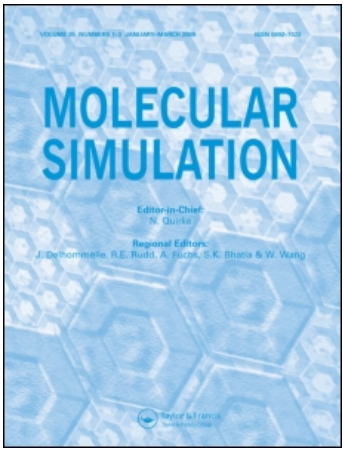

Molecular Simulation

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

Molecular Dynamics Simulations of some Small Organic Molecules

Pieter F. W. Stouten[ab]; Bouke P. Van Eijck[a]
[a] Laboratorium voor Kristal-en Structuurchemie, Rijksuniversiteit, Utrecht, The Netherlands [b] Biocomputing Group, European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany

To cite this Article Stouten, Pieter F. W. and Van Eijck, Bouke P.(1989) 'Molecular Dynamics Simulations of some Small Organic Molecules', Molecular Simulation, 4: 4, 193 — 207
To link to this Article: DOI: 10.1080/08927028908022363
URL: http://dx.doi.org/10.1080/08927028908022363

Molecular Simulation, 1989, Vol. 4, pp. 193–207
Reprints available directly from the publisher
Photocopying permitted by license only

MOLECULAR DYNAMICS SIMULATIONS OF SOME SMALL ORGANIC MOLECULES: PROBLEMS AND RESULTS OF FREE ENERGY CALCULATIONS FOR CONFORMATIONAL TRANSITIONS OF RIBOSE, MALIC ACID AND TARTARIC ACIDS

PIETER F. W. STOUTEN[1] and BOUKE P. VAN EIJCK

Laboratorium voor Kristal- en Structuurchemie, Rijksuniversiteit, Padualaan 8, 3584 CH Utrecht, The Netherlands.

(*Received April 6, 1989, accepted July 5, 1989*)

Free energy differences between different conformers of D-ribofuranose, L-malic acid and *meso*-tartaric acid in solution were calculated using Molecular Dynamics simulations. In case of ribose the $\alpha \rightarrow \beta$ transition was studied. For the acids attention was focussed on the transitions between the three possible staggered conformers with respect to the central C–C bond. In all cases a thermodynamic integration method was employed to evaluate the free energy difference. The use of an alternative technique, umbrella sampling, for ribose did not give promising results.

It was shown that one needs a fairly accurate picture of the accessible conformational space in case of flexible molecules like the ones considered here before one can determine meaningful free energy differences. Large hysteresis effects between forward and reverse simulated transitions were observed, but contrary to the general belief they are no direct measure of the accuracy of the calculated ΔG values. In all cases the ΔG values resulting from the simulations and from NMR experiments agree within the, considerable, error limits and for the different forms of D-ribose, L-malic acid and L-tartaric acid the relative order of their populations is also correctly reproduced.

KEY WORDS: Free energy, conformational transitions, ribose, malic acid, tartaric acid, molecular dynamics.

INTRODUCTION

Detailed knowledge of the conformational behaviour of biologically active molecules is necessary to assess the role they play in biochemical processes. Molecular Mechanics (MM) calculations, which in essence are energy minimizations of the isolated molecule, can provide relevant information [1]. The major disadvantage of the method is that no solvent can be included explicitly, because one would be left with a system containing a huge number of local energy minima in vast conformational space and one would never be sure whether the system under investigation had entered a real low energy minimum. And even if this were the case it is most likely that many other relevant minima exist which will never be visited. Molecular Dynamics (MD) simulations remedy this problem because of the introduction of kinetic energy which allows the system to climb barriers in the order of kT [2,3]. Another advantage

1) Present address: European Molecular Biology Laboratory, Biocomputing Group, Meyerhofstrasse 1, 6900 Heidelberg, Federal Republic of Germany.

of this method is that it yields information on the dynamic properties of the system. In addition it is also possible to calculate thermodynamic properties because a set of representative configurations is generated during the simulations. In the past MD techniques were mainly used to produce energetic and geometrical data. One of the most interesting properties, however, is not the energy but the free energy and especially when we are dealing with flexible molecules entropy effects can play a very important role [4]. Overviews of methods for the calculation of free energies have been given by van Gunsteren [5] and Straatsma *et al.*[6, 7]. Beveridge *et al.* have presented several cases where free energy calculations can be applied and be useful [8].

In the present study we are mainly interested in the aspects of free energy calculations for conformational transitions of small molecules in solution. The procedure, including the determination of the best possible settings of the simulation protocol, the problems that were encountered, and detailed results are reported for two different classes of molecules. Firstly we deal with L-malic acid (whose di-anion plays an important role in the citric acid cycle [9]) together with two other closely related but more rigid four-carbon α-hydroxy carboxylic acids: L-tartaric acid and meso-tartaric acid. Previously we have studied the conformational behaviour of L-malic acid by NMR and normal (non-ΔG) MD simulations [10], and of all three α-hydroxy carboxylic acids in more detail by MM and MD methods [4]. Now our attention is devoted to calculating free energy differences between the different rotamers of these acids with respect to the central C–C bond. Secondly we present the result of an investigation of the conformational aspects of the furanose form of D-ribose, a major constituent of RNA. The six-membered pyranose form is predominant in solution, but we chose to study the five-membered ring, because that is the form occurring in RNA. Recently we have reported non-ΔG MD simulations of ribofuranose [11] and here the results of two attempts to calculate free energy differences for the transition between its α and β forms are reported.

2. CALCULATION OF FREE ENERGY DIFFERENCES

Free energies can be calculated using MD techniques. For example the free energy of solvation has been determined quite accurately for noble gases in water [12,13]. Free energy differences for non-physical processes, like changing one substituent to another for a molecule attached to a substrate and the same molecule in solution, have been estimated in order to assess relative binding strengths [14,15].

