Mathematical Models of Folded and Unfolded Protein Ensembles

Gregory Chirikjian

Department of Mechanical Engineering
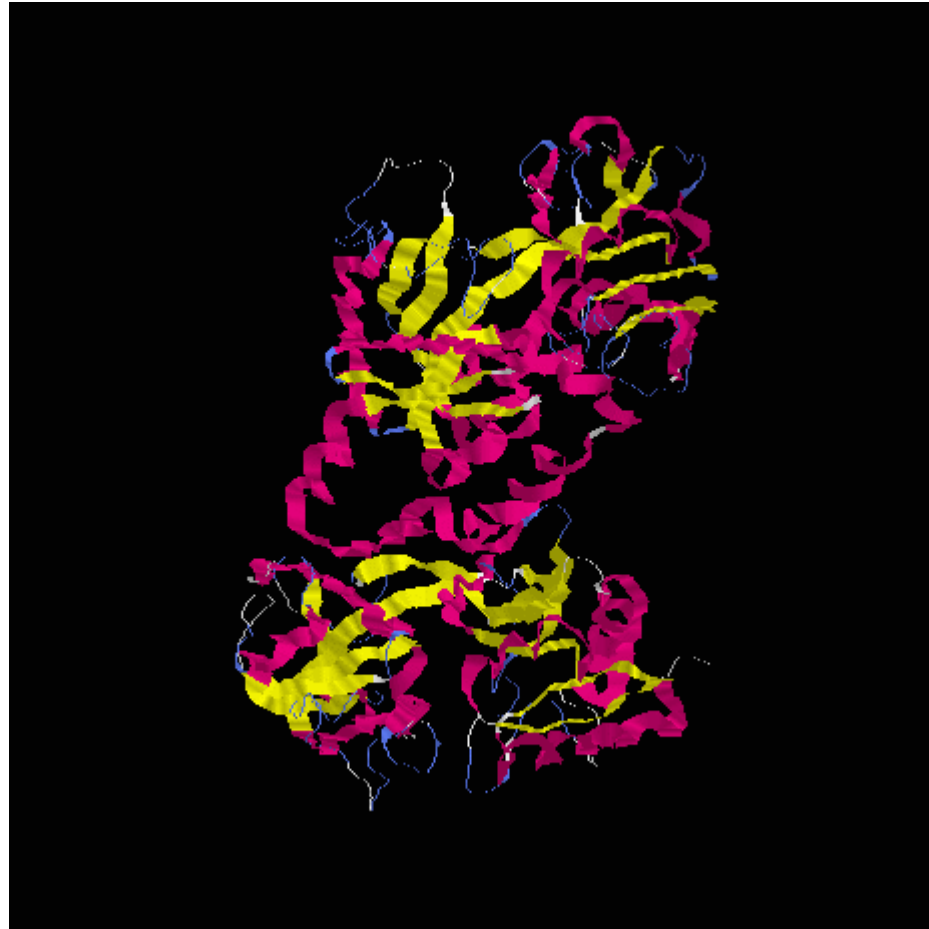
Johns Hopkins University

Topics of This Talk

- Conformational transitions between two folded states

- Statistical analysis of helix-helix interactions

- Approximating the conformational statistics of unfolded and partially folded chains

Conformational Transitions
# Using Elastic Network Interpolation

Lactoferrin Transition from 1lfg.pdb to 1lfh.pdb

M. K. Kim, R. L. Jernigan, G. S. Chirikjian. Biophysical Journal. 83:1620-1630. 2002

Rigid-cluster systems

- Many conformational changes in macromolecules can be resolved into hinge and shear motions which are associated with the collective behavior of atoms.

- Some macromolecules could be represented by a set of rigid-clusters.

$$x_i(t)$$

$$m_{i,a}$$

$$d_{i,a}(t) \qquad \text{i}^{\text{th}} \text{ cluster}$$
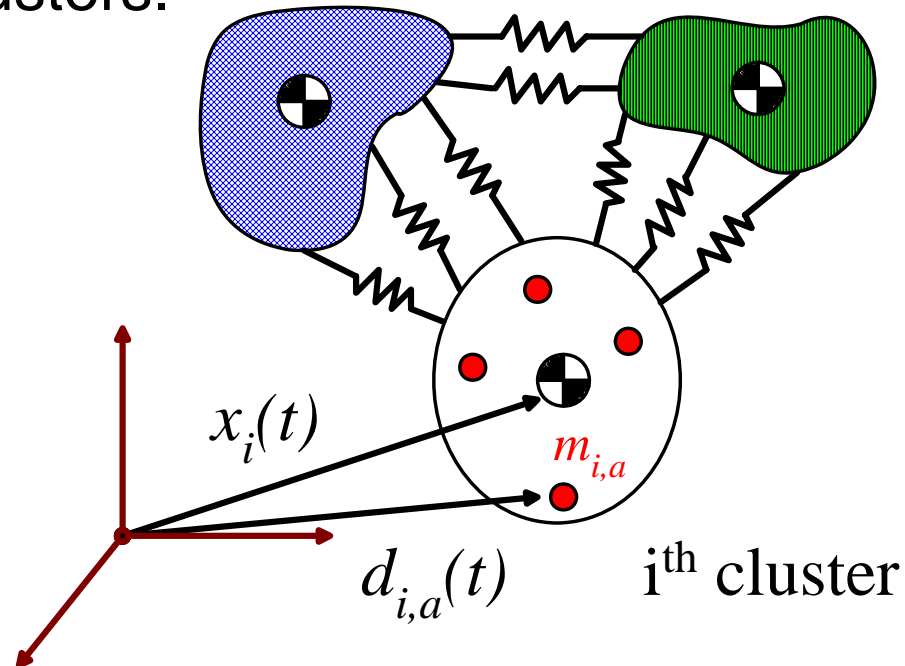
Rigid-cluster ENI

