How Structure Determines Coarse-grained Potentials and Vice Versa

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

KITP informal talk

May 17, 2012

Acknowledgements

Joe Rudzinski
Chris Ellis
Sushant Kumar
Tommy Foley
Nick Dunn

Wayne Mullinax

National Science Foundation
WHERE DISCOVERIES BEGIN

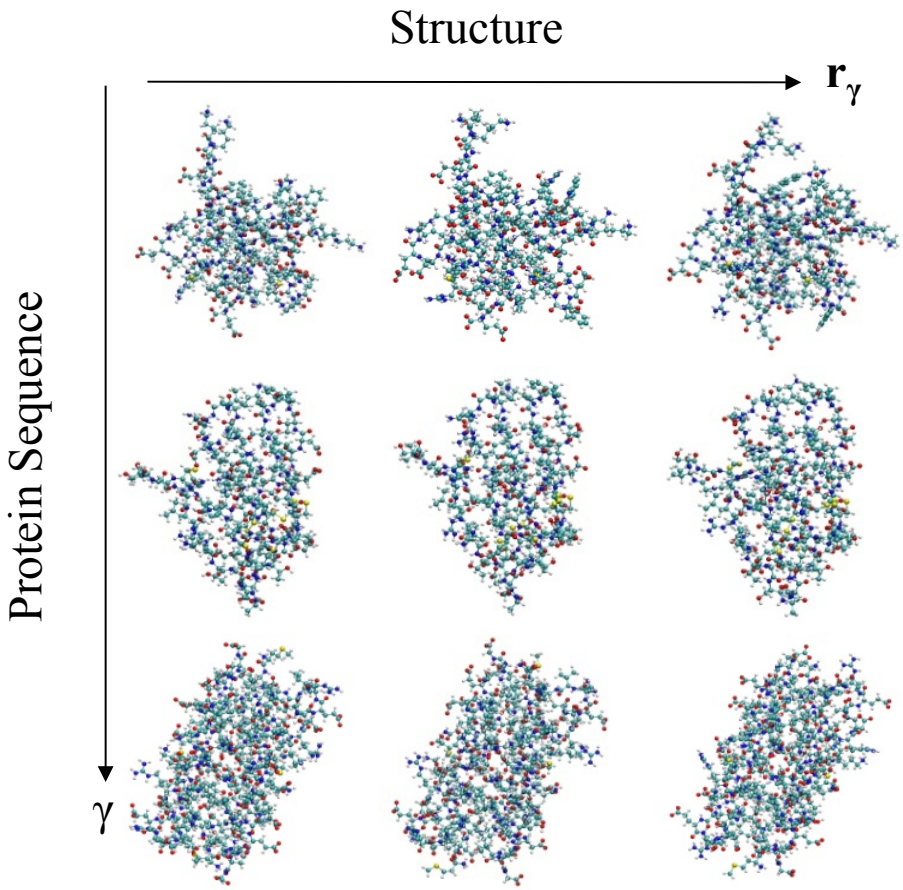
Henry and Camille Dreyfus Foundation

Alfred P. Sloan Research Foundation

Penn State Institute for Cyberscience

PENNSTATE
1855

Motivating Questions



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

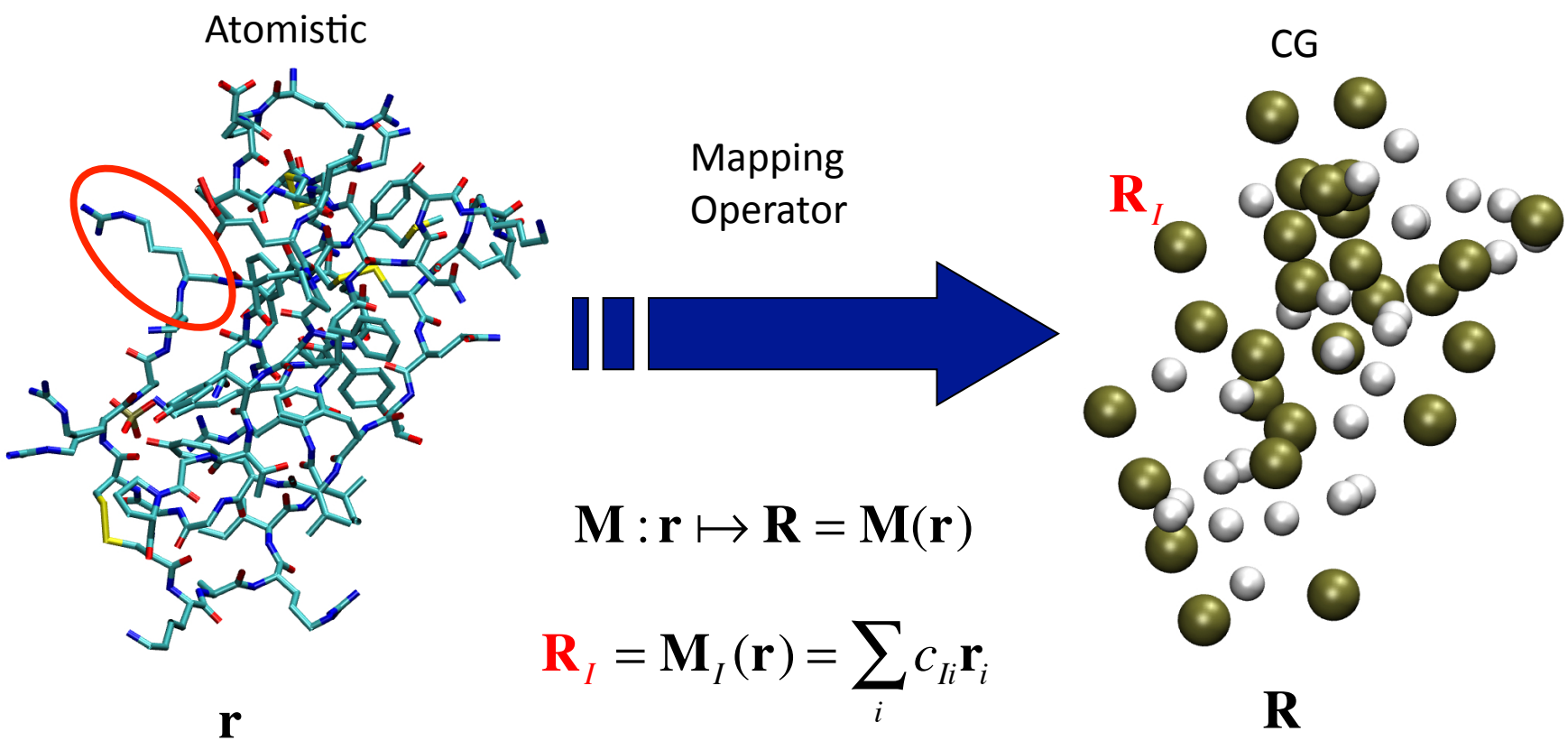
1. Given a collection of structures, what was the underlying potential?

