Conformational Sampling of Macrocyclic Alkenes Using a Kennard−Stone-Based Algorithm

Diederica D. Claeys,[†,‡] Toon Verstraelen,[†] Ewald Pauwels,[†] Christian V. Stevens,[‡] Michel Waroquier,[†] and Veronique Van Speybroeck*[,†]

Center for Molecular Modeling, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium, QCMM - alliance Ghent−Brussels, Belgium, and Research Group SynBioC, Department of Organic Chemistry, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

Received: March 12, 2010; Revised Manuscript Received: May 21, 2010

The properties and functions of (bio)molecules are closely related to their molecular conformations. A variety of methods are available to sample the conformational space at a relatively low level of theory. If a higher level of theory is required, the computational cost can be reduced by selecting a uniformly distributed set of conformations from the ensemble of conformations generated at a low level of theory and by optimizing this selected set at a higher level. The generation of conformers is performed using molecular dynamics runs which are analyzed using the MD-Tracks code [*J. Chem. Inf. Model.* **2008**, *48*, 2414]. This article presents a Kennard−Stone-based algorithm, with a distance measure based on the distance matrix, for the selection of the most diverse set of conformations. The method has been successfully applied to macrocyclic alkenes. The correct thermodynamic stability of the double-bond isomers of a flexible macrocyclic alkene containing two chiral centers is reproduced. The double-bond configuration has a limited effect on the conformation of the whole macrocycle. The chirality of the stereocenters has a larger effect on the molecular conformations.

Introduction

Macromolecules can adopt many different conformations. As there is a tight relationship between the structure of (bio)molecules and their properties and function, there is a general interest in the ensemble of energetically accessible conformations.[1] The first step is a conformational sampling that involves the generation and selection of different molecular geometries. A variety of methods is available to scan and sample the configuration space.[2]

A static approach, in which a few conformations are constructed manually based on intuition, is only applicable for the simplest molecules. The most rigorous method for locating the global minimum is to perform a full conformational search, which entails the calculation of all potentials for simultaneous rotations about every bond in the molecule. This results in a large multidimensional potential energy surface from which the lowest-energy conformer can be identified.[3−6] Systematic search protocols, like a regular grid search, considering all conformational degrees of freedom require an exponentially increasing demand of computer time and can only be applied to small molecules with only few degrees of freedom.[7] Recently, interesting new algorithms were presented to search in an efficient way for the lowest-energy conformer of a variety of molecules.[8]

An uncorrelated series of random configurations can be generated using distance geometry methods, based on distances between atoms together with their chirality. By imposing additional constraints, like experimentally derived NMR data, the search can be directed to interesting regions of the conformational space.[9,10] The obtained ensemble will not be Boltzmann-distributed.

A large variety of search procedures is based on stepping through the configurational space using energy, gradient, Hessian, randomness, and memory to determine the next step direction and size.[2] The most prominent methods are Monte Carlo (MC) and molecular dynamics (MD). In MC, the step direction is taken at random, and the correct distribution is obtained by the Boltzmann acceptance criterion, which is only feasible with moderate step sizes. In MD simulations, the step is determined by computing the forces on the atoms and by the inertia of the degrees of freedom solving Newton's or Lagrange's equations of motion. The time step is limited by the highest frequency in the molecular system.[11] Although molecular dynamics techniques are very useful, care needs to be taken that all relevant parts of the PES are sampled. Therefore, a variety of advanced techniques have been introduced, of which some that are used frequently are shortly explained below.

Eliminating these fast vibrational degrees of freedom allows larger time steps and faster sampling. This can be done by imposing distance constraints, using the SHAKE[12] or improved algorithms.[13] Another possibility is grouping atoms or molecular fragments. This coarse-graining technique requires the setup of a coarse-grained potential, based on MD or MC simulations of small representative systems.

If the barrier separating two conformations is large compared to k_BT, barrier crossing is infrequent, and the simulation gets trapped in low-energy regions. Modifications to the potential energy surface (PES) can be introduced to accelerate barrier crossing.

In umbrella sampling,[14] a set of restraints is imposed on selected (collective) variables, and the region of low sampling is bridged by a biasing potential that cancels the energy barrier. The choice of the coordinates is important, and the construction of the function that modifies the PES is a real challenge, and one should have an idea of the energy profile.

A history-dependent dynamics can be constructed in the space of the selected variables. In metadynamics, Gaussian functions

* To whom correspondence should be addressed. E-mail: Veronique.Vanspeybroeck@UGent.be.
[†] Center for Molecular Modeling, Ghent University. Tel: +32-9-2646558.
[‡] Faculty of Bioscience Engineering, Ghent University.

10.1021/jp1022778 © 2010 American Chemical Society
Published on Web 06/08/2010

are added to the free-energy surface along the trajectory, filling the minima and allowing barrier crossing.[15] The main modes of a principal component analysis (PCA) are often used as collective coordinates.[16] This requires a simulation to perform the PCA and is restricted to linear modes.

The PES can also be temporally modified by potential scaling. The energy barrier of certain interactions is lowered. Smooth rescaling during the simulation allows convergence to new low-energy regions.[17]

The sampling of the PES can also be enhanced by scaling the system parameters. In simulated annealing, the simulation is started at high temperature, allowing for barrier crossing, and is followed by gradual cooling.[18] The performance of the method, however, depends much on the initial temperature and cooling rate, but the tuning of these parameters is difficult. The whole simulation can be performed at high temperature, exploring more different states. However, the entropic effect is important for large systems, for which the free energy becomes very sensitive to the temperature.[1,2] A very elegant method is increasing the mass of specific parts of a molecule. This enhances barrier crossing as less energy is taken by the motions that do not contribute to more global conformational changes, and at the same time, this may allow for longer time steps.[19]

Genetic algorithms can be used to search the lowest-energy structures of macromolecules. Starting from a pool of different configurations, new ones are created and existing ones removed by mutating and combining parts of configurations.[20]

Another way of improving barrier crossing but maintaining a Boltzmann distributed ensemble is the replica-exchange algorithm, in which multiple copies of the system are simulated at distinct temperatures. From time to time, a MC trial move attempts to exchange copies at adjacent temperatures.[21−24]

Given the computational demands of these conformational sampling techniques, they are performed at relatively low levels of theory, and afterward, a higher level of theory (LOT) can be used on a selected training set. There are several possibilities to create the training set. The simplest procedure is to perform a random selection. However, there is a risk that geometries of certain parts of the PES are not selected and that others are very similar. To make the step to a higher LOT and limit the computational cost, one should sample the obtained ensemble in a more uniform way. This will avoid calculation of similar structures at the computationally expensive level.

