

Separation of variables in molecular-dynamics simulations: A criterion to estimate the quality of the approximation

Adolfo B. Poma and Luigi Delle Site*

Max-Planck-Institut für Polymerforschung, Ackermannweg 10, D-55128 Mainz, Germany

(Received 22 December 2007; revised manuscript received 26 August 2008; published 7 November 2008)

We propose a simple method to evaluate the approximation of separation of variables in molecular dynamics (MD) and related fields. It is based on a point-by-point evaluation of the difference between the true potential and the corresponding potential where the separation of variables is applied. The major advantage of such an approach is the fact that it requires only the analytical form of the potential as provided in most of the MD codes. We provide two examples of application, namely, a diatomic molecule adsorbing on a flat surface and an alkane (aliphatic) chain.

DOI: 10.1103/PhysRevE.78.056703

PACS number(s): 02.70.Ns, 05.10.-a

INTRODUCTION

A central question in molecular-dynamics (MD) simulations is the determination of reasonable criteria by which one can determine a specific set of degrees of freedom (DOF's) to employ in the simulation study. Such criteria must identify the relevant DOF's which are sufficient to properly characterize the phenomena under investigation. Examples of this sort are the mapping of an atomistic into a coarse-graining (CG) model, e.g., for large molecules (see, e.g., [1] and references therein) or the choice of "reaction variables" by which rare events in path sampling or similar schemes can be described (see, e.g., [2–5]). Usually the choice of a set of variables is guided by the chemical or physical intuition and does not always allow for a systematic control of the underlying approximations. In this context it would be useful to provide some criteria by which the quality of the choice regarding the optimal set of DOF's (or reaction variables) can be taken under control. In particular, for complex systems, given the analytic form of the potentials, important information is represented by the knowledge of how separable are the two DOF's; this, provided that the variables are not linked to each other by other constraints not contained in the potential, expresses the degree of independence of the DOF's. This information is very important because if they are independent their evolution occurs on two orthogonal spaces, and thus one can consider one specific DOF while the other can be neglected without altering the evolution of the system in the space of interest. Instead, if the DOF's are interconnected in a nontrivial way the dimensionality reduction is highly questionable. In this work we make a simple step in the direction of providing a criterion of evaluation regarding the approximation of separation of two DOF's. The quality of the approximation can be quantified *a priori* via an algorithm which requires only knowledge of the analytical form for the potential of interaction. Moreover, it does not require, *a posteriori*, any additional check. Finally, we provide two examples of a numerical application where the idea above is employed, namely, the interaction of a rigid diatomic molecule with a flat surface and the CG modeling of an alkane (aliphatic) chain.

I. THE BASIC IDEA

For simplicity let us consider a potential of the form $U = U(x, y)$; later on the extension to more variables will be discussed. Given the analytical form of U , if the two variables are exactly separable, under the approximation that they are not linked to each other by further constraints, then one has $U(x, y) = U_1(x) + U_2(y)$. This means that for any arbitrary fixed point x_0, y_0 , it would be possible to write

$$U(x, y) = U_1(x) + U_2(y) = U(x, y_0) + U(x_0, y) - U(x_0, y_0). \quad (1)$$

This represents the ideal case of separation of variables; a reasonable criterion to estimate the validity of the approximation of separation of variables (ASV) for a generic potential would be that of measuring its deviation from the ideal case of Eq. (1). Here we propose to introduce a simple measure Δ as the difference of the true potential and its corresponding expression when the separation is introduced in the fashion of Eq. (1). Given a potential $V(x, y)$, which is not exactly separable as $U(x, y)$, then $\Delta_{x_0, y_0}(x, y) = V(x, y) - [V(x, y_0) + V(x_0, y) - V(x_0, y_0)]$, $\forall x, y, \neq x_0, y_0$.

This is a point-by-point evaluation of how much the ASV for V resemble the true potential $V(x, y)$ (where no separation is performed). The advantage of a point-by-point evaluation lies in the fact that there may exist regions where the approximation works well and others where it does not. In this context, Δ can be used to define different spatial regions; one where the true potential must be used and another where the variables are separated, and thus the one of less interest may be disregarded, without altering the dynamical (statistical) evolution of the system in the space of the variable of interest. By dynamical (statistical) evolution here we mean that trajectories generated by the true potential that are infinitesimally close should not diverge when the ASV is applied. While the meaning and derivation of Δ are rather obvious, the treatment of its dependence on the fixed point x_0, y_0 is not. The less the ASV is valid, the stronger this dependence becomes. The treatment of this problem represents the main part of this work; in fact we show that the proposed procedure for choosing the fixed point can actually be used to complete the criterion based on Δ . Finally, it is important to settle a scale of values which quantifies the validity of the

*dellsite@mpip-mainz.mpg.de

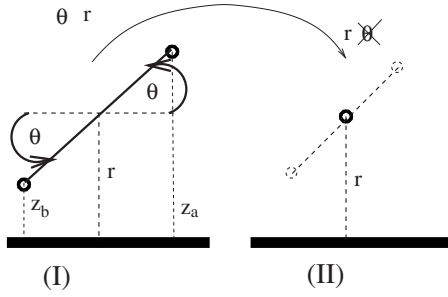


FIG. 1. Schematic representation of the system considered. In (I) the mapping from (z_a, z_b) to (r, θ) is shown pictorially. If θ and r are separable one can represent the interaction of the diatomic molecule with the surface as an effective center-of-mass–surface interaction and disregard the evolution with respect to θ , as shown in II.

ASV for a given problem. This means to determine what value of Δ one should use as a reference for estimating whether the approximation is reasonable or whether it causes the evolution of the system to deviate greatly from the one determined by the true potential. This question does not have a unique answer; rather one may fix a value for each specific problem. However, in general in MD, one deals with energies of the order of a few kT (with k being the Boltzmann constant and T the temperature of the system) and thus one may define a quality factor $Q(x, y) = \Delta(x, y)/kT$.

If Q is larger than the typical thermal fluctuations of the system, then the use of the approximation of separation of variables would necessarily lead to the wrong dynamical evolution, as this would mean that x and y are strongly correlated and do not evolve in an orthogonal space. Other criteria can be fixed by specific aspects of the problem one is interested in. For example, if one explores the conformational space of a molecule which is characterized by an energy barrier ϵ_b that separates two relevant conformational states, Δ in this case may be compared with ϵ_b (if known). In the next section we provide a numerical example of the application of Q to a test system.

II. A SIMPLE GUIDING EXAMPLE: A RIGID DIATOMIC MOLECULE INTERACTING WITH A SURFACE

Figure 1 pictorially illustrates the system we want to treat. It consists of a diatomic molecule of equivalent atoms which interacts with a uniform rigid surface via a potential

$$U(z_a, z_b) = \epsilon \left\{ \frac{2}{5} \left[\left(\frac{\sigma}{z_a} \right)^{10} + \left(\frac{\sigma}{z_b} \right)^{10} \right] - \left[\left(\frac{\sigma}{z_a} \right)^4 + \left(\frac{\sigma}{z_b} \right)^4 \right] \right\}. \quad (2)$$

The question we want to address is whether there exists a region z_a, z_b where the molecule can be treated as an “effective” pointlike particle whose interaction point is located at the center of mass (as indicated by Fig. 1). If one indicates the distance of the center of mass from the surface with r and the rotation around it by the variable θ , the terms of the problems above reduce to the question: is it possible to factorize r and θ ?

