Proteins in binary solvents

Francesco Spinozzi¹ · Paolo Mariani¹ · Maria Grazia Ortore¹

Received: 17 November 2015 / Accepted: 1 February 2016 / Published online: 18 March 2016
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2016

Abstract Proteins in living organisms exist in complex aqueous solutions or embedded in membranes. In solution, proteins are surrounded by a tightly bound hydration layer, which is more ordered and less mobile than bulk water. As a consequence, water plays a major role in controlling protein structure stability, conformational flexibility, dynamics, and functionality, but it also appears that protein surface regulates the structuring of the surrounding water. The presence of cosolvents can modify the hydration layer characteristics and then the whole protein structural and dynamical properties. Because cytoplasm or biological liquids are complex solutions, the knowledge of the solvation shell characteristics in mixed solvents should be considered as a crucial step in describing biological processes at molecular level. This review reports on recent studies on the structural and thermodynamic properties of model proteins dissolved in binary solvent mixtures by small-angle neutron scattering (SANS) and differential scanning microcalorimetry (DSC) techniques. We will show that contrast variation SANS experiments allow to acquire a direct knowledge of both protein structure and protein solvation shell (in terms of low-resolution shape and solvent/cosolvent composition), while DSC experiments provide information on all the relevant thermodynamic properties. We will focus on two main points. First, an extended description of the thermodynamic model used to define the equilibria between water and cosolvent molecules in the protein solvation shell

will be presented. Second, the determination of the peculiar characteristics of the protein solvation layer, which will be illustrated by considering different systems. As a conclusion, we will show that the investigation of structure and thermodynamics of proteins in binary aqueous mixtures is an important way to understand the role of hydration in protein stability and activity.

Keywords SANS · SAXS · DSC · Preferential solvation · Protein–solvent interaction · Protein–protein interaction

Introduction

Protein hydration is considered to play a relevant role in determining protein structure, function, and stability. Indeed, the solvent affects the thermodynamics and kinetics of numerous biological processes in which proteins are involved, including folding, dynamics, conformational changes, activity, binding capacity and specificity, diffusion, interactions, conductance as ion channels or membrane transporters (Virtanen et al. 2010). Solvation issues are however very general: in simple terms, it is impossible to think of biological processes that do not depend at all on solvation.

Hydration is primarily determined by the protein surface, *via* favorable interactions with water provided by charged and polar groups of the amino acid side chains. As a consequence, the protein in solution is surrounded by a tightly bound hydration layer, which has special characteristics. In particular, the bound hydration layer is more ordered and less mobile than bulk water, showing a 10 % greater mass density and 15 % greater heat capacity (Svergun et al. 1998). For several studies on protein stability and dynamics in solution, hydration effects have been taken into account

✉ Francesco Spinozzi
f.spinozzi@univpm.it

¹ Department of Life and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy

by including hydrophilic/hydrophobic terms in the effective interaction potentials (Vaiana et al. 2001). However, the extension to the overall landscape of protein and water molecules, e.g., an extension to a potential energy landscape that includes all degrees of freedom of both protein and water molecules, has been demonstrated to help in looking at the several aspects of solvation from a unified point of view and in extracting more rigorous conclusions (Vaiana et al. 2001; Zhou 2003). Indeed, the free energy landscape from implicit solvent models has been found to be quite different from that of the explicit solvent model (Zhou 2003; Juraszek and Bolhuis 2006), and differences pave the way for the recognition of new mechanisms and for the integrate description of biological processes at molecular level. For example, the presence of astonishingly long-lived water molecules was pointed-out during protein folding (Juraszek and Bolhuis 2006).

Hydration is then a very relevant issue in the analysis of biomolecular processes (Levinthal 1968; Sorenson et al. 1999; Mallamace et al. 2011; Cinelli et al. 2004), but the presence of cosolvents, as in biological (physiological) aqueous solutions, makes the matter even more difficult. In this regard, the study of proteins in aqueous binary mixtures deserves specific attention. Protein stability in binary mixtures has been studied from a long time (Timasheff and Inoue 1968; Gekko and Timasheff 1981; Rariy and Klibanov 1997; Timasheff 1998; Klibanov 2001; Timasheff and Xie 2003). In particular, many investigations have dealt with the mechanism underlying the protein denaturation process in aqueous solutions, also considering eventual modifications induced in the protein–protein interactions by cosolvents (Javid et al. 2007; Schroer et al. 2011). As a general result, it has to be noticed that cosolvents can play diverse and opposite roles: some cosolvents are known to stabilize the native structure (glycerol, trehalose, trimethyl amine N-oxide), others bring about denaturation (guanidinium hydrochloride, urea), and others present controversial features depending on their fraction in water (ethanol) (Ghosh et al. 2013).

Different pathways have been proposed to explain the unfolding process observed in the presence of cosolvents. As an example, a widely diffused point of view considers that urea directly attacks the polar backbone as well as nonpolar residues of a protein, helping water to penetrate into the core (Gabel et al. 2009; Timasheff and Xie 2003; Schellman 2003). On the other side, it has been proposed that urea breaks the tetrahedral network structure of water. The *chaotrope* activity then results in protein denaturation (Frank and Franks 1968; Soper et al. 2003). However, the structure-breaker role of urea in water, as well as its contribution to slow down water rotational dynamics, are besides a matter of debate (Bandhyopadhyay et al. 2014; Carr et al.

2013). In particular, a recent molecular dynamics (MD) study over a set of amino acids has shown that water and urea perform a cooperative effect for hydrophobic solutes and are in competition for hydrophilic solutes, underlining the specific role of the protein solvation shell (Ikeguchi et al. 2001; Karino and Matubayasi 2013).

Cosolvents able to stabilize protein structures are considered to act as *compatible osmolytes*, because they play an essential role in regulating the osmotic balance inside the cell (Roccatano 2013). Indeed, some of them, such as trehalose or trimethyl amine N-oxide (TMAO), are known to protect organisms in vivo from extreme conditions, like freezing or dehydration (Jain and Ipsita 2009; Zou et al. 2002). It is evident that many applications can benefit from the investigation of the processes involving these cosolvents, as the induced stabilization of proteins (and protein-like molecules) can be translated into preservation of both structure and functionality during storage and targeting. Concerning TMAO, experiments and molecular dynamics simulations have shown that this cosolvent counteracts the denaturing effect of urea in water (Graziano 2011). This result has been explained by a recent statistical thermodynamic model (Graziano 2012), which takes into account two main effects: the *direct* interaction of the protein surface with water or with the cosolvent molecules and the *indirect* effect of the solvent composition in the reversible work necessary to create the cavity suitable to host the unfolded state of the protein. Due to the strong attraction of TMAO to both water and urea, the average solvent density increases and determines an increase of the work for cavity creation, hence causing a stabilizing effect. Interestingly, the same model has been applied to explain the stabilizing effect of sucrose (Graziano 2012). Trehalose also shows a counteracting effect with respect to protein denaturation in the presence of urea. A recent all-atom MD study has shown that the direct attractive interaction of trehalose and water with the protein surface determines a removal of urea from the solvation shell, causing an overall stabilizing effect (Paul and Paul 2015). The effect of sugars in protein stabilization has been also explained using a radial-symmetric integral equation theory, where the key quantity is the solvent entropy (Oshima and Kinoshita 2013). The higher stability of the native state is mainly attributed to the increase in the solvent crowding caused by sugar addition.

Beyond the stabilizing or denaturing effects, cosolvents can also modify protein–protein interactions in aqueous solution. Albeit from a physical point of view protein–protein interactions are expected to be mainly modified by changes in the dielectric permittivity, cosolvents can act in a more complex way and a general approach to understand cosolvent effects should account for specific solvent–protein

interactions. Recently, efforts have been made to correlate macroscopic solvent effects to protein–protein interactions and solvent–protein interactions at atomic scale, and to tentatively paint a general interaction pattern in diverse biological environments (Vagenende et al. 2013). It should be noticed that also in this case the quantitative characterization of local protein solvation shell has been obtained from MD simulations combined with preferential interaction theory. It is evident that computer simulation of biomolecules in different solvents is a very powerful and versatile tool to investigate the solvation features, and the increasing number of papers published in this subject confirms this finding (Roccatano 2013). Nonetheless, it should be pointed out that experimental results are very important to confirm MD results.

Another point should be underlined: despite the number of molecular dynamics approaches, there is not likely an equivalent experimental effort to quantitatively describe the protein solvation shell in binary mixtures or in conditions resembling those existing *in vivo*. In the meantime, it is becoming clear that the role of cosolvents in important problems like amyloid aggregation is of primary relevance (Dzwolak 2006; Seeliger et al. 2013), and hence a joint effort of experimentalists and theoreticians to provide deep information is really necessary.

Small-angle scattering (SAS) is one of the more promising experimental techniques to get engaged in the investigation of protein solvation shell. SAS is widely used to study biological molecules in solution, as it probes size, shape and conformation of both macromolecules and macromolecular complexes on a length scale from ten to several thousand Angstroms (Fitter et al. 2006). SAS experiments can be carried out by using collimated x-ray (SAXS) or neutron (SANS) beam. The two techniques are very similar, but they give complementary results: indeed, using SANS it is possible to exploit the proton–deuteron isotopic effect, due to the completely different neutron scattering behavior of H and D (Fitter et al. 2006). This special feature has been demonstrated to be particularly important to enhance the effect of compositional changes in the solvent that surrounds the surface of a protein with respect to the solvent in the bulk (Spinozzi et al. 2002), fostering SANS as the best technique to directly highlight solvation effects. However, SANS and SAXS are low-resolution techniques and sometimes the information that can be extracted from experimental results is not univocal. We will show that the combination of SAS with other equilibrium techniques, such as differential scanning calorimetry (DSC), as well as the simultaneous fit of several experimental SAS curves, can allow to restrict the uncertainty in the interpretation of data and to derive precious structural and thermodynamic information often hidden behind those experimental tools.

In this review, we report on the combined use of SANS and DSC to study the structural and solvation properties of model proteins in binary solvents. In the first section, we will describe the thermodynamic model of solvent exchange developed by Schellman (1990, 1994, 2003, 2005), and its application to SAS and DSC. In the second section, relevant applications will be presented.

Theory

The thermodynamic model of solvent exchange and protein unfolding

To analyze the solvation properties of a protein in a binary mixture, we will consider the thermodynamic model derived in 1990 by Schellman (1990, 1994). The model has been successfully applied to many protein systems and proved to be adequate to understand the measured properties. In particular, Timasheff probed the physical chemistry of the denaturation reaction in binary mixtures (Gekko and Timasheff 1981; Timasheff 1998), underlying the importance of the exchanges occurring at the protein surface between the water and the denaturant molecules, Schellman re-interpreted the model including the changes of the excluded volume during the denaturation reaction (Schellman 2003, 2005), and, more recently, Graziano considered separately the role on the thermodynamic stability of the volume excluded to solvent molecules by the protein in the native and in the unfolded states (Graziano 2011, 2012). This point could be very relevant for a complete picture, but the original Schellman model, in which excluded volume effects are neglected, appeared sufficiently adequate for our purpose and for the resolution of the experimental techniques we employ.

