Biochimica et Biophysica Acta 1865 (2017) 113–120

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

What are the structural features that drive partitioning of proteins in aqueous two-phase systems?

Zhonghua Wu [a], Gang Hu [a], Kui Wang [a], Boris Yu. Zaslavsky [b,*], Lukasz Kurgan [c,**], Vladimir N. Uversky [d,e,***]

[a] School of Mathematical Sciences and LPMC, Nankai University, Tianjin, People's Republic of China
[b] Cleveland Diagnostics, 3615 Superior Avenue, Suite 4407B, Cleveland, OH 44114, USA
[c] Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
[d] Department of Molecular Medicine and Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[e] Laboratory of structural dynamics, stability and folding of proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

ARTICLE INFO

Article history:
Received 25 June 2016
Received in revised form 26 August 2016
Accepted 18 September 2016
Available online 24 September 2016

Keywords:
Aqueous two phase system
Partition
Protein structure
Protein structural feature

ABSTRACT

Protein partitioning in aqueous two-phase systems (ATPSs) represents a convenient, inexpensive, and easy to scale-up protein separation technique. Since partition behavior of a protein dramatically depends on an ATPS composition, it would be highly beneficial to have reliable means for (even qualitative) prediction of partitioning of a target protein under different conditions. Our aim was to understand which structural features of proteins contribute to partitioning of a query protein in a given ATPS. We undertook a systematic empirical analysis of relations between 57 numerical structural descriptors derived from the corresponding amino acid sequences and crystal structures of 10 well-characterized proteins and the partition behavior of these proteins in 29 different ATPSs. This analysis revealed that just a few structural characteristics of proteins can accurately determine behavior of these proteins in a given ATPS. However, partition behavior of proteins in different ATPSs relies on different structural features. In other words, we could not find a unique set of protein structural features derived from their crystal structures that could be used for the description of the protein partition behavior of all proteins in all ATPSs analyzed in this study. We likely need to gain better insight into relationships between protein-solvent interactions and protein structure peculiarities, in particular given limitations of the used here crystal structures, to be able to construct a model that accurately predicts protein partition behavior across all ATPSs.

1. Introduction

Aqueous two-phase systems (ATPSs) are formed in mixtures of two or more components in water, when concentrations of these components exceed certain threshold concentrations or at particular temperature [1–3]. The phase forming components may include two polymers, such as dextran and polyethylene glycol (PEG), single polymer and salt or organic additive, such as PEG and sodium sulfate, citrate, surfactant, e.g., octylglucoside, or osmolyte, such as glycine betaine, for example. All these ATPSs are commonly used for separation or analysis of biomacromolecules, such as proteins or nucleic acids. Separation of proteins in polymer-salt ATPSs was recently reviewed in [4], and application of this method for analysis of proteins was discussed in detail in [5].

Commonly, in order to design appropriate conditions for extraction of a particular protein from a multicomponent mixture, such as fermentation broth or cell extract, it is necessary to screen a variety of ATPSs in order to select conditions providing required recovery and purification of the target protein [6–10]. Similarly for designing conditions for analysis of a given protein it is needed to screen different ATPSs in order to establish what ATPSs provide conditions for reliable differentiating between the target protein and its structurally altered variants [5,11].

It would be very beneficial to be able to predict (even qualitatively) partition behavior of a target protein under different conditions, since it would reduce time, labor, and quantity of proteins currently used for screening different partition conditions. Several attempts to predict partition behavior of proteins based mostly on their charge and hydrophobicity have been reported in the literature [12–20]. Charge and hydrophobicity of proteins undoubtedly play an important role in the protein partition behavior in ATPSs. The definition of hydrophobicity, however, remains an open question. The overall hydrophobicity of any compound is reduced with increasing its total charge, for example,

* Correspondence to B.Y. Zaslavsky, Cleveland Diagnostics, 3615 Superior Avenue, Suite 4407B, Cleveland, OH 44114, USA.
** Correspondence to: L. Kurgan, Department of Computer Science, School of Engineering, Virginia Commonwealth University, 401 West Main Street, Richmond, VA 23284, USA.
*** Correspondence to: V.N. Uversky, Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, FL 33612, USA.
E-mail addresses: boris.zaslavsky@cleveland-diagnostics.com (B.Y. Zaslavsky), lkurgan@vcu.edu (L. Kurgan), vuversky@health.usf.edu (V.N. Uversky).

http://dx.doi.org/10.1016/j.bbapap.2016.09.010
1570-9639/

and the same trend may be expected for proteins. It has been reported [21] that partition coefficients of the peptides of the same amino acid composition but different sequence differ 3-fold in PEG-600-Na_2SO_4 ATPS. It has been established [22,23] that partition behavior of proteins in a given ATPS is governed by the nature and spatial arrangement of the solvent exposed groups; i.e., 3D–structure of the proteins. It has been also found [24–27] that protein partitioning in an ATPS is driven by electrostatic, dipole-dipole, and hydrogen bonding interactions with aqueous media in the two phases of an ATPS. The relative contributions of these different types of interactions vary for different proteins in various ATPSs.

The aim of this work was to explore what structural features of proteins are important for the protein partition behavior in dextran-PEG and PEG-salt ATPS of different ionic composition.

2. Materials and methods

2.1. Proteins

We analyzed 10 structurally and functionally diverse proteins that were used in recent related studies [28,29]. These proteins were selected based on their availability and on the availability of high resolution X-ray structures. They include α-chymotrypsin, α-chymotrypsinogen A, ribonuclease A and trypsinogen from bovine pancreas, concanavalin A from *Canavalia ensiformis* (jack beans), human hemoglobin, β-lactoglobulin A and B from bovine milk, lysozyme from chicken egg white, and papain from papaya latex. The proteins were purchased from Sigma-Aldrich. Table 1 summarizes some basic properties of these proteins.

2.2. Aqueous two-phase systems

A mixture of polymers was prepared as described elsewhere [30] 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 buffer solutions, salt additive(s) and water were added to achieve the ionic and polymer composition required for the final system (after the sample addition — see below) with total weight of 0.5 g (total volume 457 ± 2 μL). Sodium phosphate buffer solution with pH 7.4 was used.

2.3. Partitioning

Partitioning experiments were performed at 23 °C using the Automated Signature Workstation, ASW (Analiza, Inc., Cleveland, OH, USA). 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. 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. 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 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.

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) 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 <3% and in most cases lower than 1%.

