Metropolis simulations of Met-Enkephalin with solvent-accessible area parametrizations

Bernd A. Berg*
Department of Physics, Florida State University, Tallahassee, Florida 32306
and School of Computational Science and Information Technology, Florida State University, Tallahassee, Florida 32306

Hsiao-Ping Hsu†
John-von-Neumann Institute for Computing, Forschungszentrum Jülich, D-52425 Jülich, Germany
(Received 17 June 2003; published 27 February 2004)

We investigate the solvent-accessible area method by means of Metropolis simulations of the brain peptide Met-Enkephalin at 300 K. For the energy function ECEPP/2 nine atomic solvation parameter (ASP) sets are studied. The simulations are compared with one another, with simulations with a distance dependent electrostatic permittivity $\epsilon(r)$, and with vacuum simulations ($\epsilon=2$). Parallel tempering and the biased Metropolis techniques RM_1 are employed and their performance is evaluated. The measured observables include energy and dihedral probability densities, integrated autocorrelation times, and acceptance rates. Two of the ASP sets turn out to be unsuitable for these simulations. For all other systems selected configurations are minimized in the search for global energy minima, which are found for vacuum and the $\epsilon(r)$ system, but for none of the ASP models. Other observables show a remarkable dependence on the ASPs. In particular, we find three ASP sets for which the autocorrelations at 300 K are considerably smaller than those for vacuum simulations.

DOI: 10.1103/PhysRevE.69.026703 PACS number(s): 05.10.Ln, 87.15.-v, 87.14.Ee

I. INTRODUCTION

In nature biomolecules exist in the environment of solvents, thus molecule–solvent interactions must be taken into account. It is very CPU time consuming to simulate models for which the molecules of the surrounding water are treated explicitly. Therefore, a number of approximations of solvent effects have been developed. In the solvent-accessible area approach [1–3] it is assumed that protein–solvent interaction is given by the sum of the surface area of each atomic group times the atomic solvation parameter (ASP). The choice of a set of ASPs (also called hydrophobicity parameters or simply hydrophobicities) defines a model of solvation. However, there is no agreement on how to determine the universally best set of ASPs, or at least the best set for some limited purpose. For instance, eight sets were reviewed and studied by Juffer *et al.* [4] and it was found that they give rather distinct contributions to the free energy of protein folding.

In this paper we investigate how different ASP sets modify the Metropolis simulations of the small brain peptide Met-Enkephalin (Tyr-Gly-Gly-Phe-Met) at 300 K. The reason for the choice of Met-Enkephalin is that its vacuum properties define a reference system for testing numerical methods; see, e.g., Refs. [5–10]. Therefore, Met-Enkephalin appears to be well suited to set references for the inclusion of solvent effects as well, but we are only aware of few articles [11–13] which consider modifications due to including a solvent model. Related to this, the effect of ASP models on the helix-coil transition of polyalanine [14] and on the 36-residue villin headpiece subdomain HP-36 [15] have been studied recently.

We set our simulation temperature to 300 K, because room temperature is the physical temperature at which biological activity takes place. Most of the previous simulations of Met-Enkephalin in vacuum were performed at much lower temperatures or employed elaborate minimization techniques with the aim of determining the global energy minimum (GEM). Using the method in Ref. [10] we access the GEM by local minimization of properly selected configurations from an equilibrium time series at 300 K.

For our simulations we use the program package simple molecular mechanics for proteins (SMMP) [16] together with parallel tempering (PT) [17–19] and the recently introduced [10] biased Metropolis technique RM_1 (rugged Metropolis, approximation 1). SMMP implements a number of all-atom energy functions that describe intramolecular interactions and nine ASP sets [3,20–26] to model molecule solvent interactions. We use the empirical conformational energy program for peptides ECEPP/2 [27] energy function with fully variable ω angles and simulate all nine ASP sets. For comparison we simulate also Met-Enkephalin in vacuum and with the distance dependent electrostatic permittivity $\epsilon(r)$ of Ref. [28].

This paper is organized as follows: The energy functions and Metropolis methods used are explained in Sec. II. In Sec. III we present our results from simulations of the brain peptide Met-Enkephalin. A summary and conclusions are given in Sec. IV.

II. MODELS AND METHODS

A. ASP sets

In all-atom models of biomolecules the total conformational energy of the intramolecular interactions E_I is given as the sum of the electrostatic, the Lennard-Jones (van der Waals), the hydrogen bond, and the torsional contributions,

*Email address: berg@csit.fsu.edu
†Email address: h.p.hsu@fz.juelich.de

$$E_I = 332 \sum_{i<j} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{i<j} \left(\frac{A_{ij}^{\mathrm{LJ}}}{r_{ij}^{12}} - \frac{B_{ij}^{\mathrm{LJ}}}{r_{ij}^{6}} \right) + \sum_{i<j} \left(\frac{A_{ij}^{\mathrm{HB}}}{r_{ij}^{12}} - \frac{B_{ij}^{\mathrm{HB}}}{r_{ij}^{10}} \right) + \sum_k U_k [1 \pm \cos(n_k \phi_k)]. \quad (1)$$

Here r_{ij} is the distance between atoms i and j, q_i and q_j are the partial charge on atoms i and j, ϵ is the electric permittivity of the environment, A_{ij}, B_{ij}, C_{ij} and D_{ij} are parameters that define the well depth and width for a given Lennard-Jones or hydrogen bond interaction, and ϕ_k is the kth torsion angle. The units are as follows: distances are in Å, charges are in units of the electronic charge and energies are in kcal/mol.

One of the simplest ways in which to include interactions with water is to assume a distance dependent electrostatic permittivity according to the formula [28,29]

$$\epsilon(r) = D - \frac{D-2}{2}[(sr)^2 + 2sr + 2]e^{-sr}. \quad (2)$$

Empirical values for parameters D and s are chosen, so that for large distances the permittivity takes the value of bulk water, $\epsilon = 80$, and the value $\epsilon = 2$ for short distances, i.e., for the interior of the molecule. Approximating solvation effects in this way is implemented as an option in SMMP. It allows one to include solvation effects without any significant slowing down over the vacuum simulation with $\epsilon = 2$. The approach is clearly oversimplification, because atoms which are close to each other do not necessarily have to be simultaneously in the interior of the molecule. Conversely, two atoms which are separated by a large distance may still be in the interior of the molecule. More elaborated approaches are needed.

