Essential Dynamics of Proteins using Geometrical Simulation with Subspace Analysis

Presentation of Doctoral Dissertation

by Charles Christian David

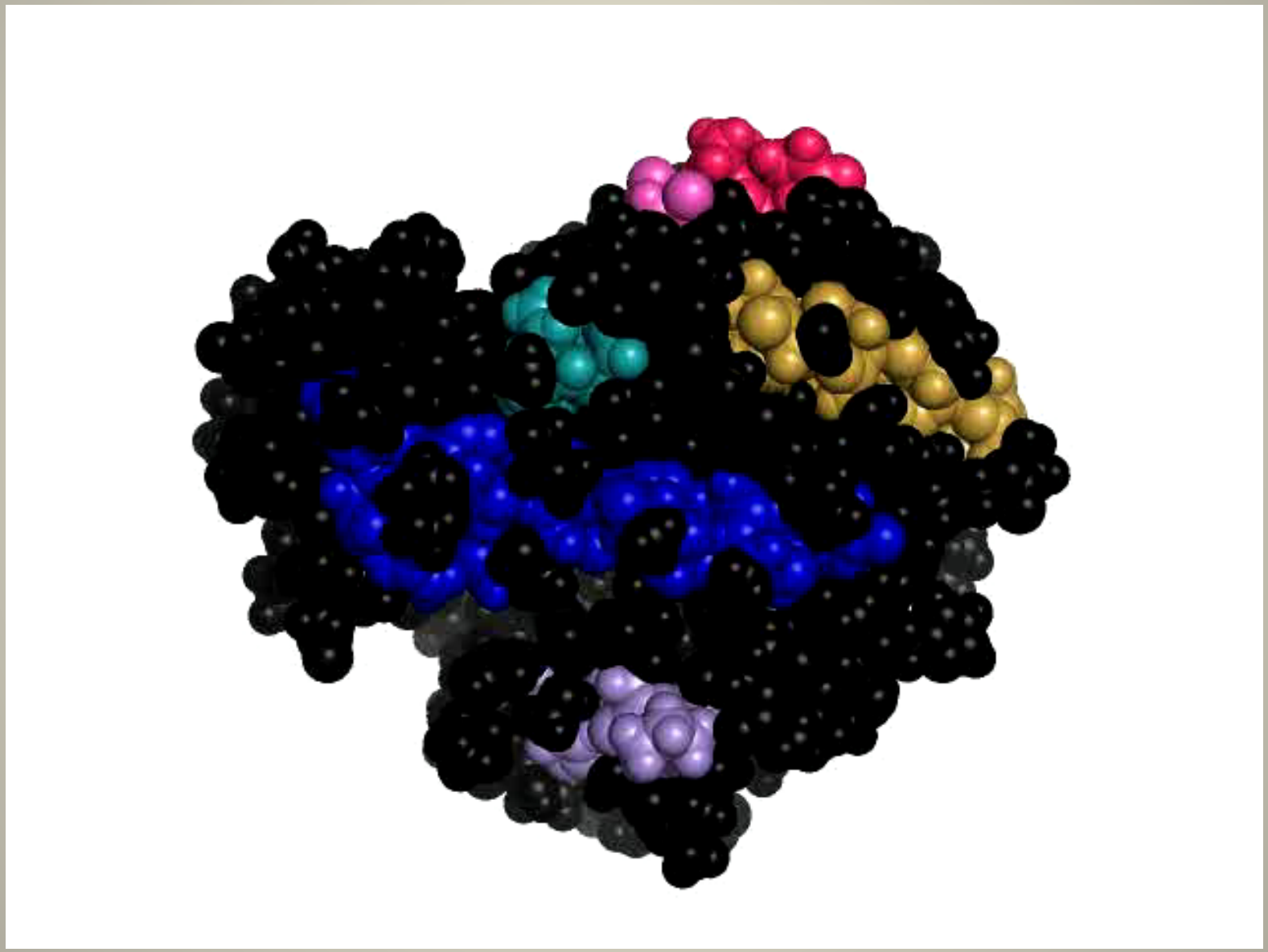
The BIG PICTURE

PREVIEW

Why BCB Loves Proteins

- Proteins are central to cellular function, i.e., LIFE
- We hope to uncover the secrets of how these complex macromolecules execute their functions
- The dream is to 'watch' proteins in action: in real time at atomic resolution
- We approach the dream with our models.

Dynamic Trajectory of Myoglobin



Overview of the research

DISSERTATION PROJECT

Three Phases

I. Benchmark the Geometrical Simulation Model (GSM)

II. Compare GSM results to molecular dynamics (MD) and elastic network model (ENM)

III. Apply the GSM to myosin V

Overview of Essential Dynamics & PCA

INTRODUCTION

Essential Dynamics (ED)

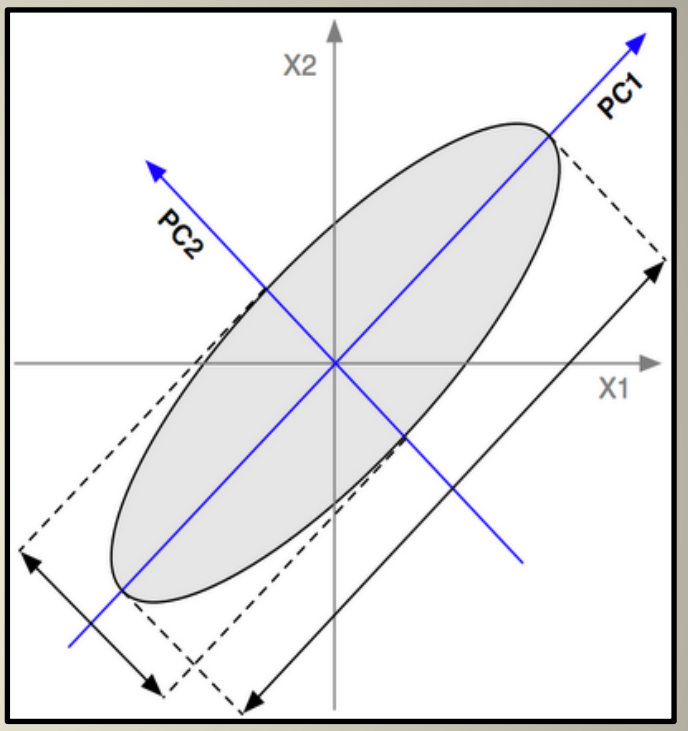
- The process of applying Principal Component Analysis (PCA) to a protein trajectory

A General Overview

PCA

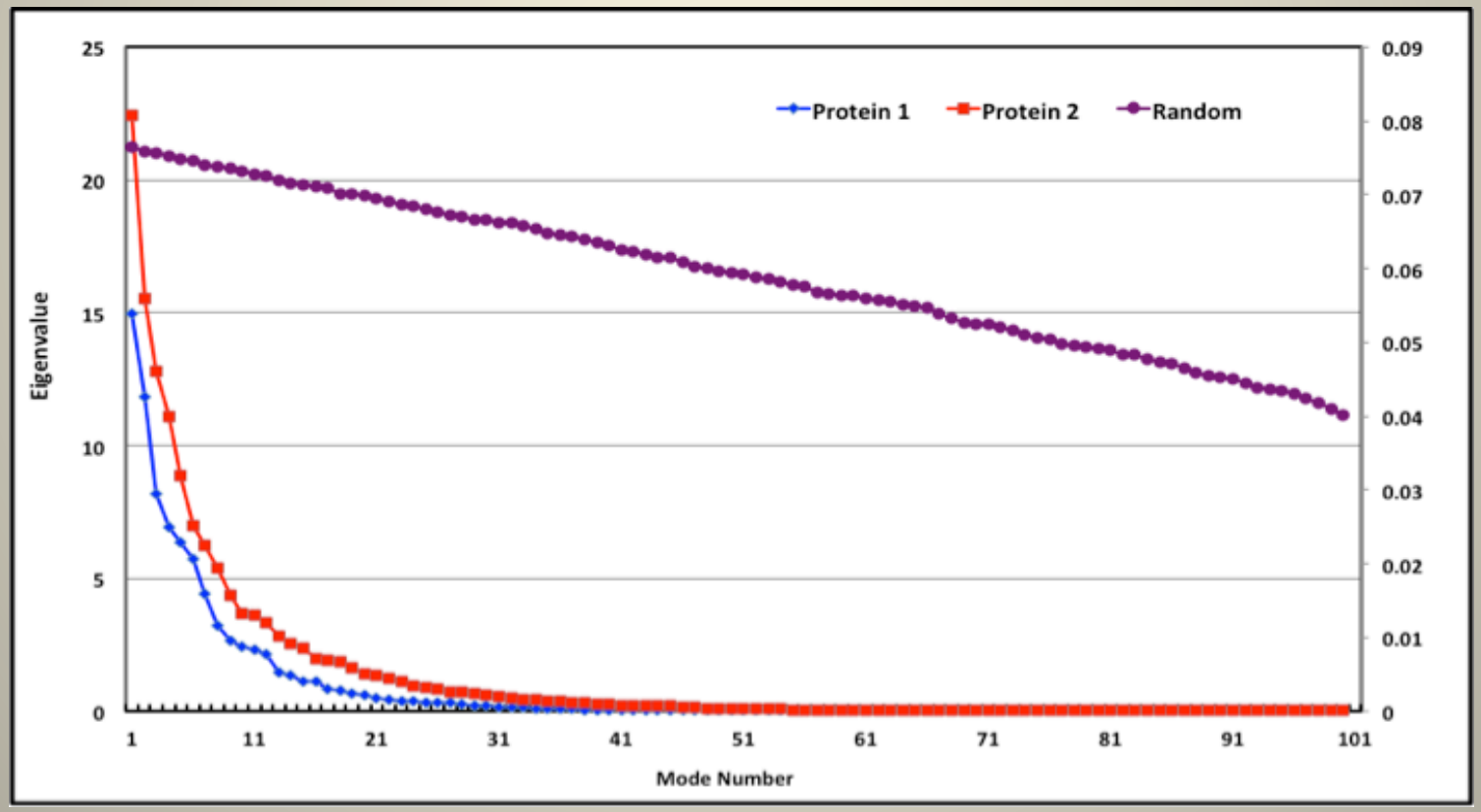
PCA I

- Identifies directions of highest variance in data
- A standard method of data compression and dimension reduction
- The analysis is done on a centered C-Matrix (C):
  - Covariance matric (Q)
  - Correlation matrix (R)



PCA II

- An **eigenvalue decomposition** is performed on **C**
- The eigenvalues are plotted on a "scree plot"
  - Scree Plots help identify the Essential Subspace



PCA III

- The primary limitations of PCA:
  - **Linear Transform**
  - **Second Moment**
  - **Orthogonal**
  - **Quality of Data Sampling**

- PCA is not limited to Cartesian DOF
  - Internal Distance Coordinates → **dPCA**

PCA IV

- Major advantage of ED is choice of subset
  - Examine the large-scale motions within the subset of residues
  - Under the influence of the entire set of residues contained in the protein
    - **Active sites**
    - Potential **allosteric pathways**

Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins.
Charles C. David and Donald J. Jacobs
[Submitted July, 2012]

Benchmarking the GSM

PART I

Overview of the paradigm

GEOMETRICAL MODEL

GSM I

- An **all atom**, athermal, **mechanical** model
- Geometrical Constraints used to define rigid regions of the input structure
- Molecularly realistic perturbation of those rigid regions yields conformers (Monte Carlo)

GSM II

- FIRST

  Floppy Inclusions and Rigid Substructure Topology

  – Covalent Bonds

  – Hydrogen Bonds (HB) (pseudo temp)

  – Hydrophobic Tethers (HP Tethers)

- Recasts the protein into a set of rigid regions joined by hinges

  – **Rigid Cluster Decomposition** (**RCD**)

GSM III

- FRODA

  Framework Rigidity Optimized Dynamics Algorithm

  – MC sampling of perturbations to the RCD

  – Molecular "realism" is enforced in FRODA
    - Constraint violations → Reject Conformer

  – Two general modes of operation:
    - Diffusion (random)
    - Momentum (biased)

Methods I

- We obtained the FIRST/FRODA implementation of the GSM from the Thorpe Group at ASU

  Generating stereochemically acceptable protein pathways.
  Farrell, D.W., Kirill, S., Thorpe, M.F.
  Proteins, 78, 2908-2921. 2010

- We designed a Java software package for the ED analysis of dynamic trajectories

  Essential Dynamics of Proteins with Subspace Analysis in Java.
  Charles C. David and Donald J. Jacobs
  [In Preparation]

Methods II

- We assessed the behavior of the model for:
  - Effectiveness
  - Efficiency
  - Consistency

Assessing Similarity of Essential Spaces

SUBSPACE METRICS

Subspace Metrics I

- Overlap: $O_{ij} = \frac{u_i \cdot v_j}{\|u_i\|\|v_j\|}$
- Cumulative Overlap: $CO(k) = \left(\sum_{j=1}^{k} O_{ij}^2 \right)^{\frac{1}{2}}$
- Root Mean Square Inner Product:

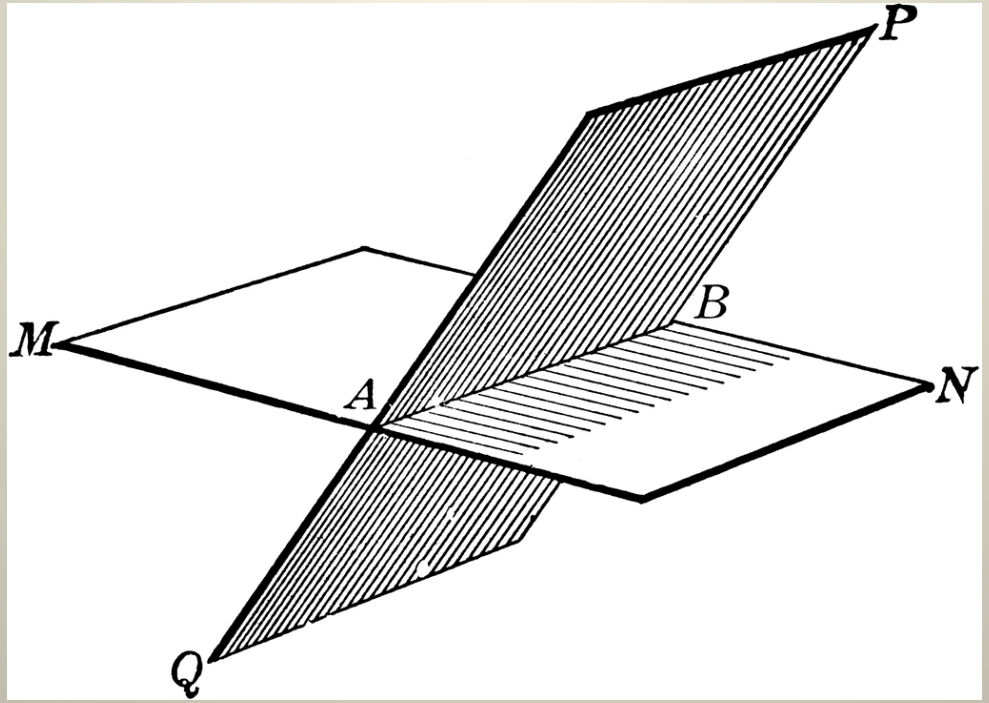
$$RMSIP(I,J) = \left(\frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{J} \left(u_i \cdot v_j \right)^2 \right)^{\frac{1}{2}}$$

- Principal Component Projections:

$$PC_i = X^T \cdot v_i$$

Subspace Metrics II

- Principal Angles
  - An optimization that gives the best alignment between the two subspaces
  - **Reveals timescales of dynamic congruency**



