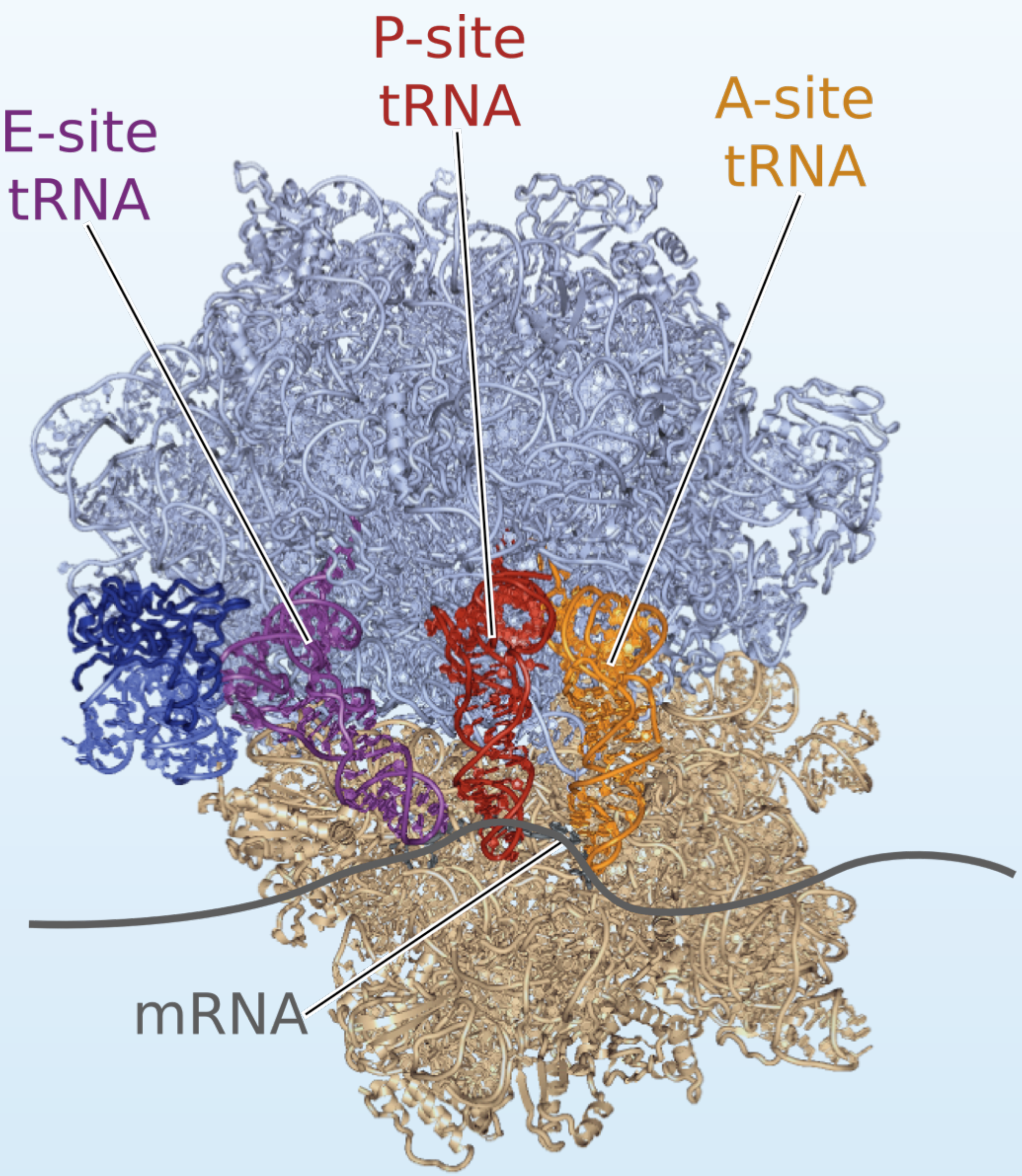
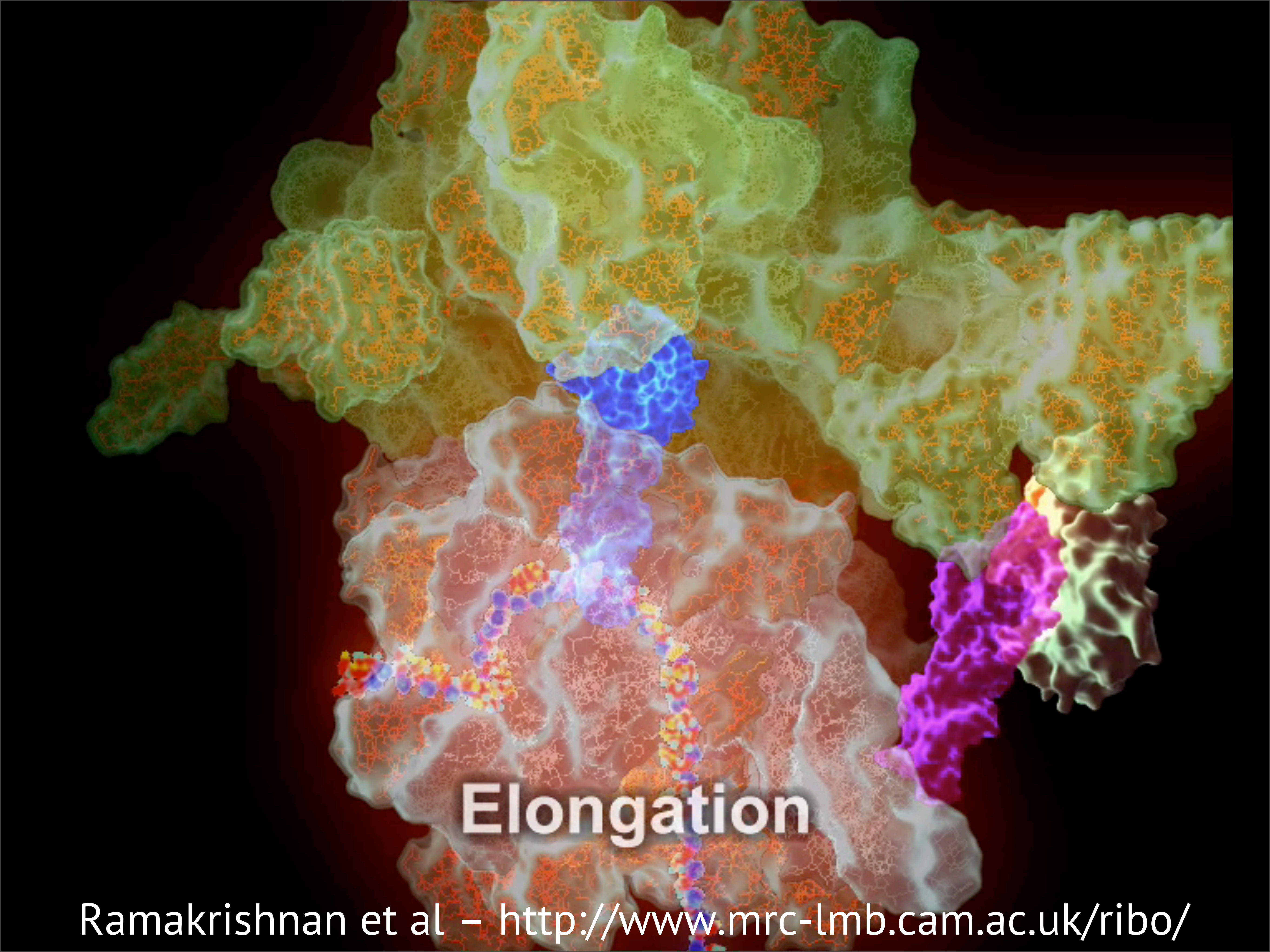
Learning Kinetic Pathways from Single-Molecule FRET Measurements

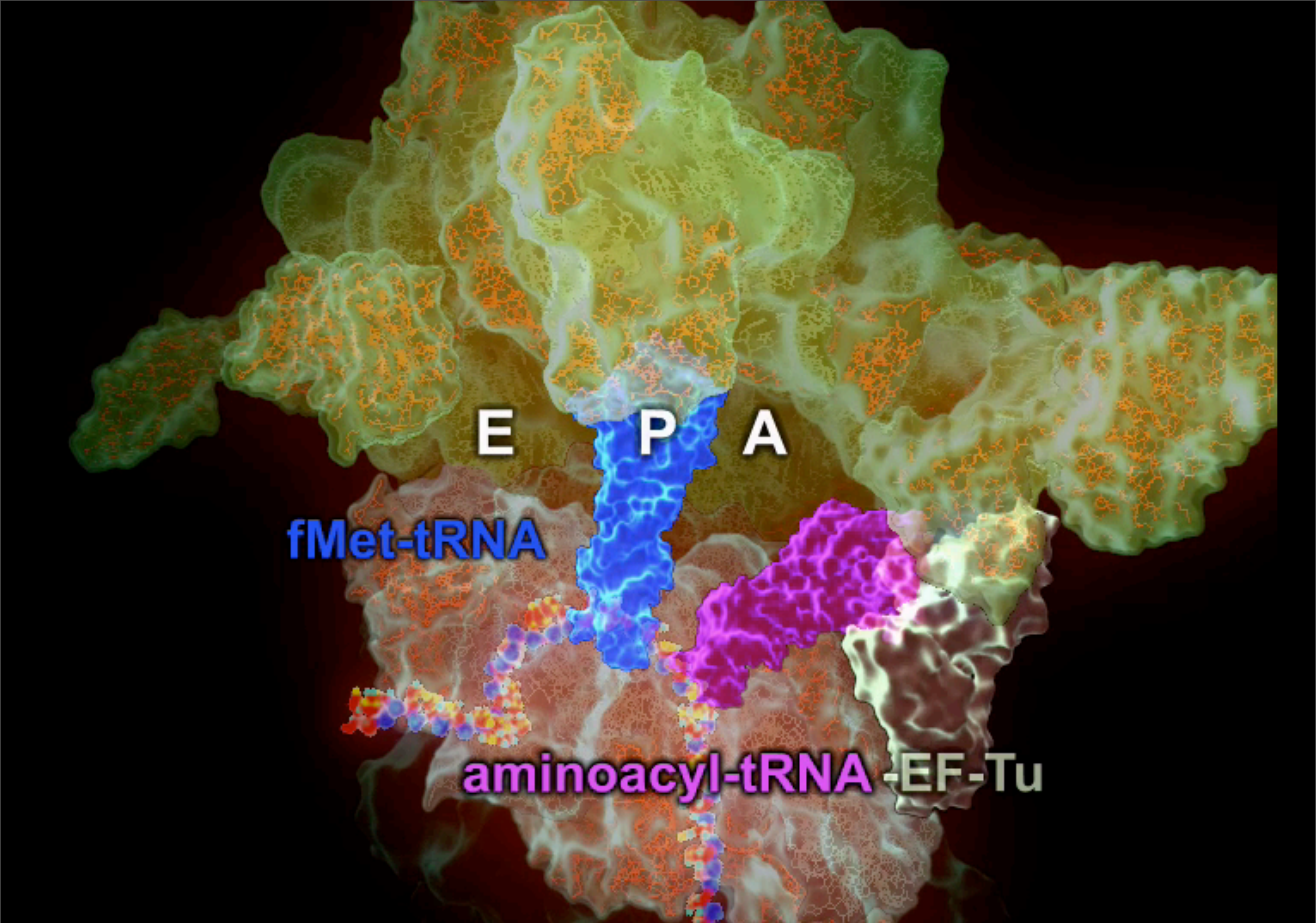


Jan-Willem van de Meent, Ruben Gonzalez, Chris Wiggins
Columbia University

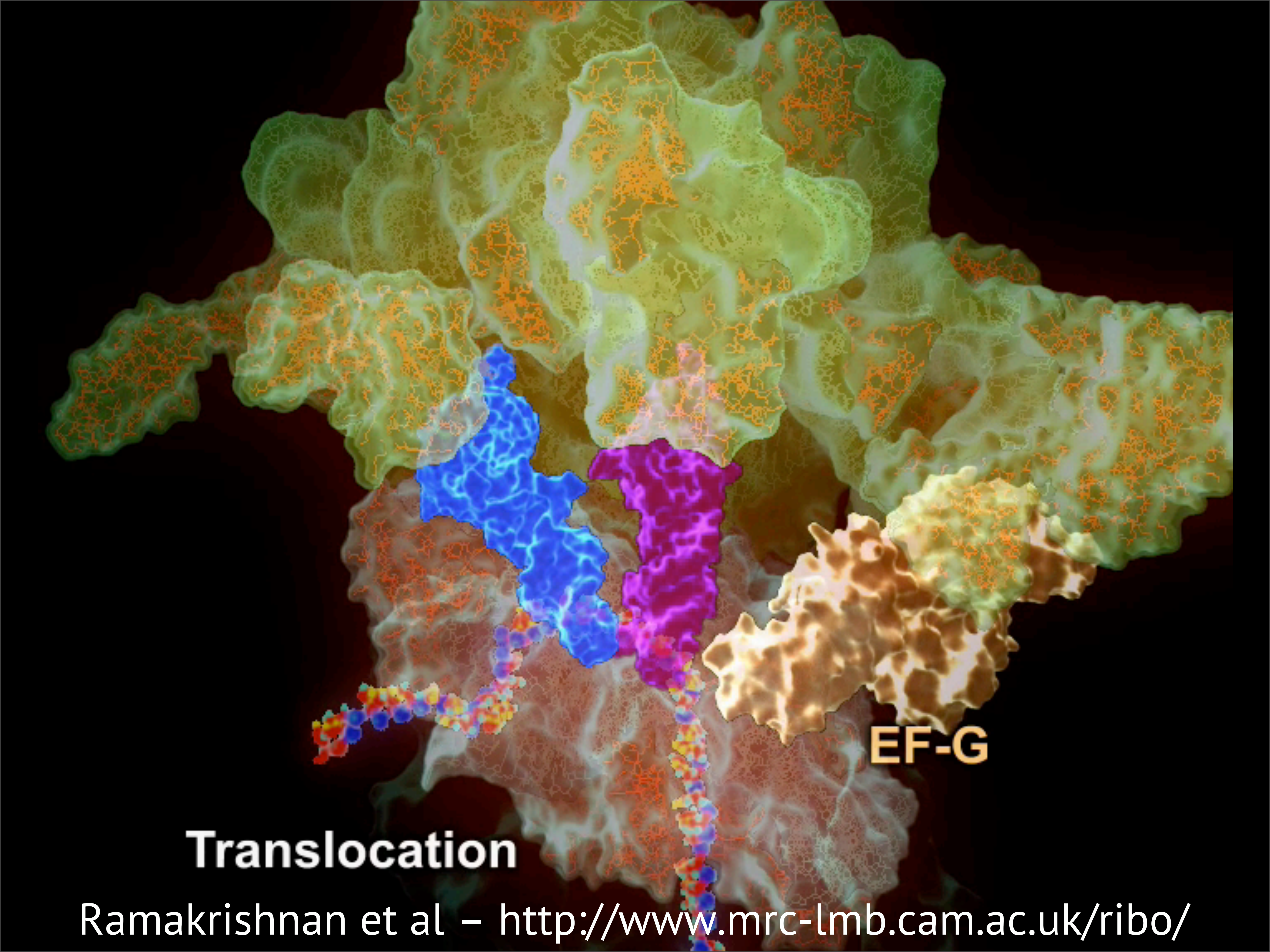
Ramakrishnan et al – http://www.mrc-lmb.cam.ac.uk/ribo/



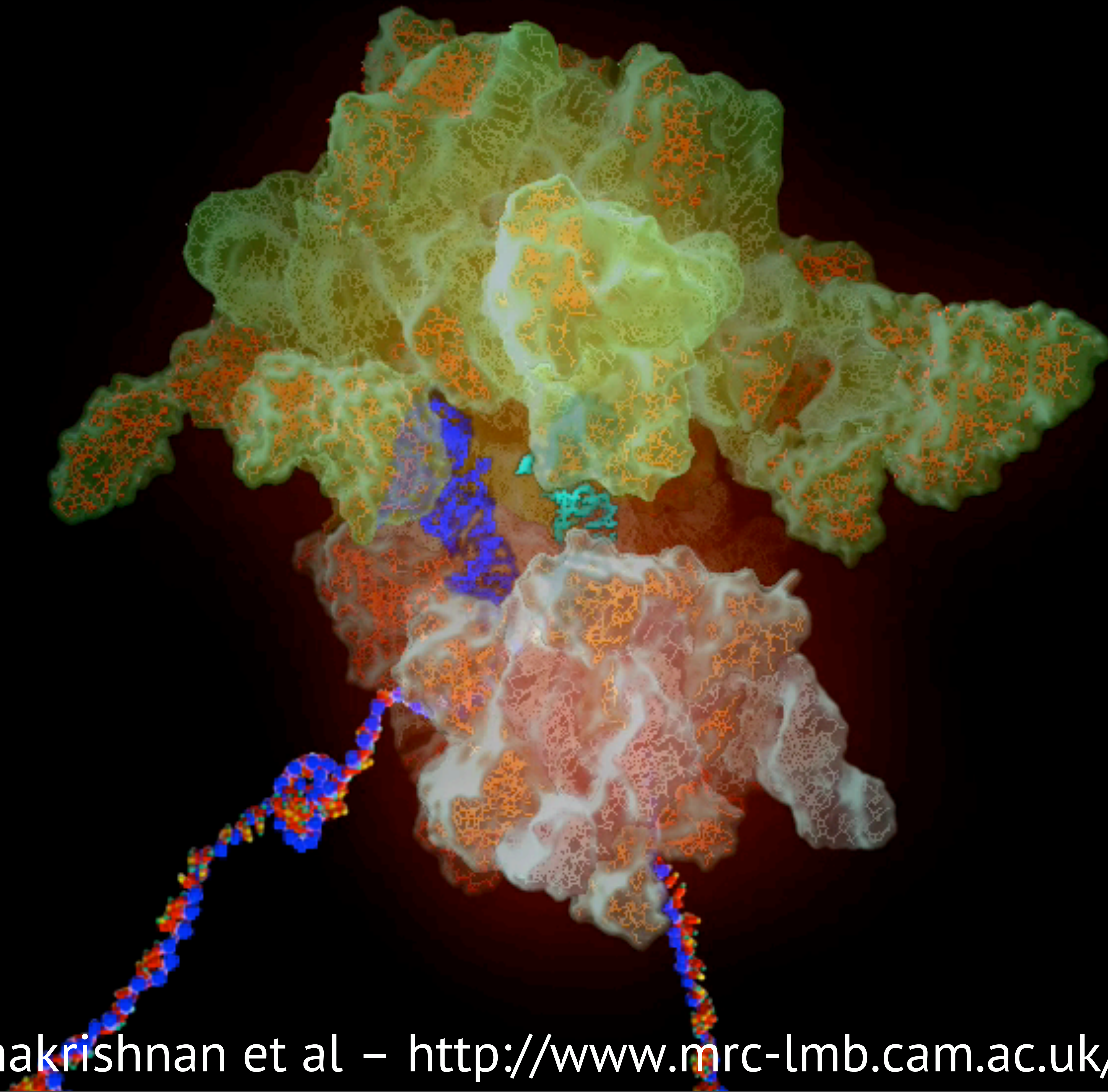
Ramakrishnan et al – http://www.mrc-lmb.cam.ac.uk/ribo/



Ramakrishnan et al – http://www.mrc-lmb.cam.ac.uk/ribo/

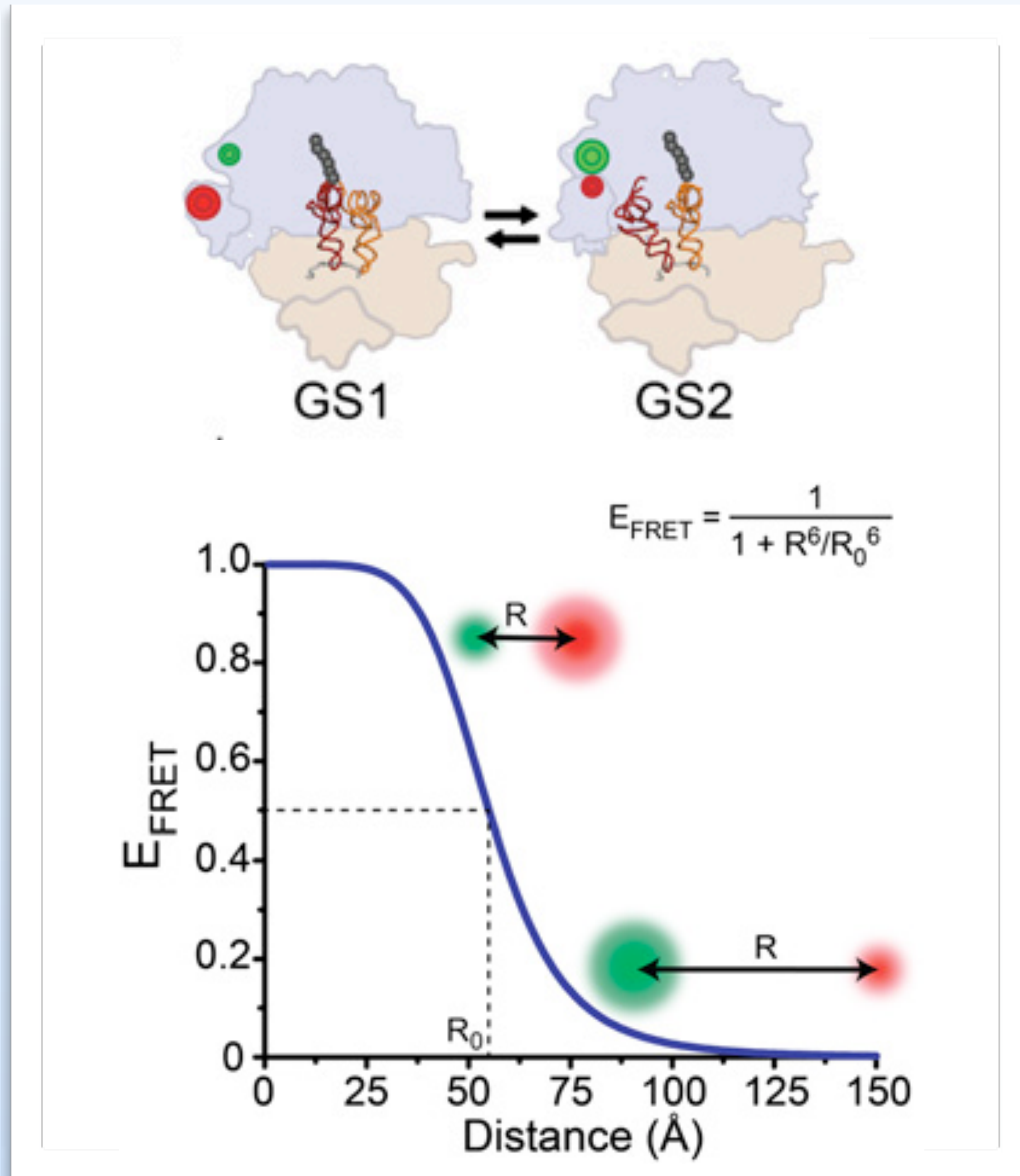


Ramakrishnan et al – http://www.mrc-lmb.cam.ac.uk/ribo/

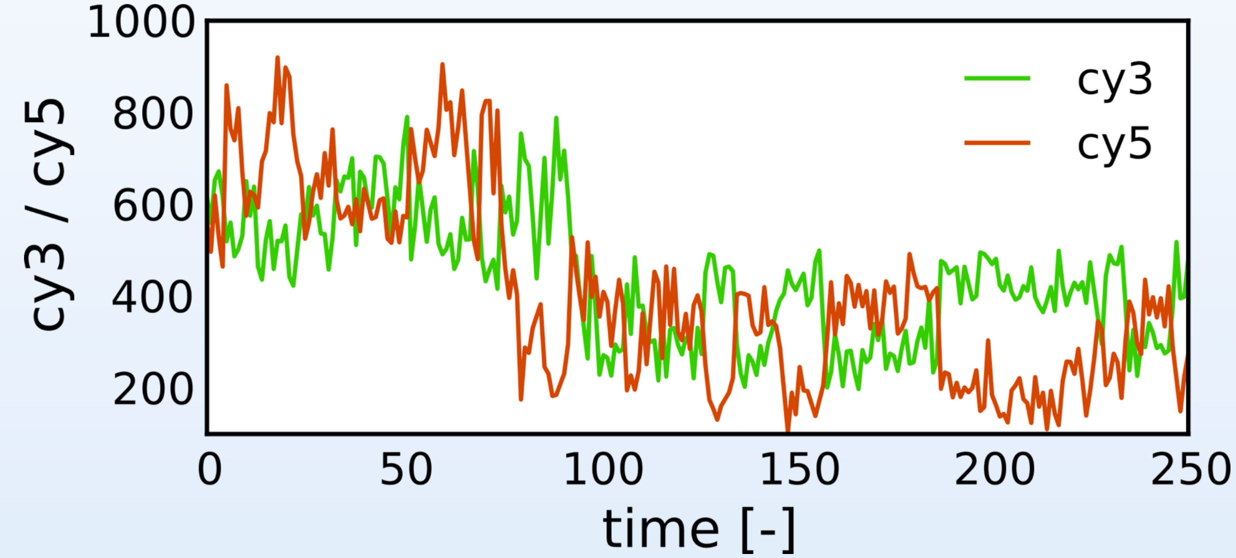


Ramakrishnan et al – http://www.mrc-lmb.cam.ac.uk/ribo/

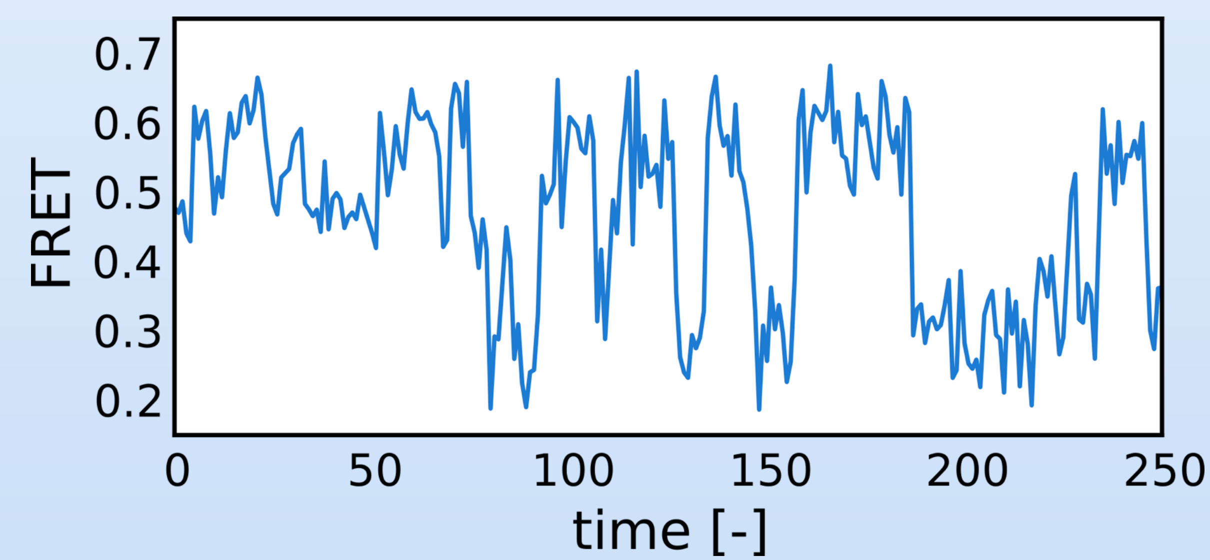
Single-Molecule FRET



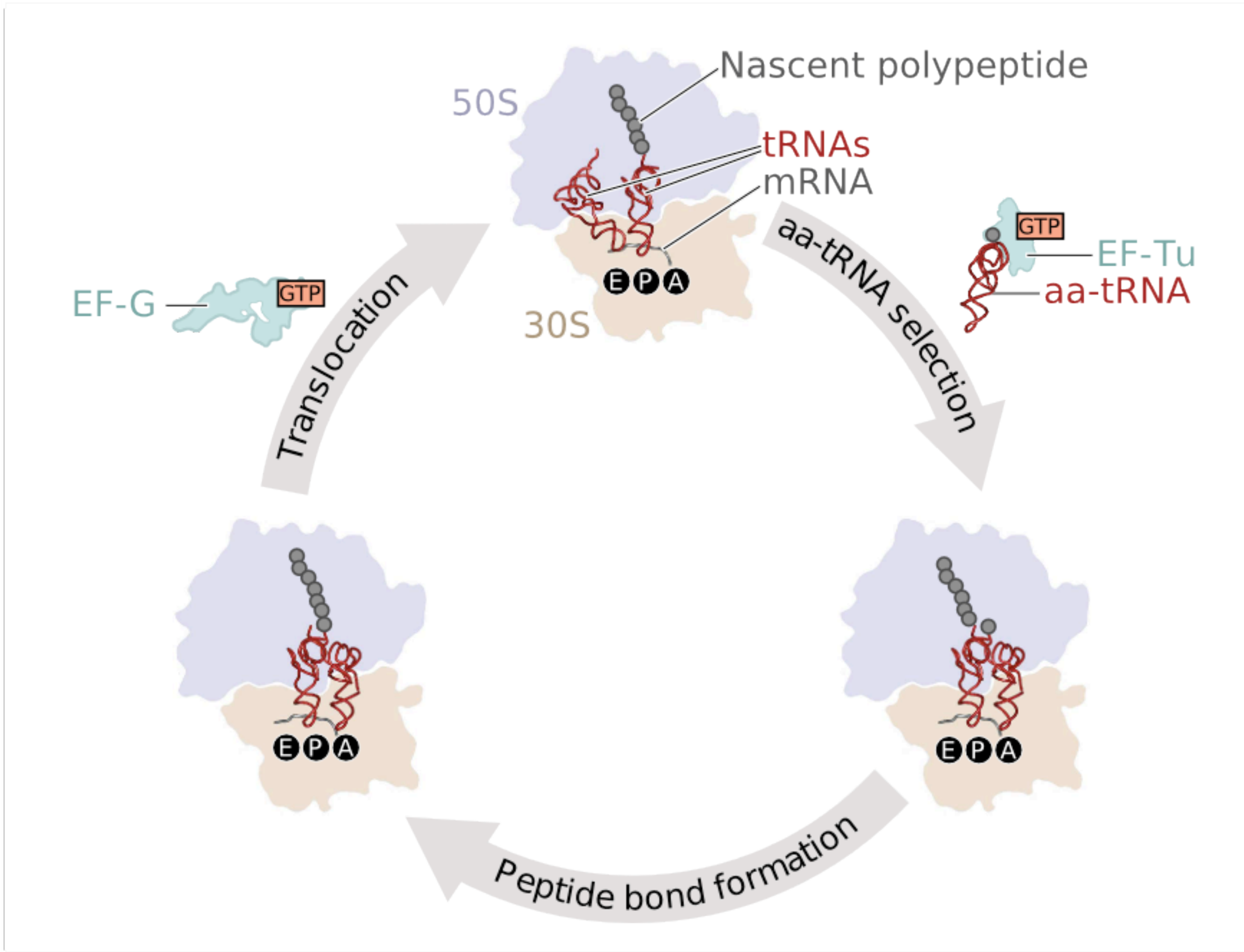
Tinoco and Gonzalez, Genes Dev, 2011



FRET = cy5 / (cy3 + cy5)

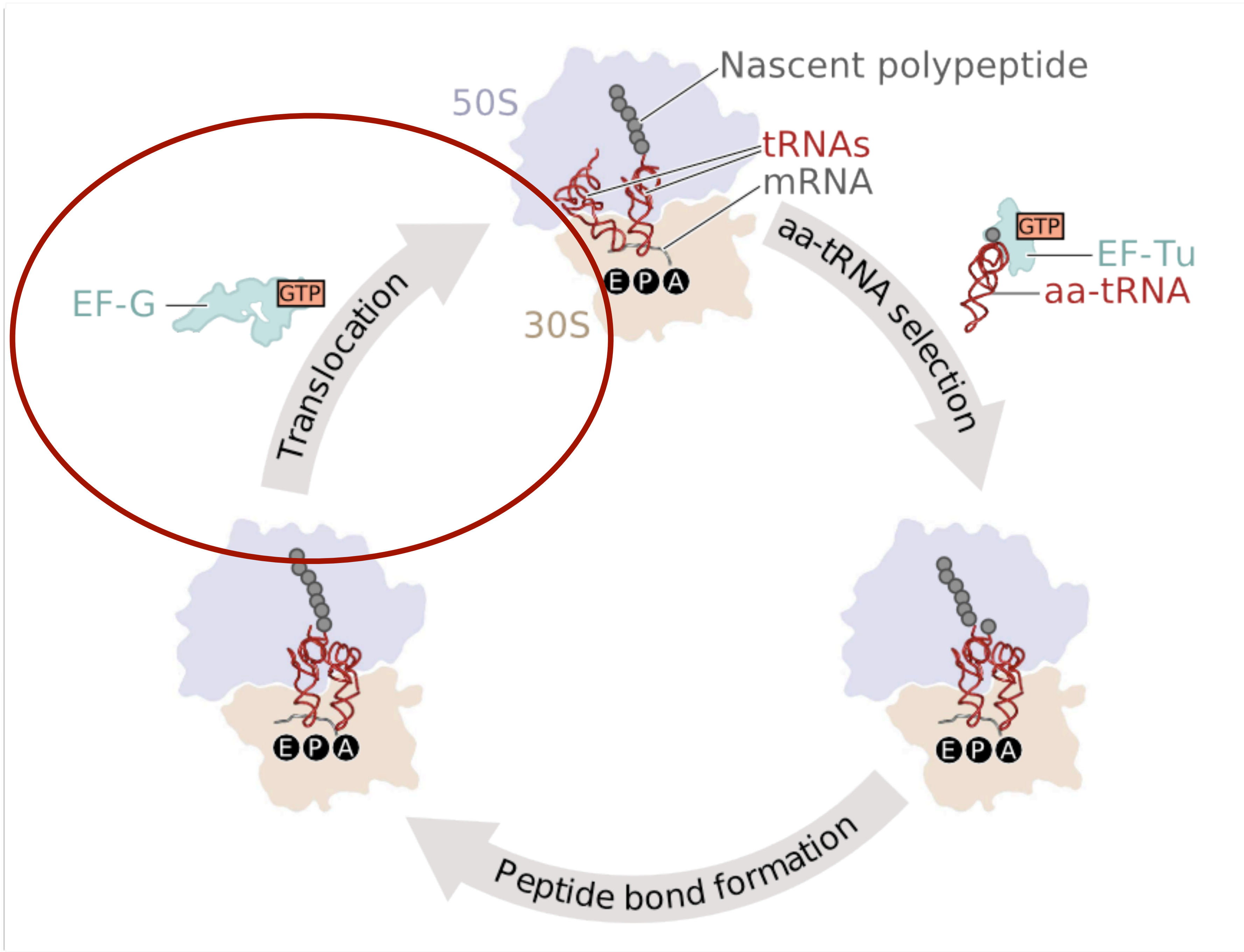


Elongation Cycle



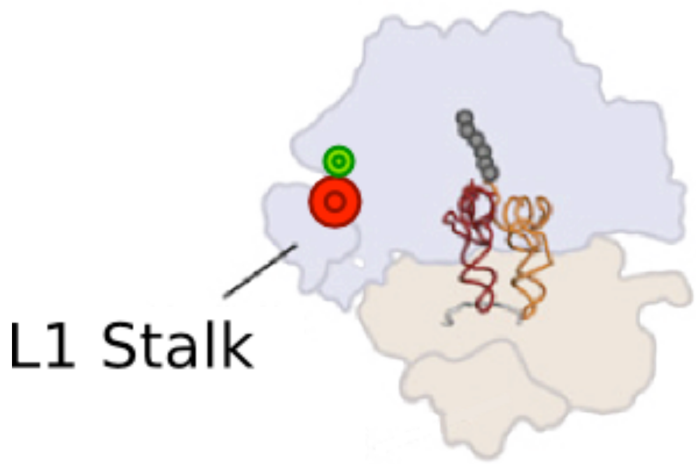
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



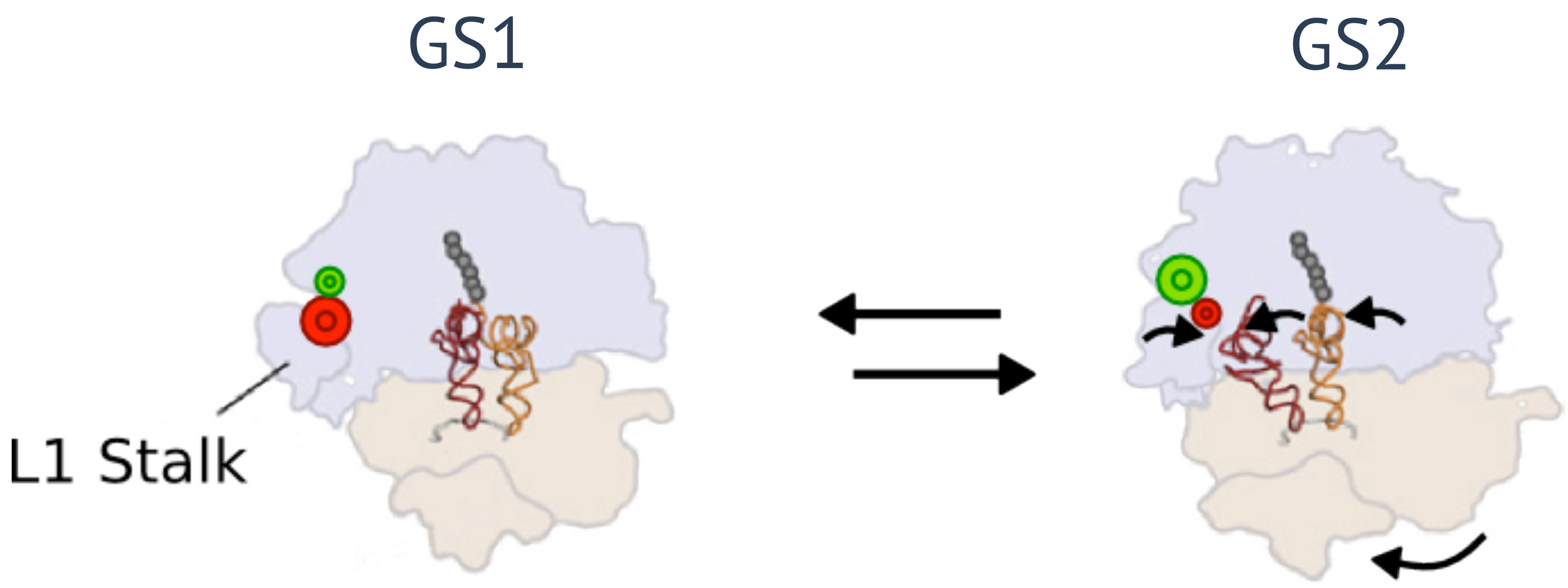
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



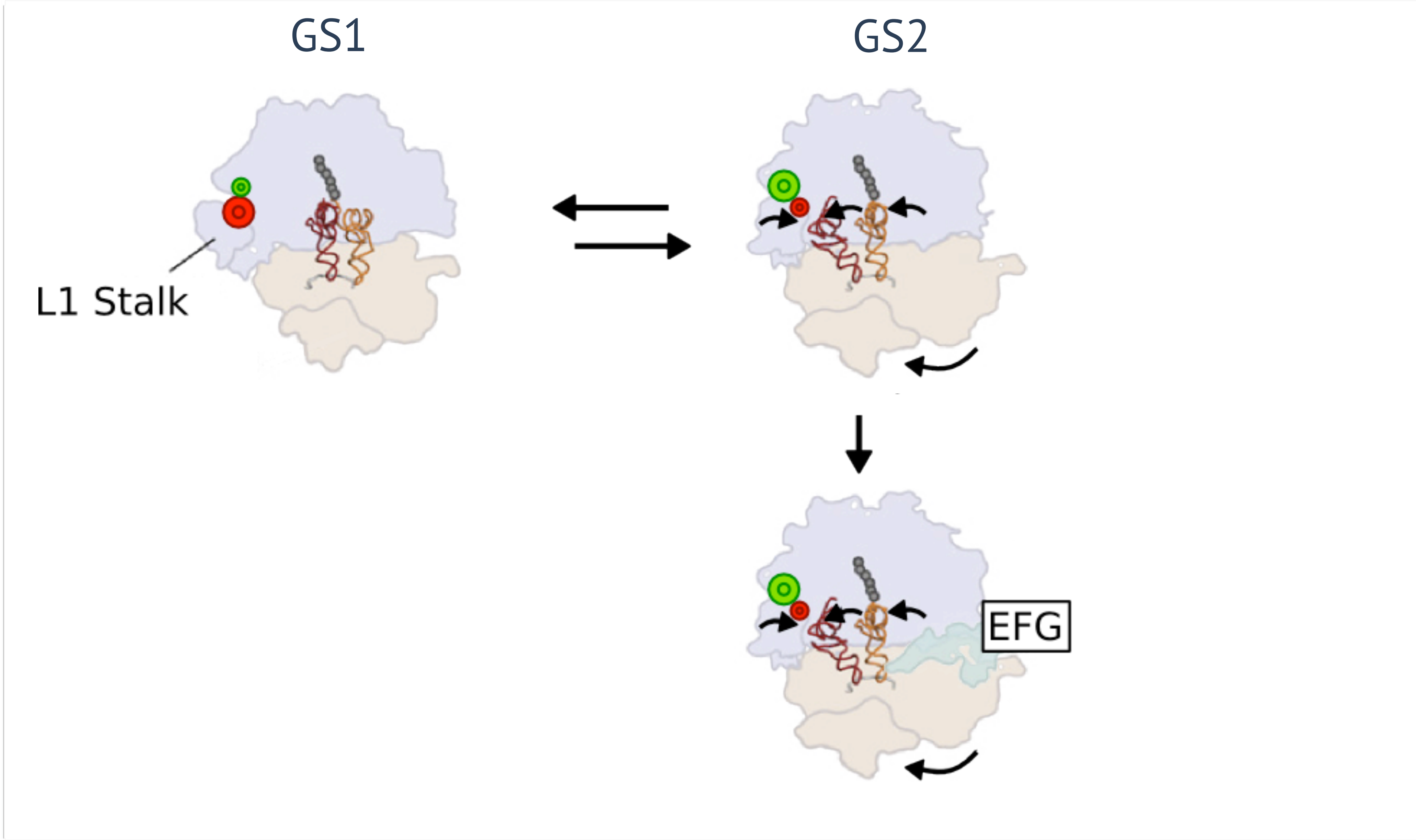
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



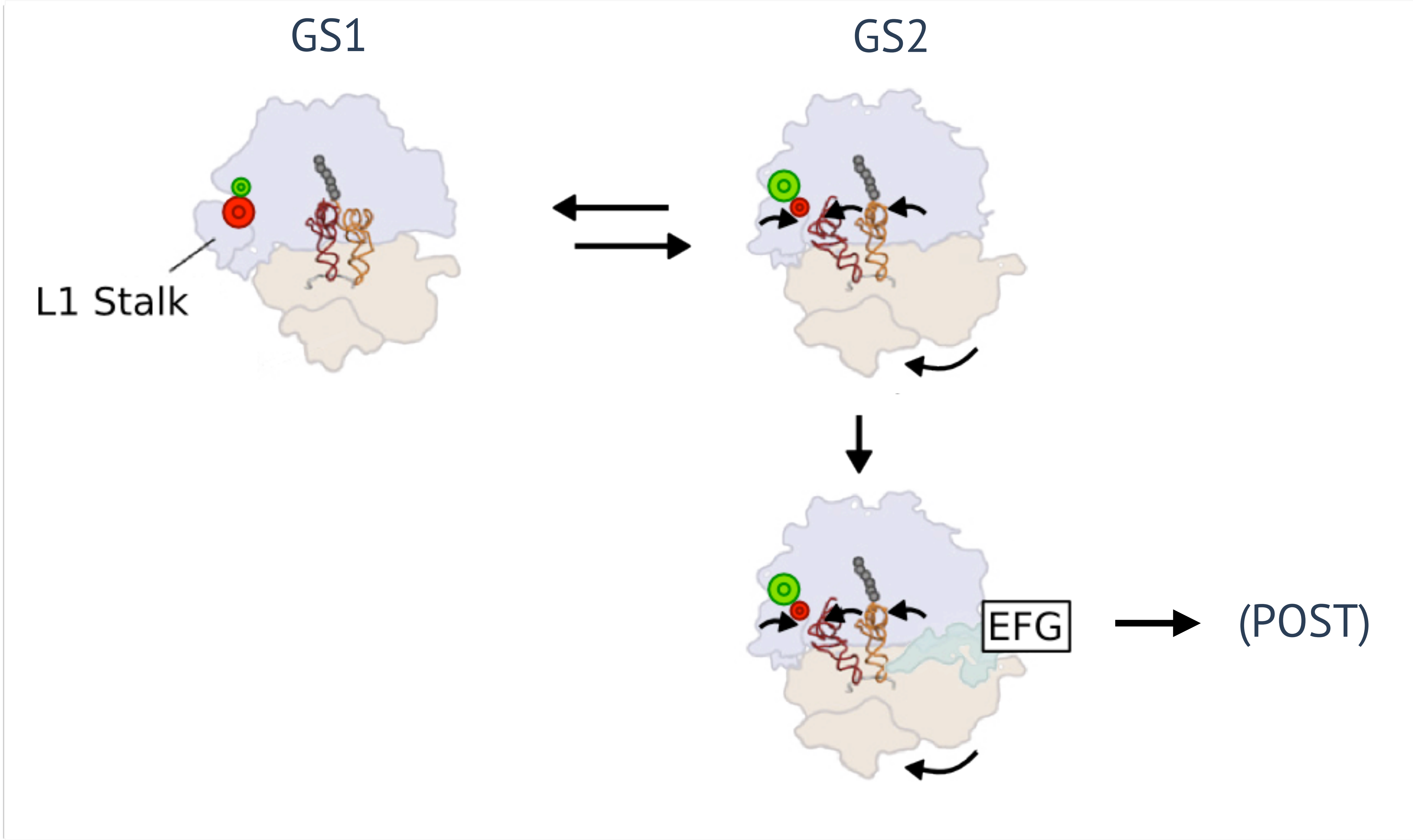
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



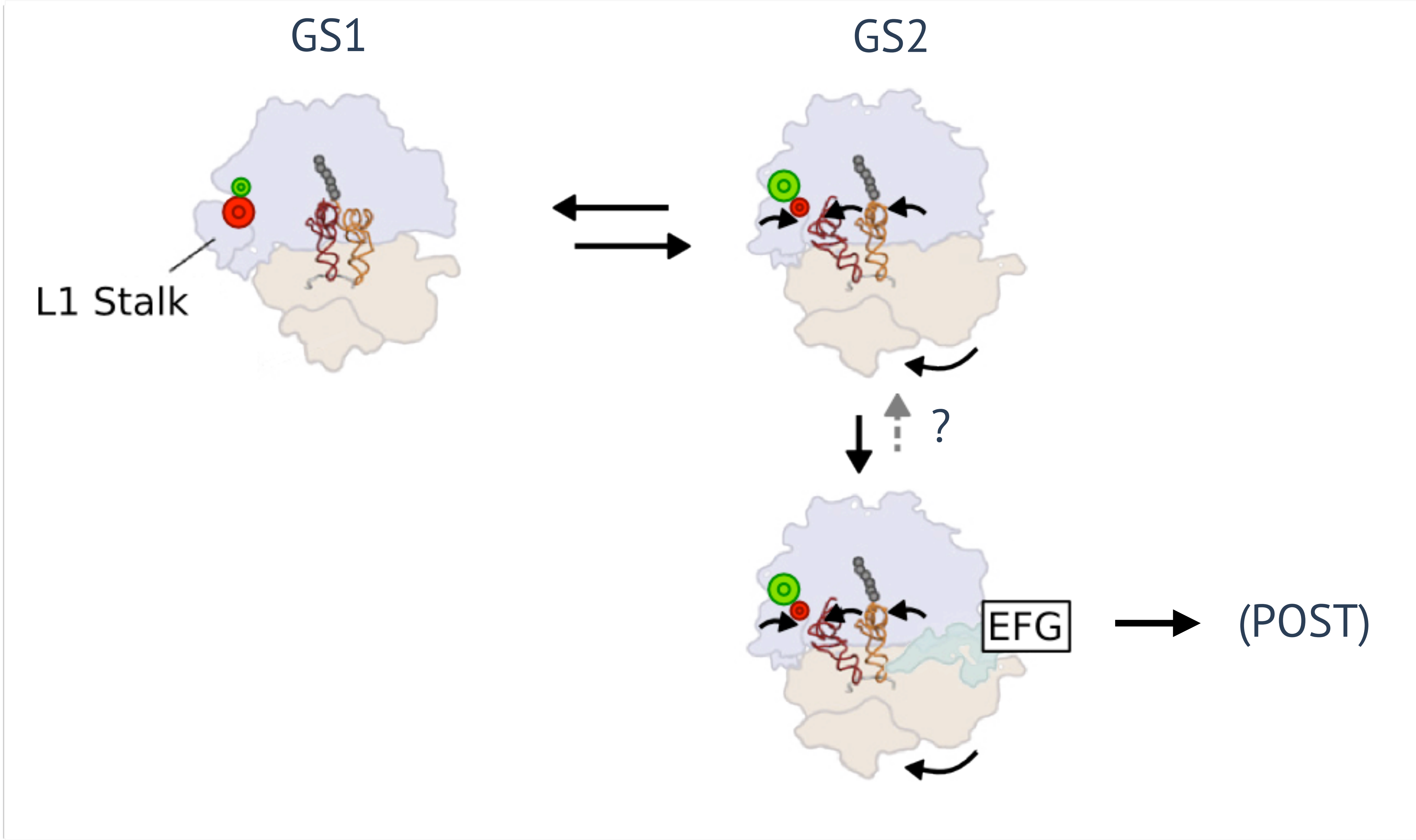
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



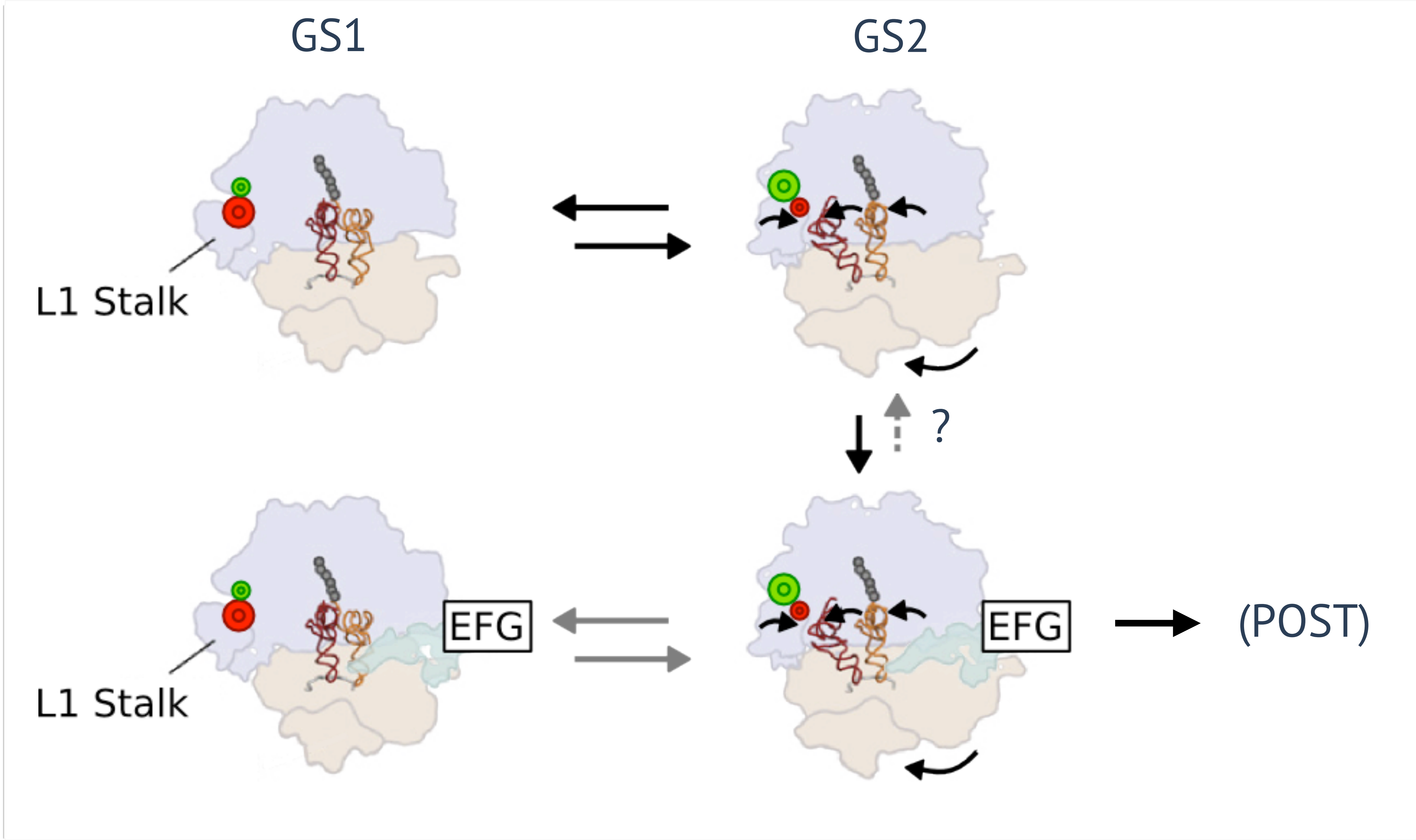
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



