

## Relative entropy as a universal metric for multiscale errors

Aviel Chaimovich and M. Scott Shell

*Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA*

(Received 28 April 2010; published 29 June 2010)

We show that the relative entropy,  $S_{\text{rel}}$ , suggests a fundamental indicator of the success of multiscale studies, in which coarse-grained (CG) models are linked to first-principles (FP) ones. We demonstrate that  $S_{\text{rel}}$  inherently measures fluctuations in the differences between CG and FP potential energy landscapes, and develop a theory that tightly and generally links it to errors associated with coarse graining. We consider two simple case studies substantiating these results, and suggest that  $S_{\text{rel}}$  has important ramifications for evaluating and designing coarse-grained models.

DOI: [10.1103/PhysRevE.81.060104](https://doi.org/10.1103/PhysRevE.81.060104)

PACS number(s): 05.20.Gg, 05.10.-a, 05.70.-a

The development of multiscale, coarse-grained (CG) physiochemical models is central to many studies of soft matter, driven by the need for models and tractable simulations at supra-atomic scales [1–5]. Such has long been the spirit of simple models of spin systems, fluids, polymers, and biomolecules [6–8]. More recent efforts have spawned sophisticated numerical coarse-graining algorithms for optimizing pseudomolecular models that enable larger scale simulations [9–11].

Despite such numerical advances, missing is a central, fundamental formalism for the *physics* of multiscale problems: how one quantifies the relevance of arbitrary CG models to “first-principles” (FP) interactions and how corresponding CG errors reflect emergent features in FP driving forces. As a basic task, consider a CG molecular model designed to mimic a reference FP one, described by an arbitrary set of fundamental interactions. We propose that errors in CG model’s ability to replicate the equilibrium FP properties are tied to a single statistical-thermodynamic quantity, the relative entropy,

$$S_{\text{rel}} = \sum_{\nu} p_{\text{FP}}(\nu) \ln \left\{ \frac{p_{\text{FP}}(\nu)}{p_{\text{CG}}[M(\nu)]} \right\} \quad (1)$$

Here,  $p$  denotes ensemble probabilities and the summation proceeds over all configurations  $\nu$  of the FP system;  $M$  is a mapping function that gives the CG microstate corresponding to a FP one. We introduced  $S_{\text{rel}}$  for multiscale purposes in a previous paper [12]; via a likelihood approach, we argued that the relative entropy measures information lost upon coarse graining. Based on the working assumption that the relative entropy should be at a minimum for optimal CG models, we developed a coarse-graining methodology that we later used to create spherically symmetric models of water [12,13]. We also found that  $S_{\text{rel}}$ -minimization recovers some important principles in statistical physics (e.g., the variational mean-field principle and the uniqueness theorem) [12], yet the physical significance and generality of this informative property has remained unclear.

Here, we now rigorously show that  $S_{\text{rel}}$  plays a more *physical* role in multiscale studies: it measures the Hamiltonian fluctuations associated with coarse graining, and more practically, it systematically predicts the errors due to coarse graining in many system properties. We show this by devel-

oping a thermodynamiclike framework for multiscale physics centered around  $S_{\text{rel}}$ . To test this theory, we examine two very simple but instructive case studies, dealing with the ideal gas and lattice gas, in which *interactions* rather than degrees of freedom are coarse grained. Both cases confirm that the relative entropy signals errors in the ability of the CG descriptions to capture true FP behavior, providing evidence for its basic role in coarse graining.

*Theory.* We first develop a basic formalism for multiscale thermodynamics in the canonical ensemble using  $S_{\text{rel}}$  [14]. Substituting the canonical configurational probabilities [12],

$$S_{\text{rel}} = \beta \langle U_{\text{CG}} - U_{\text{FP}} \rangle_{\text{FP}} - \beta (A_{\text{CG}} - A_{\text{FP}}) + S_{\text{map}} \quad (2)$$

where  $U$  is the potential energy,  $A$  the free energy, and  $\beta = 1/k_B T$ .  $S_{\text{map}}$  is the log of the number of FP configurations mapping to a single CG configuration, a property rigorously independent of either Hamiltonian and temperature (we further assume it is constant in phase space).