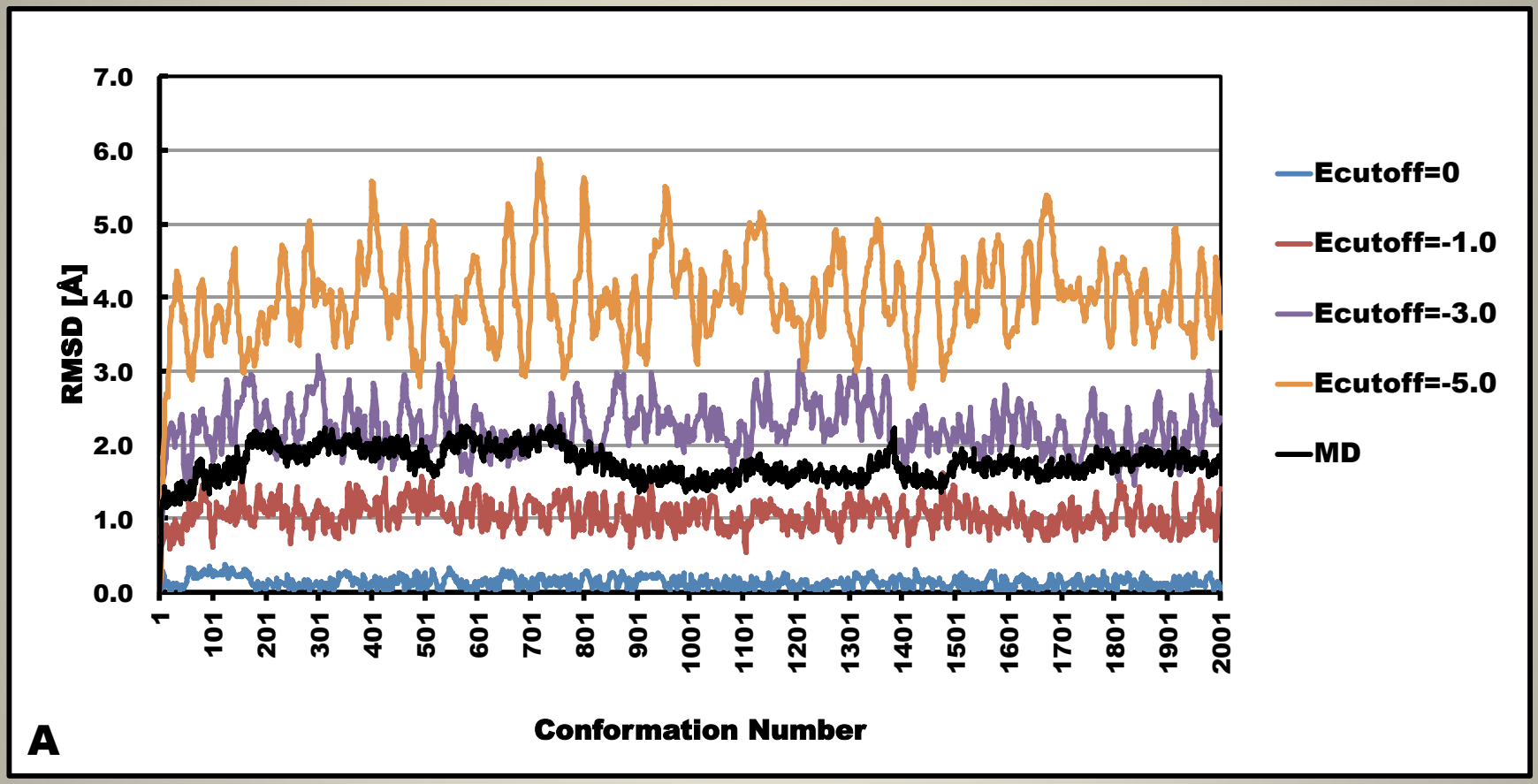
VS DIM = 3
SS DIM = 2
2 PAs

Results I

- FRODA yields trajectories very quickly

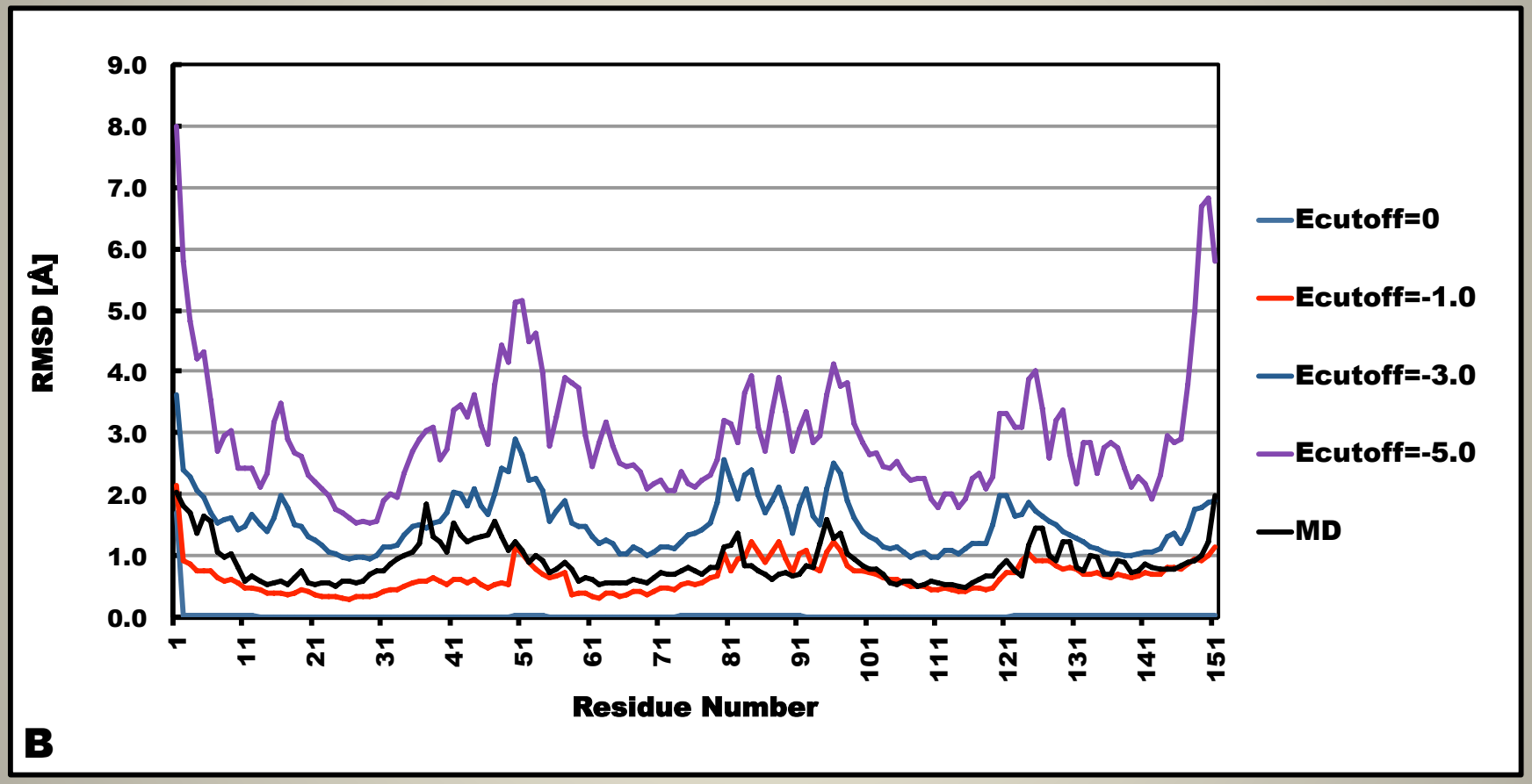
- The trajectories equilibrate rapidly when using the **momentum perturbation (MP)**

Conformation RMSD



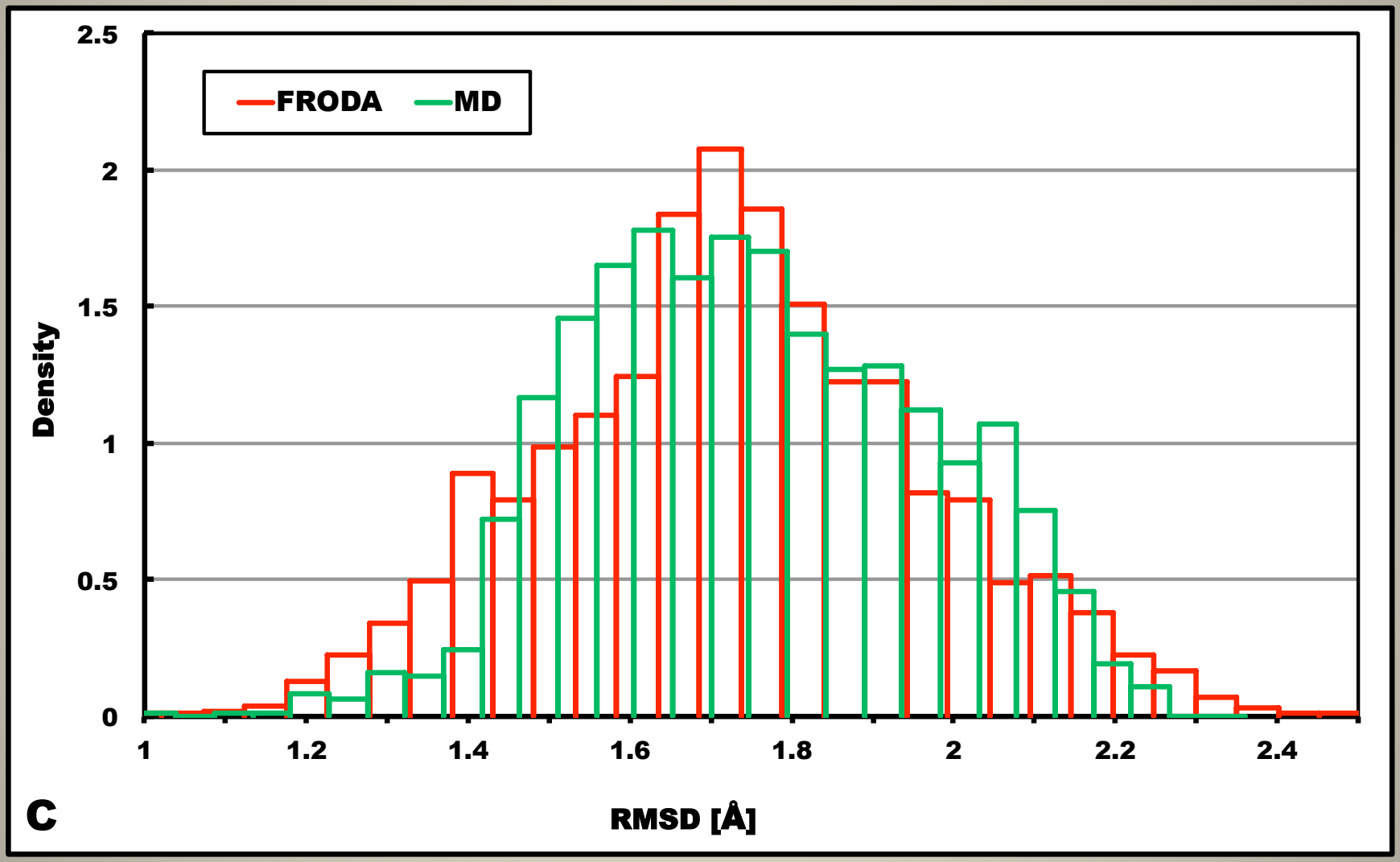
REFc

Residue RMSD



REFc

RMSD Distributions



Results II

- Momentum bias is very effective

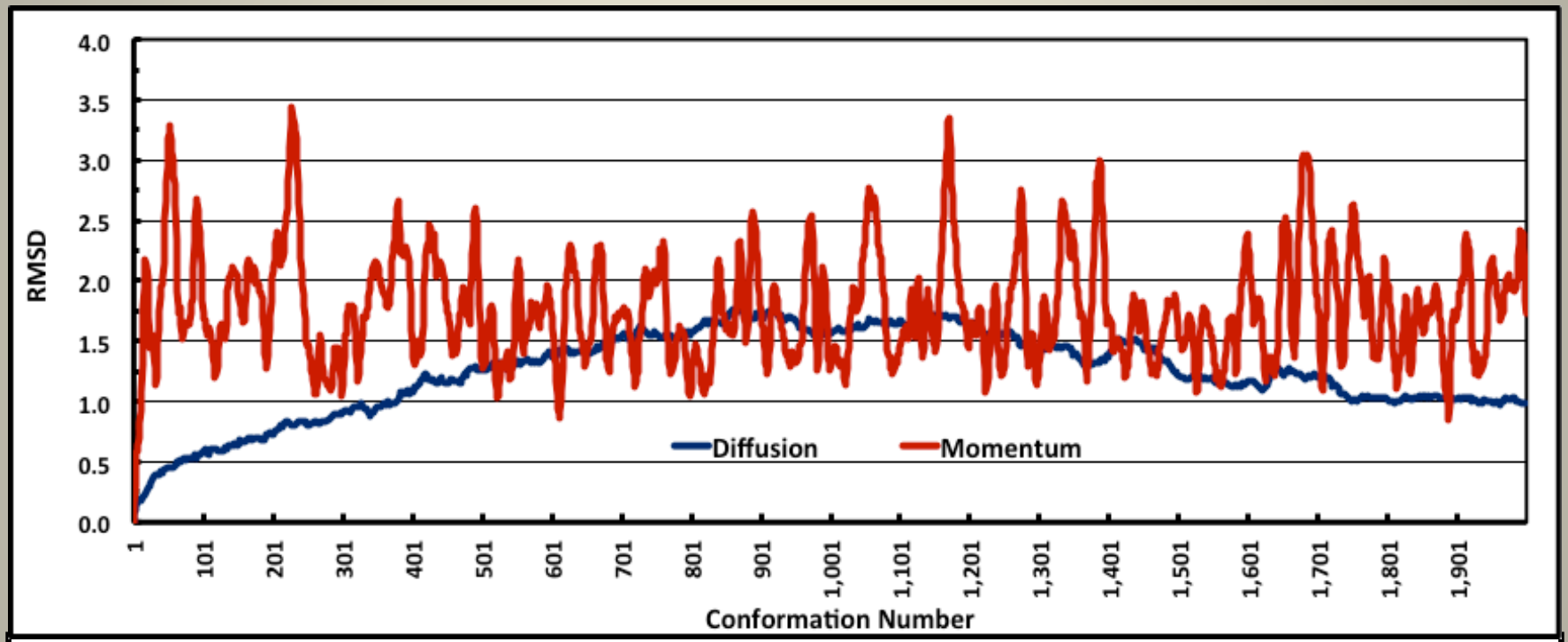
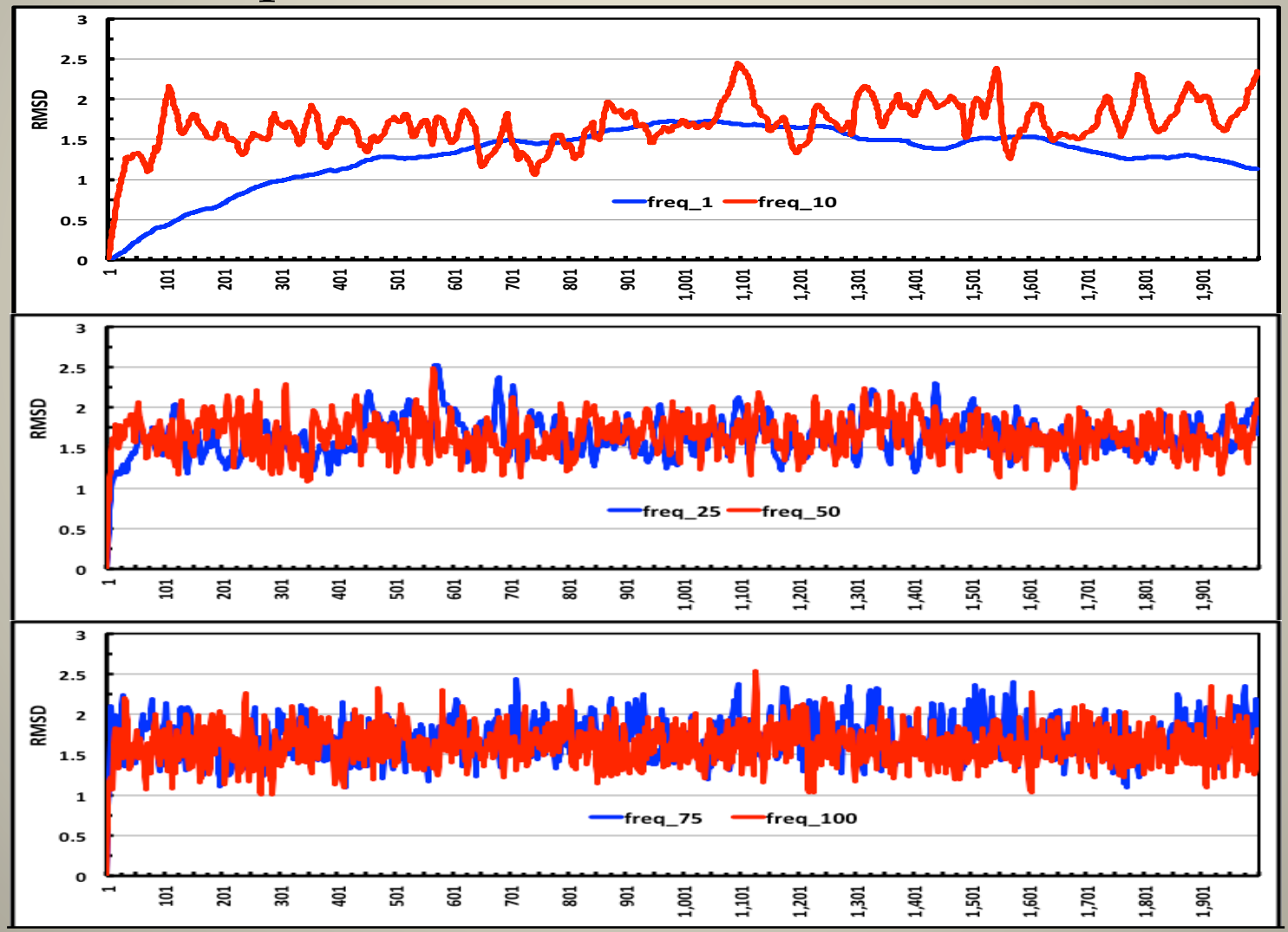


Figure 2.1 Conformation RMSD using FRODA Diffusion mode.
Running FRODA in diffusion mode yields trajectories that do not equilibrate rapidly due to low efficiency in how configuration space is sampled. These results are for MV in rigor state, which contains 946 residues.

