View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Apollo

Computer Physics Communications 201 (2016) 8–18

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Exploiting molecular dynamics in Nested Sampling simulations of small peptides

Nikolas S. Burkoff [a,1,2], Robert J.N. Baldock [b,1], Csilla Várnai [a,3], David L. Wild [a,*], Gábor Csányi [c,*]

[a] Systems Biology Centre, Senate House, University of Warwick, United Kingdom

[b] Cavendish Laboratory, University of Cambridge, United Kingdom

[c] Engineering Laboratory, University of Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 3 August 2015
Received in revised form
30 November 2015
Accepted 12 December 2015
Available online 29 December 2015

Keywords:
Nested Sampling
Alanine dipeptide
Molecular dynamics
Potential energy surface
Heat capacity

ABSTRACT

Nested Sampling (NS) is a parameter space sampling algorithm which can be used for sampling the equilibrium thermodynamics of atomistic systems. NS has previously been used to explore the potential energy surface of a coarse-grained protein model and has significantly outperformed parallel tempering when calculating heat capacity curves of Lennard-Jones clusters. The original NS algorithm uses Monte Carlo (MC) moves; however, a variant, Galilean NS, has recently been introduced which allows NS to be incorporated into a molecular dynamics framework, so NS can be used for systems which lack efficient prescribed MC moves. In this work we demonstrate the applicability of Galilean NS to atomistic systems. We present an implementation of Galilean NS using the Amber molecular dynamics package and demonstrate its viability by sampling alanine dipeptide, both in vacuo and implicit solvent. Unlike previous studies of this system, we present the heat capacity curves of alanine dipeptide, whose calculation provides a stringent test for sampling algorithms. We also compare our results with those calculated using replica exchange molecular dynamics (REMD) and find good agreement. We show the computational effort required for accurate heat capacity estimation for small peptides. We also calculate the alanine dipeptide Ramachandran free energy surface for a range of temperatures and use it to compare the results using the latest Amber force field with previous theoretical and experimental results.

© 2016 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license
(http://creativecommons.org/licenses/by/4.0/).

1. Introduction

It has been over 50 years since Ramachandran and coworkers first modelled protein peptide bonds [1]. In their work they used small peptides, containing only one or two peptide bonds, to study the sterically allowed protein dihedral angles. Using this information they developed the 'Ramachandran plot', familiar to all protein scientists today. The peptide bond is the smallest building block of proteins, and over the last few decades, it has continued to be studied intensively both experimentally [2–6] and theoretically [7–11]. Polypeptide models and force fields of varying levels of complexity have been developed, ranging from simple coarse-grained models [12], through all-atom molecular mechanics force fields [13,14], hybrid quantum mechanics molecular mechanics (QM–MM) models [15,16], up to the full quantum mechanical treatment [17]. These models have allowed the computational study of peptide thermodynamics and the exploration of their potential and free energy surfaces [10,18,19,11].

Although short peptides which occur naturally, such as the five residue neurotransmitter Met-enkephalin [20], are of particular interest, the peptide bonds in short peptides are thought to have similar properties to the peptide bonds in unfolded and unstructured proteins [21], and so their study can also inform our knowledge of proteins in their unfolded state. Peptide models have also been used to study peptide aggregation [22] and have been used in order to develop [23–25] and test [21] more general protein force field parameters and models.

Running in parallel to the development of these models and force fields, there has been considerable work in developing

* Corresponding authors.
E-mail addresses: d.l.wild@warwick.ac.uk (D.L. Wild), gc121@cam.ac.uk (G. Csányi).

[1] These authors contributed equally to this work.

[2] Current address: Tessella, Elopak House, Technology Park, Stevenage, United Kingdom.

[3] Current address: The Babraham Institute, Babraham Research Campus, Cambridge, United Kingdom.

http://dx.doi.org/10.1016/j.cpc.2015.12.005
0010-4655/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

sampling algorithms in order to fully explore the potential and free energy surfaces of proteins and peptides, and to calculate accurate thermodynamics of the force fields used. These algorithms are required because standard molecular dynamics (MD) struggles to overcome energy barriers in a computationally feasible time scale and thus cannot fully sample the conformational space of interest. The *de facto* standard algorithm for general configurational phase space exploration is replica exchange molecular dynamics (REMD) [26]. A set of canonical MD trajectories are run with each 'replica' using a different temperature parameter. Periodically, the swapping of conformations for two replicas is proposed and is accepted or rejected using the standard Metropolis–Hastings acceptance criterion. The high temperature replicas ensure the system can easily escape from local modes. Many extensions, such as allowing the temperature of the replicas to change and adapt throughout the simulation in order to improve efficiency, have been developed [27]. Subsequent to the original REMD research on the penta-peptide Met-enkephalin [26] the method has been very widely used for proteins, for example, to fold the Trp-cage miniprotein [28] and calculate the heat capacity curve of an SH3 domain [29]. Many other sampling techniques have been developed. For example, if the collective variable of interest is known *a priori* then the metadynamics technique can be used [30–32].

One of the main thermodynamical properties of interest to protein scientists is the free energy difference between different states of the system. These are used to plot the free energy surface with respect to reaction co-ordinates of interest and give key insights into the macroscopic behaviour of the system. Although in this work we focus on algorithms which do not require suitable reaction co-ordinates to be known *a priori*, if these are available, then specialised free energy algorithms can be used to calculate such differences [33–35]. One such algorithm is umbrella sampling [33], where an extra bias force is applied to keep the reaction co-ordinate at a chosen value. Originally tested on Lennard-Jones (LJ) clusters, umbrella sampling has been used to study short peptides [36] and is now a standard free energy calculation algorithm.

Sophisticated general conformational sampling algorithms have also been developed which do not require any prior knowledge about the potential energy surface. One example is accelerated molecular dynamics, where a bias function of *only* the potential energy is used to facilitate the traversal of energy barriers [37]. To initially test the algorithm, in the original work, Hamelberg et al. calculate the free energy surface of alanine dipeptide, a simple molecule with only a single peptide bond [37]. Another example is multicanonical sampling, using either Monte Carlo [38] or molecular dynamics [39] sampling. In this algorithm, instead of sampling from the Boltzmann distribution: $\mathbb{P}(\Omega) \propto \exp(-E(\Omega)\beta)$, samples are drawn from the *multicanonical* distribution: $\mathbb{P}(\Omega) \propto 1/g(E(\Omega))$ where $g(E)$ is the density of states. Multicanonical sampling was specifically designed to be efficient when sampling systems which undergo a first order phase transition [38]. Multicanonical MD has been used to study the free energy landscapes of tri-peptides [40] and a seven residue DNA binding peptide [41]. Recently, the algorithm has been applied to larger peptides and protein domains; further applications of the multicanonical MD algorithm can be found in a recent review [42]. Many variants of the multicanonical algorithm, such as the Wang–Landau algorithm [43], have also been developed. Further examples of MC algorithms include equi-energy [44] and well-tempered ensemble [45] sampling.

Recently, Skilling introduced a novel technique, Nested Sampling [46], which has distinct advantages for sampling atomistic systems. Subsequently, an algorithm similar to Nested Sampling but utilising only a single walker, originally called the "energy partitioning method", was independently developed for sampling water molecules and binary mixtures of fluids [47,48].

Nested Sampling.

Nested Sampling is an algorithm specifically designed to sample high dimensional spaces [46,49]. The algorithm is designed for systems where the bulk of the probability mass is contained in an exponentially small volume of phase space. The algorithm outputs a set of samples and associated weights from which an estimate for the partition function (also known as the marginal likelihood) and also thermodynamic variables, such as heat capacities and free energy differences, can be calculated at any temperature.

Whilst initially developed for Bayesian statistical inference [46], the algorithm is well-established in the astrophysics community [50] and has also been successfully applied in a variety of other fields including bioinformatics [51], systems biology [52] and flow model selection [53]. The Nested Sampling algorithm has also been applied to atomic systems. Pártay et al. have used it to study LJ clusters [54] and hard sphere models [55] where it significantly outperformed parallel tempering, and Burkoff et al. explored the potential energy surface of a coarse-grained protein model [56].

The original Nested Sampling algorithm is a Monte Carlo (MC) sampling algorithm, and in the work of Burkoff et al. a coarse-grained protein model used was specifically designed to allow efficient MC crankshaft moves [56]. For example, all bond lengths were fixed and the peptide bond was kept exactly planar. In the present work we apply the algorithm to an all-atom force field where the extra degrees of freedom would make MC sampling inefficient due to the high anisotropy when explicit bond stretching and angle bending degrees of freedom are taken into account.

