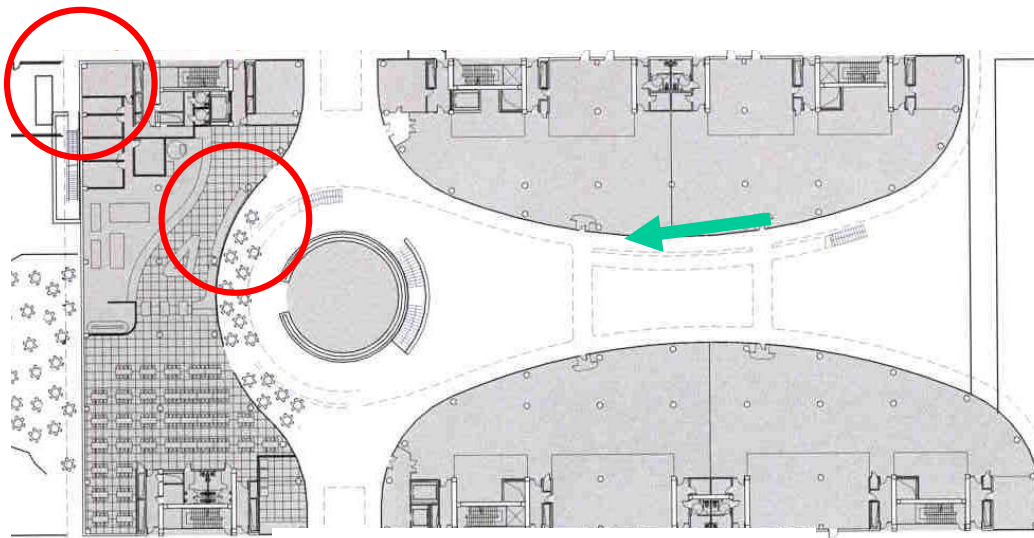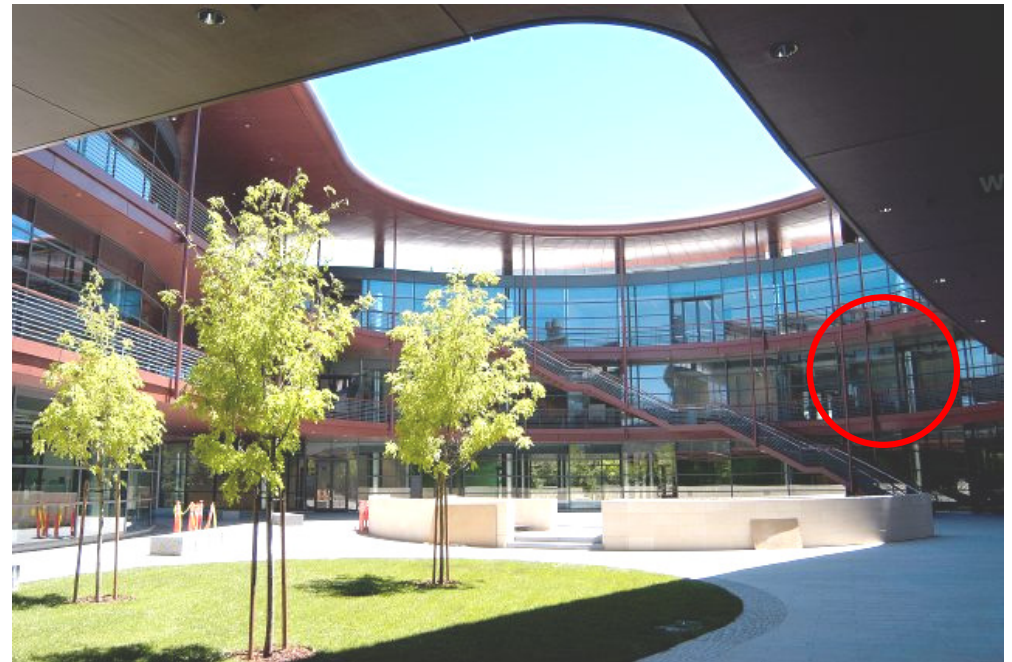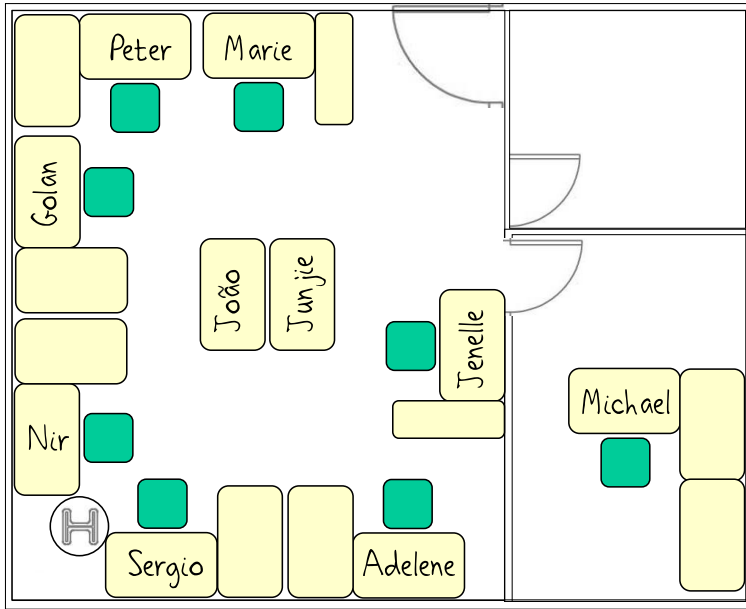# MESOSCALE MODELING OF MACROMOLECULAR MACHINES

Cryo-EM Workshop

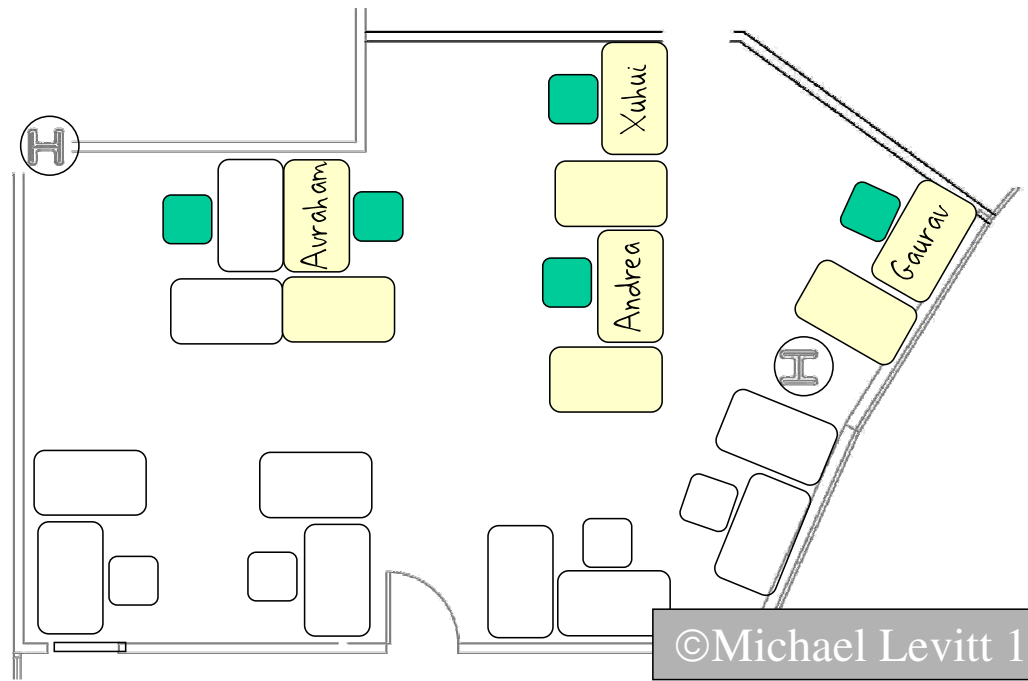16 January 2010
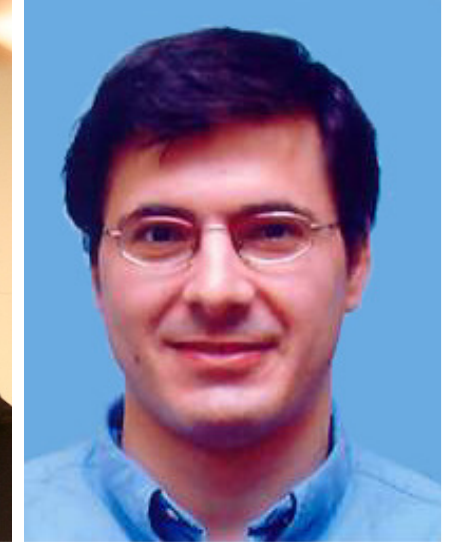
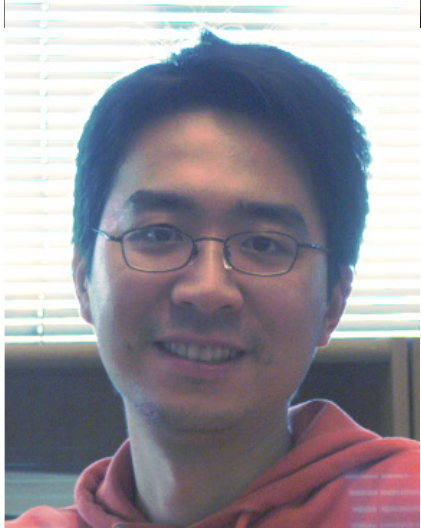Michael Levitt, Structural Biology & Computer Science, Stanford

http://csb.stanford.edu/levitt

MINI RESEARCH CENTER (MRC)

Names on floor plan (upper left): Peter, Marie, Golan, João, Junjie, Jenelle, Michael, Nir, Sergio, Adelene
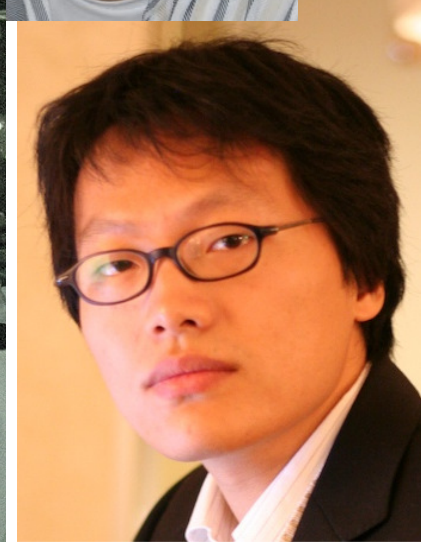
Norman Foster

Names on floor plan (lower right): Xuhui, Avraham, Andrea, Gaurav

THE PROBLEM

Venki Ramakrishnan Picture of Year 2008

Wellcome Images
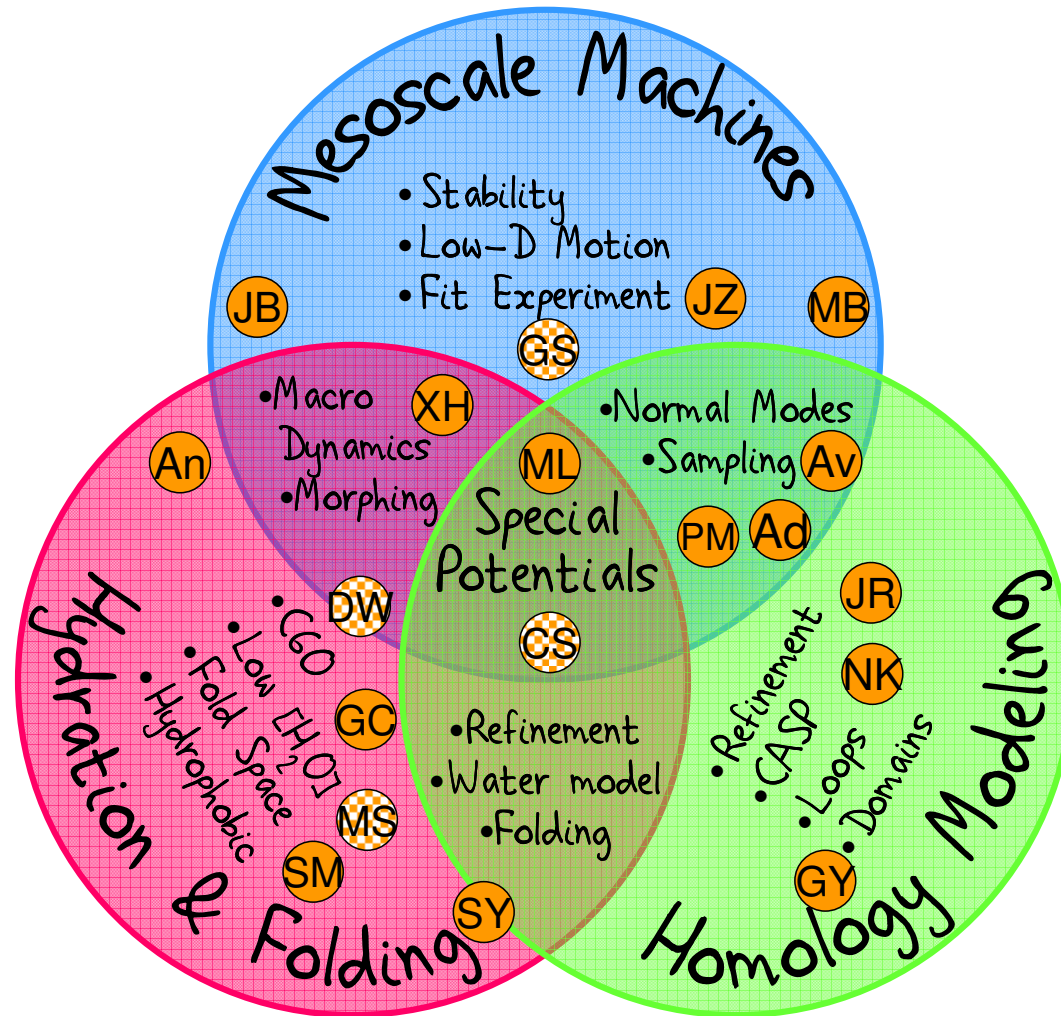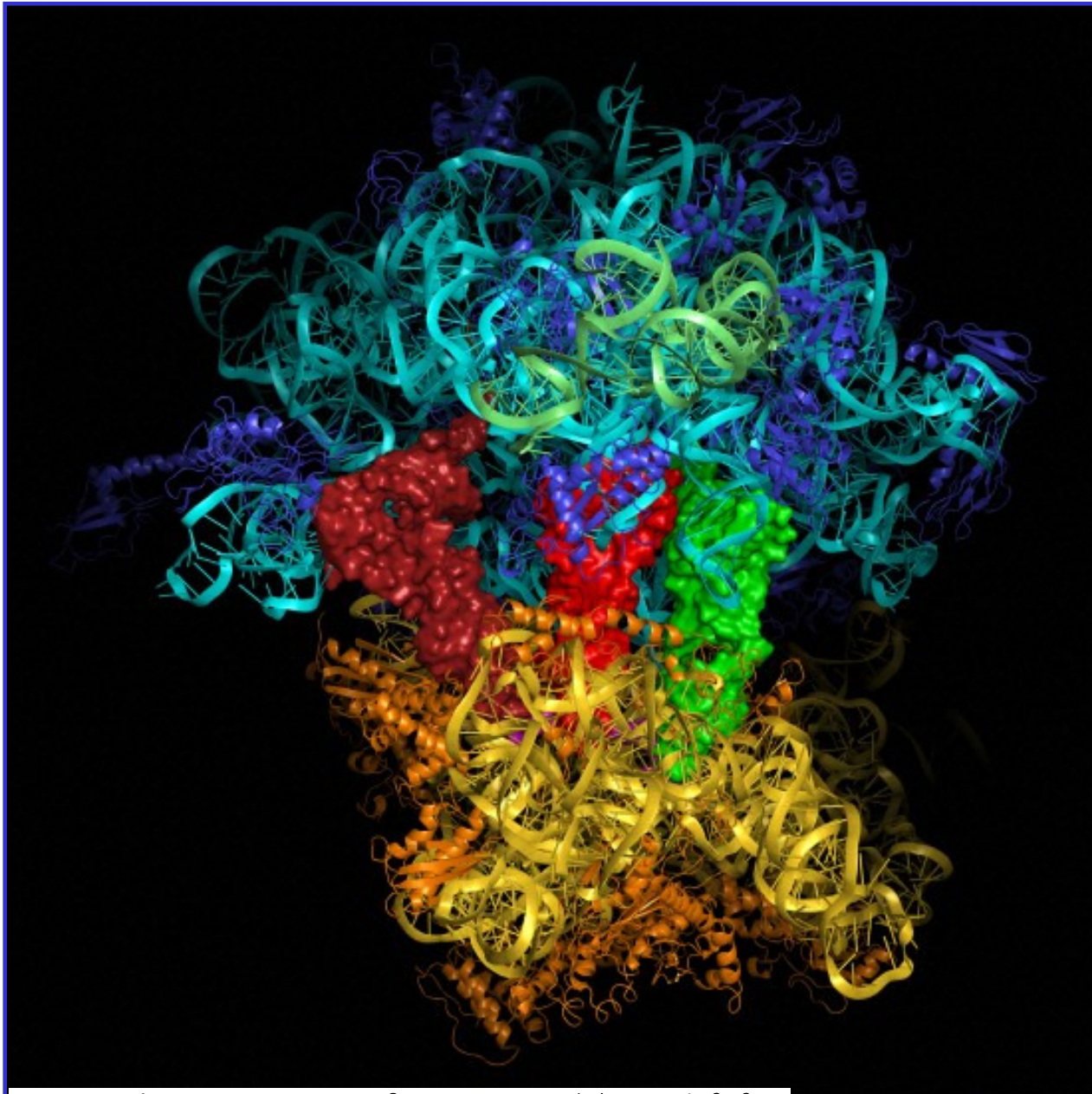
- The ribosome is a molecular machine.

- It has 56 chains
  50 protein
  6 RNA

- It has 10559 residues
  6065 protein
  4494 RNA

OUTLINE

- Few normal modes simulate suggestive motion.

- Knowledge-based potentials work for refinement.

- Simplified models are same as all-atom models.

- Simulate suggestive motion of cellular machinery.

- Knowledge-based potentials give stability & modes.

- The chain order paradox.

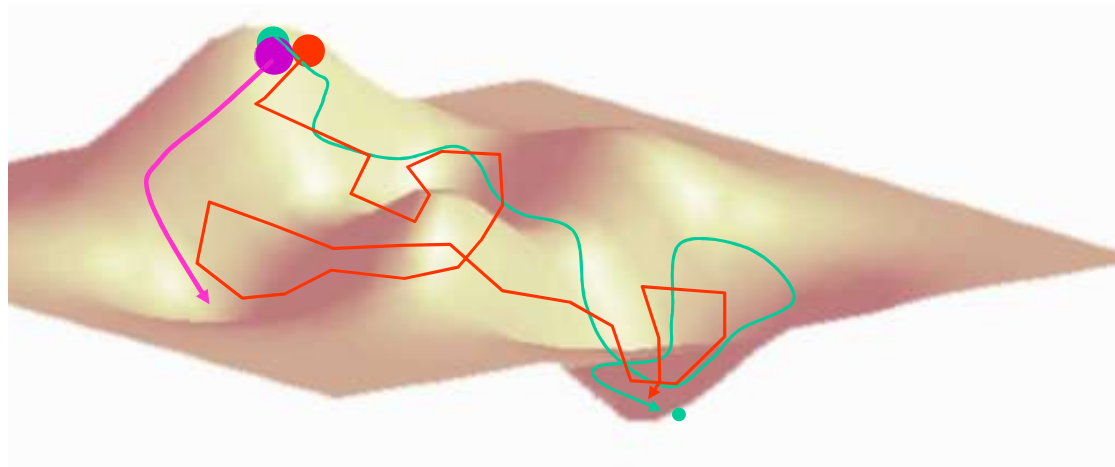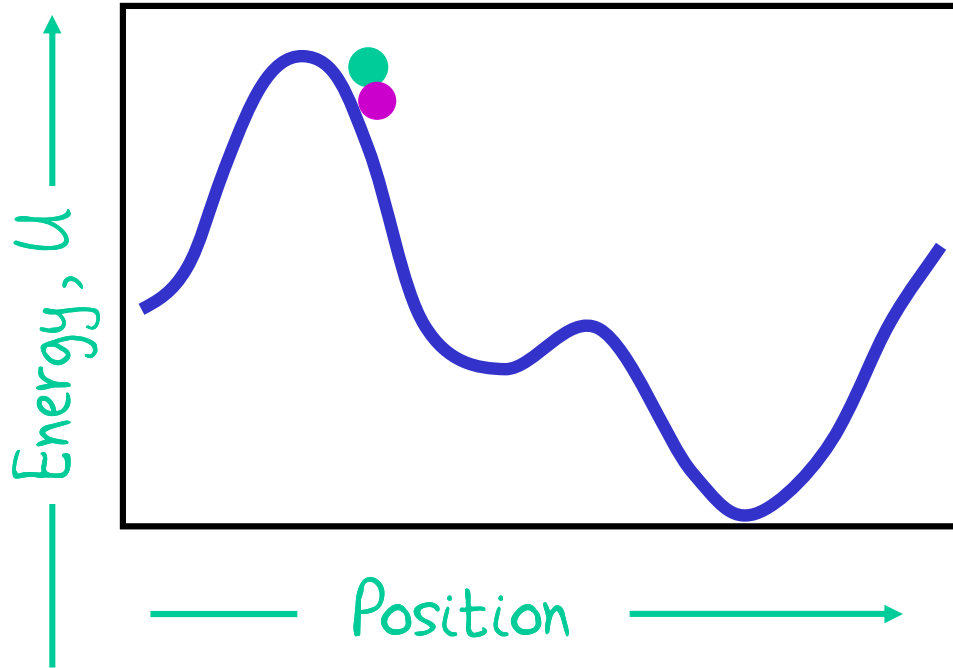- Special solution of generalized eigenvalue equation.

Molecular Motion

TWO ESSESTIAL REQUIREMENTS

- Potential Energy Surfaces.

- Methods to Move Over Surface.

MOVING OVER ENERGY SURFACE

- Energy Minimization drops into local minimum.

- Molecular Dynamics uses thermal energy to move smoothly over surface.

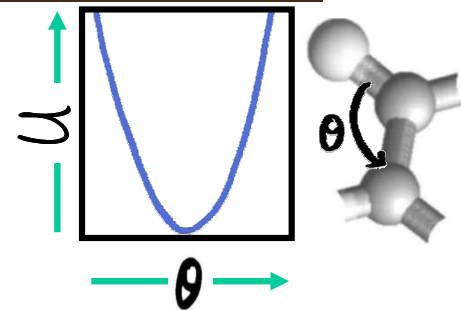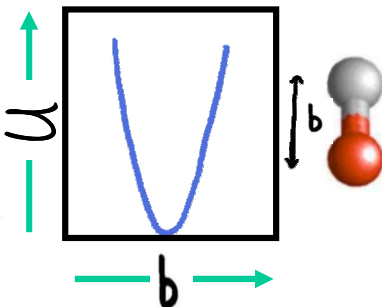- Monte Carlo Moves are random. Accept with probability $\exp(-\Delta U/kT)$.

