J Glob Optim (2009) 43:429–444
DOI 10.1007/s10898-007-9206-5

Molecular conformation of *n*-alkanes using terrain/funneling methods

Rajeswar R. Gattupalli · Angelo Lucia

Received: 3 April 2007 / Accepted: 9 June 2007 / Published online: 27 July 2007

Abstract Understanding molecular conformation is a first step in understanding the waxing (or formation of crystals) of petroleum fuels. In this work, we study the molecular conformation of typical fuel oils modeled as pure *n*-alkanes. A multi-scale global optimization methodology based on terrain methods and funneling algorithms is used to find minimum energy molecular conformations of united atom *n*-alkane models for diesel, home heating, and residual fuel oils. The terrain method is used to gather average gradient and average Hessian matrix information at the small length scale while funneling is used to generate conformational changes at the large length scale that drive iterates to a global minimum on the potential energy surface. In addition, the funneling method uses a mixture of average and point-wise derivative information to produce a monotonically decreasing sequence of objective function values and to avoid getting trapped at local minima on the potential energy surface. Computational results clearly show that the calculated united atom molecular conformations are comprised of zigzag structure with considerable wrapping at the ends of the molecule and that planar zigzag conformations usually correspond to saddle points. Furthermore, the numerical results clearly demonstrate that our terrain/funneling approach is robust and fast.

Keywords Multi-scale global optimization · Terrain methods · Funneling methods · *n*-alkane molecular conformation · Fuel oils

1 Introduction

Waxing of petroleum or hydrocarbon fuels such as home heating and diesel fuels is a nagging and costly problem in the petroleum industry. Billions of dollars are spent annually on additives to these fuels to prevent waxing or crystal formation and on production, operatio-

R. R. Gattupalli · A. Lucia (✉)
Department of Chemical Engineering, University of Rhode Island,
Kingston, RI 02881-0805, USA
e-mail: lucia@egr.uri.edu

nal, and transportation difficulties that arise due to wax formation. This problem is further exacerbated by the fact that there are a large number of phase transitions that, in turn, can lead to a variety of crystal structures. The early literature on the structure of n-alkanes (see, for example [1]) describes the molecular conformation of n-alkanes up to about C_{50} as a zigzag structure with planar chains of methylene (or CH_2) units terminated by methyl groups (CH_3). The orientation between the methyl end (or side) groups and the planar chain is either perpendicular or tilted. Crystal lattice structures are determined by the energy associated with both the horizontal and vertical arrangement of these zigzag molecules. However, more recent studies using molecular dynamics [2] show that even single crystal structures of large n-alkanes are not planar zigzag structures but show considerable wrapping at the ends of the molecule.

In this paper, we study the molecular conformation of pure n-alkanes from the perspective of global optimization. The primary motivation for this computational study is that of a first step in quantifying the interplay between molecular conformation and crystal structure in n-alkane mixtures. The long-term goal is to obtain a transition state theory description of waxing dynamics with and without additives. This paper is organized in the following way. We first provide a mathematical statement of the molecular conformation problem for n-alkanes. Next we give a brief overview of the global optimization literature and then describe the terrain and funneling methods used in this work. Following this, we present a multi-scale global optimization methodology based on terrain and funneling methods. Next, numerical results for the molecular conformation of typical fuels (diesel, home heating, and residual fuels) are presented. We end our paper with some conclusions with regard to the numerical performance of the proposed multi-scale global optimization method and its suitability for solving problems in molecular conformation. However, we remark that the reader should keep in mind that we offer no rigorous proof of convergence of the terrain/funneling approach to a global optimum.

2 Problem statement

In this section, we describe the global optimization problem for the molecular conformation of n-alkanes.

2.1 Objective function

In this work, we use a united atom potential energy function as the objective function, which simply means that we assume that any n-alkane molecule is comprised of a string of methyl (CH_3) and methylene (CH_2) groups. Thus the methylene and methyl groups form the basic units of molecular conformation. Each of these united atoms in a molecule has Cartesian coordinates and the corresponding potential energy function of this n-alkane molecule takes the general form of a sum of bond energy, angle energy, torsion energy, and non-bonded (or van der Waals) energy contributions given by

$$E = E_b + E_a + E_t + E_{vdW} \tag{1}$$

2.1.1 Bond energy

The bond energy, E_b, is a harmonic (or quadratic) function given by

$$E_b = k_b \left(r - r_{eq}\right)^2 \tag{2}$$

where r denotes Euclidean distance, $k_b = 350\,\text{kcal/mol A}^2$ and $r_{eq} = 1.53\,\text{A}$, and where the symbol A is Angstroms (10^{-10} m).

2.1.2 Angle energy

The angle energy, E_a, formed by any three adjacent united atoms is

$$E_a = K_\theta \left(\theta - \theta_{eq}\right)^2 \tag{3}$$

where θ is an angle, $k_\theta = 60\,\text{kcal/mol rad}^2$ and $\theta_{eq} = 1.91$ radians (or 109.5 degrees).

2.1.3 Torsion energy

The torsion energy, E_t, between any four united atoms is

$$E_t = 0.5\,\{k_{t1}[1 - \cos\phi] + k_{t2}[1 - \cos(2\phi)] + k_{t3}[1 - \cos(3\phi)]\} \tag{4}$$

where $k_{t1} = 1.62\,\text{kcal/mol}$, $k_{t2} = -0.867\,\text{kcal/mol}$, and $k_{t3} = 3.24\,\text{kcal/mol}$.

2.1.4 Non-Bonded or van der Waals Energy

Non-bonded energy, E_{vdW}, between united atoms separated by more than three bonds is modeled using the Lennard-Jones 6–12 potential function, which is given by

$$E_{vdW} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{5}$$

where $\varepsilon = 0.112\,\text{kcal/mol}$ and $\sigma = 4.01$ A for either CH_2 or CH_3 groups.

All parameters in the foregoing energy expressions used in this work apply to *both* methyl and methylene groups and were taken from [3], where it is shown that a united atom approximation of n-alkane molecules using methyl and methylene groups provides a good match with experimental spin lattice relaxation time and nuclear Overhauser enhancement data.