Equation (2) can be reformulated using standard free energy perturbation [15] to admit an instructive (and numerically more convenient) expression based on averages alone,

$$S_{\text{rel}} = -\langle \Delta \rangle_{\text{FP}} + \ln \langle e^{\Delta} \rangle_{\text{FP}} = \ln \langle e^{\Delta - \langle \Delta \rangle_{\text{FP}}} \rangle_{\text{FP}} \quad (3)$$

Here,  $\Delta(\nu) = \beta \{ U_{\text{FP}}(\nu) - U_{\text{CG}}[M(\nu)] \}$  is simply a dimensionless potential energy difference between a given FP configuration and a corresponding CG one. If  $\Delta$  is bounded throughout phase space, we can gain further insight by assuming that  $\Delta$  is Gaussian distributed, yielding

$$S_{\text{rel}} \cong \frac{1}{2} [\langle \Delta^2 \rangle_{\text{FP}} - \langle \Delta \rangle_{\text{FP}}^2]. \quad (4)$$

Thus, an important interpretation emerges:  $S_{\text{rel}}$  measures the fluctuations of the difference between the two potential energy landscapes; as an approximation, it is just half of the variance. For perfect replication the fluctuations vanish, but they intensify as differences in ensemble probabilities grow. The above equation is also retrieved via a cumulant expansion of Eq. (3), rendering the Gaussian approximation valid for low  $S_{\text{rel}}$ . In addition, Eq. (4) suggests that  $S_{\text{rel}}$  is a linear combination of the variances and covariance of the FP and CG potential energies (with  $\frac{1}{2}\beta^2$  as a prefactor). Note that for  $\beta \rightarrow 0$ ,  $S_{\text{rel}}$  becomes inversely proportional with the square of the temperature, indicating that CG models can better repli-

cate FP ones at higher temperatures, irrespective of other state variables.

Bearing in mind that  $S_{\text{rel}}$  fundamentally measures differences in CG and FP energy landscapes, we proceed to show that it is also an indicator of “replication errors” in various system properties. Given an arbitrary state function or average  $W$ , we define its replication error as  $W_{\text{err}} \equiv W_{\text{FP}} - W_{\text{CG}}$ ; in the following, we give specific cases for such errors. Consider a CG system whose potential contains adjustable parameters  $\{\lambda\}$  to be optimized so as to minimize replication errors. Differentiation of Eq. (2) by some  $\lambda$  (appearing in  $U_{\text{CG}}$  and  $A_{\text{CG}}$ ) yields

$$\left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{err}} \equiv \left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{FP}} - \left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{CG}} = \frac{1}{\beta} \frac{\partial S_{\text{rel}}}{\partial \lambda}. \quad (5)$$

This error obviously vanishes for all parameters at the relative entropy minimum,  $S_{\text{rel}}^* = S_{\text{rel}}(\{\lambda = \lambda^*\})$ , which is the case for  $S_{\text{rel}}$ -optimized models. Normally,  $U_{\text{CG}}$  can be expressed as a linear combination of component functions with energetic coefficients; using Eq. (5), the following then holds,

$$\langle U_{\text{CG}} \rangle_{\text{err}}^* \equiv \langle U_{\text{CG}} \rangle_{\text{FP}}^* - \langle U_{\text{CG}} \rangle_{\text{CG}}^* = 0 \quad (6)$$

which is true even without the optimization of any nonlinear parameters. This observation, together with Eq. (2), gives

$$A_{\text{err}}^* - \langle U \rangle_{\text{err}}^* = (S_{\text{rel}}^* - S_{\text{map}})/\beta, \quad (7)$$

