UNIVERSITY *of* DELAWARE

Who is afraid of I/O? Exploring I/O Challenges and Opportunities at the Exascale

Michela Taufer
Computer and Information Sciences
University of Delaware
Newark, Delaware, USA

Acknowledgements


Travis J.


Boyu Z.


Trilce E.


Adam L.


Silvia C.

Sponsors:

XSEDE Extreme Science and Engineering Discovery Environment

NSF


Dong A.


Don L.


Jim G.


Marc S.

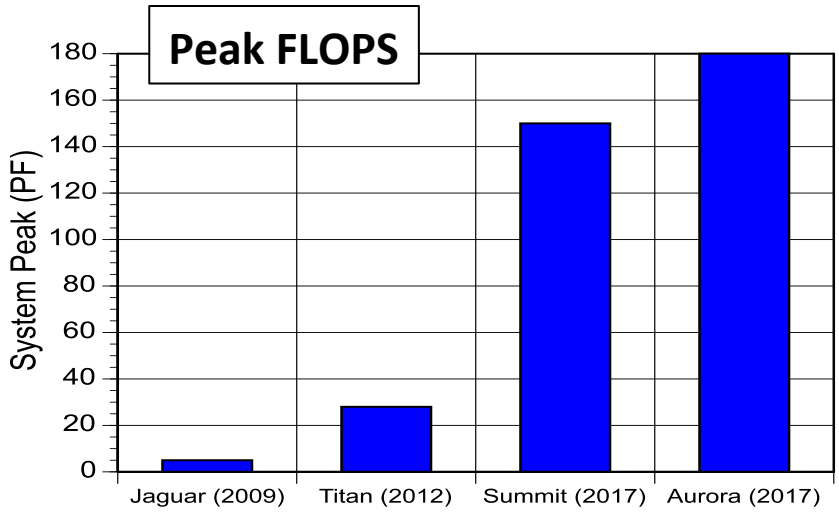

Tom S.

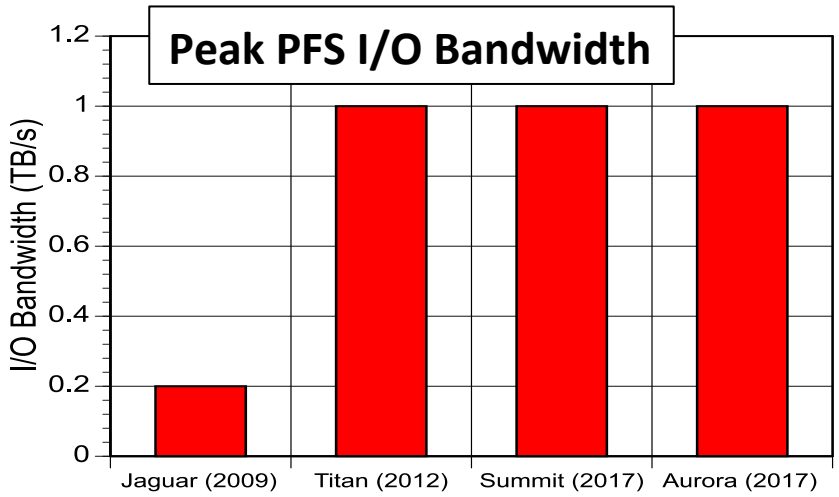

Becky S.

… and Mark G.

The GCLab@UD

Challenges at the Extreme Scale





Simulations today:

- Save all the data to analyze later!

Simulations at exascale:

- Analyze data as they are generated
- Save only what is really needed!

We must change how we run our simulaitons at the exascale

From a talk of Lucy Nowell, DoE Program Director (DoE Workflow Workshop, Rockville, MD, USA. April 20-21, 2015)

Perspective

The scientist:

"Storage technologies are advancing […] and it is really not clear at all [to me] that especially distributed storage platforms would not be able to handle […] petabyte data sets"

Anonymous Feedback

The computer architect:

"[…] there will be burst buffers on the DOE machines which will give applications much faster I/O […]"

Anonymous Feedback

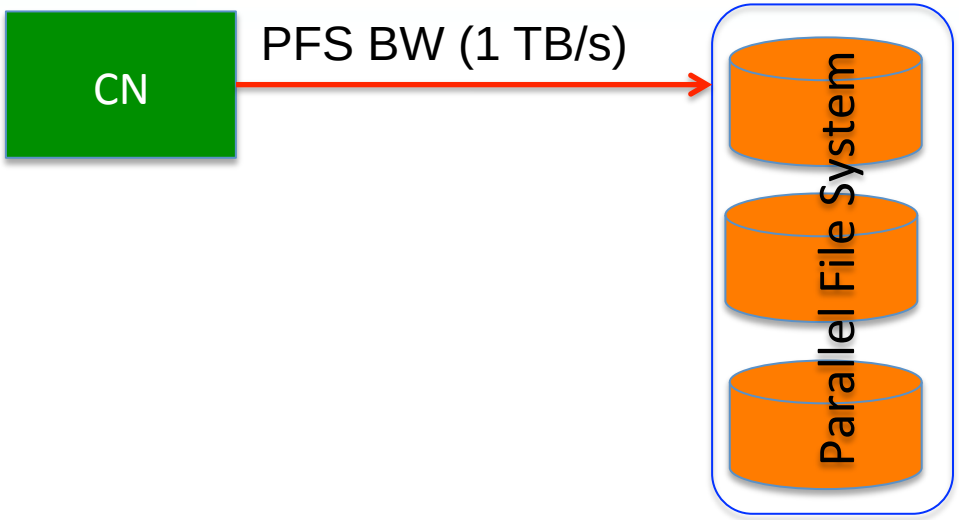
Burst Buffers

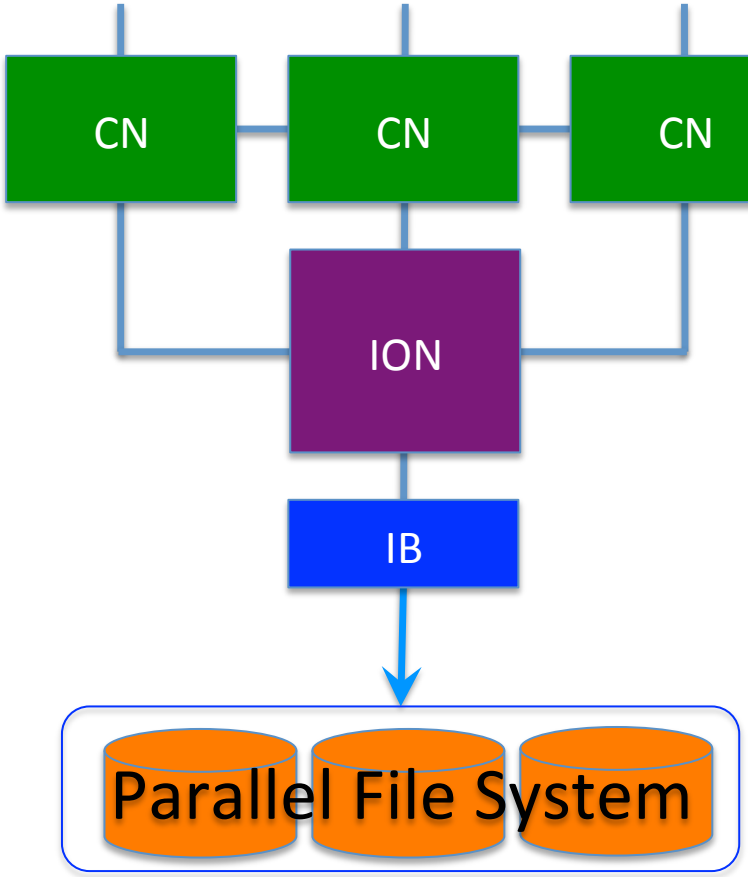
Many have heard about it,

few have seen real machines with it,

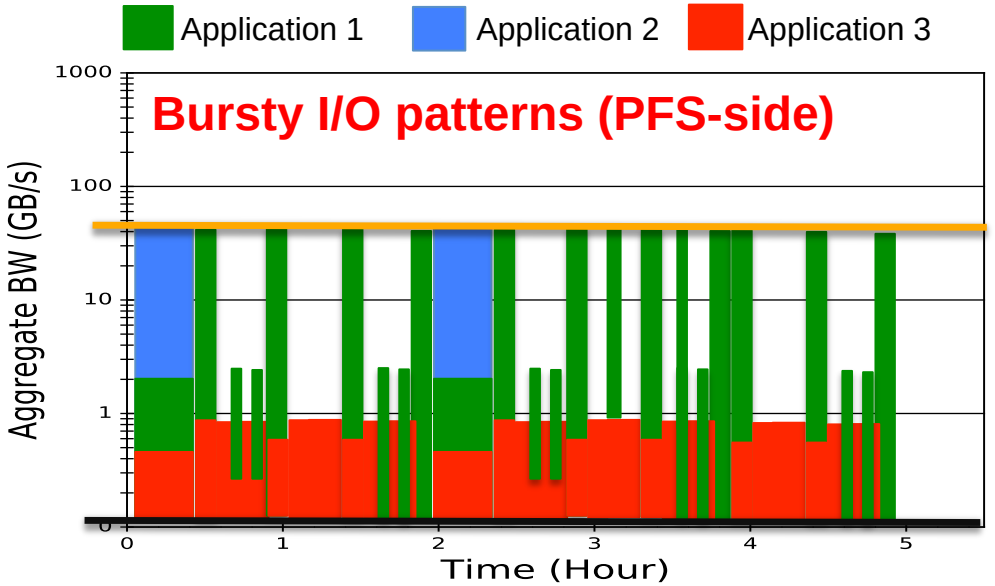
even fewer have ran applications on those machines …

Traditional System



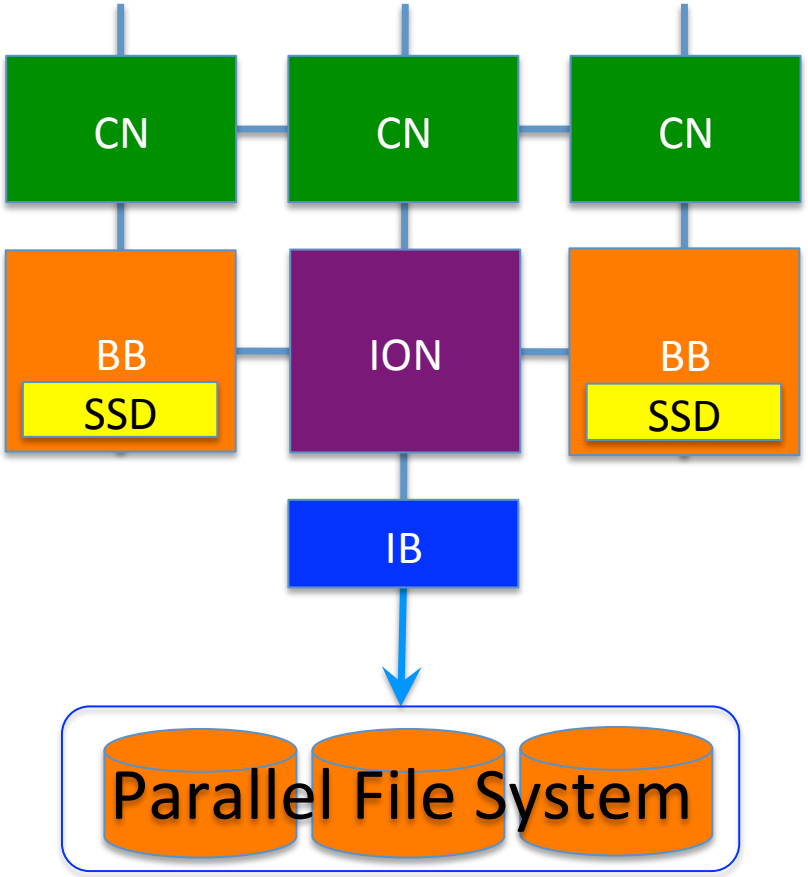


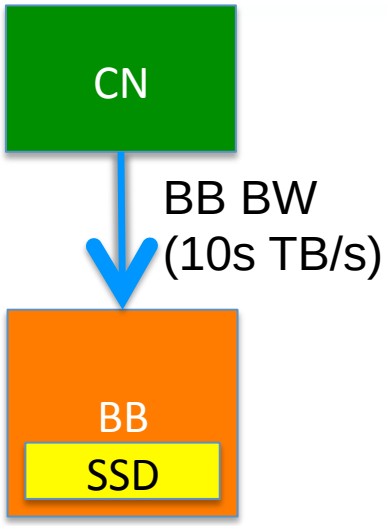
Based on: *http://www.nersc.gov/users/computational-systems/cori/burst-buffer/burst-buffer/*