2.4. Protein descriptors

Similar to refs. [28,29], the proteins were comprehensively characterized based on 57 numerical descriptors derived from the corresponding sequences and structures; Table 1 lists identifiers of their crystal structures in the Protein Data Bank [31]. These features quantify physiochemical properties, tertiary and secondary structures, surface, intrinsic disorder and flexibility and they include:

- Length of the sequence (1 feature).
- Molecular weight (1 feature).
- Isoelectric point (pI) that was computed with the ExPASy server [32] (1 feature)
- Descriptors of intrinsic disorder predicted with the MFDp method [33]: disorder content (fraction of disordered residues in the sequence), normalized (by the chain length) number of disordered segments, and the average propensity of disorder (3 features).
- Properties of the tertiary protein structure computed with the Voronoia program [34] including average packing density, van der Waals volume, solvent-excluded volume, fraction of buried atoms, and size of internal cavities in the protein structure (14 features).
- Properties of the surface generated with the CASTp software [35], such as the number, surface area and volume of pockets on the protein surface (6 features)
- The contact order that quantifies packing of the structure [36]
- Characteristics of the tertiary and secondary structures derived with the YASARA program (http://www.yasara.org/) including radius of gyration, nuclear and van der Waals radii, molecular mass, content of six secondary structure types: α-helix, 3_{10}-helix, both helix types, β-sheet, turns and coils, as well as flexibility expressed with B-factor and occupancy (12 features).
- Properties of the surface and secondary structure computed with the DSSP software [37]. These properties include fraction of surface residues; fraction of polar, nonpolar, neutral, positively charged, and negatively charged residues on the surface; hydrophobicity of surface residues that was estimated based on three amino acids scales: Kyte-Doolittle [38,39], Eisenberg [40], and Cid [41]; and content of 8 secondary structure types: α-helix, 3_{10}-helix, all helix types, β-sheet, β-bridge, both β structure types, turn, bend, and coil (18 features).

Table 1
Proteins that were utilized in this study.

Protein	Abbreviation	Molecular weight [kDa]	PDB ID
α-Chymotrypsin	CHY	25.0	1AB9
α-Chymotrypsinogen A	CHTG	25.7	1ACB
Concanavalin A	ConA	104.0	1JBC
Hemoglobin human	HHb	64.5	1BZ0
β-Lactoglobulin A	bLGA	18.3	1B8E
β-Lactoglobulin B	bLGB	18.3	1BEB
Lysozyme	HEL	14.3	194L
Papain	Pap	23.4	1PPN
Ribonuclease A	Rnase A	17.0	1BEL
Trypsinogen	TRY	24.0	1BTY

The considered features are listed in Table 2. They are categorized into five types: physiochemical properties, tertiary structure, secondary structures, surface, and intrinsic disorder and flexibility.

2.5. Multivariate modeling

We use linear regression to model a relation between the partition coefficients and structural features of the proteins. The regression was derived to minimize sum of squared errors, defined as the difference between the experimentally measured values and the values derived from regression that uses a particular set of features as the input. Given the experimentally measured (observed) values from a given two-phase system $y_i \in R^{t\times 1}$ and the set of features $X \in R^{t\times n}$ where $t = 10$ is the number of proteins, $i = 1, 2, \ldots, 29$ is the index of a particular system, and $n = 1, 2, \ldots, 57$ is the number of considered features, the regression was optimized to minimize:

$$\min_k(\|k_0 + kX - y_{i2}^2\|) \quad (1)$$

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

We empirically selected subsets of the considered 57 feature that maximize the Pearson correlation coefficients (PCC) between the outputs of the regressions that use the subsets of features and the observed values of proteins' partition coefficients for a given aqueous two-phase system. The corresponding regressions describe relationships between the selected structural features and the logarithms of partition coefficients of the proteins given that their PCC values are high and statistically significant. Since the number of samples (proteins) is relatively low, we limited the number of selected features to 1, 2 and 3 and we considered all possible combinations. For each considered set of features for which PCC is statistically significant (p-value < 0.05) based on the Pearson correlation test or Spearman rank test, depending on the result of normality test, we performed further tests to ensure that the corresponding regression does not overfit the data. We repeated the significance test 100 times by randomly sampling with replacement a subset of 5 to 10 proteins; there are in total 638 sets of 5, 6, 7, 8, 9, and 10 proteins to consider. Given the assumed significance level at 0.05, feature sets for which the fraction of results where PCC is not significant is below 0.05 are retained.

Next, we investigated whether the regressions that use the remaining feature sets could be used to predict the observed values for a given types of solute-solvent interactions. We estimated PCCs between out-of-sample predictions generated by these regressions and the observed values. We performed 100 five-fold cross validations tests on the set of the 10 proteins to calculate the corresponding out-of-sample PCC values. In this cross validation, 8 proteins were used to derive a regression and the remaining two were used to test predictions with this regression. This was repeated five times, each time selecting a different set of two proteins as the test set. Each repetition of the cross validation randomly selects these five test sets. We used the average of these 100 cross validation-based PCCs, cvPCC, to quantify the predictive performance of a regression that uses a given set of features. In our analysis, we only considered regressions characterized by strong correlation, i.e., cvPCC > 0.9. Moreover, we tested whether these correlations are statistically significant. Only the regressions for which the fraction of PCCs that are not significant out of the 100 repetitions is below 0.05 were retained.

3. Results and discussion

3.1. Regression models for proteins partition coefficients

We performed empirical analysis to find regressions of up to three features that provide statistically significant (p-value < 0.05) and high (>0.9) correlations with the measured partition coefficients for each of the considered aqueous two-phase systems. Table 3 reports the

Table 2
List of the 57 descriptors of physiochemical properties (PP), tertiary structure (TS), secondary structures (SS), surface (SF), and intrinsic disorder and flexibility (DF) of the proteins.

