Biochimica et Biophysica Acta 1844 (2014) 694–704

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Structural features important for differences in protein partitioning in aqueous dextran–polyethylene glycol two-phase systems of different ionic compositions

Luisa Ferreira [a], Xiao Fan [b], Larissa M. Mikheeva [a], Pedro P. Madeira [c], Lukasz Kurgan [b], Vladimir N. Uversky [d,e,*], Boris Y. Zaslavsky [a,**]

[a] Analiza, Inc., 3516 Superior Ave., Suite 4407B, Cleveland, OH 44114, USA
[b] Department of Electrical and Computer Engineering, University of Alberta, Canada
[c] Laboratory of Separation and Reaction Engineering, Department de Engenharia Química, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal
[d] Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[e] Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Regions, Russia

ARTICLE INFO

Article history:
Received 26 November 2013
Received in revised form 7 January 2014
Accepted 24 January 2014
Available online 30 January 2014

Keywords:
Protein structure
Protein partition
Aqueous two-phase system
Protein flexibility

ABSTRACT

Partitioning of 15 proteins in dextran–70-polyethylene glycol (PEG)-8000 aqueous two-phase systems (ATPSs) in the presence of 0.01 M sodium phosphate buffer, pH 7.4 was studied. The effect of salt additives (NaCl, CsCl, Na_2SO_4, $NaClO_4$ and NaSCN) at different concentrations on the protein partition behavior was examined. The salt effects on protein partitioning were analyzed by using the Collander solvent regression relationship between the protein partition coefficients in ATPSs with and without salt additives. The results obtained show that the presence and concentration of salt additives affect the protein partition behavior. Analysis of ATPSs in terms of the differences between the relative hydrophobicity and electrostatic properties of the phases does not explain the protein partition behavior. The differences between protein partitioning could not be explained by the protein size. The structural signatures for the proteins were constructed from partition coefficient values in four ATPSs with different salt additives, and the structural distances were calculated using cytochrome c as the reference structure. The structural distances for all the examined proteins (except lysozyme) were found to be interrelated. Analysis of about 50 different descriptors of the protein structures revealed that the partition behavior of proteins is determined by the peculiarities of their surfaces (e.g., the number of water-filled cavities and the averaged hydrophobicity of the surface residues) and by the intrinsic flexibility of the protein structure measured in terms of the B-factor (or temperature factor).

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Protein partitioning in aqueous two-phase systems (ATPSs) is well known as a convenient, inexpensive, and readily scaled-up protein separation technique [1–3]. It may also serve as an analytical method for protein analysis and structural characterization providing unique information about protein–water interactions and changes in protein structure [4–8]. ATPSs are formed in mixtures of two (or more) water-soluble polymers, such as dextran and polyethylene glycol (PEG), or a single polymer and particular salt in water above certain threshold concentrations or temperature. In such systems, two immiscible coexisting aqueous phases are formed. There is a clear interfacial boundary separating two distinct aqueous-based phases, each preferentially enriched in one of the polymers, with the aqueous solvent in both phases suitable for biological products [1–4]. These systems are unique in that each of the phases typically contains well over 80% water on a molal basis, and yet they are immiscible and differ in their solvent properties [4,8–14]. In ATPSs, each phase provides a distinct solvent environment for proteins or other solutes. Differences in solute–solvent interactions in the two phases often lead to unequal solute distribution, which is quantified by a partition coefficient, designated as *K*, and may be exploited for sensitive detection of changes in the solute structure. The partition coefficient *K* of a protein is defined as the ratio of the protein concentrations in the two phases.

Therefore, this Solvent Interaction Analysis (SIA) based on quantifying interactions of a protein with two aqueous media of different solvent properties constitutes an analytical tool to gain useful structural information (see below). This approach provides information about changes in the protein 3D structure and differences between 3D structures of closely related proteins that is very difficult to gain with conventional biophysical techniques [15]. In fact, differences between

* Correspondence to: V.N. Uversky, Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA. Tel.: +1 813 974 5816; fax: +1 813 974 7357.

** Corresponding author. Tel.: +1 216 432 9050x111; fax: +1 216 432 9050.
E-mail addresses: vuversky@health.usf.edu (V.N. Uversky), bz@analiza.com (B.Y. Zaslavsky).

1570-9639/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2014.01.016

3D structures of closely related proteins may be characterized quantitatively by analyzing partitioning behavior of these proteins in four or more ATPSs of same polymer but different ionic compositions [5]. It was also reported [16] that the specific protein–ion interactions may be detected using the so-called Collander linear solvent regression relationship observed between partition coefficients of proteins in ATPS with a given salt additive and those in ATPS without the same salt additive. As described previously [5], the potential issue related to the fact that different changes to the structure may result in the same change to *K* can be addressed by combining multiple *K*-values for the same protein using multiple ATPSs, followed by constructing a vector of information comprised of such numerical values. This vector is a numerical signature of the protein 3D structure, and different vectors can be compared using conventional mathematical tools [5].

Solvent properties of the aqueous media in the coexisting phases of ATPS formed by two nonionic polymers depend primarily on polymer and ionic composition of the phases [9,12,17]. For determination of the protein *K*-value as a sensitive descriptor of its 3D structure [18], ATPS formed by any pair of polymers may be used with the following two practical requirements: (i) lack of protein aggregation and/or precipitation at the liquid–liquid interface; and (ii) the protein *K*-value being typically in a range of 0.1–10 for robust analytical evaluation [5]. Therefore, in order to design the appropriate conditions for a given series of proteins it is necessary to perform two rounds of screening. The purpose of the first preliminary screening is to select ATPSs providing for a given protein partition coefficient within the robust analytical range (typically 0.1–10). Once the ATPSs meeting the above criterion are selected, the second final screening stage is performed. At this stage two or three proteins representing the series of proteins under analysis are examined. The purpose of this screening stage is to explore conditions providing significant differences between the partition coefficient values for the proteins tested. This screening typically uses a limited number of pre-selected ATPSs with different salt compositions. Once the suitable ATPSs are selected, the SIA for all the protein samples is performed. It would be important at this stage to know what salts and what concentration range to use to achieve the aforementioned purpose of differentiating different protein structures.

In previous study [5], we analyzed structural signatures of human, bovine and porcine insulin, bovine RNases A and B, and β-lactoglobulins A and B using the dextran-based ATPSs with different salt compositions. The purpose of this study was to extend the above approach using dextran–PEG ATPS with various salt additives and explore what salts at what concentration ranges should be used for analysis and what structural features are important for different partition behavior of structurally diverse model proteins (the list of which was also significantly increased in comparison with the previous study, see Table 1) in these systems.

2. Materials and methods

2.1. Materials

Polyethylene glycols PEG-8000 (Lot 048K00241) with an average molecular weight (M_w) of 8000 and Dextran-69 (Lot 106H0841) with an average molecular weight (M_w) 69,000 by light scattering were purchased from Sigma–Aldrich (St. Louis, MO, USA).

Albumin from human serum (fatty acid and globulin free (~99%)), α-chymotrypsin from bovine pancreas, α-chymotrypsinogen A from bovine pancreas, concanavalin A from *Canavalia ensiformis* (jack beans), cytochrome c from equine heart (>95%), hemoglobin human, hemoglobin bovine, 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, subtilisin A from *Bacillus licheniformis*, trypsinogen from bovine pancreas were purchased from Sigma–Aldrich. All proteins and the abbreviations used throughout the text are listed in Table 1.

Dinitrophenylated (DNP) amino acids—DNP-glycine, DNP-alanine, DNP-norvaline, DNP-norleucine, and DNP-α-amino-*n*-octanoic acid, were purchased from Sigma–Aldrich. The sodium salts of the DNP-amino acids were prepared by titration.

o-Phthaldialdehyde (OPA) reagent solution (complete) was purchased from Sigma–Aldrich. All salts and other chemicals used were of analytical-reagent grade and used without further purification.

2.2. Dataset

Computational analysis was performed using a set of 12 diverse proteins; PDB IDs: 1AB9, 1ACB, 1B8E, 1BEB, 1BEL, 1BTY, 1BZ0, 1HRC, 1JBC, 1PPN, 2QSS, and 3UNX (see Table 1). The corresponding sequences range between 104 and 287 residues.

2.3. Methods

2.3.1. Aqueous two-phase systems

Stock solutions of PEG 8000 (50 wt.%), Dex-69 (~35 wt.%) and salts were prepared in deionized (DI) water. Stock sodium/phosphate buffer (NaPB; 0.5 M, pH 7.4) was prepared by mixing appropriate amounts of NaH_2PO_4 and Na_2HPO_4. A mixture of polymers was prepared as described elsewhere [17] 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 and salt solutions were added to give the ionic and polymer composition required for the final system (after the sample addition — see below) with total volume of

Table 1
Proteins used in this study.

