The accurate computation of partition functions in non-rigid molecules

Alfonso Niño* and Camelia Muñoz-Caro

E.U. Informática de Ciudad Real, Universidad de Castilla-La Mancha, Ronda de Calatrava 5, 13071 Ciudad Real, Spain

Abstract—A software tool for the computation of equilibrium properties in flexible, non-rigid, molecular systems is developed. The program can use semi-rigid and anharmonic models to obtain the rotational and vibrational contributions to the partition function. In addition, a semi-empirical procedure is developed to refine the partition function using experimental information. The program is employed to determine the quantitative performance of the semi-classical closed form for the computation of the rotational partition function. At standard temperature and for medium-size molecules, the error of the usual semi-classical form is found to be smaller than 0.1%. The refinement procedure is tested using the available values for the partition function of the $(HF)_2$ hydrogen-bonded complex. The partition function is computed using *ab initio* data obtained at the MP2(Full)/6-311 + + G($2df$,$2pd$) level. Three models are tested: a pure harmonic model and two anharmonic models. In the anharmonic models, the potential for the hydrogen-bond stretching is described by a Taylor series or by a Morse function. The refinement technique is able to reproduce exactly the reference data in the three cases. © 1997 Elsevier Science Ltd. All rights reserved

1. INTRODUCTION

A useful approach to the study of bioactive compounds, e.g. drugs, is the modellization of the molecular behaviour. In this context, a model able to reproduce accurately thermodynamic properties can be applied to the prediction of molecular properties or to the study of new molecular species, even those not yet synthesized. Equilibrium thermodynamic properties can be computed from the partition function using the statistical mechanical approach. The partition function can be determined from the canonical ensemble and applied to any physical state. However, in the general case, the canonical ensemble partition function, Q, cannot be easily calculated. In the gas phase, it is usually obtained assuming ideal gas behaviour (McQuarrie, 1973; Lucas, 1991). Thus, only when interactions between molecules are very small is this approach correct. In addition, the molecular partition function is obtained within the rigid rotor–harmonic oscillator approach, which assumes very small displacements of the nuclei around their equilibrium positions. However, this approximation fails to describe molecular flexibility,

which is responsible for the conformational behaviour of bioactive molecules (Weinstein and Green, 1981; Richards, 1983). Flexibility is a consequence of anharmonic, large-amplitude vibrations. Since the vibrational energy levels appear as negative exponents in the partition function, an accurate description of these anharmonic (low frequency) vibrations is necessary for a reliable computation of equilibrium properties. In addition, if the rotational–vibrational coupling is important its effect must also be included in the partition function.

In this paper we develop a method for the calculation and refinement of canonical ensemble partition functions that include large-amplitude vibrations and rotational–vibrational coupling. The aim is to compute equilibrium thermodynamic properties quantitatively. The refinement procedure applies a semi-empirical standpoint including experimental information into the theoretical framework. We present a software tool that implements the method. The program makes use of the dynamic memory capabilities of Fortran 90 to perform the direct summation of rotational and vibrational states using dynamic lists. This program is used to test the validity of different approaches for computing the rotational contribution to the partition function. The HF hydrogen-bonded dimer is used as a test case

---

* Author for correspondence. E-mail: anino@inf-cr.uclm.es; Fax: 34-26-295354.

for analysing the stability properties of the refinement method. In this context, we use different potential forms to describe the contribution to the partition function of the hydrogen-bond stretching mode.

2. THEORY

The application of the statistical mechanical approach to flexible molecular systems involves the computation of the partition function. In this work, we want to obtain a partition function that includes, to some extent, anharmonicity and rotational–vibrational coupling.

The canonical partition function, Q, is defined as

$$Q = \sum_i^{\infty} \exp[-\varepsilon_i/kT] \tag{1}$$

where the index runs on every energy state in the ensemble. Starting with an ideal gas approach, we have

$$Q = q^N/N! \tag{2}$$

where q is the molecular partition function and N the number of molecules in the ensemble. The molecular partition function, q, can be calculated as

$$q = \sum_n^{\infty} \exp[-\varepsilon_n/kT] \cdot \sum_e^{\infty} \exp[-\varepsilon_e/kT] \cdot \sum_t^{\infty} \exp$$

$$\times [-\varepsilon_t/kT] \cdot \sum_{rv}^{\infty} \exp[-\varepsilon_{rv}/kT] \tag{3}$$

where the first three terms define q_n, q_e and q_t, the nuclear, electronic and translational partition functions. In this work, we are interested in temperatures close to the standard ambient temperature. Thus, only the ground nuclear and electronic states will be considered. In these conditions, $q_n = 1$ and q_e equals the degeneracy of the ground electronic state. The translational partition function will be calculated as usual (McQuarrie, 1973; Lucas, 1991):

$$q_t = (2\pi MkT/h^2)^{3/2} \cdot V \tag{4}$$

where M represents the total mass of the molecule, k is the Boltzmann constant, h is the Planck constant, T is the absolute temperature and V is the volume.

The last term in equation (3) corresponds to the rotational–vibrational states. In the present approach this term will be considered as a product of P groups of non-interacting states

$$q_{rv} = \prod_i^P \left[\sum_j^{\infty} \exp[-\varepsilon_{ij}/kT] \right] \tag{5}$$

The groups correspond to uncoupled vibrations, several coupled vibrations and the pure overall rotation. When rotational–vibrational coupling is important for some vibrational mode, one group will correspond to the rotational–vibrational interacting states.