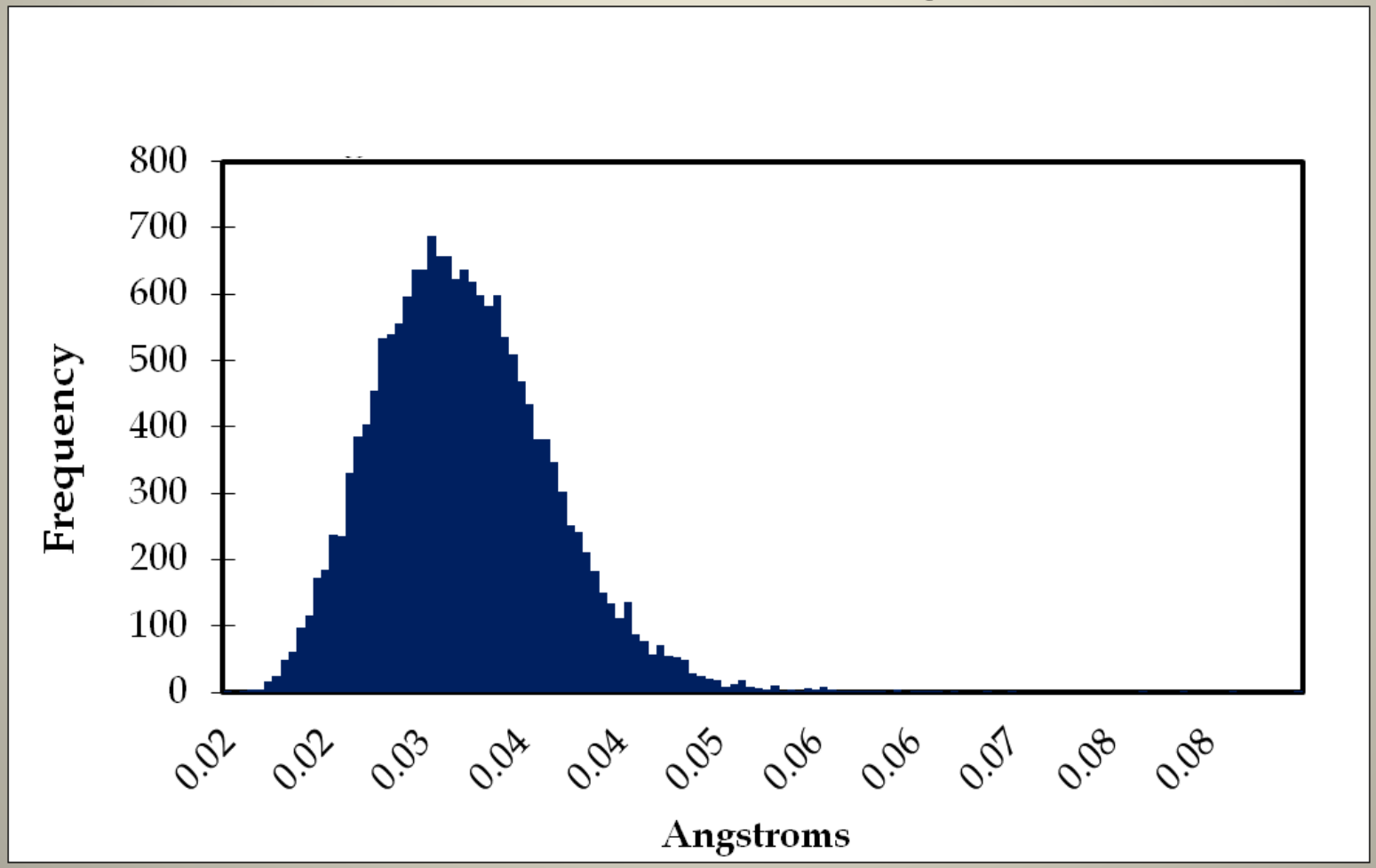
Results III

- Output frequency can be optimized
  - For most proteins: **10 ≤ FREQ ≤ 50**



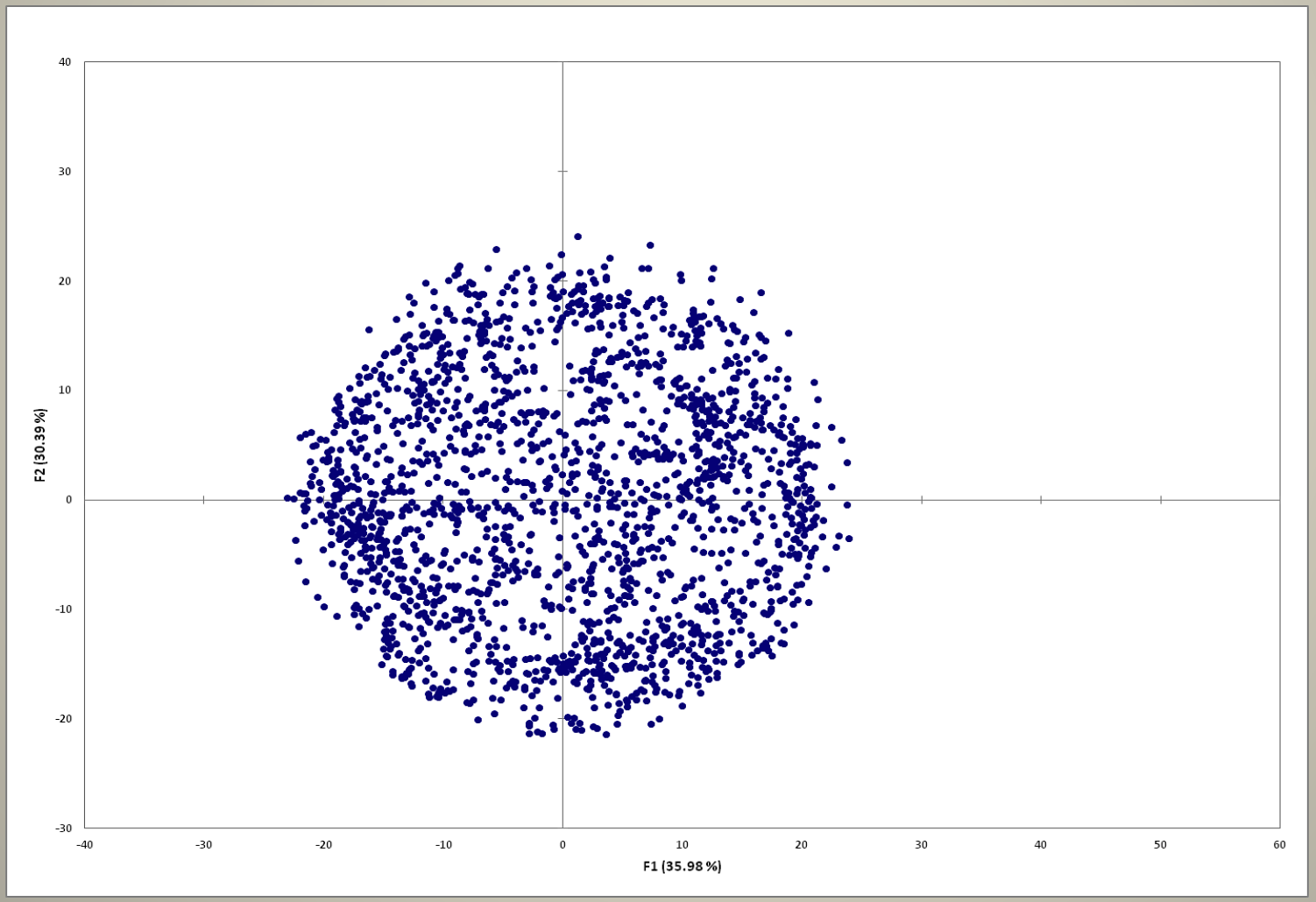
Results IV

- Constraint violations are minimized by:
  - **Step Size ≤ 0.01Å** when using MP



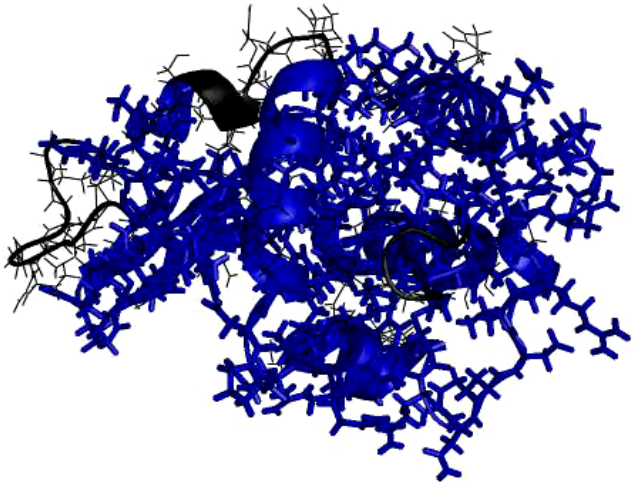
Results V

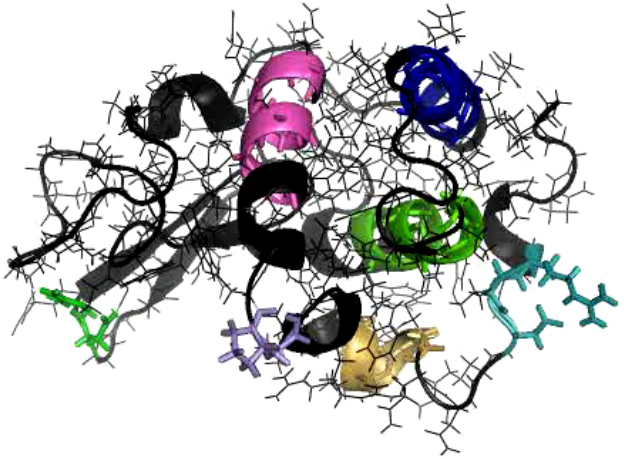
- Configuration space is well sampled based on the projections onto the top two PC modes



Results VI

- Dependency on model parameters

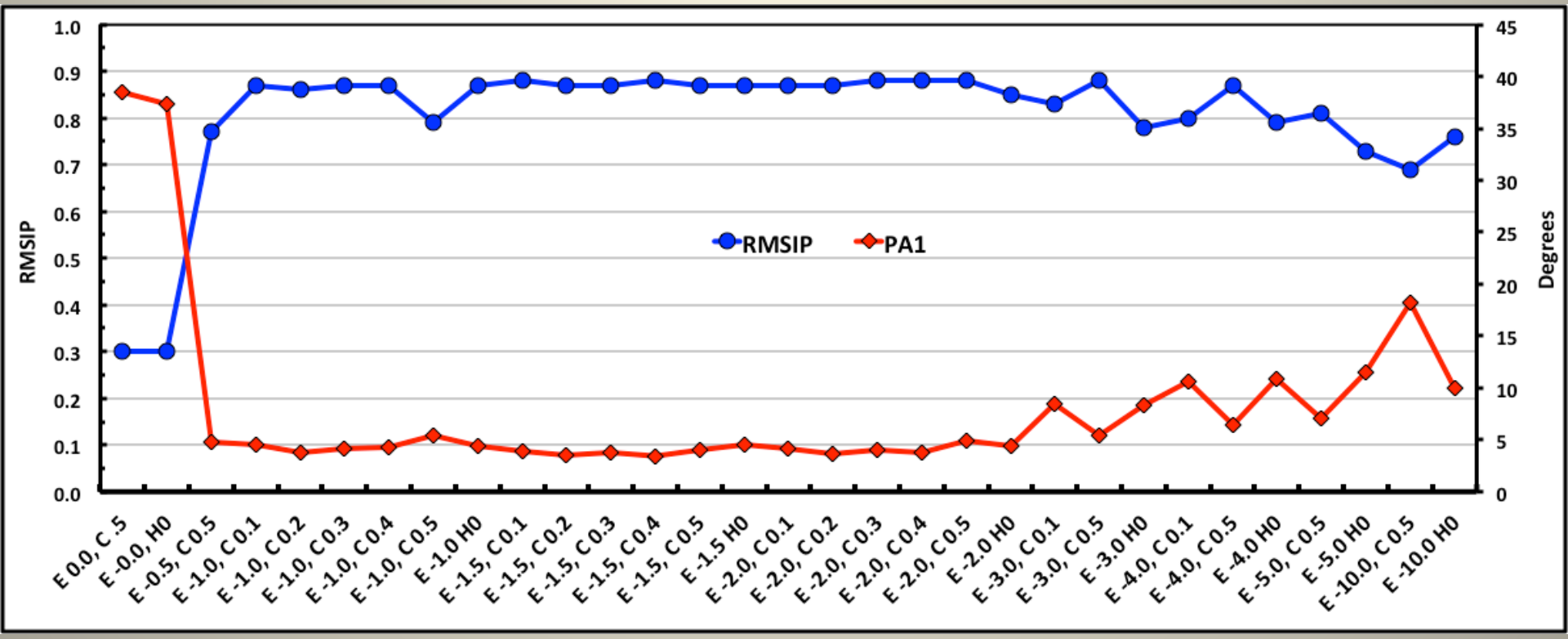




PDB ID 2nwd: Structure of chemically synthesized human lysozyme at 1 Angstrom resolution

Results VII

- There is a great degree of similarity in the essential spaces over a wide range of parameter choices
  - This suggests a **Range of Physicality**
    - **Amplitude vs. correlation in the motion**



Conclusions

- The GSM is very fast and yields trajectories that equilibrate rapidly when using the MP
- The simulations did not become irrevocably jammed nor did they yield structures containing large constraint violations
- Sampling of the native basin is efficient as assessed by projecting the DVs on the top few PCA modes
- The Essential Mode Spaces are highly conserved over a wide range of constraint assignment parameters as measured by RMSIP and PA
  - A Range of Physicality was observed
- The GSM well characterizes the native basin defined by the input structure

