Research Article

Accurate Calculation of the Density of States near the Ground-State Energy of the Peptides Met-Enkephalin and (Alanine)₅ with the Wang-Landau Method: Lessons Learned

Priya Singh¹ and Pradipta Bandyopadhyay²

¹ School of Physical Science, Jawaharlal Nehru University, New Delhi 110067, India

² School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Correspondence should be addressed to Pradipta Bandyopadhyay, praban07@gmail.com

Received 16 December 2011; Revised 7 March 2012; Accepted 7 March 2012

Academic Editor: Jan Petter Hansen

Copyright © 2012 P. Singh and P. Bandyopadhyay. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Wang-Landau method estimates the relative density of states (DOS) by performing random walk in energy space. However, estimation of the DOS near the ground state minimum is highly challenging because of the dearth of states in the low-energy region compared to that at the high-energy region. Ideally the derivative of the logarithm of the DOS with respect to energy, which is proportional to the inverse of temperature, should become steeper with decrease in energy. However, in actual estimation of the DOS for molecular systems, it is nontrivial to achieve this. In the current work, the accuracy of the Wang-Landau method in estimating the DOS near the ground state minimum is investigated for two peptides, Met-enkephalin and (Alanine)₅. It has been found that the steepness of the DOS can be achieved if the correct ground state energy is found, the bin used to discretize the energy space is extremely small (0.1 kcal/mol was used in the current case) and the energy range used to estimate the DOS is small. The findings of this work can help in devising new protocols for calculating the DOS with high accuracy near the ground state minimum for molecular systems.

1. Introduction

The density of states (DOS) is a fundamental quantity to understand the thermodynamic properties of a system. For any molecular system except for the smallest ones, calculation of the DOS is an extremely difficult problem because of the high dimensionality of the system. The Wang-Landau (WL) technique is one popular method to estimate the DOS [1]. This method performs a random walk in the energy space and estimates the DOS based on the flat histogram of energy obtained from the random walk. Initially the WL method was used for discrete systems such as Ising spin model and Potts model [1–3]. Later it has been extended to molecular systems. Till now several applications of the WL method for protein systems have been reported. For instance, the dimerization of glycophorin A, stability of proteins inside a cage, and thermodynamic properties of peptides have been investigated by the WL method [4–6].

Recently Singh et al. have done a systematic investigation on the performance of two flavours of the WL method for two small peptides [7]. However, one practical issue with the WL method is the poor quality of the DOS at the low-energy region of the system. As the DOS increases extremely rapidly with the increase of energy, low-energy estimation of the DOS is often done poorly by the WL method. Another way of saying that is that the number of states in the low-energy region is much less than that in the high-energy region and the WL method spends more time in estimating the DOS at high-energy region rendering the low-energy DOS unreliable. The current work investigates the possible ways to increase the accuracy of the low-energy DOS for small peptides. Recently the low-energy sampling problem have been investigated by Baretin and Sibani for the spin glass system with a combination of the lid method and entropic sampling, which is similar to the WL sampling [8]. The lid restricts the sampling to go beyond a specified energy

range by defining an impenetrable barrier as the upper limit of energy for sampling. The metric, which can be used to ascertain the quality of sampling near the ground state, is the derivative of the logarithm of the DOS with respect to the energy, which is proportional to the inverse of the microcanonical temperature. This derivative should become steeper with the decrease of energy and will go to infinity at the ground-state energy. In the investigation of Barretin et al., they have found that the steepness of this derivative with respect to the energy increases with the decreasing value of the lid.

