Coarse-graining entropy, forces, and structure

W. G. Noid
The Pennsylvania State University
Department of Chemistry

KITP meeting on
Modeling Soft Matter:
Linking Multiple Length and Time Scales
June 7, 2012

Acknowledgements

Joe Rudzinski
Chris Ellis
Sushant Kumar
Tommy Foley
Nick Dunn

Wayne Mullinax

KITP organizers

Scott Shell

Henry and Camille Dreyfus Foundation

Alfred P. Sloan Research Foundation

Penn State Institute for Cyberscience

Inspiration

Computer simulation of protein folding

Michael Levitt* & Arieh Warshel*

Department of Chemical Physics, Weizmann Institute of Science, Rehovoth, Israel

Nature Vol. 253 February 27 1975

"Here we tackle the [protein folding] problem differently. First, we simplify the representation of a protein by averaging over fine details. This is done both to make the calculations much more efficient and also to avoid having to distinguish between many conformations that differ only in these finer details. Second, we simulate the folding of this simple structure …"

"Our method … is based on two assumptions: (1) that much of the protein's fine structure can be eliminated by averaging, and (2) that the overall chain folding can be obtained by considering only the most effective variables (those that vary most slowly yet cause the greatest changes in conformation)."

A Warning

On the formation of protein tertiary structure on a computer

(protein folding/computer simulation/protein evolution/role of glycines)

ARNOLD T. HAGLER* AND BARRY HONIG†

"[Previous studies] have used extremely simplified representations of PTI, which, upon energy minimization, fold into globular structures that in some way resemble the native protein. … The impression generated by these various simulations is that major progress has been made … i.e., the folding problem may be far more tractable than generally been considered. …

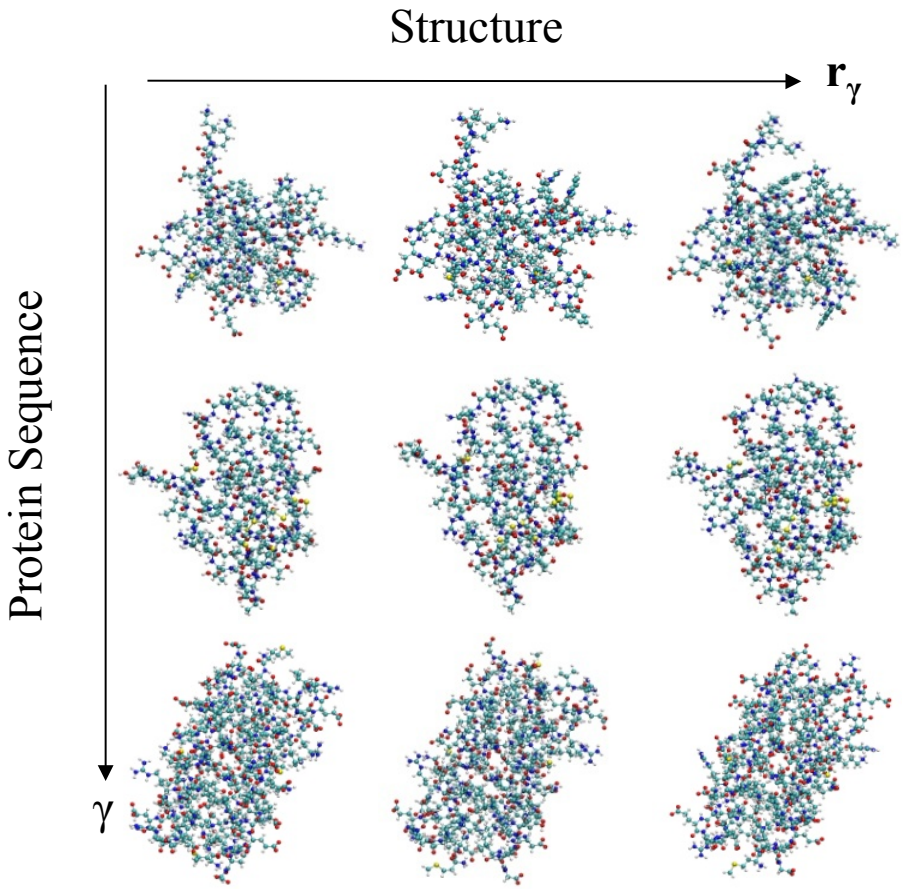
One of the major conclusions of this study is that the criteria that have been used to evaluate the success of most folding simulations has been overly permissive. … First, we show that it is possible to obtain a computed structure of PTI that satisfies all of the criteria that have been used previously to define successful folding simulations, from a sequence that would certainly not yield PTI-like conformation … Many of the positive results that have been reported are due entirely to [built-in features of the models] and may thus be regarded as artifacts.

A careful examination reveals that despite superficial similarities to the native protein, all computed structures have fundamental flaws … they fail to reproduce … important features characteristic of the tertiary structure … [and] appear sterically inaccessible from the native conformation. … "

Proc. Natl. Acad. Sci. USA
Vol. 75, No. 2, pp. 554–558, February 1978

The derailment at Gare Montparnasse, Paris, 1895.
http://phys.columbia.edu/~tutorial/

A Very Good Question



Tanaka and Scheraga (1976):

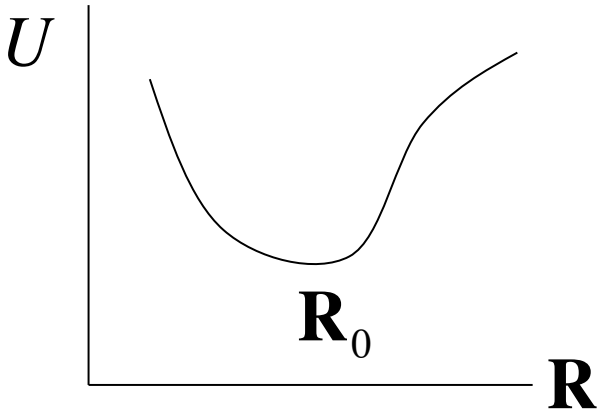
What interactions generated the PDB structures?

(At a Coarse-grained level.)

Knowledge-based approaches

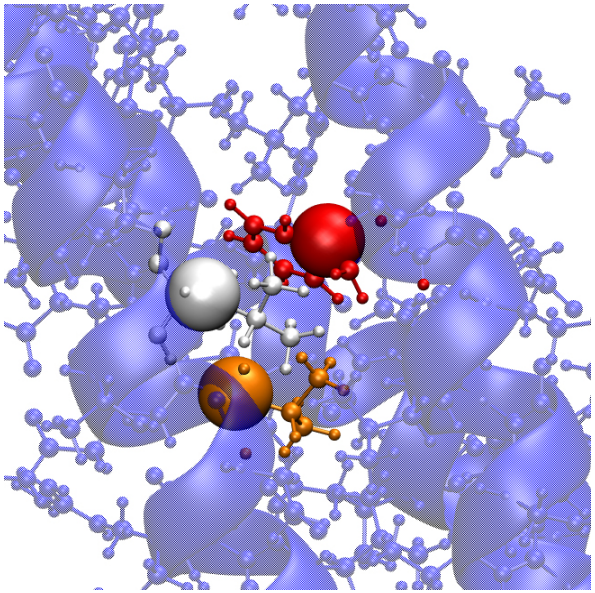
Foldability Criterion: given $\mathbf{R}_0$ for each protein

$$U(\mathbf{R}_0) = \min U(\mathbf{R})$$



Scheraga, Crippen, Wolynes, Shakhnovich, Banavar, …

Boltzmann hypothesis:



$$U(\mathbf{R}) = U_0(\mathbf{R}) + \sum_{\zeta} U_\zeta(r_\zeta)$$

Reference State

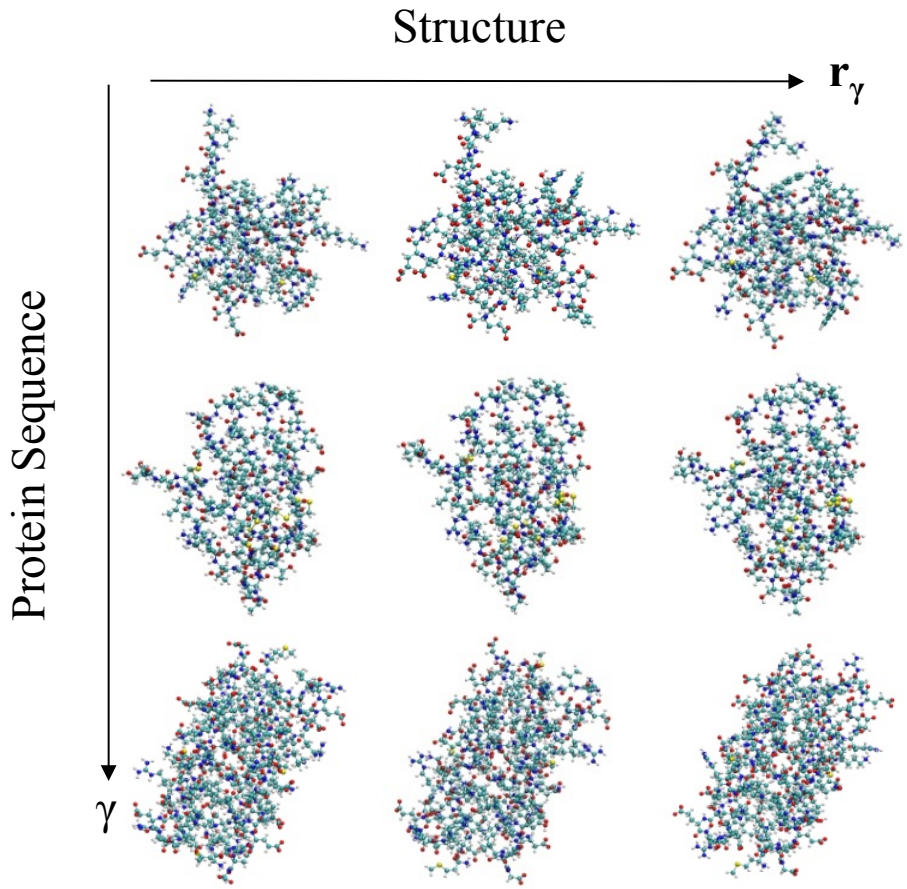
Interaction

$$p(\mathbf{R}) = p_0(\mathbf{R}) \prod_{\zeta} p_\zeta(r_\zeta)$$

$$\exp[-U_\zeta(r) / kT] = p_\zeta(r) / p_{\zeta 0}(r)$$

Scheraga, Jernigan, Sippl, Baker, Skolnick, Dill, Thirumalai, Straub …

Motivating Questions



What interactions generated the PDB structures? (Tanaka and Scheraga 1976)

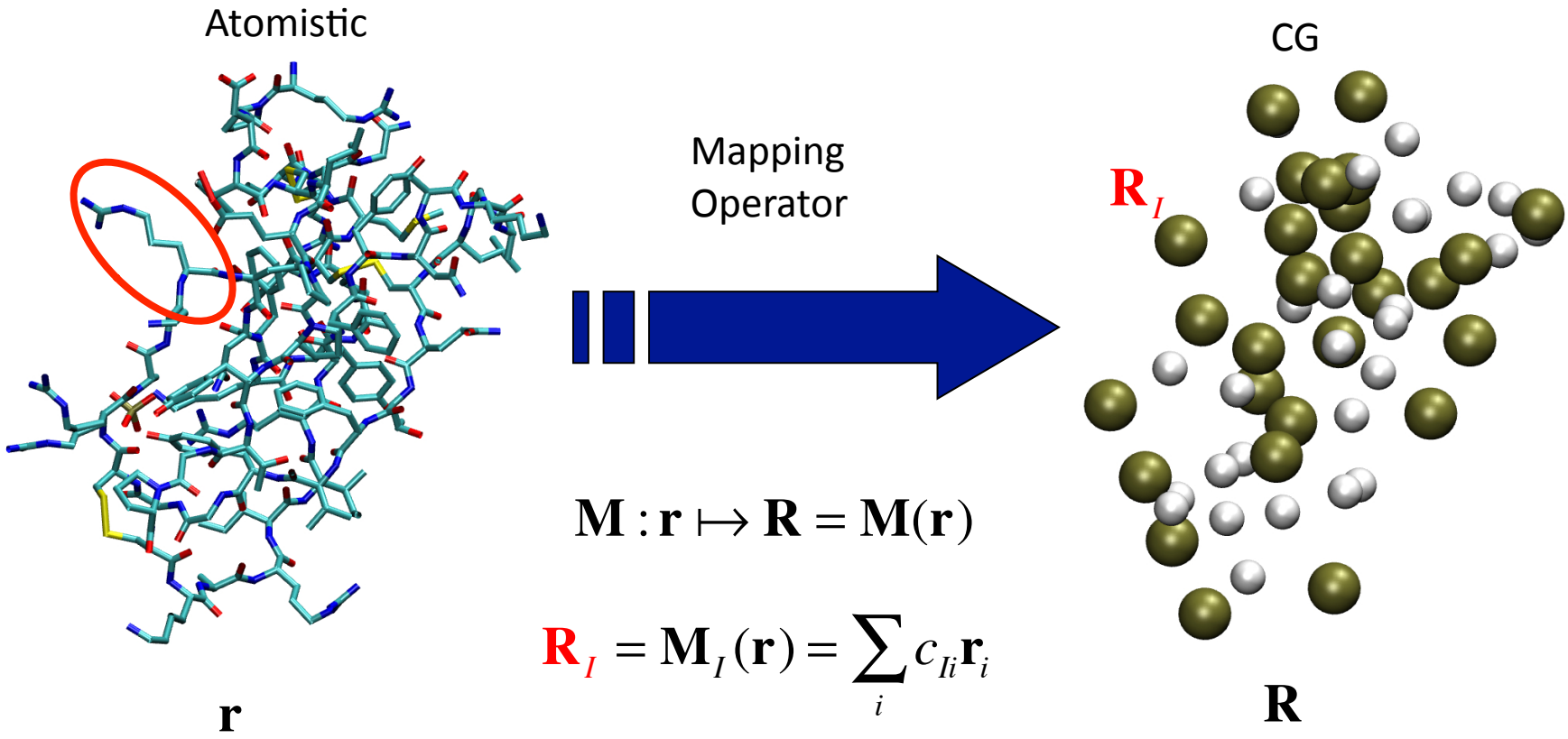
1. Given a collection of structures, what was the underlying potential?

