Systematic multiscale models and physics using the relative entropy

August 30, 2018 • Entropy, Information, and Order in Soft Matter


Avi Chaimovich


Scott Carmichael


Tanmoy Sanyal

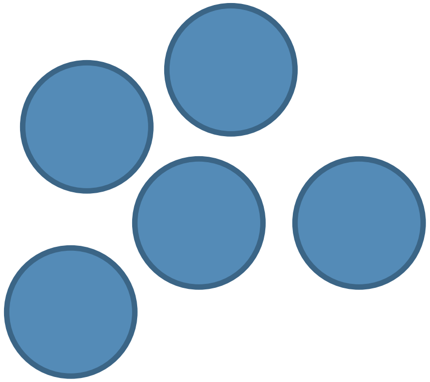

Will Noid (PSU)


Tommy Foley (PSU)

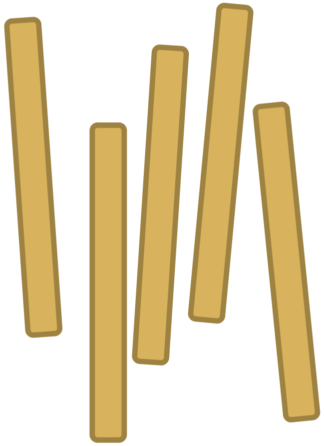
M. Scott Shell
Department of Chemical Engineering
University of California Santa Barbara

Support
Dreyfus Foundation
National Science Foundation

Model systems in soft matter physics

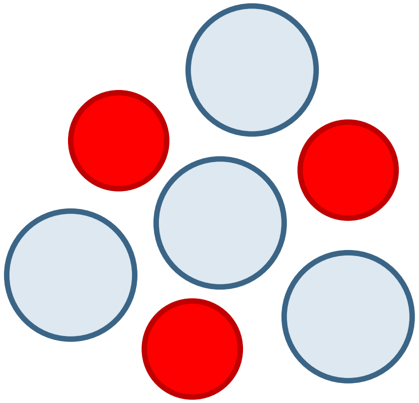


particles

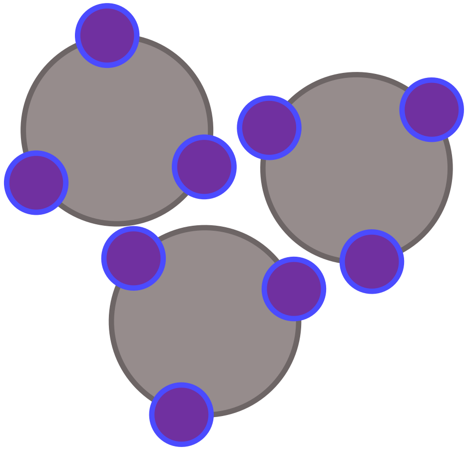


rods

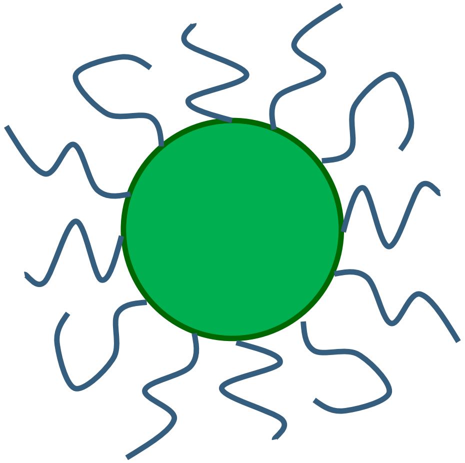
polymers



mixtures

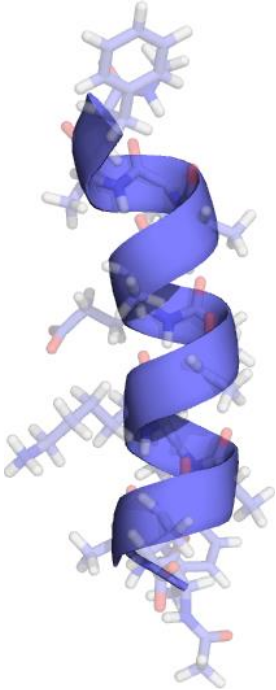


functionalized

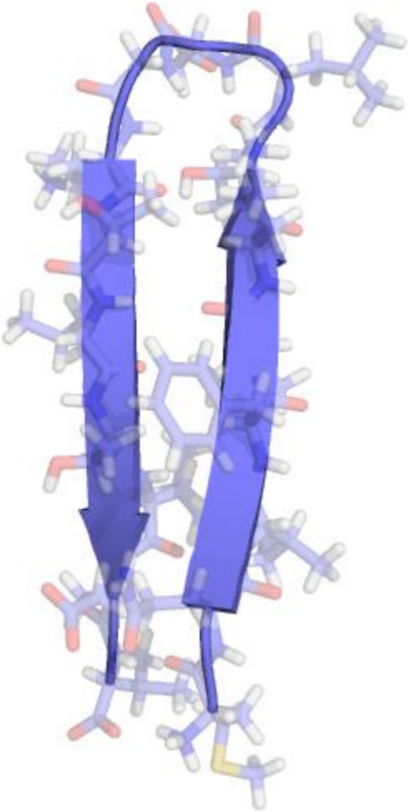


hybrids

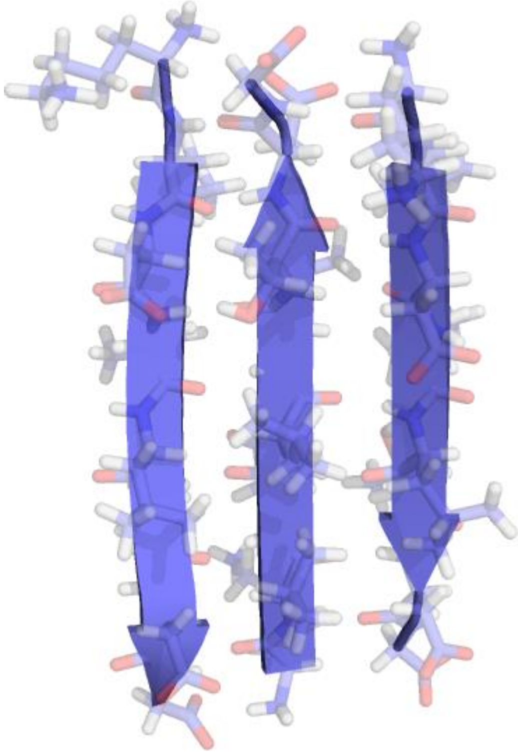
Peptides can form helices, hairpins, and beta sheets



helix

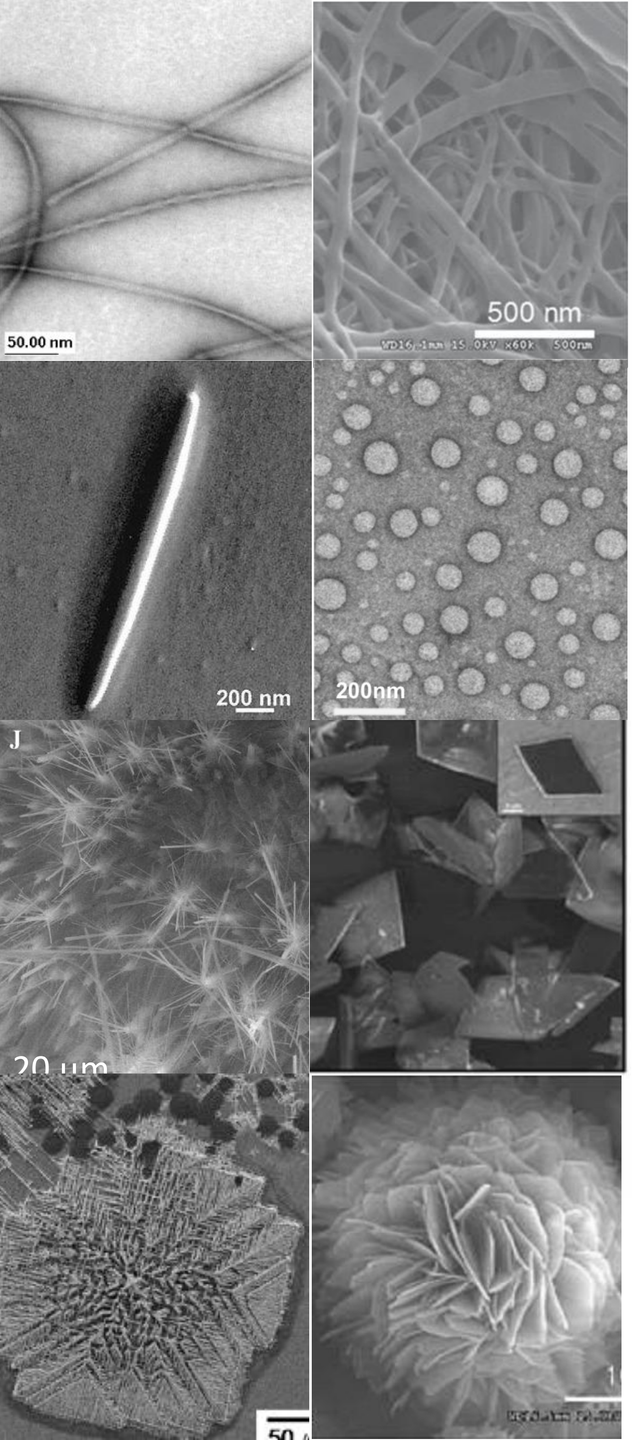


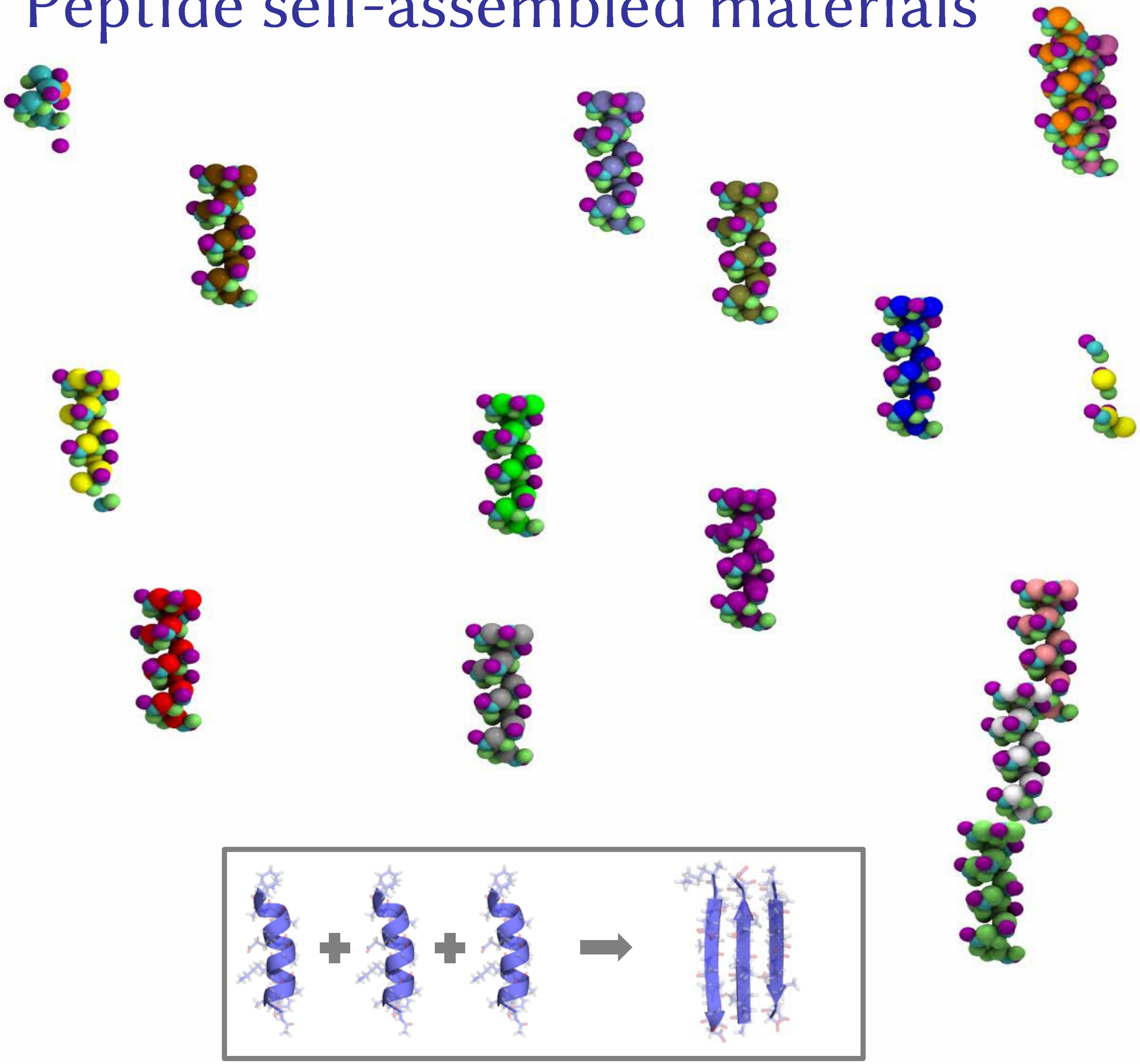
hairpin



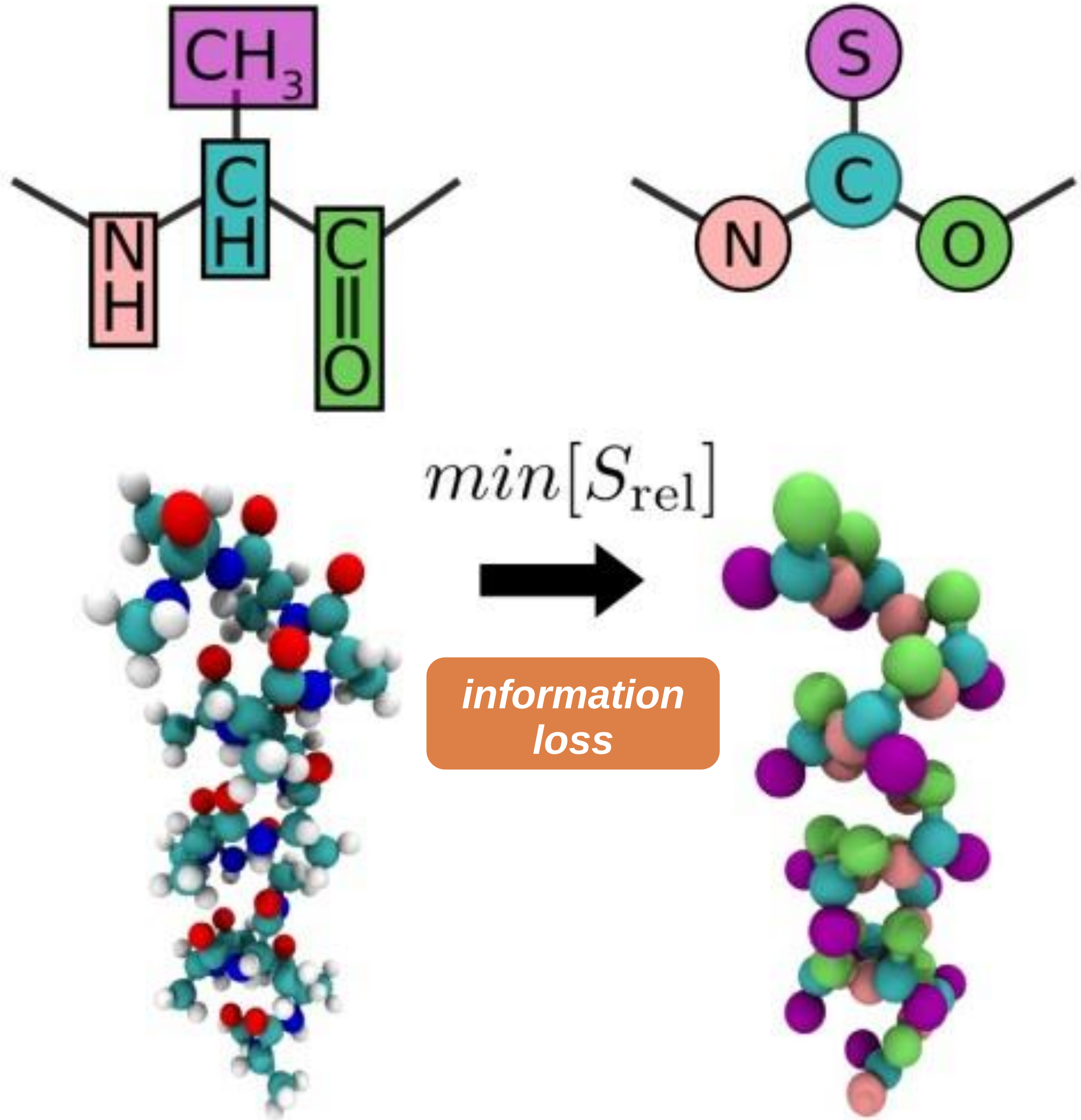
beta sheet aggregate

Peptide self-assembled materials





1. Tycko et al., Ann. Rev. of Phys. Chem. (2001)
2. Reches, et al. Science (2003)
3. Amdursky et al, Biomacromolecules (2011)
4. Han et al, Colloids and Biosurfaces B (2011)
5. Yan et al., Chem. Soc. Rev. (2010)
6. Yan et al., Angewandte Chem. Int. Ed. (2007)
7. Govindaraju et al, Supramolec. Chem. (2011)
8. Su et al, J. Mater. Chem. (2010)

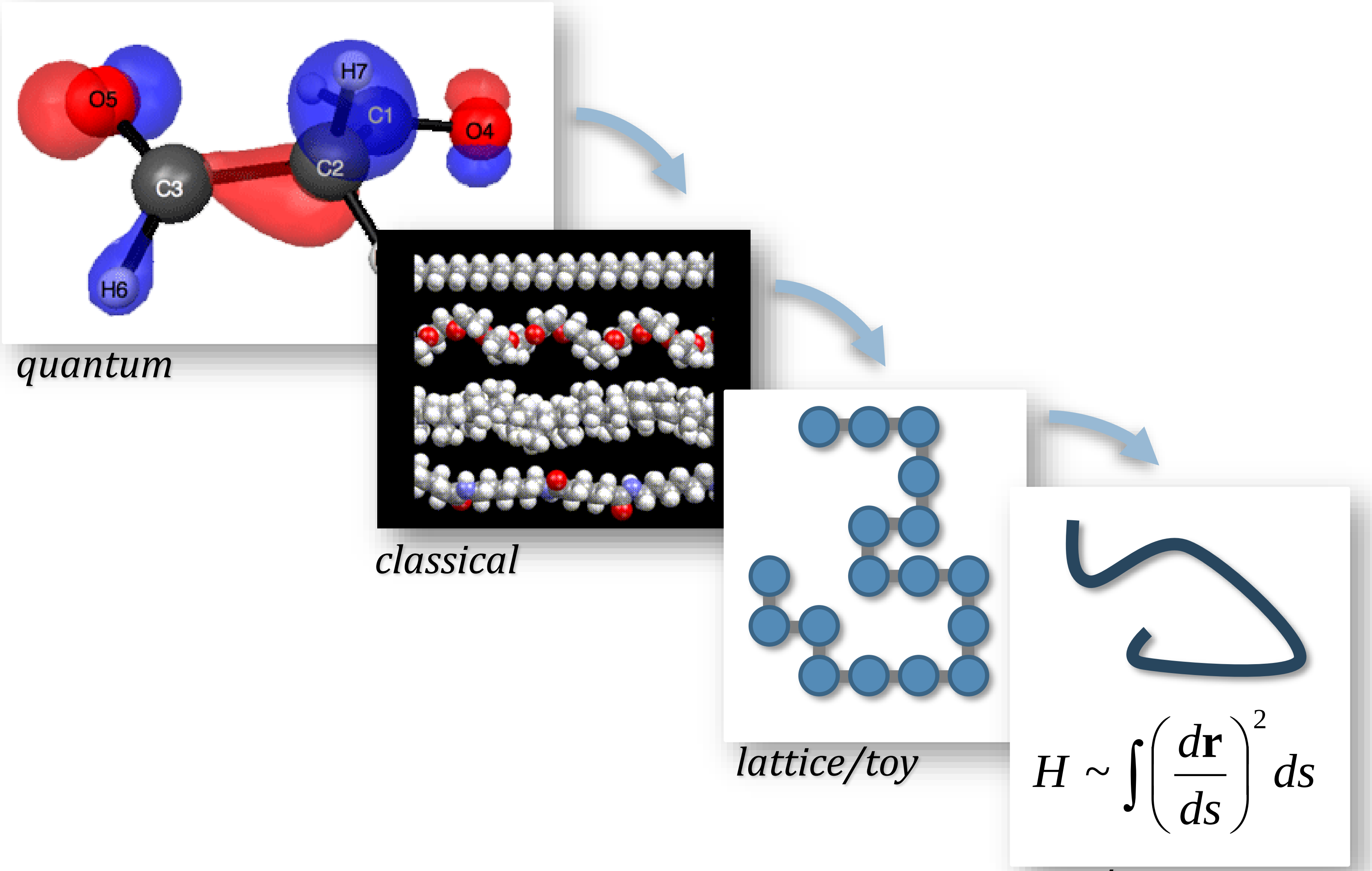


$min[S_{\mathrm{rel}}]$

information loss

Shell, JCP (2008); Carmichael and Shell, JPCB (2012); Carmichael and Shell, JCP (2015)

Many kinds of bottom-up coarse graining



quantum

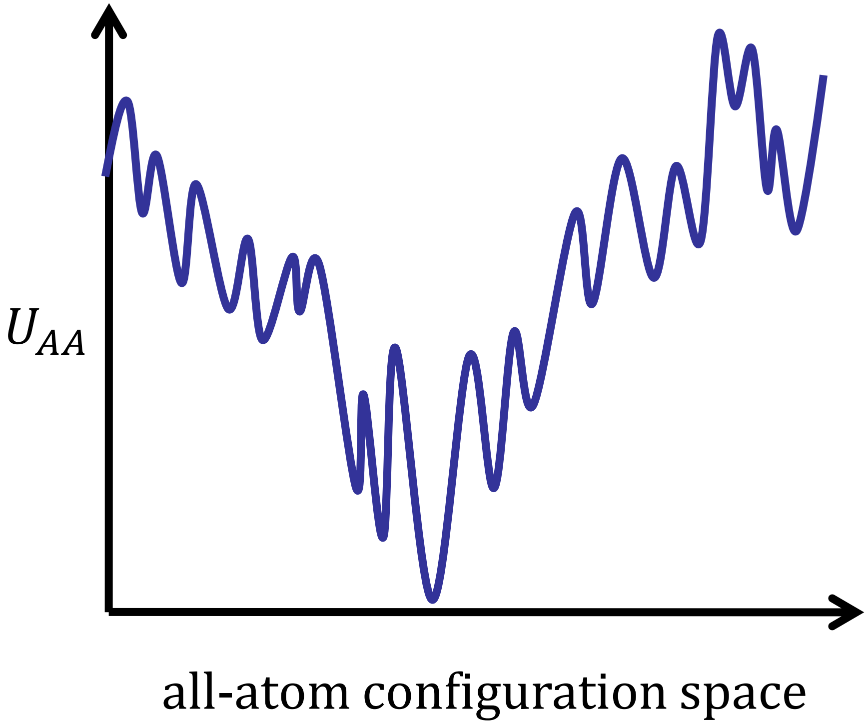
classical

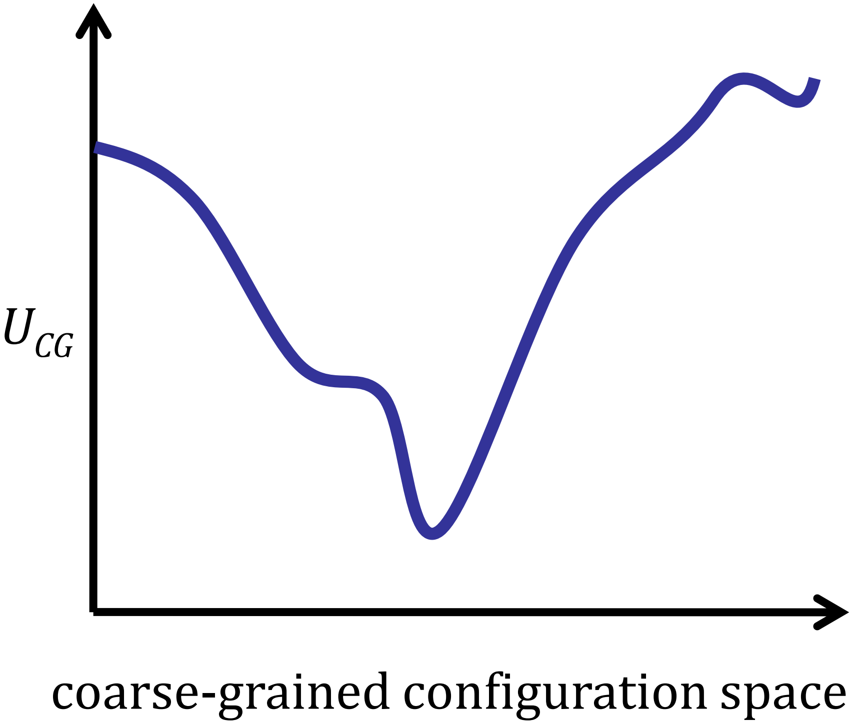
lattice/toy

$$H \sim \int \left(\frac{d\mathbf{r}}{ds} \right)^2 ds$$

continuum

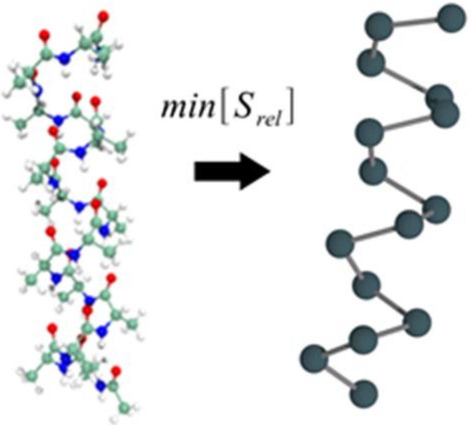
Why think about bottom-up coarse-graining?





fewer degrees of freedom,
simpler interaction potentials,
smoother energy landscapes,
emergent physics/models

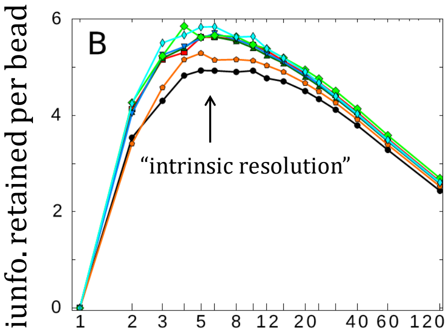
A relative entropy framework for multiscale modeling

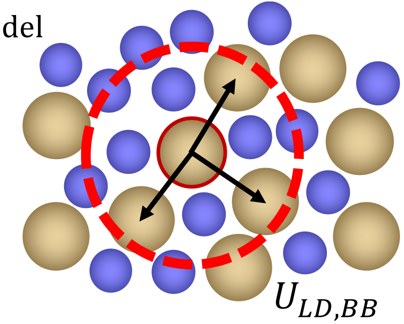


$$S_{\text{rel}} = \sum_i p_{AA}(i) \ln\left(\frac{p_{AA}(i)}{p_{CG}(i)}\right)$$

