Chemical Physics Letters 625 (2015) 1–4

Chemical Physics Letters

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

Free energy basin-hopping

K.H. Sutherland-Cash [a], D.J. Wales [a,*], D. Chakrabarti [b]

[a] University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK
[b] School of Chemistry, University of Birmingham, Birmingham B15 2TT, UK

ARTICLE INFO

Article history:
Received 13 December 2014
In final form 9 February 2015
Available online 17 February 2015

ABSTRACT

A global optimisation scheme is presented using basin-hopping with the acceptance criterion based on approximate free energy for the corresponding local minima of the potential energy. The method is illustrated for atomic and colloidal clusters and peptides to examine how the predicted global free energy minimum changes with temperature. Using estimates for the local free energies based on harmonic vibrational densities of states provides a computationally effective framework for predicting trends in structure at finite temperature. The resulting scheme represents a powerful tool for exploration of energy landscapes throughout molecular science.

1. Introduction

Global optimisation is an important tool for structure prediction throughout molecular science, as well as for soft matter and condensed matter systems. In particular, numerous methods have been described that aim to identify the global minimum on the potential energy surface (PES) [1–4]. Although the PES of a many-body system is high dimensional and often complex, searches on the PES can usually be based on accurate evaluation of the potential energy function. In contrast, if the temperature is high enough for entropy to play a role, then further sampling may be necessary. Weak interactions that correlate with low frequency vibrational modes can be important here, suggesting that a systematic approach to investigate such effects could be very useful. However, we do not wish to incur the computational expense of constructing entire free energy landscapes based upon projections onto chosen order parameters, or sample extensively in a more global fashion [5–11]. Instead we prefer to focus on the local free energy associated with particular minima on the potential energy surface, and investigate how the competition between alternative structures changes with temperature. We do not propose to compute the free energy for states associated with sets of potential energy minima. Such states would be important for analysis of phase transitions, where the high entropy phase is stabilised by landscape entropy, i.e. the configuration space associated with multiple minima that lie relatively high in potential energy. Rather, we consider the regime where structure is still well defined in terms of competition between configurations associated with distinct potential energy minima. Cross-overs in stability with temperature are then associated with the local well entropies, which will increase more rapidly with temperature for structures that have lower vibrational frequencies. In fact, we could also lump such minima, for example, using a free energy barrier threshold [12]. This sort of procedure has been used to construct free energy disconnectivity graphs [10,13], but it requires additional information to define connectivity.

In the present contribution we present a methodology that provides an estimate of free energy of the local minima on the PES on the fly as they are visited during global optimisation. To this end, we adapt the basin-hopping (BH) global optimisation method [2,14,15], which has been used for a wide variety of molecular and soft and condensed matter, to explore the *free* energy landscape, as opposed to the *potential* energy landscape. Our approach relies upon the superposition approach [16–21] to thermodynamics, which expresses the global partition function, $Z(T)$, in terms of contributions from the catchment basin of each local minimum. In the canonical ensemble, which is the focus of this letter, we have

$$Z(T) = \sum_i Z_i(T), \tag{1}$$

where $Z_i(T)$ is the partition function of minimum i at temperature T. Z_i thus only draws contributions from the catchment basin [22,18] for which steepest-descent paths converge to minimum i. Here we will employ the harmonic normal mode approximation to obtain the vibrational density of states for each minimum. This harmonic superposition approach has been used successfully for a variety of applications in previous work [21,23–26]. Various schemes to treat anharmonic corrections have been proposed [17,20,25,27–30], but we anticipate that the FEBH approach will generally be applied using harmonic vibrational densities of states

* Corresponding author.

http://dx.doi.org/10.1016/j.cplett.2015.02.015
0009-2614/

in the first instance. Once such a survey has been conducted, it would certainly be possible to admit vibrational anharmonicity, and the procedure that employs a model anharmonic form used in the basin-sampling approach to thermodynamics [25] could provide an attractive framework. We will report on such extensions in future work; the harmonic approximation adopted in the present work will probably lead to a systematic underprediction of crossover temperatures, since the softer modes of higher energy minima are expected to be more anharmonic. Corresponding shifts in heat capacity peaks and melting points have been analysed in previous work [17,27].

