Stochastic Transcription Elongation via Rule Based Modelling

Masahiro HAMANO

hamano@jaist.ac.jp

SASB, Saint Malo 2015

Purpose of This Talk

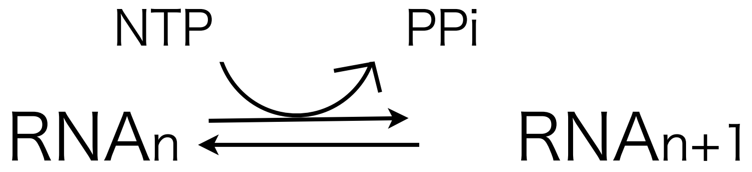
Rule based investigation for

- Equilibrium Kinetics and Steady State Dynamics of Transcription Elongation such as $\left\{\begin{array}{l}\text{detailed balance (reversibility)}\\ \text{Wegscheider condition}\end{array}\right.$

- Michaelis-Menten Enzyme Kinetics via Probability/Energy of Boltzmann Distribution

mechano-chemical TE as Rule Based Modelling

- TE is a stochastic mechano-chemical interaction consisting of
  -Brownian ratchet mechanism
  -Chemical Reactions :NTP hydrolysis/catalysis and PPi release



- This mechano-chemical principle is represented by rule based κ-calculus syntactically and semantically

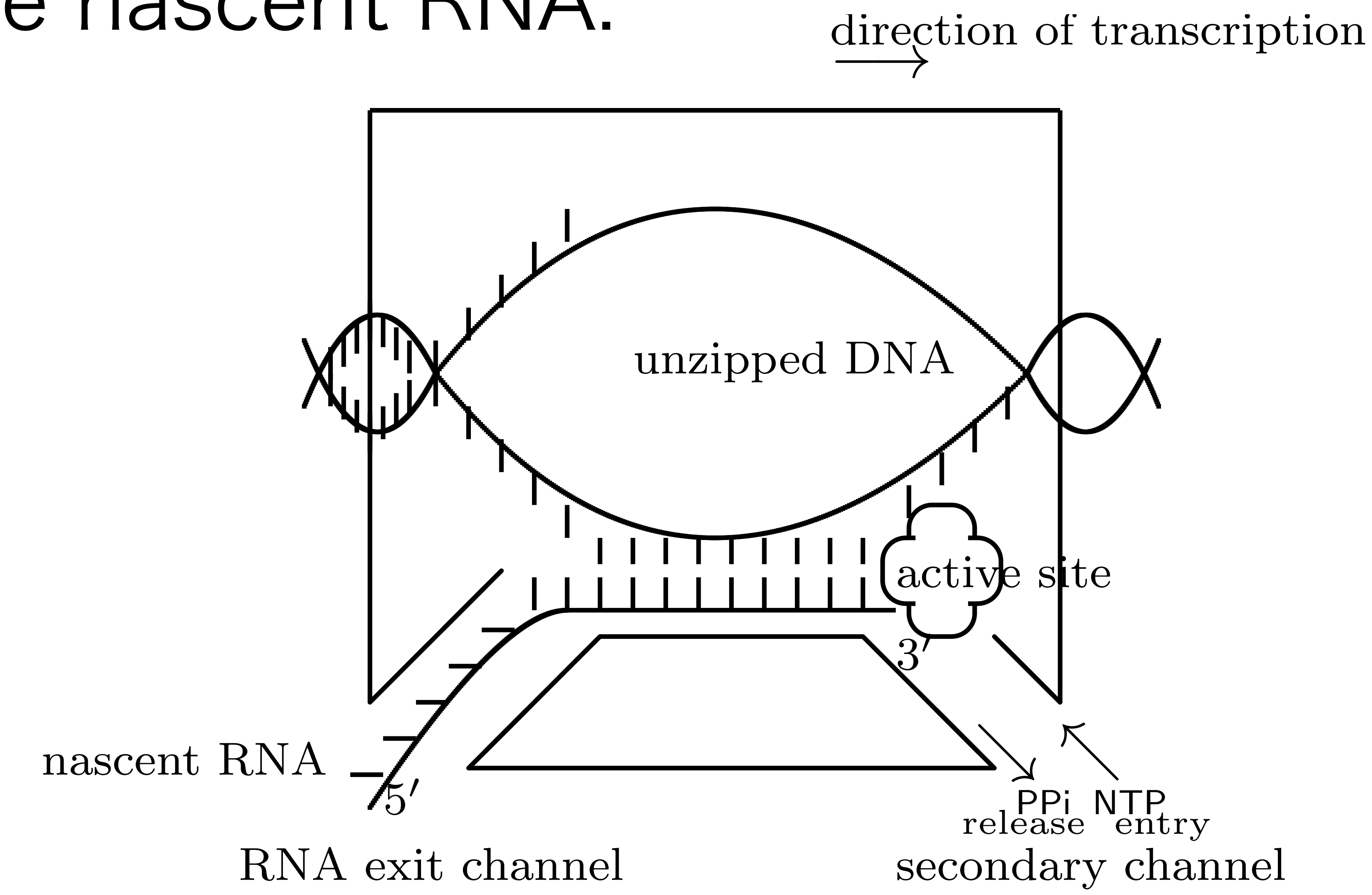
Table of Contents

Table of Contents

- Transcription Elongation Complex (TEC)
- TE is described by Rule Based κ in terms of TEC's compact active sites
- Master Equation Semantics and Abstraction
- Michaelis-Menten Kinetics for TE via Boltzmann distribution (energy/probability)

Transcription Elongation (TE)

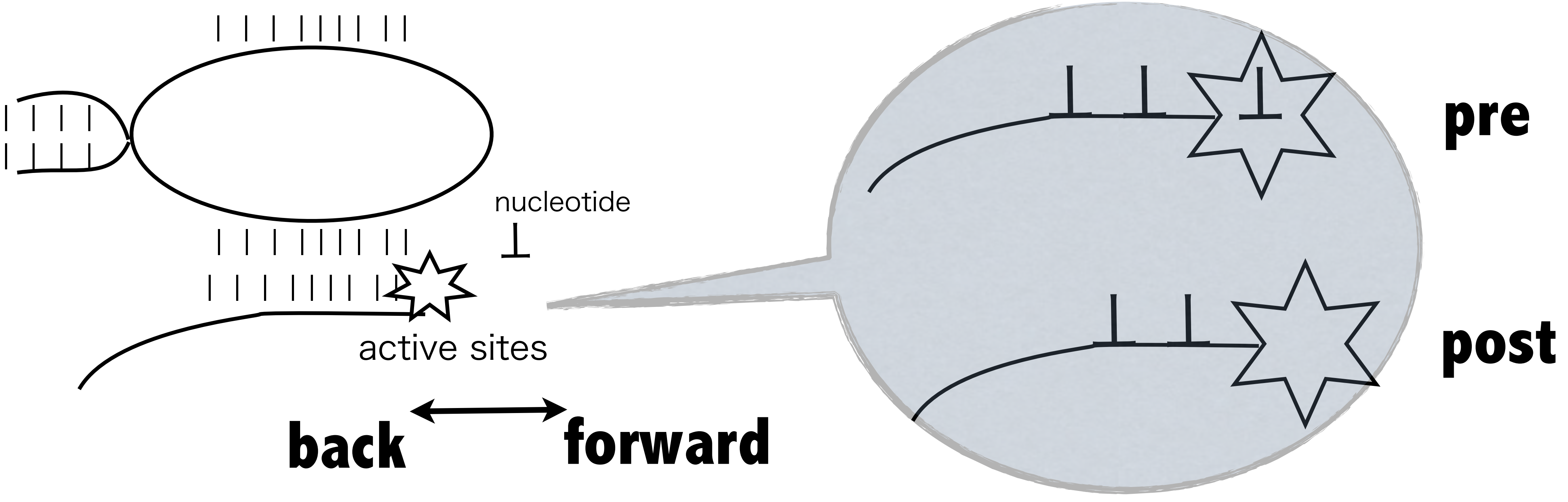
- RNAp moves along the template DNA
- (polymerization) NTP unit is added to the (3'-end of) the nascent RNA.