If the molecule–solvent interaction is proportional to the surface area of the atomic groups, it is given by the sum of contributions of a product of the surface area of each atomic group and the atomic solvation parameter [3],

$$E_{\mathrm{sol}} = \sum_i \sigma_i A_i . \quad (3)$$

Here E_{sol} is the solvation energy and the sum is over all atomic groups. A_i is the solvent-accessible surface area and σ_i the atomic solvation parameter of group i. The choice of a set of ASPs σ_i defines a model of solvation. There are nine sets of ASPs in the SMMP package, and we list them in Table I. Columns one and two of Table I give the notations used in SMMP to identify the different sets.

Eisenberg and McLachlan [3] were the first to determine a set of ASPs (`itysol`=4, EM86 in SMMP notation). For this, they considered the process of transferring atoms or groups of atoms from the interior of a protein to aqueous solution and used transfer energies of amino acids from *n*-octanol to water as reported in Ref. [30]. The ASPs are then determined by least-square fitting. Octanol is chosen, because it apparently resembles the interior of a protein. With the exception of Refs. [25] and [26], all the other authors used similar methods with the major variation that instead of transfer energies with respect to octanol-water (o/w) also transfer energies with respect to vacuum water (v/w) were used (for early determination of v/w transfer energies see Refs. [31] and [32]). The last column of Table I indicates whether the transfer energy is o/w or v/w. In chronological order Eisenberg and co-workers [3,21,24] contributed parameter sets EM86, SCH1, WE92 and SCH3. Scheraga and coworkers [20,23] contributed parameter sets OONS and JRF. Here it should be noted that some of the original ASP sets were modified over the course of time. For `itysol` $=1, \ldots, 8$ SMMP implements the parameters reviewed and tabulated in Ref. [4], where in turn sets SCH1–SCH4 are simply taken from work by Schiffer *et al.* [25]. Table I of SMMP [16] lists the implemented ASPs for `itysol` $=1, \ldots, 8$.

TABLE I. Atomic solvation parameter sets implemented in SMMP. The first column gives the value of the SMMP parameter `itysol` and the second column the letter code used in SMMP. In the author column we give also the year of publication. The last column indicates the method used that is explained in the text.

		Author	
1	OONS	Ooi *et al.* 1987 [20]	v/w
2	JRF	Vila *et al.* 1991 [23]	v/w
3	WE92	Wesson and Eisenberg 1992 [24]	v/w
4	EM86	Eisenberg and McLachlan 1986 [3]	o/w
5	SCH1	Eisenberg *et al.* 1989 [21]	o/w
6	SCH2	Kim 1990 [22]; see also Ref. [4]	o/w
7	SCH3	Wesson and Eisenberg 1992 [24]	v/w
8	SCH4	Schiffer *et al.* 1993 [25]	v/ws
9	BM	von Freyberg *et al.* 1993 [26]	cla

Somewhat special cases are ASP sets SCH4 [25] and BM [26]. SCH4 was determined by a comparison of the crystal structure in molecular dynamics simulations of small peptides and proteins explicitly in water with similar simulations using an ASP solvation term (v/ws). The BM set of SMMP relies on a specific classification (cla) of atomic groups, where for all nonhydrogen atoms the solvation coefficients are set to 1 kcal/mol per Å^2.

B. Metropolis methods

To update our systems we use PT with two processors, one running at 300 K and the other at 400 K. This builds on experience [10] with vacuum simulations of Met-Enkephalin for which the following observations were made.

(1) The integrated autocorrelation time τ_{int} (defined below) increases from 400 to 300 K by a factor of 10 for the (internal) energy and by factors of more than 20 for certain dihedral angles.

(2) The energy probability densities (PDs) at 300 and 400 K overlap sufficiently, so the PT method works and leads to an improvement factor of about 2.5 in the real time needed for simulation (see Table I of Ref. [10]).

A brief description of the PT algorithm is given in the following. PT performs n canonical Monte Carlo (MC) simu-

lations at different β values with Boltzmann weight factors,

$$w_{B,i}(E^{(k)}) = e^{-\beta_i E_i^{(k)}} = e^{-H}, \quad i = 0, \ldots, n-1, \tag{4}$$

where $\beta_0 < \beta_1 < \cdots < \beta_{n-2} < \beta_{n-1}$ and a configuration is denoted by k. PT allows the exchange of neighboring β values,

$$\beta_{i-1} \leftrightarrow \beta_i \quad \text{for} \quad i = 1, \ldots, n-1. \tag{5}$$

These transitions lead to the change

$$\begin{aligned} -\Delta H &= -\beta_{i-1}(E_i^{(k)} - E_{i-1}^{(k')}) - \beta_i(E_{i-1}^{(k')} - E_i^{(k)}) \\ &= (\beta_i - \beta_{i-1})(E_i^{(k)} - E_{i-1}^{(k')}), \end{aligned} \tag{6}$$

which is accepted or rejected according to the Metropolis algorithm, i.e., with probability one for $\Delta H \leqslant 0$ and with probability $\exp(-\Delta H)$ for $\Delta H > 0$.

For the vacuum system the performance of the PT simulation was improved by an additional factor of 2 in Ref. [10] by using a first approximation, called RM_1, to the rugged Metropolis scheme introduced there. A (short) simulation at 400 K was used to obtain estimates $\overline{\rho_j}(v_j)$, $j = 1, \ldots, 24$ of the PDs of the 24 dihedral angles, which were then fed into the simulation. For a configuration change of $k \to k'$ at temperature T_i the new configuration is accepted with probability,

$$p_{\mathrm{acpt}} = \min\left[1, \frac{\exp(-\beta_i E_i^{(k')}) \prod_{j=1}^{24} \overline{\rho_j}(v_j^{(k')})}{\exp(-\beta_i E_i^{(k)}) \prod_{j=1}^{24} \overline{\rho_j}(v_j^{(k)})}\right], \tag{7}$$

in the RM_1 updating scheme. In the present paper we report improvement due to this biased updating for some of the ASP sets.

In the vacuum simulation it is possible to determine the GEM by minimizing selected configurations of the 300 K time series. Here we apply the same procedure to our PT simulation of the ASP sets introduced in Sec. II A.

(1) We determine the lower 10% quantile $E_{0.1}$ and the upper 10% quantile $E_{0.9}$ of the energy distribution of our time series. This is done by sorting all energies in increasing order and finding the values which cut out the lower and upper 10% of the data. For statistical concepts see, e.g., Ref. [33].