The basic concept of the used model is that the protein surface, in each of possible protein j -states (typically native or unfolded states), contains a well-defined number m_j of *solvation sites* (s) that can be occupied either by a water molecule (W) or by a cosolvent (L) molecule. It has been proved that, considering cosolvents with a molecular volume larger than the one of water but of the same order of magnitude, it is not necessary to include in the model the possibility that the cosolvent molecule can occupy more than one solvation site. Hence m_j can be also seen as the number of *hydration sites* of the protein. However, since the m_j sites can be partially or completely occupied by cosolvent molecules, the thickness of the layer in the vicinity of the protein surface, defined as the *local domain*, should be considered larger than the thickness of the first hydration layer, as shown by the green line in Fig. 1. From now on, this layer in the proximity of the protein surface will be

Fig. 1 Scaled representations of a solvated protein in the native state dissolved in a cosolvent–water mixture ($C_L = 1$ M). The light- and dark-blue circles represent water molecules in the bulk and at contact with the protein, respectively. The three connected circles represent cosolvent molecules: those in the bulk are shown in yellow, while those in contact with the protein are red. The green line delimits the local domain shell with thickness larger than the thickness of the first hydration layer. *Left side*: the effective exchange constant is lower than 1 ($K_N^{\text{ex eff}} = 1/2$), determining the binding of four cosolvent molecules. *Right side*: the effective exchange constant is greater than 1 ($K_N^{\text{ex eff}} = 2$), leading to only one cosolvent molecule bound to the protein surface (Ortore et al. 2009)

indicated by the s index, while the bulk domain will be defined by 0.

Exchange equilibrium

The substitution of a cosolvent molecule that occupies one site (L_s) with a water molecule coming from the bulk solvent (W_0) is considered a mild process that involves a few $k_B T$ of free energy, k_B being Boltzmann's constant and T the absolute temperature. It can be described by the simple exchange equilibrium process

The nominal composition of the protein (P) system, which is supposed to be known, is defined in terms of molar fractions of any j -species, $X_i = \mathcal{N}_i / (\mathcal{N}_P + \mathcal{N}_W + \mathcal{N}_L)$, with $i = P, W, L$ and $\mathcal{N}_i$ representing the number of i -moles, whereas the composition of the binary solvent is expressed by the cosolvent molar fraction in the solvent, $x_L = \mathcal{N}_L / (\mathcal{N}_L + \mathcal{N}_W)$. The assumption of the Schellman model is that the thermodynamic properties of the exchange process (1) are independent on the population of the sites, and only depend on the protein j -state. The equilibrium constant of the exchange process is expressed by the thermodynamic exchange constant

$$K_j^{\text{ex}} = \frac{\phi_j}{1 - \phi_j} \frac{a_{L,0}}{a_{W,0}}. \quad (2)$$

In this definition, the activities of water and cosolvent in the bulk binary solvent ($a_{W,0}$ and $a_{L,0}$) appear: they are used to encompass the most general case of a not-ideal solution. Activities are linked to the molar fraction of the water and cosolvent in the bulk, $X_{i,0} = \mathcal{N}_{i,0} / (\mathcal{N}_P + \mathcal{N}_W + \mathcal{N}_L)$ with

$i = W, L$, by the activity coefficients γ_i , $a_{i,0} = \gamma_i X_{i,0}$. On the other hand, since it is not possible to measure activity coefficients of molecules bound to the protein surface, the concentration of water molecules in the protein surface is simply expressed by the water occupation probability ϕ_j , seen in Eq. 2. Combining the expression of K_j^{ex} with the definition of activity coefficients, we can derive the so-called effective exchange constant, simply written in terms of molar fractions,

$$K_j^{\text{ex eff}} = K_j^{\text{ex}} \frac{\gamma_W}{\gamma_L} = \frac{\phi_j}{1 - \phi_j} \frac{X_{L,0}}{X_{W,0}} = \frac{\phi_j}{1 - \phi_j} \frac{1 - x_{W,0}}{x_{W,0}} \quad (3)$$

where $x_{W,0}$ is the molar fraction of bulk water calculated in the bulk solvent, $x_{W,0} \equiv X_{W,0} / (X_{W,0} + X_{L,0})$. Notice that if $K_j^{\text{ex eff}} = 1$ it results that $x_{W,0} = \phi_j$, meaning that there is not depletion or accumulation of water in the protein surface. On the contrary, when $K_j^{\text{ex eff}} > 1$ the surface is enriched of water, while for $K_j^{\text{ex eff}} < 1$ the surface prefers to be covered by the cosolvent molecule (see Fig. 1).

Protein unfolding equilibrium

In the following we make the assumption that, from a thermodynamic point of view, only two protein states could exist: native ($j = N$) and unfolded ($j = U$). In the pure water solvent, the unfolded equilibrium is described by the following process

where NW_{sm_N} and UW_{sm_U} indicates protein molecules in the native and in the unfolded states solvated with m_N and m_U water molecules, respectively. To note, we assume that the numbers of solvation sites in both states do not change with solvent composition. The thermodynamic constant of this process depends on the molar fractions of the protein in the two states and on the water activity,

$$K_W = \frac{X_{UW_{sm_U}}}{X_{NW_{sm_N}} a_{W,0}^{m_U - m_N}}. \quad (5)$$

Since the exchange processes in the solvation sites are considered independent events, the probability that m_j sites on the surface of the protein in the j -state are occupied by n water molecules is expressed by the binomial distribution,

$$p(m_j, n) = \frac{m_j!}{n!(m_j - n)!} \phi_j^n (1 - \phi_j)^{m_j - n} \quad (6)$$

Hence, in a binary solvent the molar fraction of the protein in the j -state with all the solvation sites occupied by water is

$$X_{jW_{sm_j}} = X_j p(m_j, m_j) \equiv X_j \phi_j^{m_j} \quad (7)$$

where X_j is the molar fraction of the protein in the j -state with any population of the sites, $X_j = \sum_{n=0}^{m_j} X_{jW_{sn}L_{sm_j-n}}$.

Entering the result of Eq. 7 in the expression of the equilibrium constant in pure water (5), one finds

$$K_W = K \frac{\phi_U^{m_U}}{\phi_N^{m_N} a_{W,0}^{m_U-m_N}} \quad (8)$$

$$K = \frac{X_U}{X_N} \quad (9)$$

where K is defined as the *effective* constant of the unfolding process in the binary solvent.

Thermodynamic functions

Associated to each of the three thermodynamic constants (K_N^{ex} , K_U^{ex} and K_W , indexed by $\alpha = 1, 2, 3$), there are the changes in the free energy,

$$\Delta G_\alpha = -k_B T \log K_\alpha = \Delta H_\alpha - T \Delta S_\alpha \quad (10)$$

where the changes of enthalpy and entropy are functions of temperature on the basis of their values at the reference temperature T_o ,

$$\Delta H_\alpha = \Delta H_\alpha^\circ + \int_{T_o}^T \Delta C_{p\alpha} dT \quad (11)$$

$$\Delta S_\alpha = \Delta S_\alpha^\circ + \int_{T_o}^T \frac{\Delta C_{p\alpha}}{T} dT \quad (12)$$

The molar heat capacity variations at constant pressure can be simplified as linear functions of temperature,

$$\Delta C_{p\alpha} = \Delta C_{p\alpha}^\circ + \Delta C_{p\alpha 1}(T - T_o) \quad (13)$$

$\Delta C_{p\alpha 1}$ being the first temperature derivative of $\Delta C_{p\alpha}$ at T_o . It is worth to highlight that the present model does not enter in the description of the dependency of all the thermodynamic quantities ΔH_α° , ΔS_α° , $\Delta C_{p\alpha}^\circ$ and, eventually, $\Delta C_{p\alpha 1}$ on the detailed molecular properties of water, cosolvent and protein. They are simply considered the main fitting parameters of the model that could be estimated by analyzing calorimetric experiments. The calculation of the two unknown values, $X_{W,0}$ and $X_{L,0}$, can be numerically performed by solving the following system of equations, which represents the mass balance of water and cosolvent molecules

$$\begin{cases} X_{W,0} + X_N m_N \phi_N + X_U m_U \phi_U = (1 - X_P)(1 - x_L) \\ X_{L,0} + X_N m_N (1 - \phi_N) + X_U m_U (1 - \phi_U) = (1 - X_P)x_L. \end{cases} \quad (14)$$

Hence, taken into account the protein mass balance $X_N + X_U = X_P$ and Eq. 9, it results that $X_N = X_P/(1 + K)$ and $X_U = X_P K/(1 + K)$.

Activity coefficients

Following the standard process of physical chemistry, the activity coefficient of any j -component of the binary mixture ($j = W, L$) depends on the corresponding *excess* chemical potential, according to $\gamma_j = \exp(\mu_j^E/k_B T)$. Chemical potentials are calculated on the basis of the excess Gibbs free energy G^E according to Gibbs-Duhem equation,

$$\mu_W^E = G^E - x_{L,0} \frac{dG^E}{dx_{L,0}} \quad (15)$$

$$\mu_L^E = G^E + (1 - x_{L,0}) \frac{dG^E}{dx_{L,0}} \quad (16)$$

and depend on the molar fraction of the cosolvent in the bulk, $x_{L,0} \equiv X_{L,0}/(X_{W,0} + X_{L,0})$. Literature data of the physical-chemistry properties for binary mixtures (Ozdemir and Sadikoglu 1999) give the dependency of the reference excess Gibbs free energy (G_o^E), the reference partial excess enthalpies (H_o^E) and heat capacities (C_p^E) as a function of $x_{L,0}$, typically reported as analytic phenomenological functions. Hence, by using classical thermodynamics, values of G^E , H^E and C_p^E can be calculated at any temperature and at any molar fraction of the cosolvent in the bulk.

Effective protein unfolding equilibrium

The effective unfolding constant K expressed by Eq. 9 allows to define the effective free energy difference upon unfolding

$$\begin{aligned} \Delta G &= G_U - G_N = -k_B T \log K \\ &= -k_B T \log \left(K_W \frac{\phi_N^{m_N} a_{W,0}^{m_U}}{\phi_U^{m_U} a_{W,0}^{m_N}} \right). \end{aligned} \quad (17)$$

As a consequence, we can define the free energy of the j -state of the protein,

$$G_j = G_{W,j} + m_j k_B T \log \psi_j \quad (18)$$

where $G_{W,j}$ is the free Gibbs energy of the fully hydrated protein in the j state and $\psi_j = \phi_j/a_{W,0} = \phi_j/(\gamma_W X_{W,0})$. By using the known dependency of enthalpy on Gibbs free energy and of heat capacity on enthalpy, we derive the following expressions

$$H_j = -T^2 \left(\frac{\partial G_j/T}{\partial T} \right)_p = H_{Wj} - m_j k_B T^2 \frac{\psi_j'}{\psi_j}, \quad (19)$$

$$\begin{aligned} C_{pj} &= \left(\frac{\partial H_j}{\partial T} \right)_p \\ &= C_{pWj} - m_j k_B \left[T^2 \frac{\psi_j''}{\psi_j} - T \frac{\psi_j'(T\psi_j' - 2\psi_j)}{\psi_j^2} \right]. \end{aligned} \quad (20)$$

In order to calculate the first and second partial temperature derivative at constant pressure, ψ_j' and ψ_j'' , we need

to know the derivatives $X'_{j,0}$ and $X''_{j,0}$. These latter are calculated by numerically solving the first and the second derivative of the system of equations representing the mass balance reported in Eq. 14,

$$\begin{cases} X'_{W,0} + X'_N m_N \phi_N + X_N m_N \phi'_N + X'_U m_U \phi_U + X_U m_U \phi'_U = 0 \\ X'_{L,0} + X'_N m_N (1 - \phi_N) - X_N m_N \phi'_N + X'_U m_U (1 - \phi_U) - X_U m_U \phi'_U = 0 \end{cases} \quad (21)$$

$$\begin{cases} X''_{W,0} + X''_N m_N \phi_N + X_N m_N \phi''_N + X''_U m_U \phi_U + X_U m_U \phi''_U \\ + 2X'_N m_N \phi'_N + 2X'_U m_U \phi'_U = 0 \\ X''_{L,0} + X''_N m_N (1 - \phi_N) - X_N m_N \phi''_N + X''_U m_U (1 - \phi_U) - X_U m_U \phi''_U \\ - 2X'_N m_N \phi'_N - 2X'_U m_U \phi'_U = 0 \end{cases} \quad (22)$$