Structure of Transcription Elongation Complex (TEC)

Translocations of TEC (terminology)

- **Back- and forward-track of TEC**
  (The active sites of)TEC moves back (toward 3‘end of DNA) and forth (toward 5’-end) along the DNA template without polymerzing and depolymerizing the transcript.
- **Pre- and post-translocation of TEC**
  pre-translocation = the state immediately after the polymerization
  post-translocation = one forward track step from pre-translocation so that the active site is in position to catalyze the next nucleotide.



Brownian Ratchet Mechanism of Elongation

$$\begin{tikzcd}
& \mathsf{TEC}_{(n,fwdtrack)} \arrow[d, shift left] & & \\
\mathsf{TEC}_{(n,pre)} \arrow[r, shift left, "k_1"] \arrow[d, shift left] & \mathsf{TEC}_{(n,post)} \arrow[l, shift left, "k_{-1}"] \arrow[u, shift left] \arrow[r, shift left, "\mathsf{NTP}\ k_2"] & \mathsf{TEC}_{(n,post)} \cdot \mathsf{NTP} \arrow[l, shift left, "k_{-2}"] \arrow[r, shift left, "\mathsf{PPi}"] & \mathsf{TEC}_{(n+1,pre)} \arrow[l, shift left] \\
\mathsf{TEC}_{(n,backtrack)} \arrow[u, shift left] & & &
\end{tikzcd}$$

k_{+-3} >> k_{+-4}

$$\mathsf{TEC}_{(n,post)} \cdot \mathsf{NTP} \xrightleftharpoons[k_{-3}]{k_3} \mathsf{TEC}_{(n+1,pre)} \cdot \mathsf{PPi} \xrightleftharpoons[k_{-4}]{k_4\ \mathsf{PPi}} \mathsf{TEC}_{(n+1,pre)}$$

Bai et al, Sequence-dependent kinetic model for transcription elongation by RNA polymerase, J Mol Biol. (2004)

Table of Contents

- Transcription Elongation Complex (TEC)
- Rule Based Description in Terms of TEC's Compact Active Sites
- Master Equation Semantics and Abstraction
- Michaelis-Menten Kinetics for TE via Boltzmann distribution (energy/probability)

Stochastic Process Calculi and their Markov Semantics

Stochastic π (Priami, Regev, et al)

Rule based κ (Danos, Feret, Harmer, Krivine, et al)

Each channel (site) is equipped with an interaction rate of exponential time distribution.

Their Semantical Counterpart of
Chemical Master Equations and Markov processes

U

Stochastic Petri Nets

synchronization
blocking (inhibition)

(e.g., mass action semantics for chemical reactions among species by Danos-Oury)

U

Theory of Waiting Queues

composable but not enough descriptrive !
~~splitting of customers~~

(e.g., Poisson time consuming for customers waiting line)

Stochastic Process Calculi Provide Discrete/Stochastic Modelling in Biology

Multi-step but non-uniform molecular interaction for

- stochastic fluctuation
- non Poisson time consuming
- decay of particular events

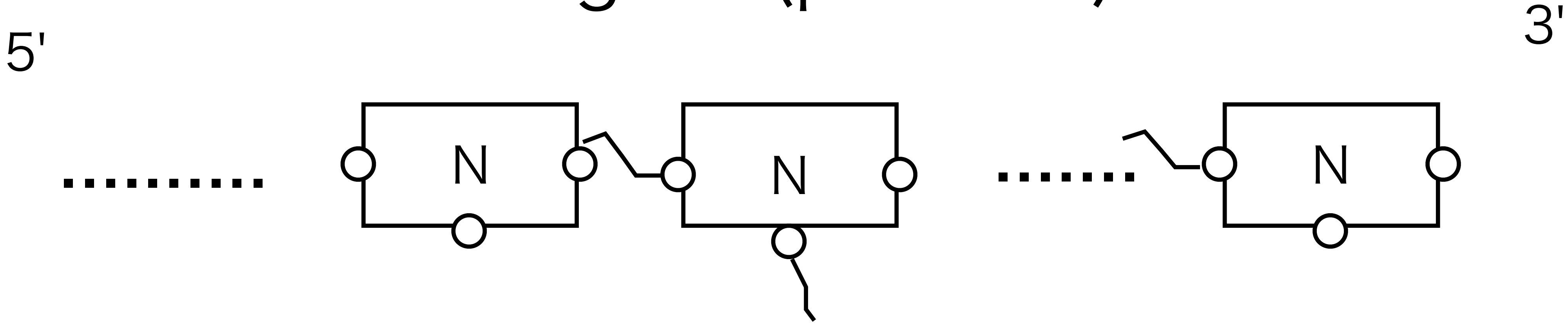
In this talk,

Transcription

- Initiation
- Elongation
- Termination

Single-nucleotide-level description

1 nucleotide as an agent (process)



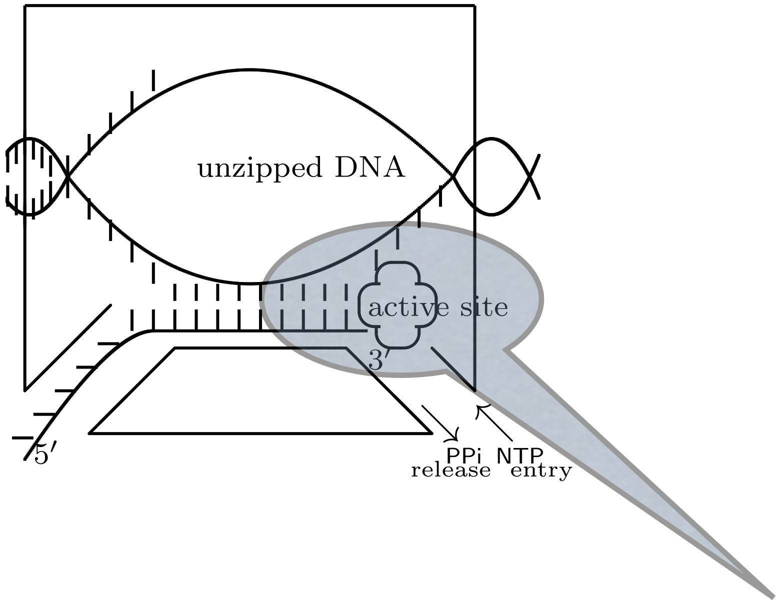
E.g., my preceding paper in SASB'12

RNAi was represented by κ:

agent = single nucleotide N with phosphate and hydrogen bondings

The description yields semantically multitype branching processes determining (non)extinction of certain classes of agents

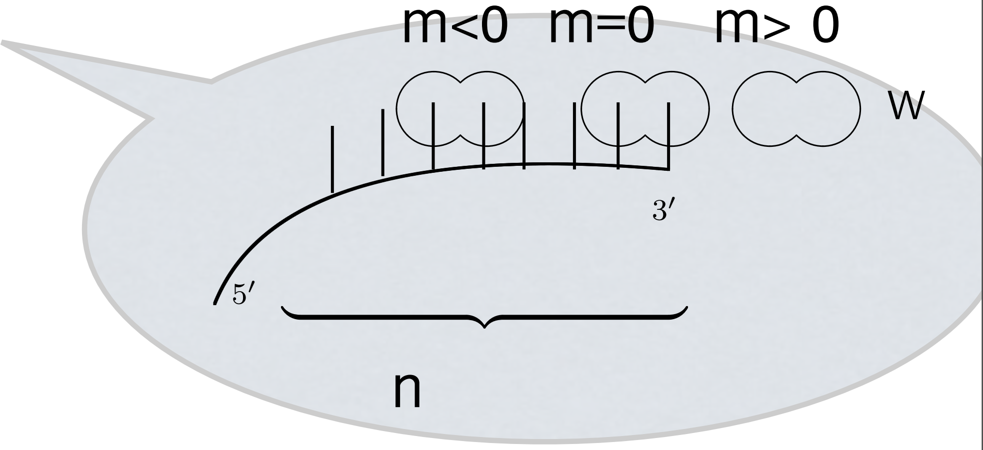
Nucleotide and TEC's Muti-Partite Active Sites as Agents

