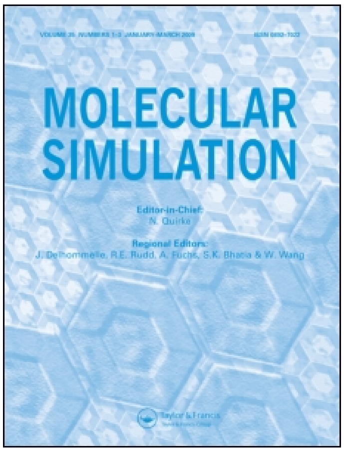
This article was downloaded by:
On: *14 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

The Effect of the Initial Guess Generator on Molecular Mechanics Calculations

O. García Pascua[a]; O. Ahumada[a]; M. Laso[ab]; M. Müller[c]
[a] Department of Chemical Engineering, ETSII, UPM, Madrid, Spain [b] Institut für Polymere, ETH Zürich, Zürich, Switzerland [c] Department of Computer Science, ETH Zürich, Zürich, Switzerland

Online publication date: 26 October 2010

To cite this Article Pascua, O. García , Ahumada, O. , Laso, M. and Müller, M.(2003) 'The Effect of the Initial Guess Generator on Molecular Mechanics Calculations', Molecular Simulation, 29: 3, 187 — 199

To link to this Article: DOI: 10.1080/0892702031000089669

URL: http://dx.doi.org/10.1080/0892702031000089669

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Molecular Simulation, 2003 Vol. 29 (3), pp. 187–199

The Effect of the Initial Guess Generator on Molecular Mechanics Calculations

O. GARCÍA PASCUA[a], O. AHUMADA[a], M. LASO[a,b,*] and M. MÜLLER[c]

[a]*Department of Chemical Engineering, ETSII, UPM, José Gutiérrez Abascal, 2, E-28006 Madrid, Spain;* [b]*Institut für Polymere, ETH Zürich, CH-8092 Zürich, Switzerland;* [c]*Department of Computer Science, ETH Zürich, CH-8092 Zürich, Switzerland*

(Received June 2002; In final form July 2002)

The effect of the initial guess generator on properties predicted by means of Molecular Mechanics has been investigated. We present a quantitative comparison of two generators, widely considered to be among the most efficient and advanced for the production of initial guesses used in the minimization of amorphous polymer structures. One of them (Polypack) is an excellent representative of and probably the most advanced generator available nowadays based on geometrical optimization strategies, whereas the other (End-Bridging Monte Carlo) has a strong statistical–mechanical footing and is based on one of the most powerful Monte Carlo schemes available for dense atomistic polymeric systems.

A large number of initial guess structures for linear polyethylene (C_{100}) in the united atom approximation were generated using Polypack and End-Bridging Monte Carlo. The potential energy, including torsional, non-bonded inter- and intramolecular energies, of these structures was subsequently minimized with respect to the torsional, single-chain positional (chain origin) and single-chain orientational (Euler angles) degrees of freedom. The predicted torsional angle distribution, inter- and intramolecular monomer pair distribution function, energy distribution and residual chemical potential were calculated for the minimized structures. Very significant differences in many of the computed properties were found between the structures generated using a geometrical optimization generator (Polypack) and those produced from a more physically founded generator (End-Bridging Monte Carlo).

In addition, a reliable probability distribution function of total energy of minimized structures for PE has been obtained for the first time.

Keywords: Molecular mechanics; Initial guess; Minimization; Geometrical optimization

INTRODUCTION

Since the pioneering work of Theodorou and Suter [1] on the prediction of structural and mechanical properties of glassy polymers, the Molecular Mechanics (MM) strategy of generating energy-minimized structures under three-dimensional periodic boundary conditions has attained the status of a standard in the field of molecular simulation. Energy-minimized structures are subsequently used in order to predict a wide variety of properties, including structural, transport, mechanical, etc. The predictions based on minimized amorphous cell structures are in many cases in excellent agreement with experimental data for a wide range of properties and detailed polymer chemistry, even though the models contain no adjustable parameters beyond those contained in the force field (which in state-of-the-art general purpose molecular simulation packages can run into the thousands). In some cases, like in the prediction of the existence of unknown PPTA crystal structures [2] some polymer structures obtained in simulations predated their experimental discovery.

The success of MM-based calculations has been so consistently remarkable, that they have swiftly found their way into commercially available software. The key components at the core of "amorphous cell" calculations are the force field (i.e. geometry plus interactions), the minimizer and the initial guess generator.

---

*Corresponding author. Address: Department of Chemical Engineering, ETSII, UPM, José Gutiérrez Abascal, 2, E-28006 Madrid, Spain.

ISSN 0892-7022 print/ISSN 1029-0435 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0892702031000089669

One of the reasons responsible for the key role played by the Initial Guess Generator in MM calculations is the strong dependence of the final minimized polymer structure upon the starting structure or initial guess. From the statistical–mechanical point of view, unlike Molecular Dynamics (MD) or Monte Carlo (MC) [3], MM does not generate an equilibrium statistical–mechanical ensemble. Nor, unlike Non-Equilibrium Molecular Dynamics (NEMD) does it generate a well-defined non-equilibrium ensemble [4]. The extraordinary success of energy minimization MM, although qualitatively justified [1], requires actually a leap of faith in the ability of the minimized structures to represent kinetically arrested glassy polymers. In a correct and efficient MD or MC calculation the starting structure should have no influence whatsoever on the properties (statistical averages) obtained during the production phase. This does not imply that the generation of the starting structure is an irrelevant issue. In fact, a poor choice of the starting structure may lead to an extremely long equilibration phase or to a "bottleneck" which prevents the desired ergodicity to be attained. However, for the majority of MD and MM applications, the generation of the starting structure is not a major issue. One of the characteristics of robust and efficient MD and MC algorithms is that they rapidly lose memory of the starting point in phase space and after an acceptably short equilibration phase, generate useful trajectories.

The situation in MM is radically different. Energy minimization does not generate a statistical mechanical ensemble, and the minimized structure closely resembles its initial guess counterpart, even if a strategy of progressive blowup of van der Waals radii allows chain backbones to cross during minimization [1]. This seemingly arbitrary procedure is however very appealing intuitively: the formation of glassy amorphous phases is widely thought to be caused or at least accompanied by a progressive freezing of molecular degrees of freedom. According to this picture, equilibrium, liquid-phase structures catastrophically lose their ability to sample configuration space upon cooling below the glass transition temperature. They essentially remain frozen in configurations which are structurally very similar to those equilibrium configurations from which they evolved. In a loose sense, the minimization procedure employed in MM attempts to mimic this effect. The results obtained amply justify this somewhat arbitrary strategy. The minimization process has been likened to a descent into the very deep wells of an "orange peel" potential energy hypersurface. Fortunately, although not entirely to be expected *a priori*, the depths of these wells have been found to lie in a relatively narrow energy band for a given polymer system and for different initial structures. It is therefore, quite natural to accept a set of minimized structures as a restricted ensemble of microstates over which to average in order to obtain the desired macroscopic average properties. This average is typically obtained by straightforward arithmetic averaging without any weighting factors, since the minimized structures cannot consistently be associated with any statistical–mechanical probability or partition function. It is only the numerical experimental observation that the potential energy minima lie quite close to each other that justifies simple averaging. On the other hand, practitioners of MM calculations often find and discard "outliers", i.e. minimized structures whose energy deviates from the simple average by an "unacceptable" margin, the interpretation of "unacceptable" largely being left to the user's discretion.

