Complete free energy landscape and statistical thermodynamics of single poly(ethylene glycol) molecules

Douglas Staple, Felix Hanke\(^1\) and Hans Jürgen Kreuzer\(^2\)
Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, B3H 3J5, Canada
E-mail: dstaple@dal.ca, hanke@fhi-berlin.mpg.de and h.j.kreuzer@dal.ca

New Journal of Physics 9 (2007) 68
Received 10 November 2006
Published 21 March 2007
Online at http://www.njp.org/
doi:10.1088/1367-2630/9/3/068

Abstract. In this paper, we revisit the statistical mechanics of poly(ethylene glycol) with the intention to check ad hoc approximations made in previous works, namely to keep all C–O bonds along the chain in the trans configuration. The theory developed here includes all the bonds on an equal footing. Starting with density functional theory calculations for the ethylene glycol monomer, dimer and trimer one can determine all the energetic and geometric parameters needed for a rotationally isomeric state model with interactions between nearest neighbour monomers included. The properties of short molecules are presented in detail, including the force-extension curves, end-to-end distributions, and also a discussion of the local helix properties. Finally, we use the transfer matrix method for long chains of 200 ethylene glycol monomers, i.e. 600 bonds. The agreement with the experimental force-extension curve is excellent.

\(^1\) Present address: Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany.
\(^2\) Author to whom any correspondence should be addressed.
1. Introduction

Poly(ethylene glycol) (PEG) is a simple polymer, yet complex enough to show a number of interesting features revealed, e.g. in atomic force microscopy (AFM) experiments. It consists of a sequence of ethylene glycol subunits (-O-CH$_2$-CH$_2$-, denoted by (EG)), with either methoxy or hydroxy terminal groups. Solvated in hexadecane, its force-extension curve shows the simple features of a freely joined chain (FJC) response, i.e. it is dominated solely by entropic effects, except at the highest extensions, where spring-like behaviour is observed [1, 2]. On the other hand, when water is used as the solvent, a plateau-like hump appears that is reminiscent of two-state behaviour usually reserved for more complex molecules such as polysaccharides.

The first models concerning the statistics of such simple polymers in general were aimed at simply accounting for a one-dimensional connected structure as well as the volume exclusion effect, i.e. making sure that atoms do not overlap. To some general degree, the mean-field calculations of Edwards [3] or the renormalization group approach by de Gennes [4] are applicable to all macromolecules in solution. The phenomenological Flory–Huggings model [5] is based on calculating the free energy of a molecule in solution by accounting for the free energy of mixing as well as a molecule–solvent interaction, generally specified by a single parameter $\chi$.

An extension of the latter theory using a continuous mean-field theory [6, 7] as well as a lattice model [8, 9] made it possible to show that the PEG phase diagram can be explained by assuming clusters of molecules, held together by hydrogen bonds. This property has been corroborated by Smith and co-workers [10]–[14], who use exact calculations on dimethoxyethane ((EG)$_2$) for the development of molecular force fields, which in return are used in molecular dynamics calculations.

In this work, we outline a theory that starts from first principles calculations and systematically introduces approximations to approach larger scales in the modelling of single PEG molecules. Thus, an explanation of the macroscopic response and properties of a long polymer chain is ultimately given in terms of the structural and energetic properties of the (EG) monomer subunits [2]. The starting point for an appropriate statistical theory should be the energy spectrum and the density of states, which can be obtained from quantum mechanical electronic structure calculations. The latter can only be done for small systems because the computational effort increases at least as the fourth power of the number of electrons in the system. This paper
aims to develop a strategy that leads to an exact statistical theory of a single PEG molecule, potentially interacting with a solvent. To treat such a long polymer molecules the following approach has been developed.

In a first step one calculates the electronic properties of a short polymer chain from first principles by, for instance, employing density functional theory (DFT) or Hartree–Fock theory complemented by Möller–Plesset perturbation theory. The result of such calculations is the energy spectrum and the density of states (the number of independent wavefunctions or conformers at a given energy) for a short chain [15]. This provides all the input required for the statistical mechanics of short chain molecules.

Although experiments are generally done with long polymers, obtaining the force-extension curve for short chains is still a useful exercise. It provides a quantitative description of the late stages of the elastic response of long chains, where the molecule is more or less fully stretched already so that entropic contributions are less important. The total response of a long chain is more or less additive over the individual monomers, provided the first principles calculations are done as accurately as possible for all the conformers of the (small) molecule.

First principles calculations were reported for \((\text{EG})_n\) with \(n = 2, 3, 4\) and its interaction with water, using DFT with nonlocal gradient corrections [15]. At that time, a tractable number of conformers was selected, chosen to be representative of those states accessible at room temperature. We recall that the conformers of a macromolecule form local minima on its total energy surface. As such, a conformer is a static concept and its relevance for a system at nonzero temperatures hinges on timescales: if the activation barrier for escape from the local minimum is so high that the time spent by the molecule in this local minimum is much larger than the time to escape from it, we can make definitive measurements on this state and assign it the status of a conformer. It was argued in the previous study that most low energy conformers have the C–O bonds in the \textit{trans} configuration and the C–C bonds in \textit{trans} and clockwise and anticlockwise gauche configurations, resulting in three states of the EG subunit, \((t|g^+|t), (t|g^-|t)\) and \((t|t|t)\), to be placed on three positions along the chain, say \((\text{EG})_3\), for a total of \(3^3 = 27\) combinations or conformers.