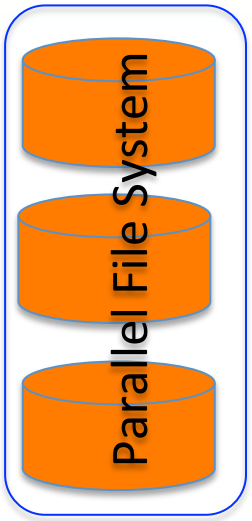


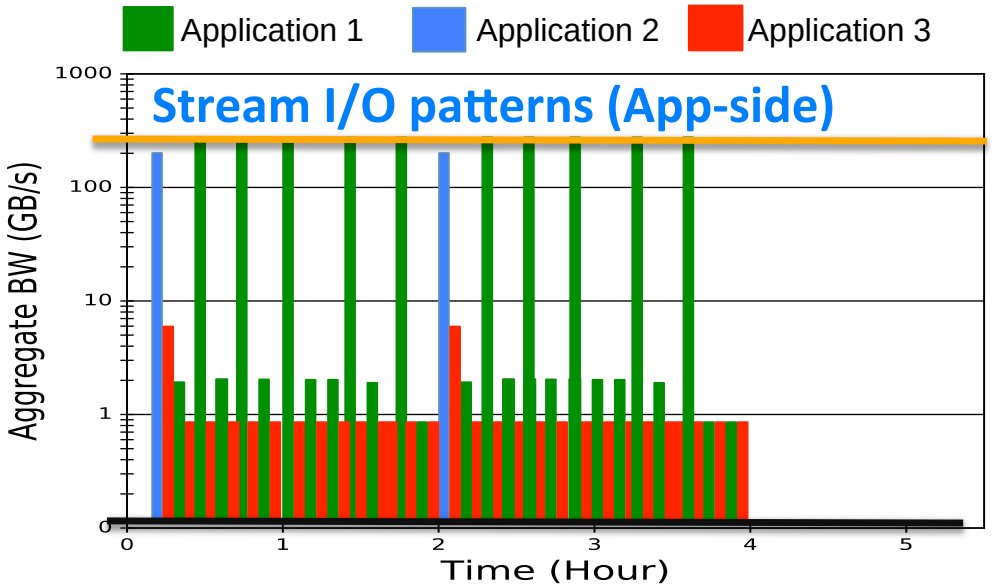
Based on: Liu, N, Cope, J, Carns, P, Carothers, C, Ross, R, Grider, G, Crume, A, Maltzahn, C . "On the Role of Burst Buffers in Leadership-class Storage Systems" MSST/SNAPI 2012

Burst Buffer System





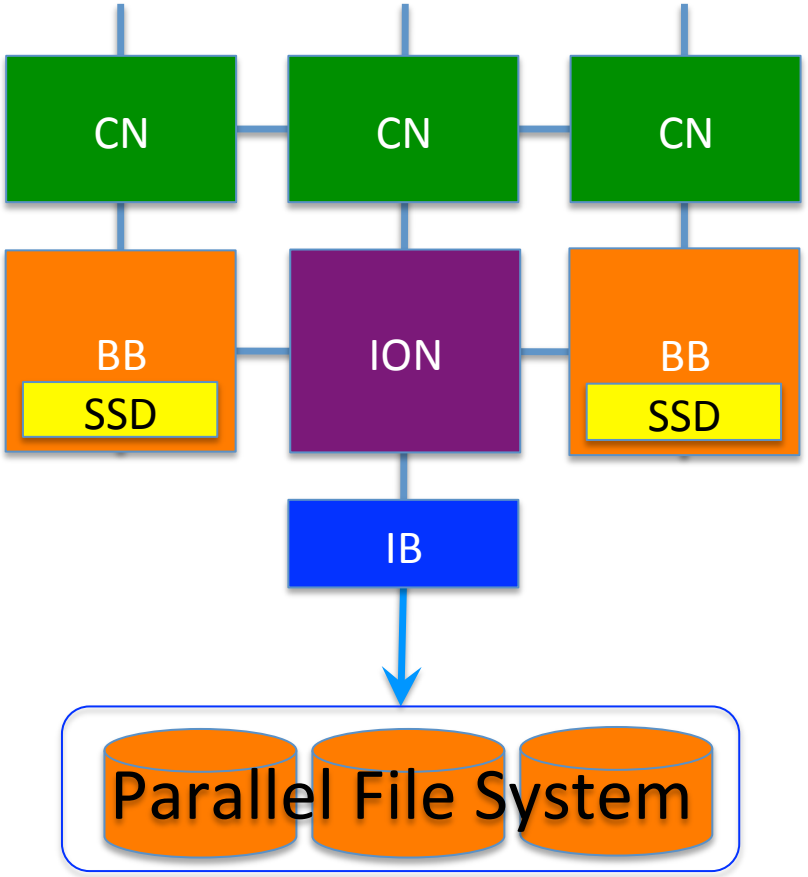




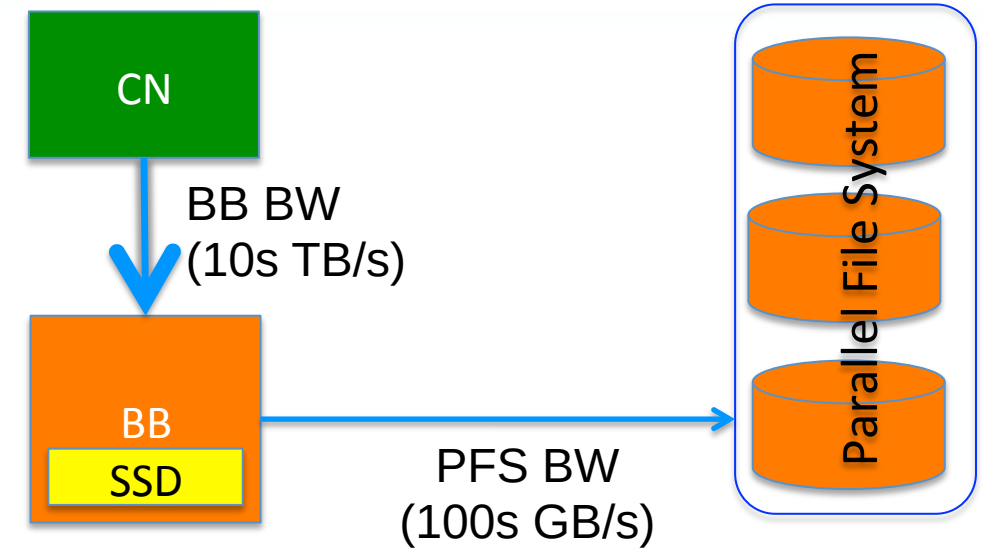
Based on: *http://www.nersc.gov/users/computational-systems/cori/burst-buffer/burst-buffer/*

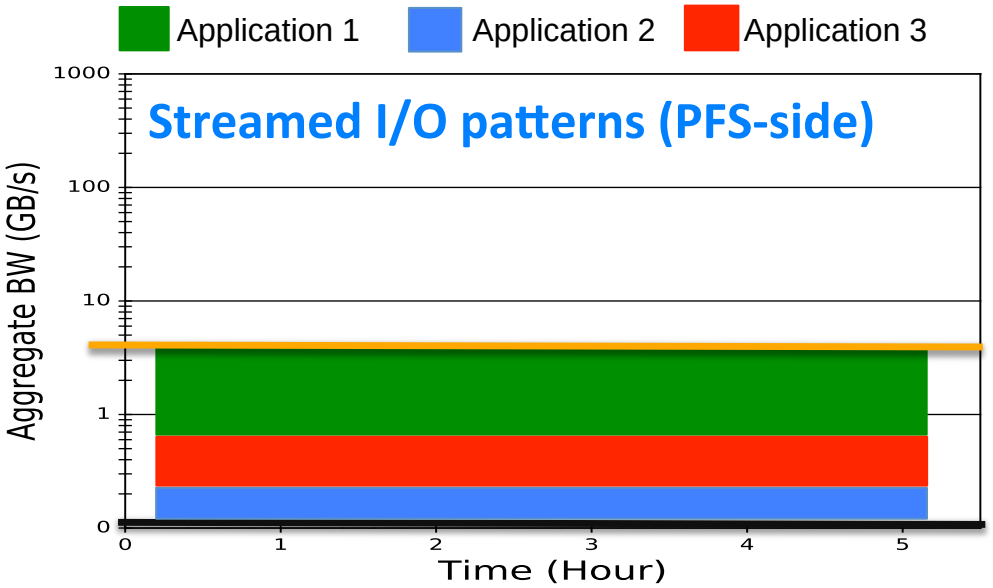
Based on: Liu, N, Cope, J, Carns, P, Carothers, C, Ross, R, Grider, G, Crume, A, Maltzahn, C . "On the Role of Burst Buffers in Leadership-class Storage Systems" MSST/SNAPI 2012

Burst Buffer System



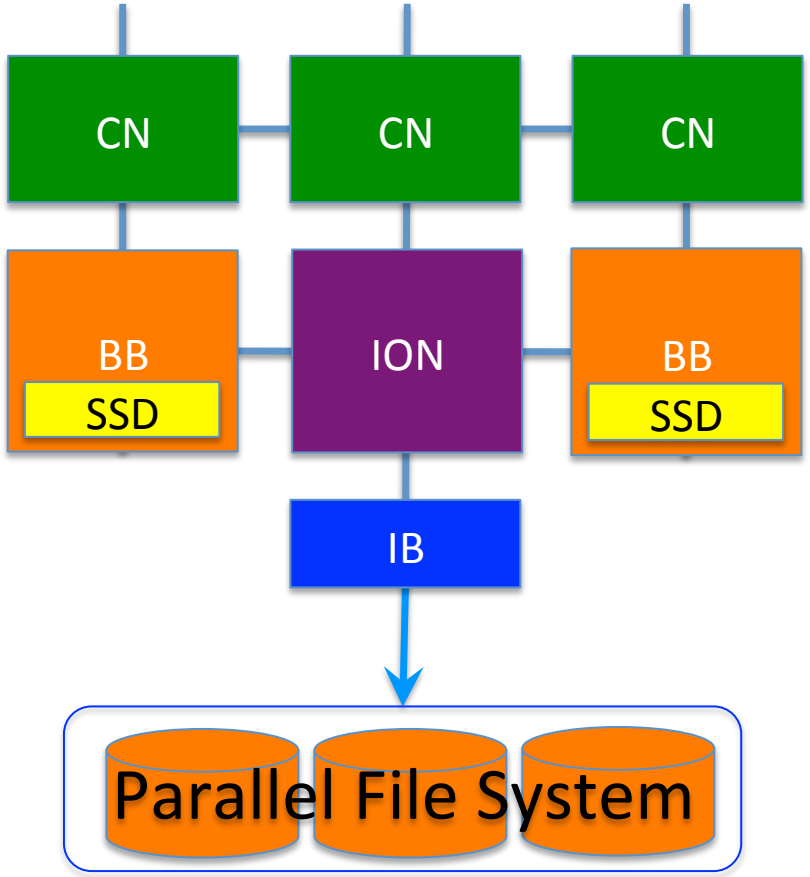
Based on: *http://www.nersc.gov/users/computational-systems/cori/burst-buffer/burst-buffer/*

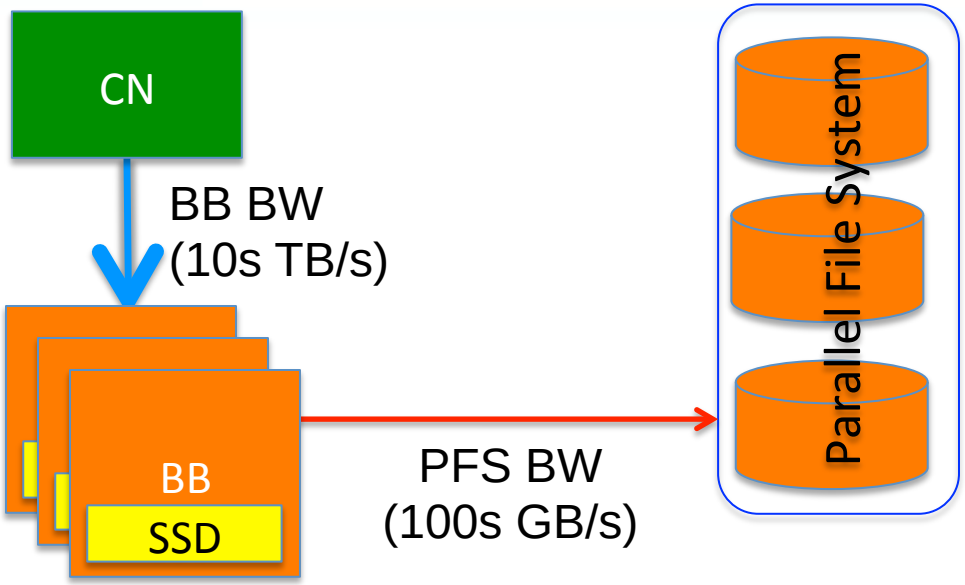




Based on: Liu, N, Cope, J, Carns, P, Carothers, C, Ross, R, Grider, G, Crume, A, Maltzahn, C . "On the Role of Burst Buffers in Leadership-class Storage Systems" MSST/SNAPI 2012

PFS Bottleneck





$$\sum \text{BB BW} > \text{PFS BW}$$

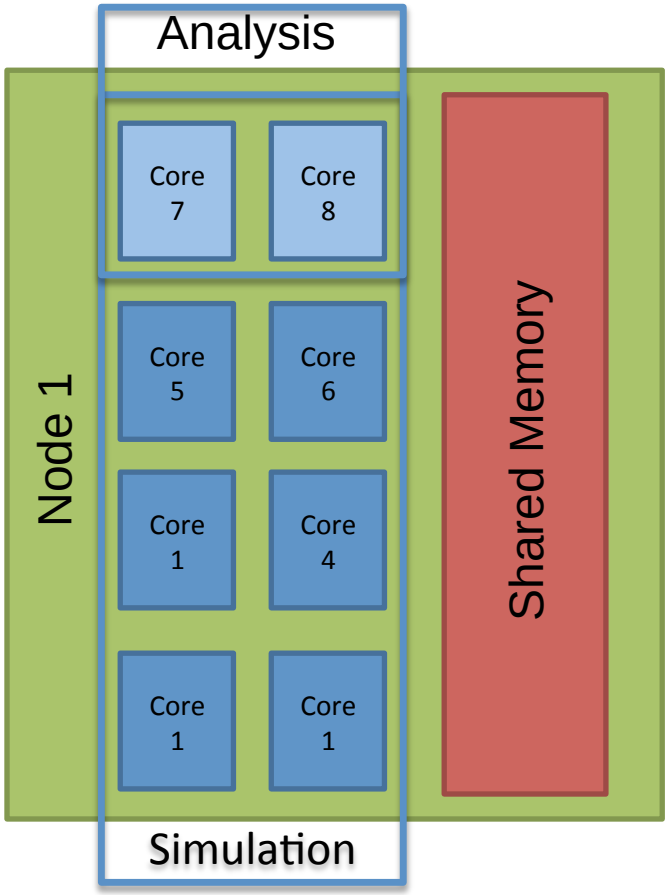
Applications' cumulative average bandwidths can exceed the PFS bandwidth, causing I/O contention