(2) We partition the time series into bunches of configurations. A bunch contains the configurations from one crossing of the upper quantile $E_{0.9}$ to the next so that at least one crossing of the lower quantile $E_{0.1}$ is located between the two crossings of $E_{0.9}$. For each bunch we then pick its configuration of lowest energy. The idea behind this procedure is to pick minima of the time series, which are to a large degree statistically independent. In Fig. 1 the arrows indicate the energy values picked in that way from the first 600 configurations recorded in the RM_1 simulation in Ref. [10].

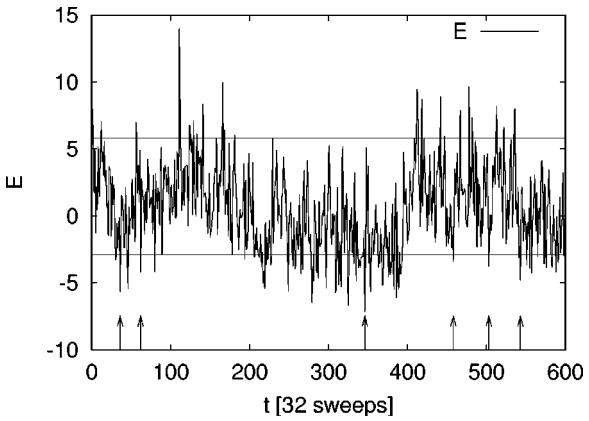

FIG. 1. Selection of configurations for local minimization from the energy (kcal/mol) time series at 300 K. The lower and upper straight lines indicate the quantiles $E_{0.1}$ and $E_{0.9}$, respectively.

(3) We run a conjugate gradient minimizer on all the selected configurations and thus obtain a set of configurations which are local energy minima. For the vacuum simulation [10] about 5%–6% of these minimized configurations agree with the GEM.

To determine the speed at which the systems equilibrate, we measure the integrated autocorrelation time τ_{int} for the energy and each dihedral angle. The integrated autocorrelation times are directly proportional to the computer run times needed to achieve the same statistical accuracy for each system. They thus determine the relative performance of distinct algorithms. For an observable f the autocorrelations are

$$C(t) = \langle f_0 f_t \rangle - \langle f \rangle^2, \tag{8}$$

where t is the computer time. Defining $c(t) = C(t)/C(0)$, the time-dependent integrated autocorrelation time is given by

$$\tau_{\mathrm{int}}(t) = 1 + 2 \sum_{t'=1}^{t} c(t'). \tag{9}$$

Formally the integrated autocorrelation time τ_{int} is defined by $\tau_{\mathrm{int}} = \lim_{t \to \infty} \tau_{\mathrm{int}}(t)$. Numerically, however, this limit cannot be reached since the noise of the estimator increases faster than the signal. Nevertheless, one can calculate reliable estimates by reaching a window of t values for which $\tau_{\mathrm{int}}(t)$ becomes flat, while its error bars are still reasonably small. This is the method we employ in Sec. III; see Ref. [34] for a more detailed discussion of the integrated autocorrelation time.

III. RESULTS

A. Autocorrelations

The PT simulations with temperatures $T_0 = 400$ K and $T_1 = 300$ K are performed on the system in vacuum ($\epsilon = 2$), with $\epsilon(r)$ given by Eq. (2) and for the nine ASP sets in Table I. The dihedral angles updated in our simulations are varied in the full range from $-\pi$–π. We keep a time series of $2^{16} = 65\,536$ configurations for each replica (i.e., each of the two processors), in which subsequent configurations are separated by 32 sweeps. A sweep is defined by updating each

TABLE II. Average energies $\langle E \rangle$(kcal/mol), acceptance rates and integrated autocorrelations times $\tau_{\rm int}$ for the energy are shown for simulations in vacuum (VAC), with $\epsilon(r)$ of Eq. (2) and with the nine ASPs introduced in Table I.

	$T=400$ K			$T=300$ K		
Set	$\langle E \rangle$	acpt	$\tau_{\rm int}$	$\langle E \rangle$	acpt	$\tau_{\rm int}$
VAC	7.07(03)	0.167	3.67(20)	1.29(06)	0.119	19.9(1.6)
$\epsilon(r)$	−12.00(03)	0.171	2.92(10)	−17.61(06)	0.121	14.35(75)
OONS	−13.80(01)	0.195	1.25(02)	−17.70(02)	0.143	2.64(14)
JRF	−311.69(44)	0.058	···	−319.08(40)	0.046	···
WE92	−15.76(02)	0.199	1.30(03)	−19.75(02)	0.145	2.94(07)
EM86	13.49(03)	0.158	4.71(21)	8.03(06)	0.116	25.0(2.9)
SCH1	10.45(03)	0.165	3.72(22)	4.95(06)	0.119	23.2(2.2)
SCH2	−18.33(02)	0.212	1.11(01)	−21.83(01)	0.160	1.89(05)
SCH3	13.33(03)	0.151	4.59(34)	8.33(06)	0.112	26.4(3.3)
SCH4	13.38(03)	0.158	4.35(17)	7.85(05)	0.115	25.1(2.1)
BM	630.4(3.9)	0.043	···	610.6(3.0)	0.037	···

dihedral angle sequentially once. Before starting with measurements $2^{18}=262\,144$ sweeps are performed to reach equilibrium. Thus, the entire simulation at one temperature relies on $2^{21}+2^{18}=2\,359\,296$ sweeps. On the Cray T3E, this takes about 14 h for the vacuum system and 5×14 h for each ASP set.

Results of the average energy, acceptance rates and integrated autocorrelations times for the energy E are shown in Table II. For the vacuum simulations and ASP sets OONS and EM86 the time-dependent integrated autocorrelations times (9) are shown in Fig. 2. In each case a window of t values is reached for which $\tau_{\rm int}(t)$ no longer increases within statistical error. In the case of the vacuum simulations it even decreases, but this is not significant due to statistical error. These windows are then used to estimate the asymptotic $\tau_{\rm int}$ values in Table II. With the exception of ASP sets JRF and BM, the integrated autocorrelations times of all other sets are determined in the same way.

From Table II we see that the acceptance rates of solvent models JRF and BM are much lower than those for the other models. In essence the simulations of these two models get stuck, which implies that their integrated autocorrelation times cannot be measured. This is illustrated in Fig. 3 for the time-dependent integrated autocorrelation time of the energy at 300 K. The function $\tau_{\rm int}(t)$ increases rapidly until it gets lost in the noise. The PDs of the dihedral angles of these two models are also erratic and the conclusion is that they cannot be used to describe Met-Enkephalin in solvent.

