RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2015, **5**, 59780

Analyzing the effects of protecting osmolytes on solute–water interactions by solvatochromic comparison method: II. Globular proteins

Luisa A. Ferreira,[a] Xiao Fan,[b] Pedro P. Madeira,[c] Lukasz Kurgan,[b] Vladimir N. Uversky*[def] and Boris Y. Zaslavsky*[a]

Received 8th May 2015
Accepted 3rd July 2015

DOI: 10.1039/c5ra08612d

www.rsc.org/advances

Partitioning of 11 globular proteins was examined in aqueous dextran–PEG–sodium/potassium phosphate buffer (0.01 M K/NaPB, pH 7.4) two-phase systems (ATPSs) containing 0.5 M sorbitol. The data obtained were analyzed together with those reported previously for the same proteins in osmolyte-free ATPS and ATPS containing 0.5 M sucrose, TMAO, or trehalose. It was found that all the partition coefficients for proteins determined in the presence of 0.5 M of different osmolytes and in the absence of osmolytes may be described in terms of the differences between solvent properties of the coexisting phases. Solute-specific coefficients characterizing different types of solute–solvent interactions were calculated for each protein. These solute-specific coefficients are linearly interrelated implying cooperativity of different types of protein–water interactions. The data obtained indicate the lack of any association of the aforementioned osmolytes at concentration of 0.5 M with proteins. Computational analysis of one of the solute-specific coefficient S_s-values characterizing dipole–dipole protein–water interactions shows that it is determined by the peculiarities of protein surface.

Introduction

Protective osmolytes are small organic compounds, typically neutral non-electrolytes or zwitterions, from different chemical classes, such as amino acids (*e.g.*, proline and glycine), methylamines (*e.g.*, sarcosine, trimethylamine *N*-oxide (TMAO), and betaine), and polyols (*e.g.*, glycerol, sucrose, trehalose, and certain other sugars). They are found at significant intracellular concentrations in many plants, animals, and microorganisms that have adapted to environmental extremes.[1–5] The primary role of osmolytes is the maintenance of the cell volume and fluid balance. To protect against harsh environmental conditions, high concentrations of osmolytes are accumulated by organisms.[6–8] In addition to their roles in regulation and protection of normal osmosis, osmolytes are known to protect cellular macromolecules against denaturation by hostile environments.[1–4,6,9,10] Since globular proteins are marginally stable and can be easily denatured by various environmental stresses and insults, osmolytes, with their ability to affect stability of proteins, play important roles in maintenance of normal proteostasis.[5,8,10,11] It is believed that osmolytes stabilizes proteins and other cellular biomolecules against the denaturing environmental stresses without significantly affecting the functional activity of the proteins and other cellular components.[4–6,12–16]

For a long time it is believed that one of the potential mechanisms explaining stabilizing action of osmolytes is preferential exclusion of osmolytes from the immediate vicinity of a protein.[5,17] Since osmolyte is preferentially excluded from the immediate protein vicinity, the protein is then preferentially hydrated.[5,18–21] In other words, the "folding" potential of protective osmolytes can be understood in terms of the model where a globular protein tends to adopt a folded conformation with a minimally exposed surface area due to the tendency of protective osmolytes to be excluded from the protein surface.[8,22]

Thermodynamically, the stabilization of a protein by protective osmolytes was attributed to the destabilization of the unfolded state of the protein in the presence of osmolyte rather than to the osmolyte-induced stabilization of the native state.[5,14,15,21,23] Therefore, protecting osmolytes push the folding equilibrium toward native state by raising the free energy of the unfolded state, whereas denaturing osmolytes push the equilibrium toward the unfolded state by lowering the free energy of

[a]*Analiza Inc, 3516 Superior Ave, Suite 4407B, Cleveland, OH, USA. E-mail: bz@analiza.com; Tel: +1-216-432-2700*

[b]*Department of Electrical and Computer Engineering, University of Alberta, Canada*

[c]*Laboratory of Separation and Reaction Engineering, Department of Chemical Engineering, Faculty of Engineering of the University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal*

[d]*Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, USA. E-mail: vuversky@health.usf.edu; Fax: +1-813-974-7357; Tel: +1-813-974-5816*

[e]*Department of Biological Science, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia*

[f]*Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia*

the unfolded state.[24] The fact that protective osmolytes act primarily on the denatured states leaving native states mostly unaffected explains the ability of these important small organic compounds to stabilize proteins against the denaturation without affecting their biological functions.[5]

Since protecting/denaturing osmolytes interact unfavorably/favorably with the unfolded state, resulting in preferential depletion/accumulation of osmolyte proximate to the protein surface,[24] the important question is by what mechanism osmolytes interact with the protein to affect its stability. It has been emphasized that osmolytes modulate protein stability predominantly affecting the protein backbone,[14,25,26] with osmolyte polar groups being able to interact with the protein backbone more favorably than the osmolyte non-polar groups.[24] This suggests that protein stabilization in the presence of osmolytes can be attributed to the net repulsive interaction between protecting osmolytes and the backbone of proteins.[26–29]

An alternative mechanism of osmolyte action is related to the potential effects of these small organic compounds on the solvent properties in the cellular environment. In this scenario, the presence of osmolytes indirectly modifies the stability of biological macromolecules *via* changes in the solvent properties.[30,31] In fact, it was emphasized that the osmolyte-induced shift in the conformational equilibrium toward the protein native state might be rooted in the ability of protective osmolytes to induce asymmetric loss of protein conformational entropy, with greater entropic loss in the unfolded state.[29,32] In this scenario, osmolytes reduce the entropy of the ensemble of unfolded conformations by increasing compactness of the unfolded state.[29]

In other words, in the presence of osmolytes, the unfolded state acquires residual partially collapsed structure characterized by the reduced number of solvent-accessible hydrophobic groups resulting in decreased number of water molecules that have to be immobilized upon unfolding.[33] This ability of osmolytes to induce collapse of the unfolded state was demonstrated for several globular proteins, such as protein S6,[34] RNase S,[32] chymotrypsin inhibitor 2,[35] and cutinase[33]. Importantly, almost in all cases studies so far, the addition of osmolytes to the unfolded proteins resulted in the rapid collapse of the unfolded state to the non-native form with the retarded refolding capabilities.[32–35] Furthermore, it has been pointed out that different osmolytes can induce differently collapsed states in a given protein.[35]

The observations that osmolytes are able to induce compaction of unfolded states provide indirect support to the idea that the addition of osmolyte might change the properties of solvent. In fact, it is well known that the hydrodynamic dimensions of unfolded polymers dramatically depend on the quality of solvent.[36,37] A poor solvent induces the attraction of macromolecular segments, resulting in the squeezing of a chain. On the other hand, in a good solvent, repulsive forces occur between segments, leading to the formation of a loose fluctuating coil.[38] It is assumed that the concentrated solutions of urea and guanidinium hydrochloride (GdmHCl) are rather good solvents for polypeptide chains, with GdmHCl being closer to the ideal one.[37,39] This difference in the solvent quality accounts for the noticeable divergences in the molecular mass dependencies of the hydrodynamic dimensions of the globular proteins unfolded by urea and GdmHCl.[40–43] From this angle, the reported osmolyte-induced compaction of unfolded proteins indicates changes in solvent quality from good in the absence of osmolytes to poor in their presence.

In brief, the mechanism of stabilizing effects of osmolytes on proteins in aqueous solution on molecular level remains unclear. Although the prevalent view is based on the preferential solvation model, according to which osmolytes are excluded from protein surface and increase the free energy of protein unfolding,[28,44] it is generally agreed that the water structure is changed in osmolyte solutions.[29,45–53]

Although the data accumulated so far are somewhat contradictory, the conclusion that the water properties in osmolyte solutions are changed relative to those of pure water seems unavoidable. This conclusion is confirmed by the data reported in the companion paper,[54] where based on the analysis of partitioning of small organic compounds in aqueous dextran–polyethylene glycol (PEG) aqueous two-phase systems containing different osmolytes (sorbitol, sucrose, trehalose, and TMAO) at concentration of 0.5 M it has been concluded that the compound partition behavior may be described in terms of the solvent properties of coexisting phases. This finding clearly indicates that partition behavior of a solute is not associated with direct osmolyte–solute interactions and reflects changes in the osmolyte-induced solvent properties.[54]

The purpose of this study was to examine partitioning of proteins in the dextran–PEG–0.01 M K/NaPB–0.5 M sorbitol ATPS, and estimate the solute-specific coefficients for all the proteins to explore if all the osmolytes employed affect the protein partition behavior solely by affecting the solvent properties of the aqueous media.

