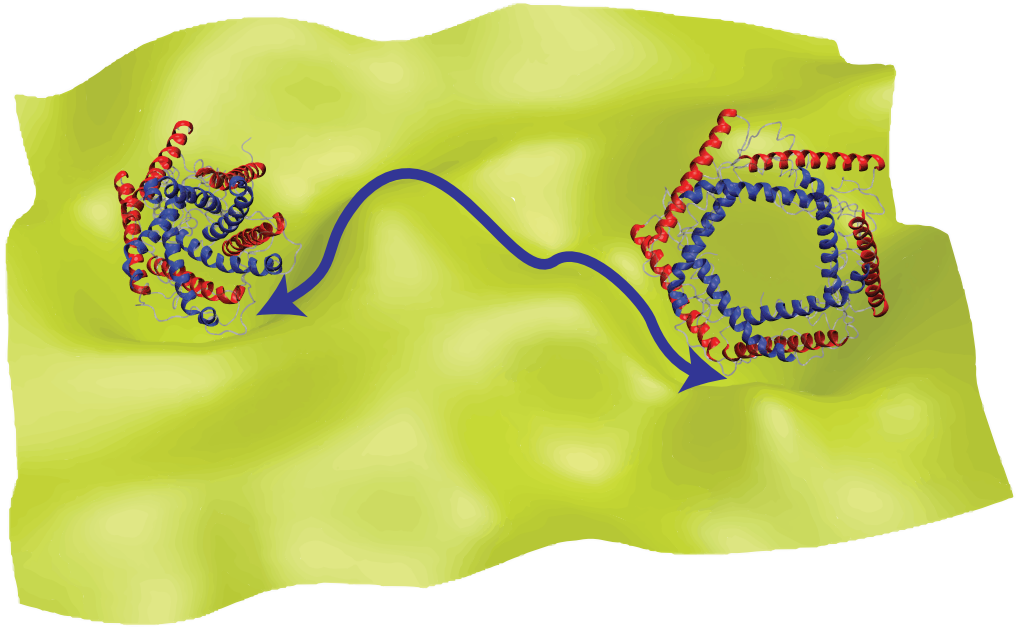
Exploring Complex Reaction Pathways

Mahmoud Moradi

Department of Chemistry and Biochemistry

University of Arkansas

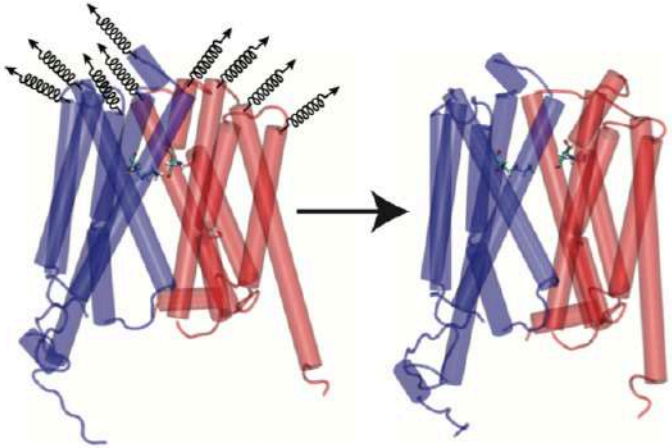


"Hands-on" Workshop on Enhanced Sampling and Free-Energy Calculation at Urbana, IL

September 14, 2018

Department of Chemistry and Biochemistry
University of Arkansas, Fayetteville

Exploring Complex Conformational Transition Pathways



Tutorial by Mahmoud Moradi

September 18 2017

A current version of this tutorial is available at
http://www.ks.uiuc.edu/Training/Tutorials/

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

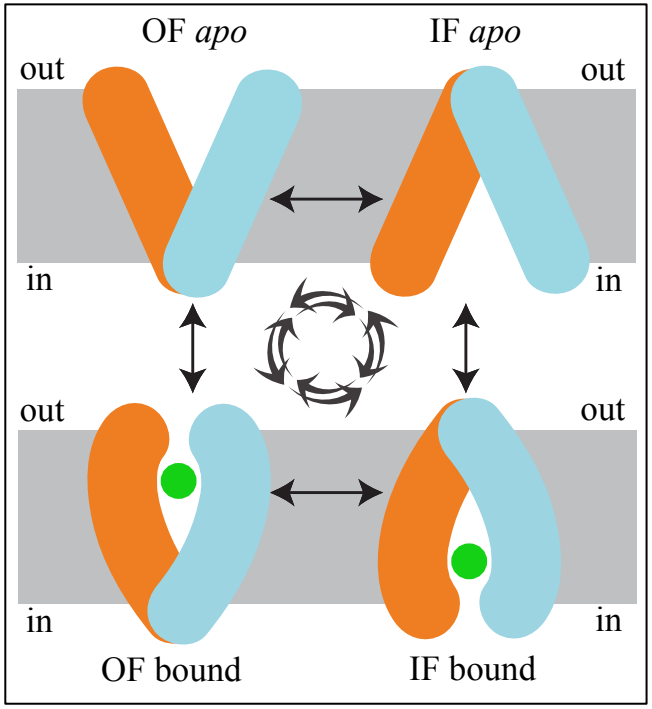
Outline

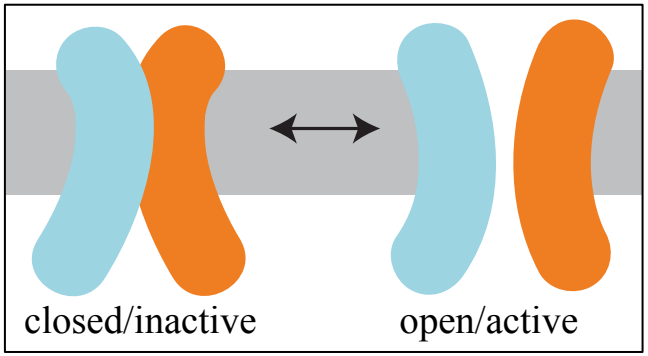
- **Introduction**
  - How to study large-scale conformational changes?

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)**

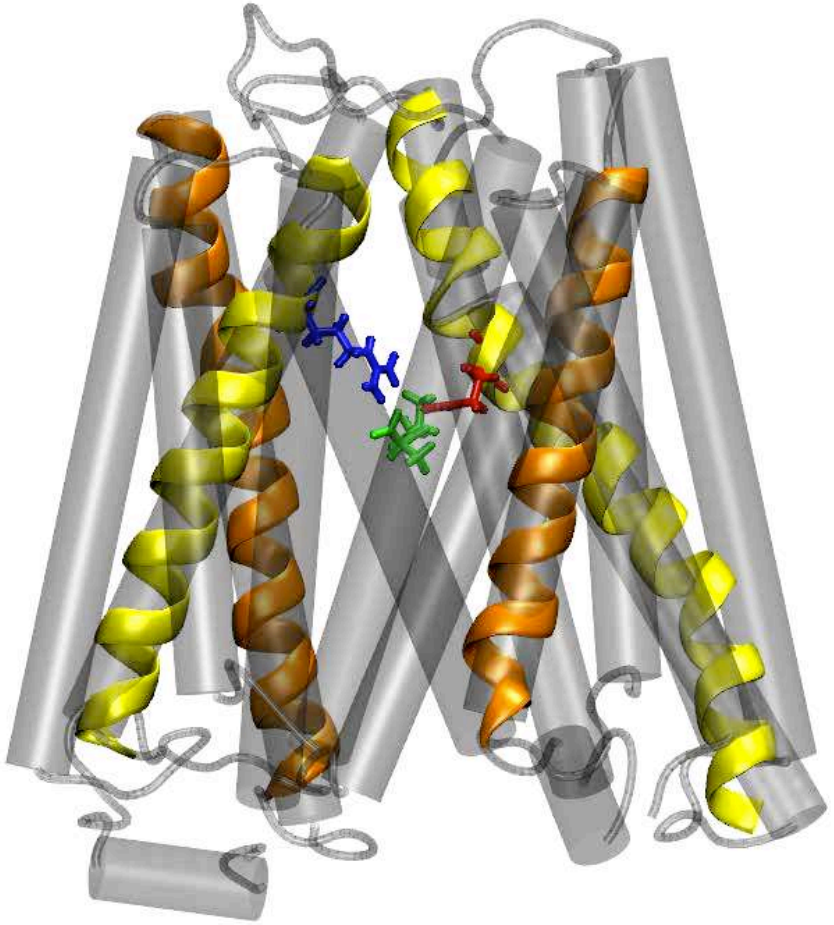
- **Channels** may require large-scale conformational changes between their **open/active** and **closed/inactive** states.





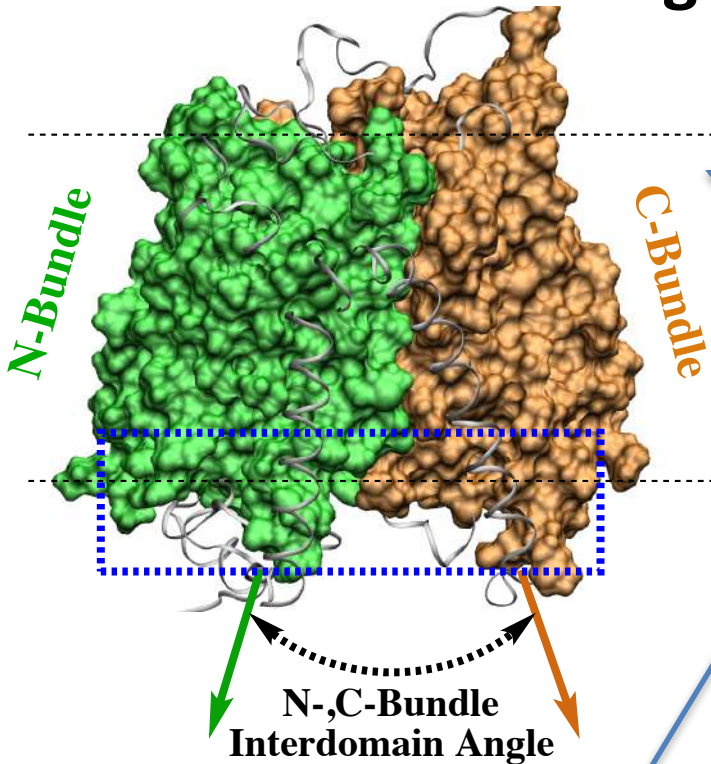
A Case Study:
Proton-coupled Oligopeptide Transporters (POTs)

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

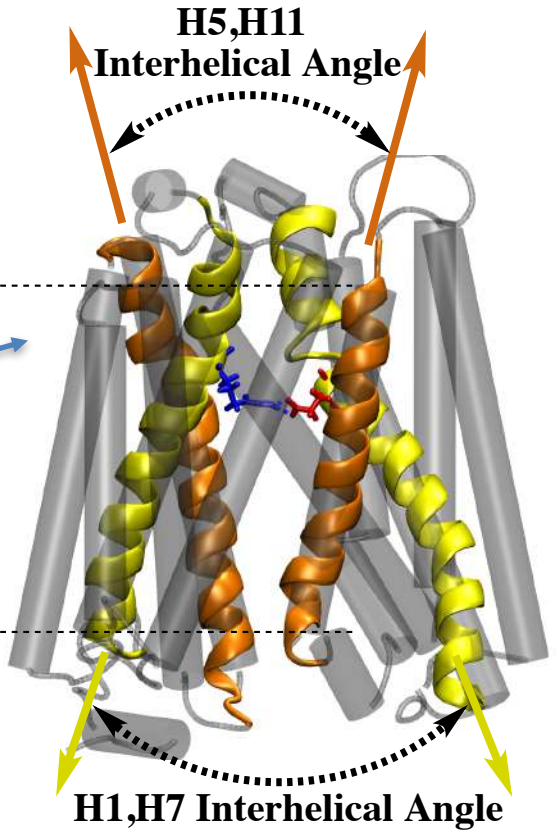


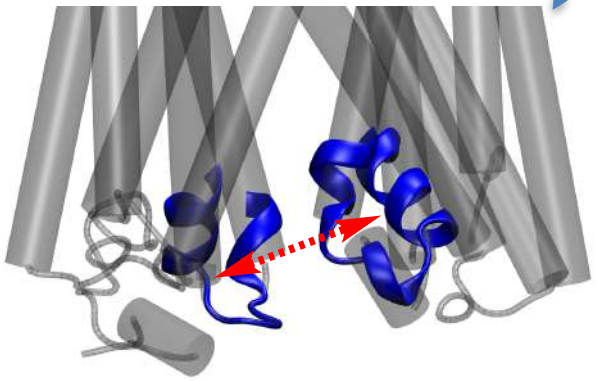
K Immadisetty, J Hettige, and M Moradi, ***What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT?*** J. Phys. Chem. B, **121**:3644-3656, 2017.

Monitoring Global and Local Conformational Changes

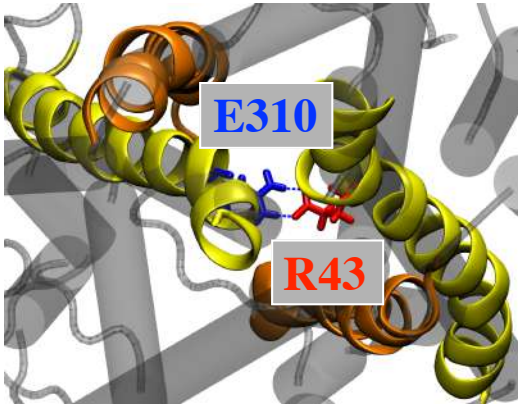


Global

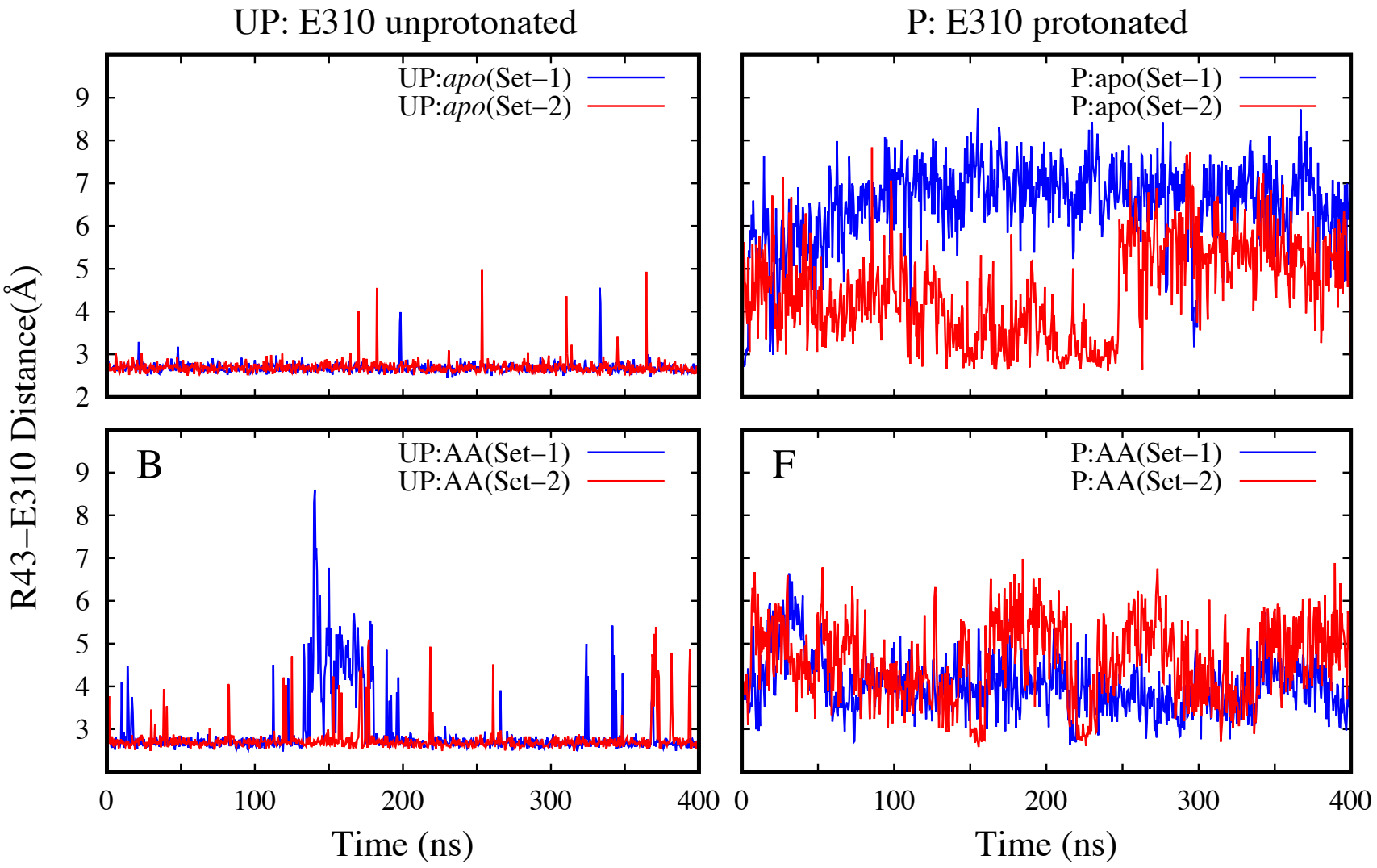




Local
(Salt Bridge)