With this purpose, let us first map the variables z_a, z_b onto the variables r, θ :

$$z_a = r + d \sin(\theta), \quad z_b = r - d \sin(\theta), \quad (3)$$

and thus, substituting the expression of Eq. (3) into Eq. (2), one writes the potential as

$$U(r, \theta) = \epsilon \left\{ \frac{2}{5} \left[\left(\frac{\sigma}{r + d \sin(\theta)} \right)^{10} + \left(\frac{\sigma}{r - d \sin(\theta)} \right)^{10} \right] - \left[\left(\frac{\sigma}{r + d \sin(\theta)} \right)^4 + \left(\frac{\sigma}{r - d \sin(\theta)} \right)^4 \right] \right\}. \quad (4)$$

θ goes from 0 to $\pi/2$ (due to the symmetry of the system) and d is the length of the interatomic axis, which is fixed. The form of the potential of Eq. (2) and the values used here for the various parameters were taken (as a guide) from an atomistic model used in a previous work where a classical force field was employed to study the adsorption of molecules on surfaces [6]. By using $Q(r, \theta)$ one can determine the minimum distance r from the surface for which the separation is reasonable, and thus for distances larger than this, one can disregard the molecular rotation and represent the molecule (with respect to the surface) solely as one effective interaction site located at the center of mass. Information of this kind is crucial, for example, in emerging adaptive resolution methods, as one can define in this way regions of different resolution among which the molecule freely diffuses (see, e.g., [1] and reference, therein). For this specific case, one can define a high-resolution region close to the surface where the specific chemical resolution of the atoms is required and thus θ plays an important role, and a region far from the surface, determined by $Q(r, \theta)$ where the single-atom resolution is not relevant and thus a CG model can be used, neglecting the dynamics associated with θ . This can be done within a single simulation scheme as for example done in the AdResS scheme [1], where the region of higher resolution and that of lower resolution are interfaced via a transition region where the force acting (in this case) between the molecule and the surface is obtained via an interpolation between the atomistic force and the one coming from the CG potential. The two forces are combined via space-dependent switching functions which allow a smooth transition between the higher- and lower-resolution regions, and vice versa. The choice of such a simple system was of course made on purpose since the physics governing the interaction of such a molecule with the rigid surface can be easily deduced from the form of the potential and thus compared with the results given by $Q(r, \theta)$. In fact one expects that, as r becomes smaller, θ starts to play an important role because when its value approaches $\pi/2$ then z_b approaches small values and thus the interaction of the closer atom with the surface becomes dominant and the center of mass can no longer be used as the only effective interaction site. This prediction is in agreement with the results obtained by analyzing Q as shown in Fig. 2. In fact as r approaches the value of 1.2–1.0 nm the potential where the separation is applied overestimates the true potential of the relevant quantity (between $2kT$ and $6kT$) and, as expected, this overestimation increases as θ increases. This, together with the results of

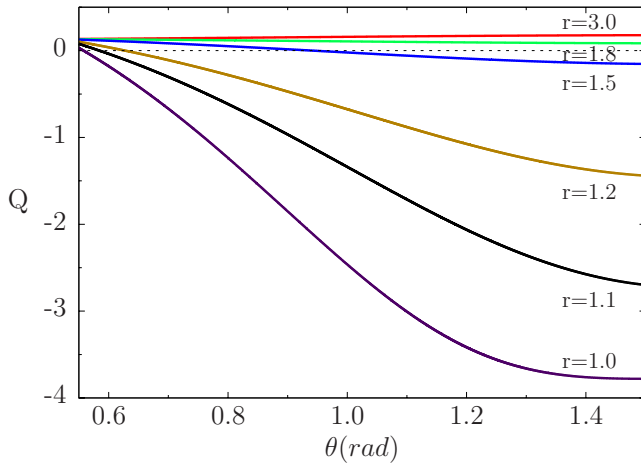


FIG. 2. (Color online) Quality factor Q studied as a function of the angle θ with r varying parametrically from 3.0 to 1.0 nm. The fixed point used in this example is $\theta_0 = \pi/6$ and $r_0 = 2.0$ nm. Here $\sigma = 0.5$ nm, $l = 0.5$ nm, and $\epsilon = 10kT$. The negative value of Q indicates that the potential term where the separation is applied overestimates the true potential. As the distance of the center of mass from the surface decreases, the dependence on θ becomes stronger, and approximately for $r \leq 1.2$ nm the ASV starts to be questionable since the error induced can be larger than $2kT$.

Fig. 2, indicates that indeed one may treat the molecule as an effective pointlike interaction site for $r \geq 1.2$ (if one chooses, for example, $2kT$ as the threshold for the potential discrepancy) without altering in a relevant way the dynamics of the adsorption process. As the molecule comes closer than this value the hypothesis of separation of θ and r is no longer valid. This conclusion was confirmed also by a study of the dependence of Q on the fixed point (θ_0, r_0) , which is discussed in the next section.

A. Dependence of Q on the fixed point as a complementary criterion

So far we have illustrated the use of Q disregarding the effects of the dependence on the fixed point (x_0, y_0) [or (r_0, θ_0) for the specific example shown]. In this section we will treat the case where this aspect is accounted for, and show that the study of such a dependence is necessary to complete the procedure of evaluation of the ASV. The natural way to investigate the dependence of Q on the fixed point (x_0, y_0) is to monitor the variation of Q upon a variation of x_0 and y_0 . This means that we define

$$\delta_{x_0} = \frac{\partial Q(x_0, y_0, x, y)}{\partial x_0}, \quad \delta_{y_0} = \frac{\partial Q(x_0, y_0, x, y)}{\partial y_0} \quad (5)$$

and calculate them over a certain range of fixed points (x_0, y_0) and on a certain (x, y) domain. In general, if the ASV is reasonable the dependence of Q on the fixed point (x_0, y_0) is negligible by construction. In fact one would have $\delta_{x_0} = (1/kT)[\partial V(x_0, y_0)/\partial x_0 - \partial V(x_0, y)/\partial x_0]$; where $\partial V(x_0, y_0)/\partial x_0 \approx \partial V(x_0)/\partial x_0$ and $\partial V(x_0, y)/\partial x_0 \approx \partial V(x_0)/\partial x_0$ because x and y are not strongly correlated and the same holds for δ_{y_0} . However, when the ASV is questionable, Q

will strongly depend on the fixed point. This, at first glance, appears as a negative aspect of the algorithm proposed; however, here we suggest that actually it can be turned into a complementary criterion to identify regions of the (x, y) space where the validity of the ASV is critical. To do so, one can study δ_{x_0} and δ_{y_0} as functions of x_0 and y_0 using x and y as the parameters to vary. This will help to identify separately a subspace of x_0 and a subspace of y_0 (which we define as “critical” and name D_0) where Q varies by a considerable amount. This subspace is actually an optimal candidates itself to be a region of the space (x, y) of high correlation. In fact a large value of the variation of Q , for example, with respect to x_0 simply means that, as one moves a little away from $y=y_0$, $V(x_0, y)$ strongly deviates from $V(x_0, y_0)$ with an infinitesimal change of x_0 . This is equivalent to saying that in the neighboring of the point (x_0, y_0) the variables x and y are strongly correlated. In this sense the analysis of δ_{x_0} and δ_{y_0} provides two different inputs: (1) It defines regions where the choice of the fixed point for Q is delicate and those where it is not; (2) it defines a region of the (x, y) space where the ASV is likely to not hold, compared with other (x, y) regions. This is complementary to the sort of information given by Fig. 2, where there was no *a priori* choice of the (x, y) region. Now, in fact, having defined the critical set of (x_0, y_0) via the analysis of δ_{x_0} and δ_{y_0} , one can go back to the study of Q , restricting the analysis to the critical region only. This means using D_0 as the (x, y) region and choosing a fixed point outside it where the dependence of Q on the fixed point is negligible (or is the minimal possible given the characteristic of the system). This part of the procedure is rather important; in fact, in addition to minimizing the fixed point dependence, it takes care also of the fact that the critical region may be small and a fixed point taken from D_0 may be too close to some of the (x, y) points, so that $V(x_0, y_0)$, $V(x, y)$, $V(x, y_0)$, and $V(x_0, y)$ are numerically very close and thus Q is small because the analysis is too local. For the specific example of the molecule at the surface, according to Fig. 3, the critical region identified by δ_{r_0} is $r \leq 1.4$ nm (by construction r cannot be less than $\sigma = 0.5$ nm) and the one identified by δ_{θ_0} is $\pi/4 \leq \theta \leq \pi/2$ (see Fig. 4). In fact Fig. 3 shows, for different values of θ and θ_0 , a trend according to which for $r_0 \leq 1.4$ nm the dependence of Q on r_0 becomes relevant. Similarly, in Fig. 4, it is shown that the critical region is $\pi/4 \leq \theta \leq \pi/2$. Thus, if one performs the analysis of Q in the region of $r \leq 1.4$ nm and $\pi/3 \leq \theta \leq \pi/2$ choosing as a fixed point $r_0 = 2.0$ nm, $\theta_0 = \pi/6$ as given in Fig. 2, it clearly emerges that for $r \leq 1.2$ nm the ASV is not valid. This shows how the complementarity of Q and δ_{r_0} , δ_{θ_0} can determine the validity of the ASV in a reasonable (robust) way.