In the current work, we have thoroughly investigated a continuous system, Met-enkephalin, which is a peptide containing five amino acids with the sequence Tyr-Gly-Gly-Phe-Met. Met-enkephalin is often used as a model system for testing performance of new simulation techniques [9]. We have explored the quality of the DOS by determining it over different energy ranges, similar to the lid method used by Barretin et al. We have also investigated the dependence of the DOS on other simulation parameters such as the bin size used to discretize the energy range. We have checked how the DOS varies as the lowest energy range for sampling was gradually moved from the ground state to higher energy. We have also investigated the low-energy DOS for another peptide, (Alanine)₅ by gradually increasing the lowest energy range as done for Met-enkephalin. From the results of this work, we have found that unless the system ground state or a structure very close to the ground state is found it is difficult to get a steep derivative of the log of the DOS. Even for an error of 1 kcal/mol in finding the ground state, the slope of the derivative deviates significantly. The energy range used, that is, the lid value used to contain the DOS, is also important to get proper sampling of low energy states. A very fine bin size (of the order of 0.1 kcal/mol) is required to get high accuracy estimation of the DOS near the ground state energy.

This paper is arranged in the following way. Next section discusses the WL method. Section three describes the details of simulation. Results and discussions are given in section four. The paper ends with a conclusion.

2. Method

The essence of the WL method is given briefly in this section. For details, the readers are referred to the original literature [1]. The WL method aims to estimate the DOS based on a random walk in energy space. At first the energy range over which the DOS is to be estimated is discretized into bins. The WL method starts the estimation with a guess of the DOS for each energy bin. It does a random walk in energy space using a Monte Carlo approach, where the acceptance/rejection of states is given by

$$\text{acc}(E_1 \rightarrow E_2) = \min\left(1, \frac{g(E_1)}{g(E_2)}\right), \quad (1)$$

where $g(E)$ is the current value of the DOS for the energy bin E . For each entry to a bin E , the $g(E)$ is updated as $g(E) * f$, where f is a convergence parameter and histogram

of energy, $H(E)$ is updated as $H(E) = H(E) + 1$. For a fixed value of f , when the flatness of energy histogram is achieved (based on a flatness criterion), f value is reduced and $H(E)$ is set to zero. The whole process is repeated, until f value becomes smaller than a specific value of f (the convergence threshold). There are other implementations of the WL method. One popular implementation is the t^{-1} method, where the flatness of histogram is not used to reduce the value of f [10]. In this method the f value is reduced as the inverse of the Monte Carlo step t . It is to be noted that there are other techniques, such as multicanonical sampling [9], parallel tempering (PT) [11], and replica exchange molecular dynamics (REMD) [12], which also aim to explore the energy landscape efficiently as done by the WL method. In the multicanonical sampling, a non-Boltzmann weight factor is obtained by an iterative procedure, which is different from the iterations of the WL method. If a production simulation is done with the fixed weight factors (obtained from the iterative procedure), then the multicanonical and WL method become identical as far as the production runs are concerned. Method like PT and REMD can be thought of doing simulation with random walk in temperature space. There has been comparison between the WL and the PT method for Met-enkephalin but for larger peptides, systematic comparison is necessary to compare the efficacy of these algorithms [13]. As methods like PT and REMD use Boltzmann weight factor, no iterative method to determine the non-Boltzmann weight factor is necessary. However, the advantage of the WL method is that the DOS is obtained directly in the iterative procedure, unlike the PT or REMD method, where the DOS can be calculated after postprocessing the obtained trajectories.

3. Details of the Simulation

Both the peptides Met-enkephalin and (Alanine)₅ were represented with the ECEP/3 force field, which has electrostatics, Van der Waals, hydrogen bonding, and torsional terms [14]. The WL algorithm has been implemented in the SMMP package [15]. No effect of solvent was taken into account in this investigation. Only torsional moves were given to the peptide by giving random values between +180 and -180 degrees. The flatness criterion used to check histogram flatness was 80% flatness of the histogram. As the convergence at a predefined value of f does not necessarily mean the DOS has converged to the level of accuracy wanted, we have cross-checked the convergence by continuing the simulation by keeping the value of f same but updating the $g(E)$. When the $g(E)$ values do not change significantly during this update process, it is assumed that $g(E)$ has converged. The ground-state energy of the Met-enkephalin was taken as -12.4 kcal/mol from a previous investigation [16]. The putative ground-state energy of (Alanine)₅ was determined by doing initial set of WL calculation, where lowest energy range was decreased continuously to search for the lowest energy structure. The error bars in the DOS were taken as the standard deviations of the DOS calculated over multiple trajectories.