M. K. Kim, R. L. Jernigan, G. S. Chirikjian. Biophysical Journal. 89:43-55. 2005

i$^{\text{th}}$ cluster
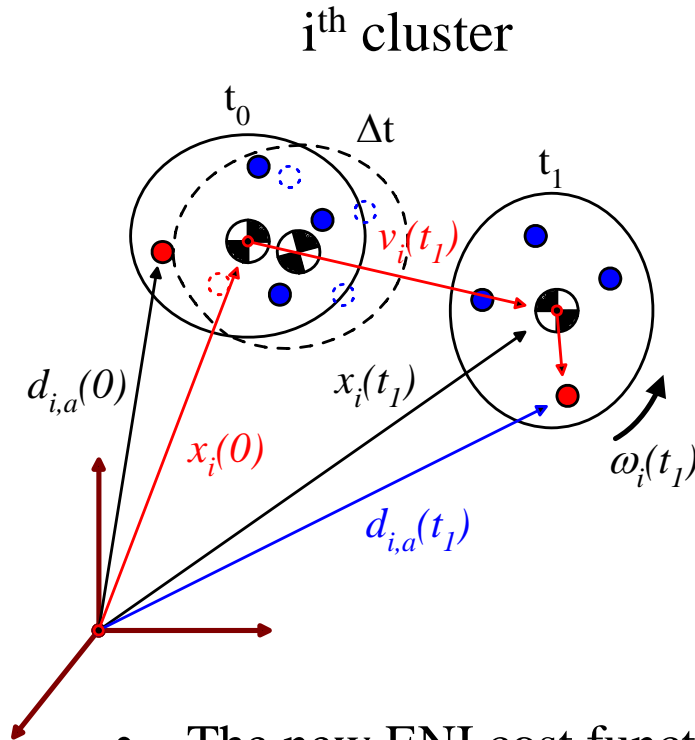
- The position of residue a at time t is

$$\vec{d}_{i,a}(t) = \vec{x}_i(0) + \vec{v}_i(t) + R(\vec{\omega}_i(t))(\vec{d}_{i,a}(0) - \vec{x}_i(0))$$

- Assuming small rigid-body displacement, motions of body i depend on the six-dimensional vector:

$$\vec{\delta}_i(t) = [\vec{v}_i^T(t)\ \vec{\omega}_i^T(t)]^T \in \mathbb{R}^6.$$

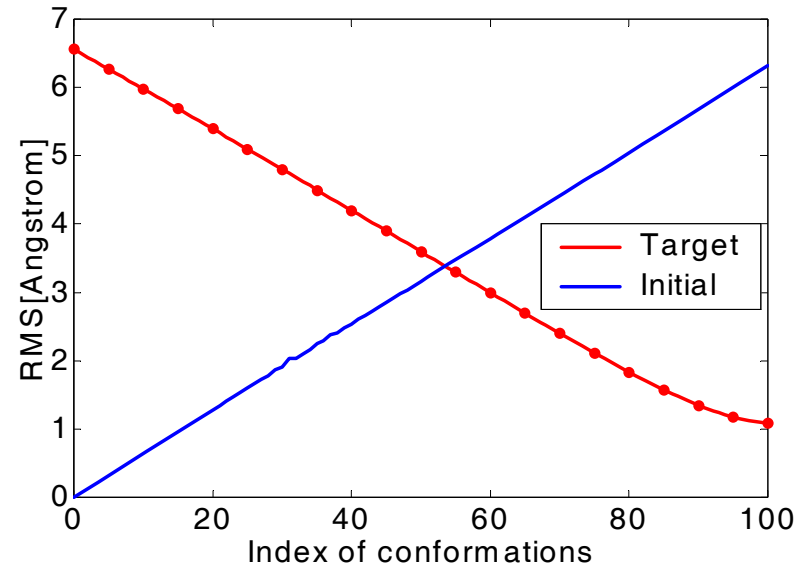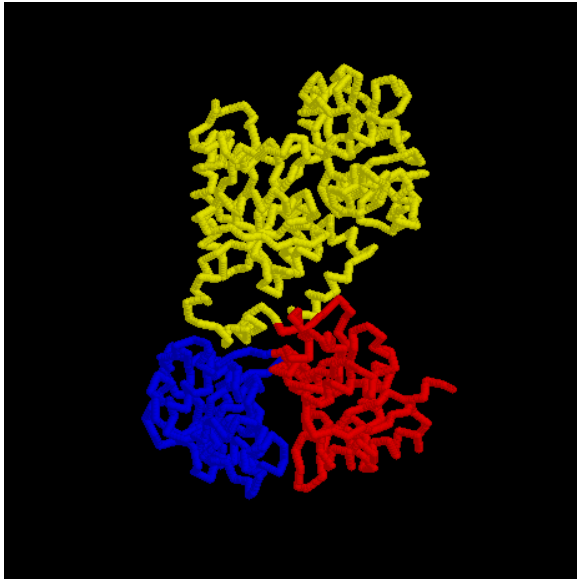translational $\quad$ orientational

- The new ENI cost function is defined as

$$C(\vec{\delta}) = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left\{\sum_{a=1}^{n(i)}\sum_{b=1}^{n(j)}k_{i,a,j,b}(\|\vec{d}_{i,a}(t) - \vec{d}_{j,b}(t)\| - l_{i,a,j,b})^2\right\}$$