Therefore, we have developed a new sampling procedure based on the Kennard−Stone algorithm. We evaluated it by investigating the relative thermodynamic stability of the double-bond isomers (*E* or *Z*) of a 24-membered cyclic alkene containing two planar hydantoin rings (Figure 1). The macrocycle **1** has two chiral centers and was obtained as a diastereomeric mixture.[25] The two chiral centers can be equal or different. Therefore, both an *S*,*S* and an *R*,*S* isomer were modeled with an *E* or *Z* double bond, denoted respectively as *E*-(*S*,*S*), *Z*-(*S*,*S*), *E*-(*R*,*S*), and *Z*-(*R*,*S*) (Figure 1).

Compound **1** was synthesized by ring closing metathesis (RCM), and only the *E* isomer was observed experimentally.[25,26] Usually, RCM results in a mixture of *E* and *Z* isomers,[27,28] but in some cases, stereoselective formation of the *E* isomers has been reported.[29] We can assume that a thermodynamic equilibrium was set. The fully saturated second-generation Grubbs's catalyst **3** (Figure 2)[30] that was used for the RCM reaction is known to give higher *E*/*Z* ratios than the use of the standard Grubbs's carbene **2**[31] as **3** is able to isomerize the initial product via backward and forward reactions during the so-called secondary metathesis process, thereby enriching the mixture in

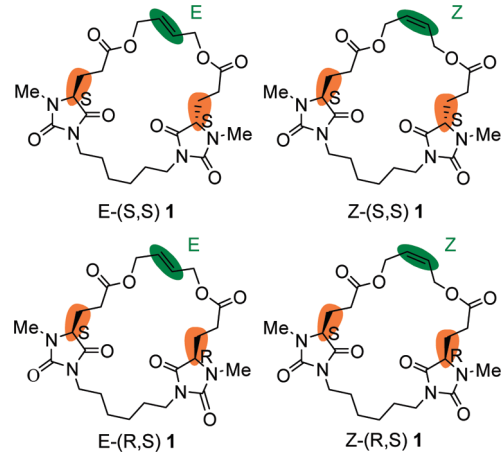


Figure 1. Different stereomers modeled by molecular dynamics.

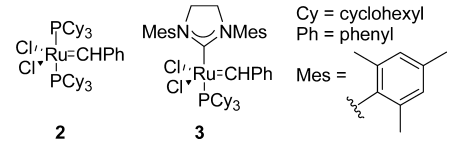


Figure 2. First (2) and second (3) generation Grubbs's catalysts.

the thermodynamically favored isomer.[32,33] As there are no substituents next to the double bond and the reaction was allowed to equilibrate by stirring the reaction overnight under reflux in benzene or dichloromethane, we can assume thermodynamic equilibrium and predict the ratio of the *E* and *Z* isomers based on their relative thermodynamic stability.

This paper first describes the conformational sampling protocol. As the main goal is to illustrate the new Kennard−Stone-based sampling technique, we have chosen the very intuitive high-temperature procedure to scan the potential energy surface at the low LOT. More advanced methods should be used to scan the PES at the low LOT if the free energy is more sensitive to the temperature. Once the ensemble of conformations is generated, they can be sampled by the new Kennard−Stone-based sampling algorithm. The suitability of the methodology is evaluated by predicting the thermodynamic stability of the double-bond isomers of macrocycle **1**. Finally, the effect of the double-bond configuration on the conformation of the macrocycle is investigated by doing a conformational analysis.

Methods

Conformational Sampling Protocol. The conformational sampling protocol outlined below is applied to four isomers of the macrocycle, two *E* and two *Z* isomers. Within each type, two stereocenters can be equal (*SS*) or different (*RS*), as illustrated in Figure 1. The end result of the sampling procedure is a set of representative and diverse conformers for each of the four isomers. These sets can be used to characterize the effect of the double bond (*E* or *Z*) on the thermodynamic stability of the macrocycle.

The sampling procedure (Figure 3) is applied to each stereomer and consists of four steps, (i) a quantum mechanical molecular dynamics simulation at a semiempirical level, (ii) a first assessment of the completeness of the simulation based on the principal component analysis, (iii) a second assessment of the completeness of the simulation based on the sampling of dihedral angles, and (iv) the selection of configurations from the molecular dynamics trajectory with the Kennard−Stone algorithm.

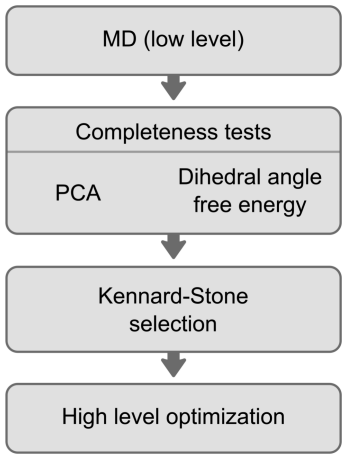


Figure 3. Flowchart of the conformational sampling protocol.

The entire analysis and postprocessing of the molecular dynamics simulation are implemented with the in-house developed trajectory analysis code, MD-Tracks, which is distributed as open source software on our Web site http://molmod.ugent.be/.[34]

An initial geometry of each conformer is constructed with Zeobuilder.[35] The molecular dynamics simulations are carried out with CP2K,[36] using the PM3 semiempirical Hamiltonian[37] to compute the interatomic forces. The conformational space is explored with an NVT molecular dynamics run of 300 ps using an integration time step of 1 fs and the Nosé−Hoover thermostat[38,39] with a temperature of 1500 K and a time constant of 100 fs. The computational efficiency of the PM3 method and the high temperature make the extensive sampling feasible. Prior to the production run, the system is initialized by means of an NVE simulation with a threshold in the instantaneous temperature, that is, as soon as the temperature drops below 1400 K or rises above 1600 K, the velocities are scaled uniformly to obtain an instantaneous temperature of 1500 K. Although the initialization algorithm does not rigorously sample the NVT ensemble, it is an efficient tool to introduce the required amount of thermal energy before the actual production run.

