Structural and thermodynamic consistency in coarse-grained models of macromolecules

This content has been downloaded from IOPscience. Please scroll down to see the full text.

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 128.223.130.222
This content was downloaded on 27/06/2017 at 00:29

Please note that terms and conditions apply.

You may also be interested in:

Computer models of bacterial cells: from generalized coarsegrained to genome-specific modular models
Evgeni V Nikolaev, Jordan C Atlas and Michael L Shuler

A thermodynamically consistent non-ideal stochastic hard-sphere fluid
Aleksandar Donev, Berni J Alder and Alejandro L Garcia

A comment on the Helmholtz free energy in the mean density approximation and thermodynamic consistency
M Itoh

Dynamics of biomolecular processes
Hans Behringer, Ralf Eichhorn and Stefan Wallin

The Tsallis distribution in proton--proton collisions at s = 0.9 TeV at the LHC
J Cleymans and D Worku

A variational approach for the free energy
J R Faleiro Ferreira and N P Silva

A Mean-Field Treatment in Studying Nuclear Matter Through a Thermodynamic Consistent Resummation Scheme
Shu Song and Li Jiarong

Stochastic thermodynamics of resetting
Jaco Fuchs, Sebastian Goldt and Udo Seifert

XXVI IUPAP Conference on Computational Physics (CCP2014) IOP Publishing
Journal of Physics: Conference Series **640** (2015) 012009 doi:10.1088/1742-6596/640/1/012009

Structural and thermodynamic consistency in coarse-grained models of macromolecules

M.G. Guenza

Department of Chemistry and Biochemistry, Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403

E-mail: `mguenza@uoregon.edu`

Abstract. Coarse-grained models of complex liquids are becoming increasingly important in simulations of complex fluids because they substantially reduce computational time while bridging information between simulations performed with atomistic resolution and continuum models. However, to be useful coarse-graining approaches need to provide physically consistent representations of the system under study, independent of the degree of coarse-graining resolution, i.e. they need to ensure consistency in both the thermodynamic and structural properties.

Developing coarse-graining approaches that ensure consistency of thermodynamic and structural properties across variable coarse-graining resolution is a challenge. The difficulty is in properly mapping many-body interactions into effective pair interactions, which are input to the fast mesoscale molecular dynamic simulations of the coarse-grained representations.

In this paper we review a Coarse-Grained approach, based on Integral Equation theory. This model conserves thermodynamics while accurately reproducing the structure of polymer liquids across variable levels of resolution. As it is solved analytically, it provides a formalism to reconstruct *a posteriori* the correct dynamics from the accelerated dynamics of the coarse-grained simulation. This approach has the advantage of being largely analytically solved and it is useful to highlight some important and general points that need to be considered when developing coarse-graining models.

1. Introduction

In recent years the field of coarse-graining (CG) has been in rapid expansion. The interest in coarse-graining methods is motivated by the need of extending the range of time and length scales that can be covered by simulations, for example molecular dynamics. Coarse-graining methods help bridge the region in resolution between atomistic simulations and continuum models by partitioning the intermediate region into subparts where the system is described at variable levels of resolution. [1, 2, 3, 4, 5, 6, 7]

During a coarse-graining procedure a number of atoms are grouped together defining a new “fictitious” coarse-grained unit that is centered in their center of mass. In doing this simple process, the dimensionality of the configurational space is reduced, smoothing the probability distribution. The new statistical unit represents now a number of microstates that have been averaged out during coarse-graining. This modifies the entropy of the system as coarser descriptions represent a larger number of “ghost” microstates, which are not present in the final coarse-grained representation. The “shape” and size of the molecule are also modified as a consequence of coarse-graining and so is its friction coefficient.

With increasing the level of coarse-graining an increasing number of fast processes are eliminated in the reduced description in favor of random fluctuations for the dynamics of the coarse-grained units. In this sense the process of coarse-graining is analogous to the transformation by projection operator techniques, e.g. Mori-Zwanzig, of the Liouville equation into a Langevin equation. The potential that in the Liouville equation is the Hamiltonian of the system becomes a potential of mean-force, i.e. a free energy of the reduced variables. And while the dynamics in the Liouville equation is deterministic it becomes dissipative in the Langevin equation, where the fluctuations of the suppressed particles enter as random forces and their correlation defines the memory function.[8]

With changes in the number of microstates that are sampled and in the "shape" of the molecule, i.e. changing the extent of surface available to the random collisions of the "solvent" and so its friction coefficient, the dynamics of the CG units accelerates. The coarser the CG description adopted the faster the simulation, and the higher the computational gain.[9]

While a coarser description is computationally more useful, all the problems and uncontrolled consequences of coarse-graining emerge more dramatically the higher the level of coarse-graining. These can affect the mesoscale CG simulations leading to unwanted incorrect predictions of structure, thermodynamics, and dynamics. Clearly it is quite important to develop coarse-grained descriptions that are statistically well controlled, or alternatively to develop methods to evaluate the extent of errors that emerge from the use of poorly controlled coarse-graining method.

