



Cite this: *Phys. Chem. Chem. Phys.*,  
2020, 22, 25848

## Can the roles of polar and non-polar moieties be reversed in non-polar solvents?<sup>†</sup>

Cedrix J. Dongmo Fomthuiem,<sup>†</sup> Manuel Carrer,<sup>‡</sup> Maurine Houvet,<sup>c</sup>  
Tatjana Škrbić,<sup>de</sup> Giuseppe Graziano<sup>†</sup> and Achille Giacometti<sup>†</sup><sup>\*gh</sup>

Using thermodynamic integration, we study the solvation free energy of 18 amino acid side chain equivalents in solvents with different polarities, ranging from the most polar water to the most non-polar cyclohexane. The amino acid side chain equivalents are obtained from the 20 natural amino acids by replacing the backbone part with a hydrogen atom, and discarding proline and glycine that have special properties. A detailed analysis of the relative solvation free energies suggests how it is possible to achieve a robust and unambiguous hydrophobic scale for the amino acids. By discriminating the relative contributions of the entropic and enthalpic terms, we find strong negative correlations in water and ethanol, associated with the well-known entropy–enthalpy compensation, and a much reduced correlation in cyclohexane. This shows that in general the role of the polar and non-polar moieties cannot be reversed in a non-polar solvent. Our findings are compared with past experimental as well as numerical results, and may shed additional light on the unique role of water as a biological solvent.

Received 1st June 2020,  
Accepted 10th October 2020

DOI: 10.1039/d0cp02948c

rsc.li/pccp

## 1 Introduction

The hydrophobic effect refers to the tendency of non-polar moieties to avoid the contact with water molecules and form an aggregate.<sup>1</sup> This is also commonly viewed as one of the main driving force underlying the folding of a protein.<sup>2,3</sup> As a polypeptide chain is formed by a sequence of amino acids taken from a 20 letter alphabet, 50% of which are roughly hydrophobic (*i.e.* tend to avoid the contact with water) and 50% are polar (*i.e.* are happy to stay in contact with water), in water

the chain tends to fold so as to bury as much as possible the hydrophobic amino acids inside the folded chain. A number of concurring effects<sup>4</sup> prevent a perfect outcome of this scenario, but this would be the optimal configuration in terms of the hydrophobic effects. Hence, water clearly plays an essential role for protein folding and protein functioning.

On the other hand, some experimental studies have pointed out that several enzymes are stable and fully active in anhydrous non-polar solvents.<sup>5</sup> Pace and collaborators<sup>6</sup> noticed that folded proteins become unstable in polar solvents such as ethanol EtOH, but prove to be very stable, albeit essentially insoluble, in non-polar solvents such as cyclohexane C<sub>6</sub>H<sub>12</sub>. The idea is that intra-chain hydrogen bonds are effectively stronger in cyclohexane and similar liquids, because there is no competition to form hydrogen bonds with the non-polar solvent molecules. The denaturing action of ethanol is not as simple to rationalize because polypeptide chains, in aqueous solution with high concentrations of ethanol, populate conformations having a high content of  $\alpha$  helices.<sup>7</sup> A similar scenario emerged from two recent numerical studies<sup>8,9</sup> using an approximate, albeit accurate, method to compute the solvation free energy. It was observed that, although the native state of globular proteins is the most stable in water compared with any other competing folds having the same sequence, this is no longer the case in ethanol and in cyclohexane, where the most stable folds are those having a high content of  $\alpha$  helices, in agreement with experimental studies.

This scenario prompts the following question: is there a liquid, different from water, in which polypeptide chains fold by burying the polar amino acids and still possess the ordered

<sup>a</sup> Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia, Campus Scientifico, Edificio Alfa, via Torino 155, 30170 Venezia Mestre, Italy. E-mail: cedrix.dongmo@unive.it

<sup>b</sup> Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, PO Box 1033 Blindern, 0315 Oslo, Norway. E-mail: manuel.carrer@kjemi.uio.no

<sup>c</sup> Polytech Nantes–Engineering School of the University of Nantes, Rue Christian Pauc, 44306 Nantes Cedex 3, France. E-mail: maurine.houvet@orange.fr

<sup>d</sup> Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia, Campus Scientifico, Edificio Alfa, via Torino 155, 30170 Venezia Mestre, Italy

<sup>e</sup> Department of Physics and Institute for Fundamental Science, University of Oregon, Eugene, OR 97403, USA. E-mail: tskrbic@uoregon.edu

<sup>f</sup> Department of Science and Technology, University of Sannio-Benevento, via Francesco de Sanctis, 82100 Benevento, Italy. E-mail: graziano@unisannio.it

<sup>g</sup> Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia, Campus Scientifico, Edificio Alfa, via Torino 155, 30170 Venezia Mestre, Italy

<sup>h</sup> European Centre for Living Technology (ECLT) Ca, Bottacin, Dorsoduro 3911, Calle Crosera, 30123 Venice, Italy. E-mail: achille.giacometti@unive.it

<sup>†</sup> Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp02948c

<sup>\*</sup> These authors equally contributed to this work.

secondary structure elements? Clearly, non-polar solvents, such as cyclohexane for instance, appear to be optimal candidates for this, thus suggesting the two processes to be mirror images of one another. To the best of our knowledge, this issue has never been addressed before. As we shall see below, the results of the present study indicate that the two processes have very different driving forces. This is also supported by recent results related to the possibility of forming micelles in non-polar solvents by surfactants having a hydrophobic (rather than polar) head, and a polar (rather than hydrophobic) tail, an issue that has been addressed in a parallel work by some of the present authors,<sup>10</sup> and that may be relevant for the existence of unconventional life forms in other planets where water is not available.<sup>11</sup> In that case too, the self-assembly processes of these polarity-inverted surfactants have different driving forces.

In order to tackle this problem in a full fledged study, it would be necessary to characterize the behavior of a complete polypeptide chain in different solvents with molecular details, a task that is beyond our current computational capabilities. On the other hand, small peptides and all isolated natural amino acids are within our present reach.

Motivated by this scenario, in this paper we study the solvation free energies of natural amino acids in water, in ethanol, and in cyclohexane, as well as the free energy differences of moving one amino acid from one solvent into another one. As natural amino acids cannot be isolated by their surrounding environment, we will replace them with their amino acid side chain equivalents that can be obtained by substituting the backbone group with a single hydrogen atom to make the molecule neutral. This can be done for all amino acids but proline and glycine: the former because it does not have a proper side chain, the latter because it does not have a side chain at all.

The problem is not new and experimental data are available – see in particular the important contributions from Wolfendens lab,<sup>12,13</sup> but experimental data for solvation in ethanol are rather scanty. There are also several computer simulation studies considering solvation of amino acid side chain equivalents in water and cyclohexane,<sup>14,15</sup> and comparing results for different water force fields.<sup>16</sup> Another study also addressed the inclusion of the amino acid backbones.<sup>17</sup>

Using thermodynamic integration,<sup>18</sup> we perform an extensive analysis of the solvation free energies of the 18 amino acid side chain equivalents, in water, cyclohexane and ethanol at different temperatures. This also allows the separation of the entropy and enthalpy contributions, thus providing an exhaustive study of the solvation thermodynamics at an unprecedented scale.

In summary, the key new elements provided in our study are (a) a comprehensive treatment in three contrasting solvents; (b) an extensive analysis of the temperature dependence, allowing entropies to be extracted; (c) a detailed discussion of the chemical-physics consequences, with a special focus on the unique role of water as a solvent for biological molecules.

The plan of the paper is as follows. Section 2 provides the general background of our analysis; results are presented in Section 3, and Section 4 will provide some summarizing take-home messages. Additional tables and figures can be found in the ESI.†

## 2 Theory and methods

### 2.1 Thermodynamic integration

The solvation free energy  $\Delta G_{\text{solv}}$  can be defined as the difference between the free energy of a single analyte molecule in a specified solvent  $G_{\text{solvent}}$  and in a vacuum  $G_{\text{vacuum}}$

$$\Delta G_{\text{solv}} = G_{\text{solvent}} - G_{\text{vacuum}} \quad (1)$$

If  $\Delta G_{\text{solv}} < 0$  ( $\Delta G_{\text{solv}} > 0$ ) the solvent is stabilizing (destabilizing) to the molecule with respect to vacuum. This concept can clearly be extended to the free energy transfer  $\Delta\Delta G(S_1 \rightarrow S_2)$  between two different solvents  $S_1$  and  $S_2$

$$\Delta\Delta G(S_1 \rightarrow S_2) = \Delta G_{S_2} - \Delta G_{S_1} \quad (2)$$

where  $\Delta G_{S_1}$  and  $\Delta G_{S_2}$  are the solvation free energy for solvents  $S_1$  and  $S_2$ , respectively.