2.2 Unknown variables

The unknown variables are the Cartesian or (x, y, z) coordinates of the united atoms in the molecule and are related to r, θ and ϕ through the following formulae. The distance between any pair of united atoms, say i and j, is given by

$$r = \left[(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right]^{1/2} \tag{6}$$

The angle formed by any three adjacent united atoms, i, j and k, is

$$\theta = \cos^{-1}\left\{\frac{[v_i - v_j]^T[v_k - v_j]}{||v_i - v_j||.||v_k - v_j||}\right\} \tag{7}$$

where $v_i = (x_i, y_i, z_i)$ are the Cartesian coordinates of the ith united atom. Finally, the torsion angle, ϕ, between any four adjacent united atoms is

$$\phi = \cos^{-1}\left\{\frac{u^T w}{||u||.||w||}\right\} \tag{8}$$

where $u = [v_i - v_j] \times [v_k - v_j]$ and $w = [v_l - v_k] \times [v_j - v_k]$.

2.3 Constraints

In any molecular conformation problem, it is helpful to fix the coordinate frame to avoid translation and rotational singularities. This can be accomplished using the following linear constraints

$$x_1 = y_1 = z_1 = 0; \quad y_2 = z_2 = 0 \quad \text{and} \quad z_3 = 0 \tag{9}$$

Equation 9 simply states that united atom 1 is fixed at the origin, united atom 2 lies along the x-axis, and united atom 3 lies in the $x - y$ plane. All other united atoms are free to lie in R^3. While it is also possible to further constrain the Cartesian coordinates of the united atoms so they lie in the positive orthant by using the additional conditions that $x_i, y_i, z_i \geq 0$, for all atoms, these conditions do not change the results because all symmetries give symmetric portions of the objective function landscape and thus all optimization calculations are the same.

Remarks It is important to note that the torsion and van der Waals energy terms have a tendency to create many minima, saddle points and singular points (or roughness) on the potential energy surface at the small length scale. However, at the large length scale, the objective function is typically funnel shaped. It is this disparity in geometry at the small and large length scales that is the primary reason that multi-scale global optimization methods are particularly useful in solving molecular conformation problems. We also note that the linear constraints given by Eq. 9 can be eliminated and the optimization problem re-cast in the general form of finding the global minimum of the potential energy function subject to simple bounds on Cartesian coordinates.

3 Optimization methods

There are many deterministic and/or stochastic methods that can be used to solve molecular conformation problems. These methods can be characterized as either general-purpose methods or algorithms that are specifically designed for the molecular conformation problems. General purpose deterministic methods include the tunneling method of Levy and Montalvo [3], the TRUST algorithm of Bahren and Protopopescu [4], and the αBB method of Maranas and Floudas [5]. Floudas and co-workers have applied the αBB method to a number of molecular conformation problems [6–10]. There are also homotopy-continuation algorithms (e.g., Sun and Seider [11]), the diffusion equation method of Piela et al. [12], the interval methods of Hansen [13] and Lin and Stadtherr [14], and terrain methods [15–17] and funneling methods [18] of Lucia and co-workers. General purpose stochastic methods like simulated annealing [19,20], genetic algorithms [21,22], as well as the stochastic differential equations of Aluffi-Pentini et al. [23] and the stochastic approach of Bilbro [24] can also be used to solve molecular conformation problems. Specialized techniques such as chain of states methods (Sevick et al. [25]), build-up procedures [26–28], the threading method [29], and the constrained simplex method [30] also find application in solving molecular conformation problems—as do the path following algorithm of Cerjan and Miller [31], Baker's modified rational function approach [32] for finding saddle points and the nudged elastic band method of Henkelman et al. [33]. In addition, specialized genetic algorithms have been developed for Lennard-Jones [34] and other types of atomic clusters [35]—as have the basin hopping methods of Wales and co-workers [36,37] and the double ended trajectory method of Matro et al. [38]. Finally there are methods that attempt to combine the best of stochastic

and deterministic methodologies such as the method of Gregurick et al. [39], which uses both a conjugate gradient method and a genetic algorithm and is quite similar to basin hopping. In this paper, we use terrain and funneling methods for multi-scale global optimization.

3.1 A terrain method for optimization at the small length scale

Terrain methods have been described in detail elsewhere in the literature [15–17] and are briefly summarized here for the purpose of continuity. Terrain methods are used to locate sets of stationary and singular points of a general C^3 objective function, say $f = f(z)$, subject to bounds on variables, $c(z)$, where z are the optimization variables. They do this by following valleys and moving up and down the landscapes of $g^T g$ and f, where $g = g(z)$ is the gradient of f. The method used for downhill calculations, either initially or when moving downhill from a saddle point to a minimum, is a standard Newton-based trust region method. Uphill calculations, on the other hand, follow the Newton vector field along the valley. The terrain method compensates for any drift from a valley, V, by intermittently solving a sequence of general nonlinearly, constrained optimization problems

$$V = \{\min g^T h^T h g \text{ such that } g^T g = L, \text{ for all } L \,\varepsilon\, \Lambda\} \tag{10}$$

where $h = h(z)$ denotes the Hessian matrix of f, L is any given value (or level) of the least-squares objective function, and Λ is some collection of contours. The nonlinear programming problem in Eq. 10 defines what are called uphill corrector steps. These corrector steps serve the purpose of forcing iterates back to the valley and are solved using a successive quadratic programming (SQP) method. At any stationary point, eigenvalues and eigenvectors are calculated in order to determine the character of that stationary point and whether the next direction should be uphill or downhill. In particular, uphill movement from any minimum is always initiated in the eigen-direction associated with the smallest positive eigenvalue while the eigen-direction associated with the largest negative eigenvalue is used to move downhill from any saddle point. Our terrain methodology also keeps track of where it has been on the objective function landscape and, in this way, avoids re-calculating previously determined stationary points. Finally, termination occurs when the algorithm encounters a pair of bounds on one or more of the optimization variables. Thus in summary, terrain methods require

(1) Reliable downhill equation solving.
(2) Reliable and efficient computation of singular points.
(3) Efficient uphill movement comprised of predictor-corrector calculations.
(4) Reliable and efficient eigenvalue-eigenvector computations.
(5) Effective bookkeeping.
(6) A termination criterion to decide when the computations have finished.

