THE JOURNAL OF CHEMICAL PHYSICS **131**, 024103 (2009)

The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures

Vinod Krishna,[a)] Will G. Noid,[a),b)] and Gregory A. Voth[c)]
Department of Chemistry and Center for Biophysical Modeling and Simulation, University of Utah, Salt Lake City, Utah 84112-0850, USA

(Received 29 March 2009; accepted 12 June 2009; published online 8 July 2009)

This work develops a method for the construction of multiscale coarse-grained (MS-CG) force fields at different temperatures based on available atomistic data at a given reference temperature. The validity of this theory is demonstrated numerically by applying it to construct MS-CG models of the Lennard-Jones liquid and simple point charge water model systems. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3167797]

I. INTRODUCTION

Molecular dynamics (MD)[1,2] is a powerful method for understanding a wide variety of physical phenomena in complex systems. It is a fundamental computational technique for investigating the equilibrium dynamics of proteins[3–7] and related biomolecular ensembles,[8] as well as for analyzing the properties of condensed phase systems.[9] However, a fundamental issue faced by MD simulation methods is the problem of sampling. Typical systems of interest involve the dynamics of hundreds to many thousands of molecules and a correspondingly large number of degrees of freedom. MD methods estimate equilibrium statistical ensemble properties by propagating each molecule's degrees of freedom using Newton's equations of motion augmented by thermo- and/or barostats. Consequently, the computational cost of MD methods grows as the number of particles involved in the physical processes of interest. In addition to this, the equilibrium dynamics of many physical systems of interest span multiple time and length scales. For example, biologically relevant properties of many proteins involve dynamics over the millisecond or greater time scales.[10–13] Self-assembly processes involving proteins occur over even longer time scales and length scales, as in the assembly of viral capsids,[14–17] which involves thousands of proteins self assembling to form structures tens of nanometers in length,[14] over a time period of minutes.[16,17]

Although the ability to compute the equilibrium properties of these systems would be of tremendous value, atomistic MD techniques are generally unable to do so. The fundamental problem of adequate ensemble sampling of these systems is further exacerbated by the large disparity in the time scales characterizing the physical processes involved in these dynamical phenomena themselves. If these processes were dynamical processes evolving on only a few time scales, it might be possible to adapt MD simulations to model them, by choosing a suitable multiple time step algorithm by which to propagate the MD trajectories. However, the dynamics of most physical systems of interest involve multiple time scales, ranging from hydrogen bond fluctuations, which occur on a subpicosecond time scale, to the collective dynamics of the system, which could take place on a millisecond time scale or even much longer. It is this inherent hierarchy of time scales and the coupling between them that make sampling the equilibrium ensemble of complex condensed phase systems an intractable problem in practice.

Strategies to overcome this problem involved the construction of reduced ["coarse-grained" (CG)] descriptions of the physical system being studied.[18–42] The general philosophy behind these approaches is to reduce the hierarchy of time and length scales present in the atomistic dynamics by incorporating fast time scale dynamics into the effective CG degrees of freedom. This approach enables the construction of reduced models that approximate only the relevant large scale dynamics. The reduction in the hierarchy of time and length scales in turn enables better sampling, either through MD simulations or by other methods, performed with the CG models to access physical processes at time and length scales, which would be inaccessible to conventional atomistic MD sampling methods.

Recently, a systematic multiscale approach has been developed for the construction of reduced CG models to represent a system's large scale dynamics.[32–42] This multiscale CG (MS-CG) method enables the construction of a CG potential function that is rigorously equivalent to the atomistic many body potential of mean force defined for the CG degrees of freedom.[40,41] Due to the rigorous thermodynamic averaging procedure, the MS-CG method can, in principle, guarantee the exact evaluation of ensemble averaged mechanical quantities through the analysis of the MS-CG model. Furthermore, the MS-CG construction can be derived as a minimal solution obtained from a variational principle. In practice, for a specified basis set expansion of the MS-CG potential energy function, the CG force field obtained from the MS-CG

[a)]These authors contributed equally to this work.
[b)]Present address: Department of Chemistry, Pennsylvania State University, University Park, PA 16802.
[c)]Author to whom correspondence should be addressed. Electronic mail: voth@chem.utah.edu.

variational principle can be shown to be an optimal approximation to the many body mean force field determined by the atomistic equilibrium ensemble.[40,41]

Even though the CG force fields constructed by this MS-CG procedure provide a statistical mechanical description of the system that is consistent with properties derived directly from the atomistic ensemble, they may not be completely transferable across different thermodynamic conditions, i.e., CG potential functions constructed at a given set of conditions (thermodynamic or otherwise) cannot be directly applied in the modeling of the underlying atomistic system placed in a different set of conditions.[43] Generically, the construction of reduced CG models depends on the conditions under which the specific phenomena of interest are being studied and hence the models may be poorly transferable across different physical conditions.[44,45] For the CG models constructed from the MS-CG approach, the lack of transferability across thermodynamic conditions arises because contributions from the atomistic degrees of freedom being averaged are dependent on the thermodynamic conditions of the atomistic system. Hence, this lack of transferability is a direct consequence of the statistical mechanical averaging procedure that is used to construct the MS-CG force fields.

Indeed the development of transferable potentials for CG models remains a significant challenge within the modeling community. Several groups employed thermodynamic properties to parametrize CG potentials.[22,46–49] In some cases CG models employing the resulting potentials semiquantitatively reproduced the structure of atomistic models under different conditions and for multiple systems. However, these methods have not been formulated to provide quantitative consistency with any particular atomistic model. Recent research has also demonstrated that the potentials calculated to ensure structural consistency between a particular atomically detailed model and a CG model of the same system can demonstrate significant and nontrivial dependence upon the particular thermodynamic state point used to parametrize the potentials.[44,45,50–56]

This research indicates that achieving complete transferability[43] of CG model parameters across different conditions is intrinsically in conflict with the CG modeling approach because constructing a reduced representation of an atomically detailed system inevitably involves including some amounts of information regarding its thermodynamic state into the CG representation. Partial transferability of a given set of model parameters can possibly be achieved, but at the cost of sacrificing quantitative consistency. Thus, an alternative strategy for addressing this inherent lack of transferability in CG model parameters without sacrificing quantitative consistency with underlying atomistic models would be to devise rigorous methods for transforming CG potentials that have been determined to provide quantitative consistency at a given thermodynamic state point into new potentials that provide quantitative consistency with the underlying atomistic system at different thermodynamic state points. It is important to note that this strategy does not attempt to construct a single set of completely transferable CG potential functions, but is rather aimed at generating different sets of CG potentials, each of which is designed to quantitatively model the underlying atomistic system at a single thermodynamic state point. This work presents such a strategy to generate sets of MS-CG model potentials each of which correspond to different thermodynamic state points, given information about ensemble averages at a single given state point. A systematic framework for deriving CG potentials, which can easily be transferred across thermodynamic conditions and ensure structural consistency between atomically detailed and CG models at each state point, will significantly advance the promise of computationally efficient low resolution models. In this work, a formally exact statistical mechanical method is presented by which a MS-CG force field that is an optimal approximation to the many body potential of mean force at a temperature, T, can be constructed through the use of ensemble averages calculated at a reference temperature, T_0.

The present work extends the MS-CG variational principle for transferring MS-CG force fields across different temperatures using a simple procedure. The variational principle is used to derive a rescaling technique to transfer CG thermodynamic averages across different temperatures. This method, implemented using the normal MS-CG equations,[40,57,58] is used to demonstrate the transferability of MS-CG force field parameters across different temperature regimes. Numerical calculations illustrate this procedure for the case of a simple condensed phase system for which the MS-CG potential is assumed to be pairwise decomposable. However, the proposed method is derived and may be numerically implemented for more complex potentials.

II. THEORY

A variational principle for temperature transferability of CG observables is derived in this section and applied to generalize the variational principle for the construction of MS-CG potentials. The notation to be used is as follows. An atomically detailed configuration with n atoms is described by the set of Cartesian vectors $\mathbf{r}^n=\{\mathbf{r}_1,\mathbf{r}_2,\cdots\mathbf{r}_n\}$. Similarly, a CG configuration with N CG sites is described by the set of Cartesian vectors $\mathbf{R}^N=\{\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{R}_N\}$. The atomistic n particle potential energy function is denoted as $u(\mathbf{r}^n)$ and the corresponding force on atom i is given by $\mathbf{f}_i(\mathbf{r}^n)=-\partial u(\mathbf{r}^n)/\partial\mathbf{r}_i$. In the following it will be assumed that the atomistic potential function is temperature independent and that no rigid intramolecular constraints are included in either the atomistic or the CG model.