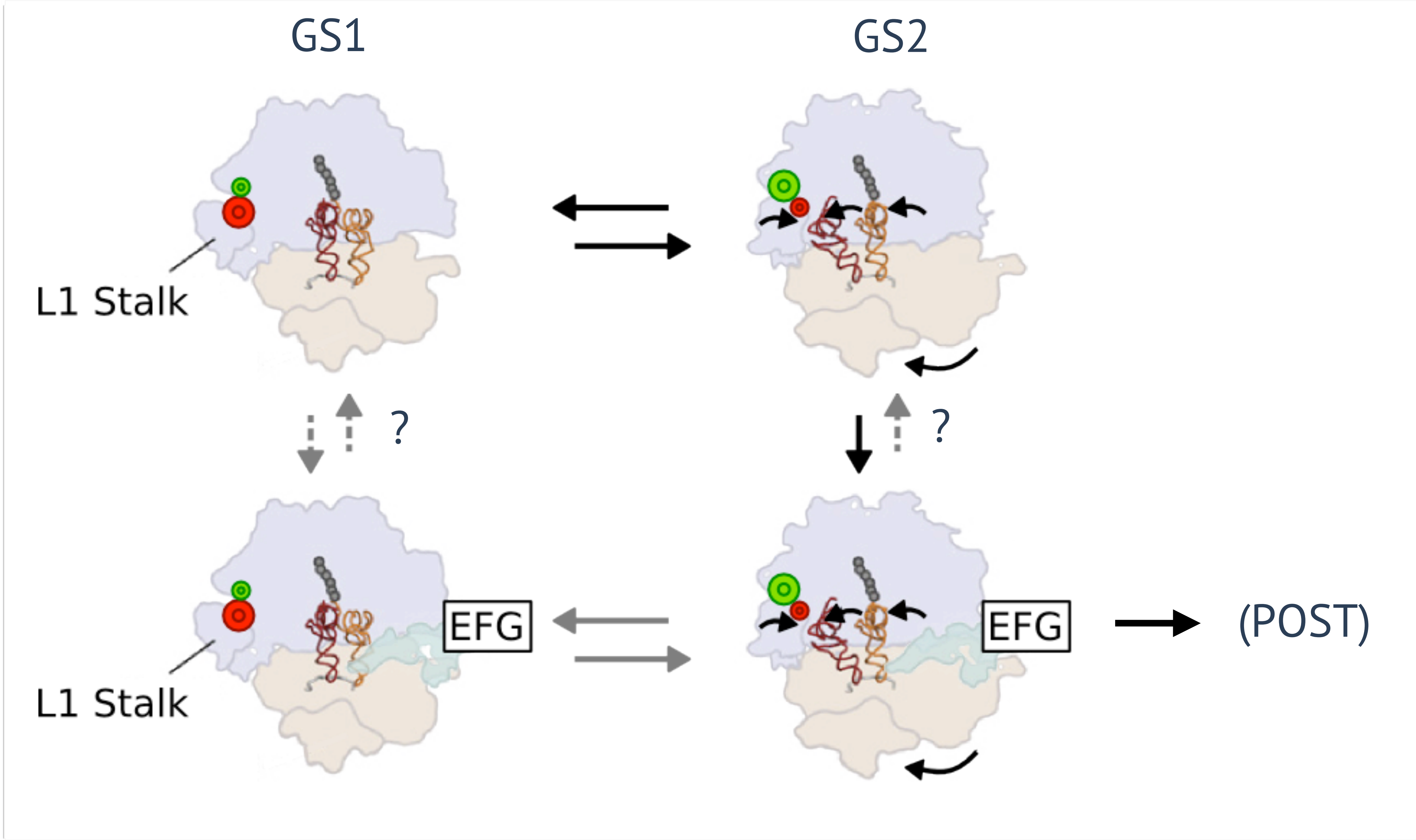
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle



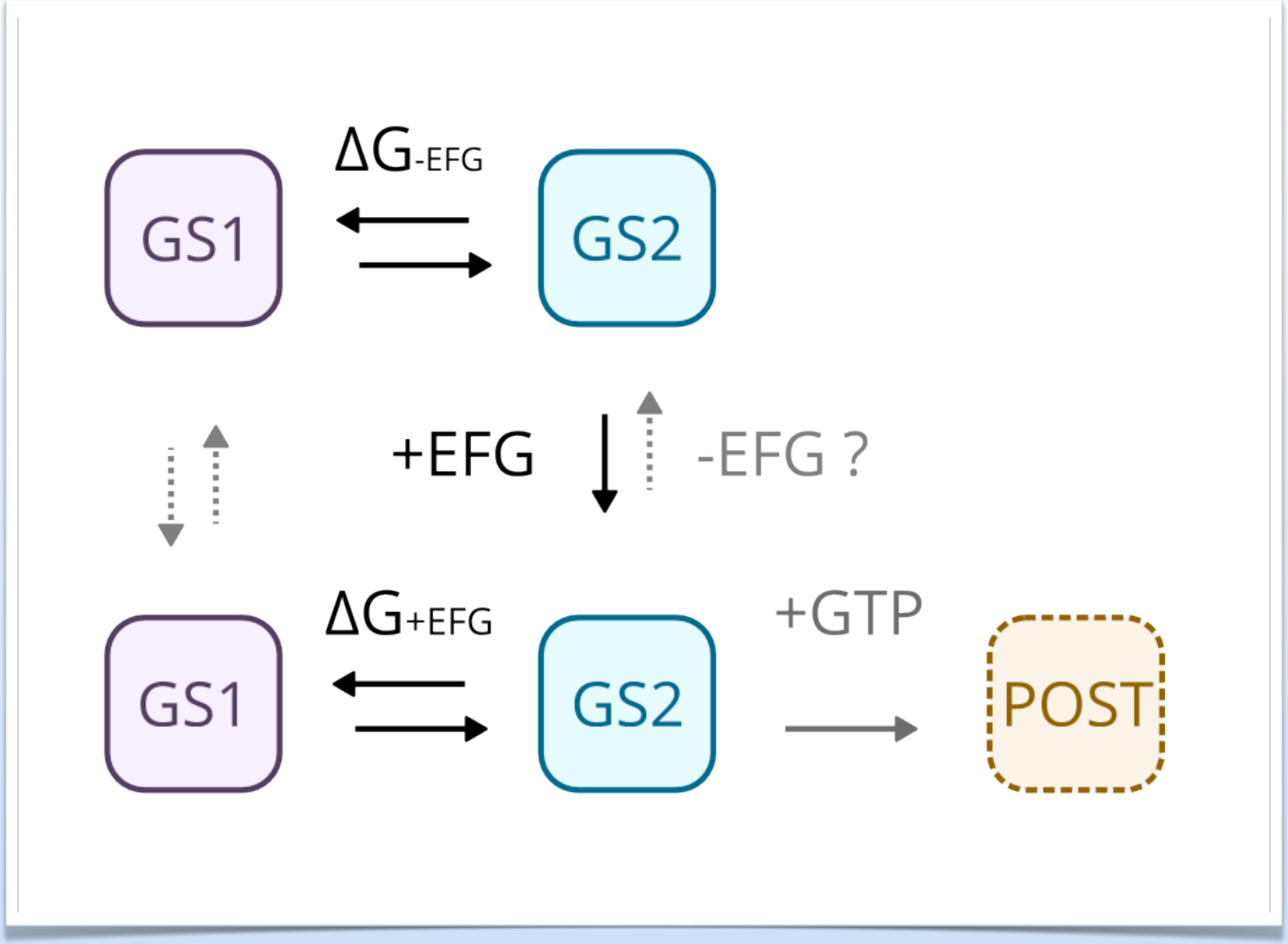
Tinoco and Gonzalez, Genes Dev, 2011

Elongation Cycle

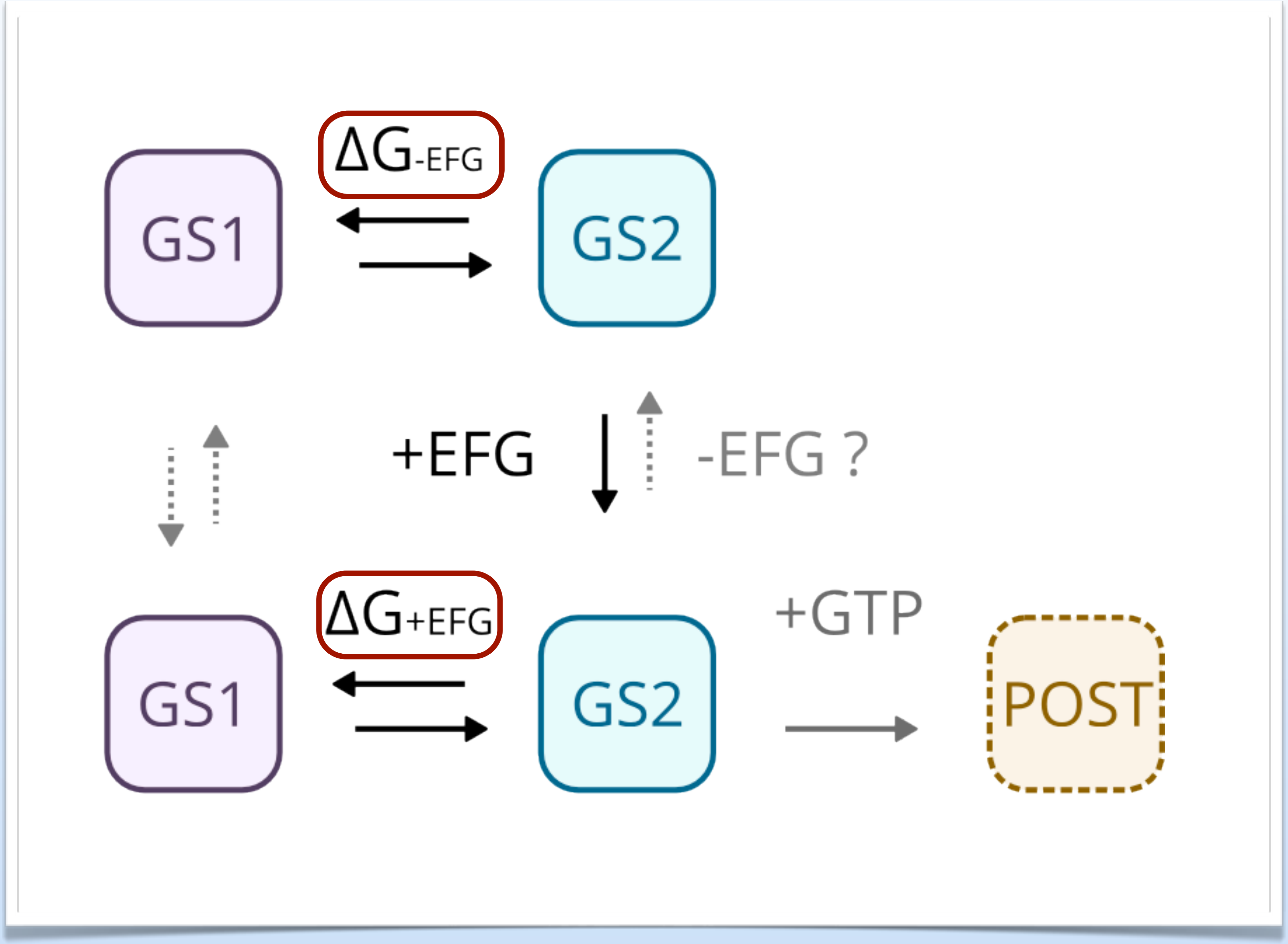


Tinoco and Gonzalez, Genes Dev, 2011

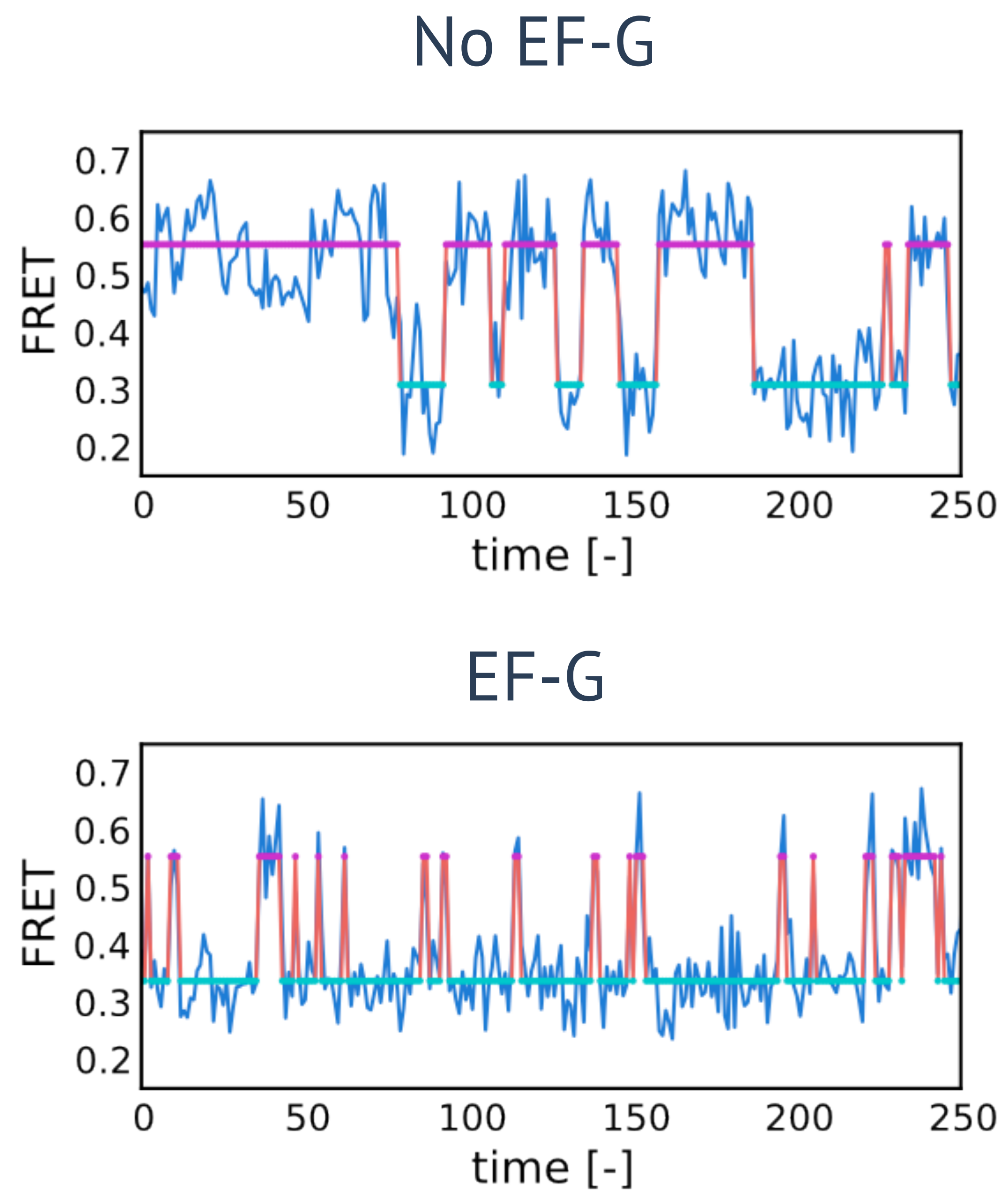
Kinetic Scheme



Kinetic Scheme



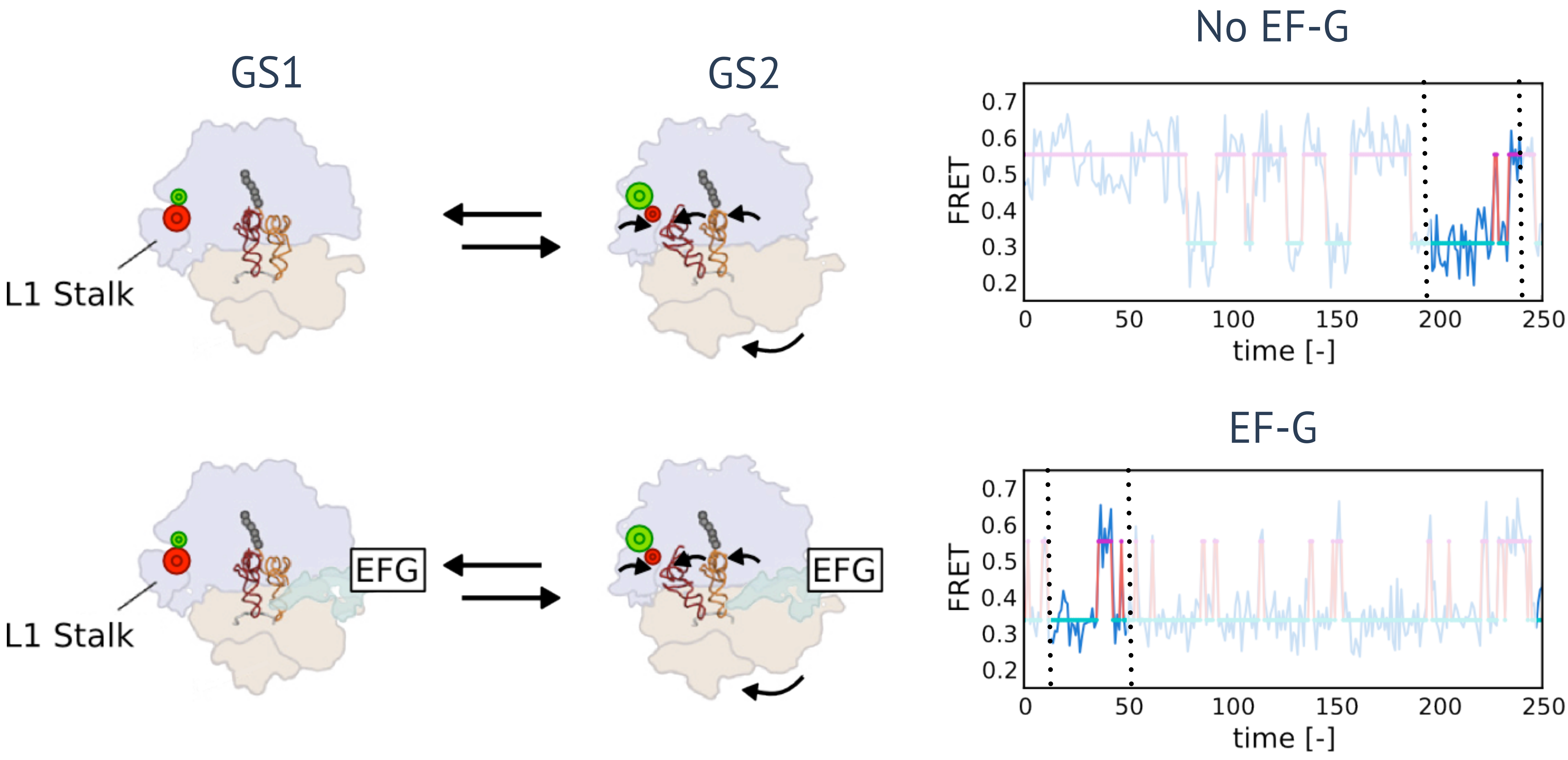
Elongation Cycle



Tinoco and Gonzalez, Genes Dev, 2011

Fei et al, PNAS, 2009

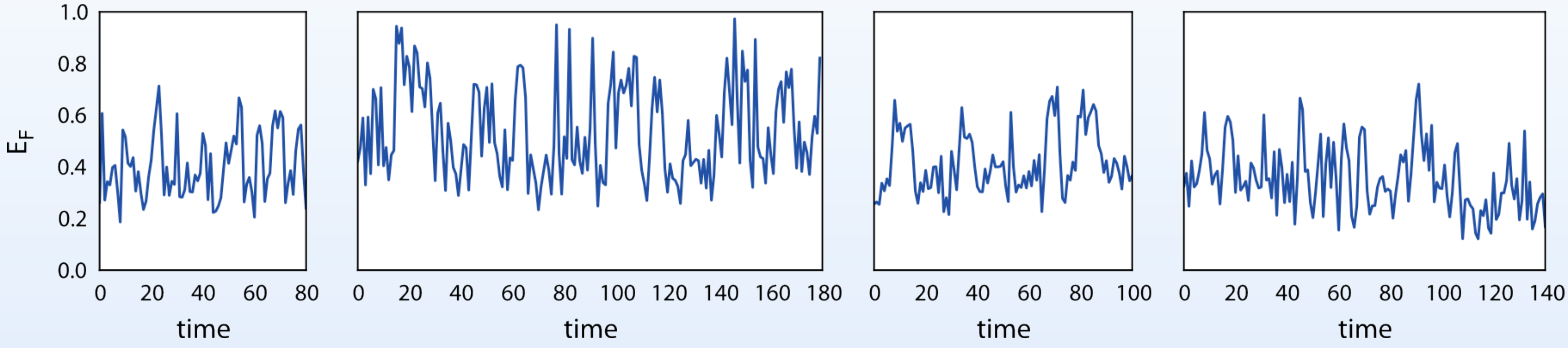
Elongation Cycle



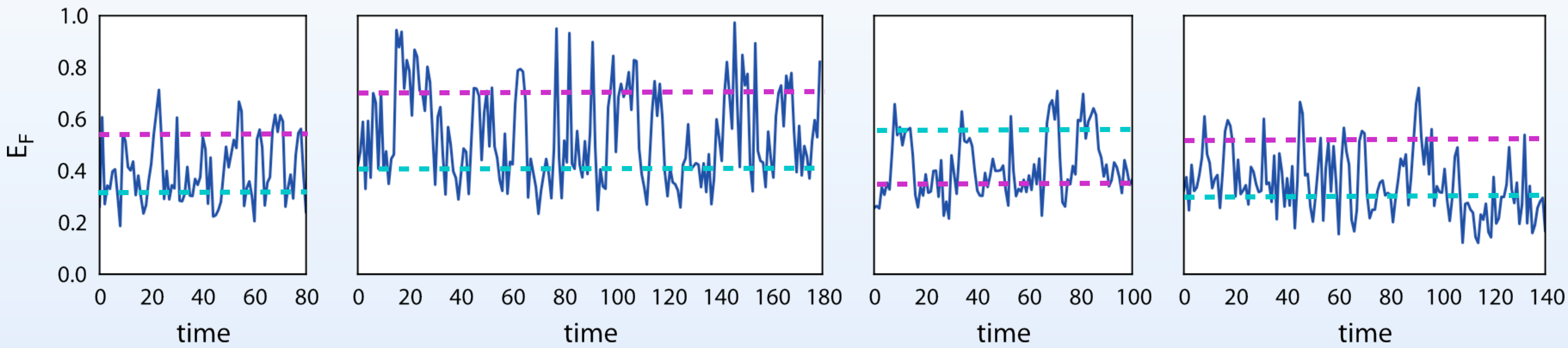
Tinoco and Gonzalez, Genes Dev, 2011

Fei et al, PNAS, 2009

Learning Kinetics from Traces

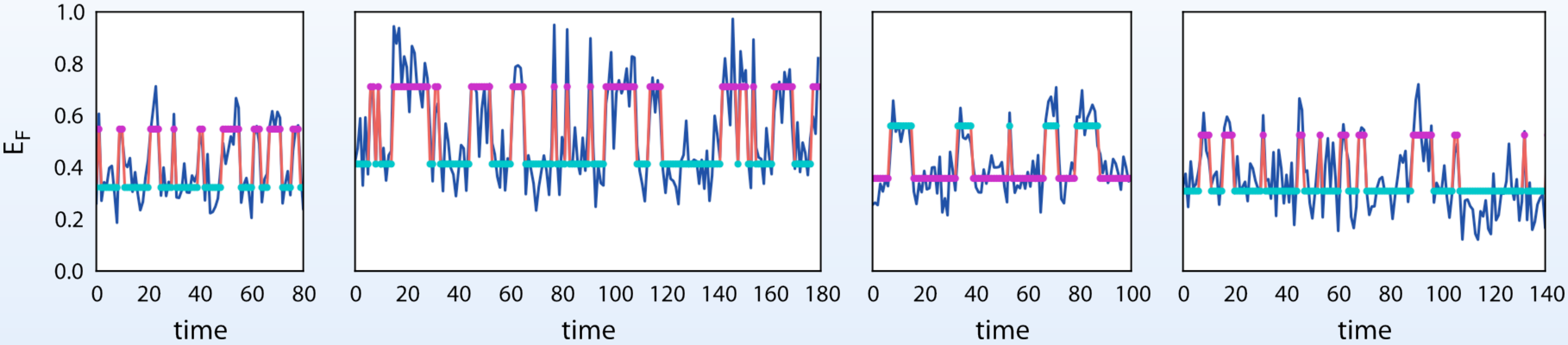


Learning Kinetics from Traces



1. Identify states

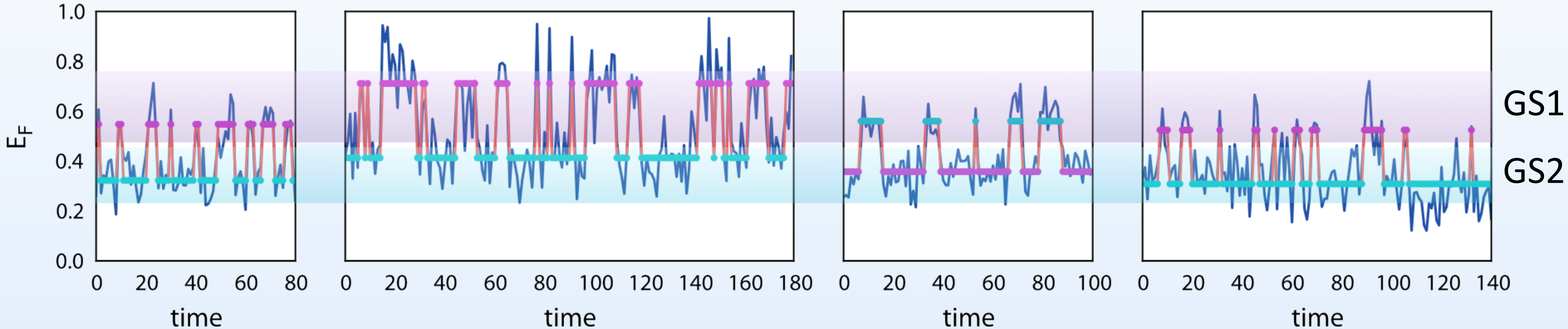
Learning Kinetics from Traces



1. Identify states

2. Calculate Kinetic Rates

Learning Kinetics from Traces

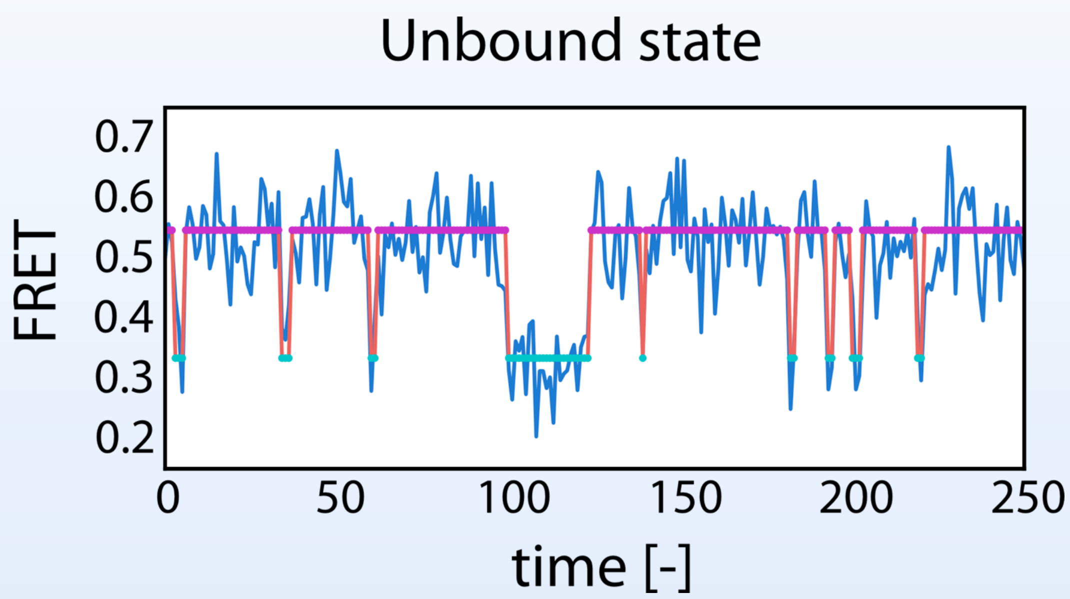


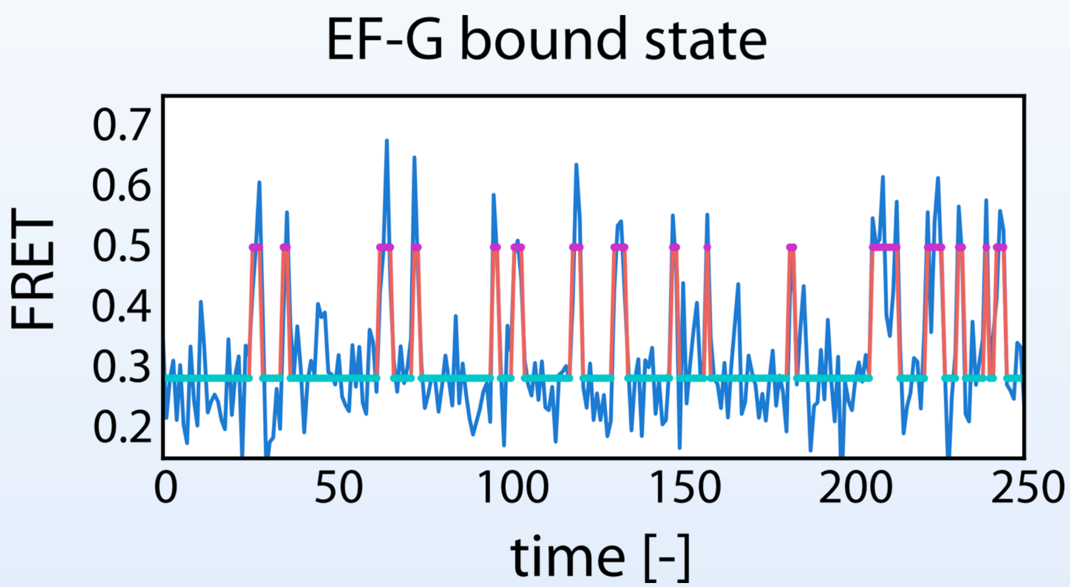
1. Identify states

2. Calculate Kinetic Rates

3. Construct Consensus Model

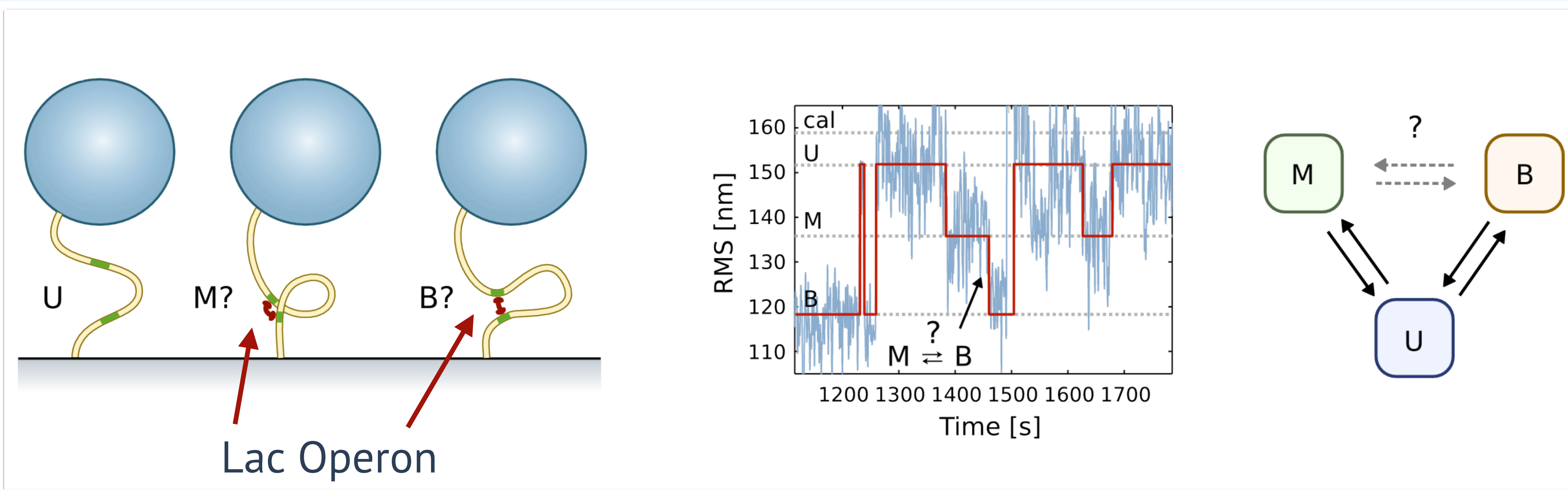
Learning Kinetics from Traces





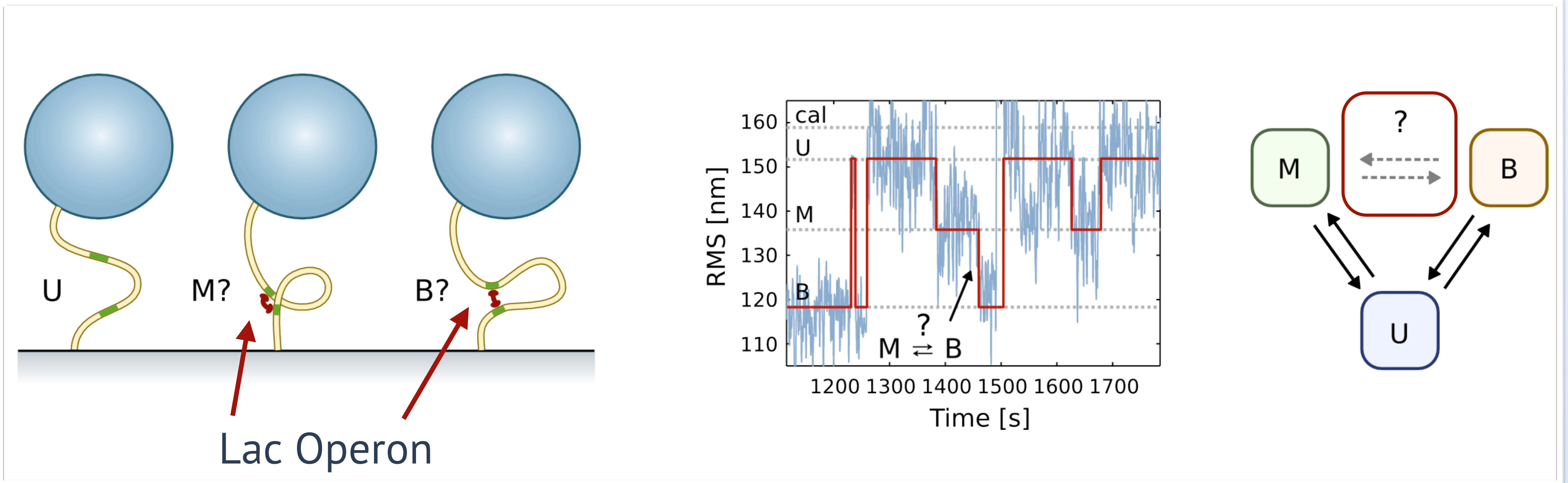
1. Identify states
2. Calculate Kinetic Rates
3. Construct Consensus Model
4. Distinguish Subpopulations

Tethered Particle Motion



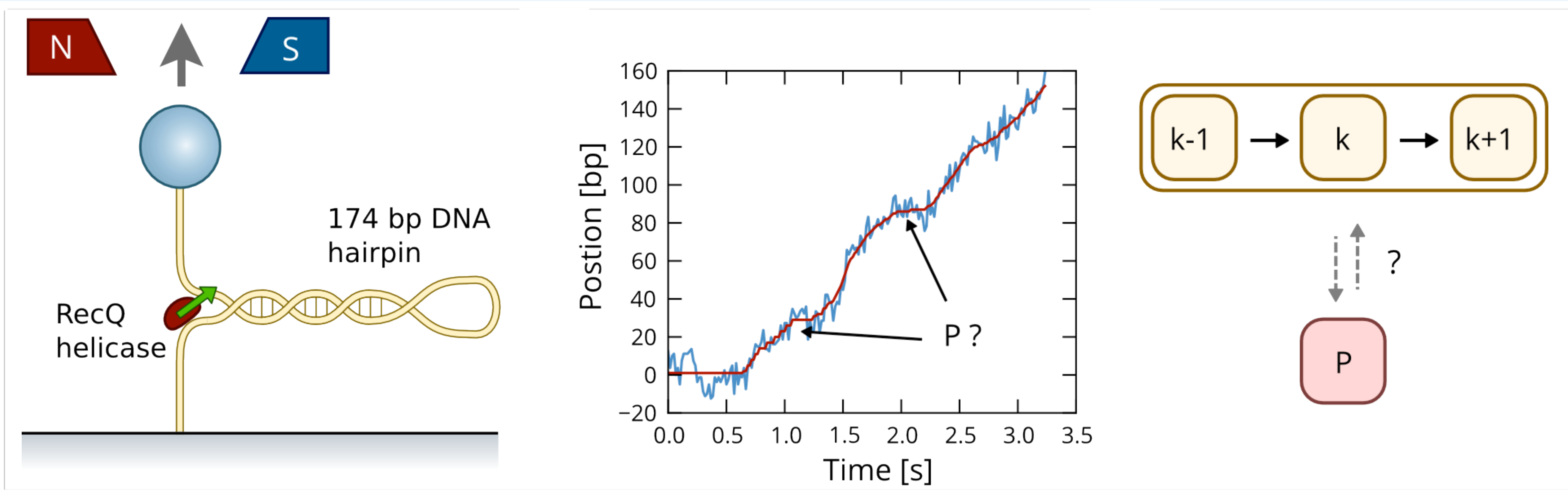
Phillips Group, Caltech

Tethered Particle Motion



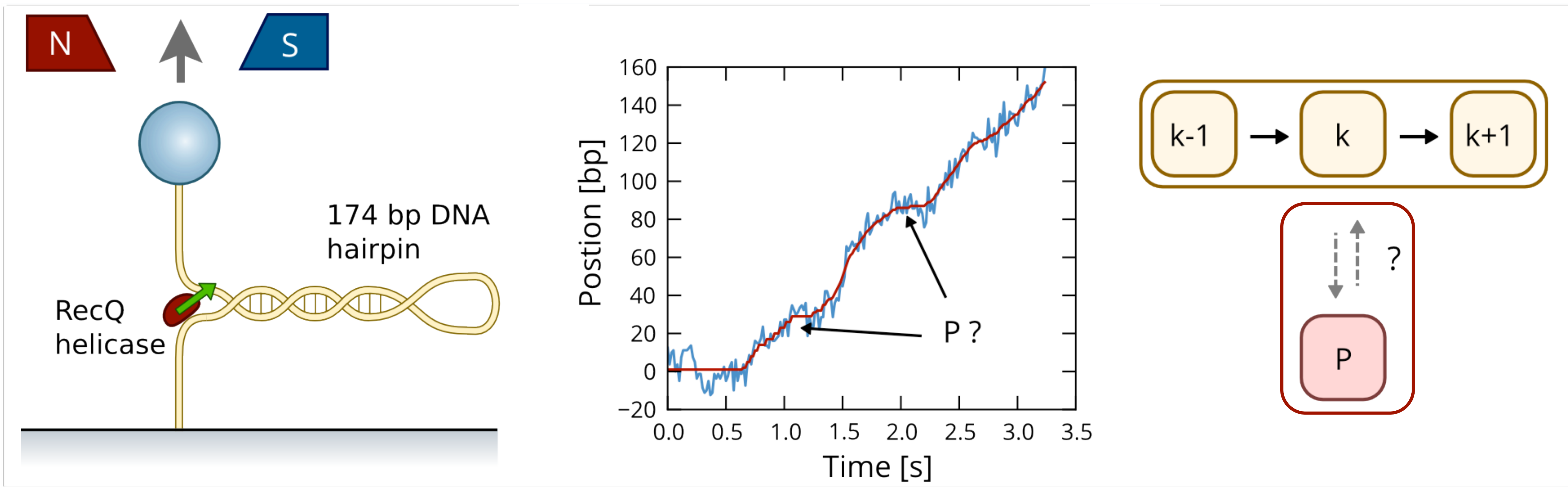
Phillips Group, Caltech

Magnetic Tweezers



Neuman Group, NIH

Magnetic Tweezers



Neuman Group, NIH

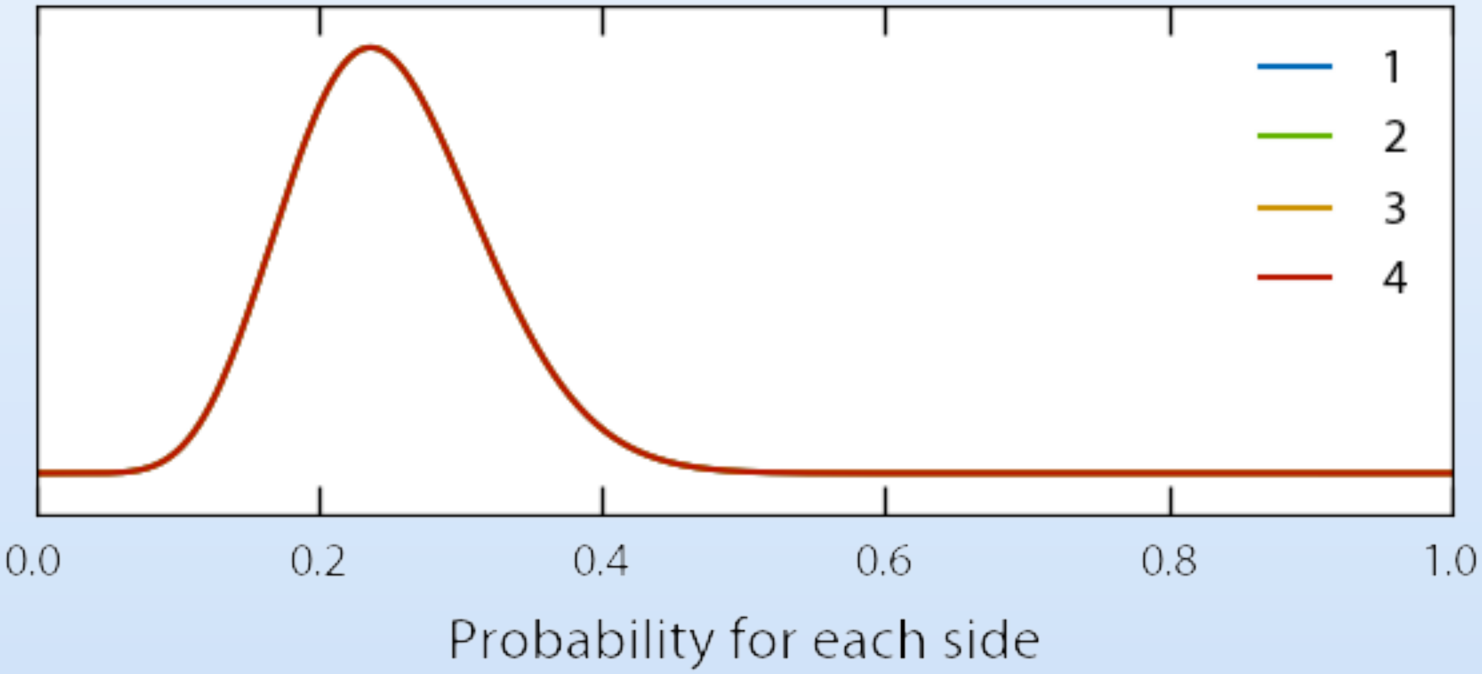
Experiment -> Kinetic Pathway

- Many molecules
- Lots of noise, thermal fluctuations
- Few transitions per molecule

Fancy Counting

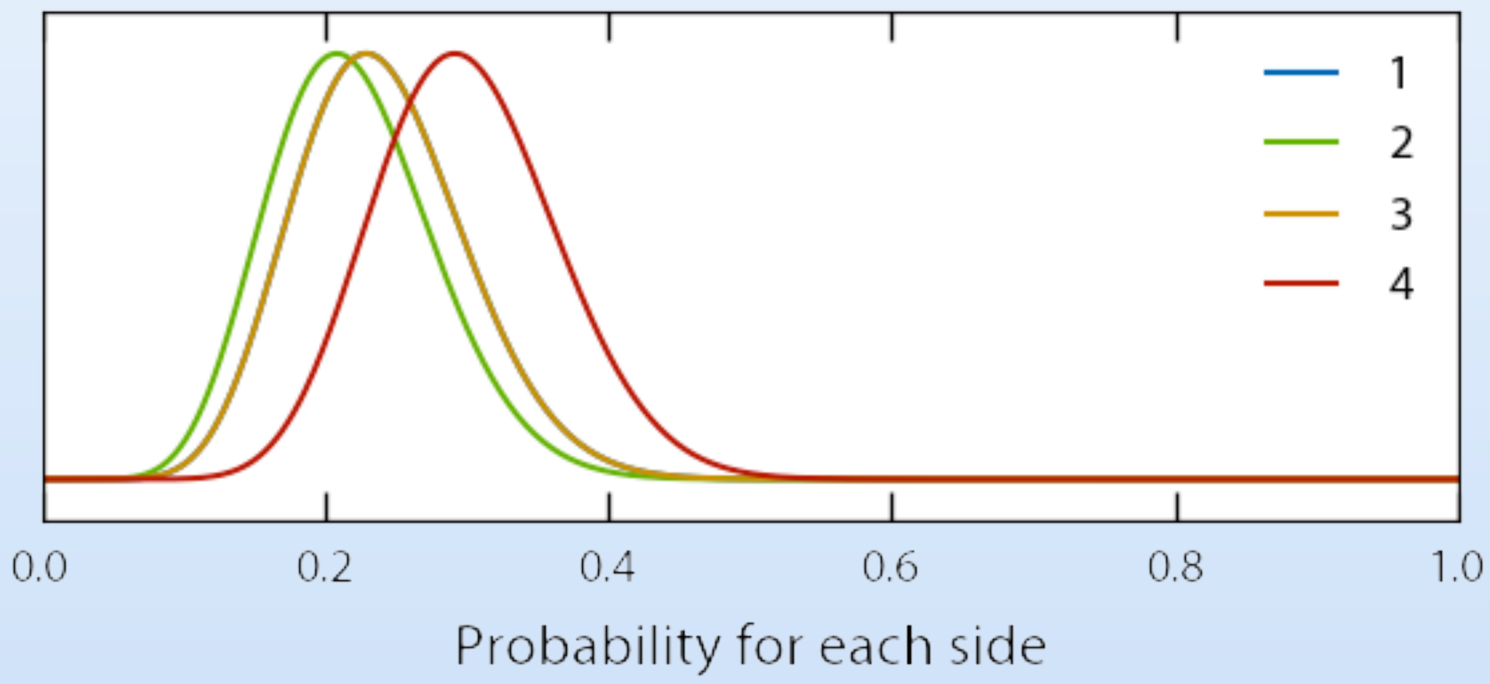
0 Rolls

1	**2**	**3**	**4**
0	0	0	0



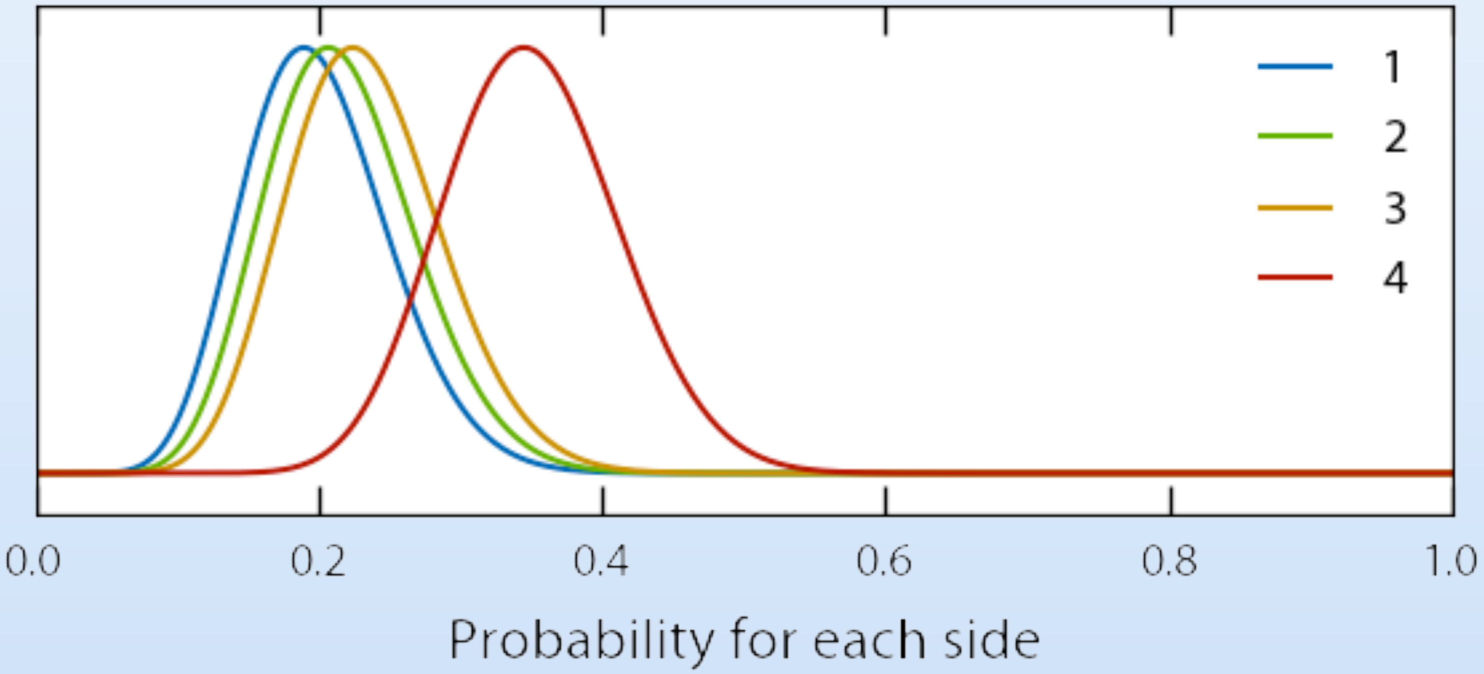
10 Rolls

1	2	3	4
2	1	2	5



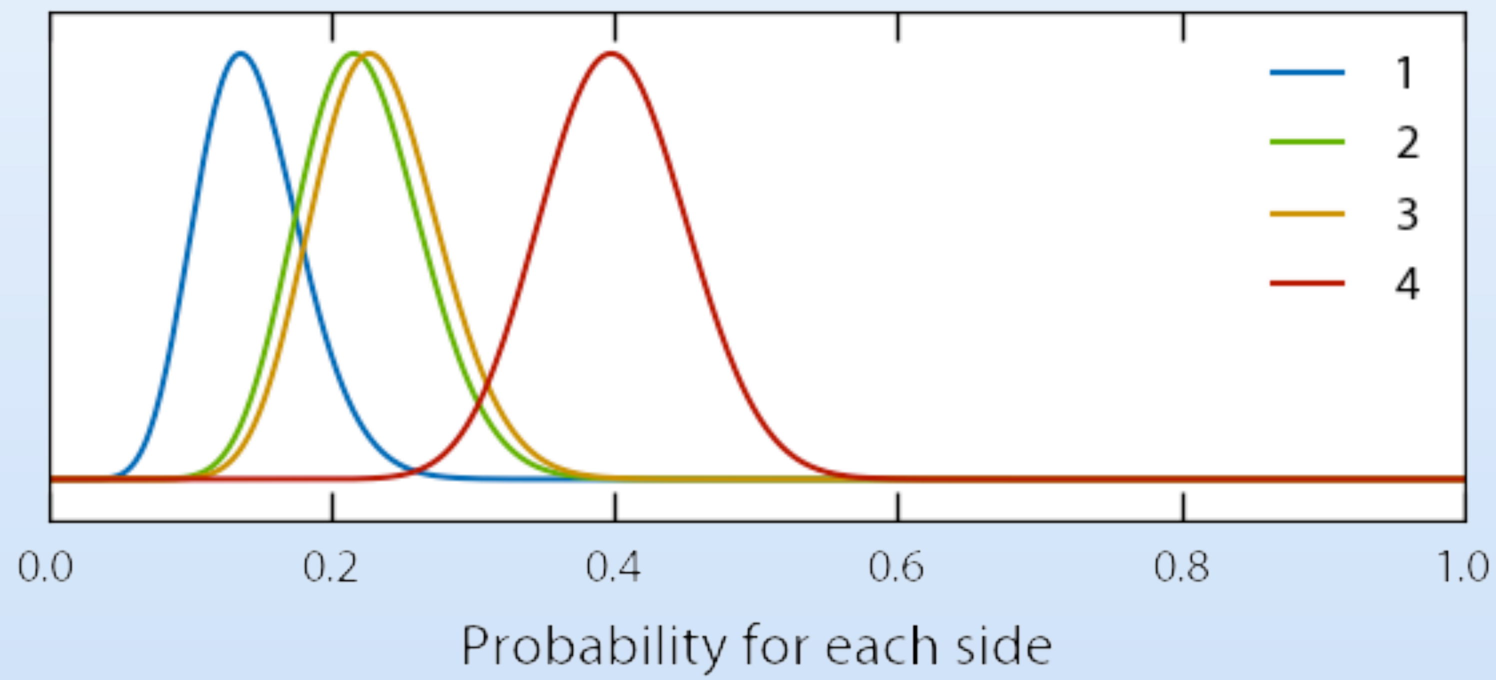
20 Rolls

1	**2**	**3**	**4**
2	3	4	11



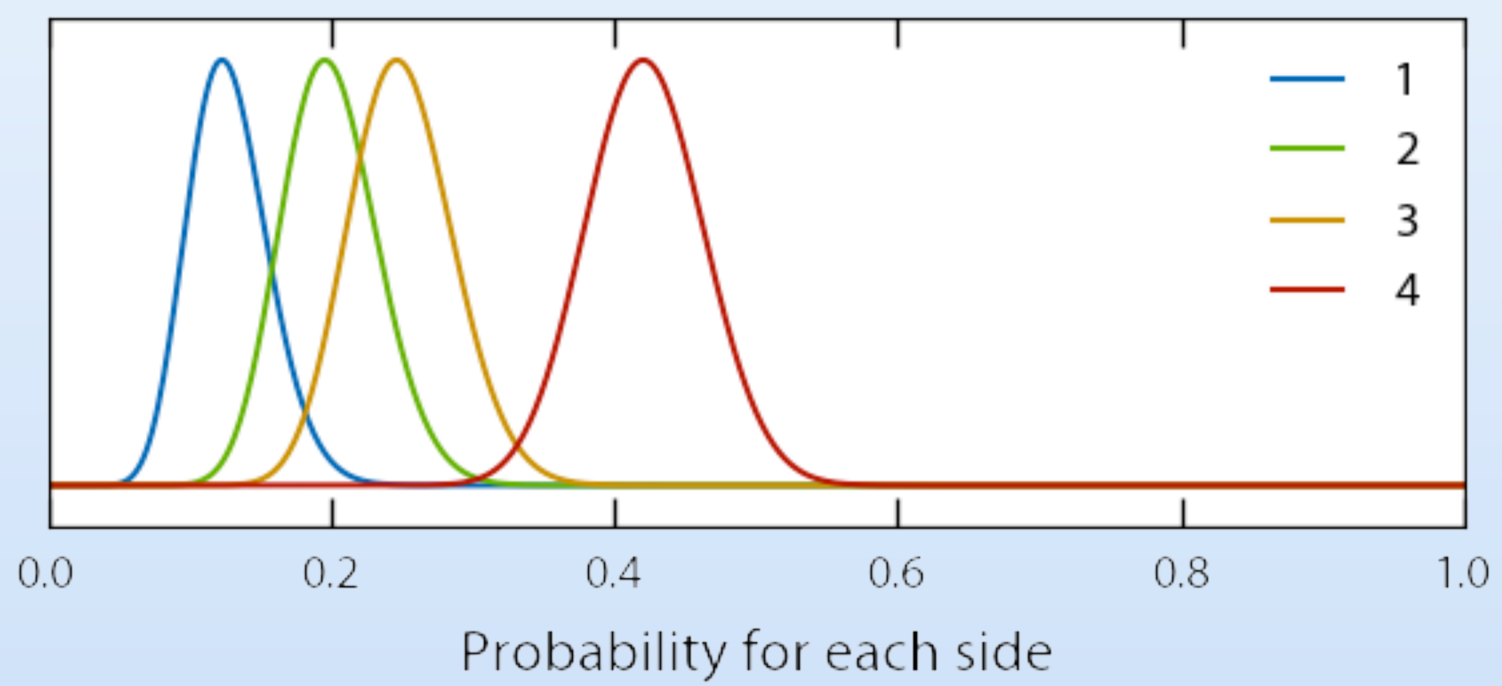
50 Rolls

1	**2**	**3**	**4**
3	10	11	26



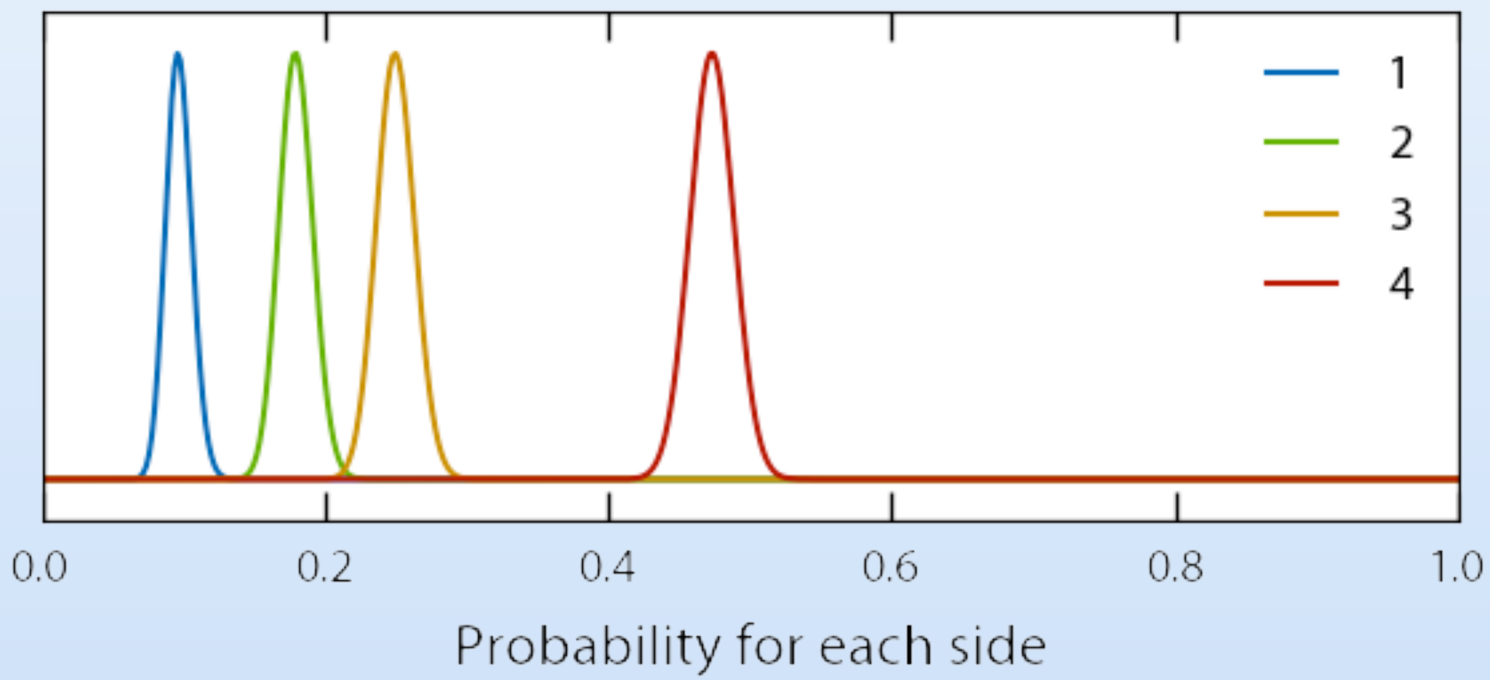
100 Rolls

1	2	3	4
8	18	25	49



1000 Rolls

1	**2**	**3**	**4**
91	177	250	482



Two Easy Pieces

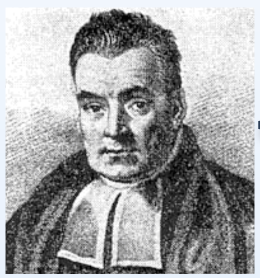
$$p(a,b) = p(a \mid b)\,p(b)$$
$$= p(b \mid a)\,p(a)$$

$$p(a) = \sum_b p(a,b)$$
$$p(b) = \sum_a p(a,b)$$

Bayes' Rule

$$p(b \mid a) = p(a \mid b)p(b)/p(a)$$

Bayes' Rule

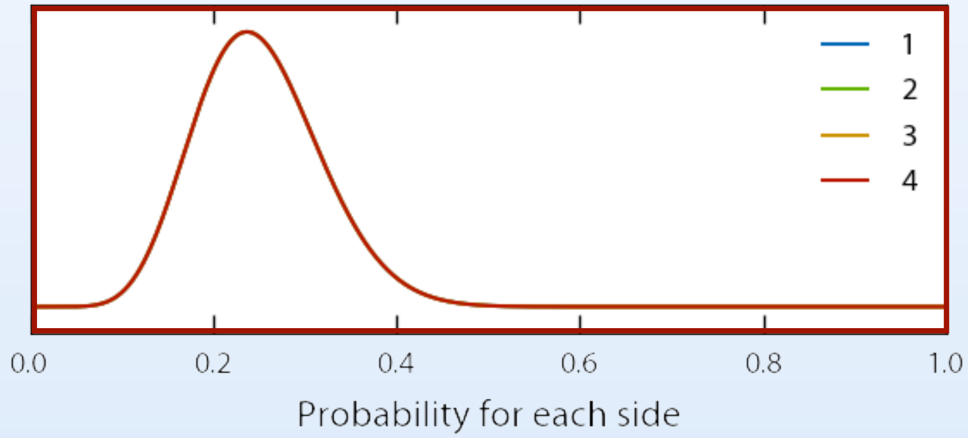


$$p(b \mid a) = p(a \mid b)p(b)/p(a)$$

$$p(w \mid n, n_0) \quad \propto \quad p(n \mid w, n_0) \quad\quad p(w \mid n_0)$$