In this or any other multi-scale application, the terrain methodology is used to find sets of stationary and singular points and to determine *average* gradient and *average* curvature (or Hessian matrix) information along a given terrain path. Average gradient and curvature information is calculated from the mean value theorem using the following equations.

$$\langle g \rangle = (1/\alpha) \int g[z(\alpha)] d\alpha \tag{11}$$

$$\langle h \rangle = (1/\alpha) \int h[z(\alpha)] d\alpha \tag{12}$$

where α is some relevant length of the smooth terrain path connecting any set of stationary and singular points. It is important for the reader to understand that it is the set of stationary and singular points as well as average gradient and Hessian matrix information that are communicated from the small length scale to the large length scale.

3.2 A funneling method for optimization at the large length scale

Funneling methods are used to build approximations to the large-scale geometry of f using the funnel function given by

$$F(z) = F_0 - \Gamma e^{-q(z)} \tag{13}$$

where $q(z) = 1/2 z^T Az + b^T z + \mathrm{c}$, and where $\Gamma > 0$, F_0 and c are scalar parameters, b is an n-dimensional vector, and A is an $n \times n$ symmetric matrix. We note that the functional form of Eq. 13 is non-convex, has a unique global minimum when A is positive definite, and contains certain inherent self-scaling characteristics.

To build iterative global funnel approximations of the objective function we match function, gradient, and second derivative information of the true objective function, f, g and h, with the function, gradient, and Hessian matrix information, F, G and H respectively, of the funnel function at various points, where $G = G(z)$ and $H = H(z)$ are given by

$$G(z) = \Gamma e^{-q(z)} [Az + b] \tag{14}$$

$$H(z) = \Gamma e^{-q(z)} [A - (Az + b)(Az + b)^T] \tag{15}$$

Note that if $f(z)$ is used in place of $F(z)$ in Eq. 13, then

$$\gamma(z) = F_0 - f(z) = \Gamma e^{-q(z)} \tag{16}$$

where $\gamma > 0$ is a positive scaling factor that depends on a single numerical measurement, $f(z)$, and the scalar parameter, F_0. Moreover, replacing $G(z)$ with $g(z)$ and $H(z)$ with $h(z)$ in Eqs. 14 and 15 respectively give the equations

$$[Az + b] = \frac{g(z)}{\gamma} \tag{17}$$

$$A = \frac{\gamma h + gg^T}{\gamma^2} \tag{18}$$

Equations 16, 17, and 18 provide a means of estimating A and b from values of $f(z)$, $g(z)$, and $h(z)$ using interpolation formula at two or more iterates.

3.2.1 Interpolating formulae

Let z_k be any value of the unknown optimization variables with corresponding objective function, gradient and Hessian matrix values f_k, g_k and h_k respectively. Also let z_{k+1} be some other arbitrary but not necessarily nearby or successive iterate with corresponding function, gradient and Hessian matrix values f_{k+1}, g_{k+1} and h_{k+1}. Writing Eq. 16 for z_k and z_{k+1} and then subtracting the latter from the former, eliminates F_0 and gives

$$\gamma_{k+1} - \gamma_k = f_k - f_{k+1} \tag{19}$$

Repeating the same algebra using Eq. 18 yields

$$\gamma_k^2[\gamma_{k+1}h_{k+1} + g_{k+1}g_{k+1}^T] - \gamma_{k+1}^2[\gamma_k h_k + g_k g_k^T] = 0 \tag{20}$$

Equations 19 and 20 form a set of $[1+n(n+1)/2]$ nonlinear equations in the two unknowns, γ_k and γ_{k+1}, when the symmetry of the associated matrices is taken into account. This together with Eq. 11 gives a total of $[1 + n(n + 1)/2]$ nonlinear equations. For $n = 1$, there are two equations and two unknowns. When $n > 1$, there are more equations than unknown variables. However, irrespective of this, two equations for which γ_k and $\gamma_{k+1} > 0$ can be determined using the Routh criterion.

3.2.2 Estimating funnel parameters

Calculated values of γ_k and γ_{k+1} can be used to determine the matrix A from Eq. 18—using gradient and Hessian matrix information either at z_k or z_{k+1}. Following this, the parameter b can be computed by simply rearranging Eq. 17 to give

$$b = \frac{g(z)}{\gamma} - Az \tag{21}$$

while F_0 can be calculated from $F_0 = f(z) + \gamma$. Like A, the values of b and F_0 can be determined using function and gradient values at either z_k or z_{k+1}.

3.2.3 Finding the funnel minimum

It is straightforward to estimate the unique global minimizer of the *funnel approximation*, say y, by simply solving

$$Ay = -b \tag{22}$$

Note that Eq. 18 shows that the matrix A is generated from a rank-one, positive semi-definite correction to $h(z)$, that the sign definiteness of A can be controlled by the parameter γ, and that γ plays the role of a self-scaling factor.

3.2.4 Communication between length scales

One of the keys to success in any multi-scale global optimization methodology is the communication between length scales. In the terrain/funneling approach, small-scale calculations communicate *average* gradient and *average* Hessian matrix information at two distinct points to the large length scale. The large length scale optimizations, on the other hand, communicate an estimate of the values of the optimization variables at a converged minimum of the funnel approximation on the true objective function surface to the small length scale and identify the next region on the objective function surface on which small-scale optimizations should be conducted.

4 A multi-scale global optimization method

The details of a multi-scale global optimization methodology based on the terrain and funneling methods are as follows.

