Simulation of drug targets and simulation analysis

Ron Dror

Jan. 16, 2018

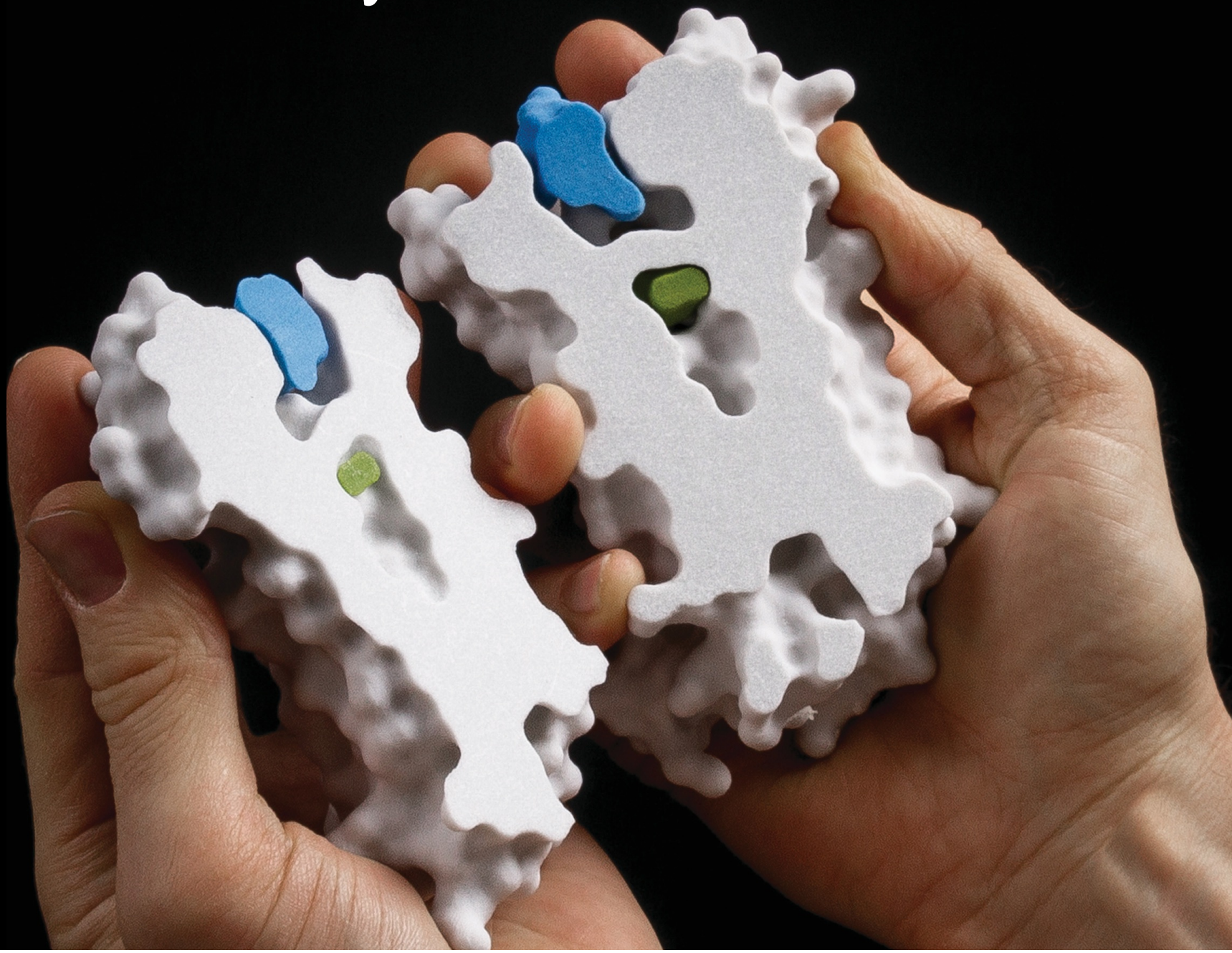


Image credit:
Ansgar Philippsen

Overview

- Molecular dynamics (MD) simulations predict the atomic-level motions of molecules

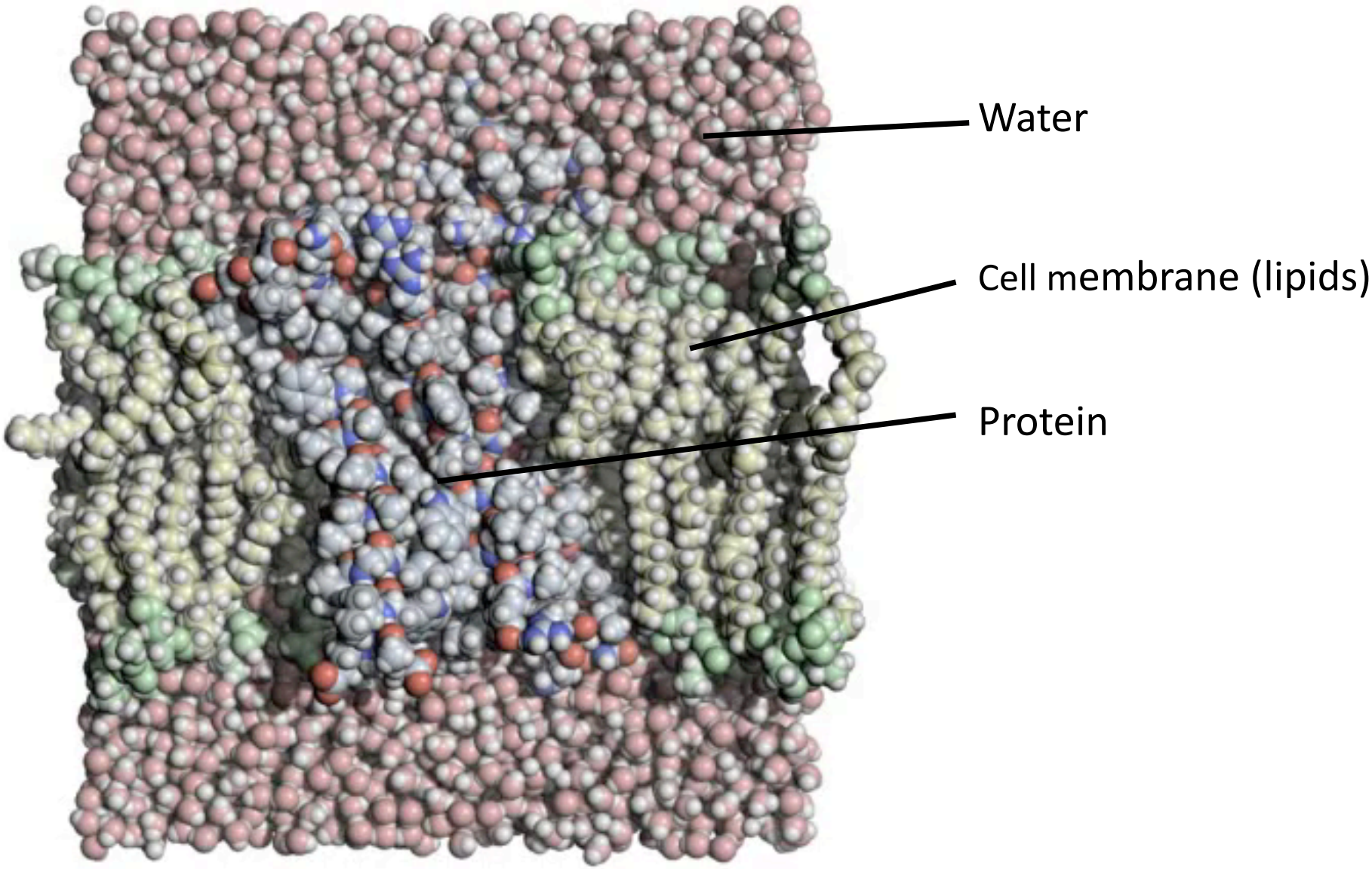
- These simulations have been around for decades, but they've become much more powerful recently, thanks to faster computers, better algorithms, and better models of the underlying physics
  - First two papers illustrates application of modern MD to problems of interest in drug discovery

- MD simulations generate a lot of data, and extracting all the important information from that data is challenging
  - Third paper presents a statistical method for detecting important "events" in MD simulations

Background:
Molecular dynamics (MD) simulations

An MD simulation computes the motion of every atom in a molecular system

For example, a protein and its surroundings

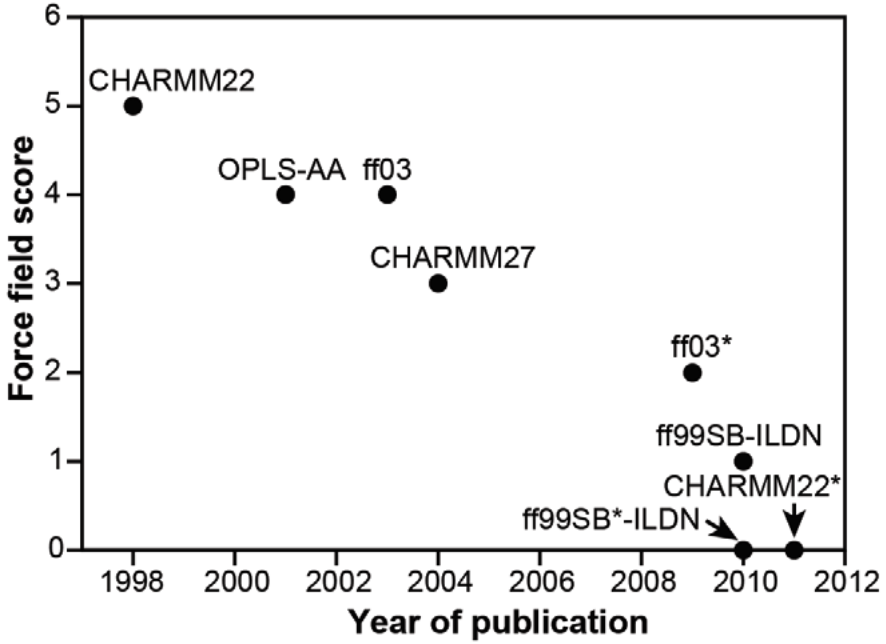


MD simulations are slow, but they've sped up substantially

- Simulations require short time steps for numerical stability
  - 1 time step ≈ 2 fs (2×10^{-15} s)
- Structural changes in proteins can take nanoseconds (10^{-9} s), microseconds (10^{-6} s), milliseconds (10^{-3} s), or longer
  - Millions to trillions of sequential time steps for nanosecond to millisecond events
- Until recently, simulations of 1 microsecond were rare
- Advances in computer power have enabled microsecond simulations (and even millisecond simulations in special cases)
- Enabling longer simulations is an active research area, involving:
  - Algorithmic improvements
  - Parallel computing
  - Hardware: GPUs, specialized hardware

MD simulations are approximate, but they've become more accurate

- Molecular mechanics force fields (the functional forms used to compute forces on atoms) are inherently approximations
- They have improved substantially over the last decade, though many limitations remain



Here force fields with lower scores are better, as assessed by agreement between simulations and experimental data. Even the force fields with scores of zero are imperfect, however!

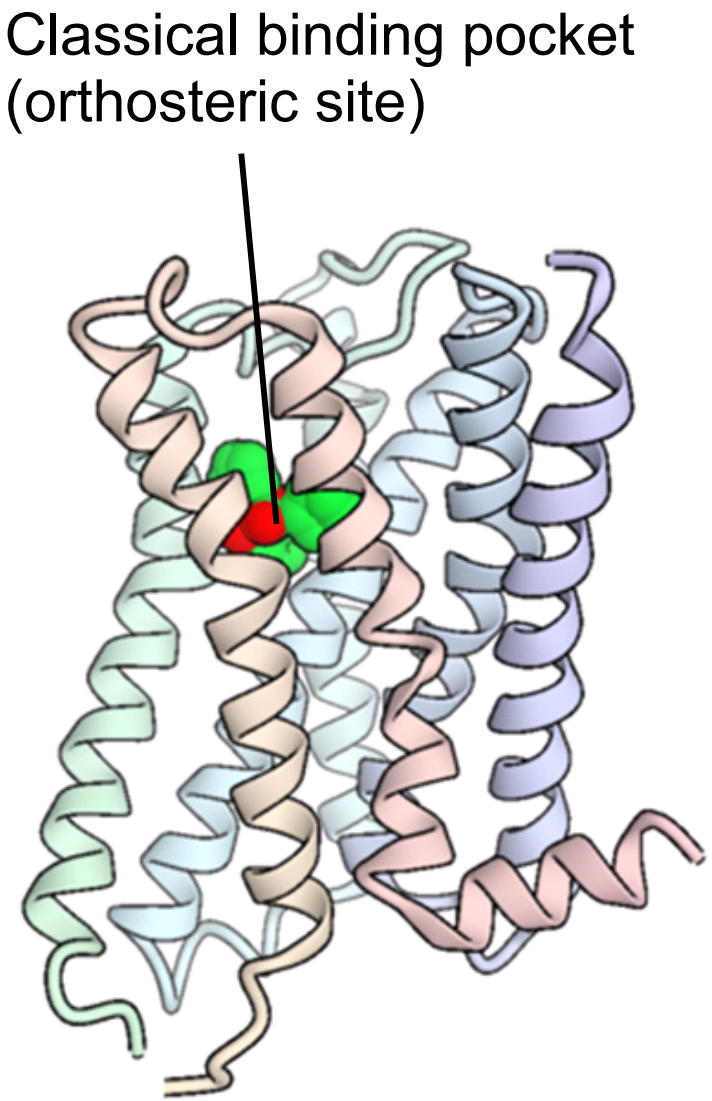
Lindorff-Larsen et al., *PLOS One*, 2012

Application #1 of MD simulation:

“Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs”