Designing coarse grained models with the relative entropy

Shell, Adv Chem Phys (2016)
review article



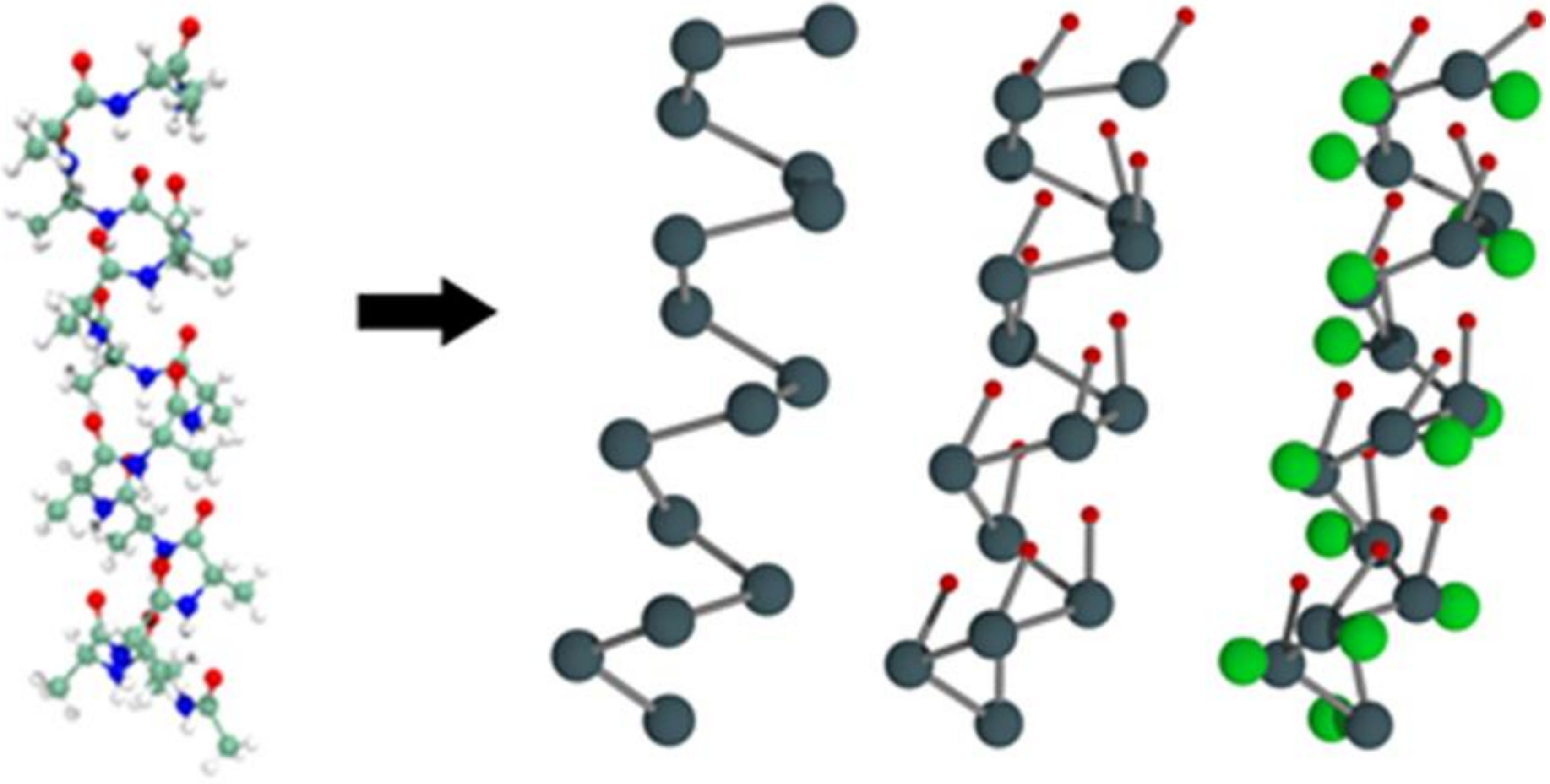


“Next-generation” coarse-grained models for simulation

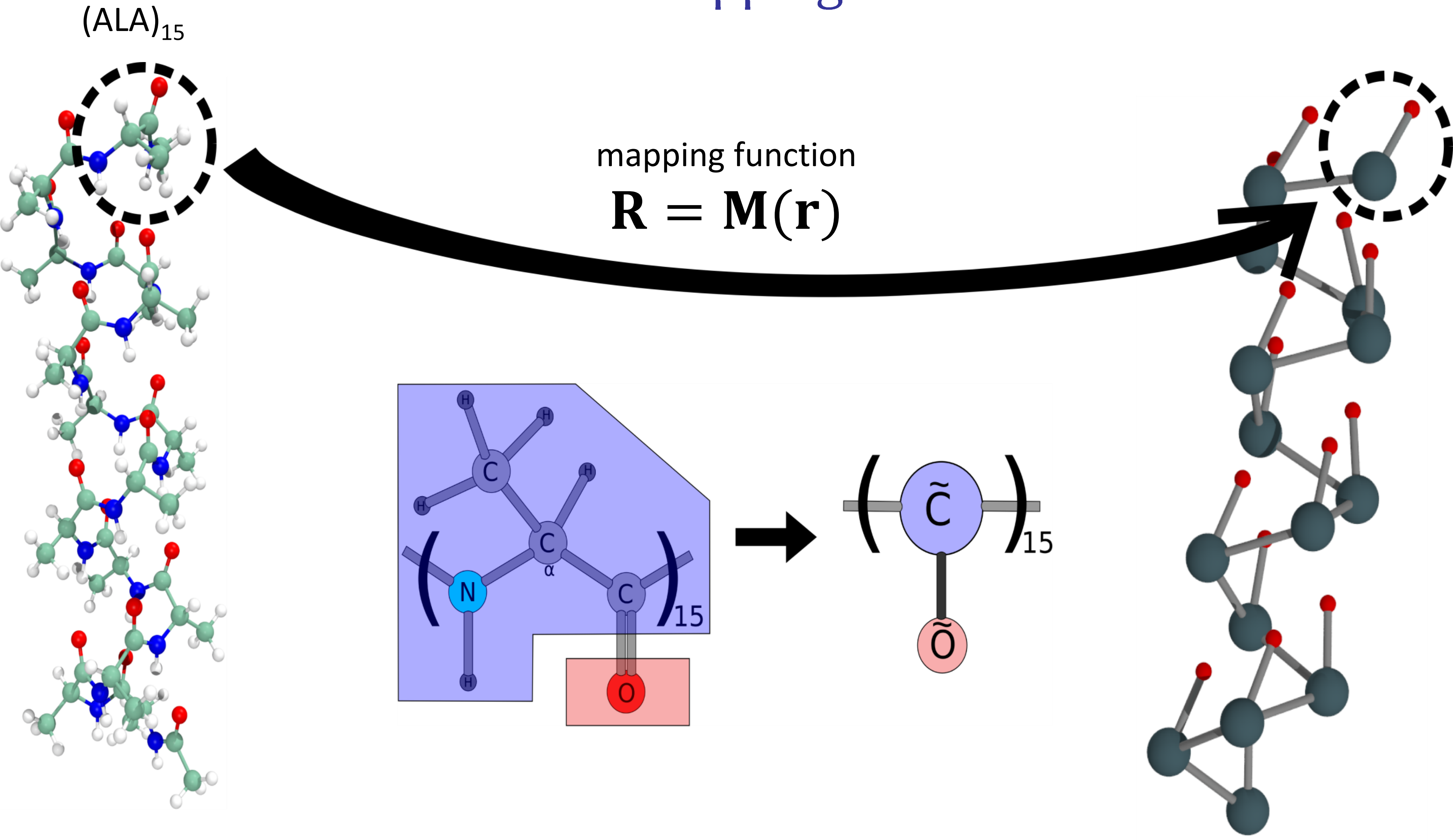
Foley, Shell, & Noid (2015)
Sanyal & Shell, JCP (2016)
Sanyal & Shell, JPCB (2018)

all-atom peptide model

coarse-grained models of varying detail



Mapping



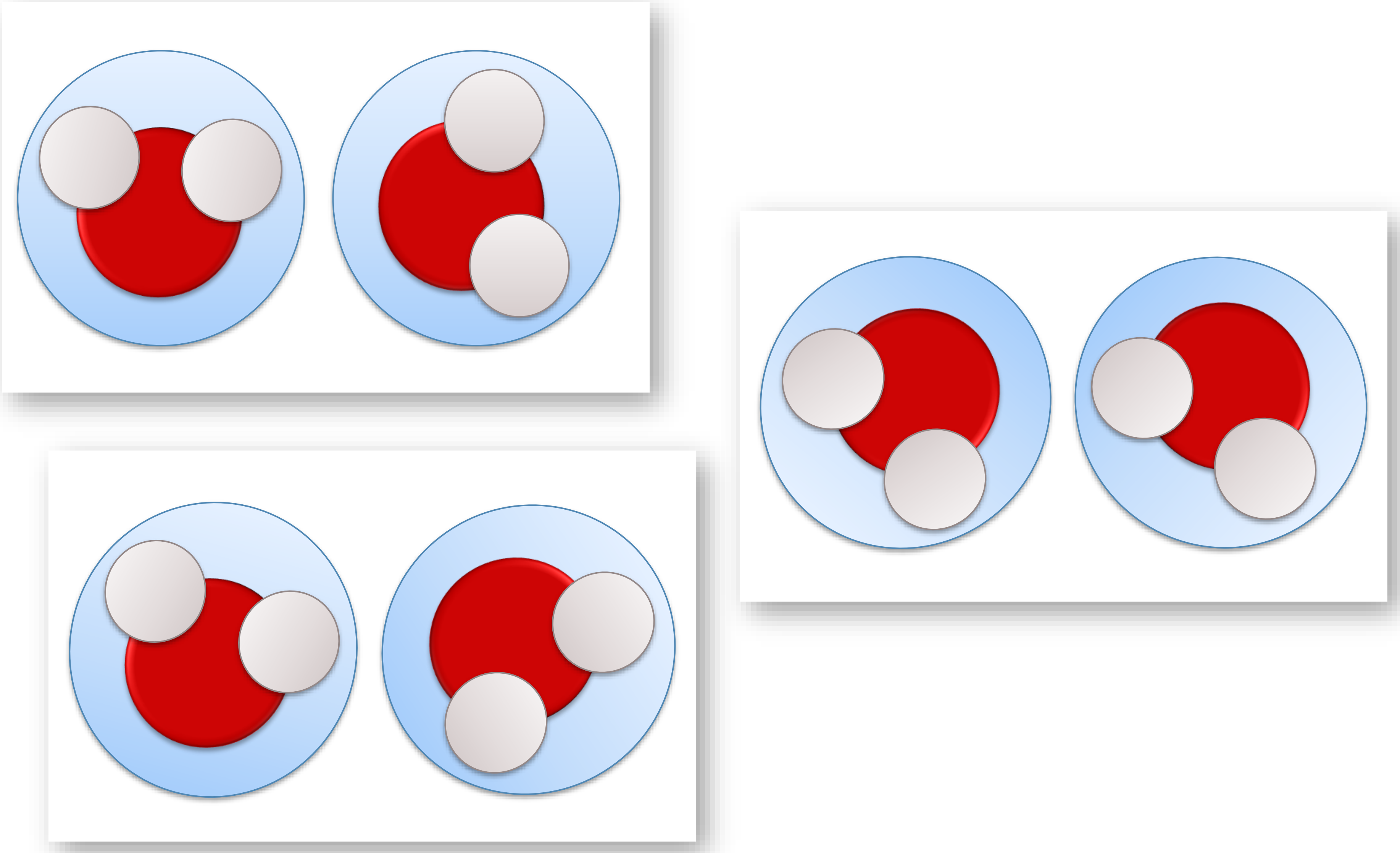
mapping function

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

atomic coordinates $\mathbf{r}$

CG coordinates $\mathbf{R}$

An important point:
multiple AA configurations map to the same CG one



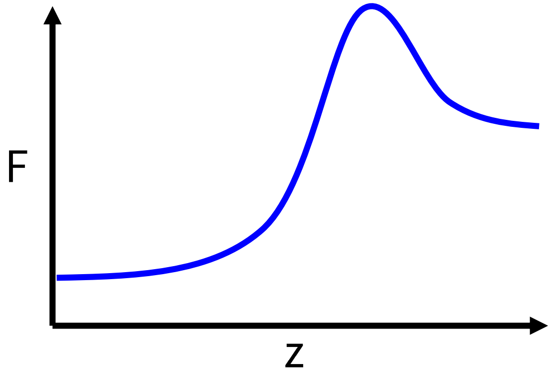
Interactions – thermodynamics to the rescue

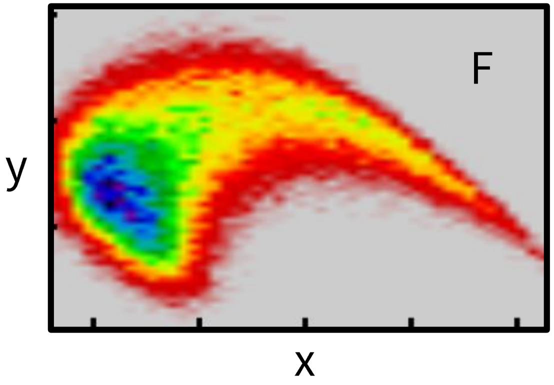
$$U_{CG}(\mathbf{R}) \quad \rightarrow \quad PMF(\mathbf{R}) = -k_B T \ln \int e^{-\beta U_{AA}(\mathbf{r})} \underbrace{\delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}]}_{\substack{\text{filters for all atomic configs. } \mathbf{r} \text{ that} \\ \text{map to the same CG config. } \mathbf{R}}} d\mathbf{r}$$

Interactions – thermodynamics to the rescue

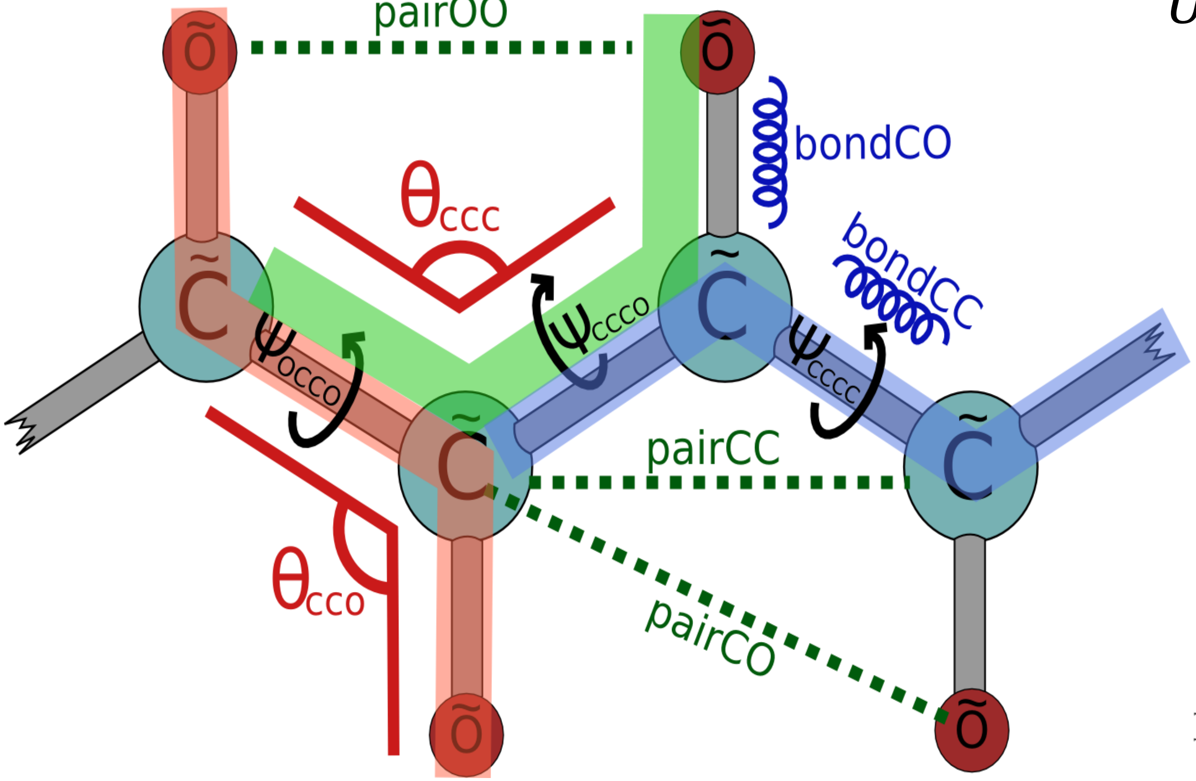
$$U_{CG}(\mathbf{R}) \;\rightarrow\; PMF(\mathbf{R}) = -k_B T \ln \int e^{-\beta U_{AA}(\mathbf{r})} \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}] d\mathbf{r}$$

Just like familiar free energy surfaces, only highly dimensional



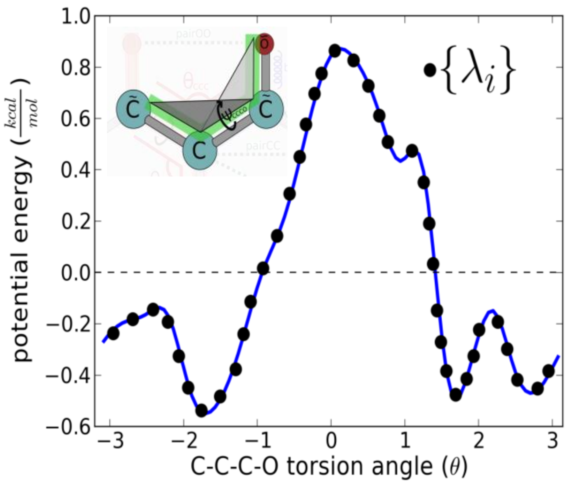


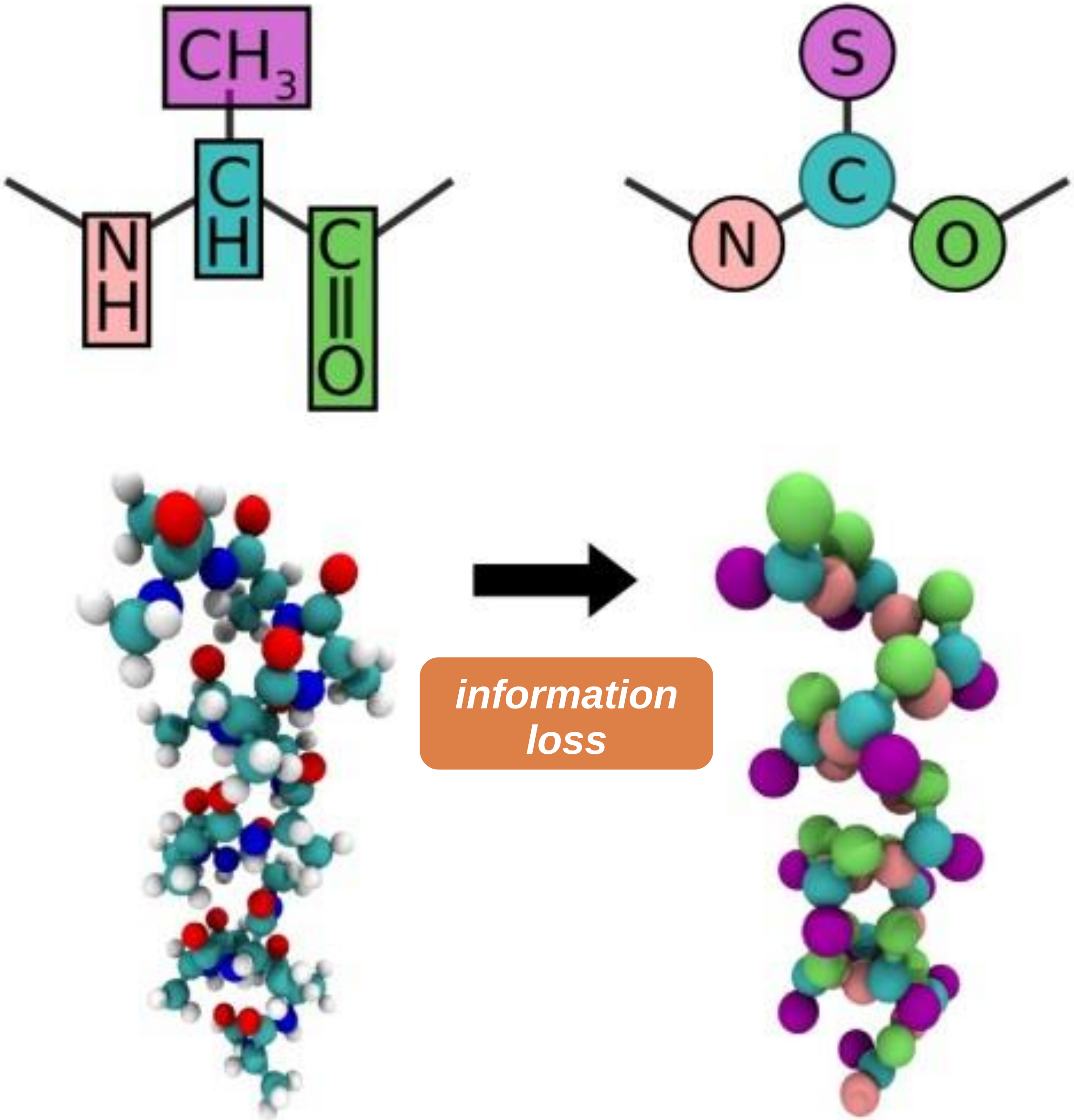
Interactions – approximating the PMF



$$U_{CG}(\mathbf{R};\underbrace{\lambda_1,\lambda_2,\ldots}_{\text{parameters}})$$

C-C-C-O dihedral potential





information loss

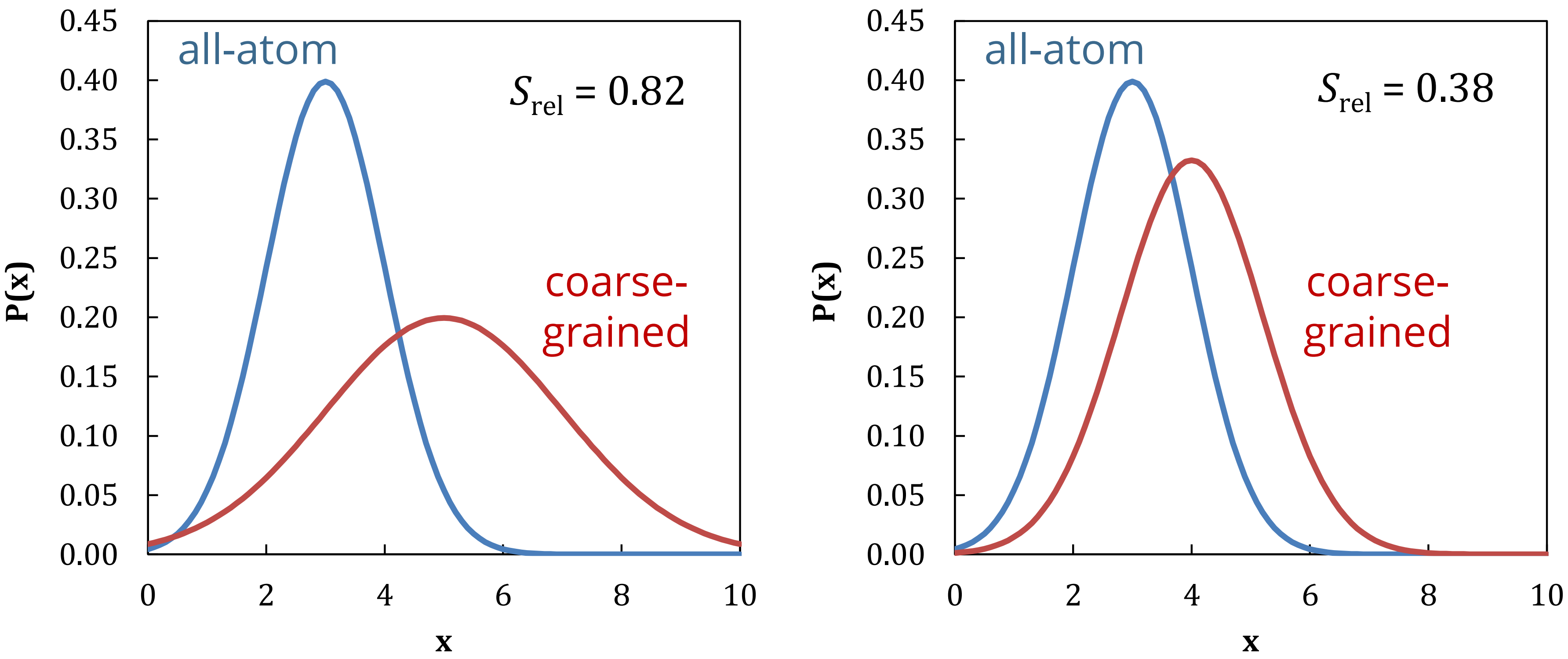
Shell, JCP (2008); Carmichael and Shell, JPCB (2012); Carmichael and Shell, JCP (2015)

Measuring information loss

$$S_{\mathrm{rel}} = \sum_{\text{configs } i} p_{AA}(i) \ln\left(\frac{p_{AA}(i)}{p_{CG}(i)}\right) \geq 0$$

Shell, JCP (2008); Chaimovich and Shell, PRE (2010); Chaimovich and Shell, JCP (2011)

Example in one dimension…



…but in reality there are 3N dimensions for $p(i)$

Measuring information loss for removing DOF

$$S_{\text{rel}} = \sum_{\text{CG configs } i} p'_{AA}(i) \ln\left(\frac{p'_{AA}(i)}{p_{CG}(i)}\right) + S_{\text{map}}$$

p'_{AA} projects all atom probabilities
onto the CG space

Shell, JCP (2008); Chaimovich and Shell, PRE (2010); Chaimovich and Shell, JCP (2011); Rudzinski and Noid, JCP (2011).

Measuring information loss for removing DOF

$$S_{\mathrm{rel}} = \underbrace{\sum_{\text{CG configs } i} p'_{AA}(i)\ln\left(\frac{p'_{AA}(i)}{p_{CG}(i)}\right)}_{\substack{\text{Zero for “ideal” interactions} \\ \text{(i.e., } U_{CG} = PMF\text{)}}} + \underbrace{S_{\mathrm{map}}}_{\substack{\text{effect of mapping,} \\ \text{independent of } U_{CG}}}$$

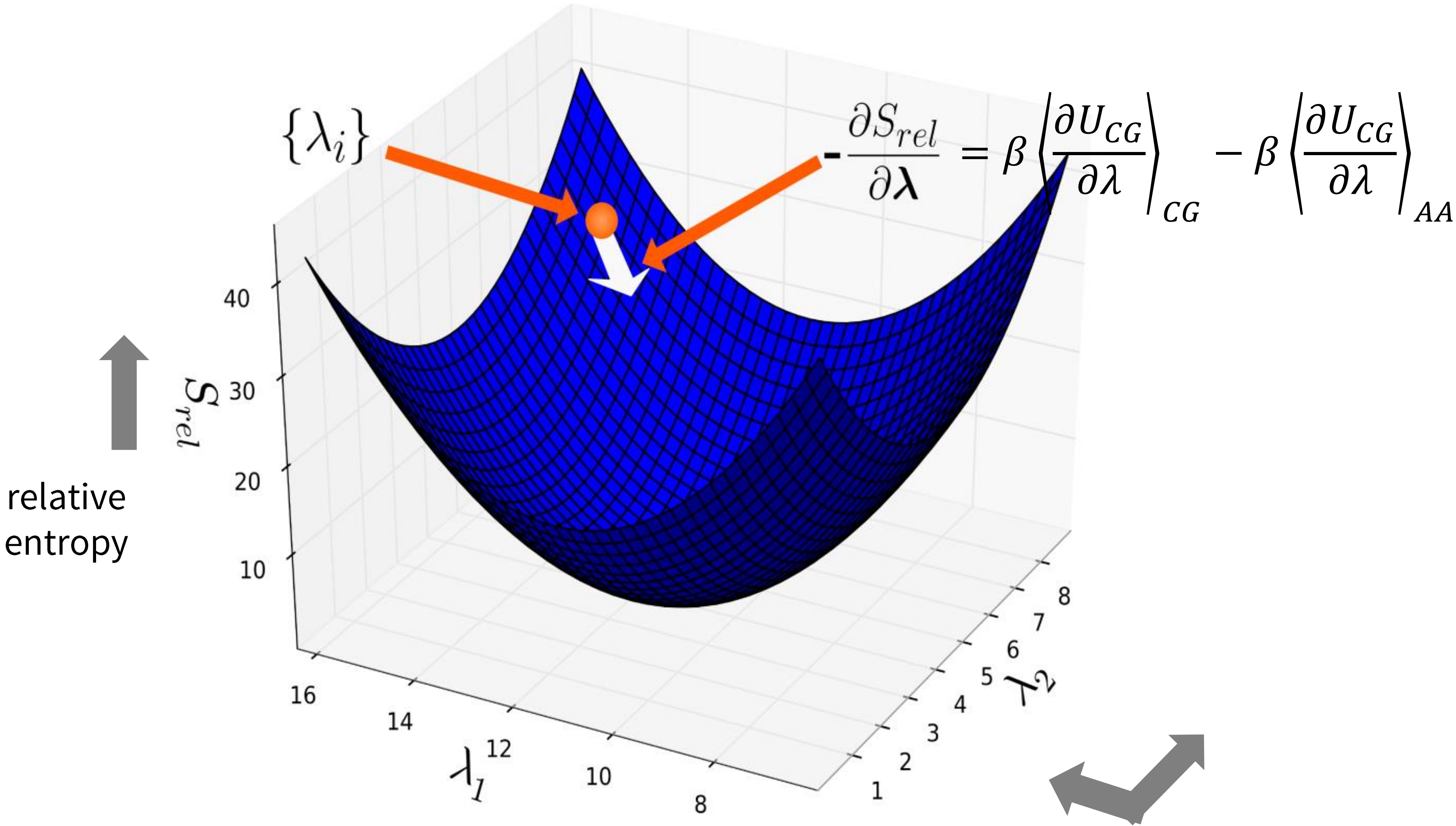
Shell, JCP (2008); Chaimovich and Shell, PRE (2010); Chaimovich and Shell, JCP (2011); Rudzinski and Noid, JCP (2011).

