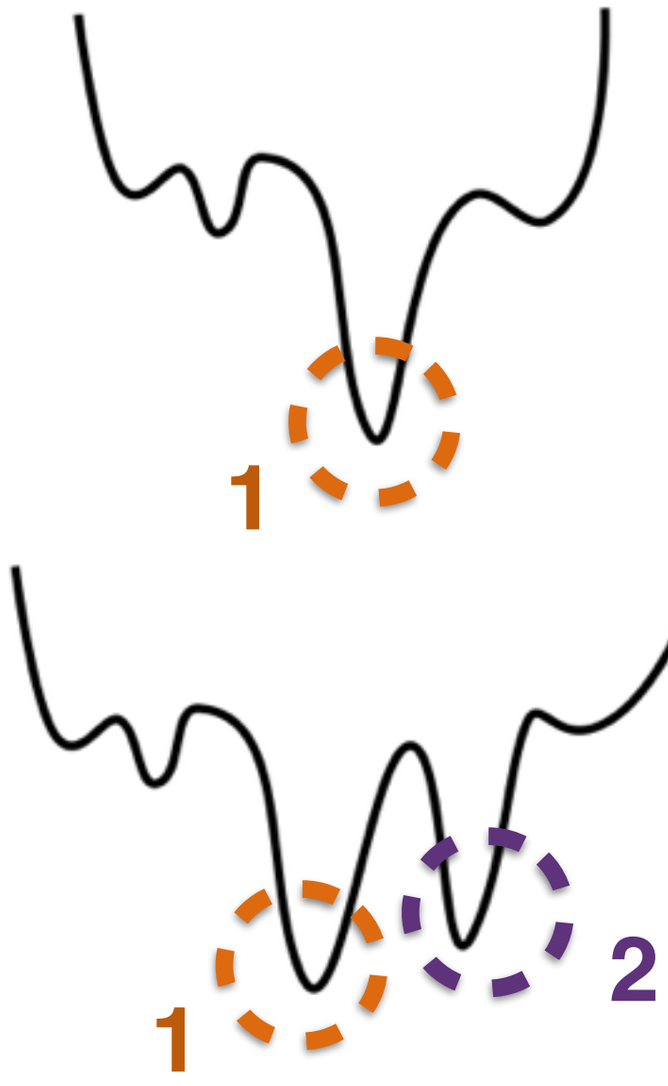
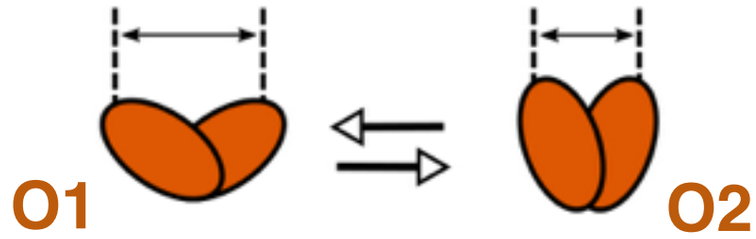
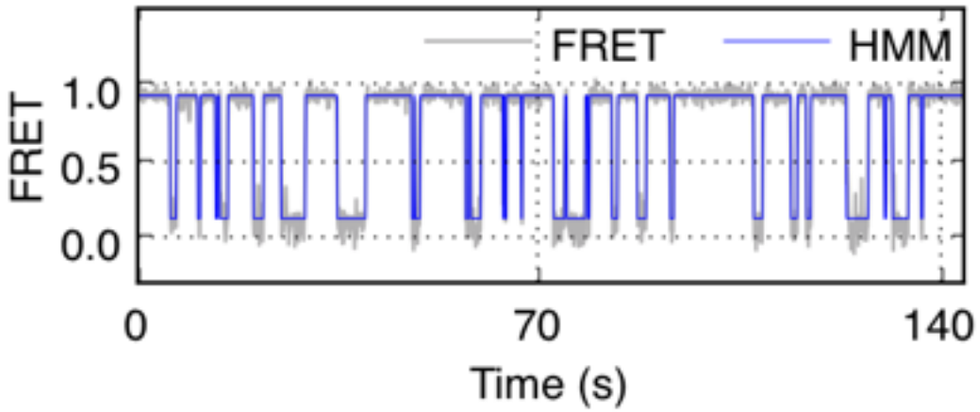
Decoding dynamic heterogeneity in single molecule data