Recently, however, Skilling introduced *Galilean* Nested Sampling [57], a variant of the Nested Sampling algorithm in which momentum variables are introduced for each degree of freedom, and system specific MC moves are not required. The momenta are then used to evolve sample points using Galilean dynamics, a novel exploration procedure, rather than using the standard Hamiltonian equations of motion. In this work we implement Galilean Nested Sampling within the Amber MD package [13] and we test the algorithm by generating thermodynamical data for alanine dipeptide, both *in vacuo* and in implicit solvent. Unlike earlier work with alanine dipeptide, we focus on calculating accurate[4] heat capacity curves and compare the Nested Sampling results to those obtained using the standard REMD procedure. We also calculate dihedral angle Ramachandran free energy surfaces, comparing the results to previous theoretical and experimental work. Finally, we discuss the properties of Galilean Nested Sampling and our expectations for the method, looking to the future.

2. Methods

Following the principles of classical statistical mechanics, the configurations of constant volume systems which are in thermal equilibrium with their surroundings are distributed according to the Boltzmann (or canonical) distribution. Specifically, at temperature T, the probability of the system adopting the configuration Ω is proportional to $\exp(-E(\Omega)\beta)$ where $\beta = 1/(k_B T)$, k_B is the Boltzmann constant ($\approx 2 \times 10^{-3}$ kcal/mol/K) and $E(\Omega)$ is the potential energy of configuration Ω.

The normalisation constant of the Boltzmann distribution, the partition function,

$$Z(\beta) = \int_{\Omega} \exp(-E(\Omega)\beta) d\Omega,$$

[4] In this work, 'accurate thermodynamic data' is shorthand for 'accurate thermodynamic data *for the force field used*'.

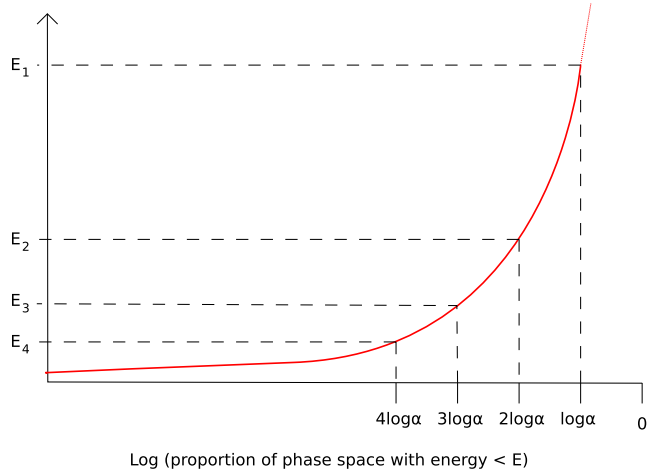

Fig. 1. The energy levels $E_1 > E_2 > \cdots$ are chosen to be equidistant in log phase space volume. Therefore, the proportion $1-\alpha$ of conformations have energy $>E_1$ and $\alpha-\alpha^2$ of conformations have energy $<E_1$ and $>E_2$.

is of fundamental importance in statistical physics, as it can be used to obtain thermodynamic quantities. For example, the internal energy,

$$U = \langle E(\Omega)\rangle_\beta \equiv -\left(\frac{\partial \ln Z}{\partial \beta}\right)_V,$$

and the constant volume configurational heat capacity

$$C_v = \langle E^2(\Omega)\rangle_\beta - (\langle E(\Omega)\rangle_\beta)^2 \equiv k_B\beta^2\left(\frac{\partial^2 \ln Z}{\partial \beta^2}\right)_V,$$

where $\langle . \rangle_\beta$ is expectation under the Boltzmann distribution.

Although it is possible to estimate the partition function using the 'harmonic mean approximation', $Z^{-1} = \langle \exp(E(\Omega)\beta)\rangle$, this estimator has infinite variance and hence should be avoided [58].

2.1. Nested Sampling algorithm

The Nested Sampling algorithm is an iterative procedure which generates a set of energy levels $E_1 > E_2 > E_3 \ldots$, where for each i, E_i is chosen so that

$$\frac{\int_\Omega \mathbb{I}\{E(\Omega) < E_i\}d\Omega}{\int_\Omega \mathbb{I}\{E(\Omega) < E_{i-1}\}d\Omega} \approx \alpha,$$

for some fixed proportion α and $\mathbb{I}$ is the indicator function. Hence the algorithm takes steps equidistant in "the logarithm of phase space volume", as illustrated in Fig. 1.

The proportion $\omega_i = \alpha^{i-1} - \alpha^i$ of conformations have energy between E_{i-1} and E_i and hence, by using numerical integration, we can estimate the partition function to be

$$Z(\beta) = \int_\Omega \exp(-E(\Omega)\beta)d\Omega \approx \sum_i \omega_i \exp(-E(\Omega_i)\beta). \tag{1}$$

The algorithm does not prescribe a specific terminating condition, only running until the estimators for the observables of interest have sufficiently converged. In previous work the algorithm was terminated at iteration j when

$$\log\left(\sum_{i=1}^{j}\omega_i \exp(-E(\Omega_i)\beta)\right) - \log\left(\sum_{i=1}^{j-1}\omega_i \exp(-E(\Omega_i)\beta)\right) < \epsilon$$

for the lowest temperature, $T_{\min}$, (respectively highest β) of interest [56]. We follow the same procedure here, and by setting $\epsilon = 10^{-5}$, we ensure the heat capacity estimate has converged at $T_{\min}$.

Generation of energy levels.

Although the original algorithm does not prescribe a specific method to calculate the energy levels, a Monte Carlo method is proposed and is described in Algorithm 1. An active set of K samples, uniformly distributed over the set of configurations with energy below the current energy level is maintained. The set is initialised with samples uniformly distributed throughout the whole phase space and the energy of the highest energy configuration in the active set is chosen to be the first energy level, E_1, and this configuration, Ω_1 is removed from the active set.

The $K-1$ samples remaining in the active set are uniformly distributed over the set of configurations with energy below the current energy level and only a single new configuration is required to replace the one that was removed. This configuration is generated by copying an existing member of the active set and using it to initialise a Markov chain with equilibrium distribution given by $\mathbb{P}(\Omega) \propto \mathbb{I}\{E(\Omega) < E_1\}$. The final configuration of the Markov chain is then placed into the active set. The second energy level E_2 is then taken to be the energy of the highest energy configuration currently in the active set, Ω_2, and the procedure repeats, generating $E_3, E_4, \ldots$.

At each iteration, the proportion of the configuration space with energy less than that of the sample with highest energy is proportional to Beta$(K+1, 1)$ and its expectation value is $K/(K+1)$, which is therefore approximately the value of α. It is straightforward to quantify the uncertainty in α when producing estimates of the partition function [49].

Algorithm 1 Monte Carlo (MC) algorithm to generate Nested Sampling energy levels

Generate K samples uniformly distributed throughout phase space, the active set
$i \leftarrow 1$
loop
  Remove sample Ω^* with highest energy, E^*, from the active set
  Output $E_i = E^*$ and $\Omega_i = \Omega^*$
  Copy randomly chosen member of active set to use as a starting conformation for a Markov chain
  Run Markov Chain Monte Carlo with equilibrium distribution $\propto \mathbb{I}\{E(\Omega) < E^*\}$
  Add the final conformation from Markov chain to the active set
  $i \rightarrow i+1$
end loop

The sample points removed from the active set, $\{\Omega_1, \Omega_2, \ldots\}$ can be used to estimate properties of the Boltzmann distribution at any temperature. Ω_i represents ω_i of configuration space, and therefore represents $\chi_i(\beta) = \omega_i \exp(-E(\Omega_i)\beta)/Z(\beta)$ of the probability mass of the Boltzmann distribution at inverse thermodynamic temperature β. Any property $Q(\Omega|\beta)$ can be estimated as

$$\mathbb{E}(Q|\beta) = \sum_i \chi_i(\beta)Q(\Omega_i).$$

For example, the heat capacity is given by

$$C_v(\beta) \approx k_B\beta^2\left[\sum_i \chi_i(\beta)E^2(\Omega_i) - \left(\sum_i \chi_i(\beta)E(\Omega_i)\right)^2\right]. \tag{2}$$

Estimates for (Helmholtz) free energy differences can also be computed: if the set of samples $\{\Omega\}$ can be split into disjoint

macrostates A and B then the free energy difference is given by

$$F_A - F_B \approx -\beta^{-1}\left[\log\left(\sum_{\{i:\Omega_i\in A\}}\omega_i\exp(-E(\Omega_i)\beta)\right) - \log\left(\sum_{\{i:\Omega_i\in B\}}\omega_i\exp(-E(\Omega_i)\beta)\right)\right]. \quad (3)$$

2.2. Galilean exploration

In our previous work sampling protein models we used a coarse-grained force field, CRANKITE [59,60]. In this model each amino acid had 3 degrees of freedom, the dihedral angles ϕ and ψ and the C_α valence angle, and we used crankshaft rotations as MC moves which efficiently sample the configurational space [56]. However, more realistic all-atom models have more degrees of freedom, and in order to sample the system, additional MC moves such as angle bending and bond stretching, must be included. These moves, especially at low temperatures, or for systems which include explicit solvent molecules, are often inefficient. For these systems, sampling using MD, which has shorter decorrelation times than MC, is often preferred.

