On the Use of the Quasi-Gaussian Entropy Theory in Systems of Polyatomic Flexible Molecules

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2001

Impact Factor: 3.3 · DOI: 10.1021/jp002805z

On the Use of the Quasi-Gaussian Entropy Theory in Systems of Polyatomic Flexible Molecules

Andrea Amadei,*,† Barbara Iacono,‡ Simone Grego,‡ Giovanni Chillemi,§ M. E. F. Apol,‡ Enrico Paci,‡ Maurizio Delfini,‡ and Alfredo Di Nola‡,||

Department of Chemical Sciences and Technology, University of Rome "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Rome, Italy, Department of Chemistry, University of Rome "La Sapienza", Rome, Italy, and Inter-University Computing Consortium (CASPUR), University of Rome, "La Sapienza" Rome, Italy

Received: August 2, 2000; In Final Form: December 4, 2000

In this article we show how the quasi-Gaussian entropy (QGE) theory can be used to treat systems of polyatomic flexible molecules, where the usual semirigid description is not always appropriate. We describe a completely general derivation of the QGE theory which does not make use of any semirigid approximation, and therefore it is very suited for large and flexible molecules. Using molecular dynamics simulations of flexible molecules in vacuo, we investigated the ability of the theory to describe intramolecular energy fluctuations and conformational equilibria of purely classical molecules in the ideal gas condition. Results show that the gamma state level of the theory and its generalization for treating conformational equilibria (multi-gamma state model) provide excellent theoretical models when applied to three polyatomic molecules of increasing conformational freedom.

1. Introduction

Fluids involving polyatomic flexible molecules are often of great scientific and technological interest but usually exhibit very complex thermodynamics. In fact, the statistical mechanics of such systems is dictated by the intramolecular energy fluctuations and their coupling with the intermolecular interactions. In recent papers[1−4] we introduced a new statistical mechanical theory, the quasi-Gaussian entropy (QGE) theory, which is basically an extension of fluctuation theory.[5] We showed[6−10] that with this theory the fluid state thermodynamics can be reproduced with high accuracy for a large variety of typical fluids (not consisting of polyatomic flexible molecules) over a very large temperature−density range. In the QGE theory the fundamental expressions of statistical mechanics are reformulated in terms of the distribution function of the fluctuations of a macroscopic property and, by modeling such a distribution, a complete solution for the statistical mechanical behavior of the system can be obtained. With the use of a few physical-mathematical principles we can restrict the set of possible distributions to a subgroup of "quasi-Gaussian" distributions of increasing complexity. In the previous articles the QGE theory was derived and utilized in the three main statistical mechanical ensembles focusing on typical fluids, without specifically addressing the case of systems involving polyatomic flexible molecules. Such systems, where large conformational fluctuations can occur, require special care in the derivation of the theory based on the excess (ideal reduced) energy fluctuations, and present specific aspects to be investigated. Among them one of the most interesting is the calculation of conformational equilibria of a molecule in the gas phase and in solution which is in general still quite unfeasible, mainly because of the difficulty in estimating the free energies involved. Computer simulation techniques, such as molecular mechanics (MM),[11−17] molecular dynamics (MD),[18−23] and Monte Carlo (MC),[24] have been used extensively to calculate conformational equilibria, and specialized sampling techniques, such as umbrella sampling, thermodynamic perturbation (TP), and thermodynamic integration (TI), coupled to MD or MC have been used both in gas and liquid phases to evaluate the Helmholtz free energy differences due to configurational changes. For small molecules, the internal energy evaluation is straightforward in the gas phase and still possible in solution using nonquantum mechanical approaches. More difficulties can arise in the calculation of the entropic contribution to the free energy, due to poor convergence of the calculations.[18] In addition, the prediction of the temperature dependence of conformational equilibria is still done on empirical basis, e.g. by perturbation expansions in temperature.

In this paper we use a more general derivation of the QGE theory concerning the excess energy fluctuations, which is appropriate also for systems of polyatomic flexible molecules, we introduce a generalization of the gamma state model (the multi-gamma state model) to treat conformational equilibria, and we specifically investigate the properties of three flexible polyatomic molecules (amiridin, cyclohexane, and the alanine dipeptide) in the ideal gas condition using pure classical Hamiltonians to treat the intramolecular potential energy, and molecular dynamics simulations to obtain "experimental" data.

2. Theory

The most general expression of the canonical partition function for a system of N molecules, with a purely classical

* To whom correspondence should be addressed. E-mail: andrea.amadei@uniroma2.it.
† Department of Chemical Sciences and Technology.
‡ Department of Chemistry.
§ Inter-University Computing Consortium.
|| E-mail: dinola@degas.chem.uniroma1.it

10.1021/jp002805z CCC: $20.00
Published on Web 02/13/2001

Hamiltonian, is

$$Q = \frac{1}{h^d N!(1+\gamma)^N}\int^* e^{-\beta U}\, d\xi d\pi = \frac{(2\pi kT)^{d/2}}{h^d N!(1+\gamma)^N}\int^* e^{-\beta(\Phi+\Psi)}\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi \quad (1)$$

where U is the Hamiltonian of the system which is a function of the generalized coordinates ξ and their conjugated momenta π, and the star on the integrals denotes as usual that we integrate only in the accessible part of the configurational space, excluding possible infinite energy configurations (as in the hard body fluids) and/or a configurational subspace which is unaccessible in the whole temperature range of interest.[2] Note that we include the electronic ground-state energy in the classical Hamiltonian. Moreover $(1+\gamma)$ is the symmetry coefficient per molecule necessary to correct the partition function for rotations and intramolecular atomic displacements which correspond to permutations of identical particles which do not change the physical state of the system,[5] and $\det \tilde{M}_j$ is the mass tensor determinant of the jth molecule. d is the total number of classical degrees of freedom, h is the Planck constant and $\beta = 1/(k\,T)$. Finally, Φ and Ψ are the classical inter- and intramolecular potential energy, respectively. It should be noted that eq 1 reduces to the usual expression,[2] valid for semirigid molecules, if $\Pi_{j=1}^{N}(\det \tilde{M}_j)^{1/2}$ is constant over the configurations, and hence the integral over the momenta can be considered independent of the coordinates, although the integrand generally is not. In that case the partition function can be factorized into two independent integrals: one over the coordinates involving the potential energy and another over the momenta involving the kinetic energy.[2,5] However, for polyatomic flexible molecules, especially in ideal gas or infinite dilution conditions where large structural fluctuations can occur because of the absence of the intermolecular potential or of the solute−solute interactions, such a factorization might be not accurate. Hence, in such cases one should use the completely general and always exact expression of the partition function, given by eq 1.

Defining as reference state a system at the same temperature and density as the actual one but with no potential energy, and so without unaccessible configurations in ξ, π phase space, with free energy

$$A_{\mathrm{ref}} = -kT\ln\left(\frac{(2\pi kT)^{d/2}}{h^d N!(1+\gamma)^N}\int\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi\right) \quad (2)$$

we can express the ideal reduced (excess) free energy $A' = A - A_{\mathrm{ref}}$ as

$$A' = -kT\ln\left(\frac{\int^* e^{-\beta U'}\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi}{\int^*\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi}\right) - kT\ln\epsilon = kT\ln\langle e^{\beta U'}\rangle - kT\ln\epsilon \quad (3)$$

with $U' = \Phi + \Psi$, and

$$\langle e^{\beta U'}\rangle = \frac{\int^* e^{-\beta U'} e^{\beta U'}\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi}{\int^* e^{-\beta U'}\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi}\int \rho(U') e^{\beta U'} dU' \quad (4)$$

$$\epsilon = \frac{\int^*\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi}{\int\prod_{j=1}^{N}(\det \tilde{M}_j)^{1/2} d\xi} \quad (5)$$