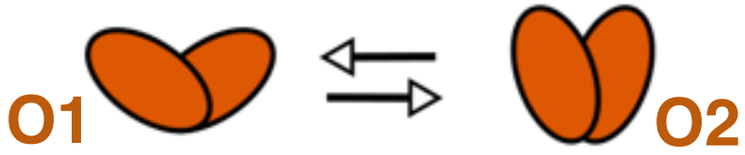
Wonseok Hwang
Post-doc
KIAS
Sep 2015

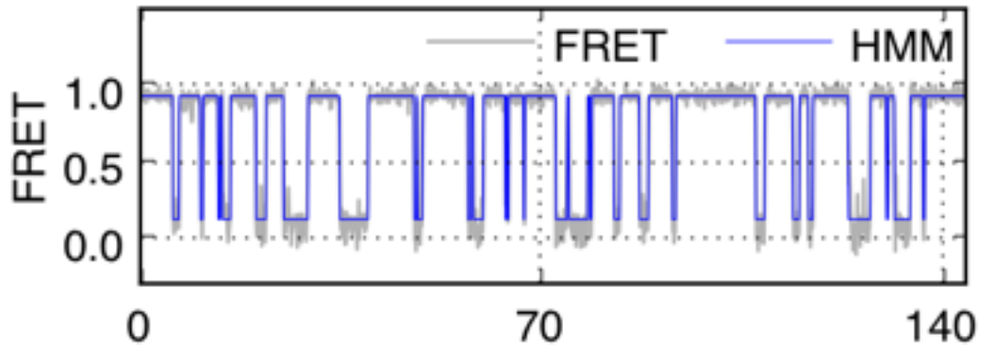
Folding energy landscape

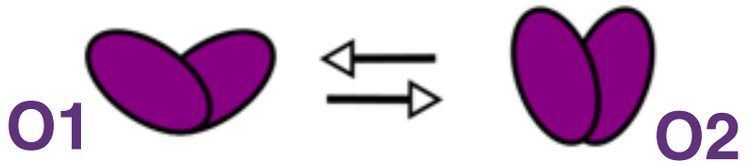


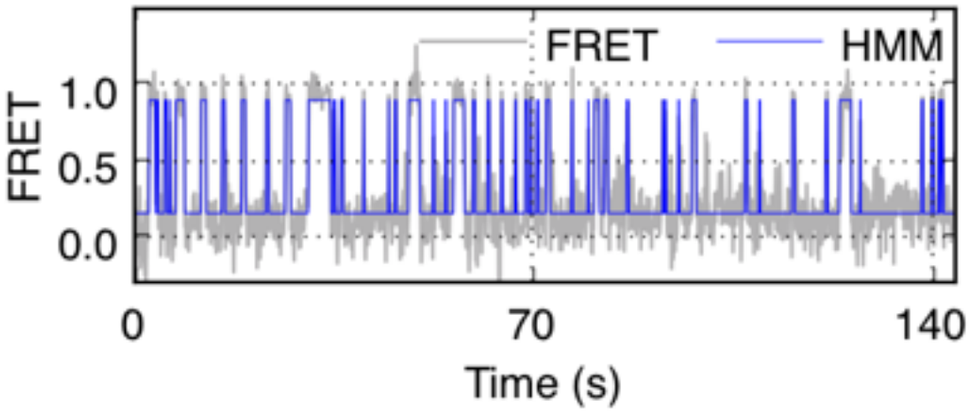


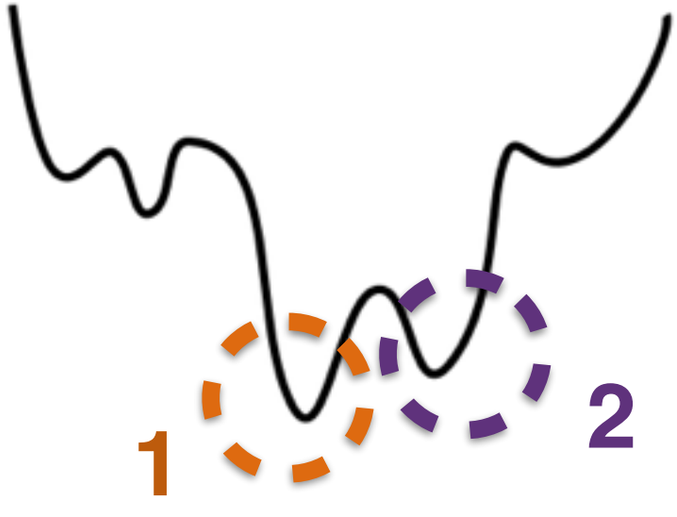


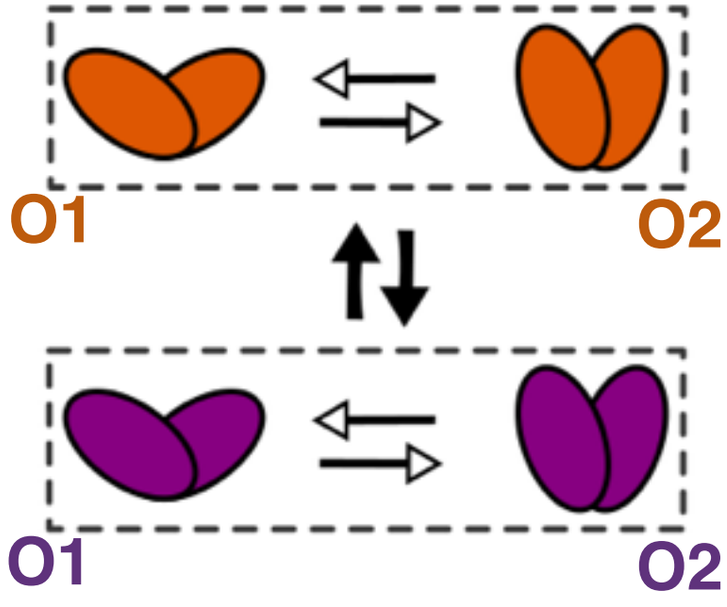


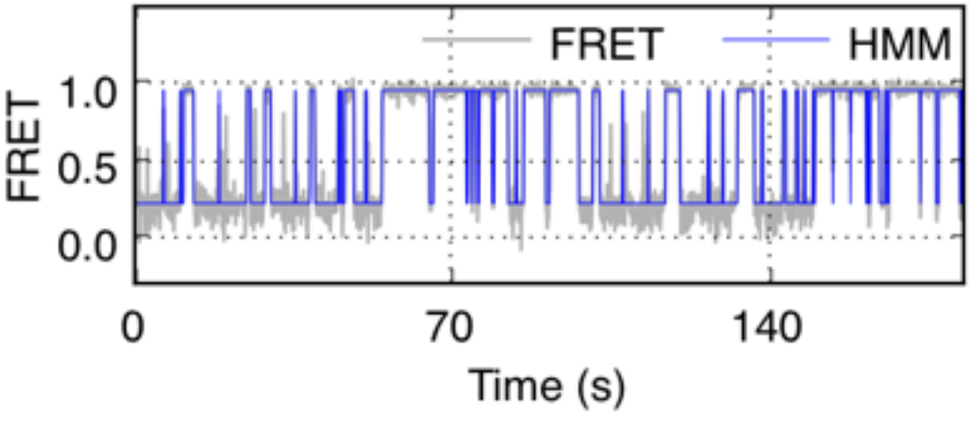


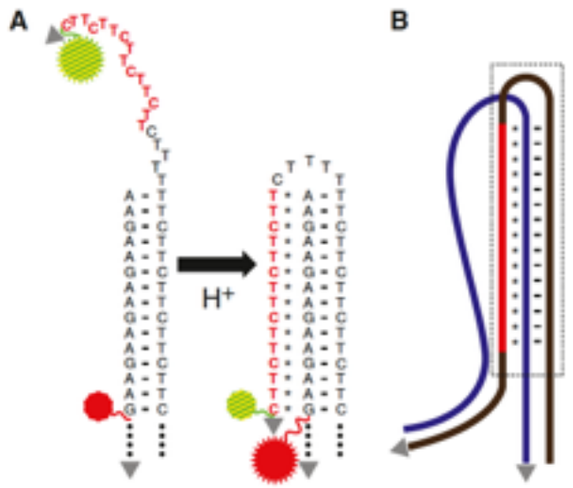




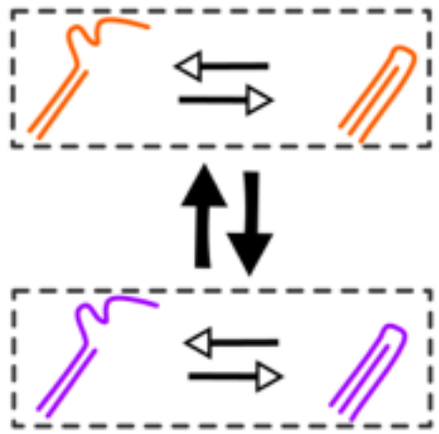




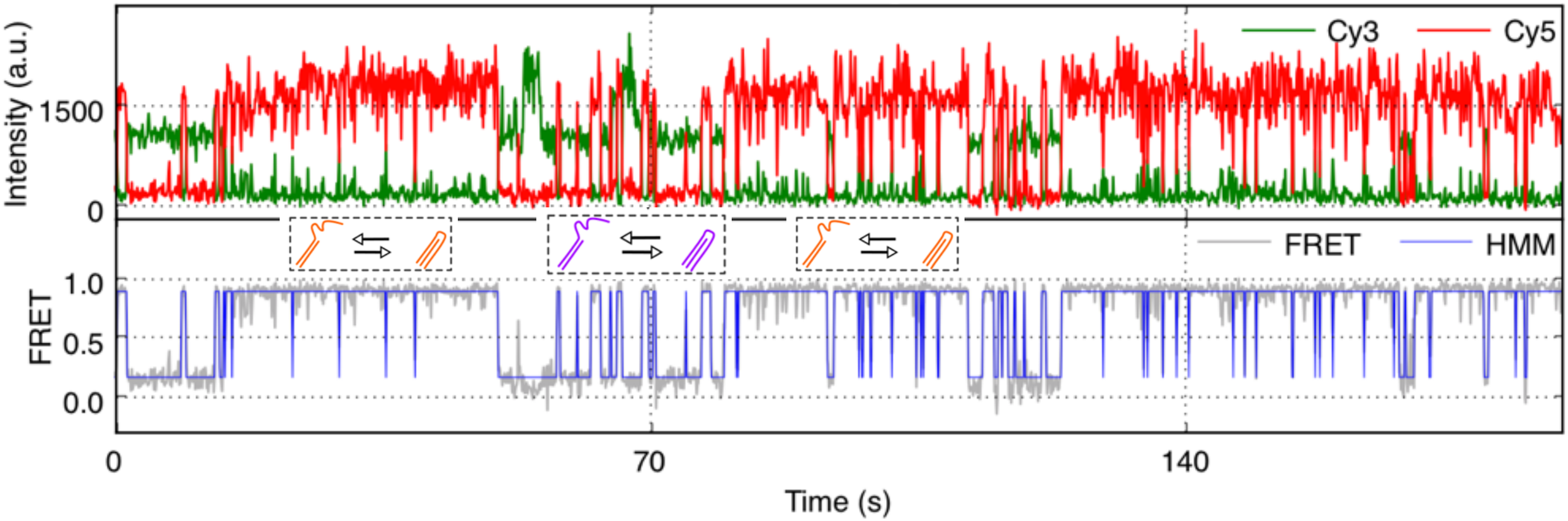


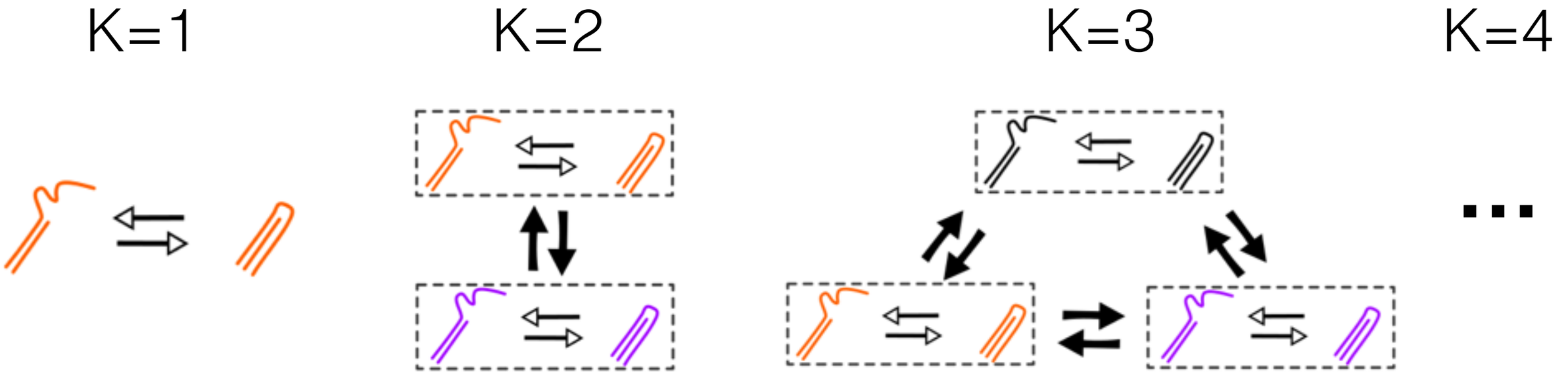


Lee et al, Biophysical J (2012)



Data from Korea Univ. (Prof. Seok-Cheol Hong's lab.)



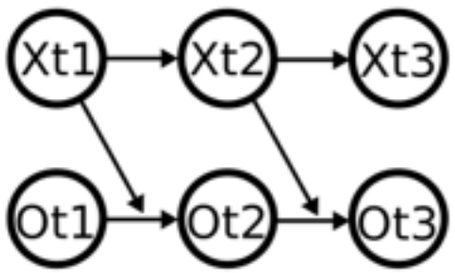


K=? boundaries? rates?

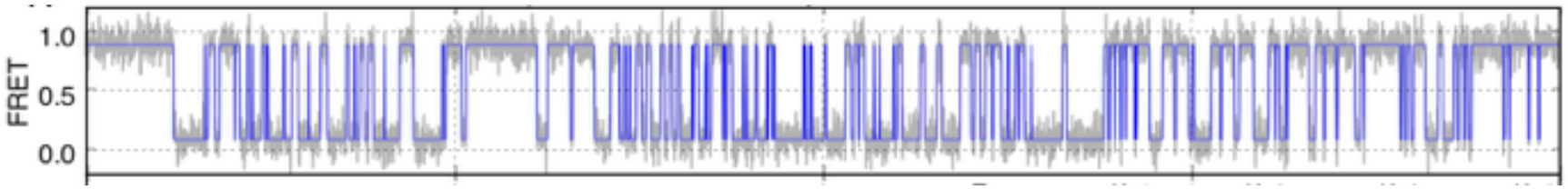
Three theoretical frameworks:

1. Double Chain Markov Model
2. Maximum evidence
3. Variational Bayes

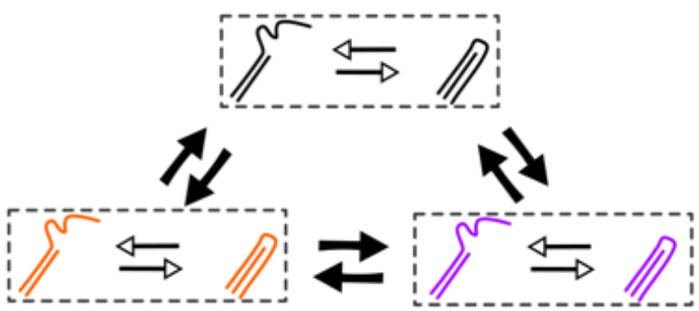
Double Chain Markov Model (Berchtold, 1999)

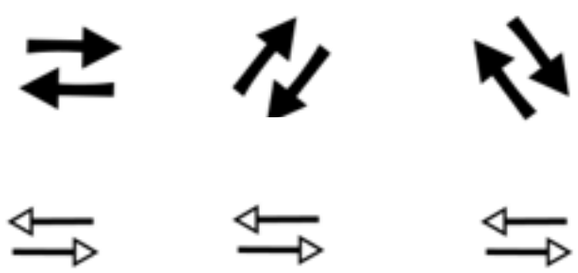


$\boldsymbol{o} =$



$\boldsymbol{x} =$



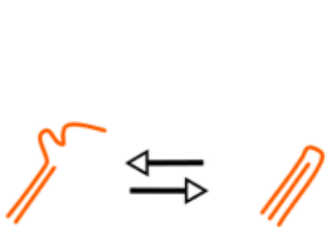
$\boldsymbol{A} =$ 

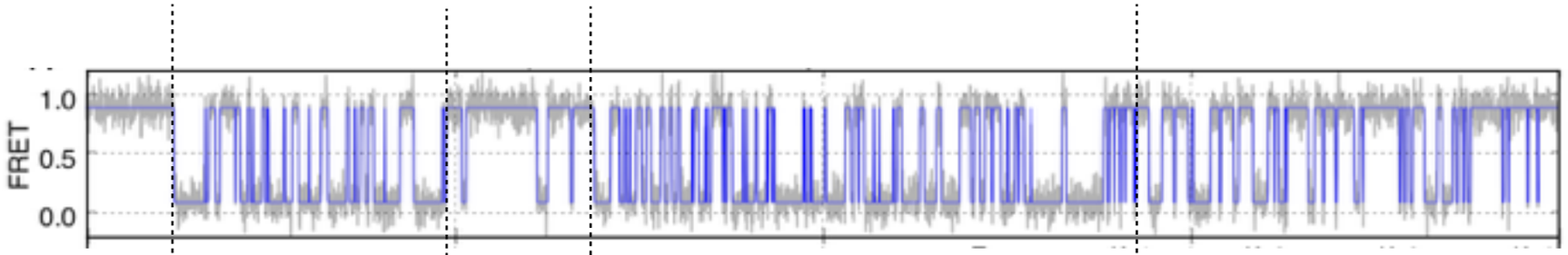
$\boldsymbol{B} =$

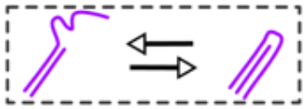
$$P(\boldsymbol{o}\,\boldsymbol{x} \mid \boldsymbol{A}\,\boldsymbol{B}\,\boldsymbol{\pi}) = \pi_{x_1} (\boldsymbol{B}^{x_1})_{o_1 \to o_2} \boldsymbol{A}_{x_1 \to x_2} (\boldsymbol{B}^{x_2})_{o_2 \to o_3} \boldsymbol{A}_{x_2 \to x_3} \cdots$$

$$P(\boldsymbol{o} \mid \boldsymbol{A}^*\ \boldsymbol{B}^*\ \boldsymbol{\pi}^*) \qquad P(\boldsymbol{x}^* \mid \boldsymbol{o}\,\boldsymbol{A}^*\ \boldsymbol{B}^*\ \boldsymbol{\pi}^*)$$

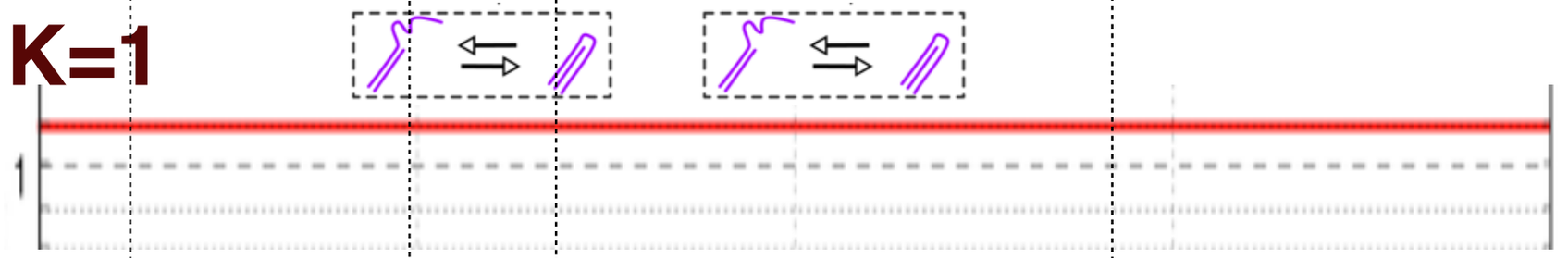
Double Chain Markov Model (Berchtold, 1999)