The energy couples to all dihedral angles and its integrated autocorrelation time is characteristic of the entire system, whereas the integrated autocorrelation times of the single dihedral angles vary greatly from angle to angle. For all of our systems except JRF and BM, we show in Fig. 4 the integrated autocorrelation times at 300 K for the energy and all dihedral angles. The notation v_i, $i=0,1,\ldots,24$ is used, where v_0 stands for energy E and the v_i for $i=1,\ldots,24$ are the dihedral angles used in the SMMP computer program. The relationship of the v_i angles to the conventional notation for dihedral angles and their residue is summarized in Table III, where it should be noted that the SMMP notation [16] differs from that in other literature [5,8].

In Fig. 4 we see that for each dihedral angle v_i the integrated autocorrelation times $\tau_{\rm int}[v_i]$ for the three solvent

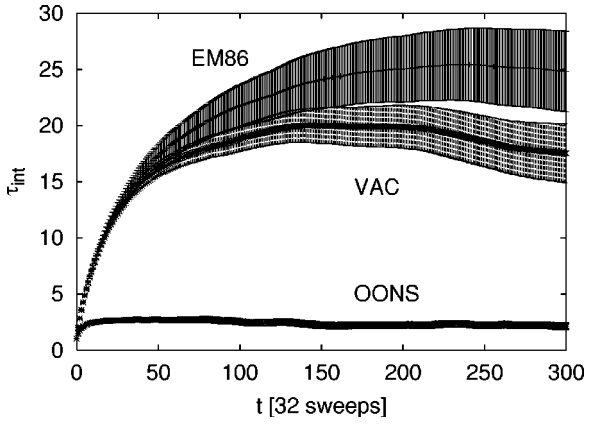

FIG. 2. Time-dependent integrated autocorrelation time for the energy at 300 K from our simulations of the vacuum system and two of the solvent models in Table I.

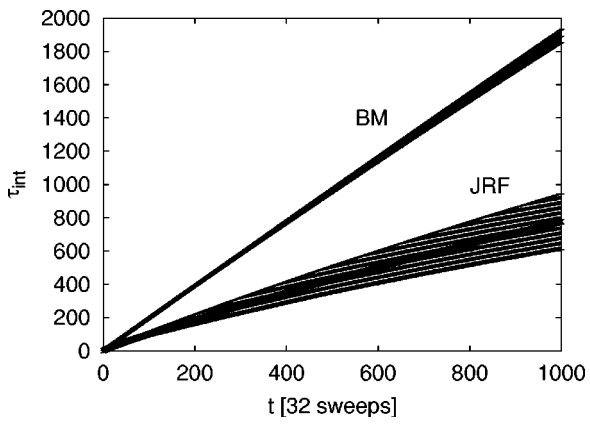

FIG. 3. Time-dependent integrated autocorrelation time for the energy at 300 K from our simulations of solvent models JRF and BM.

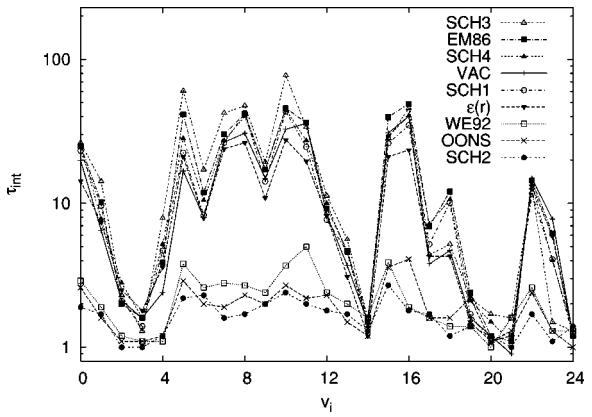

FIG. 4. Integrated autocorrelation times for the energies ($v_0 = E$) and the dihedral angles v_i, $i=1,\ldots,24$ at $T=300$ K. The up-to-down order of the curves agrees at $i=10$ with the order shown in the legend.

TABLE III. Definitions of the dihedral angles together with their integrated autocorrelation times τ_{int} at 300 K for simulations of WE92 with statistics reduced by 1/8 and configurations recorded every four sweeps. PT denotes the 400–300 K parallel tempering simulation. For PT-RM_1 the PT simulation is supplemented by RM_1 bias updating, Eq. (7), with input PDs from 400 K. The factor of the last column denotes the increase of PT τ_{int} over its values for the full WE92 simulations where configurations were recorded every 32 sweeps (8 is the upper bound for this factor).

Var	Angle	res [5,8]	res [16]	PT-RM_1	PT	Factor
v_1	χ^1	Tyr-1	Tyr-1	6.9 (1.1)	11.6 (1.6)	6.1 (0.9)
v_2	χ^2	Tyr-1	Tyr-1	2.0 (0.2)	3.1 (0.5)	2.7 (0.5)
v_3	χ^6	Tyr-1	Tyr-1	1.0 (0.1)	1.3 (0.2)	1.3 (0.2)
v_4	ϕ	Tyr-1	Tyr-1	2.1 (0.2)	2.6 (0.4)	2.4 (0.4)
v_5	ψ	Tyr-1	Gly-2	12.6 (1.7)	15.7 (2.2)	4.1 (0.7)
v_6	ω	Tyr-1	Gly-2	3.9 (0.4)	14.4 (1.2)	5.6 (0.5)
v_7	ϕ	Gly-2	Gly-2	9.1 (1.0)	13.0 (1.4)	4.6 (0.6)
v_8	ψ	Gly-2	Gly-3	10.6 (1.3)	20.4 (3.1)	7.6 (1.2)
v_9	ω	Gly-2	Gly-3	3.4 (0.2)	16.0 (1.8)	6.7 (0.8)
v_{10}	ϕ	Gly-3	Gly-3	18.2 (3.2)	31.0 (5.1)	8.4 (1.5)
v_{11}	ψ	Gly-3	Phe-4	15.6 (2.9)	52 (13)	12 (4)
v_{12}	ω	Gly-3	Phe-4	4.4 (0.6)	17.7 (2.7)	7.7 (1.3)
v_{13}	χ^1	Phe-4	Phe-4	3.3 (0.4)	6.9 (1.1)	4.4 (0.8)
v_{14}	χ^2	Phe-4	Phe-4	1.7 (0.2)	3.2 (0.4)	3.0 (0.4)
v_{15}	ϕ	Phe-4	Phe-4	8.9 (1.3)	19.6 (3.2)	6.3 (1.2)
v_{16}	ψ	Phe-4	Met-5	4.5 (0.3)	8.0 (0.9)	4.4 (0.6)
v_{17}	ω	Phe-4	Met-5	1.8 (0.2)	8.1 (1.2)	5.4 (0.8)
v_{18}	χ^1	Met-5	Met-5	2.7 (0.2)	8.3 (2.5)	6.3 (1.9)
v_{19}	χ^2	Met-5	Met-5	1.9 (0.2)	5.3 (0.4)	4.0 (0.5)
v_{20}	χ^3	Met-5	Met-5	1.1 (0.1)	2.7 (0.2)	2.5 (0.2)
v_{21}	χ^4	Met-5	Met-5	1.0 (0.1)	1.3 (0.1)	1.3 (0.1)
v_{22}	ϕ	Met-5	Met-5	36 (18)	23.8 (5.6)	9.5 (2.4)
v_{23}	ψ	Met-5	Met-5	1.4 (0.2)	1.9 (0.1)	1.9 (0.1)
v_{24}	ω	Met-5	Met-5	1.0 (0.1)	3.4 (0.2)	3.1 (0.4)
v_0	E			9.0 (1.7)	19.4 (3.1)	6.6 (1.1)