Here $\rho(U')$ is the probability distribution of the ideal reduced (or excess) energy U', and ϵ is the fraction of accessible phase space, weighted with the mass tensor determinant (confinement fraction), and the expectation value in eqs 3−4 (the moment generating function) is evaluated in the accessible part of the configurational space. Note that the last equations are a generalization of the ones given in previous papers where using the semirigid approximation the partition function could be always factorized into the momenta and the configurational integrals. In fact, when the product of the determinants of the mass tensors can be considered independent of the coordinates eqs 3−5 reduce exactly to the previous ones.[2] Note also that when the bond vibrations must be treated quantum mechanically, such a condition can be easily introduced in the derivation, similarly to what was done in previous papers for semirigid molecules.[2]

Instead of the usual perturbation expansion, in the QGE theory we solve eq 3 by modeling the probability distribution function of the excess energy U' and hence its moment generating function. As previously discussed,[1−4] any macroscopic system can be regarded as a very large collection of identical and statistically independent subsystems which define the minimal size required to exhibit a full thermodynamic behavior. In the QGE theory, such minimal thermodynamic systems are referred to as elementary systems. The fluctuation distributions of the macroscopic system are then defined by the convolution of the elementary ones and hence, from the central limit theorem, any macroscopic distribution must be close to a Gaussian at least in the vicinity of its mode (quasi-Gaussian distributions). Assuming that the probability distributions and hence moment generating functions for the fluctuation of the properties of a single subsystem, defined by a given integer number of elementary systems, can be analytically described, we can use their convolution to model the macroscopic fluctuation distributions. Note that a single elementary system includes in general a large number of atomic degrees of freedom and hence allows the use of the approximated versions of the central limit theorem to show that the fluctuation distributions of a subsystem defined even by one or a few elementary systems, could be well modeled by unimodal-like distributions. Each model distribution function of the excess energy provides via eq 4 a closure relation for the general thermodynamic equation $\partial S'/\partial T = C_V'/T$ where clearly S' and C_V' are the ideal reduced entropy and isochoric heat capacity, respectively. Hence we obtain an ordinary differential equation, the thermodynamic master equation (TME), the solution of which provides the complete temperature dependence of the system at fixed density. It should be noted that for each model distribution, via the TME, we obtain the corresponding

exact thermodynamics and so every distribution, or moment generating function, defines in a unique way a statistical state of the system.

The simplest fully physically acceptable statistical state is the gamma state defined by the gamma distribution for the macroscopic excess energy fluctuation

$$\rho(U') = \frac{b_1(1/b_1^2)^{b_0/b_1^2}}{\Gamma(b_0/b_1^2)}(b_0 + b_1\Delta U')^{b_0/b_1^2-1} \exp\left\{-\frac{b_0 + b_1\Delta U'}{b_1^2}\right\} \quad (6)$$

with $\Gamma(\cdot)$ the gamma function,[25] $b_0 = M_2$ and $b_1 = M_3/(2M_2)$. Here M_2 and M_3 are the second and third central moments of the excess energy U', respectively, and $\Delta U' = U' - \langle U' \rangle$. Note that a macroscopic gamma distribution is obtained by the convolution of gamma distributions as this distribution as the Gaussian is close to the convolution. In the case of interest of this paper (ideal gas conditions) where the elementary system is defined by a single molecule, we only deal with the intramolecular potential energy, and so the moment generating function does not depend on the volume of the system. Moreover, for a homogeneous system of purely classical mechanical molecules, eq 5 reduces to

$$\epsilon = \frac{\int^* \prod_{j=1}^{N} (\det \tilde{M}_j)^{1/2} d\xi}{\int \prod_{j=1}^{N} (\det \tilde{M}_j)^{1/2} d\xi} = \frac{(\int^* dx)^N}{(\int dx)^N} = \left(\frac{\Omega V}{V^n}\right)^N = \frac{\Omega^N}{V^{(n-1)N}} \quad (7)$$

where $\boldsymbol{x}$ are the atomic coordinates per molecule, n is the number of atoms per molecule, N the number of molecules in the system, and Ω the accessible configurational volume of the atoms in the molecule with fixed center of mass, which is temperature and density independent. Note that in this case, where the generalized coordinates $\boldsymbol{\xi}$ are the full set of the atomic Cartesian coordinates, the mass tensor determinants are exact constants and hence can be removed. The ideal reduced internal energy, entropy, heat capacity, and pressure using the gamma state model are[2]

$$U'(T) = U'_0 + (T - T_0)\, C'_{V0}\, \frac{\delta(T)}{\delta_0} \quad (8)$$

$$S'(T) = \frac{C'_{V0}}{\delta_0^2}[\delta(T) + \ln(1 - \delta(T))] + k \ln \epsilon \quad (9)$$

$$C'_V(T) = C'_{V0}\left(\frac{\delta(T)}{\delta_0}\right)^2 \quad (10)$$

$$p'(T) = \frac{NkT}{V}(1 - n) \quad (11)$$

where

$$\delta(T) = \frac{M_3}{2kTM_2} = \frac{T_0\delta_0}{T(1 - \delta_0) + T_0\delta_0} \quad (12)$$

and the zero subscript denotes as usual the value of the property at the arbitrary reference temperature T_0. It must be noted that for such a system where we assume ideal gas behavior with a purely classical Hamiltonian at every temperature, we can never encounter a phase separation along any possible isochore, and hence a single TME solution might be valid from zero Kelvin to infinite temperature. This is in contrast with real molecular systems where phase transitions produce a singularity in the TME,[3,4] implying that the solution obtained in one phase cannot be used to extrapolate to another phase or in biphasic conditions. Moreover, for real molecular systems any statistical state based on a continuous energy distribution cannot be acceptable below a certain temperature where only a pure quantum mechanical description can be valid.[26] If we use a gamma state model from 0 Kelvin on, the last equations can be simplified. In fact since for $\beta \to \infty$ any classical system behaves as a harmonic one defined by the overall potential energy minimum, for a gamma state we have

$$\lim_{\beta\to\infty} C'_V = \frac{C'_{V0}}{\delta_0^2} = \frac{1}{2}N(3n - 6)k \quad (13)$$

$$\lim_{\beta\to\infty} U' = U'_0 - T_0\frac{C'_{V0}}{\delta_0} = N\psi_{\min} \quad (14)$$

where ψ_{min} is the overall potential energy minimum of a single molecule. From the last two equations we obtain

$$C'_{V0} = \frac{1}{2}N(3n - 6)k\delta_0^2 \quad (15)$$

$$U'_0 = N\psi_{\min} + \frac{1}{2}N(3n - 6)k\delta_0^2 T_0 \quad (16)$$

and hence, upon removing C'_{V0} and U'_0, we can rewrite eqs 8−10 as

$$U'(T) = N\psi_{\min} + \frac{1}{2}N(3n - 6)kT\delta(T) \quad (17)$$

$$S'(T) = \frac{1}{2}N(3n - 6)k[\delta(T) + \ln(1 - \delta(T))] + \mathrm{N}\, k \ln \bar{\epsilon} \quad (18)$$

$$C'_V(T) = \frac{1}{2}N(3n - 6)k\delta^2(T) \quad (19)$$

with clearly $\bar{\epsilon} = \epsilon^{1/N}$ the molecular confinement fraction. Note that the last equations reduce to the pure harmonic behavior when $\delta_0 = 1$.[1] Note also that from eqs 9 and 18 the excess entropy S' diverges when T tends to zero (but A' converges to the minimum energy value) and this result is fully statistical mechanically consistent, as described in details in a previous paper.[2] A generalization of the gamma state model, the multi-gamma state model, can be easily obtained. We assume that the molecular configurational space of the internal coordinates can be partitioned into a set of L subspaces, each exactly described by a "local" gamma state. For a homogeneous system

in the ideal gas condition eq 3 then becomes

$$A' = -NkT \ln \left(\frac{\sum_{i=1}^{L} \int_i^* e^{\beta U'} (\det \tilde{M})^{1/2} dx}{\int (\det \tilde{M})^{1/2} dx} \right) = -NkT \ln \left(\sum_{i=1}^{L} \frac{\bar{\epsilon}_i}{\langle e^{\beta U'} \rangle_i} \right) \quad (20)$$

