



TECHNISCHE  
UNIVERSITÄT  
WIEN  
Vienna University of Technology

## Dissertation

# Quality by Design strategies in Inclusion Body processing

ausgeführt zum Zwecke der Erlangung des akademischen Grades  
eines Doktors der technischen Wissenschaften unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.nat.techn. Oliver Spadiut

E166

Institut für Verfahrenstechnik, Umwelttechnik und technische  
Biowissenschaften

eingereicht an der Technischen Universität Wien

Fakultät für Technische Chemie

von

Julian Ebner

Mart. Nr. 1226207

Wien, am

Unterschrift

I hereby declare that I am the sole author of this work. No assistance other than which is permitted has been used. Ideas and quotes taken directly or indirectly from other sources are identified as such. This written work has not yet been submitted in any part.

Vienna, April 2025

Julian Ebner



# Acknowledgement

First and foremost, I want to thank Oliver, for making this thesis possible. Obviously in all the essential parts, but beyond that by always being open to new ideas and projects. I immensely appreciate how he keeps an open mind in the face of the day-to-day, and how he does it with maximum support and motivation.

Furthermore, I want to thank all my Post-Docs who supervised different parts of my work over the years, especially Julian Kopp, Christoph Slouka and Diana Humer.

Stefan, who was not only a great Lab-Partner but also a great Squash-Partner.

Andreas and Karim, who showed how productive the combination of different expertise can be and who were always motivated to do one more experiment.

I want to thank all the people who worked with me on various IB projects, especially Viktor and Robert.

I want to thank EnginZyme for giving me an amazing opportunity to see research outside of academia. It was a privilege to meet so many smart and motivated people and I am still impressed how much one can learn in six months in the right environment.

Last but not least, I obviously want to thank my family and friends for their immense support in any circumstance.

## Abstract

Recombinant protein production in *E. coli* poses several significant advantages compared to other production hosts, such as short process times and high cost-efficiency. However, due to high expression levels and the lack of post-translational modification machineries, proteins of interest are often produced as insoluble aggregates, more commonly called Inclusion Bodies (IBs). Therefore, several additional unit operations are required during down-stream processing (DSP), namely IB washing, solubilization and refolding. Up to now, these unit operations are developed and optimized using empiric approaches since no universal refolding protocol is available. In recent years, several paradigm changes for production processes of pharmaceuticals took place, changing the focus to Quality by Design (QbD) and Process analytical technology (PAT) (e.g. ICH Q8-12). These strategies require scientific knowledge and process understanding to be generated during process development as well as suitable analytical methods in order to monitor and control the process. While readily implemented for small molecules, strategies are still lacking for biopharmaceuticals and, in particular, for those produced as IBs. The reason for this is twofold: (i) The production of biopharmaceuticals in living organisms presents highly complex systems. Therefore, the product requires analytical methods able to assess very specific properties of one target molecule in a highly complex matrix. (ii) Many of the analytical techniques able to fulfill the requirements listed in (i) are not applicable for the refolding process due to harsh conditions during solubilization and changing Critical Quality Attributes (CQAs) during the DSP. In this work, the development of IB processes for the enzyme horseradish peroxidase (HRP) as model protein was investigated. Screening of suitable refolding conditions was performed as Design of Experiment (DoE) in small-scale approaches. These conditions were then scaled up to the controlled environment of a bench-scale refolding vessel. Different analytical strategies were applied for the entire DSP, including high performance liquid chromatography, Quantum cascade laser-IR-spectroscopy, and enzymatic activity assays. Several of the shown methods were then employed for IB forming proteins from other, different protein families. Overall, the applicability of QbD approaches to specific parts of IB processes were demonstrated. This cumulative thesis contains four first author papers, one patent, and two book chapters. While the definition of CQAs is still an integral part of these strategies, the high structural complexity and conformation-change of IBs through various unit operations makes the identification of CQAs for intermediates challenging. Furthermore, identifying the material attributes and process parameters that can have an effect on product CQAs as well as establishing PAT tools requires significant process understanding. This can, at least partly, be generated through unit-operation spanning multivariate approaches and an iterative approach to process development requiring a “Design of Design of Experiments” concept.

# Zusammenfassung

Die Produktion von rekombinanten Proteinen in *E. coli* bietet im Vergleich zu anderen Produktionsorganismen mehrere bedeutende Vorteile, wie z. B. kurze Prozesszeiten und effiziente Kosten. Aufgrund des hohen Expressionsniveaus und des Fehlens von posttranslationalen Modifikation werden die gewünschten Proteine jedoch häufig als unlösliche Aggregate, so genannte Einschlusskörperchen (aus dem Englischen Inclusion Bodies (IBs)), produziert. Daher sind während des Down-Stream-Processing (DSP) mehrere zusätzliche Arbeitsschritte erforderlich, nämlich Waschen, Solubilisierung und Rückfaltung der IBs. Bislang wurden diese Arbeitsschritte mit empirischen Ansätzen entwickelt und optimiert, da kein universelles Rückfaltungsprotokoll zur Verfügung steht. In den letzten Jahren haben mehrere Paradigmenwechsel bei den Produktionsprozessen von Arzneimitteln stattgefunden, die den Schwerpunkt auf Quality by Design (QbD) und prozessanalytische Technologie (PAT) gelegt haben (z. B. ICH Q8-12). Diese Strategien erfordern wissenschaftliche Kenntnisse und ein Prozessverständnis, das während der Prozessentwicklung geschaffen werden muss, sowie geeignete analytische Methoden zur Überwachung und Kontrolle des Prozesses. Während diese Strategien für kleine Moleküle oft gut umsetzbar sind, fehlen sie für Biopharmazeutika und insbesondere für als IBs hergestellte Biopharmazeutika noch. Der Grund dafür ist unserer Meinung nach ein doppelter: (i) Die Herstellung von Biopharmazeutika in lebenden Organismen stellt hochkomplexe Systeme dar. Daher erfordert das Produkt Analysemethoden, die in der Lage sind, sehr spezifische Eigenschaften eines Zielmoleküls in einer hochkomplexen Matrix zu bewerten. (ii) Viele der analytischen Verfahren, die die unter (i) genannten Anforderungen erfüllen können, sind aufgrund der rauen Bedingungen während der Solubilisierung und der sich ändernden kritischen Qualitätsmerkmale (CQAs) während des DSP nicht für den Rückfaltungsprozess geeignet. In dieser Arbeit wurde die Entwicklung von IB-Prozessen für das Enzym Meerrettich Peroxidase (aus dem Englischen Horseradish Peroxidase (HRP)) als Modellprotein untersucht. Das Screening geeigneter Rückfaltungsbedingungen wurde als Design of Experiment (DoE) in kleinem Maßstab durchgeführt. Diese Bedingungen wurden dann auf die kontrollierte Umgebung eines Rückfaltungsgefäßes im Labormaßstab hochskaliert. Für das gesamte DSP wurden verschiedene Analysestrategien angewandt, darunter Hochleistungsflüssigkeitschromatographie, Quantumkaskadenlaser-IR-Spektroskopie und enzymatische Aktivitätsuntersuchungen. Mehrere der gezeigten Methoden wurden dann für andere IB-bildende Proteine aus verschiedenen Proteinfamilien eingesetzt. Insgesamt wurde die Anwendbarkeit von QbD-Ansätzen auf spezifische Teile von IB-Prozessen demonstriert. Diese kumulative Dissertation enthält vier wissenschaftliche Publikationen, ein Patent und zwei Buchkapitel. Während die Definition von CQAs immer noch ein integraler Bestandteil dieser Strategien ist, macht die hohe strukturelle Komplexität und Konformationsänderung von IBs durch verschiedene Einheitsoperationen die Identifizierung von CQAs für Zwischenprodukte zu einer Herausforderung. Darüber hinaus erfordert die Ermittlung von Materialeigenschaften und Prozessparametern, die sich auf die CQAs von Produkten auswirken können, sowie die Entwicklung von PAT-Tools ein umfassendes Prozessverständnis. Dieses kann, zumindest teilweise, durch einsatzübergreifende multivariate Ansätze und einen iterativen Ansatz für die Prozessentwicklung generiert werden, der ein „Design of Design of Experiments“-Konzept erfordert.



## Table of Content

Acknowledgement.....	3
Abstract .....	4
Zusammenfassung.....	5
1. Introduction .....	7
1.1 Recombinant protein production and choice of host.....	7
1.2 Protein folding processes – energetic basics and fundamentals <i>in vivo</i> and <i>in vitro</i> .....	8
1.3 IB formation and properties .....	11
1.4 State of the art in IB processing - Restrictions compared to the theory .....	13
1.5 Quality by Design.....	15
1.6 HRP – a challenging model protein for IB research.....	17
2 Problem formulation and goal of the thesis.....	21
2.1 Problem Formulation.....	21
2.2 Goal of the thesis.....	22
3 Results and Discussion.....	24
3.1 Scientific Question 1 .....	25
3.2 Scientific Question 2 .....	32
4 Conclusion and Outlook.....	42
4.1 Outlook HRP process .....	44
4.2 Outlook platforms.....	45
4.3 Outlook IBs in general .....	48
5 References .....	50
6 Appendix I: Scientific Publications.....	60
6.1 Book Chapter 1.....	60
6.2 Book Chapter 2.....	74
6.3 Patent.....	86
6.4 Scientific paper 1.....	87
6.5 Scientific paper 2.....	107
6.6 Scientific paper 3.....	116
6.7 Scientific paper 4.....	129
6.8 Scientific paper 5.....	137
7 Appendix II.....	146
7.1 Book Chapter: The Purification of Heme Peroxidases from <i>E. coli</i> Inclusion Bodies.....	146
7.2 Scientific paper: Recombinant Protein L .....	157
7.3 BioSpektrum Article: Inclusion Body Processing.....	165
7.4 CV .....	168

# Quality by Design strategies in Inclusion Body processing

## 1. Introduction

### 1.1 Recombinant protein production and choice of host

The development of artificial DNA synthesis in 1973 marked the first cornerstone for recombinant protein production and was followed shortly after by the first functional recombinant protein produced in 1977 [1-3]. The Nobel Prize in Chemistry awarded to Paul Berg for his work on recombinant DNA in 1980 and the approval of recombinant human insulin in 1982 by the FDA shows the rapid development and adaptation of this technology [2]. Today, more organisms, strains, and methods are available for recombinant protein production than ever before.

Due to the large size and high complexity of proteins, chemical synthesis does not provide a viable option for their production. Therefore, organisms, cells, or their molecular machinery are still used for recombinant protein production today. On a basic level, information of the protein sequence is stored in DNA which is transcribed into mRNA and then translated into the amino acid chain, i.e. the protein. Therefore, if recombinant DNA is introduced into a host organism, this organism has the information available to produce the desired recombinant protein, the protein of interest (POI). Within the 48 years since the first recombinant protein was produced, a wide variety of different host cell organisms were established. On a very fundamental level, these can be split into Prokaryotes and Eukaryotes, which show significant differences in their transcription and translation machinery [4]. The ease of introduction and stability of the introduced recombinant DNA varies between host organisms and depends mainly on two factors: (i) Well established, so called model organisms (e.g. *E. coli*, *S. cerevisiae*, *Aspergillus nidulans*), often have an easily accessible, wide variety of different genetic tools available [5]. This can include e.g. selection markers, knowledge about gene regulation sequences such as promoters or established methods for transfection. (ii) In addition, plasmids are often available for prokaryotes and lower eukaryotes, allowing for easy and fast introduction of recombinant DNA. For higher eukaryotes, DNA has to be integrated into the host's genome, requiring more complex approaches but potentially providing higher stability [6-9].

In addition, the post-translational machinery responsible for further modifications of the proteins after translation varies greatly between different host organisms. As a rough approximation, higher (eukaryotic) host cell organisms are able to produce more complex proteins, but require more challenging and complex conditions to do so. On the other hand, prokaryotic host organisms, i.e. bacteria, generally provide fast growth, simple fermentation conditions and therefore result in cheaper processes. However, they lack the ability to form many of the PTMs present in eukaryotes, potentially preventing production of more complex POIs [10]. In summary, several aspects are usually taken into account when deciding on a host organism:

- **The origin of the POI** – Since proteins are usually adapted to the environment of their original host, they can often be produced in similar host organisms without many issues (e.g. bacterial proteins in bacteria etc.).
- **Required post-translational modifications** – If certain PTMs such as glycosylation are required, usually host cells which are able to provide the required PTM machinery are employed. Alternatively, the PTMs can be introduced or further processed later in the production process. This usually happens *in vitro* and requires additional processing steps



during the downstream process (DSP). An example would be the insulin process, where disulfide bridges are formed during refolding through a cysteine / cystine (CS/CSSC) redox system [11].

- **The intended application of the recombinant protein** – Based on the intended use, the quality and quantity of the protein vary immensely. Proteins produced for research purposes (often in academia) might only be needed in small amounts but in very high purities (e.g. for structural studies through XRD), tolerating economically unfeasible processes. Industrially produced proteins show a huge range of applications and required properties. These range from bulk enzymes produced in metric ton scale to biopharmaceutical drug products with strict quality and purity profiles produced on a gram scale.

For industrial settings, the most cost-effective process delivering the desired product would be chosen. However, it should be taken into account that total costs can differ from strictly considering production costs. As an example for pharmaceutical processes, the total costs of a slightly more efficient process with a novel host organism might be higher than those of an established one, because although process costs of the established one are higher, additional costs for validation arise for the novel one. The same can be true if, e.g., new equipment is necessary in order to efficiently process a new host organism, even if productivity employing the new equipment would outperform the established host organism.

Therefore, the choice of the host organism can be highly complex and depend on many different factors. As mentioned above, this can result in processes employing bacterial hosts as cheap expression platforms, accepting that required PTMs have to be processed *in vitro* in the DSP. However, in this case, another factor plays a pivotal role, namely that proteins are folded in a complex and highly specific three-dimensional structure that is pivotal for their function. The lack of PTMs often leads to misfolding and aggregation of the POI in the expression host cell, forming so-called Inclusion Bodies (IBs) [12, 13]. In that case, it is not only required to introduce the desired PTMs in the DSP, but also to bring the IB into solution and restore the correct protein conformation – an additional process step referred to as refolding. Figure 1 shows an overview of an exemplary IB process, including the additional process steps IB isolation, solubilization and refolding.

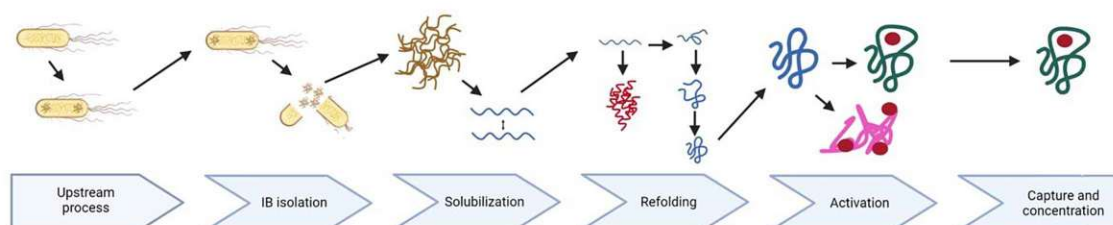


Figure 1: Example for an IB process starting with the formation of intracellular IBs during fermentation. After harvesting, the cells are disrupted and the insoluble IBs isolated. In the solubilization step, the aggregated protein is brought into solution and subsequently refolded. In case coenzymes are required, an activation step is performed after correctly folded protein is obtained. As a last step of the refolding process, capture and concentration of the correctly folded protein can be performed.

## 1.2 Protein folding processes – energetic basics and fundamentals *in vivo* and *in vitro*

The thought experiment called Levinthal's paradox played a key role to establish the fundamental theories for protein folding [14]. Based on the large number of degrees of freedom of an unfolded protein, the theoretical number of conformational combinations is exceedingly large. As an example, Levinthal assumed  $10^{300}$  for a theoretical protein containing 150 amino acids [14]. Even if these conformations would be cycled through in extremely short time periods (e.g. nanoseconds), the time required to complete the folding process would most likely exceed the existence of the known universe. However, small proteins in particular often fold spontaneously within seconds. This was the first step leading to the idea of funnel-like energy landscapes the first link in a long chain of discoveries and models trying to explain protein folding [15]. While a lot is still left unexplained, the basics are covered and well described nowadays.

The three-dimensional structure of a protein, also called conformation, is the most important part for the function of the protein. Protein functions range from providing structural integrity (e.g. tubulin) over specific detection of other (bio)-molecules (e.g. antibodies) to transport functions (e.g. hemoglobin) and allowing the light-induced electron transfer across lipid membranes (e.g. photosystem I and II) [16, 17]. Therefore, it is not surprising that an equally broad range of different protein conformations exist. The exact sequence of amino acids forming a specific protein is commonly referred to as the primary structure. While it is an essential part for determining a protein's specific conformation, the primary structure in itself does not provide information on the three-dimensional structure of the protein. Only in combination with the environment of the protein and potential modifications (e.g. PTMs) can the secondary and tertiary structure and therefore the conformation be established. Often, the "base conditions" of cellular life, i.e. an aqueous environment with moderate concentrations of dissolved matter and a temperature of 37 °C are assumed [18, 19]. However, it should be obvious that if the environment changes, the conformation of the protein would change as well. An example would be a plethora of membrane proteins which are embedded in the phospholipid bilayer of bacterial cells and keep their conformation and function in this aliphatic environment, but not in an aquatic one.

As for any (bio)chemical molecule, the behavior of proteins and therefore also their folding and conformation is dictated by thermodynamic principles. These tell us that the molecule will always strive toward a state with the minimum Gibbs free energy. As shown in Figure 2, this is also the main reason for protein folding. As soon as the protein is introduced in a changed chemical or physical environment, its conformation adapts to fulfill the requirement of the lowest energy state [20-23].

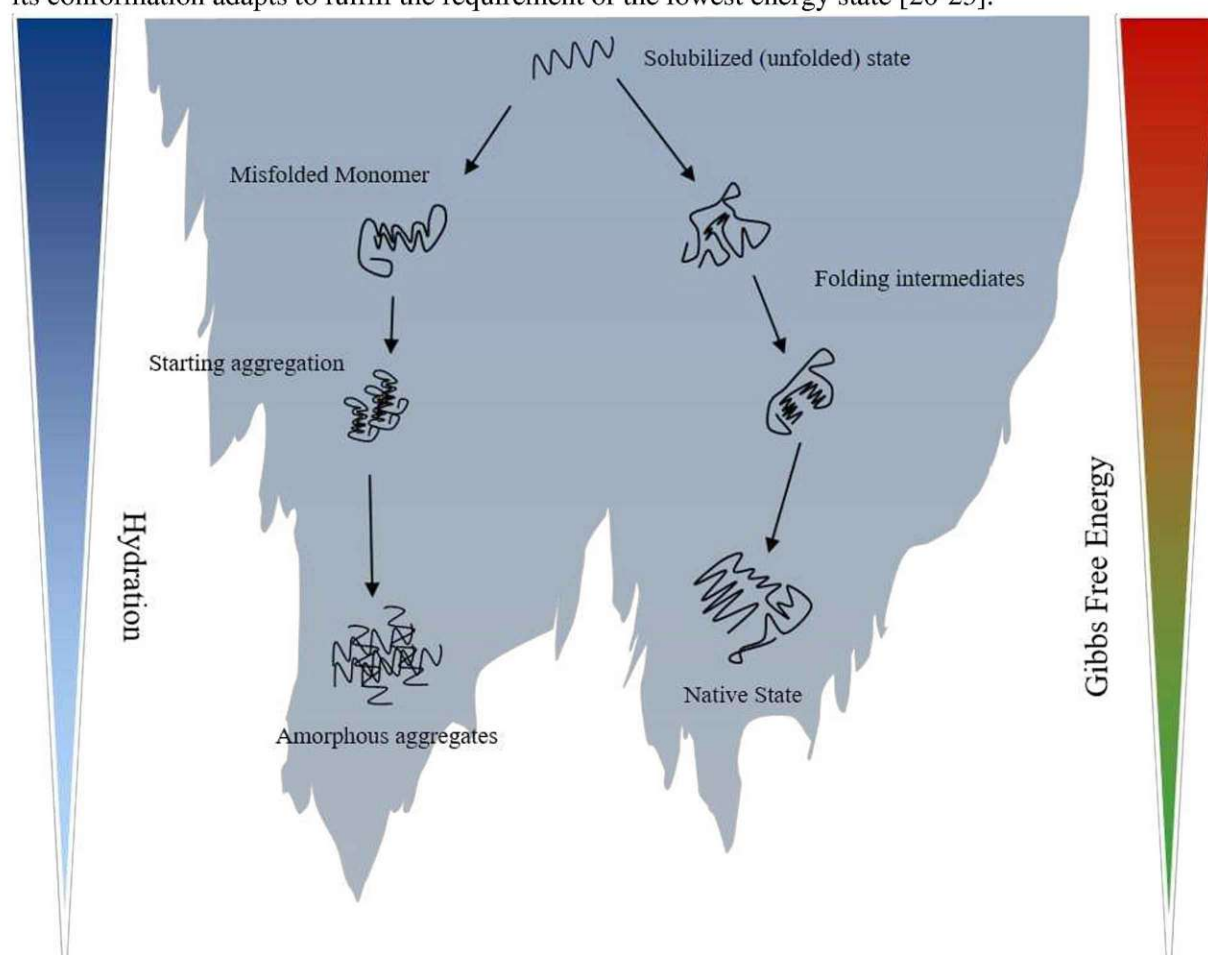


Figure 2: Graphic depiction of an exemplary free-energy landscape of protein folding. In the case of in-vitro refolding, the unfolded protein (top) is highly hydrated and is exposed to a high conformational entropy. Several conformational change pathways reduce allow for the reduction of this conformational entropy, leading either to folding intermediates and the native state or misfolded monomer, resulting in protein aggregates containing several misfolded protein units. Modified from [24].



As one of the most well-known experiments to validate this, Anfinsen showed that bovine pancreatic ribonuclease would completely denature, i.e. yield randomly coiled single polypeptide chains, if treated with a reducing agent (mercaptoethanol) in a high chaotropic (8 M urea) solution. If condition were then changed back to optimal protein concentration and pH, the protein would revert to its “functional” conformation, i.e. was able to perform the observed catalytic activity again [25]. This showed that the protein unfolding in this case was reversible, and that the native structure was contained in the primary structure of the protein itself.

However, proteins can also be irreversibly denatured. If, e.g. temperature is raised to high, the protein would be exhibiting an unordered linear structure and would not revert to its functional state after lowering the temperature, but form unsolvable aggregates (Figure 2). Therefore, the “correct” conformation of a protein presents a local minimum of energy, providing a certain stability to changes. However, if enough force in the way of energy changes of the environment is applied, the molecule can be trapped in another, more stable minimum. In the case of proteins, such a minimum is very often the state of aggregation and results in subsequent precipitation (assuming an aqueous environment) [26, 27]. The described fundamental principles are true for any protein, regardless of whether the protein is observed *in vivo* or *in vitro*. For the purpose of describing a folding system, *in-vitro* environments are often used since they are less complex and by far more controllable. However, some significant differences and implications on protein folding exist between *in-vivo* and *in-vitro* protein folding from a practical biotechnological point of view as described in detail by Hingorani and Gierasch [21]. A summary of the different factors influencing *in-vivo* folding and *in-vitro* folding is shown in Figure 3.

From an evolutionary point of view, misfolding of proteins is a highly disadvantageous event for the cell. Not only is energy wasted on the production of a protein without function, but the potential toxicity of misfolded or aggregated states poses an additional threat to the cell. Therefore, several mechanisms have evolved *in-vivo* to protect against misfolding and misfolded proteins [28]. First of all, folding can already begin during translation, a process known as co-translational folding, meaning the protein starts adopting its structure before the amino acid chain is fully synthesized. Therefore, in the ideal case, the protein spends little to no time in an unfolded state. Folding is further assisted by chaperons, macromolecules that are present *in vivo* and are believed to have evolved in tandem with hard to fold proteins. Furthermore, folding reactions are organized spatially in the cell, allowing for an environment that is advantageous for the folding of some proteins but not others. Such compartmentalization can go hand in hand with PTMs, which can be essential for the correct folding of proteins and take place in cellular compartments designed for that purpose (e.g. the endoplasmic reticulum (ER)) [29].

However, despite these advanced folding mechanisms, the folding in a cellular system might not work as desired in some circumstances, leading to aggregation of the expressed protein. This can for example occur when the POI is produced based on recombinant DNA and the chosen host cell organism is not adapted to provide an adequate folding environment. For small amounts of misfolded proteins, cells usually exhibit pathways to degrade these proteins. However, if the concentration of unfolded POI rises too fast (because of overexpression or an engineered lack of proteases), these aggregates accumulate and form so called IBs [30, 31].



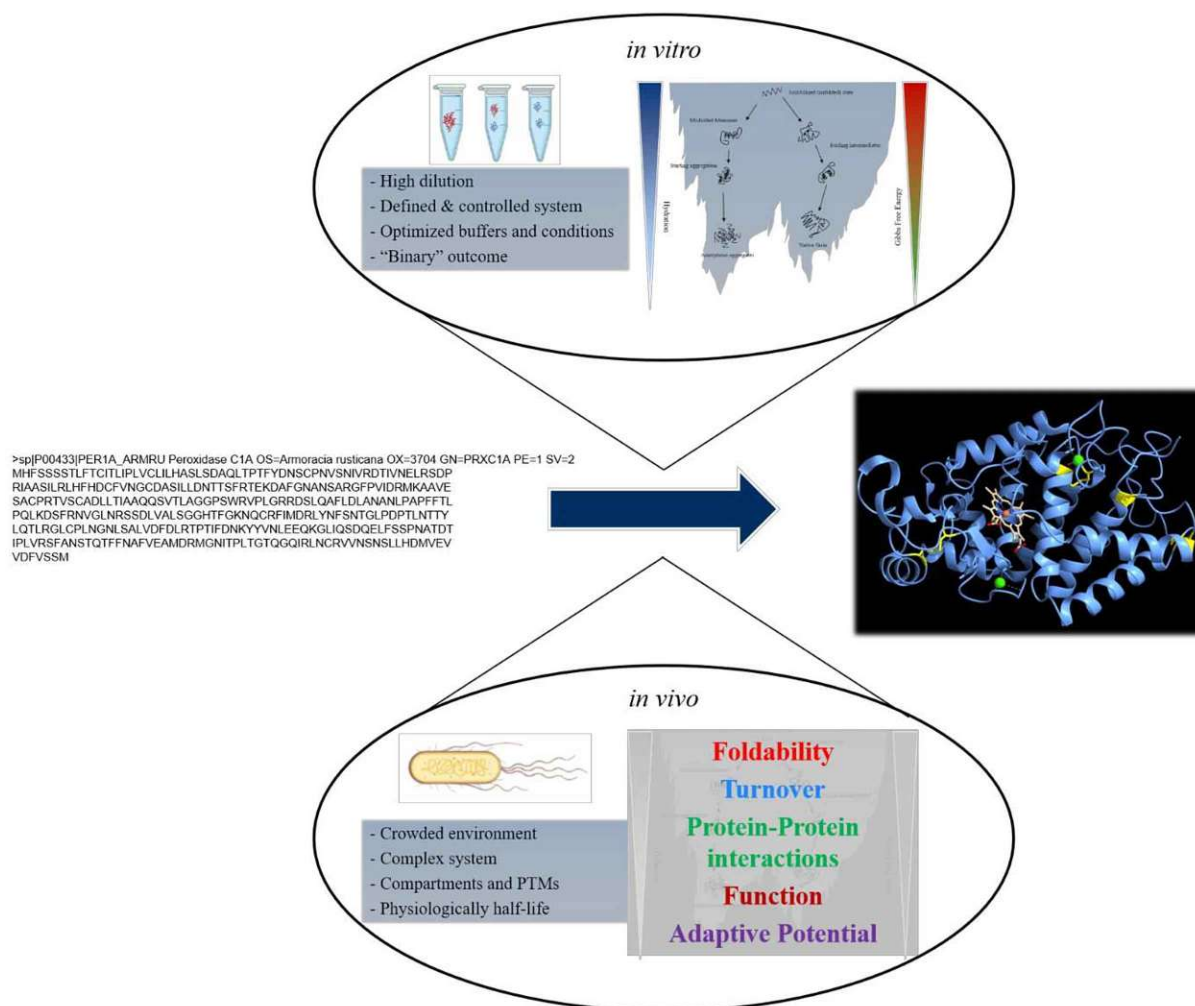


Figure 3: Graphical overview of differences between in-vivo and in-vitro protein (re)folding. For in-vitro folding (in the top part of the figure), the folding problem can be represented by an energy landscape based on the chemical and physical environment. Such a representation commonly assumes a controlled environment with low protein concentrations (high dilutions) and a binary outcome of the folding event. In the bottom part, influence factors for in-vivo folding are shown. Due to the immense complexity of present (macro)molecules and interactions, protein folding is commonly investigated based on abstracted systems, explaining one aspect or possible pathway. Middel Panel: Structure from <https://www.ncbi.nlm.nih.gov/Structure/pdb/1H58>; Image: <https://www.rbvi.ucsf.edu/chimerax> [32]. Whole Figure inspired from [21].

### 1.3 IB formation and properties

The name inclusion body stems from the fact that misfolded protein aggregates can present a significant percentage of protein in a cell, especially for systems tuned for high protein expression or overexpression. These deposits of protein form large particles of aggregated protein (yields of over 50% of the total cellular proteins have been reported [33]), which often accumulate at the polar regions of the cell and are even visible via optical microscopy [34]. This type of protein aggregation is reported for different cells and organisms, and can e.g. occur at the site of viral multiplication in eukaryotic cells. One of the most common examples is in bacterial hosts in combination with strong induction systems for recombinant protein production, e.g. high plasmid copy number IPTG induction in *E. coli* [35]. An example of IBs in *E. coli* which was recorded using scanning electron microscopy is shown in Figure 4. Two main factors influence the formation of IBs in a cell:

- I. **Overexpression** – An essential property of an IB is its macroscopic nature, i.e. it exhibits a higher density than other cellular components and is not soluble in the aqueous environment of

the cell. At least some amount of overexpression is required in order to achieve such an amount of protein and the corresponding required high intracellular concentration. In addition, the cellular system in place to deal with protein folding and misfolded protein can be overwhelmed for high expression rates. This can lead to increased protein aggregation and the formation of IBs [36].

- II. **Lack of PTMs** – While some folding issues can exclusively be caused by overexpression, the lack of a suitable environment for recombinant protein production can cause protein aggregation independent of overexpression. As one relevant example, the formation of disulfide bridges in eukaryotic cells is enabled by an oxidative environment in the ER and often assisted by ER oxidoreductases and chaperones [37-39]. For prokaryotic cells, the reducing environment of the cytoplasm prevents formation of disulfide bonds and thereby makes correct folding of the POI impossible, leading to protein aggregation [40, 41].

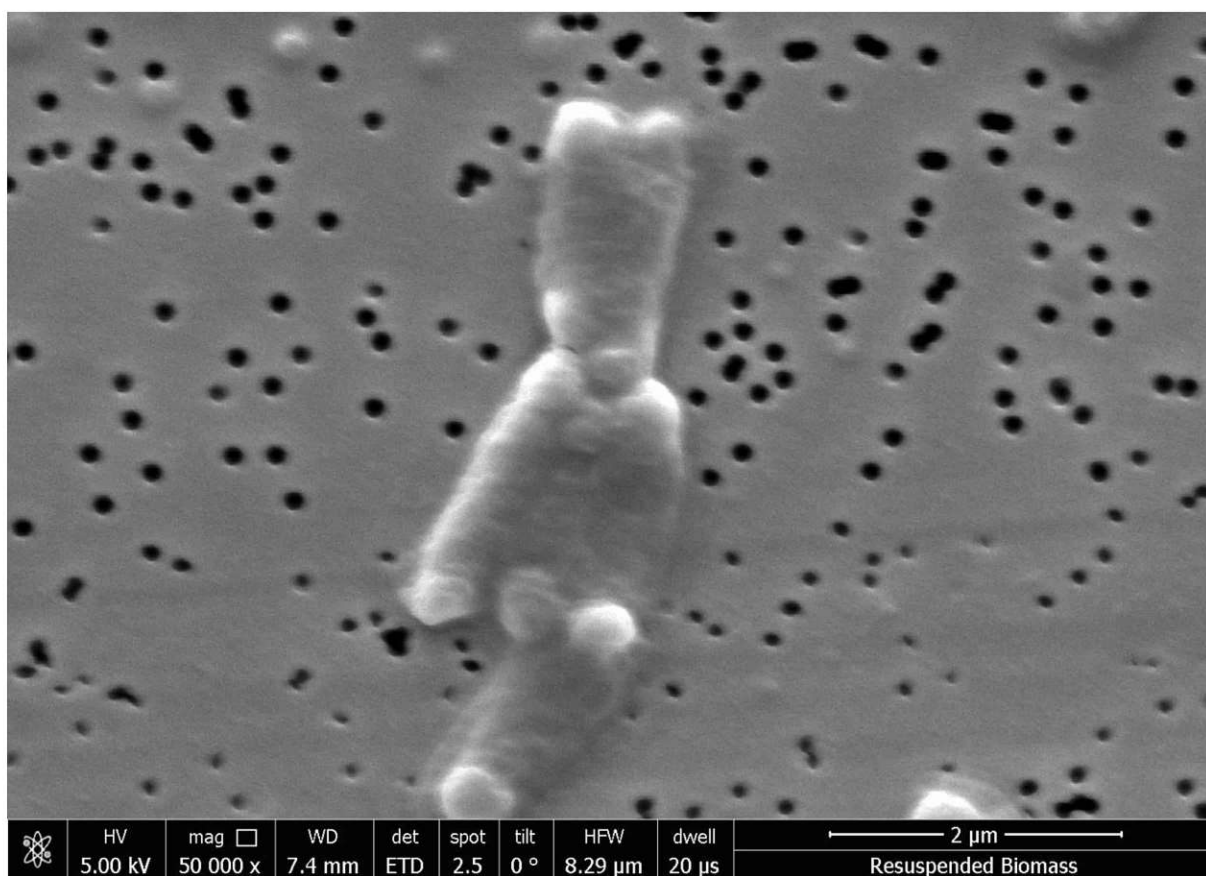


Figure 4: *E. coli* cells imaged using a scanning electron microscope. At the polar regions of the cells, IBs can be seen. Image kindly provided by Robert Klausser.

While IBs were long believed to contain completely unfolded and aggregated proteins, research in recent years has shown that some folded structures and even active protein might be present in some IBs. However, independent of the exact state of the single protein units in the IB, some properties are observed for all types of IBs. They show a higher density than the rest of the cellular environment and are not soluble in the aqueous environment [42]. Therefore, IBs are easily separable from the soluble parts of the cell after cell lysis, since they can be suspended but not dissolved in aqueous solutions. They mainly contain the POI (90-95% has been reported) and low concentrations of impurities such as host cell proteins, DNA, RNA, and parts of the ribosomal machinery [13, 43, 44]. In the case of recombinant protein production, these properties present some significant advantages over the production of soluble proteins:



- I. **Toxic POIs** – Due to the mostly inactive nature of IBs, POIs that would be toxic for the host cell if produced in soluble form can be produced as IBs.
- II. **Unstable POIs** - Proteins that exhibit short half-live times (i.e. the time it takes to half its concentration in a biological system, e.g. due to autolytic inactivation [45]) or other traits of high instability can potentially be produced as IBs. Due to the aggregated and inactive conformation formed in the IB, degradation mechanisms that require a specific protein conformation are prevented [46]. In addition, the dense structure of the IB reduces the accessibility of protein, in turn reducing degrading surrounding influences.
- III. **Protease interaction** - Native host cell proteases can pose a significant problem due to the degradation of the POI both in the cell but especially after cell lysis. While some strains provide reduced protease expression (e.g. *E. coli* BL21), this problem can be circumvented by the production of IBs, which are separated from the soluble fraction of HCPs in their aggregated form, preventing the interaction with proteases.
- IV. **Active IBs** - As mentioned above, some IBs can contain active protein, therefore exhibiting some catalytic activity. This activity is increased by the porous nature of IBs, providing sufficient surface area to allow relatively high catalytic activity [47]. Therefore, these IBs can potentially be used as immobilized proteins and provide a more economical alternative to the production and chemical immobilization of soluble protein for the use in industrial processes [48].
- V. **High POI yield and purity** – One of the biggest advantages of the production of IBs is the high protein concentration that can be achieved in combination with cost-effective production conditions during the upstream processing of bacterial hosts, e.g. *E. coli* [49-52]. Due to the easy separation of insoluble IBs from the soluble fraction (containing most HCPs, DNA and RNA etc.) after cell lysis, high initial purity can be achieved as well.

Besides the listed advantages, however, one major disadvantage is connected with the recombinant protein production via the IB route: The POI is primarily obtained in an inactive conformation and additional steps during the DSP are required to obtain the desired protein in its functional structure and conformation.

#### 1.4 State of the art in IB processing - Restrictions compared to the theory

In order to obtain correctly folded and active protein from IBs, several additional process steps are required during DSP compared to the production of the POI in its soluble form. A typical production process for classical IBs produced in a bacterial expression system is shown in Figure 5.

- I. **Cell lysis** - As for any intracellularly produced protein, the first Unit Operation (UO) after the harvesting of the biomass is cell lysis. In order to release the POI, the bacterial cell walls are disrupted. Several methods are commonly used for this process, including enzymatic lysis (e.g. using lysozyme) or mechanical methods such as sonication or high pressure homogenization (HPH). To a lesser extent, freeze-thaw cycles and grinding can also be applied. For large-scale processes, the main method applied is HPH since due to its relatively low costs and easy application and scalability [53, 54].
- II. **IB isolation and wash** - This step usually directly follows or is combined with the cell lysis step. The isolation of the IB from the soluble fraction takes advantage of the density difference and insoluble nature of the IBs in the cell lysis buffer. Therefore, separation via centrifugation or filtration is most commonly applied. Both methods show the potential to be combined with a wash step, often using a buffer with a low concentration of a chaotropic agent (e.g. 1 M urea) and/or a low to medium salt concentration (e.g. 1 M NaCl). This buffer is selected for its ability to solubilize the majority of the impurities while not dissolving the IBs. The washed IBs are often frozen until further processing, making this a standard hold step.

- III. **Solubilization** – The aim of this step is to adapt the (liquid) environment of the IBs in a way that is energetically favorable for the POI to transition from its aggregated to its soluble form. For a classic IB process, the washed IBs are suspended in a buffer with high chaotropic agent concentration (e.g. 8 M urea or 6 M Guanidinium chloride (GndHCl)) [36, 55-57]. These agents disrupt the coordinated hydrogen bonding between water molecules, weakening the hydrophobic effect, and resulting in the solubilization of the POI from the IBs. If disulfide bridges are present in the POI, a reducing agent is added at this stage in order to fully reduce the protein. While high concentrations of chaotropic agents are still considered state of the art for the solubilization step in classic IB processes, the overall goal of this UO has evolved over time. Previously, it was thought that IBs only contained completely misfolded protein, requiring full denaturation of the POI in order to enable refolding. However, since then it was discovered that the POI can show correct formation of some secondary structures or even some correctly folded protein units. Therefore, the focus shifted on finding solubilization conditions that are harsh enough to bring the POI into solution but mild enough to preserve existing correctly folded structures. This can be achieved by so-called mild solubilization approaches, often facilitating reduced chaotropic agent concentrations, alkaloid pH-values, organic solvents, high pressure, and temperature or ionic liquids [58]. In the case of active IBs, correctly folded POI can be obtained in this process, making the refolding step obsolete. However, in the case of classic IBs, the obtained solubilized and at least partly denatured protein still requires another step in order to obtain its desired conformation.
- IV. **Refolding** – During this key UO, the POI is refolded to obtain its correct and intended conformation again. This is achieved by changing the (chemical) environment in order to favor the folded state as the most stable for the single protein units. This reaction usually exhibits a competing reaction that leads to a pathway of misfolding and subsequent aggregation, which is minimized through adaption of the conditions during the refolding step. The first method to change the environment is the so-called batch dilution refolding. The high chaotropic concentration used during the solubilization step is changed by diluting the solubilization mix (containing the denatured protein) in a refolding buffer, which presents conditions favorable for protein folding. Since the refolding is a first order reaction while aggregation is a second or higher order reaction, the ratio of correctly folded protein to aggregated protein can be shifted by low protein concentrations during the refolding process [20]. Therefore, the batch dilution refolding method often requires large volumes of refolding buffer and refolding tanks. Over the years, several approaches have been described to improve or replace the batch dilution refolding method in order to reduce process volumes and the corresponding process costs. In the most simple approach, the refolding buffer and conditions are adapted in order to provide conditions that favor the refolding pathway, such as e.g. low temperature (4 °C) or the addition of sugars mimicking chaperons in order to assist protein folding. In a different approach, so-called pulsed fed-batch or fed-batch refolding can be applied. In this case, the protein solubilize is diluted into the refolding buffer in multiple steps or a constant feed, keeping the concentration of unfolded protein prone to aggregation low. Alternatively, several different chromatographic methods can be applied for the refolding process, so called on-column refolding. If the POI has a tag that allows specific binding to an affinity chromatography, the protein can be loaded in its solubilized step and the refolding buffer can be introduced afterwards. The spatial separation of the single protein units can prevent aggregation, increasing the yields while at the same time allowing for a capture and concentration step. If the POI does not contain a tag, Size-exclusion Chromatography (SEC) can be applied to achieve a similar spatial separation [59, 60]. While all of these methods can increase refolding yields, they require additional materials and processes, thereby increasing the process costs. As a result, the refolding step remains the key UO in IB processes in terms of cost and economic feasibility.
- V. **Activation** – While not necessary for all POIs, this step might be desirable if certain cofactors or coenzymes are required. These might be needed for correct folding or protein stability [61-63]. For soluble protein production, they are therefore often added either during the USP or after



cell lysis. For IB processes, however, the presence of cofactors or coenzymes can also influence the refolding process by increasing aggregation. Therefore, time and concentration of the addition can have a critical effect on the refolding yield for some POIs.

- VI. **Capture and concentration** – This is the unit operation where the IB process converges with the DSP of soluble proteins again. However, due to the high purity of IBs, the focus for this UO is often more on the separation from misfolded or aggregated proteins than from HCPs [36, 64, 65]. Especially in the case of a batch dilution refolding step, low protein concentrations and large process volumes require a method able to process these large volumes efficiently. Therefore, UF/DF or bind and elute chromatographic methods are often used for the capture and concentration step.

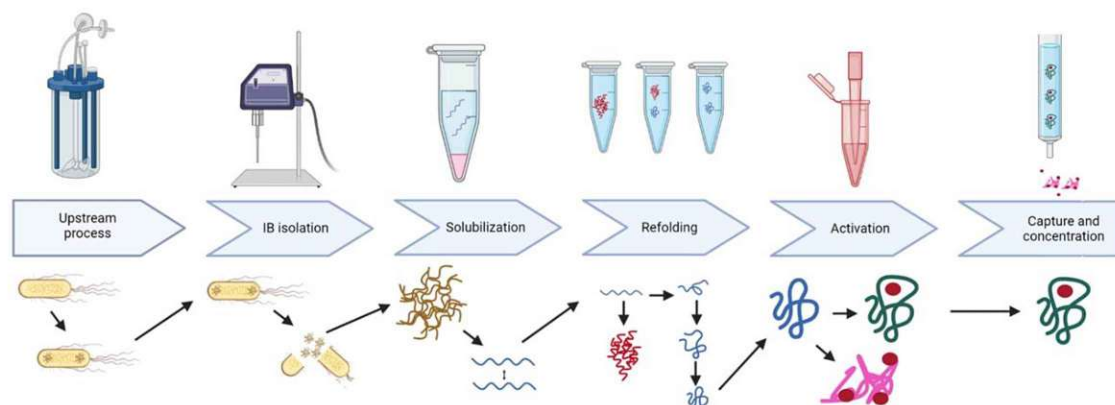


Figure 5: Typical Unit operations of an IB process and the state of POI at the single unit operations. Created with BioRender.com.

The bottleneck for IB processes are the additional DSP steps required to obtain the correctly folded and active protein. The refolding step in particular was described as requiring low protein concentrations and still often resulting in low yields, leading to increased costs for IB processes. In addition, the complex mechanisms of protein folding limit the availability of established platform technologies. Therefore, mostly empiric process development is being employed for new POIs, further increasing production costs. This results in relatively expensive processes, in turn limiting the applicability of the POI to applications for which higher costs are acceptable. One of these applications are biopharmaceutical or medical products. These products in turn require controlled production processes according to current Good Manufacturing Process (cGMP) guidelines [66, 67]. This can be achieved using the traditional, minimal approach for pharmaceutical development. However, the introduction of Quality by Design (QbD) principles allows for the use of increased process understanding and Process Analytical Technologies (PAT) tools for pharmaceutical development and pharmaceutical production processes [68].

### 1.5 Quality by Design

The general concept of QbD was described by J.M. Juran in the 1992, stating that: “Product features and failure rates are largely determined during planning for quality” [69]. The main idea of the Quality by Design approach is that quality aspects should be considered during process development and planning since most quality problems do not originate in how the process is operationally run but relate to how quality was planned. This concept was adopted by different branches of industry (e.g. telecommunication, automobile, and aviation industries) and was implemented by the FDA early in the 21<sup>st</sup> century and further refined and broadened over the last twenty years [70]. In addition, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) released several guidelines (ICH Q8 (R2), ICH Q9, ICH Q10, ICH Q11) to implement the QbD approach. The listed guidelines, and ICH Q8 (R2) in particular, give a comprehensive overview of the QbD principles as well as detailed explanations of the used terminology and comprehensive examples.

The key aspect is described in ICH Q8 (R2) as “It is important to recognize that quality cannot be tested into products; i.e., quality should be built in by design” [71]. A few of the key aspects will be discussed in the following. In general, the goal of any pharmaceutical development is a quality product and a corresponding manufacturing process. Since the introduction of the QbD concept for pharmaceutical development, a distinction between the traditional approach (used interchangeably with minimal approach in ICH Q8) and the enhanced, QbD approach is made by regulatory organs. ICH 8 also acknowledges that most processes register some aspects of the traditional and some aspects of the QbD approach, positioning them “in the middle” between the two approaches. At a minimum, following the traditional approach, Critical Quality Attributes (CQAs), an appropriate manufacturing process, and a control strategy are designed. For an enhanced, QbD approach, in addition a systematic approach to process understanding and a control strategy designed in combination with quality risk management is recommended. Therefore, some of the key aspects are used for both approaches while others, e.g. design spaces and PAT tools, are used exclusively in the QbD approach. An overview over some of these aspects is given here:

- I. **CQAs** - “A CQA is a physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality. CQAs are generally associated with the drug substance, excipients, intermediates (in-process materials) and drug product.” (Direct quote from [71]). CQAs are an essential part of the pharmaceutical development, independent if the traditional or the advanced QbD approach is chosen. It should also be highlighted that CQAs can relate to different materials and intermediates at different stages of the process, and a clear differentiation is made by ICH Q8 between those CQAs (p12 ff).
- II. **Critical process parameters (CPPs)** - CPPs are defined as process parameters that can have an effect on the drug product CQA(s). Similar to CQAs, CPPs are required independent of if a QbD approach is chosen or not. The definition of CPPs is required for defining a control strategy, which should be included also for the traditional approach. For the QbD approach, process parameters and attributes that have an effect on drug product CQA(s) can be identified through prior knowledge, risk assessment, and experimentation. Additionally, the functional link between CPPs and CQAs is used to implement an appropriate control strategy.
- III. **Control Strategy** - “A planned set of controls, derived from current product and process understanding that ensures process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring and control. (ICH Q10)” (Direct quote from [71]).
- IV. **Design Space(s)** - Design spaces can be part of the control strategy. The enhanced product and process understanding demonstrated by appropriate CQAs and CPPs and the link between them allows to not only define setpoints but ranges in which product CQAs are not changed. For the traditional approach, the range of input variables (e.g. material attributes) and process parameters in the control strategy is described one variable at a time. For the advanced, QbD approach a multivariate approach and the interaction between CPPs can be described, resulting in the design space. This describes a multidimensional space working in which is shown to provide the desired and defined quality and is therefore not considered a change.
- V. **PAT tools** - “A system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality.” (Direct quote from [71]).

ICH Q8 (R2) describes a clear minimum of elements that should be included in pharmaceutical development, which firstly includes defining the quality target product profile, then the potential CQAs of the drug product, followed by other CQAs (e.g. of intermediates) and finally the selection of an



appropriate manufacturing process and the definition of a control strategy [71, 72]. In order to employ an advanced QbD approach, scientific understanding of the process has to be demonstrated. While this can be based on prior knowledge and risk assessment, it can also include experimentation, which is the method of choice if no previous information or prior mechanistic process understanding is available for the drug product.

One option to demonstrate the scientific understanding required for QbD principles is through prior knowledge. However, for complex (biological) processes and novel products, platform technologies or mechanistic links demonstrating cause-effect relationships and correlations are often missing [73]. In this case, scientific understanding can be demonstrated through experimentation, using multivariate experiments to understand products and processes. Very few process outcomes are dependent on a single factor. Especially for complex processes, a large number of factors influence the outcome. Moreover, some or all of these factors are commonly not independent of each other, meaning that the change in one factor shows a different influence on the process depending on the setpoints of other factors. Additionally, biological processes often show non-linear behavior and unknown or uncontrollable factors, increasing complexity and leading to less robust processes [74].

Traditional approaches of pharmaceutical development were mainly empirical and experiments have often been conducted one variable at a time. Although this approach can lead to an effective production process, it generates little process and product understanding that could be exploited for process control. In contrast, advanced QbD approaches use multivariate experiments to understand product and process, as well as to establish CPPs and design spaces. These approaches should be systematic, relating mechanistic understanding of material attributes and process parameters to (drug) product CQAs. One of the commonly used methods for multivariate experimentation is the DoE approach.

## 1.6 HRP – a challenging model protein for IB research

For this thesis, many of the listed challenges for QbD principles during the DSP of IBs were studied using the enzyme Horseradish peroxidase (HRP) C1A. A combination of potential applications, prior and state-of-the-art production processes, and enzymatic properties made HRP a good model protein for this study and are summarized here. While some of these properties are specific for HRP or the peroxidase family, many of the listed properties are representative of a broad variety of IB processes and their relation to QbD principles.

The first record of HRP oxidizing an organic substrate was described as early as 1810 – using a piece of the horseradish root [75]. Since then, the reaction mechanism has been described in detail and the sequence and structure of isoenzyme C has been determined [76-79]. As one of the main applications, HRP has been isolated from its natural host, the horseradish root, in order to be used as a reporter molecule. It catalyzes the formation of colored, fluorometric, or luminescent products if incubated with the appropriate substrate and hydrogen peroxide present to catalyze the reaction. Due to the relatively small size of HRP (around 44 kDa) and its established and effective conjugation procedures, it is commonly used as a protein conjugate, often with an antibody or antibody-fragment. Typical applications of such conjugates are e.g. detection for western blotting, ELISA, or immunohistochemistry [80-82]. For these applications, HRP shows some significant advantages, such as high catalytic activity (leading to strong signals in relatively short times), easy readouts (e.g. through the measurement of absorbance using UV/VIS) and high stability due to its small size. Besides this main application, the use of HRP for the removal of aromatic organic compounds in waste water and as a catalyst for polymerization reactions has been discussed in the past [83-86].

In recent years, the potential application of HRP in drug products has emerged, increasing the relevance of a controlled production process and the investigation of QbD principles. Two main areas of medical application are currently envisioned for the use of HRP. Firstly, targeted cancer treatment, and antibody directed enzyme prodrug therapy (ADEPT) in particular, could use conjugated HRP for the catalysis of a pro-drug [87]. In this case, the pro-drug shows no toxicity (e.g. indol-3-acetic acid) but is only

converted into the toxic drug product through HRP [88]. Through conjugation of HRP to a cancer-specific antibody or fragment, the conversion into a toxic substance can be specifically targeted, thereby reducing side effects which is one of the most problematic aspects of cancer treatment today. The existing conjugation protocols for HRP for analytical applications provide an advantage since some prior knowledge exists. The second potential application uses HRP as a polymerization initiator to form so-called hydrogels *in vivo*. These gels show a highly porous structure and can therefore be used in e.g. cartilage repair. It was shown in the past that such gels support the regrowth of native cells and tissue, and degrade over time. Both polymerization speed and crosslinking are easily adjustable by the concentration of HRP and hydrogen peroxide used, allowing for easily tunable hydrogels [89, 90].

However, one potential issue for both of these applications could be the current production process for HRP. The state-of-the-art is still the isolation from plant or plant tissue (hairy root cultures), leading to a mixture of isoenzymes, all of which show extensive plant-type glycosylation patterns [91, 92]. The mixture of isoenzymes and inconsistencies in isoenzyme composition lead to batch-to-batch variations and would require testing and validation of the enzymatic activity for individual batches. While possible, this would be in stark contrast to QbD principles. A second issue that would straight up prevent the use of plant-source HRP is a potential reaction of the human immune system to the foreign glycosylation pattern of HRP. Glycosylation of drug products is usually a CQA, since it can affect both pharmacokinetics as well as immunogenicity of the biotherapeutic [93-95].

In order to understand the reason why the native host is still used as a production organism and what alternatives could exist, the particular properties of HRP should be taken into account:

- I. HRP contains 8 cysteines in its primary structure, forming a total of 4 disulfide bridges in the catalytic active conformation.
- II. Two divalent metal ion coordination sights are present in folded HRP, which are occupied with two  $\text{Ca}^{2+}$  -ions in the native HRP. These significantly increase the stability of the enzyme [96, 97].
- III. In order to be catalytically active, each HRP molecule requires a heme-group in the catalytic center. The charge change of the coordinated iron allows for the oxidation by transferring an electron in the process. The catalytic process has been extensively studied and is fully elucidated [76].
- IV. For the isoenzyme C1A, a total of nine N-glycosylation sites are present on the surface of the enzyme (if in its native conformation). Of these, eight are occupied in the plant and it has been shown that the hydrophobicity of the isoenzyme is significantly increased if this glycosylation pattern is missing. Several studies suggested that different glycosylation patterns influence both stability and enzymatic activity [98, 99].

A graphic overview of these properties is shown in Figure 6. Many of the discussed properties enhance the stability of HRP and hint at a potential evolution towards a stable enzyme which is expressed for extracellular locations. Due to the high relevance of HRP for different biotechnological applications, HRP has been recombinantly produced in a wide variety of different hosts in the past. While mammalian cell lines allow for all required PTMs and can even provide a humanized or human glycosylation pattern, they show disadvantages such as low yields and expensive production processes. A similar issue is encountered using insect cells. Several studies showed expression in fungi and yeasts, providing relatively cheap production. However, hyperglycosylation and corresponding difficulties of the purification process during DSP posed issues [100]. Glycoengineered strains could allow for the expression of HRP with humanized glycosylation patterns in yeasts, but modified yeast strains often exhibit significantly lower space-time yields in bioreactor cultivations, rendering them economically unfeasible [101, 102].

To no big surprise (taking into account the PTMs required and properties of the enzyme), expression of HRP in the cytoplasm of bacterial host cells leads to the formation of IBs, while the translocation to the periplasm was shown to result in low yields and economically unfeasible processes [103]. Several



properties of HRP cause and contribute to the formation of IBs in bacterial expression systems. The reducing environment of the cytoplasm prevents the formation of disulfide bridges. While the lack of glycosylation would be a potential advantage for the medical applications, it results in higher surface hydrophobicity of the enzyme and is therefore believed to increase aggregation and IB formation. This goes hand in hand with the strong expression system used in many of the described production processes, increasing the chance of protein aggregation and IB formation independent of the other solubility issues described. Despite these hurdles, due to the low production costs for the USP of *E. coli*, different approaches to the refolding of HRP have been described in the past and are summarized in Table 1. Previously published HRP production processes were often developed based on traditional approaches, i.e. mainly empirical with development often conducted one variable at a time. UOs were not investigated in an integrated approach but usually one UO at a time. Furthermore, limited analytical tools were available or used for the shown studies, such as Bradford, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), Reinheitszahl (spectrophotometrically) and enzyme activity using ABTS as a substrate.

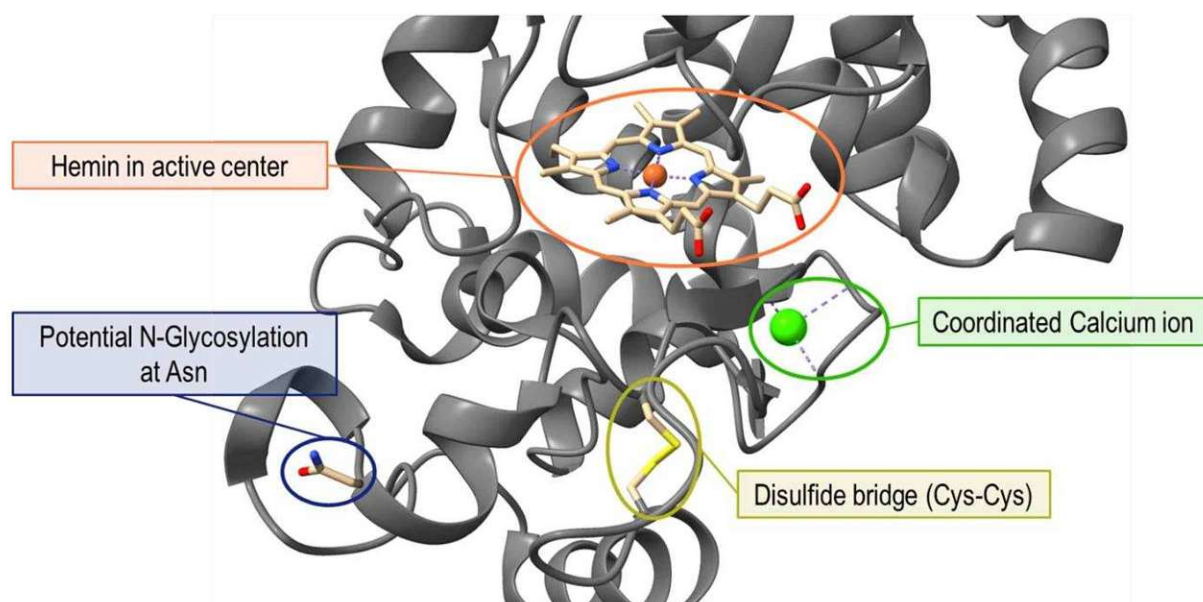


Figure 6: Schematic representation of HRP and its most relevant properties. Created with UCSF Chimera [32] and adapted from [104].

Table 1: Yields and specific activities of inclusion body refolding processes of different class II and III plant peroxidases. [105]

Enzyme	Yield (% or mg/L Culture)	Specific Activity (U/mg)	Reference
HRP	3%	630 U/mg (ABTS) 14500 U/mg (pyrogallol)	[106]
HRP	6-8 mg/L	1160 U/mg (ABTS)	[107]
HRP	24%	10 U/mg (4-aminoantipyrine)	[108]
HRP	16.7 mg/L	4000 U/mg (ABTS)	[109]
HRP	20 mg/L	2000 U/mg (ABTS)	[110]
HRP	15 mg/L	62.5 U/mg (ABTS)	[103]
CWPO_C <sup>1</sup>	0.2 mg/L	1000 U/mg (ABTS)	[111]

CWPO_C <sup>1</sup>	27.3%	1066 U/mg (syringaldazine) 120 U/mg (guaiacol)	[112]
rAtPrx71	28%	1291 U/mg (syringaldazine)	[113]
rAtPrx25	30.3%	270 U/mg (guaiacol)	[113]
TOP <sup>2</sup>	79 mg/L	2950 U/mg (ABTS)	[114]
LDP <sup>3</sup>	16.8 mg/L	70.7 U/mg (TMB) 580.7 U/mg (H <sub>2</sub> O <sub>2</sub> )	[115]
TOP <sup>2</sup>	4.6 mg/L	1100 U/mg (ABTS)	[116]
ATP N <sup>4</sup>	13 mg/L	n.m.	[117]
BP1 <sup>5</sup>	9.4 mg/L	n.m.	[117]
LiP H8 <sup>6</sup>	1%	39 µmol of veratryl alcohol ox/min/mg of protein	[118]
VPL2 <sup>7</sup>	5.5 mg/L	n.m.	[119]
LiP H2 <sup>6</sup>	3.4 mg/L	n.m.	[120]
VBPO <sup>8</sup>	40 mg/L	550 U/mg (bromination of monochlorodimedone)	[121]
LiP <sup>6</sup>	0.38 mg/L	16,300 U/mg (ABTS)	[122]
DyP <sup>9</sup>	1.5 mg/L	247 U/mg (ABTS)	[123]
MnP <sup>10</sup>	2.4%	12.9 U/mg (oxidation of Mn <sup>2+</sup> to Mn <sup>3+</sup> )	[124]
LiP H2 <sup>6</sup>	2.4%	55.6 U/mg (veratryl alcohol)	[125]
MnP <sup>10</sup>	0.275 mg/L	140 U/mg (oxidation of Mn <sup>2+</sup> to Mn <sup>3+</sup> )	[126]
BnPA <sup>11</sup>	29 mg/L	981 U/mg (ABTS)	[127]

<sup>1</sup> Cationic cell wall peroxidase from *Populus alba* L; <sup>2</sup> Tobacco peroxidase; <sup>3</sup> *Lepidium draba* peroxidase; <sup>4</sup> *Arabidopsis thaliana* peroxidase N; <sup>5</sup> Barley grain peroxidase; <sup>6</sup> Lignin peroxidase; <sup>7</sup> *Pleurotus eryngii* versatile peroxidase; <sup>8</sup> Vanadium-dependent bromoperoxidase; <sup>9</sup> Dye-decolorizing peroxidase; <sup>10</sup> Manganese peroxidase; <sup>11</sup> Turnip acidic peroxidase; n.m., not mentioned.

In summary, several aspects of HRP IBs were relevant to investigate QbD principles in IB refolding. While also relevant for current applications of HPR (e.g. medical diagnostics), the two potential medical applications for HRP provided some future relevance of a QbD approach to the recombinant production process. Due to the complexity of HRP, a broad array of different factors of the IB process could be studied. This included the formation of disulfide bridges as well as the addition of different cofactors and coenzymes. Some of these factors have been studied in the past, resulting in the existence of some prior process knowledge. This would allow the comparison of the developed process with different traditional approaches. As a last point, straightforward analytical procedures to measure the enzymatic activity allowed for a high-quality response after the refolding process and therefore provided a reference measurement. This would benefit process development and allow for the investigation of different analytical methods which showed potential for PAT-tools.



## 2 Problem formulation and goal of the thesis

### 2.1 Problem Formulation

For pharmaceutical products and their respective development and production processes, strict regulatory guidelines have been in place since the beginning of the 20<sup>th</sup> century. Starting in 2009, a paradigm shift occurred, moving focus from traditional methods (i.e. a minimal approach) for the development of a product and its manufacturing process to the QbD approach. This means that for pharmaceutical development, the goal is to design a manufacturing process that consistently delivers the intended performance of the product. ICH Q8 (R2) states: **“It is important to recognize that quality cannot be tested into products; i.e., quality should be built in by design.”** [71].

For many (biological) products, the desired properties and therefore quality are defined and present after the initial synthesis of the molecule in the USP. These attributes are not intended to change through the course of the single UOs and are measurable at the beginning as well as the end of the process. However, in the case of IBs, not all the desired properties are defined and present after the initial synthesis of the molecule in the USP. Therefore, QbD for IB processes is lagging behind other pharmaceutical processes particularly in two areas:

(1) Regarding the generation of sound process understanding, protein folding in general, and *in vitro* protein folding in particular, are not well understood processes so far. While remarkable progress was made by both fundamental and applied research to discover and elucidate protein folding pathways and principles, easily and broadly applicable principles for recombinant protein production process development (i.e. platform technologies) are still scarce or missing completely. This is not only true for the refolding step itself but includes all additionally required UOs of an IB process, such as e.g. IB isolation or solubilization, as well. Therefore, refolding processes still require empirical and individual development for each new product. In combination with missing analytical tools, this prevents the generation of sound process understanding.

(2) Regarding the establishment of suitable PAT tools, a plethora of analytical tools is already established and in use for soluble proteins. This, in turn, poses the question why so few analytical tools are available for IB processing. To answer this question, several points should be taken into consideration:

- i. The intrinsic properties of IBs, being insoluble, aggregated proteins, limit available analytical methods. Due to the heterogeneous nature of suspensions, several analytical methods used for soluble proteins (e.g. high-performance liquid chromatography (HPLC)) are not applicable for unit operations up to or including the solubilization step. Furthermore, the potentially inhomogeneous structure of IBs requires complex analytical procedures, which are not feasible for PAT applications.
- ii. For soluble protein production processes, POI concentrations are usually high, at least after the capture step [128]. Refolding methods, and especially batch dilution refolding, requires low protein concentrations in order to keep aggregation to a minimum. Although the POI is present in its soluble state at that point, its concentration falls below the limit of detection of several analytical methods (one often discussed example would be Attenuated total reflectance (ATR)-IR) which are therefore not applicable.
- iii. IBs were viewed as a non-desirable production mode in the past, putting the sole focus on development of soluble protein production processes. Therefore, analytical tools designed specifically to measure IBs properties were developed only recently.
- iv. As the last point to consider, the structure and aggregate state of the product changes until the end of the refolding step. This makes it extraordinarily difficult to identify accessible key performance indicators (KPIs) and CQAs for the single Unit Operations (UOs) up to the capture and concentration step.

Overall, the listed restrictions limit many available PAT tools for IB processing to at-line modes or soft sensor approaches, if available at all.

Summarizing, for IB processes, the conformation of the POI is intentionally changed several times in different UOs, changing even the state of matter from solid IBs to dissolved single protein units. This poses a significant difference to soluble protein production processes and a significant challenge for the generation of sound process understanding and establishing analytical and PAT tools for the USP and the early DSP of IB processes. These limitations severely restrict the development of platform technologies and the application of QbD principles for IB processes.

## 2.2 Goal of the thesis

The goal of this thesis was to address two main parts, namely 1. Process development and 2. PAT tools in order to prevent limitations of 1. The generation of sound process understanding and 2. The establishment of suitable PAT tools. For the first part, HRP IBs were used to develop systematic DoE set-ups and workflows for the IB-specific UOs. Building on the process understanding and reference analytics established in the first part, Reversed Phase Liquid Chromatography (RPLC) was investigated as a potential PAT tool allowing for monitoring and IPC during solubilization and refolding. These two main parts were addressed in the context of the following two scientific questions:

### Scientific Question 1:

What is needed for the systematic development of a **reproducible and scalable** production process of active protein from IBs?

#### Hypothesis:

A systematic and rational DoE based approach to the empiric development for IB processes allows to establish a platform technology adhering to QbD guidelines. ICH Q8 allows to either “establish independent design spaces for one or more unit operations, or to establish a single design space that spans multiple operations”. In the case of IB processes, the interaction between the unit operations solubilization and refolding requires unit operation-spanning DoEs in order to generate scientific understanding. While particular design spaces may vary for different products, the underlying principles and experiments necessary to define suitable design spaces are applicable for diverse proteins, generating broadly applicable process understanding.

#### Approach:

To address Scientific Question 1, the DSP development was separated into three parts. Initially, the process development workflow for all three parts was established using HRP IBs. Subsequently, the established process development approach was applied to and assessed with several different proteins within the context of various industrial cooperations.

- i. **Establish an IB isolation workflow to provide consistent IB starting material**  
A DoE approach was used, identifying significant factors and factor ranges for HPH and IB wash steps. A standardized sampling procedure and analytical workflow was used to consistently monitor IB quantity and quality.
- ii. **Unit Operation-Spanning Investigation of the Redox System**  
Putting a particular focus on the unit operation-spanning investigation (solubilization and refolding) of the redox system, its interaction with the pH-value was investigated.
- iii. **Develop a strategy for the required Cofactor addition (heme)**  
The influence of time of addition and concentration of cofactor were investigated in a DoE



approach. After scale-up to a refolding reactor (1.2 L), the effect of different cofactor feeding profiles on the refolding yield was evaluated.

### Scientific Question 2:

Based on the production process, which **PAT tools** can be developed for the **DSP UOs** of an **IB process**?

#### Hypothesis:

It was hypothesized that due to its tolerance of harsh conditions and denaturing agents, RPLC can be established as a broadly applicable method to measure IB quantity and quality for different UOs. In combination with short analysis times, this enables RPLC methods to be used as an at-line PAT tool for monitoring and In Process Control (IPC) during the early DSP steps. Complementary, mid-infrared (IR) spectroscopy can directly access CQAs (e.g. secondary structure) but requires soluble proteins in relatively high concentrations. While these limitations hinder its application for early DSP steps, using an in-line IR spectrometer for DSP steps after protein activation allows for the generation of process understanding and monitoring of CQAs in real time. Furthermore, required protein concentrations can be significantly lowered by the combination of larger optical path lengths and an external cavity-quantum cascade laser, applied in a novel flow-through spectrometer.

#### Approach:

Scientific Question 2 was split into 2 subparts, encompassing all UOs from IB isolation to a potential polishing step. A broadly applicable and at-line compatible RPLC method was developed with a particular focus on monitoring and control during early DSP steps. Using model systems for different chromatographic methods, the applicability of in-line IR spectroscopy for qualification and quantification of proteins was investigated.

i. **Analytical methods for product identification and quantification for different UOs and applicability for monitoring and IPC**

Based on an Analytical QbD workflow, a broadly applicable RPLC method was developed. In combination with a suitable sample preparation step, this method allows product identification and quantification for different IB proteins for all UOs from IB isolation to the capture and concentration step. The developed RPLC method was adapted to be applicable as a monitoring and IPC PAT tool in an at-line mode. Validation of the method was performed using the sound process understanding generated during the development of the HRP IB process developed for Scientific Question 1 (parts 1.ii and 1.iii).

ii. **IR for in-line monitoring**

Develop an **in-line IR measurement** for bind-elute chromatography steps:

Using a preparative ion exchange chromatography system with model proteins, a novel background compensation approach was developed. Based on the chemometric subtraction of the applied NaCl gradient during elution, real time information about secondary structures of the eluting proteins becomes accessible.

Near real time **compositional analysis** of low resolution fractions during SEC:

Separation of model proteins via preparative size exclusion chromatography was used to demonstrate the applicability of an in-line IR spectrometer for the acquisition of compositional information of co-eluting protein fractions. The developed method allows for the in-line, real-time analysis of secondary structures of eluting proteins, providing CQA data not accessible via the UV monitoring traditionally applied for preparative chromatography.

### 3 Results and Discussion

This thesis contains four first author papers, one patent, and two book chapters. These publications are structured in two parts and five sections, which are shown in Figure 7. Where applicable, the publications are complemented with information on additional research, including a book chapter as well as a scientific paper in the Appendix II. In addition, it contains a short article on the production process of HRP from IBs, published in BioSpektrum (Springer).

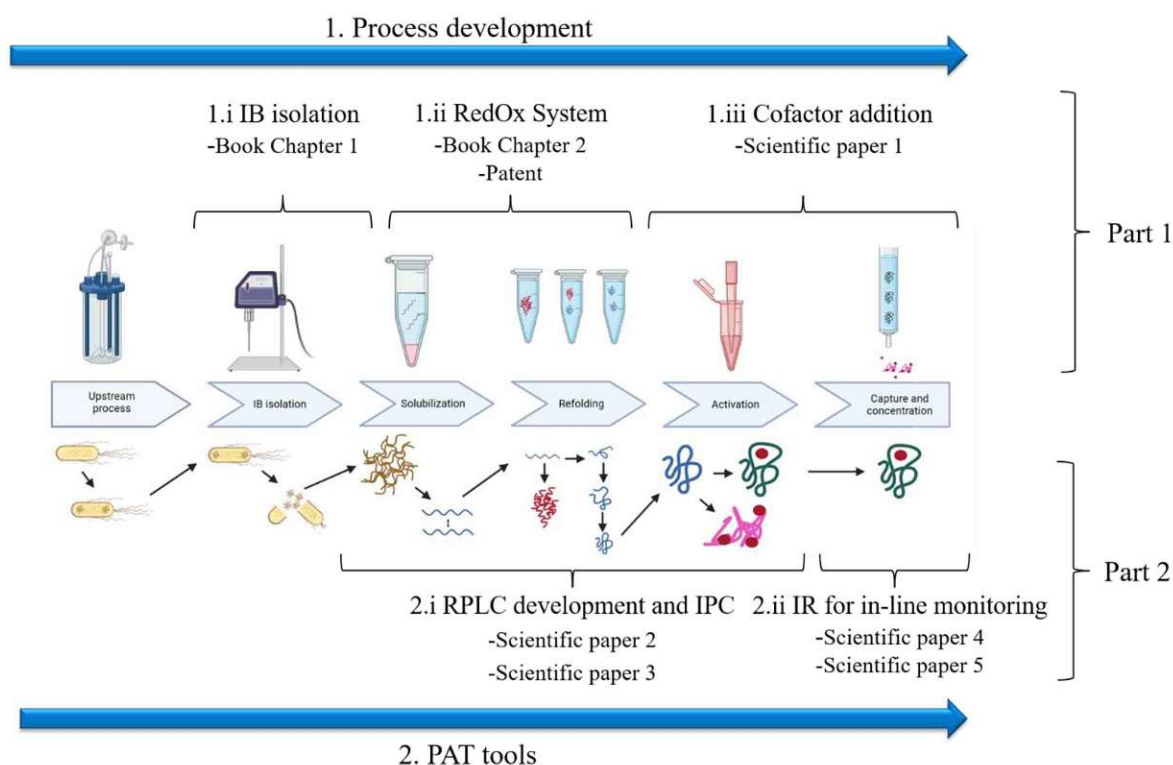


Figure 7: Overview of the two main parts discussed in this thesis in relation to QbD processes for IBs. Part 1 addresses an empiric, but structured and transferrable process development approach, aiming to generate broadly applicable process understanding. Part 2 investigates suitable PAT tools for different UOs and protein states, aiming to access KPIs and CQAs and allow monitoring and control for UOs of the IB process. Created with BioRender.com.