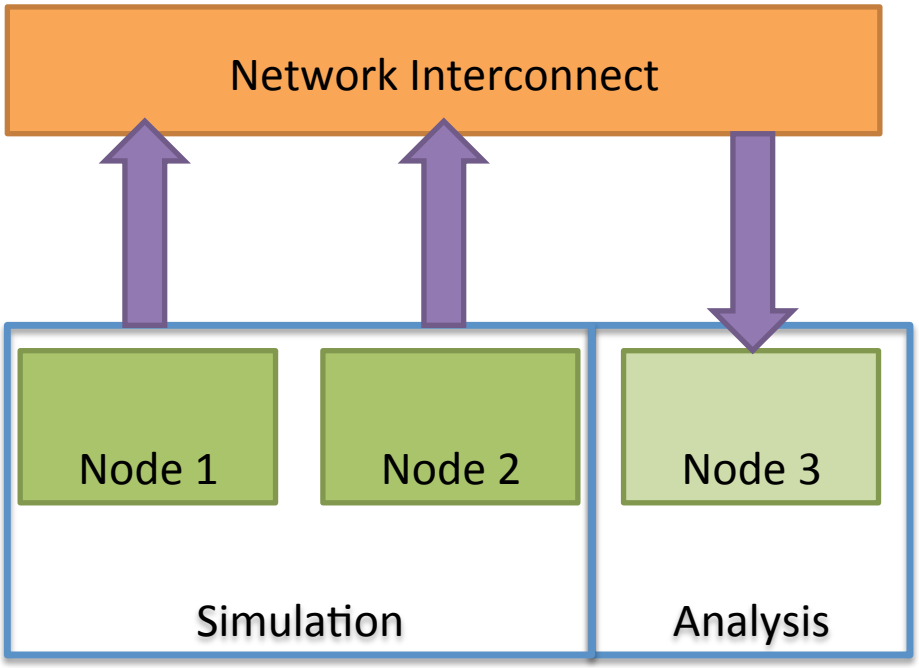
Based on: *http://www.nersc.gov/users/computational-systems/cori/burst-buffer/burst-buffer/*

Challenges

- Burst Buffers are not the magic I/O silver bullet
  - I/O contention still a problem if we exceed the burst buffer capability
  - Burst buffers improve offloading bandwidth but do **NOT** help uploading data from storage for analysis and visualization

In-situ and *In-transit* Analysis

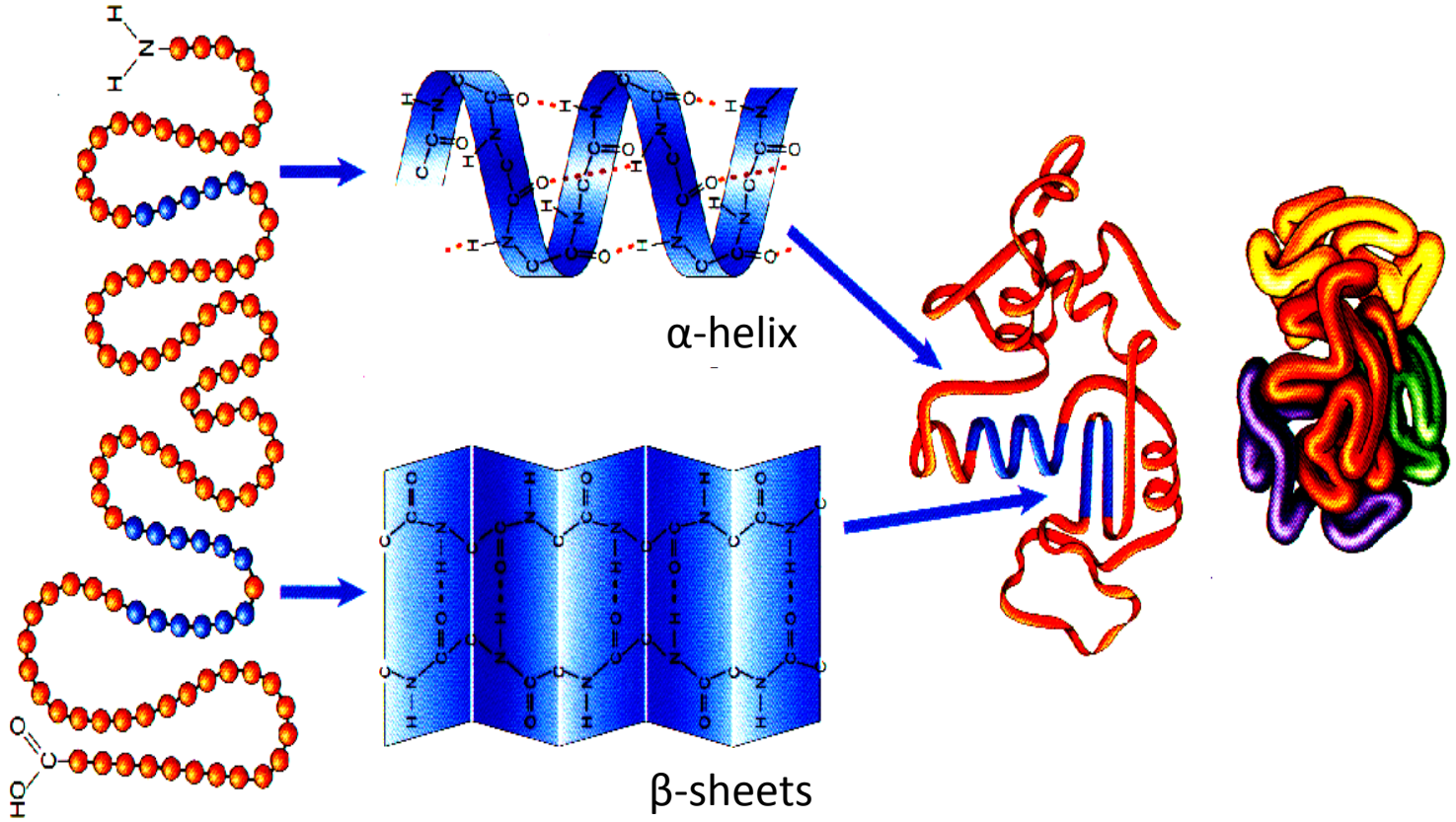




Example of tools:

- DataSpaces (Rutgers U.)
- DataStager (GeorgiaTech)

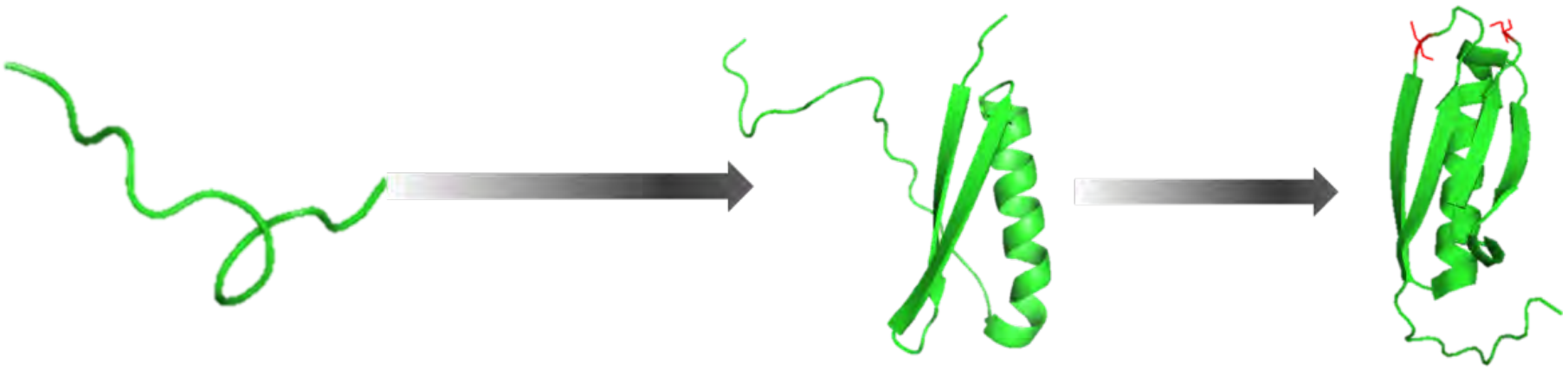
Protein Structures



From: http://schoolworkhelper.net/protein-structures-primary-secondary-tertiary-quaternary/

Molecular Dynamics (MD) Simulation

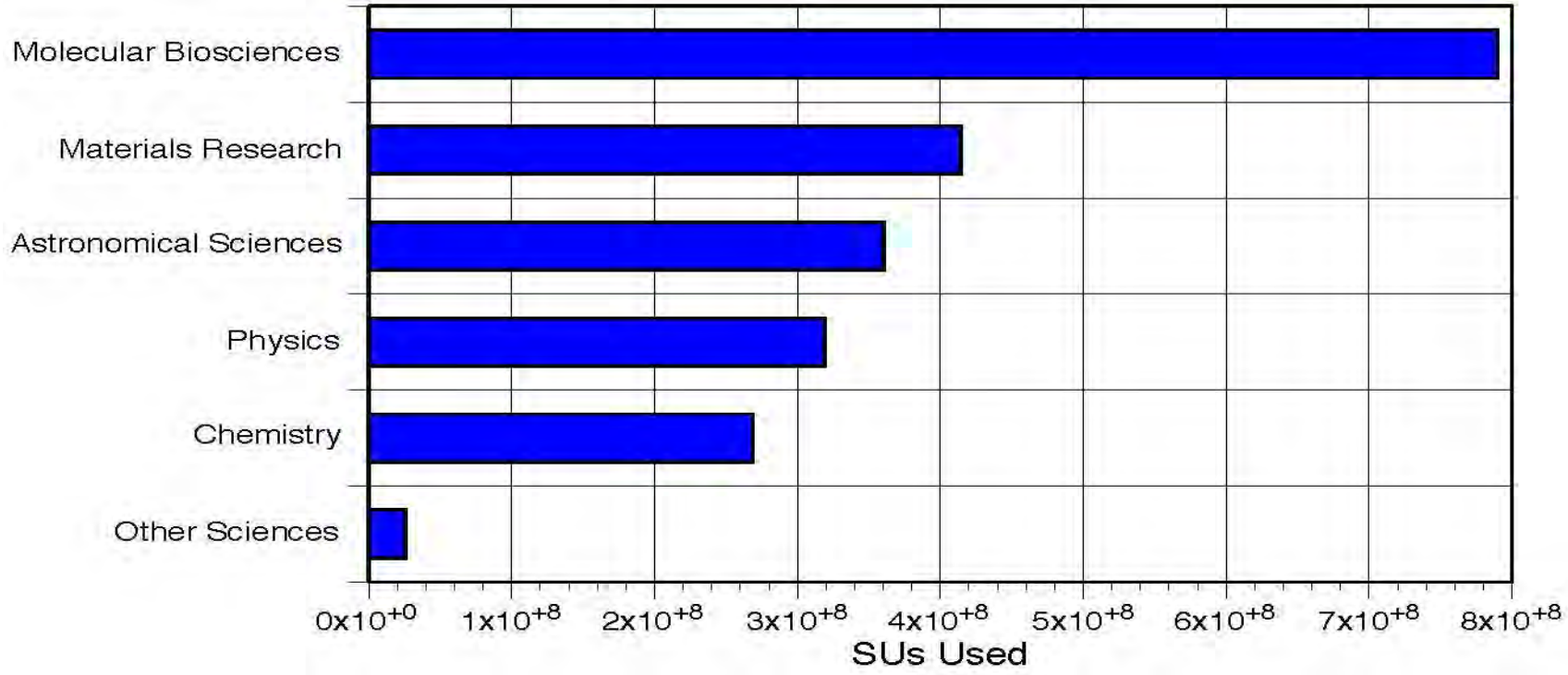
Use computer simulations to study the physical movements of atoms and molecules



From: https://www.weizmann.ac.il/chemphys/cfharan/research-activities/single-molecule-protein-dynamics-0

MD simulations are alive and kicking!