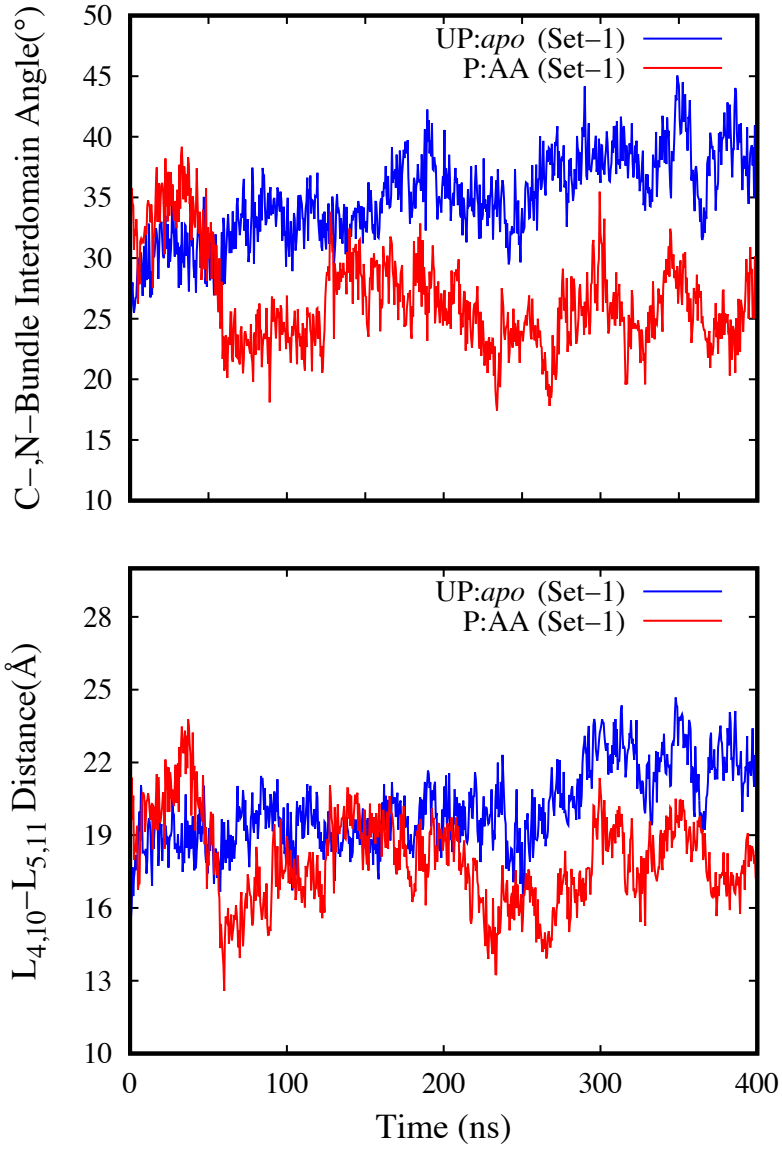


Local Conformational Changes



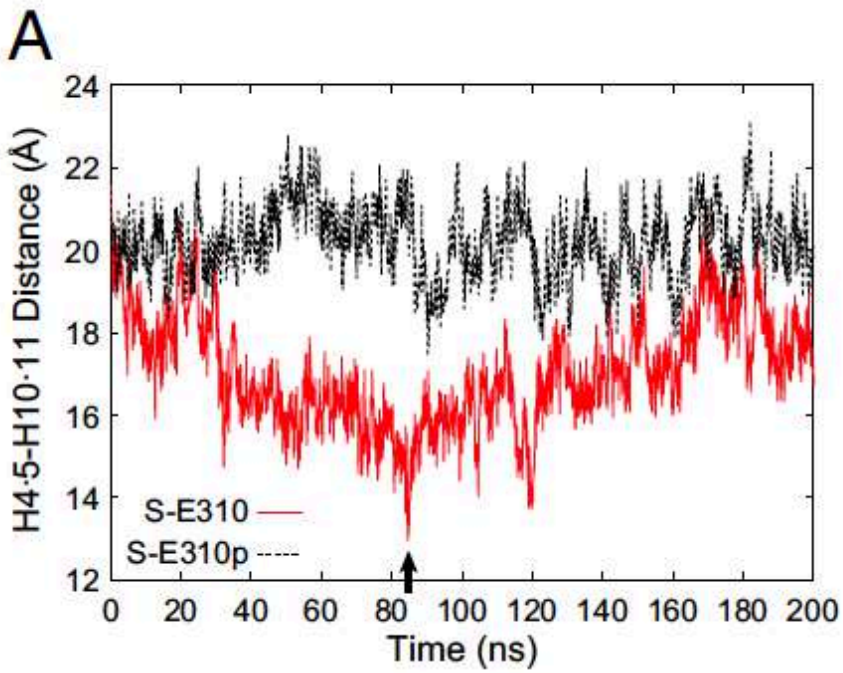
There is a clear distinction between different conditions.

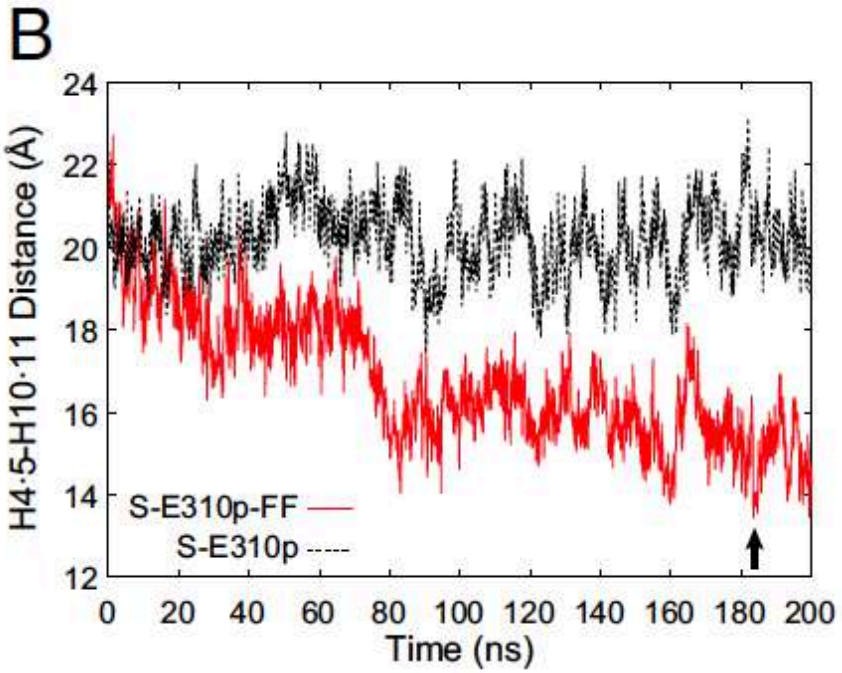
Global Conformational Changes



There is no clear distinction between different conditions.

Global Conformational Changes

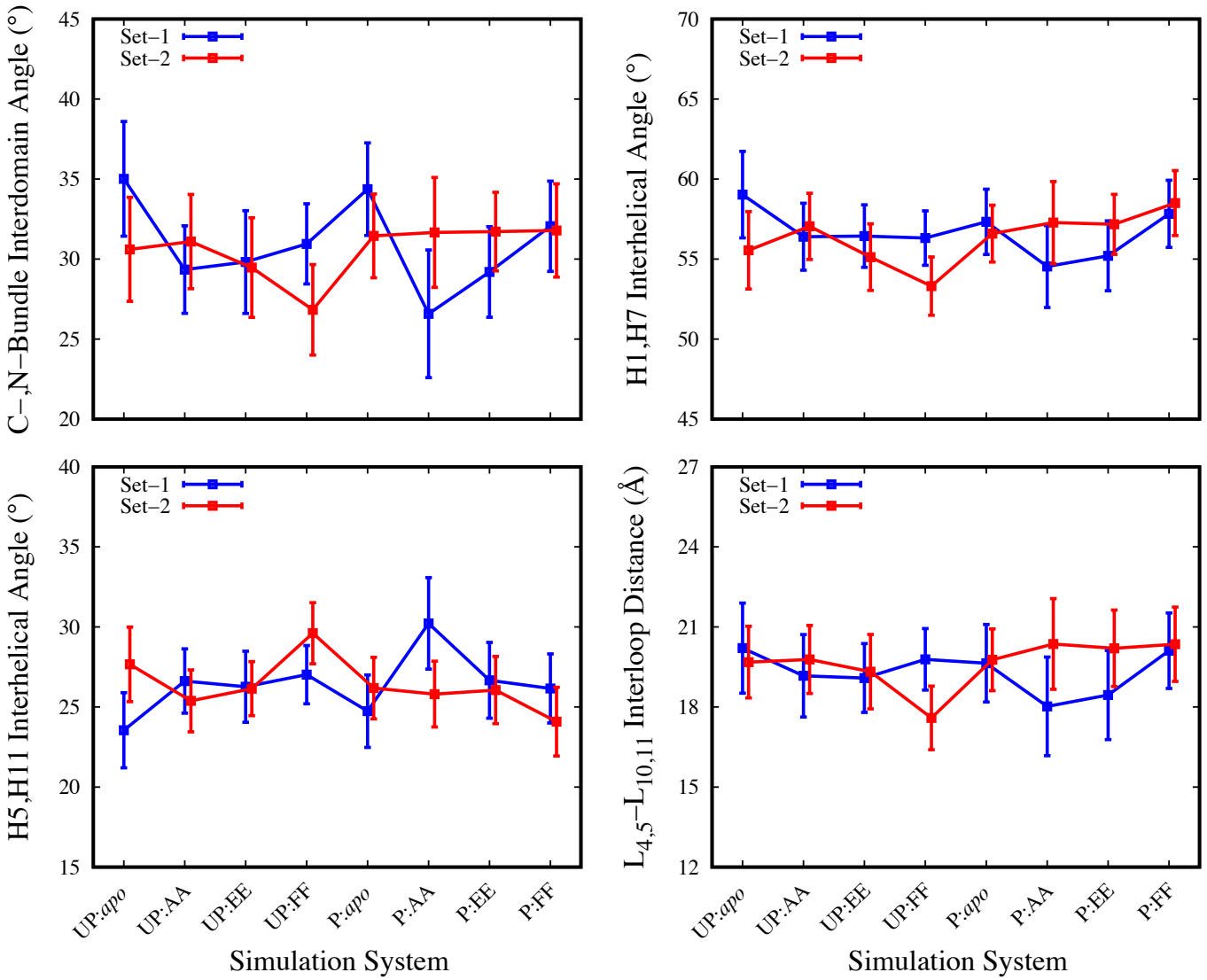




"Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT." **PNAS 2013** | vol. 110 | no. 28 | 11343–11348.

Although a common practice, statements made about millisecond-level biomolecular events based on sub-microsecond level simulations are not reliable.

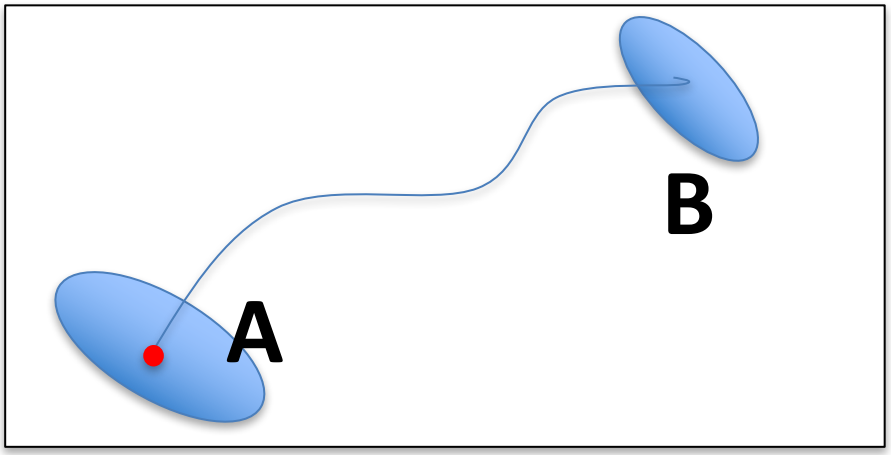
Global Conformational Changes



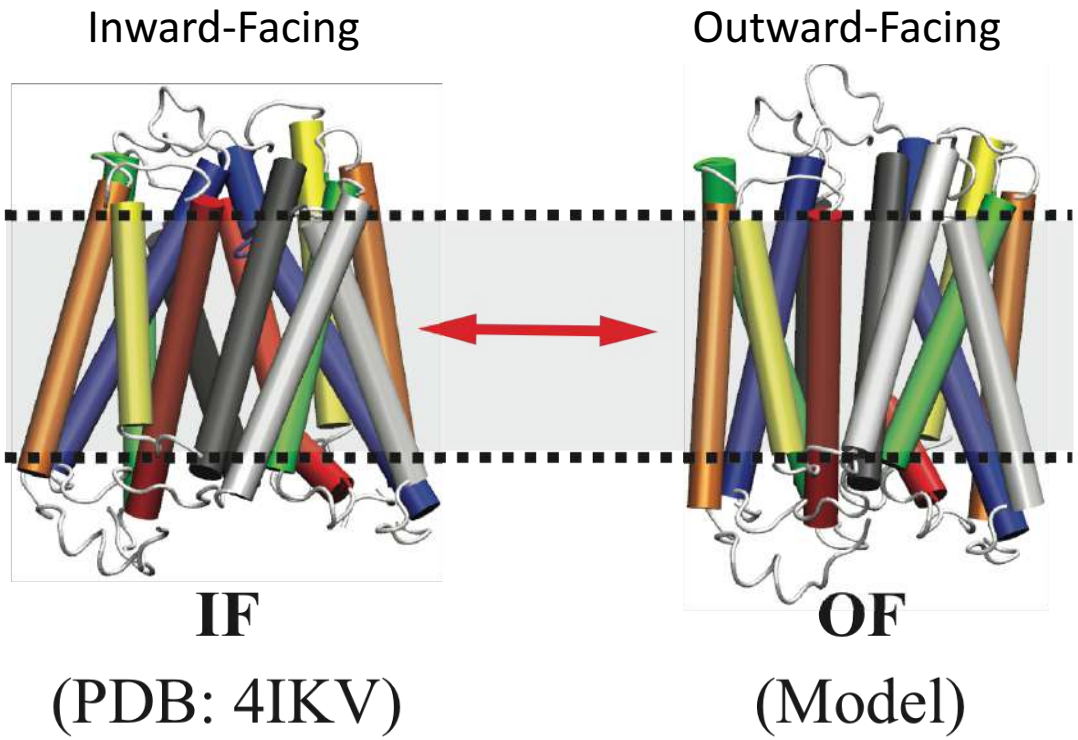
There is no statistically significant distinction between different conditions.

- **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.

