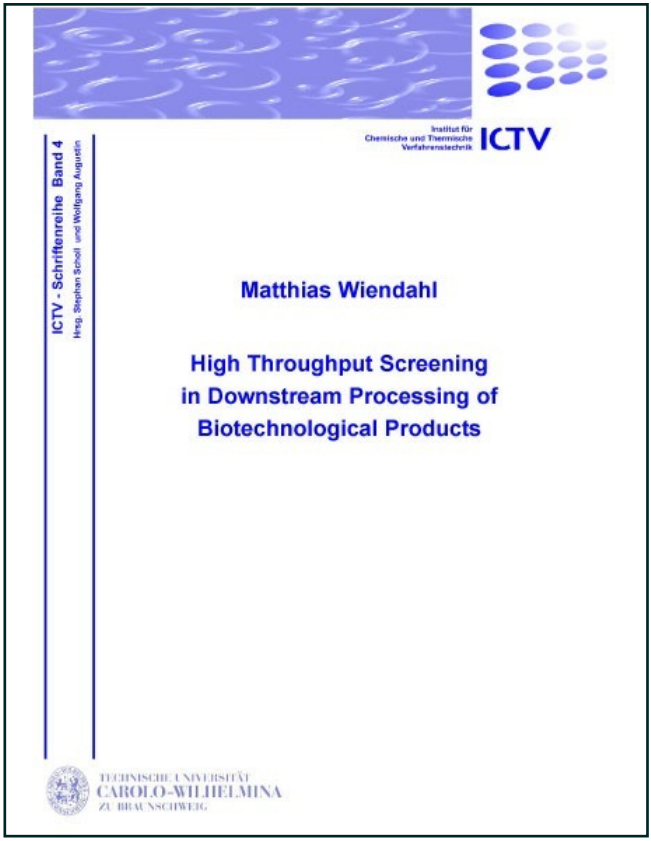
Matthias Wiendahl (Autor)

High Throughput Screening in Downstream Processing of Biotechnological Products



blockbuster drug is not released to the market, the company loses $ 3 million (Datamonitor 2001).

1.1 High throughput screening

The necessity of being ‘first on the market’ paired with the need to cut production costs highlights the demand for strategies and techniques allowing a fast process development while at the same time covering a high number of potential process parameters in order to reach an optimal overall process performance. It is obvious that early process development sets the standards for subsequent process economy and thus focus needs to be laid on this development stage (Rathore et al. 2004). The requirements to do so lie in the generation of a broad data basis describing process relevant parameters, good process understanding and the development of high performance analytics. These requirements underlie process inherent restrictions that sample material is only available in very low quantities at this early stage. The challenge lies thus in the need to create a high number of process relevant data sets with as little sample material as possible in as short a time frame as possible. A potential solution to this might lie in the development of parallel, automated and miniaturized experimental setups and strategies. A technique that combines all these three characteristics is high throughput screening (HTS) using modern liquid handling stations (LHS). HTS had its breakthrough in the late 1990s in the analytical sector and for lead discovery (Hertzberg and Pope 2000). This breakthrough only became possible due to the development of new robotic based liquid handling techniques, labware concepts such as the 96 / 384 / 1536 well formats and rapid data analysis. In its extreme case, this approach led to a daily throughput of more than 100.000 assays per day described as ultra high throughput screening 'uHTS' (Golz and Huser 2007). To date these numbers have only been reached in the analytical sector. As many different techniques have been assigned 'high throughput techniques' a definition of what the author considers to be high throughput screening and which defines the scope of the present work is given in the following:

High throughput screening is a technology that allows screening for a broad range of parameters rapidly. The three central demands automation, parallelization and miniaturization must be met. In order to be high throughput screening compatible the experiments must be able to be performed in standardized 96 / 384 / 1536 well plates; robotic liquid handling stations must be integrated and connected to a computer system.

