

Can the roles of polar and non-polar moieties be reversed in non-polar solvents?

Cedrix J. Dongmo Fomthuil^{||, 1, a)} Manuel Carrer^{||, 2, b)} Maurine Houvet,^{3, c)} Tatjana Škrbić,^{1, 4, d)} Giuseppe Graziano,^{5, e)} and Achille Giacometti^{*1, 6, f)}

¹⁾*Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia, Campus Scientifico, Edificio Alfa, via Torino 155, 30170 Venezia Mestre, Italy.*

²⁾*Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, PO Box 1033 Blindern, 0315 Oslo, Norway.*

³⁾*Polytech Nantes -Engineering school of the University of Nantes, Rue Christian Pauc, 44306 Nantes cedex 3.*

⁴⁾*Department of Physics and Institute for Fundamental Science, University of Oregon, Eugene, OR 97403, USA.*

⁵⁾*Department of Science and Technology, University of Sannio-Benevento, via Francesco de Sanctis, 82100 Benevento, Italy.*

⁶⁾*European Centre for Living Technology (ECLT) Ca' Bottacin, Dorsoduro 3911, Calle Crosera 30123 Venice, Italy.*

(Dated: 14 October 2020)

Using thermodynamics integration, we study the solvation free energy of 18 amino acid side chain equivalents in solvents with different polarity, 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 biological solvent.

I. INTRODUCTION

The hydrophobic effect refers to the tendency of non-polar moieties to avoid the contact with water molecules and form an aggregate¹. This is also commonly viewed as one of the main driving force underlying the folding of a protein^{2,3}. 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 to bury as much as possible the hydrophobic amino acids inside the folded chain. A number of concurring effects⁴ 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 solvents⁵. Pace and collaborators⁶ noticed

that folded proteins become unstable in polar solvents such as ethanol EtOH, but return to be very stable, albeit essentially insoluble, in non-polar solvents such as cyclohexane C_6H_{12} . 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 α helices⁷. A similar scenario emerged from two recent numerical studies^{8,9} 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 α 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 possessing the ordered 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 form-

^{a)}cedrix.dongmo@unive.it

^{b)}manuel.carrer@kjemi.uio.no

^{c)}maurine.houvet@orange.fr

^{d)}tskrbic@uoregon.edu

^{e)}graziano@unisannio.it

^{f)}achille.giacometti@unive.it

^{||}These authors contributed equally to this work

* Corresponding author : achille.giacometti@unive.it

ing 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 authors¹⁰, and that may be relevant for the existence of unconventional life forms in other planets where water is not available¹¹. 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 fully 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 then 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 Wolfenden’s lab^{12,13}, 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^{14,15}, and comparing results for different water force fields¹⁶. Another study also addressed the inclusion of the amino acids backbones¹⁷.

Using thermodynamic integration¹⁸, 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 on 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 II provides the general background of our analysis; results are presented in Section III, and Section IV will provide some summarizing take-home messages. Additional tables and figures can be found in Appendixes.

II. THEORY AND METHODS

A. Thermodynamic integration

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

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

If $\Delta G_{solv} < 0$ ($\Delta G_{solv} > 0$) the solvent is stabilizing (destabilizing) 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 ΔG_{S_1} and ΔG_{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 expression¹⁸

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

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

Best practices in free energies calculations^{19,20} suggest the use of alchemical transformation in the form of a thermodynamic cycle as defined in Figure 1(a)^{19–21}. Firstly, all intramolecular non-bonded interactions in the solute compound are turned off to obtain the dummy compound in vacuum. Let ΔG_1 be the free energy difference associated with this transition. Then, the dummy compound is transferred from vacuum to the solvent – the liquid in Figure 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 ΔG_3 to achieve the final compound in the solvent (liquid). Then, clearly $\Delta G_{solv} = \Delta G_1 + \Delta G_2 + \Delta G_3$, as summarized in Figure 1 (a). Note that in the presence of steric interactions only, ΔG_{solv} is purely of entropic nature and can be estimated using Scaled Particle Theory (SPT) (see Section II). In practice, however, a direct calculation ΔG_{solv} can nowadays be achieved using an efficient application of thermodynamic integration²² as schematically shown in Figure 1(b). Here the solute is inserted into a pre-equilibrated solvent, and parallel simulations are involving energy minimization, NVT and NPT equilibration, and production runs are computed for several intermediate values of the coupling parameter λ , and then combined using Bennet’s acceptance ratio²³ to finally obtain

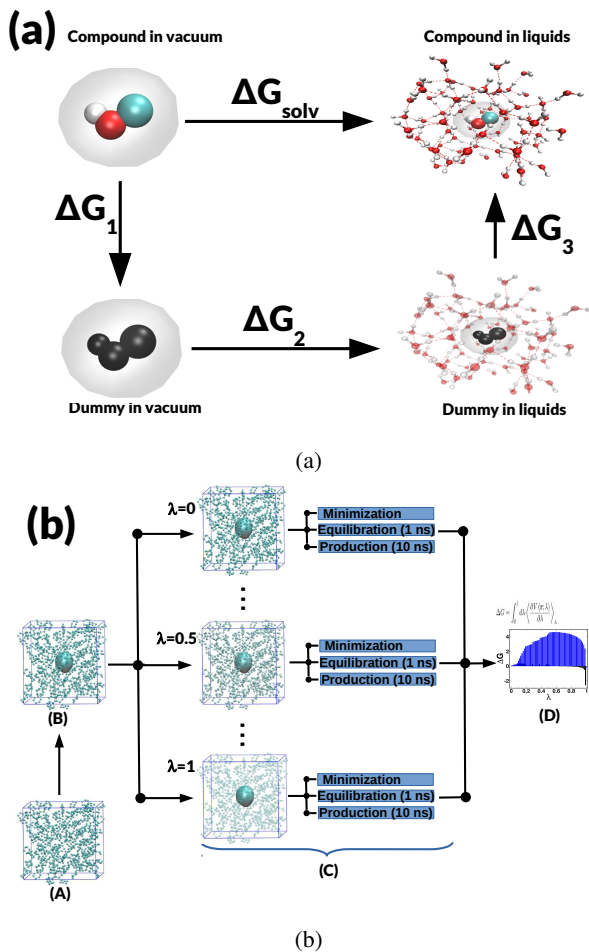


FIG. 1: **(a)** A thermodynamic cycle allowing the computation of the solvation free energy ΔG_{solv} . **(b)** Simulation workflow illustrated for the case of cyclohexane. **(A)** The simulation starts with a pre-equilibrated box of solvent; **(B)** The solute is then inserted into the equilibrated box; **(C)** Parallel simulations are performed for each value of the coupling parameter λ . This includes an energy minimization (steepest descent + 1-bfgs), an equilibration (NVT + NPT) and a production steps, as shown; **(D)** The intermediate values of lambda are combined using the 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.

the required solvation free energy. Here $\lambda = 0$ refers to fully coupled case, whereas $\lambda = 1$ to the fully uncoupled case. See Section II C for details.

Fig. 6 depicts 18 of the 20 natural amino acids that are conventionally divided in hydrophobic (non-polar) Fig. 6a, and polar (hydrophilic) Fig. 6b. 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 on these two classes. Two amino acids have special features and hence have not been in-

cluded in Fig. 6: 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 also visible in each of the 18 amino acids of Fig. 6. 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 Figure 2 where each equivalent is identified by the short hand notation of its natural side chain counterpart, as reported in Table I.

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 ϵ_r that is 78.5 in water H_2O , 24.3 in ethanol $EtOH$, and 2.0 in cyclohexane cC_6H_{12} at $T = 298 K$ ²⁴. Accordingly, cyclohexane is much more hydrophobic than water and relatively more hydrophobic than ethanol. The rationale behind this choice of course stems for 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 interactions (quadrupole, van der Waals, etc.) appearing in the absence of a permanent dipole. By computing the solvation free energy $\Delta G_w \equiv \Delta G_{H_2O}$ of each of these amino acid side chain equivalents in water, and then the solvation free energies $\Delta G_c \equiv \Delta G_{cC_6H_{12}}$ and $\Delta G_e \equiv \Delta G_{EtOH}$ 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 $\Delta\Delta G_{w>c} = \Delta G_c - \Delta G_w \equiv \Delta G_{cC_6H_{12}} - \Delta G_{H_2O}$, and water-ethanol $\Delta\Delta G_{w>e} = \Delta G_e - \Delta G_w \equiv \Delta G_{EtOH} - \Delta G_{H_2O}$. 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 ago¹.

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 an 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 a differentiation with respect to the temperature. To this aim, we assume the following temperature dependence for the free energy difference²⁵

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

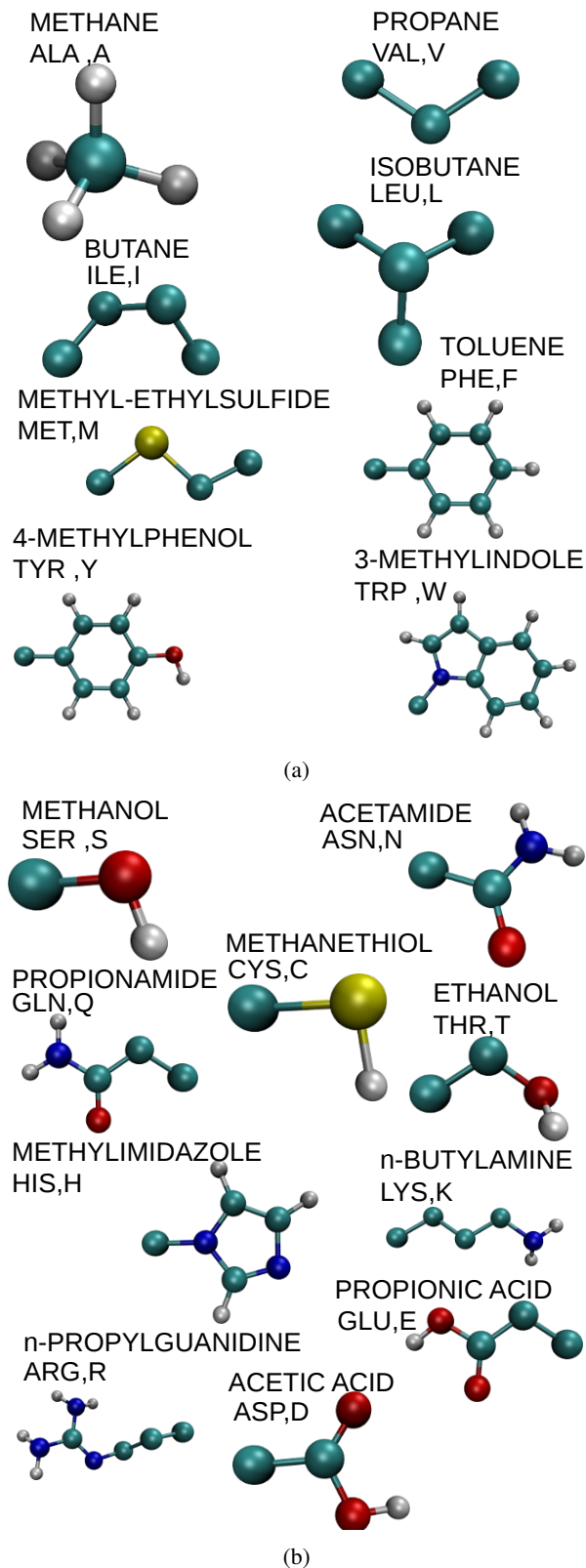


FIG. 2: (a) Hydrophobic amino acid side chain equivalents. (b) Polar amino acid side chain equivalents.

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 Eq.(4) based on the results of simulations at different temperatures, will provide the required expressions for the entropy (Eq.(5)) and for the enthalpy (Eq.(6)). Standard deviation was evaluated using error block analysis²⁵.

A word of caution is in order here. As discussed in Ref.²⁵, the functional form given in Eq.(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 – 330K in the present study. This is comparable with 278 – 338K considered in Ref.²⁵ where only water was investigated. Moreover, temperature 270K 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, 270K does not appear to be a particular outlier to the fitted curves, see Fig .11 - Fig.13.

B. Scaled Particle Theory (SPT)

According to a general theory of solvation²⁶, ΔG_{solv} can be calculated as the sum of the reversible work spent to create a cavity suitable to host the solute molecule, ΔG_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 ΔG_0 in any liquid can be calculated by means of the analytical relationships provided by classic Scaled Particle Theory (SPT)²⁷. 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 purely van der Waals attractions, whereas numerical calculations are in general necessary in the case of hydrogen bonds. Additional details on the theoretical aspects can be found in the original paper²⁶.