The difficulty of the general problem of generating densely packed polymer chains cannot be overemphasized: from the algorithmic point of view, the combinatorial chain packing problem can been shown to be NP-complete [5,6], i.e. optimal solutions can in principle be found only by exhaustive search. In practice, initial guesses for packed chain structures are generated and minimized almost routinely nowadays, although these tasks still require a very significant computational effort, especially for chemically complex polymers with bulky main chain and side groups. As an additional obstacle, it can also happen in practice that a sizeable fraction of the initial guesses do not lend themselves to minimization: the starting structure is entangled in such a way that none of the several powerful available minimizers is able to significantly reduce its potential energy down to the level of the other members of the set. Another typical source of failure is the production of excessively inhomogeneous initial structures by the generator: mers crowd in some regions of the amorphous cell while other regions are comparatively depleted. The minimizer is then unable to equalize the very inhomogeneous spatial distribution of mers, even if it successfully minimizes the total energy of the system.

As a consequence of the computation cost, typical MM restricted ensembles or structure sets contain a few tens of independent structures at the most. In practice, averaging over this relatively low number of structures yields acceptably small statistical error bars for many properties.

For all these reasons, the importance of the Initial Guess Generator is paramount: since the minimization procedure (usually a highly optimized, robust and efficient combination of several algorithms) hardly changes the overall conformational characteristics of the individual chains in the amorphous cell, the minimized structure will inherit many of the structural traits of the initial guess. Whereas it is easy to see that large-scale features of the polymer chains (like radius of gyration or characteristic ratio) are not modified very significantly during minimization, other, small-scale conformational features (like torsional angle and dyad distributions) or global properties (like total potential energy) change appreciably during minimization. Given the very strong influence of the initial guess on the final properties of the minimized structures, the choice of the Initial Guess Generator is therefore decisive.

The aim of this paper is to evaluate and compare in terms of efficiency, but above all in terms of property prediction, two of the most advanced Initial Guess Generators available nowadays, in order to assess their impact on predicted properties. To this end we have chosen polyethylene (PE) in the united atom approximation as the simplest test model system. Sets containing a very large number (1000) of initial guess structures have been generated using each of the Initial Guess Generators and their total potential energy fully minimized. Several selected properties have been calculated for each of the sets and compared.

INITIAL GUESS GENERATORS

Over the past 15 years, several alternative approaches to the generation of initial guess structures have appeared in the literature. Such Initial Guess Generators can be roughly grouped into several categories (see also Ref. [7] and references therein):

- *Chain growth methods*: perhaps the oldest methods, energy minimization by chain growth were originally performed using a modified Rotational Isomeric State scheme Initial Guess Generator [1,8], which produced satisfactory results for relatively short chains of atactic polystyrene.
- *Lattice methods*: although coming from a different context, the generation of self-avoiding random walks on lattices [9] can also be included in this category. In Ref. [10], coarse initial guesses are generated on a triangular lattice and relaxed by potential energy minimization.
- *Condensation methods*: the desired amorphous system is generated in a large box, i.e. at low density, where chain overlap is not a serious obstacle. The box is subsequently shrunk in a barostated *NpT* molecular dynamics run until the desired density is reached [11–13]. A somewhat similar idea has been put forward in Ref. [14]: monodisperse polymer chains are created in a dynamic two-step process. Configurations are initially predicted via Monte Carlo. During the second stage, the potential energy and the entropy production are allowed to relax while the global characteristics of the molecular conformations remain as unaffected as possible. The potentials, the distribution of bond lengths, the integration time step and temperature are smoothly controlled during the creation/relaxation process until they finally approach their prescribed values.
- *Polymerization*: this technique starts from a low molecular weight system which is then simulated while individual molecules “polymerize”, i.e. are connected until the desired chain length is achieved [15].
- *Geometric optimization approach*: in a recently proposed method implemented in the PolyPack code [7,16], the initial guess generation problem is cast as a geometric optimization problem which is then solved by heuristic search algorithms using five different types of moves.

One of the great advantages of this latter method (PolyPack [7]) is its generality, efficiency and elegance: thanks, among other things, to a newly developed MC move (ParRot [17]) it can generate dense packings of long chains for virtually any polymer structure, no matter how complex its architecture. Polypack structures lend themselves well to minimization. It also does an outstanding job at predicting the outcome of NMR measurements [18]. Furthermore, since the heuristics that lie at the core of Polypack draw heavily from geometry optimization and are far removed from statistical mechanics, PolyPack is a good example of an *ad hoc* generator. All these reasons have strongly motivated us to investigate in detail its suitability as an Initial Guess Generator. It is therefore also very natural to compare its performance with that of another Initial Guess Generator which has as little *ad hoc* character as possible. At this end of the spectrum lie methods which have a strong statistical–mechanical footing, like Meirovitch’s method [9], which is unfortunately not suitable for off-lattice models but which has inspired efficient off-lattice MC moves like Configurational Bias [19–21]. The appearance in recent years of even more efficient MC schemes (like EB-MC [22]) able to fully equilibrate very large systems of dense chain molecules, makes it possible to address the Initial Guess Generator question from a new point of view: we have followed the strategy of performing a long MC run using the very powerful suite of MC moves (including EB-MC) of Theodorou and coworkers [22,23] extracting molecular configurations along the way and using them as initial guess configurations for MM minimization. That is, we use an efficient and ergodic MC scheme as an Initial Guess Generator. This procedure is also intuitively appealing, since the structures drawn from the MC simulation are as good representatives of the melt as can be obtained nowadays. Driving the potential energy of these structures to a local minimum corresponds with the freezing out of molecular degrees of freedom that accompany vitrification.

We can summarize the main steps of this MC-Initial Guess Generator as follows:

- Start an EB-MC simulation, either in the *NpT*-ensemble if initial guess structures at several *NpT*-distributed densities are desired, or in the *NVT*-ensemble if structures at a prescribed density are desired. Generating a starting structure for this MC simulation is not a problem, since the efficiency of the end-bridging moves rapidly obliterates any memory of the starting point.
- Allow for a sufficiently long equilibration phase.
- During the production phase, save configurations with the desired frequency, ensuring that saved configurations are sufficiently decorrelated (for example, with the aid of statistical efficiency and block analysis, or by monitoring the decay of various autocorrelation functions in the same or in a separate preliminary MC run).
- These decorrelated structures are used as starting points of standard MM minimizations.

We have carried out our comparative study of the PolyPack and the MC-generators on a model system which consisted of $N_{\mathrm{ch}} = 5$ molecules of linear polyethylene of average length C_{100} placed in a cubic box with periodic boundary conditions. The system was deliberately chosen to be as simple as possible in order to prevent specific chemical details of the model polymer from obscuring the intrinsic performance of the Initial Guess Generators.

Polyethylene was modeled using a united-atom representation for both methylene and methyl end groups, non-bonded interactions being described by a Lennard–Jones potential:

$$V_{ij}^{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (1)$$

with r_{ij} being the scalar minimum image distance between sites $i$and j. Along a PE chain all pairs of sites separated by more than three bonds along the chain and all intermolecular sites interact via the Lennard–Jones pair potential. Potential tails are cut at $1.45\sigma_{\mathrm{PE}}$ and brought smoothly down to zero at $2.33\sigma_{\mathrm{PE}}$ using a quintic spline [1].