The high temperature is not applicable in general, and one must carefully consider entropic side effects. It is, for example, well-known that the preferred conformation of a protein is very sensitive to temperature mainly because the entropic contribution to the free energy of the folded and the unfolded states is very dissimilar.[1,2] When this is the case, more advanced (and computationally expensive) sampling schemes[1,2,11,40] are required to perform the conformational search at the semiempirical level.

It is essential to check the convergence of the molecular dynamics simulations. If the simulations were too short, the NVT ensemble of each conformer would not be sufficiently sampled, and by consequence, an ensemble average or a selection of conformers could never be representative. There is no golden standard to test convergence in general, but one can at least check that the ensemble averages of interest over the first and the second halves of the simulation are equal within the statistical error margins. Another test is the absence of cosine content in a principal component analysis.[41,42] Both types of tests are discussed below. One must be aware that all of these tests are necessary but not sufficient for the assessment of convergence. Strictly speaking, one can never avoid that a molecular dynamics simulation is trapped in a local free-energy minimum with a mean residence time that is much longer than the actual simulation.

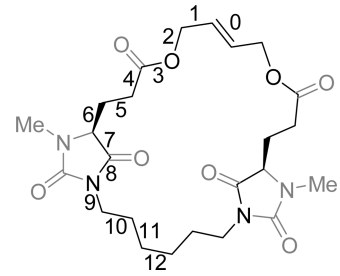


Figure 4. Indication of the backbone of macrocycle **1** in black and the dihedral angles of the *E*-(*R*,*S*) isomer.

The first convergence test is based on histograms of the dihedral angles of the macrocycle backbone, indicated in Figure 4. The backbone dihedral angles are labeled starting from the double bond. For each dihedral angle, a histogram is constructed covering the whole range from −180 to 180° with bin widths of 20°. Also, the error on the count of each bin is computed, taking into account the sampling inefficiency due to time correlations.[43] An example is given in Figure 5a. On the basis of the histogram, a free-energy profile is derived using the following relation

$$A_i = -k_B T \log(N_i) + C$$

in which A_i is the free energy of the bin with index i, N_i is the corresponding count, k_B is the Boltzmann constant, T is the temperature, and C is an arbitrary constant taken such that the lowest A_i becomes zero. First-order estimates of the statistical errors on the free energy are estimated from the errors on the counts in the histogram. An example is given in Figure 5b. The red and the cyan curves give the histogram and free-energy profiles for dihedral angle 2 positioned at one side of the double bond (closest to the *R* center) and the other side of the double bond (closest to the *S* center), respectively. If the molecular dynamics runs are complete, the two profiles should coincide as the dihedral angles are equivalent (Figure 4). The corresponding free-energy profiles are indeed nearly identical (or reflected when chirality plays a role). This procedure was checked for each pair of equivalent dihedral angles in the four conformers.

The second convergence test is based on principal component analysis.[41,42] Each geometry from the trajectory is rotated and translated with the Kabsch algorithm[44] to have a minimal root-mean-square deviation (rmsd) in Cartesian coordinates from a reference structure. The fitting procedure avoids coupling of external and internal modes in the remainder of the analysis. The Cartesian coordinates of the transformed trajectory are used as input for a principal component analysis, that is, the eigenvalue decomposition of the covariance matrix of all of the mass-weighted Cartesian coordinates is computed. One obtains the so-called principal components by transforming the Cartesian coordinates to the basis of eigenvectors. These principal components can be interpreted as time-dependent fluctuations along the eigenmodes of the covariance matrix. Hess[41] showed that the principal components of an incomplete molecular dynamics simulation can have a significant overlap with a cosine function, $\cos(n\pi t/t_{total})$, for which n is the index of the eigenmode, t is the simulation time, and t_{total} is the length of the simulation. The cosine content of a PCA mode is defined as the square of the dot product of the normalized principal component with the normalized corresponding cosine function. This is a number between 0 and 1, and any significant deviation from 0 for any of the principal components is an indicator for poor sampling. This is again only a necessary condition, and

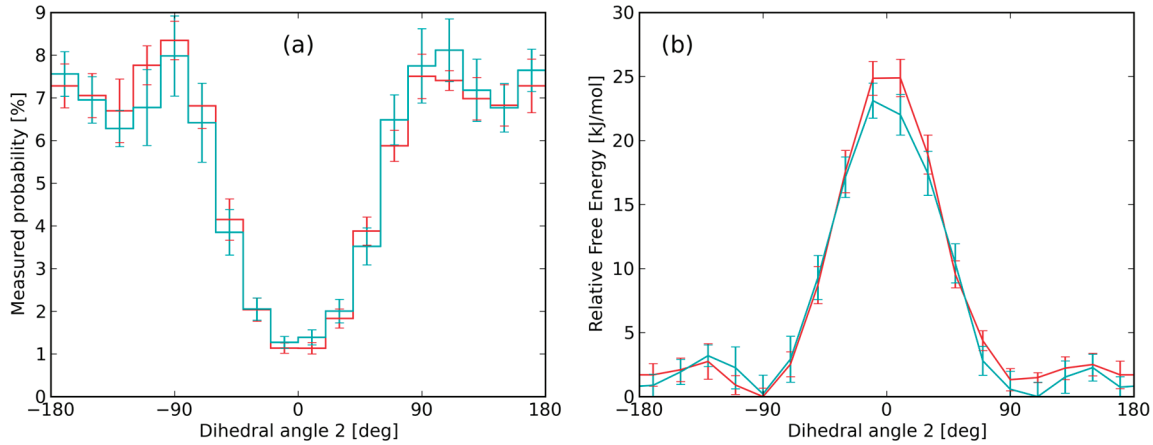


Figure 5. (a) A histogram with measured probabilities for dihedral angle 2 in conformer *E*-(*R*,*S*). The red and cyan curves correspond to the torsion along the *R* and *S* centers, respectively (see Figure 1 and 4). (b) The corresponding free-energy profiles.

one must be aware that false negatives are possible. This convergence test based on the principal component analysis will be discussed for the systems under study in the Results and Discussion section.

The last step in the protocol is the selection of distinct samples with the Kennard−Stone algorithm.[45] The goal of the Kennard−Stone algorithm is to select a maximally diverse subset from a large set of candidate samples. The algorithm assumes that one can define a "distance" between two samples, which is low when the two samples are similar and high otherwise. In this paper, the samples are molecular geometries, and we define the following distance measure between geometry μ and ν

$$D_{\mu\nu} = \sqrt{\sum_i \sum_{j>i} \left(\frac{r_{ij\nu} - r_{ij\mu}}{(r_{ij\nu} + r_{ij\mu})/2} \right)^2}$$

in which the double sum runs over all pairs of atoms, $r_{ij\nu}$ is the distance between atom i and atom j in sample ν, and $r_{ij\mu}$ is the corresponding distance in sample μ. This distance measure expresses a rmsd between the distance matrices of the two geometries weighted by the inverse of the average distance matrix. These weights are a mathematical expression of the intuitive notion that changes in short distances are more important.