Index	Type	Description	Index	Type	Description
x_1	PP	Length of the protein sequence	x_{30}	SS	Turn content computed with DSSP
x_2	PP	Molecular weight (sum of mass of all residues in the protein chain)	x_{31}	SS	Bend content computed with DSSP
x_3	PP	Isoelectric point (pI) computed with ExPASy	x_{32}	SS	Helix (including α-, 3_{10} and π) content computed with DSSP
x_4	TS	Contact number defined in ref. [36]	x_{33}	SS	β (including bridge and sheet) content computed with DSSP
x_5	TS	Nuclear radius computed with YASARA	x_{34}	SS	Coil (including turn and bend) content computed with DSSP
x_6	TS	Van der Waals radius computed with YASARA	x_{35}	TS	Fraction of atoms in partially filled internal cavities (IC) computed by Voronoia
x_7	TS	Radius of gyration computed with YASARA	x_{36}	TS	Fraction of atoms neighboring partially filled ICs by Voronoia
x_8	SS	α-helix content computed with YASARA	x_{37}	TS	Fraction of atoms in empty ICs by Voronoia
x_9	SS	β-sheet content computed with YASARA	x_{38}	TS	Fraction of atoms neighboring empty ICs by Voronoia
x_{10}	SS	Turn content computed with YASARA	x_{39}	TS	Fraction of atoms in partially filled ICs excluding HETS by Voronoia
x_{11}	SS	Coil content computed with YASARA	x_{40}	TS	Fraction of atoms neighboring partially filled ICs excluding HETS by Voronoia
x_{12}	SS	3_{10}-helix content computed with YASARA	x_{41}	TS	Fraction of atoms in filled ICs excluding HETS by Voronoia
x_{13}	SS	α- and 3_{10} helix content computed with YASARA	x_{42}	TS	Fraction of atoms neighboring filled ICs excluding HETS by Voronoia
x_{14}	PP	Molecular mass computed with YASARA	x_{43}	TS	Fraction of atoms in all ICs computed by Voronoia
x_{15}	DF	B-factor computed with YASARA	x_{44}	TS	Fraction of atoms neighboring all ICs by Voronoia
x_{16}	DF	Occupancy computed with YASARA	x_{45}	TS	Avg van der Waals volume computed by Voronoia
x_{17}	SF	Fraction of solvent accessible (SA) residues computed with DSSP	x_{46}	TS	Avg solvent excluded volume computed by Voronoia
x_{18}	SF	Fraction of polar SA residues computed with DSSP	x_{47}	TS	Fraction of buried atoms computed by Voronoia
x_{19}	SF	Fraction of non-polar SA residues computed with DSSP	x_{48}	TS	Avg packing density computed by Voronoia
x_{20}	SF	Fraction of positively charged SA residues computed with DSSP	x_{49}	DF	Disorder content predicted with MFDp
x_{21}	SF	Fraction of neutrally charged SA residues computed with DSSP	x_{50}	DF	Normalized number of disordered segments predicted with MFDp
x_{22}	SF	Fraction of negatively charged SA residues computed with DSSP	x_{51}	DF	Avg propensity of disorder predicted with MFDp
x_{23}	SF	Avg hydrophobicity (Kyte-Doolittle index) of SA residues computed with DSSP	x_{52}	SF	Number of pockets on the protein surface computed with CASTp
x_{24}	SF	Avg hydrophobicity (Eisenberg index) of SA residues computed with DSSP	x_{53}	SF	Normalized number of pockets on the protein surface computed with CASTp
$\times_{25}$	SF	Avg hydrophobicity (Cid index) of SA residues computed with DSSP	x_{54}	SF	Area of pockets on the protein surface computed with CASTp
x_{26}	SS	α-helix content computed with DSSP	x_{55}	SF	Normalized area of pockets on the protein surface computed with CASTp
x_{27}	SS	3_{10}-helix content computed with DSSP	x_{56}	SF	Volume of pockets on the protein surface computed with CASTp
x_{28}	SS	β-bridge content computed with DSSP	x_{57}	SF	Normalized volume of pockets on the protein surface computed with CASTp
x_{29}	SS	β-sheet content computed with DSSP			

regressions that secure the highest out-of-sample PCC values, cvPCC, based on 100 five-fold cross validations for each system among the regressions satisfying the above condition. The regressions for each of the 29 systems are characterized by statistically significant and very high correlations at over 0.95 for both the in-sample experiments (based on fitting the entire set of 10 proteins) and the out-of-sample cross validations. This means that just a few structural characteristics of proteins can determine the behavior of these proteins in the considered two-phase systems.

The 29 different ATPSs were grouped into four clusters:

1) ATPS formed by PEG-8000 and dextran-70 (ATPS #1–17);
2) ATPS formed by PEG-8000 and Na_2SO_4 (ATPS #18–22);
3) ATPS formed by PEG-600 and Na_2SO_4 (ATPS #23–28), and
4) ATPS #29 formed by PEG-600 and sodium phosphate buffer (NaPB).

Detailed analysis of the best performing regressions (having highest cvPCC) for one representative ATPS in each cluster (systems #3, 19, 23 and 29) is shown in Table 4; these systems are highlighted using bold font in Table 3. Analysis of the data presented in Table 4 by multiple linear regression shows that they are described as:

$$\begin{aligned}\text{ATPS}\ \#3 : \log K &= -4.491_{\pm 0.545} + 0.041_{\pm 0.006}x_{16} - 20.974_{\pm 2.644}x_{17} \\ &\quad + 0.227_{\pm 0.010}x_{28} \\ N = 10;\ r^2 &= 0.995;\ SD = 0.049;\ F = 384;\ p < 0.001\end{aligned} \tag{2a}$$

$$\begin{aligned}\text{ATPS}\#19 : \log K &= 60.528_{\pm 3.795} + 8.062_{\pm 0.582}x_{25} - -4.321_{\pm 0.272}x_{45} \\ &\quad + 0.0737_{\pm 0.0079}x_{55} \\ N = 10;\ r^2 &= 0.985;\ SD = 0.103;\ F = 134;\ p < 0.001\end{aligned} \tag{2b}$$

$$\begin{aligned}\text{ATPS}\#23 : \log K &= -1.905_{\pm 0.146} + 0.325_{\pm 0.015}x_3 - 0.157_{\pm 0.025}x_4 \\ &\quad + 0.0416_{\pm 0.010}x_{27} \\ N = 10;\ r^2 &= 0.989;\ SD = 0.077;\ F = 188;\ p < 0.001\end{aligned} \tag{2c}$$

$$\begin{aligned}\text{ATPS}\#29 : \log K &= -0.115_{\pm 0.039} - 46.522_{\pm 5.281}x_{22} + 5.655_{\pm 1.069}x_{23} \\ &\quad + 1.838_{\pm 0.235}x_{53} \\ N = 10;\ r^2 &= 0.978;\ SD = 0.044;\ F = 89;\ p < 0.001\end{aligned} \tag{2d}$$

where K — partition coefficient for a protein; x is the structural feature of the protein as defined in Table 2; N — number of proteins; r^2 — correlation coefficient; SD — standard deviation; F — ratio of variance; p — statistical significance (*p*-value for each coefficient in the above equations is mostly <0.001, and in the worst case <0.006).