As each permutation-inversion isomer of a given structure, $Z_i(T)$ is identical, the sum above can be written as the product of the partition function for one of these isomers and a factor n_i, which accounts for the permutation-inversion degeneracy. For a system containing N_A atoms of element A, N_B atoms of element B, etc., $n_i = 2N_A!\,N_B!\,N_C!\ldots/o_i$, where o_i is the order of the point group of minimum i [18,31–33]. Ignoring the rotational contribution to the partition function, which does not usually vary significantly between different isomers of the same atomic cluster, and making a harmonic approximation to the vibrational component of the partition function, one can write

$$Z(T) = \sum_i \frac{n_i \exp[-\beta V_i]}{(\beta h \bar{\nu}_i)^\kappa}, \tag{2}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, κ is the number of non-zero eigenvalues for the Hessian matrix, $\bar{\nu}_i = (\prod_{j=1}^{\kappa} \nu_i^j)^{1/\kappa}$ is the geometric mean vibrational frequency of minimum i, with ν_i^j the normal mode frequency of the j-th mode in minimum i, and V_i is the potential energy of the minimum i. The approximate free energy of minimum i is then given by $F_i(T) = -k_B T \ln Z_i(T)$. It is apparent from Eq. (2) that the effect of temperature will come into play most rapidly for competing structures with different point group symmetries or mean vibrational frequencies. The rotational partition function can easily be included if required, assuming a rigid rotor model for the density of states, but we have checked that it is not important for the examples considered below.

In the free energy basin-hopping (FEBH) framework introduced here the local free energies of the two minima are used in the accept/reject test for step-taking between local minima. The key formula is

$$F_{\rm new}(T) - F_{\rm old}(T) = V_{\rm new} - V_{\rm old} + k_B T \ln \frac{o_{\rm new} \bar{\nu}_{\rm new}^\kappa}{o_{\rm old} \bar{\nu}_{\rm old}^\kappa}. \tag{3}$$

The product of vibrational normal mode frequencies can be obtained from the product of non-zero eigenvalues of the mass-weighted Hessian.

The FEBH approach has now been implemented in the `GMIN` programme [34]. To illustrate the utility of this scheme and provide proof of principle we have tested it for various atomic and colloidal clusters, along with some peptides to provide examples for biomolecules. In future work we will report on applications to much larger biomolecules, with particular focus on systems involving ligand binding. In particular, we will optimise the efficiency of the local free energy calculations, which involve the additional expense entailed by a normal mode analysis. Although it is also possible to calculate the local harmonic free energies for a sample of potential energy minima obtained after a conventional basin-hopping run, we find examples where FEBH locates the local free energy global minimum using fewer steps, on average.

2. Applications

Lennard-Jones (LJ) clusters have been extensively studied for benchmarking global optimisation methods [15,35]. The Lennard-Jones potential can be written as $V(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, where ϵ is the pair well depth and σ is the separation r for which $V(r)$ vanishes. Figure 1a shows the free energy as a function of temperature for the minima of the seven-atom Lennard-Jones cluster, LJ_7, for which there are only four distinct structures. The global potential energy minimum corresponds to a pentagonal bipyramidal structure (D_{5h}), and the other three local minima are the capped octahedron (C_{3v}), tricapped tetrahedron (C_{3v}), and skew structure (C_2) in ascending order of potential energy [36,37]. The respective orders of the point group are 20, 6, 6, and 2. It is evident that the skew structure, which possesses the lowest symmetry, becomes relatively more stable as temperature increases, and is the second most stable structure after the pentagonal bipyramid at $T = 0.25$ in units of ϵ/k_B.