1.2 High throughput screening in bioprocess development

In the field of bioprocess development HTS has mainly been used as a screening tool for strain development and cultivation strategies (Lye et al. 2003). These approaches cover the search for strains expressing enzymes with a higher activity or enantio-selectivity (Bornscheuer and Pohl 2001; Funke et al. 2003; Reetz et al. 1999), the optimization of bioconversion processes by screening for high product formation rates (Doig et al. 2002; John and Heinzle 2001) or substrate screening experiments for the development of fermentation processes (Duetz and Witholt 2004). Being only excerpts from a wide variety of miniaturization or automation attempts in bioprocess development these examples clearly show the high potential of HTS techniques. Despite increasing efforts in the field of downstream process development there is only very little literature data published. This lack of information results mainly from the fact that initial development of these systems were mostly company driven either in-house or with external partners (Bastek et al. 2004; Eckermann 2007; Kramarczyk 2003; Mikheeva et al. 2005). However, one of the first and well documented applications targeting the development of purification processes based on HTS screening can be found in protein crystallization. Starting as a routine technique for sample preparation prior to x-ray analysis, (Klyushnichenko 2003) used this methodology for the development of crystallization processes of therapeutic proteins reaching a production scale of 150 L.

1.3 Project objective

This work aimed to develop techniques which supply key data for downstream processes with a focus on screening of protein solubility in aqueous buffers, protein partitioning in aqueous two phase systems ('ATPS') and packed bed chromatography. The techniques were implemented using the commercially available LHS Tecan Freedom Evo® 200. After having developed the algorithms and integrated peripheral devices, the resulting process results were validated and an optimization routine was integrated for the automated development of ATPS. As the three processes are very different, experiments were treated independently and the results constitute different chapters within this thesis.

2 Theoretical background

2.1 Protein solubility

2.1.1 Solubility data in biotechnological processes

The knowledge of protein solubility under various conditions belongs to the most essential data needed for a directed process development. Process steps in the biotechnological purification sector where knowledge of protein solubility is absolutely crucial comprise controlled protein precipitation and crystallization (Cheng et al. 2006; Low et al. 2007; Thommes and Etzel 2007), chromatography (Gagnon et al. 1997; To and Lenhoff 2007; Valente et al. 2006), ATPS (Berggren et al. 2002; Hachem et al. 1996) and protein formulation (Ahrer et al. 2006; Black et al. 2007). In these cases, the dependence of protein solubility on salt concentration and type, pH and temperature or additive type and concentration must be known.

2.1.2 The solubility phase diagram

Protein solubility describes the maximum amount of protein that can be dissolved in a defined amount of solvent. Proteins leaving the soluble state can crystallize, aggregate or precipitate. Crystallization describes an ordered process where seed crystals or nucleation points lead to the formation of well-structured protein entities or crystals with a distinct shape and size distribution. Neither for precipitation nor for aggregation can a single and universal definition be found in literature. In general, precipitation and aggregation are unordered processes and sometimes the terms are used synonymously (Deyoung et al. 1993). Precipitation occurs as a result of protein-protein attractive interactions (Blanch et al. 2002). Chi et al. (2003) describe aggregation as the 'assembly from initially native, folded proteins of aggregates containing nonnative protein structures'. De Young et al. (1993) state that protein aggregation 'results from hydrophobic association of entangled denatured protein chains'. Aggregation can be reversible (Oncley et al. 1952) and aggregates can rearrange to crystals, but in crystallization, aggregation is generally unwanted (Asherie 2004). The same holds true for precipitation, which is sometimes used synonymously with 'salting out' (Salis et al. 2006; Trevino et al. 2007). Nevertheless, both processes lead to insoluble protein formations that can be measured

via turbidity (Arakawa and Timasheff 1984) and separated from the liquid phase by centrifugation (Shih et al. 1992). A common graphical representation of solubility curves and crystallization processes is to plot the protein solubility as a function of precipitant concentration (Figure 2.1.1).