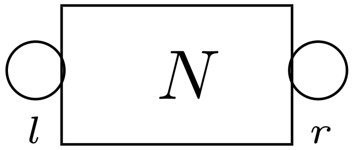
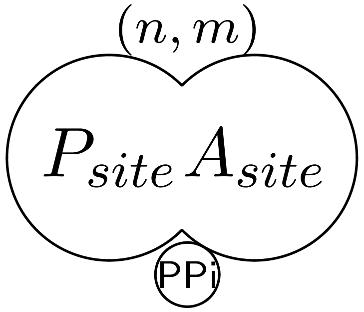


nucleotide N

window frame W for TEC m-p act sts

W with counters (n,m)

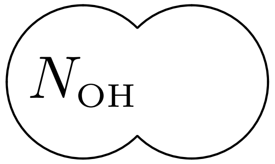


$N(\ell, r)$	W	W_n^m
		

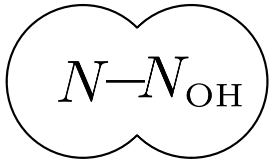
Cf. Landick, The regulatory roles and mechanism of transcriptional pausing, Biochem Society Transactions (2006)

This yields discriminating:

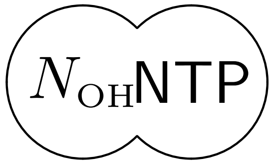
post-translocation: $\mathsf{W}\,[N(r), \emptyset]$ schematically



pre-translocation: $\mathsf{W}\left[N(r^1), N(\ell^1, r)\right]$ schematically



loading: $\mathsf{W}\,[N(r), \mathsf{NTP}]$ schematically



The ratchet pathway as Rules I

Finest grained rules $\mathsf{Elong}^{\mathbb{Z}}_{\mathbb{N}}$

$$\begin{tikzcd}
 & \vdots \arrow[d, leftrightarrow] & & \\
 & \mathsf{W}^{3}_{n}\,[\emptyset,\emptyset] \arrow[d, leftrightarrow] & & \\
 & \mathsf{W}^{2}_{n}\,[\emptyset,\emptyset] \arrow[d, leftrightarrow] & & \mathsf{W}_{n+1}\left[N(r^1),N(\ell^1,r)\right] \arrow[d, leftrightarrow] \\
\mathsf{W}_{n}\left[N(r^1),N(\ell^1,r)\right] \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \mathsf{W}_{n}\,[N(r),\emptyset] \arrow[r, leftrightarrow] & \mathsf{W}_{n+1}\,[N(r),\mathsf{NTP}] \arrow[r, leftrightarrow] & \mathsf{W}^{\mathsf{PPi}}_{n+1}\left[N(r^1),N(\ell^1,r)\right] \\
\mathsf{W}^{-1}_{n}\left[N(r^1),N(\ell^1,r^1)\right] \arrow[d, leftrightarrow] & & & \\
\mathsf{W}^{-2}_{n}\left[N(r^1),N(\ell^1,r^1)\right] \arrow[d, leftrightarrow] & & & \\
\vdots & & &
\end{tikzcd}$$

The ratchet pathway as Rules II:
To coaser/coarsets grained rules

Accordingly to forgetting the sites m and n of W

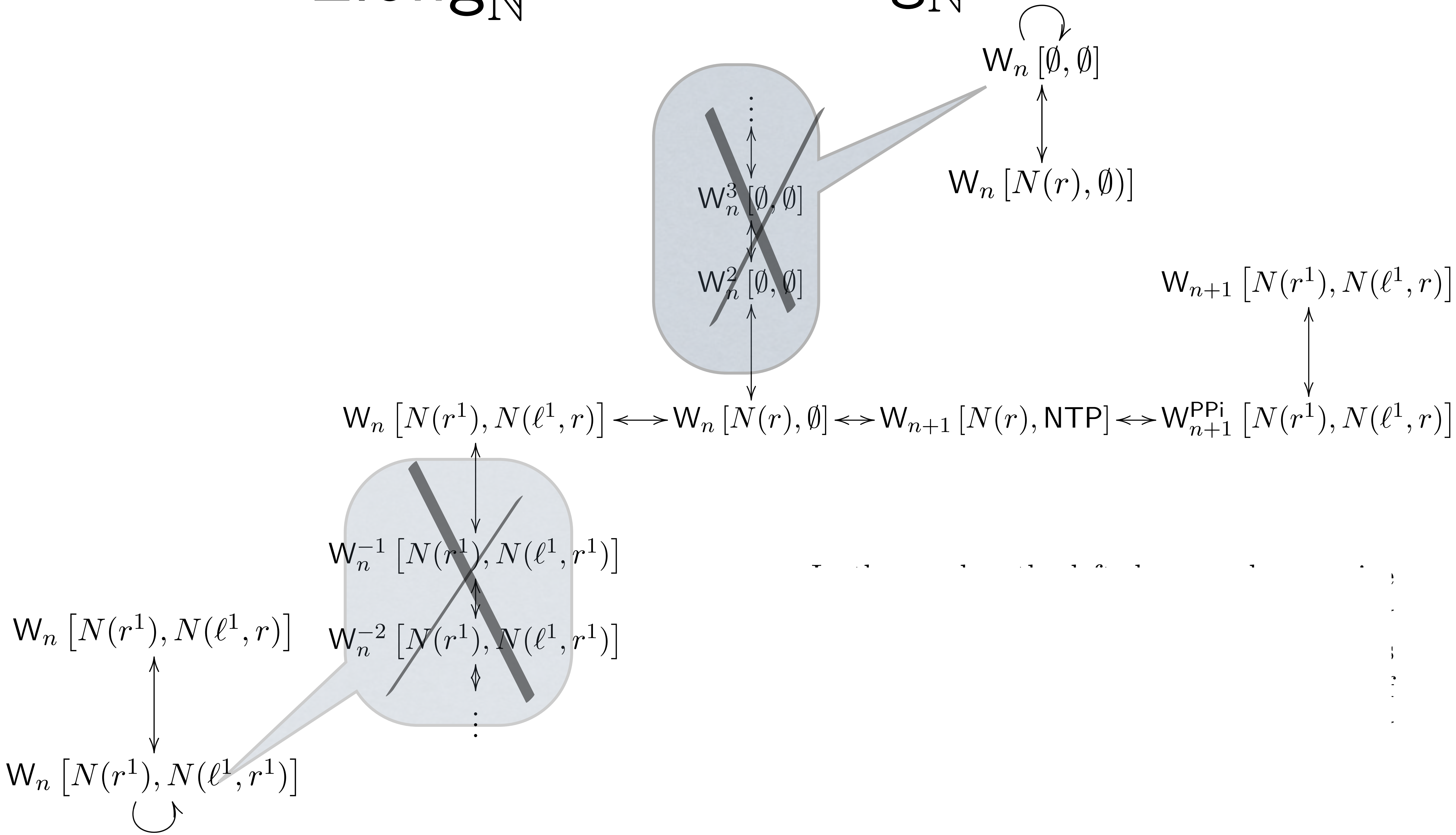
$$W^{m}_{n} \longrightarrow W_{n} \longrightarrow W,$$

rules are abstracted (collapsed) into:

$$\mathrm{Elong}^{\mathbb{Z}}_{\mathbb{N}} \longrightarrow \mathrm{Elong}_{\mathbb{N}} \longrightarrow \mathrm{Elong}$$