B. Extension to higher dimensions

This criterion can be in principle straightforwardly extended to more dimensions although the computational costs would necessarily increase. This can be achieved either by considering correlation of more variables at the same time, or by looking at two variables at a time with the others frozen. In the first case, obviously, as the number of correlated variables under investigation increases, the computational

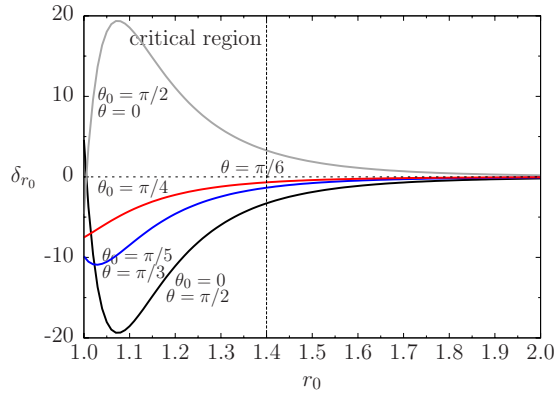


FIG. 3. (Color online) Variation of Q with respect to r_0 (δ_{r_0}) for some example values of θ_0 and θ . The symmetric curves represent the two extreme cases of values for θ_0 and θ while in between are reported two more examples. The message of this plot is that there is a general trend according to which the dependence of Q on x_0 becomes crucial for, approximately, $r_0 \leq 1.4$ nm. The vertical dashed line indicates that the region $r_0 \leq 1.4$ nm should be considered as the critical region for the variable r .

feasibility decreases, because Δ is calculated in a higher-dimensional space. The advantage in this case is that Δ would give a “direct” measure of the deviation. In the second case, suppose the potential is of the form $U(x, y, z, \dots)$; then one may concentrate on two variables at a time and see how the separation can be performed. Let us consider as an example the case of $U = U(x, y, z)$. Δ in this case would be written as $\Delta(x, y, z_0) = U(x, y, z_0) - [U(x, y_0, z_0) + U(x_0, y, z_0) - U(x_0, y_0, z_0)]$, if one is interested in how separated are x and y , under the implicit hypothesis that z can be separated from the others. In the same fashion one could analyze $\Delta(x_0, y, z)$, if one were interested in how separated are y and z , under the implicit hypothesis that x can be separated from the others, or $\Delta(x, y_0, z)$, if the interest is in the separation of x and z . Of

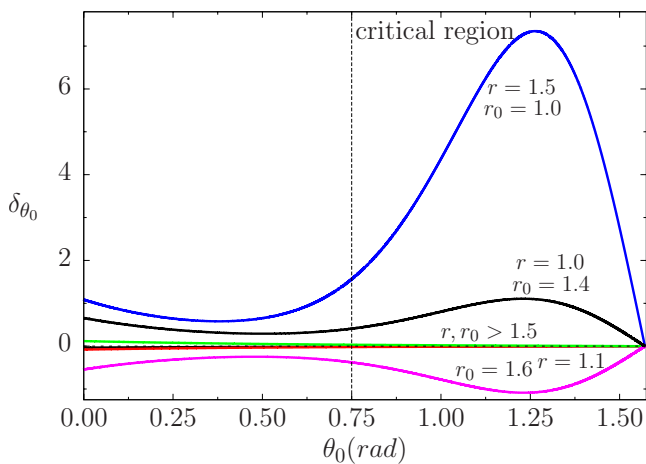


FIG. 4. (Color online) Variation of Q with respect to θ_0 (δ_{θ_0}) for some example values of r_0 and r . For values of r_0 and r outside the critical region defined by Fig. 3, the dependence of Q on θ_0 is negligible; however, for values within the critical region of r , $r_0 \leq 1.4$, the dependence becomes stronger and identifies the critical region of the variable θ . The vertical dashed line indicates that the region below $\theta_0 \geq \pi/4$ is the critical one.

course in the case of a strong three-variable correlation Δ would strongly depend on the fixed point. As shown before, such a dependence can actually be used as a further complementary criterion to that given by Δ or Q to identify regions where the correlations between variables play a crucial role. In order to provide a practical example of the case in which more than two variables are involved, in the next section we study the case of an alkane chain and use the ASV method to comment on the validity of two mapping procedures to pass from an atomistic to a CG description.

III. BEYOND TWO VARIABLES: THE CASE OF THE ALKANE CHAIN

The starting point of CG schemes for intramolecular interactions for polymers is the choice of a mapping strategy by which first a reduced set of relevant variables are identified and then are substituted for the larger set of atomistic variables [e.g., (r, θ, ϕ) where r , for example, is the distance between two atomic groups, and θ and ϕ some angles formed by three or four atomic groups], defining (in the simplest case) spheres containing chemical groups that become the “new” particles of the polymer. Next, from an all-atom simulation of a single chain *in vacuo*, the corresponding distributions of the CG variables are determined, and under the approximation that they are decoupled, one can write (see, e.g., [7])

$$P(r, \theta, \phi, T) = P(r, T)P(\theta, T)P(\phi, T), \quad (6)$$

and by Boltzmann inversion at the given temperature T one can write

$$U(r, \theta, \phi) = U(r) + U(\theta) + U(\phi). \quad (7)$$

This means that, instead of using the whole information corresponding to thousands of particles, one creates a sort of superatom approach which can still reproduce the major features of the full atomistic model but at a much lower computational cost. However, one has the disadvantage that, so far, there exists no systematic approach to control the underlying ASV implied by the CG potential derived from Eq. (7). The example shown in the next section is an attempt to (at least) partially bridge this gap.

Coarse-grained models for the alkane chain

To establish how to use the ASV criterion in polymers, we have studied in detail two different mapping schemes (MSs) for the alkane chain (AC) (see Fig. 5). They are indicated as the 1:2 and 1:3 MSs. In both cases by convention the right

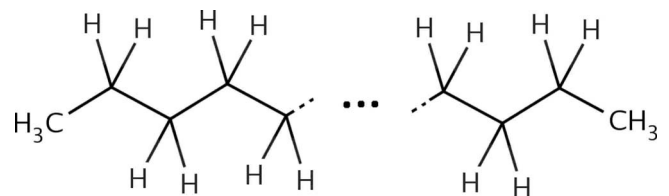


FIG. 5. Chemical structure of the alkane chain in which each vertex of the backbone represent a carbon atom.