In Table 4, the values of the structural features for proteins and the calculated logarithms of proteins partition coefficients are listed together with the corresponding observed values. Some of the structural features of the proteins have modest to high correlations with the partition coefficients. The correlation grows to very high levels once multiple (three) features are combined using the linear model. The calculated log*K* values are not only highly correlated with the observed values but also provide an accurate approximation. The average absolute error for ATPSs # 3, 19, 23 and 29 is only 0.03, 0.07, 0.05, and 0.03, respectively. These are very small errors considering that the range of the observed values is 1.5 (2% error), 2.3 (3% error), 1.6 (3% error) and 0.8 (4% error), respectively.

Fig. 1 illustrates graphically relation between the calculated and the measured partition coefficients of the proteins in the four selected ATPSs (# 3, 19, 23 and 29). The estimates generated by the regression that uses just three selected features are very similar to the actual

Table 3
List of the 29 considered solute-solvent systems together with regressions that fit values of their measured coefficients using an empirically selected set of three input features x_n (Table 2 summarizes features). For each system we show the regression that secures the highest value of cvPCC (average over 100 cross validation-based PCCs between the outputs of the regressions and the observed values), PCC on the entire protein set and *p*-value. Systems are grouped into four clusters given using shading. A representative system from each cluster is shown using bold font.

ATPS#	Description of the solute – solvent system	Regression	cvPCC	PCC	*p*-value
1	Dex PEG 8000–0.01 M NaPB	$2.793820 - 5.635150x_{21} - 26.876400x_{22} + 53.583400x_{36}$	0.95	0.99	$2.13E^{-11}$
2	Dex PEG 8000–0.11 M NaPB	$0.984056 - 115.297000x_{17} - 4.461890x_{20} - 16.819300x_{44}$	0.97	0.99	$1.40E^{-12}$
3	**0.01 M NaPB – 0.15 M NaCl**	$\mathbf{-4.490950 + 0.041001x_{16} - 20.973600x_{17} + 0.226591x_{28}}$	**0.99**	**1.00**	$\mathbf{1.45E^{-18}}$
4	0.01 M NaPB – 1.05 M NaCl	$-1.255040 + 0.452619x_{28} - 0.024110x_{52} + 0.064414x_{57}$	0.98	0.99	$5.29E^{-13}$
5	0.01 M NaPB – 0.083 M Na2SO4	$0.289463 - 93.583000x_{17} + 0.118733x_{28} - 15.071000x_{44}$	0.97	0.99	$5.59E^{-12}$
6	0.11 M NaPB – 0.083 M Na2SO4	$0.331989 - 86.959000x_{17} + 0.111770x_{28} - 15.554000x_{44}$	0.97	0.99	$1.87E^{-11}$
7	0.01 M NaPB – 0.30 M Na2SO4	$0.949488 - 0.000039x_{02} - 51.374800x_{17} + 45.969400x_{41}$	0.95	0.98	$5.31E^{-10}$
8	0.01 M NaPB – 0.17 M CsCl	$-1.158490 + 0.353765x_{28} - 0.018499x_{52} + 0.041672x_{57}$	0.99	0.99	$1.56E^{-15}$
9	0.11 M NaPB – 0.17 M CsCl	$-0.625500 - 105.114000x_{21} + 0.089028x_{28} + 24.963300x_{41}$	0.98	0.99	$6.91E^{-14}$
10	0.01 M NaPB – 0.80 M CsCl	$-1.270210 + 0.443420x_{28} - 0.025232x_{52} + 0.063328x_{57}$	0.98	0.99	$3.73E^{-14}$
11	0.01 M NaPB – 0.094 M NaClO4	$-1.189630 - 88.404500x_{21} + 0.082040x_{27} + 0.311406x_{28}$	0.96	0.99	$2.83E^{-11}$
12	0.11 M NaPB – 0.094 M NaClO4	$4.262780 - 0.099302x_{29} - 0.064040x_{32} - 9.731340x_{42}$	0.98	0.99	$1.58E^{-14}$
13	0.01 M NaPB – 0.43 M NaClO4	$-39.857700 + 0.084056x_{10} - 36.575700x_{47} + 86.959800x_{48}$	0.99	1.00	$1.55E^{-22}$
14	0.01 M NaPB – 0.17 M NaSCN	$-1.1888900 + 0.450418x_{28} - 0.030070x_{52} + 0.061609x_{57}$	0.99	0.99	$5.59E^{-15}$
15	0.11 M NaPB – 0.17 M NaSCN	$-32.677100 + 0.060262x_{10} - 24.879700x_{47} + 67.331800x_{48}$	0.96	0.99	$6.99E^{-11}$
16	0.01 M NaPB – 1.26 M NaSCN	$-36.261000 + 0.062149x_{10} - 30.658600x_{47} + 77.460600x_{48}$	0.98	0.99	$3.18E^{-15}$
17	0.11 M NaPB – 0.35 M NaSCN	$-36.745900 + 0.064237x_{10} - 29.663200x_{47} + 77.237000x_{48}$	0.99	1.00	$1.51E^{-11}$
18	PEG 8000 Na2SO4–0.01 M UB	$5.091420 + 0.040379x_{05} + 150.588000x_{19} - 0.742383x_{46}$	0.96	0.99	$6.10E^{-12}$
19	**PEG 8000 Na2SO4–0.15 M NaCl + 0.01 M UB**	$\mathbf{60.528300 + 8.062460x_{25} - 4.321020x_{45} + 0.073699x_{55}}$	**0.97**	**0.99**	$\mathbf{1.27E^{-13}}$
20	PEG 8000 Na2SO4–0.33 M NaCl + 0.01 M UB	$-14.591200 + 0.135864x_{16} + 38.145800x_{23} + 0.327895x_{28}$	0.95	0.98	$9.25E^{-10}$
21	PEG 8000 Na2SO4–0.15 M NaSCN + 0.01 M UB	$7.393290 + 0.466832x_{03} + 0.074593x_{12} - 18.500800x_{21}$	0.98	0.99	$3.96E^{-15}$
22	PEG 8000 Na2SO4–0.40 M NaSCN + 0.01 M UB	$11.506200 - 0.157056x_{13} - 0.254760x_{33} - 27.372000x_{42}$	0.92	0.98	$3.76E^{-09}$
23	**PEG 600 Na2SO4–0.15MNaCl + 0.058 M K/NaPB**	$\mathbf{-1.905130 + 0.325039x_{03} - 0.157164x_{04} + 0.041578x_{27}}$	**0.99**	**1.00**	$\mathbf{4.24E^{-15}}$
24	PEG 600 Na2SO4–0.085 M K/NaPB	$-1.572490 + 0.260153x_{03} - 2.091270x_{51} + 0.000063x_{56}$	0.99	1.00	$4.88E^{-17}$
25	PEG 600 Na2SO4–0.15 M NaCl + 0.085 M K/NaPB	$-0.975663 + 0.364431x_{03} - 0.033699x_{11} - 100.733000x_{18}$	0.98	0.99	$5.04E^{-15}$
26	PEG 600 Na2SO4–0.4 M NaSCN + 0.085 M K/NaPB	$9.477500 + 0.547497x_{03} - 0.056105x_{11} - 0.117209x_{16}$	0.97	0.99	$3.83E^{-11}$
27	PEG 600 Na2SO4–0.17 M K/NaPB	$-0.899874 + 0.342403 \times {}_{03} - 0.026658 \times {}_{11} - 97.861500x_{18}$	0.99	1.00	$3.34E^{-16}$
28	PEG 600 Na2SO4–0.15 M NaCl + 0.15 M K/NaPB	$37.838000 + 173.344000 \times {}_{25} - 2.654030x_{45} + 0.020199x_{52}$	0.97	0.99	$4.10E^{-12}$
29	**Dex PEG 600–0.14 M NaPB**	$\mathbf{-0.115000 - 46.522000x_{22} + 5.655000x_{23} + 1.838000x_{53}}$	**0.96**	**0.99**	$\mathbf{8.63E^{-12}}$