Molecular Dynamics

MOLECULAR POTENTIAL ENERGY

$$U = \sum \tfrac{1}{2} K_b (b - b_0)^2 - \sum \tfrac{1}{2} K_\theta (\theta - \theta_0)^2$$
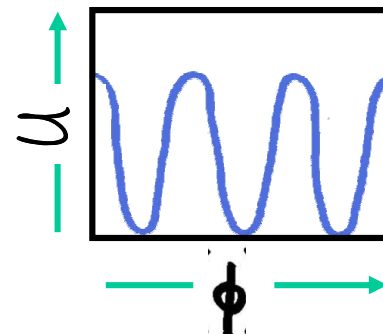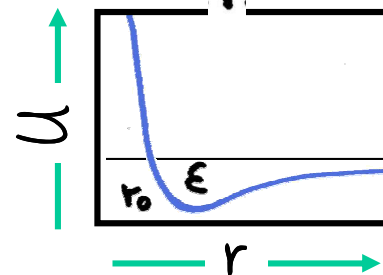
All Bonds   All Angles

$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

All Torsion Angles

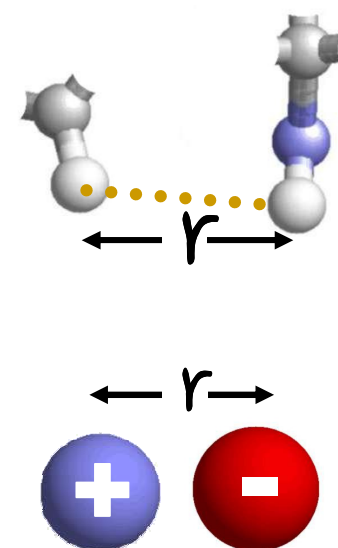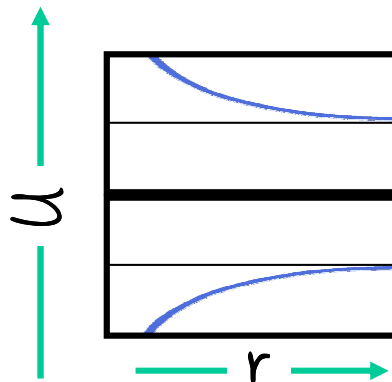$$+ \sum \epsilon \left[\left(\tfrac{r_0}{r} \right)^{12} - 2 \left(\tfrac{r_0}{r} \right)^6 \right]$$
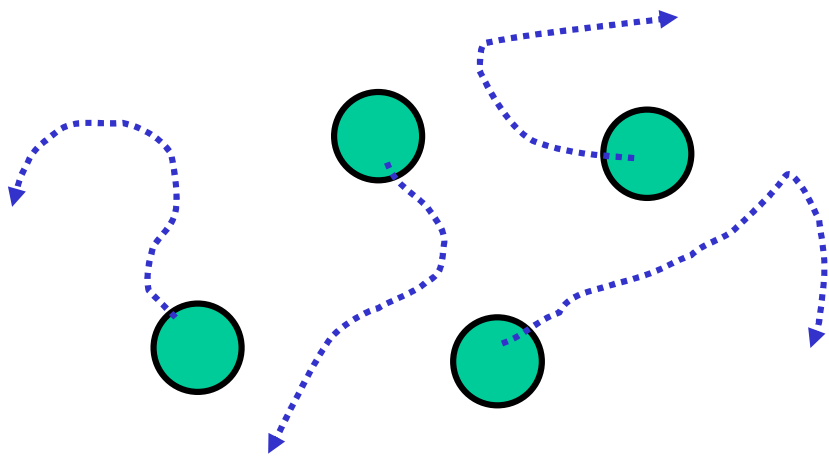
All nonbonded pairs

$$+ \sum 332 q_i q_j / r$$

All partial charges

Simple sum over many terms

MOLECULAR DYNAMICS THEORY

- All atoms move together.

- Forces between atoms change with time.

- Analytical solution to give $x(t)$ and $v(t)$ is impossible.

- Numerical solution is trivial.

$$x(t+\Delta t) = x(t) + v(t)\Delta t + [4a(t) - a(t-\Delta t)]\Delta t^2/6$$

New position   Old position   Old velocity   Acceleration

$$v(t+\Delta t) = v(t) + [2a(t+\Delta t) + 5a(t) - a(t-\Delta t)]\Delta t/6$$

New velocity   Old velocity   Acceleration

$$U_{kinetic} = \frac{1}{2}\sum m_i v_i(t)^2 = \frac{1}{2}n k_B T$$

Kinetic energy   Atomic masses, velocities

Number of coordinates (not atoms)

Temperature

Boltzmann's Constant

Time step, Δt, must be very small at 10^{-15} seconds or 0.001 ps.

Total energy ($U_{potential}$ + $U_{kinetic}$) must not change with time.

Normal Modes

NORMAL MODE DYNAMICS

- In regular Molecular Dynamics, we solve the *exact* equations of motions *approximately*.

- In Normal Mode Dynamics, we solve the *approximate* equations of motion *exactly*.
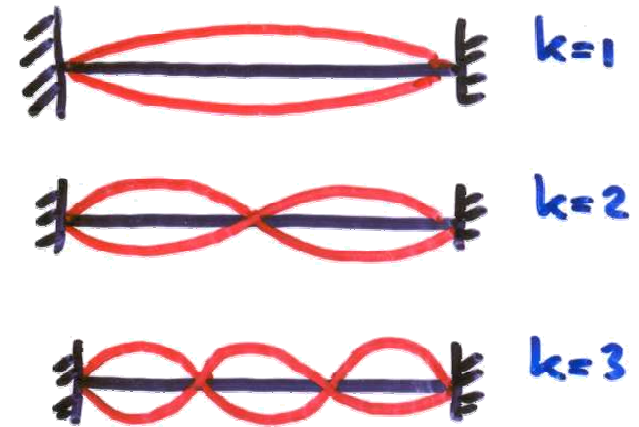
BASIC THEORY

What are normal modes?

- **A string attached at both ends:**

Get a standing wave of frequency $\nu = k\nu_0$

Amplitude is proportional to $1/\nu$

Each mode can be excited independently.

k=1

k=2

k=3

- **Discrete point masses:**

C is the spring force constant

$U(x) = 1/2\, C x^2$. Now $F = ma = -Cx$ or $m\, d^2x/dt^2 = -Cx$

Solution is $x(t) = a\cos(\omega t + \delta)$, with $\omega = \mathrm{sqrt}(C/m)$
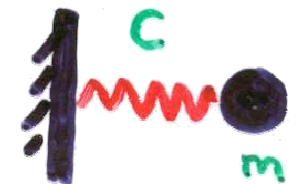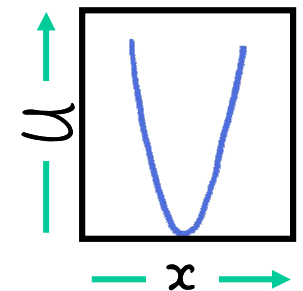
Get amplitude, a, by the equipartition theorem:

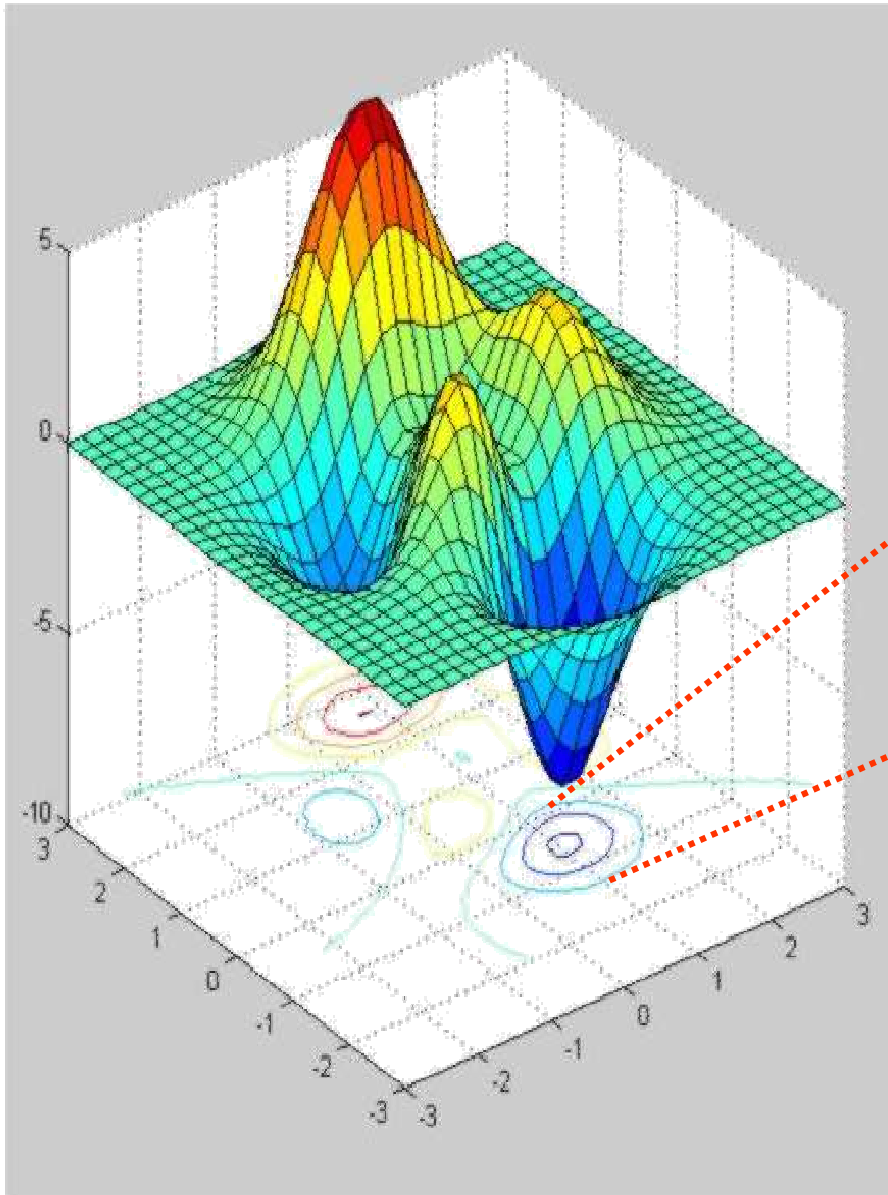$\langle E_{potential}\rangle = 1/2\, C \langle x^2\rangle = 1/2\, k_b T$

$\langle x^2\rangle$ is the mean value of x^2.

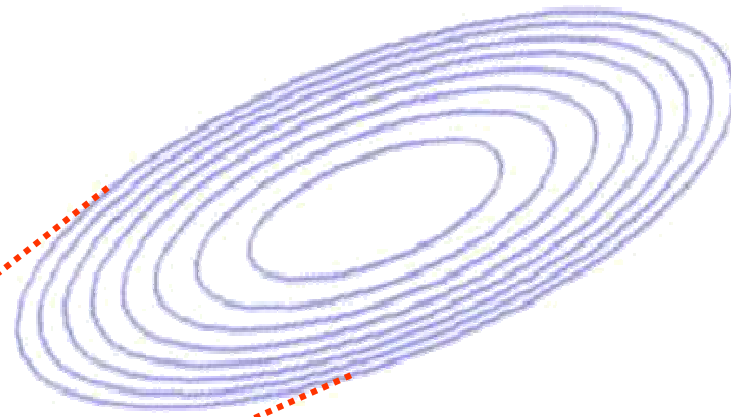Thus, $a = \mathrm{sqrt}(2k_b T/C)$ as $\langle x^2\rangle = 1/2\, a^2$ for a cosine wave.
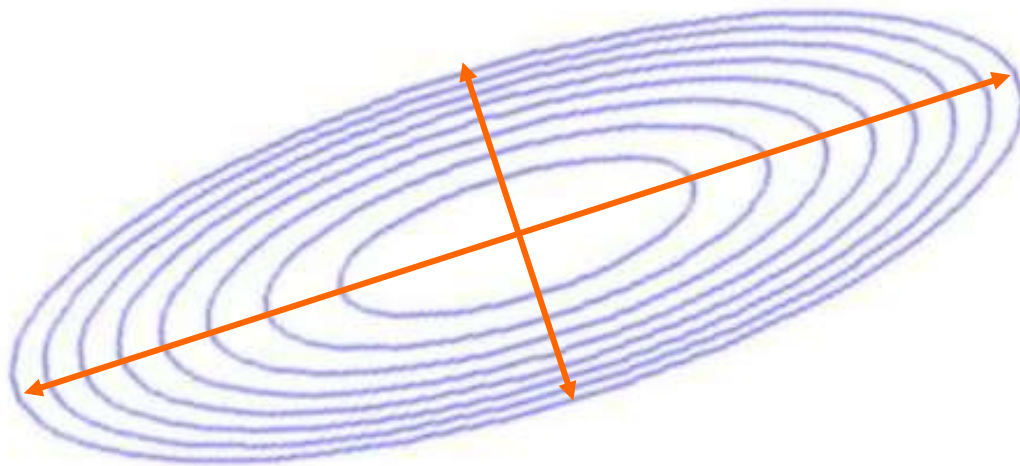
NORMAL MODES IN HIGH DIMENSION

Focus on deepest energy minimum.

- Expand energy function about minimum.

- Approximate as a quadratic function
$$F(x,y) = Ax^2 + Bxy + Cy^2.$$
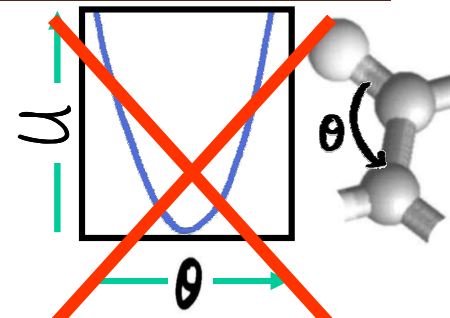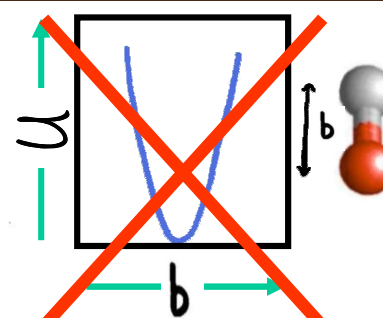
NORMAL MODES IN HIGH DIMENSIONS₃

- The normal mode directions are the major and minor axes of the ellipse.

- All other motion is a linear combination of these basic motions.

- Solving for the modes requires a matrix that is N **X** N, where N in number of degrees of freedom.

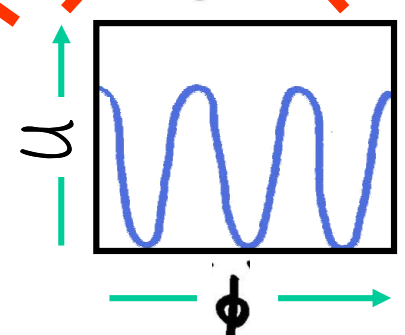For n atoms, N = 3n so it can be very big!

MOLECULAR POTENTIAL ENERGY

$$U = \sum \tfrac{1}{2} K_b (b - b_o)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_o)^2$$

All Bonds   All Angles

$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

All Torsion Angles

$$+ \sum \epsilon \left[\left(\frac{r_o}{r}\right)^{12} - 2 \left(\frac{r_o}{r}\right)^{6} \right]$$

All nonbonded pairs

$$+ \sum 332 q_i q_j / r$$

All partial charges

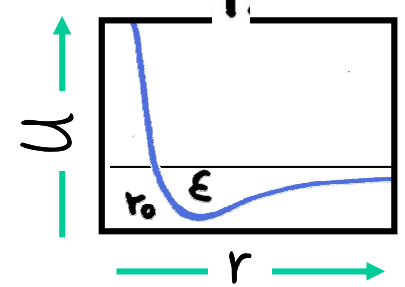Eliminate the strongest springs.

POTENTIAL ENERGY IN TORSION SPACE

$$U = \sum K_\phi [1 - \cos(n\phi + \delta)]$$

All Torsion Angles

$$+ \sum \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right]$$
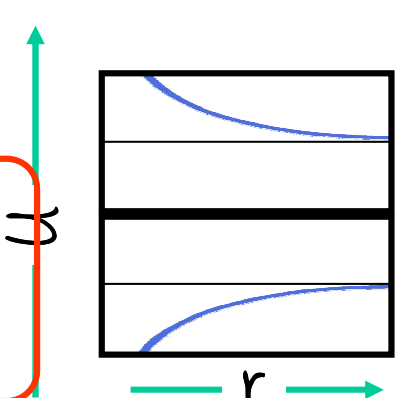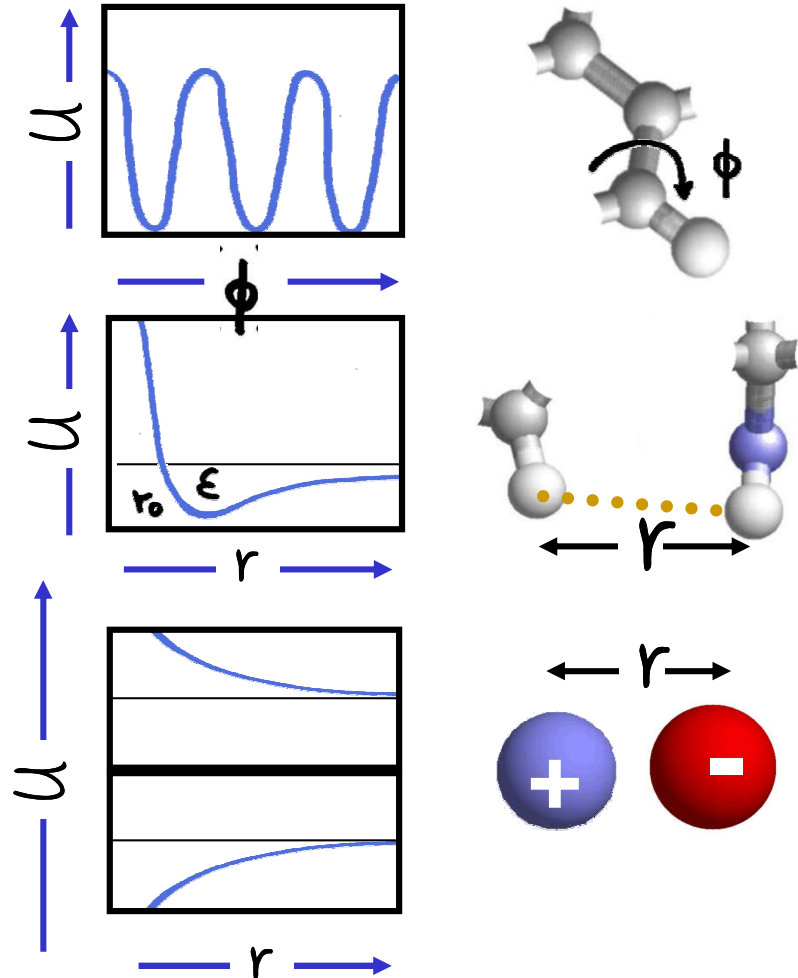
All nonbonded pairs

$$+ \sum 332 q_i q_j / r$$

All partial charges

- A protein with N residues has about 4N (φ, ψ, χ) single bond torsion angles.
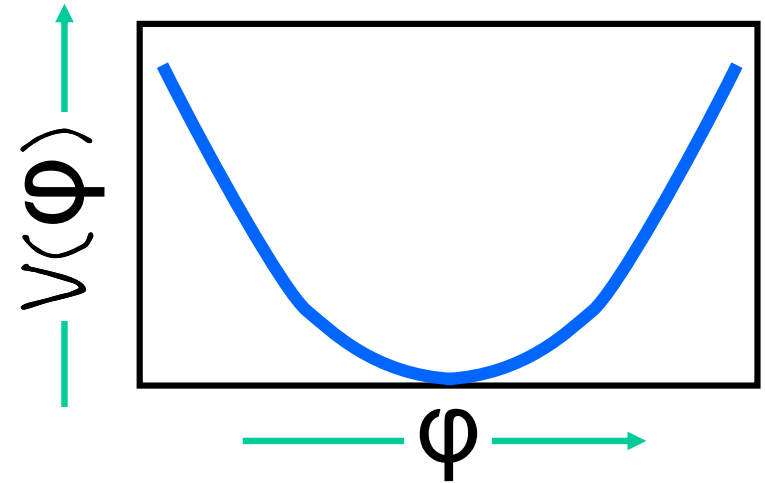- The same protein has about 50N Cartesian coordinates (x, y, z).

THEORY OF NORMAL MODES I

- Assume Potential energy, V, is quadratic function of φ.

$$V = \frac{1}{2} \sum_{i,j} V_{ij} (\phi_i - \phi_i^o)(\phi_j - \phi_j^o)$$

- This means that $V_{ij} = d^2V/d\varphi_i d\varphi_j$

- Assume Kinetic energy, T, is quadratic function of $d\varphi/dt$.

$$T = \frac{1}{2} \sum_{i,j} T_{ij} (d\phi_i/dt)(d\phi_j/dt)$$

Note the symmetry between Potential and Kinetic energy.