In this chapter, state of the art and challenges for the five sections are described. As an overview, Table 2 shows several aspects of an enhanced QbD approach and for which thesis sections these aspects are relevant and/or applicable. In part one, workflows for IB isolation using HPH and redox system investigations were established for HPR IBs. While DoEs are a common tool for QbD process development, systematic approaches for IB isolation in combination with defined sampling procedures are scarcely reported in literature. Section 1.ii addresses the development of a suitable redox system for solubilization and refolding – providing a unit operation spanning design space. This is particularly important for sound process understanding due to the interaction of reducing agent, oxidizing agent and pH value. Several previous refolding protocols have reported the batch addition of the required cofactor hemin, often at the beginning of the refolding step [96, 103, 108, 129]. In section 1.iii the hemin addition during refolding was investigated in a multivariate approach to better understand the effect of time of addition and concentration on the refolding yield. Based on the developed design space, a linear feeding profile was established, significantly increasing refolding yields and resulting in a novel refolding method for HRP IBs. Complemented by a capture and concentration via HIC, the developed process enables production of highly pure, non-glycosylated HRP in the gram-scale. The availability of these quantities allows for future exploration for different application fields as well as further research on HRP, e.g. structural analysis [130]. Building on the sound process understanding developed in part one,



part two aimed to develop complementary PAT tools. Taking advantage of the denaturing conditions and short analysis times, RPLC was used for the measurement of POI concentration in the early DSP steps. In combination with the DoE approach for the redox system developed in Section 1.ii, this enabled the identification of monomeric HRP in the solubilizate as a key quality attribute and, subsequently, the use of the method for monitoring and IPC during the solubilization and refolding steps. In the future, this method could be used to react to variations in IB quality (e.g. different IB batches), allowing to generate reproducible refolding yields independent of the upstream process. In the last section, 2.ii, IR spectroscopy was used as a tool for in-line monitoring of different preparative chromatography methods. IR for in-line monitoring has previously been used for analytical chromatography applications in isocratic elution mode. However, the varying buffer compositions throughout bind-elute runs pose significant challenges for bind-elute methods. Using a newly developed approach for background compensation we showed the possibility to generate real-time information about secondary structure of proteins during elution. The ability to identify the secondary structure was then used to analyze the composition of mixed fractions containing two proteins during SEC runs with model proteins. The described set-up and method can therefore be used for future chromatic process development either stand-alone or in combination with off-line analytics.

Table 2: Overview of selected aspects of the enhanced, Quality by Design approach and the relevance of these aspects for the sections discussed in this thesis.

Enhanced, Quality by Design Approach [71]	Relevant for Sections
Identifying potential critical quality attributes (CQAs) for intermediates (in-process materials)	All
Identifying, through e.g., prior knowledge, experimentation, and risk assessment, the material attributes and process parameters that can have an effect on product CQAs	1.i IB isolation 1.ii Redox System 1.iii Cofactor addition 2.i RPLC for IPC
Multivariate experiments to understand product and process	1.i IB isolation 1.ii Redox System 1.iii Cofactor addition 2.i RPLC for IPC
Establishment of design space	1.ii Redox System 1.iii Cofactor addition
PAT tools	2.i RPLC for IPC 2.ii IR for monitoring
Manufacturing process adjustable within design space	1.ii Redox System 1.iii Cofactor addition

### 3.1 Scientific Question 1

What is needed for the systematic development of a **robust and scalable** production process for active protein from IBs?

#### i. Establish an IB isolation workflow to provide consistent IB starting material

Title: High Pressure Homogenization for Inclusion Body Isolation

The intracellular protein production using *E. coli*, the DSP usually starts with cell lysis. Three methods are commonly employed for this purpose, namely HPH, ultrasonication and/or enzymatic lysis [131-133]. All of these methods were originally developed for soluble intracellular proteins and are therefore commonly optimized for soluble protein production. However, while the same methods are used,

suitable conditions may vary depending on the expression type (either soluble or IB). Still, process development for IBs is often focused on the solubilization and refolding steps, negating the crucial influence of cell lysis on IB quality and therefore on later unit operations. One of the reasons for this might be that the correlation of CQAs during HPH to the behavior and CQAs in later DSP unit operations has been difficult. Therefore, the KPIs often conceived as the most important remain titer and purity of the POI, although these might not lead to the best overall process [134].

The importance of an integrated development for cell lysis, IB wash and solubilization has previously been shown [135]. In this book chapter, a workflow allowing the identification of suitable cell lysis conditions for IBs using HPH with a DoE approach (Figure 8) was presented, varying the factors pressure and number of passages. While exemplified for HRP IBs, the workflow is adaptable to a broad variety of POIs. The multivariate analysis enables identification of interacting factors and quadratic interactions, surpassing the capabilities of univariate experiments. HPH is employed, allowing scale-up and achieving higher IB purity compared to ultrasonication or enzymatic lysis due to applied shear forces [135]. The protocol utilizes three analytical methods: RPLC and SDS-PAGE for measuring POI titer and purity, and DNA concentration analysis as an indicator of cell lysis efficacy. As an essential additional analytical method, assessment of the refolding yield can be incorporated if a protocol is established for the POI:

“In our experience, the achieved refolding yield is the most significant response in order to identify suitable homogenization conditions. However, as washed IBs are needed in a first step to develop a refolding protocol, this requires an iterative approach. Therefore, if no refolding protocol is established, a high purity (measured, e.g., with RPLC or SDS-PAGE) of IBs is a good starting point to develop a refolding protocol. The homogenization conditions can then be adapted after the refolding protocol is in place.” (Direct quote from [136]).

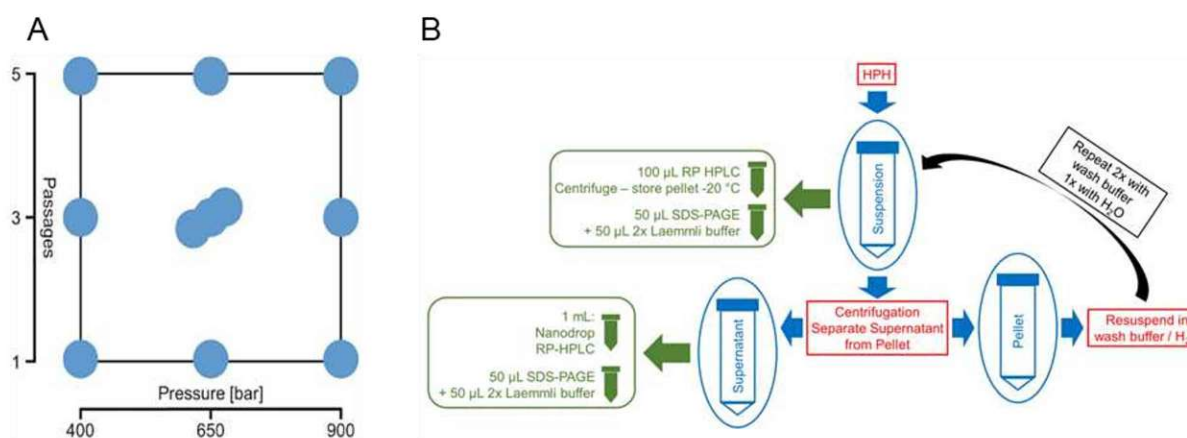


Figure 8: Setup of the factors of the DoE and the corresponding sampling plan. (A) Two factors (pressure and passages) and ranges for a suggested central composite face centered design for the investigation of high pressure homogenization. (B) A sampling plan for the presented workflow. The red boxes present operations, the blue ellipses the processed IBs and the green boxes the samples drawn. Adapted from [136].

Notably, the core elements of this approach, including sampling strategy and analytical methods, are versatile and can be applied beyond the shown DoE, allowing to screen washing buffers or scale-up IB isolation processes. Therefore, this systematic, scalable method for optimizing cell lysis conditions offers a robust framework applicable for various proteins and experimental scales, especially in regards to principles of QbD for bioprocess development.

**Authors:** Julian Ebner, Viktor Sedlmayr, Robert Klausser

**Published:** Kopp, J., Spadiut, O. (eds) Inclusion Bodies. Methods in Molecular Biology, vol 2617. Humana, New York, NY. [https://doi.org/10.1007/978-1-0716-2930-7\\_9](https://doi.org/10.1007/978-1-0716-2930-7_9)



**Contribution:** The study design was planned together with Viktor Sedlmayr and Robert Klausser. Experiments and data analysis were performed together with Viktor Sedlmayr and Robert Klausser. I wrote the book chapter. Viktor Sedlmayr and Robert Klausser gave input on the final version. All authors approved the submitted version of the manuscript.

Additional research:

- Besides the discussed use for HRP, the described workflow was successfully applied in two industry cooperation with non-peroxidase POIs. Within one of these projects, the shown workflow was combined with a small-scale solubilization and refolding approach. This enabled analysis of required refolding volume and therefore made different KPIs accessible (such as e.g. volumetric titer of correctly folded protein, titer of correctly folded protein to cell weight, etc.).
- In a separate study, we characterized an ultrasonic homogenization device for cell lysis applicable for small sample volumes (<20 mL). The method was investigated for both soluble and IB POIs and compared to enzymatic cell lysis (Lysozyme) and HPH. We showed that the use of an ultrasonic lance is suited for small-scale screening, however, suitable cell disruption conditions depend on the POI. Furthermore, we showed that high purities of HRP IBs could be achieved using HPH, while purities were significantly lower for the ultrasonic lance over the whole investigated design space. This suggests that while the method can be used in screening experiments for IBs, care should be taken as the lower purity might influence subsequent UOs, in particular the refolding step.

## ii. Unit Operation-Spanning Investigation of the Redox System

Title: Unit Operation-Spanning Investigation of the Redox System

The formation of IBs in *E. coli* is often attributed to the presence of disulfide bridges in the expressed POI. The reducing environment of the bacterial cytoplasm hinders the post-translational formation of these crucial structures, leading to protein misfolding and aggregation [35, 137]. To obtain the POI in its active form, the DSP of IBs typically involves three additional unit operations: isolation and washing of IBs, solubilization, and refolding [138-140]. For proteins containing disulfide bridges, a redox system is employed to achieve the correctly folded and active POI. This system generally involves the addition of a reducing agent during solubilization to break existing disulfide bonds formed during earlier UOs, followed by an oxidizing agent in the refolding buffer facilitating their (re)formation [141, 142]. However, the optimal concentrations of these agents are highly specific for each POI, necessitating an empirical approach to identify suitable conditions for each new protein production process [143, 144].

The shown protocol presents a workflow for identifying suitable redox conditions for batch dilution refolding using a DoE approach. The DoE varies three critical factors: the concentration of reducing agent during solubilization, the concentration of oxidizing agent during refolding, and the pH-value of the refolding buffer. This multivariate approach is essential due to the interplay between these factors. The pH-dependent kinetics of the redox system significantly influence refolding yield by affecting the rate of disulfide bridge formation, which competes with protein aggregation [145]. Moreover, the concentrations of reducing and oxidizing agents exhibit complex interactions that span multiple UOs. For the presented workflow, the specific redox agents can be adapted for different POIs. However, one of the disadvantages of the workflow is that it requires a pre-established basic refolding protocol and an analytical method to quantify correctly folded POI. As a big advantage, by employing this systematic, multivariate strategy, redox conditions can efficiently be optimized for different proteins, potentially improving yields, and developing process understanding.

The described workflow was one of the essential parts in order to develop a patentable refolding process for HRP. The described interaction between the redox system and pH and the systematic analysis played a key role in significantly increasing refolding yields. In addition, it formed an essential part of the

claims in the patent, where the used DoE and derived influence of process parameters was essential. Table 3 shows the factors and ranges investigated in three key DoEs, described in detail in the granted patent. DoE 1 was based on the assumption that the redox system was influenced by two interacting factors, namely the reducing agent (Dithiothreitol (DTT)) during solubilization and the oxidizing agent (Glutathione disulfide (GSSG)) during refolding. Results confirmed this assumption, showing that both factors as well as the interaction term have a significant influence on the correctly folded protein. In the second DoE, the protein concentration during solubilization and refolding was added as a third factor. The hypothesis was that higher protein concentrations would require higher concentrations of reducing agent during solubilization and higher concentrations of oxidizing agent during the refolding in order to counteract the higher amount of reduced protein and carry-over reducing agent. Surprisingly, DoE 2 showed that higher protein concentrations benefited from lower DTT concentrations during solubilization, while GSSG concentrations leading to the highest refolding yields stayed almost constant for different protein concentrations, i.e. GSSG showed no interaction with the protein concentration. This suggested that the reaction kinetics played a crucial role. Therefore, DoE 3 was designed in order to investigate different reaction kinetics for the refolding process. These could be adjusted by performing the reaction at different pH-values, which controls the reaction kinetics of the disulfide bond formation of both the cysteines in the protein as well as the reducing and oxidizing agent. As shown in Figure 9, this led to higher refolding yields as well as a significant broadening of the design space at faster reaction speeds at pH 10.

Table 3: The Factors and Ranges investigated in three DoEs, spanning the unit operations solubilization and refolding. Adapted from [105].

Unit Operation	Factor	Range
Solubilization	DTT	2.5 mM–28.44 mM
	Protein concentration	20 g/L–80 g/L
	pH	7–10
Refolding	GSSG	0.4 mM–3.5 mM
	Protein concentration	0.5 g/L–2 g/L
	pH	7–10

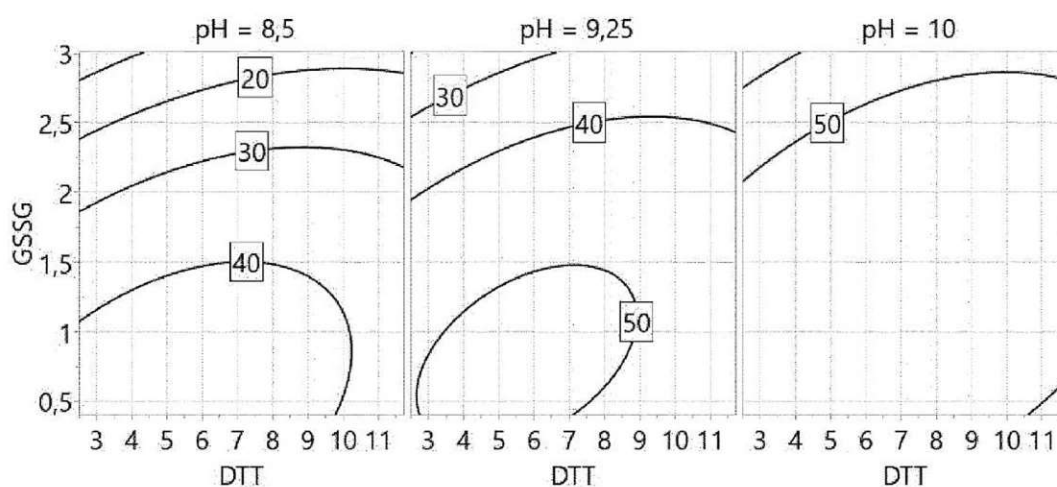


Figure 9: Response contour plot for pH 8.5, pH 9.25 and pH 10 for different DTT and GSSG concentrations (both in mM). The volumetric activity after refolding is plotted as the response. Based on: Patent No.: EP20165131.2 Methods for producing heme peroxidase.



The combination of these experiments shows how several DoEs can be combined to increase refolding yields while, at the same time, increase process knowledge and allow for the definition of design spaces. As analytical procedures to measure intermediates are often limited, an iterative approach using a "Design of Design of Experiments" could present a promising methodology towards platform technologies.

Authors: Julian Ebner, Diana Humer, Viktor Sedlmayr

Published: Kopp, J., Spadiut, O. (eds) Inclusion Bodies. Methods in Molecular Biology, vol 2617. Humana, New York, NY. [https://doi.org/10.1007/978-1-0716-2930-7\\_9](https://doi.org/10.1007/978-1-0716-2930-7_9)

Contribution: The study design was planned together with Viktor Sedlmayr. Experiments and data analysis were performed together with Viktor Sedlmayr. I wrote the book chapter. Viktor Sedlmayr and Diana Humer gave input on the final version. All authors approved the submitted version of the manuscript.

Patent: Patent No.: EP20165131.2 Methods for producing heme peroxidase

Additional research: Besides the discussed use for HRP, the described workflow was successfully applied in an industry cooperation for a non-peroxidase POI.

Based on an industrial cooperation, the developed method and the process designed for HRP was used to investigate the potential applicability of the BioLector® Pro for refolding processes development. The BioLector® Pro is an advanced microbioreactor system utilizing a 48-well plate format with several process controls. A detailed description of this project can be found in [146]. Several relevant results were discovered during this investigation and are summarized below:

- i. The redox system using CH as reducing and CSSC as oxidizing agent was established and resulted in comparable refolding yields compared to the DTT/GSSG system. The use of the CH/CSSC redox system showed vastly increased cost efficiency while comparable results were achieved both in regard to the refolding yield as well as the process development (i.e. the presented workflow was applicable for both redox systems).
- ii. Since the BioLector® plates were originally designed for fermentation process screening, a high oxygen input even at low revolutions was present, resulting in a high dissolved oxygen (DO) content. This was contrary to the above-described small-scale approach (2 mL Eppendorf tubes), where little to no oxygen input was present during refolding. A significant shift in redox conditions was therefore observed, with low reducing agent concentrations required for low oxygen input (Eppendorf tubes) and high reducing agent concentrations required for high oxygen input (BioLector® Plates). Therefore, it was concluded that the DO was a CPP and the use of the BioLector® Pro during process design had a potential advantage since it allowed for monitoring and control of the oxygen input, potentially making upscaling easier. It should be noted that both approaches led to comparable refolding yields if appropriate conditions were chosen. This shows not only the advantage of using a unit-spanning DoE approach but also how the influence of the chosen system on later upscaling effects should be considered early on in process development.
- iii. While it was not possible to establish a mechanistic link between DO and refolding yield, the in-line measurement of DO of the redox system showed potential to be used as a (soft) sensor for refolding processes [147, 148]. Since the redox potential directly influences the formation of correct disulfide bridges in HRP, monitoring the redox balance could potentially serve as an indicator of the refolding progress. By monitoring the redox system, the refolding state could be assessed and controlled via feeding of a redox pair (e.g. CS/CCSSC), providing a tool for real-time process control. This approach could help achieve a more consistent and efficient refolding process, potentially reducing the required number of experiments.

### iii. Develop a strategy for the required Cofactor addition (heme)

#### Title: Scalable High-Performance Production of Recombinant Horseradish Peroxidase from *E. coli* Inclusion Bodies

Hemin addition plays a critical role in the controlled refolding process of HRP. Previous studies showed that it was essential to obtain holo-HRP. However, somewhat counterintuitive, it was also shown that it is not only not required for the correct apo-HRP formation but actually reduces refolding yields if present in the refolding from the beginning [96, 103, 108, 109]. It was therefore assumed that both the time point of hemin addition as well as the concentration will have an influence on the refolding yield. Accordingly, a small-scale DoE was performed, using time of hemin addition as well as final hemin concentration in the refolding buffer as factors. Figure 10 shows that both factors had a significant influence on the refolding yield. Furthermore, it shows the interaction of the factors, with higher hemin concentrations having a pronounced negative effect on the refolding yield if added immediately after the refolding start. The negative effect of higher hemin concentrations on refolding yield was significantly reduced for later addition times. This further solidified the conclusion that early addition of hemin led to increased aggregation due to the hydrophobic nature of hemin and its interaction especially early in the refolding process. While the maximum refolding yield was achieved at 6  $\mu\text{M}$  hemin, these results also allowed us to choose a concentration of 20  $\mu\text{M}$  hemin for further experiments in order to avoid hemin concentration becoming the limiting factor. This was possible because the difference in refolding yield could be predicted based on the model and was shown to be minimal.

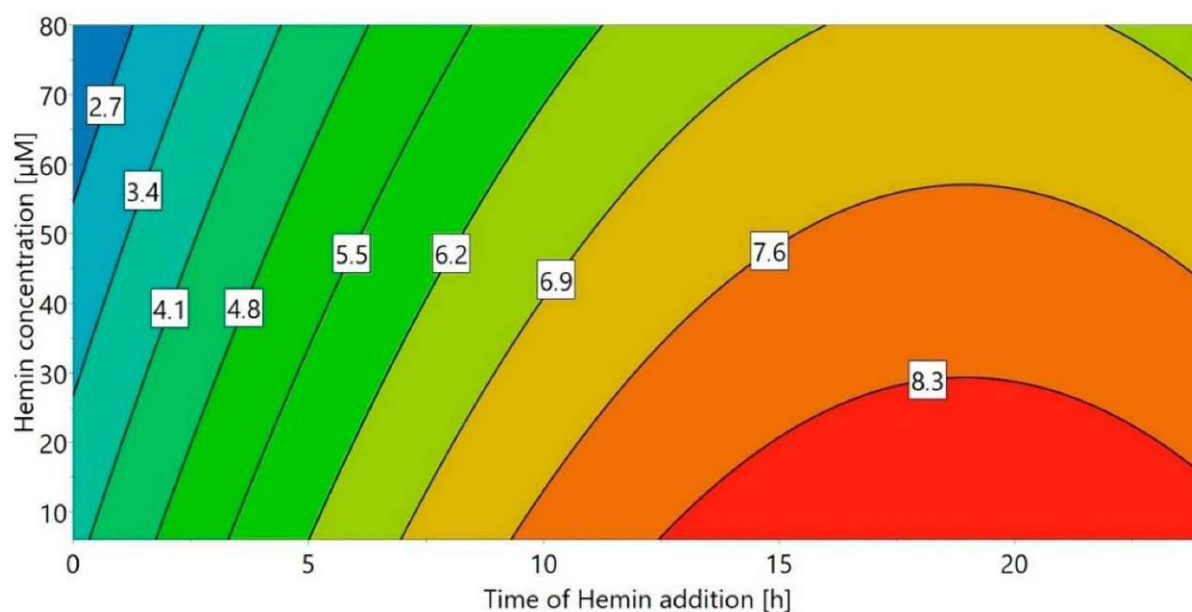


Figure 10: Response contour plot with time of hemin addition and hemin concentration as factors. The volumetric activity is shown as the response. [105]

In order to allow for the higher number of experiments required by the chosen DoE approaches, experiments were performed in small-scale (2 mL). While the smaller-scale presents a lot of advantages, such as lower material consumption, higher experiment number and simpler equipment requirements, it prevents some of the process options available for bench-scale settings. In the described DoE, it was demonstrated that hemin addition presents a trade-off between formation of the holo-HRP and concurrent protein aggregation triggered by the hydrophobicity of free hemin. Based on this developed process understanding, it was hypothesized that self-aggregation as well as aggregation with incorrectly folded protein could be kept to a minimum by adding a small amount of hemin over a longer time period, i.e. apply a linear feeding profile for hemin during refolding. This approach was tested in a 1.2 L bench-scale refolding vessel and led to increased refolding yields compared to a one-time batch addition, even if the batch addition was performed at the optimal time and hemin concentration. In a last set of



experiments, the different hemin addition methods (batch vs. linear feed) were compared at both pH 8.5 and pH 10 in order to investigate possible dependencies between the redox system and the hemin addition method. In total, 4 experiments were performed. The yield increase achieved by using a linear hemin feed compared to batch addition were similar at both pH 8.5 and pH 10. Based on these results, it was concluded that the hemin addition method was independent of the pH-value.

Hydrophobic Interaction Chromatography (HIC) was employed to purify the refolded HRP, utilizing salting out as a crucial step to enhance the process. It was shown that high salt concentrations, such as 4 M NaCl, precipitated impurities while stabilizing correctly folded HRP in solution. This pre-purification through salting out increased the efficiency of HIC by allowing impurities to be separated before the actual chromatography step.

In a last step, two bench-scale (1.2 L) runs with linear hemin feed were performed at pH 8.5 and pH 10 with all other parameters being kept identical. As suspected based on previous results, the different pH-value did indeed increase the refolding yield (from 44% to 74%), showing that the results obtained during the redox DoEs were scalable and not limited to small-scale experiments.

The final results achieved by the described process showed a significant improvement both in quality as well as quantity of the recombinantly produced HRP compared to previously described processes. A high specific activity of  $1468 \pm 24$  U/mg, excellent purity of  $\geq 99\%$  (by SEC-HPLC), and an overall yield of 28 mg active HRP per 100 mg expressed protein (Table 4) were achieved. The achieved activity levels of HRP are comparable to commercially available plant-derived HRP while being produced as a non-glycosylated single isoform of HRP with consistent quality. This can provide a significant advantage over the mixed isoforms and glycosylated HRP obtained from plant-derived sources in the future, especially for applications in the pharmaceutical area.

*Table 4: Final results for refolding performed in 1.2 L scale at pH 10. [105]*

Process variable	Final result
Specific activity [U/mg]	1468±24
Purity SEC-HPLC [%]	≥99
Overall yield active HRP per 100mg expressed protein [mg]	28
Pure HRP/L culture medium [mg]	959
Reinheitszahl	4.3

**Authors:** Julian Ebner\*, Diana Humer\* and Oliver Spadiut

\*Authors contributed equally to this work.

**Published:** Int. J. Mol. Sci. **2020**, 21, 4625; <https://doi.org/10.3390/ijms21134625>

**Contribution:** This is a shared first authorship with Diana Humer. Conceptualization, O.S.; Methodology, O.S., D.H. and J.E.; Validation, O.S., D.H. and J.E.; Formal analysis, D.H. and J.E.; Investigation, D.H. and J.E.; Resources, O.S.; Data curation, D.H. and J.E.; Writing—original draft preparation, D.H. and J.E.; Writing—review and editing, O.S.; Visualization, D.H. and J.E.; Supervision, O.S.; Project administration, O.S.; Funding acquisition, O.S. All authors have read and agreed to the published version of the manuscript.

**Patent:** Patent No.: EP20165131.2 Methods for producing heme peroxidase

**Additional research:**

- Based on the detailed investigation of the DSP for HRP IBs, a book chapter (see Appendix II - 7.1 Book Chapter: The Purification of Heme Peroxidases from *E. coli* Inclusion Bodies) was published describing a bench-scale protocol covering all process steps from IB isolation to

purification via HIC. The aim of the protocol in this book chapter was to provide a comprehensive and easy to follow description of the methods and process steps. This includes a description of all analytical methods used as well as a “Notes” section providing additional insights, minor optimizations, and practical considerations that may not be explicitly detailed in the conventional Materials and Methods section of a scientific paper. We included detailed descriptions for each unit operation, specific equipment recommendations and settings as well as process parameters influencing yield and quality of the obtained HRP throughout the protocol.

- In an unpublished study, different formulation strategies and storage conditions were investigated. We compared liquid forms and lyophilization in buffers with different salt concentrations and different storage temperatures (RT, 4 °C, -20 °C). While this gave an overview of potential formulation and storage procedures, more research in this field would be required, especially if QbD principles would be taken into account.
- Recombinant HRP produced according to the described protocol was successfully conjugated to Protein L (an immunoglobulin-binding protein) and used to detect Herceptin (a monoclonal antibody binding to the HER2 receptor), showing full functionality. More details can be found Appendix II - 7.2 Scientific paper: Recombinant Protein L.

### 3.2 Scientific Question 2

Based on the production process, which **PAT tools** can be developed for the **DSP UOs** of an **IB process**?

#### i. Analytical methods for product identification and quantification for different UOs and applicability for monitoring and IPC

Title (1<sup>st</sup> paper): Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles

Title (2<sup>nd</sup> paper): At-Line Reversed Phase Liquid Chromatography for In-Process Monitoring of Inclusion Body Solubilization

Independent of whether the POI is expressed in soluble or insoluble form (IBs), common KPIs and CQAs such as target protein titer and purity are requiring monitoring tools before quality control analysis. For IBs, the selection of such tools is limited by the fact that the state of matter of IBs changes over different UOs. Ideally, the tools should be able to access the required properties in amorphous solids, suspensions and in soluble form of the protein. Alternatively, the POI can be brought into solution by an according sample preparation step. Solubilization in high concentrations of urea or GndHCl, often combined with very high pH-values, is commonly used [36, 55-57]. This requires the analytical tool used to be able to withstand, and deliver accurate results, at these harsh conditions. Therefore, tools that denature the analytes per default are usually the first choice. Besides the very commonly used SDS-PAGE, RPLC complies with the stated requirements [149, 150]. It should be noted that while SDS-PAGE is compatible with high concentrations of urea, it is not compatible with the second commonly used chaotropic agent GndHCl, restricting its application in the IB processes of various POIs.

In a first step, the general properties of a broadly applicable reversed-phase high-performance liquid chromatography (RP-HPLC) method for protein quantification were explored. The method described in the 1<sup>st</sup> paper of this section was suitable for all five tested POIs, which included soluble as well as IB samples. The stationary phase in particular was chosen based on state-of-the art technology in order to minimize protein on-column adsorption, allowing for high efficiency and accurate quantification. Compared to SDS-PAGE, RP-HPLC offered several advantages. SDS-PAGE, while useful for



qualitative analysis and potentially allowing the detection of low concentrations through staining techniques, did not provide comparable quantitative accuracy and required longer analysis times. Therefore, the developed RP-HPLC method provided a faster and more precise alternative, allowing for robust, high-throughput analysis of protein titer and purity. As an additional advantage, the method is MS-compatible, enabling the option for more detailed protein characterization [151, 152]. Measurements of POI titer and purity are possible for both USP as well as DSP samples, which allows the method to be used over the whole production process. The use of Analytical Quality by Design (AQbD) principles in method development ensures that the process is adaptable and is a first step to meet regulatory standards, potentially enabling it as a future tool for protein analysis across various biopharmaceutical production processes and POIs [153, 154].

In a second step, the described RPLC method was used to specifically monitor and control the HRP solubilization and refolding process. The results were described in detail in the 2<sup>nd</sup> paper of this section. Three criteria were required in order to provide a RPLC method suitable for IPC: 1.) quantification capability of the chosen (intermediate) quality attribute, 2.) influence of process parameters on the quality attribute and on the refolding behavior and yield, and 3.) timely measurement of the quality attribute in relation to overall process times [155]. Therefore, in a first step, the run time was shortened to 8.1 minutes per sample. This was possible due to the in-depth RPLC method development described above and allowed for adequately short analysis times, especially during the solubilization. This is important especially for analyzing samples during solubilization as previous experiments showed that solubilization times had a significant influence on the achieved refolding yield, with short solubilization times (around 30 minutes) resulting in increased refolding yields.

Subsequently, the monomeric HRP concentration was defined as the key quality attribute during solubilization. A DoE approach was used to examine the effects of the DTT concentration and solubilization time on the monomeric HRP concentration. While it was known from previous experiments (See 1.ii. Unit Operation-Spanning Investigation of the redox System) that DTT had a significant influence on the refolding yield, no measurable quality attribute was available so far, therefore limiting process optimization to empirical investigations. It was hypothesized that the reducing agent was essential for maintaining the reduced cysteine residues, thereby preventing multimer formation and subsequent aggregation. As shown in Figure 11, samples without DTT exhibited an additional peak in RPLC analysis, while extended solubilization times of 21 hours led to complete protein degradation regardless of DTT concentration. Non-reducing SDS-PAGE analysis revealed a HRP target band at 34 kDa and a probable dimer formation at 68 kDa, with increased dimer presence observed at extended solubilization times, particularly without DTT. Both analytical methods successfully detected changes in monomeric HRP concentration. Due to orthogonal separation principles of the two methods, both approaches provided valuable information adding to the process understanding. The dimer formation of HRP was only quantifiable with SDS-PAGE. However, for the measurement of monomeric HRP concentration for future at-line monitoring and IPC applications, the rapid analysis time of under 10 minutes for RPLC proved more applicable compared to SDS-PAGE. Furthermore, multivariate data analysis revealed differences between SDS-PAGE and RPLC results (see comparison to the refolding activity in Figure 12). While SDS-PAGE suggested monomeric HRP concentration depended solely on solubilization time, RPLC indicated both DTT concentration and solubilization time significantly influenced the outcome. The enzymatic activity measurements after refolding aligned with RPLC predictions, showing optimal results for short solubilization times and increased DTT concentrations. In summary, these findings demonstrated that RPLC was better suited, compared to SDS-PAGE, for predicting refolding yields and met all criteria for an effective at-line monitoring tool in HRP IB processing.

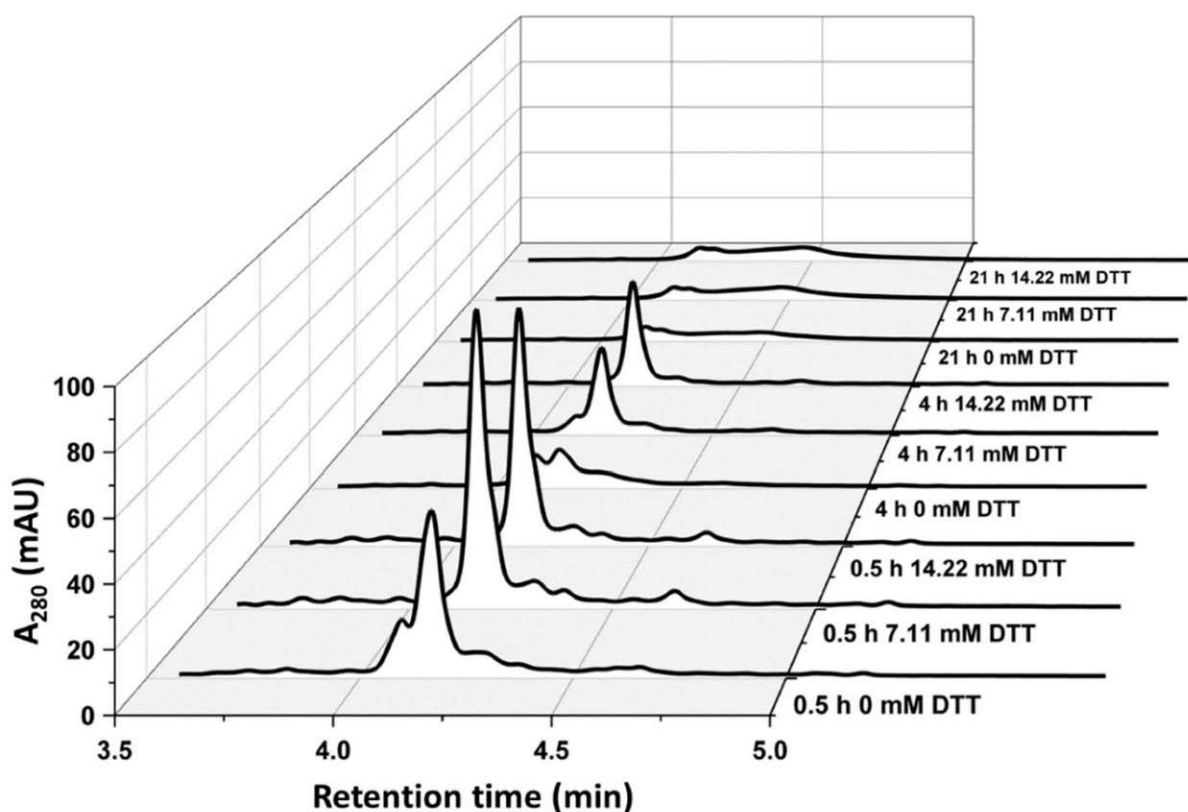


Figure 11: RPLC chromatograms at 280 nm quantifying monomeric HRP eluting at 4.18 min. The results demonstrate the trends of solubilization at three applied DTT concentrations (0 mM, 7.11 mM and 14.22 mM) for 0.5 h of solubilization, 4 h of solubilization and 21 h of solubilization. [156]

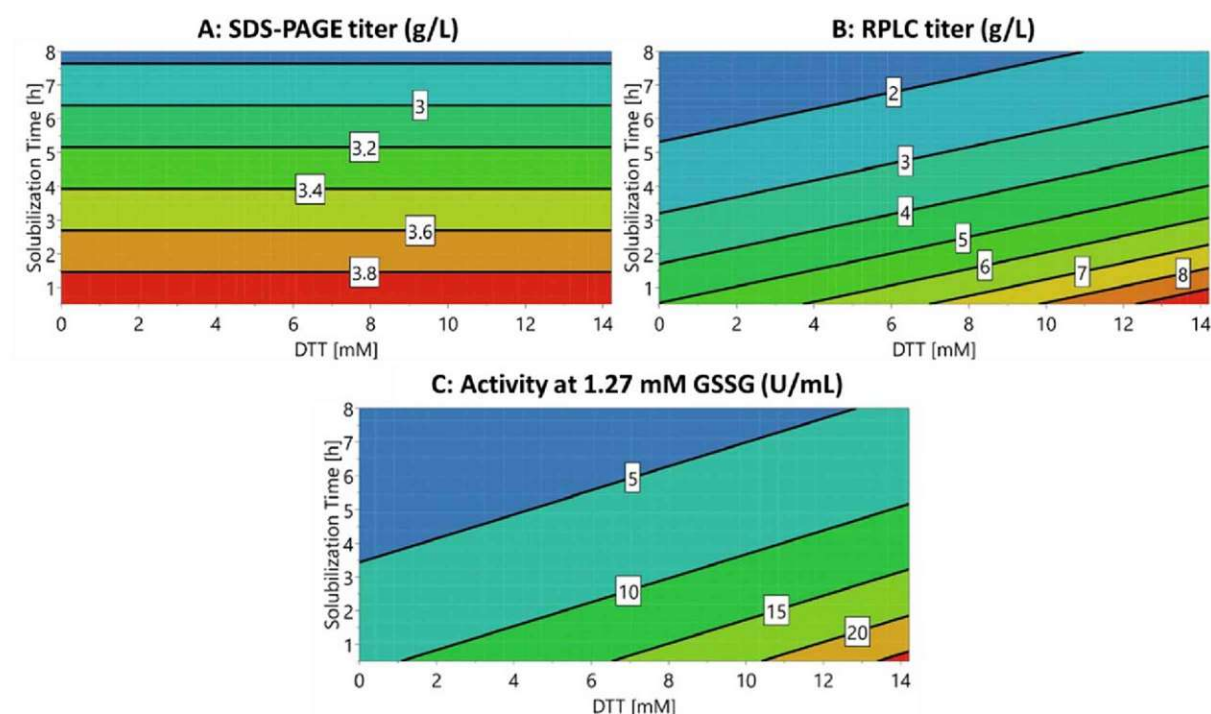


Figure 12: Response contour plots with the two factors DTT concentration and solubilization time. The following responses are shown: (A) Monomeric HRP concentration in the solubilize (g/L) analyzed using SDS-PAGE. (B) Monomeric HRP concentration in the solubilize (g/L) analyzed using RPLC. (C) Effect of the different solubilization conditions on the volumetric activity after refolding, which was performed at constant GSSG conditions of 1.27 mM. [156]



In a third step, the applicability of the method for IPC during solubilization was investigated. The protein concentration during refolding is often controlled by dissolving a fixed amount of IB-weight in solubilization buffer, followed by a set dilution in the refolding buffer. However, this approach requires consistent POI titers per IB weight and stable solubilization yields. It is therefore susceptible to variations either during the USP or the solubilization process. Negative effects of variations in protein concentration can be avoided by an applicable in-process monitoring of target protein concentration (see Figure 13). In this case, the method was tested using two different IB batches. Two refolding approaches were compared: a fixed 1:40 dilution ratio and a variable dilution based on the key quality attribute monomeric HRP concentration measured with the presented RPLC-method. The two batches of HRP IBs showed significant variation in monomeric HRP concentration at the chosen solubilization conditions, with Batch 1 containing around double the amount compared to Batch 2 (see Table 5). Unsurprisingly, when using the fixed dilution method, Batch 2 showed an over 50% decrease in enzymatic activity compared to Batch 1. However, by adjusting the dilution to 1:17 based on RPLC measurements, the variation between the batches was reduced from 54% to 11%, while simultaneously reducing refolding buffer volume by more than 50%. It was suspected that the remaining 11% variation between batches was caused by the higher concentration of DTT carry-over at lower dilutions.

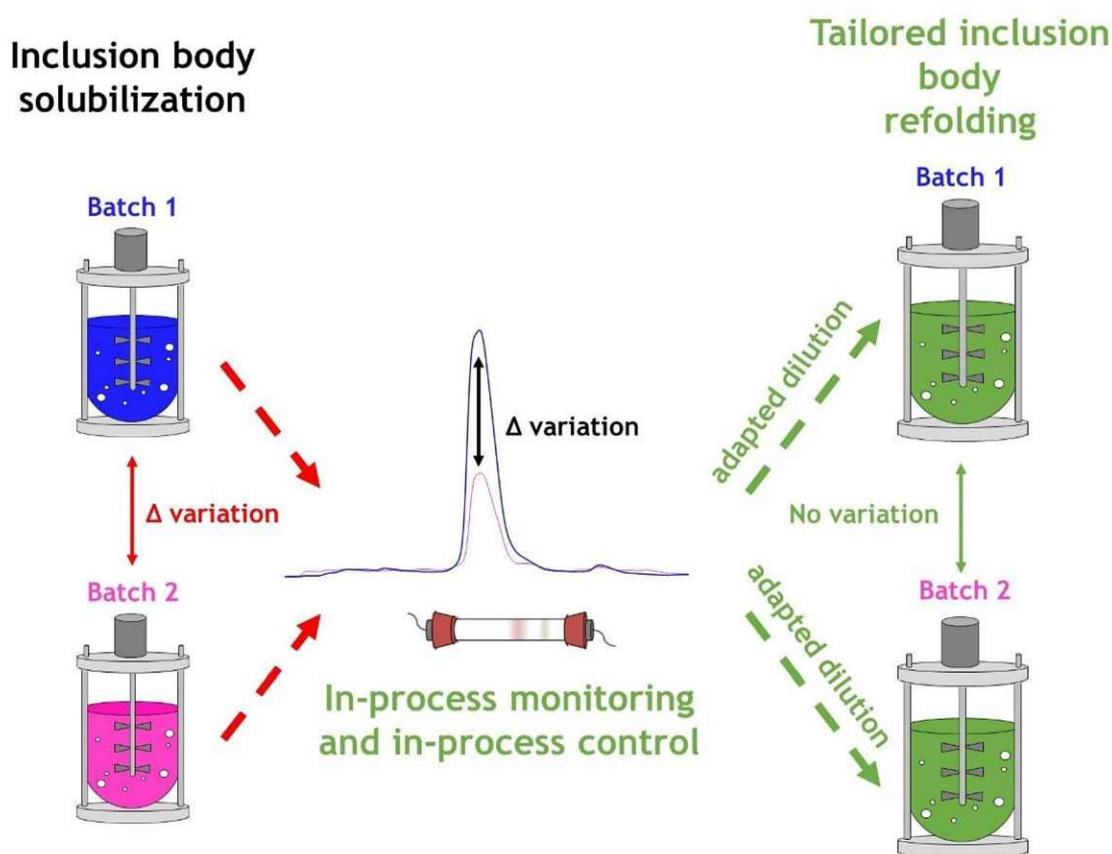


Figure 13: Schematic overview of the proposed in-process control strategy using a RPLC method to determine monomeric HRP concentration during solubilization in order to correct for concentration variations. [156]

Table 5: The concentration of HRP in the solubilize as determined via RPLC. The dilution in the refolding buffer as well as the activity after refolding are shown for: (1) IB Batch 1 with a fixed dilution (1:40); (2) Batch 2 with a fixed dilution (1:40); and (3) Batch 2 with an adapted dilution (1:17) based on IPC to achieve the same HRP concentration as for Batch 1. Adapted from [156].

IB Batch	c(HRP) (g/L) in Solubilization	Applied Dilution	Activity (U/mL) after Refolding
(1) Batch 1, Fixed Dilution	$5.27 \pm 0.11$	1:40	$89.7 \pm 6.0$
(2) Batch 2, Fixed Dilution	$2.35 \pm 0.05$	1:40	$41.9 \pm 2.8$
(3) Batch 2, IPC via RPLC	$2.35 \pm 0.05$	1:17	$79.8 \pm 5.4$

Overall, the RPLC method proved successful as an IPC tool for the HRP IB process, enabling at-line monitoring of deviations during solubilization and appropriate dilution adjustments during refolding. Additionally, the method could potentially provide valuable impurity monitoring through IB fingerprinting, adding usefulness for industrial applications.

Authors (1<sup>st</sup> paper): Julian Kopp, Florian B. Zauner, Andreas Pell, Johanna Hausjell, Diana Humer, Julian Ebner, Christoph Herwig, Oliver Spadiut, Christoph Slouka, Reinhard Pell

Published (1<sup>st</sup> paper): Journal of Pharmaceutical and Biomedical Analysis 188 (2020) 113412; <https://doi.org/10.1016/j.jpba.2020.113412>

Contribution (1<sup>st</sup> paper): JK and AP performed analytical experiments. FZ constructed the experimental robustness study. RP conducted DryLab experiments. JH, DH and JE produced and purified proteins and performed experiments. JK, FZ and RP evaluated experiments and performed data treatment. CH, OS and CS gave valuable scientific input. JK, FZ, CS and RP drafted the manuscript.

Authors (2<sup>nd</sup> paper): Julian Ebner, Diana Humer, Robert Klausser, Viktor Rubus, Reinhard Pell, Oliver Spadiut and Julian Kopp

Published (2<sup>nd</sup> paper): Bioengineering 2021, 8, 78. <https://doi.org/10.3390/bioengineering8060078>

Contribution (2<sup>nd</sup> paper): J.E., O.S. and J.K. founded the idea of this study and planned the experimental design. J.E. carried out the data treatment and the statistical analysis. J.E., D.H., R.K. and V.R. conducted the experiments. R.P. and J.K. developed the RPLC method. J.E., O.S. and J.K. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

#### Additional research:

- While the aim was to develop a broadly applicable method, we are aware of one POI (confidential industry cooperation) for which the method was not applicable, since samples showed significant tailing. We tried several adaptations of the method, however, the results were not satisfactory as the tailing behavior remained. No clear reason for this behavior could be identified. Therefore, while applicable for several investigated POIs, the presented method is not universally applicable.



## ii. IR for inline monitoring

Title (1<sup>st</sup> paper): QCL–IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography

Title (2<sup>nd</sup> paper): Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography

The successful application of IR as a PAT tool has been discussed and described for various processes and unit operations in the past. For the refolding process in particular, IR presents some apparent advantages. Through the measurement of the amid I and amid II bands, it provides direct information about the secondary structure, a protein property that changes during the refolding process, providing valuable insight for a quality attribute that is hard to access with many other analytical methods [157]. Additionally, IR is easy to integrate as an on-line or in-line measurement, providing real time information. This can either be achieved with ATR Probes or in transmission mode [158-161]. Lastly, the broad wavenumber range and highly distinctive absorption spectra of analytes allow for fingerprinting of complex mixtures [162-164]. However, one of the big disadvantages of IR spectroscopy is the high limit of detection, i.e. the high concentration of protein required in solution (5 – 10 g/L) [165, 166]. This results from the overlap of the absorption band of water with the amid I and amid II bands and makes measurements difficult especially for refolding steps where low protein concentrations are required, in particular for the classic batch dilution refolding approaches [143, 163, 167, 168].

Paper 1 describes the development of a novel in-line monitoring system for protein chromatography using quantum cascade laser-based mid-infrared (QCL-IR) spectroscopy. The system allows real-time measurement of protein secondary structure during chromatographic separation, which is a significant advancement over traditional UV detection (commonly UV 280 nm) methods that only measure total protein concentration. One of the significant challenges for this method was the overlap between protein absorption bands and changes in water absorption caused by the salt gradient commonly used in bind-elute chromatography. One option is to run an exact replicate of the gradient without protein present as a blank. While this is possible in a bench-scale set-up, it would require extensive additional material and time as well as extremely high precision in a production setting, making it technically and economically difficult. As an alternative, a novel background compensation method to compensate for the change in absorption was developed based on the monitored conductivity during the run. Three different approaches were shown: direct blank run subtraction (Case I) and two implementations of our novel reference spectra matrix method (Cases II and III). As shown in Figure 14, the background compensation methods successfully removed the interfering salt gradient effects, revealing clear protein spectra that matched well with reference measurements. This enables reliable measurement of protein secondary structure during the separation process.

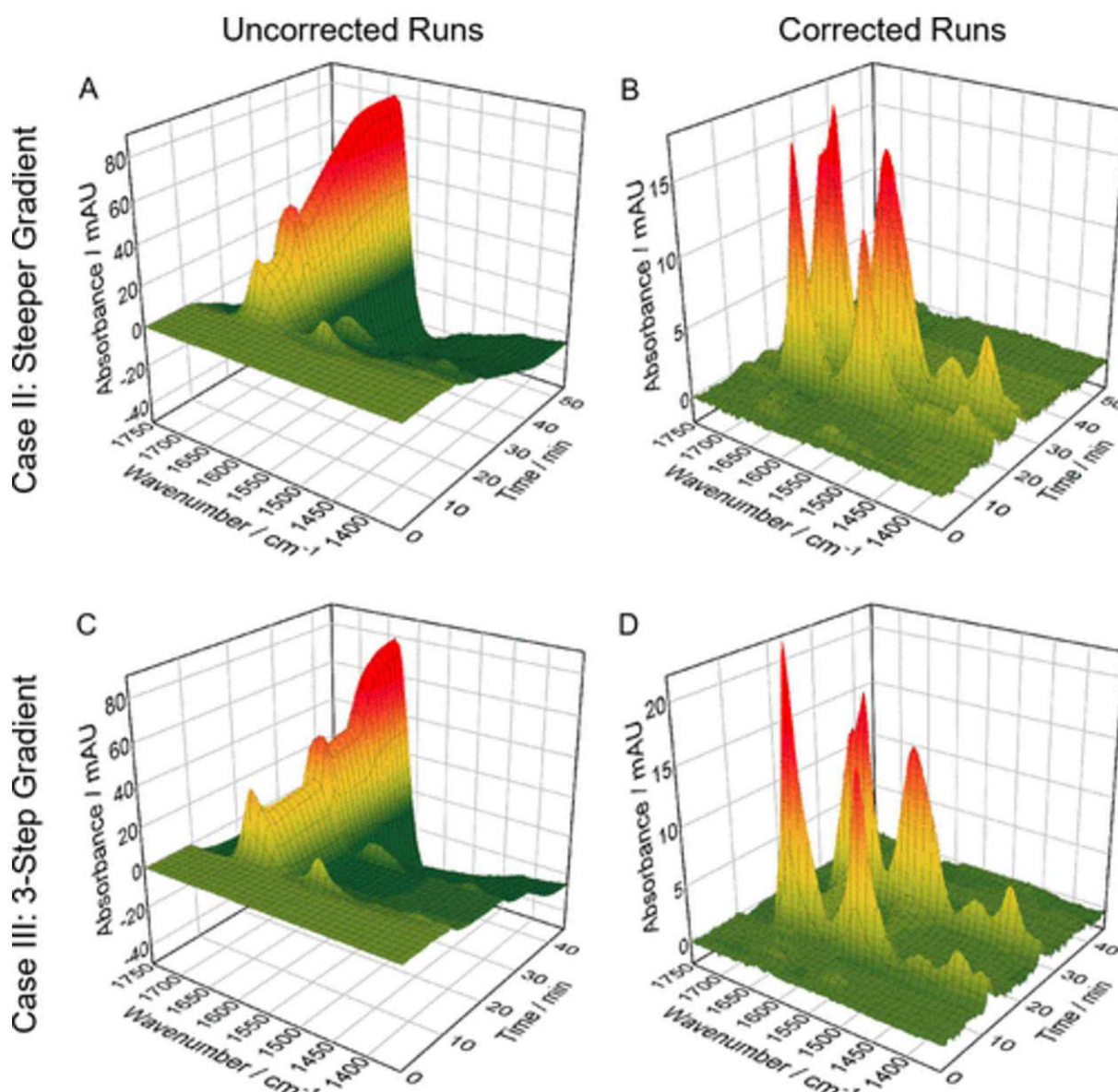


Figure 14: Spectral 3D plots for a linear gradient (Case II) and a step elution (Case III) using background compensation via reference spectra matrix method: (A,C) uncorrected sample runs and (B,D) background compensation corrected runs. [169]

As intended, the system showed effective protein identification through secondary structure analysis. The two model proteins used, namely hemoglobin and  $\beta$ -lactoglobulin, showed distinctly different spectral features in their amide I and II bands. Hemoglobin exhibited characteristic narrow bands at  $1656\text{ cm}^{-1}$  (amide I) and  $1545\text{ cm}^{-1}$  (amide II), while  $\beta$ -lactoglobulin showed broader bands at  $1632\text{ cm}^{-1}$  and  $1550\text{ cm}^{-1}$  with a shoulder at  $1680\text{ cm}^{-1}$ , consistent with their known structures [163, 170-173]. Comparing the in-line measurements to traditional offline analysis (Figure 15), the QCL-IR system provided equivalent quantitative results to UV detection. Simultaneously, it provided structural information in real-time. This can reduce or replace the need for time-consuming offline analysis methods like e.g. RPLC for protein identification. The system also showed excellent agreement with offline reference measurements, validating its accuracy.



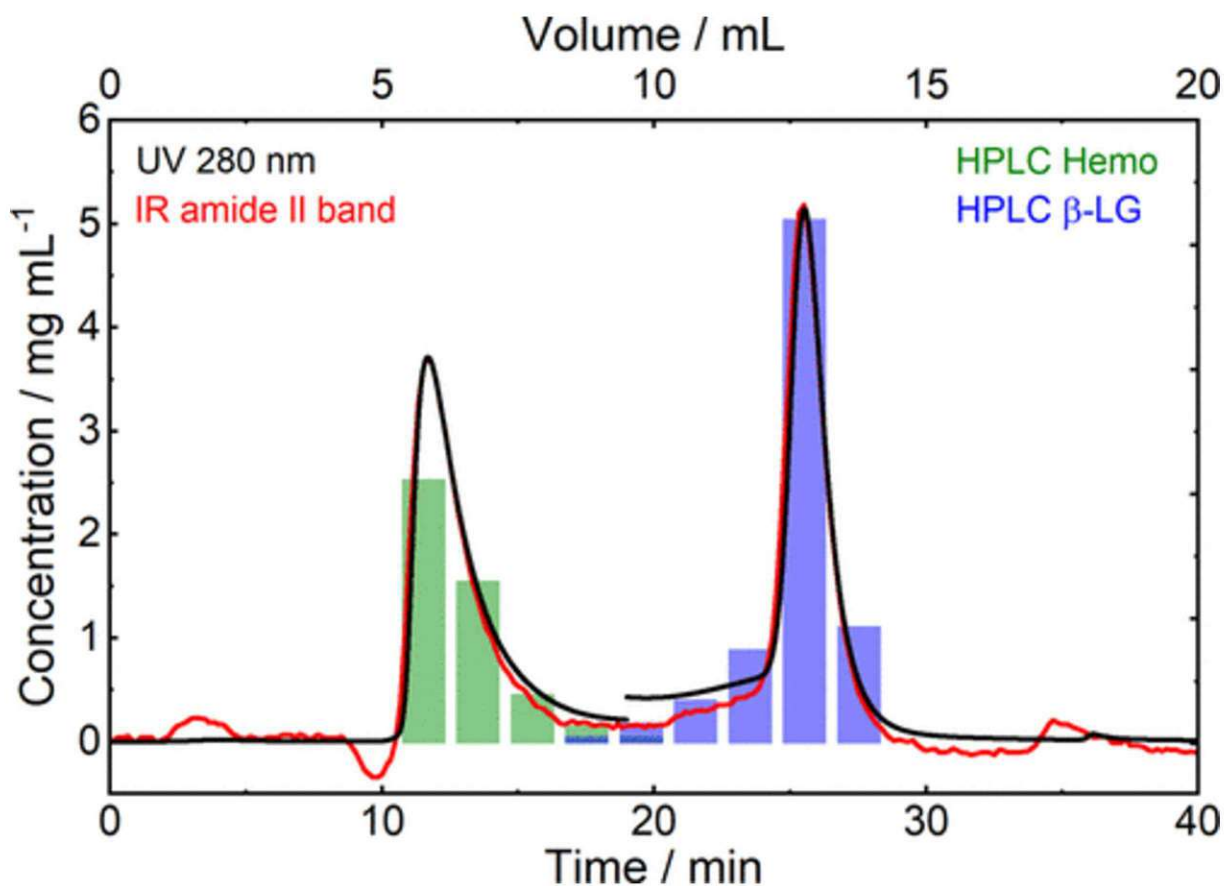


Figure 15: Comparison of protein concentrations obtained from mid-IR amide II band (red line) and UV detector signal at 280 nm (blackline) across a chromatographic run using a step elution gradient and protein reference concentrations obtained by measuring the collected fractions with reversed-phase HPLC (green and blue bars). [169]

Model proteins with well-characterized structures were used to develop the method and demonstrate the capability of the system. In a real-world application, this method would be particularly valuable for production processes in which maintaining correct protein structure is critical for product quality, as is the case e.g. in biopharmaceutical production processes [174, 175]. The ability to detect structural changes in real-time could allow immediate process adjustments, potentially improving product quality and reducing waste. An additional advantage of the method is the potential of analyzing impurities and quantifying co-eluting proteins, which is not possible using UV detection.

For the second paper, the described QCL-IR spectroscopy setup was used in combination with chemometric analysis for real-time monitoring of protein separation for SEC. As shown in Figure 16, the method generated three-dimensional data (wavenumber-time-absorbance) that captured the characteristic protein bands throughout the chromatographic separation. For resolving overlapping chromatographic peaks, different chemometric analysis were employed, particularly self-modeling mixture analysis (SMMA) and multivariate curve resolution (MCR) [176, 177]. These methods were applied for two case studies employing model proteins: one with three proteins (ovalbumin,  $\alpha$ -chymotrypsinogen A, and myoglobin) and one with four proteins showing severe peak overlap (HRP,  $\beta$ -lactoglobulin,  $\alpha$ -chymotrypsinogen A and myoglobin). The MCR analysis successfully resolved the overlapping peaks and provided both qualitative and quantitative information that matched well with conventional offline HPLC reference measurements.

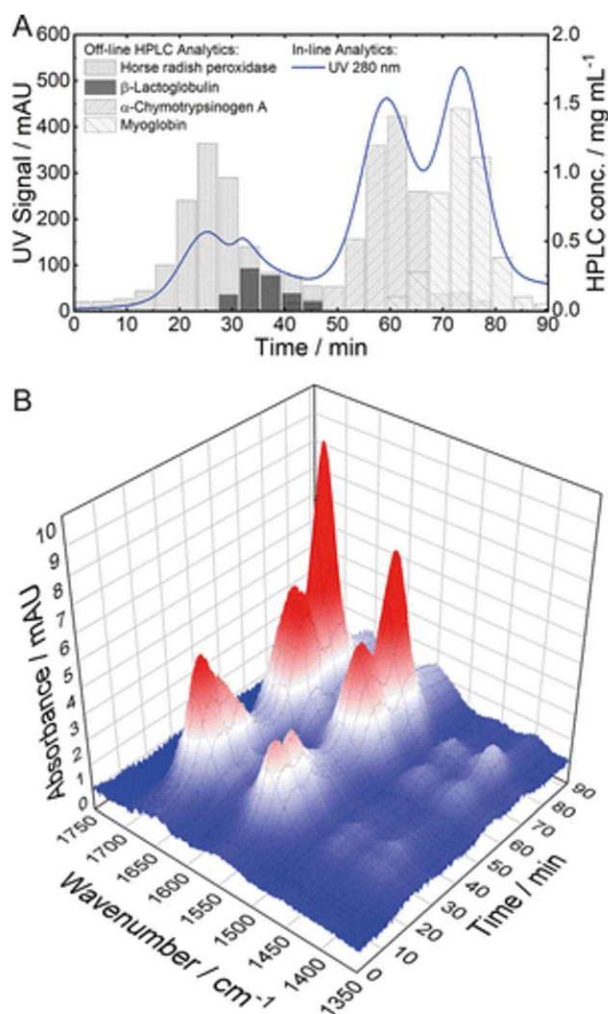


Figure 16: Experimental data obtained from the SEC run of case study II. (A) Results of in-line UV spectroscopy (left) and off-line HPLC analytics (right). (B) Spectral 3D plot recorded by the QCL-IR detector. [178]

Compared to existing methods, the presented approach offered several advantages. Traditional UV 280 nm detectors, while sensitive and robust, cannot discriminate between coeluting proteins or provide structural information [157]. Conventional FT-IR spectroscopy struggles with water absorption and requires complex solvent evaporation interfaces or very short path lengths, making it impractical for preparative LC applications. The QCL-IR method overcomes these limitations while providing real-time data on CQAs that typically require time-consuming offline analysis. In addition, the method was evaluated and compared to RPLC concerning its sustainability using the Analytical GREENness (AGREE) metric [179]. It was shown that the QCL-IR spectroscopy achieved a score of 0.84 compared to 0.43 for offline HPLC analysis (scale 0-1). This superior green chemistry performance stems from the fact that no additional solvent consumption as well as no additional sample preparation was needed, eliminating the need of additional analytical steps.

While SEC is often used as a polishing step in the DSP of soluble proteins, it was also used as a method for on-column refolding processes for several proteins [180, 181]. Initially designed with production processes in mind, the examples shown suggest that the presented method has significant potential for accelerating process development in addition to enabling real-time monitoring in production processes. The ability to provide information on protein structure and concentration almost in real time could make it particularly valuable for implementing Quality by Design (QbD) principles during process design and meeting PAT requirements in biopharmaceutical manufacturing.



Authors (1<sup>st</sup> paper): Julian Ebner\*, Christopher K. Akhgar\*, Mirta R. Alcaraz, Julian Kopp, Héctor Goicoechea, Oliver Spadiut, Andreas Schwaighofer, and Bernhard Lendl

\*Authors contributed equally to this work.

Published (1<sup>st</sup> paper): *Anal. Chem.* 2022, 94, 11192–11200;

<https://doi.org/10.1021/acs.analchem.2c01542>

Contribution (1<sup>st</sup> paper): The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript. C.K.A. and J.E. contributed equally.

Authors (2<sup>nd</sup> paper): Julian Ebner\*, Christopher K. Akhgar\*, Oliver Spadiut, Andreas Schwaighofer, and Bernhard Lendl

\*Authors contributed equally to this work.

Published (2<sup>nd</sup> paper): *Anal. Chem.* 2022, 94, 14, 5583–5590;

<https://doi.org/10.1021/acs.analchem.1c05191>

Contribution (2<sup>nd</sup> paper): The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. C.K.A. and J.E. contributed equally.

## 4 Conclusion and Outlook

The underlying goal of this thesis was to improve QbD strategies for IB processing. Table 6 shows several aspects required for an enhanced QbD approach, for which sections the approach was relevant and which specific examples are shown for the HRP process. Due to the complexity of biologics in general, QbD strategies are more challenging compared to small molecule drugs. For IBs in particular, this is further complicated by limited process understanding for IBs and their solubilization and refolding. An additional challenge is presented by a limited range of analytical tools available, restricting the development of PAT-tools. Therefore, many IB processes rely mainly on a traditional, minimal approach for pharmaceutical development. In order to address these challenges, two main parts, namely 1. Process development and 2. PAT tools were investigated.

In order to answer the first scientific question what is needed for the systematic development of a **reproducible and scalable** production process of active protein from IBs, the production process including the UOs IB isolation, solubilization, refolding, cofactor addition as well as capture and concentration were investigated based on the example of HRP IBs. A multivariate DoE approach was a key aspect to generate process understanding, especially in the case of the redox system. The developed workflow provides a tool to systematically investigate suitable redox conditions for refolding using a unit operation spanning DoE approach including the pH value. This tool was already successfully applied in an industry cooperation for a non-peroxidase POI and can provide a straight-forward protocol for future investigations of the redox system. For the cofactor addition, a hemin feeding strategy was developed in bench scale based on the process understanding gained from literature and small-scale DoEs. Such a combinatorial approach can potentially be applied for all POIs requiring cofactors, and, in particular, for hydrophobic cofactors facilitating aggregation if present at the start of the refolding.

Based on the developed HRP refolding process, the second scientific question “Based on the production process, which **PAT tools** can be developed for the **DSP UOs** of an **IB process**?” could be answered. A novel RPLC application for the at-line measurement of the POI during early DSP UOs was developed. The methods IPC capabilities were demonstrated for the solubilization step of HRP IBs. This enables the identification of different IB concentrations, allowing to adapt future refolding protocols to varying IB batches allowing for higher reproducibility in the face of upstream process variations. In the last part, potential applications for IR measurements during preparative chromatography steps were explored. While previously applied for analytical chromatography, the developed methods enabled the identification of secondary protein structures for both isocratic and bind-elute mode preparative chromatography.

Table 6: List of selected aspects of the enhanced, Quality by Design approach and their relevance for specific parts of this thesis. The relevance of certain aspects of the developed HRP refolding process in regards to QbD principles for IBs are highlighted for the relevant aspects.

Enhanced, Quality by Design Approach [71]	Relevance for the HRP IB process
Identifying <b>potential CQAs for intermediates</b> (in-process materials)	<b>All</b> - Specific activity [U/mg] of HRP and identification of monomeric HRP concentration during solubilization
Identifying, through e.g., prior knowledge, experimentation, and risk assessment, the material attributes and <b>process parameters</b> that can have an effect on product CQAs	<b>1.i IB isolation</b> - Effect of pressure and passages during HPH on refolding yield ( <i>DoE</i> ) <b>1.ii Redox System and 2.i RPLC for IPC</b> - Effect of redox system and pH ( <i>unit-operation spanning DoE</i> ) <b>1.iii Cofactor addition</b> - Effect of time and concentration of hemin addition during refolding ( <i>DoE</i> )



	Effect of salt concentration and stationary phase material ( <i>univariate and prior knowledge</i> )
	<b>1.i IB isolation</b> - <i>Factors</i> : Pressure and Passages; <i>Responses</i> : POI titer, purity, and activity
Multivariate experiments to understand product and process	<b>1.ii Redox System and 2.i RPLC for IPC</b> - <i>Factors</i> : DTT and GSSG conc., pH, solubilization time; <i>Responses</i> : Specific activity and monomeric HRP concentration during solubilization
	<b>1.iii Cofactor addition</b> - <i>Factors</i> : Time and concentration of hemin addition during refolding; <i>Response</i> : Specific activity
Establishment of <b>design space</b>	<b>1.ii Redox System</b> – Design space for DTT and GSSG concentration in combination with pH-value during solubilization and refolding
	<b>1.iii Cofactor addition</b> – Design space for time point and concentration of hemin addition
PAT tools	<b>2.i RPLC for IPC</b> - At-line RPLC for the measurement of monomeric HRP concentration during solubilization
	<b>2.ii IR for monitoring</b> - Demonstration of in-line measurement of secondary protein structures during chromatography (Model proteins)
Manufacturing process adjustable within design space	<b>1.ii Redox System</b> - The influence of the DTT/GSSG concentration on the refolding yield is reduced at pH 10
	<b>1.iii Cofactor addition</b> - Interaction of time and concentration of hemin addition allows higher concentration at later addition times

For **IB isolation** using HPH, the quality of IBs is often assessed by titer and purity of the POI. A protocol was established that allows to assess the influence of the process parameters pressure and passages not only on titer and purity, but also on the refolding yield in a multivariate approach. In addition, a straightforward sampling strategy was presented for the DoE.

A detailed investigation of the **Redox system** and its interplay during solubilization and refolding was performed, a **RPLC method was developed** and its capabilities for **IPC** was shown. As a key learning, a unit operation spanning DoE was established in order to define a suitable design space for the redox system. The interaction between the reducing agent in the solubilization mix and the oxidizing agent in the refolding mix added substantially to the process understanding. Furthermore, factors such as protein concentration and pH-value and their interaction with the redox system were described through a multivariate approach. Based on the process understanding generated through this approach for HRP and the use of a suitable RPLC method, the monomeric HRP concentration during solubilization could be identified as a CQA (of an intermediate). All of these results were combined to demonstrate the suitability of the RPLC as a PAT tool during solubilization, allowing IPC for different IB batches.

The final steps in the presented refolding process for HRP were **Cofactor addition** and a subsequent capture and concentration step. The influence of hemin on the refolding yield and the holo-HRP formation were investigated in a multivariate approach in small-scale first. Based on the established design space, a bench-scale set-up was used to access different addition modes (e.g. linear feed) of hemin. This resulted in an approach combining process understanding generated through both univariate (bench-scale) and multivariate (small-scale) experimentation with prior knowledge. In a last step, a capture and concentration step using HIC was investigated using a set of univariate experiments. In the case of HRP, excellent purities and yields could be achieved, showing that this step did not necessarily

require a multivariate approach. While this might vary for different IB processes, it demonstrated one of the biggest advantages of IBs, achieving high purities based on a high POI purity within the IBs and simple separation of IBs from soluble fractions.

As an analytical tool during process development and a potential PAT tool for preparative chromatography, **IR Spectroscopy for in-line monitoring** was applied for both an isocratic (SEC) as well as a bind-elute (IEX) mode. In order to demonstrate the full potential of the employed IR system, chromatography runs with model proteins were performed. It was shown that IR allowed to generate quantitative as well as qualitative information, allowing for compositional analysis of low-resolution fractions. For the investigated IEX system, a novel background compensation method was developed, omitting the need for precise replication blank runs.

Overall, the applicability of several QbD strategies were demonstrated for IB processes. While the definition of CQAs is still an integral part of these strategies, the high structural complexity and conformation-change of IBs through various UOs makes the identification of CQAs for intermediates challenging. Furthermore, identifying the material attributes and process parameters that can have an effect on product CQAs as well as establishing PAT tools requires significant process understanding. This can, at least partly, be generated through unit-operation spanning multivariate approaches and an iterative approach to process development requiring a “Design of Design of Experiments” concept. Compared to production processes of soluble proteins, several UOs for IB processes still lack broader knowledge and mechanistic understanding, which is also apparent by the lack of platform technologies.

#### 4.1 Outlook HRP process

One step missing for a complete production process of HRP is a potential strategy for formulation and an assessment of the storage stability. In the presented studies, the collected fractions of active HRP after HIC were stored at 4 °C. These resulted in around 1 g/L HRP in 1 M NaCl. While good stability for months was observed, this by no means presents a defined formulation and storage procedure. Besides a liquid formulation, which often requires the samples to be stored at low temperatures, lyophilization could be a feasible alternative. A commonly used method for the formulation and storage of enzymes, it can provide several advantages, such as increased long-term stability and increased shelf-life. In addition, lyophilized samples are often stable at 2-8 °C or even room temperature. With perspective of the QbD approach, additional CQAs specific for a freeze-dried product are appearance, moisture content and reconstitution time. In addition, CQAs defined for the liquid form are applicable after reconstitution of the product. Such CQAs could e.g. be appearance, potency, identity, and protein content. Commonly monitored CPPs are the shelf temperature during freezing, primary drying and secondary drying as well as the chamber pressure during primary and secondary drying [182]. In the case of recombinant HRP, prior experiments suggested that the lack of glycosylation might lead to decreased stability during freezing and drying. This could be counteracted by the addition of stabilizing agents in order to preserve enzyme activity during freezing and drying. Plant HRP is almost exclusively available in a freeze-dried form and often used in conjugation, therefore, lyophilization of the recombinant HRP would allow for a seamless integration in existing protocols.

As described in the additional research of part 1.iii, the recombinantly produced HRP was successfully conjugated to Protein L and used in an ELISA. Since HRP has been discussed as a potential candidate for targeted cancer treatment, both aspects could be combined. For antibody-directed enzyme prodrug therapy (ADEPT), HRP could be conjugated to antibodies or antibody fragments (e.g. Herceptine) specifically binding to cancer cells. *In vivo*, this conjugate is then specifically directed against the cancer cells exhibiting the antigen. In a subsequent step, an appropriate pro-drug (in the case of HRP e.g. indol-3-acetic acid) is administered, which is turned into the toxic drug substance specifically at the tumor sites. For this application, the conjugation process would have to be investigated in order to provide detailed information on the linkage mechanism, potential impurities introduced through the linkage



reaction, as well as the distribution of conjugated species [183]. In context of a QbD process, this would introduce several additional CQAs, such as homogeneity, purity, degree of conjugation, distribution profile, impurity profile of conjugation agents, and total protein concentration [183, 184]. While ample literature can be found on CQAs and analytical tools to assess these CQAs, little information is available on potential CPPs. However, some factors that have been shown to influence conjugation are e.g. molar ratio of antibody to HRP, the pH-value of the reaction and the concentration of the linker [185, 186]. Since conjugation often results in a heterogeneous distribution of different antibody to enzyme (HRP) ratios, SEC is a commonly employed UO, allowing for a combination of purification and buffer exchange [187, 188]. However, extensive off-line analytics are required for this UO, preventing the real-time release testing QbD principles aim for. This presents a strong potential future application of the IR Spectroscopy for in-line monitoring described in this thesis.