A. Variational principle

The dynamics of a system coupled to a thermostat that maintains a fixed temperature generates a canonical ensemble of structures if the dynamical trajectories are long enough for a sufficient sampling of the phase space of the system.[1,2] In mathematical terms, if $A(\mathbf{r}^n;t)$ is the value of an observable A at time t, then the time average of A is defined by

$$\langle A(\mathbf{r}^n)\rangle_t = \frac{1}{t}\int_0^t d\tau A_\tau(\mathbf{r}^n;\tau). \quad (1)$$

For a canonical ensemble at a given temperature T, with the assumption of adequate sampling, the following relationship is satisfied:

$$\lim_{t\to\infty}\langle A(\mathbf{r}^n)\rangle_t = \langle A(\mathbf{r}^n)\rangle_T. \quad (2)$$

The average in the right hand side of this equation is taken with respect to the canonical probability distribution, $p_r(\mathbf{r}^n;T)$, for the temperature T. The canonical probability distribution at a given temperature corresponding to the atomistic potential function is given as

$$p_r(\mathbf{r}^n;T) = \frac{1}{z(T)}\exp\left(-\frac{u(\mathbf{r}_n)}{k_B T}\right). \quad (3)$$

The normalization factor $z(T)$ is the canonical configurational integral given by

$$z(T) = \int d\mathbf{r}^n e^{-u(\mathbf{r}^n)/k_B T}, \quad (4)$$

where the integral in Eq. (4) and in the following represents an integral over the relevant configuration space. The hypothesis expressed in Eq. (2) is central to the validity of the MD method wherein single, long trajectories of a given, thermostatted system are generated using Newtonian dynamics and equilibrium averages estimated through the use of the ergodic hypothesis.[1,2] Consider a set of N collective (CG) variables defined by the mapping,

$$\mathbf{R}^N = \mathbf{M}_R^N(\mathbf{r}^n) \equiv \{\mathbf{M}_{\mathbf{R}_1}(\mathbf{r}^n), \ldots, \mathbf{M}_{\mathbf{R}_N}(\mathbf{r}^n)\}. \quad (5)$$

Here, the collection of N position vectors, $\mathbf{R}^N$, is the collection of position vectors for the N CG sites, with CG site I having a position given by the vector $\mathbf{R}_I$, which is determined as a function of the atomically detailed configuration, $\mathbf{r}^n$, by the mapping operator $\mathbf{M}_{\mathbf{R}_I}(\mathbf{r}^n)$. This mapping maps linear combinations of atomistic positions into an associated CG position vector. With the CG site positions determined by the atomistic configuration, a conditional probability distribution for the atomistic configuration given a fixed CG configuration and a temperature T is defined as

$$p_{r|R}(\mathbf{r}^n|\mathbf{R}^N;T) = p_r(\mathbf{r}^n;T)\,\delta(\mathbf{R}^N - \mathbf{M}_R^N(\mathbf{r}^n))/p_R(\mathbf{R}^N;T), \quad (6)$$

where

$$p_R(\mathbf{R}^N;T) = \langle \delta(\mathbf{R}^N - \mathbf{M}_R^N(\mathbf{r}^n))\rangle_T \quad (7)$$

is the equilibrium CG coordinate distribution determined by the atomistic equilibrium distribution given by Eq. (3) and the CG mapping defined in Eq. (5), and $\delta(\mathbf{R}^N - \mathbf{M}_R^N(\mathbf{r}^n))$ is the product of N delta functions, each of the form $\delta(\mathbf{R}_I - \mathbf{M}_{\mathbf{R}_I}(\mathbf{r}^n))$. Given this conditional probability distribution, one can compute a set of reduced observables that are determined by this distribution. For a general observable, $\gamma(\mathbf{r}^n)$, a corresponding reduced variable is defined as

$$\Gamma(\mathbf{R}^N;T) = \int d\mathbf{r}^n p_{r|R}(\mathbf{r}^n|\mathbf{R}^N;T)\,\gamma(\mathbf{r}^n). \quad (8)$$

Corresponding to such a reduced observable, a functional is defined for arbitrary real continuous functions of the CG configuration, $G(\mathbf{R}^N)$, as follows:

$$\chi_T^2[G] = \int d\mathbf{r}^n p_r(\mathbf{r}^n;T)\left|G(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n)) - \gamma(\mathbf{r}^n)\right|^2. \quad (9)$$

The functional defined above is always non-negative,

$$\chi_T^2[G] \geq 0. \quad (10)$$

Furthermore, χ_T^2 has a stationary point in the vector space of real continuous functions, $G(\mathbf{R}^N)$. Since χ_T^2 is quadratic in G, the stationary point corresponds to its unique minimum. The stationary point is determined by the equation

$$\frac{\delta\chi_T^2[G]}{\delta G} = 0. \quad (11)$$

When this condition is evaluated, Eq. (9) implies that

$$G_{\min}(\mathbf{R}^N) = \Gamma(\mathbf{R}^N;T), \quad (12)$$

where $\Gamma(\mathbf{R}^N;T)$ is the reduced observable given by the conditioned expectation value of $\gamma(\mathbf{r}^n)$ and defined in Eq. (8). For arbitrary $G(\mathbf{R}^N)$

$$\chi_T^2[G] = \chi_T^2[\Gamma] + \|G(\mathbf{R}^N) - \Gamma(\mathbf{R}^N;T)\|_T^2, \quad (13)$$

where Eq. (13) introduces a temperature dependent norm

$$\|G(\mathbf{R}^N) - \Gamma(\mathbf{R}^N;T)\|_T^2 = \int d\mathbf{R}^N p_R(\mathbf{R}^N;T) \times \left|G(\mathbf{R}^N) - \Gamma(\mathbf{R}^N;T)\right|^2, \quad (14)$$

and $p_R(\mathbf{R}^N;T)$ is defined in Eq. (7). The first quantity on the right hand side of Eq. (13) is independent of G, while the second quantity on the right hand side is a non-negative quantity describing the difference between the arbitrary trial function, G, and the thermally averaged CG observable, Γ. From Eqs. (12) and (13), it is clear that the CG observable $\Gamma(\mathbf{R}^N;T)$ provides the unique minimum of the functional defined in Eq. (9) for the temperature T. Therefore the minimization of the functional χ_T^2 defined in Eq. (9) provides a variational principle for calculating the CG observable at a thermodynamic temperature T using data sampled from simulations at the same temperature. Moreover, by using the norm identified in Eq. (14) to define a "distance" between two real continuous functions of the CG configuration, $\mathbf{R}^N$, then the function that minimizes χ_T^2 within a given subspace also provides an optimal approximation to Γ within that subspace, in the sense that the optimal approximation minimizes the norm given by Eq. (14). Therefore, this variational principle also provides a systematic means for computing an optimal approximation to the CG observable within a given vector subspace of functions of the CG configuration.

B. Thermal transferability

It is evident that, in the canonical ensemble, the probability of observing a given atomistic configuration at temperature T' is related to the probability of observing the same configuration at temperature T by

$$p_r(\mathbf{r}^n;T') = \omega(T',T)\mu(\mathbf{r}^n;T',T)p_r(\mathbf{r}^n;T). \tag{15}$$

The factor ω corresponds to a ratio of configuration integrals at T' and T, while μ corresponds to a ratio of Boltzmann factors, such that

$$\omega(T',T) = (z(T')/z(T))^{-1}, \tag{16}$$

and

$$\mu(\mathbf{r}^n;T',T) = \exp[-u(\mathbf{r}^n)(1/k_BT' - 1/k_BT)]. \tag{17}$$

With these definitions, the canonical average of a function $a(\mathbf{r}^n)$ at a given temperature, T', may be evaluated from sampling the ensemble at a different temperature, T,

$$\langle a(\mathbf{r}^n)\rangle_{T'} = \int d\mathbf{r}^n p_r(\mathbf{r}^n;T')a(\mathbf{r}^n) = \omega(T';T)\langle \tilde{a}(\mathbf{r}^n;T',T)\rangle_T, \tag{18}$$

where $\tilde{a}(\mathbf{r}^n;T',T)=\mu(\mathbf{r}^n;T',T)a(\mathbf{r}^n)$. Consequently, it is convenient to define a generalized functional

$$\tilde{\chi}_T^2[G;T'] = \langle \mu(\mathbf{r}^n;T',T)|G(\mathbf{M}_R^N(\mathbf{r}^n)) - \gamma(\mathbf{r}^n)|^2\rangle_T. \tag{19}$$