In the MC-generator run bond lengths were kept constant, whereas bond angles were assumed to fluctuate around an equilibrium angle θ_0 of 112° subject to the Van der Ploeg and Berendsen bending potential [24] of the form:

$$V_{\mathrm{bending}}(\theta) = (1/2)K_\theta(\theta - \theta_0)^2 \quad (2)$$

with $K_\theta = 482.23\,\mathrm{kJ/mol}$. Associated with each dihedral angle ϕ was also a torsional potential of the form [25]:

$$\begin{aligned} V_{\mathrm{torsion}}(\phi) &= c_0 + c_1\cos(\phi) + c_2(\cos(\phi))^2 \\ &\quad + c_3(\cos(\phi))^3 + c_4(\cos(\phi))^4 \\ &\quad + c_5(\cos(\phi))^5 \end{aligned} \quad (3)$$

with $c_0 = 9.28$, $c_1 = 12.16$, $c_2 = -13.123$, $c_3 = -3.06$, $c_4 = 26.25$, and $c_5 = -31.51$ in kJ/mol.

The EB-MC generator run was 4×10^8 steps long. It was performed in the isothermal-isobaric ensemble at 400 K and 1 bar using a combination of reptation, rotation, flip, volume fluctuations (one for every 10^4 *NVT* steps), concerted rotation (CONROT) and end-bridging (EB) moves (see Ref. [22] for details of the moves), which resulted in an average density of $801 \pm 0.2\,\mathrm{kg/m^3}$ (one standard deviation of the mean) with typical fluctuations of $40\,\mathrm{kg/m^3}$ (one standard deviation). Polydispersity was 1.09 with a lower cutoff chain length of C_{60}. One thousand structures were saved at regular intervals during the production phase of the MC run. Both from a block analysis and from orientational autocorrelation analyses we could conclude that the length of the EB-MC run between two consecutive savings was amply sufficient for the saved structures to be fully decorrelated.

One thousand PolyPack structures were produced at the same specific density of $800\,\mathrm{kg/m^3}$ (box edge of approximately 22 Å), using the same geometry as in the MC-Initial Guess Generator.

PolyPack is actually composed of a set of algorithms which generate polymer systems that avoid severe overlaps, and guarantee the chains to obey the proper conformational statistics. There are two main ideas behind PolyPack: first, the packing algorithm starts from a random configuration which is improved iteratively by means of a *heuristic search algorithm* which ultimately furnishes the target configuration. The second idea is the *principle of horizon*. Since it is difficult to manipulate an atomistically detailed, highly connected system, an incremental modification of the non-bonded interaction is used. A horizon value defines the scope beyond which the atoms are "ghosts" i.e. invisible to each other. It is increased to eventually attain its full range where each atom can see all others. The specific details of the interaction potential are irrelevant in PolyPack since it represents atoms and pseudoatoms as hard spheres.

In contrast to energy minimizations and simulated annealing techniques, PolyPack does not necessarily follow a path driven by gradients of physical energy potential. It rather tries to solve the combinatorial problem regardless of the physical problem it isderived from, whereas the MC-Initial Guess Generator's strategy is diametrically opposed: each individual configuration is guaranteed to be sampled from the corresponding equilibrium statistical–mechanical ensemble. Herein lies the main difference between an *ad hoc* approach and a physically founded one. As we will see, this difference has far-reaching consequences regarding the properties of the structures generated by both Initial Guess Generators.

The structures generated by both PolyPack and the MC-Initial Guess Generator were subjected to minimization using the same quasi-Newton matrix-updating algorithm of Broyden, Fletcher, Goldfarb and Shanno [1,26–28] again so as not to introduce any bias external to the Initial Guess Generator. Minimization was performed with respect to the translational (center of mass), orientational (Euler angles) and torsional internal degrees of freedom of the chains. Bond lengths and bond angles were kept fixed at the same values in all PolyPack and MC-Initial Guess Generator structures.

All computations were carried out on individual processors of a 24-CPU Beowulf Linux cluster composed of inexpensive 650 MHz Pentium III processors. The 4×10^8 step EB-MC run took 480 h, including the equilibration phase, i.e. around 1700 s per structure generated. Polypack required an average of 18 000 s to generate a structure. Full minimization of a single structure required 300 s for MC-Initial Guess Generator structures and 2100 s for PolyPack structures (all runtimes refer to a single processor and are averaged over all structures generated).

RESULTS AND DISCUSSION

All 1000 structures generated by the MC-Initial Guess Generator and 995 (out of 1000) structures generated by PolyPack could be successfully minimized to within stringent tolerances in the norm of the gradient and in the positive definiteness of the Hessian at the minimum of the total potential energy function. Such a very high rate of successful

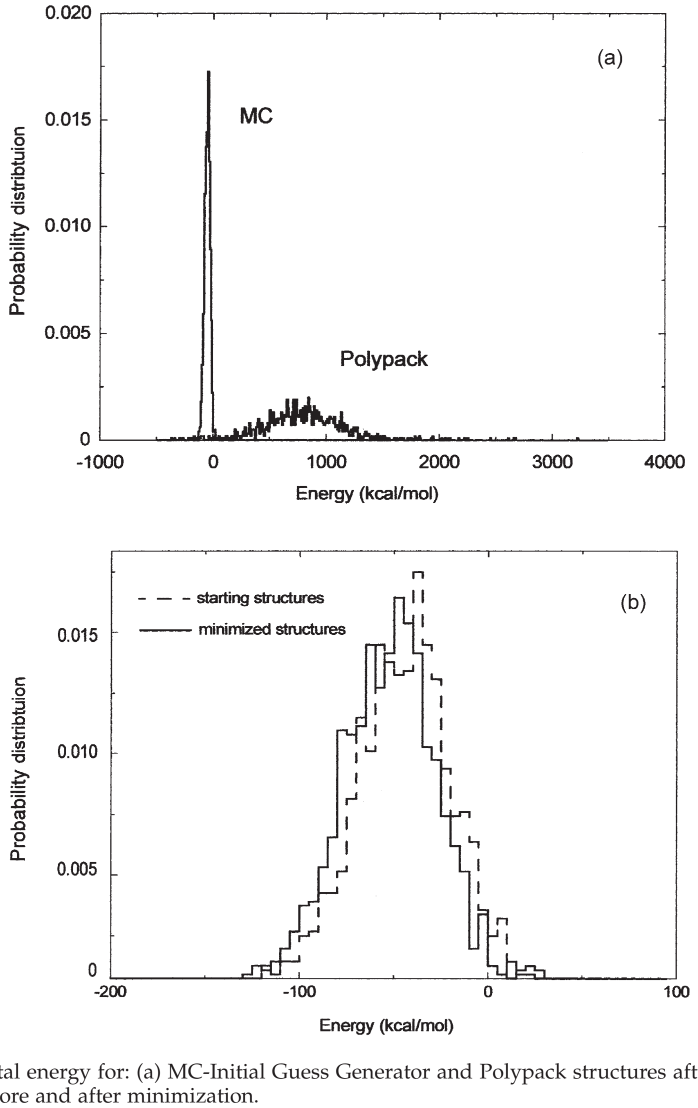


FIGURE 1 Distributions of total energy for: (a) MC-Initial Guess Generator and Polypack structures after minimization; (b) MC-Initial Guess Generator structures before and after minimization.

minimizations is in itself a remarkable achievement for both Initial Guess Generators and points to the high quality of the initial guesses. This large number of fully minimized structures is a necessary requirement for a statistically significant and unambiguous comparison of both methods. To our knowledge, it is, by at least an order of magnitude, the largest set of fully minimized structures for a given polymer to have been published to date (as a comparison, a total of 15 structures of PP were generated in Ref. [1], 20 structures of PE in Ref. [7] and less than 10 in Ref. [18]).