## 4.2 Outlook platforms

Several of the methods developed for HRP production in this thesis were applied to different IBs. In order to further develop these methods and move towards platform technologies for the process development of IB refolding, as a first step, proteins with similar properties could be considered. Based on the HRP properties and the developed process, investigation of different peroxidases might be a promising next step. Due to the different native sources of HRP (plant) and Lignin peroxidase (LiP) (fungi), they are assigned to different peroxidase classes, class III for HRP and class II for LiP. Despite their different classification, the protein properties of HRP and LiP are highly similar, as shown in Table 7. Both enzymes contain four disulfide bridges and two  $\text{Ca}^{2+}$  as well as requiring a heme-group to form the holo-enzyme (Table 7). However, due to an exposed tryptophan residue, LiP is able to oxidize bulky substrates such as Lignin [189-191]. While this changes the substrate specificity and application on a fundamental level, it can be assumed that the protein production process, and particularly the behavior during refolding, is similar for both, HRP and LiP. This is in good concordance with literature reporting refolding protocols for LiP, which are listed and compared to the HRP refolding protocol in Table 8. As for HRP, it is expected that a constant hemin feed, starting at a certain time during the refolding process (e.g. HRP: 8 h after refolding start) will increase refolding yields for LiP. Similar to the HRP refolding process, this strategy has, to our knowledge, not been reported for LiP in literature. Additionally, enzymatic activity of LiP can be measured using a modified ABTS assay, which is already established for HRP [120, 125]. As shown in Table 8, refolding protocols similar to the HRP processing method are not only reported for LiP but also for Manganese Peroxidase (MnP) and Versatile Peroxidase (VP). Therefore, if the production strategy proves to be applicable for LiP, it might further be extended to MnP and VP in the future. Both enzymes find applications in the fields of delignification and detoxification of waste streams [192-194].

Table 7: Comparison of HRP and LiP protein properties.

Property	HRP	LiP	HRP	LiP
Native source	Plant	Fungi	[101]	[191]
Class peroxidase	III	II	[101]	[191]
Size (glycosylated)	44 kDa	38-46 kDa	[101]	[118]
Size (non-glycosylated)	34 kDa	38 kDa	[101]	[118]
Isoelectric point	5.4	3.3-4.7	[195]	[191]
Disulfide bonds	4	4	[101]	[191]
Cofactors	2 $\text{Ca}^{2+}$	2 $\text{Ca}^{2+}$	[101]	[118]
Active center	Heme-group	Heme-group	[101]	[191]
Optimum pH of enzyme activity	5	3	[196]	[191]

Substrates	Diverse (aromatic phenols etc.)	Phenolic and non-phenolic substrates	[101]	[191]
Reaction mediators	-	Veratryl alcohol	[197]	[191]
Redox potential	950 mV	700-1400 mV	[198]	[199]
Macromolecular substrates?	“No net destructuring of three-dimensional network”	Yes (Lignin)	[197, 200]	[191, 197, 200]

Table 8: Comparison of HRP and LiP refolding conditions reported in literature

	pH	Redox system	Urea [M]	Hemin addition	Refolding	Ref. Yield	Reference
HRP C1A	10/10	DTT/GSSG	2	Delayed feed	Batch dil.	74%	[105]
LiP	8/9.5	DTT/GSSG	0.4	Batch	Batch dil.	2%	[122]
LiPH2	8/8	DTT/GSSG	0	Batch	Batch dil.	3.4 mg/L	[120]
LiPH2	8/8.3	DTT/GSSG	2.1	Batch	Batch dil.	2.4%	[125]
LiPH8	8/8	DTT/GSSG	0	Delayed Batch	Dialysis	n.a.	[201]
LiPH8	8/8.3	DTT/GSSG	2.1	Batch	Batch dil.	1%	[118]
rPr-MnP3	8/9.5	DTT/GSSG	0.15	Batch	Batch dil.	7%	[202]
MnP H4	8/8	DTT/GSSG	2	Batch	Batch dil.	n.a.	[203]
MnP	n.a./8.5	n.a./GSSG	1.5	Batch	Batch dil.	13%	[124]
VPs/MnP	9.5	DTT/GSSG	0.16	Batch	Batch dil.	3-28%	[204]
VP	8/9.5	DTT/GSSG	0.16	Batch	Batch dil.	n.a.	[205]

Based on the presented workflow for HRP IBs produced in *E. coli*, it is believed that several of the following problems could be addressed: Previously reported refolding protocols for LiP show significant similarities compared to HRP refolding (Table 8). However, low refolding yields and insufficient purification strategies after refolding have limited the application of these production strategies up to now. It is believed that the solution of these issues could be transferred from HRP to LiP. Furthermore, decreased stability of non-glycosylated LiP limits potential applications in industry [194]. PEGylation of HRP showed promising results in the past, as it significantly increased activity and stability. It is therefore believed that current problems in the production of LiP expressed in *E. coli* can be addressed and solved in a highly similar manner to HRP, leading to a robust and scalable production process for LiP, and in the next step, for peroxidases in general.

Besides the transfer of methods within one protein family, I believe that it will be essential to further investigate broadly applicable methods and platform technologies for IBs. Due to the highly diverse nature of different proteins forming IBs, I believe that these platform technologies will present toolkits to adapt existing processes rather than providing strict workflows. As a result of the wide variability, the perception of IBs as completely inactive protein aggregates has changed in recent years. Several POIs which form “active” IBs have been reported, meaning that some enzyme activity can already be observed in the IB state. For some of these active IBs, mild solubilization techniques are already showing promising results [36, 58, 206]. Through the use of novel solubilization techniques and agents such as e.g. high temperature and/or pressure, organic solvent extraction or addition of ionic liquids in the solubilization process, the structure of the single protein units in the IB can be kept intact, while the single protein units are brought into solution. These new solubilization approaches show that while the term IB was used for any protein that was not in solution after cell lysis, IBs are a broad species that exhibit very different properties [47]. Furthermore, taking into account that proteins adapted to remain correctly folded in vastly different environments (e.g. extremophiles), it is hard to see that a universal solubilization agent or refolding buffer, in the sense of a one size fits all approach, could be designed.



However, similar to soluble protein production, based on enhanced understanding of processes and the underlying phenomena, platform technologies with enhanced selection of suitable tools could very well be established.

Such platforms would provide large synergies with another approach that shows high potential for future applications in IB refolding: Continuous Processing. Many refolding processes today are still based on a batch dilution method because of its simple equipment requirements and handling. However, due to the low protein concentrations commonly required during refolding to reduce aggregation, continuous processes could provide significant reduction of volumes and equipment. Two previously reported options to implement continuous processing for the refolding step were tubular reactors and continuous Matrix-assisted refolding. The tubular reactor method provides some potential advantages such as: adaptable for the protein, good heat transfer, pulsed refolding, individual temperature and time profiles, higher productivity. As a potential advancement of the classic tubular reactor, microfluidic approaches might be able to further reduce protein aggregation during refolding due to precise mixing and small volumes. Similar systems have been established and are already in use for the formulation strategies for e.g. liposomes, demonstrating the at-scale capabilities [207-209]. Furthermore, the prior use could help with implementation in the field of protein refolding, especially for pharmaceutical processes. While advanced process control is one of the advantages of continuous processes, this in turn requires advanced analytical tools to allow for monitoring and control. In the ideal case, such tools provide real time information about CQAs and CPPs based on in-line measurements. However, the required analytical tools also present a challenge for IB processes. Most important process parameters (protein concentration, redox system, pH-value) for batch dilution refolding are determined at the start of the refolding and cannot be reasonably changed once the process is started. Therefore, the need and benefit of PAT tools was limited for the refolding step. In combination with low protein concentrations, only a few in-line analytics providing real time information have been investigated in the past. However, recent years showed an increase in both the use of continuous refolding and corresponding monitoring and control strategies facilitating analytical methods [210-213]. As one example, traditional ATR-IR methods are not applicable for classic batch dilution refolding due to low protein concentrations. However, the development of QCL-IR has lowered detection limits to protein concentrations of 0.5 mg/mL [165]. At the same time, higher protein concentrations during refolding can be achieved by using methods as e.g. continuous Matrix-assisted refolding. This combination could allow the use of flow QCL-IR flow cells (as was presented in section 2.ii) as a PAT tool. In context of QbD, such approaches could pave the way for the demonstrations of controlled processes, scientific understanding, and real time release testing of refolding processes in the future.

In order to fully utilize the described platforms, a digital twin can be developed, fully utilizing the collected data and process knowledge. The concept of a digital model mirroring a real-world product, system or process has been around for several centuries but has seen increased attention in the biotechnology space due to enhanced data collection and processing capacities. A digital twin consists of three distinct parts: 1. the physical object or environment, 2. the digital model and 3. the link and bidirectional interaction between the two. Part 3, the link and bidirectional interaction between the physical and digital twin is a key requirement, while approaches only modeling the physical part (but not interacting with it) are commonly called digital shadows [214]. The advantage of a full digital twin is that data from process development and manufacturing is combined, naturally aligning its goals with those of QbD principles. Based on a scalable, asset-centric data foundation including both development and manufacturing data, asset connectivity, statistical analytics, and mechanistic modeling in one environment. As a result, costs during process development are reduced, process robustness is increased, and the effort for bioprocess validation is reduced. Besides the described requirements for data collection and processing, current challenges of IB processes for the establishment of digital twins are limited analytic methods resulting in scarce data as well as low mechanistic and process understanding. Both batch dilution refolding as well as continuous processing immensely benefit from digital shadows, i.e. a model which is fed by a one-way data flow from the physical process. However, the limitations of



batch dilution refolding to change CPPs once the process is started do not negate but restrict the benefits gained by employing a digital twin. On the other hand, continuous processing allows for the adaption of CPPs also during the process, unlocking the full potential of a digital twin. Therefore, future research for the combination of continuous processing and digital twins shows a high potential for IB process development and production processes, especially with regards to regulatory requirements (e.g. ICH Q8).

### 4.3 Outlook IBs in general

The discovery of active IBs has introduced the concept of using these IBs directly, omitting the expensive steps of refolding and extensive purification of the target protein. Preparations of active IBs can be produced with minimal DSP efforts (cell lysis and IB separation via centrifugation or filtration), yielding relatively high purities compared to similar approaches for soluble proteins. However, the quality profile of these IB preparations would not be sufficient for pharmaceutical applications, but more likely closer to proteins currently used for white and grey biotechnology (e.g. large chemical processes). For many of these applications, enzyme immobilization has been investigated for ecological and economic reasons, allowing retaining of the catalyst (enzyme) and / or continuous processing. Therefore, active IBs as carrier-free immobilized enzymes provide a significant advantage compared to soluble proteins which are immobilized in an additional step [48]. This leads to protein engineering methods using short peptide tags or aggregation-inducing protein domains in order to switch the production of soluble proteins to IBs. Since milder biological methods are replacing harsher chemical processes in a variety of fields, demand for customized active IBs as catalysts could rise in the future.

For “classical” inactive IBs, a main field of application is thought to remain in pharmaceutical production processes due to relatively high costs associated. Certain aspects of personalized medicine in the future could show high synergies with IB processes. The term personalized medicine describes an approach in which people are separated into different groups and treatments are tailored to these groups, compared to a “one-drug-fits-all” approach. As an example, Herceptin (a monoclonal antibody binding to the HER2/neu receptor) is only used in the treatment of patients if the cancer is tested for over-expression of the HER2/neu receptor. Through (future) advances in the understanding of the human genome, analytics and data process as well as confirming fundamental biology and proteins, more effective treatments can be specifically tailored to fit a smaller group or even a single individual. As a conclusion, smaller batches of a drug would be required, asking for enhanced flexibility of production, and presenting challenges and new approaches to regulatory requirements and procedures. If only small changes are required of a protein (e.g. the adaption of single amino acids at the paratope in the case of antibodies or antibody-fragments), the straightforward and cheap cloning and expression in *E. coli* or other bacterial expression systems would provide large advantages for such approaches. However, in the case of antibodies or antibody-fragments, the expression would most likely lead to the formation of IBs. Under the assumption that according platform technologies are available, IBs might therefore provide significant economic advantages for the timely production of tailored drugs in the future.

Such a tailored drug production process would most likely show significant synergies with *in-silico* methods, a field of protein folding studies that gained attention especially with recent developments in the field of AI. The understanding and modeling of protein folding based on the primary structure has long been a topic in the scientific community, with different approaches developed throughout the decades. Several approaches, e.g. AlphaFold used and benefited from advancements of processing power and AI, resulting in unprecedented performances of protein structure predictions. The developments in *in-silico* possibilities of protein prediction and design could provide significant improvements not only for the understanding of protein structure, function and folding but also for QbD aspects of recombinant production processes. As stated in ICH Q8: “It is important to recognize that quality cannot be tested into products; i.e., quality should be built in by design.” (Direct quote from [71]). Enhanced *in-silico* methods could provide protein design methods able to rationally design whole



proteins to tailor them for function and behavior. In combination with scientific process understanding, a specific refolding behavior could also be designed *in-silico*, significantly reducing process costs. While this is, at current times, still a hypothetical and definitely not a certainty, it would highlight the advantages of IBs (high purity and easy separation from HCPs) while negating some of the disadvantages (cumbersome and complex refolding).

## 5 References

1. Cohen, S.N., et al., *Construction of biologically functional bacterial plasmids in vitro*. Proc Natl Acad Sci U S A, 1973. **70**(11): p. 3240-4.
2. Baeshen, N.A., et al., *Cell factories for insulin production*. Microb Cell Fact, 2014. **13**: p. 141.
3. Itakura, K., et al., *Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin*. Science, 1977. **198**(4321): p. 1056-63.
4. Schutz, A., et al., *A concise guide to choosing suitable gene expression systems for recombinant protein production*. STAR Protoc, 2023. **4**(4): p. 102572.
5. Matthews, B.J. and L.B. Voshall, *How to turn an organism into a model organism in 10 'easy' steps*. J Exp Biol, 2020. **223**(Pt Suppl 1).
6. Wurm, F.M., *Production of recombinant protein therapeutics in cultivated mammalian cells*. Nat Biotechnol, 2004. **22**(11): p. 1393-8.
7. Guzman-Herrador, D.L., et al., *DNA Delivery and Genomic Integration into Mammalian Target Cells through Type IV A and B Secretion Systems of Human Pathogens*. Front Microbiol, 2017. **8**: p. 1503.
8. Gu, P., et al., *A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies*. Sci Rep, 2015. **5**: p. 9684.
9. Wang, C., et al., *Comparison of Genome and Plasmid-Based Engineering of Multigene Benzylglucosinolate Pathway in Saccharomyces cerevisiae*. Appl Environ Microbiol, 2022. **88**(22): p. e0097822.
10. Macek, B., et al., *Protein post-translational modifications in bacteria*. Nat Rev Microbiol, 2019. **17**(11): p. 651-664.
11. Siew, Y.Y. and W. Zhang, *Downstream processing of recombinant human insulin and its analogues production from E. coli inclusion bodies*. Bioresour Bioprocess, 2021. **8**(1): p. 65.
12. Garcia-Fruitos, E., *Inclusion bodies: a new concept*. Microb Cell Fact, 2010. **9**: p. 80.
13. Slouka, C., et al., *Perspectives of inclusion bodies for bio-based products: curse or blessing?* Appl Microbiol Biotechnol, 2019. **103**(3): p. 1143-1153.
14. Levinthal, C., *How to fold graciously*. Mossbauer spectroscopy in biological systems, 1969. **67**: p. 22-24.
15. Durup, J., *On "Levinthal paradox" and the theory of protein folding*. Journal of Molecular Structure: THEOCHEM, 1998. **424**(1-2): p. 157-169.
16. Lowe, J. and L.A. Amos, *Evolution of cytomotive filaments: the cytoskeleton from prokaryotes to eukaryotes*. Int J Biochem Cell Biol, 2009. **41**(2): p. 323-9.
17. Giardi, M.T. and E. Pace, *Photosynthetic proteins for technological applications*. Trends Biotechnol, 2005. **23**(5): p. 257-63.
18. Rose, G.D., et al., *A backbone-based theory of protein folding*. Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16623-33.
19. Amani, S. and A. Naeem, *Understanding protein folding from globular to amyloid state*. Process Biochemistry, 2013. **48**(11): p. 1651-1664.
20. Kiefhaber, T., et al., *Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation*. Biotechnology (N Y), 1991. **9**(9): p. 825-9.
21. Hingorani, K.S. and L.M. Gierasch, *Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge*. Curr Opin Struct Biol, 2014. **24**: p. 81-90.
22. Clark, P.L., *Protein folding in the cell: reshaping the folding funnel*. Trends Biochem Sci, 2004. **29**(10): p. 527-34.
23. Onuchic, J.N. and P.G. Wolynes, *Theory of protein folding*. Curr Opin Struct Biol, 2004. **14**(1): p. 70-5.
24. Silva, J.L., et al., *Ligand binding and hydration in protein misfolding: insights from studies of prion and p53 tumor suppressor proteins*. Acc Chem Res, 2010. **43**(2): p. 271-9.
25. Anfinsen, C.B., et al., *The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain*. Proc Natl Acad Sci U S A, 1961. **47**(9): p. 1309-14.
26. Masson, P. and S. Lushchekina, *Conformational Stability and Denaturation Processes of Proteins Investigated by Electrophoresis under Extreme Conditions*. Molecules, 2022. **27**(20).



27. Sanchez-Ruiz, J.M., *Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry*. Biophys J, 1992. **61**(4): p. 921-35.
28. Guo, L., et al., *A cellular system that degrades misfolded proteins and protects against neurodegeneration*. Mol Cell, 2014. **55**(1): p. 15-30.
29. Zapun, A., et al., *Protein folding in a specialized compartment: the endoplasmic reticulum*. Structure, 1999. **7**(8): p. R173-82.
30. de Marco, A., et al., *Bacterial inclusion bodies are industrially exploitable amyloids*. FEMS Microbiol Rev, 2019. **43**(1): p. 53-72.
31. Bukau, B., J. Weissman, and A. Horwich, *Molecular Chaperones and Protein Quality Control*. Cell, 2006. **125**(3): p. 443-451.
32. Pettersen, E.F., et al., *UCSF ChimeraX: Structure visualization for researchers, educators, and developers*. Protein Sci, 2021. **30**(1): p. 70-82.
33. Jevsevar, S., et al., *Production of nonclassical inclusion bodies from which correctly folded protein can be extracted*. Biotechnol Prog, 2005. **21**(2): p. 632-9.
34. Clark, D.P. and N.J. Pazdernik, *Chapter 10 - Recombinant Proteins*, in *Biotechnology (Second Edition)*, D.P. Clark and N.J. Pazdernik, Editors. 2016, Academic Cell: Boston. p. 335-363.
35. Bhatwa, A., et al., *Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications*. Front Bioeng Biotechnol, 2021. **9**: p. 630551.
36. Singh, A., et al., *Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process*. Microb Cell Fact, 2015. **14**: p. 41.
37. Benham, A.M., *Protein folding and disulfide bond formation in the eukaryotic cell: meeting report based on the presentations at the European Network Meeting on Protein Folding and Disulfide Bond Formation 2009 (Elsinore, Denmark)*. FEBS J, 2009. **276**(23): p. 6905-11.
38. Robinson, P.J. and N.J. Bulleid, *Mechanisms of Disulfide Bond Formation in Nascent Polypeptides Entering the Secretory Pathway*. Cells, 2020. **9**(9).
39. Gutierrez, T. and T. Simmen, *Endoplasmic reticulum chaperones and oxidoreductases: critical regulators of tumor cell survival and immunorecognition*. Front Oncol, 2014. **4**: p. 291.
40. Derman, A.I. and J. Beckwith, *Escherichia coli alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm*. J Bacteriol, 1991. **173**(23): p. 7719-22.
41. Prinz, W.A., et al., *The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm*. J Biol Chem, 1997. **272**(25): p. 15661-7.
42. Singh, S.M. and A.K. Panda, *Solubilization and refolding of bacterial inclusion body proteins*. J Biosci Bioeng, 2005. **99**(4): p. 303-10.
43. Hoffmann, D., et al., *Reassessment of inclusion body-based production as a versatile opportunity for difficult-to-express recombinant proteins*. Crit Rev Biotechnol, 2018. **38**(5): p. 729-744.
44. Margreiter, G., et al., *Impact of different cultivation and induction regimes on the structure of cytosolic inclusion bodies of TEM1-beta-lactamase*. Biotechnol J, 2008. **3**(9-10): p. 1245-55.
45. Bodi, A., et al., *Structural determinants of the half-life and cleavage site preference in the autolytic inactivation of chymotrypsin*. Eur J Biochem, 2001. **268**(23): p. 6238-46.
46. Talafova, K., et al., *Bacterial inclusion bodies as potential synthetic devices for pathogen recognition and a therapeutic substance release*. Microb Cell Fact, 2013. **12**: p. 16.
47. Carrio, M.M., R. Cubarsi, and A. Villaverde, *Fine architecture of bacterial inclusion bodies*. FEBS Lett, 2000. **471**(1): p. 7-11.
48. Jager, V.D., et al., *Catalytically-active inclusion bodies for biotechnology-general concepts, optimization, and application*. Appl Microbiol Biotechnol, 2020. **104**(17): p. 7313-7329.
49. Ferrer-Miralles, N. and A. Villaverde, *Bacterial cell factories for recombinant protein production; expanding the catalogue*. Microbial Cell Factories, 2013. **12**(1).
50. Tripathi, N.K. and A. Shrivastava, *Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development*. Front Bioeng Biotechnol, 2019. **7**: p. 420.
51. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges*. Front Microbiol, 2014. **5**: p. 172.

52. Rosano, G.L., E.S. Morales, and E.A. Ceccarelli, *New tools for recombinant protein production in Escherichia coli: A 5-year update*. Protein Sci, 2019. **28**(8): p. 1412-1422.
53. Inguva, P., S. Grasselli, and P.W.S. Heng, *High pressure homogenization – An update on its usage and understanding*. Chemical Engineering Research and Design, 2024. **202**: p. 284-302.
54. Anand, H., et al., *The effect of chemical pretreatment combined with mechanical disruption on the extent of disruption and release of intracellular protein from E. coli*. Biochemical Engineering Journal, 2007. **35**(2): p. 166-173.
55. Rudolph, R. and H. Lilie, *In vitro folding of inclusion body proteins*. The FASEB Journal, 1996. **10**(1): p. 49-56.
56. Fischer, B., et al., *A novel sequential procedure to enhance the renaturation of recombinant protein from Escherichia coli inclusion bodies*. Protein Eng, 1992. **5**(6): p. 593-6.
57. Hatefi, Y. and W.G. Hanstein, *Solubilization of particulate proteins and nonelectrolytes by chaotropic agents*. Proc Natl Acad Sci U S A, 1969. **62**(4): p. 1129-36.
58. Klausser, R., et al., *State-of-the-art and novel approaches to mild solubilization of inclusion bodies*. Front Bioeng Biotechnol, 2023. **11**: p. 1249196.
59. Kumar, D., *Chromatographic Protein Refolding/Renaturation*, in *Reference Module in Life Sciences*. 2018, Elsevier.
60. Yamaguchi, H. and M. Miyazaki, *Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies*. Biomolecules, 2014. **4**(1): p. 235-51.
61. Foroutannejad, S., et al., *The cofactor-dependent folding mechanism of Drosophila cryptochrome revealed by single-molecule pulling experiments*. Nat Commun, 2023. **14**(1): p. 1057.
62. Chen, C. and C. Park, *Chaperone action of a cofactor in protein folding*. Protein Sci, 2020. **29**(7): p. 1667-1678.
63. Bushmarina, N.A., et al., *Cofactor effects on the protein folding reaction: acceleration of alpha-lactalbumin refolding by metal ions*. Protein Sci, 2006. **15**(4): p. 659-71.
64. Lin, W.J. and J.A. Traugh, *Renaturation of casein kinase II from recombinant subunits produced in Escherichia coli: purification and characterization of the reconstituted holoenzyme*. Protein Expr Purif, 1993. **4**(3): p. 256-64.
65. Vallejo, L.F. and U. Rinas, *Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins*. Microb Cell Fact, 2004. **3**(1): p. 11.
66. McCormick, K. and J.H. Sanders, *Chapter 2 - Good manufacturing practice*, in *Quality (Second Edition)*, K. McCormick and J.H. Sanders, Editors. 2022, Butterworth-Heinemann. p. 43-63.
67. McCormick, K. and J.H. Sanders, *Chapter 1 - Product life cycle*, in *Quality (Second Edition)*, K. McCormick and J.H. Sanders, Editors. 2022, Butterworth-Heinemann. p. 1-42.
68. Bondi, R.W. and J.K. Drennen, *5 - Quality by Design and the Importance of PAT in QbD*, in *Separation Science and Technology*, S. Ahuja and S. Scypinski, Editors. 2011, Academic Press. p. 195-224.
69. Juran, J.M., *Juran on quality by design: the new steps for planning quality into goods and services*. 1992: Simon and Schuster.
70. Sarwar Beg, M.S.H., Mahfoozur Rahman, Suryakanta Swain, *Chapter 1 - Introduction to Quality by Design (QbD): Fundamentals, Principles, and Applications*. 2019.
71. Agency, E.M., *ICH guideline Q8 (R2) on pharmaceutical development Step 5*. 2017.
72. Ter Horst, J.P., et al., *Implementation of Quality by Design (QbD) Principles in Regulatory Dossiers of Medicinal Products in the European Union (EU) Between 2014 and 2019*. Ther Innov Regul Sci, 2021. **55**(3): p. 583-590.
73. Rathore, A.S., et al., *Chapter 48 - Implementation of QbD for Manufacturing of Biologics—Has It Met the Expectations?*, in *Biopharmaceutical Processing*, G. Jagschies, et al., Editors. 2018, Elsevier. p. 1051-1073.
74. Kumar, V., A. Bhalla, and A.S. Rathore, *Design of experiments applications in bioprocessing: concepts and approach*. Biotechnol Prog, 2014. **30**(1): p. 86-99.
75. Veitch, N.C., *Horseradish peroxidase: a modern view of a classic enzyme*. Phytochemistry, 2004. **65**(3): p. 249-59.



76. Samuni, A., E. Maimon, and S. Goldstein, *Mechanism of HRP-catalyzed nitrite oxidation by H(2)O(2) revisited: Effect of nitroxides on enzyme inactivation and its catalytic activity*. Free Radic Biol Med, 2017. **108**: p. 832-839.
77. Chance, B., *The kinetics and stoichiometry of the transition from the primary to the secondary peroxidase peroxide complexes*. Arch Biochem Biophys, 1952. **41**(2): p. 416-24.
78. George, P., *Chemical nature of the secondary hydrogen peroxide compound formed by cytochrome-c peroxidase and horseradish peroxidase*. Nature, 1952. **169**(4302): p. 612-3.
79. Welinder, K.G., *Covalent structure of the glycoprotein horseradish peroxidase (EC 1.11.1.7)*. FEBS Lett, 1976. **72**(1): p. 19-23.
80. Yakovleva, J., et al., *Microfluidic enzyme immunoassay using silicon microchip with immobilized antibodies and chemiluminescence detection*. Anal Chem, 2002. **74**(13): p. 2994-3004.
81. Moody, M.D., et al., *Array-based ELISAs for high-throughput analysis of human cytokines*. Biotechniques, 2001. **31**(1): p. 186-90, 192-4.
82. Tanke, H.J., et al., *Use of Horseradish Peroxidase- and Fluorescein-modified Cisplatin Derivatives for Simultaneous Labeling of Nucleic Acids and Proteins*. Clinical Chemistry, 2002. **48**(8): p. 1352-1359.
83. Vasileva, N., et al., *Application of immobilized horseradish peroxidase onto modified acrylonitrile copolymer membrane in removing of phenol from water*. Int J Biol Macromol, 2009. **44**(2): p. 190-4.
84. Tatsumi, K., S. Wada, and H. Ichikawa, *Removal of chlorophenols from wastewater by immobilized horseradish peroxidase*. Biotechnology and Bioengineering, 2000. **51**(1): p. 126-130.
85. Kalra, B. and R.A. Gross, *Horseradish Peroxidase Mediated Free Radical Polymerization of Methyl Methacrylate*. Biomacromolecules, 2000. **1**(3): p. 501-505.
86. Singh, A., D. Ma, and D.L. Kaplan, *Enzyme-Mediated Free Radical Polymerization of Styrene*. Biomacromolecules, 2000. **1**(4): p. 592-596.
87. Rodrigues, A.P., et al., *Oxidation of acetylacetone catalyzed by horseradish peroxidase in the absence of hydrogen peroxide*. Biochim Biophys Acta, 2006. **1760**(12): p. 1755-61.
88. Wardman, P., *Indole-3-acetic acids and horseradish peroxidase: a new prodrug/enzyme combination for targeted cancer therapy*. Curr Pharm Des, 2002. **8**(15): p. 1363-74.
89. Bae, J.W., et al., *Horseradish peroxidase-catalysed in situ-forming hydrogels for tissue-engineering applications*. J Tissue Eng Regen Med, 2015. **9**(11): p. 1225-32.
90. van der Weiden, G.S., et al., *Dextran-tryamine hydrogel maintains position and integrity under simulated loading in a human cadaver knee model*. Osteoarthritis Cartilage, 2024. **32**(3): p. 100492.
91. Yang, B.Y., J.S. Gray, and R. Montgomery, *The glycans of horseradish peroxidase*. Carbohydr Res, 1996. **287**(2): p. 203-12.
92. Takahashi, N., et al., *New N-glycans in horseradish peroxidase*. Anal Biochem, 1998. **255**(2): p. 183-7.
93. Delobel, A., *Glycosylation of Therapeutic Proteins: A Critical Quality Attribute*. Methods Mol Biol, 2021. **2271**: p. 1-21.
94. Raju, S., *Glycosylation variations with expression systems*. BioProcess International, 2003. **1**: p. 44-53.
95. Kuriakose, A., N. Chirmule, and P. Nair, *Immunogenicity of Biotherapeutics: Causes and Association with Posttranslational Modifications*. J Immunol Res, 2016. **2016**: p. 1298473.
96. Pappa, H.S. and A.E. Cass, *A step towards understanding the folding mechanism of horseradish peroxidase. Tryptophan fluorescence and circular dichroism equilibrium studies*. Eur J Biochem, 1993. **212**(1): p. 227-35.
97. Chattopadhyay, K. and S. Mazumdar, *Structural and conformational stability of horseradish peroxidase: effect of temperature and pH*. Biochemistry, 2000. **39**(1): p. 263-70.
98. Skulj, S., et al., *Effect of N-glycosylation on horseradish peroxidase structural and dynamical properties*. Comput Struct Biotechnol J, 2022. **20**: p. 3096-3105.
99. Capone, S., et al., *Glyco-variant library of the versatile enzyme horseradish peroxidase*. Glycobiology, 2014. **24**(9): p. 852-63.

100. Spadiut, O., et al., *Purification of a recombinant plant peroxidase produced in Pichia pastoris by a simple 2-step strategy*. Protein Expr Purif, 2012. **86**(2): p. 89-97.
101. Spadiut, O. and C. Herwig, *Production and purification of the multifunctional enzyme horseradish peroxidase*. Pharm Bioprocess, 2013. **1**(3): p. 283-295.
102. Capone, S., et al., *Combining Protein and Strain Engineering for the Production of Glyco-Engineered Horseradish Peroxidase C1A in Pichia pastoris*. Int J Mol Sci, 2015. **16**(10): p. 23127-42.
103. Gundinger, T. and O. Spadiut, *A comparative approach to recombinantly produce the plant enzyme horseradish peroxidase in Escherichia coli*. J Biotechnol, 2017. **248**: p. 15-24.
104. Ebner, J., D. Humer, and O. Spadiut, *Von der Wurzel ins Labor: Meerrettichperoxidase produziert in E. coli*. BIOSpektrum, 2021. **27**(7): p. 773-775.
105. Humer, D., J. Ebner, and O. Spadiut, *Scalable High-Performance Production of Recombinant Horseradish Peroxidase from E. coli Inclusion Bodies*. Int J Mol Sci, 2020. **21**(13).
106. Smith, A.T., et al., *Expression of a synthetic gene for horseradish peroxidase C in Escherichia coli and folding and activation of the recombinant enzyme with Ca<sup>2+</sup> and heme*. J Biol Chem, 1990. **265**(22): p. 13335-43.
107. Grigorenko, V., et al., *New Approaches for Functional Expression of Recombinant Horseradish Peroxidase C In Escherichia Coli*. Biocatalysis and Biotransformation, 1999. **17**(5): p. 359-379.
108. Asad, S., et al., *Studies on the refolding process of recombinant horseradish peroxidase*. Mol Biotechnol, 2013. **54**(2): p. 484-92.
109. Gazaryan, I.G., et al., *Effect of single-point mutations Phe41-->His and Phe143-->Glu on folding and catalytic properties of recombinant horseradish peroxidase expressed in E. coli*. FEBS Lett, 1994. **354**(3): p. 248-50.
110. Gazaryan, I.G., et al., *Tryptophanless recombinant horseradish peroxidase: stability and catalytic properties*. Biochem Biophys Res Commun, 1999. **262**(1): p. 297-301.
111. Kim, S.J., B.K. Song, and Y.H. Kim, *Optimized refolding and characterization of S-peroxidase (CWPO\_C of Populus alba) expressed in E. coli*. Protein Expr Purif, 2011. **80**(2): p. 268-73.
112. Shigeto, J., et al., *Identification of Tyr74 and Tyr177 as substrate oxidation sites in cationic cell wall-bound peroxidase from Populus alba L.* FEBS J, 2012. **279**(2): p. 348-57.
113. Shigeto, J., et al., *Catalytic profile of Arabidopsis peroxidases, AtPrx-2, 25 and 71, contributing to stem lignification*. PLoS One, 2014. **9**(8): p. e105332.
114. Zakharova, G.S., et al., *High-yield reactivation of anionic tobacco peroxidase overexpressed in Escherichia coli*. Protein Expr Purif, 2015. **113**: p. 85-93.
115. Fattahian, Y., et al., *Heterologous Expression, Purification and Characterization of a Peroxidase Isolated from Lepidium draba*. Protein J, 2017. **36**(6): p. 461-471.
116. Hushpulan, D.M., et al., *Expression and refolding of tobacco anionic peroxidase from E. coli inclusion bodies*. Biochemistry (Mosc), 2003. **68**(11): p. 1189-94.
117. Teilum, K., L. Ostergaard, and K.G. Welinder, *Disulfide bond formation and folding of plant peroxidases expressed as inclusion body protein in Escherichia coli thioredoxin reductase negative strains*. Protein Expr Purif, 1999. **15**(1): p. 77-82.
118. Doyle, W.A. and A.T. Smith, *Expression of lignin peroxidase H8 in Escherichia coli: folding and activation of the recombinant enzyme with Ca<sup>2+</sup> and haem*. Biochem J, 1996. **315** ( Pt 1): p. 15-9.
119. Pérez-Boada, M., et al., *Expression of Pleurotus eryngii versatile peroxidase in Escherichia coli and optimisation of in vitro folding*. Enzyme and Microbial Technology, 2002. **30**(4): p. 518-524.
120. Nie, G., N.S. Reading, and S.D. Aust, *Expression of the lignin peroxidase H2 gene from Phanerochaete chrysosporium in Escherichia coli*. Biochem Biophys Res Commun, 1998. **249**(1): p. 146-50.
121. Coupe, E.E., et al., *The dodecameric vanadium-dependent haloperoxidase from the marine algae Corallina officinalis: cloning, expression, and refolding of the recombinant enzyme*. Protein Expr Purif, 2007. **52**(2): p. 265-72.
122. Miki, Y., et al., *Escherichia coli expression and in vitro activation of a unique ligninolytic peroxidase that has a catalytic tyrosine residue*. Protein Expr Purif, 2009. **68**(2): p. 208-14.



123. Linde, D., et al., *Heterologous expression and physicochemical characterization of a fungal dye-decolorizing peroxidase from Auricularia auricula-judae*. Protein Expr Purif, 2014. **103**: p. 28-37.
124. Wang, N., et al., *Expression of a fungal manganese peroxidase in Escherichia coli: a comparison between the soluble and refolded enzymes*. BMC Biotechnol, 2016. **16**(1): p. 87.
125. Lee, D. and D.H. Kim, *Heterologous expression of lignin peroxidase H2 in Escherichia coli: In vitro refolding and activation*. Journal of Biochemistry and Molecular Biology, 1999. **32**(5): p. 486-491.
126. Whitwam, R. and M. Tien, *Heterologous expression and reconstitution of fungal Mn peroxidase*. Arch Biochem Biophys, 1996. **333**(2): p. 439-46.
127. Rodriguez-Cabrera, N.A., C. Regalado, and B.E. Garcia-Almendarez, *Cloning, heterologous expression and properties of a recombinant active turnip peroxidase*. J Agric Food Chem, 2011. **59**(13): p. 7120-6.
128. Tang, S., J. Tao, and Y. Li, *Challenges and solutions for the downstream purification of therapeutic proteins*. Antib Ther, 2024. **7**(1): p. 1-12.
129. Gazaryan, I.G., et al., *Effect of single-point mutations Phe41*  
→ *His and Phe143*  
→ *Glu on folding and catalytic properties of recombinant horseradish peroxidase expressed in E. coli*. FEBS Letters, 1994. **354**(3): p. 248-250.
130. Nesa, M.L., et al., *Crystal structure of ferric recombinant horseradish peroxidase*. J Biol Inorg Chem, 2025. **30**(3): p. 221-227.
131. Benov, L. and J. Al-Ibraheem, *Disrupting Escherichia coli: a comparison of methods*. J Biochem Mol Biol, 2002. **35**(4): p. 428-31.
132. Mojsin, M., et al., *Purification and functional analysis of the recombinant protein isolated from E. coli by employing three different methods of bacterial lysis*. Journal of the Serbian Chemical Society, 2005. **70**(7): p. 943-950.
133. Ho, C.W., et al., *Comparative evaluation of different cell disruption methods for the release of recombinant hepatitis B core antigen from Escherichia coli*. Biotechnology and Bioprocess Engineering, 2008. **13**(5): p. 577-583.
134. Slouka, C., et al., *Custom made inclusion bodies: impact of classical process parameters and physiological parameters on inclusion body quality attributes*. Microb Cell Fact, 2018. **17**(1): p. 148.
135. Eggenreich, B., et al., *High pressure homogenization is a key unit operation in inclusion body processing*. J Biotechnol, 2020. **324S**: p. 100022.
136. Ebner, J., V. Sedlmayr, and R. Klausser, *High Pressure Homogenization for Inclusion Body Isolation*. Methods Mol Biol, 2023. **2617**: p. 141-154.
137. Eggenreich, B., et al., *Production strategies for active heme-containing peroxidases from E. coli inclusion bodies - a review*. Biotechnol Rep (Amst), 2016. **10**: p. 75-83.
138. Humer, D. and O. Spadiut, *Wanted: more monitoring and control during inclusion body processing*. World J Microbiol Biotechnol, 2018. **34**(11): p. 158.
139. Rathore, A.S., et al., *Refolding of biotech therapeutic proteins expressed in bacteria: review*. Journal of Chemical Technology and Biotechnology, 2013. **88**(10): p. 1794-1806.
140. Yamaguchi, S., et al., *Protein refolding using chemical refolding additives*. Biotechnology Journal, 2013. **8**(1): p. 17-+.
141. Wang, S.S.S., et al., *Effect of Glutathione Redox System on Lysozyme Refolding in Size Exclusion Chromatography*. Food and Bioproducts Processing, 2006. **84**(1): p. 18-27.
142. Gao, Y.G., et al., *Refolding of lysozyme at high concentration in batch and fed-batch operation*. Korean Journal of Chemical Engineering, 2002. **19**(5): p. 871-875.
143. Eiberle, M.K. and A. Jungbauer, *Technical refolding of proteins: Do we have freedom to operate?* Biotechnol J, 2010. **5**(6): p. 547-59.
144. Walther, C., et al., *Getting ready for PAT: Scale up and inline monitoring of protein refolding of Npro fusion proteins*. Process Biochemistry, 2014. **49**(7): p. 1113-1121.

145. Chau, M.-H. and J.W. Nelson, *Direct measurement of the equilibrium between glutathione and dithiothreitol by high performance liquid chromatography*. FEBS Letters, 1991. **291**(2): p. 296-298.
146. Rubus, V., *Application of a high-throughput microfermentation system for the screening of protein refolding*. 2021, Technische Universität Wien.
147. Meneses-Acosta, A., et al., *Effect of controlled redox potential and dissolved oxygen on the in vitro refolding of an E. coli alkaline phosphatase and chicken lysozyme*. Enzyme Microb Technol, 2013. **52**(6-7): p. 312-8.
148. Pizarro, S.A., et al., *Biomufacturing process analytical technology (PAT) application for downstream processing: Using dissolved oxygen as an indicator of product quality for a protein refolding reaction*. Biotechnol Bioeng, 2009. **104**(2): p. 340-51.
149. D'Atri, V., et al., *Current and future trends in reversed-phase liquid chromatography-mass spectrometry of therapeutic proteins*. TrAC Trends in Analytical Chemistry, 2020. **130**.
150. Kopp, J., et al., *Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles*. J Pharm Biomed Anal, 2020. **188**: p. 113412.
151. Fekete, S., J.L. Veuthey, and D. Guillarme, *New trends in reversed-phase liquid chromatographic separations of therapeutic peptides and proteins: theory and applications*. J Pharm Biomed Anal, 2012. **69**: p. 9-27.
152. Kochling, J., et al., *A platform analytical quality by design (AQbD) approach for multiple UHPLC-UV and UHPLC-MS methods development for protein analysis*. J Pharm Biomed Anal, 2016. **125**: p. 130-9.
153. Sylvester, B., et al., *A Quality by Design (QbD) approach to the development of a gradient high-performance liquid chromatography for the simultaneous assay of curcuminoids and doxorubicin from long-circulating liposomes*. J Pharm Biomed Anal, 2018. **158**: p. 395-404.
154. Tumpa, A., et al., *Quality by Design in the development of hydrophilic interaction liquid chromatography method with gradient elution for the analysis of olanzapine*. J Pharm Biomed Anal, 2017. **134**: p. 18-26.
155. Kim, E.J., et al., *Process Analytical Technology Tools for Monitoring Pharmaceutical Unit Operations: A Control Strategy for Continuous Process Verification*. Pharmaceutics, 2021. **13**(6).
156. Ebner, J., et al., *At-Line Reversed Phase Liquid Chromatography for In-Process Monitoring of Inclusion Body Solubilization*. Bioengineering (Basel), 2021. **8**(6).
157. Gupta, M., *Methods for Affinity-Based Separations of Enzymes and Proteins*. 2002.
158. Kuligowski, J., et al., *Liquid Chromatography | Fourier Transform Infrared Spectroscopy* ☆, in *Encyclopedia of Analytical Science (Third Edition)*, P. Worsfold, et al., Editors. 2019, Academic Press: Oxford. p. 75-85.
159. Kuligowski, J., et al., *Recent advances in on-line liquid chromatography - infrared spectrometry (LC-IR)*. TrAC Trends in Analytical Chemistry, 2010. **29**(6): p. 544-552.
160. Grosshans, S., et al., *In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography*. J Chromatogr A, 2018. **1547**: p. 37-44.
161. Sanden, A., et al., *Fourier-transform infrared spectroscopy as a process analytical technology for near real time in-line estimation of the degree of PEGylation in chromatography*. J Chromatogr A, 2019. **1608**: p. 460410.
162. Måge, I., et al., *Fourier-transform infrared (FTIR) fingerprinting for quality assessment of protein hydrolysates*. Lwt, 2021. **152**.
163. Barth, A., *Infrared spectroscopy of proteins*. Biochim Biophys Acta, 2007. **1767**(9): p. 1073-101.
164. Kristoffersen, K.A., et al., *Average molecular weight, degree of hydrolysis and dry-film FTIR fingerprint of milk protein hydrolysates: Intercorrelation and application in process monitoring*. Food Chem, 2020. **310**: p. 125800.
165. Schwaighofer, A., et al., *External cavity-quantum cascade laser infrared spectroscopy for secondary structure analysis of proteins at low concentrations*. Sci Rep, 2016. **6**: p. 33556.



166. Pelton, J.T. and L.R. McLean, *Spectroscopic methods for analysis of protein secondary structure*. Anal Biochem, 2000. **277**(2): p. 167-76.
167. Singh, B.R., *Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins*, in *Infrared Analysis of Peptides and Proteins*. 1999, American Chemical Society. p. 2-37.
168. Murphy, B.M., et al., *Use of the amide II infrared band of proteins for secondary structure determination and comparability of higher order structure*. Curr Pharm Biotechnol, 2014. **15**(9): p. 880-9.
169. Akhgar, C.K., et al., *QCL-IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography*. Anal Chem, 2022. **94**(14): p. 5583-5590.
170. Levitt, M. and J. Greer, *Automatic identification of secondary structure in globular proteins*. J Mol Biol, 1977. **114**(2): p. 181-239.
171. Perutz, M.F., et al., *Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis*. Nature, 1960. **185**(4711): p. 416-22.
172. van de Weert, M., et al., *Fourier transform infrared spectrometric analysis of protein conformation: effect of sampling method and stress factors*. Anal Biochem, 2001. **297**(2): p. 160-9.
173. Dousseau, F. and M. Pezolet, *Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods*. Biochemistry, 1990. **29**(37): p. 8771-8779.
174. Sane, S.U., S.M. Cramer, and T.M. Przybycien, *Protein structure perturbations on chromatographic surfaces*. Journal of Chromatography A, 1999. **849**(1): p. 149-159.
175. Mazzer, A.R., et al., *Protein A chromatography increases monoclonal antibody aggregation rate during subsequent low pH virus inactivation hold*. J Chromatogr A, 2015. **1415**: p. 83-90.
176. Tauler, R., *Multivariate curve resolution applied to second order data*. Chemometrics and Intelligent Laboratory Systems, 1995. **30**(1): p. 133-146.
177. Kucheryavskiy, S., W. Windig, and A. Bogomolov, *Chapter 3 - Spectral Unmixing Using the Concept of Pure Variables*, in *Data Handling in Science and Technology*, C. Ruckebusch, Editor. 2016, Elsevier. p. 53-99.
178. Akhgar, C.K., et al., *Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography*. Anal Chem, 2022. **94**(32): p. 11192-11200.
179. Pena-Pereira, F., W. Wojnowski, and M. Tobiszewski, *AGREE—Analytical GREENness Metric Approach and Software*. Analytical Chemistry, 2020. **92**(14): p. 10076-10082.
180. Freydel, E.J., et al., *Size-exclusion chromatographic protein refolding: fundamentals, modeling and operation*. J Chromatogr A, 2010. **1217**(49): p. 7723-37.
181. Jungbauer, A., W. Kaar, and R. Schlegl, *Folding and refolding of proteins in chromatographic beds*. Curr Opin Biotechnol, 2004. **15**(5): p. 487-94.
182. Jameel, F., et al., *Recommended Best Practices for Lyophilization Validation 2021 Part II: Process Qualification and Continued Process Verification*. AAPS PharmSciTech, 2021. **22**(8): p. 266.
183. Li, M., et al., *Antibody-Drug Conjugate Overview: a State-of-the-art Manufacturing Process and Control Strategy*. Pharm Res, 2024. **41**(3): p. 419-440.
184. de Faria, E.S.A.L. and A.G. Ryder, *Analyzing protein conjugation reactions for antibody-drug conjugate synthesis using polarized excitation emission matrix spectroscopy*. Biotechnol Bioeng, 2022. **119**(12): p. 3432-3446.
185. Wakankar, A., et al., *Analytical methods for physicochemical characterization of antibody drug conjugates*. MAbs, 2011. **3**(2): p. 161-72.
186. Grairi, M. and M. Le Borgne, *Antibody-drug conjugates: prospects for the next generation*. Drug Discov Today, 2024. **29**(12): p. 104241.
187. Matsuda, Y., *Current approaches for the purification of antibody-drug conjugates*. J Sep Sci, 2022. **45**(1): p. 27-37.
188. Jiang, Q., et al., *Structural Characterization of the Aggregates of Gemtuzumab Ozogamicin*. ACS Omega, 2019. **4**(4): p. 6468-6475.

189. Bernini, C., et al., *The nature of tryptophan radicals involved in the long-range electron transfer of lignin peroxidase and lignin peroxidase-like systems: Insights from quantum mechanical/molecular mechanics simulations*. Proteins, 2012. **80**(5): p. 1476-83.
190. Doyle, W.A., et al., *Two substrate interaction sites in lignin peroxidase revealed by site-directed mutagenesis*. Biochemistry, 1998. **37**(43): p. 15097-105.
191. Falade, A.O., et al., *Lignin peroxidase functionalities and prospective applications*. Microbiologyopen, 2017. **6**(1).
192. Lundell, T., et al., *Engineering Towards Catalytic Use of Fungal Class-II Peroxidases for Dye-Decolorizing and Conversion of Lignin Model Compounds*. Current Biotechnology, 2017. **6**(2): p. 116-127.
193. Kumar, A. and R. Chandra, *Ligninolytic enzymes and its mechanisms for degradation of lignocellulosic waste in environment*. Heliyon, 2020. **6**(2): p. e03170.
194. Lambertz, C., et al., *Progress and obstacles in the production and application of recombinant lignin-degrading peroxidases*. Bioengineered, 2016. **7**(3): p. 145-54.
195. Krainer, F.W., et al., *Purification and basic biochemical characterization of 19 recombinant plant peroxidase isoenzymes produced in Pichia pastoris*. Protein Expr Purif, 2014. **95**: p. 104-12.
196. Kay, E., L.M. Shannon, and J.Y. Lew, *Peroxidase Isozymes from Horseradish Roots*. Journal of Biological Chemistry, 1967. **242**(10): p. 2470-2473.
197. Won, K., et al., *Horseradish peroxidase-catalyzed polymerization of cardanol in the presence of redox mediators*. Biomacromolecules, 2004. **5**(1): p. 1-4.
198. Kersten, P.J., et al., *Comparison of lignin peroxidase, horseradish peroxidase and laccase in the oxidation of methoxybenzenes*. Biochem J, 1990. **268**(2): p. 475-80.
199. Plácido, J. and S. Capareda, *Ligninolytic enzymes: a biotechnological alternative for bioethanol production*. Bioresources and Bioprocessing, 2015. **2**(1).
200. Kurek, B. and B. Monties, *Oxidation of spruce lignin by fungal lignin peroxidase and horseradish peroxidase: Comparison of their actions on molecular structure of the polymer in colloidal solution*. Enzyme and Microbial Technology, 1994. **16**(2): p. 125-130.
201. Ambert-Balay, K., S.M. Fuchs, and M. Tien, *Identification of the veratryl alcohol binding site in lignin peroxidase by site-directed mutagenesis*. Biochem Biophys Res Commun, 1998. **251**(1): p. 283-6.
202. Ufot, U.F. and M.I. Akpanabiatu, *An engineered <i>Phlebia radiata</i> manganese peroxidase: expression, refolding, purification and preliminary characterization*. American Journal of Molecular Biology, 2012. **02**(04): p. 359-370.
203. Whitwam, R.E., I.G. Gazarian, and M. Tien, *Expression of fungal Mn peroxidase in E. coli and refolding to yield active enzyme*. Biochem Biophys Res Commun, 1995. **216**(3): p. 1013-7.
204. Fernandez-Fueyo, E., et al., *Ligninolytic peroxidase genes in the oyster mushroom genome: heterologous expression, molecular structure, catalytic and stability properties, and lignin-degrading ability*. Biotechnol Biofuels, 2014. **7**(1): p. 2.
205. Saez-Jimenez, V., et al., *Demonstration of Lignin-to-Peroxidase Direct Electron Transfer: A TRANSIENT-STATE KINETICS, DIRECTED MUTAGENESIS, EPR, AND NMR STUDY*. J Biol Chem, 2015. **290**(38): p. 23201-13.
206. Balta-Foix, R., E. Garcia-Fruitos, and A. Aris, *Time to consider ruling out inclusion bodies denaturing protocols for spontaneous solubilization of biologically active proteins*. Sci Rep, 2024. **14**(1): p. 26061.
207. Pittiu, A., et al., *Production of liposomes by microfluidics: The impact of post-manufacturing dilution on drug encapsulation and lipid loss*. Int J Pharm, 2024. **664**: p. 124641.
208. Yu, B., R.J. Lee, and L.J. Lee, *Microfluidic methods for production of liposomes*. Methods Enzymol, 2009. **465**: p. 129-41.
209. Buttitta, G., et al., *Scalable microfluidic method for tunable liposomal production by a design of experiment approach*. Int J Pharm, 2024. **662**: p. 124460.
210. Ferre, H., et al., *A novel system for continuous protein refolding and on-line capture by expanded bed adsorption*. Protein Sci, 2005. **14**(8): p. 2141-53.
211. Pan, S., et al., *Continuous protein refolding in a tubular reactor*. Chemical Engineering Science, 2014. **116**: p. 763-772.



212. Rajendran, V., et al., *Continuous protein refolding and purification by two-stage periodic counter-current chromatography*. J Chromatogr A, 2023. **1695**: p. 463938.
213. Igwe, C.L., et al., *Online monitoring of protein refolding in inclusion body processing using intrinsic fluorescence*. Anal Bioanal Chem, 2024. **416**(12): p. 3019-3032.
214. Bergs, T., et al., *The Concept of Digital Twin and Digital Shadow in Manufacturing*. Procedia CIRP, 2021. **101**: p. 81-84.

## 6 Appendix I: Scientific Publications

### 6.1 Book Chapter 1



## Chapter 9

### High Pressure Homogenization for Inclusion Body Isolation

Julian Ebner, Viktor Sedlmayr, and Robert Klausser

#### Abstract

High pressure homogenization (HPH) is a commonly used method for cell lysis of *Escherichia coli* in order to release intracellularly produced recombinant proteins. For misfolded proteins in *E. coli*, focus is often put on the development of a suitable solubilization and refolding protocol. However, HPH can be a critical unit operation influencing inclusion body (IB) quality and, subsequently, refolding yields. Here, a protocol for homogenization and IB washing is presented in combination with analytical methods suitable to evaluate these unit operations. The protocol is based on a multivariate approach to identify suitable conditions during HPH. Furthermore, the described workflow is easily scalable and can, therefore, also be used if fixed homogenization conditions are already established.

**Key words** High pressure homogenization, *Escherichia coli*, Cell lysis, Inclusion bodies (IBs), Design of experiment, Inclusion body isolation, Inclusion body quality

#### 1 Introduction

In recombinant protein production processes with *Escherichia coli*, the protein of interest (POI) is predominantly expressed intracellularly. In many cases, the expression leads to the formation of inclusion bodies (IBs). These insoluble protein aggregates are formed in the cytoplasm due to a lack of a post-translational modification machinery in *E. coli* [1–4]. A traditional production process for IBs comprises the following unit operations: (1) harvest of the biomass at the end of fermentation (2) cell lysis, (3) IB isolation and wash, (4) solubilization and refolding, and (5) capture and purification. Compared to the conventional production process of a soluble protein, step (3) IB isolation and wash, and step (4) solubilization and refolding are additionally required to obtain active protein from the insoluble IBs [3, 5, 6]. High pressure homogenization (HPH), ultrasonication, or enzymatic lysis are commonly used cell lysis methods, both for soluble proteins and IBs [7–9]. However, suitable conditions for HPH can significantly vary as soluble proteins are released into the supernatant, while IBs are

Julian Kopp and Oliver Spadiut (eds.), *Inclusion Bodies: Methods and Protocols*,  
Methods in Molecular Biology, vol. 2617, [https://doi.org/10.1007/978-1-0716-2930-7\\_9](https://doi.org/10.1007/978-1-0716-2930-7_9),  
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023



separated in the pellet after cell lysis. Therefore, the most important quality attributes, titer and purity of the POI after cell lysis, must be assessed for HPH. The easy separation of soluble host cell impurities via centrifugation can be a big advantage for a production strategy as IBs, attaining a high innate purity of the POI [10]. Target protein purity can then be improved even further by wash steps with appropriate buffers.

The presented protocol describes a workflow used to identify suitable cell lysis conditions in a design of experiment (DoE) approach. While Horseradish Peroxidase (HRP) IBs served as a model protein in the example displayed here, the shown methodology can be applied for any POI. The multivariate approach allows the identification of interacting factors as well as quadratic interactions, posing a significant advantage compared to univariate experiments. Using a HPH device enables the possibility of scale up steps in order to process large volumes. Compared to ultrasonication or enzymatic cell lysis, higher purities of the isolated IBs can be reached due to the applied shear forces during HPH and the resulting fractionation of otherwise insoluble cell debris [11]. For the described protocol, three key analytical methods are used to evaluate the homogenization conditions: (1) reversed phase high-performance liquid chromatography (RP-HPLC) is used to measure titer and purity of the POI, (2) SDS-PAGE is employed to measure titer and purity of the POI using a separation principle orthogonal to RP-HPLC, and (3) DNA concentration is measured to detect the release of DNA into the supernatant, serving as a good indicator of cell lysis. If a refolding protocol is already established for the POI, the refolding yield can be determined as an additional response to further evaluate the applied homogenization conditions. While the presented protocol is based on a DoE approach to determine an optimized HPH protocol, the core elements, such as sampling strategy and analytical methods, can also be used outside of the DoE context. Therefore, it can easily be adapted to screen suitable washing buffers and wash steps or for the isolation of IBs in a larger scale.

## 2 Materials

### 2.1 Biomass Containing the Target Product as IBs

For the shown workflow, around 200 g wet Cell Weight (wCW) of *E. coli* biomass, containing the POI as IBs, is required. After harvest (end of fermentation), the biomass can be isolated using centrifugation and stored at  $-20^{\circ}\text{C}$  until homogenization experiments are performed.

Find details in the book chapter **Inclusion Body Production in Fed-Batch and Continuous Cultivation**.

## 2.2 HighPressureHomogenization

1. The following buffers are given as an example and were used for HRP IBs (*see* Notes 1 and 2) [12]:
  - Homogenization buffer: 50 mM Tris HCl pH 8, 500 mM NaCl, 1.5 mM EDTA
  - Wash buffer: 50 mM Tris HCl pH 8, 500 mM NaCl, 2 M urea
2. HPH device (e.g., PANDA PLUS 2000 (GEA, Düsseldorf, Germany)) (*see* Note 3).
3. A dispersing instrument suitable for a volume of 50–1000 mL (e.g., T10 basic ULTRA-TURRAX (IKA, Staufen, Germany)) (*see* Note 4).
4. Centrifuges for 50 mL and 1.5 mL centrifugation tubes.
5. Laboratory balance.
6. Ice Box.
7. Plastic beaker (0.5–1 L) (*see* Note 5).
8. Consumables:
  - Centrifugation tubes (50 mL and 1.5 mL)

## 2.3 Analytical Tools

1. RP-HPLC (*see* Note 6):
  - An HPLC device consisting of a pump module, an auto-sampler module, a column oven, and a UV detector
  - Mobile phase A: HPLC grade water (MQ) supplemented with 0.1% trifluoroacetic acid (TFA)
  - Mobile phase B: HPLC grade acetonitrile (ACN) supplemented with 0.1% TFA
  - RP column (e.g., BioResolve RP mAb Polyphenyl column (Waters Corporation, MA, USA))
  - Rocker-Shaker
  - Solubilization buffer: 62 mM Tris HCl pH 8, 7.5 M guanidine hydrochloride, 125 mM 1,4-Dithiothreitol (DTT)
  - Software for analysis of chromatograms
2. SDS-PAGE:
  - SDS gels (e.g., Mini-PROTEAN TGX Precast Gels 4–15% (Bio-Rad, CA, USA))
  - Gel electrophoresis running system (e.g., Mini-PROTEAN Tetra System (Bio-Rad, CA, USA))
  - Laemmli buffer (e.g., 2× Laemmli Sample Buffer (Bio-Rad, CA, USA)) supplemented with a reducing agent (e.g., 2-Mercaptoethanol)
  - Coomassie staining (e.g., Bio-Safe Coomassie Stain (Bio-Rad, CA, USA))



- Imaging device and according software (e.g., Gel Doc XR+ Gel Documentation System and Image Lab Software (both Bio-Rad, CA, USA))
3. DNA concentration: Photometric device to measure the absorbance at 260 and 280 nm (e.g., NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, DE, USA)).
  4. If a refolding protocol is established for the POI, equipment to perform solubilization and refolding and analytical methods are required to determine the refolding yield (*see Note 7*).  
Find details in the book chapter **Unit Operation Spanning Investigation of the Redox System**.

## 2.4 Experimental Design and Multivariate Data Assessment

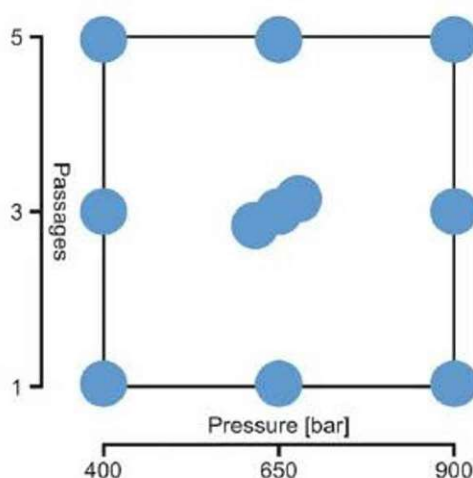
Software to design and subsequently interpret the experiments performed (e.g., MODDE 12 (Sartorius, Göttingen, Germany)) (*see Note 8*).

## 3 Methods

### 3.1 Experimental Planning and DoE

Since the unit operation HPH has been shown to significantly influence purity and quality of the isolated IBs, we recommend a multivariate approach to determine feasible conditions [13]. However, the main elements such as sampling and analytical methods are also applicable for a univariate or a constant approach (*see Note 8*). If a multivariate approach is chosen, the following points should be considered during the design of the experimental plan.

1. Decide on the factors that should be varied during the experiments. Typically investigated factors are the number of passages and the pressure which is exemplary shown in this protocol. The number of passages describes the number of times the biomass suspension is pumped through the HPH device (*see Note 9*).
2. Select feasible ranges for each factor. In the case of the pressure, the upper limit is dictated by the specifications of the used HPH device. In the case of the passages, the time per sample should be considered, especially for larger volumes (*see Note 10*).
3. Choose the design that should be used (e.g., full factorial design or central composite design). In our experience, a central composite face centered design is preferable, since potential interaction and quadratic terms are available, while the total number of experiments is still manageable for two factors (*see Note 11*).



**Fig. 1** Example for a central composite face centered design using two factors (pressure and passages)

4. Decide on the responses used for the DoE. If a refolding protocol is already established for the POI, the refolding yield is a highly significant response (*see Note 12*).

In Fig. 1, an exemplary multivariate design for the number of passages and the pressure is shown.

### 3.2 Resuspending the Biomass

1. Pre-cool the homogenization buffer (either on ice or in a 4 °C fridge).
2. Weigh in approximately 40 g wCW of frozen biomass and note the weight (*see Notes 3 and 13*).
3. Add homogenization buffer to reach a biomass concentration of 200 g wCW/L (for 40 g biomass, 0.2 L homogenization buffer is required) (*see Note 13*).
4. Resuspend the biomass using the Ultraturrax on level 3, around 60 s of resuspension followed by 30 s cooling of the beaker in an icebox (*see Notes 14 and 15*).

### 3.3 HPH Experiments for the DoE

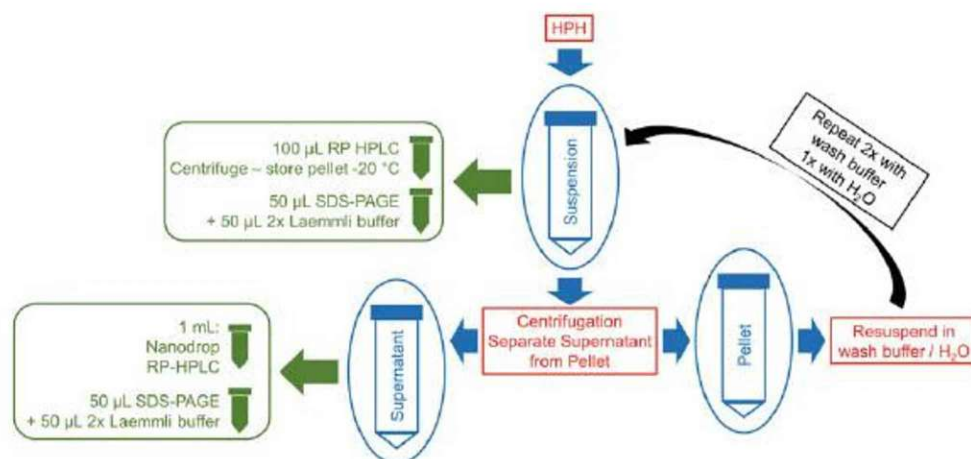
1. Fill the resuspended biomass into the feeding hopper of the used HPH device.
2. Connect the product outlet to the feeding hopper.
3. Turn on the flow to the desired flow rate.
4. Adjust the pressure to the desired value.
5. As soon as the desired pressure is reached, connect the product outlet to a 1 L centrifugation beaker (*see Note 16*).



6. When the hopper is almost empty, reduce the pressure to 0 bar and then turn off the flow (*see Note 17*).
7. Take a sample (40 mL) from the beaker connected to the product outlet (e.g., after 1/3/5 passages, depending on the used DoE) in a tared 50 mL centrifugation tube (*see Note 18*).
8. Fill the homogenized biomass into the feeding hopper.
9. Repeat steps 2–8 for the desired number of passages.
10. Flush the HPH device with water (approximately 1 L).
11. If a multivariate approach is used, then repeat steps 1–10 for the next condition (e.g., different pressure).
12. Clean and store the HPH device according to the device specifications.

### 3.4 Centrifugation Step

1. An overview which sample is drawn during the homogenization and washing process is given in Fig. 2. Sampling steps are also marked as “Sample” in the text.
2. Pre-cool centrifuge to 4 °C.
3. **Sample:** Draw suspension sample for SDS-PAGE (50 µL). Add 50 µL 2× Laemmli to sample.
4. **Sample:** Draw sample for RP-HPLC (100 µL). Centrifuge at 20,280 rcf, 20 min, 4 °C and discard the supernatant. The pellet can be stored at –20 °C until measurement on the RP-HPLC is performed.



**Fig. 2** Sampling plan for the presented protocol. Red boxes represent operations, starting at the top with the high pressure homogenization (HPH). Blue ellipses represent 50 mL centrifugation tubes containing suspension, supernatant, and pellet obtained after HPH and green boxes represent the required samples drawn

5. Centrifuge the remaining suspensions (50 mL centrifugation tube) at 20,280 rcf, 20 min, 4 °C (*see Note 19*).
6. Separate the supernatant from the cell debris pellet (*see Note 20*).
7. **Sample:** Store the supernatant for Nanodrop, SDS-PAGE, and RP-HPLC at 4 °C.
8. Weigh the centrifugation tube containing the isolated pellet and calculate the wet IB weight (wIBw).

### 3.5 Washing

1. Resuspend the pellet to a concentration of 100 g/L in wash buffer (using the Ultraturrax level 3).
2. **Sample:** Draw suspension sample for SDS-PAGE (50 µL). Add 50 µL 2× Laemmli to sample.
3. **Sample:** Draw sample for RP-HPLC (100 µL). Centrifuge at 20,280 rcf, 20 min, 4 °C and discard the supernatant. The pellet can be stored at −20 °C until measurement on the RP-HPLC is performed.
4. Centrifuge the remaining suspensions (50 mL centrifugation tube) at 20,280 rcf, 20 min, 4 °C (*see Note 19*).
5. Separate the supernatant from the cell debris pellet.
6. **Sample:** Store the Supernatant for Nanodrop, SDS-PAGE and RP-HPLC at 4 °C.
7. Note the weight of the IB (*see Note 18*).
8. Repeat the steps 1–7 once.

### 3.6 Aliquotation and Storage of Washed IBs

1. Resuspend the pellet to a concentration of 100 g/L in water (using the Ultraturrax level 3).
2. **Sample:** Draw suspension sample for SDS-PAGE (50 µL). Add 50 µL 2× Laemmli to sample.
3. **Sample:** Draw sample for RP-HPLC (100 µL). Centrifuge at 20,280 rcf, 20 min, 4 °C and discard the supernatant. The pellet can be stored at −20 °C until measurement on the RP-HPLC is performed.
4. If a known amount of IB is needed for further processing steps, aliquote the IB suspension accordingly (*see Note 21*).
5. Centrifuge the aliquoted suspensions at 20,280 rcf, 20 min, 4 °C. (*see Note 19*).
6. Separate the supernatant from the cell debris pellet.
7. **Sample:** Store the supernatant for Nanodrop, SDS-PAGE, and RP-HPLC at 4 °C.
8. Note the weight of the IB (*see Note 18*).
9. Store the pellet at −20 °C.



**3.7 Analytics****Device setup:****3.7.1 RP-HPLC**

- Prepare mobile phase A (MQ + 0.1% TFA) and mobile phase B (ACN + 0.1% TFA).
- Precool the autosampler to 5 °C.
- Prepare reversed phase column for analysis at 70 °C and a constant flow rate of 0.4 mL/min (runtime 18 min). For a detailed description of the used method, refer to [14, 15].

**Samples:**

- Supernatants: Can be measured without any further sample preparation (*see* Note 22).
- Pellets: Add solubilization buffer to reach the desired protein concentration (e.g., 10 g wBw/L). For the shown sampling plan, add 1 mL of solubilization buffer to the pellet of the RP-HPLC sample (*see* Subheadings 3.4, 3.5, and 3.6, step 3).

Dissolve at room temperature (RT) for 10 min on a rocker-shaker. Centrifuge at 20,280 rcf, 20 min, 4 °C. Use the supernatant for the RP-HPLC measurement.

- Measure samples and, if available, perform a standard calibration using the pure POI in appropriate concentrations (e.g., 0.1–2 g/L).
- Calculate protein concentration and purity using a suitable chromatographic software.

**3.7.2 SDS-PAGE**

- Dilute the sample 1:2 in 2× Laemmli (reducing) (e.g., 50 µL sample + 50 µL Laemmli buffer) (*see* Notes 23 and 24).
- Heat to 95 °C for 10 min.
- Spin down all samples.
- Dilute samples to the appropriate protein concentration (*see* Note 25).
- Perform SDS-PAGE.
- Use Coomassie staining and an appropriate imaging device. Densitometric analysis of the SDS-PAGE can be performed with a suitable software (*see* Subheading 2.3).

**3.7.3 DNA Concentration**

- Measure the absorbance of the supernatant samples at 260 and 280 nm.
- Based on the absorbance at 260 nm, calculate the DNA concentration.
- Based on the ratio of the absorbance at 260 and 280 nm, calculate the DNA purity.

**3.8 Constant DSP**

If a refolding protocol is established for the POI, refolding of the IBs obtained for the different homogenization conditions can be performed (*see* Note 26).

1. Use the frozen IB pellets obtained in step 9 of Subheading 3.6 for solubilization and refolding using constant DSP approach.
2. After refolding, measure the volumetric yield [U/mL or mg/mL], i.e., the concentration of correctly folded protein (*see* Note 27).
3. Based on the amount of wIBw or POI obtained after the homogenization, calculate the amount of correctly folded protein per gram biomass (*see* Note 28).

**3.9 Analysis of the DoE**

If a DoE approach was applied, use an appropriate software tool for data analysis and interpretation. In the following, the exemplary responses are discussed:

1. The measurement of the DNA concentration in the homogenization supernatant can be used to identify potential outliers. The DNA concentration should increase for harsher conditions (i.e., a higher percentage of cells are lysed) (*see* Note 29).
2. SDS-PAGE and RP-HPLC samples after the final washing step are used to measure quantity and purity of the POI for the applied washing protocol.
3. SDS-PAGE and RP-HPLC samples drawn after homogenization and during the washing steps are used to monitor the efficacy of the washing steps (as yield and purity can be calculated for each wash step).
4. While IB yield and especially purity can be good estimators of the refolding behavior, in our experience, the refolding yield is not always accurately predicted by these responses. Therefore, if a refolding protocol is established, the volumetric yield during refolding and the yield per gram biomass (described in Subheading 3.8, steps 2 and 3) are the best responses to identify suitable homogenization conditions.

**4 Notes**

1. The used homogenization and wash buffer are highly dependent on the POI. The shown buffers were used for HRP IBs and can be a good starting point to identify appropriate buffers as 2 M urea is a relatively low chaotropic reagent concentration. This prevents solubilization of IBs during the wash steps.
2. Depending on the POI, homogenization and wash buffer can be identical. This reduces the time needed for buffer preparation and already includes the first wash cycle in the homogenization process.



