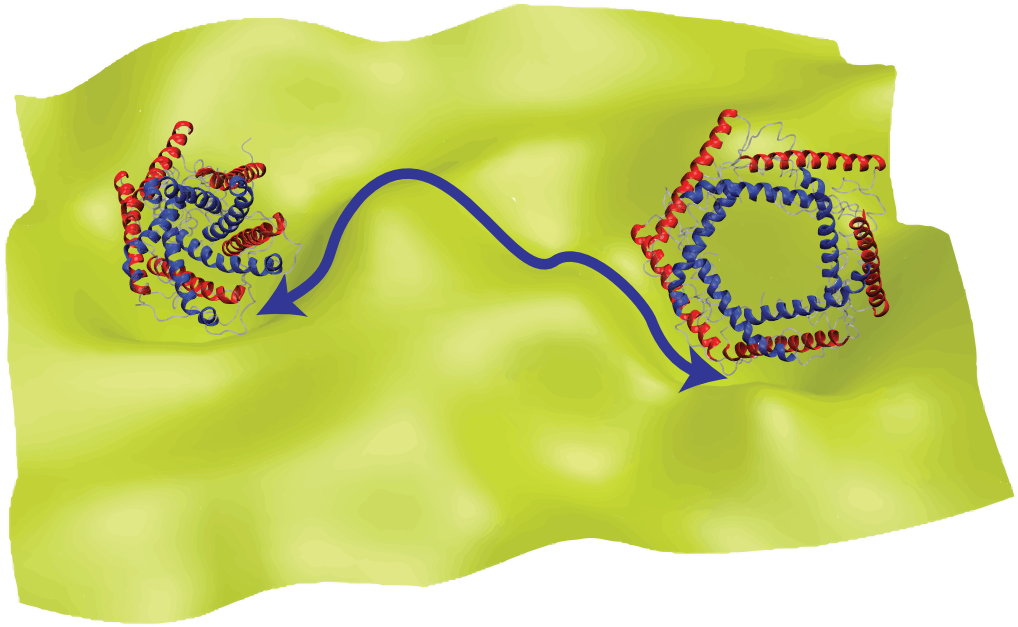
Exploring Complex Reaction Pathways

Mahmoud Moradi

Department of Chemistry and Biochemistry

University of Arkansas



"Hands-on" Workshop on Enhanced Sampling and Free-Energy Calculation at Urbana, IL

September 29, 2017

Outline

- **Introduction**
  - How to study large-scale conformational changes?
- **Methodology**
  - Empirical search for good pulling protocols
  - Iterative combination of free energy calculation methods and path-finding algorithms
- **Theoretical Framework**
  - Effective Riemannian diffusion model
- **Applications**
  - Membrane transport proteins

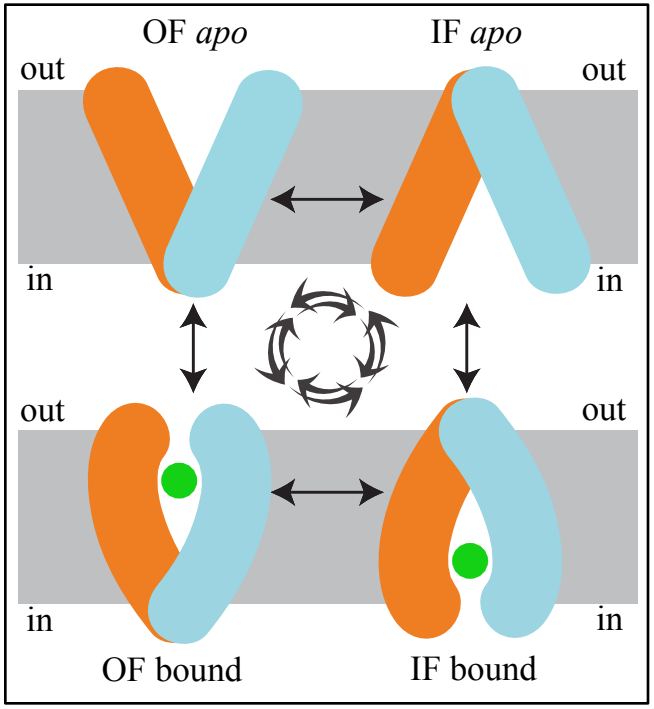
Outline

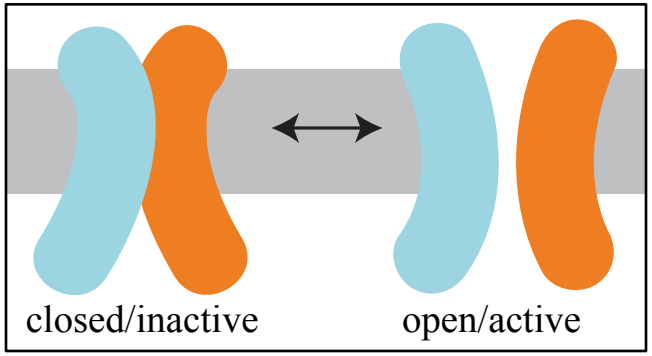
- **Introduction**
  - How to study large-scale conformational changes

Large-Scale Conformational Changes in Membrane Transport Proteins

- **Membrane transporters** rely on large-scale conformational changes between **inward-facing (IF)** and **outward-facing (OF)** states (**alternating access mechanism**)

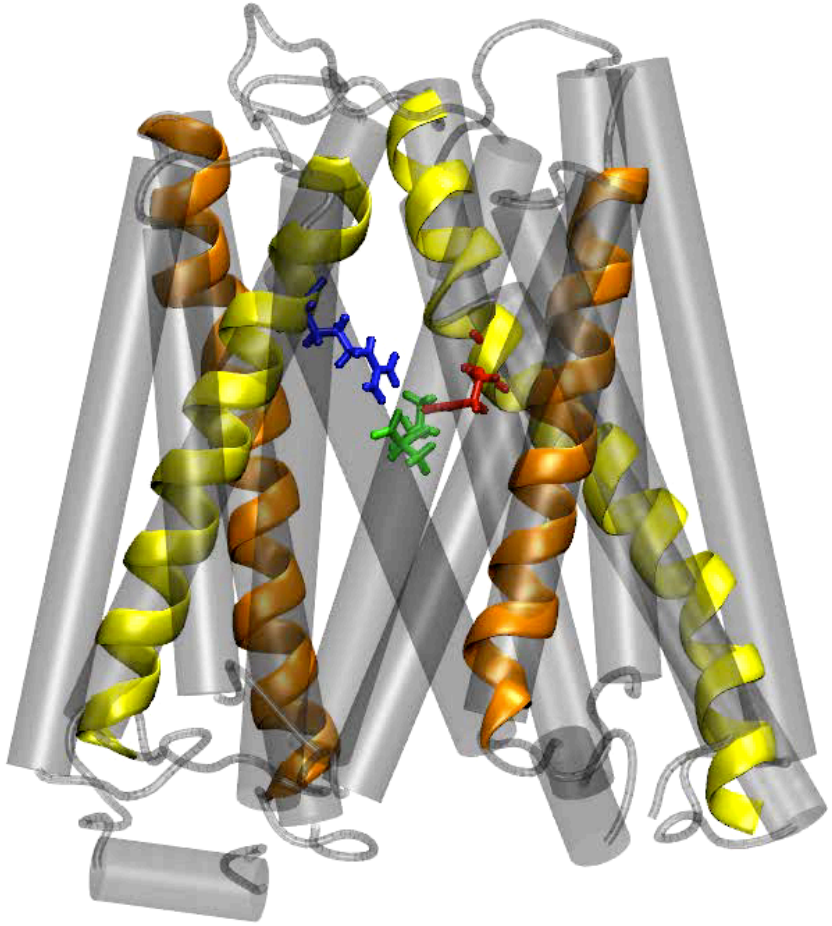
- **Channels** may require large-scale conformational changes between their **open/active** and **closed/inactive** states.





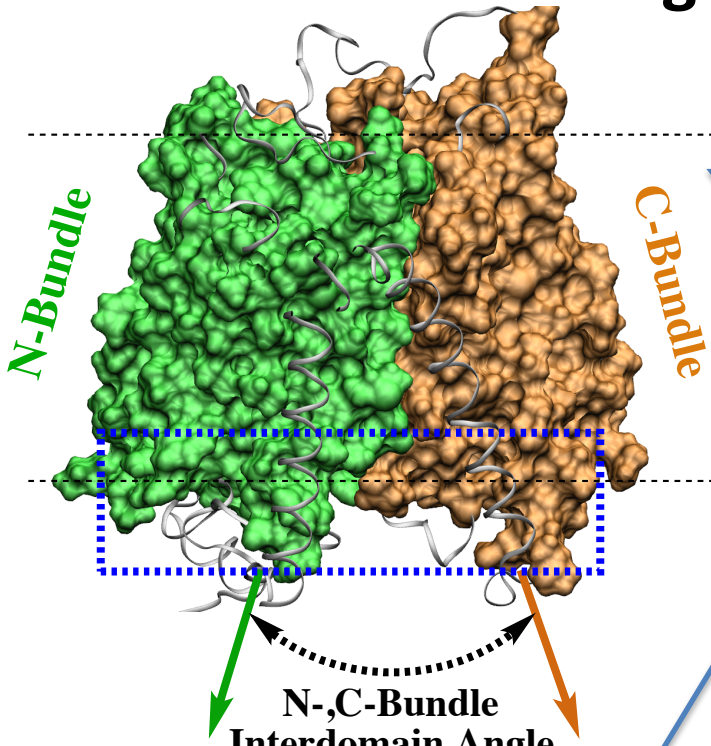
A Case Study:
Proton-coupled Oligopeptide Transporters (POTs)

GkPOT (4IKV, 1.9 Å)
~100,000 atoms
Conventional unbiased simulations performed:
8 conditions × 400 ns × 2 repeats = 6.4 μs

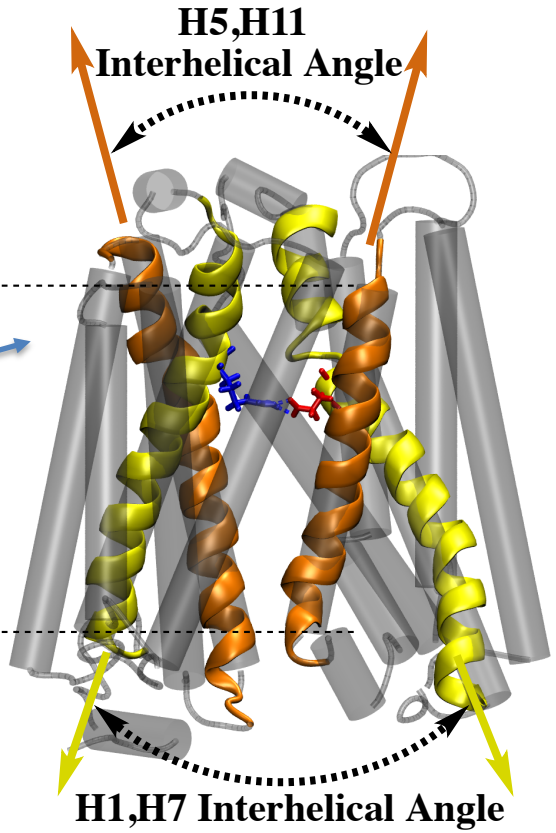


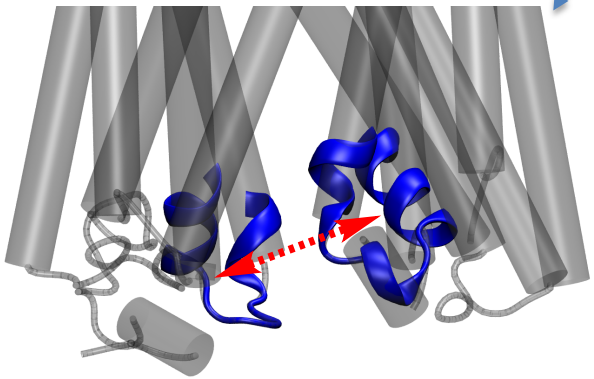
K Immadisetty, J Hettige, and M Moradi, ***What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT?*** J. Phys. Chem. B, **121**:3644-3656, 2017.

Monitoring Global and Local Conformational Changes

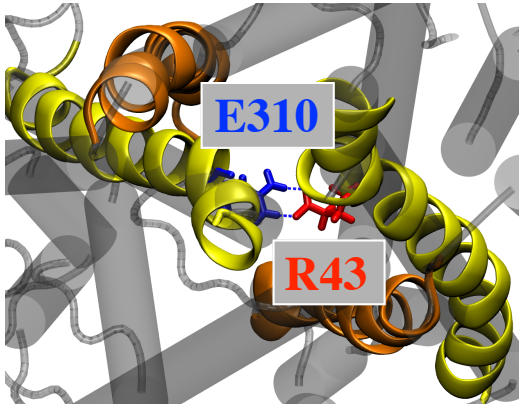


Global

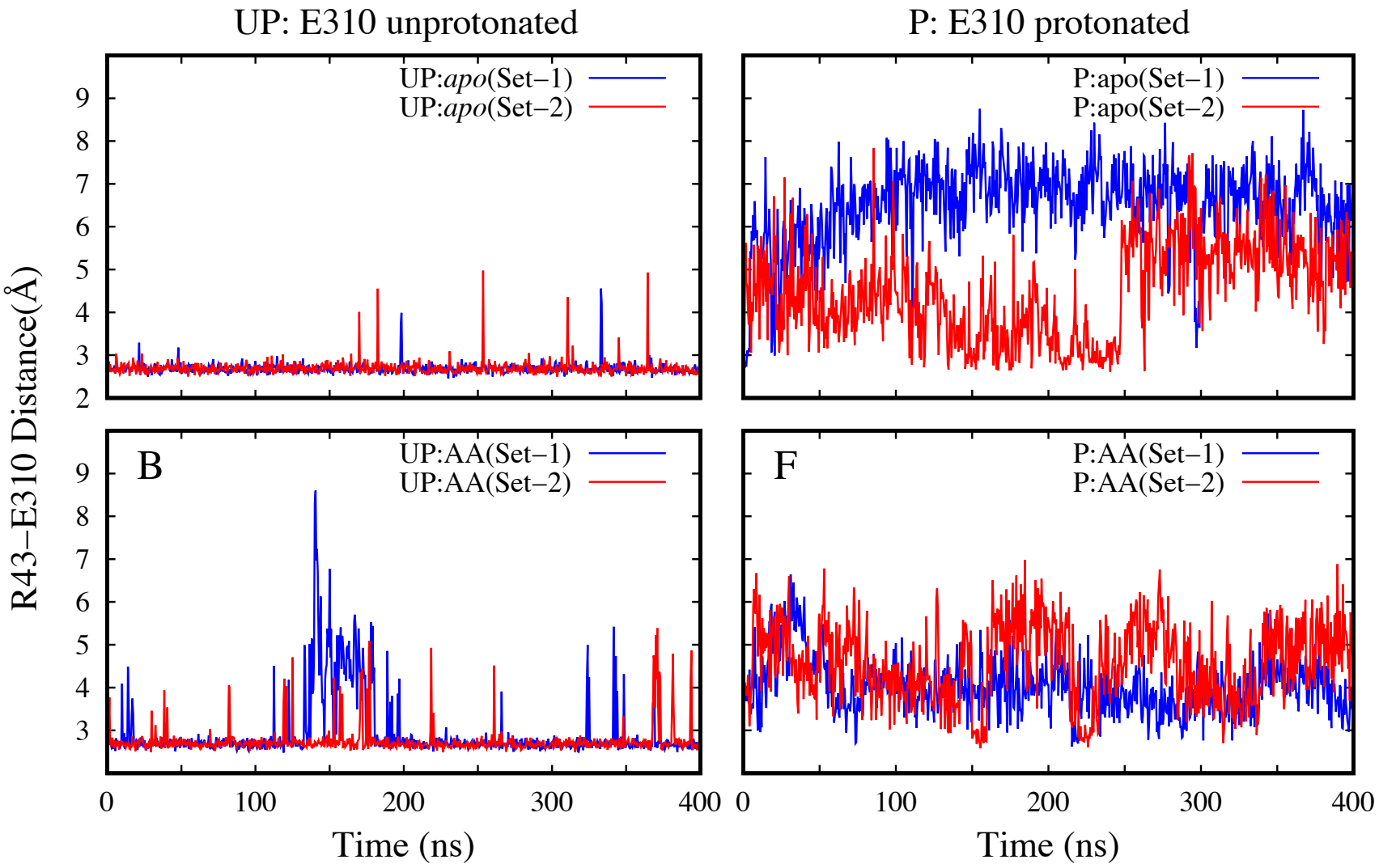




Local
(Salt Bridge)

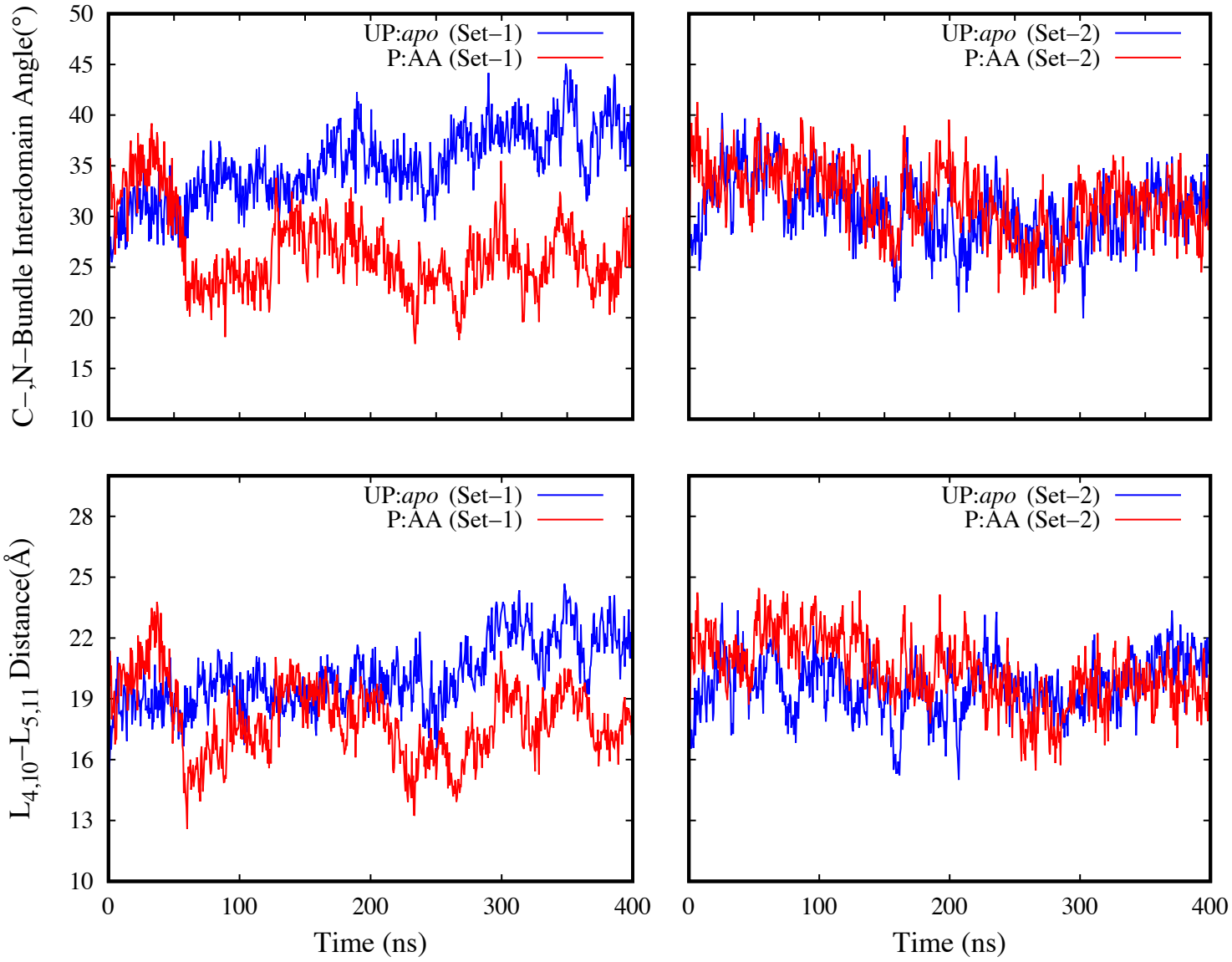


Local Conformational Changes



There is a clear distinction between different conditions.

Global Conformational Changes



There is no clear reproducible distinction between different conditions.

Global Conformational Changes

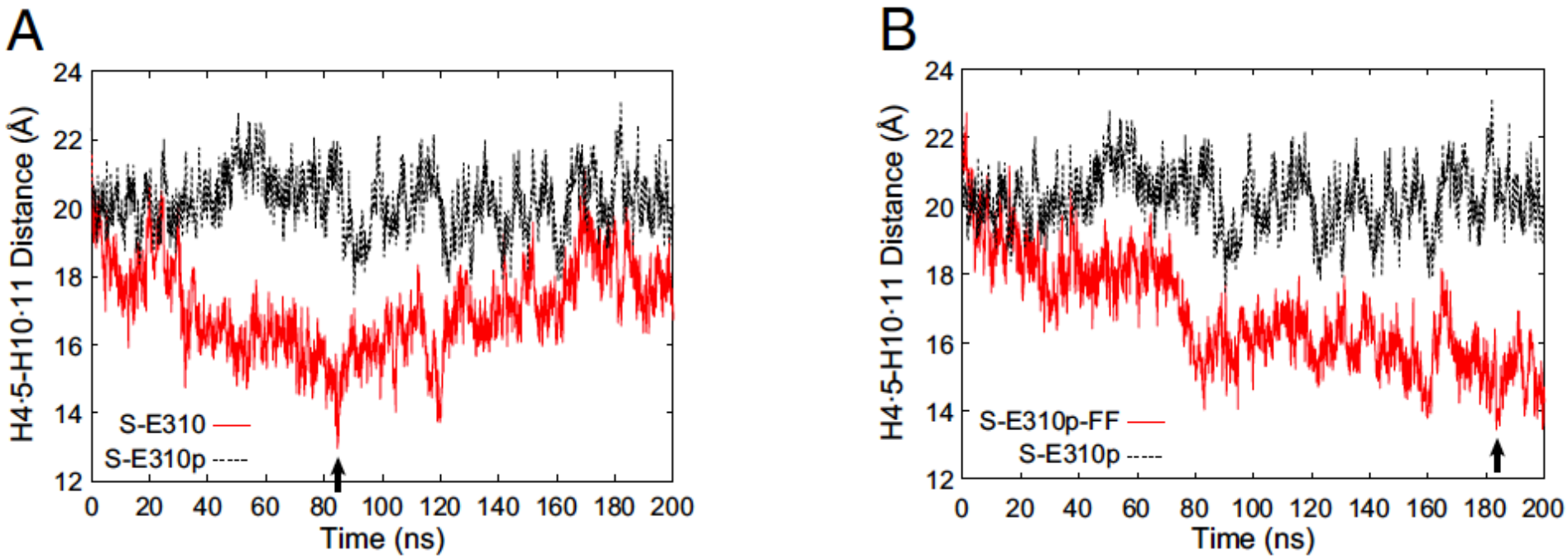
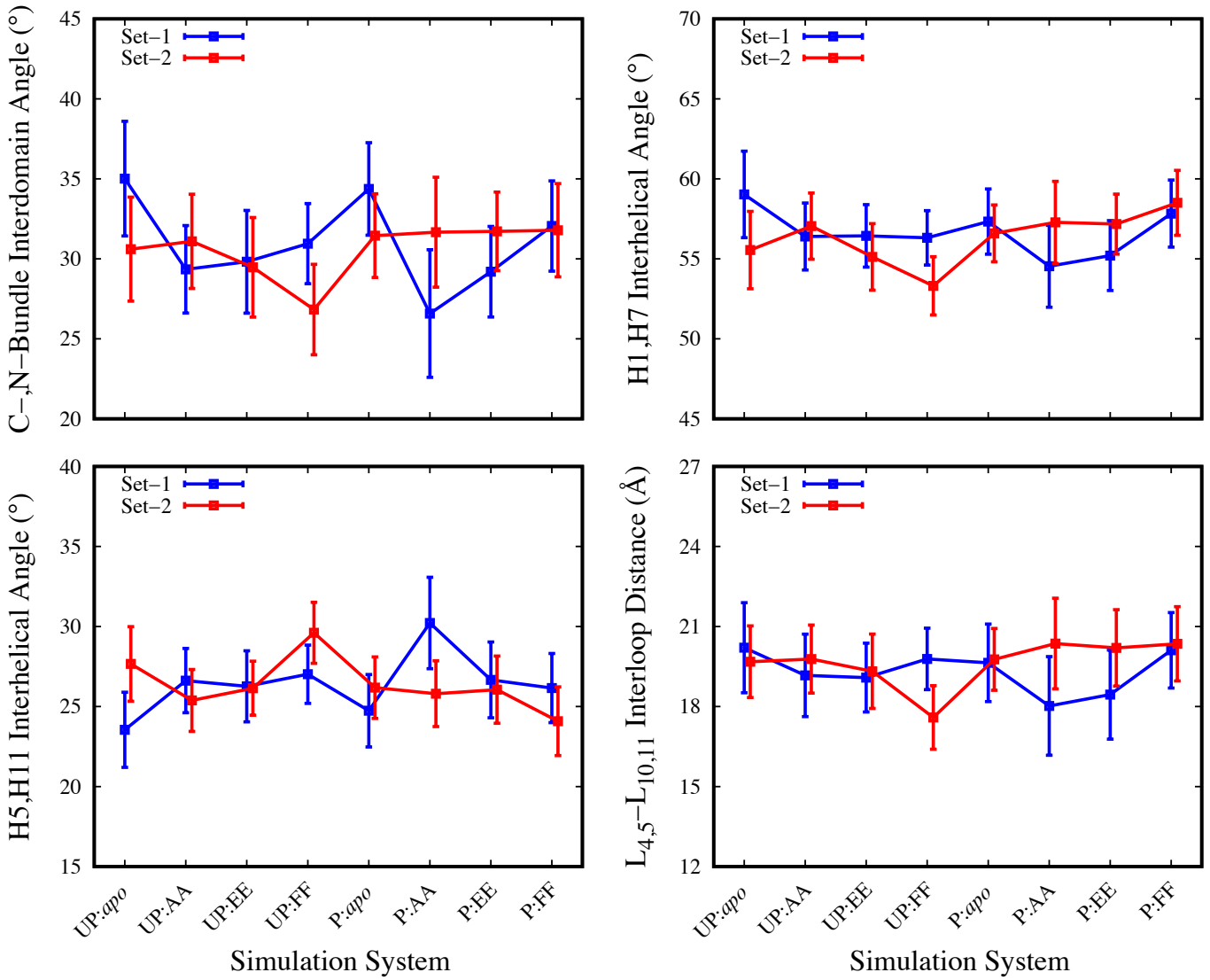


Fig. 4. Structural transitions revealed by MD simulations. Time series of the distances between the cytosolic-side helices of the N- and C-terminal bundles (residues 141–156 and 420–440, respectively) in the S-E310 and S-E310p-FF simulations.

"…",PNAS 2013, …

There is no clear reproducible distinction between different conditions even if it is claimed to be so in the literature!