2. How can one determine a transferable Coarse-Grained (CG) potential that accurately models structure for multiple proteins?

Outline

1. Introduction: Basic theory of force-matching
2. Force-matching without forces: Generalized Yvon-Born-Green Theory
3. Extended Ensemble framework: Variational approach for transferability
4. New directions:
    1. Connections to information theory and thermodynamics
    2. Mean forces as a unifying framework for understanding structure-potential relations
5. Outstanding challenges

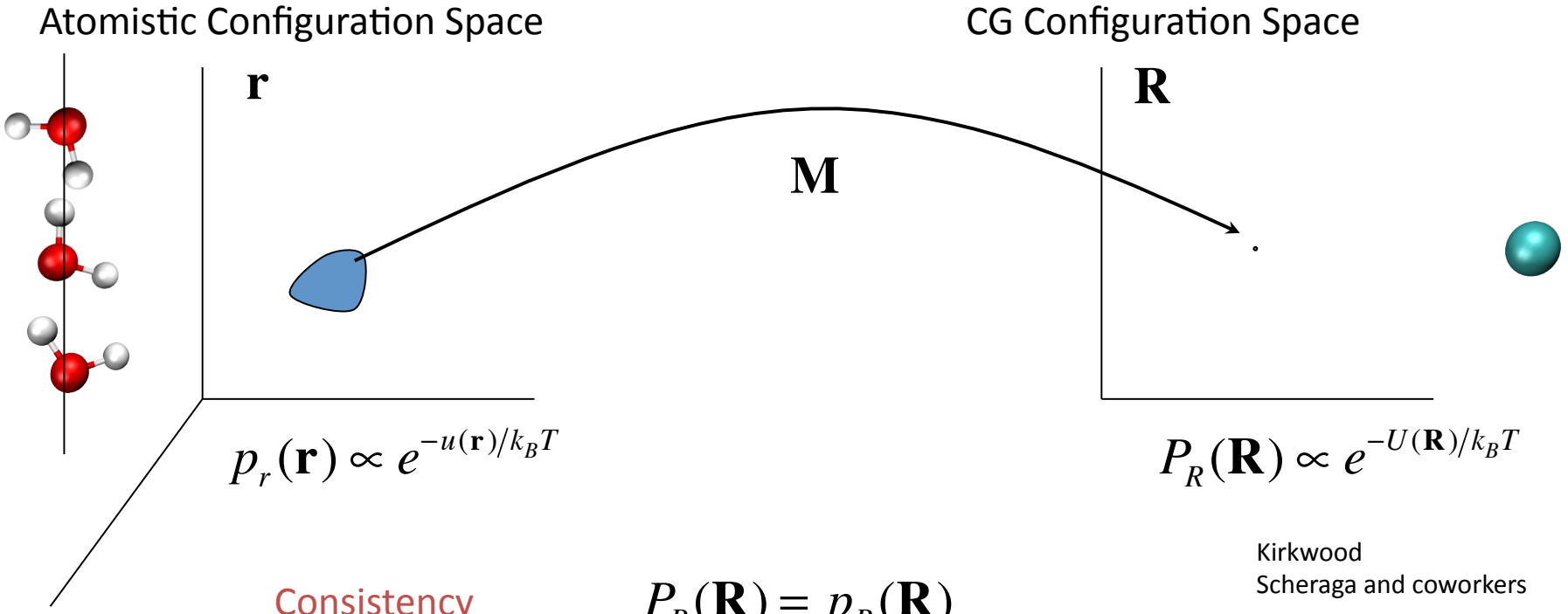
Coarse-grained (CG) Mapping



The mapping operator transforms an atomistic configuration onto a CG configuration by defining the coordinates of each site as a linear combination of the coordinates defining each site.

Noid, Chu, …, Voth, Andersen
J Chem Phys (2008)

The PMF: Structurally Consistent CG Models



Consistency $P_R(\mathbf{R}) = p_R(\mathbf{R})$

$$e^{-U(\mathbf{R})/k_BT} \propto \int d\mathbf{r}\; p_r(\mathbf{r})\,\delta\big(\mathbf{M}(\mathbf{r}) - \mathbf{R}\big)$$

Kirkwood
Scheraga and coworkers

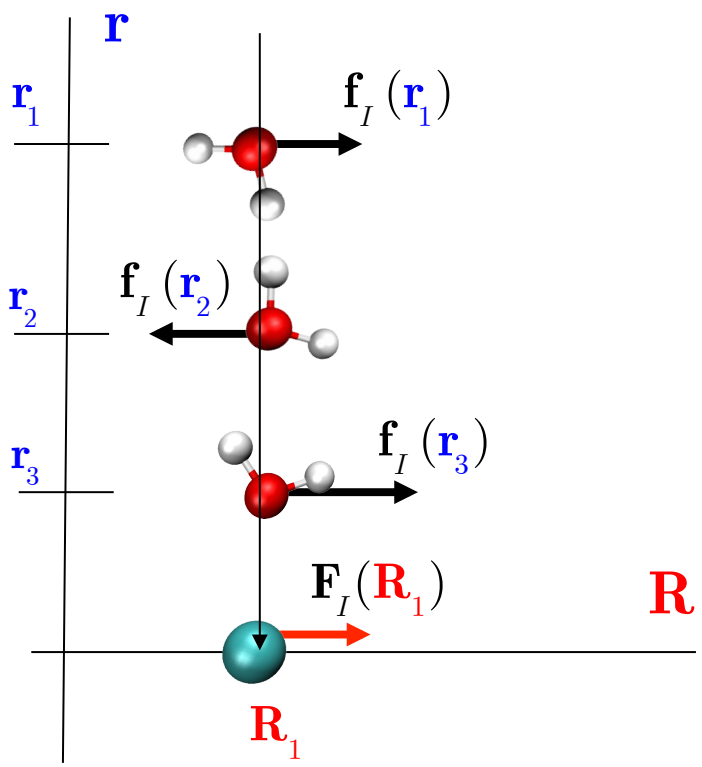
Noid, Chu, …, Voth, Andersen
J Chem Phys (2008)

For a consistent CG model that reproduces the distribution of structures generated by the atomistic model, the appropriate CG potential is a many-body PMF.

Mean Force Field

$$\mathbf{F}_I(\mathbf{R}) = \frac{-\partial U(\mathbf{R})}{\partial \mathbf{R}_I}$$

$$= \left\langle \mathbf{f}_I(\mathbf{r}) \right\rangle_{\mathbf{M}(\mathbf{r})=\mathbf{R}}$$



Atomistic FF:

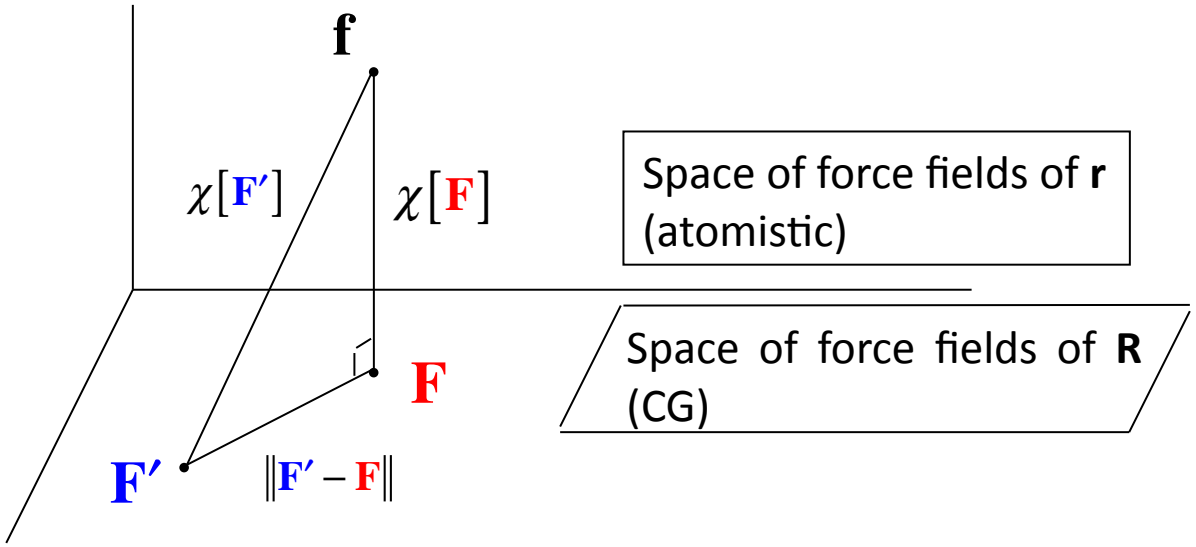
$$\mathbf{f}_I(\mathbf{r}) = \sum_{i \in I} \mathbf{f}_i(\mathbf{r})$$

$$\mathbf{R}_1 = \mathbf{M}(\mathbf{r}_1) = \mathbf{M}(\mathbf{r}_2) = \cdots$$

In a consistent model, the CG force field is the conditioned average of the atomistic force field (I.e., the mean force field). The mean force field is sufficient for a consistent CG model.

Variational Principle for Multiscale Coarse-graining

$$\chi^2[\mathbf{F}'] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \left| \mathbf{F}'_I(\mathbf{M}(\mathbf{r})) - \mathbf{f}_I(\mathbf{r}) \right|^2 \right\rangle$$
$$= \chi^2[\mathbf{F}] + \|\mathbf{F}' - \mathbf{F}\|^2$$



$$\chi[\mathbf{F}] = \|\mathbf{F} - \mathbf{f}\|$$
$$\chi[\mathbf{F}'] = \|\mathbf{F}' - \mathbf{f}\|$$

The Multiscale Coarse-graining (MS-CG) variational principle determines the many-body PMF through a geometric optimization problem in the space of CG force fields.

Izvekov and Voth.
J Phys Chem B (2005)
J Chem Phys (2005)

Noid, Chu, Ayton, Voth
J Phys Chem B (2007)
Noid, Chu, …, Voth, Andersen
J Chem Phys (2008)

See also Chorin 2003, 2006

Molecular Mechanics Basis Set

Approx. CG Potential

$$U(\mathbf{R}) = \sum_{I-J>4}^{pairs} U_{IJ}^{nb}(R_{IJ}) + \sum_{i}^{bonds} U_i^b(d_i) + \sum_{i}^{angles} U_i^\theta(\theta_i) + \sum_{i}^{dihedrals} U_i^\psi(\psi_i) + \cdots$$

Approx. CG Force field

$$\mathbf{F}_I(\mathbf{R}) = \sum_{I-J>4}^{pairs} F_{IJ}^{nb}(R_{IJ}) \frac{\partial R_{IJ}}{\partial \mathbf{R}_I} + \sum_{i}^{bonds} F_i^b(d_i) \frac{\partial d_i}{\partial \mathbf{R}_I} + \sum_{i}^{angles} F_i^\theta(\theta_i) \frac{\partial \theta_i}{\partial \mathbf{R}_I} + \cdots$$

Basis expansion

$$\mathbf{F} = \sum_{\zeta} \int dz \; F_\zeta(z) \mathbf{G}_\zeta(z)$$

Force function $F_\zeta(z) = -dU_\zeta(z)/dz$

Basis vector $\mathbf{G}_\zeta(z) = \left(\frac{\partial \psi_\zeta(\mathbf{R})}{\partial \mathbf{R}_I} \right) \delta\left(\psi_\zeta(\mathbf{R}) - z\right)$

Interactions ζ

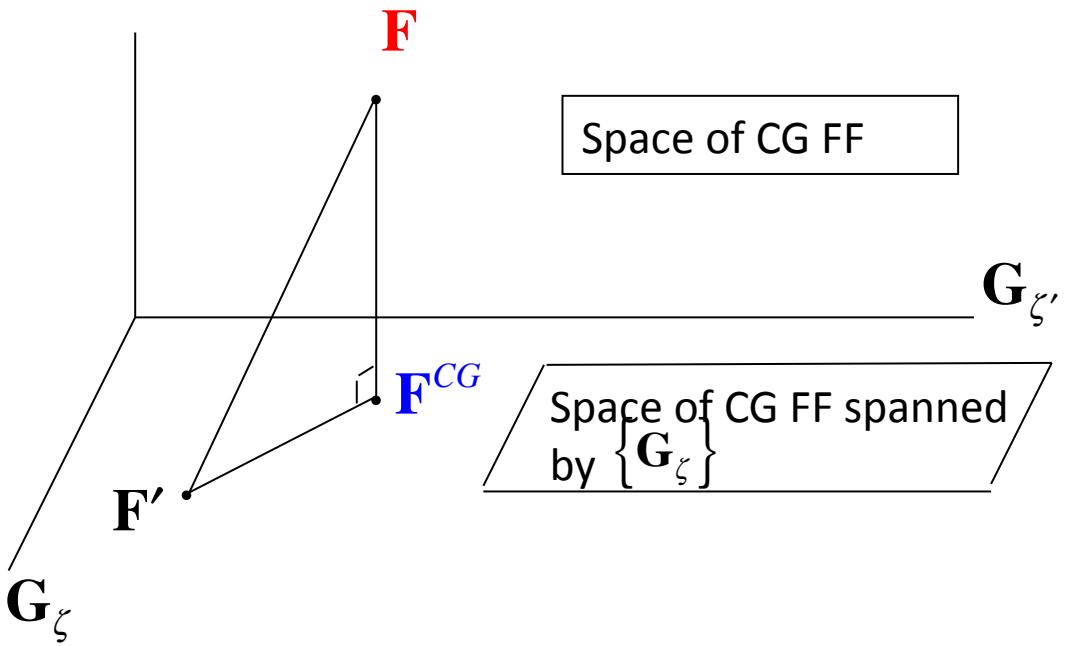
Noid, …, Chu, …, Andersen, Voth
J Chem Phys (2008)

An approximate CG potential determines a set of force field basis vectors

Linear Least Squares Problem

$$\mathbf{F}^{CG}(\mathbf{R}) = \sum_{\zeta} \int dz\, F_\zeta(z)\, \mathbf{G}_\zeta(\mathbf{R}; z)$$

$$\chi^2[F] = \frac{1}{3N} \left\langle \sum_{I=1}^{N} \left| \sum_{\zeta} \int dz\, F_\zeta(z)\, \mathbf{G}_{I;\zeta}\left(\mathbf{M}(\mathbf{r}); z\right) - \mathbf{f}_I(\mathbf{r}) \right|^2 \right\rangle$$



$\mathbf{F}^{CG}(\mathbf{R})$ CG FF

F_ζ Force functions

$\mathbf{G}_\zeta(\mathbf{R})$ FF basis fcns

The MS-CG variational principle determines F_ζ by projecting the PMF onto the space of CG force fields spanned by the given basis.