C. 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 position CB 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)²⁸.

While a united atom representation was used in setting up the systems, an in house modification of GROMOS96 (54a7) force field²⁹ was required to account for non-natively parameterized molecules. The choice of this force field is in line with past work¹⁴ where it was shown to provide good description of the solvent properties. Here we have used the latest 54a7 version of Gromos force field, while Villa *et al*¹⁴ employed the 43a2 version. A good alternative for cyclohexane would have been the most recent version of the optimized OPLS (L-OPLS)³⁰ 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.*¹⁵ 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^{16,17} provide comparable performances.

The simulations were performed in three different solvents covering a broad range of polarity, from non-polar cyclohexane C_6H_{12} , to highly polar water H_2O , through the intermediate polar ethanol EtOH . The 18 amino acid side chain equivalents were then inserted into a cubic box of 3 nm in size incorporating about 165, 290 and 881 molecules of C_6H_{12} , EtOH , and H_2O , respectively. The simulations were performed with Gromacs simulation package (versions 2018.3 and 2018.7)³¹ and all the solutes were modeled in their neutral uncharged states. As detailed in Section II, free energy differences as given by Eq.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 a 21 binning points. See Fig. 8. The initial systems were initially prepared by minimizing the solute’s potential energy and relaxing the solvent around 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 5×10^{-4} nm and a maximum convergence force of $100.0 \text{ kJmol}^{-1} \text{ nm}^{-1}$ followed by 5×10^4 steps of 1-bfgs quasi-Newtonian minimization algorithm with a minimization step size of 10^{-3} nm and a maximum convergence force of $100.0 \text{ kJmol}^{-1} \text{ nm}^{-1}$. Thereafter, an equilibration round in the canonical NVT ensemble was performed for 500ps using the accurate leap-frog stochastic dynamics integrator with the simulation time step of 2fs. Due to the large numerical fluctuations recorded, a shorter time step of 0.5fs or 1fs was used in some simulations. While long-range electrostatics interactions were accounted with the Particle Mesh Ewald summation³², 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 – 330K. Each temperature was kept around the reference value by coupling the system to an external bath using the Berendsen thermostat (for less stable systems)³³, with a

coupling constant of 1.0 ps. All simulations were replicated in a 3D bulk-like phase using the periodic boundaries conditions and all bonds involving hydrogen atoms were restrained using LINCS algorithms³⁴. For water we used the simple point charge SPC water model³⁵, whereas the parameters (or topology) for ethanol and cyclohexane were manually implemented. Small quantitative differences could be expected¹⁷ by choosing more refined force fields for water, at the expenses of a significant increase in the computational time. The second equilibration round was then performed for additional 500ps 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)³⁶ and the weak-coupling Berendsen barostat (for less stable systems)³³, with a coupling constant of 1.0ps. The isothermal compressibility (in bar^{-1}) of 4.5×10^{-5} , 1.2×10^{-4} , and 4.5×10^{-5} was used for water, ethanol and cyclohexane, respectively. The final production runs for free energy calculation were performed for 10ns 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 also the coulomb interactions were considered. In short, we have closely followed the numerical protocol by Villa *et al.*¹⁴, but we have improved it with 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*¹⁴ used 2 fs. Our simulation workflow is shown in Fig. 1b.

III. RESULTS

A. Solvation free energy ΔG_{solv}

Figure 3 displays the solvation free energy for water H_2O (Figure 3a), cyclohexane C_6H_{12} (Figure 3b), and ethanol EtOH (Figure 3c), and compares the results of the present work with both experimental and computational past works. All corresponding values can be found in II, III, IV. Broadly speaking, the solvation free energies follow the general division in hydrophobic and polar amino acids illustrated in Fig. 2 – note the sequence of the amino acids of Fig. 3 from left to right follow the same scheme hydrophobic \rightarrow polar division of Fig. 2, but there are exceptions. In water (Fig. 3a), the ΔG_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 $\Delta G_w < 0$. While odd at first sight, we note that this is in line with experimental data. For instance, for toluene $\Delta G_w = -3.7 \text{ kJmol}^{-1}$ at 25°C and 1 atm, (see V). 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 hy-

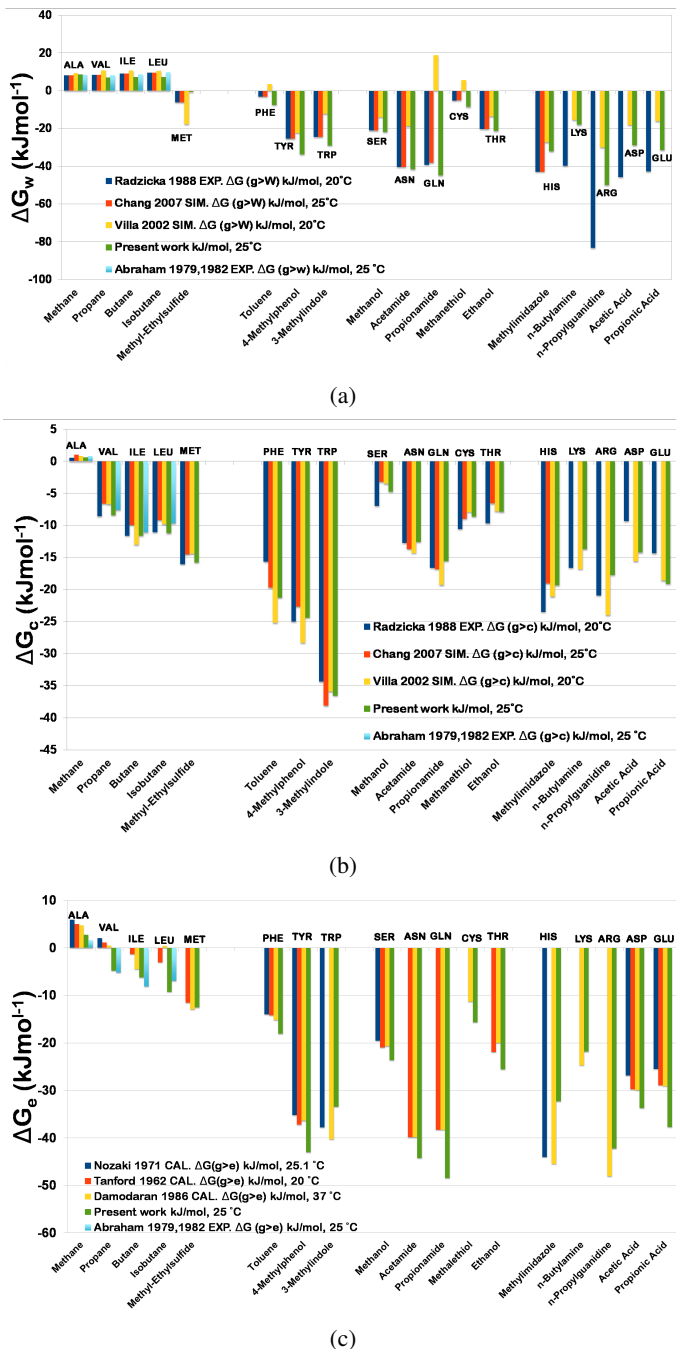


FIG. 3: (a) ΔG_w from vacuum to water H_2O at $25^\circ C$; (b) ΔG_c from vacuum to cyclohexane cC_6H_{12} ; (c) ΔG_e from vacuum to ethanol $EtOH$. Results of the present work are also contrasted with past computational work of Refs.^{14,15}, as well as with experimental work of Ref.¹². Each amino acid side chain equivalent is referred with the name (x -axis) and with the shorthand notations of the corresponding amino acid at the top of each value.

drogen atom attached to very electronegative atoms (i.e. O or N) interact favorably with the delocalized π electrons of the aromatic ring³⁷. This is a specific example of a more general class of $A-H \cdots \phi$ hydrogen bonds, where ϕ represents an aromatic ring and A may be a N, O, or C atom³⁸. In particular, this weak hydrogen bonds can form among water molecules and the aromatic side chains of PHE, TYR and TRP³⁸. 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 10 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 λ). In the case of hydrophobic amino acid side chain equivalents our findings are in general good agreement with both experimental results¹², as well as past computational results^{14,15}, with the exception of Methyl-ethylsulfide (MET) and Toluene (PHE). This could be ascribed to a more general 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 past ones, thus supporting the reliability of the adopted force field.

In parallel to the case of water, our estimates of the solvation free energy in cyclohexane 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 data¹² and previous computational investigations^{14,15}. As cyclohexane cC_6H_{12} is a nonpolar liquid, unable to form hydrogen bonds, the negative ΔG_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 cyclohexane 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 ΔG_c obtained in the present work are closer to experimental data than previously calculated values (see Table III).

The solvation free energy ΔG_e in ethanol $EtOH$ (see Fig. 3c) is found negative for all amino acid side chain equivalents, but methane (ALA), paralleling the situation in cyclohexane cC_6H_{12} (compare Figures 3b and 3c), and in line with experimental data (see Table IV). We are not aware of any previous computational study providing estimates of ΔG_e for all the amino acid side chain equivalents considered here, so only limited comparisons can be carried out³⁹⁻⁴¹. Note that here the temperatures are also different. As a further remark, we stress that experimental ΔG_e by Nozaki and Tanford³⁹ were obtained by subtracting the ΔG_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 ethanol $EtOH$ is a polar solvent able to form hydrogen bonds, its behavior resembles that of cyclohexane cC_6H_{12} . This means that here too the attractive solute-solvent energetic interactions (accounting also for hy-

drogen bonds) are able to overcome the free energy cost of cavity creation. Not surprisingly, in fact, the largest $|\Delta G_e|$ are associated with solutes, such as acetamide (ASN) and propionamide (GLN), able to be engaged in multiple hydrogen bonds with ethanol EtOH molecules.

In this respect, it proves instructive to compare present findings with results that can be obtained from SPT and related theories, as discussed in Section II B^{26,42}. Here solvation free energy can be estimated in all liquids as the sum of two contributions: (a) the reversible work ΔG_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; (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 purely energetic so that $E_a < 0$ always. In other words, E_a favors solvation, whereas ΔG_0 opposes it. Reliable estimates for ΔG_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 $\Delta G_{0,w} > \Delta G_{0,e} > \Delta G_{0,c}$ for water H₂O, ethanol EtOH, and cyclohexane cC₆H₁₂ in decreasing order, see 7. 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, that 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 water H₂O, slightly less in ethanol EtOH, and zero in cyclohexane cC₆H₁₂. Using this method sketched in Section II B, we estimated the solvation free energies for methane (ALA), propane (VAL), toluene (PHE) and methanol (SER) as $\Delta G_{0,e} + E_a$ and find them in agreement with experimental values in the three considered liquids, as shown in Table V. Hence (a) in water H₂O, $\Delta G_{0,w} > |E_a|$ for aliphatic hydrocarbons, whereas the opposite holds true for aromatic hydrocarbons and polar molecules able to form hydrogen bonds with water molecules; (b) in cyclohexane cC₆H₁₂ and ethanol EtOH, $|E_a| > \Delta G_{0,e,c}$ for all the amino acids side chain equivalents but methane (ALA), because the free energy cost of cavity creation is not so large.

B. 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. Figure 4a depicts our results for the free energy transfer $\Delta\Delta G_{w>c} \equiv \Delta\Delta G(\text{H}_2\text{O} \rightarrow \text{cC}_6\text{H}_{12})$ from water H₂O to cyclohexane cC₆H₁₂ and contrast them with the past simulations^{14,15} and experiments⁴³. 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

cyclohexane cC₆H₁₂ rather than a polar solvent such water H₂O. 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 cyclohexane cC₆H₁₂ rather than water H₂O.

The present values are quantitatively close to experimental data⁴³ and perform better than previously calculated estimates^{14,15}. For instance, $\Delta\Delta G_{w>c}$ for propionamide (GLN) is positive and in line with the experimental value, whereas a previously numerical estimate was negative (see Fig. 4a). Remarkably, $\Delta\Delta G_{w>c}$ almost perfectly divides polar amino acid side chain equivalents from 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 Wolfenden⁴⁴. This finding is also an indication that cyclohexane cC₆H₁₂, such as other non-polar organic liquids, does not act as a denaturant of the folded state of globular proteins, in agreement with experimental evidence⁶.

The values of the transfer free energy $\Delta\Delta G_{w>e} \equiv \Delta\Delta G(\text{H}_2\text{O} \rightarrow \text{EtOH})$ from water H₂O to ethanol 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 data⁴⁰, even though the latter are very small 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 V).