Since the first principles approach just outlined can only be done for short chains of a few monomers, an extension of the theory to deal with long chains must invoke some approximations, by necessity. The latter are based on selecting from the large number of quantum states of a polymer, or conformers, those that are energetically accessible at the temperature of interest and dynamically relevant. This suggests multistate interacting chain models (ICMs) of which the three-state rotationally isomeric state (RIS) model is an early version. Such models have a number of parameters such as self-energies of the various states, interaction energies between nearest and next-nearest neighbours along the chain and geometric parameters like bond lengths, bond angles and dihedral angles. The result is an ICM with all parameters determined from first principles quantum mechanical calculations, valid for polymers of arbitrary length [16]. In recent years, we have developed a Green function or transfer matrix (TM) method which allows one to solve the complete and exact statistical mechanics of ICMs in both the Gibbs and Helmholtz ensembles [16]–[19].

The restriction to calculating only a few conformers for short chains was ultimately proved to be a good (and economical) strategy in that the multistate RIS model gave perfect agreement with experiment. Not only that, this theory also explains the new two-state feature when stretching is done in an aqueous environment [1, 2, 16]. Still, the question remains whether a more complete
theory will give further information. This paper is dedicated to this endeavour. Our strategy is as follows.

1. Using DFT, we calculate the geometry and energies of all $3^3 = 27$ conformers of $(\text{EG})_1$, $3^6 = 729$ conformers of $(\text{EG})_2$ and also for a random selection of 493 (out of $3^9 = 19683$) conformers for $(\text{EG})_3$. Within the framework of quantum mechanics, a conformer is a local energy minimum in the configurational space spanned by the coordinates of the nuclear coordinates of all the atoms in these molecules. The results of such calculations are the respective densities of states, $D_N(E, L)\,dE\,dL$, i.e. the number of conformers in intervals $dEdL$ with energy $E$ and length $L$. From this we can obtain the Helmholtz and Gibbs partition functions

$$Z_H(T, L, N) = \int dEDN(E, L) \exp \left[ -\beta E \right],$$

$$Z_G(T, f, N) = \int dE dL D_N(E, L) \exp \left[ -\beta (E + fL) \right],$$

(1)

and free energies

$$F(T, L, N) = -k_B T \ln Z_H,$$

$$G(T, f, N) = -k_B T \ln Z_G,$$

from which the mechanical equation of state, i.e. force-extension curve, and the thermal equation of state can be calculated by simple derivatives. The reason why we look concurrently at the Helmholtz and Gibbs ensembles is because the former lends itself to the calculation of force fluctuations and the latter to length fluctuations, and both are accessible in AFM experiments.

2. With the density of states we can define the free energy landscape

$$V_N(T, L) = -k_B T \ln D_N(T, L),$$

(2)

which becomes apparent when we rewrite the partition functions as

$$Z_G(T, f, N) = \int dE dL \exp \left[ -\beta V_N(T, L) \right] \exp \left[ -(\beta E + fL) \right],$$

(3)

and similarly for the Helmholtz partition function.

3. Next we extract energetic and geometric parameters for the three-state RIS model including interactions between adjacent monomers. From this we recalculate the force-extension curve and find, not surprisingly, excellent agreement with the exact calculation.

4. Going beyond the shortest chains we use the RIS model to calculate the force-extension curve for $(\text{EG})_7$ for which we also use the TM method (with the same geometric and energetic parameters), again finding excellent agreement.
The TM method can be used to calculate the partition function for long chains which we do for $N = 200$. The resulting force-extension curve we then compare with that obtained in previous work based on using only 10 non-redundant conformers of $(\text{EG})_3$ as input. The agreement obtained is the ultimate justification of that simpler model as agreement with experiment is also guaranteed.

The paper is structured as follows: in the following section, we give the necessary details of the DFT calculations and results for short chains. We then construct the Hamiltonian for the RIS model, outline a method for the exact evaluation of its partition function and give details of the transfer TM. In section 3, we present force-extension curves for medium-length and long chains calculated in these frameworks. The last section contains a summary, conclusions and an outlook to future work. Some numerical details are given in appendices.

2. Theoretical approach

2.1. First principles calculations

Our calculations of all the conformers of $(\text{EG})_n$ with $n = 1, 2$ and 3 employ DFT with gradient corrections and have been done with the Gaussian 03 suite of programs [20]. The exchange-correlation potential used is BP86 and the basis set is 6-311++G**. A justification of this choice and a thorough comparison with earlier work has been given elsewhere [15].

A typical calculation of the structure of a conformer starts with standard C–C and C–O bond lengths (1.52 and 1.42 Å), C–C–O and C–O–C bond angles (109$^\circ$ and 112$^\circ$), O–C–C–O and C–C–O–C dihedral angles ($\pm 74^\circ$ or $180^\circ$), obtained from the calculation of the helical conformer. We then adjust all these parameters and allow rotation around the C–C and/or C–O bonds to find a local minimum in the total electronic energy. That this is possible for a conformer, as defined above, is not surprising when one considers that in the 84-dimensional space of the coordinates of all 30 nuclei of $(\text{EG})_3$ ($3 \times 30 - 6$ for the centre of mass translation and rotation of the molecule) these minima are far from each other with activation barriers in between. Activation barriers do not enter the calculation of equilibrium properties but of course are important to estimate the time scale of the approach of the system to equilibrium as we indicate above. We also point out that, in the various conformers, bond lengths, bond angles, and dihedral angles are all slightly different from their standard values.