To note, the temperature derivatives at constant pressure of activity coefficients and equilibrium constants are given by the following expressions

$$\gamma'_j = -\gamma_j \frac{H_j^E}{k_B T^2} \quad (23)$$

$$\gamma''_j = \gamma_j \left[\left(\frac{H_j^E}{k_B T^2} \right)^2 - \frac{T C_{p,j}^E - 2H_j^E}{k_B T^3} \right] \quad (24)$$

$$K'_\alpha = K_\alpha \frac{\Delta H_\alpha}{k_B T^2} \quad (25)$$

$$K''_\alpha = K_\alpha \left[\left(\frac{\Delta H_\alpha}{k_B T^2} \right)^2 + \frac{T \Delta C_{p,\alpha} - 2\Delta H_\alpha}{k_B T^3} \right], \quad (26)$$

with

$$H_W^E = H^E - x_{L,0} \frac{dH^E}{dx_{L,0}} \quad (27)$$

$$H_L^E = H^E + (1 - x_{L,0}) \frac{dH^E}{dx_{L,0}} \quad (28)$$

$$C_{p,W}^E = C_p^E - x_{L,0} \frac{dC_p^E}{dx_{L,0}} \quad (29)$$

$$C_{p,L}^E = C_p^E + (1 - x_{L,0}) \frac{dC_p^E}{dx_{L,0}}. \quad (30)$$

Preferential solvation coefficients and excess solvation numbers

We complete the description of the model by reporting the equations developed by Schellman (1994) for calculating the *preferential solvation coefficient* (defined on a molar fraction scale) between the protein and the water or

cosolvent molecules in the solvent, $\Delta\Gamma_{Pj} \equiv \Gamma_{Ui} - \Gamma_{Nj}$ ($i = W, L$)

$$\Gamma_{jW} = m_j \frac{a_{W,0} K_j^{\text{ex}} - a_{L,0} X_{W,0} / X_{L,0}}{a_{L,0} + a_{W,0} K_j^{\text{ex}}} \quad (31)$$

$$\Gamma_{jL} = m_j \frac{a_{L,0} - a_{W,0} K_j^{\text{ex}} X_{L,0} / X_{W,0}}{a_{L,0} + a_{W,0} K_j^{\text{ex}}}. \quad (32)$$

These expressions have been found to be equivalent to the ones obtained with the Kirkwood-Buff (KB) theory (Kirkwood and Buff 1951). Finally, we report the expressions to calculate the *excess solvation numbers*. They represent the difference between the number of j -molecules in the solvation layer and the number of j -molecules in a volume of the bulk phase equal to the volume of the solvation layer,

$$N_{j,W} = m_j \{ \phi_j - x_{W,0} [\phi_j v_{W,j,s} + (1 - \phi_j) v_L] \} \quad (33)$$

$$N_{j,L} = m_j \{ (1 - \phi_j) - x_{L,0} [\phi_j v_{W,j,s} + (1 - \phi_j) v_L] \}. \quad (34)$$

v_L is the cosolvent molecular volume, supposed to be constant, and $v_{W,j,s}$ is the water molecular volume when on protein sites.

Differential scanning calorimetry

Differential scanning calorimetry (DSC) experiments allow the determination of the heat capacity at constant pressure of all the protein species in solution. According to the Schellman model described in the previous section, the total enthalpy referred to a protein molecule is written as a combination of the enthalpies of each protein state (19), weighted by the molar fraction of the state,

$$H = (X_N H_N + X_U H_U) / X_P \quad (35)$$

Hence the heat capacity of the whole protein system, which is the content of the DSC experiment, is defined by the temperature derivative at constant pressure of H ,

$$C_p = \left(\frac{\partial H}{\partial T} \right)_p = C_{p,W_N} + \frac{X_U}{X_P} \left[\Delta C_{p,W} + \frac{(\Delta H)^2}{(K + 1) k_B T^2} \right] \quad (36)$$

where $\Delta C_{p,W} = C_{p,W_U} - C_{p,W_N}$ and $\Delta H = H_U - H_N$. All the factors of this equation depend on system temperature and composition. Hence, by performing DSC experiments in a wide range of protein concentration as well as binary solvent compositions, it will be possible to fit the unique model expressed by Eq. 36 to all the data. According to this approach, we will obtain the unknown thermodynamic parameters of the system: the changes of reference enthalpy, reference entropy, heat capacity (and eventually its first temperature derivative) for all the equilibrium processes. They are: the two water–cosolvent exchange reactions over N and U protein states and the native–unfolding transition

in pure water. The outcome of this kind of analysis is noticeable because the integration in the model of the activity coefficients of water and cosolvent molecules, together with their temperature dependency. As shown, this information is given by the calorimetric measurements of excess thermodynamic quantities of the binary mixture in a wide range of composition: Gibbs free energy (G_o^E), enthalpy (H_o^E) and heat capacity (C_p^E) (Hansen and Miller 1954; Marcus 2000). Other relevant parameters of the model are the number of solvation sites m_N and m_U . The former can be estimated whenever the tertiary structure of the protein in the native state is known, although it is still challenging to predict where water molecules prefer to bind, and which of those water molecules are displaceable during a reaction (Ross et al. 2012). The number of solvation sites in the unfolding state is generally affected by a noticeable error, since the unfolding state of the protein is quite often a collection of several disordered conformations. Hence, in our studies m_U is included in the list of parameters that can be obtained by fitting the present model to DSC data.

Small-angle scattering

In the native state, the protein structure can be obtained by x-ray crystallography or NMR measurements. The atomic positions can be read in a Protein Data Bank (PDB) file (Berman et al. 2003). Different methods have been developed to calculate SAXS or SANS profiles from a PDB file (Svergun et al. 1995; Merzel and Smith 2002). Here we propose the SASMOL method, because it allows the determination of number and position of water molecules in the first three hydration shells (Ortore et al. 2009) and to calculate the protein form factor by taking into account the contribution of the hydration shells which can have different scattering length densities, according to the cosolvent distribution in the local domain. Concerning the unfolded protein state, we make the assumption of its monomeric feature, and we suppose to be able to describe its form factor by suitable statistical models, such as the worm-like chain (Pedersen and Schurtenberger 1996), or by an ensemble of atomic structures (PDB files) coming from computational techniques. Likewise the native protein state, in the framework of the preferential solvation, it is worth that the methods used to calculate the form factor of unfolded states are able to consider the scattering contribution of the local domain, which, due to preferential solvation mechanisms, can show a different composition from the one of the bulk solution. Although from a structural point of view we can foresee different unfolding conformations depending on binary solvent composition, in the framework of the present model we assume a unique free energy for all the unfolding states (18) and, consequently, a unique number of solvation sites on unfolded proteins m_U .

For the sake of simplicity, for any j -state, we assume that all the protein molecules show the most probable composition of the solvation layer, which, according to the binomial law (6), corresponds to $m_j\phi_j$ sites occupied by water molecules and $m_j(1 - \phi_j)$ sites occupied by cosolvent molecules. We also assume that, due to the known electrostriction effect, the molecular volume of water in the bulk, $v_{W,0}$, and in the first solvation shell of the protein j -state, $v_{W,j,s}$ (depicted in cyan and in blue, respectively, in Fig. 1), could be different. On the contrary, according to the assumption of single-contacts between cosolvent molecules and protein surface, the molecular volume of the cosolvent, v_L , is considered to be the same both in the bulk and in the first solvation shell (yellow and red spheres, respectively, in Fig. 1). As already discussed, the thickness, as well as the volume of the local domain ($\delta_{j,l}$ and $v_{P,j,l}$, respectively), should be usually considered larger than those of the first hydration shell. Therefore, the local domain water composition, $x_{W,j,l}$, has to be between ϕ_j and $x_{W,0}$. By assigning to the solution that occupies the volume obtained by the difference between $v_{P,j,l}$ and the volume of the first solvation shell, corresponding to $m_j[\phi_j v_{W,j,s} + (1 - \phi_j)v_L]$, the same water molar fraction of the bulk solvent ($x_{W,0}$), we can derive the following equations for the water composition in the local domain and its average molecular volume:

$$x_{W,j,l} = \frac{[v_{P,j,l} + m_j(v_{W,0} - v_{W,j,s})\phi_j - m_j v_L]x_{W,0} + m_j v_L \phi_j}{m_j(v_{W,0} - v_L)x_{W,0} + v_{P,j,l} + m_j(v_L - v_{W,j,s})\phi_j} \quad (37)$$

$$v_{W,j,l} = \frac{\phi_j(x_{W,0} - x_{W,j,l})}{x_{W,j,l}(x_{W,0} - \phi_j)}(v_{W,j,s} - v_{W,0}) + v_{W,0} \quad (38)$$

Values of $v_{W,0}$ and v_L and their dependency on $x_{W,0}$ are easily found in the literature concerning the physical chemistry properties of binary mixtures (Lide 1996). In the SAS theory, it is also important to deal with the number density of any scattering particle in solution. In the present case, the number density of the protein in the j -state is the ratio $n_{P,j} = X_j / \langle v \rangle$, $\langle v \rangle$ being the average molecular volume calculated over any species in the system, according to

$$\begin{aligned} \langle v \rangle = & X_{W,0}v_{W,0} + X_{L,0}v_L \\ & + X_N\{v_{P,N} + m_N[\phi_N v_{W,N,s} + (1 - \phi_N)v_L]\} \\ & + X_U\{v_{P,U} + m_U[\phi_U v_{W,U,s} + (1 - \phi_U)v_L]\}. \end{aligned} \quad (39)$$

In the last equation, $v_{P,j}$ is the molecular volume of the protein core in the j -state.

SAS experiments provide the measurement of the *macroscopic differential scattering cross section*, $\frac{d\Sigma}{d\Omega}(q)$, of the system, which depends on the modulus q of the scattering vector, $q = 4\pi \sin \theta / \lambda$, 2θ being the scattering angle and λ the wavelength of the monochromatic x-ray or neutron

beam. After proper calibration and buffer subtraction, in the most general case when both native and unfolded protein could exist in the solution, $\frac{d\Sigma}{d\Omega}(q)$ is given by

$$\frac{d\Sigma}{d\Omega}(q) = \sum_{j=N,U} n_{P,j} P_j(q) + \sum_{j=N,U} \sum_{j'=N,U} (n_{P,j} n_{P,j'})^{1/2} P_{1j}^*(q) P_{1j'}(q) S_{jj'}(q) + B \quad (40)$$

B being a flat background which takes into account incoherent scattering effects, particularly relevant in SANS experiments.