Protein[a]	Abbreviation	PDB ID	Molecular weight, kDa	pI
Albumin fatty acid and globulin free	HSA		66.4	4.7
α-Chymotrypsin	CHY	1AB9	25.0	8.75
α-Chymotrypsinogen A	CHTG	1ACB	25.7	8.97
Concanavalin A	ConA	1JBC	104.0	4.5–5.5
Cytochrome c	Cyt c	1HRC	12.4	9.1
Hemoglobin bovine	BHb	2QSS	66.0	6.8
Hemoglobin human	HHb	1BZ0	64.5	6.8
β-Lactoglobulin A	bLGA	1B8E	18.3	5.3
β-Lactoglobulin B	bLGB	1BEB	18.3	5.1
Lysozyme	HEL		14.3	11.0
Papain	Pap	1PPN	23.4	8.75–9.55
Ribonuclease A	RNase A	1BEL	13.7	9.63
Ribonuclease B	RNase B		17.0	8.88
Subtilisin A	SubA	3UNX	27.0	9.4
Trypsinogen	TRY	1BTY	24.0	8.7; 9.3

[a] All proteins from Sigma–Aldrich, details see in Materials and methods.

470 ± 4 μL depending on the salt additive. All the aqueous two-phase systems used had the same polymer composition of 6.05 wt.%. PEG-8000 and 12.33 wt.% Dex-69, and different ionic compositions indicated below. The ionic compositions of the systems used are listed in Table 2.

2.3.2. 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) 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 1.25–5 mg/mL. Varied amounts (e.g. 0, 5, 10, 15, 20, and 30 μL) of protein solution and the corresponding amounts (e.g. 240, 235, 230, 225, 220 and 210 μL) of water were added to a set of the same polymer/salt/buffer mixtures. The particular volumes varied depending on the OPA assay sensitivity and concentration of the particular protein examined as well as on the particular salt additive used.

The systems were then vortexed in a Multipulse vortexer and centrifuged (Jouan, BR4i, Thermo Fisher Scientific, Waltham, MA, USA) for 30 min at 3500 ×*g* at 23 °C to accelerate phase settling. The top phase in each system was removed, the interface discarded, and aliquots of 20 to 70 μL from the top and bottom phases were withdrawn in duplicate for analysis. These aliquots were combined with 250 μL OPA solution in microplate wells. After moderate shaking for 3 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. In the experiments with cytochrome c the aliquots were diluted with water instead of OPA, shaken as above, and absorbance at 408 nm was determined.

The distribution 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 solute was determined as the slope of the concentration (fluorescence intensity or absorbance in the case of cytochrome c) in the top phase plotted as a function of the concentration (fluorescence intensity or absorbance in the case of cytochrome c) in the bottom phase averaged over the results obtained from two to four partition experiments carried out at the specified ionic composition of the system [17]. The deviation from the average *K* value was always less than 3% and in most cases lower than 1%.

2.3.3. Electrophoresis

All protein samples were characterized by SDS-PAGE electrophoresis in a microfluidic chip using Bioanalyzer 2100, Protein 200 Plus Assay (Agilent Technologies, USA) under non-reduced conditions. All proteins were observed as single bands in the electrophoregrams.

2.3.4. Protein descriptors

For each protein we collected a comprehensive set of structural descriptors that includes general descriptors derived from the analysis of protein amino acid sequences and specific descriptors derived from the analysis structures of the corresponding proteins. Complete list of 50 used structural descriptors is provided in the Results and discussion section. Next, we computed Person correlation coefficients (PCCs) between each of the 50 descriptors and the observed values derived for the given ATPS to examine whether the observed partition-based values correlate with the considered structural properties of proteins.

2.3.5. Multivariate modeling with regression

We also investigated correlation of a combination of descriptors using linear regression. We employed the minimum sum of squared errors linear regression. Given the observed data from the two-phase system $y \in R^{t \times 1}$, and a set of calculated protein descriptors $X \in R^{t \times n}$, where *t* is the number of proteins, *n* is the number of protein descriptors used in the regression model, the criterion to solve the regression model is defined as:

$$\min_{r}\left(\|Xr - y\|_2^2\right) \tag{1}$$

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

3. Results and discussion

3.1. Partition behavior of a homologous series of dinitrophenylated (DNP-) amino acids

Typical experimental data obtained for sodium salts of DNP-amino acids in dextran–PEG ATPSs with different salt additives are plotted in Fig. 1, and the linear curves observed may be described as:

$$logK^{(i)}{}_{DNP-AA} = C^{(i)} + E^{(i)}N_C \tag{2}$$

where $K_{DNP\text{-}AA}$ is the partition coefficient of a DNP-amino acid with aliphatic side-chain; superscript *(i)* denotes the particular i^{th} ATPSs used for the partition experiments; N_C is equivalent number of CH_2 groups in the aliphatic side-chain of a given DNP-amino acid; *E* is an average *logK* increment per CH_2 group; *C* represents the total contribution of the non-alkyl part of the structure of a DNP-amino acid into $logK_{DNP\text{-}AA}$ and used to characterize the difference between the electrostatic properties of the coexisting phases as described previously [3,14,16].

Table 2
Salt composition of the 12.33 wt.% Dex-69–6.05 wt.% PEG-8000-0.01 M NaPB, pH 7.4 aqueous two-phase systems and differences between the relative hydrophobicity (parameter E) and electrostatic properties (parameter C) of the coexisting phases of the systems.

#	Salt additive	C	E	−ΔG(CH_2), cal/mole CH_2	σΔG(CH_2)[a], cal/mole CH_2	σC[a]
1	–	0.075 ± 0.001	0.024 ± 0.001	33 ± 1.4		
2	0.15 M NaCl	−0.039 ± 0.001	0.026 ± 0.001	35 ± 1.4	−2 ± 2.8	−0.114 ± 0.002
3	1.05 M NaCl	−0.020 ± 0.002	0.033 ± 0.001	45 ± 1.4	−12 ± 2.8	−0.095 ± 0.003
4	0.083 M Na_2SO_4	0.166 ± 0.003	0.038 ± 0.002	51 ± 2.7	−18 ± 4.1	0.091 ± 0.004
5	0.30 M Na_2SO_4	0.204 ± 0.002	0.047 ± 0.001	64 ± 1.4	−31 ± 2.8	0.129 ± 0.003
6	0.17 M CsCl	−0.036 ± 0.001	0.025 ± 0.001	34 ± 1.4	−1 ± 2.8	−0.111 ± 0.002
7	0.80 M CsCl	−0.036 ± 0.002	0.033 ± 0.001	45 ± 1.4	−12 ± 2.8	−0.111 ± 0.003
8	0.094 M $NaClO_4$	−0.095 ± 0.002	0.028 ± 0.001	38 ± 1.4	−5 ± 2.8	−0.170 ± 0.003
9	0.43 M $NaClO_4$	−0.064 ± 0.002	0.036 ± 0.001	49 ± 1.4	−14 ± 2.8	−0.139 ± 0.003
10	0.17 M NaSCN	−0.081 ± 0.002	0.027 ± 0.001	37 ± 1.4	−4 ± 2.8	−0.156 ± 0.003
11	1.26 M NaSCN	−0.017 ± 0.001	0.042 ± 0.001	57 ± 1.4	−24 ± 2.8	−0.092 ± 0.002

[a] σΔG(CH_2) and σC represent the differences between the ΔG(CH_2) values and C values for ATPS with the indicated salt additive and those in the ATPS without salt additive.

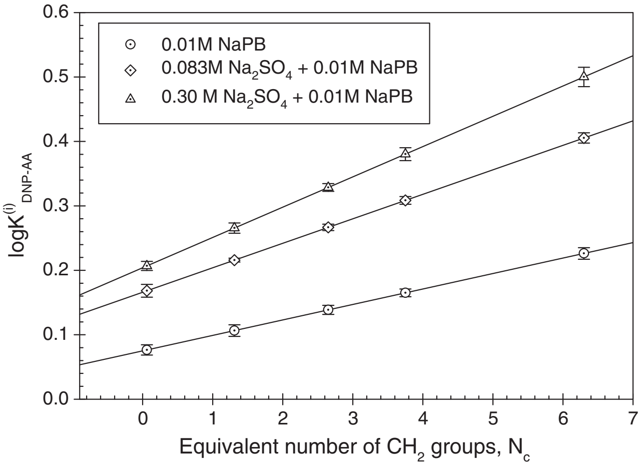

Fig. 1. Logarithm of the partition coefficient, log $K_{DNP\text{-}AA}$, value for sodium salts of DNP-amino acids with aliphatic side-chains in aqueous dextran–PEG two-phase systems as a function of equivalent length of the side-chain, N_C, expressed in terms of equivalent number of CH_2 units. in 0.01 M sodium phosphate buffer, pH 7.4; 0.083 M Na_2SO_4 in 0.01 M sodium phosphate buffer, pH 7.4; and in 0.30 M Na_2SO_4 in 0.01 M sodium phosphate buffer, pH 7.4.

The coefficients $E^{(i)}$ and $C^{(i)}$ values determined for the ATPSs examined are listed in Table 2. As the standard free energy of transfer of a solute from the bottom phase to the top phase is described as:

$$\Delta G^0 = -RTlnk \tag{3}$$

where R is the universal gas constant and T is the absolute temperature in Kelvin, it follows that

$$\Delta G^0(CH_2) = -RTE* \tag{4}$$

where $\Delta G^0(CH_2)$ is the standard free energy of transfer of a methylene group from one phase to another, E^* is expressed in natural logarithmic units. The $\Delta G^0(CH_2)$ values calculated from the experimental data with Eqs. (2)–(4) are listed in Table 2.