where

$$\langle e^{\beta U'} \rangle_i = \left(\frac{\int_i^* e^{-\beta U'} e^{\beta U'} (\det \tilde{M})^{1/2} dx}{\int_i^* e^{-\beta U'} (\det \tilde{M})^{1/2} dx} \right) \quad (21)$$

$$\bar{\epsilon}_i = \frac{\int_i^* (\det \tilde{M})^{1/2} dx}{\int (\det \tilde{M})^{1/2} dx} \quad (22)$$

with clearly U' and $\tilde{M}$ the single molecule intramolecular ideal reduced energy and mass tensor determinant. For each ith subspace we can solve an independent "local" TME equation providing the temperature dependence of a system confined in that subspace. We can then rewrite eq 20 as

$$A'(T) = -NkT \ln \left(\sum_{i=1}^{L} e^{-\beta a_i'} \right) \quad (23)$$

where

$$a_i' = \frac{A_i'}{N} = kT \ln \langle e^{\beta U'} \rangle_i - kT \ln \bar{\epsilon}_i \quad (24)$$

is the molecular ideal reduced free energy of the system when it is confined in the ith subspace. From eqs 8 and 9 we can obtain a_i' using U_0'/N, C_{V0}'/N, δ_0, and $\bar{\epsilon}$ of the ith subspace. From the last equations it is easy to derive all other properties of the system. Note that a similar derivation, providing the same expressions, can be used for infinitly diluted solutes considering now U' as the ideal reduced energy of a system defined by a single solute molecule in the solvent.

In this article, we will use the double and triple gamma state models assuming that each configurational subspace behaves as a harmonic system in the zero temperature limit (i.e., the minimum energy configuration is always a stationary point), and that the same value of δ_0 can be used for the different subspaces. These assumptions provide a relevant simplification of the model without introducing a physically unreasonable behavior. Identical δ_0 values, using eqs 17−19, mean that the energy probability distributions of different subspaces must have identical central energy moments at every temperature, implying a similar shape of the potential energy surfaces describing each subspace. As consequence, for such a simplyfied multi-gamma state model, different configurational subspaces can differ in their thermodynamics only because of the potential energy minimum and phase-space confinement fraction,

$$A_i'(T) - A_1'(T) = N(\psi_{\min,i} - \psi_{\min,1}) - NkT \ln J_i \quad (25)$$

$$U_i'(T) - U_1'(T) = N(\psi_{\min,i} - \psi_{\min,1}) \quad (26)$$

with $J_i = \bar{\epsilon}_i/\bar{\epsilon}_1$. The possible errors introduced by this approximation are likely to be not very relevant for many molecular systems. In that case the excess free energy can be written as

$$A'(T) = Na_1'(T) - NkT \ln \{ \sum_{i=1}^{M} e^{-\beta(a_i' - a_1')} \} = Na_1'(T) - NkT \ln \{ \sum_{i=1}^{M} J_i e^{-\beta(\psi_{\min,k} - \psi_{\min,1})} \} \quad (27)$$

The values $M = 1$, 2, and 3 will be used for amiridin (single gamma state), cyclohexane, and the alanine dipeptide, respectively. From the last equation we can obtain the excess internal energy and the probability p_i of each configurational subspace as a function of temperature,

$$U'(T) = N\psi_{\min,1} + \frac{1}{2}(3n - 6)NkT\,\delta(T) + N \frac{\sum_{k=2}^{M} (\psi_{\min,k} - \psi_{\min,1}) J_k e^{-\beta(\psi_{\min,k} - \psi_{\min,1})}}{1 + \sum_{k=2}^{M} J_k e^{-\beta(\psi_{\min,k} - \psi_{\min,1})}} \quad (28)$$

$$p_i(T) = \frac{J_i e^{-\beta(\psi_{\min,i} - \psi_{\min,1})}}{1 + \sum_{k=2}^{M} J_k e^{-\beta(\psi_{min,k} - \psi_{min,1})}} \quad (29)$$

3. Methods

3.1. Simulation Methods. The results discussed in the next section are obtained by MD simulations of a single molecule in vacuo, performed with canonical ensemble dynamics and applying roto-translational constraints.[27] Simulations were performed using isothermal Gaussian temperature coupling[28] in the temperature range 1−800 K for amiridin, 100−800 K for cyclohexane, and 100−600 K for the alanine dipeptide (*N*-acetylalanin-*N*′-methylamide). Simulations at 1 K were carried out also for cyclohexane and alanine dipeptide to estimate the minimum potential energy of different configurational subspaces. Both the isothermal Gaussian temperature coupling and the roto-translational constraint algorithms were implemented in the GROMACS simulation package.[29] The GROMACS force field was used in the simulations with a united atom representation for nonpolar hydrogens, and harmonic bond-stretching potentials for all bonds.[29−33] In the case of amiridin (9-amino-2,3,5,6,7,8-hexahydro-1H-cyclopenta[*b*]quinoline) the atomic charges were obtained from ab initio calculations: molecular geometry was optimized at SCF level with the Dunning Huzinaga (D95) double-ζ basis set, using the Gaussian94 package.[34] This was done to obtain a better description of the electrostatic interactions as the standard GROMACS force field is not optimized for amiridin. Note also that in order to simplify amiridin structural behavior we used improper dihedral angle potentials to inhibit the conformational transitions of the two saturated rings. Amiridin was then used as a model for polyatomic molecules dominated basically by bond length and angle fluctuations. For all simulations, an "infinite" cut off radius was used to guarantee that each atom could interact with any other in the molecules, and to obtain very accurate simulation data we always used a short time step (0.20 fs for amiridin and

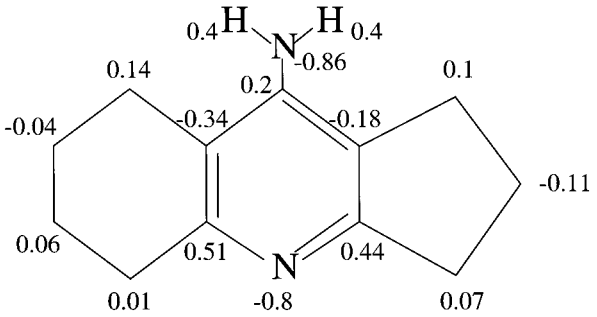

Figure 1. Chemical structure of amiridin with the atomic charges obtained from ab initio calculations. The two polar hydrogens, attached to the nitrogen, and the two nitrogens are explicitly represented. All the carbons (bare or united with the apolar hydrogens) are not explicitly represented.

0.1 fs for cyclohexane and alanine dipeptide), instead of the usual 0.30−0.40 fs time step which can be used in simulations with explicit bond length vibrations. Finally the production runs, after equilibration, were 1.0 ns for amiridin and 4.0 ns for cyclohexane and the dipeptide, with initial velocities obtained from a Maxwellian distribution at the desired temperature.

3.2. Definition of the Configurational Subspaces. For both cyclohexane and alanine dipeptide, a principal component analysis of the atomic positional fluctuations (Essential Dynamics analysis[35]) was performed and the essential subspace was defined by the first two principal component vectors (eigenvectors) that accounted for the majority of the atomic fluctuations. The conformational behavior of both molecules was monitored in this essential subspace. For the alanine dipeptide we also used an analysis based on the Ramachandran plot, i.e., the projection of the trajectory onto the plane of the ϕ and ψ torsion angles around the central C_α atom. The number of subspaces used to partition the configurational space was based for cyclohexane on the probability observed at 700 K in the space defined by the first two eigenvectors and for the alanine dipeptide on the probabilities observed at 500 K in the spaces defined by the first two eigenvectors and by the ϕ,ψ angles.

