Using force-based adaptive resolution simulations to calculate solvation free energies of amino acid sidechain analogues

Raffaele Fiorentini, Kurt Kremer, Raffaello Potestio, and Aoife C. Fogarty

Citation: The Journal of Chemical Physics **146**, 244113 (2017); doi: 10.1063/1.4989486
View online: http://dx.doi.org/10.1063/1.4989486
View Table of Contents: http://aip.scitation.org/toc/jcp/146/24
Published by the American Institute of Physics

Using force-based adaptive resolution simulations to calculate solvation free energies of amino acid sidechain analogues

Raffaele Fiorentini,[a] Kurt Kremer,[b] Raffaello Potestio,[c] and Aoife C. Fogarty[d]
Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

(Received 15 February 2017; accepted 30 May 2017; published online 28 June 2017)

The calculation of free energy differences is a crucial step in the characterization and understanding of the physical properties of biological molecules. In the development of efficient methods to compute these quantities, a promising strategy is that of employing a dual-resolution representation of the solvent, specifically using an accurate model in the proximity of a molecule of interest and a simplified description elsewhere. One such concurrent multi-resolution simulation method is the Adaptive Resolution Scheme (AdResS), in which particles smoothly change their resolution on-the-fly as they move between different subregions. Before using this approach in the context of free energy calculations, however, it is necessary to make sure that the dual-resolution treatment of the solvent does not cause undesired effects on the computed quantities. Here, we show how AdResS can be used to calculate solvation free energies of small polar solutes using Thermodynamic Integration (TI). We discuss how the potential-energy-based TI approach combines with the force-based AdResS methodology, in which no global Hamiltonian is defined. The AdResS free energy values agree with those calculated from fully atomistic simulations to within a fraction of k_BT. This is true even for small atomistic regions whose size is on the order of the correlation length, or when the properties of the coarse-grained region are extremely different from those of the atomistic region. These accurate free energy calculations are possible because AdResS allows the sampling of solvation shell configurations which are equivalent to those of fully atomistic simulations. The results of the present work thus demonstrate the viability of the use of adaptive resolution simulation methods to perform free energy calculations and pave the way for large-scale applications where a substantial computational gain can be attained. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4989486]

I. INTRODUCTION

One of the most challenging applications of computational methods in biochemistry is the accurate calculation of solvation and binding free energies. A prototypical example is provided by *in silico* drug design, where one needs to obtain, by means of computational experiments, quantitative information about the effectiveness of a new molecule or set of molecules in promoting or inhibiting a given enzyme. It is often the case that the number of viable candidates to become usable drugs is too large for experimental screenings, where the complexity of the processes under examination makes it difficult to dissect the observed system properties into its different components.

Computer simulations represent a valuable tool, as they enable the pre-screening of a large number of different systems and the comprehension of their properties at the molecular and atomic levels. This detailed information can prove crucial to identify the most promising molecules, thus allowing experimental research to focus on a reduced subset of case studies.

However, the detailed determination of ligand-enzyme binding free energies still remains a daunting task in most cases, due to the large size of the molecules under examination. In particular, a considerable bottleneck can be the simulation of the solvent, which might represent a substantial fraction of the computational cost.

A promising way of mitigating the computational overhead due to the explicit solvent molecules is to employ concurrent multi-resolution simulation methods. These use a combination of computationally expensive high-resolution potentials and cheaper low-resolution potentials simultaneously in order to facilitate the study of systems in which a large range of time and length scales play a role. The accurate high-resolution model is used to describe those parts of the system where fine-grained or chemically detailed processes take place, while use of the less expensive coarse-grained (CG) potential in the rest of the system allows bigger system sizes and longer simulations.

One such multi-resolution method is the adaptive resolution scheme (AdResS),[1] in which the simulation box is divided into atomistic (AT) and coarse-grained regions, with particles[2] smoothly changing their resolution on-the-fly as they move between regions. This resolution change is achieved by the interpolation of energies[3] or forces[4] across a transition region. The AdResS methodology allows a significant reduction in the number of degrees of freedom simulated atomistically, while still reproducing the properties of a sub-region of a fully atomistic simulation.[1] In the past decade, most works using the AdResS approach have concentrated on the study of

[a]Electronic mail: fiorentini@mpip-mainz.mpg.de
[b]Electronic mail: kremer@mpip-mainz.mpg.de
[c]Electronic mail: potestio@mpip-mainz.mpg.de
[d]Electronic mail: fogarty@mpip-mainz.mpg.de

structural and sometimes dynamical properties, as well as basic thermodynamic quantities such as density, pressure, chemical potential, or compressibility.[4–11] So far, less attention has been paid to how well free energies can be computed within an AdResS setup. Recent explorations of the thermodynamics of AdResS include Refs. 12 and 13. In particular, Agarwal *et al.* compared chemical potentials calculated as an intrinsic side-product of their grand canonical AdResS setup to free energies of solvation calculated in fully atomistic systems.[14]

Here we introduce the combination of the force-based AdResS approach and Thermodynamic Integration[15] (TI) to calculate free energies. We obtain solvation free energies of amino acid sidechain analogues in water, a set of classic systems studied notably by Shirts *et al.*[16] and also recently employed in an exploration of a non-adaptive multi-resolution technique, in which Kuhn *et al.* discussed the influence of density deviations and orientational edge effects on the solvation free energy in that approach.[17]

In our AdResS setup, we describe the solute molecule and surrounding solvent molecules using an atomistic potential, while the rest of the system is modeled at a cheaper, coarse-grained level. We explore the influence of atomistic region size, coarse-grained potential, and density control on the free energy. We also discuss how the potential-energy-based thermodynamic integration approach combines with the force-based AdResS methodology, in which no global Hamiltonian is defined. We show that because AdResS allows the sampling of atomistic configurations which are equivalent to those of fully atomistic simulations in the equivalent ensemble, we can nevertheless accurately calculate free energy values with this approach.

These results demonstrate that the usage of the force-based AdResS method in tandem with Hamiltonian-based free energy calculations is viable and quantitatively sound. This validation paves the way to large-scale applications involving large macromolecules and, therefore, large amounts of explicit solvent to be treated at dual resolution.