The difference between the relative hydrophobic character of the phases as indicated by the $\Delta G^0(CH_2)$ values in Table 2 depends on the salt additive type and concentration in the presence of 0.01 M NaPB. It should be mentioned that the salt effects in this case are similar to those observed previously [4] in the similar ATPS of different polymer composition in the presence of 0.01 M universal buffer, pH 7.4. The difference between the electrostatic properties of the phases characterized by the parameter C value (Table 2) as expected changes with the salt additive type and concentration more dramatically than the $\Delta G^0(CH_2)$ value.

3.2. Protein partitioning

Partition coefficients K for 15 different proteins examined in all ATPSs are listed in Table 3. Analysis of the K-values shows that the most of the proteins (HSA, ConA, Cyt c, BHb, HHb, bLGA, bLGB, RNase A, and RNase B) distribute predominantly into the bottom dextran-rich phase ($K < 1$) under all conditions employed. Three proteins (CHTG, Pap, and SubA) distribute predominantly into the top PEG-rich phase under all conditions used, and only three proteins (CHY, HEL, and TRY) distribute into either phase depending on the salt composition of the system. It should be noted that the PEG-8000 concentration in the top phase is ca. 12wt.% and dextran-70 concentration in the bottom phase is ca.34-25 wt.%. Therefore, the distribution of 9 proteins with molecular weight varying from 12.4 kDa for Cyt c up to 104 kDa for ConA predominantly into the bottom phase cannot be explained by the polymer excluded volume effect. The fact that partition behavior of proteins of essentially the same molecular weight, such as TRY (24 kDa) and Pap (23.4 kDa) or CHTG (25.7 kDa) and CHY (25 kDa), is so different also contradicts the viewpoint [19] that the protein size is of primary importance for protein partition behavior in ATPS and agrees with the conclusion made in our recent work [16] were it was suggested that the protein size is of secondary if any importance for the protein partition behavior.

Analysis of the data in Table 3 shows that increasing NaCl and CsCl concentrations results in increased partition coefficients for all proteins except bLGA in NaCl, and except bLGA and RNase B in CsCl, while changing CsCl concentration does not affect partition coefficient for Cyt c. Increasing Na_2SO_4 concentration results in reducing partition coefficients for CHY, ConA, Cyt c, BHb, HHb, bLGA, bLGB, and RNase B (for some proteins just slightly) and increases partition coefficients for CHTG, HEL, Pap, RNase A, SubA, and TRY. Increasing $NaClO_4$ concentration from 0.093 M to 0.43 M increases partition coefficients for all proteins except bLGA, bLGB, and TRY. Increasing NaSCN concentration from 0.17 M up to 1.26 M also increases partition coefficients for all but three (Cyt c, RNase A, TRY) proteins. Overall the data obtained imply that if the proteins being examined partition in a given ATPS with low partition coefficients, their coefficients might be increased by increasing salt additive concentration.

It was shown previously [4,20,21] that the partition coefficients for different proteins in ATPSs of different polymer but same ionic compositions are typically interrelated in accordance with the so-called Collander solvent regression equation [22–26]:

$$logK_{ji} = a_{io}logK_{jo} + b_{io} \tag{5}$$

where K_{ji} and K_{jo} are distribution coefficients for any given j^{th} solute in the i^{th} and o^{th} two-phase systems; a_{io} and b_{io} are constants, the values of which depend upon the particular composition of the i^{th} and o^{th} two-phase systems under comparison.

Table 3
Partition coefficients K-values for proteins[a] in the aqueous Dex-PEG two-phase systems[b] (error in K-value below 5%).

#	HSA	CHY	CHTG	ConA	Cyt c	BHb	HHb	bLGA	bLGB	HEL	Pap	RNase A	RNase B	SubA	TRY
1	0.076	0.580	1.164	0.125	0.067	0.158	0.245	0.375	0.386	0.225	1.020	0.225	0.555	1.137	0.422
2	*	0.859	2.420	0.133	0.212	0.136	0.207	0.072	0.138	1.939	2.091	0.583	0.589	1.469	0.812
3	*	1.346	6.71	0.151	0.230	0.267	0.355	0.054	0.256	10.4	3.034	0.757	0.644	2.023	1.553
4	0.025	0.649	1.643	0.148	0.054	0.146	0.182	0.094	0.162	0.451	1.669	0.323	0.386	1.318	0.491
5	*	0.594	2.277	0.141	0.040	0.123	0.166	0.050	0.123	0.688	2.853	0.401	0.317	1.575	0.579
6	*	0.785	2.432	0.133	0.199	0.117	0.179	0.064	0.136	2.392	2.715	0.553	0.626	1.225	0.702
7	*	1.118	5.40	0.141	0.197	0.176	0.262	0.053	0.228	7.49	3.845	0.655	0.536	1.659	0.884
8	*	0.930	3.653	0.143	0.347	0.144	0.179	0.056	0.182	14.6	3.859	0.748	0.544	1.353	1.028
9	*	1.069	5.24	0.151	0.370	0.192	0.264	0.052	0.121	53.3	4.228	0.771	0.579	1.627	1.003
10	*	0.995	3.87	0.145	0.360	0.134	0.242	0.054	0.118	10.0	3.438	0.799	0.579	1.483	0.987
11	*	1.033	4.17	0.171	0.249	0.311	0.518	0.064	0.208	23.0	4.402	0.726	0.658	1.559	0.775

* — Protein concentrates in the bottom phase (no protein was determined in the top phase).

[a] Proteins abbreviations see in Table 1.

[b] Composition of ATPSs employed see in Table 2.

The aforementioned condition of the same ionic composition in the ATPSs under comparison is important to ensure that each solute in the ATPSs employed may be viewed as being essentially the same chemical entity with partition coefficient varying only as the result of different solvent features of the phases in the two systems. The applicability of the Collander solvent regression equation [22–26] to PEG-salt ATPSs differing with regard to presence of a salt additive was reported previously [16]. Proteins demonstrating partition behavior not fitting Eq. (5) were suggested to be considered as those participating in specific protein-ion interactions [16].

The Collander equation [22–24] describes a linear correlation between distribution coefficients of solutes of the similar chemical nature in different organic solvent-water biphasic systems. The coefficients of the relationship (slope and intercept) depend on the particular systems under comparison as well as on the type of the solutes being examined. It was suggested that these coefficients represent the distinctive features of the interactions of the solute functional moieties with the solvents being compared [4]. It was established earlier [21] that the distribution coefficients for different randomly selected proteins in the Dextran–PEG and Dextran-Ucon ATPSs are correlated according to the Collander equation or solvent regression equation. The applicability of the solvent regression equation to ATPSs formed by different pairs of nonionic polymers with the same ionic composition was demonstrated in [20]. It has also been shown that the Collander relationship may exist for comparison of partition coefficients for proteins in PEG-Na_2SO_4 ATPSs with different salt additives [16]. It was also suggested that proteins fitting the linear relationship do not display specific interactions with the salt additives present in the ATPS being compared, while these specific interactions were the reason for some proteins not fitting the relationship [16]. In other words, proteins fitting the linear solvent relationship are maintained as same solutes in ATPSs with different ionic compositions, indicating that changes in the partition coefficients of these proteins in the ATPSs compared result from changes in the solvent features of the systems. On the other hand, proteins not fitting the Collander relationship display different partition behavior in the two ATPSs due to changes in their solute properties attributed to the specific protein-ion interactions.

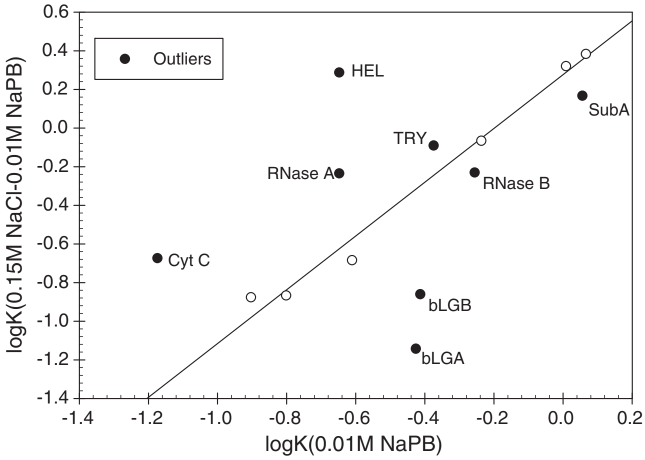

Fig. 2. Logarithms of partition coefficients for proteins in dextran–PEG-0.15 M NaCl–0.01 M NaPB versus those for the same proteins in dextran–PEG-0.01 M NaPB ATPSs. NaPB — sodium phosphate buffer, pH 4.4. Proteins not fitting the relationship are indicated as outliers.

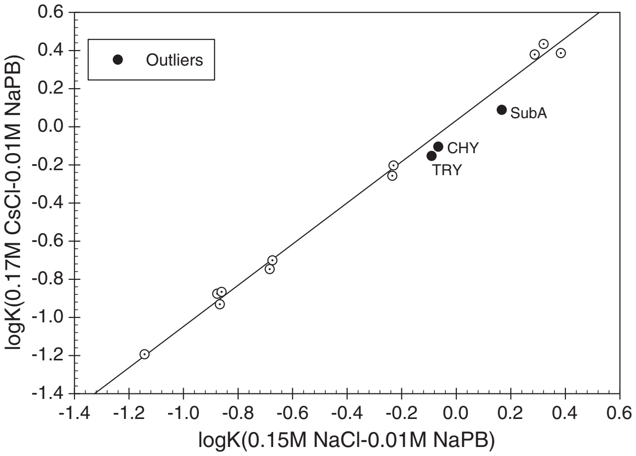

Fig. 3. Logarithms of partition coefficients for proteins in dextran–PEG-0.17 M CsCl–0.01 M NaPB versus those for the same proteins in dextran–PEG-0.15 M NaCl–0.01 M NaPB ATPSs. NaPB — sodium phosphate buffer, pH 4.4. Proteins not fitting the relationship are indicated as outliers.