Geometric Projection

Basis expansion:

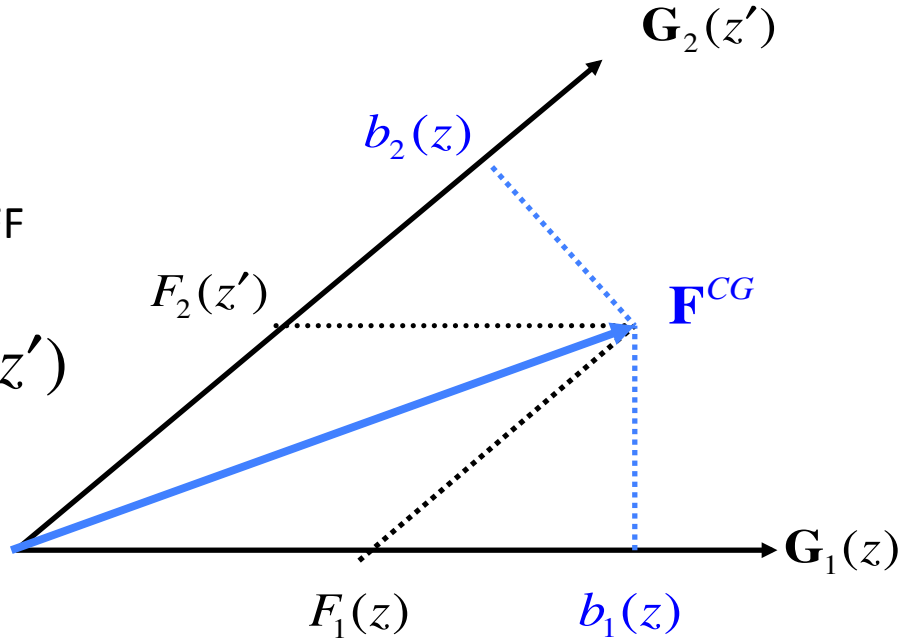
$$\mathbf{F}^{CG} = \sum_{\zeta} \int dz\, F_{\zeta}(z)\mathbf{G}_{\zeta}(z)$$

Projections:

$$b_{\zeta}(z) = \mathbf{G}_{\zeta}(z)\cdot\mathbf{F} \qquad \text{MF}$$

$$= \mathbf{G}_{\zeta}(z)\cdot\mathbf{F}^{CG} \qquad \text{Approx FF}$$

$$= \sum_{\zeta'} \int dz' G_{\zeta\zeta'}(z,z')F_{\zeta'}(z')$$



Metric Tensor:

$$G_{\zeta\zeta'}(z,z') = \mathbf{G}_{\zeta}(z)\cdot\mathbf{G}_{\zeta'}(z')$$

$$= \left\langle \sum_{I} \mathbf{G}_{I;\zeta}(\mathbf{M}(\mathbf{r});z)\cdot\mathbf{G}_{I;\zeta'}(\mathbf{M}(\mathbf{r});z') \right\rangle$$

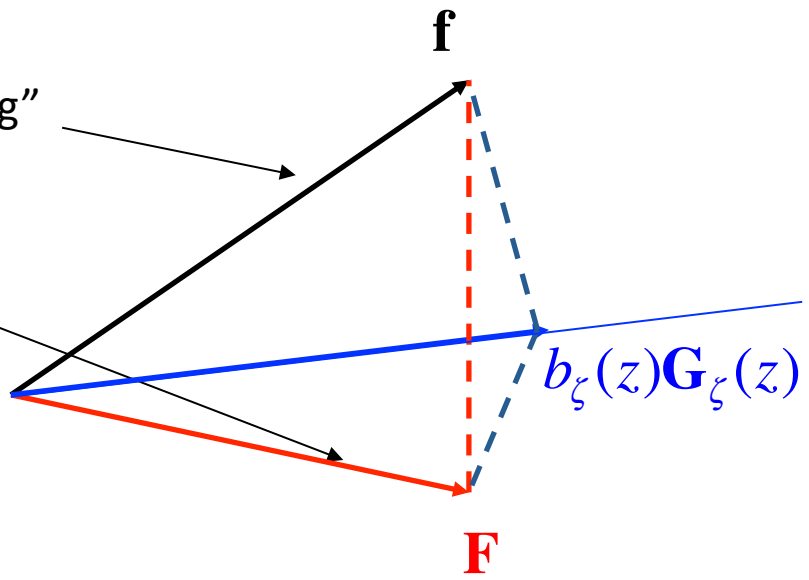
Noid, …, Andersen, Voth. *J Chem Phys* (2008)
Mullinax and Noid. *J Phys Chem C* (2010)
Mullinax and Noid *J Chem Phys* (2010)

The PMF is approximated by projecting the MF onto each basis vector, while treating the metric tensor resulting from many-body correlations.

Generalized Yvon-Born-Green Equation

Integral Eq $$b_\zeta(z) = \mathbf{G}_\zeta(z)\cdot\mathbf{F} = \mathbf{G}_\zeta(z)\cdot\mathbf{F}^{CG} = \sum_{\zeta'}\int dz' G_{\zeta\zeta'}(z,z')F_{\zeta'}(z')$$

$$\begin{aligned} b_\zeta(z) &= \mathbf{G}_\zeta(z)\cdot\mathbf{f} \quad \text{MS-CG "Force-Matching"} \\ &= \mathbf{G}_\zeta(z)\cdot\mathbf{F} \quad \text{MF} \\ &= \mathbf{G}_\zeta(z)\cdot\nabla\left(-k_BT\ln p_R(\mathbf{R})\right) \\ &= k_BT\, d\overline{g}_\zeta(z)/dz \end{aligned}$$



$$k_BT\, d\overline{g}_\zeta(z)/dz = \sum_{\zeta'}\int dz' G_{\zeta\zeta'}(z,z')F_{\zeta'}(z')$$

Mullinax and Noid.
Phys Rev Lett **103** 198104 (2009)
J Phys Chem C **114** 5661 (2010)

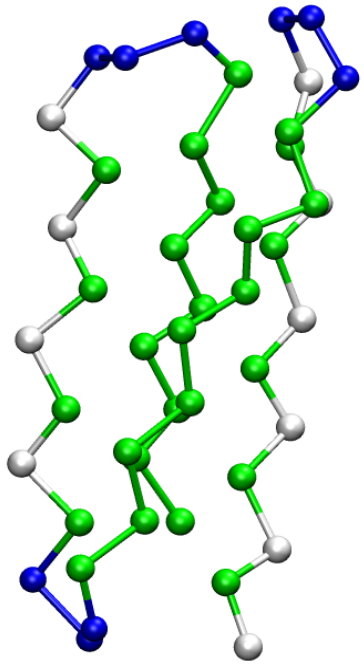
The generalized-YBG Equation determines the MS-CG potentials directly from structures!

Honeycutt-Thirumalai (HT) Model

Green: hydrophobic (B)

White: hydrophilic (L)

Blue: neutral (N)

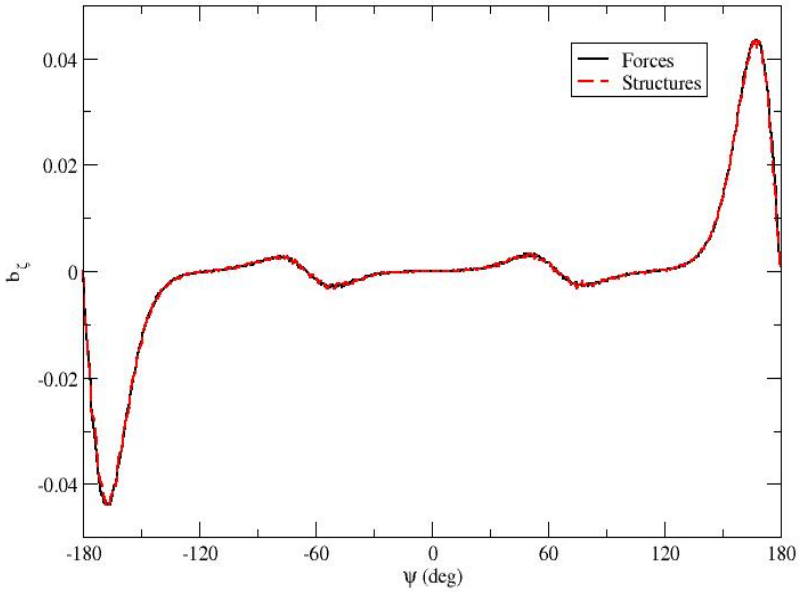


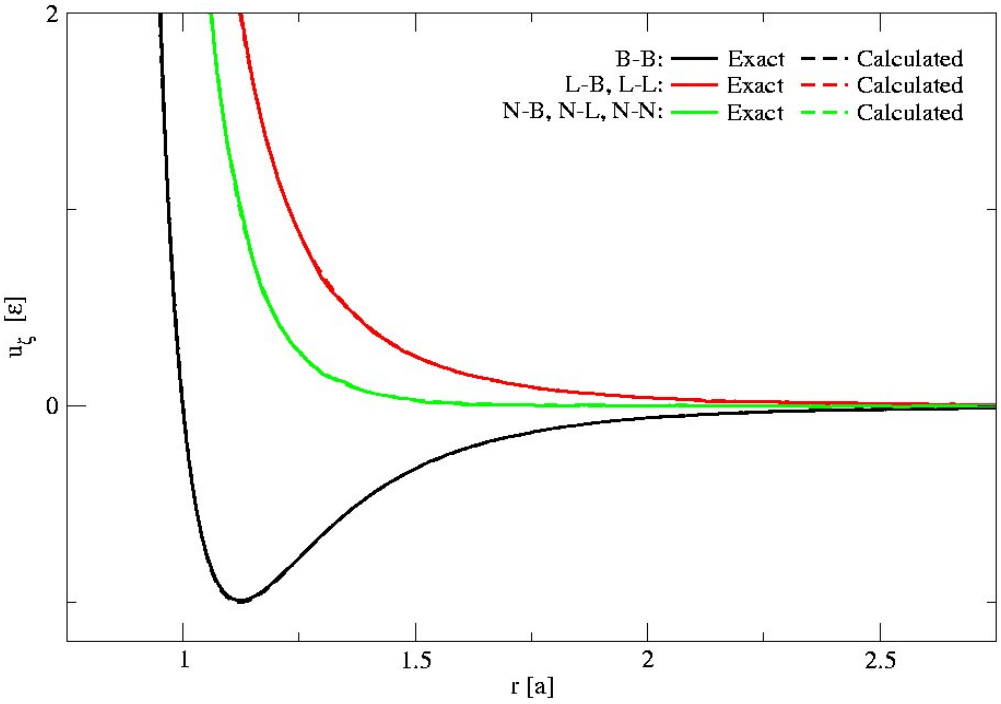
$$U(\mathbf{R}) = \sum_{I-J>4}^{pairs} U_{IJ}^{nb}(R_{IJ}) + \sum_{i}^{bonds} U_i^b(d_i) + \sum_{i}^{angles} U_i^\theta(\theta_i) + \sum_{i}^{dihedrals} U_i^\psi(\psi_i)$$

Honeycutt and Thirumalai *Biopolymers* (1992) **32**, 695

HT Results 1

$$b_\zeta(z) = k_B T \, d\overline{g}_\zeta(z)/dz$$





Mullinax and Noid
Phys Rev Lett **131** 198104 (2009)

First generalization of the YBG theory for proteins with many-body, e.g., torsional and angle, interactions.