We are interested in the relative populations of different forms of the same molecule in solution and hence in the free energy differences between them. In principle it should be sufficient to simply do normal MD simulations and count the occurrence percentages of the individual forms. In principle it should be sufficient to simply do normal MD simulations and count the occurrence percentages of the individual forms. In practice often only one form is found in simulations with a limited time span: the other forms occur insufficiently or cannot be reached due to the presence of high energy barriers. There are several ways to circumvent this problem. One can force the system under consideration to surmount energy barriers very slowly by applying an extra variable potential and integrate the free energy changes related to this transition. This method is called *thermodynamic integration* [16]. Alternatively one can compensate the energy differences and barriers by introducing an additional potential in order to make greater parts of the multi-dimensional potential surface

accessible and perform counting subsequently, which is called *umbrella sampling*. This method will be dealt with in more detail in the ribose experiment section since it was only used in that case.

In order to calculate free energy differences between two states A and B of a system by the *thermodynamic integration* method, a potential term U is added to force the system to change from state A into state B. This potential is made a function of a coupling parameter λ and depends furthermore on the generalized coordinates $\mathbf{q}$. The Hamiltonian for any state λ is now characterized by:

$$H(\mathbf{p},\mathbf{q},\lambda) = H(\mathbf{p},\mathbf{q}) + U(\mathbf{q},\lambda) \tag{1}$$

The isobaric partition function for this state is:

$$\Delta(\lambda) = c \int\int\int \exp\left\{-\frac{H(\mathbf{p},\mathbf{q},\lambda) + pV}{kT}\right\} \mathrm{d}V\, \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{q} \tag{2}$$

where c is a constant. The free energy is given by:

$$G(\lambda) = -kT \ln \Delta(\lambda) \tag{3}$$

The free energy difference between states A and B is:

$$\Delta G_{\mathrm{A}\to\mathrm{B}} = \int_{\lambda_\mathrm{A}}^{\lambda_\mathrm{B}} \frac{\partial G}{\partial \lambda}(\xi)\, \mathrm{d}\xi \tag{4}$$

Equations (1), (2) and (3) give the partial derivative of G with respect to λ for state $\lambda = \xi$:

$$\frac{\partial G}{\partial \lambda}(\xi) = -\mathrm{kT}\frac{1}{\Delta(\xi)}\frac{\partial \Delta}{\partial \lambda}(\xi) = \frac{\int\int\int \frac{\partial \mathrm{U}}{\partial \lambda}(\mathbf{q},\xi)\, P\,(\mathbf{p},\mathbf{q},\xi)\, \mathrm{d}V\, d\mathbf{p}\, d\mathbf{q}}{\int\int\int P(\mathbf{p},\mathbf{q},\xi)\, \mathrm{d}V\, d\mathbf{p}\, d\mathbf{q}}$$

$$= \left\langle \frac{\partial \mathrm{U}}{\partial \lambda}(\mathrm{q},\xi) \right\rangle_\xi \tag{5}$$

where $P\ (\mathbf{p},\mathbf{q},\xi) = \exp\{-(H(\mathbf{p},\mathbf{q},\xi) + pV)/kT\}$ and $\langle \ldots\ldots \rangle_\xi$ is the ensemble average for a specific constant value ξ of λ. Combining equations (4) and (5) gives:

$$\Delta G_{\mathrm{A}\to\mathrm{B}} = \int_{\lambda_\mathrm{A}}^{\lambda_\mathrm{B}} \left\langle \frac{\partial \mathrm{U}}{\partial \lambda}(\mathbf{q},\xi) \right\rangle_\xi \mathrm{d}\xi \tag{6}$$

Equation (6) provides the means to calculate free energy differences via MD simulations. Of course it is impossible to carry out complete averaging after each infinitesimal change of λ. The assumption is that if λ is changed very slowly the system remains essentially in equilibrium. Averaging is done implicitly because a representative part of conformation space is visited during this slow change. The derivative of $U\ (\mathbf{q},\lambda)$ with respect to λ is calculated analytically at every value of λ. Refer to [16] for a more detailed description of the method and its applications.

This method is mostly applied to processes where the potential energy functions of states A and B are different. Since G is a state function these processes need not be physically possible and indeed various kinds of chemical transformations have been studied [7, 17, 18]. We are concerned with another physical impossibility, the transition between α and β forms of ribofuranose without ring opening. A conformational transition, however, is conceptually more difficult as the initial and final states have

the same potential energy function, but differ in the region of phase space being sampled. Here the λ-dependent term is only needed to create a hypothetical intermediate state B corresponding to the top of the barrier, whereas A corresponds to one of the local free energy minima. To define the free energy for such a state A means that the integration in equations (2) and the averaging in equations (5) and (6) must only be performed over that part of phase space that we wish to consider as belonging to "state A".

3. SIMULATION EXPERIMENT

General Description MD Protocol

All simulations were performed with the GROMOS computer program package [19] on the Cyber 205 of the Amsterdam Academic Computer Centre (SARA). In all cases, for the acids and the sugar, a periodic box filled with water and one solute molecule was considered. Non-bonded interactions between pairs of charge groups (electrically neutral groups) were computed within a sphere of specified radius (cut-off). A larger cut-off radius and a larger number of solvent molecules were not feasible considering the available computer resources. The list of interacting charge group pairs was updated every 5 time steps. A time step of 2 fs was employed except in one test case. All atoms were treated explicitly except CH and CH_2 groups which were regarded as united atoms with a somewhat larger radius than a bare carbon atom. The complete potential models are detailed in [10] (acids) and [11,20] (sugar). For water the rigid SPC model was used [21]. All bond lengths and the water H–H distance were constrained to preset values with the SHAKE algorithm [22,23]. The systems were loosely coupled to a pressure bath ($p \approx 1$ atm) and a temperature bath ($T \approx 298$ K) with coupling time constants of 0.5 and 0.1 ps respectively [24, 25]. The initial equilibration periods before data sampling were at least 10 ps. The differences between the acids and sugar simulation protocols are compiled in table 1.