Analysis of the data in Table 3 shows that the protein partition coefficients are correlated under certain conditions according to Eq. (5). As an example, the linear interrelationships between the logarithms of partition coefficients for proteins in dextran–PEG-0.01 M NaPB ATPS and those for the same proteins in dextran–PEG-0.01 M NaPB-0.15 M NaCl and between similar ATPS containing 0.15 M NaCl and 0.17 M CsCl ATPS illustrated in Figs. 2 and 3 are described as:

$$logK^{0.15M\ NaCl\text{-}0.01M\ NaPB}=0.28_{\pm 0.05}+1.39_{\pm 0.09}*logK^{0.01M\ NaPB} \quad (6)$$

$$N=6\ (N_{total}=14);R^2=0.9852;SD=0.079;F=266$$

where $K^{0.15M\ NaCl-0.01M\ NaPB}$ and $K^{0.01M\ NaPB}$ are partition coefficients for the same protein in the dextran–PEG-0.15 M NaCl in 0.01 M NaPB ATPS and in dextran–PEG-0.01 M NaPB ATPS, respectively; N is the number of proteins fitting the relationship; N_{total} is the total number of proteins examined in both ATPS; R^2 is the correlation coefficient; SD is the standard deviation; and F is the ratio of variance. Eight proteins — Cyt c, bLGA, bLGB, HEL, RNase A, RNase B, SubA, and TRY do not fit the relationship described by Eq. (6) and are denoted in Fig. 2 as outliers; and

$$logK^{0.17M\ CsCl\text{-}0.01M\ NaPB}=0.03_{\pm 0.02}+1.08_{\pm 0.02}*logK^{0.15M\ NaCl\text{-}0.01M\ NaPB} \quad (7)$$

$$N=11\ (N_{total}=14);R^2=0.9959;SD=0.040;F=2200$$

where $K^{0.17M\ CsCl-0.01M\ NaPB}$ denotes partition coefficients for proteins in the dextran–PEG-0.17 M CsCl in 0.01 M NaPB ATPS; all the other parameters as defined above. Three proteins — CHY, SubA, and TRY do not fit the relationship described by Eq. (7) and are denoted in Fig. 3 as outliers. It follows from the two above relationships that while the presence of NaCl affects partition behavior for many examined proteins, replacement of NaCl for CsCl affects only 3 out of 14 proteins studied.

Proteins fitting the linear relationships (Eq. (5)) and showing lack of specific ion-protein interactions for the salts additives examined are listed in Table 4. It is essentially the same proteins: BHb, HHb, CHTG, and CHY. In the case of 0.094 M $NaClO_4$ salt additive, there is practically no reliable linear relationship. Relatively large number of proteins not fitting the Collander relationship in any two ATPS compared here demonstrates high sensitivity of the partition behavior to specific interactions.

Table 4
Coefficients a_{io} and b_{io} in solvent regression equation $\log K^{\text{Salt-0.01M NaPB}} = a_{io} + b_{io}{*}\log K^{\text{0.01M NaPB}}$ (Eq. (5)) with ATPS containing salt additive indicated (N — number of proteins fitting the relationship out of total 14 proteins examined; R^2 — correlation coefficient; SD — standard deviation; F — the ratio of variance.

Salt	a_{io}	b_{io}	N	R^2	SD	F	Proteins fitting the linear relationship
0.15 M NaCl	0.29 ± 0.03	1.50 ± 0.06	5	0.9956	0.044	680	BHb,CHTG, CHY, HHb, Pap
0.17 M CsCl	0.25 ± 0.04	1.54 ± 0.09	5	0.9904	0.060	311	BHb, CHTG, CHY, HHb, RNase B
0.083 M Na_2SO_4	0.14 ± 0.05	1.17 ± 0.06	5	0.9921	0.058	378	BHb, ConA, Cyt c, CHTG, TRY
0.17 M NaSCN	0.47 ± 0.03	1.73 ± 0.07	5	0.9946	0.056	558	BHb,CHTG, CHY, HHb, Pap

The slopes b_{io} of the Collander solvent regression relationships listed in Table 4 are correlated with the differences between the relative hydrophobicities and electrostatic properties of the coexisting phases in the corresponding ATPS as:

$$b_{io}=0.4_{\pm 0.02}–4.0_{\pm 0.3}C_i+37_{\pm 5.7}{*}E_i \tag{8}$$

$$N = 5; R^2 = 0.9878; SD = 0.046; F = 80.7$$

where b_{io} is the slope of the solvent regression relationship; C_i and E_i are the characteristics of the differences between the electrostatic and hydrophobic properties of the coexisting phases in the i^{th} ATPS; all the other parameters are as defined above. The existence of this relationship confirms the hypothesis that the proteins fitting the solvent regression relationship described by Eq. (5) under the conditions considered may be viewed as proteins lacking specific protein-ion (salt additive) interactions, and differences in their partition behavior in the systems under comparison are due to the differences in the properties of the ATPSs utilized.

Analysis of the data in Table 3 shows that the Collander solvent regression relationship between the logarithms of partition coefficients for proteins in dextran–PEG-0.01 M NaPB ATPS exists for proteins partitioning in ATPS with 0.01 M NaPB and different concentrations of the same salt as shown in Table 5. The data in Table 5 indicate that under conditions examined, changes in the NaCl concentration affects the partition behavior of BHb, HHb, bLGB, RNase B, and TRY, while changes in the CsCl concentrations affects partition of ConA, Cyt c, RNase A, and SubA. The changes in the concentration of Na_2SO_4 affect CHY, Pap, RNase A, and RNase B. The NaSCN concentration changes affect the behavior of Cyt c, BHb, HHb, bLGA, bLGB, Pap, and TRY, and changes in the $NaClO_4$ concentration affect partition behavior of BHb, HHb, bLGA, and bLGB. Lysozyme (HEL) partitioning is affected by changes in concentrations of all the salt additives examined.

3.3. Structural signatures and distances

It was shown previously [5] that the protein 3D structure in solution may be represented by the protein partition coefficients in four ATPSs of the same polymer but different ionic compositions. In order to estimate the differences between the structures of different proteins, we need to choose a reference sample. Based on the fact that the partition coefficients K determined for Cyt c are of lowest values under essentially all the conditions explored, we selected the Cyt c as a reference and normalized the partition coefficients for all proteins against the partition coefficient for Cyt c in each ATPS chosen to characterize the proteins structures (see below). Then, the normalized Euclidian distance between the normalized structural signatures in the 4-dimensional space represented by K-values in ATPSs with four different ionic compositions (for example, in ATPSs # 1, 2, 4, 8; see Table 2) for each protein and Cyt c was evaluated. This distance was calculated as:

$$d_{i,o} = \left(\sum_j \left(\frac{K_i - K_o}{K_o}\right)^2\right)^{0.5} \tag{9}$$

where $d_{i,o}$ is the distance between the structural signature of protein sample i from that of the Cyt c used as a reference, K_{ij} and K_{oj} are the partition coefficients for the sample i and the reference sample o (Cyt c) in the system j, correspondingly. The structural distances for all the proteins examined calculated using Eq. (8) and K-values measured in ATPSs # 1, 2, 4, and 8 (Table 2) are listed in Table 6. These distances characterize the differences between the structures of the proteins examined in this work. (See Table 7.)

In order to test the reliability of the structural distance values obtained using the above ATPSs we estimated the distances using K-values for the same proteins in ATPSs #1, 2, 4, and 10 (see Table 2), i.e. using K-values for the proteins in the presence of 0.17 M NaSCN (ATPS # 10) instead of those in the presence of 0.094 M $NaClO_4$ (ATPS # 8), all salt additives in 0.01 M NaPB. The resulting structural distance values are listed in Table 6. The structural distances calculated for the proteins as described above are plotted against each other in Fig. 4. The linear relationship observed may be described as:

$$D\text{-}2=0.05_{\pm 0.03}+0.984_{\pm 0.002}{*}D\text{-}1 \tag{10}$$

$$N = 13; R^2 = 0.9999; SD = 0.08; F = 285357$$

where D-1 is the structural distance calculated for protein with Cyt c as the reference based on K-values determined in ATPSs # 1, 2, 4, 8; D-2 is the structural distance calculated for protein with Cyt c as the reference based on K-values determined in ATPSs # 1, 2, 4, 10; all the parameters are as defined above, and HEL is the only protein not fitting the relationship.

In order to examine what effect the increased salt concentrations may have on the distances between their structures we estimated the distances using K-values for the same proteins in ATPSs #1, 3, 5, and 11 (see Table 2), and in ATPSs #1, 3, 5, and 9 i.e. using K-values for the

Table 5
Coefficients a_{io} and b_{io} in solvent regression equation $\log K^{\text{Salt-2–0.01M NaPB}} = a_{io} + b_{io}{*}\log K^{\text{Salt-1–0.01M NaPB}}$ (Eq. (5)) with both ATPS containing salt additive at the concentrations indicated indicated (N — number of proteins fitting the relationship out of total 14 proteins examined; R^2 — correlation coefficient; SD — standard deviation; F — the ratio of variance.