Symmetry considerations reduce the number of conformers to 10 for $(\text{EG})_1$, 196 conformers of $(\text{EG})_2$ and 5002 for $(\text{EG})_3$. Convergence of the self-consistent DFT iterations can be reached for nine conformers of the monomer, only 135 for the dimer and 493 of 1040 randomly selected conformers of the trimer. Lack of convergence is interpreted as evidence for volume exclusion, because it occurs when a conformer contains the sequence $g+g-g+$ so that the monomer doubles up on itself. Moreover, all conformers that the RIS model predicts to be excluded did not converge either.

For the construction of the RIS model we need to extract average bond lengths, bond angles and dihedral angles from all conformers, which can be done by straightforward geometric averaging. However, this procedure would overestimate the contributions from the highly deformed, high-energy conformers. We therefore also use a weighted averaging with a Boltzmann factor involving the energy of a given conformer.
We use a variation of the RIS formalism to describe the thermodynamics of a long chain. A possible conformer is described by a set of vectors

\[ \vec{n} \in \{ \vec{n}_1, \vec{n}_2, \ldots, \vec{n}_N \} \]

describing the rotational states \( s_i \in \{ g^+, t, g^- \} \) of each atomic bond (of which there are three for the EG monomer) as

\[ \vec{n}_i = (s_i, 0, -c s_i, c - c s_i, -c - c s_i) \] (4)

Table 1 presents the resulting bond lengths, bond angles and the energies relative to the all-trans conformer, for the monomer. For the dimer and trimer we give plots of all converged conformers in the energy versus length plane in figures 1 and 2. These plots represent the density of states \( D(E, L) \) which can be turned into a continuous function by appropriate binning of the discrete dataset. Important from the viewpoint of statistical mechanics is the band of conformers below about 50 meV or \( 2k_B T \) at room temperature, well separated from the higher energy conformers.

### Table 1. Geometry parameters obtained by averaging over (EG)\(_1\), (EG)\(_2\), and (EG)\(_3\): bond distances \( b \), bond angles \( \gamma \), and dihedral angles \( \phi \). The top number in each field is the Boltzmann-weighed average (used in analysis), while the bottom value in brackets denotes the arithmetic average.

|        | \( b_{cc} \) [Å] | \( b_{co} \) [Å] | \( \gamma_{cc} \) [°] | \( \gamma_{oc} \) [°] | \( \phi_{cc} \) [°] |
|--------|------------------|------------------|---------------------|---------------------|-------------------|
| (EG)\(_1\) | 1.523 ± 0.005    | 1.426 ± 0.003    | 110.7 ± 2.6         | 112.6 ± 0.9         | 81.8 ± 4.2        |
| (1.531 ± 0.007) | (1.426 ± 0.003) | (112.2 ± 2.7) | (113.2 ± 0.8) | (78.5 ± 6.3)        |
| (EG)\(_2\) | 1.523 ± 0.005    | 1.427 ± 0.004    | 110.5 ± 2.6         | 112.9 ± 1.0         | 82.3 ± 5.3        |
| (1.530 ± 0.007) | (1.428 ± 0.004) | (111.7 ± 2.8) | (113.5 ± 1.0) | (80.2 ± 7.5)        |
| (EG)\(_3\) | 1.523 ± 0.005    | 1.427 ± 0.004    | 110.5 ± 2.5         | 113.0 ± 1.1         | 82.6 ± 5.5        |
| (1.530 ± 0.007) | (1.428 ± 0.004) | (111.5 ± 2.8) | (113.6 ± 1.0) | (80.7 ± 7.1)        |

2.2. Exact calculations in the RIS formalism

We use a variation of the RIS formalism to describe the thermodynamics of a long chain. A complete description of a molecule with a given chain length necessitates a summation over all possible conformers. Each conformer with \( N \) monomers is uniquely described by a set of vectors \( \vec{n}_1, \vec{n}_2, \ldots, \vec{n}_N \) describing the rotational states \( s_i \in \{ g^+, t, g^- \} \) of each atomic bond (of which there are three for the EG monomer) as

\[ \vec{n}_i = (s_i, 0, -c s_i, c - c s_i, c - c s_i) \] (4)
Figure 1. (EG)$_2$: (top panel) comparison between the RIS prediction for energy and length of a conformer (grey $\times$) and the converged ab initio results (coloured +). Thickness and colour are used to denote the degeneracy, as shown in the legend. The red circles mark all conformers of the form $t \ast t - t \ast t$. (Bottom panel) force-extension curves for three different sets of ab initio conformers, as described in the legend.

A one-dimensional lattice gas Hamiltonian can be used to fully determine the energetics of a chain of $N$ monomers, see [16] for the details with the $(t|\ast|t)$-approximation.

$$
\mathcal{H}(\vec{\mu}) = E_{\text{term}} + \sum_{i=1}^{N} E_{\text{self}}(s_{i, \text{O} - \text{C}}s_{i, \text{C} - \text{C}}s_{i, \text{C} - \text{O}})
$$

$$
+ \sum_{i=2}^{N} V_{\text{NN}}(s_{i-1, \text{O} - \text{C}}s_{i-1, \text{C} - \text{C}}s_{i-1, \text{C} - \text{O}}; s_{i, \text{O} - \text{C}}s_{i, \text{C} - \text{C}}s_{i, \text{C} - \text{O}}).
$$

(5)

The nearest-neighbour potential $V_{\text{NN}}$ refers to the interaction of complete monomers and thus contains interactions including up to five neighbouring bonds along the backbone. This analysis allows for an exact representation of the energetics of (EG)$_2$. As for the conformers of (EG)$_1$, there are 27 self energies, $E_{\text{self}}$, 10 of which are unique. As the $(g+|g-|g+)$ and $(g-|g+|g-)$ conformers are not allowed due to volume exclusion, their corresponding self-energy has been set to infinity. Similarly, there are $3^6 = 729$ interactions in the nearest-neighbour potentials,
Figure 2. \((\text{EG})_3\): (top panel) RIS prediction for energy and length of all conformers compared to the available \textit{ab initio} results, similar to figure 1; (bottom panel) force-extension curves for three different sets of RIS-derived conformers.