Materials and methods

Materials

Polymers. Polyethylene glycol PEG-8000 (Lot 091M01372V) with an average molecular weight (M_n) of 8000 was purchased from Sigma-Aldrich (St. Louis, MO, USA) and Dextran-75 (Lot 119945) with an average molecular weight (M_w) 75 000 by light scattering was purchased from USB Corporation (Cleveland, OH, USA).

Proteins. All the proteins studied, their molecular weights, and pI values are listed in Table 1. α-Chymotrypsin from bovine pancreas, α-chymotrypsinogen A from bovine pancreas, concanavalin A from *Canavalia ensiformis* (jack beans), hemoglobin human, β-lactoglobulin A from bovine milk (>90%), β-lactoglobulin B from bovine milk (>90%), lysozyme from chicken egg white, papain from papaya latex, ribonuclease A from bovine pancreas, ribonuclease B from bovine pancreas, and trypsinogen from bovine pancreas were purchased from Sigma-Aldrich.

Other chemicals. All salts and other chemicals used were of analytical-reagent grade and used without further purification.

Table 1 Proteins used in this study

Protein*	Abbreviation	Molecular weight, kDa	pI
α-Chymotrypsin	CHY	25.0	8.75
α-Chymotrypsinogen A	CHTG	25.7	8.97
Concanavalin A	ConA	104.0	4.5–5.5
Hemoglobin human	HHb	64.5	6.8
β-Lactoglobulin A	bLGA	18.3	5.3
β-Lactoglobulin B	bLGB	18.3	5.1
Lysozyme	HEL	14.3	11.0
Papain	Pap	23.4	8.75–9.55
Ribonuclease A	RNase A	13.7	9.63
Ribonuclease B	RNase B	17.0	8.88
Trypsinogen	TRY	24.0	8.7; 9.3

Methods

Aqueous two-phase systems (ATPSs). Stock solutions of PEG 8000 (50 wt%), Dex-75 (~42 wt%), and sorbitol (2.0 M) were prepared in deionized (DI) water. Stock sodium/potassium phosphate buffer (K/NaPB; 0.5 M, pH 7.4) was prepared by mixing appropriate amounts of KH_2PO_4 and Na_2HPO_4. A mixture of polymers was prepared as described elsewhere[55] by dispensing appropriate amounts of the aqueous stock polymer solutions into a 1.2 mL microtube using a Hamilton Company (Reno, NV, USA) ML-4000 four-probe liquid-handling workstation. Appropriate amounts of stock solution of sorbitol, stock buffer solutions, and water were added to give the ionic, polymer, and osmolyte composition required for the final system (after the sample addition – see below) with total weight of 0.5 g (total volume 457 ± 2 μL). All the aqueous two-phase systems used had the same polymer composition of 6.0 wt% PEG-8000 and 12.0 wt% Dex-75 and same ionic composition of 0.01 M K/NaPB, pH 7.4 with 0.5 M sorbitol.

Partitioning. An automated instrument for performing aqueous two-phase partitioning, the Automated Signature Workstation, ASW (Analiza, Inc., Cleveland, OH, USA), was used for the partitioning experiments. The ASW system is based on the ML-4000 liquid-handling workstation (Hamilton Company, Reno, NV, USA) integrated with a FL600 fluorescence microplate reader (Bio-Tek Instruments, Winooski, VT, USA) and a UV-Vis microplate spectrophotometer (SpectraMax Plus 384, Molecular Devices, Sunnyvale, CA). Solutions of all proteins were prepared in water at concentrations of 1–5 mg mL^{-1}. Varied amounts (*e.g.* 0, 15, 30, 45, 60 and 75 μL) of protein solution and the corresponding amounts (*e.g.* 75, 60, 45, 30, 15 and 0 μL) of water were added to a set of the same polymers/buffer mixtures with sorbitol. The systems were then vortexed in a Multipulse vortexer and centrifuged (Jouan, BR4i, Thermo Fisher Scientific, Waltham, MA, USA) for 60 min at 3500 × *g* at 23 °C to accelerate phase settling. The top phase in each system was removed, the interface discarded, and aliquots from the top and bottom phases were withdrawn in duplicate for analysis.

For the analysis of the proteins (with exception of hemoglobin) partitioning, aliquots of 30 μL from both phases were transferred and diluted with water up to 70 μL into microplate wells. Then, the microplate was sealed, shortly centrifuged (2 min at 1500 rpm) and following moderate shaking for 45 min in an incubator at 37 °C, 250 μL of *o*-phthaldialdehyde reagent was combined. After moderate shaking for 4 min at room temperature, fluorescence was determined using a fluorescence plate reader with a 360 nm excitation filter and a 460 nm emission filter, with a sensitivity setting of 100–125.

For the analysis of hemoglobin partitioning, aliquots of 50 – 120 μL from both phases were diluted up to 600 μL in 1.2 mL microtubes. Water was used as diluent. Following vortexing and a short centrifugation (12 min), aliquots of 250 – 300 μL were transferred into microplate wells, and the UV-Vis plate reader was used to measure optical absorbance at wavelength previously determined to correspond to maximum absorption. In separate experiments we compared the protein partition coefficients under two condition of mixing. One condition included vortexing of protein added to the polymer mixture. The other condition included vigorous vortexing of polymer mixture to apparently homogeneous state, followed by adding protein stock solution and subsequent "gentle" mixing with very brief and mild vortexing. The partition coefficients determined for each protein under these two different mixing conditions were identical.

The partition coefficient, *K*, is defined as the ratio of the sample concentration in the top phase to that in the bottom phase. The *K*-value for each protein was determined as the slope of the concentration (fluorescence intensity or absorbance depending on the protein) in the top phase plotted as a function of the concentration in the bottom phase averaged over the results obtained from two to four partition experiments carried out at the specified composition of the system. The deviation from the average *K* value was always less than 3% and in most cases lower than 1%.

Dataset. Analysis was performed using a set of 9 diverse proteins with known 3D structures; PDB ids: 1AB9 (CHY), 1ACB (CHTG), 1B8E (bLGA), 1BEB (bLGB), 1BEL (RNase A), 1BTY (TRY), 1BZ0 (HHb), 1JBC (ConA), and 1PPN (Pap). The corresponding sequences range between 124 and 287 residues.

Protein descriptors. Similar to work reported in ref. 56 a comprehensive set of structural descriptors for each considered protein was collected. They include six descriptors that were derived directly from the amino acid sequence: chain length, molecular weight, isoelectric point (pI) obtained using the ExPASy ProtParam method (http://web.expasy.org/protparam/),[57] and three values that characterize intrinsic disorder that were predicted with MFDp (http://biomine-ws.ece.ualberta.ca/MFDp.html).[58] These disorder descriptors are disorder content (fraction of disordered residues), the number of disordered segments divided by the protein size, and the average (over the whole chains) propensity of disorder. Additional 51 descriptors were also computed from the structure which quantify the shape of the protein, surface area, cavity/pockets on the surface, packing density, secondary structure, intrinsic disorder, occupancy, and flexibility. The structure-derived descriptors were generated using several applications and algorithms:

- Voronoia (http://proteinformatics.charite.de/voronoia4rna/tools/v4rna/index)[59] that was utilized to characterize packing and pockets in the structure. Ten descriptors were computed that

characterize size and quantity of pockets and average van der Waals volume, solvent-excluded volume, fraction of buried atoms and average packing density that describe packing (14 descriptors).

- CASTp (http://sts.bioe.uic.edu/castp/)[60] that generates number, surface area and volume of pockets on the surface. Both raw and normalized by the protein size values (6 descriptors) were used.

- An algorithm based on ref. 61 to compute contact order (1 descriptor).

- YASARA (http://www.yasara.org/) that we used to compute radius of gyration, nuclear and van der Waals radii, content of six types of secondary structure (α-helix, 3_{10}-helix, both helix types, β-sheet, turns and coils), molecular mass, B-factor and occupancy (12 descriptors).

- DSSP (http://swift.cmbi.ru.nl/gv/dssp/)[62] that quantifies surface area and secondary structure. Size and properties of the surface were characterized including fraction (in the whole protein chain) of surface residues; fraction of polar, nonpolar, neutral, positively charged, and negatively charged residues on the surface; and hydrophobicity of surface residues that was estimated based on three amino acids indices: Kyte-Doolittle,[63,64] Eisenberg,[65] and Cid scales.[66] Contents of α-helix, 3_{10}-helix, both helix types, β-sheet, β-bridge, both β structure types, turn, bend, and coil secondary structures (18 descriptors) were also computed.