Given a distance measure, the Kennard−Stone algorithm works as follows. One starts with a large set of candidate samples and an empty set of selected samples. The first two selected samples are the two candidates with the maximal pair distance. All subsequent samples are selected in an iterative way until the number of selected samples reaches a desired value. In one iteration, the candidate with the largest minimum distance to all of the previously selected samples is selected. In this paper, the set of candidates counts 30000 structures, sampled from the molecular dynamics trajectory with an interval of 10 fs. From this set, 100 samples are selected with the Kennard−Stone algorithm.

The primary motivation for the Kennard−Stone algorithm is the selection of maximally diverse conformers. Duplicates or very similar structures are avoided in the computationally expensive postprocessing steps. Although the molecular dynamics simulation samples the canonical ensemble, the structures selected by the Kennard−Stone algorithm are not distributed canonically. They are rather uniformly distributed over the accessible conformational space. This basic principle of the Kennard−Stone algorithm can best be compared with the behavior pattern of asocial people entering the theater one by one. They prefer to choose a seat as far as possible from each other. When a new person enters and finds a seat in a relatively vacant region, he will immediately occupy it. Hence, all regions will be equally populated, and the distribution of occupied seats becomes uniform over the entire room. This is independent of the original distribution of the seats. The uniformity of the selected samples is in our advantage since it will cancel out relative probabilities in the conformational space that are inherent to the semiempirical method. When the samples are postprocessed at a higher level of theory, the artifacts from the semiempirical method on the selected geometries are reduced. The only remaining assumption is that both the semiempirical level and the higher levels of theory used in the postprocessing have a comparable accessible conformational space. This can be validated with an approximate correlation between PM3 energies with more accurate values. The application of a correction would be a possible extension of the methodology proposed here and would have similarities with the importance sampling method.[11]

Thermodynamic Analysis of the Selected Samples. Geometry optimizations were performed with Gaussian03[46] on all 400 selected conformations using the B3LYP[47,48] LOT with the 6-31G(d)[49] basis set. Within each set of 100 samples of a particular stereomeric configuration, a frequency calculation was performed for the 10 most stable conformations at the same B3LYP LOT. In the case where negative frequencies were present, an additional optimization was performed using the "Opt=tight" and "Int=Ultrafine" options. The frequencies were calculated at the same LOT using the normal-mode analysis. The thermal contribution to the Gibbs free energy, based on these frequencies, was also computed by Gaussian03[50] at 353.2 K, the boiling point of benzene. In addition, for the selected (40) minimum energy conformations, single-point energies were calculated at a MP2[51]/6-311G(d,p)[52] post-Hartree−Fock level as well. They should account for the dispersion interactions that are not accurately described by B3LYP. The reported Gibbs free energies, referred to as the MP2 Gibbs free energies, were calculated as the sum of the MP2 single-point electronic energy and the B3LYP thermal contributions to the Gibbs free energy.

Conformational Analysis of the MD Simulation. In addition to the conformational sampling protocol, one can also analyze the trajectory of the molecular dynamics simulation as such. The shape of the dihedral free-energy profiles is compared among the different isomers, and one can infer structural properties from these data. We also performed a similar

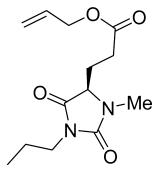


Figure 6. Half of a macrocycle used to simulate the effect of the chiral center on the dihedral angles close to the hydantoin ring.

molecular dynamics simulation on one-half of the macrocycle, the chain molecule shown in Figure 6, to estimate the strain effects imposed by closing the macrocycle and to obtain insight into the various structural data in a closed and in an open system.

The relative orientation and position of the two hydantoins in the macrocycle is further examined with four specific internal coordinates. On the basis of the article of Cremer and Pople,[53] a time-dependent axes frame is attached to the five-membered ring in each hydantoin. The centers of the rings and the axes orthogonal to the ring planes are used as parameters to characterize the orientation and position of the hydantoins. The normal vector of a five-membered ring always points toward the CH_2 group connected to the stereocenter. The following coordinates were used: α, the angle between the two normal vectors; d_{cc}, the distance between the two centers; and β_0 and β_1, the angles between each normal vector and the vector connecting the two centers.

Results and Discussion

Conformational Sampling Protocol. For each stereomer, a molecular dynamics simulation has been carried out at the PM3 level. No significant drift in the conserved quantities was present. Multiple internal rotations were observed in all dihedral angles except for the double bond and the bonds in the five-membered rings. There were no transitions between *R* and *S* stereocenters.

A full overview of all dihedral free-energy profiles is given in the Supporting Information, part 1. The numbering of the dihedral angles used in these plots is given in Figure 4. A representative example (dihedral 5 of the *E*-(*R*,*S*) isomer) is given in Figure 7, resulting from simulations with varying lengths in time. As explained in Figure 5, the symmetrical equivalent angles on both sides of the double bond are plotted. It is clear that for simulations shorter than 300 ps, statistical artifacts appear, and the error bars on the free energies increase in size. The simulations of 300 ps give equal free-energy profiles for symmetrically equivalent dihedral angles in all cases, confirming that a simulation time of 300 ps is sufficiently long for our sampling purposes. Note that due to differences in the chirality of stereocenters in the (*R*,*S*) isomers, some free-energy profiles are each other's mirror images. In these cases, the reflected curves are also included in the Supporting Information to facilitate the comparison.

TABLE 1: Cosine Content (%) of the First Five Principal Components from the 300 ps Molecular Dynamics Simulation on Each Conformer

cosine content [%]	*E*-(*S*,*S*)	*E*-(*R*,*S*)	*Z*-(*S*,*S*)	*Z*-(*R*,*S*)
mode 1	0.03	0.42	0.59	0.12
mode 2	0.01	0.36	0.12	1.46
mode 3	0.02	0.03	0.10	0.01
mode 4	0.09	0.15	0.10	0.23
mode 5	0.47	0.13	0.00	0.90

The most important results from the principal component analysis are summarized in Table 1. It is clear that all cosine content values are close to zero, and there is no indication that the molecular dynamics simulations are too short.