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. New directions:
    1. Information theoretic formulation of force-matching
    2. Mean forces as a unifying framework for understanding structure-potential relations
4. Outstanding challenges

Coarse-grained (CG) Mapping



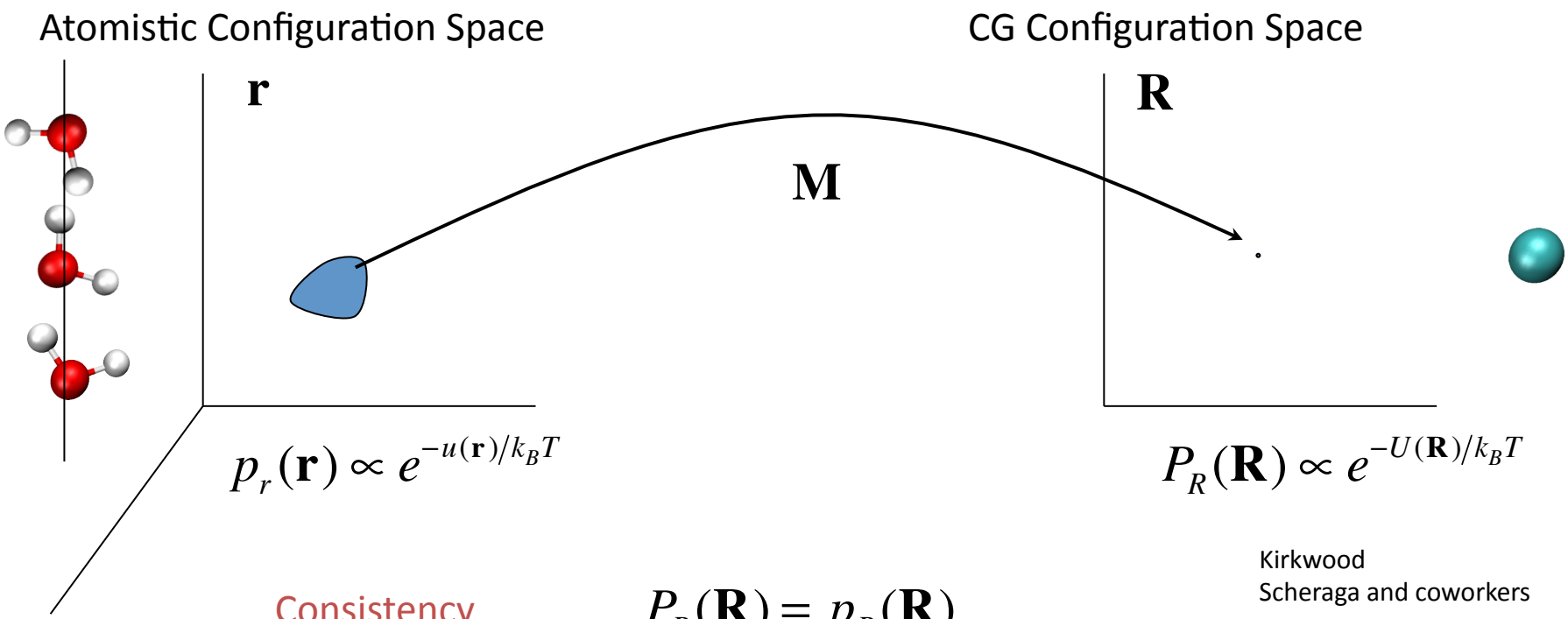
$$\mathbf{M} : \mathbf{r} \mapsto \mathbf{R} = \mathbf{M}(\mathbf{r})$$

$$\mathbf{R}_I = \mathbf{M}_I(\mathbf{r}) = \sum_i c_{Ii}\mathbf{r}_i$$

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



Kirkwood
Scheraga and coworkers

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

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

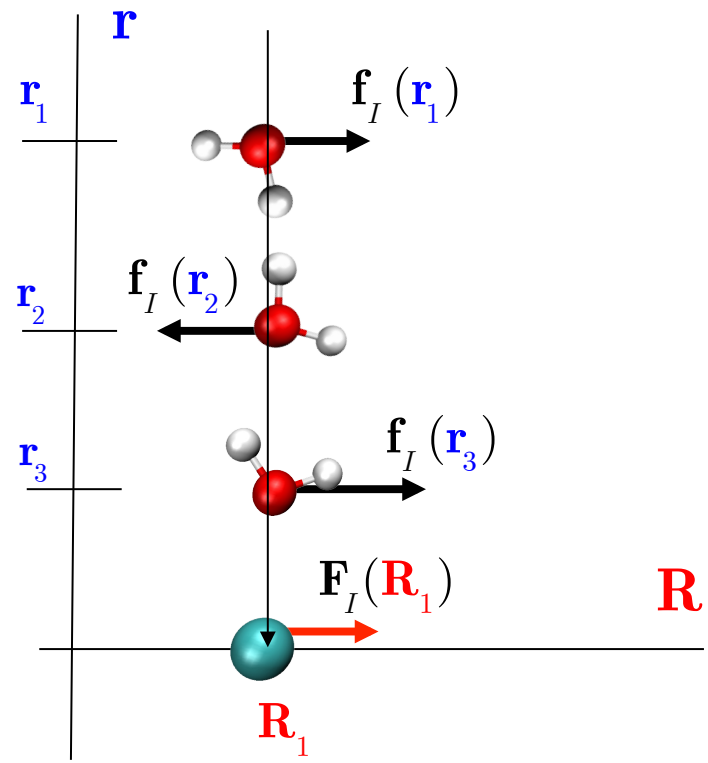
$$e^{-U(\mathbf{R})/k_B T} \propto \int d\mathbf{r}\; p_r(\mathbf{r})\, \delta\big(\mathbf{M}(\mathbf{r}) - \mathbf{R}\big)$$

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

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

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



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

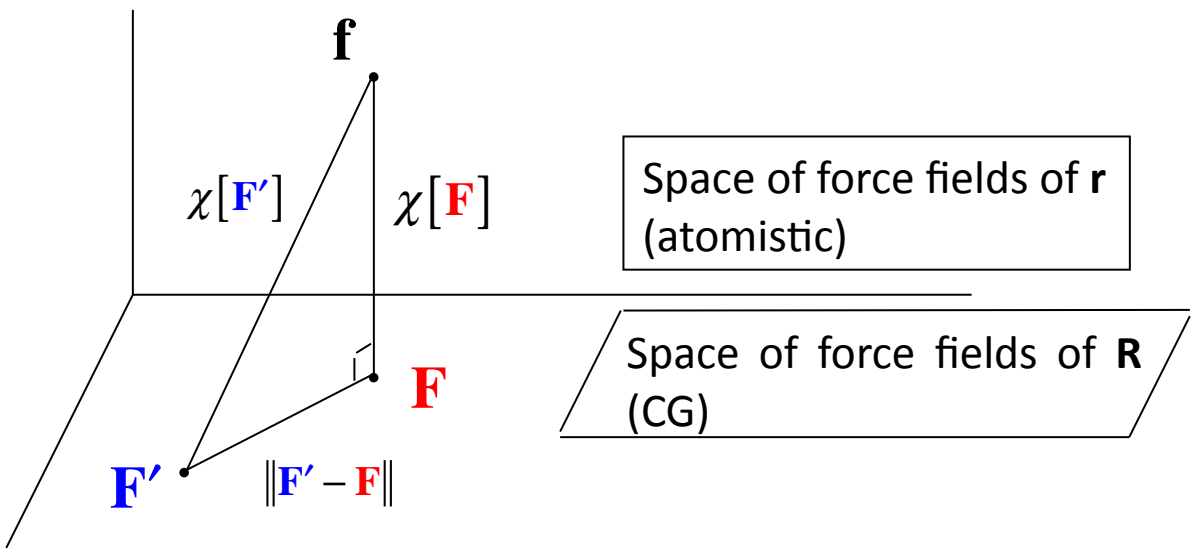
Atomistic FF:

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

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

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



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

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

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.