196 of which are unique. Sixty-one of these interaction energies are also set to infinity due to volume exclusion effects. The calculation of these parameters is described in appendix A. Our Hamiltonian (5) corresponds to a first step from pure quantum mechanics toward a statistical model and still contains a maximally possible amount of microscopic information.

The other ingredient for the RIS model is the geometry, which is based on the averages given in table 1. Given the sequence of dihedral states \(\vec{n}\), all bond vectors of a given conformer can easily be calculated with the successive application of rotation matrices. Defining the first bond to be along the \(\hat{z}\)-axis of our coordinate system, the \(n\)th bond vector \(\vec{b}_n\) is in the direction

\[
\vec{b}_n = \prod_{i=1}^{n} R_z(\varphi_i)R_y(\gamma_i)\hat{z},
\]

where the operator \(R_z(\varphi)\) rotates the coordinate system by an angle \(\varphi\) about the \(\hat{z}\)-axis and \(R_y(\gamma)\) rotates it by \(\gamma\) about the \(\hat{y}\)-axis. From equation (6) one can calculate all required geometric properties of one conformer. As an example we give the end-to-end vector \(\vec{r}_N\) for a conformer
consisting of \( N_b = 3N \) bonds (3 for each EG monomer),

\[
\mathbf{r}_N = \sum_{n=0}^{N_b} b_n \hat{b}_n = \sum_{n=0}^{N_b} \left[ \prod_{i=1}^{n} \mathcal{R}_z(\varphi_i) \mathcal{R}_y(\gamma_i) \right] b_n \hat{z}.
\]

The location of the hydrogen atoms are obtained in a similar fashion.

Since all of the atomic positions are known in this approach, it is a simple task to check whether or not a given conformer of \((\text{EG})_n\) is allowed. A conformer is rejected due to excluded volume interactions if any two atoms in the molecule overlap, as defined by their respective van der Waals radii [21, 22]. All exact RIS results presented in this work are shown twice, contrasting calculations that account for excluded volume and calculations that ‘neglect’ these effects.

Next comes the evaluation of the partition functions, equations (1), from which equilibrium thermodynamics follows straightforwardly. The computational effort for this calculation can be reduced by reducing the number of redundant conformers, which can be done with simple operations on the vector \( \vec{n} \) defined in equation (4). For our purposes, a conformer is called redundant when its sequence vector \( \vec{n} \) can be obtained from that of second conformer \( \vec{n}' \) either by inverting the sequence, by exchanging the gauche-states \( g^+ \leftrightarrow g^- \), or by using a combination of both. The resulting possible degeneracies for unique conformers are \( g_{\vec{n}} = 1, 2, \) or 4.

According to equations (1), there is a choice between Helmholtz and Gibbs ensembles for a fixed number of bonds. Given that the overall number of conformers for \((\text{EG})_n\) is still of the order \( O(3^{3n}) \), it is advisable to do all initial numerical analysis in the Helmholtz ensemble, where a complete description of the system can be obtained with each conformer taken into account exactly once. Due to the sheer number of available conformers, it is numerically sensible to bin the results of similar lengths over some small interval \( [L - \delta L, L + \delta L] \). The Helmholtz partition function now reads

\[
Z_H(L) = \sum_{\vec{n}} \delta(L - L_{\vec{n}}) \exp(-\beta E_{\vec{n}}) g_{\vec{n}} \simeq \int_{L - \delta L}^{L + \delta L} dL' \sum_{\vec{n}} \delta(L - L_{\vec{n}}) \exp(-\beta E_{\vec{n}}) g_{\vec{n}}.
\]

Using the same formalism, it is possible to write down the end-to-end distance probability distribution

\[
P_e(L) = \frac{\int_{L - \delta L}^{L + \delta L} dL' \sum_{\vec{n}} \delta(L - L_{\vec{n}}) \exp(-\beta E_{\vec{n}})}{\sum_{\vec{n}} \exp(-\beta E_{\vec{n}})}.
\]

which describes the probability of finding the length of a molecule to be equal to \( L \) and is used to calculate the mean square length of the molecule, \( \langle R^2 \rangle = \int dL L^2 P_e(L) \).

2.3. TM method

The all-inclusive RIS calculations described above, are only possible for short chains with a limited number of conformers; our longest chain was \((\text{EG})_7\). An additional approximation enables one to use the TM method for which the limit on the number of (EG)-units is at about 200 [16]. This method was developed primarily to calculate force-extension curves and various other properties of single molecules [16, 18], but it has been shown that the TM method is a powerful tool to use for modelling single molecules in external potentials or even in a self-assembled monolayer [19, 23, 24].

New Journal of Physics 9 (2007) 68 (http://www.njp.org/)
In the TM method, the state of the \( n \)th bond in a molecule is described via a Green function \( G_n(\Gamma) \) [3, 5]. The Green function for the following monomer can be obtained by defining a transfer operator \( T(\Gamma, \Gamma') \) and integrating the Chapman–Kolmogorov equation, \( G_{n+1}(\Gamma) = \int d\Gamma' T(\Gamma, \Gamma') G_n(\Gamma') \). A given bond in a molecule is completely described by specifying a set of coordinates \( \Gamma \), whose choice depends on the problem at hand and the underlying polymer model used.