Eq.	Salt	Concentration		a_{io}	b_{io}	N	R^2	SD	F	Proteins fitting the linear relationship
		Salt-1	Salt-2							
7a	NaCl	0.15 M	1.05 M	0.14 ± 0.02	1.11 ± 0.04	6	0.9960	0.037	992	CHY, ConA, Cyt c, Pap, RNase A, SubA
7b	CsCl	0.17 M	0.80 M	0.14 ±0.01	0.95 ± 0.03	7	0.9965	0.033	1426	CHY, BHb, HHb, bLGB, Pap, SubA, TRY
7c	Na_2SO_4	0.083 M	0.30 M	0.09 ± 0.02	1.18 ± 0.03	8	0.9966	0.039	1736	CHTG, ConA, Cyt c, BHb, HHb, bLGB, SubA, TRY
7d	NaSCN	0.17 M	1.26 M	0.02 ± 0.01	0.96 ± 0.02	7	0.9970	0.036	1657	CHY, CHTG, ConA, bLGA, RNase A, RNase B, SubA
7e	$NaClO_4$	0.094 M	0.43 M	0.03 ± 0.01	1.04 ± 0.02	9	0.9977	0.029	2993	CHY, ConA, Cyt c, bLGA, Pap, RNase A, RNase B, SubA, TRY

NaPB — sodium phosphate buffer, pH 7.4.

Table 6
Structural distances between examined proteins and cytochrome c used as a reference calculated with Eq. (8) from partition coefficient values in ATPS indicated[a].

Protein	D-1	D-2	D-1 & D-2 Distance-av.	D-3	D-4	D-3 & D-4 distance-av.
Cyt c	0	0	0	0	0	0
ConA	2.01	2.07	2.04 ± 0.04	2.71	2.76	2.74 ± 0.03
BHb	2.25	2.28	2.27 ± 0.02	2.50	2.53	2.52 ± 0.02
HHB	3.56	3.59	3.58 ± 0.02	4.29	4.17	4.23 ± 0.06
bLGA	4.75	4.78	4.80 ± 0.02	4.73	4.75	4.74 ± 0.01
bLGB	5.19	5.20	5.20 ± 0.01	5.20	5.24	5.22 ± 0.02
RNase A	6.05	5.90	6.0 ± 0.1	9.79	9.67	9.73 ± 0.06
RNase B	9.93	9.71	9.8 ± 0.2	10.34	10.23	10.29 ± 0.06
TRY	10.39	10.27	10.3 ± 0.1	15.72	15.67	15.70 ± 0.03
CHY	14.07	13.86	14.0 ± 0.2	16.9	16.7	16.8 ± 0.1
SubA	29.41	29.10	29.3 ± 0.2	42.6	42.4	42.5 ± 0.1
Pap	36.5	35.8	36.2 ± 0.5	74.7	73.5	74.1 ± 0.6
CHTG	36.99	36.51	36.8 ± 0.1	66.6	66.1	66.4 ± 0.2
HEL	**15.73**	**42.61**	**30 ± 19**	**103**	**151**	**127** ± 24

In bold — protein with position in the order of distances varying depending on the conditions the distances were estimated under.

[a] Distance D-1 calculated from *K*-values for proteins in ATPS# 1, 2, 4, 8; Distance D-2 calculated from *K*-values for proteins in ATPS# 1, 2, 4, 10; Distance D-3 calculated from *K*-values for proteins in ATPS# 1, 3, 5, 9; Distance D-4 calculated from *K*-values for proteins in ATPS# 1, 3, 5, 11; Distance-av — calculated as average values from D-1and D-2 and from D-3 and D-4. Compositions of various ATPSs are shown in Table 2.

proteins in the presence of 0.1.26 M NaSCN (ATPS # 11) instead of those in the presence of 0.43 M $NaClO_4$ (ATPS # 9), all salt additives in 0.01 M NaPB. The resulting structural distance values are listed in Table 6 as D-3 and D-4 values. The structural distances D-3 and D-4 calculated for the proteins as described above are plotted against each other in Fig. 4 and it can be seen that the above Eq. (10) holds for D-3 and D-4. Averaged D-values for low salt concentrations and for high salt concentrations are also listed in Table 6. The linear relationship between these avarged distances may be described as:

$$D^{av}{}_{3\text{-}4} = \text{-}0.2_{\pm 0.4} + 1.16_{\pm 0.05} {}^{*} D^{av}{}_{1\text{-}2} \tag{11}$$

$$N = 8; R^2 = 0.9871; SD = 0.65; F = 459.7$$

where $D^{av}{}_{1-2}$ and $D^{av}{}_{3-4}$ are averaged structural distances at low and high salt concentrations employed; all the other parameters as defined above. Structural distances for CHTG, HEL, RNase A, Pap, SubA, and TRY differ at low and high salt concentrations very noticeably and do not fit the above relationship.

3.4. Evaluating correlations between the partition parameters and structural descriptors

In order to find a correlation between the structural distances evaluated for various proteins based on their partition behavior and of 3D structures of 12 proteins analyzed in this study (see Table 1), various structural descriptors were derived based on the analysis of the amino acid sequences and corresponding 3D structures. Among the mentioned structural descriptors there were three descriptors, chain length, molecular weight, and isoelectric point (pI), obtained from the direct amino acid sequence analysis using the ExPASy ProtParam tool (http://web.expasy.org/protparam/) [27]. We also collected 47 descriptors that were computed from the proteins' 3D structures and which quantified 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 means:

- YASARA (http://www.yasara.org/) to generate three types of radii of the protein structure (radius of gyration, nuclear and Van der Waals radii), six measures of secondary structure (content of α-helix, 3_{10}-helix, both helix types, β-sheet, turns and coils), molecular mass, B-factors and occupancy (12 descriptors).

Table 7
Structural descriptors used to find a correlation between the protein structure peculiarities and protein partition behavior and to build a regression model.

Proteins	Structural distances	Descriptors					Output from regression
		x_1 # cavities on surface	x_2 fraction beta-bridges	x_3 hydrophobicity of surface	x_4 average B-factor	x_5 fraction beta-structures	
CHY	13.965	0.02	4.15	−1.98	17.30	38.59	17.64
CHTG	36.75	0.02	5.71	−3.70	30.70	45.71	35.26
bLGA	4.765	0.01	0.00	−0.62	40.80	41.98	6.38
bLGB	5.195	0.00	0.00	−3.65	25.00	82.10	3.47
RNase A	5.975	0.00	2.42	−3.81	20.20	35.48	10.74
TRY	10.33	0.00	3.49	−3.50	18.30	34.93	8.86
HHB	3.575	0.01	0.00	−1.90	20.00	0.00	6.59
Cyt c	0	0.00	1.92	0.00	29.10	1.92	−3.50
ConA	2.04	0.01	0.84	−2.23	15.20	46.84	1.11
Pap	36.15	0.02	4.25	−4.03	17.90	22.17	36.44
BHb	2.265	0.00	0.00	−1.67	24.30	0.00	1.76
SubA	29.255	0.02	1.82	−3.59	12.60	19.71	25.52
Correlation		0.81	0.74	−0.62	−0.19	0.07	0.98

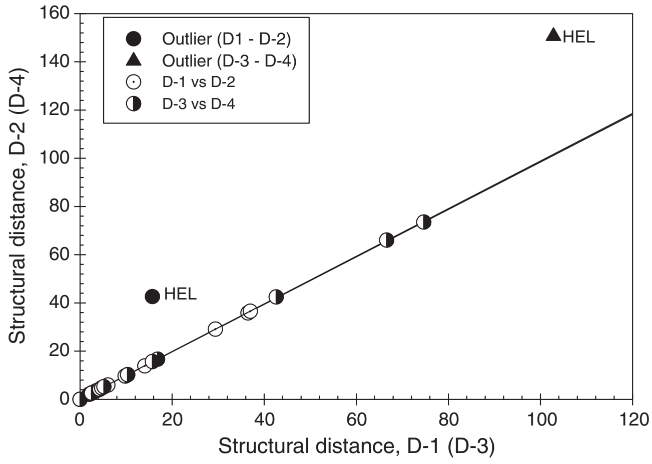

Fig. 4. Structural distances between examined proteins and cytochrome c used as a reference calculated with Eq. (8) from partition coefficient values in different ATPS indicated in Table 6.

— DSSP (http://swift.cmbi.ru.nl/gv/dssp/) [28] to derive surface area and secondary structure. We computed eight descriptors to characterize size and properties of the surface: fraction of residues on the surface; fraction of polar, neutral, positively charged, and negatively charged residues on the surface; and hydrophobicity of surface residues quantified using three amino acids indices including Kyte-Doolittle [29,30], Eisenberg [31], and Cid scales [32]; and nine measures of secondary structure (content of α-helix, 3_{10}-helix, β-sheet, β-bridge; turns, bends, coils; both helix types, and both β structure types; we note that there were no Π-helices in our protein set) (17 descriptors).

— Voronoia (http://proteinformatics.charite.de/voronoia4rna/tools/v4rna/index) [33] to characterize pockets/cavities in the structure and compute packing. We computed 10 descriptors of pockets and four descriptors of packing (average Van der Waals volume, solvent-excluded volume, fraction of buried atoms and average packing density) (14 descriptors).