4. Results

4.1. Amiridin. In Figure 1 the chemical structure of amiridin is shown. Note that the use of extra improper dihedral angle potentials in the saturated rings largely inhibited their conformational transitions up to high temperature. In Figure 2 we show the average potential energy of the simulations as a function of the temperature, and the behavior of different models. From the figure it is clear that a pure harmonic model ($\delta_0 = 1$) can only be acceptable at low temperatures (up to 100 K) and the usual second- and third-order free energy cumulant expansions fitted to the simulation data are completely wrong as shown in a previous paper.[9] This fact is not surprising as cumulant expansions cannot provide a general theoretical model but simply a local approximation valid in a certain range, preferably at high temperature. Only the use of a second-order temperature expansion of the ideal reduced internal energy, centered at 0 Kelvin,

$$U' = N\psi_{\min} + \frac{1}{2}N(3n - 6)kT + \frac{\partial C_V'}{\partial T}(T{=}0)\frac{T^2}{2}$$

gives a really good description of the system behavior in the whole temperature range, where $\psi_{\min}$ and $\partial C_V'/\partial T$ were obtained by fitting. As shown in Figure 3, the gamma state model, obtained by fitting the data with eq 17, provides an excellent description of the system in the whole temperature range and, on this scale, is undistinguishable from the second-order energy expansion. The average relative error in both cases is about 0.6%, corresponding to an average absolute error of about 0.2 kJ/mol. Note that the error bars of the average energies obtained from simulations (defined as the standard deviation of the average energy[9]) are always below 0.05 kJ/mol and hence are invisible on the figures scale. In Figure 4 we show the same energy data with a gamma state model obtained using only two energies at 300 and 400 K to solve, via eq 17, the two parameters $\psi_{\min}$ and δ_0. These new gamma state parameters are almost identical to the ones obtained by fitting the whole set of data, and indeed this new gamma state model reproduces with high accuracy the energy behavior of the molecule in its

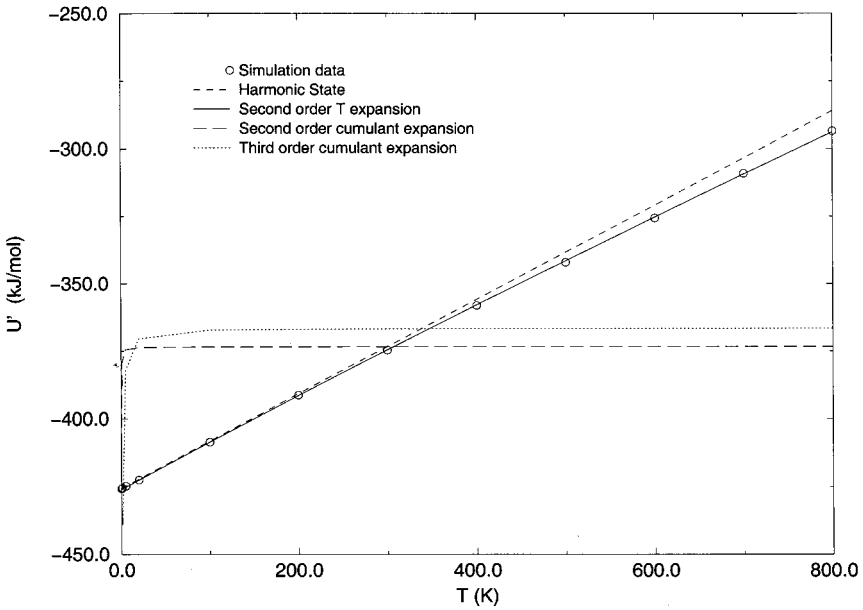

Figure 2. Average potential energy (ideal reduced internal energy) of amiridin, obtained from the simulations (circle) and different models used to describe the system: the harmonic state (dashed line), the temperature second-order expansion of ideal reduced internal energy (solid line), and the second- and third-order cumulant expansions of ideal reduced free energy (long dashed and dotted lines).

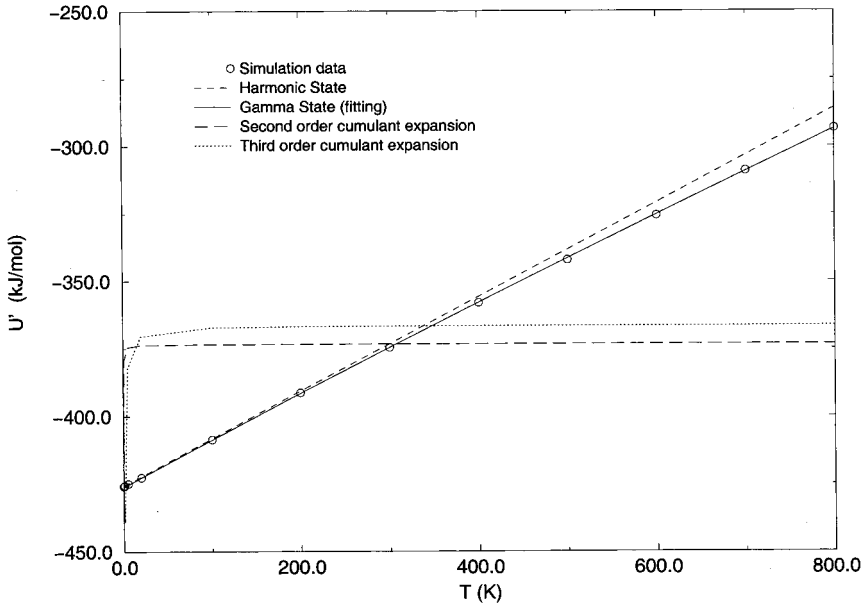

Figure 3. Average potential energy (ideal reduced internal energy) of amiridin, obtained from the simulations (circle) and different models used to describe the system: the harmonic state (dashed line), the gamma state model obtained by fitting (solid line), and the second- and third-order cumulant expansions of the ideal reduced free energy (long dashed and dotted lines).

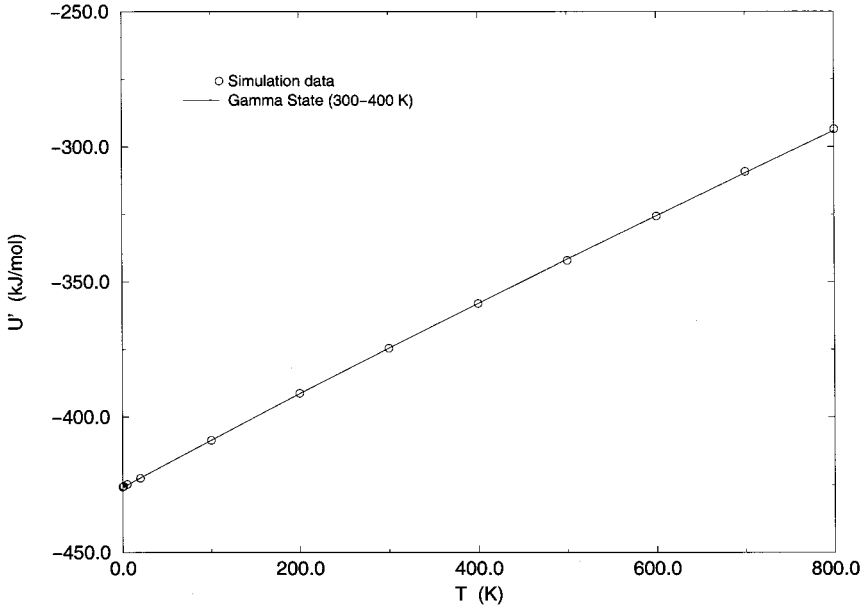

Figure 4. Average potential energy (ideal reduced internal energy) of amiridin, obtained from the simulations (circle) and the gamma state model obtained using only energy values at 300 and 400 K.

extrapolation ranges (below 300 and above 400 K). Finally, in Figure 5 the heat capacities for the potential energy obtained from the simulations via $C_V' kT^2 = M_2$, are shown as a function of the temperature. In the same figure the heat capacity predictions from the two gamma state models, together with the pure harmonic state and the second order energy expansion ones, are also reported. Again the theoretical predictions of the gamma states are rather good reproducing the data, in the whole temperature range, within 5%, corresponding to a maximum absolute error which is lower than 9 J/K mol, although no heat capacity data where used to parametrize both gamma state models. The second-order energy expansion is comparable in accuracy to the gamma state models. Note that since the standard deviations of the heat capacities[9] are relatively large, as this property is calculated from the second central moment, the models' predictions now reproduce most of the data within a couple of standard deviations (twice the error bars). Note also that the apparent oscillation of the simulation heat capacity in the figure, is likely to be due to the relatively large indetermination of the data, as also shown by the fact that the models predictions almost coincides with the heat capacities obtained as temperature derivatives of the simulation average energies.