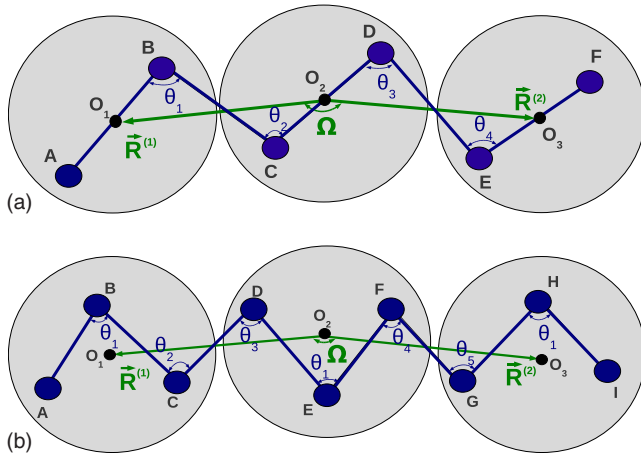


FIG. 6. (Color online) CG structure of the backbone for an AC and the underlying atomistic (carbon group) structure. Each carbon group (e.g., CH_2 or CH_3) is represented by a large dark (blue) circle close to a letter. (a) shows the MS where there are two carbon groups per bead (1:2), and (b) three per bead (1:3). The “atomistic” variables θ_i , $\forall i=1,2,3,4,5$, and the CG variables $R^{(1)}$, $R^{(2)}$, and Ω are also shown.

number denotes the number of carbon atoms which will be replaced by a spherical bead (left number). For instance, in Fig. 6, (a) corresponds to the 1:2 MS which is represented by a bead (circle in gray) and (b) corresponds to a 1:3 MS, that is, three carbon groups are replaced by one spherical bead.

Although in Fig. 6 we use the same CG variables for both mapping schemes, it is necessary to remark that each mapping scheme depends on different explicit (atomistic) variables. The CG variables are defined as (i) the distance between close beads $R^{(1)}$ and $R^{(2)}$ and (ii) the angle between vectors $\vec{R}^{(1)}$ and $\vec{R}^{(2)}$. The problem we want to address is quantification of how realistic is the hypothesis of separation

of variables, i.e., how independent $R^{(1)}$ and $R^{(2)}$, $R^{(1)}$ and Ω , and $R^{(2)}$ and Ω are of each other. To this end we apply the scheme proposed in the previous sections. The underlying atomistic potential and the procedure of coordinate transformation from the atomistic to the CG variables are reported in Appendix A.

IV. APPLICATION OF THE ASV METHOD TO THE ALKANE CHAIN: RESULTS

As stated before we are interested in the separability of the CG variables $R^{(1)}$, $R^{(2)}$, and Ω . The way we will proceed (as in the previous example) is first by considering the atomistic potential expressed in terms of the CG variables, and then by analyzing the case of dependence of two CG variables at a time with the other fixed at a parametric value. This will provide direct information about the coupling of two variables but not directly the coupling to the third one. However, later we will show also results of a study where the variable kept fixed in Q is varied parametrically and by this we will provide (at least) qualitative insight into the three-variable dependence. The procedure we follow for the “two variables at a time” analysis is the same shown for the example of the diatomic molecule, that is, (i) first choose an arbitrary fixed point; (ii) calculate the quality factor Q ; (iii) study the dependence on the fixed point by calculating the corresponding δ 's; (iv) identify the critical region; (v) chose a fixed point outside this region (the optimal choice would be where $\delta=0$; however, as explained later, this is a rather ideal situation); (vi) calculate Q once again using the fixed point of (v) and analyze Q in the critical region. Unlike the case of the diatomic molecule, in this case there are two more aspects to account for. (a) Some regions are not considered because they lie outside the domain in which the CG value are defined [e.g., an angle $\Omega=0$ is obviously a bad configu-

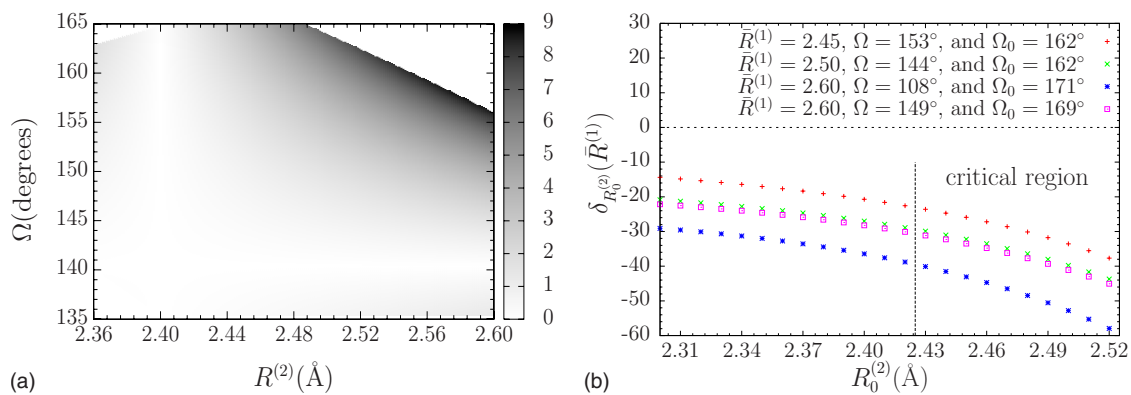


FIG. 7. (Color online) (a) Quality factor $Q_1 = Q(\bar{R}^{(1)}) = \Delta_{R_0^{(2)}, \Omega_0}(\bar{R}^{(1)}) / kT$ for the mapping 1:2 with $\bar{R}^{(1)} = 2.4$ Å as a parametric value and the fixed points equal to $R_0^{(2)} = 2.4$ Å and $\Omega_0 = 140^\circ$. In order to determine (a) we first chose an arbitrary fixed point $(R_0^{(2)}, \Omega_0)$ and then analyzed Q and its dependence on the fixed point as shown in (b). Next, according to the procedure shown for the example of a diatomic molecule, we can identify two critical regions which are given $\forall R_0^{(2)} \in [2.425, 2.525]$ and $\forall \Omega_0^{(2)} \in [145^\circ, 180^\circ]$. Accordingly, we have then chosen the fixed point of (a) outside the critical region. In (a), the light gray region shows the vicinity of the fixed point $(R_0^{(2)}, \Omega_0)$ where the CG potential is decoupled (by definition) and thus of no interest in this context. The real interest is in the critical regions determined by the plot shown in (b) [δ_{Ω_0} , not reported, shows a behavior similar to that of (b)]. Here the potential is no longer decoupled and the variables are highly correlated. Note also that the white regions are those where the CG variables are not defined and thus Q is not calculated. The same applies for all the following figures.

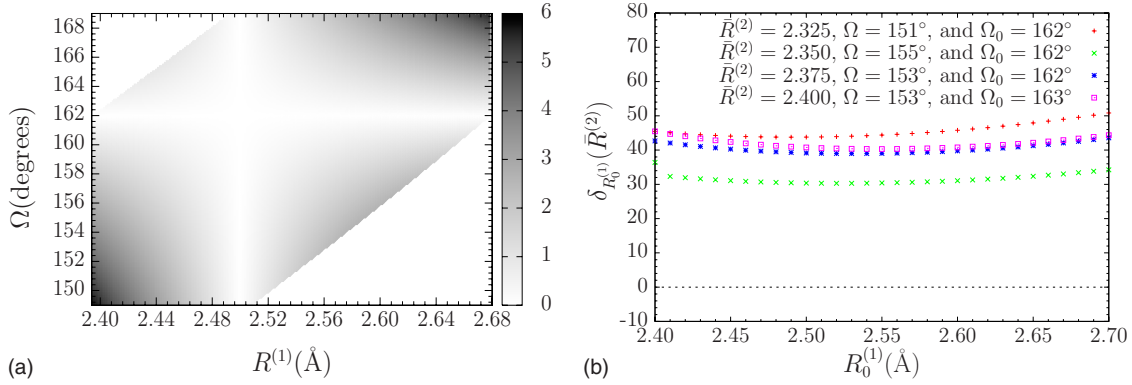


FIG. 8. (Color online) (a) Quality $Q_2=Q(\bar{R}^{(2)})=\Delta_{R_0^{(1)},\Omega_0}(\bar{R}^{(2)})/kT$ for the mapping 1:2 with $\bar{R}^{(2)}=2.35$ Å as a parametric value and the fixed points equal to $R_0^{(1)}=2.5$ Å and $\Omega_0=162^\circ$. The procedure used is the same as in the previous figure. Note that the value of the derivative shown in (b) is almost constant for each parametric curve (the same behavior holds for δ_{Ω_0} , not reported). This means that there exists no real definition of the noncritical region and in this case one can proceed by using several fixed points in the domain and averaging the quality factor resulting from each study.

