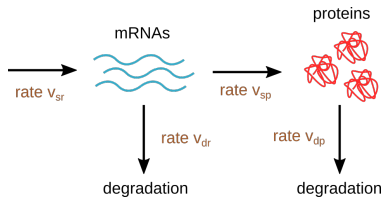
inferring model parameters from biological data (next generation sequencing, fluorescence imaging)

Quantification of gene expression



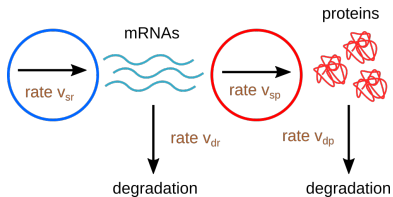
How is gene expression orchestrated **in time**?

Quantification of gene expression



How is gene expression orchestrated **in time**?

Quantification of gene expression

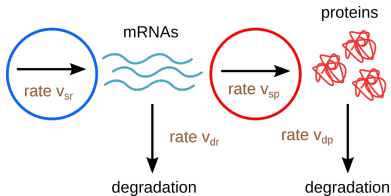


How is gene expression orchestrated **in time**?

$$\frac{dR}{dt} = v_{sr} - v_{dr}R$$

$$\frac{dP}{dt} = v_{sp}R - v_{dp}P$$

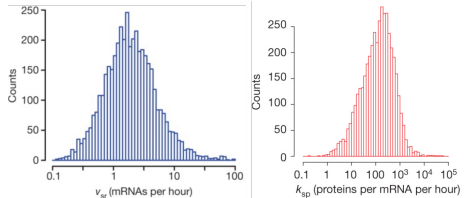
Quantification of gene expression



How is gene expression orchestrated **in time**?

$$\frac{dR}{dt} = v_{sr} - v_{dr}R$$
$$\frac{dP}{dt} = v_{sp}R - v_{dp}P$$

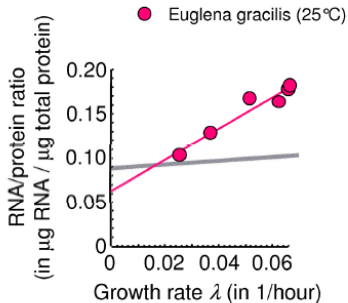
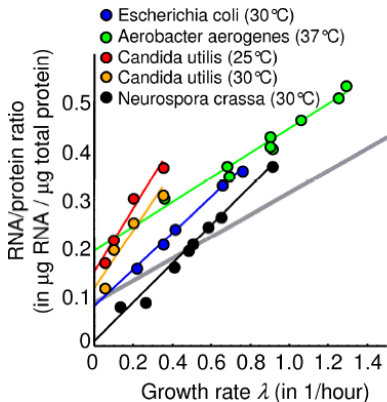
transcription and translation rates



Schwanhäusser et al, Nature 495, 126–127 (2013)

Bacterial “growth laws”

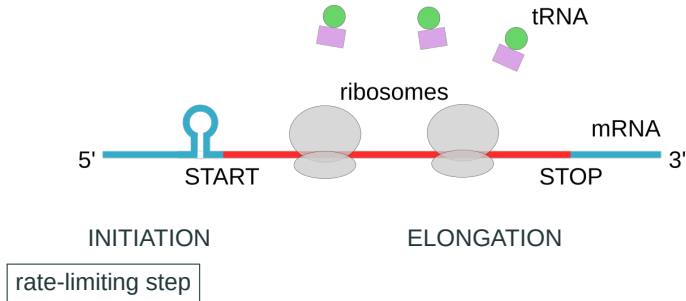
Scott et al, Science 19, 1099-102 (2010)



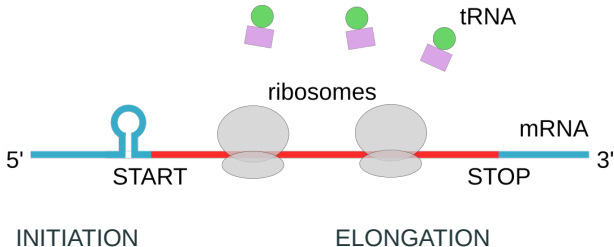
$$\text{RNA/protein ratio} = \frac{R}{P} = r_0 + \frac{\lambda}{\kappa_t}, \quad \lambda = \text{growth rate}$$

What determines transcription and translation rates?

