Condensed Matter Physics, 2005, Vol. 8, No. 2(42), pp. 271–280

A combined molecular simulation-molecular theory method applied to a polyatomic molecule in a dense solvent

L.J.D.Frink, M.Martin

Sandia National Laboratories,
Albuquerque, NM 87185

Received November 8, 2004

Simulation of small molecules, polymers, and proteins in dense solvents is an important class of problems both for processing the materials in liquids and for simulation of proteins in physiologically relevant solvent states. However, these simulations are expensive and sampling is inefficient due to the ubiquitous dense solvent. Even in the absence of the dense solvent, rigorous sampling of the configurational space of chain molecules and polypeptides with traditional Metropolis Monte-Carlo, or molecular dynamics is difficult due to long time scales associated with equilibration. In this paper we discuss a series of configurational-bias Monte-Carlo (CBMC) simulations that use a rigorous molecular theory based implicit solvent to achieve an efficient sampling of a chain molecule in a dense liquid solvent. The molecular theory captures solvent packing around the chain molecule as well as the energetic effects of solvent-polymer interactions. It also accounts for entropic effects in the solvent.

Key words: *configurational bias Monte Carlo, liquid state theory*

PACS: *05.20.Jj, 02.70.Uu*

1. Introduction

Simulation of molecules with implicit solvent has become an important computational strategy employed to increase the efficiency of calculations by effectively integrating out the solvent degrees of freedom [1–5]. These approaches have been detailed extensively elsewhere [6,8], and so we give only a brief summary.

One class of models assumes that the solvation free energy is proportional to the solvent accessible surface area (SASA) of the solute molecule of interest [9–11]. The Generalized Born (GB) method extends analytic approximations for solvation energies of ions to macromolecules [2,11,12]. Similarly, the screened coulomb potential

(SCP) models combine estimates for electrostatic energy with SASA based approximation for nonpolar solvation energy [13]. Finally, a 3-dimensional solution to Poisson's equation can be obtained in the vicinity of a solute molecule [8,12,14,15]. This approach is much more computationally demanding than the heuristic approaches cited above, but has the advantage of producing a unique and detailed 3-dimensional picture of solvation as a function of protein conformation. Even so, this approach is based on a continuum approximation for the dense solvent (water).

This letter presents implicit solvent simulations that focus on the short range structural solvation effects ignored in virtually all of the implicit solvent methods cited above. To this end we consider a simple neutral system consisting of a 10-bead tangent site chain immersed in a solvent. We efficiently sample chain conformations using a configurational-bias Monte Carlo (CBMC) approach. We compare the predictions of explicit solvent and 3 different implicit solvent models including a continuum solvent, a pair solvation potential model, and a 3-dimensional (3D) solvation model based on density functional theory (DFT).

2. Theory

In this section we briefly review the statistical mechanics underpinnings for implicit solvent approaches. Our derivation was originally developed in the context of colloidal stability [16], and differs from others that have been presented in the literature [6] in that we consider a semi-grand rather than a canonical ensemble. We begin with the partition function for a system that is held at constant volume, V and temperature, T, and that is closed with respect to the number, N, of solute molecules in the system while it is open with respect to the number, n of solvent molecules in the system. In this case, the solvent chemical potential is μ and the semi-grand partition function is

$$Z_{N\mu VT} = \int \mathrm{d}\mathbf{r}^N \mathrm{e}^{-\beta U_N\left(\mathbf{r}^N\right)} \times \left\{ \sum_{n=0}^{\infty} \int \mathrm{d}\mathbf{r}^n \mathrm{e}^{-\beta u_n(\mathbf{r}^N,\mathbf{r}^n)} \right\}, \tag{1}$$

where U_N denotes the intramolecular interactions of the solute molecule while u_n includes both the solvent-solvent and solvent-solute interactions. The term in the curly brackets above can be identified immediately as the grand canonical partition function for a fluid in an external field (Ξ), and Ξ is related to the grand potential of that inhomogeneous fluid system by $\Omega(r^N;\mu,T) = -\beta^{-1}\ln \Xi_{\mu VT}$. The partition function in equation 1 can be rewritten as

$$Z_{N\mu VT} = \int \mathrm{d}\mathbf{r}^N \mathrm{e}^{-\beta U_N(\mathbf{r}^N)} \mathrm{e}^{-\beta\Omega(\mathbf{r}^N;\mu,T)}, \tag{2}$$

where the configurational integral is now only over the configurations of the solute molecule, and the solvent is captured by the effective potential, Ω. Note that Ω is a free energy and therefore includes both energetic and entropic contributions of the fluid phase. This brief derivation provides a firm statistical mechanical basis for implicit solvation in molecular simulations.