1. Locate two distinct stationary points on the objective function surface.
2. Perform two sets of small-scale optimization calculations using the terrain methodology starting from these two stationary points on the objective function surface. Calculate *average* gradient and *average* Hessian information along the resulting terrain paths. Thus at the kth funnel iteration the following information is available—z_k, f_k, g_k and h_k and z_{k+1}, f_{k+1}, g_{k+1} and h_{k+1} such that $f_{k+1} < f_k$.
3. Conduct iterative large-scale optimization calculations with the funneling methodology initialized using z_k, f_k, g_k and h_k and z_{k+1}, f_{k+1}, g_{k+1} and h_{k+1} such that $f_{k+1} < f_k$ to find a funnel minimum that also corresponds to a stationary point on the true objective function surface. To do this,
   (a) Solve Eqs. 19 and 20 for γ_k and γ_{k+1}.
   (b) Using γ_{k+1}, calculate A and b from Eqs. 18 and 21 respectively.
   (c) Determine an estimate of funnel minimum, y, from Eq. 22.
   (d) Evaluate $f(y)$, $g(y)$ and $h(y)$.
   (e) Test $f(y)$ against f_{k+1}. If $f(y) < f_{k+1}$, then go to step 3f for the next funnel iteration. Else set $\gamma_{k+1} = \gamma_{k+1}/2$ and return to step 3b.
   (f) Set $z_{k+1} = y$, $f_{k+1} = f(y)$, $g_{k+1} = g(y)$, and $h_{k+1} = h(y)$.
   (g) If $||g(y)|| < \varepsilon$, set $y^* = y$, and go to step 4; else go to step 3a.
4. Conduct a new set of small-scale terrain calculations using the funnel minimum from step 3. Calculate *average* gradient and *average* Hessian information along the resulting terrain path such that new values of z_{k+1}, f_{k+1}, g_{k+1} and h_{k+1} satisfy the condition $f_{k+1} < f_k$.
5. Repeat step 3 using the new small-scale information and z_k, f_k, g_k and h_k from step 2.
6. Repeat steps 3 and 4 until there is no further decrease in the objective function.

We actually use the funneling method to locate the initial two stationary points on the objective function surface in step 1 of the algorithm. Step 2 of the algorithm uses the terrain methodology to calculate average gradient and average Hessian matrix information in the neighborhood of these initial stationary points. The most effective way to determine γ_{k+1} in step 3a is to rearrange Eq. 19 for γ_{k+1} in terms of γ_k and then substitute the resulting expression into Eq. 20. This gives a cubic polynomial equation in γ_k and shows that there are *three* possible values of γ_k and thus three possible sets of scaling factors (γ_k, γ_{k+1}). Using an equation solver like Newton's method, it is easy to find one solution for γ_k. The other two values of γ_k can be determined by deflation of the cubic equation to a quadratic equation and by using the quadratic formula. The correct value of γ_k is the smallest *real* valued $\gamma_k > 0$ such that $\gamma_{k+1} > \gamma_k$, where $\gamma_{k+1} = f_k - f_{k+1} + \gamma_k$. Step 3b is straightforward and step 3c requires the solution of a system of linear equations. Step 3d evaluates the *actual* function, gradient, and Hessian matrix at the funnel iterate y. Step 3e is used to ensure *monotonically decreasing* objective function values by halving γ_{k+1} until $f(y) < f_{k+1}$ while step 3f replaces the information associated with z_{k+1} with that for the funnel minimum y^*. Finally, step 3g checks the norm of the gradient of the objective function and terminates the funnel iterations once that norm of the gradient falls below the specified tolerance. Note that any point, y^*, that satisfies the convergence condition in step 3g is simultaneously a stationary point of $f(z)$ and a minimizer of the funnel function $F(z)$. It is important for the reader to recognize that the funneling method uses average derivative information on iteration 1 and a mixture of average and point-wise derivative information on all subsequent funnel iterations. Moreover, the purpose of using average gradient and Hessian matrix information is to prevent the funneling method from getting trapped at a local minimum.

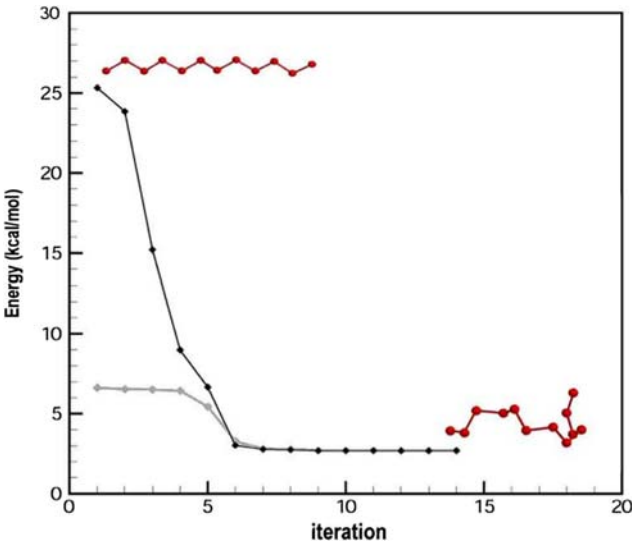

Fig. 1 Terrain/funneling calculations for n-$C_{12}H_{26}$

5 Computational results

In this section we present numerical results for the molecular conformation of pure n-$C_{12}H_{26}$, n-$C_{16}H_{34}$, and n-$C_{24}H_{50}$, as model fluids for diesel, home heating and residual fuel oils. These n-alkanes are modeled using a united atom approach and the multi-scale optimization method described in the previous section is used to solve the associated global optimization problem. The number of unknown variables in these optimization problems ranges from 30 to 66 Cartesian coordinates. All computations were performed in double precision arithmetic on a Dell 670 High Precision Workstation using a Lahey-Fujitsu LF95 compiler. The tolerance used in step 3g of the algorithm was $\varepsilon = 10^{-8}$.

5.1 n-$C_{12}H_{26}$ as a model for diesel fuel

Diesel fuel, like any fuel oil, is actually comprised of a mixture of hydrocarbons of varying carbon chain length — in this case C_{10} to C_{15}. Thus we choose pure n-$C_{12}H_{26}$ as a representative model of diesel fuel. As a result, this first example problem is the smallest of the three examples studied with 30 unknown variables. Figure 1 summarizes the results for n-$C_{12}H_{26}$ and clearly shows that the terrain/funneling approach converges to the *exact* same global minimum molecular conformation from all starting points. The Cartesian coordinates for the global minimum and corresponding potential energy are given in Table 1 at the end of this section.