Multivariate modeling with regression. Correlation of an empirically selected subset of the sequence- and structure-derived descriptors was investigated using linear regression that was derived based on minimization of the sum of squared errors. More specifically, given the observed data from the two-phase system $y \in R^{t\times 1}$ and the set of descriptors $X \in R^{t\times n}$ where t is the number of proteins and n is the number of descriptors, the criterion to solve the regression model is defined as:

$$\min_r \left(\|rX - y\|_2{}^2 \right)$$

where $r \in R^{n\times 1}$ are coefficients.

We selected the subset of the descriptors empirically using greedy search. The search maximizes Pearson correlation coefficients (PCC) between the outputs of the regression that uses a given subset of descriptors and the observed values. The PCC values were measured based on three-fold cross validation on the considered set of 9 proteins to minimize overfitting into the dataset. Each descriptor was normalized to the [−1, 1] interval using maximum absolute value and we built the regression model using a subset of i descriptors. First, i was initialized with 2, considered all pairs of two descriptors, and selected the pair that gives the highest value of PCC. Next, i was incremented by 1 until the corresponding regression that secured maximal PCC based on i descriptors increased the PCC value by a noticeable margin (0.03) compared to the PCC of the model with $i - 1$ descriptors. As a result, four descriptors to build the regression model were selected.

Results and discussion

It has been described in the accompanying paper[54] that the solvent properties of aqueous media in the coexisting phases of ATPS can be quantified using two approaches: (a) the solvatochromic comparison method[67–69] based on employing a set of solvatochromic dyes to characterize the solvent dipolarity/polarizability, solvent hydrogen bond donor (HBD) acidity, and solvent hydrogen bond acceptor (HBA) basicity of the media in the two phases, and (b) analysis of partitioning of a homologous series of charged compounds with varied length of aliphatic alkyl chain, such as sodium salts of dinitrophenylated amino acids to characterize the difference between the electrostatic and hydrophobic properties of the two phases (see in ref. 70 and 71).

It has been shown that the partition coefficient of a solute in an ATPS can be described as:[72–78]

$$\log K_i = S_s\Delta\pi_i^* + B_s\Delta\alpha_i + A_s\Delta\beta_i + C_s c_i \quad (1)$$

where K is the solute partition coefficient; $\Delta\pi^*$ is the difference between the solvent dipolarity/polarizability of the two phases, $\Delta\alpha$ is the difference between the solvent HBD acidity of the two phases, $\Delta\beta$ is the difference between the solvent HBA basicity of the two phases; c is the difference between the electrostatic properties of the two phases; S_s, A_s, B_s, and C_s are constants (solute specific coefficients) quantifying the complementary interactions of the solute with the solvent media in the coexisting phases and representing the relative contributions of these interactions into partition coefficient of the solute; the subscript s designates the solute; the subscript i denotes the ATPS used; the difference for each solvent property is determined as the one between the upper and lower phases.

The solute specific coefficients may be determined for a given compound (including proteins) by the analysis of partition coefficients of this compound in multiple ATPSs with different polymer but same ionic composition with established solvent properties of the phases. Once $\Delta\pi^*$, $\Delta\alpha$, $\Delta\beta$, and c parameters in 5–10 different ATPSs are determined, the solute specific coefficients can be calculated by multiple linear regression analysis using eqn (1). It was shown[75] also that the partition coefficient of a compound with pre-determined solute specific coefficients in a "new" ATPS with established solvent properties of the phases could be predicted with 90–95% accuracy.

It is important to emphasize that the partition coefficients of a solute in multiple ATPSs with different additives would fit eqn (1) only if the solute–solvent interactions would vary due to different solvent properties of the phases and there would be no association of additives with the solute. It was established[72,74] that while the minimal number of different ATPSs to be used for determination of solute-specific coefficients is five, using a set of 10 different ATPS provides much more reliable values of the solute-specific coefficients.

Effect of osmolytes on partition behavior of proteins

Partition coefficients of proteins in dextran–PEG–0.01 M K/NaPB ATPS with 0.5 M osmolytes additives (sorbitol, sucrose, trehalose, TMAO) and osmolyte-free ATPS are presented in Table 2. The differences between the solvent properties of aqueous media in the coexisting phases were characterized[54] by solvatochromic measurements and analysis of partitioning of

the homologous series of dinitrophenylated amino acids with aliphatic alkyl side-chains (DNP–Gly, DNP–Ala, DNP–norvaline, DNP–norleucine, and DNP–*n*-octanoic acid). The quantified differences between solvent ability to dipole–dipole interactions ($\Delta\pi^*$), hydrogen bond donor acidity ($\Delta\alpha$), hydrogen bond acceptor asicity ($\Delta\beta$), electrostatic properties (C), and hydrophobic properties (E) of the aqueous media in the coexisting phases are presented in Table 2.

Since the hydrophobic effect is one of the main driving forces of protein folding, we examined the relationships between the proteins partition behavior and difference between hydrophobic properties of the phases. However, the results did not provide any insight into mechanism of osmolyte effects, and hence are not discussed here. The likely reason seems to be that although the hydrophobic effect plays a crucial role of in protein folding, partition behavior of proteins is governed by the interactions of water with protein surface, and these interactions we explored in our study.

It has been demonstrated previously[71] that logarithms of partition coefficients of proteins in dextran–PEG–0.5 M osmolyte–0.01 M K/NaPB are interrelated according to eqn (2):

$$\ln K_i^{0.5\text{ M TMAO}-0.01\text{ M K/NaPB}} = 0.13_{\pm 0.06} + 0.29_{\pm 0.095} \times \ln K_i^{0.5\text{ M sucrose}-0.01\text{ M K/NaPB}} + 0.8_{\pm 0.13} \times \ln K_i^{0.01\text{ M K/NaPB}},\ N = 10;\ R^2 = 0.9856;\ \text{SD} = 0.08;\ F = 239 \quad (2)$$

where $K_i^{0.5\text{ M TMAO}-0.01\text{ M K/NaPB}}$ is the partition coefficient for the ith protein in the dextran–PEG–0.5 M TMAO–0.01 M K/NaPB ATPS; $K_i^{0.01\text{ M K/NaPB}}$ is the partition coefficients for the same protein in the dextran–PEG–0.01 M K/NaPB ATPS; $K_i^{0.5\text{ M sucrose}-0.01\text{ M K/NaPB}}$ is the partition coefficient for the same protein in dextran–PEG–0.5 M sucrose–0.01 M K/NaPB ATPS; N is the number of proteins examined; R^2 is the correlation coefficient; SD is the standard deviation; and F is the ratio of variance (the data for hemoglobin do not fit the above relationship).

Similar relationship illustrated graphically in Fig. 1 exists between logarithms of partition coefficients of proteins in dextran–PEG–0.5 M osmolyte–0.01 M K/NaPB. The relationship in Fig. 1 may be described as:

$$\ln K_i^{0.5\text{ M trehalose}-0.01\text{ M K/NaPB}} = -0.02_{\pm 0.04} - 0.27_{\pm 0.09} \times \ln K_i^{0.01\text{ M K/NaPB}} + 1.36_{\pm 0.07} \times \ln K_i^{0.5\text{ M sorbitol}-0.01\text{ M K/NaPB}},\ N = 11;\ R^2 = 0.9963;\ \text{SD} = 0.059;\ F = 1082 \quad (3)$$

where $K_i^{0.5\text{ M sorbitol}-0.01\text{ M K/NaPB}}$ is the partition coefficient for the ith protein in the dextran–PEG–0.5 M sorbitol–0.01 M K/NaPB ATPS; $K_i^{0.5\text{ M trehalose}-0.01\text{ M K/NaPB}}$ is the partition coefficient for the same protein in dextran–PEG–0.5 M trehalose–0.01 M K/NaPB ATPS; all the other parameters are as defined above.