- This means that $T_{ij} = d^2T/d(d\varphi_i/dt)\, d(d\varphi_j/dt)$

THEORY OF NORMAL MODES II

- Solve for $\varphi(t)$ using Lagrangian approach.

$$\sum T_{ij}(d^2\varphi_j/dt^2) = \sum V_{ij}\Delta\varphi_j$$

Analogous to $ma = F$

- Try a periodic function for $\varphi(t)$:

$$\Delta\varphi_j(t) = \sum A_{ij}\cos(\omega_i t)$$

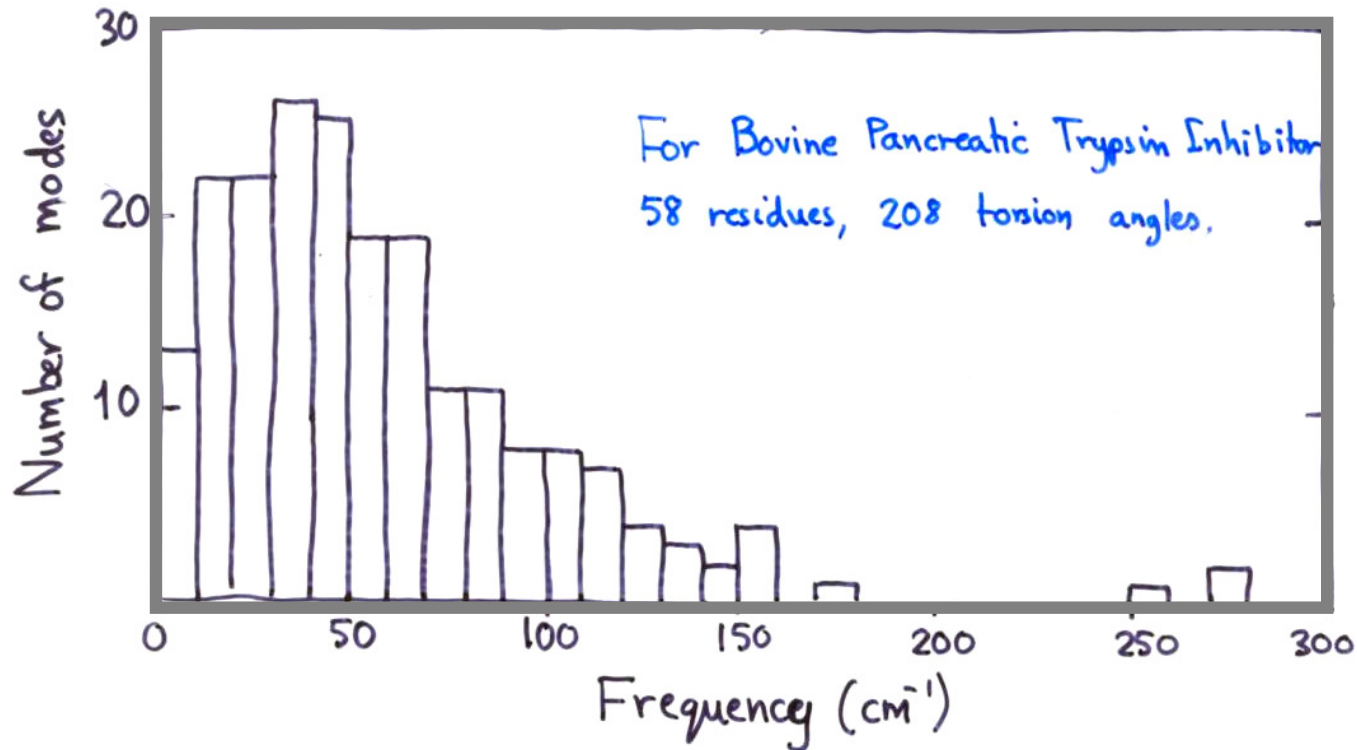$$d^2\varphi_j(t)/dt^2 = \sum A_{ij}\,\omega_i^2\cos(\omega_i t)$$

- In Matrix notation the Lagrangian equation is:

$$TA\omega^2 = VA$$

This is Eigenvalue equation that is easily solved.

Suggestive Motion
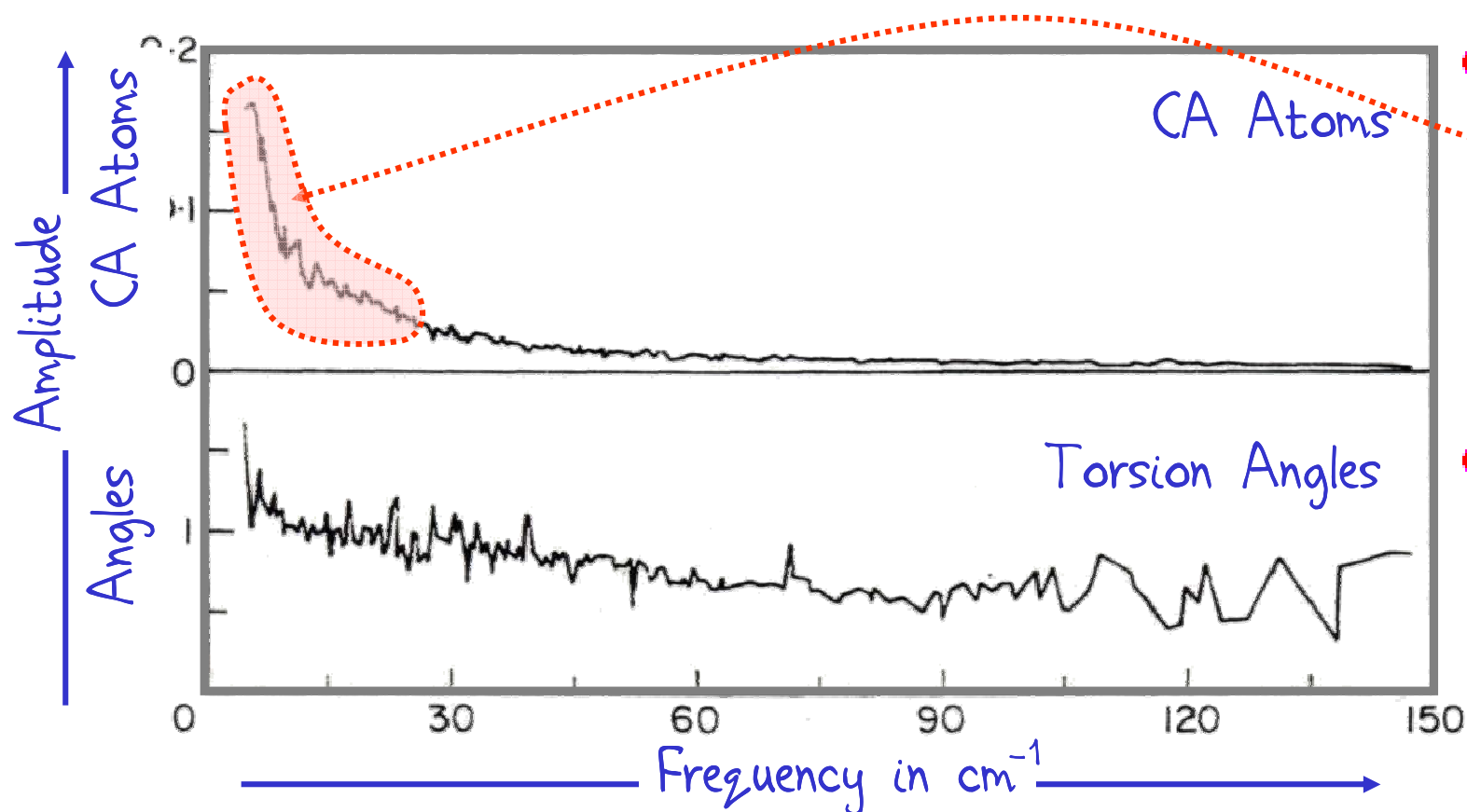
RATES OF VIBRATION

For Bovine Pancreatic Trypsin Inhibitor 58 residues, 208 torsion angles.

- There is a broad range of torsion angle mode frequencies.

- Peak near 30 cm^{-1}, which is a period of 1 ps.

- Lowest frequency is at 3 cm^{-1} or 10 ps.

- There are 12 modes below 10 cm^{-1}.

Levitt et al. J. Mol. Biol. (1985)

AMPLITUDES OF VIBRATION

CA Atoms

Torsion Angles

Almost all the motion of the CA atoms comes from the lowest frequency modes.

There is high-frequency motion of the torsion angles.

- The CA Amplitude is the RMS movement of all CA atoms as a result of activating the particular mode.

- The Torsion Angle Amplitude is the RMS movement of all torsion angles as a result of activating the particular mode.
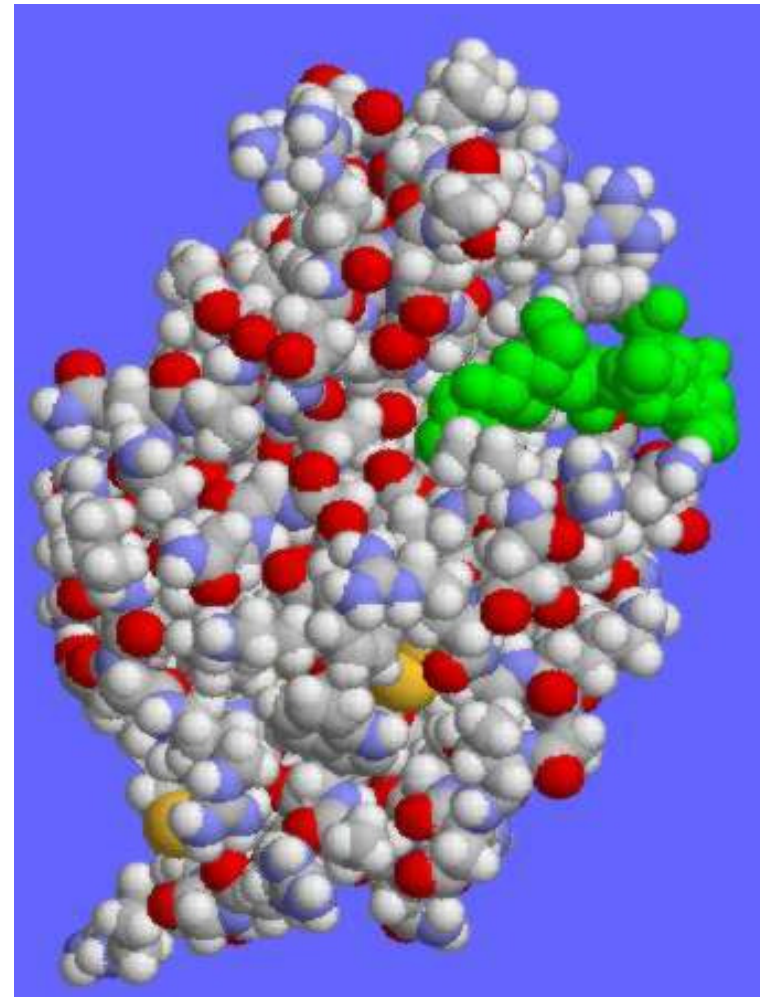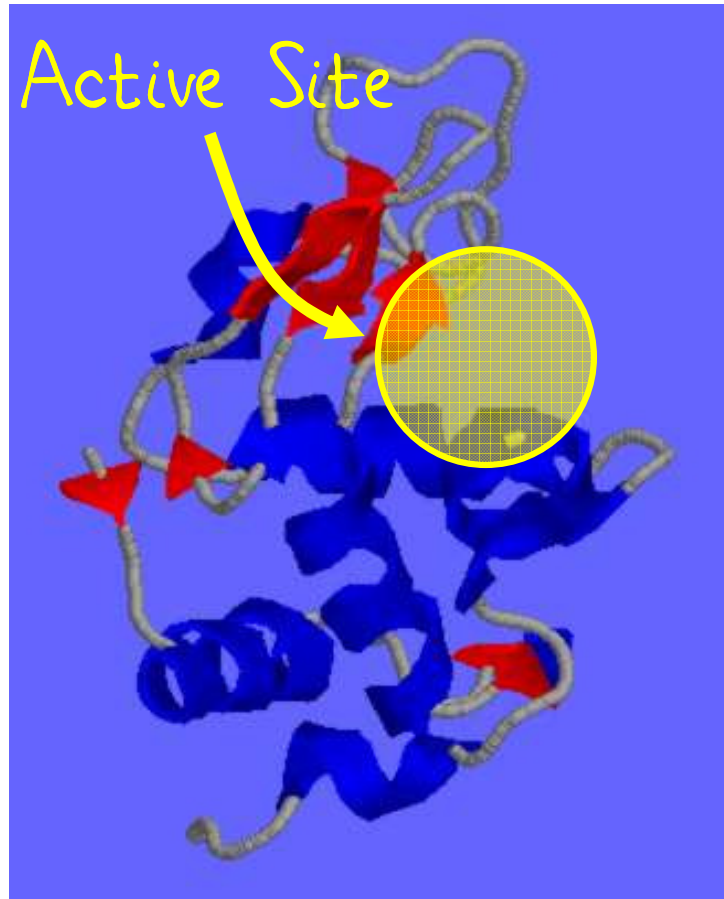
BPTI NORMAL MODES AT HIGH TEMPERATURE

Bovine Pancreatic
Trypsin Inhibitor

Three of the
four lowest modes,

the thirtieth mode

and a combination of
the eight lowest modes

LYSOZYME MODES

Active Site

An inhibitor, which is colored in green, is bound in the active site.
The inhibitor is not included in the normal mode calculations.

Hen Egg White Lysozyme

The two lowest modes

Knowledge-Based Potentials

LONG HISTORY

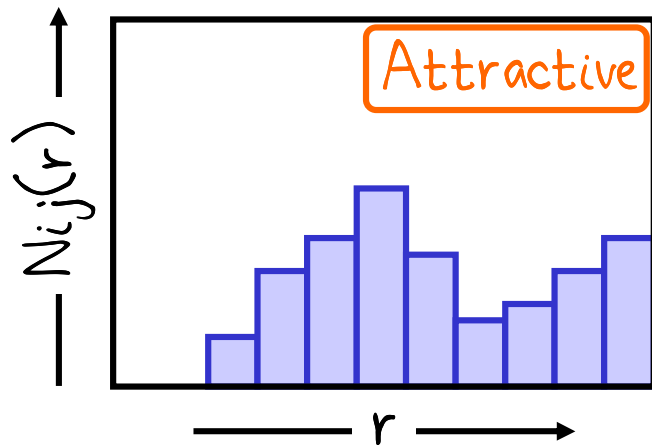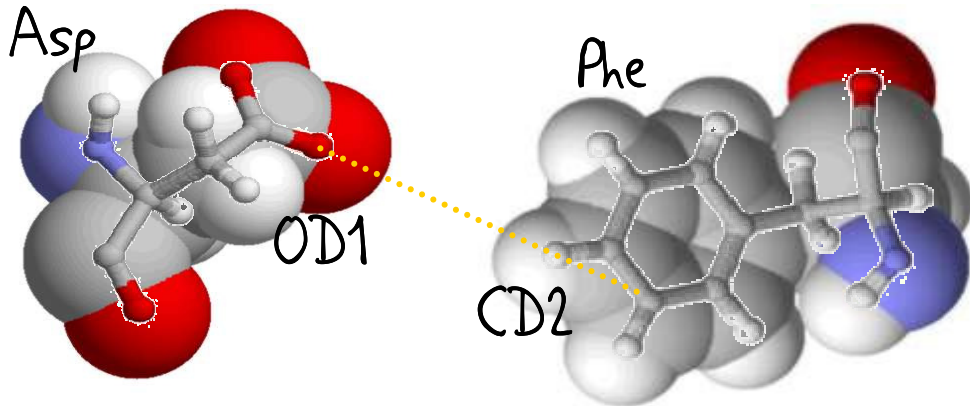| 1976 Tanaka & Scheraga	Contacts, 20 types
1985 Miyazawa & Jernigan	Contacts, 20 types
1990 Sippl	Distance-Dependent $^{20\ types}$ (DD)
1998 Samudrala & Moult	DD, 167 types
2007 Summa & Levitt	DD, 167, Continuous

POWERED BY PDB GROWTH

1976:Tanaka & Scheraga (10)   1985:Miyazawa & Jernigan (100)

1990:Sippl (120)   1998:Samudrala & Moult (1200)

2007:Summa & Levitt (10000)

BASIC IDEA

Asp

OD1

CD2

Phe

Attractive

$N_{ij}(r)$

r

$E_{ij}(r)$

r

Repulsive

$N_{ij}(r)$

r

$E_{ij}(r)$

r
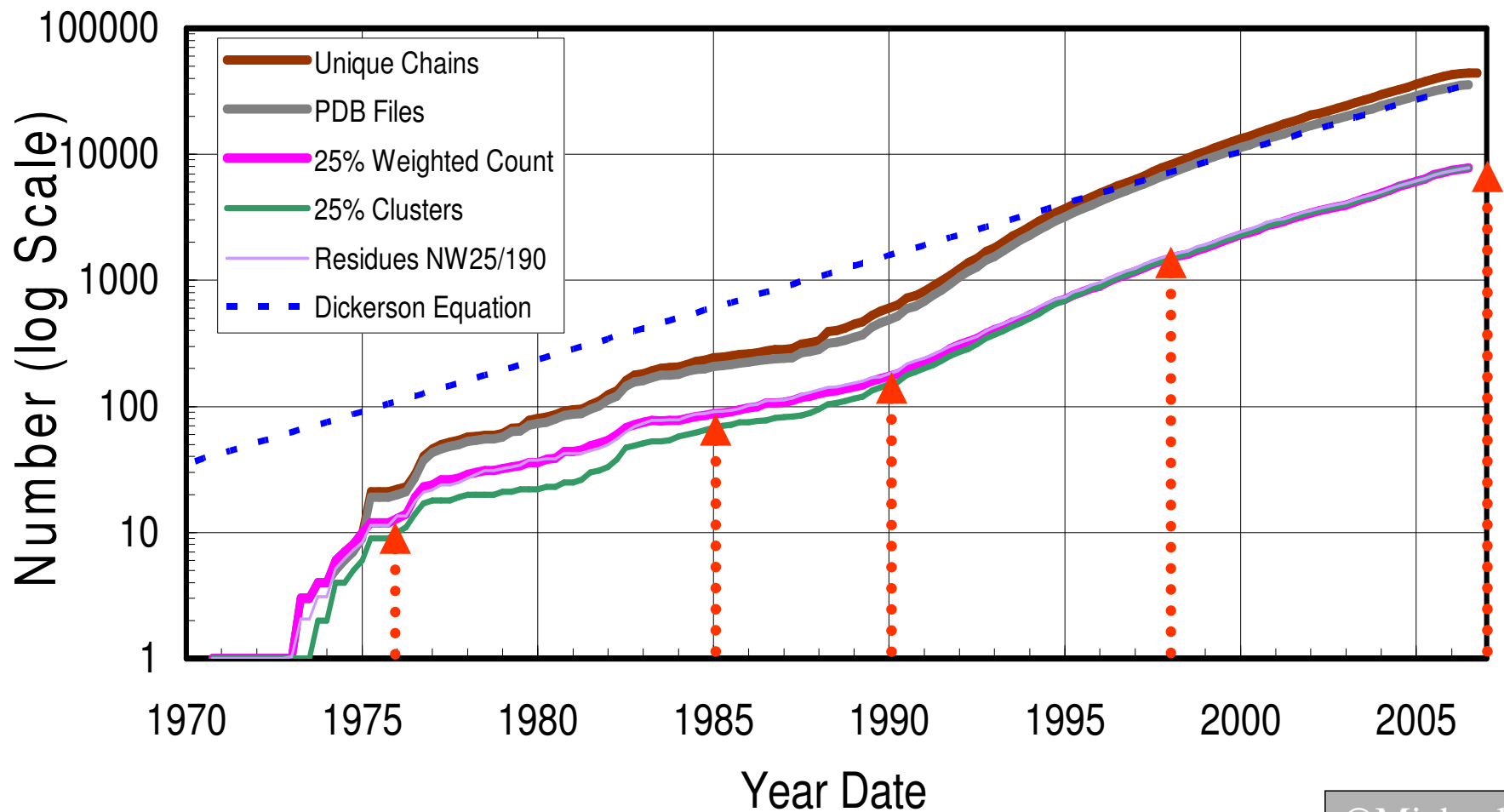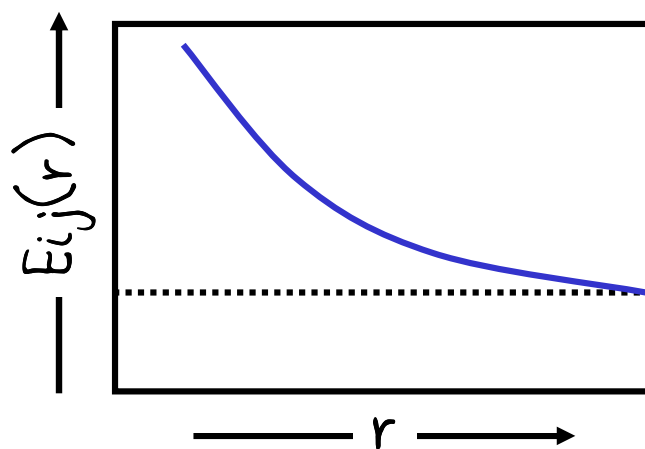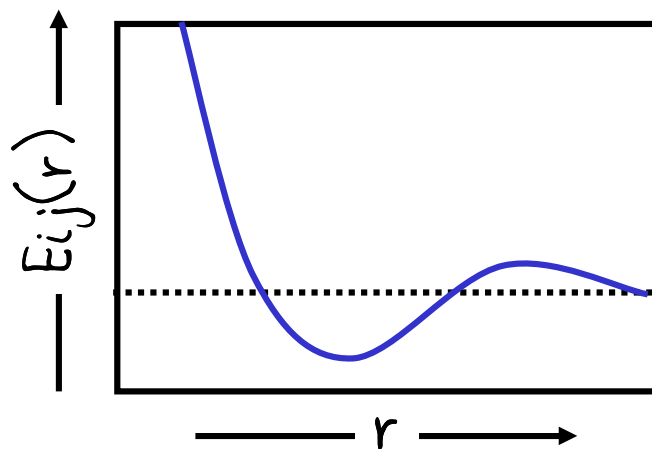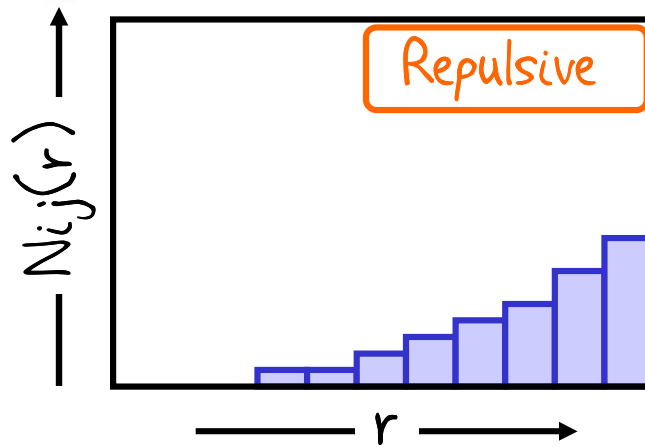
- Get distribution of distances between pairs of atom centers of a particular type, e.g. D-OD1...F-CD2.