The solvation free energy, Ω is intimately related to the potential of mean force (PMF). Specifically, the PMF is $W(r) = \Omega(\mathbf{r}^N) - \Omega^{\text{ref}}$ where Ω^{ref} is the free energy of a suitable reference system [6]. If chosen properly, the reference system is irrelevant to the simulation because it is a constant shift to the free energy. While the derivation presented above is based on the statistical mechanics of inhomogeneous fluids, a much older derivation of the PMF can be found in the literature. In its earliest description, McMillian and Mayer derived a PMF for electrolytes from a consideration of solutions [7]. In their derivation, the PMF is

$$W(r) = -kT \ln g(r), \tag{3}$$

where $g(r)$ is the radial distribution function. In the McMillian-Mayer treatment, the reference system is two atoms (or ions) separated by $r = \infty$ in the solvent (electrolyte) of interest. That choice leads to the intuitive limit $W(r) \to 0$ (or $g(r) \to 1$) as $r \to \infty$. While this is the intuitive choice for a reference system when one considers the interaction of two monatomic particles in a solvent or electrolyte, it is not the obvious choice for a solvated polyatomic chain. In this case the PMF describes solvent mediated intramolecular interactions. There is no particular interest in computing these PMFs relative to say the sum of the solvation free energies of the individual atoms that make up the chain. Rather, we find that the convenient reference system for the polyatomic case is a pure solvent. Thus our PMF is a measure of the surface excess free energy, Ω^{ex}.

We note that while it is not straightforward to directly apply equation 3 to the simulation of solvated polyatomic molecules, there is a rather obvious approximation that we will consider for comparison with the 3D implicit solvation approach developed here. This approximation is based on interactions of pairs of atoms on the polyatomic chain. To apply this approach, we first compute $g(r)$ for an atomic fluid of Lennard-Jones spheres at the density of the solvent of interest in the chain simulations. We then assume that the total solvation potential for a chain composed of those same Lennard-Jones spheres is

$$\Omega \approx -\frac{1}{2} \sum_i \sum_j \ln g(r_{ij}, \rho_b), \tag{4}$$

where the sum runs over all pairs $j \neq i$, $j \neq i+1$, and $j \neq i-1$, ignoring nearest neighbors on the tangent chain.

Before presenting results for the polyatomic solute, we briefly consider a simpler system where the PMF for both McMillian-Mayer theory and the 3D solvation approach presented here can be computed. Specifically, consider a simple Lennard-Jones fluid. The McMillian-Mayer PMF was computed from equation 3 where the radial distribution function was computed from a Monte-Carlo simulation of the atomic system. The 3D solvation PMF was computed via a series of 3D-DFT calculations where two atoms from the fluid were treated explicitly as surfaces in the calculation. The one-body external field generated by these two atoms was simply based on the same Lennard-Jones potential used to describe fluid interactions. The distance between these two explicit atoms was varied to find $W(r) = \Omega^{\text{ex}}(r) - \Omega^{\text{ex}}(\infty)$.

Figure 1 demonstrates that these two routes produce identical results, and confirms the accuracy of our 3D-DFT numerical implementation.

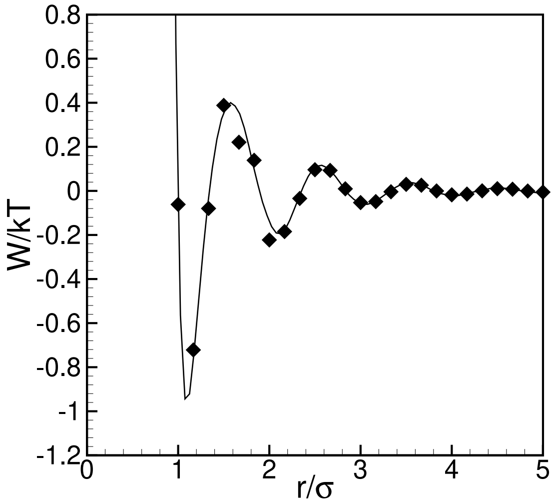

Figure 1. The PMF computed using Monte-Carlo simulation (solid line), and the PMF computed using a series of 3-dimensional DFT calculations (symbols) as described in the main text. The fluid system was an atomic Lennard-Jones fluid with a cutoff distance of $r_c/\sigma = 2.5$, at a temperature of $kT/\epsilon = 0.8$, and a density of $\rho\sigma^3 = 0.75$.