Our multi-scale optimization method starts by finding two distinct stationary points using the funneling algorithm, as described in step 1 of the algorithm. To do this for the case of n-$C_{12}H_{26}$ we chose two arbitrary starting points for which the potential energy was $E_1 = 116545.30$ kcal/mol and $E_2 = 85761.12$ kcal/mol respectively. Using the gradient and Hessian matrix information at both starting points, funnel parameters and a first estimate of funnel minimum were determined. We then replaced one of the starting points with this estimate of the funnel minimum and repeated the funnel optimization calculations until a stationary point

Table 1 Cartesian coordinates of the global minimum for n-$C_{12}H_{26}$, n-$C_{16}H_{34}$, and n-$C_{24}H_{50}$

Cartesian coordinates	Value at global minimum n-$C_{12}H_{26}$[a]	n-$C_{16}H_{34}$[b]	n-$C_{24}H_{50}$[c]
(x_1, y_1, z_1)	(0, 0, 0)	(0, 0, 0)	(0, 0, 0)
(x_2, y_2, z_2)	(1.52978, 0, 0)	(1.5299, 0, 0)	(1.52975, 0, 0)
(x_3, y_3, z_3)	(2.03783, 1.44331, 0)	(2.04964, −1.44003, 0)	(2.04909, −1.43991, 0)
(x_4, y_4, z_4)	(2.76083, 1.73036, −1.31725)	(2.87575, −1.69514, −1.26367)	(2.85560, −1.69499, −1.27531)
(x_5, y_5, z_5)	(4.23864, 2.01679, −1.04099)	(2.26127, −2.82871, −2.09091)	(2.25359, −2.85170, −2.07773)
(x_6, y_6, z_6)	(5.10547, 0.924409, −1.67263)	(1.83171, −2.29902, −3.46121)	(1.75075, −2.34254, −3.43061)
(x_7, y_7, z_7)	(5.99544, 1.52757, −2.76285)	(0.341621, −2.56374, −3.69012)	(0.276676, −2.70976, −3.61733)
(x_8, y_8, z_8)	(5.65733, 0.903789, −4.12038)	(−0.440559, −1.24761, −3.70119)	(−0.602732, −1.45618, −3.63107)
(x_9, y_9, z_9)	(5.17613, 1.97948, −5.10233)	(−0.923272, −0.940518, −5.12058)	(−1.12458, −1.21561, −5.04969)
(x_{10}, y_{10}, z_{10})	(3.73302, 1.69078, −5.52741)	(−0.438661, 0.448099, −5.55550)	(−0.826562, 0.221515, −5.49382)
(x_{11}, y_{11}, z_{11})	(2.82582, 2.89630, −5.25663)	(0.590982, 0.321678, −6.68299)	(0.307108, 0.233377, −6.52362)
(x_{12}, y_{12}, z_{12})	(3.64717, 4.09855, −4.78428)	(1.88493, 1.08332, −6.36230)	(1.41146, 1.22571, −6.12732)
(x_{13}, y_{13}, z_{13})		(1.78474, 1.84017, −5.03411)	(0.987062, 2.09389, −4.93600)
(x_{14}, y_{14}, z_{14})		(2.90461, 1.38633, −4.09287)	(1.98209, 1.99756, −3.77104)
(x_{15}, y_{15}, z_{15})		(3.83659, 2.56256, −3.80035)	(3.21152, 1.17358, −4.15806)
(x_{16}, y_{16}, z_{16})		(3.74155, 2.92394, −2.31691)	(4.47846, 2.0123, −3.97420)
(x_{17}, y_{17}, z_{17})			(5.27644, 1.44990, −2.79686)
(x_{18}, y_{18}, z_{18})			(6.58382, 0.835735, −3.30166)
(x_{19}, y_{19}, z_{19})			(6.56878, −0.675422, −3.05523)
(x_{20}, y_{20}, z_{20})			(7.64551, −1.04743, −2.03243)
(x_{21}, y_{21}, z_{21})			(6.99291, −1.63299, −0.776988)
(x_{22}, y_{22}, z_{22})			(7.29891, −0.743315, 0.430644)
(x_{23}, y_{23}, z_{23})			(6.00224, −0.127821, 0.96097)
(x_{24}, y_{24}, z_{24})			(5.70535, −0.685178, 2.35413)

[a] global min $E^* = 2.6881$ kcal/mol for n-$C_{12}H_{26}$
[b] global min $E^* = 3.233$ kcal/mol for n-$C_{16}H_{34}$
[c] global min $E^* = 3.28692$ kcal/mol for n-$C_{24}H_{50}$

on the energy surface was located. The first stationary point located was a local minimum, has a potential energy of E_1^* = 5.63197 kcal/mol, and was computed in 45 iterations and 1.6 s of computer time. In a similar way, using two other starting points with energy values of $E_3 = 2.1877 \times 10^6$ kcal/mol and E_4= 1824.31 kcal/mol, a second stationary point, a saddle point, with an energy value of E_2^* = 5.16511 kcal/mol was calculated in 17 funnel iterations and 0.43 s of computer time. Note that $E_2^* < E_1^*$.

Average gradient and average Hessian matrix information was then calculated in the neighborhood of each of these initial stationary points using the terrain methodology. This is step 2 of the multi-scale optimization algorithm. From the first stationary point, which corresponded to an energy of $E_1^* = 5.63197$ kcal/mol, average gradient and Hessian information was accumulated along the terrain path using the path integrals given by Eqs. 11 and 12. For the first set of small-scale optimization calculations, the terrain methodology used 8235 function and gradient evaluations and 871.47 s of computer time. In a similar way, average gradient and Hessian matrix information was accumulated in the neighborhood of the second stationary point, E_2^* = 5.16511 kcal/mol. Here, the terrain methodology used 18293 function and gradient evaluations and 2048 s of computer time.

To initialize iterative large-scale funnel calculations we picked two points — one from each of the two sets of terrain calculations. Using the corresponding function, average gradient and average Hessian information at these two points, we initiated funnel calculations and located a third stationary point, a minimum with an energy value of $E_3^* = 2.68811$ kcal/mol, in 14 funnel iterations and 0.23 s of computer time. The gray curve in Fig. 1 gives the funnel iteration history for this set of multi-scale global optimization calculations. Note that the potential energy decreases monotonically with each funnel iteration and finally finds the global minimum on the energy surface rather easily.