4.2. Cyclohexane. The conformational equilibria of cyclohexane were modeled with a double gamma state. The projection of the trajectory of cyclohexane onto the first two eigenvectors of the principal component analysis at 700 K (Figure 6) shows two configurational subspaces; one defined by two isoenergetic regions (chair conformations) and one by the zone in the middle with higher potential energy (twist-boat conformation).

To define the exact borders of the two subspaces we used the criterion that at 700 K the average potential energy difference between the subspaces should be equal to the difference of the corresponding minimum potential energies, evaluated at 1 K (cf. eq 26). We have chosen two slightly different partitions

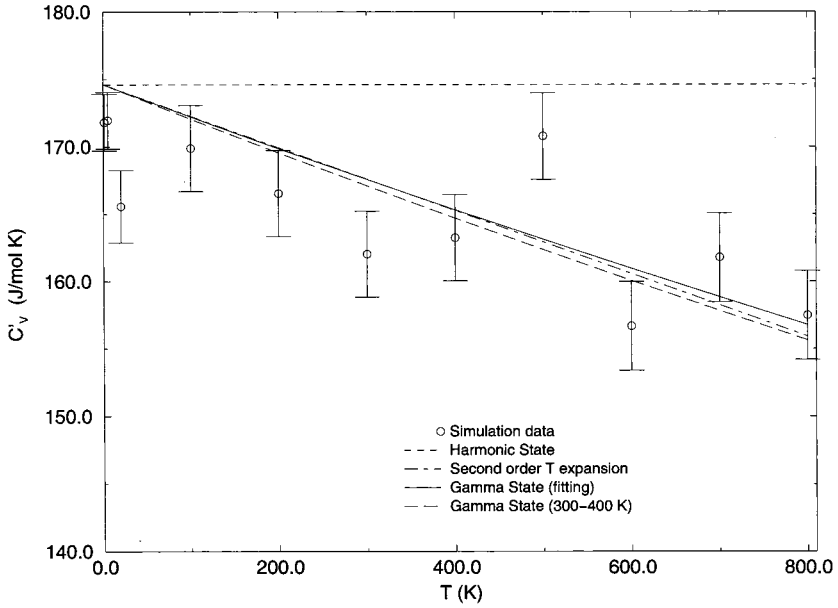

Figure 5. The potential energy heat capacity (ideal reduced heat capacity) of amiridin, obtained from the simulations (circle) and the predictions obtained by the two gamma state models (solid and long dashed lines), together with the pure harmonic state (dashed line) and the second-order temperature expansion of the ideal reduced internal energy (dashed dotted line). The error bars are given by the standard deviations.

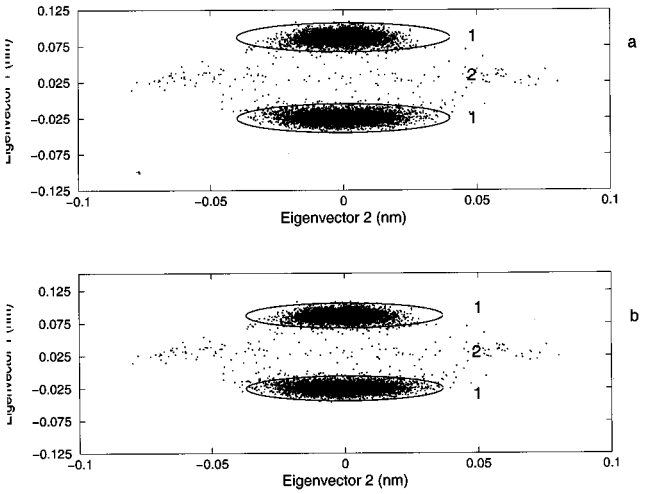

Figure 6. Projection of the trajectory of cyclohexane onto the plane of the first two eigenvectors at 700 K. Parts a and b show different partitions of configurational space. Subspaces 1 (chair) and 2 (twist-boat conformation) are indicated.

that satisfy this criterion at 700 K. For each subspace we calculated the average potential energy and relative probability; this latter quantity will be referred to as "experimental" probability. The parameters of the double gamma state model were obtained as follows. First, we fitted a single gamma state, eq 17, to the average potential energy of subspace 1 as a function of temperature in order to obtain $\psi_{\min,1}$ and δ_0. Second, the double gamma state model for the total configurational space, eq 28 with $M = 2$, was fitted to the total average potential energy, yielding J_2 and $\psi_{\min,2}$. With these parameters, eq 29 gives the "theoretical" probability for each subspace as a function of temperature. The parameters obtained in this way for the two different partitions are given in Table 1. In the same table the Helmholtz free energy differences between chair and twist-boat conformations at 300 K are also reported. It must be noted that the two different partitions of configurational space give similar parameters. Moreover, the free energy differences calculated at 300 K by eq 26 using the two different partitions differ only by 0.01 kJ/mol.

TABLE 1: Values of the Parameters δ_0, ψ_{min} and J_i Obtained with a Double Gamma State Model for Cyclohexane Using the Partition of Configurational Space as Shown in Figure 1,b

subspace	δ_0	$\psi_{\min,i}$ (kJ/mol)	J_i	$A_i'(300) - A_1'(300)$ (kJ/mol)
1[a]	0.9913	26.29	1.0	
2[a]		48.9881	0.3076	19.76[c]
1[b]	0.9913	26.3649	1.0	
2[b]		49.0592	0.3077	19.75[c]

[a] Parameters obtained with the partition of Figure 1a. [b] Parameters obtained with the partition of Figure 1b. [c] The free energy differences, evaluated at 300 K.

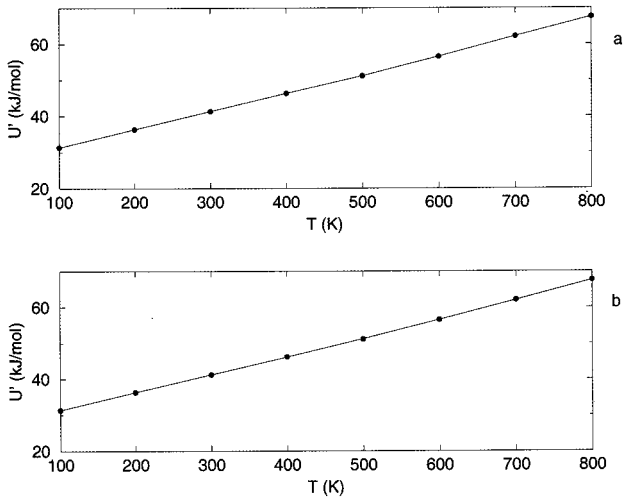

Figure 7. Average potential energy $U'(T)$ of cyclohexane as obtained by simulation (●) and double gamma state model, eq 28 with $M = 2$. (a) Partition of Figure 6a. (b) Partition of Figure 6b.