Model-to-Model Comparisons of the GSM, ANM, and MD

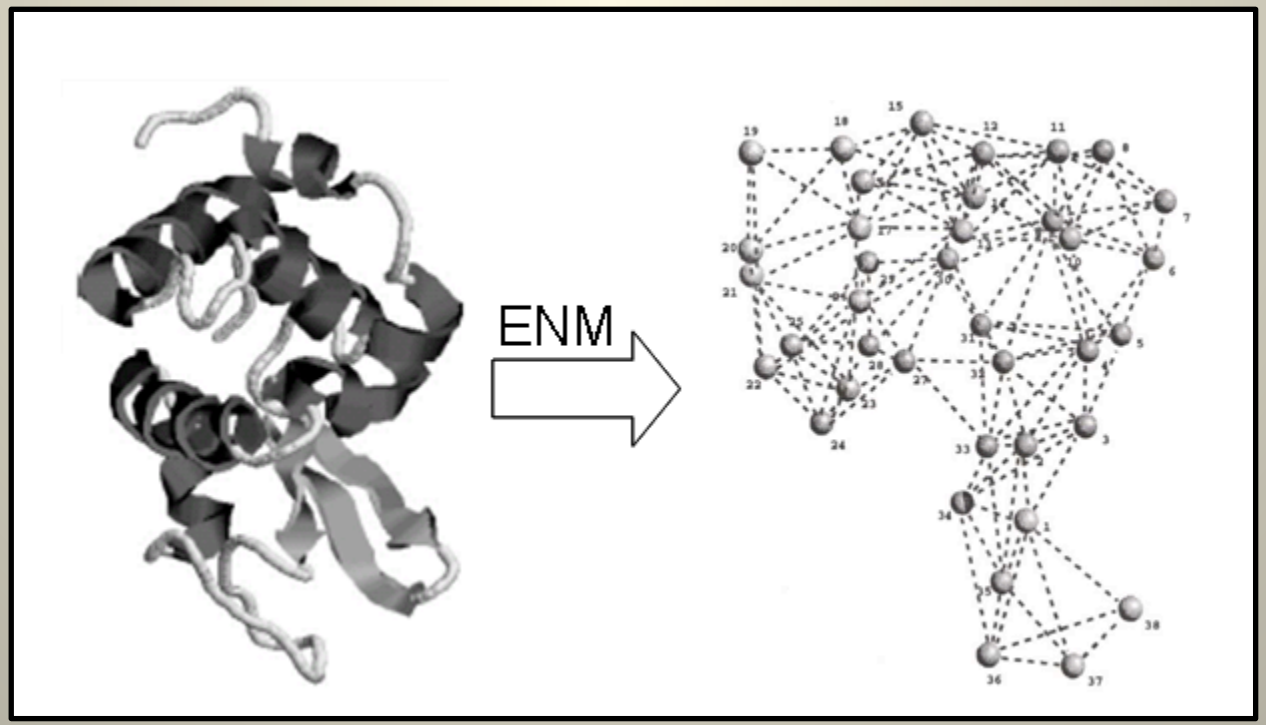
PART II

A Brief Introduction

THE MODELS

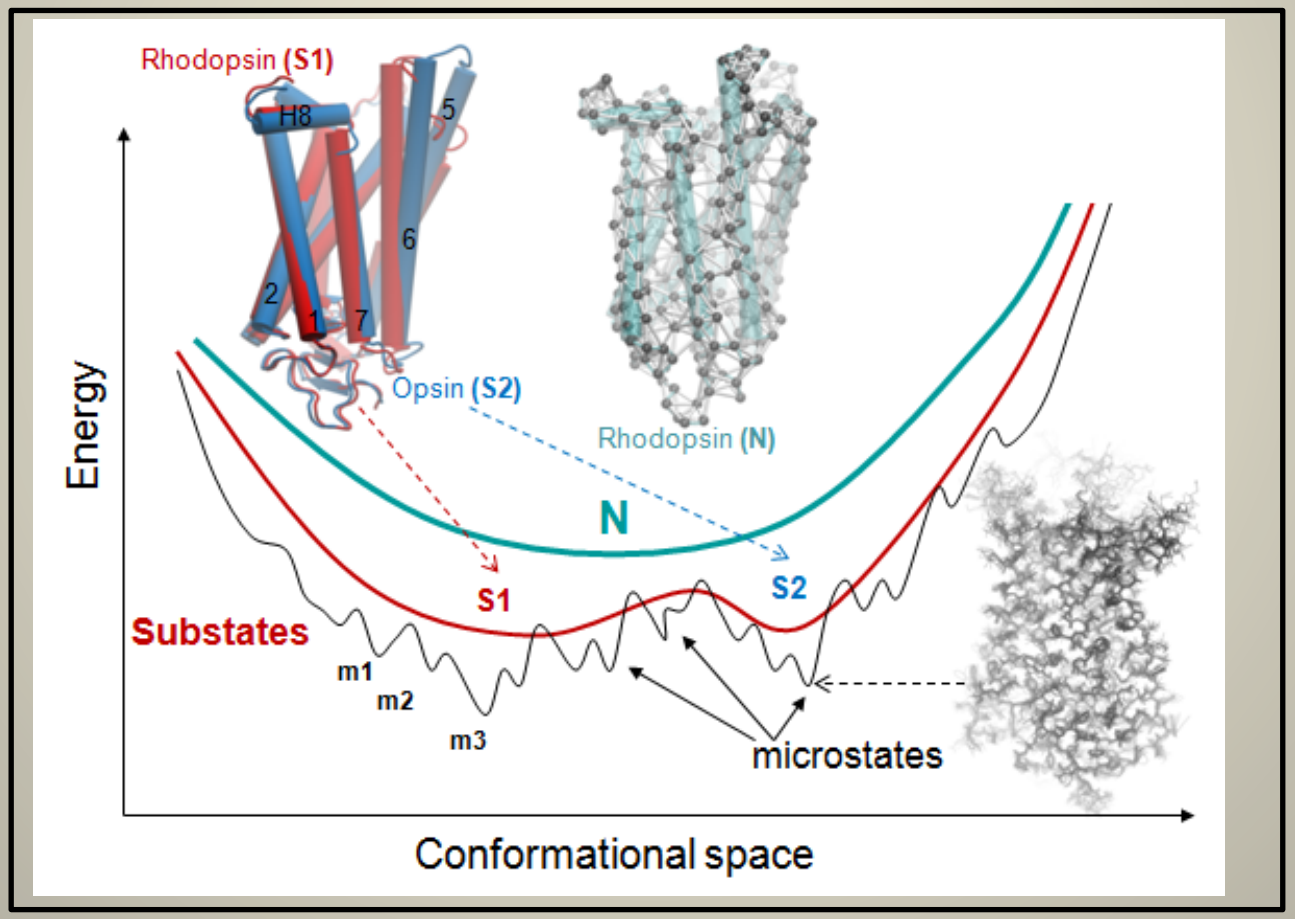
Models: ENM I

- A single point (C_α) represents each residue
- Each point forms a **node** on a graph while **edges** represent the interactions, within a cutoff distance
  - Typically a single value is used for all interactions
    - **Anisotropic Network Model (ANM)**



Models: ENM II

- The key simplifying assumption made in all ENMs is the **harmonic approximation**



Models: ENM III

- Normal Modes characterize the correlated motions of the protein
- **Longest timescales and largest amplitude motions are obtained from the lowest frequency modes**
- A coarse-grained model that requires little computational time, even for large proteins

Models: MD I

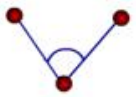
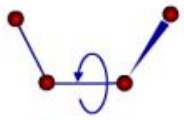
- A comprehensive all-atom model
  - **Classical**, not **QM**
- Basic assumption of the model:
  - Protein behavior can be elucidated by examining how it's molecular structure evolves through a set of molecular steps under the influence of a specified **potential** or **forcefield**

Models: MD II

- A typical MD potential:

$$V(\mathbf{r}) = \overset{(1)}{\sum_{\text{bonds}} K_b(b - b_0)^2} + \overset{(2)}{\sum_{\text{angles}} K_\theta(\theta - \theta_0)^2} + \overset{(3)}{\sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta))} + \sum_{\text{nonbonded-pairs},i,j} \left[\overset{(4)}{\frac{q_i q_j}{4\pi e_0 r_{ij}}} - \varepsilon_{ij} \left\{ \overset{(5)}{\left(\frac{R_{\min ij}}{r_{ij}}\right)^{12}} - 2\left(\frac{R_{\min ij}}{r_{ij}}\right)^{6} \right\} \right]$$

Energy dependencies on:

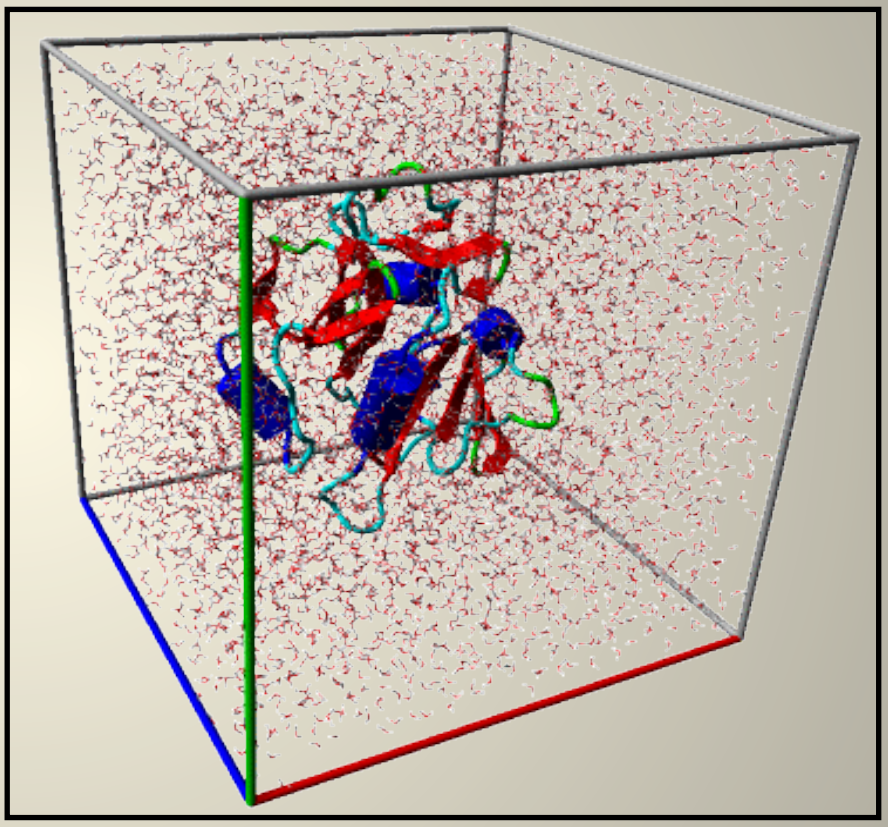
1. **Bond length** 
2. **Bond valence angle** 
3. **Bond dihedral angle** 
4. **Non-bonded electrostatic interactions** 
5. **Non-bonded van-der Waals interactions** 

Models: MD III

- **Thermal simulation:**

  Proteins are hydrated

  The entire system of protein plus solvent is equilibrated at a specific temperature



- The benefit:

  Trajectory represents an **ensemble** in the thermodynamic sense

Methods

- We selected 4 single domain proteins (monomers)
- Assessment of similarity was done by comparing the top 20 mode spaces from each of the models, plus an experimental set

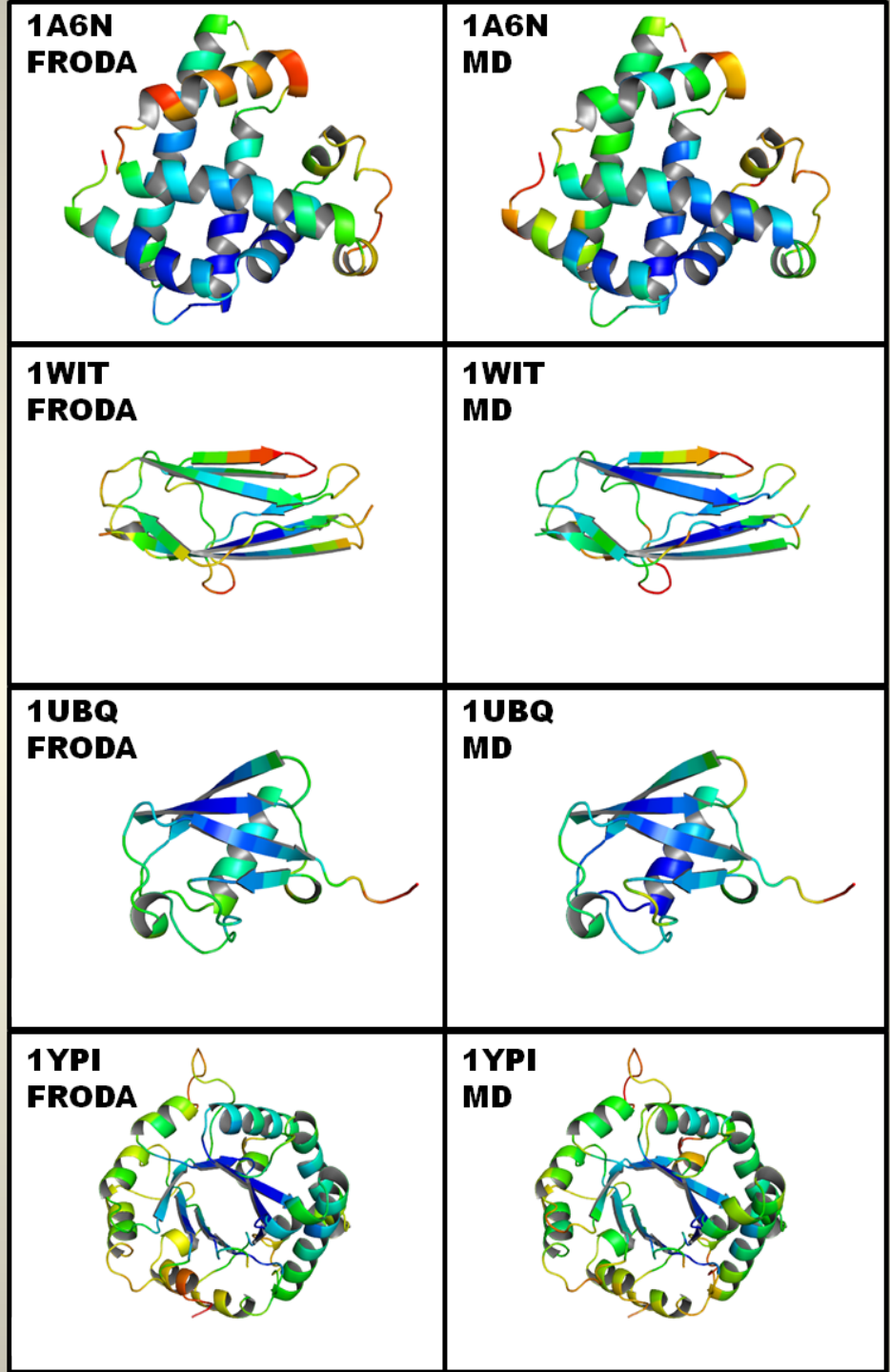
The 4 Target Proteins

- **1A6N** deoxy-myoglobin:
  SCOP class α, 151 residues
  - **The focus of these results**

- **1WIT** twitchin:
  SCOP class β, 93 residues

- **1UBQ** ubiquitin:
  SCOP class $\alpha+\beta$, 76 residues

- **1YPI** triosephosphate isomerase:
  SCOP class α/β, 247 residues

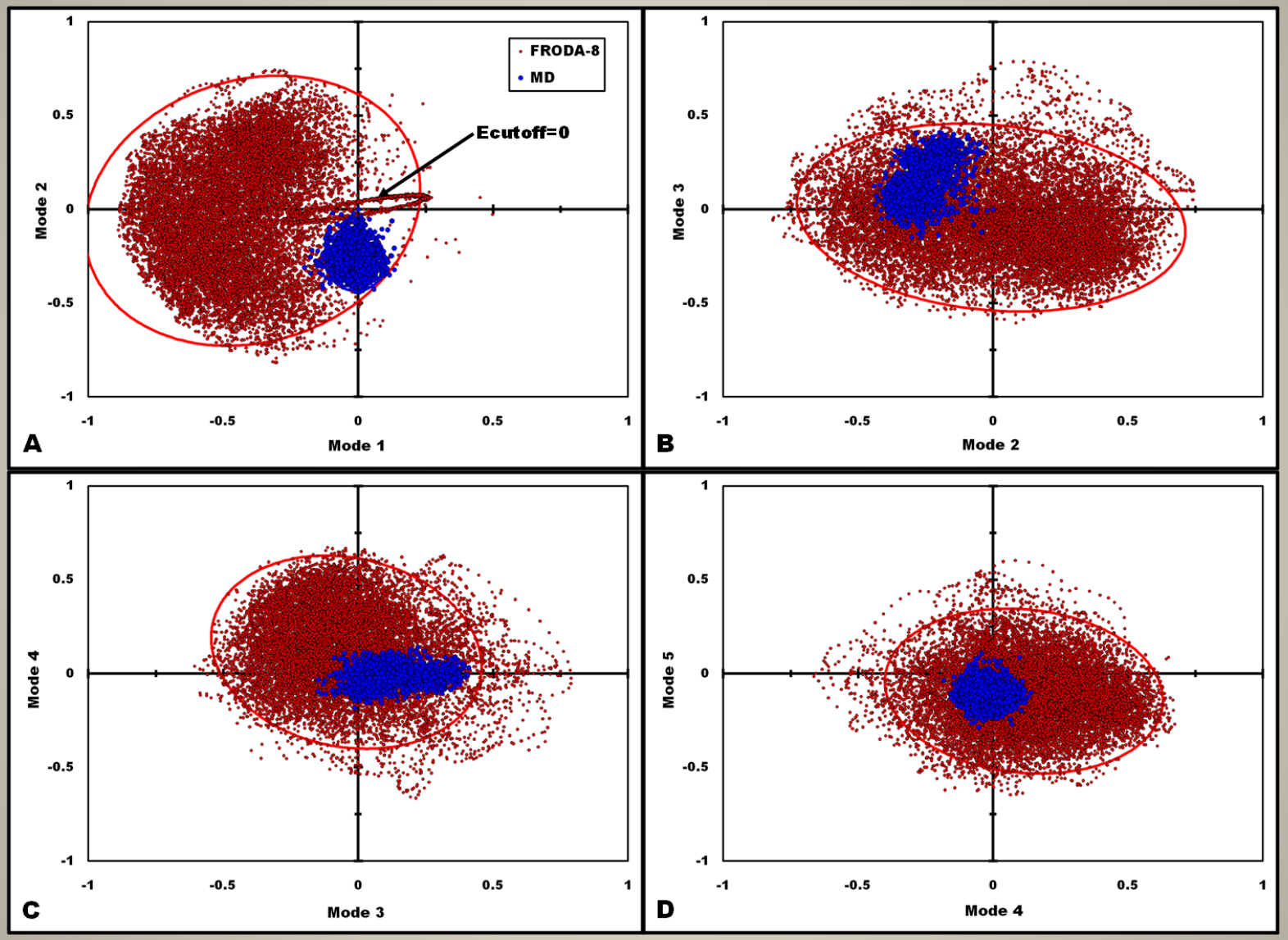


Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

Results I: PC Projections

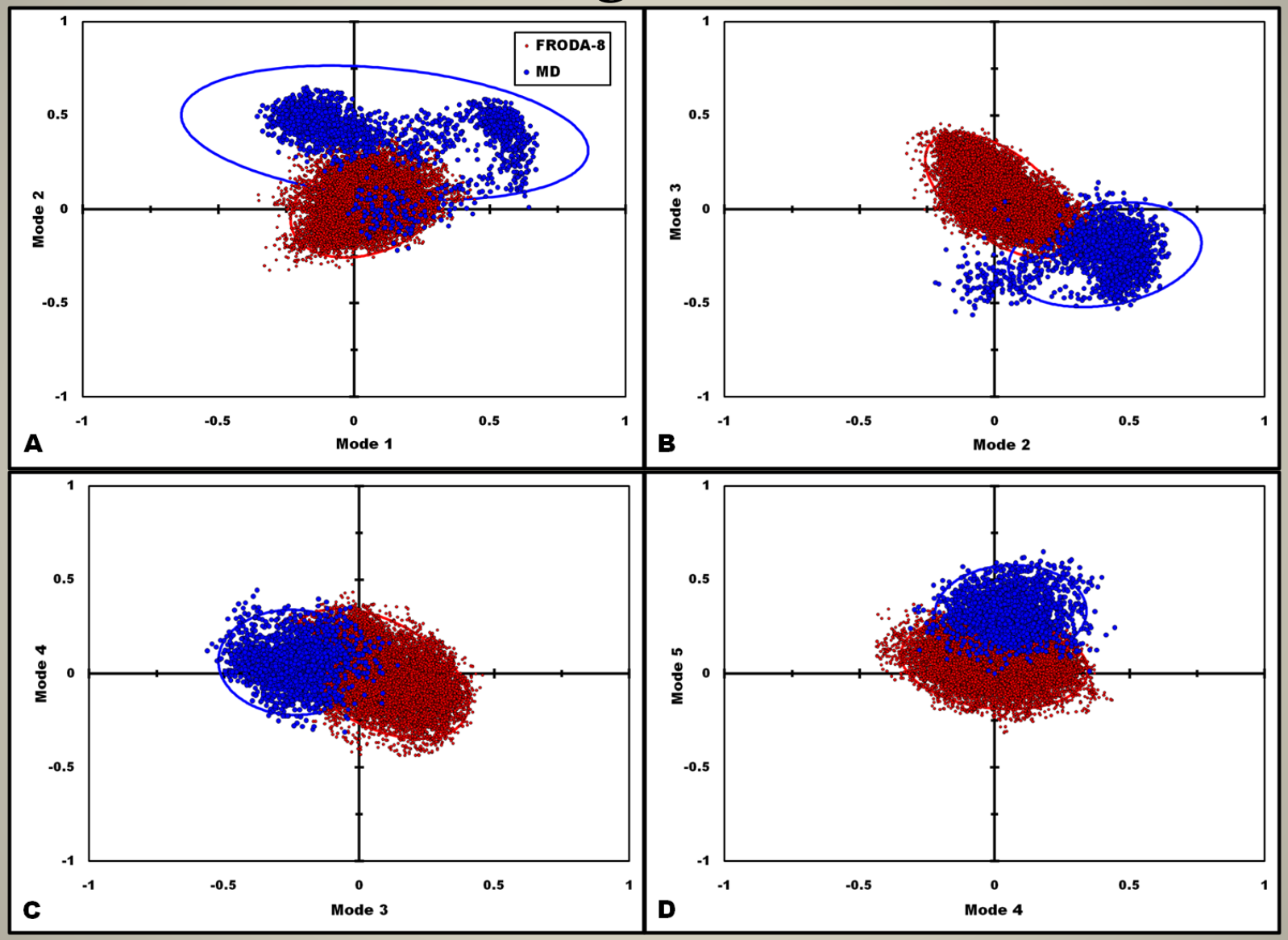
- PCs for the top 5 modes indicate degree of equilibration and mode space similarity
  - Fluctuations from FRODA and MD contrast a thermal and mechanical simulation