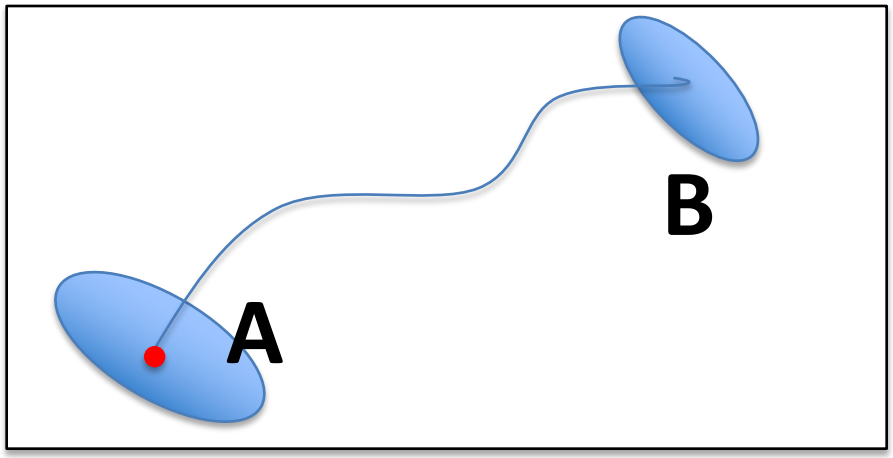
Global Conformational Changes



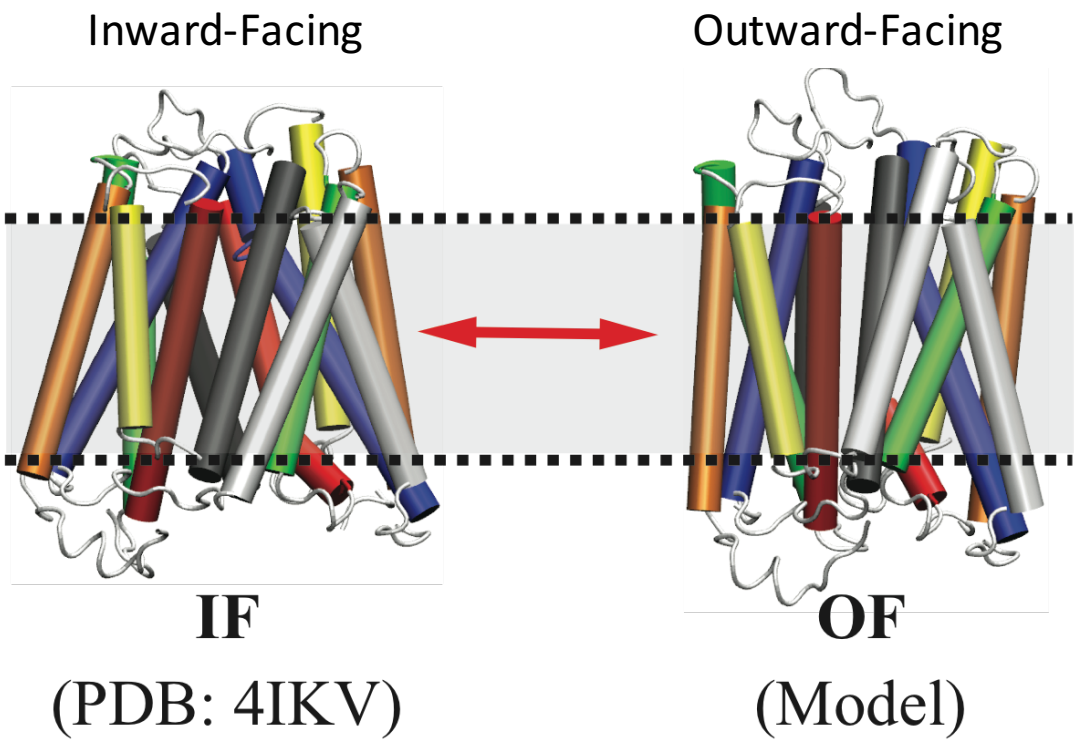
There is no statistically significant distinction between different conditions.

- **Introduction**
  - How to study large-scale conformational changes

    It is not reasonable to speculate about the conformational transition between two states based on fluctuations around one of the end points.

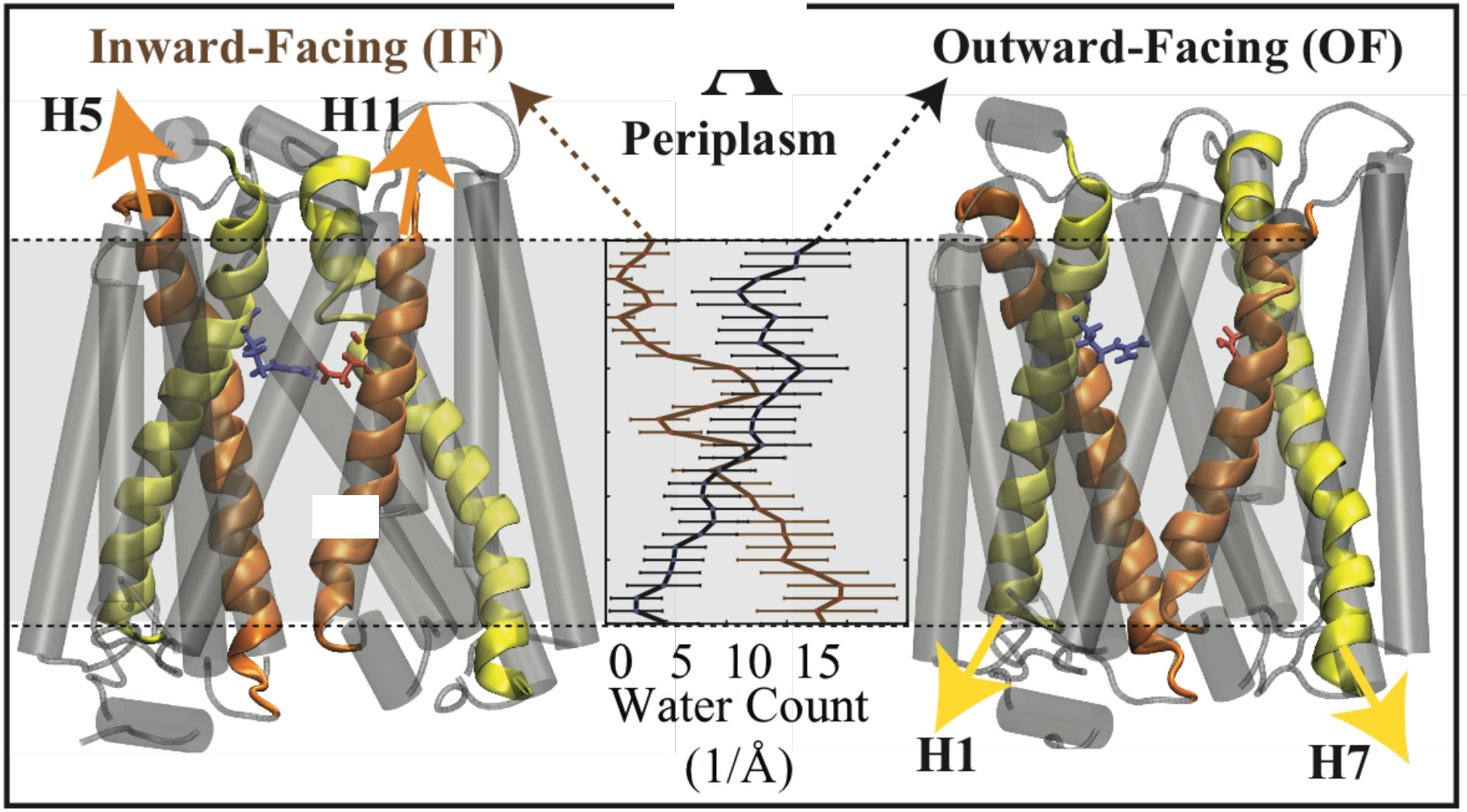


How to study large-scale conformational changes?



GkPOT

How to study large-scale conformational changes?



- **Introduction**
  - How to study large-scale conformational changes?
- **Methodology**
  - Empirical search for good pulling protocols
  - Iterative combination of free energy calculation methods and path-finding algorithms

Sampling Ideas

- Free energy calculations require **dimensionality reduction.**
- Traditionally, this is done by designing intuitive, ad-hoc, knowledge-based **collective variables.**
- Another approach is to use **data-driven** collective variables using standard **dimensionality reduction techniques** (PCA, diffusion maps, etc).
- Alternatively (or in combination with the above approaches), one can calculate **free energy along a transition path** (a 1D curve).
- The path can be obtained from **path-finding algorithms.**
- Since sampling is never perfect the procedure could be **iterative** to reach higher accuracies.

Sampling Ideas

- **Reaction coordinates**
  - System-specific collective variables
- **Searching for efficient pulling protocols**
  - An empirical approach to sampling
- **Along-the-curve free energy calculations**
  - Free energy calculations combined with path-finding algorithms
- **Iterative sampling**
  - A posteriori tests of self-consistency

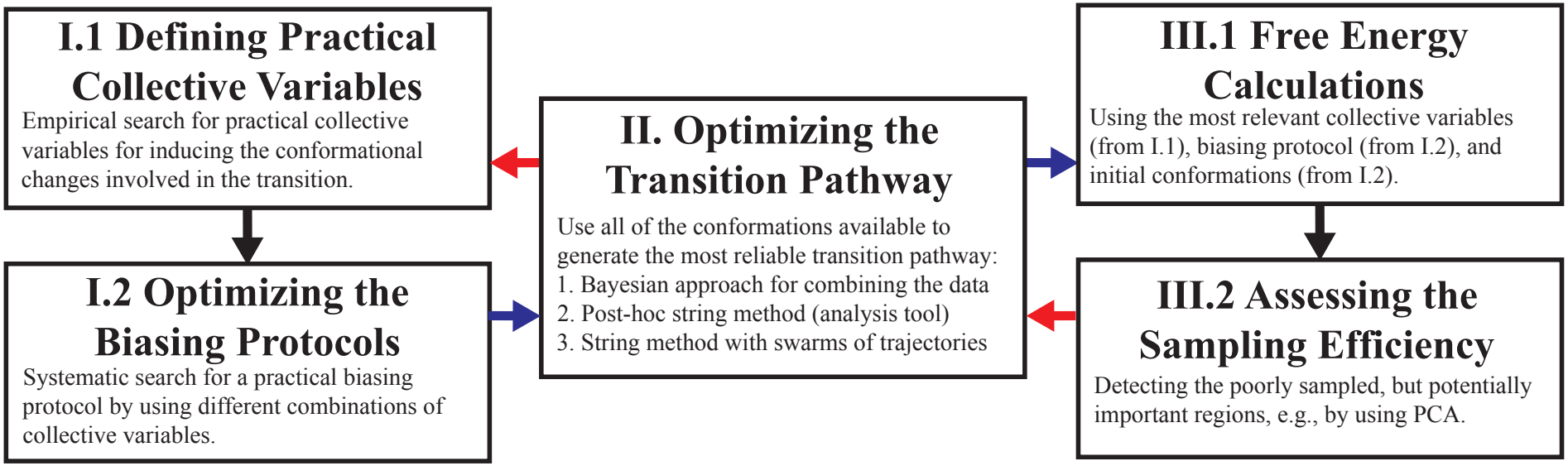
Moradi et al., **Proc Natl Acad Sci 106** 20746 (**2009**)
Moradi et al., **Chem Phys Lett 518** 109 (**2011**)
Moradi et al., **J Chem Phys 133** 125104 (**2010**)
Moradi et al., **Int J Quantum Chem 110** 2865 (**2010**)
Moradi et al., **Biophys J 100** 1083 (**2011**)
Moradi et al., **J Phys Chem B 115** 8645 (**2011**)
Moradi et al., **PLoS Comput Biol 8** e1002501 (**2012**)
Moradi et al., **Nucleic Acid Res 41** 33 (**2013**)

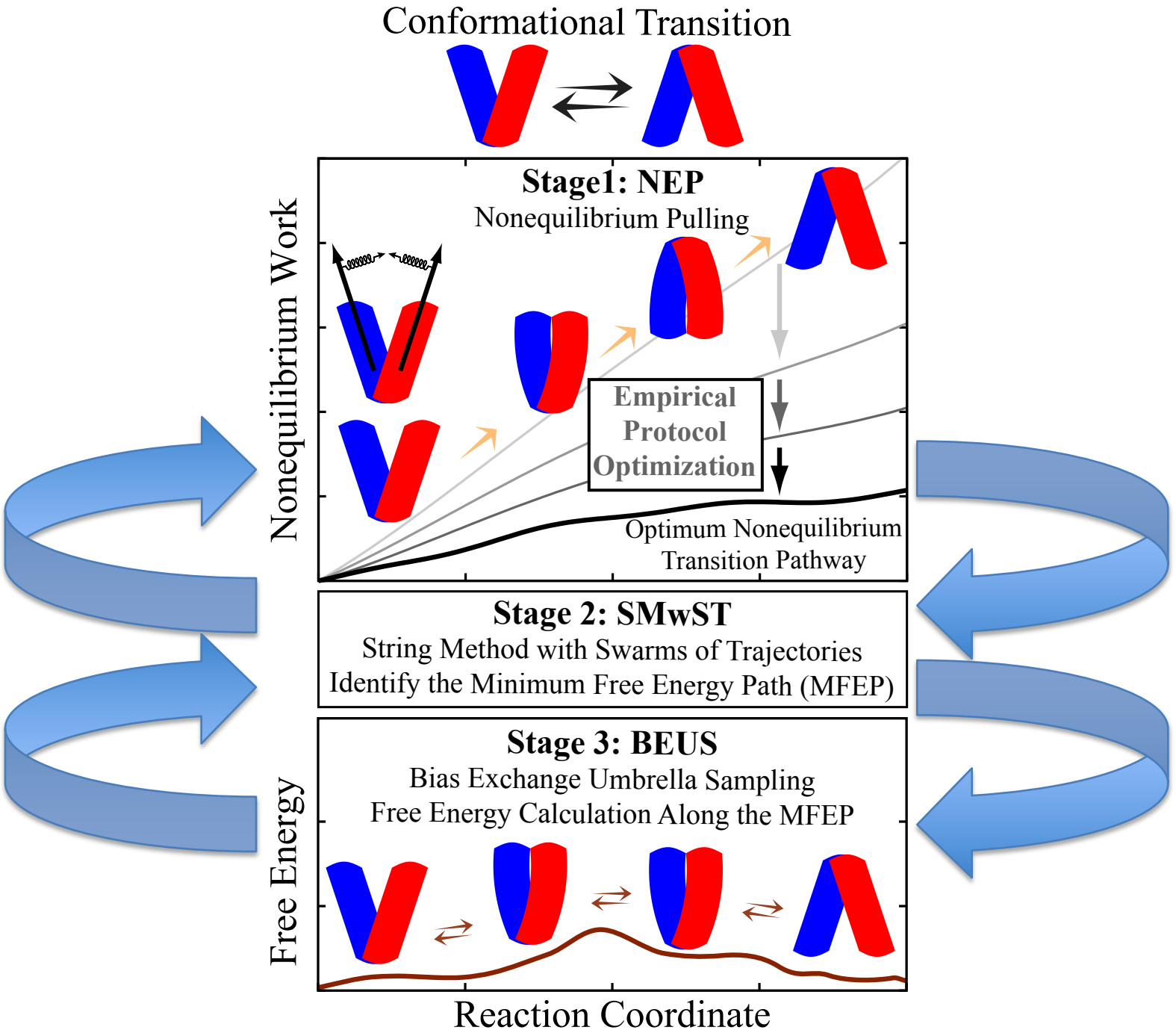
Moradi et al., **Proc Natl Acad Sci 110** 18916 (**2013**)
Moradi et al., **J Phys Chem Lett 4** 1882 (**2013**)
Moradi et al., **Methods Mol Biol 924** 313 (**2013**)
Moradi et al., **J Chem Phys 140** 034114 (**2014**)
Moradi et al., **J Chem Phys 140** 034115 (**2014**)
Moradi et al., **J Chem Theory Comput 10** 2866 (**2014**)
Moradi et al., **J Phys Conf Ser 640** 012014 (**2015**)
Moradi et al., **J Phys Conf Ser 640** 012020 (**2015**)

Moradi et al., **Nat Commun 6** 8393 (**2015**)
Fakharzadeh & Moradi, **J Phys Chem Lett 7** 4980 (**2016**)

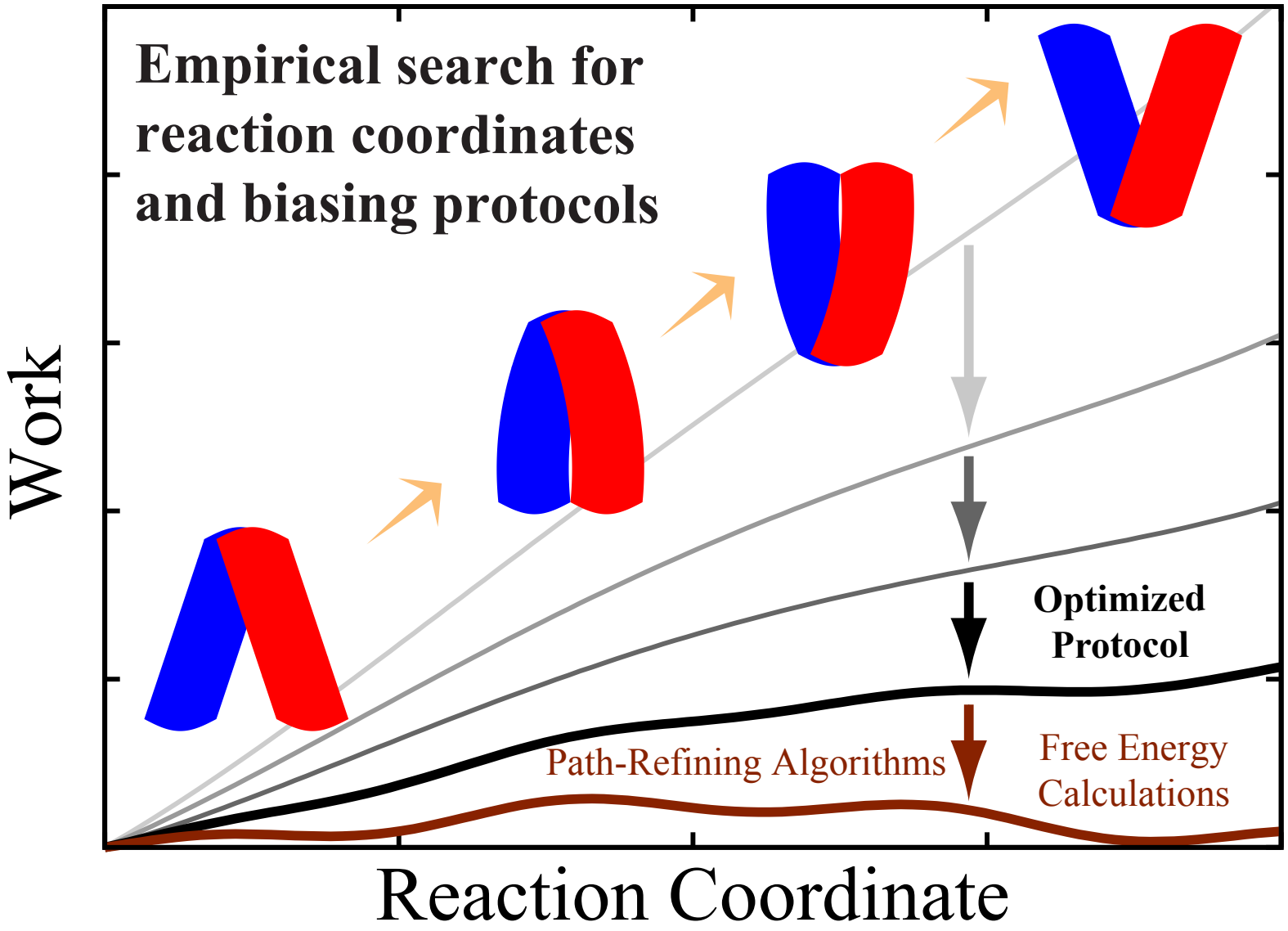
Sampling Ideas

- **Reaction coordinates**
  - System-specific collective variables
- **Searching for efficient pulling protocols**
  - An empirical approach to sampling
- **Along-the-curve free energy calculations**
  - Free energy calculations combined with path-finding algorithms
- **Iterative sampling**
  - A posteriori tests of self-consistency




Conformational Transition
Stage1: NEP
Nonequilibrium Pulling
Nonequilibrium Work
Empirical
Protocol
Optimization
Optimum Nonequilibrium
Transition Pathway
Stage 2: SMwST
String Method with Swarms of Trajectories
Identify the Minimum Free Energy Path (MFEP)
Stage 3: BEUS
Bias Exchange Umbrella Sampling
Free Energy Calculation Along the MFEP
Free Energy
Reaction Coordinate

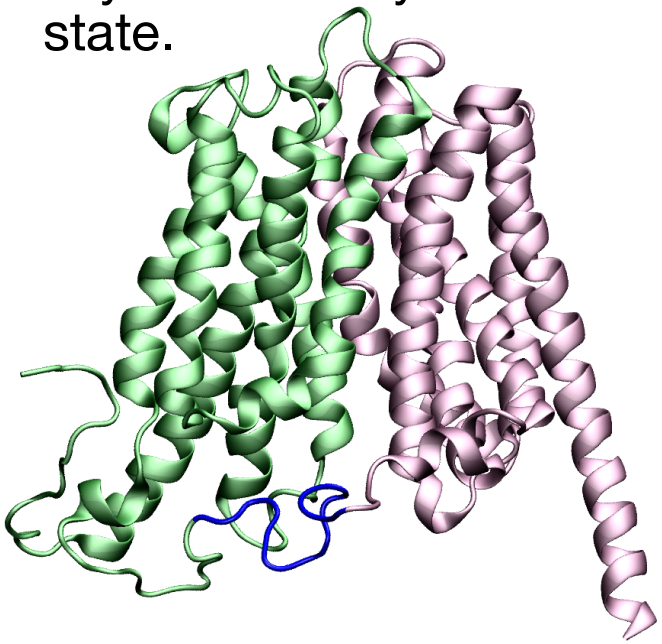
Sampling Ideas

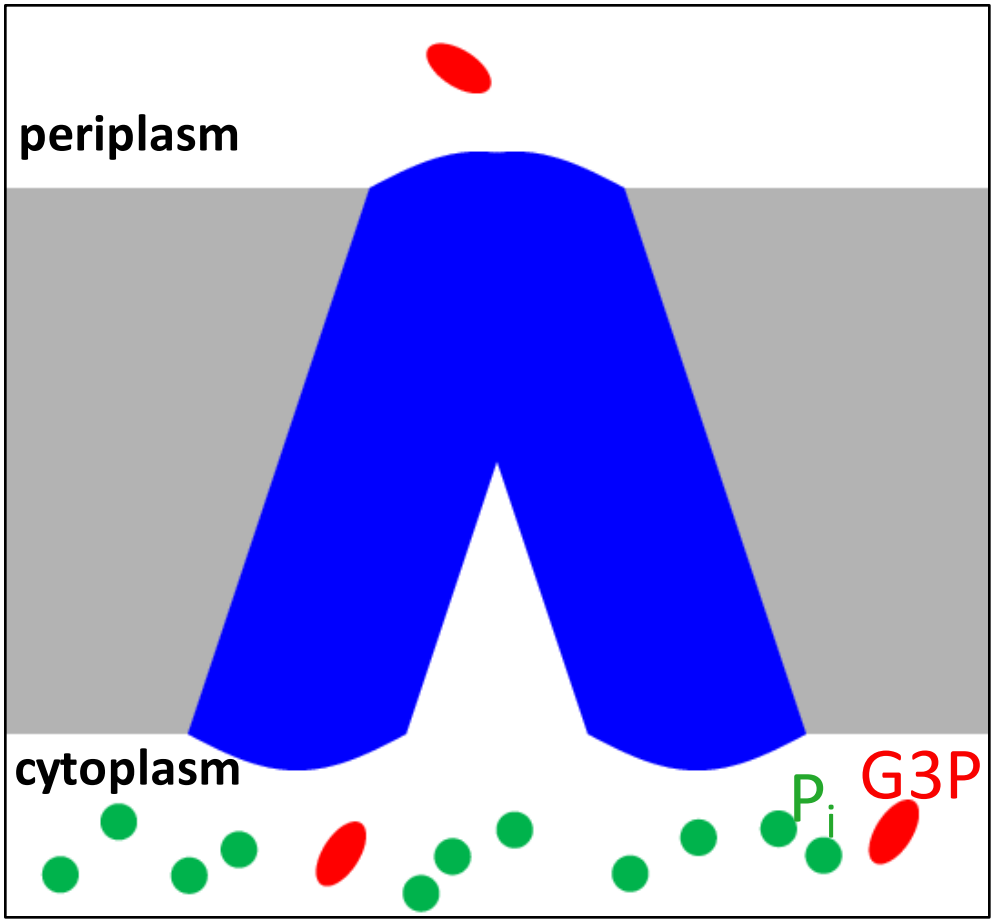


- **Introduction**
  - How to study large-scale conformational changes?
- **Methodology**
  - **Empirical search for good pulling protocols**
  - Iterative combination of free energy calculation methods and path-finding algorithms

Example: Glycerol-3-Phophate Transporter (GlpT)

- Major facilitator superfamily (MFS)
- Secondary active transporter
- Crystalized only in the IF state.