α-Hydroxy Carboxylic Acids

For malic acid and the tartaric acids (Figures 1, 2, & 3) we are interested in the free energy differences between the three possible staggered conformations (with respect to the carbon chain) in solution. The relevant variable in these cases is the dihedral angle $\Psi = C_1 - C_2 - C_3 - C_4$ which is governed by a sinusoidal function of the form:

$$V(\Psi) = k(1 + \cos 3\Psi) \tag{7}$$

where the force constant $k = 5.86$ kJ/mol. For the calculation of the free energy differences we used the thermodynamic integration method which is an integral part of the GROMOS package [19]. The acid molecule under consideration was forced to

Table 1 Settings of the simulation protocol.

	acids	sugar
cut-off radius for non-bonded interactions	7 Å	8 Å
box type	rectangualr	cubic
pressure scaling	anisotropic	isotropic
number of water molecules	216	284

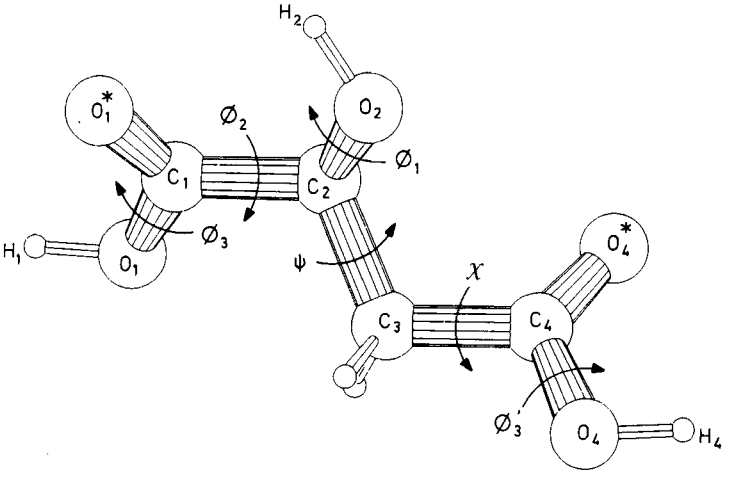

Figure 1 L-malic acid.

change from one staggered conformation to the next by applying a harmonic potential, which is defined as

$$U(\Psi,\lambda) = 1/2\ k_\lambda(\Psi - \Psi_{0,\lambda})^2 \tag{8}$$

where both the force constant k_λ and the optimum torsion angle $\Psi_{0,\lambda}$ vary in the same way with λ:

$$k_\lambda = (1 - \lambda)k_A + \lambda k_B \tag{9}$$

$$\Psi_{0,\lambda} = (1 - \lambda)\Psi_A + \lambda\Psi_B \tag{10}$$

We divided every transition into two steps. The acid molecule was moved from one energy minimum in the Ψ/E curve (state A) to the top of the inter-conformational

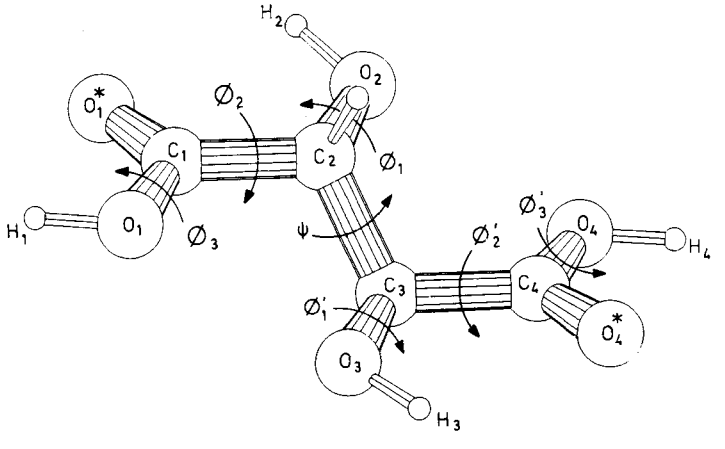

Figure 2 *meso*-tartaric acid.

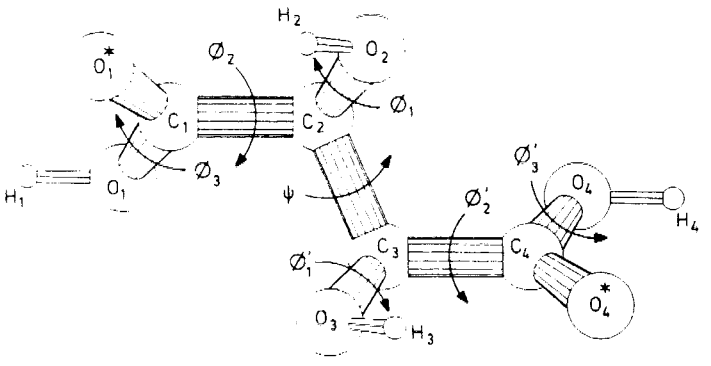

Figure 3 L-tartaric acid.

barrier (state B) and subsequently made to descend into the next potential well (state A′). Because states A and A′ are energy minima no forcing potential need be applied at the start and at the end of the simulations so the force constants k_A and $k_{A'}$ are zero. This means that no artificial potential (which is no part of the normal potential model) is present when the system is in a minimum energy state. The complete transition is depicted schematically in Figure 4 together with the relevant parameters. Uphill λ increases from 0 to 1, the well of the applied potential U becomes deeper, and its position moves from ρ to σ. Then Ψ_A is changed from ρ to τ while k_B and Ψ_B remain the same. Subsequently λ decreases from 1 to 0, the forcing potential gradually diminishes, and its minimum goes from σ to τ. With the integration method detailed above the free energy difference can be calculated now.