3. CBMC-DFT simulations

This paper presents results for both implicit and explicit solvent simulations. The specific numerical implementations have been detailed elsewhere [17,18]. In all cases, interparticle interactions are described by Lennard-Jones interactions that are cut and shifted at 2σ where σ is the characteristic size of the beads on the tangent chain as well as the size of the solvent particles. Furthermore, the energy parameter for all interactions is $\epsilon = 1$.

The implicit solvent simulations were run on the Cplant distributed memory parallel computer located at Sandia National Labs. The simulations were on an average of 32 processors and required approximately 3 days to complete. While these simulations appear expensive, the true cost of a simulation is measured by both the CPU time and the efficiency of sampling. If trial moves are rarely accepted, poor statistical sampling of the configuration space occurs and a fast algorithm is irrelevant. We return to the question of efficiency in the next section.

In the implicit 3D solvation simulations presented here, a 3D-DFT calculation was performed at each Monte-Carlo move attempt, and 5,000 total Configurational-Bias Monte Carlo (CBMC) moves were performed. Each move consisted of an attempted chain regrowth along with a calculation of the DFT based potentials of mean force on which acceptance criteria were based. CBMC generates trial moves in a biased manner that favors lower energy conformations and then removes this bias by utilizing the ratio of the new and old Rosenbluth weights as the acceptance probability. A coupled-decoupled CBMC algorithm [19] was utilized with 10 nonbond trials coupled to 360 dihedrals trials and decoupled from 1000 trials for the bending angles. The implicit solvation term was ignored during the CBMC growth procedure, computed once the growth was complete, and incorporated into the acceptance criteria as

$$P_{\mathrm{acc}} = \mathrm{Min}\left[1, \frac{\mathrm{e}^{-\beta\Omega(r^N;\mu,T)_{\mathrm{new}}} W_{\mathrm{R}}^{\mathrm{new}}}{\mathrm{e}^{-\beta\Omega(r^N;\mu,T)_{\mathrm{old}}} W_{\mathrm{R}}^{\mathrm{old}}}\right], \tag{5}$$

where W_{R} is the total Rosenbluth weight for the CBMC growth procedure.

Explicit solvent simulations were run for 10,000 CBMC cycles, and required approximately 30 hours on a single processor. An NPT Gibbs ensemble mimic of the semi-grand canonical ensemble was used in these simulations with 2500 solvent molecules. The DFT we use here is a perturbation approach that includes volume exclusion and van der Waals attraction effects. Volume exclusion is treated with Rosenfeld's fundamental measures theory [20,21], and the attractions are treated in a strict mean field approximation. Specifically, we write

$$\begin{aligned}
\Omega[\rho(\mathbf{r})] \quad = \quad & \int \rho(\mathbf{r})\{\ln\rho(\mathbf{r}) - 1\}\mathrm{d}\mathbf{r} + \int \Phi(\{n\})\mathrm{d}\mathbf{r} \\
& + \frac{1}{2}\iint \rho(\mathbf{r})\rho(\mathbf{r}')u_{\mathrm{a}}\left(|\mathbf{r}-\mathbf{r}'|\right)\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\
& + \int \rho(\mathbf{r})[V(\mathbf{r}) - \mu]\mathrm{d}\mathbf{r}\,,
\end{aligned} \tag{6}$$

where the first term is the ideal contribution, the second term is the excess hard sphere term as described by Rosenfeld's FMT theory [21], and the third term gives the attractions where the Weeks-Chandler-Anderson approximation [22] is used to define the attractive potential, u_{a}. The external field, $V(\mathbf{r})$ is given by interactions of a solvent particle with the chain in a given conformation,

$$V(\mathbf{r}) = \sum_{i=1}^{N} u_{\mathrm{cs}}(\mathbf{r} - \mathbf{r_i}), \tag{7}$$

where the "cs" subscript denotes the chain-solvent interaction.