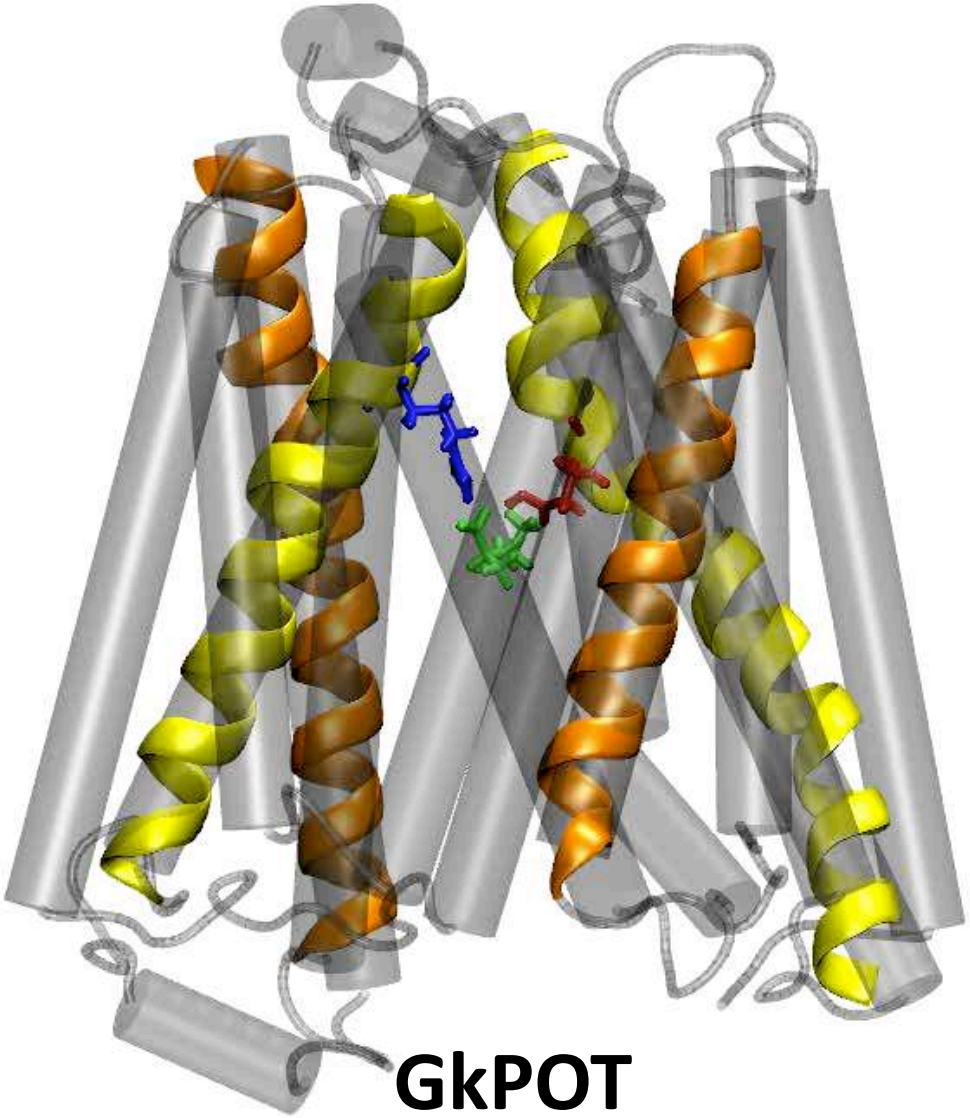


How to study large-scale conformational changes?



GkPOT

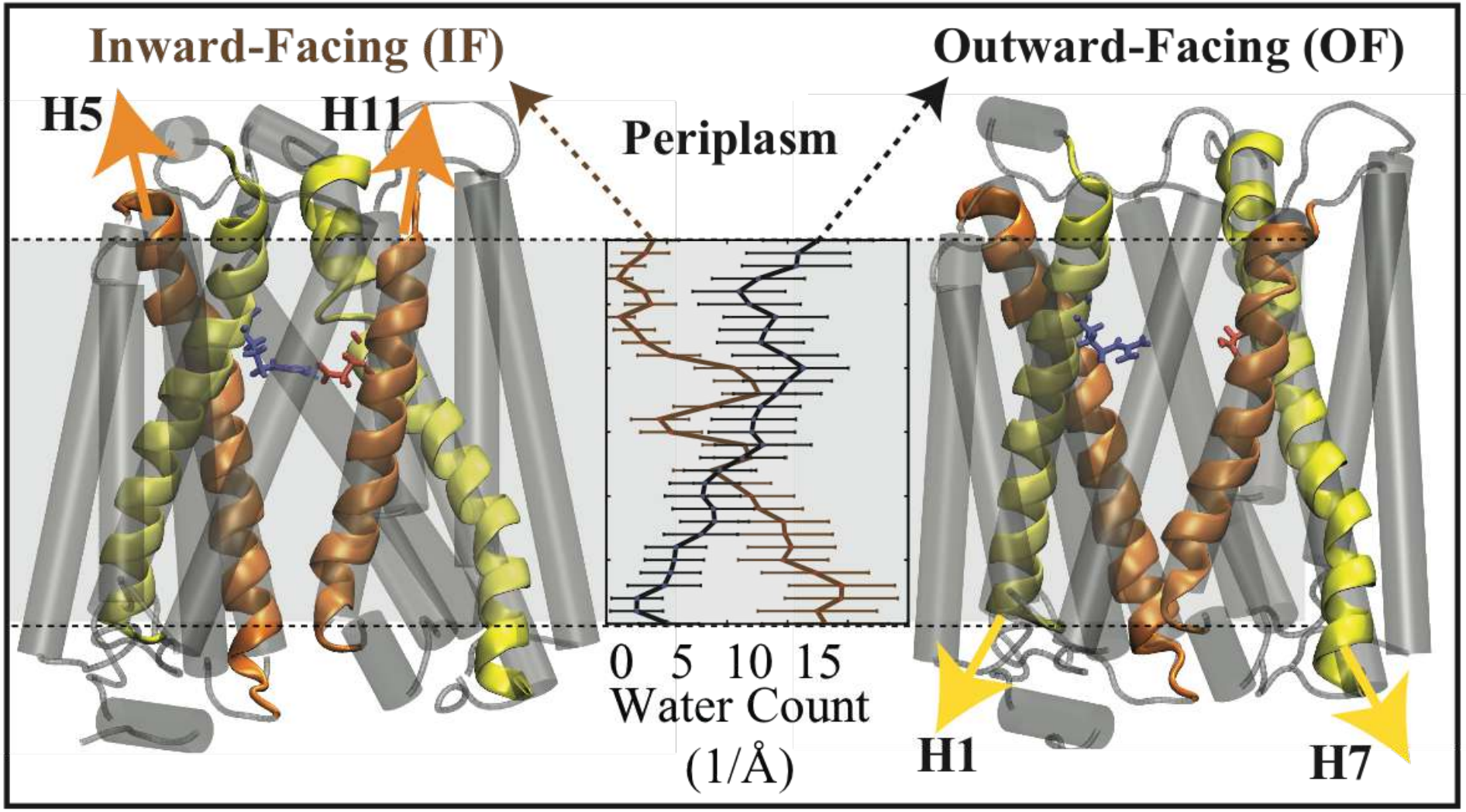
How to study large-scale conformational changes?



GkPOT

If you are not sure how good your collective variable is, start with nonequilibrium pulling.

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

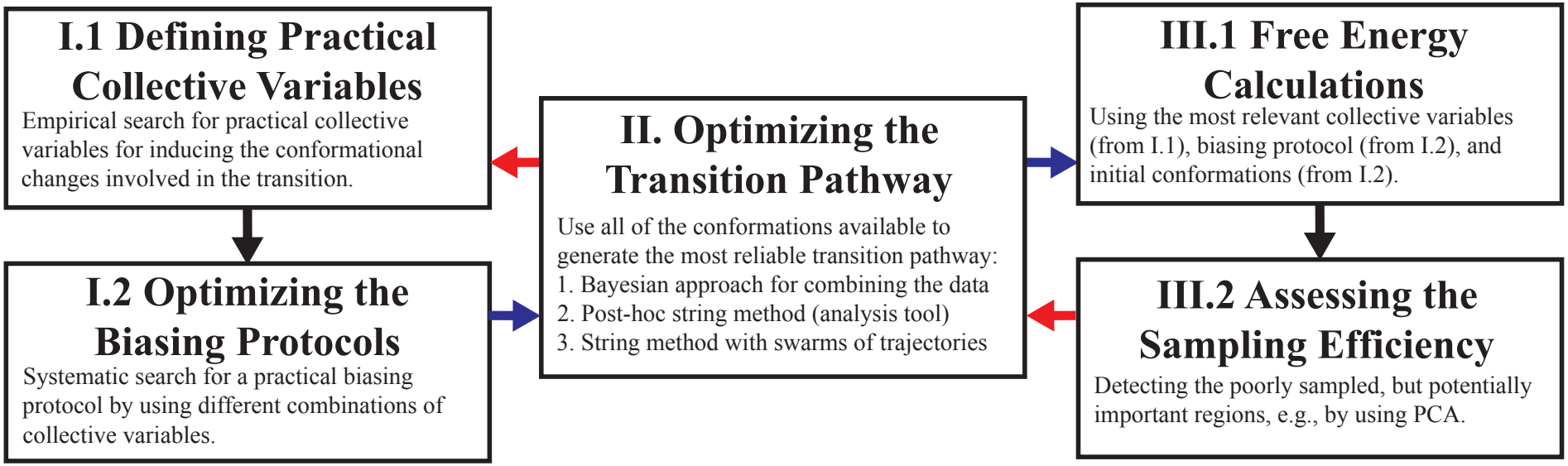
Moradi et al., **Proc Natl Acad Sci 106** 20746 (**2009**)
Moradi et al., **Chem Phys Lett 518** 109 (**2011**)
Moradi et al., **J Chem Phys 133** 125104 (**2010**)
Moradi et al., **Int J Quantum Chem 110** 2865 (**2010**)
Moradi et al., **Biophys J 100** 1083 (**2011**)
Moradi et al., **J Phys Chem 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**)

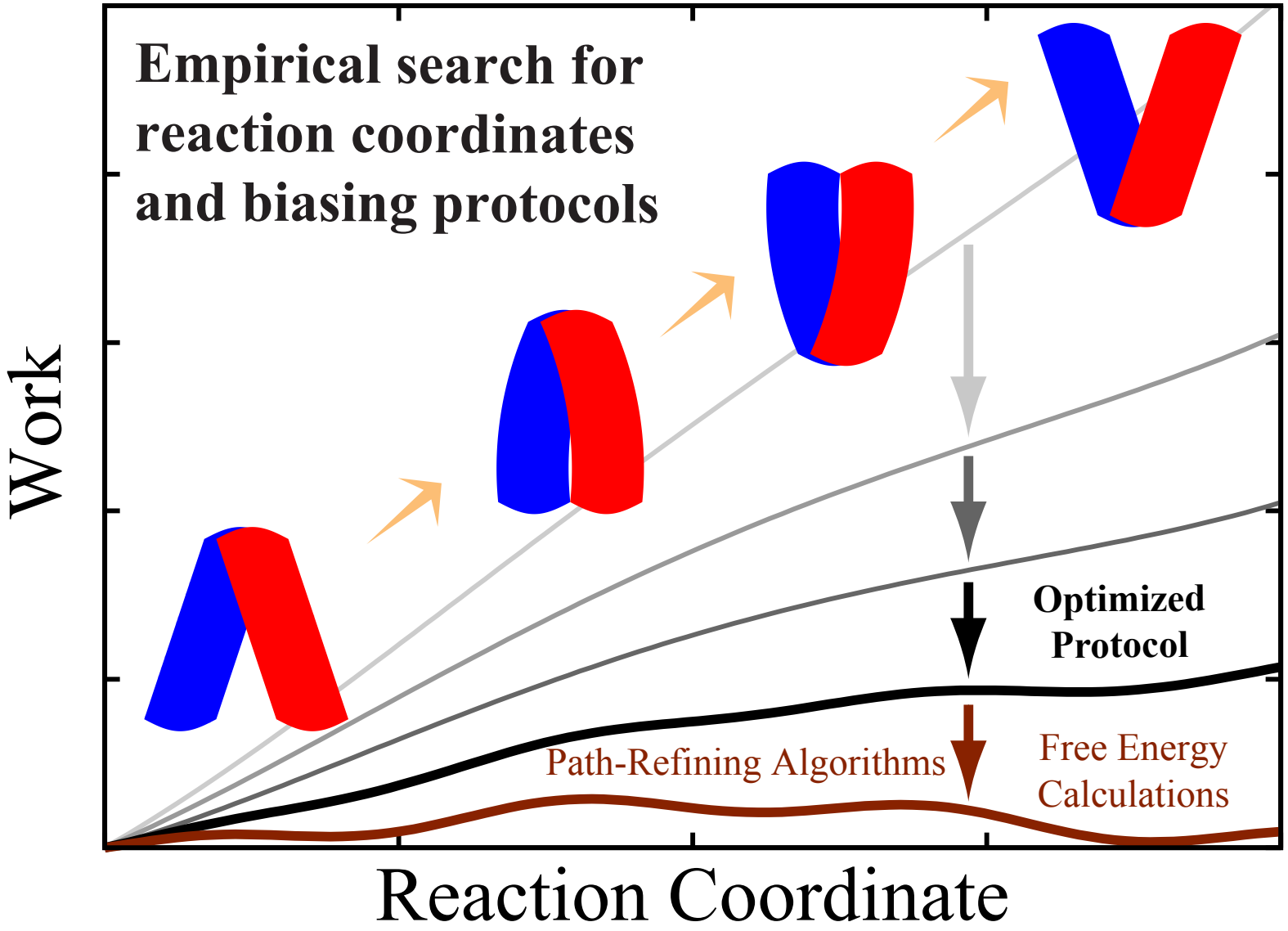
Moradi et al., **Nat Commun 6** 8393 (**2015**)
Fakharzadeh & Moradi, **J Phys Chem Lett 7** 4980 (**2016**)