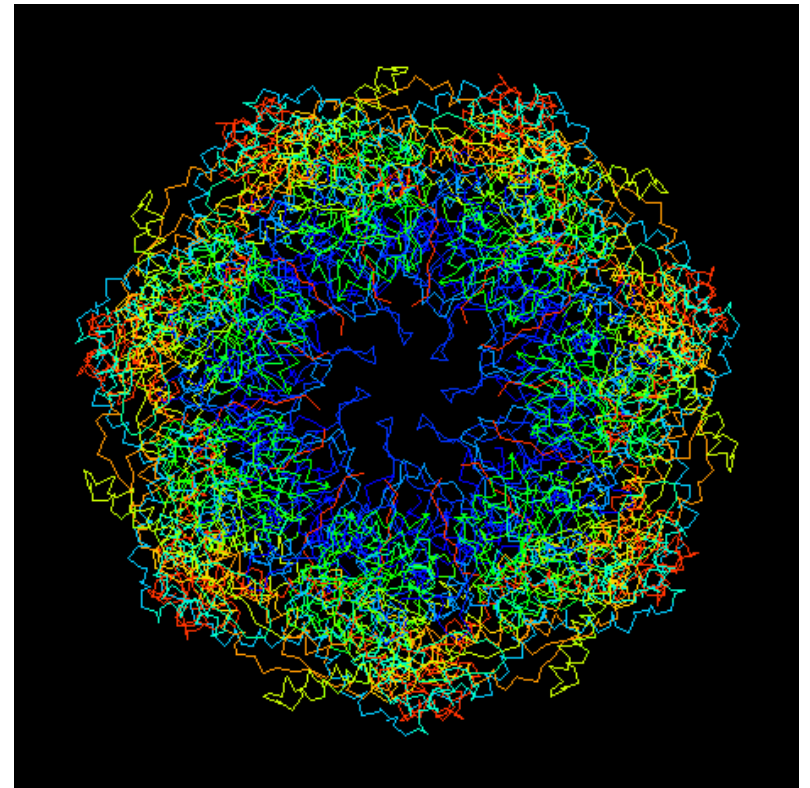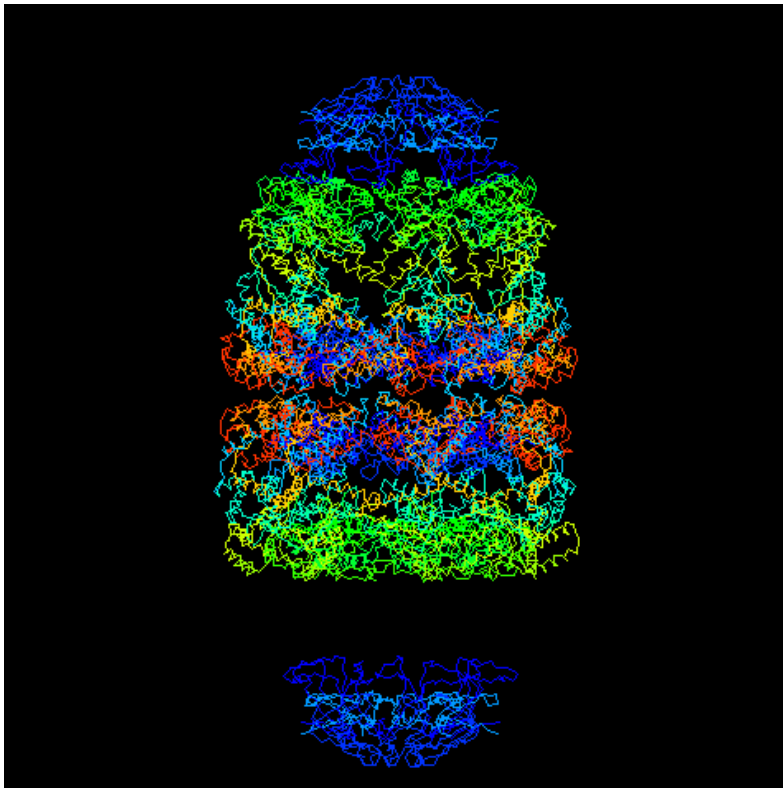# Lactoferrin

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\vec{x}_i - \vec{y}_i\|^2}$$

$$\text{Ratio of } \text{DOF} = \frac{Rigid-clusterENI}{C_\alpha ENI} = \frac{3\times6}{691\times3} < 0.01$$

ENI for the GroEL-GroES complex

Virus Capsid HK97
(1FH6 to 1IF0)
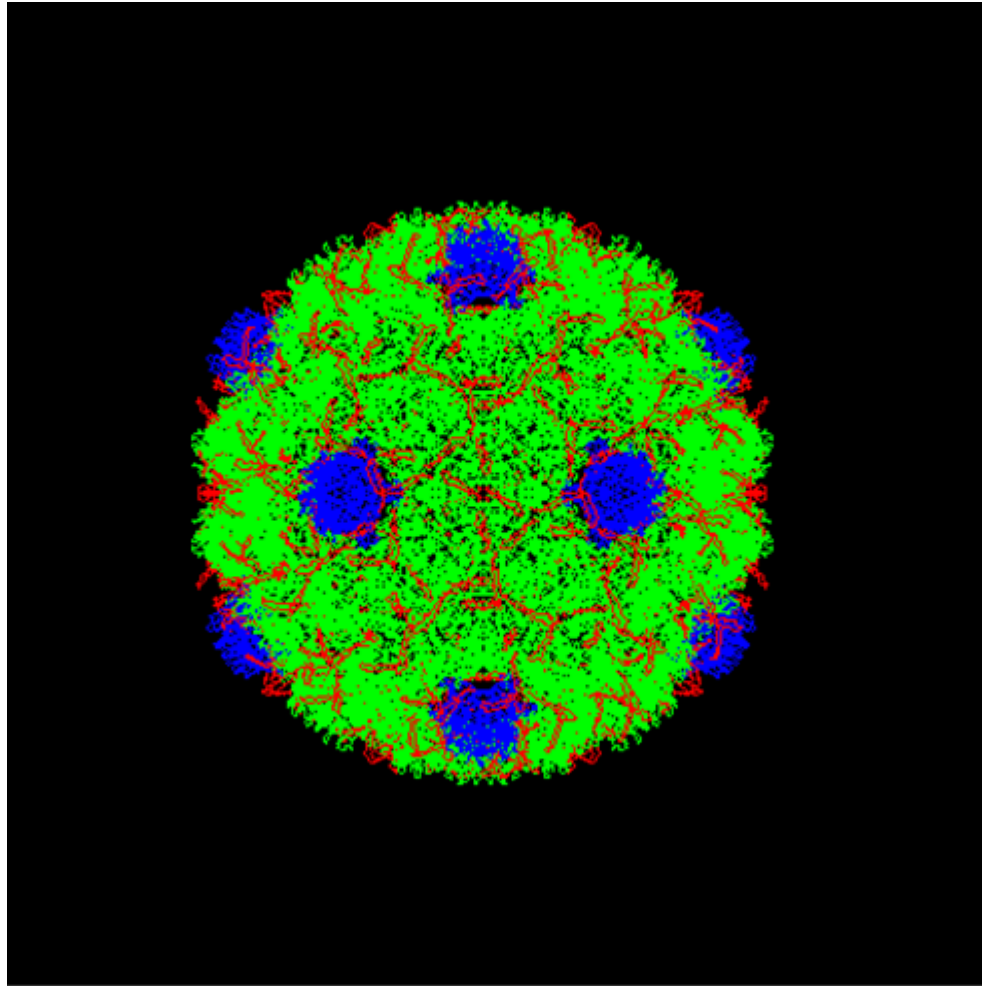
Helix-Helix Statistics

At What Angles Do Alpha-Helices like to Cross ?