Once more, it proves instructive to contrast the above findings with theoretical results stemming from the SPT analysis of Section II B. Here, we can build on the fact that the reversible work of cavity creation $\Delta G_{0,w}$ in water H₂O is larger than its counterpart $\Delta G_{0,e}$ in ethanol EtOH, that is $\Delta G_{0,w} > \Delta G_{0,e}$ (see Table V). By contrast, the reversible work of turning on solute-solvent attractions, in water $E_{a,w}$ H₂O is approximately equal to its counterpart $E_{a,e}$ in ethanol, ($E_{a,w} \approx E_{a,e}$), so that $\Delta G_{0,w} + E_{a,w} > \Delta G_{0,e} + E_{a,e}$, 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 studies³⁹.

C. 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, an useful approach is provided by the grid cell theory⁴⁵ that is a refined version of partition function methods⁴⁶. It is then of considerable interest to ask how the present study can contribute to this issue.

As anticipated in Section II A we follow the work of Schauerl et al.¹⁷, to separate the free energy of solvation in its en-

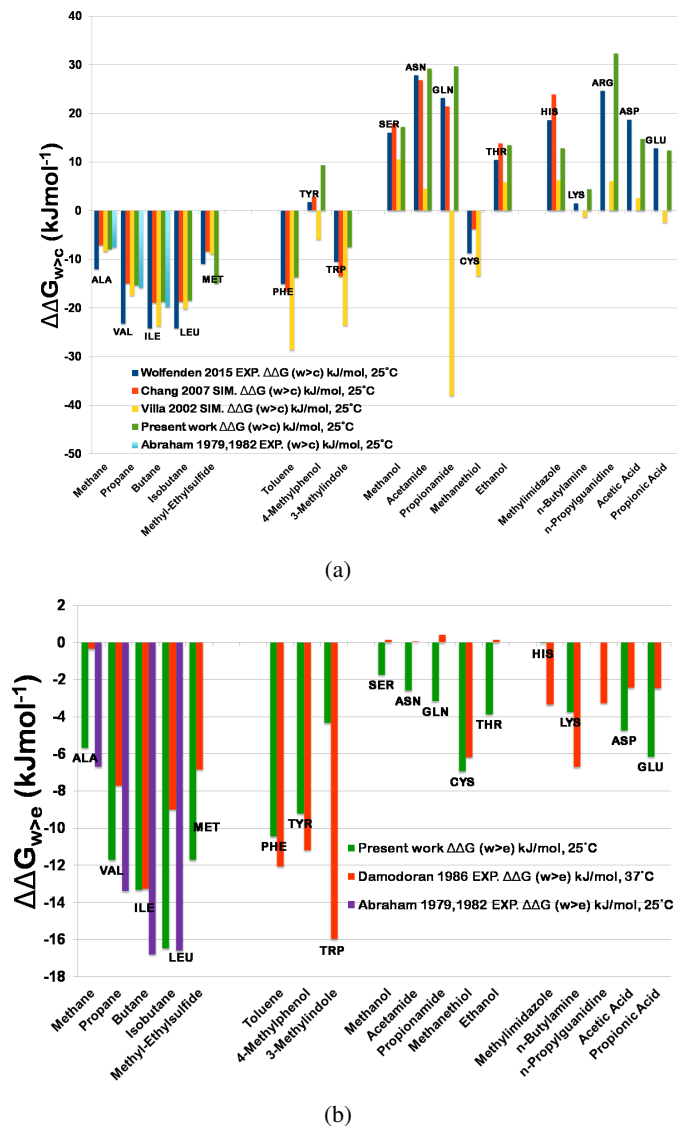


FIG. 4: (a) $\Delta\Delta G_{w>c}$ from water H₂O to cyclohexane cC₆H₁₂ ; (b) $\Delta\Delta G_{w>e}$ from water H₂O to ethanol EtOH . Results of the present work are also contrasted with past computational work of Refs.^{14,15}, as well as with experimental works of Refs.⁴³ (water H₂O to cyclohexane cC₆H₁₂) and⁴⁰ (water H₂O to ethanol EtOH). Each amino acid side chain equivalent is referred with the name (*x*-axis) and with the shorthand notations of the corresponding amino acid at the top of each value.

thalpic and entropic parts. To this aim, we compute the solvation free energy at seven different temperatures in the range 270 – 330K and then used Eq.(4) to fit the parameters *a*, *b*, and *c* and hence obtain $\Delta S(T)$ from Eq.(5). All details on these calculations can be found in XI, 11, 12, and 13. Fig.5 then reports the entropic term $-T\Delta S$ of the solvation free energy ΔG_{solv} as a function of the enthalpic term ΔH for different solvents: water H₂O (Fig. 5a), cyclohexane cC₆H₁₂ (Fig. 5b) and ethanol EtOH (Fig. 5c).

Visual inspection of the plots in Fig. 11, 12, and 13 indicates that: (a) For water H₂O $\Delta G_w(T)$ 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 data⁴⁷ (this is a further support of the reliability of the adopted force fields and calculation procedure); (b) For cyclohexane cC₆H₁₂ $\Delta G_c(T)$ does not have a temperature dependence common to all amino acid side chain equivalents; (c) For ethanol 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 Eqs.(4) and Eqs.(5) leads to the calculated solvation enthalpy and entropy values at 298.15 K listed in Table VI, Table VII, and Table VIII. These values have been used to build up plots of $-T\Delta S$ versus ΔH for all amino acid side chain equivalents in the three liquids, as displayed in Figure 5.

The general expectation is that in water a large and negative ΔH 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 water H₂O as shown in Fig. 5a. For instance, for toluene (PHE) (a) In water H₂O $\Delta H = -53\text{kJ mol}^{-1}$ and $-T\Delta S = 46\text{kJ mol}^{-1}$; (b) In cyclohexane cC₆H₁₂ $\Delta H = -46\text{kJ mol}^{-1}$ and $-T\Delta S = 24\text{kJ mol}^{-1}$; (c) In ethanol EtOH $\Delta H = -29\text{kJ mol}^{-1}$ and $-T\Delta S = 10\text{kJ mol}^{-1}$. The large difference among 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 to the isobaric thermal expansion coefficient α_p of the liquid. Indeed at room temperature 298 K and 1 atm, α_p is very small in water, but large in organic liquids ($\alpha_p = 0.257 \times 10^{-3}$ for water H₂O, 1.214×10^{-3} for cyclohexane cC₆H₁₂, and 1.089×10^{-3} for ethanol EtOH in K⁻¹)²⁴, giving a simple explanation of the different magnitude of such structural reorganization in the three solvents. The net distinction between polar and hydrophobic amino acid side chain equivalents occurring in ethanol EtOH is noteworthy and in striking contrast with the lack of such a separation in cyclohexane cC₆H₁₂ (compare Figs. 5b and 5c). Unfortunately, a detailed comparison with experimental or computational results is not possible due to a lack of such data for most of the amino acid side chain equivalents considered in the present study. As expected, in water H₂O (Fig. 5a) polar amino acid side chain equivalents have large and negative ΔH , 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 cage⁴⁹ 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

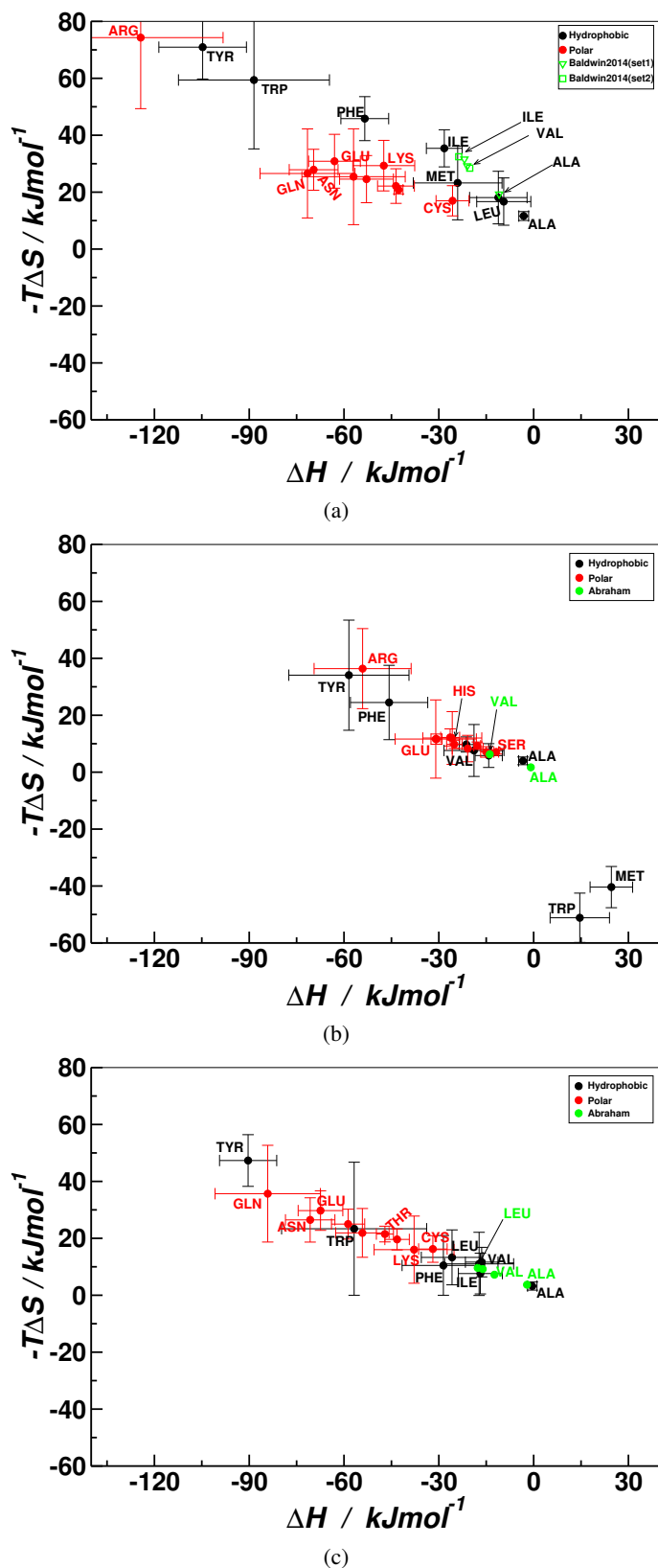


FIG. 5: Entropic part $-T\Delta S$ of the solvation free energy ΔG as a function of the enthalpic part ΔH in the case of (a) water H_2O ; (b) cyclohexane cC_6H_{12} ; (c) ethanol $EtOH$. In case of water H_2O , corresponding experimental results⁴⁸ are also included. Note that all plots are in the same scale.

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 nearly unaffected in terms of enthalpy change and show a small increase in $-T\Delta S$ originating by 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 measurements⁴⁸. Here too, 4-methylphenol (TRP) and toluene (PHE) appear to be outliers with large and negative ΔH 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.¹⁷ (see their Figure 5) because their amino acids include the backbone that in our case is represented by a single hydrogen. Their analysis further show the sensibility 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 in water, however, appears to be ≈ 0.5 consistent with them.

Fig. 5b provides the same plot in cyclohexane cC_6H_{12} . Here it is rather evident that we do not obtain the same results by inverting the role of polar and hydrophobic amino acid side chain equivalents. While a negative slope is overall visible, most of the amino acid side chain equivalents tend to cluster in a region of small enthalpic gain $\Delta H \approx -30 \text{ kJ mol}^{-1}$ and 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 $-T\Delta S \approx 30 \text{ kJ mol}^{-1}$. At 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 previous Sections, the results in the case of ethanol $EtOH$ 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 water H_2O and ethanol $EtOH$ so the gain in enthalpy is sufficient to compensate this entropy loss. On the other hand, the reverse is not true in cyclohexane 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 ones.

Finally, even in this case it is useful to perform the same calculation by transferring the molecules from water to either cyclohexane or ethanol. This is reported in Fig. 9(a) for the case from water to cyclohexane and in Fig. 9(b) for the case from water to ethanol. In the case of water to cyclohexane (Fig. 9 (a)), experimental findings⁴³ are also included for comparison. Irrespective of the polar nature of the amino acid side chain equivalent, Fig. 9(b) 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 findings⁴³. A similar trend is also visible in case from 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.