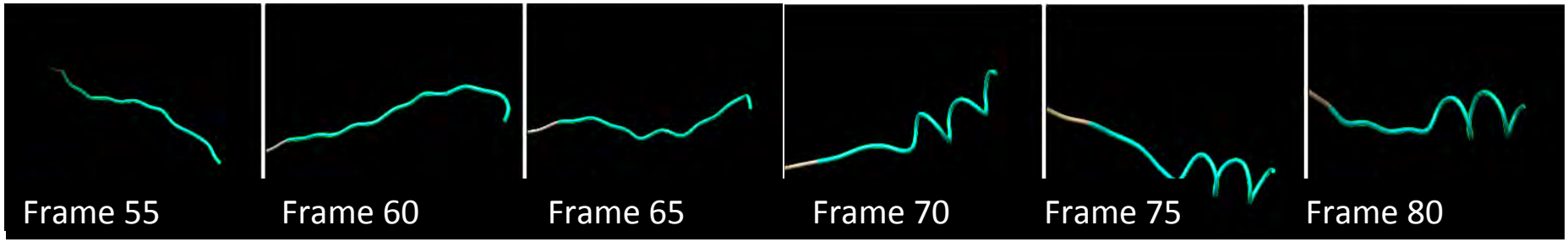
XSEDE SUs used by type of targeted science over the past 6 months (March 1, 2016 - August 31, 2016)



Four of the top 10 XSEDE users run molecular simulations (i.e., Schulten at UIUC, Feig at Michigan State U, Voth at U Chicago, and Case at Rutgers U)

Analysis Requirements

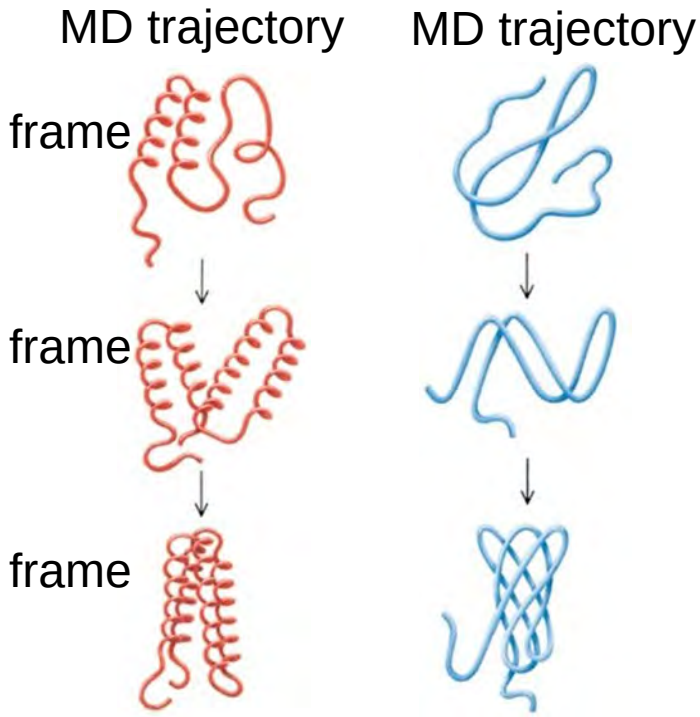
Frames of an MD trajectory:



- We want to capture what is going on in each frame **without**:
  - Disrupting the simulation (e.g., stealing CPU and memory on the node)
  - Moving all the frames to a central file system and analyzing them once the simulation is over
  - Comparing each frame with past frames of the same job
  - Comparing each frame with frames of other jobs

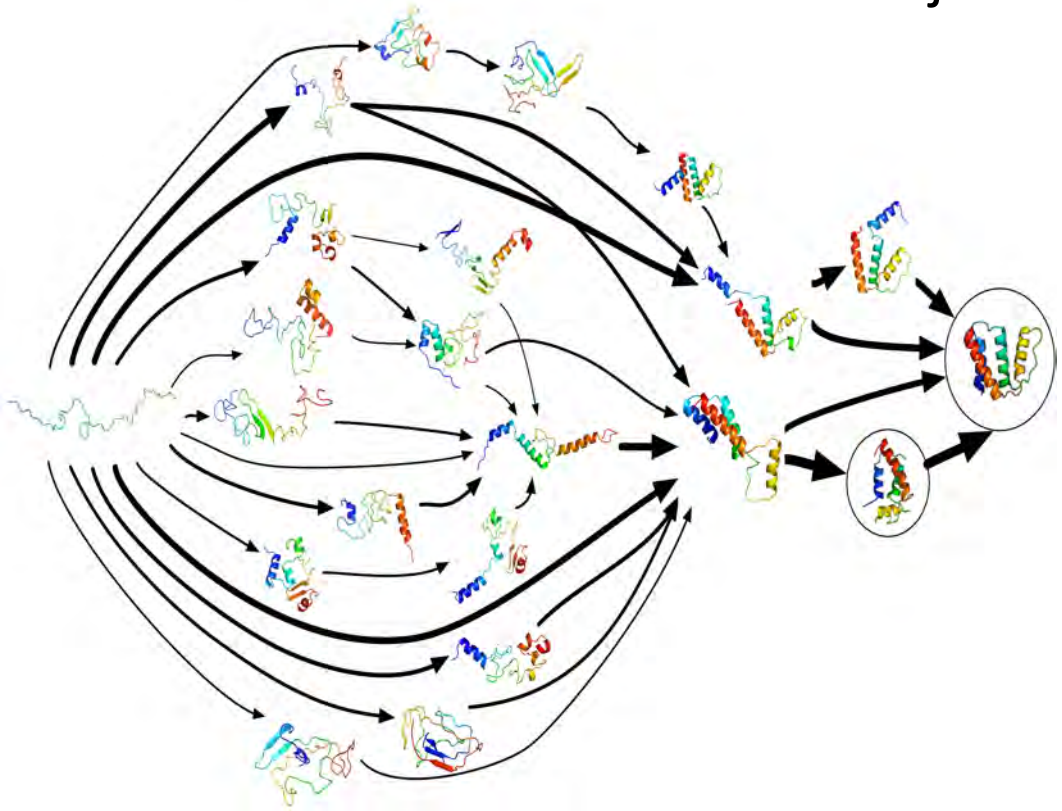
MD Simulations == Ensemble of Jobs

A MD job generates a sequence of conformational frames



From: http://images.slideplayer.com/18/5667960/slides/slide_1.jpg

A **MD simulation** comprises of hundreds of thousands of MD job



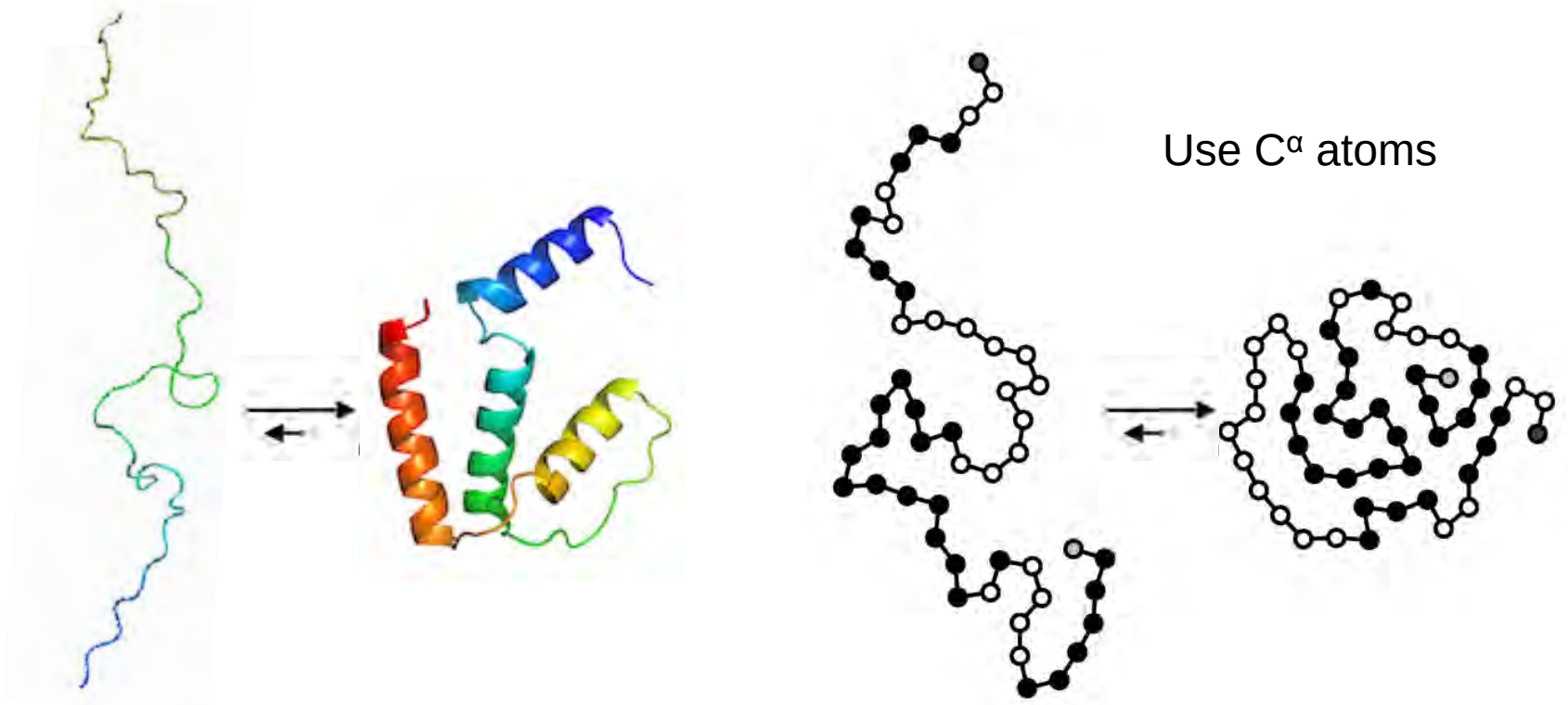
By Vincent Voelz - Sent to the uploader personally, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=218912120

MD Simulations == Ensemble of Black Boxes?

A MD job generates a sequence of conformational frames

A MD simulation comprises of hundreds of thousands MD jobs

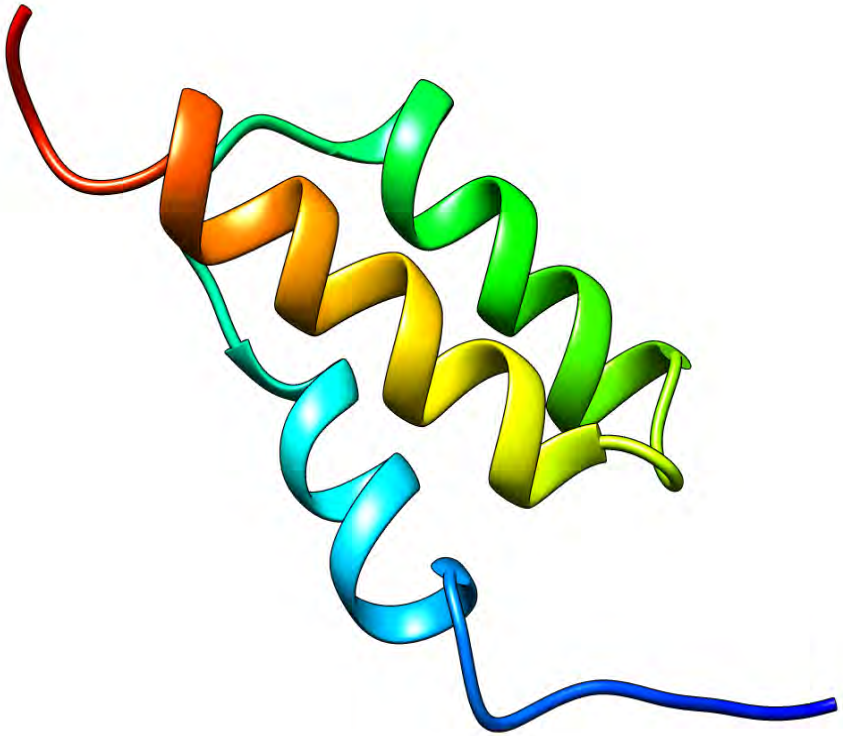
Modeling Molecules



From: By Tomixdf (talk) - Own work (Original text: self-made), Public Domain, https://commons.wikimedia.org/w/index.php?curid=23662306