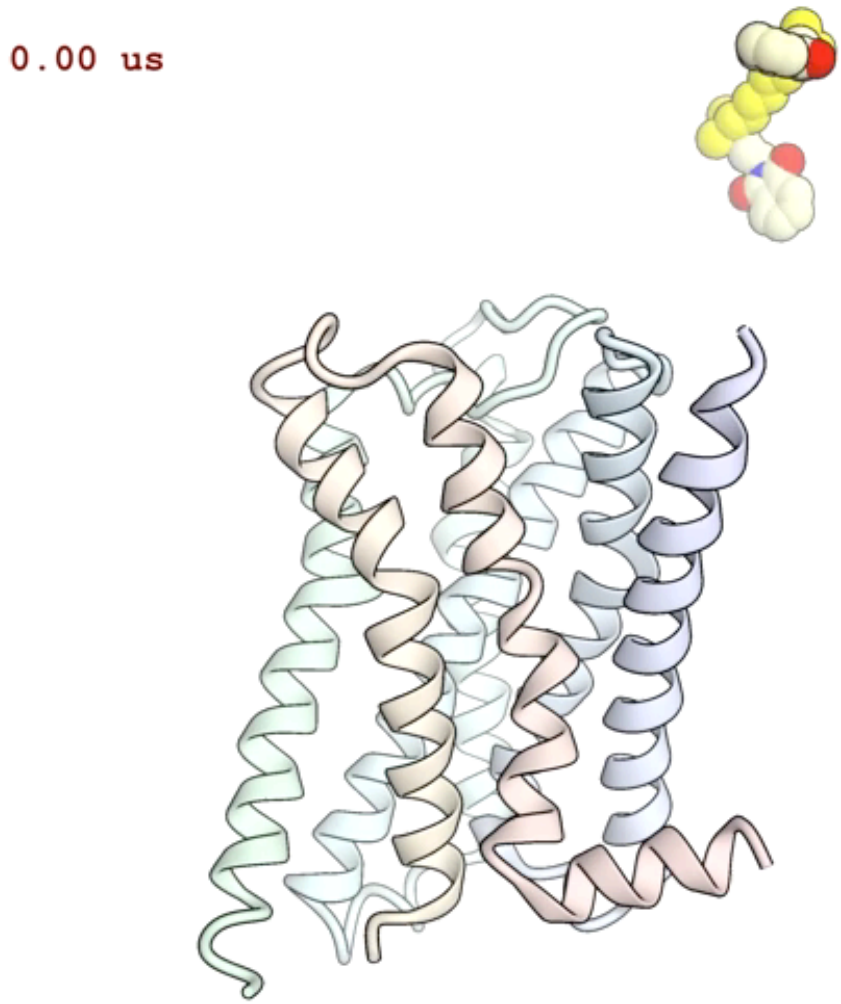
Dror et al., *Nature* 503: 295–299 (2013)

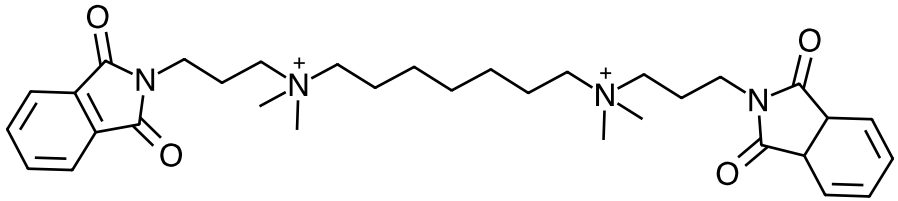
Allosteric modulators of GPCRs



- One-third of all drugs target G protein–coupled receptors (GPCRs)
- Most of these bind in the classical binding pocket (orthosteric site)
- Allosteric modulators bind *anywhere else*
- They're of great interest as drugs because they promise:
  - Selectivity between GPCR subtypes
  - Fine control of responses to body's natural signaling patterns
- Until recently, not clear how allosteric modulators bound, and even less clear how they exert their effects

Allosteric modulator binding to muscarinic acetylcholine receptor

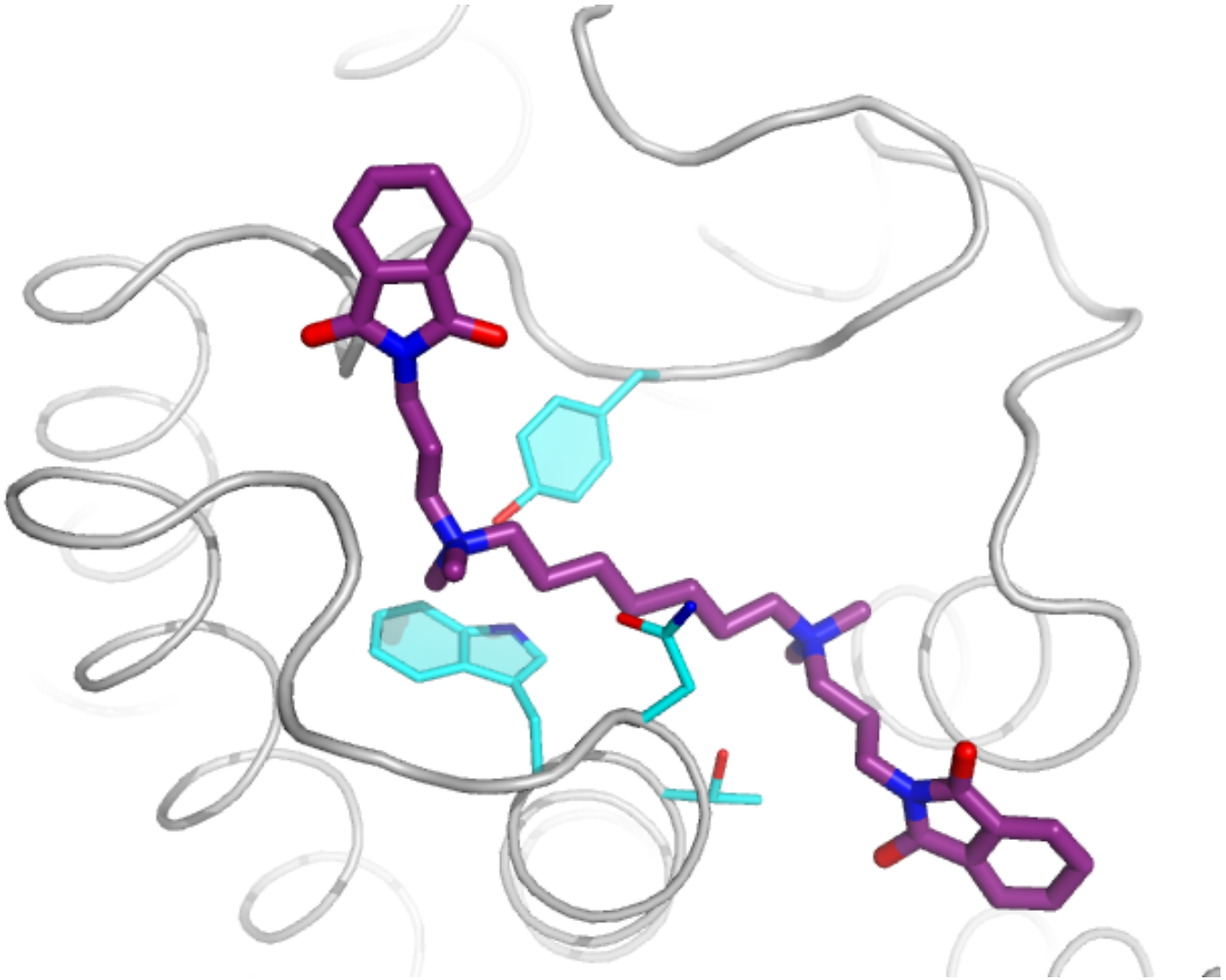




C_7/3-phth binding to M2 muscarinic receptor

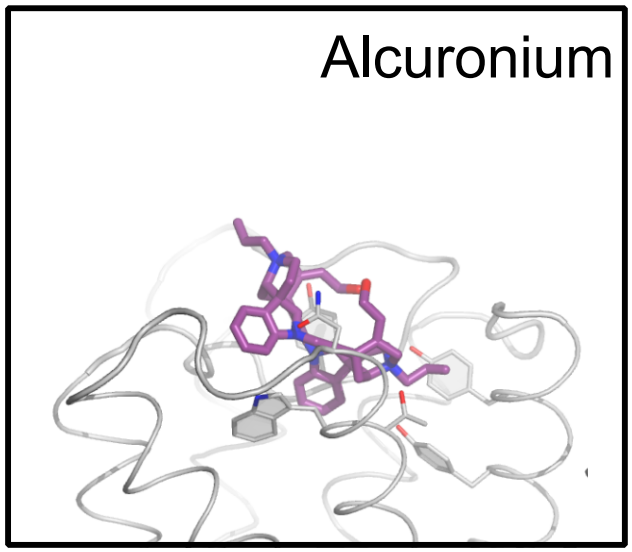
Dror et al., *Nature* 2013

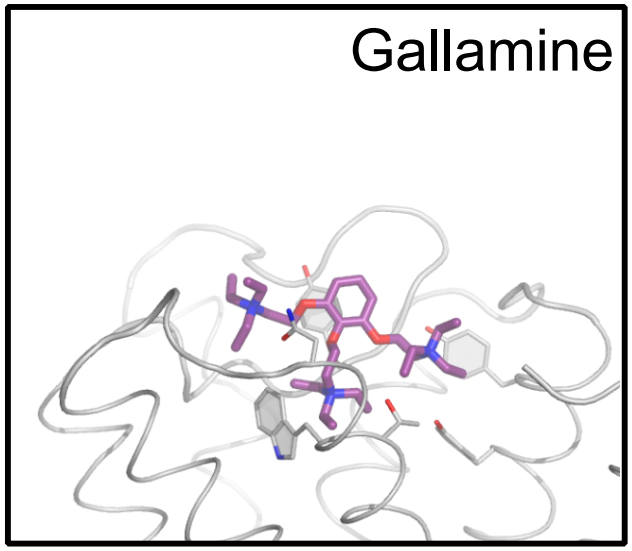
Bound pose agrees with mutagenesis data

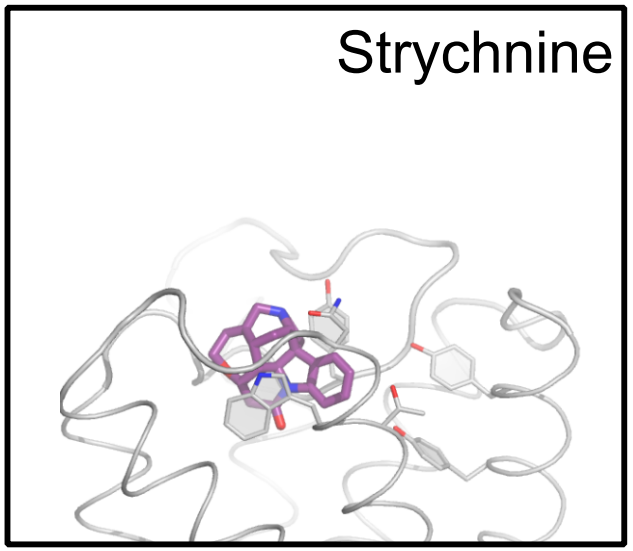


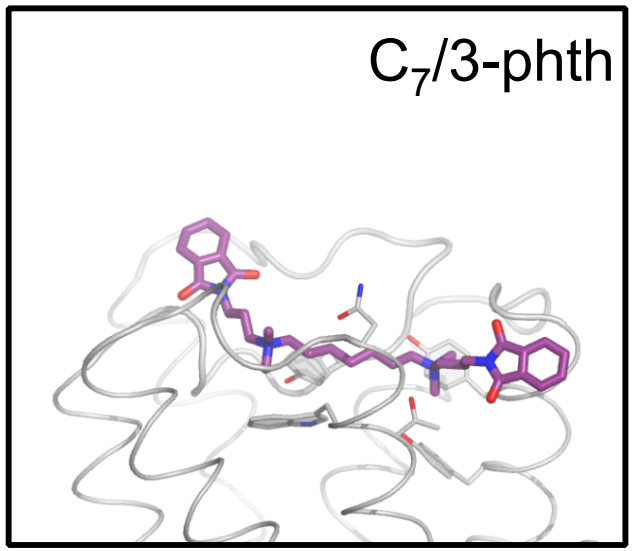
- Amino acids whose mutation caused >5-fold loss of affinity for bis-amino alkane ligands

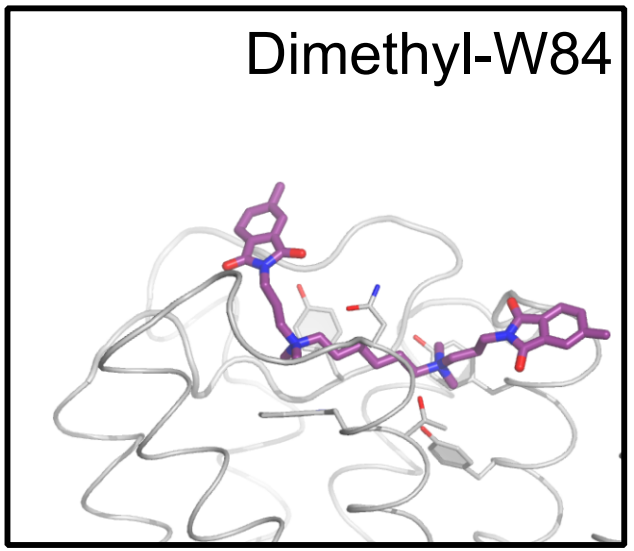
Apply the same methodology to structurally diverse allosteric modulators ...







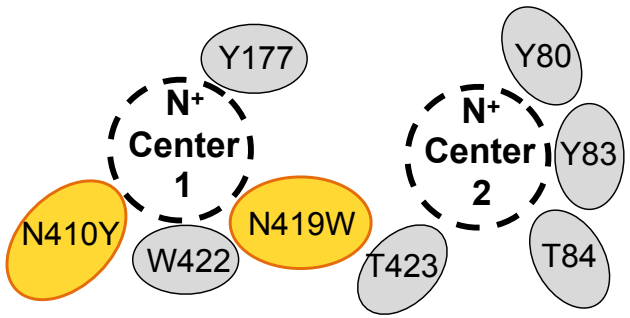




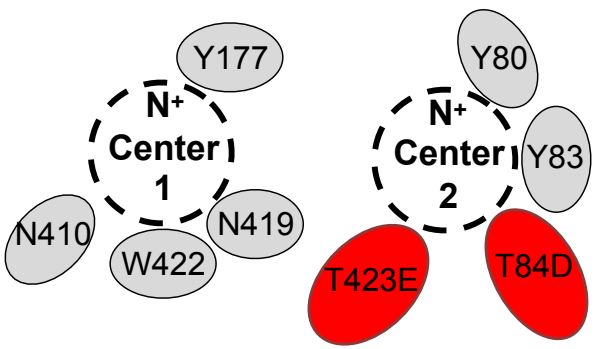
Experimental validation of new predictions

Mutagenesis predictions

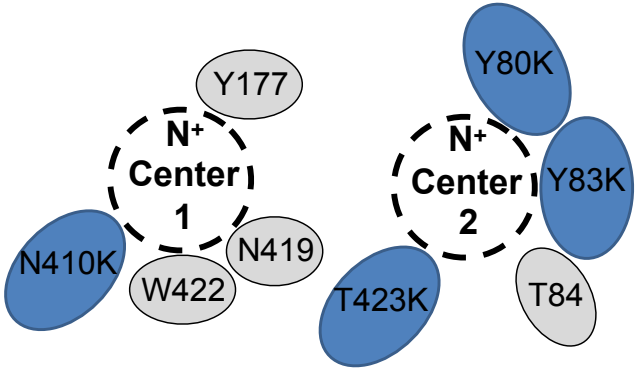
Increase affinity
Increased cation–π interactions