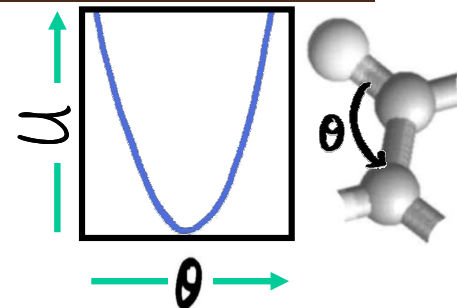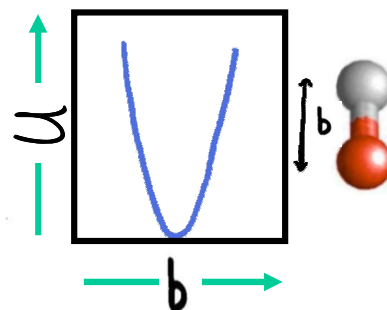- Normalize and take log to get Energy score: $E_{ij}(r) = \log(N(r_{ij})/M(r_{ij}))$

Refine Proteins

Chris Summa

KNOWLEDGE-BASED POTENTIAL ENERGY

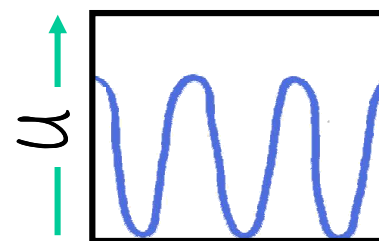$$U = \sum \frac{1}{2} K_b (b-b_0)^2 - \sum \frac{1}{2} K_\theta (\theta-\theta_0)^2$$
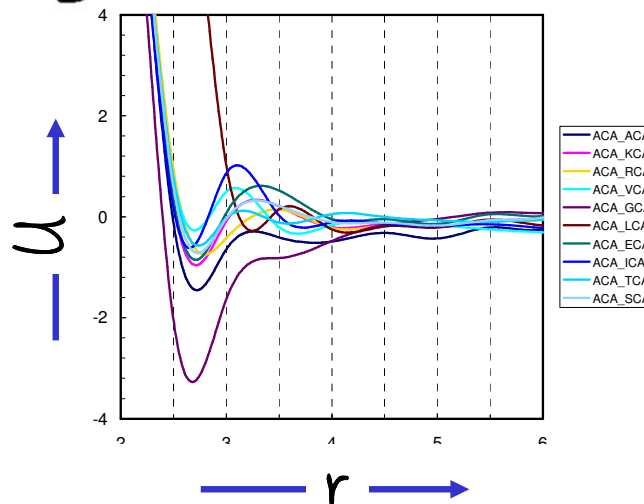
All Bonds   All Angles

$$+ \sum K_\phi \left[1 - \cos(n\phi + \delta) \right]$$
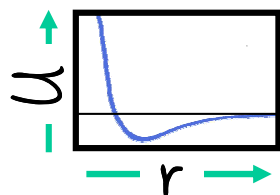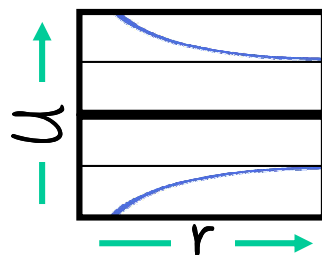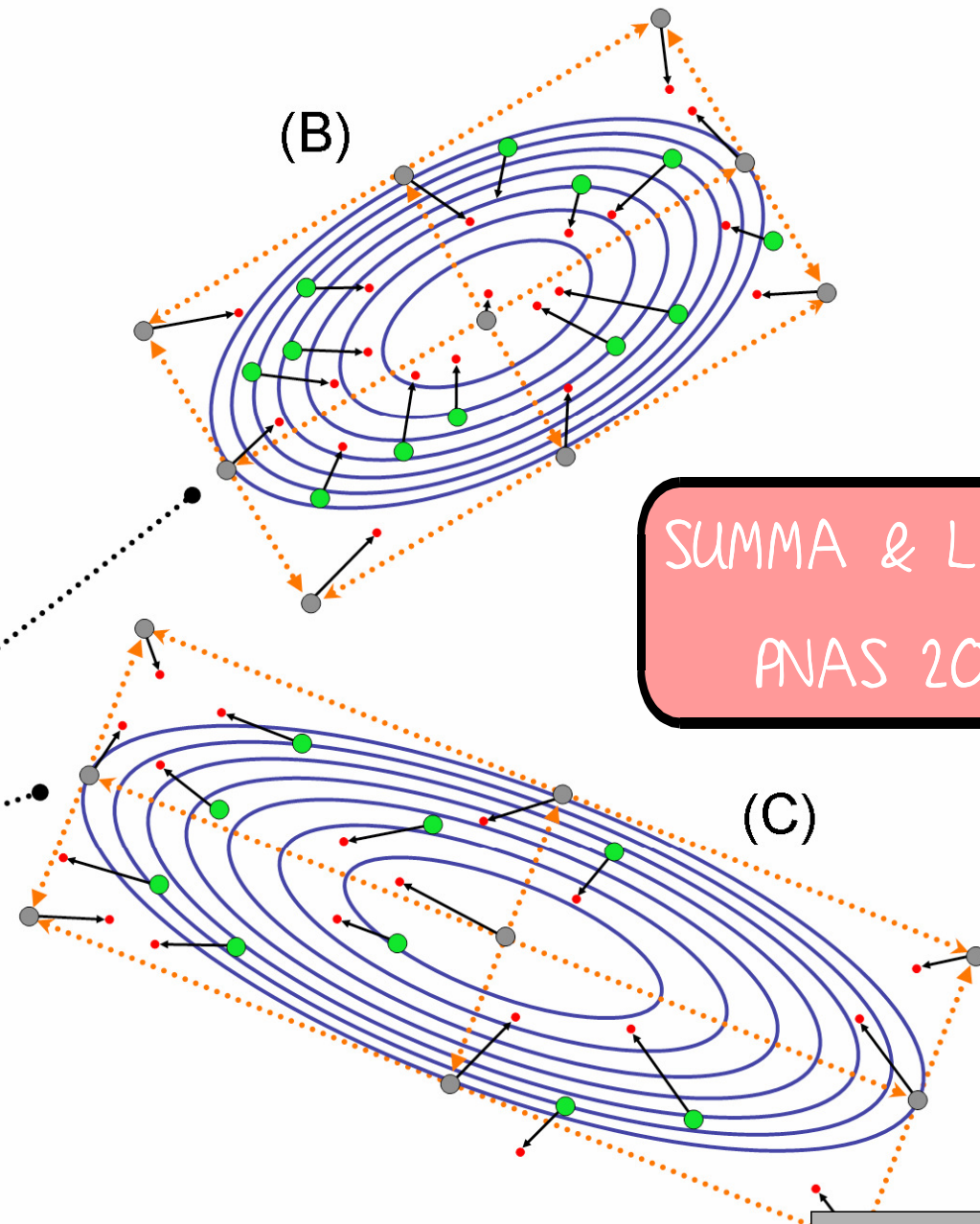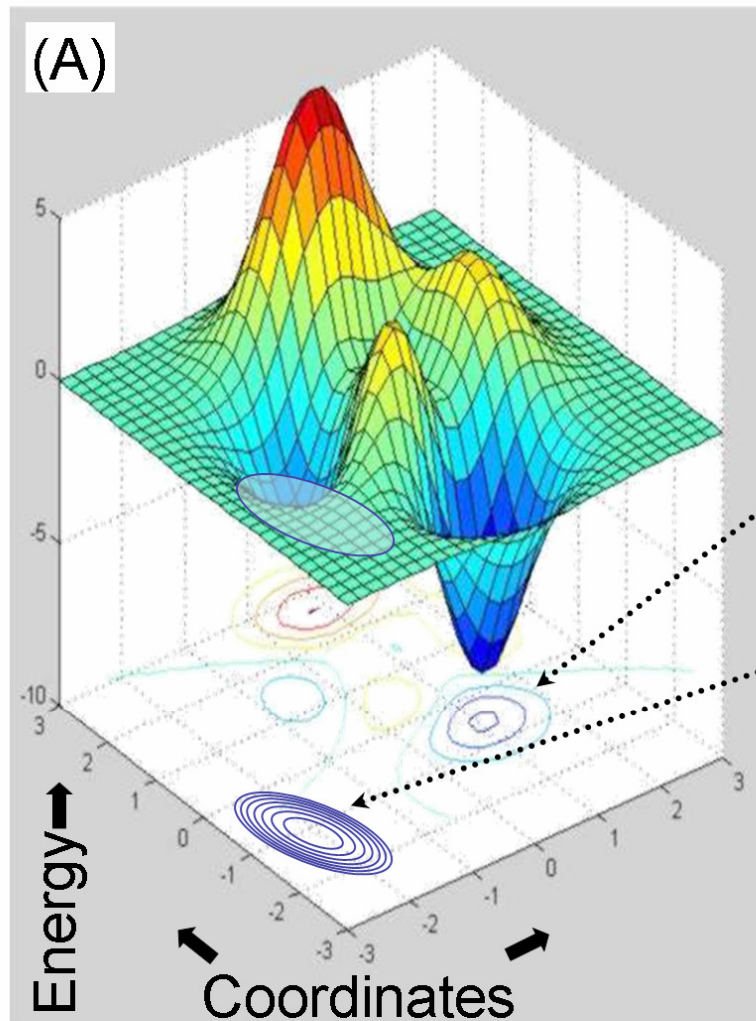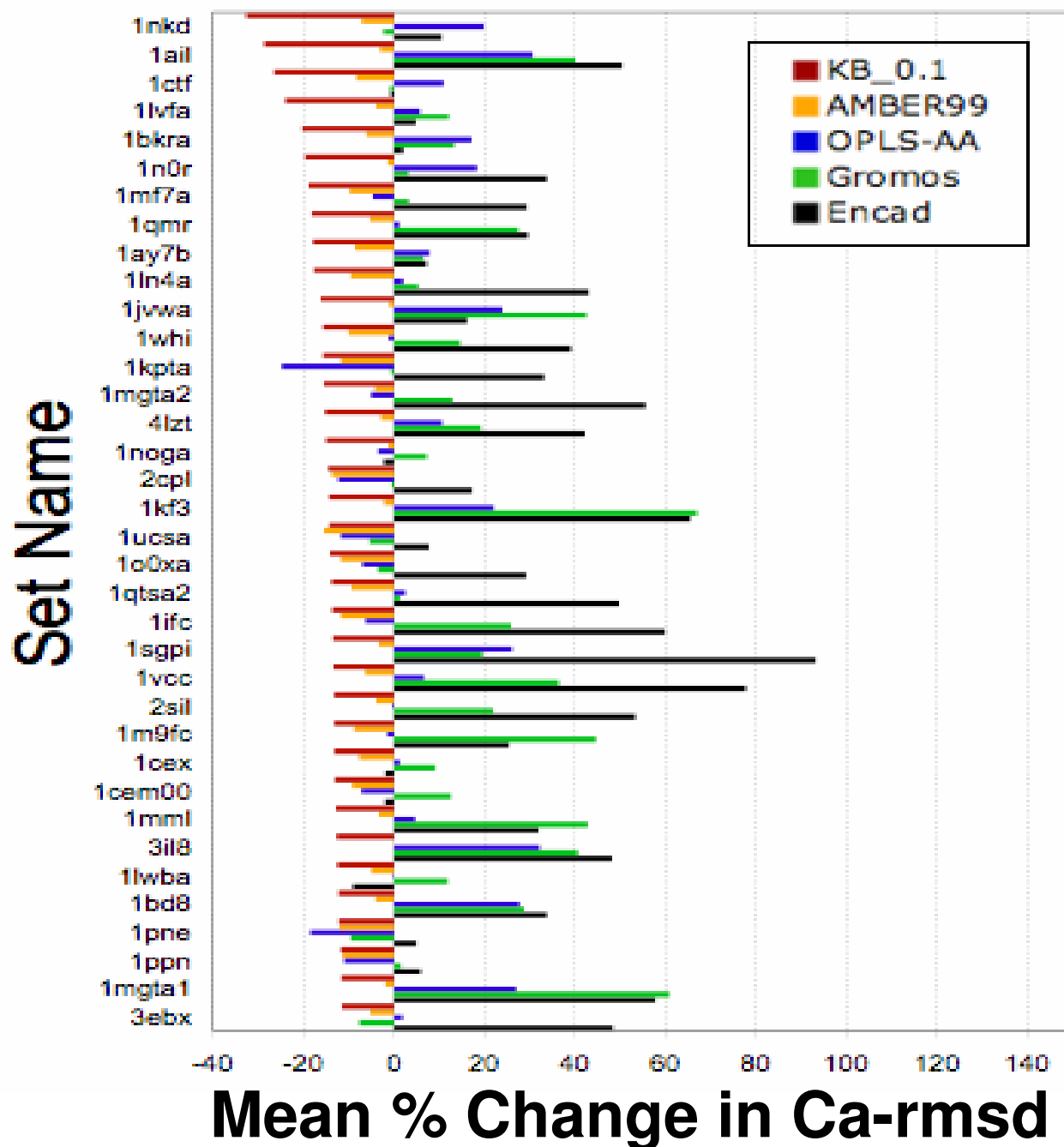
All Torsion Angles

$+$ Quintic Splines

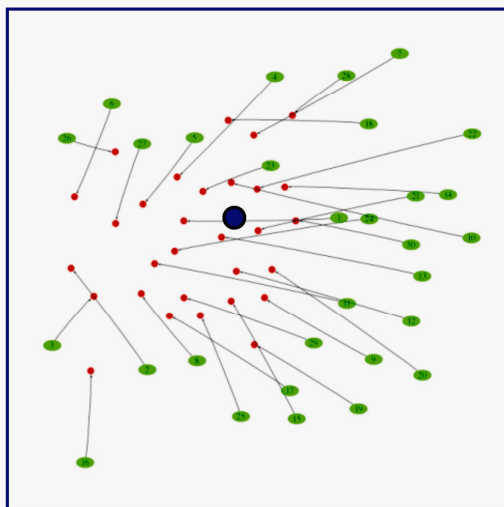All nonbonded pairs

Replace and with lookup tables

CARTOON OF REFINEMENT TEST

(A)

Energy

Coordinates

(B)

(C)

SUMMA & LEVITT

PNAS 2007

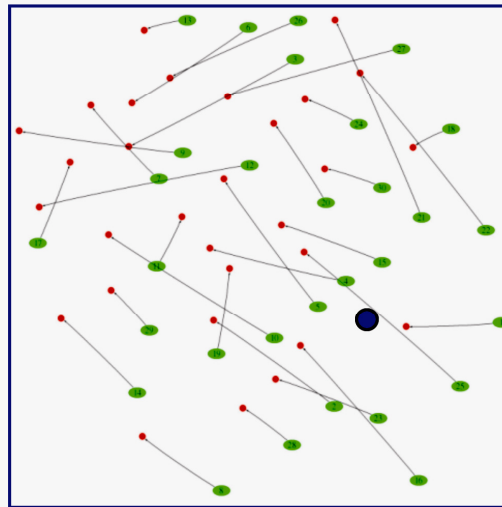# SUCCEEDS IN REFINING PROTEINS
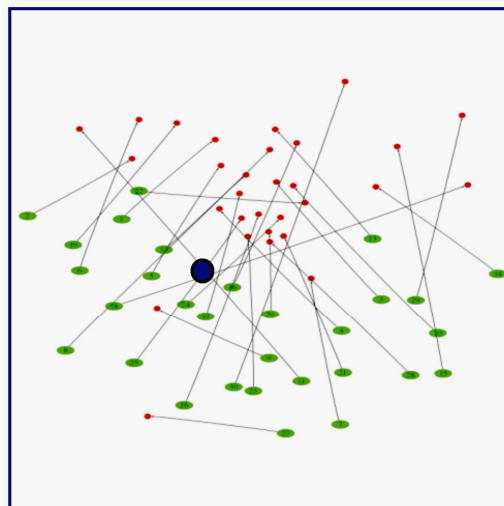

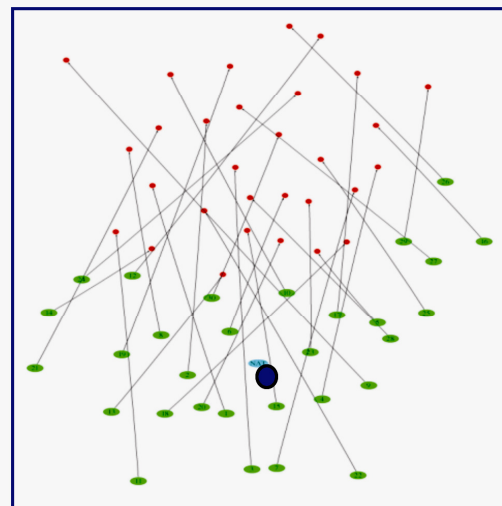**Mean % Change in Ca-rmsd**

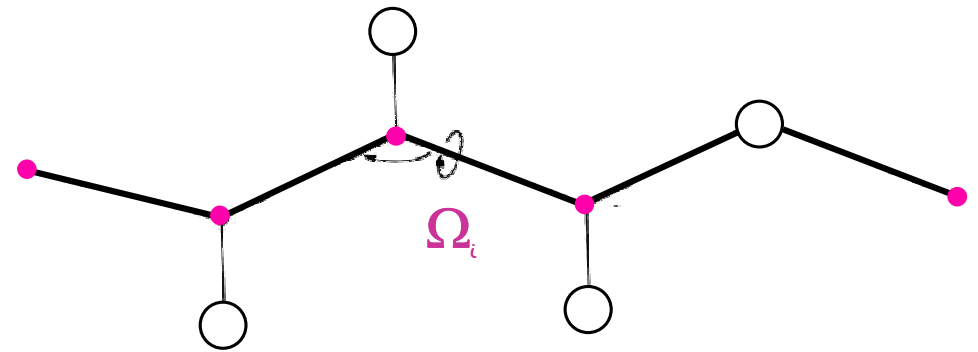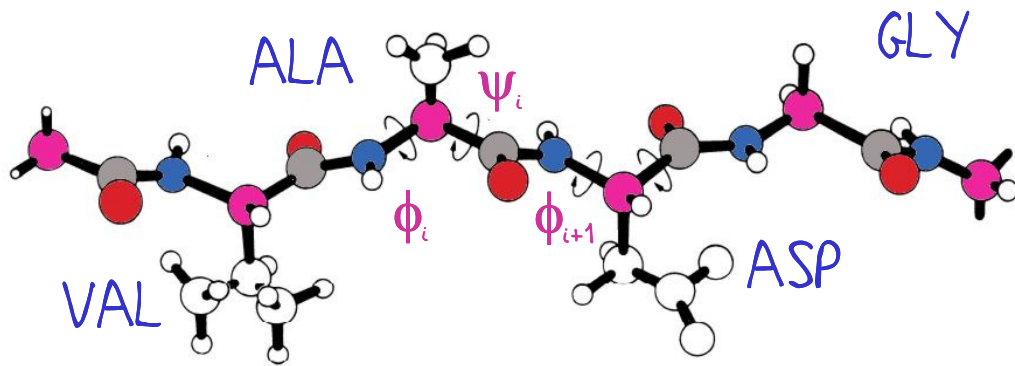# SHIFTS MAP ENERGY SURFACE

1ctf KB_0.1   1pdo KB_0.1

1ctf Encad   1pdo Encad

Reduced Models

NEEDED THEN, MORE USEFUL NOW

ALA

VAL

ψ_i

ϕ_i ϕ_{i+1}

GLY

ASP

Ω_i

1 atom/residue, 1 DOF/residue
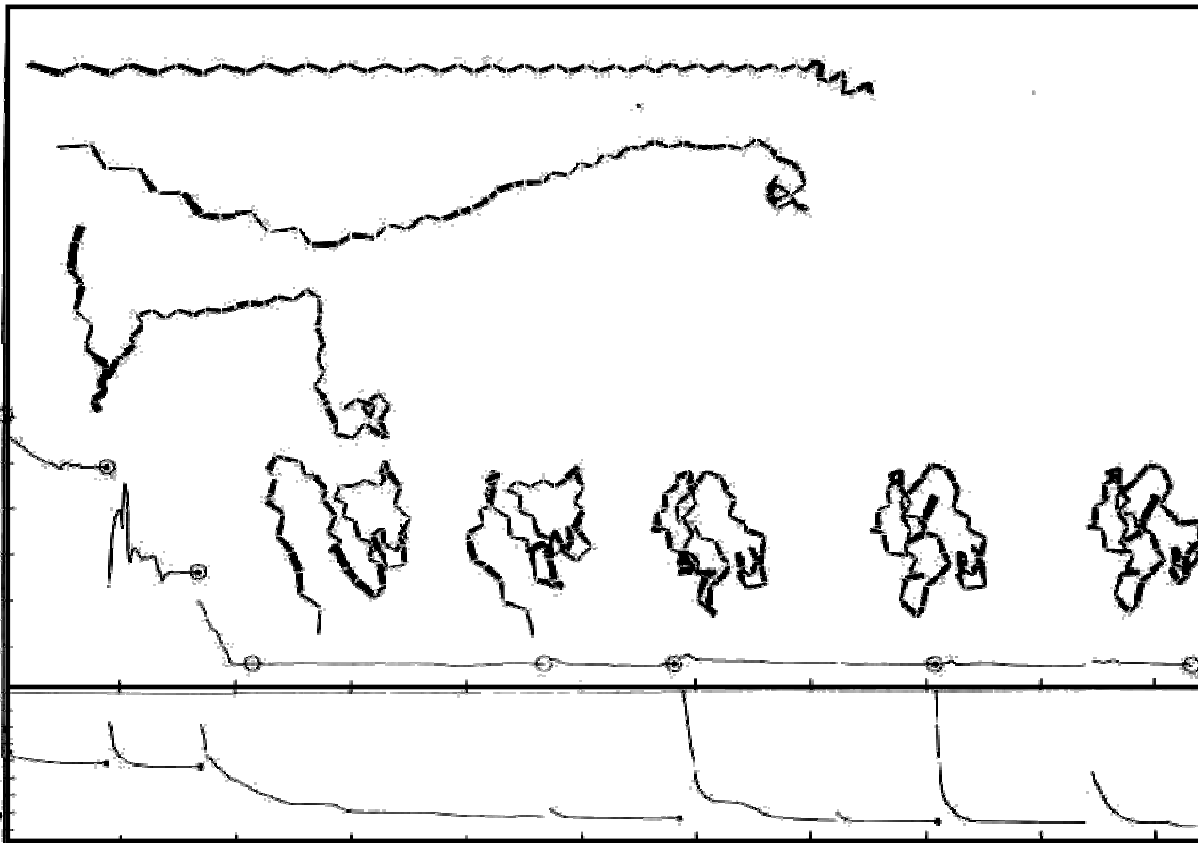
Fold protein with 1000 steps of minimization.
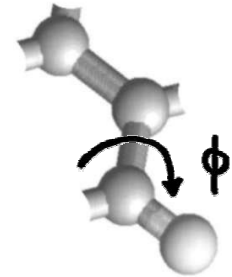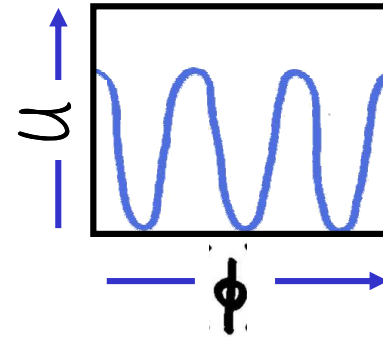
Escape from local minima with normal modes jumps.