– Biophysicists for 30 years predicted close to zero crossing angle

– Statistical analysis of Protein Data Bank (PDB) data suggested close to 90 degrees

– Kinematics/Group-Theory explains why there is a difference

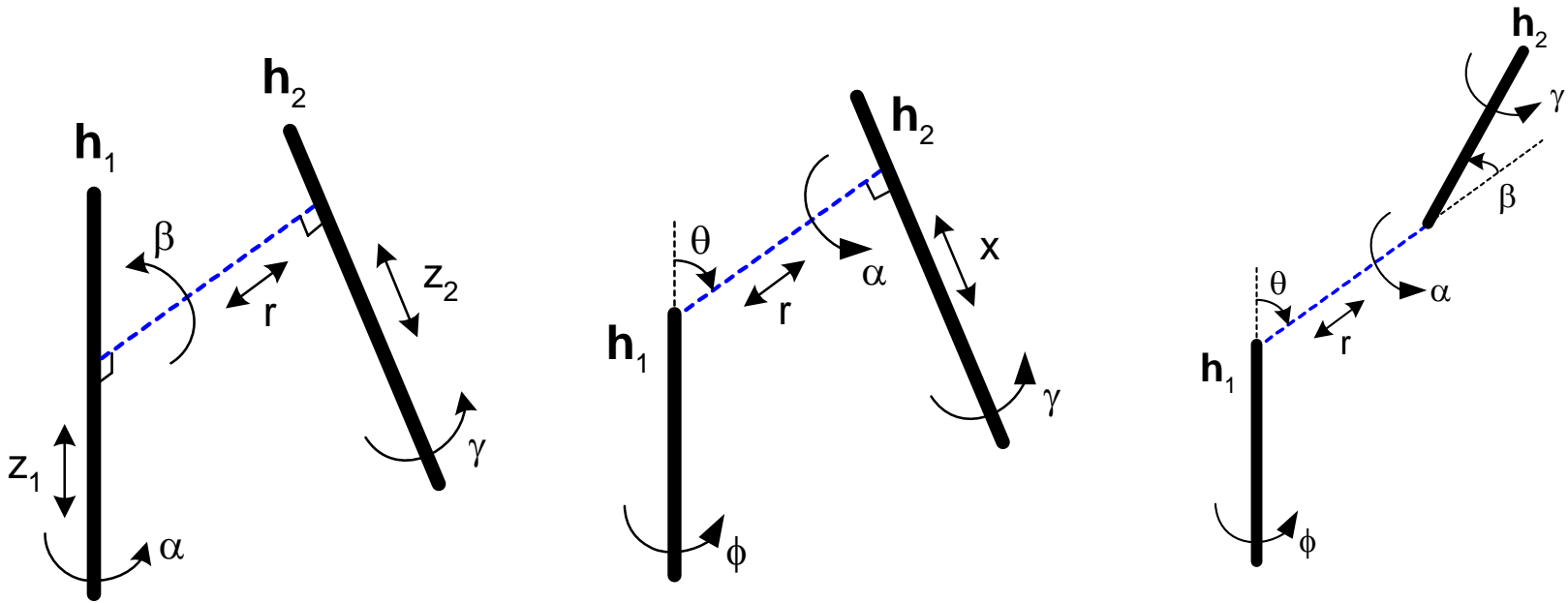# The Pose of a Rigid Body

Euclidean motion group *SE(3)*

- An element of *SE(3)*: $g = \begin{pmatrix} A & a \\ 0^T & 1 \end{pmatrix}$

- Binary operation: matrix multiplication

- Describes not only motion, but relative position and orientation (pose) as well

Helices in Proteins Can Interact in Several Ways

How Do We Analyze Statistics of the Set of Such Pairs ? First We Need Kinematics

Case 1:

$$H = \begin{pmatrix} R_3(\alpha)R_1(\beta)R_3(\gamma) & z_2 R_3(\alpha)R_1(\beta)\mathbf{e}_3 + r\,R_3(\alpha)\mathbf{e}_1 + z_1\,\mathbf{e}_3 \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix}$$

Case 2:

$$H = \begin{pmatrix} R_3(\phi)R_1(\theta)R_3(\alpha)R_1(\gamma) & r\,R_3(\phi)R_1(\theta)\mathbf{e}_3 + xR_3(\phi)R_1(\theta)R_3(\alpha)\mathbf{e}_1 \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix}$$

Case 3:

$$H = \begin{pmatrix} R_3(\phi)R_1(\theta)R_3(\alpha)R_1(\beta)R_3(\gamma) & r\,R_3(\phi)R_1(\theta)\mathbf{e}_3 \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix}$$

Computing Volume in SE(3)

- ## Infinitesimal motions

$$S = \begin{pmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{pmatrix} \implies \mathrm{vect}(S) = [s_1, s_2, s_3]^T = \vec{s}$$

$$M = \begin{pmatrix} S & \vec{v} \\ \vec{0}^T & 0 \end{pmatrix} \implies (M)^\vee = \begin{pmatrix} \vec{s} \\ \vec{v} \end{pmatrix}$$

- ## Jacobians for the motion group *SE(N)*

$$J_R(\vec{q}) = [(H^{-1} \frac{\partial H}{\partial q_1})^\vee, \cdots, (H^{-1} \frac{\partial H}{\partial q_N})^\vee]$$

$$J_L(\vec{q}) = [(\frac{\partial H}{\partial q_1} H^{-1})^\vee, \cdots, (\frac{\partial H}{\partial q_N} H^{-1})^\vee]$$

Normalization by Proper Volume Element is Important

Case 1:

$$\left| \det \vartheta_R \right| = \sin^2 \beta$$

Case 2:

$$\left| \det \vartheta_R \right| = r \sin \theta$$

Case 3:

$$\left| \det \vartheta_R \right| = r^2 \sin \beta \sin \theta$$

How Do We Take Into Account Measurement Uncertainty ?

