

NON-EQUILIBRIUM BIOMOLECULAR INFORMATION PROCESSES

Pierre GASPARD

*Centre for Nonlinear Phenomena and Complex Systems
Université libre de Bruxelles (ULB), Brussels, Belgium*

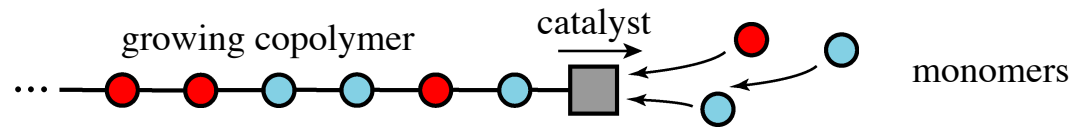
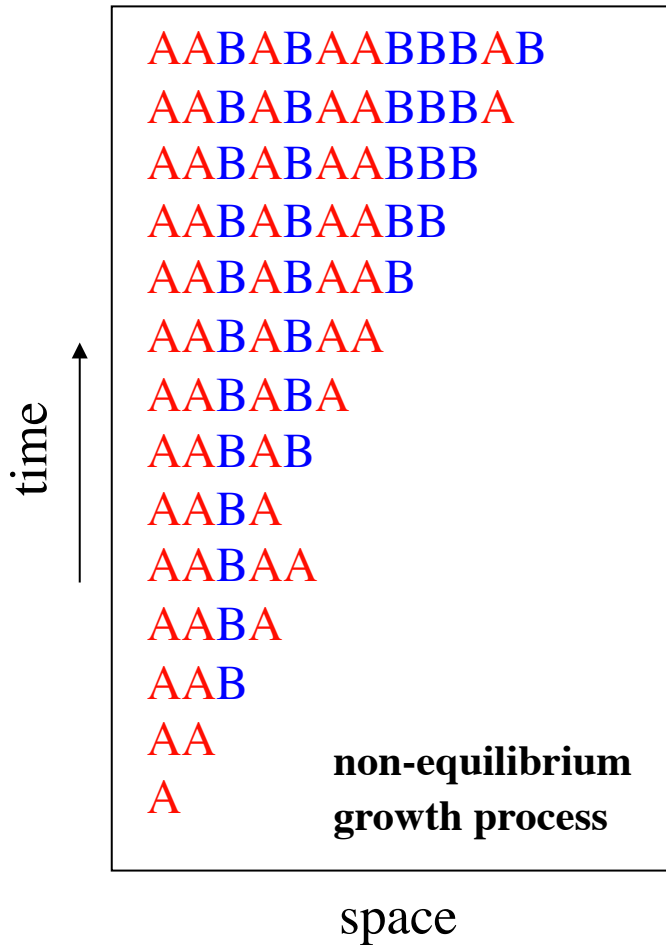
- **THERMODYNAMICS OF COPOLYMERIZATION**
- **DNA REPLICATION & ITERATED FUNCTION SYSTEM**
- **IMPLICATIONS FOR GENETIC DRIFT, MUTATIONS & EVOLUTION**

Minisymposium « Non-equilibrium Thermodynamics in Biology »,
Society for Mathematical Biology Annual Conference, June 13-17, 2021

COPOLYMERIZATION PROCESSES

statistical copolymer = spatial support of information

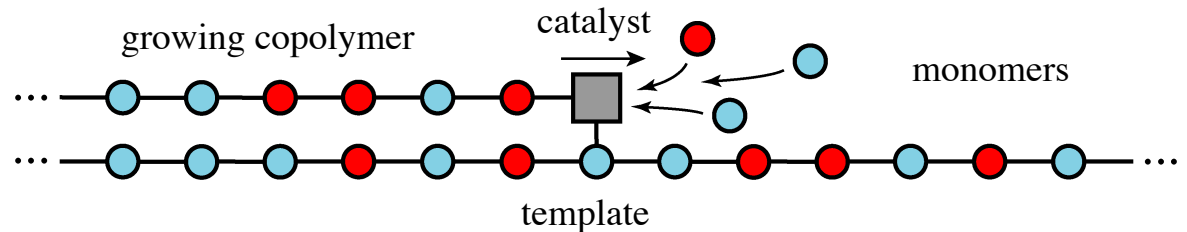
= aperiodic crystal by E. Schrödinger, *What is Life?* (1944)



free copolymerization:

statistical copolymers

ex: styrene-butadiene rubber



template-directed copolymerization:

ex: DNA replication

DNA-mRNA transcription

mRNA-protein translation

THERMODYNAMICS OF TEMPLATE-DIRECTED COPOLYMERIZATION

entropy production rate: $\frac{d_i S}{dt} = A v \geq 0$ mean growth velocity: v $k_B = 1$

entropy production
per monomer (affinity):

$$\begin{aligned} A &= \varepsilon + D(\text{copy}|\text{template}) \\ &= \varepsilon + D(\text{copy}) - I(\text{copy}, \text{template}) \end{aligned}$$

free-energy « driving force »: $\varepsilon = -\frac{\Delta G}{T}$ C. H. Bennett, Biosystems **11** (1979) 85

conditional disorder of the copy ω with respect to the template α : $D(\omega|\alpha)$ (replication errors)

overall disorder of the copy ω : $D(\omega)$ (information theory)

mutual information between the copy and the template: $I(\omega, \alpha) \equiv D(\omega) - D(\omega|\alpha) \geq 0$
(replication fidelity)

The copy can grow by the entropic effect of replication errors
in an adverse free-energy landscape with $\varepsilon < 0$, as long as $A = \varepsilon + D > 0$.
Replication errors can perform mechanical work.