Increase affinity
Charge–charge attraction

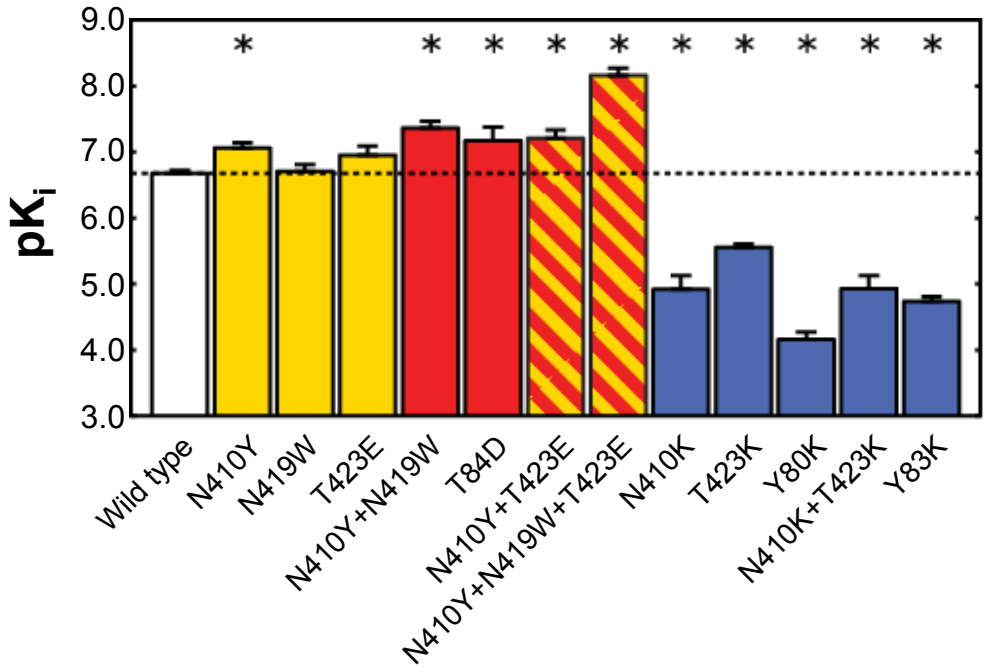


Decrease affinity
Charge–charge repulsion

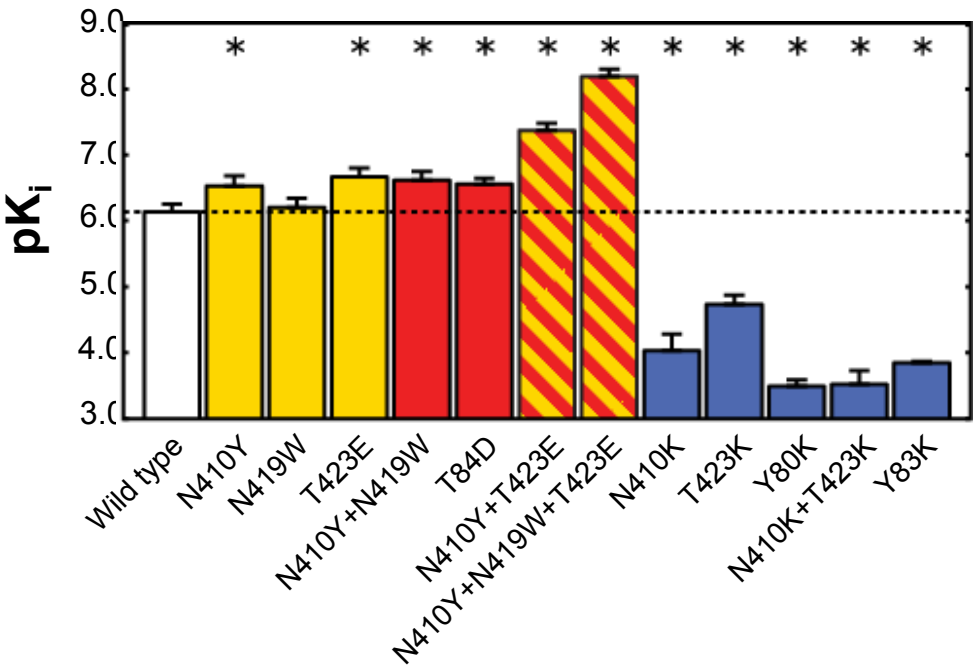


Mutagenesis results (Christopoulos lab, Monash University)

C_7/3-phth



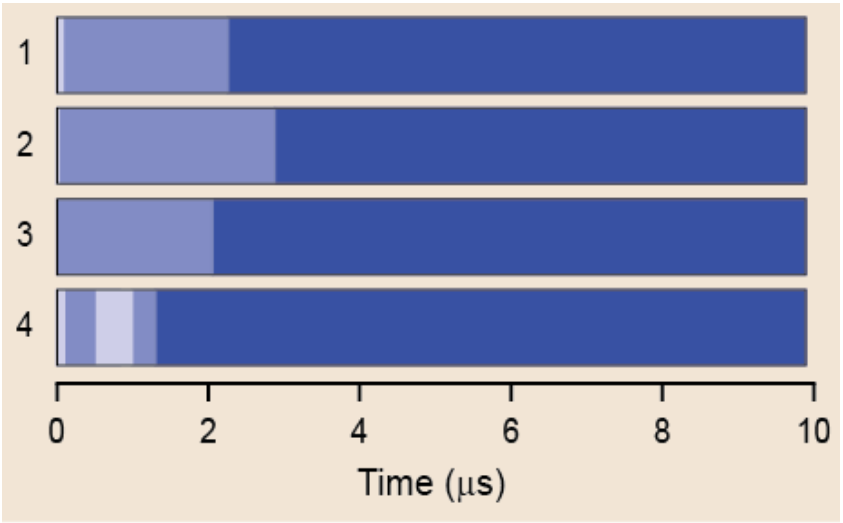
Gallamine



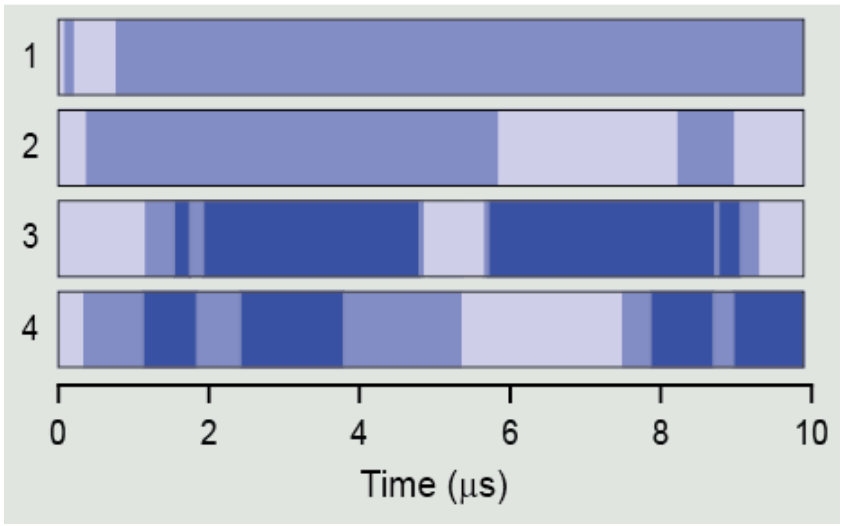
What is the basis of allosteric modulation?

- In simulation, as in experiment, allosteric ligands modulate affinity of orthosteric ligands.
  - C_7/3-phth, a negative allosteric modulator (NAM), lowers affinity of the classical antagonist NMS

No ligand in orthosteric site



NMS in orthosteric site



Allosteric ligand tightly bound

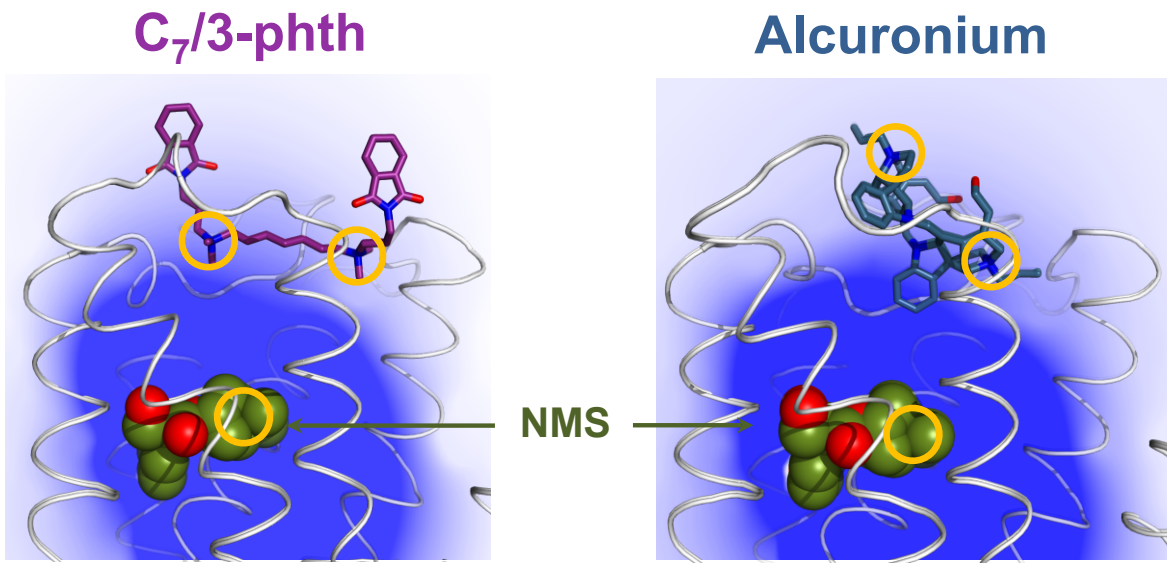
Allosteric ligand loosely bound

Allosteric ligand not bound

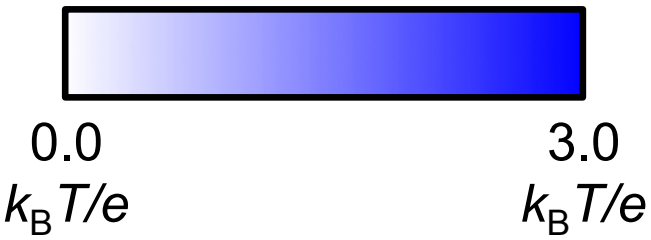
Dror et al., *Nature* 2013

Mechanism 1: Electrostatic interaction between ligands

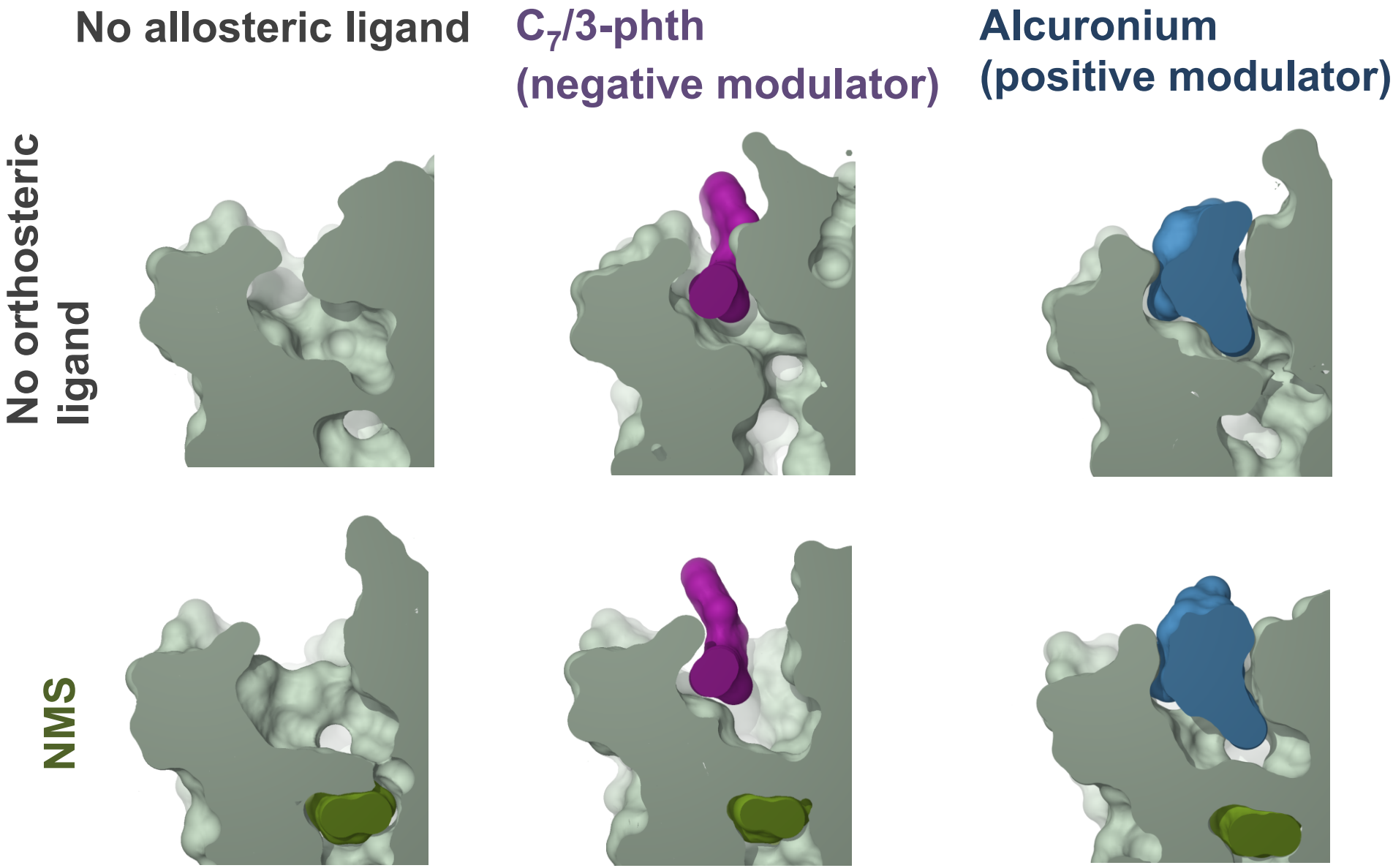
- Positively charged allosteric ligand repels positively charged orthosteric ligand