II. METHODOLOGY

In this work, we calculate the solvation free energy of amino acid sidechain analogues methanol and 3-methylindole (analogues of serine and tryptophan, respectively). These two molecules were chosen because they have significantly different sizes: methanol has a fairly similar size to water and a molar mass of 32.04 g mol^{-1}, while 3-methylindole has a molar mass of 131.18 g mol^{-1}. The radii of gyration, 0.08 nm for methanol and 0.21 nm for 3-methylindole, give an indication of the size difference. Each simulation system used contains one solute molecule in aqueous solution. We perform fully atomistic reference simulations with a range of box sizes and AdResS simulations with a range of different atomistic region sizes and two different coarse-grained potentials for water. The first coarse-grained potential used is derived via the systematic coarse-graining procedure Iterative Boltzmann Inversion (IBI)[18–20] to reproduce as closely as possible the atomistic water centre-of-mass structure. In the second case, the coarse-grained region contains a gas of non-interacting particles, i.e., an ideal gas, which can be seen as the most extreme possible coarse-grained "potential."[21]

A. Adaptive resolution scheme and thermodynamic integration

In the AdResS methodology [illustrated in Fig. 1(a)], the simulation box is divided into different regions: the atomistic (AT) region, where non-bonded interactions are modeled using an atomistic force field, and the coarse-grained (CG) region, where a coarse-grained force field is used. Between them is a hybrid (HY) region where particles smoothly change their resolution between atomistic and coarse-grained. In this work, the AT region is a sphere of radius r_{at} centered on an atom with coordinates $\mathbf{r}_{centr}$ at the center of the solute molecule. With this definition, the AT region follows the solute and moves together with it: this strategy, which relies on the translational invariance of the uniform solvent with periodic boundary conditions, makes it unnecessary to restrain the molecule in a particular point of the simulation box.

The HY region is a spherical shell of width d_{hy}, and the remainder of the system is the CG region. Water molecules diffuse freely between regions, changing resolution as a function of their instantaneous position. Particle resolution is described using a function w that varies smoothly and monotonically across the HY region, from a value of 1 in the AT region to 0 in the CG region. For a molecule α whose center of mass $\mathbf{r}_\alpha$ is at a distance $r = |\mathbf{r}_{centr} - \mathbf{r}_\alpha|$ from the center of the AT region, it has the form

$$w(r) = \begin{cases} 1, & r < r_{at} \\ \cos^2\left(\frac{\pi}{2d_{hy}}(r - r_{at})\right), & r_{at} < r < r_{at} + d_{hy} \\ 0, & r_{at} + d_{hy} < r \end{cases}. \quad (1)$$

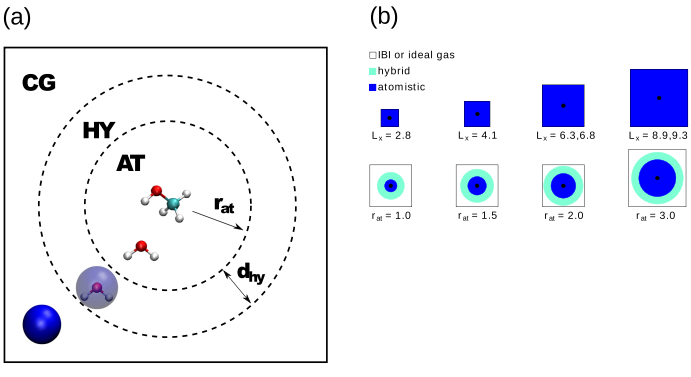

FIG. 1. (a) Illustration of the AdResS approach, showing the atomistic, hybrid, and coarse-grained regions (not to scale). (b) Schematic representation of fully atomistic and AdResS systems used. The methanol or 3-methylindole solute is represented by a black dot in the centre of the box. (c) Up: methanol chemical structure and down: 3-methylindole chemical structure. Atomistic details are shown in red, blue, cyan, and white (O, N, C, and H atoms).

Note that in many studies which use the AdResS methodology, the resolution function w is referred to using the symbol λ. We write w here to avoid confusion with the order parameter λ used in thermodynamic integration.

Non-bonded interaction forces are then modeled using a force-interpolation scheme, in which the intermolecular force between the centers of mass of molecules α and β is given by

$$\mathbf{F}_{\alpha\beta} = w(\mathbf{r}_\alpha)w(\mathbf{r}_\beta)\mathbf{F}^{AT}_{\alpha\beta} + [1 - w(\mathbf{r}_\alpha)w(\mathbf{r}_\beta)]\mathbf{F}^{CG}_{\alpha\beta}, \quad (2)$$

where

$$\mathbf{F}^{AT}_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta}\mathbf{F}^{AT}_{ij}, \quad (3)$$

where $\mathbf{F}^{AT}_{ij}$ is the atomistic non-bonded interaction between atoms i and j, and $\mathbf{F}^{CG}_{\alpha\beta}$ is the coarse-grained non-bonded interaction between molecules α and β. In this scheme, the forces interacting between two atomistic water molecules simplify to $\mathbf{F}^{AT}_{\alpha\beta}$ and between two coarse-grained water molecules to $\mathbf{F}^{CG}_{\alpha\beta}$. Water-water interactions across the resolution boundaries are treated using interpolation Eq. (2). This scheme allows simulations which are momentum-conserving but not energy-conserving. The global system Hamiltonian corresponding to Eq. (2) is not defined,[22] and a local thermostat must be applied to deal with heat production in the HY region.[4,23]

The dual-resolution treatment of the solvent allows a reduction of the computational cost of the simulations. The systems under examination here have a relatively small size and, most importantly, a relatively large *atomistic-to-total* volume ratio, meaning that the fraction of volume where molecules are treated at the atomistic level is relatively large. Because of this, in the simple test systems studied here, the computational gain is not large, and indeed a major speedup is not the goal of the present investigation. Nonetheless, we could observe that the AdResS method provided a substantial reduction in simulation time with respect to a fully atomistic simulation. This gain, quantitatively reported in Table I, is defined as the inverse of the ratio of atomistic-to-total volume. The latter is obtained as the volume of a sphere of radius $R = r_{at} + d_{hy}$ divided by the simulation box volume. The obtained speedup is somewhat lower than the corresponding ideal value; however, the discrepancy diminishes as the volume where atomistic forces are computed increases. This behavior stems from the approximation on which the definition of ideal AdResS simulation time relies, namely, that the only computational cost is due to the calculation of forces and that this takes place only in the atomistic and hybrid regions. This assumption willfully neglects surface and finite size effects, hence the deviations for systems with small atomistic regions. For the setups with the smallest atomistic regions, the speedup is between ≃3 (methanol) and ≃3.6 (methylindole).