In this work we implement *Galilean* exploration, a method of exploration used to generate Nested Sampling energy levels, recently introduced by Skilling [57]. Galilean exploration does not require system-specific MC moves. Following the MD approach, the atoms of the conformation are given momenta and the system is then evolved along a trajectory generating samples uniformly distributed over all conformations with energy less than a prescribed value. The details of Galilean exploration are given below.

In Galilean Nested Sampling, in order to generate a new sample for the active set, an existing member of the set, Ω, with atomic coordinates $\mathbf{x}$ is chosen. A set of velocities, $\mathbf{v}: v_i \sim \mathcal{N}(0, k_BT)$, for a chosen parameter,[5] T, are drawn and the move $\mathbf{x}\rightarrow\mathbf{x}'=\mathbf{x}+\tau\mathbf{v}$ is proposed, where τ is the timestep.[6] If the proposed conformation has energy below the current energy level, the move is accepted, otherwise we try to 'reflect' the conformation back into the acceptable region by choosing a unit normal vector $\mathbf{n}$ and proposing the move $\mathbf{x}\rightarrow\mathbf{x}''=\mathbf{x}'+\tau(\mathbf{v}-2\mathbf{n}(\mathbf{n}\cdot\mathbf{v}))$. In principle any unit vector $\mathbf{n}$ can be used. However, if possible, we would like to reflect off the boundary of the acceptable region, thus ensuring the move is accepted. We can estimate this orientation by taking $\mathbf{n}$ as the unit vector in the direction of $\nabla E(\mathbf{x}')$.

If the new conformation has energy less than the current energy level, the reflection is accepted and the trajectory continues with velocity $\mathbf{v}'=\mathbf{v}-2\mathbf{n}(\mathbf{n}\cdot\mathbf{v})$. If not, detailed balance insists we reject the move, thus remaining at $\mathbf{x}$, and we continue the trajectory by using the velocity $-\mathbf{v}$. See Fig. 2 for an example trajectory.

Unless the energy level boundary is crossed, the same $\mathbf{v}$ continues to be used throughout the trajectory, as the induced systematic motions can be expected to explore more efficiently than random diffusions. However, in order to decrease equilibration time, it is suggested to slightly perturb the velocity at each iteration and instead of using velocity $\mathbf{v}$, use the velocity $\mathbf{v}^{\mathbf{p}}=\mathbf{v}\cos\theta+\tilde{\mathbf{v}}\sin\theta$ where $\tilde{\mathbf{v}}$ is a newly drawn set of velocities and θ is small.

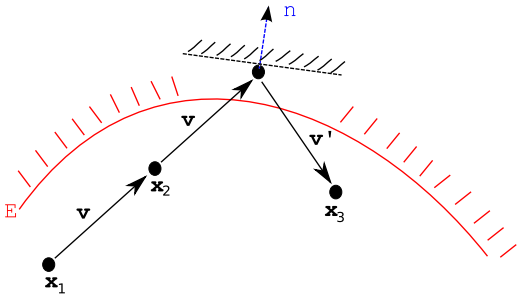

Fig. 2. At $\mathbf{x_1}$ the conformation $\mathbf{x_2}=\mathbf{x_1}+\tau\mathbf{v}$ is proposed. As the conformation remains in the acceptable region (has energy below E, shown by the red contour) it is accepted. The proposed move to $\mathbf{x}'=\mathbf{x_2}+\tau\mathbf{v}$ takes the conformation outside the acceptable region, so it is reflected to $\mathbf{x_3}=\mathbf{x}'+\tau\mathbf{v}'$, where $\mathbf{v}'=\mathbf{v}-2\mathbf{n}(\mathbf{n}\cdot\mathbf{v})$ for the unit vector $\mathbf{n}=\nabla E(\mathbf{x}')$. As $\mathbf{x_3}$ is inside the acceptable region, the move is accepted and the trajectory continues using velocity $\mathbf{v}'$. If $\mathbf{x_3}$ were to be outside the acceptable region, then the move would have been rejected, the conformation returned to $\mathbf{x_2}$ and the velocity reversed to $-\mathbf{v}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The final conformation of the Galilean MD trajectory is then saved into the active set. As the energies of conformations during Galilean MD simulations are uniformly distributed below the specified energy level, the active set will contain conformations with energies uniformly distributed below the specified energy level.

2.3. Galilean Nested Sampling for peptides

In this work we adapt the Amber molecular dynamics package [13] to perform Galilean exploration in order to generate Nested Sampling energy levels, see Supplementary material for implementation details (Appendix A). We use the Amber ff12SB protein force field with `igb=6` for vacuum simulations and for implicit solvent simulations, `igb=8`, a generalised Born solvation model [61]. We use the default Amber chirality and *trans/cis* peptide bond restraints. We do not employ a van der Waals distance cutoff and do not constrain the covalent hydrogen bond distances with SHAKE.

Due to the rapid shrinking of the available phase space volume, we find it is sufficient to set $\alpha=0.5$, and thus remove half of the available phase space each iteration. However, it is necessary to estimate the next energy level to a very high degree of accuracy, and following the original algorithm, which might use an active set with a single sample, is inappropriate. Therefore, instead, at each iteration we use Galilean exploration to generate a large set of uniformly distributed samples and use the median of these samples to estimate the next energy level. The samples from this iteration with energy less than the next energy level are still uniformly distributed, so we can, as in the original algorithm, re-use these samples in subsequent iterations. The starting conformations for the trajectories of the subsequent iteration (each trajectory resulting in a new sample point in the active set) are chosen uniformly from the set of conformations with energy less than the new energy level.

As the accessible region of phase space shrinks, it is necessary to reduce the magnitude of the velocities in order to keep the trajectories within the allowed region. We define the *mean free path* to be the average number of successful steps taken before requiring a reflection. We use the variable T in order to keep the mean free path constant throughout a Nested Sampling simulation.[7] The other parameters of the NS simulation, *i.e.* the number of Galilean MD trajectories at each NS iteration, as well as the length of the trajectories, are parameters to be optimised.

[5] The parameter T controls how fast the particle moves and hence is analogous to temperature in canonical MD simulations. However, it does not correspond to the temperature of any canonical MD simulation.

[6] With Galilean exploration, there is a direct correspondence between timestep τ and 'temperature' T: the transformation $(T,\tau)\rightarrow(aT,\tau/\sqrt{a})$ for constant a is invariant. In this work, for each simulation, we fix τ and allow T to vary as described later in the text.

[7] See the Supplementary material for details (Appendix A).

In the original algorithm, simulations are initialised by choosing samples uniformly throughout the whole of configuration space. As we are only interested in thermodynamics at relatively low temperatures, we initialise the algorithm by generating a set of samples uniformly distributed over the conformations with potential energy below a chosen initial energy level. We refer the reader to the Supplementary Material for further details concerning initialising the algorithm at a specific energy level [62].

Each reflection requires two separate force evaluations, one when the sample steps outside the acceptable region and one after it has been reflected. Therefore, when the mean free path is lower there are more reflections, and so trajectories must be shortened in order to maintain the same number of force evaluations,[8] and therefore computational expense, when comparing efficiencies. Due to the implementation within Amber, in this work we calculate the forces at each step of the trajectory. However, it is important to note that this is not strictly required as Galilean exploration only requires the calculation of the forces (*i.e.* $-\nabla E$) when outside the acceptable region. At other times, only the potential energy is required (to check whether the trajectory has left the acceptable region).

3. Results

We demonstrate the Galilean Nested Sampling algorithm by using it to calculate the thermodynamics and free energy surfaces of the small peptide alanine dipeptide both in vacuum and implicit solvent.

3.1. Alanine dipeptide in vacuo

It is over 50 years since Ramachandran and co-workers analysed the sterically allowed peptide dihedral angles ϕ and ψ. In their work they introduced the name dipeptide to describe molecules which include, besides a single amino acid, adjacent residues as far as the C_α atoms [1]. Over the last 50 years dipeptides, and particularly alanine dipeptide (N-acetyl-alanyl-N'-methylamide), have been studied experimentally, both in solution [6] and in the gas phase [4]. Alanine dipeptide has also been studied from a quantum mechanical perspective [8,9] and has previously been used to parametrise molecular force fields [63] and test their accuracy [21]. Unlike previous work, here we focus not only on calculating the free energy (or potential energy) surface, but on the accurate determination of the heat capacity of the system.

Heat capacity.

Fig. 3 shows estimates for the heat capacity of alanine dipeptide *in vacuo* for five independent Nested Sampling simulations (lines) calculated using Eq. (2). Although the potential energy at temperatures of interest is low (e.g. at 360 K, $U \approx 0 \pm 4$ kcal/mol), the initial energy cutoff was chosen to be $E = 100$ kcal/mol. This is necessary due to the extremely high energy barrier separating room-temperature accessible conformations with dihedral angle $\phi > 0$ and those with $\phi < 0$ (see Figs. 4 and 5). Although for biophysical systems we would not normally be interested in the behaviour of the system at 100 K, for this study, we choose $T_{\min} = 100$ as this allows us to capture the peak in the heat capacity curve.