Energy

Figure 1(a) represents the distribution function of the total energy of minimized MC-Initial Guess Generator and Polypack structures. Both the shape and the position of the distributions differ drastically: the distribution of energy for minimized MC-Initial Guess Generator structures is (i) much narrower and (ii) has a negative average value (−54 kcal/mol) whereas PolyPack structures have a high positive average energy (840 kcal/mol) and are much more widely distributed. As a matter of fact, the average total energy of minimized PolyPack structures is much higher than the average energy of *un*minimized MC-Initial Guess Generator structures.

Figure 1(b) illustrates how energetically close the starting MC structures are to the respective minimized structures. The averages of the two distributions shown in Fig. 1(b) differ by only 14.7 kcal/mol. The ability of the suite of MC algorithms [22,23] to generate correctly sampled, highly decorrelated, low-energy, dense PE configurations is truly phenomenal. In view of the very low value of the total energy of MC initial guesses, the "orange peel" analogy mentioned above is not entirely appropriate. The MC generator produces starting structures, which are already quite close to

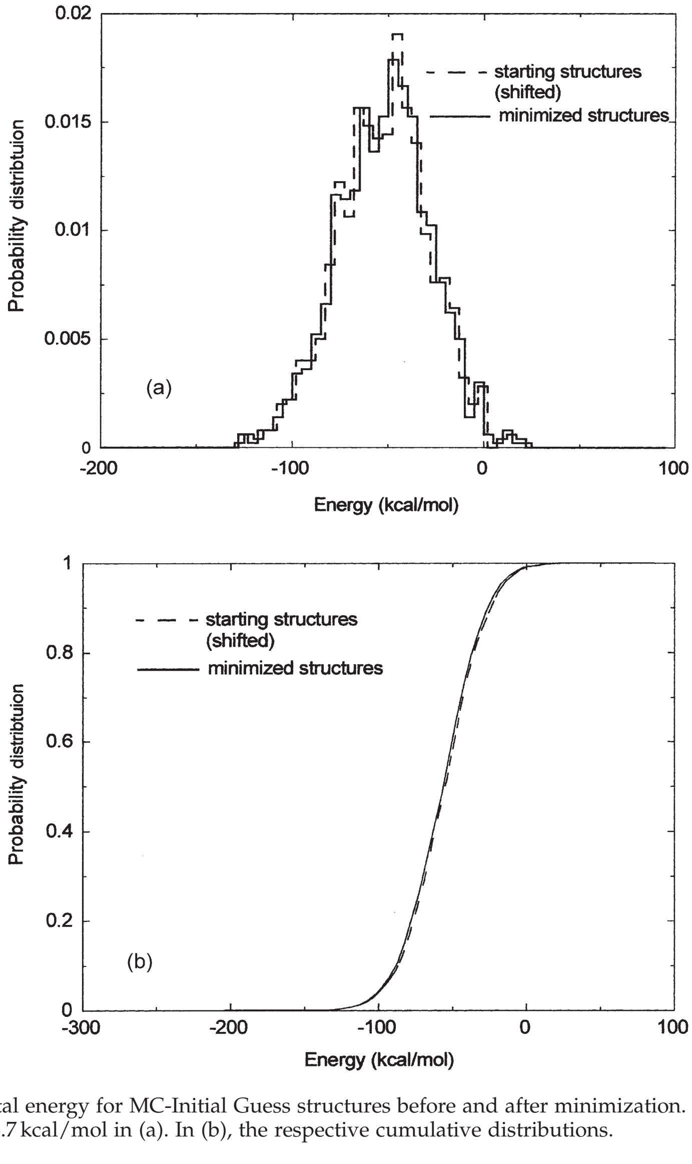


FIGURE 2 Distributions of total energy for MC-Initial Guess structures before and after minimization. The distribution of initial guess energies has been shifted by 14.7 kcal/mol in (a). In (b), the respective cumulative distributions.

a local minimum of the total potential energy. Such low starting energies also explain why 100% of the structures could be minimized and why a full energy minimization is about seven times faster than for Polypack initial guesses. No similar plots are presented for Polypack initial guesses, since due to the way Polypack generation algorithms work, its initial guesses almost always contain some overlap and therefore have unrealistically high energy.

For the MC generator, the energy distributions before and after minimization are remarkably similar, as can be seen in Fig. 2(a) and (b). The distribution of energy before minimization (dashed line) has been shifted to the left by the difference in the averages of both distributions (14.7 kcal/mol). A Kolmogorov–Smirnov test indicates that these two distributions can be considered identical to a significance level of 99.8%.

It turns out that not only are the shapes of both distributions identical but that there is a very strong correlation between the energies of the starting and the minimized MC structures as well. The parity plot of Fig. 3 shows that, apart from a shift factor, starting structures of low energy systematically lead to minimized structures of low energy. The unexpected conclusion is that the MC algorithm suite "hovers" at a constant height above the local minima. Although there is no *a priori* reason preventing a given MC initial guess from minimizing into a very deep potential well, Fig. 3 clearly indicates that MC initial guesses, drawn from a distribution representative of the melt, lead to minima which (i) are narrowly distributed and (ii) do not lie far below the starting energies. Deeper minima do indeed exists: the left tail of the Polypack distribution in Fig. 1 consists of a few structures the energy of which is lower than that of any of the MC minimized structures. On the other side, Polypack initial guesses, generated by a geometric packing algorithm, lead to minima with very widely distributed energies, most of them above the minimized MC structures.

Figures 4–6 show how the total energy before and after minimization is split into individual contributions: torsional, intramolecular and intermolecular. The respective parity plots in Figs. 4–6 show that the minimization procedure tends to reduce the intermolecular potential energy most. Intramolecular interactions are also relieved (Fig. 5) although in a lesser scale. Finally, it seems that both of these reductions are achieved through very minor adjustments in the torsional angles and therefore without incurring in any appreciable energetic penalty in the torsional contribution (the distributions of torsional angles are almost indistinguishable and the points on the parity plot of torsional energy lie virtually on the diagonal).

Finally, Figs. 7 and 8 illustrate the very significant differences in the distributions of intermolecular and torsional energy of minimized structures for both generators.

Conformation

The differences in torsional energy distributions mentioned in the previous section stem from an underlying discrepancy in the distributions of torsional angles of the minimized structures. Figure 9 does indeed reveal a severe dissimilarity between both distributions. While the MC-generator minimized structures retain a liquid-like torsional

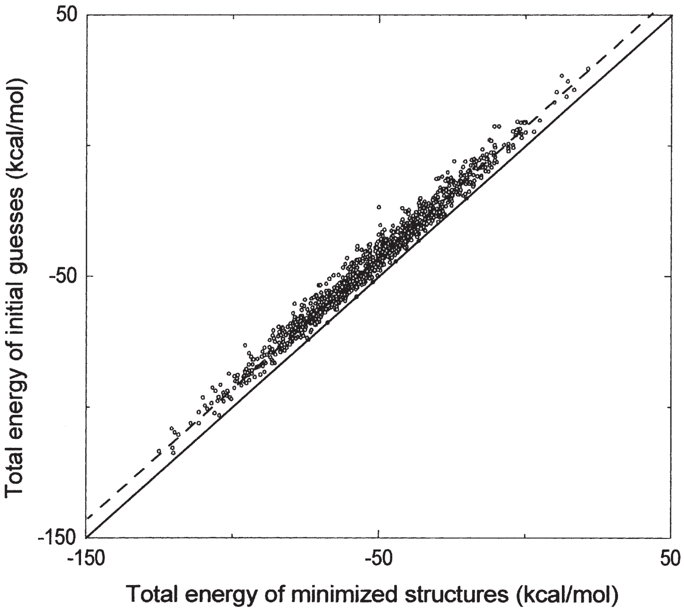


FIGURE 3 Parity plot of total energies of starting and minimized MC structures. The dotted line is drawn parallel to the diagonal and shifted 14.7 kcal/mol.