TABLE I. Comparison of the speedup in simulation time provided by AdResS simulations with respect to fully atomistic simulations of the same size (namely, 6.3 nm side for the methanol and 6.8 nm side for 3-methylindole) run on a single core. These data are obtained from 19 ps long runs, in order to minimize the idle time employed in non-run processes (system setup, memory allocation, etc.). These speedups are also compared to the ideal ones, defined as the inverse of the fraction of AdResS system volume where the calculation of atomistic forces takes place.

R	Ideal sim speedup	AdResS sim speedup
	Methanol	
1.0	5.6	3.1
1.5	3.0	2.4
2.0	1.8	1.7
	3-methylindole	
1.0	7.1	3.6
1.5	3.7	2.6
2.0	2.3	2.3

We calculate solvation free energies using the Thermodynamic Integration (TI) method.[15] For any two states A and B in which the solute-solvent interaction differs, we write the solute-solvent interaction potential U_{sw} as a function of an order parameter λ which takes values between 0 and 1, defining a pathway from state A to B. The free energy difference between the states is then given by

$$\Delta G = \int_0^1 \left\langle \frac{\mathrm{d}U_{sw}(\lambda, q)}{\mathrm{d}\lambda} \right\rangle_\lambda \mathrm{d}\lambda. \quad (4)$$

In practice this is done by discretising λ and sampling $\mathrm{d}U_{sw}(\lambda, q)/\mathrm{d}\lambda$ for a series of different λ values between 0 and 1.

We now address one perceived possible problem. TI involves derivatives of the potential energy with respect to the parameter λ, while in force-based AdResS no global Hamiltonian is defined.[22] For the calculation of solvation free energies, the energy derivative required is that of the potential energy of the interaction between the solute and solvent, since all other energy terms in the system are independent of λ. This is defined in AdResS as long as all atoms in all pairs contributing to $\mathrm{d}U_{sw}/\mathrm{d}\lambda$ fall within the AT or HY region. Moreover, the value of $\langle \mathrm{d}U_{sw}/\mathrm{d}\lambda\rangle_\lambda$ will be the same in the fully atomistic and AdResS systems as long as two conditions are fulfilled: (i) all interaction pairs contributing to U_{sw} fall within the AT region (i.e., the interaction cutoff plus the solute size is less than r_{at}) and (ii) both systems sample the same ensemble of configurations in the atomistic region. We will show below that this is indeed the case.

There also exists a formulation of AdResS (called H-AdResS) based on the interpolation of energies instead of forces.[3] In this case, a global Hamiltonian is defined and simulations are energy-conserving. We anticipate that this formulation of AdResS can also be used without problems in TI calculations. However, using H-AdResS with moving atomistic regions is inadvisable because the forces in H-AdResS involve a term which is a derivative of the resolution function; in particular cases, that force term could create additional spurious forces on that atom. This would happen if the position of the atomistic region were made mathematically dependent on the instantaneous position of a given atom, which would be necessary to have an isolated Hamiltonian with no external forces. This problem could be circumvented; however, it could overshadow the issues specific to the usage of Kirkwood TI in the context of adaptive resolution simulations. Because of these reasons, and since in the long term

we are interested in complex applications such as protein-ligand binding which will require the use of moving AT regions, we decided to validate the TI/force-based AdResS combination.

We also note at this point that in H-AdResS and in the auxiliary Hamiltonian approach of Agarwal *et al.*,[14] free energies (excess chemical potentials) can be obtained automatically as a by-product of the standard process of the system setup. However, this "by-product" approach applies only in the case of simple interactions between small molecules and can no longer be used for the calculation of free energies in more complex situations such as protein-ligand binding or interactions involving solids.

B. Thermodynamic force

In general, coarse-grained potentials cannot necessarily reproduce all thermodynamic properties of the atomistic reference potential which they are intended to represent.[24–26] In this work, we use a CG potential derived via iterative Boltzmann inversion and also simulate a system in which particles in the CG region are modeled using a gas of non-interacting particles. In both cases, the pressure of the CG potential differs significantly from that of the AT potential and would lead to an undesirable density difference between AT and CG regions. In order to avoid this, we use a thermodynamic force[27] F_T, a compensatory force which is applied within the HY region, ensuring a flat density profile along the direction of resolution change. F_T is generally obtained via an iterative procedure based on the gradient of the density profile along the direction of resolution change.[27] How straightforward it is to obtain this tabulated force depends on factors such as the thermodynamic difference between AT and CG potentials, atomistic region size and geometry, and concentration of different particle types in multicomponent systems.[28]

For the current purpose of calculating free energies, a very accurate density is required in the AT region. In order to reach this level of accuracy, even in the most difficult conditions (such as very small spherical atomistic region or large differences between AT and CG potentials in terms of thermodynamical properties such as pressure or compressibility), we have developed an upgraded algorithm to compute the thermodynamic force. Specifically, we include an additional term in the previously established procedure for obtaining F_T in the tabulated form and define the thermodynamic force at iteration $i + 1$ as

$$F_T(r)_{i+1} = F_T(r)_i - \kappa_1 \nabla \rho_i(r) + \kappa_2 (X_{ref} - X_i) \nabla w(r), \quad (5)$$

where $\rho_i(r)$ is the density profile along the direction of resolution change, calculated from a simulation using $F_T(r)_i$. In the newly added term $\kappa_2(X_{ref} - X_i)\nabla w(r)$, $w(r)$ is a function that goes smoothly from 1 to 0 across the region in which F_T is applied. We use the same functional form as given in Eq. (1) for $w(r)$ which defines the resolution change in AdResS, but this is just for convenience and there is no fundamental theoretical connection between them. The term X_i is a measure of the density throughout the atomistic region. It must be a well-defined value which can be determined with very high accuracy. In the current work, since the atomistic region is centered on an atom, the density profile is equivalent to a radial distribution, and we define X_i as the height of the first-solvation-shell peak. X_{ref} is the corresponding value in the fully atomistic reference system. Other measures of the density are possible and equally valid, for example, the average number of particles in the AT region. We found that the measure we used here [height of the first solvation-shell peak in the radial distribution function (RDF)] converged fastest as a function of simulation trajectory length and was therefore easiest to work with.

