Two-dimensional multiple-histogram method applied to isothermal–isobaric Monte Carlo simulations of molecular clusters

Aleš Vítek [a,*], René Kalus [a,b]

[a] Center of Excellence IT4Innovations, VSB - Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic
[b] Department of Applied Mathematics, VSB - Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic

ARTICLE INFO

ABSTRACT

A two-dimensional multiple-histogram method for isothermal–isobaric ensemble is discussed in detail, implemented for isothermal–isobaric Monte Carlo simulations of molecular clusters, and employed in a case study on phase changes in pure water clusters containing 15 through 21 water molecules. Full phase diagrams of these clusters are reported in the temperature–pressure plane over a broad range of temperatures ($T = 30$–800 K) and pressures $P = 10^3$–10^9 Pa. The main focus of the work is on the structural transformation occurring in the solid phase of these clusters and leading from cluster structures with all molecules on the cluster surface to cage-like structures with one molecule inside, and on how the transformation is influenced by increased pressure and temperature.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Numerical simulations play a key role in cluster science since they provide a deep insight into the physics and chemistry of these objects which is not directly accessible to experimental investigations. Even at the classical level, numerical simulations can provide valuable qualitative as well as quantitative data. Among two main branches of molecular simulations methods, molecular dynamics methods and Monte Carlo methods, the latter represent a computationally cheaper alternative and have been frequently used in the past. Up to now, however, free clusters have usually been studied using canonical Monte Carlo approaches [1–6]. This may ultimately provide complete information about the evolution of the properties of various clusters along the temperature axis for temperatures for which the evaporation of monomers from the cluster is negligible. A more detailed approach based, e.g., on isothermal–isobaric Monte Carlo simulation techniques [7] has been much less frequently applied despite the fact that it may provide a basically complete information about the cluster phase behavior in the temperature–pressure plane.

A couple of problems are closely linked to numerical simulations. The most important problems may be (a) how to achieve a sufficient convergence of simulated data and avoid quasiergodicity, which becomes particularly important if a coexistence region of two phases of a particular system is simulated, and (b) how to obtain values of the quantities of interest (like internal energy, heat capacity etc.) at thermodynamic conditions (temperatures, pressures etc.) not included in particular simulation runs. Both problems have been addressed in the past and efficient procedures have been developed to surmount them. *Parallel-tempering* techniques, proposed originally for canonical, constant-temperature simulations [8] and extended later to isothermal–isobaric calculations [7], have been developed to deal with the quasiergodicity problem and have soon become a simulation standard, at least for canonical, constant-temperature studies. To deal with latter problem, reliable interpolation schemes must be used since the size of the grids of simulated temperatures and pressures is usually strongly limited due to computational demands. *Multiple-histogram* techniques have proven to be very efficient in this respect. The original multiple-histogram approach by Ferrenberg and Swendsen [9] was first developed, similar to the evolution of the parallel-tempering methodology, for canonical, constant-temperature simulations and only recently has been generalized to be used in isothermal–isobaric calculations [10], building on the *weighted histogram analysis method* [11], an extension of the original multiple-histogram method [9] towards multi-parameter simulations. If combined with isothermal–isobaric Monte Carlo calculations, the two simulation methods, parallel tempering and multiple histograms, seem to be well suited for getting converged thermodynamic data of the system under study for basically arbitrary values of the temperature and pressure at acceptable computational costs. Since almost smooth dependences

* Corresponding author. Tel.: +420 597321441.
E-mail addresses: ales.vitek@vsb.cz (A. Vítek), rene.kalus@vsb.cz (R. Kalus).

of thermodynamic parameters on the temperature and pressure, e.g., for the system heat capacity or internal energy, can be obtained in this way, they can be used in a detailed analysis of phases and phase transitions in the system [12,2], leading eventually to a full phase diagram in the temperature–pressure plane.

The main intention of this paper is to test the performance of combined parallel-tempering and multiple-histogram techniques in a case study on selected medium-sized water clusters, $[H_2O]_N$, with N ranging between 15 and 21 water molecules. The motivation for this choice is that, according to previous studies (see, e.g., Refs. [13–17]), a structural transformation occurs in isolated water clusters with $N \approx$ 16–21 molecules at zero temperature and zero pressure. This transformation consists in a gradual transition from the most stable structures having all water molecules on the surface of the cluster (*all-surface* structures) to structures with one molecule in the center of the cluster and with the remaining molecules forming a cage around it (*cage-1* structures). Let us briefly summarize the evolution of this observation. A first systematic study on the classical equilibrium structures of water clusters at zero temperature and zero pressure was reported for cluster sizes up to $N = 21$ by Wales and Hodges [13]. In their study, the authors used the simple TIP4P potential [18] for modeling the intra-cluster interactions and the *basin-hopping* methodology for cluster structures optimizations. They concluded from their calculations that up to $N = 21$ all the classical equilibrium structures of $[H_2O]_N$ are of the all-surface type except for $N = 19$, for which a disordered cage-like structure was reported as the most stable one. This conclusion was a bit later confirmed for the TIP4P water clusters by Hartke in a study employing a modified *genetic algorithms* approach [14]. In addition to the most stable, all-surface structures, Hartke found, however, that some metastable isomers of $[H_2O]_N$, $N = 16$–18 and $N = 21$, were of the cage-1 type and that they were by 183 meV above the most stable structure for $N = 16$, but only by 22 meV, 48 meV, and 15 meV above the most stable structures for $N = 17$, 18, and 21, respectively. In the following studies [15,16], Hartke repeated the calculations using a more advanced TTM2-F potential [19] and found that, for this potential, the cage-1 structures are preferred in the $N = 16 - 21$ range. More specifically, the all-surface structures were found as the most stable ones for $N = 18$ and 20, while the cage-1 structures became most stable for $N = 17$, 19 and 21. Consequently, the all-surface to cage-1 transition was reported to start at $N = 17$, to span the interval of cluster sizes of $N = 17$–21, and probably to continue up to $N = 25$. Interestingly, another transition to cage-2 structures (two molecules inside) was reported to start at $N \approx 28$. A more recent study using another advanced interaction model (ABEEM/MM) [20], confirmed the observations of Refs. [15,16] up to $N = 24$ and only led to different predictions for $N \geq 25$. For example, the formation of cage-2 structures was predicted in this study to start at $N \approx 33$. A benchmark study on equilibrium structures of $[H_2O]_N$ for the size range of $N = 17$–21 was reported from the group of Xantheas [17]. The calculations of this work were performed at the quantum chemistry level and employed the MP2 correlation method and augmented correlation-consistent basis sets of double and triple zeta quality. This state-of-the-art calculations confirmed the earlier results obtained for the TTM2F model [15,16]; namely, that the all-surface structures are the most stable ones for $N = 18$ and 20, while for $N = 17$, 19 and 21 the most stable structures are of the cage-1 type. To our best knowledge, no results have been reported up to now on the all-surface to cage-1 transition in water clusters of this size range going beyond the zero-temperature and, in particular, beyond the zero-pressure limit. In this work, we try to partly fill this gap. Moreover, we show that the methodology based on the combination of the parallel

tempering approach and multiple histograms can provide a firm base for such an analysis. The 15-molecules water cluster is also considered here since a detailed analysis of its thermodynamic behavior has been presented in a preceding study [21] which may be useful for interpreting the phase behavior of larger cluster sizes.

The paper is organized as follows. First, a summary of the method, code implementation, and computational details are given in Section 2; second, computational data are presented and discussed in Section 3 with the main focus on phase changes in water clusters of considered sizes and, in particular, on the all-surface to cage-1 transition in the solid phase; and finally, conclusive remarks are given in Section 4.

