PCCP

Physical Chemistry Chemical Physics

Volume 19
Number 1
7 January 2017
Pages 1-896
PCCP
Physical Chemistry Chemical Physics
rsc.li/pccp
ISSN 1463-9076
ROYAL SOCIETY OF CHEMISTRY
PAPER
H.-P. Loock et al.
Determination of the thermal, oxidative and photochemical degradation rates of scintillator liquid by fluorescence EEM spectroscopy

View Article Online
DOI: 10.1039/D2CP05214H

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Solvent quality and solvent polarity in polypeptides[†]

Cedrix J. Dongmo Foumthuim, [a*] and Achille Giacometti [b*]

Using molecular dynamics and thermodynamic integration, we report on the solvation process in water and in cyclohexane of seven polypeptides (GLY, ALA, ILE, ASN, LYS, ARG, GLU). The polypeptides are selected to cover the full hydrophobic scale while varying their chain length from tri- to undeca-homopeptides provide indications on possible non-additivity effects as well as the role of the peptide backbone in the overall stability of the polypeptides. The use of different solvents and different polypeptides allows us to investigate the relation between *solvent quality* – the capacity of a given solvent to fold a given biopolymer often described on a scale ranging from "good" to "poor", and *solvent polarity* – related to the specific interactions of any solvent with respect to a reference solvent. Undeca-glycine is found to be the only polypeptides to have a proper stable collapse in water (polar solvent), with the other hydrophobic polypetides displaying in water repeated folding and unfolding events and with polar polypeptides presenting a even more complex behavior. By contrast, all polypeptides but none are found to keep an extended conformation in cyclohexane, irrespective of their polarity. All considered polypeptides are also found to have a favorable solvation free energy independently of the solvent polarity and their intrinsic hydrophobicity, clearly highlighting the prominent stabilizing role of the peptide backbone, with the solvation process largely enthalpically dominated in polar polypeptides and partially entropically driven for hydrophobic polypeptides. Our study thus reveals the complexity of the solvation process of polypeptides defying the common view " like dissolves like", with the solute polarity playing the most prominent role. The absence of a mirror symmetry upon the inversion of polarities of both the solvent and the polypeptides is confirmed.

1 Introduction

In polymer physics[1–4] the term *poor* solvent indicates that a synthetic polymer tends to collapse into a compact conformation because the effective intra-chain interactions occurring between different monomers composing the polymer overcome the monomer-solvent interactions. In the opposite limit of *good* solvent, the polymer tends to remain into an extended conformation. This effect is pictorially represented in Fig. 1a in a plot of the free energy F/k_BT, in units of the thermal energy k_BT, as a function of the mean radius of gyration R_g. In the case of poor solvent the polymer lowers its free energy by folding into a compact conformation, thus reducing R_g, whereas in the second case the free energy decreases but R_g remains large because the polymer is solvophobic. The distinction between good and bad solvent can be made more quantitative using familiar scaling arguments from polymer physics where $R_g \sim N^\nu$ where $\nu \approx 3/5$ for extended/swollen conformation and $\nu \approx 1/3$ for compact/globule conformation[1,3–6]. While this picture is very simple and handy, it clearly disregards the fact that it depends on the specific properties of the polymer as well as of the solvent. Hence *solvent quality* is used to identify the relative character of one solvent with respect to a reference one in terms of the above picture. Thus one solvent can be a good solvent for one polymer and bad for another one, and this point becomes extremely important in the framework of biopolymers and biomolecules[7].

The conformational freedom of biomolecules in general, and of proteins in particular, enables them to inter-convert between several states in solution, thereby adapting upon changing the solvent environments, for e.g. by changing from a polar to a non-polar solvent. The same flexibility allows them to perform various functions *in vivo*. However, even though water is undoubtedly the most-like biological milieu, the stability of these latter is not necessary compromised in non-polar solvents[8,9]. A protein can be regarded as a chain formed by a sequence of amino acids taken from a 20 alphabet letters, half of which have hydrophobic (H) character, so they tend to avoid contact with water, whereas the other half are polar (P) so they are happy to stay in contact

[a] Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia,Campus Scientifico, Edificio Alfa, via Torino 155,30172 Venezia Mestre, Italy. E-mail: cedrix.dongmo@unive.it

[b] Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia,Campus Scientifico, Edificio Alfa, via Torino 155,30172 Venezia Mestre, Italy and European Centre for Living Technology (ECLT) Ca' Bottacin, Dorsoduro 3911, Calle Crosera 30123 Venice, Italy. E-mail: achille.giacometti@unive.it

[†] Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

View Article Online
DOI: 10.1039/D2CP05214H

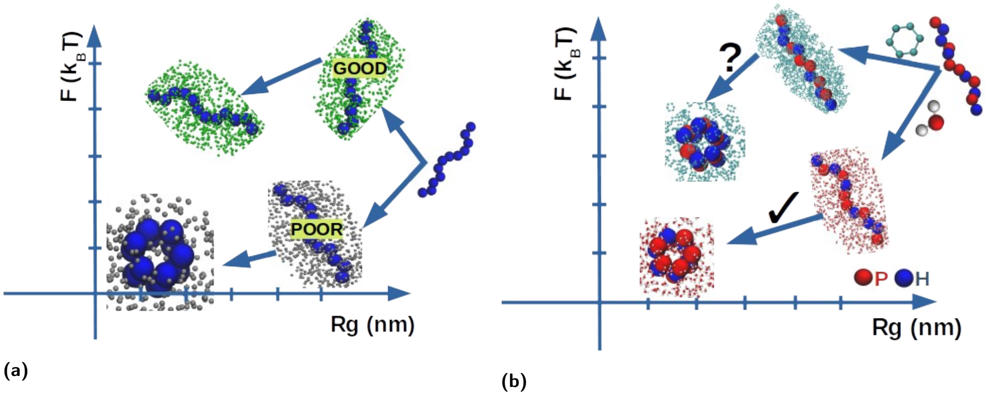


Fig. 1 Cartoon description of the solvophobic effects in different environments in the plane free energy F (units of thermal energy $k_B T$) $F/k_B T$ with respect to gyration radius R_g of the polymer. Panel (a) is for a synthetic homopolymer which collapses into a globule in a "poor" solvent and remains extended in a "good" solvent. Panel (b) displays the question of whether a heteropolymer formed by hydrophobic (H) and polar (P) monomers assumed to be collapsing in water H_2O into a unique fold with preferential exposition of the polar residues P, does collapse in a non-polar solvent such as cyclohexane cC_6H_{12} by reversing inside out its fold with H residues exposed to the solvent and P residues buried inside for the fold.

with water. Proteins in water fold reproducibly and reliably to achieve their unique native states driven by several concurring interactions, including the tendency to avoid contact with water, denoted as the *hydrophobic effect*, as indicated in Fig.1b. Note that *solvent polarity* in fact refers to the polar character of a specific solvent as compared to water that is taken as a reference scale for an optimal polar solvent, and this is clearly different from the definition of *solvent quality* defined earlier, albeit the two definitions are often interpreted as meaning the same thing. However, the presence of the hydrophobic residues might suggest a similar folding event occurring also in non-aqueous mileu, such as for instance an organic solvent. In this case, it might happen that the " protein would turn inside out with its hydrophilic or polar residues inside and hydrophobic apolar residues outside", as suggested by Peter Wolynes sometime ago[8], and pictorially represented in Fig.1b. To the best of our knowledge, no record of such event exists in literature. In a conventional surfactants framework oil form droplets in water and water form droplets in oil. However, it has been recently shown[10] that this " mirror symmetry" is not respected by using "unconventional" surfactants with hydrophobic head and polar tail that do not form micelles in apolar solvent in the same way as conventional surfactants do in polar solvents such as water. Hence there is no " mirror symmetry" in this more complex case, and the same appears to be true in proteins[11,12]. Likely, this is because this argument overlooks the character of the peptide bond, a feature that might turn the delicate balance provided by the amino acid properties[7,13]. In addition, the actual length and energy scales are different in the two cases: in water, the enthalpy gain in saturating hydrogen bonds as well as the entropy increase stemming from the additional free water molecules, have no counterpart in organic solvents where the van der Waals interactions are much weaker and the entropic gain significantly reduced[10–12]. A confirmation of this picture is the aim of the present study.

For a fully solvated analyte, the solvation free energy can be used as a good indicator of the overall stability of the studied system, in relation to the solvent considered, and we have already carried out a detailed analysis of the solvation free energy of each single amino acid side chain equivalent both in water H_2O and in cyclohexane cC_6H_{12}, as paradigmatic representative of an organic, apolar solvent[14]. It was found that the transfer free energy from water H_2O to cyclohexane cC_6H_{12}, that is the work necessary to bring one single amino acid side chain from one solvent to the other, was respecting the expected hydrophobic scale of the amino acids. Hence, hydrophobic amino acid side chains have decreasing free energy transfer, whereas polar amino acid side chains have increasing free energy, in agreement with experimental findings[15]. In this analysis, however, the backbone part of each amino acid was removed and replaced by a single hydrogen atom – obtaining what is hereafter referred to as side chain amino acid equivalents, thus hindering the effect of the backbone part that it was already argued to play an important role[16]. As experimentally the solubility of polypeptides in water H_2O decreases as the length increases[13], this dependence should also been taken into account. Both aspects will be then considered in the present study.

Polyglycine peptides (GLY_n), formed by n identical repeated residues, are a common model for the peptide units. Other polypeptides can be formed in the same way by using amino acids with different polarities as for instance those reported in Table 1. The interest in understanding the n dependence of the solvation free energy is twofold. On one hand, it constitutes one of the key ingredients of the forces stabilizing protein folding[7]. On the other hand, the solvation process is known to be significantly different above and below a critical size (of order of 1 nm), at least in water[17]. For both these reasons, there were several studies in recent literature reporting several useful results.

Tomar *et al.*[19] addressed the paradoxical difference between theory and experiments on the group-additivity of the solvation free energy in an osmolyte solution (water plus small organic cosolutes), and emphasized the importance of evaluating the transfer free energy from one solution to another.

Using calorimetric measurements of the solvation enthalpies of some dipeptide analogs, Avbelj and Baldwin[20] have suggested

View Article Online
DOI: 10.1039/D2CP05214H

Character	Amino acid	Short name	Single letter
Hydrophobic	Glycine	GLY	G
Hydrophobic	Alanine	ALA	A
Hydrophobic	Isoleucine	ILE	I
Polar	Asparagine	ASN	N
Polar	Lysine	LYS	K
Polar	Arginine	ARG	R
Polar	Glutamic acid	GLU	E

Table 1 The correspondence between the seven amino acids with their tri- and uni-code nomenclature used in this work to build the homopeptides. See e.g. Ref.[18].

that the principle of group additivity does not hold true for the interaction of the peptide group with water H_2O. According to their results, the main reason of this breakdown is the strong electrostatic interactions between neighbouring NHCO units of peptides in water H_2O.