Following our previous work [56], we choose to use a large number of independent walkers, in this case 16,000. We use the parameter T to keep the mean free path ≈ 2 and by setting $\theta = 0.2$ we allow a small amount of velocity randomisation every Galilean step. See the Discussion Section for further details concerning the chosen parameters.

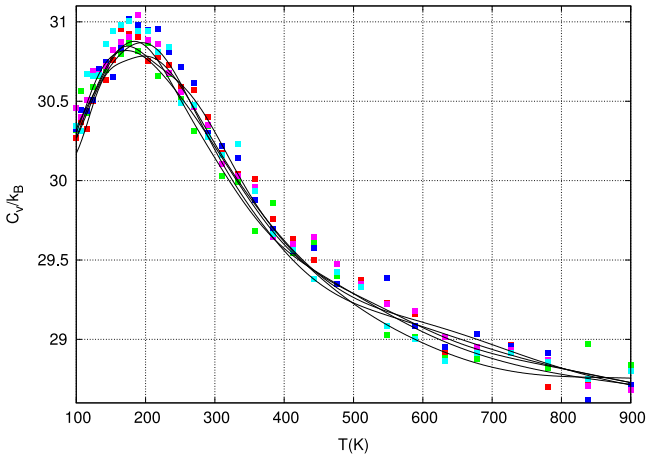

Fig. 3. The heat capacity, C_v, from 5 independent Nested Sampling simulations (lines) and 5 REMD simulations (points). All simulations used a comparable number of force evaluations ($\approx 9.6 \times 10^9$). See the text for further details. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Each Galilean trajectory runs for a total of 2700 steps outputting the potential energy every 75 steps. This implies each Nested Sampling iteration uses approximately 57 million force evaluations, which leads to a total of $\approx 9.6 \times 10^9$ force evaluations per simulation. Note that this is a *very* large number of force evaluations for such a small system. However, as the value of the heat capacity only varies by $\sim 2k_B$ over the 800 K temperature range, a very large number of force evaluations are required to reduce the statistical error to a small enough value to clearly resolve the curve.

The variance between estimates from independent simulations is very small. However, in order to show that the algorithm has converged to the *correct* value, Fig. 3 also includes heat capacity estimates from five independent REMD simulations, and there is good agreement between the methods. The REMD simulations use a similar number of total force evaluations (9.6×10^9) as the Nested Sampling simulations. The temperatures of the 32 replicas are in a geometric progression from 100 to 900 K and swaps between different replicas are attempted every 2 ps. Hydrogen atoms were unconstrained and hence a relatively small time step (0.2 fs) was used to ensure accuracy, especially for high temperature replicas; the other parameters of the REMD simulations (such as the number of replicas) have not been especially optimised. This implies that each individual REMD simulation has a length of 60 ns. Therefore, we claim only that Nested Sampling and REMD are of similar efficiencies for this system. Rigorous benchmarking of REMD and Galilean Nested Sampling on larger systems is the focus of future work.

Free energy surface.

A standard free energy reaction co-ordinate for alanine dipeptide is the pair of dihedral angles (ϕ, ψ), for example see [11]. We split the conformations generated by a Nested Sampling simulation into separate 'bins' based on their dihedral angles and then use Eq. (3) to generate the free energy surface. A Gaussian filter has then been applied to smooth the data and the result, for 300 K, is shown in Fig. 4. For comparison, the unsmoothed free energy surface is shown in the Supplementary material (Appendix A), see Figure S3 [62]. When using the original Nested Sampling algorithm, each energy level corresponded to exactly one sample point which represented ω_i of phase space. In this work, we output a whole set of samples for each energy level and, when calculating

[8] Specifically, the total number of force evaluations $= (m+2)S/(m+1)$, where S is the number of steps and m the mean free path.

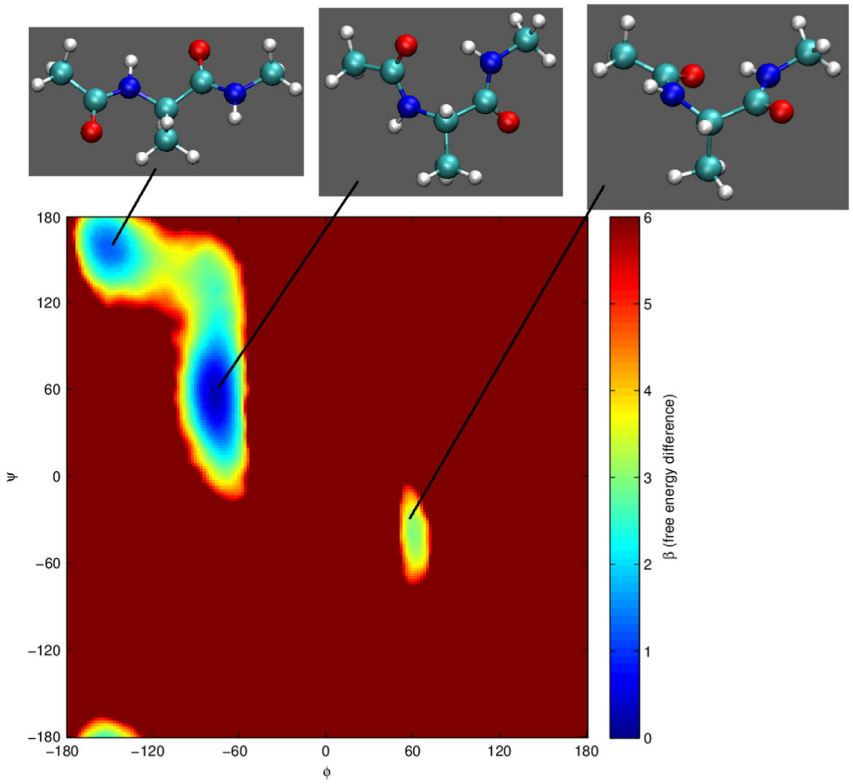

Fig. 4. Top: Three conformations of alanine dipeptide accessible (*in vacuo*) at room temperature, from left to right C_5, C_{7eq} and C_{7ax}, see [8,9,4]. Bottom: The free energy surface of alanine dipeptide *in vacuo* at 300 K. See the text for further details. Note, in this work, the dark red used for β(free energy difference) = 6 is also used where this value is greater than 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

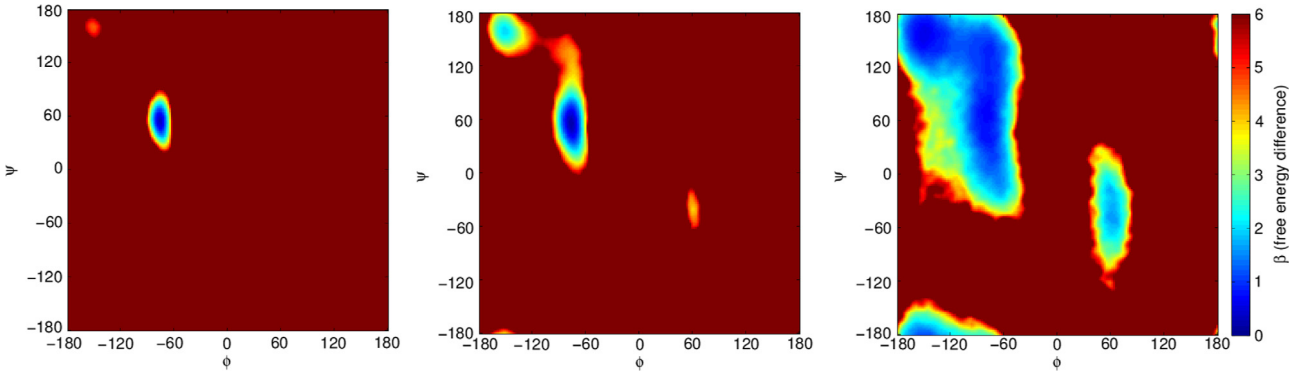

Fig. 5. Free energy surface of alanine dipeptide *in vacuo* at 100 K, 200 K and 900 K. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

free energy surfaces, we give each sample a uniform[9] share of the weighting ω_i.

Although the focus of this work is the implementation of Galilean Nested Sampling rather than force field development, it is interesting, nevertheless, to use these results to compare the Amber ff12SB force field with experimental results and quantum mechanical (QM) calculations. Pohl et al. compared alanine dipeptide QM calculations with infra-red absorption spectra in Ar and Kr isolation matrices [4]. From QM calculations they found that the two most common conformations were expected to be C_{7eq} (also named γ_L) and C_5 ($\beta_{L(D)}$). Depending on the choice of basis sets, the relative abundance of C_{7eq} (at 343 K) was between 32% and 63%. For this force field, we also find the same two common conformations with the abundance of C_{7eq} (at 343 K) ≈66%. These conformations were also identified experimentally [4]. Tobias et al. [36] compared a QM and molecular mechanics (MM) potential energy surface of alanine dipeptide and, although, they find differences in the position of local minima, they conclude the MM force field provides a very good description of alanine dipeptide *in vacuo*. We find the locations of minima agree well with the positions on the MM force field used by Tobias et al.