2. Methods and computations

2.1. 2D multiple-histogram method

In the isothermal–isobaric ensemble, the residual part of the mean value of a dynamical parameter, F, is expressed at particular temperature T and pressure P as an ensemble average,

$$\langle F \rangle_{T,P} = \frac{\int_V \int_q F(q, V) \exp\left(-\frac{E^{\text{int}}(q)+PV}{k_B T}\right) dq dV}{\int_V \int_q \exp\left(-\frac{E^{\text{int}}(q)+PV}{k_B T}\right) dq dV}, \tag{1}$$

where q and V are the system configuration and volume, respectively, E^{int} denotes the interaction energy of the system, and k_B is the Boltzmann constant. If the dynamical parameter depends on the system configuration through the interaction energy only,

$$F = F\left(E^{\text{int}}(q), V\right), \tag{2}$$

the complicated high-dimensional integral on the r.h.s. of Eq. (1) can be replaced by a computationally much cheaper two-dimensional (2D) integration,

$$\langle F \rangle_{T,P} = \frac{\int_V \int_{E^{\text{int}}} F(E^{\text{int}}, V) \exp\left(-\frac{E^{\text{int}}+PV}{k_B T}\right) \Omega(E^{\text{int}}, V) dE^{\text{int}} dV}{Z_{T,P}}, \tag{3}$$

where $\Omega(E^{\text{int}}, V)$ is the classical density of states and

$$Z_{T,P} = \int_V \int_{E^{\text{int}}} \exp\left(-\frac{E^{\text{int}} + PV}{k_B T}\right) \Omega(E^{\text{int}}, V) dE^{\text{int}} dV \tag{4}$$

is the system configuration integral. Since $\Omega(E^{\text{int}}, V)$ is a function of two independent variables, we denote it hereafter as *two-dimensional (2D) density of states* to distinguish it from the one-dimensional density of states, $\Omega(E^{\text{int}})$, considered in constant-volume calculations.

While the 2D integral on the r.h.s. of Eq. (3) can be rather simply calculated using any method of numerical quadrature, the calculation of $\Omega(E^{\text{int}}, V)$ may represent a rather tough task. However, a sufficiently accurate estimate of $\Omega(E^{\text{int}}, V)$ can be extracted at acceptable computational costs from two-dimensional energy–volume histograms collected from isothermal–isobaric simulations performed over a sufficiently broad range of temperatures and pressures. The methodology for doing so is based on the *multiple-histogram* (MH) approach proposed originally for canonical, constant-volume calculations [9] and extended later to the isothermal–isobaric case [10]. The method basically consists of two phases. First, histograms of E^{int} and V are collected from a series of isothermal–isobaric simulations carried out at various temperatures ($T = T_1, \ldots, T_N$) and pressures ($P = P_1, \ldots, P_M$). We call this initial phase the *production phase* since the production of 2D energy–volume histograms is the main task of this first step of the calculation. Second, the histograms recorded in the

production step are used in an iterative calculation of the 2D density of states over a (not necessarily equispaced) grid of interaction energies ($E^{\text{int}} = E_1, \ldots, E_A$) and volumes ($V = V_1, \ldots, V_B$) and the density of states is then used in the calculation of required ensemble averages. Since the main task of this second phase is the evaluation of the 2D density of states followed by the evaluation of corresponding ensemble averages, we call it the *evaluation phase*.

Clearly, the production phase is much more time-consuming than the evaluation phase since it usually consists of dozens or even hundreds of sufficiently long thermodynamic simulations. The evaluation of the density of states represents then the more time consuming part of the evaluation phase and requires much more computer time than the final calculation of the ensemble averages (a simple 2D integration).

In the first, *production phase*, normalized histograms of energy and volume are to be calculated for a series of temperatures T_n and pressures P_m. The histograms are defined as follows,

$$h_{nm}(E_a, V_b)$$
$$= \lim_{\Delta E_a, \Delta V_b \to 0} \frac{\int_{E \in [E_a, E_a + \Delta E_a]} \int_{V \in [V_b, V_b + \Delta V_b]} \exp\left(-\frac{E_a(q) + P_m V_b}{k_B T_n}\right)}{\Delta E_a \Delta V_b} dq dV, \quad (5)$$

and can be obtained, e.g., from isothermal–isobaric Monte Carlo (MC) simulations,

$$h_{nm}(E_a, V_b) \approx \frac{1}{\Delta E_a \Delta V_b} \frac{\mathcal{N}_{nm}^{ab}}{\mathcal{N}}, \quad (6)$$

where $\mathcal{N}_{nm}^{ab}$ is the number of samples recorded at $T = T_n$ and $P = P_m$ with the interaction energy $E^{\text{int}} \in [E_a, E_a + \Delta E_a]$ and volume $V \in [V_b, V_b + \Delta V_b]$, and $\mathcal{N}$ corresponds to the total number of samples generated during the particular simulation.[1]

Provided the energy–volume histograms have been calculated, the 2D density of states is obtained during the *evaluation phase* by iteratively solving the following set of coupled equations,

$$\Omega(E_a, V_b) = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} g_{nm}^{-1} h_{nm}(E_a, V_b)}{\sum_{n=1}^{N} \sum_{m=1}^{M} g_{nm}^{-1} \exp\left(-\frac{E_a + P_m V_b}{k_B T_n}\right) Z_{nm}^{-1}}, \quad (7)$$

$$Z_{nm} = \sum_{a=1}^{A} \sum_{b=1}^{B} \exp\left(-\frac{E_a + P_m V_b}{k_B T_n}\right) \Omega(E_a, V_b), \quad (8)$$

where $g_{nm} = 1 + 2\tau_{nm}$ and τ_{nm} is the autocorrelation time of the simulation performed at $T = T_n$ and $P = P_m$.[2] The iterative process starts from a plausible initial guess of Z_{nm} (see later in this section) and, then, the following three steps are repeated until convergence: (1) the current estimates of Z_{nm} are inserted to Eq. (7) and values of $\Omega(E_a, V_b)$ are updated; (2) the updated values of $\Omega(E_a, V_b)$ are then used in Eq. (8) to update the estimates of Z_{nm}; and finally, (3) a normalization of Z_{nm} is invoked,

e.g., by

$$Z_{nm} \to \frac{Z_{nm}}{\sqrt{\sum_{n=1}^{N} \sum_{m=1}^{M} Z_{nm}^2}}, \quad (9)$$

to prevent divergence.[3]

After a sufficiently converged estimate of the density of states has been obtained, the final calculation can be carried out by numerically evaluating the integrals on the r.h.s. of Eq. (3). This can be done, e.g., via a simple rectangle quadrature, leading eventually to

$$\langle F \rangle_{T,P} = \frac{\sum_{b=1}^{B} \sum_{a=1}^{A} F(E_a, V_b) \exp\left(-\frac{E_a + PV_b}{k_B T}\right) \Omega(E_a, V_b) \Delta E_a \Delta V_b}{\sum_{b=1}^{B} \sum_{a=1}^{A} \exp\left(-\frac{E_a + PV_b}{k_B T}\right) \Omega(E_a, V_b) \Delta E_a \Delta V_b}, \quad (10)$$

providing Ω has been pre-calculated over a sufficiently dense grid of energies E_a and volumes V_b.

2.2. Computational details

Medium-sized water clusters, $[H_2O]_N$ with $N = 15–21$, are used in the present work to demonstrate the performance of the methodology described in the previous subsection. All the simulation settings are thus related to these specific molecular systems. A simple rigid-body TIP4P potential [18] has been used to model the water–water interaction.

A parallel-tempering Monte Carlo method for the isothermal–isobaric ensemble [7] with periodic swaps of configurations and volumes between copies of the simulated system has been used in the *production phase* of the present work. A detailed description of the method was presented elsewhere [21] and, thus, only brief remarks related to the present implementation are given here. The MC simulations have been performed in parallel for 18 different temperatures ranging between 30 and 1000 K and 20 pressures spanning the interval between 10^{-3} and 10^3 MPa[4], the temperatures and pressures included are given in Table 1. In total, 360 copies of the system have been simulated in parallel for each cluster size. About 5×10^7 MC whole-cluster & volume moves have been generated for each system copy during the production phase from which only every 50th move has been used in gathering the energy–volume histograms to avoid non-physical

---

[1] In our approach, the total number of samples is the same for different simulation runs, which is enforced by using the *parallel-tempering* scheme in our calculations to accelerate their convergence, *cf.* below in this section.

[2] In this work, $g_{nm} = 1$, because, first, the parallel-tempering approach used in this work frequently mixes samples generated at different temperatures and pressures and reduces thus significantly their serial correlations, and second, a number of configuration–volume pairs are thrown away between two samples considered in the accumulation of the energy–volume histograms, a usual procedure employed in Monte Carlo calculations to suppress unphysical correlations.