From each simulation, 100 samples are selected following the Kennard−Stone procedure outlined above. All 400 geometries selected with the Kennard−Stone algorithm are optimized at the B3LYP/6-31G(d) level of theory. Figure 8a and b gives an overview of the (optimized) potential energies for the (*S*,*S*) isomers and (*R*,*S*) isomers, respectively. Before plotting, the energies are sorted, and the lowest energy is taken as the reference value. The dashed lines are an estimate of the uncertainty on these curves. The error estimates are generated with a Monte Carlo procedure that is based on the

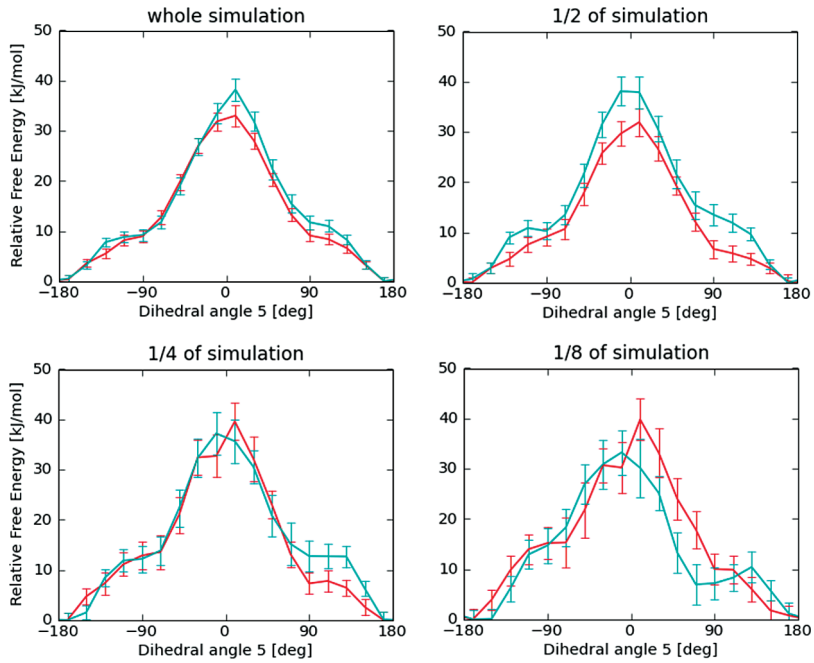


Figure 7. Free-energy profiles for dihedral angle 5 of the *E*-(*R*,*S*) conformer, based on different fractions of the trajectory (from left to right: 300, 150, 75, and 37.5 ps). The red and cyan curves correspond to the torsion along the *R* and *S* centers, respectively (see Figures 1 and 4).

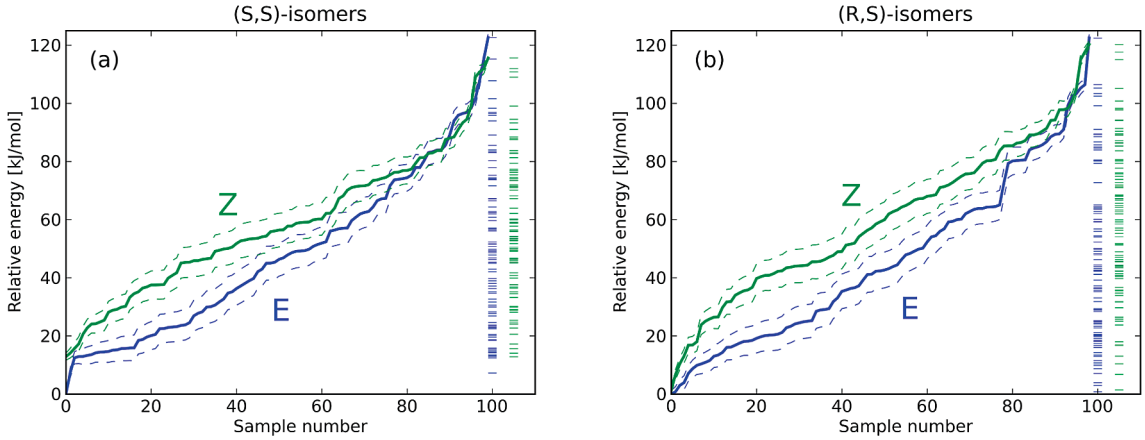


Figure 8. The solid line represents the sorted B3LYP/6-31G(d) energies for the four stereomers relative to the lowest-energy conformation. The dashed lines are error estimates on the solid lines. In the right part of both plots, the potential energy of each sample is displayed for the *E* and *Z* conformations by a single dash.

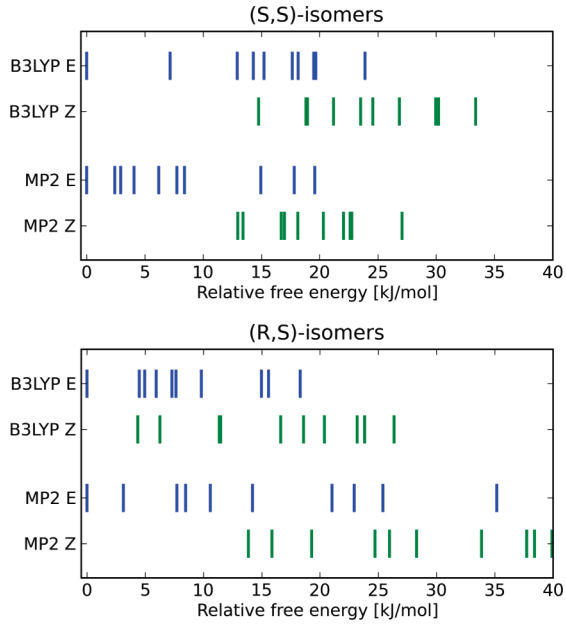


Figure 9. Comparison of B3LYP and MP2 free energies at 353.2 K for the 10 most stable samples of each conformer.

approximation that the energies within each category are distributed uniformly over a range of 120 kJ/mol. In both Figure 8a and b, the *E* isomer has significantly more low-energy conformations than the *Z* isomer.