Free energies are calculated directly from the logarithm of the partition function, without differentiation, and we find excellent agreement between independent Nested Sampling simulations when calculating free energies, see, for example, Figure S3 in the Supplementary material [62]. For alanine dipeptide, there is a

[9] Technically, as samples have slightly different energies ($E_i > E(\Omega) > E_{i+1}$), they ought to represent slightly different proportions of phase space. However, the energy gap between successive energy levels is extremely small, and this approximation is analogous to the approximation used by the original Nested Sampling algorithm when performing the numerical integration to estimate the partition function and therefore, in practice, we find this approximation adequate.

clear choice of reaction co-ordinates for a low dimensional free energy surface (the dihedral angles), and as the system is so small, Fig. 4 could easily be calculated by a specialised free energy calculation method such as umbrella sampling [33]. However, these methods typically require a reaction co-ordinate to be chosen *a priori*. This is not the case for Nested Sampling as no reaction co-ordinate is required for the sampling algorithm. The free energy surface as a function of the reaction coordinate at a given temperature is obtained by calculating the free energy using the marginal Boltzmann probability distribution as a function of only the reaction coordinate, and this can be obtained *a posteriori* for any desired collective variable. For example, a discrete order parameter corresponding to the hierarchical basin structure of the potential energy surface can actually be derived directly from clustering the samples output by a Nested Sampling simulation and we refer the reader to [54] for further details.

By reweighting the samples from the same Nested Sampling simulation, the free energy surface can be calculated for arbitrary temperatures. Fig. 5 shows the free energy surface of alanine dipeptide at 100 K, 200 K and 900 K. Although there is a clear energy barrier at $\phi \approx 0$, it is possible for canonical trajectories at 900 K to overcome this barrier. However, at 600 K it is all but impossible. This shows the importance of ensuring there are replicas which have temperatures high enough to overcome all energy barriers when running REMD. Further discussion concerning this can be found in the Supplementary material [62].

3.2. Alanine dipeptide in implicit solvent

In this section we perform Nested Sampling of alanine dipeptide in solvent and compare the results generated by the Amber ff12SB force field to the latest experimental data.

There is no theoretical barrier to using Galilean Nested Sampling algorithm with explicit solvent molecules, as each solvent molecule can be given velocities and the whole system can be evolved using Galilean exploration. However, in this work we have focused on the calculation of accurate heat capacity curves and so, in order to reduce computational expense, we have chosen to use a generalised Born [61] implicit solvent model.

The initial energy level was chosen to be 75 kcal/mol, which is high enough to allow the heat capacity to be calculated at 900 K, similarly to the *in vacuo* case. All other parameters have been kept the same except, in order to capture the peak of the heat capacity curve, we set $T_{\min} = 30$ K. Therefore we needed to calculate an additional 48 energy levels and, as we chose to use the same number of force evaluations for the simulations as previously, these additional iterations meant we had to shorten trajectory lengths from 2700 to 2100 steps.

Heat capacity.

Fig. 6 shows the heat capacity of alanine dipeptide in implicit solvent. There is, again good agreement between Nested Sampling and REMD simulations. In this case, the 32 temperatures of the REMD replicas were chosen in geometric progression from 30 to 900 K. The peak of the curve is ≈140 K lower than the *in vacuo* case.

Free energy surface.

Analogous to the *in vacuo* case, the dihedral angle free energy surface of alanine dipeptide in solvent can be calculated using the samples output from a Nested Sampling simulation. Fig. 7 shows the free energy surface at 300 K together with images of the three low energy minima, P_{II}, β and α_{R} as defined by [6].

The results presented here clearly show there are three free energy minima and their locations are given by $P_{\text{II}}(-80^\circ, 150^\circ)$, $\beta(-150^\circ, 150^\circ)$ and $\alpha_{\text{R}}(-75^\circ, -20^\circ)$. These results agree qualitatively with those from a published QM/MM force field ([21] Figure 6). Experiments cannot determine the dihedral angles to a high level of accuracy but provide probabilities of finding the system in a specific conformation (*i.e.* 'basin') [6]. In this section we use our Nested Sampling results to calculate these probabilities for the Amber force field used. Each conformation from Nested Sampling is assigned to a 'basin', P_{II}, β, α_{R} or 'other' where the basins are defined in Fig. 8. The choice for basin definitions has been guided by the free energy surface, rather than previous definitions found in the literature. However, the occupancy probabilities shown in Fig. 9 are not sensitive to the precise definitions used. Using Eq. (3), free energy differences, and hence probabilities of occupancy (*i.e.* $\mathbb{P}(\Omega \in P_{\text{II}}|T)$, with T the canonical temperature) can be calculated. Fig. 9 compares these probabilities of occupancy with probabilities derived from published ATR-absorbance spectra data [6]. The experimental results are shown by squares and the estimates calculated from the Nested Sampling simulations are shown by the error bars (mean ± sd of 5 independent simulations). The Nested Sampling probabilities of occupancy for 'other' (≈2%) are not displayed.

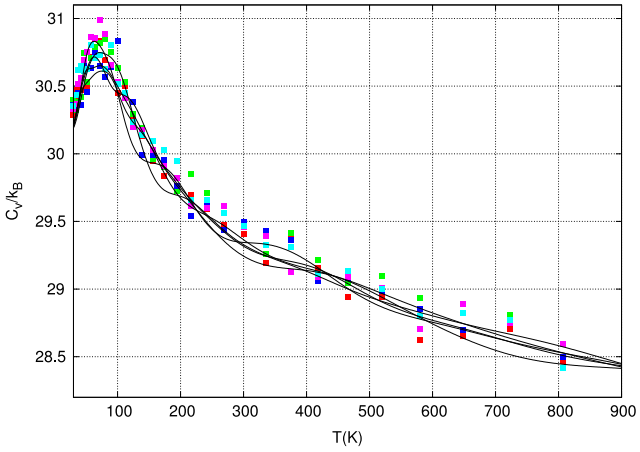

Fig. 6. The heat capacity, C_v, of alanine dipeptide in implicit solvent from 5 independent Nested Sampling simulations (lines) and 5 REMD simulations (points). All simulations used a comparable number of force evaluations (≈9.6×10^9). See the text for further details. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 9 shows that this protein force field (ff12SB), together with the generalised Born implicit solvent `igb=8`, overestimates the probability of finding the molecule in the α_R conformation. The inaccuracy of peptide force fields is detailed in the Discussion Section.

4. Discussion

Unlike proteins, where there is often a dominant free energy minimum (the native state), differences between free energy basins in peptides are typically much smaller and a distribution of states exists when the peptide is in thermal equilibrium. Therefore, a small inaccuracy in a protein force field can effect a large change in the equilibrium distribution when compared to experiments. We find that this is the case with the Amber force field used here, with the α_{R} conformation being over-represented.

The Amber force field was originally developed to study proteins in their native state, with secondary structure already formed, rather than studying the peptide bond in the unfolded state. Small peptides, which lack secondary structure, are believed to behave in similar ways to proteins in their unfolded state and previous studies have found that standard molecular force fields often struggle to reproduce peptide QM results [21]. The fact that the same peptide in different molecular force fields has different propensity to form helical or extended structures is a well known

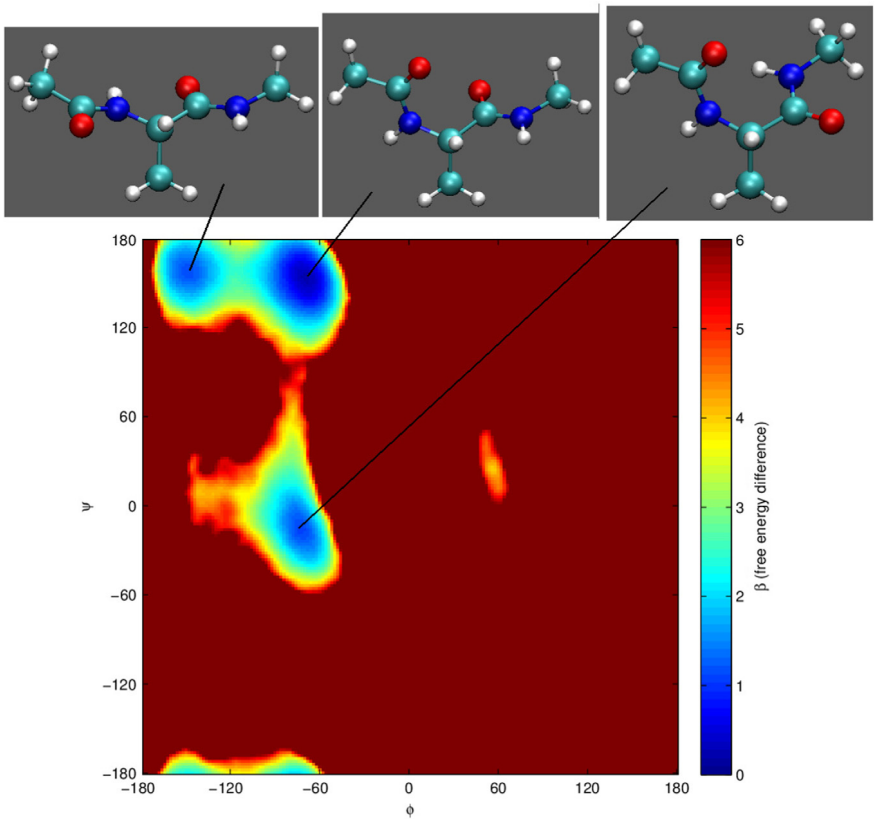

Fig. 7. Bottom: The Ramachandran free energy surface at 300 K for alanine dipeptide. Top: Conformations from the three main accessible regions of the Ramachandran plot, from left to right β, P_{II} and α_R [6]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