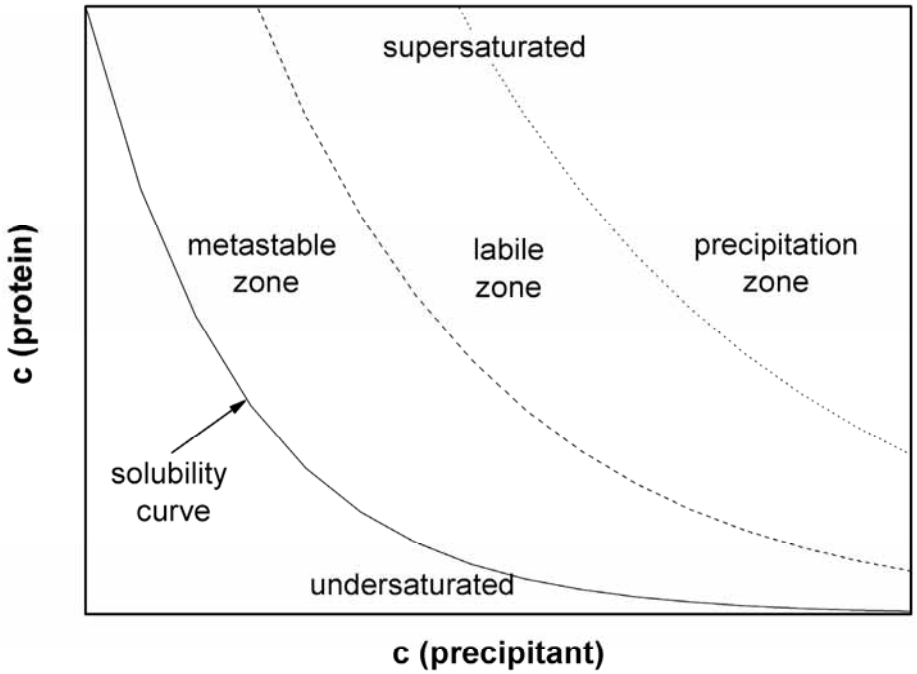


Figure 2.1.1 Schematic phase diagram. Protein solubility as a function of precipitant concentration (according to Asherie (2004))

In the undersaturated zone, crystals dissolve whereas they grow in the supersaturated zone. The precipitation zone is the zone where disordered aggregates or precipitates form instantaneously. The labile zone is the region where spontaneous crystallization without seeding with crystals occurs; and the metastable zone is the zone where no crystals will form within a reasonable amount of time. No clear distinction between the three supersaturated zones is possible because the zone limits are related to kinetic phenomena; the only line with a distinct position in the diagram is the 'thermodynamical' solubility curve which describes the thermodynamically determined transition between crystal dissolution and –formation (Asherie, 2004). In the light of the above, solubility data can only be used correctly by closely linking the interpretation to the experimental method used for its determination.

The prediction of aqueous solubility of proteins has mostly concentrated on thermodynamic models and calculations basing on the Deryaguin, Landau, Verwey and

Overbeek ('DLVO') theory. Thermodynamic models include the use of the second virial coefficient (Costenaro and Ebel 2002; Ruckenstein and Shulgin 2006) or statistical models like the UNIQUAC framework (Agena et al. 1999). In the DLVO theory, which has been derived from colloid science, the free energy between interfaces is calculated (Ninham 1999; Rowe 2001). Despite significant progress in the modeling and prediction of the solubility of small molecules in an aqueous environment, even for these structurally rather simply molecules no truly predictive and universal model is currently available (Delaney 2005). For proteins, which are structurally far more complex, few correlations between model and actually measured values have been obtained; however, models able to predict aqueous solubility from sequence and buffer information are still far out of sight. This is not only a result of the structural complexity of the protein molecules themselves but is also based on the fact that protein-protein, protein-water, protein-ion and ion-water interactions are still not fully understood.