Form factors

In Eq. 40, $P_{1j}(q) = \langle A_j(\mathbf{q}) \rangle_{\omega_q}$ is the orientational average (represented by the integration over the polar angles ω_q of the scattering vector $\mathbf{q}$) of the scattering amplitude of the protein in the j -state, whereas $P_j(q) = \langle |A_j(\mathbf{q})|^2 \rangle_{\omega_q}$, also known as *form factor*, is the orientational average of the squared modulus of the its scattering amplitude. In the SASMOL method here adopted to calculate the functions $P_{1j}(q)$ and $P_j(q)$ for the native protein state, the scattering amplitude $A_j(\mathbf{q})$ is calculated by taking into account the scattering properties of any atom of the protein. In the case of x-rays, the atomic structure factors are expressed as an expansion of atomic radial electron densities (Waasmaier and Kirfel 1995). Instead for neutrons, the coherent scattering length b_i of each atom is considered, and in the case of partially/fully deuterated proteins dissolved in a binary solvent that could contain deuterated water and/or cosolvent molecules, the proton–deuteron exchange process over the acidic hydrogen atoms of the proteins is included, too (Jacrot 1976). The calculation of the scattering amplitude of the protein also takes into account the amplitude of the displaced binary solvent and the one of the local domain. These amplitudes are depending on the *scattering length densities*, which are functions of the bulk binary solvent and of the local domain composition according to

$$\rho_0 = \frac{x_{W,0}(b_W - b_L) + b_L}{x_{W,0}(\nu_{W,0} - \nu_L) + \nu_L} \quad (41)$$

$$\rho_{W,j,l} = \frac{x_{W,j,l}(b_W - b_L) + b_L}{x_{W,j,l}(\nu_{W,j,l} - \nu_L) + \nu_L} \quad (42)$$

where b_j is the scattering length of the j molecule ($j = W, L$). For SAXS b_j depends on the number of electrons of the molecule, $b_j = r_e n_{je}$, $r_e = 0.28 \cdot 10^{-12}$ cm being the classical electron radius. For SANS, b_j is calculated by summing the atomic coherent neutron scattering lengths. If all or a part of acidic hydrogen atoms of the different molecules are deuterated, it can be assumed that due to the rapid H/D exchange, at the equilibrium the *deuteration*

grade of each acidic hydrogen corresponds to the nominal value, calculated according to

$$x_D \equiv \frac{\mathcal{N}_{D+}}{\mathcal{N}_{H+} + \mathcal{N}_{D+}} \quad (43)$$

$\mathcal{N}_{D+}$ and $\mathcal{N}_{H+}$ being the moles of all the exchangeable deuterons and protons of the system. By indicating with $b_{j,D}$ and $b_{j,H}$ the scattering lengths of the j -molecule when all its exchangeable hydrogens are deuterons or protons, respectively, the scattering length as a function of the deuteration grade can be calculated according to Jacrot (1976) and Dianoux and Lander (2003)

$$b_j = b_{j,D} x_D + b_{j,H} (1 - x_D). \quad (44)$$

SAS profiles of unfolded proteins can be calculated by an average over conformational ensembles coming from molecular dynamics or, more in general, from molecular modelling calculations. In such cases SASMOL can be applied to determine the functions $P(q)$ and $P_1(q)$ of each structure. An other way to study unfolded proteins is to adopt a statistical approach, such as the worm-like model (Pedersen and Schurtenberger 1996). In the latter case, we need to express the average scattering length density of the core protein, according to

$$\rho_{P,j} = \frac{b_P}{\nu_{P,j}}. \quad (45)$$

Structure factors

In the general expression of $\frac{d\Sigma}{d\Omega}(q)$ (40), the terms $S_{jj'}(q)$ are the *particle-particle structure factors*: they depend on the average positions into the space of the pair of molecules jj' . Formally, $S_{jj'}(q)$ is the isotropic Fourier transform of the *radial correlation function* $g_{jj'}(r)$, whose features depend on the potential energy $u_{jj'}(r)$ between the two species j and j' . Different approximated methods have been developed to calculate the radial correlation functions from the potential energies. Most of them are able to deal with monodisperse systems, in dilute or concentrate conditions (Liu et al. 2005; Broccio et al. 2006; Chen et al. 2007; Liu et al. 2011). In the case of polydisperse systems, such as in the presence of both native and unfolded protein states, a possible strategy is to write the $g_{jj'}(r)$ as a power series of the total protein number density $n_P = n_{P,N} + n_{P,U}$ (Spinozzi et al. 2002). However this method has been proved to be valid only at low or moderate protein concentrations, roughly when the whole volume fraction η of protein matter is lower than 0.1. A simpler way to deal with a polydisperse system, also valid at higher protein concentrations, is to express a unique effective protein–protein structure factor $S(q)$, i.e. by considering a unique potential energy between

two protein molecules, independently on their folding state. Following this idea, Eq. 40 can be re-formulated as

$$\frac{d\Sigma}{d\Omega}(q) = n_P P(q) S_M(q) + B \quad (46)$$

$$P(q) = \frac{1}{n_P} \sum_{j=N,U} n_{P,j} P_j(q) \quad (47)$$

$$S_M(q) = 1 + \frac{[P_{1\text{avg}}(q)]^2}{P_{\text{avg}}(q)} [S(q) - 1] \quad (48)$$

In these equations, $P(q)$ represents the average form factor of the system, depending on native and unfolding protein number densities. The function $S_M(q)$ is called the *measured* structure factor: it depends on the average particle–particle form factor $S(q)$ and on the orientational averages $P_{1\text{avg}}(q)$ and $P_{\text{avg}}(q)$ of the *mean-particle* that describes the isotropic particle–particle interaction. Here, we assume that this mean-particle is an homogeneous cylinder (Oliveira et al. 2009) with radius R and length H ,

$$P_{1\text{avg}}(q) = \int_0^{\pi/2} d\beta \sin \beta A_{\text{avg}}(q) \quad (49)$$

$$P_{\text{avg}}(q) = \int_0^{\pi/2} d\beta \sin \beta [A_{\text{avg}}(q)]^2 \quad (50)$$

$$A_{\text{avg}}(q) = 2\pi R^2 H \frac{\sin\left(\frac{1}{2}qH \cos \beta\right)}{\frac{1}{2}qH \cos \beta} \frac{J_1(qR \sin \beta)}{qR \sin \beta} \quad (51)$$

where $J_1(x)$ is the 1st Bessel function of integer order.

Protein-protein potential

The unique and average protein–protein potential $u(r)$ is in general written as a combination of three terms, $u(r) = u_0(r) + u_1(r) + u_2(r)$, where $u_0(r)$ is the hard sphere potential,

$$u_0(r) = \begin{cases} \infty & r \leq 2R \\ 0 & r > 2R \end{cases} \quad (52)$$

and the other two terms have a Yukawian form,

$$u_j(r) = A_j \frac{\exp[-\kappa_j(r - 2R)]}{r} \quad (53)$$

In all these terms, R represents the *effective* protein radius. The Yukawian term with $j = 1$ is the screened Coulombian potential. The constant A_1 depends on the protein charge z_P and on the dielectric constant ε of the bulk, in general a function of the bulk composition $x_{W,0}$:

$$A_1 = \frac{z_P^2 q_e^2}{4\pi \varepsilon_0 \varepsilon (1 + \kappa_1 R)^2}, \quad (54)$$

q_e being the absolute electron charge and ε_0 the vacuum dielectric permittivity. In Eq. 54, κ_1 is the *inverse Debye screening length*, function of the ionic strength I_S of the

solution (calculated over all microions, e.g. buffer, added salts and protein counterions):

$$\kappa_1 = q_e \left(\frac{2N_A I_S}{k_B T \varepsilon_0 \varepsilon} \right)^{1/2} \quad (55)$$

N_A being Avogadro's number. The second Yukawian term ($j = 2$) indicates the attractive potential. The constant A_2 is written in terms of the value J of the potential at protein contact ($r = 2R$)

$$A_2 = -2R J \quad (56)$$

and the decay constant depends on the *range* d of the attractive interaction, according to $\kappa_2 = 1/d$.

Approximate solutions of the Ornstein-Zernike equation

To calculate the function $S(q)$ it is necessary to solve the Ornstein-Zernike (OZ) integral equations by using proper approximate closure relationship. A relevant approximation for a two Yukawian potentials within the mean spherical approximation (MSA) has been proposed by Liu et al. (2005). Here, we follow a simple scheme (Narayanan and Liu 2003), by writing the perturbation of the structure factor $S_0(q)$ obtained with the well-known Percus-Yevick (PY) approximation in the framework of the random phase approximation (RPA),

$$S(q) = \frac{S_0(q)}{1 + n_P S_0(q) [U_1(q) + U_2(q)] / (k_B T)} \quad (57)$$

$$S_0(q) = \left\{ 1 - \frac{12\eta[\eta(3 - \eta^2) - 2]}{(1 - \eta)^4} \frac{j_1(2qR)}{2qR} \right\}^{-1} \quad (58)$$

$j_1(x)$ being the 1st-order spherical Bessel function. The volume fraction η can be written in terms of protein number density and effective protein radius, $\eta = \frac{4\pi}{3} R^3 n_P$. The functions $U_j(q)$ are the Fourier transforms of $u_j(r)$, which can be analytically derived:

$$U_j(q) = 4\pi A_j \frac{\kappa_j \sin(2qR) + q \cos(2qR)}{q(q^2 + \kappa_j^2)}. \quad (59)$$

Global fit strategy

We have built a model that combines thermodynamic and structural characteristics of a protein molecule dissolved in a binary solvent. On the basis of the thermodynamic functions that govern the unfolding process and the exchange of solvent molecules over the surfaces of native and unfolded protein states, the model is able to predict how the concentrations of both protein states vary with temperature and with solvent composition. Two complementary experiments can be described by the model: DSC and SAS, by using

both x-rays or neutrons. All thermodynamic parameters of the model can be optimized to fit both DSC and SAS experimental data. The extra availability of DSC data can restrict the field of variability of the thermodynamic parameters that are equally able to fit SAS data. On the other hand, SAS data provide information on the numbers of solvation sites over native and unfolding protein surfaces, which are important values to be included in the analysis of DSC thermograms, too. Hence, by applying the same model, a unique global fit of DSC and SAS data collected by varying protein concentration, binary solvent compositions and temperatures, can provide in a robust way the values of all the relevant parameters. To note, some structural parameters, such as protein size and shape or protein charge, are uniquely needed to analyze SAS data. It is also important to notice that we applied the model in the case of SAS data collected at ambient or physiological temperature, when the protein is native. In this case, only the thermodynamic parameters that describe the exchange equilibrium on the native protein surface are shared between DSC and SAS data. The main drawback of the model is probably the assumption that the protein unfolded state does not substantially change with solvent composition. However, from a structural point of view the form factor of the unfolded proteins can be free to change with the composition. On the contrary, the thermodynamic feature of the unfolding state, mainly the number of solvation sites over the unfolded surface (m_U) should be considered a constant, otherwise the model cannot be applied. Another limitation of the model is the absence of any excluded volume effect, which has recently been proved to be quite relevant in determining protein stabilization in the presence of osmolytes (Graziano 2011, 2012). A further limitation of the model could be related to the description of the interactions among proteins by a unique average structure factor, even at different solvent compositions. However, since the structure factor parameters describing the potential energy are not shared by DSC models, it is possible to assume that their values are not fixed, but could vary with temperature and/or with protein concentration as well as with binary solvent compositions. The model can also be applied to study only DSC data or only SAS data. In the former case, the number of solvation sites in the native state of the protein should be considered as input parameter, as already discussed, whereas the number of solvation sites in the unfolded states can be taken as a fitting parameters. When applied to analyze only SAS data, the model will allow the determination of a unique value of the thermodynamic exchange constant, which would be able to describe SAS data taken at different binary solvent compositions. To improve the quality of fitting results, it is more desirable to use SANS with a wise use of deuterated and protonated water and cosolvent molecules in order to highlight the differences in the scattering length density of core protein,

protein local domain and bulk solvent as a function of the preferential solvation (Ortore et al. 2009).

The present method is included in the GENFIT software (Spinozzi et al. 2014), a general tool for analyzing SAS data. Data regarding the calorimetric functions of the binary mixture should be provided by means of sub-routines written in Fortran. The main feature of GENFIT is the possibility to analyze many experimental data by using both single experiment fitting parameters and common fitting parameters.