DNA REPLICATION

template-directed copolymerization

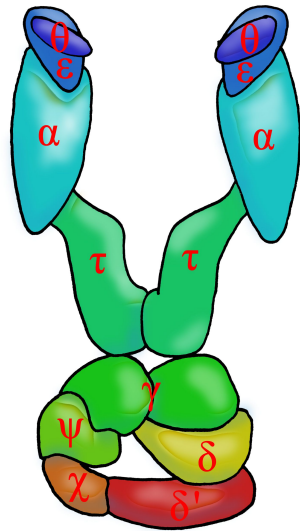
nucleotides

A: adenine T: thymine
C: cytosine G: guanine

Watson-Crick pairing:

A-T and C-G

powered by 2 ATP/nucleotide

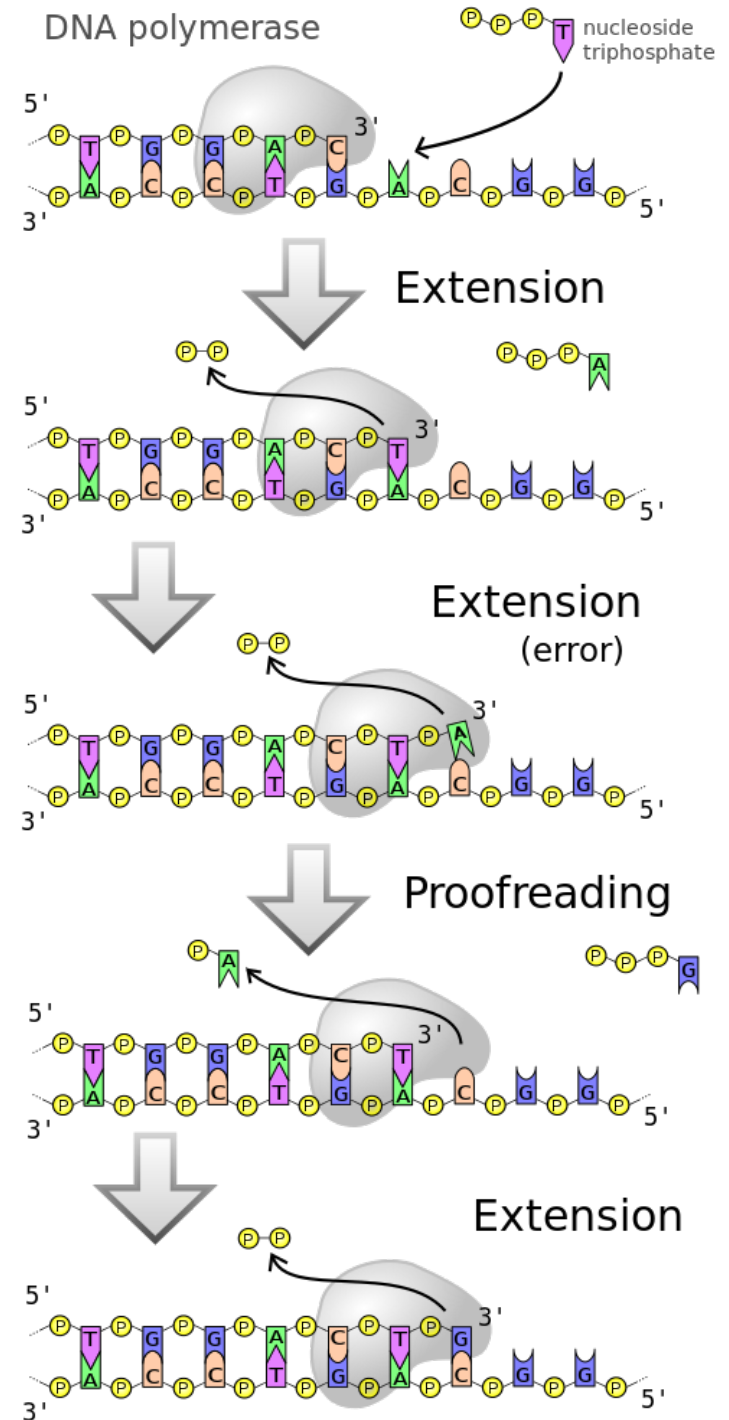


DNA polymerase III (prokaryotes)

α subunit: polymerase activity (DNA synthesis)

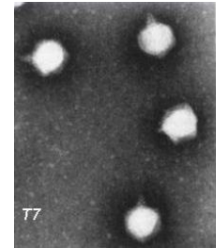
ϵ subunit: exonuclease activity (proofreading)