2.1.3 Solubility parameter dependencies

Solubility is a function of interactions occurring between protein molecules, water and the salt-ions present in solution as well as surrounding system parameters (Retailleau et al. 1997). Protein characteristics with significant influence on solubility are the net charge, the ratio of charged, polar and neutral amino acids, hydrophobic and hydrophilic properties. Surrounding parameters influencing solubility are ambient temperature and to a lesser degree air pressure. Additives can affect protein solubility due to their ability to form water structures or their interaction with the protein molecule. Salts reducing protein solubility are often used as gentle precipitating agents whereas salts increasing protein solubility are often used for the unfolding of aggregated proteins, *e.g.* inclusion bodies. The influence of salts on water structure and protein solubility has thus been a major research issue in the past, however, still awaits final conclusions. Gross system parameters can be summarized in the pH value, the concentration and the salt type. The influence of the salt type on protein solubility has first been described by Hofmeister (1888) who dissolved chicken egg white in a variety of different buffers. He sorted the salts he used in a series which turned out to be universal for most proteins ('Hofmeister series', cf. Figure 2.1.2). For a long time, anions were assumed to have a stronger effect on protein solubility (Green 1931; Green 1932) but recent studies have shown a similarly strong effect of cations (Benas et al. 2002).

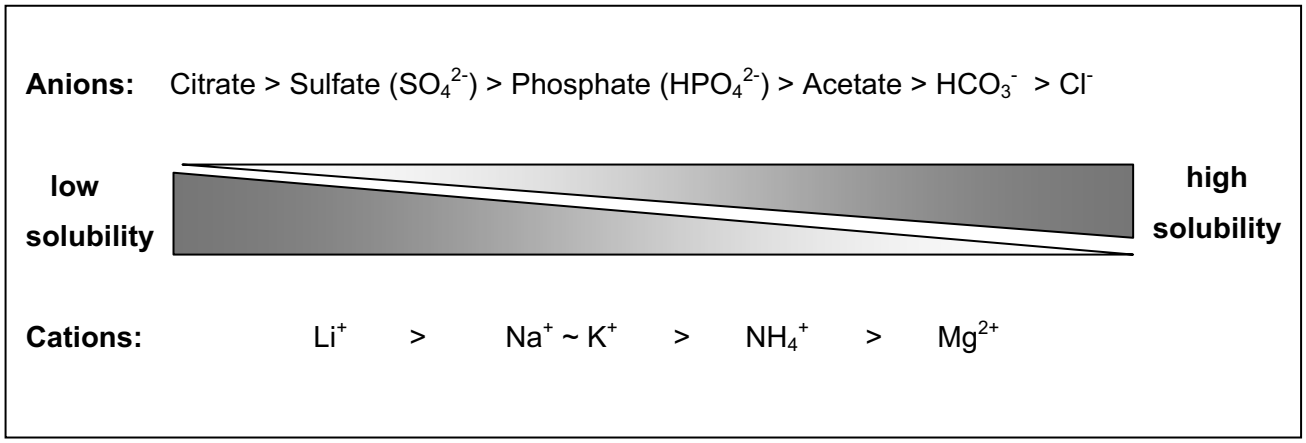


Figure 2.1.2 Hofmeister series of anions and cations (anions from Goueva et al. (2006), cations from Benas et al. (2002))