Applications

In this section, we review some recent applications of the theoretical method here described to study the protein solvation properties in binary solutions using SAS and DSC techniques.

Lysozyme in water–glycerol solutions

It is well known that cosolvents such as sugars and, more in general, poly-ols, could stabilize proteins in water solution against thermal or chemical denaturation (Gekko 1982; Yancey et al. 1982; Zou et al. 2002). The most studied system is the one of proteins in water–glycerol solution. Previous investigations have suggested that glycerol is mainly excluded from the protein surface, providing a reinforcement of the water shell around the proteins (Gekko and Timasheff 1981; Lehmann and Zaccai 1984; Xie and Timasheff 1997). The case of lysozyme, which is considered a model protein widely investigated with many theoretical and experimental techniques, is astonishing: even in nearly anhydrous glycerol this enzyme is able to maintain its biological activity (Rariy and Klivanov 1997). To understand direct or indirect molecular mechanisms behind the stabilizing role of glycerol on lysozyme many studies had been performed, including neutron scattering dynamics (Paciaroni et al. 2002; Paciaroni et al. 2003) and calorimetry (Cinelli et al. 2004).

Here we focus on two studies based on the Schellman model recently published (Sinibaldi et al. 2007; Spinozzi et al. 2008). In the first work, the SANS technique has been chosen as suitable tool to investigate preferential solvation of lysozyme water–glycerol solutions. To enhance the effect of accumulation or depletion of cosolvent molecule in the local domain, samples were prepared using pure light and heavy water and hydrogenated and deuterated glycerol. A total of 35 samples at room temperature and pH 4.5 were investigated, with protein weight concentration, c , varying between 30 and 150 g L⁻¹, nominal water molar fraction in the solvent, x_W , between 0.4 and 0.6 and three conditions of deuteration grade, x_D , namely 0, 0.05 and 0.35.

SANS measurements were performed at Forschungszentrum (FZJ) in Jülich (Germany) using the KWS1 diffractometer. Experimental curves together with their best fitting obtained with the GENFIT software are shown in Fig. 2. The interference peak responsible for protein–protein interactions is clearly visible and its features change with protein concentration. All SANS curves have been analyzed with a unique fit based on the exchange model and with the assumption that lysozyme is in the native state. The PDB entry 6LYZ (Diamond 1974) was used to calculate the form factor. In this first study, the not-ideal properties of the binary mixture water–glycerol were not considered, fixing to unity all the activity coefficients. The main fitting parameters are reported in Table 1. The thickness of the local domain $\delta_{P,N}$ turned out to be higher than 3 Å, the typical dimension of the first hydration layer, indicating that the model is able to depict the presence of glycerol molecules bound to the protein surface. The net protein charge $z_P \approx 9 q_e$ is in fully agreement with the value calculated by titration curve (Tanford and Roxby 1972). The volume of water molecules in the solvation sites, $\nu_{W,N,s}$ is lower than $\sim 30 \text{ Å}^3$, the typical value in the bulk, confirming electrostriction effects at the protein surface. The parameters of the attractive potential J and d (values not reported in Table 1) show a linear decrease with water molar fraction in the binary solvent, indicating that glycerol can prevent protein–protein aggregation (Farnum and Zukoski 1999). The main result of this study is the determination of the exchange constant: the obtained value $K_N^{\text{ex}} = 1.87 \pm 0.03$ confirms that

Table 1 Main fitting parameters of the analysis of SANS curves of lysozyme in water–glycerol solutions (Sinibaldi et al. 2007)

$\delta_{P,N}$ (Å)	z_P (e)	$\nu_{W,N,s}$ (Å ³)	K_N^{ex}
5.83 ± 0.04	9.1 ± 0.3	28.81 ± 0.04	1.87 ± 0.03

$\delta_{P,N}$, thickness of the local domain; z_P , protein charge. $\nu_{W,N,s}$, volume of water molecules in the solvation sites; K_N^{ex} , thermodynamic exchange constant

the first hydration shell of lysozyme is enriched of water and gives a quantitative estimation of this phenomenon. As a further confirmation of this results, we show in Fig. 3 that the water molar fraction in the local domain is always higher than the nominal molar fraction of water in the binary solvent. From these results, the values of the excess solvation numbers were calculated, too, and they are reported in Fig. 4. The excess solvation numbers can be compared with literature data, and clearly evidence the presence of a maximum of hydration that corresponds to the maximum protein stability and that can be explained as the best compromise between the surface affinity with water and the availability of water in solution.

In a second work (Spinuzzi et al. 2008), we investigated samples of lysozyme in water–glycerol solvents by using only DSC. Protein concentration was 1 w/w%, whereas the water molar fraction in the solvent, x_W , spans a wide range from ≈ 0 to 1.0. DSC experiments were achieved in a

Fig. 2 SANS data of lysozyme in water–glycerol solutions obtained at different protein concentrations (reported in g L^{-1} above each group of curves), water molar fraction (x_W) and deuteration grade (x_D). For the sake of clarity, each curve has been scaled by a factor multiple of 0.025 cm^{-1} (Sinibaldi et al. 2007)

Fig. 3 Dependence on the nominal water molar fraction in the solvent x_W of the water molar fraction in the local domain $x_{W,N,i}$ calculated by Eq. 37 at the minimum and the maximum protein concentration c . The dotted line representing the thermodynamic exchange constant $K_j^{\text{ex}} = 1$ case is also reported (Sinibaldi et al. 2007)

micro-DSCII (Setaram, France) at a scan rate of 1 K/min. All thermograms have been fitted with a unique analysis, carried out by means of GENFIT, based on Eq. 36. To apply the Schellman model in a wide range of compositions of the not ideal water–glycerol mixture it is imperative to deal with the excess thermodynamic functions. Here we have used the analytic functions that express the dependency of G_o^E , H_o^E and C_p^E on the glycerol molar function in the bulk

Fig. 4 Dependence on the nominal water molar fraction in the solvent (x_W) of the excess solvation number $N_{N,i}$ in the first solvation layer of lysozyme in water–glycerol solutions. Curves have been calculated with Eqs. 33 and 34 using the parameters obtained by fitting analysis: blue and red lines refer to water ($j = W$) and glycerol ($j = L$), respectively. Open symbols indicate all our investigated experimental points (Sinibaldi et al. 2007). Close symbols refer to data relative to the ribonuclease A, reported in Refs. (Gekko and Timasheff 1981; Shulgin and Ruckenstein 2005)

$x_{L,0}$ reported by Marcus (2000). By reference to the results reported in the first work (Sinibaldi et al. 2007), the number of sites for the native state of lysozyme was fixed to

Fig. 5 DSC thermal scans of lysozyme in water–glycerol mixtures at different nominal water molar fraction x_W . Circles correspond to the experimental data; solid lines represent the profiles obtained by the GENFIT analysis based on the solvent exchange model (Spinozzi et al. 2008)

$m_N = 385$. The heat capacities of the fully hydrated protein, C_{pWj} , were expanded up the first Taylor order (13), whereas it was sufficient to use the only zero-order term ($\Delta C_{p\alpha}^o$) for the heat capacity changes relative to exchange processes over native and unfolded lysozyme. Experimental data together with the optimum fit are reported in Fig. 5: it can be observed that the trend of the thermograms, including the position and the height of the peaks, is completely reproduced by the model in the full range of temperature. The few global fitting parameters, able to explain all DSC data, are reported in Table 2. Results of $\Delta G_N^{\text{ex}o}$ allow to calculate the exchange equilibrium constants at the reference temperature $T_o = 298$ K. For the native state we obtain $K_N^{\text{ex}o} = 1.89 \pm 0.02$, in almost perfect agreement with the value of 1.87 ± 0.03 (see Table 1) obtained with the SANS study previously described. For the unfolded state we found a slightly lower value, $K_U^{\text{ex}o} = 1.75 \pm 0.01$, indicating that the unfolded surface is preferentially hydrated with a preference lower than the one of the native state. As expected, the number of water–glycerol solvation sites in the unfolding state, $m_U = 425.5 \pm 0.8$, is larger than that determined in the native state: after unfolding the protein exposes an higher surface toward the solvent. To complete the picture of the obtained results, we report in Fig. 6 the excess solvation

Fig. 6 Dependence on the nominal water molar fraction x_W of the excess solvation number $N_{j,i}$ (top frame) and its change upon lysozyme unfolding (lower frame). In the upper frame, open, and close symbols refer to experimental investigated points for lysozyme in the $j = N$ and $j = U$ states, respectively. In both frames, blue and red lines refer to water ($j = W$) and glycerol ($j = L$), respectively

Table 2 Global-fitting parameters obtained by the GENFIT analysis of DSC data for lysozyme in water–glycerol mixtures (Spinozzi et al. 2008)

ΔH_W^o (kJ mol ⁻¹)	ΔS_W^o (J mol ⁻¹ K ⁻¹)	ΔG_W^o ^a (kJ mol ⁻¹)	
200 ± 40	500 ± 100	47 ± 3	
C_{pWU}^o (kJ mol ⁻¹ K ⁻¹)	C_{pWN1}^o (J mol ⁻¹ K ⁻²)	C_{pWU1}^o (J mol ⁻¹ K ⁻²)	m_U
6 ± 2	5 ± 3	20 ± 15	425.5 ± 0.8
$\Delta H_N^{\text{ex}o}$ (kJ mol ⁻¹)	$\Delta S_N^{\text{ex}o}$ (J mol ⁻¹ K ⁻¹)	$\Delta G_N^{\text{ex}o}$ ^a (kJ mol ⁻¹)	$\Delta C_{pN}^{\text{ex}o}$ (J mol ⁻¹ K ⁻¹)
-0.9 ± 0.2	2.2 ± 0.7	-1.57 ± 0.02	2 ± 3
$\Delta H_U^{\text{ex}o}$ (kJ mol ⁻¹)	$\Delta S_U^{\text{ex}o}$ (J mol ⁻¹ K ⁻¹)	$\Delta G_U^{\text{ex}o}$ ^a (kJ mol ⁻¹)	$\Delta C_{pU}^{\text{ex}o}$ (J mol ⁻¹ K ⁻¹)
-0.8 ± 0.2	2.0 ± 0.8	-1.39 ± 0.02	8 ± 2

ΔH_W^o , ΔS_W^o and ΔG_W^o : unfolding enthalpy, entropy, and Gibbs free energy of lysozyme in pure water at the reference temperature; C_{pWU}^o , C_{pWN1}^o and C_{pWU1}^o : heat capacity of lysozyme in the unfolding state and first temperature derivatives of the heat capacity of lysozyme in the native and in the unfolding states; $\Delta H_N^{\text{ex}o}$, $\Delta S_N^{\text{ex}o}$, $\Delta G_N^{\text{ex}o}$ and $\Delta C_{pN}^{\text{ex}o}$: changes of enthalpy, entropy, and Gibbs free energy at the reference temperature and of the heat capacity for the glycerol–water exchange in the solvation sites of lysozyme in the native state; $\Delta H_U^{\text{ex}o}$, $\Delta S_U^{\text{ex}o}$, $\Delta G_U^{\text{ex}o}$ and $\Delta C_{pU}^{\text{ex}o}$: changes of enthalpy, entropy and Gibbs free energy at the reference temperature and of the heat capacity for the glycerol–water exchange in the solvation sites of lysozyme in the unfolded state

^aderived parameters

number for lysozyme in the two states and its change occurring at the native–unfolding transition at T_o . For the native state, results are in a good agreement with the one obtained with the SANS study (see Fig. 4). We also see that, before $x_W \approx 0.6$, in the N-shell there are ≈ 5 water molecules more than in the U-shell and that this effect reverts after $x_W \approx 0.6$, showing a stronger preferential hydration effect on the U-state with its maximum at $x_W = 0.87$.