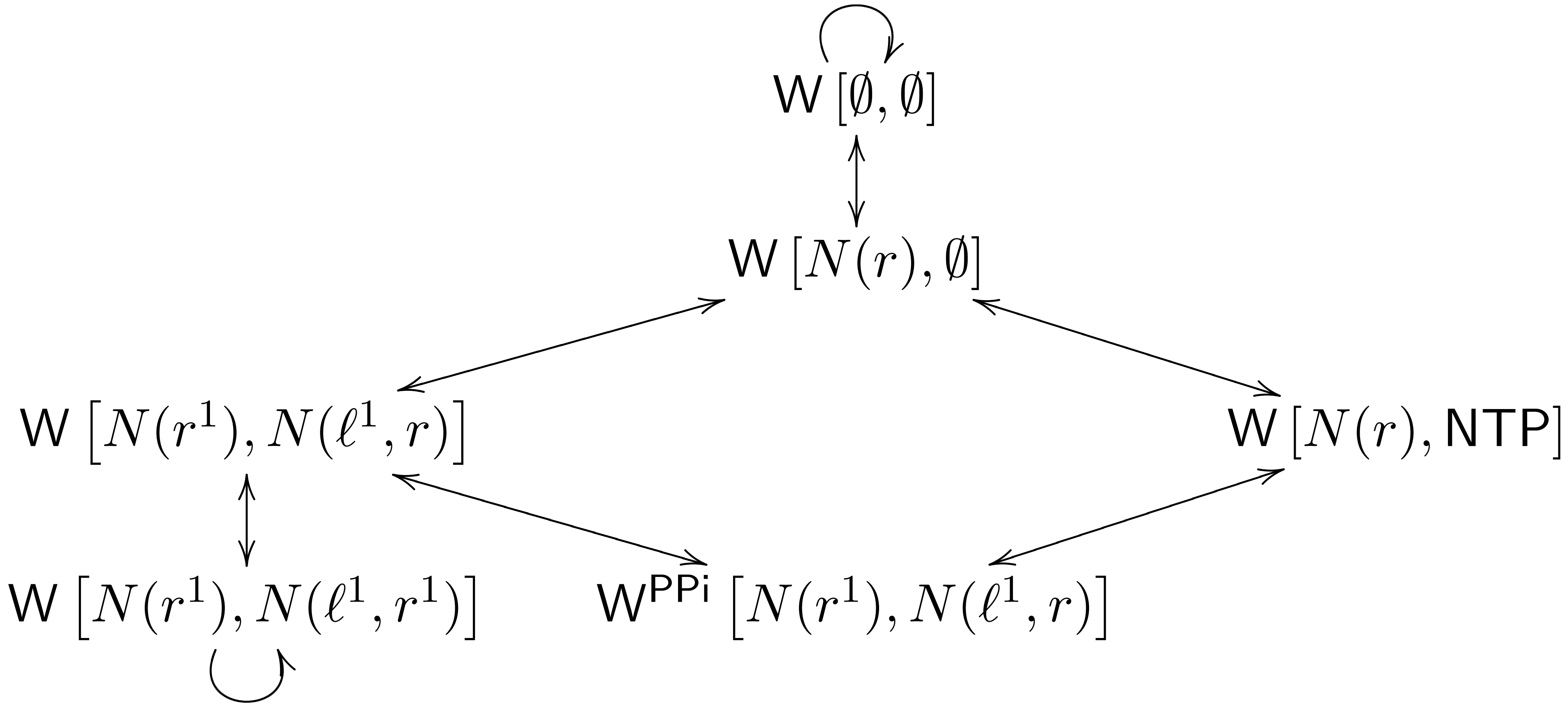
The ratchet pathway as Rules III

$$\text{Elong}^{\mathbb{Z}}_{\mathbb{N}} \longrightarrow \text{Elong}_{\mathbb{N}}$$



The ratchet pathway as Rules IV:

Finally, coarsest grained rules Elong



$\mathsf{W}\,[\emptyset, \emptyset]$

$\mathsf{W}\,[N(r), \emptyset]$

$\mathsf{W}\left[N(r^1), N(\ell^1, r)\right]$

$\mathsf{W}\,[N(r), \mathsf{NTP}]$

$\mathsf{W}\left[N(r^1), N(\ell^1, r^1)\right]$

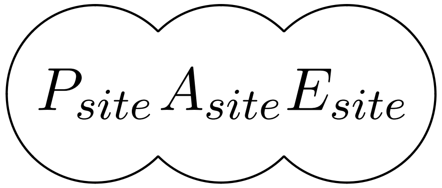
$\mathsf{W}^{\mathsf{PPi}}\left[N(r^1), N(\ell^1, r)\right]$

The ratchet pathway as Rules V:
Conversely, more complicated pathway by augmenting sites on W

$$\begin{tikzcd}
\mathsf{TEC}_{(n,pre)}\cdot\mathsf{NTP} \arrow[r, shift left, "\kappa_{i_1}"] \arrow[d, shift left] & \mathsf{TEC}_{(n,post)}\cdot\mathsf{NTP} \arrow[l, shift left, "k_{-i_1}"] \arrow[d, shift left] \\
\mathsf{TEC}_{(n,pre)} \arrow[u, shift left, "\mathsf{NTP}"] \arrow[r, shift left, "k_{i_2}"] & \mathsf{TEC}_{(n,post)} \arrow[l, shift left, "k_{-i_2}"] \arrow[u, shift left, "\mathsf{NTP}"']
\end{tikzcd}$$

Bai et al, Kinetic modeling of transcription elongation, In “ RNA polymerases as molecular motors ” in the Royal Society of Chemistry (2009)

In our rule based framework

(Agent)		(Rules)
$\mathsf{W} =$	 $P_{site}A_{site}E_{site}$	$$\begin{tikzcd} \mathsf{W}_n\left[N(r^1), N(\ell^1, r), \mathsf{NTP}\right] \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \mathsf{W}_n\left[N(r), \mathsf{NTP}, \emptyset\right] \arrow[d, leftrightarrow] \\ \mathsf{W}_n\left[N(r^1), N(\ell^1, r), \emptyset\right] \arrow[r, leftrightarrow] & \mathsf{W}_n\left[N(r), \emptyset, \emptyset\right] \end{tikzcd}$$

- Rule based κ syntax has uniformly derived variation of pathway of TEC.
- How about semantical counterpart ?

Table of Contents

- Transcription Elongation Complex (TEC)
- Rule Based Description in Terms of TEC's Compact Active Sites
- Master Equation Semantics and Abstraction
- Michaelis-Menten Kinetics for TE via Boltzmann distribution (energy/probability)

Markov Process and Stationary Distribution

Transition probability pt(i,j) of Markov process is governed by

(**Kolmogorov fwd equation**) $p_t' = p_t Q$ with an infinitesimal matrix Q.

- **stationary distribution**

π is a solution of

(invariance)

$$\pi\, p_t = \pi \qquad \text{for all } t$$

$$\Longleftrightarrow \quad \pi Q = 0$$

i.e.,

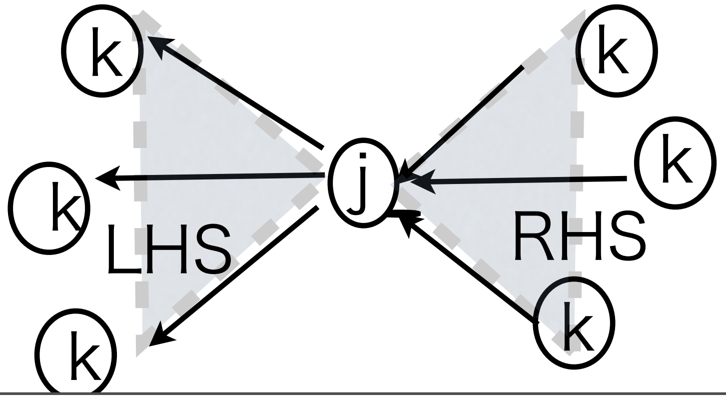
$$\pi(j) \sum_k q(j,k) = \sum_k \pi(k)\, q(k,j)$$