PCs Using FRODA Modes



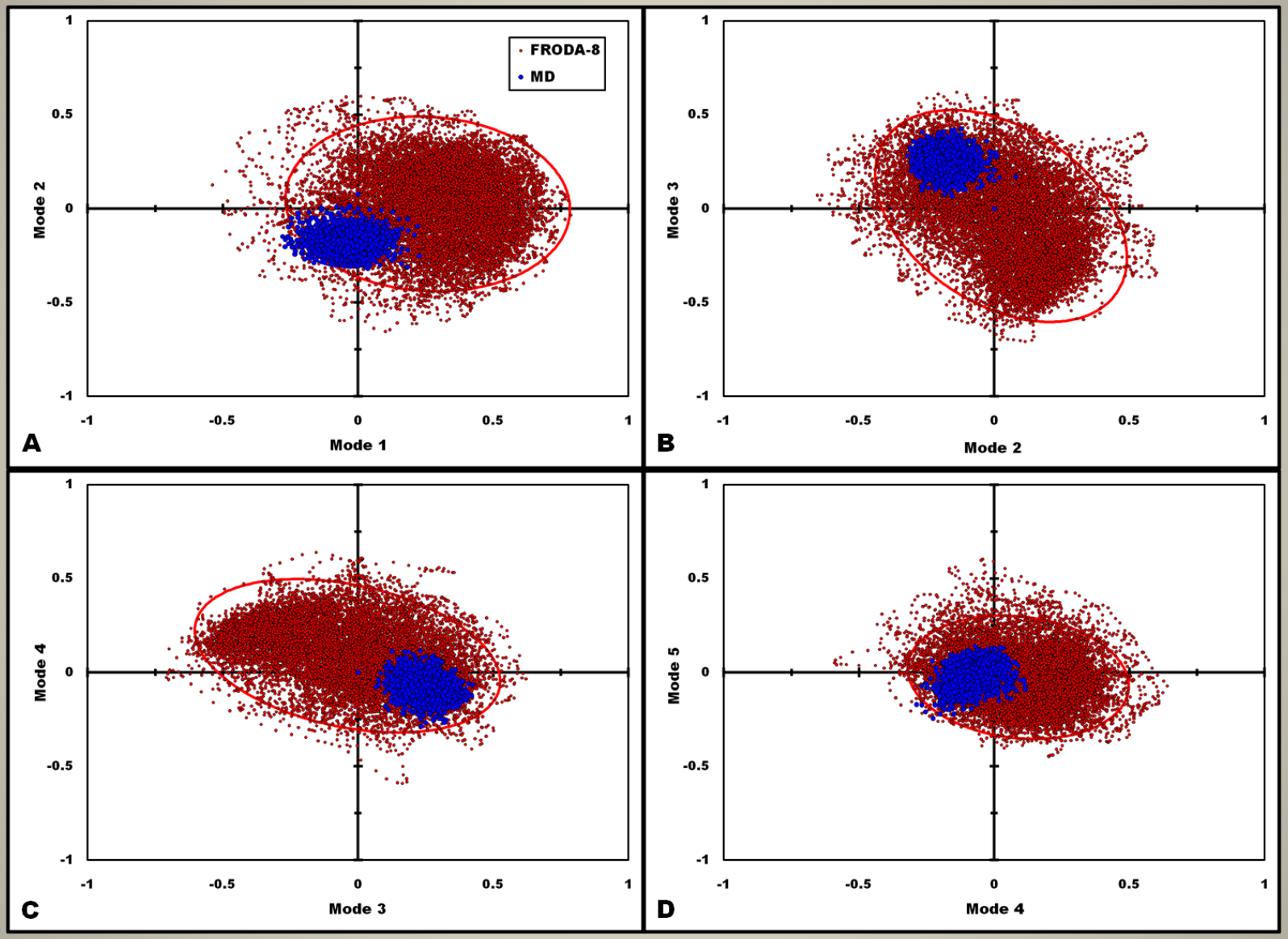
Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

PCs Using MD Modes



Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

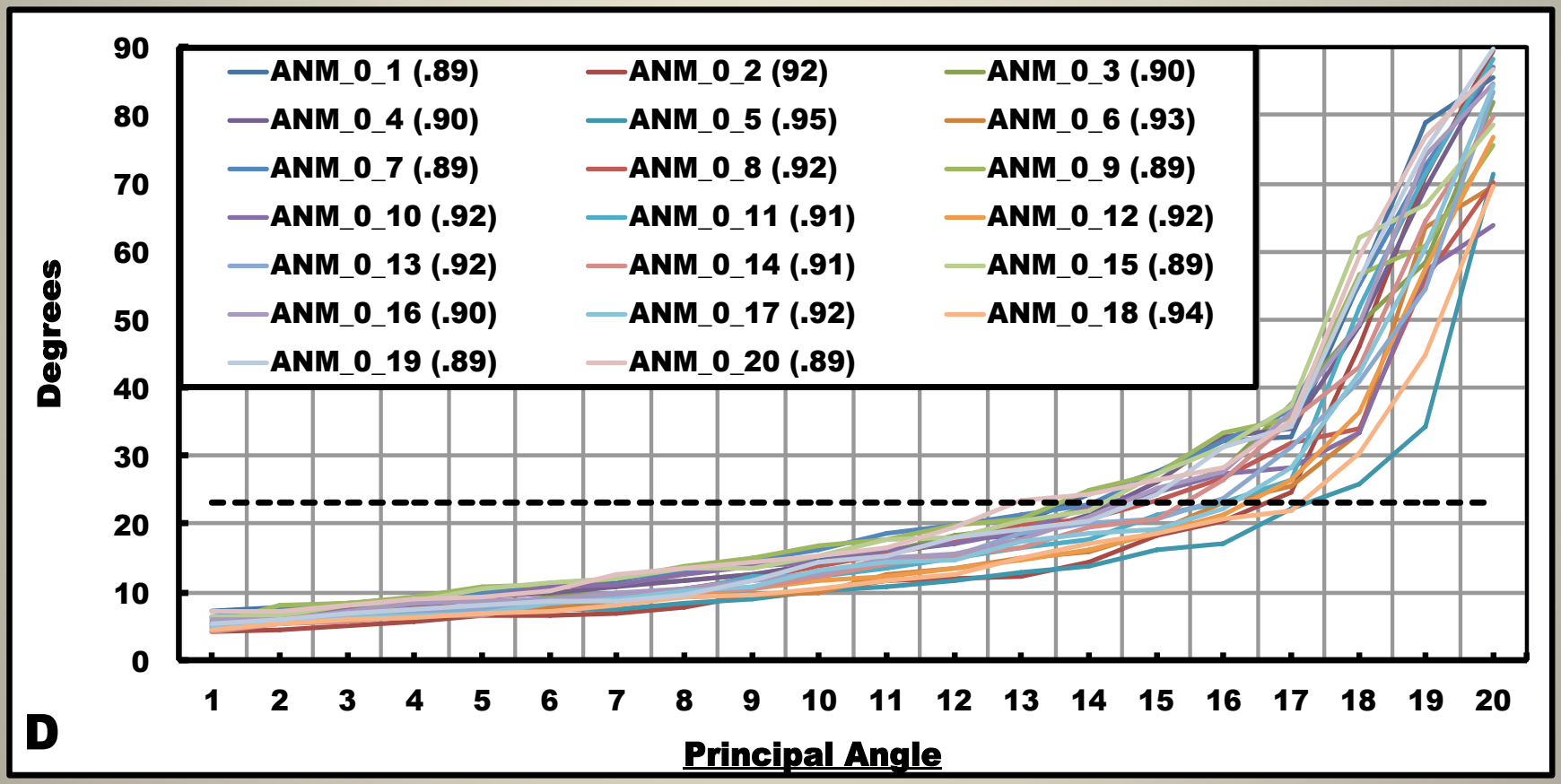
PCs Using ANM Modes



Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

ANM Consistency

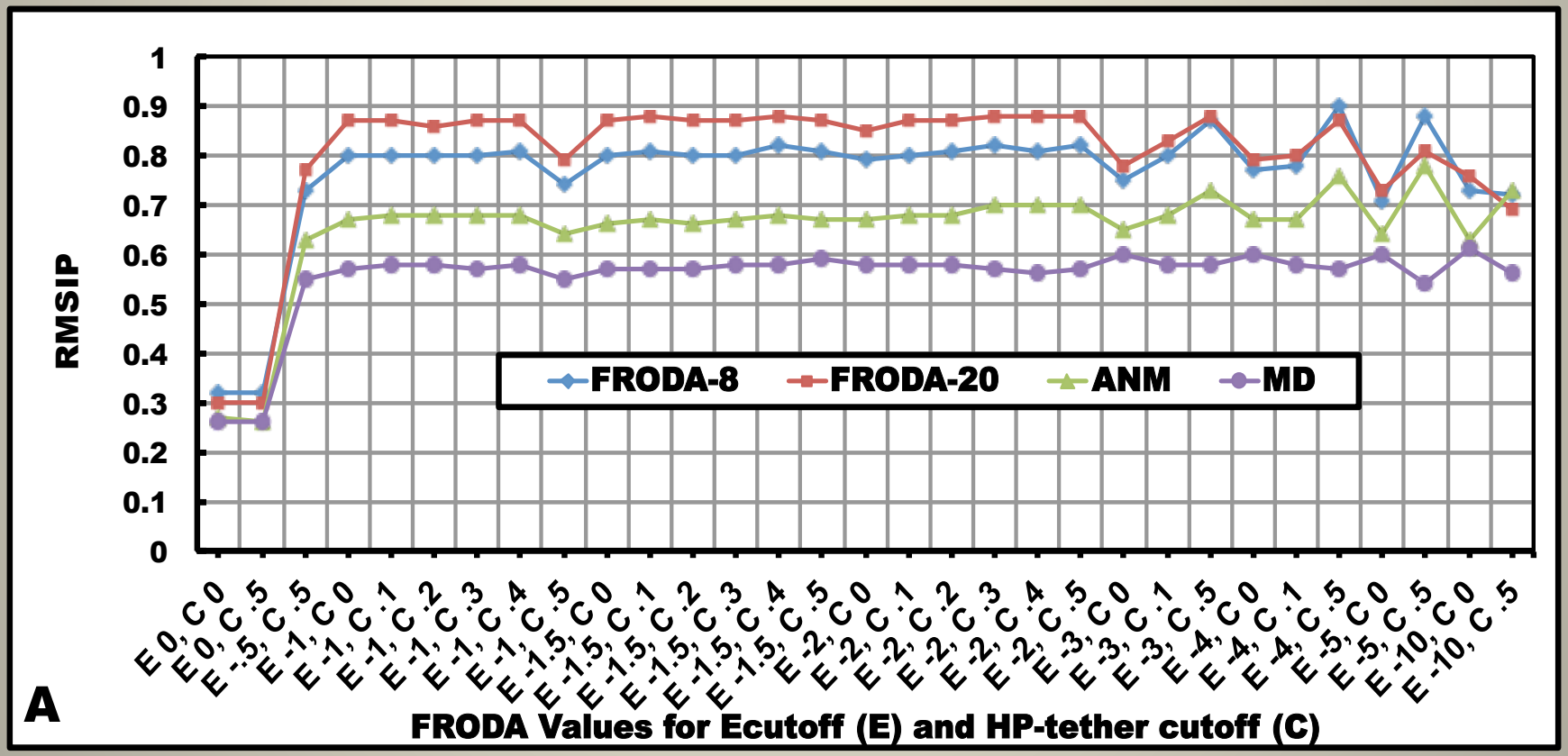
ANM essential subspaces are highly conserved when using multiple structures from a dynamic trajectory



Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

FRODA Consistency

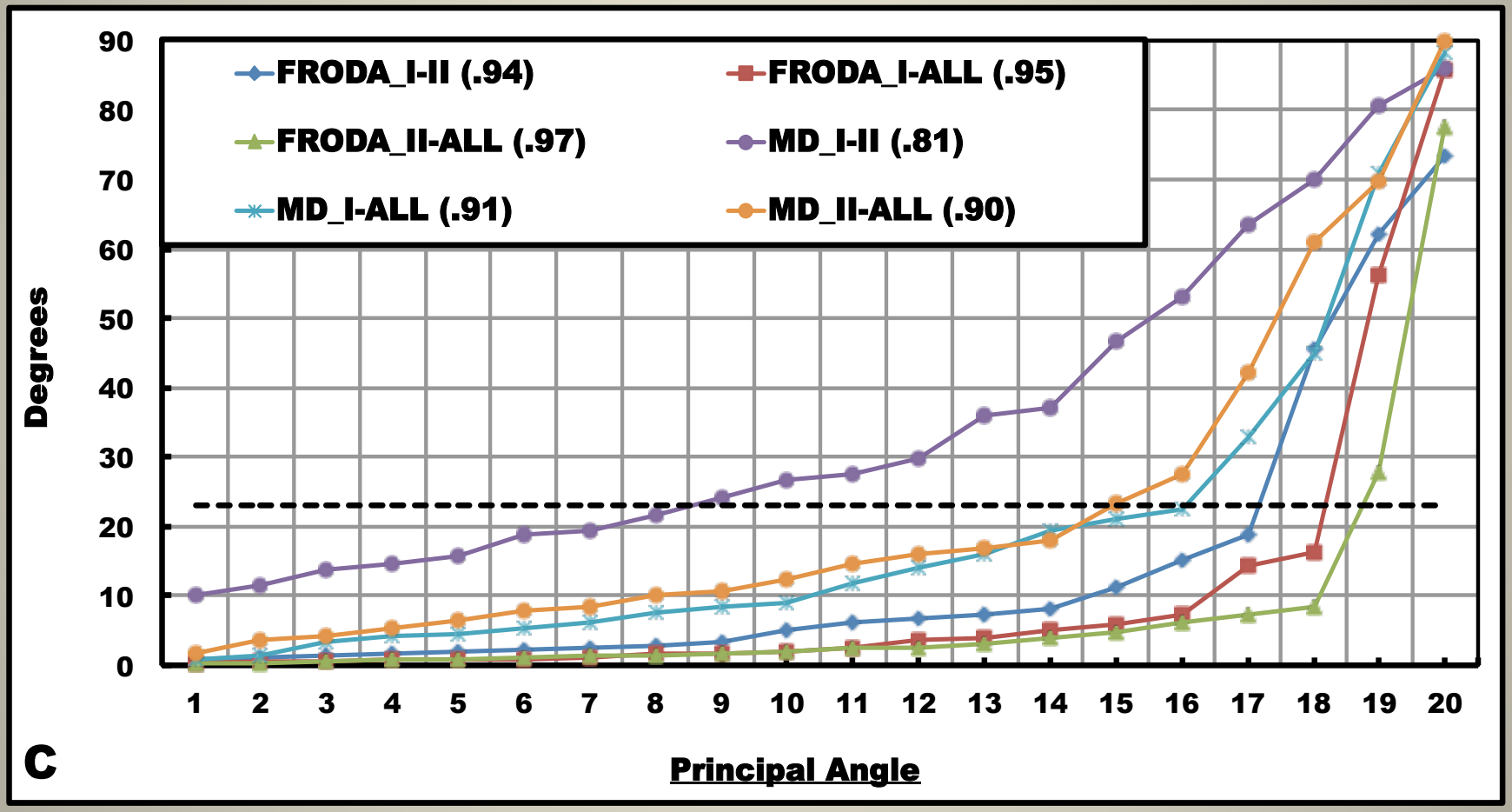
FRODA is consistent across a range of parameter choices that control rigidity/flexibility



Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

FRODA & MD Internal Consistency

FRODA and MD trajectories show internal consistency

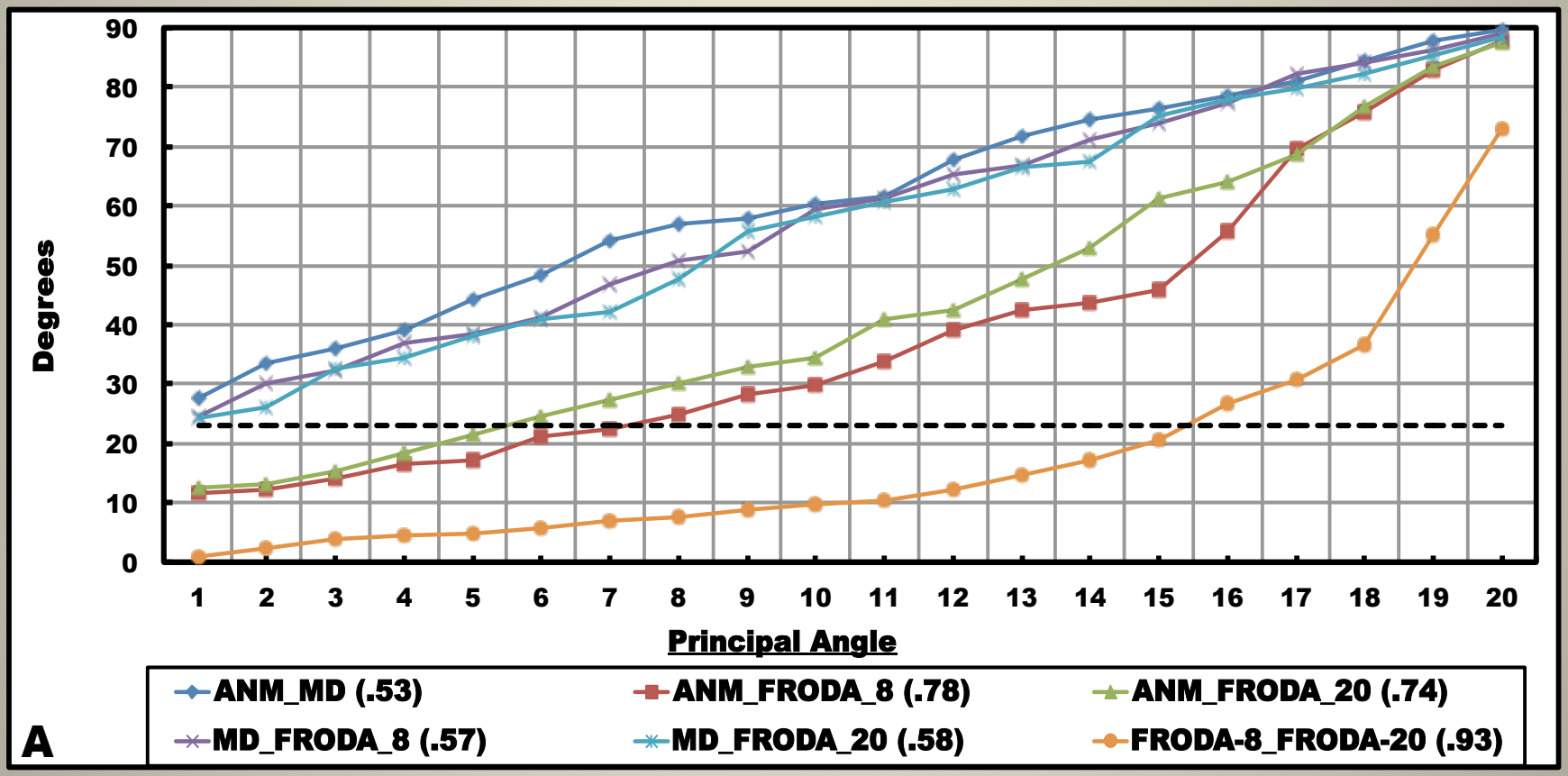


Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

Model Similarity

All three models share much in common

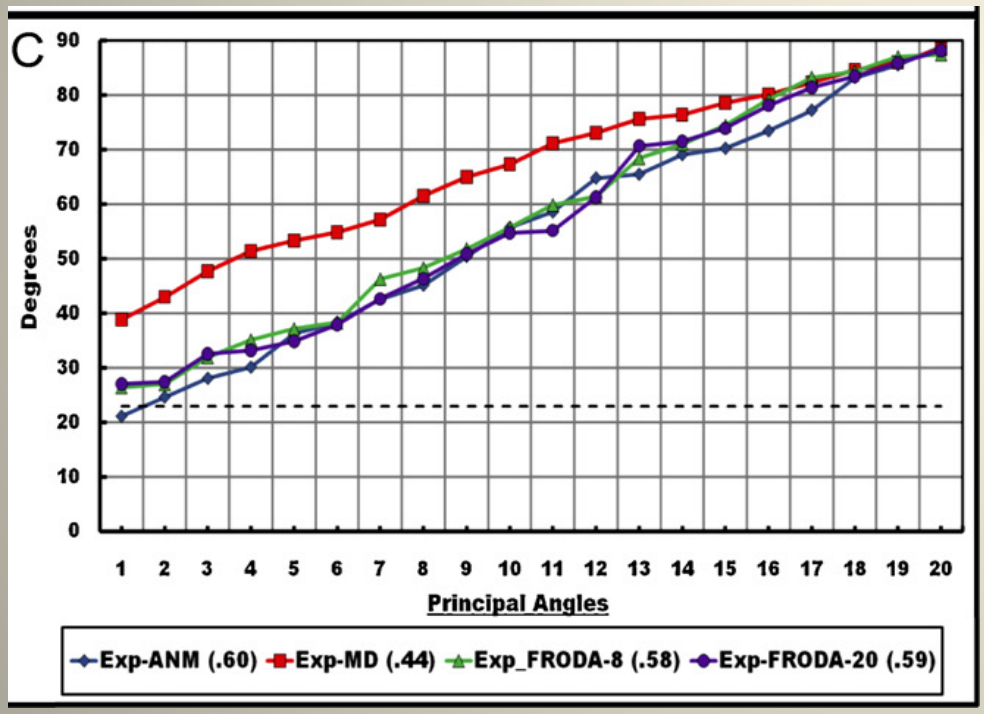
FRODA and ANM are much more alike than FRODA and MD or ANM and MD

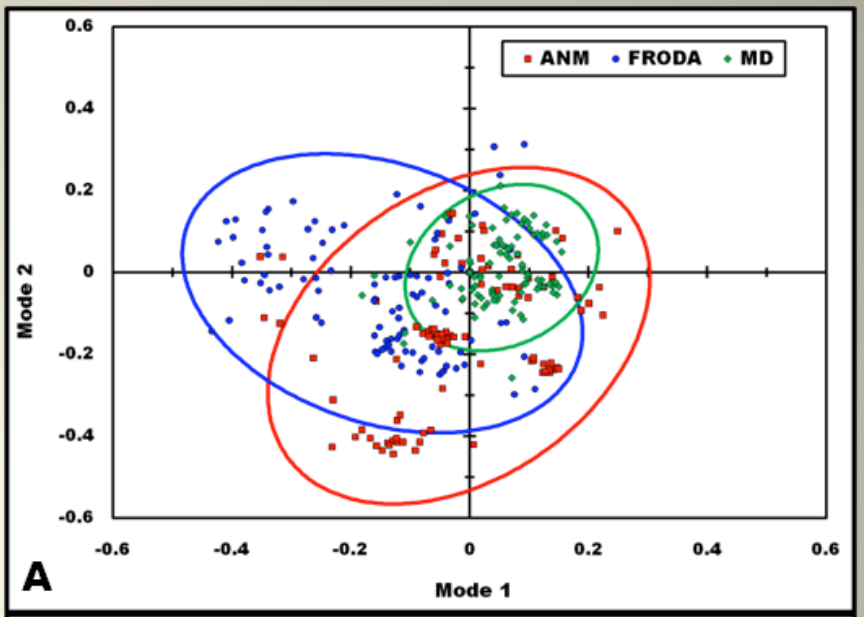


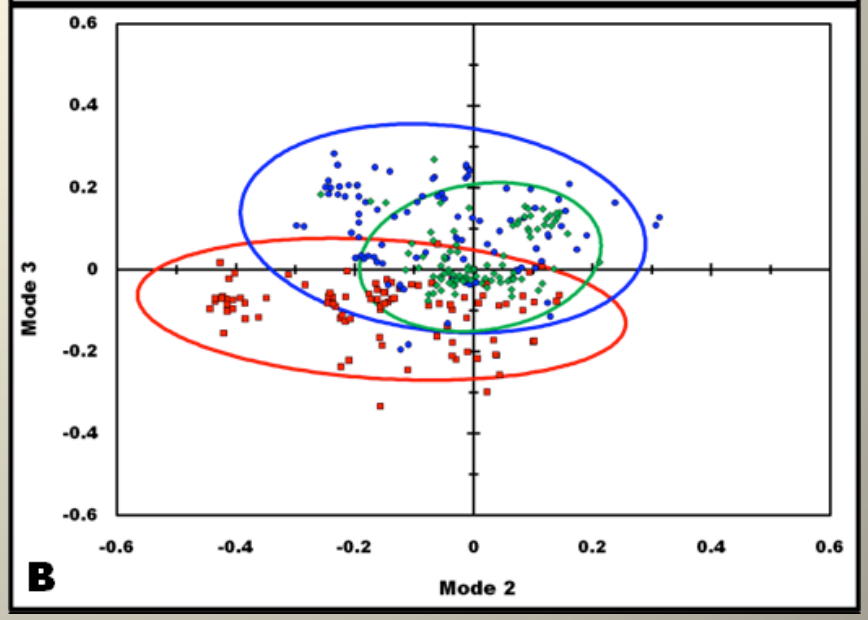
Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

Results IV

- FRODA captures a set of 100 experimental structures as well as ANM, and significantly better than MD







Characterizing Protein Motions from Structure, Charles C. David, Donald J. Jacobs, Journal of Molecular Graphics and Modelling 31 (2011) 41–56

Conclusions

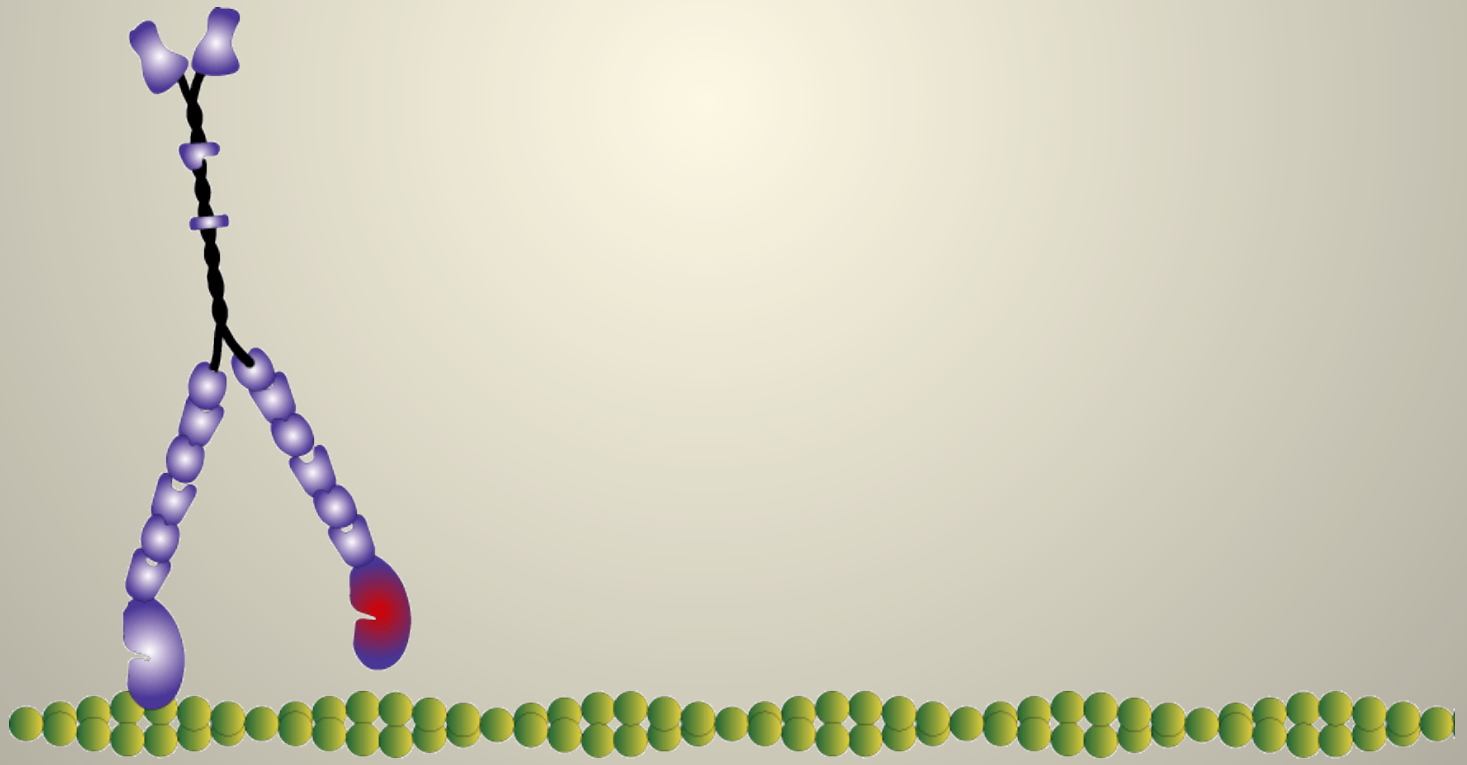
- The simulated structures project differentially into the model mode spaces, but reveal significant overlap in the Essential Subspaces
- The models show solid intra-consistency
- There is substantial inter-consistency
  - MD is unique in its ability to sample outside the native basin defined by the input structure
    - MD thus yields distinctive results
- The FRODA and ANM Essential Subspaces capture the experimental set of structures best

Application of the GSM to myosin V

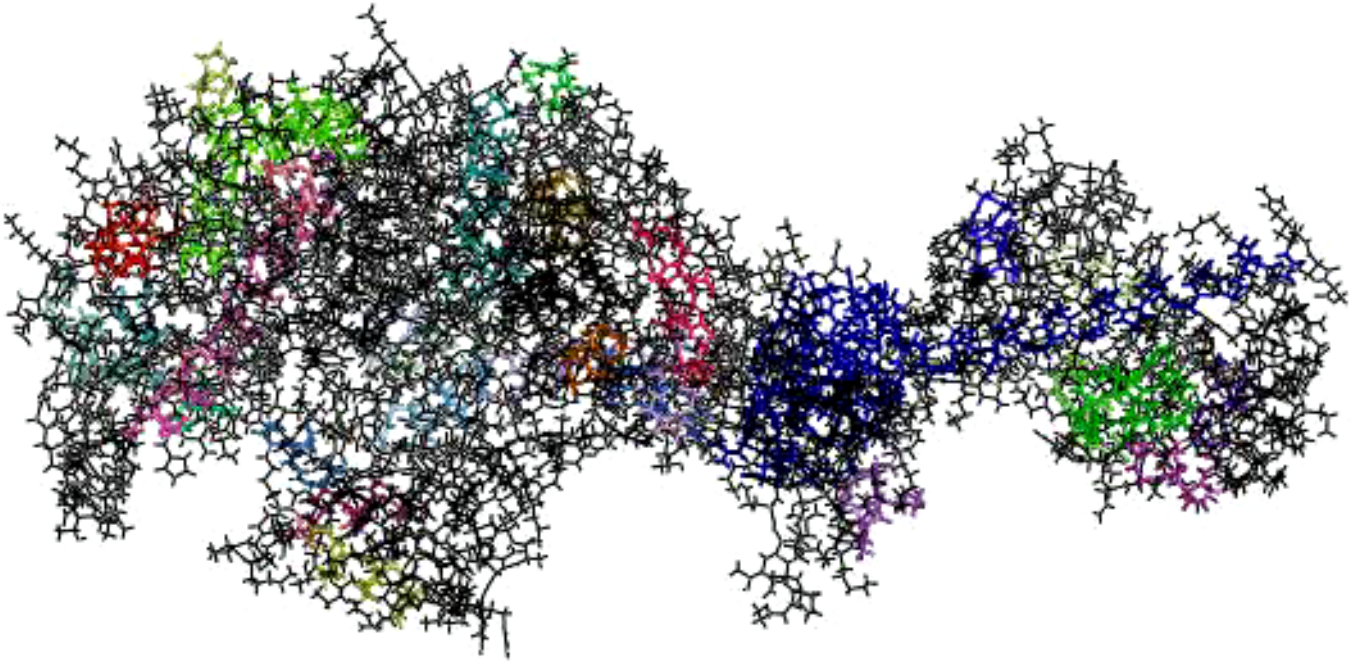
PART III

Introduction I

- Myosins are molecular motors capable of converting chemical energy into mechanical work through a cyclic interaction with actin filaments. Myosin V (MV) below:

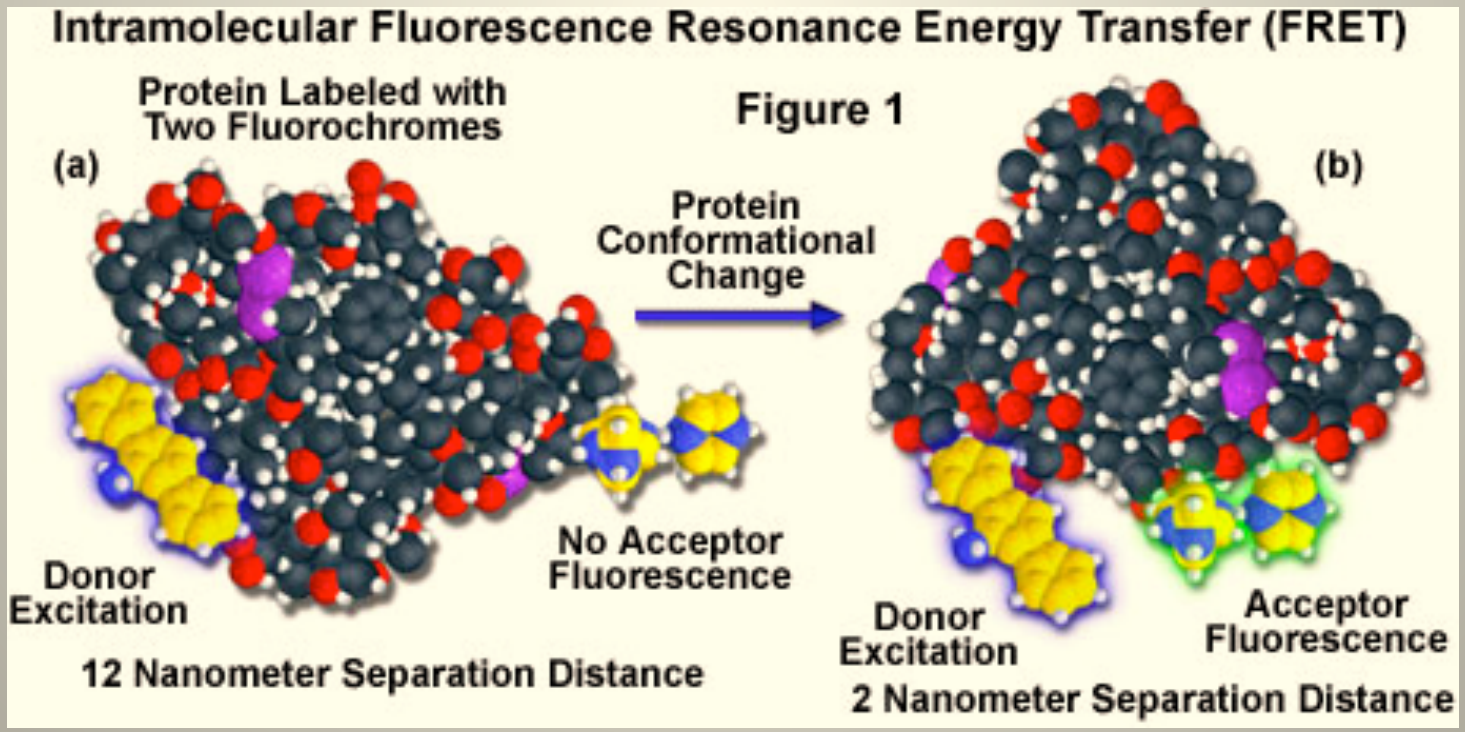


Myosin V Dynamics



Introduction II

- We collaborated with experimentalists who study myosins using FRET
- FRET allows us to determine the distance between a donor and acceptor probe



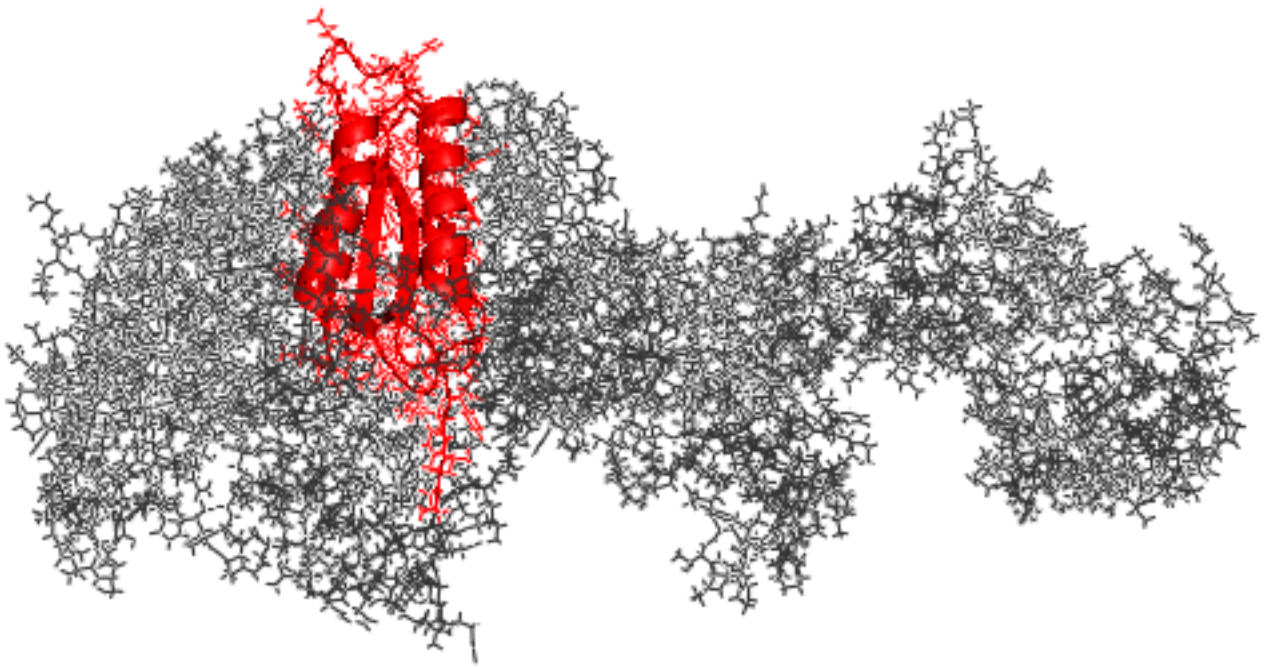
Applying the GSM to myosin V

PROJECT I

Introduction

- We selected a subset of 106 residues that defined the nucleotide-binding pocket (NBP)
- We investigated how conformational changes in the NBP are communicated to the lever arm and actin-binding cleft (ABC)

Myosin V showing NBP

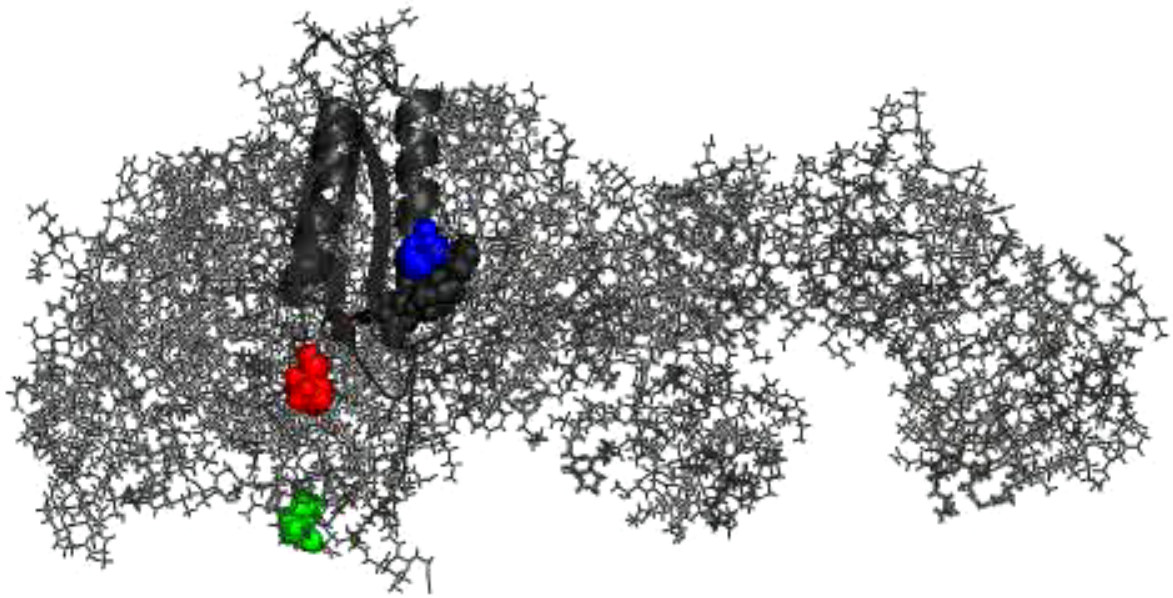


Methods

- We explore ED of MV using the GSM starting with three different X-ray crystal structures (1OE9, 1W7J, 1W7I)

- Our results are based on **dPCA**, compatible with our focus on three residues (171, 294, 525)

Myosin V Showing NBP, ATP, and residues 171, 294, 525



Results

- We obtained distinct distributions for the **171-294** distance in the MV.ATP structure and the MV.ADP structure
  - Thus, the MV.ADP structure is not capable of converting to the closed pocket conformation without significant constraint breaking

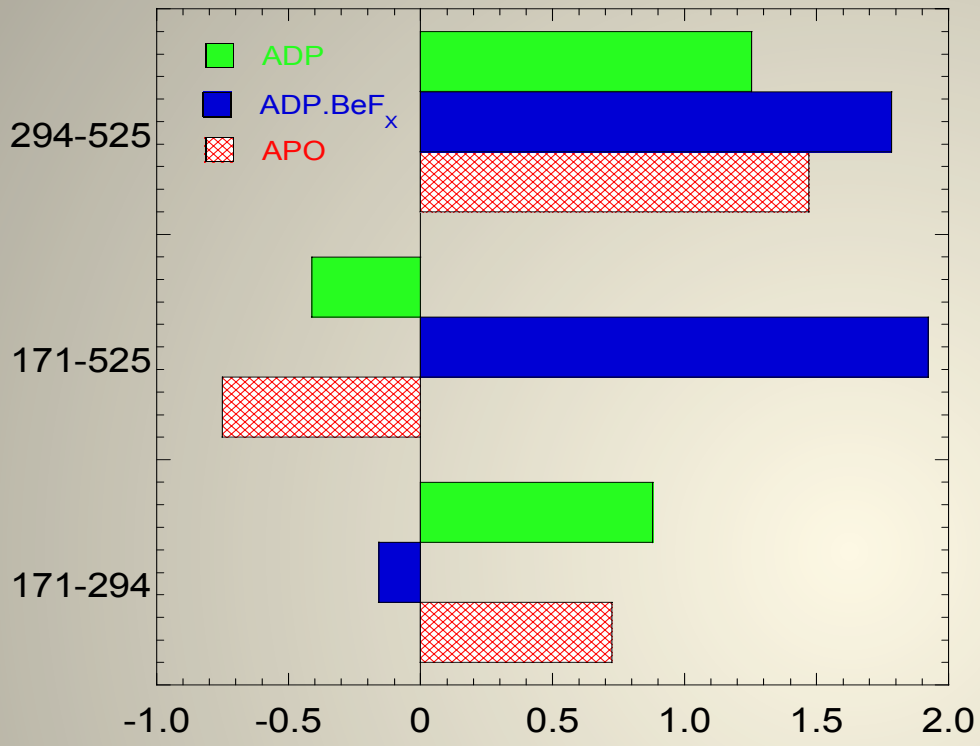
- For the MV.ADP state, very few modes of pocket opening/ closing agree with experiment
  - The crystal structure represents the **weak, open pocket MV.ADP** state

- We saw more evidence for a novel post-power-stroke MV.ADP state
  - NBP and ABC are both closed
  - This is the **strong closed pocket MV.ADP** state
  - This state has not yet been crystalized

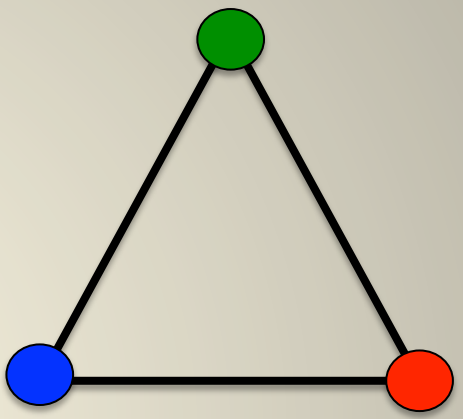
Jacobs, D.J., et al., Kinetics and thermodynamics of the rate-limiting conformational change in the actomyosin V mechanochemical cycle. J Mol Biol. 407(5): p. 716-30.

Sun, M., et al., Characterization of the pre-force-generation state in the actomyosin cross-bridge cycle. Proc Natl Acad Sci U S A, 2008. 105(25): p. 8631-6.

Mode 1 from dPCA



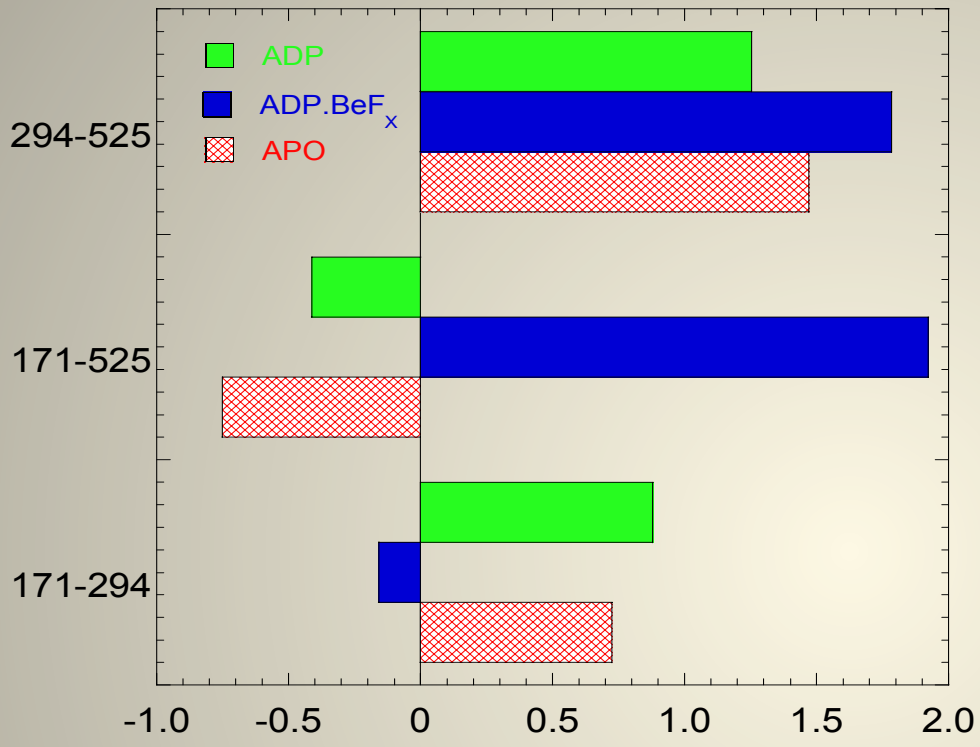
MV.ATP



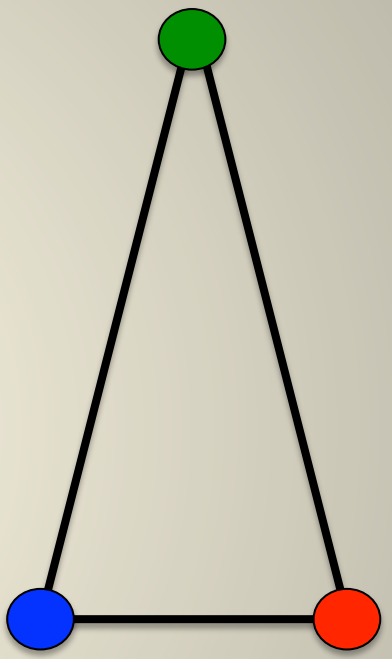
Results of **dPCA** analysis based on the relative motions of three residue pairs: 294-171, 171-525, and 294-525, in all three crystal structures. (ADP.BeF$_X$ is an ATP mimic)

Kinetics and Thermodynamics of the Rate Limiting Conformational Change in the myosin V Mechanochemical Cycle. (2011). Donald J. Jacobs, Darshan Trivedi, Charles C. David, and Christopher M. Yengo, Journal of Molecular Biology, Apr 15;407(5):716-30

Mode 1 from dPCA



MV.ATP



Results of **dPCA** analysis based on the relative motions of three residue pairs: 294-171, 171-525, and 294-525, in all three crystal structures. (ADP.BeF$_X$ is an ATP mimic)

Kinetics and Thermodynamics of the Rate Limiting Conformational Change in the myosin V Mechanochemical Cycle. (2011). Donald J. Jacobs, Darshan Trivedi, Charles C. David, and Christopher M. Yengo, Journal of Molecular Biology, Apr 15;407(5):716-30