In 2013, Kokubo et *al.*[21] analysed the effect of flexibility on the solvation free energies of alanine peptides in water H_2O. They found a linear dependence with respect to the peptide length n for both electrostatics, van der Waals cavity-formation, and total solvation free energies.

In an attempt to provide a general view on the additivity character of the solvation free energy, Staritzbichler and collaborators[22] used multiconfiguration thermodynamic integration, along with generalized-born surface area solvation model to compute the solvation free energy of different polypeptides in the form of rigid helices of various length n in water H_2O and in chloroform $CHCl_3$. They preferentially considered uncharged amino acids while tuning their backbones to fit an ideal helix conformation. Their results suggest a nonlinearity in the solvation free energy in the case of short ($n \leq 5$) peptide chains, turning to linear for longer chains.

Hajari and van der Vegt[16] performed a molecular simulation study on the temperature dependence of solvation free energy of both polar and hydrophobic tripeptides in water H_2O. They found a significant deviation from linearity in the case of hydrophobic polypeptides and a nearly linear dependence for polar polypeptides. This latter result was ascribed to a nearly perfect enthalpy-entropy compensation, leading the overall solvation free energy nearly unaltered by the peptide backbone. Contrariwise, no such compensation was found for hydrophobic tripeptides.

In their work, Konig et *al.*[23] addressed the extent to which the assumption of group additivity to the absolute solvation free energy can hold valid. In doing so, they made use of molecular dynamics–based free energy simulations to estimate the absolute solvation free energies for 15 N-acetyl-methylamide amino acids with neutral side chains. The authors have shown that values of solvation free energies of full amino acids based on group-additive approaches are systematically too negative while completely overestimating the hydrophobicity of glycine.

A work from Montgomery group[24] explored the solvation free energy of polyglycines of different length *n*, in pure water H_2O and in the osmolyte solutions, 2M urea and 2M trimethylamine N-oxide (TMAO). The solvation free energies were found linearly dependending on n and they identified the dependence on the specific interactions (van der Waals, electrostatics, etc).

While all these studies prove to be rather useful, a coherent picture of the solvation process is still lacking. Motivated by this, in the present work we first analyze the poor/good paradigm of water H_2O and cyclohexane cC_6H_{12} on polypeptides of different length n and different polarities (hydrophobic and polar), and then compute the corresponding solvation free energies, disentangling the enthalpic and entropic contributions.

The remaining of the paper is organized as follows. In Section 2 we describe the underlying theory and the simulation methods used in this study. Section 3 then includes all results and Section 4 a summary of the results along with a discussion. Supplementary Information includes additional figures and tables relative to the results reported in the main text.

2 Theory and Methods

2.1 Thermodynamic integration

The solvation free energy ΔG_{sol} 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}

$$\Delta G_{sol} = G_{solvent} - G_{vacuum} \tag{1}$$

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

$$\Delta\Delta G(S_1 \to S_2) = \Delta G_{S_2} - \Delta G_{S_1} \tag{2}$$

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 thermodynamic integration[25]

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

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 $\langle \ldots \rangle_\lambda$ 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.

Assuming a constant heat capacity, the temperature dependence of the solvation free energy can be written as

$$\Delta G(T) = a + bT + cT \ln T \tag{4}$$

so that

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

with very little dependence on the choice of the specific functional

View Article Online
DOI: 10.1039/D2CP05214H

form[16]. The enthalpy change can then be obtained from

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

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 can then be evaluated using error block analysis[16].

We remark here that this is neither the unique nor the most efficient way to compute $T\Delta S$ and ΔH. Indeed, Fogolari *et al.*[26,27] and Lai and Oostenbrink[28] looked for different ways to compute entropies and enthalpies directly thus avoiding the use of the phenomenological expression given in Eq.(4). However, this analysis is much more computational demanding and it could not be afforded for the systematic investigation that we are presenting here. We further note that Eq.(4) is known to hold true only in water H_2O within the temperature range $270-330$K consider in the present study[16], and it also appears to work for single amino acid side chain equivalents in cyclohexane cC_6H_{12}[14].

2.2 Numerical protocols

The amino acid building blocks for the polypeptides selected in this work span the full hydrophobic scale ranging from polar uncharged (ASN) to hydrophobic (GLY, ALA, ILE) through charged moieties (LYS, ARG, GLU). Moreover, most of these latter were recently shown to preferentially populate the α-helical conformational space[29], one of the major secondary structural motif found in biopolymers. The initial structures for the polypeptides were prepared using the Avogadro tool (ver 1.2.0)[30] in their extended configurations with the dihedral angles of $(\phi,\psi)=(180°,180°)$ with the N- and C- termini capped with the neutral acetate (ACE) and methylamine (NME), respectively. All the polypeptides were simulated in full atomistic details by employing the GROMOS96 (54a7) force field[31] that appears to be an optimal compromise between precision and computational cost when computing hydration enthalpies as tested against experimental data[14,32,33]. A table summarizing the amino acids used to build the homopeptides, along with their common names, and both their simplified three letters codification with the corresponding uni-letter nomenclature is shown in table 1 above.

It is worth stressing that in this work we have explicitly included charged residues, unlike previous works that avoided this case because of the tremendous effort needed to model them[23] as the charged moieties require complex parameterization for the treatment of finite-range electrostatics interactions[34,35]. This endeavour then represents a significant step forward even at the computational implementation level with respect to previous studies.

The simulations were performed in water H_2O and cyclohexane cC_6H_{12}, as paradigmatic representative of polar and hydrophobic solvents, and five polymers with length from tri- (n=3) to undeca-peptides (n=11) were considered. In all cases they were initially aligned along the z-axis as exemplified in Supplementary Figure SI in a rectangular box and subsequently solvated with the solvent. The box dimensions and the number of solvent molecules used are reported in table 2 below. The simulations were performed with Gromacs simulation package (series 2018, 2020 and 2021)[36] and all the solutes were modelled roughly at their physiological pH. Therefore, GLU was preferentially modelled in its conjugate base i.e. the singly-negative anion glutamate whilst the carboxylic acid of ARG was deprotonated and the amino and guanidino groups protonated leading to a singly-positive acid. Likewise, the carboxylic acid of LYS was deprotonated and both its α-amino and side chain lysyl groups protonated resulting to a monocation. Accordingly, Na^+ and Cl^- counterions were added to preserve the system's electroneutrality and achieve the physiological-like concentration of 0.15 M. As detailed in Section 2, 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 21 binning points. Altogether, the data discussed throughout this study are the result of approximately 10290 individual runs running up to nearly 103 μs, and thus it represents a large scale extensive computational endeavour.

n	3	5	7	9	11
box (nm^3)	3×3×3.5	3×3×4	3×3×4.5	3×3×5	3×3×5.5
H_2O	1007	1157	1251	1393	1517
cC_6H_{12}	181	210	218	241	262

Table 2 Simulation details including the unit box dimensions in nm^3 and the number of solvent molecules used in the case of H_2O and cC_6H_{12} for different polymer length. The table is meant to provide a general overview of the number of solvent molecules as subtle differences may arrive due to the size of the solute upon changing from GLY to ARG towards LYS and ILE.

The simulations described herein follow our previous protocol[14]. However unlike that case of single amino acid side chain equivalents, here the full atomistic polypeptide structures of different length has been considered, and the fully fledged thermodynamics integration has been carried out. Throughout the thermodynamics integration calculations, the polymers were kept restrained in a stretched conformation by applying a force at the two CA end-points of the polymer, as illustrated in Supplementary Figure SI. This maximizes the number of solute-solvent contacts and hence the solvation, thus allowing a direct comparison between them.

Following preliminary equilibration steps in the canonical *NVT* and isobaric-isothermal *NPT* ensembles, most of the thermodynamic integrations were performed with time step of 2×10^{-15}s, although in some cases stability tests suggested the use of time steps as low as 1×10^{-15}s.

In order to assess the enthalpic and entropic single contributions, a set of 7 different temperatures ranging from 270 K to 330 K were performed. In the case of the undeca-polypeptides, an additional set of simulations of various time-scales were performed with the same conditions as above but in this case the polymers were unrestrained, closely following previous protocol[37]. Those more conventional simulations were performed at room temperature 300 K and the conformational freedom of the

View Article Online
DOI: 10.1039/D2CP05214H

homopeptides enables them to explore the available phase space and thus adopting the most favourable conformation with respect to the solvent considered.

Standard probes such as the radius of gyration R_g [1] and the solvent accessible surface are (SASA) [38] were used to provide a quantitative assessment of the peptides behaviours in the considered solvents. It is important to remark that while calculation of SASA in folded state is unambigously defined, corresponding values in the unfolded conformation is not [39].

3 Results

3.1 Good and poor solvents

As a preliminary step, we have performed molecular dynamics simulations of polypeptides formed by 11 identical residues ranging from hydrophobic (GLY, ALA, ILE), to polar (ASN) and charged (LYS, ARG, GLU). In the following, we will denote as ASN11, ALA11, etc. polypeptides formed by 11 identical ASN, ALA, etc. Note that we are denoting them as "polypetides" even if it would not be strictly correct for a number of residues ranging from 3 to 11 as those considered here. We included also GLY as glycine has essentially no side chain (its side chain reduces to an hydrogen atom), and hence it represents a very convenient benchmark to compare with. It has been argued that water H_2O at room temperature is a poor solvent for GLY15 [40] and more generally for a protein backbone [7]. We shall confirm this results here with GLY11. By contrast, we shall see that cyclohexane cC_6H_{12} is a good solvent for the same chain, indicating the presence of preferential interactions between the backbone of GLY11 and cyclohexane molecules. Support to this interpretation stems from present calculation as well as from the linear decrease of the solvation free energy as a function of the number of repeated units, as it will be discussed further below.

We performed molecular dynamics of GLY11 and ALA11 in both water H_2O and cyclohexane cC_6H_{12} at room temperature ($T = 300\,\mathrm{K}$). In all cases the initial condition was taken to be a random swollen conformation. Self-assembly of GLY and ALA oligopeptides in water were previously studied by Pettit and collaborators [21,41] who observed a fast aggregation coherent with our results. Results for the other considered polypeptides can be found in Supplementary Information.