Table 4
Features used to quantify relations between peculiarities of protein structure and protein partition behavior for the selected four representative solute-solvent systems. We include the values of features associated with the regression that secures the highest value of cvPCC, the values outputted by this regression (Reg) and the corresponding observed values (Obs) of the solute-specific coefficients. The last row lists PCC between the observed values and the values of features and outputs of regressions.

Proteins	System 3					System 19					System 23					System 29				
	Features			Reg	Obs	Features			Reg	Obs	Features			Reg	Obs	Features			Reg	Obs
	x_{16}	x_{17}	x_{28}			x_{25}	x_{45}	x_{55}			x_3	x_4	x_{27}			x_{22}	x_{23}	x_{53}		
α-Chymotrypsin	95	0.021	4.219	−0.082	−0.066	−0.153	14.099	8.307	−1.019	−0.996	8.330	2.486	3.797	0.570	0.485	0.000	−0.015	0.158	0.091	0.130
α-Chymotrypsinogen A	96	0.008	4.979	0.399	0.384	−0.185	13.974	10.913	−0.542	−0.490	8.520	2.035	3.734	0.700	0.758	0.000	0.000	0.147	0.155	0.118
Concanavalin A	96	0.025	0.844	−0.895	−0.876	−0.154	13.970	0.000	−1.076	−1.181	5.260	2.603	3.797	−0.447	−0.472	0.008	−0.030	0.000	−0.657	−0.658
Hemoglobin human	100	0.014	0.000	−0.683	−0.684	−0.132	14.253	16.990	−0.874	−0.947	7.920	2.160	9.408	0.721	0.703	0.000	−0.003	0.125	0.097	0.163
β-Lactoglobulin A	100	0.033	0.000	−1.081	−1.143	−0.159	14.348	9.944	−2.022	−1.959	4.830	4.108	6.579	−0.707	−0.777	0.000	−0.023	0.148	0.027	0.000
β-Lactoglobulin B	100	0.026	0.000	−0.929	−0.860	−0.221	14.262	20.550	−1.365	−1.367	4.830	3.990	6.410	−0.696	−0.606	0.006	−0.023	0.247	−0.071	−0.070
Lysozyme	99	0.016	4.651	0.297	0.288	−0.188	13.789	5.046	−0.198	−0.244	9.320	4.665	10.078	0.810	0.822	0.000	−0.040	0.085	−0.183	−0.221
Papain	100	0.014	4.245	0.274	0.320	−0.072	13.925	6.468	0.254	0.353	8.880	2.887	2.830	0.645	0.592	0.000	0.000	0.108	0.083	0.053
Ribonuclease A	100	0.016	2.419	−0.181	−0.234	−0.290	13.933	3.623	−1.751	−1.638	8.640	4.986	3.226	0.254	0.255	0.000	−0.033	0.065	−0.185	−0.144
Trypsinogen	90	0.004	3.587	−0.082	−0.090	−0.227	14.017	7.508	−1.318	−1.444	8.230	2.434	3.139	0.518	0.609	0.000	0.000	0.100	0.069	0.056
PCC with observed values	−0.25	−0.75	0.96	1.00		0.59	−0.59	−0.12	0.99		0.95	−0.32	0.04	0.99		−0.74	0.68	0.64	0.99	

values, which is why the points are located near the diagonal line. This means that the structural properties of proteins can be used to accurately determine the protein partition behavior in a variety of the two-phase system that we explored in our study.

3.2. Relation between structural characteristics of proteins and their partition coefficients in aqueous two-phase systems

Analysis of the data in Table 3 shows that different structural features can be used to describe the partition coefficients of proteins in various ATPSs. It may be noticed, however, that partition coefficients of the proteins in three out of six ATPSs formed by PEG-600 and Na_2SO_4 with various salt additives (systems #25, 26 and 27) are described by the same proteins structural features, isoelectric point (x_3), coil content (x_{11}), and fraction of polar solvent accessible residues (s_{18}) or occupancy (x_{16}). The data were re-examined to find out if proteins' partition coefficients might be described in terms of the above three structural features (x_3, x_{11}, and x_{18}) in all PEG-600-Na_2SO_4 ATPSs utilized in this study. The coefficients and statistical parameters of multiple linear regression equations obtained are presented in Table 5. The data shown in Table 5 indicate that the three structural features of the proteins adequately describe protein partition coefficients in all six PEG-600-Na_2SO_4 ATPSs studied here, though the contributions of the structural features represented by the corresponding coefficients vary with salt composition of ATPS.

Furthermore, Table 4 shows that the model for system 3 relies on the features that quantify flexibility of protein structure (x_{16}), protein surface (x_{17}), and secondary structure (x_{28}). Models for the other systems utilize other types of protein characteristics. Specifically, regression for system 19 is based on characteristics of protein surface (x_{25} and x_{55}) and tertiary structure (x_{45}), system 23 based on features that quantify physiochemical properties (x_3), tertiary structure (x_4) and secondary structure (x_{27}), while model for system 29 uses information about surface (x_{22}, x_{23} and x_{53}). One should keep in mind that Table 4 shows only the models that attain the highest cvPCC while many other high quality models (having cvPCC > 0.9 and p-value < 0.05) were also produced for these systems.