[3] Note that the solution of Eqs. (7) and (8) is unique only up to a multiplicative factor. The normalization considered in this step represents just a specific choice of this factor. This arbitrary choice does not affect the value of the calculated ensemble average, however, since the factors cancel out in the denominator and the numerator of the formula on the r.h.s. of Eq. (10).

[4] The number of temperatures and pressures included in a particular simulation depends on the system size and is always a trade-off between maximizing the acceptance ratio of swapping moves and computational demands. Basically, the grids of temperatures and pressures must be dense enough so that the energy–volume histograms of neighboring temperature–pressure pairs overlap to a sufficient extent. This means, e.g., that the grids must be denser at low temperatures where the histograms are very narrow. A lot of preliminary test calculations have been performed to tune the scales of temperatures and pressures with respect to the acceptance ratio of swapping moves. In general, the average exchange rate of swapping moves decreases with the size of the simulated system. But, even for the most adverse case of $(H_2O)_{21}$, the average exchange rate for both pressure and temperature swaps has usually been above 10% in our calculations. Only for a few combinations of temperatures and pressures, the exchange rate has been lower (several per cent). This has not imposed a serious convergence problem, however, since the 2D swapping pathways in the temperature–pressure plane easily circumvent such bottlenecks. It is also worth mentioning that the acceptance ratio of swaps between different pressures does not significantly decrease at low temperatures.

Table 1
Temperatures T_n and pressures P_m considered in the present calculations. All possible combinations of indices n and m give the complete set of systems simulated in parallel.

n	T_n (K)	m	P_m (MPa)
1	30	1	0.001
2	38	2	0.003
3	47	3	0.01
4	62	4	0.021
5	80	5	0.044
6	104	6	0.1
7	132	7	0.21
8	166	8	0.44
9	203	9	1
10	240	10	2.1
11	261	11	4.4
12	294	12	10
13	330	13	30
14	400	14	100
15	490	15	150
16	600	16	200
17	750	17	300
18	1000	18	500
		19	700
		20	1000

correlations. Another 5×10^7 MC moves have been discarded at the beginning of each simulation to reach thermodynamic equilibrium. The swapping step, which interchanges configurations and volumes of two randomly selected neighboring systems,[5] has been invoked periodically after each completed whole-cluster & volume move. The external pressure has been approximately modeled by a hard-wall spherical container with a variable radius [21]. To prevent an uncontrolled growth of the volume of the container at low pressures and high temperatures, an upper limit of the container radius has been set to $R_{[H_2O]_{15}} = 150$ Å for $[H_2O]_{15}$ and to $R_{[H_2O]_N} = R_{[H_2O]_{15}} * (N/15)^{1/3}$ for larger clusters containing N water molecules.

During the *evaluation phase*, the 2D density of states has to be obtained by iteratively solving Eqs. (7)–(9). This may be a rather demanding task since, first, the number of simulated systems (and thus the number of calculated histograms entering the evaluation phase) is now considerably larger than in the case of constant-volume simulations and, second, each of the 2D energy–volume histograms is much larger than the 1D energy histograms of the isothermal–isochoric multiple-histogram (MH) method. Moreover, narrow bins are needed for both energy and volume to take into account very small fluctuations at low temperatures. In general, thousands of such bins would be needed for both energy and volume for the medium sized clusters considered in this work and, consequently, totally millions of bins will constitute the resulting 2D histograms. While the former technical issue (the number of simulated systems) can be resolved only by employing powerful computers, the latter problem (the size of the histograms) can be effectively removed by using non-uniform grids.[6] The use of the non-uniform grids is justified by a simple observation that energy fluctuations are small just at low temperatures, where only small regions of the system configuration space corresponding to the lowest energies accessible to the system are visited, and that volume fluctuations

become small only at very high pressures for which small and only slightly different volumes are sampled. Consequently, the energy grid must be dense only near the minimum of the energy of the system and the volume grid only near the minimum of its volume. Sparse grids are sufficient, on the other hand, elsewhere, in particular, in high-energy and large-volume regions. Following the results of our tests we performed previously [21] on $(H_2O)_{15}$, we expect that the size of the energy–volume histograms can be reduced by three to six orders of magnitude if non-uniform grids are employed (depending on simulated system and considered temperature–pressure interval).

In the present work, exponential spacings have been used for both energies and volumes,

$$E_A - E_{A-1} = k_E(E_2 - E_1), \qquad V_B - V_{B-1} = k_V(V_2 - V_1), \qquad (11)$$

where A and B are respectively the numbers of energies and volumes included in the grids, E_1 and V_1 are the lowest energy and volume, respectively, recorded during the simulation, and E_A and V_B the largest recorded values. Constant parameters k_E and k_V are adjusted iteratively as follows. First, instantaneous values of energy and volume are recorded during the production phase. Then, the recorded values of energy and volume are used to find the maximum and minimum values of energy, E_A and E_1, and the maximum and minimum values of volume, V_B and V_1. Next, A and B are set to a small initial value ($A \approx 100$ and $B \approx 1000$ in this work) and k_E and k_V initialized by $k_E = 10$ and $k_V = 10$. An estimate of the 2D density of states is then obtained via Eqs. (7)–(9) for this setting and for an essentially arbitrary initial guess of the configuration integrals, $Z_{nm} = 1/\sqrt{NM}$ in this work; a fairly fast calculation. The 2D density of states calculated in this way is then used in a subsequent evaluation of selected thermodynamic averages (enthalpy, heat capacity, and Pearson correlation coefficient for energy and volume) via Eq. (10) (again, a very fast calculation). The calculated averages are then compared to the averages obtained directly from MC simulations and k_E and k_V are re-adjusted according tho the differences between the MC and MH data. If the MH results deviate considerably from the MC data at low temperatures, larger non-uniformity is needed in the energy grid and k_E is increased. If, on the other hand, the MH results deviate a lot from the MC data at high temperatures, more uniform grid on the energy axis is to be used and k_E is decreased. Similar rules apply to the volume grid adjustment: considerable differences at high pressures mean that the value of k_V must be increased and differences at low pressures indicate that a smaller value of k_V should be used. After the re-adjustments of k_E and k_V, the procedure is repeated with the initial guess of the configuration integrals taken from the previous step until the MH and MC data coincide within a given error bar. At this point, one gets close-to-optimal values of k_E and k_V which are, according to our tests, only slightly dependent on the number of grid points used. For the cluster sizes and temperature and pressure intervals considered in this work, optimal values of $k_E \approx 500$ and $k_V \approx 10^5$ have been obtained. As a final step, the sizes of the energy and volume grids are increased (by a factor of ≈ 5 and the final evaluation of the classical density of states is performed starting from the initial values of the configuration integrals taken from the last iteration of the preceding calculations on the sparse grid. This notably reduces the computational demands of this last step because only a few iterations are needed for the enlarged energy–volume histograms to get a converged estimate of the 2D density of states.[7] As the very final step, the MH averages are compared to the direct MC data to verify the quality of the calculated density of states.

---

[5] Under the neighboring systems we understand any two systems considered in the simulation which differ by their temperature indices and by their pressure indices at most by one.

[6] The non-uniform grids do not change the equations used for the density of states calculations since the non-uniformity is considered in the definition of histograms, Eq. (5), as well as in the formula for the calculation of ensemble averages, Eq. (10).

[7] To provide a basic idea of the computational costs of the present calculations we briefly summarize as follows. Our simulation code, parallelized using the MPI protocol, has been compiled using the Intel compilers version 13.1.1 with the−
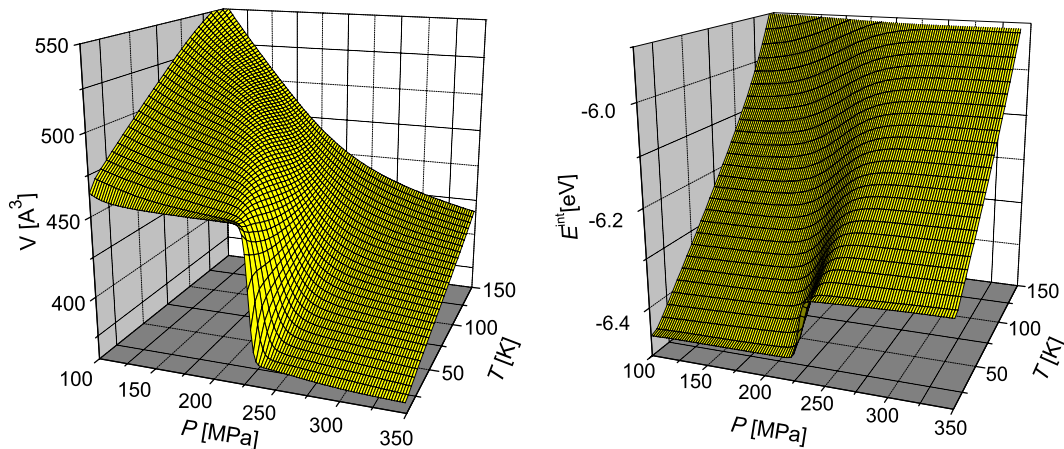
Fig. 1. Temperature and pressure dependence of the mean volume V (left) and the mean interaction energy E^{int} (right) calculated for $[H_2O]_{15}$ using the two-dimensional MH method.