This functional is expressed as a canonical ensemble average of the atomistic configuration space evaluated at a reference temperature T. The arguments of the functional are the trial function of the CG configuration, $G(\mathbf{R}^N)$, and also a target temperature T'. According to Eq. (18) $\tilde{\chi}_T^2$ is directly proportional to $\chi_{T'}^2$,

$$\tilde{\chi}_T^2[G;T'] = \frac{1}{\omega(T',T)}\chi_{T'}^2[G]. \tag{20}$$

Consequently it follows that the function $G(\mathbf{R}^N)$ that minimizes $\tilde{\chi}_T^2[G;T']$ for a given target temperature T' also minimizes $\chi_{T'}^2[G]$. This provides a variational principle for computing the CG observable at a target temperature T' using data sampled from simulations at a reference temperature, T. For a given T', $\Gamma(\mathbf{R}^N;T')$ provides the unique global minimum of $\tilde{\chi}_T^2[G;T']$. Moreover, using the distance metric defined in Eq. (14), an optimal approximation to $\Gamma(\mathbf{R}^N;T')$ within a given vector subspace of functions of the CG configuration can be obtained by minimizing $\tilde{\chi}_T^2[G;T']$ within that subspace. It is also evident from the above analysis that similar transferability can be achieved with respect to other thermodynamic parameters when the probability densities at two different values of the parameters can be related by an equation of the form of Eq. (15). This method of transferring CG quantities across different thermodynamic conditions may therefore be rather general. Similar approaches are quite common within the context of thermodynamic perturbation theory.[2] In particular, the atomistic sampling method developed by Ferrenberg and Swendsen utilizes a very similar rescaling approach.[59]

Although the procedure described above is formally exact, there are practical limitations to its applicability. A significant limitation is that the sampled regions of phase space at the given simulation conditions must have significant overlap with the regions of phase space that may be sampled at the desired conditions to which the CG quantities are to be transferred. This limitation is common to many other classical simulation techniques.

C. MS-CG variational principle

The variational procedure developed in Sec. II A can be applied to construct a CG force field, which correctly averages the forces derived from an atomistic ensemble. If an atomistic system is modeled with N CG sites, then the total "atomistic" force acting on a CG site I, $\mathcal{F}_I$, is given by the sum of atomistic forces, $\mathbf{f}_i(\mathbf{r}^n)$, from particles, i, which are mapped into the CG site. Mathematically, the atomistic force on CG site I is defined as

$$\mathcal{F}_I(\mathbf{r}^n) = \sum_{i\in I}\mathbf{f}_i(\mathbf{r}^n). \tag{21}$$

Here the set $i \in I$ indicates the set of atomistic particles, i, that are mapped into the given CG site and it has been assumed that each atomistic particle is mapped into not more than one CG site. The MS-CG variational principle,[32,33] which has recently been analyzed and generalized by Noid *et al.*,[40,41] can be obtained from the considerations leading to Eq. (13). If the function $\gamma(\mathbf{r}^n)$ in Eq. (8) is chosen to be the atomistic force, $\mathcal{F}_I$, the appropriate averaged CG observable $\Gamma(\mathbf{R}^N;T)$ is

$$\mathbf{F}_I(\mathbf{R}^N;T) = \int d\mathbf{r}^n p_{r|R}(\mathbf{r}^n|\mathbf{R}^N;T)\mathcal{F}_I(\mathbf{r}^n), \tag{22}$$

which is a many-body mean force defined by the conditioned expectation value of the atomistic force field.[41] The MS-CG functional may be considered the appropriate generalization of the functional defined in Eq. (9),

$$\chi_T^2[\mathbf{G}] = \frac{1}{3N}\left\langle \sum_{I=1}^{N}|\mathbf{G}_I(\mathbf{M}_R^N(\mathbf{r}^n)) - \mathcal{F}_I(\mathbf{r}^n)|^2\right\rangle_T, \tag{23}$$

where $\mathbf{G}=\{\mathbf{G}_1(\mathbf{R}^N),\ldots,\mathbf{G}_N(\mathbf{R}^N)\}$ is an arbitrary trial CG force field specified by a set of N vector valued functions that determine the force on each of the N CG sites as a function of the CG configuration $\mathbf{R}^N$. Because each element in the force field $\mathbf{G}$ is varied independently, it immediately follows from Eq. (13) that

$$\chi_T^2[\mathbf{G}] = \chi_T^2[\mathbf{F}(T)] + \|\mathbf{F}(T) - \mathbf{G}\|_T^2, \tag{24}$$

which employs the norm in the vector space of CG force fields introduced by Noid *et al.*:[41]

$$\|\mathbf{F} - \mathbf{G}\|_T^2 = \int d\mathbf{R}^N p_R(\mathbf{R}^N;T)\frac{1}{3N}\sum_{I=1}^{3N}|\mathbf{G}_I(\mathbf{R}^N) - \mathbf{F}_I(\mathbf{R}^N;T)|^2. \tag{25}$$

In Eq. (24), $\mathbf{F}(T)$ is the CG mean force field at the temperature T with elements $\{\mathbf{F}_1(\mathbf{R}^N;T),\ldots,\mathbf{F}_N(\mathbf{R}^N;T)\}$ given by Eq. (22) for each $I=1,\ldots,N$. Consequently, the MS-CG functional in Eq. (23) has a minimal solution $\mathbf{G}_{\min}$ that corresponds to the thermally averaged CG mean force field at temperature T, $\mathbf{F}(T)$, as defined by Eq. (22). In principle, this CG mean force field can be integrated to obtain a CG many body potential of mean force. The equilibrium canonical configurational probability distribution resulting from this potential of mean force is identical to the probability distribution, $p_R(\mathbf{R}^N;T)$, in Eq. (7). Therefore, this force field can be used to construct a CG coordinate distribution function, which is consistent with that derived for the CG configurations from the atomistic Boltzmann distribution.[40,41] The development of Sec. II B may be employed to extend the MS-CG variational principle for computing the many-body mean force field at a temperature T' given data sampled from the canonical ensemble at a temperature T. A generalized MS-CG functional of a trial CG force field $\mathbf{G}$ and a temperature T' may be defined as

$$\tilde{\chi}_T^2[\mathbf{G};T'] = \frac{1}{3N}\left\langle \mu(\mathbf{r}^n;T',T)\sum_{I=1}^{N}|\mathbf{G}_I(\mathbf{M}_R^N(\mathbf{r}^n)) - \mathcal{F}_I(\mathbf{r}^n)|^2 \right\rangle_T, \tag{26}$$

where $\mu(\mathbf{r}^n;T',T)$ has been defined in Eq. (17). It follows from Eq. (18) that

$$\tilde{\chi}_T^2[\mathbf{G};T'] = \frac{1}{\omega(T',T)}\chi_{T'}^2[\mathbf{G}]. \tag{27}$$

Thus, the many-body mean force field determined by a conditional canonical average of the atomistic force field at a temperature T', $\mathbf{F}(T')$, provides the unique minimum of the functional $\tilde{\chi}_T^2$ defined by Eq. (26). Moreover, given a vector subspace of CG force fields, the optimal approximation within this subspace to the many-body mean force field at a temperature T' may be determined by minimizing $\tilde{\chi}_T^2$ using data sampling the canonical ensemble at the temperature T.