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

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

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

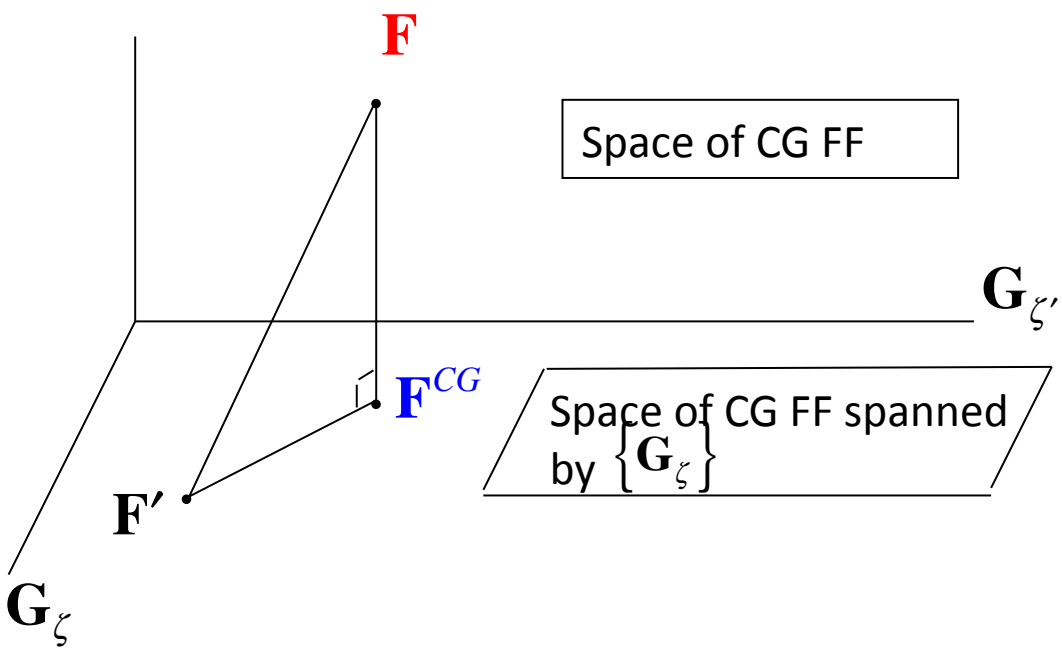
Noid, …, Chu, …, Andersen, Voth
J Chem Phys (2008)

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

Linear Least Squares Problem

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

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



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

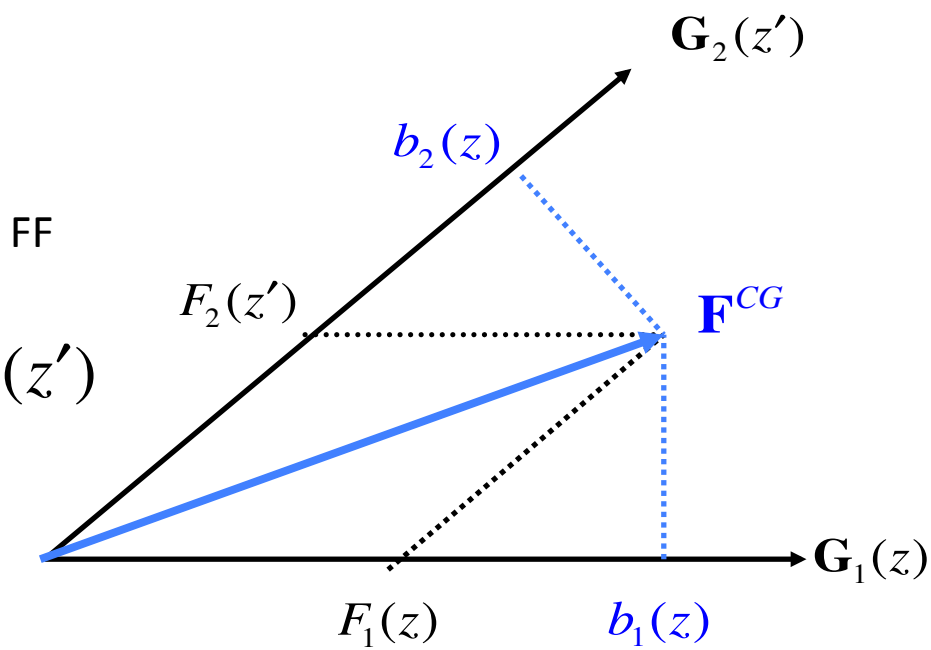
$$\boxed{\mathbf{F}^{CG} = \sum_{\zeta} \int dz\, F_{\zeta}(z)\mathbf{G}_{\zeta}(z)}$$

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

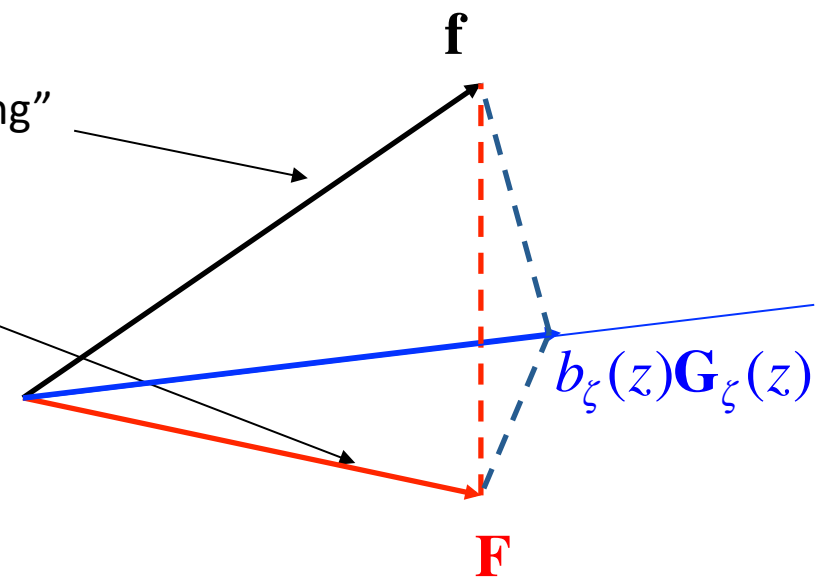
Integral Eq $$b_\zeta(z) = \mathbf{G}_\zeta(z)\cdot\mathbf{F} = \mathbf{G}_\zeta(z)\cdot\mathbf{F}^{CG} = \sum_{\zeta'}\int dz' G_{\zeta\zeta'}(z,z')F_{\zeta'}(z')$$

$$b_\zeta(z) = \mathbf{G}_\zeta(z)\cdot\mathbf{f}$$ MS-CG "Force-Matching"