From the numerical viewpoint, free energy differences can be conveniently computed by using the well-known expression<sup>18</sup>

$$\Delta G_{\text{AB}} = \int_{\lambda_A}^{\lambda_B} d\lambda \left\langle \frac{\partial V(\mathbf{r}; \lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (3)$$

where  $V(\mathbf{r}, \lambda)$  is the potential energy of the system as a function of the coordinate vector  $\mathbf{r}$ , and  $\lambda$  is a switching-on parameter allowing the switch from state A to state B by changing its value from  $\lambda_A$  to  $\lambda_B$ . The average  $\langle \dots \rangle_{\lambda}$  in eqn (3) is the usual thermal average with potential  $V(\mathbf{r}, \lambda)$ . The  $\lambda$  interval  $[\lambda_A, \lambda_B]$  is partitioned into a grid of small intervals, molecular dynamics simulations are performed for each value of  $\lambda$  belonging to each interval, and the results are then integrated over all values of  $\lambda$  to obtain the final free energy difference.

Best practices in free energies calculations<sup>19,20</sup> suggest the use of alchemical transformation in the form of a thermodynamic cycle as defined in Fig. 1(a).<sup>19–21</sup> Firstly, all intramolecular non-bonded interactions in the solute compound are turned off to obtain the dummy compound in a vacuum. Let  $\Delta G_1$  be the free energy difference associated with this transition. Then the dummy compound is transferred from vacuum to the solvent – the liquid in Fig. 1(a). As the free energy of non-interacting molecules does not depend on its environment, the corresponding free energy difference is effectively zero, so  $\Delta G_2 = 0$ . Finally, all the non-bonded interactions are turned on in the solvent with a free energy cost  $\Delta G_3$  to achieve the final compound in the solvent (liquid). Then, clearly  $\Delta G_{\text{solv}} = \Delta G_1 + \Delta G_2 + \Delta G_3$ , as summarized in Fig. 1(a). Note that in the presence of steric interactions only,  $\Delta G_{\text{solv}}$  is purely of entropic nature and can be estimated using Scaled Particle Theory (SPT) (see Section 2). In practice, however, a direct calculation  $\Delta G_{\text{solv}}$  can nowadays be achieved using an efficient application of thermodynamic integration<sup>22</sup> as schematically shown in Fig. 1(b). Here the solute is inserted into a pre-equilibrated solvent, and parallel simulations involve energy minimization, *NVT* and *NPT* equilibration, and production runs are computed for several intermediate values of the coupling parameter  $\lambda$ , and then combined using Bennet's acceptance ratio<sup>23</sup> to finally obtain the required solvation free energy. Here  $\lambda = 0$  refers to fully coupled case, whereas  $\lambda = 1$  to the fully uncoupled case. See Section 2.3 for details.

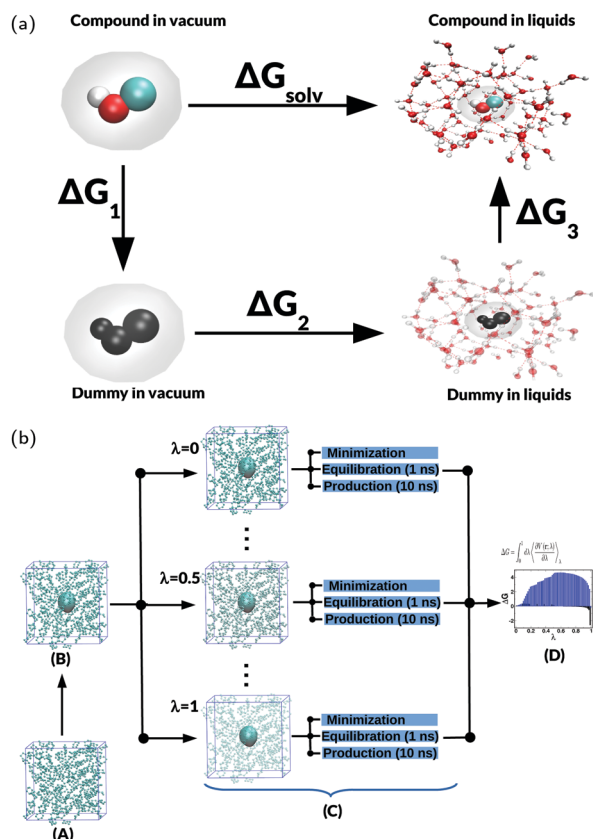


Fig. 1 (a) A thermodynamic cycle allowing the computation of the solvation free energy  $\Delta G_{\text{sol}}$ . (b) Simulation workflow illustrated for the case of cyclohexane. (A) The simulation starts with a pre-equilibrated box of the solvent; (B) the solute is then inserted into the equilibrated box; (C) parallel simulations are performed for each value of the coupling parameter  $\lambda$ . This includes an energy minimization (steepest descent + l-bfgs), an equilibration (NVT + NPT) and a production steps, as shown; (D) the intermediate values of lambda are combined using Bennet's acceptance ratio to obtain the solvation free energy. The value of lambda  $\lambda = 0$  represents the fully coupled (interacting) state while  $\lambda = 1$  is the fully uncoupled (non-interacting) state.

Fig. SI (ESI<sup>†</sup>) depicts 18 of the 20 natural amino acids that are conventionally divided into hydrophobic (non-polar) Fig. SIa, and polar (hydrophilic) Fig. SIb (ESI<sup>†</sup>). Although this division is accepted by general consensus, it relies purely on the chemical structure of the side chain. As we will see below, computational as well as experimental results based on the above rigorous definition will provide further insights into these two classes. Two amino acids have special features and hence have not been included in Fig. SI (ESI<sup>†</sup>): proline because it does not have a proper side chain, glycine because essentially it has no side chains – its side chain is a single hydrogen atom. In natural amino acids, side chains are attached to the backbone, as is also visible in each of the 18 amino acids of Fig. SI (ESI<sup>†</sup>). Molecular equivalents of these 18 natural amino acids side chains can be obtained by capping them with a single hydrogen atom replacing the backbone part. This is presented in Fig. 2 where each equivalent is identified by the short hand notation of its natural side chain counterpart, as reported in Table SI (ESI<sup>†</sup>).

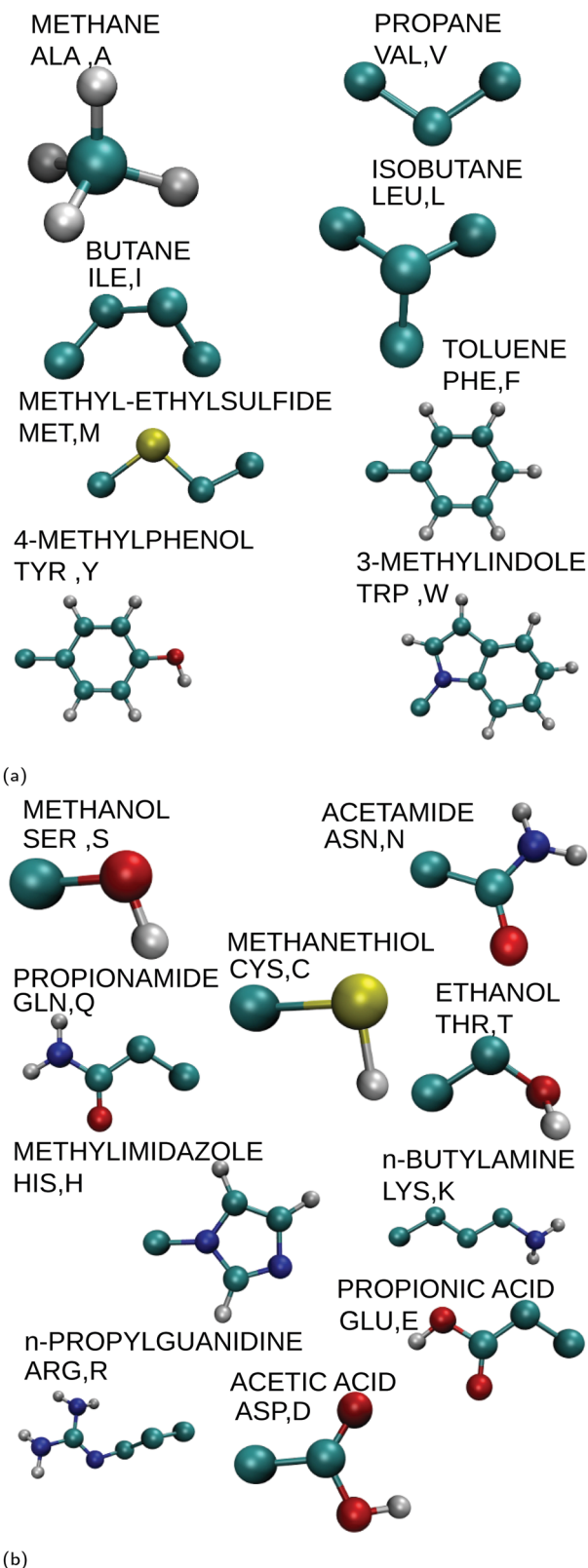


Fig. 2 (a) Hydrophobic amino acid side chain equivalents. (b) Polar amino acid side chain equivalents.