Lysozyme in water–urea solutions

The solvation properties of lysozyme have been investigated by means of SANS in the presence of low concentrations of urea, one the most widely used chemical denaturant, with the aim to shed light on the molecular mechanism that triggers the protein unfolding and on the modifications of the protein–protein interactions in the presence of urea (Ortore et al. 2008). Urea concentrations were much lower than those required for a complete denaturation (Timasheff and Xie 2003). Previous investigations based on vapor pressure osmometry showed that the urea concentration in the vicinity of lysozyme surface is higher than the one in the bulk solvent (Timasheff and Xie 2003; Shulgin and Ruckenstein 2005). Other studies based on SAXS results suggested a lower short range attraction of proteins when dissolved in water–urea solutions (Niebuhr and Koch 2005).

According to the exchange solvent model shown in the theoretical section, we have investigated SANS spectra corresponding to samples of lysozyme at weight concentration varying from 30 to 120 g L⁻¹ in 100 mM phosphate buffer (pH 7.0), at four urea concentrations comprised between 0 and 3 M (nominal water molar fraction x_W ranging from 0.94 to 1.00). A unique deuteration grade $x_D = 0.80$ was chosen in order to get optimum scattering length contrasts between water, urea and lysozyme. SANS measurements were recorded at the Hahn-Meitner-Institut in Berlin (Germany), using the V4 instrument. Since the range of x_W is quite narrow, we treated the water–urea binary solvent as an ideal solution, by fixing to unity all the activity coefficients.

All SANS curves were analyzed with a global fit by using the GENFIT software (Spinozzi et al. 2014). A preliminary analysis of SANS curves based on the Kratky representation (Glatter and Kratky 1982) showed that protein molecules in solution were globular particles, confirming that, at the low concentration of urea adopted in this study, lysozyme can be still considered in its native state. As a consequence, even in this case a unique form factor of the protein, based on the 6LYZ PDB entry (Diamond 1974), has been adopted to describe its native state. The main objective of the global-fit analysis was to get from the whole set of data the unique urea–water exchange thermodynamic constant on the surface of lysozyme in the native state. In order to highlight possible variations of the protein–protein attractive potential, we have considered a unique range d of the Yukavian potential and a depth J (56) that varies with water molar fraction in the solvent through a simple linear trend, $J = J_0 + J_m x_W$. On the contrary, based on preliminary analyses of SANS data and on preliminary ζ -potential experiments, we have considered as a free fitting parameter the net protein charge z_P , whose values can change with both protein and urea concentration.

SANS data and corresponding best fits are reported in Fig. 7. To note, the high quality of the fits, in the whole q -range: the interaction peak and its modifications with protein and urea concentration are always well reproduced by the model. The main fitting parameters resulting from the global-fit are reported in Table 3. The net charge is reported as a function of urea concentration in Fig. 8. The exchange thermodynamic constant results 0.52 ± 0.08 : it clearly indicates that urea is preferentially bound to the protein surface, independently on protein concentration. This is the basic result of this study, in full agreement with results obtained at infinite protein dilution, but obtained at finite protein concentration, hence in a condition similar to that found *in vivo*. Also the molecular volume of water in contact with the protein surface ($v_{W,N,s}$) is found to be smaller than the water volume in the bulk, in agreement with previous works (Svergun et al. 1998; Sinibaldi et al. 2007). Concerning the behavior of the attractive potential, we have found

Fig. 7 Experimental SANS data of lysozyme solutions with different urea contents, reported above each group of curves, and protein concentration, as indicated on the right (in g L⁻¹ units), respectively. Solid lines correspond to fitting curves obtained with GENFIT. For the sake of clarity, one experimental point every two has been shown and each curve has been scaled by a factor multiple of 0.04 cm⁻¹ (Ortore et al. 2008)

a positive value of J_m , indicating that the strength of the attraction decreases by adding urea in solution, hence confirming the literature result (Niebuhr and Koch 2005). The modulus of the protein charge (Fig. 8) shows little variations around 5 q_e . However, data indicate the presence of a minimum at 1 M urea followed by a small increase. Finally we report in Fig. 9 the excess solvation numbers determined by our method and the ones derived in literature (Timasheff and Xie 2003; Shulgin and Ruckenstein 2005). It can be clearly seen that urea preferentially expels water molecules from the protein surface (see also BSA case (Sinibaldi et al. 2008)). We observe that the excess solvation number of water does not vanish at zero urea concentration, an effect related to the lower molecular volume of water in the local domain (Table 3). For the sake of completeness, in Fig. 9 data calculated with the assumption that $v_{W,N,s} = v_{W,0}$ are also shown (dashed line): they are in full agreement with literature results that have been obtained without considering any variation of the molecular volume of water upon binding with protein surface.

Lysozyme in water–ethanol solutions

The last example regards a combined study of DSC and SANS to investigate the preferential solvation of lysozyme in the presence of small amount of ethanol (Ortore et al. 2011). The effect of monohydric alcohols, in particular ethanol, on water protein solutions has widely been

Table 3 Main fitting parameters of the analysis of SANS curves of lysozyme in water–urea solutions

$\delta_{N,l}$ (Å)	$\nu_{W,N,s}$ (Å ³)	K_N^{ex}	J_0 ($k_B T$)	J_m ($k_B T$)	d (Å)
3.8 ± 0.4	27.5 ± 0.5	0.52 ± 0.08	-7.7 ± 0.3	12 ± 2	2.5 ± 0.5

$\delta_{N,l}$, thickness of the local domain; $\nu_{W,N,s}$ volume of water molecules in the solvation sites; K_N^{ex} , thermodynamic exchange constant; J_0 and J_m : intercept and slope of the linear dependency of the depth of the Yukavian attractive potential J as a function of the nominal water molar fraction x_W ; d range of the Yukavian attractive potential (Ortore et al. 2008)

investigated with experimental (Onori and Santucci 1996; Wakisaka and Ohki 2005; Soper et al. 2006; Guo et al. 2003) and theoretical approaches (Suresh and Naik 2002; Allison et al. 2005; Zhang and Yang 2005; Roney et al. 2004). As a result, micro-heterogeneous structures of alcohol–water binary mixtures at the molecular level became to be generally accepted and the relationships with protein stability recognized.

The Schellman model has been applied to water–ethanol solutions of lysozyme in order to quantify the preferential binding, if any, of ethanol on the protein surface. The well-known not ideality of this binary mixture imposes the use of activity coefficients and, in general, of all the excess thermodynamic functions. Hence, we have integrated in the model the dependence on the molar fraction of ethanol in the bulk ($x_{L,0}$) of excess Gibbs free energy, excess enthalpy, excess heat capacity and excess volume by means of analytic expression available in literature (Hansen and Miller 1954; Boyne and Williamson 1967; Golier and Wilhelm 1981; Gonzalez et al. 2007).

SANS measurements were performed at the Laboratoire Léon Brillouin (LLB), Saclay (France), on the PAXE

Fig. 8 Modulus of protein effective charge determined from SANS analysis as a function of urea concentration for different concentrations of lysozyme (Ortore et al. 2008)

Fig. 9 Water and urea excess numbers calculated on the basis of our fitted parameters (solid lines (Ortore et al. 2008)) and compared with literature data derived from the preferential binding coefficients (Shulgin and Ruckenstein 2005). Triangles refer to urea, squares to water. Dotted lines represent the calculated excess numbers of the species j considering the density of water in the local domain to be equal to that in the bulk solvent, as assumed in Shulgin and Ruckenstein (2005)

diffractometer. Lysozyme weight concentration c was varied from 30 to 120 g L^{-1} at different values of the nominal molar fraction of ethanol (x_L) from 0 to 0.09. To highlight the SANS contrast among the molecular species in solution, all the samples have been prepared at a unique deuteration grade $x_D = 0.88$. All SANS data have been recorded at room temperature.

Fig. 10 SANS curves of lysozyme in water–ethanol mixtures at different compositions, as shown by the x_L values reported in the bottom of each column and by the protein concentration reported on the right of each curve in g L^{-1} unit. Solid lines are the best fits obtained by the global fit analysis. Curve are scaled by a factor 0.1 cm^{-1} for clarity (Ortore et al. 2011)

Fig. 11 DSC thermal scans of lysozyme in water–ethanol mixtures at different compositions, as shown by the indicated nominal molar fraction of ethanol in the mixture, x_L . Solid lines represent the profiles obtained by the global fit analysis. Curves are scaled by 40 kJ mol^{−1} K^{−1} for clarity (Ortore et al. 2011)

DSC experiments were carried out on a micro-DSC II (Setaram, France). Protein concentration was fixed to 20 g L^{−1} and, like in the SANS case, x_L ranged from 0 to 0.09.

Kratky plots of SANS data reveal the only presence of lysozyme in globular state, confirming that the low content of ethanol is not able to induce partial or total unfolding of the protein. Hence, as in the previous cases, the form factor of lysozyme has been calculated with the SASMOL method (included in the GENFIT software). Both kinds of experiments have been analyzed with a global fit method via the GENFIT software. Fitting parameters have been divided in three classes. i) Parameters common to both kinds of experiments and that do not depends on sample conditions. The only parameter that falls in this class is the ethanol–water exchange thermodynamic constant at the reference temperature $K_N^{\text{ex}^\circ}$. ii) Parameters that depend on the type of experiment but are not depending on sample conditions. For example, all the thermodynamics functions that allow to fit DSC data according to Eq. 36 belong to this class. iii)

Table 4 First, second and third class global fit parameters of SANS and DSC data of lysozyme dissolved in water–ethanol solutions

First class parameters

$$K_N^{\text{ex}^\circ}$$

$$m_N$$

$$2.67 \pm 0.04$$

Second class parameters and average third class parameters, SANS

$\langle z_p \rangle$ (q_e)	$\langle d \rangle$ (Å)	$\nu_{W,N,s}$ (Å ³)
16.0 ± 0.3	3.0 ± 0.4	29.3 ± 0.1

Third class parameters, SANS

$\langle J \rangle_{x_L=0.00}$ (kJ mol ^{−1})	$\langle J \rangle_{x_L=0.03}$ (kJ mol ^{−1})	$\langle J \rangle_{x_L=0.06}$ (kJ mol ^{−1})	$\langle J \rangle_{x_L=0.09}$ (kJ mol ^{−1})
12.2 ± 0.4	11.8 ± 0.3	11.4 ± 0.3	10.9 ± 0.4

Second class parameters, DSC

$K_U^{\text{ex}^\circ}$ (kJ mol ^{−1})	ΔH_W° (kJ mol ^{−1})	ΔG_W° (kJ mol ^{−1})	m_U (J mol ^{−1} K ^{−2})
2.54 ± 0.02	270 ± 10	48 ± 1	418 ± 2
C_{pWU}° (kJ mol ^{−1} K ^{−1})	C_{pWNI} (J mol ^{−1} K ^{−2})	C_{pWU1} (J mol ^{−1} K ^{−2})	
7.6 ± 0.6	130 ± 20	32 ± 7	
$\Delta H_N^{\text{ex}^\circ}$ (kJ mol ^{−1})	$\Delta H_U^{\text{ex}^\circ}$ (kJ mol ^{−1})	$\Delta C_{pN}^{\text{ex}^\circ}$ (J mol ^{−1} K ^{−1})	$\Delta C_{pU}^{\text{ex}^\circ}$ (J mol ^{−1} K ^{−1})
6.8 ± 0.5	-0.2 ± 0.1	12 ± 4	220 ± 20