3. Results and discussions

As already mentioned in previous sections, the 2D multiple-histogram methodology is employed here in a case study on medium-sized water clusters. More specifically, clusters containing 15 through 21 molecules are considered. This choice has been inspired by the fact that a structural transformation has been detected in this size range [13–17] at zero temperature and zero pressure from structures with all water molecules on the cluster surface (*all-surface* structures) to cage-like structures with one molecule inside the cluster (*cage-1* structures) (see Section 1 for more details). The main intention of the present work is to investigate how increased temperature and pressure affluence such a structural transformation. This is done through analyzing the phase transitions in the considered clusters over the temperature range of $T = 30$–800 K and the pressure range of $P = 10^{-3}$–10^3 MPa. Since the 2D multiple-histogram methodology provides values of thermodynamic parameters for which the underlying dynamical parameter obeys Eq. (2) at, in principal, arbitrary temperature and pressure, almost continuous dependences of the thermodynamic parameters used to detect the phase changes can be obtained and almost continuous coexistence curves[8] plotted in the form of two-dimensional, temperature–pressure phase diagrams.

A common tool used for the detection of phase changes along the temperature axis is the *heat capacity*. It develops, for finite systems, a more or less apparent maximum wherever the system enters a coexistence region. In this work, the residual part of the constant-pressure heat capacity is used for this purpose,

$$C_P = \left[\frac{\partial \langle H^{\text{int}} \rangle}{\partial T} \right]_P = \frac{1}{k_B T^2} \left[\left\langle \left(H^{\text{int}} \right)^2 \right\rangle - \left\langle H^{\text{int}} \right\rangle^2 \right], \quad (12)$$

where $H^{\text{int}} = E^{\text{int}} + PV$. In addition, the Pearson correlation coefficient for volume and energy,

$$\rho_{E^{\text{int}}, V} = \frac{\text{cov}(E^{\text{int}}, V)}{\sigma_{E^{\text{int}}} \sigma_V}, \quad (13)$$

with "cov" denoting the covariance operator and $\sigma_{E^{\text{int}}}$ and σ_V being the variances of the interaction energy and volume, respectively, has been proven previously [21] a sensitive parameter for pinpointing structural transitions along the pressure axis, particularly between different solid phases of a cluster. As detailed in the previous study, if such a transition occurs, the energy–volume Pearson correlation coefficient becomes negative (close to -1 at low temperatures) and develops a well apparent minimum if depicted as a function of the pressure. This is because the interaction energy and volume of the system become strongly anti-correlated in the coexistence region, the correlations being stronger at low temperatures. This behavior is illustrated for $[H_2O]_{15}$ in Fig. 1. Another feature which is well apparent on $\rho_{E^{\text{int}}, V}(T)$ dependences is a maximum developing in the evaporative region of the cluster (with $\rho_{E^{\text{int}}, V}$ being close to $+1$). This means that, with the onset of the evaporation, an abrupt increase of the energy resulting from each evaporative event is accompanied by a simultaneous increase of the volume of the cluster due to the increased distance of the evaporated particle from the rest of the cluster. As a consequence, the Pearson correlation coefficient can also be used, in addition to the heat capacity, for detecting the evaporative, liquid–gas transition in clusters.

3.1. Phase diagram of $[H_2O]_{15}$

As an illustration of the general explanation given above, plots of the cluster heat capacity and the Pearson correlation coefficient against temperature and pressure are shown for $[H_2O]_{15}$ in Fig. 2. All the features discussed in the preceding paragraph are clearly seen in this figure. First, the two ridges visible on the heat-capacity landscape (the right-hand-side panel of Fig. 2) correspond either to a solid–liquid transition (max1) or to a liquid–gas transition (max2) occurring in this cluster. Second, the sharp valley apparent on the Pearson correlation coefficient surface at low temperatures (min in the left-hand-side panel of Fig. 2) corresponds to a structural change in the solid phase of $[H_2O]_{15}$ leading from all-surface structures at low pressures to more compact cage-1 structures at high pressures, as described in detail in Ref. [21]. And finally, the maximum occurring on the Pearson correlation coefficient plot at high temperatures (max in the left-hand panel of Fig. 2) represents the liquid–gas transition in $[H_2O]_{15}$.

O2 optimization option and run subsequently on an x86-64 SandyBridge hardware architecture. About 10,000 core-hours have been consumed during the *production phase* for $(H_2O)_{15}$ which increased to about 22,000 core-hours for $(H_2O)_{21}$. The computational costs thus scale approximately as $N^{2.3-2.5}$ with the cluster size. If compared to the *production phase*, much less computer time was needed for the *evaluation phase* calculations, just about 500 core-hours per one cluster size. However, memory requirements are very high in the latter case (about 10 GB per calculation).

[8] In finite systems, the phase changes are not sharp and occur over a more or less wide region of temperatures and pressures, *coexistence region*. Following our earlier work [21], we represent these regions in a simplified way by curves corresponding to the maxima of the heat capacity along the temperature axis and the minima of the Pearson correlation coefficient for energy and volume detected along the pressure axis. We call these curves *coexistence curves*.
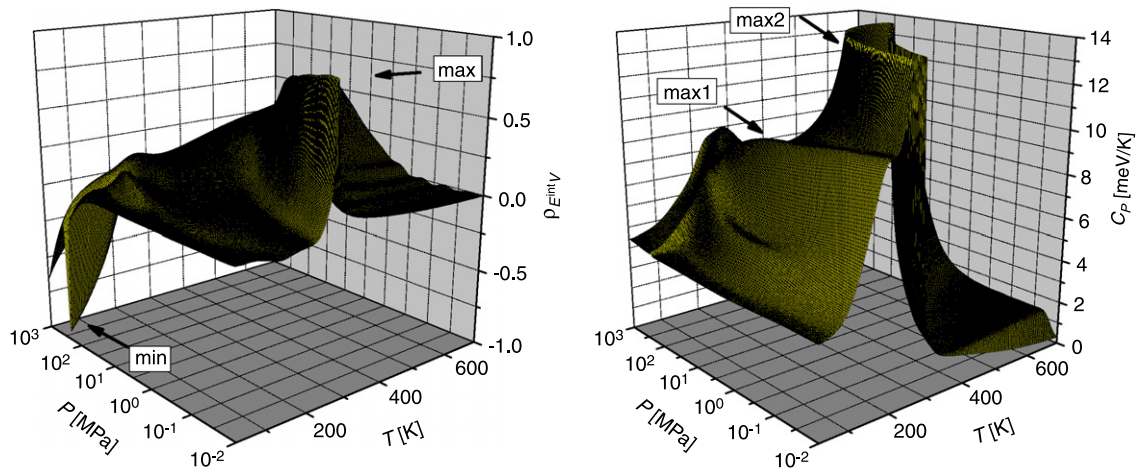
Fig. 2. Temperature and pressure dependence of the energy–volume Pearson correlation coefficient (left) and constant-pressure heat capacity (right) calculated for $[H_2O]_{15}$ using the two-dimensional MH method. Note that, pressures are given in the logarithmic scale.