Only two initial problems remain to be solved: the determination of appropriate values of the force constant k_B and the speed with which λ changes ($\lambda' = \Delta\lambda/\Delta t$). L-malic acid was chosen as the test case. We wanted a low force constant k_B, but one that would indeed force the system to surmount the barriers between the conformers. $k_B = 125.5$, $k_B = 62.8$ and $k_B = 31.4$ kJ/mol were tested with a very high λ change rate ($\lambda' = 0.4\,ps^{-1}$). The first two force constants accomplished the complete tran-

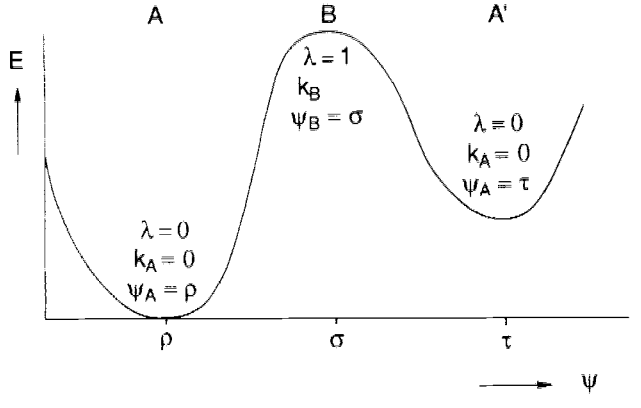

Figure 4 Schematic Ψ/E curve showing parameters for the transition A → B → A′

sition from *gauche*$^-$ to *trans* L-malic acid, the third one did not , so $k_B = 62.8$ was chosen. Subsequently several combinations of time steps and λ' values were tested for this transition. Free energy changes for subsequent λ intervals of 0.2 and the total change are detailed in Table 2. In principle we wanted to use the quickest combination whose resulting ΔG value is not significantly higher than the value delivered by slower ones. The combination with time step $\Delta t = 2$ fs and $\lambda' = 0.02\,\text{ps}^{-1}$ (second column in Table 2), corresponding with $\Delta t/\Delta\Psi = 0.83$ ps/degree, seems to fulfill this requirement and is also suitable because no irregularities in the free energy change were observed during the transition. The slowest transition shows such an irregularity at the end, which can be due to the fact that in the last part of the transition from $\lambda = 0.2$ to $\lambda = 0.0$ the *gauche*$^+$ conformation was adopted for a while too. Another, more down-to-earth, reason to reject the slowest combination and not to test even slower ones, lies in the enormous computer resources required.

Hereafter free energy differences between all conformers were calculated, starting from different equilibrated configurations and simulating forward and reverse transitions without intermediate equilibration (e.g. at the top of the barrier). The results are given in Table 3. At first sight they do not look promising: the free energy changes for the forward and reverse transitions, ΔG_F and ΔG_R respectively, should add up to zero which is far from reality as indicated by their sum (ΔG_{F+R} in Table 3). It is clear that hysteresis effects play a dominant role. This phenomenon will be discussed in more detail below. The only justifiable immediate conclusion is that the *trans* form has a lower free energy than either of the two *gauche* forms. In order to extract more accurate numerical information we first averaged ΔG_F and $-\Delta G_R$ to obtain $\Delta G'$. One might assume that the hysteresis effects in forward and reverse transitions cancel each other completely in which case the error in $\Delta G'$ is determined only by the random errors in ΔG_F and ΔG_R, leading to symmetric error bars (Approach 1). Rather conservatively one could argue that the hysteresis is another source of random error, with magnitude about $1/2 \times \Delta G_{F+R}$ (Approach 2). Regarding the hysteresis completely random again results in symmetric error bars. These two approaches lead to the two sets of error margins given in Table 3.

Table 2 ΔG (kJ/mol) of the L-malic acid *gauche*$^-$ → *trans* transition at several λ' rates: total difference and divided in λ intervals of 0.2.

$\lambda' = \Delta\lambda/\Delta t$ (ps^{-1})		0.01	0.02	0.05	0.1
time step Δt (fs)		1	2	2	2
number of MD steps / 1000		200	50	20	10
ΔG total up and down		−3.39	−2.60	+1.42	+1.71
ΔG: λ =	0.0 − 0.2	0.39	0.31	0.38	0.32
	0.2 − 0.4	1.53	1.15	1.52	1.73
	0.4 − 0.6	3.42	3.46	3.26	2.99
	0.6 − 0.8	5.76	4.98	5.11	5.32
	0.8 − 1.0	3.33	4.91	2.57	3.29
	total up	14.43	14.81	12.84	13.65
ΔG: λ =	1.0 − 0.8	−6.45	−7.13	−1.56	−3.16
	0.8 − 0.6	−5.20	−4.60	−4.70	−3.81
	0.6 − 0.4	−2.20	−3.58	−3.39	−3.32
	0.4 − 0.2	−1.46	−1.81	−1.29	−0.96
	0.2 − 0.0	−2.51	−0.29	−0.48	−0.69
	total down	−17.82	−17.41	−11.42	−11.94

Table 3 Free energy differences (kJ/mol) for all L-malic acid transitions with estimated standard deviations in parentheses

		gauche$^-$ → *trans*	*trans* → *gauche*$^+$	*gauche*$^+$ → *gauche*$^-$
Individual runs	ΔG_F	+1.9	+9.3	+5.7
		−8.8	+8.9	+2.0
			+7.2	
	ΔG_R	+14.8	−4.6	+8.5
		+2.7	+1.4	+9.7
Averages	ΔG_F	−3.4	+8.5	+3.8
	ΔG_R	+8.8	−1.6	+9.1
	ΔG_{F+R}	+5.3	+6.9	+12.9
Approach 1	$\Delta G'$	−6.1 (4.0)	+5.0 (1.6)	−2.7 (1.0)
	$\Delta G''$	−3.0 (1.7)	+5.5 (1.5)	−2.5 (1.0)
Approach 2	$\Delta G'$	−6.1 (4.8)	+5.0 (3.8)	−2.7 (6.6)
	$\Delta G''$	−5.0 (4.1)	+5.7 (3.4)	−0.7 (4.5)
NMR	ΔG	−2.0 (0.6)	+2.4 (0.6)	−0.4 (0.8)