In Figure 7a,b, the average potential energy obtained by MD simulation and by the double gamma state model (eq 28) are shown as a function of temperature. The double gamma state model gives an excellent description of the changes of the average potential energy within the total temperature range, with an accuracy which is comparable to the amiridin case. Note that just like in that case also here the average energy standard

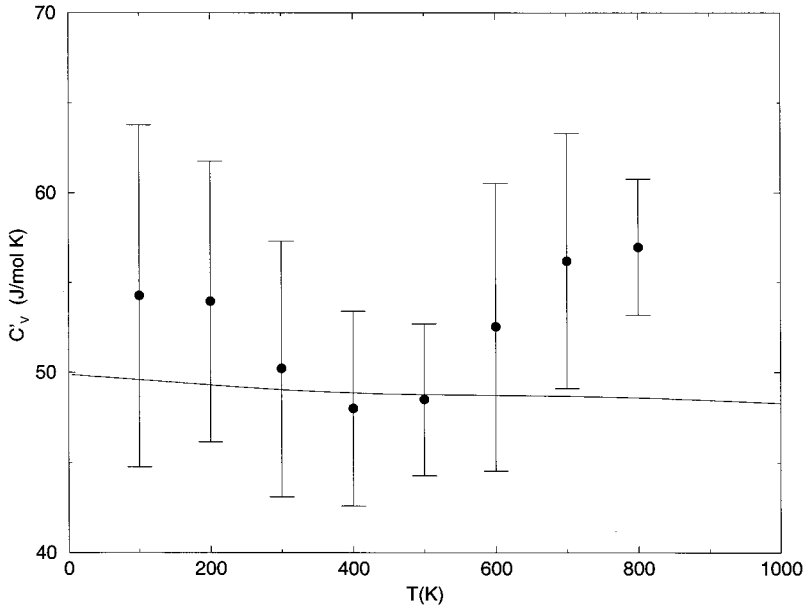

Figure 8. Heat capacities of cyclohexane obtained from simulations (using the second central moment of the potential energy) (●), and model prediction for cyclohexane. The error bars are given by the standard deviations, and the results of the two partitions are undistinguishable in the figure.

deviations are below 0.05 kJ/mol and so invisible on the figure scale. In Figure 8, the heat capacity predicted by the model is compared with the values obtained from the simulations using the second central moment of the potential energy. The theoretical model reproduces rather well the data with an accuracy that is comparable to the amiridin case, and again the deviations for almost all the data are within a couple of standard deviations (twice the error bars). The probability of each subspace, as a function of temperature, is shown in Figures 9a,b. Although the probability data were not used in the parametrization, the theoretical model can reproduce them rather well. Note that the model behavior of the two partitions cannot be distinguished in the figure. It is also worth noting that the average potential energy difference between chair and twist-boat conformations, obtained from the model, eq 26 is equal to 22.70 kJ/ mol, which is very close to the experimental value (23.0 kJ/mol[36,37]).

4.3. *N*-Acetylalanin-*N′*-methylamide. The conformational equilibria of the dipeptide were studied with a triple gamma state model. The projection of the trajectory onto the plane of the first two eigenvectors of the principal component analysis at 500 K (Figure 10) allows us to distinguish two "stable" regions (labeled in the figures as subspaces 1 and 2), corresponding to main allowed regions near $(\phi,\psi) = (-80°,70°)$ and $(\phi,\psi) = (70°, -65°)$ in the ϕ,ψ angle space, see Figure 11. Note that subspace 2 corresponding to the region near $(\phi,\psi) = (70°,-65°)$ becames to be sampled from 400 K on; In Figures 10 and 11 these two subspaces are also indicated. The exact definition of subspaces 1 and 2 has been done at 500 K following, just as for cyclohexane, the criterion that at every temperature the potential energy difference between these subspaces should be equal to the difference of the corresponding minimum potential energies evaluated at 1 K. Note that there are subspaces accessible only from 400 K, as shown by the projection of the trajectory in the ϕ,ψ angle space at 100 and 300 K, Figures 12 and 13, respectively. This confirms that conformations corresponding to regions near $(\phi,\psi) = (70°,-65°)$ do not play a very important role in protein structure;[38] moreover, poor sampling in the α-helix region demonstrates that in vacuo this part of configurational space does not correspond to a stable conformation, as reported also in the literature.[11,14,24]

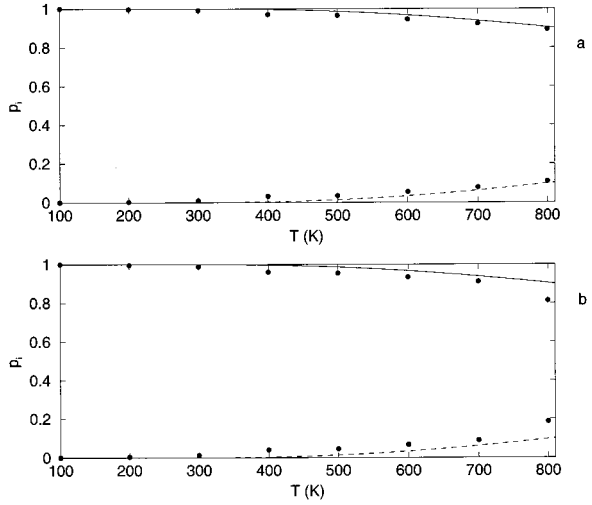

Figure 9. Probabilities p_i of cyclohexane as obtained by simulation (●) and double gamma state model for chair and twist-boat conformations. (a) Partition of Figure 6a. (b) Partition of Figure 6b.

In Figure 14 the average potential energy obtained by MD simulation and the triple gamma state model predictions are shown as a function of temperature. Again the average energy standard deviations are invisible on the figure scale, being comparable to the previous ones. For each subspace the average potential energy and the relative population were calculated within the whole temperature range. To estimate the parameters of the triple gamma state model we used two different procedures. In the first procedure, we fitted the average potential energies of subspace 1 at all temperatures with a single gamma state model, eq 17, obtaining $\psi_{\min,1}$ and δ_0. Then, the average potential energy of the sum of subspace 1 and 2 was fitted with a double gamma state model, eq 28 with $M = 2$, obtaining $\psi_{\min,2}$ and J_2. These parameters were used to fit the average potential energy corresponding to the total configurational space with a triple gamma state model, eq 28 with $M = 3$, yielding $\psi_{\min,3}$ and J_3. In the second procedure the parameters of the triple gamma state model have been obtained from energy data only

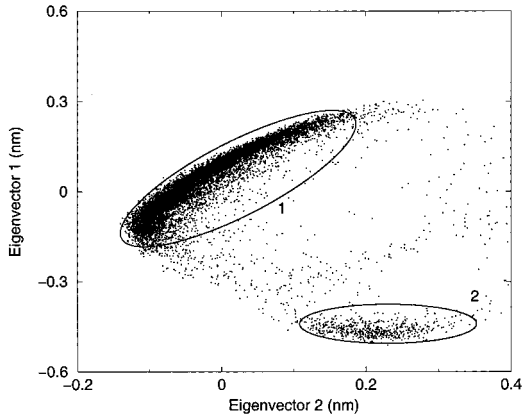

Figure 10. Projection of the trajectory of *N*-acetylalanin-*N*′-methylamide onto the plane of the first two eigenvectors at 500 K. Subspaces 1 and 2 are indicated by the two ellipses.

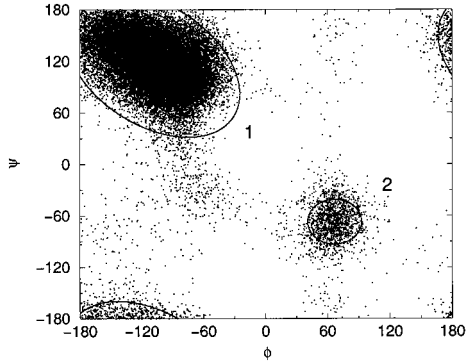

Figure 11. Projection of the trajectory of *N*-acetylalanin-*N*′-methylamide onto the ϕ,ψ space (Ramachandran plot) at 500 K. Subspaces 1 and 2 are indicated by the two ellipses.