Within the last years, the Hofmeister effect has been examined more closely in an abundance of publications which has led to several 'Hofmeister series' differing mostly in the position of citrate within the series (Benas et al. 2002; Collins and Washabaugh 1985; Curtis et al. 2002; Gouvea et al. 2006; Green 1931; Green 1932; Gurau et al. 2004; Hofmeister 1888; Zhang et al. 2005). The occurrence of the effect ions have on protein solubility has been attributed to different semi empirical quantities like salting-out coefficients (Setschenow 1889), lyotropic numbers (Voet 1937) and physicochemical properties of the ions such as Gibbs free energy of hydration (Clarke and Lupfert 1999), polarizability (Conway and Ayranci 1999), surface tension increment (Melander and Horvath 1977) or entropy change of water (Krestov 1991). Until the late 80s an indirect effect of the salt was assumed to be responsible for the influence on protein solubility: Salts were grouped as water structure-making ('kosmotropic') or structure-breaking (‘chaotropic‘). A review comprising more than 900 references summing up the knowledge until the mid 1980s can be found in Collins and Washabaugh (1985). Newer results suggest that the salt does not influence the water structure but that it interacts with the protein *via* (1) hydrogen bonding of the amide group, (2) hydrophobic hydration of the molecule and (3) direct binding of the anion to the amide group (Zhang et al. 2005). Whichever effect is dominating, probably no single parameter is solely responsible for the Hofmeister effect (Kunz et al. 2004).

2.1.4 Methods for solubility determination

In the absence of potential model-based predictions on protein solubility in aqueous environments, experimental approaches are currently the method of choice when investigating protein solubility. Due to this situation, a vast amount of methods describing ways to determine protein solubility has been published. Table 2.1-1 lists several of these methods including the measuring principle and the respective reference. The methods differ significantly in the experimental effort and time as well as the state of the protein formulation needed for the experiments. In most cases where the thermodynamic solubility curve is desired, crystals of the compound or protein are required. The experimental aim is then to determine the equilibrium state between crystal formation and dissolution. This state can be derived from enthalpy, crystal size or infrared spectral change measurements. While this information is crucial for a fundamental scientific approach, several drawbacks hinder these methods to be applied in industrial process development. The amount of crystals needed for solubility screens would be too high when taking the production costs into account. The equipment used is rather sophisticated and can not or only hardly be integrated into robotic platforms. Finally, the determination of thermodynamic solubility curves generally requires days to several weeks. Taking these points into consideration it becomes clear that the determination of the thermodynamic solubility curve can currently not be performed in a way that fulfils the demands of real high throughput screening.

2.1.5 Literature review - High throughput screening for solubility determination

High throughput screening of protein solubility imposes several requirements on the experimental and analytical setup. Of equal importance is the question what data is required, *i.e.* thermodynamically relevant data or process related estimations. In order to shorten the experimental time, not only fast analytics are necessary but also a fast transition from the soluble into the solid phase. Thermodynamic equilibrium which might occur after several days is neither process relevant when trying to describe sudden precipitation or aggregation during the purification process nor is it a HTS-compatible parameter. Industrially more relevant is the transition from the ‘labile zone’ to the ‘precipitation zone’ where immediate aggregation and precipitation occur. This sudden phase transition might impose certain hazards such as filter clogging, product loss or increasing pressure drops during chromatographic processes. The most obvious method to screen for this phase transition uses

the situation when mixing of the different components immediately leads to a precipitation indicating that the fluid composition lies in the precipitation zone. If this scenario does not prevail due to a too dilute solution, water needs to be removed to concentrate the components until precipitation or aggregation occurs (see chapter 4.1).

On the analytical side, only the optical techniques scintillation, UV-measurement and nephelometry are HTS-compatible. These methods can at least in principle all be performed on robotic LHS and experimental time lies in the range of minutes to hours for one data point. Nephelometric and UV-turbidity measurements often yield comparable results admittedly at the cost of lower precision than the thermodynamic solubility curve measurements (Pan et al. 2001). One method to quickly determine solubility values is the turbidimetric titration (Lipinski et al. 1997). Bevan et al. (2000) have used a nephelometer, which measures the intensity of scattered light in a right angle to the entering light. In a further development of this method Goodwin et al. (1999) have used a flow cytometer. Pan et al. (2001) have used quartz UV-microtiter plates in combination with a UV-photometer; Chen et al. (2002) have used plastic one-way UV-plates. Short wavelengths in the range of 300 to 600 nm are preferred over longer wavelengths as for homogenous samples; a shorter wavelength yields better signals (Pan et al. 2001). The time needed for measurement only comprises few seconds for both methods and thus a high throughput analysis is possible.