IV. CONCLUSIONS

In this paper we have addressed the following issue. Imagine to have 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 thermodynamics 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 sensible to the details of the used force fields, we believe that our evidence is sufficient in view of the reasonable agreement with past 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, 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 on the biological role and the detailed mechanism of the hydrophobic effect, that

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 zoo of 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 at 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 HP10CGFUDDT for the availability of high performance computing resources through the IS CRA 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.S), and Erasmus mobility program (M.H). The authors would like to acknowledge networking support by the COST Action CA17139.

- ¹C. Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes 2nd ed*, J. Wiley., 1980.
- ²C. 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.
- ³A. 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.
- ⁴C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo, *Scientific Reports*, 2016, **6**, 1–9.
- ⁵A. M. Klibanov, *Nature*, 2001, **409**, 241–246.
- ⁶C. N. Pace, S. Trevino, E. Prabhakaran and J. M. Scholtz, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 2004, **359**, 1225–1235.
- ⁷N. Hirota, K. Mizuno and Y. Goto, *Journal of Molecular Biology*, 1998, **275**, 365–378.
- ⁸T. Hayashi, S. Yasuda, T. Škrbić, A. Giacometti and M. Kinoshita, *The Journal of Chemical Physics*, 2017, **147**, 125102.
- ⁹T. Hayashi, M. Inoue, S. Yasuda, E. Petretto, T. Škrbić, A. Giacometti and M. Kinoshita, *The Journal of Chemical Physics*, 2018, **149**, 045105.
- ¹⁰M. Carrer, T. Škrbić, S. L. Bore, G. Milano, M. Cascella and A. Giacometti, *The Journal of Physical Chemistry B*, 2020, **124**, 6448–6458.
- ¹¹H. Sandström and M. Rahm, *Sci. Adv.*, 2020, **6**, eaax0272.
- ¹²A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664–1670.
- ¹³R. Wolfenden, *Journal of General Physiology*, 2007, **129**, 357–362.
- ¹⁴A. Villa and A. Mark, *Journal of Computational Chemistry*, 2002, **23**, 548–553.
- ¹⁵J. Chang, A. M. Lenhoff and S. I. Sandler, *The Journal of Physical Chemistry B*, 2007, **111**, 2098–2106.
- ¹⁶M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *The Journal of Chemical Physics*, 2003, **119**, 5740–5761.
- ¹⁷M. Schauerl, M. Podewitz, B. J. Waldner and K. R. Liedl, *Journal of Chemical Theory and Computation*, 2016, **12**, 4600–4610.

- ¹⁸D. Frenkel and B. Smit, *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)*, Academic Press, 2nd edn, 2001.
- ¹⁹A. Pohorille, C. Jarzynski and C. Chipot, *The Journal of Physical Chemistry B*, 2010, **114**, 10235–10253.
- ²⁰P. V. Klimovich, M. R. Shirts and D. L. Mobley, *Journal of Computer-Aided Molecular Design*, 2015, **29**, 397–411.
- ²¹A. Ben-Naim, *Hydrophobic Interactions*, Springer Science & Business Media, 1980.
- ²²A. d. Ruiter and C. Oostenbrink, *Journal of Chemical Theory and Computation*, 2016, **12**, 4476–4486.
- ²³C. H. Bennett, *Journal of Computational Physics*, 1976, **22**, 245 – 268.
- ²⁴D. R. Lide, *CRC Handbook of Chemistry and Physics*, CRC press, 2004, vol. 85.
- ²⁵T. Hajari and N. F. A. van der Vegt, *Journal of Chemical Physics*, 2015, **142**, 144502.
- ²⁶G. Graziano, *Physical Chemistry Chemical Physics*, 2019, **21**, 21418–21430.
- ²⁷R. A. Pierotti, *Chemical Reviews*, 1976, **76**, 717–726.
- ²⁸K. B. Koziara, M. Stroet, A. K. Malde and A. E. Mark, *Journal of Computer-Aided Molecular Design*, 2014, **28**, 221–233.
- ²⁹N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. van Gunsteren, *European Biophysics Journal*, 2011, **40**, 843.
- ³⁰S. W. Siu, K. Pluhackova and R. A. Bockmann, *Journal of Chemical Theory and Computation*, 2012, **8**, 1459–1470.
- ³¹M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19 – 25.
- ³²U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.
- ³³H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.
- ³⁴B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *Journal of Computational Chemistry*, 1997, **18**, 1463–1472.
- ³⁵H. 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.
- ³⁶M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.
- ³⁷S. Suzuki, P. Green, R. Bumgarner, S. Dasgupta, W. Goddard and G. Blake, *Science*, 1992, **257**, 942–944.
- ³⁸K. P. Gierszal, J. G. Davis, M. D. Hands, D. S. Wilcox, L. V. Slipchenko and D. Ben-Amotz, *Journal of Physical Chemistry Letters*, 2011, **2**, 2930–2933.
- ³⁹Y. Nozaki and C. Tanford, *Journal of Biological Chemistry*, 1971, **246**, 2211–2217.
- ⁴⁰S. Damodaran and K. Song, *Journal of Biological Chemistry*, 1986, **261**, 7220–7222.
- ⁴¹C. Tanford, *Journal of the American Chemical Society*, 1962, **84**, 4240–4247.
- ⁴²H. S. Ashbaugh and L. R. Pratt, *Reviews of Modern Physics*, 2006, **78**, 159.
- ⁴³R. Wolfenden, C. A. Lewis, Y. Yuan and C. W. Carter, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 7484–7488.
- ⁴⁴R. Wolfenden, L. Andersson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, 1981, **20**, 849–855.
- ⁴⁵G. Gerogiokas, G. Calabro, R. H. Henchman, M. W. Y. Southey, R. J. Law and J. Michel, *Journal of Chemical Theory and Computation*, 2014, **10**, 35–48.
- ⁴⁶S. J. Irudayam, R. D. Plumb and R. H. Henchman, *Faraday Discussions*, 2010, **145**, 467–485.
- ⁴⁷G. Graziano, *Chemical Physics Letters*, 2009, **479**, 56–59.
- ⁴⁸R. L. Baldwin, *Proceedings of the National Academy of Sciences of the United States of America*, 2014, **111**, 13052–13056.
- ⁴⁹H. S. Frank and M. W. Evans, *The Journal of Chemical Physics*, 1945, **13**, 507–532.

Appendix A: Appendixes

Free energy differences between initial and final states can be computed using Eq.A1 below :

$$\Delta G_{AB} = \int_{\lambda_A}^{\lambda_B} d\lambda \left\langle \frac{\partial V(\mathbf{r}; \lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (\text{A1})$$

where $V(\mathbf{r}, \lambda)$ is the potential energy of the system as a function of the coordinate vector \mathbf{r} , and λ is a switching-on parameter allowing to go from state A to state B by changing its value from λ_A to λ_B .

The λ -dependence of the potential in bonded interaction is linear while non-bonded interaction can be described with linear dependence or with Soft-core interaction. It should be noted that in our simulations we are analyzing only small molecules, so we are only interested in turning off the inter-molecular interactions such as Lennard-Jones and Coulomb potentials. We used the standard linear interpolation shown in Eq.A2

$$\begin{aligned} V &= (1 - \lambda)V^A + \lambda V^B \\ \frac{\partial V}{\partial \lambda} &= V^B - V^A \end{aligned} \quad (\text{A2})$$

However, near off-states i.e. for values of λ equal to 0 and 1 large numerical fluctuations are sometimes recorded leading to clashes between decoupling atoms, thereby preventing a smooth derivative of the potential in Eq.A2. A core softening (Eq.A3) interacting potential was used to circumvent this issue

$$\begin{aligned} V_{soft-core}(r) &= (1 - \lambda)V^A(r_A) + \lambda V^B(r_B) \\ r_A &= (\alpha R_A^6 \lambda^p + r^6)^{1/6} \\ r_B &= (\alpha R_B^6 (1 - \lambda)^p + r^6)^{1/6} \end{aligned} \quad (\text{A3})$$

where λ and p are respectively the soft-core and the soft-core power parameters, and R is the interaction radius, which is equal to the ratio between the Lennard-Jones parameters σ_{ij} .

Appendix B: Supplementary Tables

TABLE I: The correspondence between the 20 amino acids and their neutral analog equivalent.

Character	Amino acid	Short name	Single letter	Equivalent
Hydrophobic	Alanine	ALA	A	Methane
Hydrophobic	Valine	VAL	V	Propane
Hydrophobic	Isoleucine	ILE	I	Butane
Hydrophobic	Leucine	LEU	L	Isobutane
Hydrophobic	Methionine	MET	M	Methyl-ethylsulfide
Hydrophobic	Glycine	GLY	G	Hydrogen
Hydrophobic	Phenylalanine	PHE	F	Toluene
Hydrophobic	Tyrosine	TYR	Y	4-Methylphenol
Hydrophobic	Tryptophan	TRP	W	3-Methylindole
Polar	Serine	SER	S	Methanol
Polar	Asparagine	ASN	N	Acetamide
Polar	Glutamine	GLN	Q	Propionamide
Polar	Cysteine	CYS	C	Methanethiol
Polar	Threonine	THR	T	Ethanol
Polar	Histidine	HIS	H	Methylimidazole
Polar	Lysine	LYS	K	n-Butylamine
Polar	Arginine	ARG	R	n-Propylguanidine
Polar	Aspartic acid	ASP	D	Acetic Acid
Polar	Glutamic acid	GLU	E	Propionic Acid
-	Proline	PRO	P	-

TABLE II: Solvation free energies (kJ mol^{-1}) for hydrophobic and polar amino acid side chain analogs in water H_2O .

	This work, 25°C	Ref. ^a	Ref. ^b	Ref. ^c
Hydrophobic				
Methane(Ala)	8.47 ± 0.12	9.20	8.12	8.12
Propane(Val)	6.93 ± 0.50	10.70	8.33	8.33
Butane (Ile)	7.11 ± 1.83	10.70	9.00	9.00
Isobutane (Leu)	7.24 ± 1.34	10.40	9.54	9.55
Methyl-ethylsulfide (Met)	-0.80 ± 1.69	-6.19	-14.52	-6.20
3-methylindole (Trp)	-29.09 ± 2.34	-12.30	-24.60	-24.62
4-methylphenol (Tyr)	-33.79 ± 3.04	-22.40	-25.56	-25.58
Toluene (Phe)	-7.62 ± 1.12	-3.40	-3.18	-3.18
Polar				
Methanol (Ser)	-21.92 ± 0.21	-14.10	-21.17	-21.19
Ethanol (Thr)	-21.41 ± 0.35	-13.70	-20.42	-20.43
Acetamide (Asn)	-41.75 ± 0.95	-18.80	-40.50	-40.53
Propionamide (Gln)	-44.97 ± 1.41	18.70	-38.25	-39.27
Methanethiol (Cys)	-8.70 ± 2.88	5.50	-5.19	-5.28
Methylimidazole (His)	-32.16 ± 1.74	-27.40	-42.97	-43.00
n-butylamine (Lys)	-18.11 ± 1.31	-15.50	—	-39.86
n-propylguanidine (Arg)	-50.05 ± 1.47	-30.10	—	-83.40
Acetic acid (Asp)	-28.98 ± 0.50	-18.20	—	-45.85
Propionic acid (Glu)	-31.55 ± 0.88	-16.20	—	-42.87

^a Villa & Mark (2002), 20°C^b Chang *et al.* (2007), 25°C^c Radzicka & Wolfenden (1988), 20°C

TABLE III: Solvation free energies (kJ mol^{-1}) for hydrophobic and polar amino acid side chain analogs in cyclohexane cC_6H_{12} .

