Versatile approach to access the low temperature thermodynamics of lattice polymers and proteins

Thomas Wüst and David P. Landau
Center for Simulational Physics, The University of Georgia, Athens, GA 30602, USA

We show that Wang-Landau sampling, combined with suitable Monte Carlo trial moves, provides a powerful method for both the ground state search and the determination of the density of states for the hydrophobic-polar (HP) protein model and the interacting self-avoiding walk (ISAW) model for homopolymers. We obtained accurate estimates of thermodynamic quantities for HP sequences with \( > 100 \) monomers and for ISAWs up to \( > 500 \) monomers. Our procedure possesses an intrinsic simplicity and overcomes the limitations inherent in more tailored approaches making it interesting for a broad range of protein and polymer models.

PACS numbers: 87.15.ak,05.10.Ln,05.70.Fh,36.20.Ey

Coarse-grained polymer and protein models play an important role in understanding physical phenomena such as e.g. protein folding or the phase behavior of flexible macromolecules, and Monte Carlo simulation methods have become an indispensable tool for the study of such models [1]. One of the most prominent examples is the hydrophobic-polar (HP) lattice model [2], where the protein is represented as a self-avoiding chain of beads (the amino acid residues) on a lattice. The amino acids are divided into two classes - hydrophobic (H) and polar (P) - and an attractive interaction \( \epsilon \) acts between non-bonded neighboring H residues mimicking the hydrophobic force (\( \epsilon_{HH} = -1, \epsilon_{HP,PP} = 0 \)). The special case of a chain consisting entirely of H residues (homopolymer), the interacting self-avoiding walk (ISAW), is an important model for studying the statistical physics of polymers [3, 4].

Despite their formal simplicity and minimalistic framework, lattice models represent a challenging testing ground for computational methods because of their complex energy landscapes, conformational constraints and dense packings. The HP model has become a standard work, lattice models represent a challenging testing ground for computational methods because of their complex energy landscapes, conformational constraints and dense packings. The HP model has become a standard for assessing the efficiency of folding algorithms, and numerous - some very tailored - conformational ground state search strategies have been proposed, see e.g. [5, 6] and references therein.

More revealing than algorithms that merely search for low energy states, however, are methods which target the sampling of the entire conformation and energy space. They can provide an estimate of the density of states (DOS) \( g(E) \) of energy \( E \) which, in turn, gives access to thermodynamic properties (e.g. internal energy, specific heat, entropy or free energy) of a system at any temperature [7]. Only a few attempts have been undertaken to this end for the HP model, the most notable approaches being multi-self overlap ensemble Monte Carlo (MSOE) [10], multicanonical chain growth (MCCG) [11], and equi-energy sampling (EES) [12]. Although inventive and powerful, these methods also suffer from severe limitations: Large memory needs for keeping track of all sampled conformations (construction of microcanonical ensembles) (EES); (quasi-)statics, i.e. one bead of the chain is permanently fixed in space (MCCG); or the necessity to treat an expanded ensemble resulting in a large amount of computer time spent in sampling non-physical space (MSOE). Such restrictions can become increasingly important for more complex biological setups such as multi-chain systems or protein folding in heterogeneous environments (e.g. membranes) [13].

In this Letter we show that a generic algorithm - Wang-Landau sampling [14] - together with appropriate Monte Carlo trial moves, provides a powerful, yet flexible methodology for the simulation of HP-like lattice proteins and homopolymers that does not suffer from any of the above limitations.

The key to our approach is the combination of two “non-traditional” Monte Carlo trial moves, which complement one another extremely well, namely pull moves [6] and bond-rebridging moves [15], see Fig. 1. Originally proposed for the square and simple cubic lattices only, here we extended both types of trial moves to any \( n \)-dimensional space (\( n \geq 2 \)). (i) Pull moves [6] allow for the close-fitting motion of a polymer chain within a confining environment by “pulling” portions of the polymer to unoccupied neighboring sites. Pull moves are reversible and fulfill ergodicity; moreover, they provide a good balance between local and global conformational changes, as well as a “natural” dynamics of folding. These features are important to an algorithm that seeks to sample the entire conformational space such as Wang-Landau sampling and thus requires an efficient move for the continuous folding and unfolding of the polymer. (ii) Bond-rebridging moves [15]: Trial moves which displace monomers become ineffective for very compact conformations where few unoccupied neighboring sites remain available. In contrast, bond-rebridging moves allow the polymer to change its conformation even at highest den-
errors were always calculated from 15 independent DOS samples for each energy subinterval. This allows for the calculation of standard errors associated with the DOS estimate.