D. Numerical calculations: The normal MS-CG equations

It is assumed that the total MS-CG potential function $U(\mathbf{R}^N)$ may be expressed as a sum of simple CG potential functions, U_ζ, each of which is a function of a single scalar variable, ψ_ζ, which is itself a function of the coordinates for a set of CG particles, $\{\mathbf{R}\}_\gamma$, and which describes a particular interaction among that set,

$$U(\mathbf{R}^N) = \sum_\zeta \sum_\gamma U_\zeta(\psi_\zeta(\{\mathbf{R}\}_\gamma)). \tag{28}$$

Each of these potential functions, U_ζ, will be represented by a linear combinations of basis functions $U_{\zeta d}$ with coefficients, $\phi_{\zeta d}$,

$$U_\zeta(z) = \sum_d \phi_{\zeta d} U_{\zeta d}(z). \tag{29}$$

The MS-CG force on site I, $\mathbf{F}_I$, may then be expressed as a linear combination of basis functions that depend on the CG configuration, $\mathbf{R}^N$, and whose coefficients will act as force field parameters:

$$\mathbf{F}_I(\mathbf{R}^N) = \sum_\zeta \sum_d \phi_{\zeta d}\mathcal{G}_{I;\zeta d}(\mathbf{R}^N), \tag{30}$$

where

$$\mathcal{G}_{I;\zeta d}(\mathbf{R}^N) = \sum_\gamma F_{\zeta d}(\psi_\zeta(\{\mathbf{R}\}_\gamma))\frac{\partial \psi_\zeta(\{\mathbf{R}\}_\gamma)}{\partial \mathbf{R}_I} \tag{31}$$

play the role of force field basis vectors for the variational calculation and

$$F_{\zeta d}(z) = -dU_{\zeta d}(z)/dz \tag{32}$$

are the (negative) derivatives of the potential energy basis functions.[41,42] For notational simplicity the superindex D will be used to represent the combination of index ζ, indicating the particular type of interaction, and index d, indicating the particular coefficient describing the potential function governing that type of interaction. The CG force on site I in configuration $\mathbf{R}^N$ may then be re-expressed:

$$\mathbf{F}_I(\mathbf{R}^N;T) = \sum_D \phi_D(T)\mathcal{G}_{I;D}(\mathbf{R}^N). \tag{33}$$

The set of force field basis vectors $\{\mathcal{G}_D\}$ included in Eq. (33) forms a basis that spans a particular vector subspace of CG force fields. As discussed above, the CG mean force field depends on the thermodynamic temperature, T, at which the ensemble averaging was performed. Thus, the CG force field within the subspace spanned by the given basis set and also the parameters determining this CG force field depend on the thermodynamic temperature, as explicitly indicated in Eq. (33). It is important to recognize that in Eqs. (30) and (33) $\mathbf{F}_I$ represents the MS-CG approximation to the many body mean force field defined in Eq. (22) and not the mean force field itself.

By employing the linear representation for the MS-CG force field at temperature T given in Eq. (33), the MS-CG functional defined in Eq. (23) reduces to a simple function of the force field parameters and the variational calculation may be performed as a linear least-squares problem. The parameters determining the CG force field that minimizes the MS-CG functional in Eq. (23) and that therefore provide an optimal approximation to the many-body mean force at a temperature T satisfy the normal MS-CG equations,[40–42]

$$\sum_{D'} G_{DD'}(T)\phi_{D'}(T) = b_D(T), \tag{34}$$

for each parameter ϕ_D included in the MS-CG force field, where the functions

$$G_{DD'}(T) = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \mathcal{G}_{I;D}(\mathbf{M}_R^N(\mathbf{r}^n))\cdot\mathcal{G}_{I;D'}(\mathbf{M}_R^N(\mathbf{r}^n))\right\rangle_T, \tag{35}$$

$$b_D(T) = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \mathcal{G}_{I;D}(\mathbf{M}_R^N(\mathbf{r}^n))\cdot\mathcal{F}_I(\mathbf{r}^n)\right\rangle_T, \tag{36}$$

are canonical ensemble averages evaluated at temperature T.

Analogously, the CG force field parameters that provide the optimal approximation to the many-body mean force field at a temperature T' may be determined by performing the variational calculation outlined above and solving the normal system of equations, Eq. (34), using correlation functions evaluated at temperature T'. Alternatively the generalized MS-CG functional defined in Eq. (26) may be used to determine the force field parameters providing the optimal approximation to the many body mean force field at the temperature T' by reweighting configurations sampled from the canonical ensemble at temperature T. The optimal CG force field at the temperature T' may then be determined by solving the generalized set of normal MS-CG equations:

$$\sum_{D'} \tilde{G}_{DD'}(T',T)\phi_{D'}(T') = \tilde{b}_D(T',T), \tag{37}$$

where

$$\tilde{G}_{DD'}(T',T) = \frac{1}{3N}\left\langle \mu(\mathbf{r}^n;T',T)\sum_{I=1}^{N} \mathcal{G}_{I;D}(\mathbf{M}_R^N(\mathbf{r}^n)) \cdot \mathcal{G}_{I;D'}(\mathbf{M}_R^N(\mathbf{r}^n))\right\rangle_T, \tag{38}$$

$$\tilde{b}_D(T',T) = \frac{1}{3N}\left\langle \mu(\mathbf{r}^n;T',T)\sum_{I=1}^{N} \mathcal{G}_{I;D}(\mathbf{M}_R^N(\mathbf{r}^n)) \cdot \mathcal{F}_I(\mathbf{r}^n)\right\rangle_T, \tag{39}$$

are correlation functions evaluated using configurations sampled from a canonical ensemble at the temperature T.

In the following calculations, each molecule is represented by a single CG site that corresponds to the center of mass for that molecule. The CG potential energy function includes only central pair potentials that are represented by discrete delta function basis functions as described previously in Refs. 41 and 44. In this representation, the CG pair force is discretized on a grid of spacing δR and the force field parameter ϕ_D corresponds to the magnitude of the pair force between two CG sites separated by a distance R_D. The CG force field basis vector $\mathcal{G}_{I;D}$ then becomes

$$\mathcal{G}_{I;D}(\mathbf{R}^N) = \sum_{J\neq I} \hat{\mathbf{R}}_{IJ}\delta_D(R_{IJ}-R_D), \tag{40}$$

where $\hat{\mathbf{R}}_{IJ}$ is the unit vector pointing toward CG site I from CG site J and the discrete delta function,

$$\delta_D(R-R_D) = \begin{cases} 1 & R_D - \delta R/2 < R < R_D + \delta R/2 \\ 0 & \text{otherwise} \end{cases}. \tag{41}$$

The correlation functions G and b then reduce to

$$G_{DD'}(T) = \frac{1}{3N}\left\langle \int d\mathbf{R}^N \sum_{I=1}^{N}\sum_{J,K\neq I} (\hat{\mathbf{R}}_{IJ}\cdot\hat{\mathbf{R}}_{IK})\delta_D(R_{IJ} - R_D)\delta_D(R_{IK}-R_{D'})\delta(\mathbf{M}_R^N(\mathbf{r}^n)-\mathbf{R}^N)\right\rangle_T, \tag{42}$$

$$b_D(T) = \frac{1}{3N}\left\langle \int d\mathbf{R}^N \sum_{I=1}^{N}\sum_{J\neq I} (\hat{\mathbf{R}}_{IJ}\cdot\mathcal{F}_I(\mathbf{r}^n)) \times \delta_D(R_{IJ}-R_D)\delta(\mathbf{M}_R^N(\mathbf{r}^n)-\mathbf{R}^N)\right\rangle_T, \tag{43}$$

where the mapping operation $\delta(\mathbf{R}^N-\mathbf{M}_R^N(\mathbf{r}^n))$ in each correlation function has been used to express the basis vectors in terms of CG coordinates, $\mathbf{R}^N$, and $\mathbf{R}_{IJ}=\mathbf{R}_I-\mathbf{R}_J$, $R_{IJ}=|\mathbf{R}_{IJ}|$, $\hat{\mathbf{R}}_{IJ}=\mathbf{R}_{IJ}/R_{IJ}$, and $\mathbf{R}_{IK}$, R_{IK}, and $\hat{\mathbf{R}}_{IK}$ are defined analogously. The correlation functions $\tilde{G}_{D,D'}(T',T)$ and $\tilde{b}_D(T',T)$ in this basis then immediately follow from Eqs. (42) and (43) using the rescaling function $\mu(\mathbf{r}^n;T',T)$.