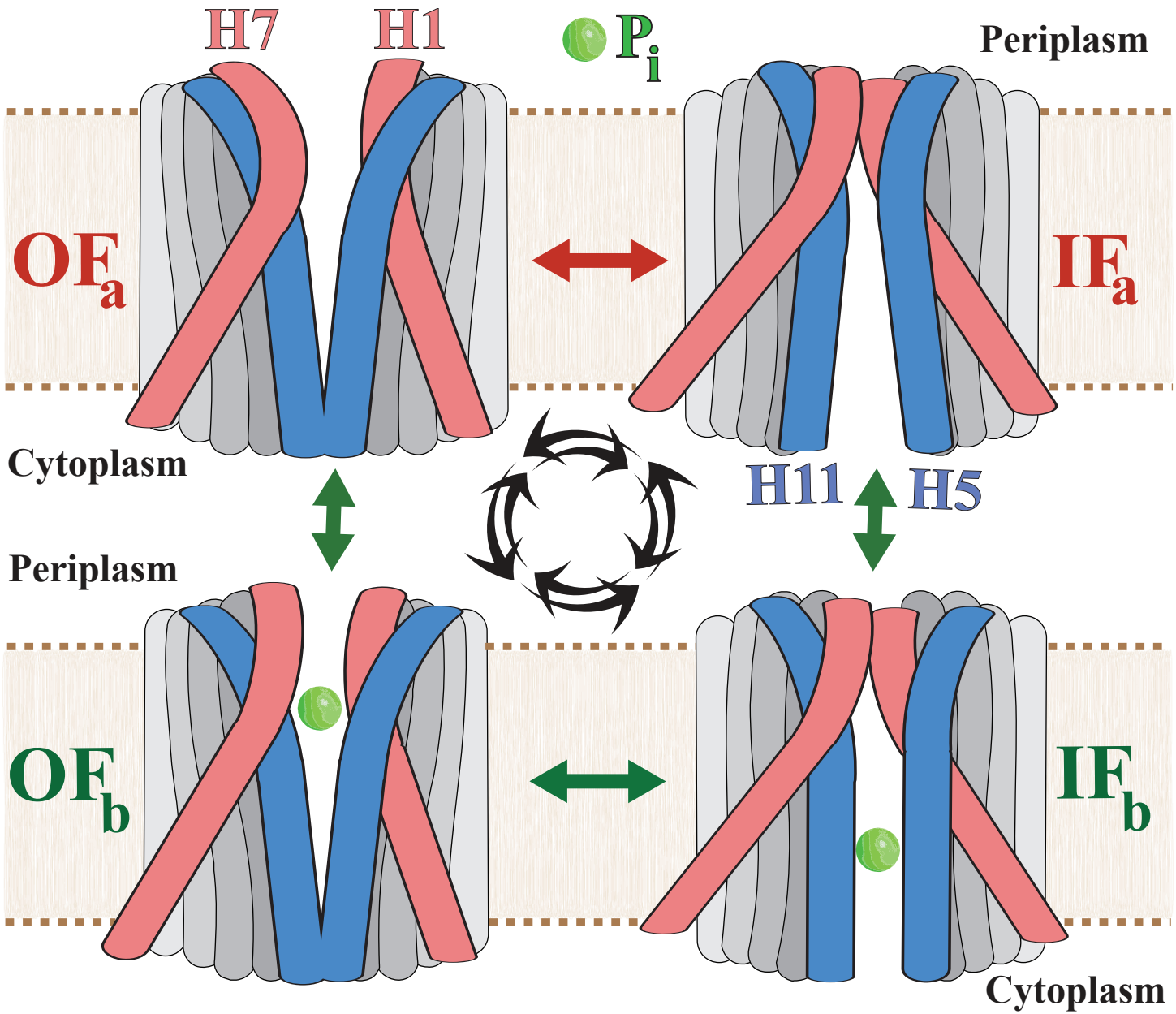




- GlpT transports **G3P** using $\mathbf{P_i}$ gradient.
- $\mathbf{P_i}$:$\mathbf{P_i}$ exchanger (in the absence of organic phosphate)

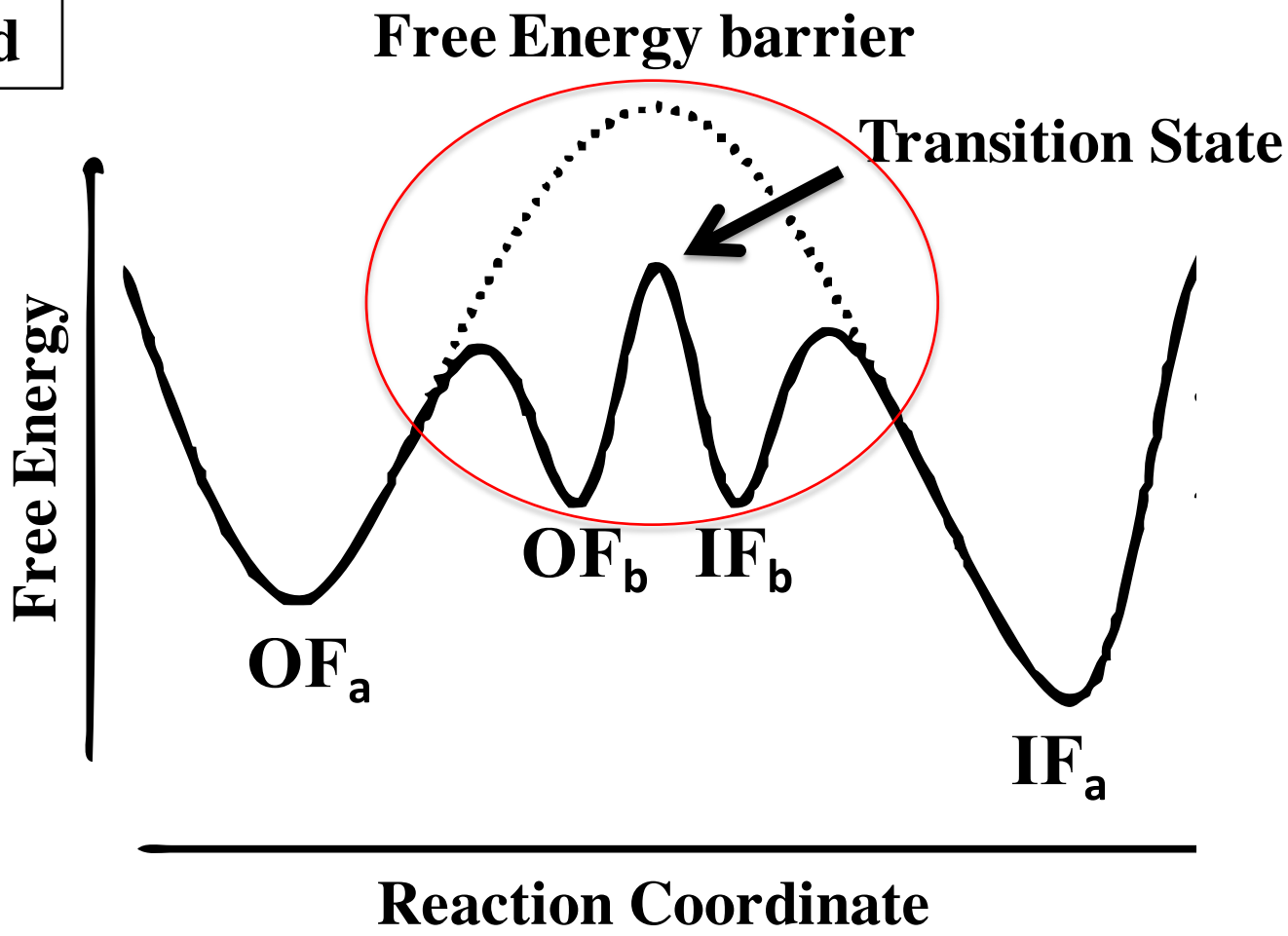
Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

Transport Thermodynamics



Transport Thermodynamics

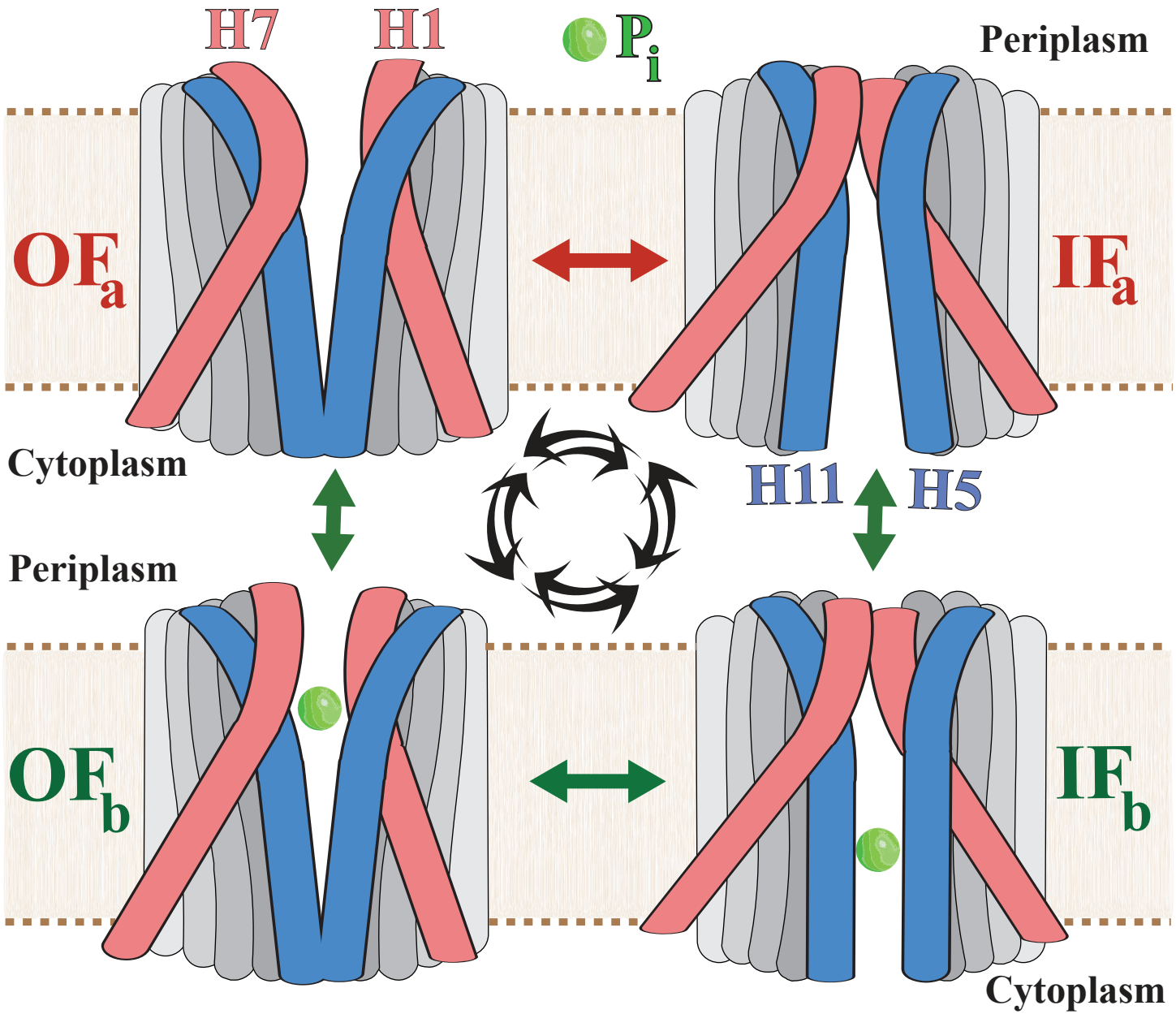
a: *apo*
b: bound



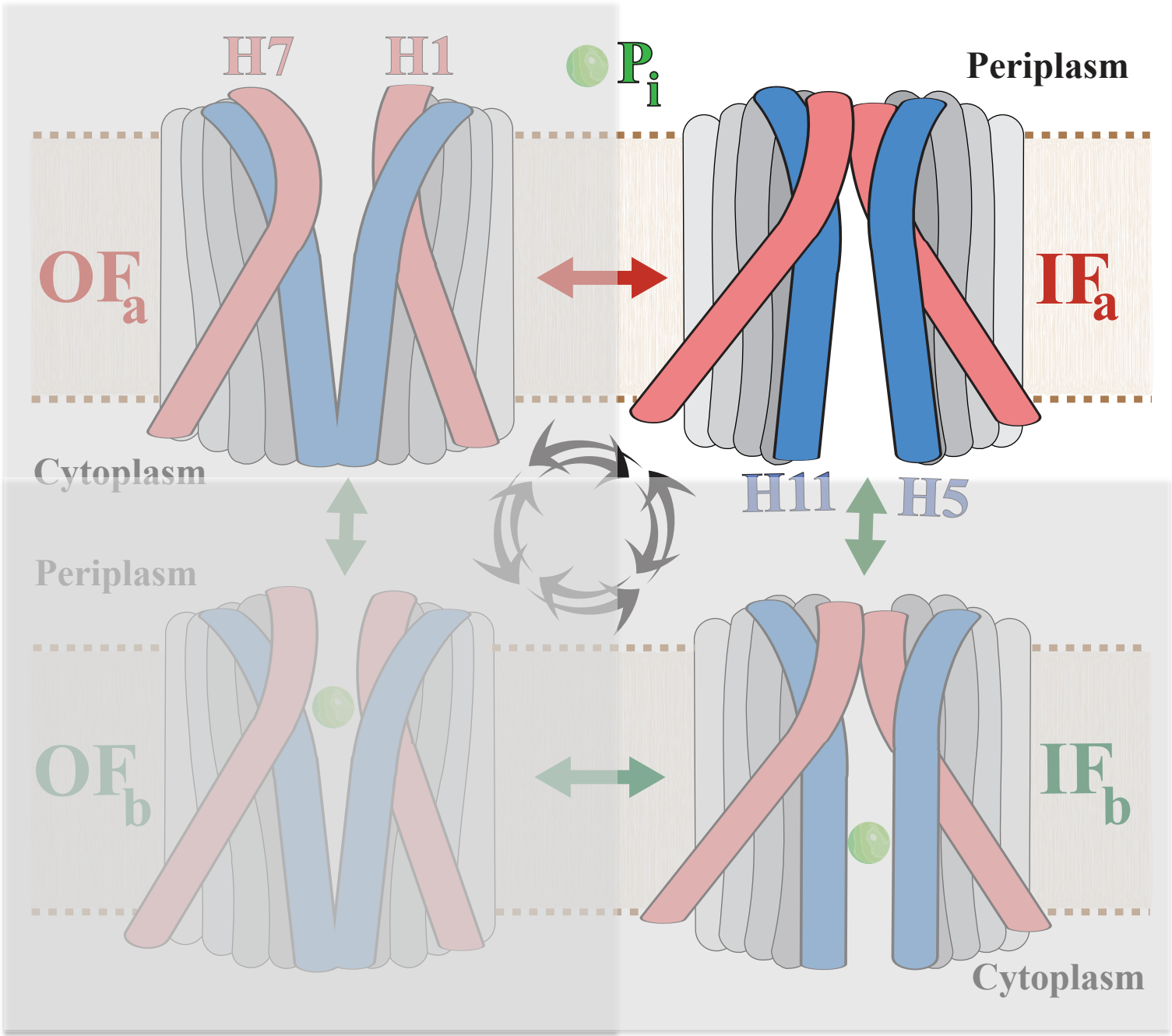
Lemieux, *et al.*, *Curr. Opin. Struct. Biol. 1***4**, 405 (2004).

Law, *et al.*, *Biochemistry* **46**, 12190 (2007).

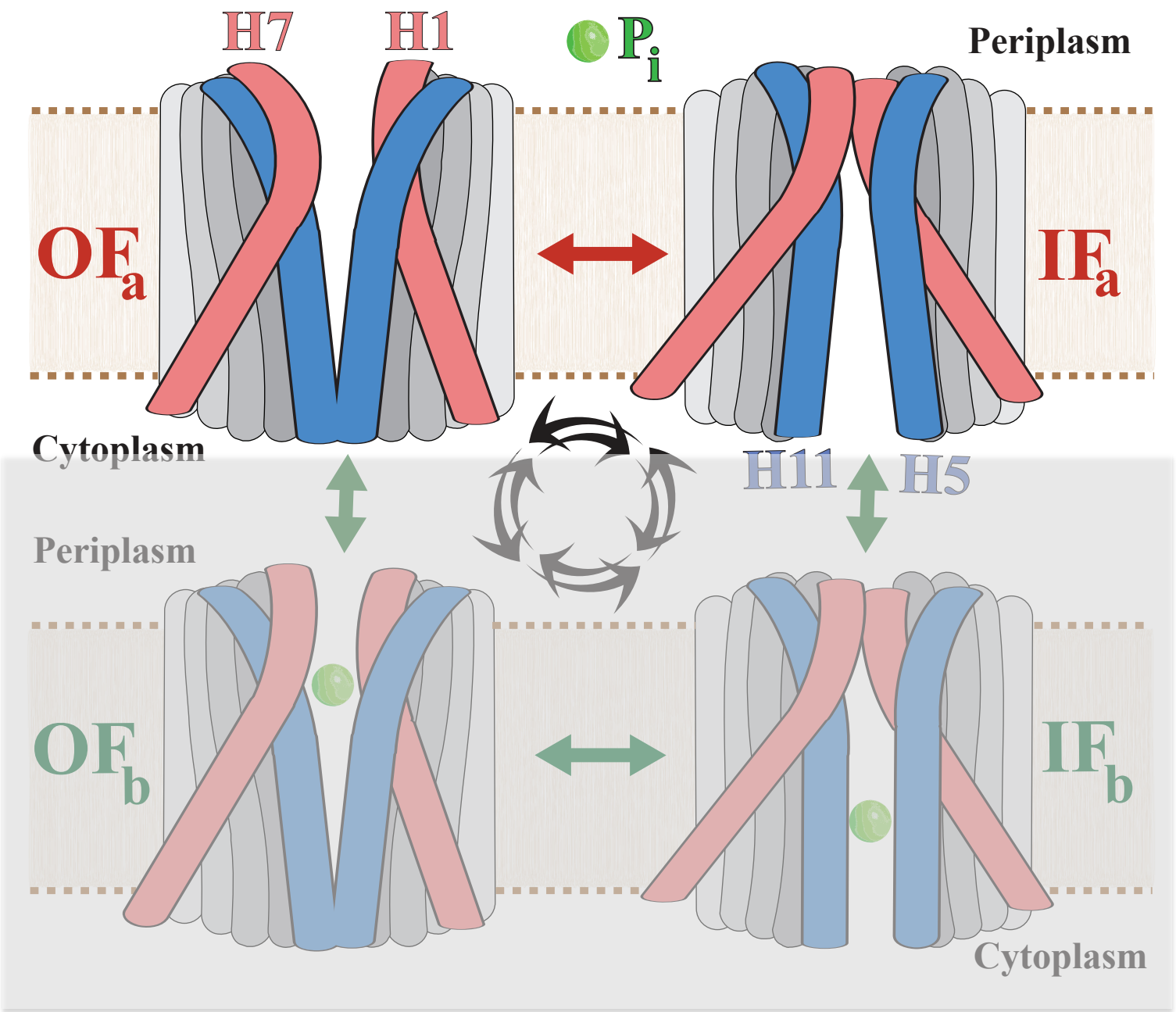
Full Thermodynamic Cycle



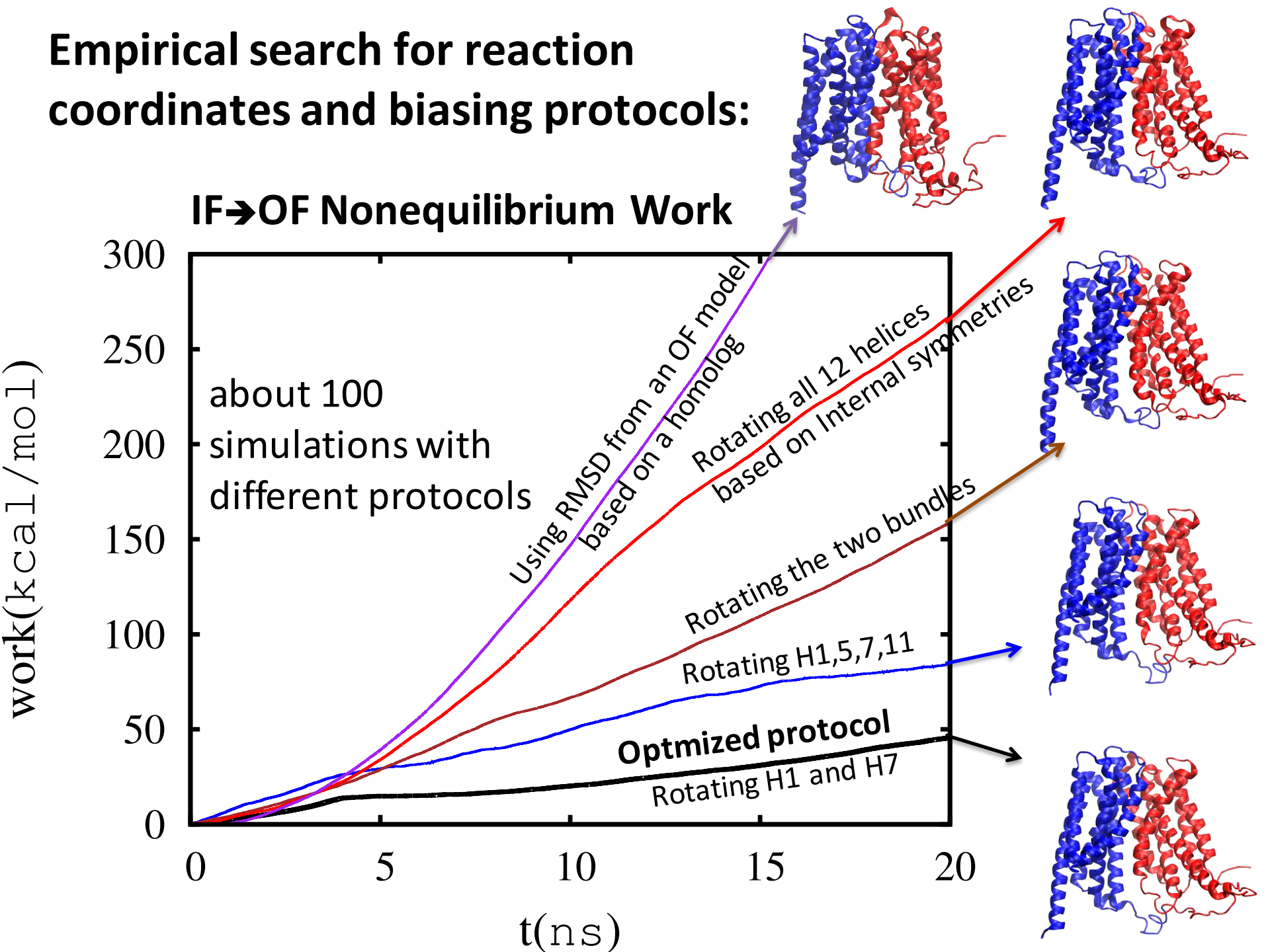
the only available crystal structure



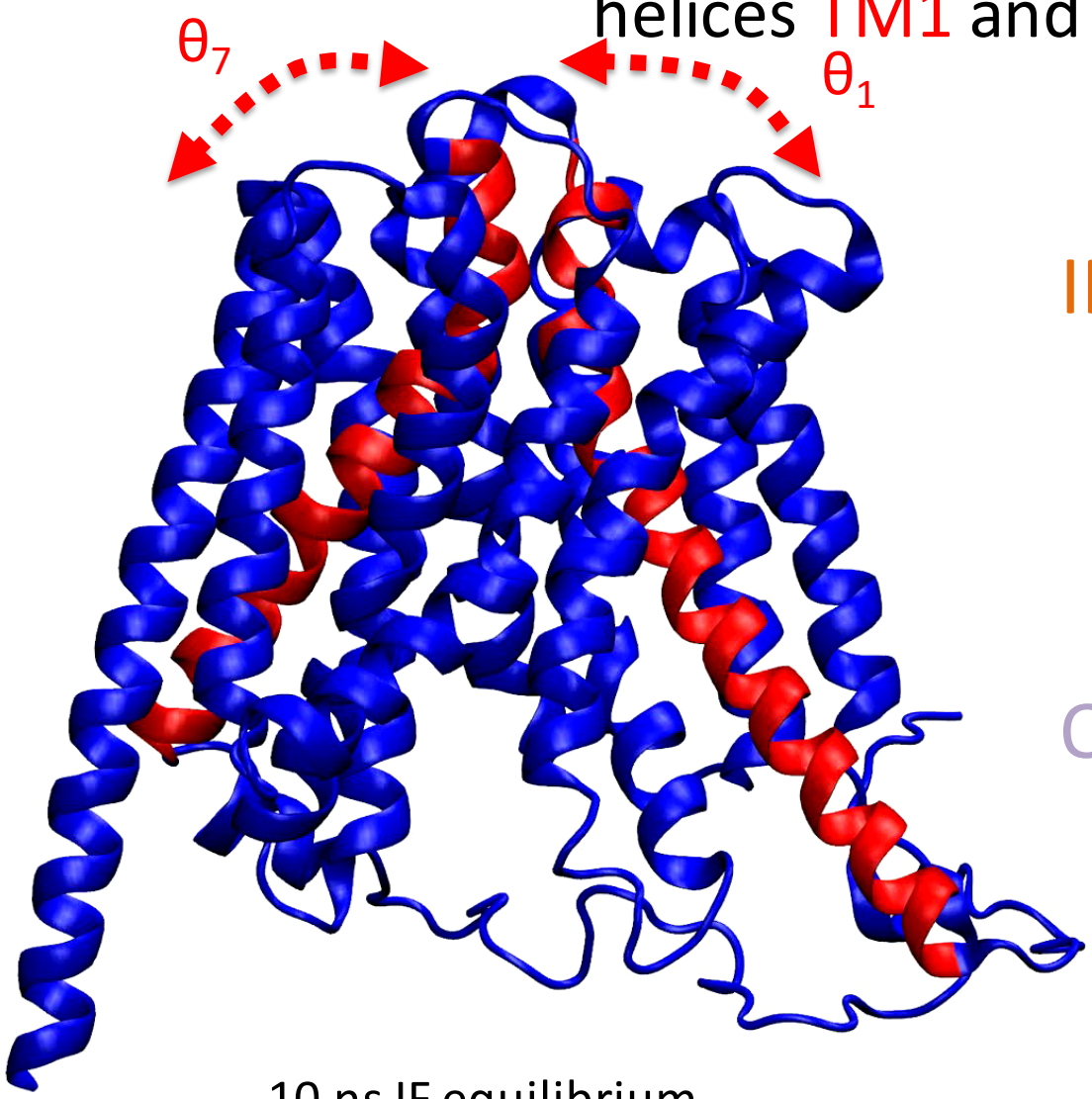
Step 1: $OF_a \leftrightarrow IF_a$

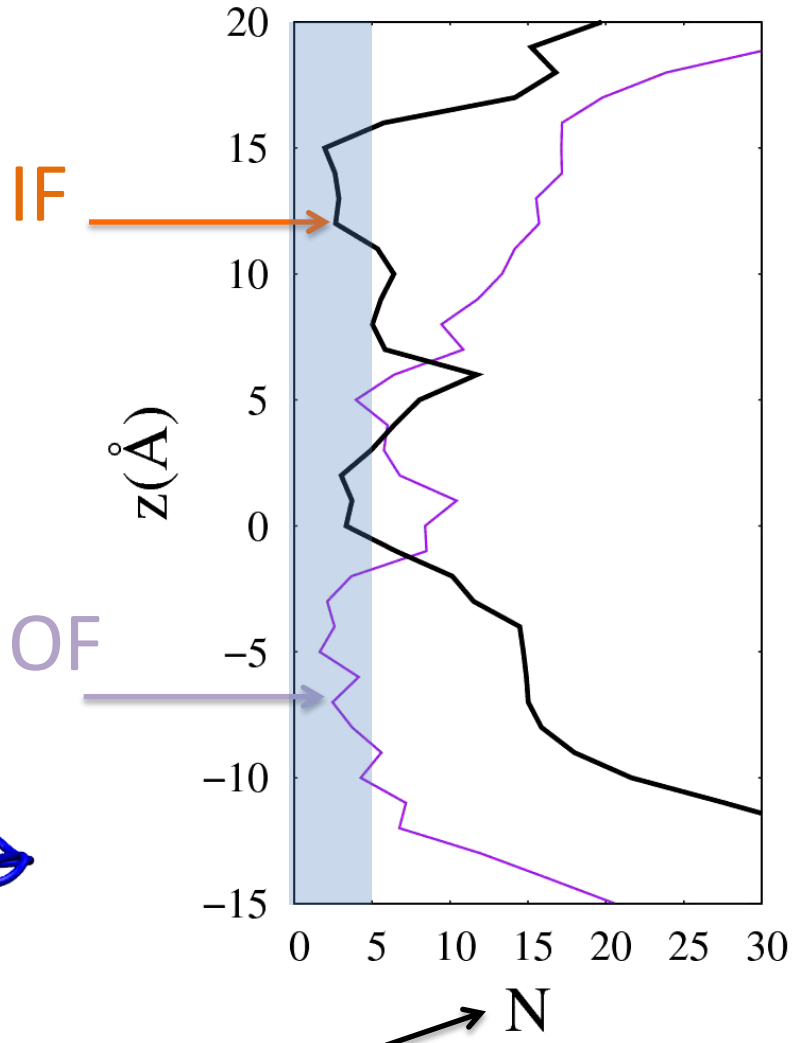


Empirical search for reaction coordinates and biasing protocols:



IF↔OF transition induced by imposing rotational change on helices TM1 and TM7



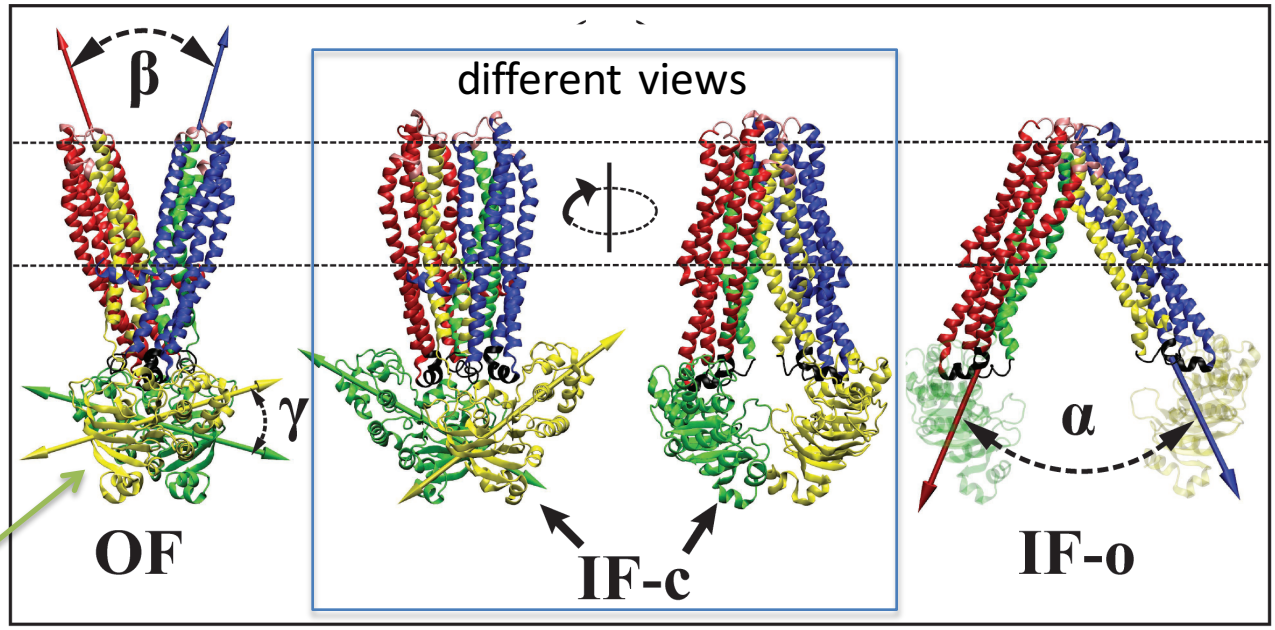


10 ns IF equilibrium
20 ns nonequilibrium (IF➔OF)
10 ns OF equilibrium

Number of water molecules per Å
(averaged over a 1 ns window)

Example: MsbA Transporter

- ATP-binding cassette (**ABC**) exporter with three crystal structures. Ward A., Reyes C. L., Yu J., Roth C. B., Chang G.. PNAS **104** 19005 (2007)
- **Reaction coordinates**:
  α, β, γ (relative orientation of different domains)

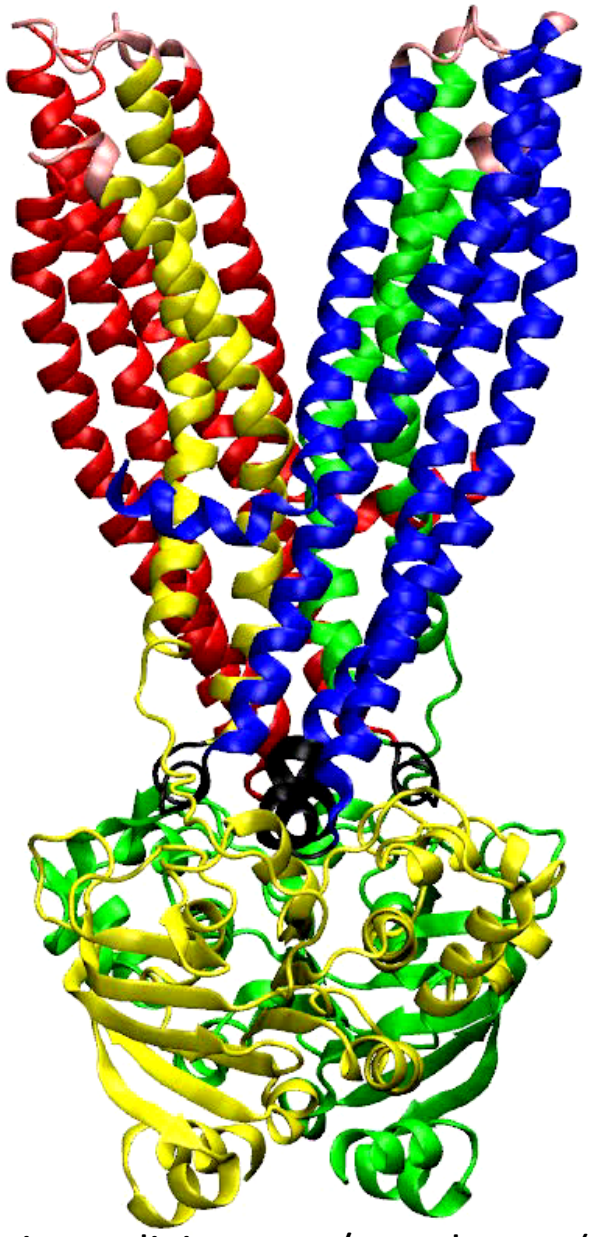


Nucleotide-binding domains (NBD)

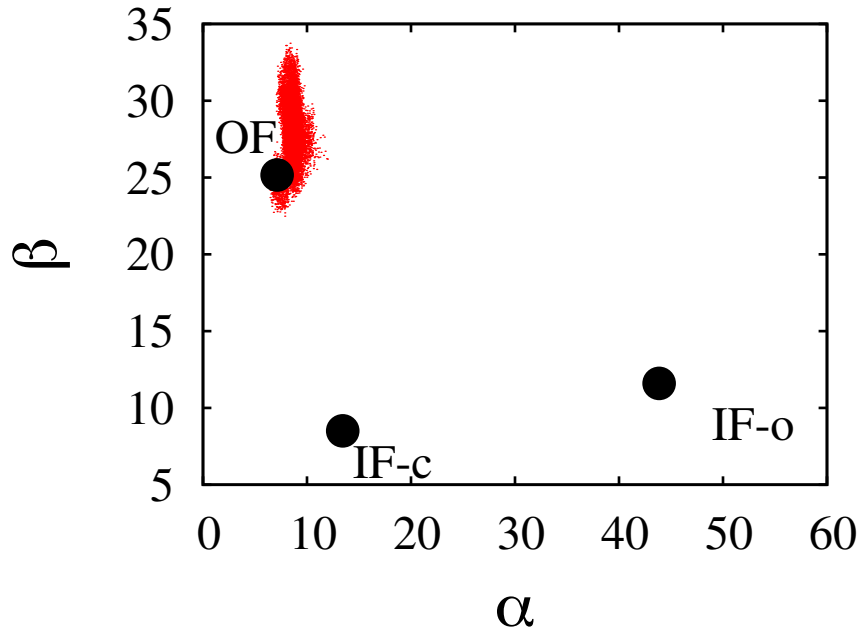
Two IF conformations:
IF-closed (IF-c) and IF-open (IF-o)

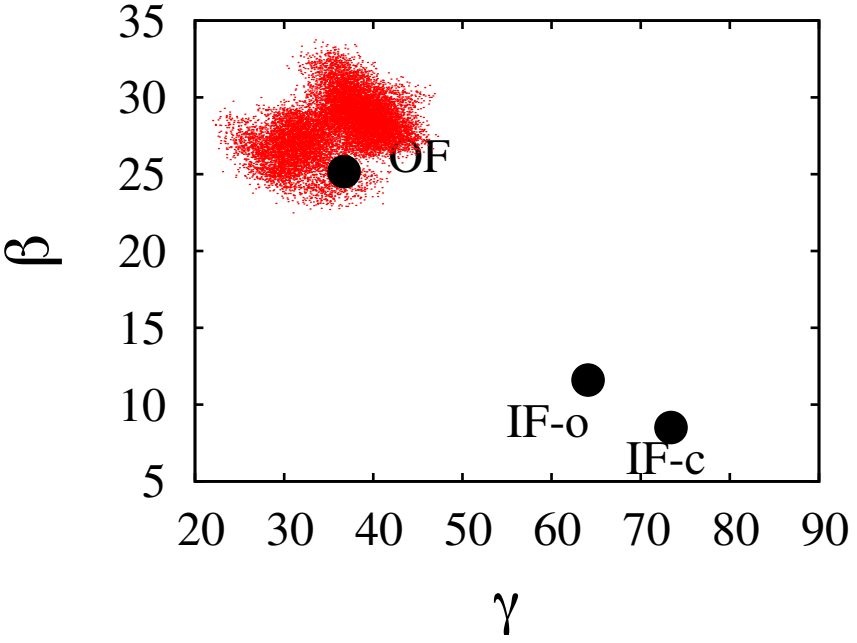
Moradi and Tajkhorshid, **PNAS 110** 18916 (**2013**)

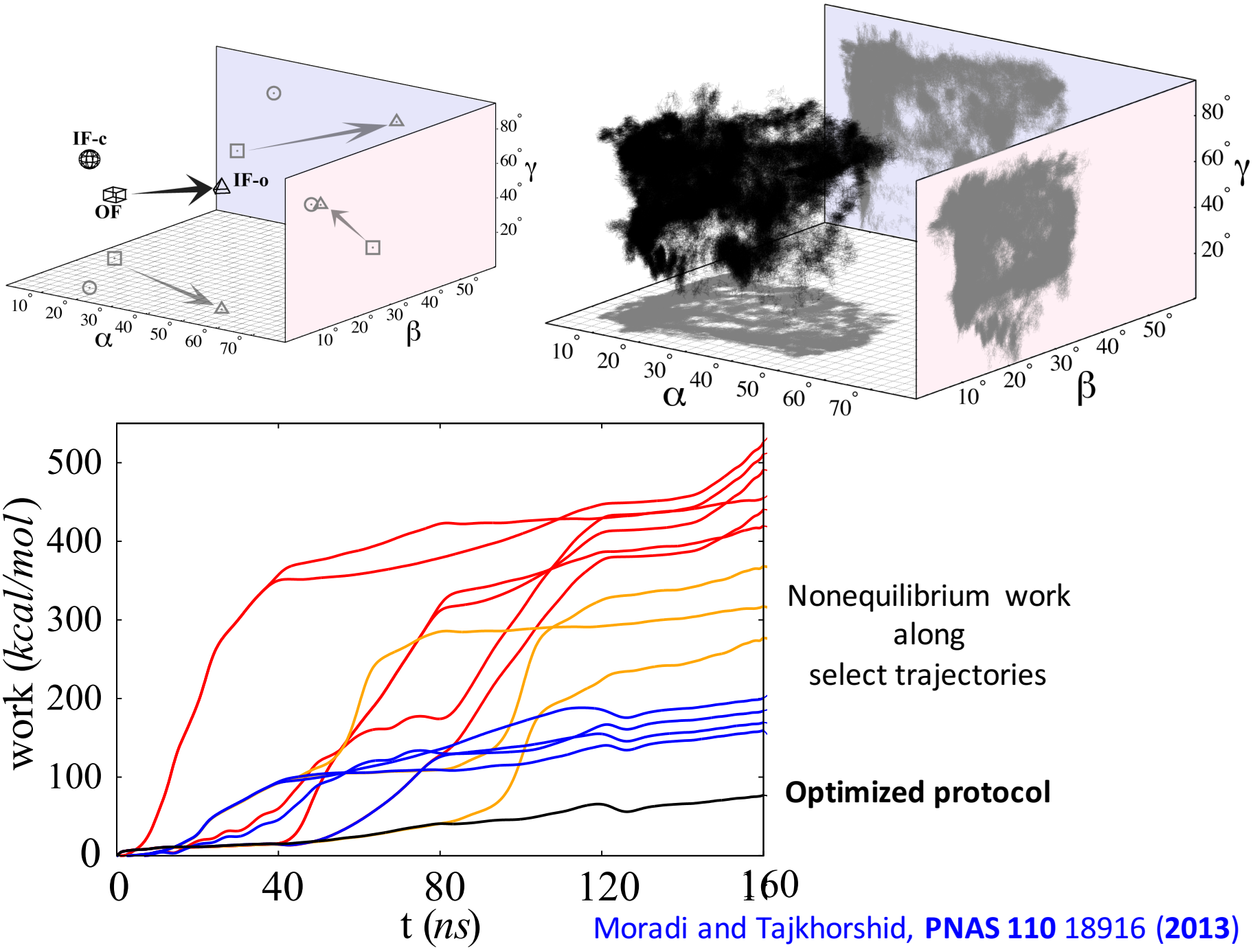
Conventional Equilibrium Molecular Dynamics



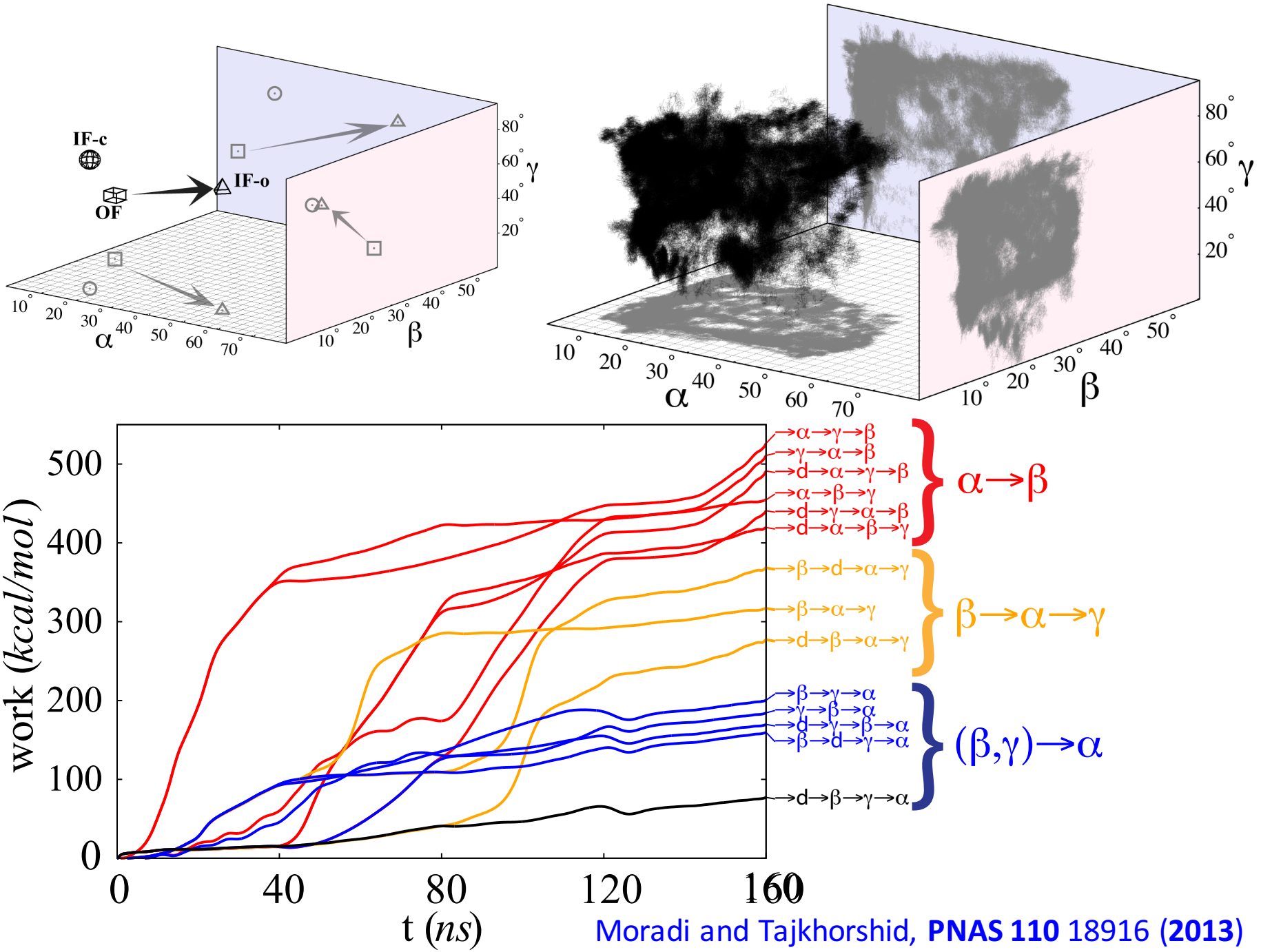
apo MsbA in explicit water/membrane (300 ns)





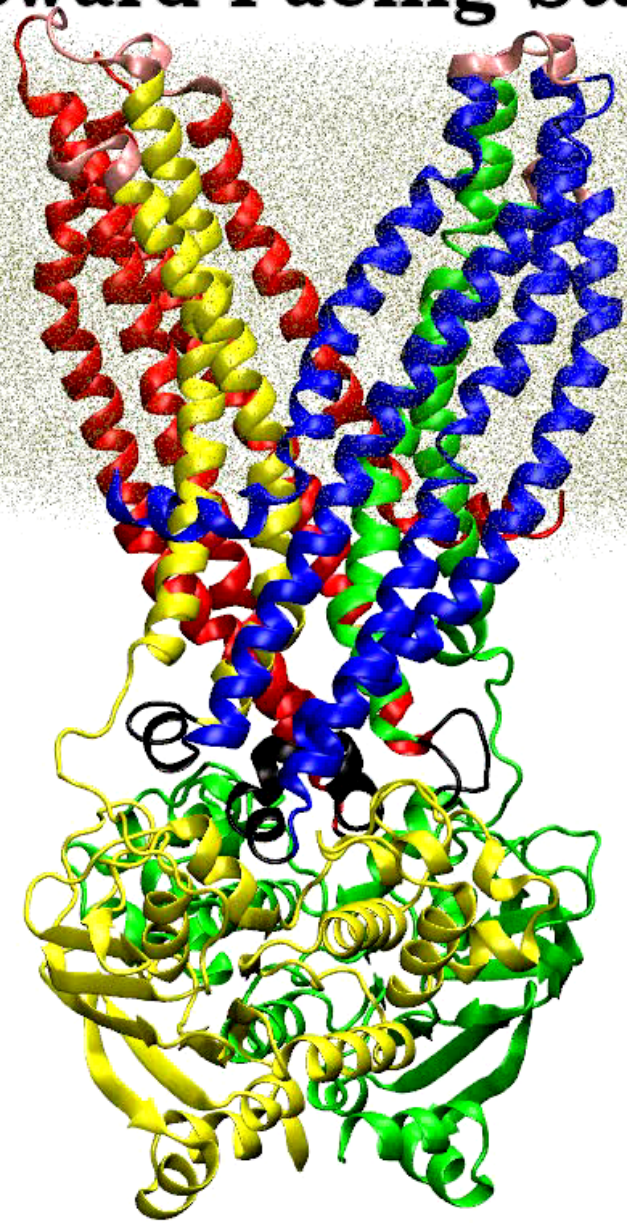


Moradi and Tajkhorshid, **PNAS 110** 18916 (**2013**)



Moradi and Tajkhorshid, **PNAS 110** 18916 (**2013**)

Outward-Facing State



OF → IF

NBD Dissociation

Periplasmic Closure

NBD Twist

Cytoplasmic Opening

IF → OF

Cytoplasmic Closure

NBD Twist

Periplasmic Opening

NBD Dimerization

R T R T R T R T R

T Transition

R Relaxation

Steering along orientation quaternion

To find the best rigid body rotation for $\{\boldsymbol{x}_k\} \rightarrow \{\boldsymbol{y}_k\}$,
find a unit quaternion q to minimize:

$$\left\langle \left\| \hat{q} x_k \hat{q}^* - y_k \right\|^2 \right\rangle$$

It turns out $q = (\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\hat{\mathbf{u}})$

θ and $\hat{\mathbf{u}}$ are the angle and axis of rotation

target quaternion at time t

Biasing Potential:

$$U_B(q_{ref}(\{\mathbf{x}_k\}), t) = \frac{1}{2} k \Omega^2(q_{ref}(\{\mathbf{x}_k\}), Q(t))$$

$$\cos(\Omega(\hat{p}, \hat{q})) = \hat{p} \cdot \hat{q}$$

Interpolation of orientation quaternion in *colvars*

At each time t+Δt:

(1) Linear interpolation:

$$(Q'(t + \Delta t) = Q(t) + \frac{Q(T) - Q(t)}{T - t}\Delta t)$$

(2) Normalization:

$$(Q(t + \Delta t) = \frac{Q'(t+\Delta t)}{||Q'(t+\Delta t)||})$$

Interpolation of orientation quaternion in NAMD

It turns out:

$$\frac{\partial}{\partial t} U_B(q_{ref}, t) = k\,\Omega(q_{ref}, Q(t))\,\frac{\partial}{\partial t}\Omega(q_{ref}, Q(t)),$$

$$= -\frac{\Omega}{\sin(\Omega)}\,(q - Q(t)\cos(\Omega)) \cdot \frac{Q(T) - Q(t)}{T - t}$$

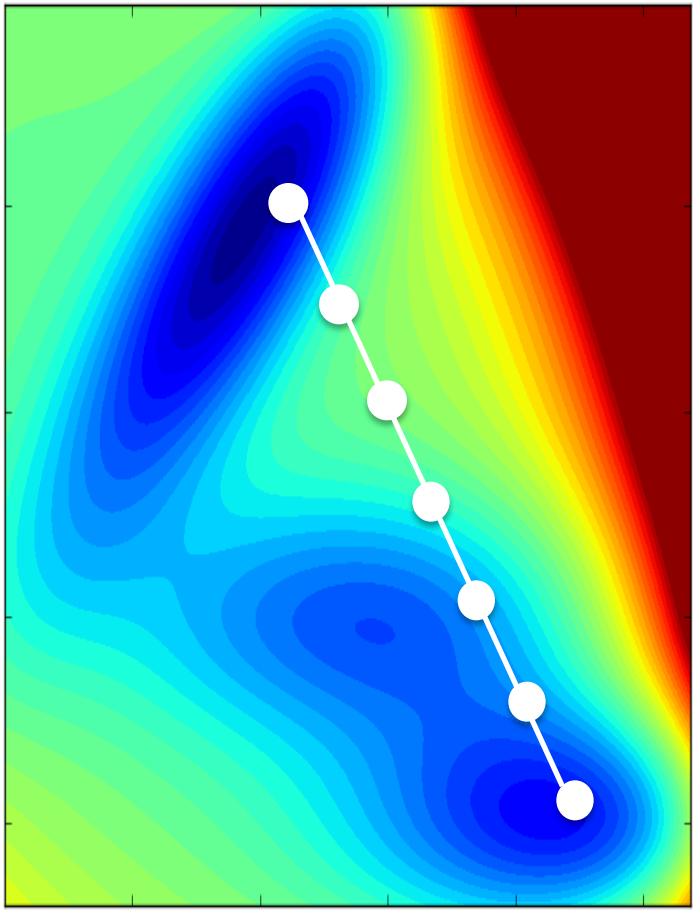
that can be used for work measurements using:

$$w^t = \int_0^t \frac{\partial}{\partial t'} U_B(q_{ref}(\{\mathbf{x}_k\}), t')\,dt'$$

- **Introduction**
  - How to study large-scale conformational changes?
- **Methodology**
  - Empirical search for good pulling protocols
  - **Iterative combination of free energy calculation methods and path-finding algorithms**

Path-Refining Algorithms

- String method (finding approximate minimum free energy pathways on high-dimensional spaces)
  - **A pathway is represented by a "string"**, i.e., an ordered series of images $\{\xi_i\}$ connecting reactant and product regions.
  - **The string is iteratively updated** according to some ``rule'' until converges to a stationary solution.

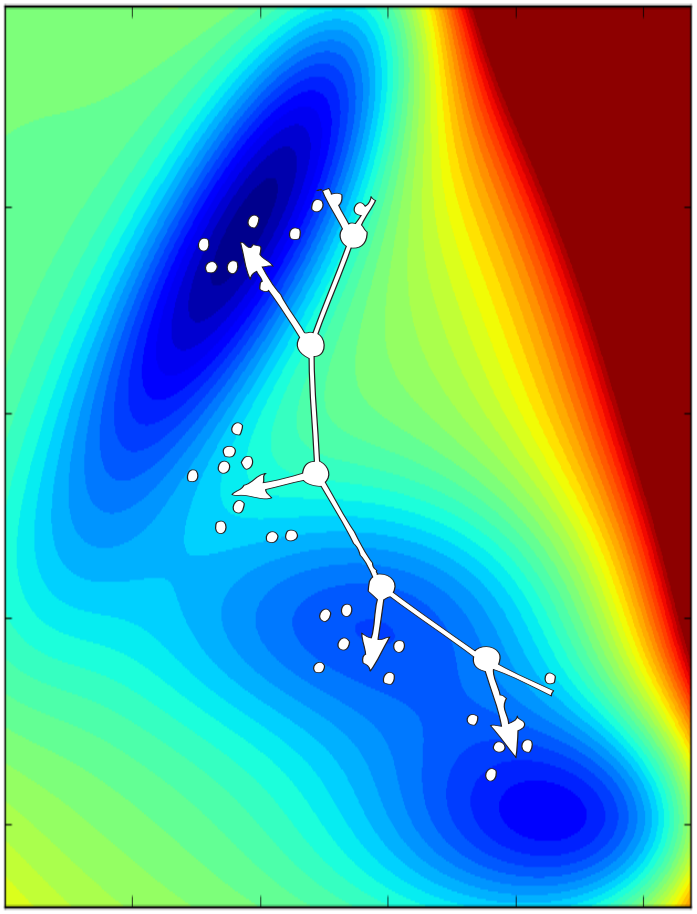


Path-Refining Algorithms

- String method with swarms of trajectories (SMwST):
  - Start from an initial string of N images ($\boldsymbol{\zeta}_i$)
  - Restrain M copies of each image for time Δt

$$U_i(\boldsymbol{\xi}) = \frac{1}{2} k(\boldsymbol{\xi} - \boldsymbol{\zeta}_i)^2$$

  - Release the restraints and run for time $\Delta t'$
  - New string ($\boldsymbol{\zeta}_i$) is determined from $\langle \boldsymbol{\xi} \rangle_i$'s
  - Iterate until converged



Drift term: $-\beta \mathbf{D}.(\boldsymbol{\nabla} \mathrm{G}) + \boldsymbol{\nabla}.\mathbf{D}$

$$\boldsymbol{\xi}(\mathrm{s}) \parallel -\beta \mathbf{D}.(\boldsymbol{\nabla} \mathrm{G}) + \boldsymbol{\nabla}.\mathbf{D}$$

SMwST gives the zero-drift path not the MFEP.

Johnson & Hummer JPCB 2012, 116, 8573.