In the DFT calculation, $\Omega[\rho(\mathbf{r})]$ in equation 6 is minimized with respect to the density distribution, $\rho(\mathbf{r})$. Given the equilibrium solution for $\rho(\mathbf{r})$, the grand free energy, Ω is then calculated from equation 6. The numerical methods behind our 3D-DFT implementation have been described elsewhere [23,24]. Here we set the cartesian mesh size to $\sigma/4$. Furthermore, we assume that some distance from the chain, t, the fluid is uniform. Beyond this distance (for $|\mathbf{r}| > t$) we set $\rho(\mathbf{r}) = \rho_b$. At every node, i in this uniform region, we replace the complex DFT residual equations with the discretized residual $\rho_i - \rho_b = 0$ when forming the matrix problem. For the calculations presented in this paper, we have taken this distance to be $t = 1.25\sigma$ from the surface of the molecule. At this distance, we find that the errors in free energy difference with conformation change are less than 30% of the $t \rightarrow \infty$ limit as shown in figure 2. The coarse mesh and the rather small t facilitate DFT solution fast enough O(100 seconds) for coupling with CBMC. One example of a 3D-DFT solution is shown in figure 3 where the dark regions indicate positions of the segments on the chain, and the light regions correspond to a density isosurface that indicates regions of adsorption of the solvent on the chain.

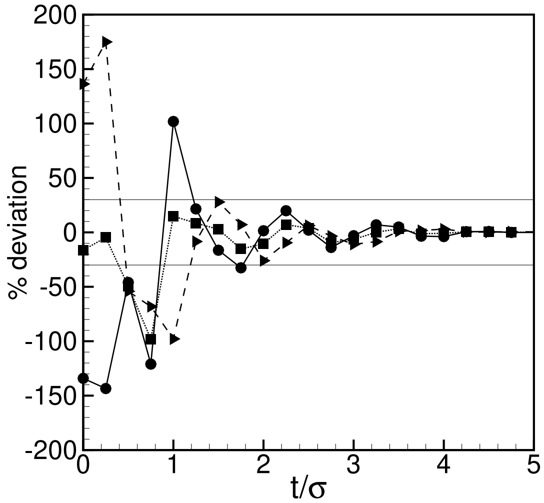

Figure 2. The deviation of the solvation free energy from the $t \rightarrow \infty$ limit for three pairs of configurations.

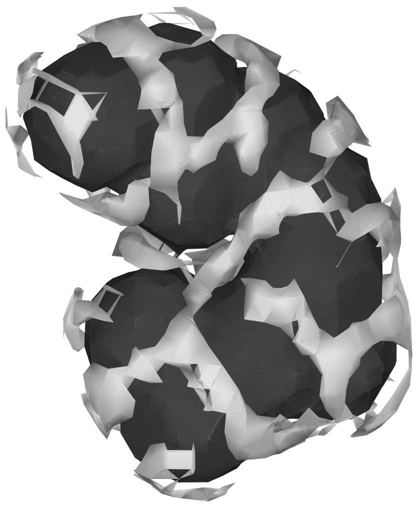

Figure 3. A density isosurface at $\rho\sigma^3 = 1.5$ that indicates regions of adsorption on the chain on particular 3D conformation.

4. Results

We present here simulations that vary the fluid from a low density gas-like state ($\rho\sigma^3 = 0.1$) to a high density liquid-like state ($\rho\sigma^3 = 0.7$) all at $kT/\epsilon = 1$. While the explicit method is reasonably efficient at generating polymer conformations at low solvent density, the liquid-like solvent densities are much more difficult. Table 1 compares the acceptance statistics of the implicit (Im) and explicit (Ex) solvent methods at the liquid-like density of $\rho\sigma^3 = 0.7$. The table shows both the acceptance rates and the ratio of those rates as a measure of the increased efficiency of the implicit solvent method from the perspective of sampling alone. In this dense solvent the implicit solvent method is O(100) times slower than the explicit method from a CPU perspective, yet it is O(1000+) more efficient than the explicit solvent simulation with respect to actual sampling of the interior atom conformations of the molecule. A complete demonstration of the method also requires a comparison of

Table 1. Acceptance statistics for explicit and implicit solvent CBMC calculations of a solvated chain at a high temperature ($kT/\epsilon = 1$) and a high solvent density ($\rho\sigma^3 = 0.7$). Regrowth statistics for various chain lengths attempted in the CBMC simulation are shown.

# grown	% Accepted Explicit Solv	% Accepted Implicit Solv	Efficiency (Im/Ex)
1	11.1	66	5.9
2	2.1	54	26
3	0.42	45	107
4	0.14	42	300
5	0.03	36	1200
6	0	28	∞
7	0	29	∞
8	0	30	∞
9	0	28	∞

physical properties of the macromolecule in explicit and implicit solvent approaches. Figure 4 shows the radius of gyration of the chain as a function of solvent density. In addition to the results for the explicit solvent and the implicit 3D solvation methods, the figure also shows the results for the pair potential approach of equation 4, and a continuum solvent approach that uses the same CBMC-DFT mechanics, but sets the solvation radius to $t = 0$ thus assuming that there is bulk solvent everywhere around the chain. This is equivalent to an incompressible fluid approximation [25].