When rotation can be separated from vibration, its contribution to the partition function can be obtained using the semi-classical expression (McQuarrie, 1973; Lucas, 1991)

$$q_r = (\pi^{1/2}/\sigma)[(kT)^3/(A \cdot B \cdot C)]^{1/2} \tag{6}$$

where σ is the symmetry number, i.e. the order of the rotational sub-group of the molecule. A more accurate estimate of q_r can be obtained by direct summation of rotational states. In the general case, the asymmetric rotor, the rotational energy levels cannot be determined analytically, but can be computed variationally using symmetric rotor wave functions. In this treatment, the rotational energy levels are classified by a J quantum number. We construct a $(2J + 1) \times (2J + 1)$ Hamiltonian matrix where the elements correspond to the $2J + 1$ different values of k, the projection of J ($K = -J$, $-J + 1,..., 0, 1,..., J$). The KK' element is obtained as (Wang, 1929; Levine, 1970)

$$H_{KK'} = \delta_{K'K}[(2C - A - B)K^2 + (A + B)J(J + 1)]/2$$

$$+ \delta_{K'(K+2)}(B - A)[J(J + 1) - K$$

$$\times (K + 1)]^{1/2} \cdot [J(J + 1) - (K + 1)(K + 2)]^{1/2}/4$$

$$+ \delta_{K'(K-2)}(B - A)[J(J + 1)$$

$$- K(K - 1)]^{1/2} \cdot [J(J + 1) - (K - 1)(K - 2)]^{1/2}/4 \tag{7}$$

where A, B and C represent the overall rotational constants at the equilibrium geometry. Each rotational state is $2J + 1$ times degenerated.

When rotational–vibrational coupling is important we will apply a less crude approximation than the rigid rotor approach above exposed. The easiest way to introduce, to some extent, rotational–vibrational coupling is by using a semi-rigid model. Thus, the stack of rotational energy levels is determined for each vibrational state. The values A, B and C for each vibrational state, Ψ_i, are obtained by quantum mechanical average,

$$\bar{A} = \langle \Psi_i(Q_1,...,Q_n)|A(Q_1,...,Q_n)|\Psi_i(Q_1,...,Q_n) \rangle$$

$$\bar{B} = \langle \Psi_i(Q_1,...,Q_n)|B(Q_1,...,Q_n)|\Psi_i(Q_1,...,Q_n) \rangle$$

$$\bar{C} = \langle \Psi_i(Q_1,...,Q_n)|C(Q_1,...,Q_n)|\Psi_i(Q_1,...,Q_n) \rangle . \tag{8}$$

In equation (8) the rotational terms must be expanded on the vibrational coordinates. Thus, if vibrations from i to j are described with a Taylor series and those from k to l are described with a Fourier expansion, we get

$$A(Q_i,..,Q_j,\theta_k,...,\theta_l) = \prod_{m=i}^{j} \sum_{s=0}^{\infty} \left[A_{ms} \cdot Q_m^s \right] \cdot \prod_{p=k}^{l} \sum_{r=0}^{\infty}$$

$$[a_{pr} \cos(r \cdot \theta_p) + b_{pr} \sin(r \cdot \theta_p)] . \tag{9}$$

Similar expressions hold for B and C.

The contribution of pure vibrations to the partition function can be obtained in several ways. The small-amplitude, harmonic vibrations contribute to equation (5) as (McQuarrie, 1973; Lucas, 1991):

$$q_v = \exp[-v_0/2kT]/(1 - \exp[-v_0/kT]) . \tag{10}$$

In this equation, v_0 is the fundamental frequency of vibration and the origin of energies is placed at the bottom of the potential well. On the other hand, the contribution of large-amplitude vibrations to equation (5) can be obtained by direct summation of states. Thus, we need the vibrational energy levels.

The Hamiltonian for large-amplitude vibrations is (Pickett, 1972; Harthcock and Laane, 1985; Niño and Muñoz-Caro, 1994)

$$\hat{H} = \sum_i^n \sum_j^n (- B_{ij} \frac{\partial^2}{\partial q_i \partial q_j}$$
$$- \frac{\partial B_{ij}}{\partial q_i} \frac{\partial}{\partial q_j}) + V(q_1,...,q_n) \quad (11)$$

where n is the number of vibrations, q are vibrational coordinates, $B_{ij} = (\hbar^2/2)g_{ij}$ are the kinetic terms, and the g_{ij} are obtained from the rotational–vibrational G matrix (Pickett, 1972; Harthcock and Laane, 1985; Niño and Muñoz-Caro, 1994). In the general case, equation (11) cannot be solved analytically. In the present approach, the Hamiltonian will be solved using a variational formalism for the simultaneous treatment of several coupled vibrations using hybrid free rotor + harmonic oscillator basis functions (Niño et al., 1995; Niño and Muñoz-Caro, 1995).

At this point, we can compute a canonical partition function that includes anharmonicity and, to some extent, rotational–vibrational coupling. To go beyond the limits of the present model, the partition function will be refined by including experimental information. The first step of this refinement procedure is the transformation of the partition function in a form easy to handle. Thus, we expand Q in a double Taylor series on the volume, V, and the inverse of the temperature, $Y = 1/T$, obtaining

$$\ln Q(V,Y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[1/(i+j)! \right.$$
$$\left. \times \left(\frac{\partial^{(i+j)} Q}{\partial V^i \partial Y^j} \right)_0 \right] V^i Y^j . \quad (12)$$

