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