Hydrophobic	This work, 25°C	Ref. ^a	Ref. ^b	Ref. ^c
Methane(Ala)	0.60 ± 0.11	0.8 ± 0.6	1.05	0.54
Propane(Val)	-8.39 ± 0.25	-6.7 ± 0.9	-6.61	-8.58
Butane (Ile)	-11.60 ± 0.89	-13 ± 1.5	-9.91	-11.59
Isobutane (Leu)	-11.23 ± 0.86	-9.8 ± 1.5	-9.16	-11.05
Methyl-ethylsulfide (Met)	-15.77 ± 0.61	-14.4 ± 1.3	-14.52	-16.02
3-methylindole (Trp)	-36.55 ± 2.37	-35.9 ± 2.8	-38.12	-34.35
4-methylphenol (Tyr)	-24.38 ± 0.77	-28.3 ± 1.2	-22.68	-24.98
Toluene (Phe)	-21.27 ± 1.65	-25.2 ± 1.0	-19.71	-15.65
Polar	This work, 25°C	Ref. ^a	Ref. ^b	Ref. ^c
Methanol (Ser)	-4.73 ± 0.13	-3.5 ± 0.9	-3.22	-6.94
Ethanol (Thr)	-7.89 ± 0.20	-7.8 ± 0.8	-6.57	-9.66
Acetamide (Asn)	-12.57 ± 0.41	-14.3 ± 1.0	-13.68	-12.72
Propionamide (Gln)	-15.32 ± 0.63	-19.3 ± 0.9	-16.82	-16.07
Methanethiol (Cys)	-8.55 ± 0.18	-7.9 ± 0.8	-8.95	-10.54
Methylimidazole (His)	-19.34 ± 1.03	-21.1 ± 1.0	-19.04	-23.47
n-butylamine (Lys)	-13.71 ± 0.55	-16.8 ± 1.8	—	-16.61
n-propylguanidine (Arg)	-17.74 ± 2.08	-24 ± 1.8	—	-20.92
Acetic acid (Asp)	-14.22 ± 0.18	-15.6 ± 1.1	—	-9.33
Propionic acid (Glu)	-19.13 ± 0.55	-18.6 ± 1.1	—	-14.35

^a Villa & Mark (2002), 20°C^b Chang et al. (2007), 25°C^c Radzicka & Wolfenden (1988), 20°CTABLE IV: Solvation free energies (kJ mol^{-1}) for hydrophobic and polar amino acid side chain analogs in ethanol EtOH. Note that here and below, results from Damodaran & Song (1986) are those from Nozaki & Tanford (1971) extrapolated at higher temperatures, and are included here for completeness.

Hydrophobic	This work, 25°C	Ref. ^a	Ref. ^b	Ref. ^c
Methane(Ala)	2.79 ± 0.17	5.92	4.75	5.05
Propane(Val)	-4.79 ± 0.71	2.03	0.50	1.13
Butane (Ile)	-6.23 ± 0.76	—	-4.46	-1.33
Isobutane (Leu)	-9.24 ± 0.47	0.01	0.40	3.03
Methyl-ethylsulfide (Met)	-12.52 ± 1.01	—	-12.93	-11.54
3-methylindole (Trp)	-33.43 ± 1.43	-33.77	-40.27	—
4-methylphenol (Tyr)	-43.01 ± 0.95	35.20	-36.38	-37.21
Toluene (Phe)	-18.07 ± 0.68	-13.98	-15.15	-14.19
Polar	This work, 25°C	Ref. ^a	Ref. ^b	Ref. ^c
Methanol (Ser)	-23.56 ± 0.37	-19.54	-20.67	-20.97
Ethanol (Thr)	-25.56 ± 0.20	—	-19.97	-21.94
Acetamide (Asn)	-44.24 ± 0.58	—	-39.86	-39.86
Propionamide (Gln)	-48.46 ± 0.42	—	-38.28	-38.28
Methanethiol (Cys)	-15.64 ± 0.41	—	-11.28	—
Methylimidazole (His)	-32.29 ± 0.57	-44.02	-45.45	—
n-butylamine (Lys)	-21.83 ± 1.80	—	-24.70	—
n-propylguanidine (Arg)	-42.24 ± 1.18	—	-48.06	—
Acetic acid (Asp)	-33.72 ± 0.62	-26.87	-29.94	-29.76
Propionic acid (Glu)	-37.71 ± 0.42	-25.51	-29.08	-28.90

^a Nozaki & Tanford (1971), 25.10°C^b Damodaran & Song (1986), 37°C^c Tanford (1962), 20°C

TABLE V: Classic SPT estimates of the Gibbs energy change, ΔG_0 , associated with the creation in water H_2O , cyclohexane cC_6H_{12} and ethanol $EtOH$ of a spherical cavity suitable to host methane, propane, toluene and methanol, at 28° and 1 atm; estimates of the solute-solvent interaction energy, consisting of a van der Waals contribution (assumed to be solvent-independent) and a H-bond contribution. A comparison between the $\Delta G_0 + E_a$ values and the experimental ΔG is shown in the last two columns (no optimization has been performed). For each solute, the first line refers to water H_2O , the second to cyclohexane cC_6H_{12} , and the third to ethanol $EtOH$. Units are ($kJ\ mol^{-1}$).

	ΔG_0	E_a	$\Delta G_0 + E_a$	ΔG
Methane(ALA) $\sigma = 3.70 \text{ \AA}$	22.9	-15.0	7.9	8.3
	16.0	-15.0	1.0	0.8
	17.7	-15.0	2.7	1.6
Propane(VAL) $\sigma = 5.06 \text{ \AA}$	38.7	-31.0	7.7	8.2
	25.7	-31.0	-5.3	-7.6
	29.0	-31.0	-2.0	-5.2
Toluene(PHE) $\sigma = 5.64 \text{ \AA}$	46.7	-50.0	-3.3	-3.7
	30.6	-50.0	-19.4	-18.7
	34.7	-50.0	-15.3	-14.2
Methanol(SER) $\sigma = 3.83 \text{ \AA}$	24.2	-45.0	-20.8	-21.4
	16.8	-22.0	-5.2	-5.3
	18.7	-39.0	-20.3	-21.0

TABLE VI: Enthalpic and entropic contributions to the solvation free energies ($kJ\ mol^{-1}$) for hydrophobic and polar amino acid side chain analogs in water H_2O .

	This work, $25^\circ C$			Baldwin (2014), $25^\circ C$		
	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$
Hydrophobic						
Methane(Ala)	8.47 ± 0.12	-3.14 ± 1.55	11.61 ± 1.55	8.29	-2.61	4.61
Propane(Val)	6.93 ± 0.50	-11.17 ± 9.09	18.10 ± 9.25	8.21(8.21)	-5.02(-4.83)	6.98(6.79)
Butane (Ile)	7.11 ± 1.83	-28.28 ± 5.66	35.39 ± 6.52	8.75	-5.66	7.75
Isobutane (Leu)	7.24 ± 1.34	-9.46 ± 8.56	16.69 ± 8.30	9.71	-5.23	7.55
Methyl-ethylsulfide (Met)	-0.80 ± 1.69	-24.04 ± 14.04	23.23 ± 12.94	--	--	--
3-methylindole (Trp)	-29.09 ± 2.34	-88.50 ± 23.89	59.41 ± 24.22	--	--	--
4-methylphenol (Tyr)	-33.79 ± 3.04	-104.73 ± 13.86	70.93 ± 11.17	--	--	--
Toluene (Phe)	-7.62 ± 1.12	-53.44 ± 7.56	45.83 ± 7.73	--	--	--
Polar						
Methanol (Ser)	-21.92 ± 0.21	-42.73 ± 1.38	20.82 ± 1.43	--	--	--
Ethanol (Thr)	-21.41 ± 0.35	-43.56 ± 5.82	22.14 ± 6.08	--	--	--
Acetamide (Asn)	-41.75 ± 0.95	-69.62 ± 7.73	27.87 ± 7.19	--	--	--
Propionamide (Gln)	-44.97 ± 1.41	-71.56 ± 15.02	26.57 ± 15.64	--	--	--
Methanethiol (Cys)	-8.70 ± 2.88	-25.67 ± 5.21	16.97 ± 5.30	--	--	--
Methylimidazole (His)	-32.16 ± 1.74	-63.02 ± 8.18	30.86 ± 9.45	--	--	--
n-butylamine (Lys)	-18.11 ± 1.31	-47.43 ± 9.80	29.32 ± 8.88	--	--	--
n-propylguanidine (Arg)	-50.05 ± 1.47	-124.34 ± 26.07	74.29 ± 24.94	--	--	--
Acetic acid (Asp)	-28.98 ± 0.50	-52.89 ± 8.53	24.59 ± 8.28	--	--	--
Propionic acid (Glu)	-31.55 ± 0.88	-56.98 ± 16.34	25.43 ± 16.84	--	--	--

TABLE VII: Enthalpic and entropic contributions to the solvation free energies (kJ mol^{-1}) for hydrophobic and polar amino acid side chain analogs in cyclohexane C_6H_{12} .

Hydrophobic	This work, 25°C			Abraham (1979,1982), 25°C		
	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$
Methane(Ala)	0.60 ± 0.11	-3.38 ± 1.37	3.98 ± 1.36	0.8	-0.9	1.7
Propane(Val)	-8.39 ± 0.25	-14.22 ± 4.31	5.84 ± 4.18	-7.6	-13.9	6.29
Butane (Ile)	-11.60 ± 0.89	-21.31 ± 2.25	9.71 ± 2.66	-11.1	---	---
Isobutane (Leu)	-11.23 ± 0.86	-18.88 ± 9.48	7.62 ± 9.12	-9.7	---	---
Methyl-ethylsulfide (Met)	-15.77 ± 0.61	24.59 ± 6.73	-40.37 ± 7.24	---	---	---
3-methylindole (Trp)	-36.55 ± 2.37	14.60 ± 9.40	-51.16 ± 8.67	---	---	---
4-methylphenol (Tyr)	-24.38 ± 0.77	-58.46 ± 19.00	34.08 ± 19.35	---	---	---
Toluene (Phe)	-21.27 ± 1.65	-45.74 ± 12.23	24.47 ± 13.06	---	---	---
Polar	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$
Methanol (Ser)	-4.73 ± 0.13	-11.83 ± 0.74	7.10 ± 0.73	---	---	---
Ethanol (Thr)	-7.89 ± 0.20	-15.00 ± 2.78	7.11 ± 2.86	---	---	---
Acetamide (Asn)	-12.57 ± 0.41	-20.88 ± 2.18	8.30 ± 1.98	---	---	---
Propionamide (Gln)	-15.32 ± 0.63	-25.10 ± 5.02	9.78 ± 4.93	---	---	---
Methanethiol (Cys)	-8.55 ± 0.18	-17.85 ± 1.08	9.30 ± 1.30	---	---	---
Methylimidazole (His)	-19.34 ± 1.03	-30.98 ± 12.85	11.64 ± 13.74	---	---	---
n-butylamine (Lys)	-13.71 ± 0.55	-25.71 ± 9.33	12.00 ± 9.27	---	---	---
n-propylguanidine (Arg)	-17.74 ± 2.08	-54.12 ± 15.35	36.38 ± 14.06	---	---	---
Acetic acid (Asp)	-14.22 ± 0.18	-26.42 ± 3.00	12.20 ± 3.00	---	---	---
Propionic acid (Glu)	-19.13 ± 0.55	-30.74 ± 3.26	11.61 ± 3.59	---	---	---

 TABLE VIII: Enthalpic and entropic contributions to the solvation free energies (kJ mol^{-1}) for hydrophobic and polar amino acid side chain analogs in ethanol EtOH.

Hydrophobic	This work, 25°C			Abraham (1979,1982), 25°C		
	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$
Methane(Ala)	2.79 ± 0.17	-0.43 ± 1.50	3.22 ± 1.43	1.6	-2.1	3.7
Propane(Val)	-4.79 ± 0.71	-16.44 ± 5.15	11.65 ± 5.19	-5.2	-12.4	7.22
Butane (Ile)	-6.23 ± 0.76	-17.25 ± 10.92	11.02 ± 11.10	-8.1	-17.7	9.6
Isobutane (Leu)	-9.24 ± 0.47	-16.86 ± 6.96	7.62 ± 7.16	-6.9	-16.1	9.21
Methyl-ethylsulfide (Met)	-12.52 ± 1.01	-25.82 ± 9.72	13.31 ± 9.65	---	---	---
3-methylindole (Trp)	-33.43 ± 1.43	-56.80 ± 22.95	23.37 ± 23.62	---	---	---
4-methylphenol (Tyr)	-43.01 ± 0.95	-90.35 ± 9.04	47.34 ± 9.08	---	---	---
Toluene (Phe)	-18.07 ± 0.68	-28.56 ± 13.07	10.49 ± 13.08	---	---	---
Polar	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$
Methanol (Ser)	-23.64 ± 0.37	-43.22 ± 3.87	19.62 ± 3.69	---	---	---
Ethanol (Thr)	-25.56 ± 0.20	-47.06 ± 2.70	21.50 ± 2.76	---	---	---
Acetamide (Asn)	-44.24 ± 0.58	-70.71 ± 7.82	26.47 ± 7.78	---	---	---
Propionamide (Gln)	-48.46 ± 0.42	-84.15 ± 16.67	35.69 ± 16.98	---	---	---
Methanethiol (Cys)	-15.64 ± 0.41	-31.89 ± 4.51	16.25 ± 4.67	---	---	---
Methylimidazole (His)	-32.29 ± 0.57	-54.20 ± 8.45	21.91 ± 8.57	---	---	---
n-butylamine (Lys)	-21.83 ± 1.80	-37.84 ± 12.64	16.01 ± 11.80	---	---	---
n-propylguanidine (Arg)	--- \pm ---	--- \pm ---	--- \pm ---	---	---	---
Acetic acid (Asp)	-33.72 ± 0.62	-58.71 ± 5.25	24.99 ± 5.24	---	---	---
Propionic acid (Glu)	-37.70 ± 0.42	-68.60 ± 7.06	30.90 ± 6.95	---	---	---

TABLE IX: Enthalpic and entropic contributions to the transfer free energies (kJ mol^{-1}) from water H_2O to cyclohexane C_6H_{12} for hydrophobic and polar amino acid side chain analogs.

