# Geometric Analysis of DNA in Molecular Dynamics Simulations of Nucleosomes
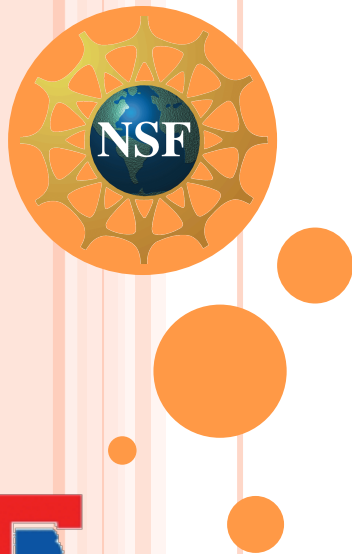
Rocky Brown

LA-SiGMA
Louisiana Alliance for Simulation-Guided Materials Applications

RU
RADFORD UNIVERSITY

# Personal Background

- Blacksburg, Virginia
- Rising senior in Physics at Radford University
- Planning to take Master's in Aerospace Engineering or Astrophysics at Virginia Tech
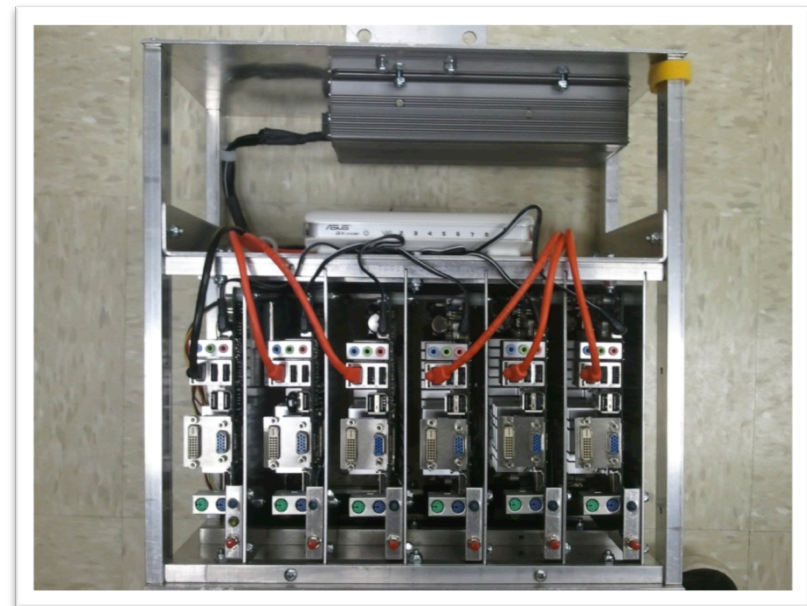
# GENERAL SYNOPSIS

- First Assignment
  - Parallel Computing (Little Fe)
- Research Objective
- The Dilemma
- HPTools
- Conclusion

# FIRST ASSIGNMENT

- Little Fe and Parallel Computing

# PARALLEL COMPUTING

- A form of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved concurrently



**Shared Memory Parallel Computer**

Processors

Central Memory shared by all processors
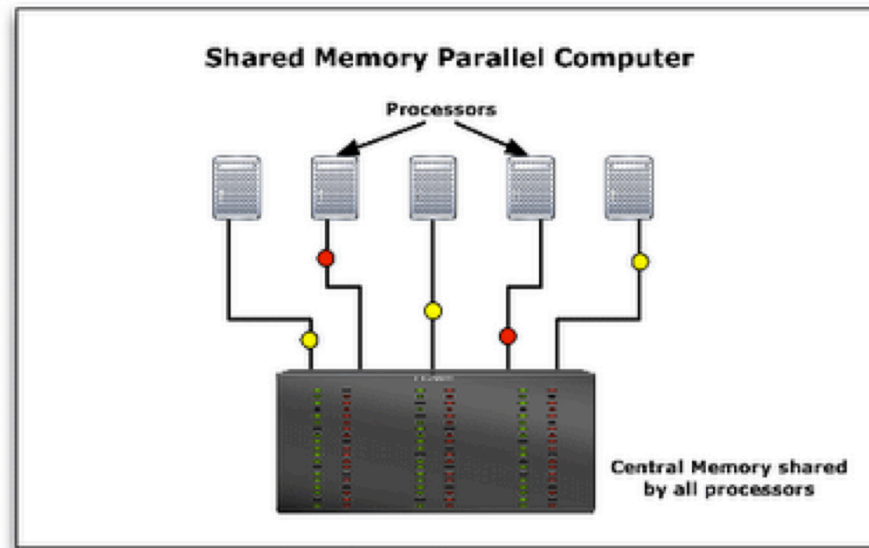
# RESEARCH OBJECTIVE

- Hope to gain insight into why nucleosomes position as they do on the genome
- Determine if the helical parameters are conserved throughout the simulations or whether the parameters are influenced by sequence
- Establish metrics for analysis that can be used in future simulations to identify DNA sequence properties