Note that $E_3^* < E_2^* < E_1^*$ as required by the methodology. Given the minimum, E_3^*, the multi-scale global optimization method performed another set of small-scale optimization calculations. That is, a new set of small-scale terrain calculations was conducted in the neighborhood of E_3^* and average gradient and Hessian information in the neighborhood of E_3^* was accumulated. This information was then used in place of E_2^* and its associated average gradient and Hessian information, as described in step 5, and a second set of funneling calculations was conducted. However, the second set of funnel calculations found the same minimum on the energy surface at E_4^* = 2.68811 kcal/mol. Since $E_4^* = E_3^*$, the multi-scale optimization algorithm terminates with E_4^* as the global minimum. Table 1 gives the Cartesian coordinates for this global minimum. From these Cartesian coordinates, it is easily determined that the conformation of the united atom model of n-$C_{12}H_{26}$ is zigzag with wrapped ends, at which the bond angles are all at their equilibrium value of θ = 1.91 rad, the bond lengths are almost all at their equilibrium values and it is the torsion and van der Waals energies that dictate conformation. See also Fig. 1. Note that without the use of average gradient and Hessian matrix information, the funnel method would be trapped at the local minimum E_1^* or the saddle point E_2^*.

To illustrate the robustness of the multi-scale methodology, small-scale terrain calculations in the neighborhood of two other stationary points on the potential energy surface – E_5^* = 43.0166 kcal/mol, which correspond to a saddle point on the energy surface with a planar (or two-dimensional) zigzag structure and a second saddle point with an energy value of E_6^* = 7.49362 kcal/mol—were performed and average gradient and average Hessian information was gathered along these terrain paths. Using this information, the funneling algorithm again calculated the same funnel minimum corresponding to a global minimum on the energy surface at E_3^* = 2.68811 kcal/mol in 14 funnel iterations and 0.24 s. The results of these calculations are shown in Fig. 1 in black. It is important to note that despite the fact that

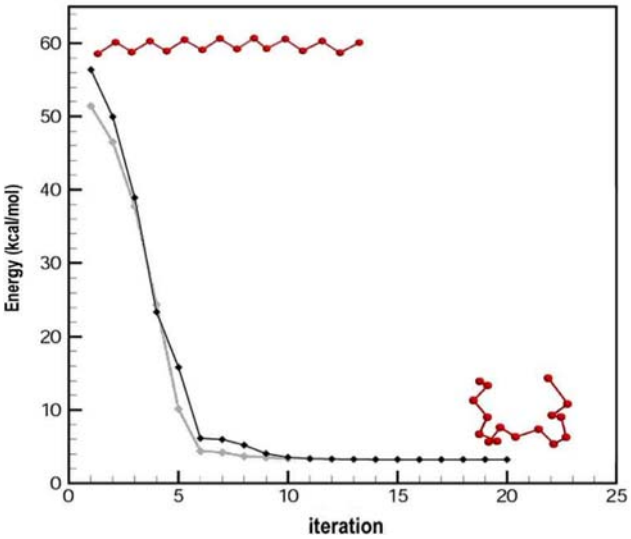

Fig. 2 Terrain/funneling calculations for n-$C_{16}H_{34}$

we started the terrain/funneling calculations from a planar structure or saddle point with energy E_5^* = 43.0166 kcal/mol, which is very high when compared to the energy at the global minimum, our multi-scale optimization approach easily found the global minimum in a reliable and efficient manner. Finally, we mention that we have performed a large number of similar multi-scale global optimization calculations using other starting points on the potential energy surface and in all cases, our methodology finds the same low energy minimum of E_3^* = 2.68811 kcal/mol and the same exact molecular conformation.

5.2 n-$C_{16}H_{34}$ as a model for home heating fuel

The model fluid for home heating fuel, which is generally a mixture of C_{10} to C_{20}, is pure hexadecane or n-$C_{16}H_{34}$. The corresponding number of unknown optimization variables for a united atom model of hexadecane is 42.

Figure 2 shows funneling results for two separate sets of multi-scale global optimization calculations for determining the molecular conformation of hexadecane. Note that both sets of funnel iterations shown in Fig. 2 produce monotonically decreasing potential energy values. Clearly both sets of terrain/funneling calculations, each initiated from different starting points on the potential energy surface, easily find the *same* global minimum at $E = 3.233$ kcal/mol with the *same* molecular conformation in 20 and 19 funnel iterations respectively.

To give the reader an understanding of how these calculations proceed, consider the funnel iterations shown in black in Fig. 2. Small-scale terrain calculations starting from points on the energy surface corresponding to $E_1^* = 62.4577$ kcal/mol, which is a saddle point on the energy surface having a flat planar zigzag structure and a local minimum with a potential energy of $E_2^* = 3.63439$ kcal/mol were performed and average gradient and average Hessian information along the terrain paths was gathered. Note that $E_2^* < E_1^*$. The first set of small-scale terrain calculations required 19415 function and gradient evaluations and 3908 s of computer time while the second set of small-scale terrain calculations used 19220 function and gradient evaluations and 3869 s of computer time. Funnel calculations were then

initiated using the information from the small-scale terrain calculations whereby the funneling algorithm located a stationary point corresponding to a global minimum on the energy surface at $E_3^* = 3.233$ kcal/mol in 20 funnel iterations and 0.25 s of computer time. The multi-scale algorithm then performed small-scale terrain calculations in the neighborhood of E_3^* in order to gather average gradient and curvature information. Using E_3^* and its average gradient and Hessian matrix information (in place of E_2^* and its average information) together with E_1^* and its average information, the funnel method computed the same minimum value of the potential energy, $E_4^* = 3.233$ kcal/mol. Since $E_3^* = E_4^*$ the multi-scale algorithm terminates with $E_3^* = E_4^* = 3.233$ kcal/mol as the global minimum. Numerical results for the gray curve shown in Fig. 2 are very similar to those just described—as are numerous other terrain/funneling calculations that have been conducted for hexadecane.