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

$$= \mathbf{G}_\zeta(z)\cdot\nabla\left(-k_B T \ln p_R(\mathbf{R})\right)$$

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

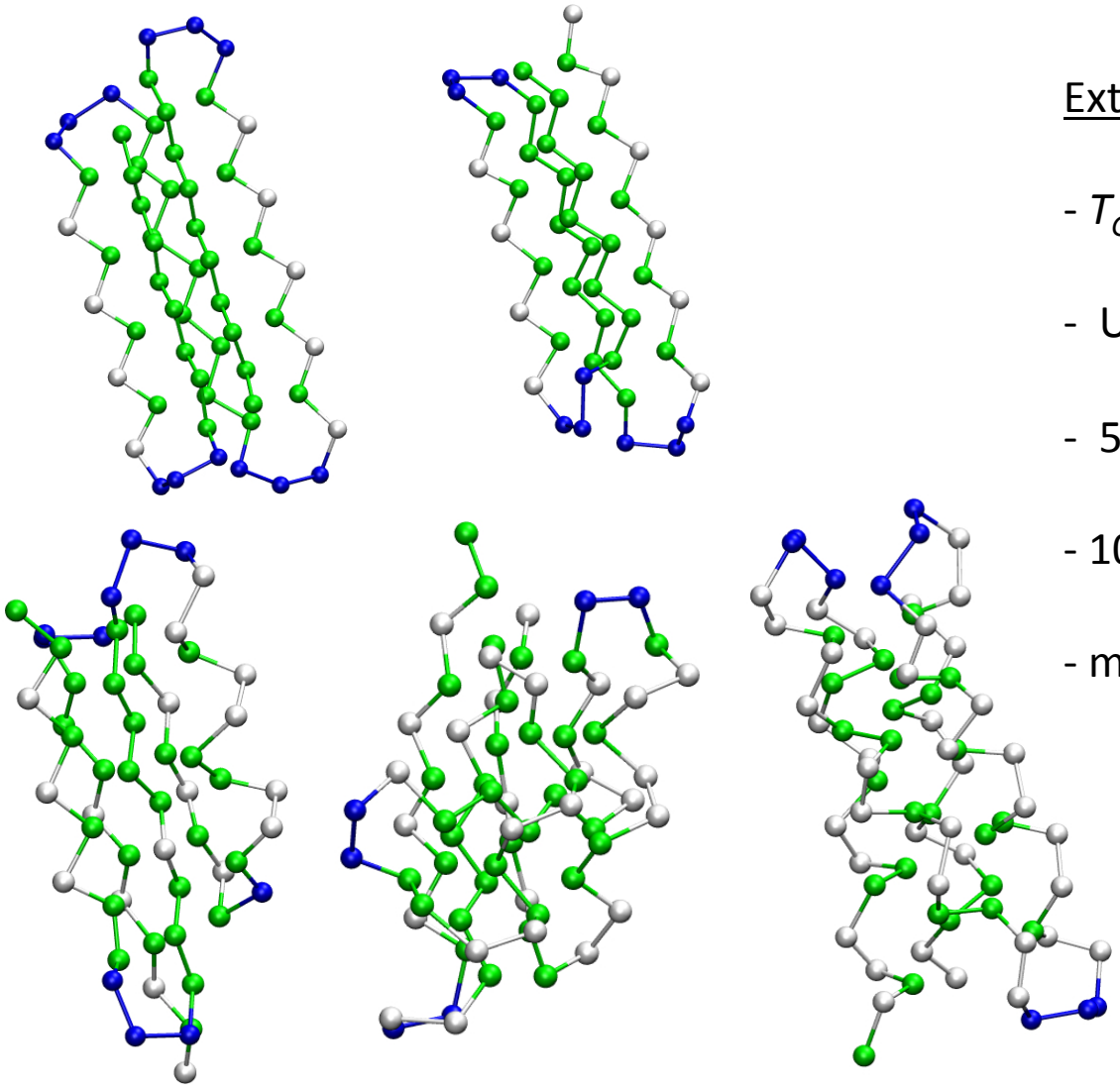


$$\boxed{k_B T\, d\bar{g}_\zeta(z)/dz = \sum_{\zeta'}\int dz' G_{\zeta\zeta'}(z,z')F_{\zeta'}(z')}$$

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!

Model Protein Databank



<u>Extended Ensemble</u>

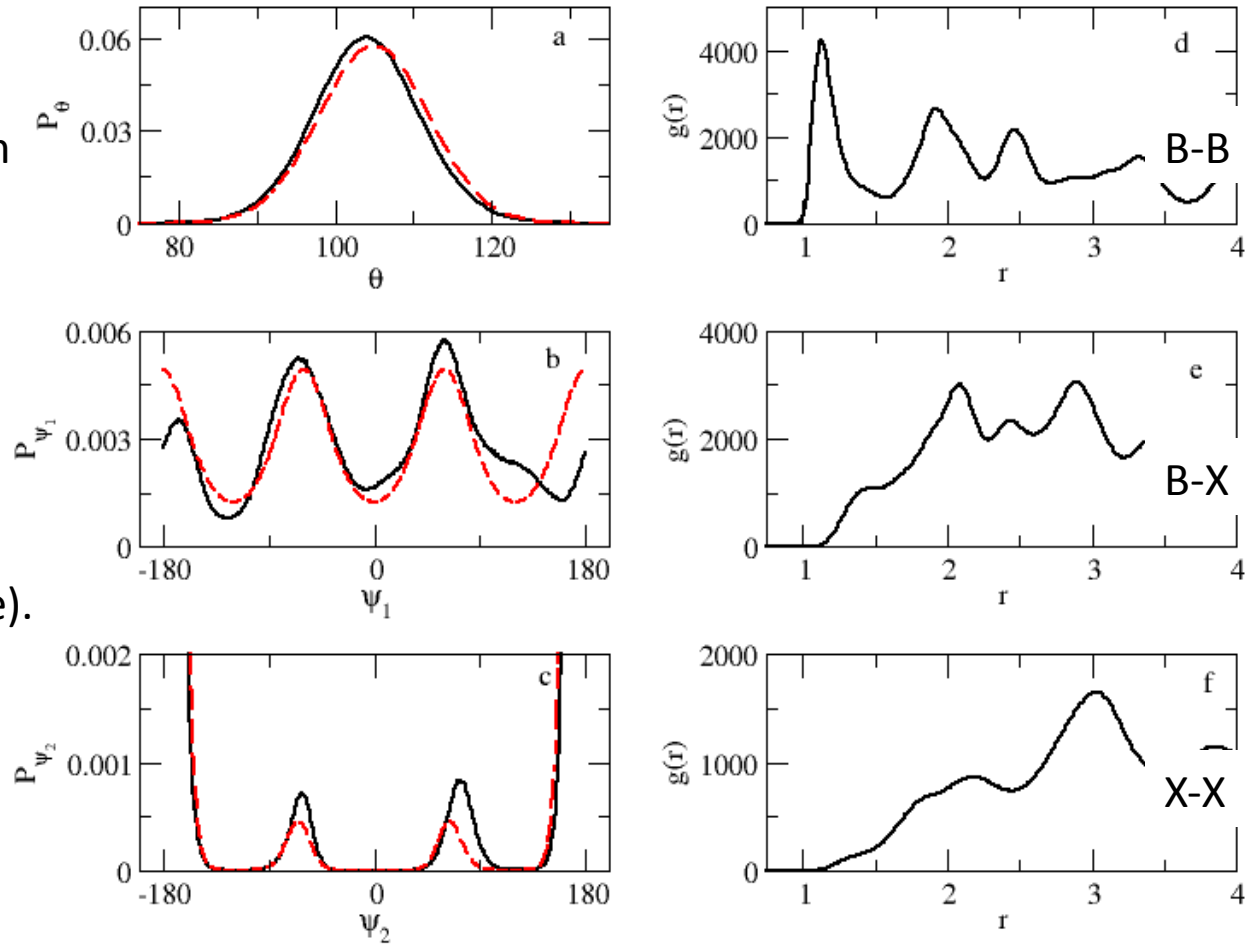
- $T_G < T < T_F$
- Uniform topology distribution
- 5 sequences
- 10^5 structures / sequence
- modified HT potential