III. METHODS

The method developed above for transferring CG potential functions between temperatures is applied and validated here for a Lennard-Jones (LJ) liquid and the simple point charge (SPC) water model.[60] In each case, MD simulations were performed in the constant NVT ensemble using an integration time step of 2 fs and employing the Nose–Hoover thermostat[61,62] to sample the canonical ensemble at the appropriate temperature. Electrostatic interactions were evaluated using the particle mesh Ewald method[63] with the real space contribution truncated at 0.9 nm for SPC water. In each case the k space contribution was evaluated using a maximum grid spacing of 0.08 and a sixth order interpolation between grid points. van der Waals interactions were truncated at 1.2 nm. For the LJ and SPC simulations, configurations, forces, and energies were sampled every 2 ps and 100 fs, respectively. All MD simulations and calculations of thermodynamic quantities were performed using the GROMACS 3.3 software package.[64]

Simulations of 216 LJ particles were performed in a cubic box of volume $(1.86\ \text{nm})^3$ at a temperature of $T=298$ K. Atomically detailed simulations of 216 SPC water molecules were performed in a cubic box of volume $(1.86\ \text{nm})^3$ at $T=278$, 298, and 318 K. All bond lengths in the SPC molecules were constrained using the LINCS algorithm.[65] Initial configurations for SPC simulations at $T=278$ and 318 K were obtained by simulated annealing in which the temperature was changed from $T=298$ K during the course of 2 ns and data from the first 5 ns of each simulation were not included in the following calculations.

CG configurations were generated from each simulation by mapping the atomic coordinates for each molecule to the center of mass for that molecule. (In the case of LJ system, the CG mapping is the trivial identity map.) Radial distribution functions (RDFs) were then calculated from the mapped configurations for each system at each temperature. The coordinates and forces from each sampled configuration were next employed to evaluate the canonical ensemble averages in Eqs. (35) and (36). Using these correlation functions, the normal MS-CG equations in Eq. (34) were solved to determine the pair forces providing the optimal approximation for the many-body mean force field for each system at each temperature.[41,42] The grid spacing for all calculations of the CG pair force was 0.001 nm. The resulting CG pair forces

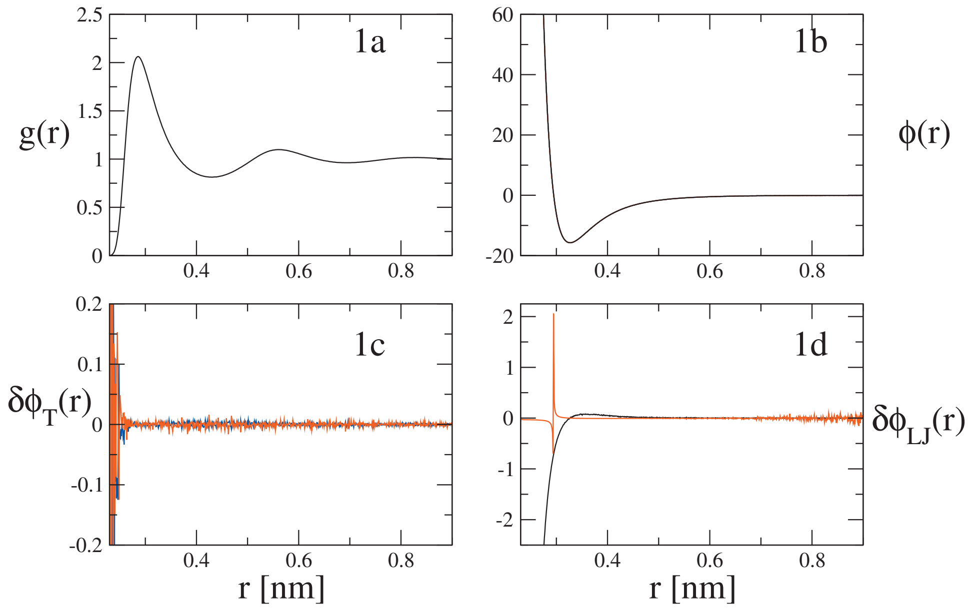

FIG. 1. Force and pair distribution curves plotted for the LJ liquid system: (a) is the RDF, $g(r)$, plotted as a function of distance, r, and computed from a simulation at 298 K. (b) compares the true LJ pair force (black) at $T=298$ K with the pair forces calculated for the LJ system at $T'=318$ K (red) and $T'=278$ K (blue) using the rescaling procedure with simulation data from $T=298$ K and also with the pair force calculated using the normal equations at $T=298$ K (dotted black). (c) displays the absolute error in the rescaling procedure, $\delta\phi_T$, by plotting the difference between the LJ pair force computed from the MS-CG normal equations at $T=298$ K and the LJ pair forces at $T'=278$ K (blue) and $T'=318$ K (red) computed via the rescaling procedure and using simulation data from $T=298$ K. (d) presents the percentage error (red) and absolute error (black) in the LJ pair force, $\delta\phi_{\mathrm{LJ}}$, calculated using the normal MS-CG equations compared to the true LJ pair force.

were then numerically integrated to obtain the pair potential. Additionally, the second derivative of the CG pair potential was approximated by evaluating the centered finite differences of the calculated pair force and then smoothing the resulting curve by performing a running average over five consecutive data points. The resulting pair potential and its second derivative were then employed in CG MD simulations at the same temperature and volume as the corresponding atomistic simulation. Averages and fluctuations in the potential energy and pressure as well as distribution functions for the CG sites were computed from configurations sampled from these CG MD simulations.

Additionally, for both systems, the configurations, energies, and forces sampled from simulations at temperature $T=298$ K were employed to evaluate the correlation functions $\tilde{G}_{DD'}(T',T)$ and $\tilde{b}_D(T',T)$ defined in Eqs. (38) and (39) for $T'=278$ and 318 K. These correlation functions evaluated using data from simulations at 298 K were then employed in the rescaled normal MS-CG equations in Eq. (37) to solve for the MS-CG pair force at 278 and 318 K. The pair force functions calculated from rescaling were compared with force functions calculated using data for the appropriate temperature. These rescaling calculations required significant sampling to obtain converged results and 45 ns worth of data were used in the SPC calculations. The pair potential and its second derivative were calculated from these forces as described above and were then employed in CG MD simulations at the temperature corresponding to the rescaling temperature T'. Again, averages and fluctuations in the potential energy and pressure, as well as distribution functions for the CG sites, were computed from configurations sampled from these simulations and compared with those obtained from the CG MD simulations using the pair forces calculated without rescaling.

IV. RESULTS

In Fig. 1 and in all subsequent figures, forces are presented in units of kJ/(mol nm). Figure 1(a) presents the RDF, $g(r)$, calculated from simulations of the LJ system at $T=298$ K. Figure 1(b) compares the true LJ pair force curve (black) with the pair force curves calculated from several MS-CG variational calculations. The dotted black curve in Fig. 1(b) corresponds to the pair force at $T=298$ K calculated from the normal MS-CG equations presented in Eq. (34). The red and blue curves in Fig. 1(b) correspond to the pair force curves at $T'=278$ and 318 K, respectively, calculated from the rescaled normal MS-CG equations in Eq. (37), while using data from LJ simulations at $T=298$ K. Because the identity map has been used to define the "CG" system, the correct CG pair force equals the true LJ pair force and should be temperature independent. These calculations are therefore "control" simulations to test the validity of the temperature rescaling method. The four pair forces calculated presented in Fig. 1(b) are seen to be virtually indistinguishable on the scale of the figure. Figure 1(c) presents the absolute errors in the rescaling procedure for this simple control system as the difference between the pair force curves calculated for $T'=278$ (blue curve) and 318 K (red curve) from the rescaled normal MS-CG equations, Eq. (37), using $T=298$ K data with the pair force curve calculated for $T=298$ K using the normal MS-CG equations, Eq. (34). Figure 1(d) presents the absolute (black) and percent error (red)

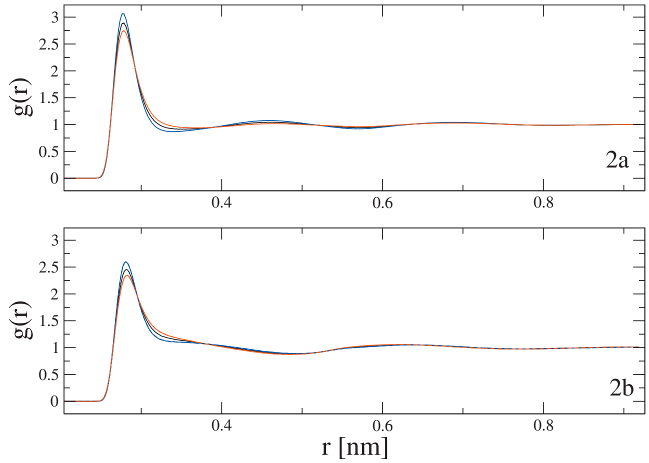

FIG. 2. Comparison of RDFs derived from MD simulations of all-atom and CG models of water at three different temperatures: (a) presents the RDFs calculated from all-atom MD simulations. The blue, black, and red solid curves correspond to RDFs calculated from all-atom MD simulations of SPC water at 278, 298, and 318 K, respectively. (b) presents RDFs calculated from CG MD simulations. The blue, black, and red solid curves correspond to RDFs calculated at 278, 298, and 318 K, respectively, using pair forces calculated via the normal MS-CG equations in Eq. (34). The blue and red dashed curves correspond to RDFs calculated from CG MD simulations at 278 and 318 K using pair forces calculated via the rescaled normal MS-CG equations in Eq. (37). The RDFs for the CG MD simulations at each temperature are indistinguishable on the scale of the figure.