The positions of the maxima on the heat-capacity profiles calculated numerically using $[\partial C_P^{res}/\partial T]_P = 0$, the positions of the minima of the Pearson correlation coefficient obtained from $[\partial \rho_{E^{int},V}/\partial P]_T = 0$, and the positions of the maxima of the Pearson correlation coefficient obtained from $[\partial \rho_{E^{int},V}/\partial T]_P = 0$ have been used to extract the coexistence curves separating different phases of $[H_2O]_{15}$. They are depicted in the form of a temperature–pressure phase diagram of $[H_2O]_{15}$ in Fig. 3. Since this phase diagram is a typical representative of analogous plots reported later in this section for larger cluster sizes, a bit more detailed description is due here. First of all, big black dots are used in Fig. 3 (and also in subsequent figures) to depict the coexistence curves obtained from the maxima on the heat capacity landscape of $[H_2O]_{15}$ (max1 and max2 in the right-hand-side panel of Fig. 2) while big gray dots correspond to the extremal points on the Pearson correlation coefficient profile (min in the left-hand-side panel of Fig. 2). In addition, small black dots correspond to maxima on the Pearson coefficient landscape (max in the left-hand-side panel of Fig. 2). The coexistence curves divide the temperature–pressure plane into several regions for which we use the same nomenclature as employed in the bulk. It is worth emphasizing here once more, however, that the coexistence curves are just a simplified representation of much broader coexistence regions typical for finite systems and that the analogy with the bulk nomenclature is thus only loose.

At low temperatures, the cluster is *solid* (S regions in Fig. 3) and only a few akin structural isomers, which are close in energy, are detected here for each particular pressure. The structure of these isomers changes suddenly, however, at pressures around $P \approx 200$–300 MPa where a transition from all-surface geometries found at low pressures (SI) to cage-1 geometries detected at high pressures (SII) takes place. This is indicated by a deep valley in the Pearson correlation coefficient landscape (min in the left-hand-side panel of Fig. 1).

At temperatures around $T \approx 200$–300 K (higher temperatures are required for higher pressures), the cluster becomes *liquid* (L), which is indicated by the first, low-temperature chain of maxima on the heat capacity surface (max2 in the right-hand-side panel of Fig. 1). In addition to the few solid-like isomers detected at low temperatures, a plethora of disordered structures emerges now in $[H_2O]_{15}$ and they become more and more abundant as the temperature increases. At even higher temperatures, *gas-like* disordered structures develop (G region in Fig. 3) and a liquid–gas transition occurs. This transition is indicated by either the second chain of maxima on the cluster heat capacity surface (max2 in
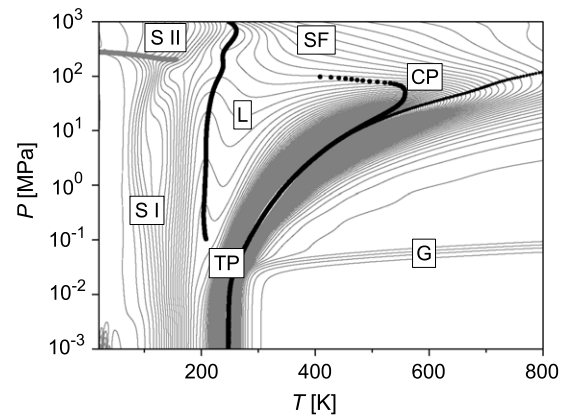

Fig. 3. Phase diagram of $[H_2O]_{15}$. Big black dots are used to depict the coexistence curves obtained from the maxima on the heat capacity landscape of $[H_2O]_{15}$ (max1 and max2 in the right-hand-side panel of Fig. 2), big gray dots correspond to the minima on the Pearson correlation coefficient profile (min in the left-hand-side panel of Fig. 2), and small black dots correspond to the Pearson coefficient maxima (max in the left-hand-side panel of Fig. 2). SI and SII denote solid phase regions, L and G are used for liquid and gas phase, respectively, SF for supercritical fluid, TP labels the triple point region, and CP denotes the critical point.

the right-hand-side panel of Fig. 1) or by the high-temperature ridge on the Pearson correlation coefficient surface (max in the left-hand-side panel of Fig. 1).

Interestingly, the two coexistence curves we obtained for the liquid–gas transition by either using the heat capacity or the Pearson correlation coefficient for energy and volume coincide well up to $P \approx 10$ MPa, but deviate significantly at higher pressures. This divergence of the two curves is not observed, however, if the coexistence curve for the liquid–gas transition is calculated from the position of the high-temperature maxima of the enthalpy fluctuation itself, $\sigma_{H^{int}}^2 = \left\langle \left(H^{int}\right)^2 \right\rangle - \left\langle H^{int} \right\rangle^2$, instead from the heat capacity, in which an additional multiplicative factor, $1/k_B T^2$, is present (cf. Eq. (12)). It can be easily shown by simple algebra that this additional factor is responsible for the shift of the heat capacity maxima with respect to the maxima of $\sigma_{H^{int}}^2$. Moreover, the flatter the maximum of $\sigma_{H^{int}}^2$ is, the larger is the shift. And this exactly occurs at high pressures where the evaporation peak observed on the temperature dependence of the heat capacity becomes extremely wide and flat. As a consequence, a divergence occurs at higher pressures between the liquid–gas coexistence curve calculated from $\sigma_{H^{int}}^2$ (or
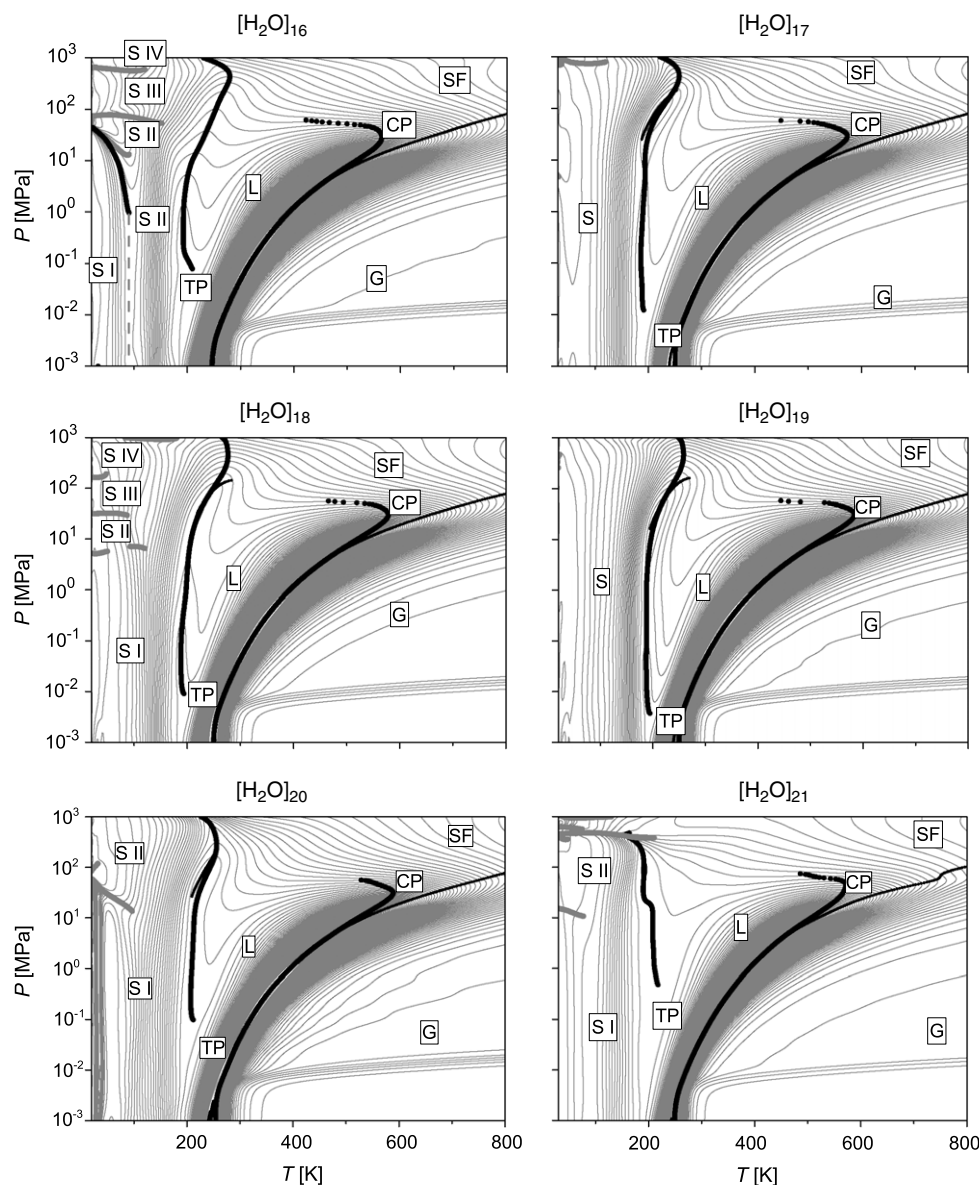
Fig. 4. Temperature–pressure phase diagrams of $[H_2O]_{16}$–$[H_2O]_{21}$. Symbols are the same as in the Fig. 3.

the curve obtained from the Pearson correlation coefficient) and the coexistence curve calculated using the cluster heat capacity. It seems thus that the enthalpy fluctuation may represent a more realistic parameter for detecting the liquid–gas transition in the high-pressure region than the heat capacity itself.

