Fluid Phase Equilibria 422 (2016) 32–42

Contents lists available at ScienceDirect

Fluid Phase Equilibria

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

Modeling and prediction of protein solubility using the second osmotic virial coefficient

Marcel Herhut, Christoph Brandenbusch, Gabriele Sadowski*

Laboratory of Thermodynamics, Department of Biochemical and Chemical Engineering, Technische Universität Dortmund, Emil-Figge-Str. 70, 44227 Dortmund, Germany

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form
9 December 2015
Accepted 11 January 2016
Available online 13 January 2016

Keywords:
Protein solubility
DLVO model
Second osmotic virial coefficient
Protein crystallization
Lysozyme

ABSTRACT

The development of a precipitation or crystallization step requires knowing the solubility of the target protein and its crystallization behavior in aqueous solutions at different pH, temperatures and in the presence of precipitating agents, especially salts. Within this work, a solubility model for proteins based on the second osmotic virial coefficient B_{22} is developed. For this, a relation between protein solubility and B_{22} was combined with the extended DLVO model. This solubility model was then used to model and also predict the protein solubility of lysozyme and monoclonal antibody for different salts, salt concentrations, and pH. The modeled as well predicted B_{22} and protein solubility data of lysozyme in the presence of sodium chloride and sodium p-toluenesulfonate and of a monoclonal antibody in the presence of ammonium sulfate at different pH shows good agreement with experimental data.

1. Introduction

The industrial protein production has gained an increased interest in industrial and academic research within the last decade. Especially within the field of medical (red) biotechnology, where pharmaceutical proteins such as monoclonal antibodies (mAB) are produced, the number of processes on an industrial scale has been steadily increasing.

One major bottleneck in state-of-the-art industrial (pharmaceutical) protein production is the downstream processing. Historically, it is often accomplished by a series of cost-intensive chromatographic steps. This results from both, costly chromatographic material, as well as the low capacity of these workup steps. This leads to the fact, that the downstream processing itself can cover up to 80% of the total production costs. For economic and efficient processes there is thus a demand for alternative downstream processing concepts [1]. One alternative to chromatographic separation steps, already widely used in the chemical industry, is crystallization. This can be used either for the initial product capture (precipitation) or final product polishing (crystallization) [2]. Besides this, protein crystals usually have a high purity and a higher stability compared to proteins in solution which makes them attractive for storage and formulation, later needed in pharmaceutical processes [3].

For developing crystallization processes, the solubility of a protein in solution is the most important information. Unfortunately this quantity is not easily accessible. Protein solubility is influenced by the type of solvent or buffer used, the pH, the precipitating agent (e.g. salt, alcohol, polymer), and temperature. In state-of-the-art investigations for crystallization processes, a high experimental effort is applied for the screening of potential crystallization conditions. For this purpose, often methods like the sitting-drop or the hanging-drop method are used [4] and applied in well-plates covering up to 384 different crystallization conditions per plate, depending on the plate used. Once a potential crystallization condition is found, the protein solubility is then measured by determining the protein concentration of the mother liquor with UV absorption after equilibration [5].

In order to decrease this experimental effort and to simplify the development of crystallization processes, the prediction of the protein solubility, as function of parameters such as pH, kind of salt, salt concentration, or temperature, by a physically-based thermodynamic model is of high interest. First approaches to model protein solubility in aqueous solutions containing a salt have been published in open literature back in 1925 by Cohn [6]. Melander and Horvath [7] developed empirical equations to correlate protein solubility and hydrophobic effects of the protein in aqueous solutions. Unfortunately, as shown by Przybycien and Bailey [8,9], these

* Corresponding author.
E-mail address: g.sadowski@bci.tu-dortmund.de (G. Sadowski).

http://dx.doi.org/10.1016/j.fluid.2016.01.020
0378-3812/

empirical equations are only valid for conformationally robust proteins such as lysozyme or chymotrypsin. In 1998 Terry Jenkins [10] proposed three empirical equations, relating protein solubility and salt concentration in terms of either salt molarity, salt activity or water activity. As described by Naik and Bhagwat, the major drawback of these empirical equations is that protein solubility is not predictable for systems or conditions different from those used for parameter fitting [11]. One of the latest approaches was presented by Agena et al. [12] who used a UNIQUAC-based approach with temperature-dependent parameters to model the solubility of lysozyme and concanavalin in aqueous salt solution. Their results agreed qualitatively with experimental data but describes only the protein solubility as function of temperature but not of pH or salt concentration [11].

All of these models have in common, that they require a high amount of experimental data in order to fit parameters used for the modeling (e.g. UNIQUAC pure-component or binary interaction parameters). Furthermore the predictive capabilities are very limited, since Agena et al. e.g. did not account for the influence of pH on solubility [11,13].

In order to provide a model which is capable of predicting protein solubility based on a minimal set of experimental data, the second osmotic virial coefficient (B_{22}) is used within this work as easily-accessible property for characterizing aqueous protein solutions containing a salt. B_{22} serves as an ideal measure, as it describes the complex interactions between two solute molecules (e.g. proteins) in solution, by at the same time accounting for the influence of salt, salt concentration, pH, and temperature. If B_{22} is negative, attractive interactions between the solute molecules, in our case proteins, in solution dominate, favoring crystallization or precipitation [14].