Levitt & Warshel
Nature, 1975

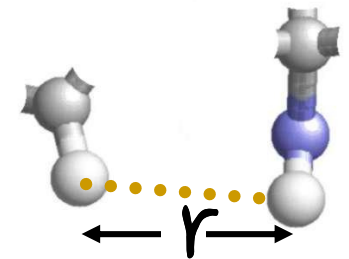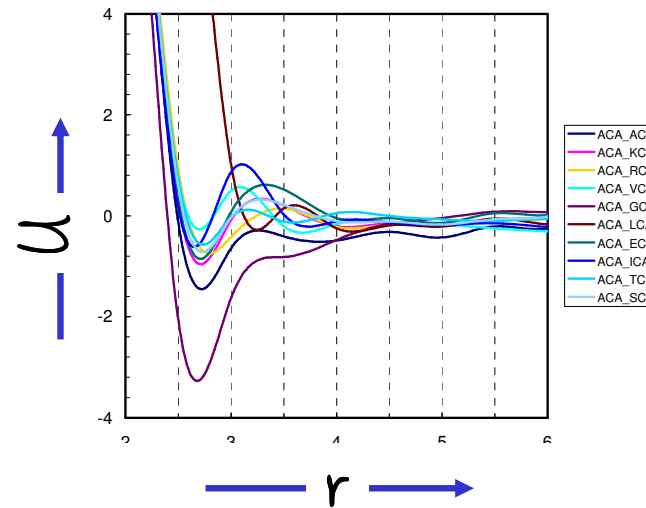# KNOWLEDGE-BASED ENERGY IN TORSION SPACE

$$U = \sum K_\phi [1 - \cos(n\phi + \delta)]$$

All Torsion Angles

+ Quintic Splines

All nonbonded pairs
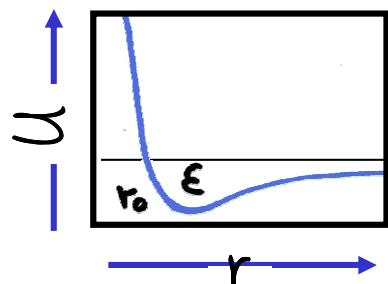
Replace with lookup tables

Myosin Motor

MYOSIN STRUCTURE

regulatory light chain

neck

essential light chain

converter

lower 50kDa

actin binding site

ATP binding site

cleft

N terminal

upper 50kDa

S1 fragment

Head

ATP

Essential light chain

Neck

Regulatory light chain

Tail

50 nm

(b)

Copyright 1999 John Wiley and Sons, Inc. All rights reserved.
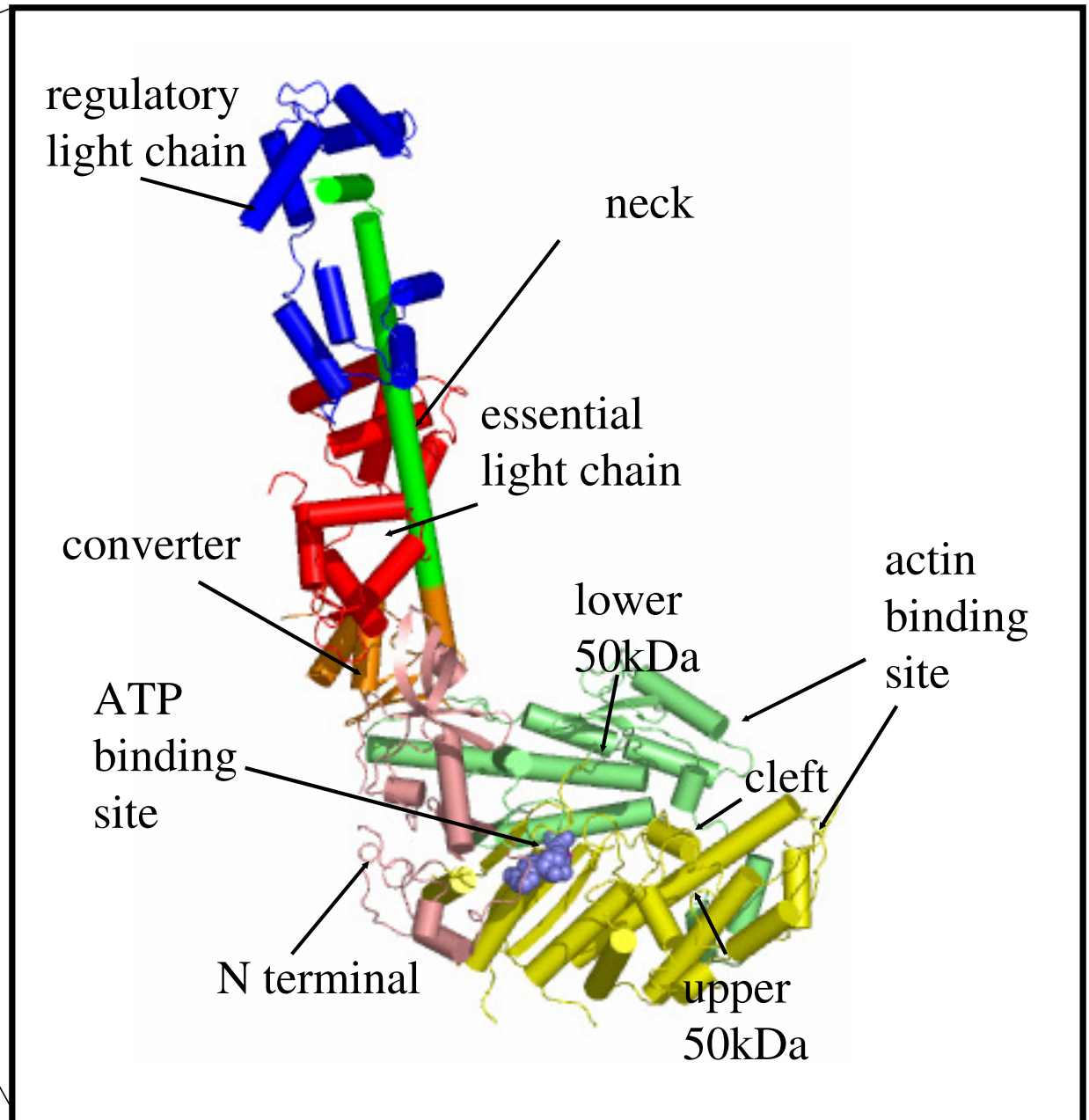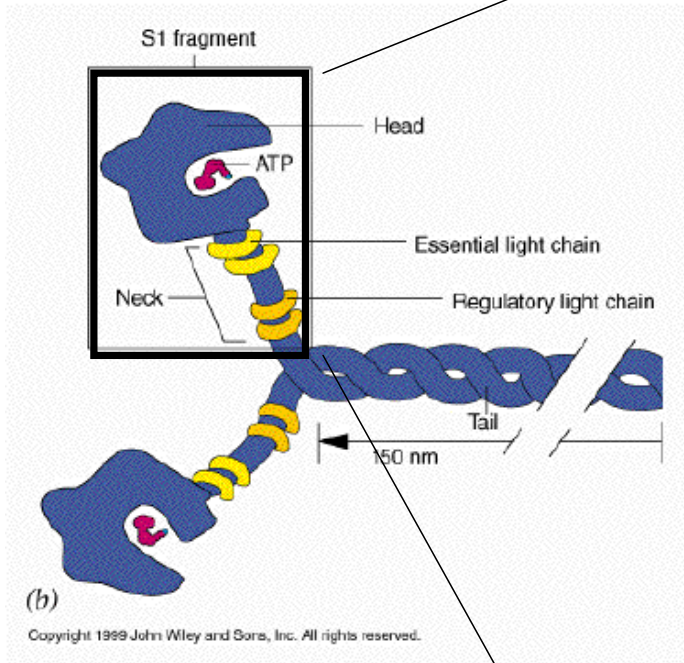
MORPHING MYOSIN

Dahlia Weiss

MYOSIN ALL-ATOM NORMAL MODE 1

MYOSIN 3PT NORMAL MODE 1

MYOSIN 1PT NORMAL MODE 1

Cellular Machinery

THE CORE MACHINERY OF LIFE
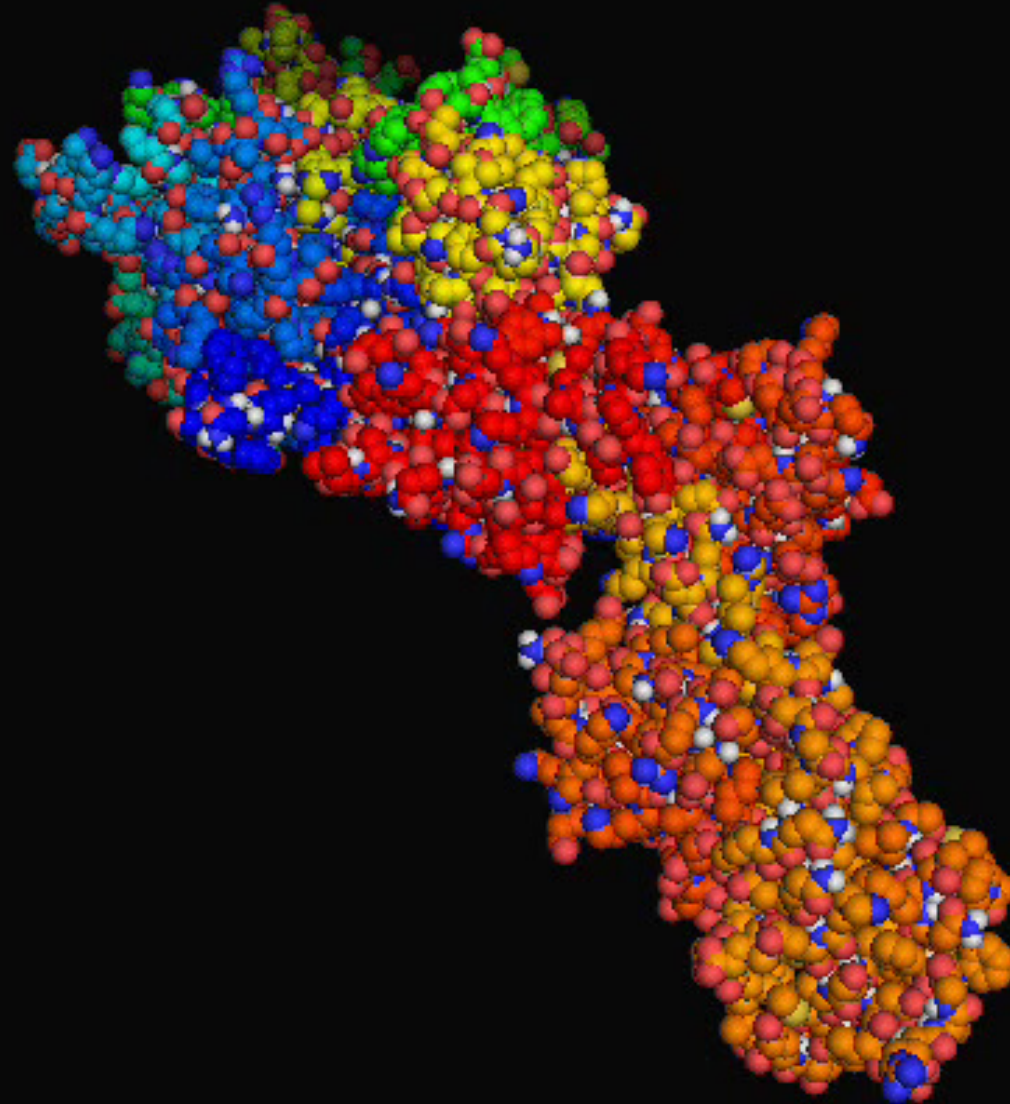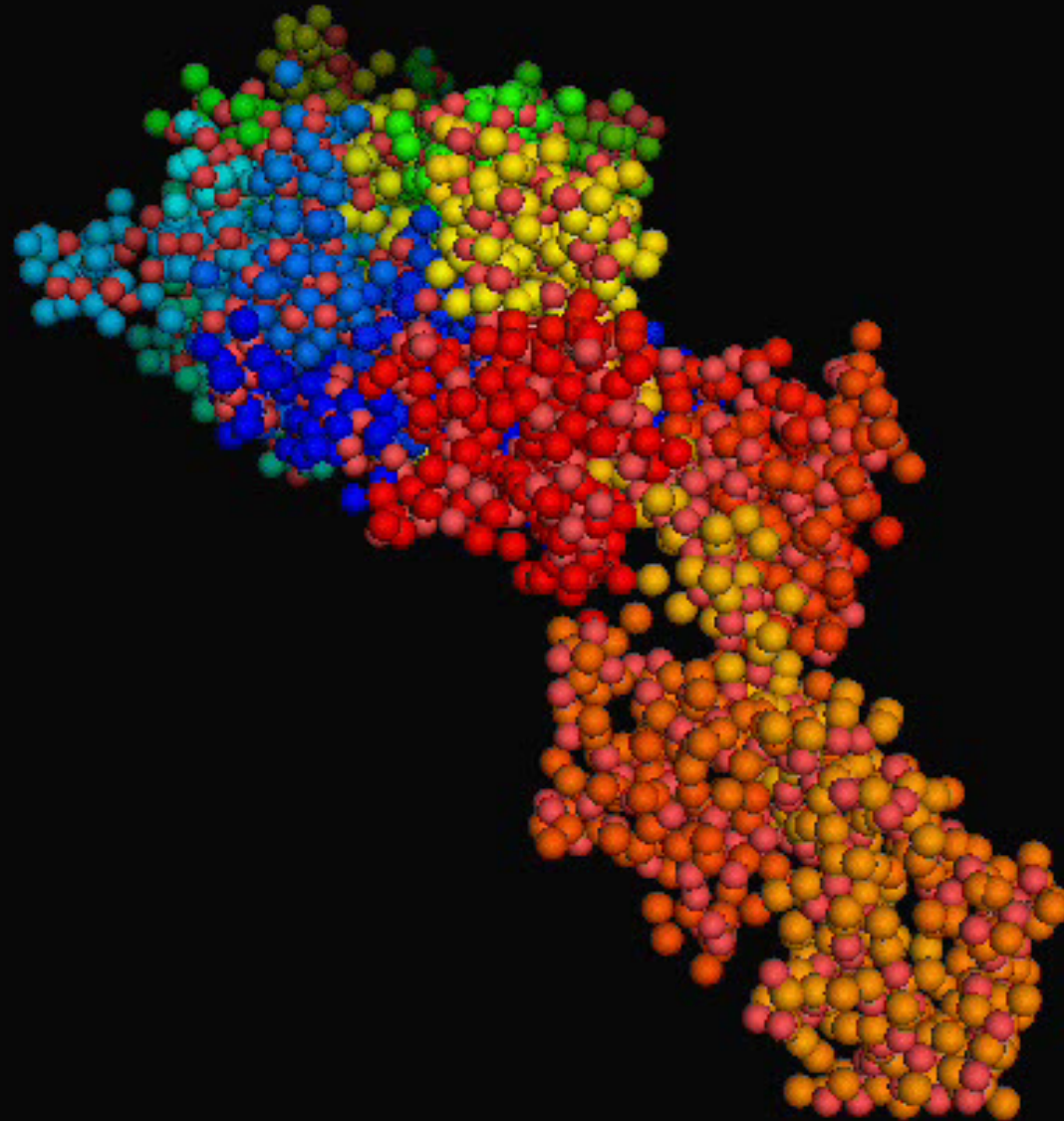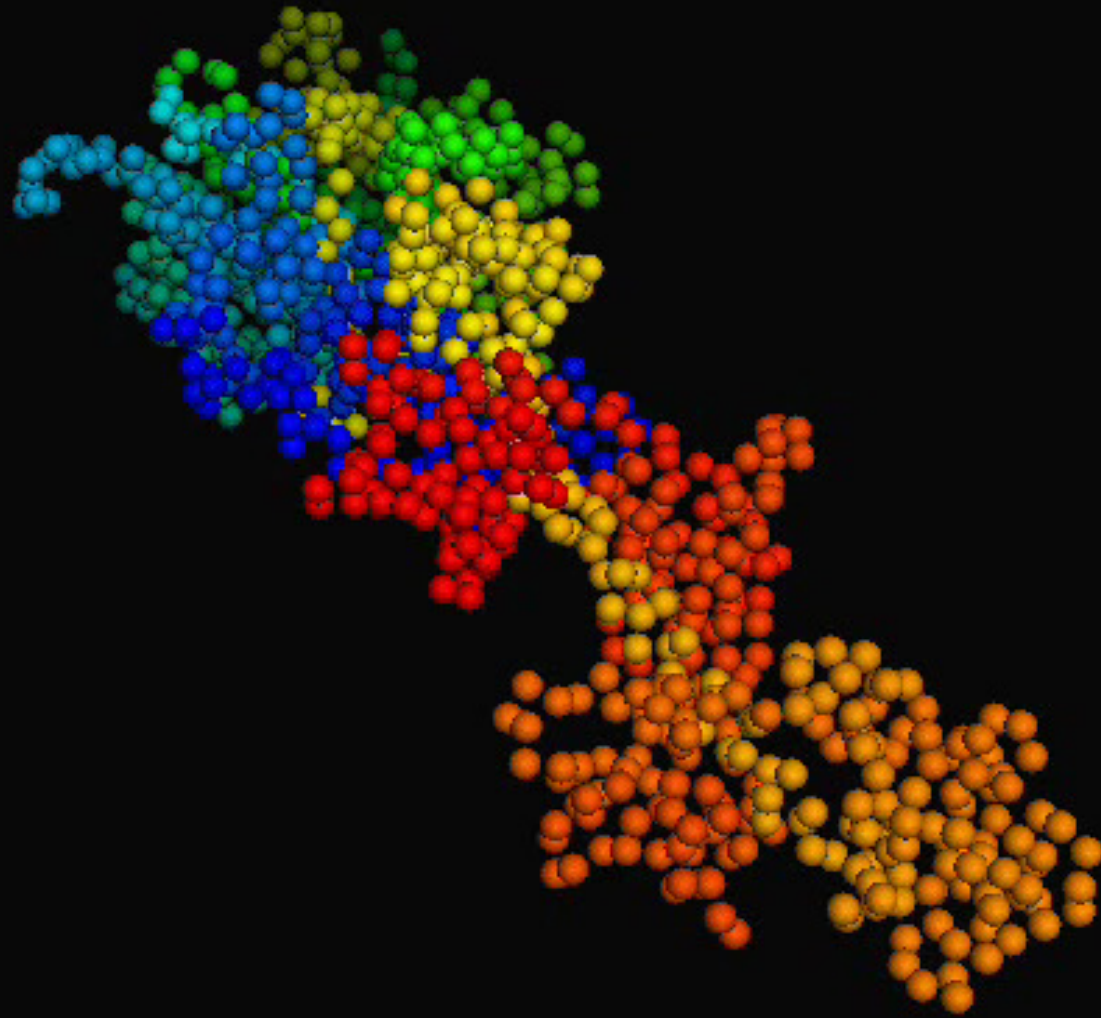
Package DNA

DNA

Polymerase

RNA Polymerase

Nucleotides

Make RNA

RNA

Fold Protein

Protein

Amino acid

mRNA

Ribosome

Make Protein
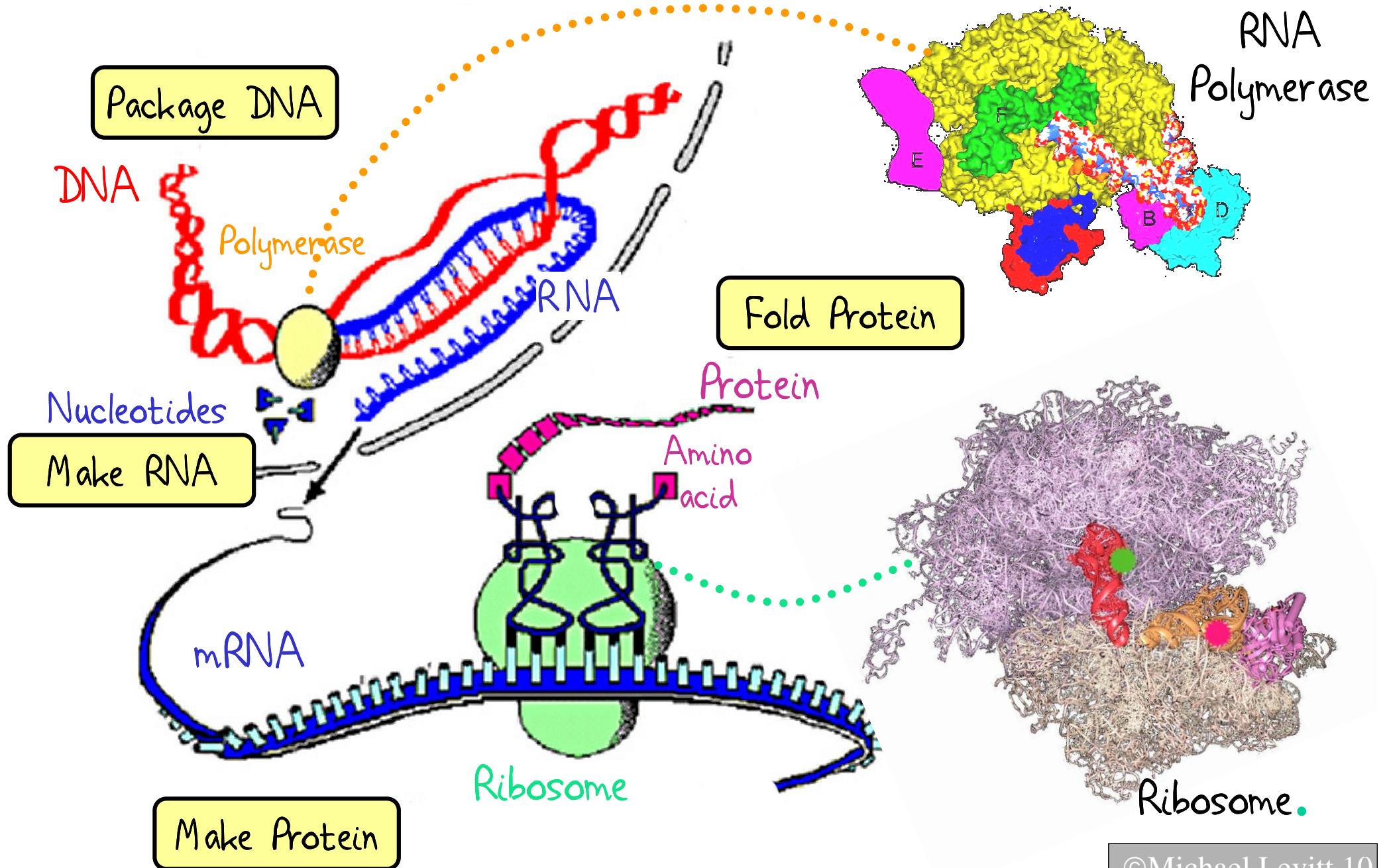
Ribosome.