The knowledge of the exact energy range is essential in the WL algorithm for the examination of the flatness of the histogram. Without prior knowledge of the energy range, the WL algorithm might fail to converge to the correct distribution. This is especially true for systems with a large number of energy levels, such as the HP model, where the energy range can be very wide.

Our approach to WL sampling is to first apply energy-guided sequential sampling (FRESS) to locate the ground state, and then use the WL algorithm to explore the conformational space around the ground state. This two-step approach has several advantages:

1. It allows for the exploration of the conformational space around the ground state, which can be difficult to access using traditional WL sampling.
2. It reduces the number of pull moves required, which can be computationally expensive.
3. It allows for the identification of metastable states, which are important for understanding the dynamics of the system.

Overall, our approach provides a robust and efficient method for WL sampling, which can be applied to a wide range of systems.
in the energy range \([-57, 0]\), within a single simulation, and with very high accuracy. It was nonetheless not possible to determine the relative magnitudes of the ground state \((E = -58)\) and 1st excited state \((E = -57)\) DOS with high precision. Fig. 2 (bottom) displays the specific heat for sequence 3D103, manifesting a peak at \(T \approx 0.51\) and a shoulder at \(T \approx 0.27\). We do not observe an additional peak in \(C_V\) at very low temperatures, contrarily to Ref. 11. However, since only conformations with energies down to \(E = -56\) were found and the estimated errors near that peak were rather large, we think that this finding was an artefact of insufficient sampling. Indeed, our \(C_V\) curves indicate that the folding transitions from unstructured globular conformations to the ground states are rather weak for both sequences - despite the difficulty in sampling their low energy regimes. By means of multicanonical sampling given our DOS estimates, we obtained the radius of gyration \(R_g\) and the Jaccard index \(q = \max\{c_{s,g}/(c_s + c_g)\}\) which measures the structural similarity between any conformation \(s\) and the ground states \(g\) of an HP sequence 17. \(c_{s,g}\) denotes the number of common (native) H-H contacts between \(s\) and \(g\), and \(c_s, c_g\) are the numbers of H-H contacts found only in \(s\) and \(g\), respectively (the maximum stems from the possible degeneracy of ground states). Fig. 2 also shows the averages \(\langle R_g \rangle\) and \(\langle q \rangle\) for sequences 2D100b and 3D103 and illustrates the complementary information in these two quantities. While \(\langle R_g \rangle\) indicates the coil-to-globule collapse, \(\langle q \rangle\) identifies the folding transition to the native state and thus may serve as a suitable structural order parameter for these kind of systems. In case of sequence 3D103, the ground state \((E = -58)\) was excluded from the sampling (due to the difficulty in finding this state) which results in \(\langle q \rangle\) saturating at a rather low value \((< 0.3)\) for \(T \rightarrow 0\). This manifests the still large structural differences between conformations with \(E = -57\) and the ground state.

TABLE I compares various methods in finding low energy conformations and, if available, the DOS for common benchmark HP sequences. We also included results from methods which were focused on the low temperature range only, i.e. FRESS 8 and the variants of PERM (pruned-enriched Rosenbluth method) 2 and hence do not provide the entire DOS. Except for the longest sequence (3D136), we could confirm all minimum energy states found previously. The superiority of FRESS for this sequence is the result of various “efficiency enhancements” towards low energy states (see 8) which become obviously the more effective the longer the chain length. However, they do not permit anymore a correct sampling, let alone an estimation of the DOS.

As a second test of performance, we applied our method to the interacting self-avoiding walk (ISAW) representing a homopolymer with nearest-neighbor attraction \((\epsilon = -1)\) on the square (sq, 2D) and simple cubic (sc, 3D) lattice. Unraveling the “phase transition” behavior of flexible macromolecules in the thermodynamic limit \((N \rightarrow \infty)\) by means of simple (lattice) models - such as e.g. the ISAW, the bond-fluctuation model or systems in the continuum - has been a challenge for decades 4, 15. Although the \(\theta\) point (coil-globule transition) could be investigated well for polymer chains with \(N \geq 10,000\) monomers, our understanding of the ISAW at very low temperatures remains elusive. Due to the very dense packings resulting for this model, accurate estimates of thermodynamic quantities below \(T_\theta\) are difficult to obtain. In the most recent computational studies, only chains with \(N \leq 125\) in 3D (multicanonical chain-growth 19) and \(N \leq 300\) in 2D (adaptive WL sampling with reptation, but without the lowest energy states 16) could be investigated.