where  $A_{\text{err}} \equiv A_{\text{FP}} - A_{\text{CG}}$  and  $\langle U \rangle_{\text{err}} \equiv \langle U_{\text{FP}} \rangle_{\text{FP}} - \langle U_{\text{CG}} \rangle_{\text{CG}}$ . This important result represents the multiscale analogy of the conventional Legendre transform, only instead involving *errors* in thermodynamic potentials. This suggests an elementary multiscale formalism for errors paralleling traditional thermal physics, in which the fundamental equation stems from the relative entropy rather than the thermal entropy. Note that  $S_{\text{map}}$  simply accounts for trivial errors associated with possible removal of degrees of freedom.

Equation (7) provides a basis for relating errors in other properties. For optimal CG models, one can connect the dimensionless derivative  $y \frac{\partial S_{\text{rel}}}{\partial y}$ , for any arbitrary state variable  $y$ , to coarse-graining errors in a corresponding dimensionless response function. For  $y = \beta$ , direct differentiation shows the response function is the dimensionless heat capacity,  $c = C_V/k_B$ . For  $y = \rho$ , where  $\rho$  is density, a Maxwell relation shows the response function is the dimensionless pressure coefficient,  $\gamma = (\partial P / \partial T)_V / k_B \rho$ . In both cases, the errors are tied to the per-particle relative entropy derivatives according to

$$c_{\text{err}}^* = \beta \frac{\partial S_{\text{rel}}^*}{\partial \beta}, \quad \gamma_{\text{err}}^* = \rho \frac{\partial S_{\text{rel}}^*}{\partial \rho}. \quad (8)$$

These connections suggest that  $s_{\text{rel}}$  systematically signals errors in the *fluctuations* of various properties of CG models, as measured by response functions.

We end this discussion by considering the replication error associated with the average dimensionless energy difference,  $\langle \Delta \rangle_{\text{err}} = \langle \Delta \rangle_{\text{FP}} - \langle \Delta \rangle_{\text{CG}}$ .  $\Delta$  measures differences in the FP and CG energy landscapes, and its error quantifies how well the CG ensemble probabilities give an accurate FP sampling of these. Employing reweighting with the Gaussian assumption,

we find  $S_{\text{rel}} \equiv -\frac{1}{2} \langle \Delta \rangle_{\text{err}}$ . For an optimized CG model, Eq. (6) can be substituted to obtain

$$\langle U_{\text{FP}} \rangle_{\text{err}}^* \equiv \langle U_{\text{FP}} \rangle_{\text{FP}} - \langle U_{\text{FP}} \rangle_{\text{CG}}^* \equiv -2S_{\text{rel}}^*/\beta. \quad (9)$$

Thus, while  $S_{\text{rel}}$  optimization guarantees that  $\langle U_{\text{CG}} \rangle$  will be the same in either ensemble, there is a bound for the reproduction of  $\langle U_{\text{FP}} \rangle$ , which can be predicted by the relative entropy. One way to interpret this approximate relation is that the intrinsic deficiency of an optimal CG model stems in its replication error of the average FP potential energy. Overall, the connection between  $S_{\text{rel}}$  and a number of specific replication errors [Eqs. (7)–(9)] might suggest that this single quantity is more generally sufficient to predict many errors from coarse graining. We proceed to address this idea with two basic case studies.

*Ideal gas.* The venerable ideal-gas model provides an approximate picture of real gaseous phenomena. Can the relative entropy measure the validity of this approximation? We let a nonideal gas be a “first-principles” system and an ideal gas be its “coarse-grained” version. There are no interaction parameters for the ideal gas, yet Eq. (7) still holds since  $U_{\text{CG}} = 0$ . In fact,  $S_{\text{rel}}$  here is identical to the excess entropy, which thus emerges as a special case of this multiscale formalism in which all *interactions* are coarse-grained away.