ration because it implies that two carbon groups overlap or a bending which would chemically break the chain (see Appendix A for the extension of the domain of definition in terms of the “atomistic” variables θ_i). (b) In this case the analysis of the fixed point dependence can still define the critical region as before, but the optimal choice of the final fixed point is not as straightforward as that of the previous example, e.g., $\delta=0$. In fact in this case δ is not likely to be zero; thus we have to extend the previously used criterion. The way we proceed in this case is the following. (1) If δ has a region where it varies slowly and then a region where its variation increases rapidly (as for example in Fig. 9 below), we define the first region as the one where we choose the final fixed point (possibly the point corresponding to the minimum value of δ , that is, the minimum dependence on the fixed point) or we sample a few different fixed points chosen in such a way that they span the entire domain. Next we determine Q as the average over them. This optimizes the dependency of Q on the fixed point in the sense that the quantitative results are as independent as possible from it. (2) If δ is constant but is characterized by a high value or it increases rapidly over the whole domain (as in Figs. 9–12),

then the whole domain is critical, which means that the ASV does not hold. However, to quantify the error introduced by the ASV one still needs to calculate Q and thus needs to choose a fixed point. In this case we proceed by choosing several fixed points all over the domain and for each separately we calculate the Q . For the final Q we take the average (plus fluctuations) of the values obtained for each study as we do for the slowly varying case (although in this case one needs a larger sample of fixed points). Although no longer rigorous, this is a practical way to solve the problem. In fact, if the ASV does not hold well, then the procedure still gives a precise message, i.e., it tells us that on average the error introduced by the ASV is larger than the accuracy we require. Of course if instead it holds well (i.e., the error introduced by the ASV is below the required accuracy), then one has to sample several fixed points; however, in this case one can prove only that the ASV is acceptable on average, that is, there may still exist a few conformations where the ASV does not hold well. In this case also, the method would in any case provide strong indications about the general validity of the approximation.

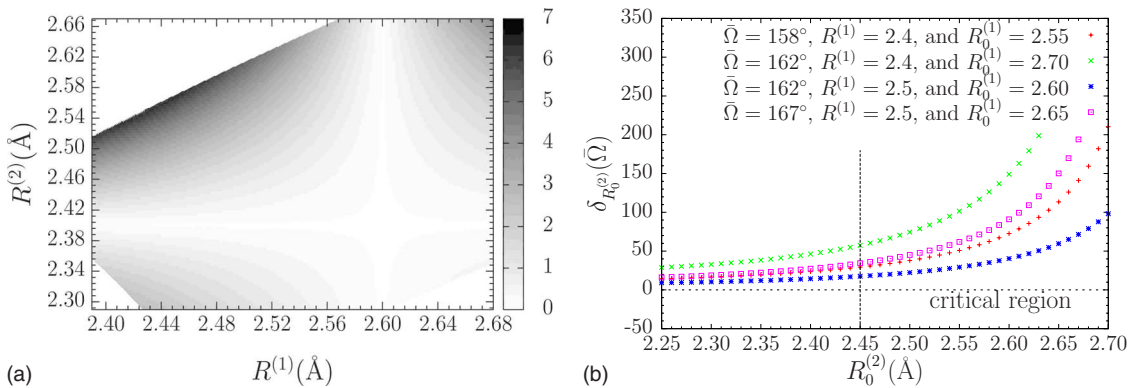


FIG. 9. (Color online) (a) Quality factor $Q_3=Q(\bar{\Omega})=\Delta_{R_0^{(1)},R_0^{(2)}}(\bar{\Omega})/kT$ for the mapping 1:2 with $\bar{\Omega}=162^\circ$ as a parametric value and the fixed points equal to $R_0^{(1)}=2.6$ Å and $R_0^{(2)}=2.4$ Å. The procedure for analyzing (b) [$\delta_{R_0^{(1)}}$, not reported, displays a similar behavior to that reported in (b)] is the same as described in the previous figures.

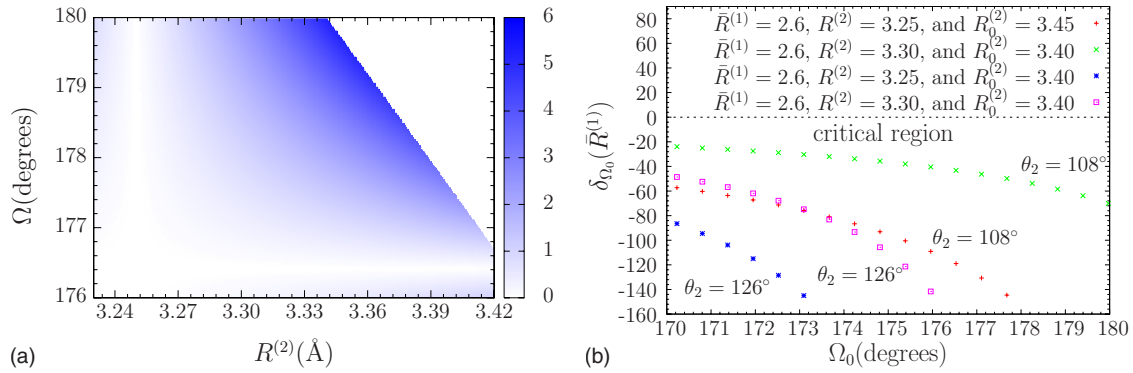


FIG. 10. (Color online) (a) Quality factor $Q_1 = Q(\bar{R}^{(1)}) = \Delta_{R_0^{(2)}, \Omega_0}(\bar{R}^{(1)})/kT$ of the mapping 1:3 with $\bar{R}^{(1)} = 2.7$ Å and $\theta_2 = 126^\circ$ as parametric values. The corresponding fixed points are $R_0^{(2)} = 3.25$ Å and $\Omega_0 = 176.5^\circ$. As for Fig. 8, but in this case due to the rapidly varying dependence on the fixed point [see (b)] one has to use several fixed points and consider the Q resulting from the average of each study, in order to have more valid quantitative information. $\delta_{R_0^{(1)}}$, not reported, shows a similar behavior.

A. Results for the 1:2 MS

Figures 7–9 report examples of Q , studied for different parametric values of the CG variable, which in turn is fixed. The general message is that indeed in each plot for the 1:2 MS there are extended regions where the error in the potential when the ASV is applied is between $6kT$ and $9kT$, which is somewhat larger than the expected thermal fluctuations. This means that indeed the ASV would not be a reasonable approximation. As underlined before, this conclusion is independent of the parametric dependence of the fixed CG variable as well as of the fixed point dependence. In fact it simply says that there exists a range of conformations for the alkane chain where the ASV would lead to the wrong potential energy (with respect to the full atomistic one) which cannot be neglected by introducing standard thermal fluctuations.