At low densities where good acceptance statistics are obtained from all methods we find that the 3D solvation method presented here agrees well with the explicit solvent CBMC calculations. At higher densities, there is a deviation in the predictions of the 3D solvation model from the fully explicit model. However, in the dense solvent, there are large error bars associated with the explicit solvent CBMC calculations due to poor acceptance of chain regrowth. Thus we conclude that the implicit solvent approach yields more accurate results than the explicit solvent approach.

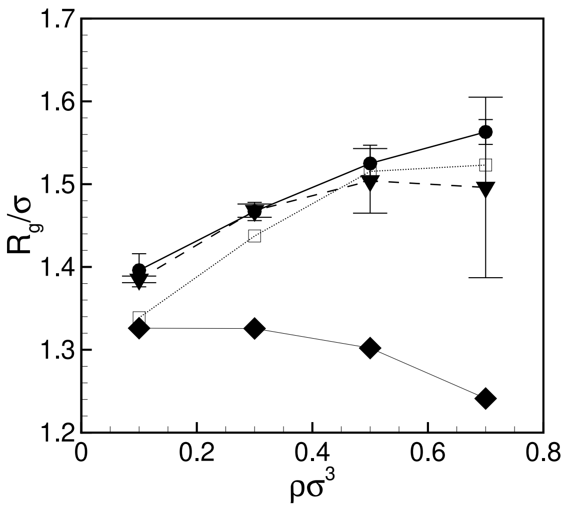

Figure 4. CBMC results for radius of gyration as a function of solvent density at $kT/\epsilon = 1$. The curves show explicit solvent (inverted triangles, wide error bars), a 3D implicit solvent (circles, narrow error bars), pair potential solvation (open squares), and continuum solvent (diamonds) models. For the last two cases, the error bars are smaller than the symbols in the figure.

The pair potential approach of equation 4 deviates from the 3D solvation approach in underpredicting the radius of gyration at all densities. At $\rho\sigma^3 = 0.1$ the pair potential approach matches the results from the continuum fluid model. Since the pair potential does not accurately treat the additive effects due to multiple chain segments interacting with a particular region of fluid, deviation from the full 3D solvation model is to be expected. If anything, this approximation seems to work better than might be expected from such a simple approach. It remains to be seen if this is a general result or is specific to this case where the polymer and fluid are composed of identical segments.

In contrast, the continuum solvent model underpredicts the radius of gyration for all the densities considered here. More importantly, the incompressible fluid model is qualitatively wrong in that it predicts *decreasing* radius of gyration with increasing solvent density.

5. Conclusions

In this work we present the first combination of CBMC with an accurate 3D-DFT based molecular theory approach where the DFT calculations are performed on-the-fly for every trial configuration generated by the CBMC algorithm. We test the approach with simulations of a 10-mer chain in solvents of varying density. These implicit solvent CBMC-DFT calculations exhibit a significantly improved efficiency with respect to sampling chain conformations over explicit solvent CBMC calculations of the same system.

With respect to the prediction of the radius of gyration, the 3D solvation model of the fully coupled CBMC-DFT system agrees well with an explicit solvent simulation. A pair potential approach based on McMillian-Mayer theory deviates from the 3D solvation model results at both high and low solvent density, but is in reasonable quantitative agreement and exhibits the same trends. In contrast, a continuum solvent approximation exhibits the wrong trend for radius of gyration as a function of solvent density. These calculations demonstrate the importance of including solvent packing effects when computing the structure of macromolecules in dense solvents. Future work will center on increasing the complexity of fluid and macromolecule models, and investigating further the application of the pair potential approach to more complex systems including polypeptides in water.

Acknowledgements

L.J.D.F. would like to thank Frank van Swol for collaboration and discussions helpful in defining the problem discussed in this paper. The funding for this work was provided by the US Department of Energy's Office of Science and Office of Industrial Technologies. This work was funded in part or in full by the US Department of Energy's Genomics: GTL program (www.doegenomestolife.org) under project, "Carbon Sequestration in Synechococcus Sp.: From Molecular Machines to Hierarchical Modeling", (www.genomes-to-life.org). Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE–AC04–94AL85000.