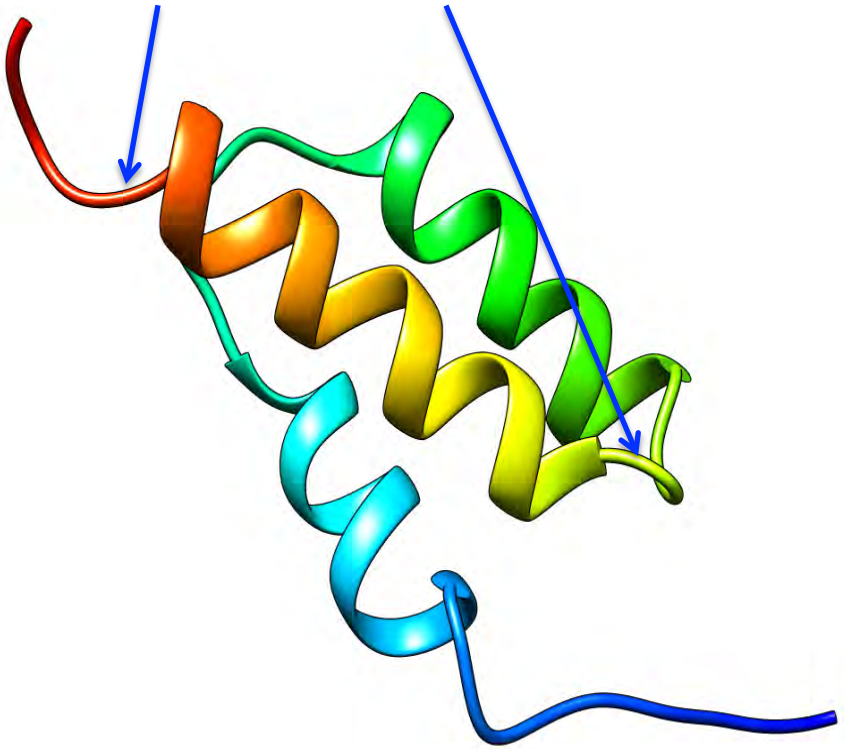
Capturing Secondary Structures

Given a frame of an MD job at time *t*



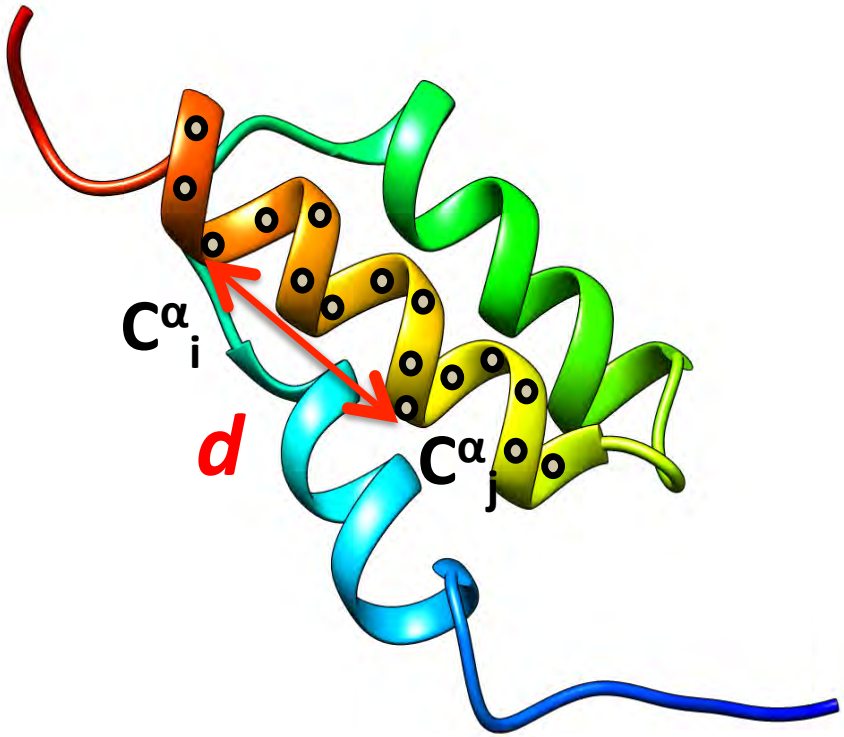
Capturing Secondary Structures

Define the substructure:
start and **stop** amino acids



Capturing Secondary Structures

Measure the distance between C^{α}_{j} and C^{α}_{i}



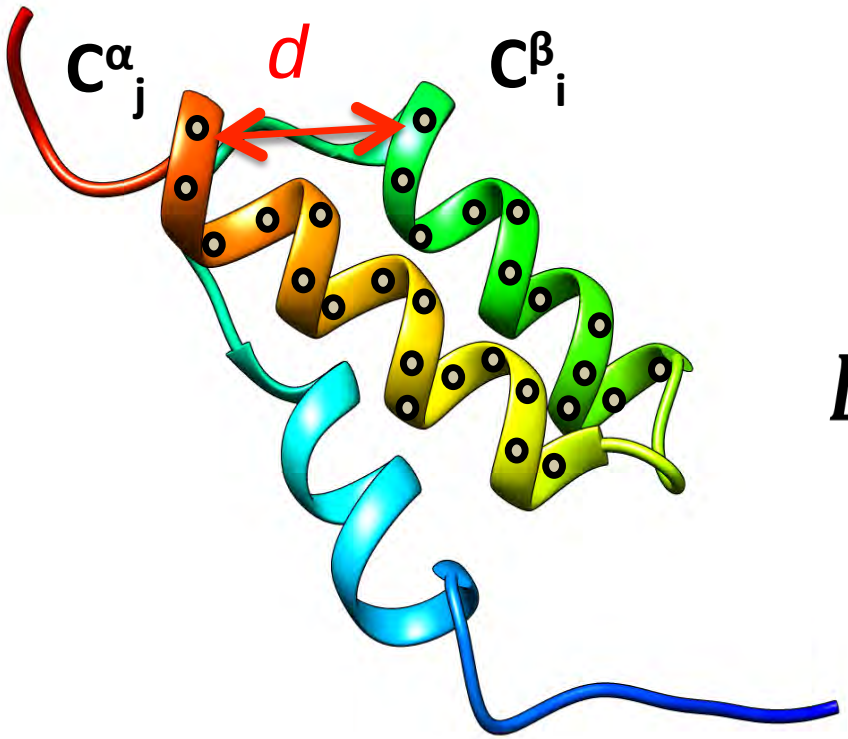
Build the **substructure Euclidean Distance Matrix (D)**

$$D = \begin{array}{c} \\ C^{\alpha}_{j} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} C^{\alpha}_{i} \\ \begin{bmatrix} 0 & x & x & x & x & x \\ x & 0 & d & x & x & x \\ x & d & 0 & x & x & x \\ x & x & x & 0 & x & x \\ x & x & x & x & 0 & x \\ x & x & x & x & x & 0 \end{bmatrix} \end{array}$$

Compute largest eigenvalue → λ_{max}

Capturing Tertiary Structures

Measure the distance between C^{α}_{j} and C^{β}_{i}



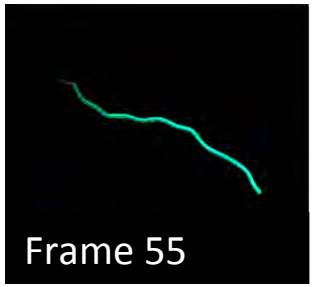
Build a **bipartite distance matrix** by comparing two substructures

$$D = \begin{array}{c} \\ j \\ \\ \\ \\ \\ \end{array} \begin{array}{c} i \\ \begin{bmatrix} 0 & 0 & 0 & X & X & X \\ 0 & 0 & 0 & d & X & X \\ 0 & 0 & 0 & X & X & X \\ X & d & X & 0 & 0 & 0 \\ X & X & X & 0 & 0 & 0 \\ X & X & X & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Compute largest eigenvalue → λ_{max}

Proxy for Conformations' Changes

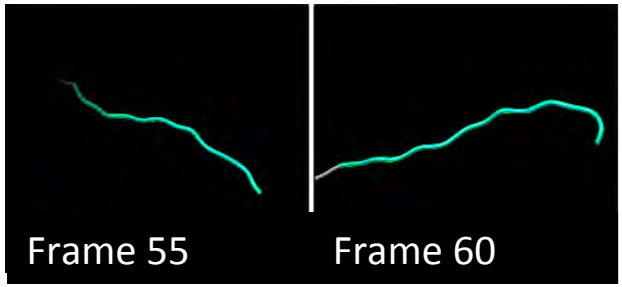
Frames of an MD job:



$\boldsymbol{\lambda_{55}}$

Proxy for Conformations' Changes

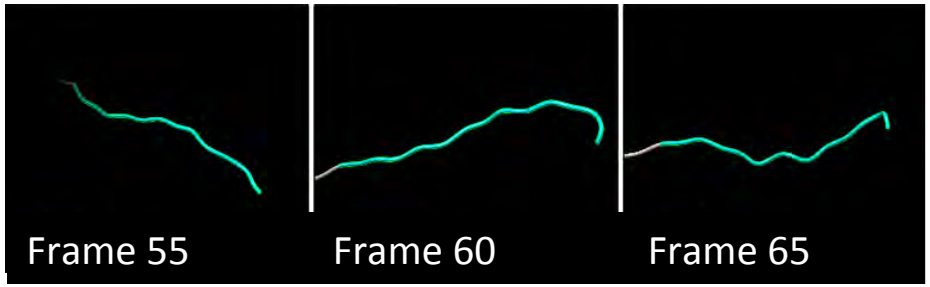
Frames of an MD job:



λ_{55} λ_{60}

Proxy for Conformations' Changes

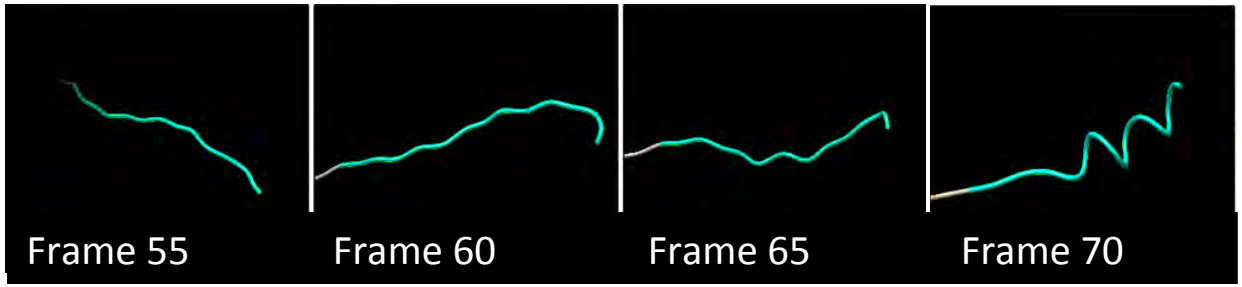
Frames of an MD job:



λ_{55} λ_{60} λ_{65}

Proxy for Conformations' Changes

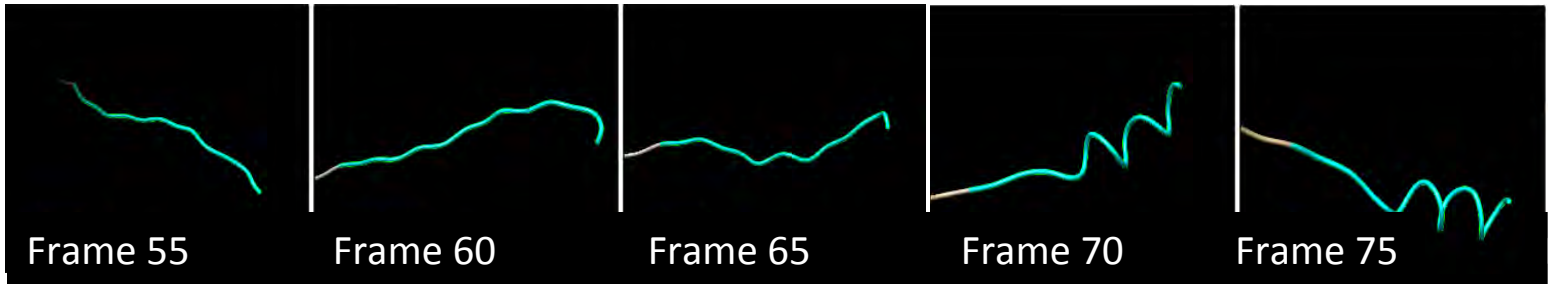
Frames of an MD job:



λ_{55} λ_{60} λ_{65} λ_{70}

Proxy for Conformations' Changes

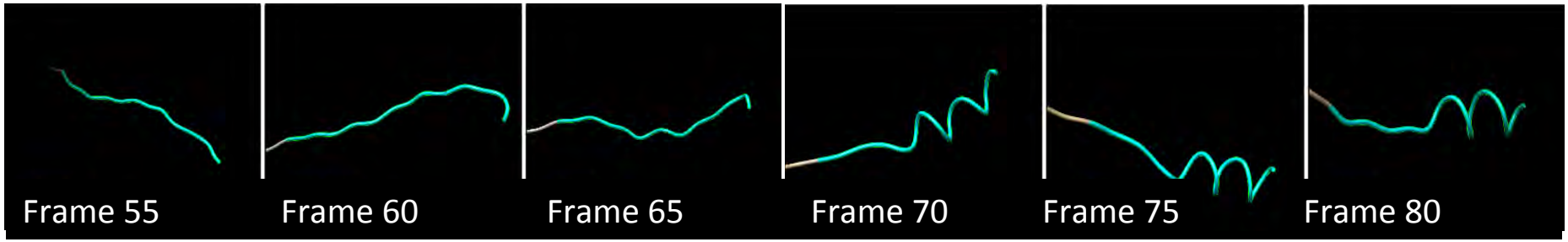
Frames of an MD job:



λ_{55} λ_{60} λ_{65} λ_{70} λ_{75}

Proxy for Conformations' Changes

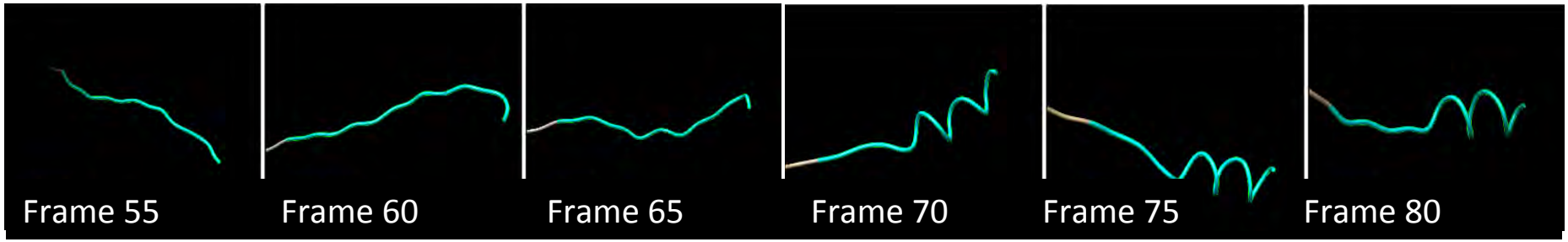
Frames of an MD job:



λ_{55} λ_{60} λ_{65} λ_{70} λ_{75} λ_{85}

Proxy for Conformations' Changes

Frames of an MD job:



λ_{55} λ_{60} λ_{65} λ_{70} λ_{75} λ_{85}

Distance between two max eigenvalues serves as a proxy for distance between the two associated conformations

Proxy for Conformations' Changes

Distance between two max eigenvalues serves as a proxy for distance between the two associated conformations