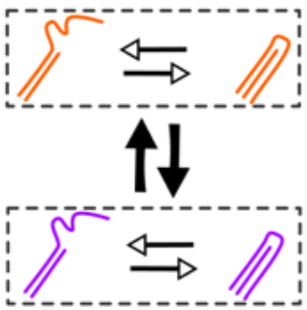




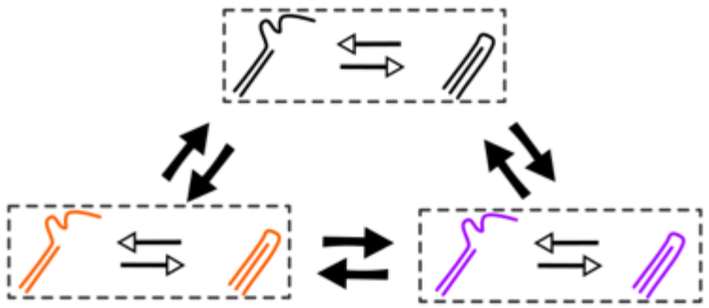


K=1

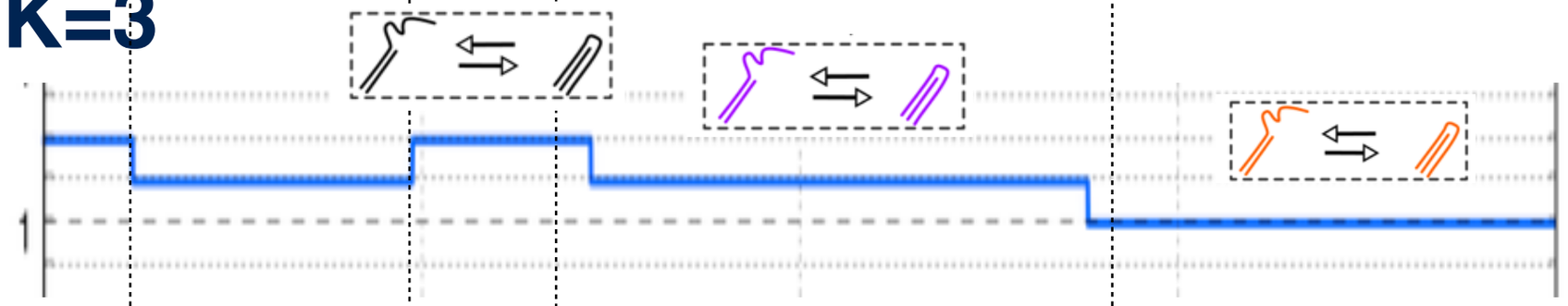




K=2



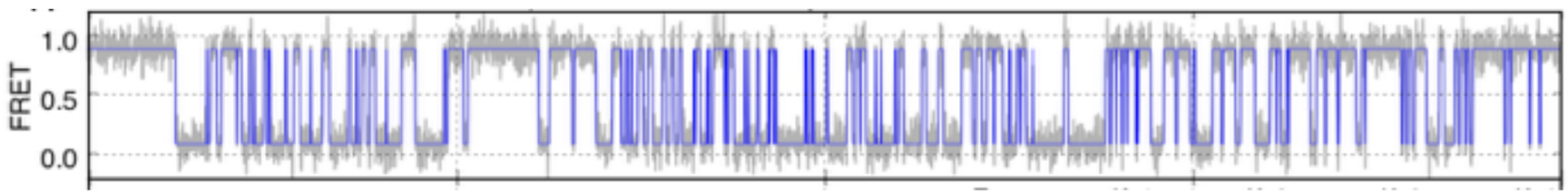
K=3



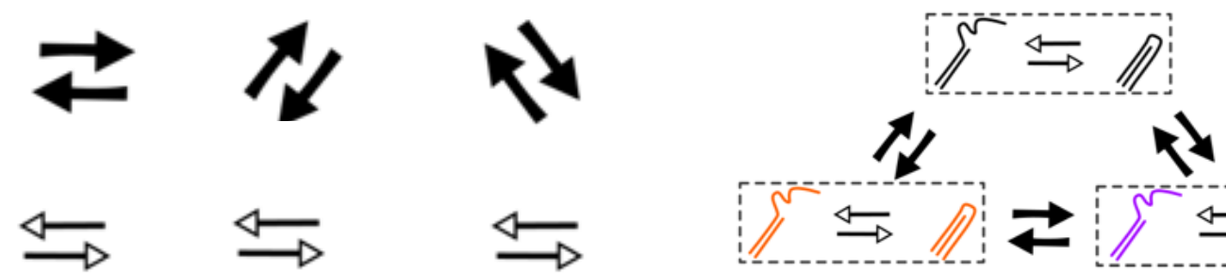
Maximum likelihood

$$P(o \mid A^* \; B^* \; \pi^* \;)$$

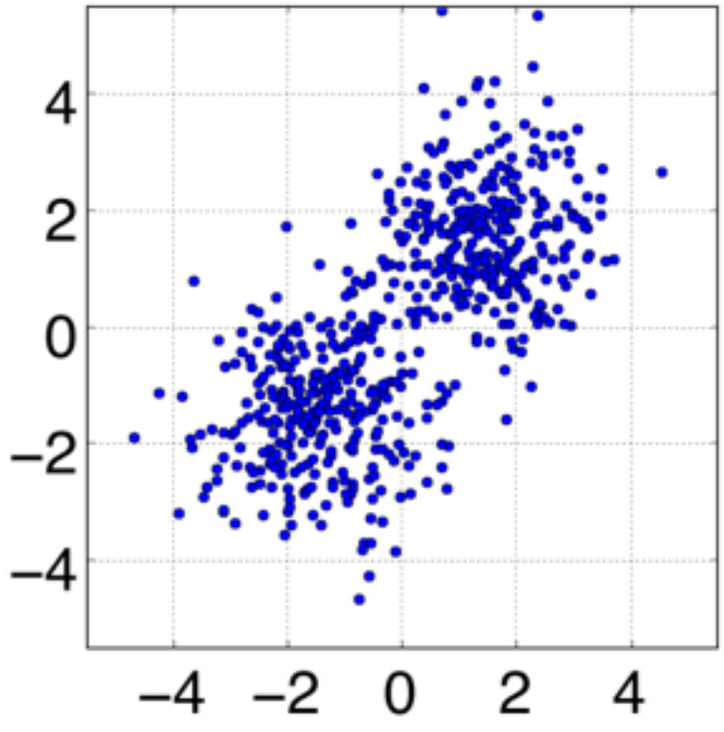
$o =$



$A =$



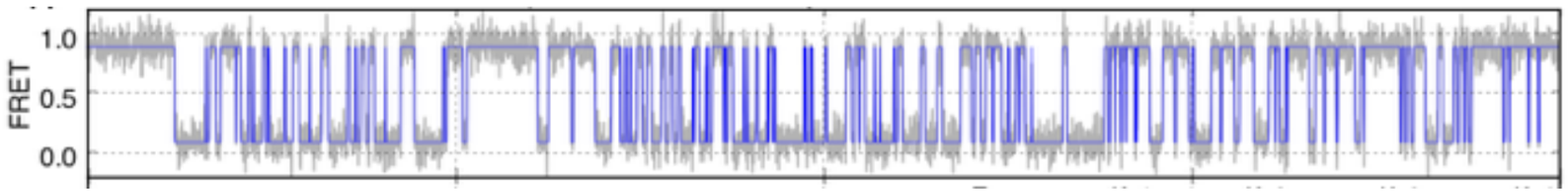
$B =$



Maximum evidence

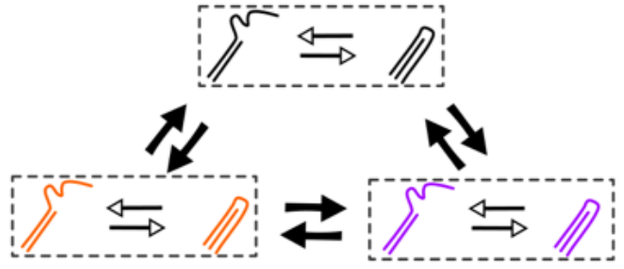
$$\int (\mathrm{d}\boldsymbol{A}\ \mathrm{d}\boldsymbol{B}\ \mathrm{d}\boldsymbol{\pi})\ \mathrm{P}(\boldsymbol{o} \mid \boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{\pi})\ \mathrm{P}(\boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{\pi} \mid K)$$

$\boldsymbol{o} =$



$\boldsymbol{A} =$

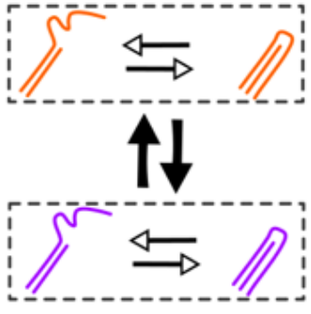
$\boldsymbol{B} =$



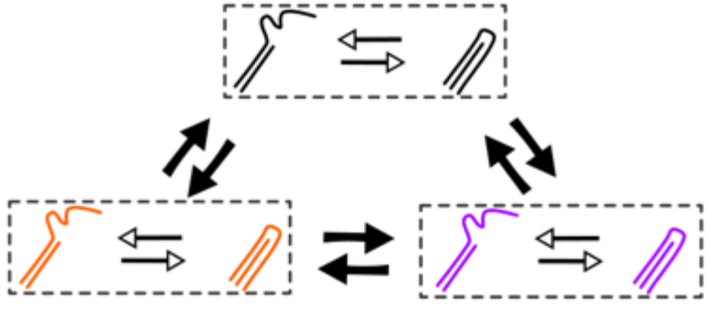
$\boldsymbol{K}$:

K=1

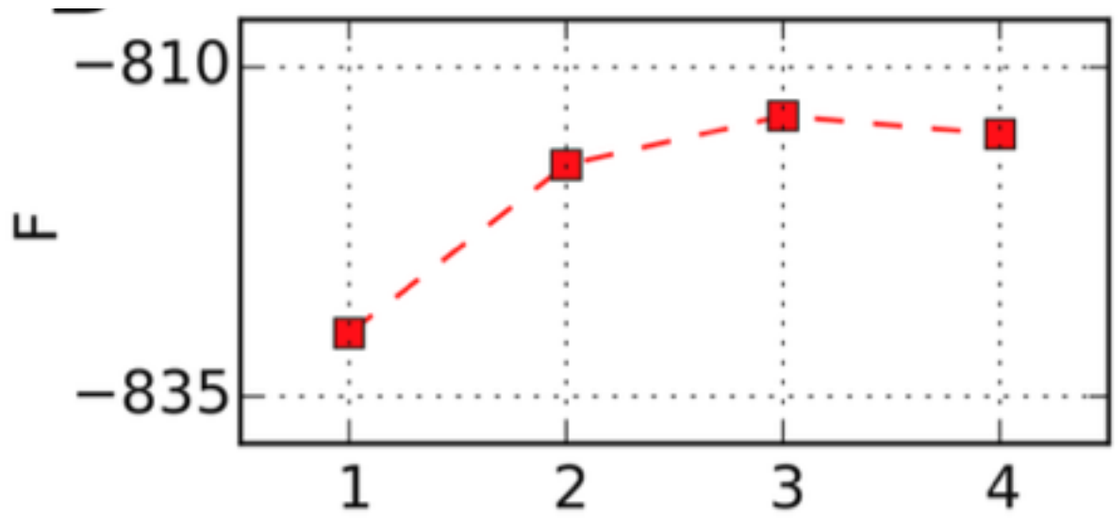
K=2



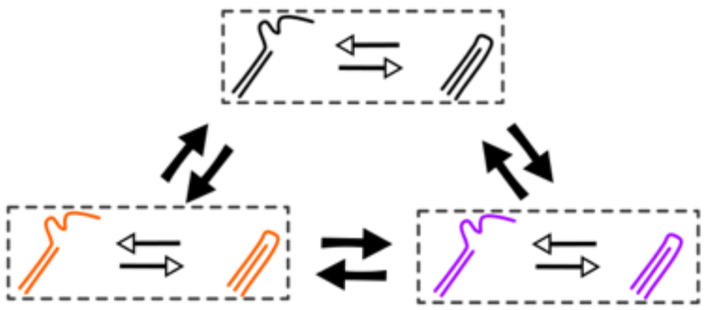
K=3

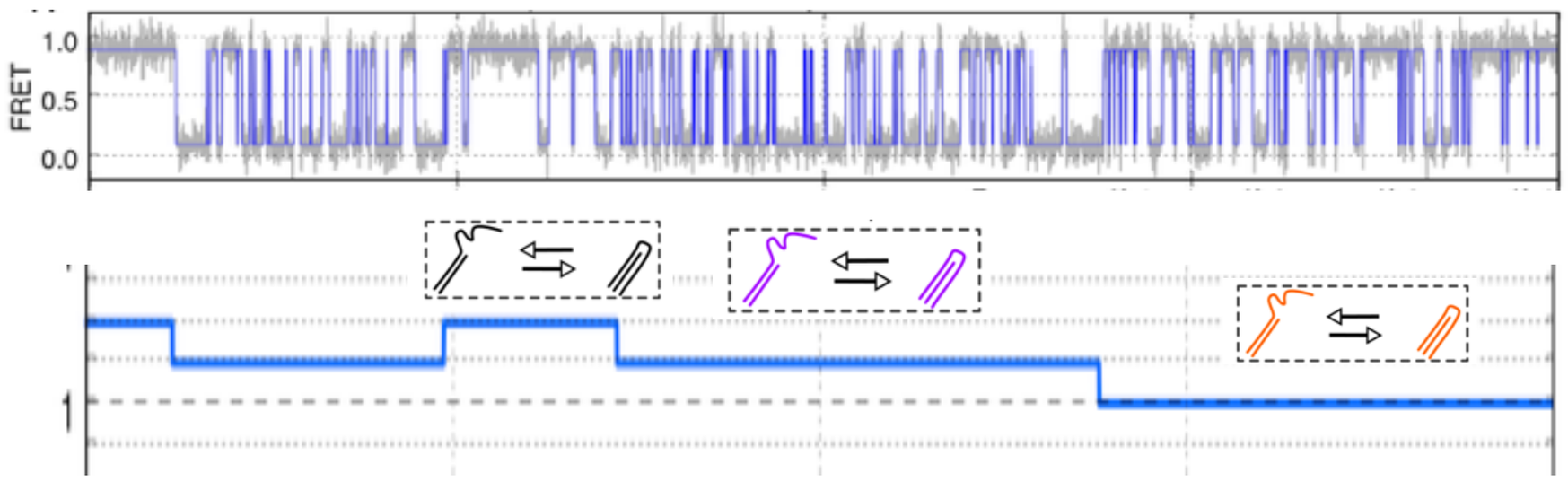


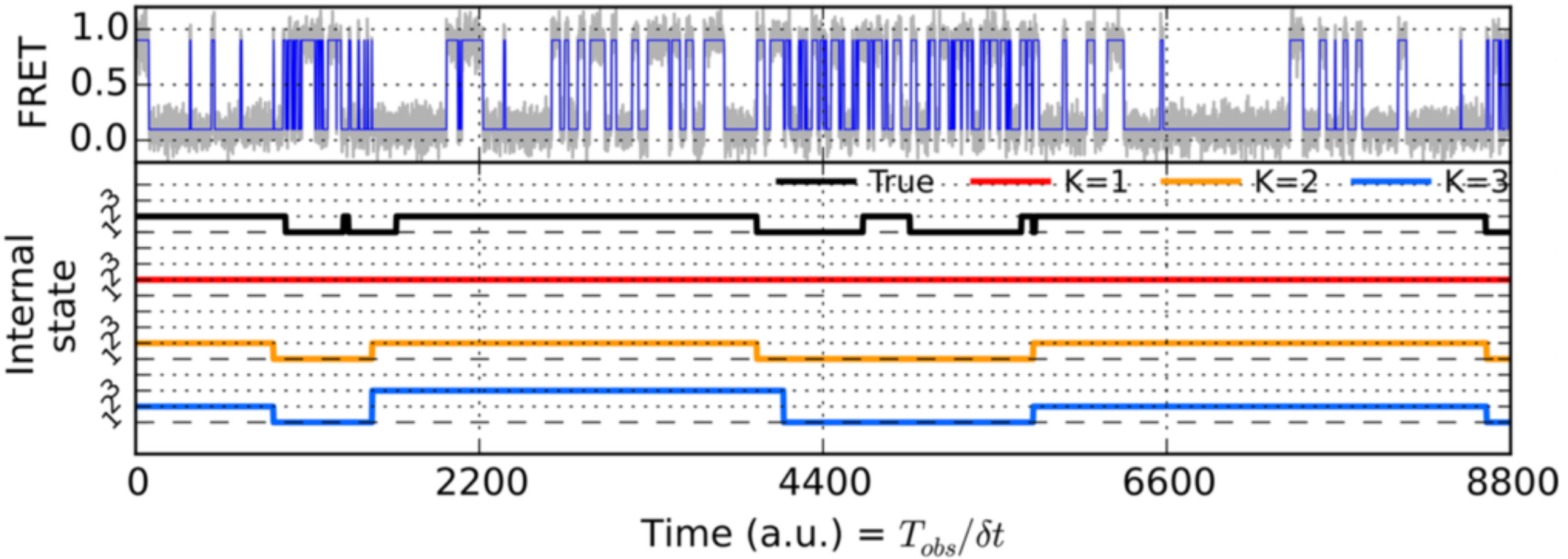
$$\int (\mathrm{d}\boldsymbol{A}\ \mathrm{d}\boldsymbol{B}\ \mathrm{d}\boldsymbol{\pi})\ \ \mathrm{P}(\boldsymbol{o} \mid \boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{\pi})\ \ \mathrm{P}(\boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{\pi} \mid K)$$

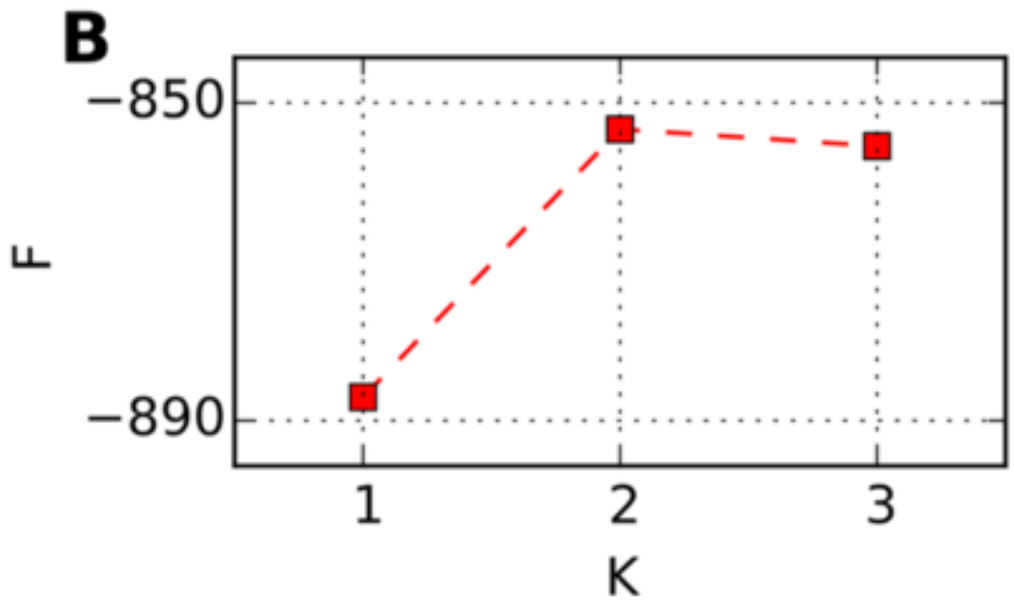


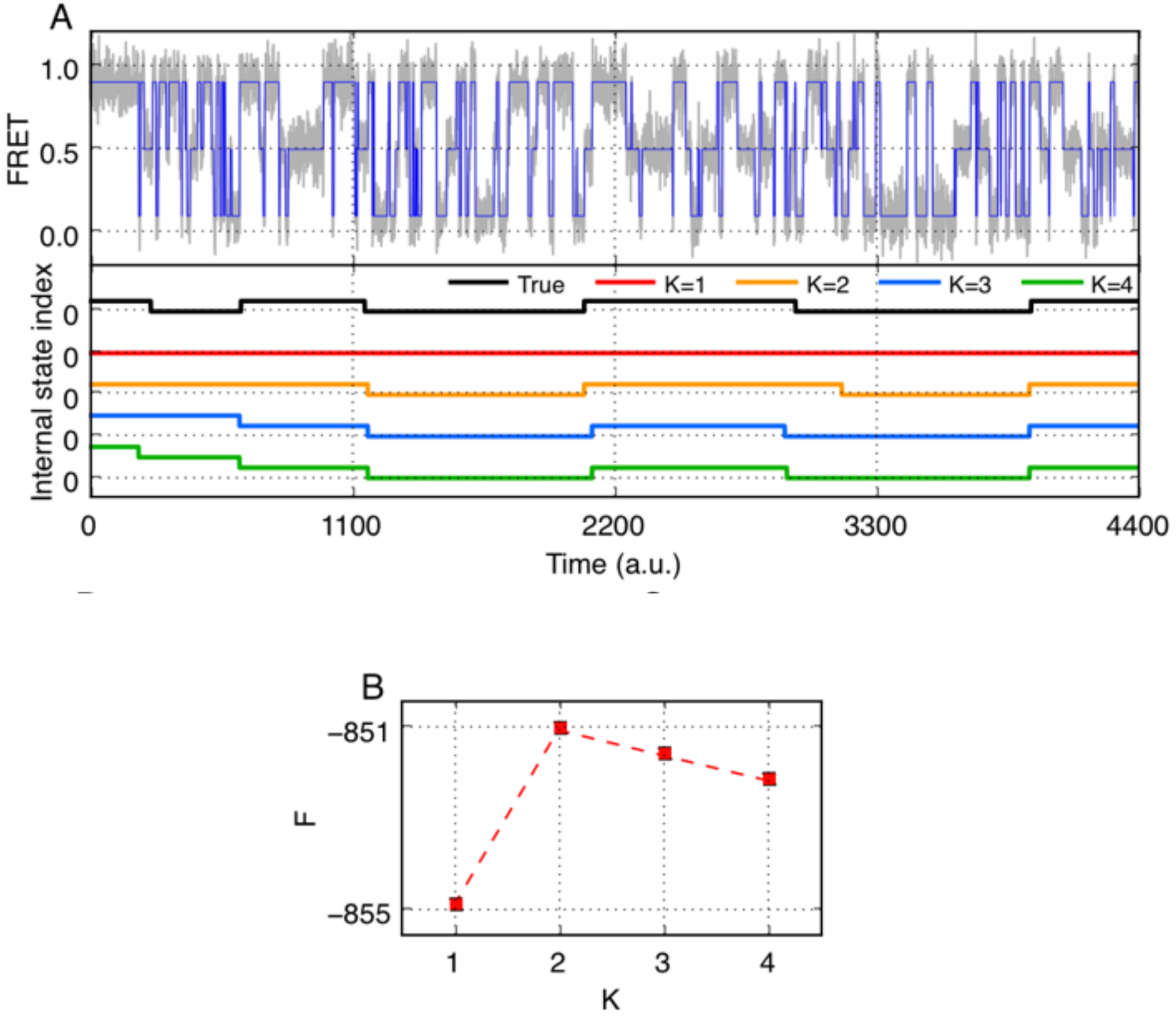
K=3






FRET
1.0
0.5
0.0
True
K=1
K=2
K=3
Internal state
0
2200
4400
6600
8800
Time (a.u.) = $T_{obs}/\delta t$


B
−850
−890
F
1
2
3
K


A
FRET
1.0
0.5
0.0
Internal state index
True
K=1
K=2
K=3
K=4
0
0
0
0
0
0
1100
2200
3300
4400
Time (a.u.)
B
−851
−855
F
1
2
3
4
K