# OVERVIEW OF NUCLEOSOMES
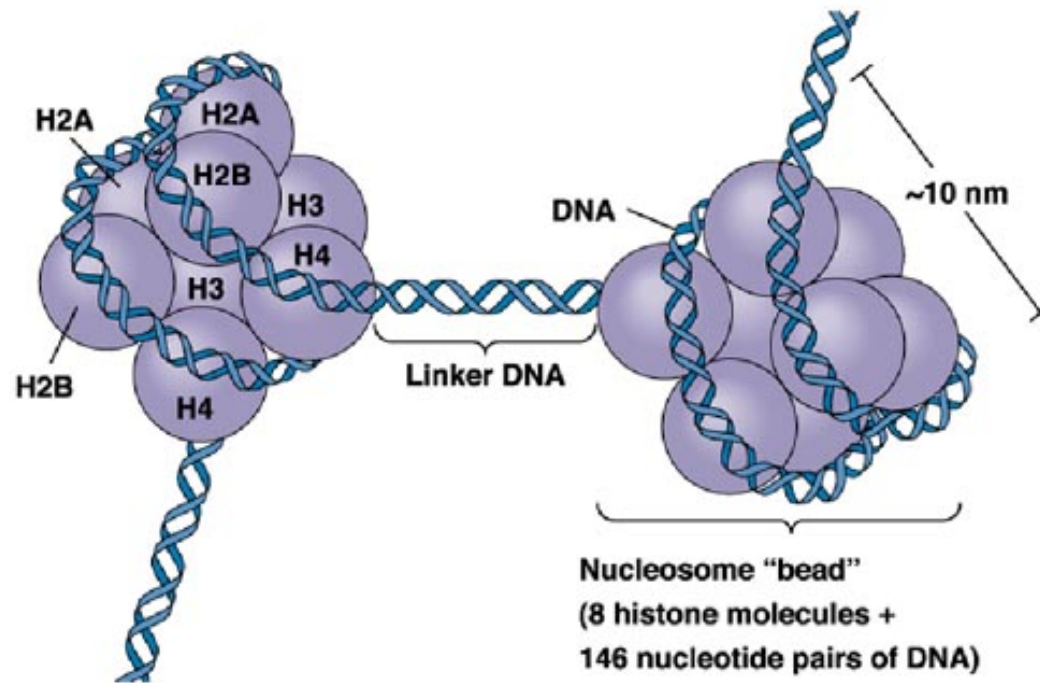


Ref: http://www.zoology.ubc.ca/~bio463/lecture_3.htm

# INTERESTING FACTS ABOUT NUCLEOSOMES

- A nucleosome core particle (NCP) is a biomolecular complex of eight histone proteins around which is wrapped 147-base pair of DNA

- Nucleosomes fold long lengths of DNA into a highly compact superhelix

- Folding- or packing- influences genetic functions such as transcription, replication, regulation and repair

- Nucleosome formation requires the 147bp of DNA to assume a specific super-helical conformation

# NUCLEOSOME FORMATION

# Base Pair Parameters
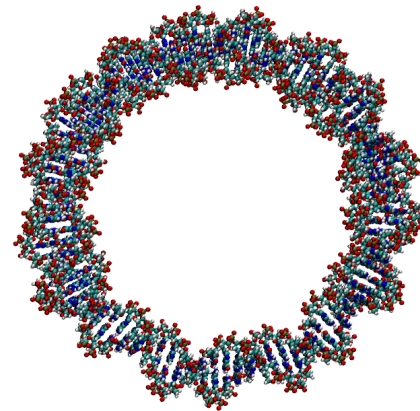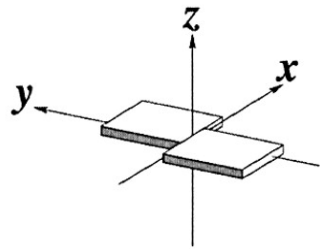
Shear (Sx)
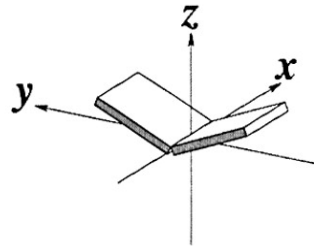
Buckle ($\kappa$)
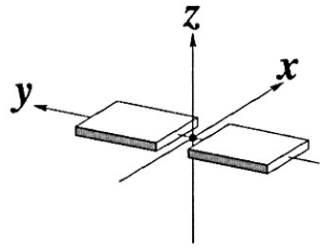
Stretch (Sy)
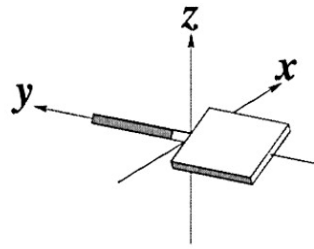
Propeller ($\pi$)
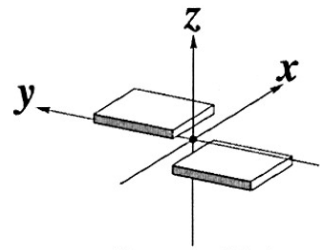
Stagger (Sz)
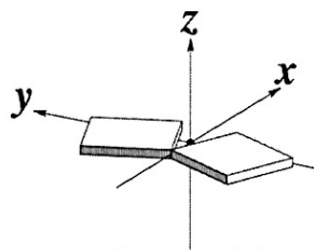
Opening ($\sigma$)

Translations
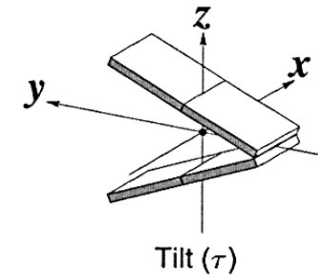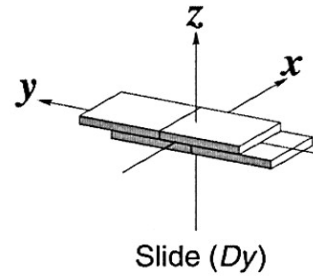
Rotations

# Dimer Step Parameters
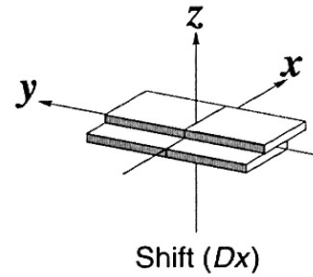
Shift (Dx)

Tilt ($\tau$)

Slide (Dy)
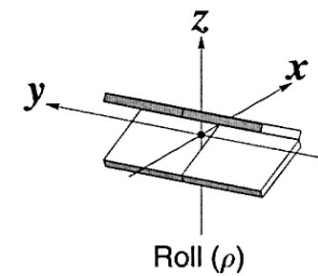
Roll ($\rho$)
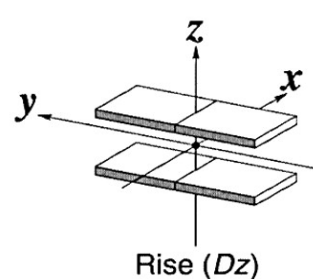
Rise (Dz)
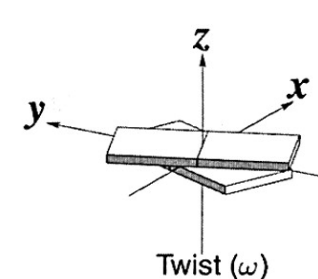
Twist ($\omega$)