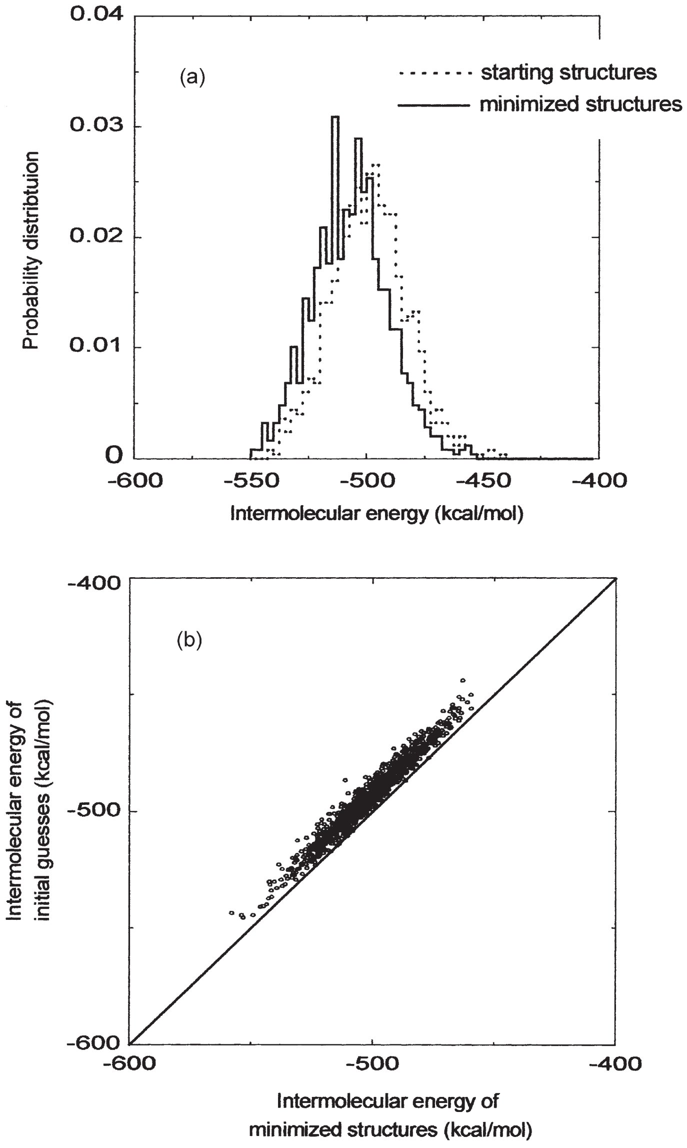


FIGURE 4 Distributions of intermolecular energy of starting and minimized MC structures (a) and parity plot (b).

angle distribution, the distribution of torsional angles of the Polypack minimized structures show rather unphysical behavior. For these structures, torsional angles are actually constrained to lie within three intervals around the *t*, g^+ and g^- bond conformations. Outside these intervals, the probability distribution function falls to zero. This unrealistic torsional angle distribution is a consequence of the packing algorithms used by Polypack: they force all torsional angles to lie within allowed intervals. This constraining strategy is the cause of the unrealistic peaks of the distribution within the allowed intervals. These peaks result from a "folding" of the distribution of torsional angles into the allowed intervals.

In view of the unphysical single-bond torsional angle distribution one would expect the consecutive bond (dyad) distribution also to behave abnormally. It is therefore all the more remarkable that the frequencies with which the possible dyads occur in Polypack minimized structures are very reasonable and, apart from a somewhat higher proportion of *tt* dyads in the minimized Polypack structures, agree well with those obtained from minimized MC-structures (Fig. 10). As a matter of fact, the correctness of the dyad distribution is a result of the packing algorithm at the core of Polypack. This algorithm has been tailored to yield structures that to a large extent fulfill rotational isomeric (RIS) probabilities [5,7]. Polypack clearly succeeds at this. The ability to reproduce RIS probabilities correctly is important if the minimized structures are to be used to predict the outcome of NMR experiments as shown by Robyr *et al.* [18]. In this study, minimized structures based on Polypack initial guesses were found to predict NMR spectra which were in better agreement with experimental data than structures generated using a modified

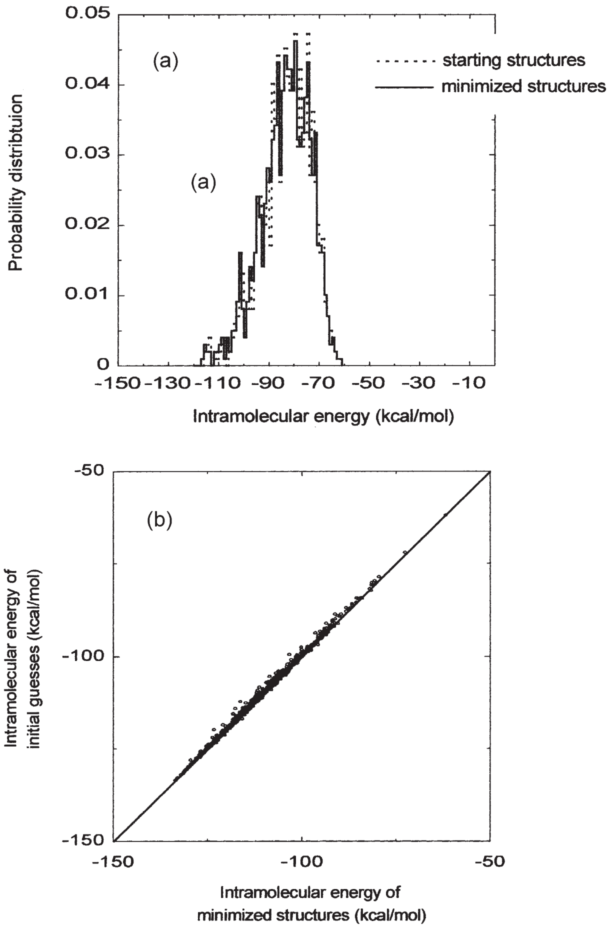


FIGURE 5 Distributions of intramolecular energy of starting and minimized MC structures (a) and parity plot (b).

Rotational Isomeric State scheme Initial Guess Generator [1,8]. Herein seems to lie one of the main strengths of Polypack: if minimized structures are to be used to investigate RIS-related properties [29], then Polypack is probably the initial guess generator of choice. The user should, however, keep in mind that other features of the structures, e.g. energy distributions, are probably less reliable or, in the case of torsional angle distributions, downright unphysical. There is of course no contradiction in the fact that RIS probabilities are well reproduced whereas the distribution of single bond states is off the mark: this is directly related to the departure of the second-order *a priori* probabilities from products of the first order probabilities in the RIS formalism [29]. Polypack manages to produce structures with the correct $p_{\zeta\eta}$ although the individual p_{ζ}, p_{η} may be incorrect.