At very high pressures ($P \approx 100$ MPa and above), the second maximum on the heat capacity landscape (or the corresponding maximum for $\sigma^2_{H_{int}}$) ceases to exist. The cluster approaches its *critical-point region* (CP in Fig. 3) where the difference between the liquid phase and the gaseous phase of the cluster disappears. Something like a *supercritical fluid* phase (SF) emerges here.

At low pressures ($P \approx 10^{-1}$ MPa for $[H_2O]_{15}$), the low-temperature maximum on the heat capacity surface (max1 in the right-hand-side panel of Fig. 1) also disappears and transforms gradually to a hump approaching gradually the second, evaporative maximum (max2). This means that the liquid phase of $[H_2O]_{15}$ becomes unstable at low pressures, the melting and evaporation coexistence intervals overlap significantly on the temperature axis, and the solid–liquid and liquid–vapor transitions merge. The $[H_2O]_{15}$ cluster approaches its *triple-point region* (TP in Fig. 3), a

finite-size analogue of the macroscopic triple point. If the pressure is once again increased, however, the mean volume of the liquid phase of the cluster decreases significantly and the liquid phase becomes again stable. The low-temperature hump on the heat capacity surface transforms back to a well pronounced peak and the system leaves its triple-point region.

3.2. Phase diagrams of $[H_2O]_{16}$–$[H_2O]_{21}$

The phase diagrams we calculated for larger cluster sizes of $[H_2O]_N$ are collected in Fig. 4. As is clear from this figure, the layouts of these diagrams are generally similar to what was recorded for $[H_2O]_{15}$. As a consequence, similar conclusions to those reported for $[H_2O]_{15}$ in the preceding subsection will also be valid for larger cluster sizes. Shortly, the coexistence curves in general divide, for each cluster, the T–P plane into three main regions, a region where the cluster is solid (S in Fig. 4) and where usually only a few structural isomers come into play, another region where the cluster becomes liquid (L) and where abundant families of disordered structural isomers become important, and

finally, a region where the cluster enters the gaseous phase (G) and where its structure becomes completely disordered. Further, the coexistence curves for the solid–liquid transition and the liquid–gas transition become very close to each other at low pressures where the clusters enter their triple-point regions (TP), and the solid–liquid coexistence curve disappears at high pressures where the clusters pass their critical-point regions (CP) and transform into supercritical fluid (SF). In this respect, the behavior of all the clusters considered in Fig. 4 is very similar to each other and the only differences between them are visible in the "fine" structure of the solid regions. Since the main focus of the present work is just on the structural transitions in the solid phase of $[H_2O]_N$, a bit more detailed analysis of these S regions is pertinent here.

3.2.1. $[H_2O]_{16}$

An all-surface structure consisting of three fused cubes was reported for $[H_2O]_{16}$ from zero-pressure and zero-temperature structure optimization calculations using the TIP4P potential [13]. Exactly only this isomer, slightly distorted by thermal oscillations, has been detected in our calculations at low pressures and low temperatures (SI region in the $[H_2O]_{16}$ panel of Fig. 4). However, if the temperature is only slightly raised (SI–SII coexistence curve), a few other local minima on the $[H_2O]_{16}$ potential energy surface (PES) come gradually into play and become finally dominant in the SII region. Since still only a few structural isomers dominate in the SII region, we conclude that the cluster undergoes a solid–solid transition. All of the structures of $[H_2O]_{16}$ detected in the SII region are of the all-surface form, and, among them, two pentagonal edge-sharing prisms, which are only 60–80 meV above the global fused-cubes minimum, completely prevail. Noteworthy, a pentagonal edge-sharing structure was reported as the most stable one in a previous study employing a more realistic interaction model [20]. The SI–SII transition develops a weak maximum on the heat capacity curve of $[H_2O]_{16}$ at pressures between $P \approx 1$–50 MPa (solid SI–SII coexistence curve in panel $[H_2O]_{16}$ of Fig. 4) and this maximum reduces to a hump on the ascending wing of the much stronger solid–liquid transition peak at lower pressures (dashed SI–SII curve).

Another solid–solid transition has been detected at pressures of $P \approx 50$–80 MPa leading from the SII phase at lower pressures to a new (SIII) phase at higher pressures. The newly formed SIII phase consists of a few cage-1 isomers and the SII→SIII transition represents thus the all-surface to cage transformation. The cage isomers of the SIII phase are much higher in energy than the all-surface structures found at low pressures (e.g., about 140 meV above the global, three-fused-cubes minimum corresponding to phase SI), but are much more compact. They are thus favored by the PV term at sufficiently high pressures. One more solid–solid transition has been detected in $[H_2O]_{16}$ in our calculations at the highest pressures considered ($P \approx 1$ GPa). This transition leads to a new family of a few compact, cage-1 minima (phase SIV) which are, however, very similar to those found in the SIII phase.

In summary, the $[H_2O]_{16}$ cluster adopts the cage-1 structure at pressures above $P \approx 75$ MPa for $T \to 0$ K or even at $P \approx 50$ MPa at higher temperatures. Below this pressure, the dominant structures are of the all-surface form; however, the most stable, three-fused-cubes isomer disappears already at much lower pressures and is thus insignificant in the all-surface to cage-1 transition. Interestingly, Hartke previously reported [14] a cage-like local minimum for TIP4P $[H_2O]_{16}$ with classical energy about 180 meV above the global minimum energy. This cage structure by Hartke is about 40 meV above the cage structures we detected in the SIII phase of $[H_2O]_{16}$ in our calculations and about 70 meV below the structures we found in the SIV phase.

3.2.2. $[H_2O]_{17}$

While an all-surface structure was previously reported as the most stable one for $[H_2O]_{17}$ and for the TIP4P potential [13], more realistic interaction models predict that a cage-1 structure is more stable [15–17,20]. According to the reports employing the latter models, the size of $N = 17$ frames the beginning of the all-surface to cage-1 transition in $[H_2O]_N$ from below, at least at zero temperature and pressure. It follows from our calculations that the solid phase of $[H_2O]_{17}$ (the S region in panel $[H_2O]_{17}$ of Fig. 4) consists of a mixture of two isomers, the TIP4P global minimum and another, the local minimum which is halfway to a cage-1 structure and approximately 20 meV above the global minimum. "Halfway" means here that one of the water molecules is significantly pushed towards the center of the cluster, but a full clathrate-like structure does not develop. The cluster is open from the side of the intruding molecule and it seems as if part of the cluster surface was just squeezed inside the cluster. Interestingly, this metastable structure has a comparable energy to that reported for the lowest clathrate-like geometry of $[H_2O]_{17}$ by Hartke [14]. Since the abundance of the halfway cage-1 isomer is significant even at low pressures, one can conclude that the $[H_2O]_{17}$ cluster represents a first step along the all-surface to cage-1 transition even for the simple TIP4P potential. In addition, the global, all-surface minimum disappears soon at pressures $P \approx 1$–10 MPa and the S phase becomes completely dominated by the halfway cage-1 isomer.

At very high pressures, an authentic cage-1 structure appears and replaces the low-pressure halfway cage-1 isomer. This is witnessed by a minimum on the Pearson energy–volume correlation coefficient plotted against the pressure at $P \approx 750$–900 MPa. The high-pressure cage-1 structure is, however, fairly similar to the halfway cage-1 structure observed at lower pressures. The high-pressure cage only differs from the low-pressure halfway cage by (a) the central molecule being fully wrapped by the surrounding molecules and (b) by a bit smaller volume.