In this paper we review a coarse-grained method for macromolecular liquids that is solved analytically, and in this way provides a mean to understand the consequences of coarse-graining in relation to the consistency of structural and thermodynamic quantities, and to the dynamics. The Integral Equation Coarse-Grained (IECG) model is solved for a variable level of coarse-graining, representing the macromolecule in the liquid as a collection of soft blobs whose number can be arbitrarily chosen.[10, 11] The theory ensures an analytical solution of the effective coarse-grained potential, and structural and thermodynamic consistency at variable CG resolution.[12] It provides also information about the general properties of coarse-graining and a method for rescaling the dynamics obtained from mesoscale simulations.[9, 13, 14]

2. Numerical versus analytical coarse-grained methods

Most coarse-grained methods are solved numerically. Numerically solved CG models assume a trial function as the starting potential, and rely on data from high resolution atomistic simulations or from experiments to optimize self-consistently the parameters of the potential until the test function is reproduced.[1, 2, 3, 4, 5, 15]

As mentioned befoere, because through the coarse-graining procedure the interaction potential becomes a free energy of the system, it depends on all the parameters that define the system including thermodynamic as well as molecular parameters.[16] In this way a numerically optimized potential is specific of the system for which it has been optimized, i.e. the molecular structure of the system and thermodynamic parameters, and of the physical quantities that have been used to optimize the model. Furthermore, because the free energy potential depends on so many parameters, a "brutal" optimization of all the quantites of interest is not a viable strategy.

In principle, a numerically optimized potential does not apply to any other system in different thermodynamic conditions or for different molecular parameters than the ones for which the potential has been optimized. Even if the calculations are performed at the correct state conditions numerical potentials optimized to reproduce one quantity, e.g. the correct pair distribution function, do not necessarily reproduce any other physical quantity, such as the pressure or energy. Furthermore, an effective pair potential optimized at one set of conditions (molecular structure and thermodynamic) will not generally be transferable to another set of conditions. These are commonly referred to as the "representability problem" and the “transferability problem” respectively. Finally, numerically optimized potentials are precise to the level of the physical quantity used to optimize them, and if the data are truncated or contain some experimental errors, the CG potential will carry this “hidden error” in any following application.

The hope for numerically optimized potential is that they are applicable in a region of the phase diagram close to the one for which they have been optimized, and for systems that are similar from a molecular point of view. We see that this is an acceptable way to proceed for CG models that group together a small number of atoms. These Numerical High Resolution CG models (NHRCG), once parameterized to reproduce the total correlation function, predict increasingly incorrect thermodynamic quantities with decreasing coarse-graining resolution.[7, 10, 11] This is the reason behind the success of fine-graining models like the united atom model. If the model is very coarse-grained the potential is specific, and it is convenient to have an analytical form of the potential because all the parameters entering the potential are explicit.

The ultimate purpose of developing coarse-graining methods is to have a mesoscale potential that can be directly applied in molecular dynamics simulations without the need of performing initial atomistic simulations to parameterize them. In this way, analytical coarse-graining methods when they can be derived appear to be a quite promising strategy to model complex systems on a low resolution scale.

3. An analytical and transferable coarse-graining method for polymer liquids

Starting from integral equation theory, we developed a coarse-graining method called the Integral Equation Coarse-Graining (IECG) approach, which allows for variable coarse-graining resolution in the molecular representation while conserving thermodynamics and structure. The IECG potential is fully transferable and the method is predictive in the sense that it is not necessary to perform atomistic simulations to parameterize the potential. The IECG formalism has been solved analytically for systems where the distribution of CG sites is Markovian and the distribution of monomers in the coarse-grained unit is Gaussian. These requirements are fulfilled in polymer liquids when the CG length scale is chosen to be larger than a number of persistence lengths. This is the length scale where the local molecular structure of the polymer in the form of localized conformational barrier and rigidity is averaged out and the central limit theorem applies to the intrablob site distribution.

The IECG method is not particularly tied to any of the approximations just described. It can be applied to a large number of systems, for example where the structure of the molecule is characterized by non-Gaussian statistics, by adopting a numerical solutions of the equation. It is important to point out that the numerical solution of the coarse-graining equation is not related at all to the numerical optimization of a parametric potential, which is typical of numerically optimized CG models. In the IECG formalism, even when the equations are solved numerically, the parameter dependence is clearly defined.

The analytical IECG model has been developed so far for homopolymer and block-copolymer liquids, and for polymer mixtures.[6, 7, 10, 17, 18, 19] The analytical solution pertains to isotropic liquids of polymers, and when simulating system undergoing phase separation it does not describe the strong segregation regime where the liquid is strongly anisotropic and macromolecules are stretched. For sake of simplicity we summarize in this paper the IECG approach for a homogeneous liquid of homopolymer chains because for this system the mathematical formalism is the simplest and a number of essential aspects of the procedure of coarse-graining already clearly emerges.