As for the amino acids, solvents too have their own hydrophobicity scale again relying essentially on indirect facts rather

than on a robust thermodynamic relative measurement. One popular way is through the relative dielectric constant  $\epsilon_r$  that is 78.5 in water  $\text{H}_2\text{O}$ , 24.3 in EtOH, and 2.0 in  $\text{C}_6\text{H}_{12}$  at  $T = 298 \text{ K}$ .<sup>24</sup> Accordingly, cyclohexane is much more hydrophobic than water and relatively more hydrophobic than ethanol. The rationale behind this choice of course stems from the fact that the dielectric constant is roughly proportional to the dipole strength that is providing the polarity of the solvent molecules, and dipole–dipole interactions are considerably stronger than any other interaction (quadrupole, van der Waals, *etc.*) appearing in the absence of a permanent dipole.

By computing the solvation free energy  $\Delta G_w \equiv \Delta G_{\text{H}_2\text{O}}$  of each of these amino acid side chain equivalents in water, and then the solvation free energies  $\Delta G_c \equiv \Delta G_{\text{C}_6\text{H}_{12}}$  and  $\Delta G_e \equiv \Delta G_{\text{EtOH}}$  in cyclohexane and ethanol, we can quantify their relative polarity. As hydrophobic molecules produce unfavourable interactions in water and favourable in cyclohexane, we expect  $\Delta G_w > 0$  and  $\Delta G_c < 0$  for hydrophobic amino acid side chain equivalents (ALA, VAL, ILE, LEU, MET, PHE, TYR, TRP), and the opposite  $\Delta G_w < 0$  and  $\Delta G_c > 0$  for polar amino acid side chain equivalents (SER, ASN, GLN, CYS, THR, HIS, LYS, ARG, APS, GLU). In addition, we can also quantify the free energy differences in the transfer water–cyclohexane  $\Delta\Delta G_{w>c} = \Delta G_c - \Delta G_w \equiv \Delta G_{\text{C}_6\text{H}_{12}} - \Delta G_{\text{H}_2\text{O}}$ , and water–ethanol  $\Delta\Delta G_{w>e} = \Delta G_e - \Delta G_w \equiv \Delta G_{\text{EtOH}} - \Delta G_{\text{H}_2\text{O}}$ . This difference provides a measure of the propensity for that particular amino acid side chain equivalent to be solvated in one or the other solvent, and hence a robust scale of relative hydrophobicity with respect to water, as already suggested by Tanford many years ago.<sup>1</sup>

Therefore we will label a particular amino acid side chain equivalent as hydrophobic (with respect to water), if  $\Delta\Delta G_{w>c} < 0$ , polar if  $\Delta\Delta G_{w>c} > 0$ . Likewise, we can have intermediate hydrophobicity values by computing the free energy difference of transferring an amino acid side chain equivalent from water to ethanol  $\Delta\Delta G_{w>e}$ .

A final interesting point is whether the particular solvation process is entropically or enthalpically dominated. This can be understood by separating out the enthalpic and the entropic contributions as obtained from the evaluation of the free energy difference  $\Delta G(T)$  at different temperatures  $T$ , and then the calculation of the entropy *via* differentiation with respect to the temperature. To this aim, we assume the following temperature dependence for the free energy difference<sup>25</sup>

$$\Delta G(T) = a + bT + cT \ln T \quad (4)$$

so that

$$\Delta S(T) = -\left(\frac{\partial \Delta G(T)}{\partial T}\right)_P = -b - c[1 + \ln T] \quad (5)$$

and then the enthalpy change can be obtained from

$$\Delta H(T) = \Delta G(T) + T\Delta S(T) \quad (6)$$

A numerical fit of the parameters  $a$ ,  $b$ , and  $c$  appearing in eqn (4) based on the results of simulations at different temperatures will provide the required expressions for the entropy

(eqn (5)) and for the enthalpy (eqn (6)). Standard deviation was evaluated using error block analysis.<sup>25</sup>

A word of caution is in order here. As discussed in ref. 25, the functional form given in eqn (4) is valid provided that the heat capacity change in solvating the amino acid side chain equivalents is approximately constant in the considered temperature range, 270–330 K in the present study. This is comparable with 278–338 K considered in ref. 25 where only water was investigated. Moreover, the temperature 270 K is below the freezing points of both water and cyclohexane and will be used here as an extrapolated value from the liquid phase. That said, we will use the same functional form even for the two other solvents considered in the present study, knowing that this is an unwarranted approximation possibly invalid in some cases. Furthermore, 270 K does not appear to be a particular outlier to the fitted curves, see Fig. SVI–SVIII (ESI†).

## 2.2 Scaled particle theory (SPT)

According to a general theory of solvation,<sup>26</sup>  $\Delta G_{\text{soln}}$  can be calculated as the sum of the reversible work spent to create a cavity suitable to host the solute molecule,  $\Delta G_0$ , and of the reversible work to turn on the attractive solute–solvent interactions,  $E_a$ , usually assumed to be a purely energetic term. Reliable estimates of  $\Delta G_0$  in any liquid can be calculated by means of the analytical relationships provided by classic Scaled Particle Theory (SPT).<sup>27</sup> It is only necessary to assign an effective hard sphere diameter to solvent and solute molecules and to use the experimental solvent density at the temperature and pressure of interest. The use of experimental density is an indirect way to take into account the true interactions existing among solvent molecules in the pure liquid. Reliable estimates of  $E_a$  can be calculated by means of simple expressions in the case of pure van der Waals attractions, whereas numerical calculations are in general necessary in the case of hydrogen bonds. Additional details of the theoretical aspects can be found in the original paper.<sup>26</sup>

## 2.3 Numerical protocols

The amino acid side chain equivalents used in this work are organic chemical moieties derived by truncating the natural amino acid side chains at the CB position and capping the tail with a hydrogen atom. In particular, the initial structures for these latter compounds along with their building topology were retrieved from the Automated Topology Builder (ATB2.0).<sup>28</sup> While a united atom representation was used in setting up the systems, an in house modification of GROMOS96 (54a7) force field<sup>29</sup> was required to account for non-natively parameterized molecules. The choice of this force field is in line with past work<sup>14</sup> where it was shown to provide good description of the solvent properties. Here we have used the latest 54a7 version of the Gromos force field, while Villa *et al.*<sup>14</sup> employed the 43a2 version. A good alternative for cyclohexane would have been the most recent version of the optimized OPLS (L-OPLS)<sup>30</sup> which yields improved values of hydrocarbon diffusion coefficients, viscosities, and gauche–trans ratios. Selection of the latter would have more faithfully compared with results of earlier simulations by Chang *et al.*<sup>15</sup> who used an older OPLS-AA force



field. In view of the highly computational requirements of the present holistic analysis, we have made the reasonable compromise of selecting GROMOS96 (54a7) which was explicitly tuned to best reproduce the experimental hydration enthalpies of the side-chain analogs as well as to better preserve the protein secondary structure. Other choices<sup>16,17</sup> provide comparable performances.