models, OONS, WE92 and SCH2, are smaller than for the remaining systems, including the vacuum system. For the integrated autocorrelation time of energy $\tau_{int}[E]$ this observation is already obvious from Table II. This means that the OONS, WE92 and SCH2 models require far fewer statistics than the vacuum run to achieve the same accuracy in the results. Using the $\tau_{int}[E]$ results in Table II, we find a factor in the range of 7–10, which more than offsets the factor of 5 by which the ASP model simulations are slower than the vacuum simulation. In the following solvation models OONS, WE92 and SCH2 define the ''fast class,'' while the other models shown in Fig. 4 constitute the ''slow class'' (models JRF and BM are omitted from this classification). ''Good'' behavior of the models OONS and WE92 has previously been observed [35].

The autocorrelation times in the fast class are so small that the resolution of 32 sweeps in our recorded time series becomes too crude. So autocorrelations over less than 32 sweeps are then not measured and the integrated autocorrelation time approaches one as soon as autocorrelations stay within the range of 32 sweeps. To investigate this point further, we performed for the OONS, WE92 and SCH2 models simulations for which the configurations were recorded every four sweeps and the total statistics were reduced by a factor of 1/8. In the new units of four sweeps the integrated autocorrelation time is larger by a factor which is bounded by $8=32/4$. The bound is assumed, if there is no improvement due to integrating additional small fluctuations out (i.e., due to the additional configuration in between the 32 sweeps, which are now kept in the time series).

For WE92 we report in the PT column of Table III the integrated autocorrelation times from the simulation with reduced statistics. For many dihedral angles the increase lies well below a factor of 8, showing that we gain in accuracy by averaging over small fluctuations within the range of 32 sweeps. On the other hand, nothing is gained by this extra averaging for several angles with large autocorrelations. In those cases the simulations yield, within statistical error, an upper bound of 8.

To supplement the vacuum results of Ref. [10], we repeated the WE92 PT simulations by using estimates of the dihedral PDs from 400 K as input for biased updating of Eq. (7). These results are reported in the PT-RM_1 column of Table III. As in the case of the vacuum simulations, we find improvement of the PT performance by a factor of approximately 2, which is also obtained for the other models of the fast class. For the slow class we checked the direct improvement of the original simulations of ASP models EM86 and SCH4 and find again acceleration by a factor of about 2 when we are using RM_1 updating.

B. Structure

For all our simulations we applied the method outlined in Sec. II B to determine local energy minima and some results are summarized in Table IV: $E_{0.1}$ and $E_{0.9}$ are the lower and upper 10% quantiles of the energy and N_{conf} denotes the number of minima of the time series prepared for further minimization. The lowest energy found in this minimization

TABLE IV. Determination of local minima: $E_{0.1}$ and $E_{0.9}$ are the lower and upper 10% quantiles of the energies of the time series recorded, and N_{conf} denotes the number of configurations prepared for further minimization, E_{min} (kcal/mol) is the lowest energy found, and N_{hits} is the number of times the lowest energy configuration was hit.

	$T=400$ K					$T=300$ K				
Set	N_{conf}	$E_{0.1}$	$E_{0.9}$	E_{min}	N_{hits}	N_{conf}	$E_{0.1}$	$E_{0.9}$	E_{min}	N_{hits}
VAC	2190	1.98	12.26	−12.91	13	1073	−2.98	5.73	−12.91	55
$\epsilon(r)$	2622	−16.95	−6.97	−31.94	8	1312	−21.85	−13.17	−31.94	27
OONS	3315	−17.83	−9.63	−27.69	1	2641	−21.40	−13.96	−28.93	1
JRF	448	−317.96	−304.98	−328.72	1	365	−323.66	−314.24	−332.87	1
WE92	3307	−19.88	−11.52	−29.44	1	2453	−23.43	−15.95	−30.39	1
EM86	2307	8.57	18.59	−4.11	1	1191	3.83	12.39	−5.47	1
SCH1	2511	5.57	15.45	−5.54	1	1147	0.71	9.33	−7.52	1
SCH2	3454	−22.17	−14.34	−31.32	1	2918	−25.31	−18.32	−32.71	1
SCH3	2315	8.57	18.32	−1.70	1	1229	4.29	12.50	−3.29	1
SCH4	2331	8.45	18.50	−4.93	1	1108	3.66	12.23	−5.16	1
BM	2	606.37	655.16	594.78	1	1	598.37	646.35	590.32	1

process is denoted by E_{min} and N_{hits} is the number of times the lowest energy configuration was hit. While the absolute values of $E_{0.1}$ and $E_{0.9}$ vary considerably from set to set, the differences in $E_{0.9}-E_{0.1}$ stay similar. The explanation is that the ASP sets differ by large additive constants to the energy.

Again, the results of the JRF and BM solvent models are erratic. The BM model is entirely frozen, $N_{\mathrm{conf}}=2$ at 400 K and $N_{\mathrm{conf}}=1$ at 300 K. Therefore, we do not give minimization results for BM. For JRF the N_{conf} numbers are more reasonable, but are still by a factor of 1/3 or less smaller than the N_{conf} numbers of each of the other systems. JRF is also disregarded in the following discussion.