Furthermore, Fig. 2 shows that analogous relationship can be obtained for the logarithms of partition coefficients of proteins in ATPSs three different osmolytes:

$$\ln K_i^{0.5\text{ M sorbitol}-0.01\text{ M K/NaPB}} = -0.05_{\pm 0.03} + 0.18_{\pm 0.07} \times \ln K_i^{0.5\text{ M TMAO}-0.01\text{ M K/NaPB}} + 0.74_{\pm 0.05} \times \ln K_i^{0.5\text{ M trehalose}-0.01\text{ M K/NaPB}},\ N = 11;\ R^2 = 0.9953;\ \text{SD} = 0.058;\ F = 853.4 \quad (4)$$

where $K_i^{0.5\text{ M sorbitol}-0.01\text{ M K/NaPB}}$ is the partition coefficient for the ith protein in the dextran–PEG–0.5 M sorbitol–0.01 M K/NaPB ATPS; $K_i^{0.5\text{ M TMAO}-0.01\text{ M K/NaPB}}$ is the partition coefficient for the ith protein in the dextran–PEG–0.5 M TMAO–0.01 M K/NaPB ATPS; $K_i^{0.5\text{ M trehalose}-0.01\text{ M K/NaPB}}$ is the partition

Table 2 Differences between the hydrophobic and electrostatic properties of the phases and partition coefficients for proteins in Dex–PEG–0.01 M K/NaPB, pH 7.4 and Dex–PEG–0.5 M osmolyte–0.01 M K/NaPB, pH 7.4 ATPS

	Difference between solvent properties of coexisting phases				
	0.01 M K/NaPB	0.5 M sorbitol	0.5 M sucrose[a]	0.5 M trehalose[a]	0.5 M TMAO[a]
$\Delta G(CH_2)^*$, cal mole^{-1}	−45 ± 1.2	−43 ± 1.1	−39.4 ± 0.44	−47.7 ± 0.6	−40.9 ± 0.6
E	0.033 ± 0.001	0.032 ± 0.002	0.029 ± 0.001	0.035 ± 0.001	0.028 ± 0.001
C	0.058 ± 0.003	0.090 ± 0.003	0.110 ± 0.002	0.113 ± 0.002	0.083 ± 0.002
$\Delta\pi^*$	−0.042 ± 0.002	−0.042 ± 0.004	−0.073 ± 0.004	−0.042 ± 0.003	−0.031 ± 0.002
$\Delta\alpha$	−0.051 ± 0.003	−0.066 ± 0.003	−0.046 ± 0.005	−0.081 ± 0.003	−0.074 ± 0.003
$\Delta\beta$	0.006 ± 0.004	0.006 ± 0.005	0.023 ± 0.006	0.006 ± 0.005	0.009 ± 0.008
Protein	**Partition coefficients**				
α-Chymotrypsin	0.42 ± 0.01	0.427 ± 0.008	0.42 ± 0.01	0.41 ± 0.01	0.42 ± 0.01
α-Chymotrypsinogen A	1.00 ± 0.01	1.5 ± 0.014	1.78 ± 0.02	1.93 ± 0.01	1.37 ± 0.02
Concanavalin A	0.236 ± 0.003	0.237 ± 0.003	0.242 ± 0.003	0.226 ± 0.003	0.233 ± 0.004
Hemoglobin human	0.129 ± 0.005	0.111 ± 0.002	0.118 ± 0.003	0.091 ± 0.002	0.208 ± 0.002
β-Lactoglobulin A	0.46 ± 0.01	0.329 ± 0.004	0.309 ± 0.004	0.255 ± 0.003	0.505 ± 0.005
β-Lactoglobulin B	0.33 ± 0.01	0.191 ± 0.003	0.211 ± 0.003	0.151 ± 0.003	0.27 ± 0.007
Lysozyme	0.23 ± 0.003	0.331 ± 0.004	0.325 ± 0.004	0.318 ± 0.002	0.255 ± 0.009
Papain	1.05 ± 0.01	1.29 ± 0.01	1.27 ± 0.01	1.37 ± 0.01	1.21 ± 0.02
Ribonuclease A	0.313 ± 0.005	0.329 ± 0.003	0.332 ± 0.006	0.311 ± 0.003	0.304 ± 0.006
Ribonuclease B	0.781 ± 0.004	0.334 ± 0.004	0.347 ± 0.005	0.318 ± 0.004	0.768 ± 0.004
Trypsinogen	0.357 ± 0.005	0.432 ± 0.009	1.463 ± 0.008	0.413 ± 0.006	0.431 ± 0.004

[a] Data reported in ref. 70 and 71 and presented for comparison.

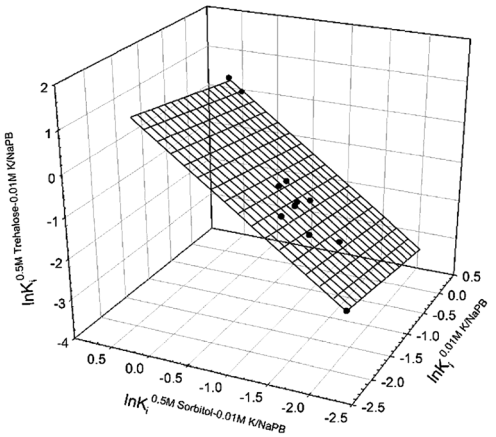

Fig. 1 Logarithms of partition coefficients for proteins in dextran–PEG–0.5 M trehalose–0.01 M K/NaPB ATPS *versus* those for the same proteins in dextran–PEG–0.01 M K/NaPB and in dextran–PEG–0.5 M sorbitol–0.01 M K/NaPB ATPS. K/NaPB – potassium/sodium phosphate buffer, pH 7.4.

coefficient for the same protein in dextran–PEG–0.5 M trehalose–0.01 M K/NaPB ATPS; all the other parameters are as defined above. Our analysis also revealed that the overall effects of trehalose and sucrose are very similar and that these two osmolytes can be used interchangeably in partition experiments.

It was suggested previously[54] that the relationships of the type represented by eqn (2)–(4) imply that the proteins respond to their environment in aqueous solutions depending on the environment and the protein structure. These relationships also seem to imply that the responses are governed by changes in the protein–water interactions and not to specific binding with the components of the environment.

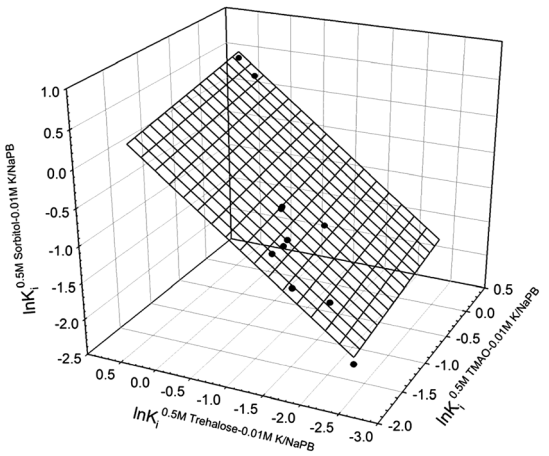

Fig. 2 Logarithms of partition coefficients for proteins in dextran–PEG–0.5 M trehalose–0.01 M K/NaPB ATPS *versus* those for the same proteins in the dextran–PEG–0.5 M TMAO–0.01 M K/NaPB ATPS and in dextran–PEG–0.5 M sorbitol–0.01 M K/NaPB ATPS. K/NaPB – potassium/sodium phosphate buffer, pH 7.4.

Protein–water interactions in the presence of polyol osmolytes and TMAO at concentration of 0.5 M

It has been shown earlier[54] that the partitioning of small organic compounds in ATPS with osmolytes additives vary in the osmolyte depending manner due to the osmolytes effects on the solvent properties of aqueous media. In order to explore if this conclusion is true for proteins in the presence of stabilizing osmolytes we examined the partition coefficients for proteins listed in Table 2 with eqn (1). The solute-specific coefficients (S_s, A_s, B_s and C_s) were calculated by multiple linear regression analysis with eqn (1), and the coefficients determined are presented in Table 3. It should be noted that we followed the procedure described by Ab Rani *et al.*[79] using the *p*-value as a test for significance for each solute-specific coefficient in eqn (1) for a given compound. In view of the extremely small number of five ATPSs utilized, we have chosen to use the maximum statistical significance value of $p \leq 0.1$. If all four coefficients (S_s, A_s, B_s and C_s) proved statistically significant ($p \leq 0.1$), then the correlation was accepted.

If one or more values reveal a *p*-value > 0.1, then equations contained different combinations of coefficients were examined. The equation with a set of coefficients providing *p*-values for all parameters below or equal to 0.1 was accepted. The solute-specific coefficients determined for each compound are presented in Table 3 together with the corresponding *p*-values (except the cases when $p < 0.001$).

The previously reported data for small polar organic compounds[54] demonstrated cooperativity between different types of solute–water interactions displayed as a linear interrelationship between different solute-specific coefficients.

Analysis of the data in Table 3 shows that similar cooperativity exists for proteins as well in agreement with the data reported previously.[72,74]

The interrelationship between solute-specific coefficients S_s, A_s, and B_s listed in Table 3 is plotted in Fig. 3. This relationship may be described as:

$$S_s^{\text{protein } i} = -0.1_{\pm 0.7} + 0.44_{\pm 0.03} \times A_s^{\text{protein } i} + 1.01_{\pm 0.08} \times B_s^{\text{protein } i},\ N = 11;\ R^2 = 0.9635;\ \text{SD} = 1.1;\ F = 105.7 \quad (5)$$

where $A_s^{\text{protein } i}$, $B_s^{\text{protein } i}$, and $S_s^{\text{protein } i}$ are the solute-specific coefficients A_s, B_s and S_s for *i*th protein, correspondingly; all the other parameters are as defined above.