To obtain the coefficients of the expansion the following approach is used. We generate a grid of $\ln Q$ values for different values of V and Y. This grid is used to derive an approximation to equation (12) by fitting the $\ln Q$ values to an expression of the form

$$\ln Q(V,Y) = \sum_{i=0}^{m} \sum_{j=0}^{n} A_{ij} \cdot V^i Y^j . \quad (13)$$

Equation (13) represents a truncated Taylor expansion of Q on V and Y. This equation can be expressed in a more useful form, from a computational standpoint. Thus, instead of two indices, i and j, only one is used. Using the new index, the corresponding powers for V and Y are stored in two integer vectors, $v(j)$ and $t(j)$. In this form we obtain

$$\ln Q(V,Y) = \sum_{j=0}^{m+n} A_j \cdot V^{t(j)} Y^{t(j)} \quad (14)$$

where $m + n$ is the number of terms in the expansion. The A_j coefficients are obtained using a least squares approach. The least squares can be performed on $\ln Q$ and its derivatives involved in the calculation of thermodynamic functions: $(\partial \ln Q/\partial T)$, $(\partial \ln Q/\partial V)$ and $(\partial^2 \ln Q/\partial T^2)$. These derivatives can be analytically calculated as

$$(\partial \ln Q/\partial V) = N_a /V$$

$$(\partial \ln Q/\partial T) = N_a[3/2T + \sum_i (\partial q_i/\partial T)/q_i]$$

$$(\partial^2 \ln Q/\partial T^2) = N_a\{ - 3/2T^2 + \sum_i [(\partial^2 q_i/\partial T^2)q_i$$
$$- (\partial q_i/\partial T)^2]/q_i^2\} \quad (15)$$

with

$$(\partial q_i/\partial T) = \sum_j [(\varepsilon_{ij}/kT^2)\exp(- \varepsilon_{ij}/kT)]$$

$$(\partial^2 q_i/\partial T^2) = \sum_j [(\varepsilon_{ij}/k)(\varepsilon_{ij}/kT^4 - 2/T^3)\exp(- \varepsilon_{ij}/kT)] .$$
$$(16)$$

In the above expressions the q_i values represent the molecular vibrational and rotational partition functions. When the rotational partition function is calculated with equation (6), equation (16) reduces to

$$(\partial q_r/\partial T) = 3q_r/2T$$

$$(\partial^2 q_r/\partial T^2) = 3q_r/4T^2 . \quad (17)$$

In a similar way, if the vibrational partition function is calculated using equation (10), the vibrational contributions to equation (15) can be obtained as

$$(\partial \ln q_v/\partial T) = (v_0/2kT^2)$$
$$\times [(1 + \exp(- v_0/kT))/(1 - \exp(- v_0/kT))]$$

$$(\partial^2 \ln q_v/\partial T^2) = (v_0/k^2T^4)$$
$$\times [\exp(-v_0/kT)/(1 - \exp(-v_0/kT))^2]$$
$$- 2(\partial \ln q_v/\partial T)/T . \quad (18)$$

To apply the least squares condition we can define an auxiliary function S

$$S = \sum_k^n [\ln Q_k - \ln Q_k^c]^2 + [T_k(\partial \ln Q_k/\partial T - \partial \ln Q_k^c/\partial T)]^2$$
$$+ [T_k^2(\partial^2 \ln Q_k/\partial T^2 - \partial^2 \ln Q_k^c/\partial T^2)]^2$$
$$+ [V_k(\partial \ln Q_k/\partial V - \partial \ln Q_k^c/\partial V)]^2 \quad (19)$$

where n represents the number of data in the grid of $\ln Q$, and $\ln Q^c$ represents values of $\ln Q$ calculated with equation (14). The A_j coefficients are computed by minimizing equation (19) with respect to every A_j:

$$(\partial S/\partial A_i) = 0, \forall i = 1,...,m . \quad (20)$$

Thus, for a given coefficient i, we obtain:

$$(\partial S/\partial A_i) = - 2\sum_k^n [\ln Q_k - \sum_j A_j V_k^{t(j)} Y_k^{t(j)}] V_k^{t(i)} Y_k^{t(i)}$$
$$+ \sum_k^n T_k[(\partial \ln Q/\partial T)_k - \sum_j t(j)A_j V_k^{t(j)} Y_k^{t(j)+1}]t(i) V_k^{t(i)} Y_k^{t(i)}$$

$$- 2\sum_k T_k^2[(\partial^2 \ln Q/\partial T^2)_k - \sum_j t(j)(t(j) + 1)A_j V_k^{t(j)} Y_k^{t(j)+2}]$$

$$t(i)(t(i) + 1)V_k^{t(i)} Y_k^{t(i)} - 2\sum_k V_k[(\partial \ln Q/\partial V)_k$$

$$- \sum_j v(j)A_j V_k^{t(j)-1} Y_k^{t(j)}]v(i)V_k^{t(i)-1} Y_k^{t(i)} . \quad (21)$$

In this form we get the following set of normal equations:

$$\sum_j \sum_k A_j[V_k^{t(j)+t(i)} Y_k^{t(j)+t(i)} - t(j)t(i)V_k^{t(j)+t(i)} Y_k^{t(j)+t(i)+1}$$

$$+ t(j)t(i)(t(j) + 1)(t(i) + 1)V_k^{t(j)+t(i)} Y_k^{t(j)+t(i)+2}$$

$$+ v(j)v(i)V_k^{t(j)+t(i)-2} Y_k^{t(j)+t(i)}]$$

$$= \sum_k [\ln Q_k V_k^{t(i)} Y_k^{t(i)} + (\partial \ln Q/\partial T)_k t(i)V_k^{t(i)} Y_k^{t(i)}$$

$$+ (\partial^2 \ln Q/\partial T^2)_k t(i)(t(i) + 1)V_k^{t(i)} Y_k^{t(i)}$$

$$+ (\partial \ln Q/\partial V)_k v(i)V_k^{t(i)-1} Y_k^{t(i)}] \quad \forall i = 1,...,m . \quad (22)$$

Equation (22) defines a system of linear equations for our least squares problem. Once solved, we have a functional expression for $\ln Q$.