3. All volumes given in this protocol are calculated for the PANDA PLUS 2000 with a minimum working volume of 40 mL. If a device with a different minimum volume is used, then the volumes can be adjusted accordingly to use the described workflow.
4. A dispersing instrument is not absolutely necessary, however, it vastly reduced time needed for resuspension. This is particularly important as each wash step includes resuspension followed by centrifugation, making the workflow very laborious if no dispersing instrument is used.
5. Essentially any beaker or flask with the required volume can be used. However, we found that plastic beakers with a large top opening are beneficial for resuspending biomass as every part of the beaker can easily be reached with the dispersing instrument.
6. The specific analytical methods used are highly protein dependent. Therefore, we showed three examples (RP-HPLC, SDS-PAGE, and UV absorbance at 260 nm to measure DNA concentration) which are applicable for a wide range of different proteins. However, many more methods are available and can be used to monitor cell lysis [13, 16–27].
7. In our experience, the achieved refolding yield is the most significant response in order to identify suitable homogenization conditions. However, as washed IBs are needed in a first step to develop a refolding protocol, this requires an iterative approach. Therefore, if no refolding protocol is established, a high purity (measured, e.g., with RP-HPLC or SDS-PAGE) of IBs is a good starting point to develop a refolding protocol. The homogenization conditions can then be adapted after the refolding protocol is in place.
8. While we would strongly recommend a multivariate approach, this is not strictly necessary. Experiments can also be performed univariate and, potentially, be included in a DoE approach later on if necessary.
9. Other important factors that could be investigated include biomass concentration and homogenization buffer composition.
10. If the process will, at some point, be performed in a larger scale, the limitations of the HPH device used after scale-up should also be considered here. Particular attention should be paid to the number of passages, as a high number of passages is often too time-consuming to be performed for large volumes.
11. Central composite circumscribed and central composite inscribed designs might potentially be used as well. However, attention should be paid in order not to exceed the device

limitations (e.g., pressure) for the star points. Furthermore, the number of passages is a discrete value, therefore fractions required for the star points are not possible.

12. This is highly product specific! While in the presented example the refolding yield, DNA concentration, purity and quantity of the POI were chosen, other responses such as turbidity, viscosity, or enzymatic activity in the supernatant can present valid responses as well.
13. It is not required to weigh in exactly 40 g of wCW as the volume of homogenization buffer added afterwards can easily be adapted to reach the desired concentration of 200 g wCW/L.
14. Depending on the used buffer and biomass, around 10 min stirring using a magnetic stirring bar before the dispersing instrument is used can make resuspension easier.
15. As larger particles can block the used homogenization device, the suspension should be homogeneous and free of visible particles. If it is not possible to fully resuspend the biomass or if frozen biomass from a plastic bag was used and parts of the bag are present in the resuspended biomass, then a common kitchen sieve can be used to separate particles before filling the resuspended biomass into the feeding hopper of the homogenization device.
16. If the used feeding hopper is not big enough for the whole volume, then use a second (empty) beaker to connect the HPH outlet to and refill the feeding hopper with the rest of the biomass suspension once it is emptying.
17. Depending on the homogenization device used, it should not run dry, especially while pressure is applied. To prevent this, a few milliliters can be left in the feeding hopper while adjusting the pressure to 0 bar and turn off the flow. Afterwards, the non-homogenized suspension left in the homogenizer can be discarded.
18. We found it to be easiest if the Tara of the used centrifugation tube is noted in a laboratory journal and on the tube itself. The Tara on the tube itself allows for a quick calculation of the required amount of buffer for the resuspension steps.
19. Either use a centrifugation tube filled with water as a counterweight or discard suspension to balance two tubes containing suspension for centrifugation. If discarding is used, take care that the suspension is homogeneous before pipetting. If buffer would be added to adjust the weight, this would dilute the supernatant, influencing the values of DNA and protein concentration in the supernatant and is therefore not recommended.



20. Especially for mild homogenization conditions (e.g., 400 bar, 1 passage), a gelatinous layer might be present between the solid pellet and the cleared supernatant. As supernatant samples should be particle free (especially for RP-HPLC measurements), we recommend that the supernatant sample is drawn with a pipette (e.g., 1 mL) and care is taken as not to disturb the pellet. The rest of the supernatant and the gelatinous layer can then be discarded safely. As an alternative, the supernatants can be filtered (0.22  $\mu\text{m}$  pore size) in order to guarantee compatibility with the HPLC measurement.
21. It is possible to perform an aliquotation step before centrifugation. If a known amount of wet IB is required for further processing steps (e.g., a solubilization and refolding protocol), then the corresponding volume of suspension can be aliquoted, as the concentration of wIBw/L is known in this step. In order to keep the suspension homogenous during aliquotation, the centrifugation tube can be stirred with a magnetic stirrer and stirring bar during pipetting. If no stirring bar fitting into the tube is available, then the suspension can also be transferred to, e.g., a 50 mL or 100 mL shot bottle with a fitting stirring bar. However, in our experience, IBs are very stable to repeated resuspension in water. Therefore, frozen IB samples can also be resuspended in water and aliquoted to the desired amount later on.
22. Depending on the amount of protease inhibitor present in the homogenization buffer and the buffer composition in general, supernatant samples might not be stable for longer time spans. It is therefore recommended to measure these samples in a timely fashion, latest a few hours after homogenization, and store them at 4 °C until the measurement is performed.
23. As the suspension samples are the pellets from the previous washing step, it is not necessary to prepare additional samples of the pellets for SDS-PAGE. In our experience, solubilizing a solid pellet using Laemmli buffer leads to a significant amount of foam formation, making it difficult to prepare samples quantitatively.
24. If guanidine hydrochloride is used in the homogenization or wash buffer, SDS-PAGE cannot be performed, as it precipitates when mixed with Laemmli buffer.
25. The desired protein concentration depends on the SDS-PAGE used. For the listed manufacturer and method, a concentration of around 1 mg/mL protein in the sample worked well. For the shown example, around 10% of the wIBw was target protein. Therefore, a 1:10 dilution is required to reach the desired 1 mg/mL concentration (of a suspension containing 100 g wIBw/L).

26. In order to achieve comparability, protein concentration can either be normalized to the wet IB concentration (wIBw/L) in the solubilizate or to the POI concentration in the solubilizate.
27. The analytical methods used to measure correctly folded protein are highly protein specific. If the correctly folded POI is an enzyme, then enzymatic activity measurement is usually performed. For non-enzyme proteins, binding assays can be performed in some cases (e.g., antibodies and fragments thereof).
28. Depending on the application of the process, the product yield per biomass might be of secondary importance (e.g., if the goal is to produce small amounts of pure protein in a bench-scale approach). However, if the process is scaled up at some point, particular attention should be paid to this response.
29. It is possible that the DNA concentration is not increasing for harsher conditions (e.g., higher pressure and number of passages), as basically all cells are lysed. DNA concentration might even decrease at harsher conditions due to shear forces fragmenting the DNA. However, IB quality attributes (such as purity) might still change because cell wall fragments are further disrupted. Therefore, it can be beneficial to investigate conditions even if the DNA concentration remains constant.

## References

1. Swartz JR (2001) Advances in *Escherichia coli* production of therapeutic proteins. *Curr Opin Biotechnol* 12(2):195–201
2. Yamaguchi H, Miyazaki M (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomol Ther* 4(1):235–251
3. Huang CJ, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J Ind Microbiol Biotechnol* 39(3):383–399
4. Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 5:172
5. Rathore AS et al (2013) Refolding of biotech therapeutic proteins expressed in bacteria: review. *J Chem Technol Biotechnol* 88(10):1794–1806
6. Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* 99(4):303–310
7. Benov L, Al-Ibraheem J (2002) Disrupting *Escherichia coli*: a comparison of methods. *J Biochem Mol Biol* 35(4):428–431
8. Mojsin M et al (2005) Purification and functional analysis of the recombinant protein isolated from *E. coli* by employing three different methods of bacterial lysis. *J Serb Chem Soc* 70(7):943–950
9. Ho CW et al (2008) Comparative evaluation of different cell disruption methods for the release of recombinant hepatitis B core antigen from *Escherichia coli*. *Biotechnol Bioprocess Eng* 13(5):577–583
10. Vallejo LF, Rinas U (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb Cell Factories* 3(1):11
11. Eggenreich B et al (2020) High pressure homogenization is a key unit operation in inclusion body processing. *J Biotechnol* 324S:100022
12. Humer D, Ebner J, Spadiut O (2020) Scalable high-performance production of recombinant horseradish peroxidase from *E. coli* inclusion bodies. *Int J Mol Sci* 21(13):4625
13. Eggenreich B et al (2017) A combination of HPLC and automated data analysis for monitoring the efficiency of high-pressure homogenization. *Microb Cell Factories* 16(1):134
14. Kopp J et al (2020) Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical



- quality-by-design principles. *J Pharm Biomed Anal* 188:113412
15. Ebner J et al (2021) At-line reversed phase liquid chromatography for in-process monitoring of inclusion body solubilization. *Bioengineering (Basel)* 8(6):78
  16. Hettwer D, Wang H (1989) Protein release from *Escherichia coli* cells permeabilized with guanidine-HCl and triton X100. *Biotechnol Bioeng* 33(7):886–895
  17. Middelberg APJ (2000) 2 Microbial cell disruption by high-pressure homogenization. In: *Downstream processing of proteins*. Springer, Cham, pp 11–21
  18. Ramanan RN et al (2009) Classification of pressure range based on the characterization of *Escherichia coli* cell disruption in high pressure homogenizer. *Am J Biochem Biotechnol* 5(1):21–29
  19. Singh RS (2013) A comparative study on cell disruption methods for release of aspartase from *E. coli* K-12. *Indian J Exp Biol* 51(11):997–1003
  20. Kastenhofer J et al (2021) Monitoring and control of *E. coli* cell integrity. *J Biotechnol* 329:1–12
  21. Shrestha P, Holland TM, Bundy BC (2012) Streamlined extract preparation for *Escherichia coli*-based cell-free protein synthesis by sonication or bead vortex mixing. *BioTechniques* 53(3):163–174
  22. Li Q et al (2012) Use of focused acoustics for cell disruption to provide ultra scale-down insights of microbial homogenization and its bioprocess impact – recovery of antibody fragments from rec *E. coli*. *Biotechnol Bioeng* 109(8):2059–2069
  23. Kastenhofer J, Spadiut O (2020) Culture medium density as a simple monitoring tool for cell integrity of *Escherichia coli*. *J Biotechnol* 324S:100017
  24. Fykse EM, Olsen JS, Skogan G (2003) Application of sonication to release DNA from *Bacillus cereus* for quantitative detection by real-time PCR. *J Microbiol Methods* 55(1):1–10
  25. Kastenhofer J et al (2021) Economic and ecological benefits of a leaky *E. coli* strain for downstream processing: a case study for staphylococcal protein A. *J Chem Technol Biotechnol* 96(6):1667–1674
  26. Fonseca LP, Cabral JMS (2002) Penicillin acylase release from *Escherichia coli* cells by mechanical cell disruption and permeabilization. *J Chem Technol Biotechnol* 77(2):159–167
  27. Van Hee P et al (2004) Relation between cell disruption conditions, cell debris particle size, and inclusion body release. *Biotechnol Bioeng* 88(1):100–110



## Chapter 11

### Unit Operation-Spanning Investigation of the Redox System

Julian Ebner, Diana Humer, and Viktor Sedlmayr

#### Abstract

Cytoplasmic expression of recombinant proteins requiring disulfide bridges in *Escherichia coli* usually leads to the formation of insoluble inclusion bodies (IBs). The reason for this phenomenon is found in the reducing environment of the cytoplasm, preventing the formation of disulfide bridges and therefore resulting in inactive protein aggregates. However, IBs can be refolded in vitro to obtain the protein in its active conformation. In order to correctly form the required disulfide bridges, cystines are fully reduced during solubilization and, with the help of an oxidizing agent, the native disulfide bridges are formed during the refolding step. Here, a protocol to identify suitable redox conditions for solubilization and refolding is presented. For this purpose, a multivariate approach spanning the unit operations solubilization and refolding is used.

**Key words** Inclusion bodies (IBs), Disulfide bridges, Inclusion body refolding, Redox system, Design of Experiment (DoE), Unit operation-spanning

#### 1 Introduction

One of the reasons for the formation of inclusion bodies (IBs) in *Escherichia coli* is the presence of disulfide bridges in the expressed protein of interest (POI). Due to the reducing environment of the cytoplasm, the posttranslational formation of disulfide bridges is severely restricted or impossible [1, 2]. This, in turn, hinders the correct folding of the POI and leads to protein aggregation and the formation of IBs. In order to obtain the POI in its active form, three additional unit operations are commonly performed during the downstream processing of IBs: (1) isolation and washing of IBs; (2) solubilization of IBs; and (3) refolding of the solubilized IBs [3–6]. If the POI contains disulfide bridges, a redox system is used in order to obtain the correctly folded and active POI [7]. Commonly, a reducing agent is added during the solubilization to reduce any formed disulfide bridges and enable complete solubilization. During the unit operation of refolding, an oxidizing agent is added to the refolding buffer to allow the formation of disulfide

Julian Kopp and Oliver Spadiut (eds.), *Inclusion Bodies: Methods and Protocols*, Methods in Molecular Biology, vol. 2617, [https://doi.org/10.1007/978-1-0716-2930-7\\_11](https://doi.org/10.1007/978-1-0716-2930-7_11), © The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023



bridges. The combination of reducing agent during solubilization and oxidizing agent during refolding (the redox system) then results in the renatured and active POI [8–10]. However, the individual concentrations of reducing and oxidizing agent required to yield high amounts of protein in its native conformation are highly specific for each POI. Therefore, an empiric approach to identify suitable conditions is necessary during the development of a production process for each new POI [11–13].

In this protocol, a workflow to identify suitable redox conditions for batch dilution refolding is presented. Based on a multivariate approach, conditions are identified using a design of experiments (DoE), varying the factors concentration of reducing agent during solubilization and concentration of oxidizing agent and pH value in the refolding. The reason to perform a DoE instead of a univariate approach is founded in two interacting factors. Firstly, the reaction kinetics of the used redox system are pH dependent. This pH dependency has a significant influence on the refolding yield as increased rates of disulfide bridges result in faster formation of correctly refolded protein and thereby reduce the competing formation of aggregation [14]. Secondly, the amounts of reducing and oxidizing agent are not only independently influencing the refolding yield, but also exhibit an interaction factor. To counteract, e.g., larger amounts of reducing agent during solubilization, larger amounts of oxidizing agent are needed during refolding. Therefore, in order to identify these interactions, a unit operation-spanning multivariate approach is necessary. For the exemplary workflow shown here, dithiothreitol (DTT) is used as the reducing and glutathione disulfide (GSSG) as the oxidizing agent to optimize the redox system for horseradish peroxidase (HRP) IBs. A basic refolding protocol (using fixed redox conditions) has previously been established. The amount of active HRP after refolding was measured via the enzymatic activity using 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) as a substrate [15]. If used for a different POI, the type of reducing and oxidizing agent can easily be adapted (using, e.g., cysteine/cystine) [16]. However, a basic refolding protocol (developed using fixed redox conditions) and an analytical method to quantify correctly folded POI after refolding are required in order to successfully perform the workflow.

## 2 Materials

### 2.1 Prerequisites

In order to perform the described workflow, several materials and methods have to be available:

1. Washed and aliquoted IBs containing the POI, stored at appropriate conditions (e.g.,  $-20^{\circ}\text{C}$ ) (*see Note 1*).
  - Further details in the book chapter **High pressure homogenization for inclusion body isolation**
2. An established refolding protocol for the POI (excluding the redox system). This should include a solubilization method and a refolding buffer suitable to stabilize the POI during refolding. Such a protocol is usually established using fixed redox conditions in order to optimize other factors before varying the redox conditions. The established conditions can then be further optimized with the found redox conditions in an iterative approach.
3. A suitable analytical method in order to measure the quantity of correctly folded protein after the refolding step (*see Note 2*).

In this case, the workflow is shown by the example of HRP IBs [15]. A working solubilization and refolding protocol was in place, which is represented by the listed buffers and conditions. Washed and aliquoted IBs were available, and the concentration (g POI/g wet IB (wIB)) was measured and calculated using a reversed phase—high-performance liquid chromatography (RP-HPLC) method [17, 18].

## 2.2 Unit Operations

1. Solubilization:
  - Solubilization buffer 1: 50 mM Tris/HCl pH 8.5; 6 M urea.
  - Solubilization buffer 2: 50 mM glycine pH 10; 6 M urea.
  - DTT stock solution: 1 M DTT dissolved in solubilization buffer 1 or solubilization buffer 2 (*see Note 3*).
  - A dispersing instrument suitable for 2 mL and 15 mL reaction tubes (e.g., T10 basic ULTRA-TURRAX (IKA, Staufen, Germany)) (*see Note 4*).
  - Table centrifuge for 2 mL and 15 mL reaction tubes.
2. Refolding:
  - Refolding buffer 1: 20 mM Tris/HCl pH 8.5; 2 M urea; 2 mM  $\text{CaCl}_2$ ; 7% v/v glycerol; varying GSSG concentrations (*see Note 3*).
  - Refolding buffer 2: 20 mM glycine pH 10; 2 M urea; 2 mM  $\text{CaCl}_2$ ; 7% v/v glycerol; varying GSSG concentrations.
  - Hemin stock solution: 1 mM hemin dissolved in 100 mM KOH (*see Note 5*).
  - Rocker-shaker (e.g., Biosan mini rocker-shaker MR-1).
  - Fridge or cold room suitable to place and run the rocker-shaker at  $4^{\circ}\text{C}$  (*see Note 6*).
  - Centrifuge for 2 mL and 15 mL reaction tubes.



### 2.3 Analytics

RP-HPLC or SDS-PAGE are primarily used in order to measure the amount of POI in the solubilizate. If the amount of POI/g wIBs is constant for the aliquoted IBs, these methods can be omitted (*see Note 1*).

#### 1. RP-HPLC:

- An HPLC device consisting of a pump module, an auto-sampler module, a column oven, and a UV detector.
- Mobile phase A: HPLC grade water (MQ) supplemented with 0.1% trifluoroacetic acid (TFA).
- Mobile phase B: HPLC grade acetonitrile (ACN) supplemented with 0.1% TFA.
- RP column (e.g., BioResolve RP mAb Polyphenyl column (Waters Corporation, MA, USA)).
- Rocker-shaker.
- Software for analysis of chromatograms.

#### 2. SDS-PAGE:

- SDS gels (e.g., Mini-PROTEAN TGX Precast Gels 4–15% (Bio-Rad, CA, USA)).
- Gel electrophoresis running system (e.g., Mini-PROTEAN Tetra System (Bio-Rad, CA, USA)).
- Laemmli buffer (e.g., 2× Laemmli sample buffer (Bio-Rad, CA, USA)) supplemented with a reducing agent (e.g., 2-mercaptoethanol).
- Coomassie staining (e.g., Bio-Safe Coomassie Stain (Bio-Rad, CA, USA)).
- Imaging device and according software (e.g., Gel Doc XR+ Gel Documentation System and Image Lab Software (both Bio-Rad, CA, USA)).

#### 3. Measurement of correctly folded protein:

The analytical method used to measure correctly folded protein after refolding is highly specific for each individual POI. Commonly used methods can include measurement of the enzyme activity (if the POI is an enzyme) or measurement of the binding kinetics (e.g., in case of antibody fragments) (*see Note 7*). In case of HRP shown in this protocol, the enzymatic activity is measured using a colorimetric assay using ABTS as substrate. The following instruments and solutions are used:

- Temperature-controlled spectrophotometer or plate reader able to record the change of absorbance at 420 nm over time.
- Multichannel pipet for a volume of 20 µL (per channel).
- Dilution buffer: 20 mM Bis-Tris pH 7, 7% v/v glycerol.

- ABTS solution: 8 mM ABTS in 50 mM phosphate-citrate buffer pH 5.
- $\text{H}_2\text{O}_2$  stock solution: 20 mM  $\text{H}_2\text{O}_2$  in water.

## 2.4 DoE

Software to design and subsequently interpret the experiments performed (e.g., MODDE 12 (Sartorius, Goettingen, Germany)) (*see Note 8*).

## 3 Methods

### 3.1 Experimental Planning and Design of Experiments

In order to identify suitable redox conditions leading to high refolding yields, the amount of reducing agent during solubilization and the amount of oxidizing agent during refolding have to be varied in combination with the pH value. To reduce the experiment number while still identifying quadratic terms and interaction terms, a multivariate approach in the form of a DoE can be used. An example for the experimental design applied for HRP is shown in Fig. 1 (*see Note 9*). However, if the DoE is to be adapted, the following points should be considered (*see Note 10*):

1. Decide on the factors that will be varied during the experiments. For the shown protocol, these are the reducing agent concentration (DTT) during solubilization and the oxidizing agent concentration (GSSG) as well as the pH value of the refolding buffer (*see Note 9*).

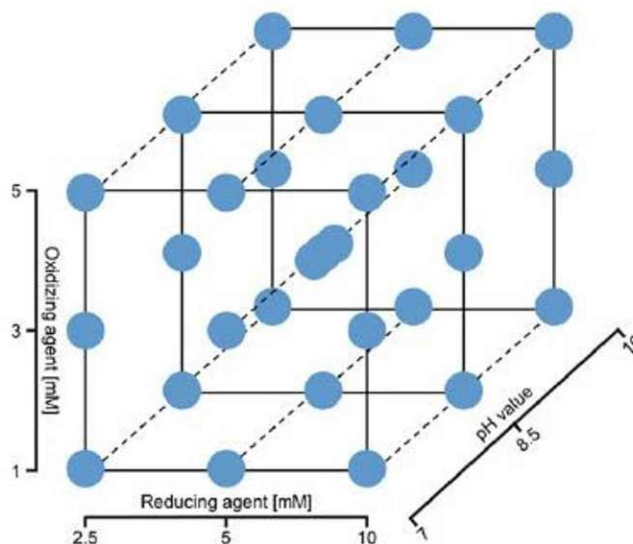


Fig. 1 Design space for factors used in the suggested DoE



2. Find feasible ranges for the factors. Figure 1 indicates ranges to test the redox system itself. However, the redox system can be highly protein dependent (influenced, e.g., by the number and location of disulfide bridges). Hence, a further adaptation of the ranges in a second DoE might be necessary (*see Note 10*).
3. Decide on the design of the DoE. For three factors, each with three levels, we would recommend a central composite design (CCD) such as the face centered design (CCF) shown in Fig. 1. The reason is that interaction and quadratic terms are expected, which cannot be analyzed by more reduced designs (e.g., fractional factorial designs). If other factors are to be included in the design, we recommend a screening design including the redox factors (DTT, GSSG, and pH) for the first estimation of factor ranges, followed by a more refined design used in a second experimental set.
4. Prior to buffer preparation (i.e., during the experimental planning phase): Think about splitting used factors accordingly and which buffers need to be prepared (*see Note 11*).

### 3.2 Performing the DoE

Buffer preparation: All buffers should be prepared fresh and prior to the start of solubilization.

- Solubilization buffer (*see Note 12*).
  - Refolding buffer:
 

After preparing all required refolding buffers, it is recommended to aliquot them into 2 mL reaction tubes according to the chosen conditions of the DoE (*see Note 13*).
  - DTT stock solution should be prepared fresh (prior to solubilization).
1. Solubilization:
    - Add solubilization buffer to the IB pellet in order to reach the desired concentration (e.g., 100 g wIB/L) (*see Note 14*).
    - Resuspend the pellet using a dispersing instrument (e.g., Ultraturrax at level 3) (*see Note 15*).
    - Add DTT stock to reach the desired concentration (*see Note 16*).
    - Place the solubilization mixtures on a rocker-shaker for the appropriate time span and condition (e.g., 0.5 h at RT for HRP).
    - Centrifuge the solubilization mixes (e.g., 20.380 rcf; 20 min; 4 °C).
    - Split the supernatant from the pellet and discard the pellet (*see Note 17*).

- **Sample:** If protein quantification in the different solubilization samples is to be performed, draw a sample of the supernatant for RP-HPLC or SDS-PAGE measurement (detailed description can be found in Subheading 3.3).
- Use the supernatant for the subsequent refolding step.

## 2. Refolding.

- Add solubilization supernatant to the 2 mL reaction tubes containing precooled refolding buffers (*see Note 13*). Invert each tube several times immediately after adding the solubilization mix in order to ensure sufficient mixing (*see Note 18*).
- Place the refolding mixes on a rocker-shaker at the desired conditions (4 °C; 16 h for HRP).
- Specific for refolding of HRP: After 16 h, add hemin stock solution to the refolding mix to reach a final concentration of 20 µM hemin (*see Note 5*).
- Centrifuge all refolding mixes (20.380 rcf; 4 °C; 20 min).
- Draw samples of all supernatants and quantify correctly folded protein (e.g., ABTS assay in case of HRP) (*see Note 7*).

## 3.3 Analytics

### 1. RP-HPLC:

#### Device Setup:

- Prepare mobile phase A (MQ + 0.1% TFA) and mobile phase B (ACN + 0.1% TFA).
- Precool the autosampler to 5 °C.
- Prepare reversed-phase column for analysis at 70 °C and a constant flow rate of 0.4 mL/min (runtime 18 min). For a detailed description of the used method, refer to [17, 18].

#### Samples:

- In general, the supernatants drawn after the solubilization and refolding step can directly be measured, if free of particles (*see Note 19*).
- Measure samples and, if available, perform a standard calibration using the pure POI in appropriate concentrations (e.g., 0.1–2 g/L).
- Calculate protein concentration and purity using a suitable chromatographic software.

### 2. SDS-PAGE:

- Dilute the sample 1:2 in 2× Laemmli (e.g., 50 µL sample + 50 µL Laemmli buffer).
- Heat to 95 °C for 10 min.



- Spin down all samples.
  - Dilute samples to the appropriate protein concentration.
  - Perform SDS–PAGE.
  - Use Coomassie staining and an appropriate Imaging device. Densitometric analysis of the SDS–PAGE can be performed with a suitable software (*see* Subheading 2.3).
3. Activity measurement: In the case of HRP, the enzyme activity after the refolding process was measured using a colorimetric assay with ABTS as substrate in a 96-well plate.
- Dilute the refolding mixes 1:20 in dilution buffer.
  - Add 175  $\mu\text{L}$  ABTS solution into each well.
  - Add 5  $\mu\text{L}$  of diluted refolding mix to each well.
  - Start the reaction by adding 20  $\mu\text{L}$  of  $\text{H}_2\text{O}_2$  stock solution to start the reaction.
  - Measure the change of absorbance at 420 nm for at least 60 s.
  - Based on (1), calculate the volumetric activity [U/mL] for each sample.

$$A \left[ \frac{\text{U}}{\text{mL}} \right] = \frac{V_{\text{total}} * \Delta A * \text{dilution}}{V_{\text{sample}} * d * \epsilon} \quad (1)$$

$A$ —Volumetric activity

$V_{\text{total}}$ —total volume in well (200  $\mu\text{L}$ )

$\Delta A$ —change in absorption [ $\Delta A$  420 nm/min]

dilution—dilution of sample

$V_{\text{sample}}$ —volume of diluted sample in well (5  $\mu\text{L}$ )

$d$ —length of light path through well (0.58 cm)

$\epsilon$ —molar extinction coefficient ABTS (36/mM/cm)

### 3.4 Analysis of the Performed DoE

1. Calculate the refolding yield based on (2):

$$\gamma [\%] = \frac{A \left[ \frac{\text{U}}{\text{mL}} \right]}{C \left[ \frac{\text{mg}}{\text{mL}} \right] * A_{\text{ref}} \left[ \frac{\text{U}}{\text{mg}} \right]} * 100 \quad (2)$$

$\gamma$ —Refolding Yield [%]

$A$ —Volumetric activity in the refolding mix

$C$ —Protein concentration in the refolding mix

$A_{\text{ref}}$ —Specific activity [U/mg] of a reference standard (pure and fully active POI)

2. Evaluate the DoE using the calculated refolding yield as a response.
3. If the highest refolding yield is measured or predicted at the edge of the chosen design space, a second DoE with an adapted design space can be performed.

### 3.5 Validation (Reactor Setup)

A validation of the found redox conditions can be performed in a bench scale bioreactor, monitoring the redox potential inline using an appropriate probe.

Further details in the book chapter **The purification of heme peroxidases from *Escherichia coli* inclusion bodies: A success story shown by the example of Horseradish peroxidase.**

## 4 Notes

1. It is recommended that a single batch of washed and aliquoted IBs is used for all experiments in order to provide reproducibility. Furthermore, the correlation between wet IB weight and POI in the solubilizate can be established, mitigating the need for at-line solubilization analytics to measure the protein concentration in the solubilizate for every solubilization mix.
2. The analytical method to measure correctly folded protein is highly dependent on the POI. Since it is a crucial part of the shown workflow, it has to be established before the workflow can be used.
3. The shown redox system with DTT and GSSG was used in the case of HRP. If a different redox system is to be used for the POI, the workflow can easily be adapted by substituting DTT with the reducing agent of choice (e.g., cysteine) and GSSG with the oxidizing agent of choice (e.g., cystine).
4. We recommend the listed dispersing instrument as it fits both 2 mL and 15 mL reaction tubes, which were used for solubilization. If different solubilization volumes are planned, the dispersing instrument can be adapted accordingly.
5. The hemin stock is specific for the refolding of peroxidases such as HRP since hemin is required to form the holoenzyme.
6. The suggested rocker-shaker can be used in combination with a commonly available fridge. The power supply cable can be connected to an electrical socket outside of the fridge and the cable is run inside through the sealing of the door. However, care should be taken in order to guarantee that the fridge door is still properly closing when the power cable is run through.
7. Care should be taken in order to ensure that the used refolding buffers do not interfere with the measurement method for correctly folded protein (e.g., the enzymatic activity might be



dependent on the pH value and the enzyme is therefore only active at a different pH value than present during refolding). If that is the case, an additional sample preparation step (e.g., buffer exchange using size exclusion chromatography or dialysis) has to be performed.

8. In general, the shown workflow can also be performed without using a multivariate data evaluation approach. However, for process understanding and accurate evaluation, we would highly recommend using such an approach including the according software. Note that one of the big advantages of using a DoE is that data from later experiments can also be included.
9. Several variations for the used factors are possible. Especially if the experimental effort for the refolding process is low and the measurement of correctly folded protein can be automated, more factors can be included.
10. For the factor of the pH value, the stability of the POI at the chosen pH should be considered. The reason for choosing high pH values can be found in the reaction kinetics of the redox partners (i.e., DTT and GSSG) [19]. If the POI is not stable at high pH values, several different redox reagents are available and can be used [20].
11. As an example, the buffer containing the 3 mM GSSG concentration can be obtained by mixing the buffer with the low GSSG concentration and the buffer with the high GSSG concentration in the appropriate ratio.
12. If the solubilization protocol requires cooling (e.g., 4 °C), the solubilization buffer should be precooled prior to use.
13. In our experience using a numbering system for the experimental conditions makes it easier to handle the large number of different experiments (e.g., for the labeling of the refolding mixes and the analytical methods). After the experimental design is finished, the conditions for all individual experiments are assigned consecutive numbers. If the experiments are sorted to cluster the different solubilization conditions, potential mistakes while pipetting can be avoided and the start of refolding is therefore made easier.
14. In our experience, using 5–10 mL solubilization mixtures in a 15 mL Falcon tube led to a reproducible solubilization procedure. If only a very limited amount of IBs is available, 2 mL reaction tubes can also be used. In the case that 2 mL reaction tubes are used, measuring the amount of solubilized protein (e.g., RP-HPLC) is highly recommended to account for varying amounts of protein that are carried over into the refolding mix.

15. Depending on the POI, the IBs might form flakes that are hard to solubilize. If it is not possible to use a dispersing instrument (e.g., due to small volumes used), a micro-spatula can be used to crush the flakes against the wall of the reaction tube, enhancing the surface area and thereby facilitating the solubilization.
16. The DTT stock should be freshly prepared and kept cool (4 °C) in order to prevent oxidation. If consecutive experiments are performed, we recommend to prepare a fresh DTT stock approximately every 2 h.
17. Depending on the POI and the solubilization protocol, a gel-like layer might be present between the pellet and the supernatant. In this case, carefully pipetting the supernatant into a fresh tube can be helpful in order to avoid carry over of the gel-like layer into the refolding mix. If this is the case, a higher solubilization volume might be necessary, as a significant amount of solubilization mix can be lost in the pellet.
18. If a large number of refolding experiments are performed at the same time, the refolding buffers should be cooled during the addition of solubilization mix (as precooling alone might not be sufficient). Therefore, either an ice bath or working in a cold room is recommended.
19. Storage of the samples at 4 °C and timely measurement is highly recommended as the solubilization might not be stable for long times. This might lead to skewed results of the protein content in the solubilization. Furthermore, if precipitation occurs over time, this might be harmful to the HPLC system and the used column.

## References

1. Eggenreich B et al (2016) Production strategies for active heme-containing peroxidases from *E. coli* inclusion bodies – a review. *Biotechnol Rep (Amst)* 10:75–83
2. Bhatwa A et al (2021) Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Front Bioeng Biotechnol* 9:630551
3. Humer D, Spadiut O (2018) Wanted: more monitoring and control during inclusion body processing. *World J Microbiol Biotechnol* 34(11):158
4. Rathore AS et al (2013) Refolding of biotech therapeutic proteins expressed in bacteria: review. *J Chem Technol Biotechnol* 88(10):1794–1806
5. Yamaguchi S, et al (2013) Protein refolding using chemical refolding additives. *Biotechnol J* 8(1): 17–+.
6. Yamaguchi H, Miyazaki M (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomol Ther* 4(1):235–251
7. Vallejo LF, Rinas U (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb Cell Factories* 3(1):11
8. Meneses-Acosta A et al (2013) Effect of controlled redox potential and dissolved oxygen on the in vitro refolding of an *E. coli* alkaline phosphatase and chicken lysozyme. *Enzym Microb Technol* 52(6–7):312–318
9. Wang SSS et al (2006) Effect of glutathione redox system on lysozyme refolding in size exclusion chromatography. *Food Bioprod Process* 84(1):18–27
10. Gao YG et al (2002) Refolding of lysozyme at high concentration in batch and fed-batch operation. *Korean J Chem Eng* 19(5):871–875



11. Eiberle MK, Jungbauer A (2010) Technical refolding of proteins: do we have freedom to operate? *Biotechnol J* 5(6):547–559
12. Walther C et al (2014) Getting ready for PAT: scale up and inline monitoring of protein refolding of Npro fusion proteins. *Process Biochem* 49(7):1113–1121
13. Hamada H, Arakawa T, Shiraki K (2009) Effect of additives on protein aggregation. *Curr Pharm Biotechnol* 10(4):400–407
14. Kiefhaber T et al (1991) Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology (N Y)* 9(9):825–829
15. Hummer D, Ebner J, Spadiut O (2020) Scalable high-performance production of recombinant horseradish peroxidase from *E. coli* inclusion bodies. *Int J Mol Sci* 21(13):4625
16. Raman B, Ramakrishna T, Rao CM (1996) Refolding of denatured and denatured/reduced lysozyme at high concentrations. *J Biol Chem* 271(29):17067–17072
17. Kopp J et al (2020) Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. *J Pharm Biomed Anal* 188:113412
18. Ebner J et al (2021) At-line reversed phase liquid chromatography for in-process monitoring of inclusion body solubilization. *Bioengineering (Basel)* 8(6):78
19. Chau M-H, Nelson JW (1991) Direct measurement of the equilibrium between glutathione and dithiothreitol by high performance liquid chromatography. *FEBS Lett* 291(2):296–298
20. Patil G, Rudolph R, Lange C (2008) In vitro-refolding of a single-chain Fv fragment in the presence of heteroaromatic thiols. *J Biotechnol* 134(3–4):218–221

## 6.3 Patent

### (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization

International Bureau

(43) International Publication Date  
30 September 2021 (30.09.2021)



(10) International Publication Number  
**WO 2021/191253 A1**

(51) International Patent Classification:

A61K 38/00 (2006.01) C07K 1/113 (2006.01)

(21) International Application Number:

PCT/EP2021/057532

(22) International Filing Date:

24 March 2021 (24.03.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

20165131.2	24 March 2020 (24.03.2020)	EP
20167716.8	02 April 2020 (02.04.2020)	EP
20177549.1	29 May 2020 (29.05.2020)	EP

(71) Applicant: **TECHNISCHE UNIVERSITÄT WIEN**  
[AT/AT]; Karlsplatz 13, 1040 Wien (AT).

(72) Inventors: **HUMER, Diana**; c/o Technische Universität Wien ICEBE Institut Gumpendorferstrasse 1A, 1060 Wien (AT). **SPADIUT, Oliver**; c/o Technische Universität Wien ICEBE Institut Gumpendorferstrasse 1A, 1060 Wien (AT). **EBNER, Julian**; c/o Technische Universität Wien ICEBE Institut Gumpendorferstrasse 1A, 1060 Wien (AT).

(74) Agent: **SONN & PARTNER PATENTANWÄLTE**;  
Riemergasse 14, 1010 Wien (AT).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: METHODS FOR PRODUCING HEME PEROXIDASES

(57) Abstract: The present invention provides methods for producing a heme pe- roxidase from inclusion bodies (IBs) comprising the steps of: providing the heme peroxidase in the form of IBs, solubilizing said IBs, transferring said solubilized IBs into a refolding buffer to obtain a refolding mix, adding a heme cofactor to said refolding mix, wherein the addition of the heme cofactor to the refolding mix is distributed over a time period of at least 1 hour. The invention further provides methods for producing heme peroxidase products.

WO 2021/191253 A1





## Article

# Scalable High-Performance Production of Recombinant Horseradish Peroxidase from *E. coli* Inclusion Bodies

Diana Humer <sup>†</sup>, Julian Ebner <sup>†</sup> and Oliver Spadiut <sup>\*†</sup>

TU Wien, Institute of Chemical, Environmental and Bioscience Engineering, Research Area Biochemical Engineering, Gumpendorfer Straße 1a, 1060 Vienna, Austria; diana.humer@tuwien.ac.at (D.H.); julian.ebner@tuwien.ac.at (J.E.)

\* Correspondence: oliver.spadiut@tuwien.ac.at

† These authors contributed equally to this work.

Received: 21 May 2020; Accepted: 26 June 2020; Published: 29 June 2020



**Abstract:** Horseradish peroxidase (HRP), an enzyme omnipresent in biotechnology, is still produced from hairy root cultures, although this procedure is time-consuming and only gives low yields. In addition, the plant-derived enzyme preparation consists of a variable mixture of isoenzymes with high batch-to-batch variation preventing its use in therapeutic applications. In this study, we present a novel and scalable recombinant HRP production process in *Escherichia coli* that yields a highly pure, active and homogeneous single isoenzyme. We successfully developed a multi-step inclusion body process giving a final yield of 960 mg active HRP/L culture medium with a purity of ≥99% determined by size-exclusion high-performance liquid chromatography (SEC-HPLC). The Reinheitszahl, as well as the activity with 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) and 3,3',5,5'-tetramethylbenzidine (TMB) as reducing substrates, are comparable to commercially available plant HRP. Thus, our preparation of recombinant, unglycosylated HRP from *E. coli* is a viable alternative to the enzyme from plant and highly interesting for therapeutic applications.

**Keywords:** *E. coli*; inclusion bodies; recombinant horseradish peroxidase; refolding; hydrophobic interaction chromatography; reversed-phase high-performance liquid chromatography (RP-HPLC); size exclusion HPLC (SEC-HPLC)

## 1. Introduction

Horseradish peroxidase (HRP) (EC 1.11.1.7) has high industrial relevance entailing an immense pool of published data, ranging from refolding of denatured plant enzyme to recombinant production in different hosts [1–7]. HRP is an oxidoreductase that contains heme as a cofactor, two calcium ions and four disulfide bridges [6,8]. This plant-derived enzyme naturally occurs in the horseradish root (*Armoracia rusticana*), where it is glycosylated usually at eight of nine asparagine sites [9,10]. Due to this glycosylation, plant HRP (pHRP) has a size of 44 kDa, but when the enzyme is produced in *Escherichia coli* (rHRP) it is unglycosylated resulting in a size of only 34.5 kDa [11]. The natural enzyme in *A. rusticana* has a large variety of isoforms, with 34 different entries in the UniProtKB database (April 2020) [1]. To date, HRP for industrial use is still extracted from the plant; therefore, the composition and abundance of the isoenzymes in the final product is diverse. The application range of HRP is immensely versatile; first and foremost, the protein is used as reporter enzyme because it reacts with many substrates to give a chromogenic, fluorogenic or electrochemical signal. Therefore, HRP is used for nucleic acid, antibody and protein labelling in immunoassays, diagnostic kits and microarrays [12–14]. The enzyme can also function as a biosensor, which can be applied for the detection of hydrogen peroxide, but also for other substances, such as glucose, L-phenylalanine, ethanol or hydroquinone [15–20].

In the field of environmental protection, HRP is used for the bioremediation of pollutants (mostly phenols) and for the decolorization of dyes in the textile industry [21–23]. In chemistry, the protein is employed for organic polymer synthesis and biocatalysis [24–29]. Another interesting field with high potential is the biomedical use of HRP, for example in medical diagnostics [30,31] or targeted cancer treatment [32,33]. However, these applications require an HRP preparation that complies with pharmaceutical standards, available as a steady supply in large amounts. These requirements are met in a recombinant production strategy, where a single isoenzyme can be produced homogeneously and, if necessary, without glycosylation. As mentioned above, HRP is unglycosylated in *E. coli* and therefore well suited for medical applications, as foreign glycosylation patterns can cause immune reactions in humans [34]. Recombinant protein production in *E. coli* follows two major routes: the protein of interest can either be soluble or insoluble in the form of protein aggregates, also known as inclusion bodies (IBs). In the *E. coli* cytoplasm, HRP forms IBs because the absence of glycans on the protein surface results in enhanced hydrophobicity and the disulfide bridges cannot be formed. Both production strategies pose several advantages as well as disadvantages and are dependent on protein properties as well as cultivation parameters [35,36]. While initially IBs were regarded as undesirable, they show many advantages for production, such as high titers, high purities after recovery and the potential to express proteins that would otherwise be toxic to the host cell [37,38]. However, the downstream process (DSP) is more elaborate as the protein must be refolded to the native state. In 2018, we published a mini-review that summarizes the current developments and analytical tools for refolding [39]. In general, IBs are harvested by centrifugation, followed by a washing step and solubilization with denaturing agents, such as chaotropic substances (urea, guanidine hydrochloride (GuHCl)) or detergents (Triton X-100, sodium dodecyl sulfate). The solubilized protein is then transferred to a buffer where the correctly folded protein is energetically favored over the unfolded protein. This can be done in different ways, with the most common methods being batch dilution, fed-batch dilution and on-column refolding. Depending on the characteristics of the protein of interest, different factors have to be considered for successful refolding. For the correct refolding of peroxidases, the establishment of the redox system and the method, as well as the time point of cofactor addition, are essential. In Table 1 we present a comparison of refolding attempts for different class II and III peroxidases. Concerning HRP, the first refolding experiments were performed more than 30 years ago by Smith et al. [40], where a yield of 3% with 630 U/mg with ABTS as substrate and a protein concentration of 0.057 mg/mL was reported. Grigorenko et al. [3] performed HRP refolding from IBs using a His tagged protein and IMAC (immobilized metal affinity chromatography) as capture step after refolding. The refolding experiment resulted in 6–8 mg HRP/L culture medium with a specific activity of 1160 U/mg for ABTS as reducing substrate.

**Table 1.** Overview of the yield and specific activity of class II and III plant peroxidases refolded from inclusion bodies.

Enzyme	Yield (% or mg/L Culture)	Specific Activity (U/mg)	Reference
HRP	3%	630 U/mg (ABTS)	Smith et al. [40]
HRP	6–8 mg/L	14500 U/mg (pyrogallol)	Grigorenko et al. [3]
HRP	24%	1160 U/mg (ABTS)	Asad et al. [41]
HRP	16.7 mg/L	10 U/mg (4-aminoantipyrine)	Gazaryan et al. [42]
HRP	20 mg/L	4000 U/mg (ABTS)	Gazaryan et al. [43]
HRP	15 mg/L	2000 U/mg (ABTS)	Gundinger et al. [44]
HRP	15 mg/L	62.5 U/mg (ABTS)	Kim et al. [45]
CWPO_C <sup>1</sup>	0.2 mg/L	1000 U/mg (ABTS)	Kim et al. [45]
CWPO_C <sup>1</sup>	27.3%	1066 U/mg (syringaldazine)	Shigeto et al. [46]
rAtPrx71	28%	120 U/mg (guaiacol)	Shigeto et al. [47]
rAtPrx25	30.3%	1291 U/mg (syringaldazine)	Shigeto et al. [47]
rAtPrx25	30.3%	270 U/mg (guaiacol)	Shigeto et al. [47]



Table 1. Cont.

Enzyme	Yield (% or mg/L Culture)	Specific Activity (U/mg)	Reference
TOP <sup>2</sup>	79 mg/L	2950 U/mg (ABTS)	Zakharova et al. [48]
LDP <sup>3</sup>	16.8 mg/L	70.7 U/mg (TMB)	Fattahian et al. [49]
TOP <sup>2</sup>	4.6 mg/L	580.7 U/mg (H <sub>2</sub> O <sub>2</sub> )	Hushpalian et al. [50]
ATPN <sup>4</sup>	13 mg/L	1100 U/mg (ABTS)	Teilum et al. [51]
BP1 <sup>5</sup>	9.4 mg/L	n.m.	Teilum et al. [51]
LiP H8 <sup>6</sup>	1%	39 µmol of veratryl alcohol ox/min/mg of protein	Doyle et al. [52]
VPL2 <sup>7</sup>	5.5 mg/L	n.m.	Pérez-Boada et al. [53]
LiP H2 <sup>6</sup>	3.4 mg/L	n.m.	Nie et al. [54]
VBPO <sup>8</sup>	40 mg/L	550 U/mg (bromination of monochlorodimedone)	Coupe et al. [55]
LiP <sup>6</sup>	0.38 mg/L	16,300 U/mg (ABTS)	Miki et al. [56]
DyP <sup>9</sup>	1.5 mg/L	247 U/mg (ABTS)	Linde et al. [57]
MnP <sup>10</sup>	2.4%	12.9 U/mg (oxidation of Mn <sup>2+</sup> to Mn <sup>3+</sup> )	Wang et al. [58]
LiP H2 <sup>6</sup>	2.4%	55.6 U/mg (veratryl alcohol)	Lee et al. [59]
MnP <sup>10</sup>	0.275 mg/L	140 U/mg (oxidation of Mn <sup>2+</sup> to Mn <sup>3+</sup> )	Whitwam et al. [60]
BnPA <sup>11</sup>	29 mg/L	981 U/mg (ABTS)	Rodriguez-Cabrera et al. [61]

<sup>1</sup> Cationic cell wall peroxidase from *Populus alba* L.; <sup>2</sup> Tobacco peroxidase; <sup>3</sup> *Lepidium draba* peroxidase; <sup>4</sup> *Arabidopsis thaliana* peroxidase N; <sup>5</sup> Barley grain peroxidase; <sup>6</sup> Lignin peroxidase; <sup>7</sup> *Pleurotus eryngii* versatile peroxidase; <sup>8</sup> Vanadium-dependent bromoperoxidase; <sup>9</sup> Dye-decolorizing peroxidase; <sup>10</sup> Manganese peroxidase; <sup>11</sup> Turnip acidic peroxidase; n.m., not mentioned.

Asad et al. [41] published a comprehensive investigation on the refolding conditions of horseradish peroxidase, focusing on the buffer system, redox conditions and additives suitable for HRP stabilization. They used one factor at a time and response surface methodology (RSM) to obtain a yield of 3.6 mg HRP from 15 mg solubilized protein. The highest enzyme activity was 10 U/mg with 4-aminoantipyrine as substrate. In 2016, we performed a comparative study on the production of HRP both in a soluble form and from IBs. We obtained a relatively high yield of 150 mg HRP/L culture medium for refolded HRP with a specific activity of 62.5 U/mg (ABTS), but unfortunately, only 15 mg/L could be recovered after concentration with spin filters [44]. The translocation of HRP to the periplasm resulted in a final yield of 28 mg HRP/L culture medium and 12.7 U/mg with ABTS as reducing substrate. Thus, all previous production strategies resulted in low yields and/or decreased enzyme activity and stability. This is also the case for many other peroxidases (Table 1), where the highest yield was reported for Tobacco peroxidase with 79 mg/L culture medium.

In this study we present a novel production process for HRP from *E. coli* IBs with upscaling potential to industrial dimensions. Moreover, we assume that this procedure can be transferred to other peroxidases of the plant peroxidase superfamily, as all enzymes of class II and III are monomers and contain a non-covalently bound heme, two calcium ions and four disulfide bridges. As shown in Table 1, several studies concerning the refolding process of HRP have previously been published, resulting in established ranges for several important parameters, for example the urea concentration during solubilization and refolding. For this study, we focused on the influence of several parameters on the overall process, resulting in an integrated approach. This integrated approach of solubilization and refolding was found to be necessary and highly beneficial. Furthermore, the time of addition and the concentration of hemin had a significant influence on the refolding yield. The best results were obtained with a slow addition of hemin to the apo-enzyme. Previous approaches of HRP purification from IBs used either IMAC, SEC (size-exclusion chromatography) or CEX (cation exchange chromatography) as a capture step after refolding [3,40–42,44]. In this study, we developed a novel protocol for salt precipitation and hydrophobic interaction chromatography (HIC). The combination of these optimized conditions resulted in a final process yield of 959 mg active HRP/L culture medium

with a purity of  $\geq 99\%$  determined by SEC-HPLC. The enzymatic activity for the substrates ABTS and TMB, as well as the Reinheitszahl (Rz), are comparable to the plant-derived enzyme. The developed process allows the scalable production of the unglycosylated, single isoenzyme HRP C1A at the highest yield reported thus far (Table 1).

## 2. Results and Discussion

In this study, we developed a novel production process for recombinant HRP from *E. coli* inclusion bodies using a series of multivariate experiments and an integrated approach. Table 2 summarizes the parameters and ranges examined for each unit operation.

**Table 2.** Overview of investigated unit operations and the corresponding process parameters.

Unit Operation	Parameters	Range
Solubilization	DTT	2.5 mM–28.44 mM
	Protein concentration	20 g/L–80 g/L
	pH	7–10
Refolding	GSSG	0.4 mM–3.5 mM
	Protein concentration	0.5 g/L–2 g/L
	pH	7–10
	Time of hemin addition	0 h–24 h after refolding start
Salt precipitation	Hemin concentration	6 $\mu$ M–80 $\mu$ M
	Type of salt	NaCl, $(\text{NH}_4)_2\text{SO}_4$
	Salt concentration	0 M–4 M
Capture step HIC	Hydrophobicity of resin	Octyl, Butyl, Phenyl
	pH value (load)	8.5, 10
	Type of elution	Step gradient, linear gradient

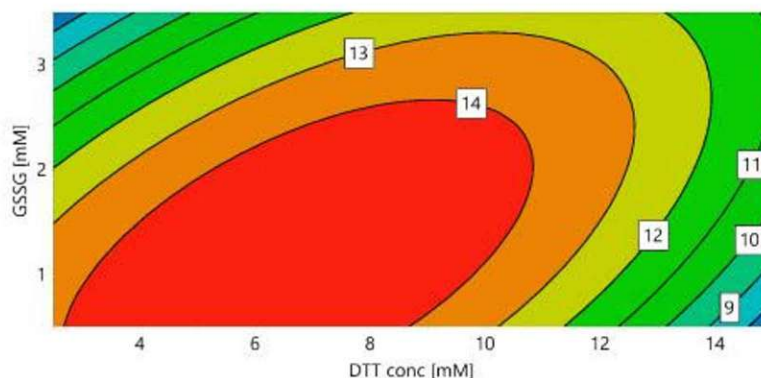
DTT, dithiothreitol; GSSG, glutathione disulfide; NaCl, sodium chloride;  $(\text{NH}_4)_2\text{SO}_4$ , ammonium sulfate.

### 2.1. Solubilization and Refolding

In order to identify the optimal redox system for solubilization and refolding, a DoE for DTT and GSSG concentration was performed (“DoE 1”). During solubilization all disulfide bridges should be reduced, granting the complete solubilization of the protein. In the subsequent refolding step, the correct disulfide bridges have to be formed, requiring an oxidizing environment. We hypothesized that these two unit operations are dependent on each other, as the DTT concentration had an influence on the refolding yield even if the solubilization yield stayed constant. In order to provide an integrated approach, the final refolding yield (determined as the volumetric activity) was used as a response for both DTT and GSSG concentration. Figure 1 shows the contour plot for DoE 1, with 7.11 mM DTT and 1.27 mM GSSG resulting in the highest yield for a protein concentration of 0.5 g/L during refolding.

For the used dilution of 1:40 solubilization mix in refolding buffer, this is equivalent to a GSSG:DTT ratio of 7:1 in the refolding mix. The interaction of DTT and GSSG is clearly visible, with increasing DTT requiring increasing GSSG concentrations. This shows the importance of optimizing the redox system in an integrated approach including solubilization and refolding and using the final refolding yield as response. Note that the absolute values of DTT and GSSG presented here might vary depending on the properties of the IBs, which are influenced by several factors, such as fermentation conditions (e.g., temperature, induction time and strength etc.) and unit operations performed before solubilization (e.g., homogenization, separation of IBs from soluble proteins, IB wash and also steps such as freezing and storage).





**Figure 1.** Response contour plot for the volumetric activity (U/mL) with ABTS as substrate, dependent on the DTT concentration in the solubilization mix and the GSSG concentration in the refolding buffer (DoE 1).

Based on these results, a second DoE ("DoE 2") was performed, using DTT, GSSG and protein concentration as factors, and volumetric activity (U/mL) and specific activity (U/mg) as responses. These responses were chosen since lower protein concentrations usually lead to higher refolding yields [62]. However, low protein concentrations require large buffer volumes and accordingly large vessels, which is a drawback for subsequent unit operations. The originating costs are a major downside of industrial IB processes, thus making the protein concentration a highly relevant factor. Initially, we assumed that the optimized redox conditions might be dependent on the protein concentration, with higher protein concentrations requiring higher DTT concentrations during solubilization. Contrary to this hypothesis, the DTT concentration resulting in the highest yield actually decreased for higher protein concentrations, with the GSSG concentration staying almost constant (Table 3). The optimized conditions for 0.5 g/L were 17.2 mM DTT and 2.2 mM GSSG, differing significantly from the results in DoE 1 (7.11 mM DTT and 1.27 mM GSSG). The reason for this is the strong contribution of the protein concentration as a factor, especially for the specific activity (Table 3 and Supplementary Figure S1). It is therefore difficult to predict optimal DTT and GSSG concentrations independent of the protein concentration. While this restricts the use of this particular experimental design for the optimization of the redox system, it clearly shows the importance of the investigation of the protein concentration during refolding. The highest specific activity in DoE 2 was achieved at a concentration of 0.5 g/L. All higher protein concentrations resulted in lower refolding yields (lower specific activities (U/mg)), therefore requiring a larger amount of IBs for the same amount of correctly folded HRP. The volumetric activity showed the opposite trend, rising with the concentration of protein in the refolding mix, with the highest volumetric activity achieved at 9.4 mM DTT, 2.2 mM GSSG and 1.6 g/L protein (26.9 U/mL) (Supplementary Figure S2). As expected, the specific activity at these conditions was low, with only 16.8 U/mg, which might result in problems during the subsequent capture step due to high concentrations of incorrectly folded protein. Furthermore, these conditions required three times more IBs for only a 20% increase in refolding yield. Based on these results, a protein concentration of 0.5 g/L during refolding was chosen for subsequent experiments. However, for an industrial production process the economically most feasible protein concentration might be different. For known USP (upstream process) and DSP costs, the presented models can be used to calculate the sweet spot between low refolding yields (requiring more IBs and raising the cost of the USP) and low concentrations of correctly folded HRP after refolding (requiring larger buffer volumes during refolding and subsequent capture steps resulting in higher DSP costs).

**Table 3.** Optimized DTT and GSSG concentrations for different protein concentrations based on the model obtained in DoE 2. The factor contribution is shown in brackets.

Protein Conc. (g/L)	Optimized DTT (mM)	Optimized GSSG (mM)	Specific Activity (U/mg)	Volumetric Activity (U/mL)
0.5 (93.2%)	17.19 (1.5%)	2.16 (5.3%)	45.1	22.5
1 (91.5%)	13.49 (1.8%)	2.17 (6.7%)	26.6	26.6
1.5 (88.5%)	9.83 (2.5%)	2.18 (9.0%)	18.9	28.4
2 (81.0%)	7.11 (6.0%)	2.21 (13.0%)	13.7	27.3

Asad et al. [41] previously reported that “the effect of the pH on HRP refolding was not significant over the range of 7–10”. These experiments were performed with a one factor at a time approach and the concentrations of DTT and GSSG were kept constant for all tested pH values. However, the reaction kinetics of DTT and GSSG are pH dependent, with slow reaction kinetics for pH values < 9. In order to cover this pH dependency of the redox pair, we performed a CCF DoE with the factors DTT concentration, GSSG concentration and pH value (pH 7–10) with the volumetric activity as response (“DoE 3”). The protein concentration during refolding was kept constant at 0.5 g/L for all experiments. Refolding at pH 7 resulted in very low refolding yields for all combinations of DTT and GSSG, so that the pH value was the dominant factor for the model. This, in turn, led to bad predictions for DTT and GSSG at pH 8.5 and 10. Therefore, the experiments at pH 7 were excluded from the model in order to obtain a correct depiction of the influence of DTT and GSSG concentrations at higher pH values. Comparing pH 8.5 and pH 10, different DTT and GSSG concentrations were required in order to achieve maximum refolding yields, underlining the importance of the multivariate approach (Supplementary Figure S3). The overall highest refolding yield was achieved at 6.7 mM DTT and 1.26 mM GSSG and pH 10, which resulted in an increase of 25% compared to refolding at pH 8.5. Furthermore, at pH 10, deviations of the DTT and/or GSSG concentration have less influence on the refolding yield. Based on the results of DoE 2 and DoE 3, we chose 7.11 mM DTT, 1.27 mM GSSG, 0.5 g/L protein concentration in the refolding buffer and pH 10 as the optimized conditions for solubilization and refolding.

The experiments in DoEs 1–3 were all performed in 2 mL reaction tubes, resulting in a maximum volumetric activity of 56 U/mL after refolding. In order to identify possible scale-up effects, these optimized conditions were scaled up to a refolding vessel with a volume of 1200 mL (“refolding vessel experiment 1”), resulting in a higher volumetric activity of 84 U/mL. Due to the tendency of protein aggregation during refolding, we assumed that favorable stirring conditions and surface to volume ratio led to this increase in activity for the upscaled process. Overall, the presented optimization approach resulted in a scalable and robust process with high refolding yields.

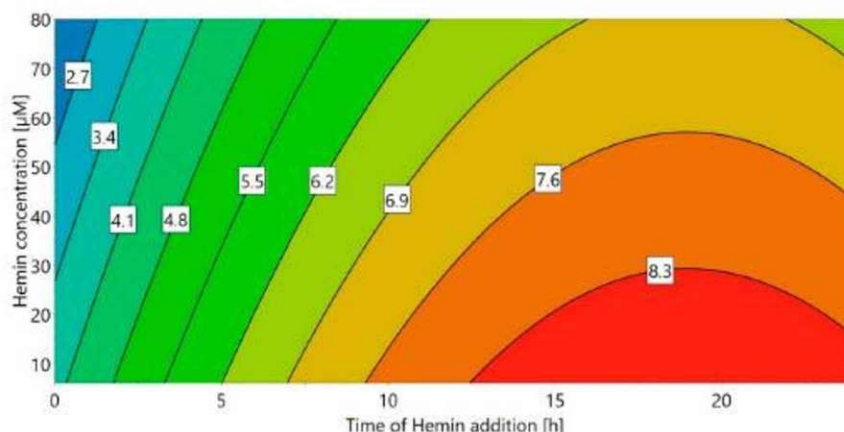
## 2.2. Hemin Addition

The concentration and time of hemin addition is another important parameter during refolding. Hemin has to be supplied in order to obtain holo-HRP; however, it is not necessary for the correct formation of apo-HRP. Previous studies showed that higher refolding yields could be achieved if hemin was added after the refolding step [41,42,44,63]. We suspect that this is the case due to the hydrophobic nature of hemin, promoting aggregation if added early on during refolding. Therefore, we assumed the time point of hemin addition, as well as the concentration, to have an influence on the refolding yield. The first experiments to test this hypothesis were performed using 2 mL reaction tubes, and optimized conditions were then validated in a bench-scale refolding vessel.

A DoE approach (“DoE 4”) was chosen for the small-scale experiments, using the concentration of hemin as well as the time of hemin addition as factors. The volumetric activity (U/mL) after refolding was used as the response (Figure 2). When hemin was added immediately after refolding start (0 h), the refolding yield strongly depended on the hemin concentration, with higher hemin concentrations leading to an over 10-fold reduced yield. This effect was decreased for later hemin addition times—in fact, when hemin was added 24 h after the start of the refolding process, the influence of the hemin



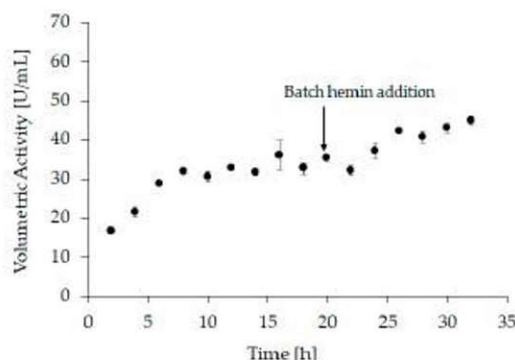
concentration was negligible. This is most likely a result of the hydrophobic nature of hemin, promoting aggregation of unfolded HRP, especially early during the refolding process. Therefore, the optimal time for hemin addition appears to be 6 h or longer after the start of refolding, with the maximum refolding yield achieved for a concentration of 6  $\mu\text{M}$  hemin after around 19 h. Although 6  $\mu\text{M}$  hemin was sufficient and higher hemin concentrations did not lead to an increase of the refolding yield, 20  $\mu\text{M}$  were chosen for all further experiments in order to prevent hemin to become a limiting factor for potential higher refolding yields in the bench-scale refolding vessel.



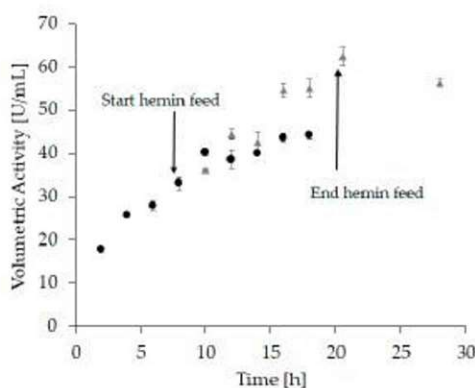
**Figure 2.** Response contour plot for DoE 4 with different times of hemin addition and hemin concentration as factors and the volumetric activity (U/mL) as response.

The optimal conditions were transferred to a bench-scale refolding vessel with a refolding volume of 1200 mL ("refolding vessel experiment 2"). For this experiment, hemin was added after 20 h of refolding to a final concentration of 20  $\mu\text{M}$ . Additionally, the refolding kinetics was monitored during this experiment. Samples were taken every 2 h to measure the activity at-line (Figure 3 and Supplementary Table S1). However, since hemin is required to form holo-HRP, 20  $\mu\text{M}$  hemin were added to each sample immediately after sampling and the activity was measured after incubation for two hours. After the addition of hemin to the refolding vessel (i.e., after 20 h), samples were still taken every 2 h, but no further hemin was added. Still, the samples were incubated for 2 h before measurement. As we see in the at-line activity measurement, refolding was completed after approximately 8 h. These observations are in good agreement with the results obtained from small-scale experiments and underline the importance of addition of hemin after the refolding process is finished.

Based on these experiments, we assumed that a linear hemin feed starting after complete apo-enzyme formation might further improve the refolding yield. The hypothesis was that hemin is prone to self-aggregation as well as aggregation with incorrectly folded protein, leading to a reduced amount of hemin available to form holo-HRP. Moreover, hemin aggregates could potentially impede further downstream applications, e.g., through irreversible binding to chromatography columns. A steady supply of small amounts on the other hand can immediately be used for holo-enzyme formation while aggregates are kept at a minimum. Therefore, in "refolding vessel experiment 3" a linear feed commencing 8 h after refolding starts with a total feeding time of 12 h was applied. Again, samples were taken every two hours (Figure 4 and Supplementary Table S2).



**Figure 3.** At-line sampling of “refolding vessel experiment 2” with volumetric activity in U/mL as response, samples 1–9 (2 h to 18 h) were taken before hemin was added to the refolding vessel therefore, 20  $\mu$ M were added afterwards and samples were incubated for another 2 h before activity measurement. Samples 10–16 (20 h to 32 h) were drawn after hemin addition but were still incubated for 2 h before activity measurement. All samples were measured in triplicates, with an average standard deviation < 6%.



**Figure 4.** At-line sampling of “refolding vessel experiment 3” with volumetric activity (U/mL) as response. The hemin feed was started after 8 h and lasted 12 h. Circles represent samples taken before and after the start of the feed, where hemin was added to each sample to reach a final concentration of 20  $\mu$ M. Samples were incubated for another 2 h before activity measurement. Triangles represent samples that were measured directly after the samples were drawn with no further hemin addition. All samples were measured in triplicates, with an average standard deviation < 6%.

At the beginning of the hemin feed, the volumetric activity (Figure 4, sample 4) was similar to the activity measured in “refolding vessel experiment 2” (Figure 3, sample 4). This demonstrates a good reproducibility and comparability between these two experiments. After an incubation time of 12 h with hemin, the volumetric activity in refolding vessel experiment 2 reached 45 U/mL (Figure 3, sample 16; Supplementary Table S1), whereas it was significantly higher after 12 h of hemin feed with 62 U/mL (Figure 4, sample 6 grey triangles; Supplementary Table S2). Moreover, the refolding time was shortened, as the maximum volumetric activity for “refolding vessel experiment 2” was reached after 44 h (46.5 U/mL, Supplementary Table S1), whereas this was achieved already after 20.5 h in this experiment (62 U/mL, Supplementary Table S2). These data substantiated our hypothesis that a linear hemin feed leads to an increased refolding yield.



All results presented above were done at pH 8.5, since we performed these experiments after optimizing the redox system, but before, we identified pH 10 as most beneficial during refolding (Supplementary Figure S3). Therefore, we wanted to verify the applicability of these results at pH 10. Applying a hemin feed instead of batch addition at pH 10 in fact resulted in a 23% increase of the refolding yield. We therefore concluded that the effect of our hemin addition strategy was independent of the pH value.

### 2.3. Capture and Concentration

The last unit operation investigated in this study was the capture step after refolding. Several different approaches, including IMAC, CEX and SEC, have been reported thus far [3,40–42,44]. In this study, a novel approach using HIC was developed. HIC is suitable for other peroxidases [46,47], but has never been used for HRP purification before. This approach turned out to be highly beneficial because the binding conditions require high salt concentrations, which precipitated impurities, such as excess hemin and protein aggregates, while correctly folded HRP was stable in solution at high salt concentrations. This has the advantage that impurities are already separated before the capture step, resulting in a higher binding capacity as well as easier cleaning and regeneration of the chromatographic resin.

First, we compared  $(\text{NH}_4)_2\text{SO}_4$  and NaCl for protein precipitation with the goal of reaching efficient impurity separation and high HRP recovery in the supernatant at the same time. Due to their position in the Hofmeister series, the concentration was varied between 0 M and 1.5 M for  $(\text{NH}_4)_2\text{SO}_4$  and 0 M and 4 M for NaCl, respectively. The recovery of the volumetric activity after protein precipitation was 96% for 1 M  $(\text{NH}_4)_2\text{SO}_4$  and 95% for 4 M NaCl. Specific activity increased 2.5-fold for  $(\text{NH}_4)_2\text{SO}_4$  and 4.5-fold for NaCl (Table 4). Due to the higher purification factor and the compatibility for potential medical applications in the future, we chose NaCl as salt.

**Table 4.** Volumetric and specific activity as well as protein concentration and purification factor for salt precipitation using either 1 M  $(\text{NH}_4)_2\text{SO}_4$  or 4 M NaCl.

Salt	Volumetric Activity (U/mL)	Protein Conc. (g/L)	Specific Activity (U/mg)	Purification Factor
$(\text{NH}_4)_2\text{SO}_4$ 1 M	38.0	0.12	317	2.5
NaCl 4 M	44.3	0.09	492	4.5

After salt precipitation, we tested different HIC resins, pH values and elution strategies (Supplementary Figures S4–S8) and finally used a resin with medium hydrophobicity (HiTrap Butyl FF), a pH of 8.5 and a step elution strategy. Using this strategy, active HRP was concentrated more than 9-fold resulting in a final concentration of 0.5 g/L active HRP in the eluate (Table 5) with a purity  $\geq 99\%$ .

**Table 5.** Volume, protein concentration, specific activity and purification factor for salt precipitation followed by hydrophobic interaction chromatography.

	Volume (mL)	Protein Conc. (mg/mL)	Specific Activity (U/mg)	Purification Factor
Refolding end	n.a.	0.51	126	1
Load (after salt precipitation)	50	0.09	726	5.8
Active HRP fraction	4	0.50	1176	9.4

n.a., not applicable.

#### 2.4. pH 8.5 vs. pH 10

Although we clearly demonstrated that refolding at pH 10 leads to the highest refolding yield, two final process runs at pH 8.5 and pH 10 were performed in order to compare the overall process yields. As demonstrated in small-scale and several bench-scale experiments, the optimal redox potential for HRP refolding was established with a GSSG:DTT ratio of 7:1. The final refolding time was 19 h, with 8 h refolding before the start of a 10 h hemin feed (final concentration 20  $\mu$ M). When refolding was done at pH 10, the refolding mix was brought to pH 8.5 before 4 M NaCl were added for salt precipitation. After centrifugation, the HRP solution was captured on a HiTrap Butyl FF column and eluted at 75% buffer B (Supplementary Figures S7 and S8).

Table 6 summarizes the results for refolding at pH 8.5 compared to pH 10. Refolding at pH 10 significantly increased the refolding yield to 74% when compared to 44% at pH 8.5. In addition, the total yield of HRP was enhanced 1.7-fold and reached 959 mg/L *E. coli* cultivation broth. The produced recombinant HRP had a very high purity ( $\geq 99\%$ ) and the specific catalytic activity with ABTS as substrate was around 1500 U/mg for both pH 8.5 and pH 10.

Table 6. Comparison of refolding at pH 8.5 and pH 10.

Process Variables	pH 8.5	pH 10
Specific activity (U/mg)	1507 $\pm$ 13	1468 $\pm$ 24
Purity SEC-HPLC (%)	$\geq 99$	$\geq 99$
Refolding yield (%)	44	74
Pure HRP/L culture medium (mg)	562	959
Rz	3.7	4.3
Total Units/refolding vessel (7 mM ABTS)	146700	209500
Overall yield active HRP per 100 mg expressed protein (mg)	19	28

The refolding yield was determined by the final amount of HRP in the refolding mix as a percentage of the total amount of IBs that was solubilized.

#### 2.5. Characterization of Refolded HRP

Finally, the kinetic parameters of the commercially available plant HRP from Sigma-Aldrich (pHRP; Cat. No.: P6782) were compared to refolded HRP (rHRP; Table 7). In fact, both enzyme preparations were comparable in terms of catalytic activity underlining the applicability and high potential of the recombinant HRP from *E. coli* inclusion bodies.

Table 7. Comparison of the commercially available plant HRP with refolded HRP concerning the kinetic parameters for the substrates ABTS and TMB.

ABTS				
	V <sub>max</sub> (U/mg)	K <sub>m</sub> (mM)	k <sub>cat</sub> (s <sup>-1</sup> )	k <sub>cat</sub> /K <sub>m</sub> (mM <sup>-1</sup> ·s <sup>-1</sup> )
pHRP	1285 $\pm$ 70	0.70 $\pm$ 0.14	734 $\pm$ 41	1043 $\pm$ 215
rHRP	1411 $\pm$ 43	0.49 $\pm$ 0.06	823 $\pm$ 25	1677 $\pm$ 205
TMB				
	V <sub>max</sub> (U/mg)	K <sub>m</sub> (mM)	k <sub>cat</sub> (s <sup>-1</sup> )	k <sub>cat</sub> /K <sub>m</sub> (mM <sup>-1</sup> ·s <sup>-1</sup> )
pHRP	7446 $\pm$ 528	0.101 $\pm$ 0.020	4343 $\pm$ 308	42830 $\pm$ 8864
rHRP	7146 $\pm$ 355	0.105 $\pm$ 0.014	4169 $\pm$ 207	39582 $\pm$ 5661

#### 2.6. Final HRP Production Process

The final process steps to prepare recombinant HRP from *E. coli* inclusion bodies are summarized in Table 8.



**Table 8.** Unit operations and the respective process parameters for the final production process of HRP from *E. coli* inclusion bodies.

Unit operation.	Parameters	Final Conditions
Solubilization	DTT	7.11 mM
	Protein concentration	20 g/L
	pH	10
Refolding	GSSG	1.27 mM
	Protein concentration	0.5 g/L
	pH	10
	Time of hemin addition	8 h after refolding start
Salt precipitation	Hemin concentration	20 $\mu$ M
	pH	Adjust to pH 8.5 with 2 M HCl
	Type of salt	NaCl
	Salt concentration	4 M (pH 8.5)
Capture step HIC	Hydrophobicity of resin	Butyl
	pH value (load)	8.5
	Type of elution	Step gradient

### 3. Materials and Methods

#### 3.1. Chemicals

GSSG was purchased from AppliChem (Darmstadt, Germany). ABTS was purchased from AppliChem or Sigma-Aldrich (St. Louis, MO, USA). Plant HRP Type VI-A (Cat. No.: P6782), hemin (hemin from bovine,  $\geq 90\%$ ) and TMB were purchased from Sigma-Aldrich. DTT and all other chemicals were purchased from Carl Roth (Karlsruhe, Germany).