2. Precise Definition of Transferability

Model:

(1) Topology
- Particles and bonds used to describe system

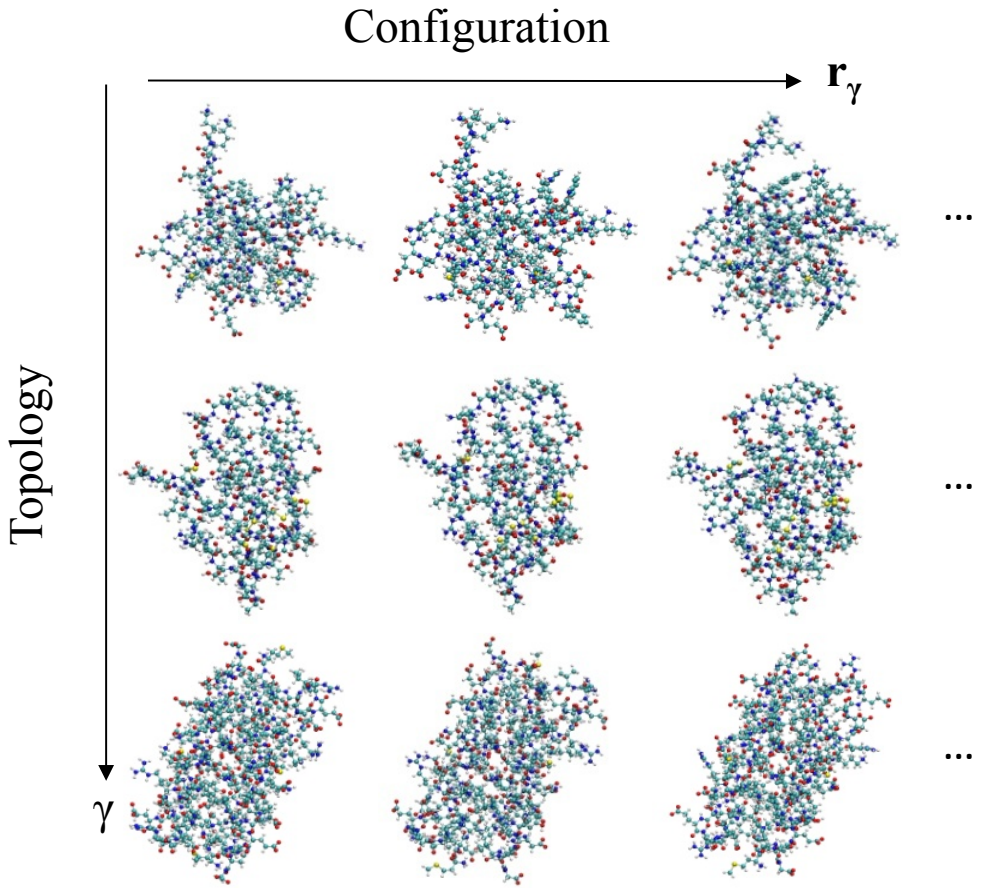
(2) Potential
- Interactions among those particles

atomistic	CG
γ	Γ
$u_\gamma(\mathbf{r}_\gamma)$	$U_\Gamma(\mathbf{R}_\Gamma)$

A potential is transferable if it can be used for describing multiple topologies.

Mullinax and Noid. *J Chem Phys* 2009.

Extended Ensemble



An **extended ensemble** is a collection of equilibrium ensembles for different **topologies.**

Distributions:

p_γ	topology
$p_{r\|\gamma}(\mathbf{r}_\gamma)$	configuration

Averages: $$\left\langle a_\gamma(\mathbf{r}_\gamma)\right\rangle = \sum_\gamma p_\gamma \int d\mathbf{r}_\gamma\, p_{r|\gamma}(\mathbf{r}_\gamma) a_\gamma(\mathbf{r}_\gamma)$$

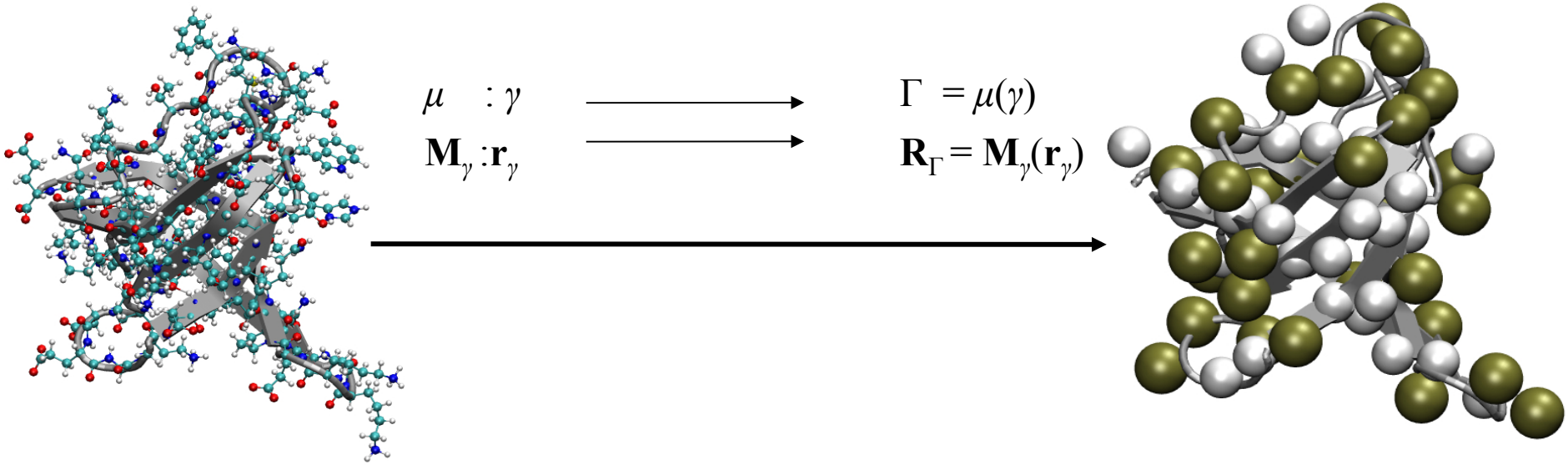
Mullinax and Noid. *J Chem Phys* 2009.

Mappings

Constructing a CG model requires **two** maps:

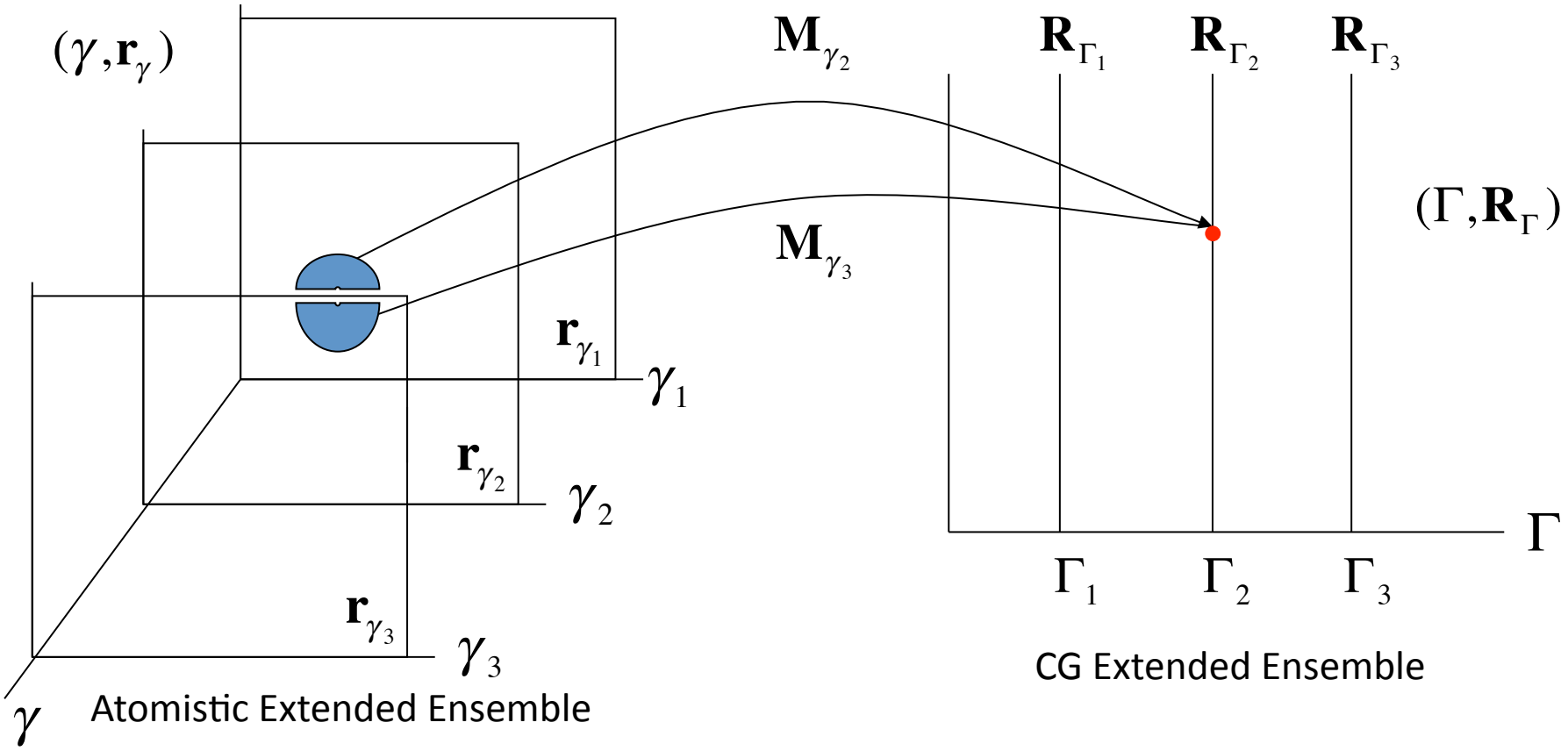
(1) Topology map - specifying site types and bonds

(2) Configuration map - specifying site coordinates



Then the remaining challenge is to determine $U_\Gamma(\mathbf{R}_\Gamma)$

Consistency between extended ensembles



Consistency:

$$P_\Gamma = \left\langle \delta_{\Gamma,\mu(\gamma)} \right\rangle$$

$$\exp[-U_\Gamma(\mathbf{R}_\Gamma)/k_B T] \propto \left\langle \delta_{\Gamma,\mu(\gamma)} \delta\left(\mathbf{R}_\Gamma - \mathbf{M}_\gamma(\mathbf{r}_\gamma)\right) \right\rangle$$

Generalized PMF

Mullinax and Noid. *J Chem Phys* 2009.

Variational Principle for the generalized PMF

$$\chi^2[\mathbf{F}'] = \left\langle \sum_{I=1}^{N_{\mu(\gamma)}} \left| \mathbf{F}'_{\mu(\gamma);I}\left(\mathbf{M}_\gamma(\mathbf{r}_\gamma)\right) - \mathbf{f}_{\gamma;I}(\mathbf{r}_\gamma) \right|^2 \right\rangle$$

$$= \chi^2[\mathbf{F}] + \|\mathbf{F}' - \mathbf{F}\|^2$$

where $\mathbf{F}_{\Gamma I}(\mathbf{R}_\Gamma) = -\nabla_{\Gamma I} U_\Gamma(\mathbf{R}_\Gamma)$ is a mean force field

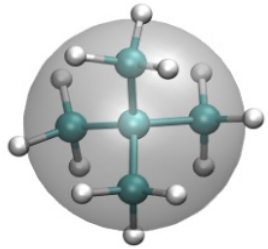
Key approximation

$$U_\Gamma(\mathbf{R}_\Gamma) \approx \sum_{\zeta \in \Gamma} U_\zeta\left(\psi_\zeta(\mathbf{R}_\Gamma)\right)$$

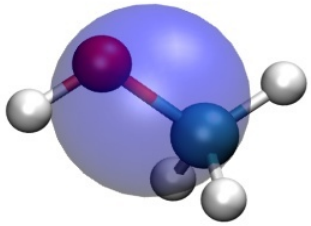
Determine optimal transferable approximation to the PMF

Methanol-Neopentane Test System

M1N1

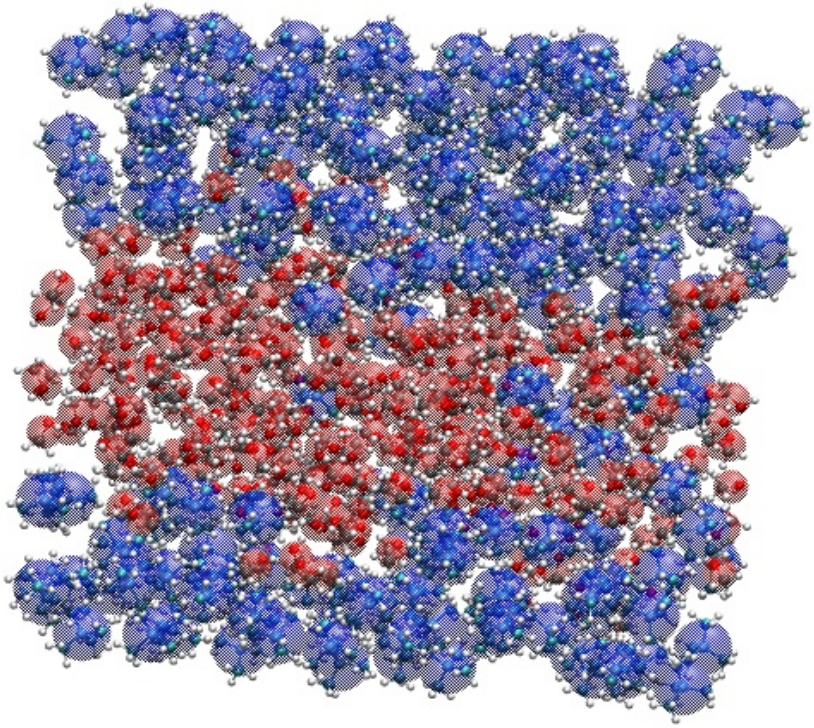


neopentane



methanol

Percent methanol (%)	Methanol	Neopentane
100	968	0
80	574	144
60	342	228
50	259	259
40	189	284
20	81	323
0	0	353



CG Potential

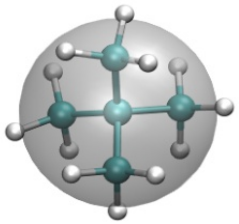
$$U_{\Gamma}(\mathbf{R}_{\Gamma}) = \sum_{\{I,J\}\in\Gamma} U^{(2)}_{\zeta(I,J)}(R_{IJ})$$

Mullinax and Noid. *J Chem Phys* 2009.