Variational Bayes

$$\int (\mathrm{d}\boldsymbol{A}\ \mathrm{d}\boldsymbol{B}\ \mathrm{d}\boldsymbol{\pi})\ \ \mathrm{P}(\boldsymbol{o} \mid \boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{\pi})\ \ \mathrm{P}(\boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{\pi} \mid K)$$

$$\begin{aligned}
\log(P(\boldsymbol{o}|\boldsymbol{K})) &= \int q(\boldsymbol{Z})\ \log(P(\boldsymbol{o}|\boldsymbol{K}))d\boldsymbol{Z} \\
&= \int q(\boldsymbol{Z})\ \log\left(P(\boldsymbol{o}|\boldsymbol{K})\frac{P(\boldsymbol{o},\boldsymbol{Z}|\mathbf{K})}{q(\boldsymbol{Z})}\frac{q(\boldsymbol{Z})}{P(\boldsymbol{o},\boldsymbol{Z}|\mathbf{K})}\right)d\boldsymbol{Z} \\
&= \int q(\boldsymbol{Z})\ \log\left(\frac{P(\boldsymbol{o},\boldsymbol{Z}|\boldsymbol{K})}{q(\boldsymbol{Z})}\right)d\boldsymbol{Z} + \int q(\boldsymbol{Z})\ \log\left(\frac{q(\boldsymbol{Z})}{P(\boldsymbol{Z}|\mathbf{o},\mathbf{K})}\right)d\boldsymbol{Z} \\
&= F[q] + KL(q||p)
\end{aligned}$$

$$\log(P(\boldsymbol{o}|\boldsymbol{K})) \geq F[q] + KL(q||p) \geq F[q]$$

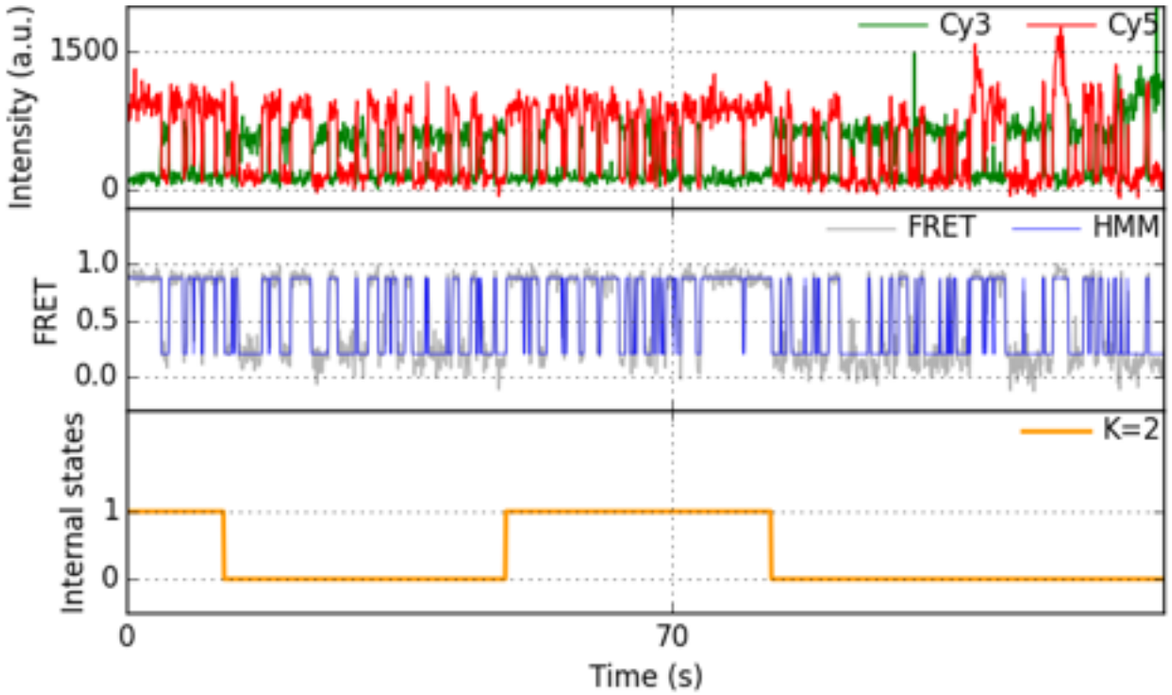
$$\begin{aligned}
P(\boldsymbol{o},\boldsymbol{Z}|\boldsymbol{K}) &= P(\boldsymbol{o},\boldsymbol{x},\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B}|\boldsymbol{K}) \\
&= P(\boldsymbol{o},\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B}|\boldsymbol{K}) \\
&= P(\boldsymbol{o},\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{\pi}|\boldsymbol{K})P(\boldsymbol{A}|\boldsymbol{K})P(\boldsymbol{B}|\boldsymbol{K}
\end{aligned}$$

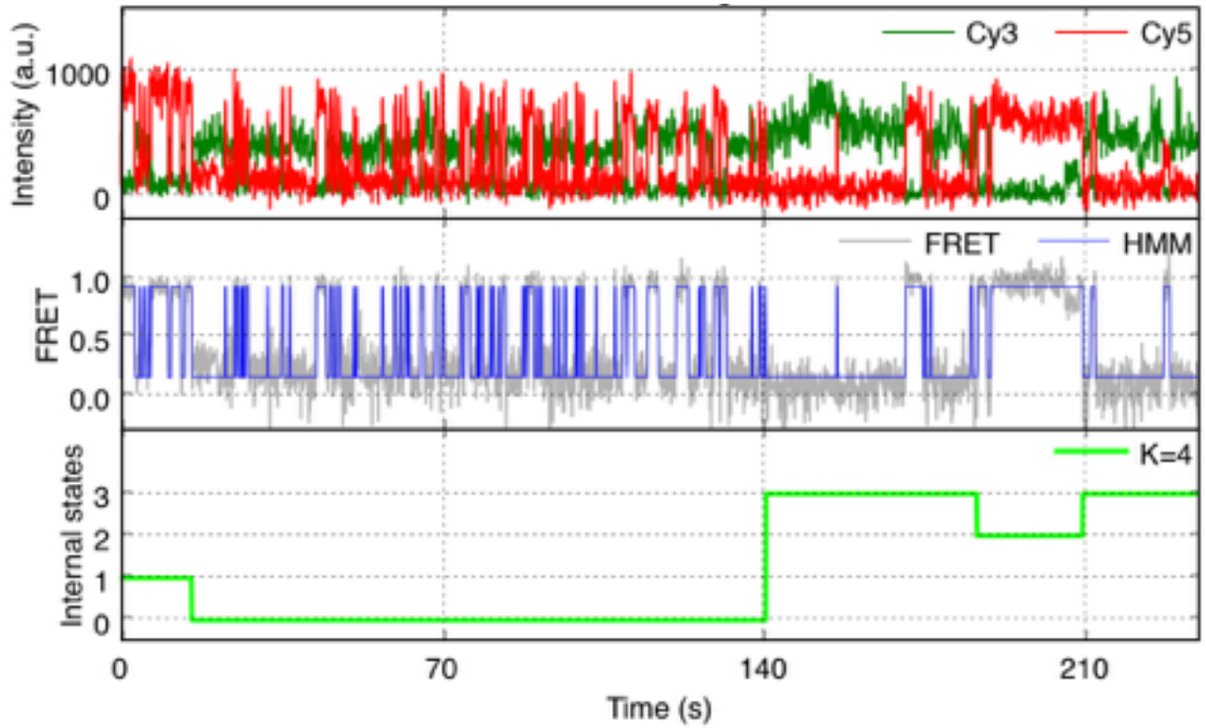
$$\begin{aligned}
P(\boldsymbol{\pi}|\boldsymbol{K}) &= Dir(\pi_1,\pi_2,...,\pi_K|u_1^{\pi},u_2^{\pi},...,u_K^{\pi}) \\
&= \frac{\Gamma(u_0^{pi})}{\prod_{k=1}^{K}\Gamma(u_k^{\pi})}\prod_{k=1}^{K}\pi_k^{u_k^{\pi}-1} \\
&\text{where } u_0^{\pi} = \sum_{k=1}^{K} u_k^{\pi} \text{ and } \sum_{k=1}^{K}\pi_k = 1
\end{aligned}$$

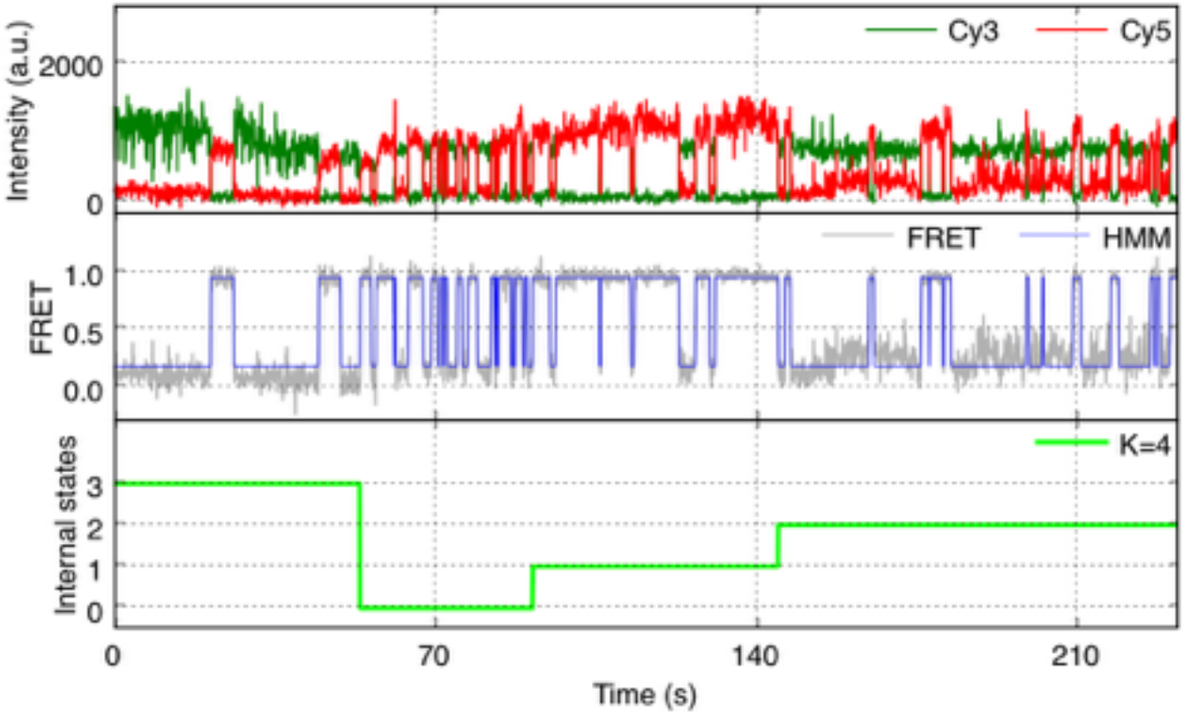
$$\begin{aligned}
P(\boldsymbol{A}|\boldsymbol{K}) &= \prod_{k=1}^{K} Dir(A_{k,1},A_{k,2},...,A_{k,K}|u_{k,1}^{A},u_{k,2}^{A},...,u_{k,K}^{A}) \\
&= \prod_{k=1}^{K}\frac{\Gamma(u_{k,0}^{A})}{\prod_{m=1}^{K}\Gamma(u_{k,m}^{A})}\prod_{m=1}^{K}A_{k,m}^{u_{k,m}^{A}-1} \\
&\text{where } u_{k,0}^{A} = \sum_{m=1}^{K} u_{k,m}^{A} \text{ and } \sum_{m=1}^{K} A_{k,m} = 1
\end{aligned}$$