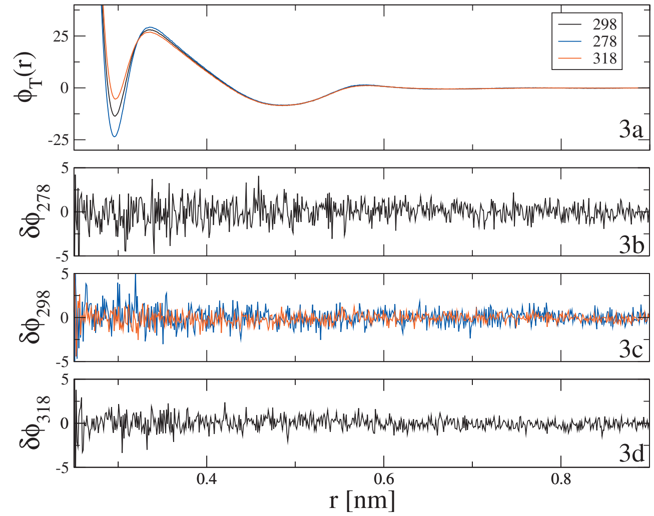

FIG. 3. CG pair forces for the SPC water model calculated via the normal MS-CG equations in Eq. (34) and the rescaled normal MS-CG equations in Eq. (37). (a) presents the MS-CG pair force for 278 (blue), 298 (black), and 318 (red) K calculated via the normal MS-CG equations. (b) presents the absolute difference between the pair force at $T=278$ K calculated via the normal equations and the pair force at $T'=278$ K calculated via the rescaled normal MS-CG equations using simulation data at $T=298$ K. (c) presents the absolute difference between the pair force at $T=298$ K calculated via the normal equations and the pair forces at $T'=298$ K calculated via the rescaled normal equations using simulation data at $T=278$ K (blue) and $T=318$ K (red). (d) presents the absolute difference between the pair force at $T=318$ K calculated via the normal equations and the pair force at $T'=318$ K calculated via the rescaled normal equations using simulation data at $T=298$ K.

in the LJ pair force curve calculated using the normal MS-CG equations as the absolute and relative difference between the pair force obtained from Eq. (34) and the true LJ pair force. The spike in the percent error curve (red) in Fig. 1(d) corresponds to the point at which the true pair force passes through zero. The magnitude of the absolute error curve (black) in Fig. 1(d) for $r<0.3$ nm corresponds to a small relative error in the magnitude of the pair repulsion at very short interparticle separations. In summary, the results of Fig. 1 clearly indicate that, for this simple control case, the MS-CG method accurately recovers the true LJ pair force and that the rescaling procedure accurately predicts the temperature independence of the MS-CG pair force for the LJ system over the temperature range from $T=278$ to 318 K using simulation data at $T=298$ K.

To further demonstrate the method, the temperature rescaling procedure was employed to determine the MS-CG pair forces for a system of 216 water molecules represented by the SPC water model.[60] Atomistic MD simulations at $T=278$, 298, and 318 K were performed for this system as described in Sec. III. Atomistic configurations were sampled from these simulations and mapped onto one-site CG water configurations. From these mapped configurations, the site-site RDF was computed for each temperature. The resulting RDFs computed from atomistic simulations at $T=278$, 298, and 318 K are presented in Fig. 2(a) as the solid blue, black, and red curves, respectively. Over this temperature range the SPC water RDFs demonstrate relatively simple temperature dependence. With decreasing temperature, the SPC water system becomes increasingly ordered. In particular, the first peak of the SPC RDF becomes increasingly large and increasingly narrow, while the second peak demonstrates slightly more order at lower temperature. The third peak appears effectively temperature independent over this temperature range.

From the mapped configurations, the correlation functions G and b defined in Eqs. (35) and (36) were computed for each temperature and the normal MS-CG equations in Eq. (34) for each temperature were numerically solved. The calculated MS-CG pair forces for $T=278$, 298, and 318 K are presented as the blue, black, and red curves in the top panel of Fig. 3, i.e., Fig. 3(a). The MS-CG pair forces for the SPC system demonstrate significant temperature dependence in the first solvation shell, but appear to be essentially temperature independent for $r>0.4$ nm. At $T=278$ K, the MS-CG pair force is strongly attractive in the first solvation shell. However, for $T=318$ K the MS-CG pair force has almost no attraction between neighboring molecules. Moreover, with decreasing temperature, the MS-CG pair force demonstrates increasing structure as the repulsive maxima in the SPC pair force increases with decreasing temperature. As discussed earlier by Noid *et al.*[40] and in contrast with iterative CG procedures such as the reverse Monte Carlo method,[9,29] the MS-CG pair force is not determined completely by the RDF and reflects not only changes in pair correlations, but also changes in three-particle correlations in the SPC fluid with temperature.

As discussed in Sec. III, the CG pair potential and its second derivative were then computed from the calculated pair force. Figure 4 illustrates the numerical procedure for the CG pair interaction at $T=318$ K. The black curve in the central panel of Fig. 4 is the MS-CG pair force calculated

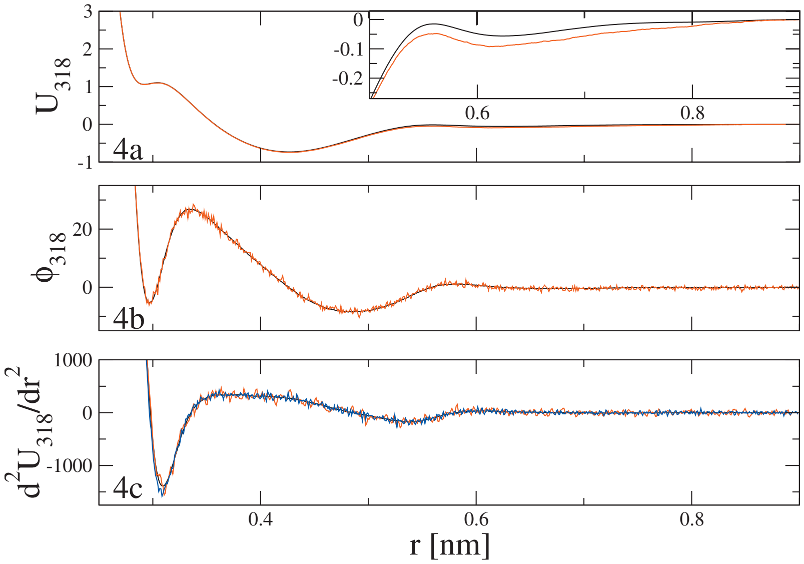

FIG. 4. The MS-CG interaction at 318 K calculated via the normal MS-CG equations in Eq. (34) (black) and via the rescaled normal MS-CG equations in Eq. (37) using simulation data at 298 K (red). (a) presents the two pair potentials. The inset of (a) compares the two potentials for r between 0.4 and 0.9 nm. (b) presents the two calculated force curves. (c) presents the second derivatives of the pair potentials calculated as smoothed centered finite differences of the pair forces shown in (b). The blue curve in (c) presents the centered finite difference of the black curve in (b) without smoothing.

from simulations at $T=318$ K. The potential obtained from integration is presented as the black curve in Fig. 4(a). The second derivative which was used in the ensuing CG MD simulations is presented as the black curve in the lower panel. For comparison, the blue curve in Fig. 4(c) is the second derivative calculated directly from the pair force without any smoothing. Clearly, the running average effectively removed much of the statistical noise, allowing an accurate and smooth evaluation of the necessary numerical derivative.