Figure 2 reports the behavior of the three selected probes to the conformational state: the root-mean-square-deviation from the initial state (RMSD) (top panels (a) for water H_2O and (b) for cyclohexane cC_6H_{12}), the radius of gyration R_g (middle panels (c) for water H_2O and (d) for cyclohexane cC_6H_{12}), and the solvent accessible surface area (SASA) (bottom panels (e) for water H_2O and (f) for cyclohexane cC_6H_{12}). The inset highlights the significant drop in all three probes in the case of GLY11 in water (magenta solid line in (A)-(a),(c),(e)) occurring within the first 1.5 nanoseconds from the initial extented conformation, followed by an equilibration around these values. A much more unstable trajectory is followed by ALA11 in water H_2O (orange line in panels (B)-(a),(c),(e)), with repeated folding and unfolding events occurring during the entire trajectory. By contrast, in cyclohexane (panels (A)&(B)-(b),(d),(f)), both GLY11 and ALA11 display a fast settling into an extended conformation essentially equivalent to the initial conformation. Note that a more quantitative assessment on the difference between compact/globule and extended/swollen can be obtained by computing the ν exponent in $R_g \sim n^\nu$ with $\nu \approx 0.6$ in the extended (Flory) regime and $\nu \approx 0.33$ in the compact/globule regime [1,3–6]. However, it should be emphasized that the above scaling is strictly valid in the $n \gg 1$ limit (as it is the case in polymer physics), so its application to small polypeptides as those treated in this paper should be taken with great care. This is indeed shown in Supplementary Figure SII where we find ν unphysically small for all considered peptides irrespective of their polarity.

All in all, the results for GLY11 in water H_2O provide support to past evidence [13,42] that water is a poor solvent for polyglycine, whereas the results for GLY11 in cyclohexane cC_6H_{12} are consistent with the presence of a long-lived metastable state for globular proteins in cyclohexane cC_6H_{12} [43]. The are also in line with the idea [7] that water H_2O is a poor solvent for the protein backbone, and that this is one of the main driving force in the collapse of the chain to a globule-shaped structure, along with solvent entropy gain and the burial of the hydrophobic side chains [42]. This is particularly effective in water H_2O because of its small size (≈ 2.8 Å of diameter) and large number density (55.3 M under standard conditions). Cyclohexane cC_6H_{12} has size more than two times larger than water H_2O and significantly smaller number density, and the solvent entropic gain is reduced accordingly.

The behavior of ALA11 in water, which displays an erratic sequence of folding and unfolding events for which no stable collapse is observed (see Fig. 2), is more surprising. ALA is usually classified as a hydrophobic amino acid (see Table 1), and hence a conformational folding akin to GLY11 may have been expected. However, ALA has a larger side chain that provides a larger steric hindrance that may hamper the collapse of the small peptides such as those considered here. In addition the energetic interactions of the two polypeptides with water is different. By contrast, the behaviour of the GLY11 and ALA11 is nearly identical in cyclohexane cC_6H_{12} with both remaining extended throughout the full trajectory. This can be interpreted as cyclohexane cC_6H_{12} being a good solvent for both, and it might provide one possible reason of the experimentally noted absence of a collapse of proteins in cyclohexane cC_6H_{12}, and more generally in any non-polar solvent [43]. Table 3 summarizes all these results in a synoptic form where water H_2O is referred to as poor+ (i.e. with stable fold) solvent for GLY11 and as poor (no stable collapse) for ALA11. Likewise cyclohexane cC_6H_{12} will be referred to as a good+ solvent for both.

For the remaining 5 considered polypeptides, the results for RMSD (top panel), the radius of gyration R_g (middle panel) and the SASA (bottom panel) for the full trajectory in water H_2O (left) and in cyclohexane cC_6H_{12} (right) are reported in Supplementary Figure SIII, and confirm a rather complex and diverse behaviour. In water H_2O, ILE11 (hydrophobic) displays an initial collapse followed by a fluctuating behaviour about a less compact conformation (black line left panel), whereas for ASN11 (polar, red line left panel) R_g remains mostly stable throughout the full trajectory following an initial drop, but with a final large fluctuation. Inter-

View Article Online
DOI: 10.1039/D2CP05214H

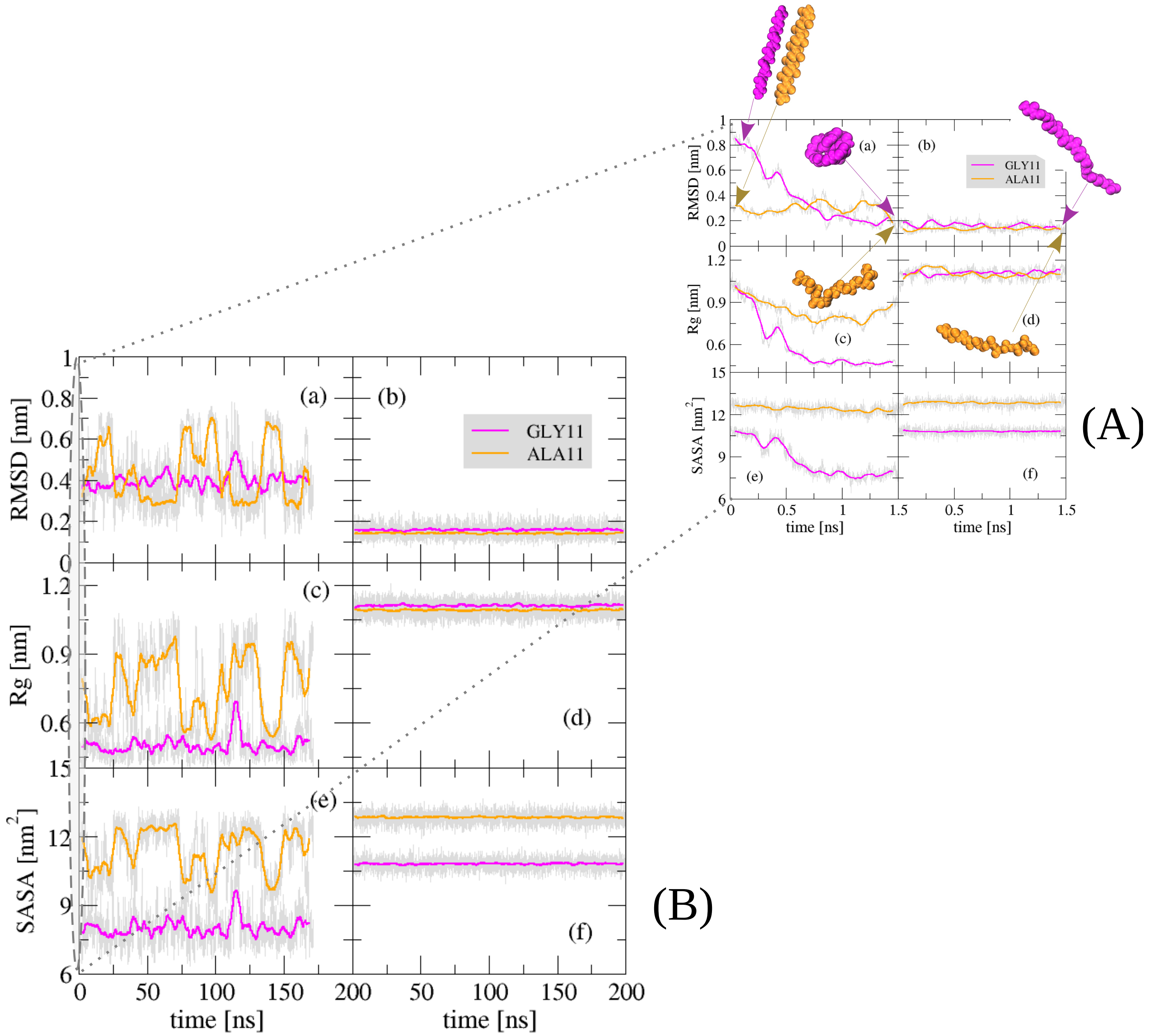


Fig. 2 Initial (inset) and equilibrium probes of the conformational behaviour of GLY11 and ALA11 at the pre-production stages (*NVT* and *NPT* equilibration). Panels (a) and (b): root-mean-square-deviation (RMSD) from the initial state in water H_2O (a) and in cyclohexane cC_6H_{12} (b). Panels (c) and (d): radius of gyration R_g in water H_2O (c) and in cyclohexane cC_6H_{12} (d). Panels (e) and (f): the solvent accessible surface area (SASA) in water H_2O (e) and in cyclohexane cC_6H_{12} (f). In all cases the inset (A) report the few initial nanoseconds of the equilibration process. Results for GLY11 are displayed in magenta and for ALA11 in orange. The insets also report representative snapshots of GLY11 (magenta) and ALA11 (orange) both at the initial and final stages. In all cases, the initial conformation is a random coil.

estingly, in cyclohexane cC_6H_{12} ILE11 remains extended (black line right panel) whereas ASN11 collapses (red line right panel). The other three polypeptides (LYS11, ARG11, and GLU11, polar because charged), display large fluctuations in water H_2O (left panel), and remain rather extended in cyclohexane cC_6H_{12} (right panel). All these findings are summarized in Table 3.

The results of these last three polypeptides (LYS11, ARG11, and GLU11) show a complex behaviour that defies any simple description in terms of poor and good solvent. Of course, this was to be expected: each residue has its own characteristics that go beyond the operative description in term of good and bad solvents, and sometimes this matters for this kind of calculation. For instance, isoleucine ILE is known to be a strong hydrophobic amino acid and the corresponding marked collapse of ILE11 in water H_2O with a noticeable structural rearrangement in the course of the simulation as depicted in the RMSD (top), R_g (middle) and SASA (bottom) plots in Supplementary Figure SIII, agrees with this picture. Also the corresponding absence of any collapse or structural

View Article Online
DOI: 10.1039/D2CP05214H

rearrangement of ILE11 in cyclohexane cC_6H_{12}, could be ascribed to the stabilizing effect of cyclohexane cC_6H_{12} in line with its hydrophobic character. However, the negatively charged GLU11 polypeptide in water H_2O, adopts a U-like shape after a long equilibration, and subsequently collapses to a globule although with less compact shape. We surmise that the length and shape of the side chain arms are major factors prohibiting the proper collapse of GLU11 in water H_2O. In cyclohexane cC_6H_{12}, after a short equilibration time a relatively steady and stable conformation is achieved, compatible with a favourable solute-solvent interactions over the solvent entropy promoting the collapse.