Translations

Rotations

# RESEARCH PROCESS

Analyze DNA helical parameter data → Examine the geometric properties of DNA sequences → Research Objective
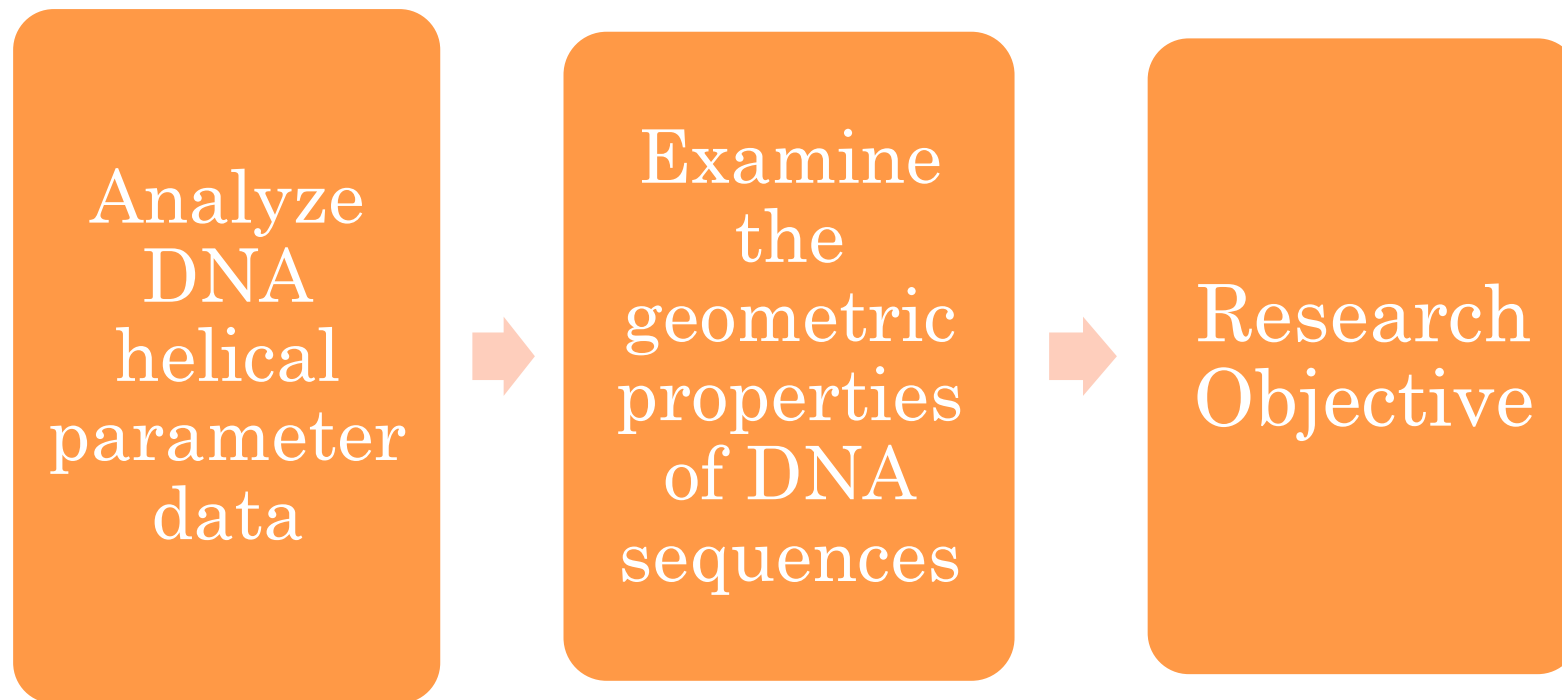
# THE DILEMMA

21 nucleosomes position  x 16 chromosomes

336 simul... ...arameter

*INFORMATION OVERLOOAAADDDDD!!!!*

# THE SOLUTION

- Necessary to develop automated software tools to assist in the task

- Requirements of the software :

  - Portable (operating system to operating system)