First shown in 1999 by Haas et al., the protein solubility can be modeled as function of B_{22} [15]. Haas et al. developed a solubility model based on the Gibbs energy of an aqueous protein solution in equilibrium with a crystalline protein containing a considerable amount of water derived from a simple lattice model. Using a value for B_{22} estimated from a square-well potential, the Gibbs energy of the liquid phase was calculated and protein solubility was estimated using this B_{22} [15]. Another solubility model was developed by Ruppert et al. to model the protein solubility from B_{22} [16]. In this model the fugacity of crystalline and dissolved protein was equalized and the activity coefficient of the protein was related to B_{22}. Mehta et al. compared both models and concluded that the model from Ruppert et al. provides better results than the model from Haas et al., since the first model has two fitting parameters whereas the latter one has only one fitting parameter [13]. The major drawback of the model by Ruppert et al. is the use of two adjustable parameters which both depend on solvent, type of salt and protein. A prediction, transferring these parameters to different systems, is not possible.

In general, using the solubility models from literature, protein solubility can only be calculated for those concentrations where experimentally determined B_{22} data is available. These limitations arise from the method used for fitting the model parameters where pairs of B_{22} and protein solubility are needed. As the protein solubility can well be measured at high salt concentrations, and B_{22} at low salt concentrations, these methods are limited to a narrow intersection of the two concentration ranges.

To avoid these problems and to improve the solubility model enabling a prediction of the protein solubility even for salt concentrations, pH, and temperature ranges where experimental B_{22} is unavailable, a new model has to be supplied.

In this work a new solubility model was developed based on a modified form of the solubility equation of Ruppert et al. combined with the xDLVO model of Asakura and Oosawa [17] to predict B_{22} data for different temperature, salt type, salt concentration and pH. The xDLVO model was used to model and to predict B_{22} data of lysozyme and monoclonal antibody (mAb) over a broad salt-concentration range from salt-free solution to saturated salt solutions. B_{22} data for the mAb were also measured at different pH. Using the B_{22} data from xDLVO, the protein solubility was then estimated applying the modified solubility equation of Ruppert et al.

This approach allows for predicting the protein solubility in a broad salt-concentration range and for different pH. This leads to a decrease of experimental effort and significantly reduces the time for developing protein-production processes.

2. Materials and methods

2.1. Materials

The proteins used in this study were lysozyme from chicken egg white (14.4 kDa) and a monoclonal antibody (144.2 kDa). Lysozyme from chicken egg white (CAS: 12650-88-3) was purchased from Sigma Aldrich (Steinheim, Germany). The monoclonal antibody, an IgG 1, was supplied in an aqueous solution of PBS buffer (10 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCl, 138 mM NaCl) by Bayer HealthCare (Wuppertal, Germany). Sodium chloride (NaCl, CAS: 7647-14-5), sodium p-toluenesulfonate (Na-p-Ts, CAS: 657-84-1), sodium acetate (NaAc, CAS: 7365-45-9), acetic acid (HAc, CAS: 64-19-7), ammonium sulfate ($(NH_4)_2SO_4$, CAS: 7783-20-2), potassium dihydrogen phosphate (KH_2PO_4, CAS: 7778-77-0), and sodium hydrogen phosphate (dodecahydrate) ($Na_2HPO_4 \cdot 12H_2O$, CAS: 10039-32-4), were purchased from Merck (Darmstadt, Germany). 2-Amino-2-hydroxymethyl-propane-1,3-diol (TRIS, CAS: 77-86-1), and potassium chloride (KCl, CAS: 7447-40-7) were obtained from Sigma (Steinheim, Germany) and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, CAS: 7365-45-9) from Prolabo (Leuven, Belgium). Lysozyme was used without purification. The monoclonal antibody was dialyzed three times against the desired buffer to purify it. The salts were purchased in 98% purity or higher and were used without purification. Water was filtered with a 0.22 μm Millipore filter.

2.2. Sample preparation

2.2.1. Protein and salt solutions

The lyophilized-delivered lysozyme was dissolved directly in buffer solution. Lysozyme solutions were buffered at pH 4.6 using a 0.05 M acetate buffer.

The monoclonal antibody (mAb) was delivered in aqueous solution with PBS buffer at pH of 7.4. To adjust the pH of the mAb solution, the protein solution was dialyzed (see next section).

Saturated aqueous salt solutions were prepared by dissolving the salt in the same buffer as the appropriate protein.

2.2.2. Dialysis

The monoclonal-antibody solution was dialyzed with a 10 kDa molecular weight cut-off membrane from SpectrumLabs (Breda, Netherlands). The sample was dialyzed three times against the desired buffer for 24 h. For dialysis the buffers used were 0.01 M TRIS buffer with a pH of 8.5 or a 0.01 M HEPES buffer with a pH of 7.7.

2.2.3. Solubility measurements

For measurements of the protein solubility, the protein solution was mixed in fixed ratios with the salt solution. The volume of the mixture was in the range of 200 μL–300 μL. The protein-salt solution was then stirred at constant temperature for at least four