$\Delta G'$ values were determined for all three transitions, but only two of them are independent because ΔG for the complete cycle *gauche*$^-$ → *trans* → *gauche*$^+$ → *gauche*$^-$ is zero. This condition can be used to correct the calculated $\Delta G'$ values. The uncorrected sum is −3.7 kJ/mol for both sets of error margins. Improved values, $\Delta G''$, which satisfy the zero sum criterion were obtained in the following way. Taking two of the three $\Delta G''$ values as independent variables the weighted least squares formalism was used to express these quantities into the three $\Delta G'$ values. The errors in $\Delta G''$ were estimated by means of the laws for error propagation. The results for the two sets of error margins in $\Delta G'$ are presented in Table 3 together with ΔG values derived from populations determined with NMR experiments [4], because that is the only experimental information that can be compared directly with our simulations.

The *gauche*$^-$ and *gauche*$^+$ conformers of *meso*-tartaric acid are mirror images so ΔG's for the two possible transitions from *trans* to *gauche* are the same. Both transitions were simulated and the results for the forward and reverse simulations are given in table 4. Again a large hysteresis effect is observed. In the same way as for malic acid, $\Delta G'$ and the lower and upper bounds of its variance (Approaches 1 and 2 respectively) were calculated. As a test, simulations of the *gauche*$^-$ → *gauche*$^+$ transition, for which ΔG is zero, were also carried out. In one case the barrier was descended at the same side as it was ascended and for the *gauche*$^+$ → *gauche*$^+$ "transition" $\Delta G = +0.02$ was calculated. Forward and reverse simulations of the *gauche*$^-$ → *gauche*$^+$ transition gave $\Delta G_F = +27.0$ and $\Delta G_R = +30.3$, resulting in $\Delta G' = -1.6$. This value is not very far from zero, but the hysteresis effect ΔG_{F+R} of 57.3 kJ/mol is enormous.

MM calculations and non-ΔG MD simulations showed that the *gauche*$^+$ rotamer of L-tartaric acid does not represent a stable conformation [4, 26] and therefore only the *gauche*$^-$ → *trans* transition via the shortest route, over the lowest barrier was considered. One forward and one reverse simulation were carried out and the result is detailed in Table 4 too. Again $\Delta G'$ was calculated by subtracting $1/2 \times \Delta G_{F+R}$ from ΔG_F, but in this case the only contribution to the estimated standard deviation is the hysteresis effect.

Looking at Tables 3 and 4 it is striking how good the numerical agreement between experiment and simulation is, even in this case where small free energy differences are involved and where the NMR results heavily depend on the interpretation model [10].

Table 4 Free energy differences (kJ/mol) for the tartaric acids transitions with estimated standard deviations in parentheses.

		meso-tartaric acid *gauche* → trans	L-tartaric acid *gauche*$^-$ → *trans*
Individual runs	ΔG_F	+7.1	+1.7
		+2.9	
	ΔG_R	+4.2	+15.5
		+8.3	
Averages	ΔG_F	+5.0	+1.7
	ΔG_R	+6.3	+15.5
	ΔG_{F+R}	+11.2	+17.2
Approach 1	$\Delta G'$	−0.6 (1.5)	
Approach 2	$\Delta G'$	−0.6 (5.8)	−6.9 (8.6)
NMR	ΔG	+2.4 (1.8)[1]	−7.3
	or ΔG	+7.0 (1.7)[1]	

[1]: On the basis of the NMR data is it impossible to discriminate between these two sets.

From Table 3 one can conclude that the $\Delta G''$ values calculated according to the two sets of error estimates, respectively, show approximately the same agreement with the NMR based ΔG values, indicating that the truth lies somewhere in between.

α/β-D-Ribofuranose

For D-ribofuranose (Figure 5) the variable of interest is the improper dihedral angle $\Psi = C_1\text{-}C_2\text{-}O_1\text{-}O_4$. In general harmonic improper dihedral potential functions of the form

$$V(\Psi) = 1/2\, k_\Psi(\Psi - \Psi_0)^2 \tag{11}$$

where k_Ψ is the Force constant and Ψ_0 the reference improper dihedral, are used to keep groups planar or to fix enantiomers. $\Psi_0 = -35°$ and $\Psi_0 = +35°$ respectively

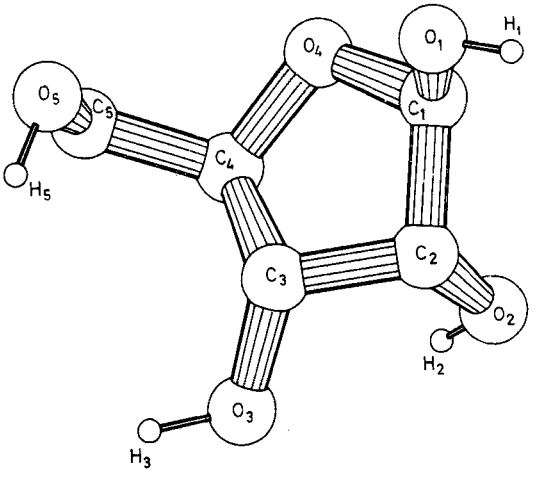

Figure 5 D-ribofuranose

determine the α and β enantiomers of ribose and normally $k_\Psi = 335\,\text{kJ/(mol.radian}^2\text{)}$ is used. In reality the transition of α to β-ribofuranose involves opening of the ring, but computationally it is much easier to determine free energy differences by means of a non-physical process, where the O_1-H_1 hydroxyl group is allowed to move from one enantiomeric position to the other. Of course this implies simultaneous movement of the hydrogen atom attached to C_1 in the opposite direction, but since aliphatic hydrogens are not explicitly described in GROMOS this poses no additional problem.