Finally, κ_1 and κ_2 are prefactors which can be varied to aid convergence. A useful procedure is to start with $\kappa_1 \neq 0$ and $\kappa_2 = 0$ and perform many iterations using relatively short, inexpensive simulations in order to rapidly obtain a good approximation of F_T. One can then set $\kappa_1 = 0$ and $\kappa_2 \neq 0$ and perform iterations with simulations long enough to determine X with high accuracy, continuing these iterations until the density in the atomistic regions is as close as desired to the reference density. These two steps can then be repeated as necessary.

Since finite-length simulations inevitably yield a density profile containing statistical noise which is then transmitted to the tabulated thermodynamic force, it can be helpful to use some procedure to smoothen the density profile $\rho(r)$, such as replacing each atom (which is a delta function, i.e., located at one defined point in space) by a triangle or Gaussian function, to smooth out its mass over several bins.[29]

We note in passing that it is also possible to obtain via IBI a coarse-grained potential with the same pressure as the atomistic reference;[19] however, this is at the cost of having the wrong compressibility in the coarse-grained region. Here, we chose to work with the non-pressure-corrected IBI potential, which has the same compressibility and structure as those for the atomistic potential. This provides a strong contrast to the other coarse-grained "potential" we use, the fluid of non-interacting particles, in which the structure, compressibility, and pressure all differ from the atomistic reference.

C. Simulation details

Fully atomistic and AdResS systems containing methanol or 3-methylindole were constructed using the simulation box sizes and atomistic region radii summarised in Table II and illustrated in Fig. 1(b). The box sizes range from a little over twice the non-bonded interaction cutoff to almost eight times the cutoff. The amino acid force field used was Amber94[30] (we note that the non-bonded parameters are the same as those in the more recent Amber force fields, which were mostly focussed on improvements in backbone parameters, not relevant here). The water model used was TIP3P.[31] Side chain analogue force fields were constructed from Amber94 amino acid residue force fields using the procedure of Ref. 16: the backbone atoms were replaced by a hydrogen of the same atom type and with the same charge as other hydrogen atoms connected to the β-carbon, and the β-carbon charge was adjusted so that the molecule was neutral overall. All other parameters were exactly as in the amino acid residue force field. The IBI coarse-grained potential was obtained using the VOTCA package.[32]

TABLE II. Simulation box length, atomistic region radius (r_{at}), and number of atomistic or atomistic-like particles in the AdResS and fully atomistic systems used to perform free energy calculations. Distances are given in nm, while w_i is defined in Eq. (1). As a function of its position in the HY region, each HY particle has a weight between 0 and 1 (the closer it is to the AT region, the bigger w_i is), whereas w_i is 0 in the CG region and 1 in the AT region. $\langle \sum w_i \rangle$ in the second column is the summation of all the weights averaged over the whole trajectory.

	AdResS		Fully atomistic	
r_{at}	$\langle \sum w_i \rangle$	≈box length	≈box length	No. of molecules
		Methanol		
1.0	608	6.3	2.8	694
1.5	1330	6.3	4.1	2 189
2.0	2486	6.3	6.3	8 212
3.0	5915	8.9	8.9	23 399
		3-methylindole		
1.5	1330	6.8	6.8	10 164
2.0	2486	6.8	...	...
3.0	5915	9.3	...	...

In free energy calculations using TI, the alchemical change was performed in two steps: first switching off Coulombic solute-water interactions (ΔG_{Coul}) and then Lennard-Jones (ΔG_{LJ}). The Coulomb step had a linear dependence of U_{sw} on λ, while for the Lennard-Jones step we used the soft-core potential of Ref. 33 with parameters $\alpha = 0.5$ and $p = 1.0$ to avoid possible singularities from overlapping atoms during the alchemical change.

The temperature was kept constant at 298 K by a Langevin thermostat with a friction constant γ of 15 ps^{-1}. The non-bonded cutoff was 1.2 nm. The integration time step was 2 fs. Electrostatics were treated using the reaction field method with a dielectric constant $\epsilon = 80$ and a cutoff of 1.2 nm; these parameters provide a good compromise between accuracy and speed, as it was verified in Ref. 16. The SETTLE[34] and RATTLE[35] algorithms for rigid water and rigid bonds to hydrogen were used.

Each system was prepared using fully atomistic minimization with the steepest descent method, 500 ps NPT equilibration and 500 ps NVT equilibration. All free energy calculations used 21 λ values per ΔG_{Coul} value and 40 equidistant λ values (with a separation of 0.025) per ΔG_{LJ}, with 1 ns of simulation per λ value, of which the first 100 ps were discarded as equilibration. Free energy calculations were performed in the NVT ensemble throughout, i.e., we approximate the Gibbs by the Helmholtz free energy, after initially verifying that the difference is negligible for systems at these concentrations. This approximation was validated through the comparison of system density with and without the solute molecule. The change in concentration is in fact on the order of 0.01%–0.1%, thus suggesting that the amount of solvent in the system is sufficiently large to absorb the effective volume change due to the decoupling from the solute. Finally, production runs for studying system properties with full solute-solvent interaction were 6 ns long each. All AdResS and most fully atomistic simulations used the ESPResSo++ simulation package,[36] in which we have implemented TI. Some preliminary fully atomistic equilibration simulations used the GROMACS simulation package.[37]

All error bars shown were calculated using the Student t distribution[38] at the 95% confidence limit, via standard deviations obtained using block averaging in which all trajectories were divided into five blocks of equal length.