To remove any ambiguity concerning the LOT dependence of the obtained results, the free energies for the 10 most stable conformations of every stereomer are computed at the MP2/6-311G(d,p) level as discussed before and are plotted in Figure 9. The preference for the *E* isomers is even more pronounced at the MP2 level than that at the B3LYP level. At MP2, there is a difference of 13 kJ/mol between the most stable *E* and *Z* isomers of the (*S*,*S*) stereomers. This is 14 kJ/mol for the (*R*,*S*) stereomers.

A Boltzmann distribution at 353.2 K is computed with these MP2 free energies for the (*S*,*S*) and the (*R*,*S*) isomers to make a crude estimate of the relative prevalence of the *E* and the *Z* isomers in the experimental setup. The (*R*,*S*) and (*S*,*S*) stereomers are not compared to each other as these are interchangeable under basic conditions, as applied during the pyroglutamate−hydantoin ring transformation but not during the RCM. The results are visualized in Figure 10 and show a clear preference for the *E* isomer, which is in accordance with the experimental results.

Figure 11 clearly shows the correlation between the B3LYP and the PM3 energy based on the 400 optimized macrocycle geometries. It is therefore acceptable to analyze the molecular dynamics simulations used for the sampling procedure as such to gain more insight into the structural properties of the bishydantoins. In the first instance, we will analyze the free-energy profiles of the backbone dihedral angles. In the remainder of the discussion, the relative orientation and position of the hydantoin pair are examined.

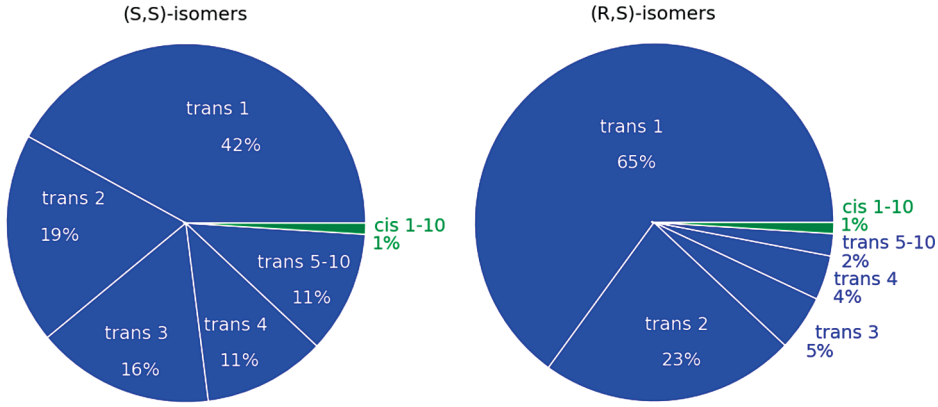


Figure 10. Boltzmann distribution of the MP2 free energies at 353.2 K of the 10 most stable conformations of both double-bond isomers. *E*1 is the most stable *E* conformation among the samples, *E*2 the second most stable, and so on. The contributions of the *Z* conformations are small and combined into one segment.

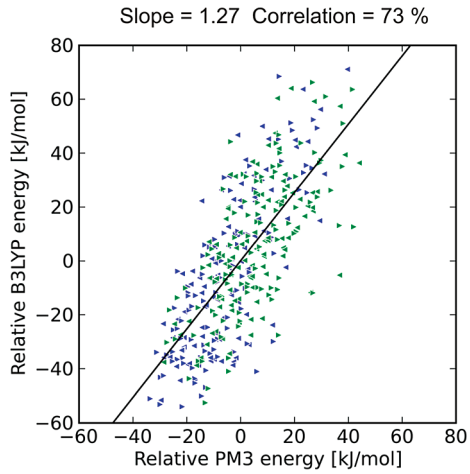


Figure 11. Correlation between PM3 and B3LYP energies. Blue and green triangles correspond to *E* and *Z* isomers, respectively. Left and right triangles correspond to (*S*,*S*) and (*R*,*S*) isomers, respectively. All energies are relative to the mean PM3 or B3LYP energy.

A complete overview of all of the dihedral angle free-energy profiles is given in the Supporting Information, part 1. Flexible dihedral angles can be recognized by low free-energy barriers, which corresponds to a nearly uniform sampling of the dihedral angle.

The major difference between the *E* and *Z* isomers was found for dihedral angles 1 (next to the carbon−carbon double bond). For the *E* isomers, the energy barrier is lower than 10 kJ/mol. The higher barrier, about 20 kJ/mol, for the *Z* isomer is mainly caused by steric hindrance with moieties at opposite positions relative to the carbon−carbon double bond. For the subsequent dihedral angles, the differences between the double-bond isomers become negligible.

A second type of differences can be observed between (*R*,*S*) and (*S*,*S*) isomers. Dihedral angles 6, 7, and 9 (Figure 12) have identical profiles in the (*S*,*S*) case, but in the (*R*,*S*) case, the free-energy profiles near the *R* stereocenter have an opposite dependence on the dihedral angle compared to those near the *S* center. The origin of these differences is obviously related to the chirality of the hydantoin, but the mechanism for each case (6, 7, and 9) is slightly different. Case 7 is simply caused by the stiff valence interactions of the ring and the bonds toward the stereocenter. In case of dihedral angle 6, the nonbonding interaction partners are chiral, which directly leads to an asymmetric dihedral angle profile. Dihedral angle 9 has no simple explanations based on local interactions. A molecular dynamics simulation on half of a macrocycle (a single chain starting from the carbon−carbon double bond and ending at the opposite end of the cycle; see Figure 6) reveals no chirality at dihedral angle 9. Therefore, the trends in this angle can only be explained by constraints due to the closed cycle or by hydantoin−hydantoin interactions.

Additional understanding of the macrocycle structure can be gained by analyzing the relative position and orientation of the two hydantoins. The average distance between the centers of the two five-membered rings is 7.82 ± 0.03 Å in the *E* case and 7.69 ± 0.03 Å in the *Z* case. Given the approximate radius of a hydantoin group, 2.5 Å, direct interactions between the hydantoin groups are unlikely to have a major effect on the structure. The time average of the angle between the normal vectors attached to the five-membered rings ranges from 80 to 100° in the four cases. This average angle of approximately 90° corresponds to a negligible correlation of the two normal vectors. This is again an indication that the two hydantoins have no significant direct interactions.