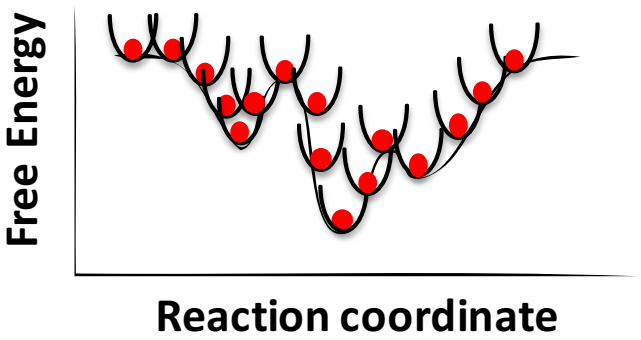
Free Energy Calculations

(Bias-Exchange Umbrella Sampling)

– Umbrella sampling

$$U_i(\xi_i^t) = \tfrac{1}{2}k(\xi_i^t - \xi_i)^2$$

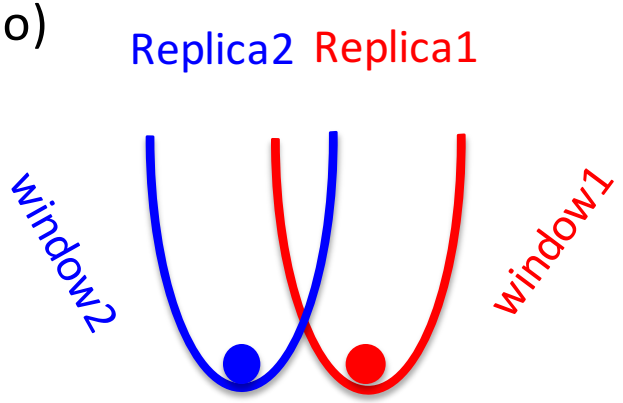
$$e^{-\beta F_i} = \left\langle \frac{e^{-\beta U_i(\xi^t)}}{\sum_j n_j e^{-\beta(U_j(\xi^t)-F_j)}} \right\rangle_{\text{all samples}}$$



– Replica exchange (Monte Carlo)

$$p(\boldsymbol{x}_1\boldsymbol{x}_2 \to \boldsymbol{x}_2\boldsymbol{x}_1) = \min\left(1, \frac{\pi_1(\boldsymbol{x}_2)\pi_2(\boldsymbol{x}_1)}{\pi_1(\boldsymbol{x}_1)\pi_2(\boldsymbol{x}_2)}\right)$$

$$\min\left(1, \frac{e^{-\beta U_1(\xi_2)}e^{-\beta U_2(\xi_1)}}{e^{-\beta U_1(\xi_1)}e^{-\beta U_2(\xi_1)}}\right)$$



Non-Parametric Reweighting

$$V_l(\boldsymbol{x}) = V(\boldsymbol{x}) + U_l(\boldsymbol{x})$$

Biased potential

Unknown

Unbiased potential

Unknown

Biasing potential

Known

$$g_l(\boldsymbol{x}) \propto g(\boldsymbol{x})\, q_l(\boldsymbol{x})$$

Biased density

Known

Unbiased density

Unknown

Biasing factor

Known

Partition function

Unknown

$$Z_l^{-1} = e^{f_l}$$

$$q_l(\boldsymbol{x}) = e^{-U_l(\boldsymbol{x})}$$

$$Z_l = \int \boldsymbol{dx}\, g(\boldsymbol{x}) q_l(\boldsymbol{x})$$

Conventional WHAM

If the $\boldsymbol{x}$ space is discrete or "binable" such that

$$U_l(\boldsymbol{x}) \approx U_l(\boldsymbol{x}^k)$$

e.g., U_l is a smooth function of a 1D coordinate

$$U_l(\boldsymbol{x}) \approx U_l(\zeta),$$

$$k \equiv \left[\frac{\zeta - \zeta_0}{\Delta\zeta}\right]$$

$$q_l^k \equiv q_l(\boldsymbol{x}^k)$$

$$g^k = \frac{H^k}{\sum_l N_l Z_l^{-1} q_l^k}$$

$$H^k \equiv \sum_l n_l^k \,, N_l \equiv \sum_k n_l^k$$

$$Z_l = \sum_k g^k q_l^k$$

S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *"The weighted histogram analysis method for free-energy calculations on biomolecules."* *J. Comput. Chem.* ***13***, 1011 (1992)

Generalizations

$$\sum_{k=1}^{K} F^k \equiv \sum_{l} \sum_{t=1}^{N_l} F \left(\boldsymbol{x}_l(t)\right)$$

Every conformation sampled is a state.

$$Z_l = \sum_k g^k q_l^k$$

$$g^k = 1 \Big/ \sum_l N_l Z_l^{-1} q_l^k$$

Solved iteratively.

C. Bartels, "*Analyzing biased Monte Carlo and molecular dynamics simulations.*" *Chem. Phys. Letters* ***331***, 446 (2000)

Generalizations

$$Z_l = \sum_k g^k q_l^k$$

$$g^k = {}^1/_{\sum_l N_l Z_l{}^{-1} q_l^k}$$

$$Z_l = \sum_k \frac{q_l^k}{\sum_m N_m Z_m{}^{-1} q_m^k}$$

$$Z_l = e^{-f_l} \qquad f_l = -log \sum_k \frac{q_l^k}{\sum_m N_m e^{f_m} q_m^k}$$

Multi-state BAR (MBAR) equation

M. R. Shirts, J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states."
J. Chem. Phys., ***129***, 124105 (2008)

Combining path-finding and free energy methods

- Potential of Mean Force:

$$G(\boldsymbol{\zeta}) = -\beta^{-1}\log\langle\delta(\boldsymbol{\xi}(\boldsymbol{x}) - \boldsymbol{\zeta})\rangle$$

$$\langle\delta(\boldsymbol{\xi}(\boldsymbol{x}) - \boldsymbol{\zeta})\rangle = \int \delta(\boldsymbol{\xi}(\boldsymbol{x}) - \boldsymbol{\zeta})\rho(\boldsymbol{x},\boldsymbol{p})\, d^{3N}x d^{3N}p$$

- Perturbed Free Energy:

$$F_i = F(\boldsymbol{\zeta}_i) = -\beta^{-1}\log Z_i$$

$$Z_i = \int e^{-\beta U_i(\boldsymbol{\xi})}\rho(\boldsymbol{x},\boldsymbol{p})\, d^{3N}x d^{3N}p = \int e^{-\beta(G(\boldsymbol{\xi})+U_i(\boldsymbol{\xi}))}\, d^{n}\xi$$

- Non-parametric MLE estimates:

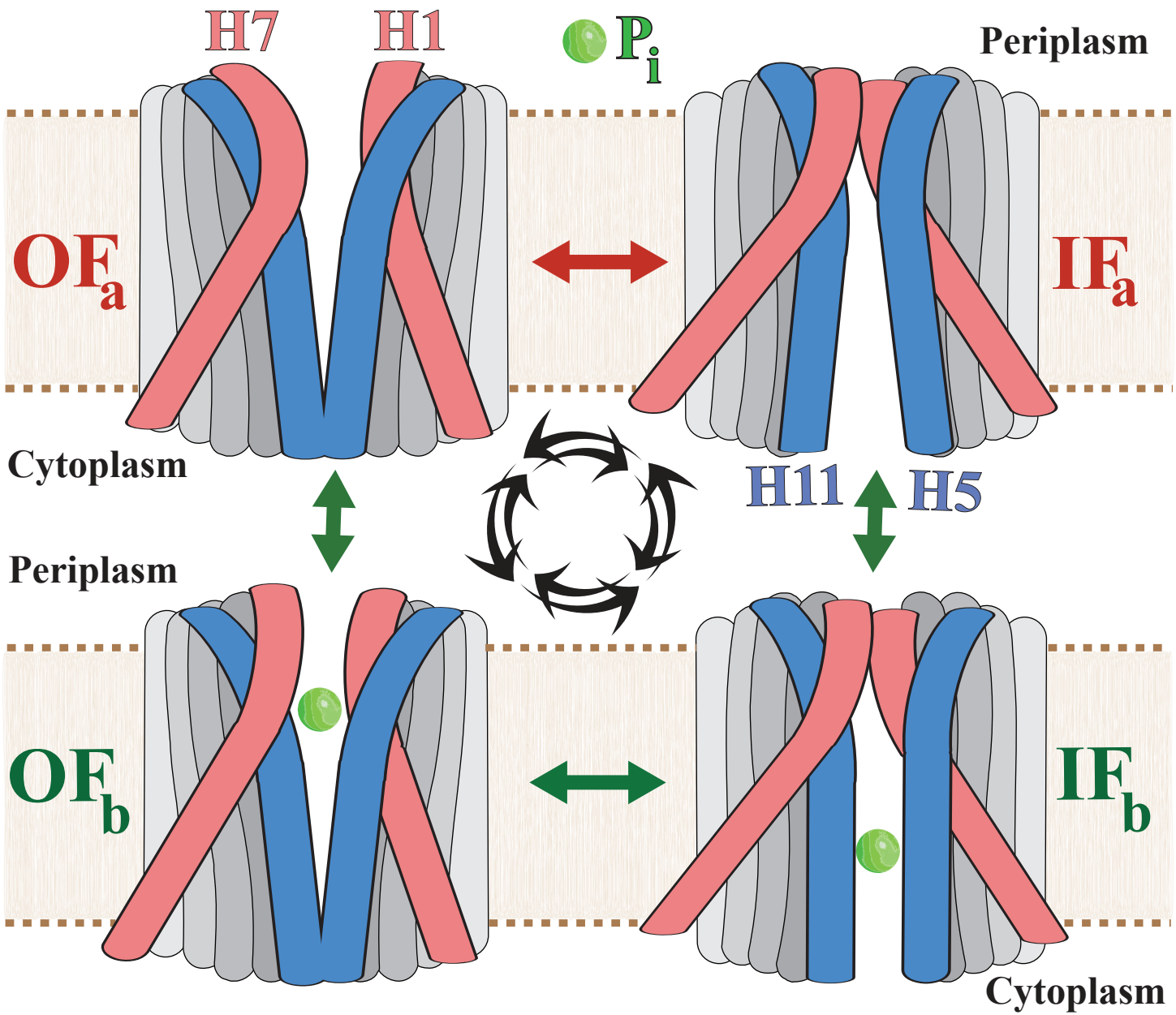
Shirts, Chodera, JCP, ***129***, 124105 (2008)

$$e^{-\beta F_i} = \left\langle \frac{e^{-\beta U_i(\xi^t)}}{\sum_j n_j e^{-\beta(U_j(\xi^t)-F_j)}} \right\rangle_{\text{all samples}}$$

Bartels, CPL, ***331***, 446 (2000)

$$e^{\beta F_i} = n_i \Big/ \sum_t w^t e^{-\beta U_i(\xi^t)}$$

$$w^t = 1 \Big/ \sum_i e^{-\beta(U_i(\xi^t)-F_i)}$$



Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

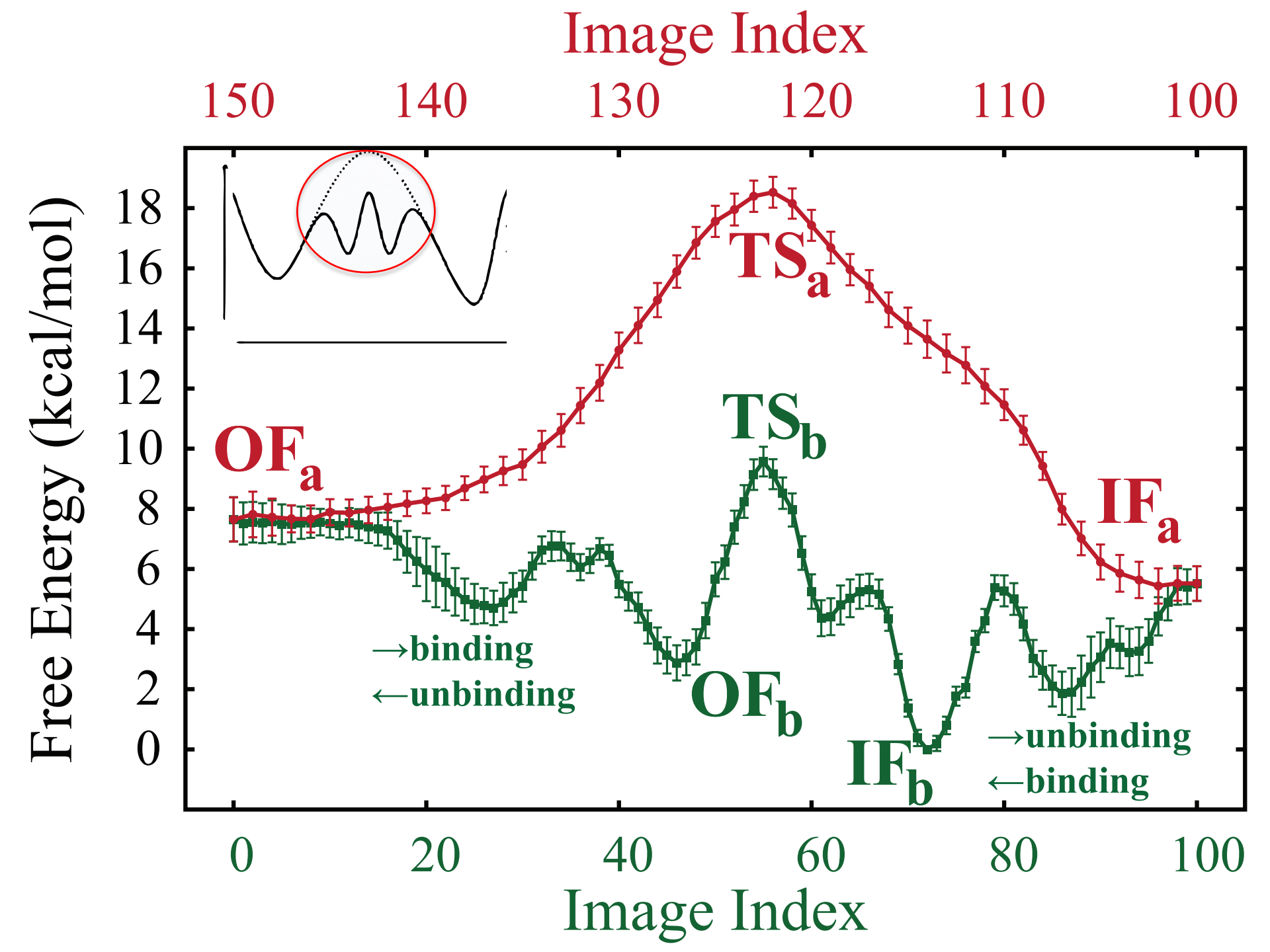
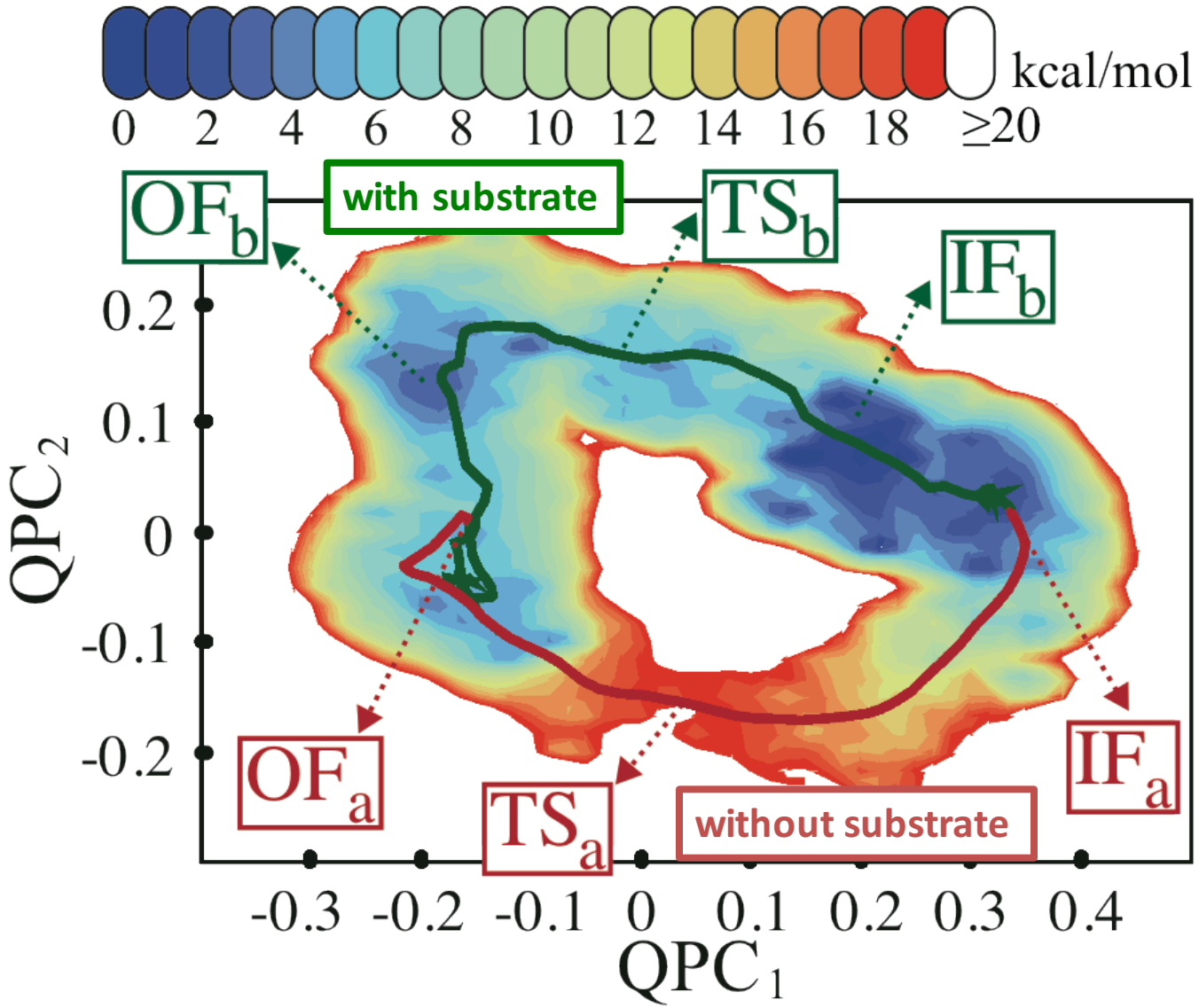


Image Index
150
140
130
120
110
100
Free Energy (kcal/mol)
18
16
14
12
10
8
6
4
2
0
TS$_a$
TS$_b$
OF$_a$
IF$_a$
→binding
←unbinding
OF$_b$
IF$_b$
→unbinding
←binding
0
20
40
60
80
100
Image Index

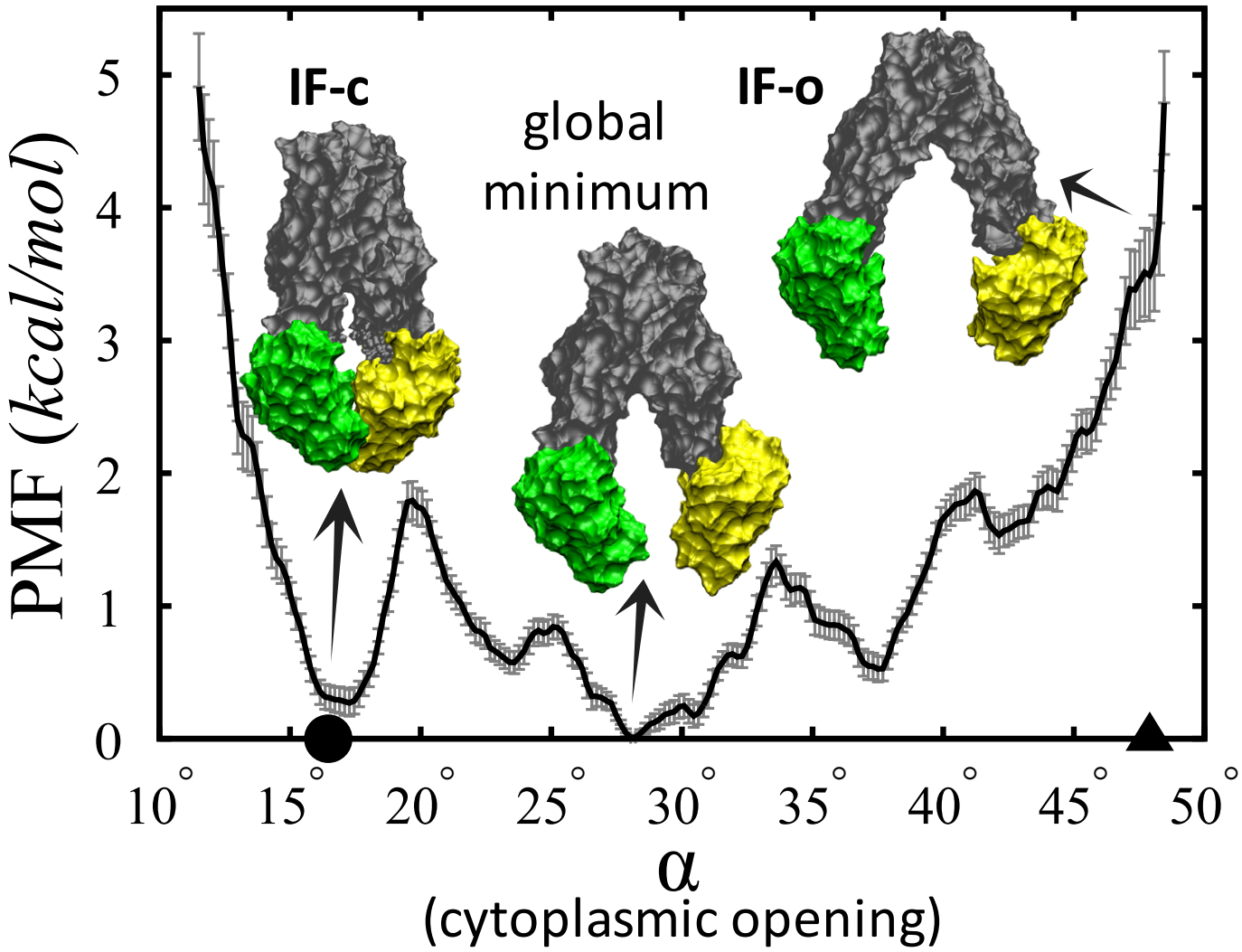
Distinct conformational transition pathways

Quaternion-based principal components (**QPCs**) represent different modes of concerted motions of transmembrane helices.

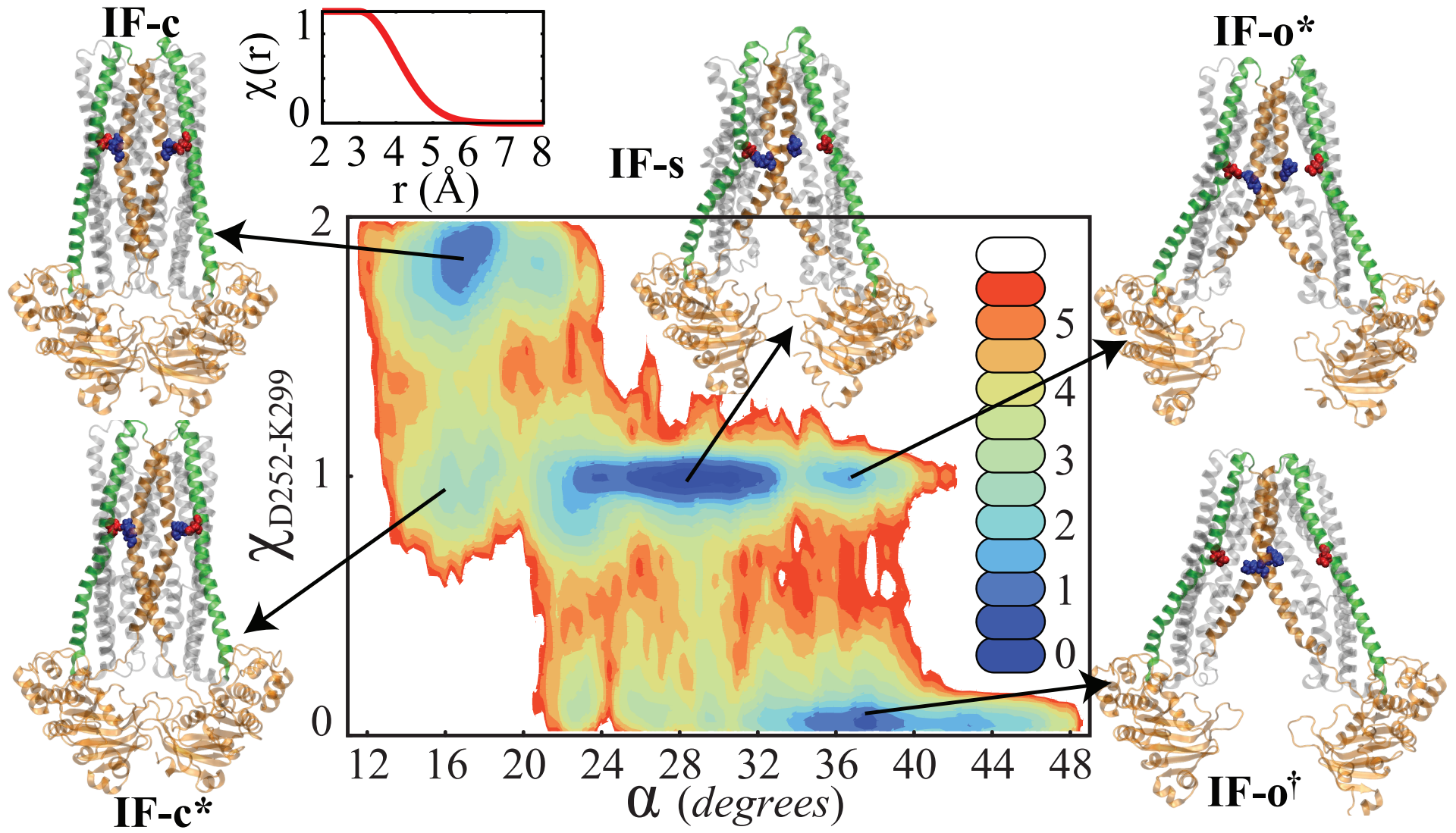


Free Energy Profile

(along one of the four stages of IF-OF transition)

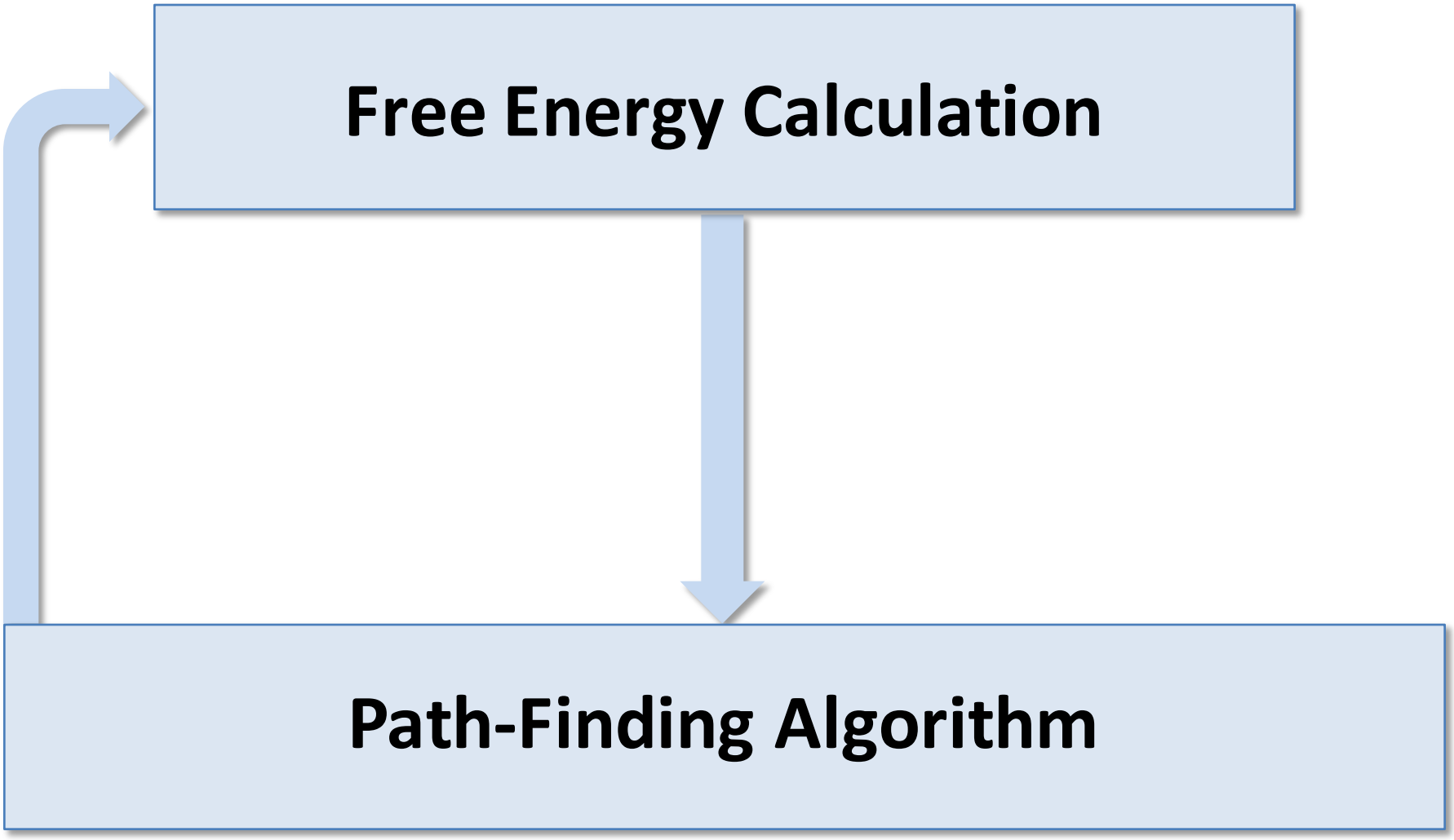


Asymmetric behavior of a homodimer?