III. RESULTS

Figure 2 shows the solvation free energy values for methanol and 3-methylindole, comparing fully atomistic systems with different simulation box sizes to AdResS systems with different atomistic region sizes and different CG potentials. The systems are those visualised in Fig. 1(b). The Coulomb [Figs. 2(a) and 2(c)] and Lennard-Jones [Figs. 2(b) and 2(d)] contributions to the free energy are plotted as a function of the number of atomistic or atomistic-like molecules in the system. For fully atomistic systems, this is simply the total number of molecules. For AdResS systems, this is the sum of the w values as defined in Eq. (1), i.e., each fully atomistic molecule contributes 1 to the sum, each fully coarse-grained molecule contributes 0, and water molecules in the hybrid region contribute in accordance with their degree of atomistic character.

In Figs. 2(a) and 2(b), for methanol, the four fully atomistic values (black diamonds) correspond to the four different simulation box sizes. Previous studies have shown that in fully atomistic systems there are no detectable finite-size effects for solvation free energies of neutral solutes as a function of simulation box size,[39] and we make the same observation here. The four values for the AdResS systems using the IBI coarse-grained potential (red circles) correspond to the four different atomistic region sizes (r_{at} = 1.0–3.0 nm), while the value for the AdResS system using a coarse-grained reservoir of non-interacting particles (green squares) has an atomistic region with r_{at} = 1.5 nm. In all cases, the AdResS free energy values agree with the fully atomistic reference to within at least 0.6 kJ mol^{-1}, or $0.2k_BT$ at 298 K, k_B being the Boltzmann constant and T the temperature. This is the case even when the radius of the atomistic region r_{at} is 1.0 nm, somewhat less than the non-bonded interaction cutoff 1.2 nm, and some of the water molecules contributing to $dU_{sw}/d\lambda$ fall within the HY region. The use of such a small atomistic region is possible because the interpolation-based AdResS approach creates a smooth transition from AT to CG region. The water molecules within the HY region close to the AT region have w values close to 1.0 [Eq. (1)] and therefore have considerable atomistic character and atomistic-like properties. Nevertheless, in practice and taking into account the non-bonded cutoff, a prudent choice for the minimum atomistic region would be closer to 1.5 nm or the solute radius of gyration, R_g (0.08 nm for methanol), and a 1.2–1.4 nm thick layer of atomistic water. This is in accordance with the rule of thumb we previously suggested based on the consideration of structural and dynamical properties, which was (R_g + 1.3) nm.[5]

Similarly, for 3-methylindole [Figs. 2(c) and 2(d)], fully atomistic and AdResS free energy values agree to within

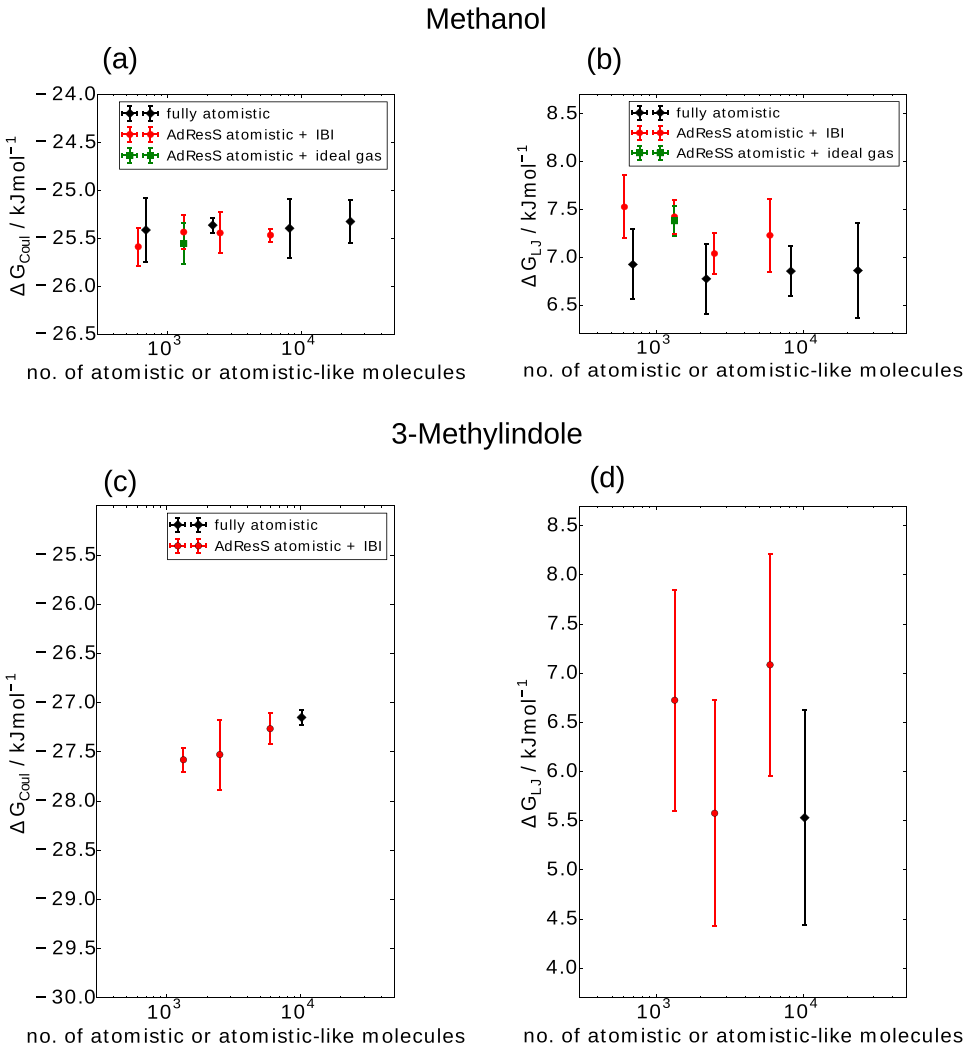

FIG. 2. [(a) and (c)] Coulomb and [(b) and (d)] LJ contributions to the free energies of solvation for [(a) and (b)] methanol and [(c) and (d)] 3-methylindole, fully atomistic versus AdResS with an IBI CG potential and with an ideal gas CG region. (a) and (b) subplots are plotted such that the y-axis covers a range of 2.5 kJ mol^{-1}, or approximately k_BT, instead (c) and (d) are plotted so that the y-axis covers a range of 5.0 kJ mol^{-1}. Note that the x-axes use a logarithmic scale. The quantity plotted on the x-axis is defined in the text. The color legend in (d) is the same as that in (c).