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



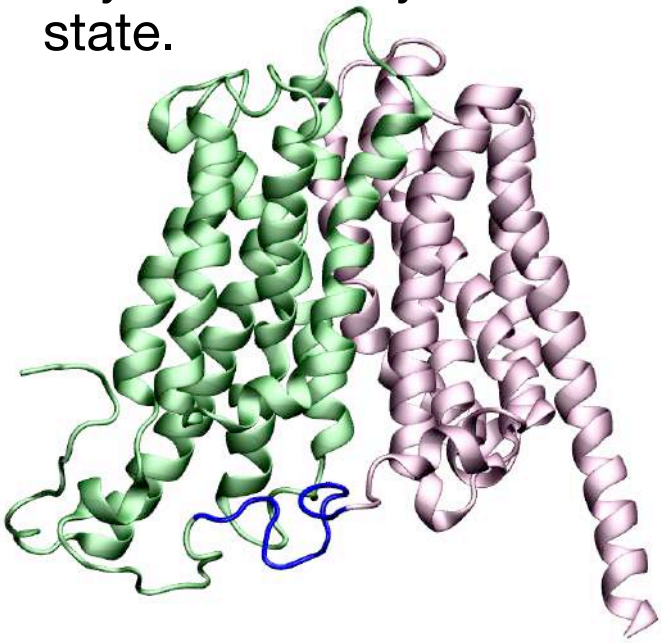
Sampling Ideas

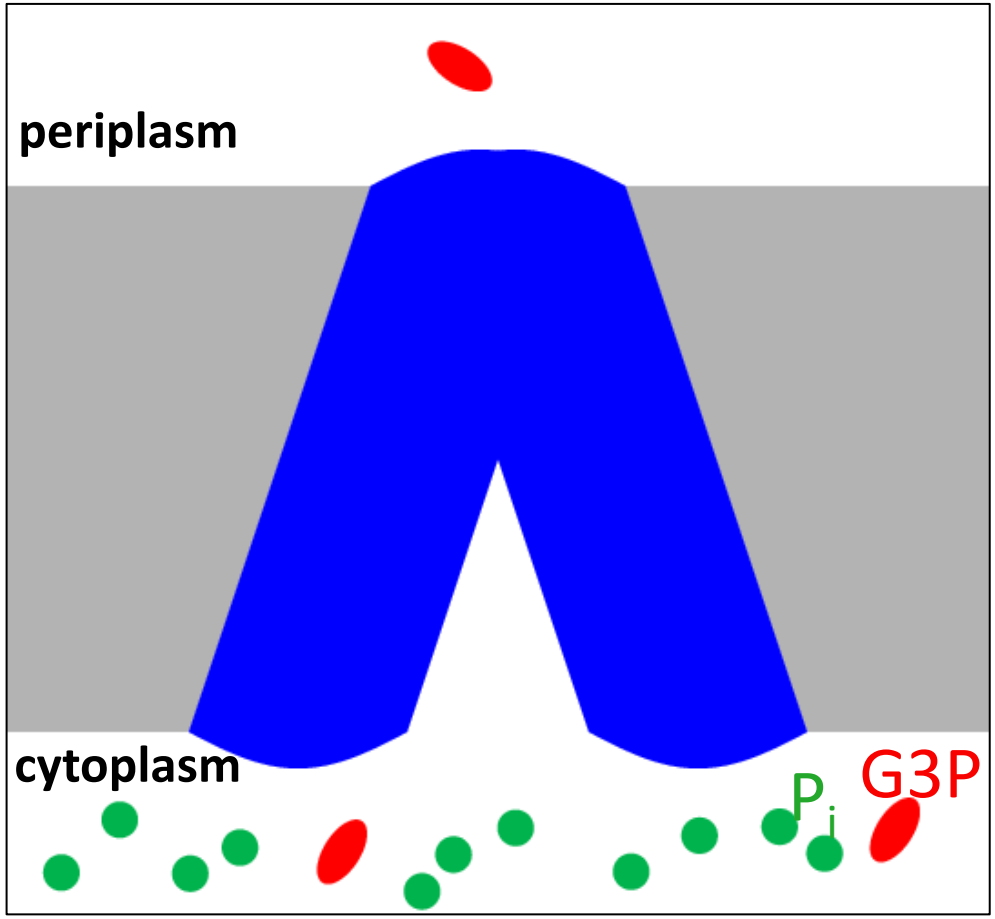


- **Introduction**
  - How to study large-scale conformational changes?
- **Methodology**
  - **Empirical search for good pulling protocols**
  - Iterative combination of free energy calculation methods and path-finding algorithms

Example: Glycerol-3-Phophate Transporter (GlpT)

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

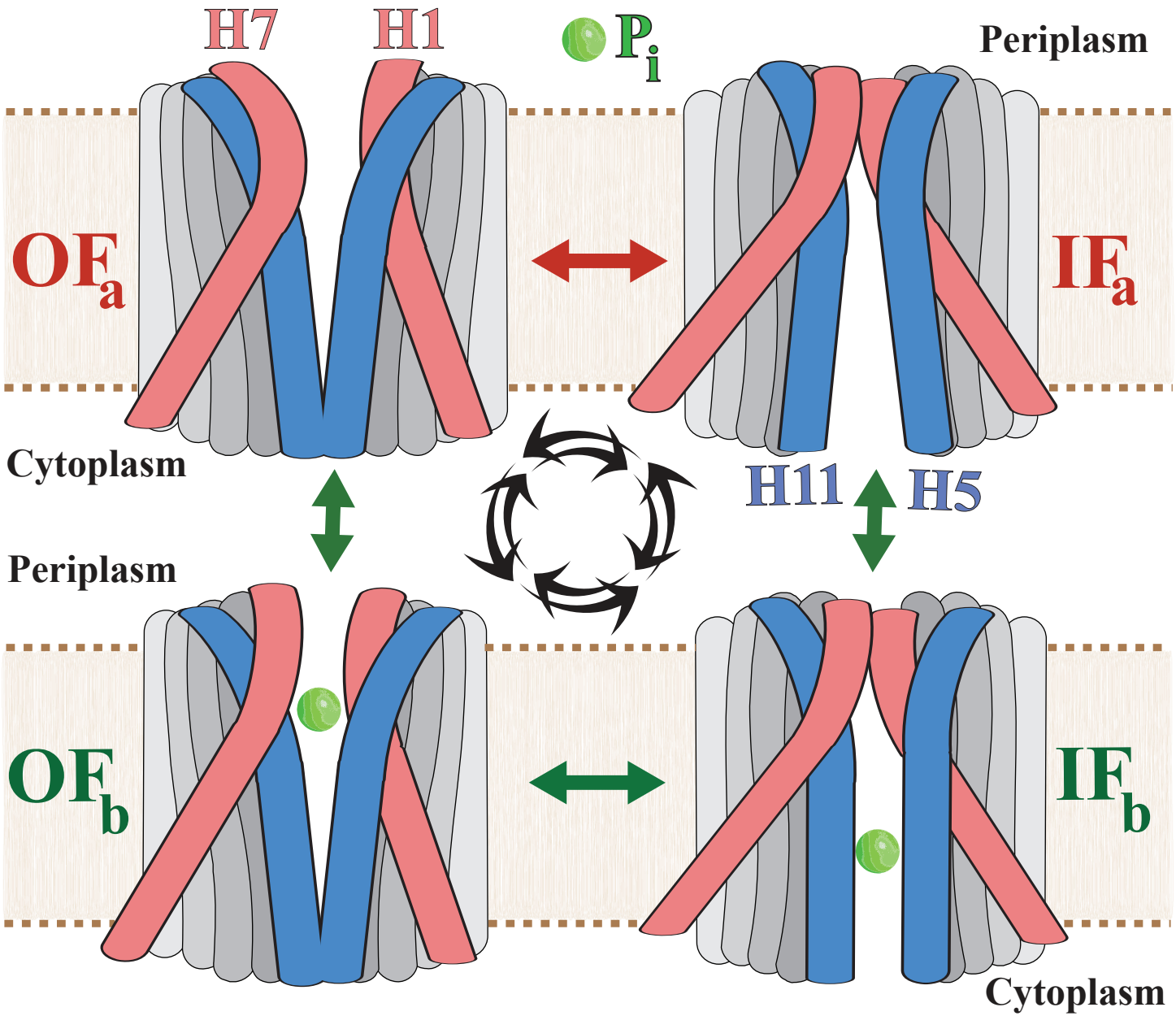




- GlpT transports **G3P** using $\mathbf{P_i}$ gradient.
- $\mathbf{P_i}$:$\mathbf{P_i}$ exchanger (in the absence of organic phosphate)

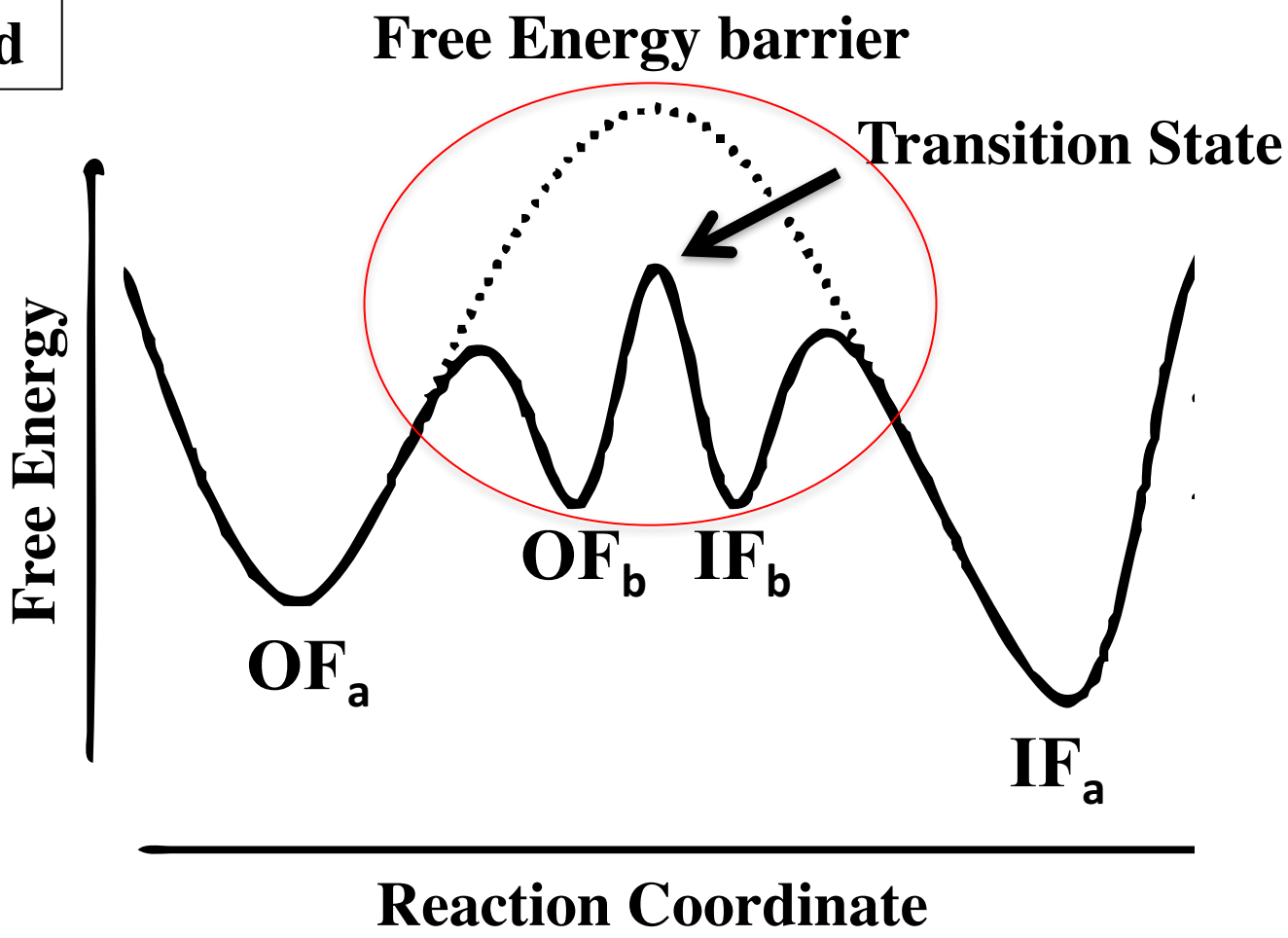
Moradi, Enkavi, and Tajkhorshid, **Nature Communications 6** 8393 (**2015**)

Transport Thermodynamics



Transport Thermodynamics

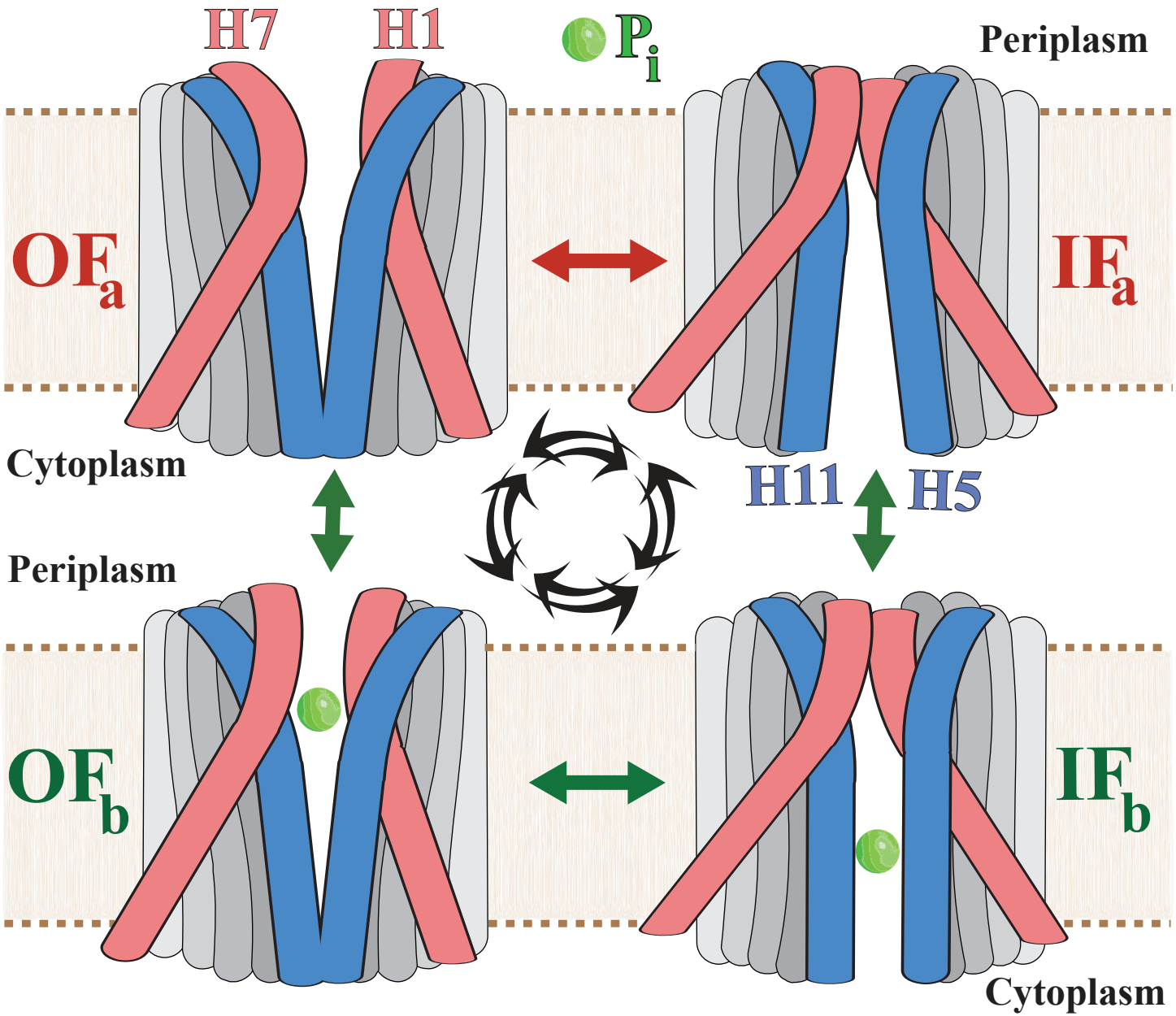
a: *apo*
b: bound



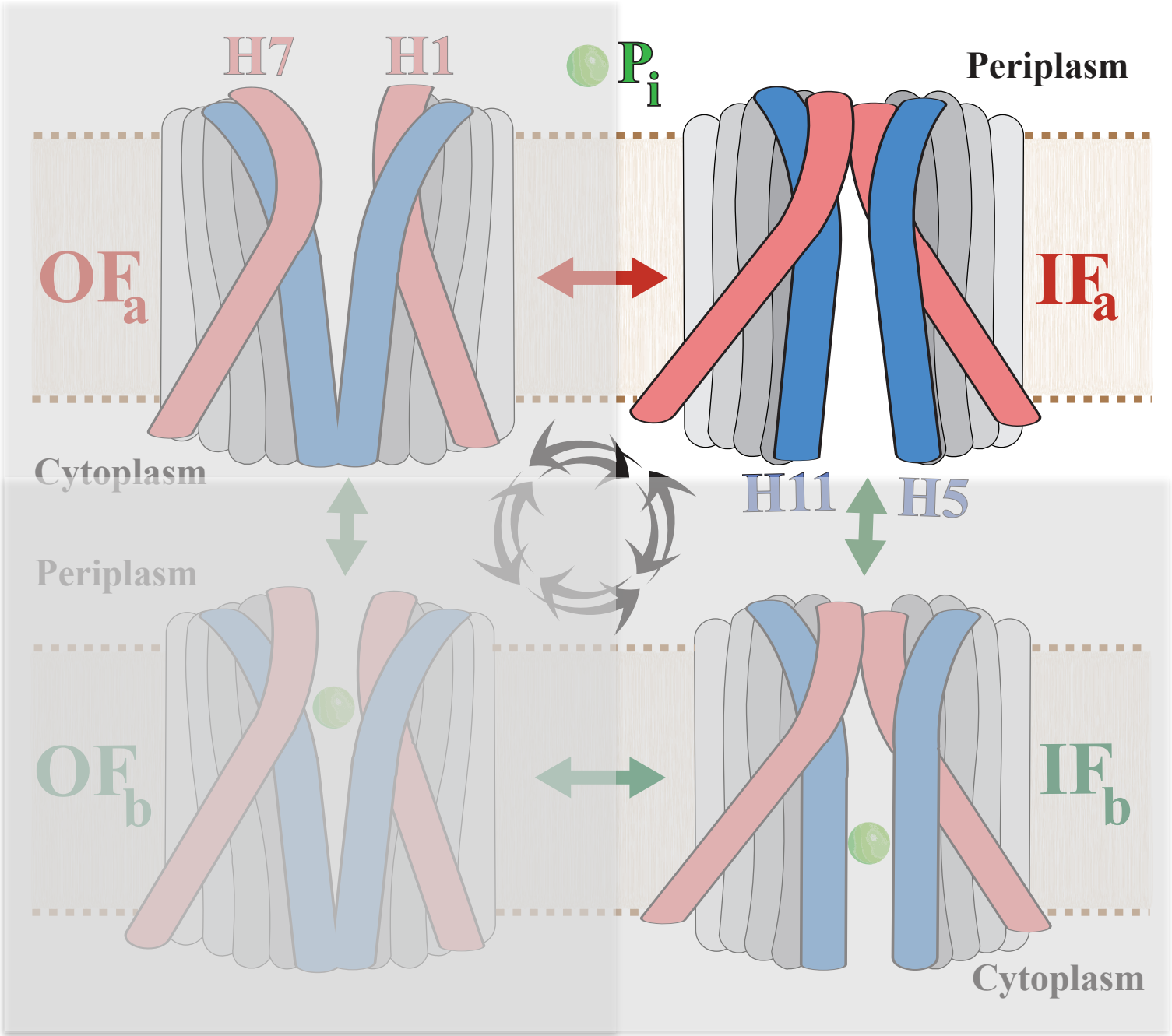
Lemieux, *et al.*, *Curr. Opin. Struct. Biol. 1***4**, 405 (2004).

