Chemometrics and Intelligent Laboratory Systems 149 (2015) 12–21

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Review

Optimization of PEG–salt aqueous two-phase systems by design of experiments

Anna Glyk, Dörte Solle, Thomas Scheper, Sascha Beutel *

Institute of Technical Chemistry, Leibniz University of Hannover, Callinstr. 5, 30167 Hannover, Germany

ARTICLE INFO

Article history:
Received 9 July 2015
Received in revised form 25 September 2015
Accepted 26 September 2015
Available online 9 October 2015

Keywords:
Aqueous two-phase systems
Design of experiments
Factorial experimental design
Response surface methodology
Central composite design
Analysis of variance

ABSTRACT

Since the mechanism governing the partitioning behavior of biomolecules, such as proteins and enzymes, in polyethylene glycol (PEG)–salt aqueous two-phase systems (ATPS) is complex and not easily predictable, many laborious experiments have to be performed for an optimization of these systems, causing increased overall cost. However, the multivariate statistical design of experiments (DoE) methodology is representing a promising and efficient optimization technique which can overcome the limitations of traditional optimization methods. Therefore, DoE has emerged as a powerful and efficient optimization tool for PEG–salt ATPS, since it is faster, more efficient and cost-effective, allowing a simultaneous and rigorous evaluation of process/system parameters. In the present review, different DoE process steps are represented to highlight the feasibility of this approach to operate as a promising and efficient optimization tool, thus facilitating the evaluation of the partitioning behavior, recovery and purification of different proteins and enzymes in PEG–salt ATPS. In this context, several experimental designs, such as factorial and response surface designs, have been discussed and evaluated by statistical regression analysis and analysis of variance (ANOVA), as well as various applications of PEG–salt ATPS using DoE have been outlined which may further promote the optimization of these systems.

Contents

1. Introduction . . . 12
2. Theory and methodology . . . 13
2.1. Screening of significant factors . . . 13
2.2. Steepest ascent/descent method . . . 15
2.3. Optimization by RSM . . . 15
2.4. Analysis of model . . . 17
2.5. Validation of model . . . 18
3. Applications of PEG–salt ATPS using DoE . . . 18
4. Conclusions . . . 20
References . . . 20

1. Introduction

Since partitioning, recovery and purification of biomolecules, such as proteins and enzymes, in polyethylene glycol (PEG)–salt aqueous two-phase systems (ATPS) are influenced by several system parameters and physicochemical/surface properties of the target biomolecules, the mechanism governing the partitioning behavior of biomolecules in ATPS is complex and cannot be easily predicted [1–6]. Thus, many experiments have to be performed for the optimization of ATPS which is tedious, leading to increased overall cost [7]. One conventional optimization method is the one-factor-at-a-time (OFAT) approach consisting

Abbreviations: PEG, polyethylene glycol; ATPS, aqueous two-phase systems; OFAT, one-factor-at-a-time; design of experiments, DoE; RSM, response surface methodology; FFD, full factorial design; fFD, fractional factorial design; ANOVA, analysis of variance; PBD, Plackett–Burman design; CCD, central composite design; CCC, central composite circumscribed; CCF, central composite face-centered; BBD, Box–Behnken design; MLR, multiple linear regression; *F*-test, Fisher's statistical test; *t*-test, Student's *t*-test; *p*-value, probability value; SS, sum of squares; LOF, lack-of-fit; MS, mean squares; DF, degrees of freedom; SOR, significance of regression; R^2, coefficient of determination; R^2_{adj}, adjusted coefficient of determination; Q^2, coefficient of predicted variation; *K*, partition coefficient; *Y*, product recovery/activity yield (%); *PF*, purification factor; *S*, selectivity; *P*, purity (%).

* Corresponding author. Tel.: +49 511 762 2868; fax: +49 511 762 3004.
E-mail address: beutel@iftc.uni-hannover.de (S. Beutel).

http://dx.doi.org/10.1016/j.chemolab.2015.09.014
0169-7439/

of a chosen starting point or baseline set of levels for each factor, the determination of significant process/system factors and a consecutive change of each factor at a time over a certain range by keeping all the other factors constant at the baseline level [7–9]. The major disadvantage of this approach is that possible interactions between the factors are not considered, and hence no combined effects of all factors are depicted [9,10]. Thus, OFAT experiments are often unreliable, usually leading to poor results or false optimal conditions/factors determined for a process/system, and this method is generally inefficient, time-consuming and laborious, since many experiments are required [7–9, 11]. Nowadays, the optimization of partitioning processes in PEG–salt ATPS is widely carried out by a multivariate statistical technique called design of experiments (DoE) which is based on a statistical factorial experimental design concept, consisting of the performance of a few experiments at a particular factor level combination [7,9,12].

In general, experiments are used to investigate the performance of a process/system [9]. Thereby a process/system (Fig. 1) can be generally described as a combination of methods, operations, machines, people and other resources inducing a transformation of inputs (often materials) into outputs [9,13]. The measured responses $Y_1, Y_2, \ldots, Y_n$ are used to characterize the performance/quality of an investigated process/system, providing information about its properties and general conditions [13]. On the other hand, factors are representing tools for the manipulation of a process/system [13]. By altering the most influential factors, the process/system features might be changed according to a desired response profile [13], at which some of the factors $X_1, X_2, \ldots, X_n$ (*e.g.*, temperature) are controllable, whereas other factors $Z_1, Z_2, \ldots, Z_n$ (*e.g.*, climate) are uncontrollable [8,9]. The performance of a designed experiment is usually related to the determination of the effects of changed controllable input factors ($X_1, X_2, \ldots, X_n$) on the corresponding varied output responses ($Y_1, Y_2, \ldots, Y_n$) of a process/system [8,9,13].