The Cartesian coordinates for the global minimum are given in Table 1. Here again, as in the case of n-$C_{12}H_{26}$, the Cartesian coordinates given in Table 1, as well as the graphical representation in Fig. 2, clearly show that the global minimum molecular conformation of hexadecane is zigzag with considerable wrapping at the ends of the molecule. In fact, there is more wrapping of the ends of the united atom molecule for hexadecane than for n-$C_{12}H_{26}$—most likely due to the increase in carbon chain length. Also, all united atoms are at their equilibrium bond angles of 1.91 rad, the bond lengths are very close to their equilibrium values and torsion and van der Waals energies play a dominant role in determining molecular conformation.

5.3 n-$C_{24}H_{50}$ as a model for residual fuel oils

Residual fuel oils vary significantly in carbon chain length from C_{20} to C_{70}. Here we choose pure n-$C_{24}H_{50}$ as a representative residual fuel oil and thus the number of unknown variables is 66. Figure 3 shows results for two separate sets of multi-scale global optimization calculations using terrain/funneling methods for pure n-$C_{24}H_{50}$. From this figure it is clearly seen that both sets of multi-scale global optimization calculations easily find the *exact same* molecular conformation with a global minimum at $E = 3.28692$ kcal/mol in 23 and 25 funnel iterations respectively.

Again we describe the numerical performance of the multi-scale global optimization algorithm for the funnel iterations shown in black in Fig. 3. First, small-scale terrain calculations from two different starting points on the potential energy surface were performed. The first starting point was a saddle point on the energy surface with a planar zigzag structure and a corresponding potential energy of $E_1^* = 101.34$ kcal/mol. The second starting point was also a saddle point and had a potential energy of $E_2^* = 11.44861$ kcal/mol. These small-scale terrain calculations were used to gather average gradient and average Hessian information along the terrain paths in the neighborhoods of these stationary points. The first set of small-scale optimization calculations required 1860 function and gradient evaluations and 945 s of computer time while the second set of small-scale terrain calculations needed 6488 function and gradient evaluations and 4349 s of computer time. Information from these small-scale optimization calculations was then used to initialize iterative large-scale funnel calculations and the funneling algorithm located a stationary point corresponding to a global minimum on the energy surface at $E_3^* = 3.28692$ kcal/mol in 24 funnel iterations and 0.5 s. Repeated small-scale calculations around E_3^* and a subsequent set of funnel iterations gave the same minimum on termination. That is, the algorithm terminates with $E_4^* = E_3^* = 3.28692$ kcal/mol as the global minimum of the potential energy for pure n-$C_{24}H_{50}$.

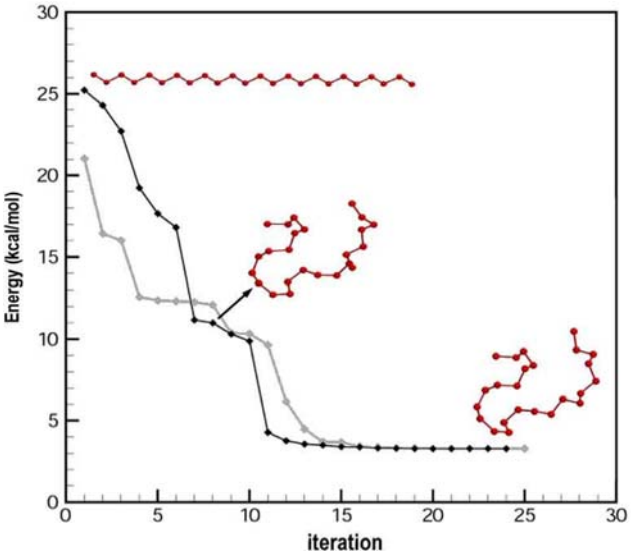

Fig. 3 Terrain/funneling calculations for n-$C_{24}H_{50}$

Similar numerical behavior was observed for the funneling calculations shown in gray in Fig. 3 and for other numerical experiments that have been performed for n-$C_{24}H_{50}$. The Cartesian coordinates for the global minimum are given in Table 1.

From the Cartesian coordinates and Fig. 3 we can again observe that there is considerable wrapping of the minimum molecular conformation for n-$C_{24}H_{50}$. In fact, it is easily seen that the degree of wrapping at the ends of the united atom model of n-$C_{24}H_{50}$ is greater than that for hexadecane, which in turn is greater than the wrapping for n-$C_{12}H_{26}$. Also all united atoms are at their equilibrium bond angles of 1.91 radians in the global minimum molecular conformation.

6 Conclusions

A multi-scale global optimization methodology that uses small-scale terrain and large-scale funneling calculations was presented. This multi-scale optimization algorithm was used to find molecular conformations for three representative fuel oils—diesel fuel, home heating fuel, and residual fuel oil—modeled as pure n-alkanes in a united atom framework. Numerical results clearly show that our multi-scale methodology is capable of reliably finding the molecular conformation corresponding to the global minimum on the potential energy surface with modest computer resources.

We close this paper with a brief indication of the direction of this research. As noted in the introduction, we are interested in the waxing of fuel oils. For this, a collection of n-alkane molecules and models of crystal lattice structures are required. Thus the next steps in this research will consist of the development of a model of the potential energy of crystal lattices, the determination of low energy crystal structures (or conformations), and a quantitative description of the interplay between molecular conformation and crystal structure. We believe that the molecular conformation of individual pure n-alkanes or n-alkane mixtures with or

without additives must change as they pack into low energy crystal structures due to the fact that there are additional *intermolecular* van der Waals forces.

7 Coda

Those interested in detailed copies of the computations are encouraged to contact the authors.

Acknowledgement The authors would like to thank the Petroleum Research Fund for support of this work under Grant No. ACS PRF # 44793-AC9.