The MC generator does a similarly good job at producing correct RIS probabilities for our pseudoatom model of polyethylene, while consistently reproducing all other expected conformational and energetic features. In this sense, such an MC generator is a very efficient and reliable tool. Unfortunately, this suite of MC algorithms currently available is not of general applicability. The acceptance rate of the suite of MC moves drops precipitously when applied to more complicated polymer structures. Even moderately bulky side groups dramatically reduce the acceptance rate, so that sampling is effectively halted. The strong reduction of the acceptance rate caused by moderately-sized side groups had already been observed for the continuous configurational bias algorithm (CCB) in a comparison involving united-atom and explicit hydrogen polyethylene [30] and more recently in polyproylene.

Site Pair Distribution Function

The differences in conformational properties between the two sets of minimized structures lead to marked deviations between the corresponding intermolecular and intramolecular radial pair distribution functions (Fig. 11). The high proportion of bonds in trans conformation in Polypack structures leads to an intramolecular pair distribution function with prominent peaks. These peaks are not characteristic of a well-equilibrated PE melt nor of a glassy PE structure.

The intermolecular pair distribution function for minimized Polypack structures does not present the characteristic damped oscillations between 0.5 and 1.5 nm. It displays a rapid rise at short distances followed by a flat plateau. The intermolecular mer distribution is therefore remarkably uniform in space, with little trace of successive neighbor shells. The very homogeneous spatial distribution of mers in Polypack structures has a profound effect on the predicted solubility of small molecules, as will be shown in the next Section.

Therefore, predicted static structure factors $Q(s)$ will also differ markedly. The suite of moves incorporated in the MC generator has been shown to produce melt structures the $Q(s)$ of which is in excellent agreement with experimental evidence [22]. Such a good agreement will be lacking in Polypack-generated structures.

Chemical Potential and Volume Available for Insertion of Small Penetrants

Finally, the ensembles of minimized structures were used as host matrices for the calculation of the chemical potential of a representative small solute (Xe) by Widom's insertion method [31]. The distribution of individual chemical potential of the Xe ghost particle (Fig. 12) is much narrower for minimized Polypack structures and its mean value lies appreciably higher than for the MC-generator ensemble. Thus, it seems that Polypack structures predict appreciably lower solubilities than MC-generator structures. This can also be interpreted as

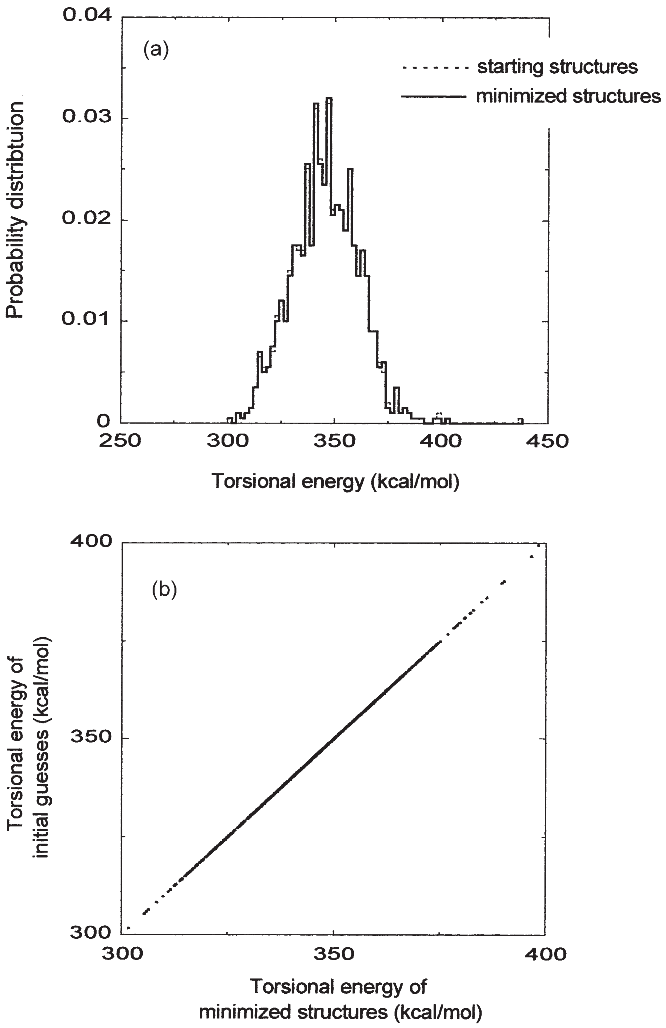


FIGURE 6 Distributions of torsional energy of starting and minimized MC structures (a) and parity plot (b).

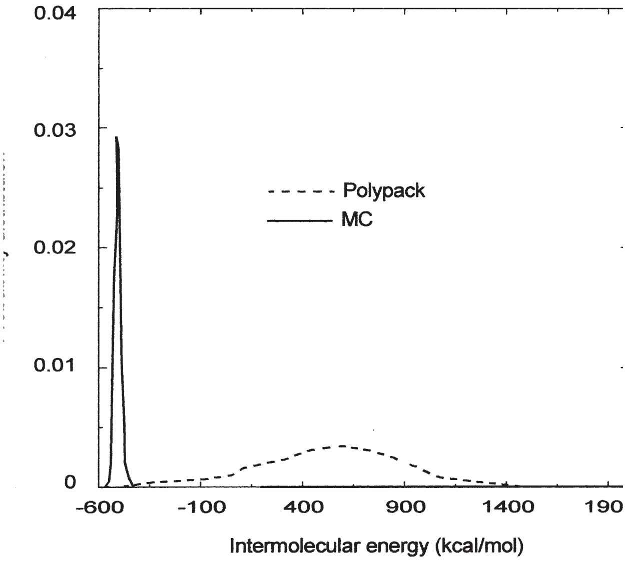


FIGURE 7 Distribution of intermolecular energy for MC-Initial Guess Generator and Polypack structures after minimization.

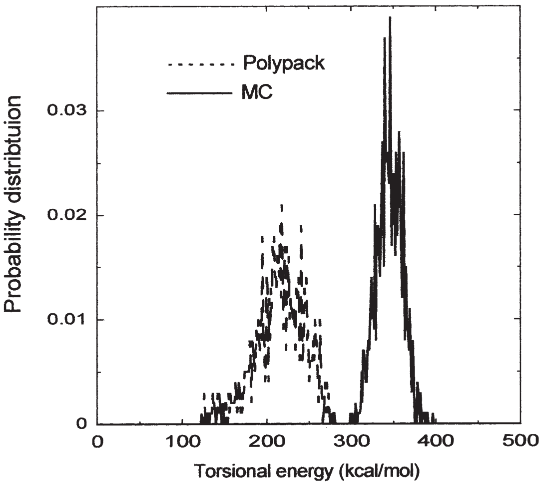


FIGURE 8 Distribution of torsional energy for MC-Initial Guess Generator and Polypack structures after minimization.