Only if we have $N_{\mathrm{hits}}>1$ do we have an indication that we found the GEM. Interestingly, this happens for none of the ASP solvent models, while it does for the case of the vacuum and the $\epsilon(r)$ simulations (notably already at 400 K). Quite some time ago Li and Scheraga [5,11] developed a Monte Carlo minimization method and applied it to Met-Enkephalin in vacuum and in solvent modeled by OONS. While for the vacuum system their method converged consistently to the GEM, all five of their runs of the solvent model led to different conformations with comparable energies. They interpreted their results in the sense that Met-Enkephalin in water at 20 °C is likely in an unfolded state for which a large ensemble of distinct conformations coexist in equilibrium. A consistent scenario was later observed in nuclear magnetic resonance (NMR) experiments [36].

Although the minimization method of Li and Scheraga is entirely different from ours, they essentially tested for valleys of attraction to the GEM at room temperature, as we do in the present paper. So, we have not only confirmed their old result, but find that it is also common to a large set of ASP models implemented in SMMP. Neither the method by which an ASP set was derived, nor whether it belongs to the fast or slow class, appears to matter with this respect.

As an example, the frequency of local energy minima of the WE92 solvation model obtained by our minimization procedure from the 300 K time series is depicted in Fig. 5. $N_{\mathrm{conf}}=2453$ minimizations are performed. Our lowest energy state is only found once and the same holds for nearby low energy states. Figure 5 should be compared with Fig. 2 in Ref. [10], where the frequency of the low energy minima of the vacuum simulation is shown. There the lowest energy state relies on 107 entries out of 1913 minimizations [37].

In a search for structural differences of Met-Enkephalin in vacuum, or in the $\epsilon(r)$ system, versus the ASP models, we looked at the PDs of the dihedral angles at 300 K. For all systems together there are $9\times24=216$ figures to consider. At first glance the PDs of the different systems are amazingly similar, independent of whether they are from systems of the fast or slow class, from an ASP model, from vacuum or from the $\epsilon(r)$ simulation. A more careful investigation reveals differences which appear to relate to distinct behavior under our minimization. For dihedral angle v_7 this is illustrated in Figs. 6 and 7. Its probability densities are compared at 300 and 400 K. For the vacuum simulation the PDs are depicted in Fig. 6 and from 400 to 300 K we observe an increase of the peak which is located close to the arrow, which indicates the vacuum GEM value of v_7. In contrast to this, the wrong peak

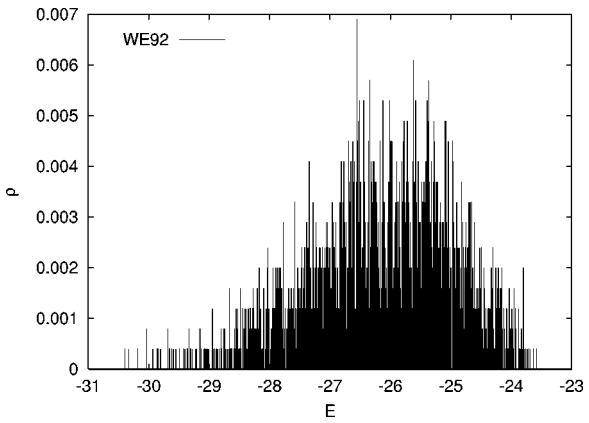

FIG. 5. Local energy minima (kcal/mol) for the WE92 solvation model obtained by our minimization method.

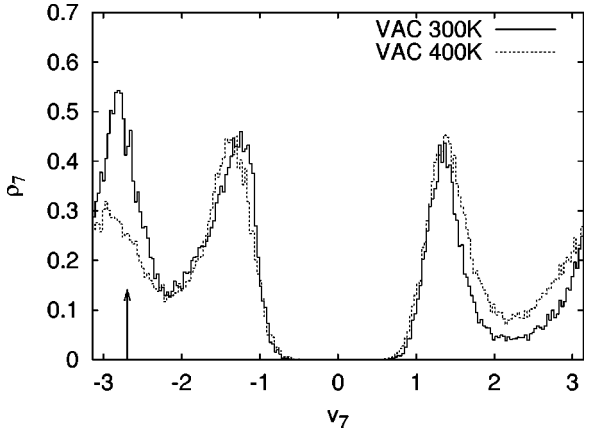

FIG. 6. Probability density of dihedral angle v_7 for the vacuum simulation. The arrow indicates the vacuum GEM value of this angle.

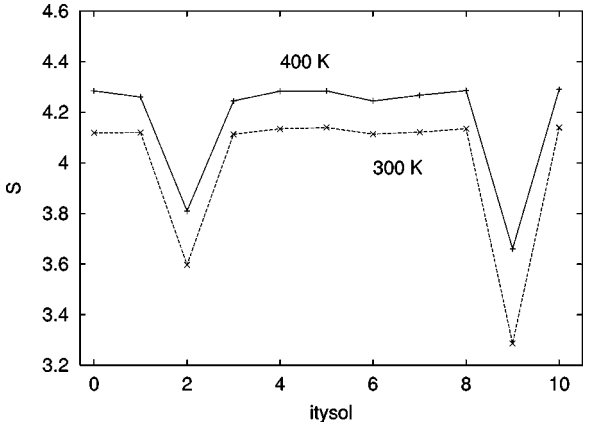

FIG. 8. Entropies of the PDs of our ASP models. The models are labeled according Table I, in addition `itysol`=0 for vacuum and `itysol`=10 for the $\epsilon(r)$ model.

increases in Fig. 7, where the PDs are shown for the WE92 solvent model.

One may suspect that the difference between the models of our fast and slow class is simply due to an effectively higher temperature for the three models of the fast class. To gain insight into this, we calculated the entropies of our PDs. Each PD is discretized as a histogram of 200 entries, ρ_{ij}, where $i=1,\ldots,24$ labels the dihedral angles according to Table III and $\Sigma_{j=1}^{200}\rho_{ij}=1$. The entropy of the PD of a dihedral is then defined by

$$S_i=-\sum_{j=1}^{200}\rho_{ij}\ln\rho_{ij}, \tag{10}$$

and the total entropy of the PDs of an ASP model is $S=\Sigma_i S_i$. In Fig. 8 the entropies obtained are depicted for all our models. The lines between the data points are simply a guide for the eye. The dips for the JRF and the BM model show, again, that their configurations are essentially frozen. For the other we see a decrease of entropy from 400 to 300 K, but we find no larger entropy for the models of the fast class than for the models of the slow class. Therefore, the effective temperature scenario is ruled out. Instead, it seems that for the models of the fast class the solvent has some kind of ''lubrication'' effect, which accelerates the simulation.