Hydrophobic	This work, 25°C			Wolfenden (2015), 25°C			Abraham (1979,1982), 25°C		
	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$
Methane(Ala)	-7.87 ± 0.23	-0.24 ± 2.92	-7.63 ± 2.91	-12.02	10.68	-22.69	-7.50	10.00	-17.5
Propane(Val)	-15.32 ± 0.75	-3.05 ± 13.40	-12.26 ± 13.44	-23.28	6.20	-29.48	-15.80	7.10	-22.29
Butane (Ile)	-18.71 ± 2.72	6.97 ± 7.91	-25.68 ± 9.18	-24.16	4.31	-28.47	-19.80	--	--
Isobutane (Leu)	-18.47 ± 2.20	-10.42 ± 18.04	-8.03 ± 17.42	-24.16	2.51	-26.67	-19.40	--	--
Methyl-ethylsulfide (Met)	-14.97 ± 2.30	48.63 ± 20.77	-66.60 ± 20.18	-10.89	3.35	-14.24	--	--	--
3-methylindole (Trp)	-7.46 ± 4.71	103.1 ± 33.29	-110.57 ± 32.89	-10.42	-0.54	-9.88	--	--	--
4-methylphenol (Tyr)	9.41 ± 3.81	46.27 ± 32.86	-36.85 ± 30.52	1.76	18.21	-16.45	--	--	--
Toluene (Phe)	-13.65 ± 2.77	7.70 ± 19.79	-21.36 ± 20.79	-15.04	-1.05	-13.98	--	--	--
Polar	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$
Methanol (Ser)	17.19 ± 0.34	30.91 ± 2.12	-13.71 ± 2.16	16.08	26.00	-9.92	--	--	--
Ethanol (Thr)	13.52 ± 0.55	28.56 ± 8.60	-15.03 ± 8.94	10.42	28.14	-17.71	--	--	--
Acetamide (Asn)	29.18 ± 1.36	48.74 ± 9.91	-19.57 ± 9.17	27.80	29.89	-2.14	--	--	--
Propionamide (Gln)	29.65 ± 2.04	46.46 ± 20.04	-16.79 ± 20.54	23.19	36.80	-13.61	--	--	--
Methanethiol (Cys)	0.15 ± 3.06	7.72 ± 6.29	-7.67 ± 6.6	-8.71	13.06	-21.77	--	--	--
Methylimidazole (His)	12.82 ± 2.77	32.04 ± 21.03	-19.22 ± 23.19	19.89	48.53	-28.64	--	--	--
n-butylamine (Lys)	4.40 ± 1.86	21.72 ± 19.13	-17.32 ± 18.15	1.55	22.23	-20.68	--	--	--
n-propylguanidine (Arg)	32.31 ± 3.55	70.22 ± 41.42	-37.91 ± 39.00	24.62	57.53	-32.91	--	--	--
Acetic acid (Asp)	14.76 ± 0.68	26.47 ± 11.53	-12.39 ± 11.28	18.71	34.92	-16.16	--	--	--
Propionic acid (Glu)	12.42 ± 1.61	26.24 ± 19.59	-13.82 ± 20.41	12.85	43.43	30.48	--	--	--

TABLE X: Enthalpic and entropic contributions to the transfer free energies (kJ mol^{-1}) from water H_2O to ethanol EtOH for hydrophobic and polar amino acid side chain analogs.

Hydrophobic	This work, 25°C			Abraham (1979,1982), 25°C		
	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$
Methane(Ala)	-5.68 ± 0.29	2.71 ± 3.05	-8.39 ± 2.98	-6.70	8.8	-15.50
Propane(Val)	-11.72 ± 1.21	-5.27 ± 14.24	-6.45 ± 14.44	-13.40	8.6	-21.97
Butane (Ile)	-13.34 ± 2.59	11.03 ± 16.58	-24.37 ± 17.62	-16.80	5.9	-22.69
Isobutane (Leu)	-16.48 ± 1.81	-7.40 ± 15.52	-9.07 ± 15.46	-16.60	5.8	-22.39
Methyl-ethylsulfide (Met)	-11.72 ± 2.70	-1.78 ± 23.76	-9.92 ± 22.59	--	--	--
3-methylindole (Trp)	-4.34 ± 3.77	31.70 ± 46.84	-36.04 ± 47.84	--	--	--
4-methylphenol (Tyr)	-9.22 ± 3.99	14.38 ± 22.90	-23.59 ± 20.25	--	--	--
Toluene (Phe)	-10.45 ± 1.80	24.88 ± 20.63	-35.34 ± 20.81	--	--	--
Polar	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$-\text{T}\Delta\Delta S$
Methanol (Ser)	-1.75 ± 0.58	-0.31 ± 5.25	-1.46 ± 5.12	--	--	--
Ethanol (Thr)	-3.87 ± 0.55	-3.47 ± 8.52	-0.39 ± 8.84	--	--	--
Acetamide (Asn)	-2.60 ± 1.53	-1.43 ± 15.55	-1.17 ± 14.96	--	--	--
Propionamide (Gln)	-3.17 ± 1.83	-20.90 ± 31.69	17.75 ± 35.62	--	--	--
Methanethiol (Cys)	-6.95 ± 3.29	-4.96 ± 9.72	-1.99 ± 9.97	--	--	--
Methylimidazole (His)	-0.05 ± 2.31	8.24 ± 16.63	-8.26 ± 18.02	--	--	--
n-butylamine (Lys)	-3.75 ± 3.11	9.72 ± 22.44	-13.47 ± 20.68	--	--	--
n-propylguanidine (Arg)	-- \pm --	-- \pm --	-- \pm --	--	--	--
Acetic acid (Asp)	-4.74 ± 1.12	-5.93 ± 13.78	0.51 ± 13.58	--	--	--
Propionic acid (Glu)	-6.16 ± 1.30	-10.47 ± 23.40	4.31 ± 23.79	--	--	--

TABLE XI: Fitting coefficients for different amino acid side chain analogs used to compute the thermodynamics parameters along with their Pearson's correlation coefficient R^2 .

	cC ₆ H ₁₂			EtOH			H ₂ O		
	a	b	R^2	a	b	R^2	a	b	R^2
Hydrophobic									
Methane(Ala)	-29.3694	0.5972	-0.0872 0.95818	10.4955	-0.2346	0.0366 0.99678	-64.9942	1.4308	-0.2079 0.99676
Propane(Val)	8.9816	-0.5017	0.0778 0.96322	-7.3071	-0.1660	0.0306 0.97862	-111.8008	2.3413	-0.3409 0.90212
Butane (Ile)	-93.0027	1.6432	-0.2405 0.97181	-104.5813	1.9986	-0.2929 0.92449	-577.1131	12.5018	-1.8502 0.95994
Isobutane (Leu)	74.7048	-2.0768	0.3139 0.94533	252.7160	-6.0302	0.9042 0.87209	-137.4634	2.9061	-0.4250 0.88111
Methyl-ethylsulfide (Met)	1117.2886	-24.6814	3.6649 0.92475	-182.5088	3.5643	-0.5255 0.98937	139.2361	-3.5924	0.5486 0.82625
3-methylindole (Trp)	2288.4173	-51.2502	7.6264 0.95952	-93.8153	0.9098	-0.1241 0.92019	-185.8273	2.3960	-0.3275 0.93339
4-methylphenol (Tyr)	-339.6722	6.4313	-0.9432 0.93764	-55.0555	-0.6341	0.1184 0.95091	307.8364	-9.0299	1.3837 0.94686
Toluene (Phe)	-52.8578	0.2420	-0.0239 0.93993	-243.1799	4.8563	-0.7198 0.80403	-35.5509	0.0252	0.0130 0.90851
Polar									
Methanol (Ser)	-27.2782	0.3709	-0.0518 0.96906	-11.5717	-0.6459	0.1063 0.99971	-35.3124	-0.0971	0.0249 0.99909
Ethanol (Thr)	-15.8606	0.0431	-0.0029 0.97949	14.7211	-1.3157	0.2072 0.98123	-110.9791	1.5442	-0.2182 0.98312
Acetamide (Asn)	55.8768	-1.6963	0.2574 0.94491	89.5388	-3.5109	0.5374 0.97194	8.7707	-1.6703	0.2635 0.98706
Propionamide (Gln)	50.8798	-1.6739	0.2548 0.93166	-39.9522	-0.8732	0.1482 0.95726	-105.0984	0.8572	-0.1151 0.92632
Methanethiol (Cys)	-52.2218	0.8032	-0.1153 0.95474	-68.3014	0.8724	-0.1221 0.98555	-116.9340	2.1153	-0.3076 0.98041
Methylimidazole (His)	-64.2742	0.7870	-0.1117 0.98703	-108.2614	1.2879	-0.1813 0.90130	-517.3281	10.3350	-1.5278 0.83670
n-butylamine (Lys)	-13.7771	-0.2279	0.0400 0.93944	-221.8295	4.1868	-0.6171 0.94929	-206.4630	3.6800	-0.5349 0.91925
n-propylguanidine (Arg)	-941.5227	20.1262	-2.9884 0.90171	--	--	--	526.5740	-14.5680	2.2182 0.84896
Acetic acid (Asp)	104.4444	1.79371	-0.26170 0.98557	-129.4953	1.6740	-0.2374 0.99669	-90.2838	0.9694	-0.1338 0.98984
Propionic acid (Glu)	34.5187	-1.4269	0.2189 0.94840	47.2406	-2.4766	0.3847 0.99048	-216.3530	3.6654	-0.5345 0.92398

Appendix C: Supplementary Figures

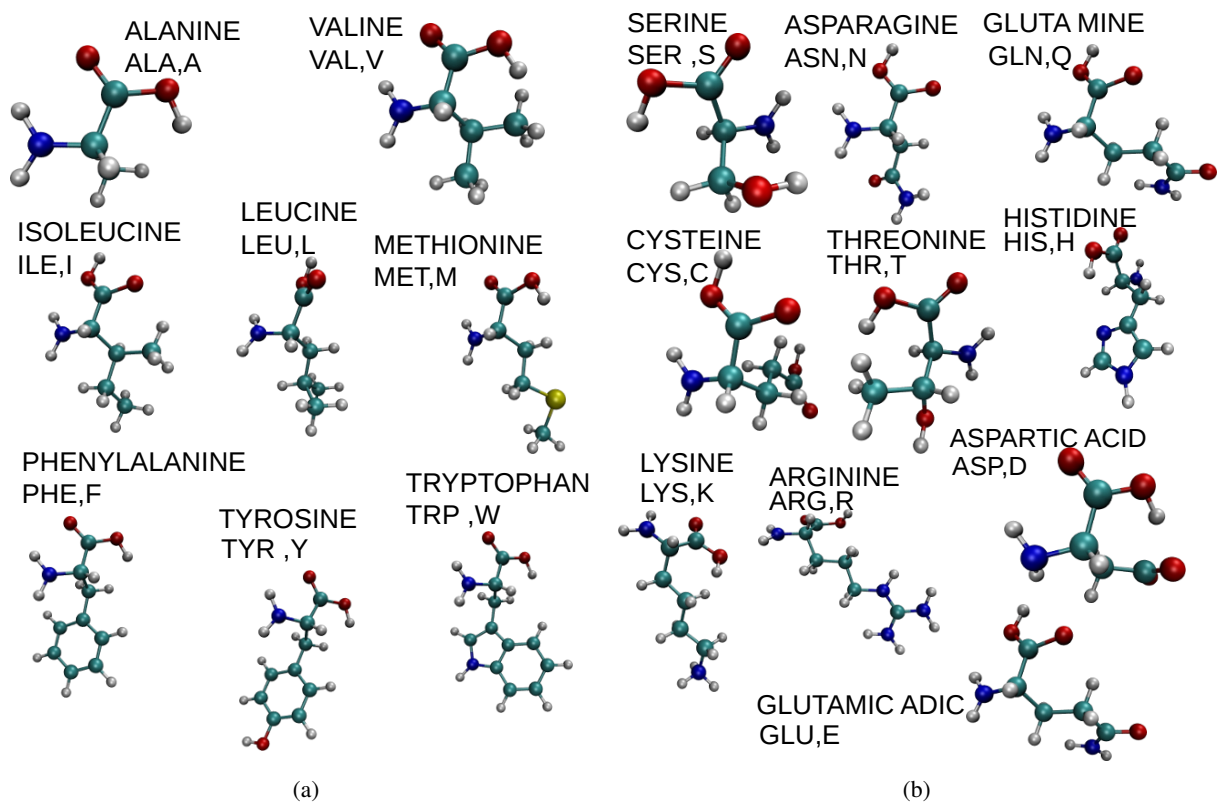


FIG. 6: (a) Hydrophobic amino acids. (b) Polar amino acids.