The approach has several steps. Starting from the Ornstein-Zernike integral equation we derived the pair distribution function of the coarse-grained model and the effective potential acting between units. Using the potential we performed simulations of the coarse-grained systems, and then compared thermodynamic quantities and structural quantities of interest from these coarse-grained simulations with united atom simulation data. The agreement of structural and thermodynamic properties between CG and atomistic descriptions is quantitative, while the analytical expressions depend on one non-trivial parameter, the direct correlation function at large distances, $c(k \to 0) = c_0$. This function is evaluated either from experiments or from theory. Other parameters are the themodynamic properties of temperature, T, density, ρ, as well as the structural parameters of the number of monomers in a chain, N, and the effective segment length, σ. The molecular specificity enters the theory through the effective segment length, which depends on the stiffness of the polymer, and the direct correlation function at large separation distances, which is also specific of the macromolecule that is simulated.

The formalism is then used to calculate the analytical correction factors that are needed to rescale the dynamics measured from the fast molecular dynamics simulations of the coarse-grained system and obtain the correct dynamical quantities for the atomistic scale. For example, the reconstructed diffusion coefficients show quantitative agreement with united atom simulations and with experiments performed using a number of different techniques. The one parameter needed to be adjusted in the dynamic reconstruction procedure is the effective hard-sphere diameter in the Lennard-Jones mapping of the molecule from atomistic simulation into a chain of hard-spheres.[9, 13, 14]

The model applies to any type of polymer, because the length scale of coarse-graining has only to be assumed to be larger than its local persistence length. By selecting a length scale larger than the persistence length, which is specific of the polymer considered, the coarse-grained units are statistically uncorrelated and follow a random walk in space. The chain of blobs can then be modeled as freely jointed. This model has unique characteristics because being analytical is fully transferable: it applies to different points in the phase diagram, and represents well any type of homopolymer liquid, independent of the molecular structure of the monomer. In this model atomic-scale simulations are used only as a test and not to optimize the numerical form of the CG potential.

4. The Integral Equation Coarse-Graining Approach for a liquid of homopolymers

In developping a coarse-grained model it is common to start from an atomistic simulation and to build the new coarse-grained description in a reduced representation by matching physical quantities in the atomistic and reduced descriptions. These numerically optimized models, however, suffer from several problems, as discussed earlier on in this paper, including the lack of transferability. To overcome this problem, we developped a coarse-graining model that approach the problem from a different perspective.

It is well known that for each system a pair distribution function is uniquely defined[20] and that when this pair distribution function is known it is possible to calculate from this function all the thermodynamic quantities of interest using the statistical mechanics of liquids.[21] A method to calculate the pair distribution function of a molecular liquid is to use integral equation theory. Specifically we start from the Ornstein-Zernike integral equation theory for the coarse-grained description and for the monomer-level description. The latter is the Polymer Reference Interaction Site Model (PRISM) approach.[22] The coarse-grained sites are represented as fictitious sites and the monomer/atomistic sites are represented as real sites. This roughly corresponds to assuming strongly repulsive or Lennard-Jones interactions between atomistic sites and bound soft potentials of much smaller amplitude than the atomistic one between fictitious and real sites, and between fictitious sites. In a low resolution coarse-grained model where the number of atomic sites represented by the coarse-grained unit is high, each CG unit represents the center-of-mass of a polymer subchain. Because two center-of-mass have some probability of superimposing their interaction potential is bound and soft.[10]

The solution of the pair distribution function from the Integral Equation theory does not require performing atomistic simulations. Instead the atomistic pair distribution function is defined starting from the molecular and the thermodynamic parameters of the system. Once the pair distribution function is derived, then the effective pair potential between the coarse-grained units is calculated using the appropriate closure, which for soft bound potentials is the HyperNetted Chain (HNC) approximation.[16] The closure approximation has the function of mapping three-body and higher-order many-body contributions, which are important in dense liquids, into the effective pair interaction potential.

The derived pair potential is an input to mesoscale simulations of the coarse-grained system, which once they are performed provide all the molecular and thermodynamic quantities of interest. Those are, for example, the pair distribution function of the coarse-grained description, the pressure, the excess free energy, as well as the internal energy and the entropy.[12] The structural and thermodynamic quantities show quantitative agreement with the ones measured in atomistic simulations for the system under study, but atomistic simulations are not used to optimize the potential. Only one non-trivial parameter, which is related to the compressibility of the liquid, needs to be adjusted as discussed later on in this paper.

The homopolymer liquid is composed of n molecules in a volume V, with each chain including a total number of monomer N. The density of chains is $\rho_{ch} = n/V$ and is related to the liquid monomer density $\rho = \rho_{ch} N$. Every chain in the liquid is partitioned in a variable number of coarse-grained units or blobs, n_b, each containing a number $N_b = N/n_b$ of monomers, with the blob density $\rho_b = \rho/N_b$. Given N monomers with a chain density ρ_{ch}, and an effective segment length $\sigma = \sqrt{6/N} R_g$ with R_g being the polymer radius of gyration and $R_{gb} = R_g/\sqrt{n_b}$ the blob radius of gyration.

We derived specific expressions for the potential for the single soft sphere representation, where each chain is represented as one sphere, as well as for the three blob and the five blob representations. For the three-blob and five-blob representations the blobs are not all equivalent, as the blobs at the end of the chains are different than the one(s) in the internal part of the chain. For chains with a number of blobs larger than five end effects become negligible and it is possible to use a blob-averaged description.[11].