References

1. Hassan S.A., Mehler E.L., Zhang D., Weinstein H., PROTEINS: Structure, Function, and Genetics, 2003, **51**, 109–125.
2. Wang T., Wade R.C., PROTEINS:, Structure, Function, and Genetics, 2003, **50**, 158–169,
3. Potter M.J., Kirchhoff P.D., Carlson H.A., Mccammon J.A., J. Comp. Chem., 1999, **20**, 956–970.
4. Yi Shen M., Freed K.F., Biophys. J., 2002, **82**, 1791–1808.
5. Zhou R., PROTEINS: Structure, Function, and Genetics, 2003, **53**, 148–161.
6. Roux B., Simonson T., Biophysical Chemistry, 1999, **78**, 1–20.
7. McMillan W.G., Mayer J.E., J. Chem. Phys., 1945, **13**, 276.
8. Simonson T., Current Opinion in Structural Biology, 2001, **11**, 243–252.
9. Fraternali F., van Gunsteren W.F., J. Mol. Biol., 1996, **256**, 939–948.
10. Ooi T., Oobatake M., Nemethy G., Scheraga H.A., Proc. Natl. Acad. Sci. USA, 1996, **84**, 3086–3090.
11. Still W.C., Tempczyk A., Hawley R.C., Hendrickson T., J. Am. Chem. Soc., 1990, **112**, 6127–6129.
12. Laurent D., Luo R., Gilson M.K., J. Comp. Chem., 2000, **21**, 295–309.
13. Hassan S.A., Guarnieri A., Mehler E.L., J. Phys. Chem. B, 2000, **104**, 6478–6489.

14. Gilson M.K., Davis M.E., Luty B.A., McCammon J.A., J. Phys. Chem., 1993, **97**, 3591–3600.
15. Borgis D., Levy N., Journal of Chemical Physics, 2003, **119**, 3516–3528.
16. Frink L.J.D., van Swol F., J. of Chem. Phys., 1994, **100**, 9106–9116.
17. http://towhee.sourceforge.net
18. http://software.sandia.gov/DFTfluids
19. Martin M., Siepmann I., J. Chem. Phys. B, 1999, **103**, 4508–4517.
20. Rosenfeld Y., Phys. Rev. Lett., 1989, **63**, 980.
21. Rosenfeld Y., Schmidt M., Lowen H., Tarazona P., J. Phys. Cond. Matt., 1996, **8**, L577.
22. Weeks J.D., Chandler D., Anderson H.C., J. Chem. Phys., 1971, **54**, 5237.
23. Frink L.J.D., Salinger A.G., J. Comp. Phys., 2000, **159**, 407–424.
24. Frink L.J.D., Salinger A.G., Sears M.P., Weinhold J.D., Frischknecht A.L.., J. Phys. Cond. Matter, 2003, **14**, 12167–12187.
25. Jendrejack R.M., de Pablo J.-J., Graham M.D., J. Non-Newt. Fluid Mech., 2002, **108**, 123–142.

Застосування об'єднаного методу молекулярна симуляція-молекулярна теорія до поліатомних молекул в густому розчиннику

Л.Дж.Фрінк, М.Мартін

Національна лабораторія Сандії, Албукерк

Отримано 8 листопада 2004 р.

Комп'ютерне моделювання малих молекул, полімерів та білків у густому розчиннику представляє важливий клас проблем як для процесів обробки матеріалів у рідинах, так і для моделювання білків у фізіологічних розчинниках. Разом з тим, такі симуляції є досить затратними, а вибірка простору не є ефективною через високу густину розчинника. Навіть при відсутності густого розчинника строга вибірка конфігураційного простору ланцюгових молекул і поліпептидів у рамках традиційного Монте Карло по рецепту Метрополіса або молекулярної динаміки є проблематичною через довгі часові масштаби, пов'язані з встановленням рівноваги. В даній роботі ми обговорюємо кілька конфігураційно покращених Монте Карло симуляцій, які використовують строгу молекулярну теорію розчинника, щоб провести ефективну вибірку ланцюгових молекул в густому рідкому розчиннику. Молекулярна теорія враховує упаковку розчинника навколо ланцюгової молекули, а також енергетичні ефекти взаємодії розчинник-полімер. Враховуються також ентропійні ефекти в розчиннику.

Ключові слова: *конфігураційно покращені Монте Карло, теорія рідкого стану*

PACS: *05.20.Jj, 02.70.Uu*