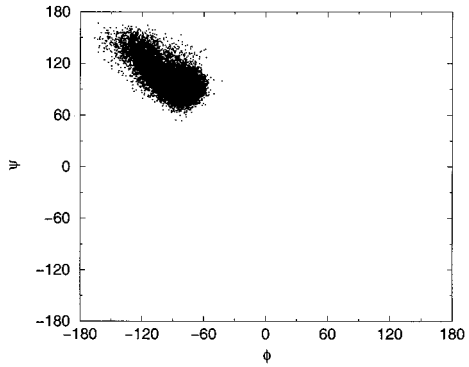

Figure 12. Projection of the trajectory of *N*-acetylalanin-*N*′-methylamide onto the ϕ,ψ space (Ramachandran plot) at 100 K.

at 400 and 500 K: the parameters $\psi_{\min,1}$ and δ_0 were obtained by equating eq 17 at both temperatures to the corresponding average potential energies of subspace 1. Next, $\psi_{\min,2}$ and J_2 were obtained by equating at both temperatures eq 28 with $M = 2$ to the average potential energy of the sum of subspaces 1 and 2, and similarly using the total potential energy at both temperatures, we obtained $\psi_{\min,3}$ and J_3. Both theoretical curves describe the behavior of the system very well and their accuracy is basically the same of the previous molecules. In Figure 15 the heat capacity predicted by the model, obtained by fitting the energy, is compared with the values obtained from the simulations using the second central moment of the potential energy. Again the theoretical model reproduces rather well the data with an accuracy that is comparable to the previous cases, and all the deviations are within a couple of standard deviations (twice the error bars). Finally, in Figure 16 we show the

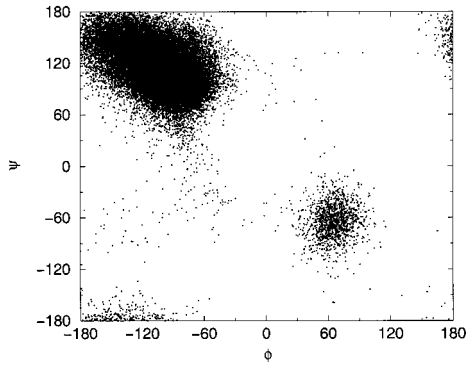

Figure 13. Projection of the trajectory of *N*-acetylalanin-*N*′-methylamide onto the ϕ,ψ space (Ramachandran plot) at 400 K.

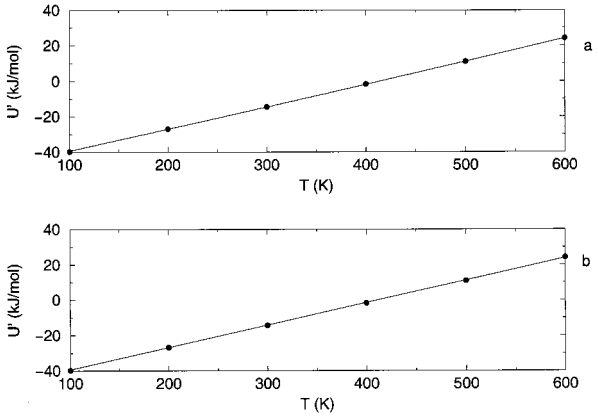

Figure 14. Average potential energy $U'(T)$ of *N*-acetylalanin-*N*′-methylamide as obtained by simulation (●) and triple gamma state model, eq 28 with $M = 3$. (a) Gamma state model using simulation data at all temperatures and (b) gamma state model using simulation data only at 400 and 500 K.

probabilities for the three configurational subspaces obtained by MD simulation as a function of temperature. In the same figure also the theoretical predictions obtained from the triple gamma state model parametrized using both procedure are shown. Also in this case the probability data were not used in the parametrization but both theoretical predictions are in good agreement with the "experimental" probabilities confirminig that the model parametrized with only a few simulation data can still provide an accurate description of the statistical mechanics of the system. Hence, it is possible to carry out accurate free energy calculations while reducing considerably the computational effort. The use of slightly different partitions of configurational space does not alter considerably these results as shown in Table 2.

4.4. Toward Applications to Macromolecules. The results shown so far indicate that conformational equilibria of polyatomic molecules in vacuo can be well described by the multi-gamma state model. However, for a large polyatomic molecule, especially a biomacromolecule, the present method could not be completely reliable. For such systems, different conformations can be present which are separated by very high free energy barriers. In that case a limited number of MD simulations at usual temperature could be unable to provide sufficient sampling to obtain all the necessary information to define the theoretical model, as the evaluation of the parameters of the multi-gamma state model requires at least one simulation which provides a rather complete sampling of the subspaces of interest.

This difficulty can be overcome if we know the accessible configurational subspaces within the temperature range of interest, and we have any mean to estimate the ratio of phase-

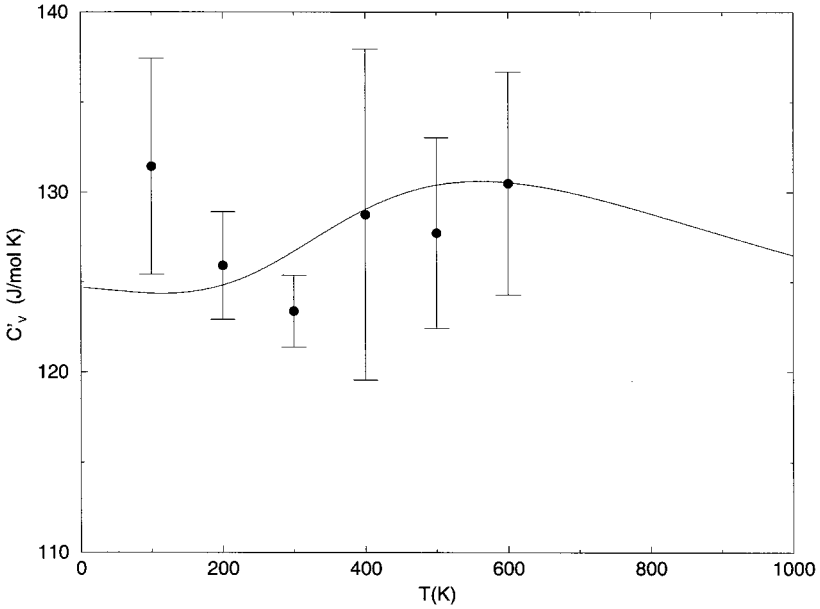

Figure 15. Heat capacities of the dipeptide obtained from simulations (using the second central moment of the potential energy) (●), and model prediction for *N*-acetylalanin-*N*′-methylamide (the error bars are given by the standard deviations). The triple gamma state obtained by fitting all the energy data is shown by solid line and the triple gamma state obtained by only two temperatures is shown by dashed line.

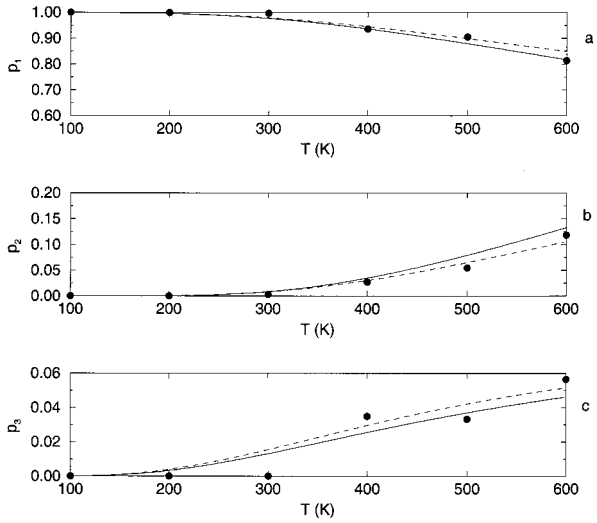

Figure 16. Probabilities p_i of *N*-acetylalanin-N′-methylamide of subspaces 1 (a), 2 (b), and 3 (c) as fuction of temperature, obtained by simulation (●) and triple gamma state model using simulation data at all temperatures or using simulation data only at 400 and 500 K.

TABLE 2: Values of the Parameters δ_0, ψ_{min}, and J_i Obtained with a Triple Gamma State Model for the Alanine Dipeptide Using the Partition of Configurational Space as Shown in Figures 4 and 5

subspace	δ_0	$\psi_{\min,i}$ (kJ/mol)	J_i	$A'_i(300) - A'_1(300)$ (kJ/mol)
1[a]	0.9956	−51.8332	1.0	
2[a]		−44.5362	0.3203	9.83
3[a]		−36.8432	2.7468	
1[b]	0.9920	−51.8186	1.0	
2[b]		−44.9171	0.2195	10.67
3[b]		−37.1486	3.0590	

[a] Parameters obtained with the partition of Figure 4. [b] Parameters obtained with the partition of Figure 5.

space confinement fractions without using any direct information on the conformational transitions. The J_i constants could be estimated by the ratio of the accessible configurational volumes of the subspaces.