#### 3.2. Strain and Growth Conditions

The *hrp* gene coding for HRP variant C1A was codon-optimized for *E. coli* and obtained from GenScript USA Inc. (Piscataway, NJ, USA). The plasmid pET21d+ was used for HRP IB production in the cytoplasm. A stop codon was introduced so that the protein was produced without any tags. HRP was produced in *E. coli* BL21(DE3) in a 10 L Biostat Cplus stainless steel bioreactor (Sartorius, Germany). The pre-culture was grown in 0.5 L DeLisa medium [64] at 37 °C, 230 rpm in a 2.5 L Ultra Yield™ Flask (UYF; Thomson Instrument company, Encinitas, CA, USA) over night. Subsequently, the pre-culture was added to 4.5 L DeLisa medium in the bioreactor vessel and batch fermentation at 35 °C was run for 6 h. The pH was adjusted constantly to 7.2 and the dissolved oxygen was kept above 20%. During the 16 h fed-batch phase the specific uptake rate ( $q_s$ ) was 0.333 g/g/h, which was set to 0.25 g/g/h after induction with 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). After an induction phase of 8 h, the biomass was harvested by centrifugation and stored at −20 °C until further processing. All data collection and control of the process was done using a process information management system (Lucullus; Biospectra; Schlieren, Switzerland). All IBs used in this study were produced in one fermentation run.

#### 3.3. Homogenization and Wash

The biomass was resuspended using an IKA T10 basic ULTRA-TURRAX (Staufen, Germany) in 5 mL buffer A/g wet biomass (Buffer A: 50 mM Tris/HCl; pH 8; 500 mM NaCl; 1.5 mM ethylenediaminetetraacetic acid (EDTA)) and homogenized at >1200 bar, three passages, cooled, using a GEA Niro Soavi Panda PLUS (Düsseldorf, Germany). The homogenized suspension was centrifuged (15,650 g; 20 min, 4 °C), the supernatant discarded and the cell debris resuspended in 10 mL buffer B/g wet cell debris (Buffer B: 50 mM Tris/HCl; pH 8; 500 mM NaCl; 2 M Urea) and centrifuged

again (15,650 g; 20 min, 4 °C). This washing step with buffer B was performed twice. Afterwards, the pellet was resuspended in water (5 mL water/g wet cell debris), the suspension aliquoted into pre-weighed 50 mL reaction tubes, centrifuged (15,650 g; 20 min, 4 °C) and the pellets were stored at −20 °C.

### 3.4. Solubilization

For solubilization, an aliquot of the washed and frozen IBs was thawed, the wet inclusion body (wIB) weight was determined and the pellet was resuspended in the appropriate solubilization buffer to reach a wIB concentration of 100 g/L (solubilization buffer 1: 50 mM Tris/HCl; pH 8.5; 6 M Urea; solubilization buffer 2: 50 mM Glycine; pH 10; 6 M Urea). After resuspension, DTT was added to reach a final concentration in the solubilization mix of 1 mM–28.44 mM and the solubilization mix was incubated (room temperature (RT); 0.5 h; slight agitation), followed by centrifugation (20,379 g; 20 min; 4 °C). The supernatant was immediately used for refolding, and the pellet was discarded.

### 3.5. Refolding

#### 3.5.1. Small-Scale

Small-scale experiments were performed using DoE approaches. Planning and analysis of DoEs were done using Umetrics MODDE 10 (Malmö, Sweden). All small-scale refolding experiments were performed using 2 mL reaction tubes. The solubilize was diluted 1:40 in the appropriate refolding buffer (refolding buffer 1: 20 mM Tris/HCl; pH 8.5; 2 M Urea; 2 mM CaCl<sub>2</sub>; 7% v/v glycerol; varying GSSG concentrations; refolding buffer 2: 20 mM glycine; pH 10; 2 M urea; 2 mM CaCl<sub>2</sub>; 7% v/v glycerol; varying GSSG concentrations), followed by incubation at 4 °C; 48 h; with slight agitation. A 1 mM hemin stock solution was prepared in 100 mM potassium hydroxide (KOH).

#### 3.5.2. DoE 1: Redox Conditions

For the first DoE, the DTT concentration for solubilization (solubilization buffer 1; pH 8) and the GSSG concentration in the refolding buffer (refolding buffer 1; pH 8.5) were varied (Table 9 and Supplementary Table S4). A CCF with the volumetric activity after refolding as a response was used and four replicates were performed for the center point (8.75 mM DTT and 2 mM GSSG).

**Table 9.** DTT and GSSG concentrations used to optimize the redox conditions in a DoE CCF approach with the volumetric activity as response.

DTT conc. (mM in Solubilize)	GSSG (mM in Refolding Buffer)
2.5	0.5
8.75	2
15	3.5

#### 3.5.3. DoE 2: Protein Concentration during Refolding

For the second DoE, the DTT concentration for solubilization (solubilization buffer 1; pH 8), the GSSG concentration in the refolding buffer (refolding buffer 1; pH 8.5) and the protein concentration in the refolding mix were varied (Table 10 and Supplementary Table S5). A CCF with the volumetric and specific activity after refolding as a response was used and four replicates were performed for the center point (14.22 mM DTT, 2.54 mM GSSG and 1 g/L protein concentration).



**Table 10.** DTT, GSSG and total protein concentrations were used to investigate interactions between the redox system and the protein concentration in the refolding mix.

DTT (mM in Solubilize)	GSSG (mM in Refolding Buffer)	Protein in the Refolding Mix (g/L)
7.11	1.27	0.5
14.22	2.54	1
28.44	5.08	2

### 3.5.4. DoE 3: Redox Conditions and pH

For the third DoE, the DTT concentration for solubilization, the GSSG concentration in the refolding buffer and the pH of the solubilization and refolding buffer were varied (Table 11 and Supplementary Table S6). For pH 8.5, solubilization and refolding buffers 1 were used; for pH 10, solubilization and refolding buffers 2 were used. A CCF with the volumetric activity after refolding as a response was used and four replicates were performed for the two center points (7.11 mM DTT and 2.54 mM GSSG for both pH 8.5 and pH 10).

**Table 11.** DoE 3 investigated the interaction between the redox system and the pH during solubilization and refolding using a CCF approach.

DTT (mM in Solubilize)	GSSG (mM in Refolding Buffer)	pH
2.5	0.4	7
7.11	1.27	8.5
11.72	3.01	10

DTT and GSSG concentrations as well as the pH value of the solubilization and refolding buffers were used as factors.

### 3.5.5. DoE 4: Hemin Addition

For this DoE, time and concentration of hemin addition were varied between 0–24 h after refolding start and 6–80  $\mu$ M hemin, respectively (Table 12). Solubilization buffer 1 (pH 8) and refolding buffer 1 (pH 8.5) were used for all experiments. The volumetric activity was used as response.

**Table 12.** Time of addition and concentration of hemin in the refolding mix were used as factors in DoE 4.

Hemin Addition (Time after Refolding Start) (h)	Final Hemin Concentration ( $\mu$ M)
0	6
6	20
12	40
24	80

## 3.6. Refolding Vessel

### 3.6.1. Refolding Vessel Set-Up

For bench-scale refolding an Infors Labfors 5 (Bottmingen, Germany) with a vessel volume of 3.6 L was used. All data collection and control of the process was done using Lucillus (Biospectra, Schlieren, Switzerland). Temperature was kept constant at 10 °C during refolding using a Lauda Alpha R8 thermostat (Lauda, Königshofen, Germany) connected to the double jacket of the glass vessel. The temperature was monitored using a sensor connected to the Infors Labfors 5. In addition, pH, dO<sub>2</sub> (dissolved oxygen) and redox potential were monitored. The redox potential was monitored using a Hamilton EasyFerm Plus ORP Arc 425 (Hamilton, Bonaduz, Switzerland). The hemin feed was applied using a LAMBDA PRECIFLOW peristaltic liquid pump (LAMBDA laboratory instruments,

Switzerland) in combination with a Sartorius Entris scale (Sartorius, Germany), enabling a PID-Feed forward control using Lucillus. Final refolding volumes for the vessel were kept constant at 1200 mL (30 mL solubilization and a dilution of 1:40).

### 3.6.2. Refolding Vessel Experiment 1

For this experiment, the solubilization mix contained 7.11 mM DTT and the refolding buffer contained 1.27 mM GSSG. Solubilization buffer 2 (pH 10) and refolding buffer 2 (pH 10) were used. Hemin was added 20 h after refolding start to a final concentration of 20  $\mu$ M and then incubated for another 5 h, resulting in a total refolding time of 25 h.

### 3.6.3. Refolding Vessel Experiment 2

For this experiment, the solubilization mix contained 7.11 mM DTT and the refolding buffer contained 1.27 mM GSSG. Solubilization buffer 1 (pH 8) and refolding buffer 1 (pH 8.5) were used. Hemin was added 20 h after refolding start to a final concentration of 20  $\mu$ M. Samples were taken every 2 h, hemin was added to a final concentration of 20  $\mu$ M (only for samples taken before hemin addition; the samples taken after hemin addition already contained 20  $\mu$ M), the samples were incubated (2 h; 4 °C, slight agitation), and then enzyme activity was measured.

### 3.6.4. Refolding Vessel Experiment 3

For this experiment, the solubilization mix contained 7.11 mM DTT and the refolding buffer contained 1.27 mM GSSG. Solubilization buffer 1 (pH 8) and refolding buffer 1 (pH 8.5) were used. A constant feed (2 mL of a 1 mM hemin stock/h; final concentration 20  $\mu$ M hemin) was applied 8 h after refolding start until 20 h (12 h feed time). Samples were drawn every 2 h and activity was measured. After the start of the hemin feed, samples were measured both directly (with a low hemin concentration at the start of the hemin feed), and again after the addition of hemin to a final concentration of 20  $\mu$ M hemin and incubation for 2 h.

## 3.7. Capture and Concentration

HIC was used as a capture step after refolding and protein precipitation. An ÄKTA Pure system (GE Healthcare, Chicago, IL, USA) was used. Three wavelengths (214 nm, 280 nm and 404 nm) as well as the conductivity were monitored. Two different salts were tested for the preceding precipitation step, namely  $(\text{NH}_4)_2\text{SO}_4$  and NaCl. Several concentrations of both salts were tested in 2 mL reaction tubes (data not shown). The results obtained for the small-scale experiments were then validated using bench-scale experiments with a volume of 150 mL. The salt was slowly added under continuous stirring within 10 min, the solution was then incubated while stirring for 20 min at RT and then centrifuged (20,379 g; 20 min; 22 °C). The capture step was developed in five HIC experiments using different resins, pH values and elution profiles (HIC experiments 1–3 are described in the Supplementary data).

### 3.7.1. HIC Experiment 4 (Small-Scale)

The load was prepared by adding 267 g NaCl/L refolding mix. A 1 mL HiTrap Butyl FF column (GE Healthcare) was used with a flow rate of  $75 \text{ cm}^{-1} \cdot \text{h}^{-1}$ . The column was equilibrated with buffer A (20 mM Bis-Tris pH 7; 7% v/v glycerol; 4 M NaCl) and 49 mL load were applied. After the load, a washing step with 8 column volumes (CV) buffer A was performed. Thereafter, a step elution was performed with 25% buffer B (20 mM Bis-Tris pH 7; 7% v/v glycerol/ 8 CV), 75% buffer B (10 CV) and 100% (17 CV) buffer B. Volumetric enzyme activity (U/mL) and protein concentration were measured for all fractions. The purity of the active pool was determined using SEC-HPLC and the Reinheitszahl.



### 3.7.2. HIC Experiment 5 (Scale-Up)

After refolding at pH 10, the pH was adjusted to 8.5 with 2 M HCl under stirring. Afterwards, 267 g NaCl/L refolding mix were added. A column packed with 80 mL Butyl Sepharose 4 Fast Flow (GE Healthcare) was used. The column was equilibrated at a flow rate of  $90 \text{ cm}^{-1} \cdot \text{h}^{-1}$  with buffer A (20 mM Bis-Tris pH 7; 4 M NaCl) and 751 mL load were applied. After the load, a wash step with 20% buffer B (20 mM Bis-Tris pH 7/1.5 CV) was performed. Thereafter, a step elution was performed at  $75 \text{ cm}^{-1} \cdot \text{h}^{-1}$  with 75% buffer B (3 CV) and 100% (3 CV) buffer B. Volumetric enzyme activity (U/mL) and protein concentration were measured for all fractions. The purity of the active pool was determined using R<sub>z</sub>.

### 3.7.3. Analytics

Enzyme activity: HRP enzyme activity was measured with a Tecan Infinite M200 PRO (Männedorf, Switzerland) using flat-bottom polystyrene 96-well plates. Depending on the concentration of correctly folded HRP, samples were diluted 1:50–1:200 in dilution buffer (20 mM Bis-Tris pH 7; 7% v/v glycerol). 170  $\mu\text{L}$  of ABTS solution (5 mM ABTS in 50 mM  $\text{KH}_2\text{PO}_4$  pH 5) were mixed with 10  $\mu\text{L}$  of diluted sample in the well, after which 20  $\mu\text{L}$  of hydrogen peroxide (1 mM final concentration) were added to start the reaction. Immediately afterwards, the change of absorption at 420 nm over 2 minutes was recorded (at 30 °C). The volumetric enzyme activity was calculated using the following Equation (1):

$$A \text{ [U/mL]} = \frac{V_{\text{total}} \cdot \Delta A / \text{min} \cdot \text{dilution}}{V_{\text{sample}} \cdot d \cdot \epsilon} \quad (1)$$

$V_{\text{total}}$  ... total volume in cuvette in ( $\mu\text{L}$ )

$\Delta A / \text{min}$  ... change in absorption ( $\Delta \text{Abs}$  420 nm/min)

$\text{Dilution}$  ... dilution of the sample

$V_{\text{sample}}$  ... volume of sample ( $\mu\text{L}$ )

$d$  ... length of the beam path through the cuvette ( $d = 0.58 \text{ cm}$ )

$\epsilon$  ... extinction coefficient ( $\epsilon_{420} = 36 \text{ mM}^{-1} \text{ cm}^{-1}$  [65]).

Determination of kinetic parameters for final HRP preparation: The procedure was performed as described in [66]. Enzyme activity parameters were determined for the substrates ABTS and TMB in a 96-well plate assay using a Tecan Infinite M200 PRO instrument. For the measurements with ABTS, the reaction mixture in each well contained a saturating hydrogen peroxide concentration of 1 mM and 7 mM ABTS in 50 mM phosphate-citrate buffer, pH 5, in a final volume of 200  $\mu\text{L}$ . The protein sample (5  $\mu\text{L}$ ) was mixed with 175  $\mu\text{L}$  ABTS-buffer mixture and the reaction was started with 20  $\mu\text{L}$  of a 10 mM hydrogen peroxide solution. The increase in absorption was followed at 420 nm at 30 °C for 120 s. For the determination of the kinetic parameters, the ABTS concentration was varied (0.1–7 mM) and calculations were performed with the Sigma Plot software (Systat Software INC., San Jose, CA, USA) and an extinction coefficient of  $\epsilon_{420} = 36 \text{ mM}^{-1} \text{ cm}^{-1}$  [65]. For the measurements with TMB, the reaction mixture contained a saturating hydrogen peroxide concentration of 1 mM and varying TMB concentrations (0.02–0.55 mM) in 50 mM phosphate-citrate buffer, pH 5, with a final volume of 200  $\mu\text{L}$ . An extinction coefficient of  $\epsilon_{652} = 39 \text{ mM}^{-1} \text{ cm}^{-1}$  was used [67].

Protein concentration: Protein concentration was determined using the method according to Bradford [68]. In total, 200  $\mu\text{L}$  Bradford solution were mixed with 5  $\mu\text{L}$  sample and the change in absorbance at 595 nm was measured with a Tecan Infinite M200 PRO instrument over the course of 10 min.

RP-HPLC: The HRP concentration in the samples was measured with RP-HPLC using a Polyphenyl BioResolve-RP-mAb 2.7  $\mu\text{m}$   $3.0 \times 100 \text{ mm}$  column (Waters, MA, USA). The method was run for 10 min with the following program: 25% line B for 0.5 min, 55% line B in a linear gradient for 8 min, 55% line B for 0.5 min and then 25% line B for 1 min (Line A: MilliQ water with 0.1% trifluoroacetic acid (TFA));

line B: acetonitrile with 0.1% TFA) at a flow rate of 1.2 mL/min. The column was kept at a constant temperature of 75 °C and the wavelengths 214 nm, 280 nm and 404 nm were monitored.

SEC-HPLC: Purity of HRP was measured using a SEC-HPLC with a BEH 200A SEC 1.7 µm 4.6 × 300 mm, 3.5 µm (Waters, MA, USA) column. The method was run at 0.3 mL/min using 100% line A (Line A: 80 mM phosphate buffer pH 6.8; 250 mM KCl) for 18 minutes. The column was kept at a constant temperature of 30 °C and the wavelengths 214 nm, 280 nm and 404 nm were monitored.

Reinheitszahl (Rz): The Reinheitszahl was calculated as the ratio of absorbance at 404 nm to 280 nm and the absorbance measurement was performed using a Hitachi Double Beam Spectrophotometer U-2900 (Tokyo, Japan).

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/21/13/4625/s1>.

**Author Contributions:** Conceptualization, O.S.; Methodology, O.S., D.H. and J.E.; Validation, O.S., D.H. and J.E.; Formal analysis, D.H. and J.E.; Investigation, D.H. and J.E.; Resources, O.S.; Data curation, D.H. and J.E.; Writing—original draft preparation, D.H. and J.E.; Writing—review and editing, O.S.; Visualization, D.H. and J.E.; Supervision, O.S.; Project administration, O.S.; Funding acquisition, O.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** Open Access Funding by the Austrian Science Fund (FWF), grant number P30872-B26.

**Acknowledgments:** Open Access Funding by the Austrian Science Fund (FWF).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

ABTS	2,2'-azino-bis (3-ethylbenzothiazoline-6-sulphonic acid)
CaCl <sub>2</sub>	Calcium chloride
CCF	Central composite face centered
CEX	Cation exchange chromatography
CV	Column volumes
dO <sub>2</sub>	Dissolved oxygen
DoE	Design of experiments
DSP	Downstream process
DTT	Dithiothreitol
GSSG	Glutathione disulfide
GuHCl	Guanidine hydrochloride
HIC	Hydrophobic interaction chromatography
HPLC	High-performance liquid chromatography
IB	Inclusion body
IMAC	Immobilized metal affinity chromatography
pHRP	Plant horseradish peroxidase
rHRP	Recombinant Horseradish Peroxidase
RP	Reversed-phase
RT	Room temperature
Rz	Reinheitszahl
SEC	Size exclusion chromatography
TFA	Trifluoroacetic acid
TMB	3,3',5,5'-tetramethylbenzidine
USP	Upstream process
wIB	Wet inclusion body

## References

1. Krainer, F.W.; Glieder, A. An updated view on horseradish peroxidases: Recombinant production and biotechnological applications. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 1611–1625. [CrossRef] [PubMed]



2. Azevedo, A.M.; Martins, V.C.; Prazeres, D.M.; Vojinovic, V.; Cabral, J.M.; Fonseca, L.P. Horseradish peroxidase: A valuable tool in biotechnology. *Biotechnol. Annu. Rev.* **2003**, *9*, 1387–2656. [\[CrossRef\]](#)
3. Grigorenko, V.; Chubar, T.; Kapeliuch, Y.; Böchers, T.; Spener, F.; Egorova, A. New approaches for functional expression of recombinant horseradish peroxidase C in *Escherichia coli*. *Biocatal. Biotransform.* **1999**, *17*, 359–379. [\[CrossRef\]](#)
4. Grigorenko, V.G.; Andreeva, I.P.; Rubtsova, M.Y.; Egorov, A.M. Recombinant horseradish peroxidase: Production and analytical applications. *Biochem. (Mosc.)* **2015**, *80*, 408–416. [\[CrossRef\]](#)
5. Spadiut, O.; Herwig, C. Production and purification of the multifunctional enzyme horseradish peroxidase. *Pharm. Bioprocess* **2013**, *1*, 283–295. [\[CrossRef\]](#)
6. Veitch, N.C. Horseradish peroxidase: A modern view of a classic enzyme. *Phytochemistry* **2004**, *65*, 249–259. [\[CrossRef\]](#)
7. Veitch, N.C.; Smith, A.T. Horseradish peroxidase. *Adv. Inorg. Chem.* **2000**, *51*, 107–162.
8. Welinder, K.G. Covalent structure of the glycoprotein horseradish peroxidase (EC 1.11. 1.7). *FEBS Lett.* **1976**, *72*, 19–23. [\[CrossRef\]](#)
9. Wührer, M.; Hokke, C.H.; Deelder, A.M. Glycopeptide analysis by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry reveals novel features of horseradish peroxidase glycosylation. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1741–1748. [\[CrossRef\]](#)
10. Wührer, M.; Balog, C.I.; Koeleman, C.A.; Deelder, A.M.; Hokke, C.H. New features of site-specific horseradish peroxidase (HRP) glycosylation uncovered by nano-LC-MS with repeated ion-isolation/fragmentation cycles. *Biochim. Biophys. Acta* **2005**, *1723*, 229–239. [\[CrossRef\]](#)
11. Berglund, G.I.; Carlsson, G.H.; Smith, A.T.; Szöke, H.; Henriksen, A.; Hajdu, J. The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **2002**, *417*, 463–468. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Van Gijlswijk, R.P.; Talman, E.G.; Peekel, I.; Bloem, J.; Van Velzen, M.A.; Heetebrij, R.J.; Tanke, H.J. Use of horseradish peroxidase and fluorescein-modified cisplatin derivatives for simultaneous labeling of nucleic acids and proteins. *Clin. Chem.* **2002**, *48*, 1352–1359. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Yakovleva, J.; Davidsson, R.; Lobanova, A.; Bengtsson, M.; Eremin, S.; Laurell, T.; Emnéus, J. Microfluidic enzyme immunoassay using silicon microchip with immobilized antibodies and chemiluminescence detection. *Anal. Chem.* **2002**, *74*, 2994–3004. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Moody, M.; Arndell, S.V.; Murphy, K.; Orencole, S.; Burns, C. Array-based ELISAs for high-throughput analysis of human cytokines. *Biotechniques* **2001**, *31*, 186–194. [\[CrossRef\]](#)
15. Lomillo, M.A.A.; Ruiz, J.G.; Pascual, F.J.M. Biosensor based on platinum chips for glucose determination. *Anal. Chim. Acta* **2005**, *547*, 209–214. [\[CrossRef\]](#)
16. Azevedo, A.M.; Prazeres, D.M.F.; Cabral, J.M.; Fonseca, L.P. Ethanol biosensors based on alcohol oxidase. *Biosens. Bioelectron.* **2005**, *21*, 235–247. [\[CrossRef\]](#)
17. Raghu, P.; Reddy, T.M.; Reddalah, K.; Jaidev, L.; Narasimha, G. A novel electrochemical biosensor based on horseradish peroxidase immobilized on Ag-nanoparticles/poly (L-arginine) modified carbon paste electrode toward the determination of pyrogallol/hydroquinone. *Enzyme Microb. Technol.* **2013**, *52*, 377–385. [\[CrossRef\]](#)
18. Kubota, K.; Mizukoshi, T.; Miyano, H. A new approach for quantitative analysis of L-phenylalanine using a novel semi-sandwich immunometric assay. *Anal. Bioanal. Chem.* **2013**, *405*, 8093–8103. [\[CrossRef\]](#)
19. Yang, H. Enzyme-based ultrasensitive electrochemical biosensors. *Curr. Opin. Chem. Biol.* **2012**, *16*, 422–428. [\[CrossRef\]](#)
20. Litescu, S.C.; Eremia, S.; Radu, G.L. Biosensors for the determination of phenolic metabolites. In *Bio-Farms for Nutraceuticals*, 1st ed.; Giardi, M.T., Rea, G., Berra, B., Eds.; Springer: New York, NY, USA, 2010; Volume 698, pp. 234–240. [\[CrossRef\]](#)
21. Vasileva, N.; Godjevargova, T.; Ivanova, D.; Gabrovska, K. Application of immobilized horseradish peroxidase onto modified acrylonitrile copolymer membrane in removing of phenol from water. *Int. J. Biol. Macromol.* **2009**, *44*, 190–194. [\[CrossRef\]](#)
22. Tatsumi, K.; Wada, S.; Ichikawa, H. Removal of chlorophenols from wastewater by immobilized horseradish peroxidase. *Biotechnol. Bioeng.* **1996**, *51*, 126–130. [\[CrossRef\]](#)
23. Bhunia, A.; Durani, S.; Wangikar, P.P. Horseradish peroxidase catalyzed degradation of industrially important dyes. *Biotechnol. Bioeng.* **2001**, *72*, 562–567. [\[CrossRef\]](#)

24. Colonna, S.; Gaggero, N.; Carrea, G.; Pasta, P. Horseradish peroxidase catalysed sulfoxidation is enantioselective. *J. Chem. Soc. Chem. Commun.* **1992**, 357–358. [\[CrossRef\]](#)
25. Colonna, S.; Gaggero, N.; Richelmi, C.; Pasta, P. Recent biotechnological developments in the use of peroxidases. *Trends Biotechnol.* **1999**, *17*, 163–168. [\[CrossRef\]](#)
26. Kalliney, S.; Zaks, A. An efficient peroxidase-catalyzed oxidation of hydroxylaminoeverminomicin in aqueous-organic media. *Tetrahedron Lett.* **1995**, *36*, 4163–4166. [\[CrossRef\]](#)
27. Kalra, B.; Gross, R.A. Horseradish peroxidase mediated free radical polymerization of methyl methacrylate. *Biomacromolecules* **2000**, *1*, 501–505. [\[CrossRef\]](#)
28. Singh, A.; Ma, D.; Kaplan, D.L. Enzyme-mediated free radical polymerization of styrene. *Biomacromolecules* **2000**, *1*, 592–596. [\[CrossRef\]](#)
29. Sakai, S.; Khanmohammadi, M.; Khoshfetrat, A.B.; Taya, M. Horseradish peroxidase-catalyzed formation of hydrogels from chitosan and poly (vinyl alcohol) derivatives both possessing phenolic hydroxyl groups. *Carbohydr. Polym.* **2014**, *111*, 404–409. [\[CrossRef\]](#)
30. Saikrishnan, D.; Goyal, M.; Rossiter, S.; Kukol, A. A cellulose-based bioassay for the colorimetric detection of pathogen DNA. *Anal. Bioanal. Chem.* **2014**, *406*, 7887–7898. [\[CrossRef\]](#)
31. Yin, H.; Wang, M.; Li, B.; Yang, Z.; Zhou, Y.; Ai, S. A sensitive electrochemical biosensor for detection of protein kinase A activity and inhibitors based on Phos-tag and enzymatic signal amplification. *Biosens. Bioelectron.* **2015**, *63*, 26–32. [\[CrossRef\]](#)
32. Wardman, P. Indole-3-acetic acids and horseradish peroxidase: A new prodrug/enzyme combination for targeted cancer therapy. *Curr. Pharm. Des.* **2002**, *8*, 1363. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Tupper, J.; Tozer, G.M.; Dachs, G.U. Use of horseradish peroxidase for gene-directed enzyme prodrug therapy with paracetamol. *Br. J. Cancer* **2004**, *90*, 1858–1862. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Brooks, S.A. Appropriate glycosylation of recombinant proteins for human use. *Mol. Biotechnol.* **2004**, *28*, 241–255. [\[CrossRef\]](#)
35. Kaur, J.; Kumar, A.; Kaur, J. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *Int. J. Biol. Macromol.* **2017**, *106*, 803–822. [\[CrossRef\]](#)
36. Slouka, C.; Kopp, J.; Hutwimmer, S.; Strahammer, M.; Strohmer, D.; Eitenberger, E.; Schwaighofer, A.; Herwig, C. Custom made inclusion bodies: Impact of classical process parameters and physiological parameters on inclusion body quality attributes. *Microb. Cell Fact.* **2018**, *17*, 148. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Eggenreich, B.; Willim, M.; Wurm, D.J.; Herwig, C.; Spadiut, O. Production strategies for active heme-containing peroxidases from *E. coli* inclusion bodies—a review. *Biotechnol. Rep.* **2016**, *10*, 75–83. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Rinas, U.; Garcia-Fruitós, E.; Corchero, J.L.; Vázquez, E.; Seras-Franzoso, J.; Villaverde, A. Bacterial inclusion bodies: Discovering their better half. *Trends Biochem. Sci.* **2017**, *42*, 726–737. [\[CrossRef\]](#)
39. Humer, D.; Spadiut, O. Wanted: More monitoring and control during inclusion body processing. *World J. Microb. Biot.* **2018**, *34*, 158. [\[CrossRef\]](#)
40. Smith, A.T.; Santama, N.; Dacey, S.; Edwards, M.; Bray, R.C.; Thorneley, R.; Burke, J.F. Expression of a synthetic gene for horseradish peroxidase C in *Escherichia coli* and folding and activation of the recombinant enzyme with Ca<sup>2+</sup> and heme. *J. Biol. Chem.* **1990**, *265*, 13335–13343.
41. Asad, S.; Dabirmarash, B.; Ghaemi, N.; Etezad, S.M.; Khajeh, K. Studies on the refolding process of recombinant horseradish peroxidase. *Mol. Biotechnol.* **2013**, *54*, 484–492. [\[CrossRef\]](#)
42. Gazaryan, I.G.; Doseeva, V.V.; Galkin, A.G.; Tishkov, V.I. Effect of single-point mutations Phe41→His and Phe143→Glu on folding and catalytic properties of recombinant horseradish peroxidase expressed in *E. coli*. *FEBS Lett.* **1994**, *354*, 248–250. [\[CrossRef\]](#)
43. Gazaryan, I.; Chubar, T.; Ignatenko, O.; Mareeva, E.; Orlova, M.; Kapeliuch, Y.L.; Savitsky, P.; Rojkova, A.; Tishkov, V. Tryptophanless recombinant horseradish peroxidase: Stability and catalytic properties. *Biochem. Biophys. Res. Commun.* **1999**, *262*, 297–301. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Gundinger, T.; Spadiut, O. A comparative approach to recombinantly produce the plant enzyme horseradish peroxidase in *Escherichia coli*. *J. Biotechnol.* **2017**, *248*, 15–24. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Kim, S.J.; Song, B.K.; Kim, Y.H. Optimized refolding and characterization of S-peroxidase (CWPO\_C of *Populus alba*) expressed in *E. coli*. *Protein Expr. Purif.* **2011**, *80*, 268–273. [\[CrossRef\]](#)



46. Shigeto, J.; Itoh, Y.; Tsutsumi, Y.; Kondo, R. Identification of Tyr74 and Tyr177 as substrate oxidation sites in cationic cell wall-bound peroxidase from *Populus alba* L. *FEBS J.* **2012**, *279*, 348–357. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Shigeto, J.; Nagano, M.; Fujita, K.; Tsutsumi, Y. Catalytic profile of Arabidopsis peroxidases, AtPrx-2, 25 and 71, contributing to stem lignification. *PLoS ONE* **2014**, *9*, e105332. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Zakharova, G.; Poloznikov, A.; Chubar, T.; Gazaryan, I.; Tishkov, V. High-yield reactivation of anionic tobacco peroxidase overexpressed in *Escherichia coli*. *Protein Expression Purif.* **2015**, *113*, 85–93. [\[CrossRef\]](#)
49. Fattahian, Y.; Riahi-Madvar, A.; Mirzaee, R.; Torkzadeh-Mahani, M.; Asadikaram, G. Heterologous expression, purification and characterization of a peroxidase isolated from *Lepidium draba*. *Protein J.* **2017**, *36*, 461–471. [\[CrossRef\]](#)
50. Hushpuliari, D.; Savitski, P.; Rojkova, A.; Chubar, T.; Fehina, V.; Sakharov, I.Y.; Lagrimini, L.; Tishkov, V.; Gazaryan, I. Expression and refolding of tobacco anionic peroxidase from *E. coli* inclusion bodies. *Biochem. (Mosc)* **2003**, *68*, 1189–1194. [\[CrossRef\]](#)
51. Teilum, K.; Østergaard, L.; Welinder, K.G. Disulfide Bond Formation and Folding of Plant Peroxidases Expressed as Inclusion Body Protein in *Escherichia coli* Thioredoxin Reductase Negative Strains. *Protein Expr. Purif.* **1999**, *15*, 77–82. [\[CrossRef\]](#)
52. Doyle, W.A.; Smith, A.T. Expression of lignin peroxidase H8 in *Escherichia coli*: Folding and activation of the recombinant enzyme with Ca<sup>2+</sup> and haem. *Biochem. J.* **1996**, *315*, 15–19. [\[CrossRef\]](#)
53. Pérez-Boada, M.; Doyle, W.; Ruiz-Dueñas, F.; Martínez, M.; Martínez, A.; Smith, A. Expression of *Pleurotus eryngii* versatile peroxidase in *Escherichia coli* and optimisation of in vitro folding. *Enzyme Microb. Technol.* **2002**, *30*, 518–524. [\[CrossRef\]](#)
54. Nie, G.; Reading, N.S.; Aust, S.D. Expression of the lignin peroxidase h2 gene from *Phanerochaete chrysosporium* in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **1998**, *249*, 146–150. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Coupe, E.; Smyth, M.; Fosberry, A.; Hall, R.; Littlechild, J. The dodecameric vanadium-dependent haloperoxidase from the marine algae *Corallina officinalis*: Cloning, expression, and refolding of the recombinant enzyme. *Protein Expression Purif.* **2007**, *52*, 265–272. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Miki, Y.; Morales, M.; Ruiz-Dueñas, F.J.; Martínez, M.J.; Wariishi, H.; Martínez, A.T. *Escherichia coli* expression and in vitro activation of a unique ligninolytic peroxidase that has a catalytic tyrosine residue. *Protein Expression Purif.* **2009**, *68*, 208–214. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Linde, D.; Coscolin, C.; Liers, C.; Hofrichter, M.; Martínez, A.T.; Ruiz-Dueñas, F.J. Heterologous expression and physicochemical characterization of a fungal dye-decolorizing peroxidase from *Auricularia auricular-judae*. *Protein Expression Purif.* **2014**, *103*, 28–37. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Wang, N.; Ren, K.; Jia, R.; Chen, W.; Sun, R. Expression of a fungal manganese peroxidase in *Escherichia coli*: A comparison between the soluble and refolded enzymes. *BMC Biotech.* **2016**, *16*, 87. [\[CrossRef\]](#)
59. Lee, D.-H.; Kim, D.-H. Heterologous expression of lignin peroxidase H2 in *Escherichia coli*: In vitro refolding and activation. *BMB Rep.* **1999**, *32*, 486–491.
60. Whitam, R.; Tien, M. Heterologous expression and reconstitution of fungal Mn peroxidase. *Arch. Biochem. Biophys.* **1996**, *333*, 439–446. [\[CrossRef\]](#)
61. Rodríguez-Cabrera, N.A.; Regalado, C.; García-Almendárez, B.E. Cloning, heterologous expression and properties of a recombinant active turnip peroxidase. *J. Agric. Food Chem.* **2011**, *59*, 7120–7126. [\[CrossRef\]](#)
62. Dong, X.-Y.; Huang, Y.; Sun, Y. Refolding kinetics of denatured-reduced lysozyme in the presence of folding aids. *J. Biotechnol.* **2004**, *114*, 135–142. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Pappa, H.S.; Cass, A.E. A step towards understanding the folding mechanism of horseradish peroxidase: Tryptophan fluorescence and circular dichroism equilibrium studies. *Eur. J. Biochem.* **1993**, *212*, 227–235. [\[CrossRef\]](#) [\[PubMed\]](#)
64. DeLisa, M.P.; Li, J.; Rao, G.; Weigand, W.A.; Bentley, W.E. Monitoring GFP-operon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. *Biotechnol. Bioeng.* **1999**, *65*, 54–64. [\[CrossRef\]](#)
65. Childs, R.E.; Bardsley, W.G. The steady-state kinetics of peroxidase with 2, 2'-azino-di-(3-ethyl-benzthiazoline-6-sulphonic acid) as chromogen. *Biochem. J.* **1975**, *145*, 93–103. [\[CrossRef\]](#)
66. Humer, D.; Spadiut, O. Improving the Performance of Horseradish Peroxidase by Site-Directed Mutagenesis. *Int. J. Mol. Sci.* **2019**, *20*, 916. [\[CrossRef\]](#)

67. Josephy, P.D.; Eling, T.; Mason, R.P. The horseradish peroxidase-catalyzed oxidation of 3, 5, 3', 5'-tetramethylbenzidine. Free radical and charge-transfer complex intermediates. *J. Biol. Chem.* **1982**, *257*, 3669–3675.
68. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



## 6.5 Scientific paper 2

Journal of Pharmaceutical and Biomedical Analysis 188 (2020) 113412



Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: [www.elsevier.com/locate/jpba](http://www.elsevier.com/locate/jpba)



### Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles

Julian Kopp<sup>a,\*</sup>, Florian B. Zauner<sup>b</sup>, Andreas Pell<sup>a</sup>, Johanna Hausjell<sup>c</sup>, Diana Humer<sup>c</sup>, Julian Ebner<sup>c</sup>, Christoph Herwig<sup>a</sup>, Oliver Spadiut<sup>c</sup>, Christoph Slouka<sup>c</sup>, Reinhard Pell<sup>b</sup>

<sup>a</sup> Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, TU Wien, Gumpendorfer Straße 1a, 1060 Vienna, Austria

<sup>b</sup> SANDOZ GmbH, Mondseestrasse 11, 4866 Unterach, Austria

<sup>c</sup> Research Division Biochemical Engineering, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Gumpendorfer Straße 1a, 1060 Vienna, Austria



#### ARTICLE INFO

##### Article history:

Received 27 February 2020

Received in revised form 13 May 2020

Accepted 2 June 2020

Available online 4 June 2020

##### Keywords:

Liquid chromatography

(U)HPLC

Protein quantification

*E. coli*

Analytical quality by design (AQbD)

Design of Experiment (DoE)

#### ABSTRACT

Biopharmaceutical drug substances are generally produced using fermentation technology and are subsequently purified in the following downstream process. For the determination of critical quality attributes (CQAs), such as target protein titer and purity, monitoring tools are required before quality control analysis. We herein present a novel reversed phase liquid chromatography method (RPLC), which enables facile and robust protein quantification during upstream and downstream processing of intracellularly produced proteins in *E. coli*.

The overall goal was to develop a fast, robust and mass spectrometry compatible method which can baseline resolve and quantify each protein of interest. Method development consisted of three steps, oriented on an Analytical Quality by Design (AQbD) workflow: (i) the stationary phase as primary parameter was chosen based on state-of-the-art technology thus minimizing protein on-column adsorption and providing high efficiency, (ii) secondary parameters (i.e. gradient conditions and column temperature) were optimized applying chromatographic modeling, and (iii) the established Method Operable Design Region (MODR) was challenged and confirmed during robustness testing, performed *in-silico* and experimentally by a Design of experiment (DoE) based approach.

Finally, we validated the RPLC method for pivotal validation parameters (i.e. linearity, limit of quantification, and repeatability) and compared it for protein quantification against a well-established analytical methodology. The outcome of this study shows (i) a protocol for RPLC development using an AQbD principle for new method generation and (ii) a highly versatile RPLC method, suited for quick and straightforward recombinant protein titer measurement being applicable for the detection of a broad range of proteins.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Abbreviations:** CQAs, critical quality attributes; RPLC, reversed phase liquid chromatography; AQbD, Analytical Quality by Design; MODR, Method Operable Design Region; DoE, Design of experiment; KPI, key performance indicator; *E. coli*, *Escherichia coli*; IB, inclusion body; SDS-PAGE, sodium-dodecyl-sulfate polyacrylamide gel electrophoresis; LC, liquid chromatography; CE, Capillary electrophoresis; MS, mass spectrometry; HIC, Hydrophobic interaction chromatography; SEC, size exclusion chromatography; IEX, ion-exchange chromatography; mAbs, monoclonal antibodies; ADCs, antibody-drug conjugates; TFA, trifluoroacetic acid; SPP, superficially porous particles; N-pro, N-pro fusion protein; GFP, green fluorescent protein; CH3H, Chalcone 3-Hydroxylase; HRP, Horseradish peroxidase; BSA, Bovine Serum Albumin; MW, Molecular weight; GRAVY, grand average of hydropathicity; FLD, fluorescence detector; UV, ultraviolet; EP, European Pharmacopeia; tG, gradient time; ATP, analytical target profile; R<sup>2</sup>, measure of fit; R<sup>2</sup>, adj, adjusted measure of fit; Q<sup>2</sup>, predictability; LOQ, limit of quantification; RSD, relative standard deviation.

\* Corresponding author.

E-mail address: [julian.kopp@tuwien.ac.at](mailto:julian.kopp@tuwien.ac.at) (J. Kopp).

<https://doi.org/10.1016/j.jpba.2020.113412>

0731-7085/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A panoply of commercially available biopharmaceuticals are manufactured as recombinant proteins using microbial or mammalian hosts [1]. To monitor the production of biopharmaceutically active compounds, product titer is a pivotal key performance indicator (KPI) [2]. Since it is necessary to react promptly on process changes, facile sample preparation and subsequent rapid analytics are of high interest [3]. The gram-negative bacterium *Escherichia coli* (*E. coli*) still is today's expression host of choice for production of approximately 30 % of recombinantly produced pharmaceuticals in industry [1]. Quantification of target proteins is challenging in *E. coli* as recombinant proteins generally are expressed in the cell interior (either soluble or as an inclusion body = IB). Necessary cell disruption leads to a high amount of host-cell associated impurities and thus to a complex sample matrix for analysis.

Today, a plethora of analytical technologies exist for the analysis of (recombinant) proteins. Favored techniques are sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which can be further used in western blotting to detect low protein concentrations, or capillary electrophoresis [4]. Samples derived from microbial hosts often suffer from impurities, favoring target protein detection via size using SDS-PAGE analysis [5]. Different stains, such as silver and comassie blue, can be employed for protein staining of gels. Silver staining PAGE enables more sensitive staining, however comassie blue SDS-PAGE is used more frequently as it is easier in its application, cheaper and also allows much faster processing compared to silver staining [5]. Still, when comparing SDS-PAGE analysis to liquid chromatography (LC) it is more time demanding and lacks accurate quantitation [6]. Capillary electrophoresis (CE) is another tool for the qualitative and quantitative analysis of biopharmaceuticals, and proteins in general [7]. CE is considered a powerful separation technique in terms of selectivity and efficiency, and can further be coupled with mass spectrometry (MS) for protein identification and characterization [8,9]. However, it has drawbacks such as limited instrument ruggedness and non-straightforward method development.

In the last decade, LC has emerged as technique of choice for characterization and quantitation of (recombinant) proteins [10–13], due to facile sample preparation, technically mature instrumentation and major advancements in stationary phase chemistry. Hydrophobic interaction chromatography (HIC) [14], size exclusion chromatography (SEC) [15] and ion-exchange chromatography (IEX) [11] enable analysis of proteins in native (non-denaturing) conditions, but are limited in their applicability. SEC is mainly carried out for quantitation of protein aggregates and IEX for separation of protein charge variants. HIC (on analytical scale) has its main application in separation and characterization of large biomolecules, such as monoclonal antibodies (mAbs) and antibody-drug conjugates (ADCs) [16]. However, HIC exhibits drawbacks like slow mass transfer resulting in broad peaks and the necessity of high mobile phase salt concentrations thus impeding hyphenation with MS.

Reversed phase liquid chromatography (RPLC) has become the predominant technique for analysis of proteins due to its versatility, robustness, high efficiency and MS compatibility [17,18]. In contrast to the aforementioned chromatographic modes, RPLC is a denaturing technique which often requires harsh conditions (e.g. high temperature and low pH) in order to eliminate irreversible on-column protein adsorption and improve peak shape. However, the advent of novel stationary phases, including wide-pore superficially porous particles (SPP) and special stationary phase coatings for suppression of silanol interactions, enabled higher peak efficiencies and application of milder elution conditions [19,20]. Recently, a wide-pore SPP column with a high coverage phenyl bonding

was introduced to the market, which addresses the last obstacles in RPLC protein analysis (i.e. on-column adsorption of large biomolecules, poor selectivity for closely related proteins) [21].

In this study, we established a widely-applicable, robust, and MS-compatible RPLC method applicable for on-line process analytics and quality control of recombinant proteins with pronounced differences in their physicochemical properties. A state-of-the-art column was used for systematic method development based on Analytical Quality by Design (AQbD) principles. The needed requirements to the method were defined according to the ATP (= analytical target profile), followed by investigation of critical method parameters and their effects onto monitored responses using a Design-of-Experiment (=DoE) approach [22]. Following the AQbD concept, the method operable design region (MODR) including retention modeling was assessed [23]. Robustness testing was performed *in-silico* and experimentally to verify the set MODR, applying a Design-of-Experiment (DoE) concept. Finally, the method performance was compared for protein quantification in a case study against one of the benchmark technologies (i.e. SDS-PAGE).

## 2. Material and methods

### 2.1. Chemicals and samples

Ultra-pure water was obtained from a Milli-Q system from Merck Millipore (Burlington, USA). Acetonitrile (HPLC grade) and trifluoroacetic acid (TFA, >99.9 %) were purchased from Carl Roth (Karlsruhe, Germany). All proteins implemented within this study are described below. Further information regarding cultivation and purification of the proteins can be found in Supplementary section A.

#### 2.1.1. N-pro fusion protein (N-pro)

N-pro fusion protein was produced in an *E. coli* BL21(DE3) with the pET30a plasmid system (kanamycin resistance). The target protein was linked to a N-pro fusion protein used for purification. Details on protein structure and purification steps cannot be given due to proprietary reasons.

#### 2.1.2. eGFP (GFP)

GFP was produced in an *E. coli* BL21(DE3) (Life technologies, Carlsbad, CA, USA), transformed with a pET21a(+) vector (ampicillin resistance) carrying the gene coding for enhanced green fluorescent protein (eGFP). Samples derived from fed-batch fermentation (as referred here [24]), were disrupted and centrifuged to receive the soluble fraction. Protein solution was filtered through pores < 0.2 µm and was further purified with an immobilized metal affinity chromatography (IMAC), binding to the expressed HIS-tag of the protein. Samples were desalted and stored in aqueous buffer composed of 20 mM TRIS, 100 mM NaCl, pH = 7.5 containing 10 % glycerol for RPLC analysis.

#### 2.1.3. Chalcone 3-Hydroxylase (CH3H)

CH3H was produced in an *E. coli* HSM174(DE3) genotype: F- recA1 hsdR(rK12- mK12+) (DE3) (Rif<sup>r</sup>) in a pET21a plasmid (Novagen, Madison, WI, USA) encoding CH3H of *D. variabilis*. The plasmid for CH3H expression was a kind donation from Christina Divne and Rosaria Gandini (Karolinska Institute, Stockholm, Sweden). In brief: sub-cloning of the CH3H sequence into pNIC-CTHO was performed using the LIC-cloning methodology. Inclusion body samples (derived from fed-batch fermentation) were analyzed as crude samples after homogenization and samples were solubilized with a buffer described elsewhere [24].



**Table 1**  
Physicochemical properties of analyzed proteins.

Target protein	MW [kDa]	Disulphide bridges [nr]	GRAVY Index [25]	expression form
N-pro	28.0	n.a.	n.a.	IB
BSA	66.5	17	-0.433	n/a
GFP	26.9	1	-0.467	soluble and IB
HRP	34.7	4	-0.177	IB
CH3H	55.0	0	-0.080	IB

#### 2.1.4. Horseradish peroxidase (HRP)

HRP (EC 1.11.1.7) was produced in an *E. coli* BL21(DE3) pET21d(+) with ampicillin resistance (Amp<sup>R</sup>). The *hrp*-gene coding for HRP variant C1A was codon-optimized for *E. coli* and obtained from GenScript USA Inc. (Piscataway, NJ, USA). Samples derived from fed-batch fermentation were disrupted and centrifuged. The aggregated fraction was further purified [22]. Inclusion bodies were washed with water, solubilized, refolded and purified using a yet unpublished procedure. Samples for analysis were stored in an aqueous buffer containing 20 mM BIS-TRIS and 7% glycerol.

#### 2.1.5. Bovine serum albumin (BSA)

BSA was purchased from Sigma-Aldrich (St. Louis, MO, USA), receiving 2 mg/mL aliquots for analysis stored in 0.9 % NaCl containing 0.05 % sodium azide (CAS Number 9048–46-8; MDL number MFCD00130384).

#### 2.1.6. Protein physicochemical properties

Physical properties and number of disulfide bonds were calculated using different software tools and are summarized in Table 1. Molecular weight (MW) and grand average of hydropathicity (GRAVY) were calculated using ExPASy ProtPar [25] as these features can be accessed on a free basis with the given amino acid composition of the target protein. Disulfide prediction was calculated using the DIANNA 1.1 webserver by F. Ferre et C. Clote [26].

#### 2.2. Equipment and software

Measurements were performed using a Dionex UltiMate 3000 system with a quaternary solvent delivery pump, an auto-sampler, UV and fluorescence (FL) detector (Thermo Fisher, Waltham, MA, USA). Dwell volume was determined as  $V_d = 0.90$  mL. Data were acquired using either an UV detector or a FL Detector. Data acquisition and instrument control were carried out by Chromeleon 7.2 software (Thermo Fisher). Calculations and data transfer were achieved with Excel (Microsoft). Resolution was calculated according to European Pharmacopeia (EP). Retention and resolution modeling was performed with DryLab® software (Version 4.3.3., Molnar-Institute, Berlin, Germany). DoE and statistical evaluations were carried out with the software MODDE (MODDE 12.1, Umetrics, Sweden).

#### 2.3. Chromatographic conditions

The HPLC was equipped with a BioResolve RP mAb Polyphenyl column (dimensions 100 × 3 mm, particle size 2.7 µm) which is designed for mAb and large protein analyses (Waters Corporation, MA, USA). In order to prolong column lifetime, a pre-column (3.9 × 5 mm, 2.7 µm) was used. The mobile phase was composed of ultrapure water (=MQ, eluent A) and acetonitrile (eluent B) both supplemented with 0.10 % (v/v) trifluoroacetic acid. The injection volume was set to 2.0 µL. The flow rate and the column temperature of the final method were 0.4 mL/min and 70 °C, respectively. Re-equilibration was conducted for six minutes, which corresponds to approximate five column volumes. The acquisition wavelength for UV detection was set to 214 and 280 nm, respectively,

for the FL detection the excitation wavelength was set to 280 nm and the emission wavelength to 360 nm. All detectors were series-connected, monitoring UV and FLD chromatograms at once, although FLD chromatograms only were used for evaluation. Sample concentration was 1.5 mg/mL for BSA. Via BSA calibration the sample nominal concentration was determined to be 2.5 mg/mL for N-pro, 3.5 mg/mL for GFP, 0.5 mg/mL for HRP and 2.0 mg/mL for CH3H. Total observed values in FLD- spectra differ as the amount of aromatic amino acids highly varies between employed target proteins.

#### 2.4. Design of experiment of input runs for software modeling

For the chromatographic modeling using DryLab, a 2D model was chosen in order to optimize gradient time (tG) and column temperature (T). The input runs required a 2 × 2 full factorial design applying a linear gradient with short (tG of 8 min) and long (tG of 24 min) gradient time (starting from 25 % B to 60 % B), and column temperature of 50 °C and 78 °C, respectively (the concept is shown in Fig. 1a). The flow rate for the input runs was set to 0.8 mL/min. All runs were carried out separately for each protein and performed in duplicate. The mean value of the retention time was used for modeling in DryLab (the raw data are given in the Supplementary data section B, Table B1).

#### 2.5. Protein assay determination (quantification)

For SDS-PAGE analysis BSA was mixed with 2-fold concentrated Laemmli solution to achieve single concentrated Laemmli buffer in the final dilution. Samples were then heated at 95 °C for 10 min. 10 µL of each sample were loaded onto pre-cast SDS gels containing 15 vials (8–16 %, Bio-Rad, Hercules, CA, USA). Gels were run in a Mini-PROTEAN Tetra System (Bio-Rad, Hercules, CA, USA) for about 60 min at 140 V and stained with Coomassie Blue. The protein bands were evaluated densitometrically using the software ImageLab (Bio-Rad, Hercules, CA, USA).

For HPLC-measurement of the soluble proteins, the sample was directly used after cell disruption and centrifugation for GFP and after a refolding step for HRP [24]. IB pellet samples were prepared according to Ref. [24], and subsequently solubilized using a buffer solution of 7.5 M guanidine hydrochloride, 62 mM TRIS, and 125 mM 1,4-Dithioerythritol (DTT) at pH = 8 (all chemicals purchased from Carl Roth, Karlsruhe, Germany). All samples were filtered through 0.2 µm syringe filters purchased from Carl Roth (Karlsruhe, Germany) prior to analysis.

#### 2.6. Validation

As BSA was stored in an aqueous solution, dilutions of the commercial purchased standard were done with ultra-pure water (Burlington, USA). Solutions for LOQ and linearity determination consisted of the following concentrations: (5, 10, 50, 100, 250, 500, 1000, 1500, and 2000 µg/mL). LOQ was determined experimentally by a six-fold injection followed by evaluation of the relative standard deviation of peak areas and the S/N ratio. Linearity was evaluated in the range of determined LOQ until 2.0 mg/mL (=con-

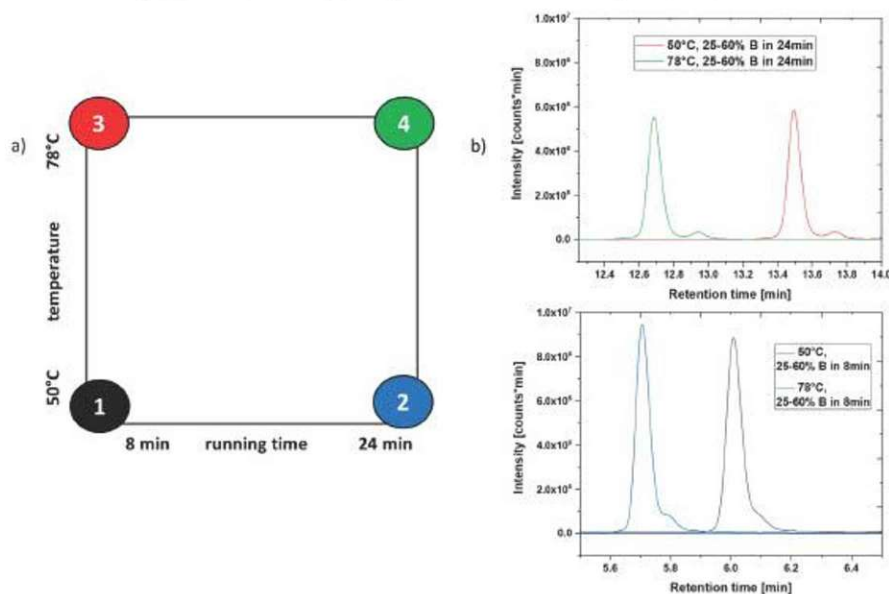


Fig. 1. a) DoE for secondary method factors varied regarding their input conditions for modeling software DryLab4 operated at a constant flow rate of 0.8 mL/min; b) Exemplary Input chromatograms for the individual DOE points, illustrated by the model protein HRP.

centration of the commercially available reference standard) at nine concentration levels.

### 3. Results and discussion

In order to develop a widely applicable LC method for protein assay determination, five model proteins were chosen which cover a broad range of protein size and hydrophobicity, as hydrophobicity is known to influence RPLC majorly. The GRAVY Index, describing hydrophobicity of target proteins, is calculated via amino acid composition of target proteins (listed in Table 1): proteins resulting in a GRAVY index close to 0 are highly hydrophobic whereas proteins are more hydrophilic, the more negative results for GRAVY Index are [25]. In biotechnology, GFP is often utilized as biosensor or as a marker protein due to easy auto-fluorescent detection [2]. BSA is often used for calibration purposes in biotechnology if there is a lack of a purified reference substance [24]. HRP, being used for enzymatic cancer treatment, was recombinantly expressed as an inclusion body but analyzed in its native form after refolding within this study. To show the applicability of the developed method independent from expression form, the two further analytes (N-pro fusion protein and CH3H) were analyzed in their incorrect folded form (=IB). All samples were evaluated separately as they were (i) derived from different purification steps (stated in Section 2.1. for each given protein) (ii) and were stored in different buffer solutions (stated in Section 2.5).

#### 3.1. Method development workflow and selection of primary factors

The analytical target profile (=ATP) was established to develop a robust and sensitive method with baseline-separation between neighboring target peaks (i.e. critical resolution greater 1.5). As samples derived from microbial cultivation might cause a highly

complex sample matrix, the focus of this study was to provide a generic RPLC method for protein quantification, being independent from state of purification and its needed buffer solution. To verify this approach, target proteins were chosen in a broad range of hydrophobicity and varied in their degree of purity, as crude samples (i.e. CH3H) and highly purified proteins (i.e. HRP) were evaluated. MS-compatibility was further chosen as target profile in order to make the developed method applicable for further elucidation possibilities. The first step was the selection of appropriate method factors (parameters) for the stationary- and mobile phase system (stationary phase and mobile phase bulk solvents are considered primary method factors in an "AQbD language" [21,27]): Considering the recent advancements in stationary phase chemistry for RPLC protein analytics, a small set of columns is commercially available which minimize protein on column adsorption due to either special end-capping of the silica surface or using an organic polymer as support material [17,21]. Further, high peak efficiency can be accomplished on account of smaller particle sizes and/or usage of superficially porous particles [19,21]. Due to aforementioned reasons, a BioResolve RP mAb Polyphenyl column was chosen.

Apart from the stationary phase chemistry, also the type of organic modifier and mobile phase pH are considered as primary factors [20,27]. As primary factors are known to exhibit a higher impact on selectivity than secondary factors (i.e. column temperature), appropriate choice is highly important [21,27]. In contrast to RPLC analyses of small organic molecules, Acetonitrile (being an aprotic solvent) is most often applied as bulk mobile phase in protein RPLC [17]. The application of protic solvents (e.g. Methanol) often creates protein adsorption, hence they are avoided or only added in small portions to Acetonitrile in order to improve selectivity. For the sake of simplicity and in order to develop an easy to use method, Acetonitrile was chosen as the only organic modifier. Hence, utilization of purified water and acetonitrile were selected



**Table 2**  
DryLab optimized method parameters.

Analysis time [min]	Eluent B [%]
0.0	25
8.0	55
10.0	75
12.0	75
12.1	25
18.0	25

Flow rate: 0.4 mL/min,  
Column temperature: 70 °C.

as mobile phase bulk solvents and TFA was used as additive to suppress secondary interactions. Also, TFA as "semi-volatile" additive enables hyphenation with MS or aerosol-forming universal detectors (i.e. charged aerosol or evaporative light scattering detectors), although detector sensitivity is compromised.

Gradient (tG) and column temperature (T), have been described to significantly enhance protein separation in RPLC [17] and thus were chosen as secondary factors. High temperature set-points were chosen to (i) decrease the chance of non-specific protein adsorption and (ii) to favor thermodynamic behavior for chromatographic interaction. Since TFA was applied because of aforementioned reasons, a systematic variation and optimization of pH was omitted.

### 3.2. DryLab assisted modeling of secondary method factors

Gradient time and column temperature were optimized using a chromatographic modeling software (i.e. DryLab [28,29]). The calculation of "ideal" parameters for the flow rate, gradient and column temperature required retention time data obtained from input runs. In this study, a  $2 \times 2$  full factorial design was applied (illustrated in Fig. 1a). tG was varied by three-fold and T was set to 50 °C for the lower limit and to 78 °C for the upper limit (Fig. 1). Limits were chosen as temperatures below 50 °C may cause irreversible adsorption of proteins on the column, whereas 80 °C was the technical limit of the employed HPLC system (thus 78 °C was chosen to allow flexibility during robustness experiments). The changes in retention time are exemplarily displayed for HRP in Fig. 1b. Supplementary Table B1 lists each retention time per protein and chromatographic condition.

Based on the obtained retention times and peak-widths (at 50 % peak height) for each condition and protein, a chromatographic model was created in DryLab. The resolution between GFP and BSA (compare Fig. 2b and c) was identified as the most critical peak pair due to their congeneric hydrophobicity [25]. By iteratively modifying the method parameters *in-silico*, a constant flow rate of 0.4 mL/min, a column temperature of 70 °C and the gradient program shown in Table 2 were determined thus establishing the method operable design region (MODR).

Fig. 2a depicts the predicted chromatogram. In order to proof the validity of the DryLab model and its prediction, a verification run was carried out in the lab (Fig. 2b). As can be deduced by comparing the predicted with the real chromatogram, the retention times and peak widths are well comparable. Thus, the DryLab model is considered to be valid and can be used for further studies.

The critical resolution, defined as the smallest calculated resolution for all peak pairs, is displayed as a heat map in dependence of the column temperature versus the gradient time (Fig. 2c). This additionally helps in the selection of the optimal combination of column temperature and gradient time by simultaneously obtaining an estimate for the robustness of the selected method factors. In Fig. 2c, the red to orange area is the result of a combination of T and tG which leads to a critical resolution bigger than 2.0. The cross shows the selection of tG and T of the optimized method. Tempera-

**Table 3**  
List of all factors altered including their set-points and their variance, as well as the variance deviation for the *in-silico* calculation.

Factor	Set-point + variance delta
Flow rate [mL/min]:	$0.4 \pm 0.1$
Dwell Volume [mL]:	$0.9 \pm 0.7$
Temp. [°C]:	$70 \pm 5$
Start %B [%]:	$25 \pm 1$
Step 1 Time [min]:	$8 \pm 1$
Step 1 %B [%]:	$55 \pm 1$
Step 2 Time [min]:	$10 \pm 1$
tG [min]:	$12 \pm 1$
End %B [%]:	$75 \pm 1$

tures below 70 °C could facilitate unspecific protein adsorption and were therefore neglected. The calculated chromatographic space around our set point is predicted to be in major parts characterized by a critical resolution > 2.0 and thus considered as robust by simultaneously minimizing the run time. Thus, the predicted method operable design region is equal to the area colored in red-orange in Fig. 2c.

### 3.3. *In-silico* and experimental robustness studies

#### 3.3.1. *In silico* robustness

Robustness evaluates the capacity of an analytical method of remaining unaffected by small, but deliberate changes of method factors (parameters) [30]. DryLab offers the possibility of *in-silico* robustness studies [28]. Nine factors in total (compare to Table 3) were varied at three levels (central point, upper and lower level) and the critical resolution was calculated for each factor combination.

A total of 13122 operating conditions were simulated and tested for their influence onto the critical resolution. Remarkably, the outcome of the *in-silico* calculation is showing that a 94.1 % of all factor combinations were above the set resolution of 2.0, indicating a highly robust method. Moreover, the *in-silico* robustness study allowed to identify pivotal method factors with the highest impact on the critical resolution (compare to Supplementary data B, Figure B1). The factor exhibiting the highest influence identified by the *in-silico* robustness study was found to be the gradient time (compare to Supplementary Fig. 1B). Further factors, exhibiting high impact on the critical resolution were identified throughout simulation and hence chosen for experimental robustness determination as described in chapter 3.3.2.

#### 3.3.2. Experimental determination of robustness

Due to the fact that the critical peak pair (i.e. BSA and GFP) is eluting during the first step gradient (compare to Fig. 2b and c), only tG of the first gradient was investigated. Further, flow rate and temperature also exhibited significant impact in the *in-silico* simulation and were therefore selected for the experimental robustness study. The influence of ion-pairing effects on selectivity and peak shape caused by certain amounts of TFA in the mobile phase cannot be predicted by the modeling software. Consequently, TFA concentration in mobile phase A and B was selected as fourth component for the experimental robustness study.

The four chosen factors (tG of the first gradient, flow rate, column temperature and TFA conc. in the mobile phase) were varied based on a response surface (central composite) design of experiment, as illustrated in Fig. 3a. The model was evaluated for linear, quadratic and interaction terms using the software MODDE. Centre-points were conducted as six-time repeated experiments. The response of the model was evaluated for the critical resolution between neighboring peaks. Factors were varied experimentally in the following ranges: duration of the first gradient from 7.5 to 8.5

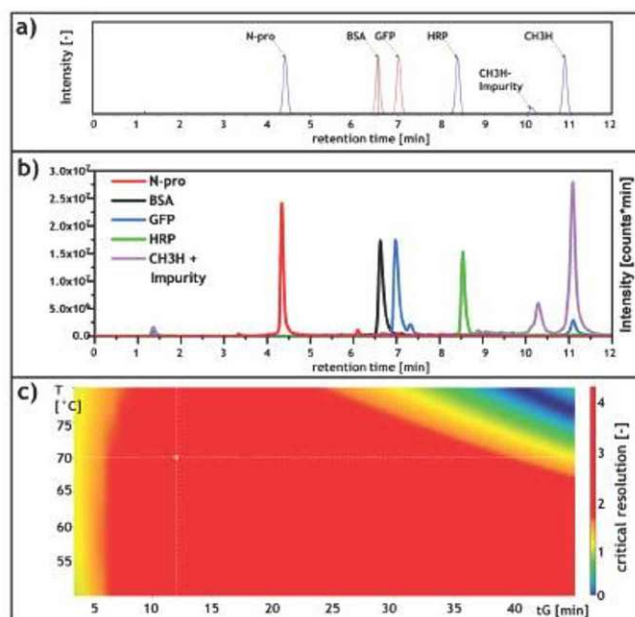


Fig. 2. a) Simulated *in-silico* chromatogram received from modeling with DryLab4; b) the chromatogram of the verification runs for all five target proteins utilizing the developed method; single chromatograms are depicted as overlay in order to show the baseline-separation of all applied target proteins c) heat map received from DryLab4 for chromatographic modeling for the corresponding input runs showing the influence of column temperature and gradient time.

Table 4  
Statistical evaluation of the model for all resolutions performed via  $R^2$ ,  $R^2$  adjusted and  $Q^2$ .

	$R^2$	$R^2$ adj.	$Q^2$
Resolution.N-pro.BSA	0.962	0.937	0.778
Resolution.BSA.GFP	0.983	0.973	0.854
Resolution.GFP.HRP	0.941	0.924	0.811
Resolution.HRP.CH3H.Impurity	0.957	0.942	0.776
Resolution.CH3H.Impurity.CH3H	0.915	0.897	0.839

min, temperature from 67–73 °C, flow rate from 0.35–0.45 mL/min and TFA concentration from 0.08–0.12 %.

The model was evaluated for its statistical terms being the measure of fit ( $R^2$ ) and the adjusted measure of fit ( $R^2$  adj.) to validate the model's variation of response, adjusted for degrees of freedom. Furthermore, the model predictability  $Q^2$  was evaluated to test for the model capability of predicting new data, shown in Table 4.

High  $R^2$ -values ( $> 0.9$ ) indicated that input data fit the model well. High values received for  $Q^2$  ( $> 0.775$ ) evaluated with MODDE showed that new data would fit the model properly and low experimental error was observed. To avoid any data specific model responses, the difference between  $R^2$  and  $Q^2$  should be as small as possible, which was found to be smaller than 0.184 for the stated data-set. The difference between  $R^2$  and  $R^2$  adj. indicates that the experimental setup performed within this study was chosen appropriate. Experiments conducted for all model proteins are given in Supplementary table 2B, whereas the resolution between the critical peak pair BSA-GFP will be discussed in more detail within this chapter, as these target proteins resulted in the smallest resolution (compare to Fig. 2b).

The DoE was evaluated by the resulting contour and interaction plots as illustrated in Fig. 3b and 3c. Even though column tempera-

tures below 70 °C were found to increase the critical resolution, 70 °C was considered as a compromise in order to avoid non-specific protein adsorption effects occurring at lower temperatures. Similar results were obtained for TFA concentration in the mobile phases: A lower TFA concentration was found to decrease the critical resolution, whereas a higher TFA concentration would improve critical resolution values but would on the other hand detrimentally impact sensitivity in mass spectrometry. A lower flow rate would improve the critical resolution but increase run time, simultaneously (see Fig. 3c). Therefore, a flow rate of 0.4 mL/min was determined as optimum. A short gradient time seemed to be beneficial, however the interaction plot in Fig. 3c showed that this was only true for smaller flow rates thus impeding a fast run time. Flow rates above 0.45 mL/min showed a greater critical resolution for long gradient times, compared to shorter gradient times.

No method factor combination resulted in a critical resolution lower than 1.5, hence selected optima were confirmed experimentally and the method can be considered robust for method variations such as  $\pm 3$  °C column temperature,  $\pm 0.02$  % of TFA concentration in mobile phase,  $\pm 0.5$  min of gradient time and  $\pm 0.05$  mL/min flow rate.

### 3.4. Method validation and comparison to a benchmark technology

#### 3.4.1. Instrument precision and repeatability

Instrument precision was tested via six subsequent injections of the same protein solution at sample nominal concentration (compare to Section 2.3). A relative standard deviation (RSD) of at most 2.0 % for peak area and retention time were set as acceptance criteria for all five proteins. For instrument precision experiments conducted with BSA, an injection of a blank solution (7.5 M



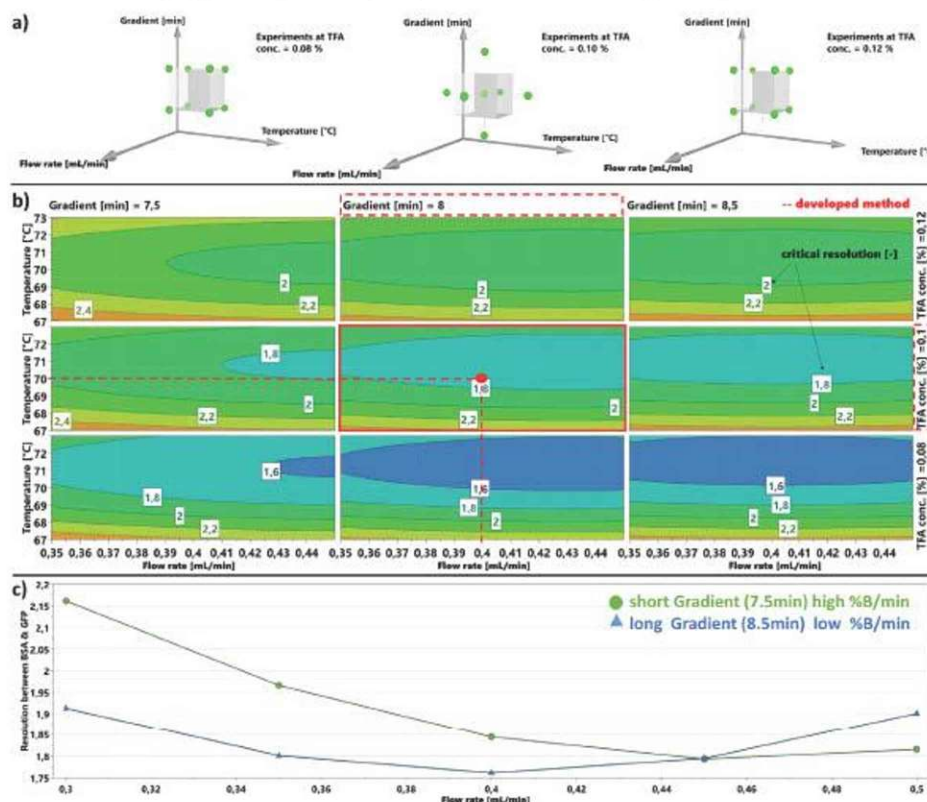


Fig. 3. a) Design of experiment and the experiments performed, exemplarily when keeping TFA conc. constant at 0.1 %; b) contour plot for all four factors indicating the best operating parameters to achieve a resolution greater than 1.5; c) an interaction plot, showing the dependencies of the variation of gradient and flow rate onto the critical resolution.

guanidine-HCl solution) was required to prevent sample carryover from one run to the next. RSD of the peak areas and the retention times were evaluated and fulfilled the criteria of not more than 2.0 % (compare to Supplementary Table C1).

In order to evaluate the impact of sample preparation, a repeatability study was performed. As BSA was commercially purchased, whereas other target proteins were derived from fermentation and subsequent purification, BSA was applied as model protein for the repeatability experiments. RSD for peak area and retention time below 2.0 % could be achieved throughout a six-fold sample preparation and subsequent analysis (compare to Supplementary Table C2).

### 3.4.2. Linearity and limit of quantitation

In biotechnology research and development model proteins are often quantified using BSA as a "generic" reference standard, in case no isolated material of the target protein is available [24]. For this reason, we applied BSA for evaluation of the Limit of Quantitation (LOQ) and linearity. A LOQ of 0.005 mg/mL could be established (RSD < 15 %, S/N ratio > 10). The method was found to be linear and precise over three orders of magnitude (correlation coefficient  $R^2 = 0.999$ , residual standard deviation = 1.7 %, prognosis interval  $P = 95$  % for the y-intercept includes the zero point). The calibration curve

is shown in Fig. 4a and further details on linearity evaluation are given Table 5.