Most publications dealing with solubility screening in high throughput mode so far concentrate on the field of proteomics where the major question is whether a given protein is soluble or mainly precipitated. Knaust and Nordlund (2001) have used a 96 well plate assay to distinguish between soluble and insoluble proteins after cell lysis. Similar systems have been described by Chambers et al. (2004) and Hammarström et al. (2002). Stenvall et al. (2005) have used a 96 well-plate compatible assay to screen for solubility data of more than 100 recombinant proteins. They diluted a protein solution (0.8 $mg \cdot mL^{-1}$ in 1 M urea) five-fold in water or in 1 M urea followed immediately by a BCA protein concentration determination assay. The samples were allowed to rest over night at 37 °C, followed by centrifugation at 2800 g for 50 min and determination protein concentration (BCA assay) in the supernatant. Doing so, the soluble protein fraction can be determined at two different protein/denaturant concentrations. This strategy is suitable to screen for conditions under which the protein can be obtained in a soluble way, however, for a more general approach where different buffers or additives must be screened, this assay is probably too slow. Methods providing the possibility to screen for a broad range of parameter combinations in a high throughput mode have not been published yet.

Table 2.1-1 Methods to determine protein solubility

Method	Procedure	HTS-compatible	Reference
Interferometry	• Protein crystals in solution in equilibrium. • Dissolution or growth of crystals as a function of temperature is determined by measuring the change of the refractive index in the medium.	-	(Gray et al. 2001; Sazaki et al. 1996)
Micro-dialysis cell and interferometry	• Protein crystals in solution in equilibrium. • Separated from precipitating agent via dialysis-membrane. • Dissolution or growth of crystals in equilibrium as a function of precipitant concentration are determined by measuring the change of the refractive index in the medium.	-	(Nakazato et al. 2004)
HPLC-Analysis	• Two protein (insulin) types are mixed and incubated for 30 min. Samples are centrifuged, the solved fraction is measured by HPLC. As a reference for total protein concentration hydrochloric acid was added to the solution, in order to solve potential precipitates. An HPLC analysis	-	(Arakawa et al. 1989)
SDS-PAGE	• Proteins and co-solvents are mixed. Dissolved and precipitated proteins are separated by filtration and both forms are detected by SDS-PAGE	-	(Bondos and Bicknell 2003)
Microcolumn technology	• Protein crystals are packed into two columns. Super- and undersaturated solutions are added to the columns until convergence of column efflux concentration occurs. Protein concentrations are determined *via* $UV_{280\ nm}$	-	(Cacioppo et al. 1991; Pusey and Gernert 1988)
Fourier Transformation Infrared Spectroscopy	• The temperature of super- and undersaturated protein solutions is varied over time and solubility is deferred from infrared spectra.	-	(Dunuwila and Berglund 1997)
Microcalorimetry	• Protein crystals are added to buffer solutions of different concentrations and the electric power needed to compensate for the crystallization or dissolution enthalpy is recorded.	-	(Darcy and Wiencek 1998)
Nephelometry	• Saturated drug solutions are diluted in series in PBS using microtiter plates. Laser light at 562 nm is scattered and measured at right angle.	+	(Bevan and Lloyd 2000)
Scintillation	• Protein is present in solution. By raising or lowering the temperature, nucleation and crystal growth occurs. Crystals are detected by light scattering at the photo-detector.	+-	(Arakawa et al. 1989; Rosenberger et al. 1993)
UV-vis	• Serial dilutions of desired (bio-)compounds. Discrimination between soluble and non-soluble samples via threshold value.	+	(Pan et al. 2001)