Posterior Observations Prior

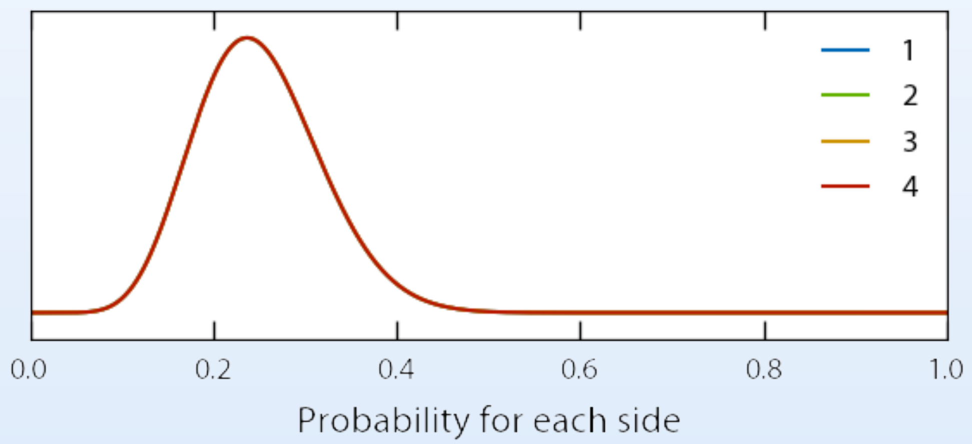
Bayes' Rule



$$p(w \mid n, n_0) \quad \propto \quad p(n \mid w, n_0) \quad p(w \mid n_0)$$

Posterior ∝ Observations · Prior

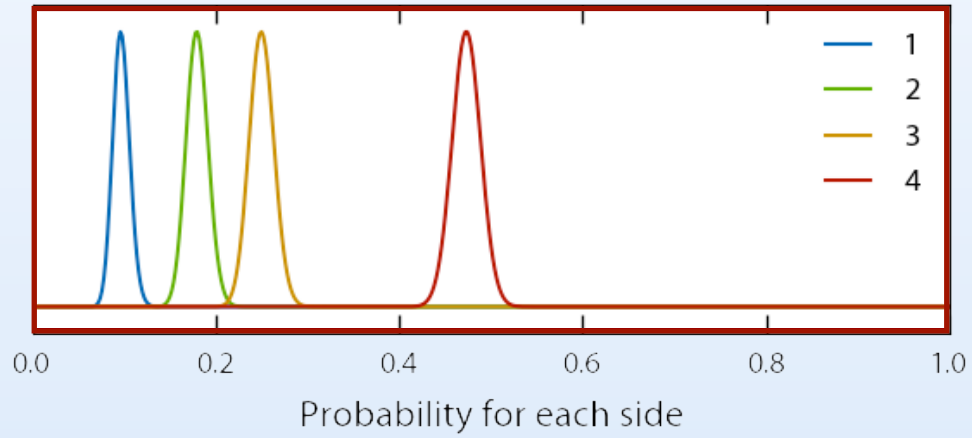
Bayes' Rule

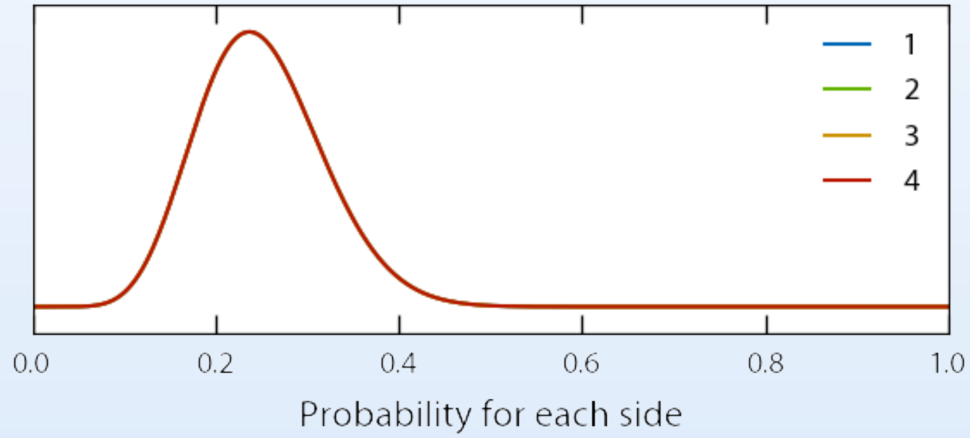


$$p(w \mid n, n_0) \propto p(n \mid w, n_0) \; p(w \mid n_0)$$

Posterior | Observations | Prior

Bayes' Rule



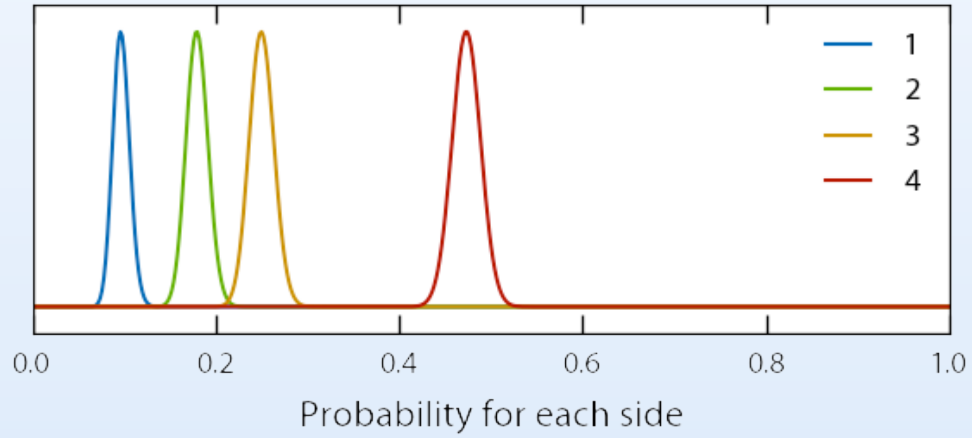


$$p(w \mid n, n_0) \propto p(n \mid w, n_0) \; p(w \mid n_0)$$

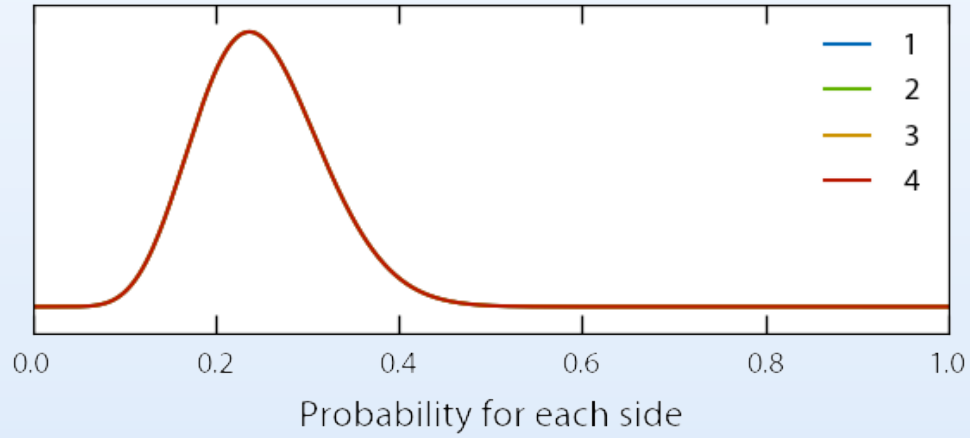
Posterior — Observations — Prior

Bayes' Rule

	1	2	3	4
n + n$_0$	101	187	260	492



	1	2	3	4
n$_0$	10	10	10	10

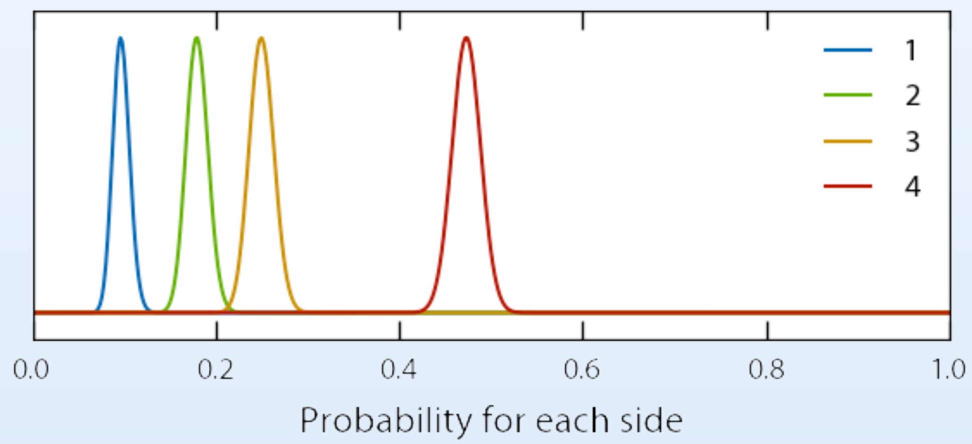


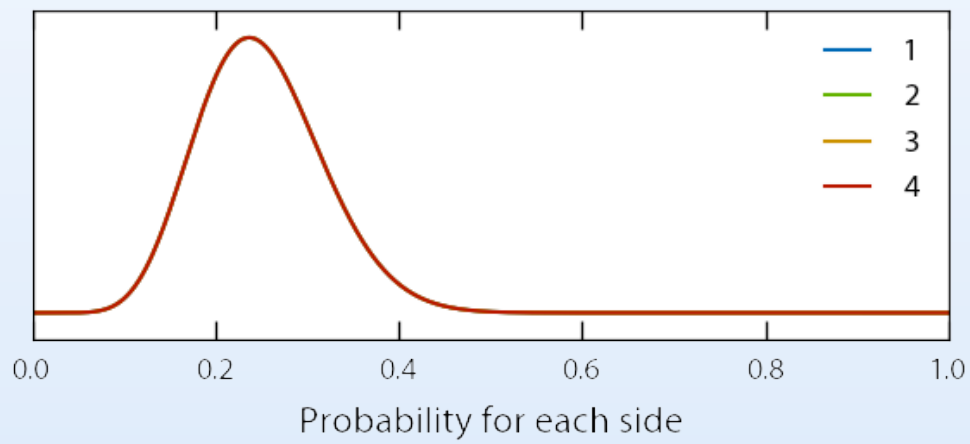
$$p(w \mid n, n_0) = p(w \mid n + n_0)$$

Bayes' Rule

	1	2	3	4
$n + n_0$	101	187	260	492

	1	2	3	4
n_0	10	10	10	10



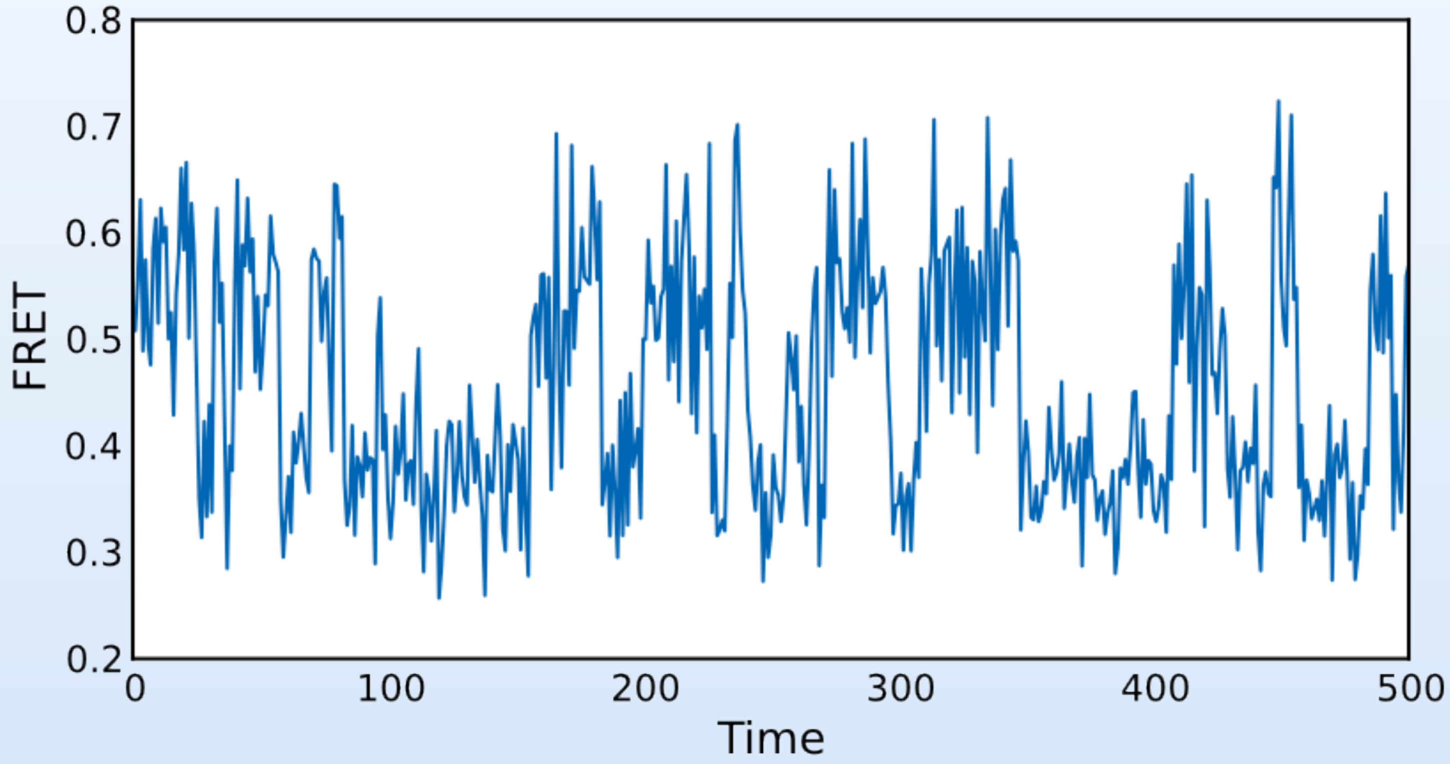


$$p(w \mid n, n_0) = p(w \mid n + n_0)$$

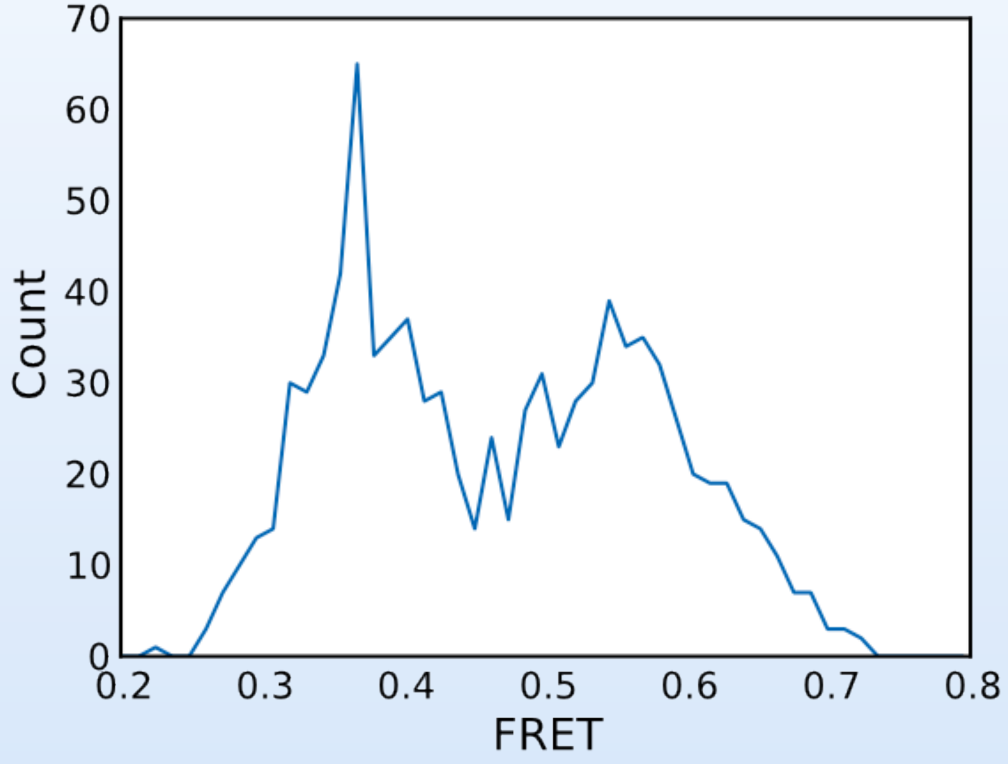
Question: What is best choice for n_0?

Finding States

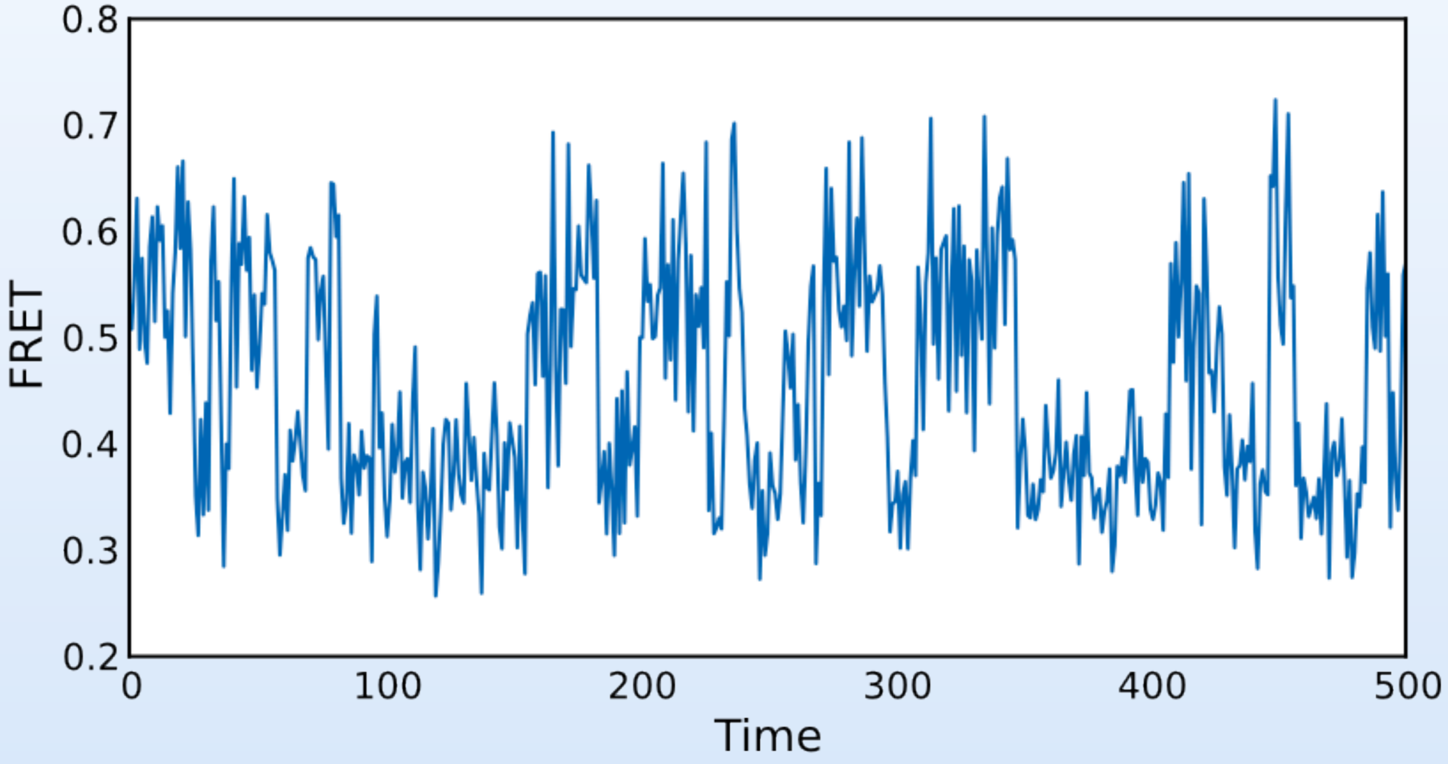
FRET Signal



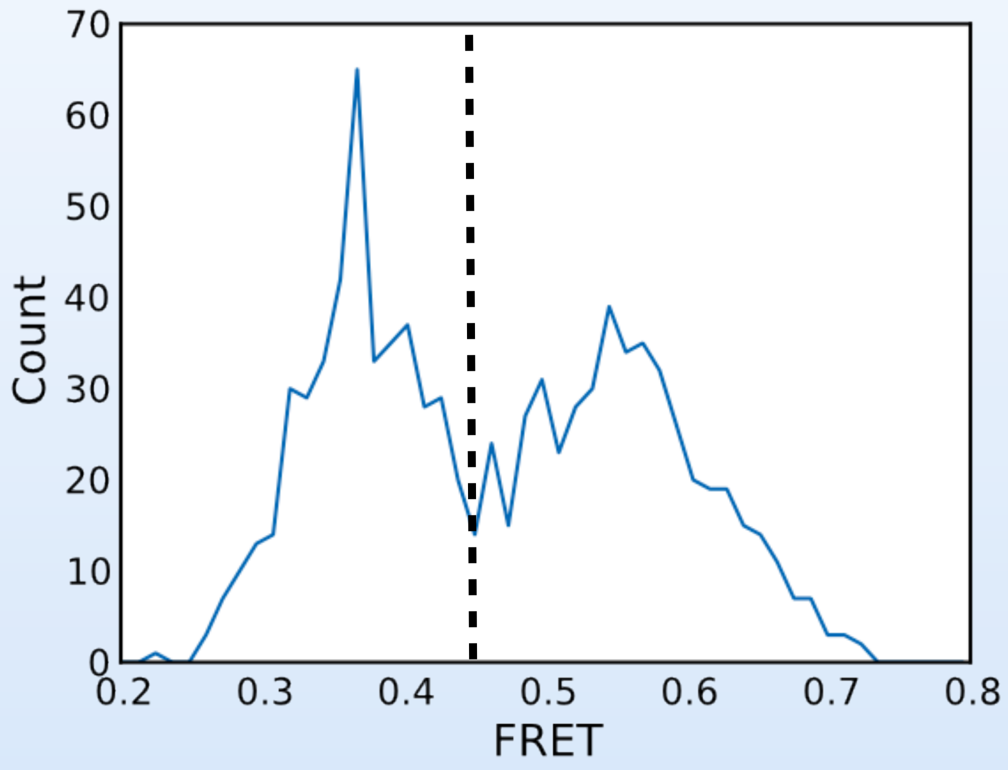
Histogram



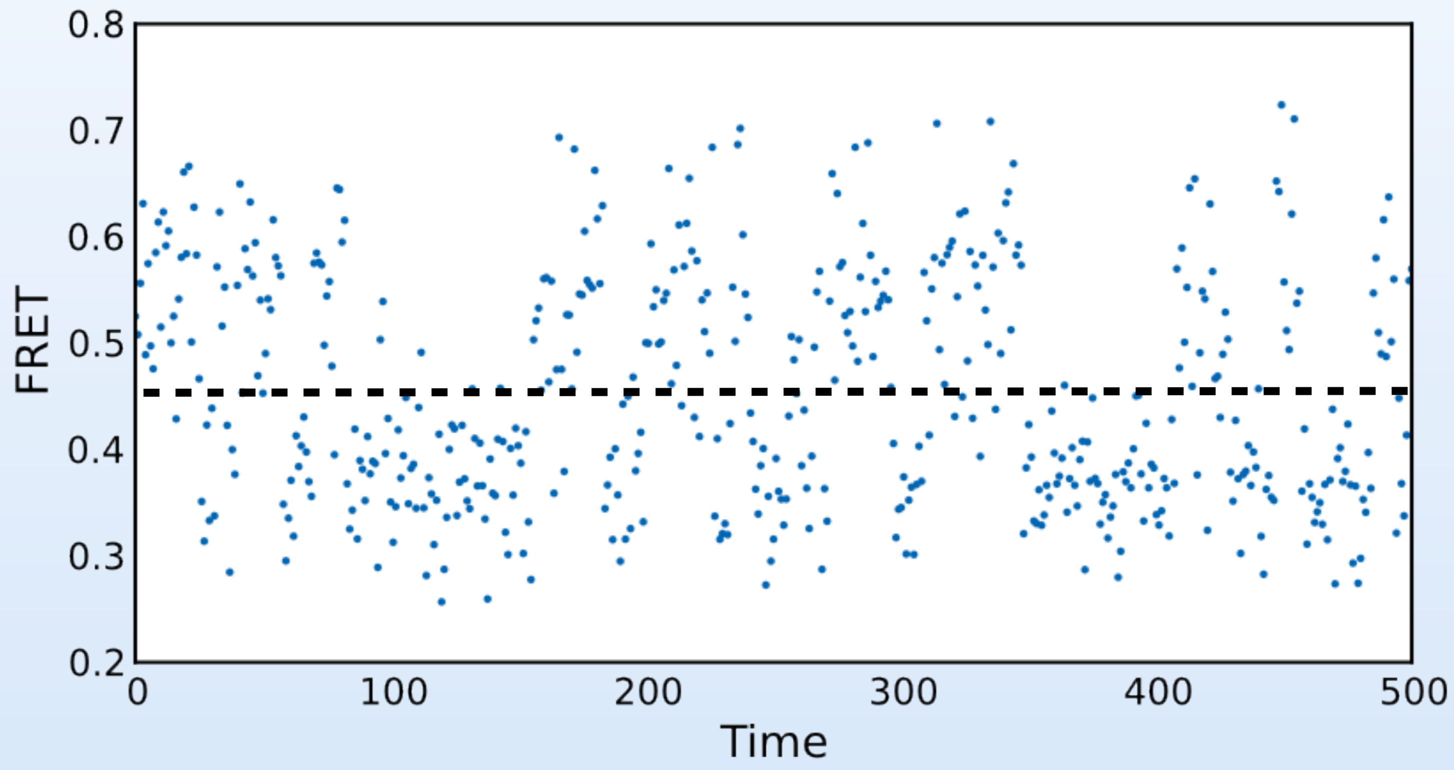
FRET Signal



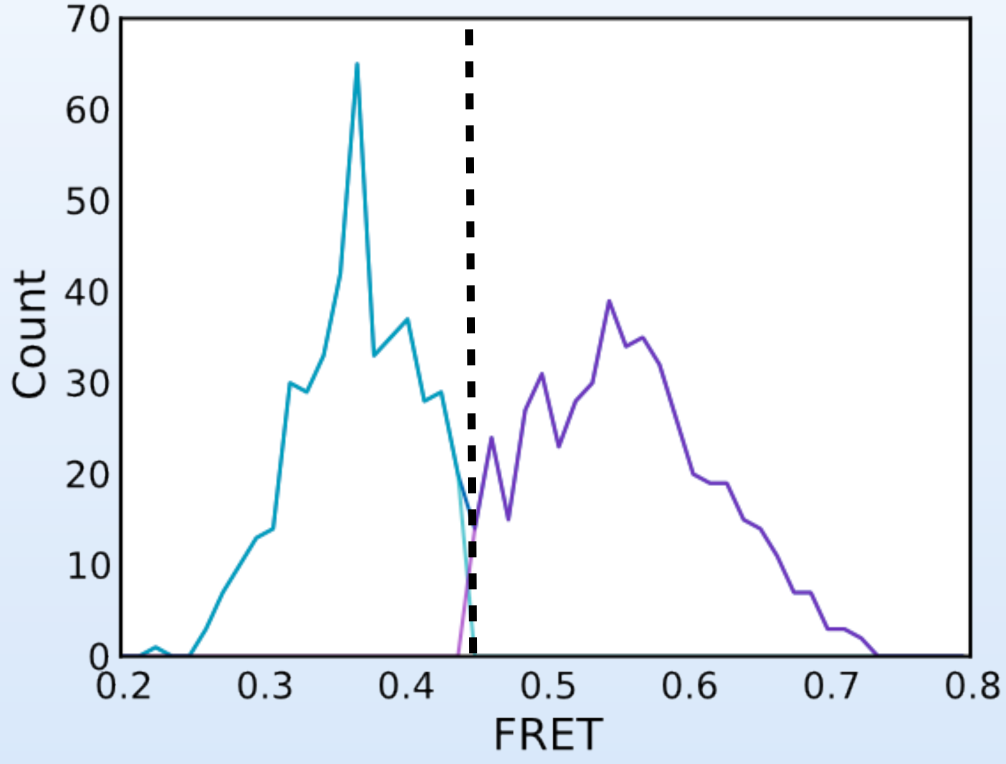
Histogram



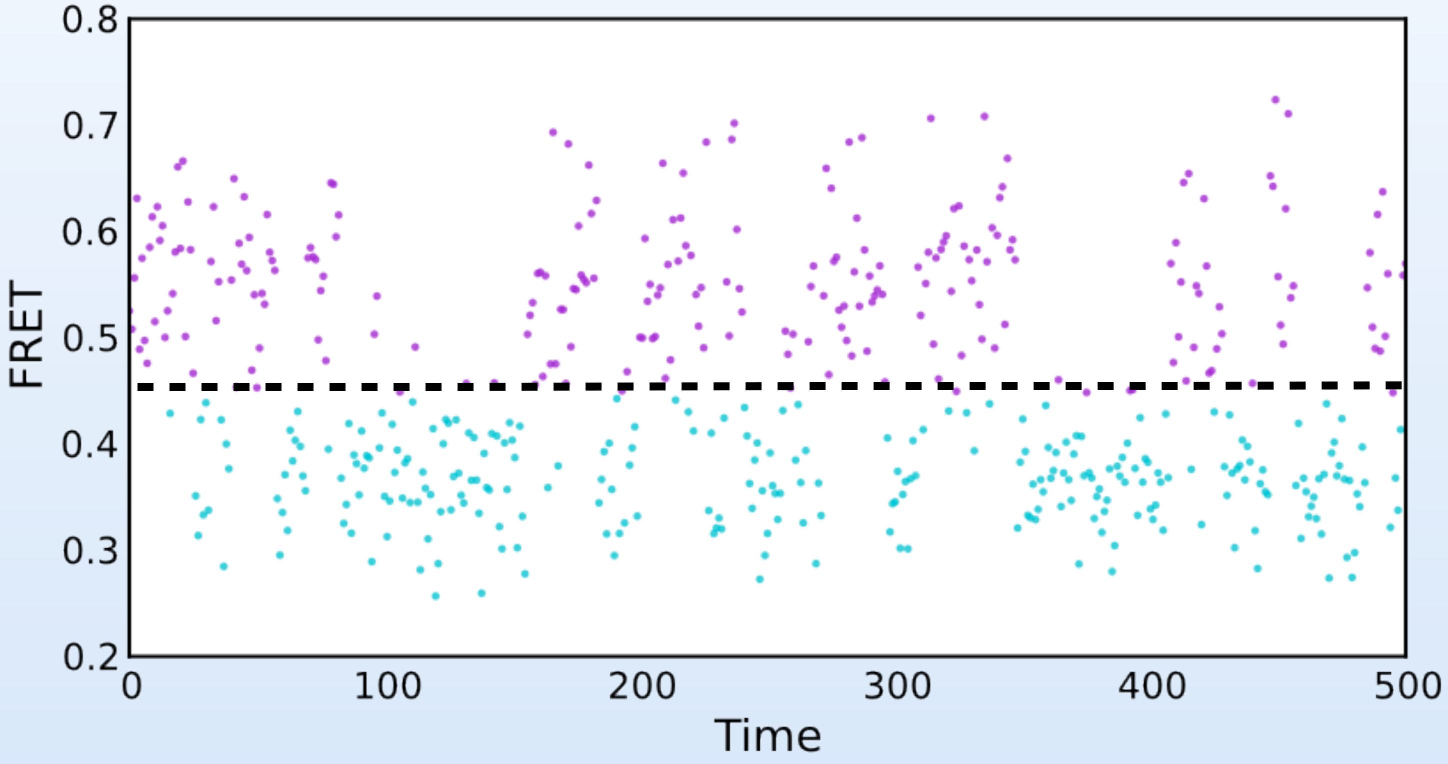
FRET Signal



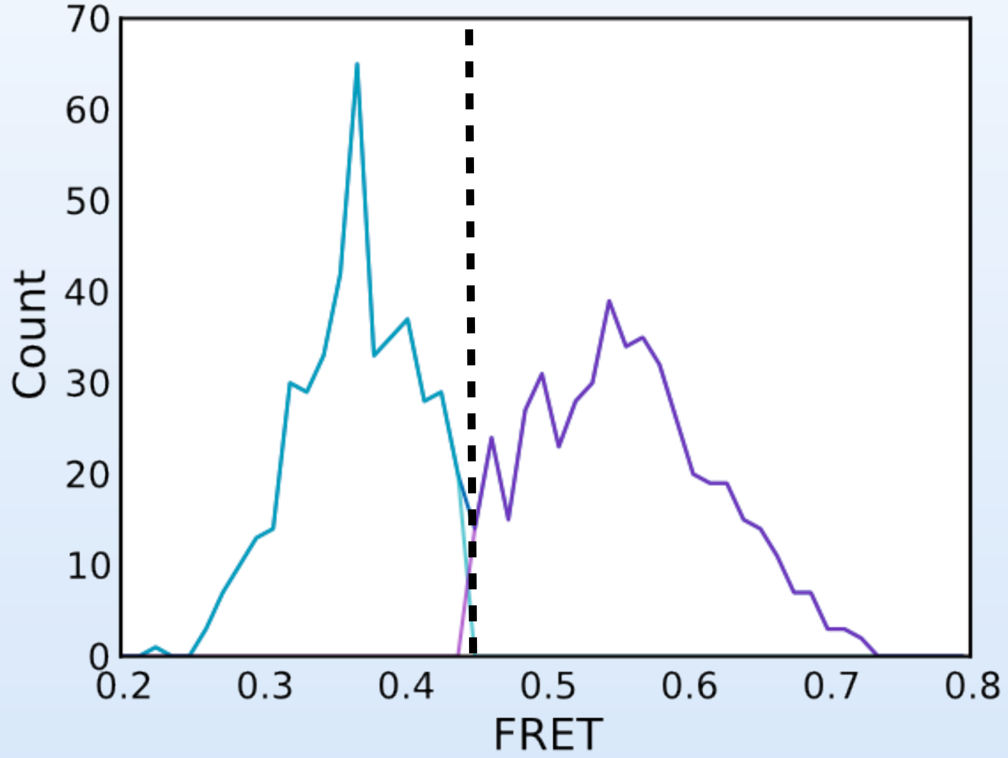
Histogram



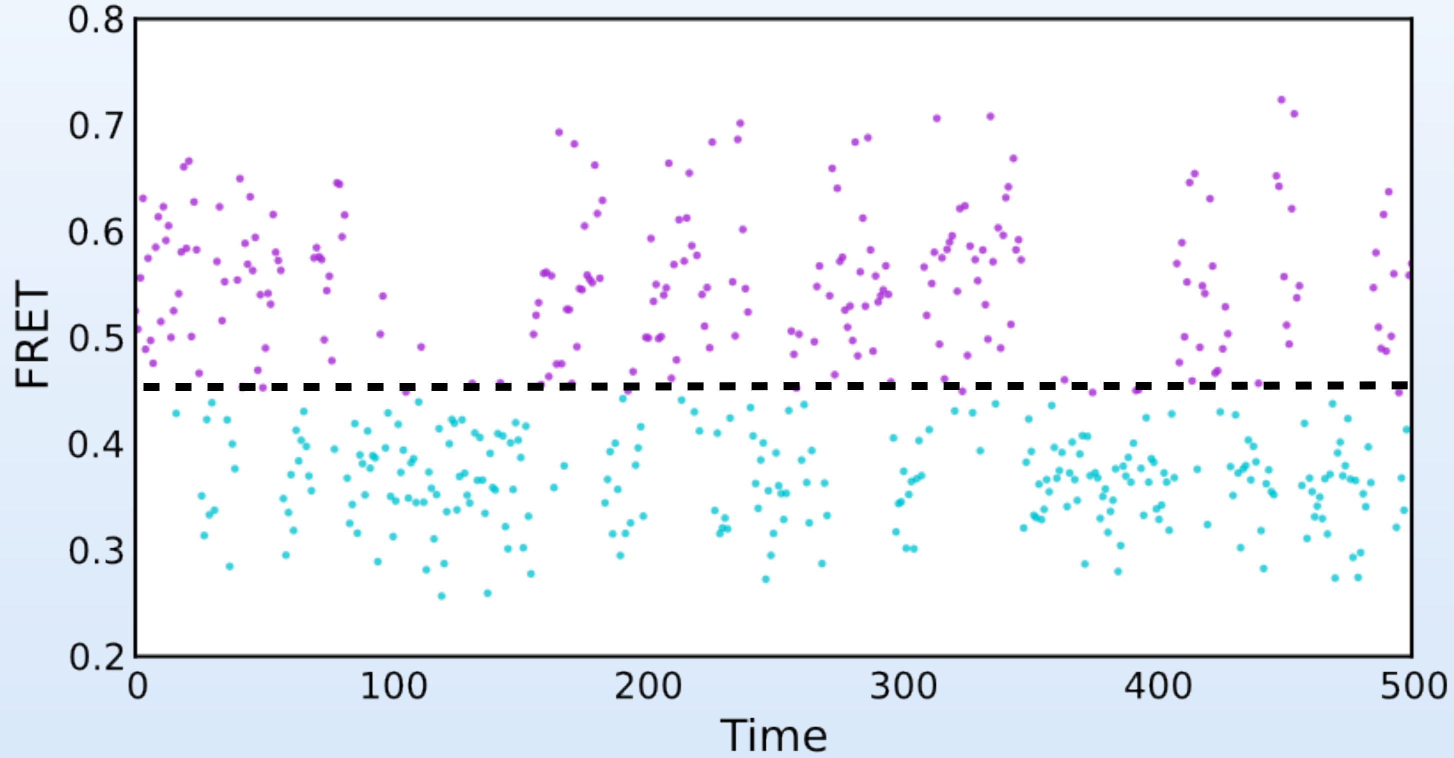
FRET Signal



Histogram

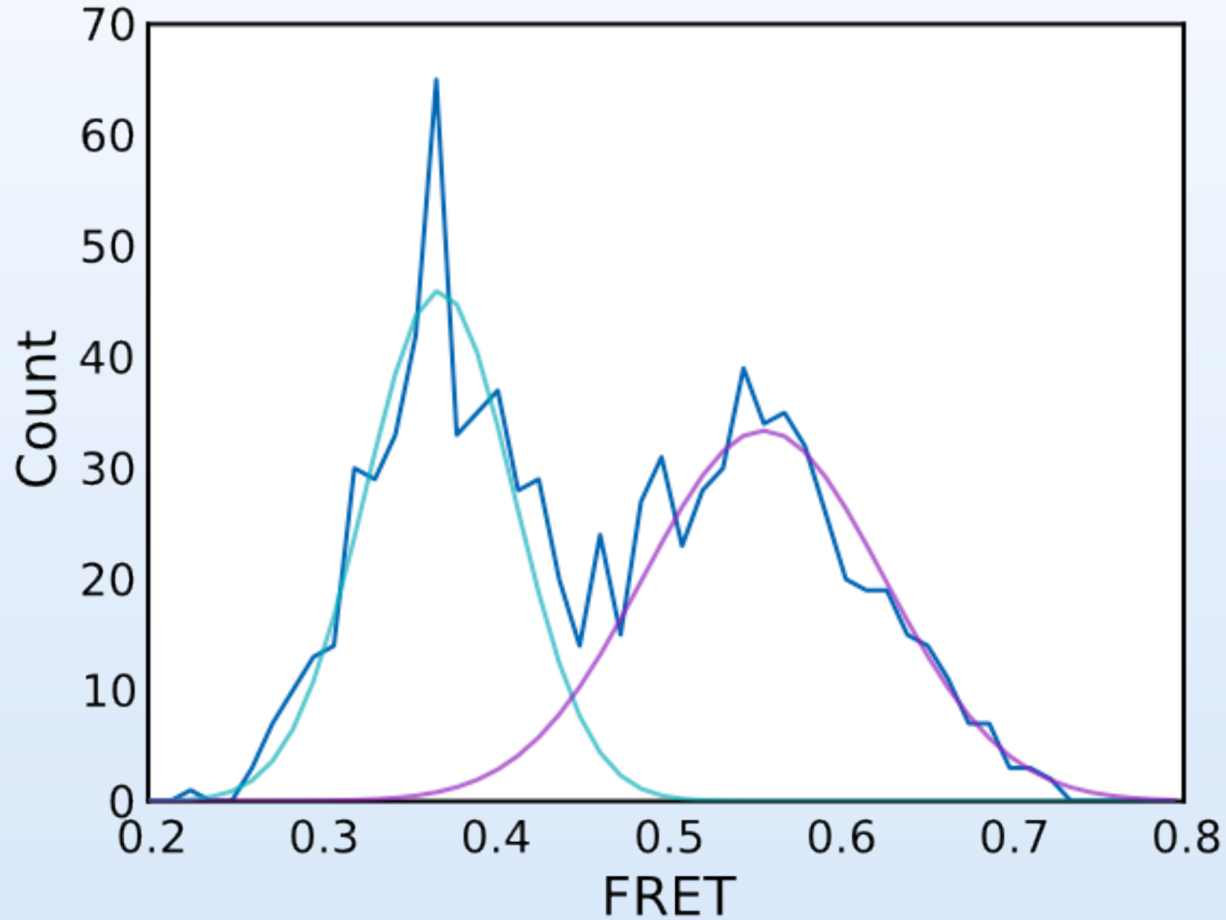


FRET Signal



Idea: Find probability of belonging to each state

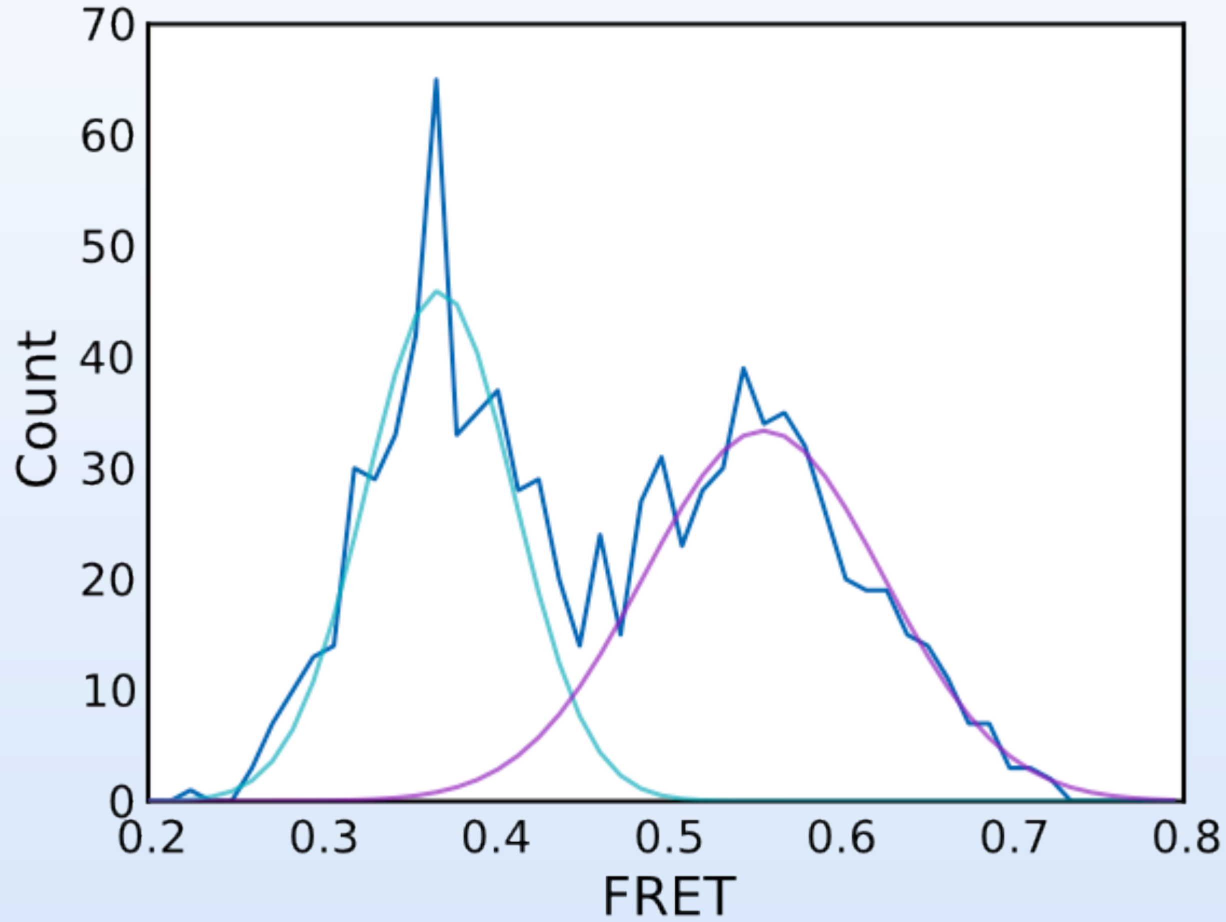
Mixture Model



$$p(x \mid z, \theta)$$

observations (FRET) — latent states — model parameters

Mixture Model



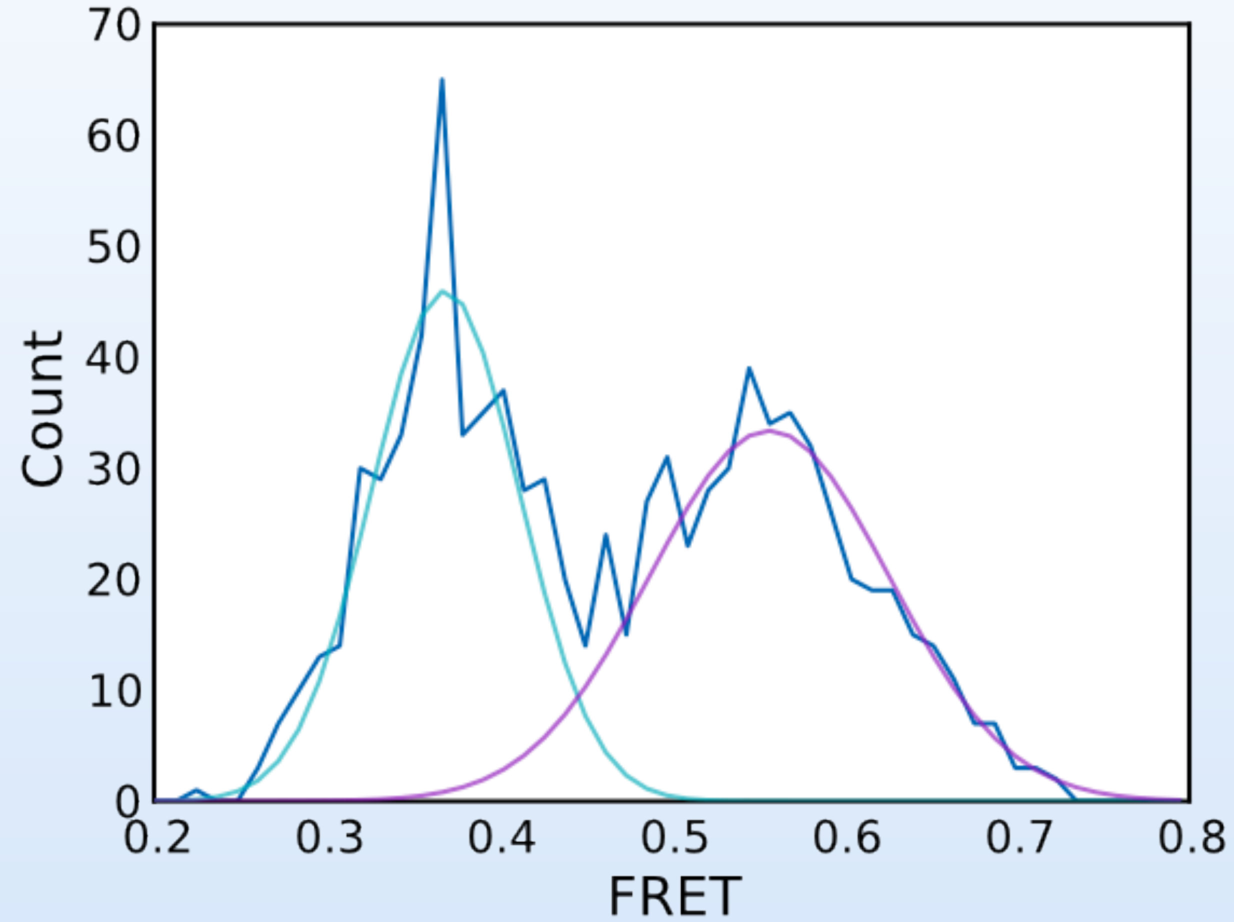
$$p(x \mid z, \theta)$$

observations (FRET) — latent states — model parameters

$$\theta = \{\mu_k, \sigma_k, \pi_k\}$$

center — width — area

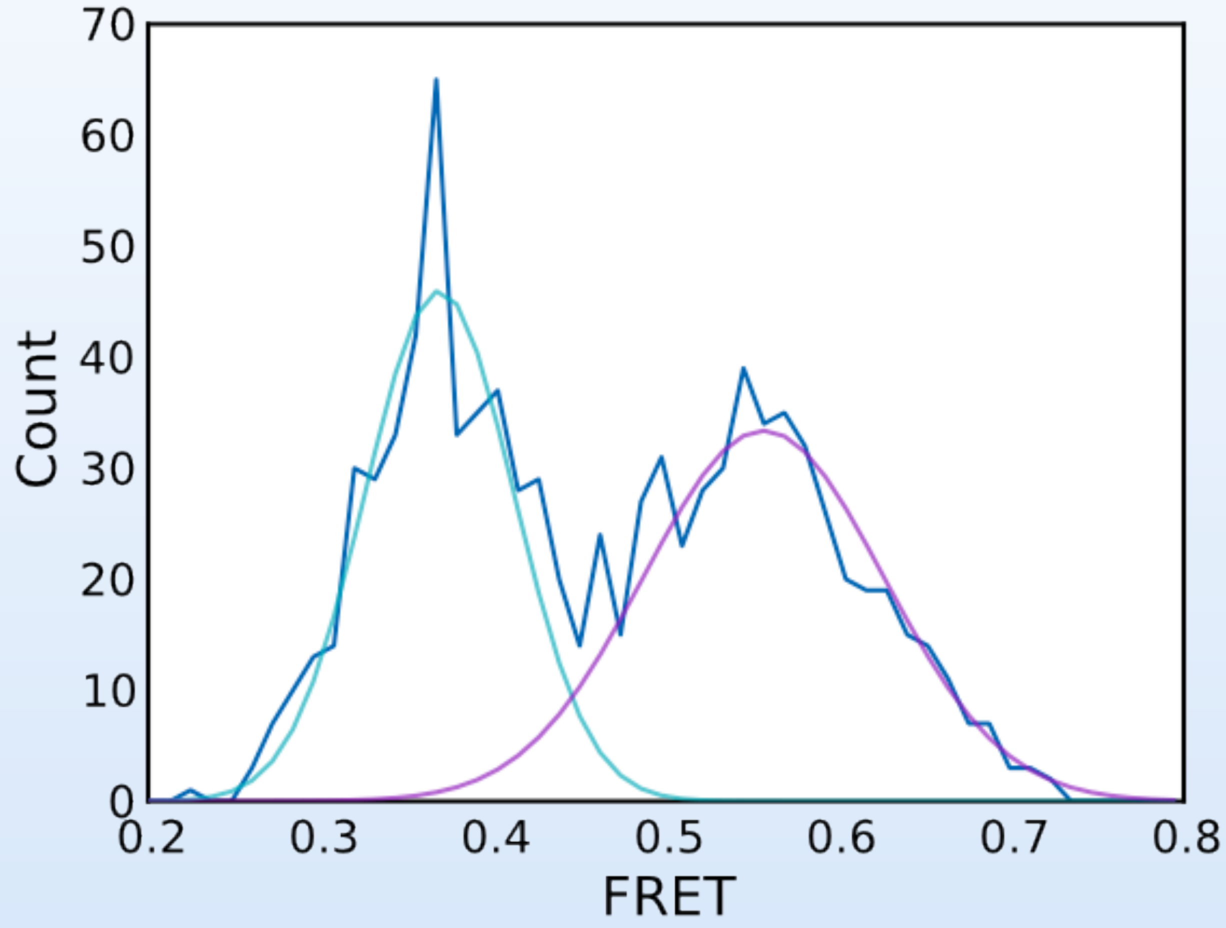
Mixture Model



$$p(z \mid x, \theta) = p(x \mid z, \theta)\, p(z \mid \theta) / p(x \mid \theta)$$