We employ two models of nonideal gases, the square-well (SW) and Redlich-Kwong (RK) [16]. The dimensionless volume of the square-well,  $\zeta$ , is set at 4 throughout, and for the SW model we invoke virial truncation at the second coefficient. The per-molecule SW and RK relative entropies are then found by Eq. (7) to be  $s_{\text{rel}}^{\text{SW}} = \zeta \rho [1 - e^{-\beta} (1 - \beta)] + \rho$  and  $s_{\text{rel}}^{\text{RK}} = \frac{1}{2} \ln(1 + \rho) \beta^{3/2} - \ln(1 - \rho)$ . Interestingly, the two models attain analogous expressions near the ideal limit,  $(\beta, \rho) \rightarrow (0, 0)$ ,

$$s_{\text{rel}}^{\text{SW}} \cong \rho \left( 1 + \frac{1}{2} \zeta \beta^2 \right), \quad s_{\text{rel}}^{\text{RK}} \cong \rho \left( 1 + \frac{1}{2} \beta^{3/2} \right). \quad (10)$$

Apparently, the relative entropy has quasiuniversal scaling as a gas departs ideality. If we compare the temperature scaling exponent in these equations with that of Eq. (4) at high temperature, it is identical at 2 for the SW model and close at 3/2 for the RK model. The result is surprising given that the Gaussian assumption for  $\Delta$ , on which Eq. (4) is based, does not apply to this case study because of singularities in  $\Delta$  where molecules overlap. The inability of the ideal gas to describe these excluded volume interactions also generates a residual  $S_{\text{rel}}$  for finite density even at infinite temperature:  $s_{\text{rel}}(\beta \rightarrow 0) = \rho$  while  $s_{\text{rel}}(\rho \rightarrow 0) = 0$ .

Errors in the response functions for the gases follow Eq. (8) and, in the ideal-gas limits of Eqs. (10), have the form,

$$c_{\text{err}} \propto [s_{\text{rel}} - s_{\text{rel}}(\beta = 0)], \quad \gamma_{\text{err}} \propto [s_{\text{rel}} - s_{\text{rel}}(\rho = 0)]. \quad (11)$$

To compare these ideal limits to the actual errors, we evaluate  $s_{\text{rel}}$ ,  $c_{\text{err}}$ , and  $\gamma_{\text{err}}$  for the SW and RK models on a grid of state points. Figure 1 shows the errors in terms of  $s_{\text{rel}}$ . Surprisingly, the behavior of the errors practically collapses onto the line of the ideal limit of Eq. (11). Moreover,  $\gamma_{\text{err}}^{\text{SW}}$  and  $c_{\text{err}}^{\text{RK}}$  have the (ideal limit) linear scaling rigorously in all of state space. In effect, when coarse-graining away interactions, the

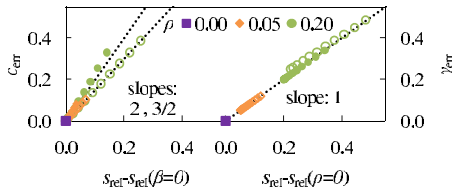


FIG. 1. (Color online) Replication errors of response functions versus  $S_{\text{rel}}$ . (LHS) Heat capacity errors, with  $s_{\text{rel}}$  offset by  $s_{\text{rel}}(\beta=0)=\rho$ . (RHS) Pressure coefficient errors, with  $s_{\text{rel}}$  offset by  $s_{\text{rel}}(\rho=0)=0$ . The square-well fluid (filled symbols) is examined on  $\beta=[0.0,0.5]$ , while the Redlich-Kwong fluid (hollow symbols) is on  $\beta=[0.0,2.0]$ . The dotted lines show the correspondingly linear relationships of Eq. (11).

relative entropy linearly signals errors in response functions and hence system fluctuations. Importantly, except for the slight discrepancy in the slope of the heat capacity line, the trends between the two models are almost identical, which might suggest broader generality in these findings.