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

Distributions from Model PDB

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

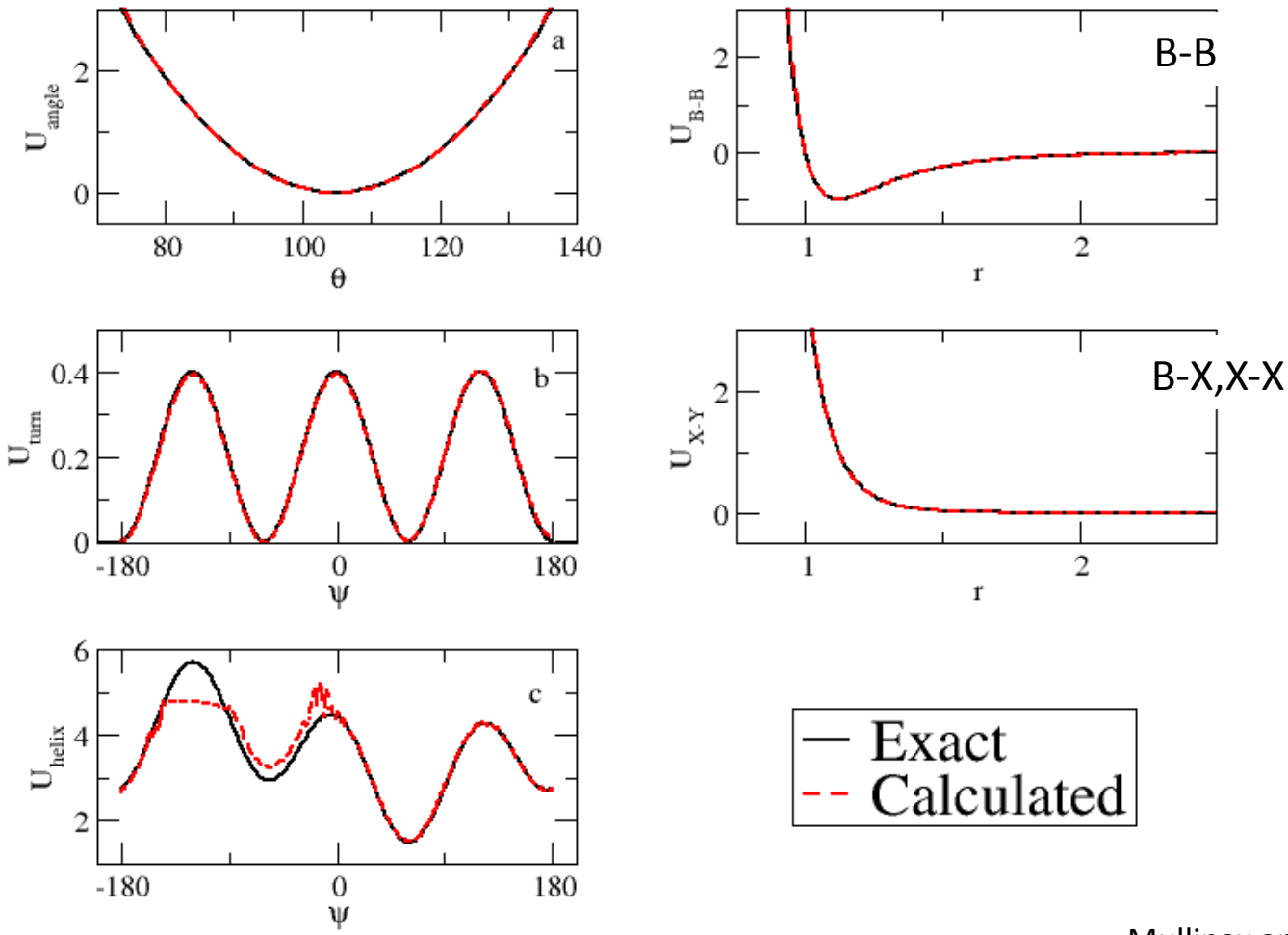
2. Chain connectivity generates long-ranged effective interactions between B-X and X-X pairs (which are purely repulsive).



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)

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

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

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**)

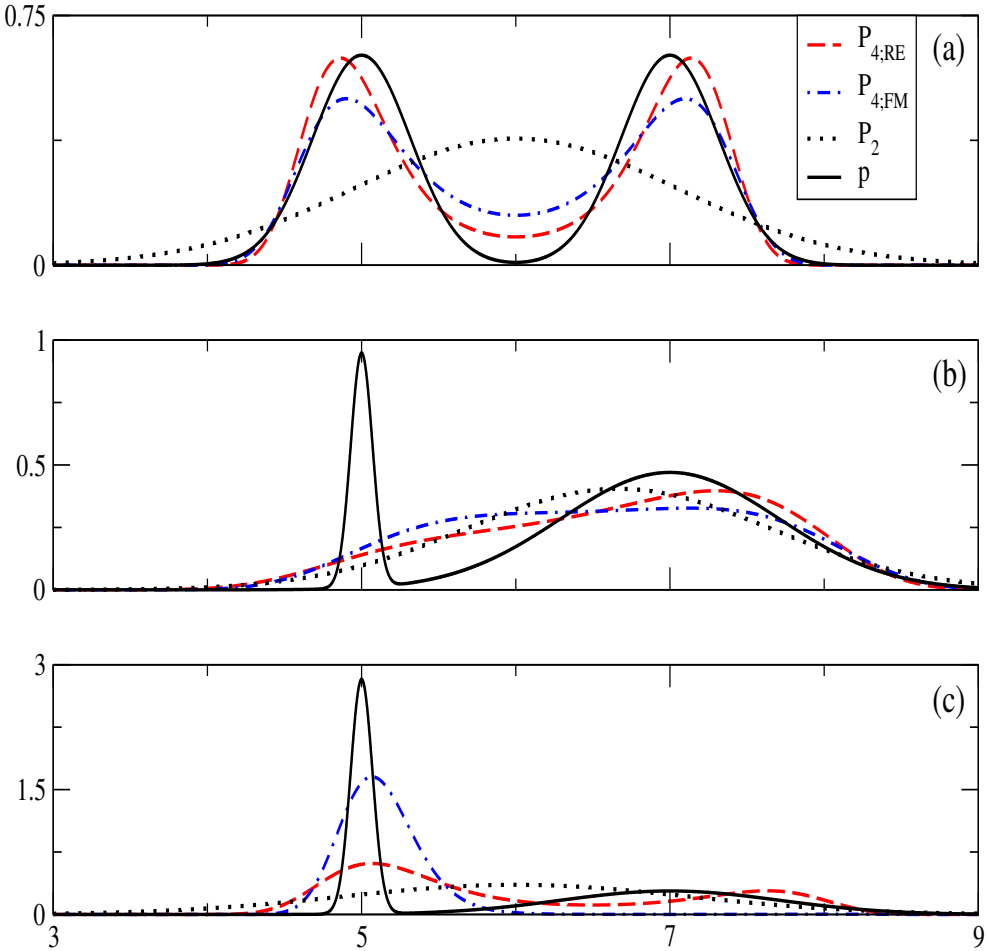
Relation to the Relative Entropy

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**

Incidental 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:

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

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

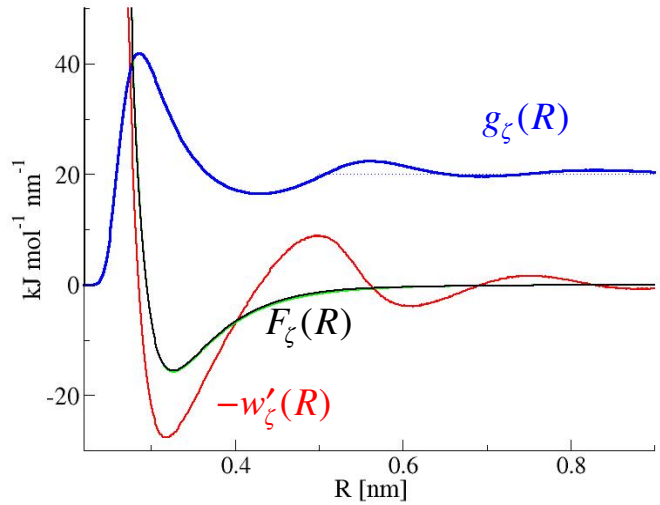
$$s_{\mathbf{r}} - s_{\mathbf{R}} \leq k_B \left\langle \ln\left[\frac{V^N}{V^n}\Omega_1\left(\mathbf{M}(\mathbf{r})\right)\right]\right\rangle$$

Rudzinski and Noid *JCP* **2011**

Mean forces

Generalized YBG theory:

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



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

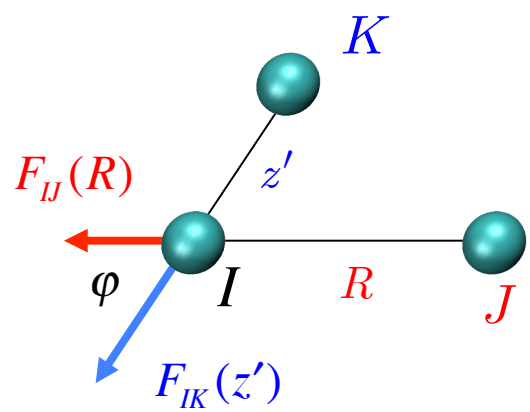
pair MF

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:

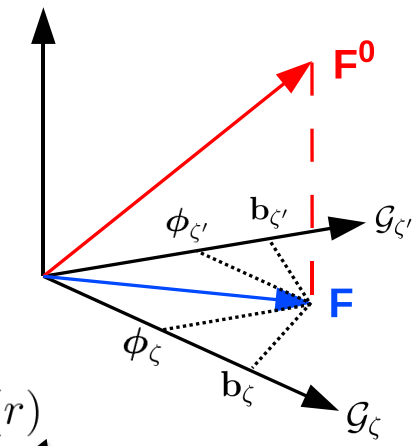
$$i = 0, \ldots \quad U_\zeta^{i+1}(z) = U_\zeta^i(z) - k_B T \ln\left[p_\zeta(z) / P_\zeta(z \mid U^i)\right]$$

Iterate to convergence !