$p_{AA}(i) \propto e^{-\beta U_{AA}(i)}$ $p_{CG}(i) \propto e^{-\beta U_{CG}(i)}$

$$S_{\mathrm{rel}} = \beta \langle U_{CG} - U_{AA} \rangle_{AA} - \beta (A_{CG} - A_{AA})$$

$$\frac{\partial S_{\mathrm{rel}}}{\partial \lambda} = \beta \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{AA} - \beta \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{CG}$$

Algorithm to minimize information loss

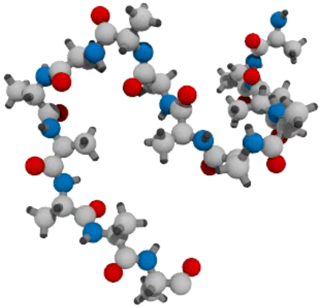


$$-\frac{\partial S_{rel}}{\partial \boldsymbol{\lambda}} = \beta \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{CG} - \beta \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{AA}$$

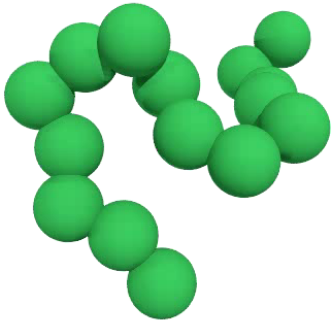
relative entropy

CG parameter space (potentials)

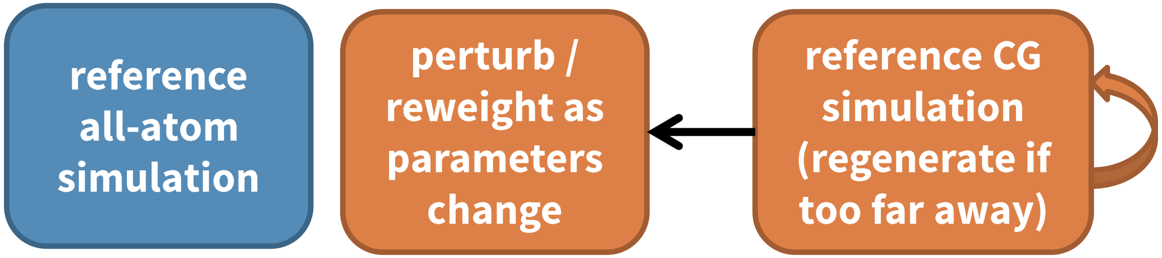
reference all-atom trajectory



adaptive coarse-grained simulations



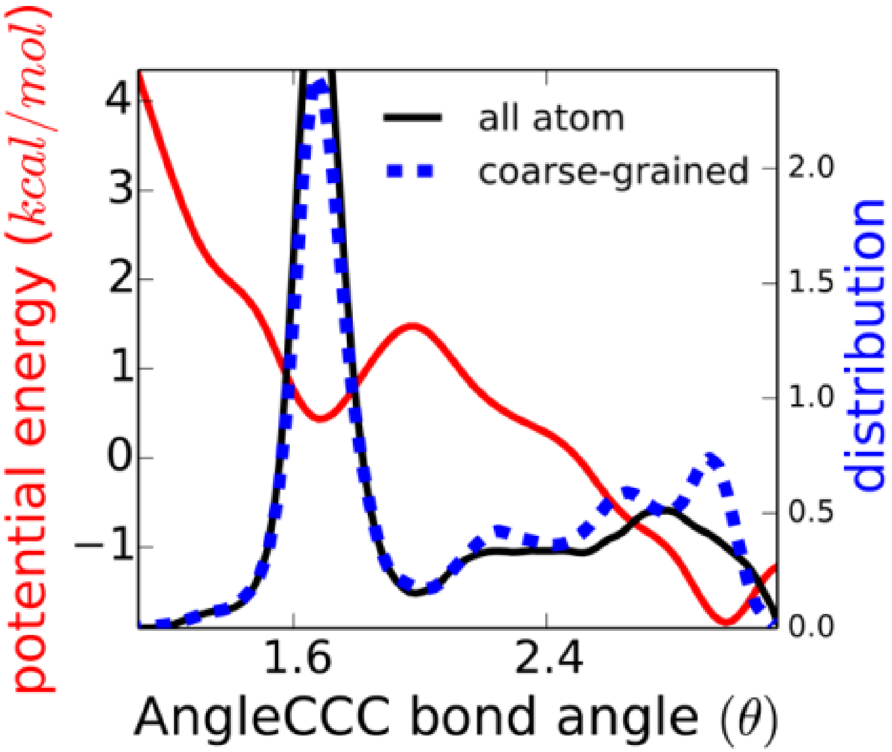
$$\frac{\partial S_{\text{rel}}}{\partial\lambda} = \beta\left\langle\frac{\partial U_{CG}}{\partial\lambda}\right\rangle_{AA} - \beta\left\langle\frac{\partial U_{CG}}{\partial\lambda}\right\rangle_{CG}$$

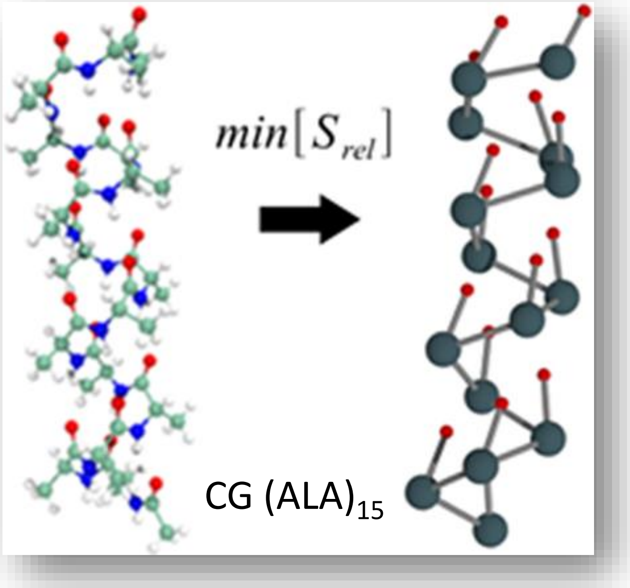


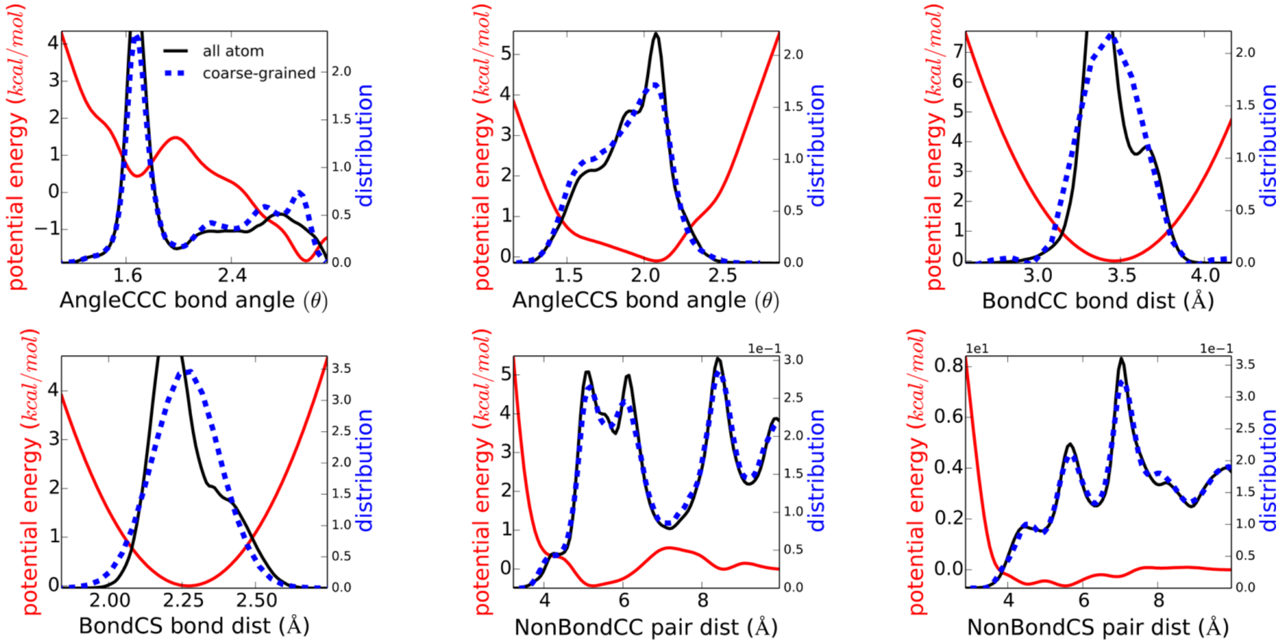
reference all-atom simulation

perturb / reweight as parameters change

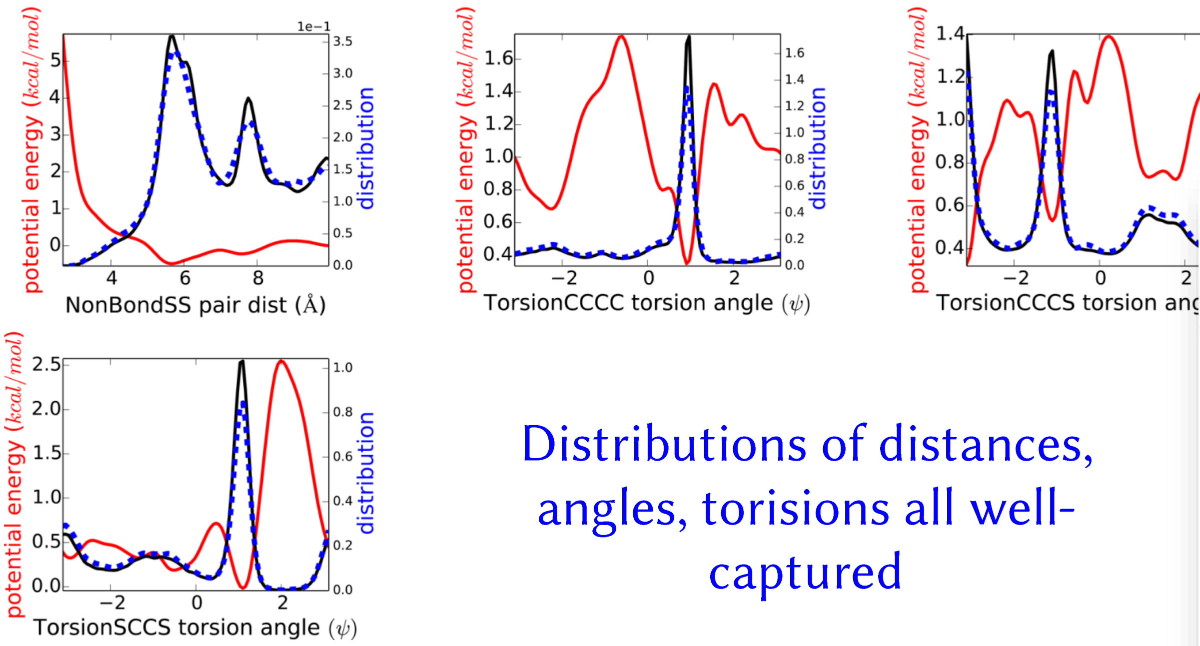
reference CG simulation (regenerate if too far away)


potential energy ($kcal/mol$)
4
3
2
1
0
−1
1.6
2.4
AngleCCC bond angle (θ)
all atom
coarse-grained
2.0
1.5
1.0
0.5
0.0
distribution

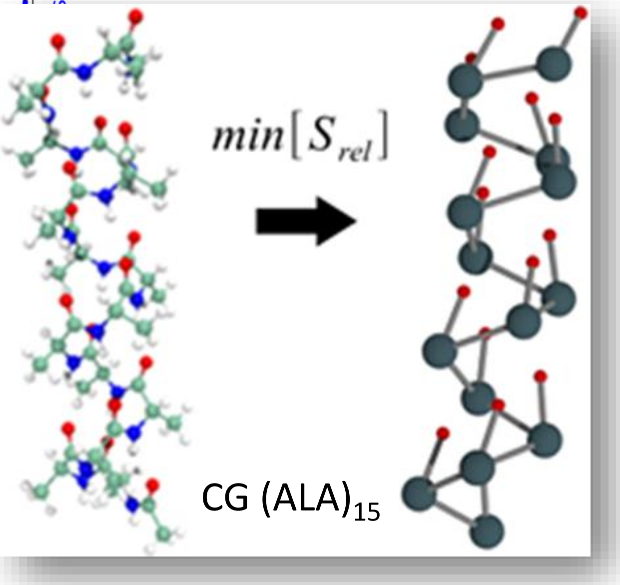

$min[S_{rel}]$
CG $(ALA)_{15}$

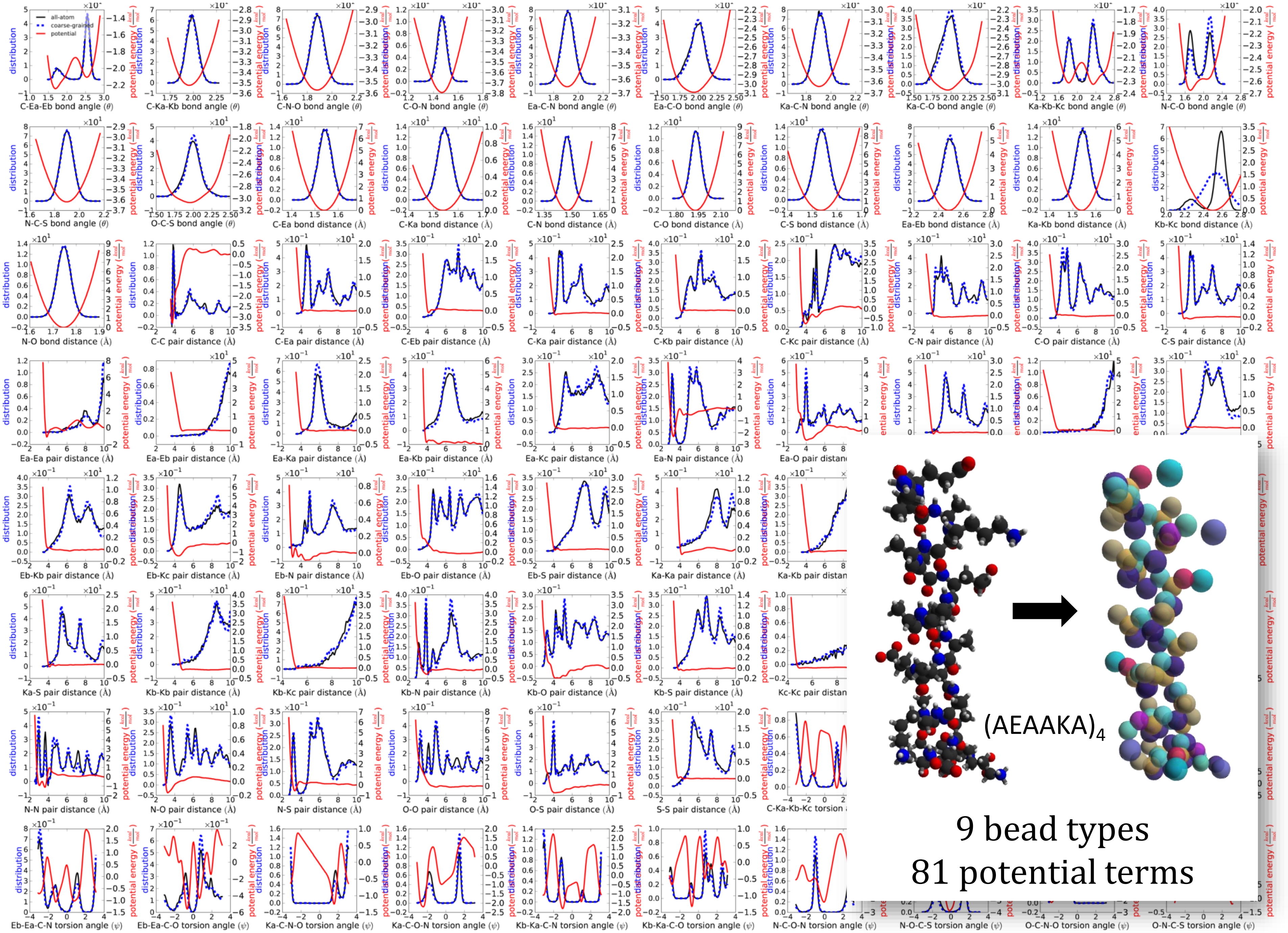


S_{rel} minimization optimizes spline pair nonbonded, angle, and dihedral potentials



Distributions of distances, angles, torisions all well-captured





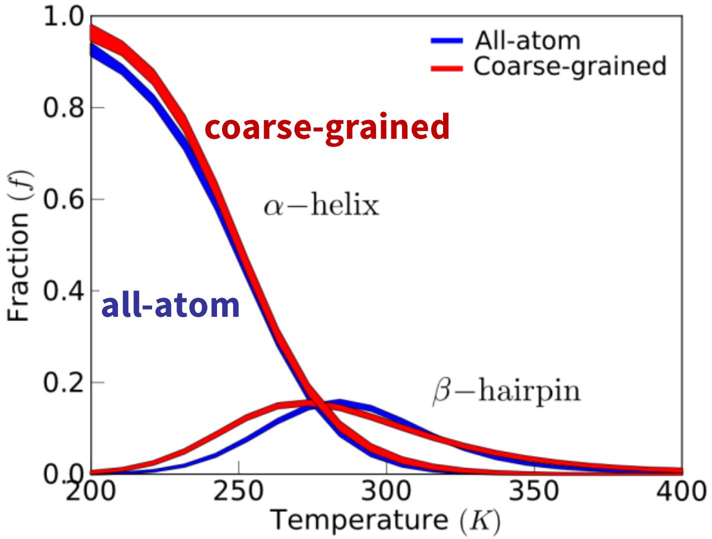
$(AEAAKA)_4$

9 bead types

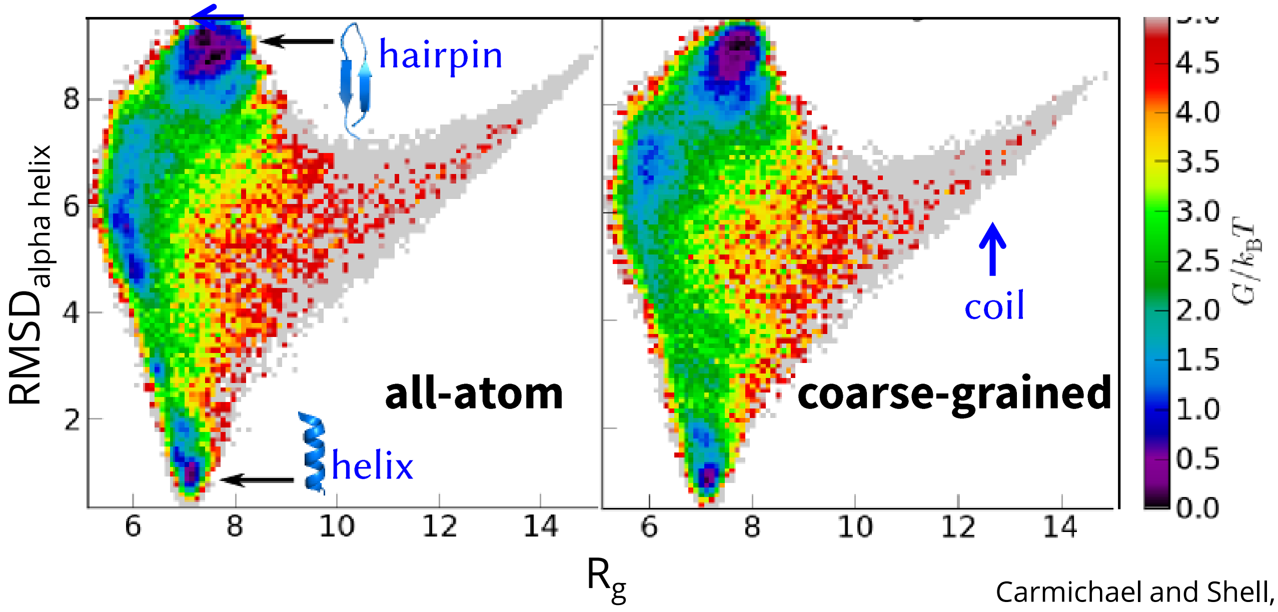
81 potential terms

CG peptide models can be surprisingly accurate

folding curve

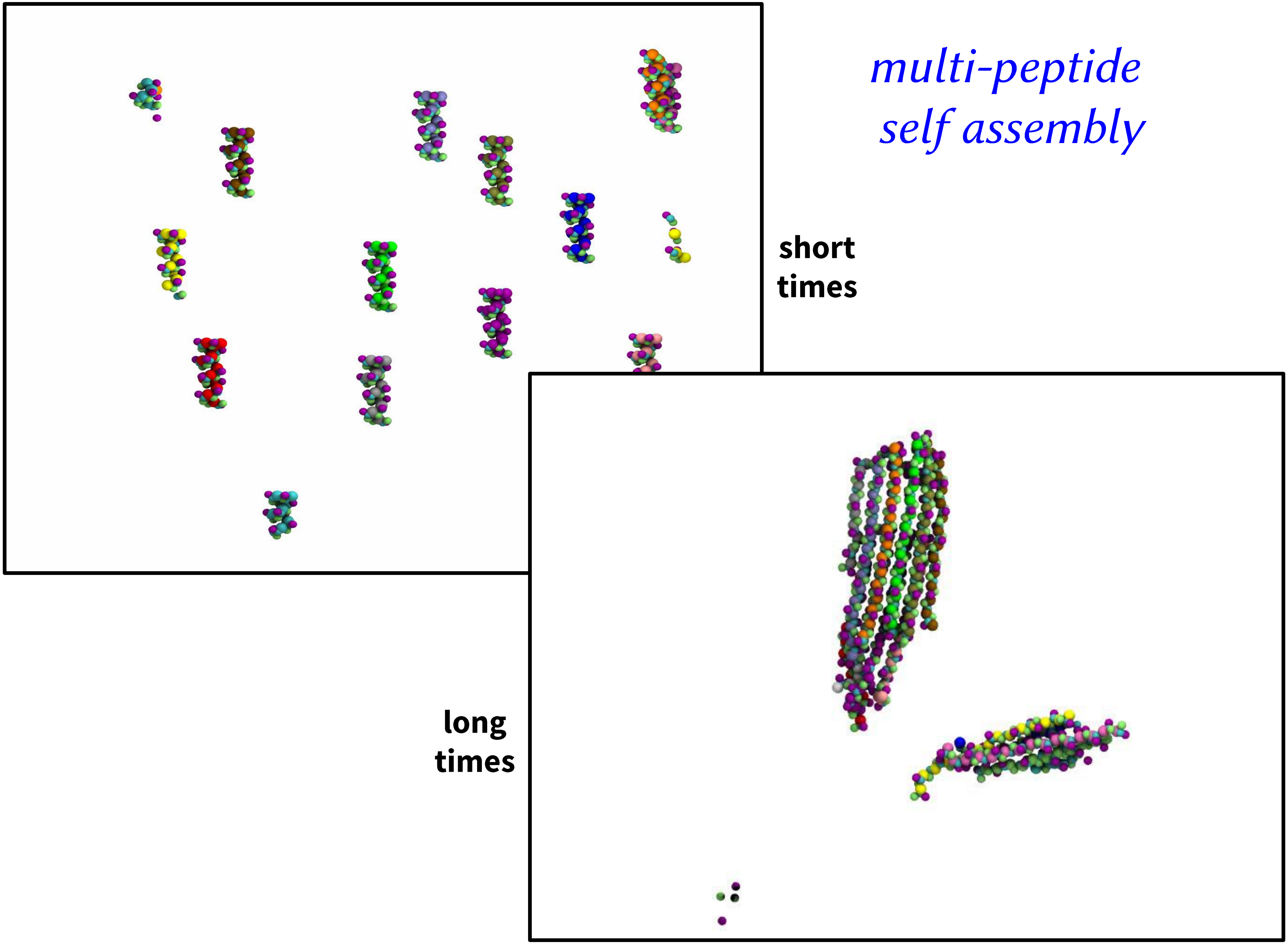


free energy surface



Carmichael and Shell, JCP (2015).

multi-peptide self assembly



short times

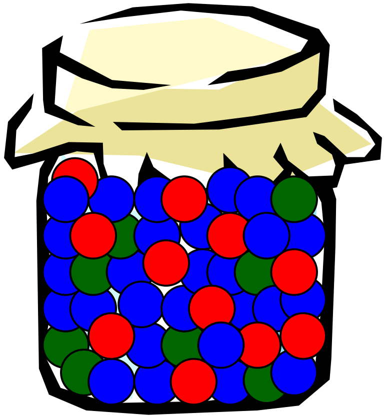
long times

Why should we consider S_{rel} a fundamental statistical-mechanical CG'ing approach?