An alternative approach to the procedure expressed above is to apply the least squares condition separately to $\ln Q$ and its derivatives. In this case, we obtain four different error functions:

$$S_1 = \sum_k^n [\ln Q_k - \ln Q_k^c]^2$$

$$S_2 = \sum_k^n [\partial \ln Q_k/\partial T - \partial \ln Q_k^c/\partial T]^2$$

$$S_3 = \sum_k^n [\partial^2 \ln Q_k/\partial T^2 - \partial \ln^2 Q_k^c/\partial T^2]^2$$

$$S_4 = \sum_k^n [\partial \ln Q_k/\partial V - \partial \ln Q_k^c/\partial V]^2 . \quad (23)$$

In this way, we obtain higher precision than with equation (19). The disadvantage is that we need to handle four different fits.

The quality of the fit is measured by the square of the multiple correlation coefficient, R^2, and the standard error of estimate, σ. They are defined as (Draper and Smith, 1981)

$$R_f^2 = \sum_i^n (f_i^c - \langle f \rangle)^2 / \sum_i^n (f_i - \langle f \rangle)^2$$

$$\sigma_j = [\sum_i^n (f_i^c - f_i)^2/(n - m)]^{1/2} \quad (24)$$

where n represents the number of data in the fit, m the number of terms in the $\ln Q$ Taylor expansion, f represents $\ln Q$ and $\langle f \rangle$ is the arithmetic average of the data.

The second step of the refinement procedure is to enhance the canonical partition function introducing experimental information. Thus, we build an error function E

$$E(\mathbf{C}) = \sum_i^n (X_i^c(\mathbf{C}) - X_i^e)^2 \quad (25)$$

where $\mathbf{C}$ is the row vector formed by the A_i coefficients of the $\ln Q$ expansion and X_i^c and X_i^e represent the calculated and experimental values of a thermodynamical magnitude, respectively. We wish to find the value of $\mathbf{C}$ that minimizes $E(\mathbf{C})$. Thus, we expand $E(\mathbf{C})$ in a Taylor series on the vicinity $\mathbf{C}_{n+1}$ of a given point $\mathbf{C}_n$. Considering a quadratic approximation to $E(\mathbf{C})$ and deriving with respect to $\mathbf{C}$ we get

$$\nabla E(\mathbf{C}_{n+1}) = \nabla E(\mathbf{C}_n) + H(\mathbf{C}_n) \cdot q(\mathbf{C}_n) \quad (26)$$

where $\nabla E(\mathbf{C})$ is the gradient of $E(\mathbf{C})$ with respect to $\mathbf{C}$, $H(\mathbf{C}_n)$ is the first derivative of the gradient, i.e. the hessian of $E(\mathbf{C})$, and q is defined as $\mathbf{C}_{n+1} - \mathbf{C}_n$. Assuming $\mathbf{C}_{n+1}$ to be a minimum we obtain:

$$\mathbf{C}_{n+1} = \mathbf{C}_n - H(\mathbf{C}_n)^{-1} \cdot \nabla E(\mathbf{C}_n) . \quad (27)$$

This expression is the basis of our refinement procedure. As usual, the search of the optimal $\mathbf{C}$ is enhanced using a parameter α that gives the minimum of the function along the search direction:

$$\mathbf{C}_{n+1} = \mathbf{C}_n - \alpha \cdot H(\mathbf{C}_n)^{-1} \cdot \nabla E(\mathbf{C}_n) . \quad (28)$$

equation (28) represents an iterative technique for the refinement of the $\mathbf{C}$ vector using the gradient and the hessian matrix defined as

$$\nabla_i E = (\partial E(\mathbf{C})/\partial A_i)$$

$$H_{ij} = (\partial^2 E(\mathbf{C})/\partial A_i \partial A_j) . \quad (29)$$

Rather than using the exact form for the hessian matrix, the BFGS updating formula for H^{-1} (Broyden, 1970; Fletcher, 1970; Goldfard, 1970; Shanno, 1970) will be used

$$A_{n+1} = A_n + \left(1 + \frac{g_n \cdot A_n \cdot g_n^T}{s_n \cdot g_n^T}\right) \cdot \frac{s_n^T \cdot s_n}{s_n \cdot g_n^T}$$

$$- \frac{s_n^T \cdot g_n \cdot A_n + A_n \cdot g_n^T \cdot s_n}{s_n \cdot g_n^T} \quad (30)$$

where $A_n = H_n^{-1}$, $s_n = \mathbf{C}_{n+1} - \mathbf{C}_n$ and $g_n = \nabla E(\mathbf{C}_{n+1}) - \nabla E(\mathbf{C}_n)$. The BFGS algorithm ensures that the hessian remains positive definite on all updates. This is a desirable property because a positive definite hessian always generates a search direction that lowers the error given by equation (25). The BFGS method also removes the need of an exact line search (equation (28)). The initial hessian is chosen as a diagonal positive definite matrix where the diagonal elements are the inverse of the numerically calculated second derivatives. Since equation (25) can be computed very quickly, the first and second derivatives are numerically calculated by central differences using a given increment ΔA_i as

$$(\partial E/\partial A_i) \simeq [E(A_i + \Delta A_i) - E(A_i - \Delta A_i)]/2\Delta A_i$$

$$(\partial^2 E/\partial A_i^2) \simeq [E(A_i + 2\Delta A_i) + E(A_i - 2\Delta A_i)$$

$$- 2E(A_i)]/4\Delta A_i^2 . \quad (31)$$

3. SOFTWARE DEVELOPMENT

The calculation of partition functions and equilibrium properties is implemented in a software tool using the described methodology. Figure 1 shows the functionality of the program organized around five high-level processes. The first process reads and validates the input information that is used in the second process for generating a grid of $\ln Q$ and its derivatives. In the third process we obtain the polynomial fit of $\ln Q$. This fit can be refined using experimental data in the fourth process, which applies BFGS. When desired, specific thermodynamic functions can be obtained from the $\ln Q$ expression, either initial or refined, through the fifth process. Each process is implemented in separate modules of several software routines.