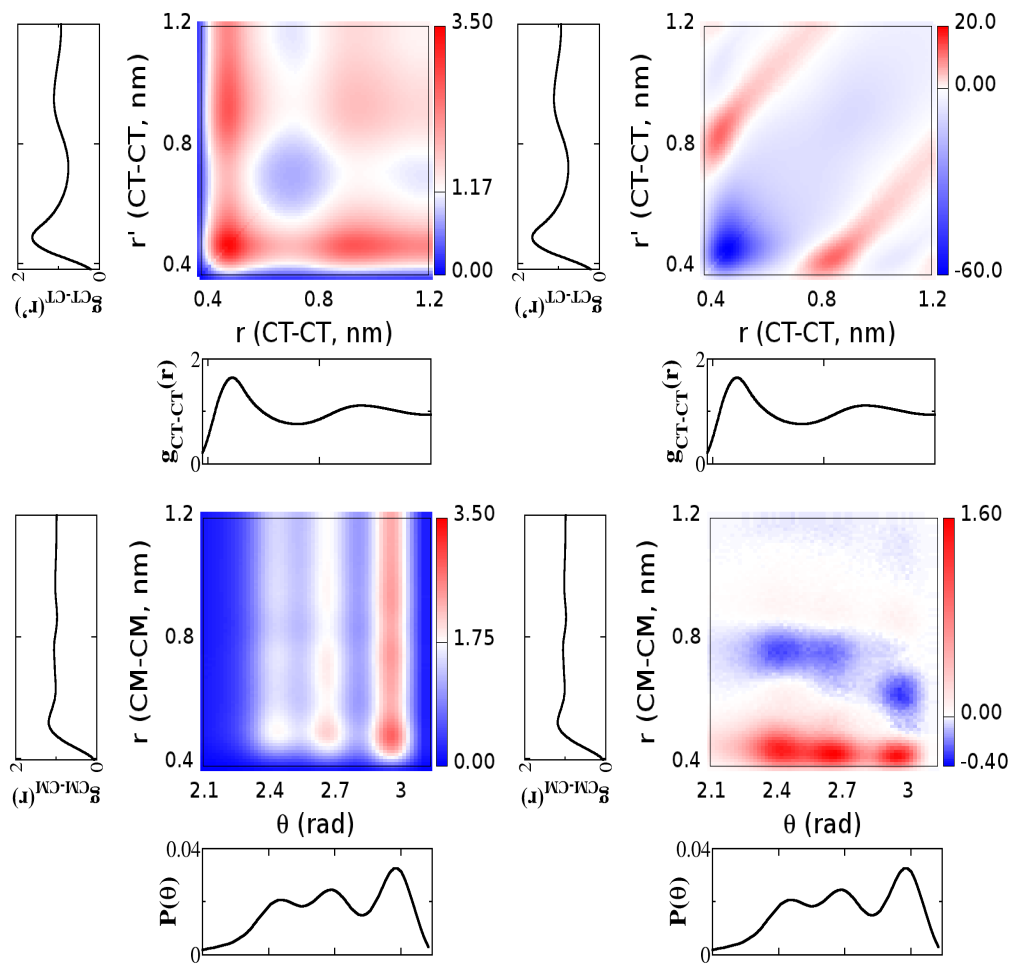
References:

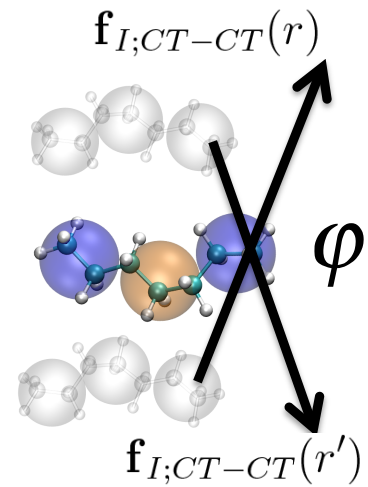
- Schommers *Phys Rev A* (1983) **28** 3599
- Soper *Chem Phys* (1996) **202** 295
- Muller-Plathe *ChemPhysChem* (2002) **9** 754
- Faller, and others
- Majek and Elber *Proteins* (2009) **76** 930