$$f(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{\delta(\theta-\theta_i)}{\sin^2\theta} = \frac{1}{N}\sum_{i=1}^{N}\frac{\delta(\theta-\theta_i)}{\sin^2\theta_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{\delta(\theta-\theta_i)}{\int_{-\pi}^{\pi}\delta(\theta'-\theta_i)\sin^2\theta'd\theta'}$$

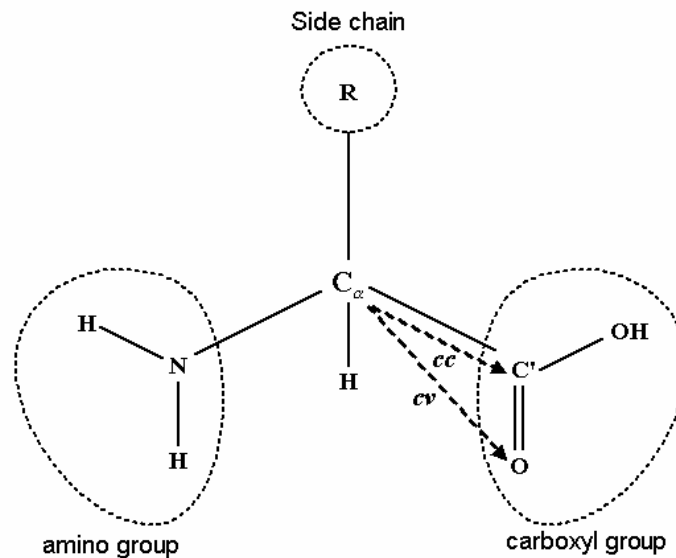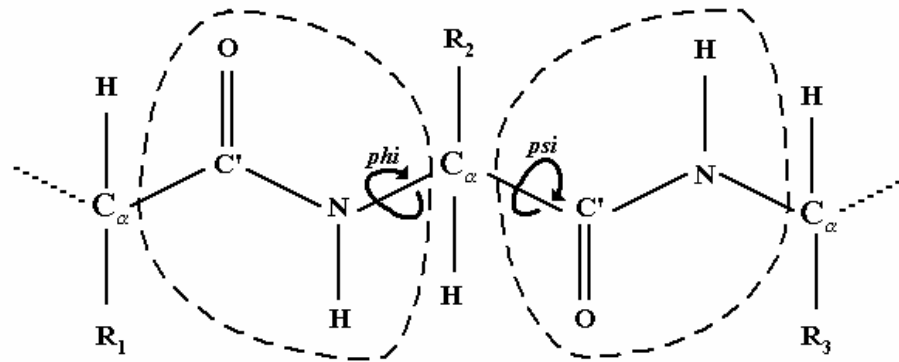$$\tilde{f}(\theta,t) = \frac{1}{N}\sum_{i=1}^{N}\frac{k(\theta-\theta_i,t)}{\int_{-\pi}^{\pi}k(\theta'-\theta_i,t)\sin^2\theta'd\theta'}$$

$$h(x,t) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t} \qquad k(\theta,t) = \sum_{n=-\infty}^{\infty}h(\theta-2\pi n,t)$$

Normalized Interhelical Angle-Distance Distribution for Case 1

A Model for Conformational Statistics of the Unfolded Ensemble
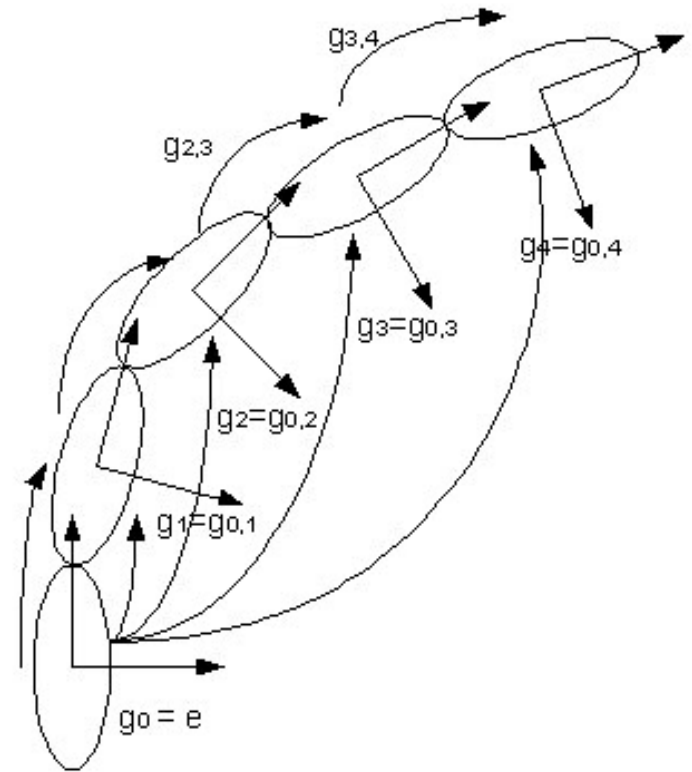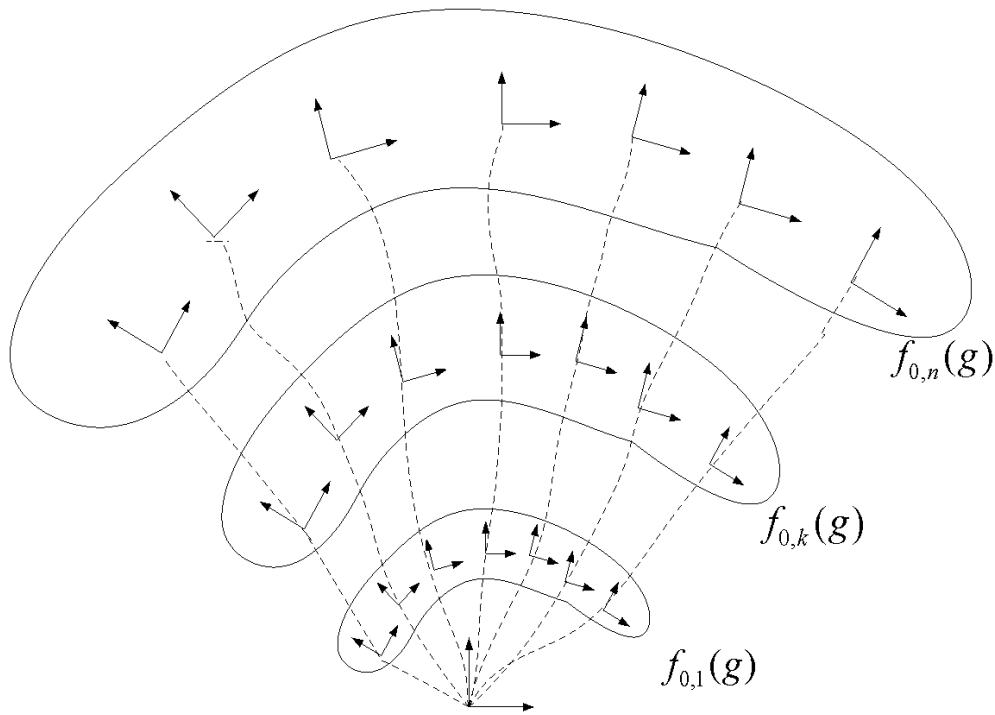
Structure of Proteins and Attachment of Reference Frames