The least squares fit of $\ln Q$ on V and T is obtained by setting up the system of normal equations that are solved applying the singular value decomposition

technique, SVD (Press *et al.*, 1986). The SVD is used for dealing with sets of equations that are either singular or numerically very close to singular, which can appear in least squares problems. The refinement of the partition function, $\ln Q$, is obtained applying the BFGS optimization algorithm, which is implemented using standard routines (Press *et al.*, 1986). A Brents' parabolic interpolation method is used to perform the line search along the minimization direction.

The characteristics of Fortran 90 permit the implementation of the different data structures in a compact and efficient form. Thus, the Taylor expansion of the partition function, $\ln Q$, is implemented as a Fortran 90 structure (Brainerd *et al.*, 1990) with three fields. The first field contains the coefficients of the expansion whereas the second and third contain the powers of V and T, respectively. In the same form, a data structure, $\ln Q$calc, with fields Coef, V and T collects the A coefficients and the V and T powers of equation (14).

The direct summation of rotational–vibrational energy levels is implemented through ordered dynamic lists using pointers. A pointer is a data structure that instead of a numerical, logical or character data contains the address to a data object (Smith, 1987). Thus, a list of data can be constructed by defining a structure (called Nodo in the program) with two fields. The first field stores the value of the data, in the present case the value of the energy level, while the second is a pointer to a new structure of the same type. Figure 2 shows how the list is dynamically generated. The first node in the list is created in run time allocating a pointer to a structure of type Nodo. The first energy level is stored in the numerical field whereas a second pointer is allocated to a new structure of type Nodo (the second node). Now, the pointer field of the first node is made to point to the second node defining the first link. The second energy level is stored in the numerical field of the second node whereas a third structure of type Nodo is allocated. The point field of the second node is made to point to the third node defining the second link. The whole process is repeated as many times as needed. The process finishes by making the pointer field of the last node to point to nothing (i.e. to nullify the pointer). A list is created for each vibrational mode and is used to store the vibrational energy levels. An additional list can appear if we compute the rotational partition function using a direct summation. When overall rotation is significantly coupled to some vibrational mode and the semi-classical approach is not applicable, the rotational levels within each vibrational state are computed using equation (7). The obtained rotational energy values are corrected for the origin by adding the corresponding value of the vibrational level, and they are inserted into the list. Figure 2(b) shows how the insertion is performed using the rotational energy levels of the first vibrational state. The insertion starts by creating a second list storing the rotational levels. This list is generated in the same way as the first one. At this point, the pointer field of the first vibrational state is made to point to the first structure of the rotational list. The insertion finishes by making the last pointer field of the rotational list to point to
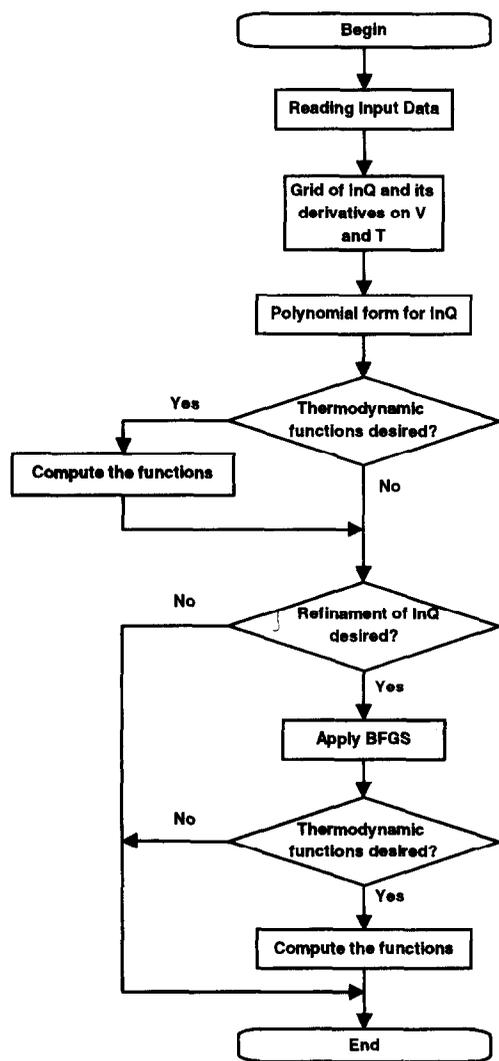

Fig. 1. Flow chart of the program for the statistical mechanical computation of equilibrium properties.
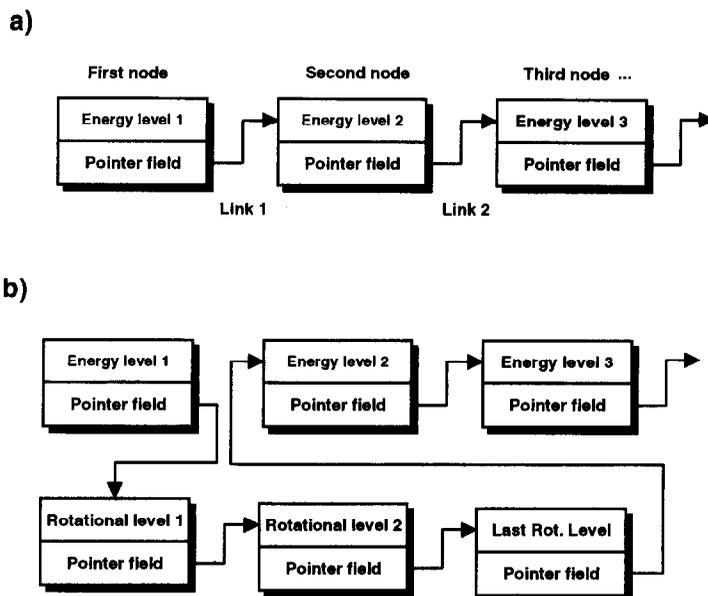
a)

b)