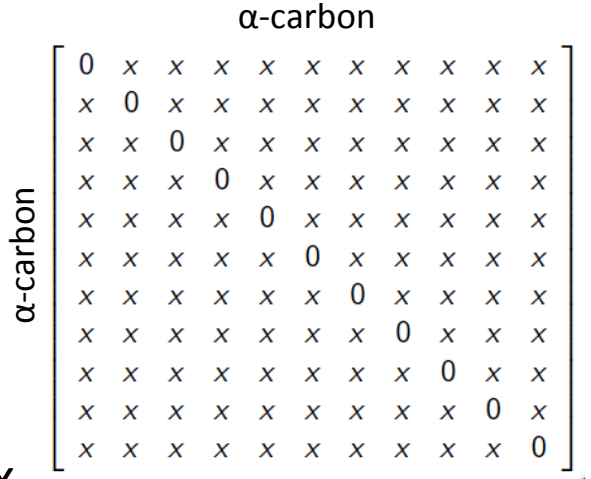
- Euclidean distance matrix D is symmetric
- Eigenvalues of symmetric, real matrices are stable
  - Small perturbations of D result in only small changes in the eigenvalues
  - Euclidean distance matrix is insensitive to rigid transformation
- Use only largest eigenvalue in distance matrix

$\lambda max = \lambda 1 < \lambda 2 < \lambda 3 < \lambda 4 < \lambda 5 = \lambda min$

$\lambda 1 + \lambda 2 + \lambda 3 + \lambda 4 + \lambda 5 = 0$

$\lambda 1 >> \lambda 2 \sim \lambda 3 \sim \lambda 4 \sim 0$

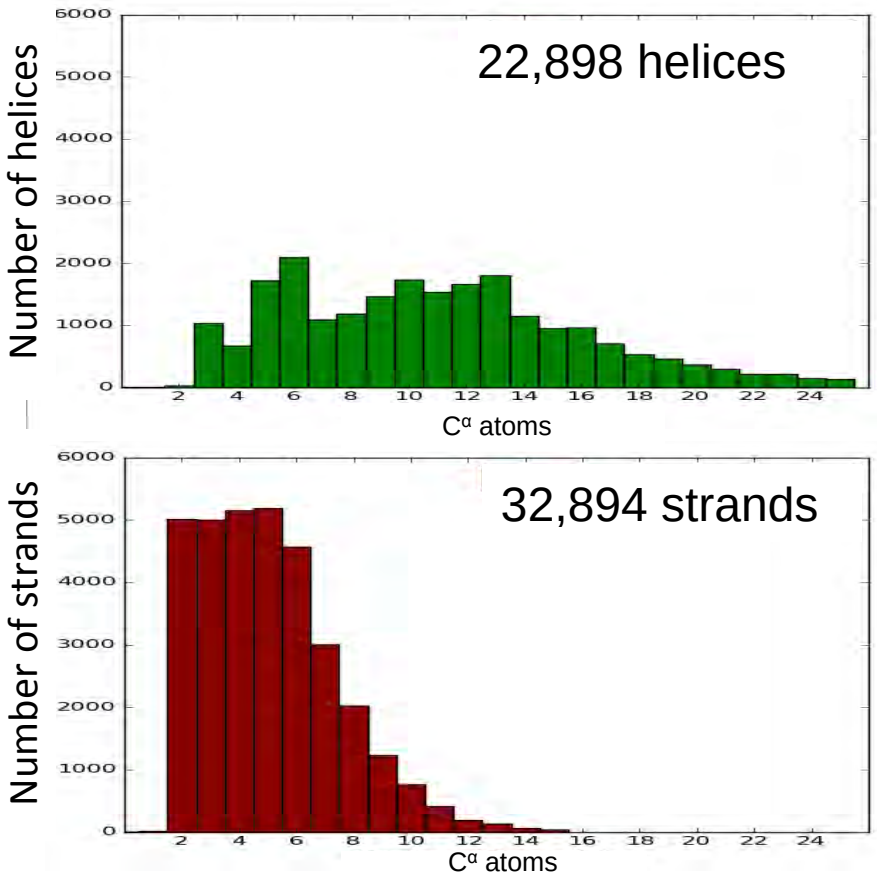
$\lambda max = \lambda 1 \sim -\lambda 5 = -\lambda min$

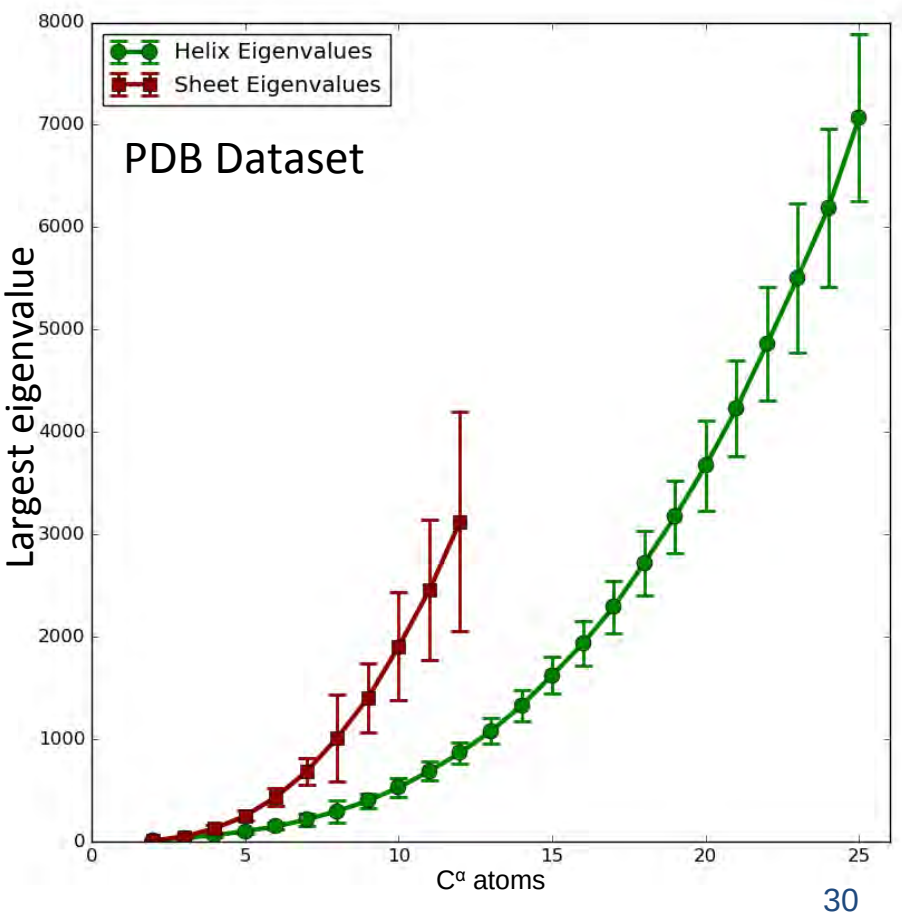


"***In-Situ Data Analysis and Indexing of Protein Trajectories***," Travis Johnston, Buyu Zhang, Adam Liwo, Silvia Crivelli, and Michela Taufer. JCC 2017.

Mapping Largest Eigenvalues to Structures

PDB dataset: 3,197 different proteins including 22,898 helices and 32,894 strands





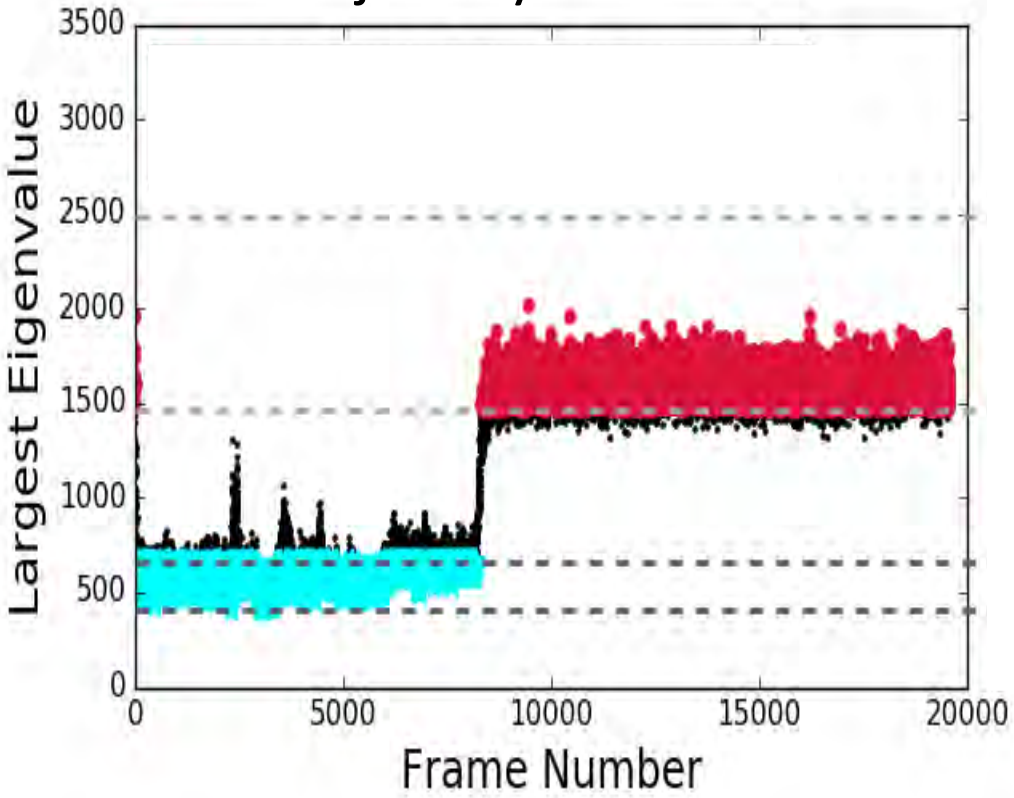
Case Study I: 2MQ8 Protein

- Canonical simulation of 2MQ8 protein including both α helices and β strands
  - After ~9M steps α helices pack tighter and change into β strands

Can the eigenvalue analysis capture the conformational change?

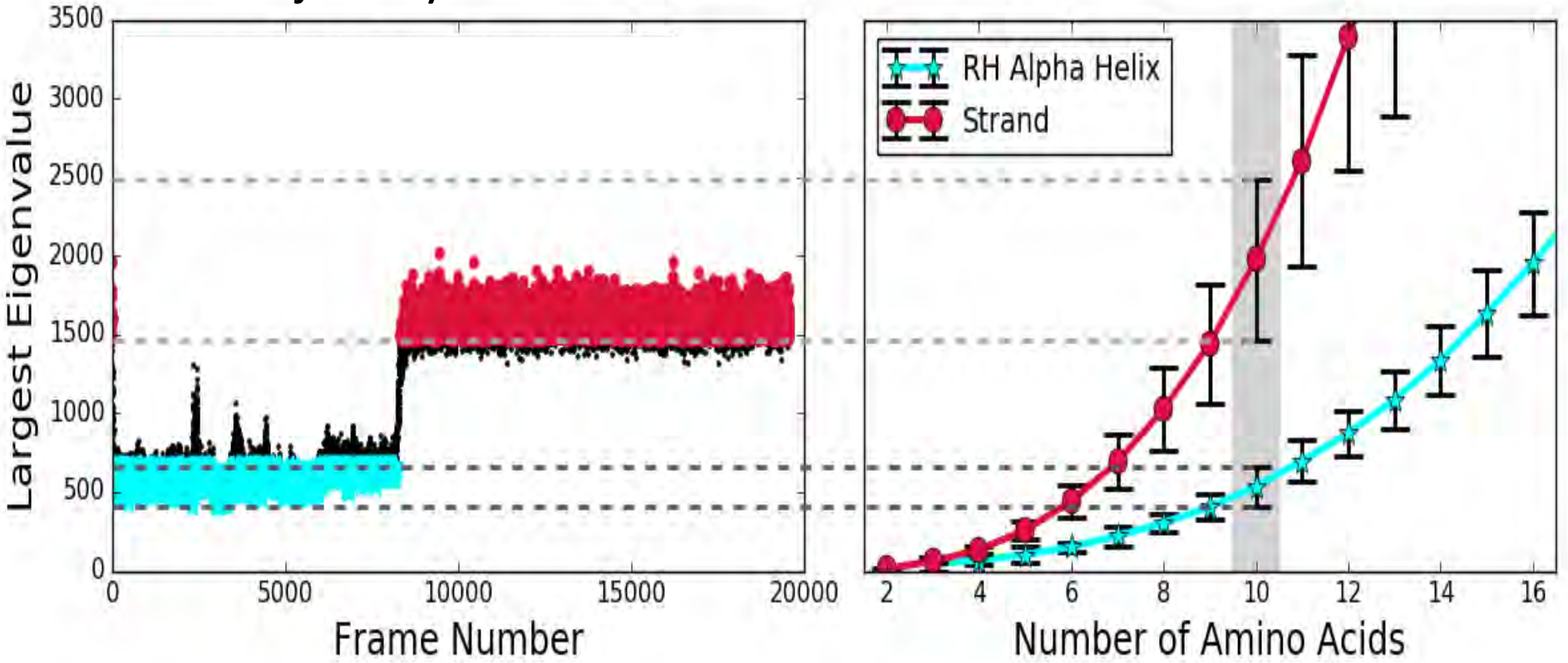
Case Study I: 2MQ8 Protein

Compute largest eigenvalue of 3rd strand (10 amino acids) for each trajectory frame