Further analysis of the data in Table 3 indicates that there is a linear relationship between the solute-specific coefficients C_s and B_s illustrated graphically in Fig. 4 and described as:

$$C_s^{\text{protein } i} = -3.5_{\pm 0.6} + 0.98_{\pm 0.08} \times B_s^{\text{protein } i},\ N = 8;\ R^2 = 0.9624;\ \text{SD} = 1.0;\ F = 153.7 \quad (6)$$

where $C_s^{\text{protein } i}$ is the solute-specific coefficient C_s for *i*th protein; all the other parameters are as defined above (it should be noted that the data for α-chymotrypsinogen, β-lactoglobulin B, and papain do not fit the relationship).

We have established[54] that there is a linear relationship between solute-specific coefficients C_s, B_s, and S_s for polar organic compounds in the presence of 0.01 M K/NaPB, pH 7.4.

Table 3 Solute-specific coefficients[a] in eqn (1) for proteins indicated (calculated by multiple linear regression analysis from data in Table 2)

Protein	S_s	A_s	B_s	C_s	SD, F
α-Chymotrypsinogen *p*-values[b]	5.05 ± 0.03	−0.60 ± 0.08, 0.017	4.42 ± 0.02	7.60 ± 0.02	0.0005; 171 443
β-Lactoglobulin A *p*-values[b]	6.0 ± 0.4, 0.0002	21.0 ± 1.0, 0.002	−3.1 ± 0.3, 0.008	−6.3 ± 0.3, 0.002	0.006; 6214
β-Lactoglobulin B *p*-values[b]	5.0 ± 1.0, 0.04	17 ± 3, 0.01	0[c]	−6.7 ± 0.5, 0.0009	0.02; 1274
RNase A *p*-values[b]	7.61 ± 0.05	−1.8 ± 0.1, 0.006	8.83 ± 0.04	4.77 ± 0.04	0.001; 360 758
RNase B *p*-values[b]	6.9 ± 0.3, 0.0002	0[c]	7.4 ± 0.3	3.5 ± 0.3, 0.001	0.007; 7036
Papain *p*-values[b]	2.0 ± 0.5, 0.02	0[c]	0.9 ± 0.4, 0.07	2.7 ± 0.4, 0.008	0.01; 153.2
Trypsinogen *p*-values[b]	8 ± 1, 0.004	0[c]	8.2 ± 0.8, 0.002	6 ± 1, 0.01	0.02; 414.9
Lysozyme *p*-values[b]	13 ± 1, 0.003	0[c]	13 ± 1, 0.001	10 ± 1, 0.005	0.03; 507.3
Chymotrypsin	6.00 ± 0.02	0[c]	5.88 ± 0.02	3.01 ± 0.02	0.0006; 748 714
Concanavalin A *p*-values[b]	9.8 ± 0.3	0[c]	10.1 ± 0.2	5.1 ± 0.2, 0.0002	0.006; 19 303
Hemoglobin *p*-values[b]	20.7 ± 0.5	35 ± 1.3, 0.0001	4.5 ± 0.2, 0.0001	0[c]	0.009; 17 992

[a] Solute specific coefficients represent the following solute–water interactions: S_s – dipole–dipole interactions; A_s – hydrogen bonding with solute as a donor; B_s – hydrogen bonding with solute as an acceptor; C_s – induced dipole–ion interactions. [b] Statistical significance *p*-value (not shown for $p < 0.0001$). [c] 0, solute-specific coefficients could not be reliably determined (with $p < 0.1$) and in subsequent calculations are taken as 0.

In the case of nonionic polar compounds the value of solute-specific coefficient C_s might be explained by dipole–ion interactions[54] and cooperativity between this type of interactions and dipole–dipole and hydrogen-bonding solute–solvent interactions. In the case of proteins with multiple ionizable and nonionic polar groups it is difficult to separate ion–ion, ion–dipole, and dipole–dipole interactions, and therefore it might be expected that the relationship under discussion would be much less clear cut and not hold for different proteins.

In order to compare the solute-specific coefficients for proteins determined in the presence of 0.01 M K/NaPB, pH 7.4 with those for the same proteins determined in different ionic environment we re-calculated the previously reported data[72] using the *p*-value based approach described above.

The solute-specific coefficients for the proteins examined in ref. 72 are presented in Table 4 with the corresponding *p*-values.

Analysis of the data in Table 4 shows the cooperativity between the solute-specific coefficients for the proteins illustrated graphically in Fig. 5A and B and described as:

$$S_{sj}^{\text{protein } i} = -0.3_{\pm 0.3} + 0.4_{\pm 0.1} \times A_{sj}^{\text{protein } i} + 0.85_{\pm 0.03} \times B_{sj}^{\text{protein } i},$$
$$N = 13;\ R^2 = 0.9915;\ \text{SD} = 0.55;\ F = 580.7 \quad (7)$$

where subscript *j* denotes the ionic composition: 0.15 M NaCl in 0.01 M NaPB, pH 7.4; all the other parameters are as defined above; and

$$C_{sj}^{\text{protein } i} = -0.7_{\pm 0.8} + 2.5_{\pm 0.5} \times B_{sj}^{\text{protein } i} - 1.9_{\pm 0.5} \times S_{sj}^{\text{protein } i},$$
$$N = 13;\ R^2 = 0.9068;\ \text{SD} = 1.5;\ F = 48.7 \quad (8)$$

where all the parameters are as defined above.

Comparison of the above relationships observed for two sets of proteins (including six same proteins) in different ionic environments shows that both regression coefficients in eqn (5) and (7) at A_s parameter are identical within the error limits, and

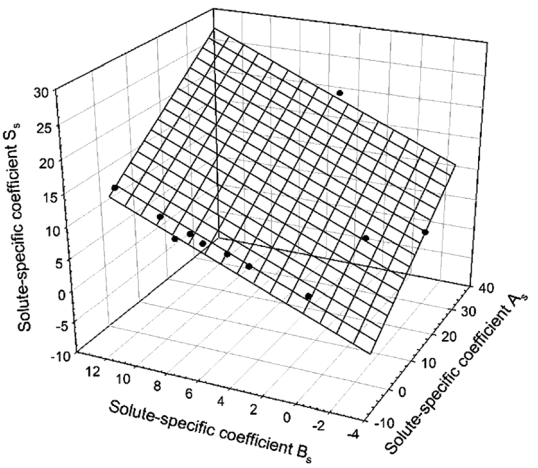

Fig. 3 Solute-specific coefficients $S_s^{\text{protein } i}$ determined for *i*th protein *versus* solute-specific coefficients $A_s^{\text{protein } i}$ and $B_s^{\text{protein } i}$ for the same proteins determined in the presence of in the presence of 0.01 M K/NaPB. K/NaPB – sodium/potassium phosphate buffer, both pH 7.4.

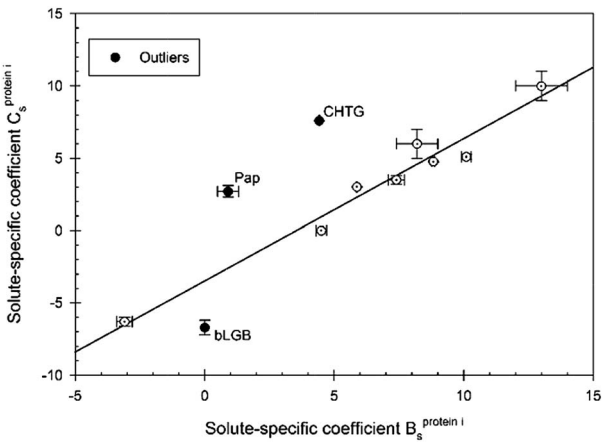

Fig. 4 Solute-specific coefficients $C_s^{\text{protein } i}$ *versus* solute-specific coefficients $B_s^{\text{protein } i}$ determined for proteins in the presence of in the presence of 0.01 M K/NaPB. K/NaPB – sodium/potassium phosphate buffer, both pH 7.4.