We use the collection of all significant regressions (cvPCC > 0.9 and p-value < 0.05) to determine a set of structural features that are associated with the measured coefficients of a given two-phase system. The degree of the association is quantified with a score based on a fraction of significant regressions that include that feature, weighted by the number of inputs for these regressions. The scores across the features used by all significant regressions for a given system sum to 100. Next, we grouped the features into the five corresponding types: physiochemical properties (PP), tertiary structure (TS), secondary structures (SS), surface (SF), and intrinsic disorder and flexibility (DF). Fig. 2 shows the highest score for each type of features across the considered 29 solute-solvent systems. Interestingly, we note that certain types of features are consistently associated with the systems from the same cluster. For instance, systems in the first cluster (systems 1 to 17) are determined by the secondary structure, surface and to a smaller extent tertiary structure. In contrast, the systems in the third cluster (systems 23 to 28) are particularly suitable to study physiochemical properties, in particular pI, surface and secondary structure. Overall, we found that partition behavior of proteins is determined by at least one of the considered structural characteristics of proteins. Peculiarities of the protein surface are one of the strongest and most consistent (across different systems) determinants. Different types of two-phase systems offer complementary way to characterize different sets of structural properties of proteins.

It should be mentioned that the protein molecular weight (feature x_2) has been found as a factor with a very small contribution just for a single ATPS in agreement with the suggested previous observation [24,27,28] that molecular weight of a solute does not play a role in solute partitioning in ATPS. The conclusion that the molecular weight of

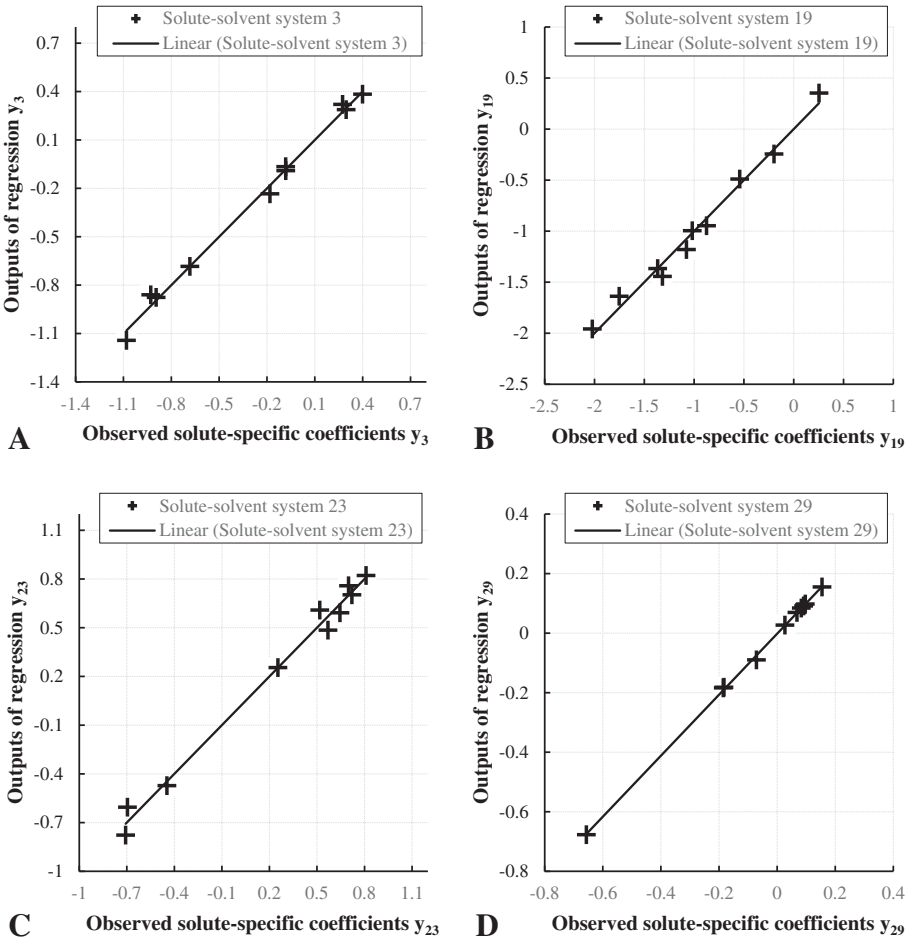

Fig. 1. Relation between values derived from the regression modeling (*y*-axis) and the corresponding observed values of the solute-specific coefficients (*x*-axis) measured for the considered 10 proteins. Lines denote linear fit. Panels A, B, C and D correspond to results for the representative four solute-solvent systems 3, 19, 23, and 29, respectively.

the solute does not influence its partition behavior was drawn previously on the basis of the facts that (a) partition coefficients of small organic compounds and proteins have been found to be equal, and (b) there was no trend between partition coefficients and molecular weight of proteins or small organic compounds.

For the two other relatively large subsets of PEG-8000-Na_2SO_4 (five systems) and dextran-PEG (17 systems) ATPSs we could not find certain structural features of proteins that could be used to describe protein partition behavior in terms of the same features. There are two likely reasons. First, it has been established [24,26–28] that the logarithm of partition coefficient of a protein in an ATPS may be described as a linear combination of different types of protein-solvent interactions in the two phases (dipole-dipole, hydrogen bonding and electrostatic interactions), and the contributions of these types of interactions vary with rather small changes in the ionic composition of ATPS. We would like to emphasize here that the previously conducted analysis of the hydrophobic properties of the phases using free energy of transfer of a methylene group revealed that there was no correlation between partition behavior of proteins and the relative hydrophobicity of the phases in all analyzed ATPSs [24–29].

Second, the majority of structural features of the proteins explored (except three physicochemical features x_1, x_2, and x_3) were computed based on the crystal structures of the proteins. The crystal structure of a protein is clearly incapable to represent changes in the protein-solvent interactions occurring in response to different ionic composition of the media. The results obtained in this study appear to confirm that analysis of protein partitioning in ATPSs of various compositions provide complimentary information about protein-solvent interactions [11]. Hence it seems unavoidable to conclude that structural features of proteins derived from the protein crystal structure cannot be used to predict protein partition behavior across different ATPSs. We likely need to gain better insight into relationships between protein-solvent

Table 5
Coefficients[a] and statistical parameters[b] for equation $\log K = k_o + k_1x_3 + k_2x_{11} + k_3x_{18}$.