In the particular case of the RIS model, \( \Gamma \) must include the spatial location and orientation of the \( n \)th bond as well as its rotational state \( (g−, t, g+) \). The spatial symmetry of the problem allows the use of a single spatial coordinate, which is chosen to be the \( z \)-coordinate in our coordinate system for computational reasons. Starting with the orientation of the \( n \)th bond, specified by \( \theta_n \) and \( \Delta_1 z_n \) in the usual cylindrical polar coordinates, the orientation of bond \( b_{n+1} \) is calculated with the same rotation matrices used in the RIS geometry, equation (6)

\[
b_{n+1}(s_{n+1}) = (b_{n+1}/b_n)R_z(\varphi_{n+1})(\gamma_{n+1})(b_n \cos \theta_n, b_n \sin \theta, \Delta_1 z_n)^T.
\] (10)

The dependence on the rotational state \( s_{n+1} \) enters through the choice of the dihedral angle \( \phi_{n+1} \) as per table 1. In the numerical implementation it proves advantageous to do the actual TM calculation by using \( z \)-projections of bonds \( n \) and \( n + 1 \) as independent variables, rather than \( \Delta_1 z_n \) and the angle \( \theta_n \). This effectively saves one (out of four) integrations in the calculation of the Green function for each bond. The independent variables for the transfer operator are \( \Gamma_i = \{z_i, \Delta z_i, \Delta z_{i-1}, s_i\} \) and the operator itself has the form

\[
T(\Gamma, \Gamma') = \sum_i \int d\zeta d\Delta z d\Delta z' \delta(z - \zeta - [b_{n+1}]_{\Delta z}) \delta(\Delta z' - [b_{n+1}]_{\Delta z}) \times \delta(\Delta z_{old} - \Delta z') \exp \left( -\frac{E_{self}(s_{n+1}) + V_N(s_n, s_{n+1})}{k_B T} \right).
\] (11)

The specification of proper initial conditions for this procedure involves an artificial bond with index zero that precedes the first real bond. This bond has to be taken into account in such a way that the Green function of the first real bond contains all possible orientations and rotational states with equal probability. Finally we should note that an appropriate discretization of the variables is necessary to actually calculate properties of a large molecule. The Green function then becomes a vector containing some \( 10^8 \) elements and the transfer operator (11) takes on the form of a square matrix of this dimension. Each integration of the Chapman–Kolmogorov equation then corresponds to a matrix multiplication.

The evaluation of statistical equilibrium properties, such as the force-extension curve from the formalism outlined in this section, follows from the fact that the Green function for the last bond in effect corresponds to the Helmholtz partition function when it is integrated over all orientation variables [16],

\[
Z_H(z_N) = \sum_{s_N} \int d\Delta z_N d\Delta z_{N-1} G_N(z_N, \Delta z_N, \Delta z_{N-1}, s_N).
\] (12)

The TM also requires microscopic geometry and energy information for the proper evaluation of equation (11). The bond lengths, bond angles, and dihedral angles can be taken directly from the DFT calculation as was done in the RIS model. However, the Hamiltonian
underlying this model is somewhat different from the RIS Hamiltonian (5). In the TM case, everything is based on single bonds and their first neighbour, rather than on monomers. Thus the Hamiltonian $H_{TM}$ has the form

$$H_{TM} = E_{term} + E_{self}(s_1) + \sum_{i=2}^{3N} [E_{self}(s_i) + V_{NN}(s_i, s_{i-1})].$$  \hspace{1cm} (13)

The interactions $V_{NN}$ only take into account the immediately neighbouring bonds, in contrast to the whole monomer that was part of the nearest neighbour interaction $V_{NN}$ in the RIS case. The nature of the energies $E_{self}$ and $V_{NN}$ does not allow for a direct calculation from DFT data. However, it can be obtained as a degeneracy-weighed best fit to all 493 available (EG)$_3$ conformers using a Hamiltonian (5) with interactions up to nearest-neighbour dihedrals. This procedure still accounts for the exact geometry up to four bonds along the chain. The results are presented in appendix A, along with the RIS energy parameters.

Although the TM in its present implementation accounts fully for interactions up to four neighbouring bonds along the chain, it does not include volume exclusion between more distant atoms as is done in the exact calculations for short chains. This neglect, however, should be of little consequence when the TM method is used to calculate the stretching response for very long molecule: when a force is applied, the relevant extended configurations are less susceptible to the excluded volume interactions. This will be shown explicitly with examples below.

3. Results

In this section, we will present the results for the thermodynamic properties (such as force-extension curves) of short and long chain molecules based on the various analytic tools that we developed in earlier sections. We begin with short chains. In the bottom panel of figure 1 we show the force extension curves for (EG)$_2$ calculated at various levels of approximation. The solid line results from the inclusion of all conformers obtained from the converged DFT calculations. For the dash-dotted line we only included those conformers with energies less than 50 meV or $2k_B T$ (at room temperature). As this excludes some of the high-energy short conformers the curve goes into the compressive mode a little earlier. Lastly, the dashed line results when we only account for the conformers where the CO bonds are kept in the trans configuration, i.e. conformers of the form $(t \ast t - t \ast t)$ where $\ast$ stands for trans or the two gauche configurations of the C–C bonds. This excludes almost all the short conformers and the compressional regime starts very early. The latter approximation was used in our previous work [16]. This clearly shows that for (EG)$_2$ this is not a good approximation. Note however, that in the regime of high stretching all curves agree because there only the longer trans configurations are important. The same conclusions are reached for (EG)$_3$ as shown in the bottom panel of figure 2.