Table 4 Solute-specific coefficients[a] in eqn (1) for proteins in the presence of 0.15 M NaCl in 0.01 M NaPB, pH 7.4 (calculated by multiple linear regression analysis from data in ref. 72)

Protein	S_s	A_s	B_s	C_s
α-Chymotrypsinogen *p*-values[b]	−4 ± 1, 0.01	−4.0 ± 1.4, 0.03	−1.5 ± 0.8, 0.1	0[c]
γ-Globulin human *p*-values[b]	9 ± 2.7, 0.02	0[c]	12 ± 2.1, 0.001	13 ± 2.7, 0.003
γ-Globulin bovine *p*-values[b]	10 ± 2.8, 0.02	0[c]	13 ± 2.1, 0.002	12 ± 3, 0.01
Hemoglobin bovine *p*-values[b]	7 ± 1.3, 0.002	0[c]	8 ± 1, 0.0002	7 ± 1.3, 0.001
Hemoglobin human *p*-values[b]	5 ± 1.2, 0.006	0[c]	6.1 ± 0.9, 0.0005	6 ± 1.2, 0.002
β-Lactoglobulin *p*-values[b]	8 ± 1.7, 0.003	0[c]	9 ± 1.3, 0.0005	5 ± 1.7, 0.02
Lipase *p*-values[b]	1.8 ± 0.1, 0.0002	0.6 ± 0.2, 0.02	1.8 ± 0.1	0.6 ± 0.1, 0.006
Lysozyme *p*-values[b]	−4.3 ± 0.5, 0.0002	−3.7 ± 0.9, 0.006	−3.1 ± 0.4, 0.0004	2.5 ± 0.5, 0.003
Myoglobin *p*-values[b]	6.3 ± 0.7, 0.0003	3 ± 1.1, 0.08	6.7 ± 0.6	2.9 ± 0.8, 0.01
RNase A *p*-values[b]	2.4 ± 0.3, 0.0003	0[c]	3.4 ± 0.3	2.5 ± 0.3, 0.0002
RNase B *p*-values[b]	2.4 ± 0.5, 0.002	0[c]	3.6 ± 0.4	2.3 ± 0.4, 0.0002
Transferrin *p*-values[b]	14 ± 2, 0.0004	0[c]	16 ± 1.5	12 ± 1.9, 0.0009
Trypsinogen *p*-values[b]	−1.5 ± 0.7, 0.08	−3 ± 1.1, 0.03	0[c]	2.4 ± 0.8, 0.02

[a] Solute specific coefficients represent the following solute–water interactions: S_s – dipole–dipole interactions; A_s – hydrogen bonding with solute as a donor; B_s – hydrogen bonding with solute as an acceptor; C_s – induced dipole–ion interactions. [b] Statistical significance *p*-value (not shown for $p < 0.0001$). [c] 0, solute-specific coefficients could not be reliably determined (with $p < 0.1$) and in subsequent calculations are taken as 0.

the regression coefficient at B_s parameter increases ~1.2-fold with increasing salt concentration in the protein environment. The possible reason of the difference between the regression coefficients at B_{sj} in eqn (5) and (7) may be that the partitioning of proteins described by eqn (5) and (7) was examined under different ionic composition of the media – in the presence of 0.01 M K/Na–phosphate buffer (eqn (5)) and in the presence of 0.15 M NaCl in 0.01 M Na–phosphate buffer (eqn (7)). Both relationships (eqn (6) and (8)) for C_s parameter are less reliable than those for S_s parameter described by eqn (5) and (7), likely due to the aforementioned reasons.

Correlations between the proteins partition behavior and protein structure

It should be noted that analysis of the proteins partition coefficients in the presence or absence of osmolytes shows that there is no relationship between partition behavior and molecular weight and pI of the proteins studied in this work. It agrees with the data reported[55,56] previously for protein partitioning in PEG–sodium sulfate and dextran–PEG ATPS with different salts additives.

We computed Pearson correlation coefficients (PCCs) between each of the 57 descriptors that characterize structural properties of the considered proteins (see Materials and Methods) and the observed values of S_s to investigate whether the observed partition-based solute-specific coefficients correlate with these structural properties. In view of linear relationships observed between different solute-specific coefficients for proteins the analysis under consideration may be performed for any single solute-specific coefficient.

We chose solute-specific coefficient S_s as the one with non-zero values for all the proteins. Seven of the structural descriptors have modest correlations above 0.3 and three of them have high correlations above 0.5.

Most of the correlated structural parameters are related to some characteristics of the protein surface. They include the amount of positively charged, neutral, and negatively charged residues on the surface (Fig. 6A), hydrophobicity of the surface residues (Fig. 6B), size and volume of pockets on the surface, and content of β-sheets in the protein fold. Fig. 6A shows that normalized (by protein size) amount of the positively charged residues on the surface is negatively correlated with S_s (PCC = −0.43), while the normalized amount of negatively charged surface residues is positively correlated (PCC = 0.73). Average hydrophobicity of the surface residues is positively correlated with S_s (PCC = 0.53), suggesting that higher value of S_s corresponds to higher hydrophobicity of the surface.

In short, our empirical analysis suggests that the partition behavior of a given protein is determined by the peculiarities of its surface.

Regression model for solute-specific coefficient S_s

The regression modeling (see Materials and Methods for details) results in the model that uses four descriptors (Table 5): x_1 molecular mass of the protein computed with YASSARA; x_2 normalized (by the sequence length) number of positively charged amino acids on the surface (similar to ref. 56, a residue is defined to be on the surface if its solvent accessible surface area computed with DSSP > 0.746; this threshold was selected empirically to maximize the PCC with the observed data using cross validation); x_3 normalized (by the sequence length) area of protein pockets extracted with CASTp; x_4 sum of volume of all protein pockets generated by CASTp. The corresponding regression is formulated as follows:

$$y = -0.0004_{\pm 39.96}x_1 - 781.6914_{\pm 45.05}x_2 - 0.9639_{\pm 69.86}x_3 + 0.0040_{\pm 103.19}x_4 + 21.1173_{\pm 68.79} \quad (9)$$

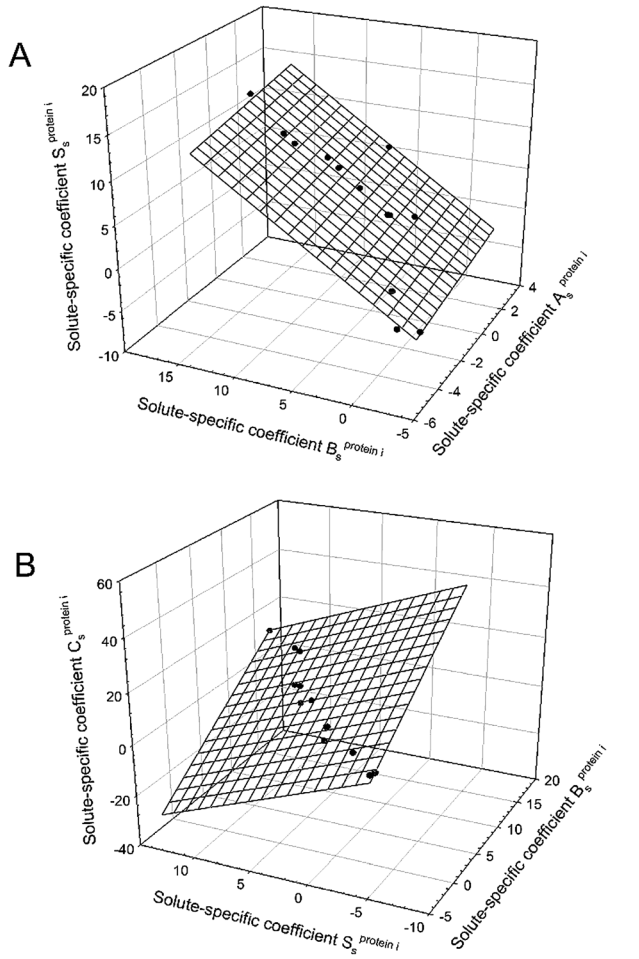

Fig. 5 A) Solute-specific coefficients $S_s^{protein\ i}$ determined for *i*th protein *versus* solute-specific coefficients $A_s^{protein\ i}$ and $B_s^{protein\ i}$ for the same proteins determined in the presence of in the presence of 0.15 M NaCl in 0.01 M NaPB. NaPB – sodium/potassium phosphate buffer, both pH 7.4. (B) Solute-specific coefficients $C_s^{protein\ i}$ determined for *i*th protein *versus* solute-specific coefficients $B_s^{protein\ i}$ and $S_s^{protein\ i}$ for the same proteins determined in the presence of in the presence of 0.15 M NaCl in 0.01 M NaPB. NaPB – sodium/potassium phosphate buffer, both pH 7.4.