Comparatively, ASN11 and ARG11 behave symmetrically, with water H_2O acting as a good solvent whereas cyclohexane cC_6H_{12} as a poor one. Indeed, ASN11 in H_2O seems to remain marginally extended and undergo a number of noticeable conformational fluctuations as reported by the minor changes seen in its solvent accessible surface area plot and the root mean square deviation analysis, respectively. Transiently formed globular-like conformations are identified in the trajectory signalled by the significant decrease in the radius of gyration R_g reported.

In cyclohexane cC_6H_{12}, after a short equilibration period corresponding to the coil-to-globule adaptation, all RMSD, R_g and SASA level off and remain steady flat throughout the simulation timescale, implying undoubtedly a favourable and stable ASN11 - cC_6H_{12} interactions. Furthermore, we monitored an increase in the number of intramolecular hydrogen bonds (see also further below) in ASN11 as shown in figure Supplementary Figure SIV, a sign of an increased compactness of the globular shape obtained. ARG11, albeit simulated on a shorter time span, displays a behaviour in both water H_2O and cyclohexane cC_6H_{12} mirroring that reported for ASN11. Again, as already mentioned for GLU11, the long arms of ARG11 side chains are forming a cage-like network around the backbone, thus restraining the degrees of freedom of the latter thereby shielding its proper collapse to a globular state. Meanwhile, in cyclohexane cC_6H_{12} a fast structural reorganization of ARG11 is seen wherein the polymer's side chains are preferentially folded back inside towards the core and the backbone rather exposed to the bulk.

In summary, we observe a general tendency for those undeca-polypeptides folding in water not folding in cyclohexane, and conversely those folding in cyclohexane not folding in water. However, LYS11 fails to follow this general rule as it remains essentially extended in both water H_2O and cyclohexane cC_6H_{12}, albeit with side chains more parallel to the backbone in the latter case, see Fig. 3. This behaviour might be ascribed to a steric hindrance of the long arms side chain densely parked around the relatively short undeca-homopeptide backbone, thus significantly reducing its conformational space, not allowing the proper collapse of the polymer within the simulated time considered here.

In principle, the relative stability of each polypeptide with respect to a specific solvent can be also quantified by a direct calculation of the solvation free energy in both water H_2O and cyclohexane cC_6H_{12}. This will be carried out in the next Section. However, in interpreting a comparison with the data reported here, the differences in the flexibility conditions (fully flexible here, fully constrained in the solvation free energy calculation reported below), plays an important role as noted earlier[21].

Figure 3 reports snapshots of the most representative conformers in all considered cases, and Table 3 summarizes these results in a synoptic form.

Additional insights can be obtained by monitoring the evolution in the fractions of peptide-solvent and intra-peptide hydrogen bonds. Confining our attention to the initial equilibration stage of few nanoseconds first, we report the total number of hydrogen bonds with water H_2O for both GLY11 in Fig.4 (black line of the top panel) and ALA11 (Fig.4 black line of the top panel). Correspondingly, the total number of intra-chain hydrogen bonds are also reported in for GLY11 (Fig.4 red line in top panel) and for ALA11 (Fig.4 red line in bottom panel). For GLY11, the number of hydrogen bonds with water shows a fast drop (Fig.4 black line top panel) consistent with a folding of GLY11 being further stabilized by an increase of the number of intra-chain hydrogen bonds (Fig.(4 red line top panel). This does not seem the case for ALA11 where the number of hydrogen bonds with water does not show any drop with time (Fig.(4) black line bottom panel) and the number of intra-chain hydrogen bonds remains essentially unchanged (Fig.(4 red line bottom panel). It is worth noticing that on assuming an approximate average value of ($20\,\mathrm{kJ\,mol^{-1}}$) for each hydrogen bond, a typical total energy involved for approximately 30 bonds (see Figure 4) is of the order of $600\,\mathrm{kJ\,mol^{-1}}$ which is comparable with the solvation free energy discussed in the next Section. This confirms the fundamental role played by the hydrogen bonds in stabilizing the protein fold as discussed in detail in Ref.[44]. At equilibrium, the above findings are confirmed. Fig.4-(a) and (c) report the fluctuations of the number solute-water H_2O and solute-solute hydrogen bonds, respectively. Black lines refer to GLY11 and red lines to ALA11. Note that the total number of hydrogen bonds with water H_2O is of the order of 25 for both GLY11 and ALA11 whereas the total number of internal hydrogen bonds is stably of the order of 2.5 and for GLY11 and it is highly fluctuating between 0 and 2.5 in the case of ALA11, coherently with the lack of a stable fold for ALA11 in water. Fig.4(b) display the distributions of the total number of hydrogen bonds of both GLY11 (black) and ALA11 (red) with water that turns out to be nearly identical, as visible.

In cyclohexane cC_6H_{12} the behaviour is clearly different. Fig.4(d) shows the fluctuations of the number of solute-solute hydrogen bonds in cyclohexane cC_6H_{12} for GLY11 (black line) and ALA11 (red line). Here the total number of intra-chain hydrogen bonds is significantly higher for GLY11 (black line) than for ALA11 (red line), indicating a much more stable fold in the GLY11 case. When compared to water H_2O, the total number of intra-chain hydrogen bond for ALA11 is smaller in cyclohexane cC_6H_{12} than in water H_2O (compare the red lines in Figure 4(c) and Figure 4)(d), so ALA11 is still relatively less stable in cC_6H_{12} than in H_2O.

SI report the same quantities for ILE11, ASN11, LYS11, ARG11, and GLU11. Supplementary Figure SIV(a) display the total number of hydrogen bonds of ILE11 (black line), ASN11 (red line), LYS11 (green line), ARG11 (blue line), and GLU11 (magenta line) with water H_2O. While a nearly constant trend is observed in all cases, the actual total number decreases from GLU11

View Article Online
DOI: 10.1039/D2CP05214H

polymers	GLY11	ALA11	ILE11	ASN11	LYS11	ARG11	GLU11
H_2O	poor+	poor	poor	good	good	good	poor
cC_6H_{12}	good+	good+	good+	poor+	good	poor	good

Table 3 Summary of the solvent property in relation to the polymers (undeca-mer) considered here in water H_2O and cyclohexane cC_6H_{12}. Good and poor are used to point out whether the solvent tends to promote the extension or the collapse of the solute, respectively. Furthermore, the sign $+$ is an indication of either a fully extended or fully compact conformation, without any significant structural fluctuations that characterize those cases without the $+$ sign.

(the largest) to ILE11 (the smallest), with Supplementary Figure SIV(b) displaying the corresponding equilibrium distribution. The total number of solute-solute intrachain hydrogen bonds, depicted in Supplementary Figure SIV(c), also shows a constant trend with slightly variable absolute number. This number increases in the case of cyclohexane cC_6H_{12}, again due to the absence of an alternative provided by the solvent, and again decreases from GLU11 (the largest) to ILE11 (the smallest), thus confirming a stabilization effect of cyclohexane cC_6H_{12} decreasing from the charged GLU11 to the hydrophobic ILE11.

3.2 Solvation free energy

In Section 3.1 we have seen how different polypeptides behave in solvents with different polarities. This analysis highlights that the definition of 'good' and 'poor' solvent is not an absolute property but has to be related to the specificities of the polypeptides. For example, water H_2O is a poor solvent for polyglycine, polyanaline, polyisoleucine and polyglutamic acid, but it is a good solvent for polyasparagine, polylysine and polyarginine. Conversely, cyclohexane cC_6H_{12} is a poor solvent for polyasparagine and polyarginine, and it is a good solvent for polyglycine, polyalanine, polyisoleucine, and polylysine. In most cases these findings agree with our intuition and with the common view that " like dissolves like" but this is not always the case. For instance, polyglutamic acid collapses in water H_2O and remains extended in cyclohexane cC_6H_{12}, whereas a reversed behavior could be expected on the basis of the charged nature of the glutamic acid GLU residue. An even more notable exception is provided by polylysine which shows no collapse in either cyclohexane cC_6H_{12} or water H_2O, in spite of the charged nature of the lysine residue.

In drafting these conclusions two additional points must be born in mind. First, none of the investigated homo polypeptides are really hydrophobic irrespective of the polarities of their residues. Indeed, we have shown that each of the considered polypeptides form a number of hydrogen bonds with the solvent ranging from $2-3$ bonds/residue for ILE11 to more than 10 hydrogen bonds/residue for GLU11 (see Supplementary Figure SIV(a)). This is also evident from the snapshots of the initial conformation that shows in all cases a significant hydrogen bonding with the solvent, as explicitly displayed in Supplementary Figure SVI. Accordingly, none of them with the exception of GLY11 is shown to have a stable fold in water H_2O (see representative snapshots in Fig.3), although clearly ILE11 has a stronger tendency to fold compared to GLU11. The second point that is worth stressing is that the difference between extended/swollen and compact/globule is well defined only for sufficiently longer polypeptides compared to those analyzed in the present work.

Next, we turn our attention to the corresponding solvation free energies that can be computed via thermodynamic integration. As anticipated, the aims here are twofold. First, we would like to extend our previous calculation[14] for a single amino acid side chain equivalent – a single amino acid where the backbone part of the amino acid has been replaced with a single hydrogen atom, to include the effect of the backbone as well as the dependence of the number n of included residues. Relevant questions here are possible non-linear effects of the solvation free energy as a function of the number of repeated units, and whether there is a mirror symmetry in by changing a highly polar solvent such as water H_2O to an apolar organic solvent such as cyclohexane cC_6H_{12}. For instance, is the solvation free energy of $(G)_n$ equal to n times the solvation free energy of a single amino acid $(G)_1$? And is this depending on the polarity properties of the amino acids and/or the polarity of the solvent? Both questions will be addressed in the present section.

A second issue of paramount importance is what is the main driving force to solvation. A conventional simple accepted picture is that solvation includes two different and competing processes: the entropically unfavourable creation of a cavity, and the enthalpically favourable attractive dispersion contributions arising by the introduction of the solute. Note that in water this picture is known to be affected above a critical solute size of 1 nm in view of the fact that sufficiently small solutes (smaller that 1 nm) do not affect the water hydrogen bond network[17].

While we will not consider all the 18 side chains studied in Dongmo Foumthuim *et al.*[14], our representative results will be sufficient to understand the emerging pattern.

Figure 5 displays the solvation free energy for water H_2O (Figure 5a) and cyclohexane cC_6H_{12} (Figure 5b), at room temperature (25°C) in both cases. All corresponding values can be found in Supplementary Table SII, and Supplementary Table SIII. The ordering is according the nominal character of the amino acid from hydrophobic (left) to polar (right). Glycine GLY is listed first as the simplest case.