1000 nt/s



exo⁻ T7 DNA polymerase

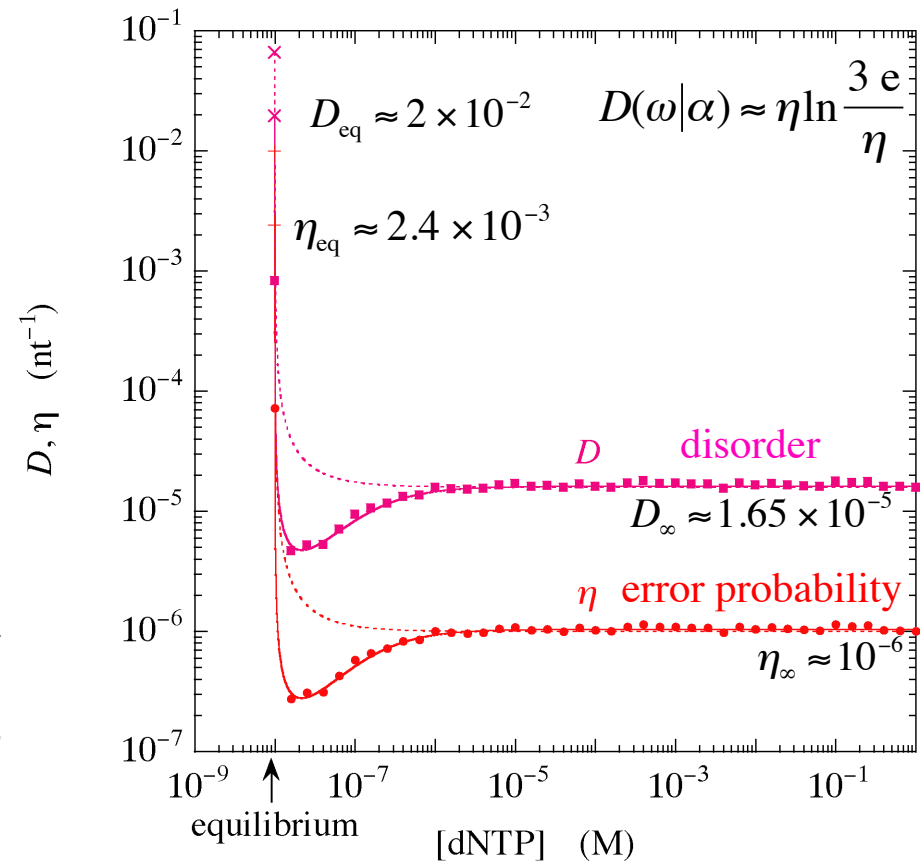
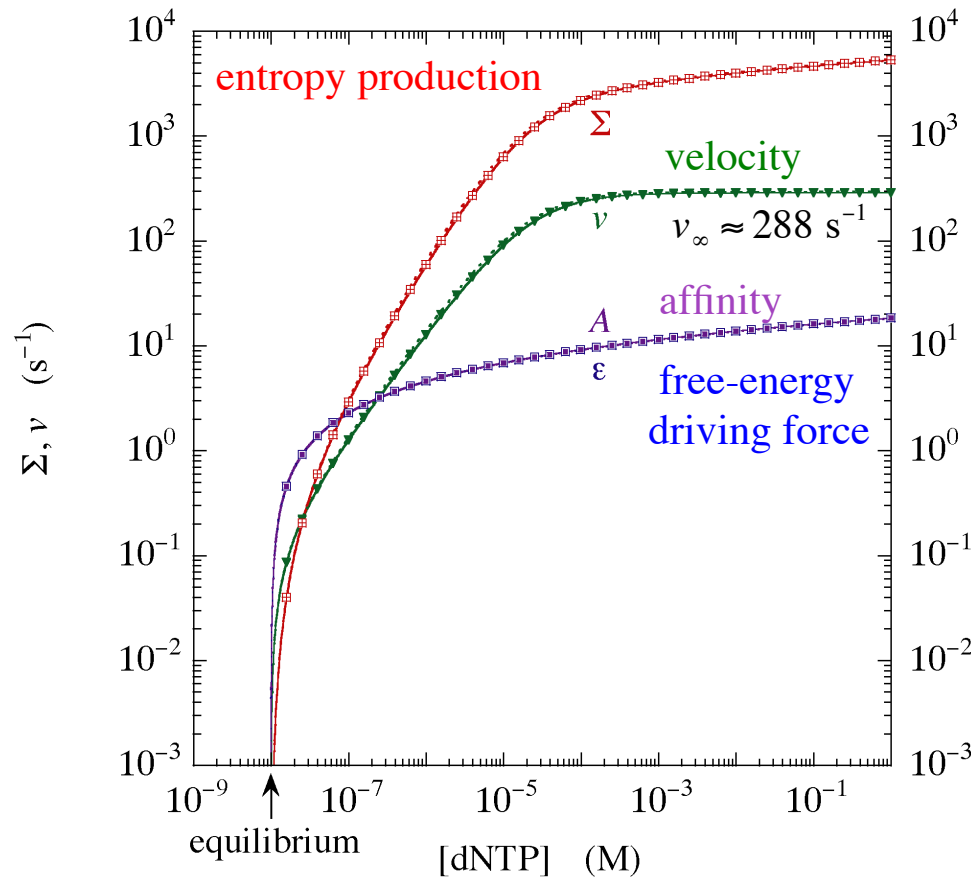
Data: K. A. Johnson, Annu. Rev. Biochem. **62** (1993) 685-713 $K_p \approx 200$ mM $[PP_i] = 10^{-4}$ M



Wikipedia

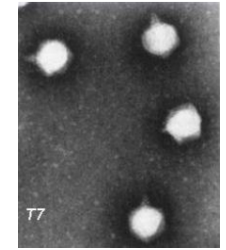
Rates for correct or incorrect nucleotide incorporation, depending on the previously incorporated nucleotide.