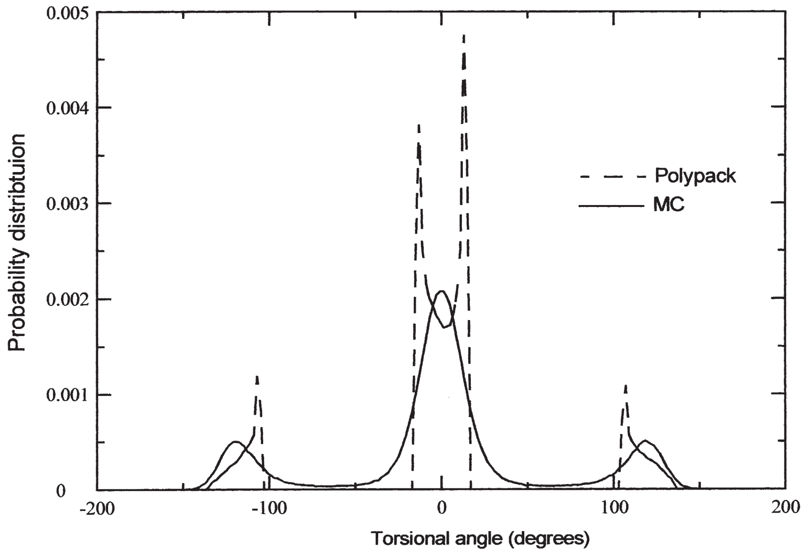


FIGURE 9 Torsional angle distributions for MC-Initial Guess Generator and Polypack structures after minimization.

a sign that the space available for insertion in Polypack structures is more homogeneously distributed than in MC-generator structures. The higher uniformity (narrower distribution) is a consequence of the geometric packing approach Polypack uses: it tends to fill the amorphous cell as homogeneously as possible while complying with the RIS probabilities. Since this constraint is far less severe than the one the MC generator has to fulfill (ergodic sampling from a well defined ensemble), it is only natural that mers are spatially more uniformly distributed in Polypack structures. Low-energy pockets available for insertion are therefore quite similar energetically throughout the structure. Small solutes in these pockets are on average surrounded by a larger number of mers than in MC structures. Their energy is therefore higher on average. This view is substantiated by

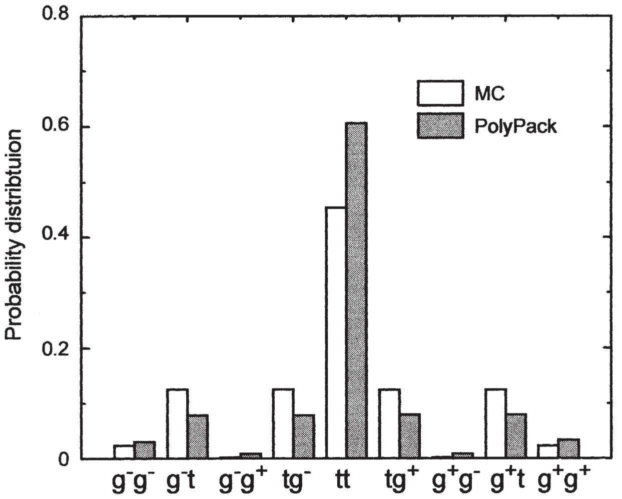


FIGURE 10 Bar chart of dyad probability for MC-Initial Guess Generator and Polypack structures after minimization.

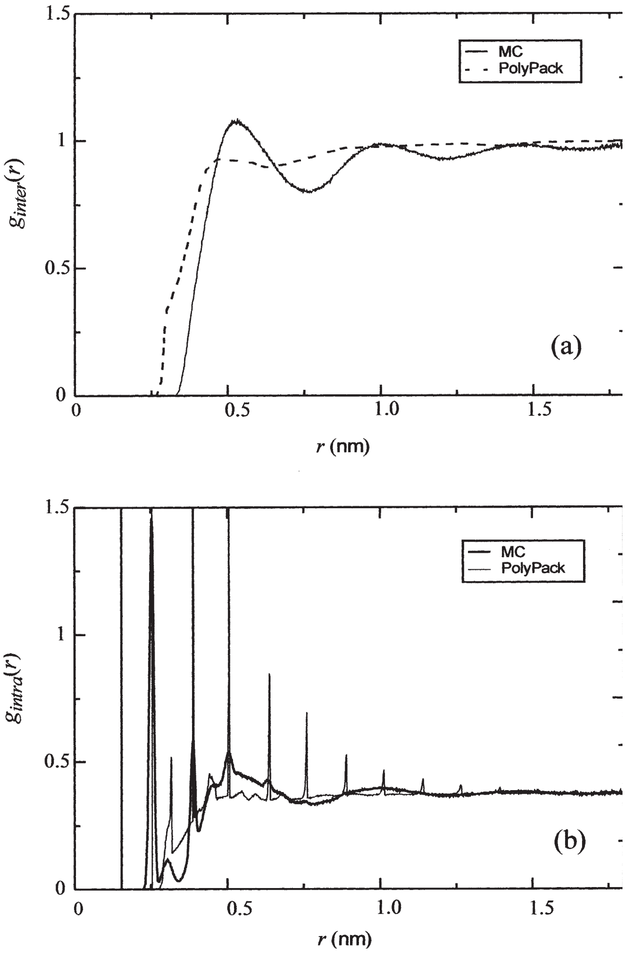


FIGURE 11 Intermolecular (a) and intramolecular (b) radial pair distribution functions for MC-Initial Guess Generator and Polypack structures after minimization.

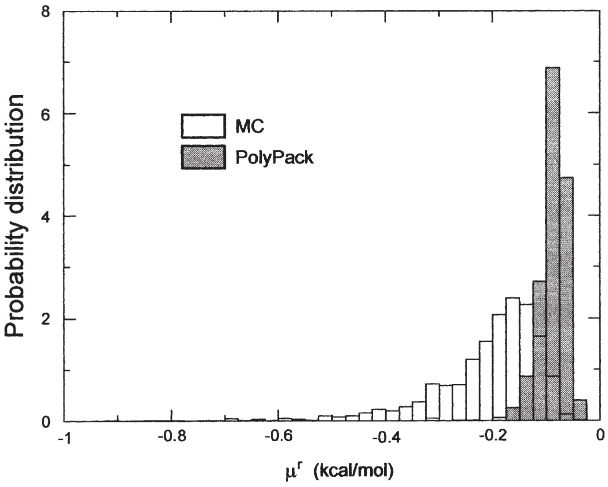


FIGURE 12 Distribution of residual chemical potential of xenon in minimized MC-Initial Guess Generator and Polypack structures.

the steeper rise of the intermolecular pair distribution function in Polypack structures. Just as chain mers are on average more closely surounded by other mers (Fig. 11a), so are inserted small solutes.

The energy of the space available for insertions in MC-generator structures is more widely distributed. The less homogeneous spatial distribution of mers leads to the appearance of regions where insertion is facilitated and which show up as the long tail of the residual chemical potential for MC-generator structures.

An interesting consequence of these results is that spatially very homogeneous minimized dense polymeric systems can be generated even if the initial structures are constrained to satisfy RIS probabilities and optimal geometrical packing (Polypack). More severe constraints (such as those imposed on efficient and ergodic MC algorithms) lead to spatially more heterogeneous structures in which small solutes can find a larger fraction of low-energy insertion sites. The unavoidable conclusion is that polymeric systems, or at the very least PE melts and probably glasses, pack in a fashion that does not have spatial homogeneity as the main criterion to be fulfilled [32]. A certain (but not arbitrary) degree of inhomogeneity seems to be a characteristic feature of realistic minimized structures.

CONCLUSIONS

The comparison of two highly efficient initial guess generators has shown them to be able to produce dense initial structures which can be successfully minimized in virtually all cases. The two generators represent opposite methodological extremes: while one (Polypack) is based on a series of packing and minimization algorithms with mostly geometrical or mathematical footing, the other one (MC-generator) is largely based on physical or statistical mechanical analogies.