NUCLEOSOME NORMAL MODE 2
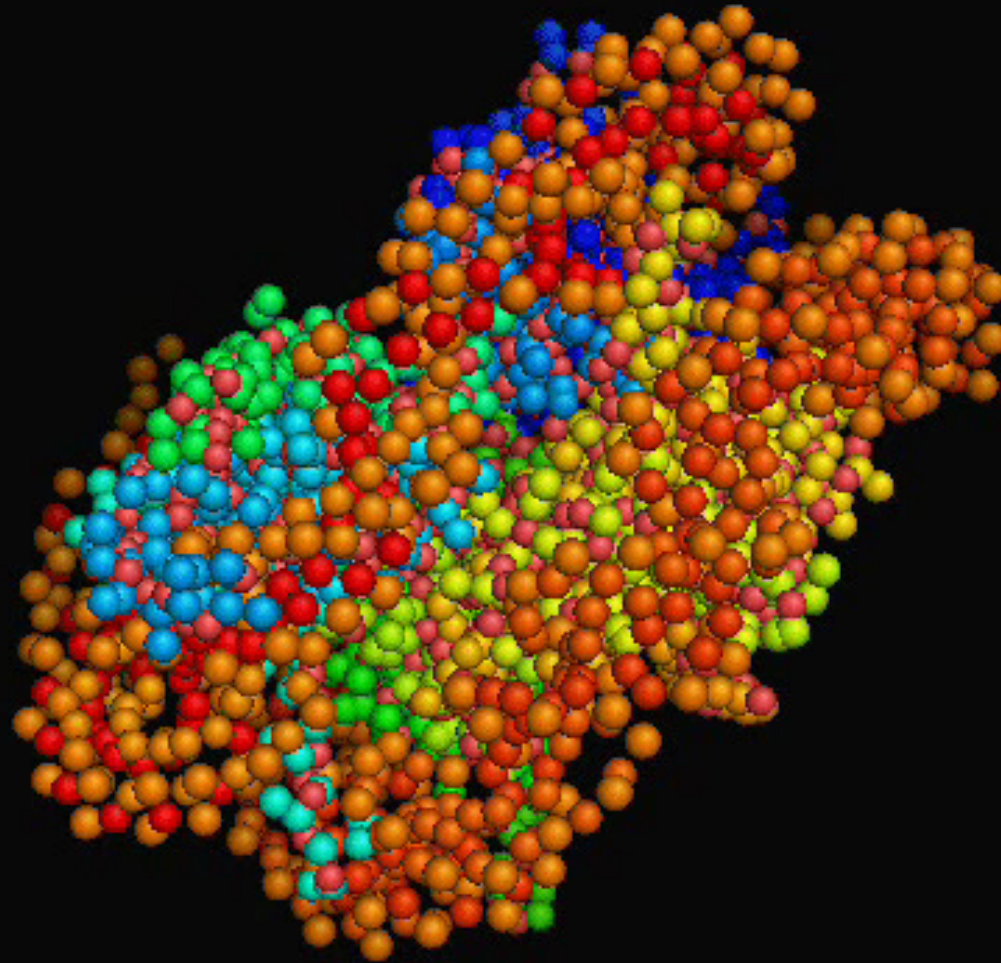
70S RIBOSOME NORMAL MODE 2
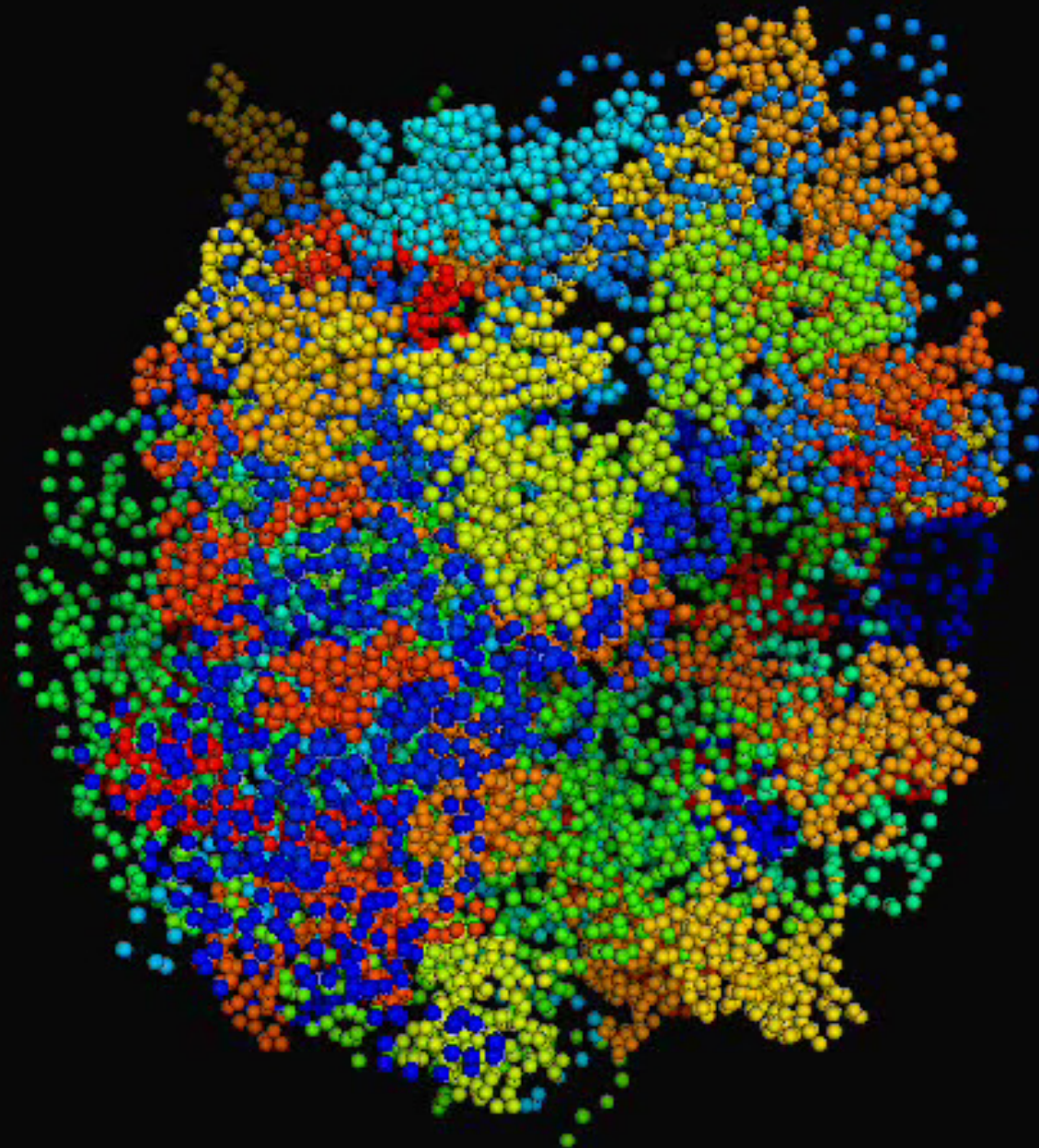
Chaperonin Bac NORMAL MODE 2
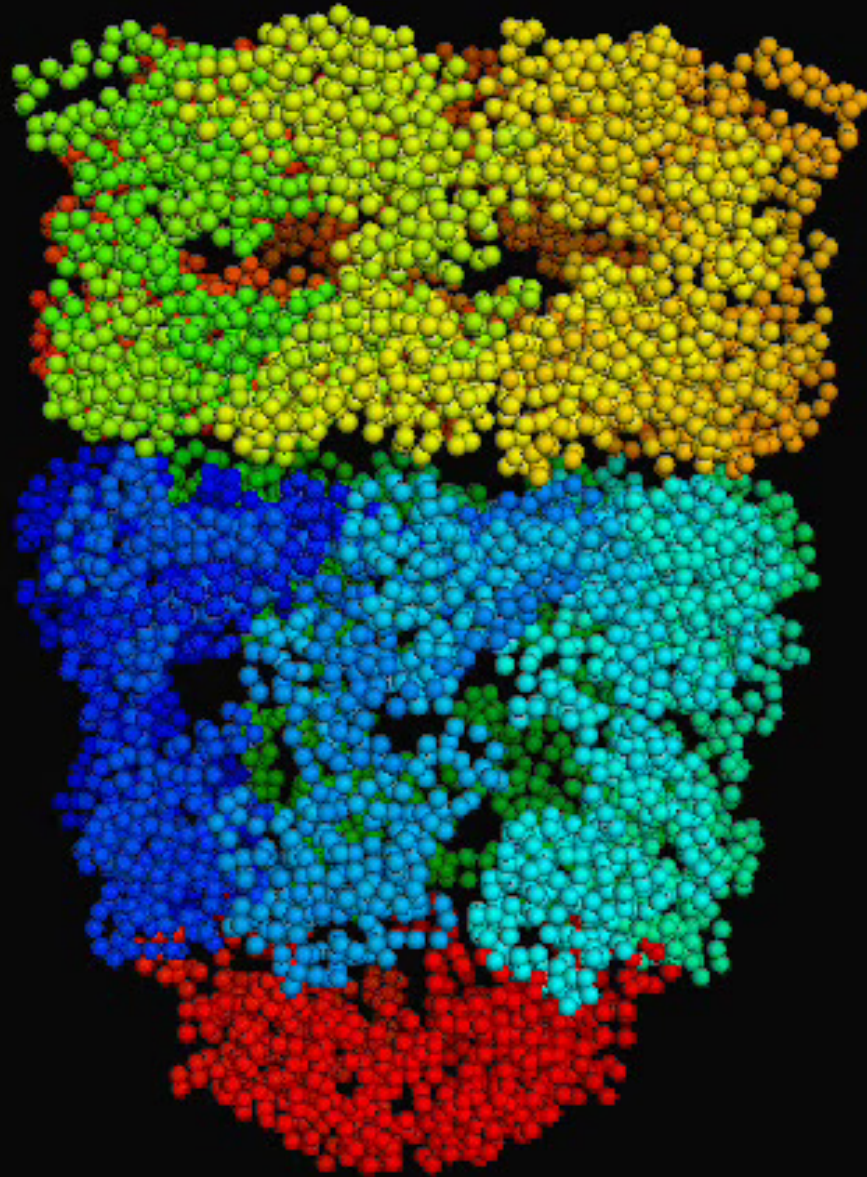
Chaperonin Eu NORMAL MODE 1
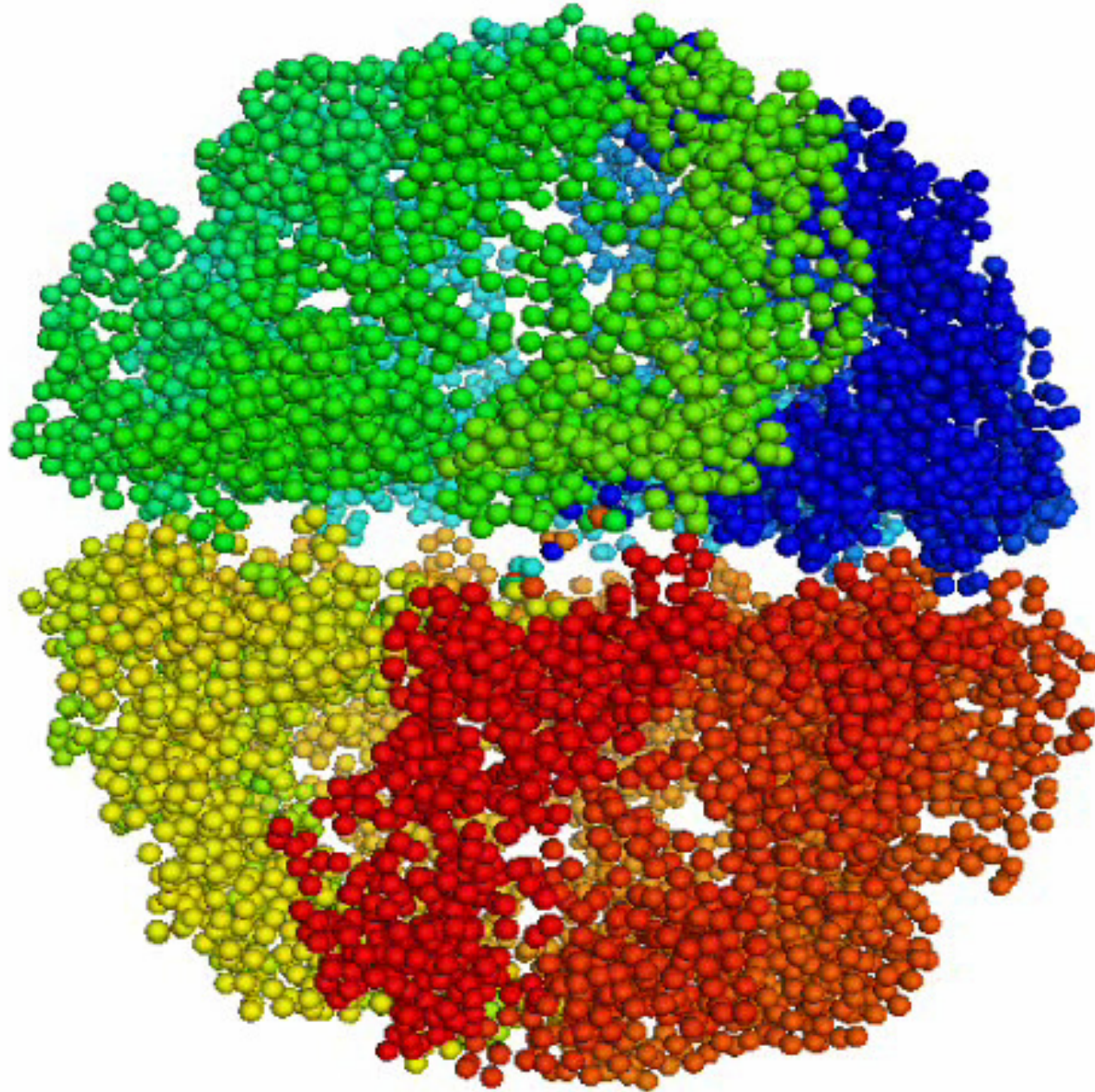
ALL 1PT NORMAL MODE RUNS SO FAR

PDB_ID	NRES	%Prot	RMS(Å)
1CTF	74	100.0	1.08
1VBX	176	56.8	1.07
1QA6	250	53.6	1.79
1MFQ	364	65.1	3.82
1M5K	424	47.2	2.51
1AOI	1128	74.1	1.94
1KK7	1149	100.0	3.98
1SFO	4197	99.4	2.48
1Q3S	8768	100.0	2.21
7OSR	10517	58.0	2.40