Translational control in gene expression



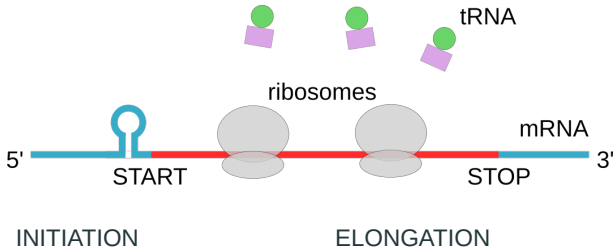
Translational control in gene expression



rate-limiting step

mRNA secondary structures

Translational control in gene expression

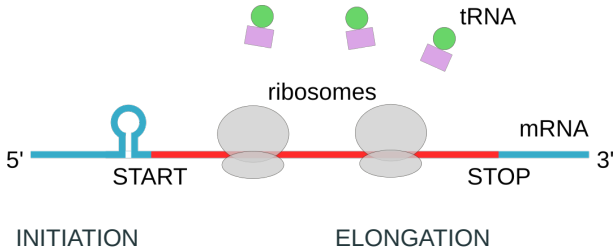


rate-limiting step

mRNA secondary structures

Kozak sequence (eukaryotes)

Translational control in gene expression



rate-limiting step

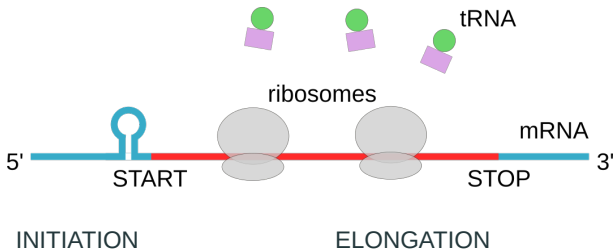
mRNA secondary structures

Kozak sequence (eukaryotes)

Shine-Dalgarno sequence

(prokaryotes)

Translational control in gene expression



rate-limiting step

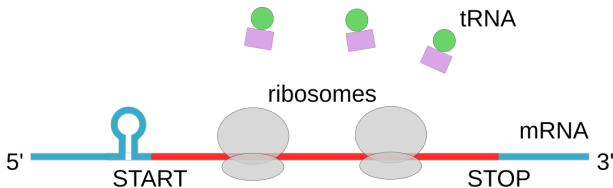
mRNA secondary structures

Kozak sequence (eukaryotes)

Shine-Dalgarno sequence
(prokaryotes)

5' upstream ORFs

Translational control in gene expression



INITIATION

rate-limiting step

mRNA secondary structures

Kozak sequence (eukaryotes)

Shine-Dalgarno sequence

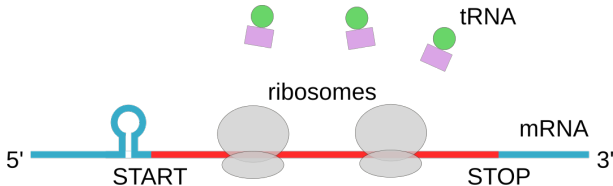
(prokaryotes)

5' upstream ORFs

ELONGATION

controversial

Translational control in gene expression



INITIATION

rate-limiting step

mRNA secondary structures

Kozak sequence (eukaryotes)

Shine-Dalgarno sequence
(prokaryotes)

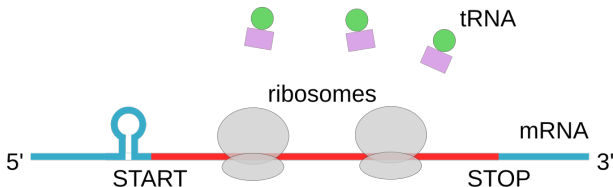
5' upstream ORFs

ELONGATION

controversial

codon choice

Translational control in gene expression



INITIATION

rate-limiting step

mRNA secondary structures

Kozak sequence (eukaryotes)

Shine-Dalgarno sequence
(prokaryotes)