Difference in electrostatic potential due to addition of NMS

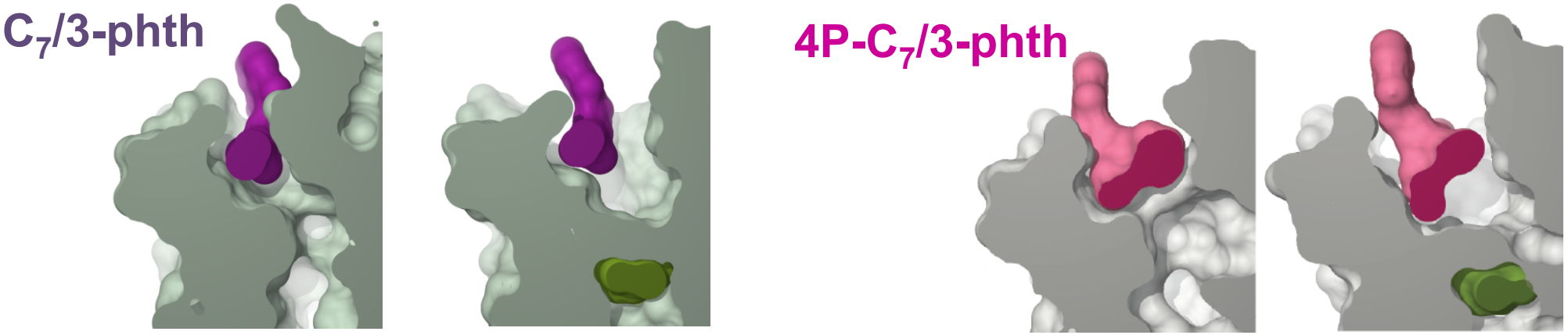


Mechanism 2: Coupled conformational change of orthosteric and allosteric sites

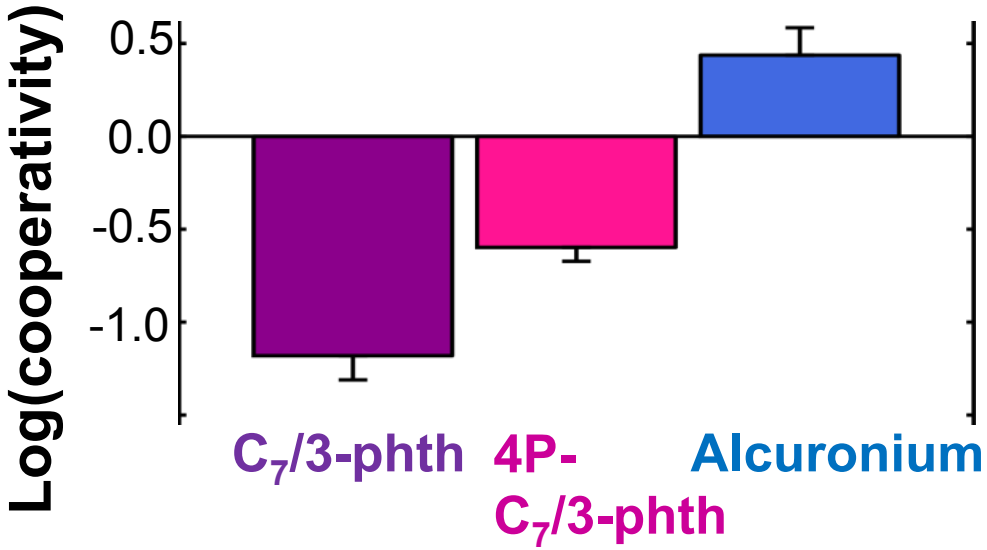


Designing an allosteric modulator

- **Prediction**: Computationally designed modulator 4P-C_7/3-phth binds like C_7/3-phth but forces open allosteric site, making cooperativity more positive (less negative)



- **Experimental validation**: 4P-C_7/3-phth has less negative cooperativity than C_7/3-phth despite binding more tightly

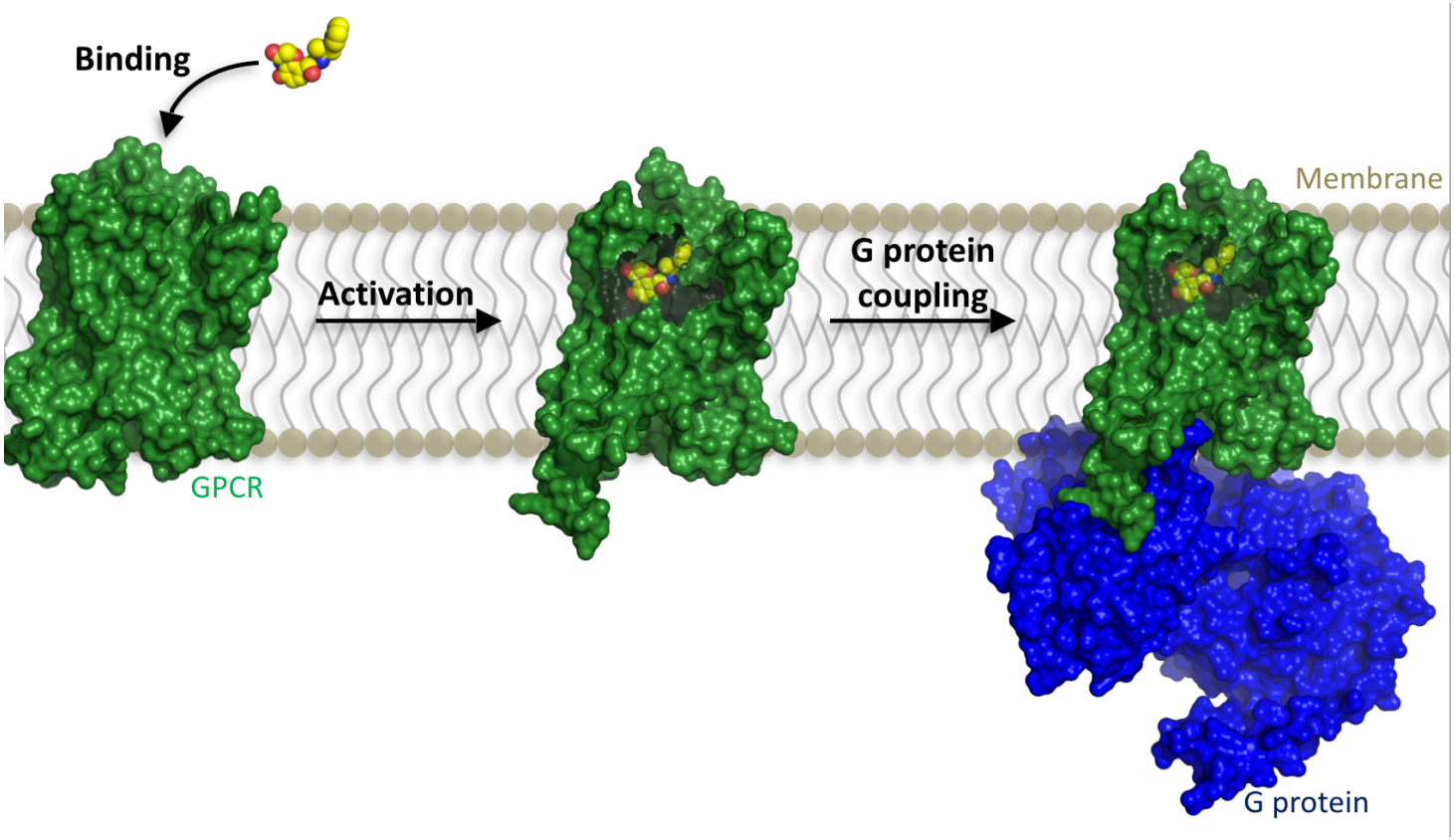


Limitations of study

- Computational:
  - We didn't explicitly compute cooperativity between allosteric and orthosteric ligands
  - We have good evidence for negative cooperativity between the allosteric modulator C_7/3-phth and the orthosteric ligand, but weaker evidence for cooperativity of most other allosteric modulators
- Experimental:
  - We didn't perform experimental validation of the electrostatic mechanism
  - We didn't solve a crystal structure with one of these modulators bound

Limitations of study

- The big one (in my opinion):
  We did not study allosteric modulation of acetylcholine, the neurotransmitter that is the natural ligand for this receptor
  - When developing an allosteric modulator as a drug, you generally care most about cooperativity with the natural ligand
  - Acetylcholine is an agonist: it favors activation of the receptor
  - But we had only an inactive-state structure of the receptor, so we studied modulation of NMS, which does not favor activation



Application #2 of MD simulation:

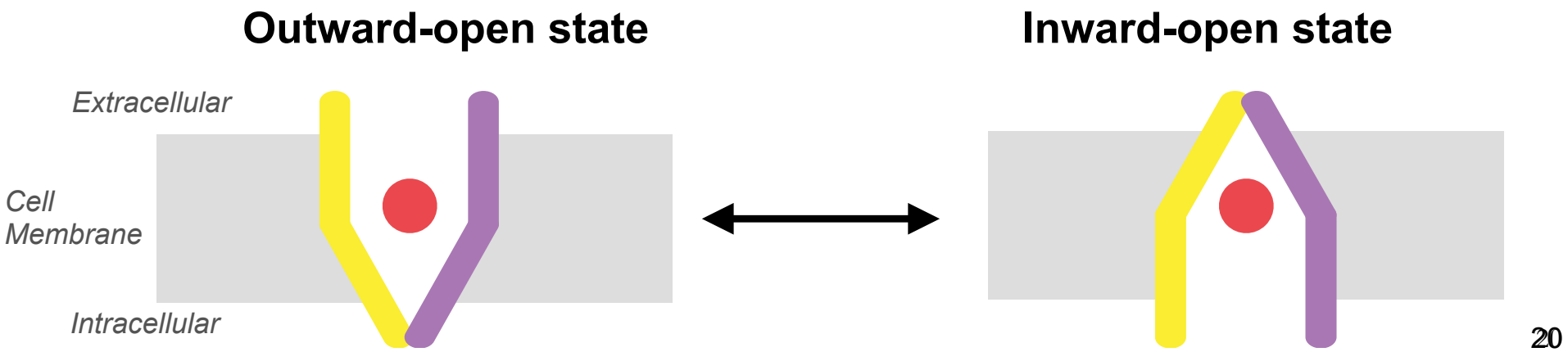
“Mechanism of Substrate Translocation in an Alternating Access Transporter”

Latorraca et al., *Cell* 169: 96–107 (2017)

How do membrane transporters work?

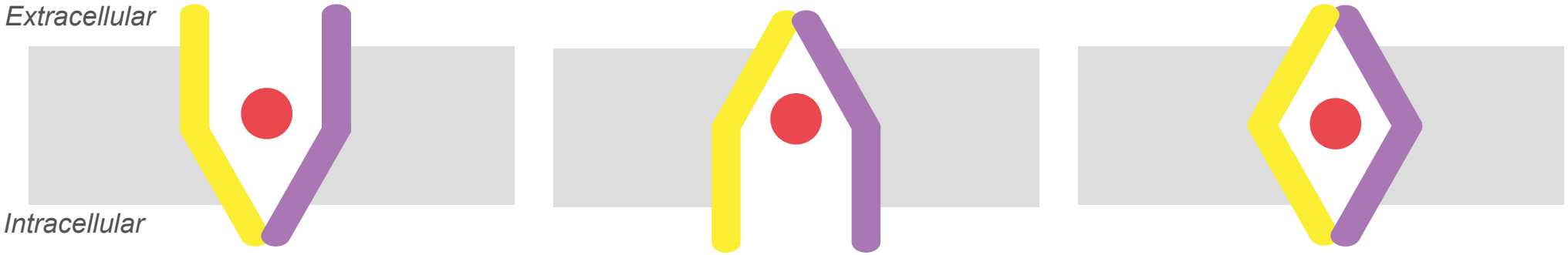
- Transporters are:
  - Membrane proteins that shuttle molecules into and out of the cell
  - Key drug targets
  - Responsible for drug efflux (e.g., antibiotic resistance)

- Jardetzky's hypothesis, 1966: Transporters shuttle molecules across the cell membrane by **alternating access** mechanism

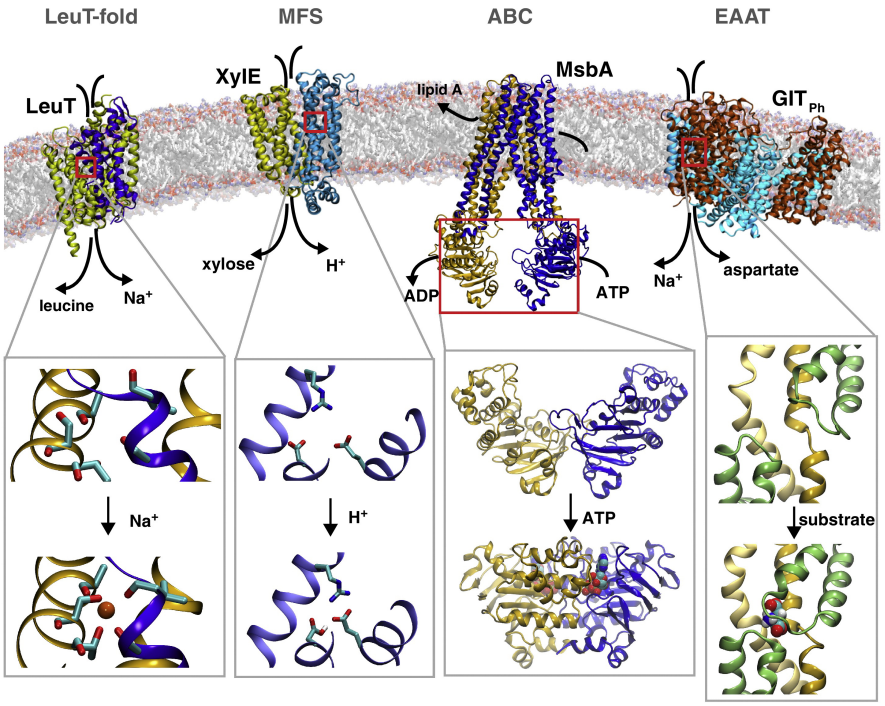


Alternating access transport

- Crystallography has captured outward-open, inward-open, and occluded states of transporters



Outward-open state **Inward-open state** **Occluded state**



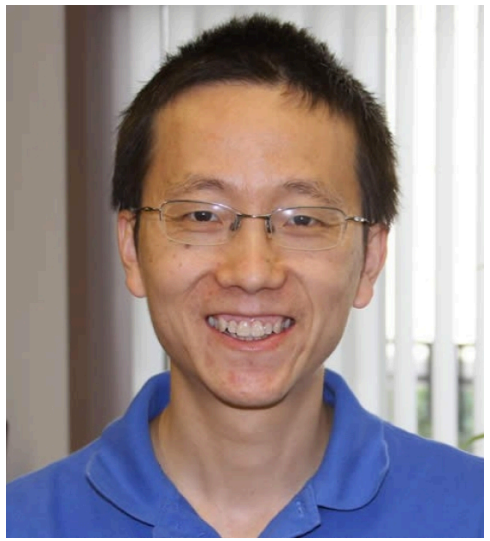
Li, …, Tajkhorshid, COSB 2015

Can we observe the transport process?