Since all variables involved in a process/system have a significant influence on the experimental performance, a designed experiment may include the following objectives according to [8,9]:

- Determination of the most influential variables on responses Y_1, $Y_2, \ldots, Y_n$,
- Determination of optimal settings of influential $X_1, X_2, \ldots, X_n$ resulting in Y which is almost always near the desired nominal value
- Determination of optimal settings of influential $X_1, X_2, \ldots, X_n$ resulting in a small variability in $Y_1, Y_2, \ldots, Y_n$
- Determination of optimal settings of influential X_n minimizing the effects of uncontrollable factors $Z_1, Z_2, \ldots, Z_n$.

Moreover, several significant factors can be changed and optimized simultaneously in the multivariate DoE approach in contrast to the OFAT method [7,9]. Therefore, the DoE approach is faster, more efficient and cost-effective for a rigorous evaluation of the significant factors on the selected responses and their possible interactions using a mathematical model (usually a quadratic/second-order or higher polynomial function) [13], thus overcoming the limitations of the OFAT method. Overall, a DoE process for the optimization of PEG–salt ATPS is consisting of the following steps:

- Screening of significant factors
- Steeptest ascent/descent method
- Optimization by response surface methodology (RSM)
- Analysis of model
- Validation of model

All of these steps are highlighted in Fig. 2 and discussed in the following chapters.

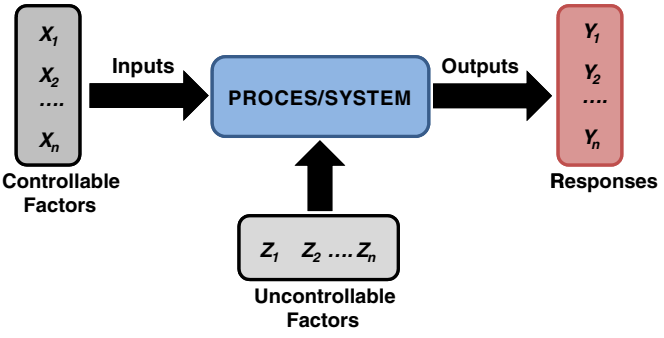

Fig. 1. Schematic representation of a general process/system, adapted from [8,9,13].

2. Theory and methodology

2.1. Screening of significant factors

Initially, in a DoE process for the optimization of PEG–salt ATPS a screening of a large number of factors k is carried out in a few experiments in order to reveal the most important factors having a statistical significant influence on the output responses or performance of a process/system and investigate their appropriate ranges [7,8,13,14]. Furthermore, the purpose of a screening design is the identification and selection of those factors demanding a more thorough investigation in further experiments [8,9]. In general, screening of significant factors in PEG–salt ATPS is widely carried out by the full factorial design (FFD) and fractional factorial design (fFD). These experimental design approaches are generally performed by assigning all factors k at two-levels, usually denoted by a high (+1 or simply +) and low (−1 or just −) level for each factor [7–9,13]. These notations are often called the orthogonal/effect coding and are consequently used to construct an orthogonal arrangement or design of experiments [9,13].

The two-level FFD is carried out at all possible factor level combinations for all factors and is consisting of a set of experimental runs with a total number of 2^k [7–9,13]. For instance, the 2^2 FFD for two factors (*e.g.*, A and B), is resulting in four possible experimental runs at the four factor combinations: (−, −), (+, −), (−, +) and (+, +), according to the first four rows in the design matrix in Fig. 3a [7–9,13]. Furthermore, this design can be described as a square containing an experimental region of a regular geometry (Fig. 3b), at which each row in the design matrix is corresponding to one experiment depicted as a point in the two-dimensional factor space [13]. In addition, a 2^3 FFD for three factors (*e.g.*, A, B and C) can be designed similarly to the 2^2 FFD using high (+) and low (−) factor levels leading to eight possible experiments at eight different factor combinations, as summarized in the eight rows in the design matrix in Fig. 3a, while the corresponding experimental region can be illustrated by a cube of a regular geometry (Fig. 3c) [9,13].

Furthermore, the interaction effects among factors on a response can be determined in FFD's by a simultaneous alteration of factor levels, at which a linear response over a selected factor level range is assumed in these designs because each factor has only two-factor levels [7–9]. However, with five or more factors in a 2^k factorial design the number of experiments is drastically increased, resulting in a fast outgrowing of the resources of most experimenters in terms of too demanding experiments [7–9,13]. Hence, the two-level fFD is usually used as a more appealing experimental screening design which is based on the FFD and may be constructed by choosing only fractions of corner experiments [8,13–15]. Generally, the fFD is denoted by 2^{k-1}, 2^{k-2} and 2^{k-4} containing a total number of experimental runs reduced to a one-half, one-quarter or a higher fraction of the FFD [7–9,13]. For instance, a 2^{3-1} fFD is resulting in four experimental runs of three factors (*e.g.*, A, B and C), as shown exemplarily in a cube of a regular geometry with coded units in Fig. 3d, according to the rows 1, 4, 6 and 7 in the design matrix in Fig. 3a [9,13].

Moreover, FFD and fFD usually include 3–5 replicated center-point experiments which are carried out for an evaluation of the pure