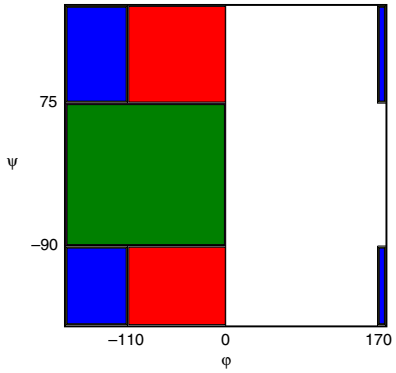

Fig. 8. The dihedral angle definitions of P_{II} (red), β (blue), α_R (green) and 'other' (white). The choice for basin definitions has been guided by the free energy surface, rather than previous definitions found in the literature, however, the occupancy probabilities shown in Fig. 9 are not sensitive to the precise definitions used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

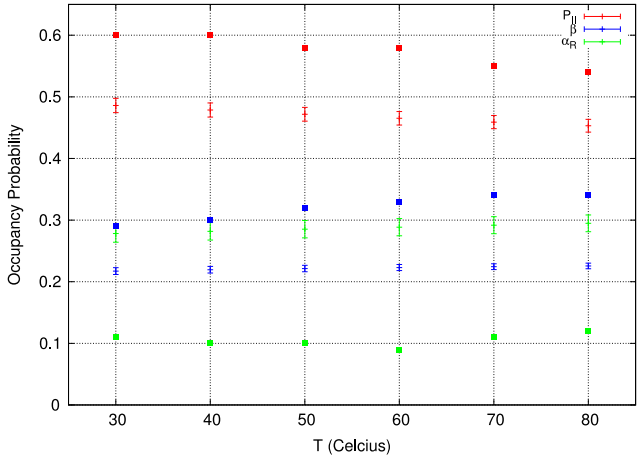

Fig. 9. The occupancy probabilities for the three main conformations P_{II} (red), β (blue), α_R (green) as a function of temperature. The squares are ATR-absorbance spectra data [6] and the error bars are mean ± sd of 5 independent Nested Sampling simulations. The Nested Sampling probabilities of occupancy for 'other' (≈2%) are not displayed. The 'other' refer to the small population of α_L (left-handed helical) conformations, the free energy minimum with $\phi > 0$ in Fig. 7. No direct test to detect these conformations was possible using the experimental techniques used in [6]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phenomenon [21,25,64] and corrections to existing force fields, to accurately reproduce helix propensity, have been developed [25].

Galilean Nested Sampling.

The Galilean Nested Sampling simulations in the Results Section used $\theta = 0.2$, which introduced a small amount of randomisation at every Galilean step. This randomisation is important in order to efficiently sample the system; Fig. 10 (left) shows three Nested Sampling simulations with exactly the same parameters as those in Fig. 3 except $\theta = 0.01$ rather than 0.2. The same REMD data are shown for ease of comparison.

We also find that having a large number of short trajectories is beneficial as shown in Fig. 10 (right). In this figure, the same parameters were used as in Fig. 3 except instead of 16,000 trajectories of 2700 steps each, 160 trajectories of 270,000 were used each iteration. As with the case of Monte Carlo Nested Sampling in our previous work, the number of independent trajectories and their length are convergence parameters. The key quantity controlling the error is the total number of force evaluations. However, beyond a certain trajectory length, the samples are fully decorrelated, and making the trajectories even

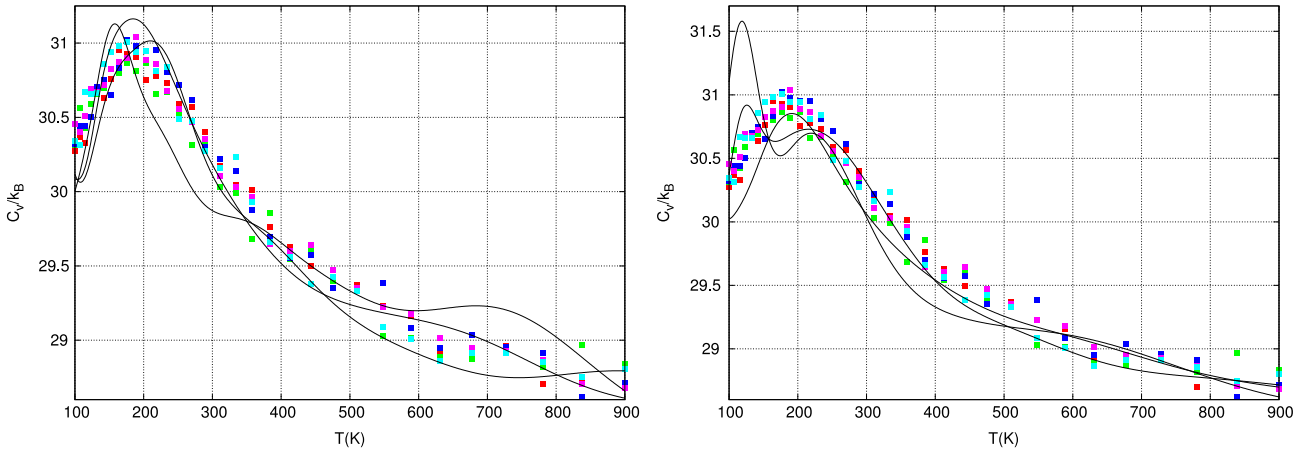

Fig. 10. Nested Sampling simulations of alanine dipeptide *in vacuo* with the same number of force evaluations as those in Fig. 3. The same parameters were used except in left: $\theta = 0.01$ rather than 0.2 and in right: 160 trajectories of 270,000 were used each iteration instead of 16,000 trajectories of 2700 steps each. The REMD data from Fig. 3 are shown for ease of comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

longer has no benefit. For the moment, the particular level of effort required to converge the heat capacity seems to be highly system dependent.

In previous studies, accurate heat capacities of alanine dipeptide have not been calculated, and because the curve is *almost* constant varying by only $\sim 2k_B$ over the 800 K temperature range, we believe that a large number of force evaluations are required in order to clearly resolve the curve. Fig. 11 shows the heat capacity estimates using an order of magnitude fewer force evaluations[10] for Nested Sampling, as compared to Fig. 3. Fig. 11 (left) reduces the number of trajectories of each Nested Sampling iteration by a factor of ten and this clearly reduces the quality of the curves generated. Fig. 11 (right), instead, reduces the length of each trajectory by a factor of ten. Although the general shape of the heat capacity can still be resolved, individual C_v curves are of a lower quality than those of Fig. 3. Once more, the previous REMD data is shown for ease of comparison between figures. It is important to note that $\approx 10^9$ force evaluations is still a large number of force evaluations for a system with only 60 internal degrees of freedom, of which very few (notably the dihedrals ϕ and ψ) are not highly constrained.

In the future, we expect to test Galilean Nested Sampling with proteins. If a protein has a well-defined and known tertiary structure, *i.e.* a single dominant free energy minimum, then by starting all replicas of an REMD simulation from this minimum, the amount of equilibration is significantly reduced, as the protein does not need to be folded before investigating its thermodynamics. For example Yeh et al. [65] calculate the heat capacity of an SH3 domain in different implicit solvents, starting trajectories from the crystal structure of the protein. In this case the computational expense associated with Nested Sampling having to discover the native state would be wasted.

However, there has been a lot of recent interest in intrinsically disordered proteins, that is proteins which do not have a well-defined fixed structure, which may, for example, only take well-defined structure upon binding. This interest is because it is now understood that they are significantly more common and important than first thought and perform a variety of biological functions, often related to human disease [66,67]. For these proteins, in equilibrium, there is a distribution over a set of possible macrostates, as is the case for alanine dipeptide. In this case, there is not a single obvious starting conformation for REMD replicas, and hence we believe Nested Sampling, with its top down approach, might be particularly beneficial for the study of the thermodynamics of intrinsically disordered proteins.