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

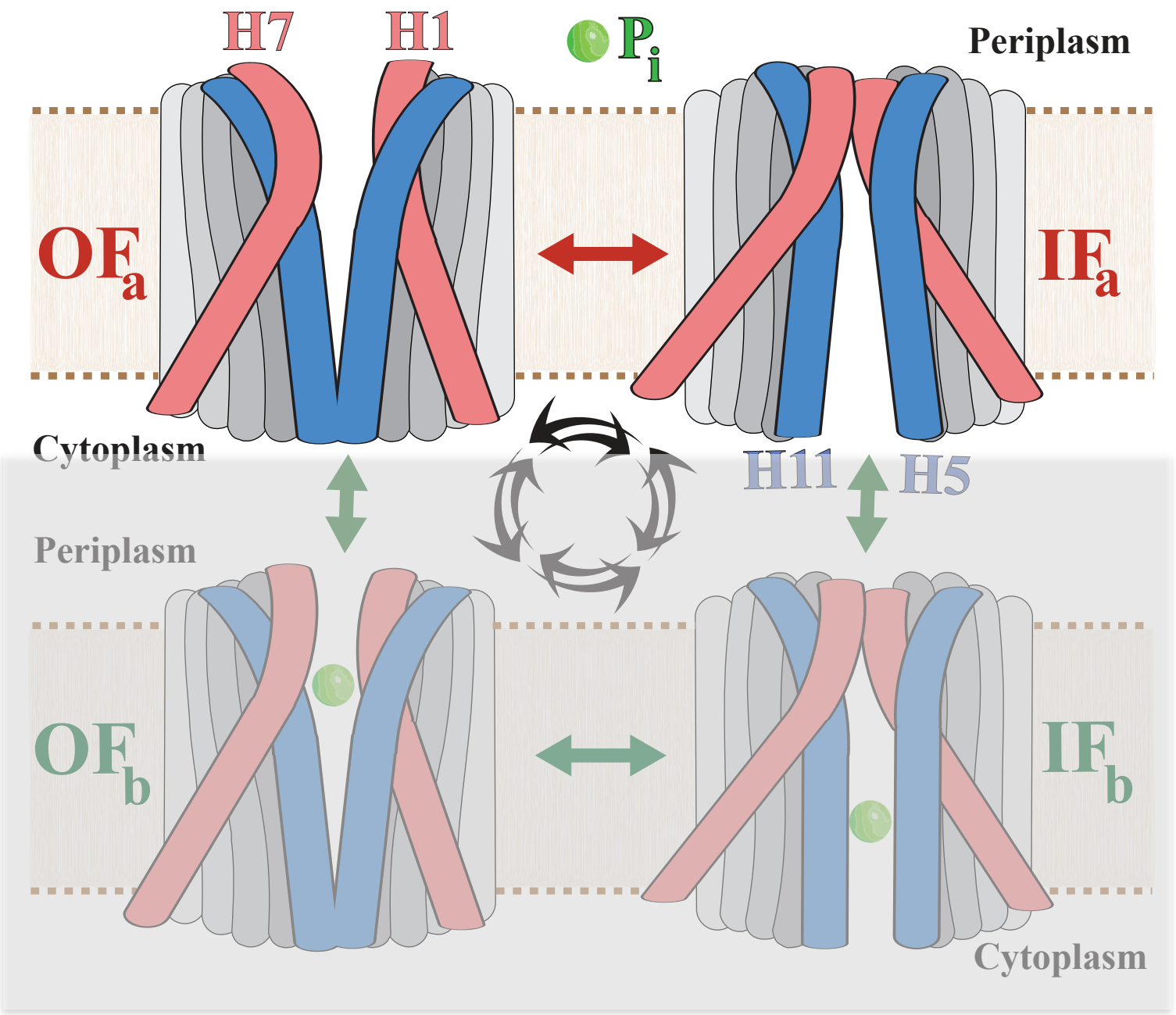
Full Thermodynamic Cycle



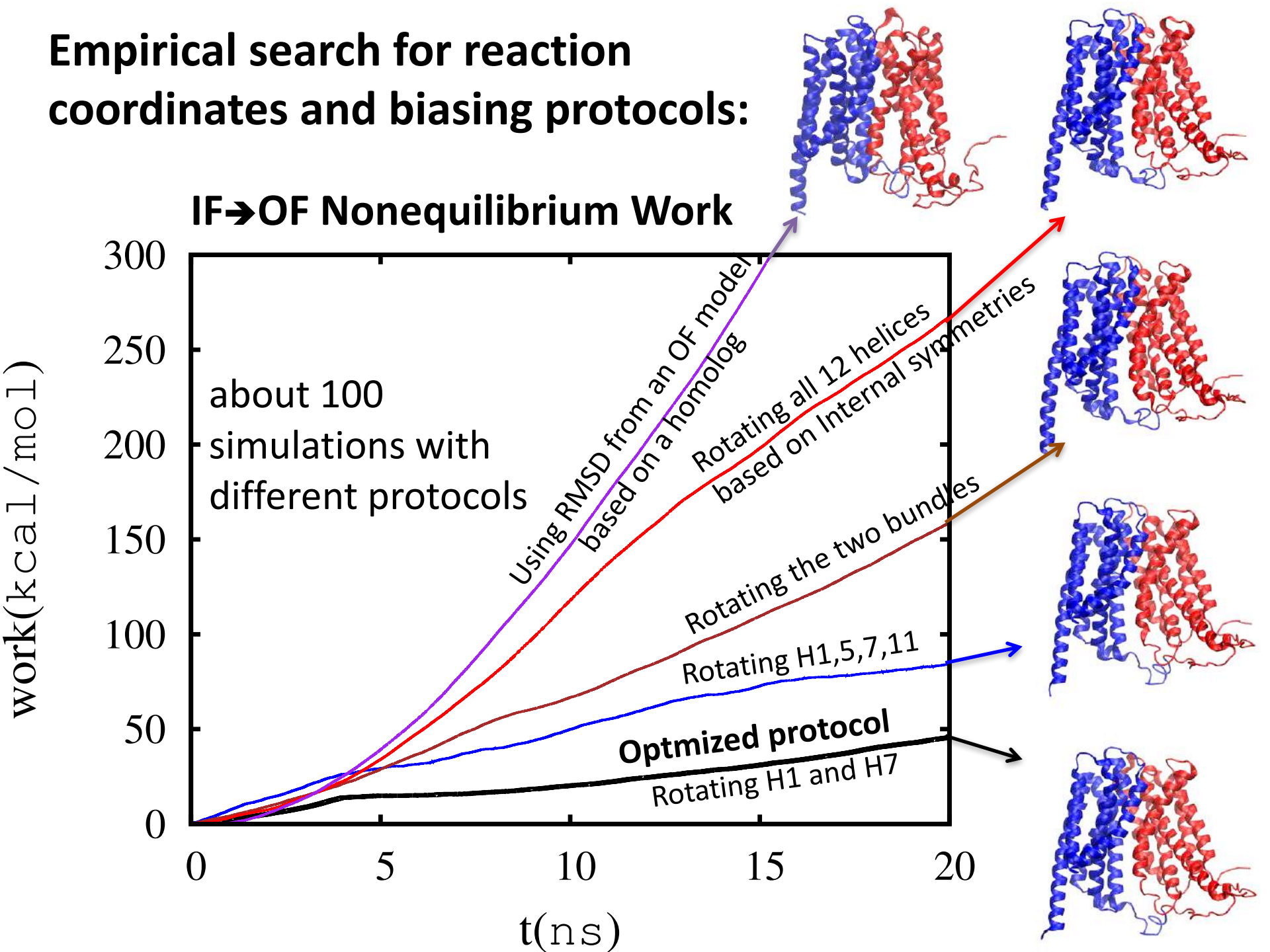
the only available crystal structure



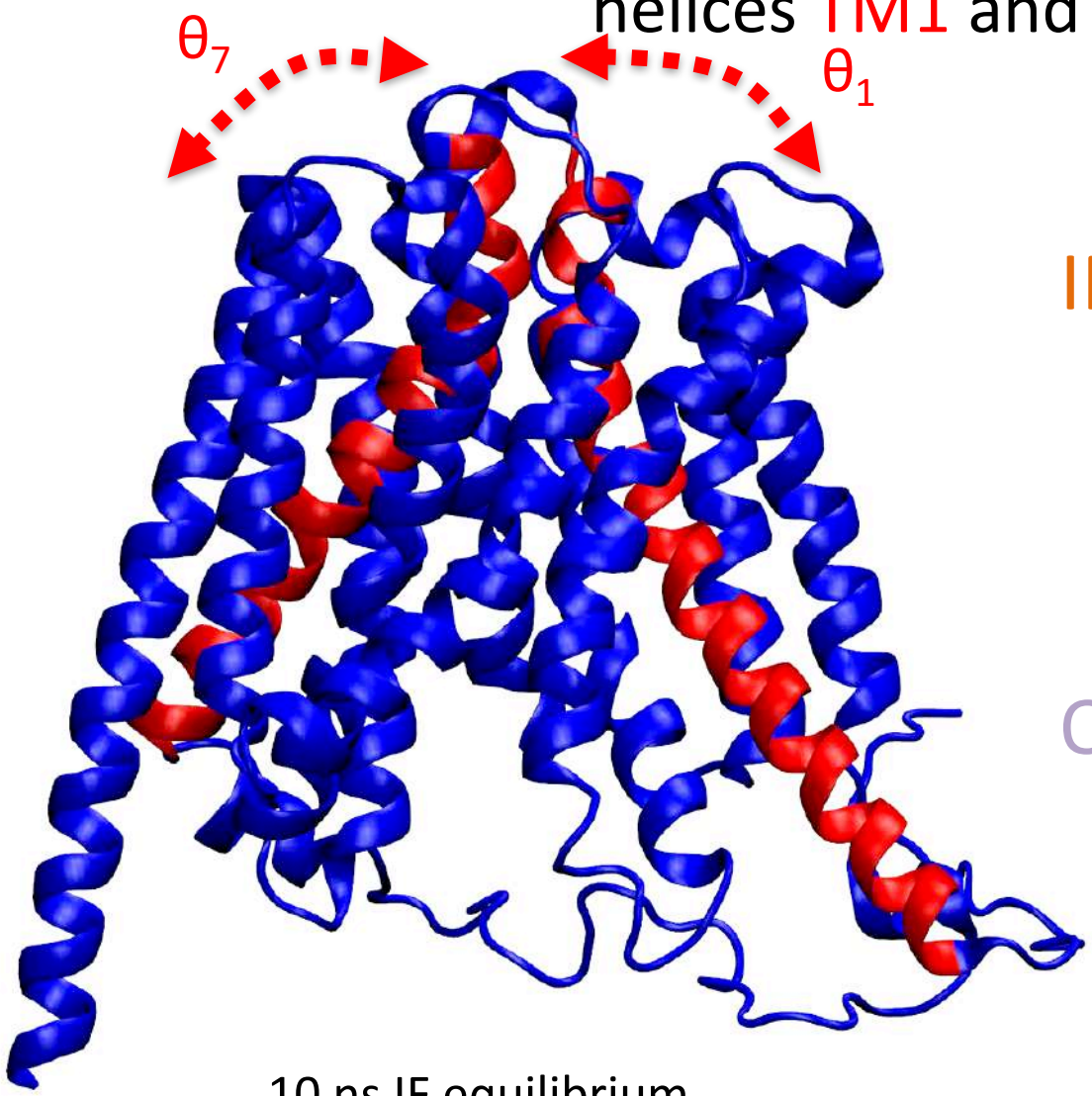
Step 1: $OF_a \leftrightarrow IF_a$

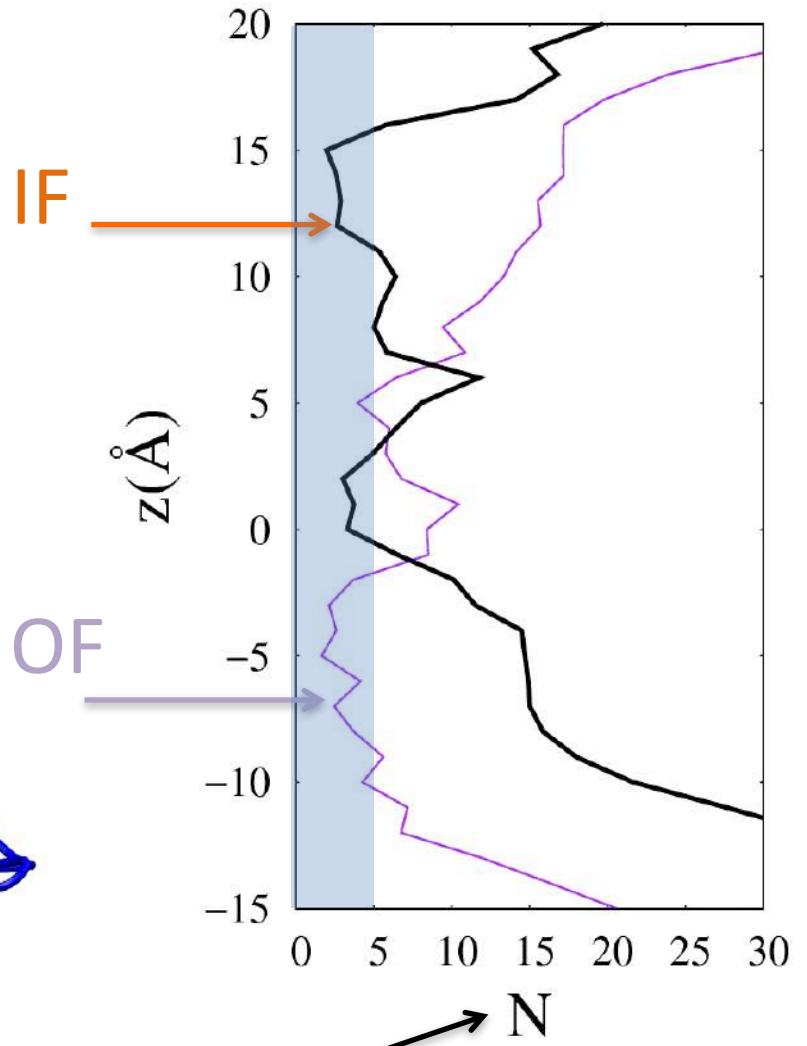


Empirical search for reaction coordinates and biasing protocols:



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



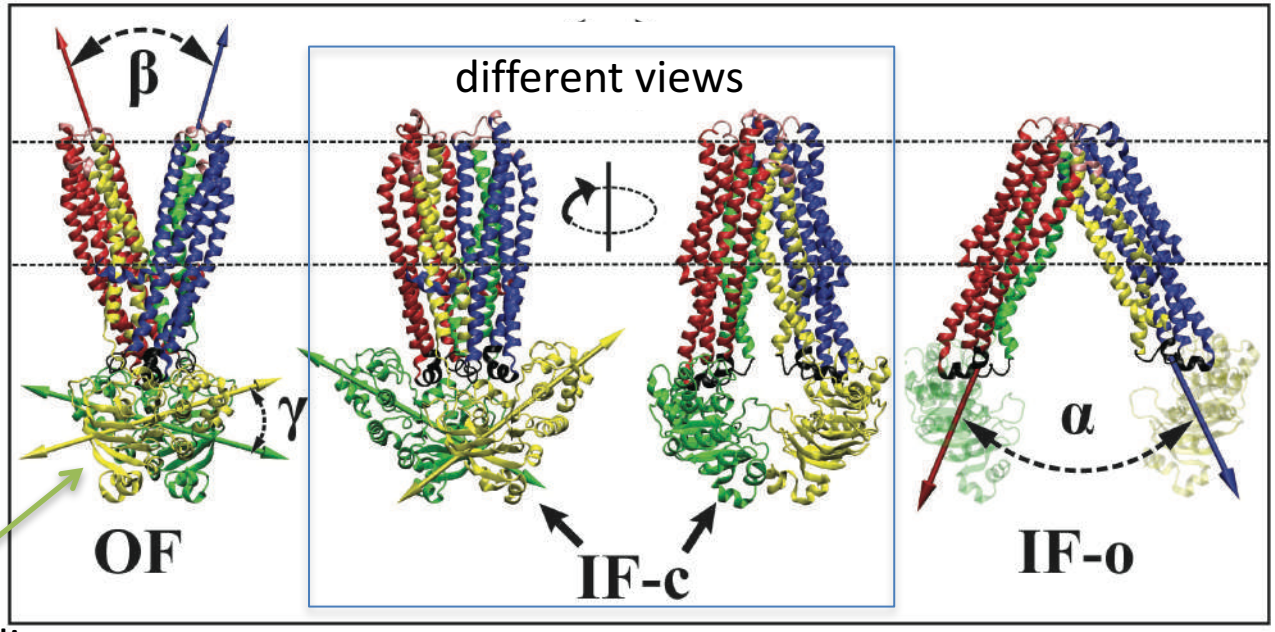


10 ns IF equilibrium
20 ns nonequilibrium (IF➔OF)
10 ns OF equilibrium

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)

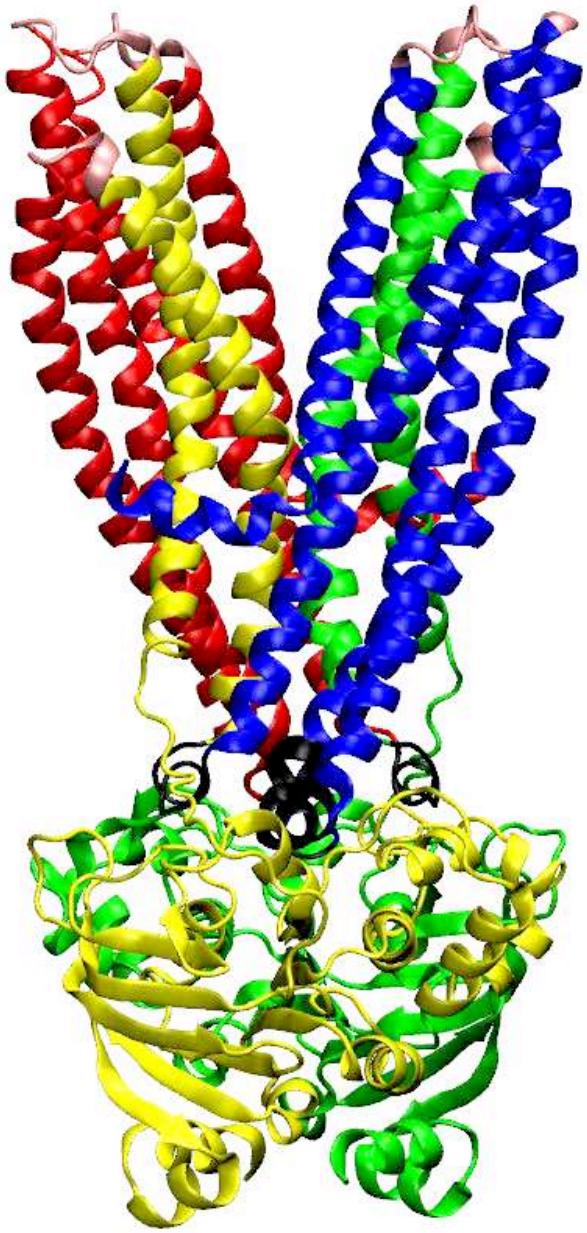


Nucleotide-binding domains (NBD)

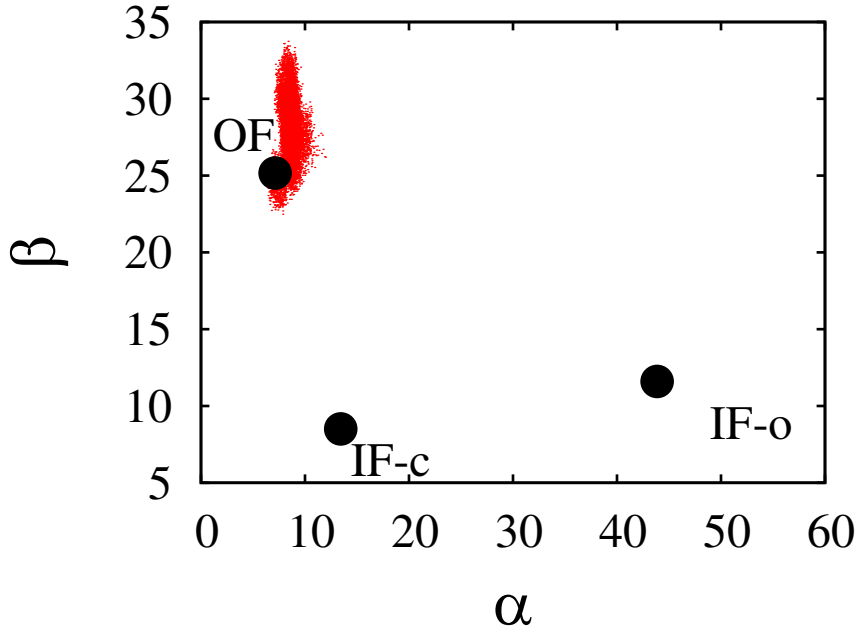
Two IF conformations:
IF-closed (IF-c) and IF-open (IF-o)

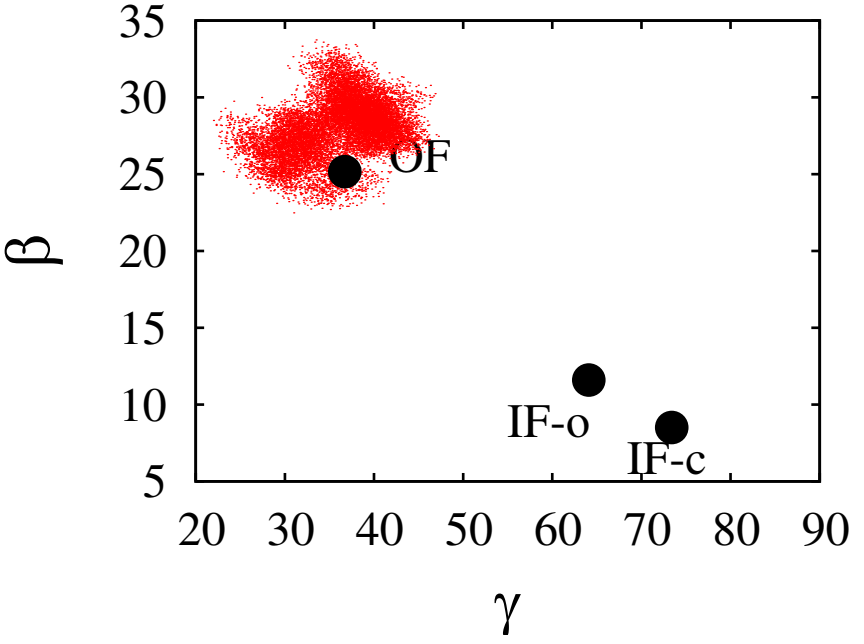
Moradi and Tajkhorshid, **PNAS 110** 18916 (**2013**)

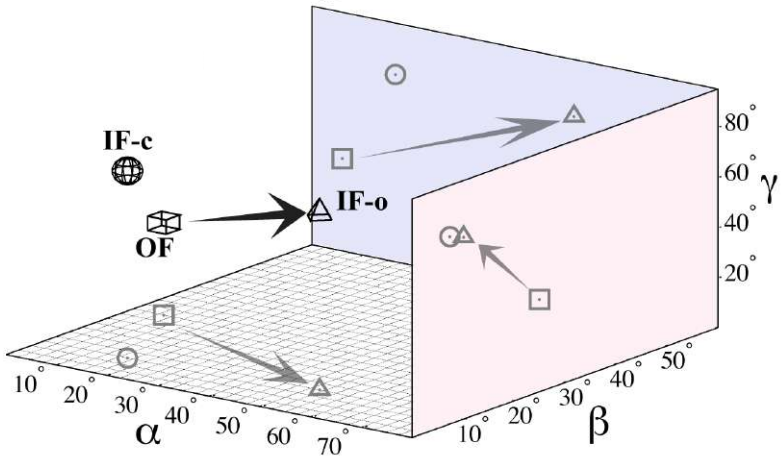
Conventional Equilibrium Molecular Dynamics

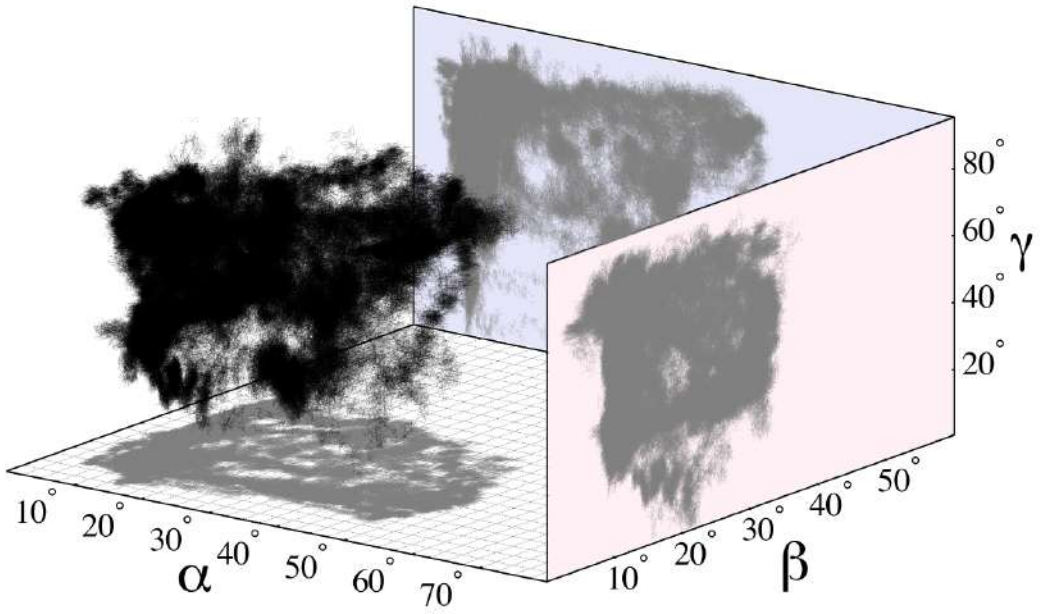


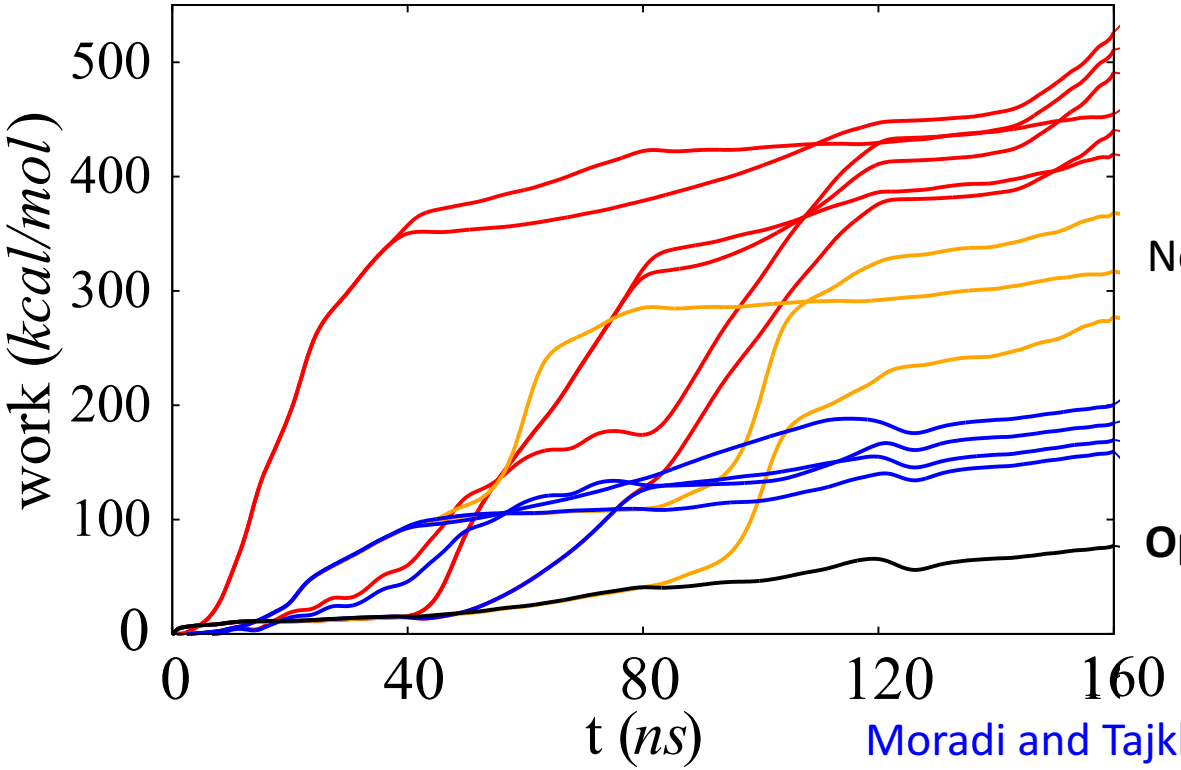
apo MsbA in explicit water/membrane (300 ns)