— MFDp (http://biomine-ws.ece.ualberta.ca/MFDp.html) [34] to quantify propensity of a given protein for intrinsic disorder. We computed the disorder content (fraction of disordered residues), number of disordered segments normalized by the protein size, and average disorder score (3 descriptors).

— Based on [35], we also computed contact order (1 descriptor).

Supplementary Table 1 lists all the structural parameters (or descriptors) derived for 12 proteins. Then, we quantified correlation between the individual structural descriptors and the partition-based structural distances by computing the Person correlation coefficients (PCCs) between each of the 50 structural descriptors listed above and the experimentally observed values derived for the given ATPS. The highest PCC value for a correlation between an individual descriptor and different sets of observed data for partition behavior was 0.81 (for D-1, D-2, and 0.01 M NaPB). The structural descriptor with the highest correlation to the partition behavior of proteins consistently was normalized (by the sequence size) number of water-filled cavities with heteroatoms removed (see Fig. 5A). The other protein properties

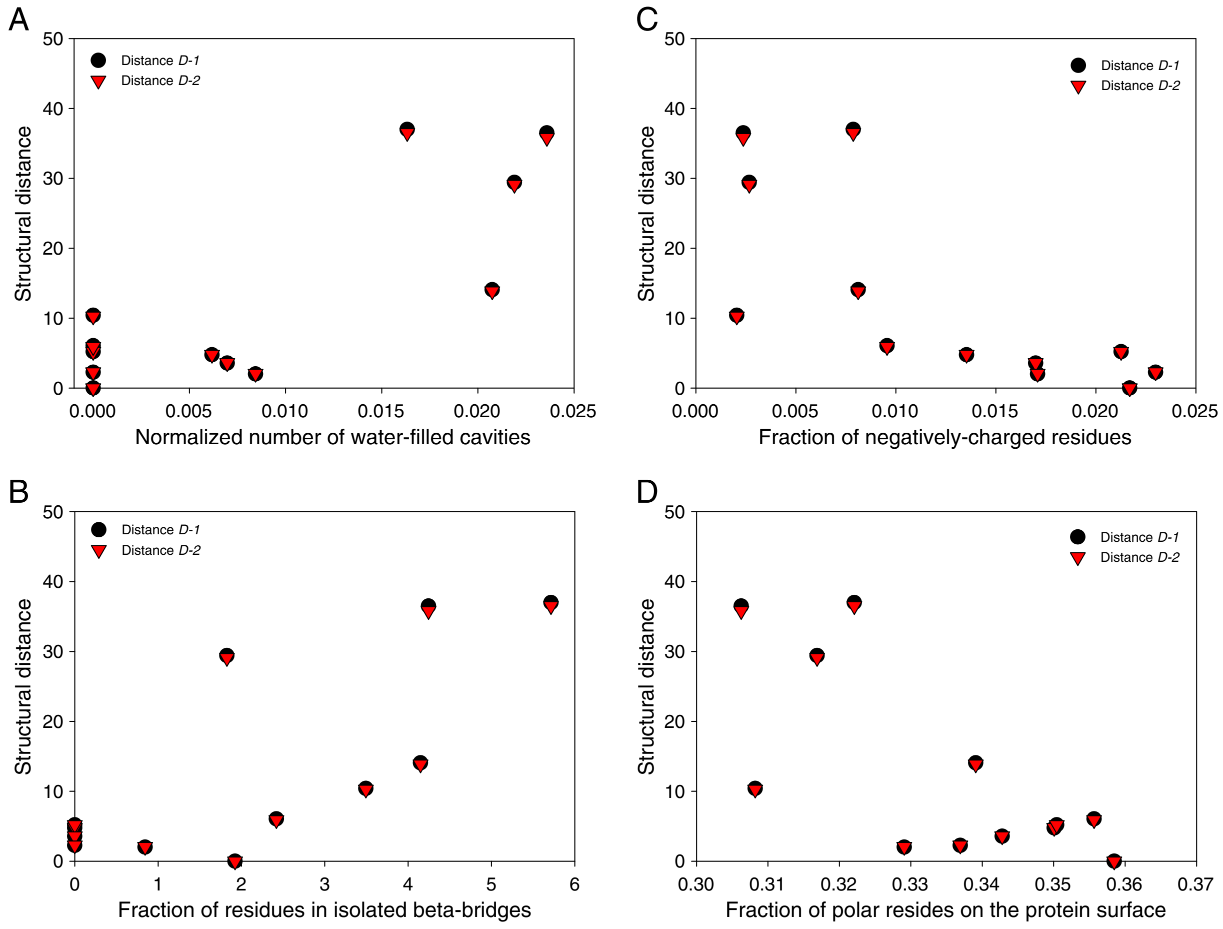

Fig. 5. Comparison of four structural distances measured for 12 proteins in various ATPSs with some structural descriptors derived based on the proteins' 3D structure analysis. Four structural descriptors possess highest PCC values with experimentally determined structural distances: A. Normalized (by the sequence size) number of water-filled cavities with heteroatoms removed. B. The fraction of residues in isolated β-bridges. C. The fraction of negatively charged residues on the protein surface. D. The fraction of polar resides on the protein surface.

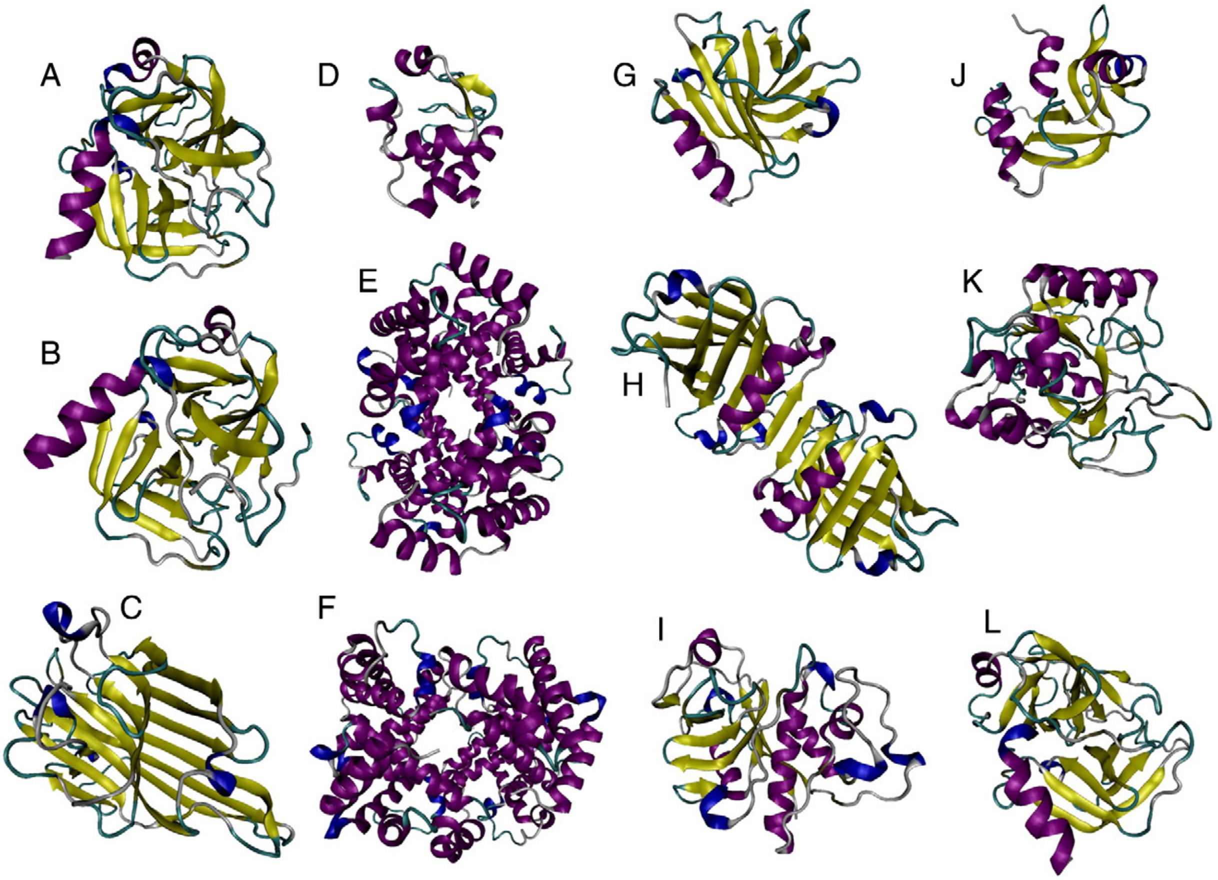

Fig. 6. The X-ray structure gallery of 12 model proteins analyzed in this study. A. α-chymotrypsin (PDB ID: 1AB9); B. α-chymotrypsinogen A (PDB ID: 1ACB); C. concanavalin A (PDB ID: 1JBC); D. cytochrome c (PDB ID: 1HRC); E. bovine hemoglobin (PDB ID: 2QSS); F. human hemoglobin (PDB ID: 1BZ0); G. β-lactoglobulin A (PDB ID: 1B8E); H. β-lactoglobulin B (PDB ID: 1BEB); I. papain (PDB ID: 1PPN); J. ribonuclease A (PDB ID: 1BEL); K. subtilisin A (PDB ID: 3UNX); and L. trypsinogen (PDB ID: 1BTY).

with high PCC values include fraction of residues in isolated β-bridges (Fig. 5B), fraction of negatively charged residues on the protein surface (Fig. 5C), and fraction of polar resides on the protein surface (Fig. 5D). Curiously, all these structural parameters with high PCC values are related to some features of the protein surface (see Table 7). This observation suggests that the partition behavior of a given protein is mostly determined by the peculiarity of its surface.