The intramolecular distributions in the blob-averaged limit are normalized as $\Omega(k) = \omega(k)/N$, for the blob-blob (bb), blob-monomer (bm), and monomer-monomer (mm) distributions. The normalized blob-monomer and the blob-blob distributions are given respectively as

$$\hat{\Omega}^{bm}(k) = \frac{1}{n_b}\left[\frac{\sqrt{\pi}}{kR_{gb}} Erf\left(\frac{kR_{gb}}{2}\right) e^{-\frac{k^2R_{gb}^2}{12}} - 2\left(\frac{e^{-n_b k^2 R_{gb}^2} - n_b e^{-k^2R_{gb}^2} + n_b - 1}{k^2 R_{gb}^2 n_b (e^{-k^2R_{gb}^2} - 1)}\right) e^{-k^2R_{gb}^2/3}\right] , \tag{1}$$

and

$$\hat{\Omega}^{bb}(k) = \frac{1}{n_b} + 2\left[\frac{e^{-n_b k^2 R_{gb}^2} - n_b e^{-k^2R_{gb}^2} + (n_b - 1)}{n_b^2 (e^{-k^2R_{gb}^2} - 1)^2}\right] e^{-2k^2R_{gb}^2/3} . \tag{2}$$

The monomer distribution $\Omega^{mm}(k)$ is normalized as

$$\Omega^{mm}(k) = \hat{\omega}^{mm}(k)/N = \frac{2}{n_b^2 k^4 R_{gb}^4}(k^2 R_{gb}^2 n_b - 1 + e^{-n_b k^2 R_{gb}^2}) . \tag{3}$$

Given the Ornstein-Zernike relation for the coarse-grained blob representation

$$\hat{h}^{bb}(k) = n_b \hat{\Omega}^{bb}(k) \hat{c}^{bb}(k) \left[n_b \hat{\Omega}^{bb}(k) + \rho_b \hat{h}^{bb}(k)\right] , \tag{4}$$

the bob-blob direct correlation function is given as

$$\hat{c}^{bb}(k) = \frac{\hat{h}^{bb}(k)}{n_b \hat{\Omega}^{bb}(k) \left[n_b \hat{\Omega}^{bb}(k) + \rho_b \hat{h}^{bb}(k)\right]} . \tag{5}$$

The effective pair potential between blob CG units is calculated from the Fourier transform of Eq. 4 and Eq. 5 and the Hyper Netted Chain closure equation

$$\frac{v^{bb}(r)}{k_B T} = -\ln\left[h^{bb}(r)+1\right] + h^{bb}(r) - c^{bb}(r) \ . \tag{6}$$

The solution of Eq.6 can be obtained analytically when the intramolecular distributions are represented by simple formal equations, or numerically in the other cases.

When $|h^{bb}(r)| << 1$, which always holds at large separations ($r >> 1$ in units of R_{gb}) and at any separation for representations with large N_b and high densities, the potential further simplifies to

$$V^{bb}(r) \approx -k_b T c^{bb}(r) \ . \tag{7}$$

This formula is referred to as the mean spherical approximation (MSA) in the literature, and applies to low compressible polymer liquids.[23] If this formalism is improperly used to treat low density liquids, where the mean spherical approximation does not hold, this approximation would lead to unphysical behavior. For low density liquids the full closure approximation has to be used.

We now focus on the effective direct correlation function, $c^{bb}(r)$, and the related MSA potential. In the limit of large separations in real space, where $r >> 1$ (in R_{gb} units), the inverse transform integral is sufficiently dominated by $\hat{c}^{cc}(k)$ (in the small wave vector limit) that the large wave vector contribution can be entirely neglected. Furthermore, since the expansion for small wave vectors is bounded at large k, the error incurred in using the small k form with the integral bounds extended to infinity is small. This approximation leads, for $r >> 1$ (in R_{gb} units), to

$$\begin{aligned} c^{bb}(r) \ &\approx \ \tfrac{-N_b\Gamma_b}{2\pi^2\rho_m R_{gb}^3 r}\int_0^\infty \left(k\sin(kr)\left[\tfrac{45}{45+\Gamma_b k^4} + \tfrac{5k^2}{28}\tfrac{13\Gamma_b k^4-3780}{(\Gamma_b k^4+45)^2}\right]\right)dk \\ &= \ [-\left(\tfrac{45\sqrt{2}N_b\Gamma_b^{1/4}}{8\pi\sqrt{3}\sqrt[4]{5}\rho_m R_{gb}^3}\right)\tfrac{\sin(Q'r)}{Q'r}e^{-Q'r} + \left(\tfrac{\sqrt{5}N_b}{672\pi\rho_m\Gamma_b^{1/4}R_{gb}^3}\right)[(13Q^3(Q'r-4))cos(Q'r) \\ &\quad + \left(\tfrac{945+13Q^4}{\Gamma_b^{1/4}}\right)r\sin(Q'r) + \tfrac{945r}{\Gamma_b^{1/4}}cos(Q'r)]\tfrac{e^{-Q'r}}{Q'r}] \ , \end{aligned} \tag{8}$$

where $Q' = 5^{1/4}\sqrt{3/2}\Gamma_b^{-1/4}$ and $Q \equiv Q'\Gamma_b^{1/4}$. This is the main result of our theory as it provides an anlytical, albeit approximated, expression for the effective pair potential between CG units in a liquid of homopolymers. Some relevant points emerge from the analysis of this equation.