at least 1.5 kJ mol^{-1} or $0.6k_BT$. The three values for the AdResS systems using the IBI coarse-grained potential (red circles) have atomistic region sizes r_{at} = 1.5, 2.0, 3.0 nm. Of course the minimum advisable AT region size is bigger for 3-methylindole (radius of gyration = 0.21 nm) than for the smaller molecule, methanol. Error bars are larger for 3-methylindole than for methanol because the solvation shell of the larger molecule has a more complex configurational space. Moreover, error bars are larger for Lennard-Jones than for Coulomb contributions to the free energy because the linear dependence of the Coulomb energy on λ produces a smoother, more easily integrated curve than the non-linear soft-core potential used for the Lennard-Jones alchemical step.

Finally, Table III summarizes the comparison between experimental solvation free energy values and those calculated in this work. We note that simulated solvation free energy values for these amino acid sidechain analogue systems are known to differ by roughly 1 kcal mol^{-1} or 4 kJ mol^{-1} from experimental values,[16] something we also see here. Simulated free energy values also depend sensitively on the force field chosen and the method used to treat non-bonded interactions.[16] We stress that our main goal here is the comparison of AdResS free energy values to the equivalent fully atomistic reference, for a given force field and set of simulation parameters, and that for this comparison the differences are within the statistical error bars of the simulations and well below k_BT.

The AdResS approach can yield such accurate free energy values relative to the fully atomistic reference because in the atomistic region the AdResS simulations sample configurations from the same ensemble as the fully atomistic

TABLE III. Experimental solvation free energy values in kJ mol^{-1} compared to total solvation free energies ($\Delta G = \Delta G_{Coul} + \Delta G_{LJ}$) calculated in this work. In the second and third rows of the table we report the values for methanol and 3-methylindole obtained in fully atomistic reference simulations and in adaptive resolution simulations with IBI CG potential, respectively. The last row shows the value of ΔG for methanol obtained with ideal gas CG potential. It is useful to point out that the latter model has in fact been employed only in the case of methanol for testing purposes and therefore the value of total free energy solvation for AdResS with ideal gas CG potential for 3-methylindole was not computed.

	Methanol	3-methylindole
Experimental[16,40]	−21.1 to −21.5	−24.6
Fully atomistic, this work	−18.5 to −18.6	−21.6
AdResS + IBI, this work	−18.0 to −18.4	−20.2 to −22.0
AdResS + ideal gas, this work	−18.1	...