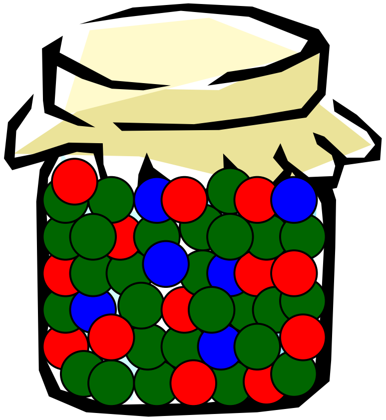
S_{rel} measures the likelihood of sampling correct configurations from the CG model

Shell, JCP (2008); Shell, Adv Chem Phys (2016)

$$S_{\text{rel}} \sim \ln \mathcal{L}$$



all-atom



coarse-grained

marbles → configurations

S_{rel} measures the likelihood of sampling correct configurations from the CG model

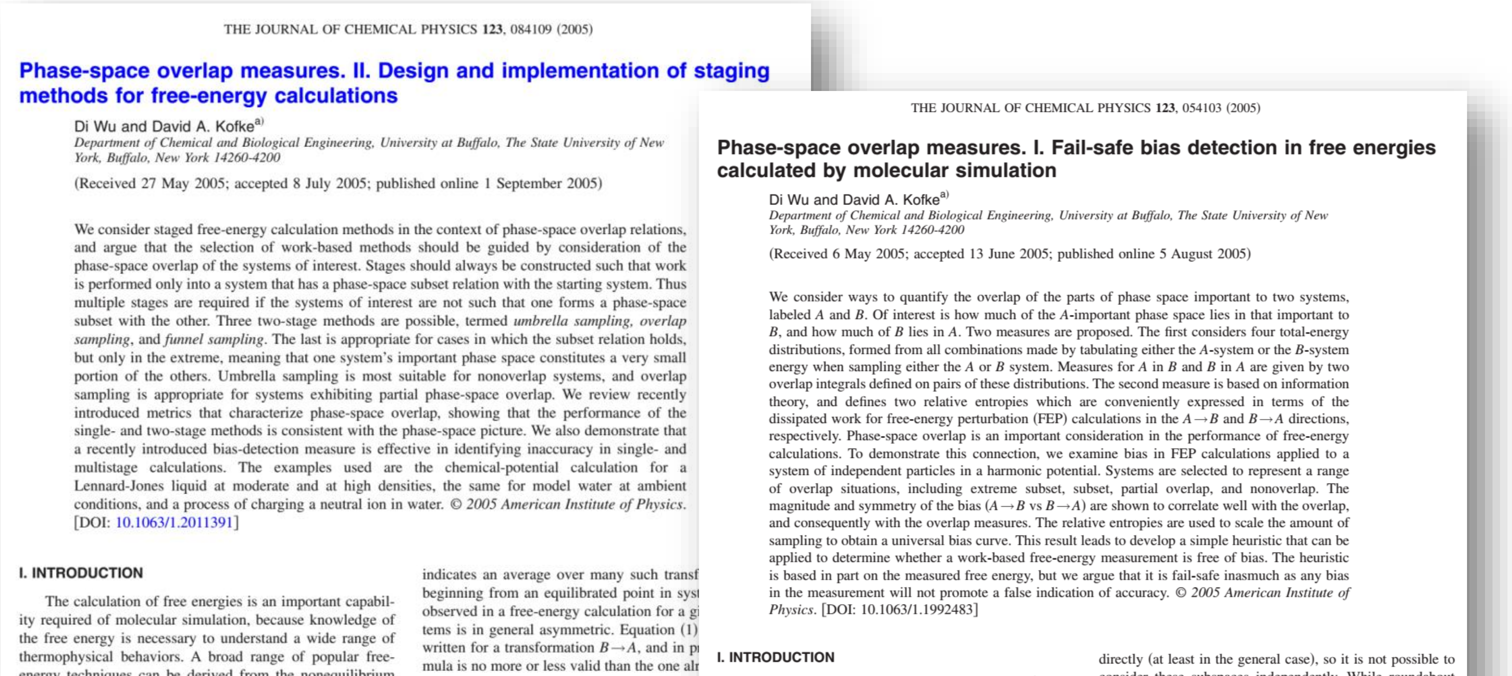
Shell, JCP (2008); Shell, Adv Chem Phys (2016)

effective number of frames during reweighting

S_{rel} measures how hard it is to reweight CG simulations to the true AA ensemble

Wu and Kofke, JCP (2005 x 2); Shell, Adv Chem Phys (2016)

$$\ln \frac{n_{\text{eff}}}{n} = -S_{\text{rel}}$$



THE JOURNAL OF CHEMICAL PHYSICS **123**, 084109 (2005)

Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations

Di Wu and David A. Kofke[a)]
Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200

(Received 27 May 2005; accepted 8 July 2005; published online 1 September 2005)

We consider staged free-energy calculation methods in the context of phase-space overlap relations, and argue that the selection of work-based methods should be guided by consideration of the phase-space overlap of the systems of interest. Stages should always be constructed such that work is performed only into a system that has a phase-space subset relation with the starting system. Thus multiple stages are required if the systems of interest are not such that one forms a phase-space subset with the other. Three two-stage methods are possible, termed *umbrella sampling, overlap sampling*, and *funnel sampling*. The last is appropriate for cases in which the subset relation holds, but only in the extreme, meaning that one system's important phase space constitutes a very small portion of the others. Umbrella sampling is most suitable for nonoverlap systems, and overlap sampling is appropriate for systems exhibiting partial phase-space overlap. We review recently introduced metrics that characterize phase-space overlap, showing that the performance of the single- and two-stage methods is consistent with the phase-space picture. We also demonstrate that a recently introduced bias-detection measure is effective in identifying inaccuracy in single- and multistage calculations. The examples used are the chemical-potential calculation for a Lennard-Jones liquid at moderate and at high densities, the same for model water at ambient conditions, and a process of charging a neutral ion in water. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2011391]

I. INTRODUCTION

The calculation of free energies is an important capability required of molecular simulation, because knowledge of the free energy is necessary to understand a wide range of thermophysical behaviors. A broad range of popular free-energy techniques can be derived from the nonequilibrium

indicates an average over many such transf
beginning from an equilibrated point in syst
observed in a free-energy calculation for a gi
tems is in general asymmetric. Equation (1)
written for a transformation $B \rightarrow A$, and in pr
mula is no more or less valid than the one alr

THE JOURNAL OF CHEMICAL PHYSICS **123**, 054103 (2005)

Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation

Di Wu and David A. Kofke[a)]
Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200

(Received 6 May 2005; accepted 13 June 2005; published online 5 August 2005)

We consider ways to quantify the overlap of the parts of phase space important to two systems, labeled A and B. Of interest is how much of the A-important phase space lies in that important to B, and how much of B lies in A. Two measures are proposed. The first considers four total-energy distributions, formed from all combinations made by tabulating either the A-system or the B-system energy when sampling either the A or B system. Measures for A in B and B in A are given by two overlap integrals defined on pairs of these distributions. The second measure is based on information theory, and defines two relative entropies which are conveniently expressed in terms of the dissipated work for free-energy perturbation (FEP) calculations in the $A \rightarrow B$ and $B \rightarrow A$ directions, respectively. Phase-space overlap is an important consideration in the performance of free-energy calculations. To demonstrate this connection, we examine bias in FEP calculations applied to a system of independent particles in a harmonic potential. Systems are selected to represent a range of overlap situations, including extreme subset, subset, partial overlap, and nonoverlap. The magnitude and symmetry of the bias ($A \rightarrow B$ vs $B \rightarrow A$) are shown to correlate well with the overlap, and consequently with the overlap measures. The relative entropies are used to scale the amount of sampling to obtain a universal bias curve. This result leads to develop a simple heuristic that can be applied to determine whether a work-based free-energy measurement is free of bias. The heuristic is based in part on the measured free energy, but we argue that it is fail-safe inasmuch as any bias in the measurement will not promote a false indication of accuracy. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1992483]

I. INTRODUCTION

directly (at least in the general case), so it is not possible to consider these subspaces independently. While roundabout

S_{rel} measures the likelihood of getting correct configurations from the CG model

Shell, JCP (2008); Shell, Adv Chem Phys (2016)

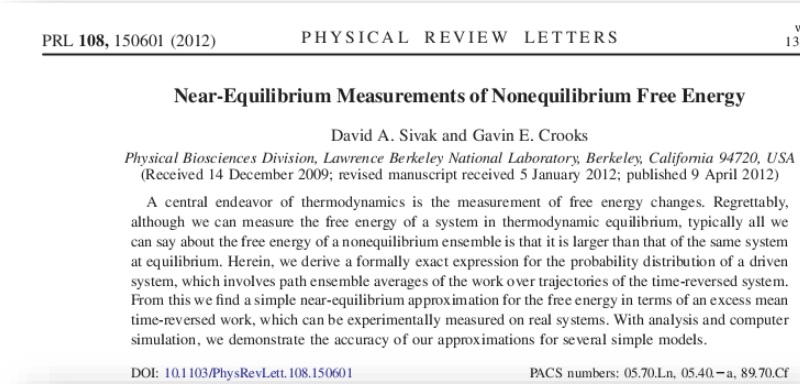
S_{rel} measures how hard it is to reweight CG simulations to the true AA ensemble

Wu and Kofke, JCP (2005 x 2); Shell, Adv Chem Phys (2016)

S_{rel} measures the minimum thermodynamic work to correct for errors in the CG model

Qian, PRE (2001); Crooks & Sivak J Stat Mech (2011); Sivak & Crooks PRL (2012)

$$\frac{W_{\text{rev}}}{k_B T} = \frac{\Delta F}{k_B T} = S_{\text{rel}}$$



PRL **108,** 150601 (2012) PHYSICAL REVIEW LETTERS

Near-Equilibrium Measurements of Nonequilibrium Free Energy

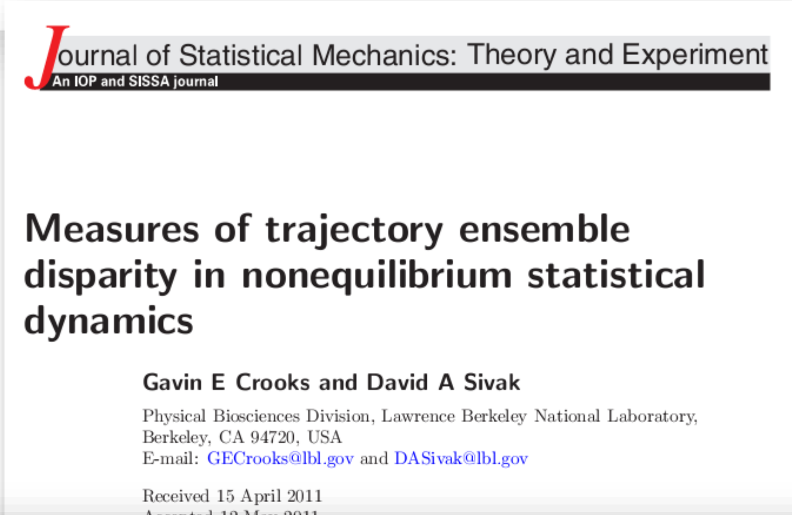
David A. Sivak and Gavin E. Crooks

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

(Received 14 December 2009; revised manuscript received 5 January 2012; published 9 April 2012)

A central endeavor of thermodynamics is the measurement of free energy changes. Regrettably, although we can measure the free energy of a system in thermodynamic equilibrium, typically all we can say about the free energy of a nonequilibrium ensemble is that it is larger than that of the same system at equilibrium. Herein, we derive a formally exact expression for the probability distribution of a driven system, which involves path ensemble averages of the work over trajectories of the time-reversed system. From this we find a simple near-equilibrium approximation for the free energy in terms of an excess mean time-reversed work, which can be experimentally measured on real systems. With analysis and computer simulation, we demonstrate the accuracy of our approximations for several simple models.

DOI: 10.1103/PhysRevLett.108.150601 PACS numbers: 05.70.Ln, 05.40.−a, 89.70.Cf



Journal of Statistical Mechanics: Theory and Experiment

An IOP and SISSA journal

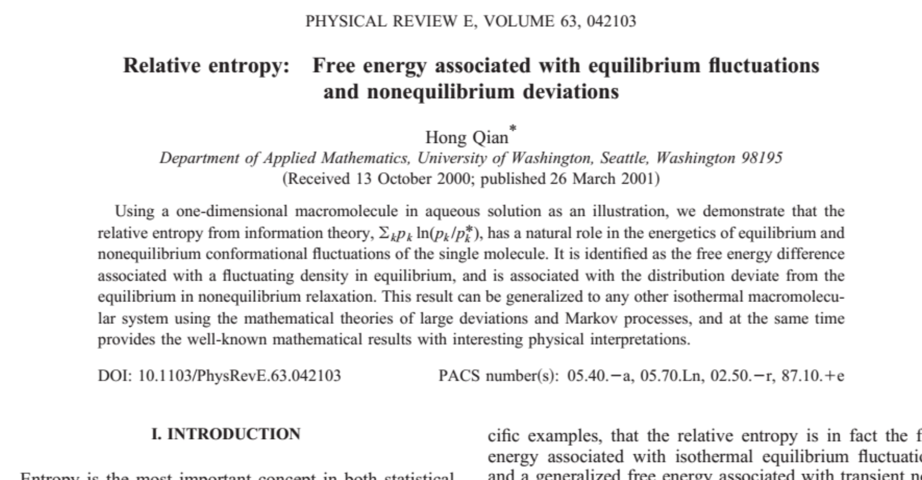
Measures of trajectory ensemble disparity in nonequilibrium statistical dynamics

Gavin E Crooks and David A Sivak

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

E-mail: GECrooks@lbl.gov and DASivak@lbl.gov

Received 15 April 2011



PHYSICAL REVIEW E, VOLUME 63, 042103

Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations

Hong Qian*

Department of Applied Mathematics, University of Washington, Seattle, Washington 98195

(Received 13 October 2000; published 26 March 2001)

Using a one-dimensional macromolecule in aqueous solution as an illustration, we demonstrate that the relative entropy from information theory, $\Sigma_k p_k \ln(p_k/p_k^*)$, has a natural role in the energetics of equilibrium and nonequilibrium conformational fluctuations of the single molecule. It is identified as the free energy difference associated with a fluctuating density in equilibrium, and is associated with the distribution deviate from the equilibrium in nonequilibrium relaxation. This result can be generalized to any other isothermal macromolecular system using the mathematical theories of large deviations and Markov processes, and at the same time provides the well-known mathematical results with interesting physical interpretations.

DOI: 10.1103/PhysRevE.63.042103 PACS number(s): 05.40.−a, 05.70.Ln, 02.50.−r, 87.10.+e

I. INTRODUCTION

cific examples, that the relative entropy is in fact the

energy associated with isothermal equilibrium

S_{rel} measures the likelihood of getting correct configurations from the CG model

Shell, JCP (2008); Shell, Adv Chem Phys (2016)

S_{rel} measures how hard it is to reweight CG simulations to the true AA ensemble

Wu and Kofke, JCP (2005 x 2); Shell, Adv Chem Phys (2016)

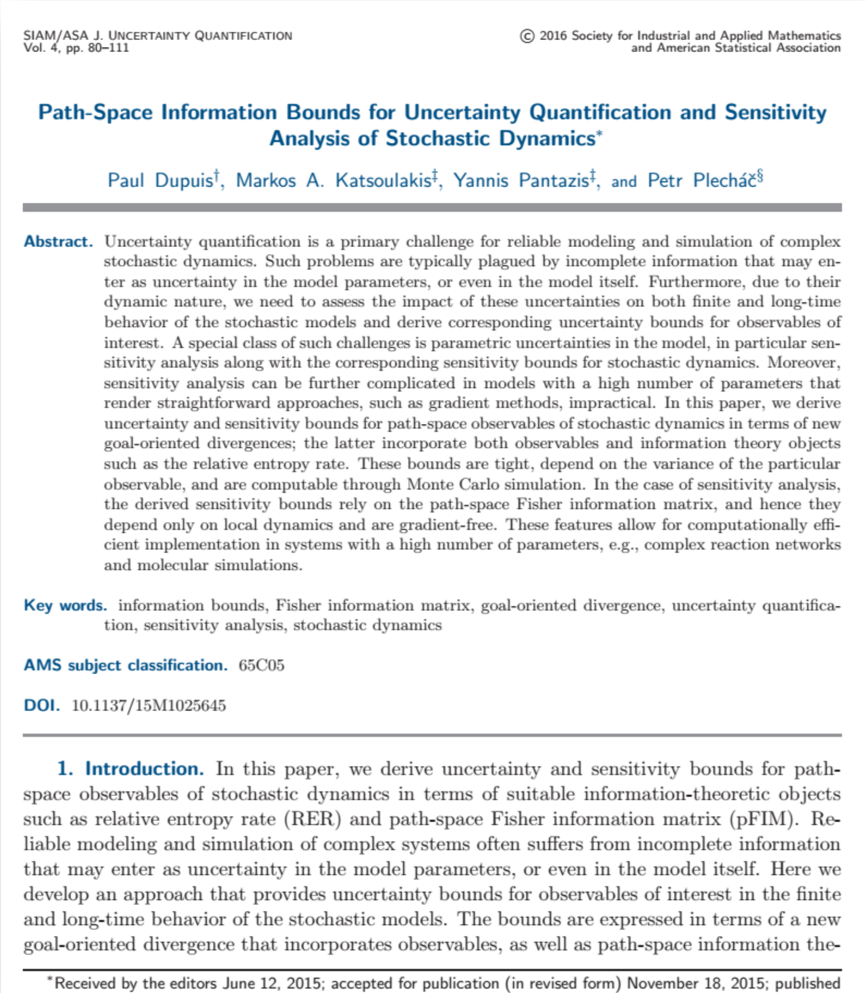
S_{rel} measures the minimum thermodynamic work to correct for errors in the CG model

Qian, PRE (2001); Crooks & Sivak J Stat Mech (2011); Sivak & Crooks PRL (2012)

S_{rel} predicts bounds on errors in observables and properties of the CG model

Chaimovich & Shell, PRE (2010); Chaimovich & Shell JCP (2011);
Dupis, Katsoulakis, Pantazis, & Plechac, SIAM UQ (2016)

$$|\langle X_{AA}\rangle - \langle X_{CG}\rangle| \leq c_X \sqrt{2S_{\text{rel}}}$$



SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 4, pp. 80–111

Path-Space Information Bounds for Uncertainty Quantification and Sensitivity Analysis of Stochastic Dynamics*

Paul Dupuis[†], Markos A. Katsoulakis[‡], Yannis Pantazis[‡], and Petr Plecháč[§]

Abstract. Uncertainty quantification is a primary challenge for reliable modeling and simulation of complex stochastic dynamics. Such problems are typically plagued by incomplete information that may enter as uncertainty in the model parameters, or even in the model itself. Furthermore, due to their dynamic nature, we need to assess the impact of these uncertainties on both finite and long-time behavior of the stochastic models and derive corresponding uncertainty bounds for observables of interest. A special class of such challenges is parametric uncertainties in the model, in particular sensitivity analysis along with the corresponding sensitivity bounds for stochastic dynamics. Moreover, sensitivity analysis can be further complicated in models with a high number of parameters that render straightforward approaches, such as gradient methods, impractical. In this paper, we derive uncertainty and sensitivity bounds for path-space observables of stochastic dynamics in terms of new goal-oriented divergences; the latter incorporate both observables and information theory objects such as the relative entropy rate. These bounds are tight, depend on the variance of the particular observable, and are computable through Monte Carlo simulation. In the case of sensitivity analysis, the derived sensitivity bounds rely on the path-space Fisher information matrix, and hence they depend only on local dynamics and are gradient-free. These features allow for computationally efficient implementation in systems with a high number of parameters, e.g., complex reaction networks and molecular simulations.

Key words. information bounds, Fisher information matrix, goal-oriented divergence, uncertainty quantification, sensitivity analysis, stochastic dynamics

AMS subject classification. 65C05

DOI. 10.1137/15M1025645

1. Introduction. In this paper, we derive uncertainty and sensitivity bounds for path-space observables of stochastic dynamics in terms of suitable information-theoretic objects such as relative entropy rate (RER) and path-space Fisher information matrix (pFIM). Reliable modeling and simulation of complex systems often suffers from incomplete information that may enter as uncertainty in the model parameters, or even in the model itself. Here we develop an approach that provides uncertainty bounds for observables of interest in the finite and long-time behavior of the stochastic models. The bounds are expressed in terms of a new goal-oriented divergence that incorporates observables, as well as path-space information the-

---

*Received by the editors June 12, 2015; accepted for publication (in revised form) November 18, 2015; published

Relative entropy minimization and “matching” strategies

structure

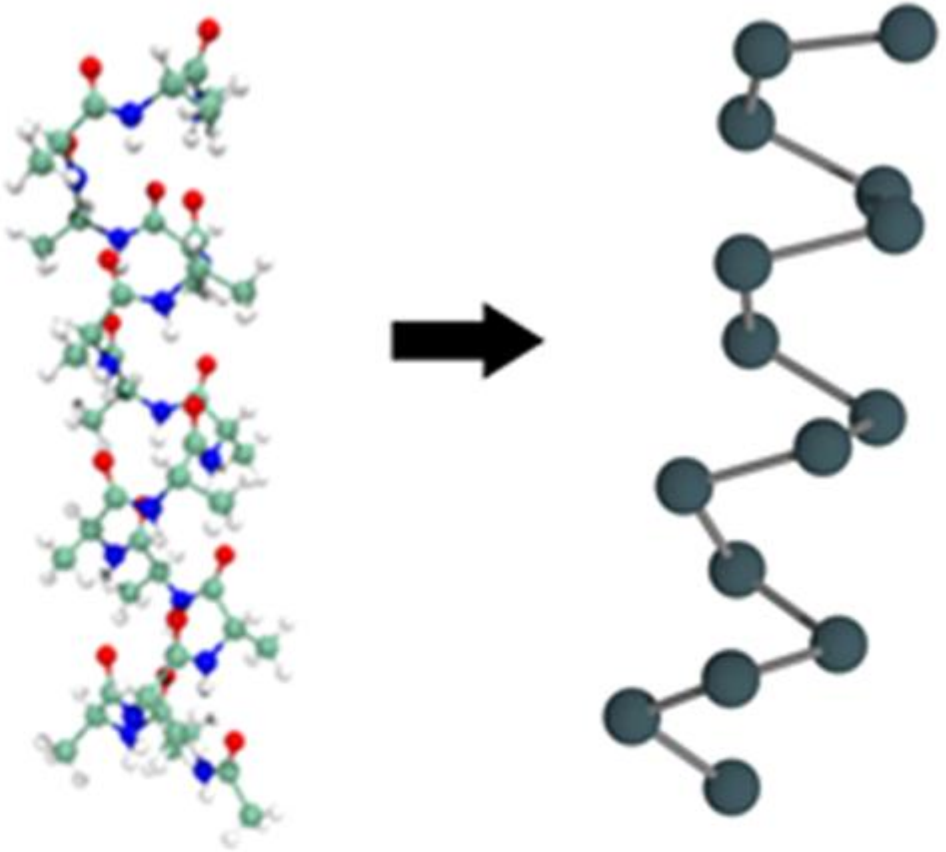
assume pair interactions

energies

assume high T

forces

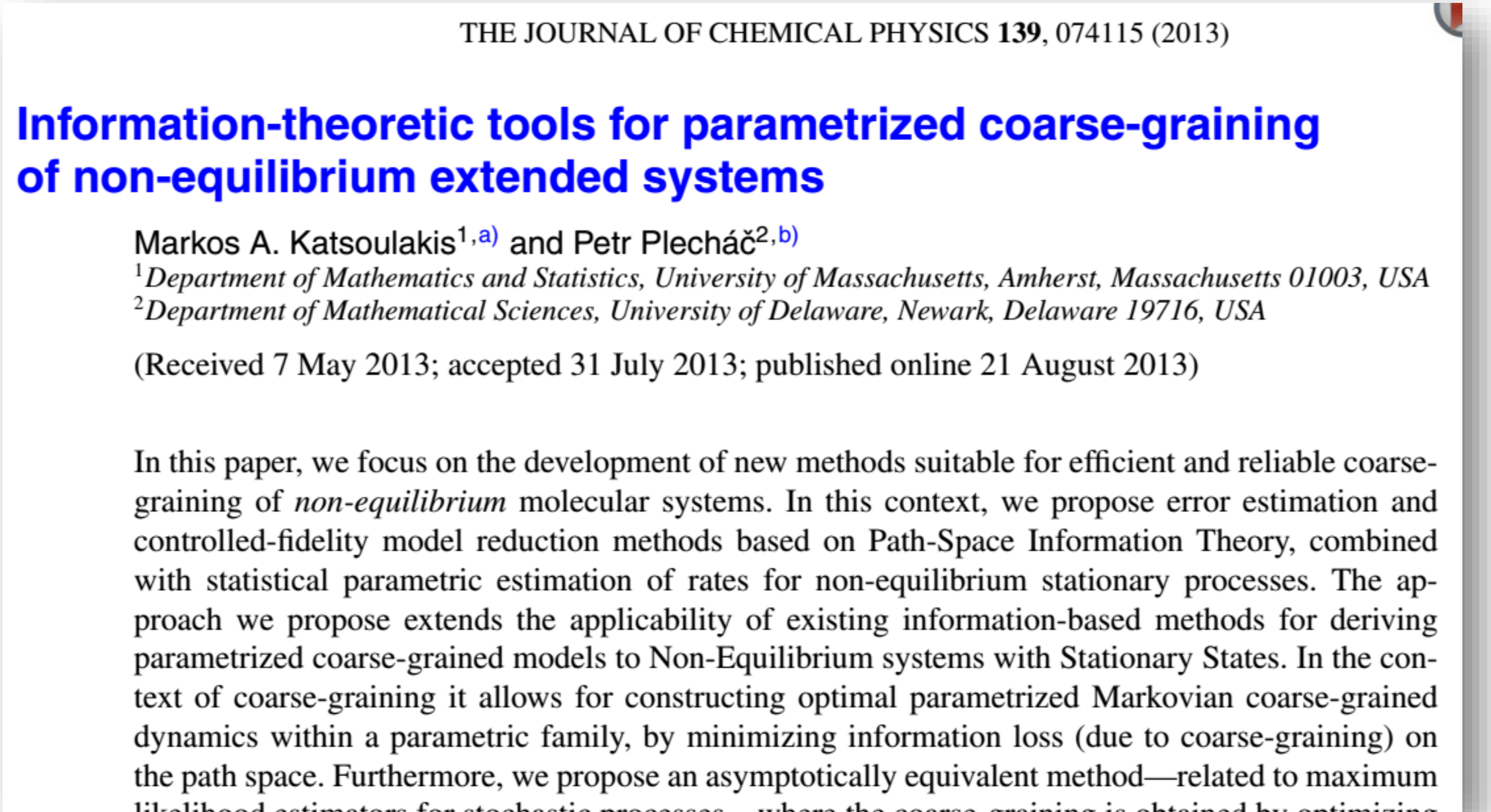
assume multibody interactions



Chaimovich and Shell, JCP (2011)
see also: Rudzinski and Noid, JCP 135, 214101 (2011)

Optimizing coarse dynamic models with S_{rel}

$$S_{\text{rel,path}} = \sum_{\text{paths } i} \mathcal{P}_{AA}(i) \ln\left(\frac{\mathcal{P}_{AA}(i)}{\mathcal{P}_{CG}(i)}\right) \approx \sum_{x} p_{AA}(x) \sum_{x'} p_{AA}(x \to x') \ln\left(\frac{p_{AA}(x \to x')}{p_{CG}(x \to x')}\right)$$



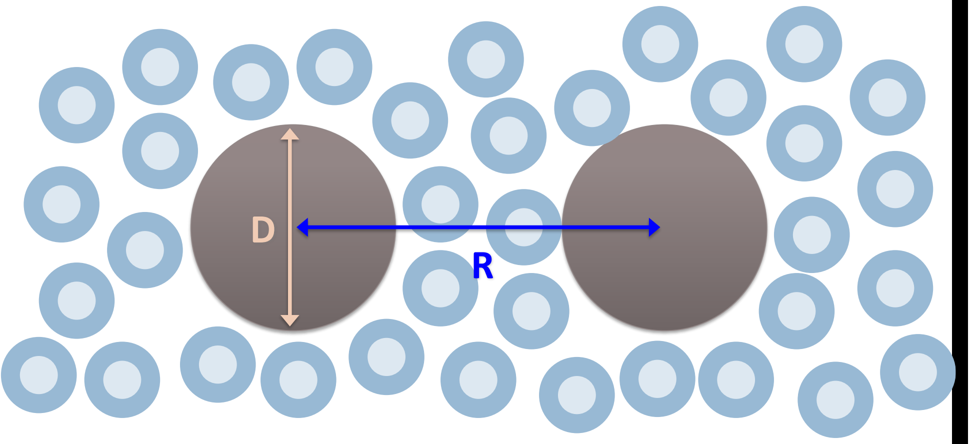
THE JOURNAL OF CHEMICAL PHYSICS **139**, 074115 (2013)

Information-theoretic tools for parametrized coarse-graining of non-equilibrium extended systems

Markos A. Katsoulakis[1,a)] and Petr Plecháč[2,b)]
[1]Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003, USA
[2]Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19716, USA

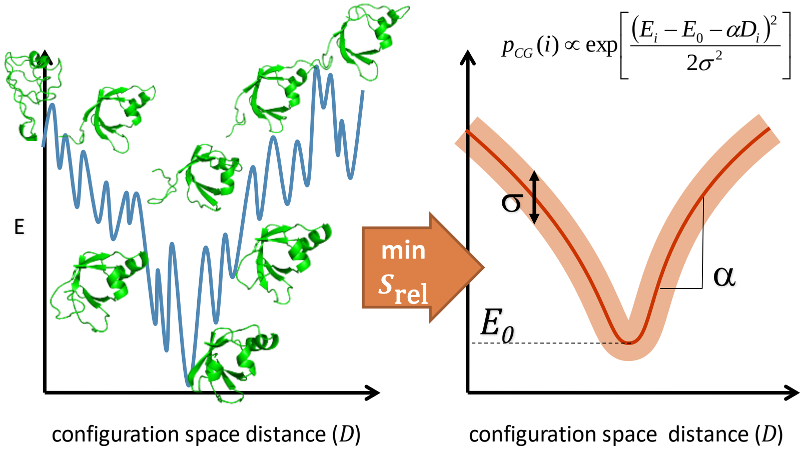
(Received 7 May 2013; accepted 31 July 2013; published online 21 August 2013)

In this paper, we focus on the development of new methods suitable for efficient and reliable coarse-graining of *non-equilibrium* molecular systems. In this context, we propose error estimation and controlled-fidelity model reduction methods based on Path-Space Information Theory, combined with statistical parametric estimation of rates for non-equilibrium stationary processes. The approach we propose extends the applicability of existing information-based methods for deriving parametrized coarse-grained models to Non-Equilibrium systems with Stationary States. In the context of coarse-graining it allows for constructing optimal parametrized Markovian coarse-grained dynamics within a parametric family, by minimizing information loss (due to coarse-graining) on the path space. Furthermore, we propose an asymptotically equivalent method—related to maximum



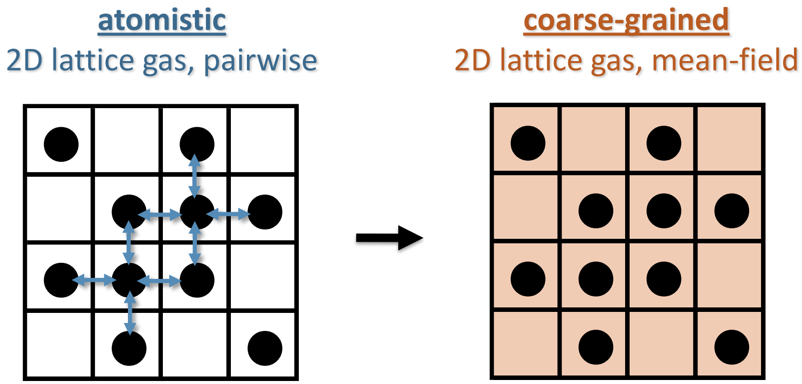
Coarse-graining water and hydrophobic interactions