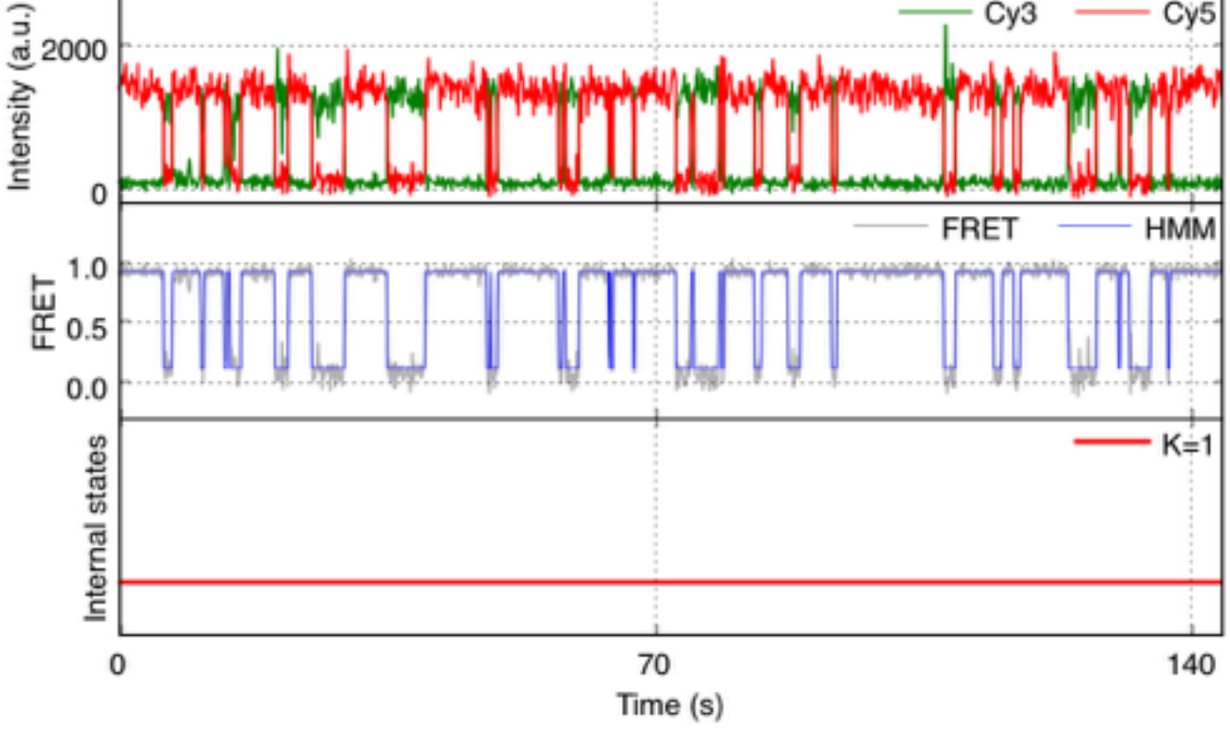
$$\begin{aligned}
P(\boldsymbol{B}|\boldsymbol{K}) &= \prod_{k=1}^{K}\prod_{i=1}^{L} Dir(B_{k,i,1},B_{k,i,2},...,B_{k,i,L}|u_{k,i,1}^{B},u_{k,i,2}^{B},...,u_{k,i,L}^{B}) \\
&= \prod_{k=1}^{K}\prod_{i=1}^{L}\frac{\Gamma(u_{k,i,0}^{B})}{\prod_{j=1}^{L}\Gamma(u_{k,i,j}^{B})}\prod_{j=1}^{L}B_{k,i,j}^{u_{k,i,j}^{B}-1} \\
&\text{where } u_{k,i,0}^{B} = \sum_{j=1}^{L} u_{k,i,j}^{B} \text{ and } \sum_{j=1}^{L} B_{k,i,j} = 1
\end{aligned}$$

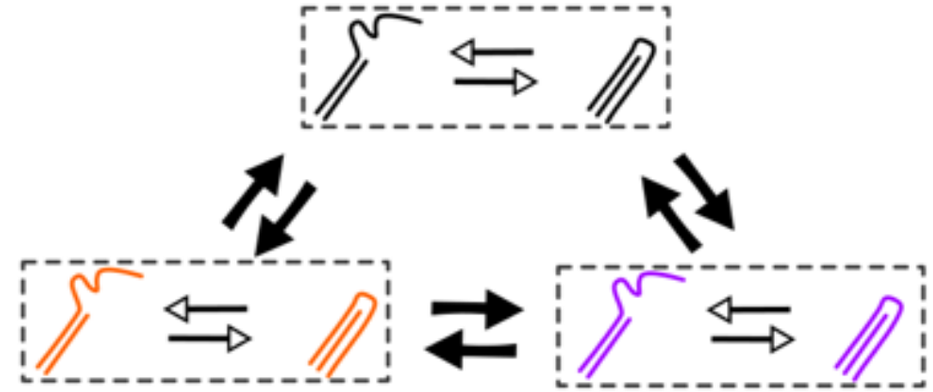
. . .

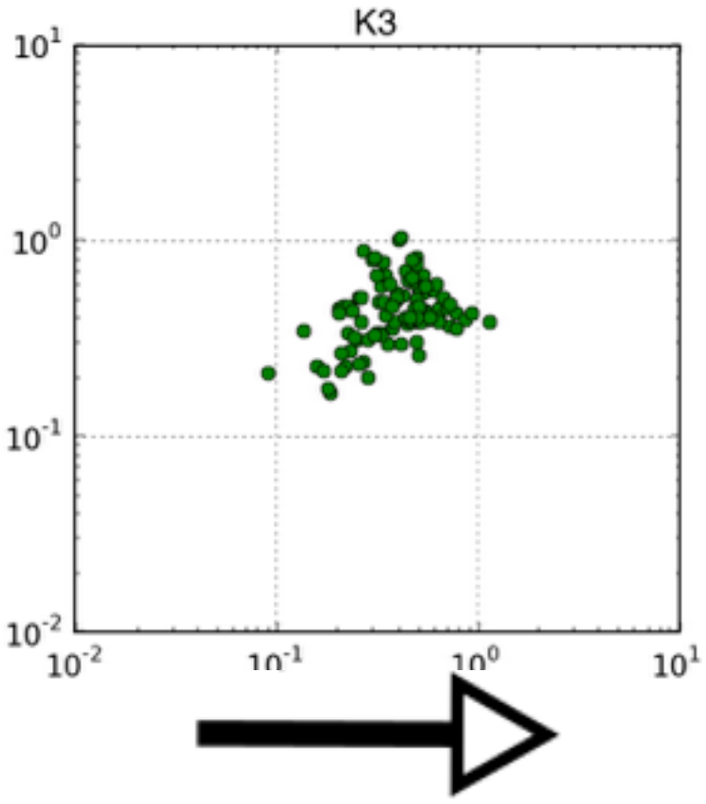

Intensity (a.u.)
1500
0
Cy3
Cy5
FRET
1.0
0.5
0.0
HMM
Internal states
1
0
K=2
0
70
Time (s)

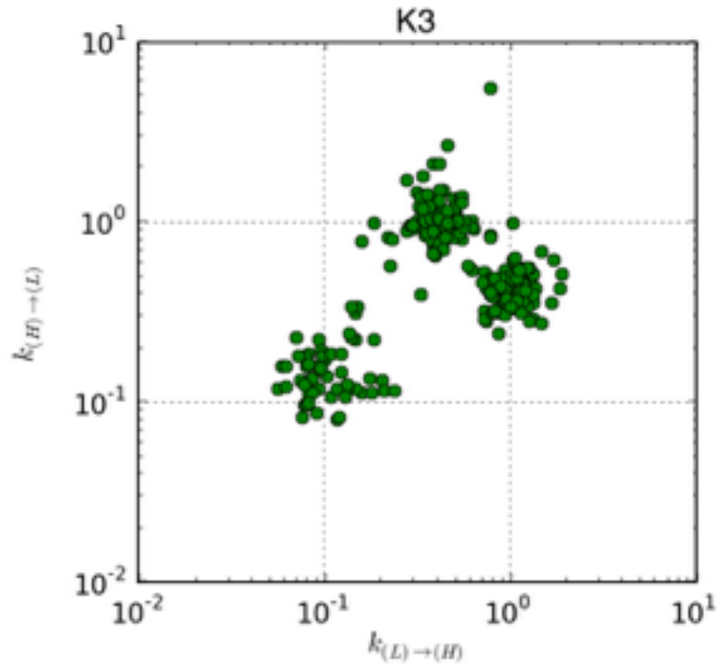

Intensity (a.u.)
1000
0
Cy3
Cy5
FRET
1.0
0.5
0.0
HMM
Internal states
3
2
1
0
K=4
0
70
140
210
Time (s)

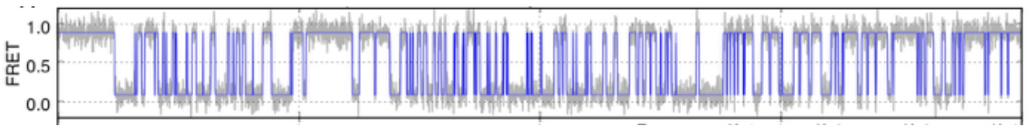

Intensity (a.u.)
2000
0
Cy3
Cy5
FRET
1.0
0.5
0.0
HMM
Internal states
3
2
1
0
K=4
0
70
140
210
Time (s)

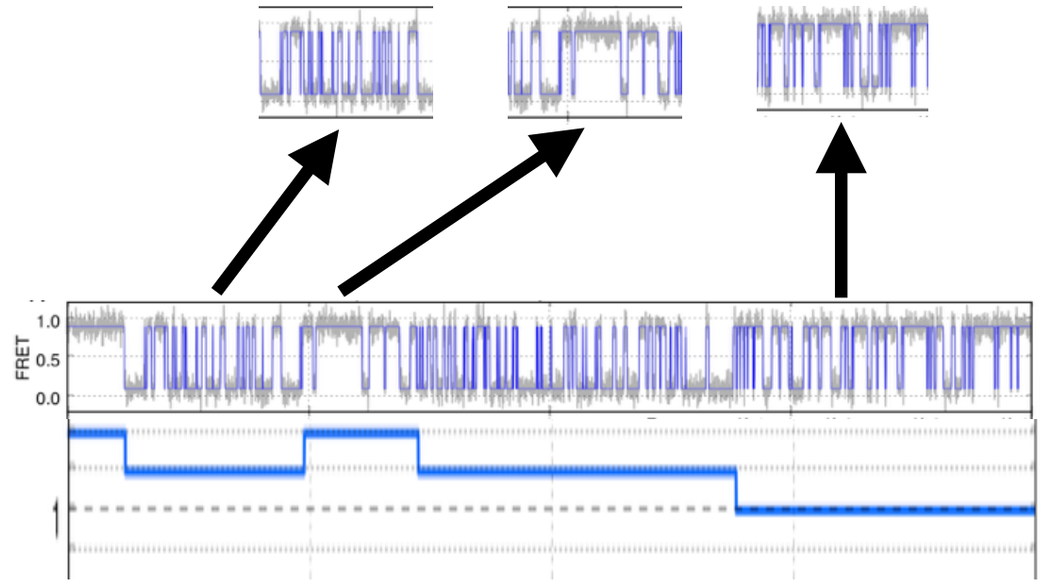

Intensity (a.u.)
2000
0
Cy3
Cy5
FRET
1.0
0.5
0.0
HMM
Internal states
K=1
0
70
140
Time (s)



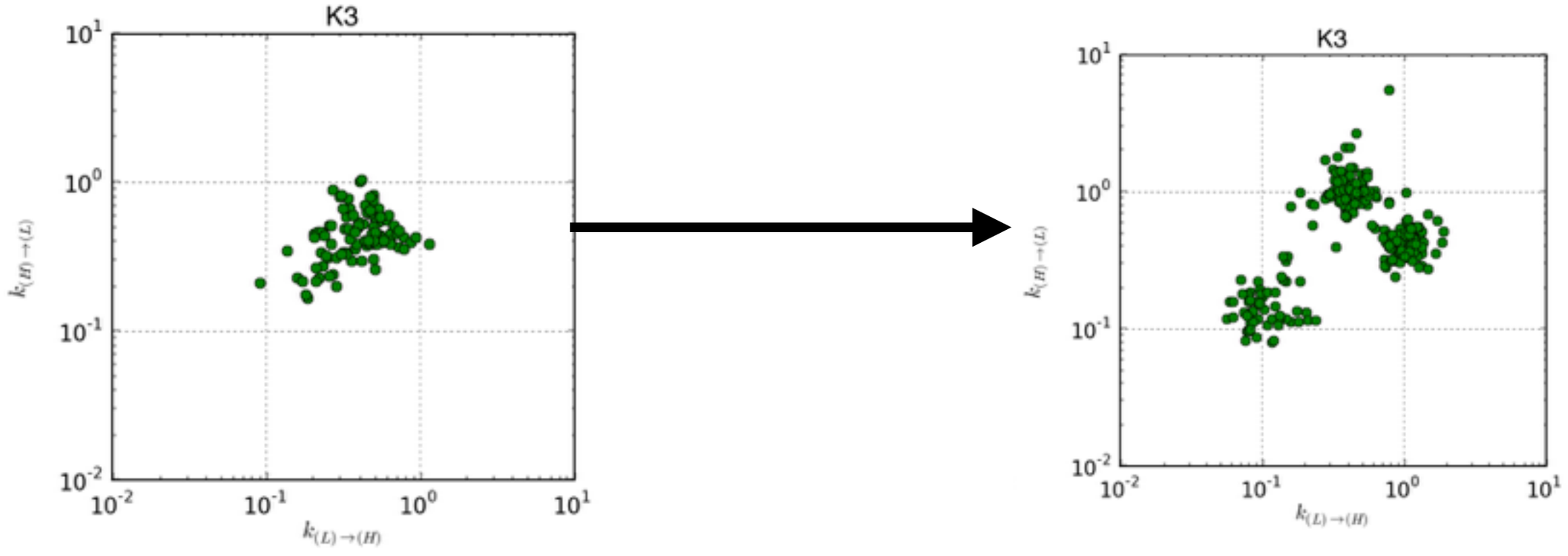

K3
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-2}
10^{-1}
10^{0}
10^{1}