5. Conclusion

In this study we have implemented Galilean Nested Sampling for use with the widely used Amber MD package. We have demonstrated the algorithm by sampling alanine dipeptide both *in vacuo* and using a generalised Born implicit solvent model. We have calculated heat capacity curves, and, by comparing our results with those generated by REMD, we have shown that it is possible to achieve good agreement between different sampling algorithms when estimating peptide heat capacity curves.

In this work we sampled Galilean velocities $\mathbf{v} = \mathbf{S}\mathbf{r}$ where $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{S} = \sqrt{k_B T}\mathbf{I}$ with the identity matrix $\mathbf{I}$. In the original description of the algorithm, Skilling suggests that certain choices of 'semimetric' $\mathbf{S}$ could be used to improve Galilean exploration [57]. The reflection formula is then adapted to preserve detailed balance. Specifically, Skilling suggests the semimetric $\mathbf{S} \approx (-\nabla\mathbf{F})^{-1/2}$, where $\mathbf{F}$ are the forces, at a preferred configuration [57]. This semimetric takes into account the curvature of the space when choosing velocities. We believe this improvement would be essential for using Galilean Nested Sampling with larger molecular systems. This is because in molecular systems, certain degrees of freedom, such as the stretching of covalent bonds, are very highly constrained, whereas others, such as the dihedral angles ϕ and ψ, are not very constrained at all. It is clear that the magnitude of velocities in the highly constrained directions should be smaller than those in other directions in order to maximise efficiency. Preliminary results using the isotropic algorithm (*i.e.* $\mathbf{S} \propto \mathbf{I}$) for the penta-peptide Met-enkephalin (not shown) suggest an appropriate semimetric would be essential when using Galilean Nested Sampling with larger biophysical systems.

We conclude that Galilean Nested Sampling, with an appropriate semimetric, is a promising conformational sampling algorithm for biophysical atomistic systems, and we look forward to investigating its performance compared to other general-purpose sampling algorithms (*i.e.* those where no prior knowledge of the PES is required) such as REMD, accelerated MD and multicanonical MD, when sampling larger peptides and proteins.

[10] Note though that we did not reduce the initial equilibration period, as we did not want an unequilibrated initial set to affect the comparison.

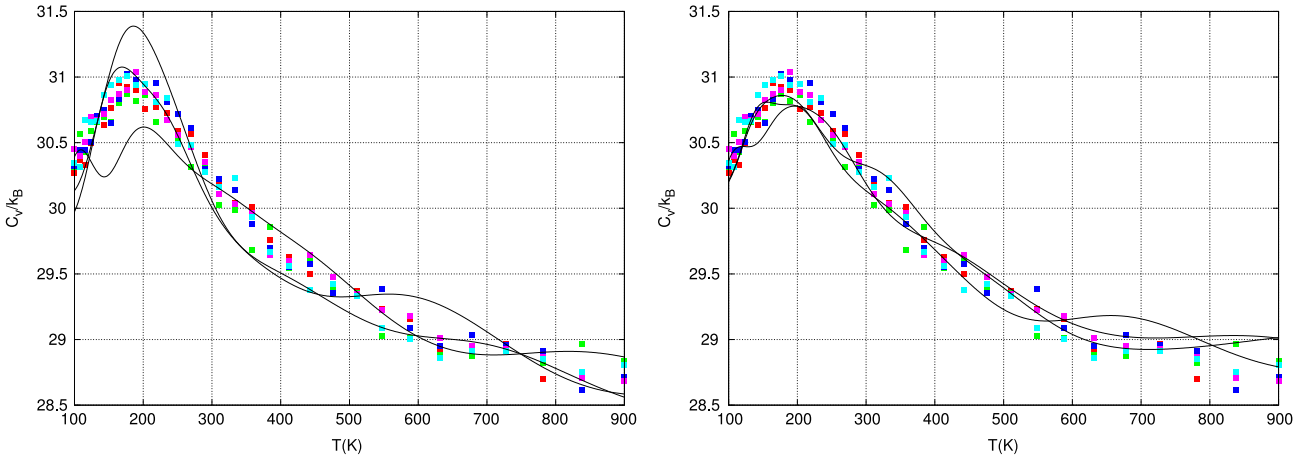

Fig. 11. Heat capacity estimates from Nested Sampling simulations using an order of magnitude fewer force evaluations, as compared to Fig. 3. Left: the number of trajectories of each Nested Sampling iteration by a factor of ten. Right: the length of each trajectory is reduced by a factor of ten. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Acknowledgements

The authors thank John Skilling for helpful discussions. We acknowledge support from the Leverhulme Trust (Grant F/00 215/BL(NSB, CV and DLW)) and the EPSRC (Grants EP/J020281/1 (DLW), EP/J010847/1 (GC) and a Doctoral Training Award (RJB)).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.cpc.2015.12.005.

References

[1] G. Ramachandran, C.T. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1) (1963) 95–99.
[2] V. Madison, K.D. Kopple, Solvent-dependent conformational distributions of some dipeptides, J. Am. Chem. Soc. 102 (15) (1980) 4855–4863.
[3] R. Schweitzer-Stenner, F. Eker, Q. Huang, K. Griebenow, Dihedral angles of trialanine in D_2O determined by combining FTIR and polarized visible raman spectroscopy, J. Am. Chem. Soc. 123 (39) (2001) 9628–9633.
[4] G. Pohl, A. Perczel, E. Vass, G. Magyarfalvi, G. Tarczay, A matrix isolation study on ac-gly–nhme and ac-l-ala–nhme, the simplest chiral and achiral building blocks of peptides and proteins, Phys. Chem. Chem. Phys. 9 (33) (2007) 4698–4708.
[5] R. Schweitzer-Stenner, Distribution of conformations sampled by the central amino acid residue in tripeptides inferred from amide i band profiles and nmr scalar coupling constants, J. Phys. Chem. B 113 (9) (2009) 2922–2932.
[6] J. Grdadolnik, V. Mohacek-Grosev, R.L. Baldwin, F. Avbelj, Populations of the three major backbone conformations in 19 amino acid dipeptides, Proc. Natl. Acad. Sci. USA 108 (5) (2011) 1794–1798.
[7] W.L. Jorgensen, J. Gao, Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution, J. Am. Chem. Soc. 110 (13) (1988) 4212–4216.
[8] T. Head-Gordon, M. Head-Gordon, M.J. Frisch, C. Brooks, J. Pople, A theoretical study of alanine dipeptide and analogs, Inter. J. Quantum Chem. 36 (S16) (1989) 311–322.
[9] D.J. Tobias, C.L. Brooks III, Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results, J. Phys. Chem. 96 (9) (1992) 3864–3870.
[10] B. Montgomery Pettitt, M. Karplus, The potential of mean force surface for the alanine dipeptide in aqueous solution: a theoretical approach, Chem. Phys. Lett. 121 (3) (1985) 194–201.
[11] Y. Miao, W. Sinko, L. Pierce, D. Bucher, R.C. Walker, J.A. McCammon, Improved reweighting of accelerated molecular dynamics simulations for free energy calculation, J. Chem. Theory Comput. arXiv:http://pubs.acs.org/doi/pdf/10.1021/ct500090q, http://dx.doi.org/10.1021/ct500090q. URL http://pubs.acs.org/doi/abs/10.1021/ct500090q.
[12] O.M. Becker, M. Karplus, The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics, J. Chem. Phys. 106 (4) (1997) 1495–1517.
[13] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (9) (2004) 1157–1174.
[14] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., Charmm: the biomolecular simulation program, J. Comput. Chem. 30 (10) (2009) 1545–1614.
[15] S. Antonczak, G. Monard, M.F. Ruiz-López, J.-L. Rivail, Modeling of peptide hydrolysis by thermolysin. a semiempirical and QM/MM study, J. Am. Chem. Soc. 120 (34) (1998) 8825–8833.
[16] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus, A qm/mm implementation of the self-consistent charge density functional tight binding (scc-dftb) method, J. Phys. Chem. B 105 (2) (2001) 569–585.
[17] T. Oie, G.H. Loew, S.K. Burt, J.S. Binkley, R.D. MacElroy, Quantum chemical studies of a model for peptide bond formation: formation of formamide and water from ammonia and formic acid, J. Am. Chem. Soc. 104 (23) (1982) 6169–6174.
[18] D.A. Evans, D.J. Wales, Folding of the gb1 hairpin peptide from discrete path sampling, J. Chem. Phys. 121 (2004) 1080–1090.
[19] I.-C. Yeh, A. Wallqvist, Structure and dynamics of end-to-end loop formation of the penta-peptide cys-ala-gly-gln-trp in implicit solvents, J. Phys. Chem. B 113 (36) (2009) 12382–12390.
[20] J. Hughes, T. Smith, H. Kosterlitz, L. Fothergill, B. Morgan, H. Morris, Identification of two related pentapeptides from the brain with potent opiate agonist activity, Nature 258 (314) (1975) 577–579.
[21] H. Hu, M. Elstner, J. Hermans, Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine dipeptides (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution, Proteins: Struct. Funct. Genet. 50 (3) (2003) 451–463. http://dx.doi.org/10.1002/prot.10279.
[22] H.D. Nguyen, C.K. Hall, Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides, Proc. Natl. Acad. Sci. USA 101 (46) (2004) 16180–16185.
[23] J.L. Banks, G.A. Kaminski, R. Zhou, D.T. Mainz, B.J. Berne, R.A. Friesner, Parametrizing a polarizable force field from ab initio data. i. the fluctuating point charge model, J. Chem. Phys. 110 (2) (1999) 741–754.
[24] O. Winther, A. Krogh, Teaching computers to fold proteins, Phys. Rev. E 70 (3) (2004) 030903.
[25] R.B. Best, G. Hummer, Optimized molecular dynamics force fields applied to the helix–coil transition of polypeptides, J. Phys. Chem. B 113 (26) (2009) 9004–9015.
[26] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1) (1999) 141–151.
[27] H.G. Katzgraber, S. Trebst, D.A. Huse, M. Troyer, Feedback-optimized parallel tempering Monte Carlo, J. Stat. Mech. Theor. Exp. 2006 (03) (2006) P03018.
[28] C. Simmerling, B. Strockbine, A.E. Roitberg, All-atom structure prediction and folding simulations of a stable protein, J. Am. Chem. Soc. 124 (38) (2002) 11258–11259.
[29] M.S. Lee, M.A. Olson, Comparison of two adaptive temperature-based replica exchange methods applied to a sharp phase transition of protein unfolding-folding, J. Chem. Phys. 134 (2011) 244111.
[30] A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method, Phys. Rev. Lett. 100 (2008) 020603. http://dx.doi.org/10.1103/PhysRevLett.100.020603. URL http://link.aps.org/doi/10.1103/PhysRevLett.100.020603.
[31] J. Vymětal, J. Vondrášek, Metadynamics as a tool for mapping the conformational and free-energy space of peptides:the alanine dipeptide case study, J. Phys. Chem. B 114 (16) (2010) 5632–5642.
[32] K. Walczewska-Szewc, E. Deplazes, B. Corry, Comparing the ability of enhanced sampling molecular dynamics methods to reproduce the behavior of fluorescent labels on proteins, J. Chem. Theory Comput. 11 (7) (2015) 3455–3465.
[33] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, J. Comput. Phys. 23 (2) (1977) 187–199.