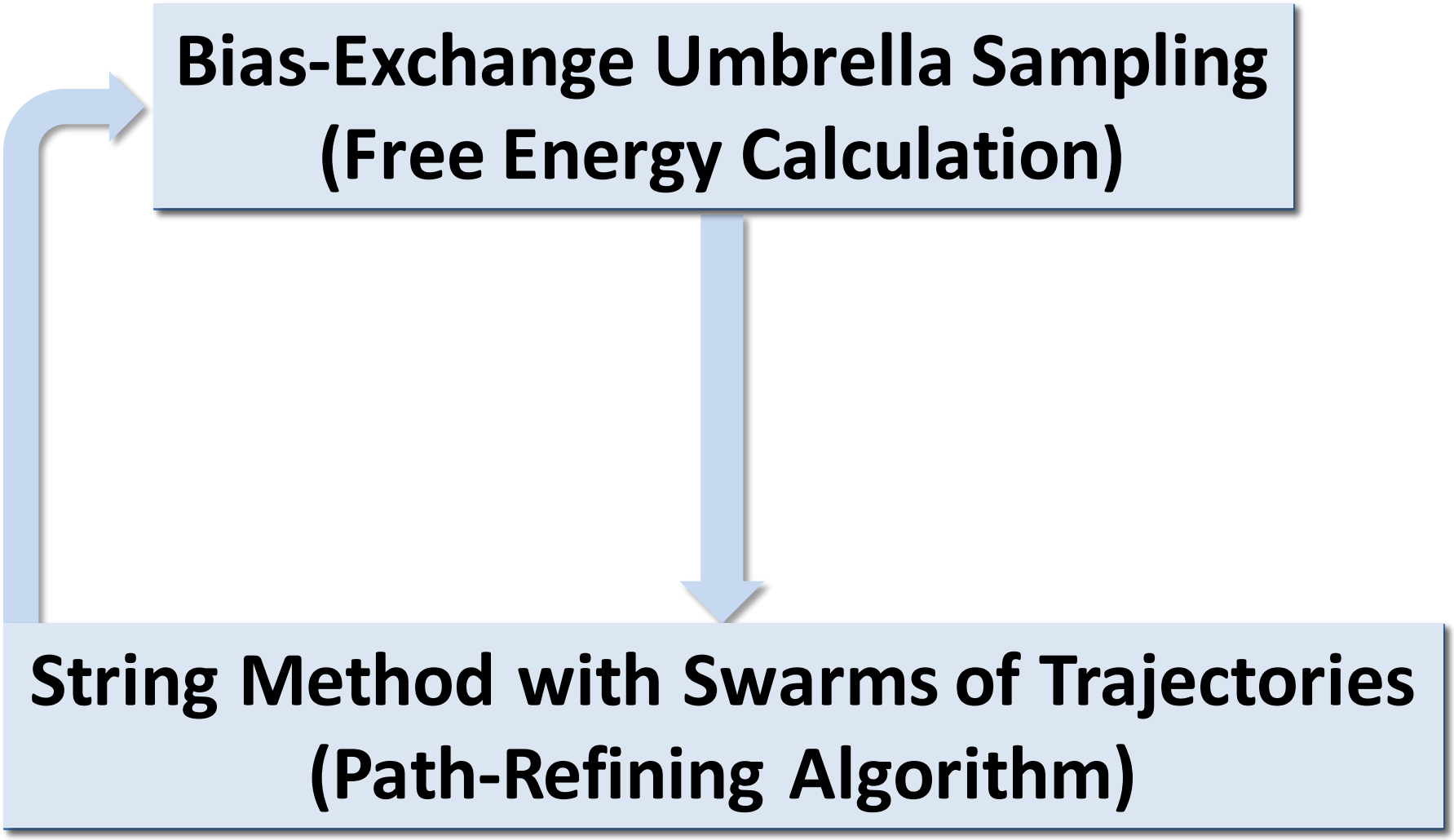


Moradi and Tajkhorshid, **JCTC 10** 2866 (**2014**)

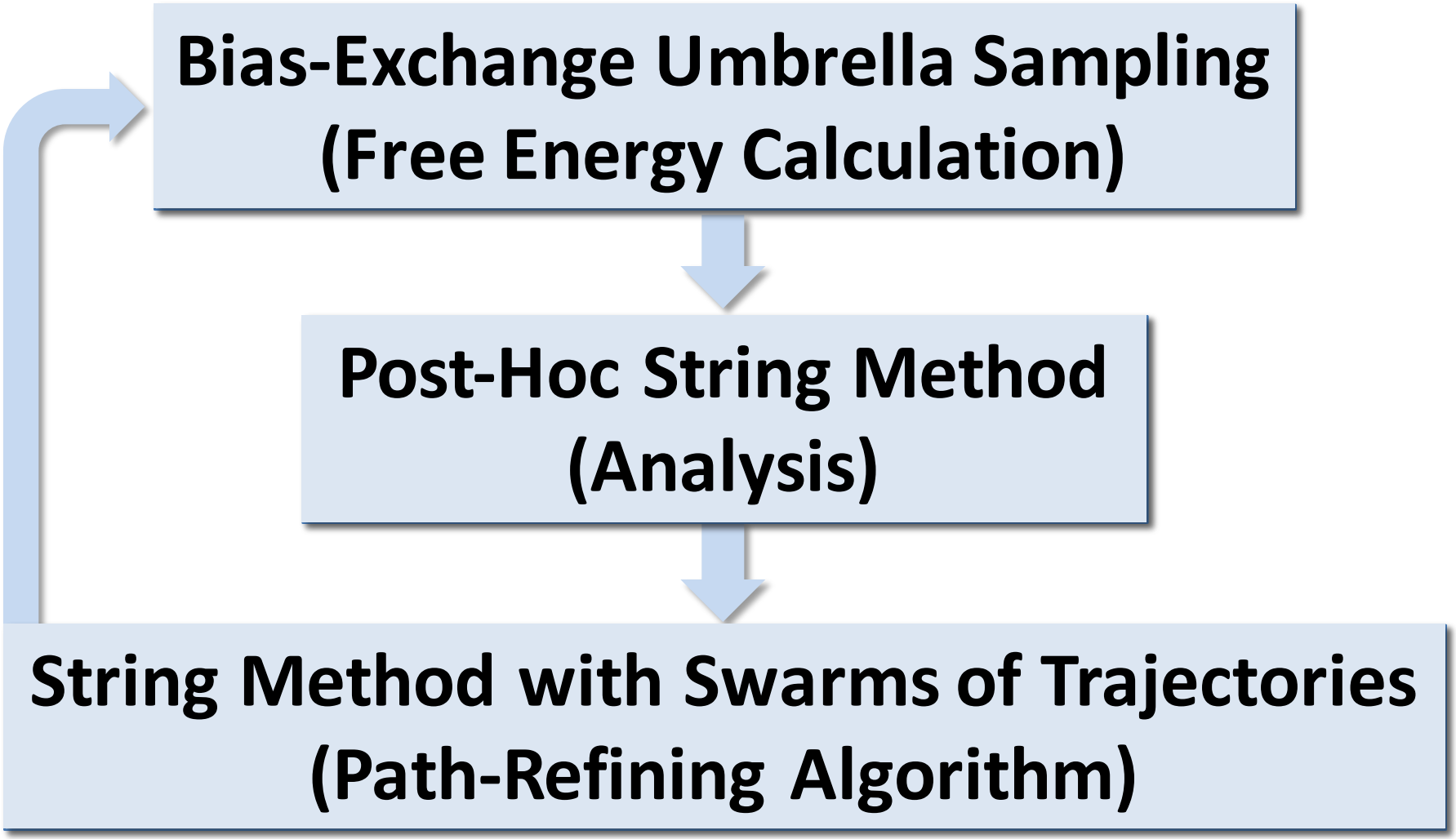
Iterative path-finding algorithms and free energy calculations



Iterative path-refining algorithms and free energy calculations



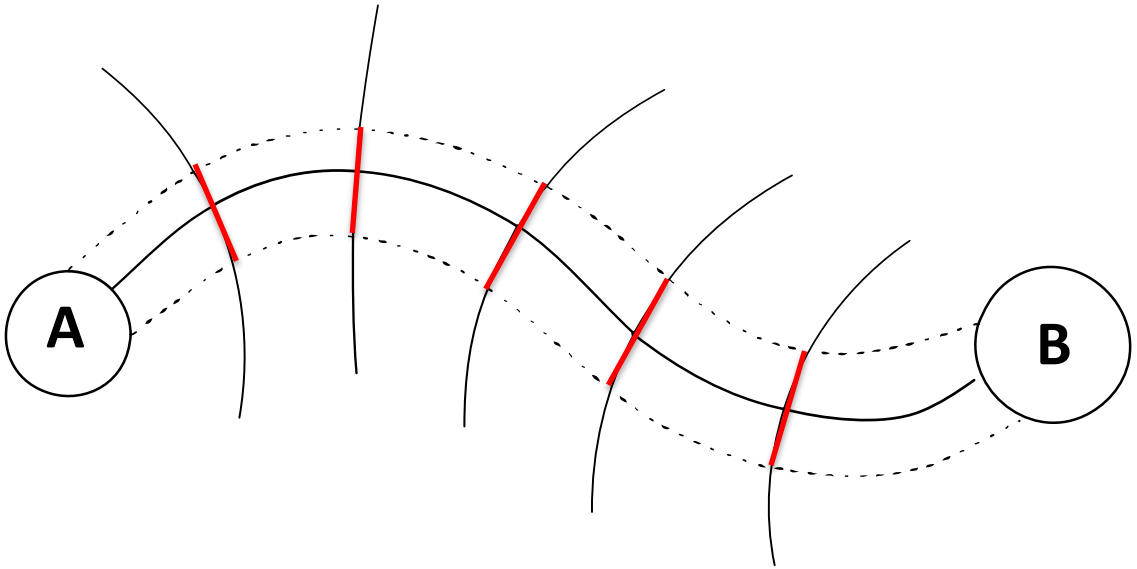
Iterative path-refining algorithms and free energy calculations



Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

Post-hoc string method (PHSM)

- Suppose we have already sampled a particular **"continuous"** region of configuration space (or some multi-dimensional collective variable space $\{\xi_i\}$) and estimated the **weight** of each sample $\{w^t\}$.
- PHSM finds the **principal curve** in the ξ space using available samples as an **approximate minimum free energy pathway**.



Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

Post-hoc string method (PHSM)

- Suppose we have already sampled a particular **"continuous"** region of configuration space (or some multi-dimensional collective variable space $\{\xi_i\}$) and estimated the **weight** of each sample $\{w^t\}$.
- PHSM finds the **principal curve** in the ξ space using available samples as an **approximate minimum free energy pathway**.

- For each image i with center ξ_i, find all samples that are closer to $\{\xi_i\}$than any other image center (Voronoi tessellation):

$$B_i = \left\{ \xi^t \middle| \quad |\xi^t - \xi_i| \; < \; |\xi^t - \xi_j| \;\; \forall j \neq i \;\; \& \;\; |\xi^t - \xi_i| \; < \; \varepsilon \right\}$$

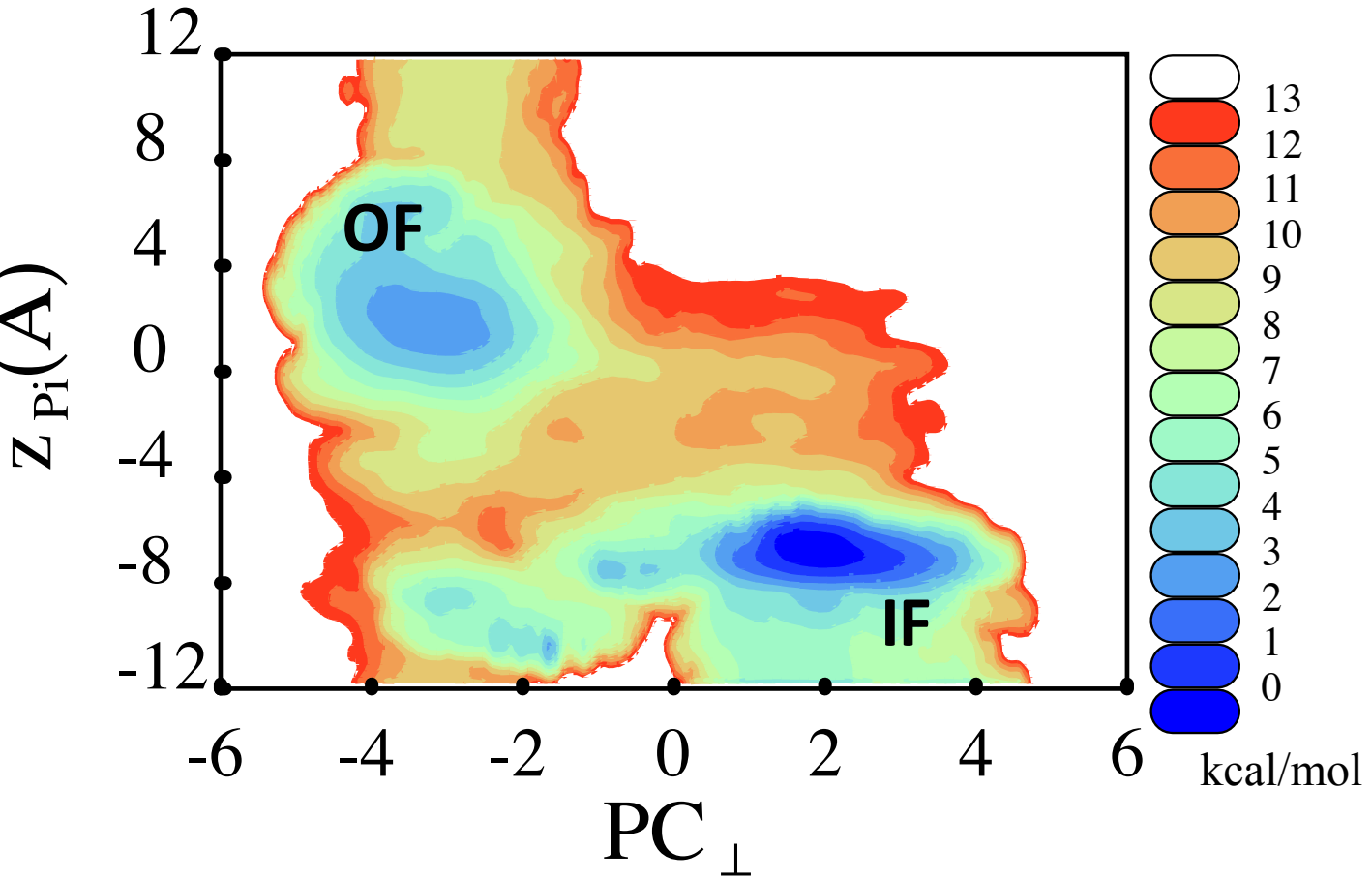
- Find the new image center by weighted averaging over all the samples in the Voronoi cell:
- Smooth and reparametrize.

$$\xi_i = \left\langle w^t \xi^t \chi_i^t \right\rangle \Big/ \left\langle w^t \chi_i^t \right\rangle$$

indicator function

Post-hoc string method (PHSM)

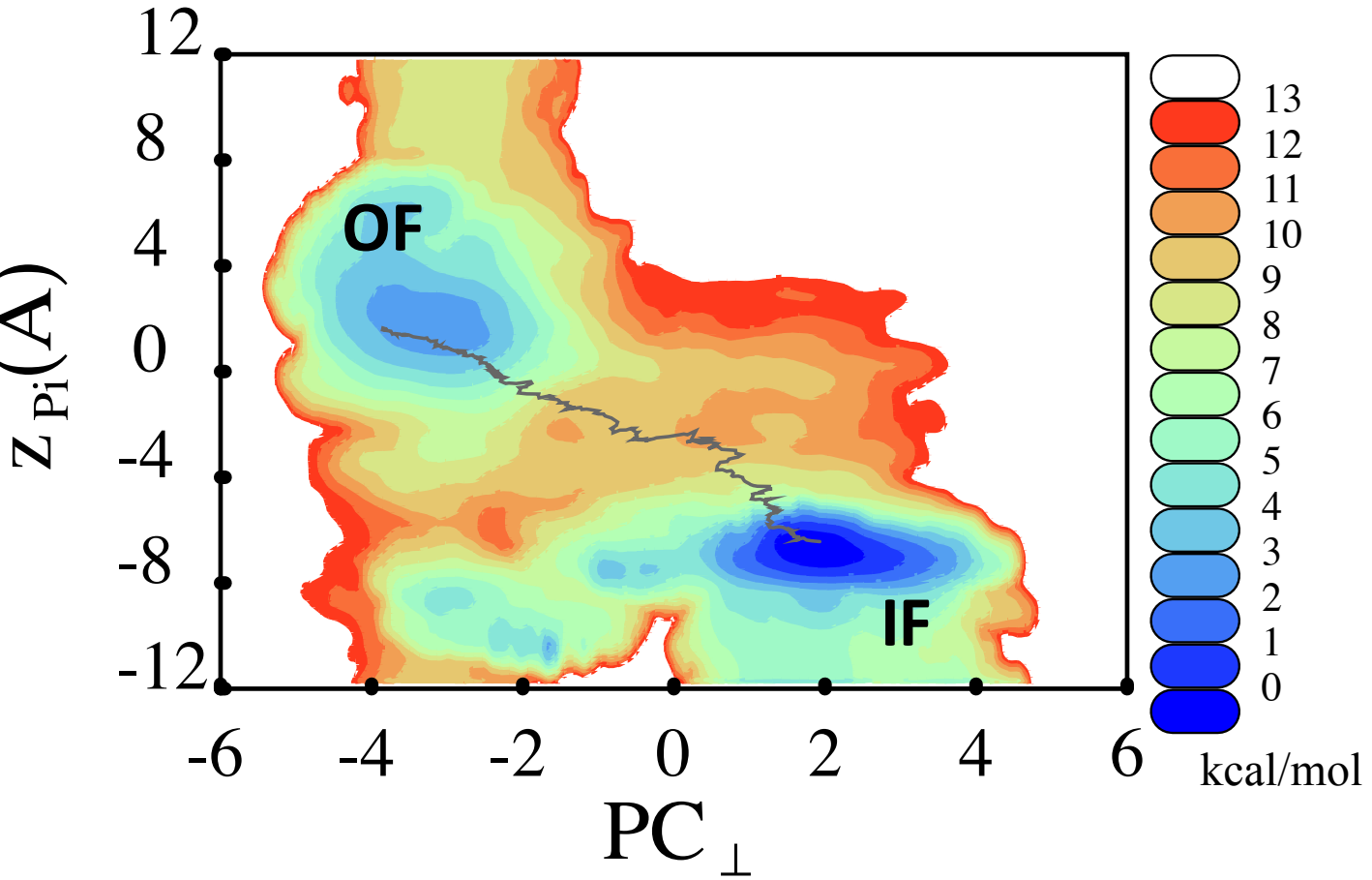
- Free energy landscape in a given 2D collective variable space
- Note: The algorithm is non-parametric, i.e., the pathway is not optimized in a particular low-dimensional space.



Moradi, Enkavi, and Tajkhorshid, **Nature Communications** 6 8393 (**2015**)

Post-hoc string method (PHSM)

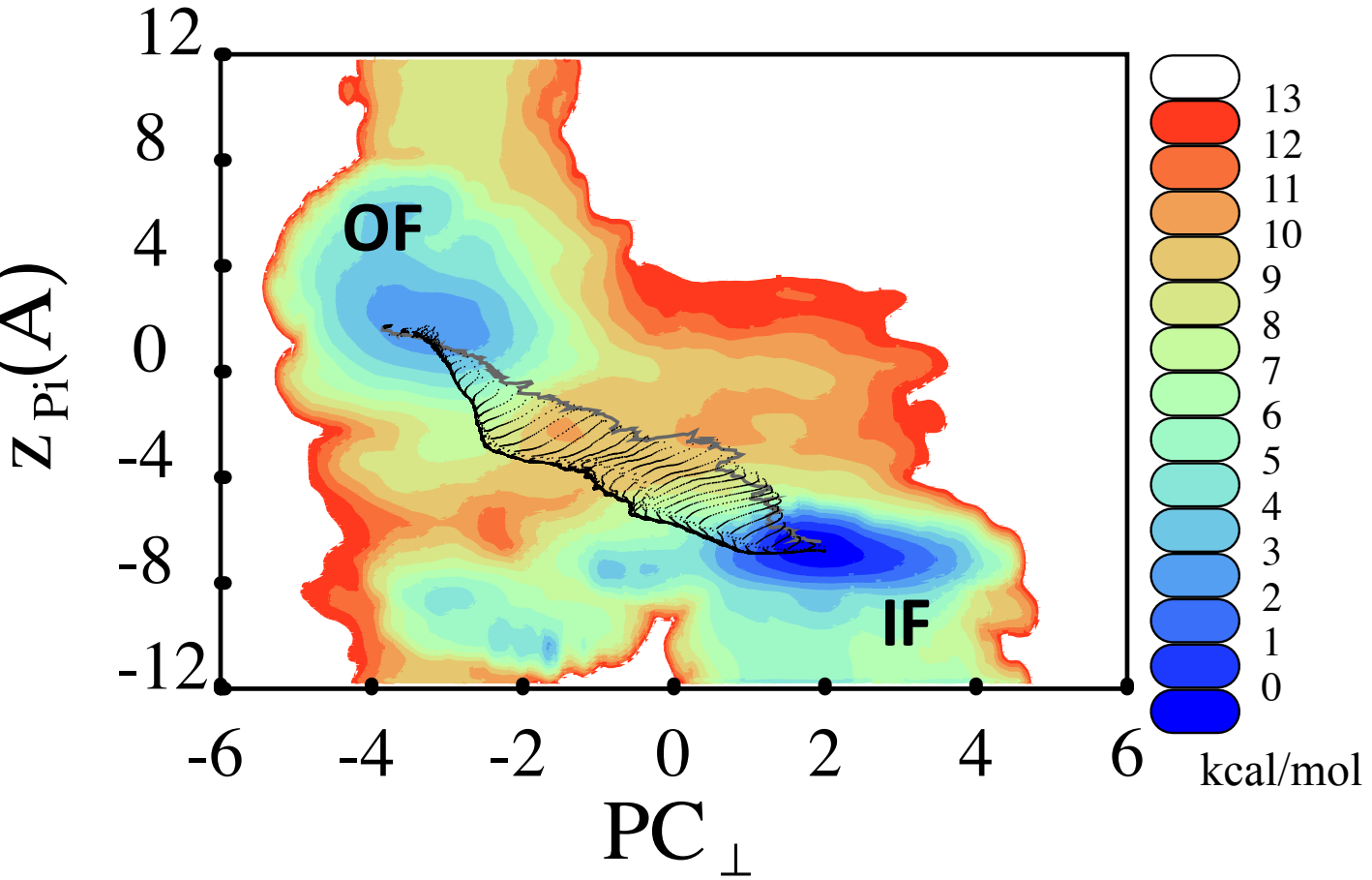
- Free energy landscape in a given 2D collective variable space
- Note: The algorithm is non-parametric, i.e., the pathway is not optimized in a particular low-dimensional space.



Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

Post-hoc string method (PHSM)

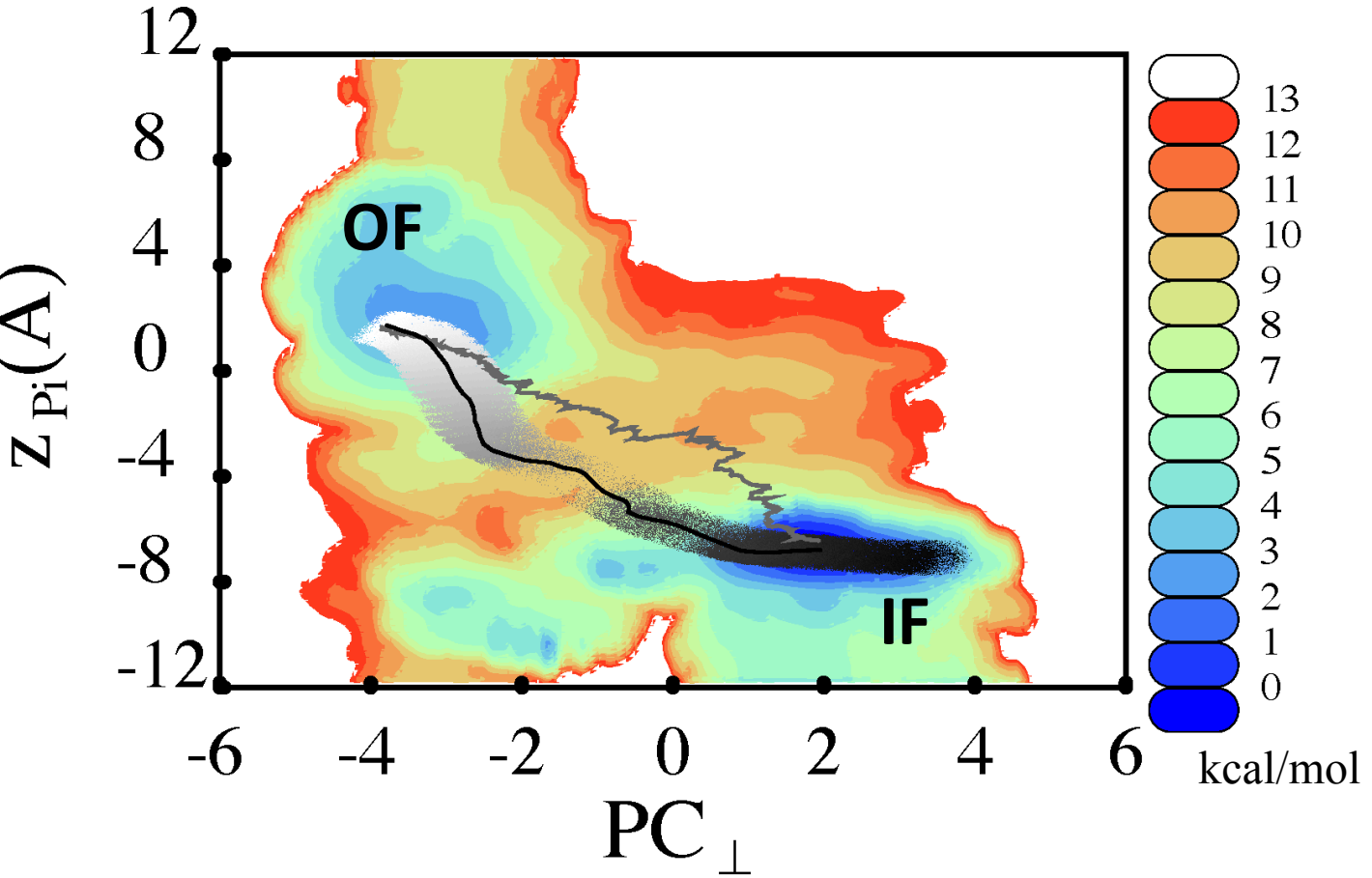
- Free energy landscape in a given 2D collective variable space
- Note: The algorithm is non-parametric, i.e., the pathway is not optimized in a particular low-dimensional space.



Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

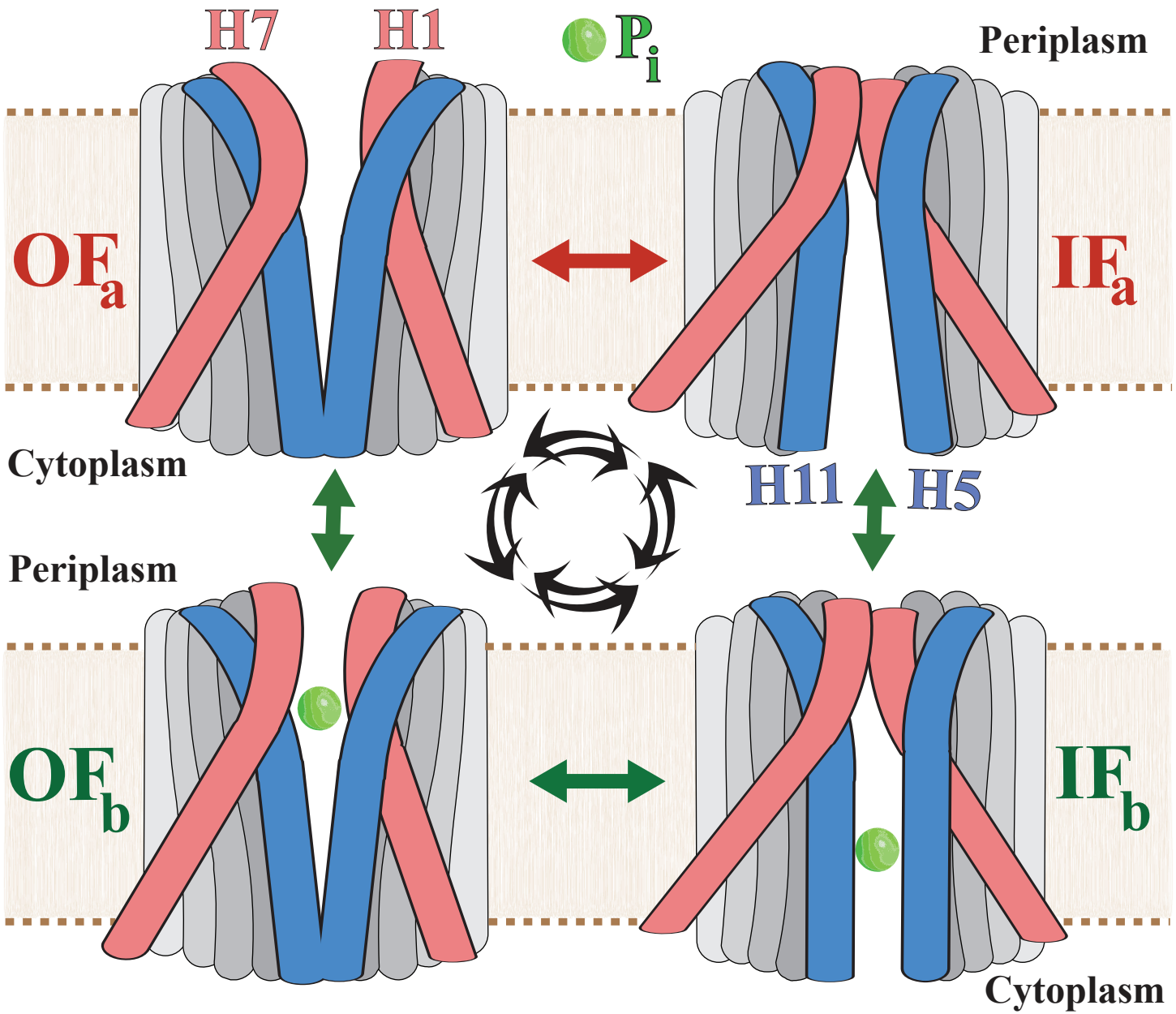
Post-hoc string method (PHSM)

- Free energy landscape in a given 2D collective variable space
- Note: The algorithm is non-parametric, i.e., the pathway is not optimized in a particular low-dimensional space.



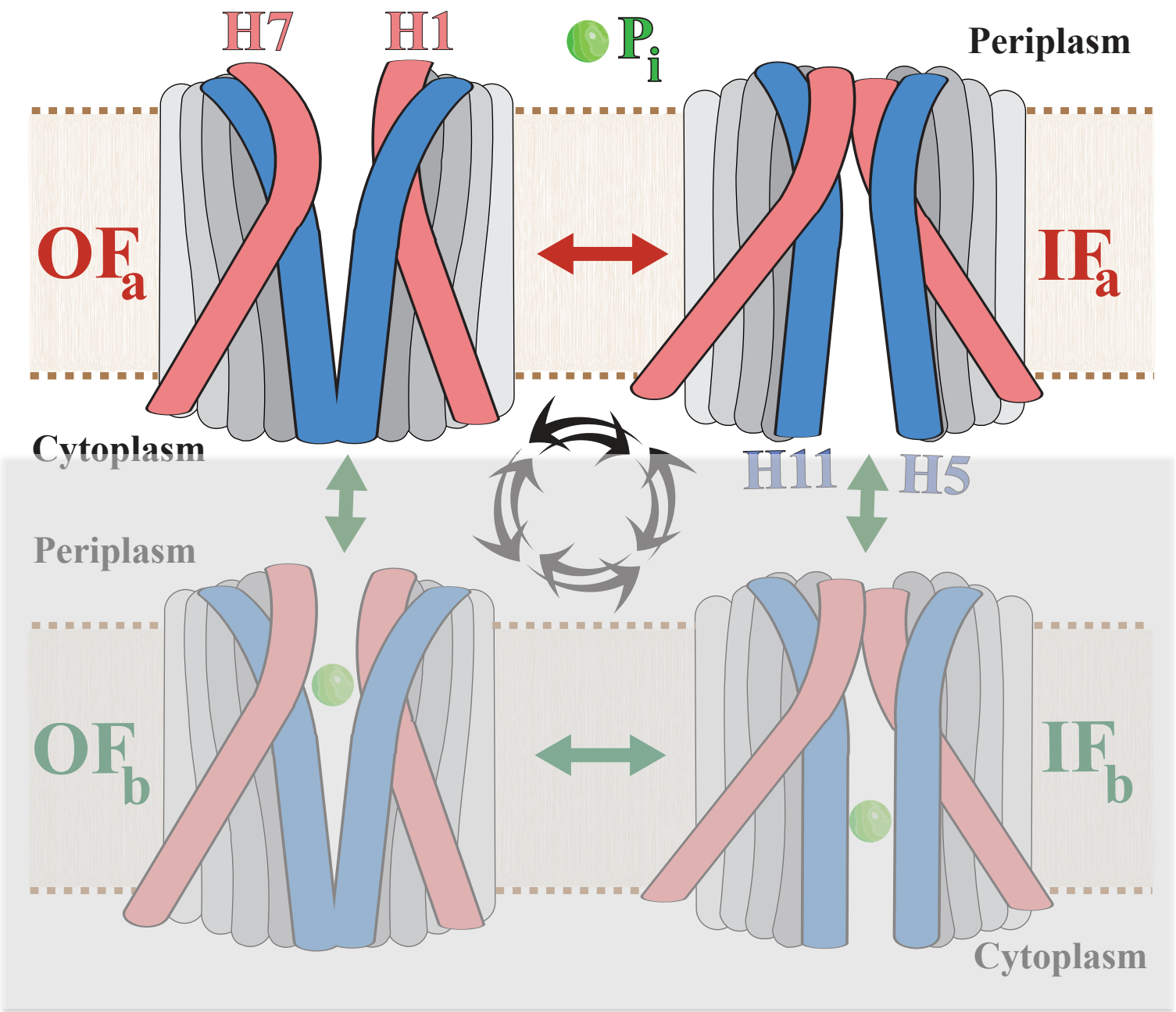
Moradi, Enkavi, and Tajkhorshid, **Nature Communications** 6 8393 (**2015**)

Example: Transport Cycle of GlpT

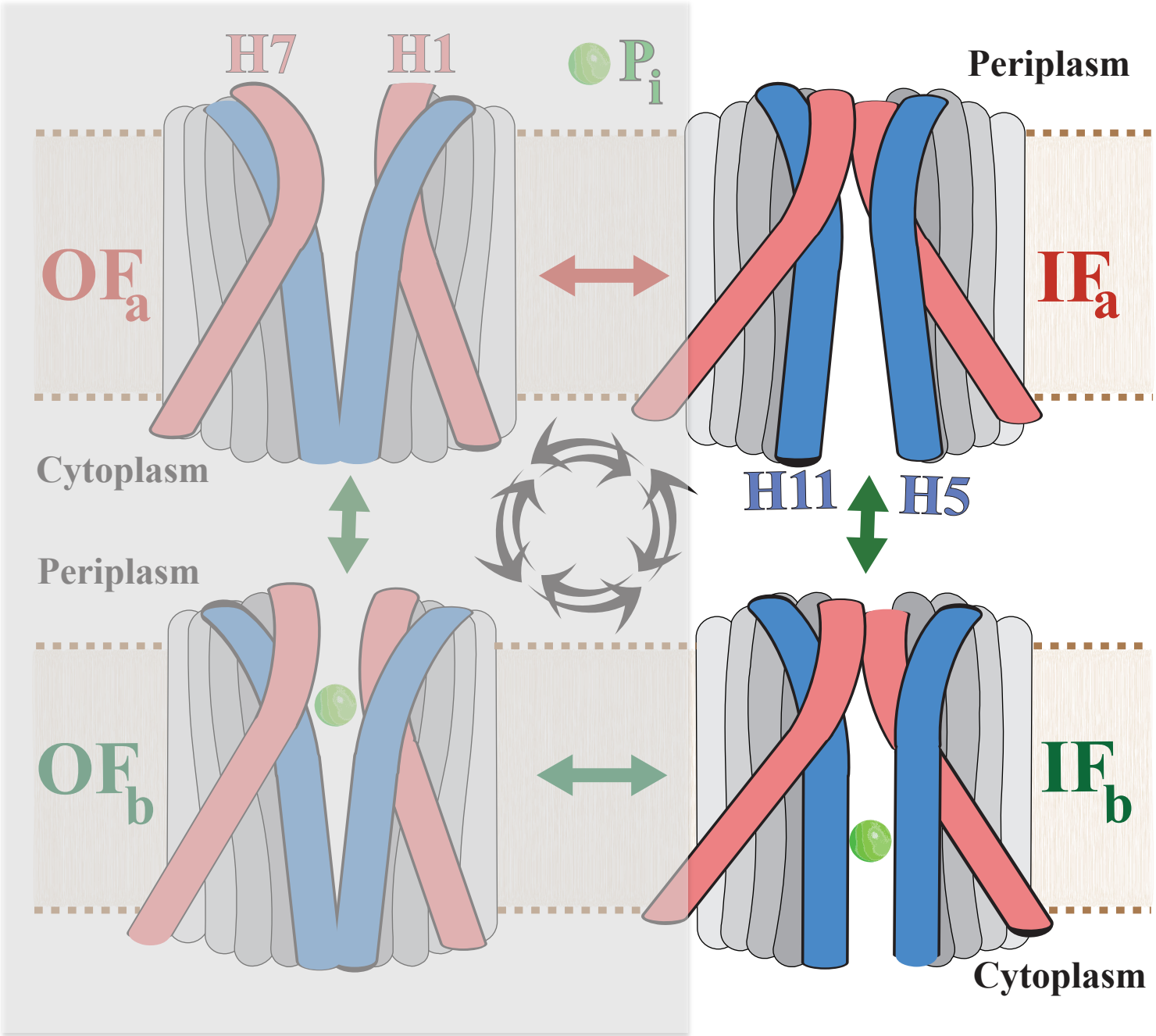


Moradi M., Enkavi G., and Tajkhorshid E., **Nature Communications 6** 8393 (**2015**)

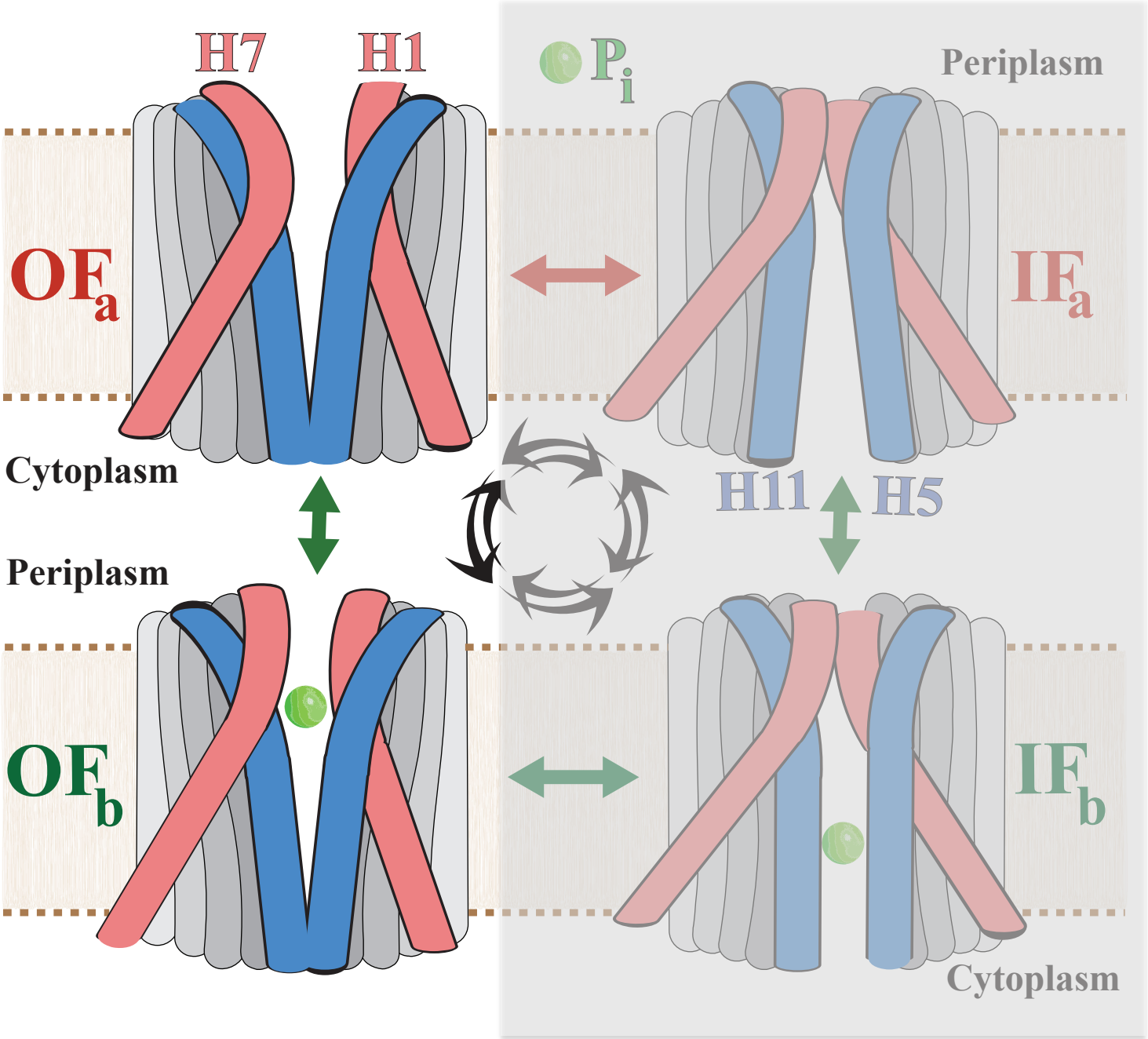
Step 1: $OF_a \leftrightarrow IF_a$



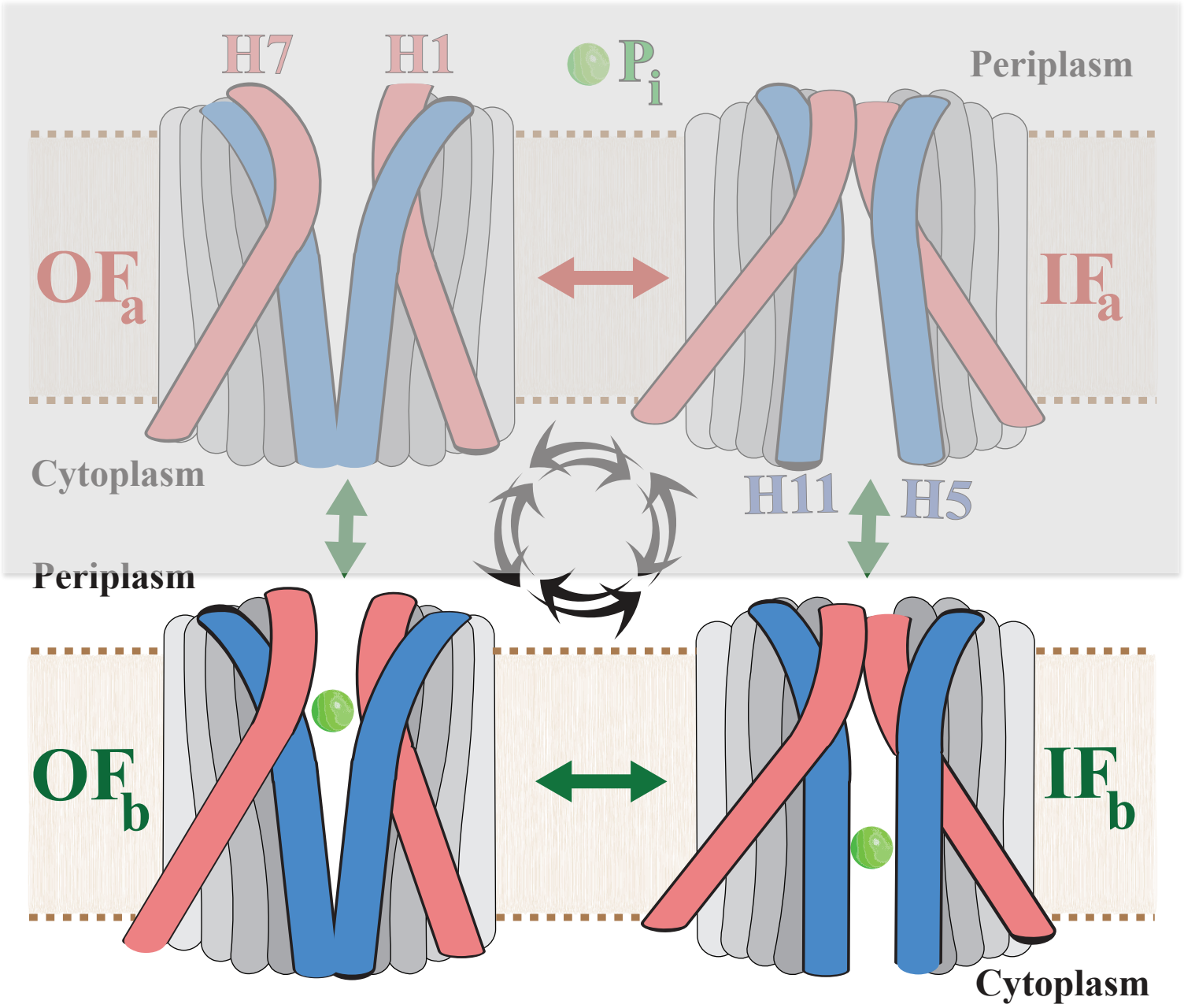
Step 2: $IF_a \leftrightarrow IF_b$

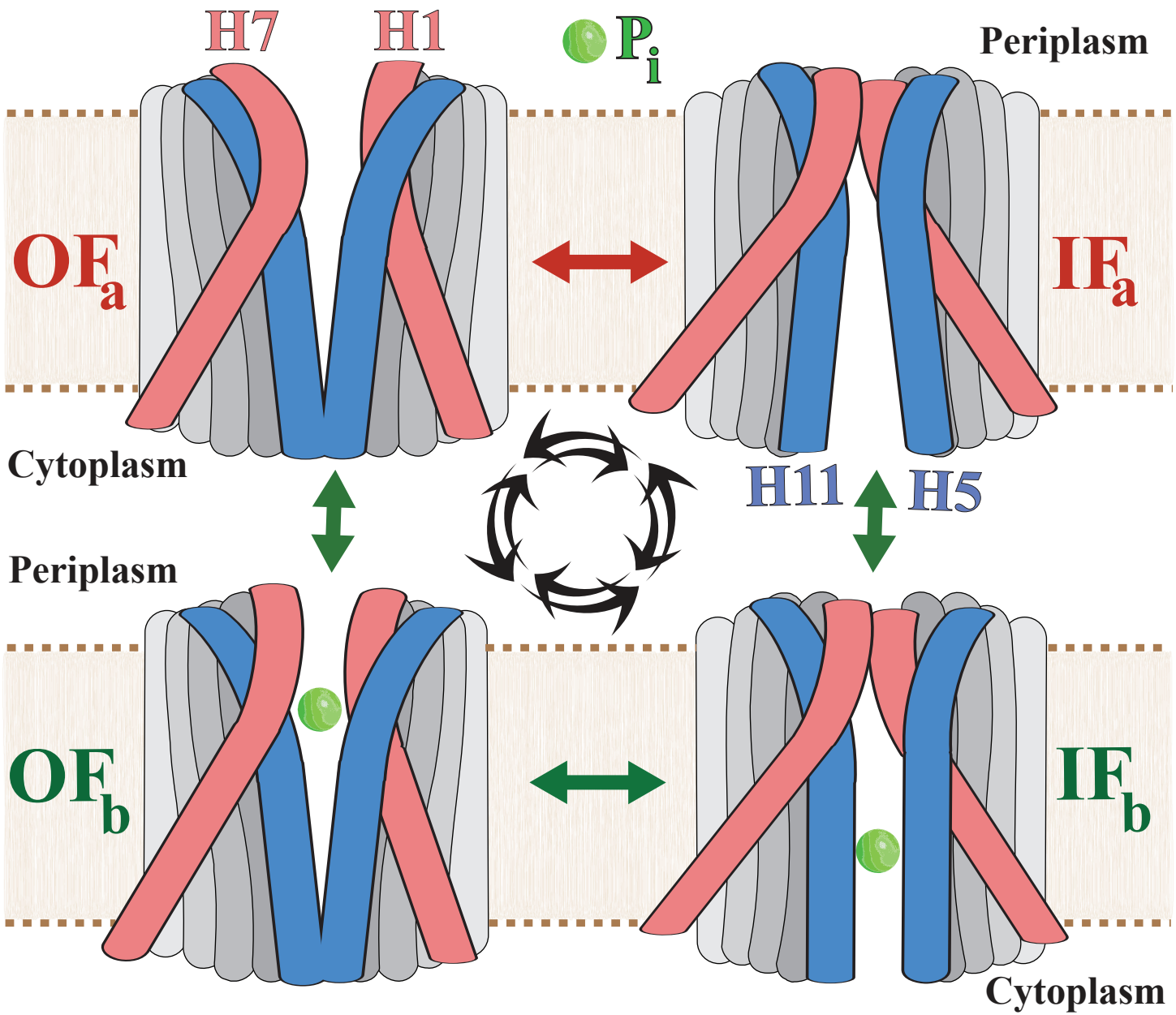


Step 3: $OF_a \leftrightarrow OF_b$



Step 4: $OF_b \leftrightarrow IF_b$

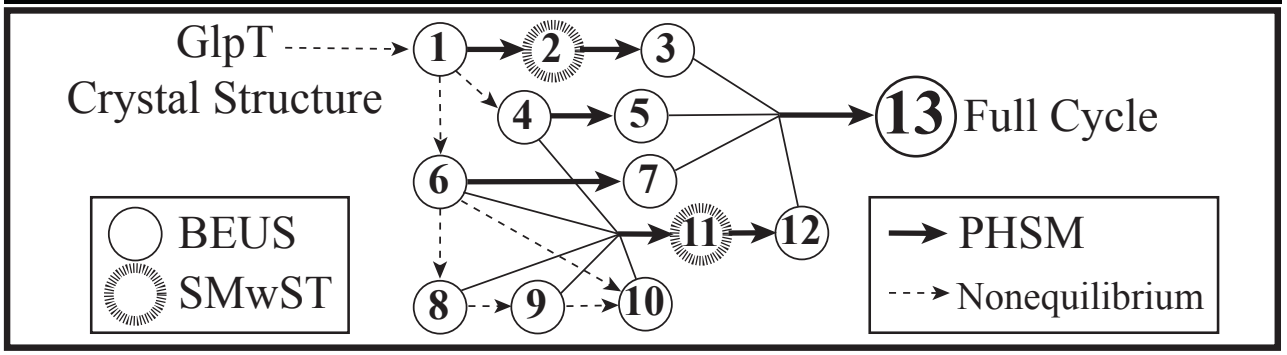




Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

Simulation Protocols

	Transition	Technique	Collective Variables	# of Replicas × Runtime		
1	$IF_a \leftrightarrow OF_a$	BEUS	(Q_1,Q_7)	12 × 40 ns	=	0.5 μs
2		SMwST	{Q}	1000 × 1 ns	=	1 μs
3		BEUS	{Q}	50 × 20 ns	=	1 μs
4	$IF_a \leftrightarrow IF_b$	BEUS	Z_{Pi}	30 × 40 ns	=	1.2 μs
5		BEUS	$(\{Q\}, Z_{Pi})$	30 × 40 ns	=	1.2 μs
6	$OF_a \leftrightarrow OF_b$	BEUS	Z_{Pi}	30 × 40 ns	=	1.2 μs
7		BEUS	$(\{Q\}, Z_{Pi})$	30 × 40 ns	=	1.2 μs
8	$IF_b \leftrightarrow OF_b$	BEUS	(Q_1,Q_7)	24 × 20 ns	=	0.5 μs
9		BEUS	Z_{Pi}	15 × 30 ns	=	0.5 μs
10		2D BEUS	$(\Delta RMSD, Z_{Pi})$	200 × 5 ns	=	1 μs
11		SMwST	$(\{Q\}, Z_{Pi})$	1000 × 1 ns	=	1 μs
12		BEUS	$(\{Q\}, Z_{Pi})$	50 × 20 ns	=	1 μs
13	**Full Cycle**	**BEUS**	**$(\{Q\}, Z_{Pi})$**	**150 × 50 ns**	**=**	**7.5 μs**
			Total Simulation Time			**18.7 μs**