3.2.3. $[H_2O]_{18}$

For $[H_2O]_{18}$, an all-surface structure consisting of a two-layer pentagonal prism with two edge-sharing cubes attached from opposite sides is predicted to be most stable for both the TIP4P potential [13] as well as for more realistic interaction models [15–17].[9] It follows from our calculations that this most stable isomer dominates at low temperatures up to pressures between $P \approx 5$–7 MPa (the SI region in panel $[H_2O]_{18}$ of Fig. 4) and is gradually replaced by another all-surface structure (two edge-sharing pentagons with a double-edge-sharing cube) at slightly higher pressures (the SII region). This first solid–solid structural change in $[H_2O]_{18}$ is clearly witnessed by the first chain of minima on the Pearson energy–volume correlation coefficient which appears at low temperatures and projects to the T–P plane as the SI–SII coexistence borderline.

At $P \approx 30$ MPa, another structural transformation takes place which leads now from the low-pressure all-surface structures to a cage-1 structure and which corresponds to another valley on the Pearson correlation coefficient surface (SII–SIII borderline). Above this pressure (the SIII region), the cage-1 structure becomes dominant and only another transformation to a more compact arrangement (still of the cage-1 like form) takes place at ($P \approx 160$–200 MPa) (the SIV region of the $[H_2O]_{18}$ phase diagram). The cage-1 like structures observed in the SIII and SIV regions are however, alike for smaller cluster sizes, very similar to each other.

---

[9] Except for the ABEEM/MM model for which a hexagonal prism was reported as the global minimum structure [20]. This has not been reproduced, however, by other studies.

The cage structure reported for the $[H_2O]_{18}$ cluster by Hartke [14] has the binding energy between the energies of the two cage structures found in our calculations, the SIII structure is about 20 meV below Hartke's cage and the SIV structure about 20 meV above.

3.2.4. $[H_2O]_{19}$

Both the TIP4P empirical potential [13] as well as more advanced interaction models [15–17] predict a disordered cage-1 like structure as the most stable form of $[H_2O]_{19}$ at zero pressure and zero temperature. Our calculations allow to extend this conclusion to pressures up to $P = 1$ GPa. Moreover, no solid–solid phase change has been detected in this range of pressures (see panel $[H_2O]_{19}$ of Fig. 4) and the zero-pressure isomer survives thus as the dominant structure up to very high external pressures.

3.2.5. $[H_2O]_{20}$

For $[H_2O]_{20}$, four families of energetically close minima have been reported for the TIP4P potential [13]. Within each family, the oxygen skeleton geometry is preserved and the members differ only in the orientations of dangling hydrogens. All of these minima are of the all-surface form, however. The structures built on the lowest minimum oxygen-skeleton geometry consist of three edge-sharing pentagonal prisms and are about 0–37 meV above the lowest minimum. Other two families of low-energy $[H_2O]_{20}$ minima consist of four fused cubes or face-sharing pentagonal prisms and are respectively about 37–42 and 59–64 meV above the lowest minimum. The last family of local minima reported for $[H_2O]_{20}$ in Ref. [13] are of the dodecahedron geometry. They are nevertheless energetically too rich (about 485–490 meV above the global minimum) and are expected to play only a marginal role at low temperatures.

In our calculations, we have mostly observed isomers belonging to the global minimum family at low pressures (the SI region of panel $[H_2O]_{20}$ in Fig. 4). It is the only structure populated at the lowest temperature considered ($T = 30$ K) and it survives as a completely dominating structural motif up to $T \approx 60$ K. In addition to this most abundant isomer, the fused-cubes structures have also been detected in our calculations at a slightly increased temperature ($T \geq 40$ K) as well as a (very) small amount of face-sharing pentagonal prisms at $T \geq 50$ K. As expected, the dodecahedron structures have not occurred in our simulations in the solid phase of $[H_2O]_{20}$. It is worth mentioning that fairly high barriers separating different isomers were reported previously for $[H_2O]_{20}$ [22] which might lead to non-ergodic behavior and loose convergence of thermodynamic simulations. Since all the three low-lying isomers have been found in our calculations, one can conclude that the parallel-tempering methodology we employed is very efficient for preventing these adverse effects.

At pressures between $P = 15$ MPa and $P = 60$ MPa (higher pressures are required for lower temperatures), a transition occurs from the all-surface structures to cage-1 structures and is accompanied by a chain of minima on the Pearson correlation coefficient landscape (the SI–SII gray borderline in the $[H_2O]_{20}$ panel of Fig. 4). Above these pressures, the cage-1 like forms completely prevail in $[H_2O]_{20}$ and no further structural transformation has been detected in this cluster in our calculations up to $P = 1$ GPa.

3.2.6. $[H_2O]_{21}$

The classically most stable structure of $[H_2O]_{21}$ reported for the TIP4P potential is of the all-surface form [13,14] while more realistic interaction potentials as well as correlated *ab initio* calculations predict the cage-1 form to be more stable than the all-surface one [15–17,20]. But, Hartke [14] found that there are local minima on the TIP4P PES of $[H_2O]_{21}$ which have the cage-1 shape and are only by ≥ 15 meV above the most stable, all-surface structure. Since the cage-like geometries corresponding to these local minima are much more compact than the all-surface global minimum structure, they can be expected to become important even at not very high pressures.

In fact, exactly this has been observed in our calculations. While the global minimum structure prevails in the solid phase of $[H_2O]_{21}$ for the lowest pressures (region SI in the $[H_2O]_{21}$ panel of Fig. 4), the lowest cage-1 isomer found by Hartke emerges already at $P \approx 15$ MPa and becomes completely dominant slightly above this pressure (region SII in the $[H_2O]_{21}$ panel of Fig. 4). The structural transition in the solid phase of $[H_2O]_{21}$ gives rise to a valley in the landscape of the Pearson correlation coefficient for volume and energy along the $P \approx 15$ MPa line. However, this valley is rather broad and shallow in this case, particularly for higher temperatures (though not too much). The roots of such behavior may be either due to the fact that (a) the pressure at which the transition occurs is not high enough to give rise to sufficiently strong anti-correlations between energy and volume of the cluster or (b) that the energy difference between the all-surface isomer and the lowest cage-1 isomer is small enough for the cage-1 isomer to become abundant on the lower-pressure side of the SI–SII coexistence curve even at slightly elevated temperatures. Since the boundary between the SI phase and the SII phase of $[H_2O]_{21}$ is not well defined, the corresponding coexistence curve is plotted in a dashed pattern in panel $[H_2O]_{21}$ of Fig. 4.

Much more compact cages come into play at higher pressures (they are about 40 meV above the global minimum isomer for $P \approx$ 200–300 MPa or even 280–500 meV above the global minimum isomer for pressures above $P \approx 500$ MPa) and give rise to a series of structural transitions in $[H_2O]_{21}$ represented in the $[H_2O]_{21}$ panel of Fig. 4 by a series of coexistence curves lying between $P \approx$ 300 MPa and $P \approx 1$ GPa. Noteworthy, these coexistence curves correspond to deep and narrow valleys in the energy–volume Pearson correlation coefficient landscape and are, thus, distinctly defined.

4. Conclusions

A two-dimensional (2D) [10] multiple-histogram (MH) [9] methodology for isothermal–isobaric simulations has been discussed in detail and employed in a case study of phase changes in pure water clusters containing 15–21 molecules. The main motivation for performing this specific investigation consists in an observation that, at zero temperature and zero pressure, a structural change occurs in water clusters of these sizes leading from structures with all molecules on the cluster surface (*all-surface* structures) to cage structures with one molecule inside and the remaining molecules forming an envelope around it (*cage-1* structures) [13–17,20]. The main focus of the present work is twofold. First, the power of the 2D MH method is demonstrated by calculating smooth numerical data over extended ranges of temperatures and pressures at practicable computational costs, and, second, a detailed analysis is provided of how the structural transformation in the medium-size water clusters is influenced by increased pressures and temperatures. For this purpose, full phase diagrams have been constructed for the considered clusters and for extended ranges of temperature and pressure ($T = 30$–800 K and $P = 0$–1000 MPa).