M-N: Results 1

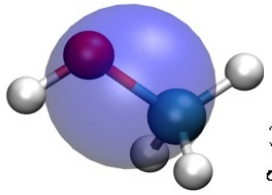
M1N1

N

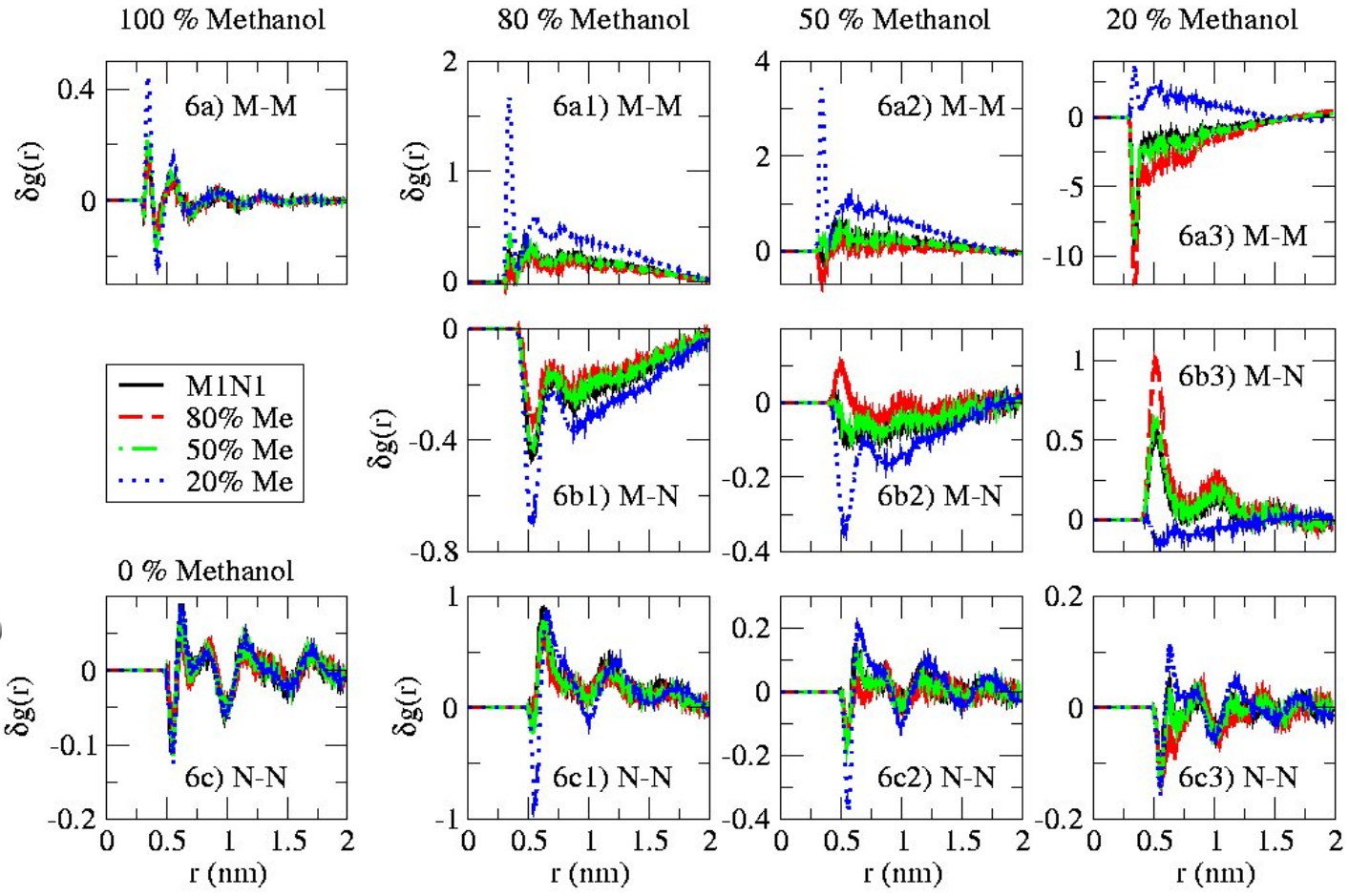


neopentane

M



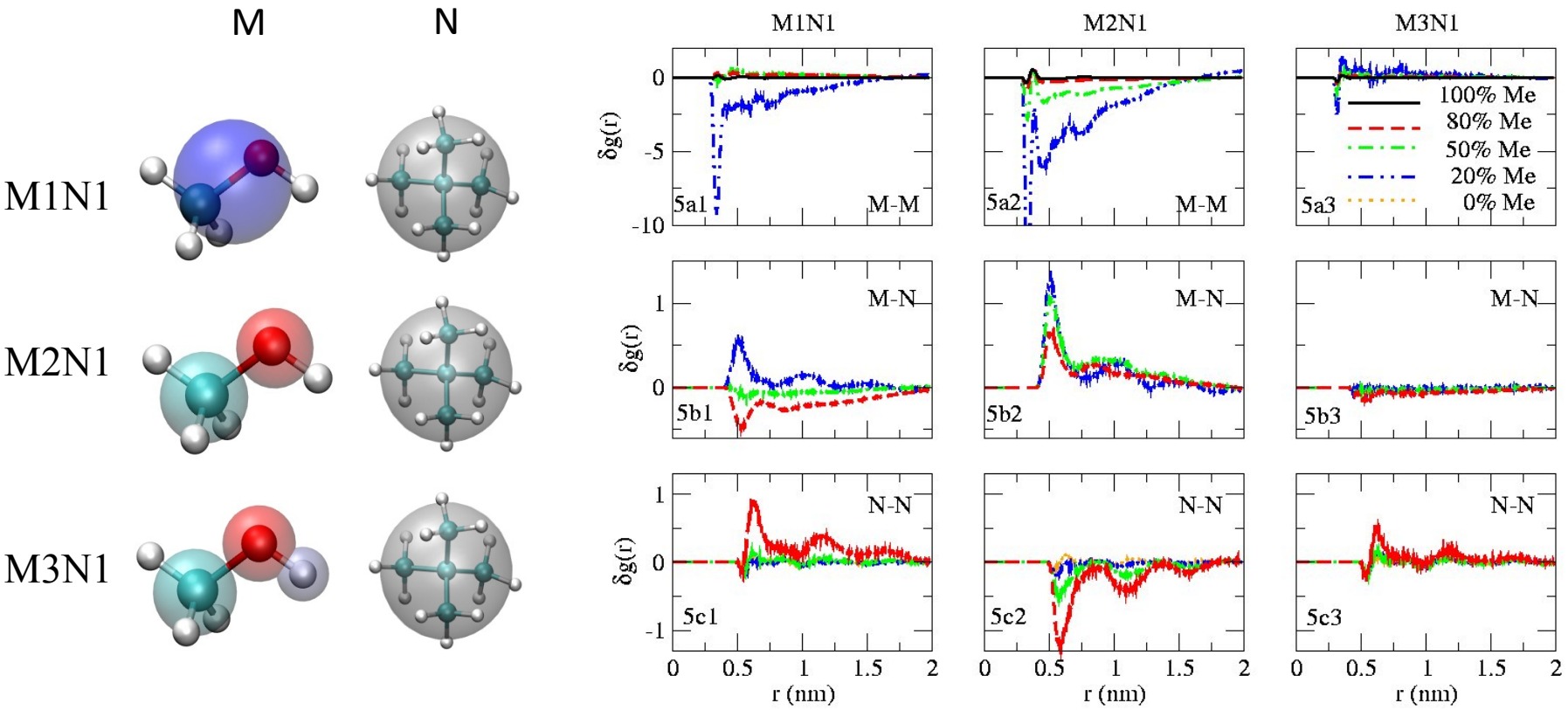
methanol



Mullinax and Noid. *J Chem Phys* 2009.

Extended ensemble potentials provide improved transferability.

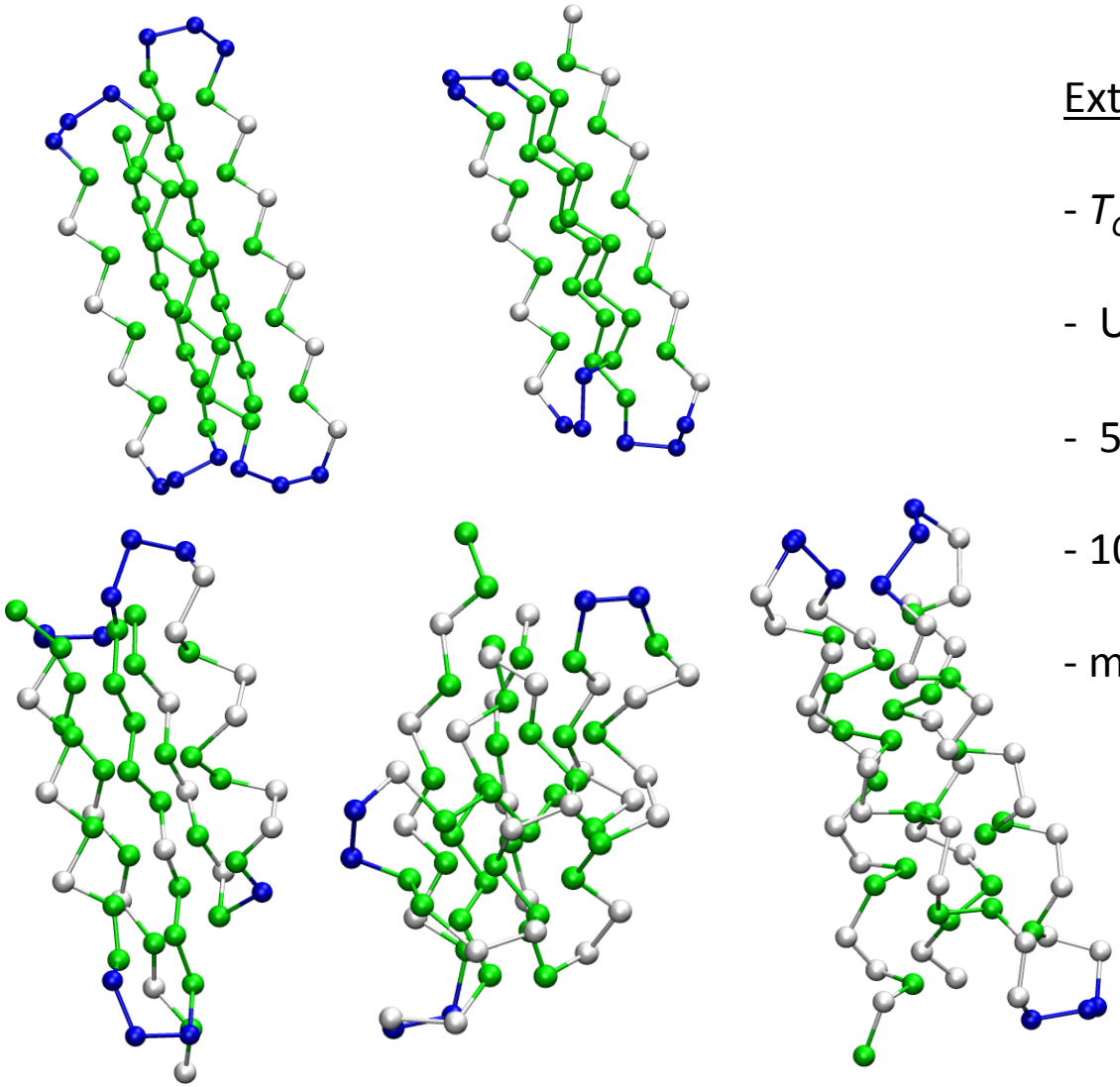
M-N: Results 2



Mullinax and Noid. *J Chem Phys* 2009.

The accuracy and transferability of the potentials are sensitive to the topology mapping.