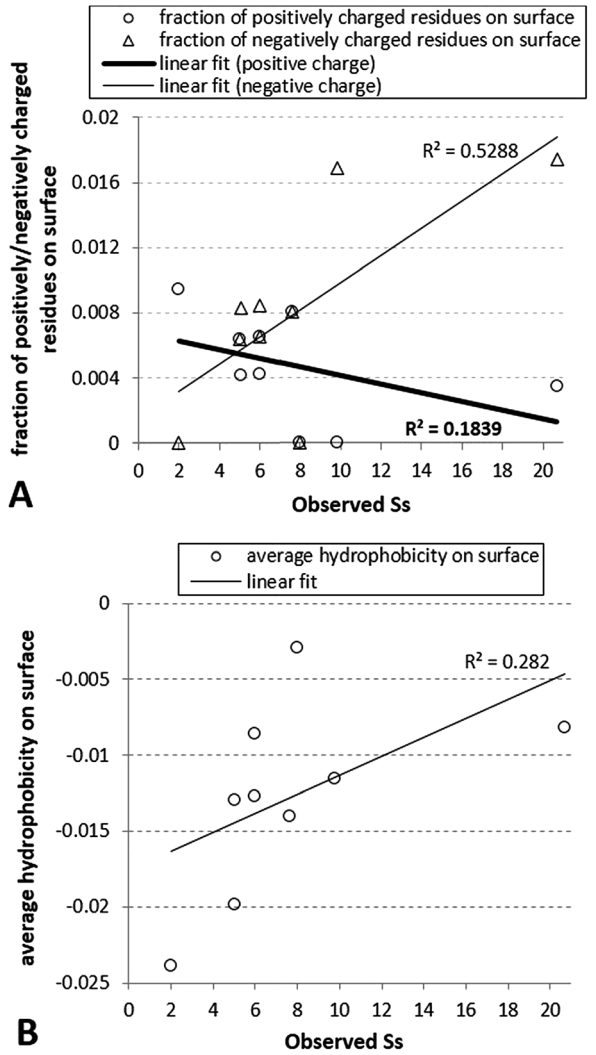

Fig. 6 Relation between the observed S_s values and values of selected descriptors derived based on the structural analysis of the protein structure. (A) Fraction of positively (circles) and negatively (triangles) charged residues on the protein surface. (B) Average hydrophobicity measured using the Eisenberg scale on the protein surface. Lines denote linear fit together with the corresponding R^2 value.

We report the standard errors for each estimated coefficient in the regression model. The five coefficients are statistically significant with *p*-values below 0.001. We also estimated the relative contributions of individual descriptors in the regression by normalizing their values and scaling the corresponding absolute values of coefficients to sum to 1. The recomputed absolute coefficients are −0.1774, −0.1046, −0.2811, and 0.4369, respectively. These values indicate that the three descriptors that characterize the surface (x_2, x_3 and x_4) are the main determinants of the value computed by the regression; their values sum up to $|-0.1046| + |-0.2811| + |0.4369| = 0.8226$ out of 1. This suggests that the value of S_s is primarily influenced by the properties of the surface of the protein including the positive charge and area and volume of pockets.

The outputs generated by the regression have high PCC with the observed data that equals 0.999 (0.998 based on the three-fold cross validation). This value is substantially larger than the PCC of 0.73 calculated for the best single descriptor, fraction of negatively charged residues on the surface (Fig. 6A).

The relation between the observed values of S_s and the values generated using regression is illustrated graphically in Fig. 7. The corresponding data for the three-fold cross validation analysis are also shown for comparison.

Our computational analyses reveal that the partition behavior of proteins is primarily determined by the peculiarities of their surfaces including positive charge and the area and volume of cavities on the surface. We observe that higher values of S_s correlate with lower numbers of positively charged surface residues and similarly they correlate with higher areas and volumes of pockets; last row in Table 5 shows the corresponding correlations.

Table 5 Structural descriptors used to find a correlation between the protein structure peculiarities and protein partition behavior and to build the regression model

		Descriptors				
Proteins	S_s	x_1 mass	x_2 fraction of positively charged residues on the surface	x_3 normalized area of pockets on the surface	x_4 volume of pockets on the surface	Output from regression
CHY	6.00	26 253.159	0.004219	8.307	1512.6	5.79
CHTG	5.05	32 534.909	0.004149	10.913	2529.6	5.00
bLGA	6.00	17 502.003	0.006579	9.944	1555.8	5.91
bLGB	5.00	17 850.429	0.006410	20.550	3891.0	5.06
RNaseA	7.61	14 390.182	0.008065	3.623	459.6	7.64
TRY	8.00	28 568.447	0.000000	7.508	1305.7	8.14
ConA	9.80	29 439.827	0.000000	0.000	0.0	9.80
Pap	2.00	25 539.630	0.009434	6.468	1101.6	2.11
HHb	20.70	31 490.614	0.003484	16.990	7669.8	20.70
Correlation with S_s		0.36	−0.43	0.23	0.70	0.99

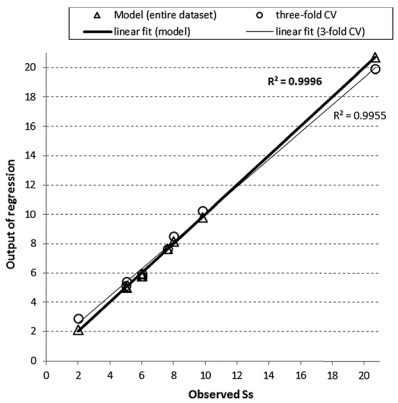

Fig. 7 Relation between values derived from the regression modeling and the corresponding observed values of S_s measured for 9 proteins. We compare regression models generated on the entire dataset (triangles) and based on the three-fold cross validation (CV) (circles). Lines denote linear fit together with the corresponding R^2 value.

Conclusions

The most important conclusion from the fact that the partition coefficients for proteins obtained in the presence of different osmolytes (sorbitol, sucrose, trehalose, TMAO)[54] and those obtained in the osmolyte-free ATPS is that at concentration of 0.5 M all these different osmolytes do not associate with the proteins examined and affect their partition behavior solely by changing solvent properties of aqueous media in the coexisting phases. Furthermore, the results obtained in our study show that the protective mechanism of stabilizing osmolytes is determined by the effects of osmolytes on the solvent properties of water translating into changes of protein–water interactions.

Acknowledgements

This work was supported in part by a grant from Russian Science Foundation RSCF No. 14-24-00131 (V. N. U.).

Notes and references

1 L. J. Borowitzka and A. D. Brown, *Arch. Microbiol.*, 1974, **96**, 37–52.
2 A. D. Brown and J. R. Simpson, *J. Gen. Microbiol.*, 1972, **72**, 589–591.
3 A. Pollard and R. G. W. Jones, *Planta*, 1979, **144**, 291–298.
4 P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus and G. N. Somero, *Science*, 1982, **217**, 1214–1222.
5 D. W. Bolen, *Methods Mol. Biol.*, 2001, **168**, 17–36.
6 P. H. Yancey and G. N. Somero, *Biochem. J.*, 1979, **183**, 317–323.
7 T. Arakawa, D. Ejima, Y. Kita and K. Tsumoto, *Biochim. Biophys. Acta, Bioenerg.*, 2006, **1764**, 1677–1687.
8 S. H. Khan, N. Ahmad, F. Ahmad and R. Kumar, *IUBMB Life*, 2010, **62**, 891–895.
9 J. A. Stewart and L. Ouellet, *Can. J. Chem.*, 1959, **37**, 744–750.
10 T. Arakawa, S. J. Prestrelski, W. C. Kenney and J. F. Carpenter, *Adv. Drug Delivery Rev.*, 2001, **46**, 307–326.
11 I. Baskakov and D. W. Bolen, *J. Biol. Chem.*, 1998, **273**, 4831–4834.
12 P. H. Yancey and G. N. Somero, *J. Exp. Zool.*, 1980, **212**, 205–213.
13 M. M. Santoro, Y. Liu, S. M. Khan, L. X. Hou and D. W. Bolen, *Biochemistry*, 1992, **31**, 5278–5283.
14 Y. Liu and D. W. Bolen, *Biochemistry*, 1995, **34**, 12884–12891.
15 A. Wang and D. W. Bolen, *Biochemistry*, 1997, **36**, 9101–9108.
16 G. N. Somero, *Am. J. Physiol.*, 1986, **251**, R197–R213.
17 J. C. Lee and S. N. Timasheff, *J. Biol. Chem.*, 1981, **256**, 7193–7201.
18 T. Arakawa and S. N. Timasheff, *Biochemistry*, 1982, **21**, 6536–6544.
19 T. Arakawa and S. N. Timasheff, *Arch. Biochem. Biophys.*, 1983, **224**, 169–177.
20 T. Arakawa and S. N. Timasheff, *Biophys. J.*, 1985, **47**, 411–414.