Chaimovich & Shell PCCP (2009); JCP (2013); JCP (2014)



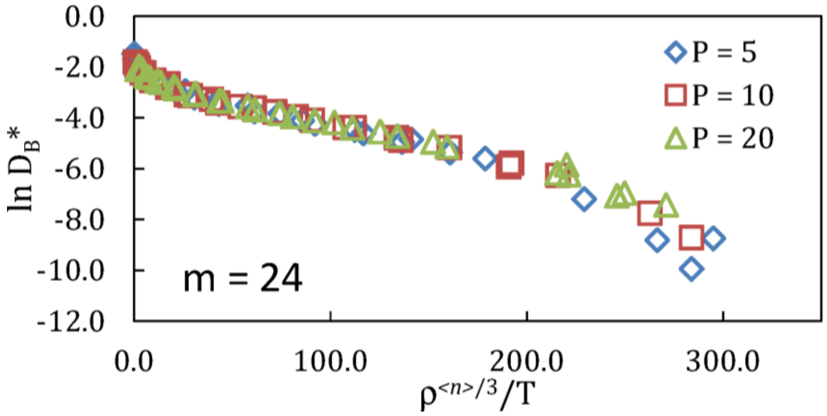
Predicting protein structures by coarse-graining to analytical folding funnels

Pritchard-Bell & Shell, Biophys J (2011)



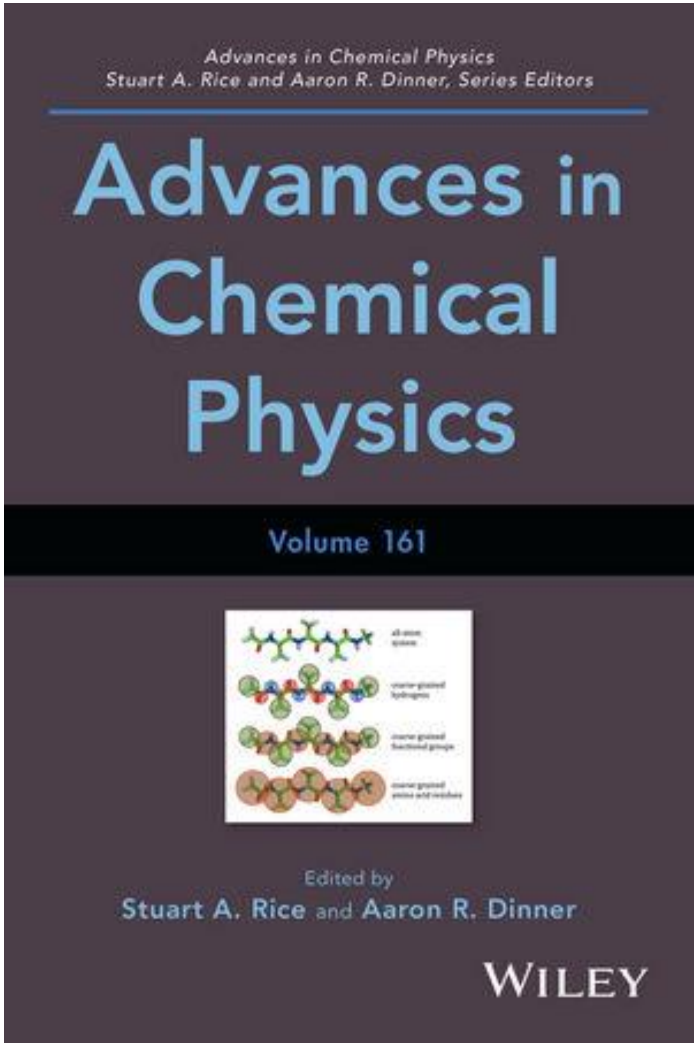
Finding and scoring mean-field closures in simple lattice models

Chaimovich & Shell PRE (2010); JCP (2011)



Finding liquid-state scaling laws by coarse-graining to simpler models

Shell JCP (2012)



Advances in Chemical Physics
Stuart A. Rice and Aaron R. Dinner, Series Editors

Advances in Chemical Physics

Volume 161

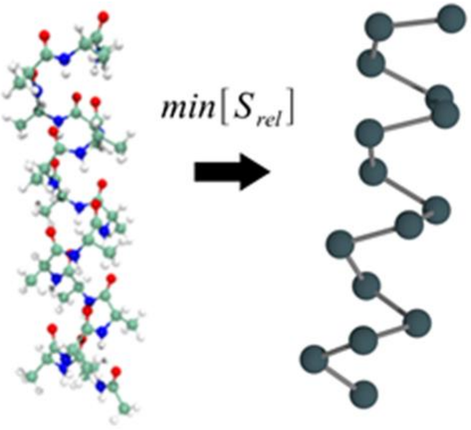
Edited by
Stuart A. Rice and Aaron R. Dinner

WILEY

for theoretical and algorithmic details:

M.S. Shell, *Coarse graining with the relative entropy,* Advances in Chemical Physics 161 (2016)

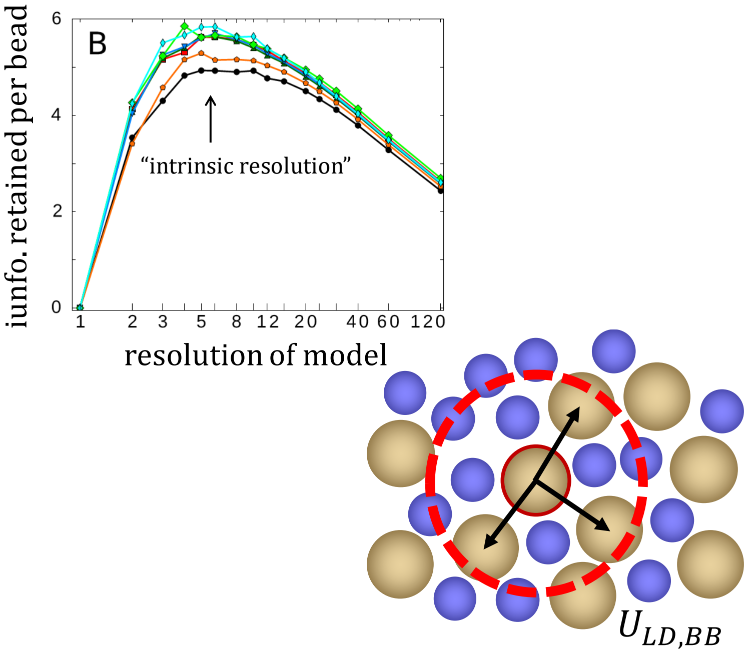
A relative entropy framework for multiscale physics



$$S_{\text{rel}} = \sum_i p_{AA}(i) \ln\left(\frac{p_{AA}(i)}{p_{CG}(i)}\right)$$

Coarse-graining
with the relative entropy

Shell, Adv Chem Phys (2016)
review article



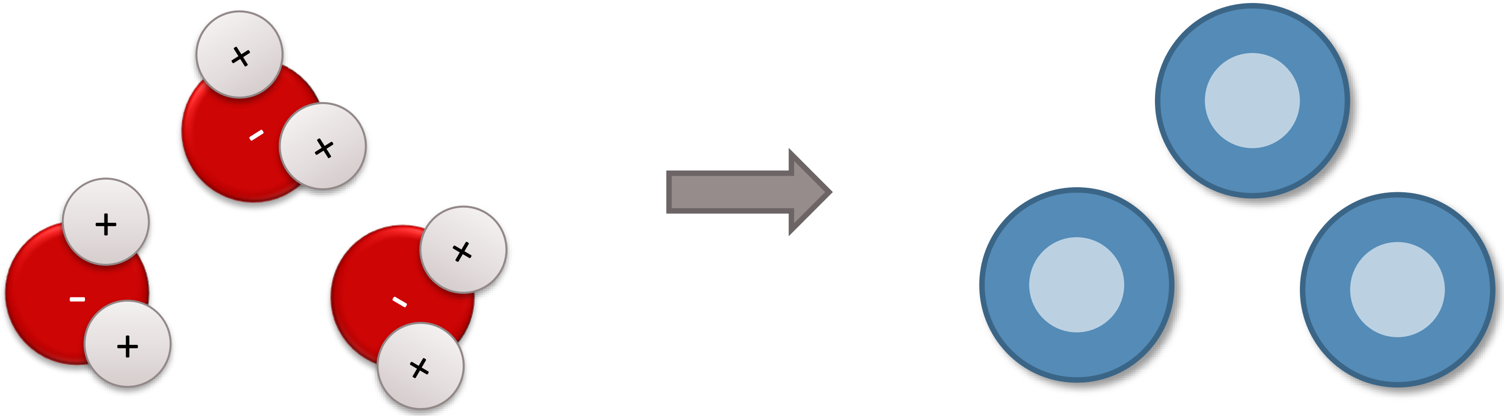
“Next-generation” coarse-grained models for simulation

Foley, Shell, & Noid (2015)
Sanyal & Shell, JCP (2016)
Sanyal & Shell JPCB (2018)

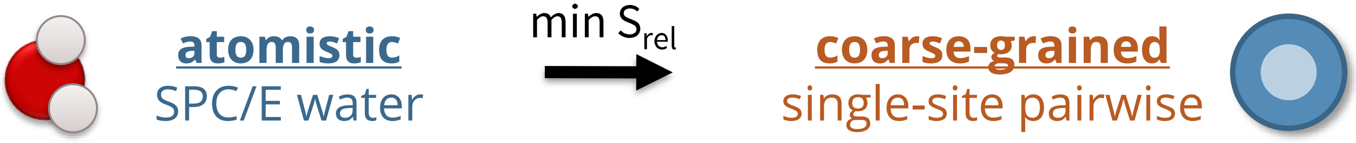
How can we improve CG models of complex liquid-state systems?

A major current challenge:
it’s very difficult to represent the ‘ideal’ CG interactions

$$U_{CG}(\mathbf{R}) \;\rightarrow\; PMF(\mathbf{R}) = -k_B T \ln \int e^{-\beta U_{AA}(\mathbf{r})} \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}] d\mathbf{r}$$



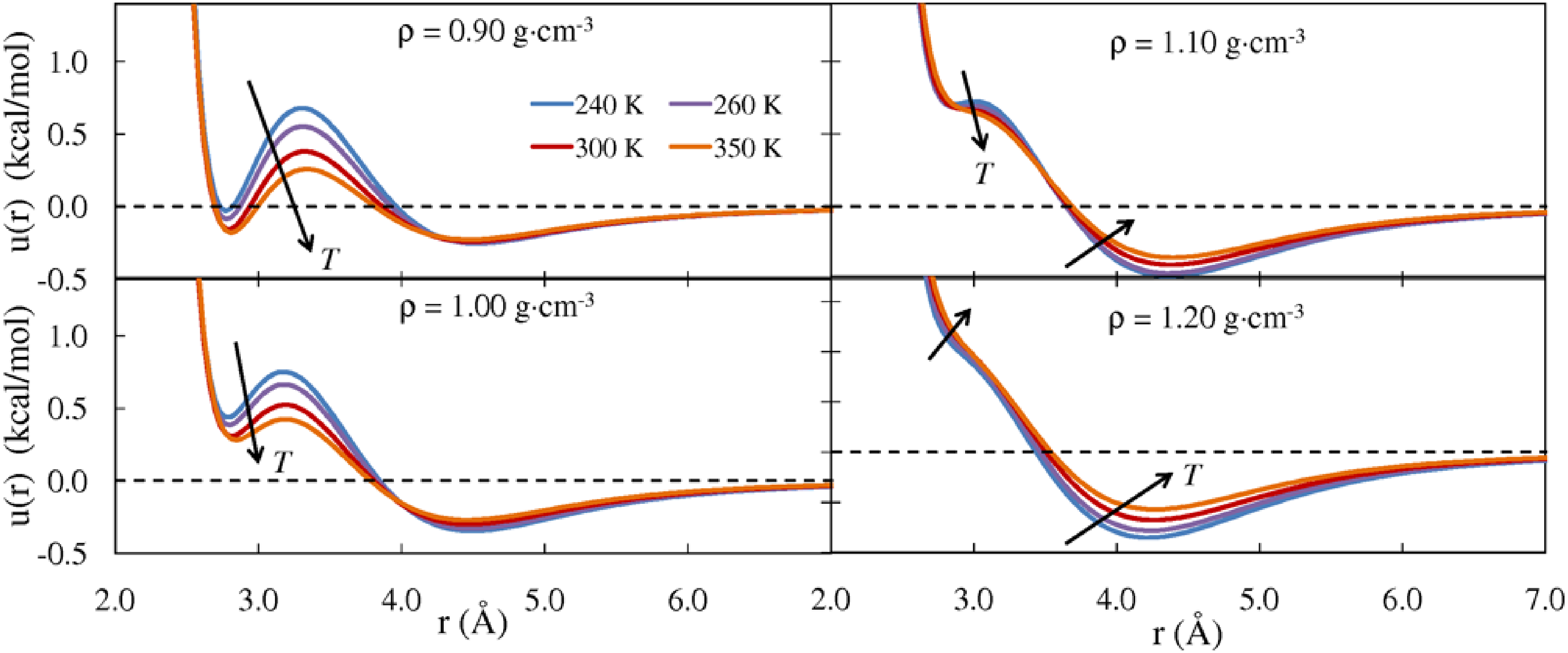
Pairwise interactions compromise transferability



atomistic
SPC/E water

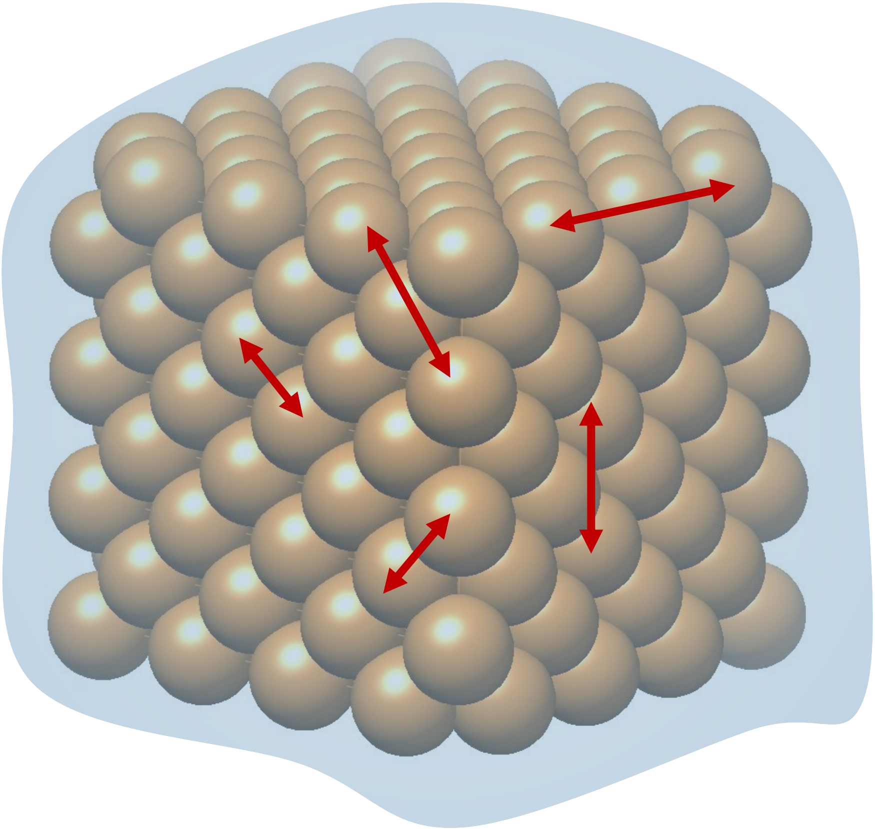
min S_{rel}

coarse-grained
single-site pairwise



Chaimovich and Shell, PCCP (2009)

Multibody interactions inspired by embedded atom models



spatially varying
mean-field
electron density

embedding
function

$$U = \sum_{i<j} u_{pair}(r_{ij}) + \sum_{i} F(\rho_i)$$

$\sum_j \phi(r_{ij})$ count of neighbors

Daw and Baskes (1984)

THE JOURNAL OF CHEMICAL PHYSICS **130**, 034904 (2009)

Evaluating the transferability of coarse-grained, density-dependent implicit solvent models to mixtures and chains

Erik C. Allen and Gregory C. Rutledge[a)]
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

(Received 9 September 2008; accepted 16 November 2008; published online 21 January 2009)

Previously, we described a coarse-graining method for creating local density-depe
solvent (DDIS) potentials that reproduce the radial distribution function (RDF) an
chemical potential across a range of particle concentrations [E. C. Allen and G. C
Chem. Phys. **128**, 154115 (2008)]. In this work, we test the transferability of th
derived from simulations of monomeric solute in monomeric solvent, to mixtures of
solute chains in the same monomeric solvent. For this purpose, "transferability"
predictive capability of the potentials without additional optimization. We f
transferability to mixtures is very good, while RDF errors in systems of chains increas
chain length. Excess chemical potential transferability is good for mixtures
concentration, chains, and chains of mixed composition; at higher solute concentratio
chemical potential transferability fails due to the nature of the DDIS potentials, in
insertion directly affects the interaction potential. With these results, we demonstr
potentials derived for pure solutes can be used effectively in the study of many imp
including those involving mixtures, chains, and chains of mixed composition
solvent. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3055594]

I. INTRODUCTION

Coarse-graining techniques have received growing interest as methods to extend the time and length scales of molecular simulations. In its most essential form, a coarse-graining algorithm is a statistical fitting process that systematically reduces the number of degrees of freedom from that of an all-atom simulation. This reduction is achieved either by aggregating multiple atomic coordinates into a single coarse-grained (CG) particle or by the deletion of particles, as in an implicit solvent simulation. A coarse-graining scheme usually starts with a short simulation in which every particle is explicitly included; this is used to generate descriptive "data" about the exact system. This simulation is followed by a matching procedure in which the CG potential is created and applied to a system involving a reduced number of particles to reproduce the data of the underlying all-particle simulation. Various types of data have been used for CG potential fitting, including forces,[1–7] reversible work,[8] radial distribution functions[9–12] (RDFs), and experimental results.[13–15] Additionally, a wide variety of fitting procedures have been proposed, including simplex optimization,[16–18] RDF inversion,[19–31] wavelet transform,[32,33] and semigrand canonical Monte Carlo (MC).[13]

The value of a CG potential is determined in large part by its utility, or "transferability," to situations outside of the one to which it was fitted because transferability is the only way that the up-front cost of performing the fit can be recouped. As suggested by Johnson *et al.*,[34] transferability can

be classified as one of two t
and "state-point transferabi
(called "representability" by
ity of a CG potential that i
observables to reproduce acc
observables. State-point tran
ability of a CG potential tha
state point (temperature, de
tion) to predict the same obs
state points.

Both types of transfera
dressed in the literature.[35–4
servable transferability of a
in water. The potential utiliz
sequently shown to reprod
profile of a lipid bilayer. No
pic, homogeneous materials
also reproduce the system
algorithms are used to simu
are inaccessible via all-ato
demonstrating observable t
underlying molecular simul
atom simulation cannot be
parison. Further, theoretical
that observable transferabil
cases; for example, he show
RDFs do not reproduce sys
result, he suggested that onl
incorporating density depen
may overcome these rep
Merabia and Pagonabarrag

[a)] Author to whom correspondence should be addressed. Electronic mail: rutledge@mit.edu.

THE JOURNAL OF CHEMICAL PHYSICS **128**, 154115 (2008)

A novel algorithm for creating coarse-grained, density dependent implicit solvent models

Erik C. Allen and Gregory C. Rutledge[a)]
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 15 January 2008; accepted 27 February 2008; published online 18 April 2008)

Implicit solvent simulations are those in which solvent molecules are not explicitly simulated, and the solute-solute interaction potential is modified to compensate for the implicit solvent effect. Implicit solvation is well known in Brownian dynamics of dilute solutions but offers promise to speed up many other types of molecular simulations as well, including studies of proteins and colloids where the local density can vary considerably. This work examines implicit solvent potentials within a more general coarse-graining framework. While a pairwise potential between solute sites is relatively simple and ubiquitous, an additional parametrization based on the local solute concentration has the possibility to increase the accuracy of the simulations with only a marginal increase in computational cost. We describe here a method in which the radial distribution function and excess chemical potential of solute insertion for a system of Lennard–Jones particles are first measured in a fully explicit, all-particle simulation, and then reproduced across a range of solute particle densities in an implicit solvent simulation. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2899729]

I. INTRODUCTION

Implicit solvent models are those in which the explicit treatment of solvent-solvent and solvent-solute interactions is replaced by a modified solute-solute interaction potential designed to capture some of the solvent-mediated behavior. Typically, the solvent is assumed to be much smaller, and to equilibrate much faster, than the solute. These models have received a great deal of interest in recent years, from arenas including proteins,[1] surfactants,[2] membranes,[3] tissues,[4] and macromolecules.[5–7] Because many of the systems of interest are solvent rich, the implicit solvent formulation offers the possibility to improve greatly the efficiency of computer simulations of such systems, because it permits a vast reduction in the number of degrees of freedom.

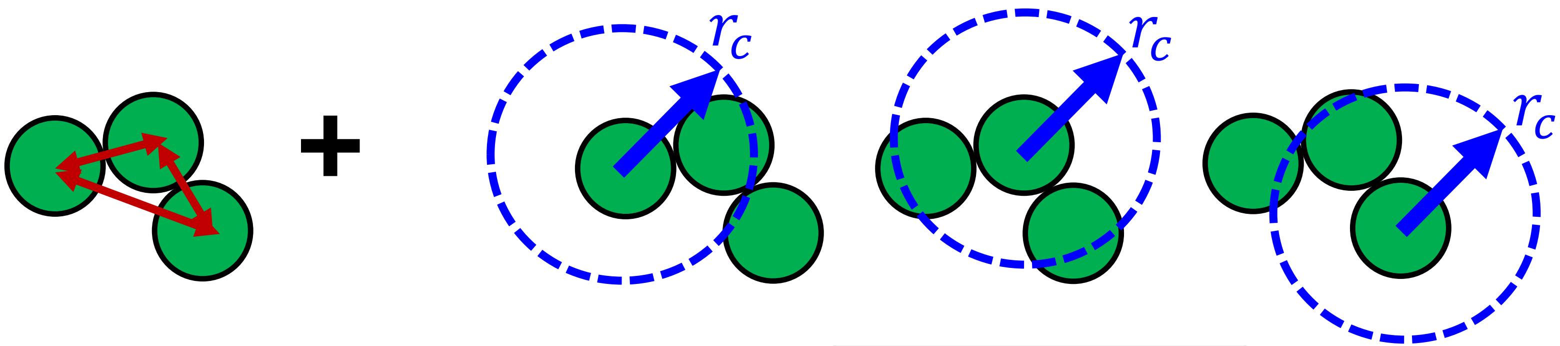
Here, we examine the process of creating an implicit solvent potential within the coarse-graining framework.[8] In coarse graining, one seeks to create a reduced model of a molecular system that matches some properties of an underlying all-atom simulation. A coarse-graining scheme usually involves a short simulation in which every particle is explicitly included, which is used to generate descriptive "data" about the exact system. This simulation is followed by a matching procedure in which the coarse-grained potential is created and applied to a reduced number of particles to reproduce the data of the underlying all-particle simulation. In many coarse-graining applications, a small number of particles or degrees of freedom are combined and represented number of particles (the "solvent") while retaining others (the "solute") in unaltered form, except for the requisite correction to the solute interaction potential to reflect solvent-mediated interaction. The resulting coarse-grain interaction potential is a potential of mean force between solute particles that serves to preserve the solution structure. Reproducing the full thermodynamic behavior of the all-particle simulation, however, requires some additional accounting for the lost solvent-solvent interactions. In contrast to the traditional assumptions of Brownian dynamics that allow one to treat the solvent implicitly as a heat bath, the coarse-graining method is inherently a fitting process that seeks to optimize a trade-off between accuracy and computational efficiency through a reduction of degrees of freedom between interacting particles.

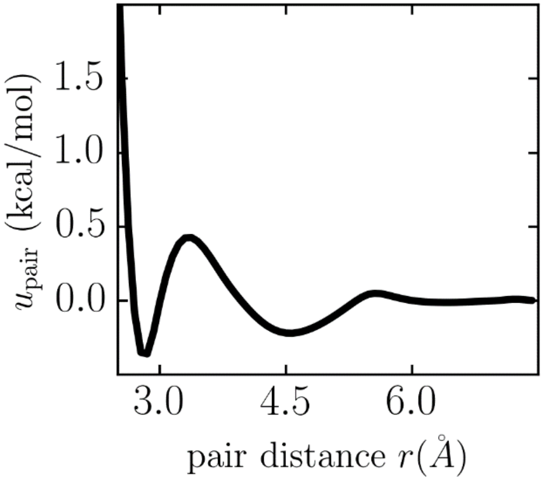
To explain our approach here, we begin with three basic questions that have guided this work:

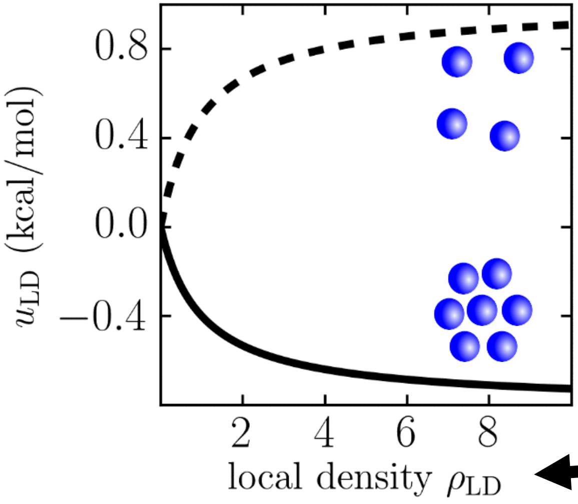
(1) What property or properties does one seek to conserve during the coarse-graining process to maximize the utility of the resulting potential?
(2) What is the simplest possible potential form that can sufficiently capture the property or properties of interest?
(3) Once a potential form has been selected, how does one optimize the potential to reproduce the properties of interest?