Model Protein Databank



Extended Ensemble

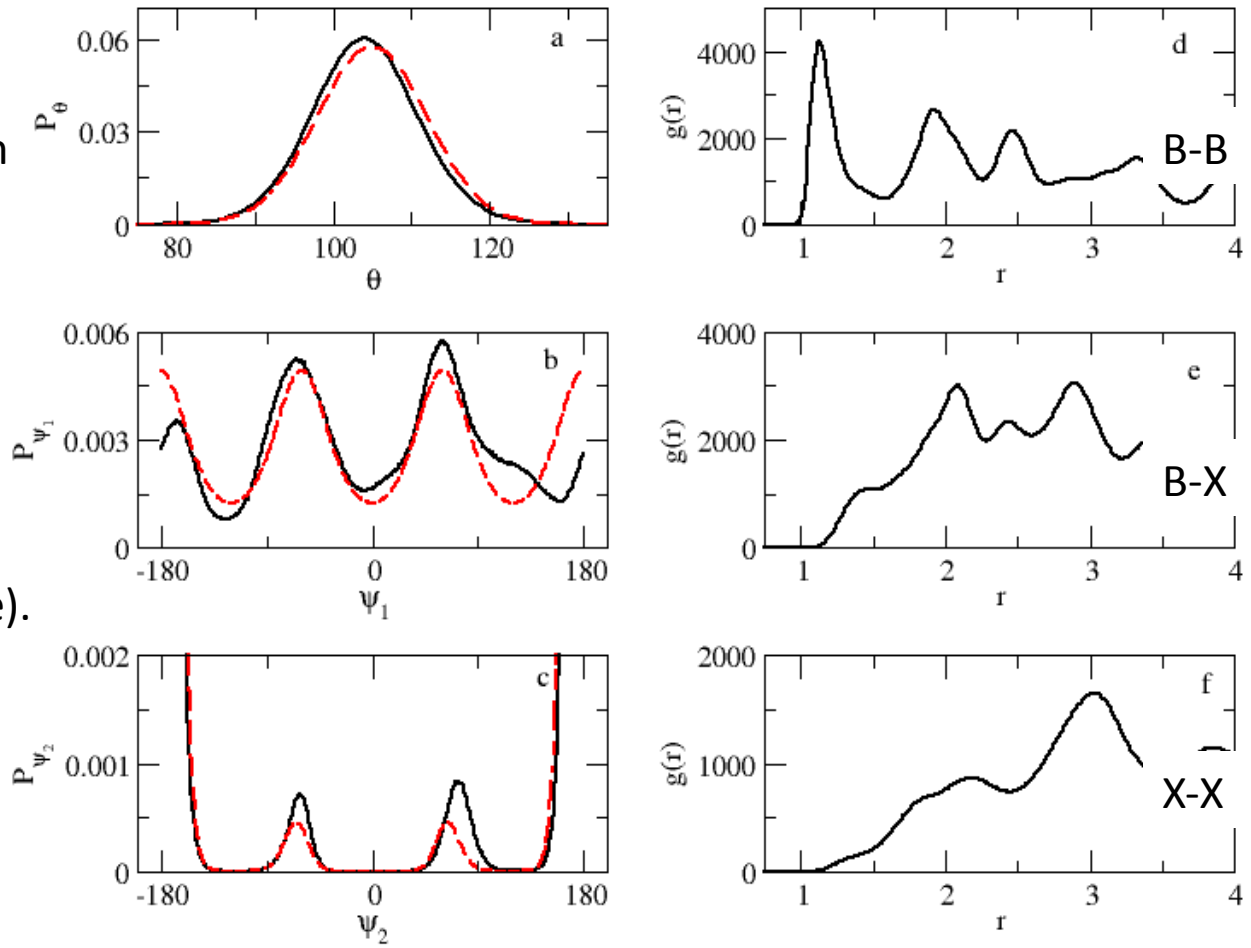
- $T_G < T < T_F$
- Uniform topology distribution
- 5 sequences
- 10^5 structures / sequence
- modified HT potential

Honeycutt and Thirumalai
Biopolymers (1992) **32**, 695

Distributions from Model PDB

1. Soft degrees of freedom couple to other degrees of freedom.

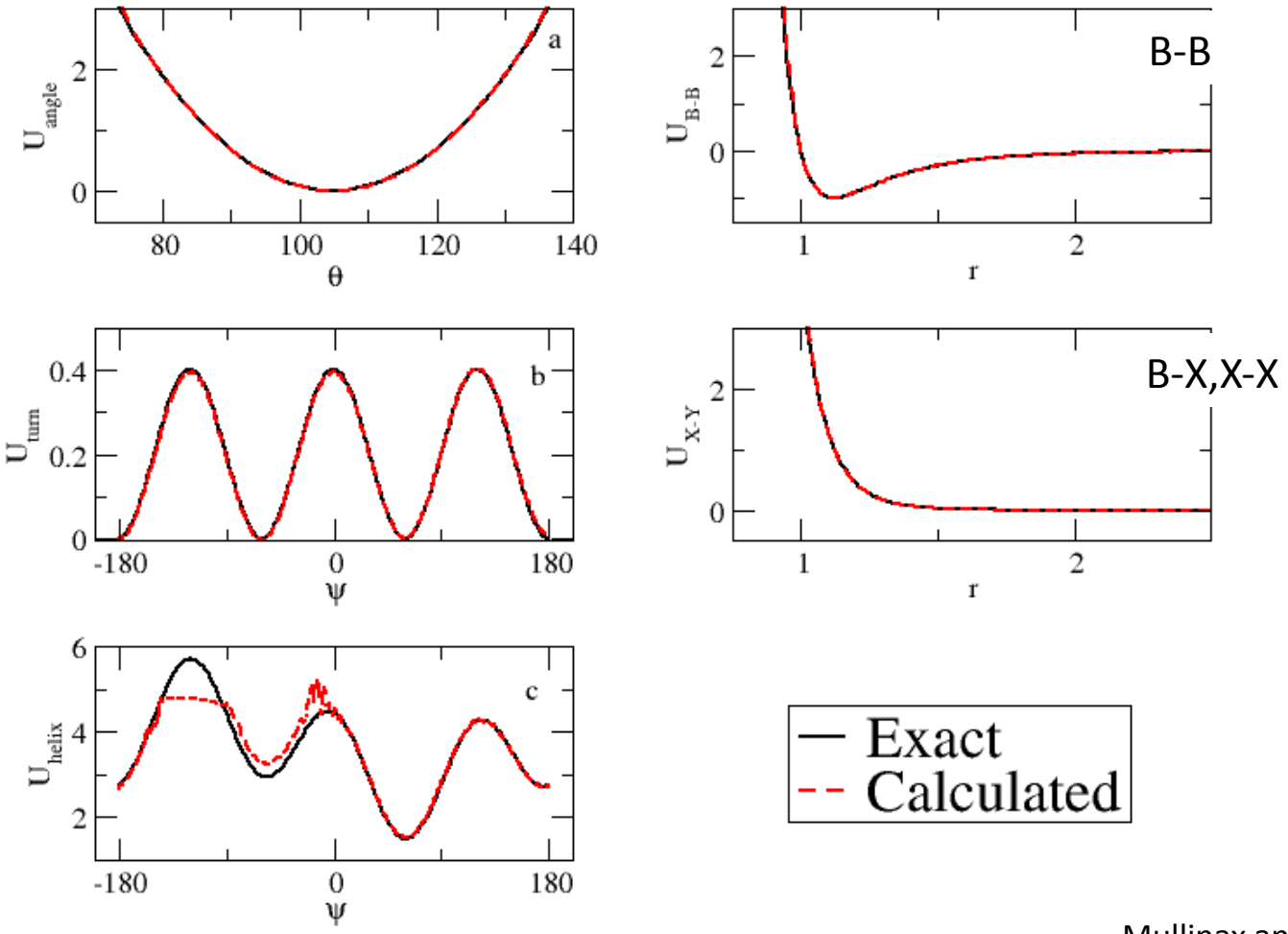
2. Chain connectivity generates long-ranged effective interactions between B-X and X-X pairs (which are purely repulsive).



Mullinax and Noid. *PNAS* **107** 19867 (2010)

Soft degrees of freedom are strongly coupled and cannot be treated independently.

Validation



The generalized-YBG theory quantitatively determines the underlying potentials for a model extended ensemble of folded protein structures.

Mullinax and Noid.
JCP **131** 104110 (2009)
PRL **103** 198104 (2009)
PNAS **107** 19867 (2010)

Relative Entropy

$\Phi(\mathbf{R})$ Information content in configuration **R** for distinguishing atomistic and CG distributions

Atomistic

$$\Phi(\mathbf{R}|U) = \ln\left[\frac{p_R(\mathbf{R})}{P_R(\mathbf{R}|U)}\right]$$

CG

$$0 \quad \text{if} \quad p_R(\mathbf{R}) = P_R(\mathbf{R}|U)$$

$$\pm\infty \quad \text{if} \quad p_R(\mathbf{R})/P_R(\mathbf{R}|U) \to \infty \text{ or } 0$$

Relative Entropy:
(Kullback-Leibler divergence)

$$S_{\mathrm{Rel}}[U] = \int d\mathbf{R}\; p_R(\mathbf{R})\Phi(\mathbf{R}|U)$$

$$\delta S_{\mathrm{Rel}}[U]/\delta U_\zeta(z) = \left(p_\zeta(z) - P_\zeta(z|U)\right)/k_BT$$

Considering variations w.r.t. CG potential $U_\zeta(z)$

1. The Relative Entropy is minimized when the conjugate distribution is reproduced
2. Minimizing the Relative entropy via Newton's method leads to IMC equations

References: Kullback & Leibler *Ann Math Stat* (**1951**); Shell *JCP* (**2008,2010**); Murtola et al. *JCP* (**2009**)

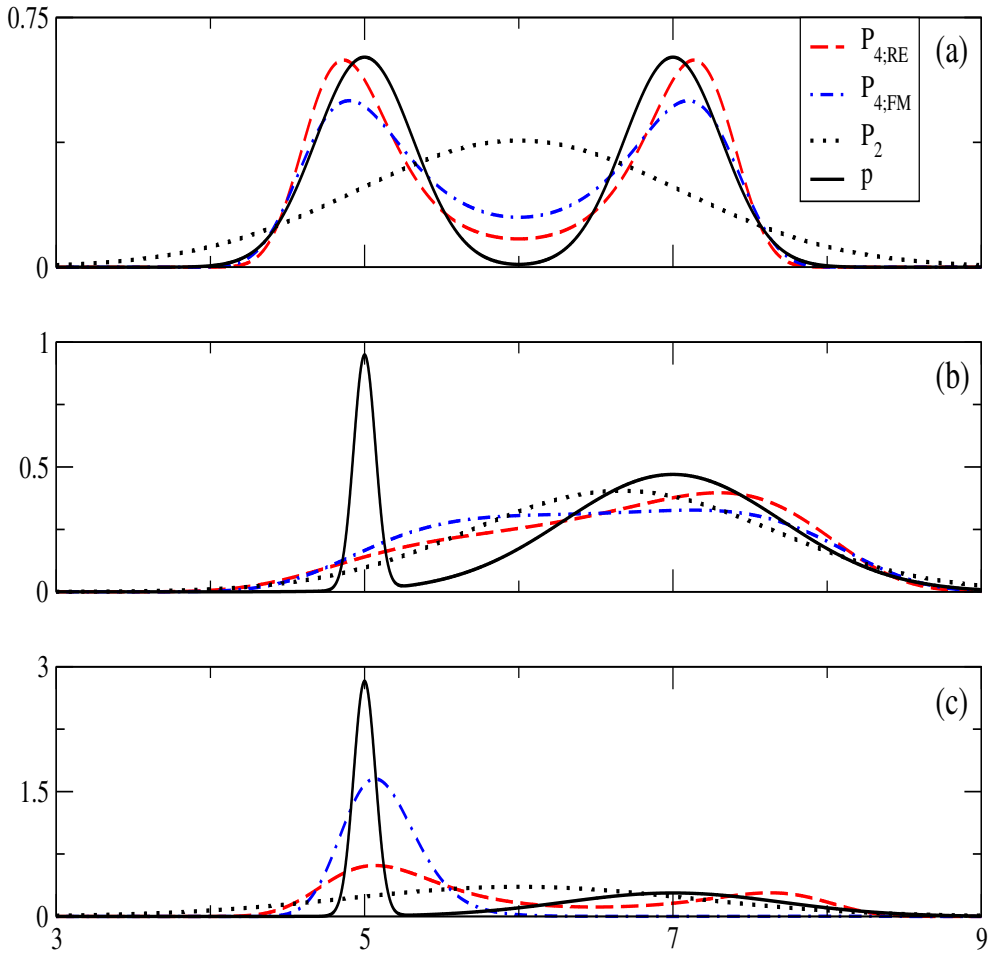
Relation to the Relative Entropy

Inverse Monte Carlo (Relative Entropy) functional:

$$S_{\mathrm{Rel}}[U] = k_B \int d\mathbf{R}\; p_R(\mathbf{R}) \Phi(\mathbf{R}|U)$$

Multiscale Coarse-graining "force-matching" functional

$$\chi^2[U] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \left|\mathbf{F}'_I(\mathbf{M}(\mathbf{r})) - \mathbf{f}_I(\mathbf{r})\right|^2 \right\rangle$$
$$= \chi^2\left[U^0\right]$$
$$+ \frac{(k_B T)^2}{3N} \int d\mathbf{R}\; p_R(\mathbf{R}) \left|\nabla \Phi(\mathbf{R}|U)\right|^2$$



Both the MS-CG "force-matching" and Inverse Monte Carlo approaches can be expressed in terms of the Kullback-Leibler information function.