B. Results for the 1:3 MS

Figures 10–12 show the calculation of Q for the 1:3 MS. In comparison to the 1:2 MS, there is an improvement since the maximum value of Q is between $6kT$ and $7kT$. This leads

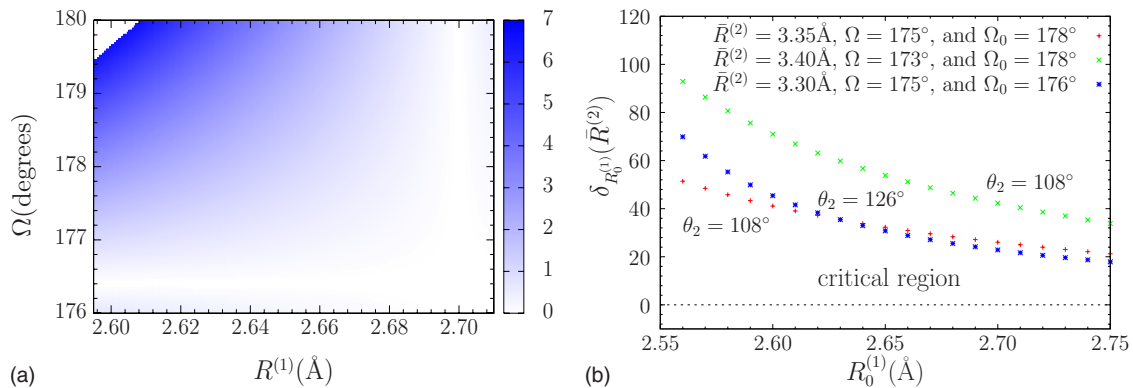


FIG. 11. (Color online) (a) Quality factor $Q_2 = Q(\bar{R}^{(2)}) = \Delta_{R_0^{(1)}, \Omega_0}(\bar{R}^{(2)})/kT$ for the mapping 1:3 with $\bar{R}^{(2)} = 3.25$ Å and $\theta_2 = 126^\circ$ as parametric values. The corresponding fixed points are $R_0^{(1)} = 2.7$ Å and $\Omega_0 = 176.5^\circ$. As in the previous figure, there is no noncritical region [see (b)]. $\delta_{R_0^{(1)}}$, not reported, shows the same behavior as in (b). This means that one must explore critical points over the whole domain in order to estimate Q quantitatively.

us to conclude that the ASV is a better approximation for the 1:3 MS than for the 1:2 MS; this somehow is expected because the coarser the system becomes the closer it is to a freely jointed chain. However, implicitly, this suggests that the 1:3 MS is a better CG model than the 1:2 (see also the analysis of the average values of Q reported in the next section). Methodologically, the relevant aspect of Figs. 10–12 is that a noncritical region cannot be identified by the analysis of δ and thus one has to proceed as suggested before by sampling several fixed points and use a statistical definition of Q . However, the message is rather clear: in this case also the error introduced by the ASV would be of the order of $(6-7)kT$ and thus still not negligible, and once again this is independent of the parametric value of the fixed CG variable and of the fixed point.

C. Average of Q in fixed point and CG variable parametric space

Finally, in order to show in practice how to deal with the situation where a noncritical region cannot be defined, we report in Tables I and II the values of Q averaged over sev-

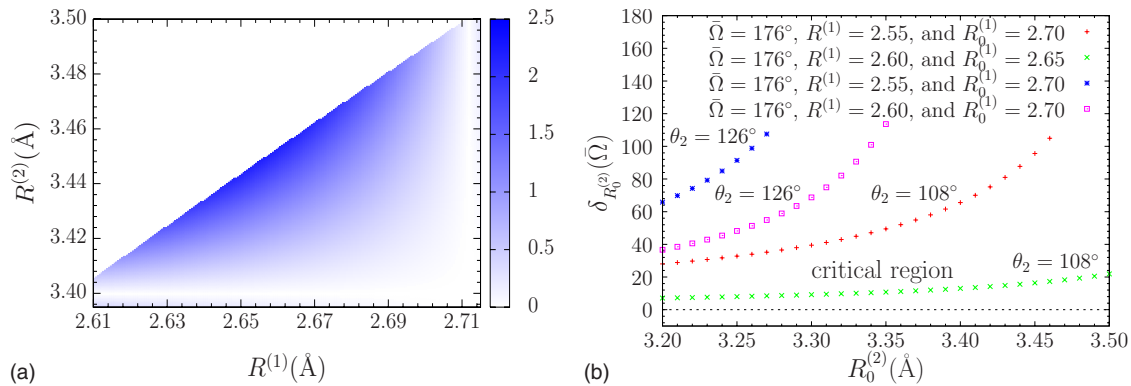


FIG. 12. (Color online) (a) Quality factor $Q_3 = Q(\bar{\Omega}) = \Delta_{R_0^{(1)}, R_0^{(2)}}(\bar{\Omega})/kT$ for the mapping 1:3 with $\bar{\Omega} = 174^\circ$ and $\theta_2 = 126^\circ$ as parametric values. The corresponding fixed points are $R_0^{(1)} = 2.71$ Å and $R_0^{(2)} = 3.4$ Å. As in the other cases for Q_1 , Q_2 , and now for Q_3 noncritical regions within the domains of the CG variables cannot be found [see (b)] and thus the same considerations apply as in Figs. 10 and 11. $\delta_{R_0^{(1)}}$, not reported, shows a similar behavior to that reported in (b).

eral fixed points. As can be noticed, as for the case of Q calculated with one single fixed point, the 1:2 MS is characterized by values much above the expected standard thermal fluctuations. For example, the value of 16.0 ± 4.60 tells us that on average the error introduced is $16kT$ with a maximum of even $20.6kT$ and a minimum of $11.4kT$, i.e., it gives a clear message about the validity of the ASV. The same could be said about the 1:3 MS; however, in this case the error introduced by the ASV is much lower. Finally, in Table III we show the average of Q over the parametric values of the CG variables, which is in turn fixed. This is an indirect indication of the three-variable dependence. According to Table III the additional correlations due to the third variable do not change the situation. The single values of Q vary by an amount that is not sufficient to change the previous conclusions, that is, the two-variable correlation is dominant with respect to the three-variable correlation.

V. CONCLUSIONS

We have proposed a systematic procedure to estimate the validity of the ASV. The advantage of such a method consists of the fact that only the analytical form of the potential is required. This means that it allows one to define regions where the ASV is reasonable *a priori* without performing any preliminary MD study for calculating possible correlations among the DOF's. We have shown its application for a simple system, namely, the diatomic molecule on a flat surface, and the application to any other system is conceptually straightforward; this is certainly true for the cases where the variables in question explicitly appear in the potential. For polymers, in general, in CG-based studies, the nonbonded

interactions are well represented via simple excluded volume; this means that the separation of variables is relevant only for intramolecular interactions [1]. These, in turn, are usually determined by local interactions (e.g., bond potentials, angular potentials, torsional potentials) involving only one specific chemical moiety of the polymer at a time. This means that the parametrization of a large molecule can be, in a rather practical way, checked piecewise by the method presented here. We have shown the application of the procedure for the case of an alkane chain and two different CG mapping procedures. In conclusion, this method may represent a first step toward the development of a more systematic and practical way to control the problem of reducing the number of DOF's in MD simulations.

ACKNOWLEDGMENTS

We are grateful to C. Peter, A. Villa, V. Harmandaris, L. M. Ghiringhelli, M. Praprotnik and K. Kremer for helpful comments on the manuscript; discussions with C. Junghans and M. Baptista are also acknowledged.

APPENDIX A: THE ATOMISTIC POTENTIAL

The atomistic model used in this study involves an interaction between atomic sites describes by a bond-bending potential [8,9] of the form

$$V_{\text{bending}}(\theta_i) = \frac{1}{2}K(\theta_i - \theta_i^0)^2 \quad (\text{A1})$$

TABLE I. Quality factor for 1:2 MS (average in fixed point space).