The simulations were performed in three different solvents covering a broad range of polarities, from non-polar  $\text{C}_6\text{H}_{12}$ , to highly polar  $\text{H}_2\text{O}$ , through the intermediate polar EtOH. The 18 amino acid side chain equivalents were then inserted into a cubic box of 3 nm size incorporating about 165, 290 and 881 molecules of  $\text{C}_6\text{H}_{12}$ , EtOH, and  $\text{H}_2\text{O}$ , respectively. The simulations were performed using the Gromacs simulation package (versions 2018.3 and 2018.7)<sup>31</sup> and all the solutes were modeled in their neutral uncharged states. As detailed in Section 2, free energy differences as given by eqn (3) have been computed from the fully coupled ( $\lambda = 0$ ) to the fully uncoupled  $\lambda = 1$  system, by gradually switching off all non-steric interactions. A grid of  $\Delta\lambda = 0.05$  has been used in all cases, resulting into 21 binning points. See Fig. SIII (ESI<sup>†</sup>). The initial systems were initially prepared by minimizing the solute's potential energy and relaxing the solvent around the solute atoms before running free energy molecular dynamics (MD) simulations. Two cycles of minimization rounds were performed embedding  $10^5$  steps of steepest descent minimization algorithm with a minimization step size of  $5 \times 10^{-4}$  nm and a maximum convergence force of  $100.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  followed by  $5 \times 10^4$  steps of l-bfgs quasi-Newtonian minimization algorithm with a minimization step size of  $10^{-3}$  nm and a maximum convergence force of  $100.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . Thereafter, an equilibration round in the canonical *NVT* ensemble was performed for 500 ps using the accurate leap-frog stochastic dynamics integrator with a simulation time step of 2 fs. Due to the large numerical fluctuations recorded, a shorter time step of 0.5 fs or 1 fs was used in some simulations. While long-range electrostatics interactions were accounted with the Particle Mesh Ewald summation,<sup>32</sup> short-range electrostatics and van der Waals interactions were truncated with a single-range cutoff at 12 Å with the pair list updated every 20 steps. The simulations were performed at seven different temperatures in the range 270–330 K. Each temperature was kept around the reference value by coupling the system to an external bath using the Berendsen thermostat (for less stable systems),<sup>33</sup> with a coupling constant of 1.0 ps. All simulations were replicated in a 3D bulk-like phase using the periodic boundary conditions and all bonds involving hydrogen atoms were restrained using LINCS algorithms.<sup>34</sup> For water we used the simple point charge SPC water model,<sup>35</sup> whereas the parameters (or topology) for ethanol and cyclohexane were manually implemented. Small quantitative differences could be expected<sup>17</sup> by choosing more refined force fields for water, at the expense of a significant increase in the computational time. The second equilibration round was then performed for additional 500 ps in the isobaric–isothermal *NPT* ensemble using the same parameters as in *NVT*. The pressure was equilibrated to the reference value of 1 bar using the Parrinello–Rahman pressure coupling (for more stable systems)<sup>36</sup> and the weak-coupling Berendsen

barostat (for less stable systems),<sup>33</sup> with a coupling constant of 1.0 ps. The isothermal compressibility (in  $\text{bar}^{-1}$ ) of  $4.5 \times 10^{-5}$ ,  $1.2 \times 10^{-4}$ , and  $4.5 \times 10^{-5}$  was used for water, ethanol and cyclohexane, respectively. The final production runs for free energy calculations were performed for 10 ns using 21 equidistant lambda points with a step size of 0.05. In the case of cyclohexane only the dispersive interactions were decoupled, while for ethanol and water the coulomb interactions were also considered. In short, we have closely followed the numerical protocol by Villa *et al.*,<sup>14</sup> but we have improved it using an updated forcefield and extended the simulation timescales by performing longer free energy samplings. Moreover, in many cases we have employed a smaller time-step where Villa *et al.*<sup>14</sup> used 2 fs. Our simulation workflow is shown in Fig. 1b.

## 3 Results

### 3.1 Solvation free energy $\Delta G_{\text{solv}}$

Fig. 3 displays the solvation free energy for  $\text{H}_2\text{O}$  (Fig. 3a),  $\text{C}_6\text{H}_{12}$  (Fig. 3b), and EtOH (Fig. 3c), and compares the results of the present work with both experimental and computational previous studies in the literature. All corresponding values can be found in Tables SII, SIII and SIV (ESI<sup>†</sup>). Broadly speaking, the solvation free energies follow the general division in hydrophobic and polar amino acids illustrated in Fig. 2 – note that the sequence of the amino acids of Fig. 3 from left to right follows the same scheme hydrophobic → polar division of Fig. 2, but there are exceptions. In water (Fig. 3a), the  $\Delta G_{\text{w}}$  values of polar amino acid side chain equivalents are negative, whereas they are positive for the hydrophobic methane (ALA), propane (VAL), butane (ILE) and isobutane (LEU), as largely expected. However, for MET, PHE, TYR and TRP we find  $\Delta G_{\text{w}} < 0$ . While seemingly odd at first sight, we note that this is in line with experimental data. For instance, for toluene  $\Delta G_{\text{w}} = -3.7 \text{ kJ mol}^{-1}$  at 25 °C and 1 atm, (see Table SV, ESI<sup>†</sup>). This means that aromatic side chains cannot be classified as purely hydrophobic, because they have favorable interactions with water molecules. This is a very interesting point. Indeed, it is known that benzene forms weak hydrogen bonds with two water molecules located over the two faces of the planar aromatic ring. In general, the partial positive charge of the hydrogen atom attached to very electronegative atoms (*i.e.* O or N) interact favorably with the delocalized  $\pi$  electrons of the aromatic ring.<sup>37</sup> This is a specific example of a more general class of  $\text{A-H} \cdots \phi$  hydrogen bonds, where  $\phi$  represents an aromatic ring and A may be a N, O, or C atom.<sup>38</sup> In particular, these weak hydrogen bonds can form among water molecules and the aromatic side chains of PHE, TYR and TRP.<sup>38</sup> While the present forcefield was not devised to address this problem, it still proves instructive to test for this prediction in the case of the TRP amino acid side chain equivalent. This is reported in Fig. SV (ESI<sup>†</sup>) where for both water and ethanol we report the number of hydrogen bonds as a function of the simulation time at various stages of the decoupling process (*i.e.* different values of  $\lambda$ ). In the case of hydrophobic amino acid side chain equivalents our findings are, in general, in good

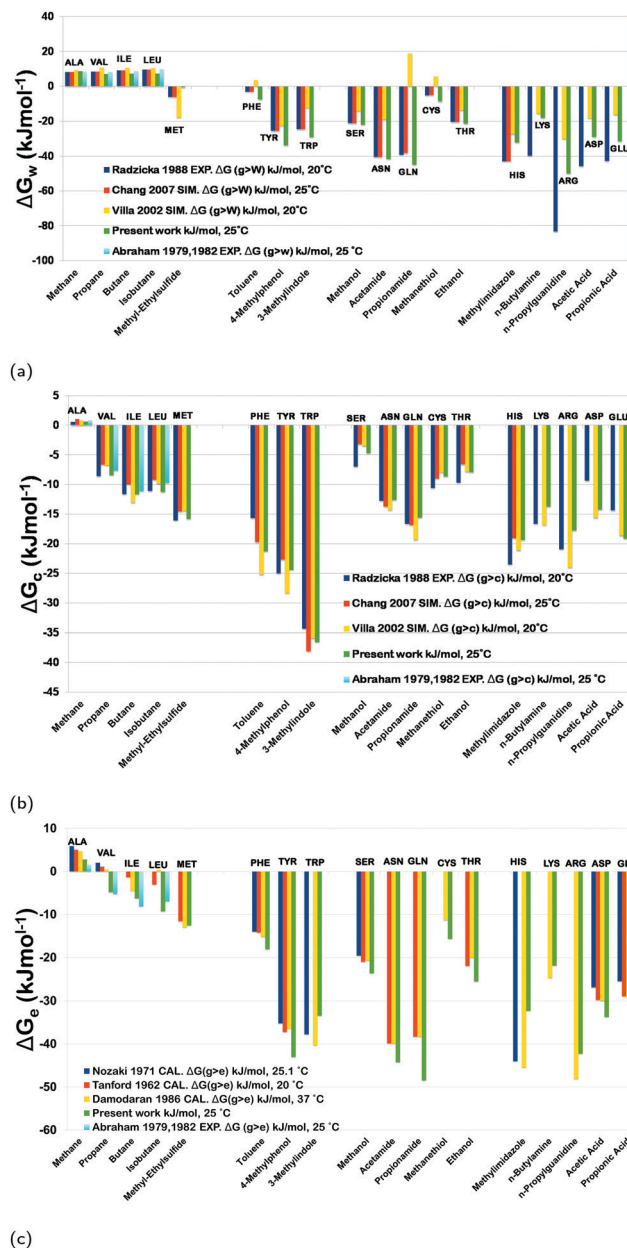


Fig. 3 (a)  $\Delta G_w$  from vacuum to water  $H_2O$  at 25 °C; (b)  $\Delta G_c$  from vacuum to  $cC_6H_{12}$ ; (c)  $\Delta G_e$  from vacuum to EtOH. The results of the present work are also contrasted with the past computational work of ref. 14 and 15, as well as with the experimental work of ref. 12. Each amino acid side chain equivalent is referred to with the name (*x*-axis) and with the shorthand notations of the corresponding amino acid at the top of each value.

agreement with both experimental results,<sup>12</sup> as well as past computational results,<sup>14,15</sup> with the exception of methyl-ethylsulfide (MET) and toluene (PHE). This could be ascribed to a generally greater difficulty in simulating the aromatic rings compared to the other acyclic compounds. By contrast, well grounded estimates were obtained for 3-methylindole (TRP) in agreement with the earlier ones, thus supporting the reliability of the adopted force field.