Fig. 2. Dynamic lists used to simulate the stacks of energy levels. Case (a): dynamic list structure used to store pure rotational or vibrational energy levels. Case (b): insertion of a list of rotational energy levels into a list of vibrational levels.

the second node of the vibrational list. The procedure is repeated for the different vibrational states. In this form, we obtain an ordered list of rotational–vibrational energy levels for each vibrational mode. Using dynamic lists, it is not necessary to know beforehand the number of energy levels. Using a simple algorithm the lists are created and the partition function, q, is computed always in the same form independently of the nature of the levels.

The data types defined for the program and the value of physical constants are encapsulated within a Fortran 90 module (Brainerd et al., 1990). For the physical constants the 1986 least squares adjustment is used (Cohen and Taylor, 1987). The module is invoked by the routines where the defined data types or the constants are used. The program has been developed and runs on a 64-bit DEC AlphaStation 200 4/166 machine on the OSF/1(UNIX) operating system.

4. APPLICATIONS

4.1. The Semi-classical Computation of Rotational Partition Functions

It is assumed generally that the semi-classical expression for the rotational partition function, equation (6), works well "for usual temperatures and molecules larger than H_2". The present program permits the quantification of this question. Thus, we compute the percentage difference between the semi-classical (equation (6)) and the exact (calculated through equation (7)) values of the rotational partition function for an asymmetric rotor. The error is calculated for a grid of temperatures ranging from 25 to 400 K. The size of the molecule is defined through the A rotational constant. We used values from 200 GHz (corresponding approximately to the water dimer) down to 20 GHz. The B and C

rotational constants are taken as 3/5 and 2/5 of A, respectively. Figure 3 shows that the error increases very quickly for small molecules and low temperatures. The error is always smaller than 1% for temperatures higher than 100 K. For medium size molecules and/or $T \simeq 273.15$–373.15 K, the error is smaller than 0.1% (shadowed region).

4.2. Performance of the Refinement Procedure

A point of interest is the ability of the refinement procedure to reproduce a set of reference data using
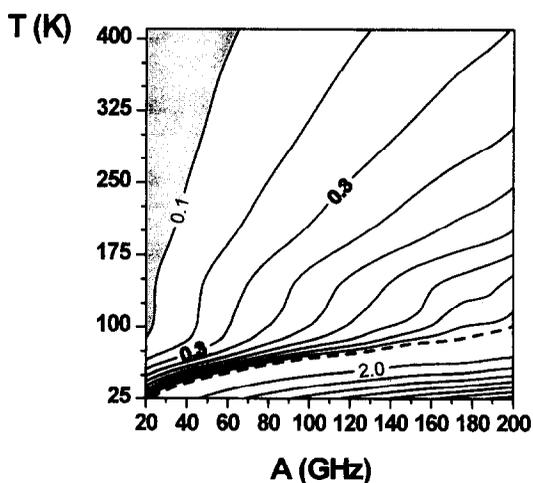

Fig. 3. Error (%) of the rotational contribution to the partition function calculated with the semi-classical expression. The error is represented as a contour map as a function of the temperature and the rotational constant A. The shaded region corresponds to an error smaller than 0.1% and the dashed line to an error of 1%. The interval between contour lines is 1% and 0.1% above and below the line of 1%, respectively.

harmonic and anharmonic models. As a test case we have selected the hydrogen fluoride dimer. The hydrogen fluoride dimer, $(HF)_2$, is one of the simplest hydrogen-bonded systems. This system is so small that it has been possible to describe totally its six vibrational degrees of freedom using different semi-empirical potentials (Quack and Suhm, 1990, 1991). In particular, the SQSBDE potential (Quack and Suhm, 1991) permits the accurate calculation of $\ln Q$ at different temperatures. These data will be used in the present work to test the refinement procedure. The present study could be of interest for dealing with molecules larger than $(HF)_2$, where a whole description of the system is not feasible at all.

In a hydrogen-bonded complex of the type X • • •H–Y, one of the highest contributions to the partition function arises from the anharmonic X • • •Y stretching mode of the hydrogen bond. The motion along this mode is directly related to the dissociation of the complex. Two useful approaches can be used for modelling this anharmonic mode—a Taylor expansion of the potential on the vibrational coordinate or a Morse potential form. The rotational, vibrational and structural data needed for our study can be obtained from *ab initio* calculations. To describe properly the hydrogen bond, polarization and diffuse functions are introduced in the basis set (Scheiner, 1991). Correction for correlation energy is accounted for at the all-electrons MP2 level. Thus, a triple split MP2(full)/6-311 + + G(2df,2pd) basis set was selected. All the calculations were performed with the GAUSSIAN 94 package (Frisch *et al.*, 1995).

A full molecular geometry optimization followed by a normal modes analysis is performed on the hydrogen fluoride dimer. Owing to the existence of soft intermolecular degrees of freedom, the optimization is performed using the "Tight" option. Thus, the limits in the optimization procedure are reduced below the standard values. The optimized geometry yields rotational constants of 6.72, 6.78 and 724.33 GHz. The harmonic frequencies are found at 155.9, 223.0, 505.0, 576.6, 4063.2 and 4146.7 cm^{-1}. The first frequency corresponds to the intermolecular F• • •F stretching. To study this stretching mode, we need to identify the corresponding vibrational coordinate. This goal can be achieved by decomposing the normal modes in internal coordinates. Thus, the hessian matrix obtained in the calculation of frequencies is introduced in the GAMESS program (Schmidt *et al.*, 1993), where a useful algorithm for normal modes decomposition is implemented (Boatz and Gordon, 1989). The results show that in $(HF)_2$ the F• • •F stretching is mainly described by the F–F separation. Thus, we select the increments from the equilibrium position, R_{FF}, as the vibrational coordinate.