As a simple test for such a procedure we used subspaces 1 and 2 of the dipeptide molecule, shown in Figure 10. Assuming that the accessible configurational volume for the coordinates orthogonal to the first two eigenvectors is the same for the two subspaces, the J_2 constant can be readily obtained from the ratio of the areas of the two ellipses shown in the figure. In this way we obtain a value $J_2 \sim 0.29$, which is very close to the value 0.32 obtained from the energy fitting, see Table 2. With the present procedure the free energy difference between the two subspaces at 300 K is 10.05 kJ/mol, while from the energy fitting we obtained 9.83 kJ/mol (Table 2). This means that without any direct information about the conformational transitions, it might be possible to infer for large molecules the value of the J_i parameters.

5. Conclusions

In this paper we introduced a more general derivation of the QGE theory in the canonical ensemble which allows the exact treatment of polyatomic flexible molecules. Such a generalization of the theory can be of great interest for studying large molecules and particularly biomolecules where the statistical mechanics is determined by the internal coordinates, and opens the possibility for the QGE theory to describe conformational equilibria. In the paper we applied the theory to three polyatomic molecules, of increasing flexibility, in vacuo using purely classical Hamiltonians. Results clearly show that the multi-gamma state model can be used as an excellent theoretical model to describe the statistical mechanics of these systems. This approach provides the temperature dependence of the thermodynamic properties of the system including the conformational equilibria of the molecule. Note that although the models reproduce accurately (small average relative and absolute errors) the energy data, their deviations are statistically significant as the average energy standard deviations are very small (lower than 0.05 kJ/mol). This fact is common[10] and simply means, assuming no systematic errors in the simulations, that any theoretical or semiempirical model cannot reproduce the "exact" behavior of the simulated system.

In the case of amiridin and the alanine dipeptide we also demonstrated that the parametrization based on simulation data

at two temperatures allows one to predict the behavior of the molecule within the whole temperature range used, hence reducing the required amount of input information. We also suggest how to apply the method to macromolecules where it is not possible to obtain a complete sampling of configurational space by molecular simulations. In that case, the ratio of the phase-space confinement fractions might be estimated from the sampled configurational volumes as explained in the Results section. This opens the way to investigate and describe conformational equilibria in macromolecules, like proteins. Finally, from the results of amiridin it is also clear that the behavior of the internal coordinates, in a single gamma state subspace, can be well approximated using a second order temperature expansion of the internal energy, and this means that only a moderate anharmonicity is present, at least up to 800 K (similar results were obtained also for different molecules, data not shown). Such a moderate anharmonicity of the intramolecular energy fluctuations implies that the second order energy expansion can be used as an alternative to the gamma state model up to rather high temperatures. However a pure expansion cannot provide a physical model of the system and hence it can be used only as a "local" description, in this case valid in a given range from zero Kelvin and always incorrect in the infinite temperature limit. On the contrary, the gamma state models can really provide a physically coherent description of the system, and hence we expect that the use of the multi-gamma state model will be very useful in theoretical studies of flexible molecules based on molecular simulations.

Acknowledgment. This work was supported by the EC TMR Network Project No. ERBFMRXCT960013 and by the Italian Ministero dell'Universita' e della ricerca scientifica e tecnologica, MURST (National Project "Structural Biology and Dynamics of redox proteins").

References and Notes

(1) Amadei, A.; Apol, M. E. F.; Di Nola, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1996**, *104*, 1560−1574.

(2) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1997**, *106*, 1893−1912.

(3) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 3004−3016.

(4) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 3017−3027.

(5) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics*, 3rd ed.; Pergamon Press: Oxford, 1980; Vol. 5, Part 1 (Statistical Physics).

(6) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1996**, *104*, 6665−6678.

(7) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *Chem. Phys. Lett.* **1996**, *256*, 172−178.

(8) Amadei, A.; Roccatano, D.; Apol, M. E. F.; Berendsen, H. J. C.; Di Nola, A. *J. Chem. Phys.* **1996**, *105*, 7022−7025.

(9) Roccatano, D.; Amadei, A.; Apol, M. E. F.; Di Nola, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 6358−6363.

(10) Amadei, A.; Apol, M. E. F.; Chillemi, G.; Berendsen, H. J. C.; Di Nola, A. *Mol. Phys.* **1999**, *96*, 1469−1490.

(11) Stern, P. S.; Chorev, M.; Goodman, M.; Hagler, A. T. *Biopolymers* **1983**, *22*, 1885−1900.

(12) Aleman, C.; Perez, J. J. *Int. J. Peptide Proteins Res.* **1994**, *43*, 258−263.

(13) Fletcher, M. D.; Campbell, M. M. *Chem. Rev.* **1998**, *98*, 763−795.

(14) Dauber-Osguthorpe, P.; Campbell, M. M.; Osguthorp, D. J. *Int. J. Pept. Proteins Res.* **1991**, *38*, 357−377.

(15) Rodríguez, A. M.; Baldoni, H. A.; Suvire, F.; Vázquez, R. N.; Zamarbide, G.; Enriz, R. D.; Farkas, Ö.; Perczel, A.; McAllister, M. A.; Torday, L. L.; Papp, J. G.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **1998**, *455*, 275−301.

(16) Stern, P. S.; Chorev, M.; Goodman, M.; Hagler, A. T. *Biopolymers* **1983**, *22*, 1901−1917.

(17) Dixon, D. A.; Komornicki, A. *J. Phys. Chem.* **1990**, *94*, 5630−5636.

(18) Di Nola, A.; Brunger, A. T. *J. Comput. Chem.* **1998**, *19*, 1229−1240.

(19) Resat, H.; Mezei, M. *J. Chem. Phys.* **1993**, *99*, 6052−6061.

(20) Mezei, M. *Mol. Simul.* **1993**, *10*, 225−239.

(21) Tobias, D. J.; III, C. L. B. *J. Phys. Chem.* **1992**, *96*, 3864−3870.

(22) Christensen, I. T.; Jørgensen, F. S. *J. Comput.-Aided Mol. Design* **1997**, *11*, 385−394.

(23) Pettitt, B. M.; Karplus, M. *J. Phys. Chem* **1988**, *92*, 3994−3997.

(24) Mezei, M.; Mehrotra, P. T.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2239−2245.

(25) Abramowitz, M.; Stegun, I. A. *Handbook of Mathematical Functions*; Dover: New York, 1972.

(26) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C.; Di Nola, A. *J. Chem. Phys.* **1999**, *111*, 4431−4441.

(27) Amadei, A.; Chillemi, G.; Ceruso, M.; Grottesi, A.; Di Nola, A. *J. Chem. Phys.* **2000**, *112*, 9−23.

(28) Evans, D. J.; Morris, G. P. *Phys. Lett. A* **1983**, *98*, 433−436.

(29) van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; van Drunen, R.; Berendsen, H. J. C. Gromacs User Manual, Version 1.51; University of Groningen: Nijenborgh 4, 9747 AG Groningen, The Netherlands, 1997.

(30) Gromos-87 manual. van Gunsteren, W. F.; Berendsen, H. J. C. Biomos BV; University of Groningen: Nijenborgh 4, 9747 AG Groningen, The Netherlands.

(31) van Gusteren, W. F.; Karplus, M. *Macromolecules* **1982**, *15*, 1528−1544.

(32) Herman, J.; Berensen, H. J. C.; van Gusteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513−1518.

(33) van Buren, A. R.; Marrink, S. J.; Berensen, H. J. C. *J. Phys. Chem* **1993**, *97*, 9206−9212.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, PA, 1994.

(35) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(36) Anet, F. A. L.; Bourn, A. J. R. *J. Am. Chem. Soc.* **1967**, *89*, 760−768.

(37) Squillacote, M.; Sheridan, R. S.; Chapman, O. L.; Anet, F. A. L. *J. Am. Chem. Soc.* **1975**, *97*, 3244−3246.

(38) Anderson, A. G.; Hermans, J. *Proteins: Struct., Funct., Genet.* **1988**, *3*, 262−265.