It is instructive to compare the case of a seven-particle colloidal cluster, bound by a short-ranged Morse potential, M_7. The Morse potential is given by $V(r) = \epsilon \exp[\rho(1 - r/r_e)](\exp[\rho(1 - r/r_e)] - 2)$, where ϵ is the pair well depth, r_e is the pair equilibrium separation, and ρ is a parameter that controls the range of the potential. With the range parameter $\rho = 30$, the Morse potential was found to provide a quantitative description of occupation probability distribution for competing structures of polystyrene microspheres as measured by optical microscopy [38,39]. For $\rho = 30$, the global minimum on the PES is also the pentagonal bipyramid, as shown in Figure 2a, whilst three other minima have the same number of nearest neighbour contacts and are of comparable potential energies. Our global optimisation runs confirm that the M_7 cluster in this case has more minima as compared to the LJ_7 cluster, which is due to the short-ranged pair interactions for $\rho = 30$. The relative stabilities of these structures are therefore more sensitive to

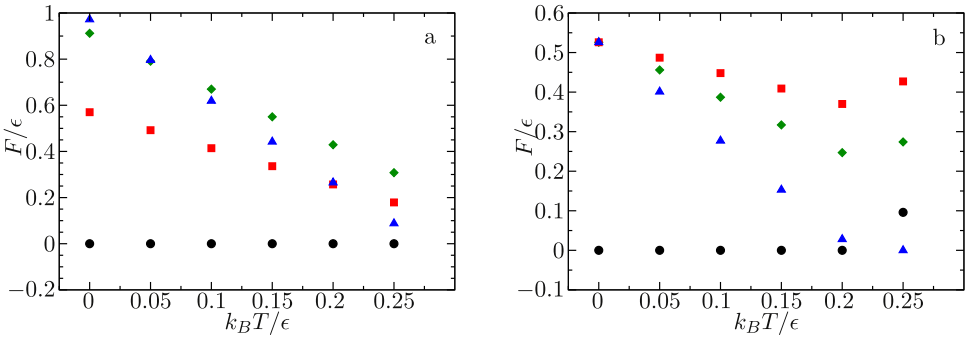

Figure 1. Free energy of the minima as a function of temperature for the Lennard-Jones cluster LJ_7 and the Morse cluster M_7 with range parameter $\rho = 30$. (a) The LJ_7 cluster has four distinct minima: the pentagonal bipyramid (circle), capped octahedron (square), tricapped tetrahedron (diamond), and skewed structure (triangle). (b) The M_7 cluster has four low-lying minima with the same number of nearest neighbour contacts: pentagonal bipyramid (circle), capped octahedron (square), tricapped tetrahedron (diamond), and bicapped trigonal bipyramid (triangle). The free energies are presented relative to that of the most stable structure, which is set to zero at all temperatures.

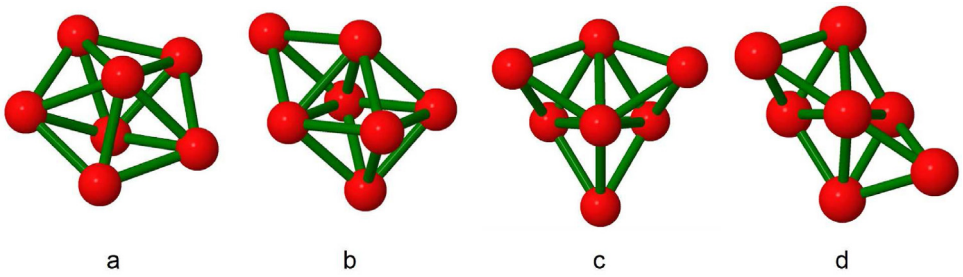

Figure 2. The four minima of the M_7 cluster: pentagonal bipyramid (a), capped octahedron (b), tricapped tetrahedron (c) and bicapped trigonal bipyramid (d).

temperature than for LJ_7, as reflected in the corresponding free energies (Figure 1b). At $T = 0.25$ in reduced units, the bicapped trigonal bipyramid becomes the most stable structure. For both the LJ_7 and M_7 clusters, the structure with lowest symmetry becomes relatively more stable as the temperature increases because of the larger number of distinct permutation-inversion isomers, and the correspondingly greater landscape entropy [38,39].

For biological systems, two peptides were chosen to evaluate the performance of FEBH, namely alanine dipeptide and the serine-lysine peptide. Alanine dipeptide is an alanine amino acid with two methyl-capped peptide bonds. It has been the subject of many simulation efforts, owing to the fact that it possesses very few degrees of freedom and its free energy landscape is relatively simple. Serine-lysine was chosen because it has the additional potential for hydrogen-bonding and other interactions between its sidechains, whilst still having relatively few degrees of freedom.