  - Able to execute high level functions such as Fourier Transforms

  - Supportive of graphing and plotting utilities

# THE SOLUTION REVEALED!

# python™ FOR DUMMIES 101

**High-Level Programming Language**

**Large Standard Library**

**FREE!!!**
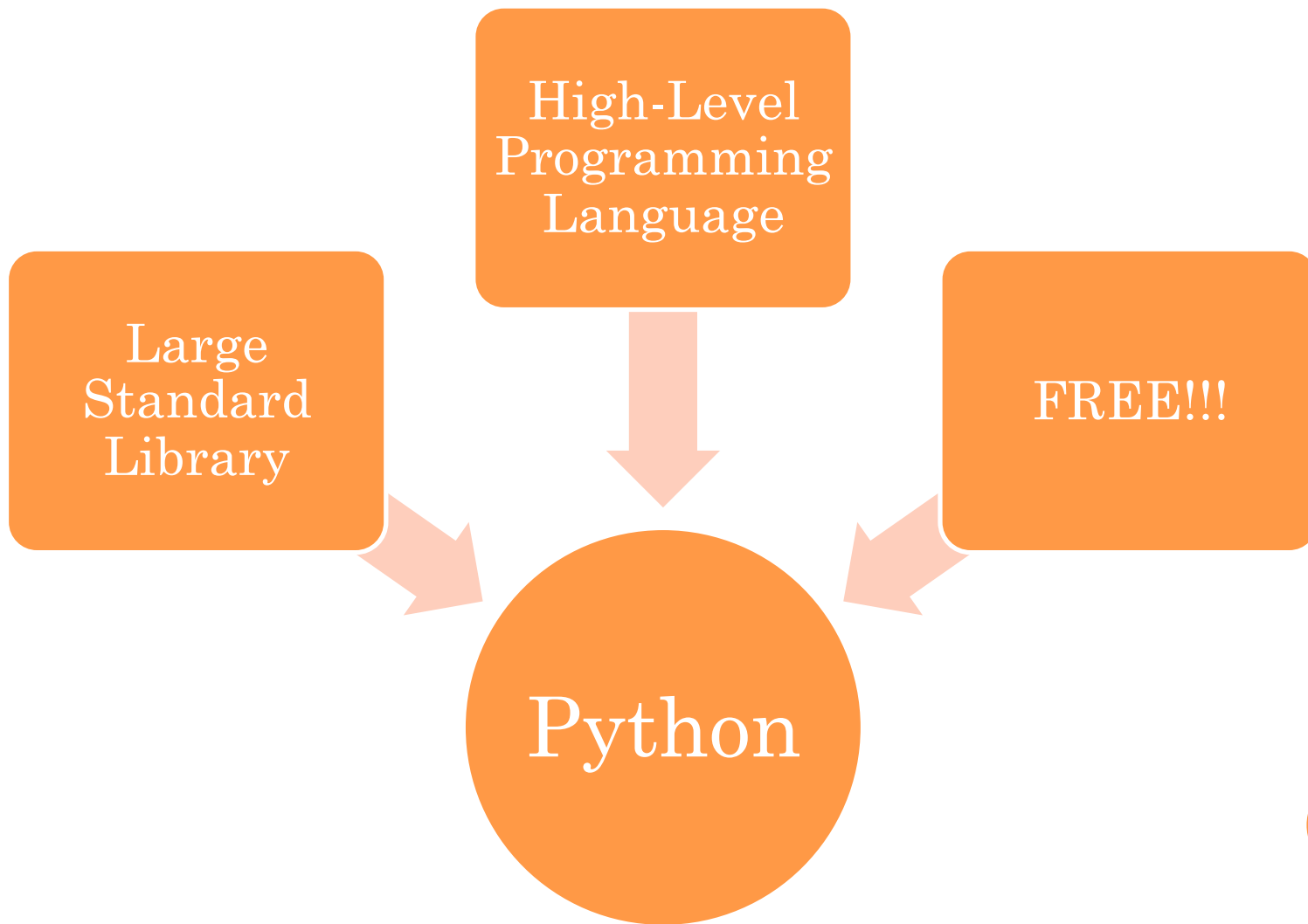
**Python**

# LET THE CODING BEGIN!

- We began by doing simple codes to quickly learn the syntax

```
#!/usr/bin/python
print ('Hello World')
```

- "It's that simple. Now just make HPTools." (Dr. Thomas C. Bishop, REU 2012)

# WHAT IS HPTOOLS?

- A module of independent functions used to make the entire nucleosome simulation workflow more efficient
- Benefits :
  - Real-time analysis
  - More efficient workflow
- Utilities :
  - File collection utilities
  - Fourier filtering
  - Multiple plotting commands

# COMPREHENSIVE LOOK AT HPTOOLS

```python
#!/usr/bin/python
###
### Python Tools to manipulate, analyze and plot DNA helical parameter data
###     TcB @ Louisiana Tech 2012
###     with REU 2012 contributions from Victorial Bamburg, Rocky Brown, and
###     and additional assistance from James Liman(Tech)

import sys
import numpy
import re
import csv
import math
import os
import string
from pylab import *
#import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D


########### h
### helper routine to read column data
##############

def readdat(fn,skiprows = 0):
def readseq(dirfp):
def readpar(dirfp,skiprows=0):


def writepar(parfn, hpdata):

def stats(hpdata, calc):

def kink():
    ## THis will take an Hparray and determine if there are kinks
    ## and return an hparray with 0's and 1's for kinks or not kinks, irrespectively.

def fft(hpdata):

def filter(hpdata):


def ifft(hpdata,filter=0):

def plot(hpdata,nhp=12,abcmult=2):


def stdplot(hpdata,nhp=12,abcmult=1):


def contour(hpdata,nhp=12,abcmult=2):
```

# CLOSER LOOK AT A FUNCTION

```python
########################3
### FUnctions to plot HP data
####################

def plot(hpdata,nhp=12,abcmult=2):
    plist=[1,4,7,10,2,5,8,11,3,6,9,12]
    filelist=['She','Str','Sta','Buc', 'Pro', 'Ope','Shi', 'Sli','Ris', 'Til', 'Rol','Twi' ]

    print"### the abc values are the average and standard devications observed for sims of dna free in solution"
    abcavg=[0.02, 0.03, 0.09, 1.2,-11.0, 2.1, -0.05, -0.44, 3.32, -0.3, 3.6, 32.6]
    abcstd=[0.31, 0.12, 0.41, 12.4, 9.3, 4.6, 0.76, 0.68, 0.37, 4.6, 7.2, 7.3]

    seq=hpdata[0]
    hparray=hpdata[1]
    nbp = len(seq)
    print "#give the time steps and tell in what increments it needs to skip"
    figure(figsize=(20,10))
    for id in range(1,13):
        hp = plist[id-1]
        y = hparray[hp-1]
        if nhp == 12:
            subplot(3,4,id)
            plt.plot(y)
            title(filelist[hp-1],fontsize=10)
            if abcmult != 0:
                cvals = (0,nbp,-2*abcmult*abcstd[hp-1]+abcavg[hp-1],2*abcmult*abcstd[hp-1]+abcavg[hp-1])
                plt.axis(cvals)
                axhline(y=-abcmult*abcstd[hp-1]+abcavg[hp-1],color='r',linestyle=':')
                axhline(y= abcmult*abcstd[hp-1]+abcavg[hp-1],color='r',linestyle=':')

        if nhp == 6:
            subplot(2,3, id)
        if nhp == 1:
            surf = ax.plot_surface(X,Y,Z)
            cvals = arange(-abcmult*abcstd[hp-1]+abcavg[hp-1],abcmult*abcstd[hp-1]+abcavg[hp-1],abcstd[hp-1]*2)
            contour(X,Y,Z,cvals)
            plt.colorbar(shrink=0.8,format='% 4.1f')
            title(filelist[hp-1],fontsize=10)
    show()
    return()
```