During the entire simulation of the four isomers, the methyl substituents on the hydantoin rings point to the outer side of the macrocycle. The angle between the normal vector of a five-membered ring and the vector connecting the two ring centers is practically identical in all cases, on average, 67.8 ± 0.7°. Note that the sense of the normal vector always points toward the CH_2 group connected to the stereocenter. Figure 13 gives a schematic representation of the average hydantoin orientation

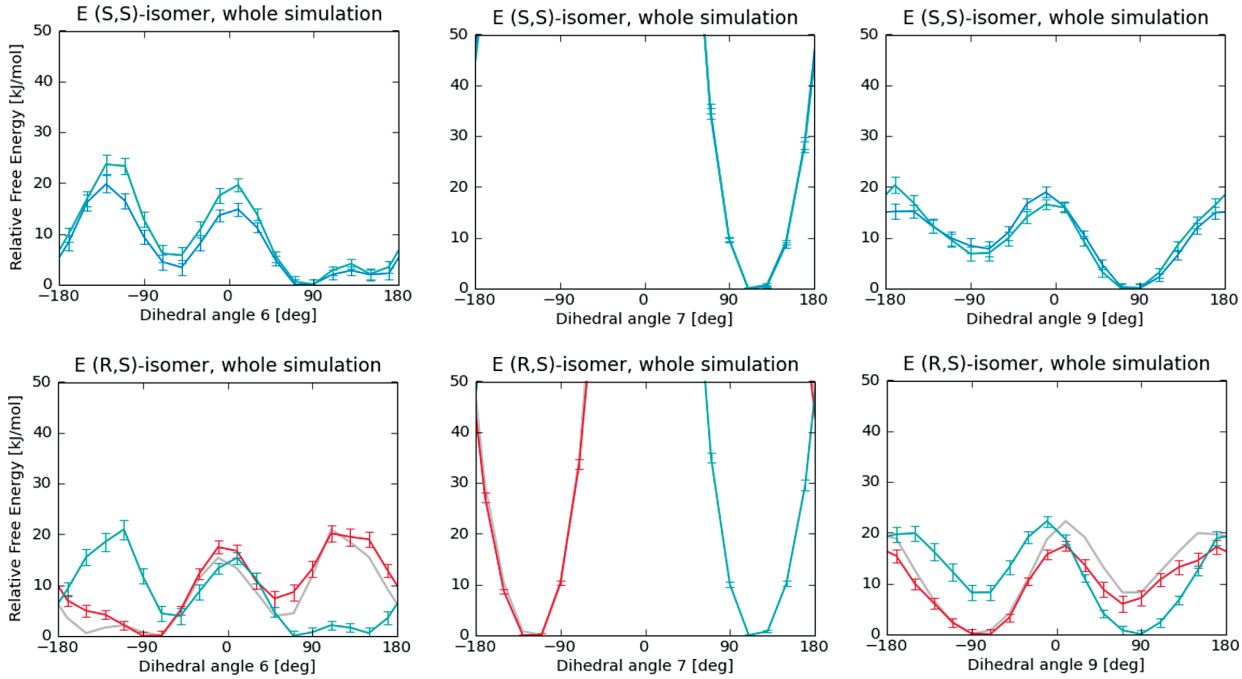


Figure 12. Free-energy profiles of dihedral angles 6, 7, and 9 in conformers *E*-(*S*,*S*) and *E*-(*R*,*S*). The cyan or blue curves correspond to *S* stereocenters. The red curves correspond to *R* stereocenters. To facilitate the comparison with the red curve, the gray curve in the plots of the *E*-(*R*,*S*) isomer is the mirror image of the cyan curve.

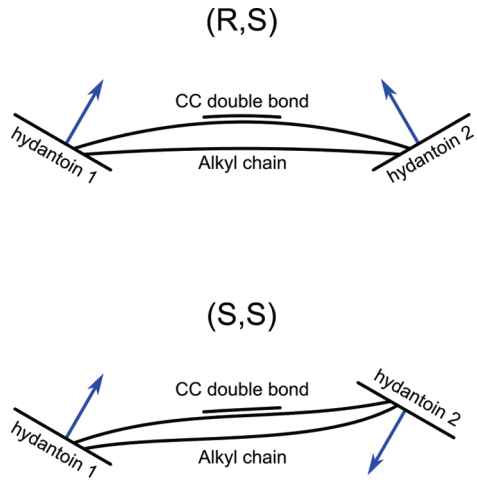


Figure 13. Schematic representation of the relative orientation of the hydantoin rings and the (*S*,*S*) and (*R*,*S*) case.

in both the (*S*,*S*) and (*R*,*S*) cases. It can be seen that the hydantoins are, on average, tilted with the CH_2 group on the stereocenters toward the double bond. This orientation imposes the least strain on the chain between the two hydantoins on the side of the carbon−carbon double bond. The alkyl chain that connects the two other ends of the hydantoins tries to relieve its own strain too. This is only possible when dihedral angle 9 synchronizes with the chirality of the stereocenters.

Conclusions

This work presented a suitable methodology to easily generate molecular conformations that span the potential energy surface. We used it to validate the experimentally observed thermodynamic stability of the double-bond isomers of large macrocycles, formed by RCM, with many degrees of freedom. Due to the large flexibility, this is a computationally challenging task. Our method starts with a fast generation of molecular geometries covering the potential energy surface by a molecular dynamics simulation with a semiempirical method at high temperature. Dihedral free-energy distribution plots and principal component analysis are used to test the completeness of the molecular dynamics trajectory. Next, a diverse set of molecules is selected by a Kennard−Stone algorithm with a distance measure based on the distance matrix. This selection of conformations is then optimized at a DFT or post-Hartree−Fock level. By calculating probability distributions based on the MP2 calculations that include effects of electron correlation, more insight is obtained in the number of conformers contributing to the experimentally observed findings. Indeed, the *E* isomers are 10 kJ/mol more stable than the *Z* isomers and determine the overall distribution. This corresponds with the *E* isomer being the only experimentally observed double-bond stereoisomer.

Furthermore, the rotational free-energy barriers were investigated for all dihedral angles. It was found that the bonds next to the double bond reflect the main difference between the double-bond isomers. The other dihedral free-energy profiles are independent of the carbon−carbon double bond. The larger part of the ring backbone is rather flexible, except for the few bonds in the hydantoins, which is the reason for the large number of different conformations. Each stereocenter has three related dihedral angles that show asymmetric behavior. Two can be explained by local interactions with the stereocenter, but the asymmetry of the third dihedral angle is only present when the macrocycle is closed. Further analysis shows that this preference is caused by relieving strain in the chains that connect the two hydantoins.