Conceptually new CG models with mean-field multibody interactions

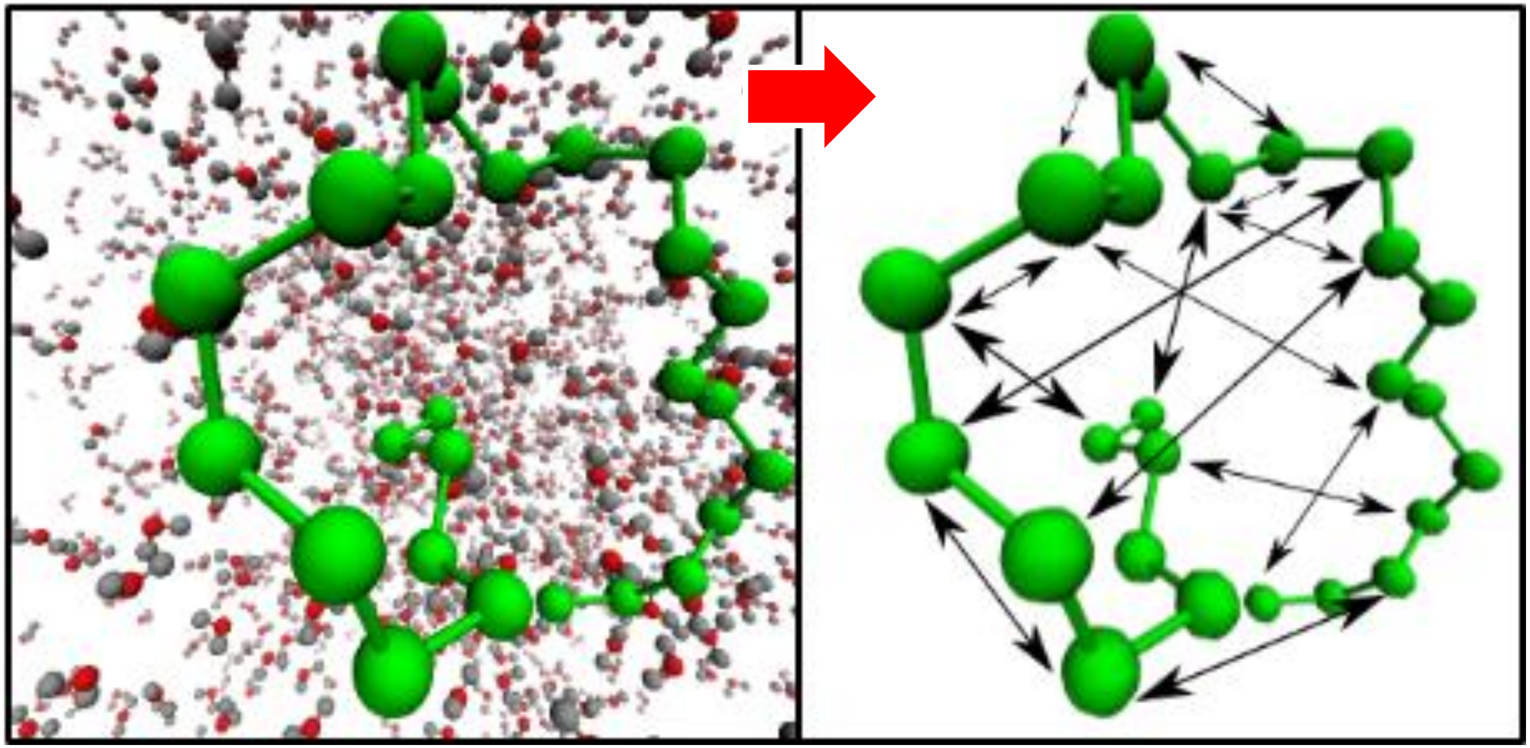
$$U = \sum_{i<j} u_{pair}(r_{ij}) + \sum_{i} f(\rho_i)$$







local coordination number

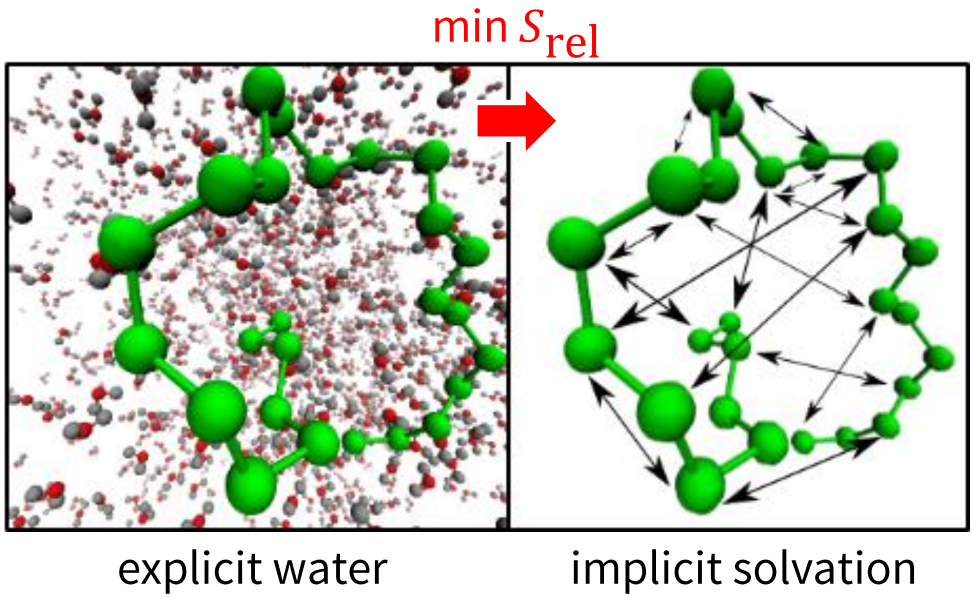


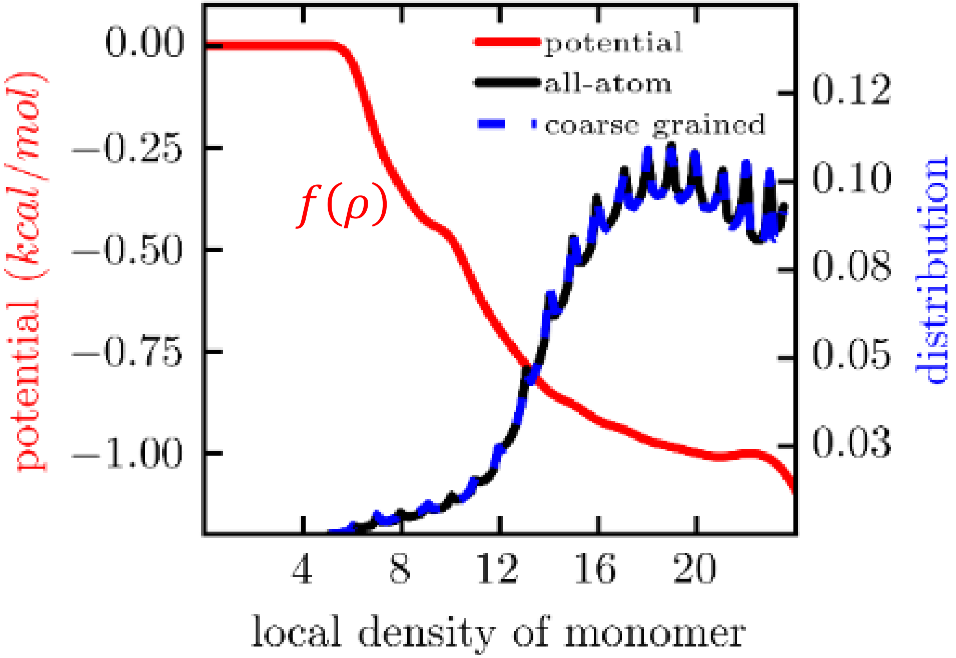
min S_{rel}

explicit water

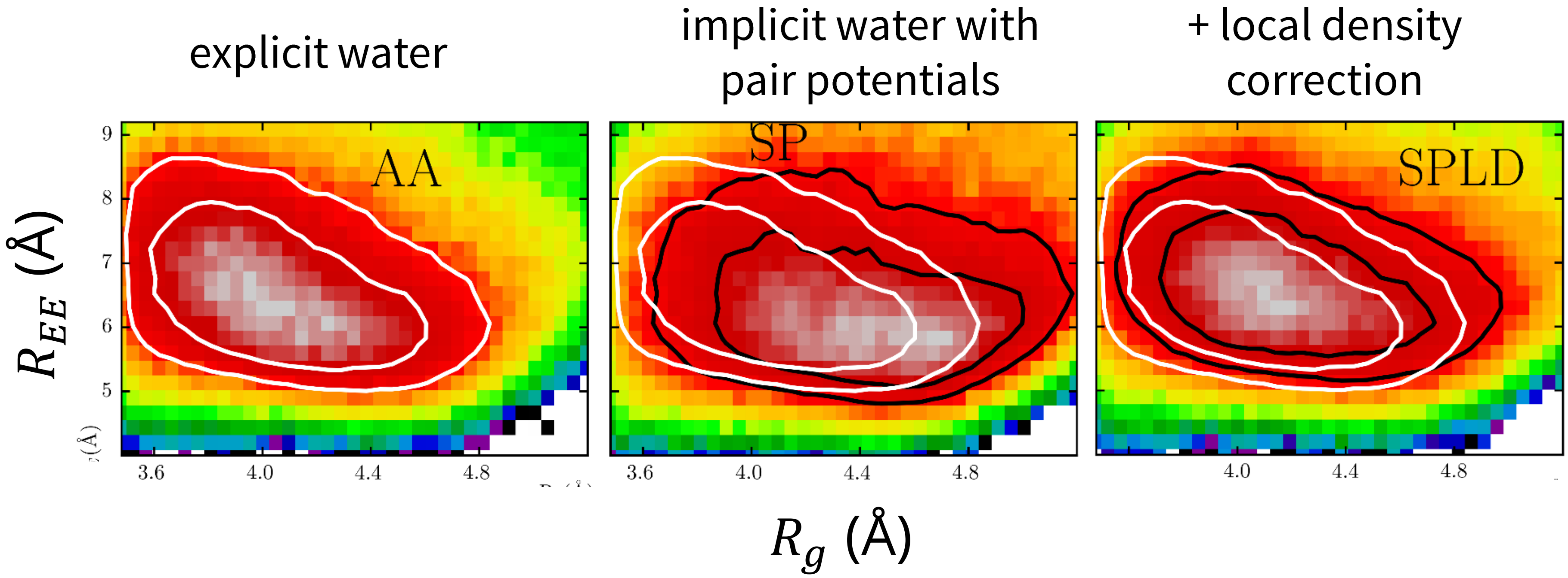
implicit solvation

hydrophobic 25mer alkane
polymer in water
Athalwe et al, PNAS 2007


min S_{rel}
explicit water
implicit solvation


potential
all-atom
coarse grained
$f(\rho)$
potential $(kcal/mol)$
distribution
local density of monomer
0.00
−0.25
−0.50
−0.75
−1.00
0.12
0.10
0.08
0.05
0.03
4
8
12
16
20

Hydrophobic polymer folding in water



explicit water

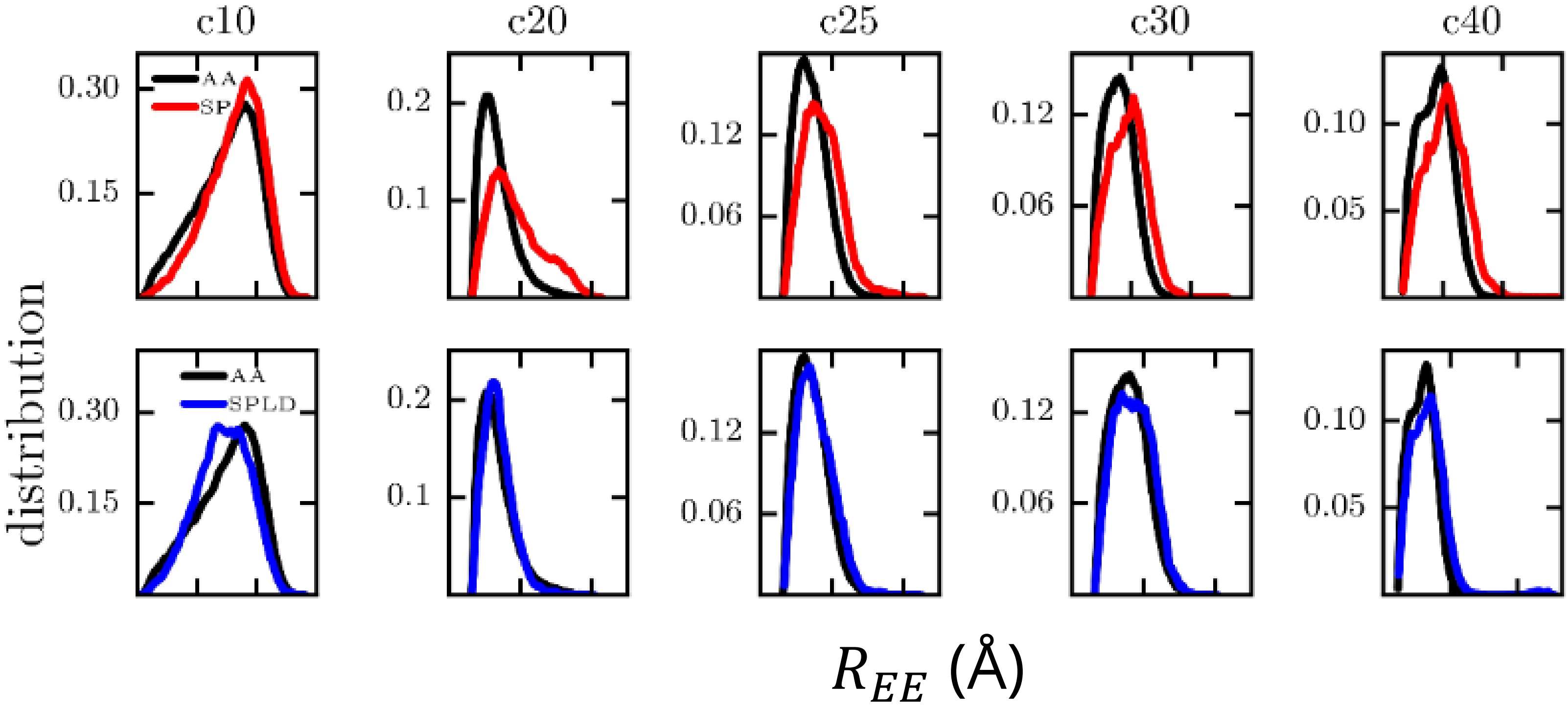
implicit water with pair potentials

+ local density correction

Sanyal and Shell, J Chem Phys (2016)

Transferability to different polymer lengths

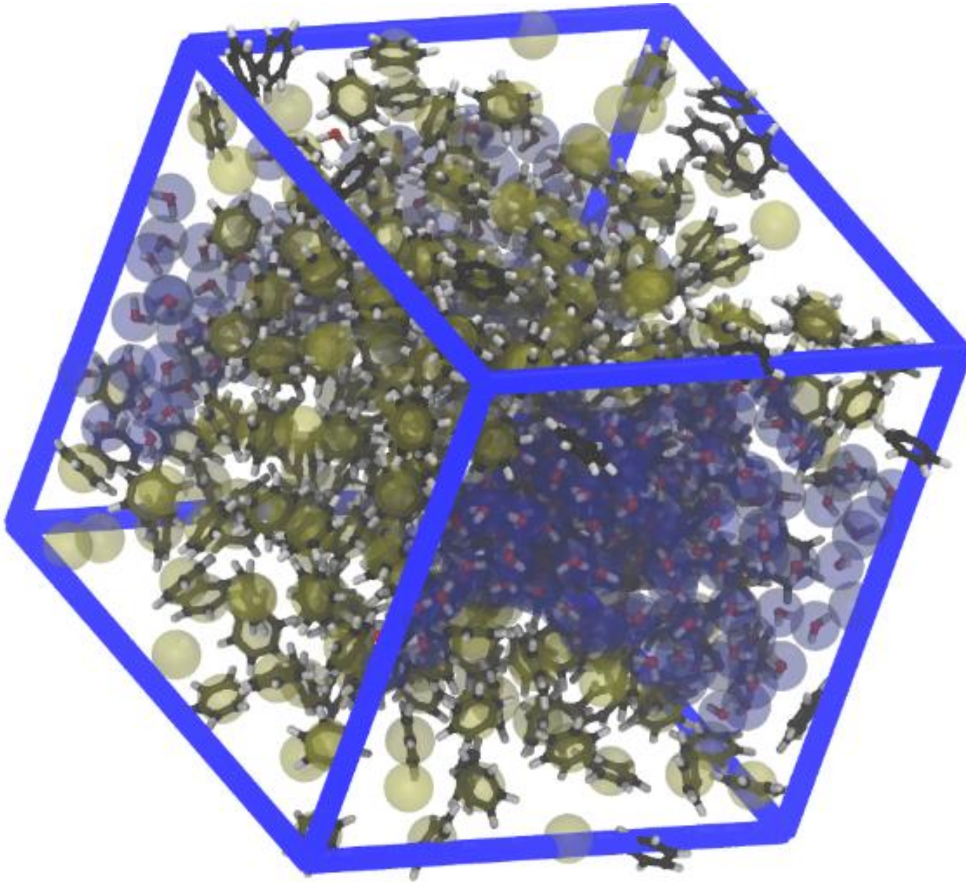
shorter polymers **original 25mer** longer polymers

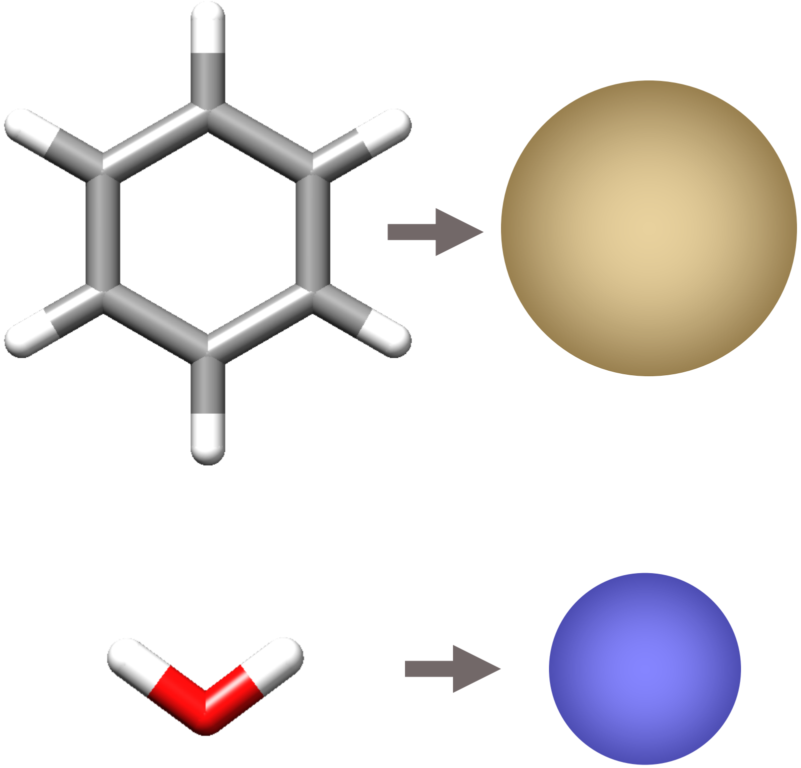


implicit with pair potentials

+ local density correction

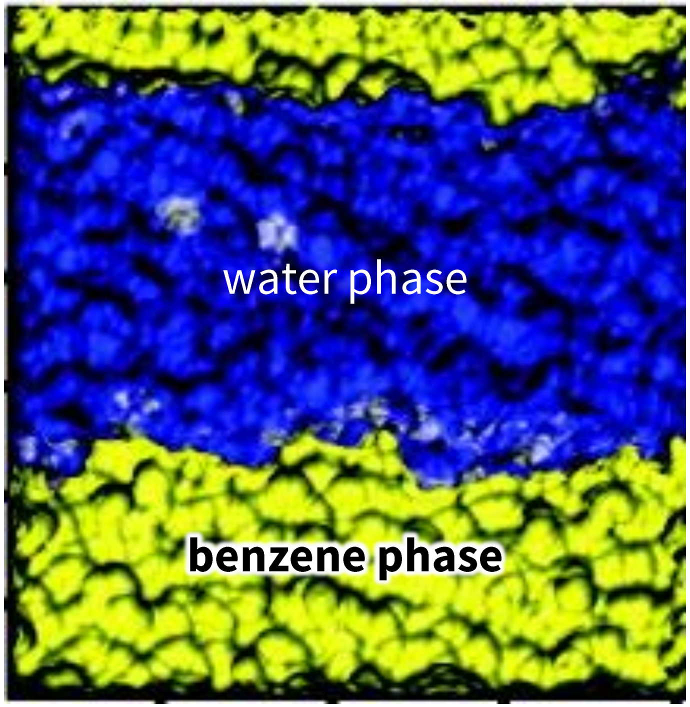
Expanding CG models to phase equilibria



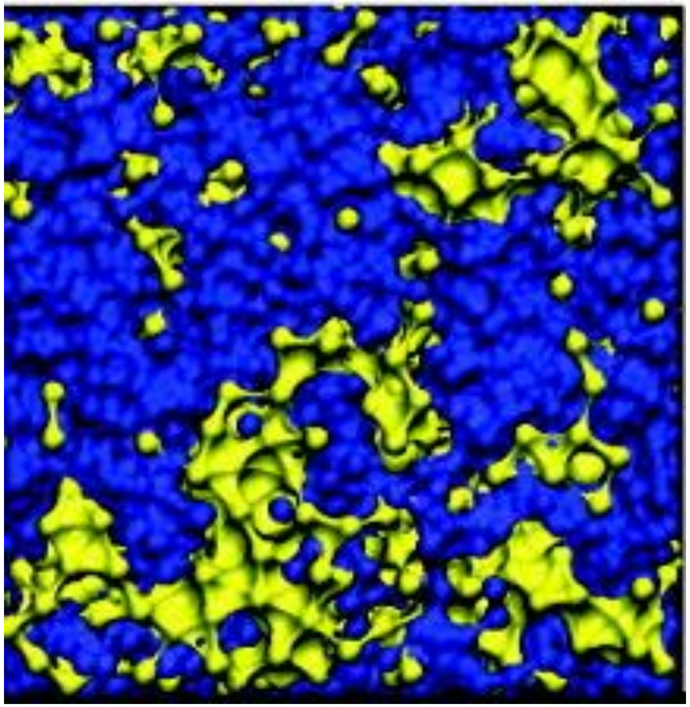


model CG system studied by Villa, Peter, & van der Vegt, JCTC (2010)

Challenges in modeling phase separation without multibody interactions



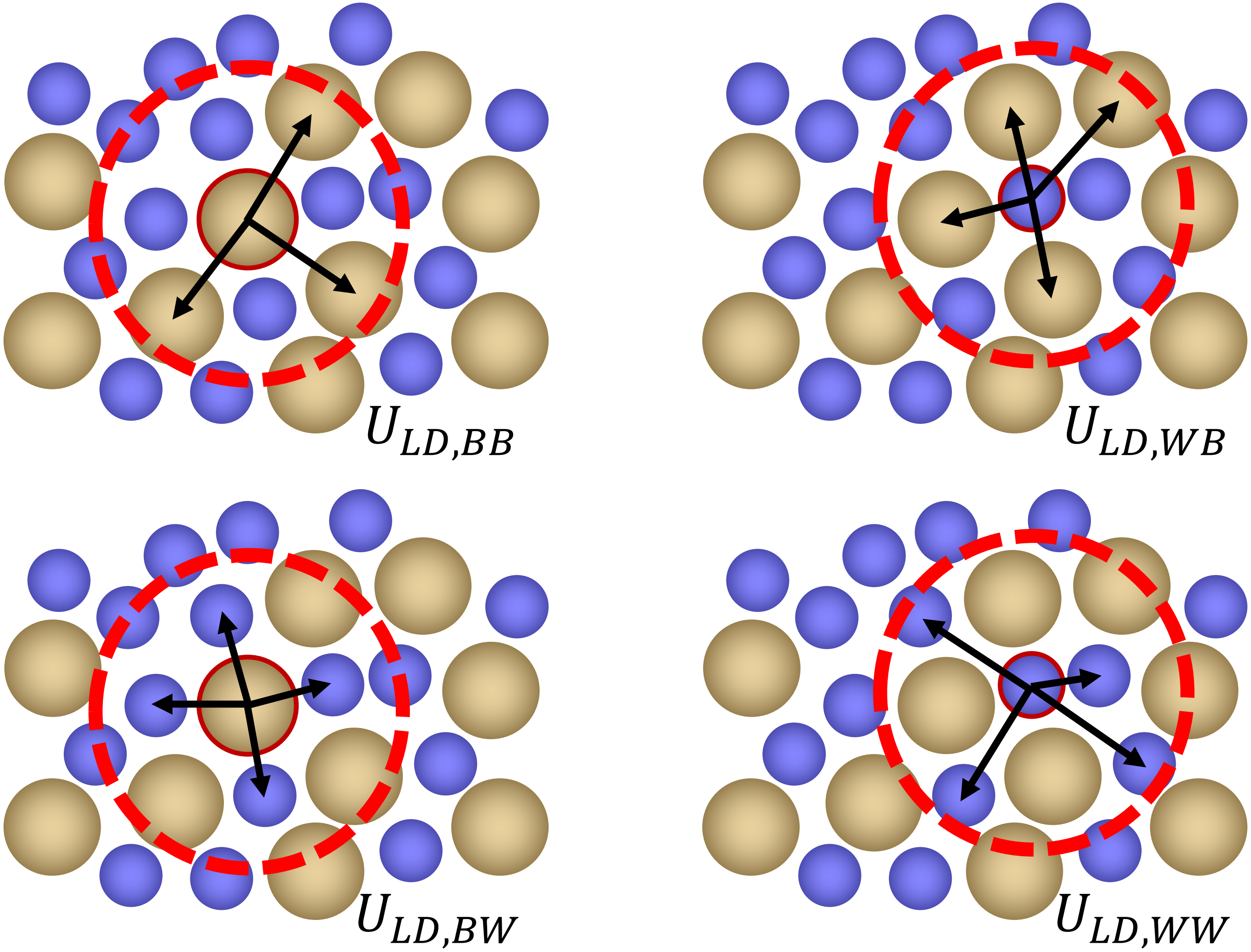
atomistic model – bulk phase separation



CG pair potential model – clustering

Villa, Peter, & van der Vegt, JCTC (2010)

Four different kinds of local density potentials



benzene center

water center

benzene neighbor

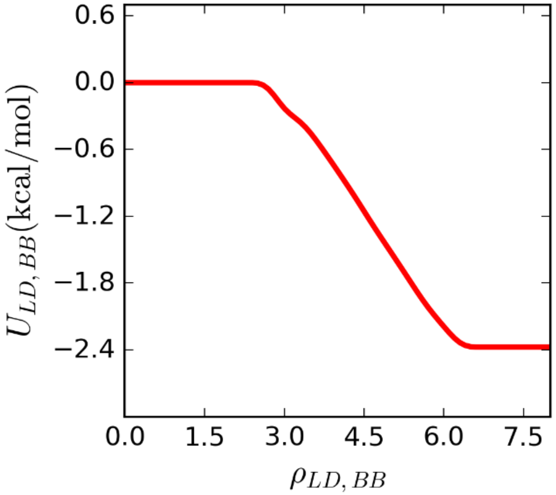
water neighbor

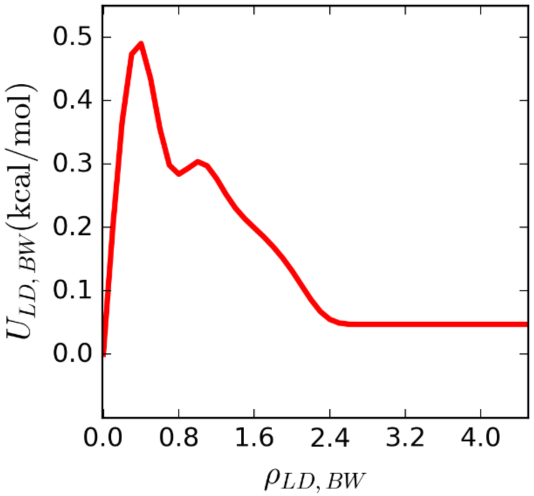
Four different kinds of local density potentials