# PROCESS FLOWCHART
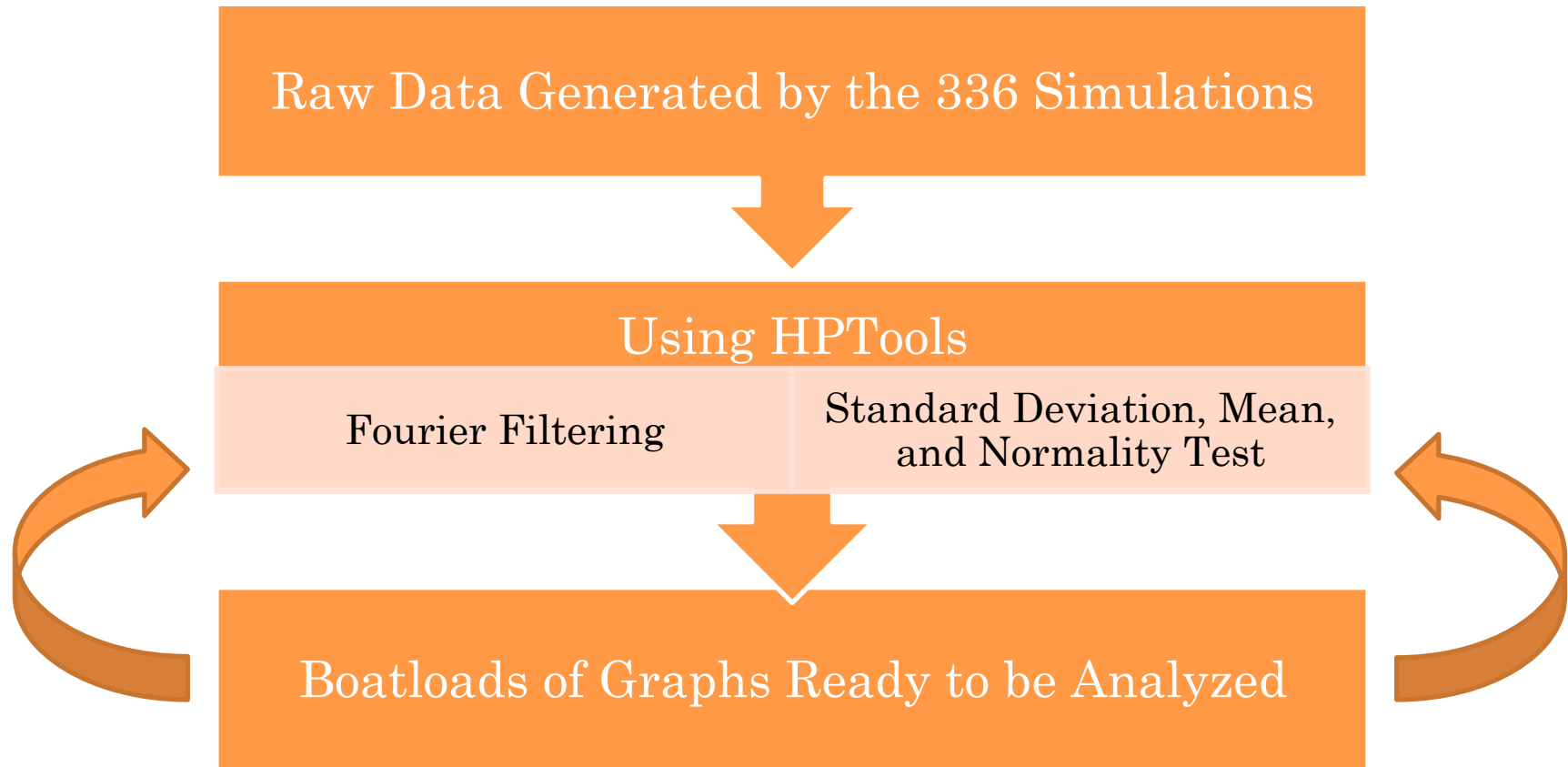
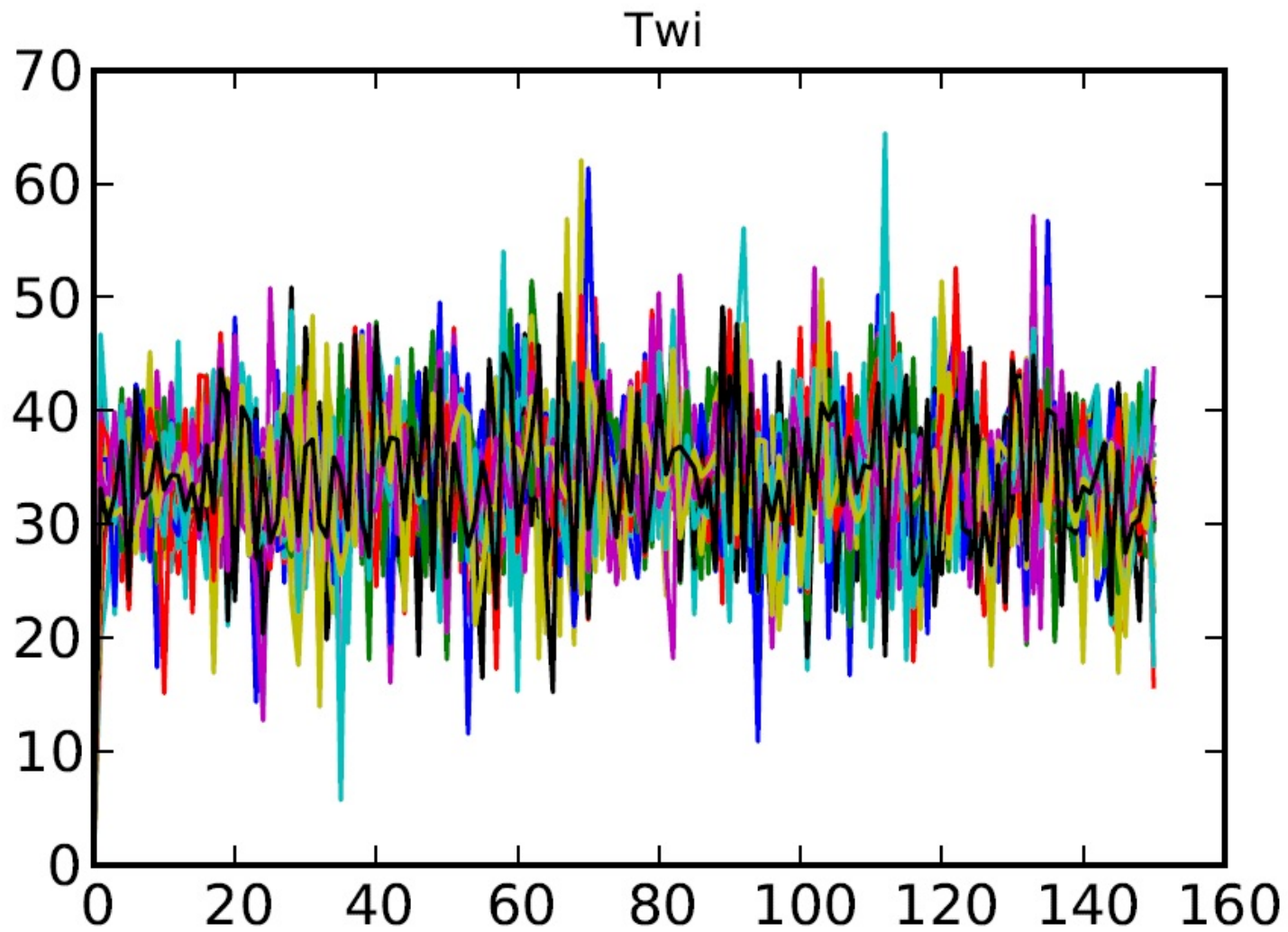Raw Data Generated by the 336 Simulations
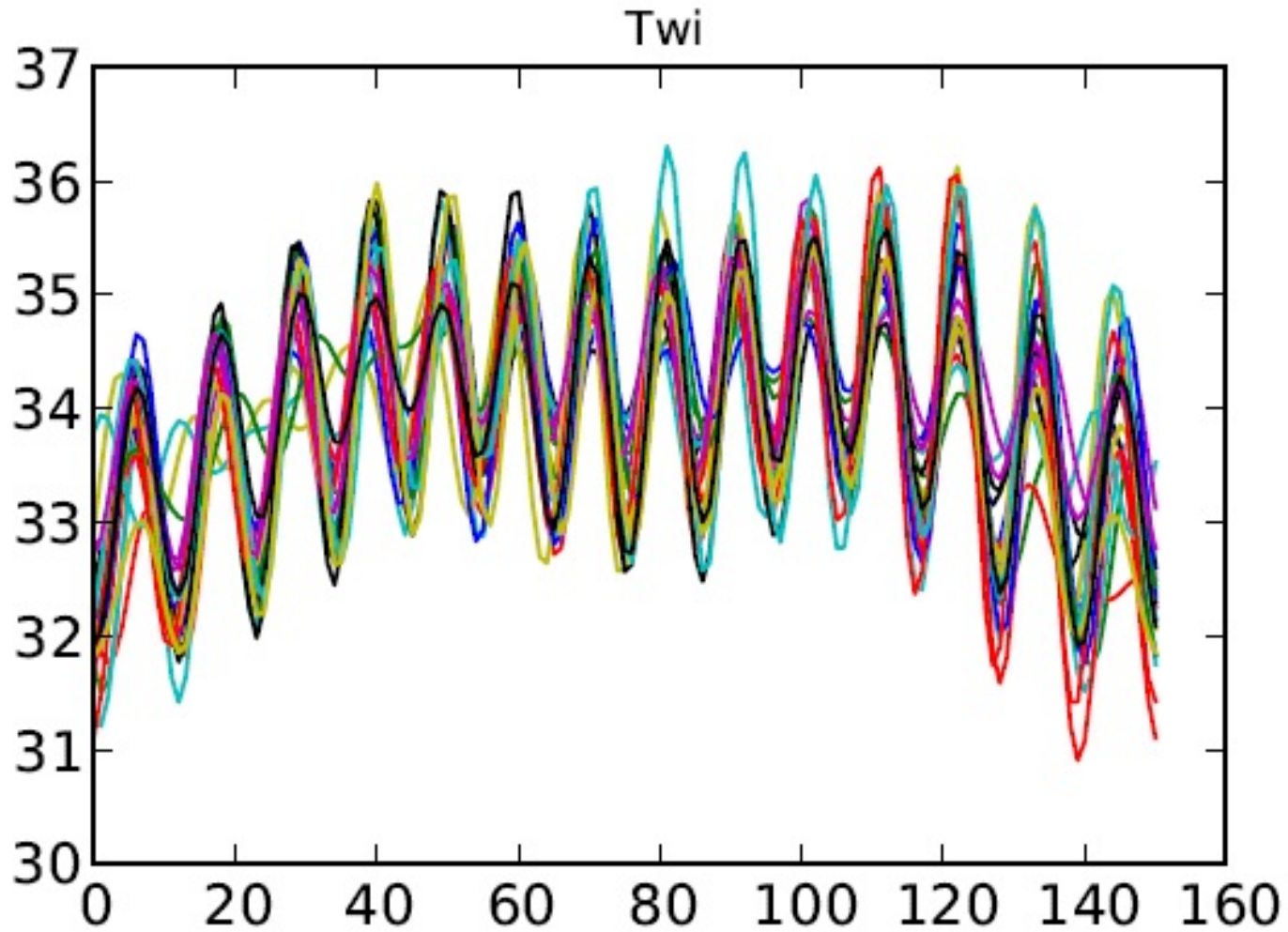
Using HPTools