Strong similarities between the ASP models of the fast class, on the one hand, and the ASP models of the slow class on the other are found for the solvation energies, the gyration radii and the end to end distances.

The solvation energies E_{sol}, Eq. (3), measured during our simulations of ASP models are solvent–vacuum (s/v) transfer energies. There are structural differences between the typical configurations of an ASP model time series and the vacuum time series. Consequently, the average s/v transfer energies are not identical with the average vacuum–solvent (v/s) transfer energies, which are obtained by calculating E_{sol} of the solvent models on the configurations of the vacuum time series. The s/v as well as the $-$v/s average transfer energies are given in Table V. The averages are taken for the canonical time series at 300 and 400 K. At 300 K averages are also taken for the time series minima (indicated by arrows in Fig. 1) and for the local minima (which are obtained by running the conjugate gradient minimizer on the time series minima). For gyration radii R_{gy} and end to end distances $R_{\mathrm{e\text{-}e}}$ the same averages are given in Table VI (definitions and software are given in SMMP).

For the transfer energies the overall effect is hydrophilic for the ASP models of the fast class and hydrophobic for the ASP models of the slow class. Within each class the values are quite similar, despite differences in interaction coefficient (see Table I of Ref. [16]). As expected the overall transfer energies of the JRF and BM models are out of the reasonable range, JRF to the hydrophilic and BM to the hydrophobic side. Our Table VI shows that we observe the extended structures found in previous simulations [11,12] and in NMR experiments [36] only for the ASP models of the fast class.

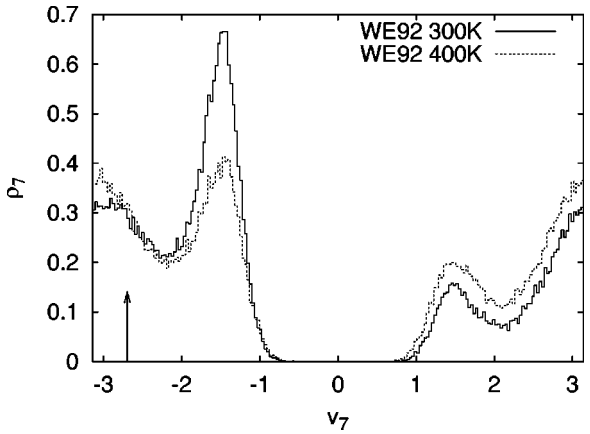

FIG. 7. Probability density of dihedral angle v_7 for the WE92 simulation. The arrow indicates the vacuum GEM value of this angle.

IV. SUMMARY AND CONCLUSIONS

We have performed Met-Enkephalin simulations at room temperature (300 K) for the solvation models of Table I. Quantitative results obtained in that way cannot be trusted,

TABLE V. Transfer energies.

	$T=300$ K						$T=400$ K	
	$\langle E^T\rangle$(local minima)		$\langle E^T\rangle$(time series minima)		$\langle E^T\rangle$		$\langle E^T\rangle$	
Set	s/v	−v/s	s/v	−v/s	s/v	−v/s	s/v	−s/v
OONS	−24.85	−18.71	−24.63	−18.53	−24.84	−19.07	−24.94	−20.19
JRF	−346.95	−185.65	−346.72	−187.30	−343.56	−196.96	−337.68	−209.46
WE92	−28.84	−17.46	−28.52	−17.62	−27.87	−18.76	−27.47	−20.61
EM86	6.19	6.93	6.29	7.01	6.45	7.08	6.60	7.11
SCH1	2.76	3.89	2.86	3.92	2.94	3.80	2.97	3.60
SCH2	−31.03	−20.61	−30.77	−20.80	−30.61	−21.97	−30.45	−23.85
SCH3	4.03	9.73	4.28	9.80	4.74	9.50	5.15	9.00
SCH4	6.08	6.86	6.17	6.93	6.32	6.95	6.46	6.94
BM	⋯	715.29	⋯	721.43	581.60	745.22	597.61	776.03

apparently because the methods to derive the ASPs are quite crude. Also, our simulations do not give information that would allow us to pick the best ASP set for the intended purpose of simulating Met-Enkephalin at 300 K. Nevertheless, we obtain a qualitative overview of a number of interesting consequences, which one might expect by including solvation effects via an ASP model in Metropolis calculations.

Two of the ASP sets (JRF [23] and BM [26] implemented in SMMP [16]) suffer from such large autocorrelations that for them Metropolis simulations at 300 K are in essence impossible. Their dihedral angles are essentially frozen. These two ASP sets are certainly erratic, since 300 K is the temperature at which thermodynamic fluctuations of the systems are expected (also these two sets perform badly at 400 K).

The remaining nine models, seven ASP sets, $\epsilon=2$ vacuum, and an $\epsilon(r)$ system [28], fall into a fast and a slow class with respect to their integrated autocorrelation times; see Fig. 4. Vacuum simulations are in the slow class. This leads to the interesting feature that it takes less computer time to estimate physical observables at room temperature in fast solvation models OONS [20], WE92 [24], and SCH2 [22,25], than it takes for vacuum, despite the substantial increase in computer time per sweep of a factor of about 5 for the solvation models over the vacuum system. We have no clear clue why some models have fast and others slow dynamics. To derive the parameters of OONS and WE92 v/w transfer energies were used, but for SCH2 it was o/w. Also, the slow class features v/w as well as o/w ASP models.

We applied the minimization procedure of Ref. [10] in an attempt to locate the GEM for the nine systems, which are reasonably well behaved under Metropolis simulations at 300 K. The GEM is unambiguously found for the vacuum system and for the simulation with a distance-dependent electrostatic permittivity. No true GEM is found for any of the remaining seven ASP models. This confirms an old result of Li and Scheraga [11], who concluded that at room temperature Met-Enkephalin in water is likely in an unfolded state. To get better understanding of this result, we studied at 300 K the dihedral PDs in some detail. At first glance they look quite similar for all the models in the fast as well as in the slow class. Differences are found in a number of details, which may allow one to explain why the 300 K configurations of the ASP models behave entirely different under our

TABLE VI. Gyration radii and end to end distance.