benzene center

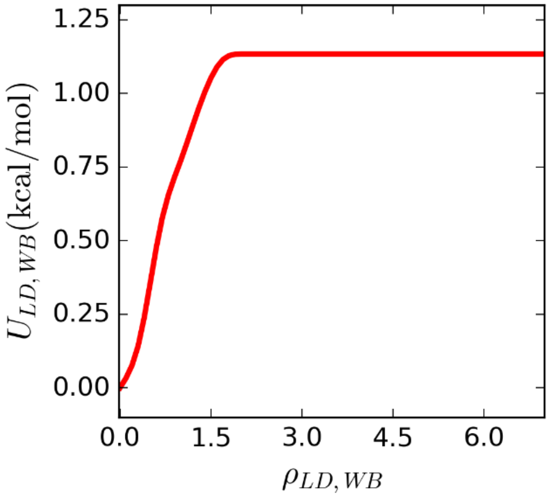
water center

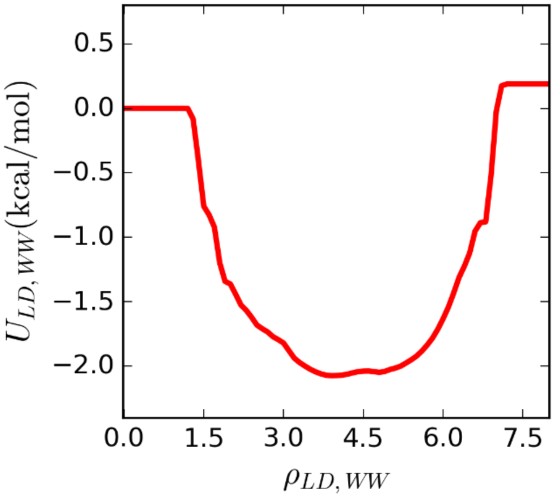
benzene neighbor



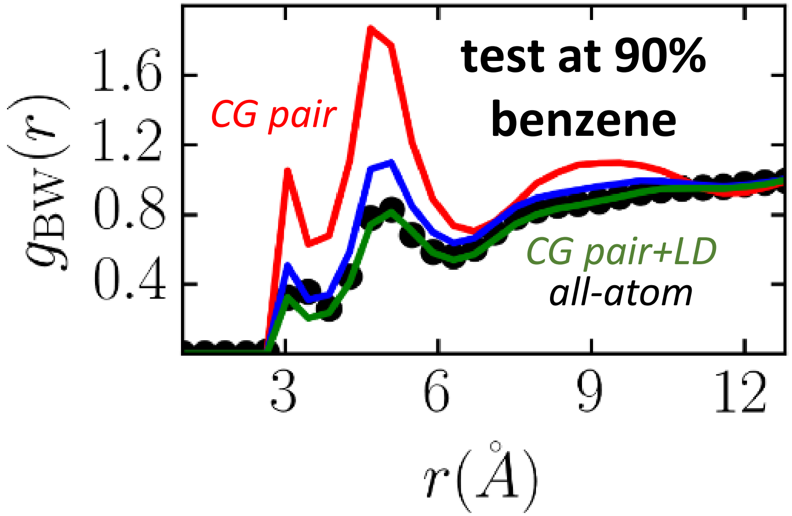


water neighbor



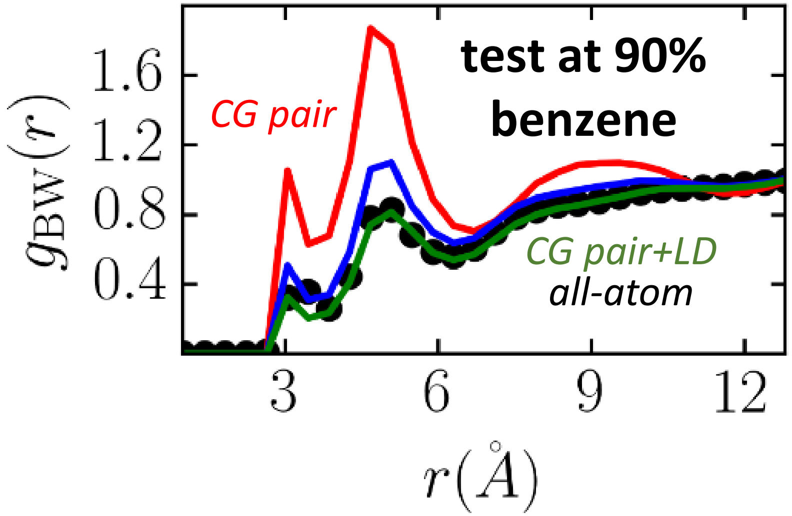


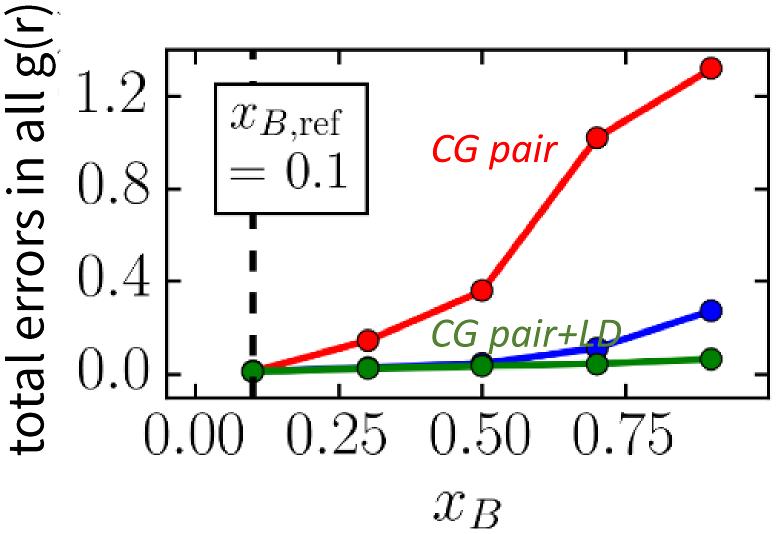
model developed at 10% benzene



Sanyal & Shell, JPCB (2018)

model developed at 10% benzene

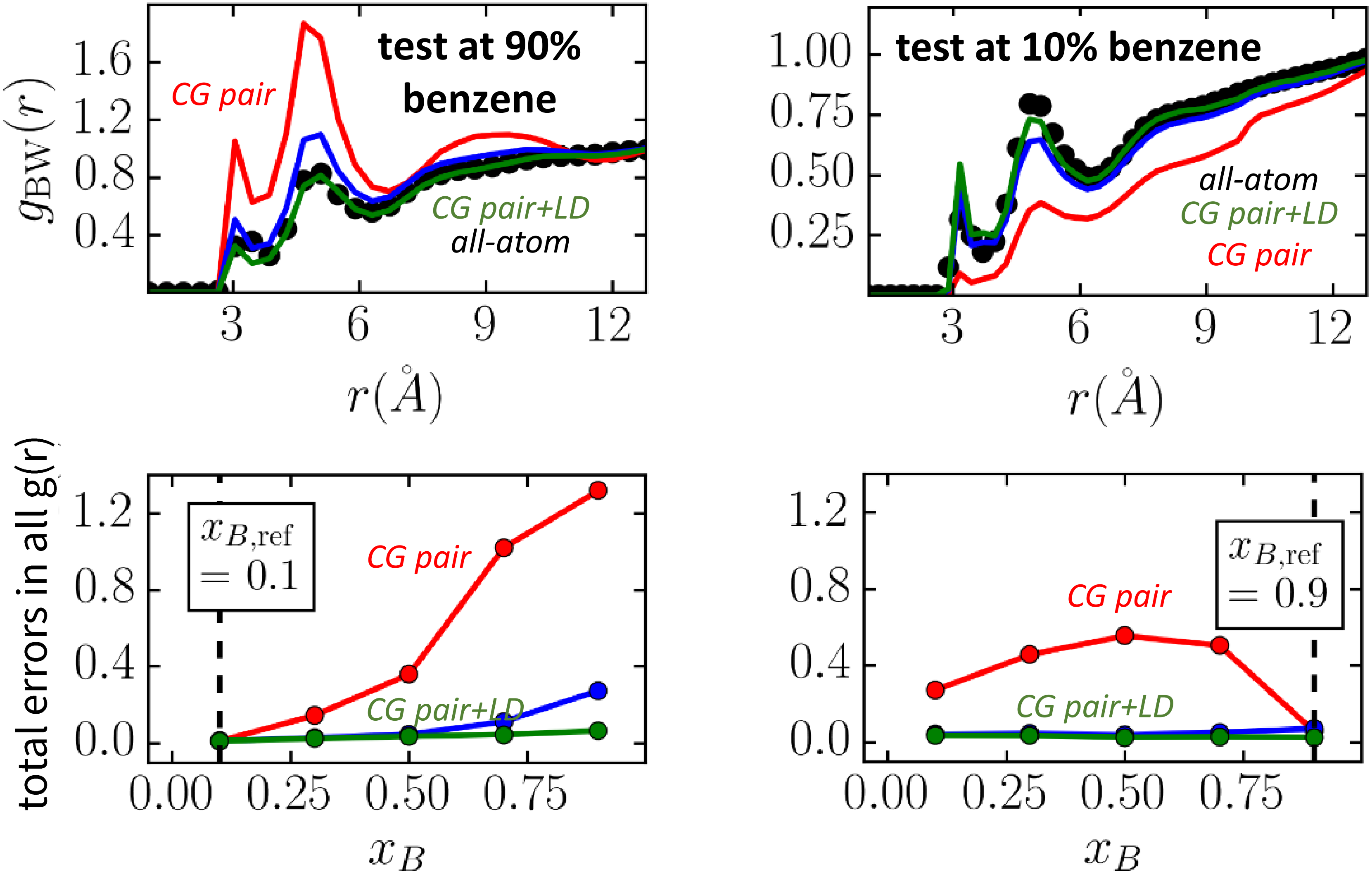




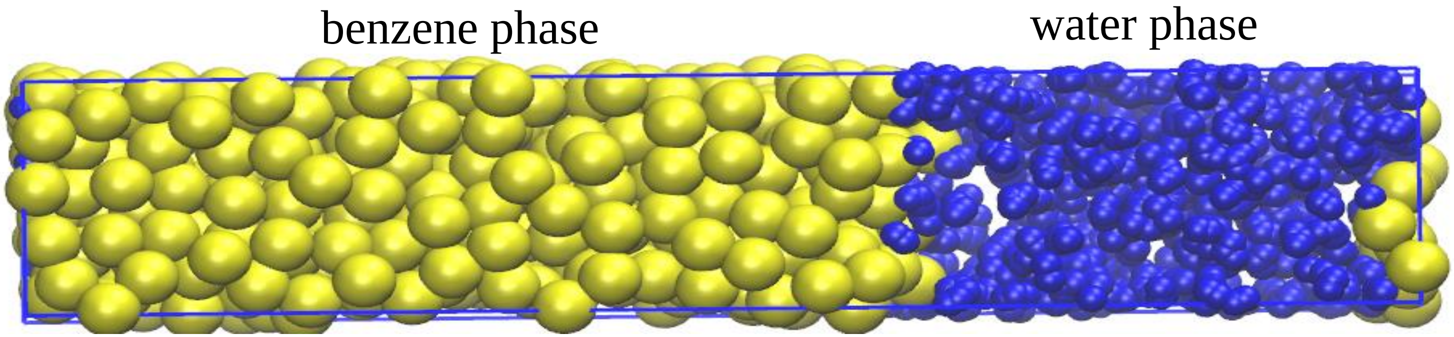
Sanyal & Shell, JPCB (2018)

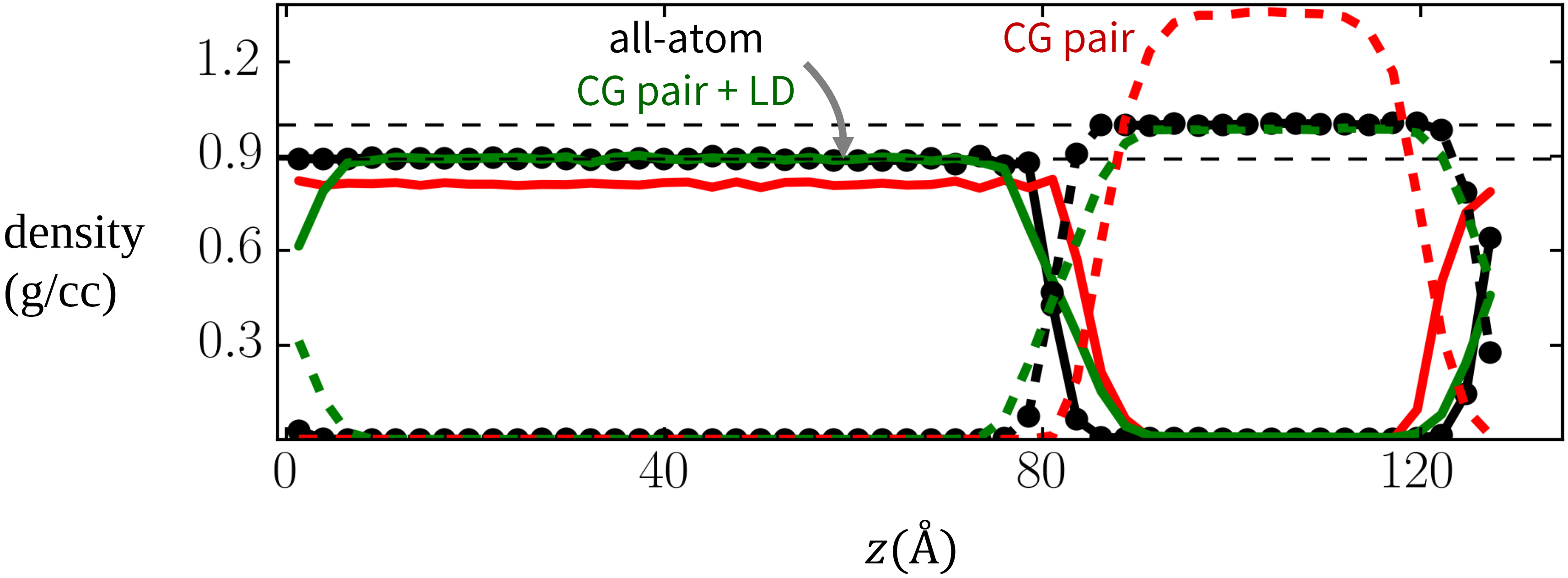
model developed at 10% benzene

model developed at 90% benzene



Sanyal & Shell, JPCB (2018)

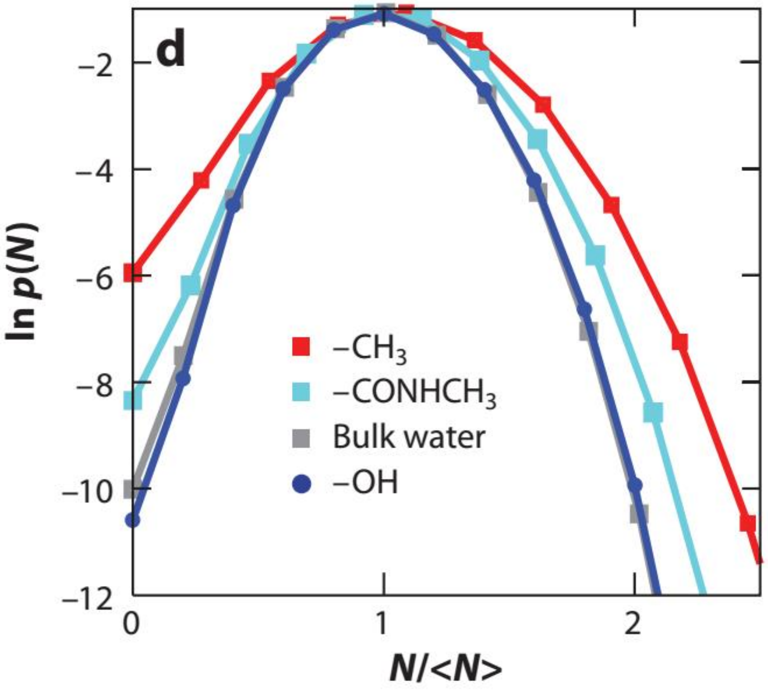


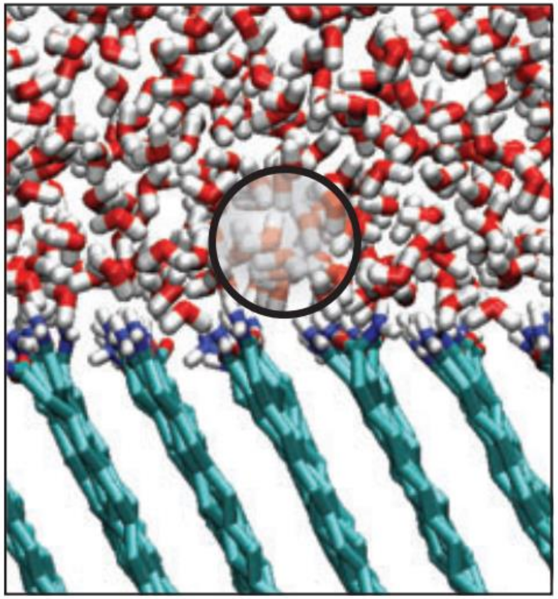


Sanyal & Shell, JPCB (2018)

Why do local density potentials work so well?

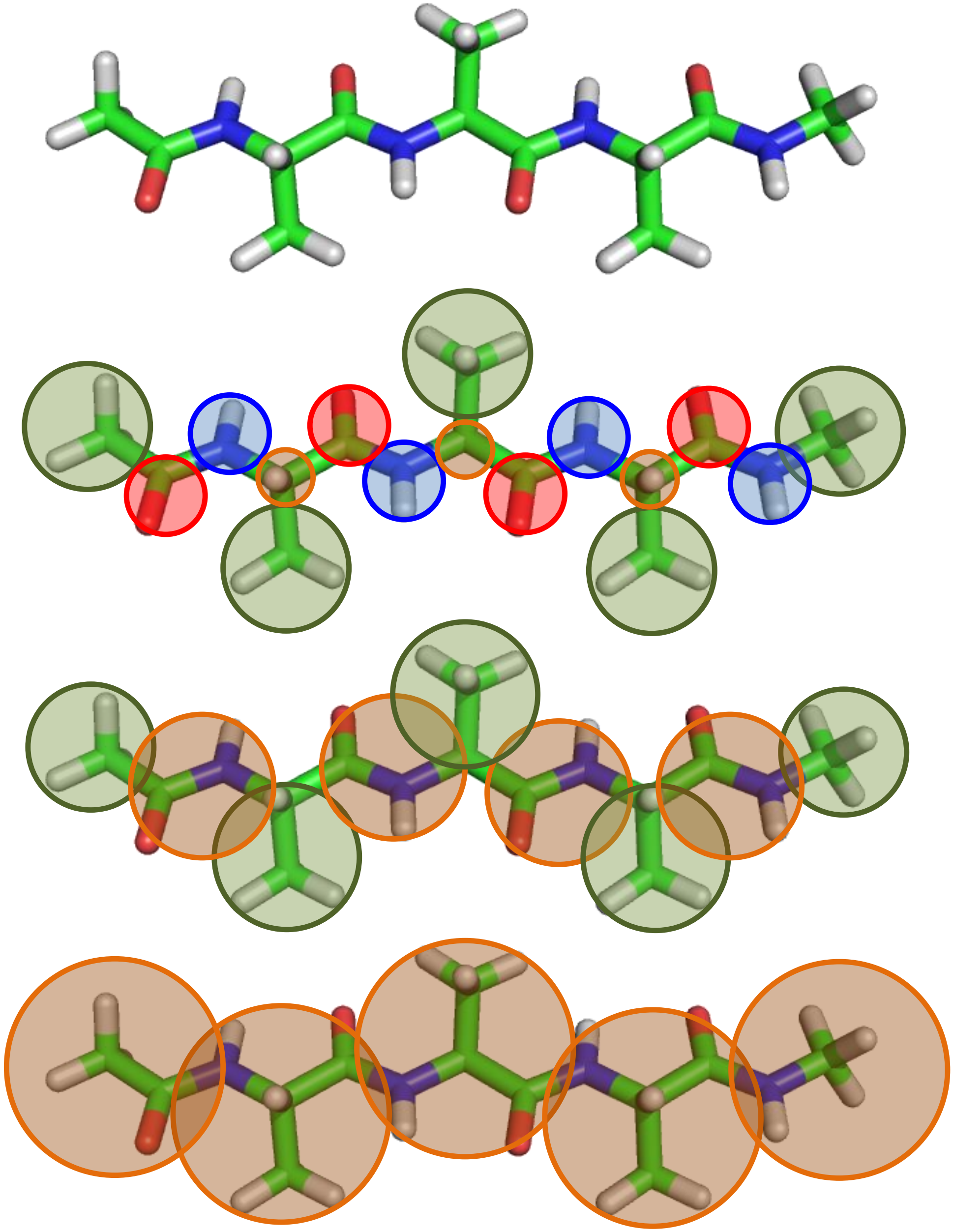
Garde and co-workers, Ann. Rev. Chem. & Biomol. Eng. (2011)





Local density fluctuations drive water's unique solvation properties

What about the design of the CG model?

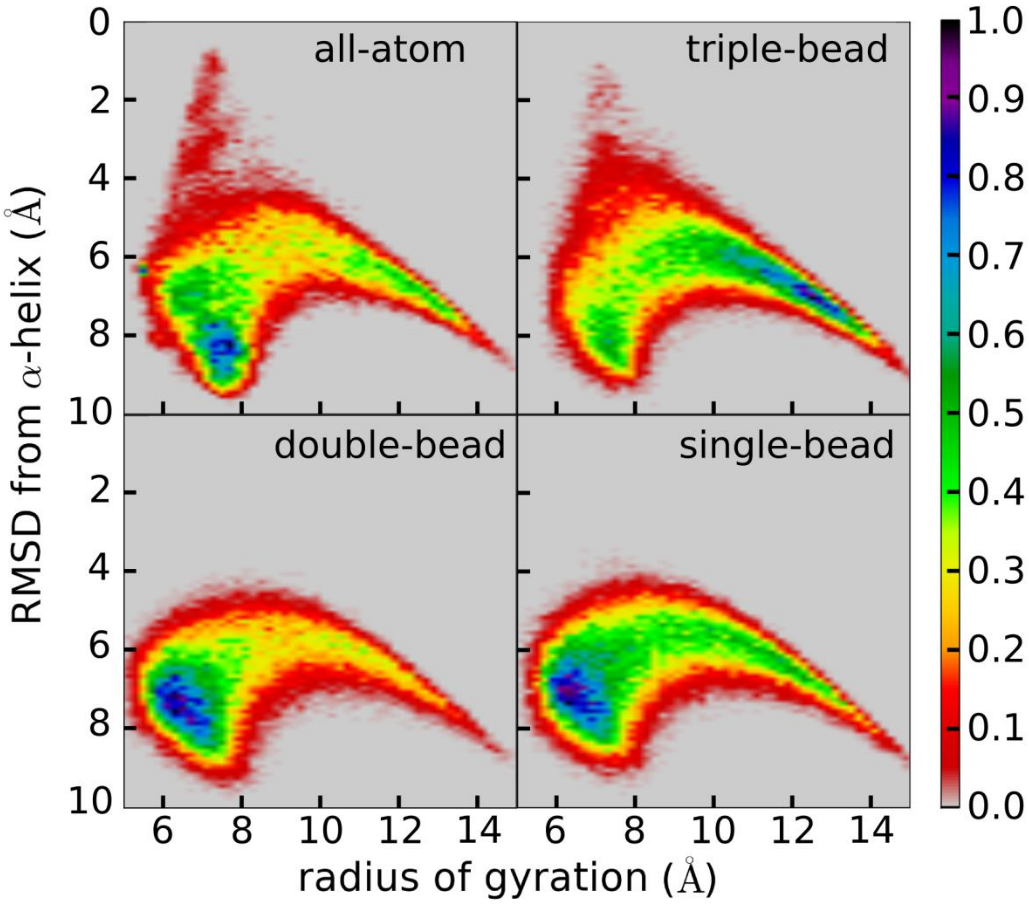


all-atom
system

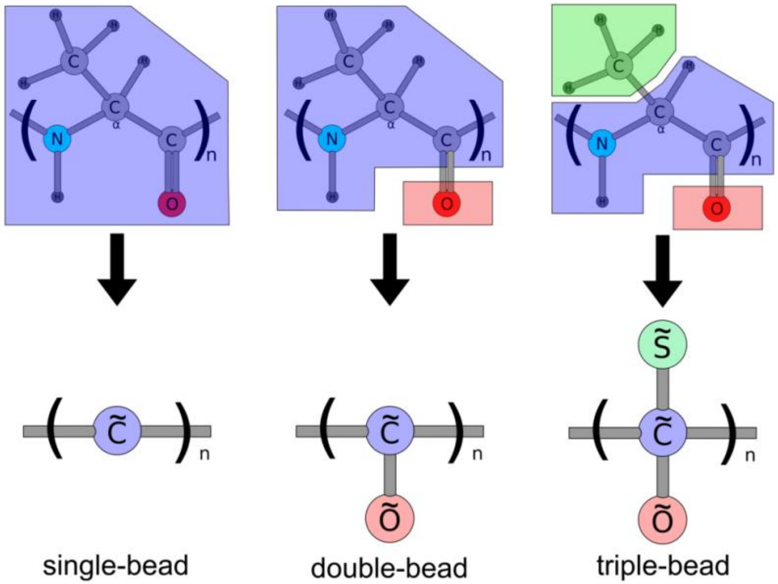
coarse-grained
hydrogens

coarse-grained
functional groups

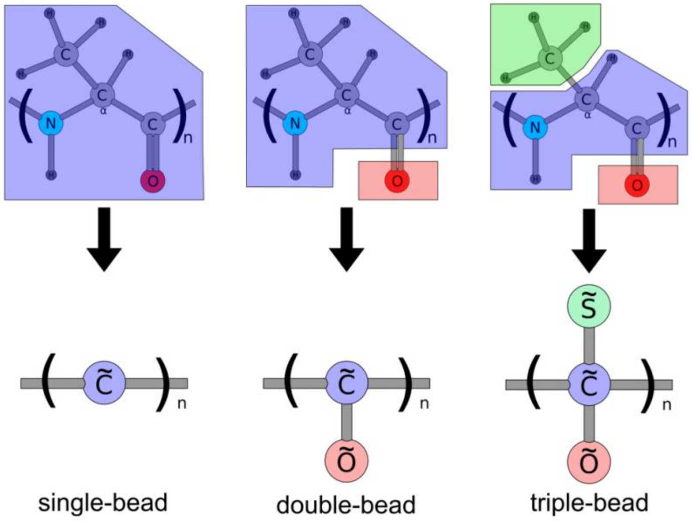
coarse-grained
amino acid residues

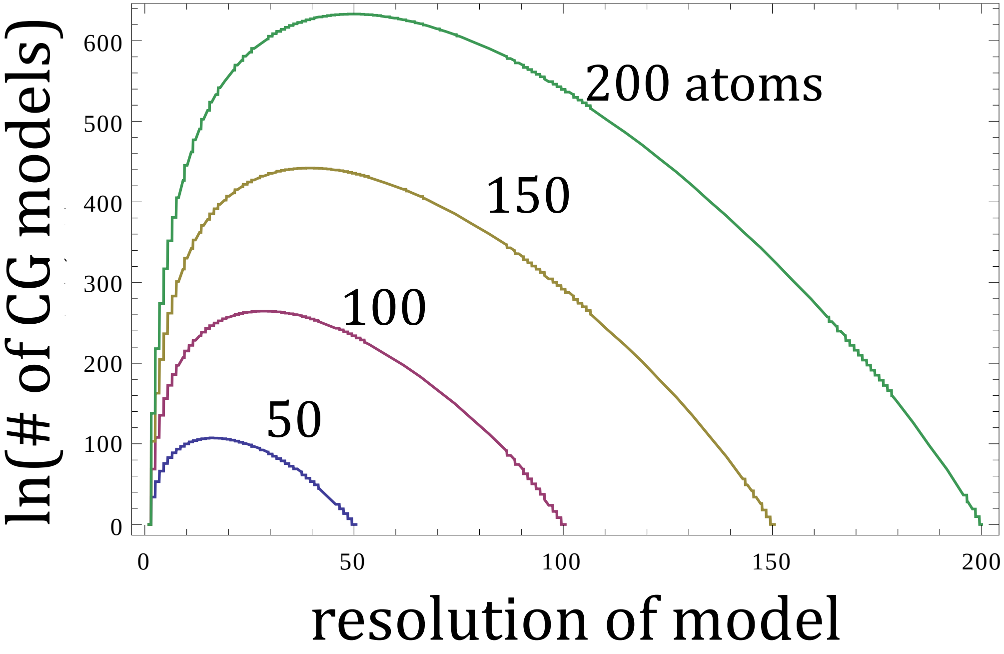


The three bead model does worse than the two-bead one!



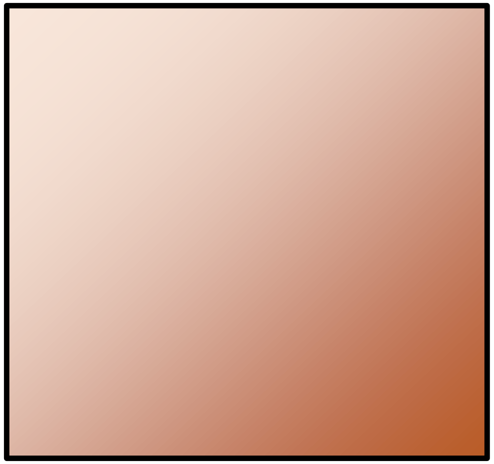
The vast space of all possible CG models



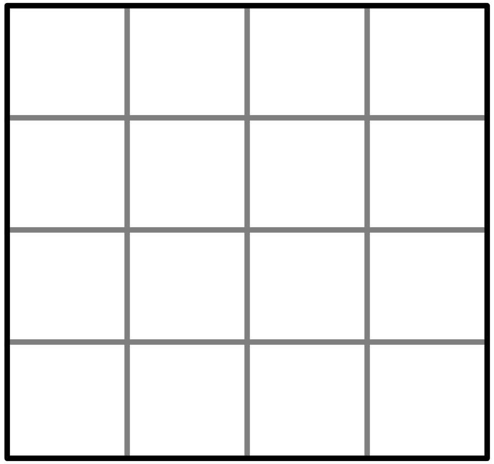


$$S_{\mathrm{map}} = \sum_i p_{AA}(i) \ln \frac{p_{AA}(i)}{\bar{p}_{AA}(i)}$$

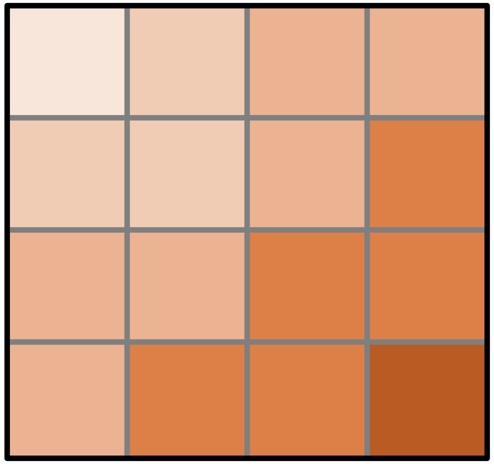
average probability of all AA configurations mapping to the same CG state



p_{AA}



$\mathbf{M}$



$\bar{p}_{AA}$

Rudzinski and Noid, JCP (2011); Foley, Shell, and Noid JCP (2015).

$$S_{\mathrm{map}} = \sum_i p_{AA}(i) \ln \frac{p_{AA}(i)}{\bar{p}_{AA}(i)}$$

$$PMF = \langle U_{AA} \rangle + TS_{\mathrm{map}}$$

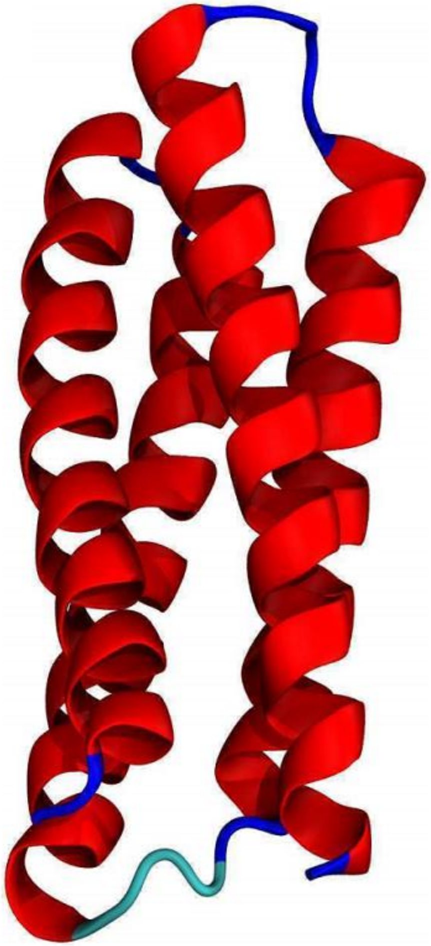
average AA energies
within the CG state

all of the complex stuff!