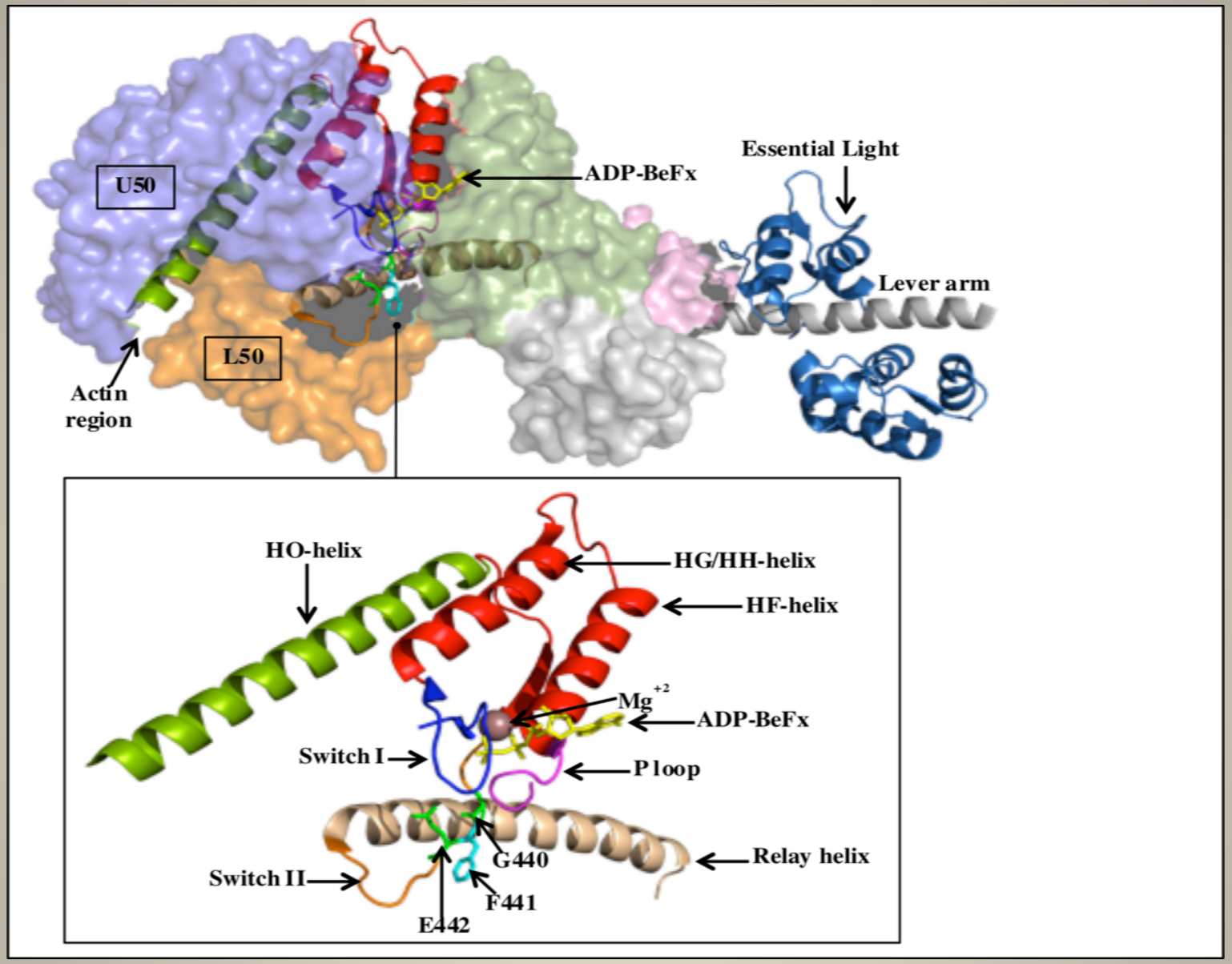
Applying the GSM to myosin V

PROJECT II

Introduction I

- Understanding the mechanism of force generation in MV requires elucidating allosteric communication pathways critical to motor function

- Three well conserved regions involved:
  - **The P-loop**
  - **Switch I**
  - **Switch II**



Switch II Mutants Reveal Coupling Between the Nucleotide- and Actin-Binding Regions in Myosin V
Darshan V. Trivedi, Charles C. David, Donald J. Jacobs, and Christopher M. Yengo,
Biophysical Journal Volume 102 Issue 11 pp.2545-2555. doi:10.1016/j.bpj.2012.04.025

Introduction II

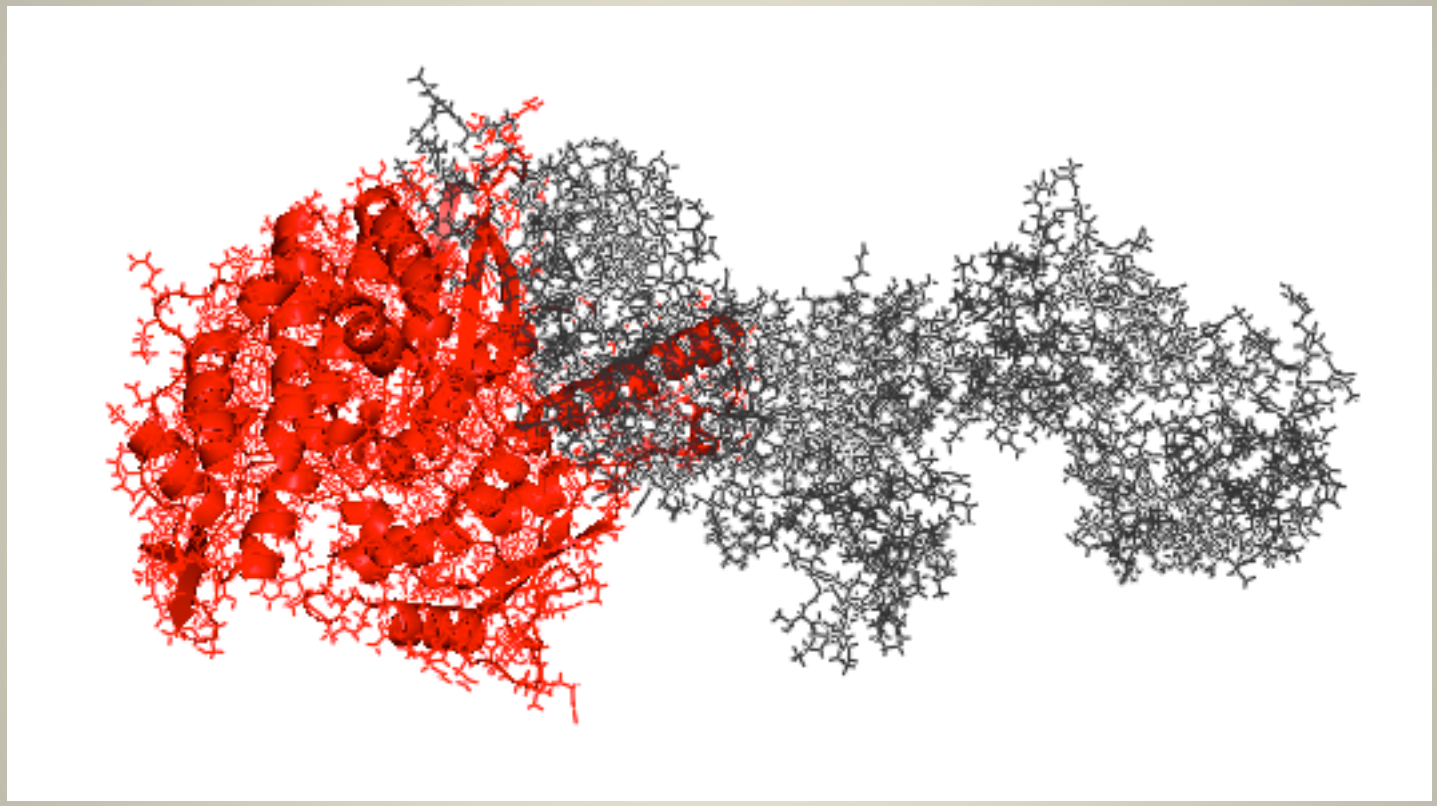
- Switch I is chiefly involved in the communication pathways between the active site and the ABC
- Switch II mediates the communication to the converter-lever arm domain
- To understand how Switch II affects the conformational dynamics of the NBP and ABC, we introduced two single site mutations
  - **G440A**: Eliminates a highly conserved hydrogen bond to the gamma phosphate of ATP
  - **E442A**: Eliminates a highly conserved salt bridge between Switch I and Switch II

Methods

- We defined three subsets:
  - NBP [106]
  - Actin Binding Region (ABR) [455]
  - A proposed Communication Pathway (CP) [89]

- Conformational changes in these subsets were investigated using GSM with ED

Myosin V showing ABR



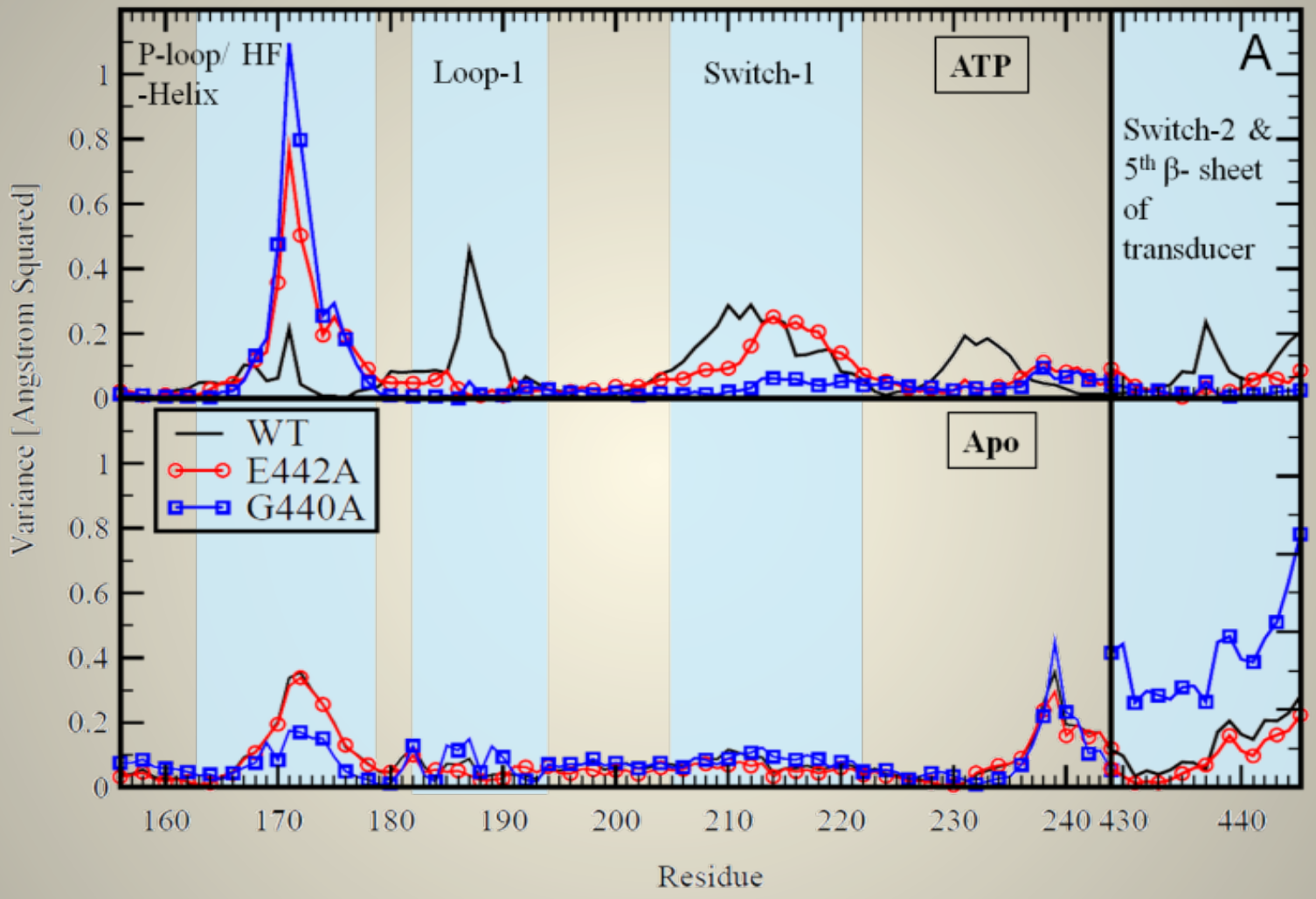
Results I

- Switch I-Switch II interactions stabilize the closed nucleotide binding pocket conformation in the absence of actin

- The NBP in the ATP state had reduced dynamics of Switch I in the G440A mutant
  - This hinders Switch I-Switch II interactions
  - Reduces the stability of the closed NBP conformation

- The mobility of the P loop is increased by both Switch II mutants

The first PC mode of the NBP of mutant and wild-type myosin V



Switch II Mutants Reveal Coupling Between the Nucleotide- and Actin-Binding Regions in Myosin V
Darshan V. Trivedi, Charles C. David, Donald J. Jacobs, and Christopher M. Yengo,
Biophysical Journal Volume 102 Issue 11 pp.2545-2555. doi:10.1016/j.bpj.2012.04.025

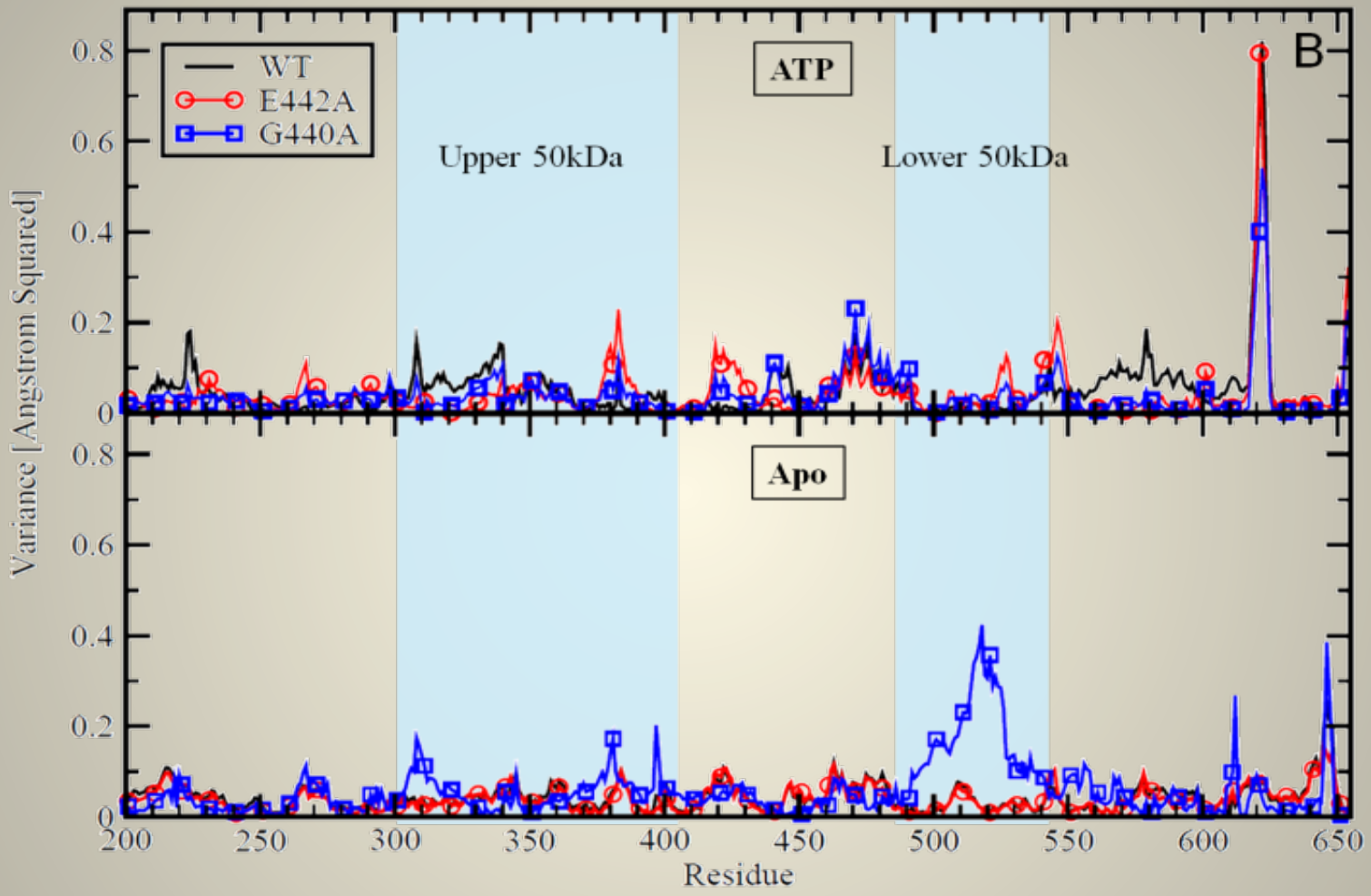
Results II

- The G440A mutation alters ABR dynamics
  - In the rigor state, the G440A mutant increases dynamics of a region of the U50 domain:
    - The cardiomyopathy loop
    - The C-terminus of the HO-helix
  - There is a dramatic increase in the mobility of helix-loop-helix region of the L50 domain

- The G440A mutation makes F441 highly dynamic in the ATP state
  - Disruption of Switch II rotation by G440A alters the mobility of F441
  - This hinders F441 interaction with the surrounding hydrophobic environment

- **These two alterations result in disruption of the communication pathways between the active site and the U50 and L50 regions**

The first PC mode of the ABR of mutant and wild-type myosin V



Switch II Mutants Reveal Coupling Between the Nucleotide- and Actin-Binding Regions in Myosin V
Darshan V. Trivedi, Charles C. David, Donald J. Jacobs, and Christopher M. Yengo,
Biophysical Journal Volume 102 Issue 11 pp.2545-2555. doi:10.1016/j.bpj.2012.04.025

Summary of the Work

CONCLUSION

Summary

- We have established the GSM as a viable alternative and/or co-model to either an ENM or MD
- The GSM is both efficient and effective at determining the native state dynamics of a protein
  - The GSM is **qualitatively** and **quantitatively** similar to MD and ANM
- The GSM is scalable
  - Applied to a wide range of proteins with success including small, single domain proteins and large multi-domain, multi-state proteins like MV
- ED of GSM trajectories is an effective method to tease out the biological motions of a subset of residues in a protein

Thank You!

- Mentors, Colleagues, Friends, & Family
- UNCC for my financial aid
- NIH for supporting my research