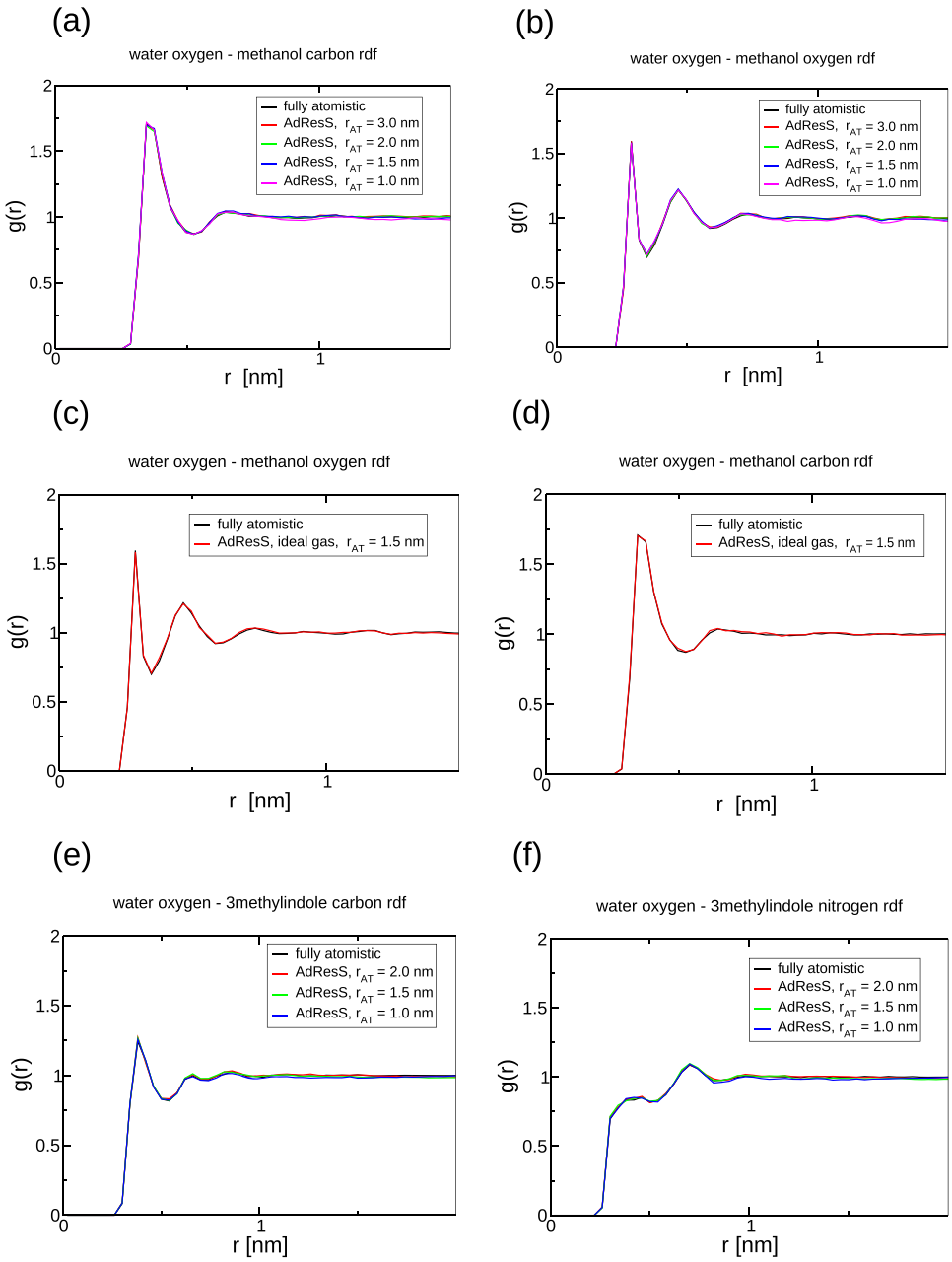

FIG. 3. Radial distribution functions between water oxygen atoms and selected solute heavy atoms, compared to the fully atomistic reference: [(a) and (b)] methanol, fully atomistic versus AdResS with IBI coarse-grained potential, [(c) and (d)] methanol, fully atomistic versus AdResS with ideal gas coarse-grained region, [(e) and (f)] 3-methylindole, fully atomistic versus AdResS with IBI coarse-grained potential: in particular, in (e) we used the carbon atom with sp3 hybridization as the solute heavy atom.

simulations. We now examine some of the structural and thermodynamic properties of the atomistic region in the methanol system.

Figure 3 shows the radial distribution functions (RDFs) of water oxygen atoms around selected solute heavy atoms, comparing various AdResS systems to the fully atomistic reference. In every case, the structure of the solute's solvation shell is perfectly reproduced in the AdResS systems, as has been shown before for a variety of other solutes.[5,7,41]

In Fig. 4 we plot the molecular fluctuations $(\langle N^2\rangle - \langle N\rangle^2)/\langle N\rangle$ along the direction of resolution change, where N is the instantaneous number of particles in a given bin, and all bins have the same surface-to-volume ratio. The molecular fluctuations are proportional to the compressibility. Figures 4(a) and 4(c) show the AdResS systems using the IBI CG potential for methanol and 3-methylindole, respectively, which is parametrised to have the same structure, and therefore the same compressibility, i.e., the same molecular fluctuations, as the atomistic reference. In these AdResS systems, therefore, the molecular fluctuations across the entire system including atomistic and coarse-grained regions correspond to those measured in the fully atomistic system. More striking is the case shown in Fig. 4(b) for the system where the coarse-grained region contains a fluid of non-interacting particles (ideal gas) only for methanol. The molecular fluctuations there are considerably larger than those in the atomistic model and the coarse-grained fluid is completely structureless. Nevertheless, even in this extreme case, the properties of the atomistic region remain unperturbed and hence the atomistic solvation free energy values are still reproduced in this system.

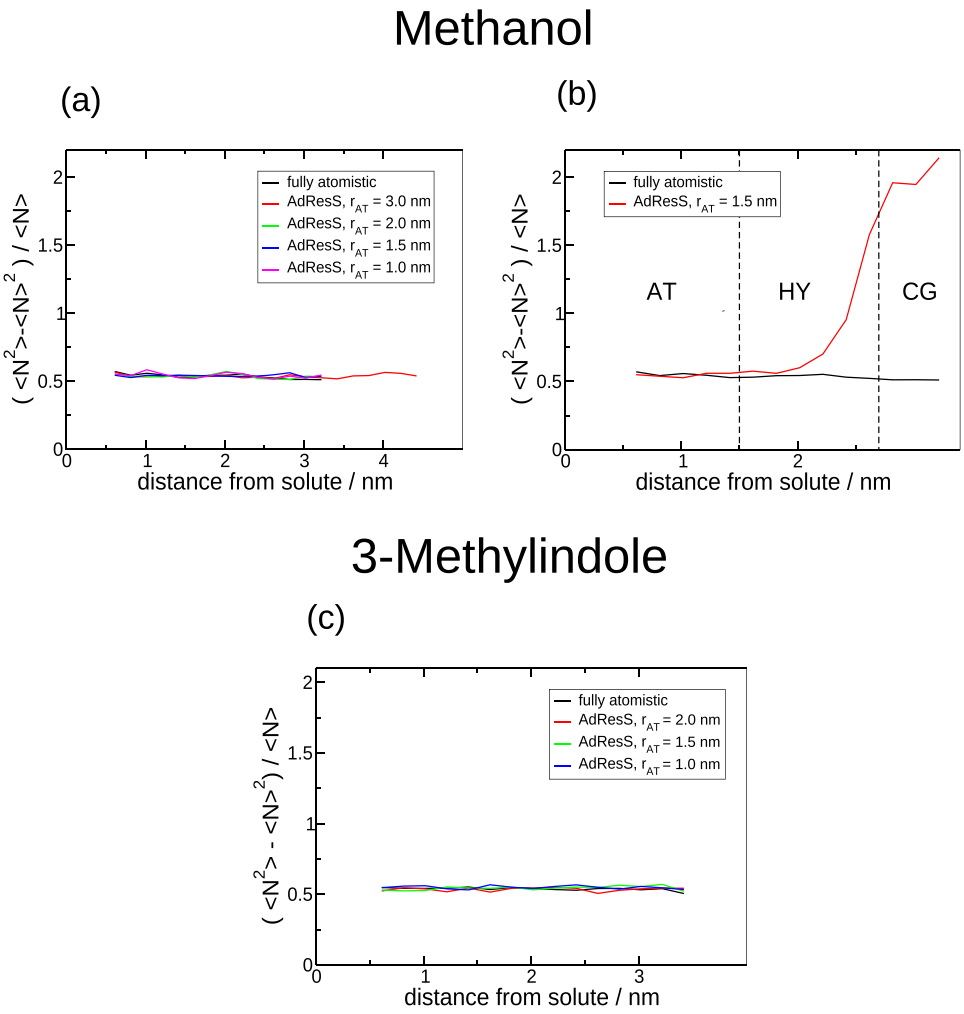

FIG. 4. Molecular fluctuations in spherical concentric bins of equal surface-to-volume ratio, as a function of distance from the solute atom defining the center of the atomistic region, [(a) and (c)] AdResS with IBI coarse-grained potential for methanol and 3-methylindole, respectively, (b) AdResS with ideal gas coarse-grained region for methanol.