An equilibrated configuration of β-D-ribofuranose in water had been obtained in earlier work [11]. The reference angle Ψ_0 was transformed from $+35°$ (β) to $-35°$ (α) in 50 ps ($\lambda' = 0.02\,\text{ps}^{-1}$) with constant k_Ψ and the free energy difference calculated with the *thermodynamic integration* method. The α-form thus obtained was equilibrated for 10 ps, after which the reverse process was simulated. Here a complication arose due to the occurrence of transitions about $C_4 - C_5$, the net result being a change from the α-*gg* to the β-*gt* conformation. To complete the cycle the process *gg* → *gt* in the β-form was simulated using the same method as described for the acids, again with $k_\Psi = 62.8\,\text{kJ/mol}$. The results are given in Table 5. The violation of the sum rule amounts to 2.2 kJ/mol, comparable to what was found for L-malic acid. However, for the cycle described here no reverse simulations were performed. Experimentally the free energy difference $\Delta G_{\beta\to\alpha}$ for the furanose form can be estimated from the tautomeric equilibrium at 30°C [27] to be 1.5 kJ/mol. Considering that the latter number relates to a mixture of the *gg*, *gt* and *tg* forms and that the calculated ΔG values are uncertain by about 4 kJ/mol we can only state that the sign and the order of magnitude are correctly reproduced.

As a possible alternative to *thermodynamic integration* we considered the use of *umbrella sampling* [28, 29, 30]. Here an artifical potential energy term U is added, which is meant to diminish differences in energy so as to approximately equilize the times that the solute (or system) spends in the forms of interest. For a one-dimensional problem with coordinate Ψ the modified distribution $P'(\Psi)$ can then be found by counting and the original distribution can be reconstructed from

$$P(\Psi) = cP'(\Psi) \exp\,(U(\Psi)/kT) \tag{12}$$

where c is a normalization constant. The problem is of course to find a working umbrella function $U(\Psi)$. It must be constructed by trial and error and the problem may only be solvable by dividing the Ψ region into smaller parts ("windows") and combining the results afterwards [31, 32]. For ribose we chose

$$U(\Psi) = -1/2\,k_\Psi\,(\Psi - \Psi_0)^2 + 1/2\,k_1\,\Psi^2 - k_2 \cos\Psi \tag{13}$$

so that the standard improper dihedral potential is replaced by one that has its minimum at $\Psi = 0°$. Thus we stimulate in essence a hypothetical molecule where the O_1 atom is in the plane of the furanose ring and derive the α/β population ratio from the excursions of the solute into regions around $\Psi = -35°$ and $\Psi = +35°$. It is essential that these regions are sampled adequately and so k_1 should not be taken too

Table 5 Free energies (kJ/mol) for the ribofuranose transitions

$\beta_{gg} \to \alpha_{gg}$	+2.7
$\alpha_{gg} \to \beta_{gt}$	−5.9
$\beta_{gg} \to \beta_{gt}$	−1.0

large. On the other hand, when $k_1 = k_2 = 0$ the solute got stuck in either the α or the β-region. Promising results were obtained for $k_1 = k_2 = 42$ kJ/mol. During 12 ps a more or less Gaussian distribution of Ψ was maintained with extreme values around $-40°$ and $+40°$ and a slight preference for the α-form. But after that the picture changed dramatically: the sugar molecule passed into the β-form and remained there for a further 12 ps. It is our opinion that this application of *umbrella sampling*, attractive though the method may seem due to its conceptual simplicity, would probably be at least as costly in computer time as *thermodynamic integration*, without any guarantee for more reliable results.

4. DISCUSSION

Hysteresis

As has also been observed by other authors [17, 18, 33] hysteresis effects can play an important role when determining free energy differences, especially when conformational changes are involved. ΔG can be calculated very accurately when only small changes of the solute are involved that do not act as a major perturbation on the surrounding solvent. For example the solvation of noble gases, the transformation of Cl^- into Br^-, or the replacement of substituents as in the transition $CH_3OH \rightarrow CH_3CH_3$ are regarded as small changes, although Singh *et al.* regard the latter as a *large* change [17]. In our case the changes are much larger because for the acids complete rotations of the glycolic acid moieties are involved. Anderson *et al.* observed that the hysteresis (ΔG_{F+R}) for the conformational transition of a dipeptide in vacuo is only 0.3 kJ/mol while the same simulation in solution gives several kJ/mol [33], so the hysteresis can be attributed almost completely to solvent effects, meaning that water does not adapt fast enough to the perturbation (or alternatively, hinders a smooth change of Ψ) to regard the system as being in thermodynamical equilibrium all the time.

A solution can be to employ even slower transition rates. It is not sure, however, that the result will markedly improve (see Table 2), but it is sure that more computer time is required. Alternatively, since the uphill trajectories turn out to give similar results for different transition rates (Table 2) one may consider to employ lower transition rates for the downhill part only. Or one can use lower rates for the parts of the transition where the free energy changes are the largest, around $\lambda = 0.5$ for both the uphill and downhill trajectories. A fourth solution may be to increase the force constant of the forcing potential, although this increases the errors, as the difference between two larger numbers is taken. The cases where high inter-conformational barriers are crossed (e.g. *gauche*$^-$ *gauche*$^+$ of L-malic acid and *meso*-tartaric acid and *trans* $\rightarrow$ *gauche*$^-$ of L-tartaric acid [4]) exhibit the largest hysteresis. If a higher force constant is used the climbing of the barrier would be facilitated. This would imply that at all times the variable of interest Ψ is closer to the imposed value Ψ_0 and fluctuates less. No sudden jumps of Ψ would occur and the water molecules would be able to adapt easier, resulting in a transition where the condition of thermodynamic equilibrium is approximately met. In retrospect our initial idea to use the lowest possible force constant may not have been so good: it may make the absolute values of ΔG for the uphill and downhill parts relatively small and the corresponding sum rather precise, but at the cost of considerable hysteresis.