- multibody interactions
- temperature dependence
- minimize these by min S_{map}

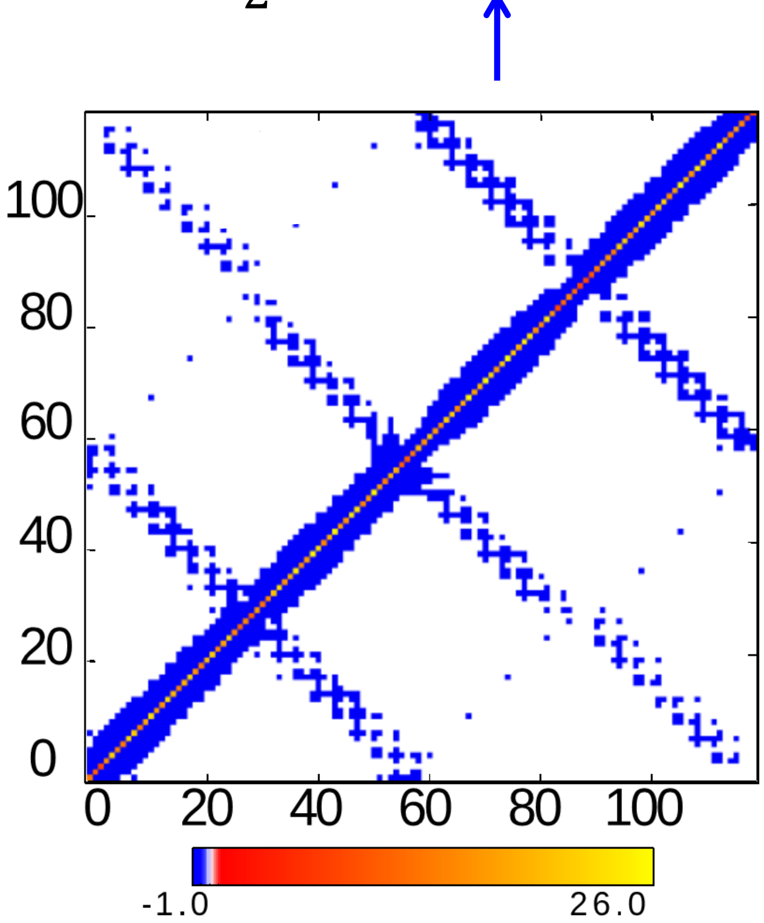
Foley, Shell, Noid, J. Chem. Phys. (2015)

Case study: Gaussian Network Model



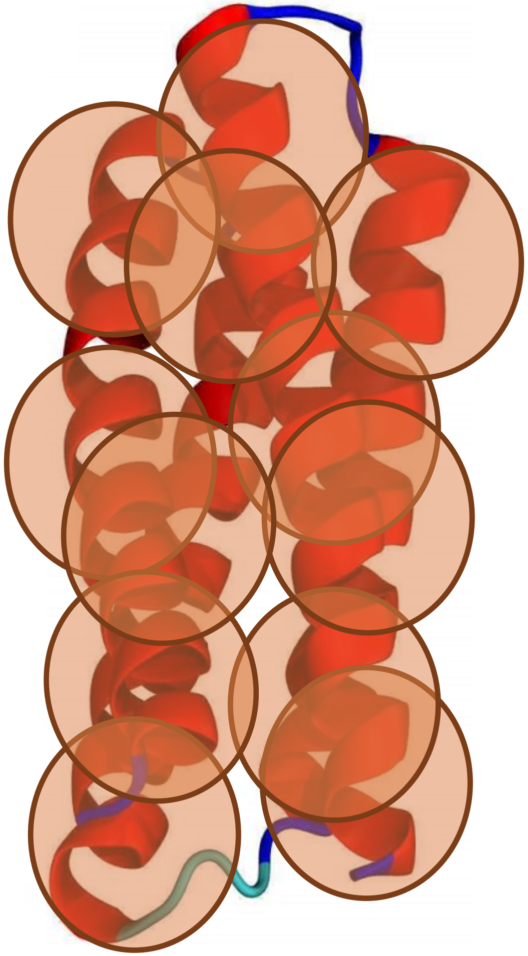
"all-atom"
120 amino acid sites

$$U_{AA} = \frac{1}{2}(\mathbf{r}-\mathbf{r}_0)^t\mathbf{a}(\mathbf{r}-\mathbf{r}_0)$$



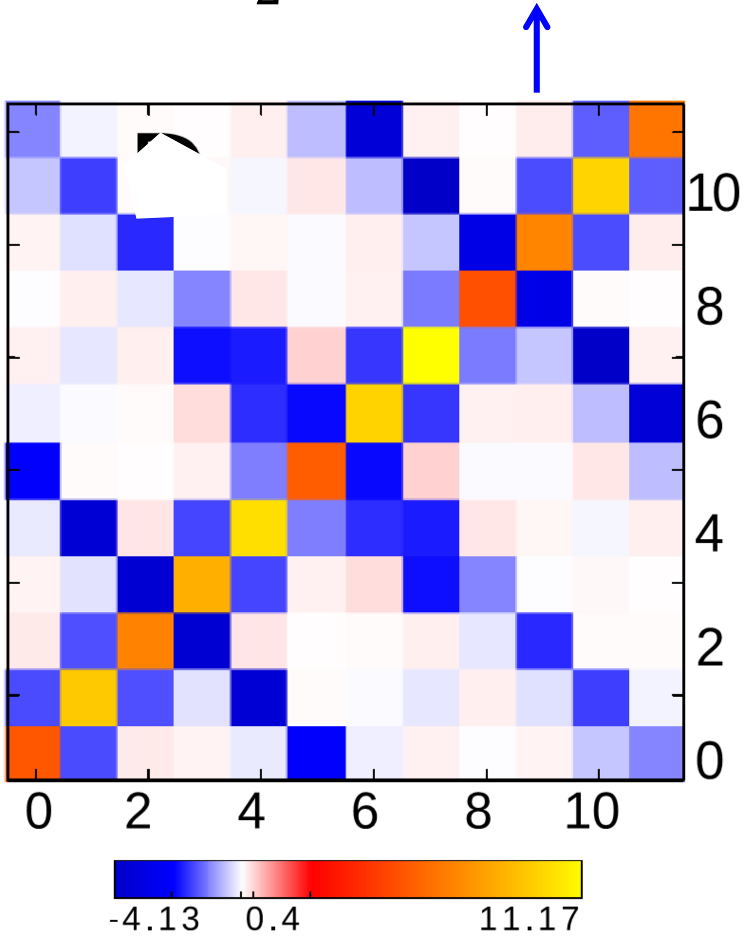
Foley, Shell, Noid, J. Chem. Phys. (2015)

Case study: Gaussian Network Model

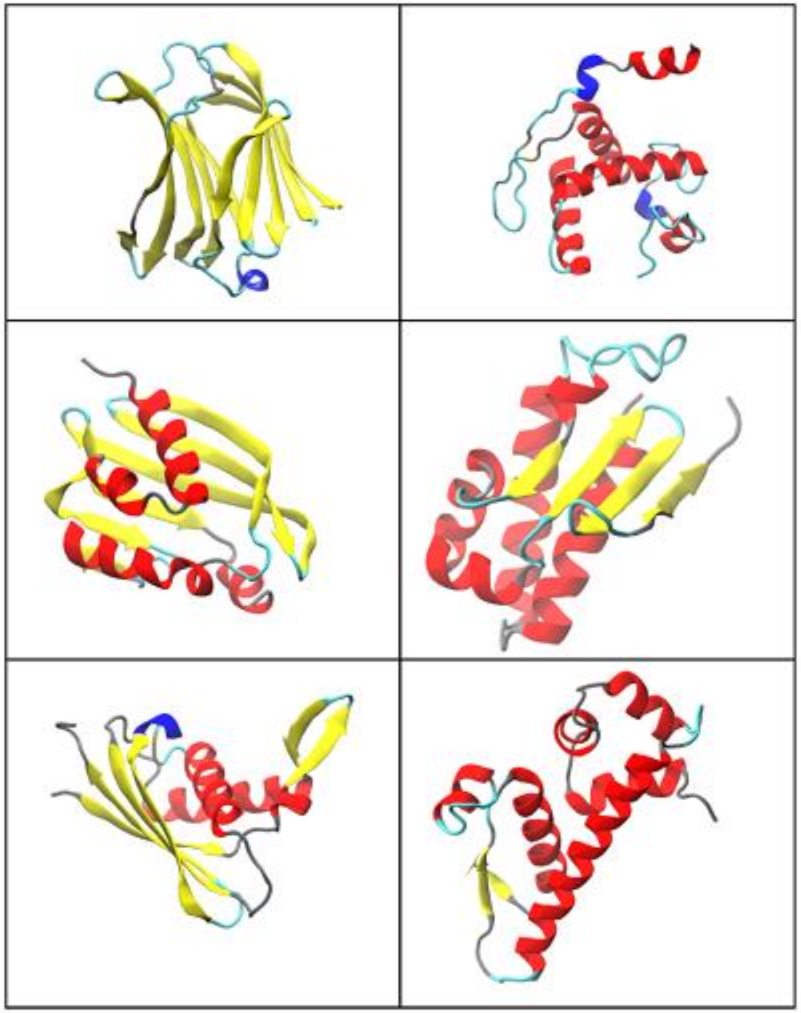


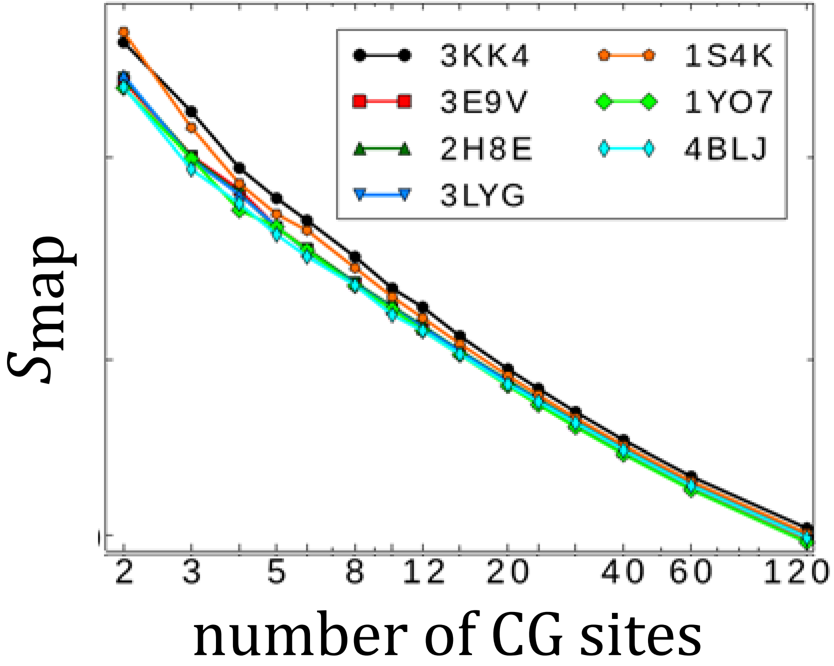
coarse-grained
12 pseudoatom sites

$$U_{CG} = PMF = \frac{1}{2}(\mathbf{R}-\mathbf{R}_0)^t\mathbf{A}(\mathbf{R}-\mathbf{R}_0)$$

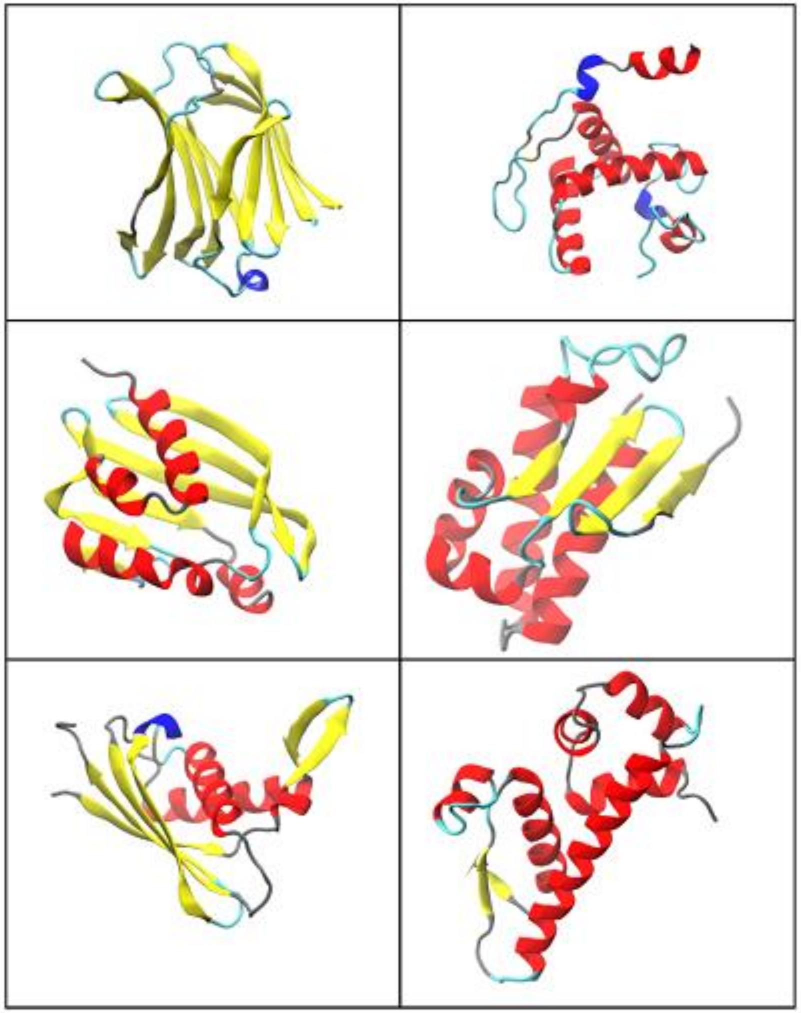


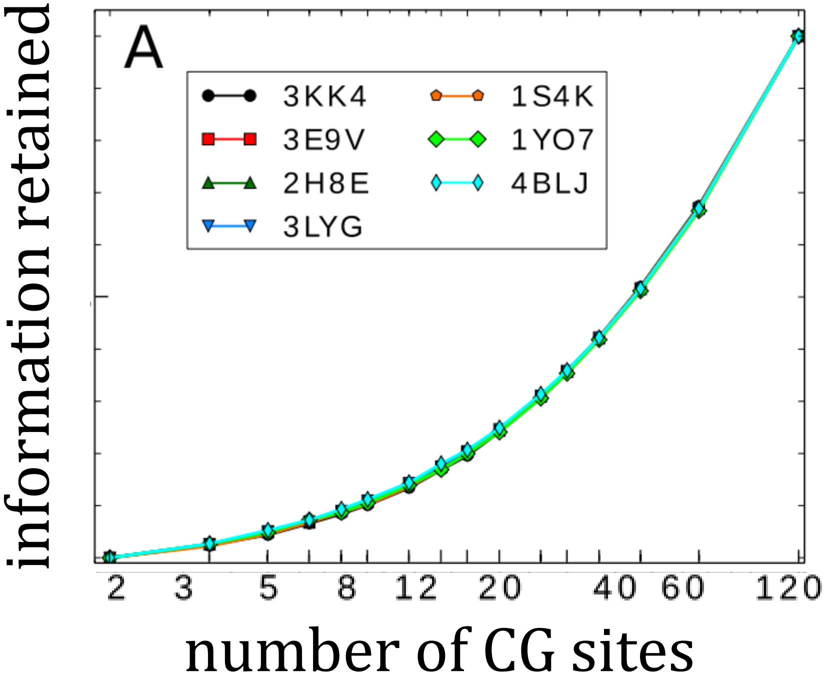
Foley, Shell, Noid, J. Chem. Phys. (2015)





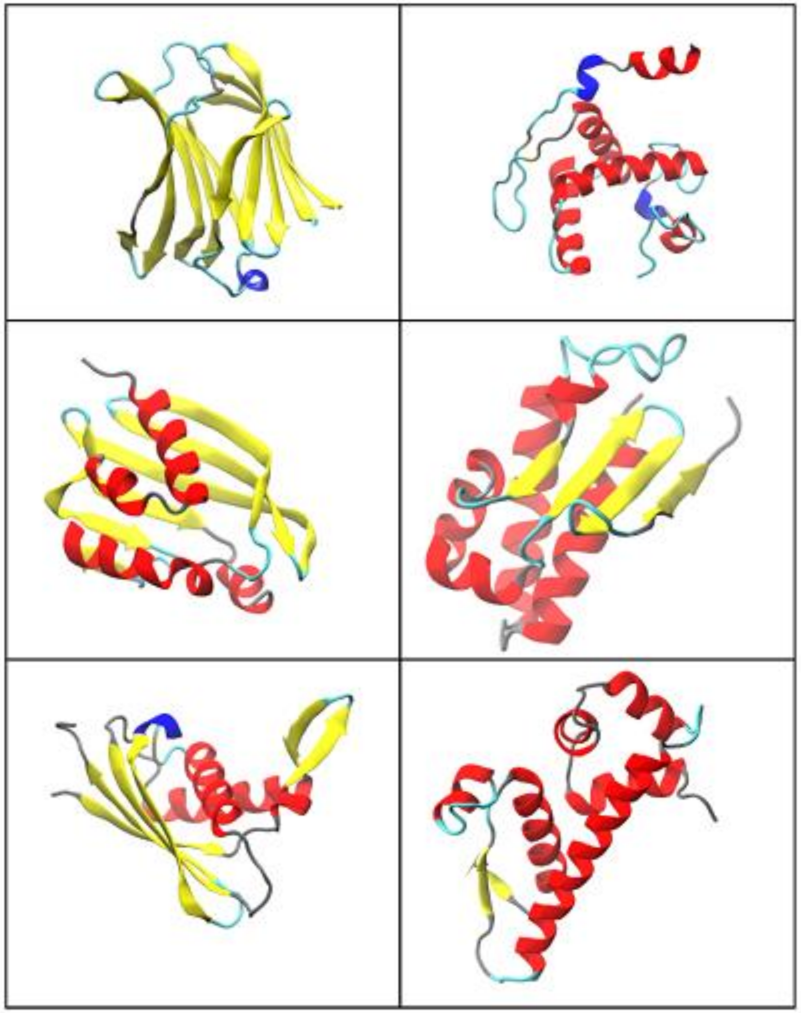
Foley, Shell, **Noid**, J. Chem. Phys. (2015)

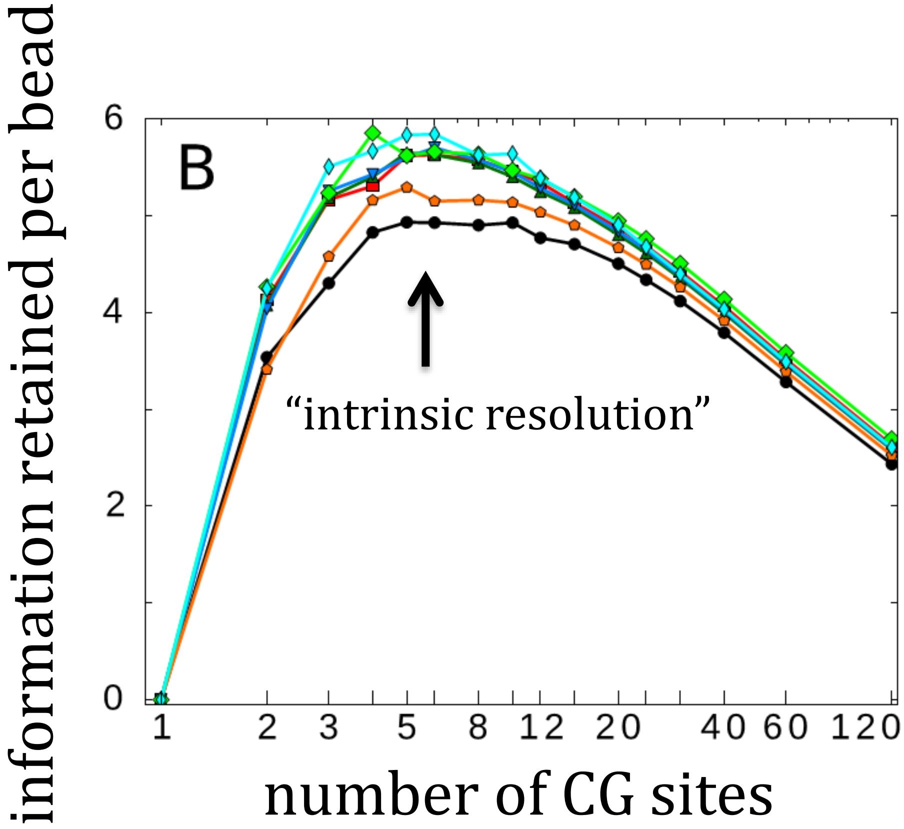




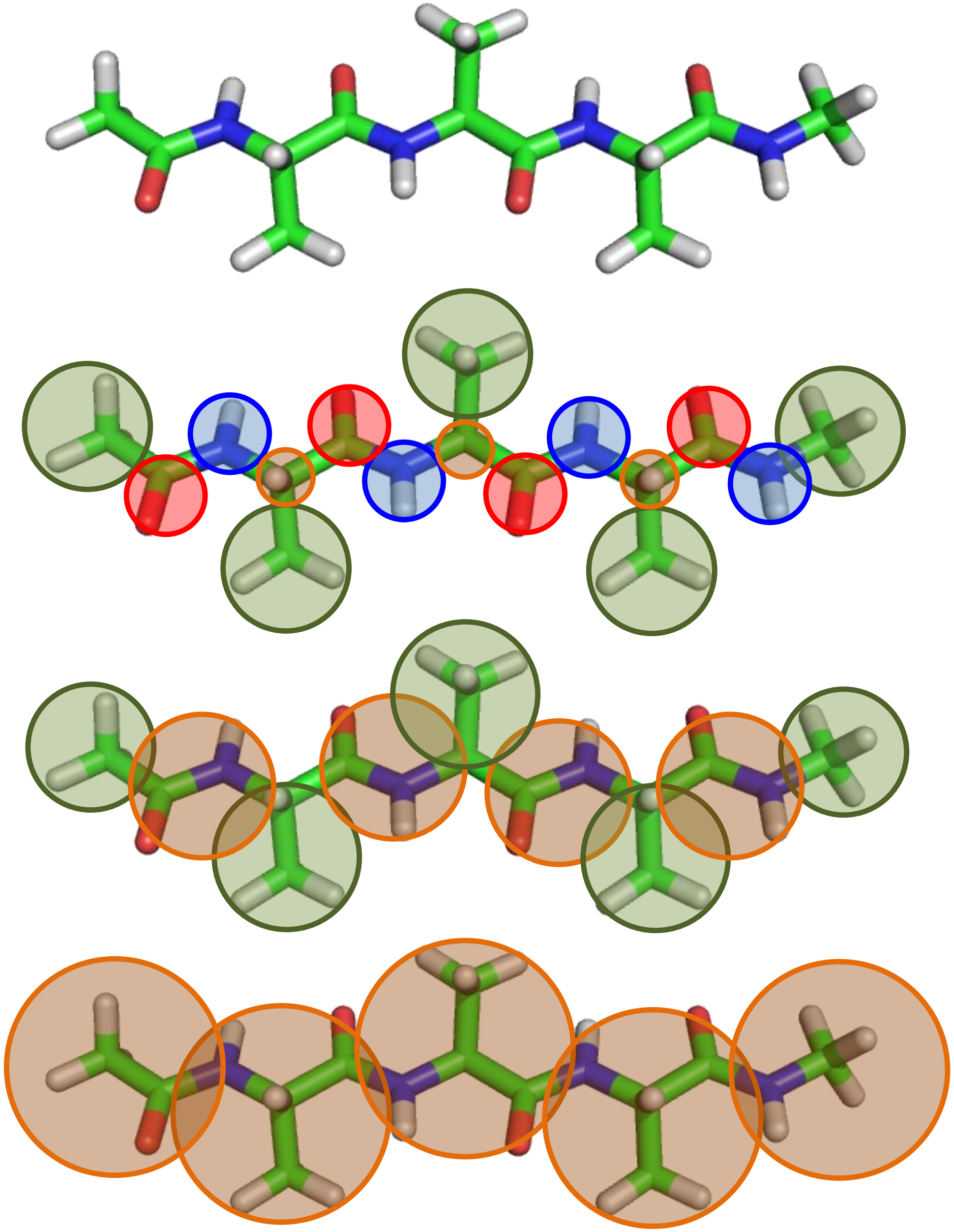
Foley, Shell, **Noid**, J. Chem. Phys. (2015)

Information theoretic signals for *emergent* resolutions





Foley, Shell, **Noid**, J. Chem. Phys. (2015)



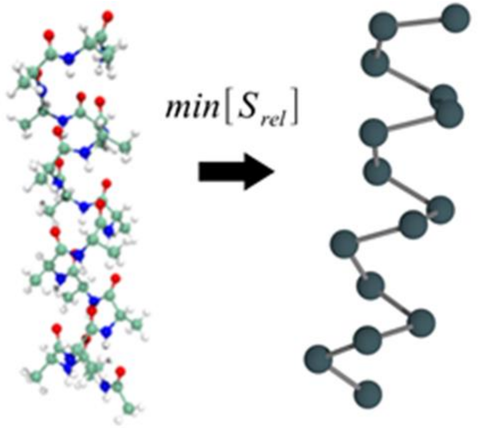
Potential process for a complete CG procedure:

(1) Find an intrinsic resolution by maximizing per-site information.

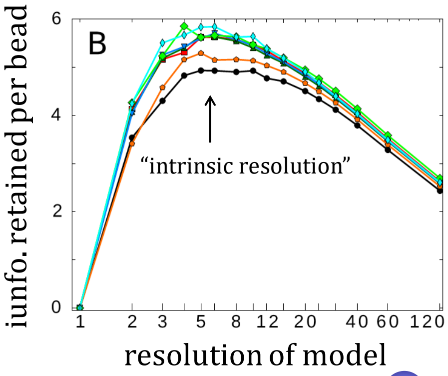
(2) Pick CG degrees of freedom so as to minimize S_{map}.

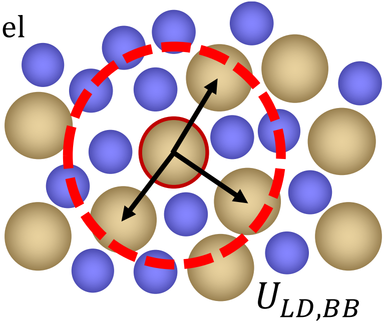
(3) Optimize force field / interactions U_{CG}.

A relative entropy framework for multiscale modeling



$$S_{\mathrm{rel}} = \sum_i p_{AA}(i) \ln\left(\frac{p_{AA}(i)}{p_{CG}(i)}\right)$$





The relative entropy provides both fundamental theory and practical CG’ing strategies for all kinds of models.

Shell, Adv Chem Phys (2016)
review article

New ways to design and represent CG models should broaden their utility in chemical thermodynamics problems.

Foley, Shell, & Noid (2015)
Sanyal & Shell, JCP (2016)
Sanyal & Shell, JPCB (2018)