In our implementation, the 2D MH method consists of two basic steps. First, a series of parallel isothermal–isobaric parallel-tempering Monte Carlo [7] simulations are run for each particular cluster size for selected pressures and temperatures. During each simulation run, energies and volumes visited by the simulated system are recorded for subsequent calculations of energy–volume histograms (Eq. (6)). Typically, about 20 pressures and the same
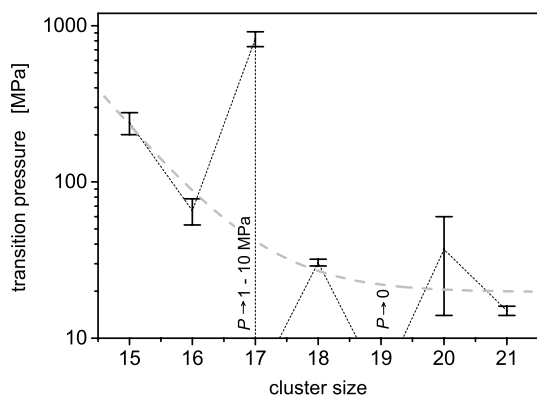
Fig. 5. Transition pressures for the *all-surface* to *cage-1* structural transformation in $[H_2O]_N$. Vertical segments are used to show the bounds of the transition pressures recorded at various temperatures, the gray dashed line is added to guide eyes and to illustrate the overall trend in the transition pressure evolution with cluster size.

number of temperatures have been considered in this work, which means that several hundred Monte Carlo simulations have been run in parallel. Second, the 2D classical density of states is calculated over a sufficiently dense grid of cluster energies and volumes by numerically integrating Eqs. (7)–(9) and following the procedure described in Section 2.2. After a reasonably accurate estimate of the 2D density of states has been obtained, a broad range of thermodynamic parameters can easily be calculated (Eq. (10)) over basically any set of temperatures and pressures. This is particularly important if one is interested in pinpointing phase changes in the system under study when either maxima on the heat capacity vs. temperature plots are to be found or, as shown in a preceding work [21], minima of the Pearson correlation coefficient for energy and volume are to be detected along the pressure axis. Particularly, the latter calculations require extremely dense grids of pressures since the minima on the Pearson correlation coefficient curves may be extremely narrow at low temperatures and can be easily missed.

Several implementation improvements have been applied in this work to save the computational time, memory, and disk space, in particular in the second, evaluation phase. For example, non-uniform grids of the cluster energy and volume, with the density of points decreasing exponentially with increasing energies and volumes, have been shown to reduce the required disk space considerably while preserving the accuracy of the resulting 2D density of states achievable only for impractically large uniform grids. Moreover, it has been shown that the degree of the non-uniformity of the energy–volume grids can be safely adjusted using sparse grids and only then employed in generating a large non-uniform grid to be used in the final calculation of the 2D density of states. This enables significant savings of the computer time.

As a specific test of the 2D MH methodology, numerical modeling of phase changes in TIP4P water clusters has been performed and phase diagrams of these clusters have been constructed in the temperature–pressure (T–P) plane. The overall structure of these phase diagrams is basically the same for all the cluster sizes considered. The T–P plane is in general partitioned into several regions corresponding to the solid phase, liquid phase, and gaseous phase of the clusters. The solid–liquid and liquid–gas coexistence curves merge at low pressures in a triple-point region and the liquid–gas curve disappears at high pressures in the critical-point region (beyond which the clusters exist in the form of supercritical fluid). The only significant difference between the phase diagrams of $[H_2O]_{15}$–$[H_2O]_{21}$ consists in different "fine structures" of the solid-phase regions. More specifically, these regions are further divided into several parts by coexistence curves corresponding to pressure-induced solid–solid phase changes among which

the phase changes leading from all-surface structural motifs found at low pressures to cage-like structures which become dominant at high pressures are most important. This compares well with the observation reported previously from zero-temperature and zero-pressure structure optimization calculations that the global minima on the TIP4P surfaces of $[H_2O]_N$ are of the all-surface type [13,14]. However, one finds that the cage structures become more stable (and much more abundant) than the all-surface structures, even for the TIP4P potential, if non-zero external pressures are considered. For the smallest cluster size considered in this work, $[H_2O]_{15}$, rather high pressures are needed for such a transformation, but as the cluster size increases the transition pressure decreases considerably. For a summary of pressures at which $[H_2O]_N$ change their structure from all-surface geometries to cages see Fig. 5.

The evolution of the transition pressure is not smooth. There are two markable exceptions to the general picture described in the preceding paragraph. First, $[H_2O]_{19}$ has been reported to have, even at $T = 0$ K and $P = 0$ Pa, a distorted cage-1 like global minimum [13]. This has been confirmed by our calculations which further show that this global minimum structure survives as the dominating low-temperature structural motif up to very high pressures. No solid–solid phase change has been detected in $[H_2O]_{19}$ in our calculations up to $P = 1$ GPa. Second, the $[H_2O]_{17}$ cluster has been reported to have an all-surface global minimum on the TIP4P potential energy surface at $T = 0$ K and $P = 0$ Pa [13]. However, it follows from our calculations that even at very low pressures another isomer, corresponding to a local minimum of a "halfway" cage-1 shape, is present in the solid phase of $[H_2O]_{17}$, in addition to the most stable all-surface isomer, and becomes completely prevalent at not too high pressures of $P \approx 1$–10 MPa. As a consequence, one can conclude that the "halfway" cage structure, though not being the global minimum structure of $[H_2O]_{17}$, is an important component of the solid phase of this cluster at non-zero, though rather low pressures. Noteworthy, a transition to a full cage-1 structure then occurs in $[H_2O]_{17}$ only at very high pressures ($P \approx 750$–900 MPa).

In summary, the 2D MH method as implemented here represents an efficient tool for detailed thermodynamic modelings of complex molecular systems, the investigation of phase changes in molecular clusters being a particularly illustrative example. Nonetheless, even though reduced significantly, computational demands still remain rather high for such studies and highly powerful computers are to be used. This particularly holds for larger cluster sizes (usually encountered in experiments) and for more realistic interaction models (polarizable and/or flexible molecules) than that used in our work. Further simulations are being prepared to enter this numerically demanding field.

Acknowledgments

References

[1] C.J. Tsai, K.D. Jordan, J. Chem. Phys. 95 (1991) 3850.
[2] C.J. Tsai, K.D. Jordan, J. Chem. Phys. 99 (1993) 6957.
[3] D.J. Wales, I. Ohmine, J. Chem. Phys. 98 (1993) 7245.
[4] J.M. Pedulla, K.D. Jordan, Chem. Phys. 239 (1991) 593.
[5] A.N. Tharrington, K.D. Jordan, J. Chem. Phys. 107 (2003) 7380.
[6] J. Yin, D.P. Landau, J. Chem. Phys. 134 (2011) 074501.
[7] T. Okabe, M. Kawata, Y. Okamoto, M. Mikami, Chem. Phys. Lett. 335 (2001) 15.
[8] C.J. Geyer, Proceedeings of the 23rd Symposium on the Interface in Computing Science and Statistics, American Statistical Association, New York, 1991.
[9] A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. 63 (1989) 1195.
[10] A. Pal, S.K. Roy, Phys. Rev. E 69 (2004) 021709.
[11] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, J. Comput. Chem. 13 (1992) 1011.
[12] A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. 61 (1988) 2635.
[13] D.J. Wales, M.P. Hodges, Chem. Phys. Lett. 286 (1998) 65.
[14] B. Hartke, Zeit. Phys. Chem. 214 (2000) 1251.
[15] B. Hartke, Eur. Phys. J. D 24 (2003) 57.
[16] B. Hartke, Phys. Chem. Chem. Phys. 5 (2003) 275.
[17] A. Lagutschenkov, G.S. Fanourgakis, G. Niedner-Shatteburg, S.S. Xantheas, J. Chem. Phys. 122 (2005) 194310.
[18] W.L. Jorgensen, J.J. Tirado-Rivers, J. Am. Chem. Soc. 110 (1988) 1657.
[19] C.J. Burnham, S.S. Xantheas, J. Chem. Phys. 116 (2002) 5115.
[20] P. Qian, L. Lu, W. Song, Z. Yang, Theor. Chem. Acc. 123 (2009) 487.
[21] A. Vitek, R. Kalus, Phys. Chem. Chem. Phys. 14 (2012) 15509.
[22] D.J. Wales, Mol. Phys. 102 (2004) 891.