As visible in Fig.5a all considered polypeptides display negative solvation free energy, indicating that in water H_2O the onset of attractive energies originating upon the insertion of the polypeptides overwhelms the entropic cost of creating a cavity. The effect is more pronounced for polar and charged amino acids, with the ΔG decreasing with the increase of the number of identical amino acids n from 3 to 11 (i.e. the length of the peptide).

The same trend is observed for the solvation free energy in cyclohexane cC_6H_{12}, as reported in Fig. 5b. While in water H_2O this behaviour is in marked contrast with that of single amino

View Article Online
DOI: 10.1039/D2CP05214H

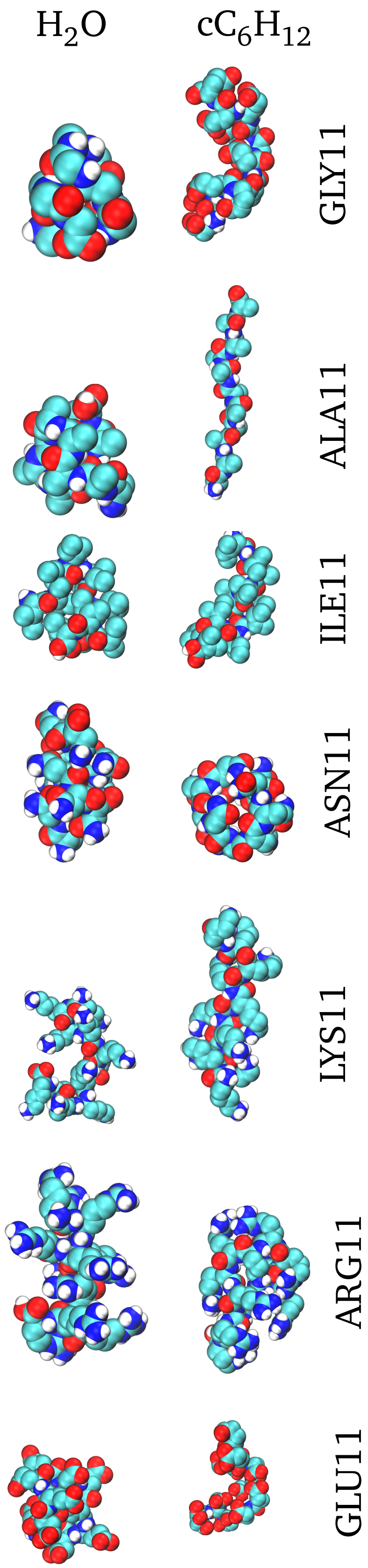
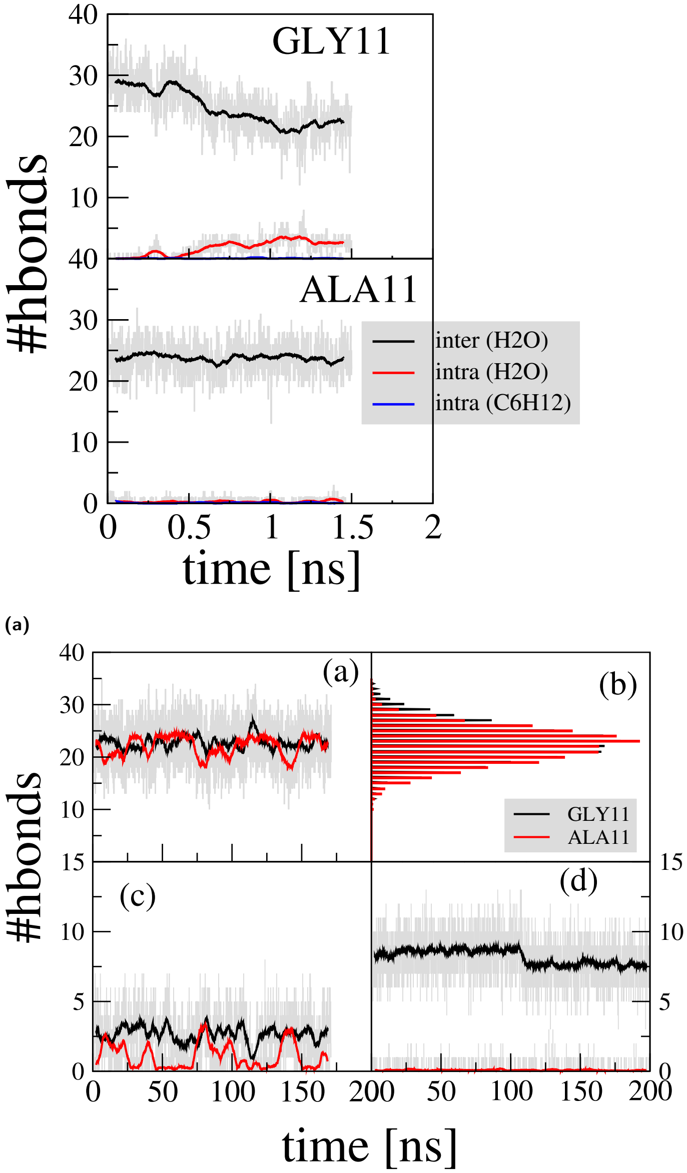


Fig. 3 Representative snapshots of the smallest R_g conformers i.e. the most collapsed conformations. On the left the structures obtained in water H_2O and on the right those obtained in cyclohexane cC_6H_{12}. From top to bottom the corresponding structures for GLY11, ALA11, ILE11, ASN11, LYS11, ARG11 and GLU11, respectively.



(a)

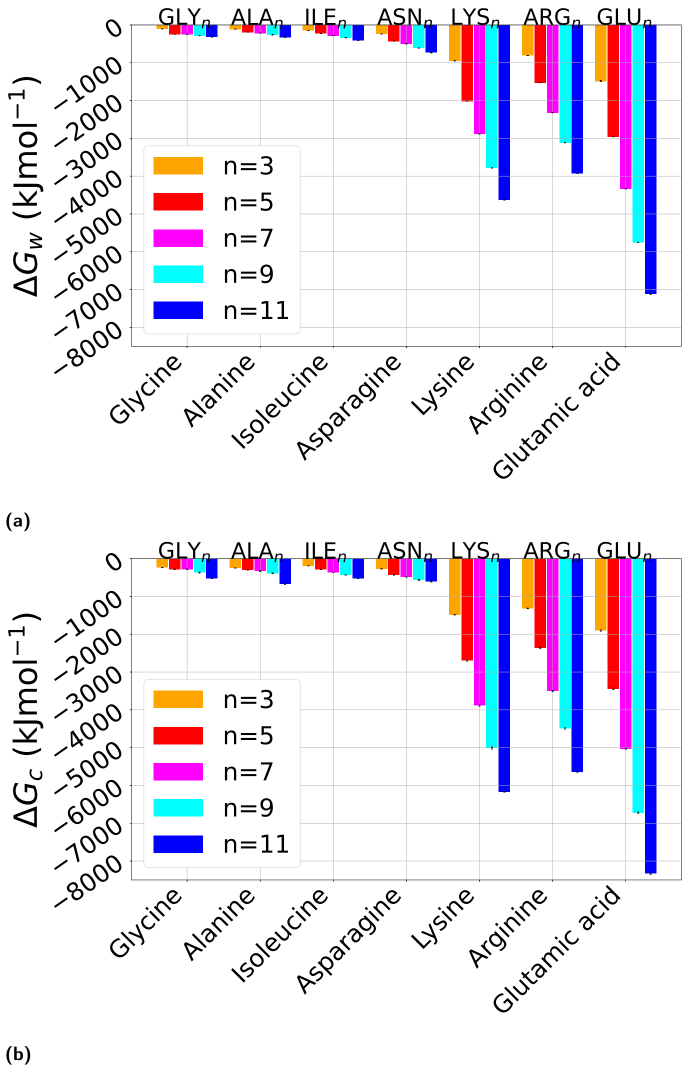
(b)

Fig. 4 Top panel: Initial stage evolution of the number of hydrogen bonds for GLY11 (top) and ALA11 (bottom). Both inter- (H_2O-solute black line) and intra- (solute-solute in H_2O red line and cC_6H_{12} blue line) molecular hydrogen bonds are plotted. Bottom panel: Long time evolution of the number of hydrogen bonds of GLY11 (black line) and ALA11 (red line). (a) Solute-H_2O hydrogen bonds (GLY11 black line, ALA11 red line); (b) Histogram distribution of (a) (GLY11 black line, ALA11 red line); (c) Solute-Solute hydrogen bonds in H_2O (GLY11 black line, ALA11 red line); (d) Solute-Solute hydrogen bonds in cC_6H_{12} (GLY11 black line, ALA11 red line)

acids equivalents[14] where the solvation free energy ΔG_w is found to be large and positive for hydrophobic amino acids side chains equivalent, and large and negative for polar ones[14], it is in accord with a similar computational study of tripeptides in water[16].

In cyclohexane cC_6H_{12}, however, this behaviour is more intriguing. We note that *both* hydrophobic *and* polar peptides have negative solvation free energy ΔG_c in cyclohexane cC_6H_{12}, more negative for polar than for hydrophobic ones[14]. A calculation of

View Article Online
DOI: 10.1039/D2CP05214H



(a)

(b)

Fig. 5 Solvation free energy ΔG_{solv} : (a) ΔG_w from vacuum to water H_2O at 25°C and (b) ΔG_c from vacuum to cyclohexane cC_6H_{12}. The polypeptides shown in the x-axis are representative of the full hydrophobic scale following previous work [14]. Their lengths vary from tri- ($n = 3$) to undeca- ($n = 11$) polypeptides. Note that all plots are in the same scale.

the transfer free energy $\Delta\Delta G_{w>c}$ from water H_2O to cyclohexane cC_6H_{12}, however restores our intuitive picture in terms of the relative stability.

Fig.6 reports $\Delta\Delta G_{w>c}$ for polypeptides from water H_2O to cyclohexane cC_6H_{12} with the same arrangement and ordering of Figs. 5, hydrophobic (left) and polar (right), at different peptide length n. With the exclusion of asparagine (ASN), all $\Delta\Delta G_{w>c}$ are negative, significantly larger for polar than for hydrophobic polypeptides although $n = 3$ is clearly an outlier for hydrophobic polypeptides likely due to its small size. As anticipated, previously alluded to in Fig. 2 and reported in Supplementary Table SI, all tripeptides ($n = 3$) have sizes smaller that 1 nm that is known to be a critical value for solvation in water [17], whereas all peptides with $n > 3$ have sizes larger than this value. In this respect, present results are complementary to those on tripeptides reported in Ref. [16].