[34] T. Straatsma, H. Berendsen, Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations, J. Chem. Phys. 89 (9) (1988) 5876–5886.
[35] X. Kong, C.L. Brooks III, λ-dynamics: A new approach to free energy calculations, J. Chem. Phys. 105 (6) (1996) 2414–2423.
[36] D.J. Tobias, S.F. Sneddon, C.L. Brooks III, Reverse turns in blocked dipeptides are intrinsically unstable in water, J. Mol. Biol. 216 (3) (1990) 783–796.
[37] D. Hamelberg, J. Mongan, J.A. McCammon, Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, J. Chem. Phys. 120 (24) (2004) 11919–11929.
[38] B.A. Berg, T. Neuhaus, Multicanonical ensemble: A new approach to simulate first-order phase transitions, Phys. Rev. Lett. 68 (1) (1992) 9–12.
[39] U.H. Hansmann, Y. Okamoto, F. Eisenmenger, Molecular dynamics, Langevin and hydrid Monte Carlo simulations in a multicanonical ensemble, Chem. Phys. Lett. 259 (3) (1996) 321–330.
[40] N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Free energy landscapes of peptides by enhanced conformational sampling, J. Mol. Biol. 296 (1) (2000) 197–216.
[41] J. Higo, N. Ito, M. Kuroda, S. Ono, N. Nakajima, H. Nakamura, Energy landscape of a peptide consisting of α-helix, 310-helix, β-turn, β-hairpin, and other disordered conformations, Prot. Sci. 10 (6) (2001) 1160–1171.
[42] J. Higo, J. Ikebe, N. Kamiya, H. Nakamura, Enhanced and effective conformational sampling of protein molecular systems for their free energy landscapes, Biophys. Rev. 4 (1) (2012) 27–44.
[43] F. Wang, D.P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states, Phys. Rev. Lett. 86 (10) (2001) 2050–2053.
[44] S. Kou, Q. Zhou, W.H. Wong, Equi-energy sampler with applications in statistical inference and statistical mechanics, Ann. Statist. 34 (4) (2006) 1581–1619.
[45] M. Bonomi, M. Parrinello, Enhanced sampling in the well-tempered ensemble, Phys. Rev. Lett. 104 (19) (2010) 190601.
[46] J. Skilling, Bayesian inference and maximum entropy methods in science and engineering, AIP Conf. Proc. 735 (2004) 395–405.
[47] H. Do, J.D. Hirst, R.J. Wheatley, Rapid calculation of partition functions and free energies of fluids, J. Chem. Phys. 135 (17) (2011) 174105.
[48] H. Do, J.D. Hirst, R.J. Wheatley, Calculation of partition functions and free energies of a binary mixture using the energy partitioning method: application to carbon dioxide and methane, J. Phys. Chem. B 116 (15) (2012) 4535–4542.
[49] J. Skilling, Nested Sampling for general Bayesian computation, J. Bayesian Anal. 1 (4) (2006) 833–859.
[50] P. Mukherjee, D. Parkinson, A.R. Liddle, A nested sampling algorithm for cosmological model selection, Astrophys. J. Lett. 638 (2) (2006) L51.
[51] M. Doğruel, T.A. Down, T.J. Hubbard, NestedMICA as an ab initio protein motif discovery tool, BMC Bioinform. 9 (1) (2008) 19–31.
[52] N. Pullen, R.J. Morris, Bayesian model comparison and parameter inference in systems biology using nested sampling, PLOS ONE 9 (2) (2014) e88419.
[53] A. Elsheikh, M. Wheeler, I. Hoteit, Nested sampling algorithm for subsurface flow model selection, uncertainty quantification, and nonlinear calibration, Water Resour. Res. 49 (12) (2013) 8383–8399.
[54] L.B. Pártay, A.P. Bartók, G. Csányi, Efficient sampling of atomic configurational spaces, J. Phys. Chem. B 114 (32) (2010) 10502–10512.
[55] L.B. Pártay, A.P. Bartók, G. Csányi, Nested sampling for materials: The case of hard spheres, Phys. Rev. E 89 (2) (2014) 022302.
[56] N.S. Burkoff, C. Várnai, S.A. Wells, D.L. Wild, Exploring the energy landscapes of protein folding simulations with Bayesian computation, Biophys. J. 102 (4) (2012) 878–886.
[57] J. Skilling, Bayesian computation in big spaces-nested sampling and galilean Monte Carlo, in: Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, vol. 1443, AIP Publishing, Melville, NY, USA, 2012, pp. 145–156.
[58] A.E. Raftery, M.A. Newton, J.M. Satagopan, P.N. Krivitsky, Estimating the Integrated Likelihood via Posterior Simulation using the Harmonic Mean Identity, Oxford University Press, Oxford, UK, 2007.
[59] A.A. Podtelezhnikov, D.L. Wild, Crankite: A fast polypeptide backbone conformation sampler, Source Code Biol. Med. 3 (1) (2008) 1–7.
[60] C. Várnai, N.S. Burkoff, D.L. Wild, Efficient parameter estimation of generalizable coarse-grained protein force fields using contrastive divergence: a maximum likelihood approach, J. Chem. Theory Comput. 9 (12) (2013) 5718–5733.
[61] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics, J. Am. Chem. Soc. 112 (16) (1990) 6127–6129.
[62] Supplementary material for details can be found online at http://dx.doi.org/10.1016/j.cpc.2015.12.005.
[63] Z.-X. Wang, W. Zhang, C. Wu, H. Lei, P. Cieplak, Y. Duan, Strike a balance: optimization of backbone torsion parameters of amber polarizable force field for simulations of proteins and peptides, J. Comput. Chem. 27 (6) (2006) 781–790.
[64] K. Lindorff-Larsen, P. Maragakis, S. Piana, M.P. Eastwood, R.O. Dror, D.E. Shaw, Systematic validation of protein force fields against experimental data, PLoS One 7 (2) (2012) e32131.
[65] I.-C. Yeh, M.S. Lee, M.A. Olson, Calculation of protein heat capacity from replica-exchange molecular dynamics simulations with different implicit solvent models, J. Phys. Chem. B 112 (47) (2008) 15064–15073.
[66] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic, Intrinsically disordered protein, J. Mol. Graphics Modell. 19 (1) (2001) 26–59.
[67] D. Eliezer, Biophysical characterization of intrinsically disordered proteins, Curr. Opin. Struct. Biol. 19 (1) (2009) 23–30.