↑ posterior ↑ observations ↑ prior ↑ likelihood

Mixture Model



$$p(z \mid x, \theta) = p(x \mid z, \theta)\, p(z \mid \theta) / p(x \mid \theta)$$

↑ posterior ↑ observations ↑ prior ↑ likelihood

Maximum Likelihood

$$p(x \mid \theta) = \sum_z p(x, z \mid \theta)$$

Likelihood

Expectation Maximization

1. calculate $p(z \mid x, \theta^i)$
2. calculate θ^{i+1} from $p(z \mid x, \theta^i)$

Maximum Likelihood

$$L = \log p(x \mid \theta) = \log \left[\sum_{z} p(x, z \mid \theta) \right]$$

Log-Likelihood

Expectation Maximization

1. calculate $p(z \mid x, \theta^{i})$
2. calculate θ^{i+1} from $p(z \mid x, \theta^{i})$

Maximum Likelihood

$$L = \log p(x \mid \theta) = \log\left[\sum_z p(x, z \mid \theta)\right]$$

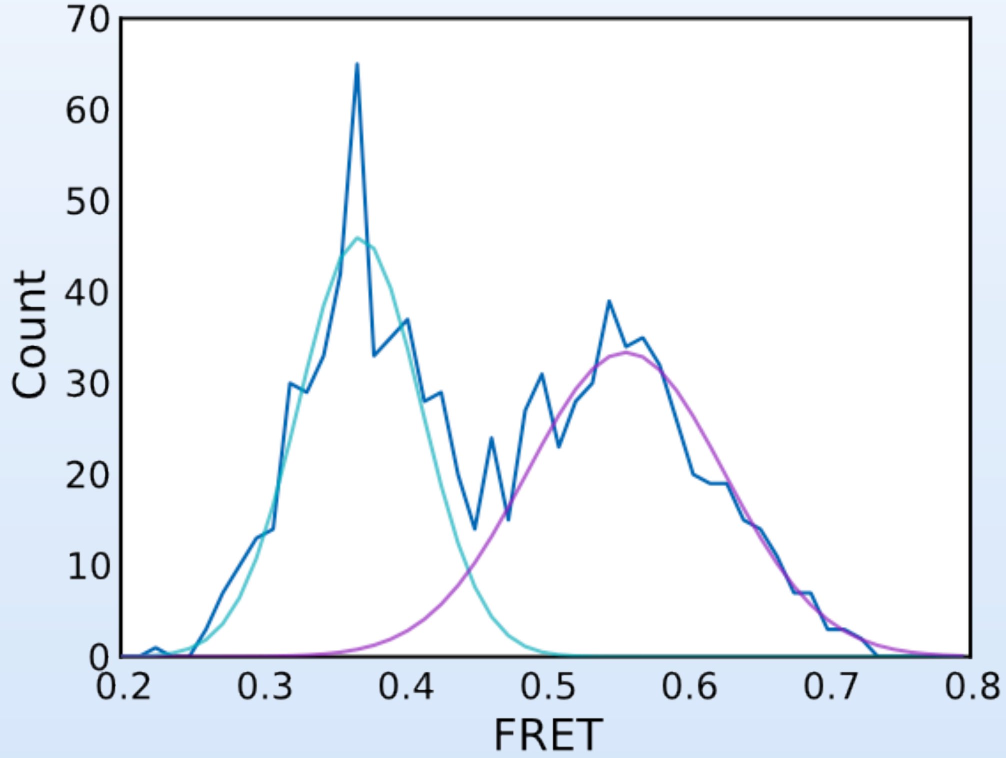
Log-Likelihood

Expectation Maximization

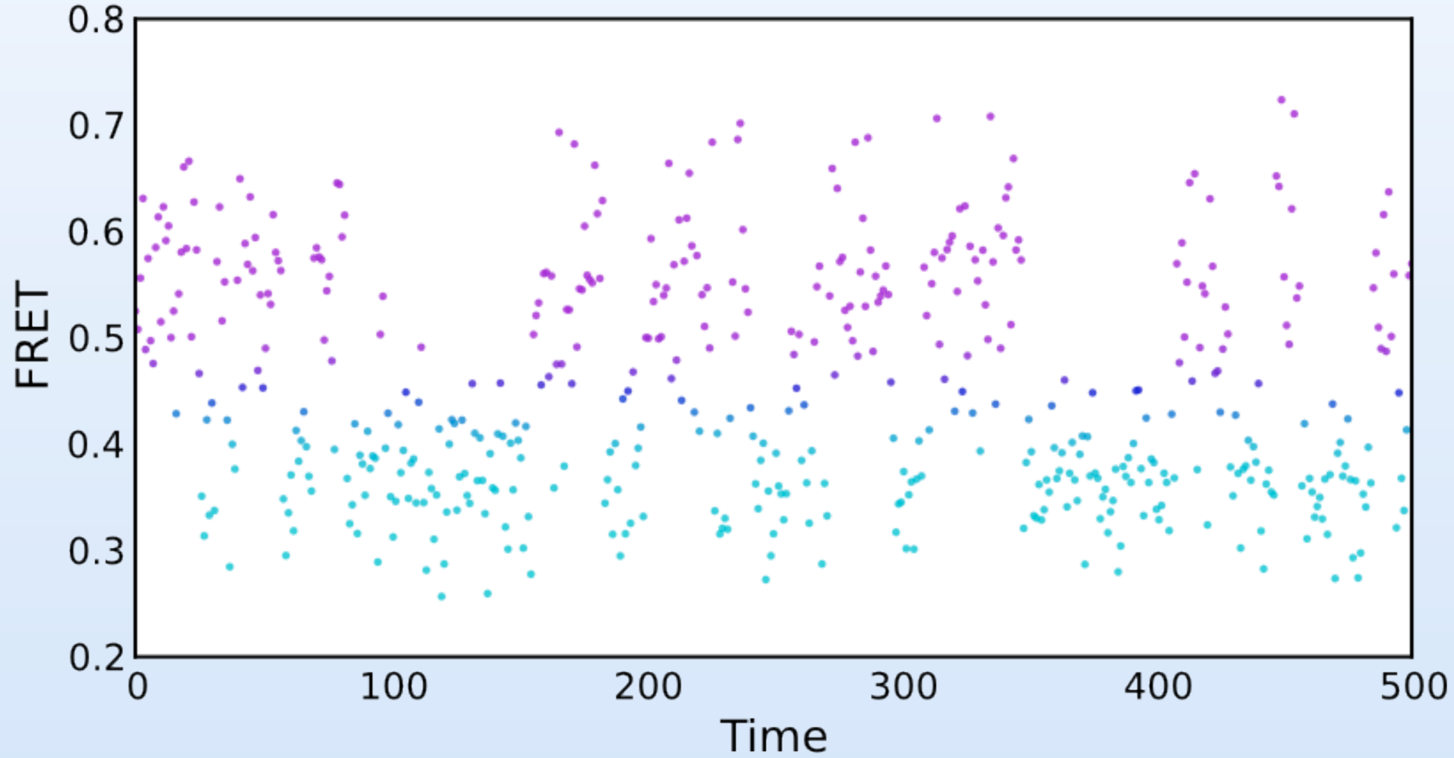
1. calculate $p(z \mid x, \theta^i)$
2. solve $\partial L / \partial \theta = 0$

Gaussian Mixture Model

Histogram



FRET Signal

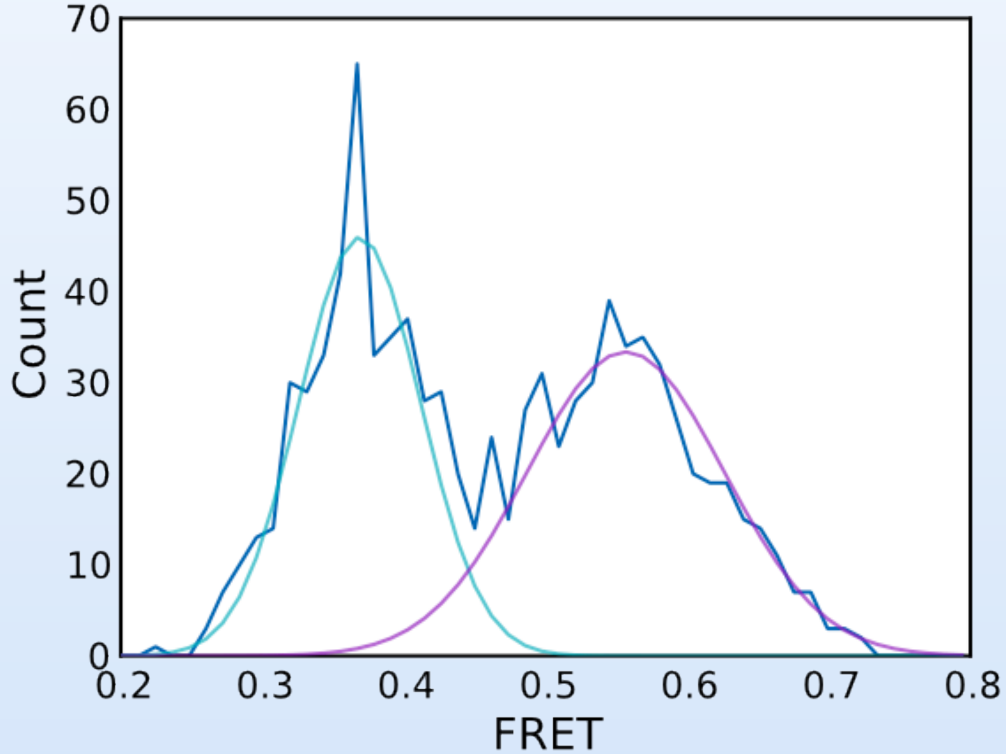


We've learned:

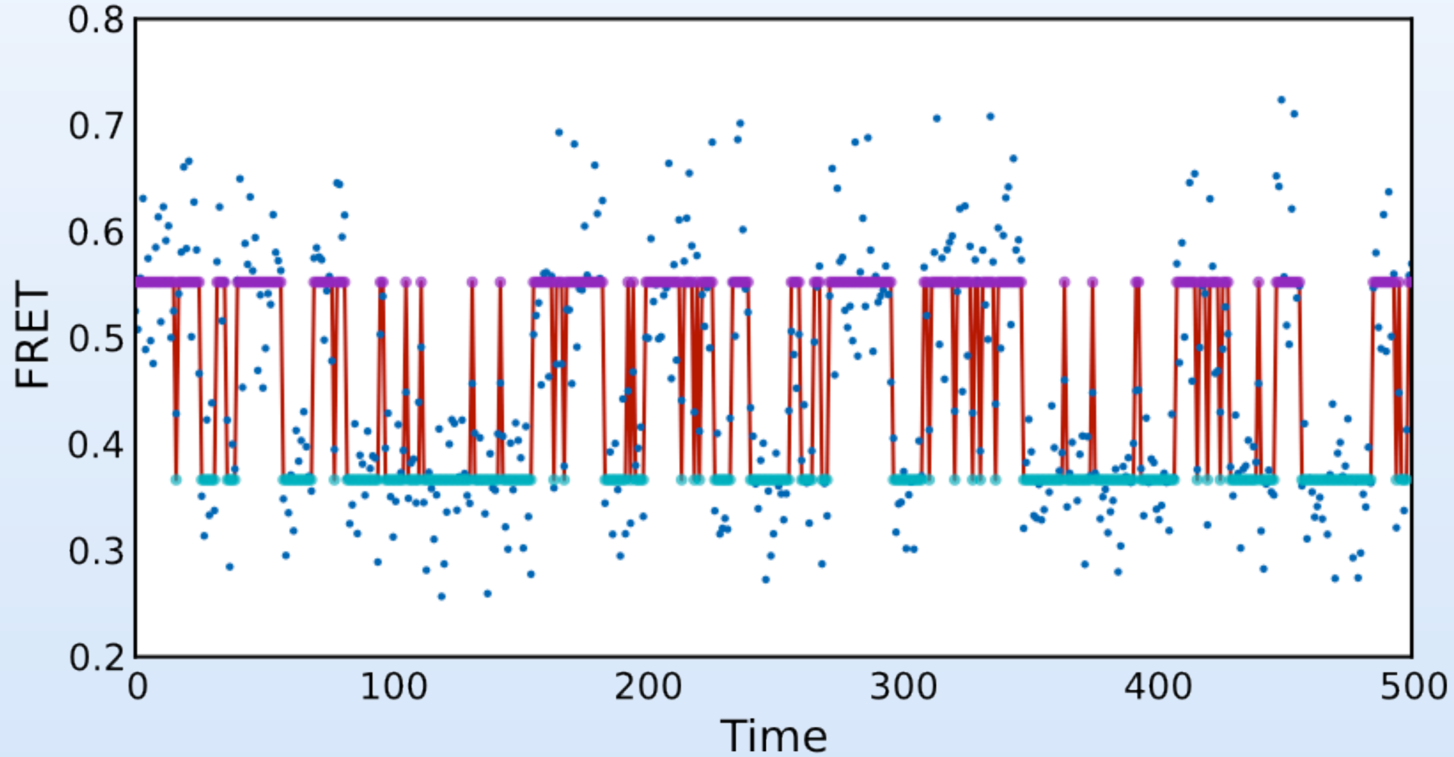
parameters: $\theta = \{\mu, \sigma, \pi\}$ states: $p(z \mid x, \theta)$

Gaussian Mixture Model

Histogram



FRET Signal

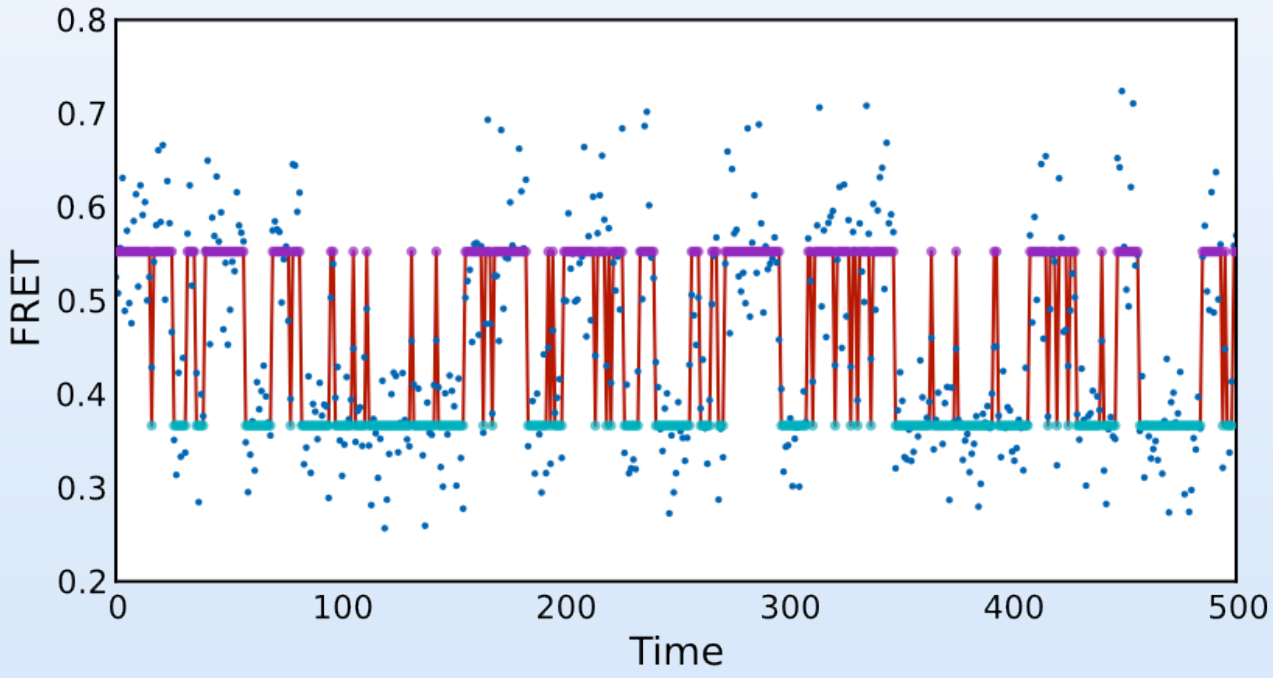


We've learned:

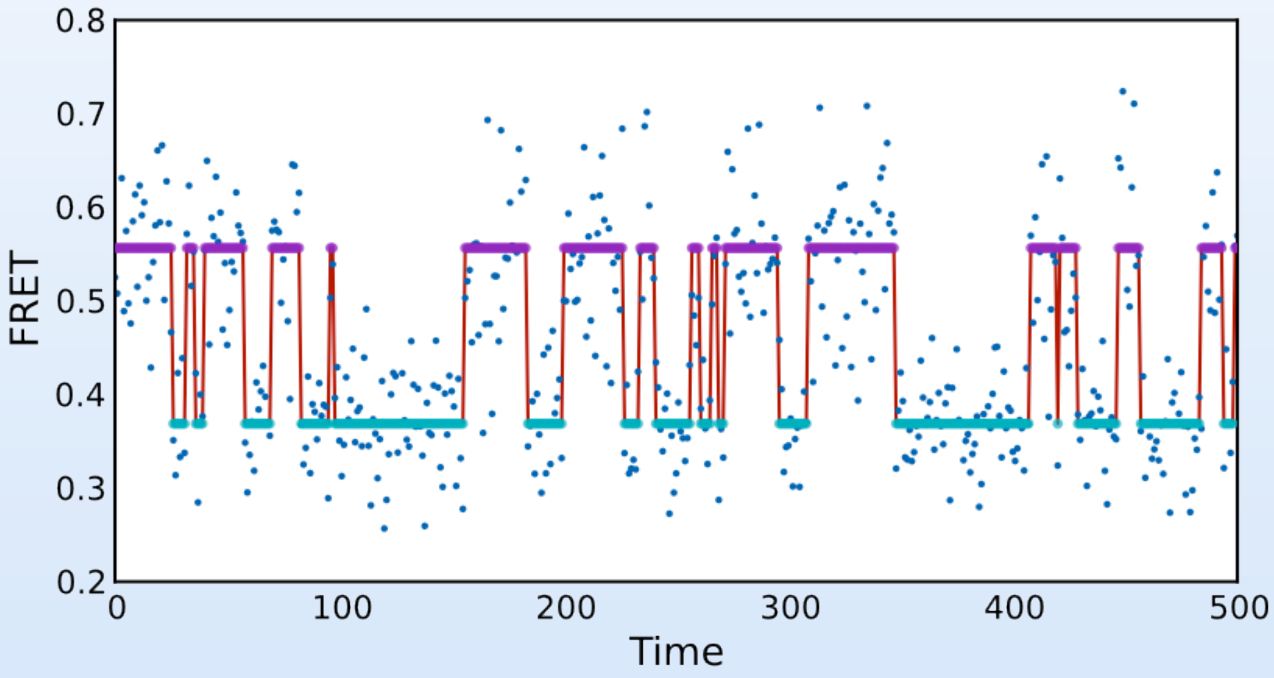
parameters: $\theta = \{\mu, \sigma, \pi\}$ states: $p(z \mid x, \theta)$

Gaussian Mixture Model

Learned



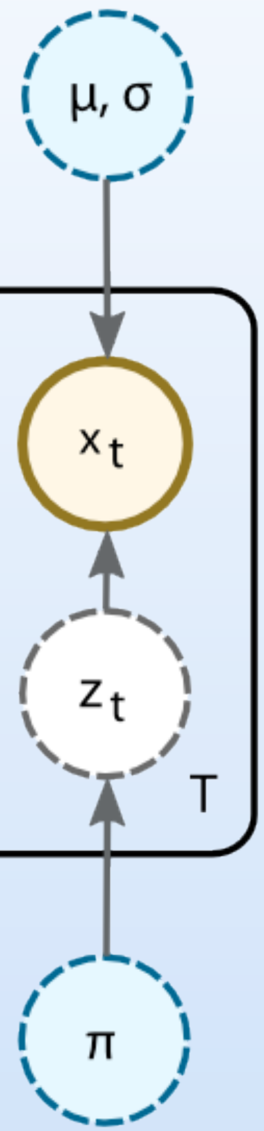
Truth



Accurate for occupancy of states,
not so good for rate estimates

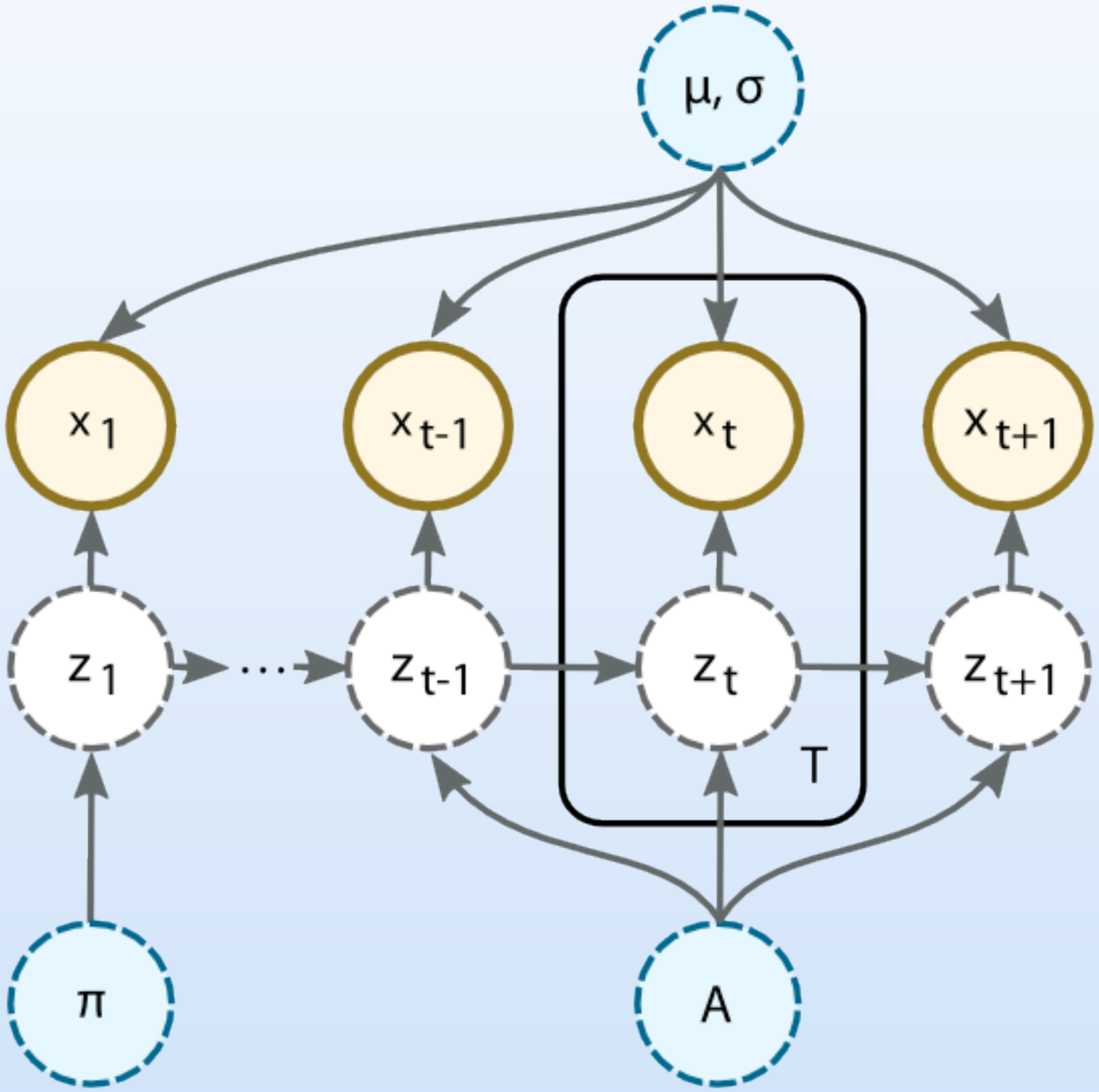
Learning Rates

Graphical Models



$$p(x, z \mid \mu, \sigma, \pi) = p(x \mid z, \mu, \sigma)\, p(z \mid \pi)$$

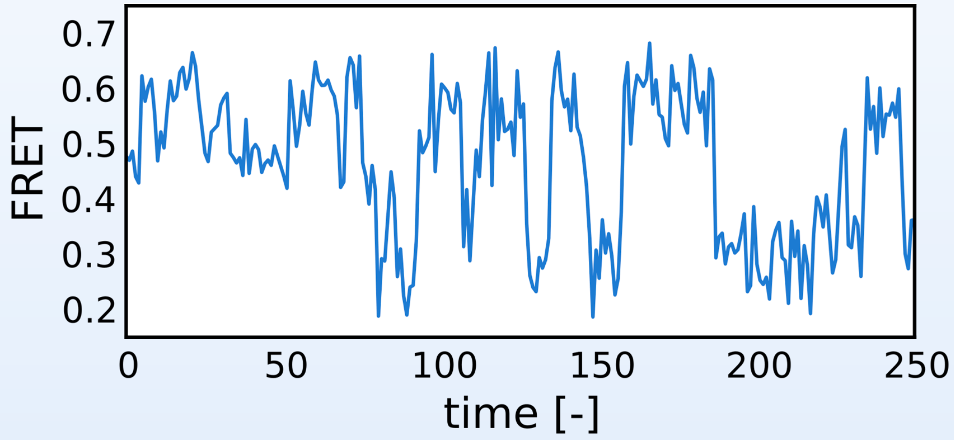
Hidden Markov Model

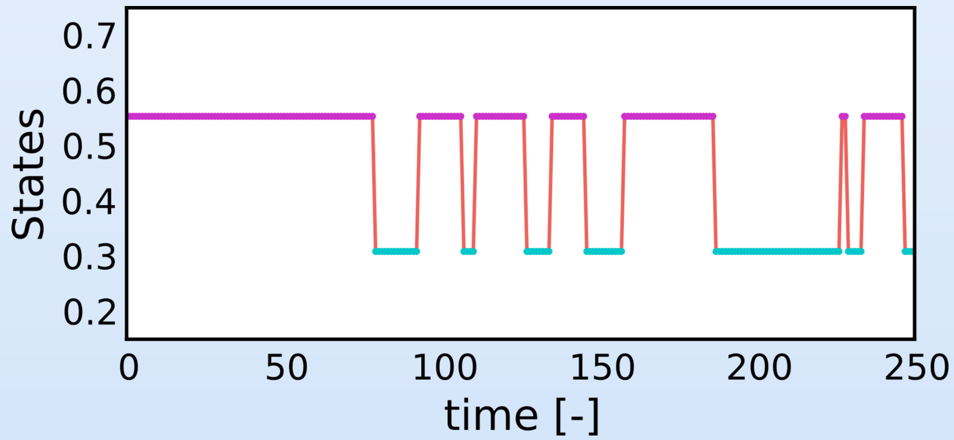


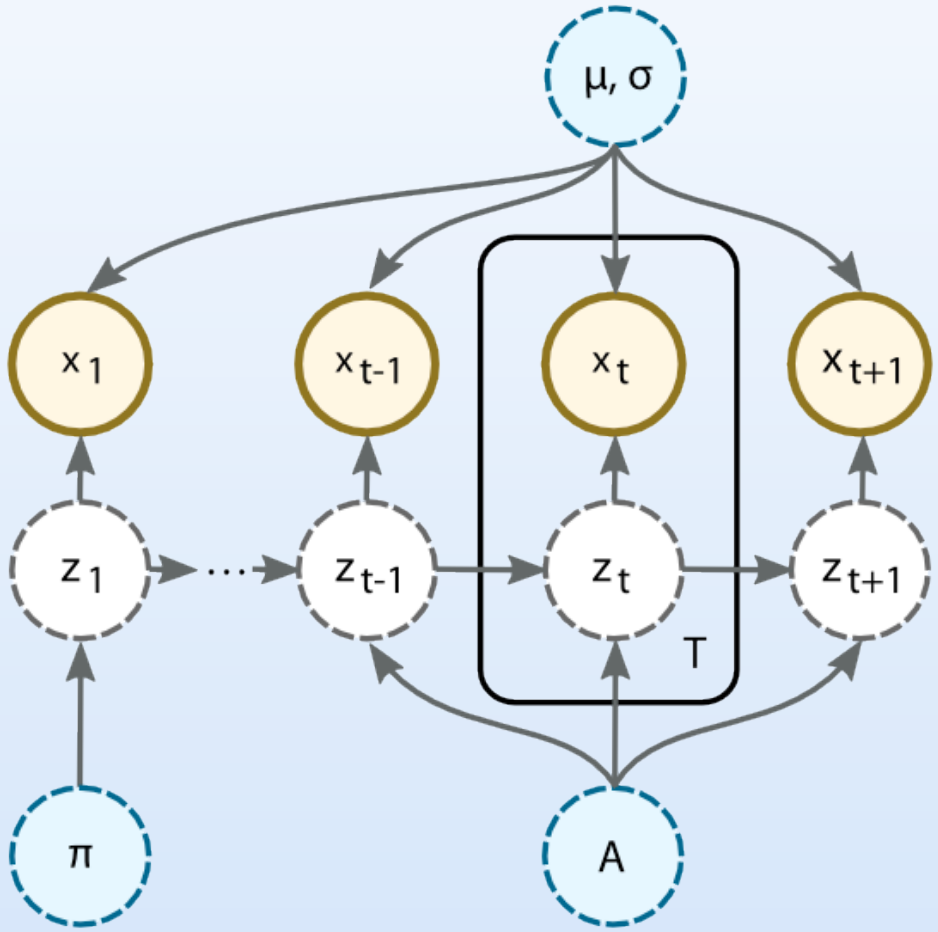
probability of state depends on previous state