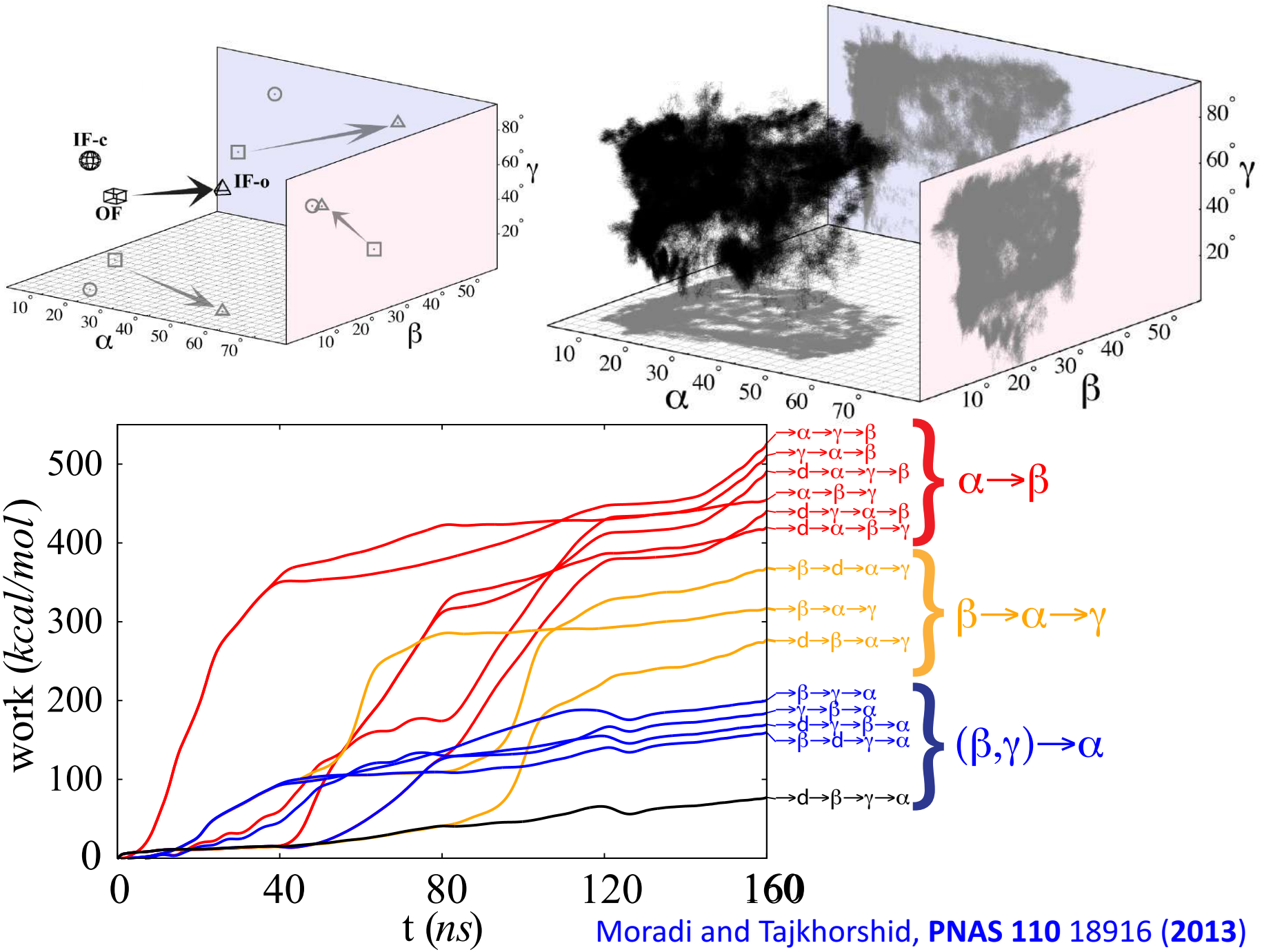




Nonequilibrium work
along
select trajectories

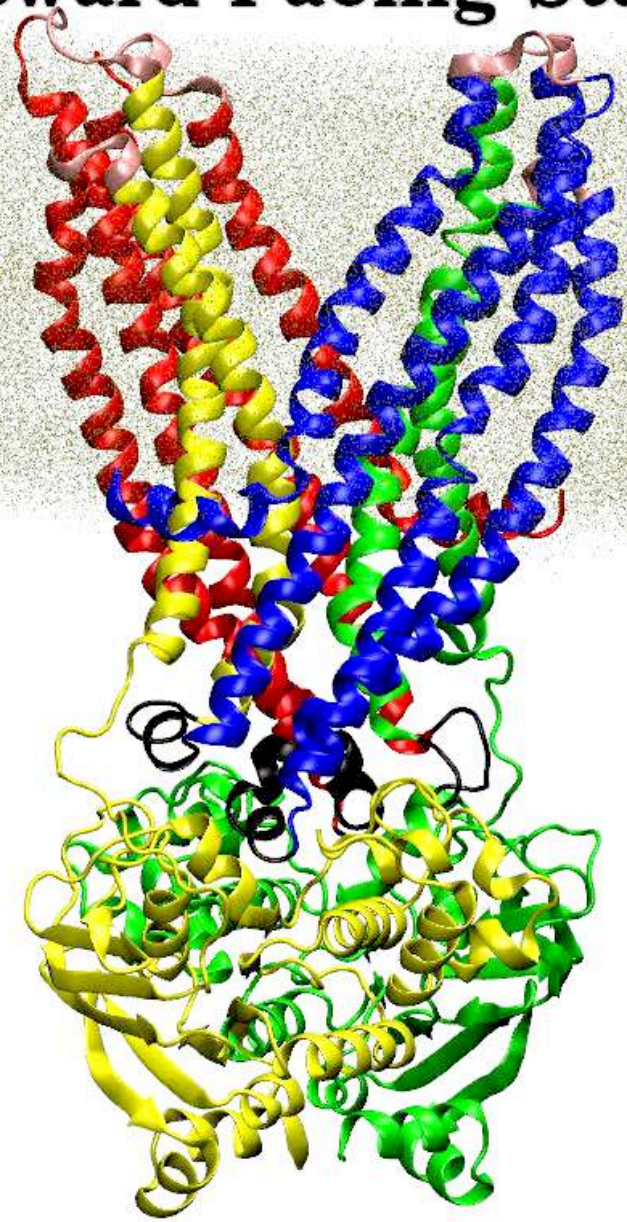
Optimized protocol

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

A quaternion is a 4-component mathematical object:

$$q = (q_0, q_1, q_2, q_3)$$

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

$$\langle |\hat{q}(0, \boldsymbol{x}_k)\hat{q}^* - (0, \boldsymbol{y}_k)|^2 \rangle$$

It turns out:

$$q = (\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\hat{\boldsymbol{u}})$$

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

Biasing Potential:

target quaternion
at time t

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

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

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:

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

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

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**

Non-Parametric Reweighting

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

Biased potential | Unbiased potential | Biasing potential

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

Biased density — Known

Unbiased density — Unknown

Biasing factor — Known

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

Partition function — Unknown

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

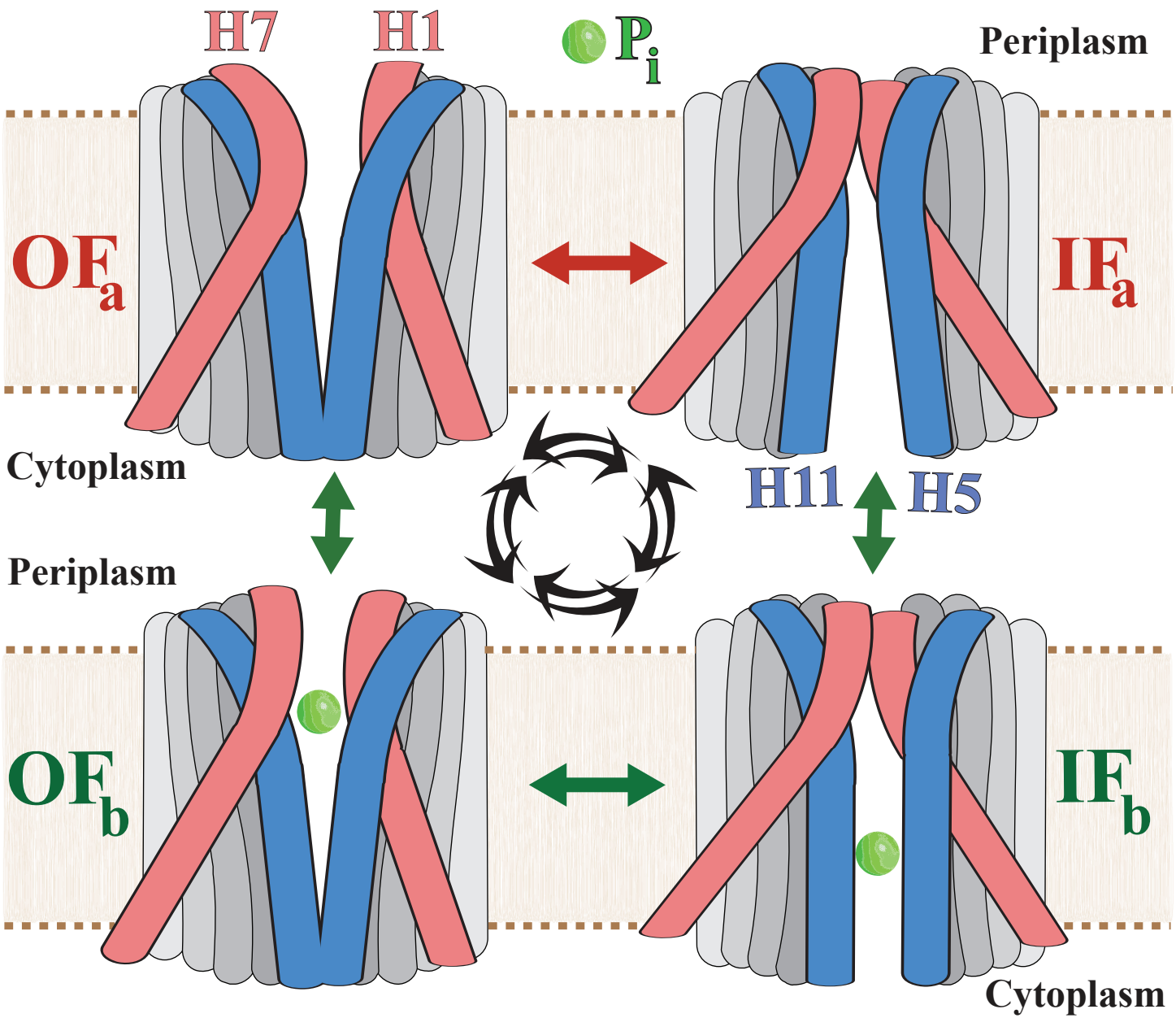
Shirts, Chodera, JCP, ***129***, 124105 (2008)

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

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**)

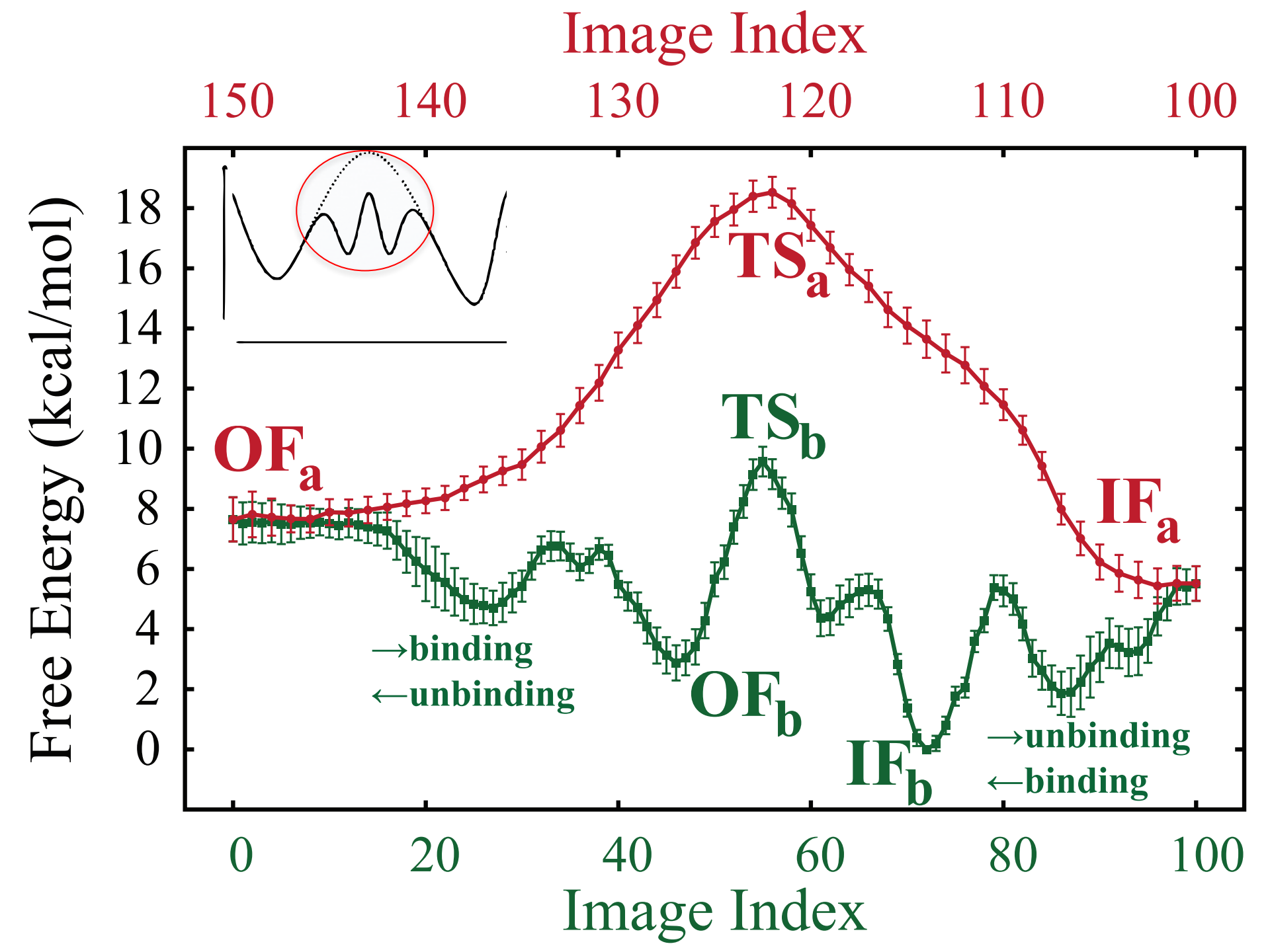
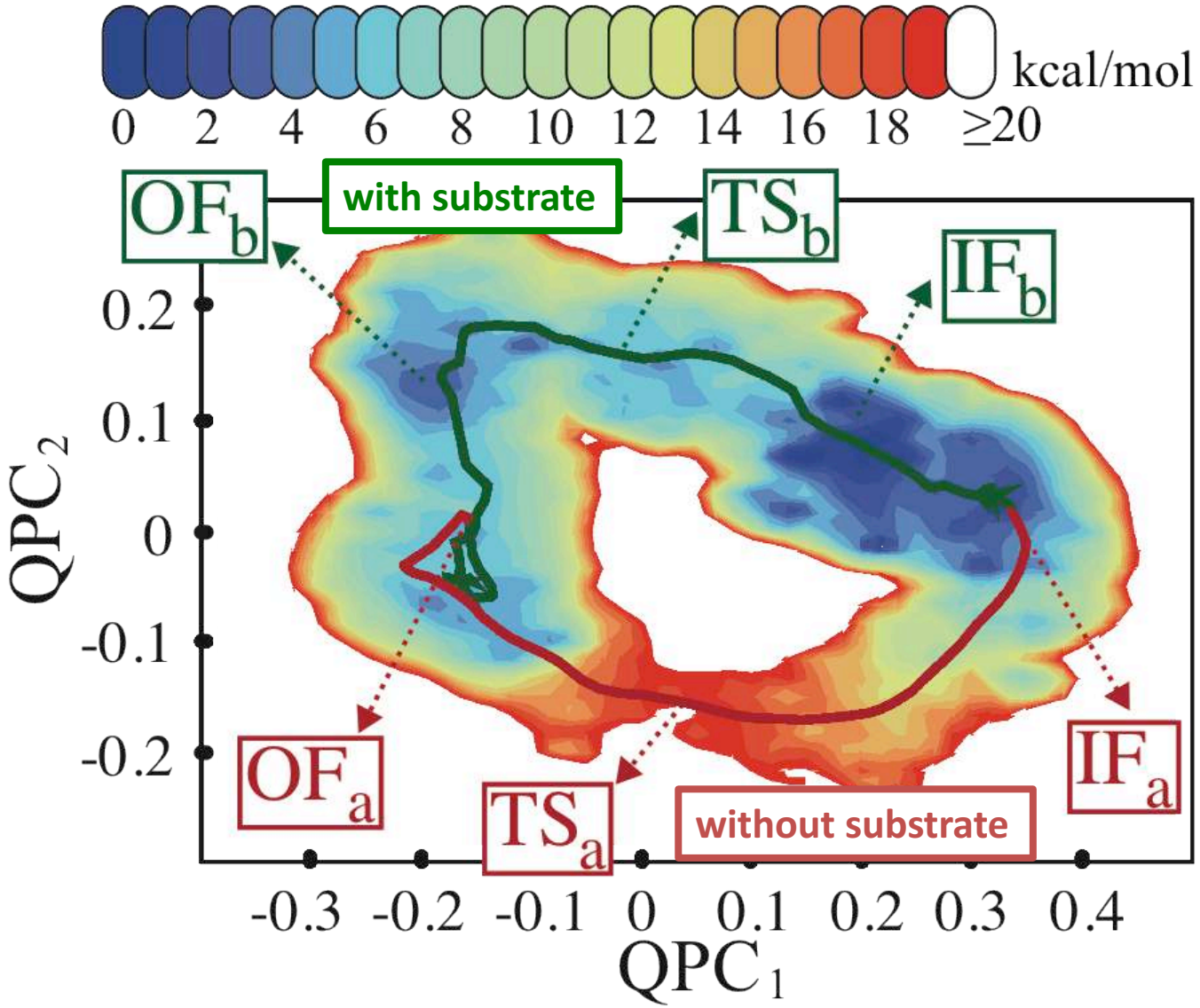