### 3.4.3. Comparison of the developed RPLC method with SDS-PAGE analysis

Even though analysis of microbial proteins is commonly performed with SDS-PAGE, proper LC analytics should be superior regarding accuracy and analysis time [5]. To verify this hypothesis, a standard calibration curve established with (i) SDS-PAGE analysis (using comassie blue staining) was compared to (ii) a calibration curve measured with the herein developed RPLC method. Method attributes such as analysis time, LOQ, correlation coefficient, y-intercept and RSD are compared in Table 5.

The standard calibration curves (compare to Table 5 and Fig. 4a and b) show that higher accuracy and a decrease in analysis time can be achieved using the established RPLC method. Due to method limitations low protein concentrations could not be evaluated with a gel-based quantification, as indicated in Fig. 4b. As integration of SDS-PAGE-derived bands is challenging and needs to be performed visually (being highly operator dependent), high deviations in evaluation can be observed, which is in accordance with literature [6]. To test whether host cell impurities would impact the resolution of target proteins, crude samples were compared with purified ones (see Supplementary data section D, figure D1). The respective target

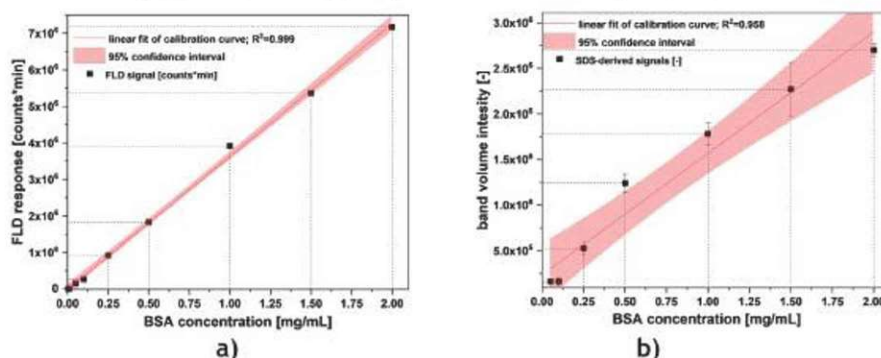


Fig. 4. Calibration of BSA standards with a confidence interval set for 95 % comparing a) the developed RPLC method with b) a state of the art SDS-PAGE using comassie blue staining.

Table 5

Parameters for the evaluation of the two compared analytical methods, using LOQ, correlation coefficient, y-intercept, and residual standard deviation referred to nominal concentration.

Method	Analysis time [min/sample]	LOQ BSA [mg/mL]	correlation coefficient [-]	y-intercept (ref. to sample nominal concentration) [%]	residual standard deviation (ref. to sample nominal concentration) [%]
RPLC	18	0.005	0.999	0	1.7
SDS-PAGE	150	0.1	0.979	8.7	8.6

peaks could be separated from their impurities, thus eliminating the need for SDS-page derived analysis throughout any process step.

Summarizing the comparison of both methods, RPLC measurements outperform the SDS-PAGE in terms of analysis speed (18 min versus 150 min) and limit of quantification (LOQ 0.005 mg/mL versus 0.1 mg/mL). This makes RPLC an attractive technique facilitating pharmaceutical up- and downstream development where fast and accurate responses are a prerequisite for screening conditions that potentially increase recovery yields and/or purity of the final product.

#### 4. Conclusion

A generic RPLC method for quantification of commonly applied model proteins expressed by the microbial host *E. coli* was developed. The method development workflow applied analytical quality-by-design principles by using retention modeling and establishing a method operable design region. The MODR was challenged by both *in-silico* and experimental robustness experiments, using a multivariate experimental design and analysis approach. Method performance of the established RPLC method was compared to a SDS-PAGE method, which is still a benchmark technology for the analysis of microbial derived samples. Results indicate the superiority of the RPLC method regarding analysis time, ease of use, linear range and LOQ.

In addition, the herein described AqBD workflow can be used as a generic method development protocol for RPLC analysis of proteins. Furthermore, the presented method has found direct utilization in the analysis of the chosen target proteins, indicating (i) its applicability for the analysis of a broad range of proteins with different hydrophobicities and (ii) its independency of applied protein purification procedures and the therefore needed buffer solutions. Finally, the established RPLC method is also MS-compatible thus enabling further opportunities for mass elucidation.

#### Author contributions

JK and AP performed analytical experiments. FZ constructed the experimental robustness study. RP conducted DryLab experiments. JH, DH and JE produced and purified proteins and performed experiments. JK, FZ and RP evaluated experiments and performed data treatment. CH, OS and CS gave valuable scientific input. JK, FZ, CS and RP drafted the manuscript.

#### Funding

The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program. Furthermore, the authors thank the Christian Doppler Society for funding of this project.

#### Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jpba.2020.113412>.

#### Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Furthermore, the authors declare that they have no competing interests.

#### References



- [1] G. Walsh, Biopharmaceutical benchmarks 2018, *Nat. Biotechnol.* 36 (12) (2018) 1136–1145.
- [2] J. Kopp, C. Slouka, O. Spadiut, C. Herwig, The rocky road from Fed-Batch to continuous processing with *E. coli*, *Front. Bioeng. Biotechnol.* 7 (2019) 328.
- [3] A.S. Rathore, QbD/PAT for bioprocessing: moving from theory to implementation, *Curr. Opin. Chem. Eng.* 6 (2014) 1–8.



- [4] M.H. Caparon, K.J. Rust, A.K. Hunter, J.K. McLaughlin, K.E. Thomas, J.T. Herberg, R.E. Shell, P.B. Lanter, B.F. Bishop, R.L. Dufield, Integrated solution to purification challenges in the manufacture of a soluble recombinant protein in *E. coli*, *Biotechnol. Bioeng.* 105 (2) (2010) 239–249.
- [5] J. Sasse, S.R. Gallagher, Staining proteins in gels, *Curr. Protoc. Mol. Biol.* 85 (1) (2009), 10.6.1–10.6.27.
- [6] V. Dolnik, W.A. Gurske, Chemical modification of proteins to improve the accuracy of their relative molecular mass determination by electrophoresis, *Electrophoresis* 32 (20) (2011) 2893–2897.
- [7] E. Domínguez-Vega, R. Haselberg, G.W. Somsen, Capillary zone electrophoresis–Mass spectrometry of intact proteins, in: N.T. Tran, M. Taverna (Eds.), *Capillary Electrophoresis of Proteins and Peptides: Methods and Protocols*, Springer, New York, New York, NY, 2016, pp. 25–41.
- [8] S. Fekete, D. Guilleme, P. Sandra, K. Sandra, Chromatographic, electrophoretic, and mass spectrometric methods for the analytical characterization of protein biopharmaceuticals, *Anal. Chem.* 88 (1) (2016) 480–507.
- [9] R.L.C. Voeten, I.K. Ventouri, R. Haselberg, G.W. Somsen, Capillary electrophoresis: trends and recent advances, *Anal. Chem.* 90 (3) (2018) 1464–1481.
- [10] K. Sandra, I. Vandenheede, P. Sandra, Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization, *J. Chromatogr. A* 1335 (2014) 81–103.
- [11] S. Fekete, A. Beck, J.L. Veuthey, D. Guilleme, Ion-exchange chromatography for the characterization of biopharmaceuticals, *J. Pharm. Biomed. Anal.* 113 (2015) 43–55.
- [12] S. Fekete, D. Guilleme, Ultra-high-performance liquid chromatography for the characterization of therapeutic proteins, *Trac Trends Anal. Chem.* 63 (2014) 76–84.
- [13] S. Fekete, J.-L. Veuthey, D. Guilleme, Modern Column Technologies for the Analytical Characterization of Biopharmaceuticals in Various Liquid Chromatographic Modes, *LCGC Europe*, 2015.
- [14] M. Baca, J. De Vos, G. Bruylants, K. Bartik, X. Liu, K. Cook, S. Eeltink, A comprehensive study to protein retention in hydrophobic interaction chromatography, *J. Chromatogr. B* 1032 (2016) 182–188.
- [15] S. Fekete, A. Goyon, J.-L. Veuthey, D. Guilleme, Size exclusion chromatography of protein biopharmaceuticals: past, present and future, *Am. Pharm. Rev.* (2018) 1099–8012.
- [16] T. Janaratne, X.C. Wang, C. Becker, Y. Zhao, R. Leanna, W. Pritts, Hydrophobic interaction chromatography (HIC) method development and characterization of resolved drug-load variants in site-specifically conjugated pyrrolbenzodiazepine dimer-based antibody drug conjugates (PBD-ADCs), *J. Pharm. Biomed. Anal.* 179 (2020) 113027.
- [17] S. Fekete, J.-L. Veuthey, D. Guilleme, New trends in reversed-phase liquid chromatographic separations of therapeutic peptides and proteins: theory and applications, *J. Pharm. Biomed. Anal.* 69 (2012) 9–27.
- [18] J. Kochling, W. Wu, Y. Hua, Q. Guan, J. Castaneda-Merced, A platform analytical quality by design (AQbD) approach for multiple UHPLC–UV and UHPLC–MS methods development for protein analysis, *J. Pharm. Biomed. Anal.* 125 (2016) 130–139.
- [19] B. Bobály, V. D'Atri, M. Lauber, A. Beck, D. Guilleme, S. Fekete, Characterizing various monoclonal antibodies with milder reversed phase chromatography conditions, *J. Chromatogr. B* 1096 (2018) 1–10.
- [20] B. Bobály, J.-L. Veuthey, D. Guilleme, S. Fekete, New developments and possibilities of wide-pore superficially porous particle technology applied for the liquid chromatographic analysis of therapeutic proteins, *J. Pharm. Biomed. Anal.* 158 (2018) 225–235.
- [21] B. Bobály, M. Lauber, A. Beck, D. Guilleme, S. Fekete, Utility of a high coverage phenyl-bonding and wide-pore superficially porous particle for the analysis of monoclonal antibodies and related products, *J. Chromatogr. A* 1549 (2018) 63–76.
- [22] B. Sylvester, L. Tefas, L. Vlase, I. Tomuță, A. Porfire, A Quality by Design (QbD) approach to the development of a gradient high-performance liquid chromatography for the simultaneous assay of curcuminoids and doxorubicin from long-circulating liposomes, *J. Pharm. Biomed. Anal.* 158 (2018) 395–404.
- [23] A. Tumpa, A. Stajić, B. Jančić-Stojanović, M. Medenica, Quality by Design in the development of hydrophilic interaction liquid chromatography method with gradient elution for the analysis of olanzapine, *J. Pharm. Biomed. Anal.* 134 (2017) 18–26.
- [24] D.J. Wurm, J. Quehenberger, J. Mildner, B. Eggenreich, C. Slouka, A. Schwaighofer, K. Wieland, B. Lendl, V. Rajamanickam, C. Herwig, O. Spadiut, Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in *E. coli*, *Appl. Microbiol. Biotechnol.* (2017).
- [25] E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, the Proteomics Protocols Handbook, Springer, 2005, pp. 571–607.
- [26] F. Ferré, P. Clote, DiANNA: a web server for disulfide connectivity prediction, *Nucleic Acids Res.* 33 (suppl.2) (2005) W230–W232.
- [27] S. Fekete, V. Sadat-Noorbakhsh, C. Schelling, I. Molnar, D. Guilleme, S. Rudaz, J.L. Veuthey, Implementation of a generic liquid chromatographic method development workflow: application to the analysis of phytocannabinoids and Cannabis sativa extracts, *J. Pharm. Biomed. Anal.* 155 (2018) 116–124.
- [28] R.M. Krisko, K. McLaughlin, M.J. Koenigbauer, C.E. Lunte, Application of a column selection system and DryLab software for high-performance liquid chromatography method development, *J. Chromatogr. A* 1122 (1) (2006) 186–193.
- [29] I. Molnar, Computerized design of separation strategies by reversed-phase liquid chromatography: development of DryLab software, *J. Chromatogr. A* 965 (1) (2002) 175–194.
- [30] B. Dejaegher, Y.V. Heyden, Ruggedness and robustness testing, *J. Chromatogr. A* 1158 (1) (2007) 138–157.

## Article

# At-Line Reversed Phase Liquid Chromatography for In-Process Monitoring of Inclusion Body Solubilization

Julian Ebner <sup>1</sup>, Diana Humer <sup>1</sup>, Robert Klausser <sup>1</sup>, Viktor Rubus <sup>1</sup>, Reinhard Pell <sup>2</sup>, Oliver Spadiut <sup>1</sup>   
and Julian Kopp <sup>1,\*</sup> 

<sup>1</sup> Research Division Integrated Bioprocess Development, Institute of Chemical, Environmental and Bioscience Engineering, Vienna University of Technology, 1060 Vienna, Austria; julian.ebner@tuwien.ac.at (J.E.); diana.humer@tuwien.ac.at (D.H.); robert.klausser@tuwien.ac.at (R.K.); viktor.rubus@tuwien.ac.at (V.R.); oliver.spadiut@tuwien.ac.at (O.S.)

<sup>2</sup> SANDOZ GmbH, Mondseestrasse 11, 4866 Unterach, Austria; reinhard.pell@gmx.at

\* Correspondence: julian.kopp@tuwien.ac.at; Tel.: +43-1-5880-1166-485



**Citation:** Ebner, J.; Humer, D.; Klausser, R.; Rubus, V.; Pell, R.; Spadiut, O.; Kopp, J. At-Line Reversed Phase Liquid Chromatography for In-Process Monitoring of Inclusion Body Solubilization. *Bioengineering* 2021, 8, 78. <https://doi.org/10.3390/bioengineering8060078>

Academic Editor: Susan Sharfstein

Received: 7 May 2021

Accepted: 5 June 2021

Published: 7 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Refolding is known as the bottleneck in inclusion body (IB) downstream processing in the pharmaceutical industry: high dilutions leading to large operating volumes, slow refolding kinetics and low refolding yields are only a few of the problems that impede industrial application. Solubilization prior to refolding is often carried out empirically and the effects of the solubilization on the subsequent refolding step are rarely investigated. The results obtained in this study, however, indicate that the quality of the IB solubilization has a severe effect on subsequent refolding. As the solubilization contains chaotropic reagents in high molarities, it is commonly analyzed with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE, however, suffers from a long analysis time, making at-line analytical implementation difficult. In this study, we established an at-line reversed phase liquid chromatography method to investigate the time-dependent quality of the solubilization. To verify the necessity of at-line solubilization monitoring, we varied the essential solubilization conditions for horseradish peroxidase IBs. The solubilization time was found to have a major influence on subsequent refolding, underlining the high need for an at-line analysis of solubilization. Furthermore, we used the developed reversed phase liquid chromatography method for an in-process control (IPC). In conclusion, the presented reversed phase liquid chromatography method allows a proper control of IB solubilization applicable for tailored refolding.

**Keywords:** inclusion bodies; inclusion body solubilization; tailored refolding; reversed phase liquid chromatography; process analytical technology tools; in-process monitoring; in-process control

## 1. Introduction

To date, approximately 20–30% of all approved biopharmaceuticals are produced in microbial hosts [1,2]. Insoluble aggregates, better known as inclusion bodies (IBs), produced by the gram-negative bacterium *Escherichia coli*, present a dominant fraction of the microbial production segment [3]. This is mainly because cultivation with *E. coli* can be carried out at very low costs in short fermentation run-times and high target protein concentrations at a high purity can be achieved [4–6]. Early downstream steps in IB processing, however, are notorious for their high complexity and low yields [7,8]. In particular, refolding is regarded as a major bottleneck in IB processing. Solubilization and refolding strategies are commonly developed empirically with protocols being highly dependent on the target protein [9,10]. Protein hydrophobicity, for instance, affects the required molarity of the chaotropic agent and pH in solubilization and refolding [11]. The addition of reducing agents during solubilization is required for proteins containing disulfide bonds [8], subsequently influencing the amount of oxidizing compounds to be added in refolding [12]. Refolding yields are highly dependent on the protein in question as well as the protein concentration during the refolding process: only 15–25% of refolding



yields are reported for many therapeutic proteins [13] compared with a refolding yield of 97% for the enzyme lysozyme [14].

Singh et al. reported that mild solubilization boosts the refolding yield [15]. The reduction of chaotropic reagent molarity was compensated by either highly alkaline conditions or via the addition of solubilization-enhancing chemicals (i.e., n-propanol) [16,17]. The analyses in these studies, however, were performed with the commonly known sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [17]. SDS-PAGE is frequently used for a solubilization analysis as it tolerates the harsh conditions [18,19]. Various staining protocols (e.g., silver staining or Coomassie staining) with a different selectivity and treatment time have been established [20] and subsequent immunoblotting can further increase the sensitivity of gel-based analytical methods [21].

Unfortunately, this gel-based method is not suitable for in-process control (IPC) as the sample treatment, method running time and staining protocols are very time-consuming. However, according to quality by design (QbD) principles in pharmaceutical manufacturing, an IPC must be applied [22] and, for this purpose, process analytical technology (PAT) tools are of a great advantage as they allow a timely process intervention [23,24]. High molarities of chaotropic reagents required for solubilization, however, interfere with the measuring principles of many available PAT tools (i.e., near-infrared and Raman spectroscopy) [25,26].

Several liquid chromatography (LC) separation principles can tolerate the high molarities of the chaotropic agents required for the sample dissolution [27,28]. LC is implemented for the quantification of diverse biopharmaceuticals due to the straightforward and facile sample preparation as well as a rapid and accurate analysis [29–31]. Hydrophobic interaction chromatography (HIC), ion exchange chromatography (IEX) and size exclusion chromatography (SEC) are frequently used in a recombinant protein analysis [32,33]; however, high molarities of the chaotropic reagent required in IB solubilization again complicate the implementation of these chromatographic techniques. Reversed phase liquid chromatography (RPLC) is a denaturing chromatographic technique tolerating these harsh conditions [18]. Furthermore, RPLC is known for its robustness, high selectivity and compatibility with a mass spectrometric analysis [34,35] making it a highly suitable technique for in-process monitoring in IB processing.

In this study, we implemented an RPLC method applicable for an at-line solubilization analysis. In this context, at-line defines a measurement that is performed in close proximity to the process stream as specified by regulatory agencies [36]. To demonstrate the need for in-process monitoring, we used HRP (horseradish peroxidase) as a model protein. HRP contains four disulfide bonds, thus requiring a complex solubilization and refolding strategy. We hypothesize that RPLC is faster and more precise than SDS-PAGE to determine the optimal solubilization conditions for tailored refolding. Hence, the solubilization conditions for HRP were varied and an SDS-PAGE analysis was compared with an RPLC analysis. The obtained results confirmed the effect of the solubilization on the subsequent refolding step. In addition, the in-process control based on at-line RPLC measurements was demonstrated. The results of this study demonstrate that solubilization quality influences the refolding yield and RPLC is suitable for the IPC of IB solubilization.

## 2. Materials and Methods

### 2.1. Production and Isolation of HRP IB

HRP was produced in *E. coli* BL21 (DE3) (Life Technologies, Carlsbad, CA, USA) with details stated in previous publications [37,38]. After a successful HRP expression [37,39,40], the cell broth was harvested via centrifugation and the biomass was stored at  $-20^{\circ}\text{C}$ .

Cell disruption was carried out at 1200 bar for 3 passages using a high-pressure homogenizer (PANDA+ 2000, GEA, Biberach, Germany). After centrifugation at 10,000 rpm,  $4^{\circ}\text{C}$  and 20 min (Eppendorf, Hamburg, Germany), the soluble fraction was discarded. The resulting IB pellet was washed two times with a buffer (50 mM Tris, pH 8, 500 mM NaCl,

2 M Urea). The IB pellets were then resuspended in water, aliquoted to a defined wet IB weight, centrifuged (20,379 rcf, 4 °C, 20 min) and stored at −20 °C until further use.

## 2.2. Solubilization and Refolding

### 2.2.1. In-Process Monitoring

The aliquoted IBs were thawed and resuspended in an HRP solubilization buffer (50 mM Tris/HCl pH 8; 6 M Urea). After the resuspension, DTT (1 M DTT stock) was added to a final concentration of 0 mM, 7.11 mM and 14.22 mM, respectively. The solubilization was performed at RT and a slight agitation and the samples were drawn after 0.5 h, 2 h, 4 h, 6 h, 8 h and 21 h. The samples were centrifuged (20,379 rcf, 4 °C, 20 min) and the supernatant was used for refolding, the RPLC analysis and diluted 1:2 in 2x Laemmli buffer for the SDS-PAGE analysis. For refolding, the solubilize was diluted 1:40 in a pre-cooled HRP refolding buffer (20 mM Tris/HCl pH 8.5; 2 M Urea; 2 mM CaCl<sub>2</sub>; 7% v/v glycerol) containing either 0 mM, 1.27 mM or 2.54 mM GSSG (glutathione disulfide) [37]. Refolding was performed for 48 h at 4 °C on a rocker-shaker. Hemin was added 24 h after the refolding start to a final concentration of 20 µM (1 mM Hemin stock in 100 mM KOH). After refolding, the enzyme activity was measured as described previously [37].

### 2.2.2. Demonstration of At-Line RPLC for IPC

In order to demonstrate the applicability of the at-line RPLC method for the IPC, we produced two different IB batches. Fermentations for both batches were conducted as described here [39] only the specific feeding rate ( $q_s$ ) during the induction was varied. Batch 1 was conducted at a  $q_s$  of 0.25 g/g/h during the induction whereas Batch 2 was performed at a  $q_s$  of 0.35 g/g/h during the induction. The harvest, cell disruption, IB wash, aliquoting and storage were done identically for both batches.

The inclusion bodies from both batches were solubilized at a concentration of 100 g wet IB/L solubilization mix. The solubilization buffer consisted of 50 mM glycine pH 10; 6 M Urea. DTT was added to a final concentration of 7.11 mM (1 M DTT stock). The solubilization was performed for 0.5 h at RT on a rocker-shaker. After centrifugation (20,379 rcf, 4 °C, 20 min), the supernatant was analyzed using the described RPLC method. Batch 1 solubilize was diluted 1:40 in a refolding buffer (20 mM glycine pH 10; 2 M Urea; 2 mM CaCl<sub>2</sub>; 7% v/v glycerol; 1.27 mM GSSG). For IB Batch 2, two different refolding approaches were performed: the solubilize was diluted 1:40 (= "fixed dilution") and the dilution was adapted based on the RPLC results in order to achieve the same monomeric HRP concentration as in Batch 1. In this case, the dilution was reduced to 1:17 due to a lower concentration of the target protein HRP in IB Batch 2.

## 2.3. Analytical Techniques

### 2.3.1. Reversed Phase Liquid Chromatography Measurements

The RPLC measurements were performed using a Dionex UltiMate 3000 system with a quaternary solvent delivery pump, an auto-sampler with a sample thermostat and a UV detector (Thermo Fisher, Waltham, MA, USA). The wavelength for the UV detection was set to 280 nm in order to monitor the protein absorption allowing the quantification of the target protein and its impurities. The instrument control and data acquisition were carried out via Chromeleon 7.2 software (Thermo Fisher). Prior to the RPLC measurement, all samples were centrifuged (20,379 rcf, 4 °C, 20 min) to separate the aggregates from the soluble fraction. We used a BioResolve RP mAb Polyphenyl column (dimensions 100 mm × 3 mm, particle size 2.7 µm) (Waters Corporation, MA, USA) connected to a pre-column (3.9 mm × 5 mm, 2.7 µm) of the same stationary phase. The mobile phase was composed of ultrapure water (MQ; eluent A) and acetonitrile (eluent B) both supplemented with 0.1% (v/v) trifluoroacetic acid. Ultrapure water was acquired from a Milli-Q system from Merck Millipore (Darmstadt, Germany). Acetonitrile (HPLC-grade) and trifluoroacetic acid (TFA, >99.9%) were obtained from Carl Roth (Karlsruhe, Germany). For the analysis, a recently published RPLC method, which had been developed



and validated according to QbD principles, was modified in terms of gradient, column temperature and flow rate [18]. The method was optimized empirically, reducing the overall running time to 8.1 min in order to allow for short analysis times during solubilization. The runs were conducted at 78 °C and a flow rate of 1.2 mL/min using the gradient displayed in Table 1.

**Table 1.** Gradient used for the RPLC analysis, with eluent A being ultrapure water and eluent B acetonitrile both supplemented with 0.1% (v/v) trifluoroacetic acid.

Time (min)	Percent Eluent B (%)
0	25
3.1	62
5.1	62
5.2	25
8.1	25

The HRP concentration in the solubilize was determined using a bovine serum albumin (BSA) standard calibration ranging from 0.0625 g/L to 1.0000 g/L. The BSA was used for the quantification as no non-glycosylated standard was available for HRP.

### 2.3.2. SDS-PAGE Measurements

For the SDS-PAGE analysis, the samples were mixed with 2× concentrated Laemmli buffer to achieve a 1× concentration of Laemmli buffer in the final dilution. The used buffer solution did not contain β-mercaptoethanol (non-reducing conditions) in order to analyze the solubilization quality in regard to disulfide bridge formation. The samples were heated to 95 °C for 10 min. A total of 5 µL of each sample was loaded onto pre-cast SDS gels (4–15%, Bio-Rad, Hercules, CA, USA). The gels were run in a Mini-PROTEAN Tetra System (Bio-Rad) for 30 min at 180 V and stained with Coomassie Blue. The protein bands were evaluated densitometrically using ImageLab software (Bio-Rad).

### 2.3.3. HRP Enzymatic Activity Assay

The HRP enzyme activity was measured with a Tecan Infinite M200 PRO (Männedorf, Switzerland) using flat-bottom polystyrene 96-well plates, as described previously [37]. The samples after refolding were diluted in the range of 1:2–1:25 in a dilution buffer (20 mM Bis-Tris; pH 7; 7% v/v glycerol) depending on their volumetric activity. A total of 170 µL of ABTS solution (50 mM KH<sub>2</sub>PO<sub>4</sub>; pH 5; 5 mM ABTS) was mixed with 10 µL of the respective diluted sample in the well. The reaction was started by adding 20 µL hydrogen peroxide (the final concentration in the well was 1 mM). Immediately after the start of the reaction, the change in the absorbance at 420 nm was recorded at 30 °C for 2 min. The volumetric enzyme activity was calculated using Equation (1):

$$A \left[ \frac{U}{mL} \right] = \frac{V_{total} * \Delta \frac{A}{min} * Dilution}{V_{sample} * d * \epsilon} \quad (1)$$

$V_{total}$ —total well volume (µL).

$\Delta A/min$ —change in absorption ( $\Delta A_{420}$  nm/min).

$Dilution$ —dilution of the sample.

$V_{sample}$ —volume of the sample (µL).

$d$ —length of the beam path through the well ( $d = 0.58$  cm).

$\epsilon$ —extinction coefficient ( $\epsilon_{420} = 36 \text{ mM}^{-1} \text{ cm}^{-1}$ ).

## 2.4. Experimental Design

The experiments were conducted using a full factorial design, as shown in Table 2, varying the solubilization time and DTT concentrations in a multivariate data approach. All combinations of the solubilization time and DTT concentration are listed in Supplementary

Table S1. As responses for this design of experiment (DoE), the concentration of monomeric HRP was quantified with SDS-PAGE and RPLC. Additionally, all solubilization conditions were refolded at a GSSG concentration of either 0 mM, 1.27 mM or 2.54 mM GSSG to verify the effect on the refolding yield.

**Table 2.** Full factorial experimental design for HRP solubilization. The listed DTT concentrations were used in combination with each listed time-point.

Time (h)	DTT Concentrations (mM)
0.5	0
2	7.11
4	14.22
6	-
8	-
21	-

## 2.5. Multivariate Data Assessment of Solubilization and Refolding

An analysis of the used design of experiments (DoEs) was done using a multivariate data assessment program (MODDE 12, Umetrics, Sweden) with the model being based on a multiple linear regression. The results were analyzed for the statistical relevance of the model by the measure of fit ( $R^2$ ) and the model predictability ( $Q^2$ ).

## 3. Results and Discussion

### 3.1. In-Process Monitoring

In order to be applicable as an in-process monitoring method for solubilization, the respective analytical method had to meet three criteria:

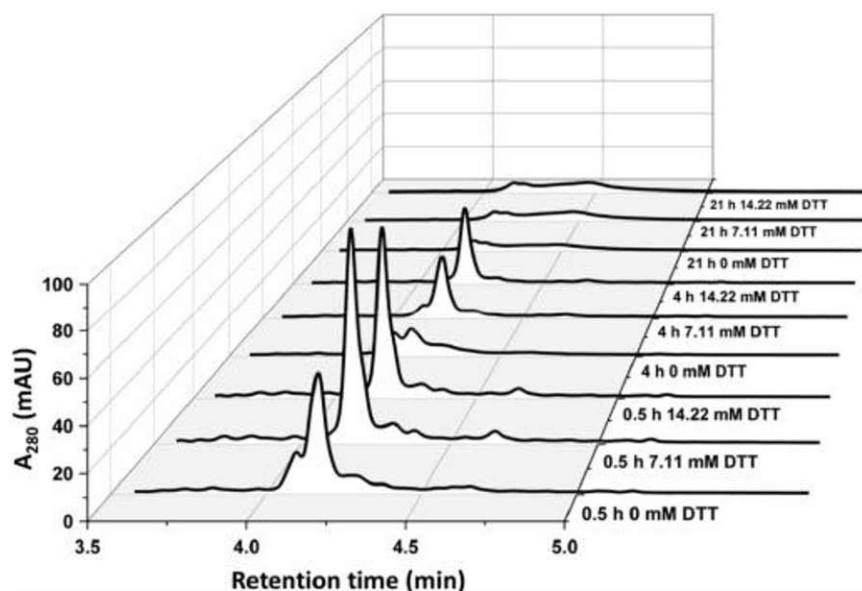
1. The ability to quantify a defined quality attribute;
2. The quality attribute had to influence the refolding behavior;
3. Timely measurement of the respective quality attribute.

In order to determine whether at-line RPLC met the criteria of an in-process monitoring tool, DTT concentration and the solubilization time of HRP IBs were varied in a DoE approach (Table 2). DTT was required during solubilization in order to keep the cysteines (eight contained in HRP) in a reduced state and therefore prevent aggregation caused by an intermolecular disulfide bridge formation. It was expected that DTT concentration and the solubilization time would influence the titer of monomeric HRP in the solubilize.

The key quality attribute (=DoE response) was defined as the concentration of monomeric HRP. We hypothesized that the enzymatic activity after refolding would directly correlate with the concentration of monomeric HRP in the solubilize. The aggregated HRP was believed to result in a structure not applicable for refolding whereas monomeric HRP was defined as completely reduced and denatured during solubilization. Therefore, a monomeric HRP titer was chosen as the target response for the DoE. To test whether both RPLC and SDS-PAGE could predict the targeted solubilization for tailored refolding, all solubilizes were refolded and the volumetric activity (U/mL) after refolding was recorded.

Figure 1 shows that the solubilize containing no DTT displayed an additional peak before the HRP target peak (at a 4.1 min retention time). This peak could already be monitored after 0.5 h of solubilization and increased with longer solubilization times. A complete degradation of the target peak could be observed after 21 h of solubilization independent of the supplied DTT concentration. The pellets resulting from the centrifugation prior to the RPLC analysis after 21 h solubilization increased compared with pellets received from shorter solubilization times, indicating an enhanced protein aggregate formation [7].

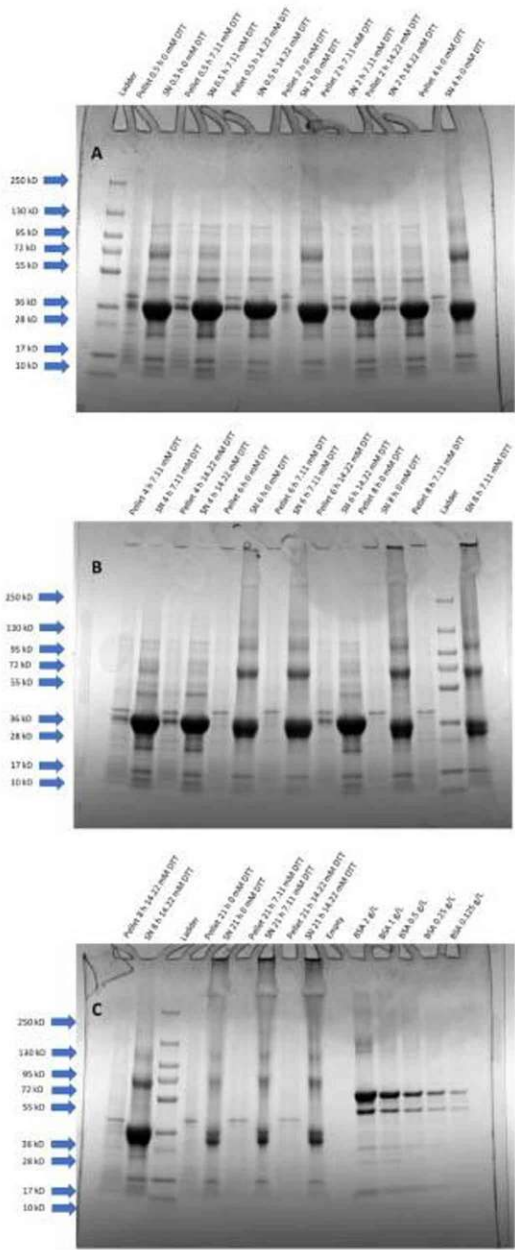




**Figure 1.** RPLC chromatograms at 280 nm quantifying monomeric HRP eluting at 4.18 min. The results demonstrate the trends of solubilization at three applied DTT concentrations (0 mM, 7.11 mM and 14.22 mM) for 0.5 h of solubilization, 4 h of solubilization and 21 h of solubilization. After 21 h of solubilization, a strong degradation of the target peak in the solubilize is visible.

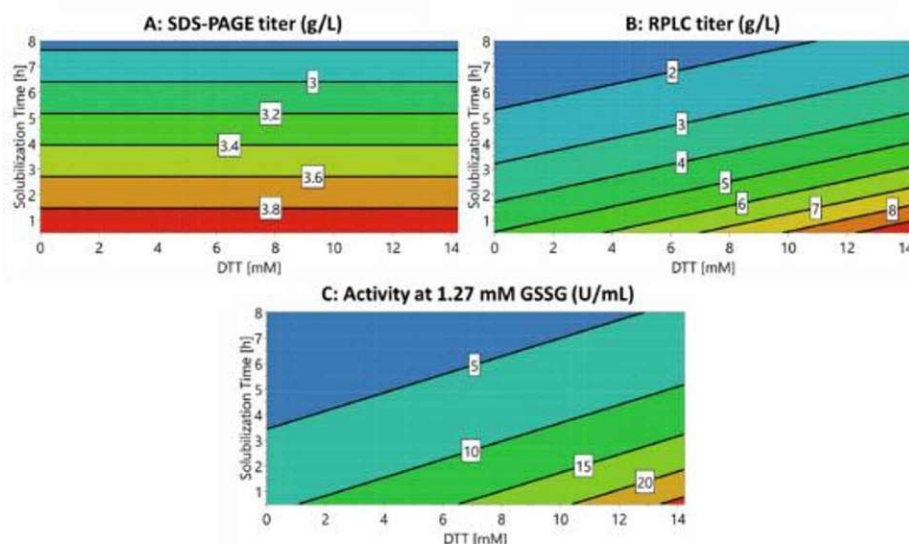
For the SDS-PAGE analysis, the HRP target band was found at 34 kDa (Figure 2). We hypothesized that the protein band at approximately 68 kDa was a product dimer due to an intermolecular disulfide bridge formation. The 68 kDa band tended to increase with extended solubilization times, especially beyond 4 h and for experiments without DTT. In good agreement with the RPLC measurements, we found that HRP was substantially degraded after 21 h of solubilization (Figure 2C). It was therefore concluded that both factors varied for this DoE (DTT concentration and solubilization time) had an influence on the quality attribute “monomeric HRP concentration” in the solubilize. Furthermore, SDS-PAGE and RPLC were able to measure these changes, with RPLC being applicable as an at-line monitoring tool due to its short analysis time of less than 10 min.

A multivariate data approach was applied in order to quantify the effects of the varied DTT concentrations and solubilization times on the monomeric HRP concentration during solubilization. The contour plots (i.e., model responses) of the multivariate data analysis for SDS-PAGE and RPLC in solubilization and the enzymatic activity after refolding are shown in Figure 3. All concentrations of HRP during solubilization (including respective purity) for the shown experiment are listed in Supplementary Table S1 for both SDS-PAGE and RPLC. The results for 21 h of solubilization (Figures 1 and 2, Supplementary Table S1) were excluded from the model as they led to a response distortion due to product degradation. The model terms ( $R^2$  and  $Q^2$ ) are summarized in Supplementary Figure S1. Furthermore, ANOVA plots for the responses and significant factors used for the models are shown in Supplementary Figures S2 and S3, respectively.



**Figure 2** SDS-PAGE analysis of HRP solubilizes with varying DTT concentrations and time factors as depicted in Table 2; non-glycosylated HRP is visible at 34 kDa. The potential dimer formed due to the intermolecular disulfide bridge formation can be seen at 68 kDa. The samples are displayed in the following order from (A–C): pellet and supernatant (SN) for each time-point varying the three altered DTT concentrations. Additionally to samples, in (C), a standard protein calibration with BSA was performed. Protein ladders were added to confirm protein size.





**Figure 3.** MODDE contour plots with the two factors of DTT concentration on the X-axis and the solubilization time on the Y-axis. The following responses are shown: (A) Monomeric HRP concentration in the solubilize (g/L) analyzed using SDS-PAGE. (B) Monomeric HRP concentration in the solubilize (g/L) analyzed using RPLC. (C) Effect of the different solubilization conditions on the volumetric activity (U/mL) after refolding, which was performed at constant GSSG conditions of 1.27 mM.

The results obtained for the model responses varied between the SDS-PAGE and the RPLC analysis for the conducted DoE (Figure 3A,B). Surprisingly, the SDS-PAGE analysis indicated that the monomeric HRP concentration was solely dependent on the solubilization time (Figure 3A). The raw data of both SDS-PAGE and RPLC (Figures 1 and 2) showed that short solubilization times below 4 h were superior over longer solubilization times. However, the raw data also indicated that high DTT concentrations were beneficial for the concentration of monomeric HRP during solubilization whereas lower DTT concentrations led to a lower solubilization yield. This effect was especially true for longer solubilization times. The model prediction for the SDS-PAGE analysis was very low in comparison with the RPLC analysis (Supplementary Figure S1). We hypothesized that this trend occurred due to different separation principles leading to non-linear model responses. The model response for the RPLC analysis with the monomeric HRP concentration indicated that both DTT concentration and the solubilization time had a significant influence (Figure 3B).

The effects on the product purity during solubilization can be found in Supplementary Table S1 and Figure S4 for both SDS-PAGE and RPLC, indicating the same trends as for the monomeric HRP concentration shown in Figure 3A,B. The differences in the total titer between the SDS-PAGE analysis and RPLC might result from the high sample concentrations chosen for SDS-PAGE. This was done to determine the impurities and monomeric HRP titer in the SDS-PAGE analysis. For the RPLC analysis, monomeric HRP and purity could be well assessed within one chromatogram.

To assess which analytical method was better suited for refolding yield prediction, the volumetric activity (U/mL) after refolding was determined (Figure 3C). The GSSG concentration was varied according to previous experiments [37]; however, no alterations in the model trend were obtained (Supplementary Table S1). The effects are exemplarily shown for 1.27 mM GSSG in Figure 3C. The enzyme activity after refolding was found to be highest at short solubilization durations (0.5 h, Figure 3C) and an increase in DTT concentration led to increased enzyme activity. Hence, the enzyme activity showed the same trends as

RPLC solubilize prediction (Figure 3B,C) while solubilizes quantified via SDS-PAGE led to a different prediction than enzymatic activity after refolding (Figure 3A,C).

Therefore, the presented RPLC method met the criteria required for an at-line monitoring tool for the solubilization of HRP IBs. In comparison with the commonly used SDS-PAGE, the models based on the RPLC data were able to predict the influence of DTT concentration and time during solubilization on the subsequent refolding step and the refolding yield correctly. Furthermore, due to the versatility and short analysis time [18], RPLC could be applied as a suitable technique for the in-process control of IB processes.

### 3.2. Demonstration of At-Line RPLC for IPC

The refolding yield highly depends on the protein concentration during refolding with lower protein concentrations favoring higher refolding yields [6,9]. The protein concentration that leads to a maximum space-time yield of a correctly folded protein is empirically determined during process development of refolding [37]. The protein concentration in solubilization is commonly controlled by a fixed amount of wet IB weight dissolved in a solubilization buffer and a subsequent fixed dilution in the refolding buffer [10]. However, this approach is highly dependent on a rigid HRP titer per wet IB weight as well as a constant solubilization yield. It guarantees the desired concentration of the target product during solubilization and the subsequent refolding. In-process monitoring of the target protein concentration in the solubilize is necessary to counteract batch to batch variations derived from upstream processing (USP) and varying yields during solubilization.

In order to evaluate if the presented RPLC method was suited for the IPC, we conducted solubilization and refolding experiments comparing two IB batches. Early DSP until the refolding step were kept constant for both IB batches. The refolding was performed by two different approaches:

1. Solubilizes were diluted at a fixed ratio of 1:40, as previously determined [37];
2. Solubilizes were diluted according to the product quantity assessed by RPLC (Table 3).

**Table 3.** HRP concentration in the solubilize was determined via RPLC. Furthermore, the dilution in the refolding buffer as well as the activity after refolding are given for the two different IB batches. Two different dilutions were done for Batch 2: (1) solubilize was diluted with a fixed dilution (1:40) and (2) dilution was adapted to achieve the same HRP concentration as for Batch 1 (1:17).

IB Batch	c(HRP) (g/L) in Solubilization	Applied Dilution	Activity (U/mL) after Refolding
Batch 1, Fixed Dilution	5.27 ± 0.11	1:40	89.7 ± 6.0
(1) Batch 2, Fixed Dilution	2.35 ± 0.05	1:40	41.9 ± 2.8
(2) Batch 2, IPC via RPLC	2.35 ± 0.05	1:17	79.8 ± 5.4

As shown in Table 3, the concentration of monomeric HRP in the solubilize varied from Batch 1 (5.27 g/L) to Batch 2 (2.35 g/L). This led to an over 50% decrease of enzymatic activity after refolding for Batch 2 if the empirically determined fixed dilution of 1:40 was applied. When using the RPLC method as an in-process control tool, the dilution for Batch 2 could be adapted to 1:17. While the refolding yield for Batch 2 only decreased minimally when adjusting the dilution from 1:40 to 1:17 (Table 3), the refolding buffer volume could be reduced by more than 50%. Therefore, using a correction via RPLC, the 54% variation caused by USP could be reduced to only 11%. However, for the IPC in solubilization and refolding, the following factors still needed to be considered to elucidate the activity deviation of 11%:

1. The corrected dilution in refolding led to a variation in the redox system because a higher DTT carry-over occurred at lower dilutions. This shift of the redox system potentially influences the refolding yield [37].
2. For the demonstration purpose of the IPC via RPLC, drastic deviations from the USP were targeted (i.e. 54% of titer deviation). However, the protein concentration



adjustment in refolding via at-line RPLC from solubilization might be even less error-prone for smaller deviations.

3. Furthermore, standard deviations resulting from the RPLC measurements and enzymatic assay could explain further deviations (Table 3).

In the case of HRP IBs, the presented RPLC method was applied successfully as an IPC tool. Based on the rapid analysis time of the developed RPLC method, the deviations caused by the USP could be monitored at-line and the dilution was adapted to minimize the deviations during the refolding step. In addition, the received IB fingerprint (i.e., impurity monitoring) obtained via RPLC could provide valuable information using a reference impurity pattern for industrial applications.

#### 4. Conclusions

The solubilization of IBs prior to refolding is essential to obtain the desired protein conformation and protein concentrations in refolding. In this study, we developed an at-line RPLC method to monitor the target protein concentration during the solubilization unit operation of IB processing. DTT concentration and the solubilization time were varied for HRP IBs and the monomeric HRP concentration was recorded using SDS-PAGE and RPLC. The short analysis time (8.1 min), facile sample preparation and the high accuracy of RPLC (as demonstrated for the generic method [18]) allowed for a precise prediction of the monomeric HRP concentration on the refolding yield. The results thus favored the RPLC analysis over the SDS-PAGE analysis as the former could also be used for in-process monitoring.

Moreover, we demonstrated the IPC making use of the developed RPLC method, determining USP alterations. As downstream operations are performed sequentially in industry, subsequent unit operations are influenced by initial deviations. In this study, the protein concentration in refolding could be adapted by adjusting the dilution factor based on the at-line RPLC analysis. This allowed for a more robust refolding process against the deviations contrived from the USP and a reduction of the refolding buffer compared with empirical dilutions thus facilitating a more economic process.

Concluding, the developed RPLC method can be applied to accelerate process development in IB solubilization and for in-process monitoring therefore allowing IPC, which facilitates tailored IB refolding.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/bioengineering8060078/s1>, Table S1: Conditions and raw data for the HRP solubilization DoE. Figure S1: Showing the measure of fit ( $R^2$ ) of the model and the model predictability ( $Q^2$ ) for the multivariate data analysis (based on multiple linear regression) conducted. SDSHRP conducts for titer measurements in solubilization performed with SDS-PAGE whereas SDSpurity shows the model for SDS impurity measurement in solubilization. It can be seen that these models show a low model predictability and a low measure of fit compared to other models in supplementary Figure S2. RPHRP and RPPurity cope for solubilization models of HRP titer and impurities respectively. Models after refolding are abbreviated according to their GSSG concentration i.e., 0 mM GSSG in refolding = vAct 0 mM GSSG, 1.27 mM GSSG in refolding = vAct 1.27 mM GSSG and 2.54 mM GSSG in refolding = vAct 2.54 mM GSSG; All model except for SDS-PAGE prediction (i.e. SDSHRP and SDSImp) show a  $R^2$  close to 0.8 and  $Q^2$  close to 0.7 and can thus be regarded as models describing input data appropriately. Figure S2: ANOVA plots are displayed for utilized responses. SDSHRP displays the concentration of monomeric HRP measured using SDS-PAGE, whereas SDSpurity shows the purity of the monomeric HRP analyzed via SDS-PAGE. RPHRP displays the concentration of monomeric HRP measured using RPLC and RPPurity shows the purity of monomeric HRP measured using RPLC analysis. vAct 1.27 mM GSSG shows the volumetric activity [U/mL] after refolding with 1.27 mM GSSG contained in the refolding buffer. For each response, SD-regression shows the variation of the response explained by the model while the RSD shows the variation of the response which is not explained by the model. Both values are adjusted for the respective degrees of freedom.  $RSD \cdot \sqrt{F(crit)}$  shows RDS multiplied by the square root of the critical F (statistically significant at the 95% confidence level). Figure S3: Significant factors contributing for the models

for the used responses. SDSHRP displays the concentration of monomeric HRP measured using SDS-PAGE, whereas SDSpurity shows the purity of the monomeric HRP analyzed via SDS-PAGE. RPHRP displays the concentration of monomeric HRP measured using RPLC and RPhpurity shows the purity of monomeric HRP measured using RPLC analysis. vAct 1.27 mM GSSG shows the volumetric activity [U/mL] after refolding with 1.27 mM GSSG contained in the refolding buffer. Abbreviated factors are: Tim is the solubilization time [h], DTT is the DTT concentration during solubilization [mM]. For both SDS-PAGE responses (SDSHRP and SDSpurity), only the solubilization time was identified as a significant factor. For the RPLC responses as well as the volumetric activity after refolding both the DTT concentration during solubilization and the solubilization time were significant factors. Figure S4: Comparison of the two quality attributes monomeric HRP concentration [g/L] and purity [%] measured with SDS-PAGE and RPLC, respectively. MODDE contour plots with the two factors DTT concentration on the X-axis and the solubilization time on the Y-axis. The following responses are shown: A: Monomeric HRP concentration in the solubilize [g/L] analyzed using SDS-PAGE. B: Purity [%] of the monomeric HRP concentration in the solubilize analyzed via SDS-PAGE. C: Monomeric HRP concentration in the solubilize [g/L] analyzed using RPLC. D: Purity [%] of the monomeric HRP concentration in the solubilize analyzed via RPLC.

**Author Contributions:** J.E., O.S. and J.K. founded the idea of this study and planned the experimental design. J.E. carried out the data treatment and the statistical analysis. J.E., D.H., R.K. and V.R. conducted the experiments. R.P. and J.K. developed the RPLC method. J.E., O.S. and J.K. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Austrian Research Promotion Agency (Österreichische Forschungsförderungsgesellschaft = FFG), grant number 874206. Furthermore, the authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to cooperation with an industrial partner.

**Acknowledgments:** The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program. Furthermore, the authors want to thank the Austrian Research Promotion Agency (FFG) for their funding facilitating this study (874206).

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be interpreted as a potential conflict of interest.

## References

- Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **2018**, *36*, 1136–1145. [\[CrossRef\]](#) [\[PubMed\]](#)
- Baeshen, M.N.; Al-Hejin, A.M.; Bora, R.S.; Ahmed, M.M.; Ramadan, H.A.; Saini, K.S.; Baeshen, N.A.; Redwan, E.M. Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives. *J. Microbiol. Biotechnol.* **2015**, *25*, 953–962. [\[CrossRef\]](#) [\[PubMed\]](#)
- Slouka, C.; Kopp, J.; Spadiut, O.; Herwig, C. Perspectives of inclusion bodies for bio-based products: Curse or blessing? *Appl. Microbiol. Biotechnol.* **2018**. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rinas, U.; Garcia-Fruitós, E.; Corchero, J.L.; Vázquez, E.; Seras-Franzoso, J.; Villaverde, A. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem. Sci.* **2017**, *42*, 726–737. [\[CrossRef\]](#)
- García-Fruitós, E.; Vázquez, E.; Díez-Gil, C.; Corchero, J.L.; Seras-Franzoso, J.; Ratera, I.; Veciana, J.; Villaverde, A. Bacterial inclusion bodies: Making gold from waste. *Trends Biotechnol.* **2012**, *30*, 65–70. [\[CrossRef\]](#)
- Jungbauer, A. Continuous downstream processing of biopharmaceuticals. *Trends Biotechnol.* **2013**, *31*, 479–492. [\[CrossRef\]](#) [\[PubMed\]](#)
- Palmer, I.; Wingfield, P.T. Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. *Curr. Protoc. Protein Sci.* **2012**. [\[CrossRef\]](#) [\[PubMed\]](#)
- Humer, D.; Spadiut, O. Wanted: More monitoring and control during inclusion body processing. *World J. Microbiol. Biotechnol.* **2018**, *34*, 158. [\[CrossRef\]](#)
- Eiberle, M.K.; Jungbauer, A. Technical refolding of proteins: Do we have freedom to operate? *Biotechnol. J.* **2010**, *5*, 547–559. [\[CrossRef\]](#)



10. Walther, C.; Mayer, S.; Jungbauer, A.; Dürner, A. Getting ready for PAT: Scale up and inline monitoring of protein refolding of Npro fusion proteins. *Process. Biochem.* **2014**, *49*, 1113–1121. [\[CrossRef\]](#)
11. Wagner, J.R.; Sorgentini, D.A.; Ahón, M.C. Relation between Solubility and Surface Hydrophobicity as an Indicator of Modifications during Preparation Processes of Commercial and Laboratory-Prepared Soy Protein Isolates. *J. Agric. Food Chem.* **2000**, *48*, 3159–3165. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Uchimura, H.; Kim, Y.; Mizuguchi, T.; Kiso, Y.; Saito, K. Quantitative evaluation of refolding conditions for a disulfide-bond-containing protein using a concise  $^{18}\text{O}$ -labeling technique. *Protein Sci.* **2011**, *20*, 1090–1096. [\[CrossRef\]](#)
13. Datar, R.V.; Cartwright, T.; Rosen, C.-G. Process Economics of Animal Cell and Bacterial Fermentations: A Case Study Analysis of Tissue Plasminogen Activator. *Biotechnology* **1993**, *11*, 349–357. [\[CrossRef\]](#)
14. Ma, F.-H.; An, Y.; Wang, J.; Song, Y.; Liu, Y.; Shi, L. Synthetic Nanochaperones Facilitate Refolding of Denatured Proteins. *ACS Nano* **2017**, *11*, 10549–10557. [\[CrossRef\]](#)
15. Singh, A.; Upadhyay, V.; Panda, A.K. Solubilization and Refolding of Inclusion Body Proteins. In *Insoluble Proteins: Methods and Protocols*; García-Fruitós, E., Ed.; Springer: New York, NY, USA, 2015; pp. 283–291.
16. Singh, S.M.; Panda, A.K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **2005**, *99*, 303–310. [\[CrossRef\]](#)
17. Singh, S.M.; Sharma, A.; Upadhyay, A.K.; Singh, A.; Garg, L.C.; Panda, A.K. Solubilization of inclusion body proteins using n-propanol and its refolding into bioactive form. *Protein Expr. Purif.* **2012**, *81*, 75–82. [\[CrossRef\]](#)
18. Kopp, J.; Zauner, F.B.; Pell, A.; Hausjell, J.; Humer, D.; Ebner, J.; Herwig, C.; Spadiut, O.; Slouka, C.; Pell, R. Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. *J. Pharm. Biomed. Anal.* **2020**, 113412. [\[CrossRef\]](#)
19. Sasse, J.; Gallagher, S.R. Staining Proteins in Gels. *Curr. Protoc. Mol. Biol.* **2009**, *85*, 10.16.11–10.16.27. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Gallagher, S.; Sasse, J. Protein Analysis by SDS-PAGE and Detection by Coomassie Blue or Silver Staining. *Curr. Protoc. Pharmacol.* **1998**, *2*, A.3B.1–A.3B.10. [\[CrossRef\]](#)
21. Kurien, B.T.; Scofield, R.H. Western blotting. *Methods* **2006**, *38*, 283–293. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Pramod, K.; Tahir, M.A.; Charoo, N.A.; Ansari, S.H.; Ali, J. Pharmaceutical product development: A quality by design approach. *Int. J. Pharm. Investig.* **2016**, *6*, 129–138. [\[CrossRef\]](#)
23. Kopp, J.; Slouka, C.; Spadiut, O.; Herwig, C. The Rocky Road from Fed-Batch to Continuous Processing with *E. coli*. *Front. Bioeng. Biotechnol.* **2019**, *7*, 328. [\[CrossRef\]](#)
24. Esmonde-White, K.A.; Cuellar, M.; Lenain, B.; Lewis, I.R. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Anal. Bioanal. Chem.* **2017**, *409*, 637–649. [\[CrossRef\]](#)
25. Ferrari, M.; Mottola, L.; Quaresima, V. Principles, techniques, and limitations of near infrared spectroscopy. *Can. J. Appl. Physiol.* **2004**, *29*, 463–487. [\[CrossRef\]](#)
26. Eberhardt, K.; Stiebing, C.; Matthäus, C.; Schmitt, M.; Popp, J. Advantages and limitations of Raman spectroscopy for molecular diagnostics: An update. *Expert Rev. Mol. Diagn.* **2015**, *15*, 773–787. [\[CrossRef\]](#)
27. Gustavsson, R.; Mandenius, C.-F. Soft sensor control of metabolic fluxes in a recombinant *Escherichia coli* fed-batch cultivation producing green fluorescence protein. *Bioprocess. Biosyst. Eng.* **2013**, *36*, 1375–1384. [\[CrossRef\]](#)
28. Chemmalil, L.; Prabhakar, T.; Kuang, J.; West, J.; Tan, Z.; Ehamparanathan, V.; Song, Y.; Xu, J.; Ding, J.; Li, Z. Online/at-line measurement, analysis and control of product titer and critical product quality attributes (CQAs) during process development. *Biotechnol. Bioeng.* **2020**, *117*, 3757–3765. [\[CrossRef\]](#)
29. Sandra, K.; Vandenhede, I.; Sandra, P. Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. *J. Chromatogr. A* **2014**, *1335*, 81–103. [\[CrossRef\]](#)
30. Fekete, S.; Guilleme, D. Ultra-high-performance liquid chromatography for the characterization of therapeutic proteins. *TrAC Trends Anal. Chem.* **2014**, *63*, 76–84. [\[CrossRef\]](#)
31. Fekete, S.; Veuthey, J.-L.; Guilleme, D. Modern Column Technologies for the Analytical Characterization of Biopharmaceuticals in Various Liquid Chromatographic Modes. *Spec. Issues* **2015**, *34*, 6–13.
32. Baca, M.; De Vos, J.; Bruylants, G.; Bartik, K.; Liu, X.; Cook, K.; Eeltink, S. A comprehensive study to protein retention in hydrophobic interaction chromatography. *J. Chromatogr. B* **2016**, *1032*, 182–188. [\[CrossRef\]](#)
33. Fekete, S.; Goyon, A.; Veuthey, J.-L.; Guilleme, D. Size Exclusion Chromatography of Protein Biopharmaceuticals: Past, Present and Future. *Am. Pharm. Rev.* **2018**, 1–4.
34. Kochling, J.; Wu, W.; Hua, Y.; Guan, Q.; Castaneda-Merced, J. A platform analytical quality by design (AQbD) approach for multiple UHPLC-UV and UHPLC-MS methods development for protein analysis. *J. Pharm. Biomed. Anal.* **2016**, *125*, 130–139. [\[CrossRef\]](#)
35. Fekete, S.; Veuthey, J.-L.; Guilleme, D. New trends in reversed-phase liquid chromatographic separations of therapeutic peptides and proteins: Theory and applications. *J. Pharm. Biomed. Anal.* **2012**, *69*, 9–27. [\[CrossRef\]](#)
36. Rathore, A.S.; Bhambure, R.; Ghare, V. Process analytical technology (PAT) for biopharmaceutical products. *Anal. Bioanal. Chem.* **2010**, *398*, 137–154. [\[CrossRef\]](#)
37. Humer, D.; Ebner, J.; Spadiut, O. Scalable High-Performance Production of Recombinant Horseradish Peroxidase from *E. coli* Inclusion Bodies. *Int. J. Mol. Sci.* **2020**, *21*, 4625. [\[CrossRef\]](#)

38. Humer, D.; Spadiut, O. Improving the Performance of Horseradish Peroxidase by Site-Directed Mutagenesis. *Int. J. Mol. Sci.* **2019**, *20*, 916. [\[CrossRef\]](#)
39. Slouka, C.; Kopp, J.; Hutwimmer, S.; Strahammer, M.; Strohmer, D.; Eitenberger, E.; Schwaighofer, A.; Herwig, C. Custom made inclusion bodies: Impact of classical process parameters and physiological parameters on inclusion body quality attributes. *Microb. Cell Fact.* **2018**, *17*, 148. [\[CrossRef\]](#)
40. Kopp, J.; Kittler, S.; Slouka, C.; Herwig, C.; Spadiut, O.; Wurm, D.J. Repetitive Fed-Batch: A Promising Process Mode for Biomanufacturing with *E. coli*. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1312. [\[CrossRef\]](#)



# QCL-IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography

Christopher K. Akhgar, Julian Ebner, Oliver Spadiut, Andreas Schwaighofer,\* and Bernhard Lendl\*



Cite This: <https://doi.org/10.1021/acs.analchem.1c05191>



Read Online

ACCESS |

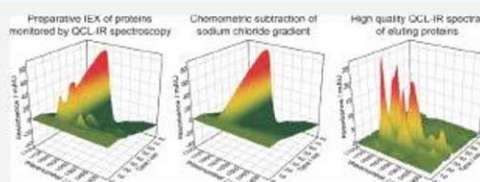
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** In this study, an external cavity-quantum cascade laser-based mid-infrared (IR) spectrometer was applied for in-line monitoring of proteins from preparative ion-exchange chromatography. The large optical path length of 25  $\mu\text{m}$  allowed for robust spectra acquisition in the broad tuning range between 1350 and 1750  $\text{cm}^{-1}$ , covering the most important spectral region for protein secondary structure determination. A significant challenge was caused by the overlapping mid-IR bands of proteins and changes in the background absorption of water due to the NaCl gradient.

Implementation of advanced background compensation strategies resulted in high-quality protein spectra in three different model case studies. In Case I, a reference blank run was directly subtracted from a sample run with the same NaCl gradient. Case II and III included sample runs with different gradient profiles than the one from the reference run. Here, a novel compensation approach based on a reference spectra matrix was introduced, where the signal from the conductivity detector was employed for correlating suitable reference spectra for correction of the sample run spectra. With this method, a single blank run was sufficient to correct various gradient profiles. The obtained IR spectra of hemoglobin and  $\beta$ -lactoglobulin were compared to off-line reference measurements, showing excellent agreement for all case studies. Moreover, the concentration values obtained from the mid-IR spectrometer agreed well with conventional UV detectors and high-performance liquid chromatography off-line measurements. LC-QCL-IR coupling thus holds high potential for replacing laborious and time-consuming off-line methods for protein monitoring in complex downstream processes.



Preparative liquid chromatography (prep-LC) remains an essential unit operation in downstream processing of complex biopharmaceuticals.<sup>1</sup> The most widespread detection method for monitoring the corresponding protein concentrations is single-wavelength UV/vis spectroscopy, offering a broad dynamic range and excellent sensitivity. The obtained univariate signal, however, does not give any information about the protein structure and purity during elution.<sup>2</sup> Thus, the collected fractions have to be additionally analyzed by laborious and time-consuming off-line methods in order to measure critical quality attributes (CQAs). Commonly used off-line methods are high-performance liquid chromatography (HPLC), sodium dodecyl sulphate-polyacrylamide gel electrophoresis, Western blotting and biological activity assays, such as enzyme-linked immunosorbent assays. In recent years, quality by design (QbD) principles were established in the pharmaceutical manufacturing sector.<sup>3</sup> In order to comply with these regulatory guidelines, process analytical technology (PAT) tools are required, allowing for in-process control and timely adaption of process parameters. Therefore, analytical methods able to provide information about CQAs in an in-line or on-line measurement setup are required.<sup>4,5</sup>

Mid-infrared (IR) spectroscopy provides detailed information about rotational-vibrational transitions of proteins. The established technique in this spectral region is Fourier

transform infrared (FTIR) spectroscopy, which is routinely used for quantitative and qualitative analysis of infrared absorption spectra. The most important mid-IR regions for protein quantification and secondary structure determination are the amide I (1700–1600  $\text{cm}^{-1}$ )<sup>6,7</sup> and amide II (1600–1500  $\text{cm}^{-1}$ )<sup>8</sup> band. Coupling of IR spectroscopy and LC were successfully demonstrated,<sup>9,10</sup> in most cases using organic solvent gradients. However, flow-through mid-IR transmission measurements of proteins in aqueous matrix have several challenges. One of these constraints arises from the overlap of the HOH-bending band of water near 1643  $\text{cm}^{-1}$  with the protein amide I band. A second limitation is the low emission power provided by thermal light sources (Globars) that are used in FTIR spectrometers. Thus, using FTIR spectroscopy, the optical pathlength for aqueous protein solutions is restricted to <10  $\mu\text{m}$  for transmission measurements in order to avoid total IR light absorption of water.<sup>11,12</sup> These low path-

Received: November 30, 2021

Accepted: March 21, 2022



lengths are unsuitable for in-line flow-through measurements due to low robustness and limited sensitivity. Consequently, mid-IR spectroscopy found its way into in-line detection of preparative protein chromatography effluents only recently.<sup>13,14</sup> In these studies, attenuated total reflection-FTIR spectroscopy was applied, offering robust spectra acquisition, but limited sensitivity. As an alternative approach, complex solvent-removal setups were developed in order to enable protein secondary structure analysis in LC effluents.<sup>15</sup> Here, the eluent is evaporated, while the analyte is (almost) simultaneously deposited onto a suitable substrate prior to FTIR spectra acquisition. Typical challenges for solvent-evaporation interfaces are, however, the morphology of certain analytes that can change over time and possible spatial inhomogeneity.<sup>9</sup> Moreover, this destructive type of sample preparation prevents fractionation of the effluent after detection.

Quantum cascade lasers (QCLs) are new light sources in the mid-IR spectral region that provide polarized and coherent light with spectral power densities higher by a factor of  $10^4$  or more compared with thermal light sources.<sup>16,17</sup> In combination with an external cavity (EC), QCLs enable tuning over several hundred wavenumbers, thus offering high potential for mid-IR transmission spectroscopy of liquids. It was shown that the high available spectral power of EC-QCLs opens a wide range of applications, including robust in-line detection of LC effluents.<sup>18</sup> For protein analysis, the higher spectral power densities enabled to increase the optical path lengths for transmission measurements by a factor of 3–4, thus considerably improving the ruggedness by significantly lowering the backpressure at the cell.<sup>19</sup> In this framework, academic setups were reported that applied EC-QCLs to investigate the amide I<sup>20,21</sup> and amide I + amide II band,<sup>22,23</sup> finally achieving a limit of detection almost 10 times lower than high-end FTIR spectroscopy at similar measurement conditions.<sup>24</sup> These setups were successfully combined with different chemometric methods in order to quantify individual protein content in complex mixtures<sup>25–28</sup> and to monitor changes in the protein secondary structure after denaturation.<sup>21,29,30</sup> Most recently, a commercial EC-QCL-based spectrometer (ChemDetect Analyzer, Daylight Solutions) was introduced, offering robust and sensitive spectra acquisition across the wavenumber region between 1350 and 1770  $\text{cm}^{-1}$ .<sup>31</sup>

Another challenge in in-line LC-IR is the compensation of possible eluent absorbance bands, which can be higher by several orders of magnitude compared to actual analyte bands. Even though direct subtraction of a background spectrum is possible under isocratic conditions, background correction during gradient elution is more challenging. For this purpose, a method based on a “reference spectra matrix” (RSM) was introduced.<sup>32</sup> In this approach, the spectra acquired in-line during the LC run are viewed as “sample matrix” (SM), whereas the RSM is recorded during a blank run or the re-equilibration phase of the system. Based on this acquired information, spectral regions that show absorbance bands characteristic for eluent composition are identified, which are located in different spectral regions when compared to the analyte spectrum. Consequently, each spectrum in the SM is corrected with the RSM spectrum that has the closest eluent composition. This method was successfully applied for gradient compensation in a wide range of different reversed-phase

HPLC-FTIR applications, including analysis of carbohydrates,<sup>33,34</sup> nitrophenols,<sup>35,36</sup> and pesticides.<sup>32,37</sup>

In the present work, the ChemDetect Analyzer was coupled to a lab-scale prep-LC system. The large optical path length of the transmission cell of 25  $\mu\text{m}$  enabled in-line monitoring of proteins without solvent evaporation steps. This LC–QCL–IR coupling was employed to analyze systems based on ion-exchange chromatography (IEX) and different NaCl gradient profiles. For this application, the laser-based spectrometer offers advantages regarding acquisition of protein spectra in aqueous solution. However, due to the limited spectral region of QCL-IR spectra, gradient compensation is more challenging than with FTIR spectra, which cover the entire mid-IR region. In the studied analytical problems, absorption bands from the salt gradient highly overlap with protein absorptions. Three different case studies were performed, covering various real-life conditions used in chromatographic protein downstream processing. In Case I, a reference blank run was directly subtracted from a sample run with the same linear gradient. For Case II and Case III, two different gradient profiles were employed when compared to the one of the reference blank run. Here, a modified RSM-based approach was devised by incorporating the signal of the conductivity detector. With this approach, each spectrum in the SM was corrected with the RSM spectrum that had the closest conductivity value. Thus, a singular measurement of the RSM blank run was sufficient to correct sample runs with different salt gradient profiles, revealing high quality protein spectra. The obtained results demonstrate the successful coupling of a laser-based IR spectrometer with a LC system and present a novel approach for LC-IR-based gradient correction of highly overlapping eluent and analyte mid-IR spectra, indicating high flexibility for future in-line monitoring of the protein secondary structure in chromatographic effluents.

## EXPERIMENTAL SECTION

**Reagents and Samples.** Tris, hydrochloric acid (HCl), and NaCl used for eluents of the prep-LC were purchased from Carl Roth (Karlsruhe, Germany). Hemoglobin (Hemo) from bovine blood and  $\beta$ -lactoglobulin ( $\beta$ -LG) from bovine milk ( $\geq 90\%$ ) were purchased from Sigma-Aldrich (Steinheim, Germany). For LC-IR and protein secondary structure reference measurements, proper amounts of lyophilized protein powder were dissolved in the corresponding buffer. Ultrapure water (MQ) was from a Milli-Q system from Merck Millipore (Darmstadt, Germany). HPLC-Grade acetonitrile (ACN) and trifluoroacetic acid (TFA) were both purchased from AppliChem (Darmstadt, Germany).

**LC–QCL–IR Setup.** The flow path of LC-IR measurements is illustrated in Figure 1. An ÄKTA pure system (Cytiva Life Sciences, MA, USA) equipped with an U9-M UV monitor, a C9 conductivity monitor, and a F9-C fraction collector was used for the prep-LC runs. All runs were performed with a 1 mL HiTrap Capto Q column (Cytiva Life Sciences, MA, USA). Laser-based mid-IR spectra were acquired with a ChemDetect Analyzer (Daylight Solutions Inc., San Diego, USA). The delay volume between the conductivity detector and ChemDetect Analyzer was determined by injection of 1 M NaCl solution.

**Preparative Chromatography Conditions.** For all four prep-LC runs, the setup described in Figure 1 was used. The flowrate was kept constant at 75  $\text{cm}^3/\text{h}$ , with Buffer A (50 mM Tris/HCl, pH 8.5) being used for equilibration and wash and a



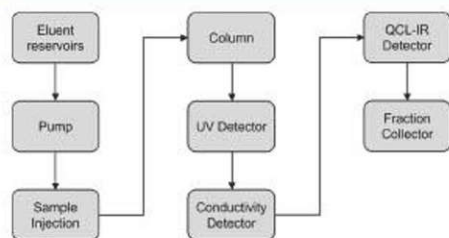


Figure 1. Schematic of the flow path in the LC-QCL-IR setup.

gradient of Buffer A and Buffer B (50 mM Tris/HCl, pH 8.5, 1 M NaCl) being used for elution. Injections were performed using a 1 mL loop, and absorbance at 280 nm as well as conductivity were recorded for all runs. Fractions of 1 mL were collected over the whole run-time and the respective protein concentrations of each fraction measured using the described reversed phase (RP)-HPLC method. The specific elution profiles are shown in Figure S1 and described in detail below.

**Blank run:** a volume of 1 mL Buffer A was injected. A linear gradient elution was performed over 60 min (30 CVs) from 0% Buffer B to 100% Buffer B.

**Case I:** as load, 1 mL of 10 mg/mL Hemo and 10 mg/mL  $\beta$ -LG in Buffer A was injected. A linear gradient elution identical to the blank run was performed (0–100% Buffer B in 60 min).

**Case II:** as load, 1 mL of 5 mg/mL Hemo and 5 mg/mL  $\beta$ -LG in Buffer A was injected. A linear gradient elution was performed over 30 min (15 CVs) from 0% Buffer B to 100% Buffer B.

**Case III:** as load, 1 mL of 5 mg/mL Hemo and 5 mg/mL  $\beta$ -LG in Buffer A was injected. A step gradient elution was performed with 25% Buffer B 6 min after injection, 50% Buffer B 20.5 min after injection, and 100% Buffer B 30.5 min after injection.

**EC-QCL-Based Mid-IR Measurements.** The ChemDetect Analyzer, equipped with an EC-QCL (tunable between 1350 and 1750  $\text{cm}^{-1}$ ) and a diamond transmission flow cell (25  $\mu\text{m}$  optical path length), was used for acquisition of laser-based mid-IR spectra. An external water-cooling unit was operated at 17  $^{\circ}\text{C}$  in order to ensure thermal stabilization of the laser head during operation. Spectra acquisition was performed with the ChemDetect Software package. For flow-through LC-QCL-IR measurements, first, a reference background spectrum was recorded by averaging 121 scans (60 s), followed by continuous acquisition of 20 averaged scans, resulting in a measurement time of 10 s per spectrum. Protein secondary structure off-line reference measurements were acquired by averaging 91 scans during a measurement time of 45 s.

**HPLC Reference Measurements.** As an off-line analytical method for the collected fractions, a previously published RP-HPLC method was applied.<sup>38</sup> In short, a BioResolve RP mAb Polyphenyl column (Waters, MA, USA) was used with MQ as mobile phase A and ACN as mobile phase B, both supplemented with 0.1% v/v TFA. Column temperature was kept constant at 70  $^{\circ}\text{C}$  and an injection volume of 2  $\mu\text{L}$  was used for all samples. Total run time for one injection was 18 min with a flow of 0.4 mL/min and UV/vis absorbance at 214, 280, and 404 nm was recorded for the whole run. A linear gradient from 25% B to 75% B (10 min) was performed, followed by 2 min with 75% B, after which the column was re-equilibrated with 25% B for 6 min.

**Quantification of IR and UV Signals.** Protein concentrations ( $c$ ) across the chromatographic run were calculated from mid-IR and UV signals, according to the Beer–Lambert law

$$c = \frac{A}{\epsilon d} \quad (1)$$

where  $A$  denotes the recorded absorbance values,  $d$  is the transmission path, and  $\epsilon$  indicates the molar decadic absorption coefficient.

For laser-based IR spectroscopy, values for  $A$  were obtained by integrating the spectra across the amide II region (1500–1600  $\text{cm}^{-1}$ ). The absorption coefficients ( $\epsilon$ ) for the two proteins were received by integrating the same area region from off-line recorded reference spectra with known protein concentrations.