The question now is whether the hysteresis effect in itself introduces large uncertain-

ties in the calculated ΔG values. Certainly for the greater part the hysteresis is systematic i.e. the effect is approximately the same for forward and reverse simulations. The problem is to separate the random and systematic contributions. The *gauche*$^-$ → *gauche*$^+$ transition of *meso*-tartaric acid provides a perfect test case because ΔG is zero by definition, but is calculated as -1.6kJ.mol. The standard deviation, if estimated on the basis of hysteresis, of 28.6 kJ/mol is a clear exaggeration. How to deal with this problem in general cases where ΔG is not known *a priori* is not obvious. When one also considers the fact that the smallest hysteresis does not implicate the smallest standard deviation (see Table 3), it becomes clear that it is not correct to take the hysteresis effect as a direct measure for the uncertainty in the resulting ΔG values.

Conformational Freedom

Other major problems apart from hysteresis are related to conformational freedom. The best situation would be 1) high barriers (i.e. > kT) between the energy minima in the curve of the variable of interest Ψ and 2) a very flat or very restricted surface for all other conformational degrees of freedom. Condition 2) must be satisfied because in order to arrive at reliable free energy differences between two conformations all relevant configurational space must have been sampled during the transition. For both the acids and the sugar rotational transitions of the hydroxyl groups were observed frequently, so this is no cause for concern. But energy minimizations of L-malic acid *in vacuo* showed that for one conformer there is one minimum with respect to the dihedral $\chi = C_2\text{-}C_3\text{-}C_4\text{-}O_4$*, for the others there are two [4]. When the barrier between two χ minima is high and only one of them is visited, one calculates the free energy difference between two Ψ conformers given their respective sampled χ ranges. In the above case of malic acid this would roughly make a difference of RT ln 2 in the estimated free energy. In the solvated case the χ/E curves seem flatter and both non-ΔG [4] and ΔG MD simulations indicate there is only one more or less wide region of χ-space accessible in all cases. So here no problem exists but in other, similar cases one must be alert. The tartaric acids have two rather rigid, approximately planar, glycolic acid moieties [4], leaving $\Psi = C_1\text{-}C_2\text{-}C_3\text{-}C_4$ the only important degree of freedom. If one finds that condition 2) is not satisfied there are two options. As we have done in case of ribose one can accept that the problem involves more separate minima now and calculate ΔG values for all possible transitions. This has the additional advantage that more information about the potential surface becomes available, which can increase the understanding of the properties of the system under consideration. The second option is to constrain one or more degrees of freedom, but we advise against it, because in essence the complete interaction model has been changed: one does not know what the side effects are and furthermore configurational space is not sampled correctly anymore. When studying solutes with many conformational degrees of freedom, unwanted conformational changes can occur. For example van Eerden *et al.* studied the transformation of one ion into another in complexes with crown ethers [18]. After having carried out forward and reverse transitions, during which may conformational changes took place, the crown ether had not readopted its original confirmation. In such cases it is virtually impossible to act according to either of both options.

If condition 1) is not satisfied and the Ψ barriers are not very high, spontaneous transitions can occur and instead of determining ΔG between two conformers one

might end up with the difference between one conformer and a mixture of two others. The situation is even worse when the barrier between the conformations of interest is low and crossed spontaneously in the unperturbed simulations, but not frequently enough to make direct counting possible. This means that one calculates the free energy difference between a solute molecule adopting one range of values for dihedral Ψ and another one adopting a second range, overlapping the first. We dare not venture an opinion on the meaning of the resulting ΔG value. Only in case of malic acid this could have presented a problem but since in the non-ΔG MD simulations only very few spontaneous transitions occurred the calculated free energy differences have a physical meaning.

5. CONCLUSIONS

Free energy differences between conformers of the same molecule can be determined by the *thermodynamic integration* method using MD simulations, provided that one studies small or not very flexible molecules. Otherwise one must have a fairly good picture of the accessible conformational space. The results from the simulations can also improve the knowledge about accessible regions.

Hysteresis effects play an important role, but contrary to the general opinion, they provide no direct measure for the accuracy of the calculated ΔG values, because they are partly random, partly systematic in nature. However, one can use hysteresis effects together with the estimated standard deviations of the individual transitions to estimate upper bounds for the errors.

The relative order of the populations of the staggered conformers of L-malic acid and L-tartaric acid and of the α and β enantiomers of ribose is the same as derived from experiments. The calculated ΔG values agree with experimental ones indicating that the GROMOS forcefield performs rather well. Unfortunately, the error margins that can be obtained with the present resources have the same order of magnitude as most ΔG values of interest in conformational equilibria.

Preliminary calculations on ribose show that the *umbrella sampling* method does not provide any advantages compared with *thermodynamic integration.*

Acknowledgement

We thank Wilfred van Gunsteren for generously supplying us with the GROMOS package. We acknowledge the support of the SURF foundation which provided us with facilities on the Cyber 205 through a grant from the Dutch National Fund for the use of Supercomputers (NFS).

References

[1] D.B. Boyd and K.B. Lipkowitz, "Molecular mechanics - the method and its underlying philosophy", *J. Chemical Education,* **59**, 269 (1982).

[2] W.F. van Gunsteren and H.J.C. Berendsen, "Molecular dynamics: perspectives for complex systems", *Biochem. Soc. Trans,* **10**, 301 (1982).

[3] W.F. van Gunsteren and H.J.C. Berendsen, "The power of dynamic modelling of molecular systems", submitted to *J. Comp. Aid. Mol. Des.*.

[4] P.F.W. Stouten, L.M.J. Kroon-Batenburg and J. Kroon, "Conformational aspects of α-hydroxy carboxylic acids, the heralds of stereochemistry", submitted to *J. Mol. Struct.*.