MD simulations of the one-site CG SPC model were then performed at each temperature using the MS-CG pair forces calculated for that temperature. The site-site RDFs were computed from the CG MD simulations at $T=278$, 298, and 318 K and are presented as the solid blue, black, and red curves, respectively, in Fig. 2(b). It can be seen from Fig. 2(b) that the one-site MS-CG model for SPC water qualitatively reproduces structural features of the water model derived from all-atom simulations. For each temperature, the first peak of the CG RDF is at approximately the correct distance. However, although the relative ordering of the CG RDFs with temperature follows the same trend as the RDFs calculated for the atomistic model, the structure of the second solvation shell in the CG model and the temperature dependence of the CG RDFs are quantitatively different from the corresponding structural properties calculated for the SPC model. Since the objective of this work is to test the temperature transferability methodology, the quantitative accuracy of this CG model of SPC water is not a significant concern.

Using the CG configurations mapped from atomistic MD simulations at each temperature, the rescaled correlation functions $\tilde{G}$ and $\tilde{b}$ defined by Eqs. (38) and (39) were computed. The rescaled normal MS-CG equations in Eq. (37) were then solved to determine (1) the MS-CG pair force at $T'=278$ K using simulation data from $T=298$ K, (2) the MS-CG pair force at $T'=298$ K using simulation data from $T=278$ K, (3) the MS-CG pair force at $T'=298$ K using simulation data from $T=318$ K, and (4) the MS-CG pair force at $T'=318$ K using simulation data from $T=298$ K. These four MS-CG pair forces calculated via the rescaling method were then compared with the MS-CG pair forces computed for each temperature from simulation data at the same temperature, which are presented in Fig. 3(a) and described above. Figures 3(b)–3(d) then present the absolute numerical difference between the MS-CG pair force computed for each temperature ($T=278$, 298, and 318 K, respectively) from the normal MS-CG equations and the pair force computed for each temperature from the rescaled normal MS-CG equations. In particular, Fig. 3(c) presents the difference between the MS-CG pair force at $T=298$ K calculated from Eq. (34) and the pair forces at $T'=298$ K calculated from Eq. (37) by rescaling simulation data from $T=278$ K and $T=318$ K simulations as the blue and red curves, respectively. In both cases, the pair force calculated from the rescaled normal equations agrees quite well with the pair force calculated directly from the normal equations. However, the statistical error in the pair force calculated by rescaling data from 278 K is slightly larger than the statistical error in the pair force calculated by rescaling data from 318 K. This is not unexpected, since simulations at a lower temperature can be expected to sample the Boltzmann distribution less efficiently than a high temperature simulation. As a result, a rescaling of simulation data down a temperature gradient (i.e., from a higher to a lower temperature) would be expected to be more efficient and reliable than a rescaling estimate performed for a given temperature using simulation data from a lower temperature. In summary, it is clear from Fig. 3 that, although the rescaled force curves are somewhat more noisy than the force curves calculated directly from simulation data at the appropriate temperature, the error in the rescaled force curves represents statistical error arising from lack of sampling and not systematic error. Within a relatively small statistical error, the current method accurately calculates the MS-CG pair force at a particular temperature by rescaling simulation data sampled from simulations at reference temperatures ± 20 K from the target temperature.

As described in Sec. III and above, the pair forces calculated for each temperature via rescaling were then em-

TABLE I. Comparison of thermal averages and root mean squared deviation (RMSD) fluctuations in the potential and pressure computed from CG MD simulations performed with the MS-CG pair forces of 278 and 318 K obtained from the (unscaled) normal and rescaled normal MS-CG equations.

$T=278$ K	Average	RMSD
Potential (kcal/mol)		
Unscaled	−75.96	23.26
Rescaled	−90.70	23.33
Pressure (bar)		
Unscaled	7233.69	570.41
Rescaled	7180.92	575.90
$T=318$ K		
Potential (kcal/mol)		
Unscaled	−73.73	25.60
Rescaled	−265.07	25.68
Pressure (bar)		
Unscaled	8338.49	575.69
Rescaled	7843.17	577.25

ployed to obtain the CG pair potential and its second derivative. Figure 4 further compares the pair potential and its derivatives at $T=318$ K computed via the normal MS-CG equations (black) and the rescaled normal MS-CG equations (red). The direct comparison of the two force curves presented in Fig. 4(b) clearly demonstrates that the force curve calculated by rescaling simulation data from $T=298$ K provides an accurate, if somewhat noisy, approximation to the CG pair force at $T=318$ K. [The difference between these two force curves is presented in Fig. 3(d).] The potentials computed from the two force curves are virtually indistinguishable on the scale of Fig. 4(a). However, the inset, which compares the two pair potentials for r between 0.5 and 0.9 nm, demonstrates that the rescaled pair potential (red) slightly underestimates the true pair potential (black). Additionally, Fig. 4(c) demonstrates that, despite the statistical noise present in the rescaled force curve, upon performing a running average, the derivative of the rescaled force curve [red curve in Fig. 4(c)] provides a reasonably accurate approximation to the second derivative of the pair potential [black curve in Fig. 4(c)].

CG MD simulations were then performed at $T'=278$ and 318 K using the pair potentials (and second derivatives) calculated by rescaling simulation data sampled at $T=298$ K. The RDF for the CG model at both temperatures was then computed from configurations sampled by these simulations. The RDFs computed from these CG MD simulations performed at $T'=278$ and 318 K are presented as the dashed blue and red curves, respectively, in Fig. 2(b). On the scale of Fig. 2(b), the RDFs computed from simulations using the pair potentials calculated by rescaling are indistinguishable from the RDFs computed from simulations using the pair potentials calculated directly from the normal MS-CG equations shown as solid lines. Additionally, Table I presents the average and standard deviation of the potential and pressure calculated from these CG MD simulations. Table I indicates that simulations employing the pair potentials calculated by rescaling approximately reproduce the thermodynamic properties calculated from simulations employing the MS-CG pair potentials determined by the normal equations for the given temperature. The largest discrepancy arises in the potential energy at $T=318$ K and results from the slight error in the pair potential shown in the inset of Fig. 4(a).

V. DISCUSSION

The results presented above demonstrate that the thermal rescaling approach proposed in this work may provide a viable and systematic method by which MS-CG force fields can be transferred across different temperatures. As has been demonstrated from the model studies presented here, the rescaling procedure accurately determines the MS-CG force field at a particular target temperature by rescaling either from lower or higher temperatures than the target temperature. As noted in the results section, the rescaling calculations for the SPC system required extensive sampling at the reference temperature in order to rescale the force curve to the target temperature. Consequently, more careful numerical procedures using higher order basis functions,[42] perturbation expansion approaches, and/or numerical preconditioning techniques may be necessary for implementing the rescaling procedure for more complex biomolecular systems.

The results suggest that the rescaling method may be used in a symmetric fashion when the target temperature is either higher or lower than the reference temperature and may be extended to provide accurate approximations to CG force fields for other liquid systems, more generally for systems which can be well sampled over the time scale of the simulation trajectories. However, the rescaling method presented here is possibly an asymmetric method, with a rescaling downward from a higher temperature ensemble being potentially more accurate. This asymmetry is related to the fact that higher temperatures could enable a more exhaustive sampling of phase space.

The method presented in this work offers a promising approach to transferring CG potential functions across a range of thermodynamic conditions. Although this work provided a numerical demonstration of the method to transfer force parameters across temperatures, in principle, this method has more general applicability to the transferability of other thermodynamically averaged quantities. This can be seen clearly from the formulation of the generalized variational principle on which the method is based. However, this method becomes considerably simplified when applied to the transfer of force parameters. This simplification is made possible by the use of the normal MS-CG equations, Eqs. (34)–(36), which relate the CG forces directly to statistical mechanical distribution functions of the system and to simple kinetic equations for the liquid state.[40,57,58] The strategy by which the variational principle is combined with the formal hierarchy for the distribution functions has considerable generality and promises to provide a powerful tool for the development of easily transferable MS-CG potentials for complex condensed phase systems, where the challenge of adequate sampling becomes especially severe.

ACKNOWLEDGMENTS

This research was supported by a Collaborative Research in Chemistry grant from the National Science Foundation (Grant No. CHE-0628257). W.G.N. acknowledges funding from the National Institutes of Health through a Ruth L. Kirschstein National Research Service Award postdoctoral fellowship (Grant No. 5 F32 GM076839-02). Discussions with Dr. S. Iuchi, Dr. Y. Wang, and Dr. P. Liu are gratefully acknowledged. This work benefited greatly from our ongoing collaboration with Professor Hans Andersen and his group at Stanford University. We acknowledge many stimulating discussions with them on the problem presented in this paper.