- Atomic-level simulation seems like a natural approach
- But it's proven very difficult
  - No unguided simulation of spontaneous outward-open to inward-open transition and substrate translocation

Naomi Latorraca

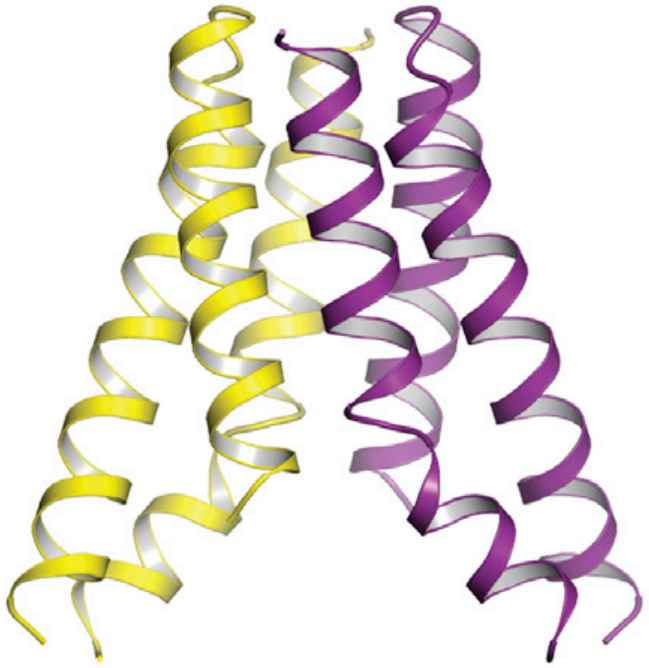
Nathan Fastman



Liang Feng

Latorraca et al., *Cell* 169:96-107, 2017

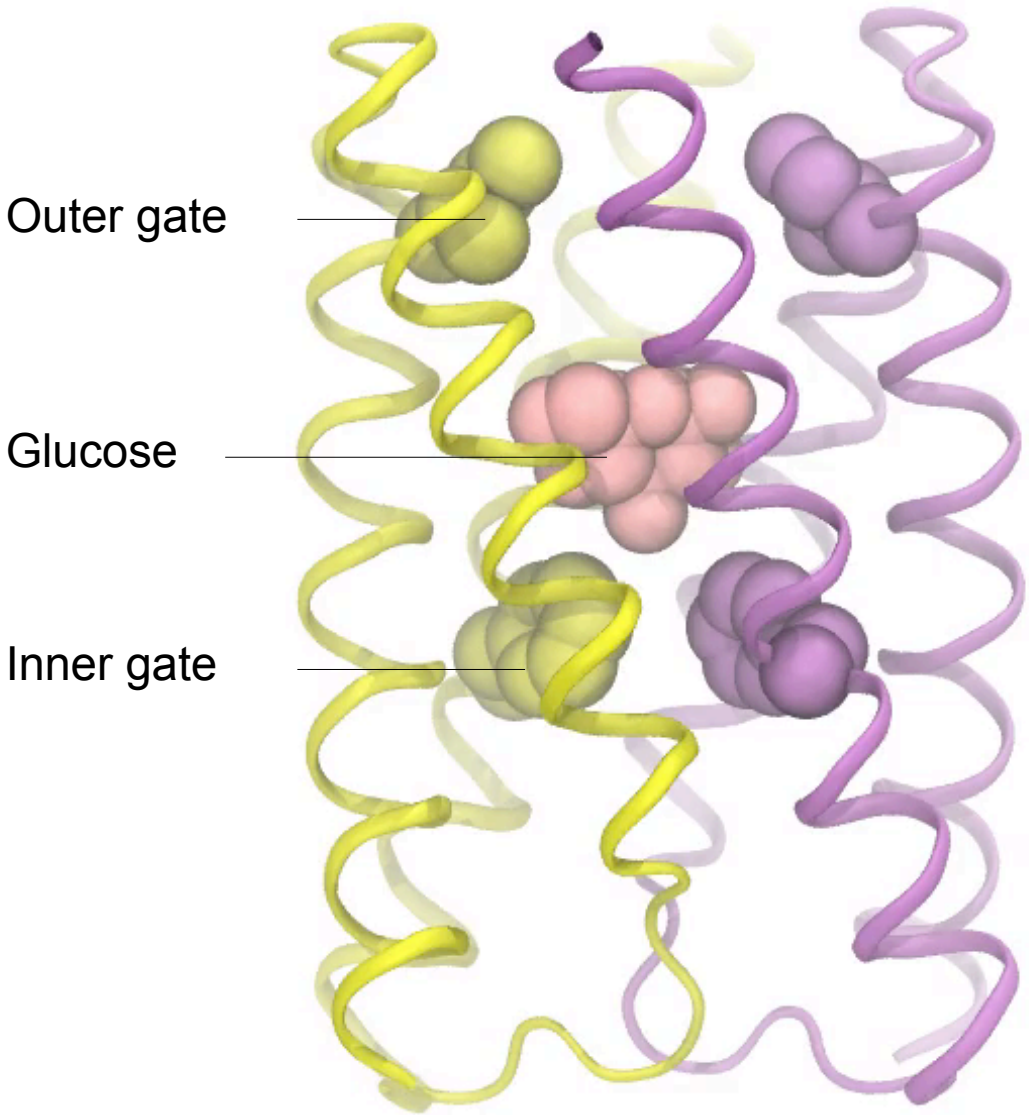
The SemiSWEET sugar transporter



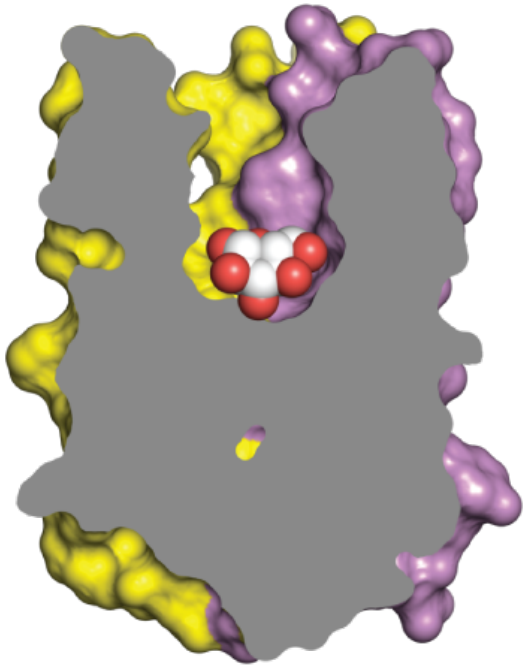
- A minimal alternating access transporter
  - Small (<20 kDa)
  - Simple geometry (symmetric dimer)
  - Uniporter (single substrate)

Liang Feng lab

Spontaneous transport in simulation

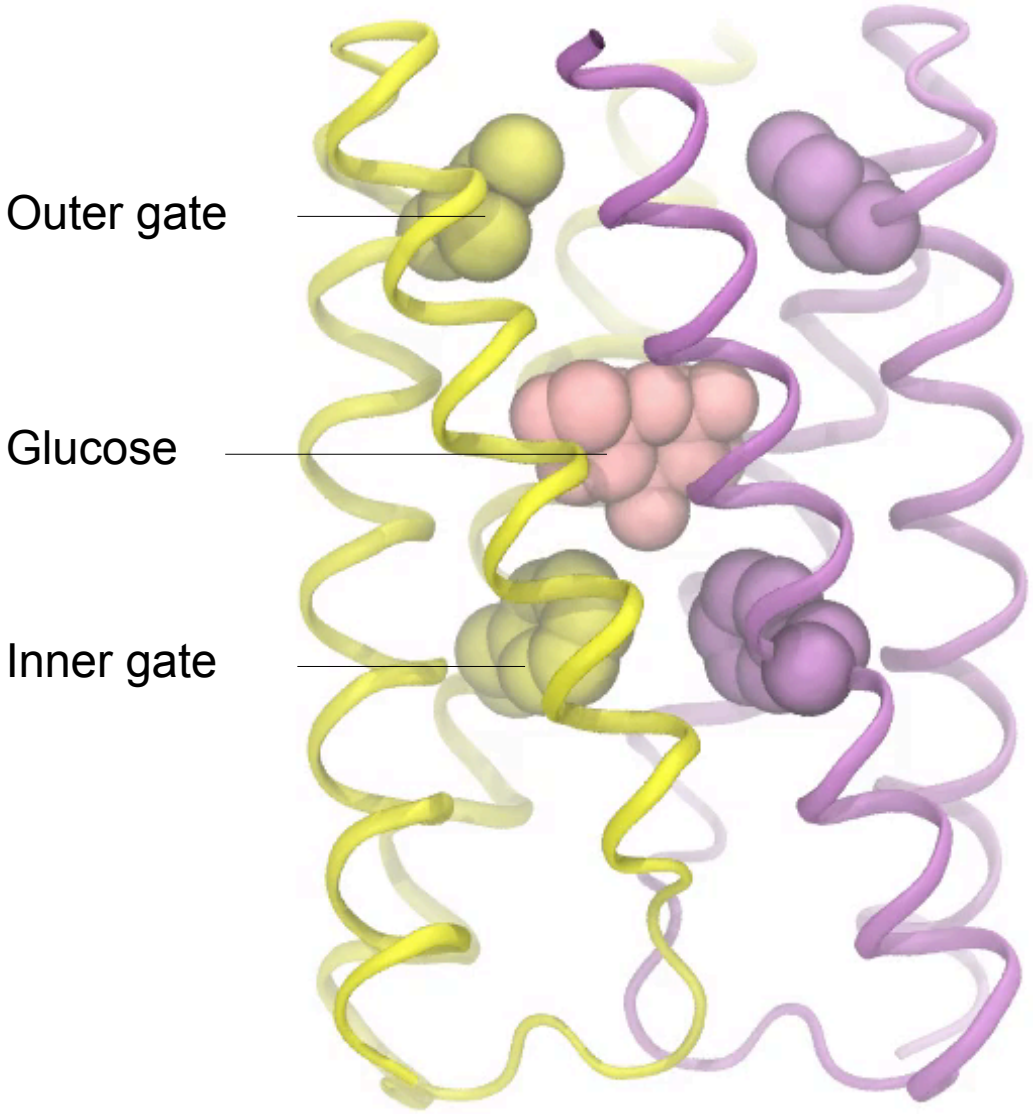


We start with a crystal structure of the SemiSWEET sugar transporter in its outward-open state, bound to glucose

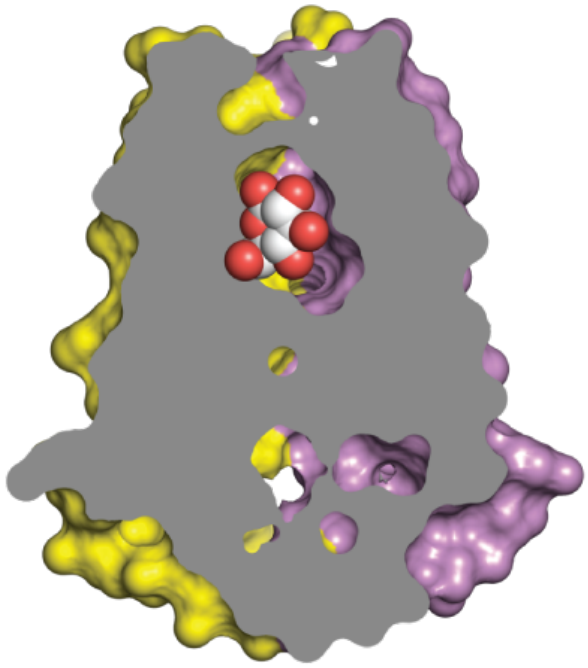


Latorraca, et al., *Cell*, 2017 (collaboration with Liang Feng's lab)

Spontaneous transport in simulation

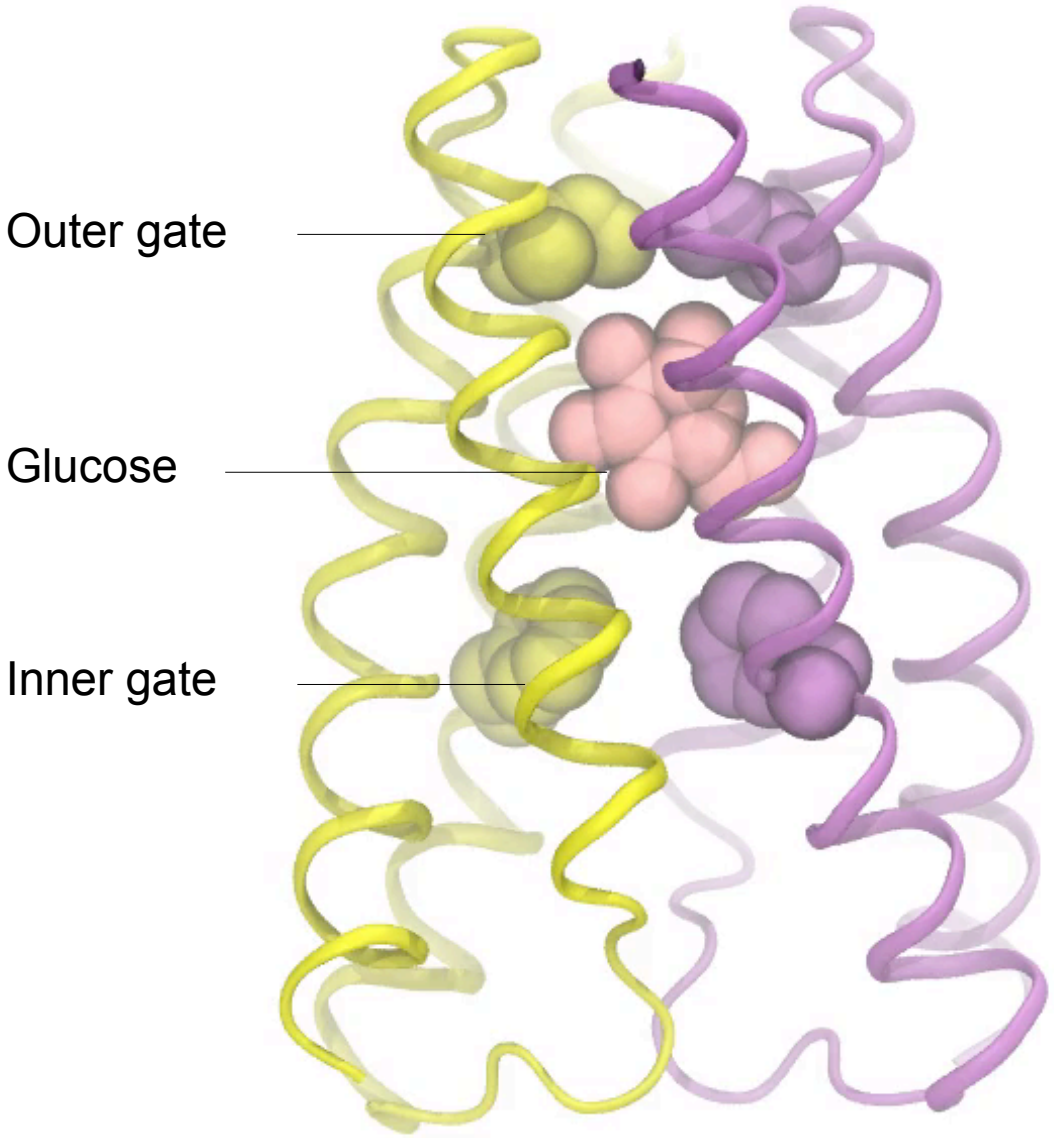


The transporter transitions spontaneously to an occluded state, then an inward-open state