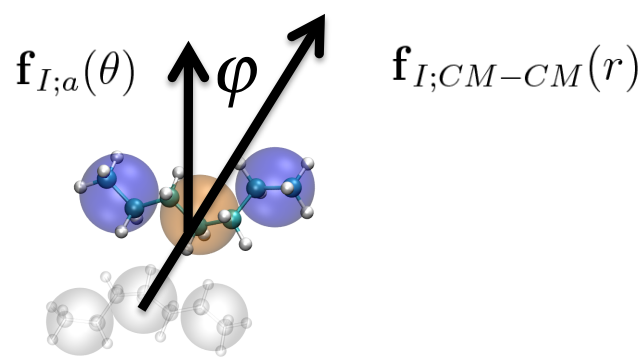
Understanding the Metric Tensor



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





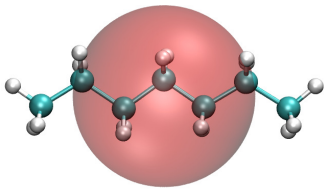


Rudzinski and Noid *JPCB* **2012**

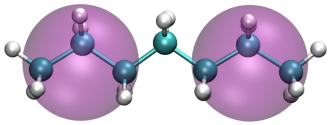
Robust features

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

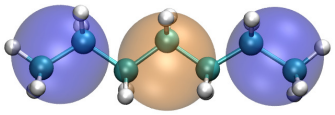
(a) CC

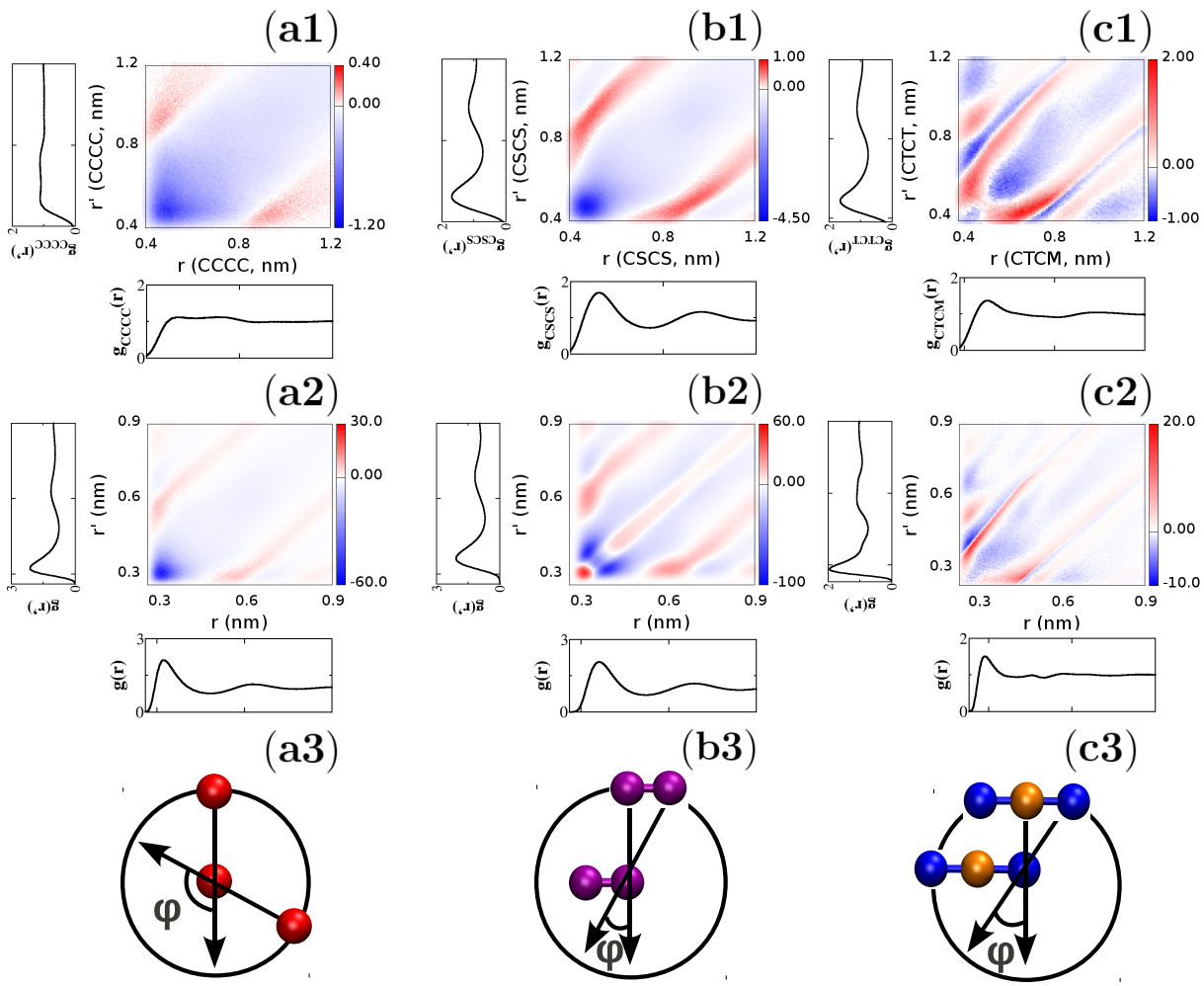


(b) CS-CS



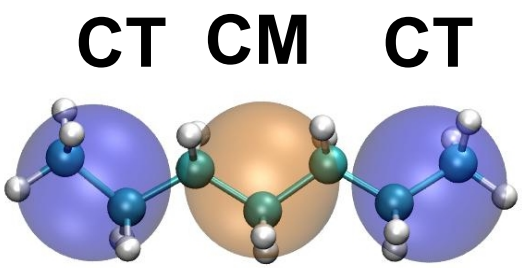
(c) CT-CM-CT

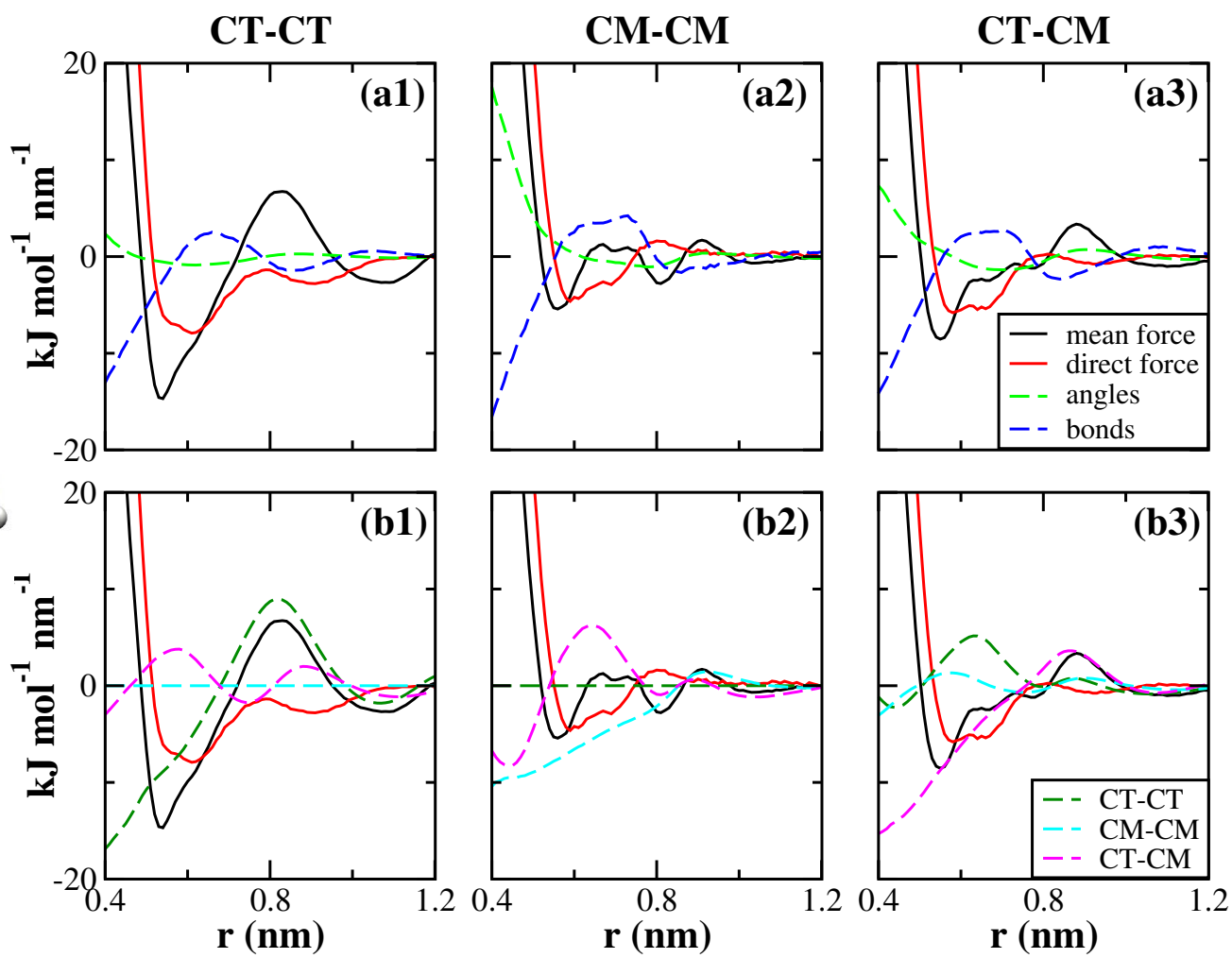




Rudzinski and Noid *JPCB* **2012**

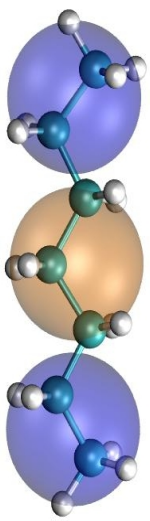
Decomposition of mean forces





Rudzinski and Noid *JPCB* **2012**

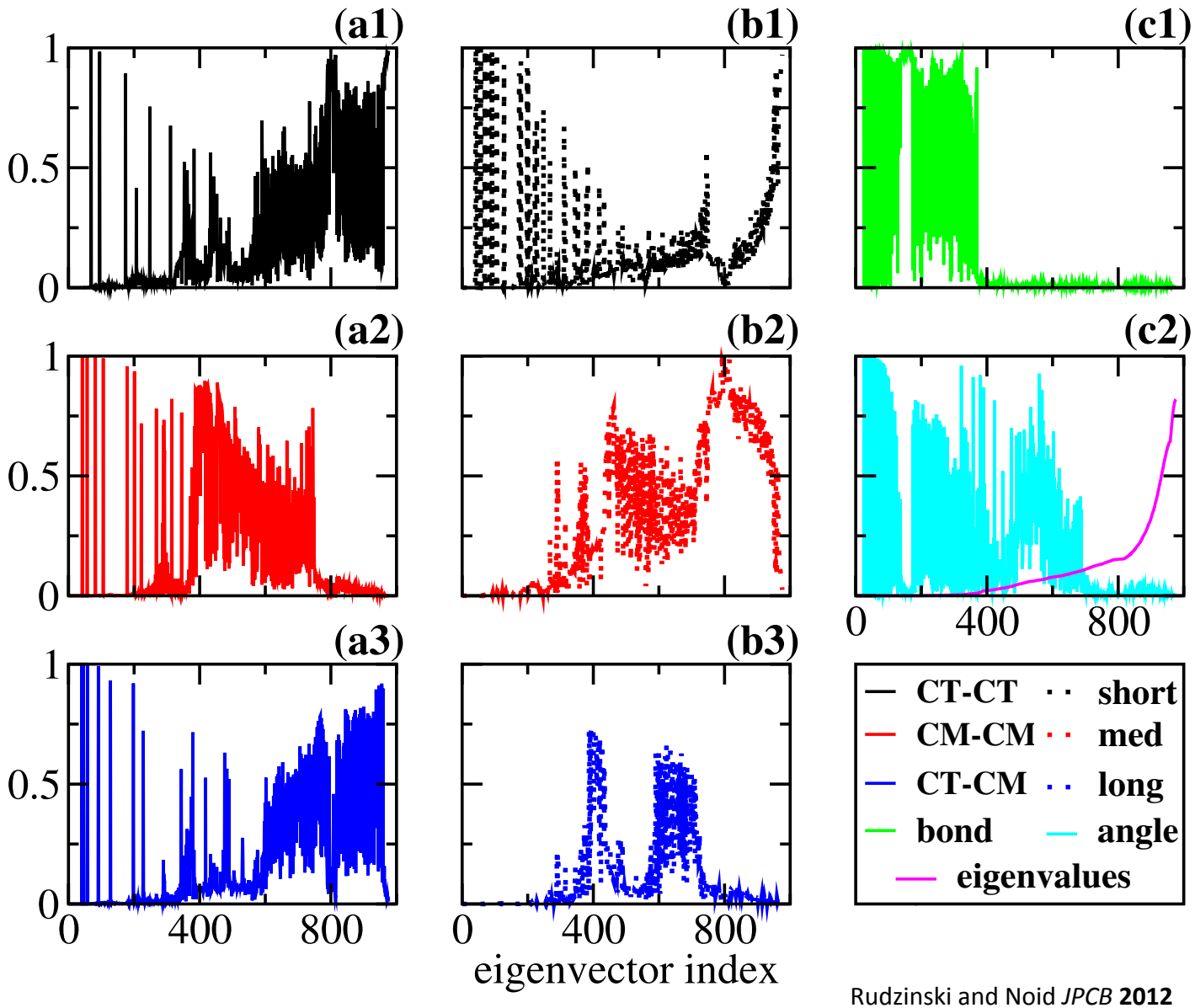
Eigenspectrum of metric tensor



CT

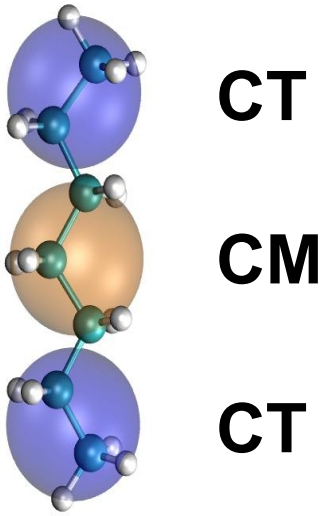