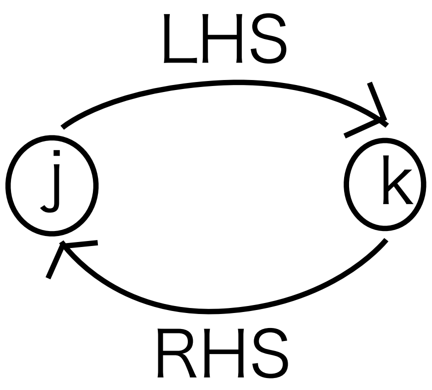
⇐

- **detailed balance,** i.e.,
- **reversibility** (invariance under time-reversing)

$$\pi(j)\, q(j,k) = \pi(k)\, q(k,j)$$

rate of leaving j = rate of arriving j

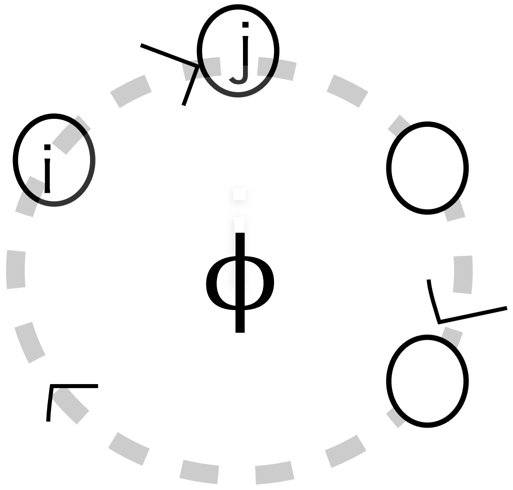




Wegscheider Condition Is Chemist's Algorithm to Check Reversibility

For every cycle path ϕ

$$\prod_{(i,j)\in\phi} \frac{q(i,j)}{q(j,i)} = 1$$



rewritten by

$$\sum_{(i,j)\in\phi} \log \frac{q(i,j)}{q(j,i)} = 0$$

cf. Danos-Oury's undecidability of the W condition

Chemical Master Equation (CME) : Kolmogorov Eq Arisen from Chemical Reactions

CME

arriving **x** by the reaction k

leaving **x** by the k

$$\frac{d}{dt}p_t(\mathbf{x}) = \sum_k \{\lambda_k(\mathbf{x} - (\boldsymbol{\nu}'_k - \boldsymbol{\nu}_k))\, p_t(\mathbf{x} - (\boldsymbol{\nu}'_k - \boldsymbol{\nu}_k)) - \lambda_k(\mathbf{x})\, p_t(\mathbf{x})\}$$

$(\nu'_K - \nu_K)_k$ stoichiometry (state-change) matrix

λ_K propensity function

Though analytically unsolvable CME

Behaviour of average yields a continuous model

$$< \mathbf{x}_i(t) >= \sum_{\mathbf{x}} \mathbf{x}_i p_t(\mathbf{x})$$

i for a chemical specie

$$\frac{d}{dt} < \mathbf{x}_i(t) >= \sum_k (\boldsymbol{\nu}'_{ik} - \boldsymbol{\nu}_{ik}) < \lambda_k(\mathbf{x}(t)) >$$

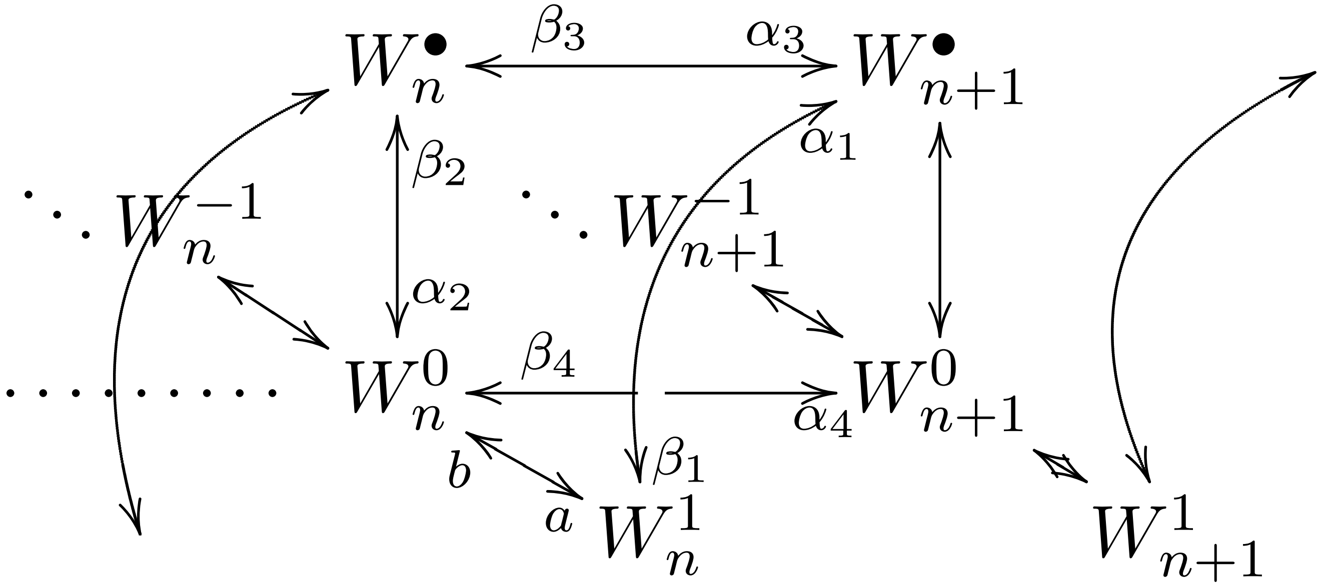
$$\Big\Downarrow$$

no fluctuation hypothesis

reaction rate equation

$$= \sum_k (\boldsymbol{\nu}'_{ik} - \boldsymbol{\nu}_{ik}) \lambda_k(\mathbf{x}(t))$$

State Transition for $\mathrm{Elong}_{\mathbb{N}}^{\mathbb{Z}}$



Evolution on Main Pathway

$$\frac{d}{dt}w_n^{[0]} = \left(aw_n^{[-1]} + \alpha_4 w_{n-1}^{[0]} + \beta_4 w_{n+1}^{[0]} + \alpha_2 w_n^{[\bullet]} + bw_n^{[1]}\right) - (\alpha_4 + \beta_2 + a + b + \beta_4)w_n^{[0]}$$

$$\frac{d}{dt}w_n^{[\bullet]} = \left(\alpha_3 w_{n-1}^{[\bullet]} + \beta_3 w_{n+1}^{[\bullet]} + \beta_2 w_n^{[0]} + \alpha_1 w_{n-1}^{[1]}\right) - (\alpha_3 + \beta_3 + \alpha_2 + \beta_1)w_n^{[\bullet]}$$