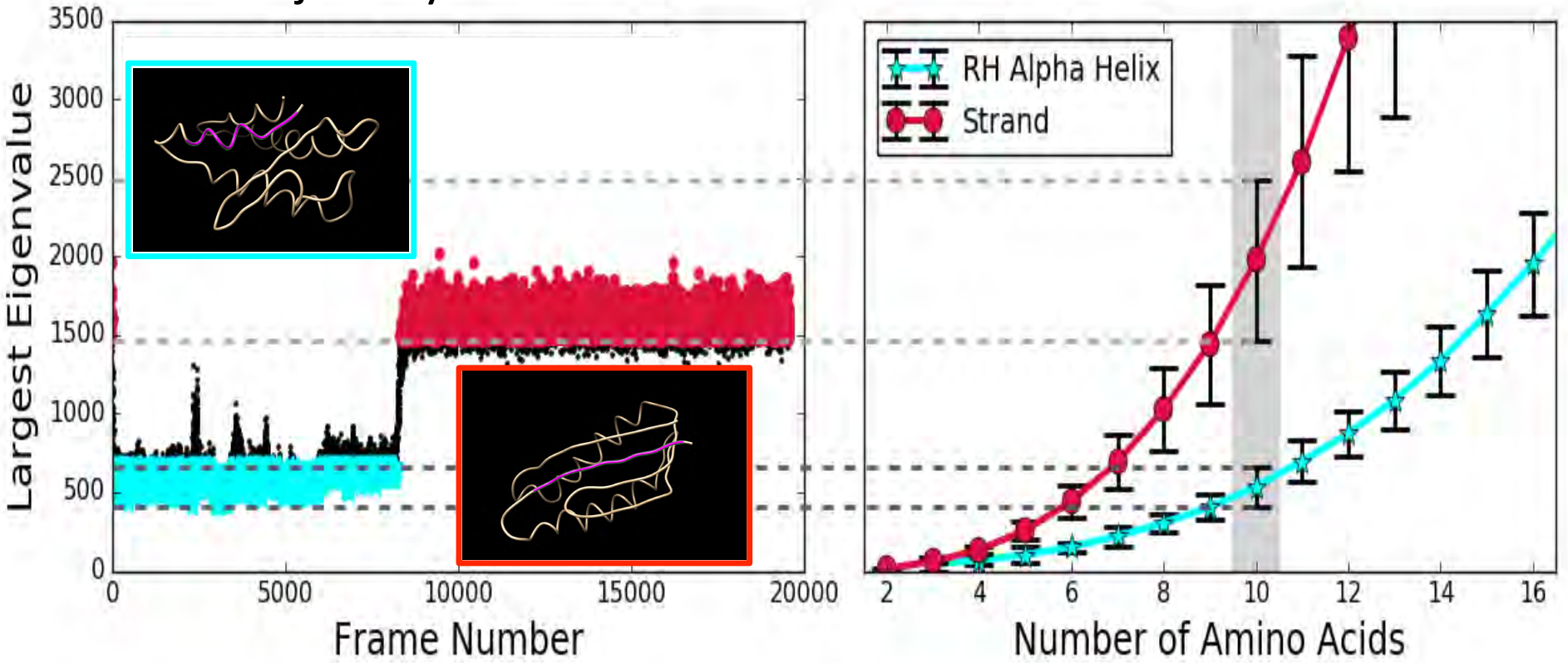
Case Study I: 2MQ8 Protein

Compute largest eigenvalue of 3rd strand (10 amino acids) for each trajectory frame



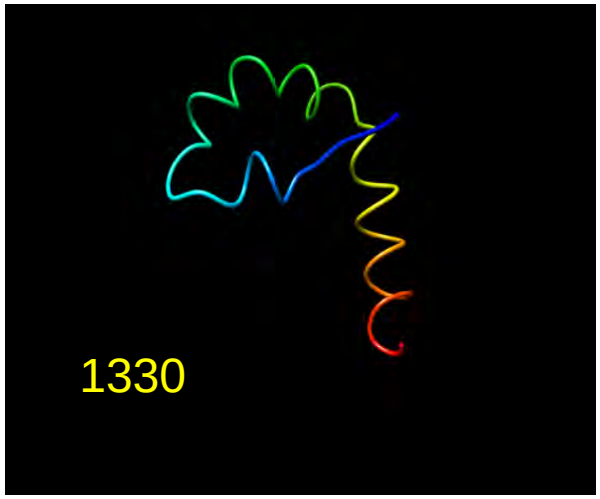
Case Study I: 2MQ8 Protein

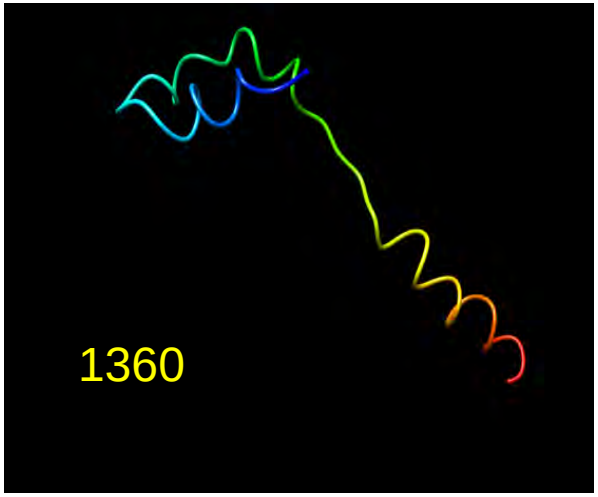
Compute largest eigenvalue of 3rd strand (10 amino acids) for each trajectory frame

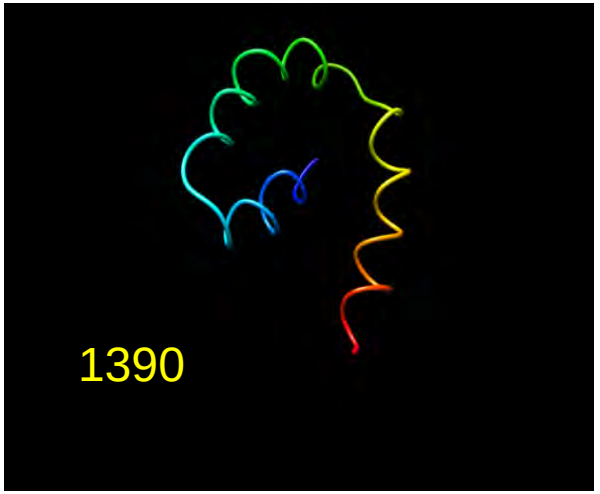


Case Study II: Capturing Movement of α-helices

Capture movement of structures with respect to each other

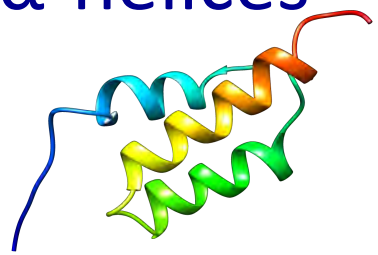




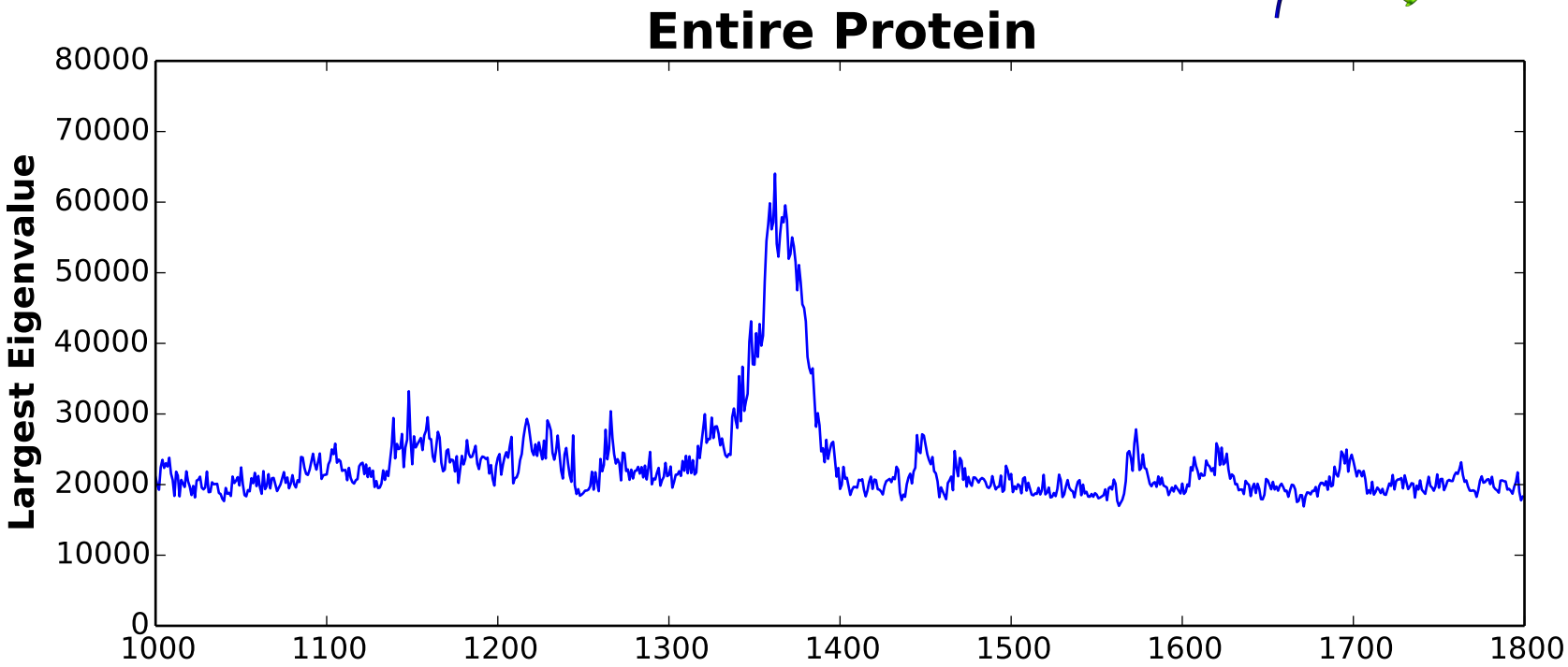


Can the eigenvalue analysis capture the movement of helices ?

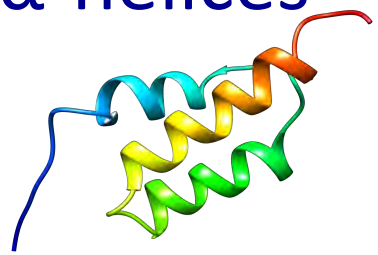
Case Study II: Capturing Movement of α-helices



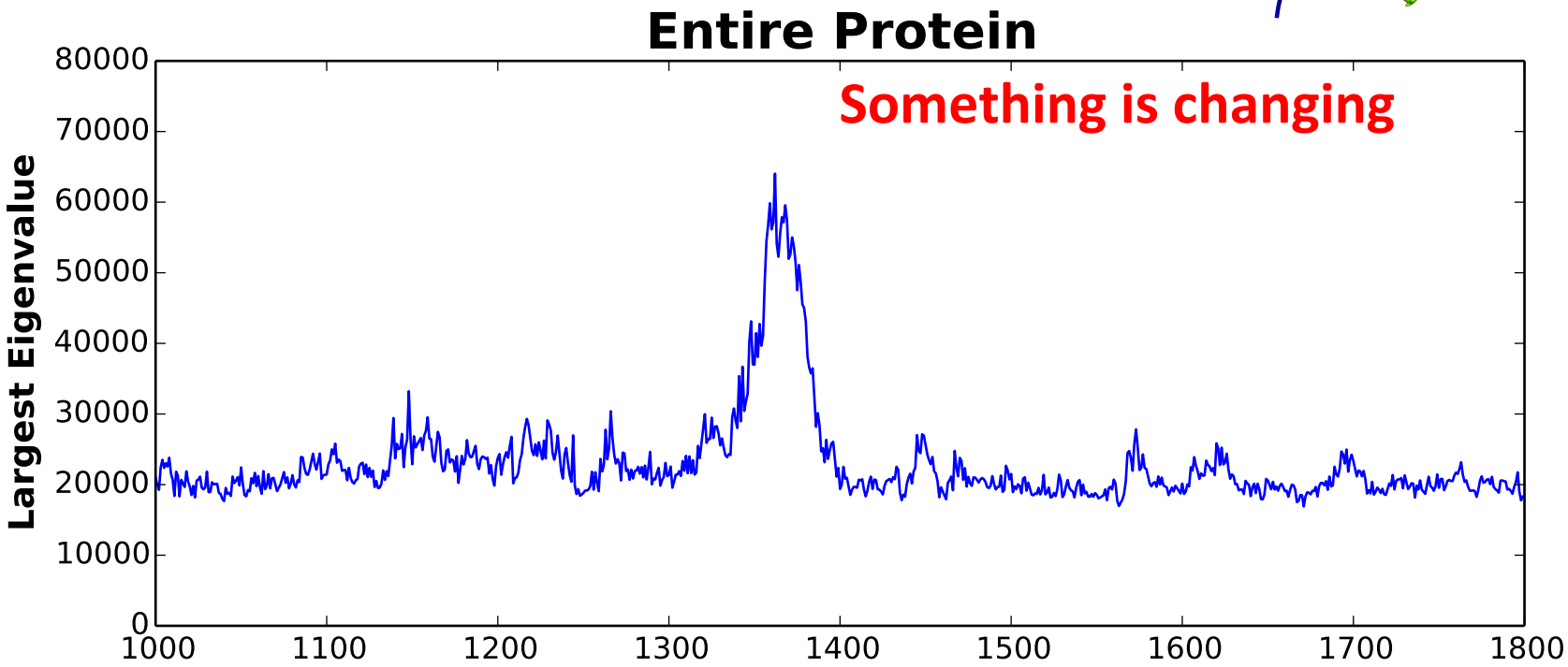
Monitor largest eigenvalue of entire protein