To generate the different potential forms, we obtain a grid of points on the F • • •F distance from 1.941 to 3.941 Å in increments of 0.2 Å. The geometry is fully optimized at each point and the total energy results are fitted to the appropriate potential forms. For the Taylor series expansion we obtain

$$V = 4142.15 \cdot R_{FF}^2 - 7043.68 \cdot R_{FF}^3$$
$$+ 5700.04 \cdot R_{FF}^4 - 1829.19 \cdot R_{FF}^5 \quad (32)$$

with correlation $R = 1.00000$ and standard deviation $\sigma = 9.17$ cm^{-1}.

In a similar way, the Morse form is obtained by fitting the energy data to a Morse potential. After a non-linear least squares fit we obtain

$$V = 1555.94[1 - \exp(-1.5415 R_{FF})]^2 \quad (33)$$

with correlation $R = 0.99963$. The dissociation energy computed from the previous potential is $D_e = 18.6$ kJ mol^{-1}. This value is in good agreement with the 18.9 kJ mol^{-1} experimental estimate (Quack and Suhm, 1996).

To determine the effect of the previous potential forms on the partition function, it is necessary to obtain the vibrational energy levels. The kinetic term, B_{FF}, is obtained from the rotational–vibrational G matrix (Pickett, 1972; Harthcock and Laane, 1985; Niño and Muñoz-Caro, 1994). B_{FF} is computed for the fully optimized geometries at each point of the grid on the F• • •F distance. It is found that the variation from the equilibrium value is only 0.002 cm^{-1}. Thus, the equilibrium result, 1.686 cm^{-1}, is selected for all the calculations. For the potential of equation (32), the Schrödinger equation is variationally solved in the harmonic oscillator basis. In the case of the Morse potential the analytic expression for the energy levels is applied (Pauling and Wilson, 1935) using equation (33) and the previous result for the B_{FF} term. Thus, the energy levels can be obtained as

$$v(\text{cm}^{-1}) = 2\sqrt{D_e a^2 B}(v + 1/2)$$
$$- (a^2 B/2)(v + 1/2)^2 \quad (34)$$

where v represents the vibrational quantum number, B is the kinetic term, D_e the dissociation energy and a is the exponential factor in the Morse potential. Table 1 collects the energy levels obtained from the two potential forms. It can be observed that the Morse potential exhibits its usual behaviour with a separation of energy levels decreasing with the quantum number. In contrast, the Taylor series yields energy levels with a separation changing in an erratic

Table 1. Vibrational energy levels (in cm^{-1}) for the F• • •F stretching mode of $(HF)_2$

v	E (case a)*	E (case b)†
0	0.0	0.0
1	149.9	155.9
2	291.8	300.7
3	425.7	338.7
4	551.6	435.1
5	669.4	560.4
6	779.2	677.9
7	881.0	788.4
8	974.8	889.8
9	1060.6	906.1
10	1138.4	985.8
11	1208.2	1067.6
12	1270.0	1150.4
13	1323.7	1244.4
14	1369.4	1348.1
15	1407.2	1459.9

*Morse potential, equation (33).
†Potential expressed as a Taylor series, equation (32).

Table 2. Least squares fits, $\ln Q = A + B/T + C/T^3$, for the canonical ensemble partition function. Correlation and standard deviation, as a percentage of the mean, are included

Term	$\ln Q$ (case a)*	$\ln Q$ (case b)†	$\ln Q$ (case c)‡	$\ln Q$ (case d)§
A	0.21224×10^{26}	0.21385×10^{26}	0.21474×10^{26}	0.223552×10^{26}
B	-0.59766×10^{28}	-0.60196×10^{28}	-0.60342×10^{28}	-0.593530×10^{28}
C	0.22860×10^{32}	0.23581×10^{32}	0.23439×10^{32}	0.232727×10^{32}
R	1.00000	1.00000	1.00000	1.00000
σ (% mean)	0.04326	0.04384	0.04231	0.00489

*Results obtained using harmonic frequencies.
†Results obtained using harmonic frequencies and the energy levels of case a in Table 1.
‡Results obtained using harmonic frequencies and the energy levels of case b in Table 1.
§From a fit of the data of Quack and Suhm (1991). See text.

way. This can be attributed to the wrong behaviour of the polynomial form in the limits ($\pm\ \infty$ for the F•••F distance).

After calculation of the energy levels, we have computed the partition function through equation (23) in four cases for a temperature range of 0–100°C and a constant pressure of 1 atm. The first case corresponds to the harmonic frequencies. The second case is obtained using the energy levels calculated from the Morse potential (case a of Table 1). The third case uses the energy levels (case b of Table 1), obtained from the Taylor expansion of the F•••F potential. The last case is the reference. It is obtained from the accurate molecular partition function values, q_{evr} (including only electronic, vibrational and rotational contributions) collected in Table XII of Quack and Suhm (1991). To compute the canonical $\ln Q$ for the last case, we have used the q_{evr} values for $T = 270, 280, 290, 300$ and 350 K. As zero-point energy we used the 4591 cm^{-1} result of Table VI of Quack and Suhm (1991) obtained from the six-dimensional SQSBDE potential. The translational partition function was calculated using a molecular mass of 40.0125 amu. Table 2 collects the four functional forms for $\ln Q$. It is observed that the two anharmonic models yield similar results. These results are closer to the reference than the harmonic case. In all cases, the principal difference from the reference appears in the independent term, A. That means that the $\ln Q$-curves for the three cases are approximately parallel to the reference.