5' upstream ORFs

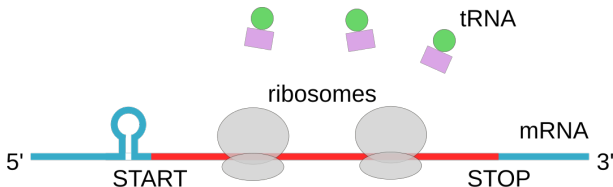
ELONGATION

controversial

codon choice

tRNA abundance

Translational control in gene expression



INITIATION

rate-limiting step

mRNA secondary structures

Kozak sequence (eukaryotes)

Shine-Dalgarno sequence
(prokaryotes)

5' upstream ORFs

ELONGATION

controversial

codon choice

tRNA abundance

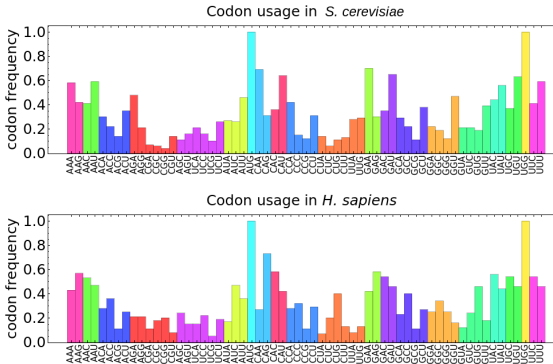
mRNA secondary structures

Degeneracy of the genetic code

codon usage bias = non-uniform frequencies of synonymous codons

Degeneracy of the genetic code

codon usage bias = non-uniform frequencies of synonymous codons



Codon Adaptation Index (CAI)

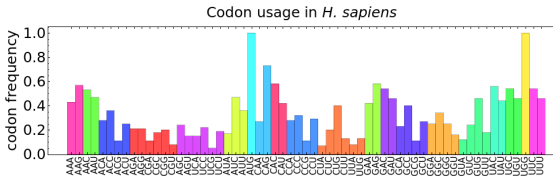
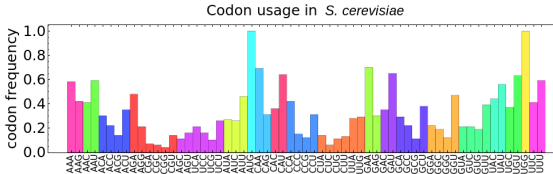
Sharp and Li, Nucleic Acids
Res.15, 1281-95 (1986)

$$\text{CAI} = \left(\prod_{k=1}^L w_k \right)^{1/L}$$

$$w_k = \frac{f_k}{\max(f_k)}$$

Degeneracy of the genetic code

codon usage bias = non-uniform frequencies of synonymous codons



Codon Adaptation Index (CAI)

Sharp and Li, Nucleic Acids

Res.15, 1281-95 (1986)

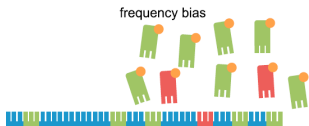
$$CAI = \left(\prod_{k=1}^L w_k \right)^{1/L}$$

$$w_k = \frac{f_k}{\max(f_k)}$$

How is the choice of codons affecting translation?

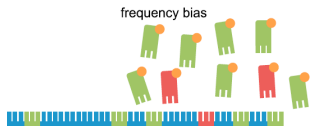
Codon optimisation (controversial)

rare codons = less abundant
tRNA (slower decoding rate)



Codon optimisation (controversial)

rare codons = less abundant
tRNA (slower decoding rate)

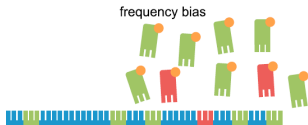


Protein	Host	Improvement
Ig κ chain	<i>S. cerevisiae</i>	> 50-fold
LLO	<i>M. musculus</i>	100-fold
L1 and L2	<i>Mammalian</i>	> 1000-fold
GFP	<i>H. sapiens</i>	22-fold
Gp120	<i>H. sapiens</i>	> 40-fold

Gustafsson et al, Trends Biotechnol 22, 346–53 (2004)

Codon optimisation (controversial)