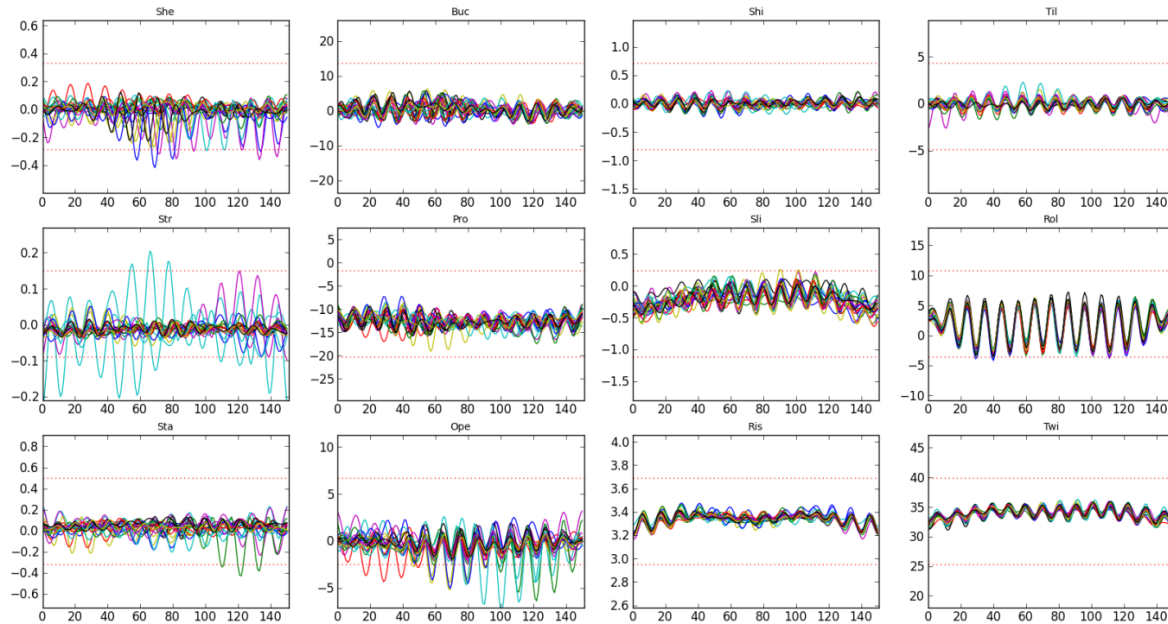| Fourier Filtering | Standard Deviation, Mean, and Normality Test |
|---|---|