With our generic approach we were able to obtain accurate DOS estimates for ISAWs up to chain lengths \(N = 400\) (2D) and \(N = 512\) (3D) over the entire energy range (including ground states) and we could then determine reliable thermodynamic quantities even at lowest temperatures \((T \rightarrow 0)\), see Fig. 3. The possibility to compare the specific heat \(C_V(T)\) for various system sizes up to these chain lengths allowed us to draw interesting conclusions which apply for the ISAW on both the sq and sc lattice: At high \(T\), the collapse transition \((\theta \text{ point})\) indicates a clear phase transition manifested by cooperative structural rearrangements from the coil to the globular state and \(C_V(T_\theta) \rightarrow \infty\) for \(N \rightarrow \infty\). At very low \(T\), a pronounced peak appears due to various ground state excitations (here, the ground states form either regular squares or cubes). These excitations are induced by local rearrangements at the surface and therefore, the magnitude of the peak decreases systematically with chain length (2D) or becomes constant to within statistical errors bars (3D). The breaking up of the ground

| Seq. | WLS | EES | MCCG | MSOE | FRESS | PERM |
|------|-----|-----|------|------|-------|------|
| 2D100a | -48 | -48 | -47 | -48 | -48 |
| 2D100b | -50 | -49 | -50 | -50 | -50 |
| 3D88 | -72 (-69) | -72 | -69 |
| 3D103 | -58 (-57) | -56 | -57 | -55 |
| 3D124 | -75 (-74) | -75 | -71 |
| 3D136 | -82 (-81) | -83 | -80 |

*See e.g. dr2u2ldbdudbdflfdflflf2f2dr2dbulub2dr fual2dfuufu fdlbd2u2fud2b2ulblbf s2f3 unal2bd f1f3 fddfd bd (encoded as sequence left[l], right[r], up[u], down[d], forward[f], backward[b]).

*DOS not attained.
the overall thermodynamic scenario remains the same

(\textit{such a transition}. Whereas a variation of chain length
(i.e. symmetry breaking) which would be necessary for
the model does not permit a further cooperative effect
els \cite{4, 18}. Once in the globular phase, the rigidity of
observed for other lattice and off-lattice polymer mod-
for \(T \rightarrow \infty\). Most notably, the ISAW on the sq/sc lat-
ture ranges of such models. A major advantage of our
method is that it remains rather simple and flexible be-
side its proven performance which has not been achieved
earlier, by more elaborate attempts \cite{10, 12, 19}. These
features make it readily applicable to the study of com-
plex biological phenomena such as e.g. protein aggrega-
gation or protein insertion into a membrane \cite{13}. Since
both trial moves are usable for lattice and off-lattice mod-
el systems with conformational constraints should also benefit from our self-adaptive WL procedure.
We thank K. Binder and W. Paul as well as C. Gervais and D. T. Seaton for fruitful discussions. This work was
supported in part by NSF Grant DMR-0810223.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig3}
\caption{(Color online) Specific heat \(C_V/N\) as a function of
temperature \((T)\) for ISAWs of various chain lengths \(N\) on
square (\textit{top}) and simple cubic (\textit{bottom}) lattice. Numbers in
parentheses denote corresponding energy minima. \textit{Bottom}
rows: Representative structures at specific temperatures for
\(N = 64\) (2D) and \(N = 125\) (3D).}
\end{figure}

state structure bears similarity to surface roughening on
crystal facets, i.e. the formation of kinks and edges at
the surface of a compact core without vacancies (indeed,
bulk vacancies appear at much higher \(T\) only). At
intermediate temperatures, metastable (and chain length
dependent) phases emerge but they gradually diminish
for \(N \rightarrow \infty\). Most notably, the ISAW on the sq/sc lat-
tice does not undergo a true crystallization transition as
observed for other lattice and off-lattice polymer mod-
els \cite{1, 18}. Once in the globular phase, the rigidity of
the model does not permit a further cooperative effect
(i.e. symmetry breaking) which would be necessary for
such a transition. Whereas a variation of chain length
\((N \neq \text{“magic” number})\) has an influence on the magnitude
and the position of the excitation peak at low \(T\),
the overall thermodynamic scenario remains the same
for sufficiently large \(N\). Note that it was essential to
have data for chains that were longer than other meth-
ods could treat in order to ascertain the low \(T\) behavior
of the ISAW in the thermodynamic limit.