CM

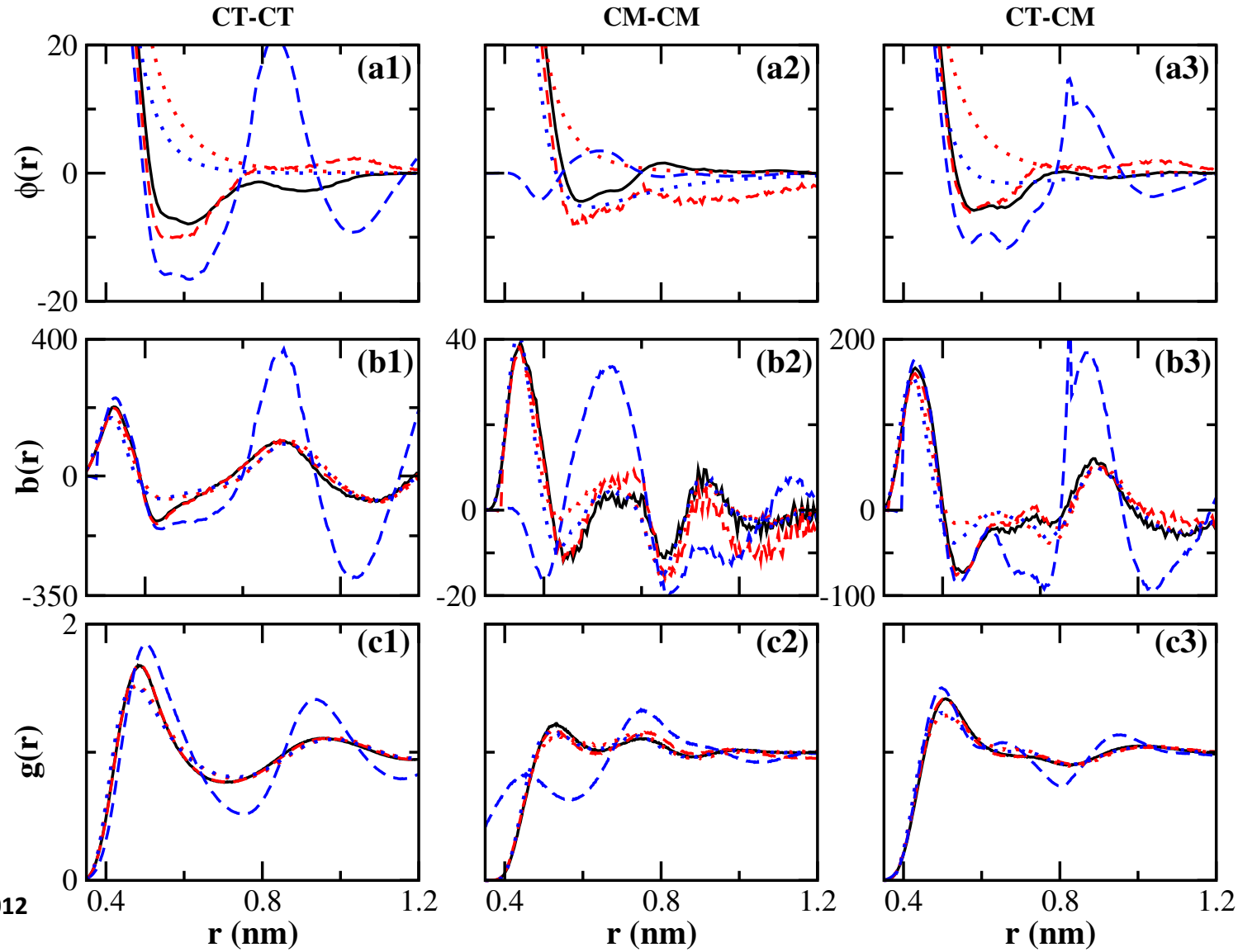
CT



Rudzinski and Noid *JPCB* **2012**

Impact upon CG structure





Rudzinski and Noid *JPCB* **2012**

Outstanding Challenges for Structure-Motivated Coarse-graining

1. Better metrics? – Are rdfs good enough?
2. Expanded basis sets?
3. Optimized mappings?
4. Transferability? – State Point, Systems,
5. Thermodynamics? – Pressure, Phase transitions?
6. Explicit solvent?
7. Explicit simulations? – Experimental data or theory?
8. Predictiveness and error bounds? – CG Amber?
9. Lunch theorems? - What do we want and is it possible?
    1. Does lunch exist?
    2. How much does it cost?

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

http://www.greendiary.com/entry/futuristic-trains-change-face-public-transportation/