Boatloads of Graphs Ready to be Analyzed

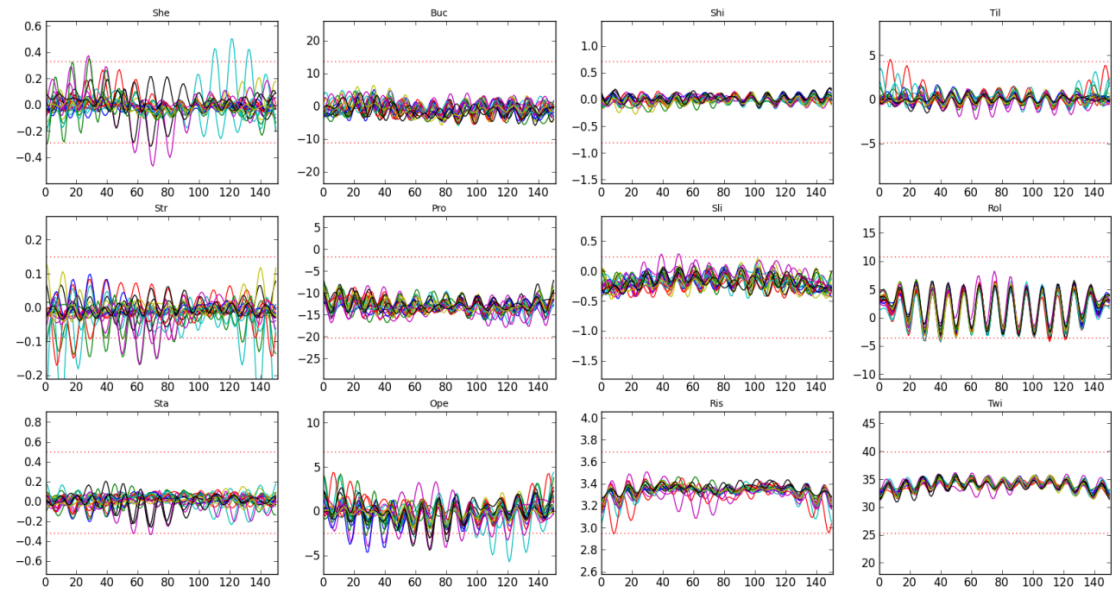# UNFILTERED TWIST DATA
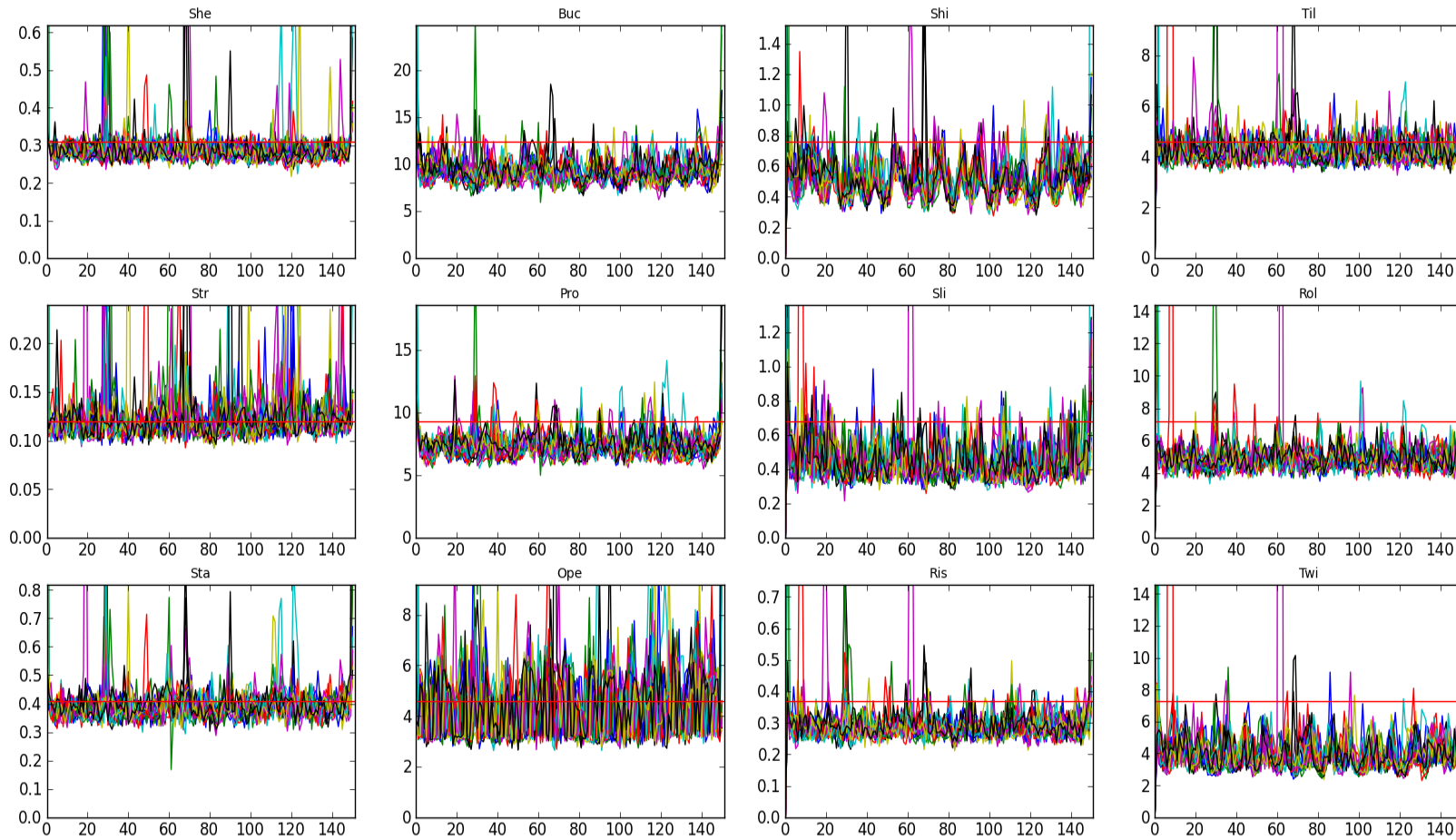