Properties of Rigid-Body Motions
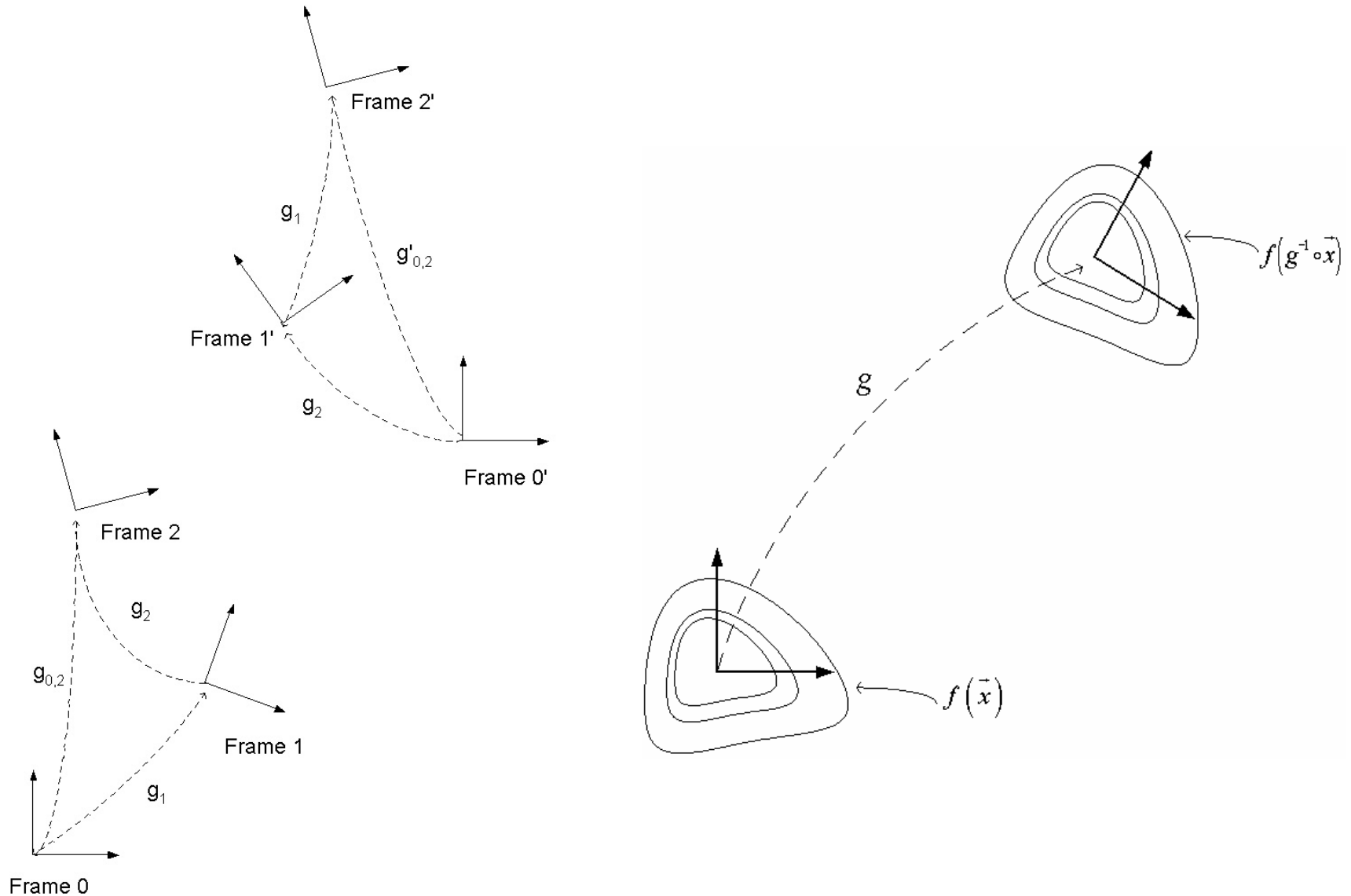
The Ensemble of Freely Moving Conformations