$$p(z_{t+1}=l \mid z_t=k) = A_{kl}$$

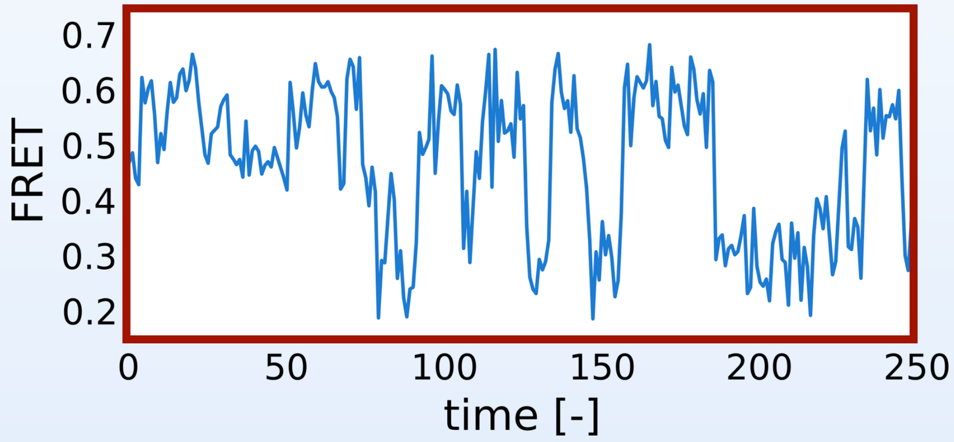
Hidden Markov Model

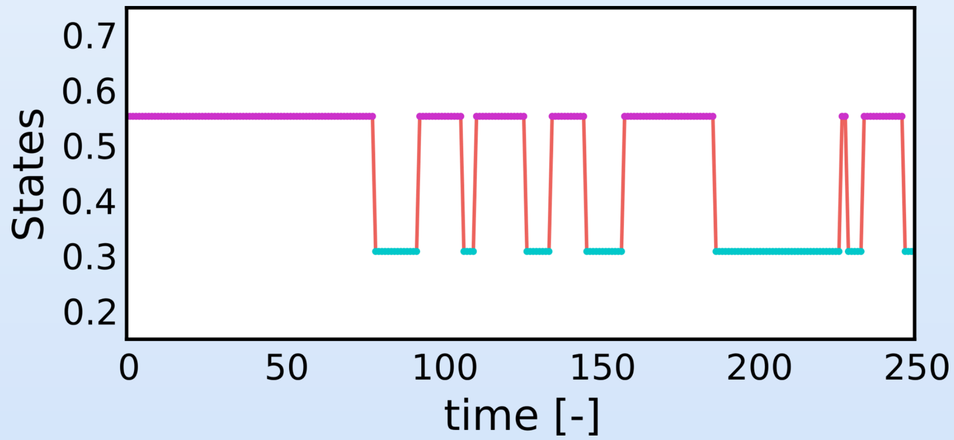


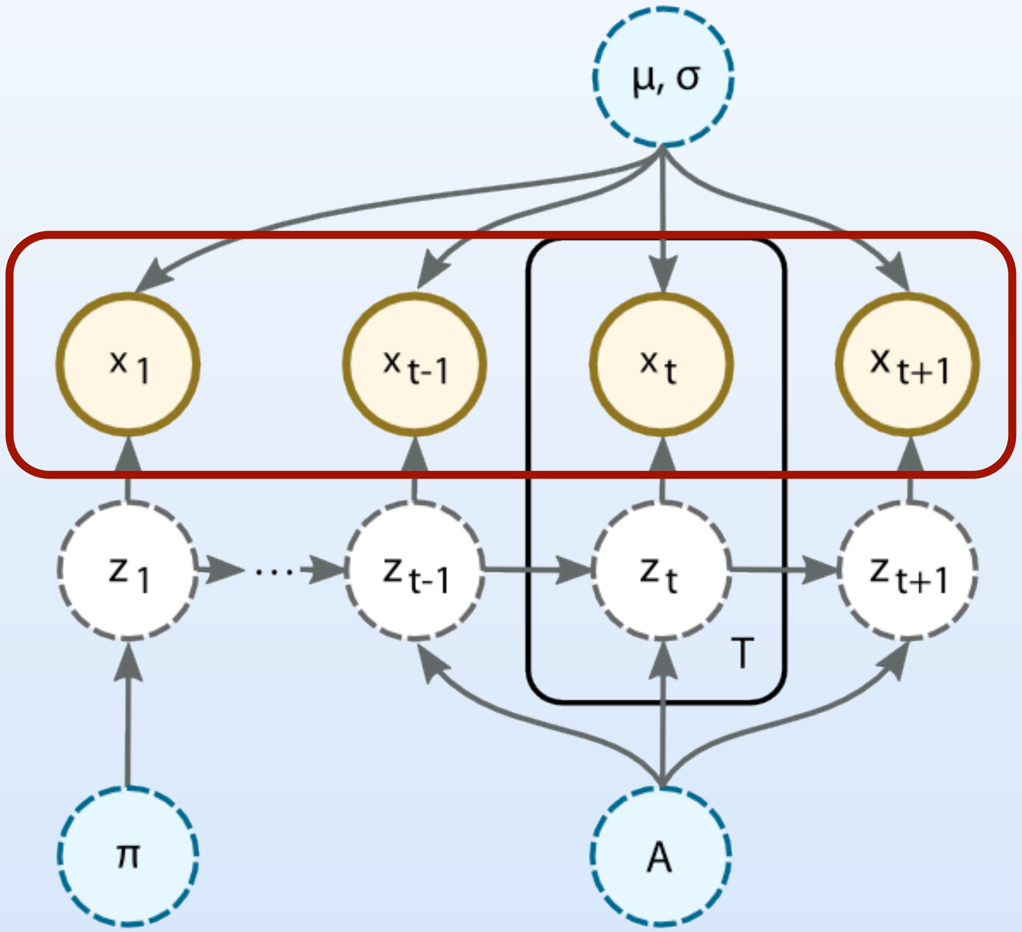




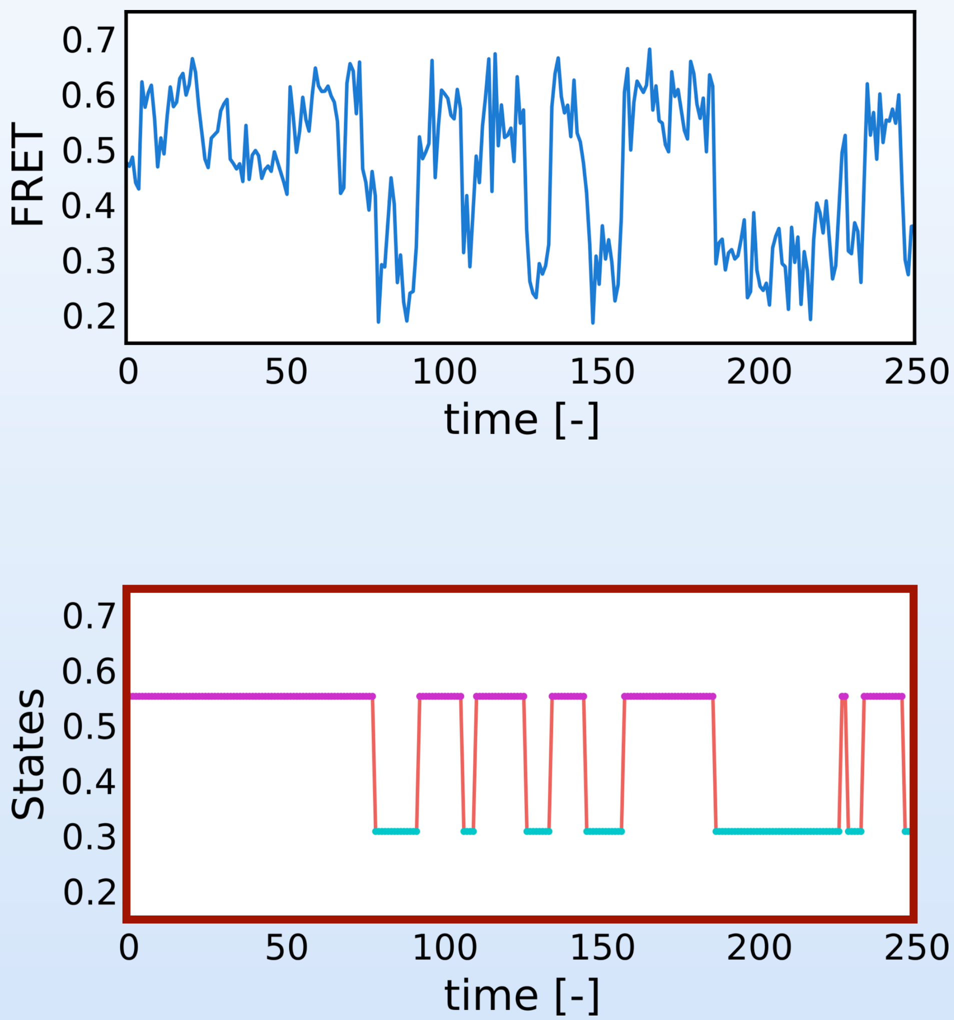
Hidden Markov Model

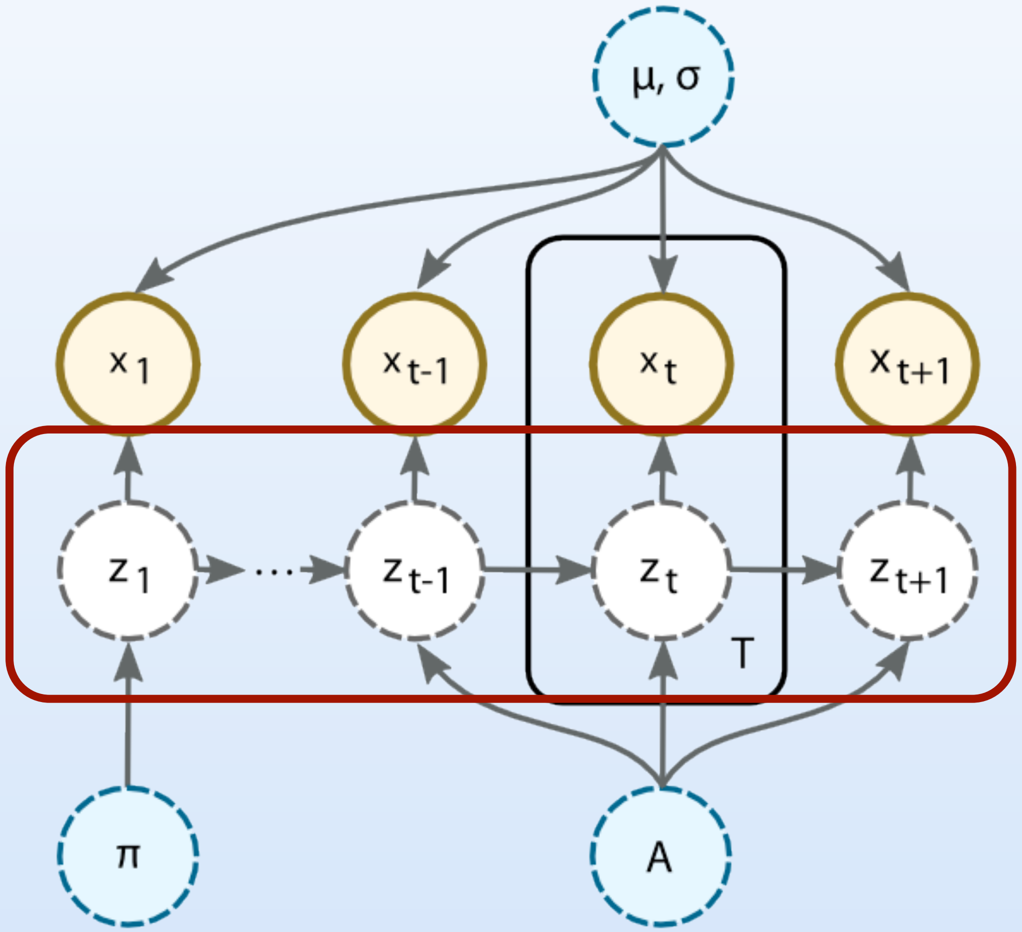




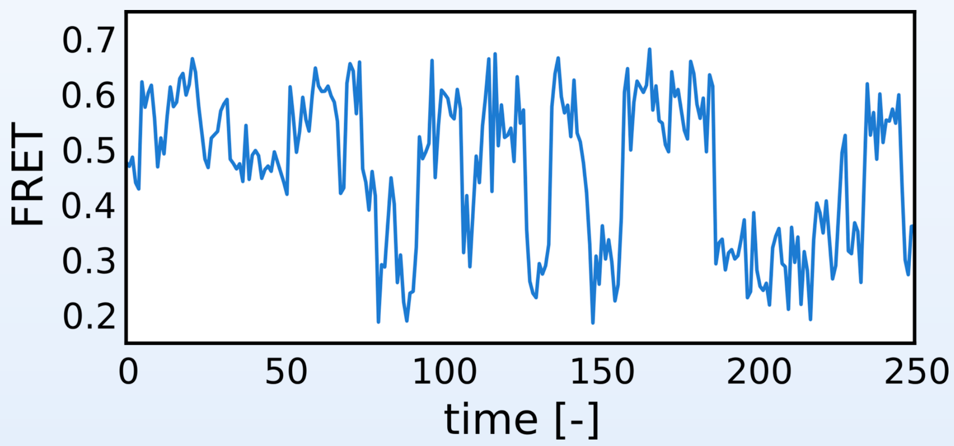


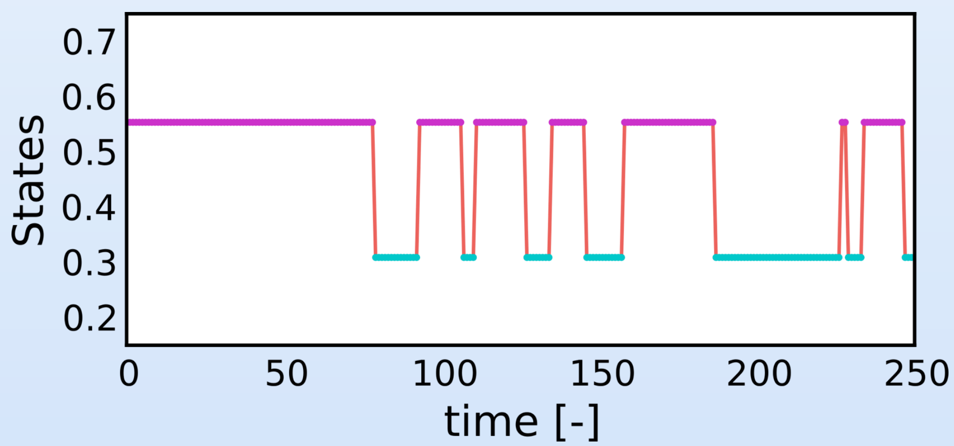
Hidden Markov Model

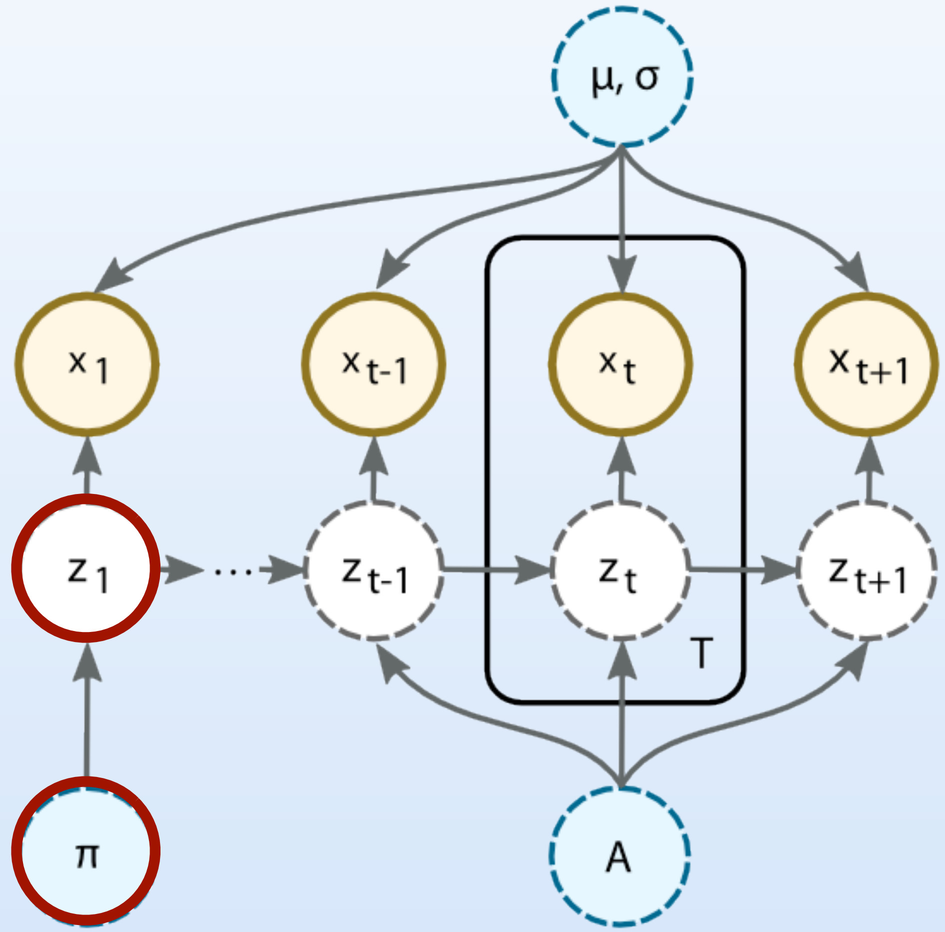




Hidden Markov Model

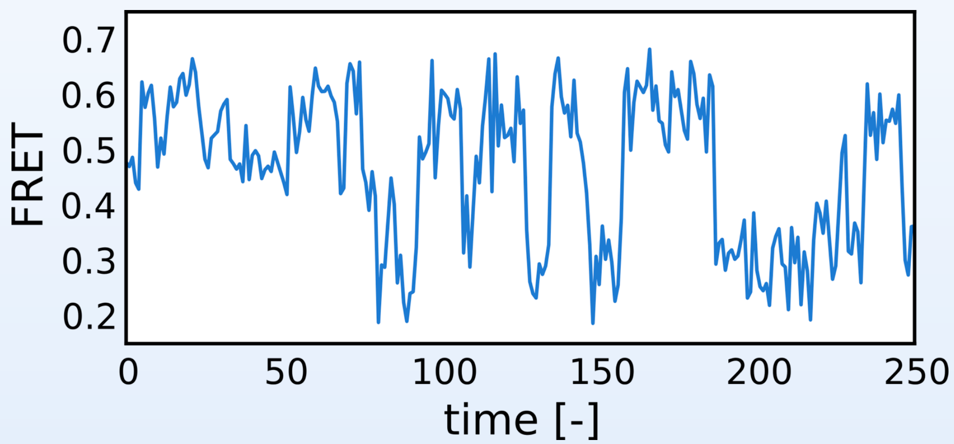


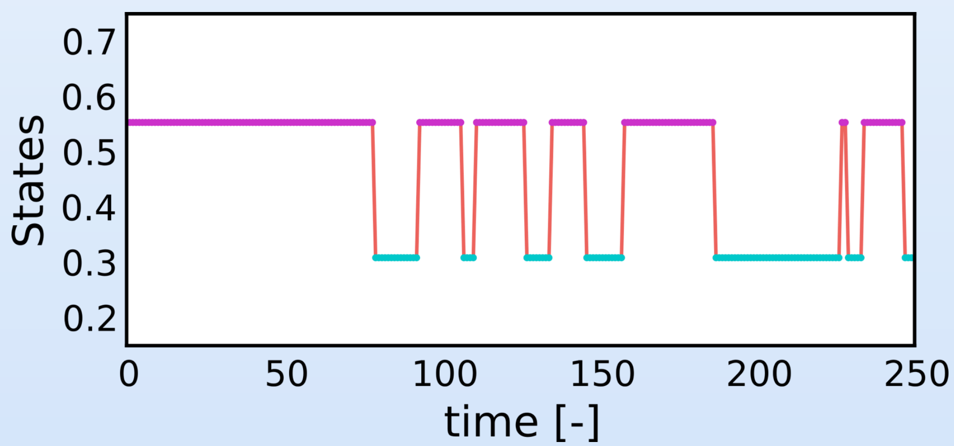


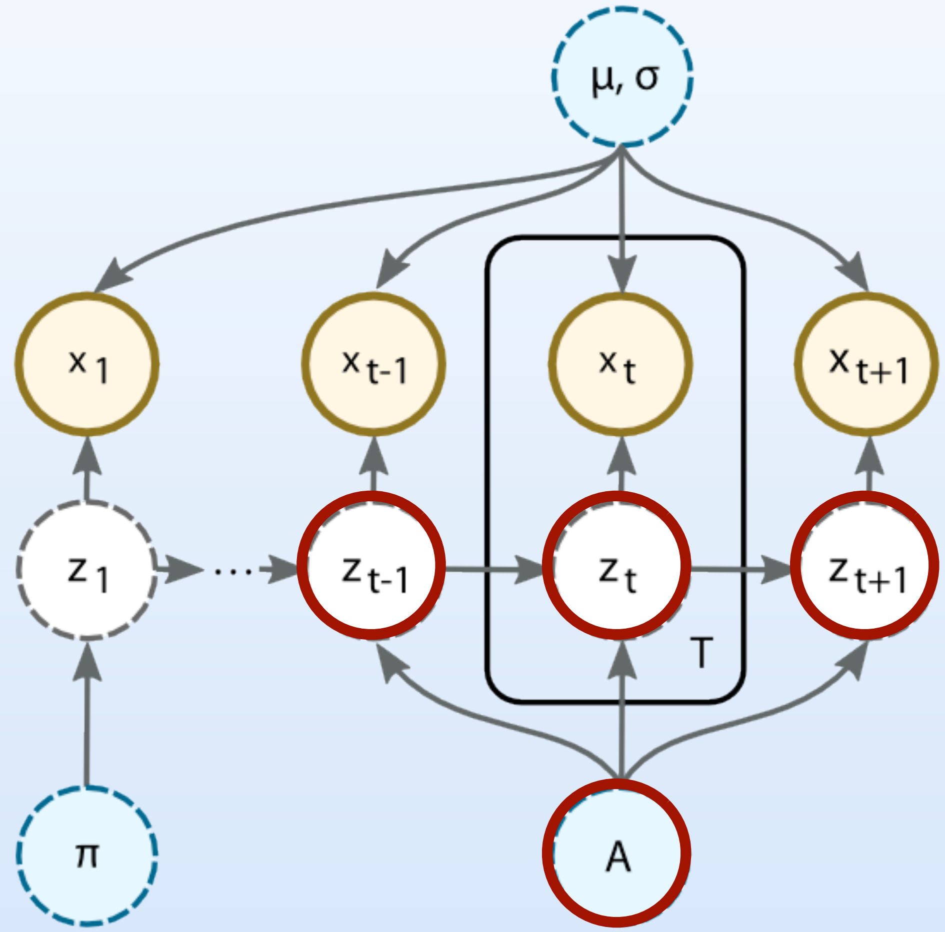


$$p(z_1 = k) = \pi_k$$

Hidden Markov Model

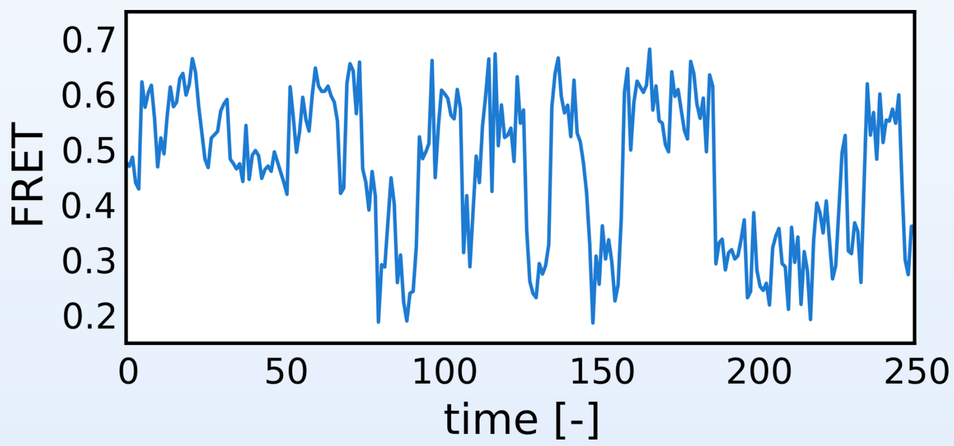


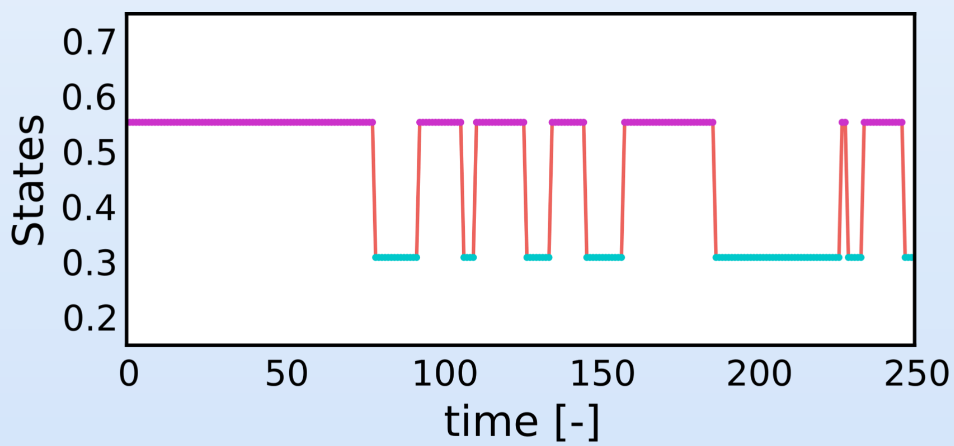


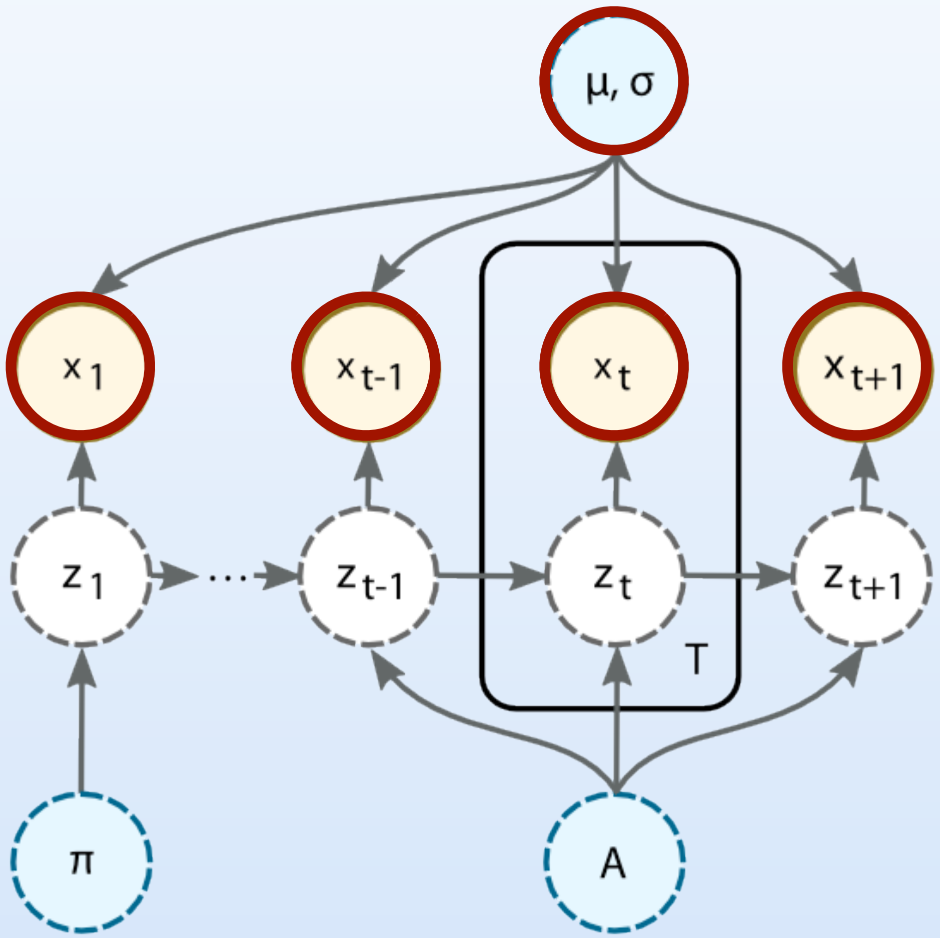


$$p(z_{t+1} = l \mid z_t = k) = A_{kl}$$

Hidden Markov Model



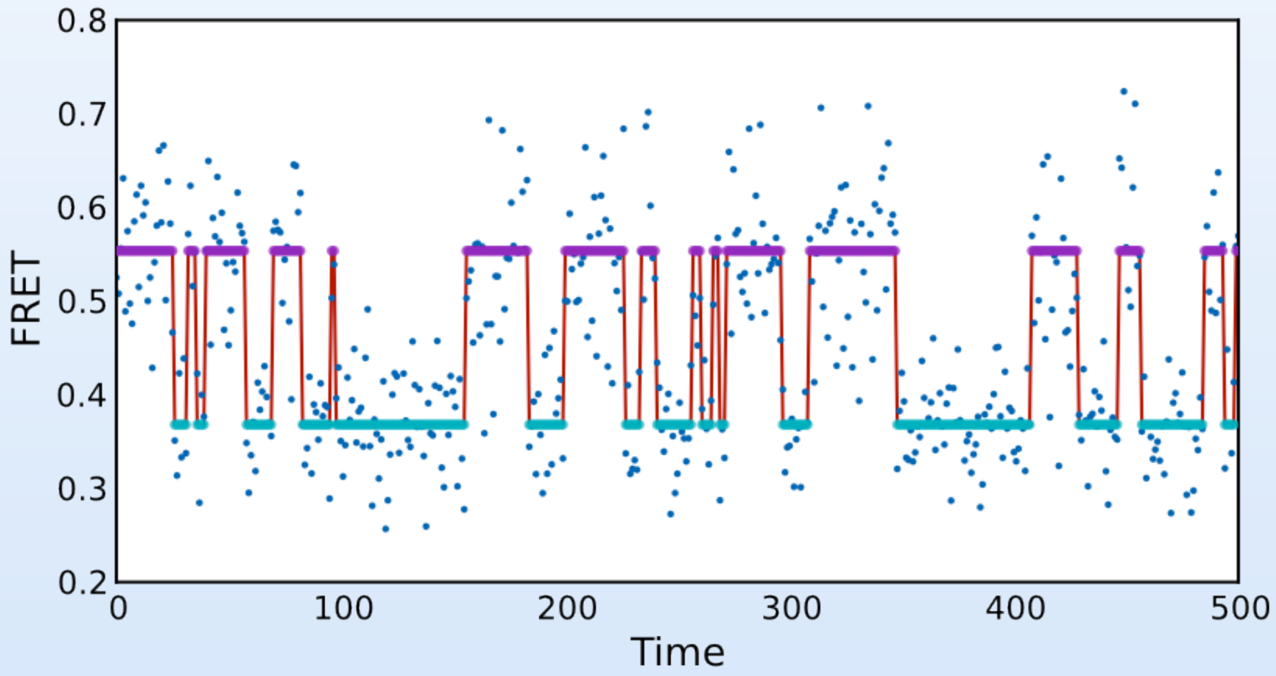




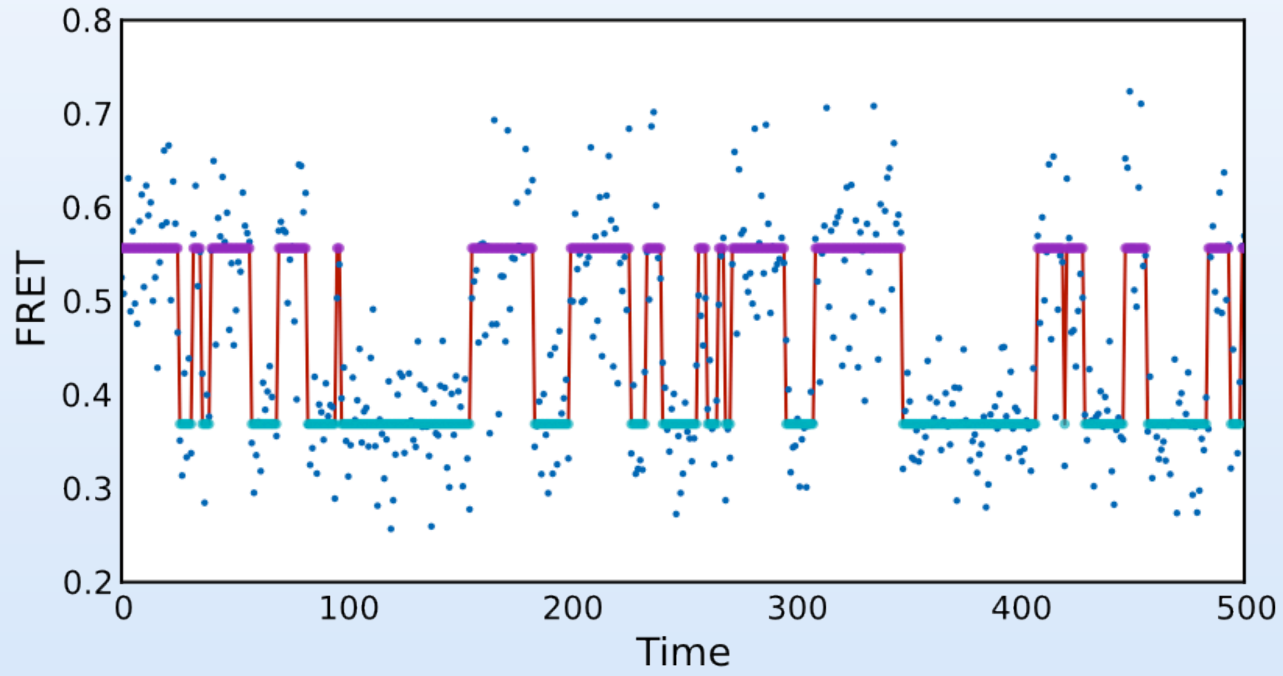
$$p(x_t \mid z_t = k) = N(x_t \mid \mu_k, \sigma_k)$$

Hidden Markov Model

Learned



Real



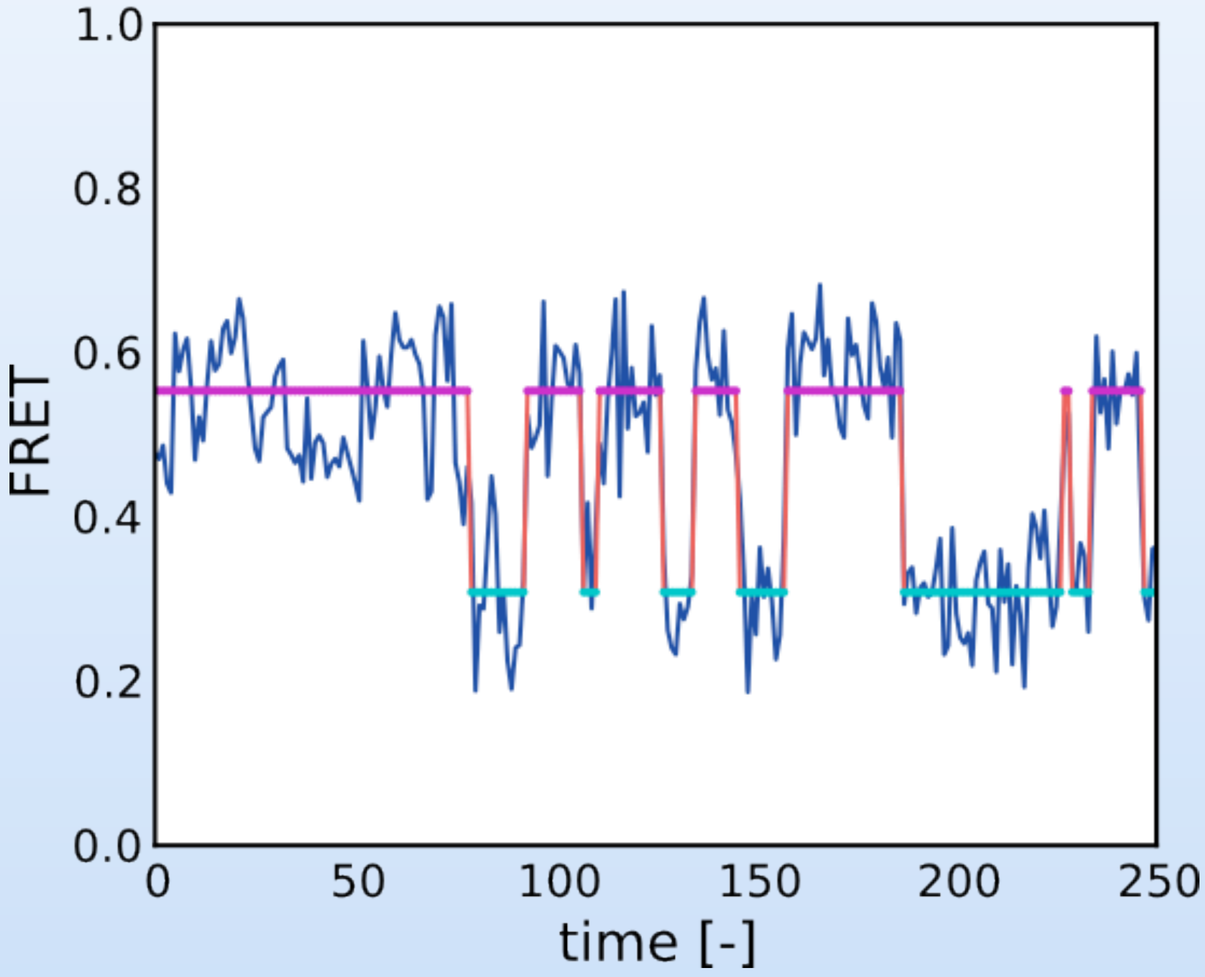
We've learned:

parameters: $\theta = \{\mu, \sigma, \pi, A\}$ states: $p(z \mid x, \theta)$

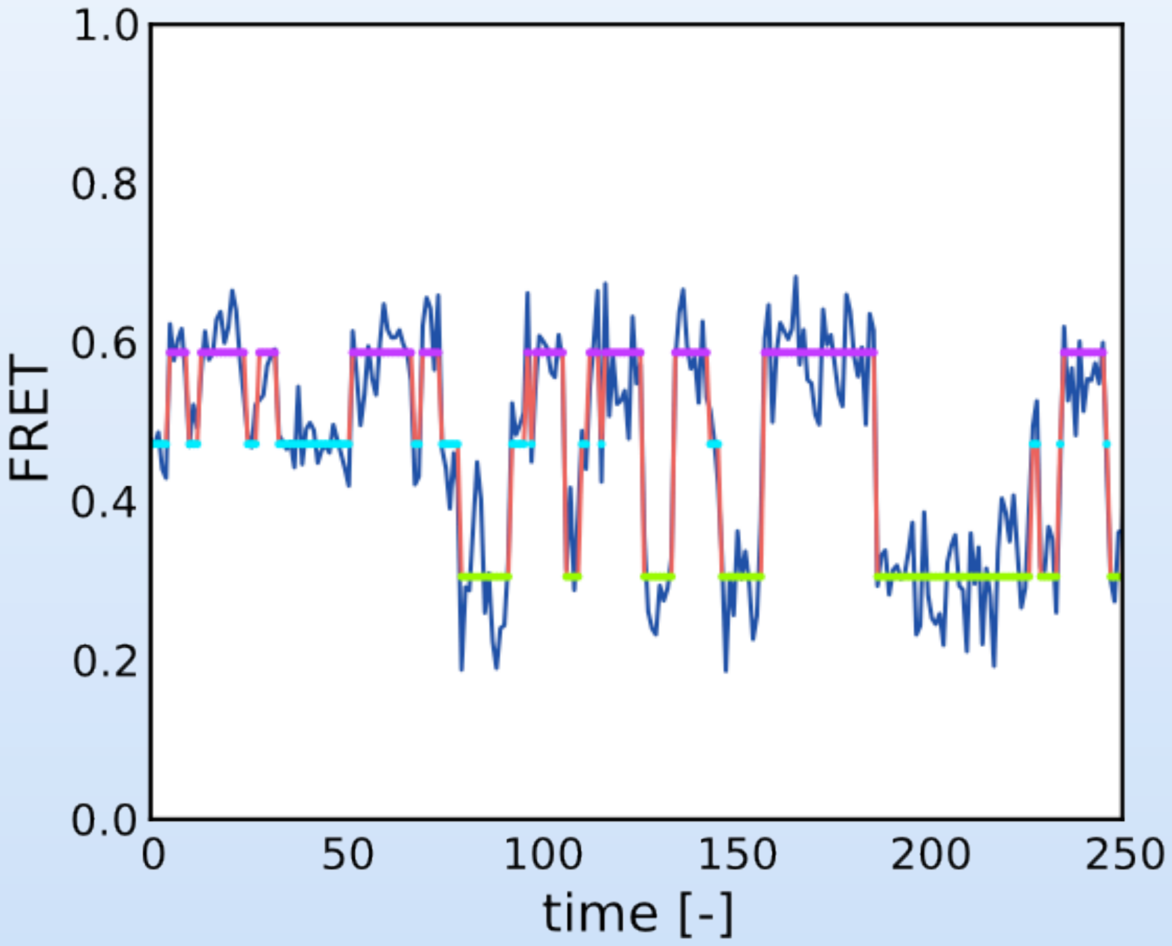
How Many States?

Model Complexity

2 States



3 States



Maximum Evidence

Log-Likelihood

$$L = \log p(x \mid \theta) = \log \left[\sum_z p(x, z \mid \theta) \right]$$

Log-Evidence

$$L = \log p(x \mid u) = \log \left[\sum_z \int d\theta \; p(x, z \mid \theta) p(\theta | u) \right]$$

Maximum Evidence

Log-Likelihood

$$L = \log p(x \mid \theta) = \log \left[\sum_z p(x, z \mid \theta) \right]$$

Log-Evidence

$$L = \log p(x \mid u) = \log \left[\sum_z \int d\theta \; p(x, z \mid \theta) p(\theta|u) \right]$$

Prior

Maximum Evidence

Log-Likelihood

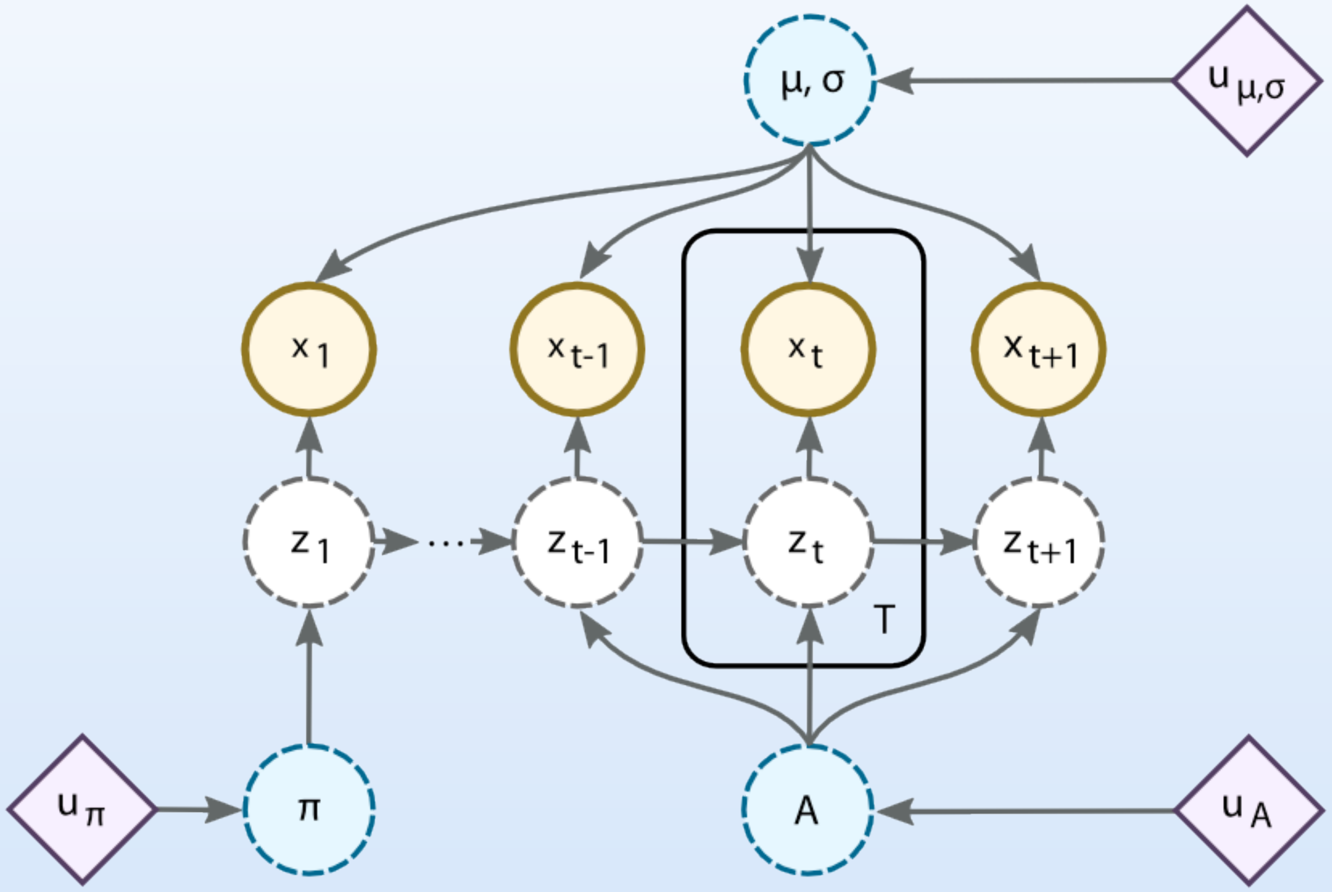
$$L = \log p(x \mid \theta) = \log\left[\sum_z p(x, z \mid \theta)\right]$$

Log-Evidence

$$L = \log p(x \mid u) = \log\left[\sum_z \int d\theta \; p(x, z \mid \theta) p(\theta|u)\right]$$

best model has highest *average* likelihood

Variational Bayes



VBEM Updates

$$\frac{\delta \mathcal{L}_n}{\delta q(z_n)} = 0$$

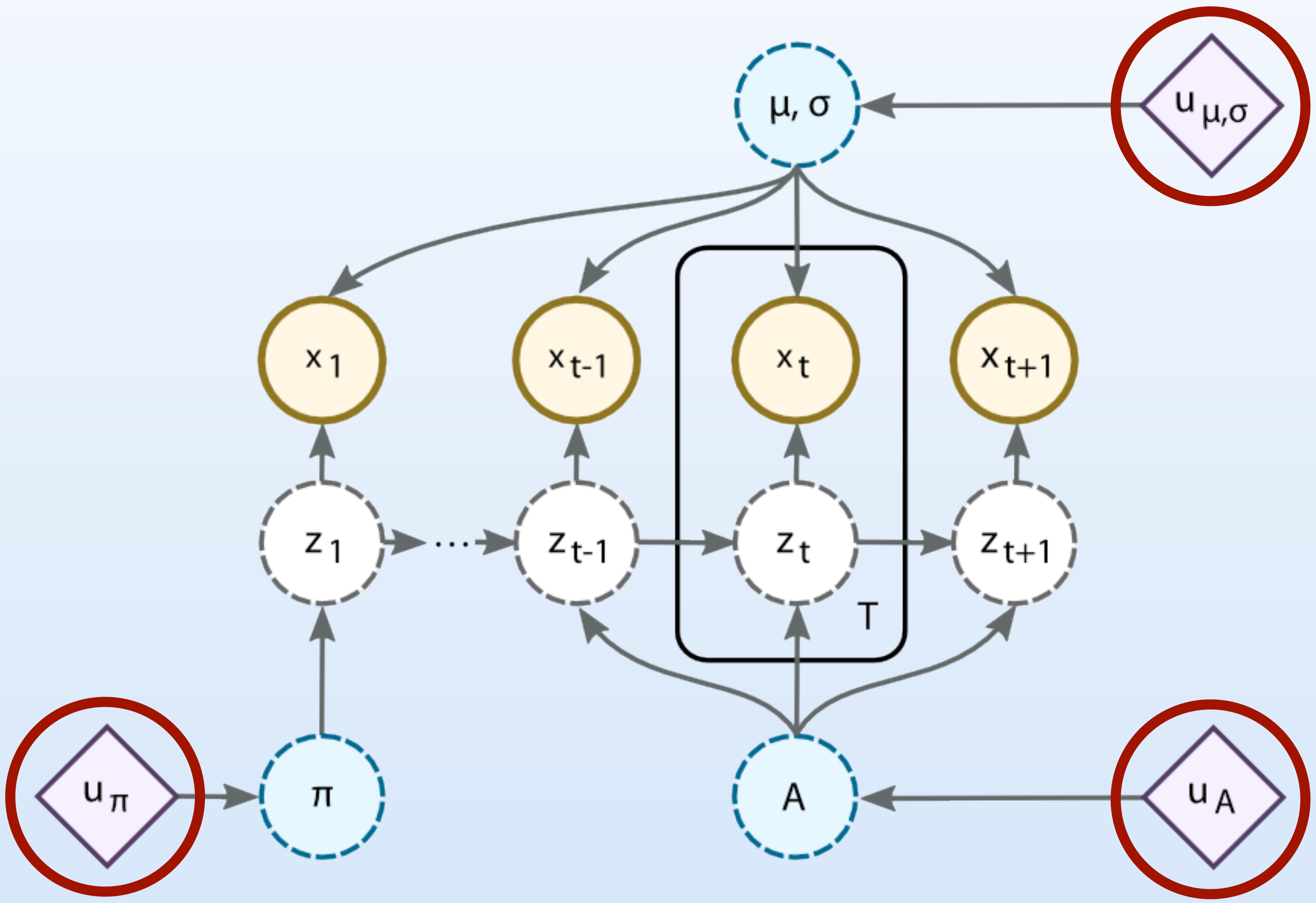
$$\frac{\delta \mathcal{L}_n}{\delta q(\theta_n)} = 0$$

We've learned:

parameters: q(θ | w)

states: p(z | x, θ)

Variational Bayes



VBEM Updates

$$\frac{\delta \mathcal{L}_n}{\delta q(z_n)} = 0$$

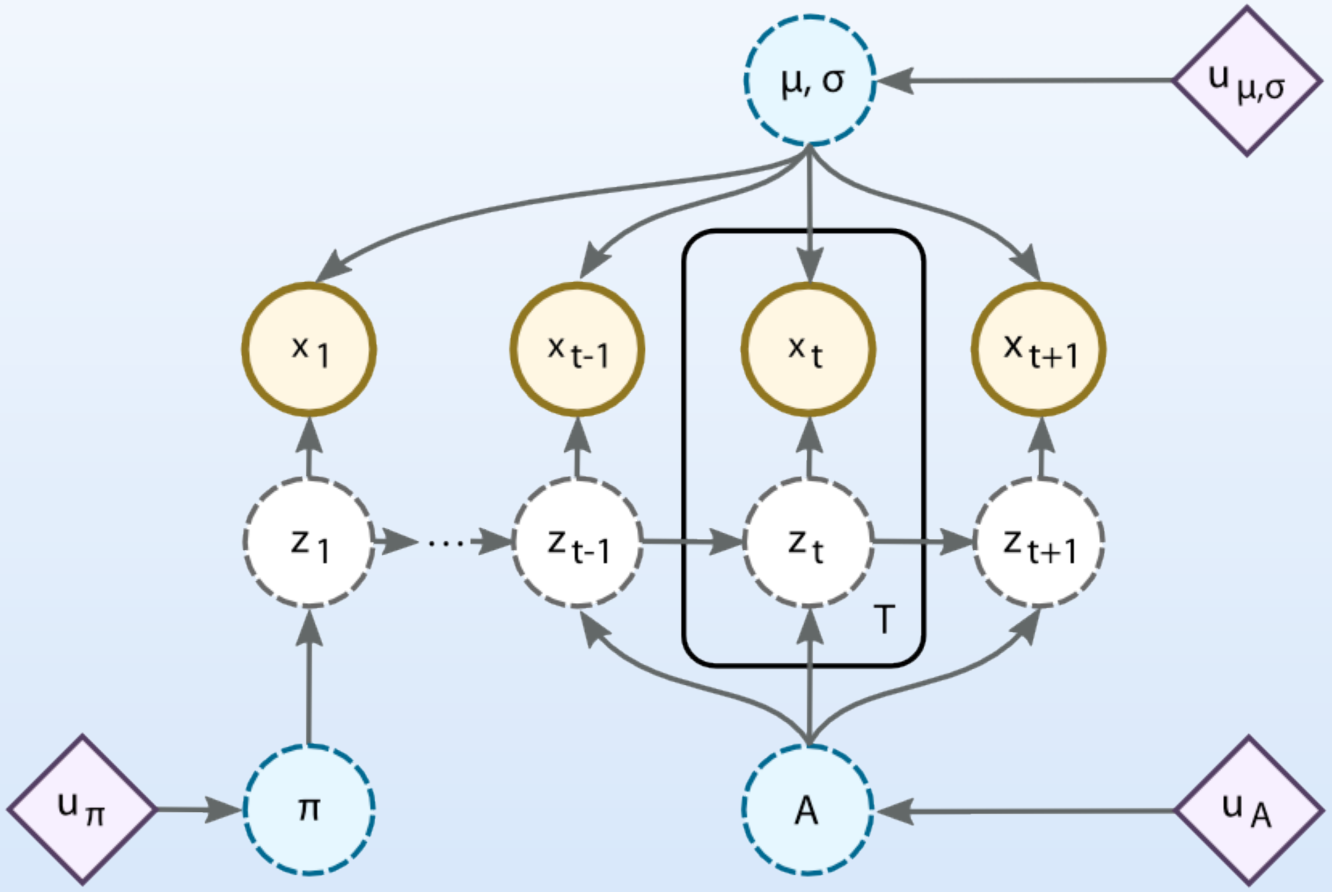
$$\frac{\delta \mathcal{L}_n}{\delta q(\theta_n)} = 0$$

We've learned:

parameters: q(θ | w)

states: p(z | x, θ)

Variational Bayes



VBEM Updates

$$\frac{\delta \mathcal{L}_n}{\delta q(z_n)} = 0$$

$$\frac{\delta \mathcal{L}_n}{\delta q(\theta_n)} = 0$$

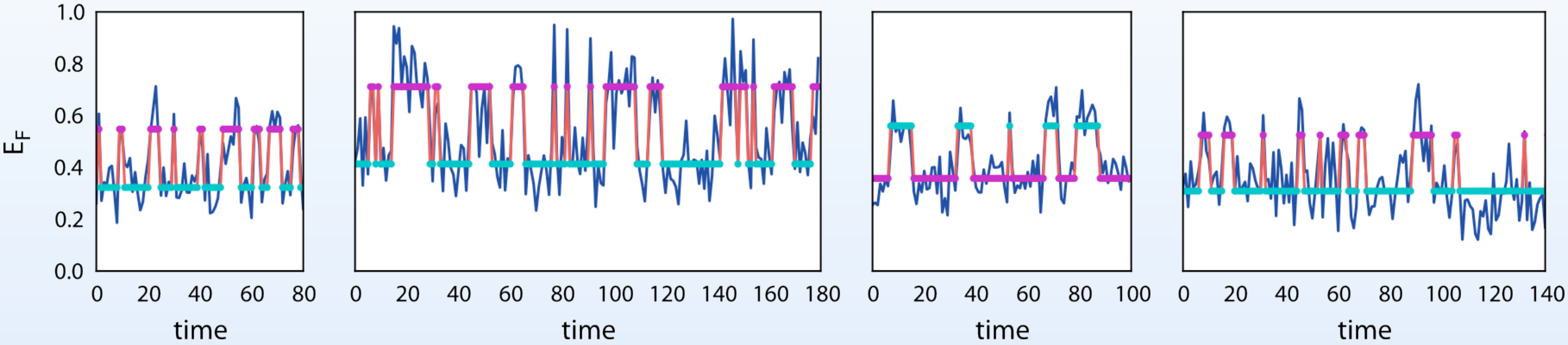
We've learned:

parameters: q(θ | w)

states: p(z | x, θ)

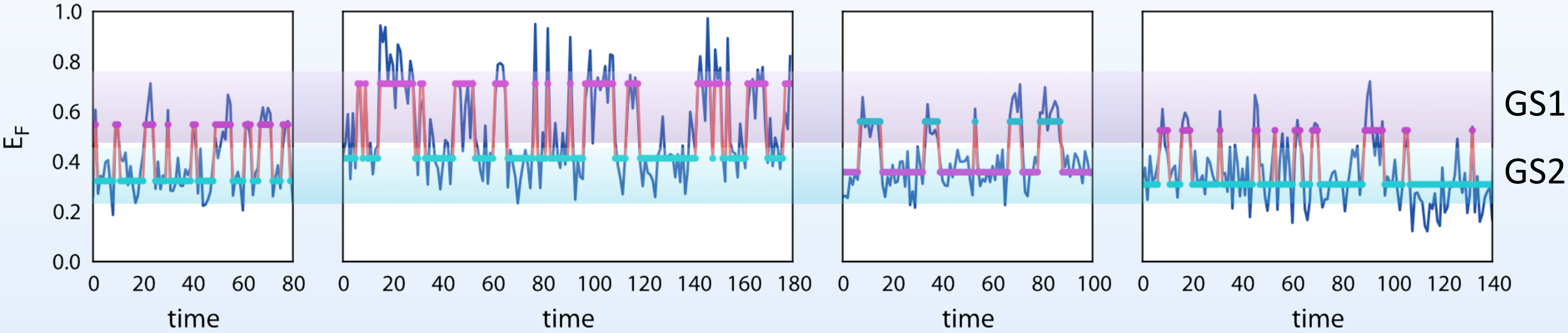
Consensus Analysis

Learning Kinetics from Traces



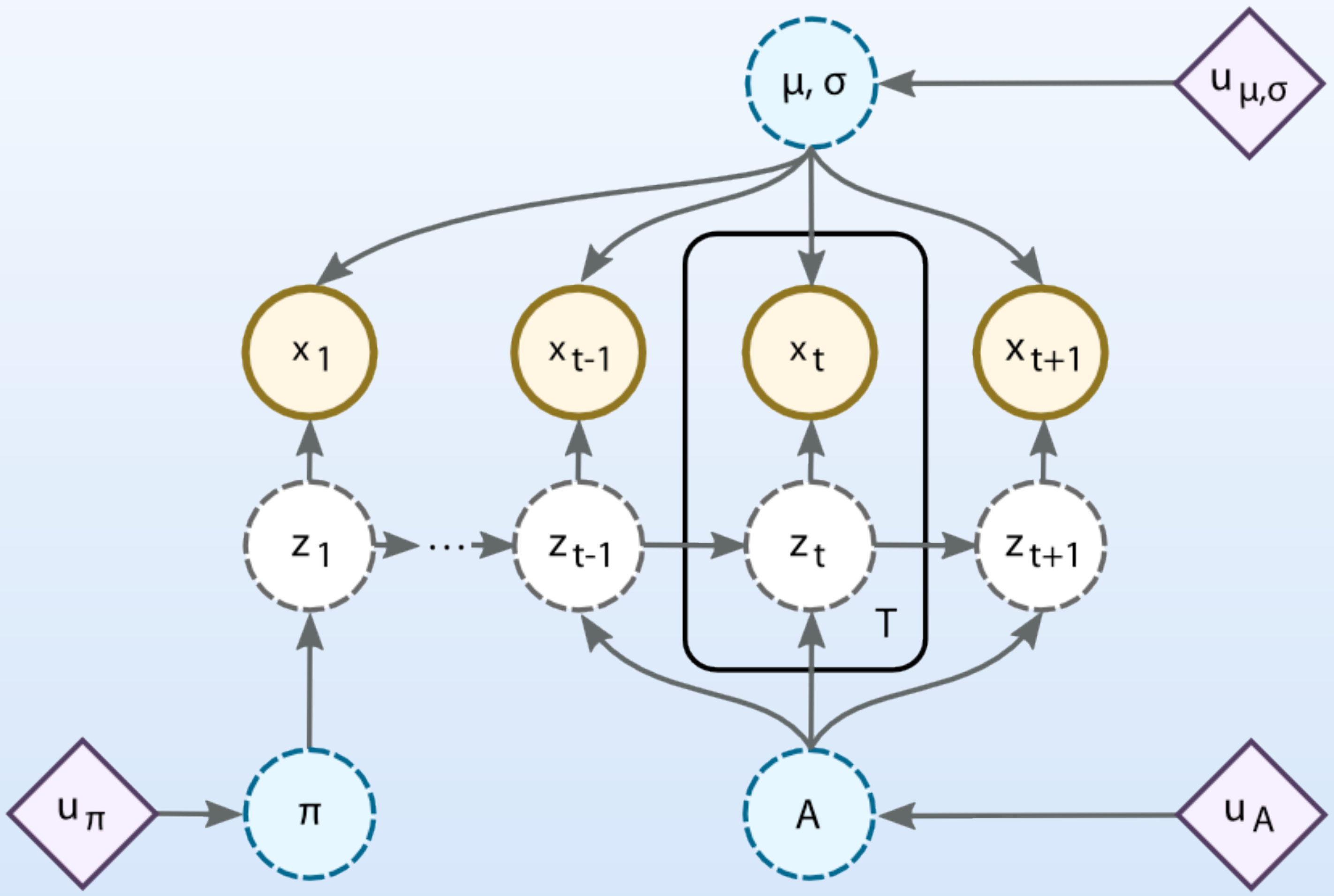
1. Identify states
2. Calculate Kinetic Rates