$K_N^{\text{ex}^\circ}$: ethanol–water exchange thermodynamic constant at the reference temperature for the native state of lysozyme; $\langle z_p \rangle$: mean value of the effective charge of lysozyme in the native state; $\langle d \rangle$: mean value of the decay length of the attractive Yukavian potential; $\nu_{W,N,s}$, volume of water molecules in the solvation sites; $\langle J \rangle_{x_L=0.00}$, $\langle J \rangle_{x_L=0.03}$, $\langle J \rangle_{x_L=0.06}$ and $\langle J \rangle_{x_L=0.09}$: mean values of the depth of the attractive Yukavian potential; $K_U^{\text{ex}^\circ}$: ethanol–water exchange thermodynamic constant at the reference temperature for the unfolded state of lysozyme; ΔH_W° and ΔG_W° : unfolding enthalpy and Gibbs free energy of lysozyme in pure water at the reference temperature; C_{pWU}° , C_{pWNI} and C_{pWU1} : heat capacity of lysozyme in the unfolding state and first temperature derivatives of the heat capacity of lysozyme in the native and in the unfolding states; m_U : number of solvation sites in the unfolding state of lysozyme; $\Delta H_N^{\text{ex}^\circ}$ and $\Delta C_{pN}^{\text{ex}^\circ}$: changes of enthalpy at the reference temperature and of the heat capacity for the ethanol–water exchange in the solvation sites of lysozyme in the native state; $\Delta H_U^{\text{ex}^\circ}$ and $\Delta C_{pU}^{\text{ex}^\circ}$: changes of enthalpy, at the reference temperature and of the heat capacity for the ethanol–water exchange in the solvation sites of lysozyme in the unfolded state (Ortore et al. 2011)

The last class involves parameters that are specific of single experiments. All the experimental SANS curves, together with their best fits are reported in Fig. 10. Concerning DSC, data and fits are represented in Fig. 11. In all cases the agreement with the experiments is very good, confirming that the

unique model here adopted is able to catch all the information contained in the data. Fitting parameters, divided in the three classes introduced above, are reported in Table 4. Some results deserve proper comments. The protein effective charge of the native lysozyme is almost independent on protein and ethanol concentration: its mean value is found $\langle z_P \rangle = 16.0 \pm 0.3 q_e$, corresponding to the charge of lysozyme at pH 3 according to its amino acid composition (Toldo 1997). The decay length of the attractive Yukavian potential is also almost constant, with an average value $\langle d \rangle = 3.0 \pm 0.4 \text{ \AA}$. On the contrary, the depth J is decreasing with x_L , confirming the observation that the increase of the effective lysozyme repulsive forces in the presence of ethanol is mainly determined by the reduction of attractive hydrophobic interactions (Liu et al. 2004). Clearly, the main results of this study is the determination of the two thermodynamic exchange constants K_N^{ex} and K_U^{ex} . Due to the use of activity coefficients, from their numerical values (both greater than unity) is not possible to establish whether there is preferential accumulation or depletion of ethanol from the surface of the protein. The answer to the question is given by the calculation of the ethanol molar fraction in the local domain ($x_{L,j,l} = 1 - x_{W,j,l}$), according to Eq. 37. Results are represented in the graph of Fig. 12. For both native and unfolded lysozyme, the experimental points are above the dashed line representing the absence of any interaction, indicating a preferential interaction of the surfaces of both protein states with ethanol. Quantitative results, here confirmed by both experiments, indicate a slightly higher preference of the N-sites with respect the U-site. This achievement should be read in parallel with the increase of the number of solvation sites on going from N to U-state, which transforms from $m_N = 385$ to $m_U = 418 \pm 2$.

Fig. 12 Ethanol molar fraction in the local domain as a function of the nominal ethanol molar fraction x_L . The thin dashed line represents the effective constant $K_j^{\text{ex}} \gamma_L / \gamma_L = 1$. Continuous lines with points result from the global fit of native (open circles) and unfolded (close circles) lysozyme dissolved in water/ethanol mixtures (Ortore et al. 2011)

Fig. 13 Number of excess molecules of the species j in the lysozyme solvation layer as a function of the nominal ethanol molar fraction x_L . Circles refers to ethanol and squares to water. Empty and filled symbols refer to the native and to the unfolded protein, respectively (Ortore et al. 2011)

The combination of these effects is reflected in the behavior of the excess solvation numbers for both states, reported in Fig. 13. We observe that the excess solvation numbers for native lysozyme show a critical point at $x_L \approx 0.05$, corresponding to the molar fraction in which a transition in water–ethanol mixtures has been detected. This important result confirms the general idea that a protein is a good probe to evidence binary mixtures properties. While at the lowest values of x_L a standard equilibrium between free water and free ethanol molecules probed by the protein surface exists, it is possible that for $x_L > 0.05$ the formation of ethanol clusters introduces a sort of a partition of free ethanol molecules between protein surface and ethanol clusters. We have to add that this phenomenology could be implicitly included in our model, which makes use of experimental activity coefficients: for this reason, a unique thermodynamic exchange constant succeeds in analyzing all the investigated experimental data.

Conclusions

The quantification of protein preferential solvation in the presence of stabilizing or de-stabilizing cosolvents is a cutting-edge issue, mostly studied with molecular dynamics simulations (Roccatano 2013). The complexity of this phenomenon is also due to the interplay between the interactions of proteins with solvent molecules and the structures that arise in the binary solvent at the nano- or micro-scale. To get information from an experimental point of view on preferential solvation, it is necessary to exploit a technique able to detect structural information on the protein and on its solvation shell. We have shown that SANS, combined

with a wise choice of deuterated and hydrogenated solvent molecules, is one of the most powerful tool to get insight in this kind of systems, investigating proteins at concentrations possibly resembling those found in vivo. Because SANS curves can be only slightly modified by preferential solvation effects, in order to extract quantitative information, it is necessary from one hand to investigate a large landscape of experimental conditions in terms of protein concentration and solvent composition and, from the other hand, to analyze the whole set of data by applying advanced methods based on proper structural as well as thermodynamic models. We have shown that the simple exchange model early developed by Schellman is well suitable to catch the main preferential solvation mechanisms. However, this method should be combined with experimental calorimetric information on the binary solvent in order to be successfully applied in a wide range of compositions. Moreover, the combined use of DSC and SANS allows to cross-validate structural features with thermodynamic properties and to describe how these features change with temperature, protein concentration and solvent composition. The picture that emerges from the combination of SANS and DSC involves not only preferential solvation, but also the changes of protein–protein interactions induced by cosolvents, a very important biophysical and biotechnological issue.

The investigation of structure and thermodynamics of proteins in binary aqueous mixtures can be considered another way to understand the role of hydration in protein stability. How the molecules in the solvent interact with the molecules they dissolve is the basic question raising from the issue we now call solvation science (Morgens-tern et al. 2015). Due to the wide interdisciplinary network possibly involved by this basic research, new experimental approaches and synergic theoretical investigations on this topic can be expected in the next future.

Acknowledgments The authors gratefully acknowledge Raffaele Sinibaldi, who dedicated his PhD thesis to this issue, Stefania Cinelli for her accuracy in most samples preparation, and Flavio Carsughi for his stimulating discussions. The authors remember with gratitude and affection Giuseppe Onori who inspired and motivated their research on preferential solvation.

Compliance with Ethical Standards

Conflict of interest Francesco Spinozzi declares that he has no conflict of interest.

Paolo Mariani declares that he has no conflict of interest.

Maria Grazia Ortore declares that she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Allison SK, Fox JP, Hargreaves JP, Bates SP (2005) Clustering and microimmiscibility in alcohol–water mixtures: evidence

- from molecular-dynamics simulations. *Phys Rev B* 71:024,201–024,205
- Bandyopadhyay D, Mohan S, Ghosh SK, Choudhury N (2014) Molecular dynamics simulation of aqueous urea solution: is urea a structure breaker? *J Phys Chem B* 118(40):11,757–11,768
- Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 10:980–980
- Boyne JA, Williamson AG (1967) Enthalpies of mixture of ethanol and water at 25 °C. *J Chem Eng Data* 12:318–318
- Broccio M, Costa D, Liu Y, Chen SH (2006) The structural properties of a two-Yukawa fluid: simulation and analytical results. *J Chem Phys* 124:084,501–084,509
- Carr JK, Buchanan LE, Schmidt JR, Zanni MT, Skinner JL (2013) Structure and dynamics of Urea/Water mixtures investigated by vibrational spectroscopy and molecular dynamics simulation. *J Phys Chem B* 117:10,1021
- Chen SH, Broccio M, Liu Y, Fratini E, Baglioni P (2007) The two-Yukawa model and its applications: the cases of charged proteins and copolymer micellar solutions. *J Appl Cryst* 40:s321–s326
- Cinelli S, Francesco AD, Onori G, Paciaroni A (2004) Thermal stability and internal dynamics of lysozyme as affected by hydration. *Phys Chem Chem Phys* 6:3591–3595
- Diamond R (1974) Real-space refinement of the structure of hen egg-white lysozyme. *J Mol Biol* 82:371–391
- Dianoux AJ, Lander G (2003) Neutron data booklet. OCP Science, Grenoble, France
- Dzwolak W (2006) Tuning amyloidogenic conformations through cosolvents and hydrostatic pressure: when the soft matter becomes even softer. *Biochim Biophys Acta (BBA), Proteins Proteomics* 1764(3):470–480
- Gabel F, Ringkjbing Jensen M, Zaccai G, Blackledge M (2009) Quantitative model-free analysis of urea binding to unfolded ubiquitin using a combination of small-angle X-ray and neutron scattering. *J Am Chem Soc* 131:8769–8771
- Farnum M, Zukoski C (1999) Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophys J* 76:2716–2726
- Fitter J, Gutberlet T, Katsaras J (2006) Neutron scattering in biology. Springer, Berlin
- Gekko K (1982) Calorimetric study on thermal denaturation of lysozyme in polyol–water mixtures. *J Biochem* 91:1197–1204
- Ghosh R, Roy S, Bagchi B (2013) Solvent sensitivity of protein unfolding: dynamical study of chicken villin headpiece subdomain in water/ethanol binary mixture. *J Phys Chem B* 117(49):15,625–15,638
- Glatter O, Kratky O (1982) Small-angle x-ray scattering. Academic Press, New Press
- Golier JPE, Wilhelm E (1981) Excess volumes and excess heat capacities of water + ethanol at 298.15 K. *Fluid Phase Equilib* 6:283–287
- Gonzalez B, Calvar N, Gomez E, Dominguez A (2007) Density, dynamic viscosity, and derived properties of binary mixtures of methanol or ethanol with water, ethyl acetate, and methyl acetate at $T = (293.15, 298.15, \text{ and } 303.15)$ K. *J Chem Thermodyn* 39:1578–1588
- Graziano G (2011) How does trimethylamine N-oxide counteract the denaturing activity of urea. *Phys Chem Chem Phys* 13:17,689–17,695
- Graziano G (2012) How does sucrose stabilize the native state of globular proteins? *Int J Biol Macromol* 50:230–235
- Guo JH, Luo Y, Augustsson A, Kashtanov S, Rubensson JE, Shuh DK, Ågren H, Nordgren J (2003) Molecular structure of alcohol–water mixtures. *Phys Rev Lett* 91:1574,011–1574,014
- Frank HS, Franks F (1968) Structural approach to the solvent power of water for hydrocarbons; urea as a structure breaker. *J Chem Phys* 48:4746–4757