In parallel with the case of water, our estimates of the solvation free energy in  $cC_6H_{12}$  (Fig. 3b) confirm the results that

all amino acid side chain equivalents have favourable solvation free energy ( $\Delta G_c < 0$ ), with quantitative agreement with experimental data<sup>12</sup> and previous computational investigations.<sup>14,15</sup> As  $cC_6H_{12}$  is a nonpolar liquid, unable to form hydrogen bonds, the negative  $\Delta G_c$  values are due to the action of van der Waals attractions among the solute molecule and surrounding solvent molecules whose magnitude overcomes the free energy cost for creating the cavity. This is likely to be ascribed to the  $cC_6H_{12}$  large molecular polarizability, a fundamental player of London dispersion interactions. This is confirmed by the finding that the largest  $|\Delta G_c|$  value is found for 3-methylindole (TRP), that is the largest solute molecule in terms of surface area among those considered in this study. In general, the values  $\Delta G_c$  obtained in the present work are closer to experimental data than the previously calculated values (see Table SIII, ESI†).

The solvation free energy  $\Delta G_e$  in EtOH (see Fig. 3c) is found negative for all amino acid side chain equivalents but methane (ALA), paralleling the situation in  $cC_6H_{12}$  (compare Fig. 3b and c), and in line with experimental data (see Table SIV, ESI†). We are not aware of any previous computational study providing estimates of  $\Delta G_e$  for all the amino acid side chain equivalents considered here, so only limited comparisons can be carried out.<sup>39–41</sup> Note that here the temperatures are also different. As a further remark, we stress that experimental  $\Delta G_e$  values by Nozaki and Tanford<sup>39</sup> were obtained by subtracting the  $\Delta G_e$  value of GLY from those of the amino acids (*i.e.* including backbones) under an unwarranted additivity assumption. Indeed this assumption usually breaks down for very polar solutes, such as amino acids. Interestingly, even though EtOH is a polar solvent able to form hydrogen bonds, its behavior resembles that of  $cC_6H_{12}$ . This means that here too the attractive solute–solvent energetic interactions (accounting also for hydrogen bonds) are able to overcome the free energy cost of cavity creation. Not surprisingly, in fact, the largest  $|\Delta G_e|$  values are associated with solutes, such as acetamide (ASN) and propionamide (GLN), which can be engaged in multiple hydrogen bonds with EtOH molecules.

In this respect, it proves instructive to compare the present findings with results that can be obtained from SPT and related theories, as discussed in Section 2.2.<sup>26,42</sup> Here the solvation free energy can be estimated in all liquids as the sum of two contributions: (a) the reversible work  $\Delta G_0$  to create a cavity in the liquid, suitable to host the solute molecule. This contribution is always positive so  $\Delta G_0 > 0$  always; and (b) the reversible work  $E_a$  to turn on solute–solvent energetic attractions, both van der Waals interactions and hydrogen bonds. This second contribution can be considered to be purely energetic so that  $E_a < 0$  always. In other words,  $E_a$  favors solvation, whereas  $\Delta G_0$  opposes it. Reliable estimates for  $\Delta G_0$  of simple geometric shapes are calculated by means of classic SPT. By assigning an effective hard sphere diameter to solvent molecules, and using the experimental density of the three liquids at 298 K and 1 atm, we find that  $\Delta G_{0w} > \Delta G_{0e} > \Delta G_{0c}$  for  $H_2O$ , EtOH, and  $cC_6H_{12}$  in decreasing order, see Fig. SII (ESI†). The ranking order can be easily rationalized by the fact that water molecules are the smallest (a cyclohexane molecule has a van der Waals volume roughly 5 times larger than that of a water molecule)

and so liquid water is characterized by the largest number density, which in turn increases the entropy loss associated with cavity creation, due to the solvent-excluded volume effect. On the other hand, the energetic  $E_a$  term consists of a van der Waals contribution, essentially of the same magnitude in the three liquids, and a hydrogen bond contribution, whose magnitude is large in  $H_2O$ , slightly less in  $EtOH$ , and zero in  $cC_6H_{12}$ . Using the method sketched in Section 2.2, we estimated the solvation free energies for methane (ALA), propane (VAL), toluene (PHE) and methanol (SER) as  $\Delta G_{0e} + E_a$  and find them to be in agreement with experimental values in the three considered liquids, as shown in Table SV (ESI†). Hence (a) in  $H_2O$ ,  $\Delta G_{0w} > |E_a|$  for aliphatic hydrocarbons, whereas the opposite holds true for aromatic hydrocarbons and polar molecules able to form hydrogen bonds with water molecules; (b) in  $cC_6H_{12}$  and  $EtOH$ ,  $|E_a| > \Delta G_{0e,c}$  for all the amino acid side chain equivalents but methane (ALA), because the free energy cost of cavity creation is not so large.

### 3.2 Transfer free energies between solvents

Additional insights can be achieved by computing the change in the solvation free energy between different solvents. We shall refer to them as the transfer free energies in the following. This is shown in Fig. 4. Fig. 4a depicts our results for the free energy transfer  $\Delta\Delta G_{w>c} \equiv \Delta\Delta G(H_2O \rightarrow cC_6H_{12})$  from  $H_2O$  to  $cC_6H_{12}$  and contrasts them with the past simulations<sup>14,15</sup> and experiments.<sup>43</sup> Rather evidently, here all hydrophobic amino acid side chain equivalents (except TYR) have  $\Delta\Delta G_{w>c} < 0$  indicating their increased propensities in being solvated by a non-polar solvent such as  $cC_6H_{12}$  rather than a polar solvent such as  $H_2O$ . Likewise, we find that  $\Delta\Delta G_{w>c} > 0$  for all polar amino acid side chain equivalents indicating their decreased propensities in being solvated by  $cC_6H_{12}$  rather than  $H_2O$ .

The present values are quantitatively close to experimental data<sup>43</sup> and perform better than previously calculated estimates.<sup>14,15</sup> For instance,  $\Delta\Delta G_{w>c}$  is positive for propionamide (GLN) and in line with the experimental value, whereas a previous numerical estimate was negative (see Fig. 4a). Remarkably,  $\Delta\Delta G_{w>c}$  almost perfectly divides polar amino acid side chain equivalents from the hydrophobic ones, thus prompting the possibility of being used as a correct measure of hydrophobicity for amino acid side chains, as claimed a long time ago by Wolfenden.<sup>44</sup> This finding is also an indication that  $cC_6H_{12}$ , like other non-polar organic liquids, does not act as a denaturant of the folded state of globular proteins, in agreement with experimental evidence.<sup>6</sup>

The values of the transfer free energy  $\Delta\Delta G_{w>e} \equiv \Delta\Delta G(H_2O \rightarrow EtOH)$  from  $H_2O$  to  $EtOH$  are shown in Fig. 4b. They are negative for all amino acid side chain equivalents, regardless of their polarity. This is in line with available experimental data,<sup>40</sup> even though the latter are very slightly positive for methanol (SER), acetamide (ASN) propionamide (GLN), and ethanol (THR). In particular, the calculated  $\Delta\Delta G_{w>e}$  values for methane (ALA), propane (VAL), butane (ILE) and isobutane (LEU) are negative and fully consistent with experimental data (see Table SV, ESI†).