Moving on to longer chains we look in detail at (EG)$_7$. This is the longest chain for which we can, in a reasonable time, do exact calculations with excluded volume interactions. We start in figure 3 with a plot of the density of states obtained from the RIS model showing the results without and with volume exclusion. Volume exclusion reduces the number of viable conformers, in particular for higher energies. This is simply due to the fact that with excluded volume conformers with multiple gauche defects are suppressed. Of interest here are the bands with a high density of states at fairly low energies, i.e. around 50 meV. Those are, of course,
the conformers that will dominate the thermodynamics of a single PEG molecule. Clearly the probability distributions for the end-to-end length with and without volume exclusion must differ with the former peaking at higher lengths, as seen in figure 4. The effect of volume exclusion is almost negligible in the force-extension curve. This is due to the fact that in the low-force regime the curve is dominated by entropic effects for which the exclusion of a few conformers is not crucial. On the other hand, in the high-force regime the chain is more or less fully stretched so that again volume exclusion does not play a role. We have also employed the TM method to calculate the force-extension curve and find that with all conformers included the result is almost identical to the curve calculated with the exact calculation apart from small numerical discrepancies. On the other hand, including only the trans configurations of the C-O bonds, i.e. only EG monomers of the form \((t \ast t - t \ast t)\), the polymer is softer. This is due to two facts: with fewer conformers entropy is lower so that less force is needed to unfold the polymer; and these conformers are among the longest so that the unfolding immediately results in longer conformers.

Next, we discuss the experimental claim that local helical structures are an important contribution to the statistics of single PEG molecules. A local helix of \(n\) monomers can be defined as a sequence \((tgt)_n\) of consecutive gauche bonds (all of the same helicity) anywhere in the backbone. As an example, we choose the \(2/7\) helix that was used as one of the two fundamental conformers to model the force response in previous studies [1], based on the argument that this type of helix is observed in crystalline PEG. Within the RIS formalism, it is an easy task to sort out these local helical contributions to the radial and other distribution functions. Figure 5 shows the details of the radial distribution functions with at least \(n = \{2, 3, 4\}\) consecutive helical monomers. These curves demonstrate that even sequences with \(n = 4\) are highly improbable and have a negligible contribution to \(P(R)\) and hence for all thermodynamic properties.
Figure 4. (a) The end-to-end distance distributions for our RIS calculations of (EG)$_7$ and (b) the corresponding force-extension relations, also contrasted with TM results.

Figure 5. The excluded volume radial distribution functions of (EG)$_7$ as in figure 4 (black circles) are shown together with the contributions molecules that contain at least two, three or four consecutive helical monomers.

More data is available to support this claim. The relative probability of finding two neighbouring bonds among all possible rotational states has been calculated directly from averaging over all conformers in (EG)$_7$. For the O–C–C and C–C–O bond pairs we find $P_{gt} = 0.531$, $P_{gg'} = 0.149$, $P_{tt} = 0.232$, $P_{tg} = 0.040$, and $P_{gg} = 0.048$, while the results for the C–O–C bond pairs are $P_{tt} = 0.585$, $P_{tg} = P_{gt} = 0.376$, $P_{gg'} = 4 \times 10^{-4}$ and $P_{gg} = 0.038$. From the above probabilities it can be estimated that the probabilities of finding one, two, or three consecutive (EG) units in the same helical state are $7 \times 10^{-2}$, $3 \times 10^{-3}$ and $2 \times 10^{-4}$ respectively. Smith et al [11] have measured this relative occurrence of bond states in PEG melts using small-angle neutron diffraction and their results agree well with our estimates. There is a simple physical
Figure 6. The force extension relations for (EG)_{200} as derived from TM calculations for all possible conformers and taking into account only the conformers of the form (t\ast t). The experimental data comes from a stretching experiment of a PEG molecule in an AFM, done by Oesterhelt et al [1].

(entropic) reason why one should not find large local helical sequences in nature. As indicated in the introduction, a piece of (EG)_{3} has several thousand possible configurations, only two of which are of the type (tgr)_{3}.

All results presented in this paper are for PEG in vacuo. There have been speculative arguments in the literature [1] that solvation in water stabilizes local helical structures. We briefly want to address this problem. In a detailed ab initio study [2] of the energetics and geometry of short PEG molecules closely related to the 2/7 helix, it was shown that in the absence of water the helical structure is the most stable one. This result has been confirmed here, but we also show that this conformer is among many of similar energetics and length. In the presence of water it was found [2] that gauche defects (rather than continuous helices) are stabilized greatly by water cross-linking between three to four neighbouring monomers. Thus it appears that solvation in water would destroy rather than enhance local helical structures. Although this argument is sound it needs further elaboration which we hope to do with further ab initio calculations of EG dimers and trimers in the presence of water.

In our last calculation we compute the force-extension curve for a long chain stretched in an AFM, (EG)_{200}, of 200 monomers, see figure 6. Again the curve obtained with all conformers included lies above the one where only the (t \ast t \ast t) conformers are accounted for. The reason is once more that the restriction to a few conformers, i.e. 3\^{N} instead of 33\^{N} significantly decreases the entropy. In our earlier work we had advanced two scenarios to correct for this flaw, namely to use (EG)_{3} as the base unit in a FJC model but modified in such a way that the base unit itself is stretched appropriately with the applied force [2]. The FJC model then obviously supplies the missing entropy. Another scenario was to enlarge the three-state RIS model to a multistate ICM. The motivation was the following: in the RIS model one takes into account only the configuration at the minimum of the energy for a particular conformer. However, small changes in its length will result in a stretched or compressed conformer of the same overall configuration, albeit with somewhat higher energies. The multistate ICM takes a few additional points (five or seven,
typically) in this potential trough thus enhancing the total number of contributing configurations and results in a larger entropy.