LINEAR CPU REQUIREMENTS
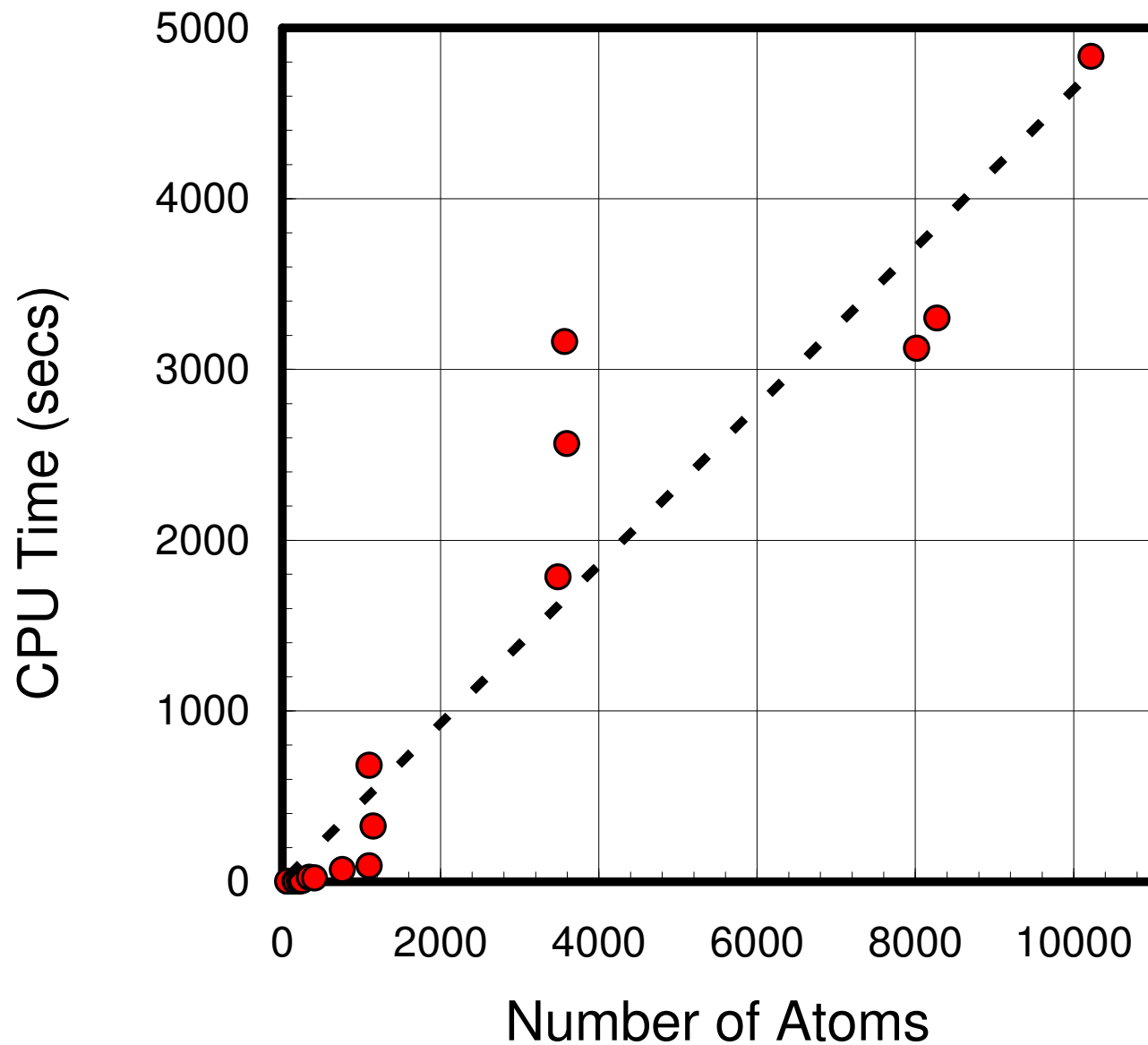
Problems

RNA POLYMERASE NORMAL MODE 1
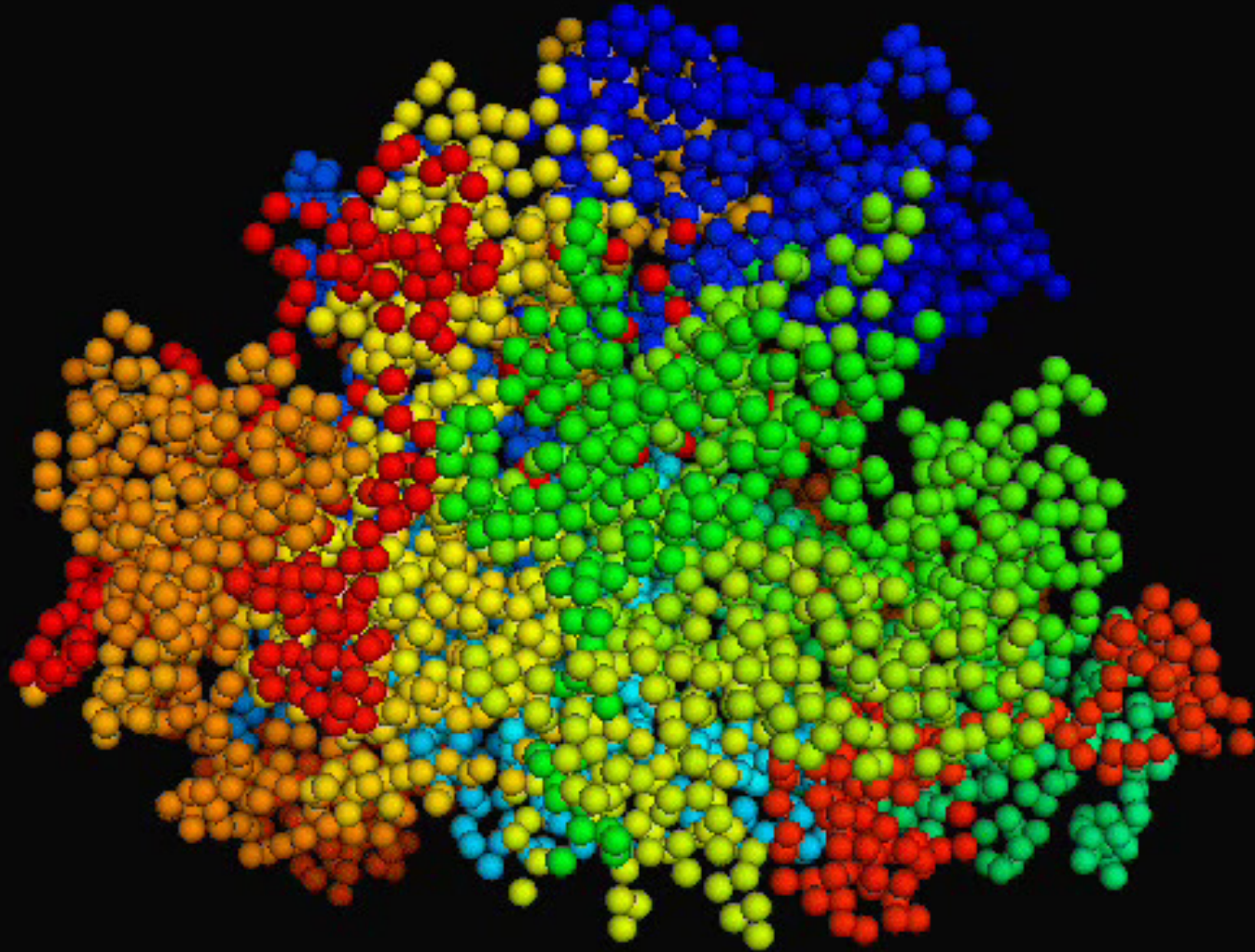
USE GAPS TO MAKE ONE CHAIN

Start

End

1

2

3

4
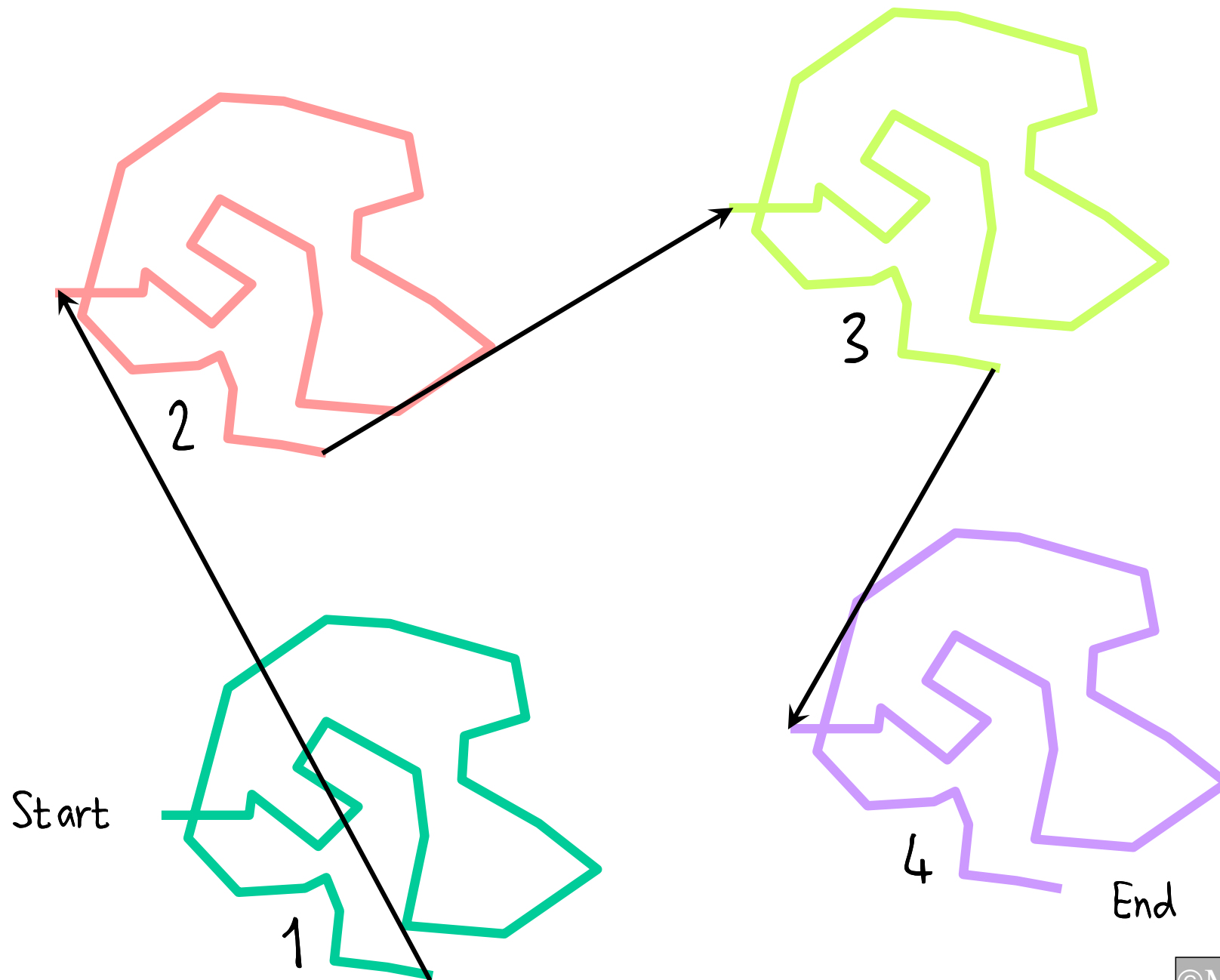
USE ARM TO LINK TO FIXED POINT

t_x	1 - 2
t_y	2 - 3
t_z	3 - 4
θ_x	5 − 6 − 7 - 8
θ_y	8 − 9 − 10 - 11
θ_z	6 − 7 − 8 - 9
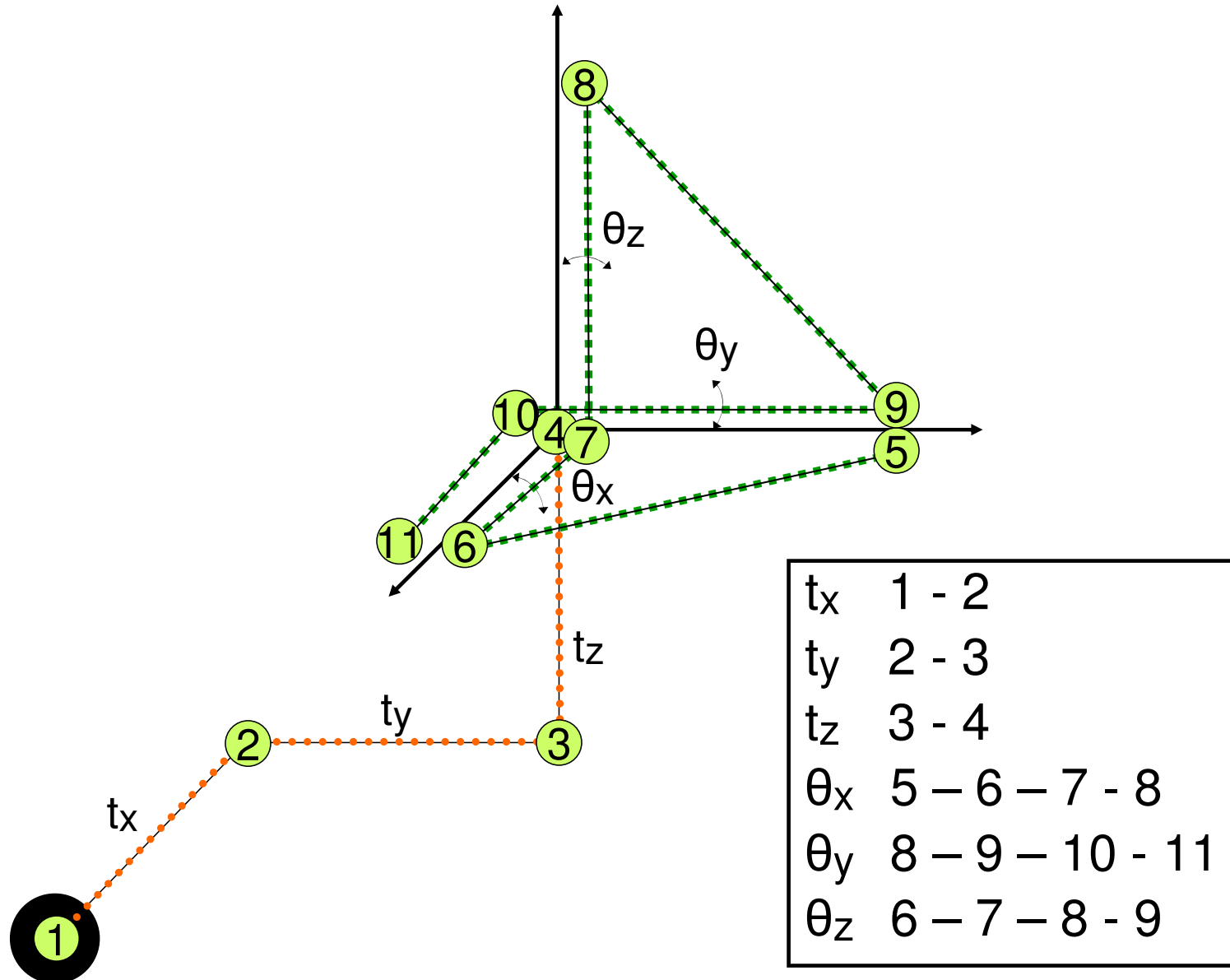
EACH CHAIN HAS ITS OWN ARM

©Michael Levitt 10
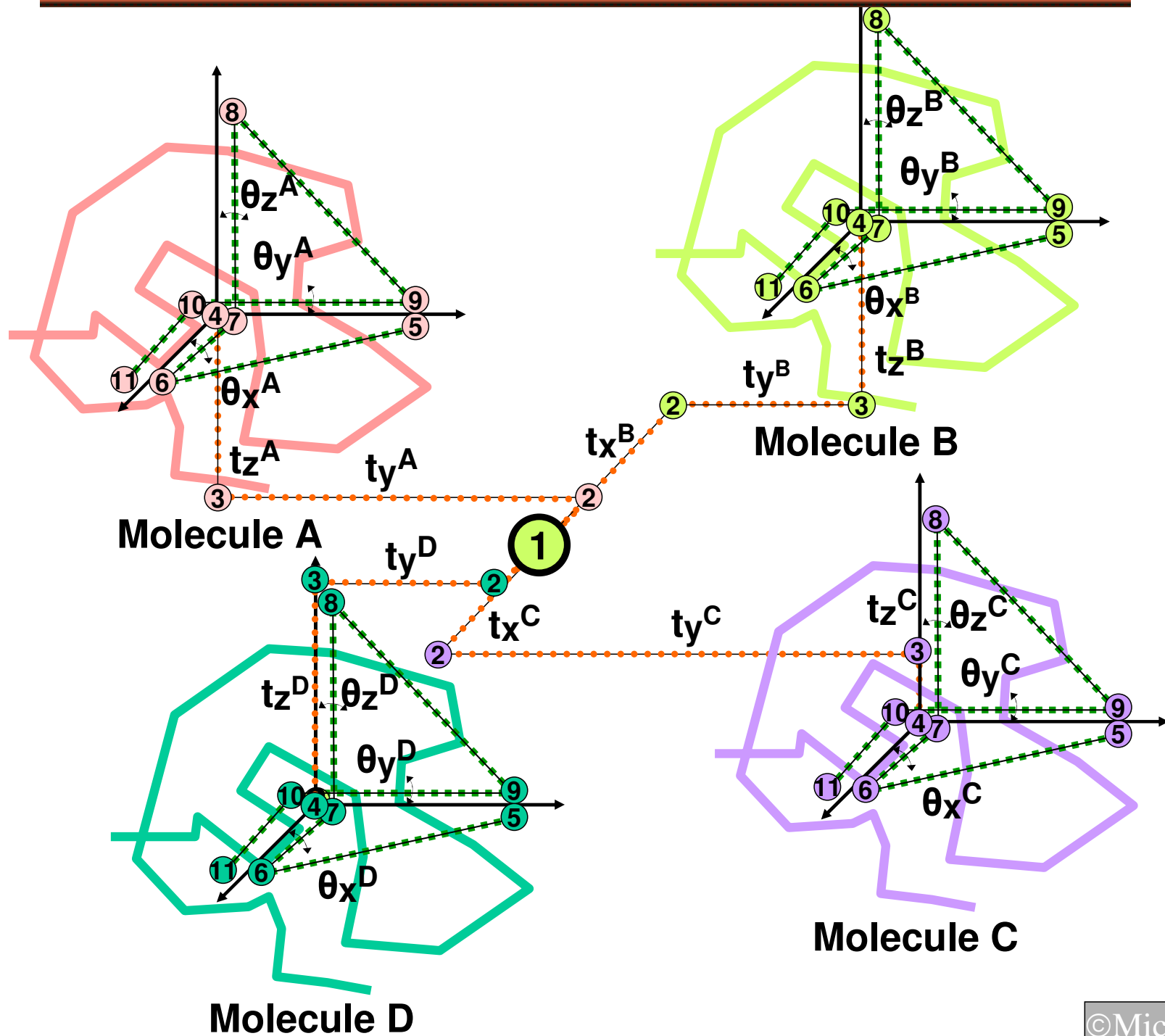
GENERALIZED EIGENVALUE PROBLEM

$$Vx = \lambda Tx$$

- V (Potential Energy) matrix is singular.

- T (Kinetic Energy) matrix is singular.

EIGENVECTORS LOOK WRONG

Different recommended methods give different

eigenvectors.

    RGG (QZVEC) from EISPACK

    DDGEVX from LAPACK

    FO2AEF from Harwell (NAG)

None!

What is correct?

Solution

SOLVE TWO NORMAL EIGENVALUE PROBLEMS

$\mathbf{VX = TX\Lambda}$ $\mathbf{V}$ and $\mathbf{T}$ are real, symmetric but singular.

First solve $\mathbf{TZ = Z\Omega}$ to give $\mathbf{T = Z\Omega Z^T}$.

Now calculate the square root of the generalized inverse of $\mathbf{T}$.

$\mathbf{T^{-1/2} = Z\Omega^{-1/2}Z^T}$ where $\Omega_{ii}^{-1/2} = 0$ if Ω_{ii} is 0.

Transform the original eigenvalue equation as follows:

$\mathbf{VX = TX\Lambda}$ or $\mathbf{VT^{-1/2}T^{1/2}X = T^{1/2}T^{1/2}X\Lambda}$

as $\mathbf{T^{-1/2}T^{1/2} = 1}$ and $\mathbf{T^{1/2}T^{1/2} = T}$ ($\mathbf{T}$ is real and symmetric).

Multiply both sides by $\mathbf{T^{-1/2}}$ to get

$\mathbf{(T^{-1/2}VT^{-1/2})(T^{1/2}X) = (T^{-1/2}T^{1/2})(T^{1/2}X)\Lambda}$ or $\mathbf{HZ = Z\Lambda}$

where $\mathbf{H = T^{-1/2}VT^{-1/2}}$. Solve for $\mathbf{Z}$ and $\mathbf{\Lambda}$, calculate

$\mathbf{X = T^{-1/2}Z}$, and test by $\mathbf{X^TVX = \Lambda}$ and $\mathbf{X^TTX = 1}$.
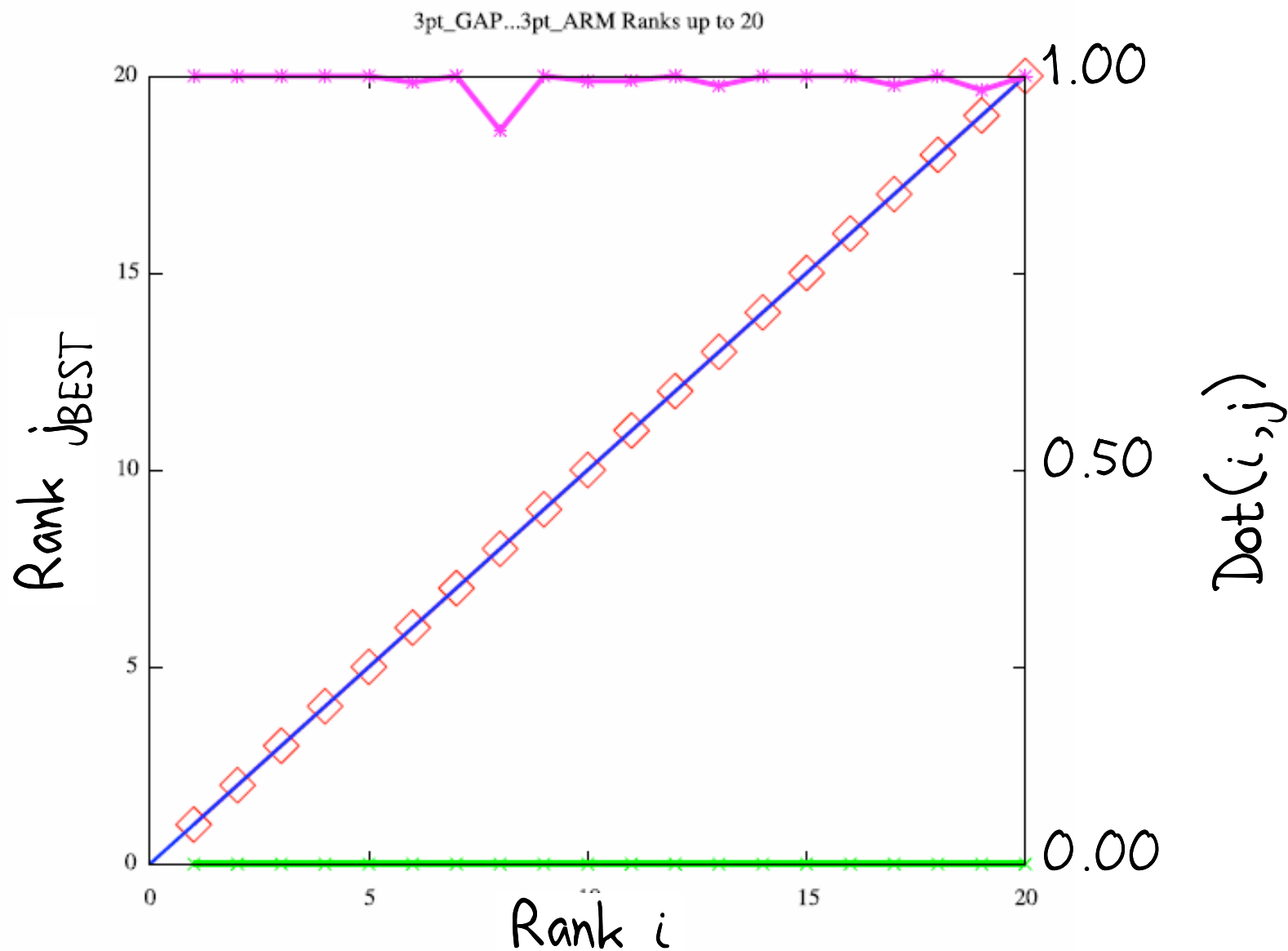
NOW ALL WORKS

- The eigenvectors are always real.

- Solutions are always obtained.

- Insensitive to singularities.

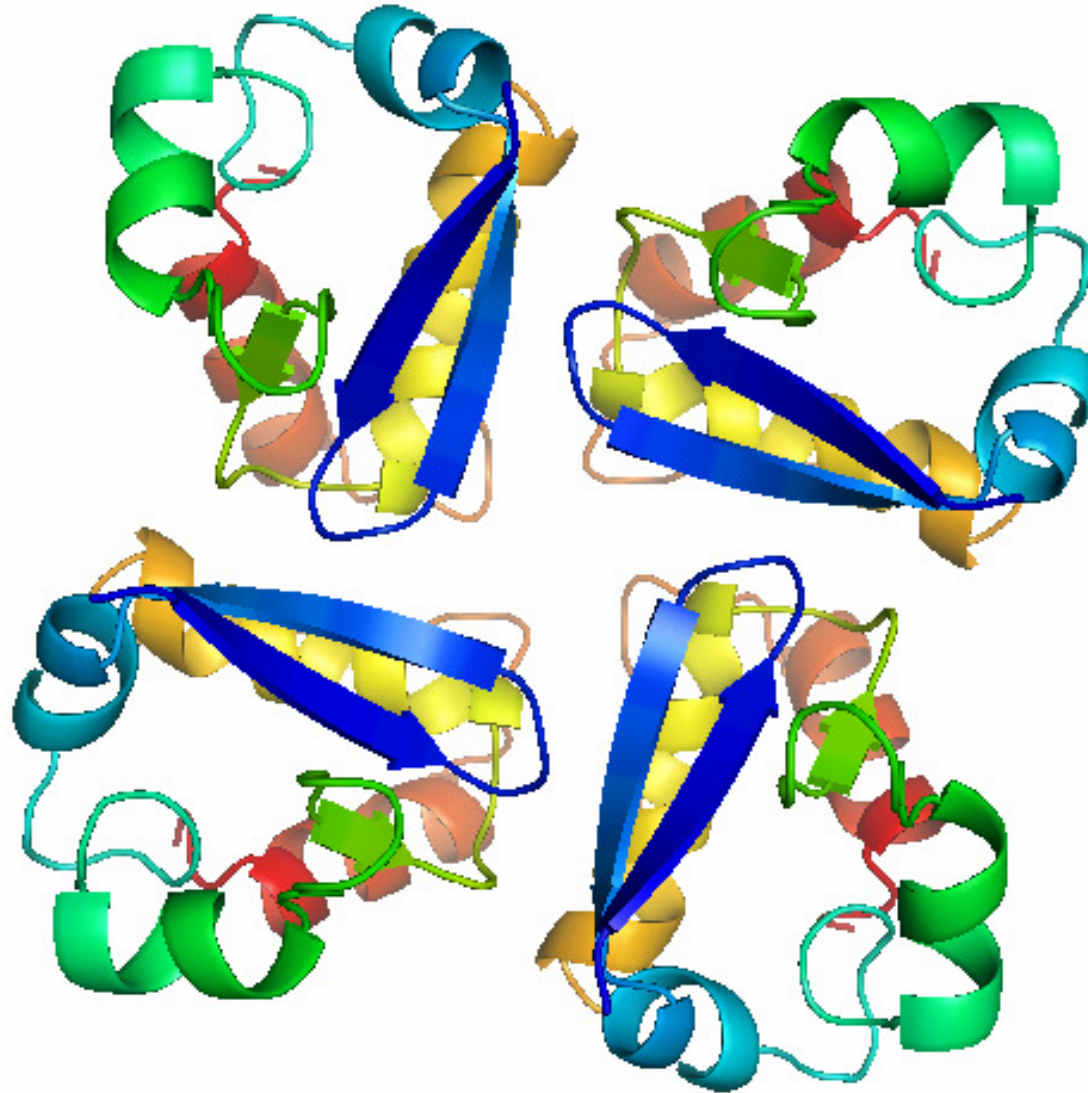- Insensitive to round-off error in V or T.