4. Results and Discussion

4.1. Met-Enkephalin. At first we investigated the difference between the two flavours of the WL method, the original WL implementation and the t^{-1} implementation, for our problem. In a previous work it was found that the t^{-1} implementation gave results with smaller standard deviation of the DOS, average energy, and specific heat at constant volume (C_v) compared to the original implementation of the WL method for the two peptides, Met-enkephalin, and Trp-cage considered in that work [7]. However, that investigation was done for a wide energy range, while the current work focuses only on the low energy states. Hence, we have compared the performance of the t^{-1} and the original WL implementation by running both for a low energy segment of 5 kcal/mol for Met-enkephalin with 0.5 kcal/mol bin size. CPU time of about 38 hours was taken by the t^{-1} method, while the original WL implementation took about 6 hours for the f value 1×10^{-8} as the convergence threshold in a Xeon@2.7 GHz machine. We have found that the DOS calculated by the original WL method converges smoothly with the lowering of the convergence parameter f in each iteration, while in the t^{-1} implementation, the DOS varies widely with the f parameter. This is shown in Figures 1(a) and 1(b) for the original WL and the t^{-1} implementations, respectively. It can be seen that in the original WL method, the DOS for the four f values are essentially same, while for the t^{-1} case, the DOS decreased on going from the f value of 5×10^{-7} to 1×10^{-7} . For the f value of 5×10^{-8} it decreased even more but increased for the f value of 1×10^{-8} . Since, the t^{-1} implementation does not need flatness of the histogram of energy to reduce the f value, the DOS can vary substantially between different values of f . Moreover, for small system like Met-enkephalin considered here, getting flat histogram at each value of f was usually fast. Based on these observations, original implementation of the WL method was used in this work. However, it is to be noted that a full comparison between the t^{-1} and the original WL method is beyond the scope of the current work and hence not attempted. Hence, no general conclusions regarding comparison between the t^{-1} and WL method should be inferred from our results. In the rest of this paper, WL method would imply the original implementation of the WL method.

Next we investigated the dependence of the DOS on the bin size. Figure 2 shows the DOS calculated for the energy range -12.4 kcal/mol to -7.5 kcal/mol with three different bin sizes, 0.1, 0.5, and 1.0 kcal/mol. The starting values in the plot are different for each curve, since the midpoint of the first bin (which is taken as the first point) differs for these three calculations. Five independent trajectories were run for each case. It can be seen from the figure that near the ground state, the steepness of $\log(\text{DOS})$ clearly differs for these three bin sizes. The DOS calculated with 0.1 kcal/mol bin size has the steepest $\log(\text{DOS})$, followed by the DOS calculated by the 0.5 and 1.0 kcal/mol bin size. Also, there is difference of the DOS values for these plots and in the error bars. Essentially the sharp rise of DOS could not be captured with the 0.5 and 1.0 kcal/mol

FIGURE 1: (a) The variation of the DOS for Met-enkephalin for different f values (see text) for the original implementation of the Wang-Landau algorithm. (b) The variation of the DOS for Met-enkephalin for different f values (see text) for the t^{-1} implementation of the Wang-Landau algorithm.

bin size calculations. These two calculations show lower value of DOS for this energy range. However, the relative values of DOS in each bin are similar for the three cases shown, which suggests that the thermodynamic quantities calculated with these bin sizes will have less variation as compared to the variation among the DOS calculated. This was checked by calculating the average energy at 50 K. The difference in the average energy value differs only about 4% between the 0.1 and 1.0 kcal/mol bin sizes. Although it will be interesting to further investigate the bin size dependence on other thermodynamic properties, this was not investigated in detail in the current work. It is clear that with a standard convergence criterion, the 0.1 kcal/mol bin size calculation

FIGURE 2: The DOS calculated for Met-enkephalin from its ground state energy with three bin sizes.