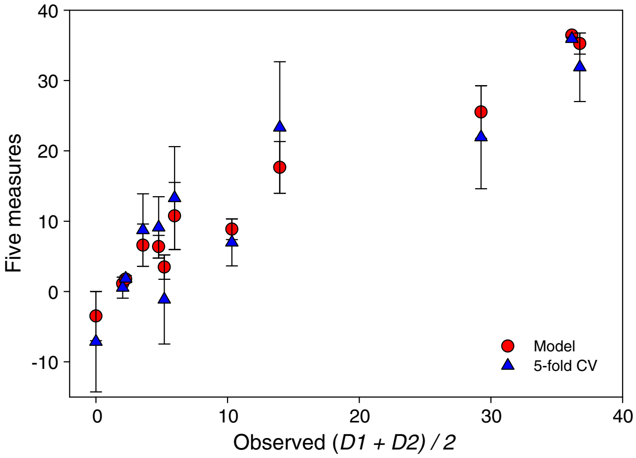

Fig. 7. Dependence of the values derived from the regression modeling with five structural parameters on the corresponding averaged structural distances measured for 12 proteins in various ATPSs in the presence of 0.01 M NaPB (red circles). The averaged structural distances were calculated as (D-1 + D-2)/2. The corresponding data for the five-fold cross validation analysis are also shown by blue triangles.

3.4.1. Regression model for 0.01 M NaPB (average of D-1 and D-2)

Next, we performed best-first search to find a subset of descriptors that maximize PCC with the observed values. First, each descriptor was normalized to the [−1, 1] interval using maximum absolute value, and next these descriptors were sorted in the descending order by their absolute PCC values with the observed data. We initialized the set of selected descriptors with the descriptor with the highest PCC value. We added a subsequently ranked descriptor into the set of selected descriptors if its inclusion increased the PCC value with the observed values by at least 0.02 when compared with the PCC obtained on the set of descriptors without this descriptor. We scanned the entire list of descriptors once and we measured the PCC values based on five-fold cross validation on the considered set of 12 proteins. To visualize the structural differences between these model proteins, Fig. 6 represents an X-ray structure gallery of these proteins and clearly shows that the model proteins belong to different structural classes.

For the ATPSs containing 0.01 M NaPB, this regression modeling resulted in selection of five descriptors (see Table 7): x_1 normalized (by the sequence size) number of water-filled cavities with hetero atoms removed; x_2 fraction of isolated β-bridges; x_3 average hydrophobicity value (in terms of the Kyte-Doolittle scale [29,30]) for the residues on the surface (here, a given residue is defined to be on the surface if its solvent accessible surface area (computed with DSSP) is larger than 0.75; we empirically selected this threshold to maximize the PCC with the observed data); x_4 average B-factor of the protein molecule; and x_5 fraction of residues involved in the formation of β-structure (including β-sheets and β-bridges) extracted with DSSP. The corresponding regression is formulated as follows:

$$y=917.30_{\pm148.19}x_1+1.27_{\pm0.75}x_2-6.49_{\pm1.29}x_3+0.71_{\pm0.19}x_4-0.14_{\pm0.06}x_5-26.35_{\pm5.99}. \quad (12)$$

The above regression utilizes raw values of the five descriptors. We estimate the relative contributions of individual descriptors in the regression by normalizing values of these descriptors and scaling the corresponding absolute values of coefficients to sum to 1. The recomputed absolute coefficients are 0.226, 0.076, 0.274, 0.303, and 0.122, respectively, and they indicate that x_4, x_3 and x_1 are the main determinants $(0.303 + 0.274 + 0.226 = 0.803$ out of 1) of the structural distance defined by the regression. This reveals that the value of the structural distance is primarily influenced by the average flexibility (measured with B-factors) and two properties of the protein surface: number of cavities and hydrophobicity. We also note that all six coefficients are statistically significant with *p*-values below 0.001.

The outputs of regression are characterized by relatively high PCC value with the observed data that equals 0.98 (0.91 based on the five-fold cross validation), which is noticeably larger that the PCC of 0.81 calculated for the best single descriptor, x_1.

These results are further illustrated by Fig. 7, which represents the dependencies of the values derived by the regression modeling with five structural parameters on the corresponding averaged structural distances measured for these 12 proteins in the ATPSs containing 0.01 M NaPB. The averaged structural distances were calculated as (D-1 + D-2)/2. The corresponding data for the five-fold cross validation analysis are also shown for comparison.

Based on the results of these computational analyses, it is obvious that partition behavior of proteins is determined by the peculiarities of their surfaces (e.g., the number of water-filled cavities and the averaged hydrophobicity of the surface residues). Another important point is that partition of proteins in ATPSs with low NaPB content (0.01 M) is also dependent on the intrinsic flexibility of the protein structure, measured in terms of the B-factor, which in crystal structures of macromolecules reflects the uncertainty in atom positions in the model and often represents the combined effects of thermal vibrations and static disorder [36]. Therefore, the B-factor of the α-carbon and the B-factor averaged over the four backbone atoms are the commonly used measures of residue flexibility of folded proteins [37–39]. It is known that besides the regions of missing electron density, crystallized proteins often contain regions with high B-factor. In order to differentiate between flexible but ordered regions and intrinsically disordered regions, comparisons were made among four categories of protein flexibility: low-B-factor ordered regions, high-B-factor ordered regions, short disordered regions, and long disordered regions (with two last categories being selected as the short and long regions of missing electron density, respectively) [40]. This analysis revealed that the high-B-factor regions were more similar to intrinsically disordered regions than to ordered regions with low-B-factor. Furthermore, the observed distinctive amino acid biases of high-B-factor ordered regions, short disordered regions, and long disordered regions clearly indicated that the sequence determinants for these flexibility categories differ from one another, suggesting that the amino acid attributes that specify flexibility and intrinsic disorder are distinct and not merely quantitative differences on a continuum [40].

Data represented in this and previous studies (e.g., see [41–45]) suggest that the *K*-values retrieved based on the protein partition in different ATPSs is highly sensitive to the structural changes in proteins. In fact, this parameter was shown to reflect interactions between the solvent-exposed groups of the protein with the two aqueous solvent environments in ATPS [12,46–48] and with co-solutes present in two aqueous phases. For example, the partition behavior of the prostate-specific antigen (PSA) in aqueous Dextran–Ficoll two-phase system was shown to be sensitive to the presence of other proteins, such as bovine or human serum albumin, human transferrin, and human gamma-globulin [6]. Curiously, no specific interactions between the PSA and these proteins were found, suggesting that the effect of protein-additives on the partition behavior of free PSA can be explained by the existence of non-specific PSA-protein interactions (formation of the PSA-protein encounter complexes) affecting the PSA conformation [6]. Furthermore, the propensity for intrinsic disorder (i.e., the propensity for high conformational dynamics) was related to the PSA partition-modulating capability of the proteins [6].

It was pointed out that application of ATPSs implies the use of high concentrations of two polymers in water when a certain threshold concentration of the polymers is exceeded, and that these levels of polymer concentrations are similar to those commonly used to mimic the effects of macromolecular crowding on proteins [49]. Recent experimental and computational analyses of the effects of macromolecular crowding on the dynamics of several intrinsically disordered proteins (IDPs, such as prothymosin α, α-synuclein, and TC1) [50] supported the model where IDPs retained the segmental motions on the nanosecond timescale under crowded conditions and function as dynamic structural ensembles in cellular environments. This conclusion is based on the observations that IDPs remained at least partially disordered in the crowded environment, and that crowding possessed differential effects on the conformational propensity of the different regions of IDPs, with some of these regions being unaffected by crowding, and with other regions (potentially related to certain target-binding motifs) being selectively stabilized due to the presence of high concentration of other macromolecules [50]. Therefore, crowding might cause limited structural changes in IDPs, and the degree of these structural changes reflects the functional requirements of these highly mobile and promiscuous proteins [51,52]. Not only IDPs might be affected by macromolecular crowding. For example, computational molecular dynamics analysis revealed that the molecular crowing possesses large effects on the enzymatic conformational dynamics, the average enzymatic cycle time, characteristic times of internal conformational motions and transport properties of the adenylate kinase, and that the corresponding effects were dependent on the concentration and size of crowding agents [53].

We believe that the partition coefficient *K* can be used as a general-purpose numerical index to characterize the 3D structure and that the partition behavior-based method described in this study represents an important addition to the set of existing experimental and computational tools for the analysis of structural and dynamic properties of proteins as well as for the accurate description of the effects of ions and co-solvents on the structural features and conformational behavior of proteins.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2014.01.016.

References

[1] P.A. Albertsson, Partition of Cell Particles and Macromolecules, 3rd ed. Wiley, New York, 1986.

[2] In: H. Walter, D.E. Brooks, D. Fisher (Eds.), Partitioning in Aqueous Two-Phase Systems: Theory, Methods, Use, and Applications to Biotechnology, Academic Press, Orlando, FL, 1985.