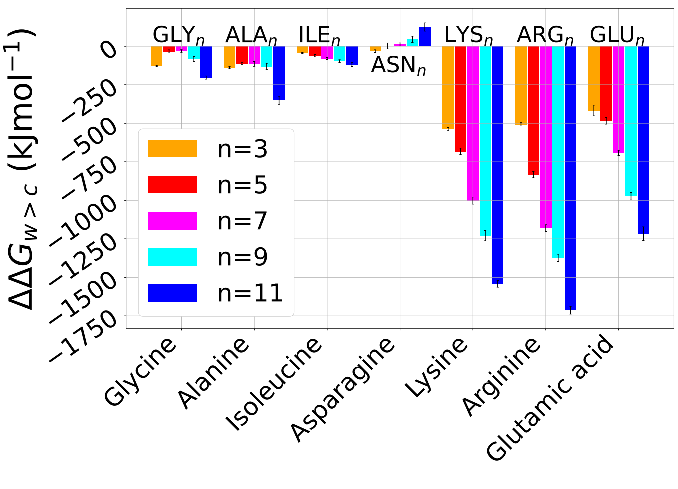


Fig. 6 $\Delta\Delta G_{w>c}$ from water H_2O to cyclohexane cC_6H_{12} at 25°C. Ordering is the same as in Fig. 5.

Consider GLY$_n$ first (the outermost left in Fig.6). Here $\Delta\Delta G_{w>c}$ is small and negative, indicating a stabilizing effect of cyclohexane cC_6H_{12} compared to water H_2O. This agrees with the calculations of Section 3.1 and confirms findings from previous studies [21,40]. However, the trend is not linear: $\Delta\Delta G_{w>c}$ increases from $n = 3$ to $n = 7$ and then decreases again for higher $n = 9, 11$. Polyalanine ALA$_n$ and polyisoleucine ILE$_n$ show a more regular increasing trend, whereas polyasparagine ASN$_n$ switches from negative to positive $\Delta\Delta G_{w>c}$ as n increases. Polar and charged polypeptides, on the other hand, display a much more significantly negative $\Delta\Delta G_{w>c}$ with a monotonic increase with n, a result that defies with our physical intuition, but it is again in agreement with results on tripeptides [16].

The emerging scenario is then that the stability of a (homo) polypeptide is mainly dictated by the polarity of the solute, with the polarity of the solvent playing a minor role

3.3 Entropy-enthalpy compensation

Two remaining issues are left from the results of previous sessions. The first issue is whether any observed process is predominantly enthalpically or entropically driven, and it will be discussed in the present Session. This can be conveniently obtained by the analysis of the solvation free energy at different temperatures that allows to separate out the entropy and the enthalpy contributions, as anticipated in Sect.2.

As anticipated, the solvation free energy ΔG can be factorized in two terms. First, the creation of a cavity in the solvent to accomodate the solute. This process is clearly entropically unfavourable so $T\Delta S < 0$ ($-T\Delta S > 0$). However, attractive interactions may form upon inserting the solute in the cavity, thus leading to a favourable process with $\Delta H < 0$. If the two processes happen to balance each other, then $\Delta G \approx 0$ and $-T\Delta S = -\Delta H$, thus leading to a perfect anticorrelation in the $-T\Delta S$ versus ΔH plane, known as "entropy-enthalpy compensation" with a slope $= -1$ (see Supplementary Figure SVII). If the slope is > -1, then the system is entropically driven, conversely is enthalpically driven.

View Article Online
DOI: 10.1039/D2CP05214H

Supplementary Figure SVIII and Supplementary Figure SIX display the temperature dependence of ΔG_w in water H_2O and ΔG_c in cyclohexane cC_6H_{12} respectively. Both are increasing function of the temperature as expected since both $T\Delta S_w$ and $T\Delta S_c$ are entropic positive costs irrespective of the solvent polarity, in agreement with the results from the single amino acid side chain equivalents as well as past experimental results[14]. Curvatures are however different depending of the specific solvent and also on the length n of the polypeptide, indicating a very complex patchwork of interactions that in water may also depend on the size of the polypeptide[17].

In Ref.[14], we reported this calculation for each single amino acid side chain equivalent. In water H_2O, hydrophobic amino acid side chain equivalents were found to comply the entropy-enthalpy compensation rule reasonably well, with a wide distribution of values along the line with slope ≈ -1 in the $-T\Delta S$ vs ΔH plane, depending on the specificity of each single residue. Polar amino acid side chain equivalents showed instead a tendency to lump together around a specific region of this line, with the exception of arginine ARG. In cyclohexane cC_6H_{12} the tendency to lump around similar state points was found to be even more pronounced for both polar and hydrophobic amino acids[14].

The values of the slopes along with the intercepts to origin and the corresponding correlation coefficients are reported in Supplementary Table SVII for all considered polypeptides and for both H_2O and cC_6H_{12}. Interestingly, all slopes are found < 1 indicating that all these solvation processes are largely enthalpically dominated.

Fig.7 reports the results of this analysis, where the entropic part of the free energy $-T\Delta S$ is plotted as a function of the enthalpic part ΔH. Each panel 7a-7g includes points computed at different lengths from $n = 3$ to $n = 11$ for all the considered polypeptides. In all cases, data for water H_2O are in black, those for cyclohexane cC_6H_{12} are in red.

Consider the glycine GLY case first, see Fig.7a. In water H_2O (black), nearly all different points G_3-G_{11} lump very closely one another along a line with slope approximately -1. By contrast, in cyclohexane cC_6H_{12} (Fig.7a red) there is a very clear anti-correlation in the sense that ΔH is decreasing with increasing length n, with a corresponding increase of $-T\Delta S$. That is, a gain in enthalpy translates into a corresponding loss of entropy. This corresponds exactly to the entropy-enthalphy compensation usually found in water H_2O (see e.g.[11,12], this time in cyclohexane cC_6H_{12} rather than in water, and it reflects the fact that cyclohexane cC_6H_{12} is a good solvent for polyglycine whereas water H_2O is poor one, in agreement with the results of Section 3.1. The cases of alanine ALA (Fig. 7b) and isoleucine ILE (Fig. 7c) are expected to follow a similar pattern on the basis of their hydrophobic character (Table 1), but they appear to present a more complex behaviour. In the case of polyalanine ALA (Fig. 7b) a rather similar behaviour in water H_2O (black) and cyclohexane cC_6H_{12} (red) is found (note the two scales of Figs. 7a and 7b are nearly equivalent), suggesting a similar behaviour for polyglycine and GLY and polyalanine ALA. An additional notable feature of polyalanine ALA in water H_2O is the irregular dependence as a function of n, with $n = 11$ very different from all others, in line with the same trend displayed for $\Delta\Delta G_{w>c}$ (Fig.6). Polyisoleucine ILE (Fig. 7c) also shows an entropy-enthalpy compensation for both water H_2O and cyclohexane cC_6H_{12}, but with a much more linear dependence on n. Interestingly, polyasparagine ASN also displays a similar pattern (Fig.7d) where for polylysine LYS (Fig.7e), polyarginine ARG (Fig.7f), and polyglutamic acid GLU (Fig.7f) a rather different trend is observed for water H_2O and cyclohexane cC_6H_{12}, in all cases with a slope significantly smaller than -1, indicating a predominant enthalpic role. Here, we emphasize again that the assumed temperature dependence reported in Eq. 4 is phenomenological and it might break down for some of the cases reported here, although it has been found to work rather well in past similar studies on single amino acid side chain equivalents both in water H_2O[14,16] and in cyclohexane cC_6H_{12}. More robust direct calculations are possible[28] albeit much more computational demanding.

3.4 Chain length dependence of solvation free energy, ΔG : implication on additivity

The second point is related to the n dependence of ΔG in H_2O and in cyclohexane cC_6H_{12} that was anticipated in Fig.5. Here the relevant question is whether $\Delta G_n \propto n$ (linear dependence on the length) or there exist non-linear effects due to the backbone, as it was observed in the case of tripeptides[16]. Note that in water marked change is expected when a hydrophobic solute size increases from below to above 1 nm because below 1 nm a cavity able to accommodate the solute can be created without affecting the hydrogen bond network[17] and the tripeptides considered in Ref.[16] were all smaller that 1 nm.

Fig.8 reports our results for ΔG (black circles), and it includes also the corresponding dependence of ΔH (blue triangles) and $T\Delta S$ (magenta squares), in water H_2O panels (a)-(g) and in cyclohexane cC_6H_{12} panels (h)-(n). In all cases solid lines represent a linear fit. Note that $T\Delta S$ and ΔH both decrease as a function of n indicating an enthalpic gain and an entropic loss. Fig.5 already suggested a linear dependence on n of both ΔG_w and ΔG_c for all consider polypeptides. This is indeed confirmed by Fig.8 (black lines) but with different slopes, smaller for the hydrophobic polypeptides (GLY, ALA, ILE, top three row panels (a) to (c) for water H_2O and (h) to (j) for cyclohexane cC_6H_{12}), as well as for ASN ((d) for water H_2O and (k) for cyclohexane cC_6H_{12}), larger in all cases for the polar polypeptides (LYS, ARG, GLU) (lower four panels (e) to (g) in water H_2O and (l) to (n) in cyclohexane. Upon splitting in the enthalpic and entropic terms, reveals however a rather different weight of the two contributions in the different cases. For GLY, ALA, ILE and ASN, the relative weights of ΔH and $T\Delta S$ appears to be comparable and results into a weak increase of ΔG_w and ΔG_c as a function of n (Figs. 8a-8d), in agreement with the findings of Fig.5. By contrast, LYS, ARG, GLU have a much stronger n dependence stemming from ΔH as is clearly visible in Figs. 8e-8g, so its additivity is purely enthalpically driven. While this is clearly consistent with the different trends observed in the enthalpy-entropy plots of Fig.7e-7g, the very similar behaviour in water H_2O and cyclohexane cC_6H_{12} is rather surprising and it will require further analysis that are