FIGURE 3: The DOS calculated for Met-enkephalin by shifting the lowest energy range of energy. All the calculations are done with 0.1 kcal/mol bin size.

captures the DOS at low-energy region most efficiently. It is to be noted that going from 1.0 to 0.1 kcal/mol bin size makes the DOS determination about twice more expensive.

One question is how the steepness of the derivative of the $\log(\text{DOS})$ with respect to energy changes as we move the lowest energy range gradually from the ground state minimum to higher energy value. This might be important in general investigation of protein systems, where knowledge of global minimum is not always present. Figure 3 shows how the steepness of the $\log_e(\text{DOS})$ changes as a function of shifting the lowest energy state. All the calculations were done for 1 kcal/mol range with 0.1 kcal/mol as the bin width. Five different trajectories were run to calculate the error

FIGURE 4: (a) The DOS calculated for Met-enkephalin for 0.9 kcal/mol energy range with 0.1 kcal/mol bin size. (b) The DOS calculated for Met-enkephalin for 4.9 kcal/mol energy range with 0.1 kcal/mol bin size.

bars for each energy range. As we shift the lowest energy range from the known lowest energy state of -12.4 , the steepness monotonically decreases. We also found that the DOS has maximum increase for the first energy bin of 0.1 kcal/mol from the ground state. The rate of change of the DOS with energy decreases as we sample higher energy bins. The findings from this plot indicate that if the ground-state energy is not known for the system, it will be difficult to get steep derivative of the entropy with respect to the energy.

Finally we have focused on how the $\log_e(\text{DOS})$ changes as we shift the maximum of the energy range (or the lid value in the language of reference [8]) by keeping the minimum of the energy range (the ground state) intact. Figures 4(a) and 4(b) show the DOS calculated with maximum energy range -11.5 kcal/mol and -7.5 kcal/mol, respectively with bin size 0.1 kcal/mol. Results for five different calculations are shown in the figures. From the current data of five trajectories, we have fitted the $\log_e(\text{DOS})$ for first 0.9 kcal/mol energy to a straight line for both the plots. It was found from the fitted

FIGURE 5: The DOS calculated for (Alanine)₅ by shifting the lowest energy range of energy. All the calculations are done with 0.1 kcal/mol bin size.

straight lines that calculations done with the smaller energy range gave a slightly higher gradient than that came from the larger energy range.

4.2. (Alanine)₅. To reconfirm our findings for Met-Enkephalin, we have performed the estimation of the DOS at low energy for another system, (Alanine)₅. The steepness of the DOS for 1 kcal/mol range by shifting the energy range is shown in Figure 5. It can be seen that the plot has the similar variation of steepness as was found for Met-Enkephalin. Based on the large value of steepness for the 4.5 to 5.5 kcal/mol energy range, it is very likely that simulation has reached structures close to the ground state-energy minimum.

From the results of this work, we can qualitative say that to get high-accuracy DOS near the ground-state energy of small peptides, the knowledge of ground state minimum or a structure very close to the ground state is necessary. Moreover, the calculation should be done with very fine bin size. The energy range has an effect on the steepness of the derivative but the difference is not significant for Met-enkephalin. The conclusions that came from this work should be applicable to other molecular systems as well. However, the results of this work should be taken as preliminary, as investigations on more systems are required to get a quantitative understanding.