K3
10^{1}
10^{0}
10^{-1}
10^{-2}
$k_{(H)\rightarrow(L)}$
10^{-2}
10^{-1}
10^{0}
10^{1}
$k_{(L)\rightarrow(H)}$

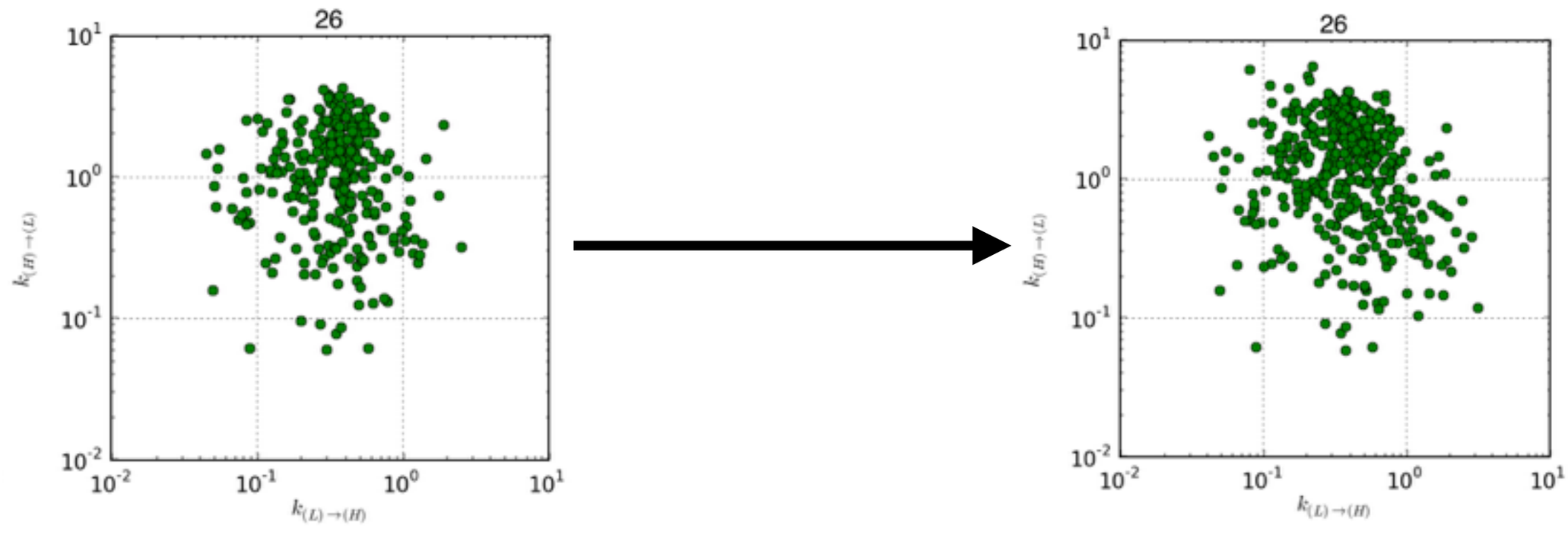

FRET
1.0
0.5
0.0

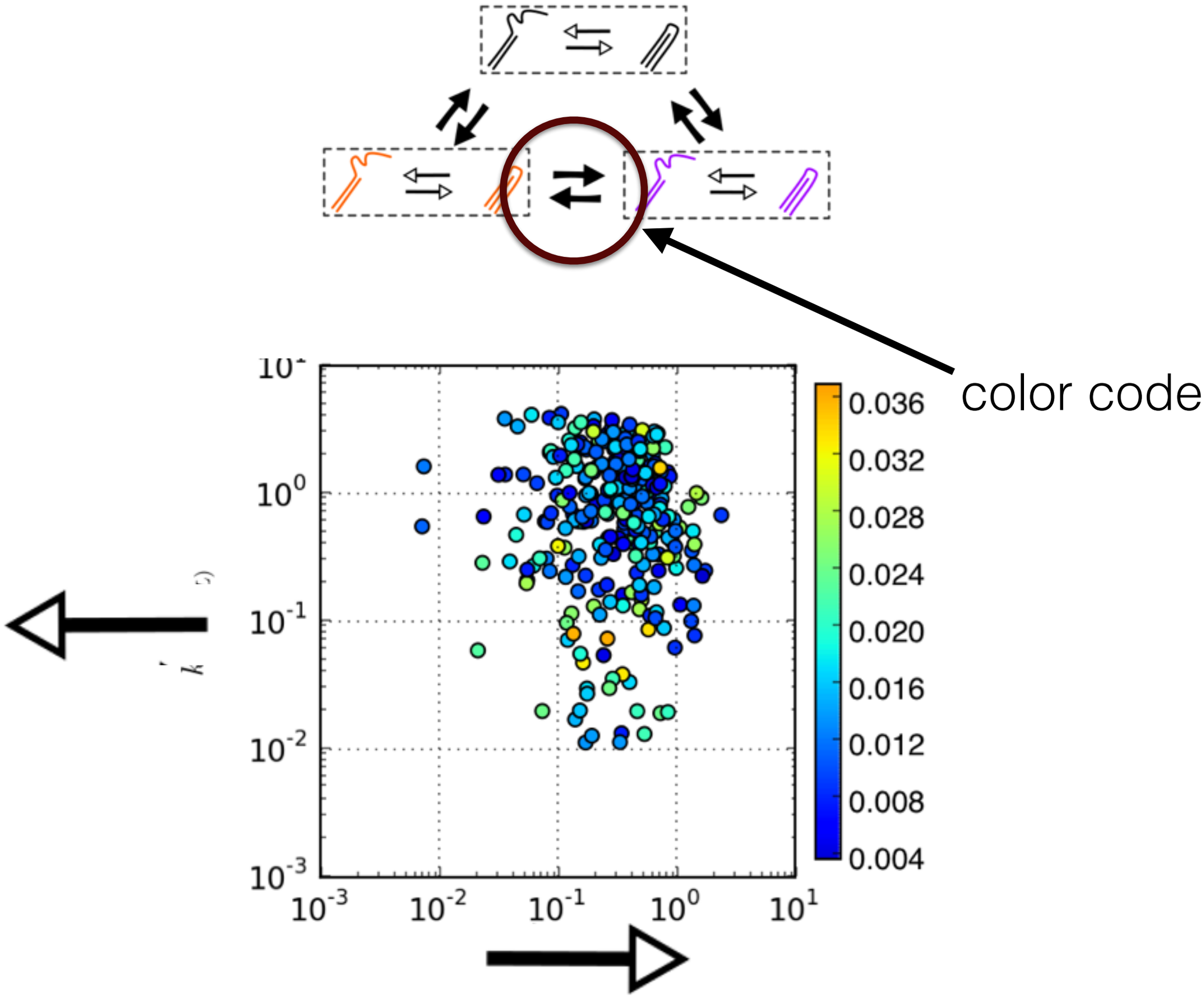

FRET
1.0
0.5
0.0

Synthetic data



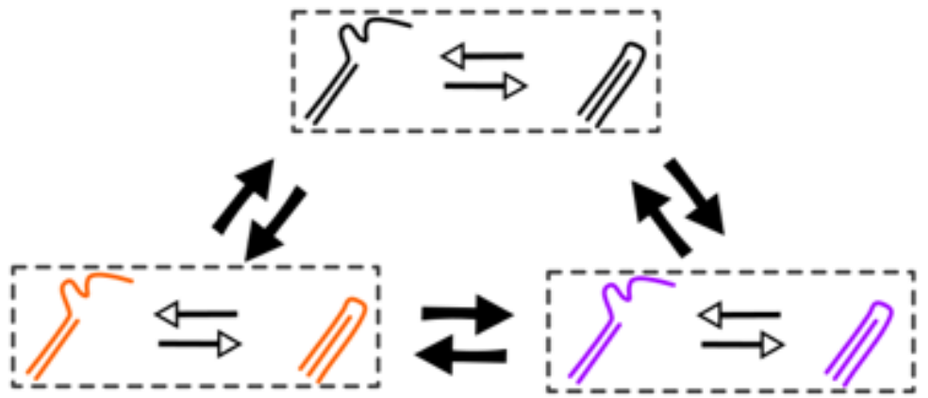
H-DNA




color code
0.036
0.032
0.028
0.024
0.020
0.016
0.012
0.008
0.004
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}

Summary

1. To analyse dynamic disorder, we have developed VB-DCMM
by combining
three theoretical frameworks:
DCMM, maximum evidence, and variational Bayes.



2. H-DNA have many basins of attractions.

3. More analysis on H-DNA is ongoing.

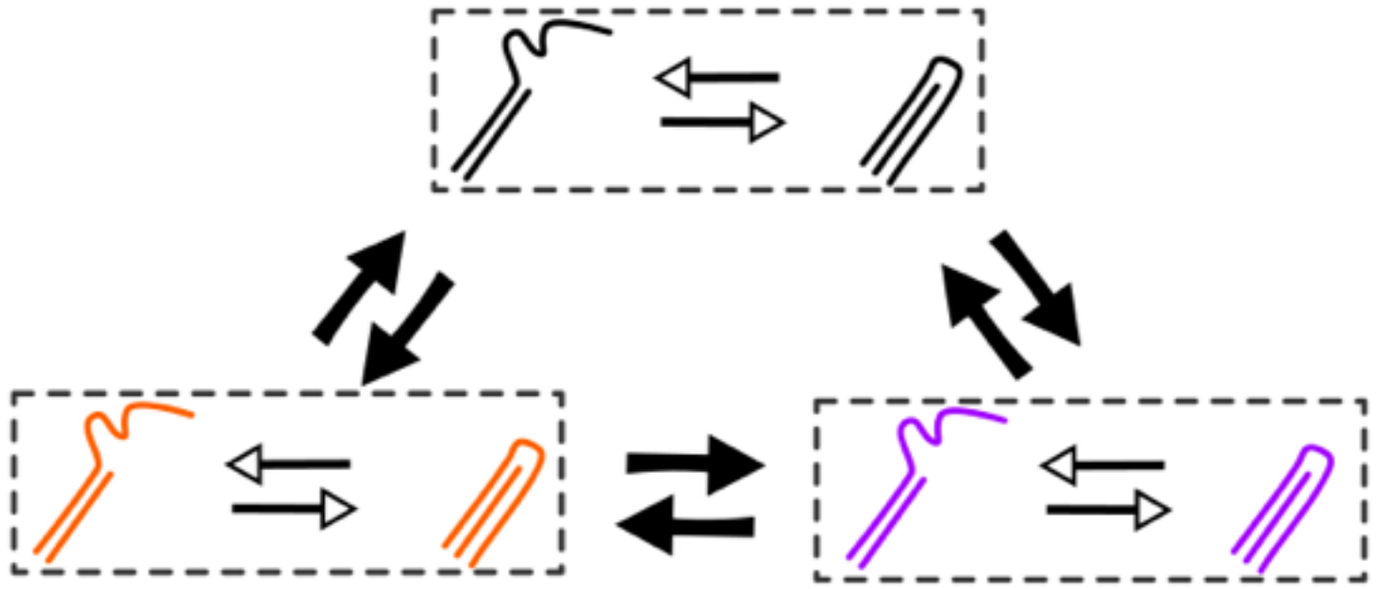
Acknowledgements

Prof. Changbong Hyeon

Prof. Seok-Cheol Hong

Il-Buem Lee (PhD candidate)
Moon-Hyeon Min (PhD candidate)

Thank you for the attention.