As a consequence, structures minimized using one or the other differ very significantly in most of the properties investigated here (which are quite representative of the uses minimized structures are put to). The best agreement between both and the best performance for Polypack is found for RIS-related properties, since the correctness of RIS probabilities has one of the highest built-in priorities in Polypack. For all other properties, the MC generator produces initial guesses that, upon minimization, have properties in consistently better agreement with experimental evidence.

Thanks to the large restricted ensembles generated, a precise probability distribution of total energy (Fig. 1b) of minimized structures for PE has been obtained, to our knowledge, for the first time. For the physically-founded MC generator, this distribution is (within stringent statistical limits) a shifted but otherwise undistorted replica of the corresponding distribution in the melt.

The MC generator is however limited to very simple molecular architectures. Although improvements are to be expected, for the time being even moderately bulky side groups, branching, etc. reduce the sampling efficiency of the method so drastically as to render it useless. On the other hand Polypack is able to produce dense structures for polymers of virtually any molecular architecture but at the expense of rather unrealistic predictions in a number of important properties. MM would greatly benefit from an algorithm which combines the best of the two generators investigated here.

Acknowledgements

The authors are indebted to Prof D.N. Theodorou for very kindly making several of their MC codes available. M.L. gratefully acknowledges partial support by CICYT grant MAT 1999-0972.

References

[1] Theodorou, D.N. and Suter, U.W. (1985) "Detailed molecular structure of a vinyl polymer glass", *Macromolecules* **18**, 1467.
[2] Rutledge, G.C., Suter, U.W. and Papaspyrides, C.D. (1991) "Analysis of structure and polymorphism in poly (*p*-phenyleneterephthalamide) through correlation of simulation and experiment", *Macromolecules* **24**, 1934.
[3] Allen, M.P. and Tildesley, D.J. (1987) *Computer Simulation of Liquids* (Oxford University Press, Oxford).
[4] Evans, D.J. and Morriss, G.P. (1990) *Statistical Mechanics of Nonequilibrium Liquids* (Academic Press, New York).
[5] Müller, M. (1999) *The Structure of Dense Polymer Systems; Geometry, Algorithms, Software* PhD Thesis 13096, ETH Zürich.

[6] Garey, M.R. and Johnson, D.S. (1979) *A guide to the Theory of NP-Completeness* (Freeman, San Francisco).

[7] Müller, M., Nievergelt, J., Santos, S. and Suter, U.W. (2001) "A novel geometric embedding algorithm for efficiently generating dense polymer structures", *J. Chem. Phys.* **114**, 9764.

[8] Rapold, R.F. and Suter, U.W. (1994) "Conformational Characteristics of Polystyrene", *Macromol. Theory Simul.* **3**, 1.

[9] Meirovitch, H. (1983) "Computer simulation of self-avoiding walks: testing the scanning method", *J. Chem. Phys.* **79**, 502.

[10] Brostow, W. and Kubat, J. (1993) "Molecular-dynamics simulation of stress relaxation on a triangular lattice", *Phys. Rev. B* **47**, 7659.

[11] Gusev, A.A., Zehnder, M.M. and Suter, U.W. (1994) "Elasticity of solid polymers as a result of thermal motions", *Macromolecules* **27**, 615.

[12] McKechnie, J.I., Brown, D. and Clarke, J.H.R. (1992) "Methods of generating dense relaxed amorphous pollymer samples for use in dynamics simulations", *Macromolecules* **25**, 1562.

[13] Brown, D., Clarke, J.H.R., Okuda, M. and Yamazaki, T. (1994) "The preparation of polymer melt samples for computer simulation studies", *J. Chem. Phys.* **100**, 6011.

[14] Kröger, M. (1999) "Efficient hybrid algorithm for the dynamic creation of womlike chains in solutions, brushes, melts and glasses", *Comput. Phys. Commun.* **118**, 278.

[15] Kotelyanskii, M., Wagner, N.J. and Paulaitis, M.E. (1996) "Building large amorphous polymer structures: atomistic simulation of glassy polystyrene", *Macromolecules* **29**, 8497.

[16] Avelar, S. and Müller, M. (2000) "Generating Topologically Correct Schematic Maps". In Proceedings of 9th International Symposium on Spatial Data Handling, 4a.28–4a.35, Peking.

[17] Santos, S., Suter, U.W., Müller, M. and Nievergelt, J. (2001) "A novel parallel rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses", *J. Chem. Phys.* **114**, 9772.

[18] Robyr, P., Müller, M. and Suter, U.W. (1999) "Atomistic simulations of glassy polystyrenes with realistic chain conformations", *Macromolecules* **32**, 8681.

[19] Rosenbluth, M.N. and Rosenbluth, A.W. (1953) "Equations of state calculations by fast computing machines", *J. Chem. Phys.* **21**, 1087.

[20] Siepmann, J.I. and Frenkel, D. (1992) "Configurational bias Monte Carlo: a new sampling scheme for flexible chains", *Mol. Phys.* **75**, 59.

[21] de Pablo, J.J., Laso, M. and Suter, U.W. (1992) "Simulation of polyethylene above and below the melting point", *J. Chem. Phys.* **96**, 2395.

[22] Mavrantzas, V.G., Boone, T.D., Zervopoulou, E. and Theodorou, D.N. (1999) "End-Bridging Monte Carlo: an ultrafast algorithm for atomistic simulation of condensed phases of long polymer chains", *Macromolecules* **32**, 5072.

[23] Pant, P.V.K. and Theodorou, D.N. (1995) "Variable connectivity method for the atomistic Monte Carlo simulation of polydisperse polymer melts", *Macromolecules* **28**, 7224.

[24] Van der Ploeg, P. and Berendsen, H.J.C. (1982) "Molecular dynamics simulation of a bilayer membrane", *J. Chem. Phys.* **76**, 3271.

[25] Ryckaert, J.P. and Bellemans, A. (1975) "Molecular dynamics of liquid *n*-butane near its boiling point", *Chem. Phys. Lett.* **30**, 123.

[26] Coleman, T.F. (1984) "Large sparse numerical optimization", *Lecture Notes in Computer Science* (Springer-Verlag, New York) **Vol. 165**.

[27] Dennis, J.E. and Moré, J.J. (1977) "A quasi-Newton method for function minimization", *SIAM Rev.* **19**, 46.

[28] Hillstrom, K. (1976) "Nonlinear Optimization Routines in AMDLIB > ". Technical Memorandum No. 297, Argonne National Laboratory, Applied Mathematics Division (subroutine GQBFGS in AMDLIB, Argonne, III).

[29] Flory, P.J. (1969) *Statistical Mechanics of Chain Molecules* (Interscience, New York).

[30] Leontidis, E., De Pablo, J.J., Laso, M. and Suter, U.W. (1994) "A critical evaluation of novel algorithms for the off-lattice Monte Carlo simulation of condensed polymer phases", *Adv. Polym. Sci.* **116**, 283, (See also "Atomistic Modelling of Physical Properties", Chapter VIII, Springer-Verlag).

[31] Widom, B. (1963) "Some topics in the theory of fluids", *J. Chem. Phys.* **39**, 2808.

[32] Arizzi, S., Mott, P.H. and Suter, U.W. (1992) "Space available to small diffusants in polymeric glasses: analysis of unoccupied space and its connectivity", *J. Polym. Sci., Polym. Phys. Ed.* **30**, 415.