Learning Kinetics from Traces

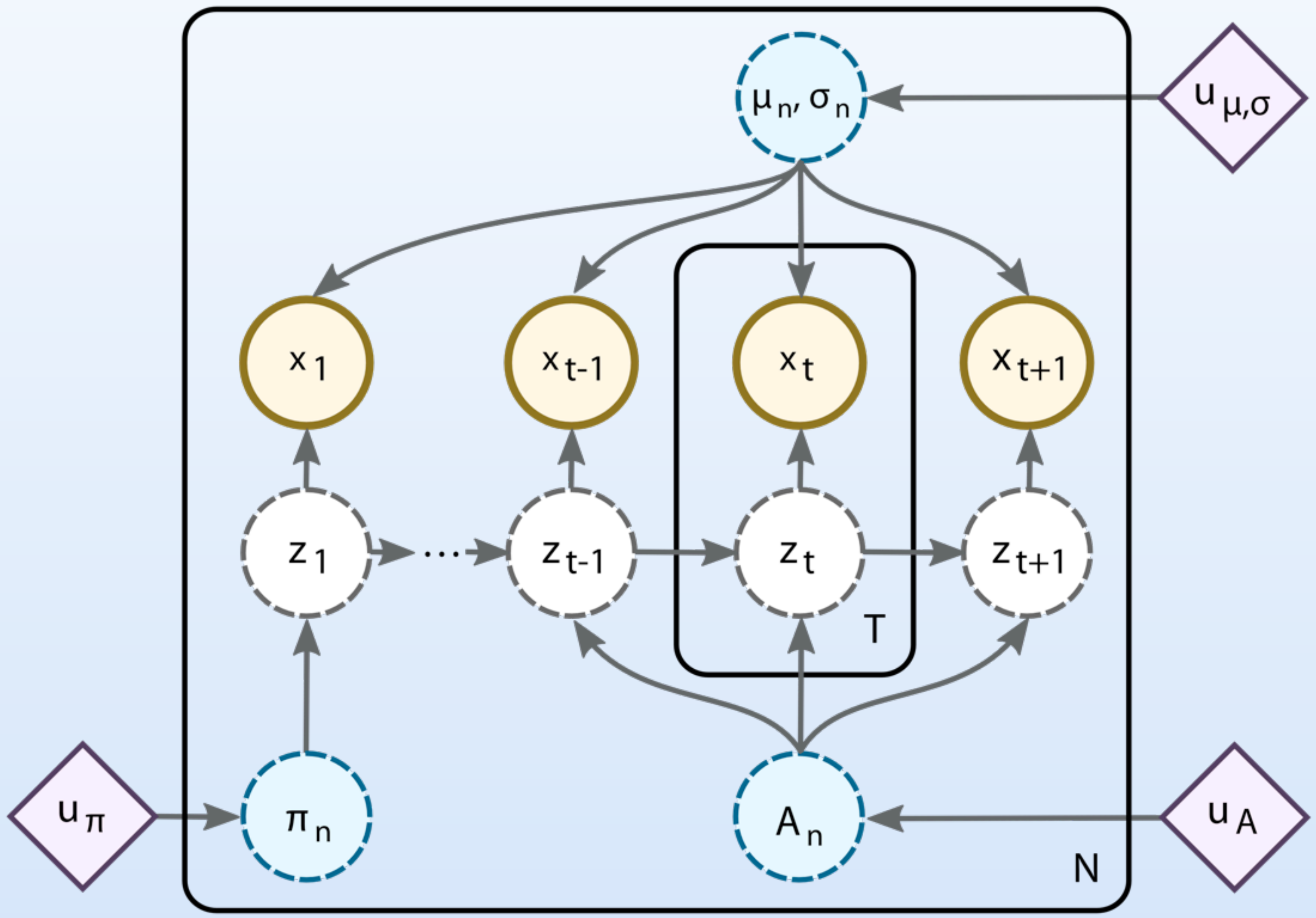


1. Identify states
2. Calculate Kinetic Rates
3. Construct Consensus Model

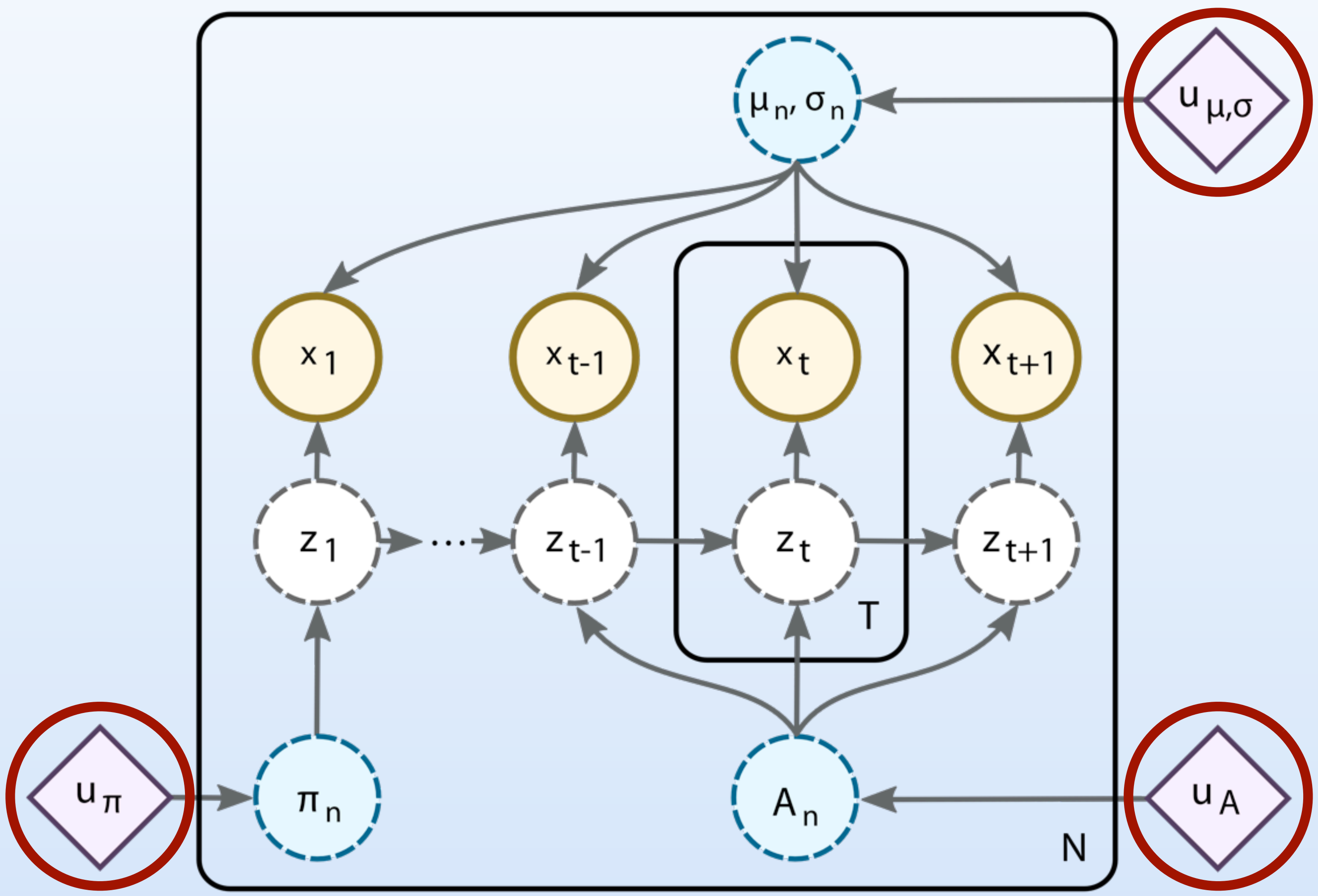
Learning Ensembles



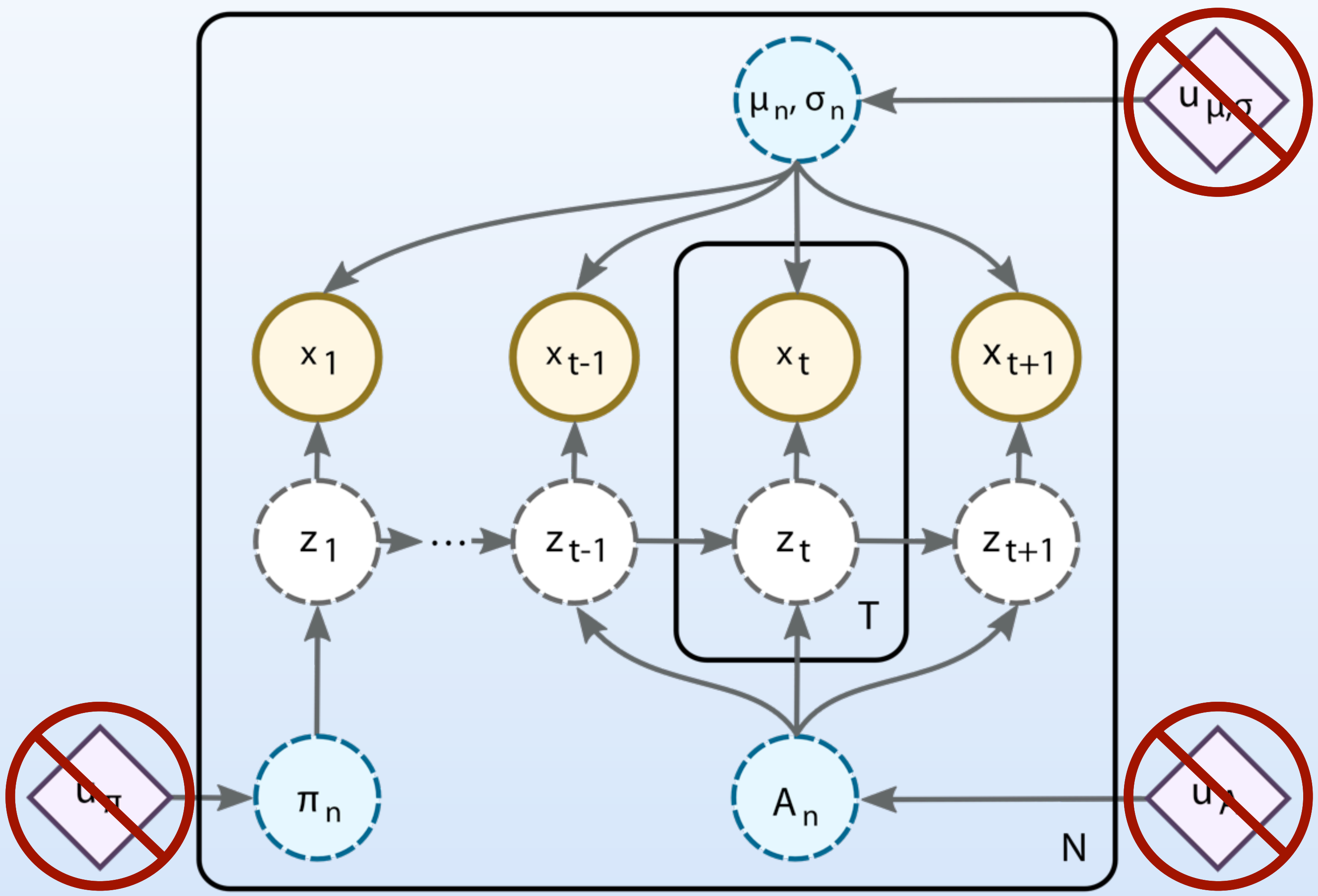
Learning Ensembles



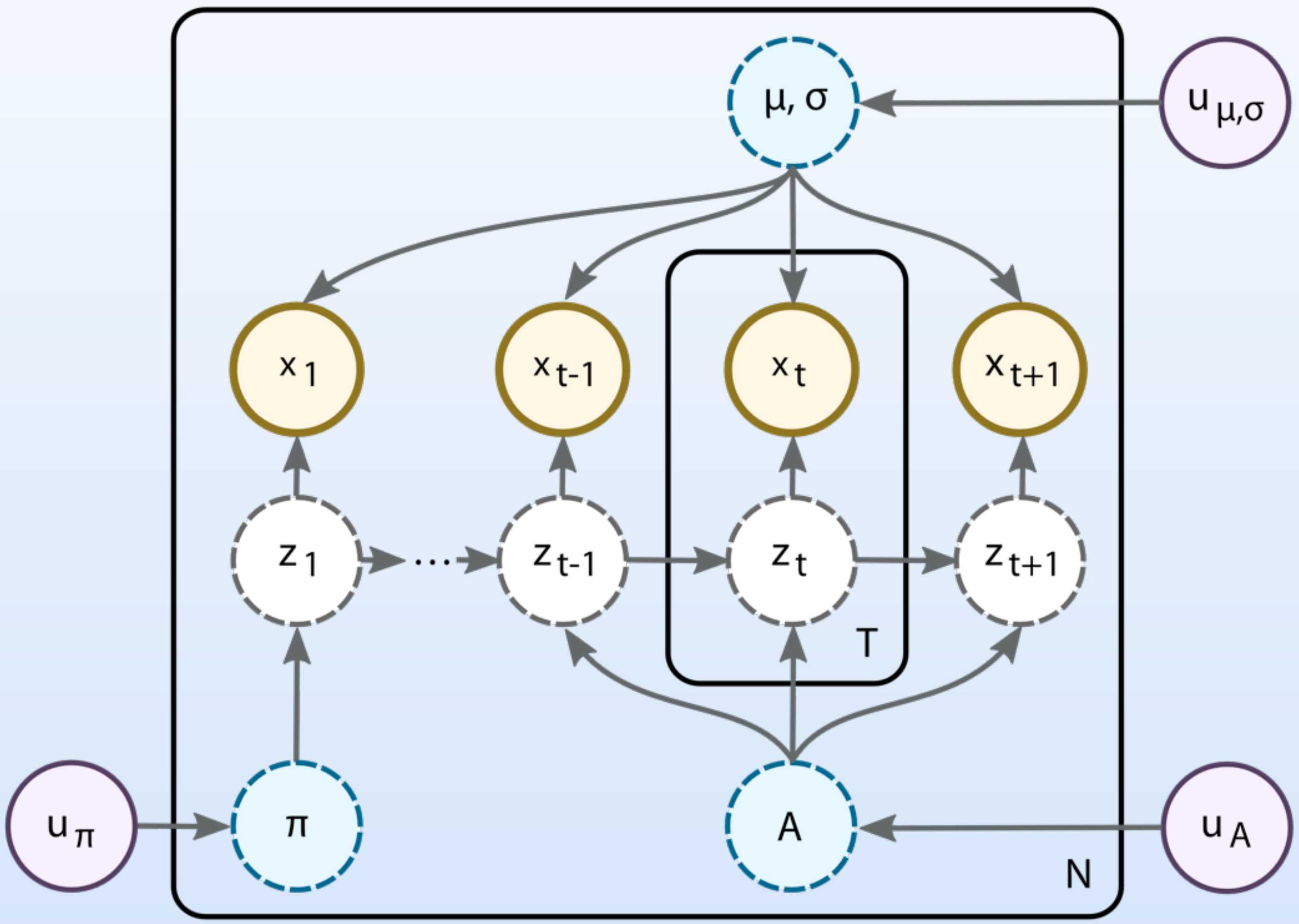
Learning Ensembles



Learning Ensembles



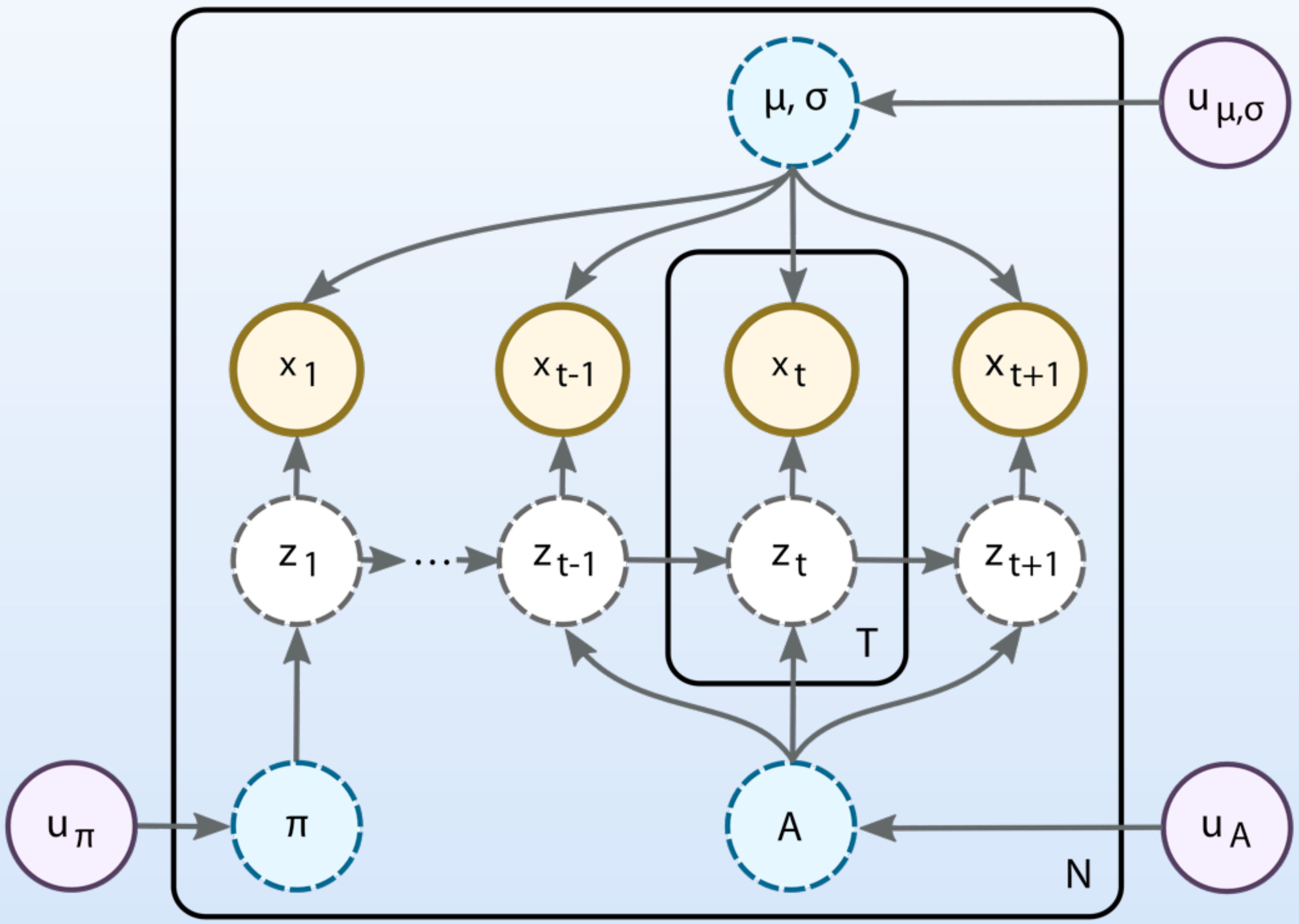
Learning Ensembles



Hierarchical Updates

$$\frac{\partial}{\partial u} \sum_{n} \mathcal{L}_n = 0$$

Empirical Bayes on HMM's



1. Run VBEM on each trace
    - Update $q(z_n)$
    - Update $q(\theta_n \mid w_n)$

    Until L_n converges

2. Update $p(\theta \mid u)$

Until $\Sigma\ L_n$ converges

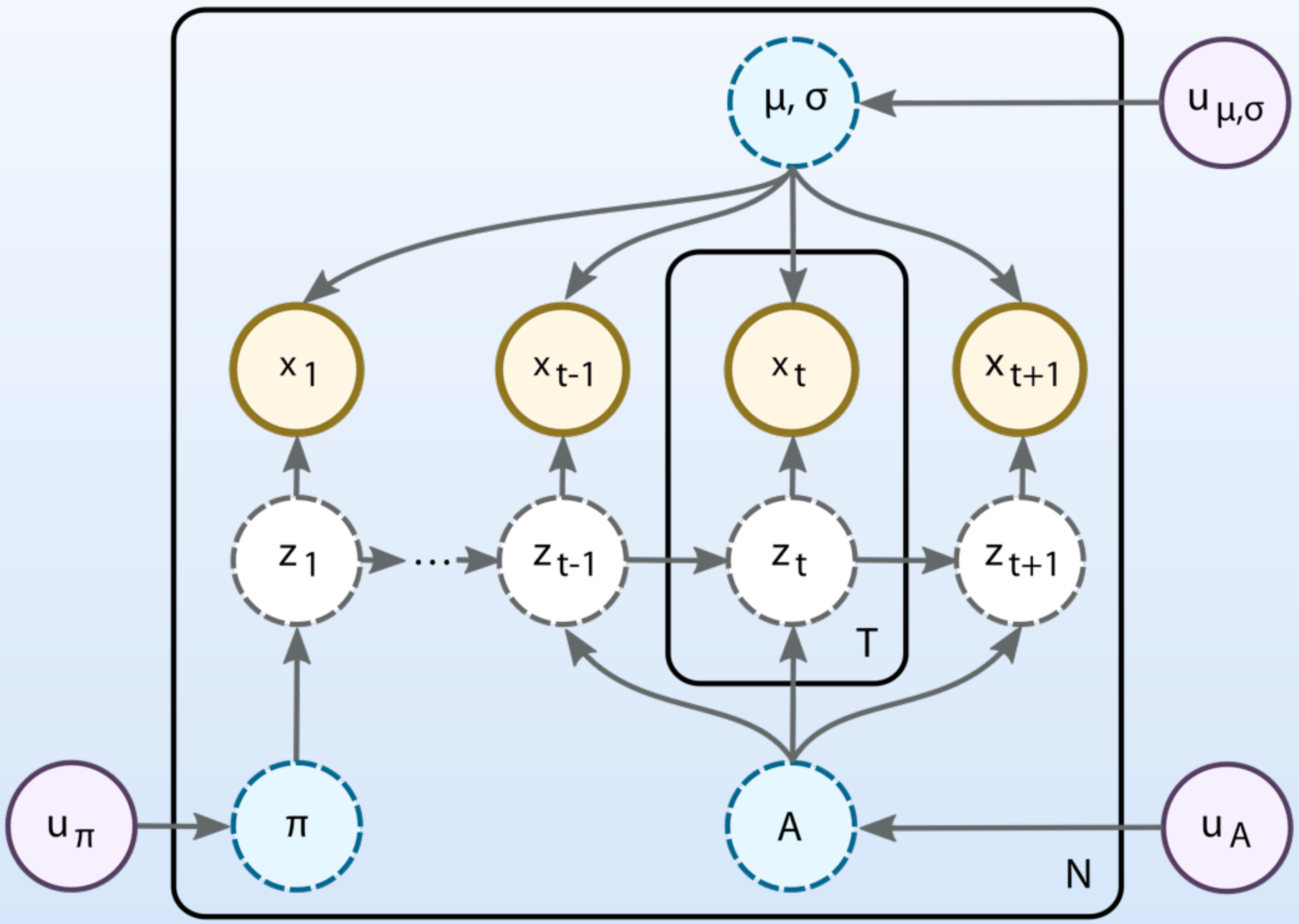
VBEM Updates

$$\frac{\delta \mathcal{L}_n}{\delta q(z_n)} = 0 \qquad \frac{\delta \mathcal{L}_n}{\delta q(\theta_n)} = 0$$

Hierarchical Updates

$$\frac{\partial}{\partial u} \sum_n \mathcal{L}_n = 0$$

Empirical Bayes on HMM's



1. Run VBEM on each trace
   - Update $q(z_n)$
   - Update $q(\theta_n \mid w_n)$

   Until L_n converges

2. Update $p(\theta \mid u)$

Until $\Sigma\, L_n$ converges

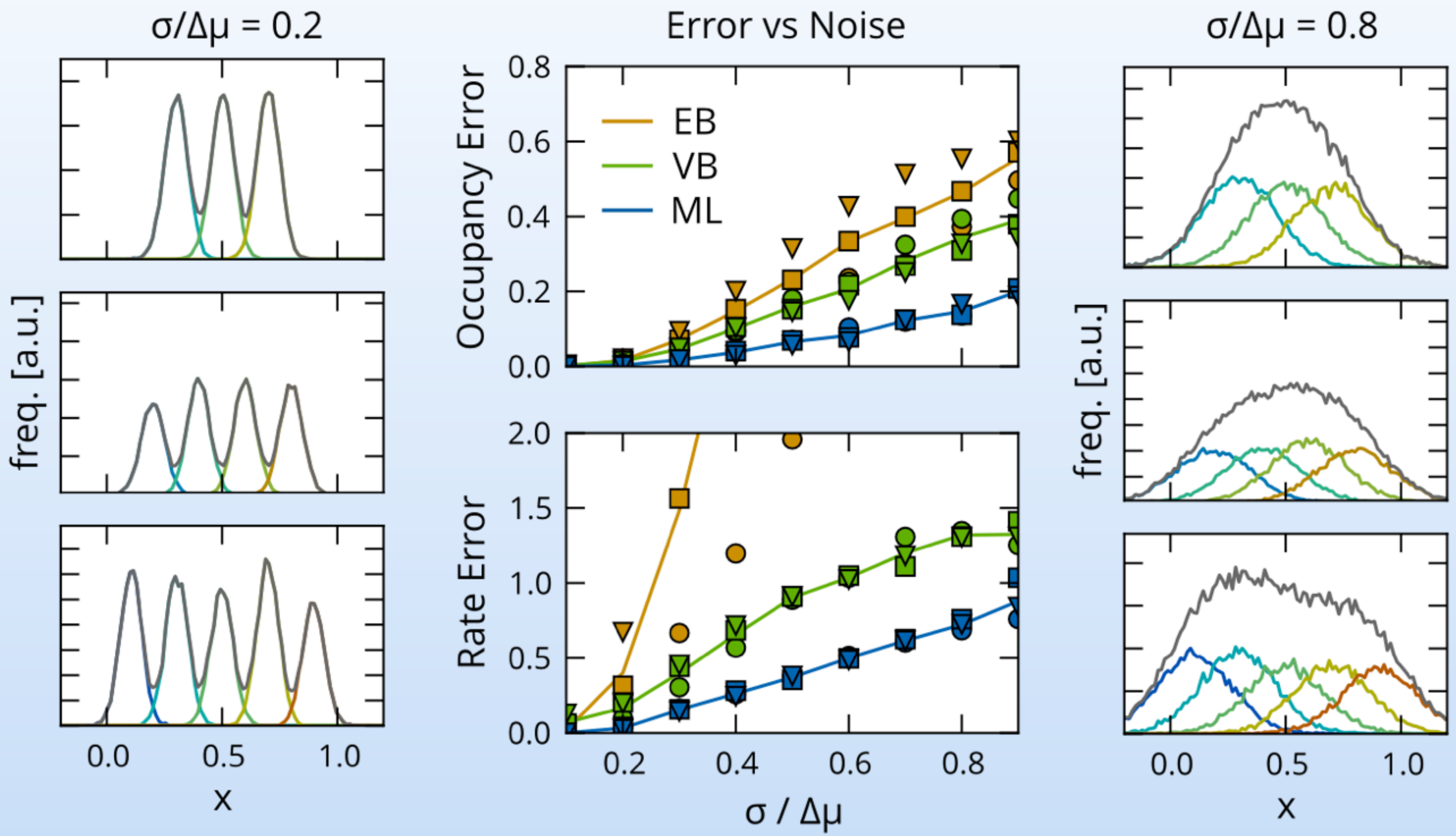
We've learned:

$p(\theta_n, z_n \mid x_n) \simeq q(\theta_n)\, q(z_n)$
(for each trace)

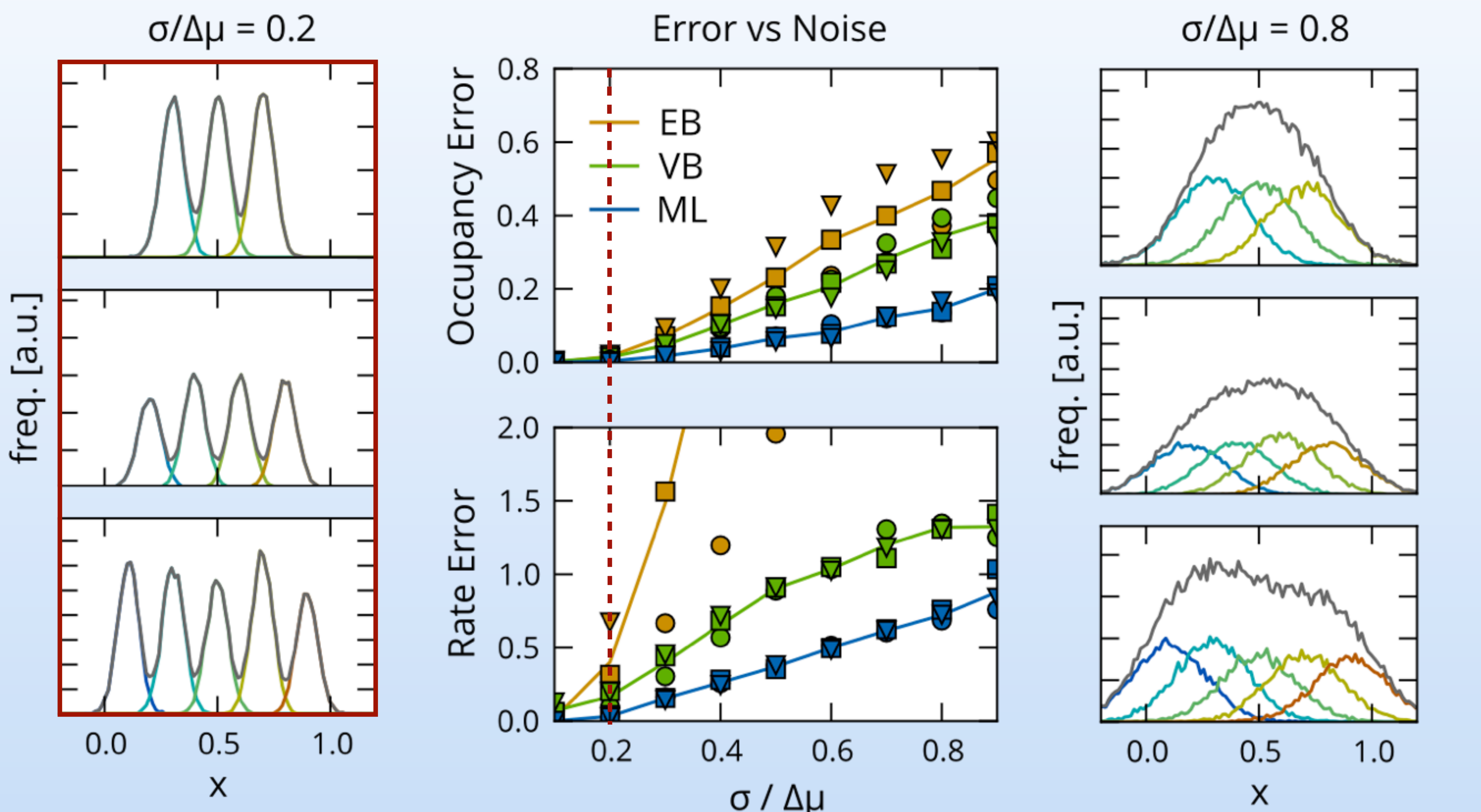
$p(\theta \mid u)$
(for ensemble)

Validation

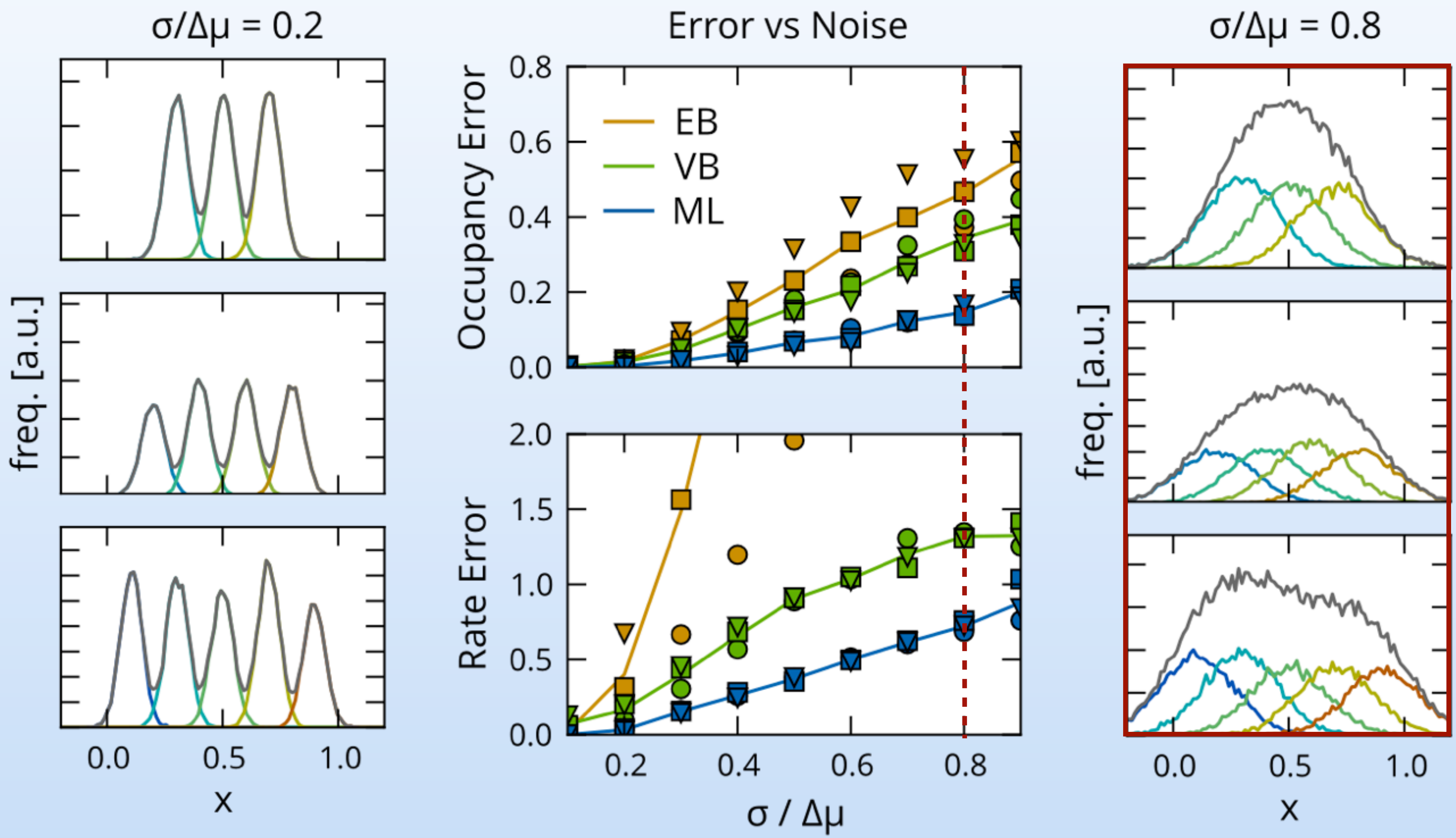
Accuracy vs Noise



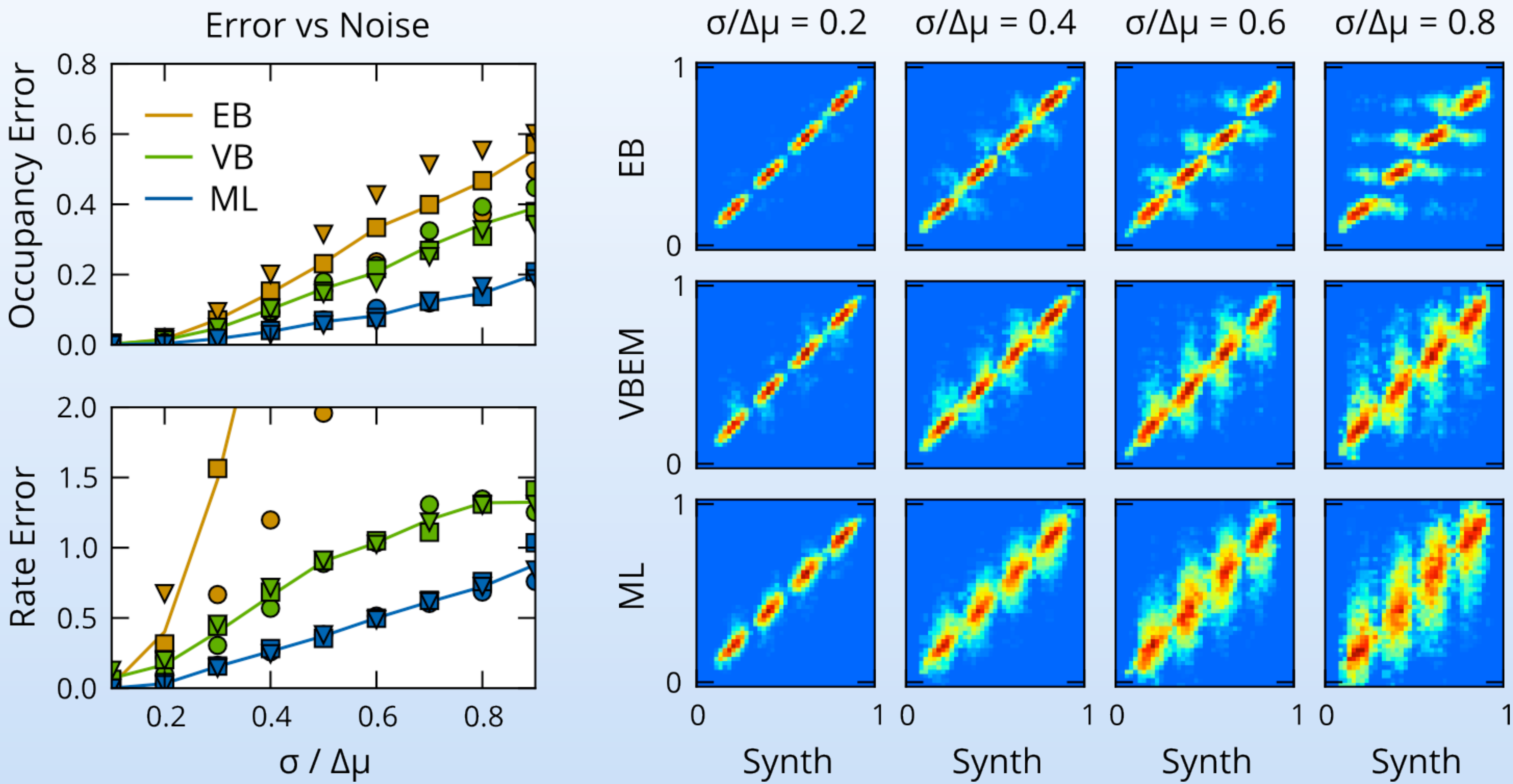
Accuracy vs Noise



Accuracy vs Noise

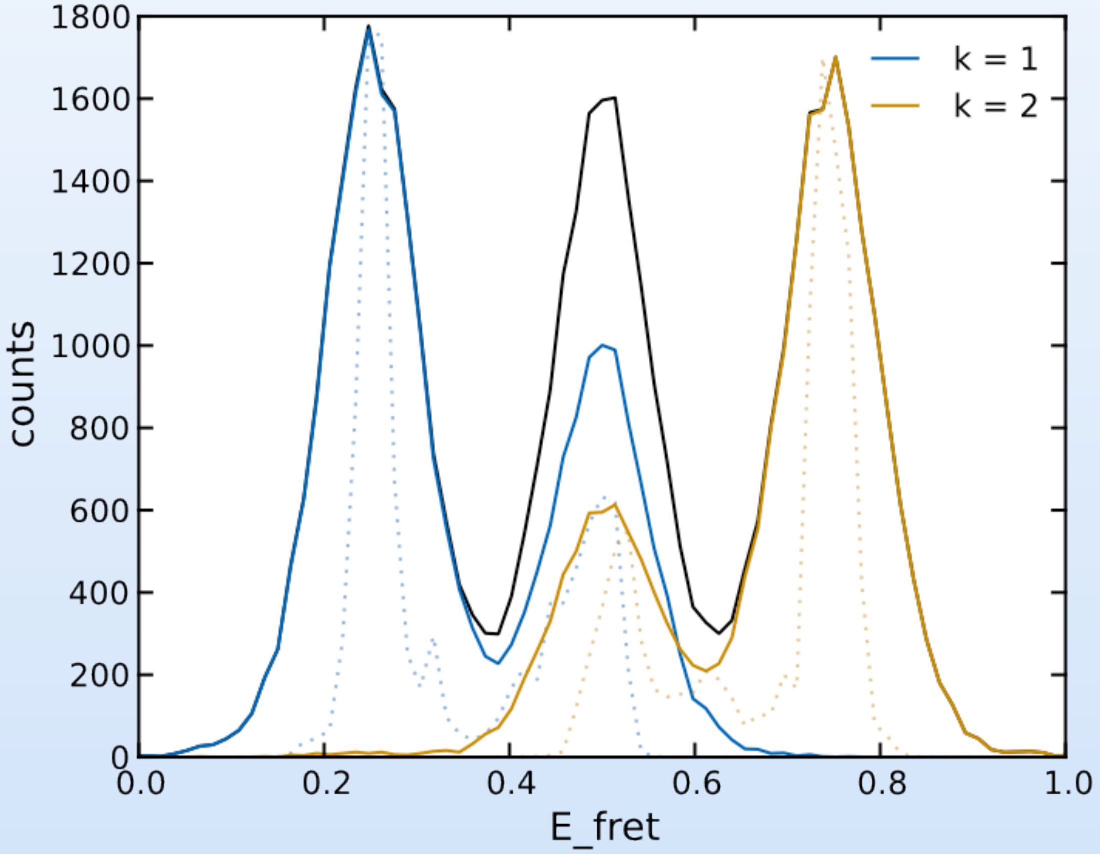


Accuracy vs Noise

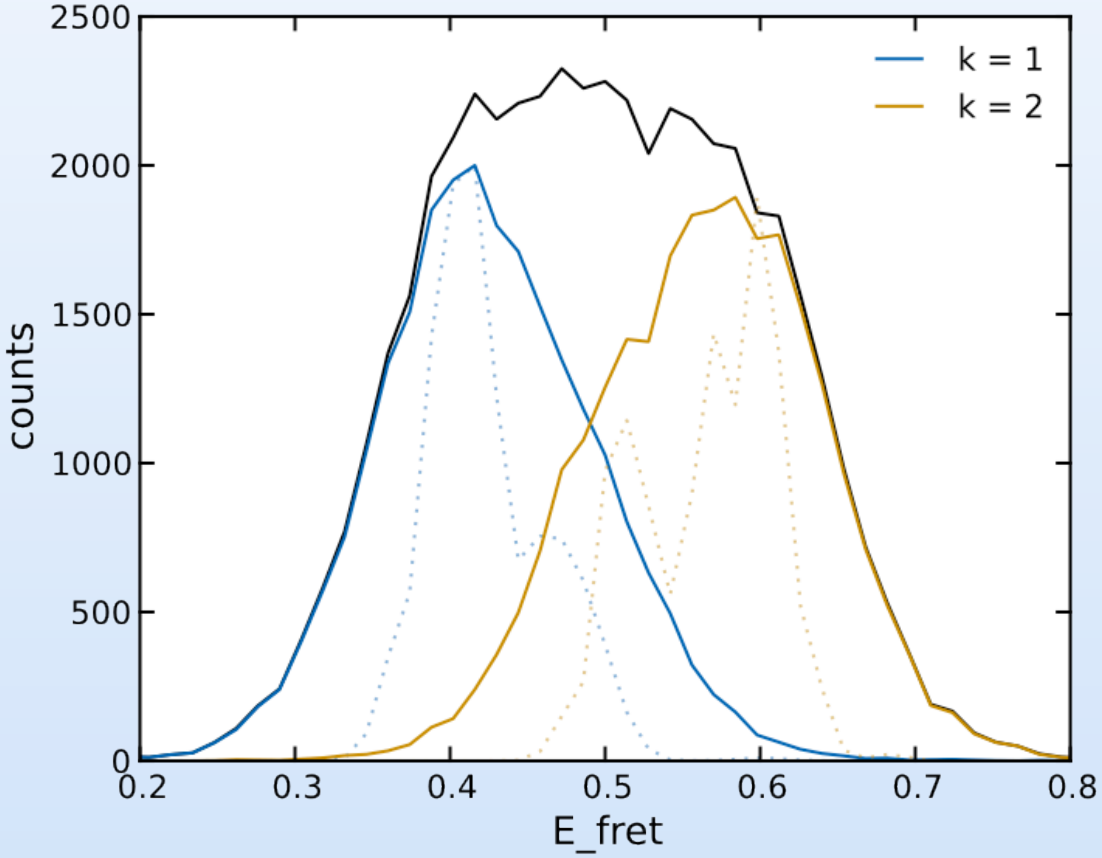


Model Selection

Low Noise, 2 States

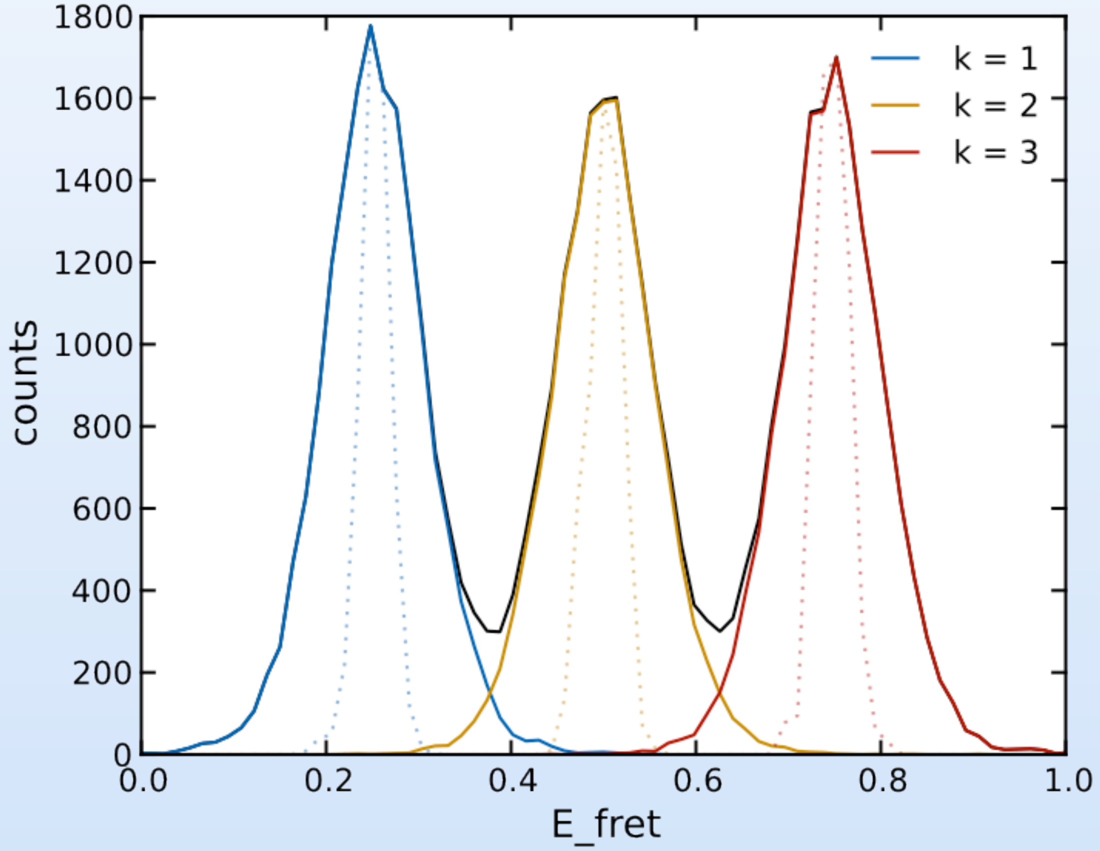


High Noise, 2 States

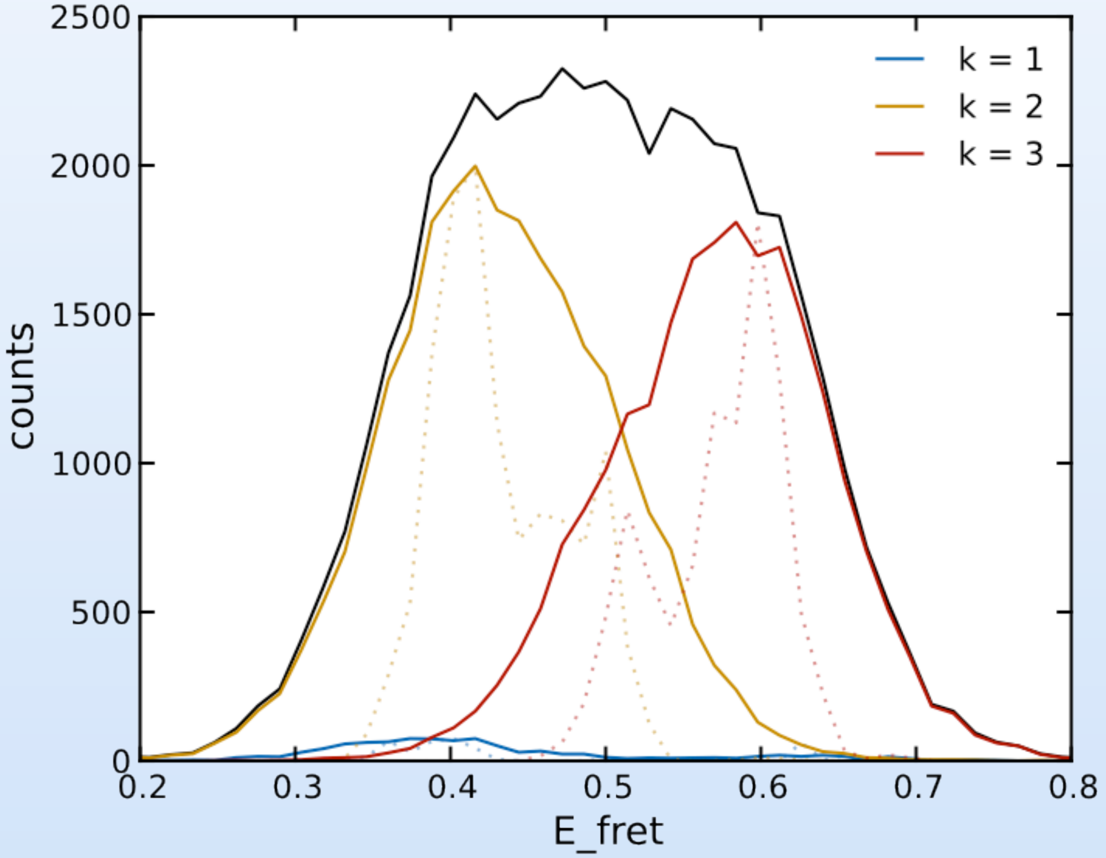


Model Selection

Low Noise, 3 States

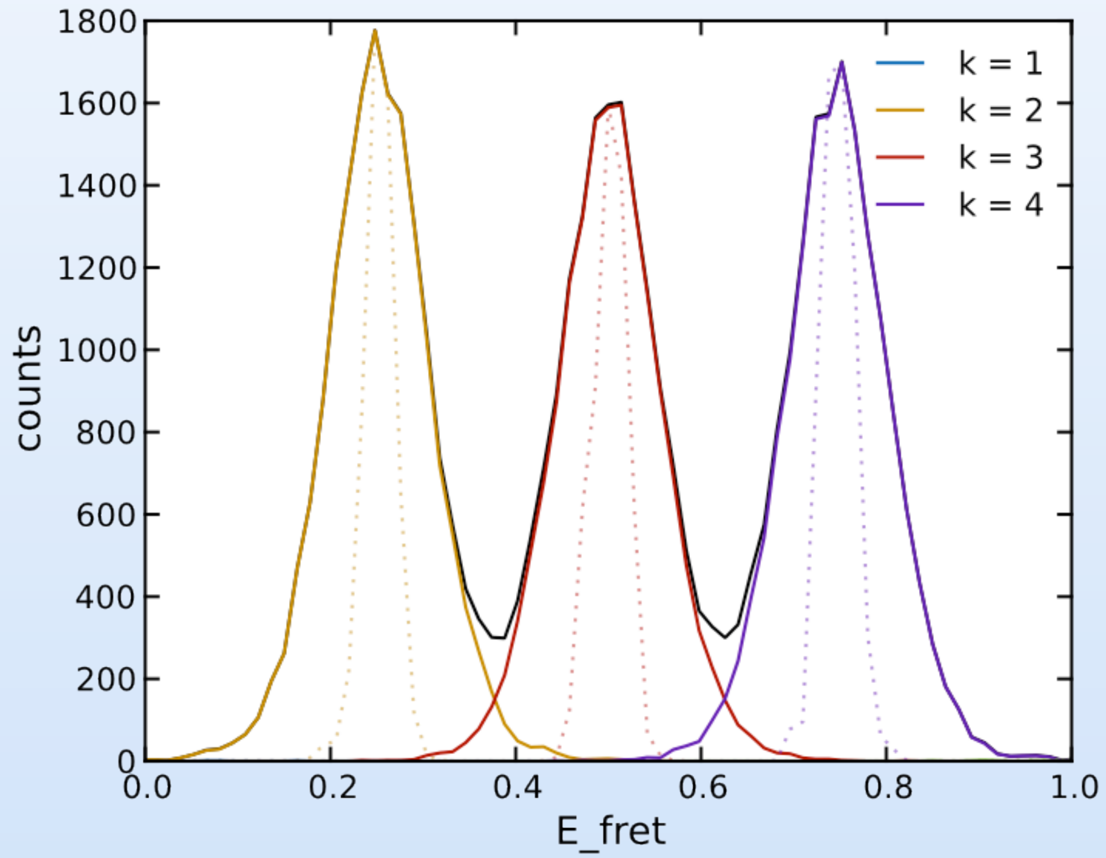


High Noise, 3 States

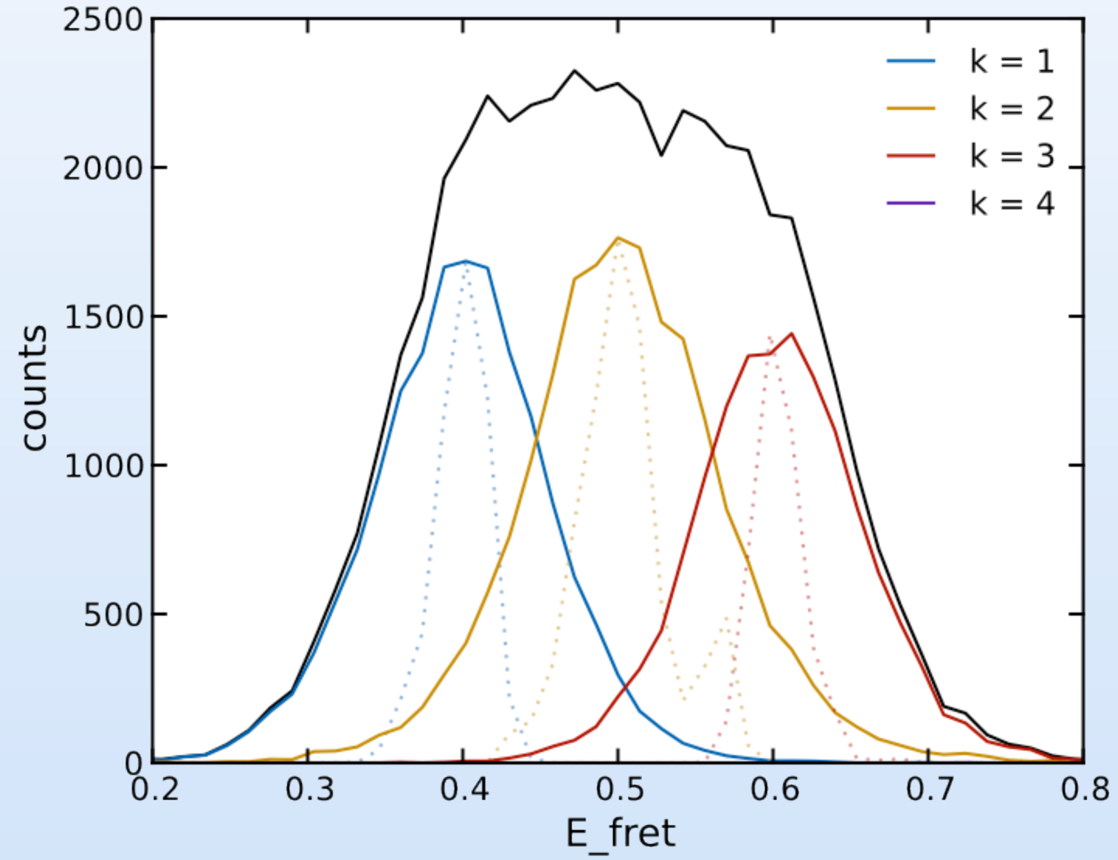


Model Selection

Low Noise, 4 States

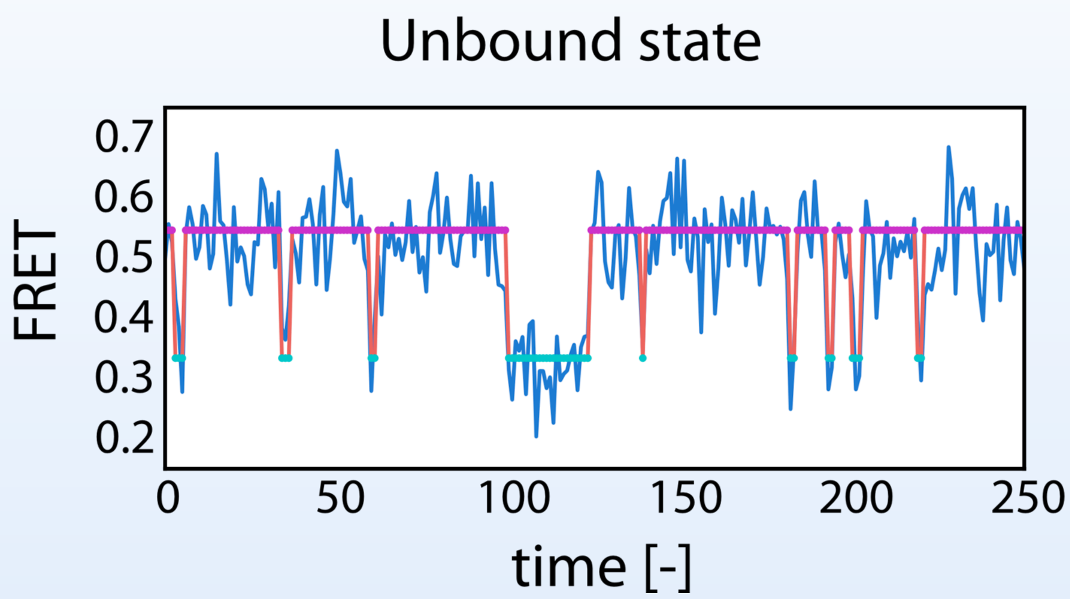


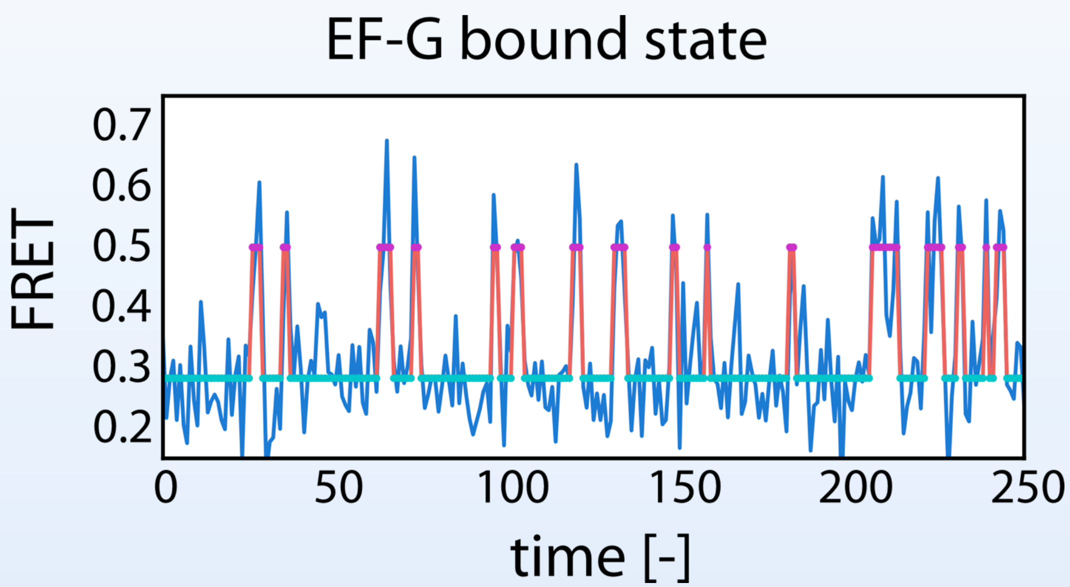
High Noise, 4 States



Sub-Populations

Learning Kinetics from Traces





1. Identify states
2. Calculate Kinetic Rates
3. Construct Consensus Model
4. Distinguish Subpopulations

Detecting Subpopulations

Use mixture model of priors

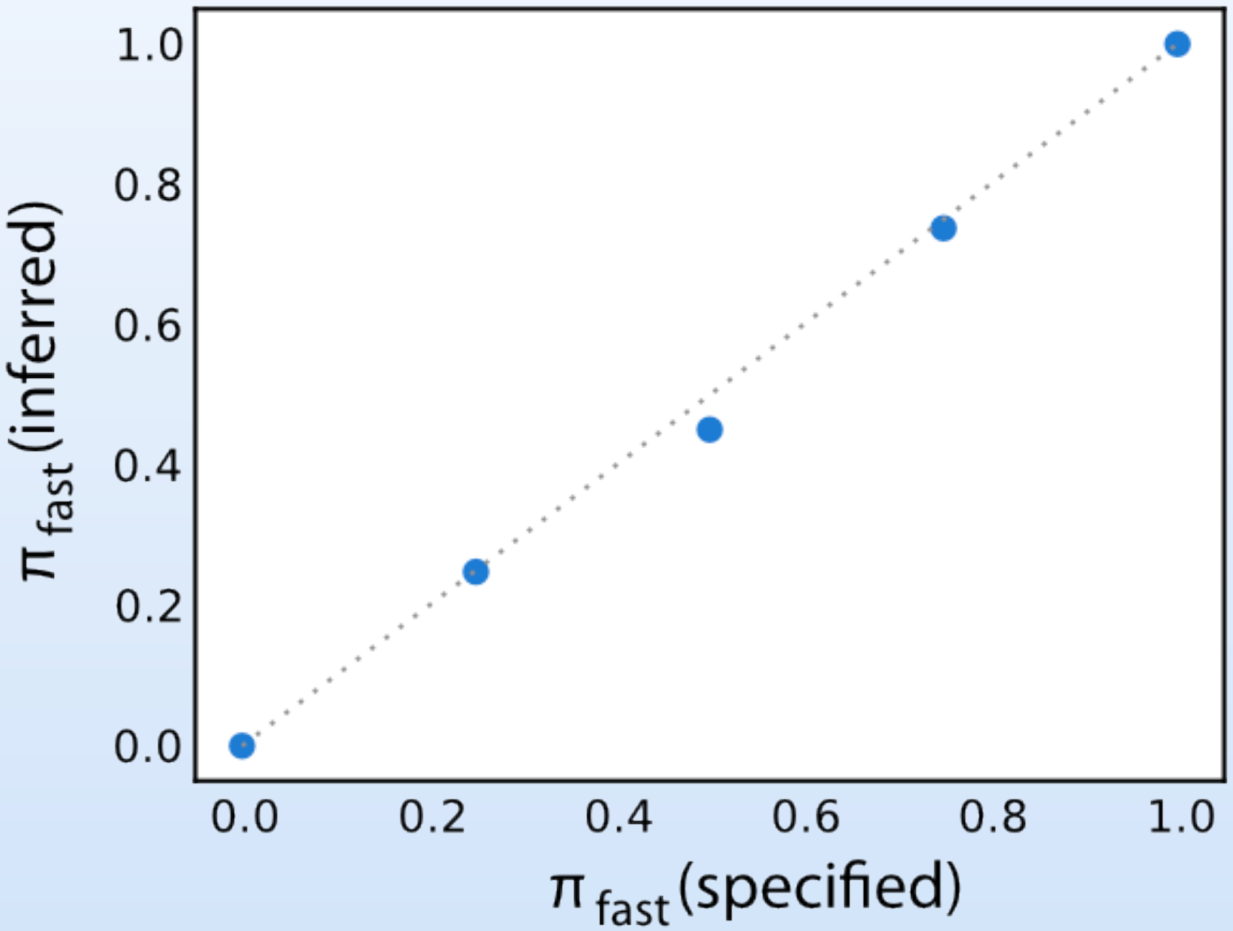
$$p(x \mid u) = \sum_m p(x \mid u_m)\, p(y = m \mid v)$$