We sampled temperatures in the range from 0 K to 503 K in increments of 50.3 K. This range was chosen because our units of temperature correspond to the energy function, in this case kcal mol^{-1} (1 kcal mol^{-1} = 50.3 K). We tuned the BH parameters (maximum Cartesian step size, Metropolis acceptance temperature, and target acceptance ratio) to minimise the mean first encounter time of the global minimum for simulations at 302 K, the value that is most biologically relevant.

For comparison with earlier results [40], we used the AMBER ff03 force field [41] with the Generalised Born implementation developed by Onufriev et al. [42] and smoothed cut-offs. This smoothing is implemented in the programme GMIN [34] and is necessary to remove discontinuities in the potential and first derivatives, which would otherwise hinder convergence to local minima. We also employed a properly symmetrised representation, where the energy is invariant to feasible exchange of atoms of the same element [43].

The criterion used for convergence (i.e. whether we have found a true local minimum) is typically based on root mean square (RMS) force. However, in these simulations, it was important to ensure that minima possessed non-zero eigenvalues that were well separated from the six zero eigenvalues corresponding to overall translation and rotation. This check was achieved by calculating the lowest eigenvalues and reminimising structures with too low a separation (a factor of 10^3 in units of frequency was chosen) using a tighter RMS convergence condition. This approach has the added advantage of allowing us to identify and remove any transition states or saddles from our results, as they possess one or more negative, non-zero eigenvalues.

For the two peptides all the local minima exhibit C_1 symmetry. Thus, temperature effects on stability arise exclusively from the differences in the normal mode spectra.

Within the range of temperatures chosen, alanine dipeptide exhibits very simple behaviour. This result is due to the relatively small number of physically relevant minima (seven were characterised using this force field and solvent model) and the larger potential energy spacing between the global minimum and higher energy minima. Indeed, cross-overs in the approximate global free energy minimum for this system (where the free energy of the second-lowest *potential* energy minimum becomes lower than that

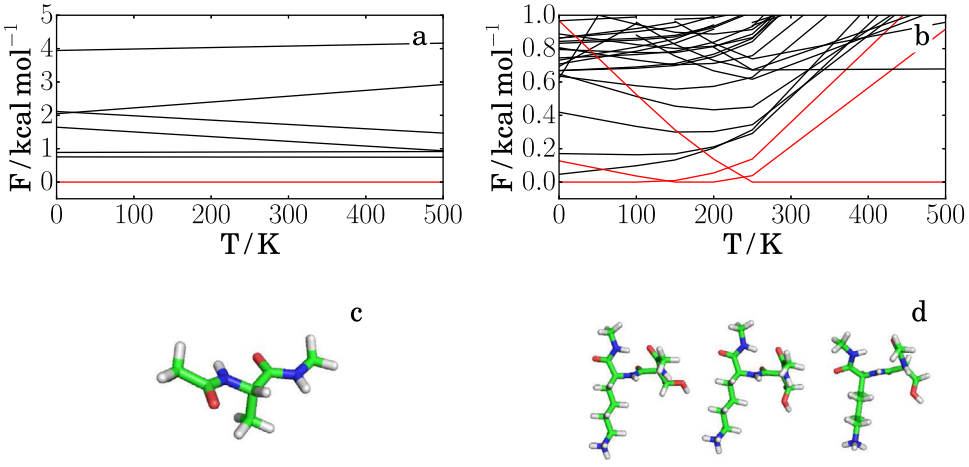

Figure 3. Free energy of the minima as a function of temperature for alanine dipeptide and the serine-lysine dipeptide. Lines for the minima that become the global minimum at different temperatures are highlighted in red. (a) Alanine dipeptide has a relatively simple energy spectrum, with only 7 physically relevant local minima and two instances of minima reordering. (b) Minima for serine-lysine dipeptide with energies within 1.0 kcal mol^{-1} of the global minimum at the corresponding temperature. It can clearly be seen that the minima undergo significant reordering and three different global minima are observed over this temperature range. As before, the free energies are presented relative to the global minimum, which is chosen to have zero energy at all temperatures. (c) The structure of the alanine dipeptide global minimum. (d) The structures of the serine-lysine global minima at increasing temperatures from left to right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

of the global potential energy minimum) in the harmonic approximation would occur at 1170 K. In general, the harmonic prediction for the crossover temperature between two minima is

$$k_B T_{xo} = \frac{V_1 - V_2}{\ln((o_2 \bar{\nu}_2^{\kappa})/(o_1 \bar{\nu}_1^{\kappa}))}, \quad (4)$$

from Eq. (3), which clearly illustrates the balance between potential energy and well entropy.