*Lattice gas.* The mean-field treatment of the 2D lattice gas offers a useful scenario, as the traditional variational mean-field (VMF) solution can be evaluated against a  $S_{\text{rel}}$ -optimized mean-field (SMF) model. Importantly, the mean-field description breaks down near the critical point, and  $S_{\text{rel}}$ 's ability to signal this failure provides an elementary test of it.

We consider  $M=16^2$  lattice sites, by which the analysis is normalized. With  $n_i$  molecules at lattice site  $i$ , the ‘‘first-principles’’ lattice gas interacts through nearest-neighbors,  $U_{\text{FP}}(\nu)=-\sum_i n_i n_{i+1}$ , and the ‘‘coarse-grained’’ one interacts by a mean-field,  $U_{\text{CG}}(\nu)=-\lambda \sum_i n_i$ . The value for the dimensionless mean-field parameter,  $\lambda$ , is the sole difference between the VMF and SMF models. The lattice gas is coupled to a molecule-bath through the chemical potential  $\mu$ , and a grand potential is developed for the canonical framework presented above. Note that all properties of interest (including  $S_{\text{rel}}$ ) behave symmetrically about the critical chemical potential,  $\mu_c=-2$ . The critical temperature is at  $\beta_c \cong 1.6$ .

For the lattice gas,  $S_{\text{rel}}$  minimization [using Eq. (6)] provides the following optimality condition for the SMF parameter:

$$\lambda^* = -\beta^{-1} \ln(\langle n \rangle_{\text{FP}}^{-1} - 1) - \mu, \quad (12)$$

which requires the average occupancy in the FP ensemble. Conversely, the VMF parameter is determined by the familiar transcendental equation, which entails the average occupancy in the CG (mean-field) ensemble [7]. For a given value of  $\lambda$ , determined by either approach, we compute the relative entropy by Eq. (3) using a Monte Carlo simulation of the FP system. This calculation also evaluates various errors in replication by comparing numerical FP averages with analytical CG ones. We analyze a grid of state points, separately performing ten replicates of this scheme for each. This guarantees statistical errors in the relative entropy less than 2%.

Figure 2 shows the relative entropy in state space for both CG models, which have comparable behavior. At infinite temperature, these curves collapse onto a distinct  $\beta$  power law, having the expected exponent of 2 (within 1% error).

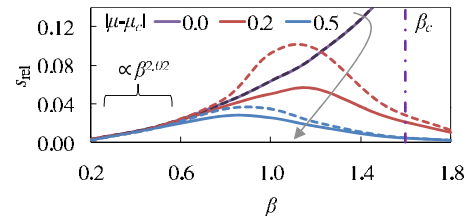


FIG. 2. (Color online)  $S_{\text{rel}}$  as a function of temperature for the lattice gas. The  $S_{\text{rel}}$ -optimized mean-field (solid lines) and the conventional variational mean-field (dotted) models are shown; the two coincide at  $\mu_c$ . We fit a power law to the data of  $\mu_c$  to estimate the behavior at  $\beta \rightarrow 0$ . Arrow indicates increasing  $|\mu - \mu_c|$ .

For the critical  $\mu$ -isoline, this unique (sharply increasing) trend continues toward the critical point; for other  $\mu$ , a maximum is observed just below  $\beta_c$ . Thus, the growth of the relative entropy in the vicinity of the critical point naturally signals the inadequacy of the mean-field description there. Interestingly, both mean-field models attain the same value for  $\lambda$  at  $\mu_c$ , yielding equivalent relative entropies for this isoline. For other  $\mu$ -isolines, the SMF relative entropy is always lower than its VMF counterpart, as expected by  $S_{\text{rel}}$  minimization.