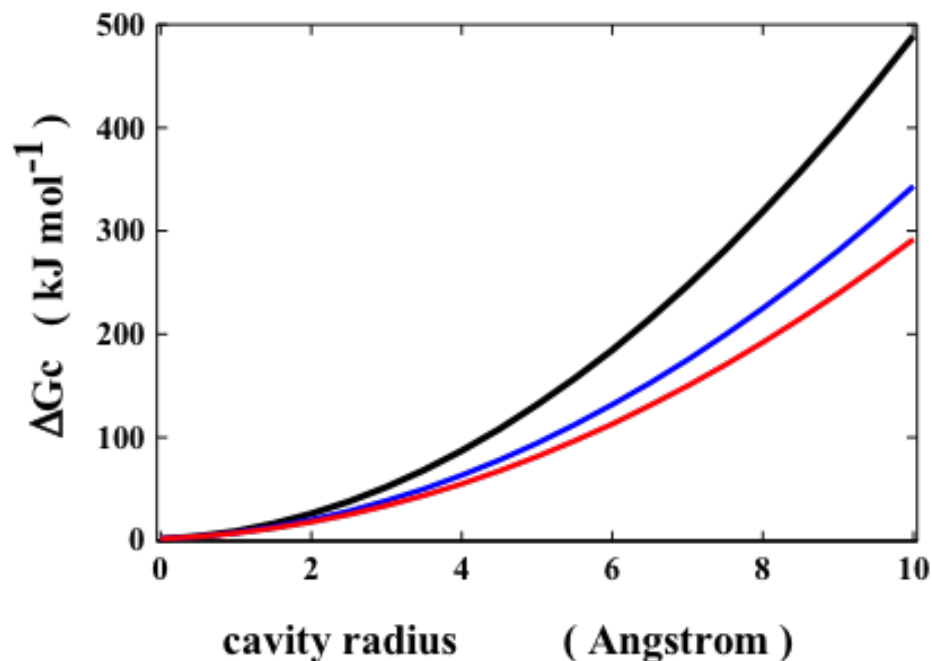


FIG. 7: Trend of the free energy of cavity creation in the three liquids versus the radius of the spherical cavity, calculated by means of classic SPT at 28°C and 1 atm; black line refers to water H_2O , blue line refers to ethanol EtOH , and red line refers to cyclohexane cC_6H_{12} .

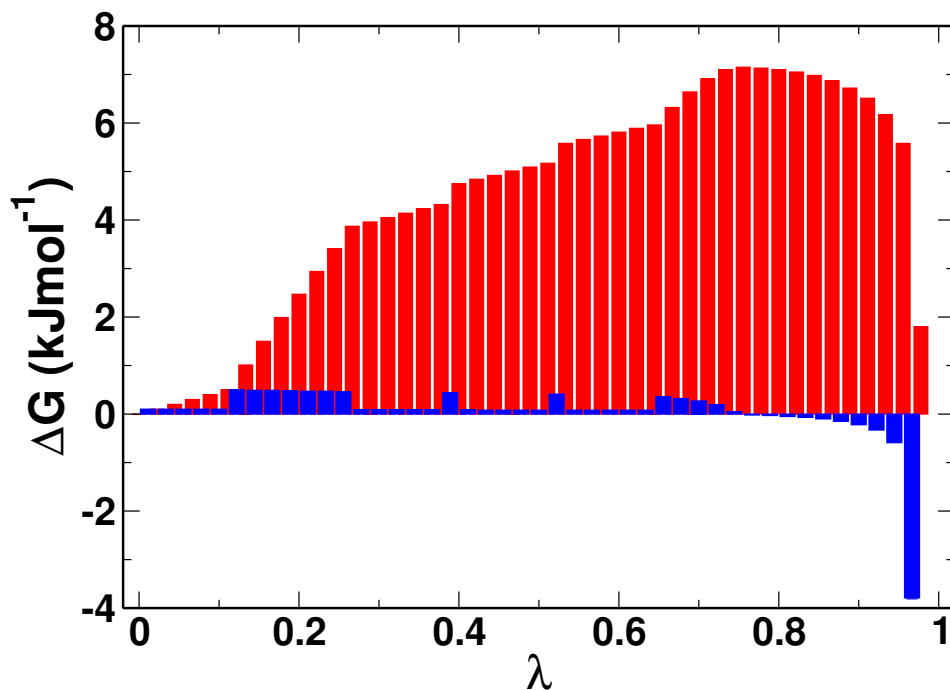
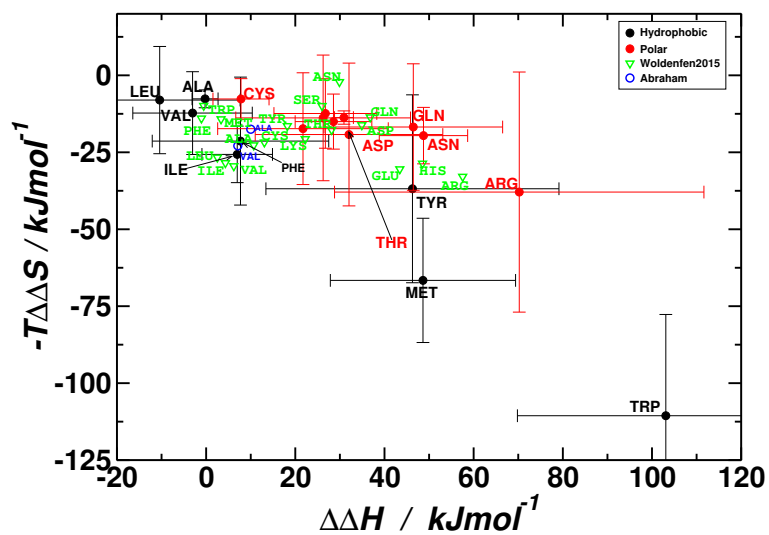
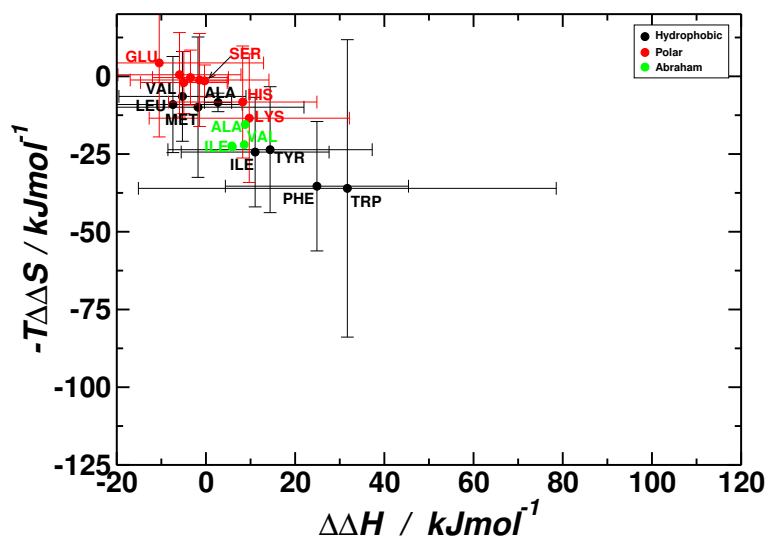


FIG. 8: Illustrative case of the decoupling process for Methanol, SER in cyclohexane, cC_6H_{12} . **Blue** histograms show the free energy difference between two consecutive lambda points while **red** ones display the integral i.e. the cumulative free energy change as a function of lambda. While throughout the work 21 lambda points were used, in this particular case the plot is displayed for 45 lambda points.



(a)



(b)

FIG. 9: Change in the entropic term $-T\Delta\Delta S$ as a function of the change in the entropic part $\Delta\Delta H$ in the case of (a) water to cyclohexane; (b) water to ethanol. In the case of water to cyclohexane, results from Woldenfien *et al* are also included. Units are in kJmol^{-1} .

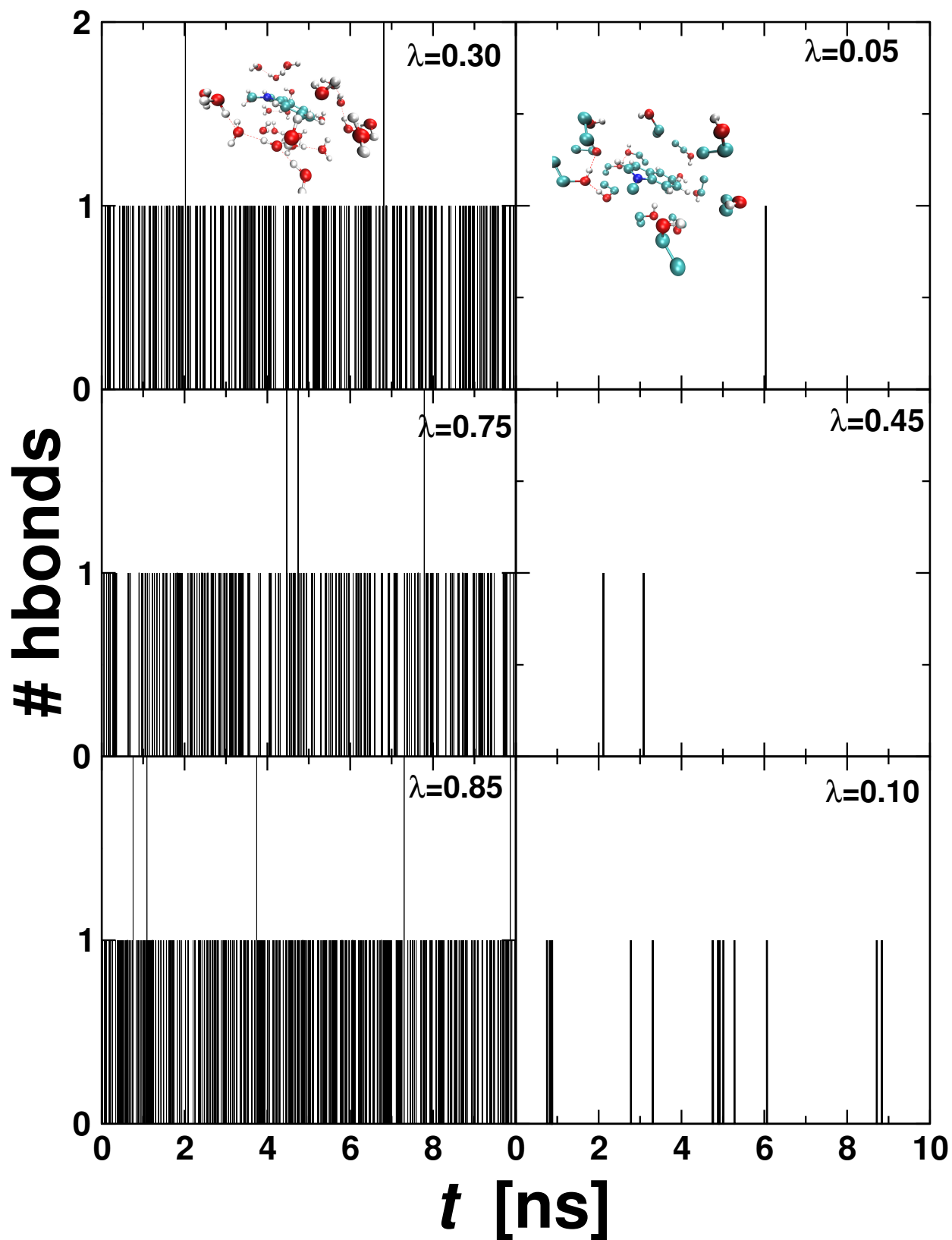


FIG. 10: Time-based number of hydrogen bonds change for 3-methylindole in water, H_2O (**left panel**) and in ethanol, EtOH (**right panel**) at three different temperatures 280, 290 and 300 K upon moving from the **top** to the **bottom**, respectively. Insets are representative snapshots. Hydrogen bonds are computed using *gmx hbond* tool of Gromacs package implying that both faces of the phenyl rings are potentially involved in the geometric consideration for Hbond existence.