Once more, it proves instructive to contrast the above findings with the theoretical results stemming from the SPT analysis of

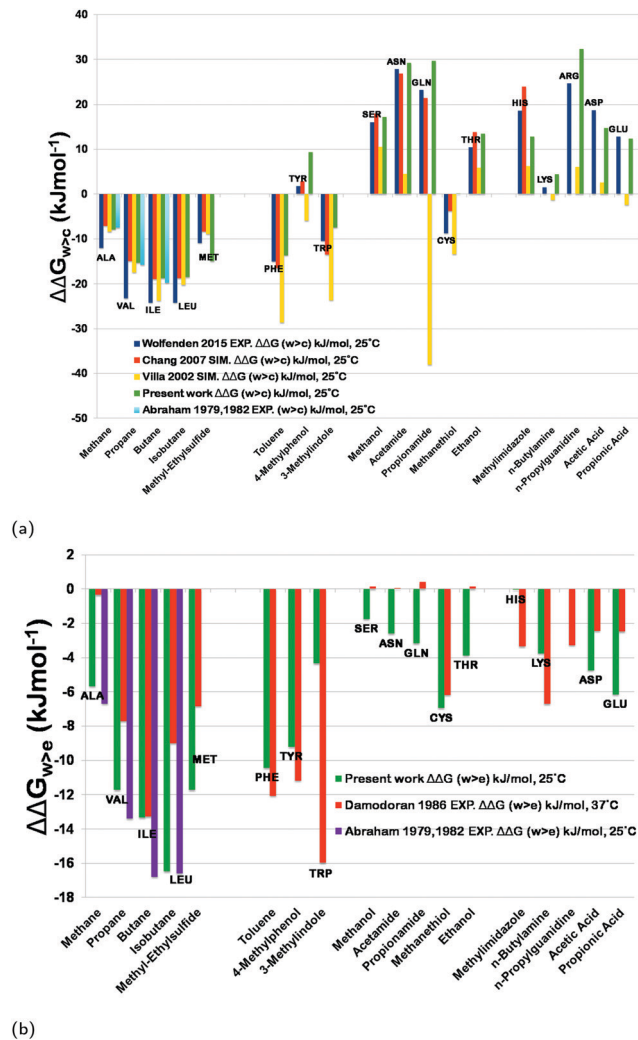


Fig. 4 (a)  $\Delta\Delta G_{w>c}$  from  $H_2O$  to  $cC_6H_{12}$ ; (b)  $\Delta\Delta G_{w>e}$  from  $H_2O$  to  $EtOH$ . The results of the present work are also contrasted with the past computational work of ref. 14 and 15, as well as with the experimental work of ref. 43 ( $H_2O$  to  $cC_6H_{12}$ ) and<sup>40</sup> ( $H_2O$  to  $EtOH$ ). Each amino acid side chain equivalent is referred to with the name ( $x$ -axis) and with the shorthand notations of the corresponding amino acid at the top of each value.

Section 2.2. Here, we can build on the fact that the reversible work of cavity creation  $\Delta G_{0w}$  in  $H_2O$  is larger than its counterpart  $\Delta G_{0e}$  in  $EtOH$ , that is  $\Delta G_{0w} > \Delta G_{0e}$  (see Table SV, ESI†). By contrast, the reversible work of turning on solute–solvent attractions in water  $E_{aw}$  in  $H_2O$  is approximately equal to its counterpart  $E_{ae}$  in ethanol ( $E_{aw} \approx E_{ae}$ ) so that  $\Delta G_{0w} + E_{aw} > \Delta G_{0e} + E_{ae}$ , thus predicting  $\Delta\Delta G_{w>e} < 0$  in agreement with the above numerical results. The fact that the  $\Delta\Delta G_{w>e}$  is negative for almost all side chains indicates that (a) it cannot be a correct measure of hydrophobicity; (b) ethanol has a denaturing action towards the folded state of globular proteins as confirmed by experimental studies.<sup>39</sup>

### 3.3 Entropy–enthalpy compensation

While solvation free energy is certainly the most insightful quantity for understanding a molecule's interaction with a solvent, a deeper understanding can be achieved by singling

out its entropy and enthalpy components. In this case, experiments struggle to provide a detailed description and theoretical and numerical simulations prove to be very effective. In this framework, a useful approach is provided by the grid cell theory<sup>45</sup> that is a refined version of partition function methods.<sup>46</sup> It is then of considerable interest to ask how the present study can contribute to this issue.

As anticipated in Section 2.1 we follow the work of Schauerl *et al.*,<sup>17</sup> to separate the free energy of solvation into its enthalpic and entropic parts. To this aim, we computed the solvation free energy at seven different temperatures in the range 270–330 K and then used eqn (4) to fit the parameters  $a$ ,  $b$ , and  $c$  and hence obtain  $\Delta S(T)$  from eqn (5). All details of these calculations can be found in Table SXI, Fig. SVI, SVII and SVIII (ESI†). Fig. 5 then reports the entropic term  $-T\Delta S$  of the solvation free energy  $\Delta G_{\text{solv}}$  as a function of the enthalpic term  $\Delta H$  for different solvents: H<sub>2</sub>O (Fig. 5a), cC<sub>6</sub>H<sub>12</sub> (Fig. 5b) and EtOH (Fig. 5c).

A visual inspection of the plots in Fig. SVI, SVII, and SVIII (ESI†) indicates that (a) for H<sub>2</sub>O  $\Delta G_w(T)$  is an increasing function of temperature for all amino acid side chain equivalents and so the calculated hydration entropy change is always negative, regardless of the solute polarity, in line with experimental data<sup>47</sup> (this is a further support of the reliability of the adopted force fields and the calculation procedure); (b) for cC<sub>6</sub>H<sub>12</sub>  $\Delta G_c(T)$  does not have a temperature dependence common to all amino acid side chain equivalents; (c) for EtOH  $\Delta G_e(T)$  is an increasing function of temperature for almost all amino acid side chain equivalents, closely resembling the situation for water. A quantitative analysis performed using eqn (4) and (5) leads to the calculated solvation enthalpy and entropy values at 298.15K listed in Tables SVI, SVII and SVIII (ESI†). These values have been used to build up plots of  $-T\Delta S$  versus  $\Delta H$  for all amino acid side chain equivalents in the three liquids, as displayed in Fig. 5.

The general expectation is that in water a large and negative  $\Delta H$  term (*i.e.*, strong solute–solvent energetic attractions) is associated with a large and positive  $-T\Delta S$  term (*i.e.*, a decrease of entropy). In other words, an enthalpy gain leads to an entropy loss, and a correlation with a negative slope emerges. This feature is commonly denoted as ‘entropy–enthalpy compensation’. This is indeed confirmed by our results reported in Fig. 5a. While qualitatively similar in the three solvents, the entropy–enthalpy compensation is quantitatively much more relevant in H<sub>2</sub>O as shown in Fig. 5a. For instance, for toluene (PHE) (a) in H<sub>2</sub>O  $\Delta H = -53$  kJ mol<sup>-1</sup> and  $-T\Delta S = 46$  kJ mol<sup>-1</sup>; (b) in cC<sub>6</sub>H<sub>12</sub>  $\Delta H = -46$  kJ mol<sup>-1</sup> and  $-T\Delta S = 24$  kJ mol<sup>-1</sup>; (c) in EtOH  $\Delta H = -29$  kJ mol<sup>-1</sup> and  $-T\Delta S = 10$  kJ mol<sup>-1</sup>. The large difference between the three liquids is mainly due to the structural reorganization of solvent molecules upon solute insertion, that should provide positive contributions to both the solvation enthalpy and entropy changes. This structural reorganization can be correlated with the isobaric thermal expansion coefficient  $\alpha_p$  of the liquid. Indeed, at room temperature 298 K and 1 atm,  $\alpha_p$  is very small in water, but large in organic liquids ( $\alpha_p = 0.257 \times 10^{-3}$  for H<sub>2</sub>O,  $1.214 \times 10^{-3}$  for cC<sub>6</sub>H<sub>12</sub>, and  $1.089 \times 10^{-3}$  for EtOH in K<sup>-1</sup>),<sup>24</sup> giving a simple explanation for the different magnitudes of such structural