Rudzinski and Noid *JCP* **2011**

Additional results

1. Equivalence of Force- and Structure-based potentials for quadratic potentials

2. Remarkable parallels in formulation:
   Variational problems in linear space with bases that are related by differentiation

3. Generalization of Henderson's uniqueness theorem

   1. Conditions – Linear independence of conjugate density operators
   2. Relation to force-matching uniqueness:
      Uniqueness of force-matching implies uniqueness of structure-based potential

4. Generalization of force-matching and g-YBG theory for arbitrary potentials

5. Entropy changes in coarse-graining:

$$s_{\mathbf{r}} = -k_B \int d\mathbf{r}\; p_r(\mathbf{r}) \ln\left[V^n p_r(\mathbf{r})\right]$$

$$s_{\mathbf{R}} = -k_B \int d\mathbf{R}\; p_R(\mathbf{R}) \ln\left[V^N p_R(\mathbf{R})\right]$$

$$S_{map} = s_{\mathbf{r}} - s_{\mathbf{R}}$$

$$= \int d\mathbf{R}\; p_R(\mathbf{R})\; k_B \left\langle \ln\left[\frac{V^N p_R(\mathbf{R})}{V^n p_r(\mathbf{r})}\right]\right\rangle_{\mathbf{R}}$$

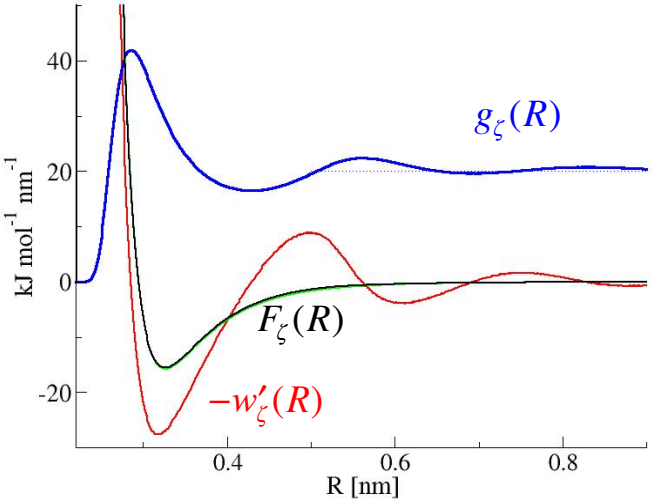
$$\leq 0$$

Rudzinski and Noid *JCP* **2011**
Noid and Shell In progress

Mean forces

Generalized YBG theory:

$$b_\zeta(z) = k_B T \, d\bar{g}_\zeta(z)/dz = \sum_{\zeta'} \int dz' G_{\zeta\zeta'}(z,z') F_{\zeta'}(z')$$



$$-w'_\zeta(R) = F_\zeta(R) + \sum_{\zeta'} \int dz' \, F_{\zeta'}(z') \; \bar{G}_{\zeta\zeta'}(R,z')/\bar{g}_\zeta(R)$$

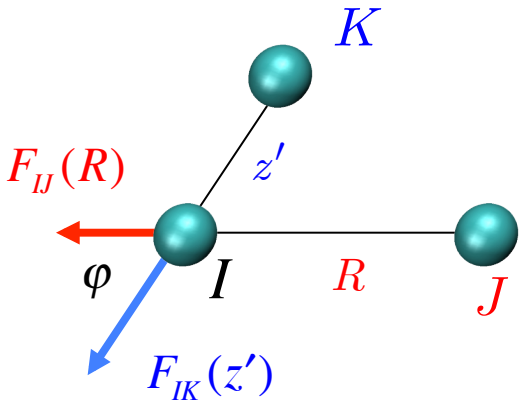
pair MF

direct

indirect

CG pair force

conditioned
3-particle density



The generalized Yvon-Born-Green equation determines the CG potential that reproduces the mean force (when using atomistic configurations).

Ellis, Rudzinski, and Noid
Macromol Sim Theory (2011)

Iterative Boltzmann Inversion

First estimate:

$$i = 0 \quad U_\zeta^0(z) = w_\zeta(z) = -k_B T \ln\left(p_\zeta(z) / J_\zeta(z) \right)$$

Corresponding pmf

CG MD

$$P_\zeta(z \mid U^i) \neq p_\zeta(z)$$

Error in distribution

Error in pmf:

$$w_\zeta(z) - W_\zeta^i(z) = -k_B T \ln\left[p_\zeta(z) / P_\zeta(z \mid U^i) \right]$$

Improve pmf:

$$i = 0,\ldots \quad U_\zeta^{i+1}(z) = U_\zeta^i(z) - k_B T \ln\left[p_\zeta(z) / P_\zeta(z \mid U^i) \right]$$

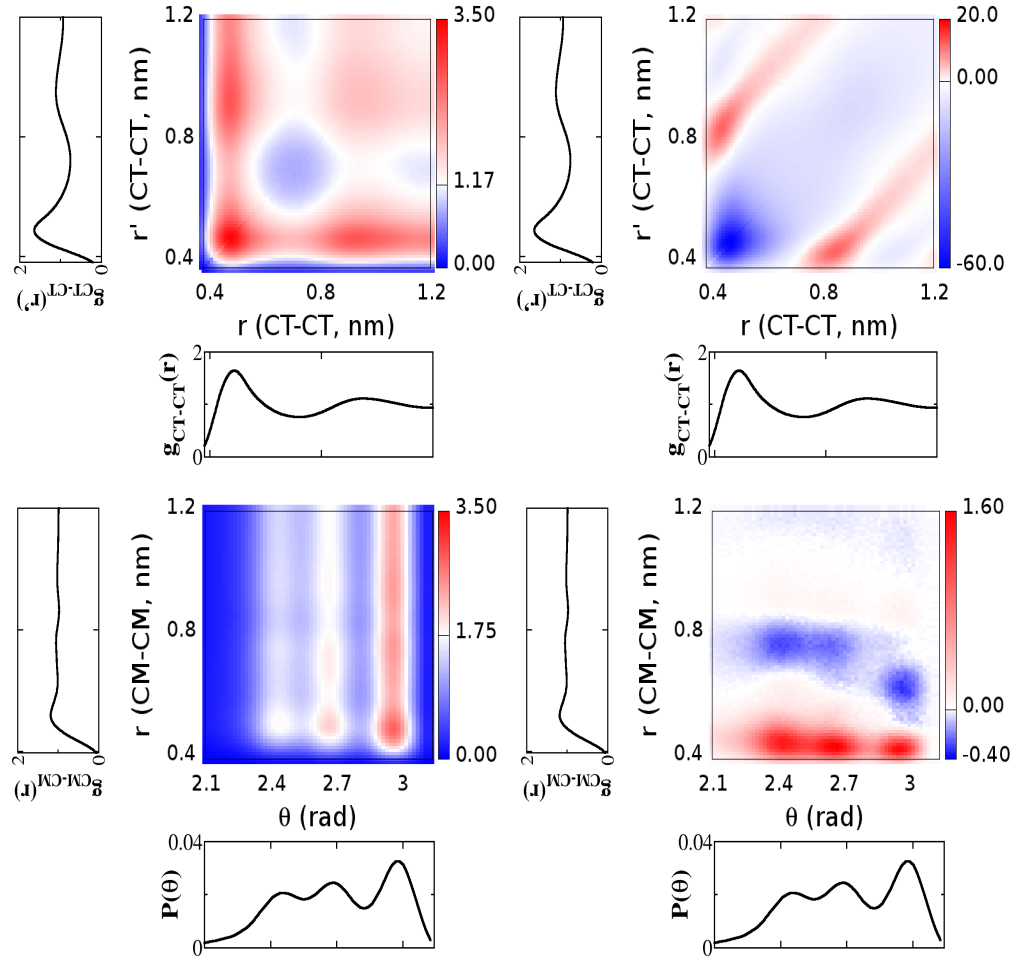
Iterate to convergence !

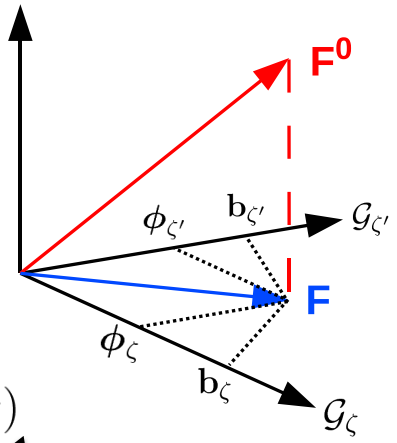
References:

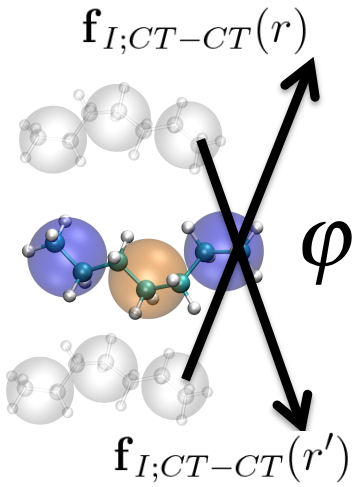
- Schommers *Phys Rev A* (1983) **28** 3599
- Soper *Chem Phys* (1996) **202** 295
- Muller-Plathe *ChemPhysChem* (2002) **9** 754
- Faller, and others
- Majek and Elber *Proteins* (2009) **76** 930

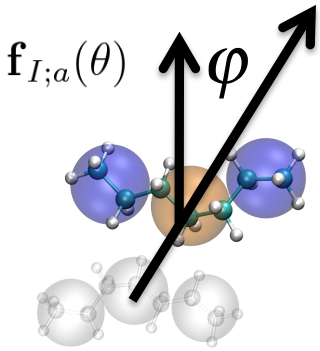
Understanding the Metric Tensor

$P_{\zeta\zeta'}(r,r')$ $\quad \bar{G}_{\zeta\zeta'}(r,r') = \langle \cos\varphi \rangle_{r,r'}$