In contrast, the greater number of possible interactions and larger number of degrees of freedom mean that the serine-lysine dipeptide has many more minima within a narrower range of energies than alanine dipeptide. This structure results in a complex energy spectrum with frequent reordering of higher energy minima and changes in the global free energy minimum three times within the temperature range explored (Figure 3). However, this complexity increases the likelihood of overlap between neighbouring harmonic basins, which would result in overestimation of the true density of states from double counting, and a correspondingly lower free energy.

It is worth noting that whilst harmonic contributions to the free energy are approximately linear with respect to temperature, the *relative* energies of the higher energy minima with respect to the global minimum are non-linear. This effect occurs because the identity of the global minimum changes and with it, the reference zero energy structure.

3. Conclusions

We have described a free energy basin-hopping (FEBH) global optimisation approach based on a harmonic approximation to the local vibrational partition function of individual potential energy minima. The FEBH scheme allows for the free energy surface to be explored efficiently in terms of structures that correspond to potential energy minima, maintaining all the advantages of basin-hopping, whilst approximating free energy contributions in a computationally affordable manner. For systems where analytical second derivatives of the potential are relatively simple, the FEBH scheme is especially efficient. For example, for the M_7 cluster and serine-lysine dipeptide, FEBH runs take about 10% more time than the corresponding potential energy based basin-hopping runs. This increased time arises both from the requirement to calculate analytical second derivatives, and from the time required to diagonalise the Hessian matrix. The scaling of the diagonalisation routine ($\mathcal{O}(N^3)$) could dominate the simulation time for larger, force field based simulations. Of course, the overhead for calculating second derivatives with potential energy functions of greater computational complexity, such as quantum mechanical approaches, could be significant, especially if numerical second derivatives are used. The harmonic approximation will also lead to systematic underestimation of vibrational densities of states for low frequency modes, and hence the predicted crossover temperatures are likely to be overestimates, and should therefore be regarded as likely upper bounds for guiding more accurate calculations of thermodynamic properties. Corresponding shifts in heat capacity peaks and melting points within the harmonic approximation have been characterised in previous work [17,27]. In fact, for some systems, such as the M_7 cluster, the free energy surface can differ significantly from the potential energy surface even at relatively low temperatures.

The results presented here were all obtained from basin-hopping runs at the free energy level. Of course, it is also straightforward to calculate the corresponding local free energies for an existing sample of local potential energy minima, and we have checked all the present results in this way. However, locating the low free energy minima can take longer if we only explore the potential energy landscape and perform the local free energy analysis a posteriori. For example, the global free energy minimum for serine-lysine at 500 K was found after an average of 1478 steps in the corresponding FEBH run. At the lower temperature of 250 K, finding this minimum took an average of 2185 steps and, using only the potential energy, 4011 steps. Clearly there will be a balance between the additional overhead of calculating free energy (dependent on the potential) and efficient landscape exploration.

We anticipate that free energy basin-hopping will find widespread application in molecular, soft matter, and condensed matter science.

Acknowledgements

We are grateful to the EPSRC and the ERC for financial support under grants EP/1001352/1 and 267369, respectively.