IV. CONCLUSIONS

We have shown how the force-based adaptive resolution approach can be used to calculate solvation free energy values, even when using a coarse-grained region or reservoir with extremely different thermodynamic properties. The free energy values obtained in the AdResS setup are accurate to within a fraction of k_BT compared to fully atomistic reference values. These calculations highlight one of the strengths of the AdResS approach, in that it allows accurate control of the atomistic region density and a smooth transition between atomistic and coarse-grained regions, with no perturbation of the structural and thermodynamic properties of the solute and its solvation shell even for atomistic regions whose size is on the order of the correlation length. We also discussed how the energy derivative is defined in the case of a system with no global Hamiltonian.

The speed-up obtained via the AdResS approach compared to fully atomistic simulations depends on the ratio of the coarse-grained and atomistic region volumes and the relative computational cost of atomistic and coarse-grained potentials. In this work, we studied relatively small systems where the atomistic region occupies a large proportion of the total simulation box and where fully atomistic simulations are also feasible. Studying these small systems allowed us to validate the AdResS approach via comparison to fully atomistic reference values. Our long-term goal is the calculation of free energies in large, complex systems where fully atomistic simulations are unfeasible because of system size or indeed because not all system components have been characterised to within atomistic resolution.[10] This includes, for example, ligand binding processes in high-molecular-weight proteins, ligand intercalation in DNA, or small molecule-surface interactions. In such systems, the AdResS approach can be used to simulate at an atomistic level only those solvent molecules in the vicinity of the process of interest, thus significantly reducing the number of atomistic degrees of freedom in the system. The current work forms the basis for such calculations.

ACKNOWLEDGMENTS

K.K., R.F., and A.C.F. acknowledge research funding through the European Research Council under the European Union's Seventh Framework Programme (No. FP7/2007-2013)/ERC Grant Agreement No. 340906-MOLPROCOMP. We thank the John von Neumann Institute for Computing at the Jülich Supercomputing Center for allocating computer time on JURECA. We are grateful to Tristan Bereau and Maziar Heidari for a critical reading of the manuscript.

[1]R. Potestio, C. Peter, and K. Kremer, Entropy **16**, 4199 (2014).
[2]J. H. Peters, R. Klein, and L. Delle Site, Phys. Rev. E **94**, 047701 (2016).
[3]R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, Phys. Rev. Lett. **110**, 108301 (2013).
[4]M. Praprotnik, L. Delle Site, and K. Kremer, J. Chem. Phys. **123**, 224106 (2005).
[5]A. C. Fogarty, R. Potestio, and K. Kremer, J. Chem. Phys. **142**, 195101 (2015).
[6]J. Zavadlav, R. Podgornik, and M. Praprotnik, J. Chem. Theory Comput. **11**, 5035 (2015).
[7]S. Fritsch, C. Junghans, and K. Kremer, J. Chem. Theory Comput. **8**, 398 (2012).
[8]M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, and C. Clementi, J. Phys.: Condens. Matter **19**, 292201 (2007).
[9]S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. Delle Site, J. Chem. Phys. **128**, 024503 (2008).
[10]A. C. Fogarty, R. Potestio, and K. Kremer, Proteins: Struct., Funct., Bioinf. **84**, 1902 (2016).
[11]F. Stanzione and A. Jayaraman, J. Phys. Chem. B **120**, 4160 (2016).
[12]R. Delgado-Buscalioni, Philos. Trans. R. Soc., A **374**, 20160152 (2016).
[13]J. Zhu, R. Klein, and L. Delle Site, Phys. Rev. E **94**, 043321 (2016).
[14]A. Agarwal, H. Wang, C. Schütte, and L. Delle Site, J. Chem. Phys. **141**, 034102 (2014).
[15]J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[16]M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).
[17]A. B. Kuhn, S. M. Gopal, and L. V. Schäfer, J. Chem. Theory Comput. **11**, 4460 (2015).
[18]W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, Acta Polym. **49**, 61 (1998).
[19]D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[20]A. Soper, Chem. Phys. **202**, 295 (1996).
[21]K. Kreis, A. C. Fogarty, K. Kremer, and R. Potestio, Eur. Phys. J.: Spec. Top. **224**, 2289 (2015).
[22]L. Delle Site, Phys. Rev. E **76**, 047701 (2007).
[23]H. Wang, C. Schtte, and L. Delle Site, J. Chem. Theory Comput. **8**, 2878 (2012).
[24]M. E. Johnson, T. Head-Gordon, and A. A. Louis, J. Chem. Phys. **126**, 144509 (2007).
[25]H. Wang, C. Junghans, and K. Kremer, Eur. Phys. J. E **28**, 221 (2009).
[26]A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[27]S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer, Phys. Rev. Lett. **108**, 170602 (2012).
[28]S. Bevc, C. Junghans, K. Kremer, and M. Praprotnik, New J. Phys. **15**, 105007 (2013).
[29]M. Heidari, R. Cortes-Huerto, D. Donadio, and R. Potestio, Eur. Phys. J.: Spec. Top. **225**, 1505 (2016).
[30]W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[31]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[32]V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, J. Chem. Theory Comput. **5**, 3211 (2009).
[33]M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and GROMACS Development Team, GROMACS User Manual version 5.0.4, 2014, www.gromacs.org.
[34]S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).
[35]H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[36]J. D. Halverson, T. Brandes, O. Lenz, A. Arnold, S. Bevc, V. Starchenko, K. Kremer, T. Stuehn, and D. Reith, Comput. Phys. Commun. **184**, 1129 (2013).
[37]B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
[38]Student, Biometrika **6**, 1 (1908).
[39]S. Parameswaran and D. L. Mobley, J. Comput.-Aided Mol. Des. **28**, 825 (2014).
[40]R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate, Biochemistry **20**, 849 (1981).
[41]M. Praprotnik, L. Delle Site, and K. Kremer, J. Chem. Phys. **126**, 134902 (2007).