$\langle Q(\bar{R}^{(1)} = 2.44 \text{ \AA}) \rangle$	$\langle Q(\bar{R}^{(2)} = 2.47 \text{ \AA}) \rangle$	$\langle Q(\bar{\Omega} = 162^\circ) \rangle$
9.10 ± 0.40	16.00 ± 4.60	6.20 ± 1.60

TABLE II. Quality factor for 1:3 MS (average in fixed point space).

$\langle Q(\bar{R}^{(1)} = 2.7 \text{ \AA}) \rangle$	$\langle Q(\bar{R}^{(2)} = 3.29 \text{ \AA}) \rangle$	$\langle Q(\bar{\Omega} = 170^\circ) \rangle$
3.50 ± 1.20	7.20 ± 2.70	2.10 ± 0.60

TABLE III. Quality factor (average in parametric space).

MS	$\langle Q(\bar{R}^{(1)}) \rangle$	$\langle Q(\bar{R}^{(2)}) \rangle$	$\langle Q(\bar{\Omega}) \rangle$
1:2	8.20 ± 1.50	15.0 ± 6.00	5.40 ± 2.50
1:3	4.90 ± 2.90	6.50 ± 3.70	3.00 ± 1.60

with $K=115.2$ kcal/mol and $\theta_i^0=112^\circ$ for $T=450$ K. θ_i is the angle formed by three consecutive carbon groups, as shown in Fig. 6. Some DOF's, such as the torsion angles, are disregarded, as we focus on the separation of the variables $R^{(1)}$, $R^{(2)}$, and Ω which (for symmetry reasons) are independent of the torsion. In both systems with 1:2 or 1:3 mapping, we have used a stiff bond length between carbon atoms equal to $l_1=1.54$ Å. The difference between the two cases studied lies in the choice of the center for each bead and the number of carbon groups per bead. For instance, in the first case we take the center of the distance between two groups ($\text{CH}_3\text{-CH}_2$) as the center of the bead and in the second we fix the internal angle formed by three molecules ($\text{CH}_2\text{-CH}_3\text{-CH}_2$) and then use the barycenter of this triangle as the center of the bead. In both cases the angle θ_1 is kept fixed at its equilibrium value, that is, $\theta_1=108^\circ$, while the other angles are allowed to vary in such a way that $\theta_i=115^\circ \pm 10^\circ$, $\forall i=2,3,4$, as suggested by atomistic simulations [8,9].

APPENDIX B: THE CG VARIABLE AS FUNCTION OF THE ATOMISTIC VARIABLES

Once the relevant CG variables are determined one has to proceed to express the CG variables as a function of the explicit (atomistic) variables. After that one can apply systematically the criterion of the ASV and calculate the quality factor for this particular mapping scheme. Here, for technical convenience, we proceed first by expressing the atomistic variables as a function of the new CG variables and then use

the inverse function in order to get the CG variables as a function of the explicit dependencies. Due to the high complexity of the system we have fixed some atomistic variables (as reported before) and consider them as possible parametric variables. This is a useful procedure and a common way to treat a complex system with a high number of DOF's where some variables are more relevant than others. In order to obtain a set of equations, we analyze the geometrical properties of the MS and search for suitable relations between the explicit and CG variables. The geometrical conditions that we have used to obtain a system of equations involve the square of the absolute values of $\bar{R}^{(1)}$ and $\bar{R}^{(2)}$, which are expressed as the resultants of the sequence of collinear vectors for each case. For the first MS, now we define $\vec{r}_1=\overrightarrow{O_1B}$, $\vec{r}_2=\overrightarrow{BC}$, $\vec{r}_3=\overrightarrow{CO_2}$, $\vec{r}_4=\overrightarrow{DE}$, and $\vec{r}_5=\overrightarrow{EO_3}$ (see Fig. 6). So we have

$$\bar{R}^{(1)} = -(\vec{r}_1 + \vec{r}_2 + \vec{r}_3),$$

$$\bar{R}^{(2)} = \vec{r}_3 + \vec{r}_4 + \vec{r}_5.$$

Then the squares of the absolute values are given by

$$(R^{(1)})^2 = (r_1)^2 + (r_2)^2 + (r_3)^2 + 2[-r_1r_2 \cos(\theta_1) + r_1r_3 \cos(\theta_2 - \theta_1) - r_2r_3 \cos(\theta_2)],$$

$$(R^{(2)})^2 = (r_3)^2 + (r_4)^2 + (r_5)^2 + 2[-r_3r_4 \cos(\theta_3) + r_3r_5 \cos(\theta_4 - \theta_3) - r_4r_5 \cos(\theta_4)]. \quad (\text{B1})$$

Similarly for the second MS, having defined $\vec{r}_1=\overrightarrow{O_1B}$, $\vec{r}_2=\overrightarrow{BC}$, $\vec{r}_3=\overrightarrow{CD}$, $\vec{r}_4=\overrightarrow{DE}$, $\vec{r}_5=\overrightarrow{EO_2}$, $\vec{r}_6=\overrightarrow{EF}$, $\vec{r}_7=\overrightarrow{FG}$, $\vec{r}_8=\overrightarrow{GH}$, and $\vec{r}_9=\overrightarrow{HO_3}$,

$$\bar{R}^{(1)} = -(\vec{r}_1 + \vec{r}_2 + \vec{r}_3 + \vec{r}_4 + \vec{r}_5),$$

$$\bar{R}^{(2)} = -\vec{r}_5 + \vec{r}_6 + \vec{r}_7 + \vec{r}_8 + \vec{r}_9,$$

once again the squares of the absolute values are obtained:

$$(R^{(1)})^2 = (r_1)^2 + (r_2)^2 + (r_3)^2 + (r_4)^2 + (r_5)^2 + 2 \left[-r_1r_2 \cos\left(\frac{\theta_1}{2}\right) + r_1r_3 \cos\left(\theta_2 - \frac{\theta_1}{2}\right) - r_1r_4 \cos\left(\theta_3 + \frac{\theta_1}{2} - \theta_2\right) + r_1r_5 \cos(\theta_2 - \theta_3) - r_2r_3 \cos(\theta_2) + r_2r_4 \cos(\theta_2 - \theta_3) - r_2r_5 \cos\left(\frac{\theta_1}{2} + \theta_2 - \theta_3\right) - r_3r_4 \cos(\theta_3) + r_3r_5 \cos\left(\theta_3 - \frac{\theta_1}{2}\right) - r_4r_5 \cos\left(\frac{\theta_1}{2}\right) \right],$$

$$(R^{(2)})^2 = (r_5)^2 + (r_6)^2 + (r_7)^2 + (r_8)^2 + (r_9)^2 + 2 \left[-r_5r_6 \cos\left(\frac{\theta_1}{2}\right) + r_5r_7 \cos\left(\theta_4 - \frac{\theta_1}{2}\right) - r_5r_8 \cos\left(\theta_5 + \frac{\theta_1}{2} - \theta_4\right) + r_5r_9 \cos(\theta_4 - \theta_5) - r_6r_7 \cos(\theta_4) + r_6r_8 \cos(\theta_5 - \theta_4) - r_6r_9 \cos\left(\frac{\theta_1}{2} + \theta_4 - \theta_5\right) - r_7r_8 \cos(\theta_5) + r_7r_9 \cos\left(\theta_5 - \frac{\theta_1}{2}\right) - r_8r_9 \cos\left(\frac{\theta_1}{2}\right) \right]. \quad (\text{B2})$$

Although the scalar product of $\vec{R}^{(1)} \cdot \vec{R}^{(2)}$ is a valid relation to obtain Ω , we noticed that it is not well handled by the conventional inverse procedure because it involves no simple argument dependencies. Hence we use a particular relation for each MS as shown below.