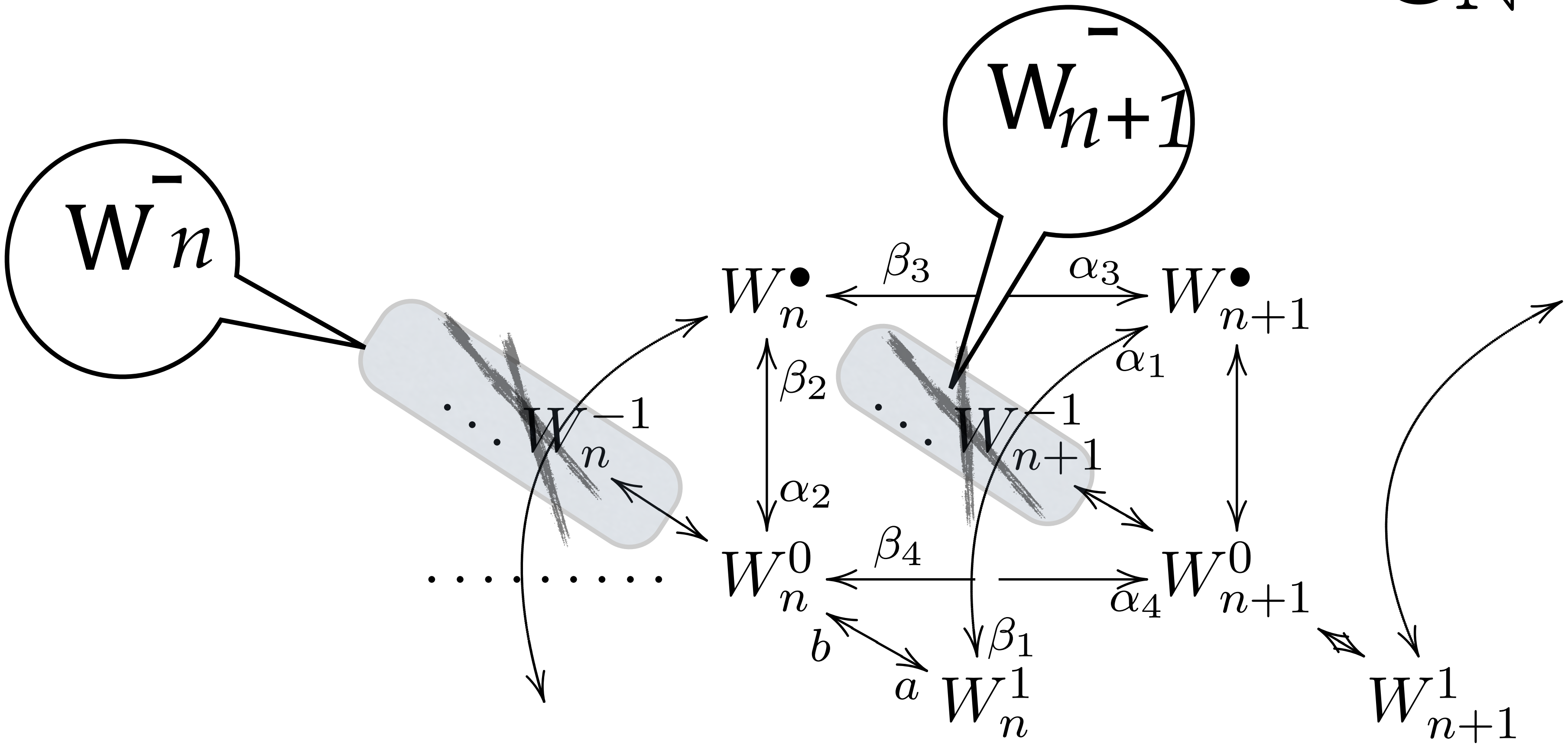
$$\frac{d}{dt}w_n^{[1]} = \left(\beta_1 w_{n+1}^{[\bullet]} + aw_n^{[0]}\right) - (b + \alpha_1)w_n^{[1]}$$

Evolution on Back-Track Pathway

$$\frac{d}{dt}w_n^{[-(n-1)]} = bw_n^{[-(n-2)]} - aw_n^{[-(n-1)]}$$

$$\frac{d}{dt}w_n^{[-j]} = \left(aw_n^{[-(j+1)]} + bw_n^{[-(j-1)]}\right) - (a + b)w_n^{[-j]} \qquad \text{for } 1 \leq j < n-1$$

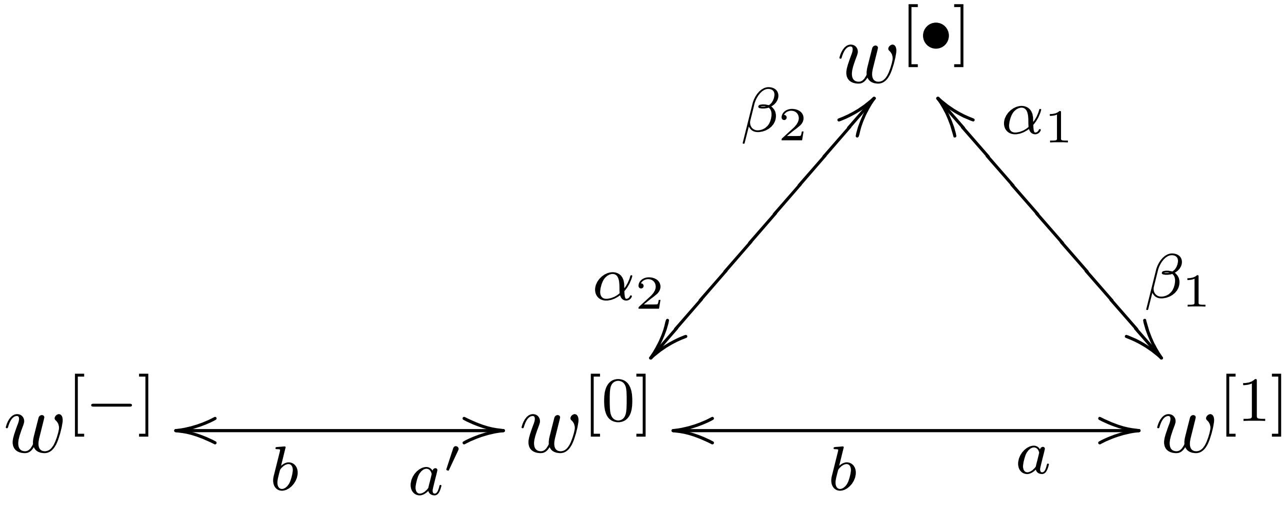
State Transition for $\mathrm{Elong}_{\mathbb{N}}$



Evolution on back-track is reduced by summing over the positions j's

$$w_n^{[-]} = \sum_{1\le j\le n-1} w_n^{[-j]} \quad \text{and} \quad \frac{d}{dt}w_n^{[-]} = \sum_{1\le j\le n-1} \frac{d}{dt}w_n^{[-j]} = bw_n^{[0]} - aw_n^{[-1]}$$

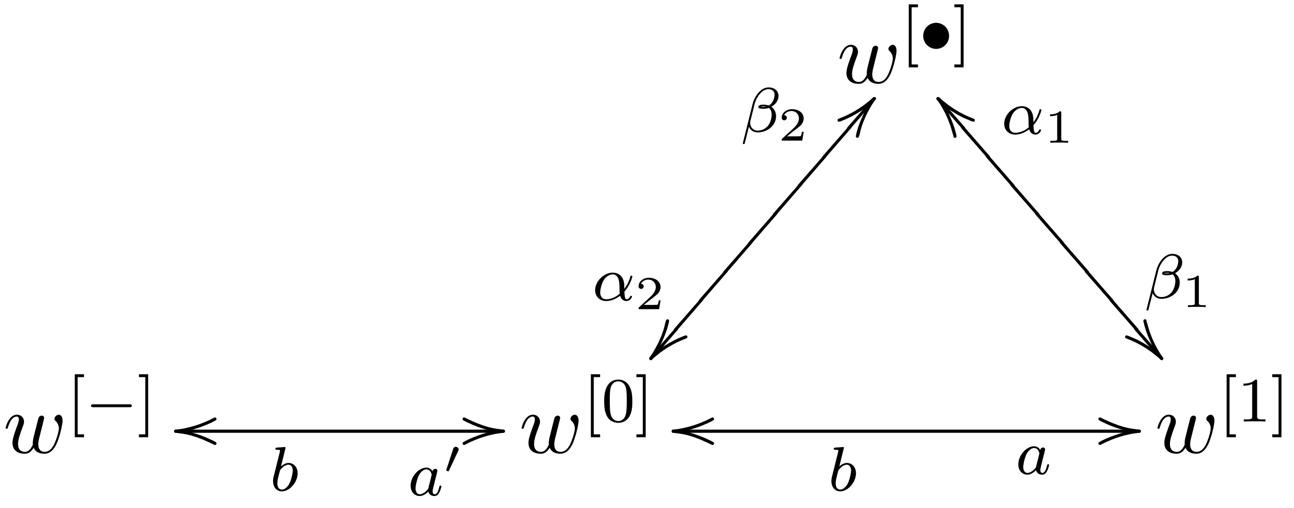
Finally, State Transition for Elong



Evolution on main-track is reduced by summing over the lengths n's

$$
\begin{aligned}
\frac{d}{dt}w^{[0]} &= aw^{[-1]} + \alpha_2 w^{[\bullet]} + bw^{[1]} - (\beta_2 + a + b)w^{[0]} \\
\frac{d}{dt}w^{[\bullet]} &= \qquad \beta_2 w^{[0]} + \alpha_1 w^{[1]} - (\alpha_2 + \beta_1)w^{[\bullet]} \\
\frac{d}{dt}w^{[1]} &= \qquad \beta_1 w^{[\bullet]} + aw^{[0]} - (b + \alpha_1)w^{[1]}
\end{aligned}
$$

Equilibrium Distribution of Elong



Wegscheider condition:

$$(a/b)\,(\alpha_1/\beta_1)\,(\alpha_2/\beta_2) = 1$$

so that equilibrium is given with a relax time τ

$$w_*^{[0]} = b\beta_1\tau \qquad w_*^{[1]} = a\beta_1\tau \qquad w_*^{[\bullet]} = a\alpha_1\tau \qquad \tau = 1/(b\beta_1 + a\beta_1 + a\alpha_1)$$

Quasi-steady state approximation

How sound is the model abstraction arisen from forgetting sites ?

$$\begin{tikzcd}
\llbracket \mathrm{Elong}^{\mathbb{Z}}_{\mathbb{N}} \rrbracket \arrow[r] \arrow[rr, bend left, "\text{model abstraction}"] & \llbracket \mathrm{Elong}_{\mathbb{N}} \rrbracket \arrow[r] & \llbracket \mathrm{Elong} \rrbracket \arrow[ll, bend left, "\text{QSS}"]
\end{tikzcd}$$

The finest grained rules may be approximately retrieved from the coarsest ones under certain biological assumption.