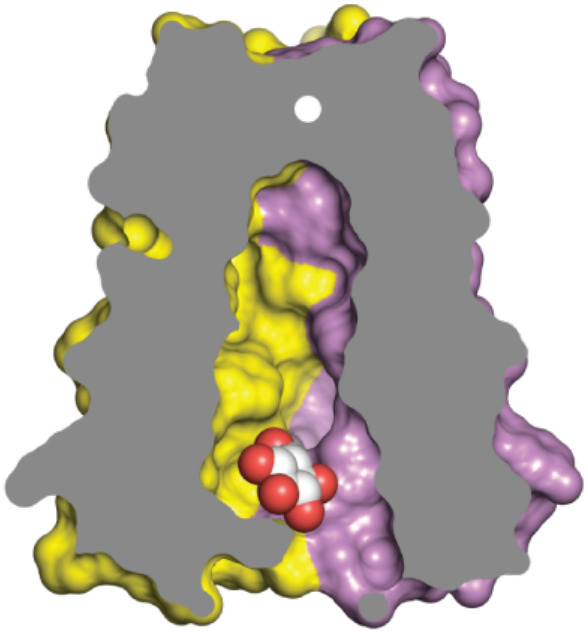


Latorraca, et al., *Cell*, 2017 (collaboration with Liang Feng's lab)

Spontaneous transport in simulation

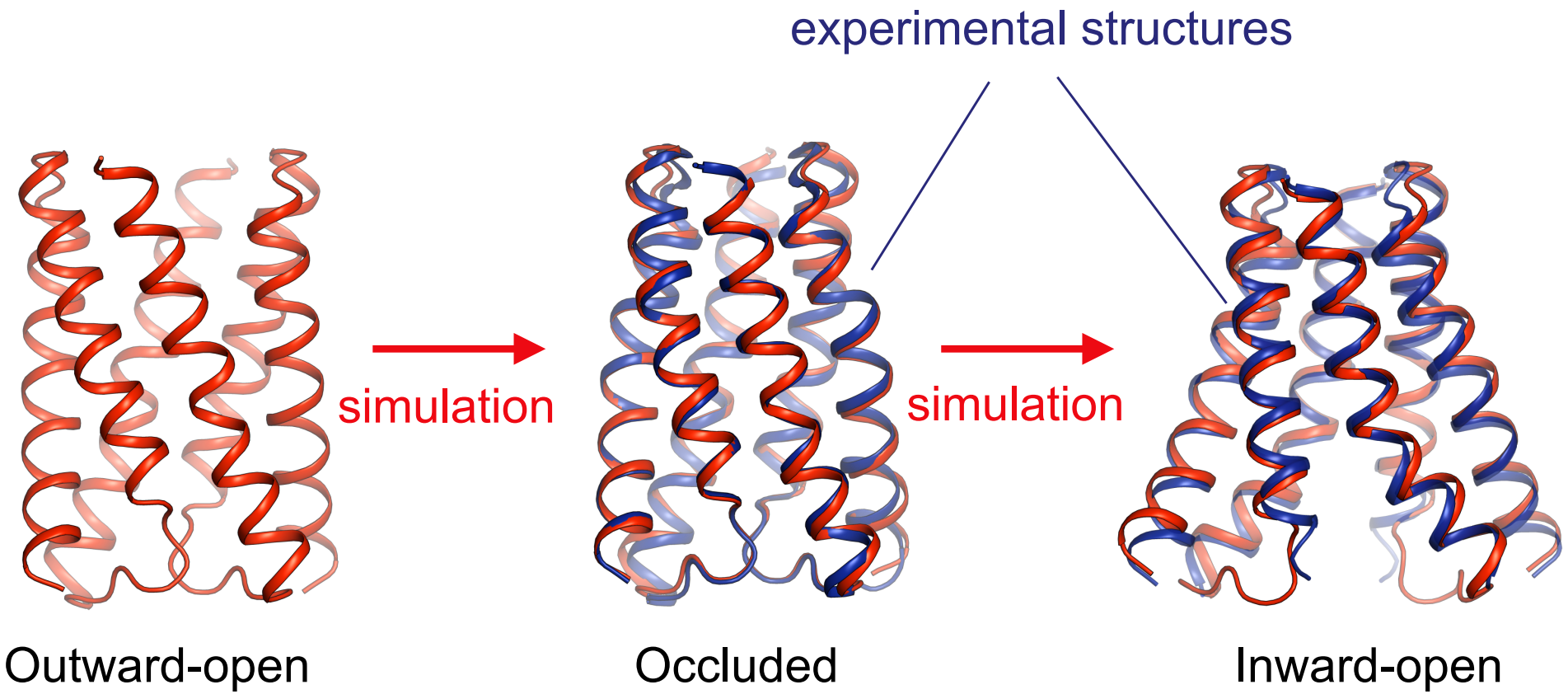


The transporter transitions spontaneously to an occluded state, then an inward-open state



Latorraca, et al., *Cell*, 2017 (collaboration with Liang Feng's lab)

Simulated conformations match experimental structures

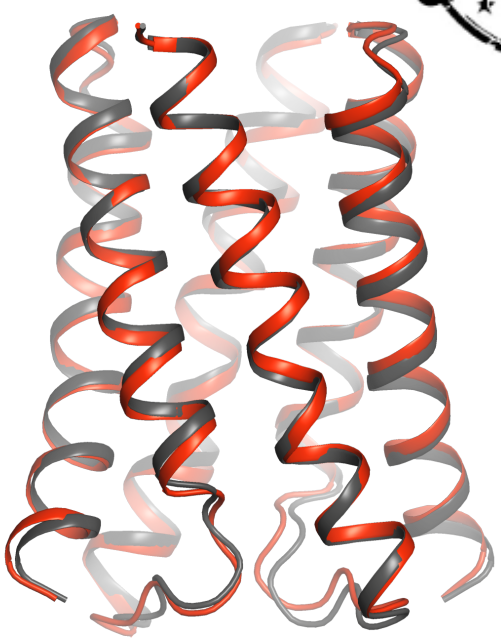


Simulations allow us to answer longstanding questions about transport mechanism

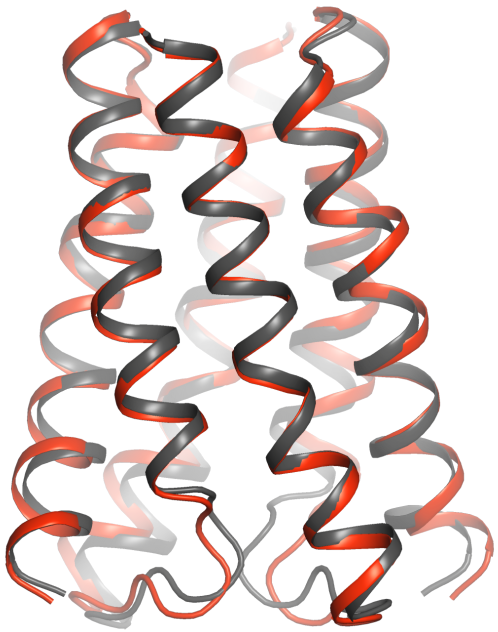
- How does substrate influence the transporter's structure?

no sugar

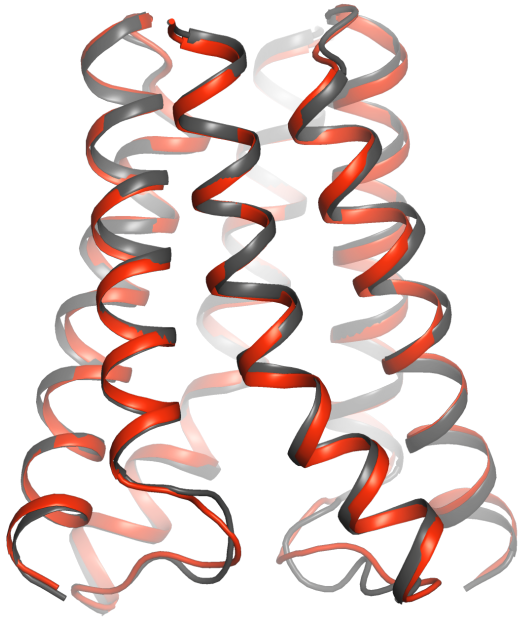
sugar bound



Outward-open



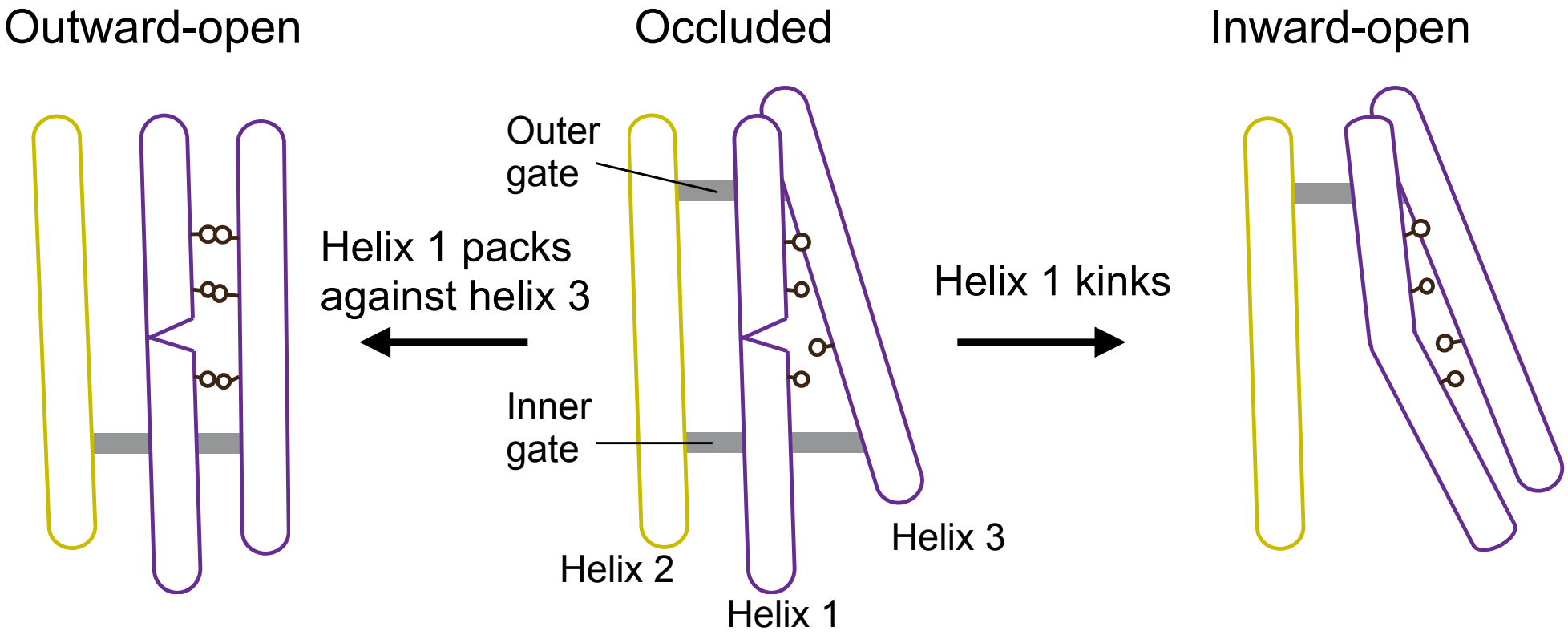
Occluded



Inward-open

Simulations allow us to answer longstanding questions about transport mechanism

- What is the role of the occluded state?
- How do the inner and outer gates avoid opening simultaneously?



Opening either gate stabilizes helix 1, but only if the other gate is closed.

Limitations: what we didn't do

- We did not simulate the inward-open to outward-open transition
- We did not address substrate selectivity
- We did not simulate other transporters
  - They may or may not behave similarly

Analysis of simulation data:

"Identifying localized changes in large systems: change-point detection for biomolecular simulations"

Fan et al., *PNAS* 112:7454–7459 (2015)