ATPS #	k_o	k_1	k_2	k_3	N	r^2	SD	F	p
23	$-1.40_{0.22}$	$0.323_{0.018}$	$-0.017_{0.005}$	$-63.7_{12.3}$	10	0.984	0.095	124	<0.001
24	$-0.75_{0.21}$	$0.264_{0.018}$	$-0.024_{0.005}$	$-59.9_{11.9}$	10	0.979	0.092	91.2	<0.001
25	n/a*	$0.353_{0.042}$	$-0.040_{0.013}$	$-84.4_{29.1}$	10	0.942	0.213	32.4	<0.001
26	$-1.83_{0.68}$	$0.545_{0.057}$	$-0.043_{0.016}$	$-107.1_{38.6}$	10	0.947	0.298	35	<0.001
27	n/a*	$0.330_{0.036}$	$-0.034_{0.011}$	$-86.9_{24.9}$	10	0.952	0.182	39.9	<0.001
28	n/a*	$0.339_{0.027}$	$-0.034_{0.007}$	$-151.5_{18.1}$	10	0.977	0.140	86	<0.001

[a] Standard errors shown as subscripts; [b] N — number of proteins; r2 — correlation coefficient; SD — standard deviation; F — ratio of variance; p — statistical significance; *n/a — low statistical significance ($p > 0.05$).

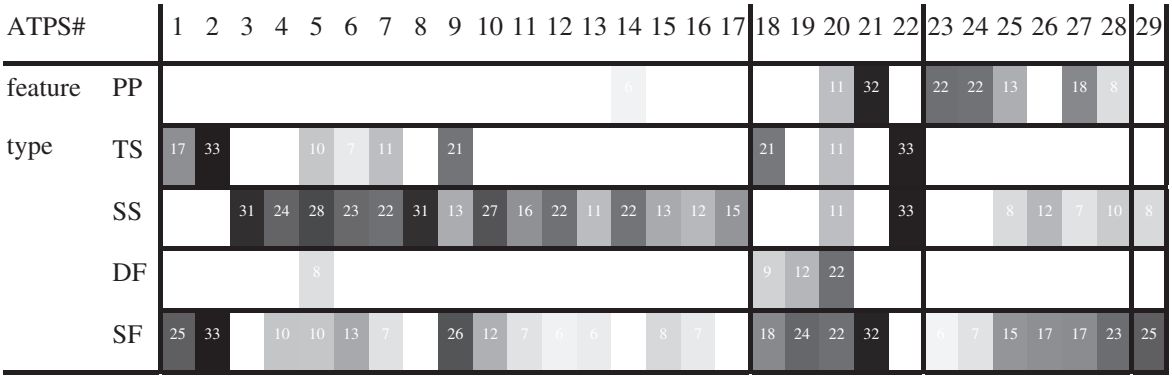

Fig. 2. Relation between the peculiarities of the protein structures and the observed solute-specific coefficients for the considered 29 systems. Shading indicates strength of the relation where white, gray and black correspond to lack, modest and strong relation. The strength values are quantified by the frequency of inclusion of the corresponding protein features among all regressions characterized by significant correlations (cvPCC > 0.9, p-value < 0.05) with the observed coefficients for a given system. The systems are divided into four groups: systems 1 to 17, 18 to 22, 23 to 28, and 29. Features are divided into five categories: physiochemical properties (PP), tertiary structure (TS), secondary structures (SS), surface (SF), and intrinsic disorder and flexibility (DF); see Table 2.

interactions and protein structure peculiarities gleaned from more suitable structural models to be able to predict these protein partition behaviors.

4. Conclusions

Analysis of partition coefficients of 10 different proteins in 29 various ATPS in terms of 57 different structural features derived from the proteins crystal structures shows that three different features describe protein partition behavior in each ATPS quite satisfactory. These features, however, vary for different ATPS. It appears that analysis of the protein partition behavior in different ATPS may provide different complementary information. There is no a unique set of protein structural features derived from crystal structures of query proteins that can describe the partition behavior of all proteins in all ATPSs. It is suggested that information gained from the crystal structure of a protein cannot provide insight in regard of the protein responsiveness to different aqueous environment.

Transparency document

The Transparency document associated with this article can be found, in online version.