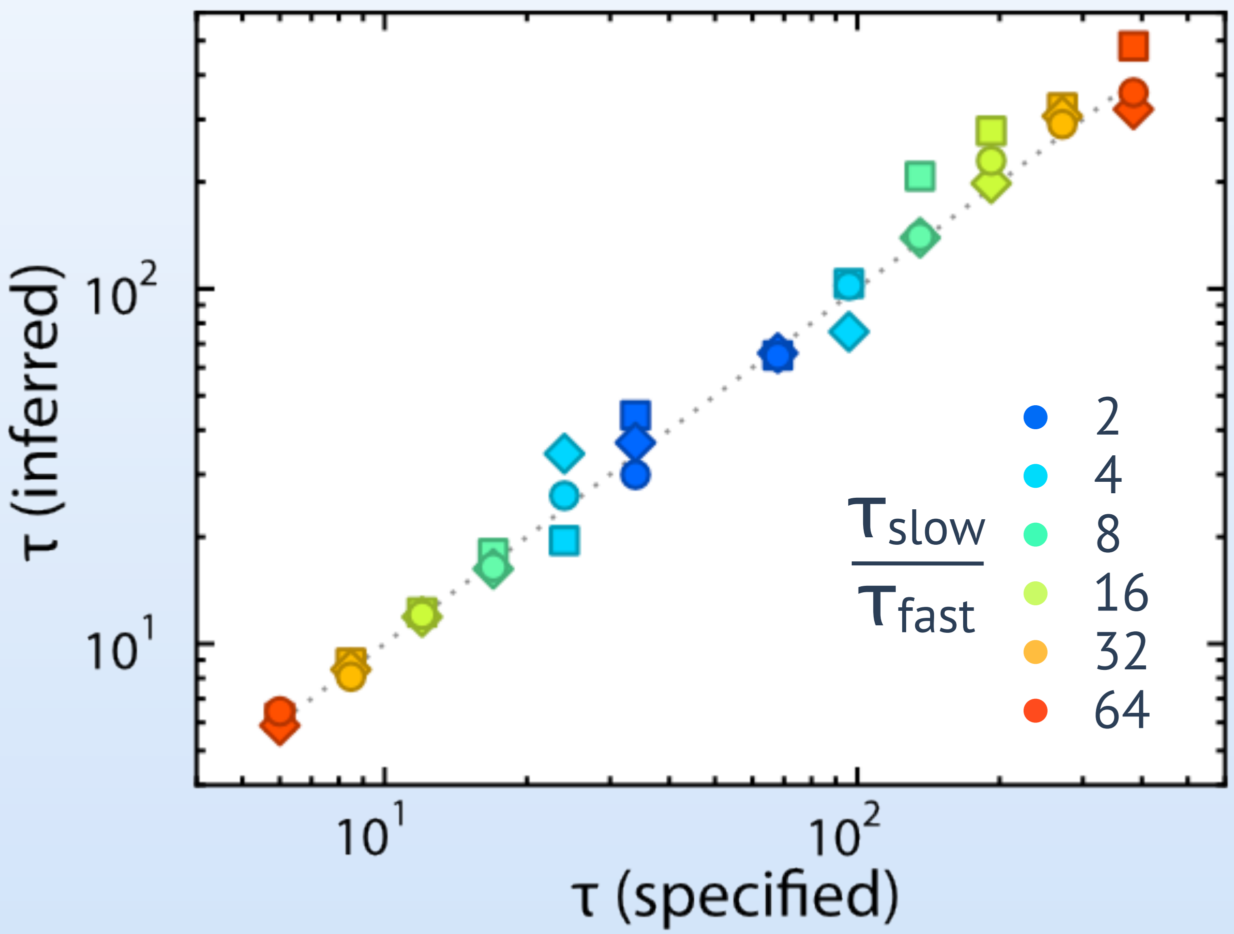
Detecting Subpopulations

Use mixture model of priors

$$p(x \mid u) = \sum_{m} p(x \mid u_m)\, p(y = m \mid v)$$

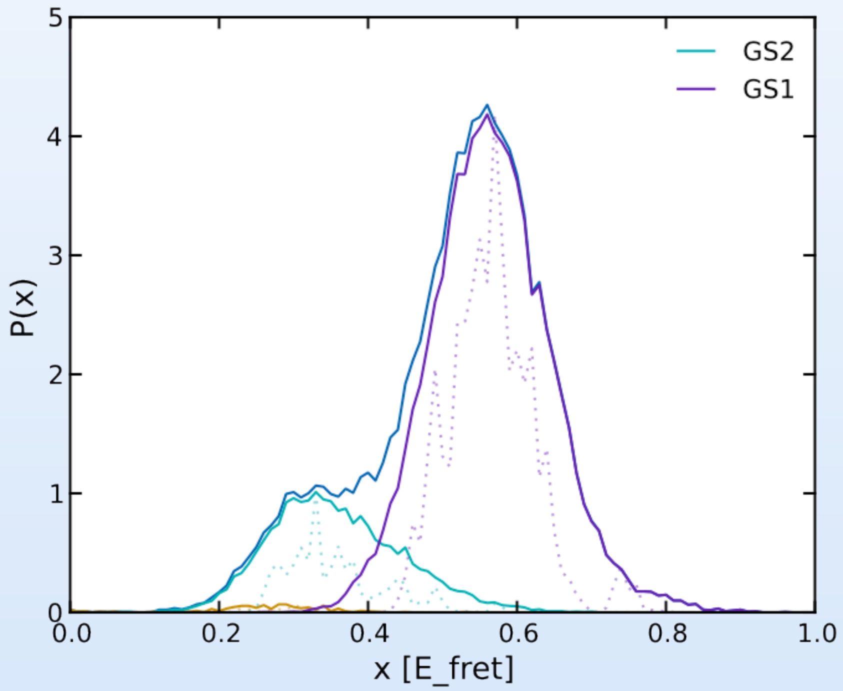
Validation on Synthetic Data



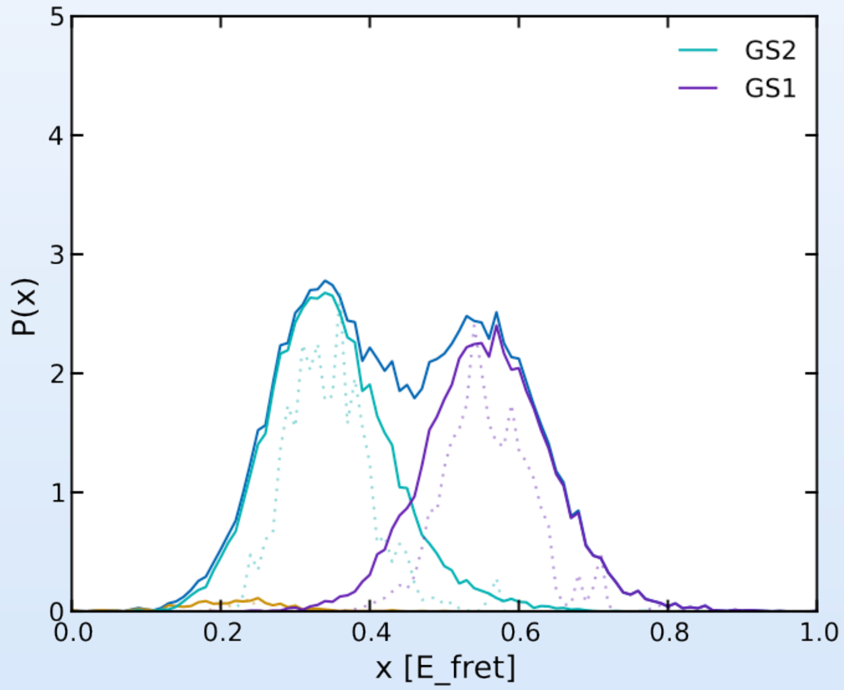


The role of EF-G binding

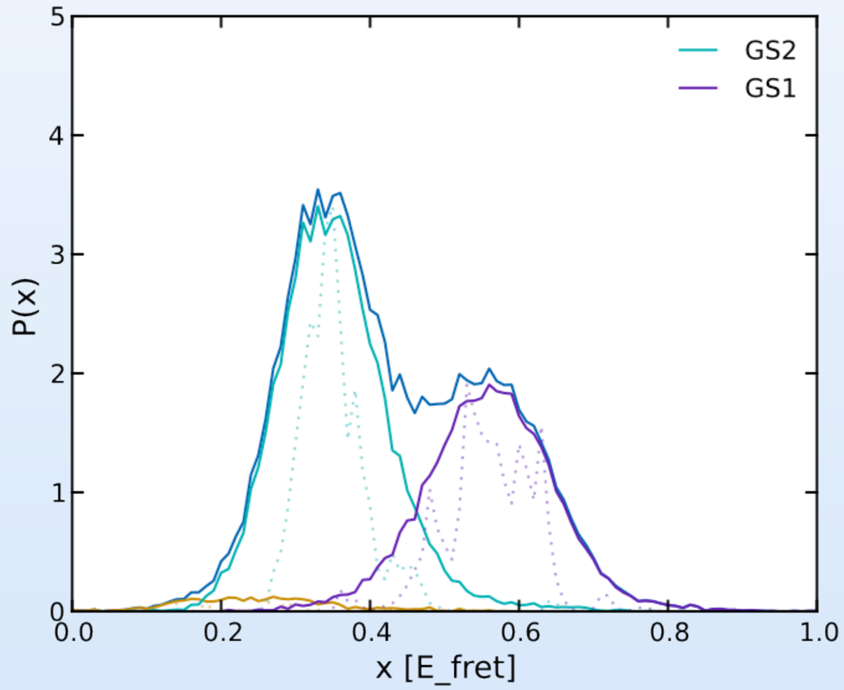
no EF-G



50 nM EF-G



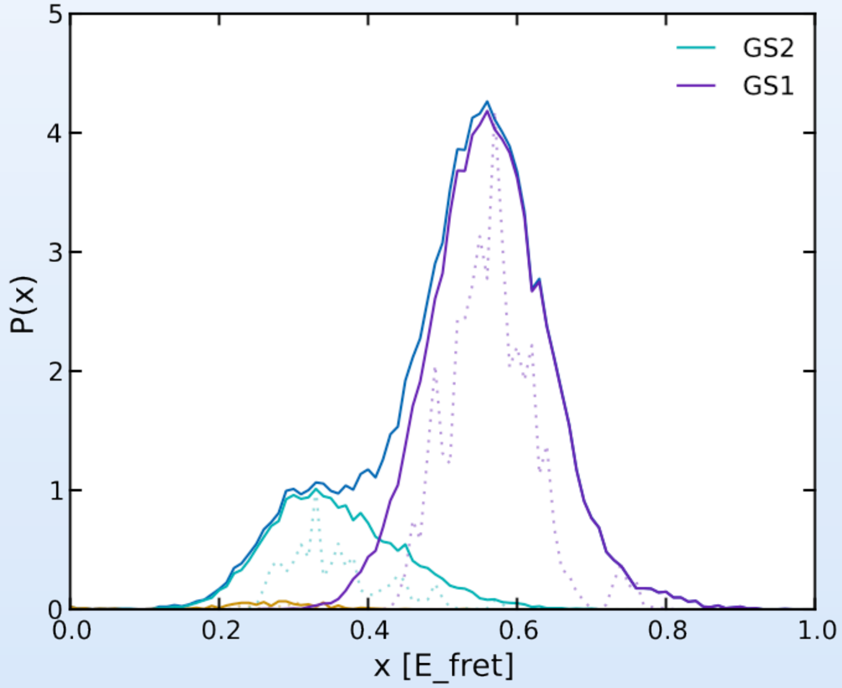
500 nM EF-G



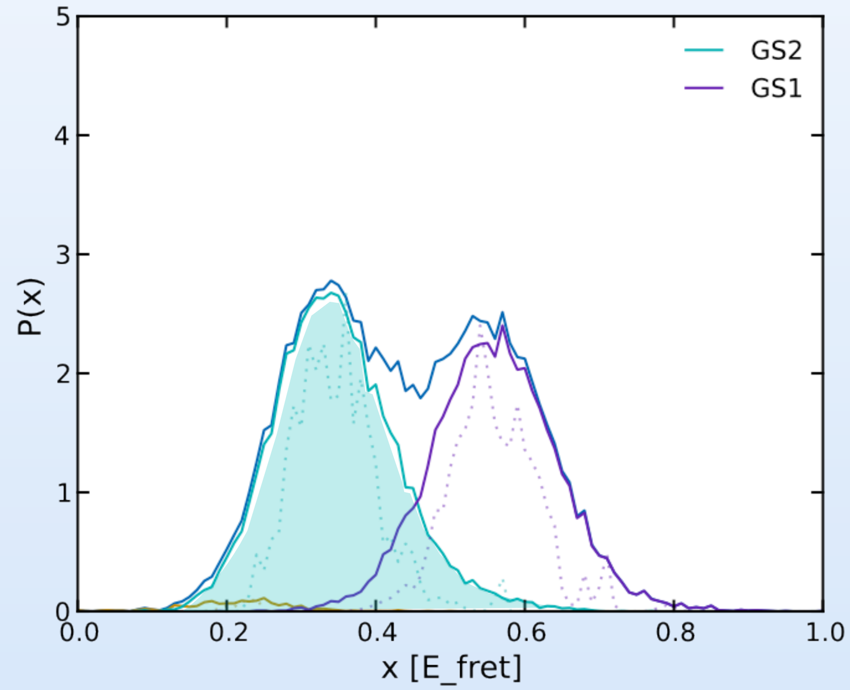
Fei, Bronson, Hofman, Srinivas, Wiggins, Gonzalez, PNAS, 2009

The role of EF-G binding

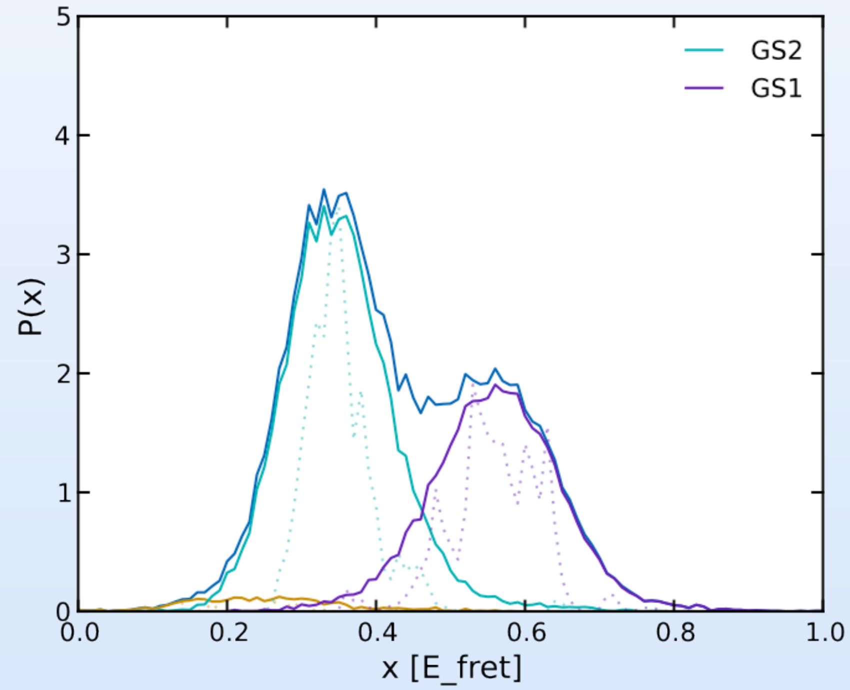
no EF-G



50 nM EF-G



500 nM EF-G

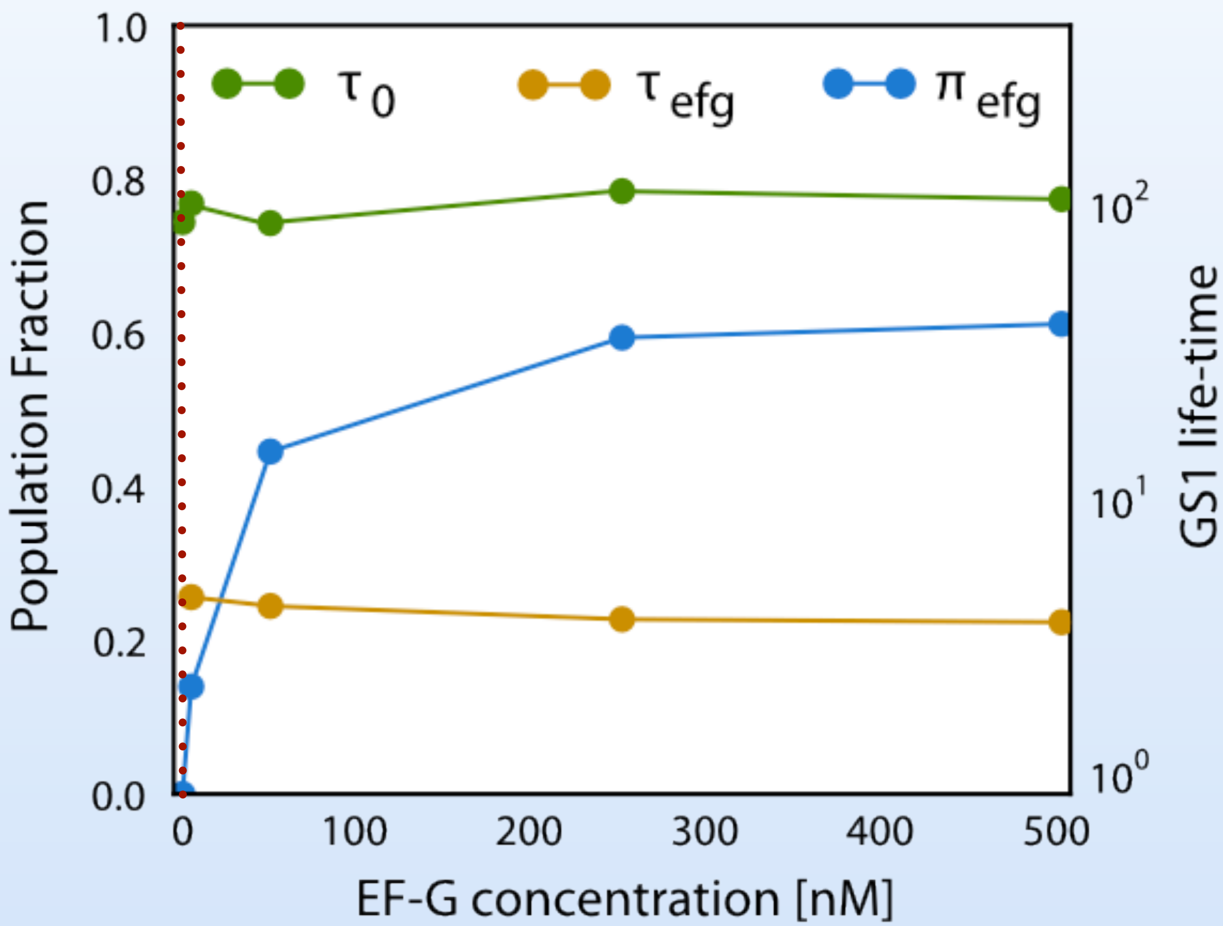


$$p(z_k) \sim e^{-G_k/k_B T}$$

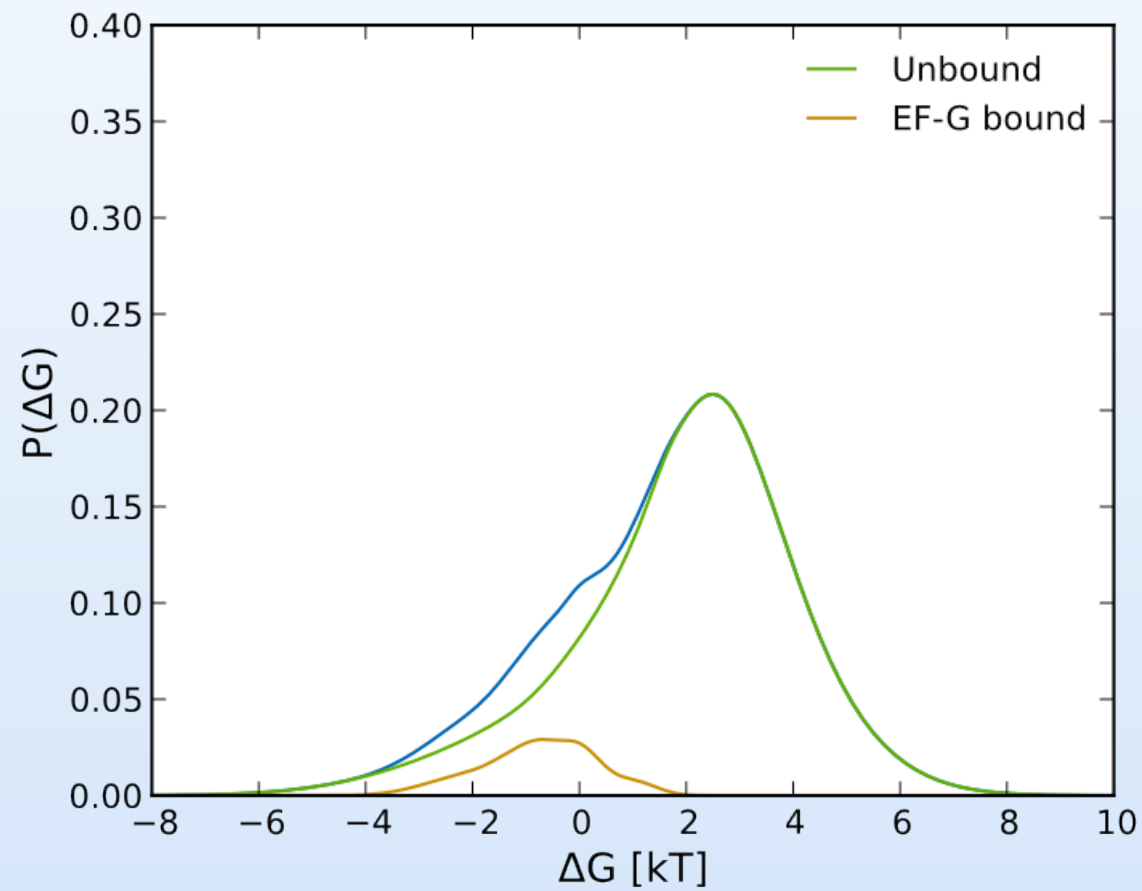
$$\log p(z_k) - \log p(z_l) = -(G_k - G_l)/k_B T + \text{cst.}$$