For quantification of UV/vis spectra, absorption coefficients of the proteins were obtained by performing reference measurements using a Cary 50 Bio UV/vis spectrometer (Agilent Technologies, Santa Clara, California, USA). Spectra were recorded using the Cary WinUV software. Cuvettes with a path length of 10 mm were used to measure 0.5 mg/mL protein solutions. The thereby obtained absorption coefficients at 280 nm agreed well with those calculated via ExPASy ProtParam tool<sup>39</sup> and were used to calculate protein concentrations.

**Data Processing.** Data processing and analysis was conducted with in-house codes developed in Matlab R2020b (MathWorks, Inc., Natick, MA, 2020). During the first preprocessing step, absorbance bands of water vapor from the atmosphere were automatically corrected by subtraction of a water vapour reference spectrum, if necessary. Gradient correction was performed by direct blank run subtraction (Case I) and a modified procedure based on RSM (Case II and Case III).<sup>32</sup> Finally, absorption spectra were smoothed with a second order Savitzky–Golay filter (window = 15 points). The spectral resolutions of 1.2 and 3.6  $\text{cm}^{-1}$  were determined for smoothed and unsmoothed spectra by comparing the bandwidth of water vapor spectra to FTIR reference spectra.

## RESULTS AND DISCUSSION

**Relation between Salt Gradient, Conductivity, and Mid-IR Spectra.** In-line monitoring of prep-LC effluents was performed with the ChemDetect Analyzer as well as with conventional UV/vis and conductivity detectors (Figure 1). In addition to the routinely recorded signals, laser-based mid-IR spectroscopy offered robust spectra acquisition in the most important wavenumber range for protein secondary structure analysis.

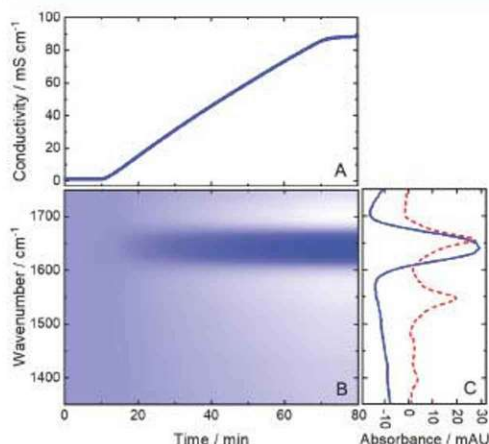
In the present LC-QCL-IR application, IEX was employed as the separation mechanism. Here, the pI of the analyte and the pH of the mobile phase are decisive for the degree of retention. Bound analytes are commonly eluted by utilizing a gradient of increasing salt concentration. Conductivity detection is a widespread method for measuring the salt concentration in IEX.<sup>40</sup>

Figure 2 displays the signals of the mid-IR and conductivity detector on the example of an IEX blank run with a linear NaCl gradient, and thus, the corresponding signal (A) linearly increases with the NaCl concentration.

The blue line in Figure 2C shows the distinctive mid-IR spectrum of NaCl (blue line) with the NaCl-free starting buffer as the reference. Increasing absorbance can be observed

C

<https://doi.org/10.1021/acs.analchem.1c05191>  
Anal. Chem. XXXX, XXX, XXX–XXX



**Figure 2.** Detector signals from an ion-exchange preparative-liquid chromatography blank run with a NaCl gradient (0–1 M between 10 and 70 min). (A) Conductivity detector response. (B) Heatmap of the corresponding laser-based mid-IR spectra. White areas indicate negative absorbance and dark blue areas positive absorbance. (C) Absorbance spectra of 0.3 M NaCl (retention time: 31.3 min, blue line) and 5 mg/mL hemoglobin (red, dashed line).

between approximately 1610 and 1680  $\text{cm}^{-1}$ , while the absorbance in the remaining spectrum decreases with higher salt concentrations. Even though  $\text{Na}^+$  and  $\text{Cl}^-$  are not IR active, reorganization of the water molecules due to the presence of these ions is responsible for a change in the mid-IR spectrum compared to the NaCl-free initial buffer.<sup>41</sup> The maximum of the HOH bending band of water (at approximately 1643  $\text{cm}^{-1}$ ) is not shifted in position; however, its intensity increases, accompanied by narrowing of the band with increasing NaCl concentrations.<sup>42</sup> The negative absorbance in the other spectral regions can be related to partial displacement of  $\text{H}_2\text{O}$  molecules with salt.

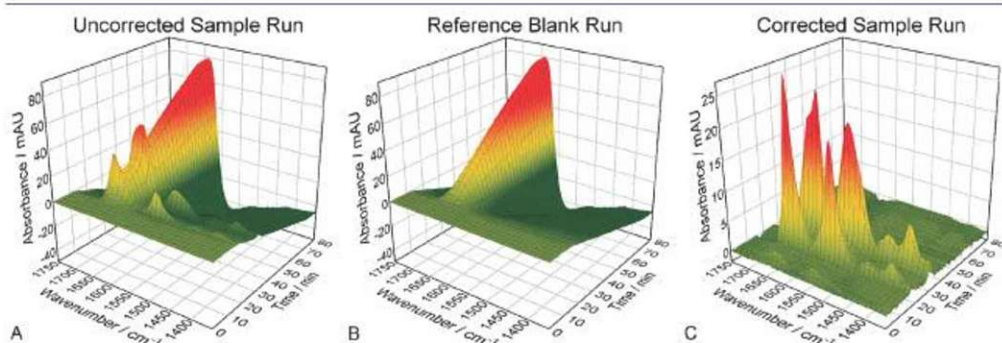
Figure 2B shows a 2D heat map depicting the response of the NaCl gradient recorded by the ChemDetect analyzer.

White areas indicate negative absorbance, whereas dark blue areas highlight positive absorbance.

The red dashed line in Figure 2C shows the IR spectrum of Hemo, illustrating the spectral overlap of protein bands with the one of NaCl (blue line). Particularly, the amide I band, representing the most important mid-IR region for protein secondary structure determination, is impeded by the modified HOH bending band of water. The described overlap of analyte and changing solvent band indicate significant challenges when employing the ChemDetect Analyzer for in-line monitoring of IEX. In order to obtain feasible mid-IR protein spectra, the influence of the changing salt concentration has to be eliminated. Consequently, feasible gradient compensation strategies are discussed and demonstrated based on model protein systems in the following subchapters.

**Case I: Direct Blank Run Subtraction to Enable In-Line Monitoring of Proteins.** In this approach, for background compensation, an analyte-free reference blank run was performed with the same gradient as the sample run. Based on the retention time, the reference spectra were subsequently subtracted from the sample spectra.

In the sample run, Hemo and  $\beta$ -LG were included, due to their difference in the pI and secondary structure. Figure 3 shows the spectral 3D plots of the (A) uncorrected sample run and the (B) reference blank run, as well as the (C) corrected sample run after direct background subtraction. The mid-IR spectra of the blank run were already discussed above. Even though the amide I and amide II bands are visible in the uncorrected sample run, the effect of the NaCl gradient on the absorbance spectra is clearly dominating. Thus, no reliable information about the protein secondary structure can be obtained. In contrast, the corrected sample run shows a stable baseline and distinct protein spectra. Figure 4 depicts the absorbance spectra extracted from the peak maxima of the differently corrected sample runs as well as off-line recorded reference IR spectra of the investigated proteins. For Case I (blue lines), the first chromatographic peak with a maximum at approximately 17 min can be related to Hemo due to its higher pI of 7.1.<sup>43</sup> Hemo mainly contains  $\alpha$ -helical secondary structures and shows the characteristic narrow amide I band with a maximum at approximately 1656  $\text{cm}^{-1}$  and a narrow amide II band at 1545  $\text{cm}^{-1}$ .<sup>7,44,45</sup>  $\beta$ -LG is predominantly composed of  $\beta$ -sheet secondary structures<sup>46</sup> and has a pI of

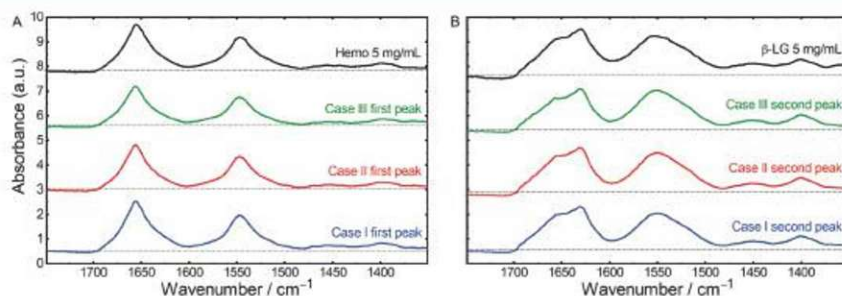


**Figure 3.** Spectral 3D plots for Case I: (A) uncorrected sample run, (B) reference blank run, and (C) corrected sample run after direct background subtraction.

D

<https://doi.org/10.1021/acs.analchem.1c05191>  
Anal. Chem. XXXX, XXX, XXX–XXX





**Figure 4.** Laser-based mid-IR absorbance spectra, extracted from the (A) first and (B) second peak maxima of the corrected chromatographic runs from Case I (blue lines), Case II (red lines), Case III (green lines), and off-line recorded reference IR spectra (black lines) of 5 mg/mL hemoglobin and  $\beta$ -lactoglobulin. The individual spectra were offset for better visibility. The pointed black lines indicate zero absorbance for every spectrum.

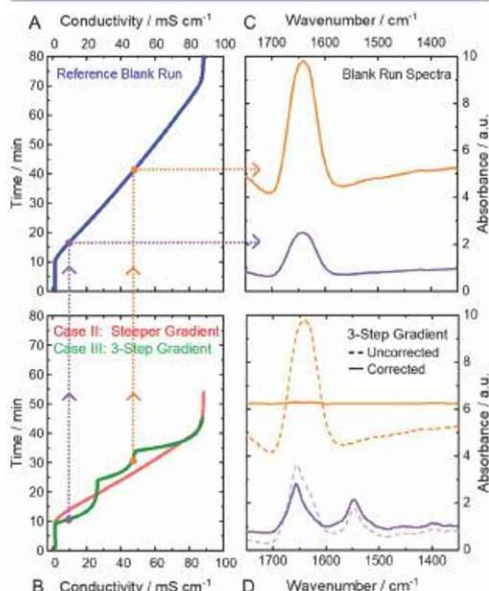
5.1<sup>47</sup> and thus elutes second under these IEX conditions. The corresponding absorbance spectra show broader amide I and II bands with  $\beta$ -sheet typical maxima at 1632 and 1550  $\text{cm}^{-1}$  and a shoulder at 1680  $\text{cm}^{-1}$ .<sup>22,48,49</sup> The obtained high-quality protein spectra thus indicate excellent long-term stability of the ChemDetect Analyzer.

A drawback of this direct gradient compensation approach is the need for a blank run that is performed identical to the sample run. Consequently, any adaptations in the gradient profile require acquisition of an additional blank run and very high reproducibility of the chromatographic system is required in general. The additional time consumption of the blank run and the re-equilibration phase of the column make this approach methodically rigid and impractical to realize for bench-scale as well as industrial applications. Hence, a more flexible strategy is presented in the next subchapter.

**Case II and III: Flexible Gradient Compensation with Adapted RSM.** Even though direct blank run subtraction revealed excellent protein spectra, more flexible and less laborious gradient compensation strategies are favorable. Gradient compensation based on RSM requires only a single blank run, which can be used for correcting various gradient profiles.<sup>32</sup> This approach was successfully applied for HPLC-FTIR, where the entire mid-IR range and eluent specific absorbance bands were available. These specific bands were used as mobile phase identification parameter (IP) in order to select the blank spectrum to be subtracted. This approach, however, is not feasible for monitoring proteins in IEX due to lack of specific absorbance bands and limited sensitivity of FTIR spectroscopy. Due to significantly higher spectral power densities of the EC-QCL, the ChemDetect Analyzer overcomes the typical limitations of FTIR spectroscopy and offers robust and highly sensitive flow-through measurements in the most important wavenumber range for protein analysis.

In the present study, a modified RSM method based on an external IP is proposed and successfully applied. For this purpose, the signal from the conductivity detector was related to the mid-IR spectra. The measured conductivity represents an adequate IP, due to its high dependency on the NaCl concentration, whereas the influence of the protein concentration on the conductivity is negligible (Figure S2). For this purpose, the blank run with the linear NaCl gradient (same as in Case I) was taken as the reference. In order to evaluate the applicability of this approach, two chromatographic sample runs with different gradients compared to the one from the

reference run were conducted: Case II included a linear gradient with a steeper profile, while Case III featured a 3-step gradient. A specific conductivity value was assigned to each mid-IR spectrum in the SM and RSM, based on the retention time. Figure 5 depicts the principle of the applied compensation method on the example of a baseline and protein spectrum in Case III. Based on the comparison of the measured conductivity in the (C) sample run and (A) the



**Figure 5.** Principle of the applied RSM-based background compensation method. (A) Conductivity detector signals of the reference blank run and (B) sample runs of Case II and Case III. Dotted lines indicate the relation between selected conductivity values and the corresponding mid-IR spectra of background (orange) and protein (purple) of Case III. The absorbance spectra of the (C) blank run were subtracted from (D) the 3-step gradient sample run spectra, based on the most similar conductivity value, revealing distinctive protein spectra (purple) and a flat baseline (orange).

E

<https://doi.org/10.1021/acs.analchem.1c05191>  
Anal. Chem. XXXX, XXX, XXX–XXX

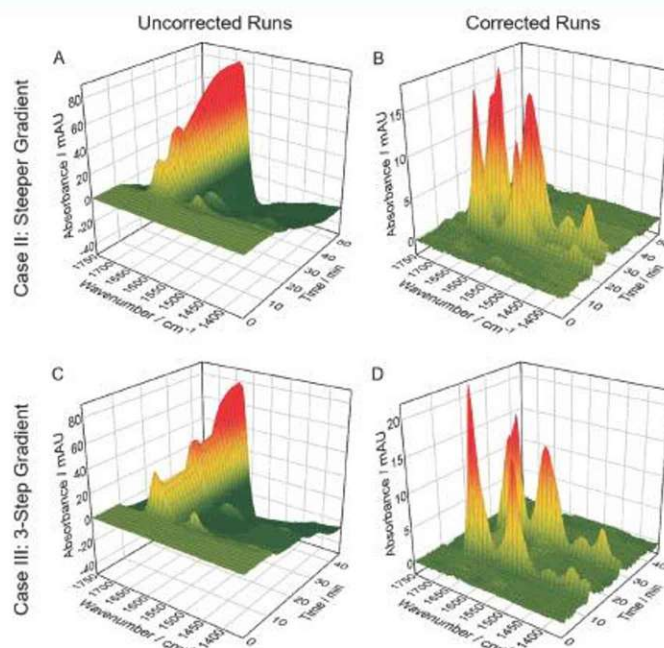


Figure 6. Spectral 3D plots for Case II and Case III: (A,C) uncorrected sample runs and (B,D) corrected sample runs after background correction with adapted RSM.

reference blank run, the corresponding (B) blank run spectra are identified. Finally, the (D) sample spectra are corrected by subtracting the identified blank run spectra with the most similar conductivity value.

Figure 6 displays the spectral 3D plots of the uncorrected (A,C) and corrected (B,D) sample runs. The uncorrected run from Case II constitutes a similar profile as the reference blank run with a shorter measurement time, whereas the run from Case III shows a clearly different profile due to the 3-step gradient. Both corrected runs comprise stable baselines and excellent protein spectra, comparable to those from direct blank run subtraction (Figure 4). Consequently, acquisition of a singular blank run is sufficient for correcting sample runs of widely different gradient profiles. Due to high flexibility and little additional time consumption, the applied gradient compensation approach shows high potential for industrial application.

**Comparison of Mid-IR In-Line Measurements to Conventional Off-Line Analysis.** In order to verify the signal of the ChemDetect Analyzer, protein concentrations calculated from mid-IR absorbance were compared to those calculated from the UV detector signal. Figure 7 displays the concentrations, determined from the ChemDetect Analyzer (red line) and UV detector (black line) over the chromatographic run from Case III. In mid-IR spectroscopy, the total protein content is best represented by the amide II band because it is less influenced by water absorption than the amide I band. Thus, the wavenumber region between 1500 and 1600  $\text{cm}^{-1}$  was integrated to obtain the absorbance values. For comparison, protein concentrations were also calculated from

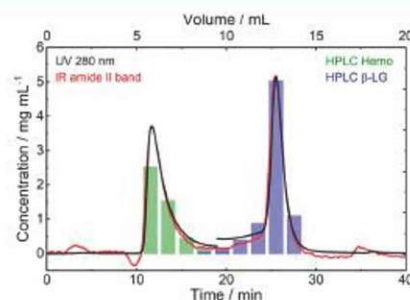


Figure 7. Comparison of protein concentrations obtained from mid-IR amide II band (red line) and UV detector signal at 280 nm (black line) across the chromatographic run from Case III and protein reference concentrations obtained by measuring the collected fractions with reversed-phase HPLC (green and blue bars).

the UV signals at 280 nm recorded by the UV detector of the LC instrument. UV absorption at this wavelength detects aromatic protein residues and disulfide bonds and is most commonly used to monitor proteins in chromatographic applications.<sup>1</sup> Absorption coefficients for Hemo and  $\beta$ -LG for IR and UV spectroscopy were obtained from reference measurements with known protein concentrations. Because these values are different for the two proteins, for conversion of the recorded signals to concentration values, the chromatogram was split into two parts and the absorption coefficients of

F

<https://doi.org/10.1021/acs.analchem.1c05191>  
Anal. Chem. XXXX, XXX, XXX–XXX



Hemo were used from 0 to 19 min, whereas the coefficients of  $\beta$ -LG were applied between 19 and 40 min. Figure 7 depicts the highly overlapping protein concentrations derived from the signals of ChemDetect and UV detector in the time axis as well as for the protein concentration, demonstrating that equivalent quantitative signals can be obtained by the two methods.

For protein identification, IR spectra from the two chromatographic peak maxima were extracted and compared to pure Hemo and  $\beta$ -LG reference spectra, respectively. Figure 4 reveals excellent agreement between in-line and off-line ChemDetect measurements. Absorbance band positions as well as band shapes show excellent comparability between mid-IR spectra of the chromatographic peak maxima and those of the pure reference solutions. In contrast, with conventional LC instrumentation employing an UV detector, real-time information about the protein secondary structure, and consequently type of protein cannot be obtained. In practical applications where the protein identity or purity are usually unknown, highly biased results would be obtained from direct quantitation of the UV signal because absorption coefficients at 280 nm can be different by more than an order of magnitude between proteins.<sup>50</sup> Then, protein identification must be performed by off-line RP-HPLC measurements of the collected fractions. In this regard, the green bars shown in Figure 7 indicate the concentration of the protein that was identified as Hemo by RP-HPLC, and the blue bars show the concentration of  $\beta$ -LG. Consequently, the obtained LC-QCL-IR chromatograms agree well with the conventionally applied quantification methods, while offering the additional advantage of providing real-time information about the protein secondary structure.

## CONCLUSIONS AND OUTLOOK

In this work, a laser-based mid-IR spectrometer was successfully applied for in-line monitoring of proteins from preparative IEX. A major challenge was the highly overlapping absorbance bands of proteins and NaCl gradient that dominated the recorded sample run spectra. Advanced background compensation strategies based on adapted RSM were developed and their implementation in three different case studies resulted in high-quality protein spectra. In Case I, a reference blank run was directly subtracted from a sample run with the same gradient profile. The obtained protein spectra indicated excellent long-term stability of the ChemDetect Analyzer. Case II and III included sample runs with a steeper linear gradient and a 3-step gradient, respectively. Here, a novel gradient compensation approach based on RSM and conductivity IP was introduced. With this method, a single blank run was sufficient for compensating the NaCl gradient in sample runs with distinctively different profiles. The thereby obtained protein spectra showed excellent comparability to off-line reference measurements.

It was shown that the protein concentrations evaluated from signals obtained by the ChemDetect Analyzer are equivalent to UV spectroscopy at 280 nm, which is the standard quantification method for proteins in chromatographic systems. Furthermore, compared to conventional LC detectors, the laser-based mid-IR spectrometer offers the major advantage of providing real-time information about the protein secondary structure, comparable to high-end off-line measurements. The ChemDetect Analyzer thus holds high potential for complementing laborious and time-consuming off-line methods and provides an easily accessible in-line method. In combination with the presented adapted RSM method to

compensate for varying gradient profiles, QCL-IR presents a powerful tool for in-process monitoring and control. Especially, in the light of QbD principles, a near real-time PAT tool able to give information about protein secondary structures and corresponding CQAs presents high potential. In the future, LC-QCL-IR coupling can be employed for chemometrics-based analysis of possible impurities and individual quantification of co-eluting proteins.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c05191>.

Elution profiles of the recorded chromatographic runs, conductivity detector signal of reference blank run and sample run with the same NaCl gradient, and IR spectra and calibration lines for hemoglobin and  $\beta$ -lactoglobulin (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Andreas Schwaighofer – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; [orcid.org/0000-0003-2714-7056](https://orcid.org/0000-0003-2714-7056);  
Email: [andreas.schwaighofer@tuwien.ac.at](mailto:andreas.schwaighofer@tuwien.ac.at)

Bernhard Lendl – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; [orcid.org/0000-0003-3838-5842](https://orcid.org/0000-0003-3838-5842);  
Email: [bernhard.lendl@tuwien.ac.at](mailto:bernhard.lendl@tuwien.ac.at)

### Authors

Christopher K. Akhgar – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; [orcid.org/0000-0001-8266-043X](https://orcid.org/0000-0001-8266-043X)

Julian Ebner – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Oliver Spadiut – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Complete contact information is available at:  
<https://pubs.acs.org/doi/10.1021/acs.analchem.1c05191>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. C.K.A. and J.E. contributed equally.

### Funding

Open Access is funded by the Austrian Science Fund (FWF).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work has received funding from the COMET Center CHASE (project no. 868615), funded within the COMET—Competence Centers for Excellent Technologies programme by the BMK, the BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET programme is managed by the Austrian Research Promotion Agency (FFG). Additional funding was provided by the European Union's Horizon 2020 research and innovation program through NutriShield project under grant agreement no. 818110. This research was further

G

<https://doi.org/10.1021/acs.analchem.1c05191>  
Anal. Chem. XXXX, XXX, XXX–XXX

funded by the Austrian Research Promotion Agency (FFG) (project no. 874206) and by the Austrian Science Fund FWF (project no. P32644-N).

## REFERENCES

- (1) Carta, G.; Jungbauer, A. *Protein Chromatography. Process Development and Scale-Up*; John Wiley & Sons: Hoboken, USA, 2010; Vol. 5.
- (2) Gupta, M. *Methods for Affinity-Based Separations of Enzymes and Proteins*; Birkhäuser: Basel, Switzerland, 2002.
- (3) Ali, J.; Pramod, K.; Tahir, M. A.; Charoo, N. A.; Ansari, S. H. *Int. J. Pharm. Invest.* **2016**, *6*, 129–138.
- (4) Esmonde-White, K. A.; Cuellar, M.; Uerpmann, C.; Lenain, B.; Lewis, I. R. *Anal. Bioanal. Chem.* **2017**, *409*, 637–649.
- (5) Ebner, J.; Humer, D.; Klausner, R.; Rubus, V.; Pell, R.; Spadiut, O.; Kopp, J. *Bioengineering* **2021**, *8*, 78.
- (6) Singh, B. R. Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins. *Infrared Analysis of Peptides and Proteins*; American Chemical Society: Washington D.C., USA, 1999; pp 2–37.
- (7) Barth, A. *Biochim. Biophys. Acta Bioenerg.* **2007**, *1767*, 1073–1101.
- (8) Murphy, B.; D'Antonio, J.; Manning, M.; Al-Azzam, W. *Curr. Pharm. Biotechnol.* **2014**, *15*, 880–889.
- (9) Kuligowski, J.; Quintás, G.; Guardia, M.; Lendl, B. *Liquid Chromatography—Liquid Chromatography—Fourier Transform Infrared. Encyclopedia of Analytical Science*, 3rd ed.; Elsevier: Amsterdam, Netherlands, 2019; pp 75–85.
- (10) Kuligowski, J.; Quintás, G.; Garrigues, S.; Lendl, B.; de la Guardia, M.; Lendl, B. *Trends Anal. Chem.* **2010**, *29*, 544–552.
- (11) Fabian, H.; Mantele, W. *Infrared Spectroscopy of Proteins. Handbook of Vibrational Spectroscopy*; John Wiley & Sons: Hoboken, USA, 2006.
- (12) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. *Nat. Protoc.* **2015**, *10*, 382–396.
- (13) Großhans, S.; Rüdt, M.; Sanden, A.; Brestrich, N.; Morgenstern, J.; Heissler, S.; Hubbuch, J. *J. Chromatogr. A* **2018**, *1547*, 37–44.
- (14) Sanden, A.; Suhm, S.; Rüdt, M.; Hubbuch, J. *J. Chromatogr. A* **2019**, *1608*, 460410.
- (15) Turula, V. E.; de Haseth, J. A. *Anal. Chem.* **1996**, *68*, 629–638.
- (16) Faist, J.; Capasso, F.; Sivco, D. L.; Sirtori, C.; Hutchinson, A. L.; Cho, A. Y. *Science* **1994**, *264*, 553–556.
- (17) Schwaighofer, A.; Brandstetter, M.; Lendl, B. *Chem. Soc. Rev.* **2017**, *46*, 5903–5924.
- (18) Beskers, T. F.; Brandstetter, M.; Kuligowski, J.; Quintás, G.; Wilhelm, M.; Lendl, B. *Analyst* **2014**, *139*, 2057–2064.
- (19) Schwaighofer, A.; Lendl, B.; Ozaki, Y.; Baranska, M.; Lednev, I. Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In *Vibrational Spectroscopy in Protein Research*; Wood, B., Ed.; Academic Press: Cambridge, USA, 2020; pp 59–88.
- (20) Alcaráz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. *Anal. Chem.* **2015**, *87*, 6980–6987.
- (21) Schwaighofer, A.; Alcaraz, M. R.; Araman, C.; Goicoechea, H.; Lendl, B. *Sci. Rep.* **2016**, *6*, 33556.
- (22) Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. *Anal. Chem.* **2018**, *90*, 7072–7079.
- (23) Chon, B.; Xu, S.; Lee, Y. J. *Anal. Chem.* **2021**, *93*, 2215–2225.
- (24) Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. *Anal. Chem.* **2020**, *92*, 9901–9907.
- (25) Kuligowski, J.; Schwaighofer, A.; Alcaráz, M. R.; Quintás, G.; Mayer, H.; Vento, M.; Lendl, B. *Anal. Chim. Acta* **2017**, *963*, 99–105.
- (26) Schwaighofer, A.; Kuligowski, J.; Quintás, G.; Mayer, H. K.; Lendl, B. *Food Chem.* **2018**, *252*, 22–27.
- (27) Schwaighofer, A.; Alcaráz, M. R.; Kuligowski, J.; Lendl, B. *Biomol. Spectrosc. Imag.* **2018**, *7*, 35–45.
- (28) Montemurro, M.; Schwaighofer, A.; Schmidt, A.; Culzoni, M. J.; Mayer, H. K.; Lendl, B. *Analyst* **2019**, *144*, 5571–5579.
- (29) Alcaráz, M. R.; Schwaighofer, A.; Goicoechea, H.; Lendl, B. *Anal. Bioanal. Chem.* **2016**, *408*, 3933–3941.
- (30) Schwaighofer, A.; Alcaraz, M. R.; Lux, L.; Lendl, B. *Spectrochim. Acta, Part A* **2020**, *226*, 117636.
- (31) Schwaighofer, A.; Akhgar, C. K.; Lendl, B. *Spectrochim. Acta, Part A* **2021**, *253*, 119563.
- (32) Quintás, G.; Lendl, B.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2008**, *1190*, 102–109.
- (33) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Anal. Chim. Acta* **2008**, *624*, 278–285.
- (34) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2009**, *1216*, 3122–3130.
- (35) Quintás, G.; Kuligowski, J.; Lendl, B. *Anal. Chem.* **2009**, *81*, 3746–3753.
- (36) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Talanta* **2010**, *80*, 1771–1776.
- (37) Quintás, G.; Kuligowski, J.; Lendl, B. *Appl. Spectrosc.* **2009**, *63*, 1363–1369.
- (38) Kopp, J.; Zauner, F. B.; Pell, A.; Hausjell, J.; Humer, D.; Ebner, J.; Herwig, C.; Spadiut, O.; Slouka, C.; Pell, R. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113412.
- (39) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571–607.
- (40) Haddad, P. R. Conductivity Detection. *Journal of Chromatography Library*, Elsevier, 1990, Chapter 9; pp 245–289.
- (41) Max, J.-J.; Chapados, C. *J. Chem. Phys.* **2001**, *115*, 2664–2675.
- (42) Max, J.-J.; Trudel, M.; Chapados, C. *Appl. Spectrosc.* **1998**, *52*, 234–239.
- (43) Zhan, G.; Li, C.; Luo, D. *Bull. Korean Chem. Soc.* **2007**, *28*, 1720–1724.
- (44) Levitt, M.; Greer, J. *J. Mol. Biol.* **1977**, *114*, 181–239.
- (45) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. *Nature* **1960**, *185*, 416–422.
- (46) Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. *J. Mol. Biol.* **1987**, *197*, 695–706.
- (47) Engelhardt, K.; Lexis, M.; Gochev, G.; Konnerth, C.; Miller, R.; Willenbacher, N.; Peukert, W.; Braunschweig, B. *Langmuir* **2013**, *29*, 11646–11655.
- (48) van de Weert, M.; Haris, P. L.; Hennink, W. E.; Crommelin, D. J. A. *Anal. Biochem.* **2001**, *297*, 160–169.
- (49) Dousseau, F.; Pezolet, M. *Biochemistry* **1990**, *29*, 8771–8779.
- (50) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. *Protein Sci.* **1995**, *4*, 2411–2423.

H

<https://doi.org/10.1021/acs.analchem.1c05191>  
Anal. Chem. XXXX, XXX, XXX–XXX



# Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography

Christopher K. Akhgar, Julian Ebner, Mirta R. Alcaraz, Julian Kopp, Héctor Goicoechea, Oliver Spadiut, Andreas Schwaighofer,\* and Bernhard Lendl\*

Cite This: *Anal. Chem.* 2022, 94, 11192–11200

Read Online

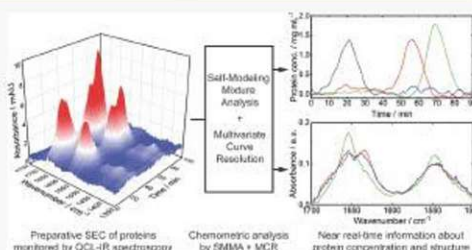
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** An external-cavity quantum cascade laser (EC-QCL)-based flow-through mid-infrared (IR) spectrometer was placed in line with a preparative size exclusion chromatography system to demonstrate real-time analysis of protein elutions with strongly overlapping chromatographic peaks. Two different case studies involving three and four model proteins were performed under typical lab-scale purification conditions. The large optical path length (25  $\mu\text{m}$ ), high signal-to-noise ratios, and wide spectral coverage (1350 to 1750  $\text{cm}^{-1}$ ) of the QCL-IR spectrometer allow for robust spectra acquisition across both the amide I and II bands. Chemometric analysis by self-modeling mixture analysis and multivariate curve resolution enabled accurate quantitation and structural fingerprinting across the protein elution transient. The acquired concentration profiles were found to be in excellent agreement with the off-line high-performance liquid chromatography reference analytics performed on the collected effluent fractions. These results demonstrate that QCL-IR detectors can be used effectively for in-line, real-time analysis of protein elutions, providing critical quality attribute data that are typically only accessible through time-consuming and resource-intensive off-line methods.



Protein purification and polishing protocols typically include diverse process unit operations based on liquid chromatography (LC).<sup>1</sup> This technique separates analytes in a liquid mobile phase by interactions with a solid stationary phase according to different physico-chemical properties. In size exclusion chromatography (SEC), compounds are separated by their size and shape which offers several advantages, such as straight-forward operation, nondenaturing conditions, and isocratic elution, compared to other separation principles.<sup>2</sup> Protein concentrations in chromatographic effluents are routinely monitored in-line by univariate UV/vis or evaporative light scattering detectors, offering excellent sensitivity, high robustness, and a broad linear range. A major drawback of these detectors is, however, that the obtained signals do not provide information that allows for the discrimination or quantitation of different coeluting proteins. Critical quality attributes (CQAs), thus, have to be obtained by analyzing the collected fractions off-line. During process development, this can lead to significant time delays. Moreover, it hinders development based on quality by design (QbD) principles. QbD requires the application of process analytical technology (PAT) tools, facilitating in-process monitoring and in-process control. In-line or on-line measurements providing real-time or

near real-time information on CQAs are required to allow timely adaption of set-points during the purification step.

Mid-infrared (IR) spectroscopy is a well-established technique for nondestructive analysis of diverse compounds, including polypeptides and proteins.<sup>3</sup> Conventional Fourier-transform IR (FT-IR) spectrometers are equipped with thermal light sources that emit low-power radiation across the entire mid-IR region (400–4000  $\text{cm}^{-1}$ ). Even though LC-FT-IR hyphenation was successfully demonstrated for the analysis of numerous analytes including nitrophenols,<sup>4,5</sup> carbohydrates,<sup>6,7</sup> and pesticides,<sup>8,9</sup> mid-IR flow-through measurements of proteins remain challenging. The most important IR bands for protein secondary structure determination and quantitation are the amide I (1600–1700  $\text{cm}^{-1}$ ) and amide II (1500–1600  $\text{cm}^{-1}$ ) band, respectively.<sup>10</sup> Substantial light absorption by the HOH bending band of

Received: April 8, 2022  
Accepted: July 25, 2022  
Published: August 4, 2022



water at approximately  $1645\text{ cm}^{-1}$  makes investigations of the overlapping amide I band with FT-IR instrumentation a cumbersome task. In order to avoid total absorption of IR radiation in this spectral region, optical path lengths of 6 to  $8\text{ }\mu\text{m}$  are typically applied.<sup>11,12</sup> Such limited path lengths are not suitable for LC-IR hyphenation as they lead to distinctly impaired robustness and sensitivity. For this purpose, complex schemes were developed that evaporate the solvent and deposit the protein almost simultaneously onto a substrate before FT-IR analysis.<sup>13</sup> Even though these setups enabled protein secondary structure analysis from LC effluents,<sup>14,15</sup> solvent evaporation interfaces can bear major challenges such as spatial heterogeneity and changes in analyte morphology over time.<sup>16</sup> Moreover, in preparative LC runs, the effluent is usually fractionated after detection, making a preceding solvent evaporation step inapplicable. More recently, attenuated total reflection-FT-IR spectroscopy<sup>17</sup> was coupled to an LC system for in-line monitoring of proteins.<sup>18–20</sup> This configuration overcomes the limitations regarding ruggedness, but still requires high protein concentrations due to its limited sensitivity.

Significant progress in quantum cascade lasers (QCLs)<sup>21</sup> has challenged conventional FT-IR spectrometers for biochemical sensing applications.<sup>22</sup> Properties such as  $\geq 10^4$  times higher brightness compared to thermal light sources and tunability over several hundred wavenumbers in external cavity (EC) configurations make QCLs highly beneficial for the analysis of proteins.<sup>23</sup> In this context, diverse academic setups were developed that employed EC-QCLs for protein investigations. Here, it has been demonstrated that the intense power outputs of QCLs allow to significantly increase the path length for transmission measurements and, thus, the ruggedness for protein amide I band analysis in aqueous solutions.<sup>24</sup> Due to the particular characteristics of water absorption in the protein amide I spectral region and emission properties of EC-QCLs,<sup>25</sup> it has turned out to be a challenging task to develop setups also covering the protein amide II region. However, simultaneous analyses of amide I + II bands were realized by different approaches, for example, by combining EC-QCLs with either mercury cadmium telluride detectors (MCTs) and optical filters,<sup>25,26</sup> MCTs and acousto-optic modulators,<sup>27</sup> or quantum cascade detectors (QCDs).<sup>28</sup> Furthermore, the implementation of an advanced noise compensation strategy based on balanced detection led to robust protein measurements with limits of detection almost an order of magnitude lower than those from high-end FT-IR spectrometers.<sup>26</sup>

Parallel to the rapid advances of laser-based optical setups in academic research, a commercially available QCL-IR spectrometer, the ChemDetect Analyzer (Daylight Solutions), was recently introduced.<sup>29</sup> This device covers a broad wavenumber range beyond protein amide I and II bands and offers robust and sensitive spectra acquisition with an optical transmission path of  $25\text{ }\mu\text{m}$ . In a recent piece of work, the ChemDetect Analyzer was successfully applied for in-line monitoring of proteins from preparative LC.<sup>30</sup> Compared to conventionally used LC detectors, laser-based mid-IR spectroscopy offers the major advantage of providing near real-time information about protein quantity and secondary structure, which can otherwise merely be obtained by off-line measurements. LC-QCL-IR coupling, thus, bears a high potential for in-line analysis of CQAs, such as protein purity, which is further investigated in the present study.

For analysis of complex experimental data, chemometrics is typically applied to extract chemical information about individual analytes from spectroscopic data of multicomponent systems. Multivariate spectroscopic monitoring of dynamic processes, such as in LC-QCL-IR, generates two-way data matrices that comprise the information about the occurring spectral changes and the chemical perturbation profiles of the system. Multicomponent spectroscopic signals generally follow Beer–Lambert's Law and fulfill the concept of the so-called bilinear models.<sup>31</sup> Among the most used chemometric techniques based on bilinear decomposition are self-modeling mixture analysis (SMMA)<sup>32</sup> and multivariate curve resolution (MCR).<sup>33</sup> The advantage of these methods is that they do not require any a priori knowledge about the system, for example, the number or spectra of components, and all information can be deduced from the recorded data set. Even though the obtained pure variables do not represent a pure component, for systems with a reduced number of components, this approach can serve as a good and fast estimator of the chemical behavior of the system that can be readily compared to recorded spectra and, thus, allows straightforward interpretation by nonchemometricians.<sup>34,35</sup>

In this work, LC-QCL-IR hyphenation was performed for in-line monitoring of proteins from coeluting chromatographic peaks. Two case studies involving three and four proteins, respectively, were performed based on SEC, and real-life conditions used in protein purification protocols were applied. The goal of this work is to employ in-line QCL-IR spectroscopy for obtaining qualitative and quantitative information, which can be conventionally only received by work- and time-intensive off-line high-performance LC (HPLC) analytics. For this purpose, chemometric analysis based on a bilinear decomposition model was performed to (i) extract IR absorption spectra of the individual proteins from the recorded multidimensional data set as well as to (ii) retrieve their concentration profiles over the chromatographic run. Achieved results were benchmarked against reference off-line IR spectra of pure protein solutions and HPLC measurements of the collected fractions, showing excellent agreement in both cases.

## ■ EXPERIMENTAL SECTION

**Reagents and Samples.** Ovalbumin (Ova,  $\geq 90\%$ ),  $\alpha$ -chymotrypsinogen A ( $\alpha$ -CT) from bovine pancreas, myoglobin (Myo) from equine skeletal muscle ( $\geq 95\%$ ), horseradish peroxidase type VI-A (HRP), and  $\beta$ -lactoglobulin ( $\beta$ -LG) from bovine milk ( $\geq 90\%$ ) were obtained from Sigma-Aldrich (Steinheim, Germany). Appropriate amounts of protein powder were dissolved in SEC buffer. Ultrapure water (MQ) was obtained with a Milli-Q system from Merck Millipore (Darmstadt, Germany). Trifluoroacetic acid and acetonitrile, both HPLC-grade, were purchased from AppliChem (Darmstadt, Germany). All other chemicals used for the preparation of mobile phases were obtained from Carl Roth (Karlsruhe, Germany).

**LC-QCL-IR Flow Path.** The applied LC-QCL-IR setup is depicted in Figure 1. An ÄKTA pure preparative chromatographic system (Cytiva Life Sciences, MA, USA), equipped with a U9-M UV monitor and an F9-C fraction collector was used for all SEC runs. A HiLoad 16/600 Superdex 200 pg (Cytiva Life Sciences, MA, USA) was used as the SEC column for both Case study I and Case study II. A ChemDetect



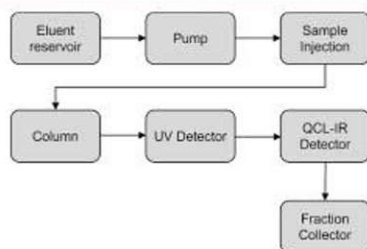


Figure 1. Scheme of the flow path in the LC-IR setup.

Analyzer (Daylight Solutions Inc., San Diego, USA) was used to record QCL-IR spectra.

**Size Exclusion Chromatography Conditions.** For preparative LC runs, the setup described in Figure 1 was used. Both runs were performed in the isocratic mode with a 50 mM phosphate buffer pH 7.4 (SEC buffer) with a constant flow of 7.5 cm/h (=0.25 mL/min). For Case study I, 0.5 mL of SEC buffer containing 10 mg/mL Ova, 10 mg/mL  $\alpha$ -CT, and 10 mg/mL Myo were injected. For case study II, 0.5 mL of SEC buffer containing 10 mg/mL HRP, 10 mg/mL  $\beta$ -LG, 10 mg/mL  $\alpha$ -CT, and 10 mg/mL Myo were injected. UV absorbance (280 nm) was recorded over the whole run and fractions with a volume of 1 mL were collected. For case study I, the protein concentration of the individual proteins in the collected fractions was measured using reversed-phase (RP) HPLC. For case study II, the concentration of HRP and Myo in the collected fractions were quantified using the described RP-HPLC method, while the concentrations of  $\beta$ -LG and  $\alpha$ -CT were obtained using a cation exchange (CEX) HPLC method.

**Laser-Based Mid-IR Measurements.** All mid-IR measurements were acquired with a ChemDetect Analyzer. The equipped EC-QCL was operated between 1350 and 1750  $\text{cm}^{-1}$  and thermally stabilized with an external water-cooling unit (set to 17  $^{\circ}\text{C}$ ). A custom-built, temperature-stabilized  $\text{CaF}_2$  flow cell with an optical path length of 25  $\mu\text{m}$  was used for all transmission measurements. The provided ChemDetect software package was used for spectra acquisition. For LC-QCL-IR in-line measurements, a background spectrum was acquired within 60 s by averaging 121 scans, followed by spectra acquisition every 10 s (averaging of 20 scans). Off-line reference measurements of pure protein solutions were performed by averaging 91 scans within 45 s. During spectra acquisition, the ChemDetect Analyzer was flushed with dry air to decrease the influence of water vapor from the atmosphere.

**HPLC Reference Measurements.** As an off-line analytical method to qualify and quantify proteins contained in the collected fractions, a previously published RP-HPLC method was used.<sup>36</sup> Because it was not possible to achieve satisfactory peak separation for  $\beta$ -LG and  $\alpha$ -CT using the RP-HPLC method (case study II), additionally, CEX HPLC measurements were performed. For that purpose, an UltiMate 3000 HPLC system (Thermo Fisher, MA, USA) equipped with a quaternary pump module, a temperature-controlled autosampler, a column oven, and a UV/vis detector module was used. The method used a MabPac SCX-10 (250 mm) column (Thermo Fisher, MA, USA) with a constant column temperature of 35  $^{\circ}\text{C}$  and a constant flow rate of 1 mL/min. In total, three mobile phases (mobile phase A: 20 mM

phosphate citrate buffer pH 4; mobile phase B: 20 mM phosphate citrate buffer pH 4 with 1 M NaCl; and mobile phase C: 50 mM phosphate buffer pH 7.4 with 1 M NaCl) were used, and the exact gradient profile is shown in Figure S1. In order to achieve sufficient separation, the pH value of all samples was adjusted to pH 4 (10 M phosphoric acid) prior to the measurement. An injection volume of 20  $\mu\text{L}$  was used for all samples and concentrations were calculated based on peak integration and comparison to measured standards with a known concentration. Standards were treated in the same way as samples, that is, dissolved to the desired concentration in SEC buffer and adjusted to pH 4 using 10 M phosphoric acid.

**Data Analysis.** In the present work, the separation of proteins by SEC was monitored with QCL-IR spectroscopy. For data analysis, the spectral range of the data matrix was cut to 1500–1700  $\text{cm}^{-1}$ , corresponding to the protein amide I and amide II bands, and the temporal range was limited to cover periods of protein elution. Prior to chemometric resolution, aiming to improve the S/N ratio, averaging of two data points was performed in the spectral axis. For case study I, an additional averaging of two spectra was performed in the time axis. Finally,  $273 \times 174$  (case study I) and  $271 \times 176$  (case study II) matrices were obtained and subjected to chemometric analysis.

Multicomponent spectroscopic signals generally follow Beer–Lambert's Law, hence they fulfill the concept of bilinear models described by

$$X = CS^T + E, \quad (1)$$

where  $X$  describes the two-way data matrix, and  $S$  and  $C$  contain the bilinear description of the data for both spectral profile and their relative concentrations, respectively;  $E$  contains the residuals of the model. In the applied workflow, spectral estimates for spectral profiles were obtained by SMMA. This group of techniques estimates the purest chemical factors and their contribution requiring any specific information about the data. In this regard, the pure variable-based methods, such as the simple-to-use interactive SSMA approach (SIMPLISMA), seek to obtain the selective spectral (or concentration) variables through the calculation of a purity value. The subsequently applied MCR, on the other hand, is a family of soft modeling techniques able to solve the bilinear description of the data for both spectral ( $S$ ) profiles and their relative concentrations ( $C$ ) through bilinear decomposition of the two-way data matrix  $X$  either by noniterative or iterative methods.

Data processing and chemometric analysis were performed in MATLAB R2020b (Mathworks, Inc., Natick, MA, 2020). MCR-ALS algorithms are available online at <http://www.mcrals.info/>.

**Protein Quantitation of Reconstituted Individual LC-QCL-IR Chromatograms.** Based on the concentration and spectral profiles obtained by the chemometric analysis, individual chromatogram matrices  $X_n (=c_n s_n^T)$  were reconstituted for every protein. These reconstituted IR spectra were employed to calculate protein concentrations ( $c$ ) across the chromatographic run according to the Beer–Lambert law

$$c = \frac{A}{\epsilon d} \quad (2)$$

Here,  $d$  is the path length of the transmission cell. The absorbance values ( $A$ ) were obtained by integrating the amide II bands (1500–1600  $\text{cm}^{-1}$ ) of the reconstituted QCL-IR



spectra. Absorption coefficients ( $\epsilon$ ) of the selected proteins were obtained by integrating the same spectral region of off-line acquired QCL-IR spectra of reference solutions with known protein concentrations.

## RESULTS AND DISCUSSION

To demonstrate the potential of the presented approach, monitoring of preparative SEC by a QCL-IR detector with subsequent chemometric analysis was performed with two-protein model systems exhibiting partial coelution of the proteins.

### Case Study I: A Model System with Three Proteins.

For case study I, a SEC run with three different proteins was performed. For this purpose, Ova,  $\alpha$ -CT, and Myo were identified to be proteins with different molecular weights and secondary structures. Figure 2A shows the results of in-line UV

results agree with the separation principle of SEC, where large molecules elute first. The ChemDetect Analyzer was used to record mid-IR spectra of the LC effluent across the chromatographic run.

Figure 2B displays the 3D plot (wavenumber-time-absorbance) of the performed LC-QCL-IR measurements. The plot shows stable baseline and chromatographic peaks with the characteristic amide I and II bands at retention times corresponding to the three proteins. For rather basic qualitative interpretation and discrimination between the three eluting proteins, chromatograms at wavenumbers characteristic for individual secondary structures can be extracted from the 3D data set and compared.<sup>40</sup> However, to gain more insight into the qualitative and quantitative information, an in-depth chemometric analysis needed to be performed. For this purpose, first, the number of components was estimated by singular value decomposition. Then, the purest spectral profiles were received by using a SIMPLISMA-like approach.<sup>41</sup> Subsequently, unconstrained MCR was applied to determine the corresponding time-dependent concentration profiles. At this point, it should be emphasized that for this analysis, no initial knowledge, for example, about the number and type of proteins, is required and all information can be derived from the recorded 3D QCL-IR data set. The obtained lack of fit (LOF, 2.5%) indicates a good description of the experimental data by the MCR model. By this approach, five components were determined to be needed to explain the experimental data, three of them were attributed to the eluting proteins in the chromatographic run, whereas the last two were associated with background signals. Figure 3A,B show the spectral and time-resolved concentration profiles, respectively, of the identified proteins. The retention times at the maximum of the concentration profiles agree very well with the peak maxima observed by reference techniques.

Analysis of the spectral profiles allows assigning secondary structures to the eluted proteins, even if no further reference information is available. The shapes of the amide I and amide II bands indicate that the first two eluting proteins are composed of mixed or  $\beta$ -sheet secondary structures, whereas the third protein mainly contains  $\alpha$ -helices. In case reference spectra are available (Figure 3C), identification of the eluted proteins is also possible. Ova features both  $\alpha$ -helical and  $\beta$ -sheet secondary structures, resulting in an amide I band maximum at 1656  $\text{cm}^{-1}$  with a shoulder at 1638  $\text{cm}^{-1}$  and a broad amide II band with the maximum at approximately 1545  $\text{cm}^{-1}$ .<sup>42,43</sup>  $\alpha$ -CT also contains  $\alpha$ -helices but is predominantly composed of  $\beta$ -sheets showing a characteristic broad amide I band with a maximum at 1635  $\text{cm}^{-1}$  and shoulders at 1650 and 1680  $\text{cm}^{-1}$ . The amide II band features a broad shape with a maximum at approximately 1548  $\text{cm}^{-1}$ .<sup>44</sup> Even though the spectral profiles of the first and second chromatographic peaks appear similar, evaluation of the different amide I band maxima and bandwidths allows the assignment of the first peak to Ova and the second peak to  $\alpha$ -CT. Finally, the third resolved spectral profile can be assigned to Myo, which is mainly composed of an  $\alpha$ -helical secondary structure.<sup>45</sup> The corresponding absorption spectrum shows a distinct amide I band at approximately 1656  $\text{cm}^{-1}$  and a narrow amide II band with a maximum at 1547  $\text{cm}^{-1}$ .

Two additional components obtained from the chemometric analysis were identified in the recorded QCL-IR data set (Figure S2). One concentration profile featured negative dips at the same retention times as the chromatographic protein

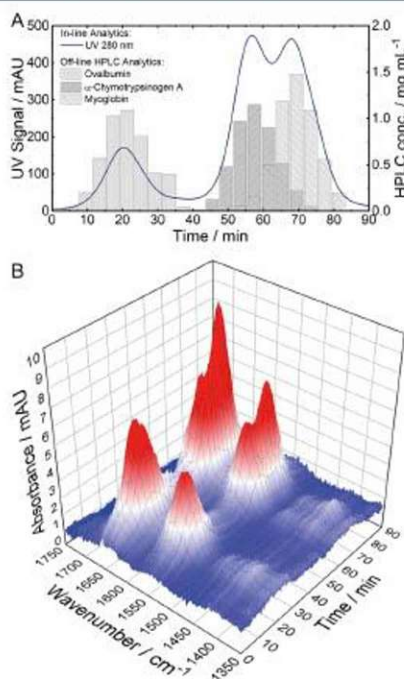
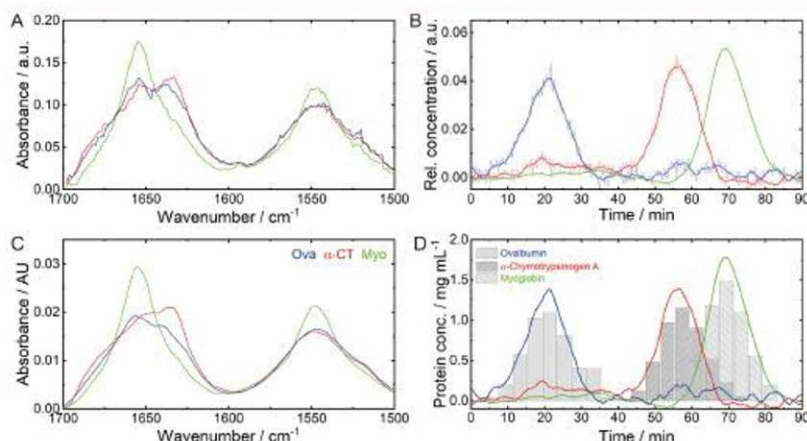


Figure 2. Experimental data obtained from the SEC run of case study I. (A) Results of in-line UV spectroscopy (left) and off-line HPLC analytics (right). (B) Spectral 3D plot recorded by the QCL-IR detector.

spectroscopy at 280 nm, indicating three chromatographic peaks. This signal is the most common for protein detection but does not provide any information regarding the secondary structure. Thus, in order to obtain qualitative and quantitative information about the eluting proteins, off-line HPLC analytics need to be performed (Figure 2A). The first chromatographic peak at 22 min can be related to Ova with a molecular weight of 44.5 kDa.<sup>37</sup>  $\alpha$ -CT and Myo have more similar molecular masses of 25.6 kDa<sup>38</sup> and 17 kDa,<sup>39</sup> respectively, and show overlapping peaks at approximately 57 and 70 min. These





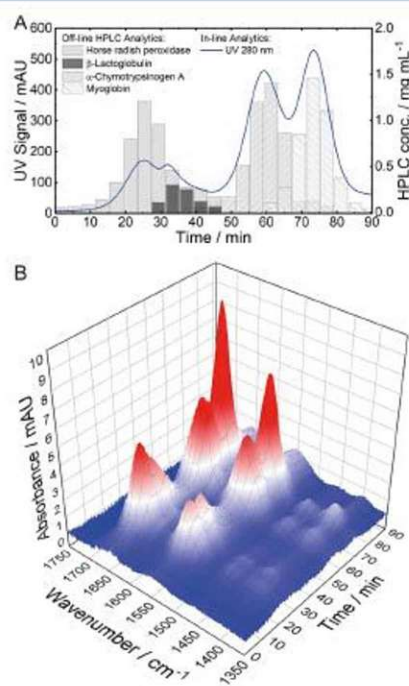
**Figure 3.** Results obtained by chemometric analysis of the bidimensional QCL-IR data set of case study I. (A) Spectral and (B) time-dependent concentration profiles retrieved by chemometric analysis. The obtained concentration profiles (thin) were smoothed by a Savitzky–Golay filter (thick). (C) Reference laser-based IR spectra of Ova,  $\alpha$ -CT, and Myo. (D) Protein concentrations obtained by in-line QCL-IR analysis (lines) and off-line reference HPLC analytics (bars).

peaks. Thus, this profile was assigned to the dilution of the buffer during protein elution, as the presence of proteins reduces the relative water content as compared to the background spectrum (pure buffer). Thus, due to the high absorption coefficient of the HOH-bending band, even subtle variations of the water content can introduce an observable effect on the IR spectra.<sup>46</sup> One further component was assigned to varying baseline and instrumental responses.

#### Case Study II: A Model System with Four Proteins.

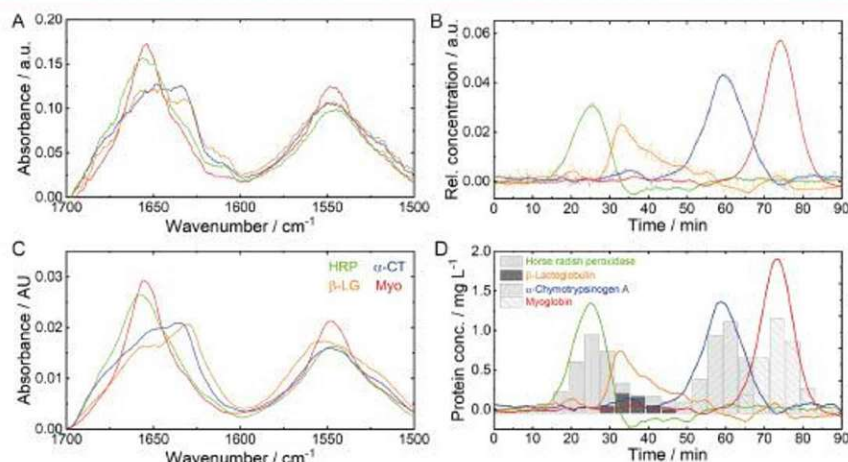
After the successful application of QCL-IR spectroscopy combined with chemometric analysis for protein structure identification and resolving of individual protein chromatograms with the model system comprising three proteins, a further, even more challenging case study was devised to validate the potential and versatility of the introduced method. To this end, a SEC run including HRP,  $\beta$ -LG,  $\alpha$ -CT, and Myo was performed. Figure 4A shows the results of in-line UV spectroscopy at 280 nm as well as off-line HPLC analytics, indicating four chromatographic peaks with a severe overlap of the first two protein peaks at 25 and 32 min. The first of these chromatographic peaks can be attributed to HRP with the highest molecular weight of 44 kDa.<sup>47</sup> The second peak is related to  $\beta$ -LG, which is present at the employed pH conditions in its dimeric form with a molecular mass of 36.7 kDa.<sup>48</sup> Due to the similar masses of these two proteins, they show highly coeluting behavior. Finally, the two remaining peaks at 59 and 72 min are assigned to  $\alpha$ -CT and Myo with molecular masses of 25.6<sup>38</sup> and 17 kDa,<sup>39</sup> respectively.

Figure 4B shows the 3D plot (wavenumber-time-absorbance) of the performed LC-QCL-IR measurement. This plot shows amide I and amide II maxima at retention times comparable with the reference techniques. Also, in this case, the data set was subjected to chemometric analysis to retrieve spectral and time-dependent concentration profiles of the eluting proteins. After obtaining initial estimates, MCR was performed obtaining an LOF of 2.3%, indicating a good fit of the MCR model to the experimental data. For this data set, six components were identified of which four could be assigned to



**Figure 4.** Experimental data obtained from the SEC run of case study II. (A) Results of in-line UV spectroscopy (left) and off-line HPLC analytics (right). (B) Spectral 3D plot recorded by the QCL-IR detector.

proteins in the chromatographic effluent. The spectral and time-resolved concentration profiles of the identified proteins



**Figure 5.** Results obtained by chemometric analysis of the bidimensional QCL-IR data set of case study II. (A) Spectral and (B) time-dependent concentration profiles retrieved by chemometric analysis. The obtained concentration profiles (thin) were smoothed by a Savitzky–Golay filter (thick). (C) Reference laser-based IR spectra of HRP,  $\beta$ -LG,  $\alpha$ -CT, and Myo. (D) Protein concentrations obtained by in-line QCL-IR analysis (lines) and off-line reference HPLC analytics (bars).

are shown in Figure 5A,B. The maximum positions of the concentration profiles agree very well with the retention times observed by the reference techniques. Without any prior knowledge, the secondary structures of the first and fourth eluting proteins could be assigned to be mostly  $\alpha$ -helical, while the second and third eluting proteins show spectral features of a mixed or  $\beta$ -sheet secondary structure. Comparison with the reference spectra allows identification of the first chromatographic peak as HRP with an amide I band maximum at  $1656\text{ cm}^{-1}$  with shoulders at  $1640$  and  $1680\text{ cm}^{-1}$  and an amide II band maximum at approximately  $1545\text{ cm}^{-1}$ .<sup>49,50</sup> It can be distinguished from Myo due to the slightly shifted amide I band maximum and the overall broader shape. The spectral profiles of the second and third identified chromatographic peaks have very similar shapes. However, comparison with reference spectra allows identifying the second peak as  $\beta$ -LG due to the narrower amide I band and the broader amide II band shapes.  $\beta$ -LG is predominantly composed of  $\beta$ -sheet secondary structures and shows a distinct amide I band with a maximum at  $1632\text{ cm}^{-1}$  and a shoulder at  $1660\text{ cm}^{-1}$  and a broad amide II band with a maximum at  $1550\text{ cm}^{-1}$ .<sup>51,52</sup> Finally, the spectral profile of the fourth chromatographic peak can be unanimously attributed to Myo.

Evaluation of the concentration profile reveals negative relative concentrations of HRP at retention times between 30 and 50 min. Impaired resolution of the bidimensional data set by MCR in this time region is caused by the severe overlap in retention times of HRP and  $\beta$ -LG as also shown by the reference HPLC results. Furthermore,  $\alpha$ -CT which also starts to elute in this period features similar IR spectral features as  $\beta$ -LG.

Chemometric analysis of this data set further identified two additional components (Figure S3), as occurred for case study I. Due to the shape of the concentration profiles, one was attributed to varying baseline and buffer dilution in the presence of proteins. The concentration profile of the other component shows a zigzag shape at rather low relative

concentrations. This periodic noise characteristic is partly also visible in the concentration profile of  $\beta$ -LG. It is present in the IR measurements, but not observable in the in-line UV measurements. The origin of these features was traced back to inconstant dry air supply throughout the QCL-IR measurements. The periodic behavior is introduced by a switching valve in the purge gas generator. Furthermore, the involvement of IR absorption of water vapor in this repetitive noise pattern is also supported by the narrow absorption bands in the related spectral profile.

Notwithstanding this instrumental issue, owing to the challenging model system and experimental difficulties, it was decided to present these results as proof to demonstrate the potency of the presented workflow to obtain reliable estimations by the combination of QCL-IR spectroscopy and chemometrics.

#### In-Line Protein Quantitation by QCL-IR Spectroscopy.

The reconstituted IR spectra obtained from chemometric analysis were employed for the calculation of the absolute protein concentrations. For this purpose, the area of the amide II band was integrated because this spectral region is less prone to water absorption-related intensity variations than the amide I band. Absorption coefficients of the proteins included in case studies I and II were obtained from QCL-IR off-line measurements with known protein concentrations. Figure 3D shows a comparison between the calculated concentrations based on QCL-IR spectroscopy and reference HPLC results of the collected fractions for case study I. The graph demonstrates highly overlapping concentration profiles between in-line and off-line reference measurements, indicating high validity of the presented approach based on QCL-IR spectroscopy and chemometrics. This evaluation was also performed for case study II (Figure 5D). Here, the elution profiles of all four proteins agree well between the two methods, even though absolute concentrations appear slightly shifted. These differences might be explained by the highly overlapping spectral features of  $\beta$ -LG and  $\alpha$ -CT and the



pronounced overlap of the chromatographic peaks of HRP and  $\beta$ -LG. Those challenging circumstances may adversely affect chemometric analysis and lead to underestimation of the HRP content at retention times between 30 and 50 min while overestimating the  $\beta$ -LG content. Nevertheless, these good results, in spite of the complex data sets, indicate high flexibility and robustness of the presented LC-QCL-IR approach.

Case study II also was challenging to resolve for off-line RP-HPLC analytics. The insufficient peak resolution of  $\beta$ -LG and  $\alpha$ -CT required the introduction of an additional CEX-HPLC method in order to accurately identify and quantify the proteins in the collected fractions. As two different off-line HPLC methods are required to analyze fractions in case study II, this further emphasizes the difficulties to establish straightforward reference analytics. Consequently, QCL-IR in-line detectors hold significant potential for providing near-real-time protein concentrations from chromatographic separation processes by achieving similar results as conventionally applied time- and cost-intensive off-line methods.

## CONCLUSIONS AND OUTLOOK

In this work, an EC-QCL-based mid-IR spectrometer was successfully hyphenated to a preparative SEC system for in-line discrimination of proteins from highly overlapping chromatographic peaks. The advantages of QCL-IR detectors over conventional LC detectors were demonstrated in two case studies, involving mixtures of three and four different proteins, respectively. Due to similar molecular weights of the proteins, highly overlapping chromatographic peaks were obtained that could not be distinguished with a standard UV detector. In contrast, QCL-IR detection enabled the acquisition of multivariate data sets, containing mid-IR absorbance spectra across the chromatographic runs that provide information regarding protein secondary structure, thus allowing protein identification. These data sets were investigated by chemometrics to obtain spectral profiles of the individual proteins as well as their relative concentration profiles. The obtained spectra agree well with reference off-line spectra of pure protein solutions. Furthermore, absolute protein concentrations were calculated according to the Beer–Lambert law, showing high agreement with HPLC reference measurements of the collected effluent fractions. Consequently, QCL-IR in-line detectors can provide qualitative and quantitative information about proteins, comparable to time- and labor-intensive off-line methods that are inaccessible with conventional UV detectors. Hence, the in-line QCL-IR system has the capability to (i) accelerate process development and to (ii) monitor production processes on-line, allowing in-process control. Furthermore, the presented system enables QbD principles and concurs with the requirements of a PAT tool, providing information about protein secondary structure in real-time.

Finally, an important property of the presented in-line QCL-IR detection of preparative LC is its accordance with green analytical chemistry (GAC) principles.<sup>53</sup> For a comprehensive comparison of in-line QCL-IR spectroscopy and off-line HPLC analysis, the previously introduced Analytical GREENness (AGREE) metric approach,<sup>54</sup> a straightforward assessment approach based on the 12 principles of GAC (SIGNIFICANCE),<sup>55</sup> was applied. The results for both methods are shown in Figure S4. The scores of 0.84 for QCL-IR spectroscopy and 0.43 for off-line HPLC indicate a clear

superiority of the presented in-line QCL-IR method in terms of the greenness of the analytical procedure.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.2c01542>.

Gradient profile used in the CEX-HPLC method; additional spectra and time-resolved concentration profiles retrieved by chemometric analysis of the chromatographic run of case study I and case study II; results of evaluation according to the Analytical GREENness metric approach for in-line QCL-IR spectroscopy and off-line HPLC analysis; and selected options for the Analytical GREENness evaluation are presented here (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Andreas Schwaighofer – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; [orcid.org/0000-0003-2714-7056](https://orcid.org/0000-0003-2714-7056); Email: [andreas.schwaighofer@tuwien.ac.at](mailto:andreas.schwaighofer@tuwien.ac.at)

Bernhard Lendl – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; [orcid.org/0000-0003-3838-5842](https://orcid.org/0000-0003-3838-5842); Email: [bernhard.lendl@tuwien.ac.at](mailto:bernhard.lendl@tuwien.ac.at)

### Authors

Christopher K. Akhgar – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; [orcid.org/0000-0001-8266-043X](https://orcid.org/0000-0001-8266-043X)

Julian Ebner – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Mirta R. Alcaraz – Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, S3000ZAA Santa Fe, Argentina; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQB CABA, Argentina

Julian Kopp – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Héctor Goicoechea – Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, S3000ZAA Santa Fe, Argentina; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQB CABA, Argentina; [orcid.org/0000-0001-7145-0082](https://orcid.org/0000-0001-7145-0082)

Oliver Spadiut – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.2c01542>

### Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript. C.K.A. and J.E. contributed equally.



## Funding

Open Access is funded by the Austrian Science Fund (FWF).

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work has received funding from the COMET Center CHASE (project no. 868615), funded within the COMET—Competence Centers for Excellent Technologies program by the BMK, the BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is managed by the Austrian Research Promotion Agency (FFG). Additional funding was provided by the European Union's Horizon 2020 research and innovation program through the NUTRISHIELD project under grant agreement no. 818110. This research was further funded by the Austrian Research Promotion Agency (FFG) (project no. 874206) and by the Austrian Science Fund FWF (project no. P32644-N). M.R.A. and H.G. acknowledge CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina) for funding.

## REFERENCES

- Deutscher, M. P. Chapter 5 Setting Up a Laboratory. In *Methods in Enzymology*; Burgess, R. R., Deutscher, M. P., Eds.; Academic Press: 2009; pp 37–42.
- Sun, Y.; Shi, Q.; Zhang, L.; Zhao, G. F.; Liu, F. F. 247—Adsorption and Chromatography. In *Comprehensive Biotechnology* (2nd ed.); Moo-Young, M., Ed.; Academic Press: Burlington, 2011; pp 665–679.
- Barth, A. *Biochim. Biophys. Acta, Bioenerg.* 2007, 1767, 1073–1101.
- Quintás, G.; Kuligowski, J.; Lendl, B. *Anal. Chem.* 2009, 81, 3746–3753.
- Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Talanta* 2010, 50, 1771–1776.
- Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Anal. Chim. Acta* 2008, 624, 278–285.
- Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* 2009, 1216, 3122–3130.
- Quintás, G.; Lendl, B.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* 2008, 1190, 102–109.
- Quintás, G.; Kuligowski, J.; Lendl, B. *Appl. Spectrosc.* 2009, 63, 1363–1369.
- Singh, B. R. Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins. *Infrared Analysis of Peptides and Proteins*; American Chemical Society: Washington D.C., USA, 1999; pp 2–37.
- Fabian, H.; Mantele, W. Infrared Spectroscopy of Proteins. In *Handbook of Vibrational Spectroscopy*; John Wiley & Sons: Hoboken, USA, 2006.
- Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. *Nat. Protoc.* 2015, 10, 382–396.
- Turula, V.; de Haseth, J. *Appl. Spectrosc.* 1994, 48, 1255–1264.
- Turula, V. E.; de Haseth, J. A. *Anal. Chem.* 1996, 68, 629–638.
- Turula, V.; Bishop, R.; Ricker, R.; de Haseth, J. *J. Chromatogr. A* 1997, 763, 91–103.
- Kuligowski, J.; Quintás, G.; Guardia, M.; Lendl, B. Liquid Chromatography—Liquid Chromatography—Fourier Transform Infrared. *Encyclopedia of Analytical Science*, 3rd ed.; Elsevier: Amsterdam, Netherlands, 2019; pp 75–85.
- Ramer, G.; Lendl, B. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy. In *Encyclopedia of Analytical Chemistry*; Meyers, R., Meyers, R., Eds.; John Wiley & Sons: Hoboken, USA, 2013.
- Großhans, S.; Rüdert, M.; Sanden, A.; Brestrich, N.; Morgenstern, J.; Heissler, S.; Hubbuch, J. *J. Chromatogr. A* 2018, 1547, 37–44.
- Sanden, A.; Suhm, S.; Rüdert, M.; Hubbuch, J. *J. Chromatogr. A* 2019, 1608, 460410.
- Walch, N.; Scharl, T.; Felföldi, E.; Sauer, D. G.; Melcher, M.; Leisch, F.; Dürner, A.; Jungbauer, A. *Biotechnol. J.* 2019, 14, No. e1800521.
- Faist, J.; Capasso, F.; Sivco, D.; Sirtori, C.; Hutchinson, A.; Cho, A. *Science* 1994, 264, 553–556.
- Schwaighofer, A.; Brandstetter, M.; Lendl, B. *Chem. Soc. Rev.* 2017, 46, 5903–5924.
- Schwaighofer, A.; Lendl, B. Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In *Vibrational Spectroscopy in Protein Research*; Ozaki, Y.; Baranska, M.; Lednev, I.; Wood, B., Eds.; Academic Press: Cambridge, USA, 2020; pp 59–88.
- Alcaráz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. *Anal. Chem.* 2015, 87, 6980–6987.
- Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. *Anal. Chem.* 2018, 90, 7072–7079.
- Akhgar, C. K.; Ramer, G.; Zbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. *Anal. Chem.* 2020, 92, 9901–9907.
- Chon, B.; Xu, S.; Lee, Y. J. *Anal. Chem.* 2021, 93, 2215–2225.
- Dabrowska, A.; David, M.; Freitag, S.; Andrews, A. M.; Strasser, G.; Hinkov, B.; Schwaighofer, A.; Lendl, B. *Sens. Actuators, B* 2022, 350, 130873.
- Schwaighofer, A.; Akhgar, C. K.; Lendl, B. *Spectrochim. Acta, Part A* 2021, 253, 119563.
- Akhgar, C. K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. *Anal. Chem.* 2022, 94, 5583–5590.
- de Juan, A.; Jaumot, J.; Tauler, R. *Anal. Methods* 2014, 6, 4964.
- Kucheryavskiy, S.; Windig, W.; Bogomolov, A. Chapter 3—Spectral Unmixing Using the Concept of Pure Variables. In *Data Handling in Science and Technology*; Ruckebusch, C., Ed.; Elsevier, 2016; pp 53–99.
- Tauler, R. *Chemom. Intell. Lab. Syst.* 1995, 30, 133–146.
- Hu, B.; Sun, D.-W.; Pu, H.; Wei, Q. *Talanta* 2020, 217, 120998.
- Mansoldo, F. R. P.; Berrino, E.; Guglielmi, P.; Carradori, S.; Carta, F.; Secci, D.; Supuran, C. T.; Vermelho, A. B. *Spectrochim. Acta, Part A* 2022, 267, 120602.
- Kopp, J.; Zauner, F. B.; Pell, A.; Hausjell, J.; Humer, D.; Ebner, J.; Herwig, C.; Spadiut, O.; Slouka, C.; Pell, R. *J. Pharm. Biomed. Anal.* 2020, 188, 113412.
- Strixner, T.; Kulozik, U. 7—Egg proteins. In *Handbook of Food Proteins*; Phillips, G. O., Williams, P. A., Eds.; Woodhead Publishing, 2011; pp 150–209.
- Wilcox, P. E. [5] Chymotrypsinogens—chymotrypsins. *Methods in Enzymology*; Academic Press, 1970; pp 64–108.
- Zaia, J.; Annan, R. S.; Biemann, K. *Rapid Commun. Mass Spectrom.* 1992, 6, 32–36.
- Akhgar, C.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography. *Biomedical Vibrational Spectroscopy 2022: Advances in Research and Industry*; SPIE, 2022; Vol. 11957.
- Windig, W.; Bogomolov, A.; Kucheryavskiy, S. Two-Way Data Analysis: Detection of Purest Variables. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S.; Tauler, R.; Walczak, B., Eds.; Elsevier, 2020; pp 275–307.
- Stein, P. E.; Leslie, A. G. W.; Finch, J. T.; Turnell, W. G.; McLaughlin, P. J.; Carrell, R. W. *Nature* 1990, 347, 99–102.
- Dong, A.; Meyer, J. D.; Brown, J. L.; Manning, M. C.; Carpenter, J. F. *Arch. Biochem. Biophys.* 2000, 383, 148–155.
- Freer, S. T.; Kraut, J.; Robertus, J. D.; Wright, H. T.; Nguyen-Huu-Xuong, X. *Biochemistry* 1970, 9, 1997–2009.