References

[1] P.A. Albertsson, Partition of Cell Particles and Macromolecules, 3rd ed. Wiley, New York, 1986.
[2] Aqueous two-phase systems, in: R. Hatti-Kaul (Ed.), Methods and Protocols, Humana Press, Totowa, 2000.
[3] B.Y. Zaslavsky, Aqueous two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications, CRC Press, 1994.
[4] R.R. Soares, A.M. Azevedo, J.M. Van Alstine, M.R. Aires-Barros, Partitioning in aqueous two-phase systems: analysis of strengths, weaknesses, opportunities and threats, Biotechnol. J. (2015).
[5] B.Y. Zaslavsky, V.N. Uversky, A. Chait, Analytical applications of partitioning in aqueous two-phase systems: exploring protein structural changes and protein-partner interactions in vitro and in vivo by solvent interaction analysis method, Biochim. Biophys. Acta 1864 (2016) 622–644.
[6] F. Rahimpour, F. Feyzi, S. Maghsoudi, R. Hatti-Kaul, Purification of plasmid DNA with polymer-salt aqueous two-phase system: optimization using response surface methodology, Biotechnol. Bioeng. 95 (2006) 627–637.
[7] J.T. de Faria, F.C. Sampaio, A. Converti, F.M. Passos, V.P. Minim, L.A. Minim, Use of response surface methodology to evaluate the extraction of *Debaryomyces hansenii* xylose reductase by aqueous two-phase system, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 877 (2009) 3031–3037.
[8] T.P. Nascimento, A.E. Sales, C.S. Porto, R.M. Brandao, G.M. de Campos-Takaki, J.A. Teixeira, T.S. Porto, A.L. Porto, A. Converti, Purification of a fibrinolytic protease from Mucor subtilissimus UCP 1262 by aqueous two-phase systems (PEG/sulfate), J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 1025 (2016) 16–24.
[9] L.P. Malpiedi, B.B. Nerli, M.E.S. Taqueda, D.S.P. Abdalla, A. Pessoa, Optimized extraction of a single-chain variable fragment of antibody by using aqueous micellar two-phase systems, Protein Expr. Purif. 111 (2015) 53–60.
[10] R.L. Perez, D.B. Loureiro, B.B. Nerli, G. Tubio, Optimization of pancreatic trypsin extraction in PEG/citrate aqueous two-phase systems, Protein Expr. Purif. 106 (2015) 66–71.
[11] A. Zaslavsky, P. Madeira, L. Breydo, V.N. Uversky, A. Chait, B. Zaslavsky, High throughput characterization of structural differences between closely related proteins in solution, Biochim. Biophys. Acta 1834 (2013) 583–592.
[12] M.A. Eiteman, J.L. Gainer, A model for the prediction of partition coefficients in aqueous two-phase systems, Bioseparation 2 (1991) 31–41.
[13] M.A. Eiteman, C. Hassinen, A. Veide, A mathematical model to predict the partitioning of peptides and peptide-modified proteins in aqueous two-phase systems, Biotechnol. Prog. 10 (1994) 513–519.
[14] T.T. Franco, A.T. Andrews, J.A. Asenjo, Use of chemically modified proteins to study the effect of a single protein property on partitioning in aqueous two-phase systems: Effect of surface charge, Biotechnol. Bioeng. 49 (1996) 309–315.
[15] T.T. Franco, A.T. Andrews, J.A. Asenjo, Use of chemically modified proteins to study the effect of a single protein property on partitioning in aqueous two-phase systems: Effect of surface hydrophobicity, Biotechnol. Bioeng. 49 (1996) 300–308.
[16] W.Y. Fan, U. Bakir, C.E. Glatz, Contribution of protein charge to partitioning in aqueous two-phase systems, Biotechnol. Bioeng. 59 (1998) 461–470.
[17] A. Olivera-Nappa, G. Lagomarsino, B.A. Andrews, J.A. Asenjo, Effect of electrostatic energy on partitioning of proteins in aqueous two-phase systems, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 807 (2004) 81–86.
[18] B.A. Andrews, A.S. Schmidt, J.A. Asenjo, Correlation for the partition behavior of proteins in aqueous two-phase systems: effect of surface hydrophobicity and charge, Biotechnol. Bioeng. 90 (2005) 380–390.
[19] J.C. Salgado, B.A. Andrews, M.F. Ortuzar, J.A. Asenjo, Prediction of the partitioning behaviour of proteins in aqueous two-phase systems using only their amino acid composition, J. Chromatogr. A 1178 (2008) 134–144.
[20] J.C. Salgado, I. Rapaport, J.A. Asenjo, Is it possible to predict the average surface hydrophobicity of a protein using only its amino acid composition? J. Chromatogr. A 1075 (2005) 133–143.
[21] W.-Y. Chen, C.-G. Shu, J.Y. Chen, J.-F. Lee, The effects of amino acid sequence on the partition of peptides in aqueous two-phase system, J. Chem. Eng. Jpn. 27 (1994) 688–690.
[22] K. Berggren, A. Wolf, J.A. Asenjo, B.A. Andrews, F. Tjerneld, The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1596 (2002) 253–268.
[23] K. Berggren, M.R. Egmond, F. Tjerneld, Substitutions of surface amino acid residues of cutinase probed by aqueous two-phase partitioning, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1481 (2000) 317–327.
[24] N.R. da Silva, L.A. Ferreira, P.P. Madeira, J.A. Teixeira, V.N. Uversky, B.Y. Zaslavsky, Analysis of partitioning of organic compounds and proteins in aqueous polyethylene glycol-sodium sulfate aqueous two-phase systems in terms of solute-solvent interactions, J. Chromatogr. A 1415 (2015) 1–10.
[25] L.A. Ferreira, P.P. Madeira, V.N. Uversky, B.Y. Zaslavsky, Analyzing the effects of protecting osmolytes on solute-water interactions by solvatochromic comparison method: I. Small organic compounds, RSC Adv. 5 (2015) 59812–59822.
[26] P.P. Madeira, C.A. Reis, A.E. Rodrigues, L.M. Mikheeva, A. Chait, B.Y. Zaslavsky, Solvent properties governing protein partitioning in polymer/polymer aqueous two-phase systems, J. Chromatogr. A 1218 (2011) 1379–1384.
[27] L.A. Ferreira, N.R. da Silva, S.R. Wlodarczyk, J.A. Loureiro, P.P. Madeira, J.A. Teixeira, V.N. Uversky, B.Y. Zaslaysky, Interrelationship between partition behavior of organic compounds and proteins in aqueous dextran-polyethylene glycol and polyethylene glycol-sodium sulfate two-phase systems, J. Chromatogr. A 1443 (2016) 21–25.
[28] L.A. Ferreira, X. Fan, P.P. Madeira, L. Kurgan, V.N. Uversky, B.Y. Zaslavsky, Analyzing the effects of protecting osmolytes on solute-water interactions by solvatochromic comparison method: II. Globular proteins, RSC Adv. 5 (2015) 59780–59791.

[29] L. Ferreira, X. Fan, L.M. Mikheeva, P.P. Madeira, L. Kurgan, V.N. Uversky, B.Y. Zaslavsky, Structural features important for differences in protein partitioning in aqueous dextran-polyethylene glycol two-phase systems of different ionic compositions, Biochim. Biophys. Acta, Proteins Proteomics 1844 (2014) 694–704.
[30] L. Mikheeva, P. Madeira, B. Zaslavsky, Protein characterization by partitioning in aqueous two-phase systems, in: V.N. Uversky, A.K. Dunker (Eds.), Intrinsically Disordered Proteins, Experimental Techniques, vol. I, Humana Press, Totowa, NJ 2012, pp. 351–361.
[31] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[32] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, D.F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, Methods Mol. Biol. 112 (1999) 531–552.
[33] M.J. Mizianty, W. Stach, K. Chen, K.D. Kedarisetti, F.M. Disfani, L. Kurgan, Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources, Bioinformatics 26 (2010) i489–i496.
[34] K. Rother, P.W. Hildebrand, A. Goede, B. Gruening, R. Preissner, Voronoia: analyzing packing in protein structures, Nucleic Acids Res. 37 (2009) D393–D395.
[35] J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang, CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, Nucleic Acids Res. 34 (2006) W116–W118.
[36] K.W. Plaxco, K.T. Simons, D. Baker, Contact order, transition state placement and the refolding rates of single domain proteins, J. Mol. Biol. 277 (1998) 985–994.
[37] W. Kabsch, C. Sander, Dictionary of protein secondary structure - pattern-recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[38] D. Juretic, D. Zucic, B. Lucic, N. Trinajstic, Preference functions for prediction of membrane-buried helices in integral membrane proteins, Comput. Chem. 22 (1998) 279–294.
[39] D. Juretic, L. B., D. Zucic, N. Trinajstic, Protein transmembrane structure: recognition and prediction by using hydrophobicity scales through preference functions. J. Theor. Comput. Chem. 5 (1998) 405–445.
[40] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, The hydrophobic moment detects periodicity in protein hydrophobicity, Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 81 (1984) 140–144.
[41] H. Cid, M. Bunster, M. Canales, F. Gazitua, Hydrophobicity and structural classes in proteins, Protein Eng. 5 (1992) 373–375.