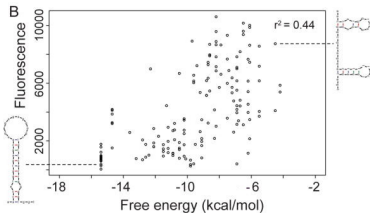
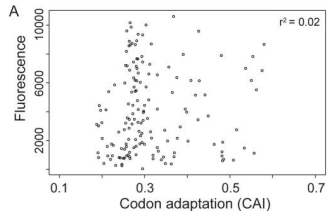
rare codons = less abundant
tRNA (slower decoding rate)



Protein	Host	Improvement
Ig κ chain	<i>S. cerevisiae</i>	> 50-fold
LLO	<i>M. musculus</i>	100-fold
L1 and L2	<i>Mammalian</i>	> 1000-fold
GFP	<i>H. sapiens</i>	22-fold
Gp120	<i>H. sapiens</i>	> 40-fold

Gustafsson et al, Trends Biotechnol 22, 346–53 (2004)

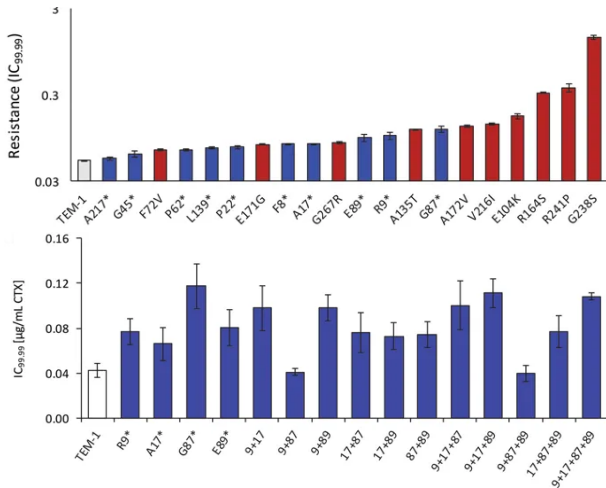
154 gene variants of GFP
in *E. coli*



Kudla et al, Science 324, 255–258 (2009)

Synonymous mutations in TEM-1 β -lactamase

inactivates penicillin and cephalosporin antibiotics

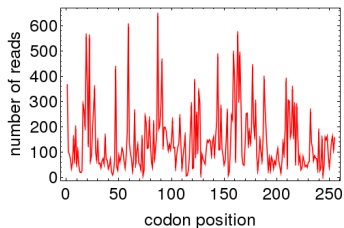
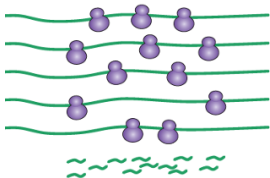


Zwart et al, Heredity 121, 406–421 (2018)

Probing ribosome dynamics

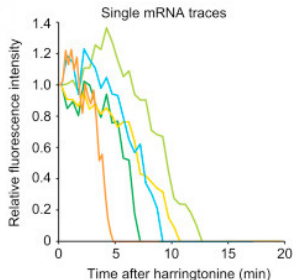
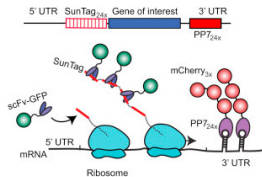
Ribosome profiling

Ingolia et al, Science 324, 218–223 (2009)



Fluorescence imaging

Yan et al, Cell 165, 976–989 (2016)



**How do we reconstruct ribosome dynamics
from the gene expression data?**

Biophysical modelling of mRNA translation

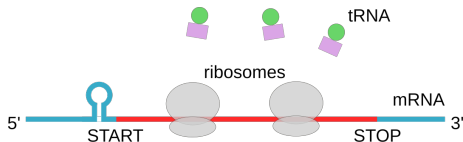
Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

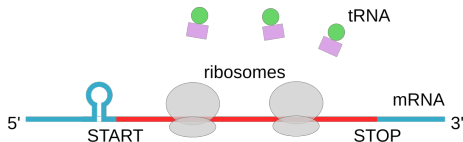
$$\text{elongation: } A_i \emptyset_{i+l} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

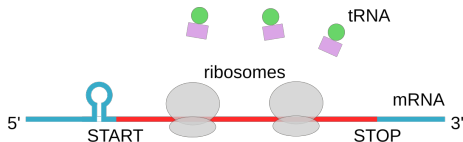
$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

stochastic: random decoding times (exponential distribution)