The key quantity of interest here is the universal parameter $\Gamma_b = N_b\rho|c_0|$, which is defined once we decide the level of coarse-graining, N_b, as well as the molecular and thermodynamic parameters of our system. This quantity also depends on the direct correlation function at $k = 0$, c_0, which is in principle not known. This function is however related to the potential between atomistic units and to the isothermal compressibility of the liquid. It can be determined numerically or from experiments as we discuss later in this paper.

The range of the potential, in units of the radius of gyration of the blob, scales as $\Gamma_b^{1/4} \propto N_b^{1/4}$. This scaling behavior describes how the interaction between effective units propagates through the liquid. By increasing the density of the liquid or decreasing the resolution in the coarse-grained description, i.e. increasing the number of monomeric units in the blob, the potential becomes increasingly longer ranged. The characteristic scaling with the number of monomers in the blob emerges from the many-body interaction pathway following a random walk ($N_b^{1/2}$) in the space defined by the lengthscale of the blob-blob interpenetration, which also scales with the degree of polymerization as $N_b^{1/2}$.

Interestingly, the range of the potential increases with the number of monomers comprised in the coarse-grained unit, i.e. blob or soft sphere, while the potential at contact decreases.

However the interblob potential does not vanish even when the length of the polymer chain becomes infinity, indicating that intermolecular interactions between polymers are important even for infinitely long chains. This result disagrees with the conventional assumption in polymer physics that intramolecular interactions are dominant over intermolecular contributions, and a polymer melt can simply be described by mean-field approaches of the single chain.[24]

The potential becomes longer-ranged with increasing the lengthscale of coarse-graining. For highly coarse-grained models the gain in computational time is maximized. However the presence of long-ranged forces in the molecular dynamic simulations makes the use of a large box necessary, given that the simulation box is usually chosen to be at least twice the range of the effective potential. Then the simulation is slowed down by the need of simulating a larger number of CG units. In practice the two effects tend to compensate and their balance will dictate the choice of the level of coarse-graining. Finally, the inconvenience of having long-range interactions and large boxes can be alleviated by the use of simulations in reciprocal space, as it is conveniently done in the case of electrostatic interactions using the Ewald summation.[16]

The potential has a long-range, slowly decaying repulsive component and a second attractive part that is smaller in comparison with the repulsive part. This attractive contribution is important when one evaluates the thermodynamic properties of the system and cannot be discarded. Higher order terms, which are present in the equation of the potential, tend to give increasingly more negligible contributions. The potential in our simulations is often truncated after the first attractive well, or more rarely the second repulsive contribution, depending on error minimization.

It is interesting to note that the attractive contribution is present in the effective potential even when the intermolecular atomistic potential, from which the coarse-grained potential is derived, is purely repulsive. This indicates that the attractive contribution to the intermolecular potential is, at least partially, a consequence of coarse-graining and of the propagation of the interactions through the liquid. Being the resultant of the projection of many-body interactions onto the pair of coarse-grained units, the attractive component of the potential contains contributions that are of entropic origin.

When already at the atomistic-scale the interaction potential includes an attractive part, i.e. for example it is a Lennard-Jones potential, the attractive part of the atomistic potential also provides a contribution to the total attractive component of the CG potential. In that case the attractive CG component is enhanced with respect to purely entropic one arising from the hard-sphere monomer-monomer interaction.

5. Thermodynamics: equation of state, compressibility, and Helmholtz free energy

Given the analytical form of the potential for the multiblob coarse-grained description we formally derived the thermodynamic quantities of interest, starting form the equation of state.[12] The normalized pressure is derived using the virial expression of a fluid under the assumption of pairwise additivity as

$$\frac{P}{\rho_{ch} k_B T} = 1 - \frac{2\pi\rho_b}{3k_B T}\int_0^\infty \frac{dV^{bb}(r)}{dr} g^{bb}(r) r^3 dr \approx 1 - \frac{N\rho c_0}{2} \ , \tag{9}$$

with ρ_b the number blob density.[25] The equation of state Eq.9 is independent of the level of coarse-graining adopted in the multiblob description, as it is expected given that the pressure is a macroscopic thermodynamic property of the liquid. All the non-ideal contributions that arise from system-specific interactions are contained in the non-trivial parameter c_0, which is specific of the system under consideration, depends on the density of the system, and is independent of the resolution of the coarse-grained model.

The isothermal compressibility is derived from the equations for the blob and monomer total correlation functions. For the different CG models, which have variable level of resolution, the isothermal compressibility is independent of the level of coarse-graining and is given by

$$\rho k_B T \kappa_T = \frac{N}{1-\rho N c_0} \ . \tag{10}$$

The excess Helmholtz free energy per monomer, calculated relatively to the energy of the system in its gas phase, is simply given by

$$\frac{F-F_0}{nk_BT} = \frac{\rho_b}{2k_BT}\int_0^\infty V^{bb}(r)g^{bb}(r)4\pi r^2 dr = -\frac{Nc_0\rho}{2} \ , \tag{11}$$

with ρ_b the number density of blobs. Both the isothermal compressibility and the excess free energy are depending on the density and number of monomer in the polymer, as well as on the non-universal parameter c_0, however they are independent of the level of coarse-graining, as it is reasonable to expect for macroscopic liquid properties.