[5] W.F. van Gunsteren, "Methods for calculation of free energies and binding constants: successes and problems", in: *Computation of Free Energy for Biomolecular Systems* (W.F. van Gunsteren and P. Weiner eds.), Escom Science Publishers, Leiden (1988).
[6] T.P. Straatsma, Thesis *Free Energy Evaluation by Molecular Dynamics Simulations*, University of Groningen, The Netherlands (1987) Ch. 3.
[7] T.P. Straatsma and H.J.C. Berendsen, "Free energy of ionic hydration, analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations", *J. Chem. Phys.*, **89** 5876 (1988).
[8] D.L. Beveridge, M. Mezei, G. Ravishankar and B. Jayaram, "Free energy simulations: applications to the study of liquid water, hydrophobic interactions and solvent effects on conformational stability", *Int. J. of Quantum Chemistry*, **39**, 1513 (1986).
[9] L. Stryer, *Biochemistry*, W.H. Freeman and Company, New York (1988) Ch. 16.
[10] P.F.W. Stouten, B.R. Leeflang, B.P. van Eijck, J. Kroon and J.R. Mellema, "Conformational aspects of malic acid: a multidisciplinary approach", *J. Mol. Struct.*, **189**, 65 (1988).
[11] B.P. van Eijck and J. Kroon, "Molecular-dynamics simulations of β-D-ribose and β-D-deoxyribose solutions" *J. Mol. Struct.*, **195**, 133 (1989).
[12] J.P.M. Postma, H.J.C. Berendsen & J.R. Haak, "The thermodynamics of cavity formation in water, a molecular dynamics study", *Faraday Symp. Chem. Soc.*, **17** 55 (1982).
[13] T.P. Straatsma, H.J.C. Berendsen and J.P.M. Postma, "Free energy of hydrophobic hydration: a molecular dynamics study of noble gases in water", *J. Chem. Phys.*, **85** 6720 (1986).
[14] P.A. Bash, U.C. Singh, F.K. Brown, R. Langridge and P.A. Kollman, "Calculation of the relative change in binding free energy of a protein-inhibitor complex", *Science*, **235**, 574 (1987).
[15] C.F. Wong and J.A. McCammon, "Dynamics and design of enzymes and inhibitors", *J. Am. Chem. Soc.*, **108** 3830 (1986).
[16] W.F. van Gunsteren and H.J.C. Berendsen, "Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry", *J. Comp. Aided Mol. Des.* **1** 171 (1987).
[17] U.C. Singh, F.K. Brown, P.A. Bash and P.A. Kollman,"An approach to the application of free energy perturbation methods using molecular dynamics: applications to the transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, glycine $\rightarrow$ alanine, and alanine $\rightarrow$ phenylalanine in aqueous solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the gas phase", *J. Am. Chem. Soc.*, **109**, 1607 (1987).
[18] J. van Eerden, S. Harkema and D. Feil, "Molecular dynamics of 18-crown-6 complexes with alkali-metal cations: calculation of relative free energies of complexation". *J. Phys. Chem.*, **92**, 5076 (1988)
[19] W.F. van Gunsteren, GROMOS, *Groningen Molecular Simulation computer program package*. University of Groningen, The Netherlands (1987).
[20] J. Koehler, W. Saenger and W.F. van Gunsteren, "Molecular dynamics simulation of crystalline β-cyclodextrin at 293 K and 120 K", *Eur. Biophys. J.* **15**, 211 (1987).
[21] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren and J. Hermans, "Interaction models for water in relation to protein hydration", In *Intermolecular Forces* (B. Pullman. ed.), Reidel, Dordrecht (1981) p. 331.
[22] W.F. van Gunsteren and H.J.C. Berendsen, "Algorithms for macromolecular dynamics and constraint dynamics", *Mol. Phys.*, **34**, 1311 (1977).
[23] J.P. Ryckaert, G. Cicotti and H.J.C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes", *J. Comput. Phys.*, **23**, 327 (1977).
[24] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola and J.R. Haak, "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.*, **81**, 3684 (1984).
[25] P.F.W. Stouten and J. Kroon, "Hydrogen bonds in concreto and in computro", *J. Mol. Struct.*, **177**, 467 (1988).
[26] L.M.J. Kroon-Batenburg, Thesis *Hydrogen Bonding and Molecular Conformation*, University of Utrecht, The Netherlands (1985) Ch. 8.
[27] D. Horton and Z. Walaszek, "Tautomeric equilibria of some sugars by partially relaxed, ^{13}C pulse Fourier-transform, nuclear magnetic resonance spectroscopy", *Carbohydr. Res.*, **105**, 145 (1982).
[28] D.W. Rebertus, B.J. Berne and D. Chandler, "A molecular dynamics and Monte Carlo study of solvent effects on the conformational equilibrium of n-butane in CCl_4", *J. Chem. Phys.*, **70**, 3395 (1979).
[29] C.D. Bell and S.C. Harvey, "Comparison of free energy surfaces for extended-atom and all-atom models of n-butane", *J. Phys. Chem.*, **90**, 6595 (1986).

[30] S.C. Harvey and M. Prabhakaran, "Umbrella sampling: avoiding possible artefacts and statistical biases", *J. Phys. Chem.*, **91**, 4799 (1987).
[31] G.M. Torrie and J.P. Valleau, "Nonphysical sampling distributions in Monte Carlo free energy estimation: umbrella sampling", *J. Comp. Phys.* **23**, 187 (1977).
[32] S.H. Northrup, M.R. Pear, C. Lee, J.A. McCammon and M. Karplus, "Dynamical theory of activated processes in globular proteins", *Proc. Nat. Acad. Sci. USA* **79**, 4035 (1982).
[33] A. Anderson, M. Carson and J. Hermans, "Molecular dynamics simulation study of polypeptide conformational equilibria: a progress report", *Ann. N.Y. Acad. Sci.*, **482**, 51 (1986).