The novel methodology presented in this work is completely implemented in the MD-Tracks code, which is distributed as open source software on our Web site http://molmod.ugent.be/. This further proves the usefulness of this code as an adequate tool for advanced analysis and postprocessing of molecular dynamics simulations.

Acknowledgment. The authors thank the FWO (Fonds voor Wetenschappelijk Onderzoek − Vlaanderen, Fund for Scientific Research − Flanders), the research board of the Ghent University, and the IAP-BELSPO project in the frame of IAP 6/27 for financial support of this research. Computational resources and services used in this work were provided by Ghent University.

Supporting Information Available: The dihedral angle free-energy profiles and optimized configurations. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Liwo, A.; Czaplewski, C.; Oldziej, S.; Scheraga, H. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134–139.
(2) Christen, M.; van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *29*, 157–166.
(3) Lipton, M.; Still, W. C. *J. Comput. Chem.* **1988**, *9*, 343–355.
(4) Goodman, J. M.; Still, W. C. *J. Comput. Chem.* **1991**, *12*, 1110–1117.
(5) Kolossvary, I.; Guida, W. C. *J. Comput. Chem.* **1993**, *14*, 691–698.
(6) Wang, C. S. *J. Comput. Chem.* **1997**, *18*, 277–289.
(7) Lipton, M.; Still, C. W. *J. Comput. Chem.* **1988**, *9*, 343–355.
(8) Izgorodina, E. I.; Lin, C. Y.; Coote, M. L. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2507–2516.
(9) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *Bull. Math. Biol.* **1983**, *45*, 665–720.
(10) Hendrickx, P. M. S.; Corzana, F.; Depraetere, S.; Tourwé, D. A.; Augustyns, K.; Martins, J. C. *J. Comput. Chem.* **2010**, *31*, 561–572.
(11) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.
(12) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(13) Bailey, A. G.; Lowe, C. P.; Sutton, A. P. *J. Comput. Phys.* **2008**, *227*, 8949–8959.
(14) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
(15) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
(16) Amadei, A.; Linssen, A. B. M.; de Groot, B. L.; van Aalten, D. M. F.; Berendsen, H. J. C. *J. Biom. Struct. Dyn.* **1996**, *13*, 615–626.
(17) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
(18) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671–680.
(19) Mao, B. *Biophys. J.* **1991**, *60*, 611–622.
(20) Parent, B.; Kökösy, A.; Horvath, D. *Soft Comput.* **2007**, *11*, 63–79.
(21) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776–1783.
(22) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.
(23) Geyer, C. J.; Thompson, E. A. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.
(24) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(25) Claeys, D. D.; Stevens, C. V.; Dieltiens, N. *Eur. J. Org. Chem.* **2008**, 171–179.
(26) Dieltiens, N.; Claeys, D. D.; Stevens, C. V. *J. Org. Chem.* **2006**, *71*, 3863–3868.
(27) For general reviews on RCM, see: (a) Schmalz, H.-G. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1833–1836. (b) Schuster, M.; Blechert, S. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2036–2056. (c) Dragutan, I.; Dragutan, V.; Filip, P. *Arkivoc* **2005**, 105–129. (d) Fürstner, A. *Angew. Chem., Int. Ed.* **2000**, *39*, 3012–3043. (e) Grubbs, R. H.; Chang, S. *Tetrahedron* **1998**, *54*, 4413–4450. (f) Drozdzak, R.; Allaert, B.; Ledoux, N.; Dragutan, I.; Dragutan, V.; Verpoort, F. *Coord. Chem. Rev.* **2005**, *249*, 3055–3074. (g) Nicolaou, K. C.; Bulger, P. G.; Sarlah, D. *Angew. Chem., Int. Ed.* **2005**, *44*, 4490–4527. (h) Colacino, E.; Martinez, J.; Lamaty, F. *Coord. Chem. Rev.* **2007**, *251*, 726–764.

(28) For a review on the synthesis of oxygen- and nitrogen-containing heterocycles by RCM, see: Deiters, A.; Martin, S. F. *Chem. Rev.* **2004**, *104*, 2199–2238.

(29) Fürstner, A.; Thiel, O. R.; Kindler, N.; Bartkowska, B. *J. Org. Chem.* **2000**, *65*, 7990–7995.

(30) Scholl, M.; Ding, S.; Lee, C. W.; Grubbs, R. H. *Org. Lett.* **1999**, *1*, 953–956.

(31) Schwab, P.; France, M. B.; Ziller, J. W.; Grubbs, R. H. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2039–2041.

(32) Lee, C. W.; Grubbs, R. H. *Org. Lett.* **2000**, *2*, 2145–2147.

(33) Allaert, B.; Ledoux, N.; Dieltiens, N.; Vander Mierde, H.; Stevens, C. V.; Van Der Voort, P.; Verpoort, F. *Catal. Commun.* **2008**, *9*, 1054–1059.

(34) Verstraelen, T.; Van Houteghem, M.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Model.* **2008**, *48*, 2414–2424.

(35) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Model.* **2008**, *48*, 1530–1541.

(36) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(37) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601–1622.

(38) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(39) Nose, S. *Mol. Phy.s* **1984**, *52*, 255–268.

(40) von der Lieth, C. W.; Kozar, T.; Hull, W. E. *J. Mol. Struct.: THEOCHEM* **1997**, *395*, 225–244.

(41) Hess, B. *Phys. Rev. E* **2000**, *62*, 8438–8488.

(42) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910/1–031910/10.

(43) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1989; Chapter 2.

(44) Kabsch, W. *Acta Crystallogr.* **1976**, *32*, 922–923.

(45) Kennard, R. W.; Stone, L. A. *Technometrics* **1969**, *11*, 137–148.

(46) Frish, M. J.; et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford CT, 2004; see Supporting Information.

(47) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(48) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frish, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(49) (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261. (c) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*, 213–222. (d) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209–214. (e) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163–168. (f) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

(50) Ochterski, J. W. Thermochemistry in Gaussian; http://www.gaussian.com/g_whitepap/thermo.htm (2000).

(51) (a) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622. (b) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503–506. (c) Saebø, S.; Almlöf, J. *Chem. Phys. Lett.* **1989**, *154*, 83–89. (d) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275–280. (e) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 281–289. (f) Head-Gordon, M.; Head-Gordon, T. *Chem. Phys. Lett.* **1994**, *220*, 122–128.

(52) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.

(53) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

JP1022778