Generating the Ensemble of Conformations by Convolution

$$f_{0,2} = f_{0,1} * f_{1,2}$$

$$f_{0,1}(h)$$
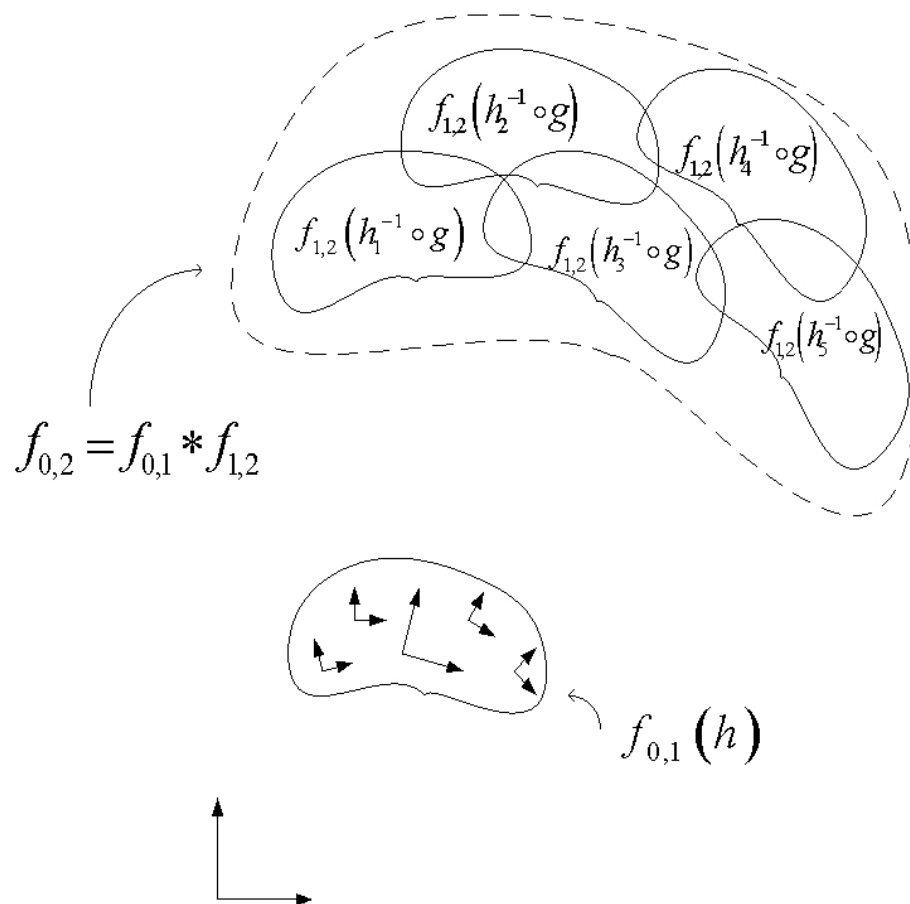
G. S. Chirikjian. Computational and Theoretical Polymer Science. 11:143-153. 2001

J. S. Kim, G. S. Chirikjian. Polymer 46:11904-11917. 2005

Propagating By Convolution

$$f(g, L_1 + L_2) = \int_G f(h, L_1) f(h^{-1} \circ g, L_2) dh$$

$$\int_G dg = \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_0^{\infty} r^2 \sin\beta \sin\theta \, dr \, d\phi \, d\theta \, d\alpha \, d\beta \, d\gamma$$

$$f(r, L_1 + L_2) = \int_0^{2\pi} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_0^{\pi} f(g, L_1 + L_2) d\omega$$

Fourier Analysis of Motion

- Fourier transform of a function of motion, *f(g)*

$$F(f) = \hat{f}(p) = \int_G f(g)U(g^{-1}, p)\,dg$$
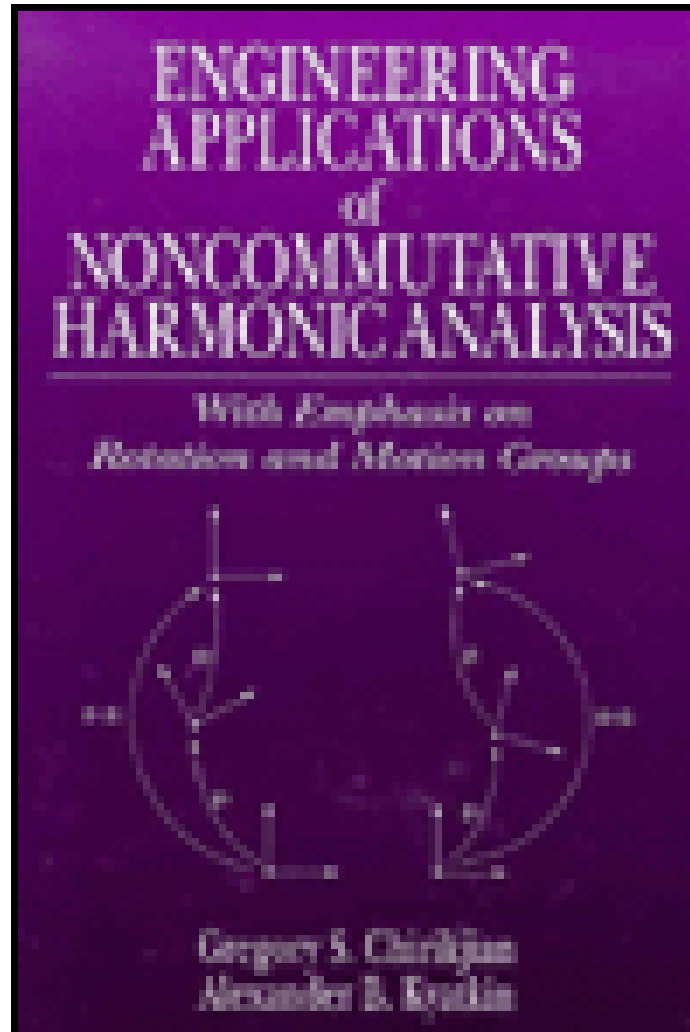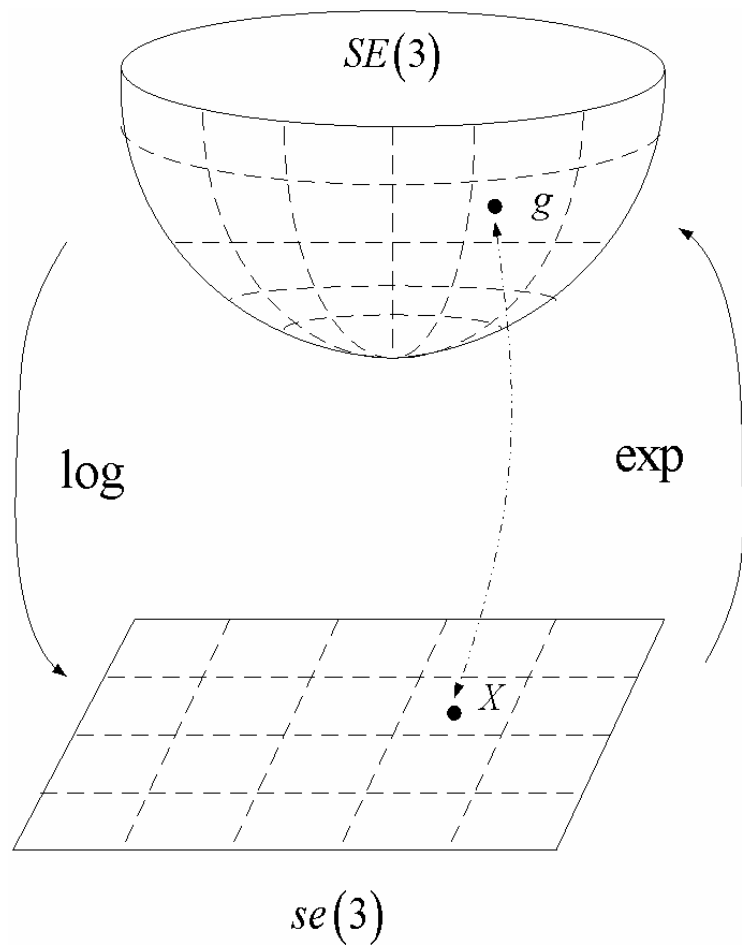
- Inverse Fourier transform of a function of motion