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

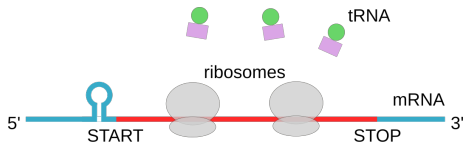
stochastic: random decoding times (exponential distribution)

non-equilibrium: GTP-driven movement $5' \rightarrow 3'$

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

stochastic: random decoding times (exponential distribution)

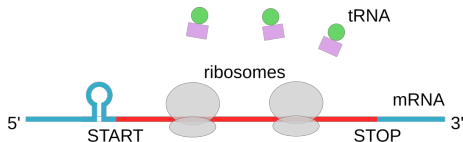
non-equilibrium: GTP-driven movement $5' \rightarrow 3'$

many-body: multiple ribosomes allowed per mRNA

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

stochastic: random decoding times (exponential distribution)

non-equilibrium: GTP-driven movement $5' \rightarrow 3'$

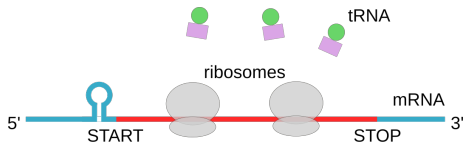
many-body: multiple ribosomes allowed per mRNA

interacting: excluded-volume interactions between ribosomes

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

stochastic: random decoding times (exponential distribution)

non-equilibrium: GTP-driven movement $5' \rightarrow 3'$

many-body: multiple ribosomes allowed per mRNA

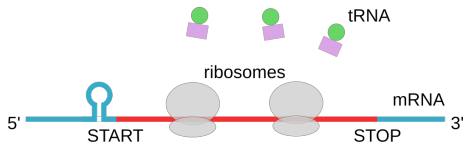
interacting: excluded-volume interactions between ribosomes

inhomogeneous: codon-dependent elongation rates

Biophysical modelling of mRNA translation

Totally Asymmetric Simple Exclusion Process (TASEP)

MacDonald, Gibbs, Pipkin, Biopolymers 6, 15 (1968)



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

stochastic: random decoding times (exponential distribution)

non-equilibrium: GTP-driven movement $5' \rightarrow 3'$

many-body: multiple ribosomes allowed per mRNA

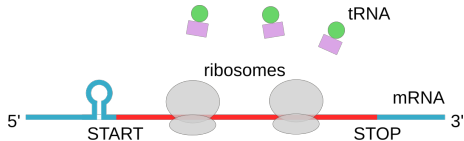
interacting: excluded-volume interactions between ribosomes

inhomogeneous: codon-dependent elongation rates

Reviews: van der Haar, CSBJ 1, e201204002 (2012), Zur and Tuller, Nucleic Acids Res 44, 9031–9049

(2016)

Steady-state solution

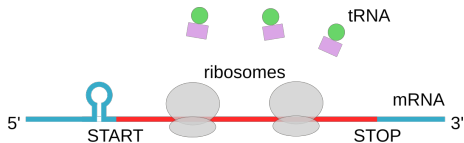


initiation: $\emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$

elongation: $A_i \emptyset_{i+l} \xrightarrow{k_i} \emptyset_i A_{i+1}$

termination: $A_L \xrightarrow{\beta} \emptyset_L$

Steady-state solution



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$

$$\text{elongation: } A_i \emptyset_{i+\ell} \xrightarrow{k_i} \emptyset_i A_{i+1}$$

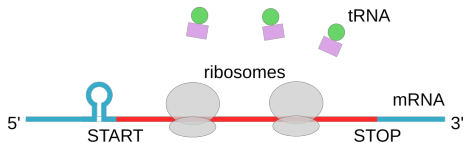
$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

Mathematical solution as a power series in the initiation rate α

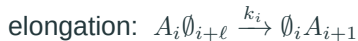
Szavits-Nossan, Ciandrini and Romano, *Phys. Rev. Lett.* **120** 128101 (2018)

Scott, Szavits-Nossan, bioRxiv 687335 (2019)