4. Conclusion

In this paper we have revisited the statistical mechanics of single PEG with the intention to check ad hoc approximations made in previous works, namely that of keeping all C–O bonds along the chain in the trans configurations. The theory developed here includes all the bonds on an equal footing. Starting with DFT calculations for the (EG) dimer and trimer, one can determine all the energetic and geometric parameters needed for a RIS model with interactions between nearest neighbour monomers included. The results are force-extension curves and end-to-end distributions for short chains. We use the TM method for long chains of 200 EG monomers, i.e. 600 bonds.

A comparison of our theoretical force-extension curve to experiment, figure 6, shows that the agreement of our best theory, namely accounting for all degrees of freedom, is excellent. Recall that this theory is parameter-free because all energetic and geometric parameters were obtained from the DFT calculations. We have also shown that some previously employed approximation schemes are justified. We have also included a discussion of the issue of helical dominance.

The exact analysis approach employed in this work does not, at present, take into account the interaction between PEG and the surrounding solvent as well as intermolecular interactions (such as hydrogen bonds) between different PEG molecules. Based on the results in section 2.2 and section 3 it will be possible to do exact calculations of such a system in the future. An adaptation of an all-conformer sum in the RIS model will facilitate exact calculations on small PEG clusters, say up to three short molecules. Once the interactions between neighbouring molecules is settled, a mean-field theory along the lines developed by Szleifer and co-workers [25] can be used to model the proper thermodynamics at larger scales. Using ab initio calculations such as presented in [15], it might even be possible to account for intermolecular hydrogen bonding in such a theory. This work will have to involve similar methods as employed in the current paper and rigorous checks of all approximations can and must be made along the way.

Acknowledgments

This work was supported in part by the Office of Naval Research. DS acknowledge scholarships from the National Science and Engineering Research Council of Canada, while FH has been supported through the Killam Trusts. The numerical work was carried out on the Ace-net and Westgrid computing resources.

Appendix A. Extracting RIS and TM energies from ab initio results

The results presented in this work depend heavily on being able to obtain model interaction energies from ab initio data. This appendix serves to describe this extraction, first for the all-conformer RIS analysis for short chains and then for the TM method.
Figure A.1. (a) The RIS model and TM energetics for (EG)$_3$ compared to the DFT predictions. The black line $E_{\text{model}} = E_{\text{DFT}}$ serves as a guide to the eye. The correlation coefficient for the RIS and TM energetics are $r^2 = 0.995$ and 0.946 respectively. (b) Prediction of the conformer lengths for all calculated (EG)$_3$ conformers compared to the DFT results. The correlation coefficient is $r^2 = 0.870$. The RIS geometry is the same in the exact conformational sums and the TM calculations.

A.1. RIS model input parameters

The Hamiltonian (5) for the RIS model is fully determined by the DFT calculation of all non-redundant conformers of (EG)$_1$ and (EG)$_2$ that were described in subsection 2.1. The (EG)$_1$ energies for each conformer are equivalent to the term $E_{\text{self}}$ in the Hamiltonian. Given these numbers, it is straightforward to calculate the interaction energies $V_{\text{NN}}$, taking into account the complete set of (EG)$_2$ results:

$$V_{\text{NN}}(s'_{O-C'}-C'-s'_{C-O}; s_{O-C-C-C'-O}) = E_{\text{DFT}}(s'_{O-C'-C'-C-O}; s_{O-C-C-C-O}) - E_{\text{self}}(s'_{O-C'-C'-C-O}) - E_{\text{self}}(s_{O-C-C-C-O}).$$ (A.1)

To check the accuracy of such a calculation, the energies of all 493 calculated (EG)$_3$ conformers was predicted and compared to the actual values. This comparison is presented in figure A.1 and shows that the agreement over all energy ranges is excellent.

A.2. TM method input parameters

The energies of the (EG)$_3$ conformers that were obtained from our initial DFT calculations have to be expanded into the components of $E_{\text{self}}$ and $V_{\text{NN}}$ defined in equation (13). For example, the conformer $(t|t|t) - (t|t|t) - (t|t|t)$ has an energy

$$E_{(t|t|t)-(t|t|t)-(t|t|t)} = 3E_{\text{self},CC}(t) + 6E_{\text{self},CO}(t) + 3V_{\text{NN},CC} + 6V_{\text{NN},CO}.$$ (A.2)

Due to symmetry, there are four first-order interactions that make up the vectors $E_{\text{self}}$ and six independent values for the second-order energies $V_{\text{NN}}$. These 16 numbers were obtained by
least-square fitting to all 493 equations of type (A.2), with the results presented graphically in figure A.1(a). The self-energies of all the bonds can be written in vector form as

\[ E_{\text{self}} = \begin{pmatrix} E_{g+} \\ E_t \\ E_{g-} \end{pmatrix} \]

with \( E_{\text{self,CC}} = \begin{pmatrix} 28.01 \\ -21.46 \\ 28.01 \end{pmatrix} \) and \( E_{\text{self,CO}} = \begin{pmatrix} -5.21 \\ 15.49 \\ -5.21 \end{pmatrix} \).