The role of EF-G binding

bound fraction and life-times

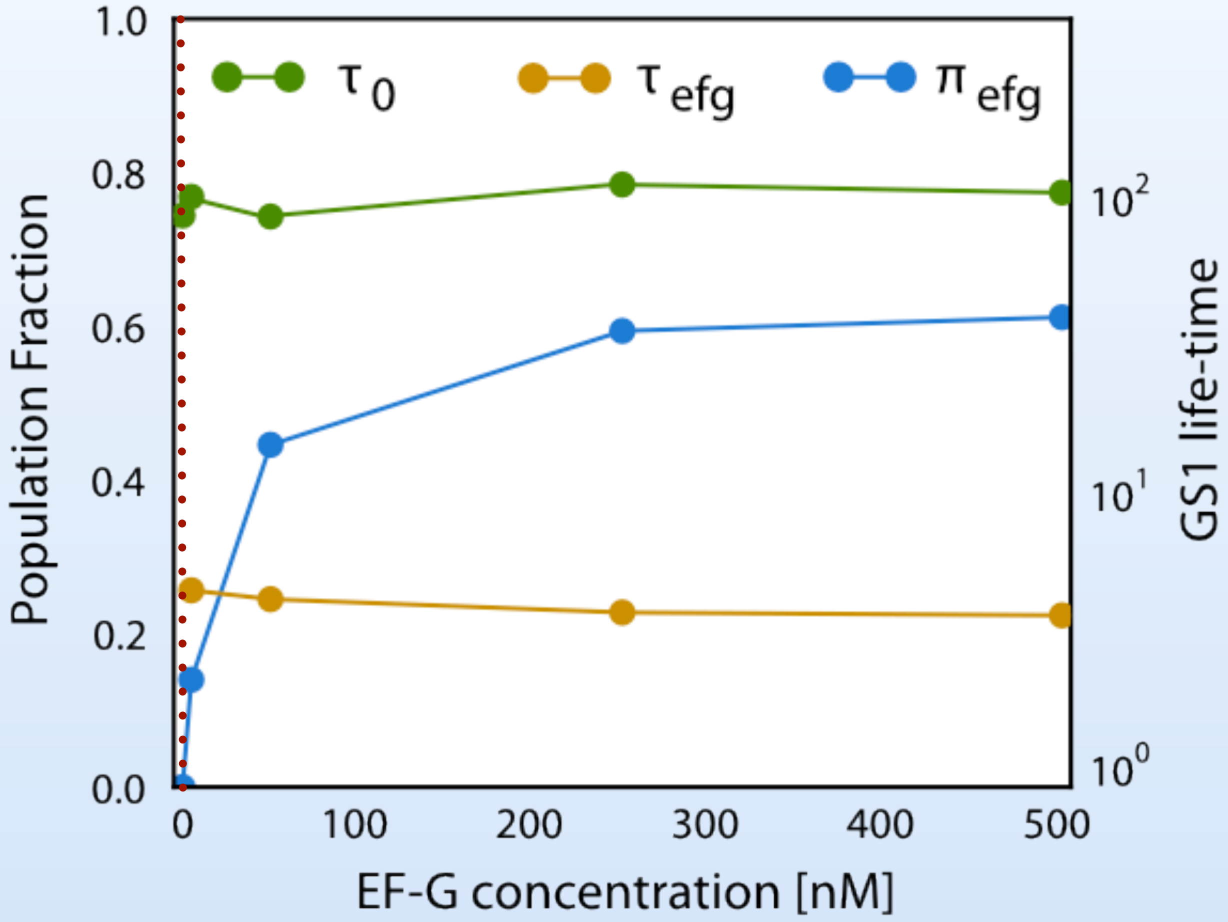


no EF-G

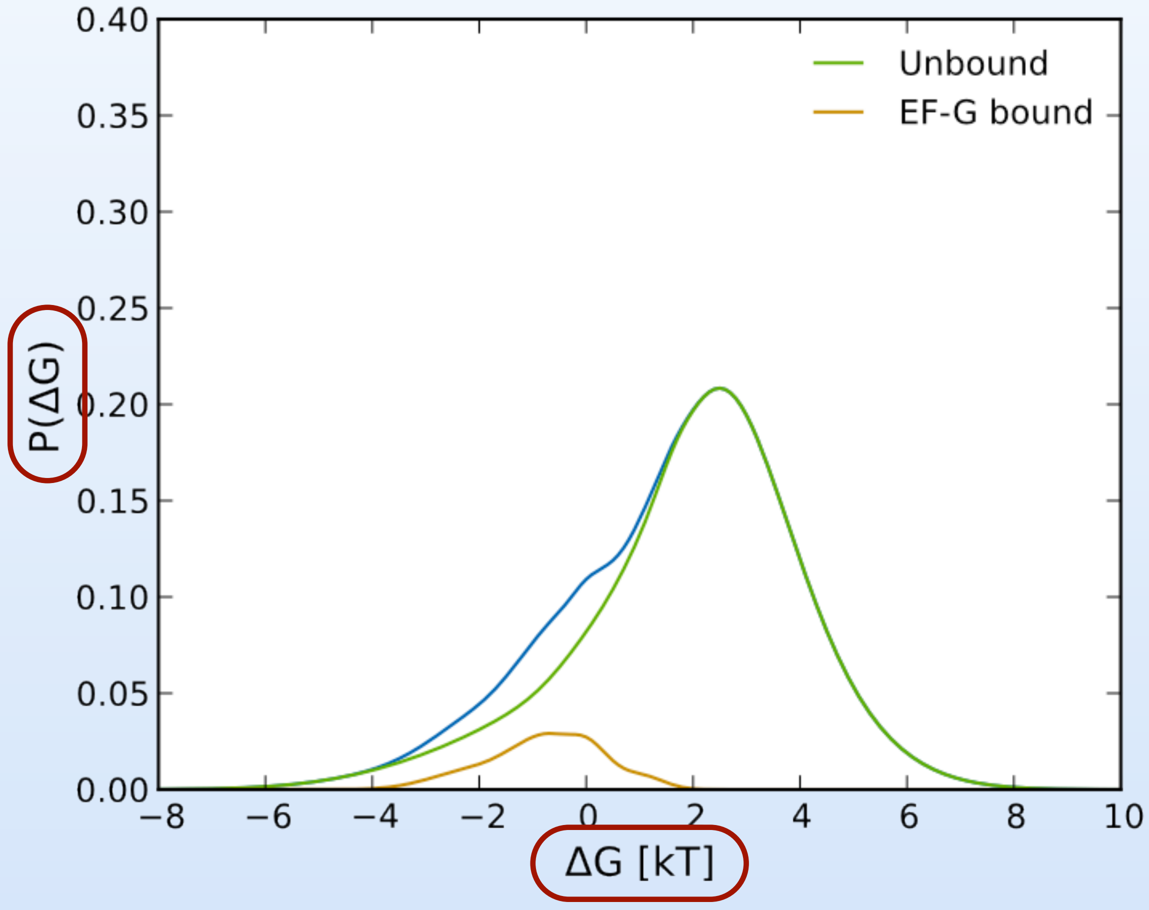


The role of EF-G binding

bound fraction and life-times

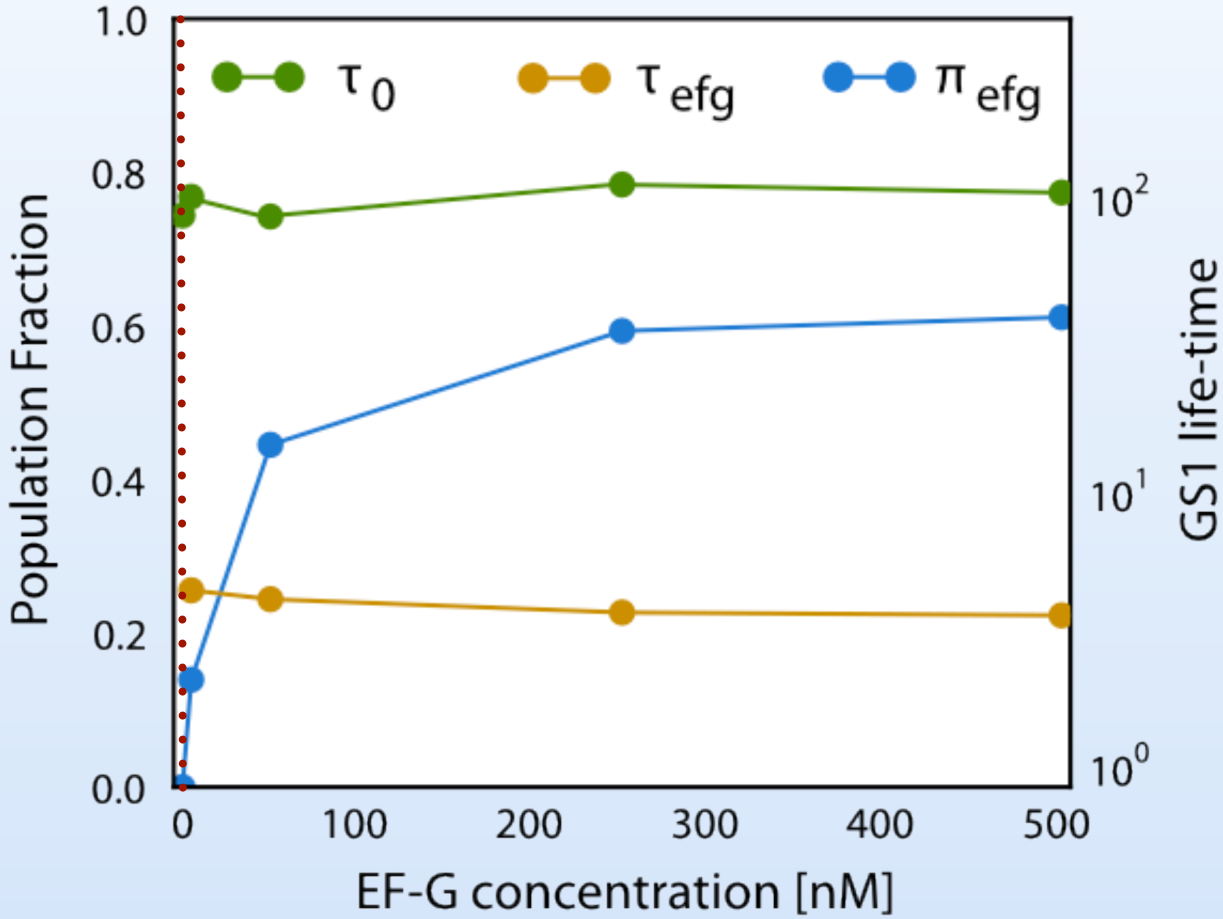


no EF-G

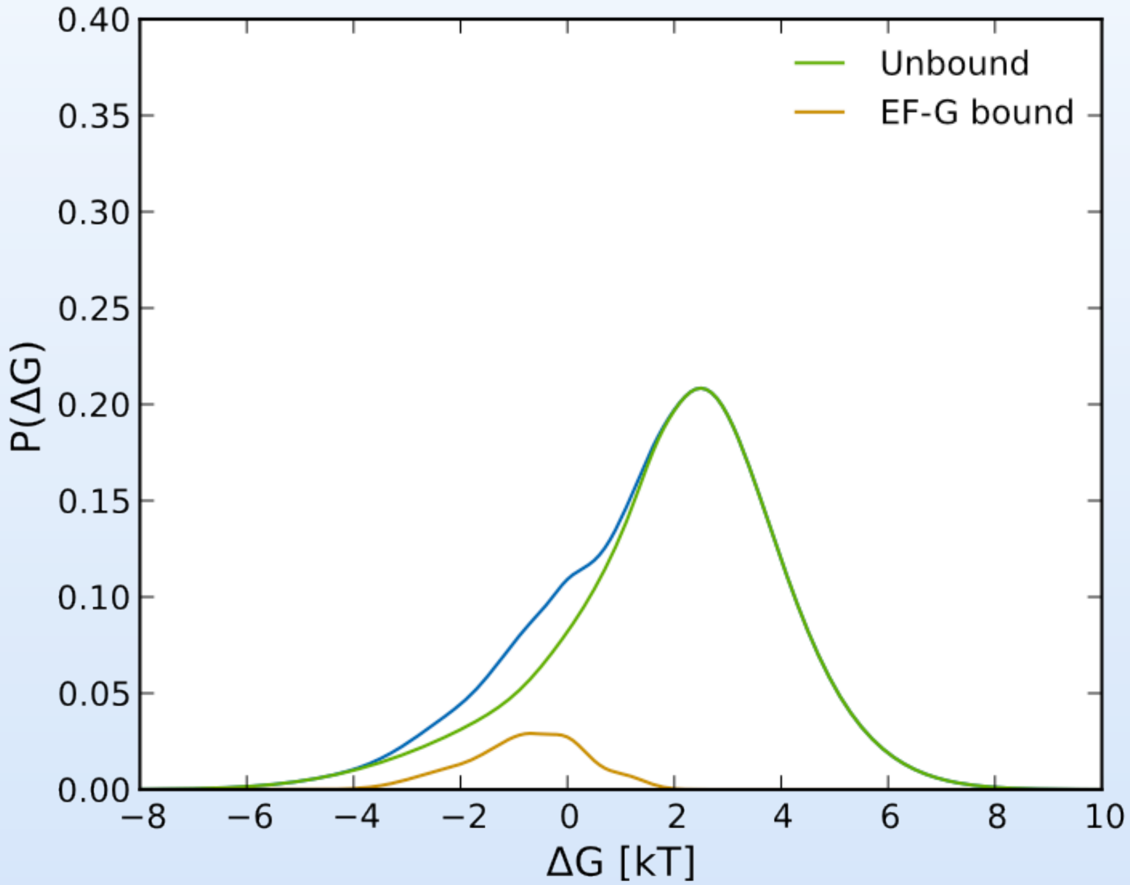


The role of EF-G binding

bound fraction and life-times

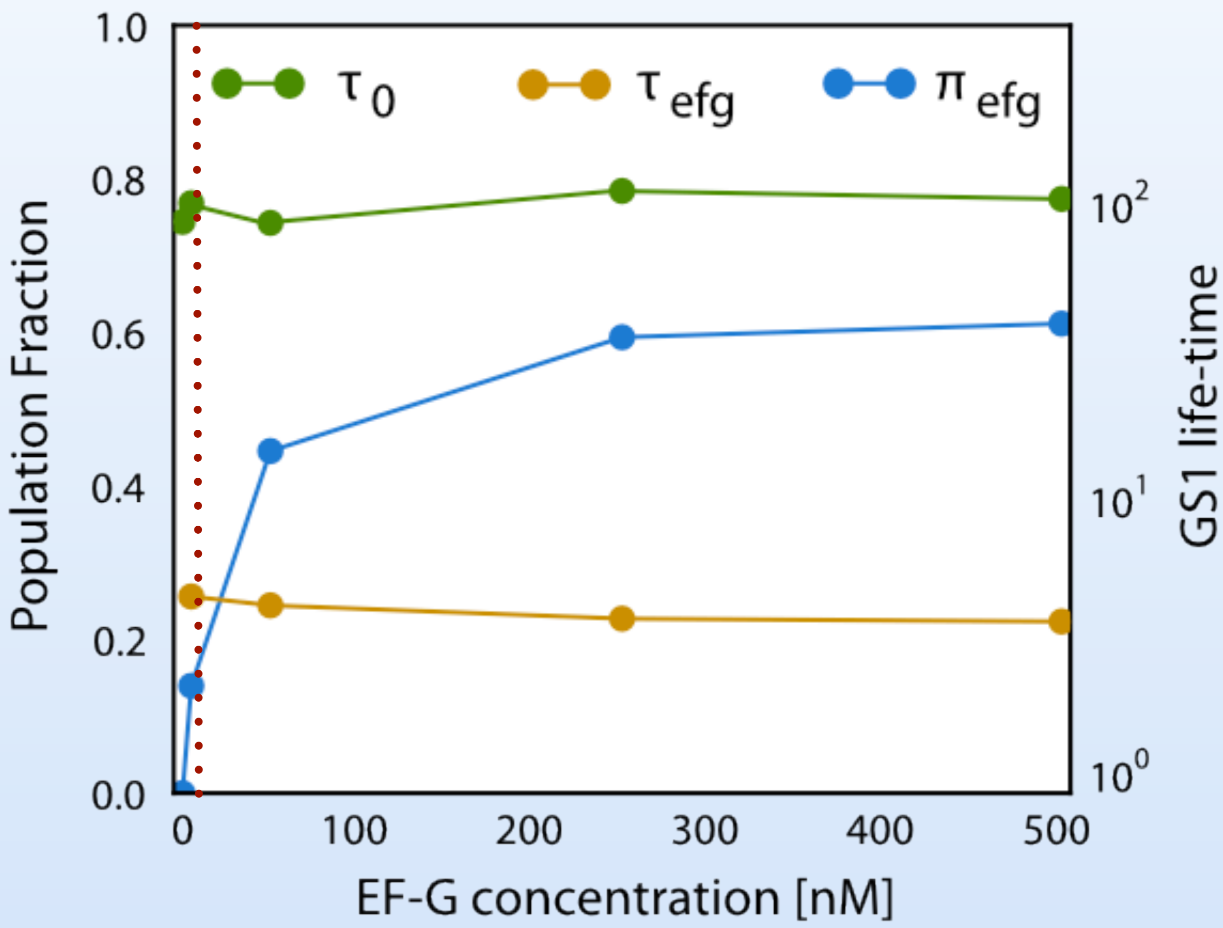


no EF-G

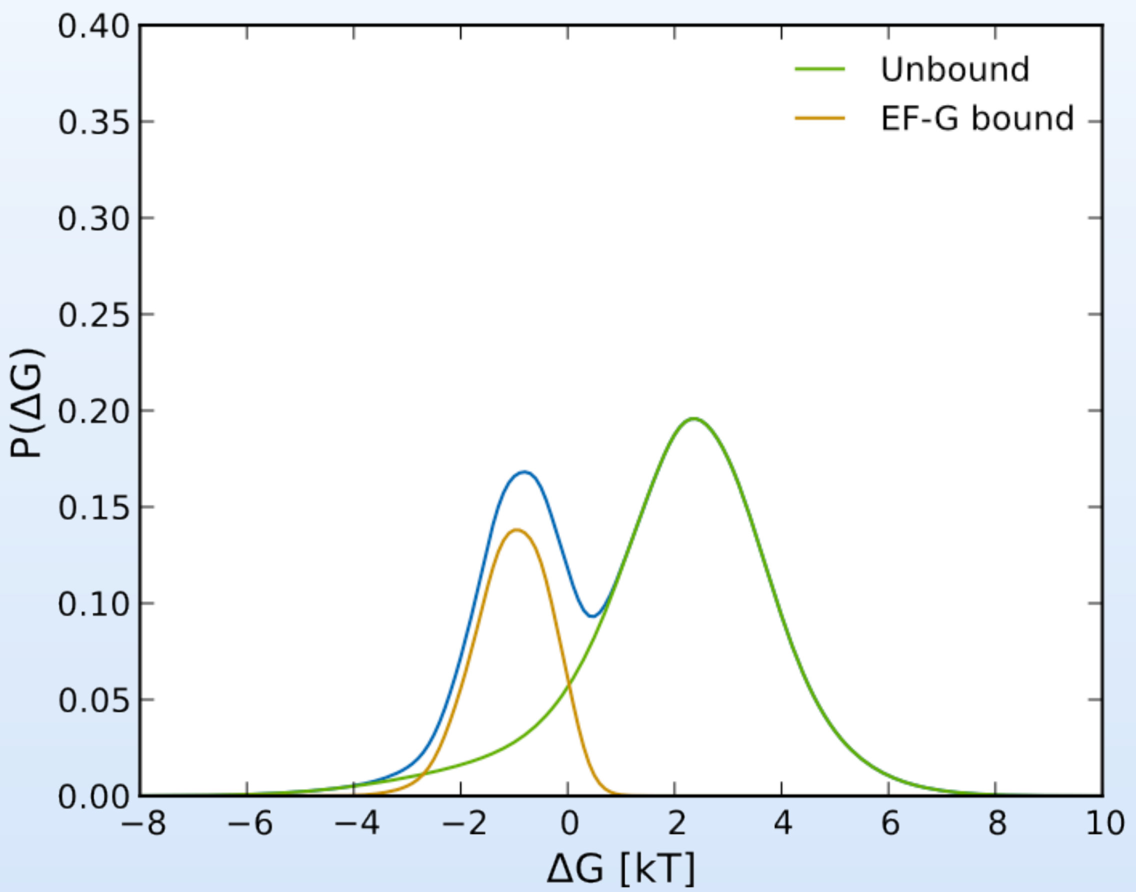


The role of EF-G binding

bound fraction and life-times

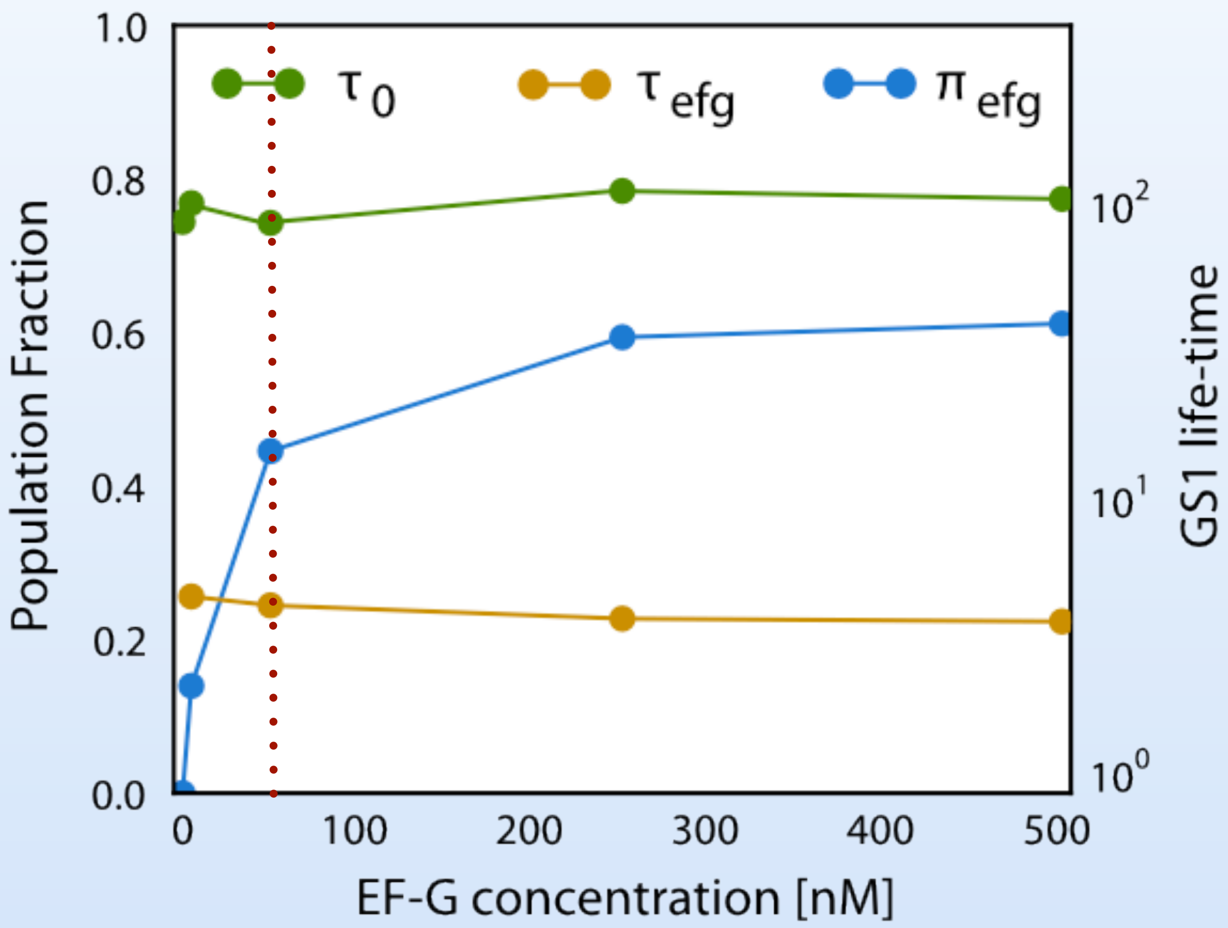


5 nM EF-G

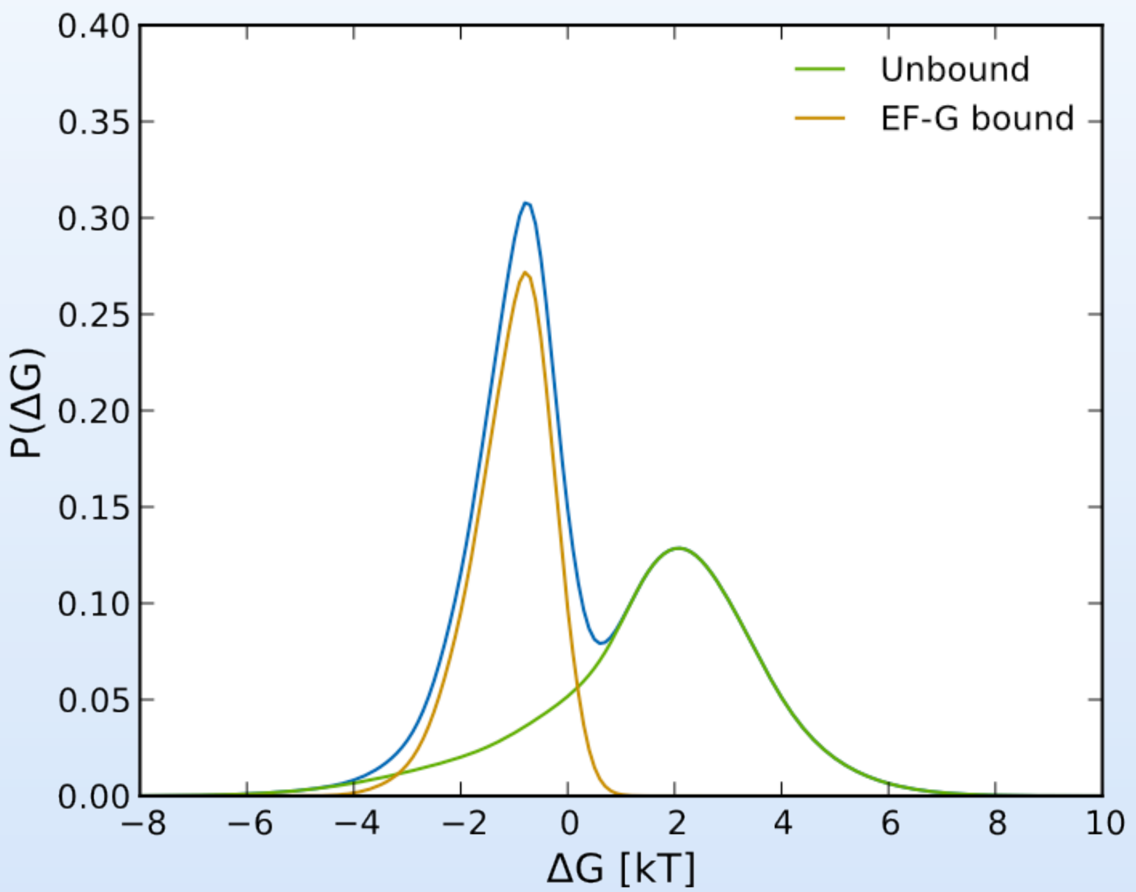


The role of EF-G binding

bound fraction and life-times

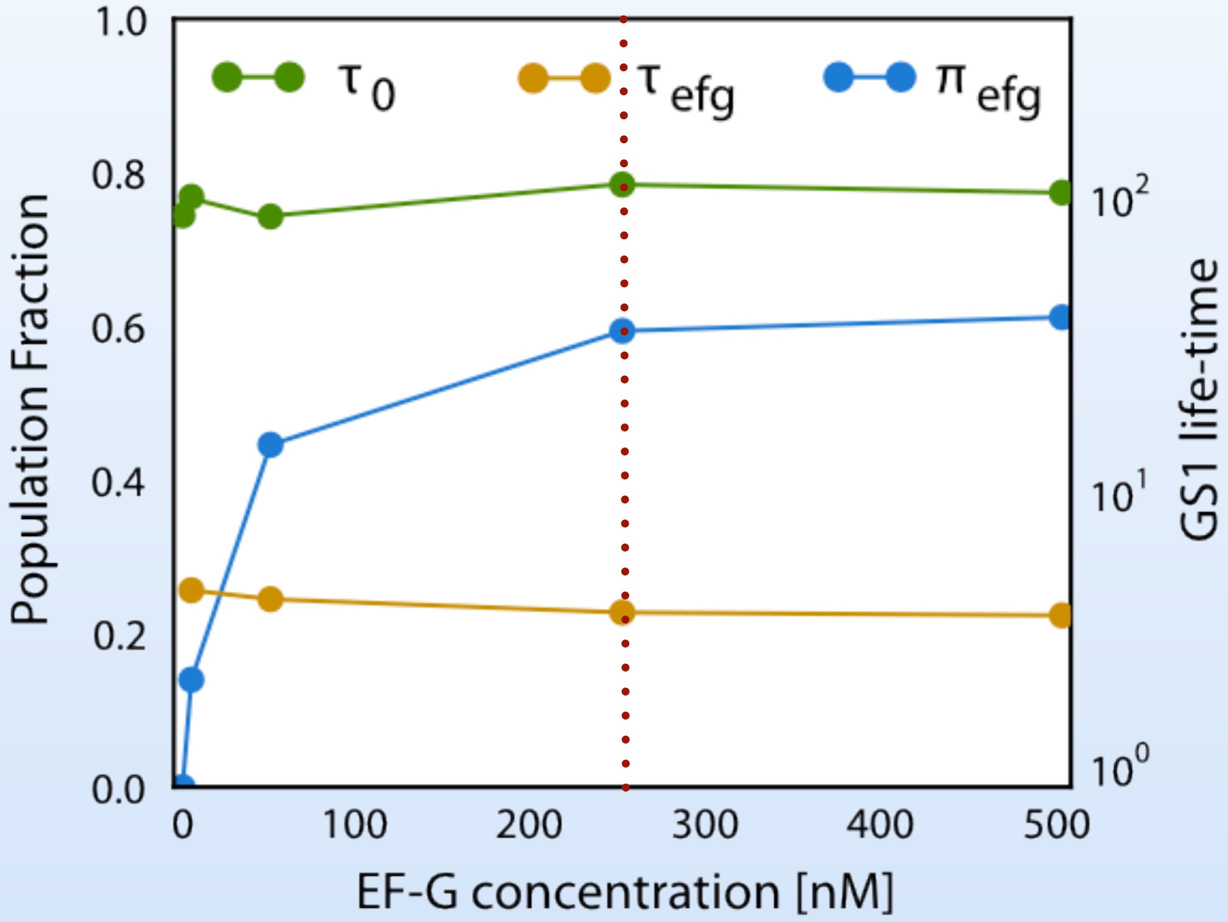


50 nM EF-G

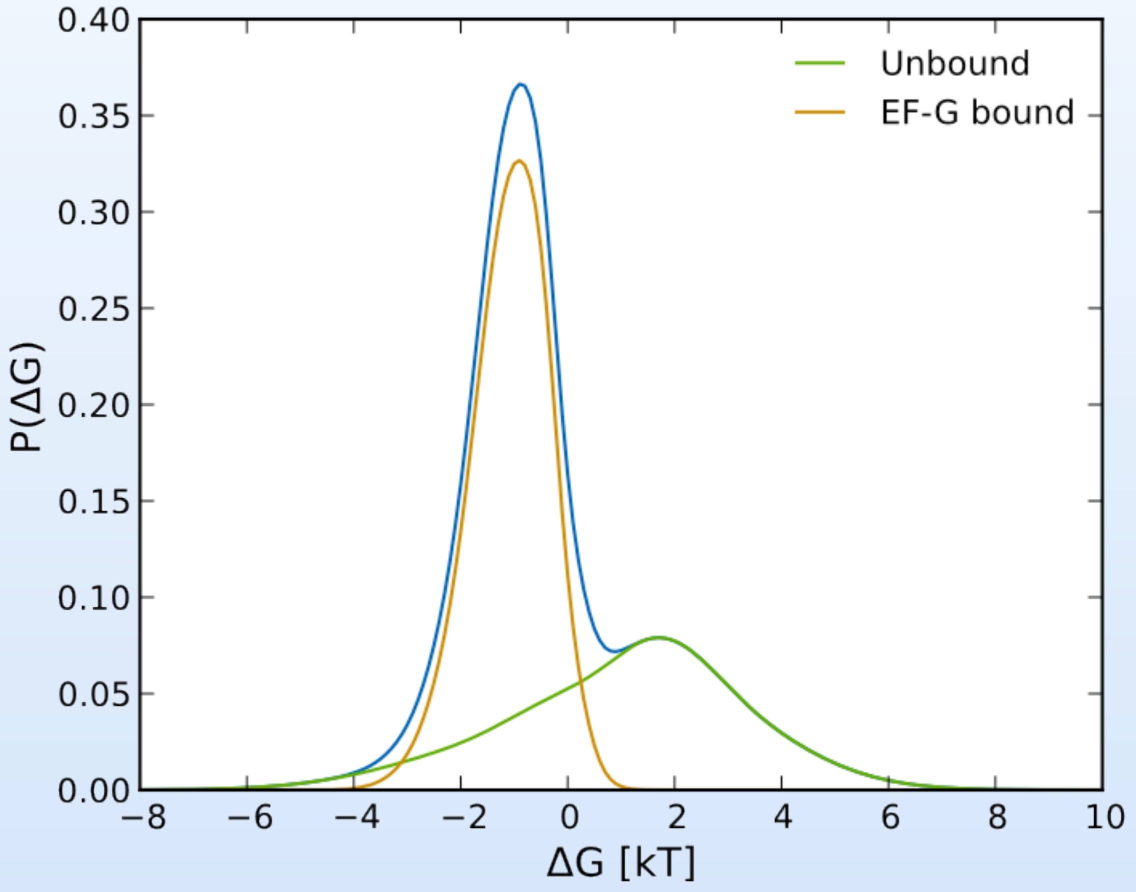


The role of EF-G binding

bound fraction and life-times

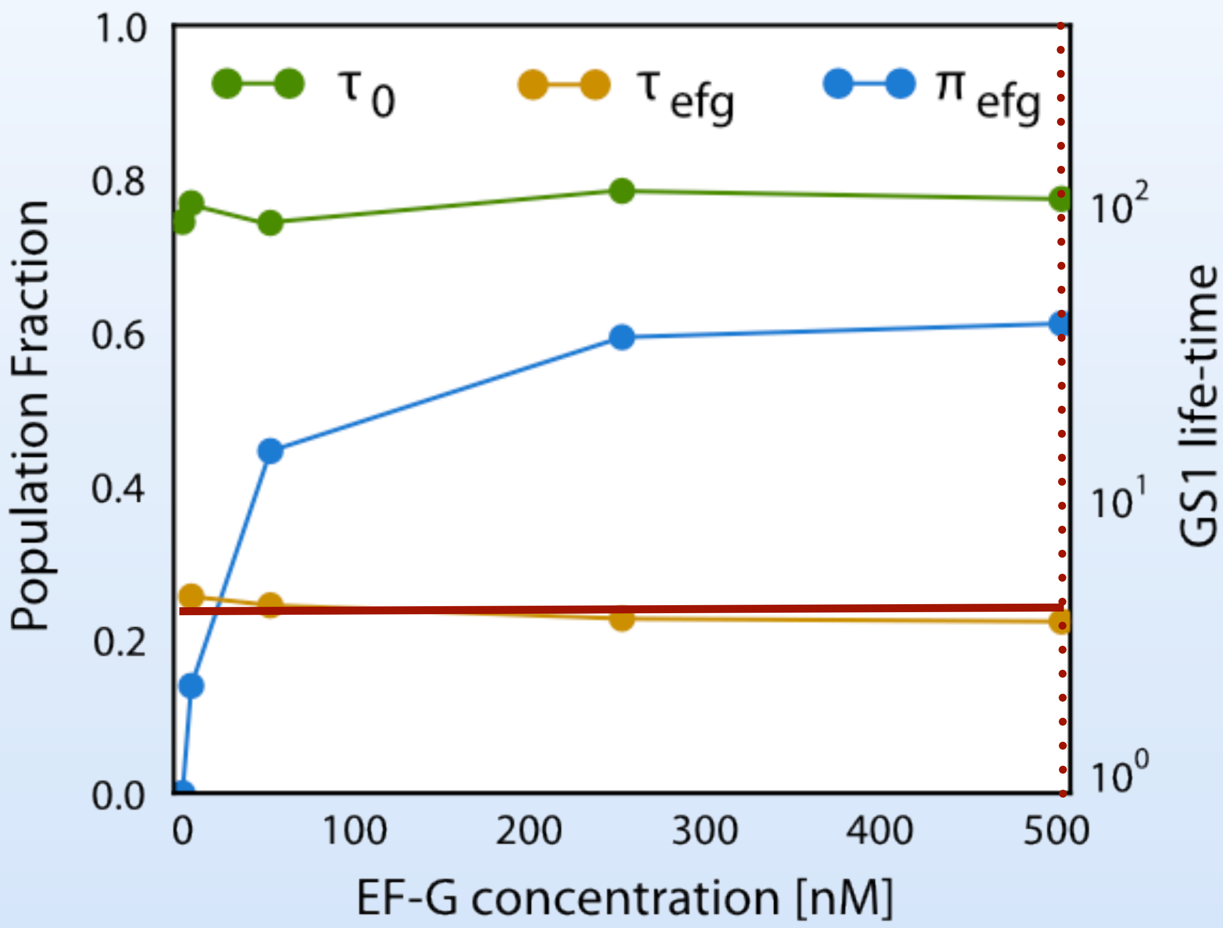


250 nM EF-G

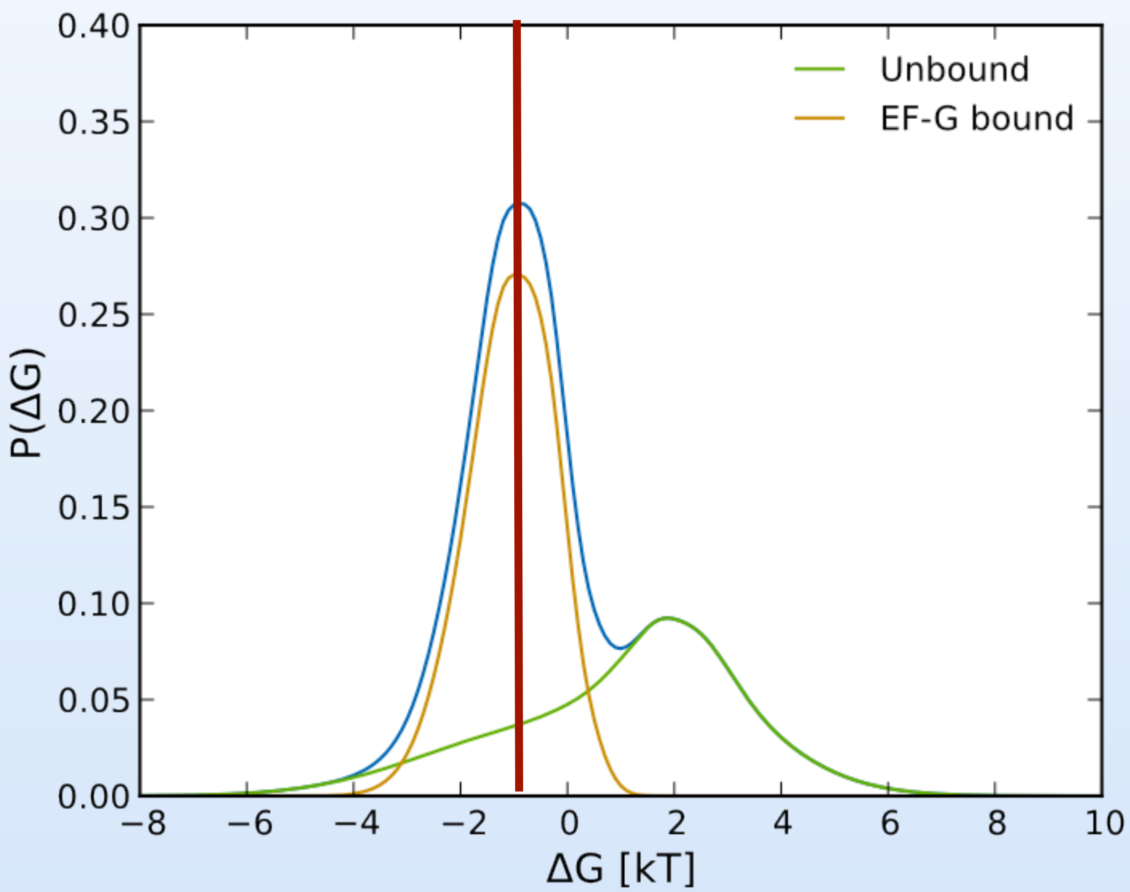


The role of EF-G binding

bound fraction and life-times

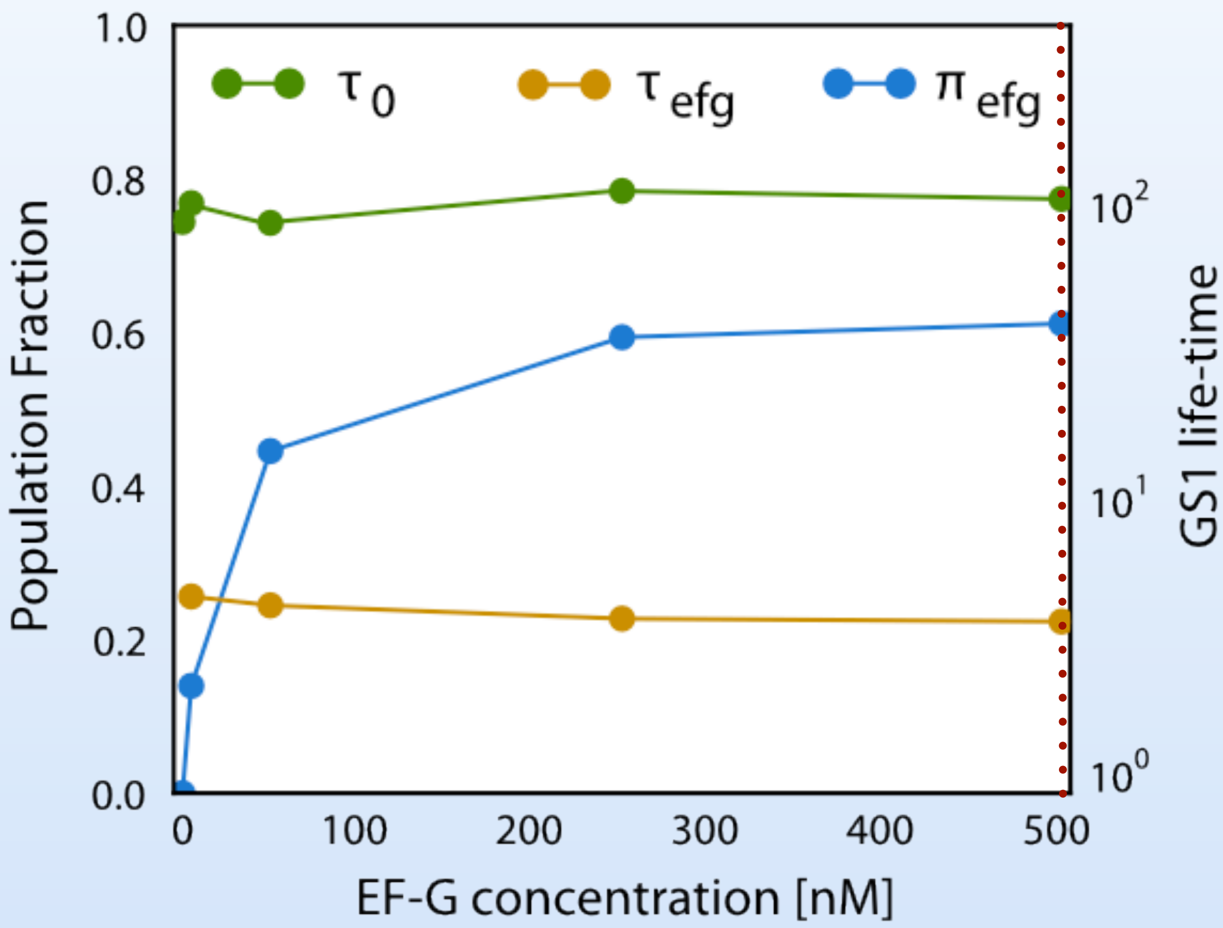


500 nM EF-G

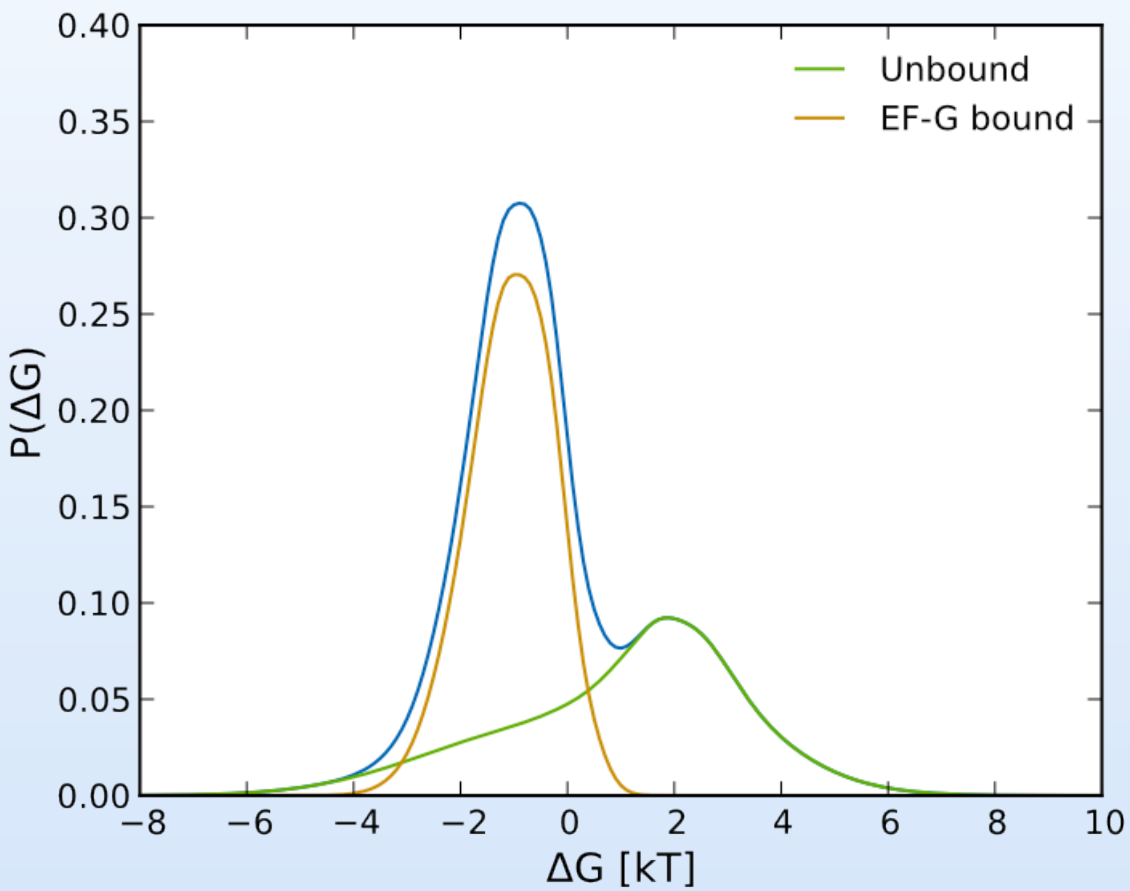


The role of EF-G binding

bound fraction and life-times

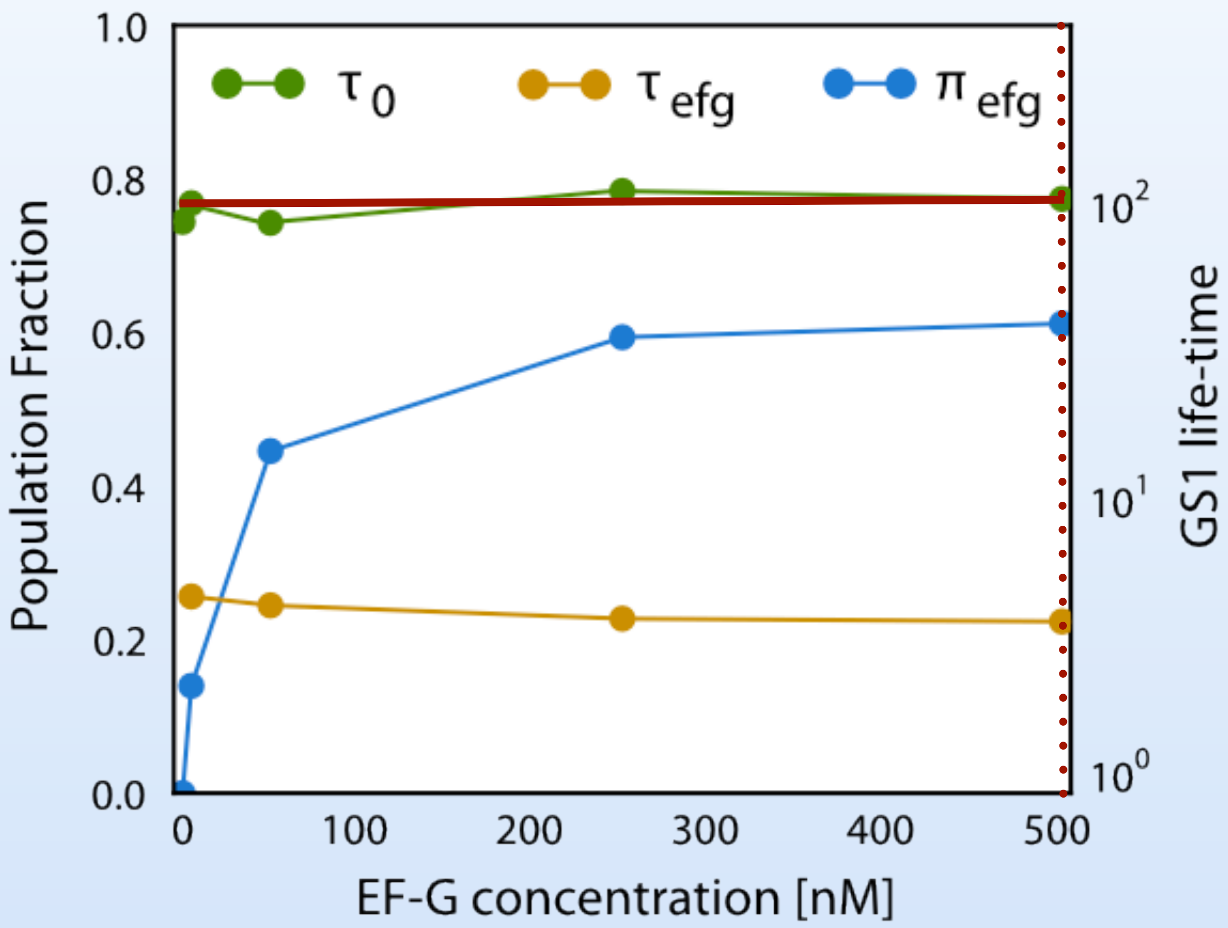


500 nM EF-G

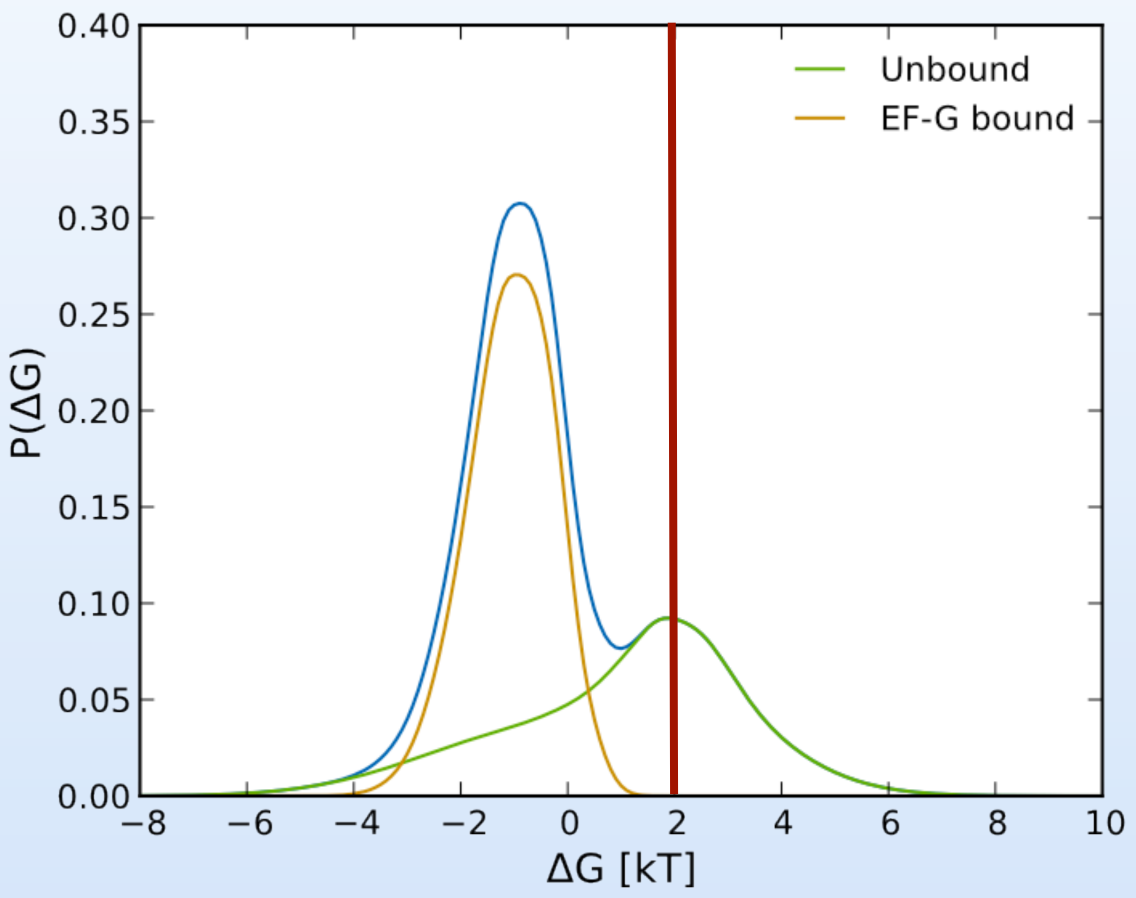


The role of EF-G binding

bound fraction and life-times

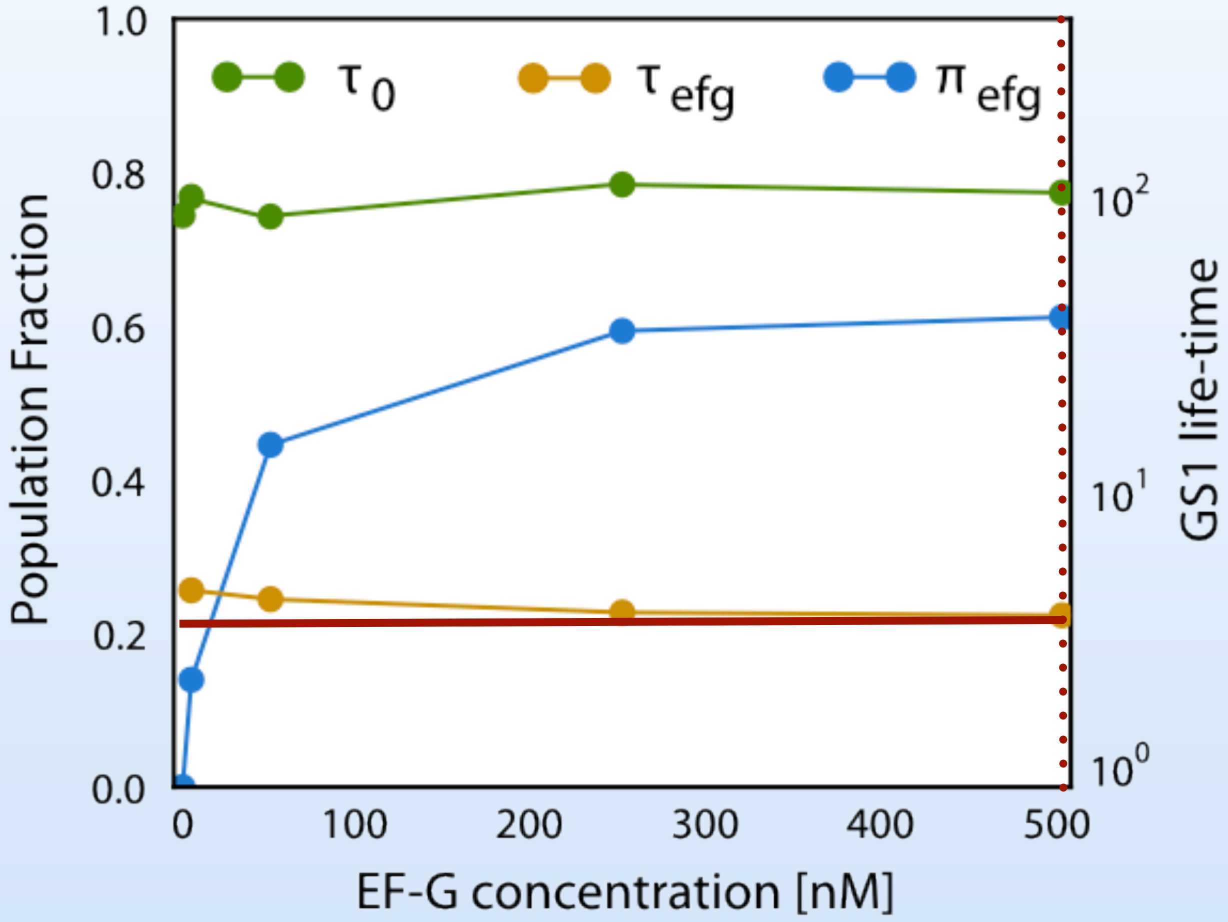


500 nM EF-G

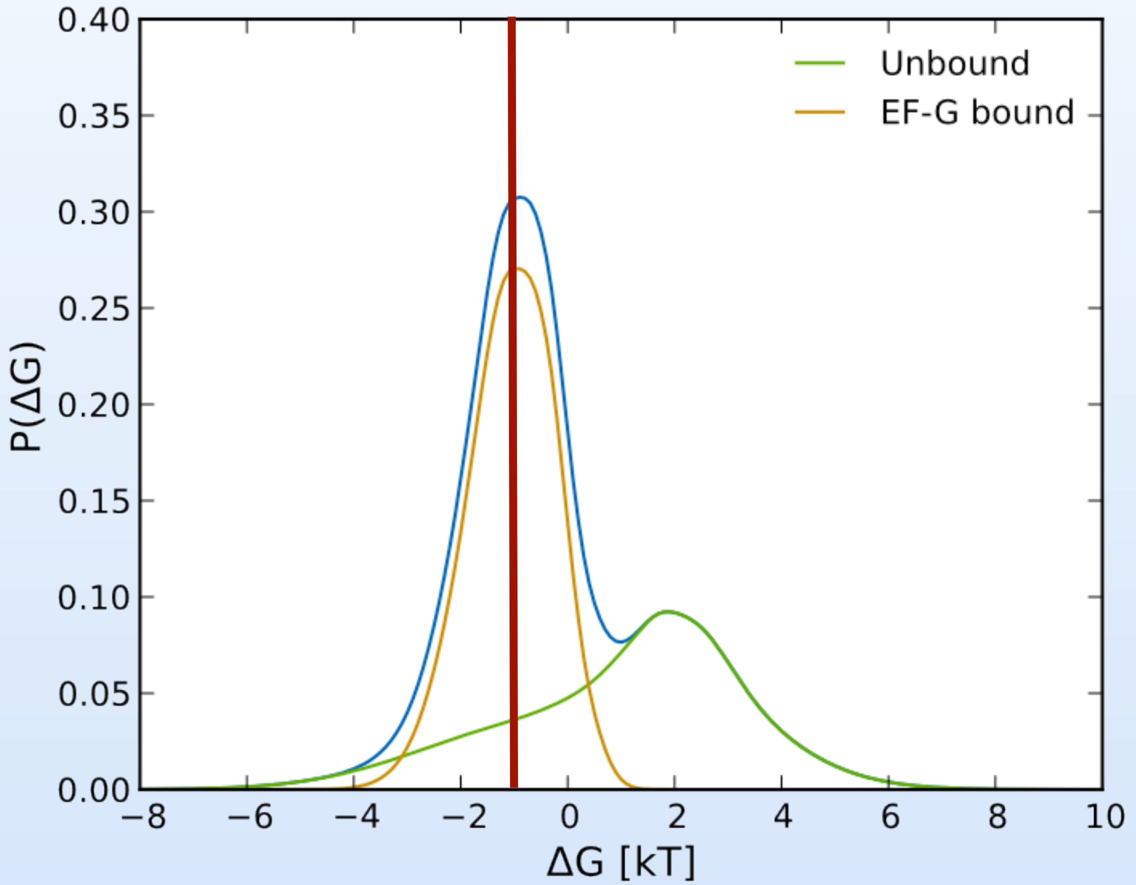


The role of EF-G binding

bound fraction and life-times

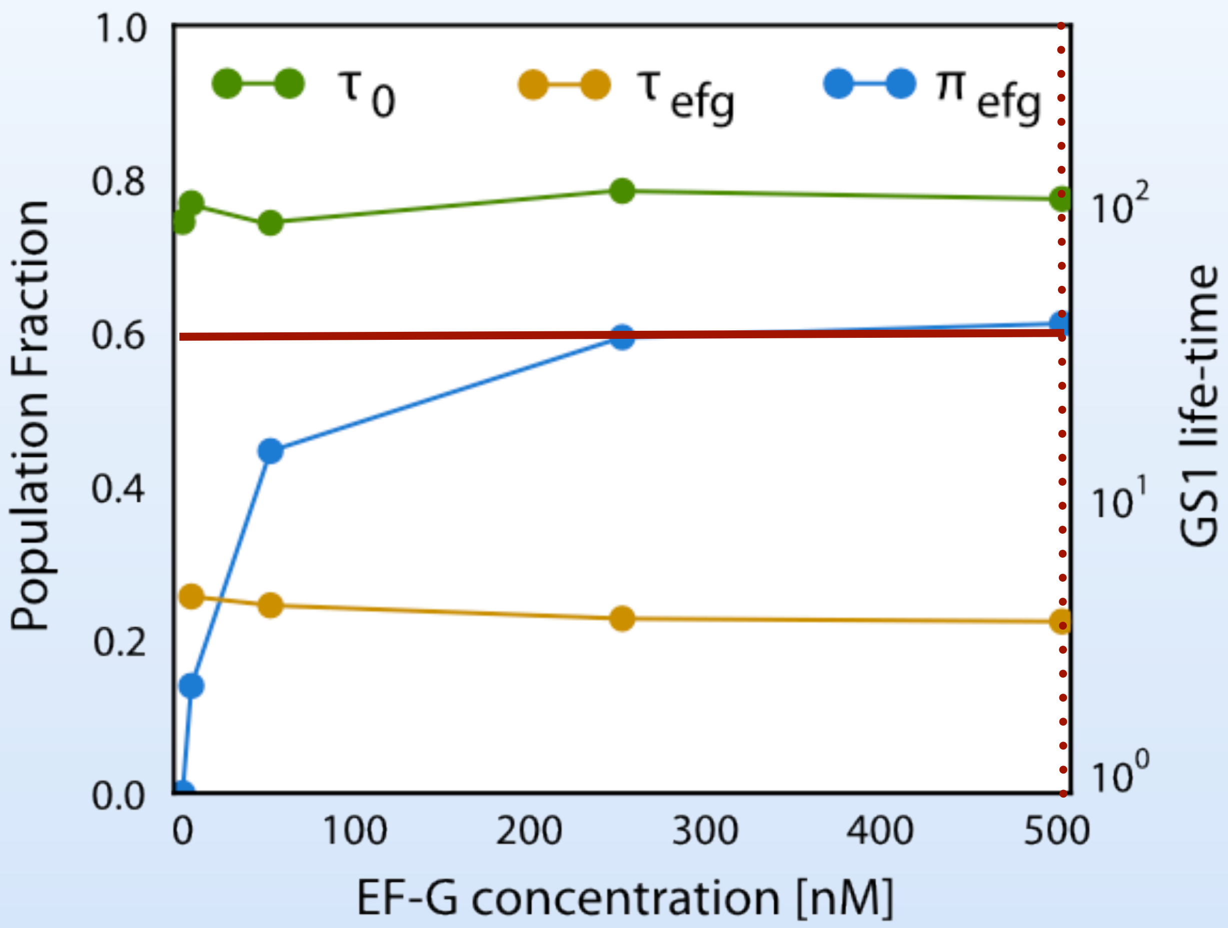


500 nM EF-G

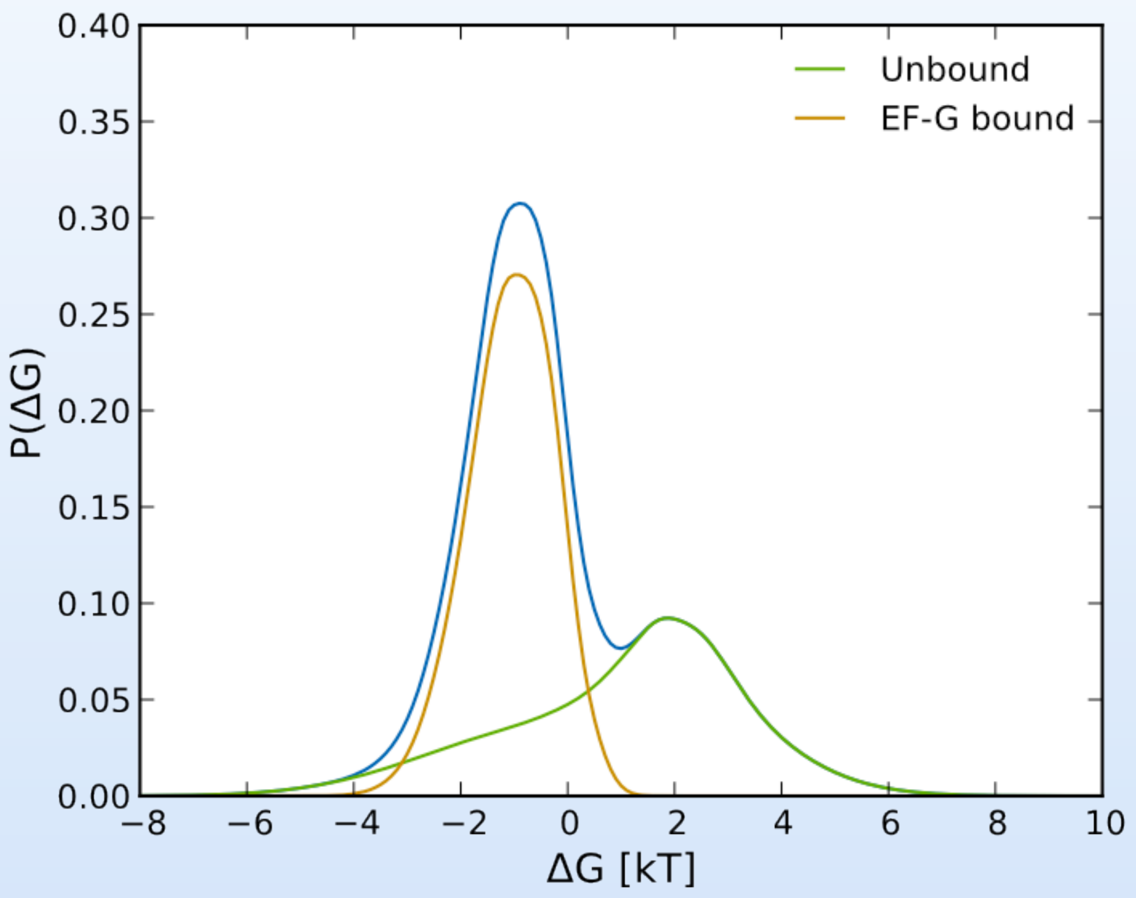


The role of EF-G binding

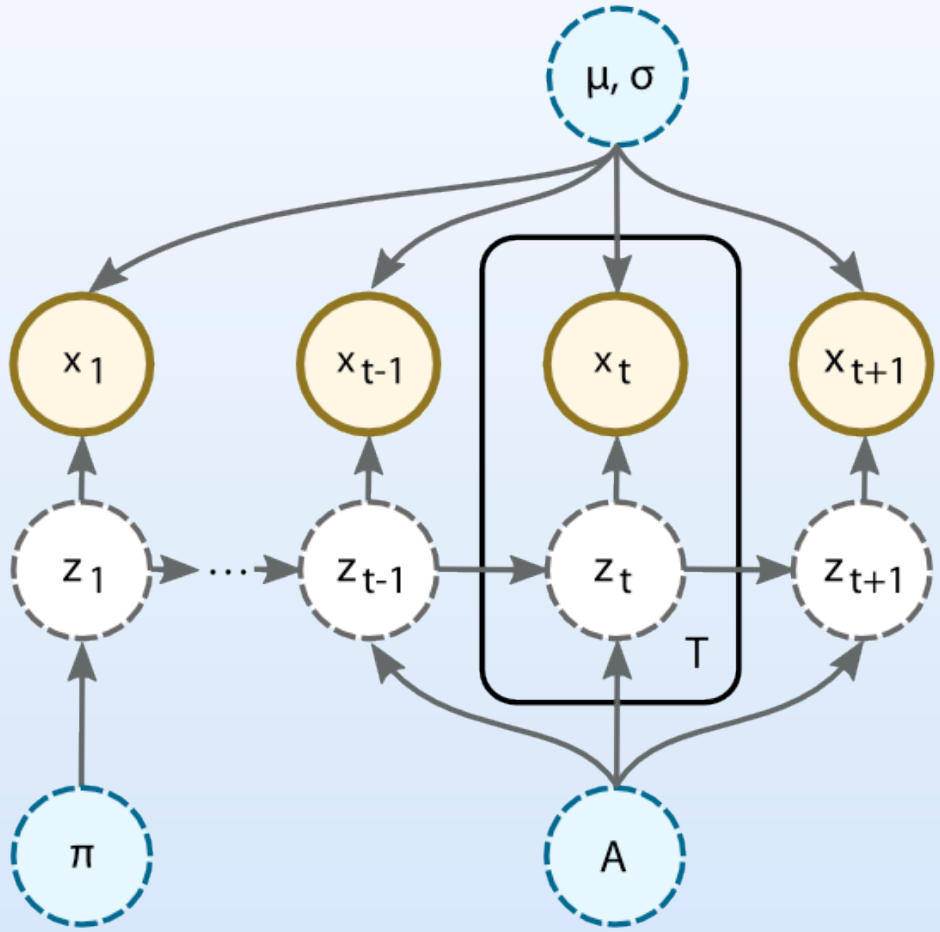
bound fraction and life-times



500 nM EF-G

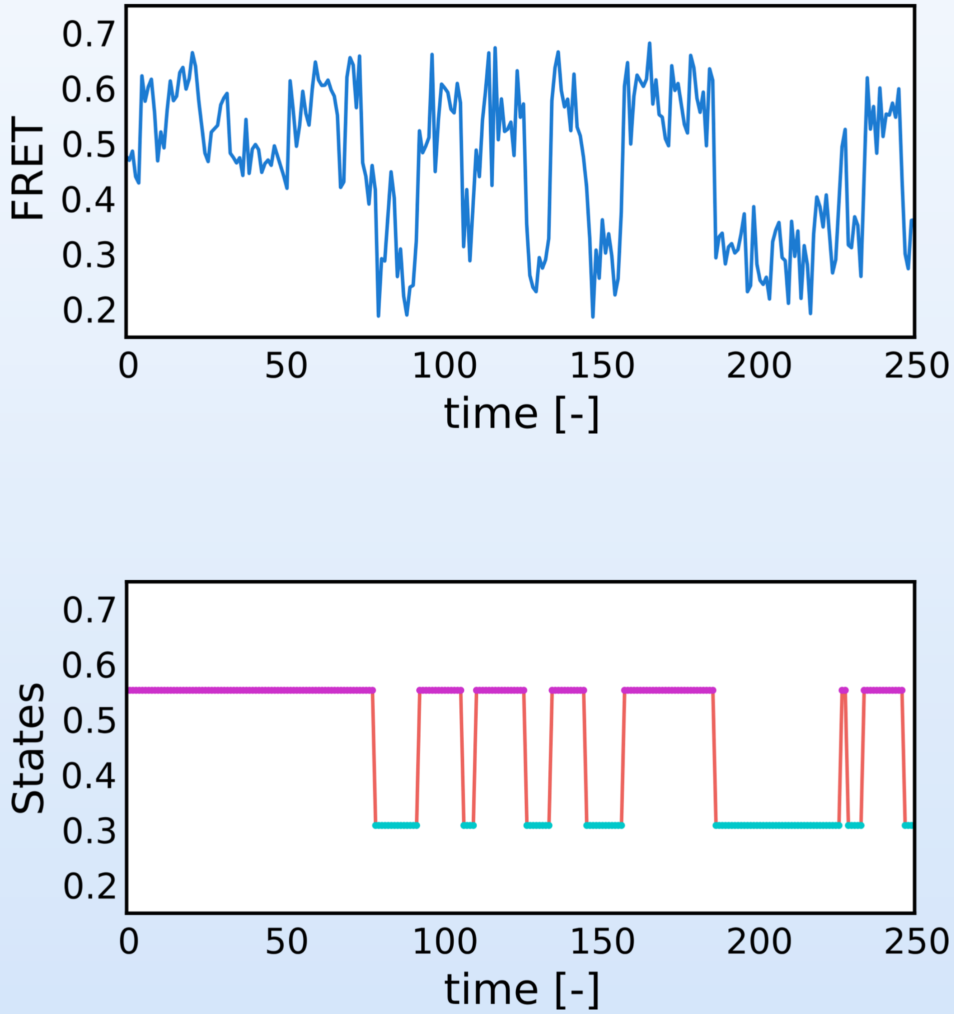


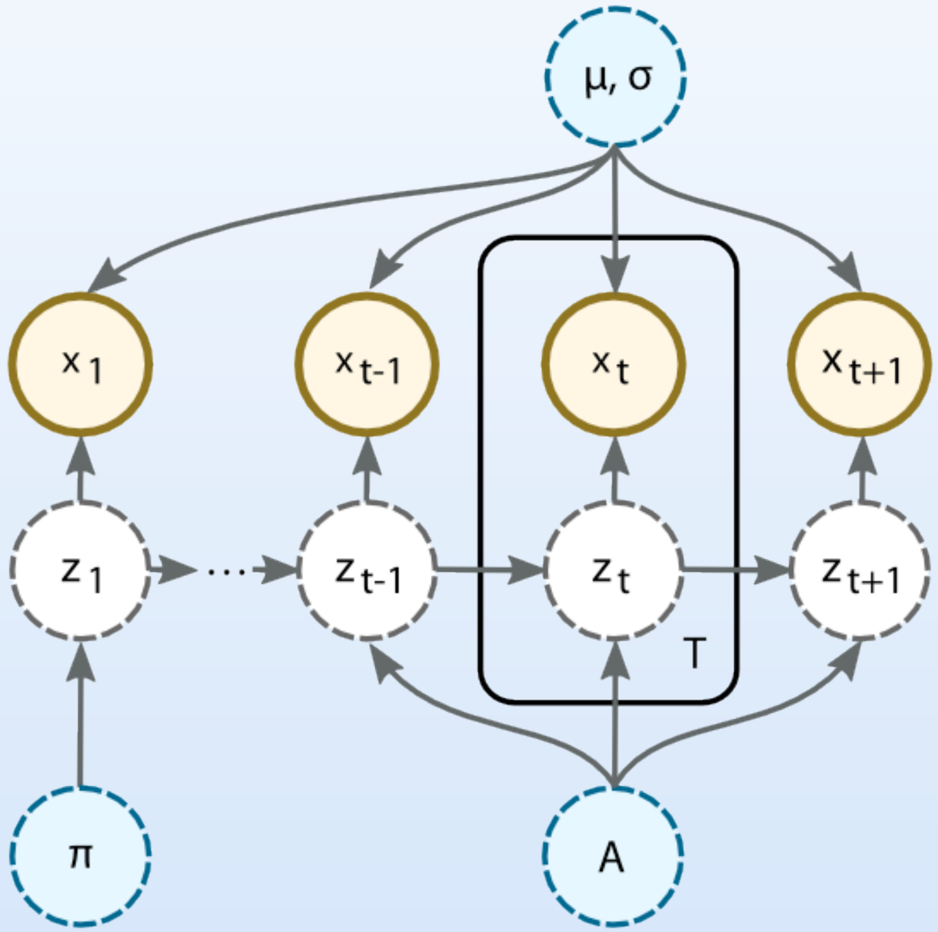
The Basic Idea



Graphical Model: Encodes Assumptions

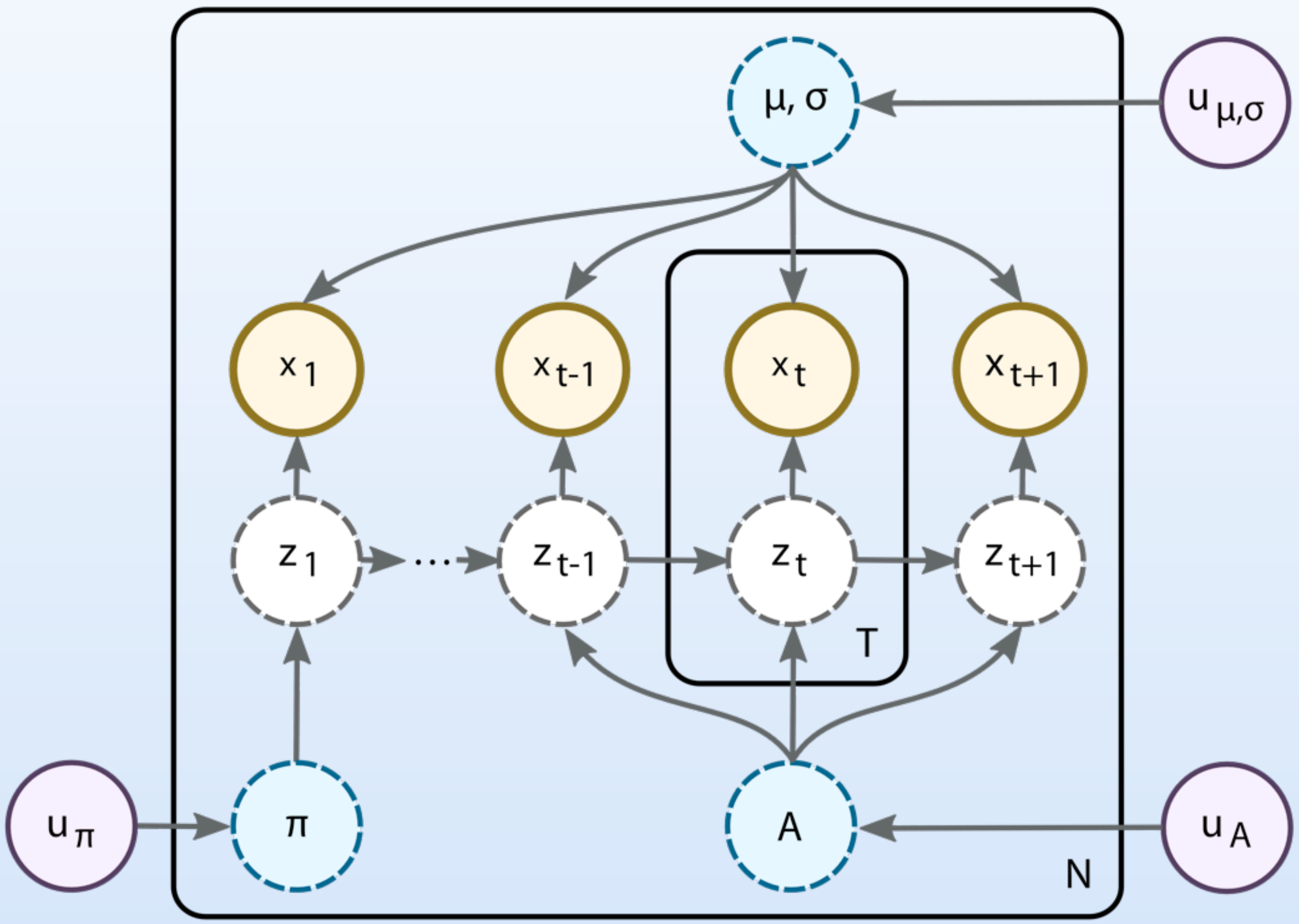
The Basic Idea





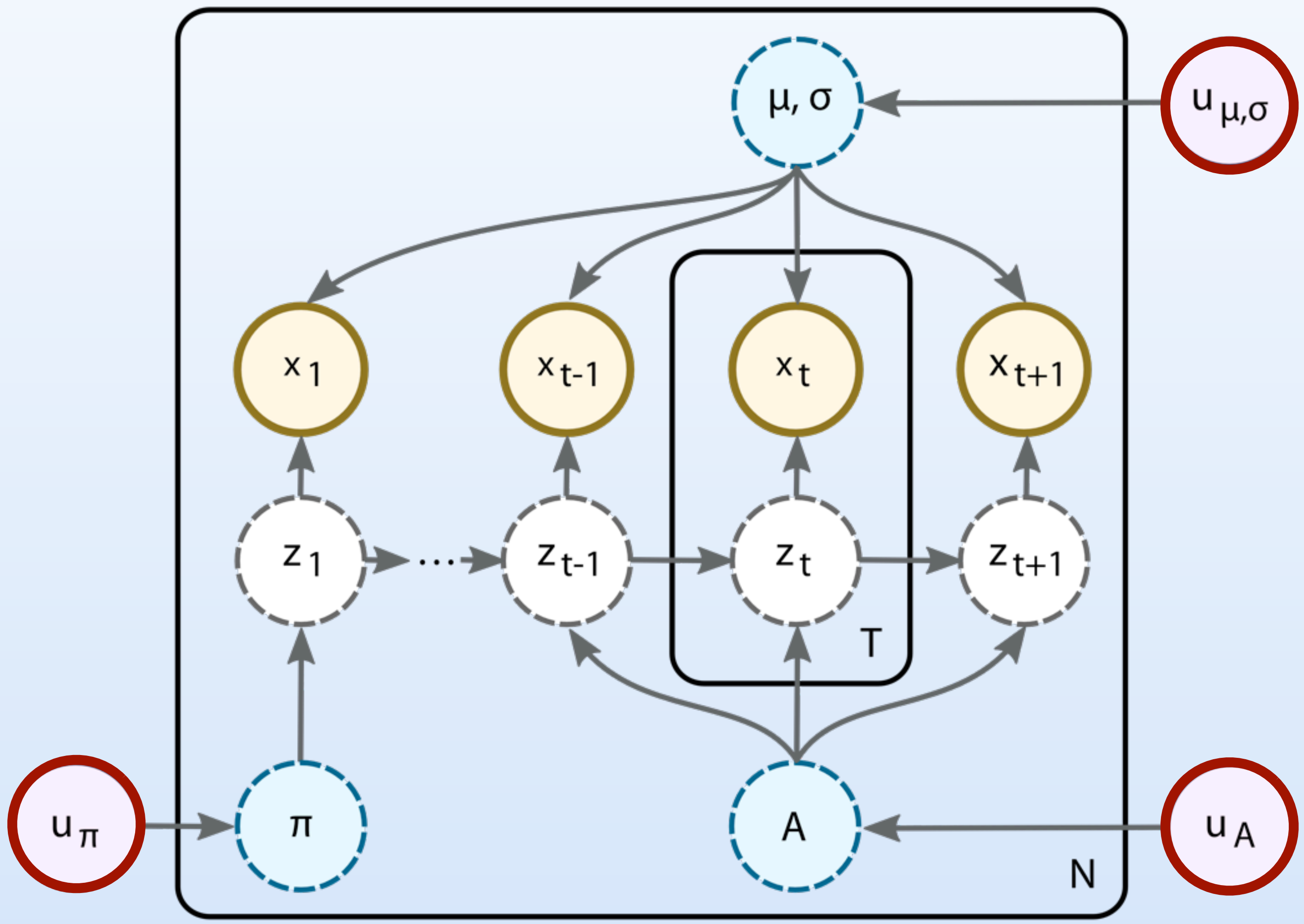
Inference: Estimate most probable states and rates

The Basic Idea



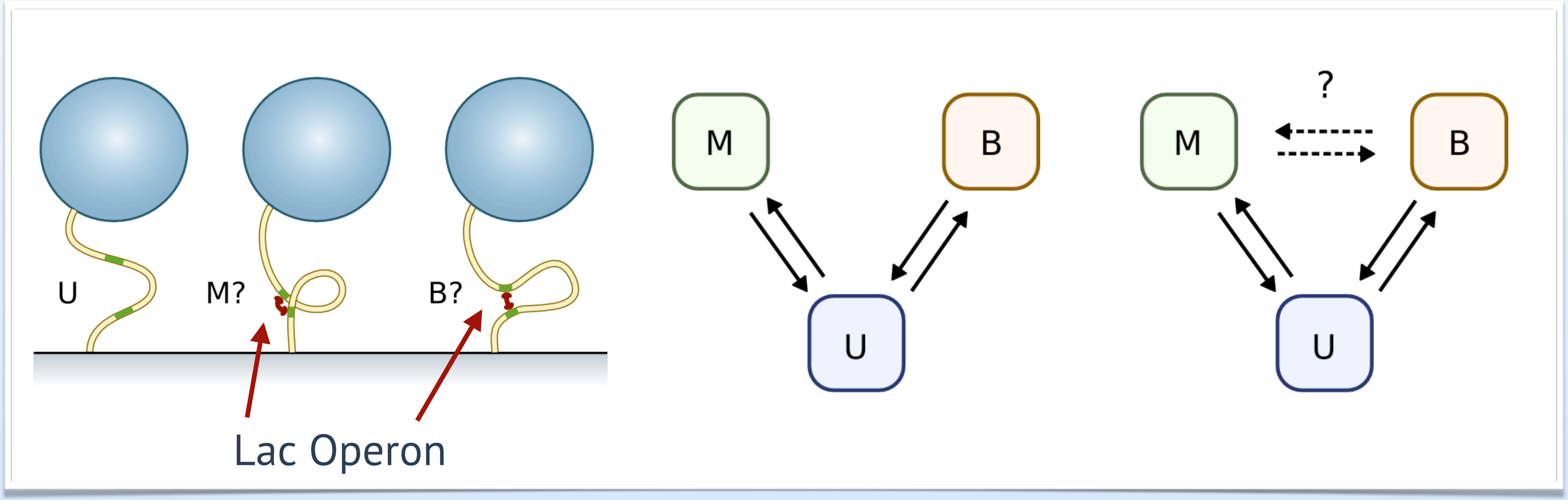
Empirical Bayes: Common Features in Ensemble

The Basic Idea



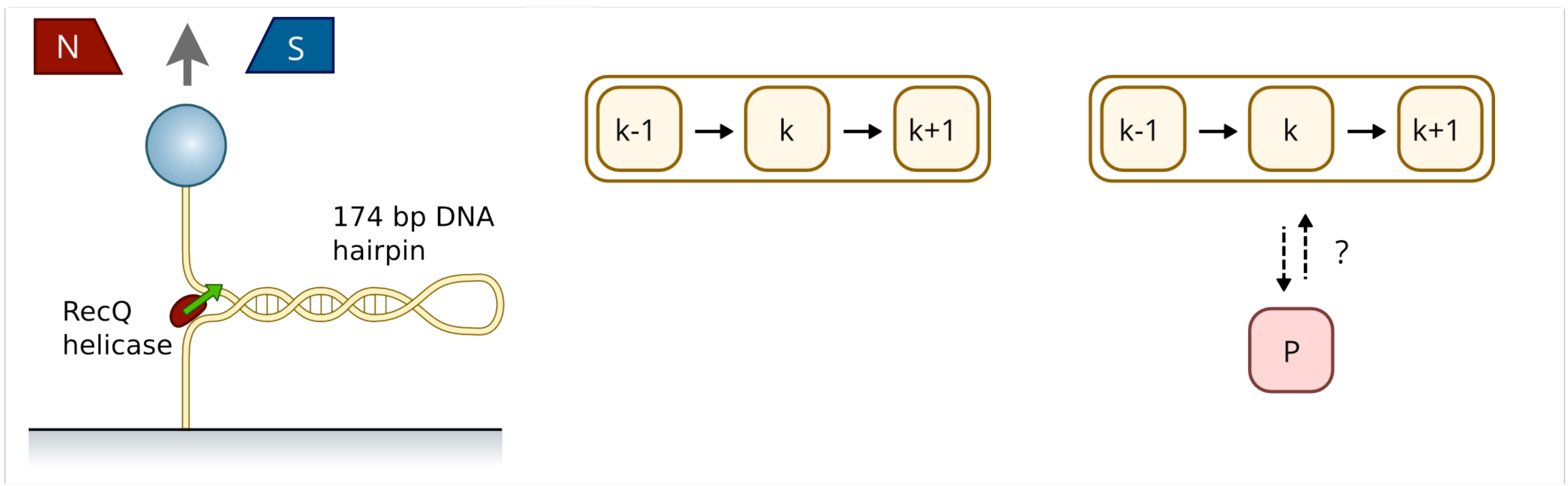
Empirical Bayes: Common Features in Ensemble

The Basic Idea



Model comparison: Mechanistic Hypothesis Testing

The Basic Idea



Model comparison: Mechanistic Hypothesis Testing

Empirical Bayes

- Parameter-free
- Learns from all data at once
- Extremely robust under noise
- Access to heterogeneous kinetics

Co-conspirators

Chemistry

Ruben
Gonzalez

Jingyi
Fei

Applied Maths

Chris
Wiggins

Jonathan
Bronson

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

NWO

Variational Bayes

Log-Evidence

$$L = \log p(x \mid u) = \log\left[\sum_z \int d\theta\, p(x, z \mid \theta) p(\theta|u)\right]$$

Lower Bound

$$\mathcal{L} = \sum_z \int d\theta\, q(z) q(\theta \mid w) \log\left[\frac{p(x, z, \theta \mid u)}{q(z) q(\theta \mid w)}\right]$$

$$\geq \log p(x \mid u)$$

$$q(z) q(\theta \mid w) \simeq p(z, \theta \mid x)$$

Variational Bayes

Lower bound tight for true posterior

$$L = \sum_z \int d\theta \, p(z, \theta \mid x) \log \left[\frac{p(x, z, \theta \mid u)}{p(z, \theta \mid x)} \right]$$

$$= \sum_z \int d\theta \, p(z, \theta \mid x) \log [p(x \mid u)]$$

$$= \log p(x \mid u)$$

$$\mathcal{L} = \log p(x \mid u) - D_{kl}[q(z) q(\theta \mid w) \,\|\, p(z, \theta \mid x)]$$