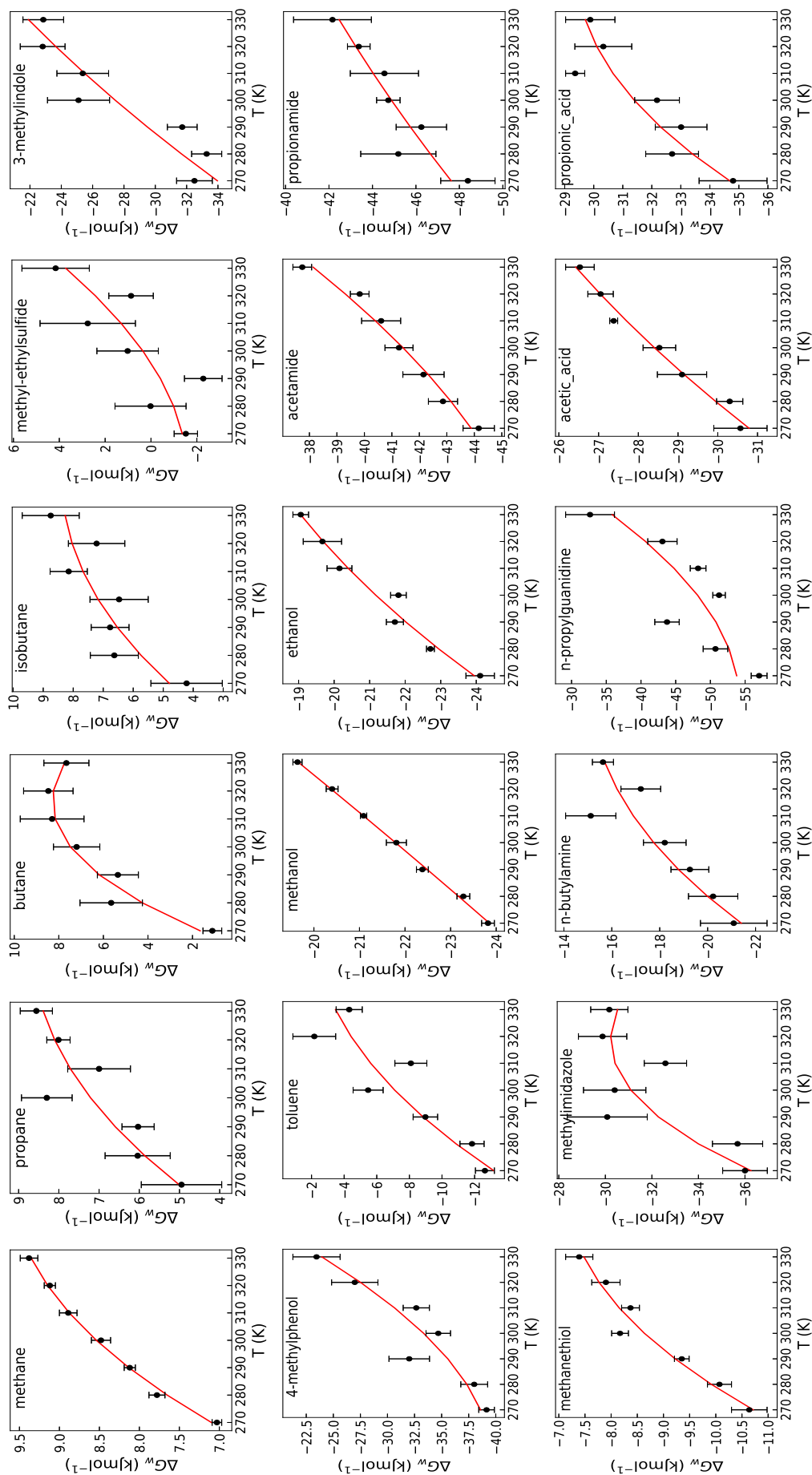


FIG. 11: Temperature dependence of the solvation free energy from gas to water H_2O , ΔG_w .

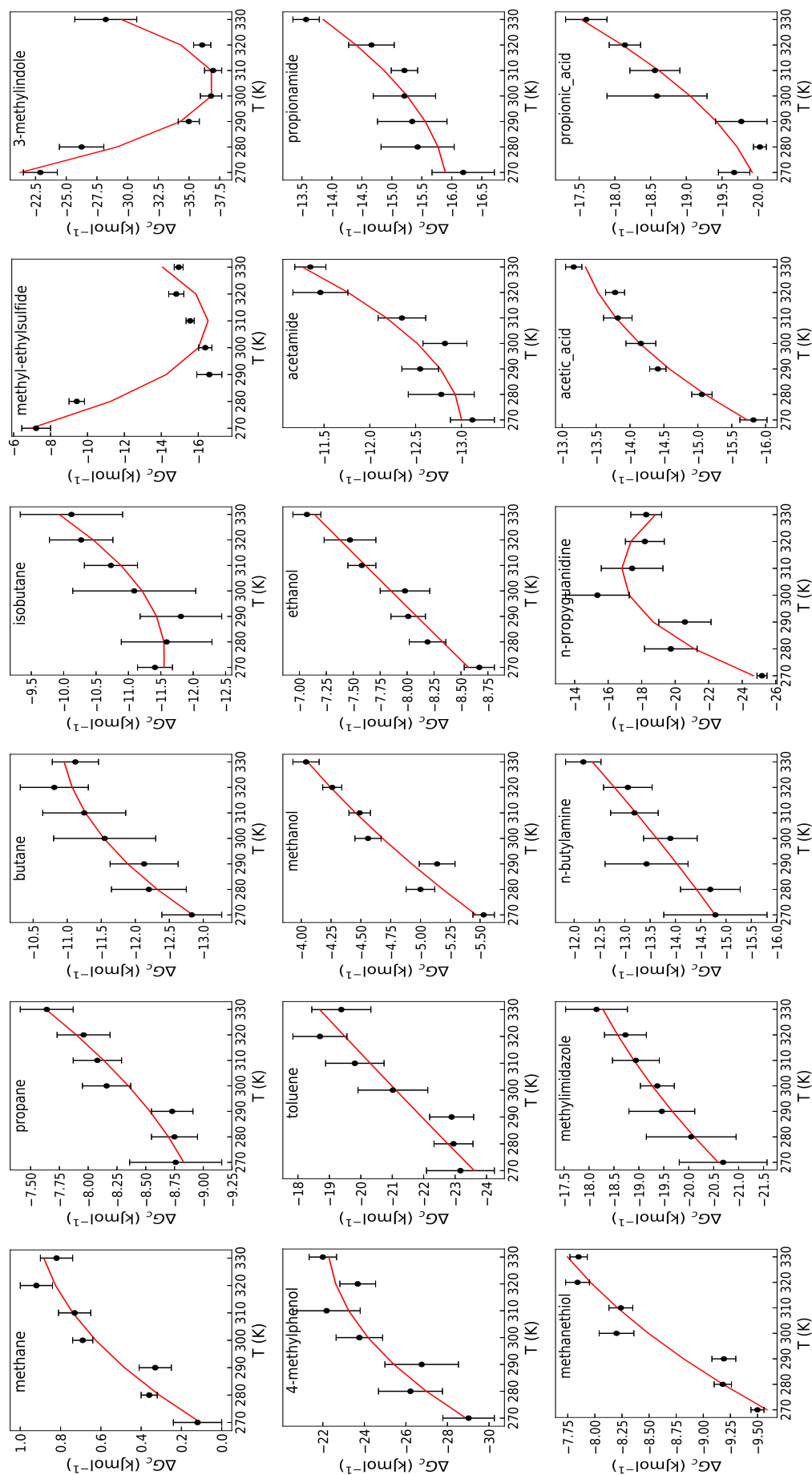


FIG. 12: Temperature dependence of the solvation free energy from gas to cyclohexane cC_6H_{12} . ΔG_c

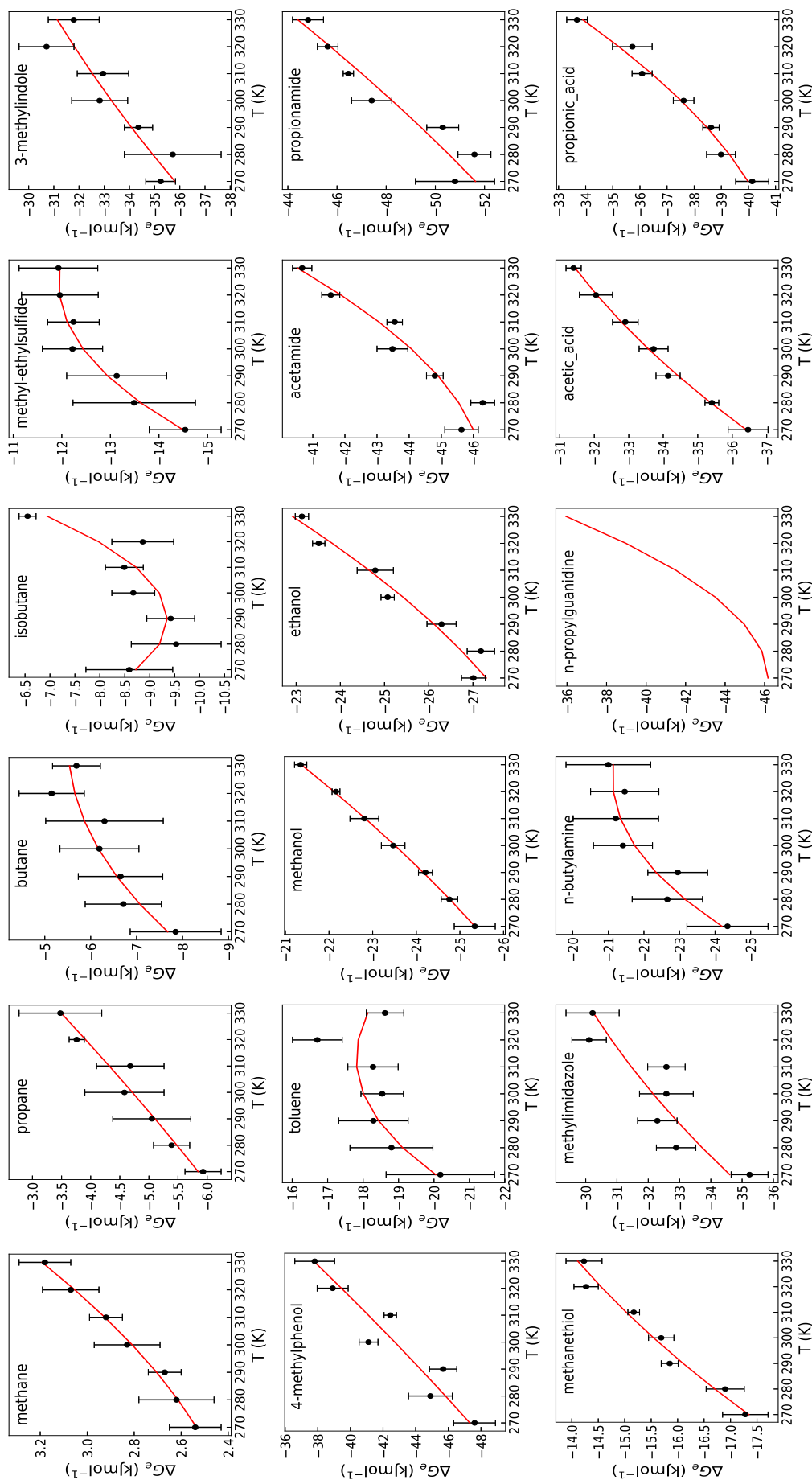


FIG. 13: Temperature dependence of the solvation free energy from gas to ethanol EtOH, ΔG_e . It should be noted that large numerical fluctuations prevent the data collection in the case of n-propylguanidine (ARG). Therefore the corresponding plot should be viewed as an *empirical*-like fitting which doesn't enable reliable extraction of thermodynamics constants.