Cf. Voliotis et al, Fluctuations, Pauses, and Backtracking in DNA Transcription, Biophysical Journal (2008)

Quasi-Steady State approximation

Under the assumption that

poly/depoly merization rates << translocation rates,

the translocation rates may be run off in our time scale.

$$w_n^{[j]}(t) \approx w_*^{[j]} \times w_n(t) \qquad \text{where} \quad w_n = \sum_{j \in \{1,0,-,\bullet\}} w_n^{[j]}$$

Summing evolutions on main (res. backtrack) pathways res. of $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$ (res. $\mathsf{Elong}_{\mathbb{N}}$) yields a simple birth-death master equation with the effective rates θ_1/θ_2 for poly/depoly merization

$$\frac{d}{dt} w_n(t) \qquad = \theta_+ w_{n-1} - (\theta_+ + \theta_-) w_n + \theta_- w_{n+1}$$

$$\text{in which} \quad \theta_+ = \alpha_4 b \beta_1 + \alpha_3 a \alpha_1 + \alpha_1 a \beta_1 \qquad \text{and} \quad \theta_- = \beta_4 b \beta_1 + \beta_3 a \alpha_1 + \beta_1 a \alpha_1$$

Table of Contents

- Transcription Elongation Complex (TEC)
- Rule Based Description in Terms of TEC's Compact Active Sites
- Master Equation Semantics and Abstraction
- Michaelis-Menten Kinetics for TE via Boltzmann distribution (energy/probability)

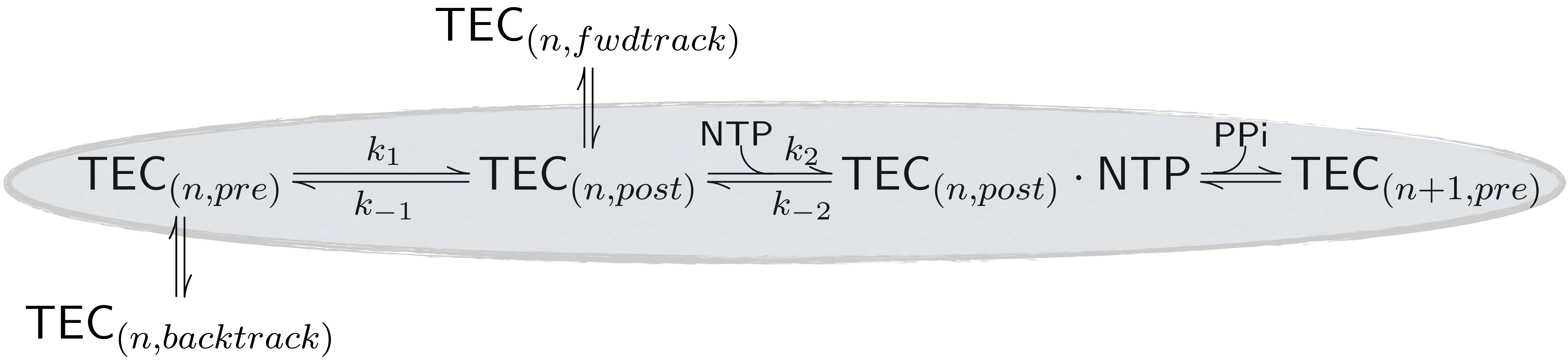
Boltzmann distribution

correspondence of probability π and energy ε

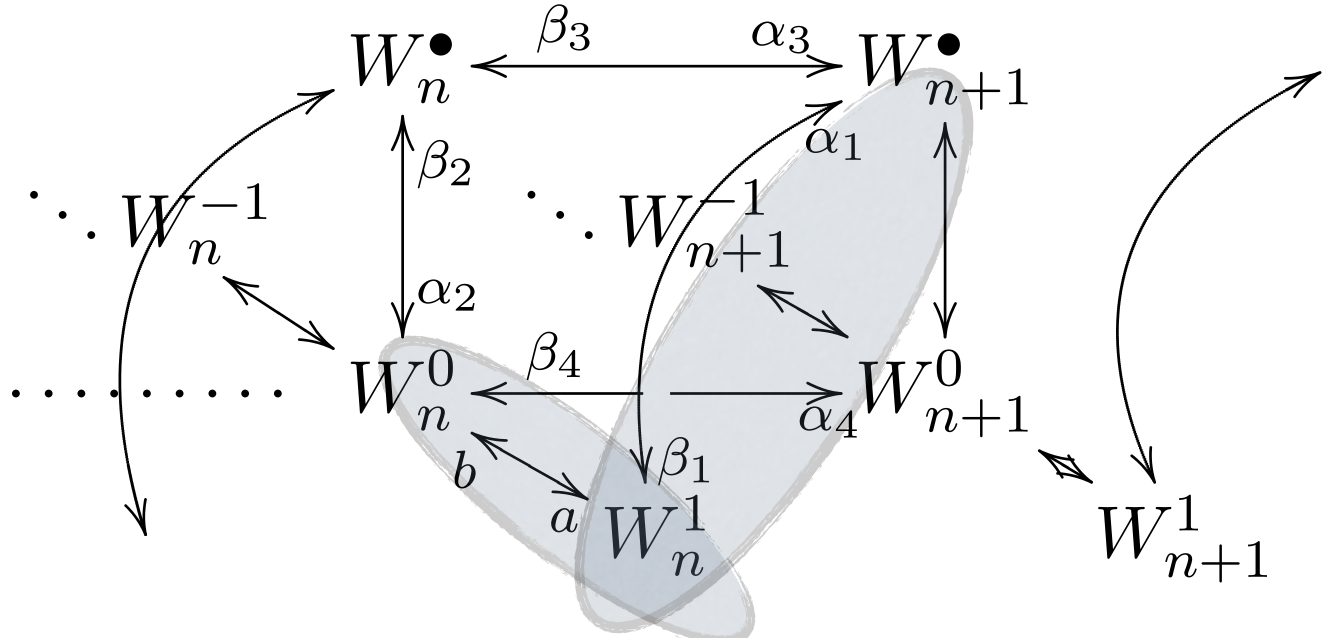
$$w_n^{[j]} = \pi(W_n^j) = \frac{1}{Z_n} \exp\left(\frac{-\mathcal{E}(W_n^j)}{k_B T}\right)$$

$$\text{with} \qquad Z_n = \sum_j \exp\left(\frac{-\mathcal{E}(W_n^j)}{k_B T}\right)$$

The main reaction pathway is the following part



In our transition diagram with a=k$_1$, b=k$_{-1}$ α=k $_2$ β= k$_2$ [NTP]



The generator Q of Kolmogorov back eqn p'(t)=Qp(t) is given for a fixed n;

$$Q = \begin{pmatrix} -(q_{\bullet 1}+q_{01}) & q_{1\bullet} & q_{10} \\ q_{\bullet 1} & -q_{1\bullet} & 0 \\ q_{01} & 0 & -q_{10} \end{pmatrix} \qquad p = (p_i)_{i=1,\bullet,0} \qquad p_1 \text{ is } W_n^i$$