Finally, the $\ln Q$ functions shown as cases a, b and c in Table 2 are refined using the five reference values of $\ln Q$ obtained from Quack and Suhm (1991). After the refinement, the harmonic and the two anharmonic cases yield exactly the same result. We obtain a functional form that equals case d of Table 2. It is remarkable that three different models yield the same result even with the same error (a difference of 0.22238×10^{12} between calculated and reference $\ln Q$ data).

5. CONCLUSIONS

A program for computing equilibrium properties is presented. The program is useful for dealing with flexible molecular systems since it incorporates semi-rigid and anharmonic models. In addition, a semi-empirical procedure is developed for the refinement of partition functions using experimental data. The program permits quantitative analysis of the accuracy of different approaches for the computation of the rotational contribution to the

partition function. Using an asymmetric rotor, we found that the usual semi-direct expression yields an error smaller than 0.1% for standard temperature and normal size molecules. In all cases, the error is found smaller than 1% for temperatures higher than 100 K. Analysing the hydrogen-bond stretching mode of (HF)$_2$, we found that a Taylor expansion of the potential does not reproduce correctly the stack of energy levels. In contrast, a Morse potential reproduces accurately the dissociation energy of (HF)$_2$. Computation of the partition functions with both models shows that the Taylor expansion overestimates the result. Using the accurate data available for the partition function of (HF)$_2$, we investigate the performance of the refinement procedure. We find that the procedure reproduces exactly the reference data, even starting with a pure harmonic approximation to the vibrational modes. This result shows that, using reliable experimental information, it could be possible to study accurately equilibrium properties of molecular systems starting with simple theoretical models. Thus, using the available experimental information, we can refine the partition function, and use it to compute thermodynamic magnitudes not easily amenable to experiment. From a practical point of view this fact means that low *ab initio* levels of theory can be used when modelling the thermodynamics of flexible molecular systems without deviating too far from the harmonic approach. An open question concerns the usefulness of this refinement procedure for reproducing data in the liquid phase.

6. PROGRAM AVAILABILITY

The program is available as PARTI from QCPE (Niño and Muñoz-Caro, 1997). Included in the distribution are a manual and several examples.

REFERENCES

Boatz, J. A. and Gordon, M. S. (1989) *Journal of Physical Chemistry* **93**, 1819.
Brainerd, W. S., Goldberg, C. H. and Adams, J. C. (1990) *Programmer's Guide to Fortran* 90. Intertex Publications, McGraw-Hill Book Company, New York.
Broyden, C. G. (1970) *Journal of the Institute of Mathematical Applications* **6**, 76.
Cohen, E. R. and Taylor, B. N. (1987) *Physics Today* **40**, 11.

Draper, N. R. and Smith, H. (1981) *Applied Regression Analysis*, 2nd edn. John Wiley & Sons, New York.

Fletcher, R. (1970) *Computer Journal* **13**, 317.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T. A., 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. and Pople, J. A. (1995) *Gaussian* 94 (*Revision B*.3). Gaussian, Inc., Pittsburgh, PA.

Goldfard, D. (1970) *Mathematics and Computers* **24**, 23.

Harthcock, M. A. and Laane, J. (1985) *Journal of Physical Chemistry* **89**, 4231.

Levine, I. N. (1970) *Quantum Chemistry*, Vol. II, chap. 4. Allyn & Bacon, Boston.

Lucas, K. (1991) *Applied Statistical Thermodynamics*. Springer, Berlin.

McQuarrie, D. A. (1973) *Statistical Mechanics*. Harper & Row, New York.

Niño, A. and Muñoz-Caro, C. (1994) *Computers and Chemistry* **18**, 27–32.

Niño, A. and Muñoz-Caro, C. (1995) *Computers and Chemistry* **19**, 371–378.

Niño, A. and Muñoz-Caro, C. (1997) Quantum Chemistry Program Exchange, Vol. 17 (in press).

Niño, A., Muñoz-Caro, C. and Moule, D. C. (1995) *Journal of Physical Chemistry* **99**, 8510.

Pauling, L. and Wilson, E. B. Jr. (1935) *Introduction to Quantum Mechanics*. McGraw-Hill Book Company, Inc., New York and London.

Pickett, H. M. (1972) *Journal of Chemical Physics* **56**, 1715.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986) *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Quack, M. and Suhm, M. A. (1990) *Molecular Physics* **69**, 791.

Quack, M. and Suhm, M. A. (1991) *Journal of Chemical Physics* **95**, 28.

Quack, M. and Suhm, M. A. (1996) *Theoretica Chimica Acta* **93**, 61.

Richards, W. G. (1983) *Quantum Pharmacology*, chap. 15. Butterworths, London.

Scheiner, S. (1991) In *Reviews in Computational Chemistry*, eds K. B. Lipkowitz and D. B. Boyd, Vol. II, chap. 5. VCH Publishers, Inc., New York.

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Theresa, L. W., Dupuis, M. and Montgomery, J. A. Jr. (1993) *Journal of Computational Chemistry* **14**, 134.

Shanno, D. F. (1970) *Mathematics and Computers* **24**, 647.

Smith, H. F. (1987) *Data Structures*. Harcourt Brace Jovanovich, London.

Wang, S. C. (1929) *Physical Review* **34**, 243.

Weinstein, H. and Green, J. P. (1981) Molecular conformation and structure–activity relationships. In *Annals of the New York Academy of Sciences. Quantum Chemistry in Biomedical Sciences*, eds H. Weinstein and J. P. Green, Vol. 367, Section 3. The New York Academy of Sciences, New York.