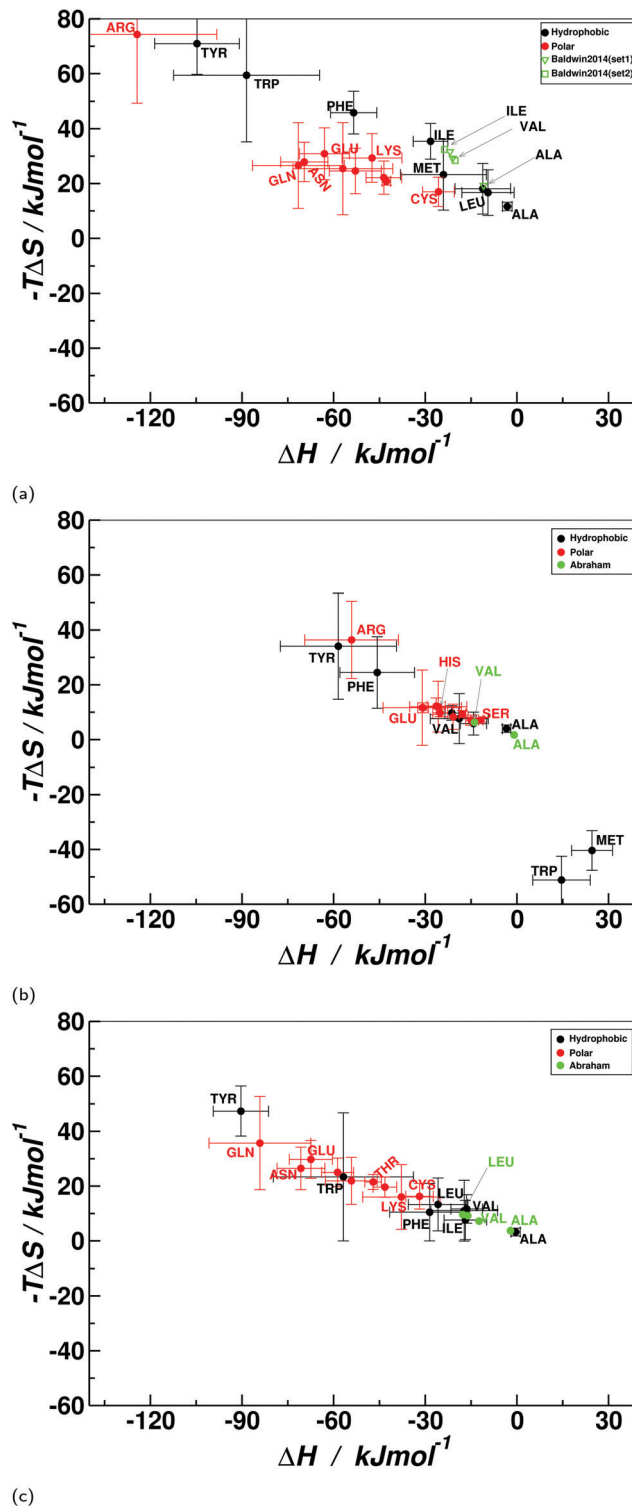


Fig. 5 Entropic part  $-T\Delta S$  of the solvation free energy  $\Delta G$  as a function of the enthalpic part  $\Delta H$  in the case of (a) H<sub>2</sub>O; (b) cC<sub>6</sub>H<sub>12</sub>; (c) EtOH. In the case of H<sub>2</sub>O, the corresponding experimental results<sup>48</sup> are also included. Note that all plots are in the same scale.

reorganization in the three solvents. The net distinction between polar and hydrophobic amino acid side chain equivalents occurring in EtOH is noteworthy and is in striking contrast



to the lack of such a separation in  $cC_6H_{12}$  (compare Fig. 5b and c). Unfortunately, a detailed comparison with experimental or computational results is not possible due to the lack of such data for most of the amino acid side chain equivalents considered in the present study.

As expected, in  $H_2O$  (Fig. 5a) polar amino acid side chain equivalents have large and negative  $\Delta H$ , mainly induced by the possibility of forming hydrogen bonds with water, compensated by an equally large and positive  $-T\Delta S$  stemming from the solvent entropy reduction in the water cage<sup>49</sup> around a polar solute. Note that this trend is particularly emphasized in the case of *n*-propylguanidine (ARG) and much more reduced in methanethiol (CYS) for which this enthalpy–entropy compensation mechanism is very limited. By contrast, hydrophobic amino acid side chain equivalents cannot form hydrogen bonds with water and hence do not trigger the re-orientation of the first solvation shell of water around them. As a consequence, they are almost unaffected in terms of enthalpy change and show a small increase in  $-T\Delta S$  originating from the entropy loss of forming the cavity for accommodating the solute (that is always present). The obtained results in the case of methane (ALA), propane (VAL), and butane (ILE) are in reasonable agreement with calorimetric measurements.<sup>48</sup> Here too, 4-methylphenol (TRP) and toluene (PHE) appear to be outliers with large and negative  $\Delta H$  and large and positive  $-T\Delta S$ . We remark that the results presented here are different from those appearing in the analogue plot of Schauerl *et al.*<sup>17</sup> (see their Fig. 5) because their amino acids include the backbone that in our case is represented by a single hydrogen. Their analyses further show the sensitivity of the obtained results with respect to a change of the water model, so care must be exercised in using them for quantitative comparisons. The slope of the negative correlation between enthalpy and entropy change in water appear to be  $\approx 0.5$  consistent with their results.

Fig. 5b provides the same plot in  $cC_6H_{12}$ . Here it is rather evident that we do not obtain the same results by reversing the role of polar and hydrophobic amino acid side chain equivalents. While an overall negative slope is visible, most of the amino acid side chain equivalents tend to cluster in a region of small enthalpic gain  $\Delta H \approx -30 \text{ kJ mol}^{-1}$  with negligible entropic loss  $-T\Delta S \approx 0$ . Intriguingly, this appears to be independent of the polar character of the amino acid side chain equivalents, as both polar and hydrophobic molecules belong to this cluster. There are however outliers in both senses. Hydrophobic 4-methylphenol (TYR) and polar *n*-propylguanidine (ARG) show a much more marked enthalpy gain with  $\Delta H \approx -60 \text{ kJ mol}^{-1}$  compensated by a significant entropy loss  $-T\Delta S \approx 30 \text{ kJ mol}^{-1}$ . On the opposite side, hydrophobic 3-methylindole (TRP) and methyl-ethylsulfide (MET) present a significant entropic gain  $-T\Delta S \approx -40 \text{ kJ mol}^{-1}$  compensated by a corresponding enthalpy loss  $\Delta H \approx 20 \text{ kJ mol}^{-1}$ .

Somewhat surprisingly, but in line with the discussion presented in the previous sections, the results in the case of EtOH appear to be the cleanest ones, as reported in Fig. 5c. Here a nearly perfect negative correlation is found, with all polar amino acid side chain equivalents gaining in enthalpy and losing in entropy upon being solvated. This can be rationalized by recalling that in all cases a negative enthalpy of

solvation, which provides a favourable contribution to the free energy is compensated by a positive solvation entropy. This is due to the fact that for the insertion of a solute inside a solvent a cavity must be created and the solvent molecules have to rearrange themselves around it, independently of their polarity. Polar solutes however can form hydrogen bonds in  $H_2O$  and EtOH so the gain in enthalpy is sufficient to compensate this entropy loss. On the other hand, the reverse is not true in  $cC_6H_{12}$  where there is no general tendency of the hydrophobic amino acid side chain equivalent to display a favourable solvation in  $cC_6H_{12}$  compared to the polar solvents.

Finally, even in this case it is useful to perform the same calculations by transferring the molecules from water to either cyclohexane or ethanol. This is reported in Fig. SIV (a) for the case of water to cyclohexane and in Fig. SIV (b) (ESI<sup>†</sup>) for the case of water to ethanol. In the case of water to cyclohexane (Fig. SIV(a), ESI<sup>†</sup>), experimental findings<sup>43</sup> are also included for comparison. Irrespective of the polar nature of the amino acid side chain equivalent, Fig. SIV(b) (ESI<sup>†</sup>) generally shows a significant entropy gain with large variations from one amino acid side chain equivalent to another, and an equivalently large enthalpy loss especially for polar molecules, as expected. These results are also in reasonable agreement with experimental findings.<sup>43</sup> A similar trend is also visible in the case of water to ethanol, but in this case a splitting of the cluster of hydrophobic molecules from the cluster of polar ones is very appreciable, in line with the results of Fig. 5c.