While the better performance of the SMF over the VMF model is not surprising considering that the former uses the exact solution, the important point is that  $S_{\text{rel}}$  can be used to predict errors in either model. By our hypothesis that errors are tied to  $S_{\text{rel}}$ , the SMF system should perform better than the VMF one in replicating true features of the lattice gas. By Eq. (5), the SMF model attains perfect reproduction of the average occupancy,  $\langle n \rangle$ , while the VMF model does not. In general, the effect of perturbing a model parameter about its relative entropy-optimal value can be captured by a second-order expansion of  $S_{\text{rel}}$  with Eq. (5). Thus, the nonzero occupancy error in the VMF model can be captured to second order by,

$$|\langle n \rangle_{\text{err}}| \cong \sqrt{2\beta^{-2} \frac{\partial^2 S_{\text{rel}}^*}{\partial \lambda^2} (s_{\text{rel}} - s_{\text{rel}}^*)}. \quad (13)$$

For this case study,  $\beta^{-2} \frac{\partial^2 S_{\text{rel}}^*}{\partial \lambda^2} = M(\langle n^2 \rangle_{\text{CG}} - \langle n \rangle_{\text{CG}}^2)$ . Figure 3 shows the simulation-determined error in the average occupancy for the conventional mean-field model in terms of the relative entropy, agreeing quite well with Eq. (13)'s square root trend. Regardless, the SMF system consistently exhibits

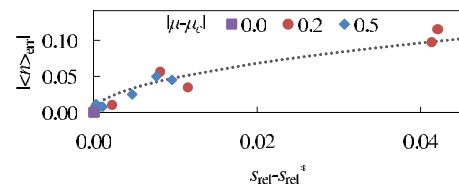


FIG. 3. (Color online) Replication error for the average occupancy versus  $S_{\text{rel}}$ , for the VMF model. For each point,  $s_{\text{rel}}$  is offset by its minimum value at the same state point. The dotted line shows the rough scaling of Eq. (13), in which we approximate  $\langle n^2 \rangle_{\text{CG}} - \langle n \rangle_{\text{CG}}^2$  by its average value over all state points, 0.3393.

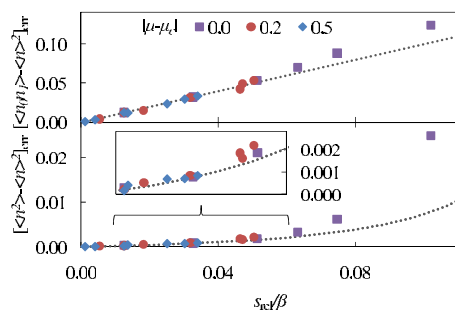


FIG. 4. (Color online) Replication errors for local and global correlations versus the temperature-scaled relative entropy, for the SMF model. Nearest-neighbor (top) and bulk occupancy (bottom) correlations are shown. The dotted lines show Eqs. (14) and (15), respectively; in the latter, we approximate  $\langle n^2 \rangle_{\text{FP}} - \langle n \rangle_{\text{FP}}^2$  by its average over all state points, 0.1861. The inset zooms-in on the lower region as indicated.

better replication than the VMF system for other properties as well (e.g.,  $\langle n^2 \rangle$ ); we restrict the remainder of our discussion to the former model, meaning that  $s_{\text{rel}} = s_{\text{rel}}^*$ .

We now show that errors in the SMF model are tightly linked with the value of the relative entropy. This model can perfectly replicate  $\langle U_{\text{CG}} \rangle$  by Eq. (6), but the inadequate replication of  $\langle U_{\text{FP}} \rangle$  is guaranteed through Eq. (9), which gives

$$[\langle n_{\bar{0}} n_{\bar{1}} \rangle - \langle n \rangle^2]_{\text{err}} \cong (s_{\text{rel}}/\beta). \quad (14)$$

This shows that the error of replication in the correlation between nearest-neighbors is largely given by the value of  $s_{\text{rel}}/\beta$ . Figure 4 compares this approximate scaling law with numerical results. Remarkably, the linear law describes the data superbly, despite the fact that the Gaussian approximation invoked for Eq. (14) overestimates the true  $s_{\text{rel}}$  value by a factor as high as 2 for several state points. Also, it is clear that the error, just as the relative entropy, is pronounced toward the critical point. Equation (14) suggests that the rela-

tive entropy increases with a growing correlation length that renders the failure of the mean-field approach in describing critical phenomena. This observation strengthens our hypothesis that  $S_{\text{rel}}$  measures multiscale model quality; in this case, it signals the critical point through the correlation function.