References

[1] S. Kirkpatrick, J.C.D. Gellat, M.P. Vecchi, Science 220 (1983) 671.
[2] D.J. Wales, H.A. Scheraga, Science 285 (1999) 1368.
[3] R.L. Johnston, Dalton Trans. (2003) 4193.
[4] S.M. Woodley, R. Catlow, Nat. Mater. 7 (2008) 937.
[5] G.M. Torrie, J.P. Valleau, J. Comp. Phys. 23 (1977) 187.
[6] A.P. Lyubartsev, A.A. Martinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, J. Chem. Phys. 96 (1992) 1776.
[7] B.A. Berg, T. Neuhaus, Phys. Rev. Lett. 68 (1992) 9.
[8] F. Wang, D.P. Landau, Phys. Rev. Lett. 86 (2001) 2050.
[9] A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12562.
[10] S.V. Krivov, M. Karplus, J. Chem. Phys. 117 (2002) 10894.
[11] D. Gfeller, P.D.L. Rios, A. Caflisch, F. Rao, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1817.
[12] J.M. Carr, D.J. Wales, J. Phys. Chem. B 112 (2008) 8760.
[13] D.A. Evans, D.J. Wales, J. Chem. Phys. 118 (2003) 3891.
[14] Z. Li, H.A. Scheraga, Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 6611.
[15] D.J. Wales, J.P.K. Doye, J. Phys. Chem. A 111 (1997) 5111.
[16] F.H. Stillinger, T.A. Weber, Science 225 (1984) 983.
[17] D.J. Wales, Mol. Phys. 78 (1993) 151.
[18] D.J. Wales, Energy Landscapes, Cambridge University Press, Cambridge, 2003.
[19] F.H. Stillinger, Science 267 (1995) 1935.
[20] B. Strodel, D.J. Wales, Chem. Phys. Lett. 466 (2008) 105.
[21] V.A. Sharapov, D. Meluzzi, V.A. Mandelshtam, Phys. Rev. Lett. 98 (2007) 105701.
[22] P.G. Mezey, Theor. Chim. Acta 58 (1981) 309.
[23] V.A. Sharapov, V.A. Mandelshtam, J. Phys. Chem. A 111 (2007) 10284.
[24] S. Martiniani, J.D. Stevenson, D.J. Wales, D. Frenkel, Phys. Rev. X 4 (August) (2014) 031034.
[25] D.J. Wales, Chem. Phys. Lett. 584 (2013) 1.
[26] K. Mochizuki, C.S. Whittleston, S. Somani, H. Kusumaatmaja, D.J. Wales, Phys. Chem. Chem. Phys. 16 (2014) 2842.
[27] J.P.K. Doye, D.J. Wales, J. Chem. Phys. 102 (1995) 9659.
[28] J.P.K. Doye, D.J. Wales, J. Chem. Phys. 102 (1995) 9673.
[29] F. Calvo, J.P.K. Doye, D.J. Wales, J. Chem. Phys. 115 (2001) 9627.
[30] I. Georgescu, V.A. Mandelshtam, J. Chem. Phys. 137 (2012) 144106.
[31] F.G. Amar, R.S. Berry, J. Chem. Phys. 85 (1986) 5943.
[32] M.K. Gilson, K.K. Irikura, J. Phys. Chem. B 114 (2010) 16304.
[33] F. Calvo, J.P.K. Doye, D.J. Wales, Nanoscale 4 (2012) 1085.
[34] D.J. Wales, GMIN, 2010 (Updated 19.02.10).
[35] M.T. Oakley, R.L. Johnston, D.J. Wales, Phys. Chem. Chem. Phys. 15 (2013) 3965.
[36] J.B. Kaelberer, R.-B. Etters, J.C. Raich, Chem. Phys. Lett. 41 (1976) 580.
[37] S. Schelstraete, H. Verschelde, J. Chem. Phys. 108 (1998) 7152.
[38] G. Meng, N. Arkus, M.P. Brenner, V.N. Manoharan, Science 327 (2010) 560.
[39] D.J. Wales, ChemPhysChem 11 (2010) 2491.
[40] H. Kusumaatmaja, C.S. Whittleston, D.J. Wales, J. Chem. Theory Comput. 8 (2012) 5159.
[41] Y. Duan, et al., J. Comput. Chem. 24 (2003) 1999.
[42] A. Onufriev, D. Bashford, D.A. Case, Proteins: Struct. Funct. Bioinform. 55 (2004) 383.
[43] E. Małolepsza, B. Strodel, M. Khalili, S. Trygubenko, S. Fejer, D.J. Wales, J. Comp. Chem. 31 (2010) 1402.