21 T. Arakawa, R. Bhat and S. N. Timasheff, *Biochemistry*, 1990, **29**, 1924–1931.
22 S. N. Timasheff, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9721–9726.
23 Y. Qu, C. L. Bolen and D. W. Bolen, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 9268–9273.
24 T. O. Street, D. W. Bolen and G. D. Rose, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 13997–14002.
25 M. Auton and D. W. Bolen, *Biochemistry*, 2004, **43**, 1329–1342.
26 M. Auton and D. W. Bolen, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15065–15068.
27 D. W. Bolen and I. V. Baskakov, *J. Mol. Biol.*, 2001, **310**, 955–963.
28 D. W. Bolen and G. D. Rose, *Annu. Rev. Biochem.*, 2008, **77**, 339–362.
29 A. Linhananta, S. Hadizadeh and S. S. Plotkin, *Biophys. J.*, 2011, **100**, 459–468.
30 K. Gekko and S. N. Timasheff, *Biochemistry*, 1981, **20**, 4667–4676.
31 K. Gekko and S. N. Timasheff, *Biochemistry*, 1981, **20**, 4677–4686.
32 G. S. Ratnaparkhi and R. Varadarajan, *J. Biol. Chem.*, 2001, **276**, 28789–28798.
33 R. P. Baptista, S. Pedersen, G. J. Cabrita, D. E. Otzen, J. M. Cabral and E. P. Melo, *Biopolymers*, 2008, **89**, 538–547.
34 L. Chen, G. J. Cabrita, D. E. Otzen and E. P. Melo, *J. Mol. Biol.*, 2005, **351**, 402–416.
35 M. Silow and M. Oliveberg, *J. Mol. Biol.*, 2003, **326**, 263–271.
36 P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
37 C. Tanford, *Adv. Protein Chem.*, 1968, **23**, 121–282.
38 A. Y. Grossberg and A. R. Khohlov, *Statistical Physics of Macromolecules*, Nauka, Moscow, 1989.
39 C. Tanford, *Physical Chemistry of Macromolecules*, Willey, New York, 1961.
40 O. Tcherkasskaya, E. A. Davidson and V. N. Uversky, *J. Proteome Res.*, 2003, **2**, 37–42.
41 O. Tcherkasskaya and V. N. Uversky, *Protein Pept. Lett.*, 2003, **10**, 239–245.
42 V. N. Uversky, *Protein Sci.*, 2002, **11**, 739–756.
43 V. N. Uversky, *Eur. J. Biochem.*, 2002, **269**, 2–12.
44 J. Rosgen, *Methods Enzymol.*, 2007, **428**, 459–486.
45 J. D. Batchelor, A. Olteanu, A. Tripathy and G. J. Pielak, *J. Am. Chem. Soc.*, 2004, **126**, 1958–1961.
46 M. Sajadi, Y. Ajaj, I. Ioffe, H. Weingartner and N. P. Ernsting, *Angew. Chem.*, 2010, **49**, 454–457.
47 M. Heyden, E. Brundermann, U. Heugen, G. Niehues, D. M. Leitner and M. Havenith, *J. Am. Chem. Soc.*, 2008, **130**, 5773–5779.
48 K. A. Sharp, B. Madan, E. Manas and J. M. Vandewrkooi, *J. Chem. Phys.*, 2001, **114**, 1791–1796.
49 M. V. Athawale, J. S. Dordick and S. Garde, *Biophys. J.*, 2005, **89**, 858–866.
50 S. Choudhary and N. Kishore, *J. Chem. Thermodyn.*, 2011, **43**, 1541–1551.
51 R. Politi, L. Sapir and D. Harries, *J. Phys. Chem. A*, 2009, **113**, 7548–7555.
52 S. Magazu, F. Migliardo and M. T. F. Telling, *Eur. Biophys. J.*, 2007, **36**, 163–171.
53 L. Larini and J. E. Shea, *J. Phys. Chem. B*, 2013, **117**, 13268–13277.
54 L. A. Ferreira, P. P. Madeira, V. N. Uversky and B. Y. Zaslavsky, *RSC Adv.*, 2015, DOI: 10.1039/c5ra08610h.
55 L. Ferreira, P. P. Madeira, L. Mikheeva, V. N. Uversky and B. Zaslavsky, *Biochim. Biophys. Acta, Bioenerg.*, 2013, **1834**, 2859–2866.
56 L. Ferreira, X. Fan, L. M. Mikheeva, P. P. Madeira, L. Kurgan, V. N. Uversky and B. Y. Zaslavsky, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 694–704.
57 M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser, *Methods Mol. Biol.*, 1999, **112**, 531–552.
58 M. J. Mizianty, W. Stach, K. Chen, K. D. Kedarisetti, F. M. Disfani and L. Kurgan, *Bioinformatics*, 2010, **26**, i489–496.
59 K. Rother, P. W. Hildebrand, A. Goede, B. Gruening and R. Preissner, *Nucleic Acids Res.*, 2009, **37**, D393–D395.
60 J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang, *Nucleic Acids Res.*, 2006, **34**, W116–W118.
61 K. W. Plaxco, K. T. Simons and D. Baker, *J. Mol. Biol.*, 1998, **277**, 985–994.
62 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
63 D. Juretic, D. Zucic, B. Lucic and N. Trinajstic, *Comput. Chem.*, 1998, **22**, 279–294.
64 D. Juretic, B. Lucic, D. Zucic and N. Trinajstic, *J. Theor. Comput. Chem.*, 1998, **5**, 405–445.
65 D. Eisenberg, R. M. Weiss and T. C. Terwilliger, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 140–144.
66 H. Cid, M. Bunster, M. Canales and F. Gazitua, *Protein Eng.*, 1992, **5**, 373–375.
67 M. J. Kamlet and R. W. Taft, *J. Am. Chem. Soc.*, 1976, **98**, 377–383.
68 R. W. Taft and M. J. Kamlet, *J. Am. Chem. Soc.*, 1976, **98**, 2886–2894.
69 M. J. Kamlet, J. L. Abboud and R. W. Taft, *J. Am. Chem. Soc.*, 1977, **99**, 6027–6038.
70 L. Breydo, A. E. Sales, L. Ferreira, O. Fedotoff, M. P. Shevelyova, S. E. Permyakov, K. G. Kroeck, E. A. Permyakov, B. Y. Zaslavsky and V. N. Uversky, *Arch. Biochem. Biophys.*, 2015, **570**, 66–74.
71 L. Ferreira, O. Fedotoff, V. N. Uversky and B. Y. Zaslavsky, *RSC Adv.*, 2015, **5**, 27154–27162.
72 P. P. Madeira, C. A. Reis, A. E. Rodrigues, L. M. Mikheeva, A. Chait and B. Y. Zaslavsky, *J. Chromatogr. A*, 2011, **1218**, 1379–1384.
73 P. P. Madeira, C. A. Reis, A. E. Rodrigues, L. M. Mikheeva and B. Y. Zaslavsky, *J. Phys. Chem. B*, 2010, **114**, 457–462.
74 P. P. Madeira, A. Bessa, L. Alvares-Ribeiro, M. R. Aires-Barros, C. A. Reis, A. E. Rodrigues and B. Y. Zaslavsky, *J. Chromatogr. A*, 2012, **1229**, 38–47.
75 P. P. Madeira, A. Bessa, D. P. C. de Banos, M. A. Teixeira, L. Alvares-Ribeiro, M. R. Aires-Barros, A. E. Rodrigues,

A. Chait and B. Y. Zaslavsky, *J. Chromatogr. A*, 2013, **1271**, 10–16.
76 P. P. Madeira, A. Bessa, L. Alvares-Ribeiro, M. R. Aires-Barros, A. E. Rodrigues and B. Y. Zaslavsky, *J. Chromatogr. A*, 2013, **1274**, 82–86.
77 P. P. Madeira, A. Bessa, L. Alvares-Ribeiro, M. R. Aires-Barros, A. E. Rodrigues, V. N. Uversky and B. Y. Zaslavsky, *J. Biomol. Struct. Dyn.*, 2014, **32**, 959–968.
78 P. P. Madeira, A. Bessa, M. A. Teixeira, L. Alvares-Ribeiro, M. R. Aires-Barros, A. E. Rodrigues and B. Y. Zaslavsky, *J. Chromatogr. A*, 2013, **1322**, 97–104.
79 M. A. Ab Rani, A. Brant, L. Crowhurst, A. Dolan, M. Lui, N. H. Hassan, J. P. Hallett, P. A. Hunt, H. Niedermeyer, J. M. Perez-Arlandis, M. Schrems, T. Welton and R. Wilding, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16831–16840.