It is important to note that these equations are derived from the approximated analytical solution of the effective potential, which is accurate in the mean-field limit of a nearly incompressible liquid. The potential so derived should not be used to describe systems that are highly compressible, for example system at low density. In that case the equation would correctly predict an instability of the liquid, but the predictions will be qualitative.

Because in the range of thermodynamic parameters, temperature and density, of a regular polymeric liquid the theory predicts thermodynamic properties that are in quantitative agreement with the atomistic simulations, this result is indicative that many-body interactions, which are typical of liquids, are well accounted for in the two-body effective potential. While intergal equation theories are expressed as expansions in density of the many-body interactions, the closure approximation projects onto effective pair interactions the many-body terms. In this case the HNC closure equation, which is appropriate for bound soft potentials, is shown to be a valid approximation.

Mostly because of the need of a closure equation an integral equation theory inspired coarse-grained model like the one we developed is by its nature approximated. However, the corrections to the closure approximation are contained in the effective parameter of the direct correlation function at $k \to 0$. Once the c_0 parameter is defined for the system under study, we see that the thermodynamic quantities are found to be fully consistent across different levels of coarse-graining, and with the atomistic description. The equation of state was tested against united atom simulations of polyethylene liquids at variable chain length and different state points. The consistency between atomistic and coarse-grained descriptions both at the formal level and when compared with simulation data support the approximations adopted in the analytical solution of the potential.

From the equation of state of the pressure, all the thermodynamic quantities can be calculated.[16] However to do so it is necessary to know the dependence of the non-trivial parameter c_0 on the thermodynamic parameters for the given system that is simulated. To do so, we derived an equation for c_0 as a function of the liquid packing fraction, as described in the following section.

6. Determination of the monomer direct correlation function $c(k \to 0) = c_0$.

6.1. Soft-sphere Equation of State

As mentioned earlier the only non-trivial parameter in our theory is the direct correlation function c_0, which is related through the Ornstein-Zernike equation to the compressibility of the liquid and to the equation of state. The direct correlation function is a parameter specific of the system that is simulated and has a non-trivial dependence on the liquid density, while it is independent of the level of coarse-graining.

Because in the IECG fromalism the equation of states and the compressiblity and the excess free energy do not depend on the number of coarse-grained units in which the molecule is partitioned, it is convenient to fromally derived an equation of state for the polymer liquid in the most coarse-grained, and easiest to model, representation, i.e. the soft sphere model. This is the simplest, most reduced, representation where the whole macromolecule is described as a point particle interacting through an effective long-ranged potential. Effectively a liquid of soft spheres is a simple liquid, and expressions for the thermodynamic properties of this liquid can be easily derived following and further implementing the existing literature of hard-sphere fluids.

We derived for the soft-sphere liquid an equation of state in of the form of a Carnahan-Starling expression in analogy with the equation for the hard-sphere liquid. The equation of state is expressed as a function of an effective soft-sphere packing fraction

$$\eta_{eff} = \frac{\pi}{6}\rho d^3 \ , \tag{12}$$

where d the effective soft sphere diameter, as

$$\frac{P}{\rho k_B T} = \frac{4(\eta_{eff} + c_1\eta_{eff}^2 + c_2\eta_{eff}^3)}{(1-\eta_{eff})^3} \ . \tag{13}$$

The equation of state includes three numerical parameters (d, c_1, and c_2), which reflect the specificity of the macromolecular structure, e.g. the chain connectivity, and the fact that the real potential is not of the simple hard-sphere form.

As an example, the numerical values of these parameters c_1, c_2, and d have been calculated for liquids of polyethylene chains. Specifically we displayed data of normalized pressure as a function of an effective packing fraction. We used data for polyethylene from simulations performed at $T = 400\ K$ and increasing degree of polymerization ($N = 44$, 66, 78, 100, and 200) and variable density, as well as simulations for polyethylene samples at $T = 509\ K$, density $\rho = 0.03153\ sites/A^3$, and degree of polymerization $N = 36$, 44, 66, 78, 100, 192, 224, and 270. All the data fall onto an universal curve, which is well represented by the equation of state for soft spheres, Eq.13 when the parameters are defined as following: the optimized effective sphere diameter is $d = 2.5\ \AA$, while the other two parameters are $c_1 = -11.9$ and $c_2 = 31.11$. It is interesting to see that the soft sphere effective diameter is actually of the order of the Lennard-Jones sigma parameter, indicating that polymer chains can highly interpenetrate and the centers-of-mass of a pair of polymers can superimpose in space.

Following the same procedure the direct correlation function at $k \to 0$ is also defined as a function of the same parameters that enter the virial equation of state

$$c_0 = -\frac{4\pi d^3}{3}\frac{1 + c_1\eta_{eff} + c_2\eta_{eff}^2}{(1-\eta_{eff})^3} \ . \tag{14}$$

Once the parameters of the equation of state are defined, the non-trivial parameter c_0 is also defined. Eq.14 shows how the density dependence of c_0, through the effective packing fraction, is not simply linear. When thermodynamic properties are calculated as a function of a variable liquid density, the dependence of c_0 on the thermodynamic and structural parameters ensures the prediction of the correct thermodynamic quantities of interest.