- Hansen RS, Miller FA (1954) A new method for determination of activities of binary solutions of volatile liquids. *J Phys Chem* 58:193–196
- Ikeguchi M, Nakamura S, Shimizu K (2001) Molecular dynamics study on hydrophobic effects in aqueous urea solutions. *J Am Chem Soc* 123:677–682
- Kirkwood JG, Buff FP (1951) The statistical mechanical theory of solutions. *J Chem Phys* 19:774–777
- Narayanan J, Liu XY (2003) Protein interactions in undersaturated and supersaturated solutions: a study using light and X-Ray scattering. *Biophys J* 84:523–532
- Jacrot B (1976) The study of biological structures by neutron scattering from solution. *Reg Prog Phys* 39:911
- Jain NK, Ipsita R (2009) Effect of trehalose on protein structure. *Protein Science: A Publication of the Protein Society* 18:24–36
- Javid N, Vogt K, Krywka C, Tolan M, Winter R (2007) Protein-protein interactions in complex cosolvent solutions. *Chem Phys Chem* 8:679–689
- Juraszek J, Bolhuis PG (2006) Sampling the multiple folding mechanisms of trp-cage in explicit solvent. *Proc Natl Acad Sci* 103(43):15,859–15,864
- Gekko K, Timasheff SN (1981) Thermodynamic and kinetic examination of protein stabilization by glycerol. *Biochemistry* 20:4667–4676
- Karino Y, Matubayasi N (2013) Interaction-component analysis of the urea effect on amino acid analogs. *Phys Chem Chem Phys* 15:4377–4391
- Klibanov AM (2001) Improving enzymes by using them in organic solvents. *Nature* 409:241–246
- Lehmann MS, Zaccai G (1984) Neutron small-angle scattering studies of ribonuclease in mixed aqueous solutions and determination of the preferentially bound water. *Biochemistry* 23:1939–1942
- Levinthal C (1968) Are there pathways for protein folding. *J Chem Phys* 65:44–45
- Lide DR (1996) *Handbook of chemistry and physics*, 77th ed. CRC Press, Cleveland
- Liu W, Bratko D, Prausnitz JM, Blanch HW (2004) Effect of alcohol on aqueous lysozyme-lysozyme interactions from static light-scattering measurements. *Biophys Chem* 107:289–298
- Liu Y, Chen WR, Chen SH (2005) Cluster formation in two-Yukawa fluids. *J Chem Phys* 122(044):507–13
- Liu Y, Porcar L, Chen J, Chen WR, Falus P, Faraone A, Fratini E, Baglioni KHP (2011) Lysozyme protein solution with an intermediate range order structure. *J Phys Chem B* 115:7238–7247
- Mallamace F, Corsaro C, Mallamace D, Baglioni P, Stanley HE, Chen SH (2011) A possible role of water in the protein folding process. *J Phys Chem B* 115(48):14,280–14,294
- Marcus Y (2000) Some thermodynamic and structural aspects of mixtures of glycerol with water. *Phys Chem Chem Phys* 2:4891–4896
- Merzel F, Smith J (2002) Sasim: a method for calculating small-angle scattering from explicit-atom models. *Acta Cryst D* 58:242–249
- Morgenstern K, Marx D, Havenith M, Muhler M (2015) Editorial of the pccp themed issue on “solvation science”. *Phys Chem Chem Phys* 17:8295–8296
- Niebuhr M, Koch MHJ (2005) Effects of urea and Trimethylamine-N-Oxide (TMAO) on the interactions of lysozyme in solution. *Biophys J* 89:1978–1983
- Oliveira CLP, Behrens MA, Pedersen JS, Erlacher K, Otzen D, Pedersen JS (2009) A SAXS study of glucagon fibrillation. *J Mol Biol* 387:147–161
- Onori G, Santucci A (1996) Dynamical and structural properties of water/alcohol mixtures. *J Mol Liq* 69:161–181
- Ortore MG, Mariani P, Carsughi F, Cinelli S, Onori G, Teixeira J, Spinozzi F (2011) Preferential solvation of lysozyme in water/ethanol mixtures. *J Chem Phys* 135:245,103–245,111
- Ortore MG, Sinibaldi R, Spinozzi F, Carbin A, Carsughi F, Mariani P (2009) Looking for the best experimental conditions to detail the protein solvation shell in a binary aqueous solvent via small-angle scattering. *J Phys Conf Ser* 177:012,007
- Ortore MG, Sinibaldi R, Spinozzi F, Carsughi F, Clemens D, Bonincontro A, Mariani P (2008) New insights into urea action on proteins: SANS and Zeta Potential study of lysozyme. *J Phys Chem B* 112:12,881–12,887
- Ortore MG, Spinozzi F, Mariani P, Paciaroni A, Barbosa LRS, Amenitsch H, Steinhart M, Ollivier J, Russo D (2009) Combining structure and dynamics: non-denaturing high-pressure effect on lysozyme in solution. *J R Soc Interface* 6:S619–S634
- Oshima H, Kinoshita M (2013) Effects of sugars on the thermal stability of a protein. *J Chem Phys* 138:245,101–245,112
- Ozdemir M, Sadikoglu H (1999) Use of unifac in the prediction of water activity values of aqueous polyol solutions. *J Food Process Preserv* 23:109–120
- Paciaroni A, Cinelli S, Onori G (2002) Effect of the environment on the protein dynamical transition: a neutron scattering study. *Biophys J* 83:1157
- Paciaroni A, Orecchini A, Cinelli S, Onori G, Lechner RE, Pieper J (2003) Protein dynamics on the picosecond timescale as affected by the environment: a quasielastic neutron scattering study. *J Chem Phys* 292:397–404
- Paul S, Paul S (2015) Investigating the counteracting effect of trehalose on Urea-Induced protein denaturation using molecular dynamics simulation. *J Phys Chem B* 119:10,975–10,988
- Pedersen JS, Schurtenberger P (1996) Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules* 29:7602–7612
- Rariy RV, Klibanov AM (1997) Correct protein folding in glycerol. *Proc Natl Acad Sci* 94:13,520–13,523
- Roccatano D (2013) Computer simulations of biomolecules in non-aqueous and semi-aqueous solvent conditions. *Advances in Protein and Peptide Sciences* 1:318–381
- Roney AB, Space B, Castner EW, Napoleon RL, Moore PB (2004) A molecular dynamics study of aggregation phenomena in aqueous n-propanol. *J Phys Chem B* 108:7389–7401
- Ross GA, Morris GM, Biggin PC (2012) Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS ONE* 7(3):e32,036
- Schellman JA (1990) A simple model for solvation in mixed solvents. Applications to the stabilization and destabilization of macromolecular structures. *Biohpys Chem* 37:121–140
- Schellman JA (1994) The thermodynamics of the solvent exchange. *Biopolymers* 34:1015–1026
- Schellman JA (2003) Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophys J* 85:108–125
- Schellman JA (2005) Destabilization and stabilization of proteins. *Q Rev Biophys* 38:351–361
- Schroer MA, Zhai Y, Wieland DCF, Sahle CJ, Nase J, Paulus M, Tolan M, Winter R (2011) Exploring the piezophilic behavior of natural cosolvent mixtures. *Angew Chem Int Ed* 50:11,413–11,416
- Seeliger J, Estel K, Erwin N, Winter R (2013) Cosolvent effects on the fibrillation reaction of human IAPP. *Phys Chem Chem Phys* 15:8902–8907
- Shulgin IL, Ruckenstein E (2005) A protein molecule in an aqueous mixed solvent: fluctuation theory outlook. *J Chem Phys* 123(054):909

- Sinibaldi R, Ortore M, Spinozzi F, Funari SS, Teixeira J, Mariani P (2008) SANS/SAXS study of the BSA solvation properties in aqueous urea solutions via a global fit approach. *Eur Biophys J* 37:673–681
- Sinibaldi R, Ortore MG, Spinozzi F, Carsughi F, Frielinghaus H, Cinelli S, Onori G, Mariani P (2007) Preferential hydration of lysozyme in water/glycerol mixtures: a small-angle neutron scattering study. *J Chem Phys* 126:235,101–235,109
- Soper AK, Castner EW, Luzar A (2003) Impact of urea on water structure: a clue to its properties as a denaturant? *Biophys Chem* 105:649–666
- Soper AK, Dougan L, Crain J, Finney JL (2006) Excess entropy in alcohol-water solutions: a simple clustering explanation. *J Phys Chem* 110:3472–3476
- Sorenson JM, Hura G, Soper AK, Pertsemlidis A, Head-Gordon T (1999) Determining the role of hydration forces in protein folding. *J Phys Chem B* 103(26):5413–5426
- Spinozzi F, Ferrero C, Ortore MG, Antolinos ADM, Mariani P (2014) GENFIT: Software for the analysis of small-angle X-ray and neutron scattering data of macromolecules in-solution. *J App Cryst* 47:1132–1139
- Spinozzi F, Gazzillo D, Giacometti A, Mariani P, Carsughi F (2002) Interaction of proteins in solution from small-angle scattering: a perturbative approach. *Biophys J* 82:2165–2175
- Spinozzi F, Ortore MG, Sinibaldi R, Mariani P, Esposito A, Cinelli S, Onori G (2008) Microcalorimetric study of thermal unfolding of lysozyme in water/glycerol mixtures: an analysis by solvent exchange model. *J Chem Phys* 129:35,101–35,109
- Suresh SJ, Naik VM (2002) Theory of dielectric constant of aqueous solutions. *J Chem Phys* 116:4212–4220
- Svergun D, Barberato C, Koch MHJ (1995) CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28:768–773
- Svergun D, Richard S, Koch MHJ, Sayers Z, Kuprin S, Zaccai G (1998) Protein hydration in solution: experimental observation by X-ray, neutron scattering. *Proc Natl Acad Sci USA* 95:2267–2272
- Tanford C, Roxby R (1972) Interpretation of protein titration curves. Application to lysozyme. *Biochemistry* 11:2192–2198
- Timasheff SN (1998) Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. *Adv Protein Chem* 51:355–432
- Timasheff SN, Inoue H (1968) Preferential binding of solvent components to proteins in mixed water-organic solvent systems. *Biochemistry* 7(7):2501–2513
- Timasheff SN, Xie G (2003) Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys Chem* 105:421–448
- Toldo LIG (1997) *JaMBW* Chapter 3.1.6
- Vagenende V, Han AX, Pek HB, Loo BLW (2013) Quantifying the molecular origins of opposite solvent effects on protein–protein interactions. *PLoS Computational Biology* 9:e1003,072
- Vaiana SM, Manno M, Emanuele A, Palma-Vittorelli MB, Palma MU (2001) The role of solvent in protein folding and in aggregation. *J Biol Phys* 27:133–145
- Virtanen JJ, Makowski L, Sosnick TR, Freed KF (2010) Modeling the hydration layer around proteins: HyPred. *Biophys J* 99(5):1611–1619
- Waasmaier D, Kirfel A (1995) New analytical scattering-factor functions for free atoms and ions. *Acta Cryst A* 51:416–431
- Wakisaka A, Ohki T (2005) Phase separation of water-alcohol binary mixtures induced by the microheterogeneity. *Faraday Discuss* 129:231–245
- Xie G, Timasheff SN (1997) The thermodynamic mechanism of protein stabilization by trehalose. *Biophys Chem* 64:25
- Yancey PH, Clark ME, Hand S, Bowlus R, Somero G (1982) Living with water stress: evolution of osmolyte systems. *Science* 217:1214–1222
- Zhang C, Yang X (2005) Molecular dynamics simulation of ethanol/water mixtures for structure and diffusion properties. *Fluid Phase Equilib* 231:1–10
- Zhou R (2003) Free energy landscape of protein folding in water: Explicit vs. implicit solvent. *Proteins: Struct, Funct, Bioinf* 53(2):148–161
- Zou Q, Bennion B, Daggett V, Murphy K (2002) The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *J Am Chem Soc* 124:1192