	$T=300$ K						$T=400$ K	
Set	Local minima		Time series minima		Time series		Time series	
	$\langle R_{gy}\rangle$	$\langle R_{e\text{-}e}\rangle$	$\langle R_{gy}\rangle$	$\langle R_{e\text{-}e}\rangle$	$\langle R_{gy}\rangle$	$\langle R_{e\text{-}e}\rangle$	$\langle R_{gy}\rangle$	$\langle R_{e\text{-}e}\rangle$
VAC	4.56	5.67	4.60	5.83	4.72	6.83	4.97	8.50
ϵr	4.53	6.20	4.57	6.33	4.71	7.39	4.99	8.94
OONS	4.92	10.30	4.94	10.34	5.32	11.71	5.60	12.45
JRF	5.75	13.00	5.75	13.00	5.78	13.05	5.70	13.20
WE92	5.06	12.06	5.07	12.09	5.43	13.02	5.72	13.47
EM86	4.47	6.97	4.48	7.02	4.62	7.94	4.86	9.28
SCH1	4.51	6.96	4.52	7.02	4.68	7.96	4.96	9.45
SCH2	5.18	12.48	5.20	12.53	5.63	13.47	5.86	13.66
SCH3	4.54	8.88	4.55	8.90	4.72	9.56	4.95	10.50
SCH4	4.46	6.82	4.47	6.87	4.62	7.83	4.87	9.17
BM	⋯	⋯	⋯	⋯	4.13	7.30	4.21	7.66

minimization procedure than the vacuum and the $\epsilon(r)$ systems.

The main question which remains to be settled is whether ASP models will ultimately allow accurate Metropolis simulations of biomolecules like Met-Enkephalin in solvent or not. In principle, this could be decided by determining whether ASPs exist which reproduce accurately mean energies of explicit solvent simulation around a large number of fixed Met-Enkephalin configurations.

ACKNOWLEDGMENTS

The authors would like to thank Professor P. Grassberger for useful discussions and generous support of this work. The computer simulations were carried out on the Cray T3E at the John von Neumann Institute for Computing. One author (B.A.B.) acknowledges partial support by the U.S. Department of Energy under Contract No. DE-FG02-97ER41022.

[1] B. Lee and M. Richards, J. Mol. Biol. **55**, 379 (1971).
[2] C. Cothia, Nature (London) **248**, 338 (1974).
[3] D. Eisenberg and A.D. McLachlan, Nature (London) **319**, 199 (1986).
[4] A.H. Juffer, F. Eisenhaber, S.J. Hubbard, D. Walther, and P. Argos, Protein Sci. **4**, 2499 (1995).
[5] Z. Li and H.A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **85**, 6611 (1987).
[6] Y. Okamoto, T. Kikuchi, and H. Kawai, Chem. Lett. **1992**, 1275 (1992).
[7] U.H. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993).
[8] H. Meirovitch, E. Meirovitch, A.G. Michel, and M. Vásquez, J. Phys. Chem. **98**, 6241 (1994).
[9] U.H. Hansmann, Y. Okamoto, and J.N. Onuchic, Proteins: Struct., Funct., Genet. **34**, 472 (1999).
[10] B.A. Berg, Phys. Rev. Lett. **90**, 180601 (2003).
[11] Z. Li and H.A. Scheraga, J. Mol. Struct.: THEOCHEM **179**, 333 (1988).
[12] M. Kinoshita, Y. Okamoto, and F. Hirata, J. Am. Chem. Soc. **120**, 1855 (1998).
[13] Min-yi Shen and K.F. Freed, Biophys. J. **82**, 1791 (2002).
[14] Y. Peng, U.H. Hansmann, and N.A. Alves, J. Chem. Phys. **118**, 2374 (2003); Y. Peng and U.H. Hansmann, Biophys. J. **82**, 3269 (2003).
[15] C.-Y. Lin, C.-K. Hu, and U.H. Hansmann, Proteins: Struct., Funct., Genet. **52**, 436 (2003).
[16] F. Eisenmenger, U.H. Hansmann, S. Hayryan, and C.-K. Hu, Comput. Phys. Commun. **138**, 192 (2001).
[17] G.J. Geyer, in *Proceedings of the 23rd Symposium on the Interface*, edited by E.M. Keramidis (Interface Foundation, Fairfax, VA, 1991), pp. 156–163.
[18] K. Hukusima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[19] U.H. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[20] T. Ooi, M. Obatake, G. Nemethy, and H.A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **84**, 3086 (1987).
[21] D. Eisenberg, M. Wesson, and M. Yamashita, Chem. Scr. **29**, 217 (1989).
[22] A. Kim, dissertation, University of California at San Francisco, San Francisco, 1990.
[23] J. Vila, R.L. Williams, M. Vásquez, and H.A. Scheraga, Proteins: Struct., Funct., Genet. **10**, 199 (1991).
[24] L. Wesson and D. Eisenberg, Protein Sci. **1**, 227 (1992).
[25] C.A. Schiffer, J.W. Caldwell, P.A. Kollman, and R.M. Stroud, Mol. Simul. **10**, 121 (1993).
[26] B. von Freyberg, T.J. Richmond, and W. Braun, J. Mol. Biol. **233**, 275 (1993).
[27] M.J. Sippl, G. Nemethy, and H.A. Scheraga, J. Phys. Chem. **88**, 6231 (1984), and references therein.
[28] B. Hingerty, R.H. Richie, T.L. Ferrel, and J. Turner, Biopolymers **25**, 427 (1985).
[29] Y. Okamoto, Biopolymers **34**, 529 (1994).
[30] J.L. Fauchere and V. Pliska, Eur. J. Med. Chem. **18**, 369 (1983).
[31] S. Cabani, P. Gianni, V. Mollica, and L. Lepori, J. Solution Chem. **10**, 563 (1981).
[32] R. Wolfenden, L. Andersson, P.M. Cullis, and C.C.B. Southgate, Biochemistry **20**, 849 (1981).
[33] S. Brandt, *Statistical and Computational Methods in Data Analysis* (North-Holland, Amsterdam, 1983).
[34] A. Sokal, in *Functional Integration: Basics and Applications*, edited by C. DeWitt-Morette, P. Cartier, and A. Folacci (Plenum, New York, 1997), pp. 131–192.
[35] M. Masuya and Y. Okamoto, quoted in Y. Okamoto, Recent Res. Dev. Pure Appl. Chem. **2**, 1 (1998).
[36] W.H. Graham, E.S. Carter II, and R.P. Hicks, Biopolymers **32**, 1755 (1992).
[37] By mistake 1357 instead of 1913 was reported in Ref. [10].