It is worth noticing that a number of simulations had to be perfomed to calibrate the curve of the equation of state as a function of the effective packing fraction, and to define in this way the parameters in the equation of state. However once this procedure is completed for a given macromolecular liquid in a range of temperatures and densities, the parameters are defined and can be used to calculate c_0 for any other liquid of the macromolecule of interest and in the range of temperatures and densities where the optimized parameters apply. In this limits the theory, once it is calibrated, is quantitative and fully predictive.

The direct correlation function, c_0, appears to be a function of the density and is specific of the monomeric structure of the polymer, but it does not depend on the degree of coarse-graining selected, i.e. it is not a function of the number of blobs or the number of monomers in a blob. In this way the formalism ensures full structural and thermodynamic consistency across the different coarse-graining resolution.

6.2. Other methods to evaluate c_0

While the definition of the equation of state for soft spheres is a way to calculate c_0, this is by no means the only possible way to do so. For example we have shown how to calculate c_0 from the numerical solution of the PRISM equation starting from a realistic representation of the polymer chain. The solution of the PRISM equation provides results that are consistent with the equation of state method described above.[12]

A third method to evaluate the direct correlation function at $k = 0$ is to directly use the experimental isothermal compressibility of the liquid under study. Using this method the theory directly connects the coarse-graining method with the experimental data, without the need of performing atomistic simulations.

The isothermal compressibility, κ_T, which is also preserved during coarse-graining, is related to the static structure factor $\hat{S}(k = 0)$ as

$$\rho k_B T \kappa_T = \hat{S}(k = 0) = [N + \rho \hat{h}^{mm}(0)] \ , \tag{15}$$

with $\hat{S}(k = 0)$ the $k \to 0$ limit of $\hat{S}(k) = [\hat{\omega}^{mm}(K) + \rho \hat{h}^{mm}(K)]$. The isothermal compressibility in the blob description is formally identical to the isothermal compressibility in the monomer description. The value of c_0 is then determined, for example in the monomer description, as

$$c_0 = \frac{\hat{h}^{mm}(0)}{\rho N \hat{h}^{mm}(0) + N^2} \ , \tag{16}$$

with $\hat{h}^{mm}(0)$ related to the isothermal compressibility through Eq. 15.

7. Potential energy and entropy

In the multiblob description the *potential energy* has both inter- and intra-molecular contributions. The intermolecular part is calculated as an extension of the formula for the soft sphere

$$\frac{E^{bb}}{n k_B T} = \frac{2\pi \rho_b}{k_B T} \int_0^\infty v^{bb}(r) g^{bb}(r) r^2 dr \approx -\frac{\rho N c_0}{2} \ , \tag{17}$$

and gives a contribution that even in the multiblob description is a constant. The potential energy, however, contains in this case also contributions from the intramolecular structure.

Since the bond energy is a harmonic potential with a Gaussian probability distribution, the average bond energy is simply the equipartition result,

$$\left\langle \frac{E_{bond}}{n k_B T} \right\rangle = \frac{3(n_b - 1)}{2} \ . \tag{18}$$

For the angular contribution to the energy we add an additional $E_{angle} \approx (n_b - 2)/2 k_B T$ contribution per chain. In this way the potential energy depends on the level of coarse-graining. Tests of the equation for the potential energy as measured from the united atom simulations show that the theory and the simulations are quantitatively in agreement.[12]

The basic procedure of any coarse-graining formalism is the averaging of the microscopic states that are then represented by effective units, with the consequence that the *entropy* of the system in a given coarse grained representation is different with respect to the atomistic description. The extent of the change in entropy depends on the level of detail maintained in the coarse-grained representation, which determines the number of atomistic configurations that can be mapped into a single coarse-grained configuration. It can be quite large when the level of coarse-graining is extreme and the underlying chain is flexible. This is commonly called as the "mapping entropy," and is simply the difference in entropy of the atomistic model when viewed from the atomistic configurations and the coarse-grained configurations.[26]

If the chain is assumed to have a statistical distribution of monomers in space that follows a Gaussian form, the entropy associated with increasing the number of blobs in the coarse-graining procedure is given as

$$\frac{S_{bb}}{nk_B} \approx \frac{3}{2}n_b + \frac{3}{2}(n_b - 1) - \frac{3}{2}(n_b - 1)\ln\left(\frac{3n_b}{8\pi R_g^2}\right) + \ln\left(\frac{Ve}{\Lambda^3 n}\right). \tag{19}$$

The first two terms in Eq. 19 arise from the kinetic energy and bond potential energy, while the final two terms are the ideal translational and vibrational free energy.

Importantly, there is no contribution in Eq. 19, or in the intramolecular contribution to the potential energy, from the potential or c_0. In both cases the increasing entropy and potential energy with the number of blobs n_b is due solely to the increasing configurational degrees of freedom and not the interaction potential itself.