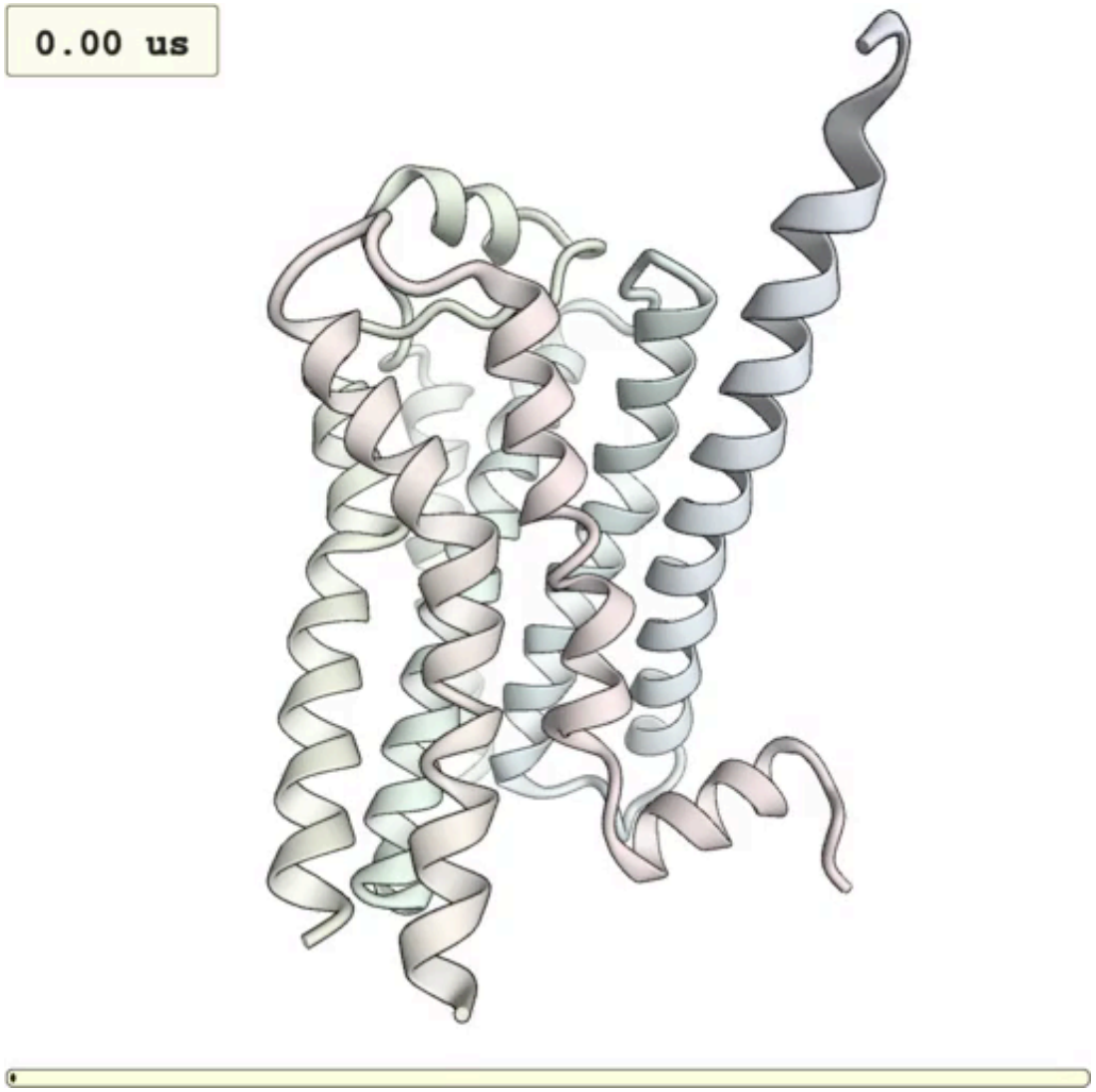
The challenge

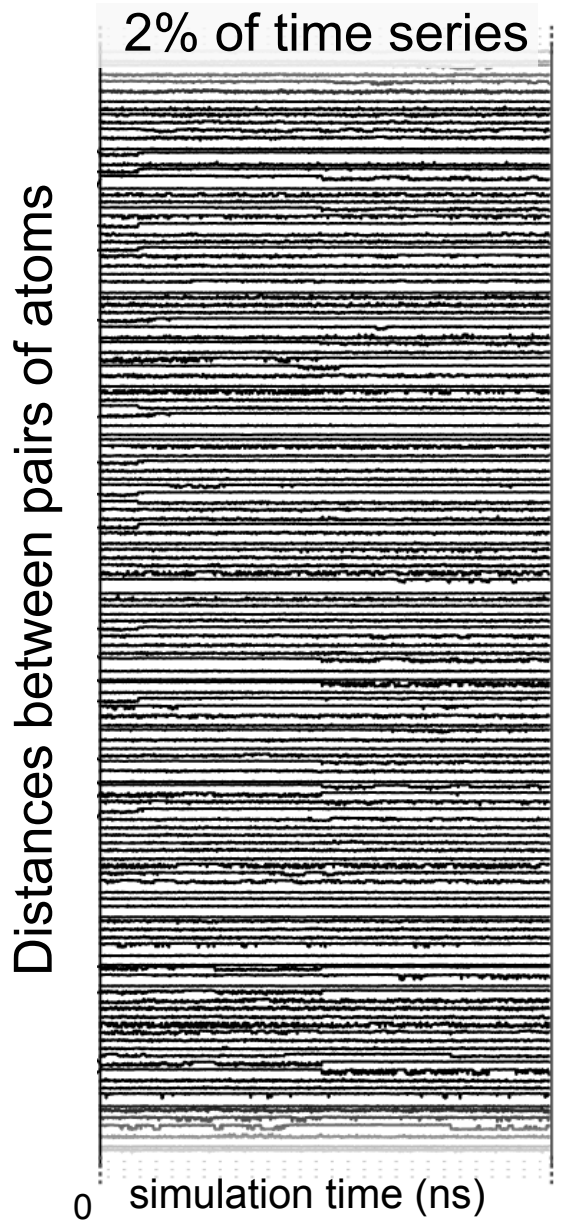
- MD simulations generate a lot of data
  - Example: simulate a 50,000-atom system for 1 μs
  - That's half a billion time steps
  - The simulation calculates the position and velocity of every atom at every time step
- Sometimes one knows precisely what to look for
- In other cases—particularly when using simulations to understand functional mechanisms—extracting meaningful information from simulations involves staring at the results for a long time
- Can we automate this process?

What are the “important events”?

- They usually involve conformational (that is, structural) changes
- These changes can be subtle: they might involve only a very small part of the protein
- The protein is moving *constantly*
- We tend to care most about *rare* changes

Can we approach this as a changepoint detection problem?





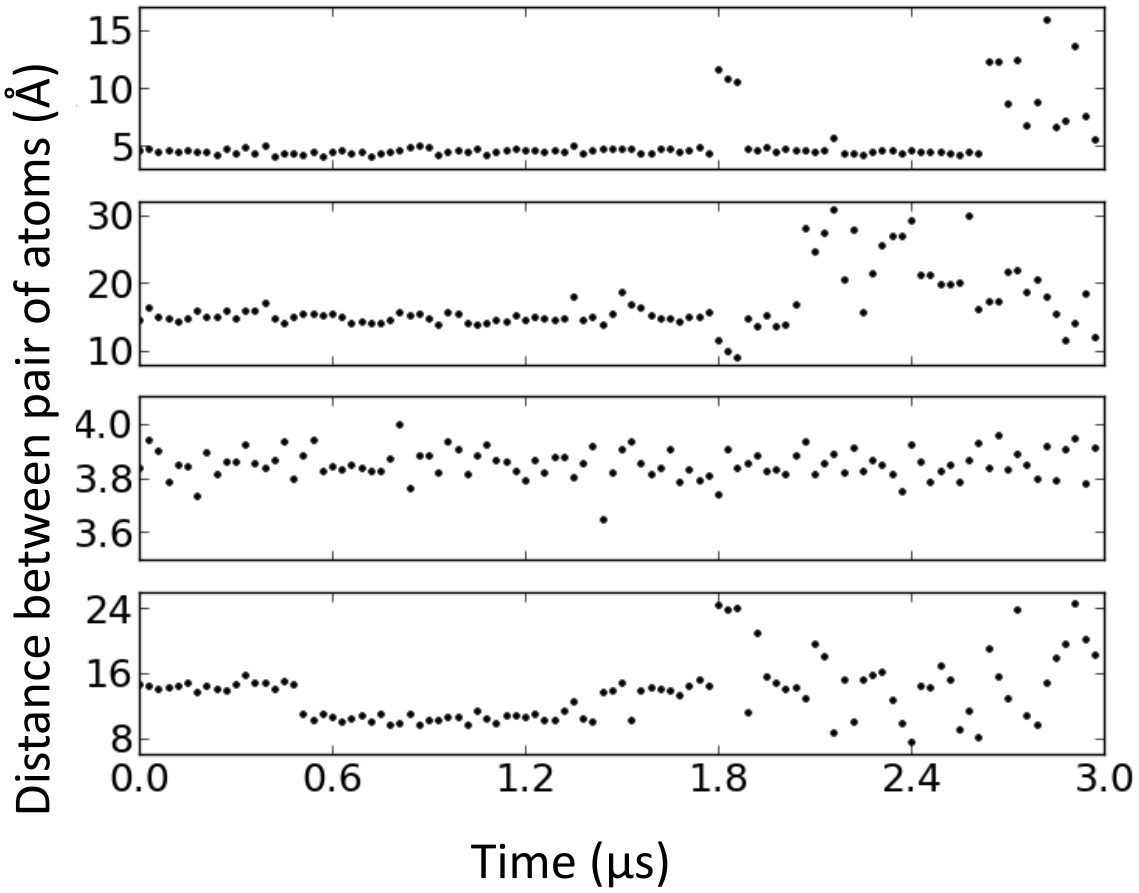
Challenge: Each change affects only a few of the thousands of time series that describe a simulation

Fan et al., *PNAS*, 2015

Detection of simultaneous change points

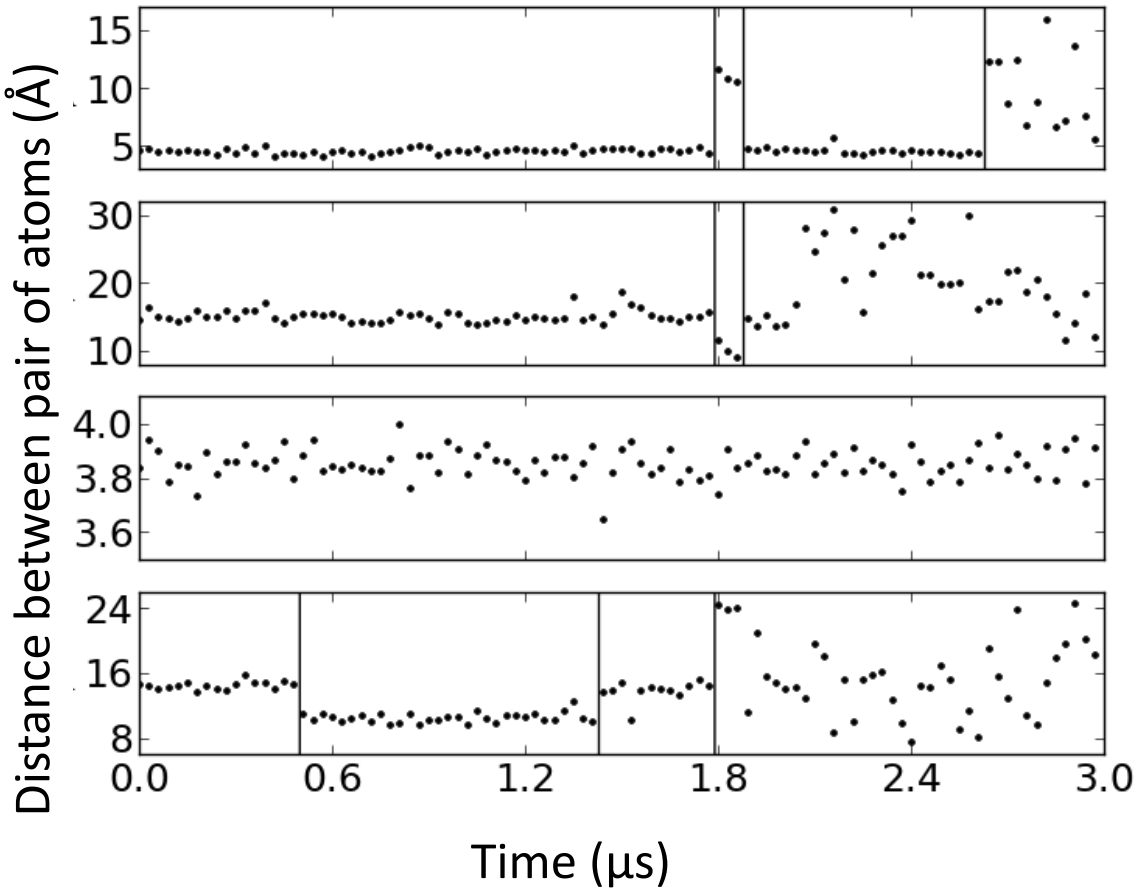
- Approach this as a statistical change-point detection problem, but:
  - Determine both change times and the observables that change at each time
  - Search, in particular, for simultaneous changes of multiple observables
- We can formulate this as a giant optimization problem
  - We can solve this problem efficiently by iterative application of recently introduced dynamic programming algorithms

Simultaneous changepoint detection: A simple example



Four sample time series (out of thousands), each corresponding to the distance between a pair of protein atoms in a simulation

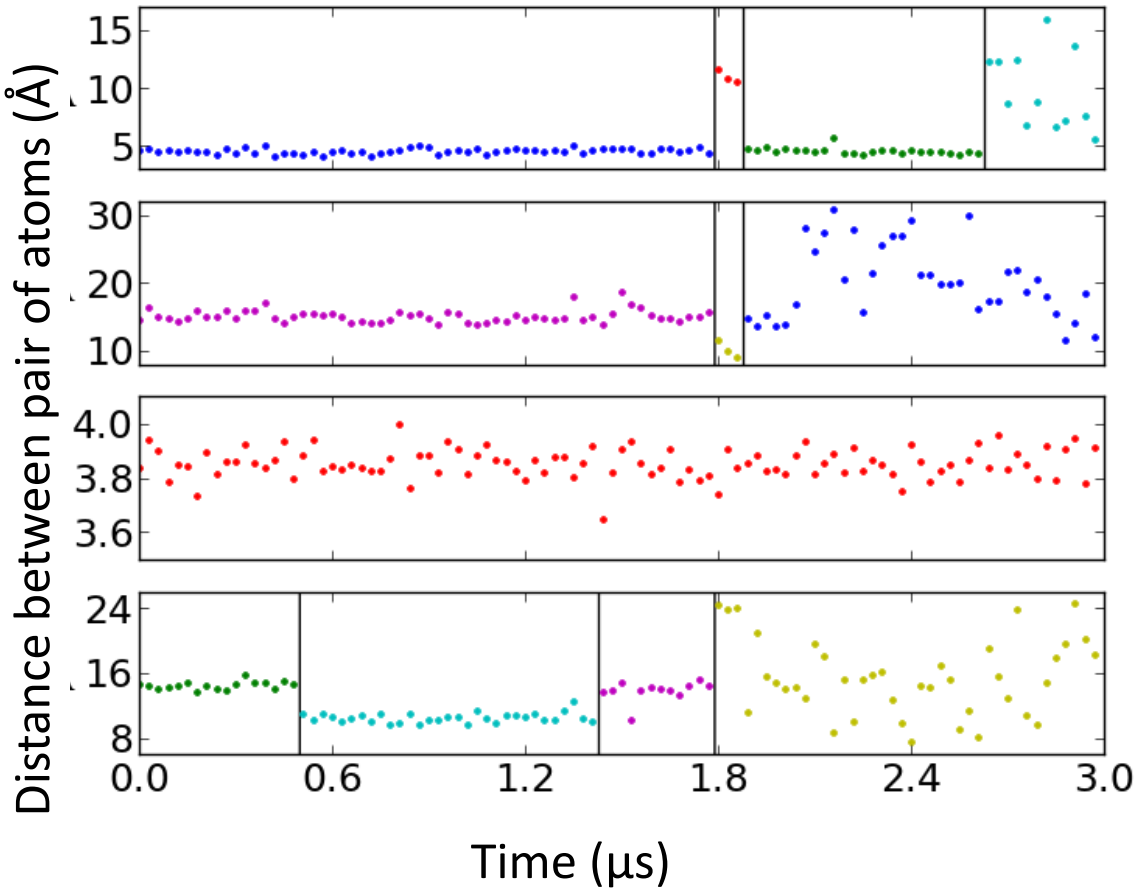
Simultaneous changepoint detection: A simple example



We wish to choose a small set of changepoints such that each observable has constant statistics between changepoints.