5. Conclusion

Poor sampling of the DOS near the ground state of a molecular system by the WL method is well known. In the current work, we have investigated the accuracy of the WL method to get highly accurate DOS for a peptide, Met-Enkephalin, near its ground-state minimum. The findings for Met-enkephalin was reconfirmed by calculating the DOS for (Alanine)₅. Our results indicate that it is necessary to

know the ground-state minimum or a structure very close to the ground-state minimum of the system and only stringent convergence criteria can increase the sampling efficiency of the DOS at low-energy region. The dependence on bin size, energy range to use for the DOS estimation has also been investigated in this work. However, it is to be mentioned that in general for larger peptides and proteins, knowledge of ground-state energy minimum is rarely known. The steepness of the DOS can be used to check if the system is close to the ground-state minimum albeit with increasing computational cost.

References

- [1] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letters*, vol. 86, no. 10, pp. 2050–2053, 2001.
- [2] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and Ph. Dugourd, "Performances of Wang-Landau algorithms for continuous systems," *Physical Review E*, vol. 73, no. 5, Article ID 056704, 11 pages, 2006.
- [3] A. N. Morozov and S. H. Lin, "Accuracy and convergence of the Wang-Landau sampling algorithm," *Physical Review E*, vol. 76, no. 2, Article ID 026701, 7 pages, 2007.
- [4] N. Rathore, T. A. Knotts, and J. J. de Pablo, "Density of states simulations of proteins," *Journal of Chemical Physics*, vol. 118, no. 9, pp. 4285–4290, 2003.
- [5] P. Ojeda, M. E. Garcia, A. Londoño, and N. Y. Chen, "Monte carlo simulations of proteins in cages: influence of confinement on the stability of intermediate states," *Biophysical Journal*, vol. 96, no. 3, pp. 1076–1082, 2009.
- [6] C. Gervais, T. Wust, D. P. Landau, and Y. Xu, "Application of the Wang-Landau algorithm to the dimerization of glycoporphin A," *Journal of Chemical Physics*, vol. 130, no. 21, Article ID 215106, 7 pages, 2009.
- [7] P. Singh, S. K. Sarkar, and P. Bandyopadhyay, "Understanding the applicability and limitations of Wang-Landau method for biomolecules: Met-enkephalin and Trp-cage," *Chemical Physics Letters*, vol. 514, no. 4–6, pp. 357–361, 2011.
- [8] D. Baretin and P. Sibani, "Entropic algorithms and the lid method as exploration tools for complex landscapes," *Physical Review E*, vol. 84, no. 3, Article ID 036706, 7 pages, 2011.
- [9] U. H. E. Hansmann and Y. Okamoto, "Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem," *Journal of Computational Chemistry*, vol. 14, no. 11, pp. 1333–1338, 1993.
- [10] R. E. Belardinelli and V. D. Pereyra, "Fast algorithm to calculate density of states," *Physical Review E*, vol. 75, no. 4, Article ID 046701, 5 pages, 2007.
- [11] U. H. E. Hansmann, "Parallel tempering algorithm for conformational studies of biological molecules," *Chemical Physics Letters*, vol. 281, no. 1–3, pp. 140–150, 1997.
- [12] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters*, vol. 314, no. 1–2, pp. 141–151, 1999.
- [13] C. Junghans and U. H. E. Hansmann, "Numerical comparison of Wang-Landau sampling and parallel tempering for met-enkephalin," *International Journal of Modern Physics C*, vol. 17, no. 6, pp. 817–824, 2006.
- [14] G. Nemethy, K. D. Gibson, K. A. Palmer et al., "Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3

- algorithm, with application to proline-containing peptides,” *Journal of Physical Chemistry*, vol. 96, no. 15, pp. 6472–6484, 1992.
- [15] F. Eisenmenger, U. H. E. Hansmann, S. Hayryan, and C. K. Hu, “[SMMP] A modern package for simulation of proteins,” *Computer Physics Communications*, vol. 138, no. 2, pp. 192–212, 2001.
- [16] L. Zhan, J. Z. Y. Chen, and W. K. Liu, “Conformational study of Met-enkephalin based on the ECEPP force fields,” *Biophysical Journal*, vol. 91, no. 7, pp. 2399–2404, 2006.