In summary we have shown that Wang-Landau sampling
with suitable Monte Carlo trial moves (pull and bond-
rebridging moves combined) offers a powerful solution
for studying the thermodynamics of lattice homo- and
heteropolymers even in the very demanding low temper-
perature ranges of such models. A major advantage of our
method is that it remains rather simple and flexible be-
side its proven performance which has not been achieved
earlier, by more elaborate attempts \cite{10, 12, 19}. These
features make it readily applicable to the study of com-
plex biological phenomena such as e.g. protein aggrega-
gation or protein insertion into a membrane \cite{13}. Since
both trial moves are usable for lattice and off-lattice mod-
el systems with conformational constraints should also benefit from our self-adaptive WL procedure.
We thank K. Binder and W. Paul as well as C. Gervais and D. T. Seaton for fruitful discussions. This work was
supported in part by NSF Grant DMR-0810223.

\begin{thebibliography}{99}

\bibitem{1} K. A. Dill \et al., Protein Sci. 4, 561 (1995); A. Kolinski
and J. Skolnick, Polymer 45, 511 (2004).
\bibitem{2} K. A. Dill, Biochemistry 24, 1501 (1985); K. F. Lau
and K. A. Dill, Macromolecules 22, 3986 (1989).
\bibitem{3} A. D. Sokal, in \textit{Monte Carlo and Molecular Dynamics
Simulations in Polymer Science}, edited by K. Binder
(Oxford University Press, New York, 1995), p. 47.
\bibitem{4} K. Binder and W. Paul, Macromolecules 41, 4537 (2008).
\bibitem{5} H. Frauenkron \et al., Phys. Rev. Lett. 80, 3149 (1998);
J.-P. Hsu \et al., Phys. Rev. E 68, 021113 (2003).
\bibitem{6} N. Lesh \et al., in \textit{RECOMB} (2003), p. 188.
\bibitem{7} R. Backofen and S. Will, Constraints 11, 5 (2006).
\bibitem{8} J. Zhang \et al., J. Chem. Phys. 126, 225101 (2007).
\bibitem{9} D. P. Landau and K. Binder, \textit{A Guide to Monte Carlo
Simulations in Statistical Physics} (Cambridge University
Press, Cambridge, UK, 2005), 2nd ed.
\bibitem{10} Y. Iba \et al., J. Phys. Soc. Jpn. 67, 3327 (1998);
G. Chikenji \et al., Phys. Rev. Lett. 83, 1886 (1999).
\bibitem{11} M. Bachmann and W. Janke, Phys. Rev. Lett. 91, 208105
(2003); J. Chem. Phys. 120, 6779 (2004); T. Prellberg
and J. Krawczyk, Phys. Rev. Lett. 92, 120602 (2004).
\bibitem{12} S. C. Kou \et al., J. Chem. Phys. 124, 244903 (2006).
\bibitem{13} See e.g. P. M. Harrison \et al., J. Mol. Biol. 286, 593
(1999); R. Bonaccini and F. Seno, Phys. Rev. E 60, 7290
(1999); L. Zhang \et al., Biophys. Chem. 133, 71 (2008).
\bibitem{14} F. Wang and D. P. Landau, Phys. Rev. Lett. 86, 2050
(2001); Phys. Rev. E 64, 056101 (2001).
\bibitem{15} J. M. Deutsch, J. Chem. Phys. 106, 8849 (1997).
\bibitem{16} A. G. Cunha-Netto \et al., Phys. Rev. E 78, 055701(R)
(2008).
\bibitem{17} R. Fraser and J. I. Glasgow, in \textit{ICANNGA (1)} (2007),
p. 758.
\bibitem{18} D. F. Parsons and D. R. M. Williams, J. Chem. Phys.
124, 221103 (2006); W. Paul \et al., Phys. Rev. E 75, 060801(R) (2007); D. Seaton \et al., Comput. Phys. Comm.
180, 587 (2009).
\bibitem{19} T. Vogel \et al., Phys. Rev. E 76, 061803 (2007).
\bibitem{20} P. V. K. Pant and D. N. Theodorou, Macromolecules 28,
7224 (1995).
\end{thebibliography}