Case Study II: Capturing Movement of α-helices

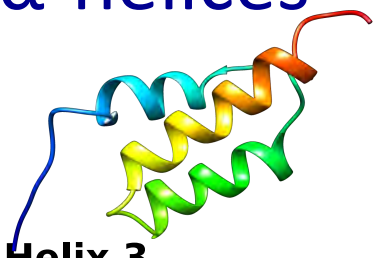


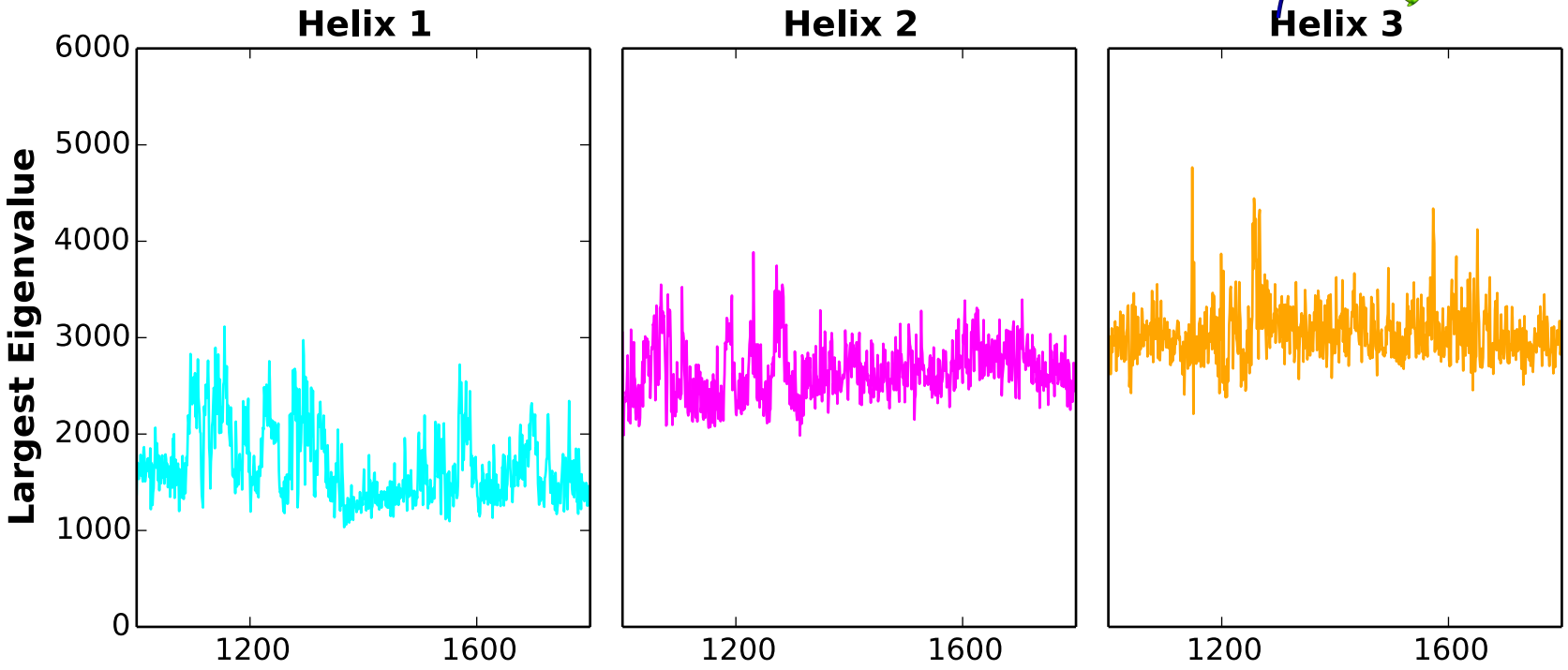
Monitor largest eigenvalue of entire protein



Case Study II: Capturing Movement of α-helices

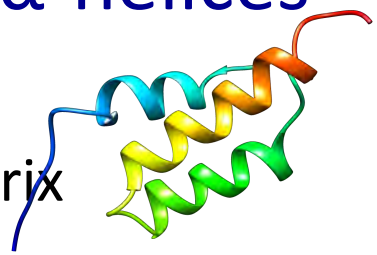
Monitor largest eigenvalue of single helices



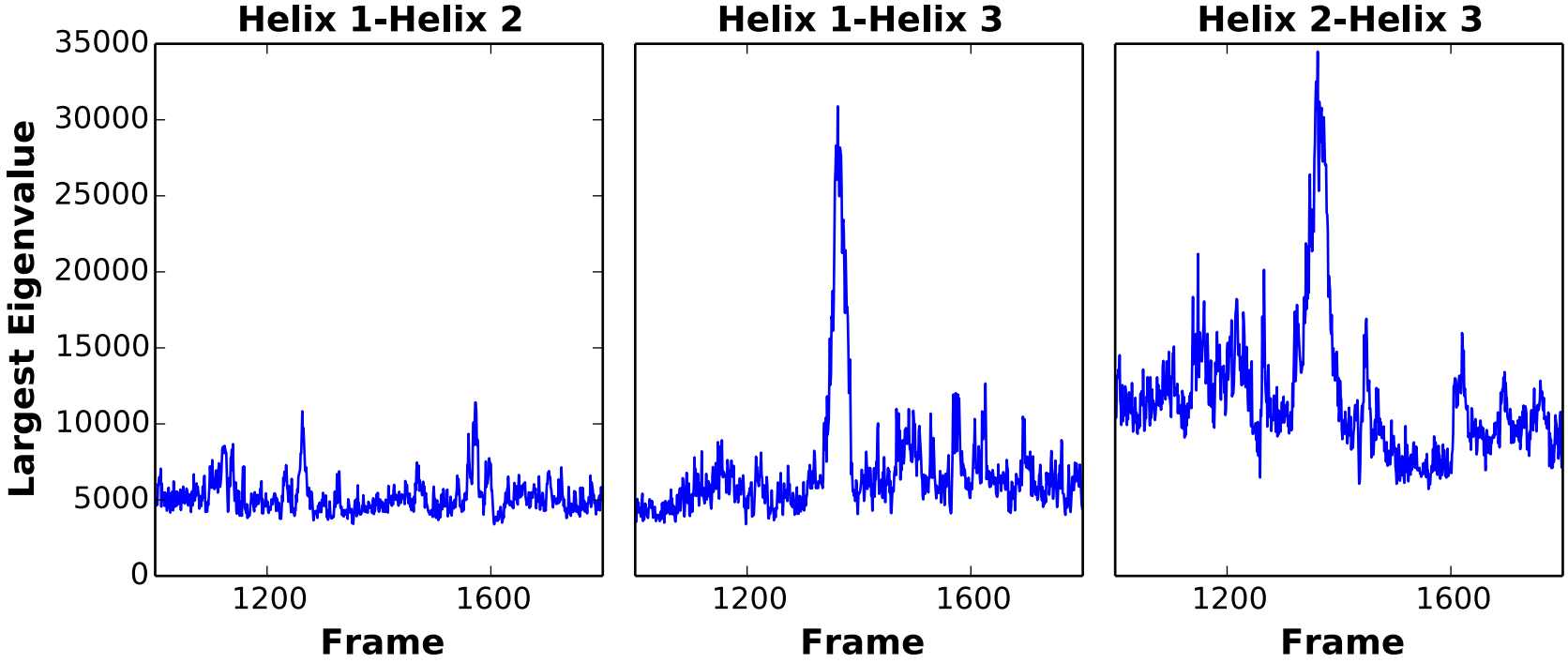


Individual α-helices (Helix 1, Helix 2, and Helix 3) appear stable

Case Study II: Capturing Movement of α-helices

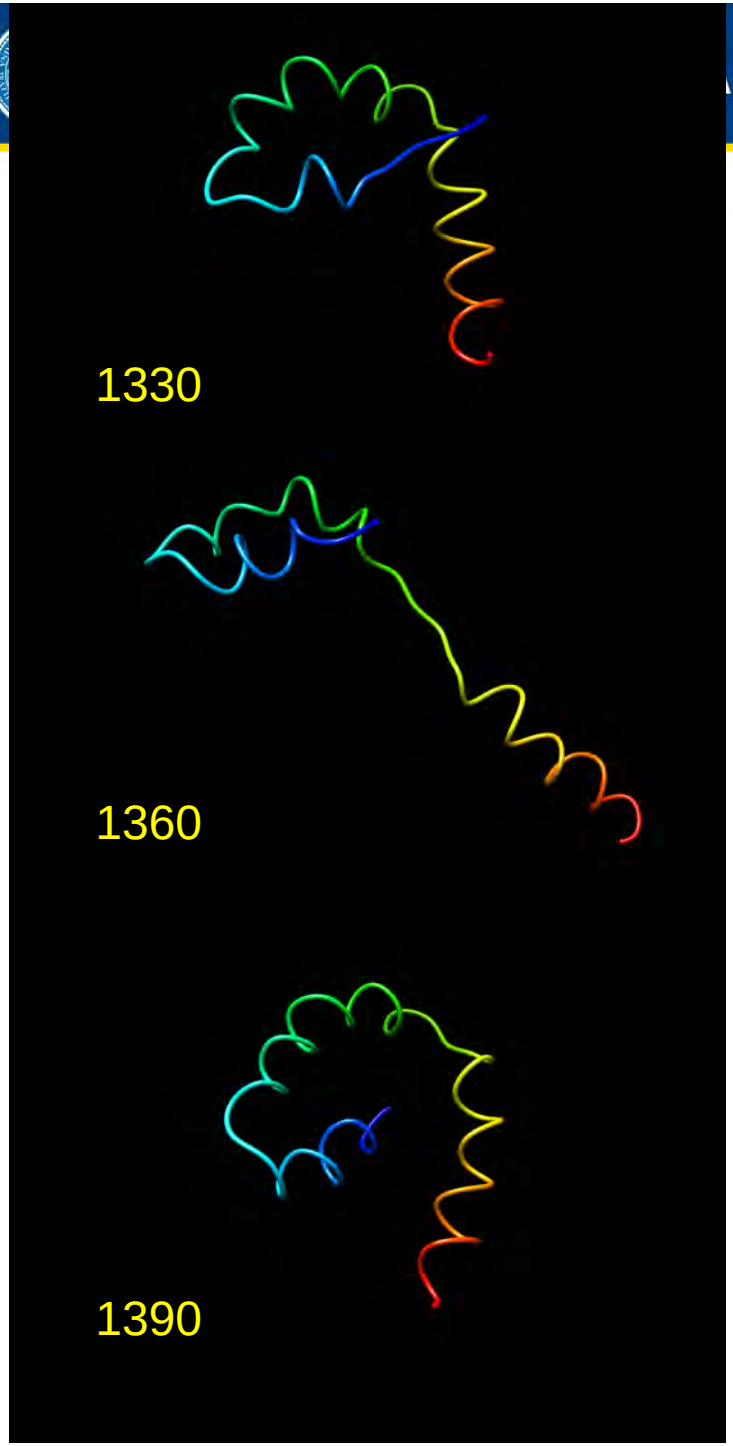


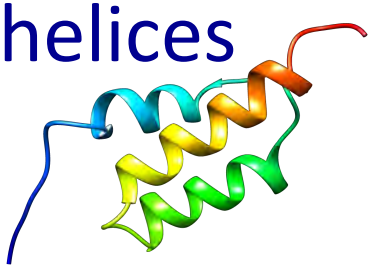
Monitor largest eigenvalue of bipartite distance matrix

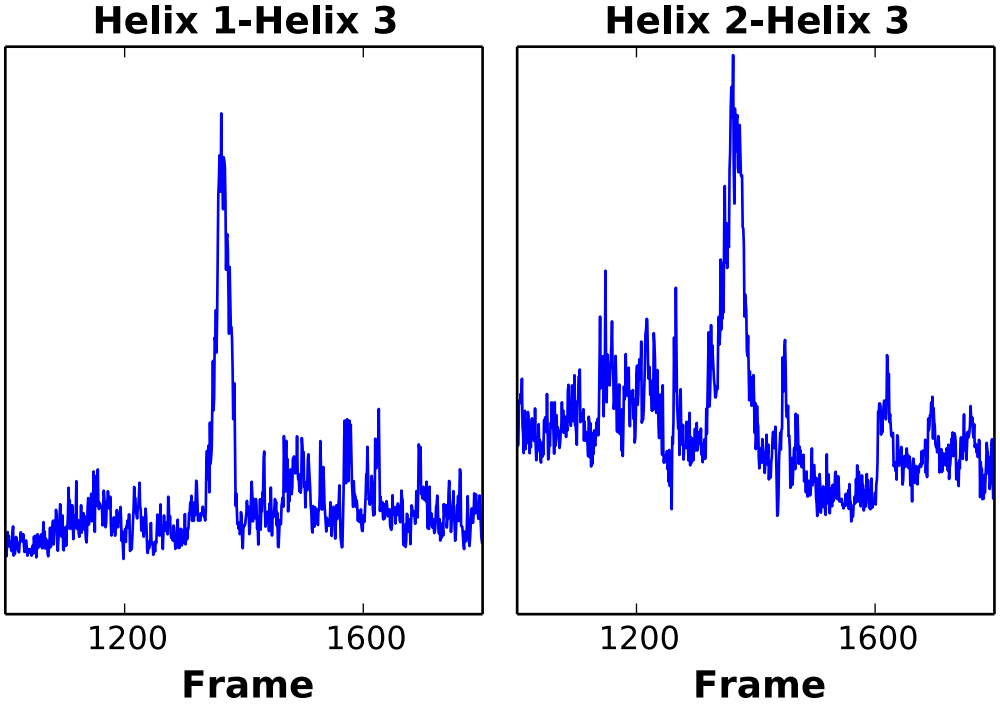


First and second α-helices appear stable; third helix moves

Case Study II: Capturing Movement of α-helices







Large relative change between two pairs of α-helices

"Storage technologies are advancing […] and it is really not clear at all [to me] that especially distributed storage platforms would not be able to handle […] petabyte data sets"

Anonymous Feedback

Yes, new technologies will be able to handle data at the extreme scale but *only* if we integrate new software paradigms.
In-situ and in-transit analysis are here to stay!