By measuring neighbor correlations, the relative entropy also signals errors in bulk molecule fluctuations. Assuming that the correlation function decays exponentially with distance between neighbors while expanding  $\langle n^2 \rangle$  (see SI), we attain the following expression using Eq. (14),

$$[\langle n^2 \rangle - \langle n \rangle^2]_{\text{err}} \cong \frac{4}{M} (s_{\text{rel}}/\beta) \left[ 1 - \frac{(s_{\text{rel}}/\beta)}{\langle n \rangle_{\text{FP}} - \langle n \rangle_{\text{FP}}^2} \right]^{-2}. \quad (15)$$

Thus, bulk correlation is also tied to  $s_{\text{rel}}/\beta$ . Figure 4 shows the actual computational results for this error. Despite the crude approximation for the correlation distance decay, Eq. (15) captures the trend in the simulation results, where the error, just as the relative entropy, is pronounced toward the critical point. Thus, errors in bulk fluctuations are also signaled by  $S_{\text{rel}}$ , in turn indicating proximity to the critical point.

In summary, for the first time, we rigorously show the relative entropy is of fundamental relevance to multiscale studies: it is tightly coupled to errors incurred upon coarse graining. The origin of this connection likely stems in the potential energy landscape fluctuations that  $S_{\text{rel}}$  measures, propagating into replication errors in various other properties. This suggests that the relative entropy should be a powerful guide in multiscale studies, as a single measurement able to predict those conditions under which simple analytic and simulation models are accurate or break down.

The authors gratefully acknowledge B. Knott and N. Duff for helpful comments, and the support of the Camille & Henry Dreyfus Foundation and the National Science Foundation (Award No. CBET-0845074).

- 
- [1] J. J. de Pablo, *AIChE J.* **51**, 2371 (2005).  
 [2] G. S. Ayton, W. G. Noid, and G. A. Voth, *Curr. Opin. Struct. Biol.* **17**, 192 (2007).  
 [3] M. Praprotnik, L. D. Site, and K. Kremer, *Annu. Rev. Phys. Chem.* **59**, 545 (2008).  
 [4] P. Sherwood, B. R. Brooks, and M. S. P. Sansom, *Curr. Opin. Struct. Biol.* **18**, 630 (2008).  
 [5] *Phys. Chem. Chem. Phys.* **11** (2009), Themed issue on Coarse-Grained Modeling of Soft Condensed Matter, edited by R. Faller.  
 [6] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, Mass., 1960).  
 [7] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).  
 [8] K. A. Dill *et al.*, *Protein Sci.* **4**, 561 (1995).  
 [9] A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* **52**, 3730 (1995).  
 [10] S. Izvekov and G. A. Voth, *J. Phys. Chem. B* **109**, 2469 (2005).  
 [11] S. Izvekov and G. A. Voth, *J. Chem. Phys.* **123**, 134105 (2005).  
 [12] M. S. Shell, *J. Chem. Phys.* **129**, 144108 (2008).  
 [13] A. Chaimovich and M. S. Shell, *Phys. Chem. Chem. Phys.* **11**, 1901 (2009).  
 [14] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevE.81.060104> for detailed derivations for Eqs. (2)–(15).  
 [15] R. W. Zwanzig, *J. Chem. Phys.* **22**, 1420 (1954).  
 [16] O. Redlich and J. N. S. Kwong, *Chem. Rev.* **44**, 233 (1949).