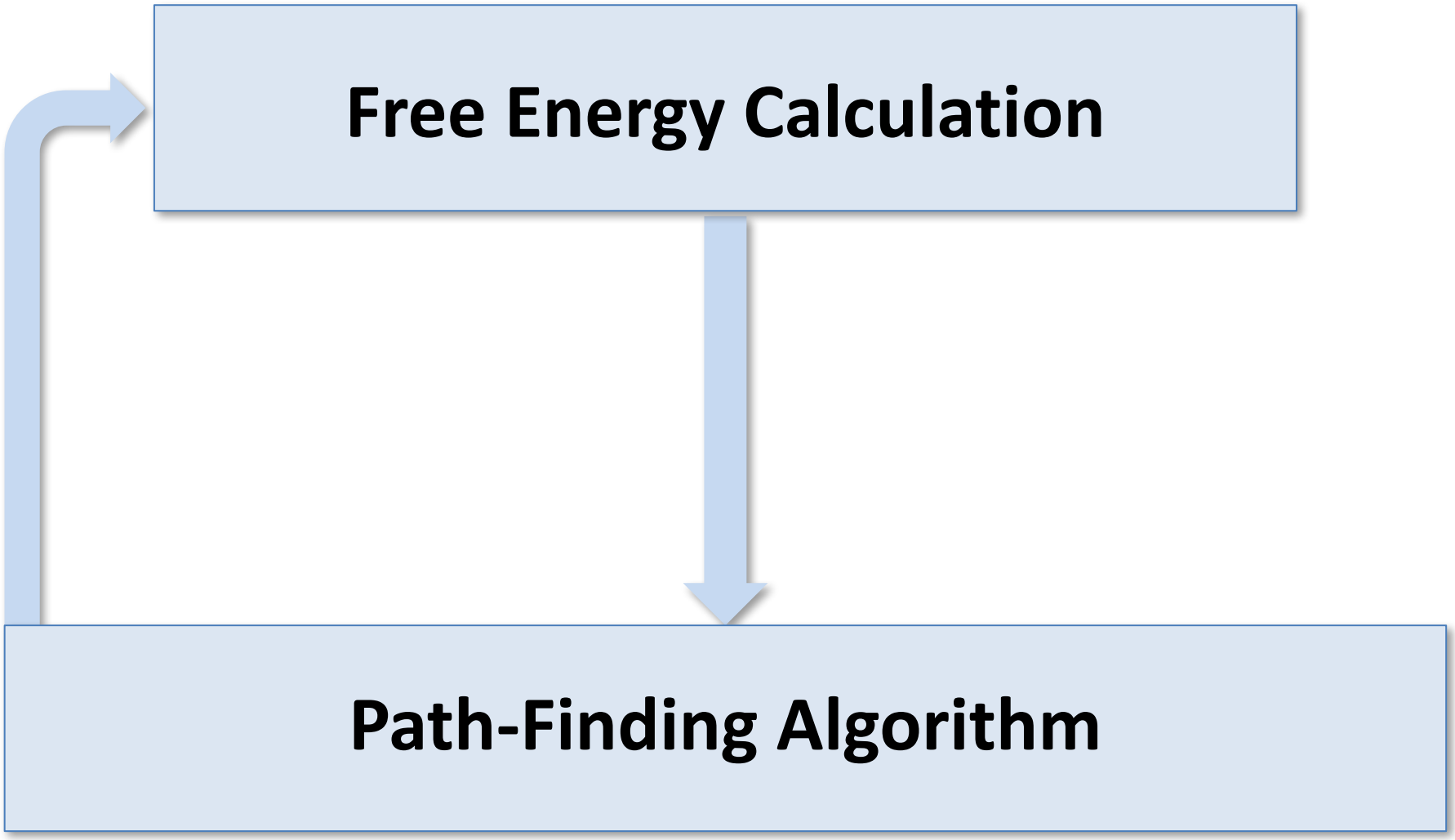
Image Index
150
140
130
120
110
100
Free Energy (kcal/mol)
18
16
14
12
10
8
6
4
2
0
TS$_a$
TS$_b$
OF$_a$
IF$_a$
→binding
←unbinding
OF$_b$
IF$_b$
→unbinding
←binding
0
20
40
60
80
100
Image Index

Distinct conformational transition pathways

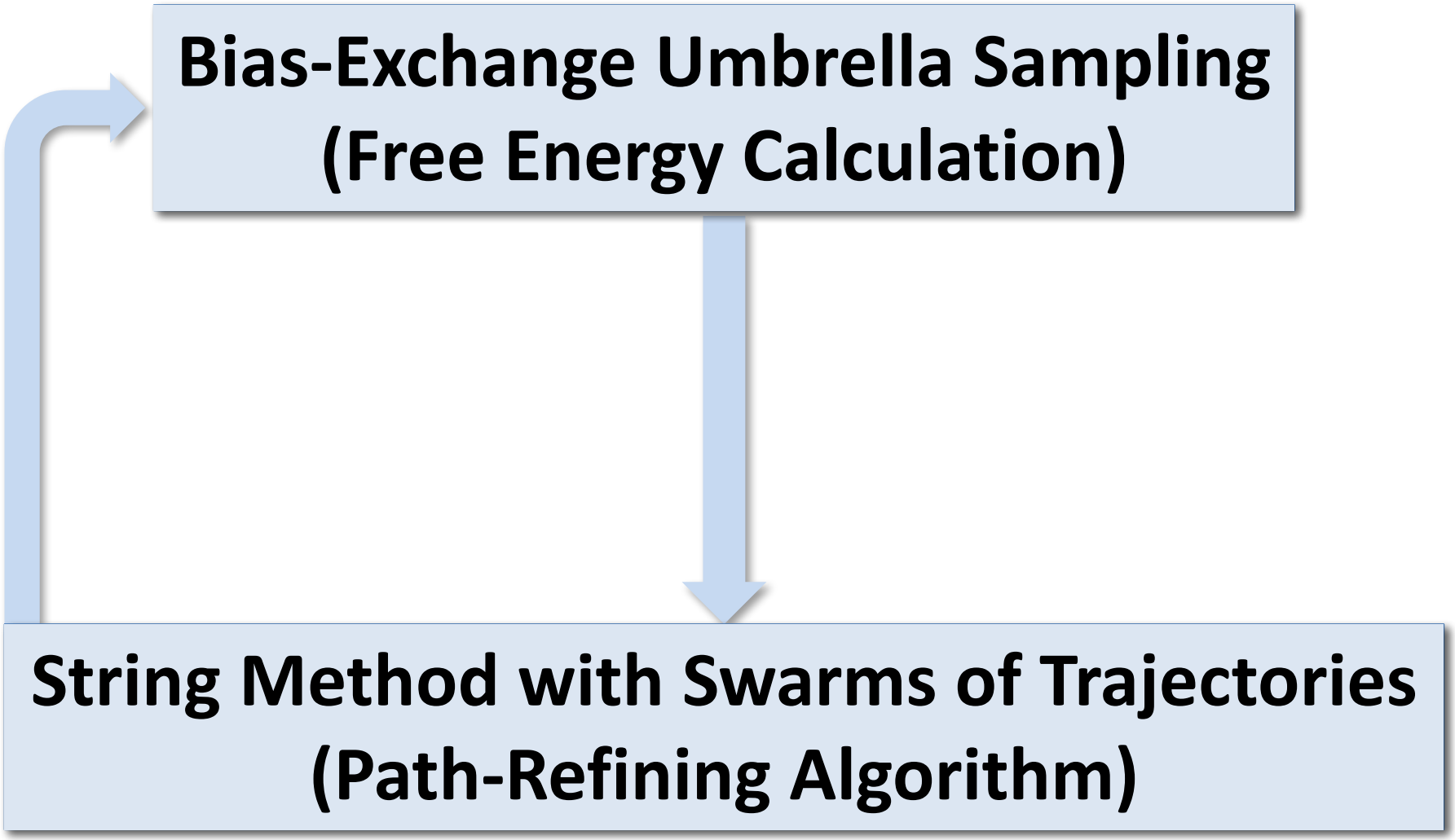
Quaternion-based principal components (**QPCs**) represent different modes of concerted motions of transmembrane helices.



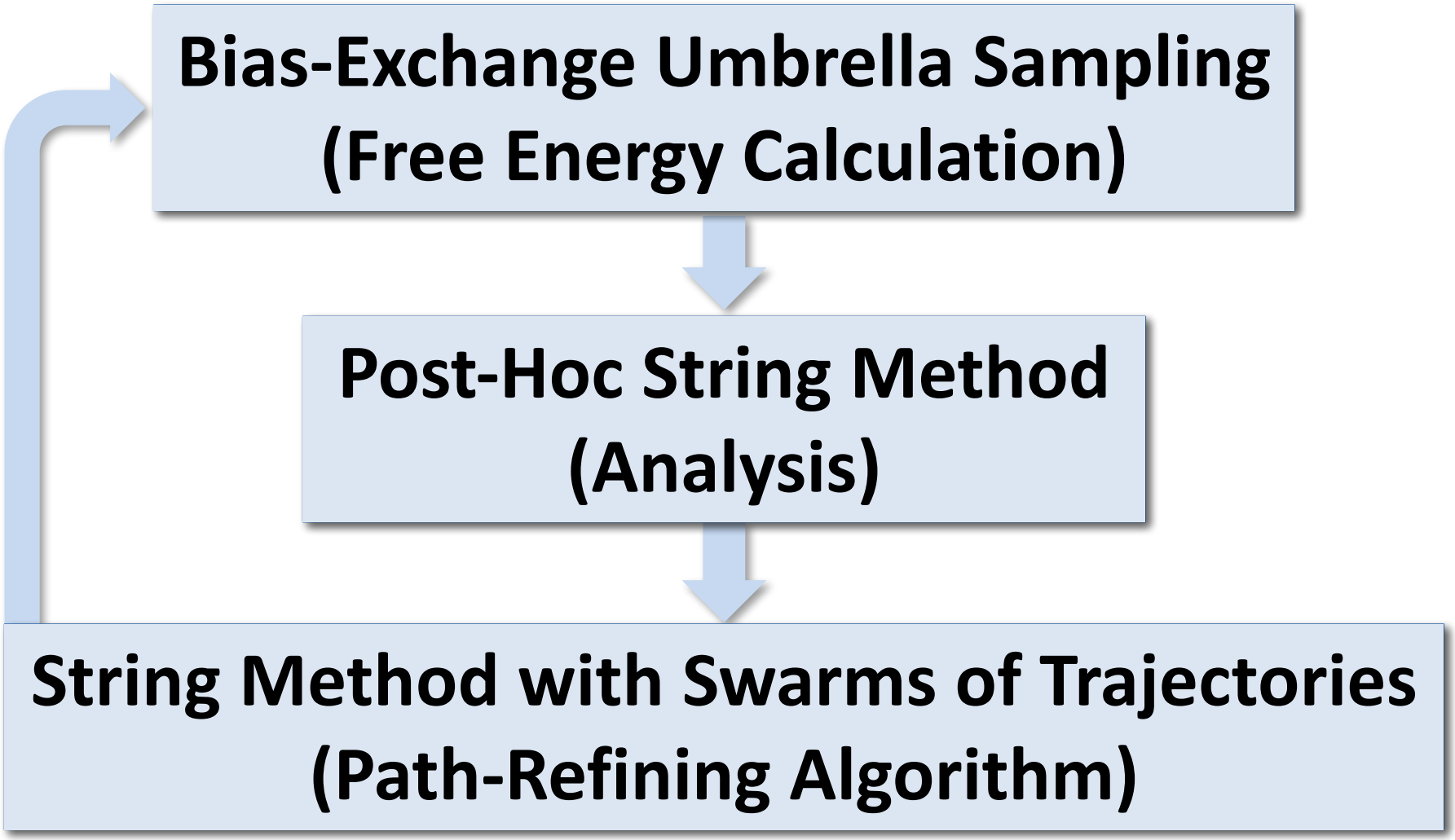
Iterative path-finding algorithms and free energy calculations



Iterative path-refining algorithms and free energy calculations



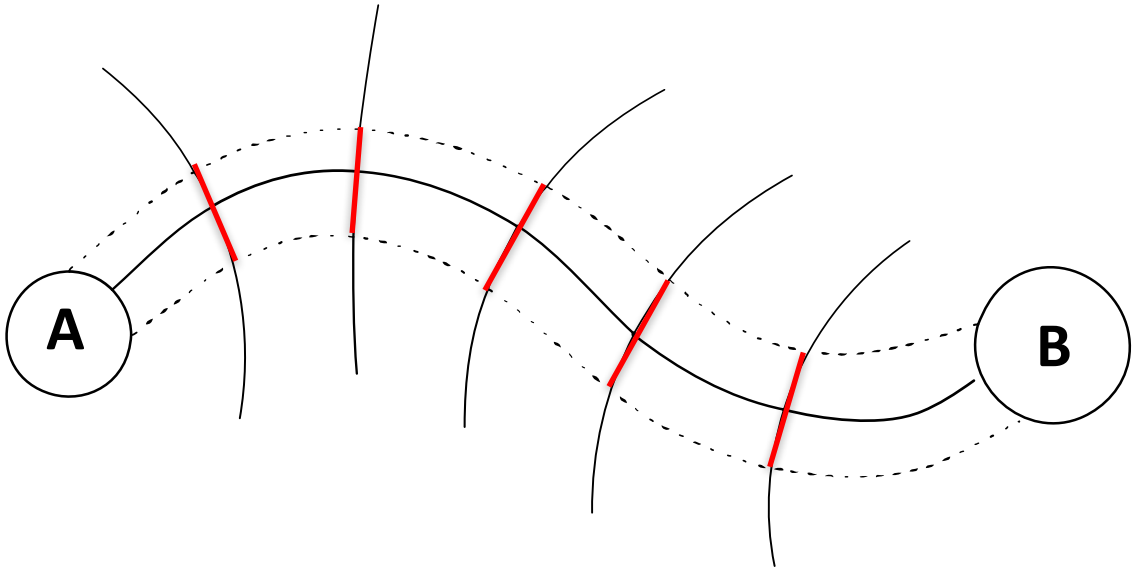
Iterative path-refining algorithms and free energy calculations



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**)