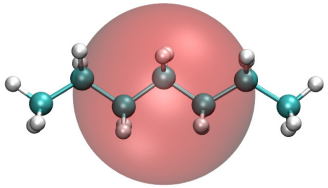
$\mathbf{f}_{I;CM-CM}(r)$

Rudzinski and Noid *JPCB* **2012**

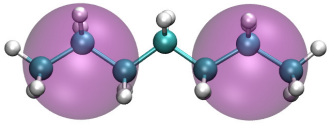
Robust features

$$\bar{G}_{\zeta\zeta'}(r,r') = \langle \cos\varphi \rangle_{r,r'}$$

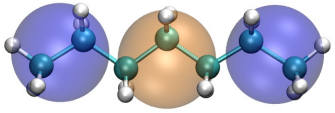
(a) CC

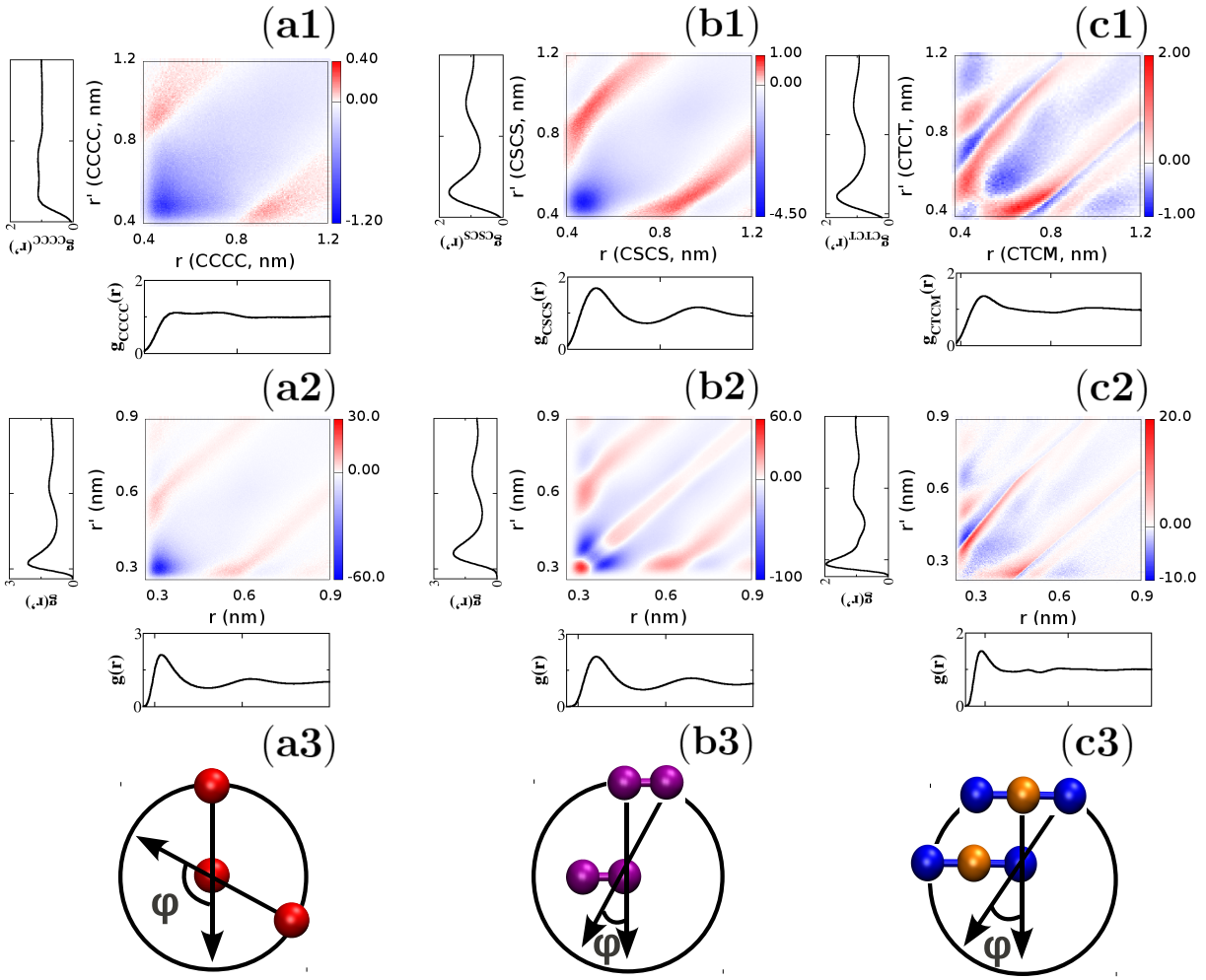


(b) CS-CS



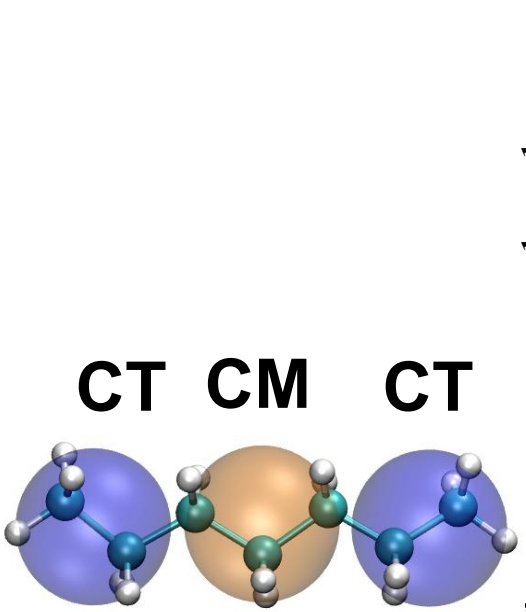
(c) CT-CM-CT

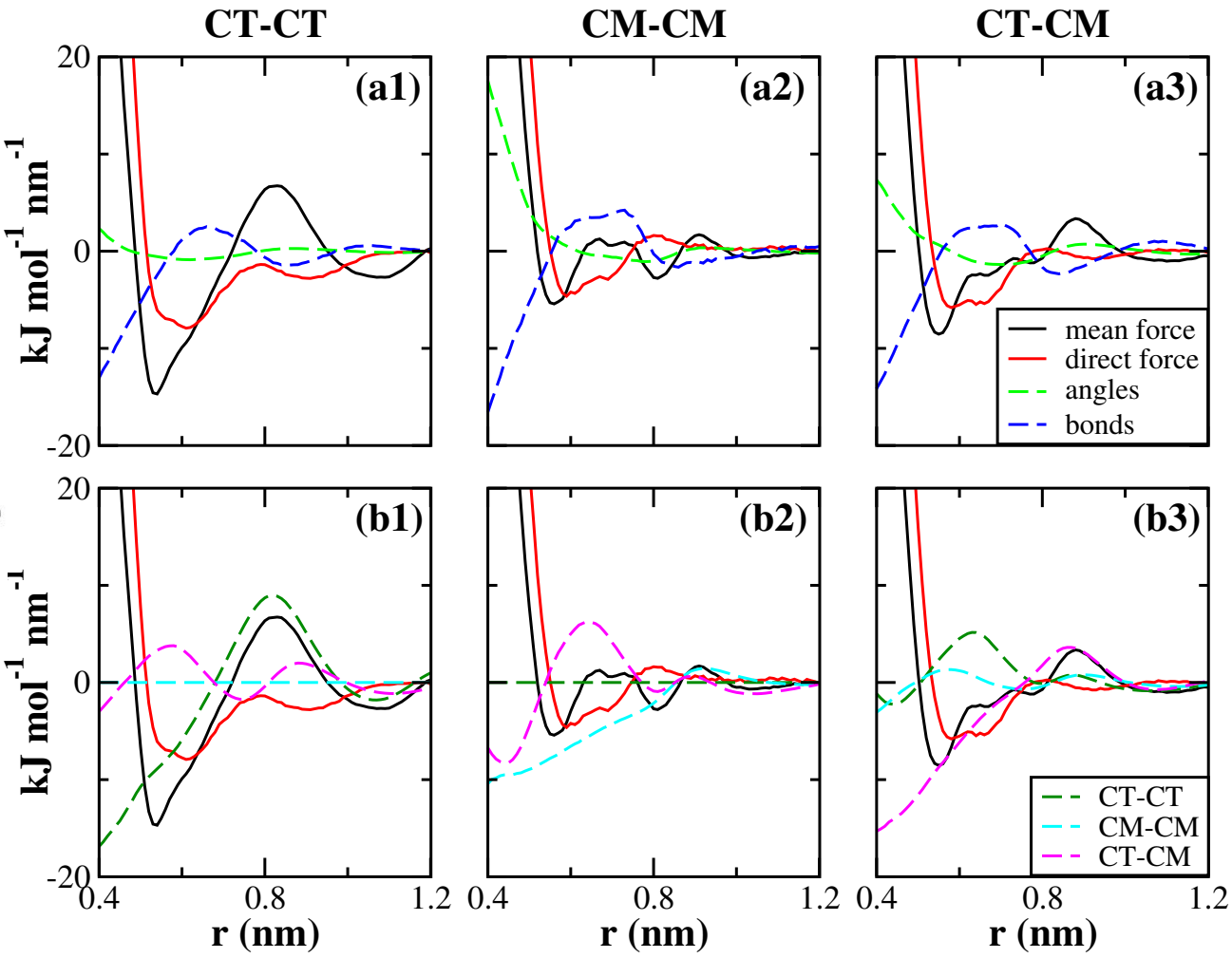




Rudzinski and Noid *JPCB* **2012**

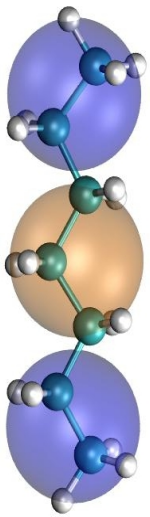
Decomposition of mean forces





Rudzinski and Noid *JPCB* **2012**

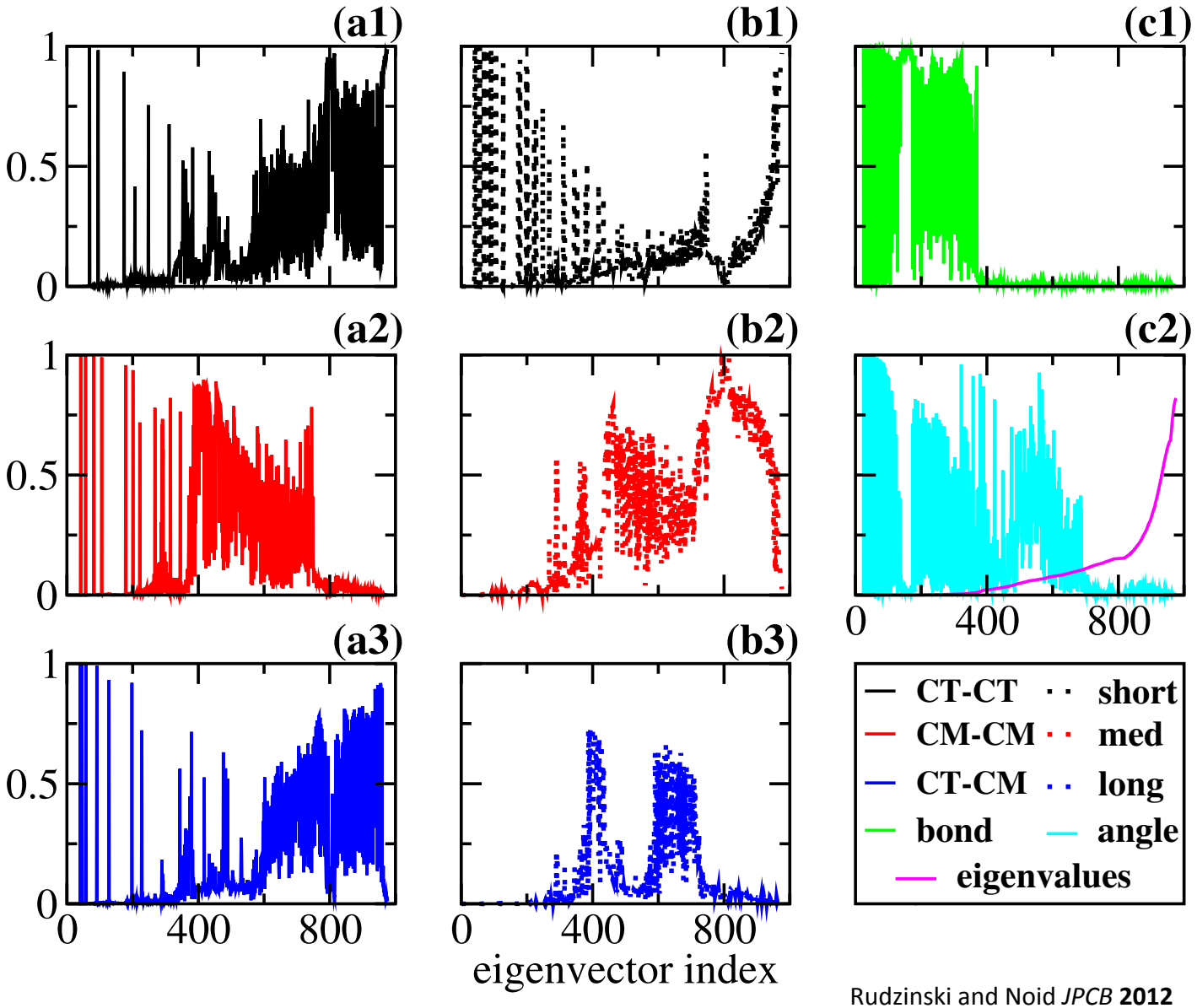
Eigenspectrum of metric tensor



CT

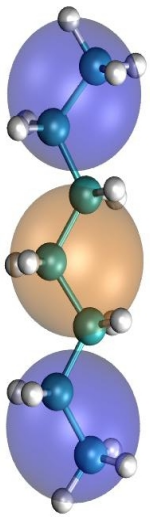
CM

CT



Rudzinski and Noid *JPCB* **2012**

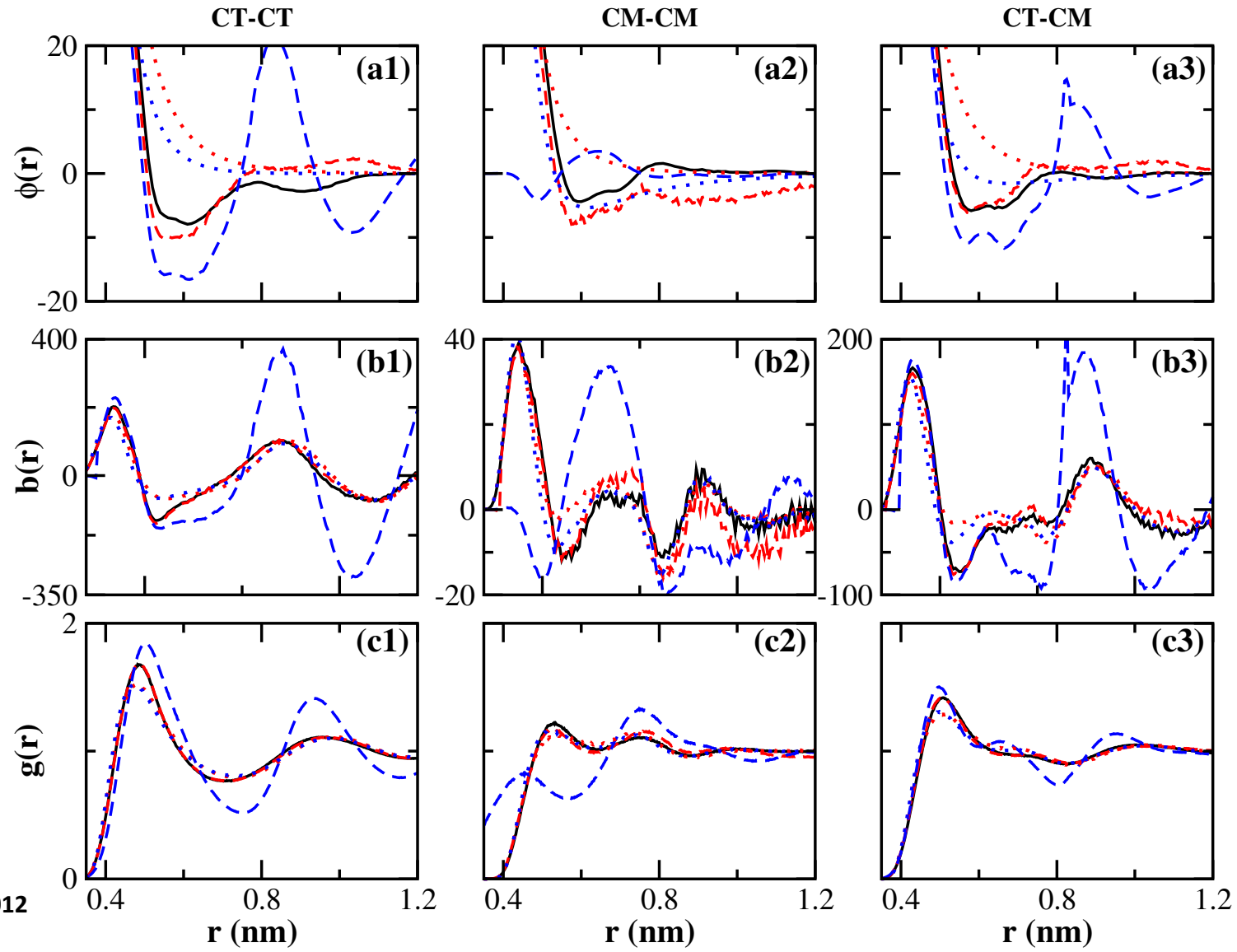
Impact upon CG structure



CT

CM

CT



Rudzinski and Noid *JPCB* **2012**

Conclusions

1. The generalized-YBG theory determines variationally optimal potentials (i.e., MS-CG potentials) directly (i.e., noniteratively) from structures.

2. The extended ensemble framework systematically and quantitatively improves the transferability of CG potentials for accurately modeling multiple chemically distinct systems.

3. In combination, these approaches provide a rigorous and accurate approach for determining physics-based potentials from a databank of protein structures.

4. The MS-CG/g-YBG method can be related to gradients of the information function that provides a variational basis for structure-based coarse-graining.

5. Mean forces provide a basis for connecting force and structure-based coarse-graining approaches.