Another type of entropy of interest is the relative entropy.[5] This function is based on the "information" that is lost during corse-graining, which has to be minimized to optimize the coarse-grained model. Our coarse-grained formalism, based on liquid state theory, is devised to reproduce the correct distribution function, so that the relative entropy between the coarse-grained sites and monomer sites is minimized, and the potential is optimized, without need for any variational approach. This is equivalently to say that the relative entropy, whch is based on the information function that discriminates between coarse-grained configurations sampled in the two levels of representation, is zero.

8. Concluding remarks

We have presented a brief overview of the thermodynamic properties in the Integral Equation Coarse-Graining theory (IECG) for macromolecular systems. The Hyper-Netted Chain closure equation has been adopted to account for the projection of many-body interactions, typical of liquids, into effective pair interactions between coarse-grained sites. Pair distribution functions and effective potentials for the coarse-grained units have been calculated for models that have variable levels of coarse-graining resolution. For isotropic macromolecular liquids the IECG approach can be solved analytically and provides the analytical coarse-grained pair potential input to mesoscale simulations.

From the potential structural and thermodynamic properties of the coarse-grained description can be calculated as a function of the resolution of the model, or level of coarse-graining. Structural and thermodynamc properties, such as pressure and excess free energy, are consistently described across variable levels of coarse-graining. Internal energy and entropy are instead depending on the degree of resolution of the coarse-grained model.

All the quantities are functions of one non-trivial parameter, i.e. the total correlation function at $k \to 0$, c_0, which can be directly determined from the experimental isothermal compressibility of the liquid. This parameter is system specific, depends on the thermodynamic conditions, but does not depend on the resolution of the selected coarse-grained description.

The IECG theory differs from most alternative coarse-graining approches because it does not require performing high-resolution simulations to numerically parametrize the coarse-grained model. Atomistic simulations are used only *a posteriori* as a test of the consistency of the coarse-grained description. In this way the IECG approach is fully predictive of the structural and thermodynamic properties of the system simulated on the length scale of the coarse-graining description and larger.

9. Acknowledgements

This material is based upon work partially supported by the National Science Foundation under Grant No. CHE-1362500. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

10. References

[1] P. Ganguly, and N. F. A. van der Vegt J. Chem. Theory and Computation **9**, 5247(2013).
[2] W. Shinoda, R. Devane, M. L. Klein Mol. Sim. **33**, 27 (2007)
[3] J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth J. Chem. Theory Comput. **9**, 2466 (2013).
[4] P. Ganguly, and N. F. A. van der Vegt J. Chem. Theory Comput. **9**, 5247 (2013).
[5] M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[6] G. Yatsenko, E. J. Sambriski, M. A. Nemirovskaya, and M. Guenza, Phys. Rev. Lett **93** 257803 (2004).
[7] A. J. Clark; J. McCarty; I. Y. Lyubimov; M. G. Guenza, Phys. Rev. Lett., **109** 168301 (2012).
[8] R. Zwanzig *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).
[9] I. Y. Lyubimov, M. G. Guenza J. Chem. Phys. **138**, 12A546 (2013).
[10] A. J. Clark, and M. G. Guenza J. Chem. Phys. **132**, 044902 (2010).
[11] A. J. Clark, J. McCarty, and M. G. Guenza, J. Chem. Phys. **139** 124906 (2013).
[12] J. McCarty, A. Clark, J. Copperman, and M. G. Guenza J. Chem. Phys. **140**, 204913 (2014).
[13] I. Y. Lyubimov, J. McCarty, A. Clark, and M. G. Guenza, J. Phys. Chem. **133**, 094904 (2010).
[14] I. Y. Lyubimov, and M. G. Guenza, Phys. Rev. E, **84**, 031801 (2011).
[15] C. Dalgicdir, O. Sensoy, C. Peter, M. Sayar, J. Chem. Phys. **139**, 234115 (2013).
[16] J. -P. Hansen and I. R. McDonald,*Theory of Simple Liquids*, Academic Press, New York, N.Y. 2nd Ed., 1990.
[17] E. J. Sambriski and M. G. Guenza, Phys. Rev. E **76** 051801 (2007).
[18] G. Yatsenko, E. J. Sambriski, and M. G. Guenza, J. Chem. Phys. **122** 054907 (2005).
[19] E. J. Sambriski, G. Yatsenko, M. A. Nemirovskaya, M. G. Guenza, J. Chem. Phys., **125**, 234902 (2006).
[20] R. L. Henderson Phys. Lett. A **49**, 197 (1974).
[21] D. A. McQuarrie *Statistical Mechanics*; University Science: Sausalito, CA, 2000.
[22] K. S. Schweizer and J. G. Curro, Chemical Physics **149**, 105 (1990).
[23] C. N. Likos, A. Lang, M. Watzlawek and H. Lowen, Phys Rev. E **63**, 031206 (2001).
[24] M. Doi and S. F. Edwards *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1986).
[25] J. McCarty; A. J. Clark; I. Y. Lyubimov; M. G. Guenza, Macromol, **45** 8482 (2012).
[26] J. F. Rudzinski and W. G. Noid, J. Chem. Phys. 135 214101 (2011).