Steady-state solution



$$\text{initiation: } \emptyset_1 \dots \emptyset_\ell \xrightarrow{\alpha} A_2$$



$$\text{termination: } A_L \xrightarrow{\beta} \emptyset_L$$

Mathematical solution as a power series in the initiation rate α

Szavits-Nossan, Ciandrini and Romano, *Phys. Rev. Lett.* **120** 128101 (2018)

Scott, Szavits-Nossan, bioRxiv 687335 (2019)

mRNA
translation rate

$$J = \alpha - \sum_{i=1}^{\ell} \frac{\alpha^2}{k_i} + \dots$$

total
ribosome density

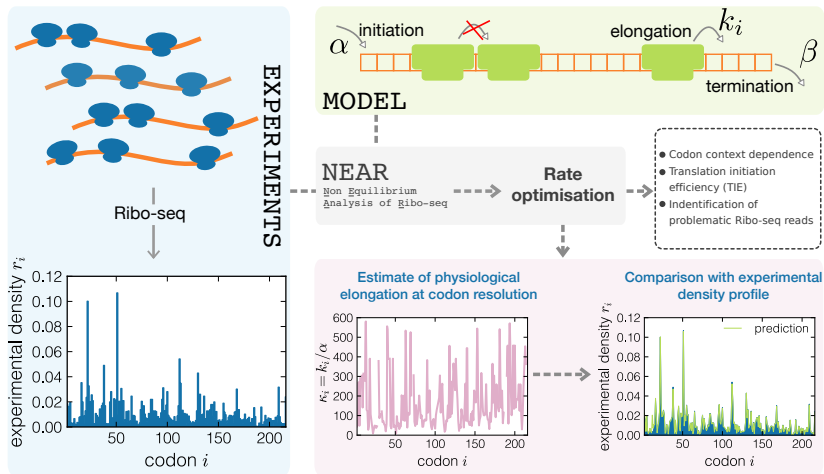
$$\rho = \frac{1}{L} \sum_{i=1}^L \frac{\alpha}{k_i} + \dots$$

local
ribosome density

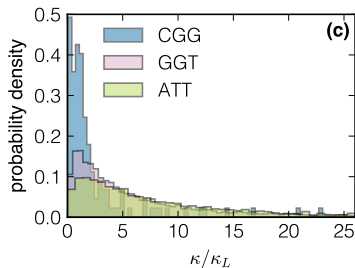
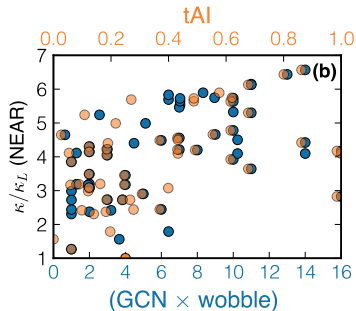
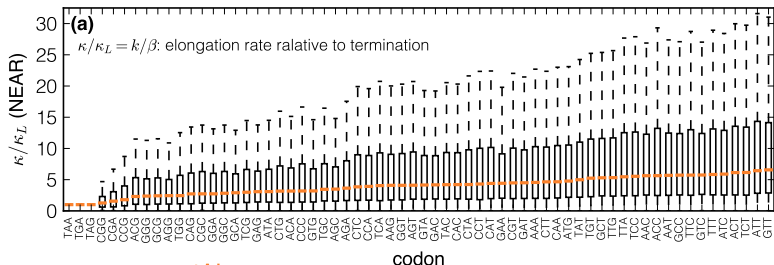
$$\rho_i = \frac{\alpha}{k_i} + \dots$$

Solving the inverse problem

Szavits-Nossan and Ciandrini, bioRxiv 719302 (2019)



Genome-wide elongation rates for *S. cerevisiae*



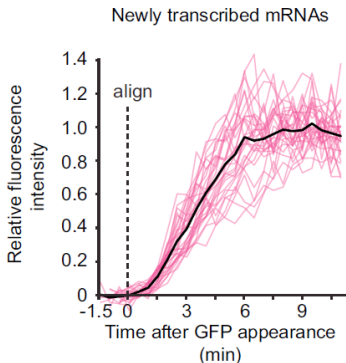
**How do we find sequence determinants of
the elongation rates?**

A little (final) problem...

most TASEP-based studies assume translation under
steady-state conditions

but

mRNA has a finite lifetime



What is the
relaxation time of
the TASEP?

Thank you for your attention!