There is no effect of sequence heterogeneity: the template appears homogeneous.



exo⁺ T7 DNA polymerase

Data: K. A. Johnson, Annu. Rev. Biochem. **62** (1993) 685-713 $K_p \approx 200$ mM $[PP_i] = 10^{-4}$ M

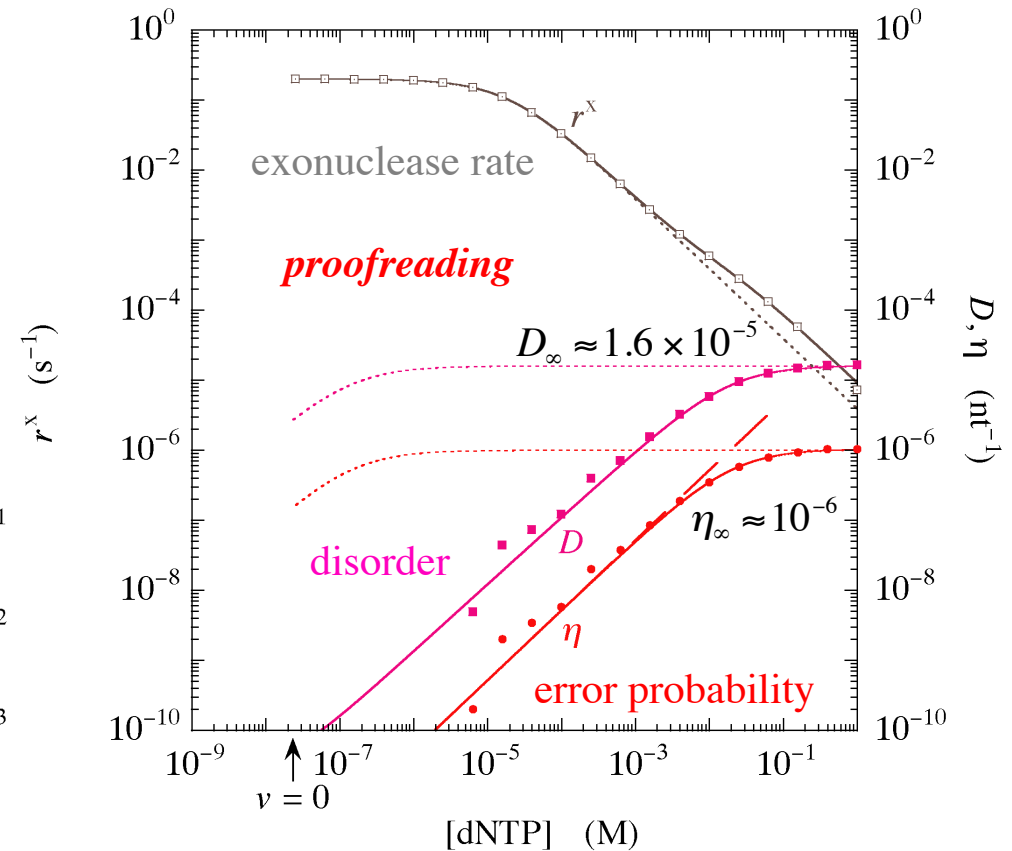
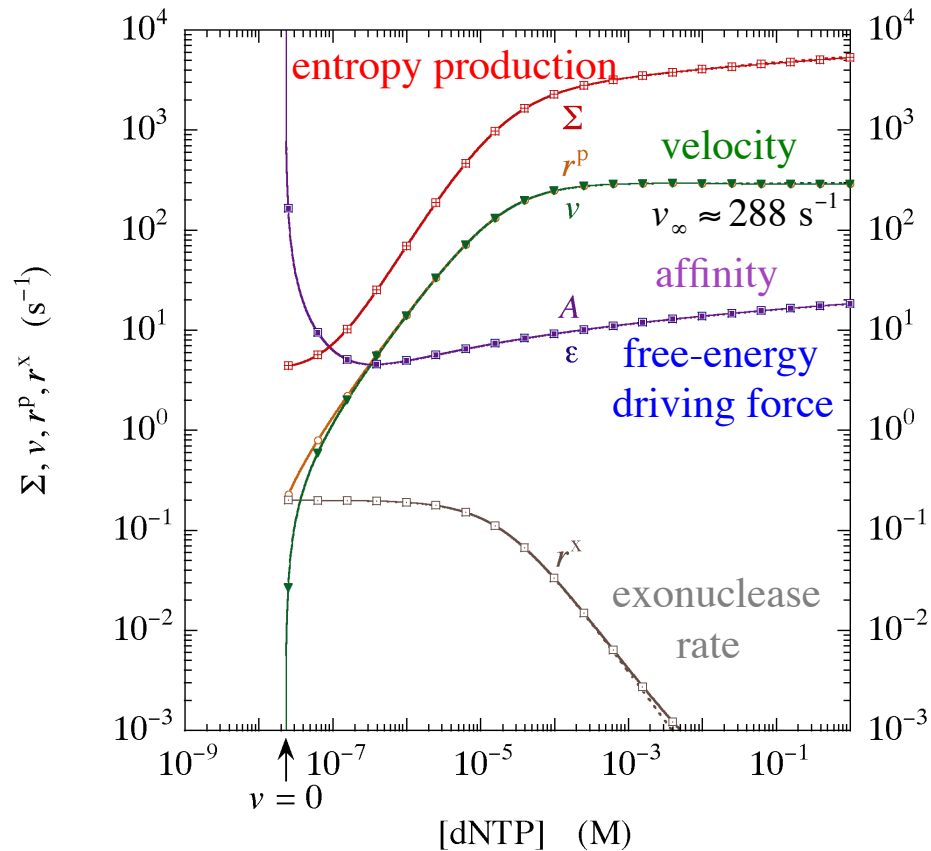