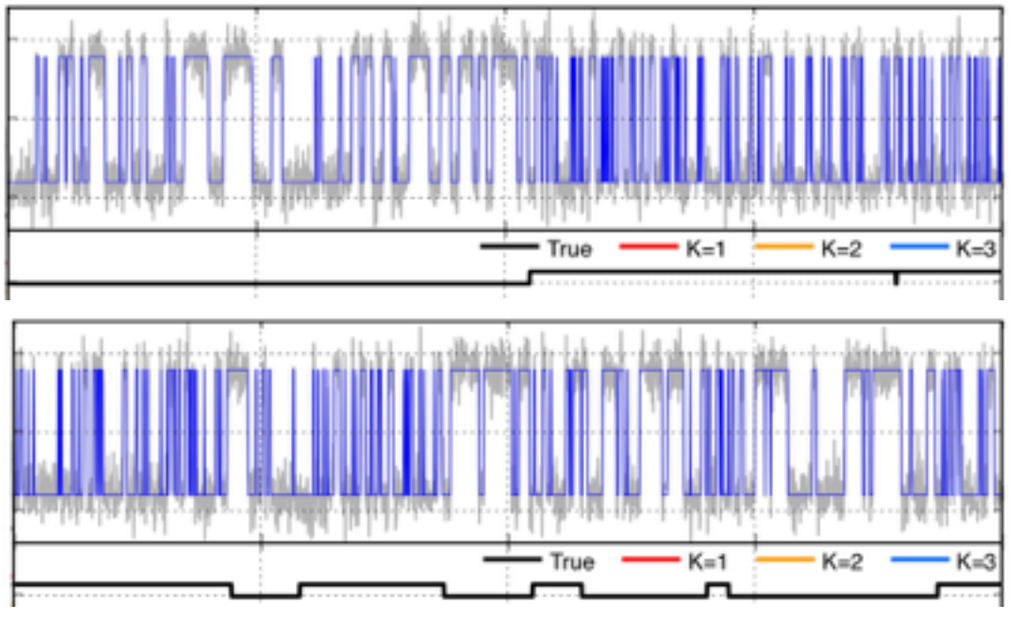


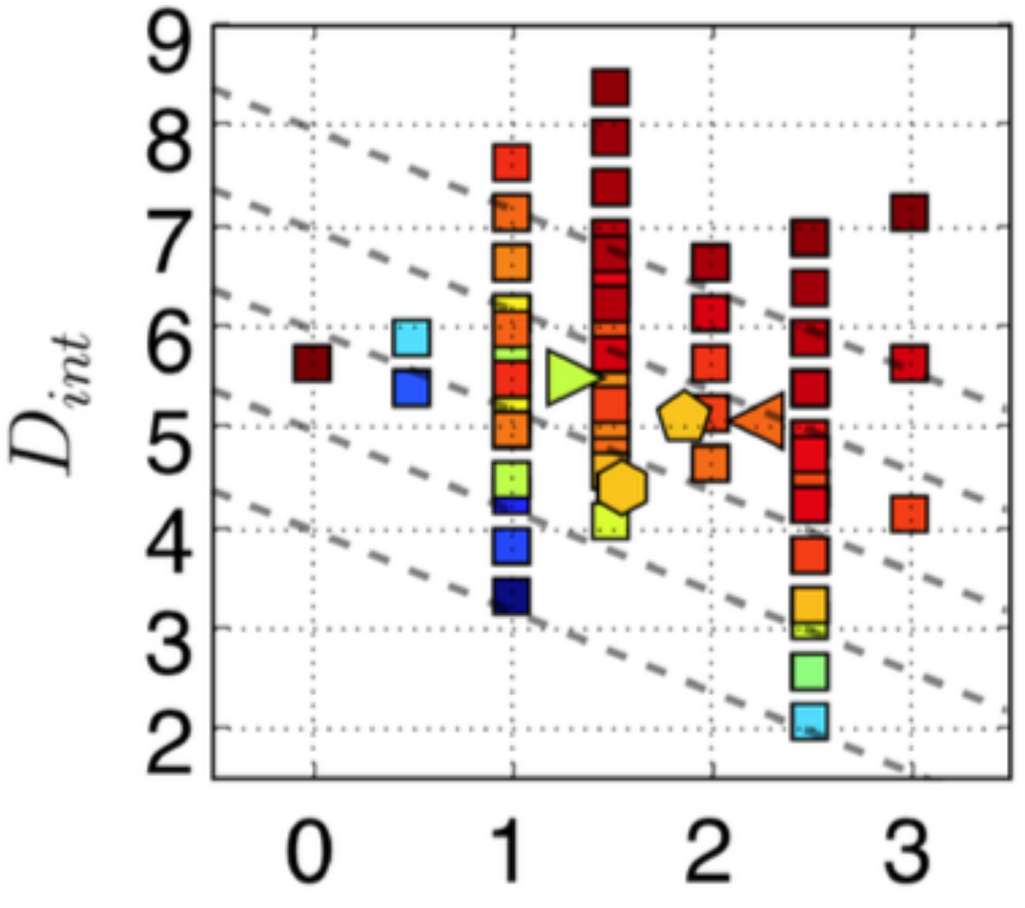
$$D_{conf} = \frac{2}{K(K-1)/2} \sum_{\substack{\mu,\nu \\ \mu \neq \nu}}^{K} \left(\frac{1}{L(L-1)} \sum_{\substack{a,b \\ a \neq b}}^{L} \left| \log_2 \frac{k_{a \to b}^{(\mu)}}{k_{a \to b}^{(\nu)}} \right| \right)$$

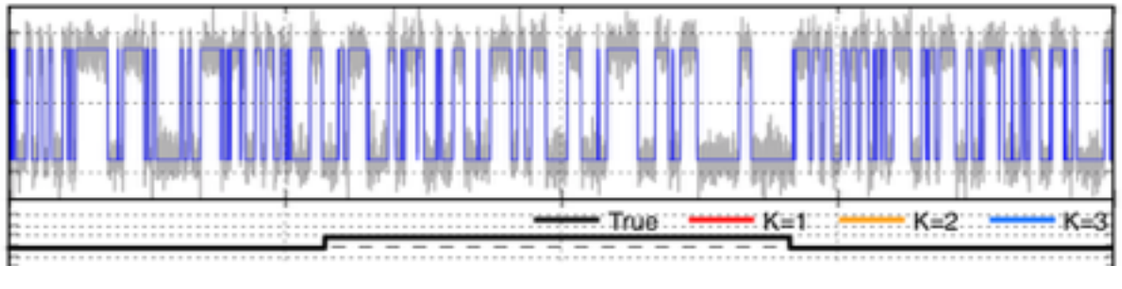
$$D_{int} = \frac{1}{K} \sum_{\mu}^{K} \left(\frac{1}{L(L-1)} \sum_{\substack{a,b \\ a \neq b}}^{L} \left| \log_2 \frac{k_{a \to b}^{(\mu)}}{\sum_{\mu' \neq \mu} \gamma^{(\mu) \to (\mu')}} \right| \right)$$

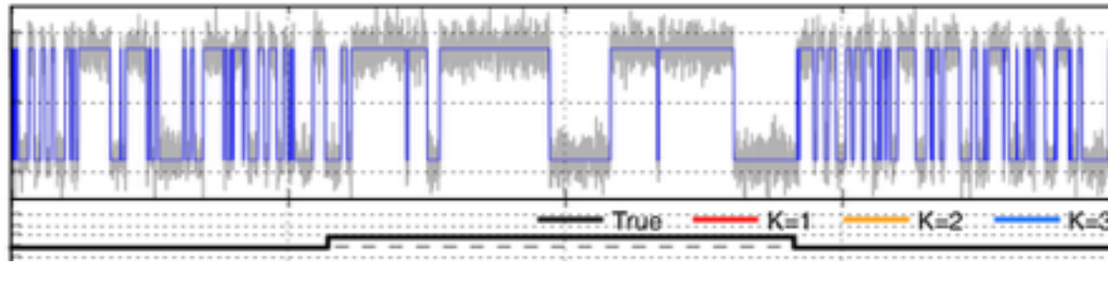
D_{conf}

D_{int}


True
K=1
K=2
K=3
True
K=1
K=2
K=3


D_{int}
9
8
7
6
5
4
3
2
0
1
2
3
D_{conf}


True
K=1
K=2
K=3

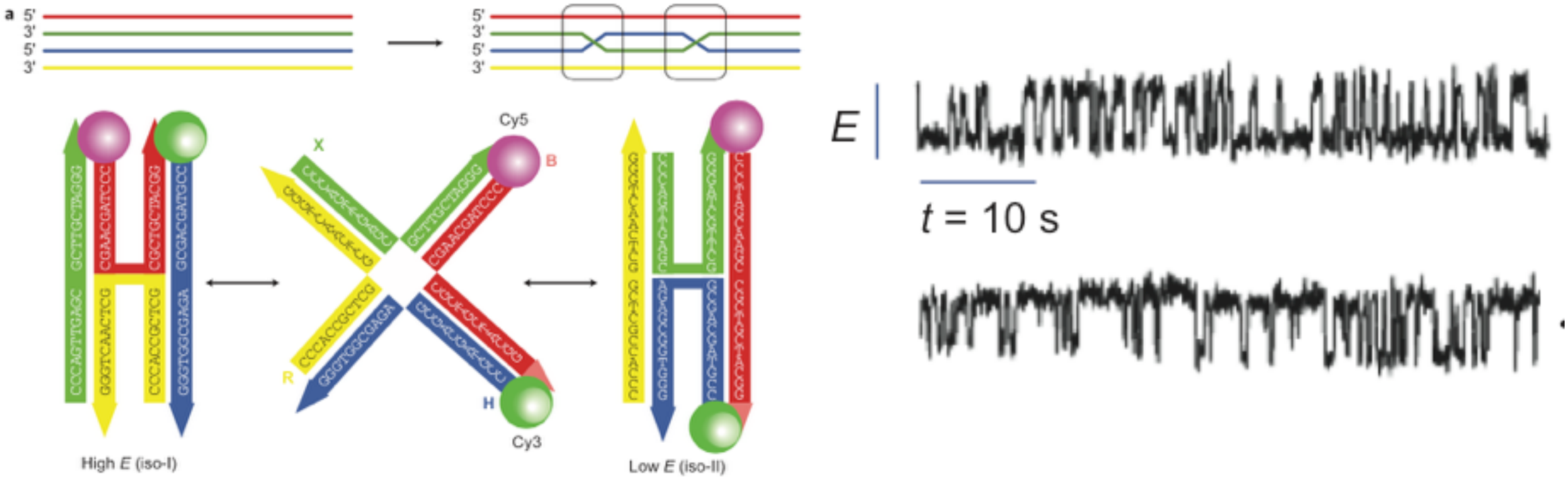

True
K=1
K=2
K=3

Folding energy landscapes of bio-molecules are rugged

Many (functionally competent) folded-states

Each folded-states can have different kinetic rates in their conformational dynamics

“static heterogeneity”



Hyeon et al (Nat. Chem. 2012)

“dynamic heterogeneity”?

It may be observed depending on

Height of energy barrier & Observation time

Vol 463|4 February 2010|doi:10.1038/nature08717

nature

LETTERS

Multiple native states reveal persistent ruggedness of an RNA folding landscape

Sergey V. Solomatin[1], Max Greenfeld[2], Steven Chu[3,4]† & Daniel Herschlag[1]

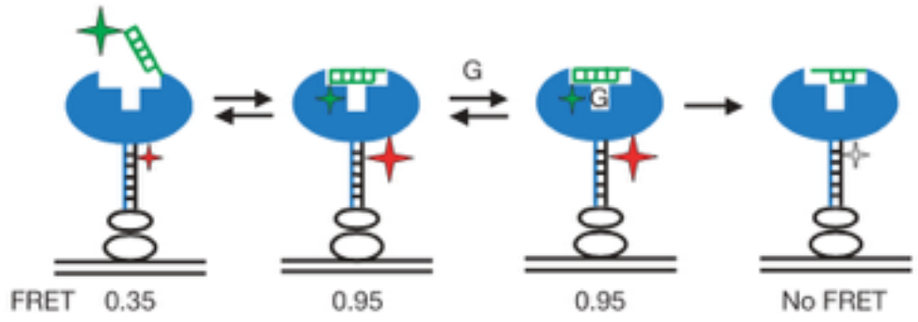


Figure 1 | Docking and cleavage of the oligonucleotide substrate by the *Tetrahymena* ribozyme, observed using single-molecule FRET. The