$$F^{-1}(\hat{f}) = f(g) = \int trace(\hat{f}(p)U(g, p))\,p\,dp$$

where $g \in SE(N)$, p is a frequency parameter,
$U(g,p)$ is a matrix representation of *SE(N)*, and
dg is a volume element at g.

For More on Harmonic Analysis …

Geometry of Lie Groups

Lie Algebra Basis Elements and Exponential Parametrization for $SE(3)$

$$E_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad E_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$E_3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad E_4 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$E_5 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad E_6 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Example of exponential map:

$$\exp(\theta E_3) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \exp(yE_5) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

The 'exponential parametrization'

$$g = g(\chi_1, \chi_2, ..., \chi_6) = \exp\left(\sum_{i=1}^{6} \chi_i E_i\right) \tag{1}$$

is a useful way to describe relatively small rigid-body motions because, unlike the Euler angles, it does not have singularities near the identity.

One defines the 'vee' operator, $\vee$, such that for any

$$X = \sum_{i=1}^{6} \chi_i E_i,$$

$$X^{\vee} = \begin{pmatrix} \chi_1 \\ \chi_2 \\ \chi_3 \\ \vdots \\ \chi_6 \end{pmatrix}.$$

The Adjoint Matrix:

$$Ad_g(X^\vee) = (gXg^{-1})^\vee,$$

If $g = (\mathbf{a}, A)$ then

$$Ad_g = \begin{pmatrix} A & 0 \\ \mathbf{a} \times A & A \end{pmatrix}$$

The Jacobian Matrix:

$$J(\boldsymbol{\chi}) = \left[\left(g^{-1} \frac{\partial g}{\partial \chi_1} \right)^\vee, \cdots, \left(g^{-1} \frac{\partial g}{\partial \chi_6} \right)^\vee \right]$$

The 'Vee' operation:

$$\boldsymbol{\chi} = (\log g)^\vee.$$

The volume element:

$$dg = |J(\boldsymbol{\chi})| d\chi_1 \cdots d\chi_6$$

Convolution:

$$f_{0,2}(g) = (f_{0,1} * f_{1,2})(g) = \int_G f_{0,1}(h) f_{1,2}(h^{-1} \circ g) dh$$

Computing Bounds on the Entropy of the Unfolded Ensemble Using Gaussians on $SE(3)$

We can define the Gaussian in the exponential parameters as

$$f(g(\boldsymbol{\chi})) = \frac{1}{(2\pi)^3 |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}\boldsymbol{\chi}^T \Sigma^{-1} \boldsymbol{\chi}) \tag{1}$$

Given two distributions that are shifted as $f_{i,i+1}(g_{i,i+1}^{-1} \circ g)$, each with 6×6 covariance $\Sigma_{i,i+1}$, then it can be shown that the mean and covariance of the convolution $f_{0,1}(g_{0,1}^{-1} \circ g) * f_{1,2}(g_{1,2}^{-1} \circ g)$ respectively will be of the form $g_{0,2} = g_{0,1} \circ g_{1,2}$ and

$$\Sigma_{0,2} = Ad_{g_{1,2}}^{-1} \Sigma_{0,1} Ad_{g_{1,2}}^{-T} + \Sigma_{1,2}. \tag{2}$$

$$f(g_1, g_2, ..., g_n) = \prod_{i=0}^{n-1} f_{i,i+1}(g_i^{-1} \circ g_{i+1}) \tag{3}$$

where $g_0 = e$, the identity.

The full pose entropy of a phantom chain:

$$S_g = -\int_G \cdots \int_G f(g_1, g_2, ..., g_n) \log f(g_1, g_2, ..., g_n) dg_1 \cdots dg_n. \tag{4}$$

Marginal and conditional entropies can also be computed.

References

- A. Skliros, G. S. Chirikjian. *Torsional Random-Walk Statistics on Lattices using Convolution on Crystallographic Motion Groups*. Polymer. 48:2155-2173. 2007

- J. S. Kim, G. S. Chirikjian. *A Unified Approach to Conformational Statistics of Classical Polymer and Polypeptide Models*. Polymer 46:11904-11917. 2005

- S. Lee, G. S. Chirikjian , *Interhelical Angle and Distance Preferences in Globular Proteins*. Biophysical Journal. 86:1105-1117. 2004

- M. K. Kim, R. L. Jernigan, G. S. Chirikjian. *Rigid-Cluster Models of Conformational Transitions in Macromolecular Machines and Assemblies*. Biophysical Journal. 89:43-55. 2005

- G.S. Chirikjian, *Modeling Conformational Disorder in Polypeptide Coils:New Perspectives from Robotics and Polymer Theory (draft)*

Summary

- Kinematics and Lie group theory can be used in new ways to model protein motions

- Differential geometry has a place in the analysis of statistical data extracted from the pdb

- Polymer-like models with quasi-closed-form solutions can provide insights from a different perspective than massive computations