All the values are in meV. Similarly, the nearest neighbour interactions, can be written in matrix form

\[ V_{\text{NN}} = \begin{pmatrix} V_{g+g+} & V_{g+t} & V_{g-g} \\ V_{g+t} & V_t & V_{g-t} \\ V_{g-g} & V_{g-t} & V_{g-g} \end{pmatrix}. \]

Symmetry requires that the energetics of all \( g+ \) and \( g- \) states are equal. Moreover, the order of the sequence should not matter. These conditions lead to the following equalities: \( V_{g+g+} = V_{g-g-} \), \( V_{g+g-} = V_{g-g+} \), \( V_{g+t} = V_{g-t} \), \( V_{g+} = V_{g-} \). The absolute values for the two matrices (A.4) are

\[ V_{\text{NN,O–C}} = \begin{pmatrix} 23.29 & 13.25 & -16.61 \\ 13.25 & 2.65 & 13.25 \\ -16.61 & 13.25 & 23.29 \end{pmatrix}, \]

\[ V_{\text{NN,C–C}} = \begin{pmatrix} -5.75 & 4.56 & \infty \\ 4.56 & 8.73 & 4.56 \\ \infty & 4.56 & -5.75 \end{pmatrix}. \]

Note that some combinations were never observed in any of our DFT calculations due to the excluded volume effects discussed previously. Thus, the matrix elements \( V_{\text{NN,C–C}} \) have been set to infinity.

**References**

[1] Oesterhelt F, Rief M and Gaub H E 1999 Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water *New J. Phys.* 1 6.1

[2] Kreuzer H J and Grunze M 2001 Stretching of single polymer strands: a first-principles theory *Europhys. Lett.* 55 640

[3] Doi M and Edwards S F 1986 *The Theory of Polymer Dynamics* (Oxford: Oxford University Press)

[4] de Gennes P-G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)

[5] Flory P J 1988 *Statistical Mechanics of Chain Molecules* (Cincinnati: Hanser/Gardner)

[6] Matsuyama A and Tanaka F 1990 Theory of solvation-induced reentrant phase separation in polymer solutions *Phys. Rev. Lett.* 65 341

[7] Bekiranov S and Bruinsma R and Philip P 1997 Solution behavior of polyethylene oxide in water as a function of temperature and pressure *Phys. Rev. E* 55 577–85

[8] Linse P and Björling M 1991 Lattice theory for multicomponent mixtures of copolymers with internal degrees of freedom in heterogeneous systems *Macromolecules* 24 6700

[9] Linse P 1993 Phase behavior of poly(ethylene oxide)-poly(propylene oxide) block copolymers in aqueous solution *J. Phys. Chem.* 97 13896

*New Journal of Physics* 9 (2007) 68 (http://www.njp.org/)
[10] Smith G D and Jaffe R L 1996 Quantum chemistry study of conformational energies and rotational energy barriers in $n$-alkanes J. Phys. Chem. 100 18718
[11] Smith G D, Yoon D Y, Jaffe R L, Colby R H, Krishnamoorti R and Fetters L J 1996 Conformations and structures of poly(oxyethylene) melts from molecular dynamics simulations and small-angle neutron scattering experiments Macromolecules 29 3462
[12] Bedrov D and Smith G D 1998 Anomalous conformational behaviour of poly(ethylene oxide) oligomers in aqueous solutions. A molecular dynamics study J. Chem. Phys. 1998 8118
[13] Borodin O and Smith G D 2003 Development of quantum chemistry-based force fields for poly(ethylene oxide) with many-body polarization interactions J. Phys. Chem. B 107 6801
[14] Borodin O, Douglas R, Smith G D, Trouw F and Petrucci S 2003 MD simulations and experimental study of structure, dynamics, and thermodynamics of poly(ethylene oxide) and its oligomers J. Phys. Chem. B 107 6813
[15] Wang R L C, Kreuzer H J and Grunze M 2000 The interaction of oligo(ethylene oxide) with water: a quantum mechanical study Phys. Chem. Chem. Phys. 2 3613
[16] Livadaru L, Netz R R and Kreuzer H J 2003 Interacting chain model for poly(ethylene glycol) from first principles—stretching of a single molecule using the transfer matrix approach J. Chem. Phys. 118 1404
[17] Livadaru L, Netz R R and Kreuzer H J 2003 Stretching response of discrete semiflexible polymers Macromolecules 36 3732
[18] Livadaru L and Kreuzer H J 2004 Statistical mechanics of an $n$-alkane chain in $\theta$-condition: going beyond the RIS model Phys. Chem. Chem. Phys. 6 3872
[19] Livadaru L and Kreuzer H J 2003 Confinement of a polymer chain in a tube New J. Phys. 5 95
[20] Frisch M J et al 2004 Gaussian 03 (Wallingford, CT: Gaussian)
[21] Bondi A 1964 van der Waals volumes and radii J. Phys. Chem. 68 441
[22] Lide D R, 2005 CRC Handbook of Chemistry and Physics 86th edn (Boca Raton, FL: CRC Press)
[23] Himmelhaus M, Bastuck T, Tokumitsu S, Grunze M, Livadaru L and Kreuzer H J 2003 Growth of a dense polymer brush layer from solution Europhys. Lett. 64 378
[24] Hanke F, Livadaru L and Kreuzer H J 2005 Adsorption forces on a single polymer molecule in contact with a solid surface Europhys. Lett. 69 242
[25] Carignano M A and Szleifer I 1992 Statistical thermodynamic theory of grafted polymer layers J. Chem. Phys. 98 5006–18