APPENDIX: GENERALIZED THERMAL TRANSFERABILITY EQUATIONS

This appendix presents the thermal transferability equations in a general, matrix form. Let $\underline{\underline{G}}(T)$ and $\underline{b}(T)$ be the temperature dependent matrix and vector of correlation functions determined for a temperature T with an associated vector of force parameters, $\underline{\phi}(T)$. Then, the relation between $\underline{\underline{G}}$, $\underline{b}$, and $\underline{\phi}$ at two temperatures, T and T', is

$$\underline{\underline{G}}(T)\underline{\phi}(T) = \underline{b}(T),$$

$$\underline{\underline{G}}(T')\underline{\phi}(T') = \underline{b}(T'). \tag{A1}$$

Define

$$\underline{\delta\phi} = \underline{\phi}(T') - \underline{\phi}(T),$$

$$\underline{\underline{G}}(T') = \underline{\underline{\Gamma_G}}\,\underline{\underline{G}}(T),$$

$$\underline{b}(T') = \underline{\underline{\Gamma_b}}\,\underline{b}(T). \tag{A2}$$

Note that $\underline{\underline{G}}$ is symmetric by definition and that Γ_b is a diagonal matrix with elements $(\underline{\underline{\Gamma_b}})_{DD'} = \delta_{DD'} b_D(T')/b_D(T)$. The quantities $(\underline{\underline{\Gamma_G}}, \underline{\underline{\Gamma_b}})$ contain the effects of a temperature change on the correlation functions, $(\underline{\underline{G}}, \underline{b})$. From these definitions, Eq. (A2) can be rewritten as

$$\underline{\delta\phi} = \underline{\underline{G}}^{-1}(T)\underline{\underline{\Delta}}\,\underline{b}(T), \tag{A3}$$

where

$$\underline{\underline{\Delta}} = \underline{\underline{\Gamma_G^{-1}}}\,\underline{\underline{\Gamma_b}} - \underline{\underline{1}}. \tag{A4}$$

The quantity $\underline{\underline{\Delta}}$ is zero when $T=T'$ and hence the right hand side of Eq. (A3) represents the "perturbation" to the CG force resulting from changes in the temperature. Consequently, this form of the thermal transferability equations, Eq. (A3), may be useful in constructing approximate perturbative treatments estimating the effect of changes in thermodynamic state on the MS-CG force fields.

[1] M. P. Allen and D. P. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, New York, 2002).
[3] M. Karplus and J. A. McCammon, Nat. Struct. Mol. Biol. **9**, 646 (2002).
[4] H. L. Scott, Curr. Opin. Struct. Biol. **12**, 495 (2002).
[5] J.-W. Chu and G. A. Voth, Proc. Natl. Acad. Sci. U.S.A. **102**, 13111 (2005).
[6] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, Structure (London) **14**, 437 (2006).
[7] K. Y. Sanbonmatsu, S. Joseph, and C. S. Tung, Proc. Natl. Acad. Sci. U.S.A. **102**, 15854 (2005).
[8] T. Hansson, C. Oostenbrink, and W. F. van Gunsteren, Curr. Opin. Struct. Biol. **12**, 190 (2002).
[9] F. Muller Plathe, ChemPhysChem **3**, 754 (2002).
[10] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[11] K. Simons and D. Toomre, Nat. Rev. Mol. Cell Biol. **1**, 31 (2000).
[12] T. Pawson and P. Nash, Science **300**, 445 (2003).
[13] J. Onuchic and P. G. Wolynes, Curr. Opin. Struct. Biol. **14**, 70 (2004).
[14] S. C. Harrison, Curr. Opin. Struct. Biol. **11**, 195 (2001).
[15] W. R. Wikoff and J. E. Johnson, Curr. Biol. **9**, R296 (1999).
[16] B. Ganser-Pornillos, M. Yeager, and W. I. Sundquist, Curr. Opin. Struct. Biol. **18**, 203 (2008).
[17] S. Casjens and J. King, Annu. Rev. Biochem. **44**, 555 (1975).
[18] M. Levitt and A. Warshel, Nature (London) **253**, 694 (1975).
[19] M. Levitt, J. Mol. Biol. **104**, 59 (1976).
[20] V. Tozzini, Curr. Opin. Struct. Biol. **15**, 144 (2005).
[21] G. S. Ayton, W. G. Noid, and G. A. Voth, Curr. Opin. Struct. Biol. **17**, 192 (2007).
[22] S. J. Marrink, A. H. de Vries, and A. E. Mark, J. Phys. Chem. B **108**, 750 (2004).
[23] E. Villa, A. Balaeff, L. Mahadevan, and K. Schulten, Multiscale Model. Simul. **2**, 527 (2004).
[24] E. Villa, A. Balaeff, and K. Schulten, Proc. Natl. Acad. Sci. U.S.A. **102**, 6783 (2005).
[25] S. O. Nielsen, C. F. Lopez, I. Ivanovic, P. B. Moore, J. C. Shelley, and M. L. Klein, Biophys. J. **87**, 2107 (2004).
[26] M. J. Stevens, J. Am. Chem. Soc. **127**, 15330 (2005).
[27] H. Gohlke and M. F. Thorpe, Biophys. J. **91**, 2115 (2006).
[28] R. Goetz and R. Lipowsky, J. Chem. Phys. **108**, 7397 (1998).
[29] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).
[30] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[31] I. G. Kevrekidis, C. W. Gear, and G. Hummer, AIChE J. **50**, 1346 (2004).
[32] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).
[33] S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).
[34] S. Izvekov and G. A. Voth, J. Chem. Phys. **125**, 151101 (2006).
[35] Y. Wang, S. Izvekov, T. Y. Yan, and G. A. Voth, J. Phys. Chem. B **110**, 3564 (2006).
[36] Q. Shi, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **110**, 15045 (2006).
[37] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, Biophys. J. **92**, 4289 (2007).
[38] S. Izvekov, A. Violi, and G. A. Voth, J. Phys. Chem. B **109**, 17019 (2005).
[39] P. Liu, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **111**, 11566 (2007).
[40] W. G. Noid, J.-W. Chu, G. S. Ayton, and G. A. Voth, J. Phys. Chem. B **111**, 4116 (2007).
[41] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).
[42] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, J. Chem. Phys. **128**, 244115 (2008).
[43] W. L. Jorgensen, J. Am. Chem. Soc. **103**, 335 (1981).
[44] A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[45] M. E. Johnson, T. Head-Gordon, and A. A. Louis, J. Chem. Phys. **126**, 144509 (2007).
[46] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, J. Phys. Chem. B **111**, 7812 (2007).
[47] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, J. Chem. Theory Comput. **4**, 819 (2008).
[48] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandhyopadhyay, and M. L. Klein, J. Phys. Chem. B **105**, 4464 (2001).
[49] W. Shinoda, R. Devane, and M. L. Klein, Mol. Simul. **33**, 27 (2007).
[50] H.-J. Qian, P. Carbone, X. Chen, H. A. Karimi-Varzaneh, C. C. Liew, and F. Müller-Plathe, Macromolecules **41**, 9919 (2008).
[51] P. Carbone, H. A. Karimi-Varzaneh, X. Chen, and F. Müller-Plathe, J. Chem. Phys. **128**, 064904 (2008).
[52] T. Vettorel and H. Meyer, J. Chem. Theory Comput. **2**, 616 (2006).
[53] J. Ghosh and R. Faller, Mol. Simul. **33**, 759 (2007).
[54] J. Silbermann, S. H. L. Klapp, M. Shoen, N. Channasetty, H. Block, and

K. E. Gubbins, J. Chem. Phys. **124**, 074105 (2006).
[55] J. Fischer, D. Paschek, A. Gieger, and G. Sadowski, J. Phys. Chem. B **112**, 13561 (2008).
[56] S. Garde and H. S. Ashbaugh, J. Chem. Phys. **115**, 977 (2001).
[57] M. Born and H. S. Green, Proc. R. Soc. London, Ser. A **188**, 10 (1946).
[58] J. Yvon, *Fluctuations en Densité*, Actualités Scientifiques et Industrielle, No. 542 (Hermann, Paris, 1937), pp. 203.
[59] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[60] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331–342.
[61] S. Nose, J. Chem. Phys. **81**, 511 (1984).
[62] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[63] T. A. Darden, D. M. York, and L. G. Petersen, J. Chem. Phys. **98**, 10089 (1993).
[64] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[65] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).