1CTF A VERY EASY TEST
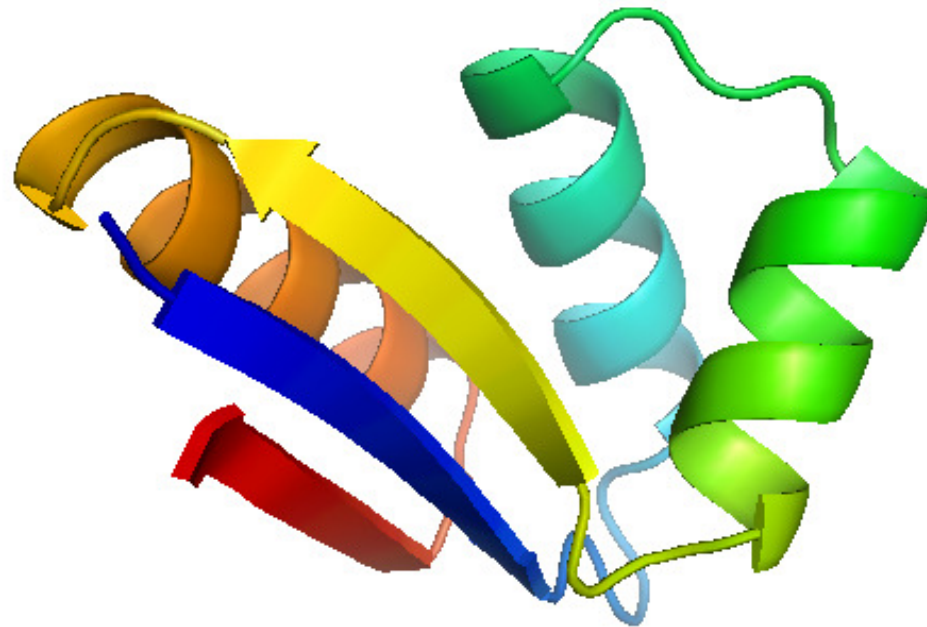
One chain with 54 residues

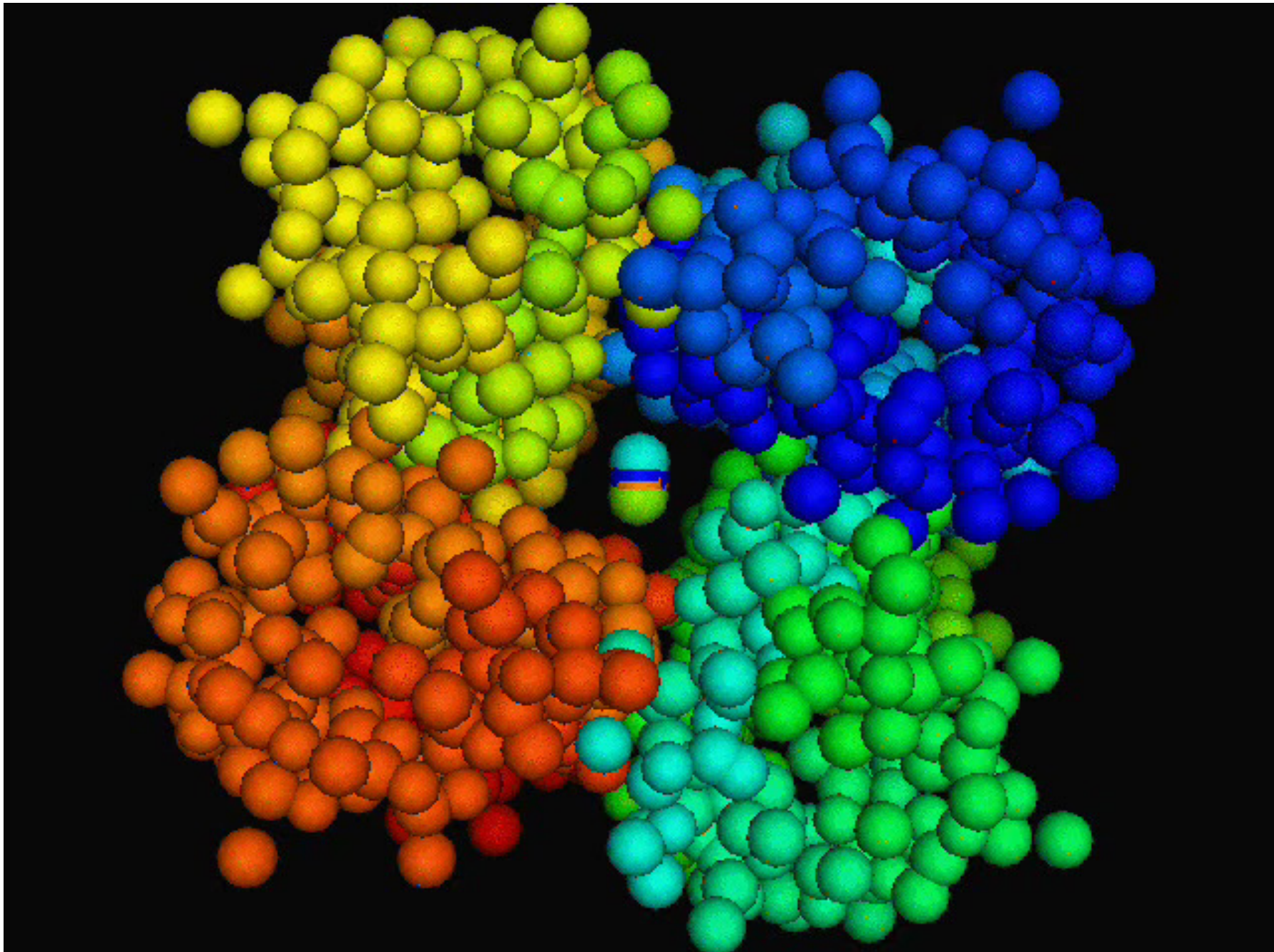# ARM HAS NO EFFECT ON MONOMER

3pt_GAP...3pt_ARM Ranks up to 20
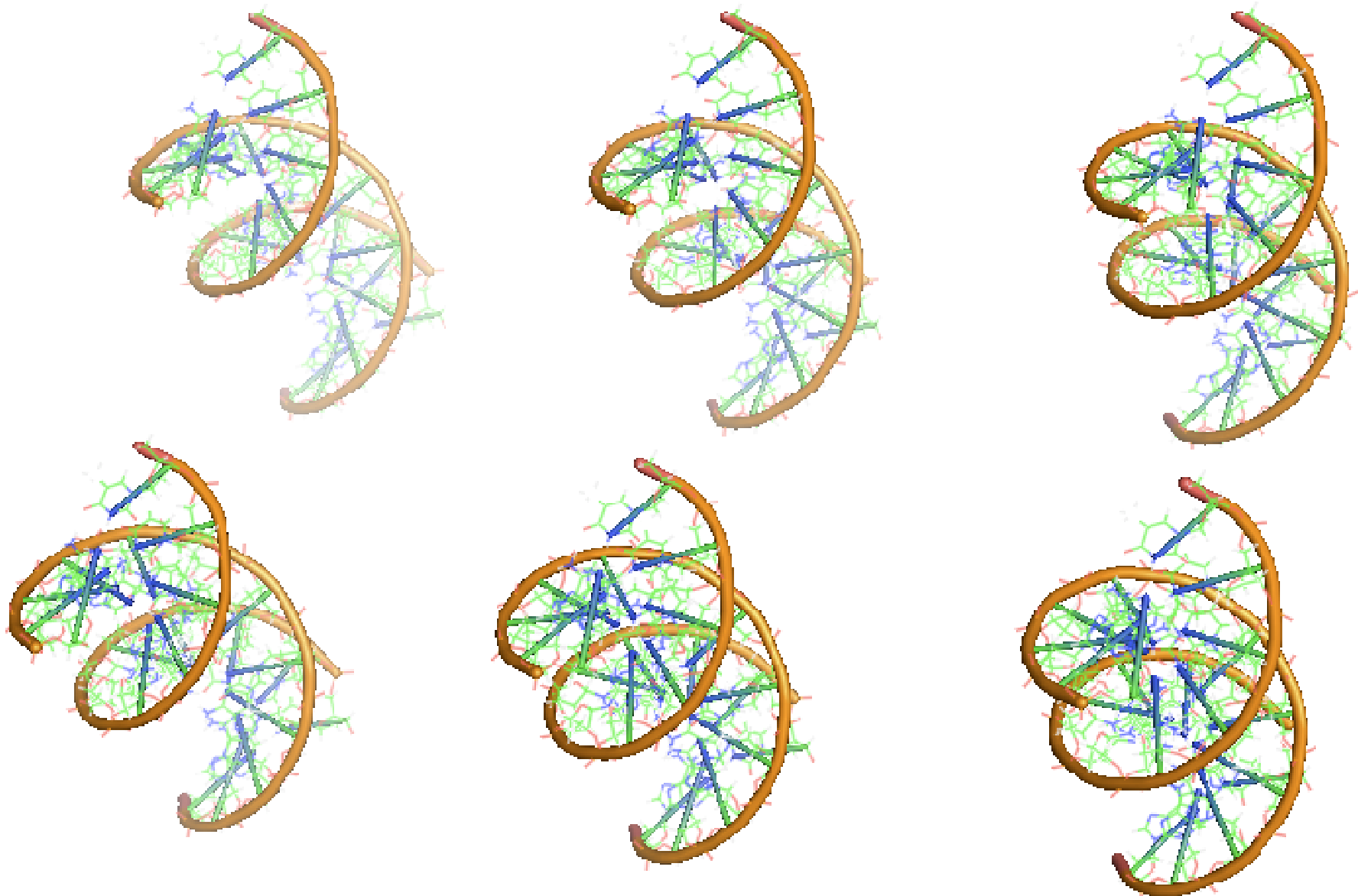
MODEL SYSTEM: TETRAMER 1AG8

Four chains in small tetramer (87 each)

MODE 01 IS SYMMETRIC

MODEL SYSTEM: 6xAAAATTTTAAAA

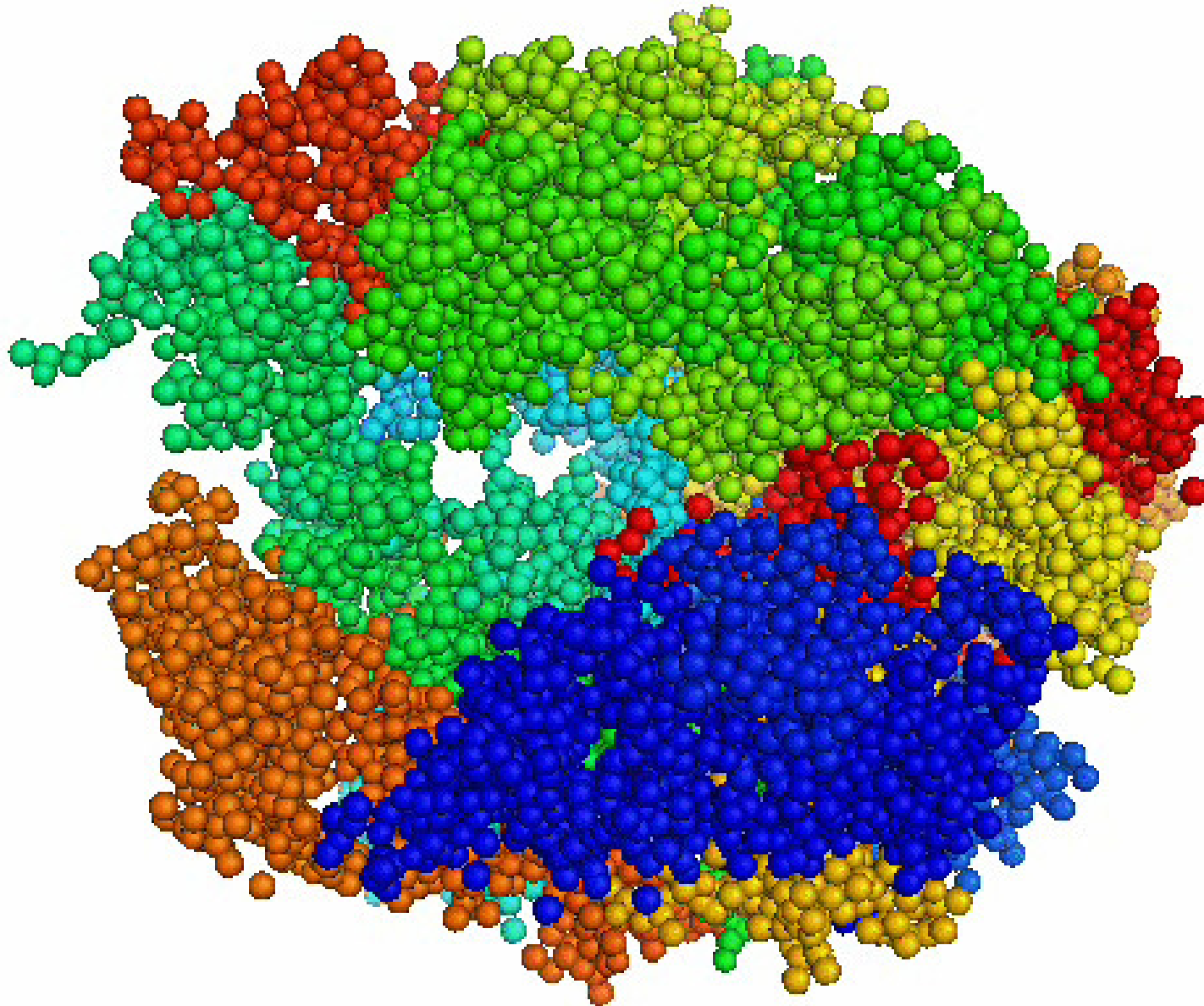Twelve chains in 6 independent molecules

6xAAAATTTTAAAA ARMS

Expect groups of 6 as 6 independent rigid bodies

Only see this for ARMs not GAPs

36 zero eigenvalues

ARM

GAP
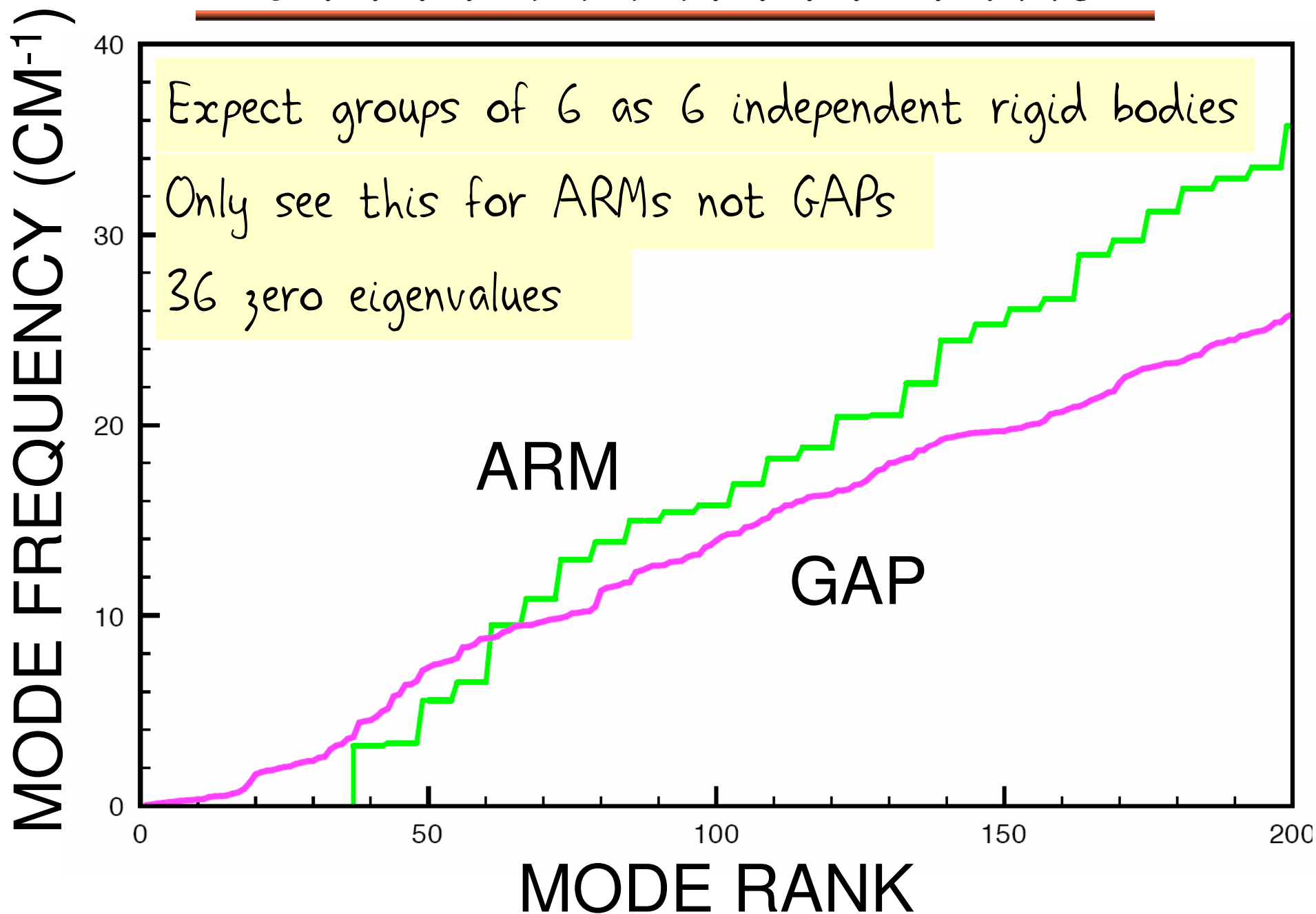
# RNA POLYMERASE NORMAL MODE 1

Future

METHODOLOGICAL TESTS

- A unique part of the present work is the calculation of real normal modes of large complexes in torsion angle space. Others have used quasi elastic modes (Tirion modes) in Cartesian space to look at similar systems.

- Test done to establish the advantages of our method include:
    - Is energy more harmonic with torsion angle variables?
    - Do Tirion modes with torsion angle differ from normal modes?
    - How important is damping?
    - Are results better with 3pt and all-atom KB functions?
    - Does number of degrees of freedom matter?

Chaperonin Mn-cpn NORMAL MODE 1
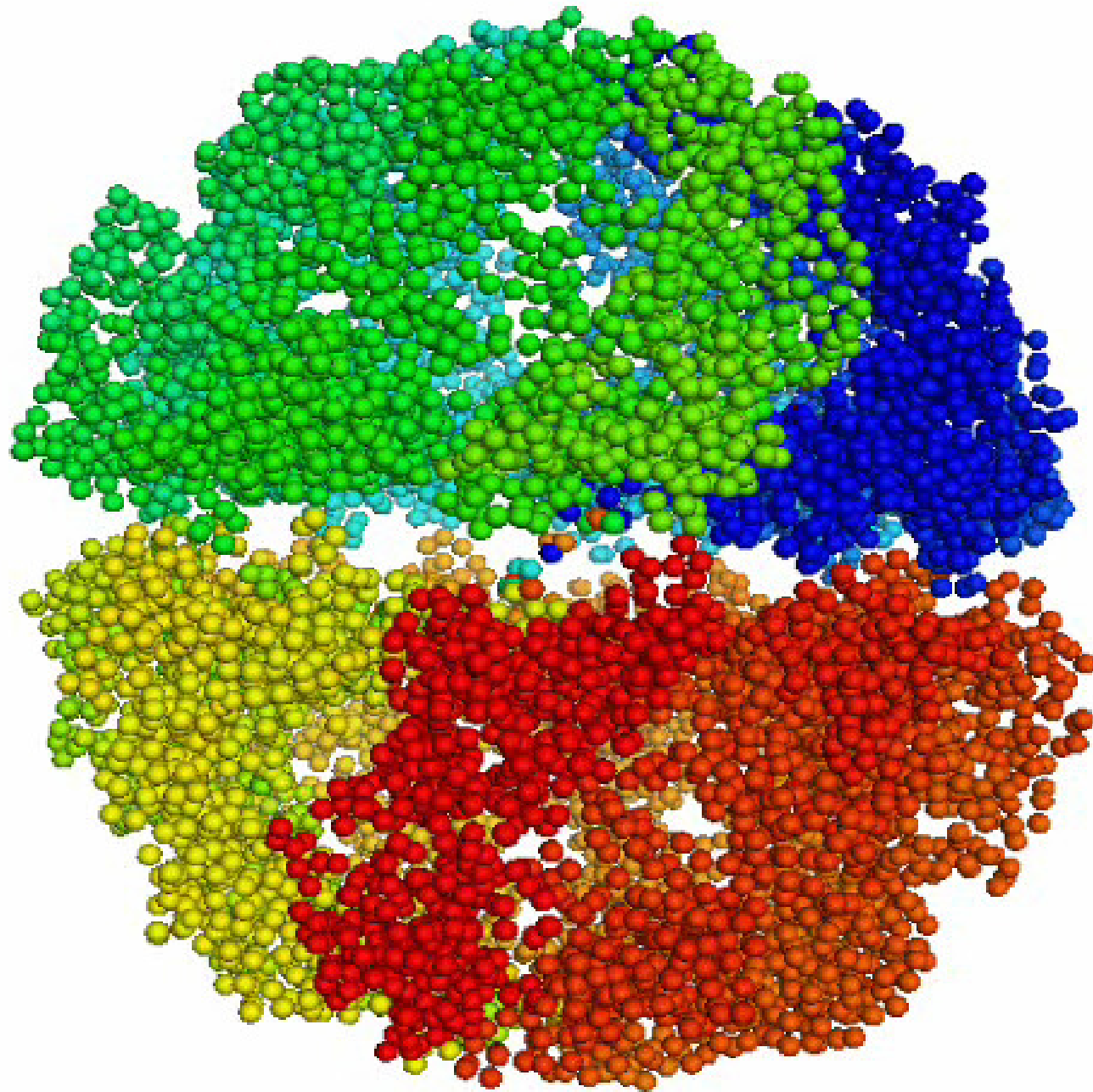
Chaperonin Mn-cpn NORMAL MODE 4
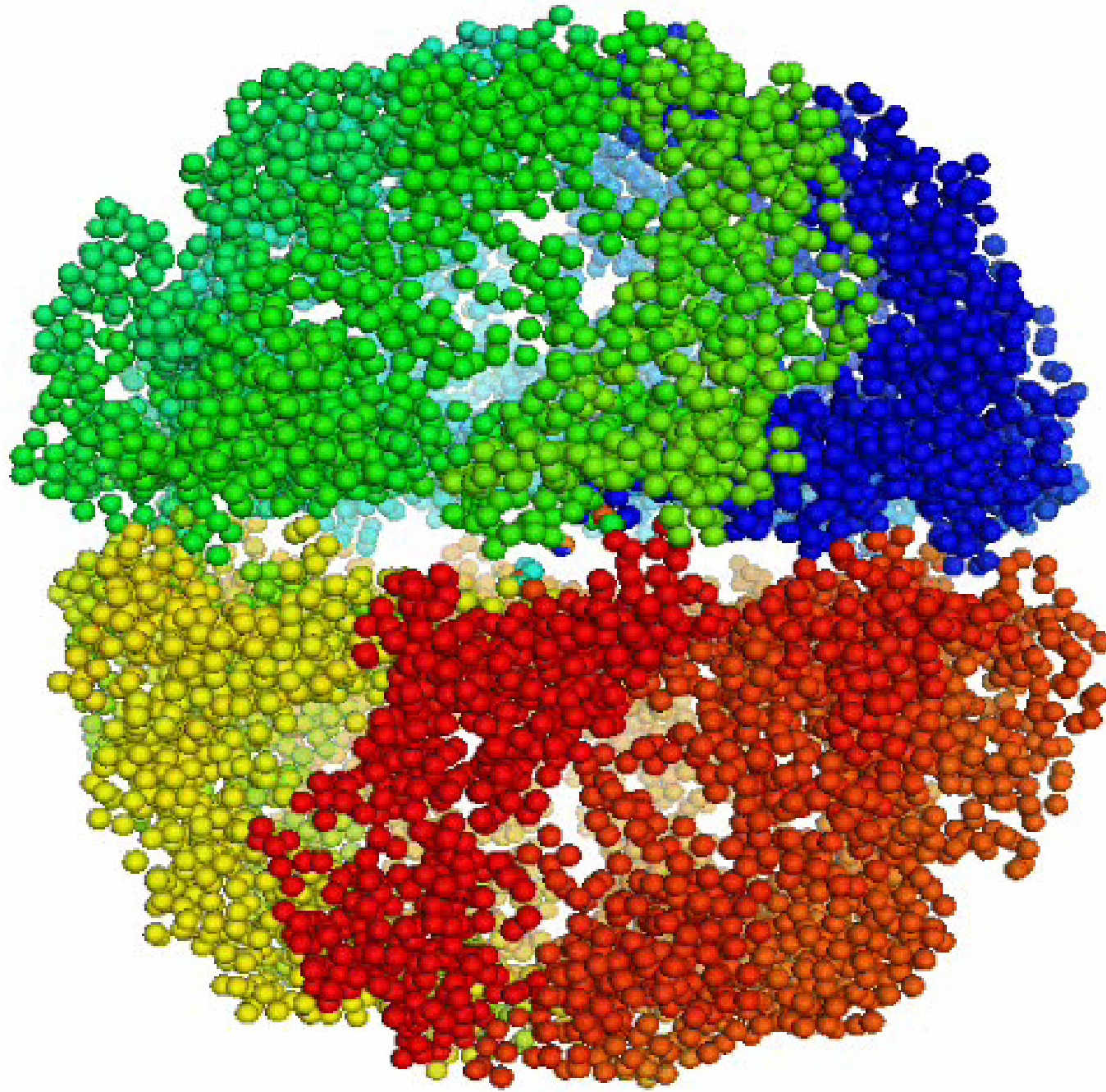
OUTLINE

- Few normal modes simulate suggestive motion. ✓
- Knowledge-based potentials work for refinement. ✓
- Simplified models are same as all-atom models. ✓
- Simulate suggestive motion of cellular machinery. ✓
- Knowledge-based potentials give stability & modes. ✓
- The chain order paradox. ✓
- Solution in solution of generalized eigenvalue equation. ✓

GENERAL ACKNOWLEDGMENTS

WEB

http://csb.stanford.edu/levitt

{ Papers
Lectures
PhD Thesis
Talklets

The End