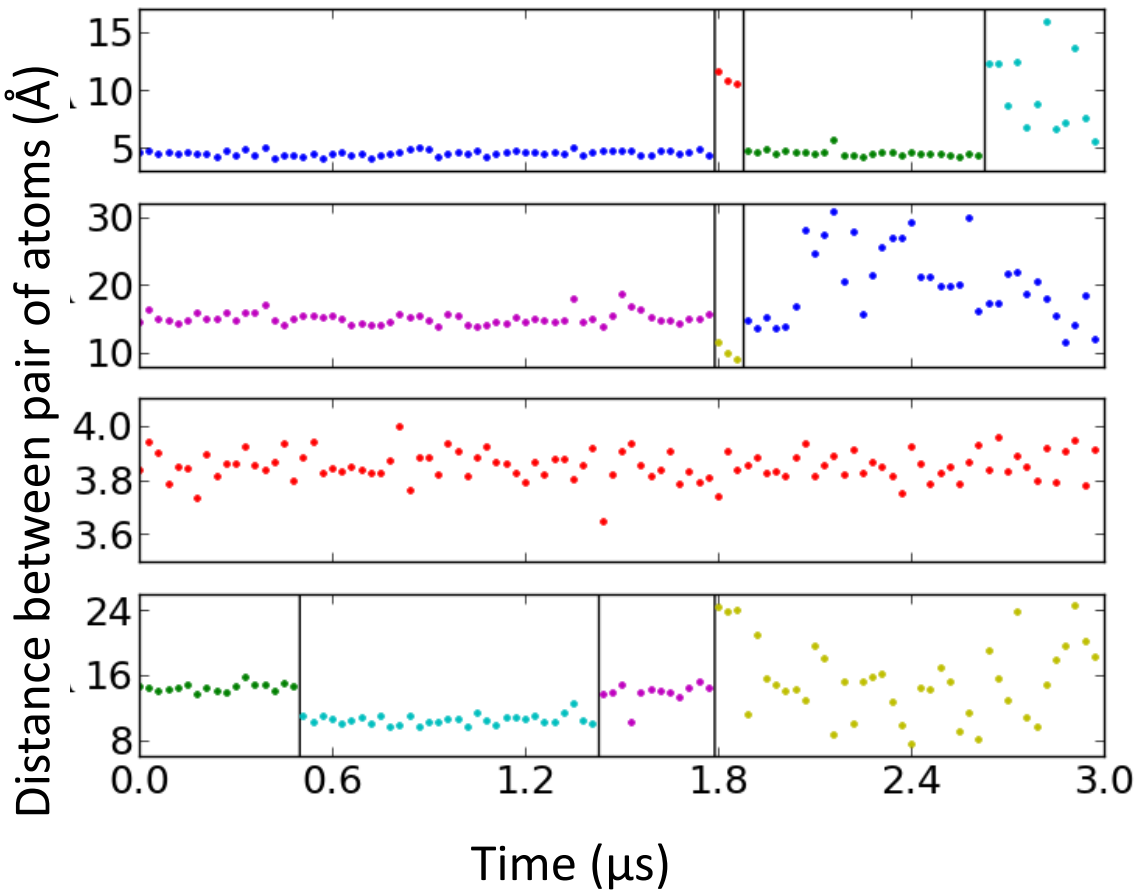
Exploit the fact that changes in different observables are likely to occur simultaneously, especially if the atoms involved are nearby.

Our approach



Solve a big optimization problem to determine when changes occur, which observables they affect, and how the statistics of those observables change (i.e., what parts of the protein change at what times, and how).

Our approach



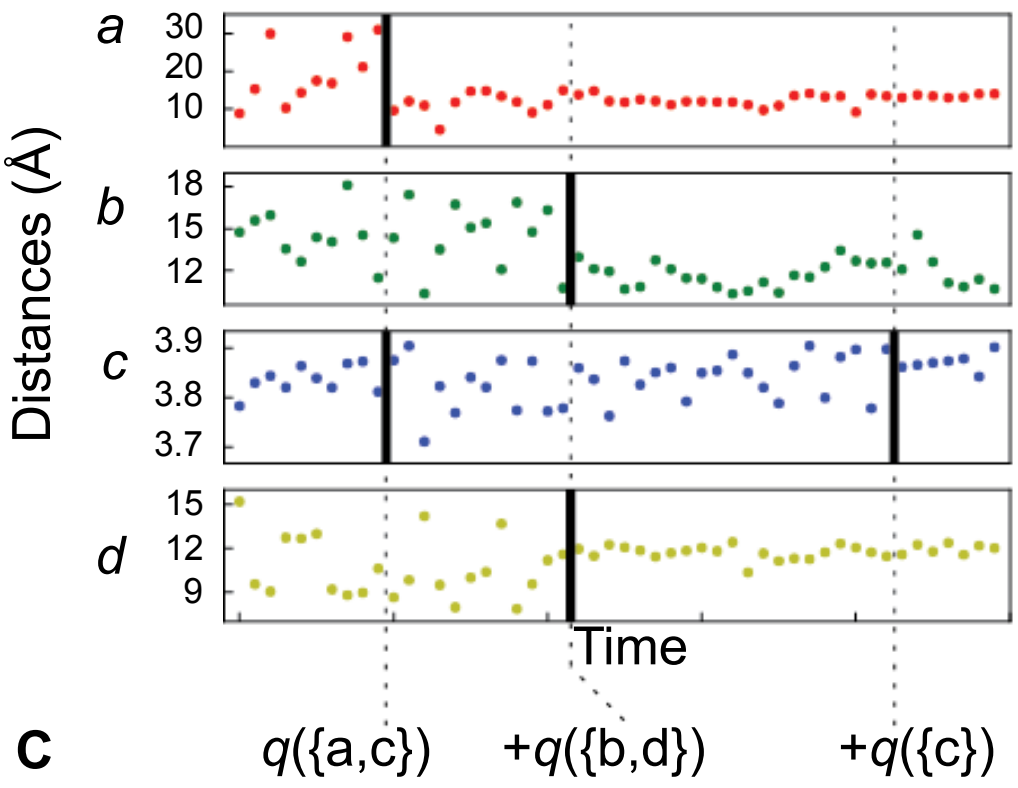
Choose changepoints, and model parameters for each segment, to maximize:

[Likelihood of data given model] – [Penalty function]

where the penalty function increases with the number of changepoints, but less so if multiple changepoints are simultaneous (especially if the atoms involved are nearby one another).

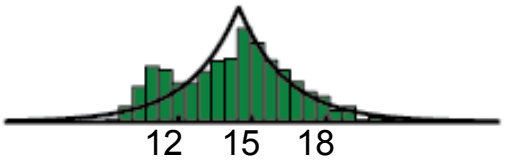
Our approach: Simultaneous penalized likelihood estimation (SIMPLE) changepoint detection

A Candidate change points



B Data likelihood given change points

$$\hat{l}(\;\;)+\hat{l}(\;\;)+\hat{l}(\;\;)+\ldots$$



$\hat{l}$ = log likelihood under best-fit model

C

$q(\{a,c\})$ $+q(\{b,d\})$ $+q(\{c\})$

Change-point penalties

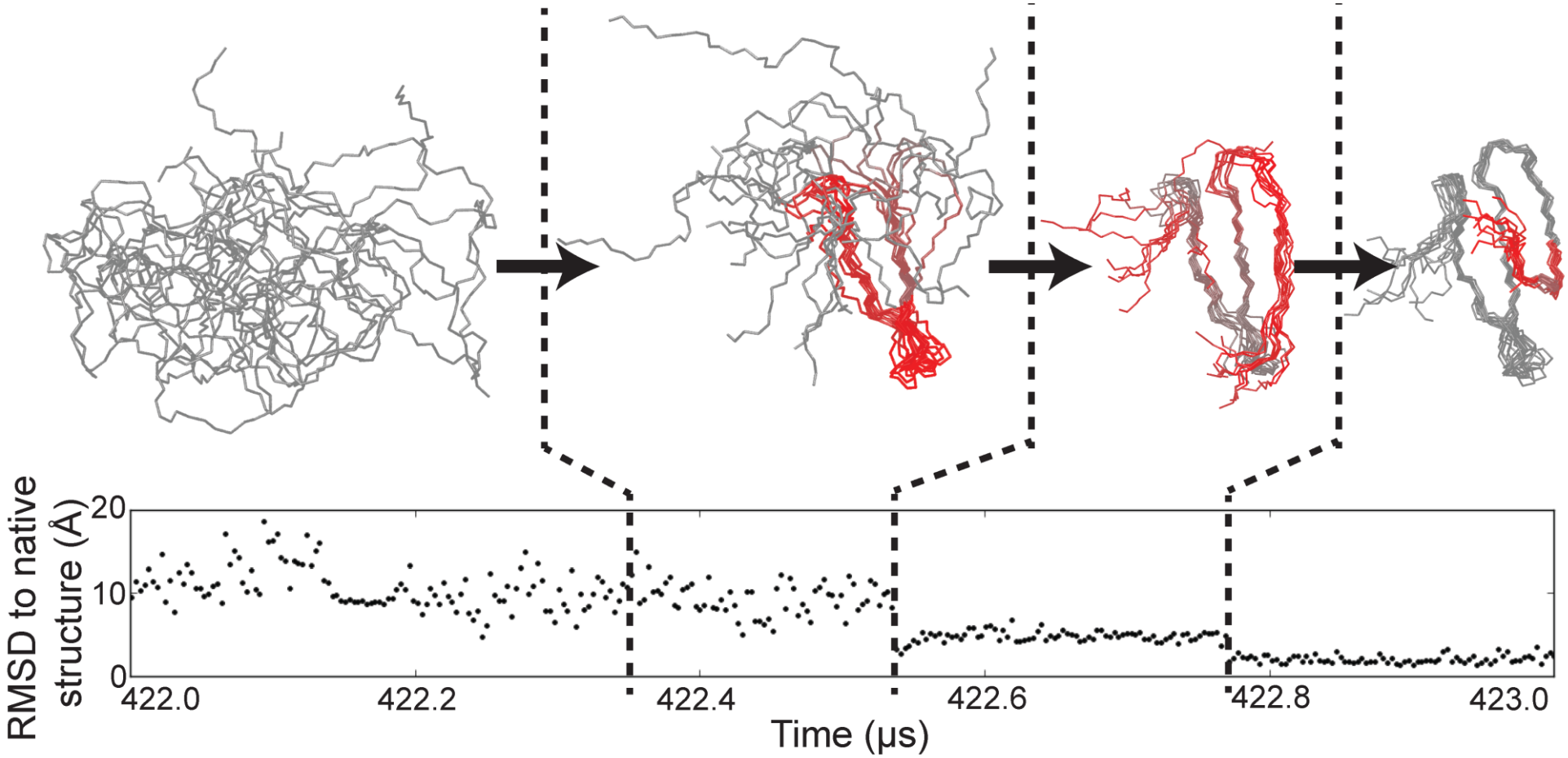
D

$$\sum \hat{l} - \lambda \sum q$$

Objective function to maximize

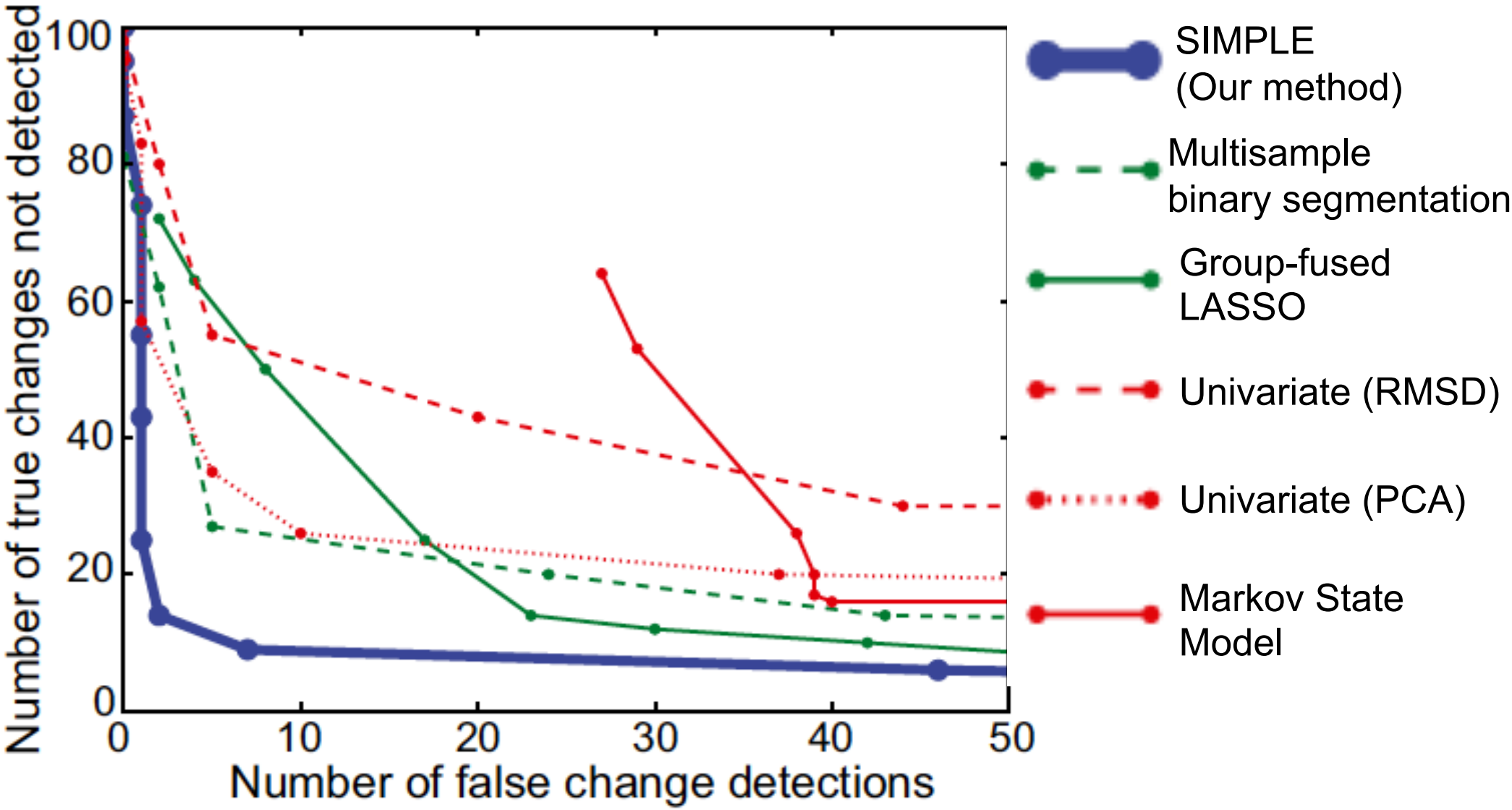
Penalty increases with number of changes, but less so if changes are simultaneous

Detected change points in WW domain folding simulation



Performance comparison

(on synthetic data for which we know the true change times)



Limitations of the study

- Basic problem definition: is one always looking for “sudden” changes of this sort?
- No guarantee that optimization algorithm will converge to global optimum
- Lack of good software for visualizing the results
  - We actually developed (and released) such software, but its portability is poor
- Although the paper describes a general method, it doesn’t demonstrate application of the method to other types of data