View Article Online
DOI: 10.1039/D2CP05214H

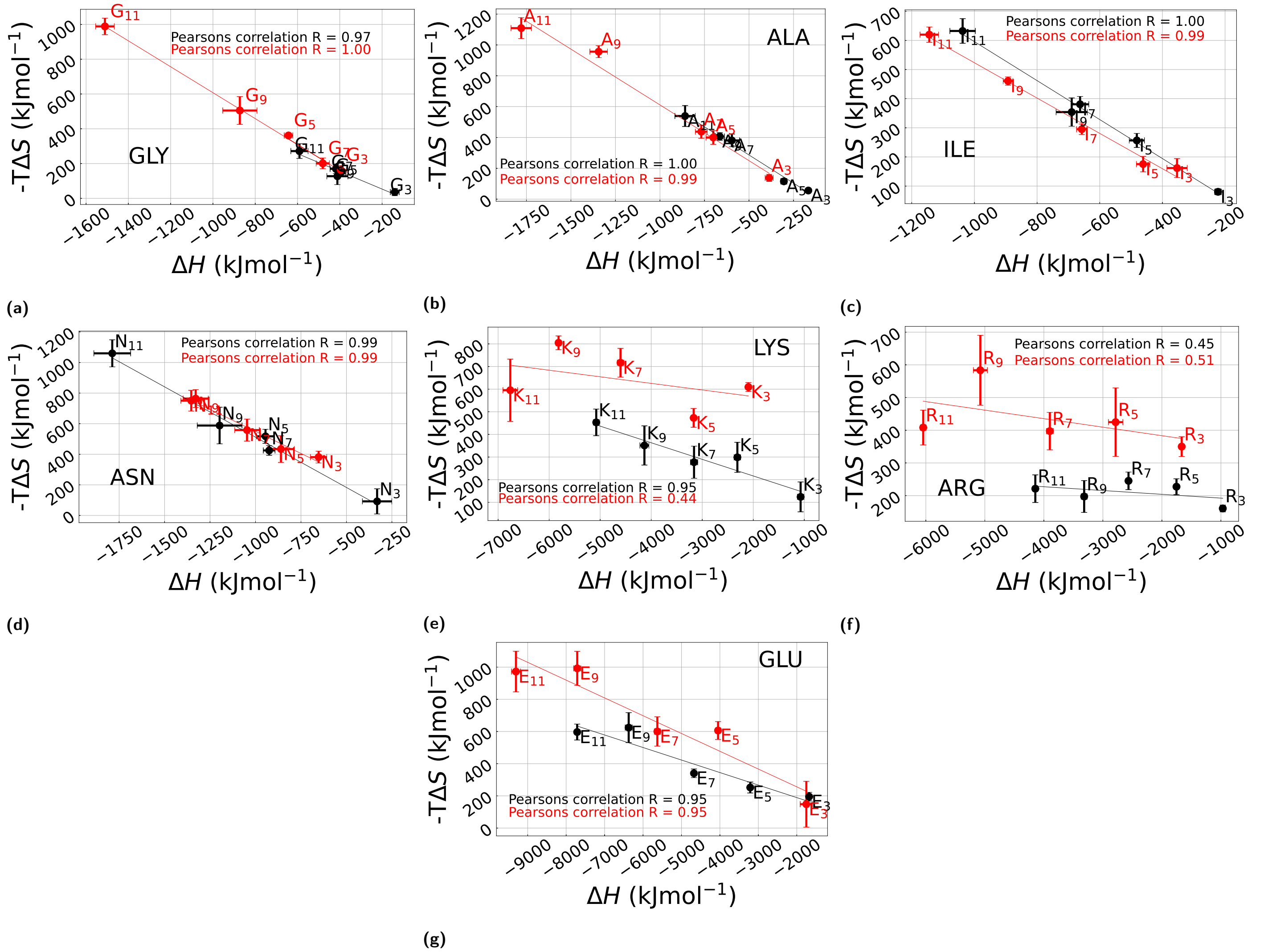


Fig. 7 Entropic contribution $-T\Delta S$ of the solvation free energy ΔG as a function of the enthalpic counterpart ΔH in the case of water H_2O and cyclohexane cC_6H_{12} for different polymers length. The solvation data in water H_2O are displayed in black and those in cyclohexane cC_6H_{12} are plotted in red while the error bars represent the standard deviations. The subplots annotated from (a) to (g) correspond to each of the polypeptides used here. These are GLY, ALA, ILE, ASN, LYS, ARG, and GLU. Furthermore, the continuous lines represent the linear fits of the simulation data. Please note that different scales have been used in different cases.

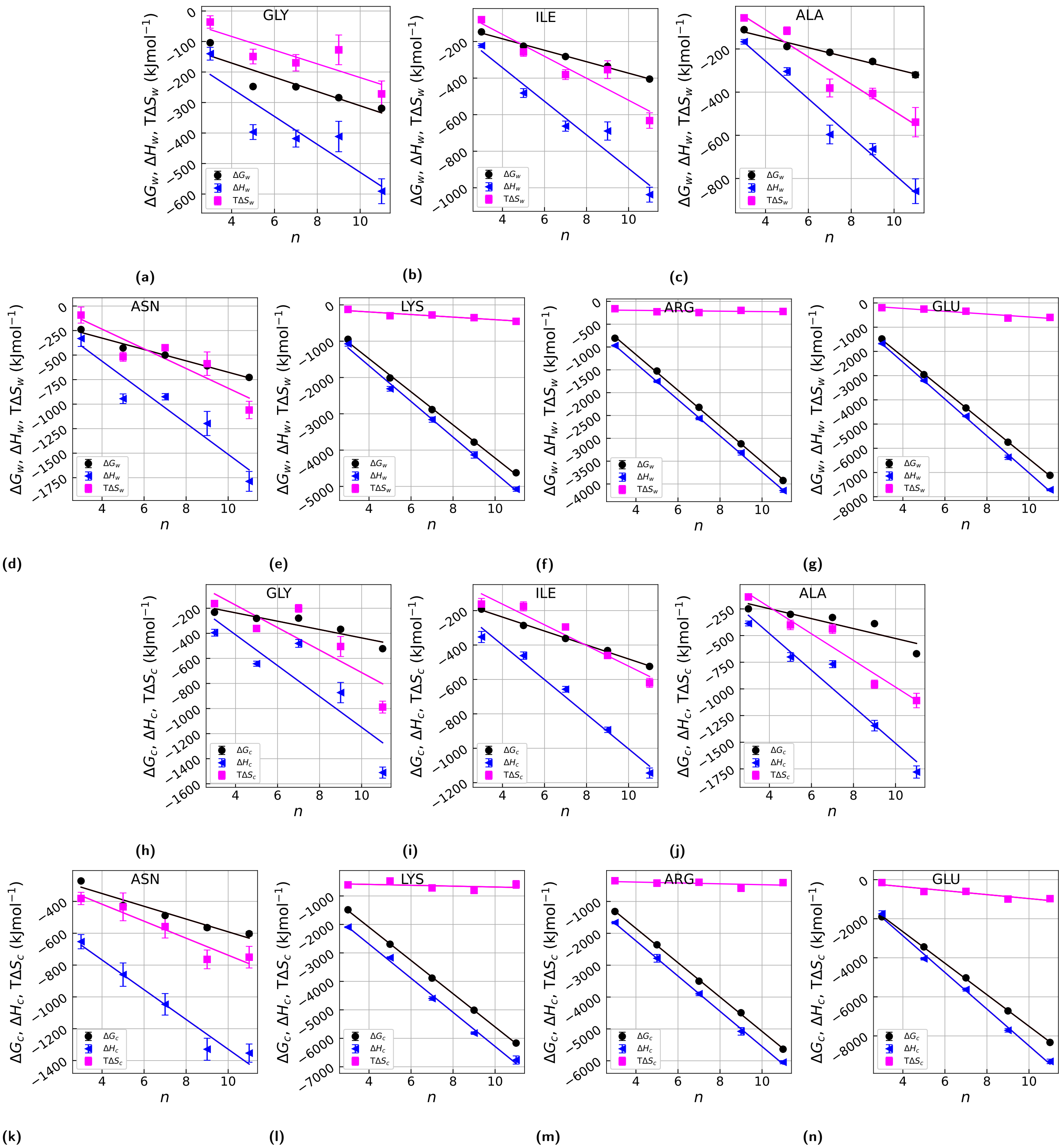
planned in the future.

Another related relevant issues concerns the relation with past results referring to the solvation free energy ΔG_1 for a single amino acid side chain equivalent[14], that is a single amino acid with the backbone part replaced with a single hydrogen atom. We show this analysis in Supplementary Figure SV where ΔG_n is plotted versus $n \times \Delta G_1$ both in water H_2O and in cyclohexane cC_6H_{12} for all considered polypeptides, with the exception of GLY for which there is clearly no amino acid side chain equivalent since it does not have a proper side chain. The results highlights rather clearly the importance of the backbone in particular for ALA and ILE for which a significant deviation from the naive expectation $\Delta G_n \propto n(\Delta G_1)$ is observed. Again, this is consistent with the relevant role of the backbone in the case of nominally hydrophobic polypeptides.

4 Conclusions

In this paper, we have addressed the issue of "good" and "poor" solvents in the framework of polypeptides of different polarities, both hydrophobic and polar, and including polyglycines as a reference point. Here the definition of good and poor solvent refers to the common view of " like dissolves like": polar solutes dissolve in polar solvents and hydrophobic solutes dissolve in hydrophobic (apolar) solvents. Polar solvents have typically large dipole moments and high dielectric constants, a feature that can be easily rationalized by the fact that high dielectric constants favour the tendency to dissociate and hence forming dipoles. A paradigmatic example of polar solvent is water (dielectric constant ≈ 80), and hence polar solvents typically mix with water. As a representative example of hydrophobic solvent, we have considered cyclohexane that has dielectric constant ≈ 2 and hence can be considered at the opposite end of water. A similar reasoning applies to so-

View Article Online
DOI: 10.1039/D2CP05214H



(a) (b) (c)

(d) (e) (f) (g)

(h) (i) (j)

(k) (l) (m) (n)

Fig. 8 Solvation free energy, enthalpy, and entropy (ΔG, ΔH, $T\Delta$) changes with the polymer chain length n in H_2O and in cyclohexane cC_6H_{12} at 25°C for each of the polypeptides considered in this work (GLY, ALA, ILE, ASN, LYS, ARG, GLU). The continuous lines connecting the points are the representative linear fitting. The data representing the hydration thermodynamics in water H_2O are shown from (a) to (g), whilst those corresponding to cyclohexane cC_6H_{12} are plotted from (h) to (n) for each of the polypeptides, respectively. Negative ΔG and ΔH represent an energetic gain upon solvation, whereas a negative $T\Delta S$ represent an entropic loss, upon solvation.

lutes that can be classified in polar and hydrophobic based on the same rationale as the solvent. Hence good and poor solvents are to be defined with respect to a specific solute. A fully hydrophobic polypeptide is expected to collapse in water (water is a poor solvent), but it remains extended in cyclohexane (cyclohexane is a then good solvent). Conversely, a fully polar polypeptide usually folds in its own poor solvent such as cyclohexane, while remaining extended in its own good solvent such as water. Hence, the

View Article Online
DOI: 10.1039/D2CP05214H

solvent quality always requires the solvent polarity to be unambiguously defined. This is especially important for polypeptides as they are formed by an identical backbone part, plus a sum of single side chains that provide their hydrophobic/polar character. While the polarity character is usually attributed to the side chains on the basis of their chemical characters, a much more robust indication is given by their solvation free energies in solvents with different polarities, and we have studied their properties in the present paper.