Twi

# FILTERED TWIST DATA

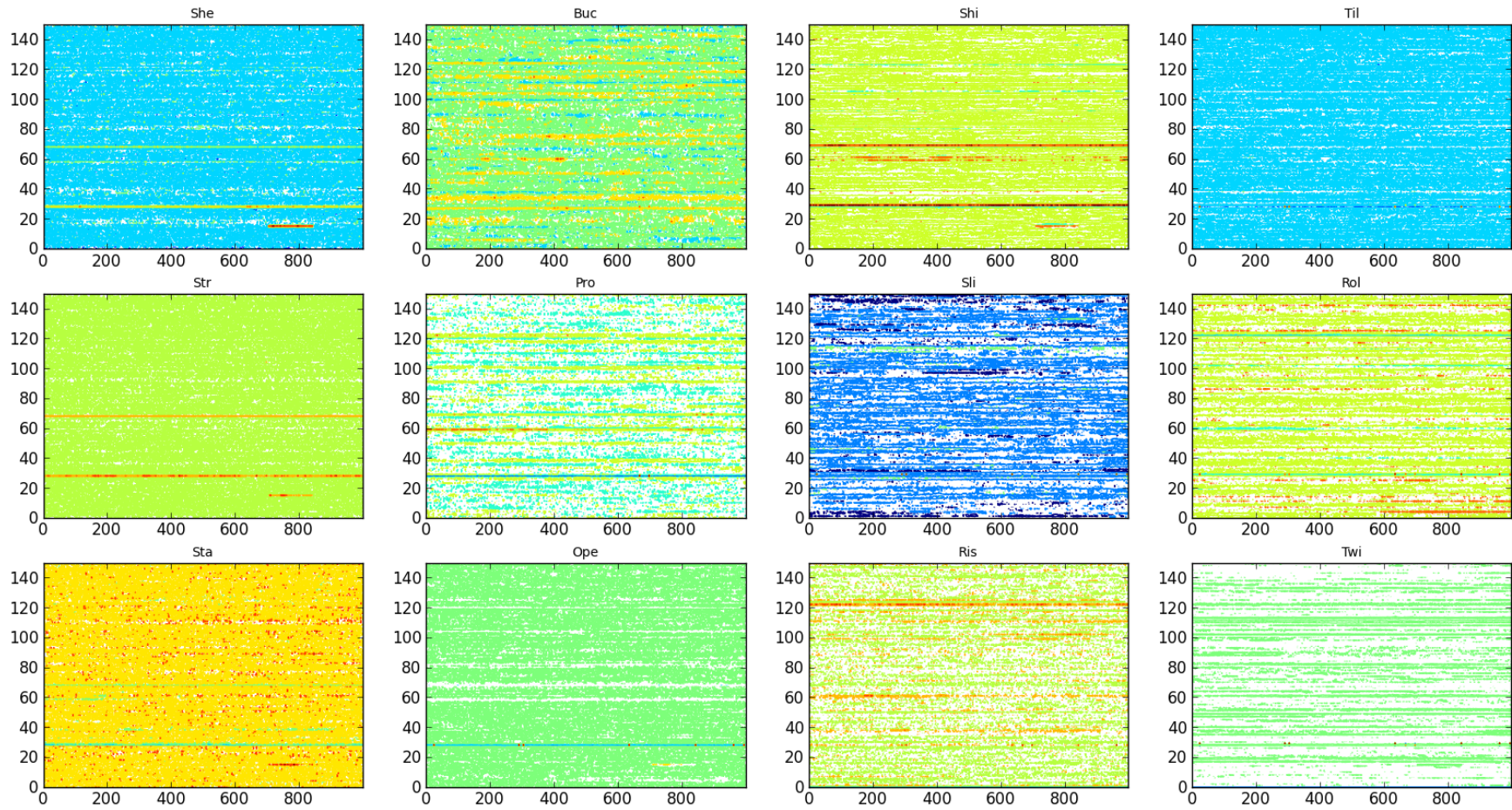# RESULTS

Chromosome I

Chromosome II

# STANDARD DEVIATION GRAPH



The solid red line is the corresponding value of standard deviation of DNA free in solution

# Contour Graph



Chromosome XV

# RESULTS RELATED TO PREVIOUS ANALYSIS

➢ Values for Roll, Twist, and Slide were highly conserved, as expected

➢ This finding extends the previous results (Bishop 2005) which were based on only one simulation

➢ Evidence of kinks

# CONCLUSION

➢ It is clear that the patterns of Roll, Twist, and Slide that are necessary and sufficient for nucleosome formation are not affected by DNA sequence

➢ Based on the simulations that we have analyzed, there is only one conformation of DNA superhelix

➢ Nucleosomal DNA is less flexible than free DNA as indicated by values of standard deviations

# FUTURE WORK USING HPTOOLS

- Further analyze the remaining 13 chromosomes helical parameter data
- Incorporate real-time analysis into our simulation workflow
- Study kinks in detail

# ACKNOWLEDGEMENTS

THANK YOU FOR YOUR ATTENTION