The main reaction pathway as Michaelis-Menten kinetics

M-M kinetics is derivable from the stationary distribution of Kolmogorov

(Kolmogorov backward equation)

$$p'(t) \;=\; Q\,p(t)$$

First, stationary distribution p'(t)=0 yields via $q_{1\bullet} = k_{-2}$ and $q_{\bullet 1} = k_2[\mathsf{NTP}]$

$$p_\bullet = \frac{1}{\frac{q_{1\bullet}}{q_{\bullet 1}}(1+\frac{q_{01}}{q_{10}})+1} = \frac{[\mathsf{NTP}]}{\frac{k_{-2}}{k_2}(1+\frac{q_{01}}{q_{10}})+[\mathsf{NTP}]}$$

Second, the velocity v of TEC(n+1,pre) is given so that $v/v_{max} = p_\bullet$

Michaelis-Menten

$$v = \frac{v_{max}[\mathsf{NTP}]}{\frac{k_{-2}}{k_2}(1+\frac{q_{01}}{q_{10}})+[\mathsf{NTP}]}$$

=

a/b =K_{eq} equilibrium const for translocation

More generally combining M-M and Boltzmann

Assuming bck/fwd translocations are in equilibrium:

Steady-state elongation rate

$$v = \frac{v_{max}[\text{NTP}]}{K_n + [\text{NTP}]} \quad \text{with} \quad K_n = \tfrac{k_{-2}}{k_2} \log \prod_j \frac{w_n^{[j]}}{w_n^{[0]}}$$

$$\textstyle\sum_j \exp\left((\mathcal{E}(W_n^0) - \mathcal{E}(W_n^j))/k_B T\right)$$

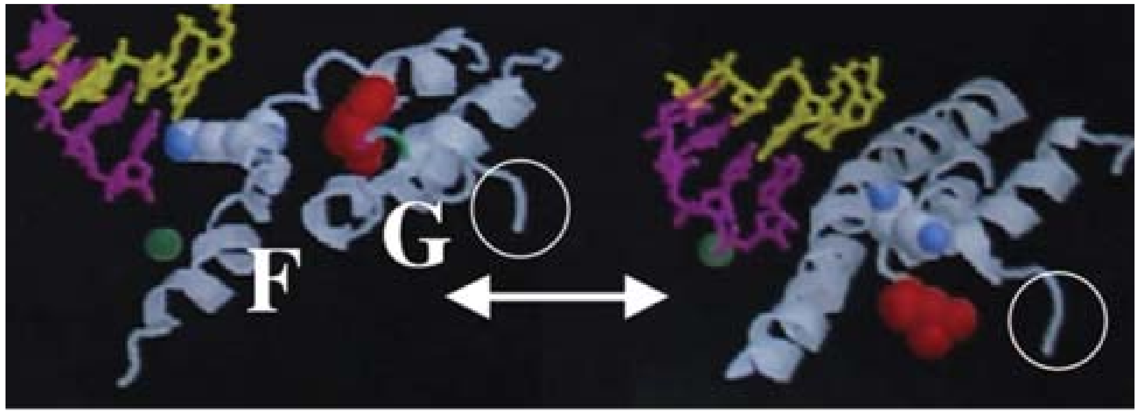
energy difference

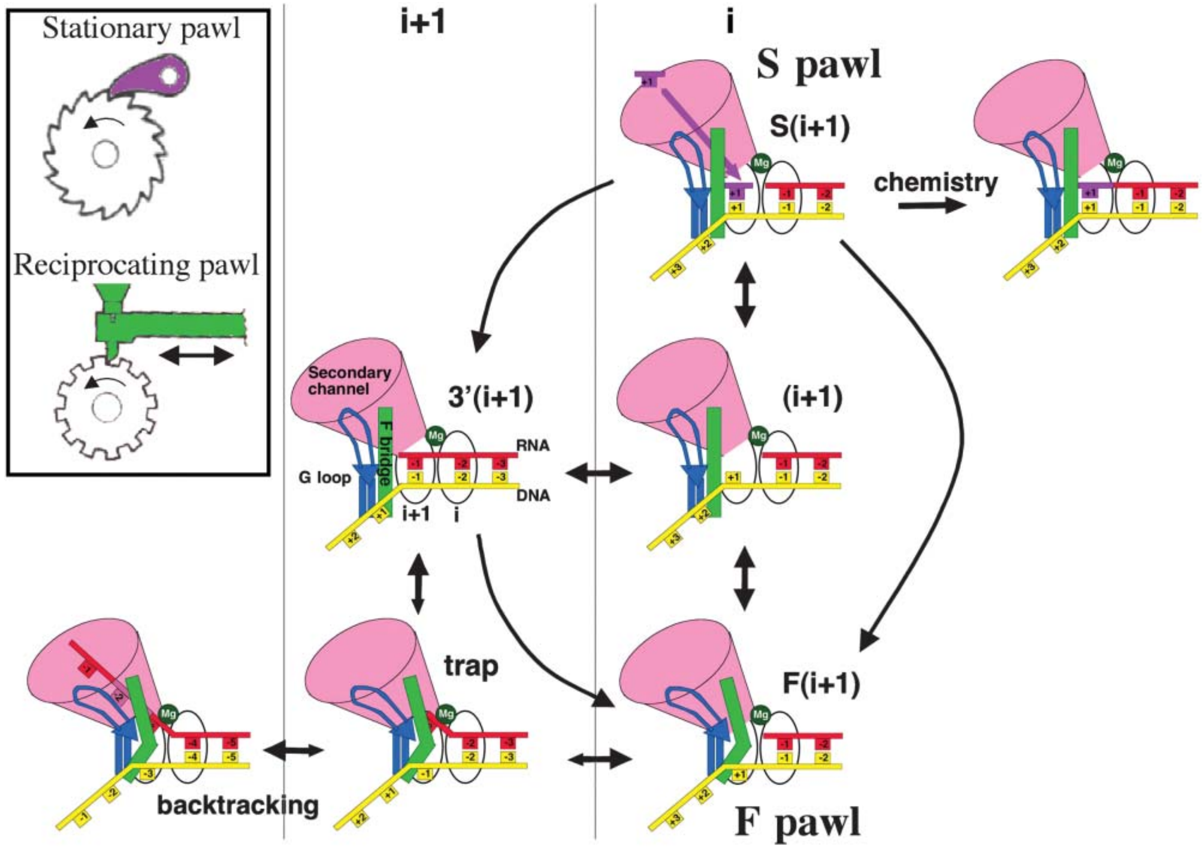
Future Work1

More conformationally faithful details to realize various pawls in TEC

E.g. , G-loop configuration to control bending/straightening F-helix

Bar-Nahum, et al, “A Ratchet Mechanism of Transcription Elongation and its Control”, Cell 120(28) 2005





Future Work 2

RNAp as molecular motor

How to augment drift (thermal force) on top of κ description of Brownian motion ??

probability $p(x,t)$

probability flux given by Fick's law wiith drift;

$$j(x,t) \;=\; -D\frac{\partial p}{\partial x} + \frac{F}{\gamma}p(x,t)$$

diffusion (random motion) + drift (directed motion)

where diffusion coef. D and external force $F(x,t)$ applied to RNAp with γ drag coef.

Fokker-Planck eqn

$$\frac{\partial p}{\partial t} \;=\; -\frac{\partial j}{\partial x} \;=\; D\frac{\partial^2 p}{\partial x^2} + \frac{\partial}{\partial x}(\frac{\partial U}{\partial x}p)$$

1st eqn is flux continuity eqn

(Note: $F = -\frac{\partial U}{\partial x}$ with energy U)

Future Work 2 (contd.)

Why Fokker-Planck ?

At equilibrium of FP-eqn;

$$0 = D\frac{dp}{dx} + \frac{1}{\gamma}\frac{dU}{dx}p$$

Boltzmann distribution is a solution

$$p(x) \quad \propto \quad \exp\left(-\frac{U(x)}{D\gamma}\right)$$

Towards more energy $\mathrm{U}(x)$ efficient descprption of elongation (e.g., chemist's use of energy-landscape)