## 4 Conclusions

In this paper we have addressed the following issue. Imagine there are two molecules, one polar and one hydrophobic, and two solvents, one polar and one non-polar (hydrophobic). In water we expect a negative solvation free energy for the polar molecule, indicating the propensity of the molecule for being hydrated, as well as a weakly positive solvation free energy for the hydrophobic one, because in this case there is no gain in being hydrated. What happens in a non-polar (hydrophobic) solvent? Is the opposite true? To this aim we have performed detailed thermodynamic integration calculations to compute the solvation free energy of 18 amino acid side chain equivalents in water, cyclohexane and ethanol, the latter representing an intermediate case between a paradigmatic polar solvent such as water, and an equally paradigmatic non-polar solvent such as cyclohexane. Our findings strongly suggest the answer to the above question to be negative, as we did not find any indication of a symmetry between the two cases. We ascribed this to the different interactions, polar–polar in the case of polar amino acid side chain equivalents in water, van der Waals/quadrupolar in the case of hydrophobic amino acid side chain equivalents in cyclohexane. While these numerical simulations are notoriously difficult and very sensitive to the details of the used force fields, we believe that our evidence is sufficient in view of the reasonable agreement with the available results, to make the above statement relatively sound. By repeating the calculations at

different temperatures, we have been able to discriminate between the entropic and the enthalpic contributions. In water we found in all cases an entropy–enthalpy compensation, albeit with some unexpected and intriguing anomalies, to be in agreement with our expectations and past literature. No such compensation appears in the case of cyclohexane, thus supporting the above claim. Remarkably, a cleaner trend with no anomalies is found in the case of ethanol, with the hydrophobic and polar amino acid side chain equivalents arranging in two clearly separated clusters.

Our findings provide new insights into the biological role and the detailed mechanism of the hydrophobic effect, which is known to play a fundamental role in essentially all biological processes. In addition, they also suggest the possibility of defining a robust scheme to identify the relative polarities of the natural amino acids, thus rationalizing the use of the different scales of hydrophobicity that have been proposed in the literature.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We are indebted to Paul Dupire and Emanuele Petretto for their help in the initial stage of the project. The use of the SCSCF multiprocessor cluster at the Università Ca' Foscari Venezia and of the high performance computer Talapas at the University of Oregon is gratefully acknowledged. We also acknowledge the CINECA projects HP10CYJPYK and HP10CGFUUDT for the availability of high performance computing resources through the ISCRA initiative. The work was supported by MIUR PRIN-COFIN2017 Soft Adaptive Networks grant 2017Z55KCW (A. G.), Marie Skłodowska-Curie Fellowship No. 894784-EMPHABIOSIS and a Knight Chair to Prof. Jayanth Banavar at University of Oregon (T. Š.), and Erasmus mobility program (M. H.). The authors would like to acknowledge the networking support by the COST Action CA17139.

## Notes and references

- 1 C. Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, J. Wiley, 2nd edn, 1980.
- 2 C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry: Part II: The Behavior of Biological Macromolecules (Their Biophysical Chemistry; PT. 2)*, W. H. Freeman, 1st edn, 1980.
- 3 A. V. Finkelstein and O. Ptitsyn, *Protein Physics, Second Edition: A Course of Lectures (Soft Condensed Matter, Complex Fluids and Biomaterials)*, Academic Press, 2nd edn, 2016.
- 4 C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo, *Sci. Rep.*, 2016, **6**, 1–9.
- 5 A. M. Klibanov, *Nature*, 2001, **409**, 241–246.
- 6 C. N. Pace, S. Trevino, E. Prabhakaran and J. M. Scholtz, *Philos. Trans. R. Soc. London, Ser. B*, 2004, **359**, 1225–1235.
- 7 N. Hirota, K. Mizuno and Y. Goto, *J. Mol. Biol.*, 1998, **275**, 365–378.
- 8 T. Hayashi, S. Yasuda, T. Škrbić, A. Giacometti and M. Kinoshita, *J. Chem. Phys.*, 2017, **147**, 125102.
- 9 T. Hayashi, M. Inoue, S. Yasuda, E. Petretto, T. Škrbić, A. Giacometti and M. Kinoshita, *J. Chem. Phys.*, 2018, **149**, 045105.
- 10 M. Carrer, T. Škrbić, S. L. Bore, G. Milano, M. Cascella and A. Giacometti, *J. Phys. Chem. B*, 2020, **124**, 6448–6458.
- 11 H. Sandström and M. Rahm, *Sci. Adv.*, 2020, **6**, eaax0272.
- 12 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664–1670.
- 13 R. Wolfenden, *J. Gen. Physiol.*, 2007, **129**, 357–362.
- 14 A. Villa and A. Mark, *J. Comput. Chem.*, 2002, **23**, 548–553.
- 15 J. Chang, A. M. Lenhoff and S. I. Sandler, *J. Phys. Chem. B*, 2007, **111**, 2098–2106.
- 16 M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *J. Chem. Phys.*, 2003, **119**, 5740–5761.
- 17 M. Schauerl, M. Podewitz, B. J. Waldner and K. R. Liedl, *J. Chem. Theory Comput.*, 2016, **12**, 4600–4610.
- 18 D. Frenkel and B. Smit, *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)*, Academic Press, 2nd edn, 2001.
- 19 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.
- 20 P. V. Klimovich, M. R. Shirts and D. L. Mobley, *J. Comput.-Aided Mol. Des.*, 2015, **29**, 397–411.
- 21 A. Ben-Naim, *Hydrophobic Interactions*, Springer Science & Business Media, 1980.
- 22 A. D. Ruiter and C. Oostenbrink, *J. Chem. Theory Comput.*, 2016, **12**, 4476–4486.
- 23 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.
- 24 D. R. Lide, *CRC Handbook of Chemistry and Physics*, CRC press, 2004, vol. 85.
- 25 T. Hajari and N. F. A. van der Vegt, *J. Chem. Phys.*, 2015, **142**, 144502.
- 26 G. Graziano, *Phys. Chem. Chem. Phys.*, 2019, **21**, 21418–21430.
- 27 R. A. Pierotti, *Chem. Rev.*, 1976, **76**, 717–726.
- 28 K. B. Koziara, M. Stroet, A. K. Malde and A. E. Mark, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 221–233.
- 29 N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. van Gunsteren, *Eur. Biophys. J.*, 2011, **40**, 843.
- 30 S. W. Siu, K. Pluhackova and R. A. Bockmann, *J. Chem. Theory Comput.*, 2012, **8**, 1459–1470.
- 31 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19–25.
- 32 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 33 H. J. Berendsen, J. V. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- 34 B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 35 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, in *Interaction Models for Water in Relation to Protein Hydration*, ed. B. Pullman, Springer, Netherlands, Dordrecht, 1981, pp. 331–342.
- 36 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

- 37 S. Suzuki, P. Green, R. Bumgarner, S. Dasgupta, W. Goddard and G. Blake, *Science*, 1992, **257**, 942–944.
- 38 K. P. Gierszal, J. G. Davis, M. D. Hands, D. S. Wilcox, L. V. Slipchenko and D. Ben-Amotz, *J. Phys. Chem. Lett.*, 2011, **2**, 2930–2933.
- 39 Y. Nozaki and C. Tanford, *J. Biol. Chem.*, 1971, **246**, 2211–2217.
- 40 S. Damodaran and K. Song, *J. Biol. Chem.*, 1986, **261**, 7220–7222.
- 41 C. Tanford, *J. Am. Chem. Soc.*, 1962, **84**, 4240–4247.
- 42 H. S. Ashbaugh and L. R. Pratt, *Rev. Mod. Phys.*, 2006, **78**, 159.
- 43 R. Wolfenden, C. A. Lewis, Y. Yuan and C. W. Carter, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 7484–7488.
- 44 R. Wolfenden, L. Andersson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, 1981, **20**, 849–855.
- 45 G. Gerogiokas, G. Calabro, R. H. Henchman, M. W. Y. Southey, R. J. Law and J. Michel, *J. Chem. Theory Comput.*, 2014, **10**, 35–48.
- 46 S. J. Irudayam, R. D. Plumb and R. H. Henchman, *Faraday Discuss.*, 2010, **145**, 467–485.
- 47 G. Graziano, *Chem. Phys. Lett.*, 2009, **479**, 56–59.
- 48 R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 13052–13056.
- 49 H. S. Frank and M. W. Evans, *J. Chem. Phys.*, 1945, **13**, 507–532.