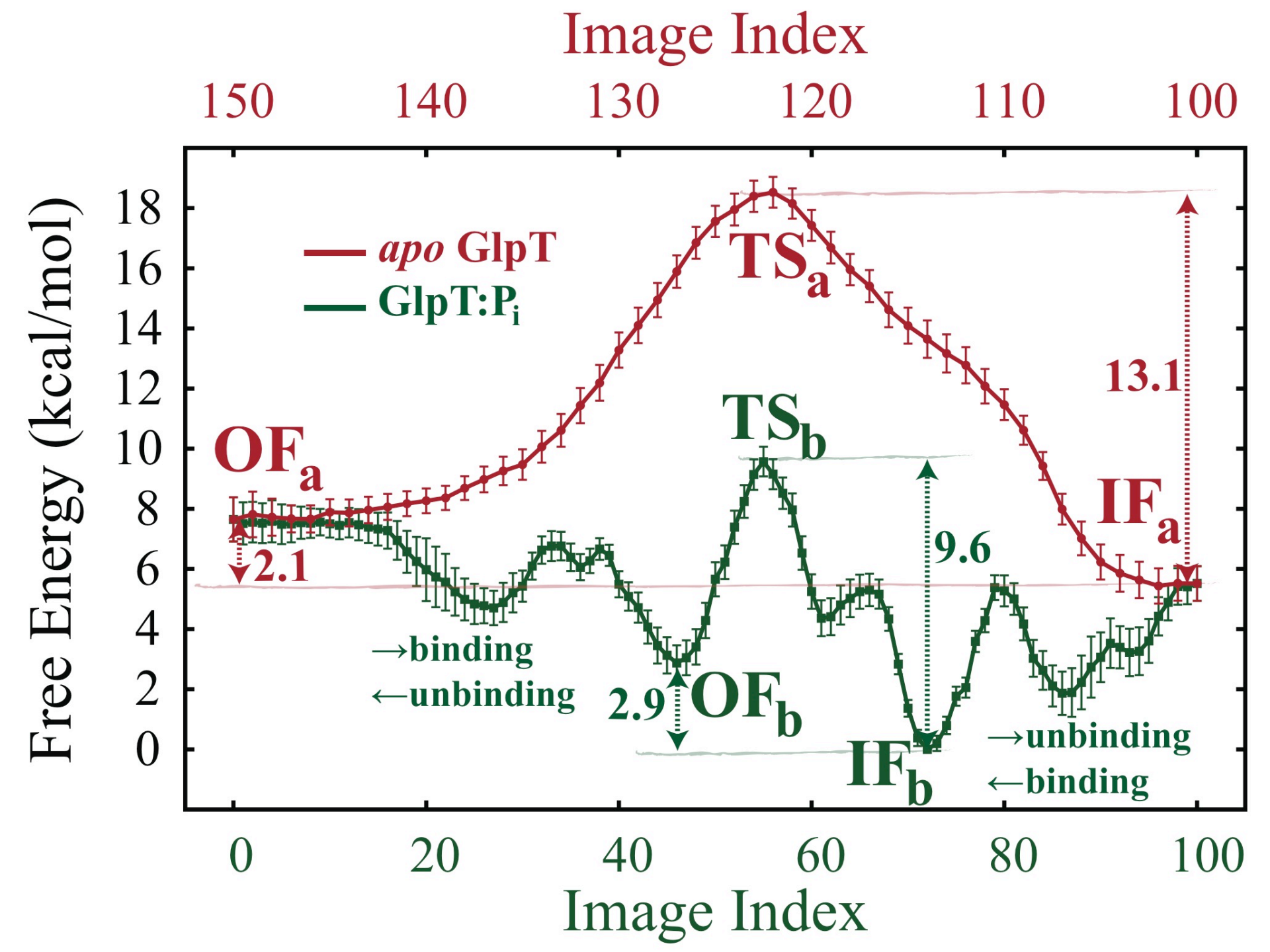


Image Index
150
140
130
120
110
100
Free Energy (kcal/mol)
18
16
14
12
10
8
6
4
2
0
apo GlpT
GlpT:Pi
TSa
TSb
OFa
IFa
13.1
2.1
9.6
2.9
OFb
IFb
→binding
←unbinding
→unbinding
←binding
0
20
40
60
80
100
Image Index

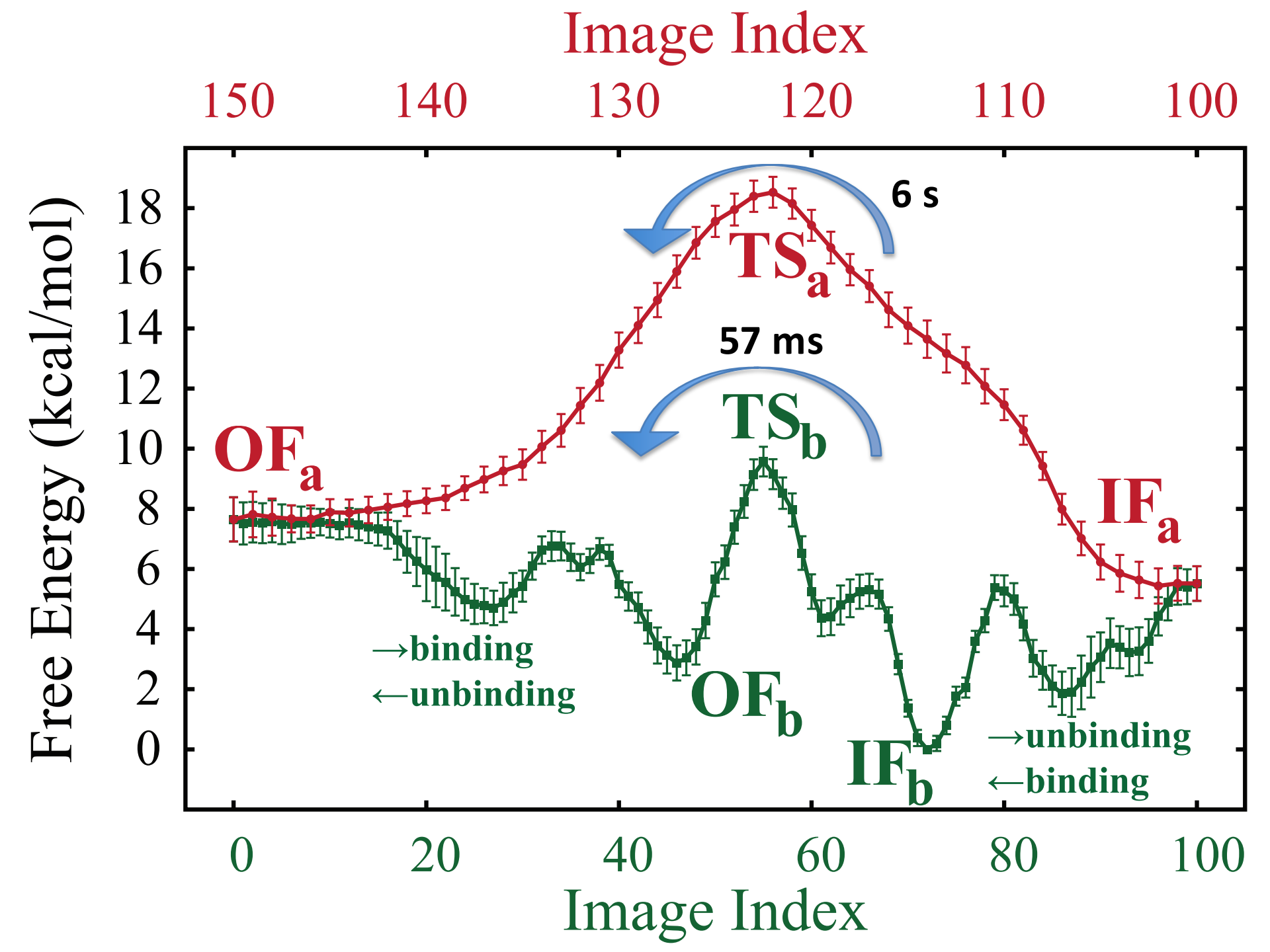


Image Index
150
140
130
120
110
100
Free Energy (kcal/mol)
18
16
14
12
10
8
6
4
2
0
6 s
TS$_a$
57 ms
TS$_b$
OF$_a$
IF$_a$
→binding
←unbinding
OF$_b$
IF$_b$
→unbinding
←binding
0
20
40
60
80
100
Image Index

- **Introduction**
  - How to study large-scale conformational changes?
- **Methodology**
  - Empirical search for good pulling protocols
  - Iterative combination of free energy calculation methods and path-finding algorithms
- **Theoretical Framework**
  - Effective Riemannian diffusion model

Dynamics of the Atomic System

- Langevin equation:

$$dp_i = \left(-\frac{\partial}{\partial x_i} V(\boldsymbol{x}) + \gamma_i m_i^{-1} p_i\right) dt + \sqrt{2\gamma_i \beta^{-1}}\, dB_i$$

$\boldsymbol{x} \in \mathbb{R}^{3N}$ $\quad \gamma_i = \gamma m_i$

$\boldsymbol{p} = \boldsymbol{m}\dot{\boldsymbol{x}}$ $\quad V(\boldsymbol{x})$ $\quad$ Wiener process

$$\langle B_i(t) \rangle = 0$$
$$\langle B_i(t) B_j(t) \rangle = \delta_{ij} t$$

- Overdamped Langevin equation:

$$dx_i = -\gamma_i^{-1} \frac{\partial}{\partial x_i} V(\boldsymbol{x}) dt + \sqrt{2\gamma_i^{-1}\beta^{-1}}\, dB_i$$

βD_i $\quad D_i$

Effective Dynamics (Euclidean Geometry)

$$\boldsymbol{x} \longmapsto \boldsymbol{\xi}(\boldsymbol{x})\ (3N \to n)$$

- Overdamped Langevin equation:

generally a position-dependent anisotropic tensor

$$d\boldsymbol{\xi} = (-\beta \mathbf{D}.\boldsymbol{\nabla} A(\boldsymbol{\xi}) + \boldsymbol{\nabla}.\mathbf{D})dt + \sqrt{2}\boldsymbol{\sigma}\, d\boldsymbol{B}$$

$$d\xi_i = \sum_j \left(-\beta D_{ij} \frac{\partial}{\partial \xi_j} A(\boldsymbol{\xi}) + \frac{\partial}{\partial \xi_j} D_{ij} \right) dt + \sqrt{2} \sum_j \sigma_{ij}\, dB_j$$

Drift rate

Note: zero drift path is not the same as MFEP.

$$\sum_j \sigma_{ij}\sigma_{jk} = D_{ik}$$

Johnson & Hummer JPCB 2012, 116, 8573.

- $A(\boldsymbol{\xi}(\boldsymbol{x}))$: the potential of mean force of the atomic system

$$A(\boldsymbol{\zeta}) = -\beta^{-1} \log\langle \delta(\boldsymbol{\xi}(\boldsymbol{x}) - \boldsymbol{\zeta}) \rangle$$

- $A(\boldsymbol{\xi})$: the "effective" potential energy of the reduced system

$$A(\boldsymbol{\zeta}) = -\beta^{-1} \log\langle \delta(\boldsymbol{\xi} - \boldsymbol{\zeta}) \rangle$$

Effective Dynamics (Euclidean Geometry)

$\boldsymbol{x} \longmapsto \boldsymbol{\xi}(\boldsymbol{x})$ ($\boldsymbol{x}$:3N-d, $\boldsymbol{\xi}$: n-d)

- Fokker-Planck/Smoluchowski equation:

$$\frac{\partial}{\partial t}\mu(\boldsymbol{\xi},t) = \sum_{ij}\left(\beta\frac{\partial}{\partial \xi_i}\left(\mu(\boldsymbol{\xi},t)\,D_{ij}\frac{\partial}{\partial \xi_j}A(\boldsymbol{\xi})\right) + \frac{\partial}{\partial \xi_i}\left(D_{ij}\frac{\partial}{\partial \xi_j}\mu(\boldsymbol{\xi},t)\right)\right)$$

$p(\boldsymbol{\xi},t|\boldsymbol{\xi}_0,0)$

$$\frac{\partial}{\partial t}p(\boldsymbol{\xi},t|\boldsymbol{\xi}_0,0) = \mathcal{L}p(\boldsymbol{\xi},t|\boldsymbol{\xi}_0,0)$$
$$\mathcal{L} \equiv \boldsymbol{\nabla}.\boldsymbol{D}.(\beta\boldsymbol{\nabla}A + \boldsymbol{\nabla})$$

- In this formalism, neither the the PMF (which is the "effective" potential energy), nor the MFEP (which is the minimum-"effective"-potential-energy path) is invariant under coordinate transformations.

Riemannian Reformulation

$\boldsymbol{\xi}: \mathbb{R}^{3N} \to \mathrm{M}$ $(M, \boldsymbol{g})$: Riemannian manifold (n-d)

- Overdamped Langevin equation:

$$d\boldsymbol{\xi} = (-\beta \mathrm{D} \boldsymbol{\nabla} G(\boldsymbol{\xi}) + \mathbf{b})dt + \sqrt{2D}\, d\boldsymbol{W}$$

$$d\xi^i = (-\beta D g^{ij} \partial_j G(\boldsymbol{\xi}) + b^i)dt + \sqrt{2D}\, dW^i$$

Riemannian Wiener process

drift rate

geometric drift $-D g^{jk} \Gamma^i_{jk}$

Christoffel symbols

$$\langle W^i(t) \rangle = 0$$
$$\langle W^i(t) W^j(t) \rangle = g^{ij} t$$

D: a position-independent scalar

$\boldsymbol{g}$: Riemannian metric tensor $G(\boldsymbol{\xi})$: Riemannian PMF

Fakharzadeh & Moradi Effective Riemannian diffusion model for conformational dynamics of biomolecular systems. **J Phys Chem Lett.** 2016;7(24):4980-4987.

Riemannian Reformulation

Riemannian PMF or "effective" potential:

$$G(\boldsymbol{\zeta}) = -\beta^{-1}\log\langle\delta_{\zeta}(\boldsymbol{\xi}(\boldsymbol{x}))\rangle \text{ atomic system}$$

$$G(\boldsymbol{\zeta}) = -\beta^{-1}\log\langle\delta_{\zeta}(\boldsymbol{\xi})\rangle \text{ reduced system}$$

$$G(\boldsymbol{\zeta}) = -\beta^{-1}\log\langle\delta_{\zeta}(\boldsymbol{\xi})\rangle$$
$$A(\boldsymbol{\zeta}) = -\beta^{-1}\log\langle\delta(\boldsymbol{\xi}-\boldsymbol{\zeta})\rangle$$
$$G(\boldsymbol{\zeta}) = A(\boldsymbol{\zeta}) + \frac{1}{2}\beta^{-1}\log g$$

Riemannian
Dirac delta function

$$\mathcal{F}(\boldsymbol{\zeta}) = \int d\Omega_{\xi}\delta_{\zeta}(\boldsymbol{\xi})\mathcal{F}(\boldsymbol{\xi})$$

$$d\Omega_{\xi} = \sqrt{g}d^{n}\xi$$

Gibbs measure:

$$e^{-\beta G(\xi)}d\Omega_{\xi} = e^{-\beta A(\xi)}d^{n}\xi$$

invariant

Conventional
Dirac delta function

$$\mathcal{F}(\boldsymbol{\zeta}) = \int d^{n}\xi\,\delta(\boldsymbol{\xi}-\boldsymbol{\zeta})\mathcal{F}(\boldsymbol{\xi})$$

Riemannian Reformulation

$\boldsymbol{\xi} : \mathbb{R}^{3N} \rightarrow \mathrm{M}$ $(M, \boldsymbol{g})$: Riemannian manifold (n-d)

- Overdamped Langevin equation:

$$d\boldsymbol{\xi} = (-\beta \mathrm{D} \boldsymbol{\nabla} G(\boldsymbol{\xi}) + \mathbf{b})dt + \sqrt{2D}\, d\boldsymbol{W}$$

$$d\xi^i = (-\beta D g^{ij} \partial_j G(\boldsymbol{\xi}) + b^i)dt + \sqrt{2D}\, dW^i$$

Potential/FE gradient (invariant)

$-D g^{jk} \Gamma^i_{jk}$ Geometric drift (non-invariant)

In a geodesic normal coordinate system, b^i disappears.

non-invariant

$$d\boldsymbol{\xi} = (-\beta \mathbf{D}.\boldsymbol{\nabla} A(\boldsymbol{\xi}) + \boldsymbol{\nabla}.\mathbf{D})dt + \sqrt{2}\boldsymbol{\sigma} d\boldsymbol{B}$$

non-invariant

Riemannian Reformulation

- Fokker-Planck/Smoluchowski equation:

$$\frac{\partial}{\partial t}u(\boldsymbol{\xi},t) = \beta D \frac{1}{\sqrt{g}}\partial_i\left(u(\boldsymbol{\xi},t)\sqrt{g}g^{ij}\partial_j G(\boldsymbol{\xi})\right) + D\frac{1}{\sqrt{g}}\partial_i\left(\sqrt{g}g^{ij}\partial_j u(\boldsymbol{\xi},t)\right)$$

$$\frac{\partial}{\partial t}u(\boldsymbol{\xi},t) = \beta D\boldsymbol{\nabla}.(u(\boldsymbol{\xi},t)\boldsymbol{\nabla}G) + D\Delta u(\boldsymbol{\xi},t)$$

$$\frac{\partial}{\partial t}p(\boldsymbol{\xi},t|\boldsymbol{\xi}_0,0) = \mathcal{L}p(\boldsymbol{\xi},t|\boldsymbol{\xi}_0,0)$$

$$\mathcal{L} \equiv D\boldsymbol{\nabla}.(\beta\boldsymbol{\nabla}G + \boldsymbol{\nabla})$$

It can be shown the conventional diffusion with potential $A(\boldsymbol{\xi})$ and diffusion tensor $\boldsymbol{D}$ and the Riemannian diffusion with potential $G(\boldsymbol{\xi})$ and metric $\boldsymbol{g}$ are equivalent if:

$$\boldsymbol{D} = D\boldsymbol{g}^{-1} \text{ and } G = A + \frac{1}{2}\beta^{-1}\log g$$

Riemannian Reformulation

- The Riemannian formulation allows for developing more **robust** free energy calculation methods and path-finding algorithms (due to the **"invariance"**).
- Methods such as (BE)US and SM(wST) can be modified by using **geodesic distance** instead of **Euclidean distance** (and modifying the definitions of differential operators, etc)
- For certain collective variables, the geodesic distance can be approximated by analytic expressions, e.g., for **orientation quaternions**:

$$d(\boldsymbol{q}_1, \boldsymbol{q}_2) \approx \cos^{-1}(\boldsymbol{q}_1 . \boldsymbol{q}_2)$$

- For other collective variables, on-the-fly metric calculations maybe necessary.

Free Energy Calculation Methods Revisited

- Biasing potential: $U_c(\boldsymbol{\xi}) = \frac{1}{2} k\, d(\boldsymbol{\xi}, \boldsymbol{\xi}_c)$
- Perturbed free energy: $e^{-\beta f(\xi_c)} = \int e^{-\beta(G(\boldsymbol{\xi}) + U_c(\boldsymbol{\xi}))}\, d\Omega_{\boldsymbol{\xi}}$
- Connecting perturbed free energy to PMF:

$$e^{-\beta G(\boldsymbol{\xi})} = \left(\frac{2\pi}{\beta k}\right)^{-\frac{n}{2}} e^{-\frac{1}{2\beta k}\Delta} e^{-\beta f(\boldsymbol{\xi})}$$

Fakharzadeh & Moradi **J Phys Chem Lett.** 2016;7(24):4980-4987.

- This relation was first proven for a 1D case.

Hummer, G.; Szabo, A. PNAS. 2010, 107, 21441.

- Stiff-spring approximation (large k):

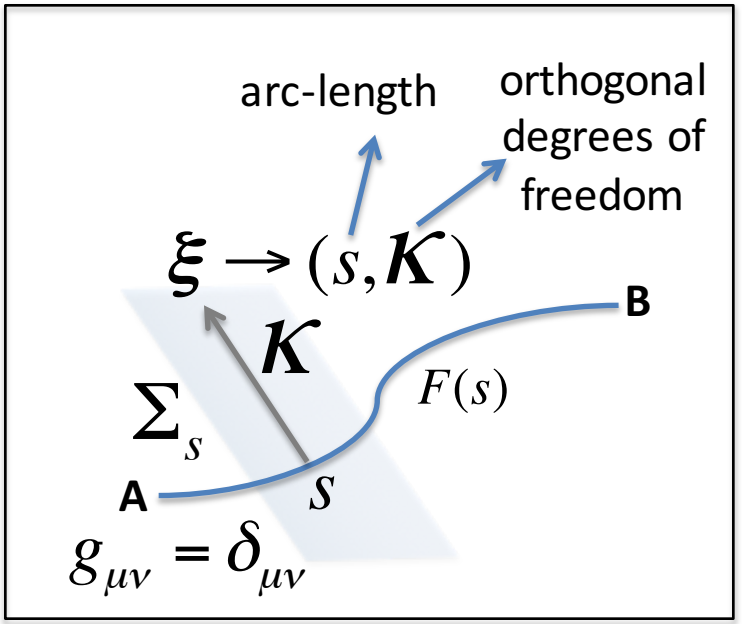
$$G(\boldsymbol{\xi}) = f(\boldsymbol{\xi}) + \frac{1}{2\beta k}(\beta \nabla f(\boldsymbol{\xi}).\nabla F(\boldsymbol{\xi}) - \Delta f(\boldsymbol{\xi})) + O((\frac{1}{\beta k})^2) + \text{constant}$$

Free Energy on a Curve

- A curve $\xi(s)$ parameterized by its arc-length s: $ds^2 = g_{\mu\nu} d\xi^\mu d\xi^\nu$
- Free energy on the curve: $F(s) = f(\xi(s))$
- Stiff-spring correction term ($\times 2\beta k$):

$$\beta \nabla f(\xi).\nabla F(\xi) - \Delta f(\xi)$$
$$= \beta\left(\frac{d}{ds}F(s)\right)^2 - \frac{d^2}{ds^2}F(s)$$
$$+\beta \nabla_\perp f(\xi).\nabla_\perp f(\xi) - \Delta_\perp f(\xi)$$

Frenet frame



- Minimum free energy path:

$$\xi(s) \parallel \nabla f(\xi(s)) \qquad (\nabla : g^{\mu\nu}\partial_\nu)$$

$$G(s) = F(s) + \frac{1}{2\beta k}\left(\beta\left(\frac{d}{ds}F(s)\right)^2 - \frac{d^2}{ds^2}F(s)\right) + O\left(\left(\frac{1}{\beta k}\right)^2\right) + \text{constant}$$

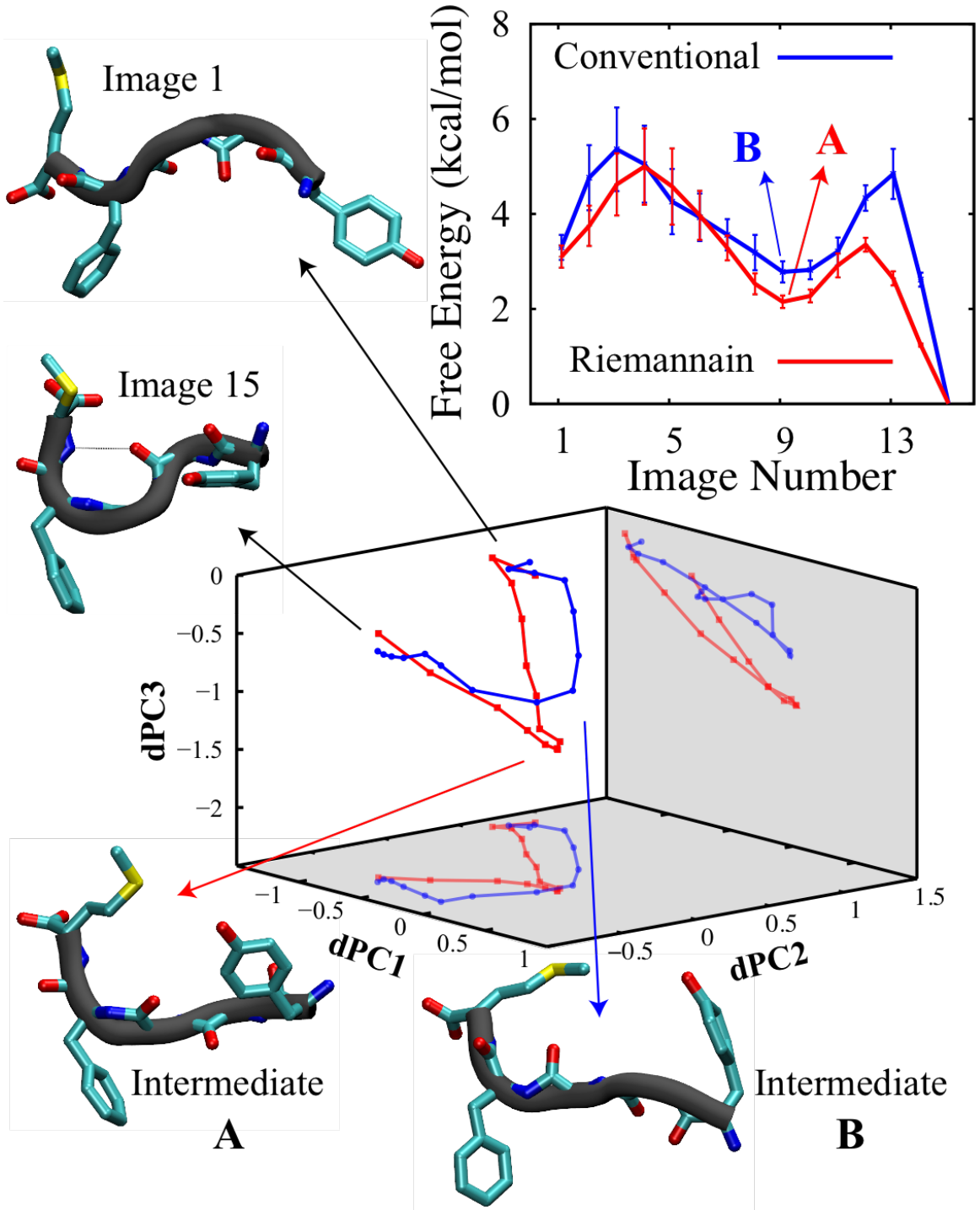
Metric estimation

$$\left\langle \delta\xi^i \delta\xi^j \right\rangle_{s,\delta_t} = 2D\left\langle \delta W^i \delta W^j \right\rangle \delta t + O(\delta t^2) = 2D g^{ij} \delta t + O(\delta t^2)$$
$$\left\langle \delta\xi^i \delta\xi^j \right\rangle_{s,\delta_t} \approx 2\sum_l D_l \left\langle \frac{\partial}{\partial x^l}\xi^i \frac{\partial}{\partial x^l}\xi^j \right\rangle_s \delta t + O(\delta t^2)$$

$$g^{ij} \approx \sum_l \frac{m}{m_l}\left\langle \frac{\partial}{\partial x_l}\xi^i \frac{\partial}{\partial x_l}\xi^j \right\rangle$$

Conformational transition between two configurations of Met-enkephalin peptide

- All backbone dihedral angles were used as collective variables
- The trajectories are projected onto the first three dPCs (dihedral pricipal components)

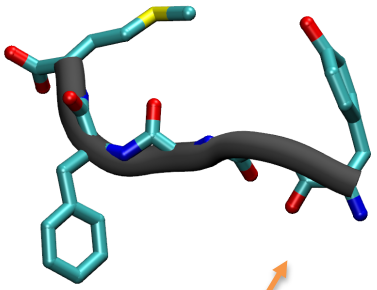


Coordinate transformation changes the results of conventional SMwST/BEUS substantially.

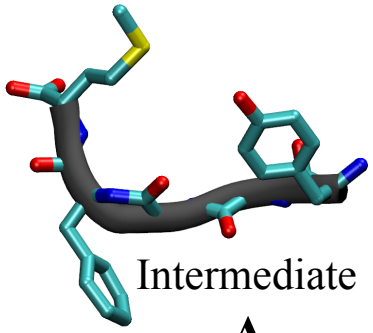
Riemannian SMwST/BEUS results in an invariant PMF and MFEP.

$$dPC_i = \sum_j a_{ij} \cos\theta_j + a_{ij} \sin\theta_j$$

$\theta_j's$ are the dihedral angles



Intermediate B



Intermediate A