(1) Case of Fig. 6(a). For Ω we have $\Omega = \pi - \widehat{DO_2O_3} + \widehat{O_1O_2C}$; then using the scalar product of $\vec{r}_3 \cdot \vec{r}_4 = -r_3r_4 \cos(\theta_3)$ and since $\vec{r}_4 = \vec{R}^{(2)} - \vec{r}_3 - \vec{r}_5$, we have that

$$\begin{aligned} -r_3r_4 \cos(\theta_3) &= \vec{r}_3 \cdot (\vec{R}^{(2)} - \vec{r}_3 - \vec{r}_5) \\ &= [r_3R^{(2)} \cos(\widehat{DO_2O_3}) - r_3^2 - r_3r_5 \cos(\theta_4 - \theta_5)]; \end{aligned}$$

thus by inverting this relation we obtain $\widehat{DO_2O_3}$. The remaining angle $\widehat{O_1O_2C}$ can be expressed as a function of $R^{(1)}$ and is equal to

$$\widehat{O_1O_2C} = \arccos\left(\frac{(R^{(1)})^2 - 4(r_1)^2[1 - 2 \cos(\theta_1)]}{2R^{(1)}r_1}\right).$$

(2) Case of Fig. 6(b). $\Omega = 2\pi - \widehat{DO_2F} - \widehat{O_1O_2D} - \widehat{O_3O_2F}$, and due to the symmetry of the system the last two angles on the right-hand side of the equation are mathematically similar. We now show the geometrical procedure to determine the expression for one of them, namely, $\widehat{O_1O_2D}$.

$$\Delta O_1DC: \quad \overline{O_1D}^2 = \overline{O_1C}^2 + \overline{CD}^2 - 2\overline{O_1CCD} \cos(\widehat{O_1CD}),$$

$$\begin{aligned} \Delta O_1DO_2: \quad \overline{O_1D}^2 &= \overline{O_1O_2}^2 + \overline{DO_2}^2 \\ &\quad - 2\overline{O_1O_2DO_2} \cos(\widehat{O_1O_2D}), \end{aligned}$$

where the symbol Δ indicates the triangle under consideration further defined by the letters of its vertices (e.g., ΔO_1DC is the triangle whose vertices are the points O_1 , D , and C). Next we solve for $\widehat{O_1O_2D}$ and, considering that $\overline{O_1C} = \overline{DO_2}$, we have

$$\widehat{O_1O_2D} = \frac{\arccos[\overline{O_1O_2}^2 - \overline{CD}^2 + 2\overline{O_1CCD} \cos(\widehat{O_1CD})]}{2\overline{O_1O_2O_1C}},$$

by analogy, for the other angle $\widehat{O_3O_2F}$,

$$\widehat{O_3O_2F} = \frac{\arccos[\overline{O_2O_3}^2 - \overline{FG}^2 + 2\overline{GO_3FG} \cos(\widehat{FGO_3})]}{2\overline{O_2O_3GO_3}},$$

where $\widehat{O_1CD} = \theta_2 + \widehat{O_1CB}$ and $\widehat{O_3O_2F} = \theta_5 + \widehat{HGO_3}$. Due to the symmetry we have that $\widehat{O_1CB} = \widehat{HGO_3}$, this angle can be calculated using the geometrical properties of the triangles and is equal to

$$\widehat{O_1CB} = \frac{3}{8} + \frac{1}{2} \left(\frac{5}{4} - \cos(\theta_1) \right)^{1/2}.$$

From our notation for CG variables we have

$$\overline{O_1O_2} = R^{(1)},$$

$$\overline{O_2O_3} = R^{(2)},$$

and

$$\overline{CD} = \overline{FG} = l_1.$$

Despite the appearance of a complicated mathematical procedure, technically those relations are not difficult to obtain and by using standard computational tools as MATHEMATICA [10], one can obtain “numerical” expressions of the transformation which can be directly plugged into a computer code. The explicit expressions are rather lengthy (but easy to use the numerical procedure of the ASV) and would occupy too much space, thus they are not reported here. In any case the formal procedure reported above is sufficient to reproduce all the calculations we are performing. We have tested the correctness of the explicit expressions obtained by calculating several values of the potential using both the explicit atomistic coordinates and the corresponding CG values of $R^{(1)}$, $R^{(2)}$, and Ω . Below the formal relations corresponding to the two MS's are reported; for the 1:2 MS, we have the following dependencies with θ_1 and l_1 as parametric values:

$$\theta_1 = \bar{\theta}_1,$$

$$R^{(1)} = R^{(1)}(\bar{l}_1, \bar{\theta}_1, \theta_2) \rightarrow \theta_2 = \theta_2(\bar{l}_1, \bar{\theta}_1, R^{(1)}),$$

$$R^{(2)} = R^{(2)}(\bar{l}_1, \bar{\theta}_1, \theta_2, \theta_3, \theta_4) \rightarrow \theta_3 = \theta_3(\bar{l}_1, \bar{\theta}_1, R^{(1)}, R^{(2)}, \Omega),$$

$$\Omega = \Omega(\bar{l}_1, \bar{\theta}_1, \theta_2, \theta_3, \theta_4) \rightarrow \theta_4 = \theta_4(\bar{l}_1, \bar{\theta}_1, R^{(1)}, R^{(2)}, \Omega).$$

(B3)

For the 1:3 MS, we fixed θ_1 , θ_2 , and l_1 as parametric values and obtain

$$\theta_1 = \bar{\theta}_1,$$

$$\theta_2 = \bar{\theta}_2,$$

$$R^{(1)} = R^{(1)}(\bar{l}_1, \bar{\theta}_1, \bar{\theta}_2, \theta_3) \rightarrow \theta_3 = \theta_3(\bar{l}_1, \bar{\theta}_1, \bar{\theta}_2, R^{(1)}),$$

$$R^{(2)} = R^{(2)}(\bar{l}_1, \bar{\theta}_1, \theta_4, \theta_5) \rightarrow \theta_4 = \theta_4(\bar{l}_1, \bar{\theta}_1, \bar{\theta}_2, R^{(1)}, R^{(2)}, \Omega),$$

$$\Omega = \Omega(\bar{l}_1, \bar{\theta}_1, \bar{\theta}_2, \theta_3, \theta_4, \theta_5) \rightarrow \theta_5 = \theta_5(\bar{l}_1, \bar{\theta}_1, \bar{\theta}_2, R^{(1)}, R^{(2)}, \Omega).$$

(B4)

- [1] M. Praprotnik, L. Delle Site, and K. Kremer, *Annu. Rev. Phys. Chem.* **59**, 545 (2008).
- [2] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- [3] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
- [4] E. Vanden Eijnden, in *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, edited by M. Ferrario, G. Ciccotti, and K. Binder, *Lecture Notes in Physics* Vol. 703 (Springer, Berlin, 2006), Part 2.
- [5] A. Mossa and C. Clementi, *Phys. Rev. E* **75**, 046707 (2007).
- [6] P. Schravendijk, L. M. Ghiringhelli, L. Delle Site, and N. F. A. van der Vegt, *J. Phys. Chem. C* **111**, 2631 (2007).
- [7] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, *Acta Polym.* **49**, 61 (1998); W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Büüger, *ibid.* **49**, 75 (1998).
- [8] V. Harmandaris, V. G. Mavrantzas, and D. N. Theodorou, *Macromolecules* **31**, 7934 (1998).
- [9] V. A. Harmandaris, N. P. Adhikari, N. van der Vegt, and K. Kremer, *Macromolecules* **39**, 6708 (2006); V. Harmandaris, D. Reith, N. van der Vegt, and K. Kremer, *Macromol. Chem. Phys.* **208**, 2109 (2007).
- [10] Computer code MATHEMATICA, Version 5.1 (Wolfram Research, Inc., Champaign, IL, 2004).