Our main findings can be summarized as follows.

1) There is no general mirror symmetry between the behaviour of hydrophobic/polar polypeptides in water/cyclohexane, due to the presence of the backbone, as well as of the different energy scales involved. Hence hydrophobic polypeptides in water do not behave as polar peptides in cyclohexane, nor the other way around. Polyglycine is formed by n different residues having a single hydrogen atom as a side chain, and it is usually regarded as a rough model for the peptide backbone of a protein. We find that it collapses in water and it remains extended in cyclohexane, so water is a poor solvent for polyglycine (in line with past studies), and cyclohexane a good one. Accordingly, the solvation free energy in cyclohexane ΔG_c is negative and decreases approximately linearly with the number n of residues. Interestingly, a similar trend is also found for the solvation free energy in water ΔG_w with the transfer free energy $\Delta\Delta G_{w>c}$ negative and decreasing with the length of the polypeptide. Additional hydrophobic polypeptides, such as alanine ALA and isoleucine ILE, behave similarly to GLY, with some small differences. These results can be rationalized as follows. Firstly, none of these polypeptides are really hydrophobic irrespective of the polarities of the single side chain. This is evident since they all form at least 2-3 hydrogen bonds/residue. Secondly, the solvation free energy is composed by an entropically unfavourable term associated with the creation of a cavity, and an enthalpic favourable term originating upon the insertion of the polypeptides in the solvent. Our results indicate that the latter always dominate the former leading to negative solvation free energies.

2) Polar polypeptides such as ASN$_n$, LYS$_n$, ARG$_n$, and GLU$_n$ markedly deviate from the mirror symmetry. GLU$_n$ collapses in water but not in cyclohexane, whereas both water and cyclohexane are good solvents for LYS$_n$. Accordingly, the transfer free energy $\Delta\Delta G_{w>c}$ from water to cyclohexane is found negative, linearly decreasing with the number n of residues, and significantly more negative than the hydrophobic counterparts. LYS$_n$, ARG$_n$, and GLU$_n$ are mostly enthalpically driven, whereas in ASN$_n$, as well as all the hydrophobic polypeptides, the driving force is a mixture of enthalpic and entropic contributions. These results suggest that the solvation process is mainly dominated by the polarity of the solute, with the solvent playing a minor role.

3) For all hydrophobic polypeptides as well as for ANS$_n$, there is nearly a similar entropy-enthalpy compensation in both water and cyclohexane, whereas for the other polar polypeptides LYS$_n$, ARG$_n$, and GLU$_n$ there is a marked difference. Combined with previous point, this shows that ANS$_n$ hardly belongs to the same class as LYS$_n$, ARG$_n$, and GLU$_n$, and more generally that the rough polar/hydrophobic division of the amino acids scale is not representing well the complexity of the interactions, and additional features (e.g. charge, size, etc.) should be taken into account. The peculiar properties of ASN$_n$ reported throughout this study might also be related to its marked propensity together with aspartic acid ASP to populate loop regions in protein structures thus most often with no defined secondary structure[29].

While the present work is focused specifically on the solvation process of polypeptides and its dependence on both the solvent and peptide polarities, a similar study has been tackled by the present authors also for a specific synthetic polymer displaying a coil-helix transition and it will be presented elsewhere. Coupled with the present findings, the general scenario presents still some missing points requiring further studies. One promising route that has been already addressed in past studies[28], is the quantification of the individual entropic and enthalpic solute-solvent and solvent-solvent contributions, thus allowing a quantitative assessment on the exact putative cancellation of the solvent-solvent enthalpy and solvent-solvent entropy in water and not in cyclohexane. We are planning to explore this possibility in a future dedicated study. All together it is hoped that a systematic analysis as those outlined above will provide new insights on the nuances solvation mechanism in different solvents, a process which is ubiquitous in biological systems.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors would like to acknowledge useful discussions and correspondence with Giuseppe Graziano, George Rose and Chris Oostenbrink, as well as networking support by the COST Action CA17139. The use of the SCSCF multiprocessor cluster at the Università Ca’ Foscari Venezia is gratefully acknowledged. We also acknowledge the CINECA projects HP10CGFUDT and HP10C1XOOJ 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).

View Article Online
DOI: 10.1039/D2CP05214H

Notes and references

1 P. Flory, *Statistical mechanics of chain molecules*, Interscience Publishers, 1969.

2 M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics (International Series of Monographs on Physics)*, Clarendon Press, 1988.

3 A. R. Khokhlov, A. Y. Grosberg and V. S. Pande, *Statistical Physics of Macromolecules (Polymers and Complex Materials)*, American Institute of Physics, 1994th edn, 2002.

4 M. Rubinstein and R. H. Colby, *Polymer Physics (Chemistry)*, Oxford University Press, 1st edn, 2003.

5 P. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979.

6 S. M. Bhattacharjee, A. Giacometti and A. Maritan, *JOURNAL OF PHYSICS-CONDENSED MATTER*, 2013, **25**, 503101/1–15.

7 D. W. Bolen and G. D. Rose, *Annu. Rev. Biochem.*, 2008, **77**, 339–362.

8 P. G. Wolynes, *Proceedings of the National Academy of Sciences*, 1995, **92**, 2426–2427.

9 T. Meyer, V. Gabelica, H. Grubmüller and M. Orozco, *WIREs Computational Molecular Science*, 2013, **3**, 408–425.

10 M. Carrer, T. Skrbic, S. L. Bore, G. Milano, M. Cascella and A. Giacometti, *The Journal of Physical Chemistry B*, 2020, **124**, 6448–6458.

11 T. Hayashi, S. Yasuda, T. Škrbić, A. Giacometti and M. Kinoshita, *The Journal of Chemical Physics*, 2017, **147**, 125102.

12 T. Hayashi, M. Inoue, S. Yasuda, E. Petretto, T. Škrbić, A. Giacometti and M. Kinoshita, *The Journal of Chemical Physics*, 2018, **149**, 045105.

13 D. Karandur, K.-Y. Wong and B. M. Pettitt, *The Journal of Physical Chemistry B*, 2014, **118**, 9565–9572.

14 C. J. Dongmo Foumthuim, M. Carrer, M. Houvet, T. Škrbić, G. Graziano and A. Giacometti, *Phys. Chem. Chem. Phys.*, 2020, **22**, 25848–25858.

15 R. 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.

16 T. Hajari and N. F. A. van der Vegt, *Journal of Chemical Physics*, 2015, **142**, 144502.

17 D. Chandler, *Nature*, 2005, **437**, 640–647.

18 D. Voet and J. G. Voet, *Biochemistry*, John Wiley & Sons, 2010.

19 D. S. Tomar, D. Asthagiri and V. Weber, *Biophysical Journal*, 2013, **105**, 1482–1490.

20 F. Avbelj and R. L. Baldwin, *Proceedings of the National Academy of Sciences*, 2009, **106**, 3137–3141.

21 H. Kokubo, R. C. Harris, D. Asthagiri and B. M. Pettitt, *The Journal of Physical Chemistry B*, 2013, **117**, 16428–16435.

22 R. Staritzbichler, W. Gu and V. Helms, *The Journal of Physical Chemistry B*, 2005, **109**, 19000–19007.

23 G. König, S. Bruckner and S. Boresch, *Biophysical Journal*, 2010, **104**, 453–463.

24 C. Y. Hu, H. Kokubo, G. C. Lynch, D. W. Bolen and B. M. Pettitt, *Protein Science*, 2010, **19**, 1011–1022.

25 D. Frenkel and B. Smit, *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)*, Academic Press, 2nd edn, 2001.

26 F. Fogolari, C. J. Dongmo Foumthuim, S. Fortuna, M. A. Soler, A. Corazza and G. Esposito, *Journal of Chemical Theory and Computation*, 2016, **12**, 1–8.

27 F. Fogolari, O. Maloku, C. J. Dongmo Foumthuim, A. Corazza and G. Esposito, *Journal of Chemical Information and Modeling*, 2018, **58**, 1319–1324.

28 B. Lai and C. Oostenbrink, *Theoretical Chemistry Accounts*, 2012, **131**, 1–13.

29 T. Škrbić, A. Maritan, A. Giacometti and J. R. Banavar, *Protein Science*, 2021, **30**, 818–829.

30 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, **4**, 1–17.

31 N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. van Gunsteren, *European Biophysics Journal*, 2011, **40**, 843.

32 C. Oostenbrink, A. Villa, A. E. Mark and W. F. Van Gunsteren, *Journal of computational chemistry*, 2004, **25**, 1656–1676.

33 A. Villa and A. Mark, *Journal of Computational Chemistry*, 2002, **23**, 548–553.

34 M. M. Reif, P. H. Hünenberger and C. Oostenbrink, *Journal of Chemical Theory and Computation*, 2012, **8**, 3705–3723.

35 M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *The Journal of Chemical Physics*, 2003, **119**, 5740–5761.

36 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19 – 25.

37 D. Foumthuim, C. J., A. Corazza, R. Berni, G. Esposito and F. Fogolari, *BioMed Research International*, 2018, **2018**, 1–14.

38 F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander and M. Scharf, *Journal of computational chemistry*, 1995, **16**, 273–284.

39 H. Gong and G. D. Rose, *Proceedings of the National Academy of Sciences*, 2008, **105**, 3321–3326.

40 H. T. Tran, A. Mao and R. V. Pappu, *Journal of the American Chemical Society*, 2008, **130**, 7380–7392.

41 D. Karandur, R. C. Harris and B. M. Pettitt, *Protein Science*, 2016, **25**, 103–110.

42 A. Merlino, N. Pontillo and G. Graziano, *Physical Chemistry Chemical Physics*, 2017, **19**, 751–756.

43 C. 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.

44 G. D. Rose, P. J. Fleming, J. R. Banavar and A. Maritan, *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 2006, **103**, 16623–16633.