References

1. Turner, W.R.: Normal alkanes. Ind. Eng. Chem. Prod. Res. **10**, 238–260 (1971)
2. Larini, L., Barbieri, A., Prevosto, D., Rolla, P., Leporini, D.: Equilibrated polyethylene single-molecule crystals: molecular dynamics simulations and analytic model of the global minimum of the free energy landscape. J. Phys.: Condens. Matter. **17**, L199–L208 (2005)
3. Levy, A.V., Montalvo, A.: The tunneling algorithm for the global minimization of functions. SIAM J. Sci. Stat. Comp. **6**, 15–29 (1985)
4. Bahren, J., Protopopescu, V.: Generalized TRUST algorithms for global optimization. In: Floudas, C.A., Pardalos, P.M. (eds.) State of the Art in Global Optimization., pp. 163–180. Kluwer, Dordrecht (1996)
5. Maranas, C.D., Floudas, C.A.: Finding all solutions to nonlinearly constrained systems of equations. J. Global Optim. **7**, 143–182 (1995)
6. Maranas, C.D., Floudas, C.A.: A global optimization approach to Lennard-Jones microclusters. J. Chem. Phys. **97**, 7667–7678 (1992)
7. Androulakis, I.P., Maranas, C.D., Floudas, C.A.: Prediction of oligopeptide conformations via deterministic global optimization. J. Global Optim. **11**, 1–34 (1997)
8. Maranas, C.D., Floudas, C.A.: Global optimization in generalized geometric programming. Comput. Chem. Eng. **21**, 351–369 (1997)
9. Westerberg, K.M., Floudas, C.A.: Locating all transition states and studying the reaction pathways of potential energy surfaces. J. Chem. Phys. **110**, 9259–9295 (1999)
10. Klepeis, J.L., Floudas, C.A., Morikis, D.: Predicting peptide structures using NMR data and deterministic global optimization. J. Comput. Chem. **20**, 1354–1370 (1999)
11. Sun A.C., Seider, W.D.: Homotopy-continuation algorithm for global optimization. In Floudas, C.A., Pardalos, P.M. (eds.) Recent Advances in Global Optimization, pp. 561–592. Princeton Univ. Press (1992)
12. Piela, L., Kostrowicki, J., Scheraga, H.A.: The multiple minima problem in conformational analysis of molecules. Deformation of the potential energy hypersurface by the diffusion equation method. J. Phys. Chem. **93**, 3339–3346 (1989)
13. Hansen, E.R.: Global optimization using interval analysis—the multidimensional case. Numer. Math. **34**, 247–270 (1980)
14. Lin, Y., Stadtherr, M.A.: Deterministic global optimization of molecular structures using interval analysis. J. Comput. Chem. **26**, 1413–1420 (2005)
15. Lucia, A., Yang, F.: Global terrain methods. Comput. Chem. Eng. **26**, 529–546 (2002)
16. Lucia, A., Yang, F.: Multivariable terrain methods. AIChE J. **49**, 2553–2563 (2003)
17. Lucia, A., DiMaggio, P.A., Depa, P.: A geometric terrain methodology for global optimization. J. Global Optim. **29**, 297–314 (2004)
18. Lucia, A., DiMaggio, P.A., Depa, P.: Funneling algorithms for multi-scale optimization on rugged terrains. Ind. Eng. Chem. Res. **43**, 3770–3781 (2004)
19. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)
20. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 671–680 (1983)
21. DeJong K.: Ph.D. Dissertation, University of Michigan (1976)
22. Holland, J.H.: Genetic algorithms. Sci. Am. **267**, 66 (1992)

23. Aluffi-Pentini, F., Parisi, V., Zirilli, F.: Global optimization and stochastic differential equations. J. Opt. Theory and Appl. **47**, 1–16 (1985)
24. Bilbro, G.L.: Fast stochastic global optimization. IEEE Trans. Sys. Man. Cyber **4**, 684–689 (1994)
25. Sevick, E.M., Bell, A.T., Theodorou, D.N.: A chain of states method for investigating infrequent events in processes occurring in multistate, multidimensional systems. J. Chem. Phys. **98**, 3196–3212 (1993)
26. Scheraga, H.A.: Prediction of protein conformation. In: Anfinsen, C.B., Schechter, A.N. (eds.) Current Topics in Biochemistry, pp. 1. Acad. Press, New York (1974)
27. Pincus, M.R., Klausner, R.D., Scheraga, H.A.: Calculation of the three dimensional structure of the membrane-bound portion of melittin from its amino acid sequence. Proc. Nat. Acad. Sci. **79**, 5107–5110 (1982)
28. Gibson, K.D., Scheraga, H.A.: Revised algorithm for the build-up procedure for predicting protein conformation by energy minimization. J. Comput. Chem. **8**, 826–834 (1987)
29. Jones, D.T., Taylor, W.R., Thornton, J.M.: A new approach to protein folding recognition. Nature **358**, 86–89 (1992)
30. Muller, K., Brown, L.D.: Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. Theor. Chim. Acta. **53**, 75–93 (1979)
31. Cerjan, C.J., Miller, W.H.: On finding transition states. J. Chem. Phys. **75**, 2800–2806 (1981)
32. Baker, J.: An algorithm for the location of transition states. J. Comput. Chem. **7**, 385–395 (1986)
33. Henkelman, G., Johannesson, G., Jonsson, H. : Methods for finding saddle points and minimum energy paths. In: Schwartz, S.D. Progress in Theoretical Chemistry and Physics, pp. 269–300. Kluwer, Dordrecht Netherlands (2000)
34. Deaven, D.M., Tit, N., Morris, J.R., Ho, K.M.: Structural optimization of Lennard-Jones clusters by a genetic algorithm. Chem. Phys. Lett. **256**, 195–200 (1996)
35. Niesse, J.A., Mayne, H.R.: Global geometry optimization of atomic clusters using a modified genetic algorithm in space-fixed coordinates. J. Chem. Phys. **105**, 4700–4706 (1996)
36. Wales, D.J., Doye, J.P.K.: Global optimization by basin hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. A. **101**, 5111–5116 (1997)
37. Doye, J.P.K., Wales, D.J.: Saddle points and dynamics of Lennard-Jones clusters, solids and supercooled liquids. J. Chem. Phys. **116**, 3777–3788 (2002)
38. Matro, A., Freeman, D.L., Doll, J.D.: Locating transition states using double-ended classical trajectories. J. Chem. Phys. **101**, 10458–10463 (1994)
39. Gregurick, S.K., Alexander, M.H., Hartke, B.: Global geometry optimization of $(Ar)_n$ and $B(Ar)_n$ clusters using a modified genetic algorithm. J. Chem. Phys. **104**, 2684–2691 (1996)