[3] P.A. Rosa, A.M. Azevedo, S. Sommerfeld, W. Backer, M.R. Aires-Barros, Aqueous two-phase extraction as a platform in the biomanufacturing industry: economical and environmental sustainability, Biotechnol. Adv. 29 (2011) 559–567.

[4] B.Y. Zaslavsky, Aqueous Two-phase Partitioning: Physical Chemistry and Bioanalytical Applications, Marcel Dekker, New York, 1994.

[5] 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.

[6] O. Fedotoff, L.M. Mikheeva, A. Chait, V.N. Uversky, B.Y. Zaslavsky, Influence of serum proteins on conformation of prostate-specific antigen, J. Biomol. Struct. Dyn. 29 (2012) 1051–1064.

[7] M. Stovsky, L. Ponsky, S. Vourganti, P. Stuhldreher, M.B. Siroky, V. Kipnis, O. Fedotoff, L. Mikheeva, B. Zaslavsky, A. Chait, J.S. Jones, Prostate-specific antigen/solvent interaction analysis: a preliminary evaluation of a new assay concept for detecting prostate cancer using urinary samples, Urology 78 (2011) 601–605.

[8] P.P. Madeira, A. Bessa, L. Alvares-Ribeiro, M. Raquel Aires-Barros, A.E. Rodrigues, V.N. Uversky, B.Y. Zaslavsky, Amino acid/water interactions study: a new amino acid scale, J. Biomol. Struct. Dyn. (32) (2014).

[9] P.P. Madeira, C.A. Reis, A.E. Rodrigues, L.M. Mikheeva, B.Y. Zaslavsky, Solvent properties governing solute partitioning in polymer/polymer aqueous two-phase systems: nonionic compounds, J. Phys. Chem. B 114 (2010) 457–462.

[10] L.A. Ferreira, P. Parpot, J.A. Teixeira, L.M. Mikheeva, B.Y. Zaslavsky, Effect of NaCl additive on properties of aqueous PEG-sodium sulfate two-phase system, J. Chromatogr. A 1220 (2012) 14–20.

[11] P.P. Madeira, A. Bessa, L. Alvares-Ribeiro, M.R. Aires-Barros, C.A. Reis, A.E. Rodrigues, B.Y. Zaslavsky, Salt effects on solvent features of coexisting phases in aqueous polymer/polymer two-phase systems, J. Chromatogr. A 1229 (2012) 38–47.

[12] 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.

[13] M.L. Moody, H.D. Willauer, S.T. Griffin, J.G. Huddleston, R.D. Rogers, Solvent property characterization of poly(ethylene glycol)/dextran aqueous biphasic systems using the free energy of transfer of a methylene group and a linear solvation energy relationship, Ind. Eng. Chem. Res. 44 (2005) 3749–3760.

[14] H.D. Willauer, J.G. Huddleston, R.D. Rogers, Solvent properties of aqueous biphasic systems composed of polyethylene glycol and salt characterized by the free energy of transfer of a methylene group between the phases and by a linear solvation energy relationship, Ind. Eng. Chem. Res. 41 (2002) 2591–2601.

[15] 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-Proteins Proteomics 1834 (2013) 583–592.

[16] L. Ferreira, P.P. Madeira, L. Mikheeva, V.N. Uversky, B. Zaslavsky, Effect of salt additives on protein partition in polyethylene glycol-sodium sulfate aqueous two-phase systems, Biochim. Biophys. Acta. 2014 (1834) 2859–2866.

[17] 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, Vol, vol. I, Humana Press, Totowa, NJ, Experimental Techniques, 2012, pp. 351–361.

[18] A. Chait, B. Zaslavsky, in: USPTO (Ed.), Characterization of Molecules, vol. 7,968,350, Analiza, Inc., Bay Village, OH, USA, 2011.

[19] J.A. Asenjo, B.A. Andrews, Aqueous two-phase systems for protein separation: phase separation and applications, J. Chromatogr. A 1238 (2012) 1–10.

[20] P.P. Madeira, J.A. Teixeira, E.A. Macedo, L.M. Mikheeva, B.Y. Zaslavsky, "On the Collander equation": protein partitioning in polymer/polymer aqueous two-phase systems, J. Chromatogr. A 1190 (2008) 39–43.

[21] P. Madeira, J.A. Teixeira, E.A. Macedo, L.M. Mikheeva, B.Y. Zaslavsky, Correlations between distribution coefficients of various biomolecules in different polymer/polymer aqueous two-phase systems, Fluid Phase Equilib. 267 (2008) 150–157.

[22] R. Collander, On lipoid solubility, Acta Physiol. Scand. 13 (1947) 363–381.

[23] A. Leo, C. Hansch, Linear free energy relations between partitioning solvent systems, J. Org. Chem. 36 (1971) 1539–1544.

[24] A. Leo, C. Hansch, D. Elkins, Partition coefficients and their uses, Chem. Rev. 71 (1971) 525–616.

[25] C. Hansch, W.J. Dunn III, Linear relationships between lipophilic character and biological activity of drugs, J. Pharm. Sci. 61 (1972) 1–19.

[26] B.Y. Zaslavsky, L. Mikheeva, S. Rogozhin, Comparison of conventional partitioning systems used for studying the hydrophobicity of polar organic compounds, J. Chromatogr. 216 (1981) 103–113.

[27] 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.

[28] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.

[29] D. Juretic, B. Lucic, 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.

[30] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.

[31] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, The hydrophobic moment detects periodicity in protein hydrophobicity, Proc. Natl. Acad. Sci. U. S. A. 81 (1984) 140–144.

[32] H. Cid, M. Bunster, M. Canales, F. Gazitua, Hydrophobicity and structural classes in proteins, Protein Eng. 5 (1992) 373–375.

[33] K. Rother, P.W. Hildebrand, A. Goede, B. Gruening, R. Preissner, Voronoia: analyzing packing in protein structures, Nucleic Acids Res. 37 (2009) D393–D395.

[34] 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.

[35] 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.

[36] G. Rhodes, Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models, Academic Press, San Diego, CA, 1993.

[37] P.A. Karplus, G.E. Schulz, Prediction of chain flexibility in proteins, Naturwissenschaften 72 (1985) 212–213.

[38] M. Vihinen, E. Torkkila, P. Riikonen, Accuracy of protein flexibility predictions, Proteins 19 (1994) 141–149.

[39] S. Kundu, J.S. Melton, D.C. Sorensen, G.N. Phillips Jr., Dynamics of proteins in crystals: comparison of experiment with simple models, Biophys. J. 83 (2002) 723–732.

[40] P. Radivojac, Z. Obradovic, D.K. Smith, G. Zhu, S. Vucetic, C.J. Brown, J.D. Lawson, A.K. Dunker, Protein flexibility and intrinsic disorder, Protein Sci. 13 (2004) 71–80.

[41] F.D. Raymond, D.W. Moss, D. Fisher, Phase partitioning detects differences between phospholipase-released forms of alkaline phosphatase—a GPI-linked protein, Biochim. Biophys. Acta 1156 (1993) 117–122.

[42] C. Hassinen, K. Kohler, A. Veide, Polyethylene glycol-potassium phosphate aqueous two-phase systems. Insertion of short peptide units into a protein and its effects on partitioning, J. Chromatogr. A 668 (1994) 121–128.

[43] A. Sakurai, M. Katai, T. Miyamoto, K. Ichikawa, K. Hashizume, Ligand- and nuclear factor-dependent change in hydrophobicity of thyroid hormone beta1 receptor, Thyroid 8 (1998) 343–352.

[44] C. Ramsch, L.B. Kleinelanghorst, E.A. Knieps, J. Thommes, M.R. Kula, Aqueous two-phase systems containing urea: influence of protein structure on protein partitioning, Biotechnol. Bioeng. 69 (2000) 83–90.

[45] K. Becker, J. Van Alstine, L. Bulow, Multipurpose peptide tags for protein isolation, J. Chromatogr. A 1202 (2008) 40–46.

[46] B.Y. Zaslavsky, D. Chaiko, A new analytical methodology for quality control testing of biological and recombinant products, in: J. Shillenn (Ed.), Validation Practices for Biotechnology Products, ASTM STP 1260, American Society for Testing and Materials, Philadelphia, 1996, pp. 107–122.

[47] 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 1596 (2002) 253–268.

[48] 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.

[49] L. Breydo, L.M. Mikheeva, P.P. Madeira, B.Y. Zaslavsky, V.N. Uversky, Solvent interaction analysis of intrinsically disordered proteins in aqueous two-phase systems, Mol. Biosyst. 9 (2013) 3068–3079.

[50] E.A. Cino, M. Karttunen, W.Y. Choy, Effects of molecular crowding on the dynamics of intrinsically disordered proteins, PLoS One 7 (2012) e49876.

[51] V.N. Uversky, Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding, Protein J. 28 (2009) 305–325.

[52] C.S. Szasz, A. Alexa, K. Toth, M. Rakacs, J. Langowski, P. Tompa, Protein disorder prevails under crowded conditions, Biochemistry 50 (2011) 5834–5844.

[53] C. Echeverria, R. Kapral, Molecular crowding and protein enzymatic dynamics, Phys. Chem. Chem. Phys. 14 (2012) 6755–6763.