Wikipedia

Polymerase & exonuclease activities: **velocity** $v = r^P - r^X$

Proofreading by the exonuclease: **error probability** $\eta \ll \eta_\infty \approx 10^{-6}$

The polymerase is out of equilibrium, even if $v = 0$.



KINETICS WITH SEQUENCE HETEROGENEITY

nucleotides $m, n = A, C, G, T = 1, 2, 3, 4$

DNA copy: $\omega = m_1 m_2 \cdots m_{l-1} + m_l \Leftrightarrow m_1 m_2 \cdots m_{l-1} m_l$

DNA template: $\alpha = n_1 n_2 \cdots n_{l-1} n_l n_{l+1} \cdots n_1 n_2 \cdots n_{l-1} n_l n_{l+1} \cdots$

Kinetic equations:
$$\frac{d}{dt} P_t \begin{pmatrix} m_1 \cdots m_l \\ n_1 \cdots n_l n_{l+1} \cdots \end{pmatrix} = W_{+m_l, n_l} P_t \begin{pmatrix} m_1 \cdots m_{l-1} \\ n_1 \cdots n_{l-1} n_l \cdots \end{pmatrix} + \sum_{m_{l+1}} W_{-m_{l+1}, n_{l+1}} P_t \begin{pmatrix} m_1 \cdots m_l m_{l+1} \\ n_1 \cdots n_l n_{l+1} \cdots \end{pmatrix} - \left(W_{-m_l, n_l} + \sum_{m_{l+1}} W_{+m_{l+1}, n_{l+1}} \right) P_t \begin{pmatrix} m_1 \cdots m_l \\ n_1 \cdots n_l n_{l+1} \cdots \end{pmatrix}$$

attachment and detachment rates: $W_{\pm m, n}$

growth of a **Bernoulli chain**: $P_t \begin{pmatrix} m_1 \cdots m_l \\ n_1 \cdots n_l n_{l+1} \cdots \end{pmatrix} \approx p_t(l) \times \prod_{j=1}^l \mu(m_j, j)$

local velocity: x_l

mean growth velocity:

$$\frac{1}{v} = \left\langle \frac{1}{x} \right\rangle$$

iterated function system:
(Barnsley & Demko, 1985)

$$x_{l-1} = f_{n_l}(x_l) \quad \text{with}$$

$$f_n(x) \equiv x \sum_{m=1}^4 \frac{W_{+m, n}}{W_{-m, n} + x}$$

$$n \in \{1, 2, 3, 4\}$$

pairing probability:

$$\mu(m_l, l) = \frac{x_l}{x_{l-1}} \frac{W_{+m_l, n_l}}{W_{-m_l, n_l} + x_l}$$

$$\sum_{m_l} \mu(m_l, l) = 1$$

***10⁵-10⁶ faster than
Monte Carlo simulations***

exo⁻ human mitochondrial DNA pol γ

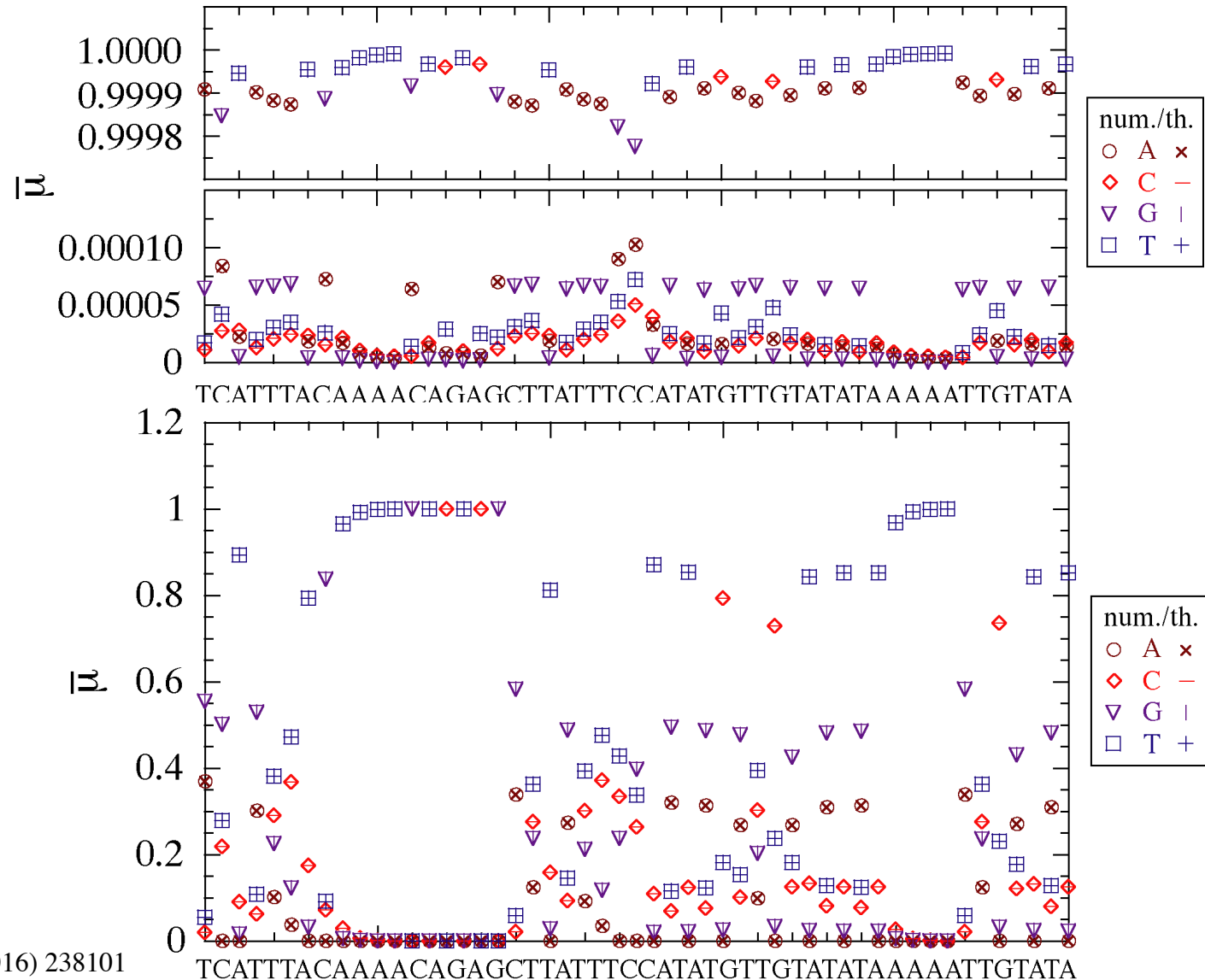
Data: A. A. Johnson and K. A. Johnson, J. Biol. Chem. **276** (2001) 38090-38096

H. R. Lee and K. A. Johnson, J. Biol. Chem. **281** (2006) 36236-36240

general
physiological
concentrations:
[dATP] = 24 μ M
[dCTP] = 29 μ M
[dGTP] = 5.2 μ M
[dTTP] = 37 μ M
error probability:
 $\eta \approx 9 \times 10^{-5}$

dATP imbalance
with respect to
physiological
concentrations:
[dATP] = 1 nM
error probability:
 $\eta \approx 0.27$

pairing probabilities



exo⁻ human mitochondrial DNA pol γ

Data: A. A. Johnson and K. A. Johnson, J. Biol. Chem. **276** (2001) 38090-38096

H. R. Lee and K. A. Johnson, J. Biol. Chem. **281** (2006) 36236-36240

normal cell
physiological
concentrations:

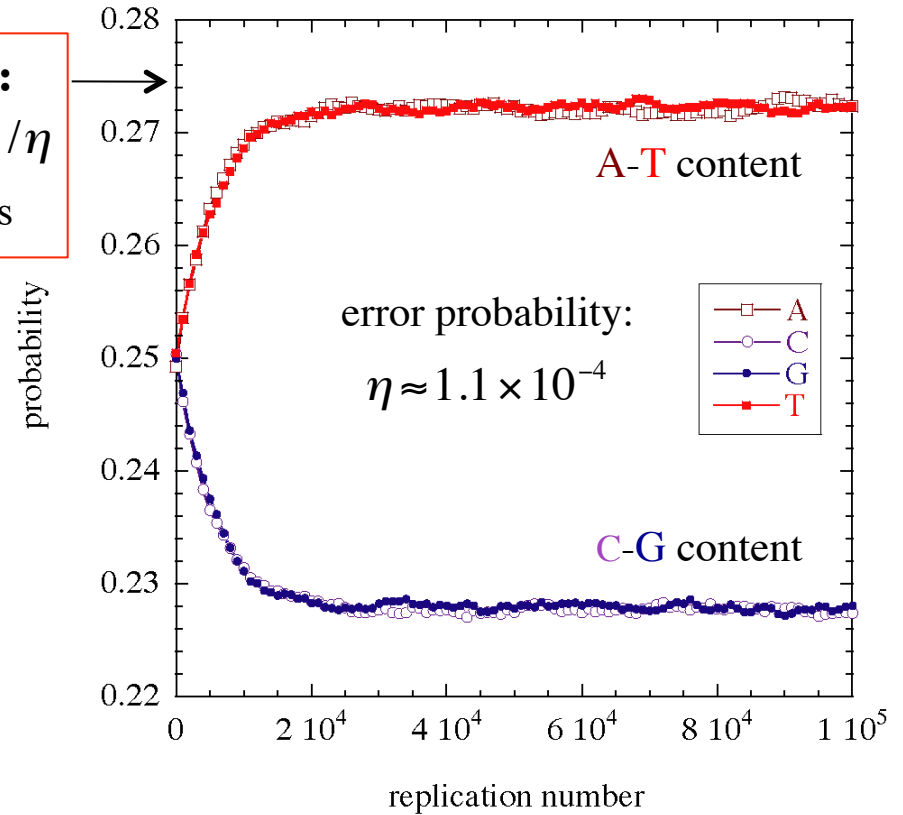
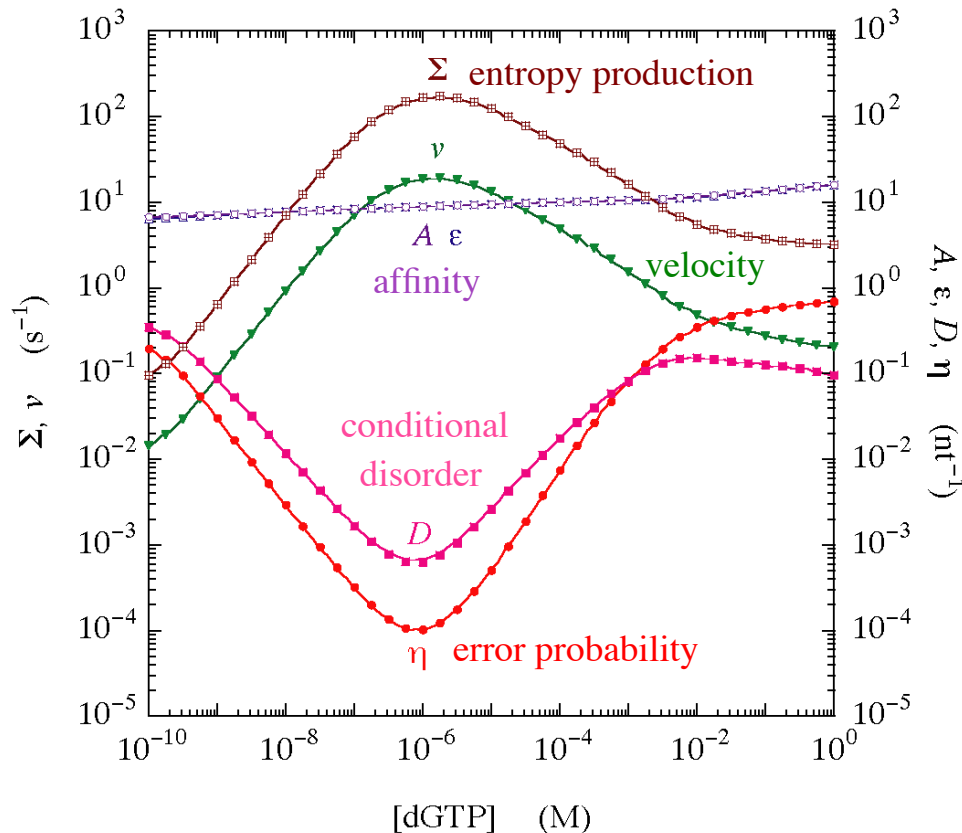
[dATP] = 3.2 μ M
[dCTP] = 2.1 μ M
[dGTP] = 1.5 μ M
[dTTP] = 5.4 μ M

Genetic drift:

$$t_{\text{drift}} \approx t_{\text{replication}} / \eta$$

Chargaff's rules

dGTP imbalance:



Human DNA:

nucleotide bulk probabilities:

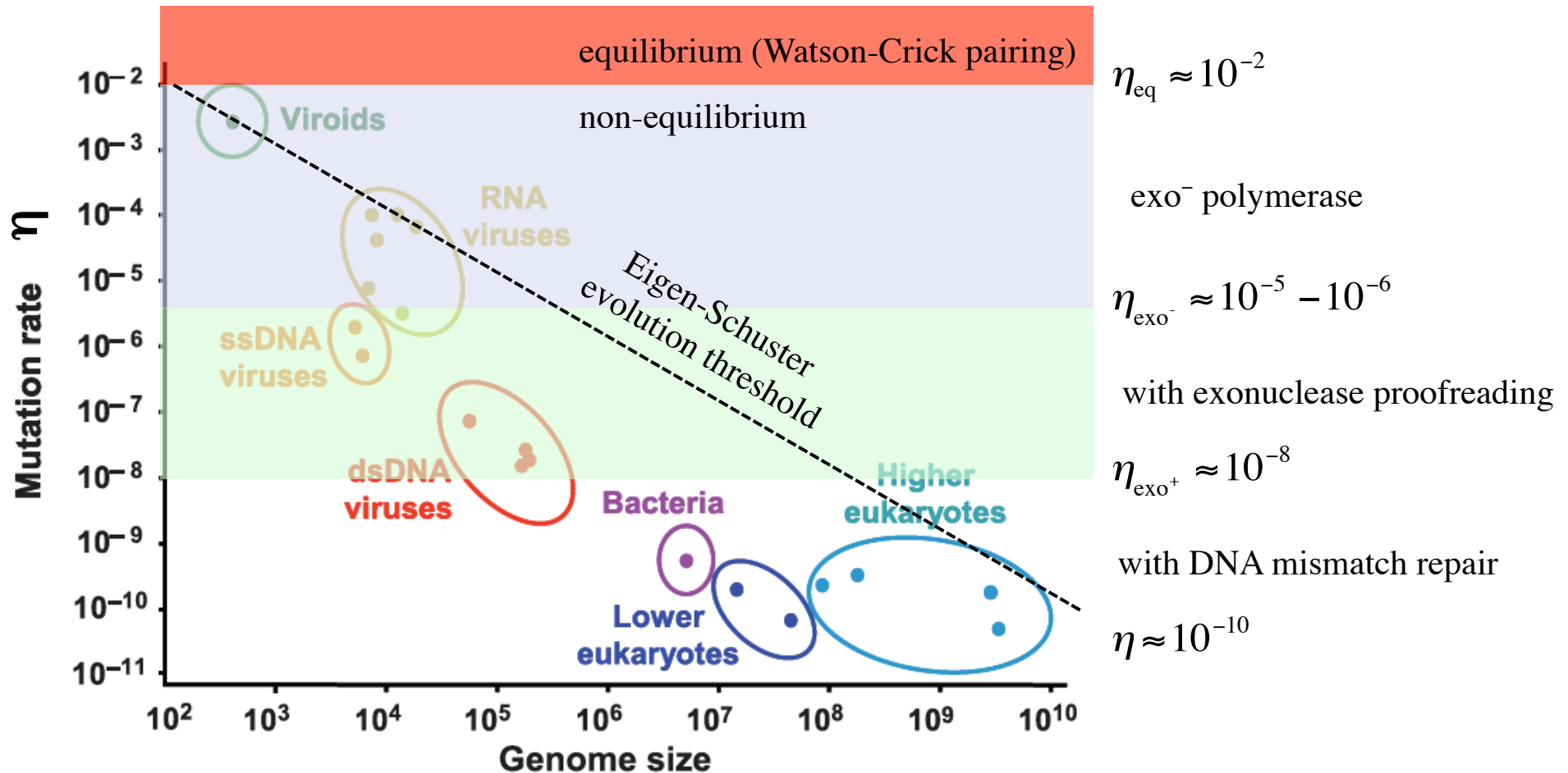
$$\left\{ \begin{array}{l} \bar{\mu}(A) \approx 0.2946 \\ \bar{\mu}(C) \approx 0.2051 \\ \bar{\mu}(G) \approx 0.2055 \\ \bar{\mu}(T) \approx 0.2949 \end{array} \right.$$

A. Provata, C. Nicolis, and G. Nicolis,
Comput. Biol. Chem. **53** (2014) 5-14

REPLICATION, MUTATIONS & EVOLUTION

Molecular fluctuations are sources of mutations

S. Gago, S. F. Elena, R. Flores, and R. Sanjuan, *Science* **323** (2009) 1308



M. Eigen and P. Schuster, *Naturwissenschaften* **64** (1977) 541: maximum genome size $\propto \frac{1}{\text{mutation rate}}$

CONCLUSIONS

Copolymers are natural supports of information, e.g., DNA, RNA.

Thermodynamics of template-directed copolymerization:

The entropy production rate depends on the mutual information I between copy and template:

$$\frac{1}{k_B} \frac{d_i S}{dt} = v \left[-\frac{g}{k_B T} + D(\text{copy}) - I(\text{copy}, \text{template}) \right] \geq 0$$

Powerful method to solve the kinetic equations: 10^5 - 10^6 faster than Monte Carlo simulations.

$$\frac{d}{dt} P_t \left(\begin{matrix} m_1 \cdots m_l \\ n_1 \cdots n_l n_{l+1} \cdots \end{matrix} \right) = W_{+m_l m_{l-1}, l} P_t \left(\begin{matrix} m_1 \cdots m_{l-1} \\ n_1 \cdots n_{l-1} n_l \cdots \end{matrix} \right) + \sum_{m_{l+1}} W_{-m_{l+1} m_l, l+1} P_t \left(\begin{matrix} m_1 \cdots m_l m_{l+1} \\ n_1 \cdots n_l n_{l+1} \cdots \end{matrix} \right) - \left(W_{-m_l m_{l-1}, l} + \sum_{m_{l+1}} W_{+m_{l+1} m_l, l+1} \right) P_t \left(\begin{matrix} m_1 \cdots m_l \\ n_1 \cdots n_l n_{l+1} \cdots \end{matrix} \right)$$

iterated function system:
for the local partial velocities

$$v_{m_{l-1}, l-1} = \sum_{m_l} \frac{W_{+m_l m_{l-1}, l}}{W_{-m_l m_{l-1}, l} + v_{m_l, l}} v_{m_l, l} \quad \text{rates: } W_{\pm m_l m_{l-1}, l}$$

- exact solutions of kinetic equations for DNA replication, transcription, and translation
- mean growth velocity, error probabilities, effects of *sequence heterogeneity*

Perspectives to understand dynamical aspects of information in molecular biology:

- Fundamental thermodynamic link between *metabolism* and *self-replication*.
- *Mutation rates* in DNA replication without and with proofreading.
- New light on the *molecular mechanisms* underlying *biological evolution*.