- (45) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* 1958, 181, 662–666.
- (46) Kuligowski, J.; Schwaighofer, A.; Alcaráz, M. R.; Quintás, G.; Mayer, H.; Vento, M.; Lendl, B. *Anal. Chim. Acta* 2017, 963, 99–105.
- (47) Wisdom, G. B. Horseradish Peroxidase Labeling of IgG Antibody. In *The Protein Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2009; pp 681–683.
- (48) Madureira, A. R.; Pereira, C. L.; Gomes, A. M. P.; Pintado, M. E.; Xavier Malcata, F. *Food Res. Int.* 2007, 40, 1197–1211.
- (49) Gajhede, M.; Schuller, D. J.; Henriksen, A.; Smith, A. T.; Poulos, T. L. *Nat. Struct. Mol. Biol.* 1997, 4, 1032–1038.
- (50) Tavares, T. S.; da Rocha, E. P.; Esteves Nogueira, F. G.; Torres, J. A.; Silva, M. C.; Kuca, K.; Ramalho, T. C. *Molecules* 2020, 25, 259.
- (51) Dousseau, F.; Pezolet, M. *Biochemistry* 1990, 29, 8771–8779.
- (52) Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. *J. Mol. Biol.* 1987, 197, 695–706.
- (53) Armenta, S.; Garrigues, S.; de la Guardia, M. *Trends Anal. Chem.* 2008, 27, 497–511.
- (54) Pena-Pereira, F.; Wojnowski, W.; Tobiszewski, M. *Anal. Chem.* 2020, 92, 10076–10082.
- (55) Galuszka, A.; Migaszewski, Z.; Namieśnik, J. *Trends Anal. Chem.* 2013, 50, 78–84.

## 7 Appendix II

### 7.1 Book Chapter: The Purification of Heme Peroxidases from *E. coli* Inclusion Bodies



## Chapter 16

### The Purification of Heme Peroxidases from *Escherichia coli* Inclusion Bodies: A Success Story Shown by the Example of Horseradish Peroxidase

Diana Humer and Julian Ebner

#### Abstract

In the following chapter a purification process for recombinant Horseradish peroxidase (HRP) produced in *Escherichia coli* is described. This enzyme is a secretory plant oxidoreductase belonging to the large peroxidase family III within the peroxidase-catalase superfamily of enzymes. It has high biotechnological significance, however, the isolation of the enzyme from its natural source, the horseradish root, has several shortcomings, which makes the development of a recombinant production strategy interesting. The presented protocol covers all process steps from isolation to the final chromatography step; the enzyme is solubilized from insoluble inclusion bodies, refolded and concentrated to yield a high purity enzyme preparation which is comparable to the commercially available plant-derived HRP. Moreover, we believe that this procedure can also be used to process other peroxidases of family II and III of the plant peroxidase superfamily, as they all share the same relevant features like disulfide bonds and a heme group.

**Key words** Inclusion bodies, Horseradish peroxidase, Heme peroxidases, *Escherichia coli*, Refolding, Hydrophobic interaction chromatography, Recombinant, Reversed phase high-performance liquid chromatography

#### 1 Introduction

The isolation of plant enzymes from their natural source can be cumbersome and the yield depends on several unpredictable environmental factors. This can be detrimental when a steady supply of the enzyme in question is required for biotechnological or molecular biological applications. Recombinant protein production can circumvent these problems; however, this is not trivial for enzymes with a complex structure like horseradish peroxidase (HRP). This oxidoreductase from the horseradish root requires the establishment of four disulfide bonds and the incorporation of two calcium ions as well as a protoporphyrin ring (called the “heme group”) for enzyme function. Moreover, about 20% of the enzymes total

Julian Kopp and Oliver Spadiut (eds.), *Inclusion Bodies: Methods and Protocols*, Methods in Molecular Biology, vol. 2617, [https://doi.org/10.1007/978-1-0716-2930-7\\_16](https://doi.org/10.1007/978-1-0716-2930-7_16), © The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023



molecular weight of 44 kDa amount to Asn-bound sugar chains on the surface, the so-called glycosylation. In general, glycosylation is host-specific and cannot be performed by prokaryotes like *Escherichia coli* (*E. coli*) as they lack the respective cell organelles. When glycans are absent, the thermal stability is reduced and the hydrophobicity of the enzyme increases [1, 2]. Currently, HRP is isolated from the horseradish plant (*Armoracia rusticana*). This isolate consists of a variety of HRP isoforms that diverge in their biochemical properties (for example molecular weight, isoelectric point or substrate affinity) [1, 3, 4]. Hence, a recombinantly produced single isoform with consistent quality would be a desirable alternative for industrial purposes. The present chapter describes a refolding process for recombinant HRP from inclusion bodies (IBs) produced in *E. coli*. Moreover, this protocol can also be applied for the expression of other plant heme peroxidases. The renaturation of HRP from IBs comprises several distinct steps, the isolation of IBs from the host cells; which is most often performed by high pressure homogenization (for detailed information refer to the Chapter “High pressure homogenization for inclusion body isolation”); washing, solubilization, refolding, and purification/capture. These unit operations have all been investigated during the development of this protocol. Especially the interaction of solubilization and refolding conditions is of great importance, as the formation of the four disulfide bonds requires a redox system, which consists of dithiothreitol (DTT) and L-glutathione oxidized (GSSG) in this case (for further information refer to the Chapter “Unit operation-spanning investigation of the redox system”). Another highly relevant process step is the addition of the cofactor hemin, which was continuously added for several hours to avoid aggregation. Finally, HRP was purified by hydrophobic interaction chromatography (HIC) and the resulting enzyme preparation showed high purity and enzymatic activity comparable to the commercially available, glycosylated HRP isolated from the horseradish root.

## 2 Materials

The solutions and buffers should be prepared using ultrapure water and can be stored at room temperature (RT) if not indicated otherwise.

### 2.1 Equipment

- IKA T10 basic ULTRA-TURRAX (Staufen, Germany).
- Homogenizer GEA Niro Soavi Panda PLUS (Düsseldorf, Germany).
- Thermo Scientific™ Sorvall™ LYNX™ centrifuge (Waltham, MA, US).

- Sigma 4-16KS refrigerated benchtop centrifuge (St. Louis, MO, USA).
- Infors Labfors 5 vessel (Bottmingen, Germany) 3.6 L.
- Lucullus process control system (Biospectra, Schlieren, Switzerland).
- LAMBDA PRECIFLOW peristaltic liquid pump (LAMBDA laboratory instruments, Switzerland) in combination with a Sartorius Entris scale (Sartorius, Germany).
- Lauda Alpha R8 thermostat (Lauda, Königshofen, Germany).
- Dionex UltiMate 3000 HPLC system with a quaternary solvent delivery pump, an auto-sampler with a sample thermostat and an UV detector (Thermo Fisher, Waltham, MA, USA).
- pH meter.
- ÄKTA Pure system (Cytiva, Marlborough, MA, US).
- Cytiva Butyl Sepharose 4 Fast Flow resin and a HiScale™ 26/40 column.
- Hitachi Double Beam Spectrophotometer U-2900 (Tokyo, Japan).
- Tecan Infinite M200 PRO (Männedorf, Switzerland).

## 2.2 Isolation and Solubilization

1. Homogenization buffer: 50 mM Tris, pH 8, 500 mM NaCl, 1.5 mM EDTA. Weigh in 6.06 g tris(hydroxymethyl)aminomethane (Tris), 29.2 g sodium chloride, and 0.56 g ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA), add 900 mL of water and mix. After all components are dissolved, adjust the pH to 8 with HCl, then fill to 1 L with water.
2. Resuspension buffer: 50 mM Tris, pH 8, 500 mM NaCl, 2 M Urea. Weigh in 6.06 g Tris, 29.2 g sodium chloride, and 120.1 g urea, add 700 mL of water and mix (*see Note 1*). After all components are dissolved adjust the pH to 8 with HCl, then fill to 1 L with water.
3. Solubilization buffer: 50 mM glycine, pH 10, 6 M urea. Weigh in 1.88 g glycine and 180.2 g urea, add 900 mL of water and mix (again *see Note 1*). After all components are dissolved, adjust the pH to 10 with NaOH, then fill to 0.5 L with water.
4. DTT stock solution: 1 M DTT in water. Weigh in 0.1543 g DTT in a 1.5 mL reaction tube and add 1 mL water. Store at 2–8 °C.

## 2.3 Refolding

1. Refolding buffer: 20 mM glycine, pH 10, 2 M urea, 2 mM CaCl<sub>2</sub>, 7% v/v glycerol, 1.27 mM GSSG. Weigh in 3 g glycine, 240.2 g urea, 0.588 g calcium chloride dihydrate (CaCl<sub>2</sub>), 142.9 mL glycerol 98%, 1.56 g GSSG (*see Note 2*), add 1.5 L



water and mix. After all components are dissolved, adjust the pH to 10 with NaOH, then fill to 2 L with water. Store at 2–8 °C.

2. Hemin stock solution: 1 mM hemin in 100 mM KOH. Weigh in 0.0196 g hemin, add 10 mL water, and then 600  $\mu$ L 5 M potassium hydroxide (KOH) (*see Note 3*), mix and then fill to 30 mL with water. Store at 2–8 °C.

## 2.4 Chromatography

1. Equilibration and wash buffer: 20 mM Bis-Tris, pH 7, 4 M NaCl. Weigh in 20.9 g of Bis-(2-hydroxyethyl)-imino-tris (hydroxymethyl)-methane (Bis-Tris) and 1169 g sodium chloride, add 4 L water and mix. After all components are dissolved, adjust the pH to 7 with HCl, then fill to 5 L with water.
2. Elution buffer: 20 mM Bis-Tris, pH 7. Weigh in 12.6 g of Bis-Tris and add 2.5 L water and mix. After all components are dissolved, adjust the pH to 7 with HCl, then fill to 3 L with water.
3. Wash solution 1: 1 M NaOH. Dissolve 40 g NaOH in 1 L of water.
4. Wash solution 2: 70% EtOH. Dilute 729.2 mL ethanol 96% with 270.8 mL water.
5. Storage solution: 20% EtOH. Dilute 208.3 mL ethanol 96% with 791.7 mL water (*see Note 4*).

## 2.5 Enzyme Activity Measurements

1. Measurement buffer: 50 mM phosphate-citrate, pH 5. Weigh in 6.7 g disodium phosphate heptahydrate, add 0.3 L of water and mix then adjust the pH to 5 with 2 M citric acid and fill to 0.5 L with water.
2. 2 M citric acid stock solution: Weigh in 210.14 g citric acid monohydrate and dissolve in 0.5 L water.
3. 10 mM hydrogen peroxide stock solution: Add 10.2  $\mu$ L of a 30% hydrogen peroxide solution (9.8 M) and fill to 10 mL with water.
4. Enzyme dilution buffer: 20 mM Bis-Tris, pH 7. Weigh in 2.1 g of Bis-Tris, dissolve in 0.4 L water, and then adjust the pH to 7 with HC and fill to 0.5 L with water.

## 2.6 RP-HPLC

1. Mobile phase A: Ultrapure water with 0.1% trifluoroacetic acid (TFA). Add 1 mL TFA after sonication (*see Note 5*) to 1 L water.
2. Mobile phase B: Acetonitrile (ACN) with 0.1% TFA. Add 1 mL TFA after sonication (also *see Note 5*) to 1 L ACN.
3. Sample preparation (for IBs): 62 mM Tris, 7.5 M GdnHCl, 125 mM DTT. Weigh in 35.8 g guanidine hydrochloride, 0.964 g DTT, 0.376 g Tris, dissolve in 20 mL water (*see Note 6*), and fill up to 50 mL afterwards.

### 2.7 SE-HPLC

1. Mobile phase: 20 mM phosphate buffer, pH 6.8, 250 mM KCl. Weigh in 18.6 g of potassium chloride (KCl), 3.5 g of dipotassium hydrogen phosphate, and 2.7 g of potassium dihydrogen phosphate, dissolve in 900 mL water, check if pH is  $6.8 \pm 0.1$ , if not adjust with phosphoric acid or KOH, then fill to 1 L with water.

## 3 Methods

An overview of the chosen parameters for the refolding process is given in Table 1.

### 3.1 Isolation and Solubilization

1. Resuspend the biomass thoroughly (*see Note 7*) in homogenization buffer ( $200 \text{ g L}^{-1}$ ) biomass concentration) using an ULTRA-TURRAX device (IKA T10 basic ULTRA-TURRAX). Then homogenize at  $\geq 1200$  bar, cooled, three passages (GEA Niro Soavi Panda PLUS) and centrifuge the homogenized suspension ( $15,650 \text{ g}$ ; 20 min,  $4^\circ\text{C}$ ). Discard the supernatant and resuspend the pellet in 10 mL resuspension buffer per g wet cell debris using the ULTRA-TURRAX and centrifuge ( $15,650 \text{ g}$ ; 5 min,  $4^\circ\text{C}$ ) (*see Note 8*). Perform this washing step with resuspension buffer twice (*see Note 9*). Then resuspend the pellet in water (*see Note 10*) ( $200 \text{ g L}^{-1}$ ) and aliquot into preweighed 50 mL reaction tubes, centrifuge ( $15,650 \text{ g}$ ; 5 min,  $4^\circ\text{C}$ ) and store at  $-20^\circ\text{C}$ .

**Table 1**  
Parameters for each unit operation

Unit operation	Parameters	Final setpoints
Solubilization	DTT	7.11 mM
	Protein concentration	$20 \text{ g L}^{-1}$
	pH	10
Refolding	GSSG	1.27 mM
	Protein concentration	$0.5 \text{ g L}^{-1}$
	pH	10
	Time of hemin addition	8 h after refolding start
	Duration hemin feed	10 h
	Final hemin concentration	$20 \mu\text{M}$
	Total refolding time	19 h
Salt precipitation	Type of salt	NaCl
	Salt concentration	4 M
	Hydrophobicity of resin	Butyl
Capture step HIC	pH value (load)	8.5
	Type of elution	Step gradient

Adapted from [6]



2. For solubilization, resuspend the preweighed pellet in solubilization buffer using the ULTRA-TURRAX device to reach a wet inclusion body concentration of  $100 \text{ g L}^{-1}$  (see Note 11). Hence, for a refolding volume of 1.2 L use 3.5 g wet IBs and resuspend in solubilization buffer to a final volume of 35 mL. After resuspension, add 7.11 mM DTT and incubate at room temperature (RT) for 30 min with slight agitation, then centrifuge at  $20,379 \text{ g}$ ; 20 min,  $4^\circ\text{C}$ . Afterwards discard the pellet and immediately use the supernatant for refolding (see Note 12).

### 3.2 Refolding

1. The refolding procedure should ideally be performed in a stirred refolding vessel at 200 rpm and  $4\text{--}10^\circ\text{C}$  (e.g., 3.6 L Infors Labfors 5 vessel). For a total refolding volume of 1.2 L, fill the vessel with 1146 mL (0.955 L refolding buffer/L refolding batch) refolding buffer. Then connect the sparger to the gas outlets on the tower and the double jacket to the cooling device.
2. Calibrate the pH and Redox probe and install them in the vessel, then install the  $\text{dO}_2$  probe, the temperature sensor, and the stirrer engine. Next, start a new process operation using a process information management system (e.g., Lucullus). Now the  $\text{dO}_2$  probe can be calibrated using air and  $\text{N}_2$ .
3. Afterwards slowly add 30 mL solubilize with a serological pipette (see Note 13). Refolding should be performed for 8 h at  $10^\circ\text{C}$ , 200 rpm and pH, temperature,  $\text{dO}_2$ , and redox potential (see Note 14) are monitored using Lucullus. After 8 h, add the cofactor hemin over 10 h (see Note 15) using a PID-Feed forward control (Lucullus) with a Lambda preciflow peristaltic pump and a Sartorius scale (24 mL feed for 1.2 L refolding =  $20 \text{ g L}^{-1} \text{ h}^{-1}$  for 10 h, resulting in a final concentration of  $20 \mu\text{M}$  hemin). After the 10 h feed, the process continues for an additional hour, hence the total refolding time is 19 h.

### 3.3 Chromatography

1. Prior to HIC, the pH of the refolding solution is adjusted to 8.5 with HCl (see Note 16). As a preparation for the chromatography step, impurities and aggregates are removed by salt precipitation. Add 0.267 g sodium chloride per mL refolding solution in small portions over 10 min while stirring at RT, then stir for additional 20 min, and centrifuge at  $17,568 \text{ g}$  for 25 min at  $4^\circ\text{C}$ . Discard the pellet and immediately use the supernatant for chromatography.
2. Subsequently, HRP is concentrated and purified using HIC. Connect a chromatography column (e.g., HiScale™ 26/40 column) packed with 80 mL Butyl Sepharose 4 Fast Flow to an ÄKTA pure 25 system (see Note 17). An overview of the chromatography parameters is given in Table 2. The UV-VIS signals are monitored at 280 and 404 nm (see Note 18). Equilibrate the column with the buffer for equilibration and wash at a linear flow rate of  $90 \text{ cm h}^{-1}$  for 2 column volumes

**Table 2**  
**Chromatography parameters**

Step	Solution	CV	Linear flow rate [cm h <sup>-1</sup> ]
Equilibration	Equilibration and wash buffer	2	90
Load	Refolding supernatant	–	75
Wash	Equilibration and wash buffer	3	90
Elution step 1	80% equilibration and wash buffer and 20% elution buffer	3	90
Elution step 2	25% equilibration and wash buffer and 75% elution buffer	3	90
Elution step 3	100% elution buffer	3	90
Cleaning step 1	1 M NaOH	10	75
Cleaning step 2	70% EtOH	10	90
Storage	20% EtOH	5	90

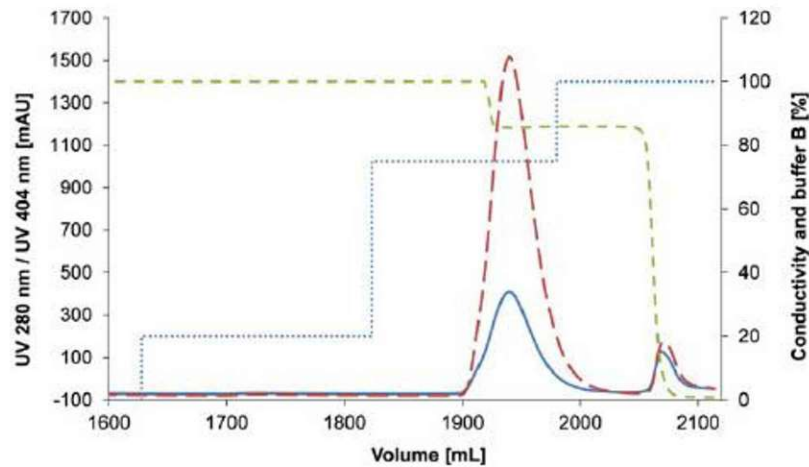
(CV). Then load the refolding supernatant at a flow rate of 75 cm h<sup>-1</sup> and afterwards wash the column with 3 CV equilibration and wash buffer at 90 cm h<sup>-1</sup>. After washing, the UV signal (280 nm) should be below 10 mAU, then step elution can be started. The percentage of elution buffer is increased at a flow rate of 90 cm h<sup>-1</sup> to 20, 75, and finally 100% (3 CV each), whereby HRP elutes from the column at 75% elution buffer (Fig. 1). Finally, clean the column as described in Table 2. The final process results which can be achieved are shown in Table 3.

### 3.4 Enzyme Activity Measurements

Measure the volumetric enzyme activity (U mL<sup>-1</sup>) using the ABTS assay in a Tecan 96-well plate reader with flat-bottom polystyrene 96-well plates. The reaction mixture should contain a saturating hydrogen peroxide concentration of 1 mM, and 7 mM ABTS in 50 mM phosphate-citrate buffer pH 5 (*see Note 19*). Prior to the measurements, dilute the protein samples accordingly in enzyme dilution buffer so that the linear increase in absorption at 420 nm can be followed for 120 s at 30 °C. For the measurement, mix 5 µL of the protein sample with 175 µL ABTS-buffer mixture and then start the reaction by adding 20 µL of a 10 mM hydrogen peroxide stock solution (*see Note 20*). Calculate the volumetric enzyme activity using the following Eq. (1):

$$A [\text{U mL}^{-1}] = \frac{V_{\text{total}} * \Delta A \text{ min}^{-1} * \text{dilution}}{V_{\text{sample}} * d * \epsilon} \quad (1)$$





**Fig. 1** Elution of active HRP and impurities using hydrophobic interaction chromatography. Active HRP elutes at 75% buffer B, while hydrophobic impurities elute at 100% buffer B. Blue solid line, UV 280 nm [mAU]; red long dashed line, UV 404 nm [mAU]; blue dotted line, buffer B [%]; green short dashed line, conductivity [mS/cm]

**Table 3**  
**Final process results**

Process variables	Final results
Specific activity [U mg <sup>-1</sup> ]	1468 ± 24
Purity SEC-HPLC [%]	≥99
Overall yield active HRP per 100 mg expressed protein [mg]	28
Pure HRP/L culture medium [mg]	959
Rz	4.3

Adapted from [6]

$V_{\text{total}}$  ... total volume in cuvette in [μL]

$\Delta A \text{ min}^{-1}$  ... change in absorption [ $\Delta \text{Abs } 420 \text{ nm min}^{-1}$ ]

Dilution ... dilution factor of the sample

$V_{\text{sample}}$  ... volume of the sample [μL]

$d$  ... length of the beam path through the cuvette ( $d = 0.58 \text{ cm}$ )

$\epsilon$  ... extinction coefficient ( $\epsilon_{420} = 36 \text{ mM}^{-1} \text{ cm}^{-1}$  [5])

### 3.5 Reinheitszahl

Measure the Reinheitszahl (ratio of absorbance at 404 nm to 280 nm) of enzyme samples with a spectrophotometer using quartz cuvettes (absorbance scan 190–500 nm).

**Table 4**  
**RP-HPLC conditions**

Solution	Time [min]	Flow rate [mL min <sup>-1</sup> ]
75% Mobile phase A and 25% Mobile phase B	0.5	1.2
25–55% Mobile phase B in a linear gradient	8	1.2
45% Mobile phase A and 55% Mobile phase B	0.5	1.2
75% Mobile phase A and 25% Mobile phase B	1	1.2

**3.6 RP-HPLC**

Determine the HRP concentration using RP-HPLC with a Polyphenyl BioResolve-RP-mAb 2.7  $\mu\text{m}$  3.0  $\times$  100 mm column. Run the method for 10 min at a flow rate of 1.2 mL min<sup>-1</sup> as described in Table 4. Column oven temperature should be set to 75 °C and the wavelengths 214, 280, and 404 nm are monitored. To evaluate the process, measure the HRP concentration in IBs, solubilization, after refolding and in the load as well as the pools from HIC chromatography. For the determination of the HRP concentration in IBs, dissolve the IB pellet in ultrapure water at 100 g L<sup>-1</sup> and remove 3  $\times$  100  $\mu\text{L}$ , centrifuge and discard the supernatant. Then solubilize these pellets in 1 mL sample preparation buffer (62 mM Tris, 7.5 M GdnHCl, 125 mM DTT), resuspend thoroughly and incubate for 30 min at RT with slight agitation, then centrifuge at 20,379 g; 15 min, 4 °C, and filter the supernatant using a 0.2  $\mu\text{m}$  syringe filter. For the sample preparation of the solubilize, use 100  $\mu\text{L}$  solubilize and 900  $\mu\text{L}$  sample preparation buffer, mix and then centrifuge at 20,379 g; 15 min, 4 °C, and filter the supernatant using a 0.2  $\mu\text{m}$  syringe filter. All other samples should just be filtered prior to HPLC analysis.

**3.7 SE-HPLC**

Determine the purity of the HRP samples using SE-HPLC with a BEH 200A SEC 1.7  $\mu\text{m}$  4.6  $\times$  300 mm, 3.5  $\mu\text{m}$  column. Run the method for 18 min at a flow rate of 0.3 mL min<sup>-1</sup> using 100% 20 mM phosphate buffer pH 6.8 as mobile phase. Column oven temperature should be set to 30 °C and the wavelengths 214, 280, and 404 nm are monitored.

**4 Notes**

1. To bring the components in solution faster, heat the water to 37 °C prior to mixing.
2. Add urea first and dissolve it in preheated water (37 °C), then add all other components.



3. Hemin is not soluble in water at neutral pH, 100 mM KOH has to be added to bring it in solution.
4. If the EtOH is denatured, the solution has to be filtered after dilution, as the denaturing agent precipitates at lower concentrations of EtOH.
5. Add TFA only after sonication, as volatile and highly toxic hydrogen fluoride is formed when TFA is heated or sonicated.
6. Due to the very high molarity of GdnHCl, the water has to be heated to 37 °C on a hot plate during mixing to dissolve the GdnHCl, be careful not to add too much water, as the 35.8 g already take up a considerable amount of volume in the beaker at a final volume of 50 mL. In addition, it is advisable to use high purity grade GdnHCl ( $\geq 99\%$ ) for this solution.
7. It is essential that no chunks of biomass remain and the suspension is homogeneous, as remaining particles can lead to severe damage of the homogenizer.
8. Thorough resuspension is recommended to increase the removal of impurities.
9. A short centrifugation step of 5 min is sufficient to pellet the IBs, while many impurities on the other hand are still in solution and can then be separated easily during washing.
10. The last wash should be performed with water to remove all excess urea, because this chemical impedes resuspension of the pellet after freezing at  $-20\text{ }^{\circ}\text{C}$ .
11. Again, thorough resuspension is key, at least 30–60 s are recommended.
12. Depending on the IB quality, the solubilize might be very viscous, turbid, and hard to separate from the pellet, if this is the case, the centrifugation time can be increased to 45 min. Avoid stirring up the highly viscous and turbid part of the solubilize, it is easiest to remove the supernatant with a serological pipette.
13. The dilution factor solubilize to refolding is 1:40, 35 mL solubilize will result in 30–35 mL supernatant after centrifugation depending on IB quality.
14. The IB quality is not completely constant during different fermentation runs, certain batch to batch variations have to be expected, therefore the redox system should be robust to cover these variations so that no adjustments of DTT or GSSG concentration have to be made. At pH 10, the redox system covers a wide range of DTT:GSSG concentrations, which ensures a consistent refolding yield.

15. The slow and continuous addition of the hydrophobic hemin solution ensures that protein and hemin aggregation are diminished, which provides more suitable conditions for refolding of HRP to its native structure.
16. Binding of HRP to the column is more efficient at pH 8.5 than at pH 10.
17. The column can also be smaller, a column volume of around 20–30 mL is sufficient for this batch size, depending on the amount of HRP per g IB.
18. During the chromatography, HRP can easily be identified at 404 nm using the VIS detector signal due to the fact that the cofactor hemin acts as a chromophore (absorption in the Soret band region).
19. A pH value of 5 was chosen for the assay as the enzyme activity is higher than at neutral pH and ABTS is less soluble and stable at pH values  $\geq 7$ .
20. It is very important to mix each well thoroughly after addition of hydrogen peroxide by pipetting up and down 2–3 times using a multichannel pipette.

## References

1. Krainer FW, Glieder A (2015) An updated view on horseradish peroxidases: recombinant production and biotechnological applications. *Appl Microbiol Biotechnol* 99:1611–1625. <https://doi.org/10.1007/s00253-014-6346-7>
2. Gundinger T, Spadiut O (2017) A comparative approach to recombinantly produce the plant enzyme horseradish peroxidase in *Escherichia coli*. *J Biotechnol* 248:15–24. <https://doi.org/10.1016/j.jbiotec.2017.03.003>
3. Krainer FW, Pletzenauer R, Rossetti L, Herwig C, Glieder A, Spadiut O (2014) Purification and basic biochemical characterization of 19 recombinant plant peroxidase isoenzymes produced in *Pichia pastoris*. *Protein Expr Purif* 95:104–112. <https://doi.org/10.1016/j.pep.2013.12.003>
4. Naatsaari L, Krainer FW, Schubert M, Glieder A, Thallinger GG (2014) Peroxidase gene discovery from the horseradish transcriptome. *BMC Genomics* 15:227. <https://doi.org/10.1186/1471-2164-15-227>
5. Childs RE, Bardsley WG (1975) The steady-state kinetics of peroxidase with 2,2'-azino-di-(3-ethyl-benzthiazoline-6-sulphonic acid) as chromogen. *Biochem J* 145:93–103. <https://doi.org/10.1042/bj1450093>
6. Humer D, Ebner J, Spadiut O (2020) Scalable high-performance production of recombinant horseradish peroxidase from *E. coli* inclusion bodies. *Int J Mol Sci* 21:4625. <https://doi.org/10.3390/ijms21134625>



## 7.2 Scientific paper: Recombinant Protein L

Journal of Biotechnology 359 (2022) 106–115



Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: [www.elsevier.com/locate/jbiotec](http://www.elsevier.com/locate/jbiotec)



### Recombinant Protein L: Production, Purification and Characterization of a Universal Binding Ligand

Stefan Kittler<sup>a,b,1</sup>, Julian Ebner<sup>a,b,1</sup>, Mihail Besleaga<sup>a</sup>, Johan Larsbrink<sup>c</sup>, Barbara Darnhofer<sup>d</sup>, Ruth Birner-Gruenberger<sup>d,e</sup>, Silvia Schobesberger<sup>f</sup>, Christopher K. Akhgar<sup>g</sup>, Andreas Schwaighofer<sup>g</sup>, Bernhard Lendl<sup>g</sup>, Oliver Spadiut<sup>a,1</sup>

<sup>a</sup> Research Division Integrated Bioprocess Development, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Gumpendorfer Strasse 1a, 1060 Vienna, Austria

<sup>b</sup> Alfred Gruber GmbH, Nordstrasse 6, 5301 Eugendorf, Austria

<sup>c</sup> Wallenberg Wood Science Center, Division of Industrial Biotechnology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

<sup>d</sup> Research Division Functional Proteomics and Metabolic Pathways, Diagnostic and Research Institute of Pathology, Medical University of Graz, Stiftgasse 24, 8010 Graz, Austria

<sup>e</sup> Research Division Bioanalytics, Institute of Chemical Technology and Analytics, TU Wien, Getreidemarkt 9/164, 1060 Vienna, Austria

<sup>f</sup> Research Division Organic & Biological Chemistry, Institute of Applied Synthetic Chemistry, TU Wien, Getreidemarkt 9/163, 1060 Vienna, Austria

<sup>g</sup> Research Division Environmental Analytics, Process Analytics and Sensors, Institute of Chemical Technology and Analytics, TU Wien, Getreidemarkt 9/164, 1060 Vienna, Austria

#### ARTICLE INFO

##### Keywords:

Protein L  
E. coli  
Recombinant production  
Bioreactor  
Affinity ligand

#### ABSTRACT

Protein L (PpL) is a universal binding ligand that can be used for the detection and purification of antibodies and antibody fragments. Due to the unique interaction with immunoglobulin light chains, it differs from other affinity ligands, like protein A or G. However, due to its current higher market price, PpL is still scarce in applications. In this study, we investigated the recombinant production and purification of PpL and characterized the product in detail. We present a comprehensive roadmap for the production of the versatile protein PpL in *E. coli*.

#### 1. Introduction

The affinity protein protein L (PpL) originates from *Finlayella magna* (formerly *Peptostreptococcus magnus*) and was identified by Myhre and Erntell as a membrane protein (Björck, 1988; Nilsson et al., 1992; Rosenthal et al., 2012). PpL has up to 5B (binding) domains, which selectively bind to kappa light chains of immunoglobulins (Igs) and, unlike protein A and G, do not interfere with the Fc region during binding (Kittler et al., 2021; Rodrigo et al., 2015; Zheng et al., 2012). In biotechnology, PpL is of high interest due to its ability not only to bind whole Igs, but also antibody fragments containing light chains, such as single chain variable fragments (scFvs) and fragments antigen binding (Fabs) (Kittler et al., 2021; Nilsson et al., 1993; Rodrigo et al., 2015;

Zheng et al., 2012). Different commercial versions (4 or 5B domains) are available, as the fifth binding domain has only minor effects on the binding affinity of the protein (Kittler et al., 2021). It was shown that PpL binds to kappa subtypes 1, 3 and 4, and is therefore applicable for more Ig classes compared to protein A and G (Nilsson et al., 1992; Rodrigo et al., 2015). However, all three proteins are currently used in downstream processing (DSP) and bioanalytics due to their binding abilities. Ig binding proteins enable the purification of high value products (i.e. antibodies and antibody fragments) in the pharmaceutical industry, where most of the processes use protein A, G or L affinity columns as a first chromatography capture step. Other applications encompass site-specific immobilization of Igs to maintain high functionality, including enzyme-linked immunosorbent assay (ELISA) and

\* Corresponding author.

E-mail addresses: [stefan.kittler@tuwien.ac.at](mailto:stefan.kittler@tuwien.ac.at) (S. Kittler), [julian.ebner@tuwien.ac.at](mailto:julian.ebner@tuwien.ac.at) (J. Ebner), [mihail.besleaga@tuwien.ac.at](mailto:mihail.besleaga@tuwien.ac.at) (M. Besleaga), [johan.larsbrink@chalmers.se](mailto:johan.larsbrink@chalmers.se) (J. Larsbrink), [b.darnhofer@medunigraz.at](mailto:b.darnhofer@medunigraz.at) (B. Darnhofer), [ruth.birner-gruenberger@tuwien.ac.at](mailto:ruth.birner-gruenberger@tuwien.ac.at) (R. Birner-Gruenberger), [silvia.schobesberger@tuwien.ac.at](mailto:silvia.schobesberger@tuwien.ac.at) (S. Schobesberger), [christopher.akhgar@tuwien.ac.at](mailto:christopher.akhgar@tuwien.ac.at) (C.K. Akhgar), [andreas.schwaighofer@tuwien.ac.at](mailto:andreas.schwaighofer@tuwien.ac.at) (A. Schwaighofer), [bernhard.lendl@tuwien.ac.at](mailto:bernhard.lendl@tuwien.ac.at) (B. Lendl), [oliver.spadiut@tuwien.ac.at](mailto:oliver.spadiut@tuwien.ac.at) (O. Spadiut).

<sup>1</sup> These authors contributed equally to the work.

<https://doi.org/10.1016/j.jbiotec.2022.10.002>

Received 21 August 2022; Received in revised form 28 September 2022; Accepted 1 October 2022

Available online 4 October 2022

0168-1656/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

immunoprecipitation (Bohinski, 2000; Shen et al., 2017). Furthermore, they are used as binding ligand for surface plasmon resonance (SPR) and biolayer interferometry (BLI) to determine binding kinetics of Igs and antibody fragments (Douzi, 2017; Kittler et al., 2021; Sultana and Lee, 2015).

However, the number of applications of PpL lag behind those of protein A and G, even though the use of PpL is the only viable alternative for binding antibody fragments missing the Fc region. To our knowledge, there is no literature describing the production of PpL available to date (process mode and conditions, purification, etc.), even though recombinant PpL versions can be purchased. Furthermore, the commercial price is approximately twice as high as for protein G and even three times higher than that of protein A (Kittler et al., 2021). We hypothesize that this is caused by the lower number of approved antibody fragments on the biopharmaceutical market compared to antibodies, reducing the need of PpL.

The goal of this study was to recombinantly produce SB domain His<sub>6</sub>-tagged PpL in *Escherichia coli* and test its functionality in comparison with a commercially available PpL with 4B domains (purchased from Sigma-Aldrich). In the upstream process (USP), product localization (inclusion body (IB) vs. soluble) and protein quantity were investigated by altering specific substrate uptake rate ( $q_s$ ) and temperature. The downstream processing aimed to obtain a high purity (>80%) without using an expensive affinity column. In this respect, the attachment of a His<sub>6</sub>-tag presents several advantages: It enables simple DSP, composed of Immobilized Metal Affinity Chromatography (IMAC), resulting in an efficient capture step for proteins produced in soluble form (Bornhorst and Falke, 2000; Ley et al., 2011). Furthermore, the tag enables the immobilization in a specific orientation which benefits subsequent applications such as electrochemical assays, SPR and lateral flow assays (Andrescu et al., 2001; Kröger et al., 1999; Ley et al., 2011). In the last step the purified PpL was characterized to investigate any potential negative impact of the His<sub>6</sub>-tag on the binding characteristics of the protein.

## 2. Material and Methods

### 2.1. Strain

The cultivations were performed using an *E. coli* BL21(DE3) strain transformed with a pET-24a(+) plasmid carrying the codon-optimized gene for the SB domain PpL (GenBank accession no. AAA67503) with a C-terminal His<sub>6</sub>-tag (restriction sites: *NheI/XhoI*). The encoded PpL has a theoretical size of 41.98 kDa and a theoretical pI of 4.82 (supplementary information: 1. Sequencing Result). The sequence for the SB domains originates from *F. magna* and was adapted from ProSpec (PROTEIN-1, 2019).

### 2.2. Media

The bioreactor cultivations were carried out using a defined medium described by DeLisa et al. (DeLisa et al., 1999), supplemented with 0.02 g/L kanamycin to prevent plasmid loss (Table 1). Glycerol was used as sole carbon source since recent results showed that this can result in higher product titers due to a higher amount of accessible energy (Kopp et al., 2017).

### 2.3. Bioreactor Cultivations

For pre-culture, 500 mL of DeLisa medium (Table 1) in a 2500 mL high yield shake flask were inoculated with 1.5 mL frozen bacterial stock. The pre-culture was grown at 37 °C and 230 rpm in an Infors HR Multitron incubator (Infors, Bottmingen, Switzerland) for 16 h. The cultivations were carried out in a DASGIP parallel reactor system (Eppendorf, Hamburg, Germany) with four vessels having 2 L working volume each. The culture broth was aerated with 2 L/min and stirred at

Table 1

DeLisa medium used for all cultivation phases. 0.02 g/L kanamycin were added for all phases.

Component	Pre-culture	Batch	Feed	Sterilization
	Concentration [g/L]			
Citric acid	13.3	–	–	autoclave
(NH <sub>4</sub> ) <sub>2</sub> HPO <sub>4</sub>	4	–	–	autoclave
Citric acid	1.7	–	–	autoclave
MgSO <sub>4</sub> ·0.7 H <sub>2</sub> O (stock 500 x)	1.2	–	10.00	autoclave
Fe(III) citrate (stock 100 x)	0.1	–	0.02	autoclave
EDTA (stock 100 x)	0.0084	–	0.0065	autoclave
Zn(CH <sub>3</sub> COO) <sub>2</sub> ·H <sub>2</sub> O (stock 200 x)	0.013	–	0.008	filter sterile
TE <sup>a</sup> (stock 200 x)	5 mL/L	–	7.27 mL/L	filter sterile
Thiamine HCl (stock 1000 x)	0.0045	–	–	filter sterile
Glycerol	8	20	400	autoclave

<sup>a</sup> TE stock: CoCl<sub>2</sub>·0.6 H<sub>2</sub>O (2.5 mg/L); MnCl<sub>2</sub>·0.4 H<sub>2</sub>O (15 mg/L); CuCl<sub>2</sub>·0.2 H<sub>2</sub>O (1.2 mg/L); H<sub>3</sub>BO<sub>3</sub> (3 mg/L); Na<sub>2</sub>MoO<sub>4</sub>·0.2 H<sub>2</sub>O (2.5 mg/L)

1400 rpm. The pH was monitored with an Easyferm electrode (Hamilton, Reno, NV, USA) and kept constant at 6.7 via addition of NH<sub>4</sub>OH (12.5%). The dissolved oxygen was monitored using a Visiferm fluorescence dissolved oxygen electrode (Hamilton, Reno, NV, USA) and kept above 30% by supplying a mixture of pressurized air and pure oxygen if necessary. Off-gas was monitored using a DASGIP-GA gas analyzer (Eppendorf, Hamburg, Germany). The temperature was controlled with a heat jacket and cooling finger and kept at 37 °C for the batch and at 35 °C during the non-induced fed-batch phase. For process control and monitoring the DAS-GIP-control system (DASware-control) was used. The batch phase (volume = 1 L) was started by inoculating the reactor with 10% (v/v) of the pre-culture. Once all glycerol was depleted, as indicated by a drop of the CO<sub>2</sub> signal and vice versa a rise in the dO<sub>2</sub> signal, substrate was fed to reach a cell dry weight concentration of approx. 25 g/L. After the fed-batch, expression of PpL was induced by addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM (induced fed-batch). During the induction phase, the specific substrate uptake rate ( $q_s$ ) and temperature were altered using a design of experiment (DoE) approach. These process parameters have been shown to be crucial factors for protein quantity and localization (IB vs. soluble) (Kopp et al., 2017; Slouka et al., 2018). During the induced fed-batch, samples were drawn every two hours to monitor product formation. The factor  $q_s$  was adjusted at the start of induction and was altered in the range of 0.1 g/g/h to 0.5 g/g/h. The temperature was investigated in a range of 27–35 °C, while the center-point conditions (CP) ( $q_s$  = 0.3 g/g/h; T = 31 °C) were performed three times to investigate reproducibility. We decided to perform a central composite circumscribed design to adequately describe potential quadratic interactions. In a first step the product localization and feasibility of a tunable production of IB or soluble PpL was investigated. For the design of DoE the volumetric product concentration in g/L was chosen as process response. The software MODDE 10 (Sartorius, Göttingen, Germany) was used for model design and to develop a multilinear regression model describing volumetric titer of PpL as a function of  $q_s$  and temperature over the induction time.

### 2.4. Bioreactor Cultivation Analytics

#### 2.4.1. Biomass

Biomass was quantified via optical density by measuring OD<sub>600</sub> using a UV/VIS photometer (Genisys 20, Thermo Scientific, Waltham, MA, USA) to monitor biomass growth during the process. Additionally, dry cell weight (DCW) was determined gravimetrically by centrifuging 1 mL of culture broth (9000 rcf, 10 min), subsequently washing the pellet with 0.9% (w/v) NaCl and centrifuging again. Afterwards the pellets were dried at 105 °C for 72 h.



### 2.4.2. Metabolite Analysis

The concentration of glycerol in the supernatant was analyzed via high performance liquid chromatography (HPLC) (UltiMate 3000; Thermo Fisher, Waltham, MA) using a Supelco C-610 H column (Supelco, Bellefonte, PA, USA) (Kopp et al., 2017). For this method 0.1%  $H_3PO_4$  with a constant flow rate of 0.5 mL/min was used as eluent. The sugars were detected and quantified using respective standards, by a refractive index detector (Shodex RI-101, Ecom, Prague, Czech Republic) (Kopp et al., 2017). The  $q_s$  was calculated according to Eq. 1, taking accumulated glycerol into account.

$$q_s \left[ \frac{g}{g \cdot h} \right] = \frac{V_{in,feed,\Delta t} [L] * c_{feed} \left[ \frac{g}{L} \right] - V_{reactor,\Delta t} [L] * c_{out,\Delta t} \left[ \frac{g}{L} \right]}{\Delta t [h] * X_{\Delta t} [g]} \quad (1)$$

- $q_s \left[ \frac{g}{g \cdot h} \right]$  ... specific substrate uptake rate.  
 $V_{in,feed,\Delta t} [L]$  ... feed volume in timespan  $\Delta t$ .  
 $c_{feed} \left[ \frac{g}{L} \right]$  ... glycerol concentration in the feed (400 g/L).  
 $V_{reactor,\Delta t} [L]$  ... reactor volume at time point i.  
 $c_{out,\Delta t} \left[ \frac{g}{L} \right]$  ... concentration of glycerol in the supernatant at time point i.  
 $\Delta t [h]$  ... timespan ( $t_i - t_{i-1}$ ) for the calculation of  $q_s$ .  
 $X_{\Delta t} [g]$  ... average biomass in the reactor in the timespan  $\Delta t$ .

### 2.4.3. Product Analysis

For determining the PpL concentration, 10 mL of the cultivation broth were centrifuged at 9000 rcf for 10 min at 4 °C. The supernatant was discarded and the product samples were stored at – 20 °C until further use. The cell pellet was suspended in 40 mL lysis buffer (100 mM Tris, 10 mM EDTA, pH = 7.4) and homogenized using a high-pressure homogenizer at 1200 bar for 7 passages (PandaPLUS, Gea AG, Germany). Subsequently, the cell suspension was centrifuged at 20,380 rcf for 20 min at 4 °C, and supernatant and IB pellet were further analyzed as described in the sections below. a) SDS-PAGE/For SDS-PAGE, Mini-PROTEAN® TGX Stain-Free™ (BioRad, Hercules, CA, USA) gels were used. The IB pellet were dissolved in 1x Laemmli buffer containing 100  $\mu$ L  $\beta$ -mercaptoethanol (reducing conditions) (Laemmli, 1970). The soluble fraction was mixed in a 1–2 ratio with 2x Laemmli buffer. All samples were incubated at 95 °C for 10 min and subsequently spun down. Five  $\mu$ L of a protein molecular weight standard (precision plus protein standard dual color, BioRad) and 10  $\mu$ L of each sample were loaded in the respective wells. The gel was run at 180 V for 30 min. Staining was performed using Coomassie brilliant blue and gels were analyzed using a Gel Doc (Universal Hood II, BioRad, Hercules, CA, USA) and the ImageLab software (Version 6.0.1, BioRad, Hercules, CA, USA). b) HPLC/Reversed Phase-HPLC: The product concentration was determined using a BioResolve reversed phase (RP) Polyphenyl column (dimensions 100  $\times$  3 mm, particle size 2.7  $\mu$ m) (Waters Corporation, MA, USA) equipped with a pre-column (3.9  $\times$  5 mm, 2.7  $\mu$ m) (Kopp et al., 2020). Eluent A was ultrapure water (MQ) and eluent B was acetonitrile, both supplemented with 0.1% (v/v) trifluoroacetic acid. A sample volume of 2  $\mu$ L was injected. The flow was kept constant at 0.4 mL/min and the measurement performed at 70 °C. A detailed description of the used method is given by Kopp et al. (Kopp et al., 2020). The soluble fractions were filtered with a 0.2  $\mu$ m syringe filter (CHROMAFIL® Xtra PVDF-20/25, Pall, New York, USA), while the IB pellets were first solubilized with 7.5 M guanidine hydrochloride, 62 mM Tris and 100 mM DTT at pH = 8 and filtered afterwards. To determine the protein concentration, BSA (bovine serum albumin) standards with concentrations between 0.1 and 2 g/L were measured. Size Exclusion Chromatography-HPLC: PpL concentration and purity during the DSP was measured using a size exclusion chromatography (SEC)-HPLC method. A BEH 200 A SEC 1.7  $\mu$ m 4.6  $\times$  300 mm, 3.5  $\mu$ m (Waters Corporation, MA, USA) column was run isocratically with SEC buffer (80 mM phosphate, 250 mM KCl, pH = 6.8) (Kittler et al., 2020). The method run time was 18 min with a flow rate of 0.5 mL/min and an

injection volume of 2  $\mu$ L was used. Column temperature was kept constant at 25 °C and absorbance was recorded at 280 nm and 214 nm. For quantification purposes, BSA protein standards were measured between 0.125 and 2 g/L and PpL concentrations calculated based on the calibration curve.

### 2.5. Downstream Processing

The biomass of each cultivation was harvested 12 h after induction, centrifuged at 17,000 rcf, 4 °C, 30 min and the biomass pellet stored at – 20 °C until further use. For the described DSP protocol, biomass produced at  $q_s = 0.3$  g/g/h and  $T = 31$  °C was used. For preparative chromatography, an ÄKTA pure™ system (Cytiva Life Sciences, MA, USA) was used, monitoring conductivity and UV absorbance at three wavelengths (280 nm, 260 nm, 214 nm).

The DSP comprised the following steps:

- Cell lysis via high pressure homogenization
- Capture via IMAC chromatography
- Purification via anion exchange chromatography (AEC)

#### 2.5.1. High Pressure Homogenization

In order to release intracellularly produced soluble PpL, cell lysis was performed via high pressure homogenization using a PandaPLUS (Gea AG, Germany). Frozen biomass was resuspended in Buffer A (50 mM phosphate, pH = 7.4) containing protease inhibitor (cOmplete™ Mini, EDTA-free, Roche, Switzerland) to a final concentration of 13 g DCW/L buffer A. Homogenization was performed at 1200 bar for 7 passages and the homogenized sample was kept at 4 °C afterwards. In order to separate cell debris from the soluble fraction containing PpL, the sample was centrifuged at 20,380 rcf for 20 min at 4 °C. The supernatant was used in the following chromatography step and the pellet discarded.

#### 2.5.2. Capture Chromatography: IMAC

As a capture step for the His<sub>6</sub>-tagged PpL, a 5 mL HiTrap™ IMAC FF (Cytiva Life Sciences, MA, USA) with a flowrate of 0.2 column volumes (CV)/min was used. The column was equilibrated with buffer A (50 mM phosphate, pH = 7.4) until all signals were stable. The supernatant (35 mL) after homogenization was loaded onto the column, followed by a wash step with 4 CVs of buffer A. Elution was performed using a step gradient with 40% buffer B (50 mM phosphate, 500 mM imidazole, pH = 7.4). During elution, fractions of 1 mL were collected, pooled based on the UV 280 nm signal and analyzed for their respective concentration and purity using SEC-HPLC.

#### 2.5.3. Purification Chromatography: AEC

As purification step, a 1 mL HiTrap™ Capto Q (Cytiva Life Sciences, MA, USA) column with a flowrate of 1 CV/min was used. The column was equilibrated with buffer A (50 mM phosphate, pH = 7.4). The pooled fractions containing PpL obtained from the capture chromatography step (IMAC) were used as load (6 mL load volume) (Cytiva, 2021a, 2021b). After loading was completed, the column was washed with 5 CVs buffer A. PpL was eluted using a step gradient with 25% buffer C (50 mM phosphate, 1 M NaCl, pH = 7.4). As for IMAC, fractions of 1 mL were collected, pooled based on the UV 280 nm signal and analyzed for their respective concentration and purity using SEC-HPLC.

### 2.6. Characterisation and Protein Functionality

#### 2.6.1. Mass Spectrometry

The primary structure and mass of the purified His<sub>6</sub>-tagged PpL was confirmed using digestion followed by LC/MS (liquid chromatography/mass spectrometry). Additionally, the total mass of the produced PpL and the commercial PpL (for comparison) was measured using intact protein mass spectrometry. A detailed description of the performed



measurements can be found in the [supplementary information \(supplementary information: 6. Mass spectrometry\)](#).

### 2.6.2. Infrared Spectroscopy

Both PpL variants were measured using laser-based mid-infrared spectroscopy to obtain structural information. A detailed description of the applied external cavity-quantum cascade laser (EC-QCL) setup has previously been reported (Akhgar et al., 2020). Briefly, the laser (Hedgehog, Daylight Solutions Inc., San Diego USA) was operated in a tuning range between 1470  $\text{cm}^{-1}$  and 1730  $\text{cm}^{-1}$  with a scan speed of 3600  $\text{cm}^{-1}$  and pulse rate and width of 1 MHz and 200 ns, respectively. The IR light was attenuated by optical filters, divided into two beams and directed into a two-path transmission flow cell with a path length of 26  $\mu\text{m}$ . Approximately 500  $\mu\text{L}$  of sample solution were injected into the signal cell, while the reference cell was filled with the pure buffer solution. The intensity of both beams was detected by a thermoelectrically cooled mercury cadmium telluride balanced detector (Vigo System S.A., Poland) to compensate the noise introduced by the EC-QCL. A previously described pre-processing routine (Schwaighofer et al., 2018), including similarity index evaluation, scan averaging (300 single scans = 45 s acquisition time) and fast Fourier transform (FFT) filtering, was applied in order to obtain the final protein spectra with a spectral resolution of 2.6  $\text{cm}^{-1}$ .

### 2.6.3. Surface Plasmon Resonance

Surface plasmon resonance spectroscopy (SPR) was performed at the core facility at the University of Natural Resources and Life Sciences in Vienna. For the measurement, a Bioacore T200 (Cytiva Life Sciences, MA, USA) was used. For measuring the binding affinity of both PpL samples, a commercial protein A chip was used. In the first step, a 10  $\mu\text{g}/\text{mL}$  Herceptin solution was applied to bind the antibody to the protein A chip. Subsequently, to monitor the binding affinity, different PpL concentrations (0.617, 1.85, 5.55, 16.6, and 50 nM respectively) were loaded for 10 min. The flow rate for all loading steps was 10  $\mu\text{L}/\text{min}$ . The general running buffer was HBS-EP buffer (Cytiva Life Sciences, MA, USA). The obtained RPU (response) was plotted versus the time and fitted with a one site saturation model (Eq. 1) using Sigma plot (Systat Software GmbH) to determine the response at equilibrium. The RPU at equilibrium of all measurements was then plotted against the respective concentration and fitted with a one site saturation equation to determine the  $K_D$  value (Eq. 2) (Moschetti et al., 2017; Sparks et al., 2019).

$$y = \frac{B_{\max} \cdot x}{K_D + x} \quad (2)$$

y ... y - value e.g. response of the SPR measurement  $B_{\max}$  ... maximal y-value x ... x - value e.g. concentration of the analyte  $K_D$  ... analyte concentration at  $B_{\max}/2$ .

### 2.6.4. Structure Analysis

Finally, the 3D structure of PpL was predicted using AlphaFold (Jumper et al., 2021; Varadi et al., 2022). For the analysis, the full amino acid sequence including the His6-tag was used and the prediction computed using the Alvis cluster within the Chalmers Centre for Computational Science and Engineering (C3SE), Sweden.

### 2.6.5. ELISA

To evaluate the functionality of the purified PpL, the purified protein was conjugated to recombinant horseradish peroxidase (HRP) and ELISA was performed. A detailed description of the protocol for conjugation, which was adapted from Nygren et al. and Molin et al. is given in the [supplementary information \(supplementary information: 2. ELISA Conjugation\)](#) (Molin So Fau - Nygren et al., 1978a, 1978b; Nygren et al., 1981).

For the ELISA, 100  $\mu\text{L}$  Herceptin (200  $\mu\text{g}/\text{mL}$ ) in phosphate buffered saline (PBS, pH = 7.4) was incubated at 4 °C in high binding ELISA

plates (Greiner Bio-One, Kremsmünster, Austria) for 16 h. Afterwards, the Herceptin solution was removed and the wells were washed four times with PBS containing 0.05% Tween20 (=wash buffer). Subsequently, to block open binding positions, the wells were incubated with 200  $\mu\text{L}$  1% BSA in PBS (=blocking solution) for 60 min at RT. Then the wells were washed again four times with wash buffer. Subsequently, the respective detection complex with a concentration of 0.2  $\mu\text{g}/\text{mL}$  was incubated at RT for 30 min (Table 2). The solution in the wells was discarded and wells were washed seven times with wash buffer to remove non-specifically bound HRP.

Bound HRP-PpL conjugates were quantified immediately after the last washing step using a 5,5'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) assay. 180  $\mu\text{L}$  of 8 mM ABTS in 50 mM phosphate-citrate buffer, pH = 5, were pipetted in each well and the reaction was started by adding 20  $\mu\text{L}$  10 mM  $\text{H}_2\text{O}_2$ . The change of absorbance at 420 nm was monitored for 45 min using a Tecan plate reader (Spark®, Tecan, Männedorf, Switzerland) at 30 °C. For each well, the volumetric activity was calculated according to Eq. 3.

$$A \left[ \frac{\text{U}}{\text{mL}} \right] = \frac{V_{\text{total}} [\mu\text{L}] \cdot \Delta A / \text{min} \cdot \text{dilution}}{V_{\text{sample}} [\mu\text{L}] \cdot d [\text{cm}] \cdot \epsilon [\text{mM}^{-1} \cdot \text{cm}^{-1}]} \quad (3)$$

$V_{\text{total}}$  [ $\mu\text{L}$ ] ... total volume in well.

$\Delta A / \text{min}$  ... change in absorption ( $\Delta A$  420 nm/min).

Dilution ... dilution of the sample.

$V_{\text{sample}}$  [ $\mu\text{L}$ ] ... volume of sample.

$d$  [cm] ... length of the beam path through the cuvette ( $d = 0.58$  cm).

$\epsilon$  [ $\text{mM}^{-1} \cdot \text{cm}^{-1}$ ] ... extinction coefficient ( $\epsilon_{420} = 36 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ ).

## 3. Results and Discussion

### 3.1. Production and Purification of Recombinant Protein L

The goal of this study was to develop a production process for recombinant PpL that enables production of a high amount of biologically active protein with high purity (>80%). The investigated variables in the USP and DSP are listed in Table 3.

For the recombinant production of His<sub>6</sub>-tagged PpL, the specific substrate uptake rate and temperature ( $q_s$ , T) were investigated in a DoE approach. In previous studies it was shown that these process variables influence localization (IB or soluble) and quantity of produced recombinant protein (Kopp et al., 2017; Slouka et al., 2018). For all tested fermentation conditions, an excess of soluble PpL compared to IBs (9 g/L to <0.5 g/L) was observed (supplementary information: Fig. S1). Since PpL originates from a bacterial host, we believe that expression has a low burden on *E. coli* due to low protein complexity (no disulfide bridges or other post-translational modifications) and therefore only low amounts of IB were produced (Bhatwa et al., 2021). As depicted in Fig. 1a, high specific substrate uptake rates ( $\geq 0.5$  g/g/h) led to an increase of the volumetric product titer in the beginning of the induction phase. However, after four to six hours, glycerol started to accumulate (data not shown) as the biomass growth stopped (DCW decreased) and the volumetric productivity declined. The observed drop in the specific growth rate correlated with the mentioned accumulation of glycerol, caused by a high metabolic stress using a  $q_s$  higher than 0.5 g/g/h. Lower temperatures (25.3 °C and 27 °C) did not increase the amount of

Table 2

Tested combination of samples and conjugation-complex. HRP: horseradish peroxidase; PpL: protein L; BSA: bovine serum albumin.

Immobilized protein	Detection complex
Herceptin	HRP
Herceptin	PpL
Herceptin	PpL-HRP
BSA	PpL-HRP



**Table 3**

Overview of investigated factors in the production process of recombinant protein L. USP: upstream processing; DSP: downstream processing;  $q_s$ : specific substrate uptake rate; IMAC: immobilized metal affinity chromatography.

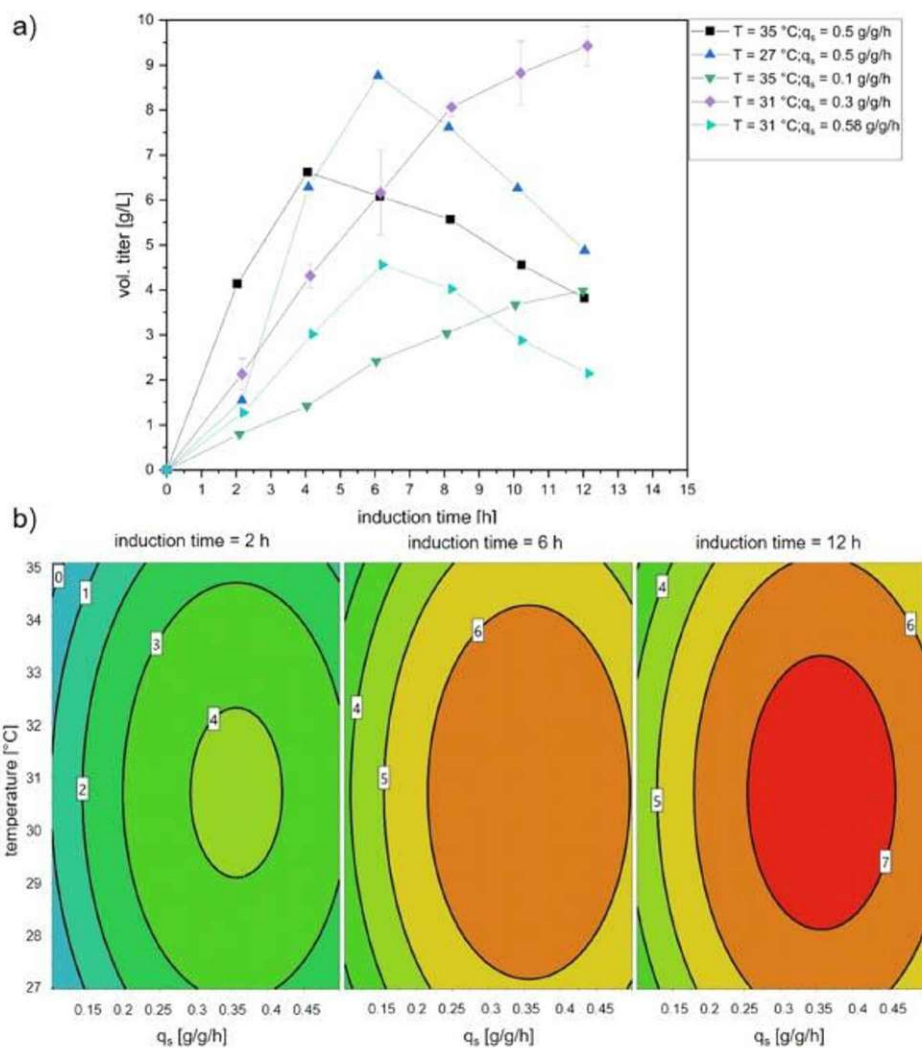
Unit operation	Factors	Range
USP	$q_s$	0.1 g/g/h – 0.5 g/g/h
	Temperature	27–35 °C
DSP - IMAC	NaCl	0 mM or 500 mM

PpL, independent of the adjusted  $q_s$ . The CP conditions ( $q_s = 0.3$  g/g/h, temperature 31 °C) resulted in an almost linear increase of the volumetric titer, while higher temperatures did not lead to an increase of

product formation.

The regression model for the volumetric titer showed that longer induction time had a positive influence on productivity (Fig. 1b). Furthermore, increasing temperature led to an increase of the volumetric titer, while the quadratic terms of all tested parameters had a negative impact on product formation (supplementary information: Fig. S2). The highest volumetric product titer (> 9 g/L) was achieved at a  $q_s = 0.3$  g/g/h, a temperature of 31 °C and an induction time of 12 h, leading to a specific product titer of 0.17 g/g with a final biomass concentration of 55 g DCW/L (process data: supplementary information: Figs. S3 and S4).

After cell lysis, the His<sub>6</sub>-tagged PpL was purified to achieve purities



**Fig. 1.** a: Time resolved trend of volumetric titer [g/L] of performed cultivations according to the design of experiment (DoE). b: Contour plot of the volumetric titer [g/L]. The optimum in the tested design space is at a specific substrate uptake rate ( $q_s$ ) = 0.3 g/g/h and temperature 31 °C after 12 h. These conditions led to a product concentration of about 9 g/L.

similar to the commercial PpL (i.e. 80%, determined by SEC). A standard protocol for IMAC, including 20 mM imidazole in the sample and the binding buffer, led to unexpectedly low recoveries (<30%, data not shown). In order to improve binding capacity and recovery, sample and binding buffers were prepared without imidazole. This adaptation resulted in improved recovery (59%) and purity (67%). Still, a second chromatography step was required to reach the targeted purity of > 80%. Due to the pI of 4.82, an anion exchange chromatography (AEC) step was chosen. However, the used salt (500 mM NaCl) in the IMAC buffers requires a buffer exchange to ensure sufficient binding in the subsequent AEC step. Therefore, we investigated whether NaCl can be omitted in the IMAC step without negative effects on recovery and purity. Summarized in Table 4, the final process using IMAC buffers without salt showed increased recovery and purity, resulting in a final PpL purity of 92% and a final yield of 5 g PpL / L fermentation broth (57%) after the second AEC step. As an additional advantage, the eluate of the capture step (IMAC) could directly be used to load the column in the second purification step, circumventing additional unit operations such as e.g. pH-adjustment or buffer exchange (Cytiva, 2021a, 2021b). The presented process might be a viable alternative to industrial production processes in which far more expensive Ig containing resins are used for the capture step (Cytiva). Furthermore, the purified 5B domain His<sub>6</sub>-tagged PpL had a higher purity (92%) compared to the commercial protein (80%). However, it is important to mention that DNA and endotoxin concentrations were not analyzed.

### 3.2. Protein Structure and Functionality Analysis

First, the primary structure of both PpL variants (TU Wien: 5B domain His<sub>6</sub>-tagged PpL and Sigma-Aldrich: 4B domain PpL) was confirmed using enzymatic digestion followed by LC/MS. However, due to the existence of repetitive B domains, the exact mass of the proteins could not be determined by analyzing the fragments. For this purpose, whole protein LC/MS was performed of each PpL variant. The 5B domain His<sub>6</sub>-tagged PpL, had a mass of 41816 Da (supplementary information: Fig. S5), while the commercial 4B domain PpL (one binding domain less) had a mass of 36038 Da (supplementary information: Fig. S6). Moreover, for both PpL variants, a distribution of different mass fragments was observed. This size heterogeneity has been reported for different Ig-binding proteins of Gram-positive bacteria (Kastern et al., 1992).

Laser-based mid-IR spectroscopy was applied to record absorbance spectra across the amide I (1600–1700 cm<sup>-1</sup>) and amide II (1500–1600 cm<sup>-1</sup>) bands, since these represent the most important wavenumber regions for protein secondary structure determination (Barth, 2007). Compared to conventional Fourier-transform infrared (FTIR) instrumentation, EC-QCLs operate at significantly higher spectral power densities, thus allowing the application of larger optical path lengths that lead to improved robustness and sensitivity (Schwaighofer and Lendl, 2020). Fig. 2a shows a comparison of PpL from Sigma-Aldrich (4B domains) and PpL produced here (TU Wien; 5B domains His<sub>6</sub>-tagged). Positions and shapes of the IR absorbance bands show excellent comparability between the spectra. The maxima at 1640 cm<sup>-1</sup> in the amide I region, as well as the broad amide II bands at approximately 1550 cm<sup>-1</sup> indicate a high share of  $\beta$ -sheet secondary structure (Barth, 2007).

**Table 4**  
Recovery and purity of the His<sub>6</sub>-tagged PpL comparing the IMAC runs with and without 500 mM NaCl.

Step	Recovery [%]	Purity [%]
Capture (IMAC) with NaCl	66	45
Whole DSP	44	72
Capture (IMAC) without NaCl	49	67
Whole DSP	57	92

For testing functionality and binding affinity, SPR measurements were performed to determine the K<sub>D</sub> value of both PpL variants. However, PpL has a very high affinity to Igs and antibody fragments, indicated by a small K<sub>D</sub> and a slow dissociation reaction. Therefore, the dissociation could not be determined with the used experimental set up, since no steady state was achieved (data not shown). As an alternative approach, the maximum signals for steady state conditions were calculated based on a one site saturation model. In Fig. 2b the maximum fitted response (RPU) for each tested concentration is plotted against the respective concentration (Moscetti et al., 2017; Sparks et al., 2019). The obtained values were fitted again using a one site saturation model to determine the K<sub>D</sub> values (Fig. 2b).

The determined K<sub>D</sub> value for the recombinant 5B His<sub>6</sub>-tagged PpL variant (K<sub>D</sub> = 2.93 × 10<sup>-10</sup> M) was lower than for the commercial 4B domain PpL (K<sub>D</sub> = 1.55 × 10<sup>-9</sup> M). This can likely be attributed to the additional B domain in the construct, which slightly increases the affinity (Kastern et al., 1992; Kittler et al., 2021).

After similarity of the secondary structure was validated for both variants and the functionality was demonstrated, the tertiary structure of the 5B His<sub>6</sub>-tagged PpL variant was modeled using AlphaFold (Jumper et al., 2021; Varadi et al., 2022). Since no amino acid sequence was available for the commercial PpL variant, a comparison of the modeled structures was not possible. However, the single B domains could be predicted with high confidence and are in accordance with literature (Kittler et al., 2021; Wikstroem et al., 1994). The linkers between the binding domains were found to have lower prediction scores and are most likely flexible and without a defined structure (Fig. 3, supplementary information: Fig. S7). These linkers as well as the unstructured N- and C- terminal ends of the protein (the latter incorporating the His<sub>6</sub>-tag) could be susceptible to some proteases, possibly explaining the size heterogeneity observed in mass analysis and the low recovery during the capture step (IMAC) (Shen et al., 1991).

### 3.3. Application of Protein L in an ELISA

For the purpose of using PpL in an ELISA, both PpL variants were conjugated to recombinant HRP and used to detect Herceptin. First, solely the HRP activity without the binding to Herceptin was measured. The recombinant His<sub>6</sub>-tagged 5B domain PpL conjugate showed an activity of 6.82 ± 0.27 × 10<sup>3</sup> U/mol, which was similar to the 4B domain conjugate PpL that showed an activity of 7.49 ± 0.91 × 10<sup>3</sup> U/mol. However, it was expected that the 5B domain PpL would exhibit a higher amount of conjugated HRP due to the higher number of primary amines (e.g. lysine), which act as conjugation partners. However, we assume that these additional primary amines were not accessible due to their position within the binding domains, resulting in similar amounts of conjugated HRP. In the ELISA, the His<sub>6</sub>-tagged PpL with 5B domains achieved a volumetric activity of 3.99 ± 0.86 U/mL, while for the commercial PpL (Sigma-Aldrich) with 4B domains showed an activity of 4.96 ± 0.92 U/mL. Conjugation of PpL to HRP without loss of activity and detection of an antibody was thus possible for both variants, showing that the tested HRP-PpL complex can be successfully used for ELISAs applied as in-process control e.g. in the early DSP or for refolding processes.

### 4. Conclusion

In this study we were able to successfully produce and purify recombinant His<sub>6</sub>-tagged PpL with 5B domains in *E. coli* with a final product yield of 5 g product per L fermentation broth. The produced protein was characterized and compared to a commercially available PpL regarding secondary structure and activity (Table 5). The initial DSP resulted in unexpectedly low recoveries and purities, but these could be significantly improved by optimization of the IMAC capture step and performing a second AEC step. The results from the performed MS indicate a size heterogeneity of PpL, which could possibly explain the



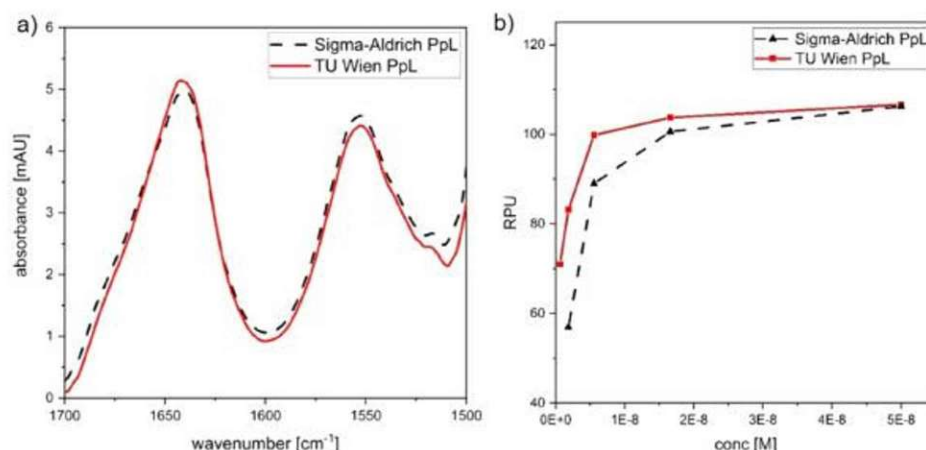


Fig. 2. a) IR spectrum of both protein L (PpL) variants. Both 4B domain and 5B domain His<sub>6</sub>-tagged PpL, from Sigma-Aldrich and TU Wien, respectively, show highly similar absorbance spectra indicating comparable secondary structure. b) Results of the SPR measurement of both protein L (PpL) variants. The calculated saturation values for each concentration is plotted against the respective concentration. The curve was fitted with a one-site saturation equation to determine the K<sub>D</sub> value. 4B domain PpL (Sigma-Aldrich PpL) K<sub>D</sub> = 1.5 × 10<sup>-9</sup> M; 5B domain His<sub>6</sub>-tagged PpL (TU Wien PpL): K<sub>D</sub> = 2.93 × 10<sup>-10</sup> M.

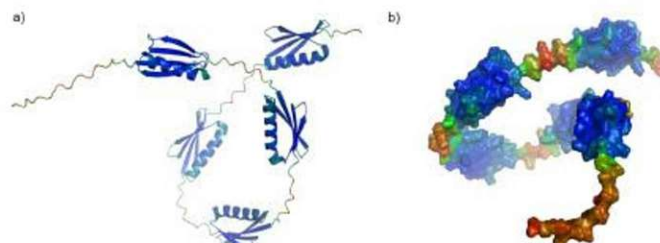


Fig. 3. a) Cartoon models colored by the prediction score, red-green-blue, from low to high, with the N-terminus on the left. b) Space-fill model of His<sub>6</sub>-tagged protein L, colored as in a), and with the N-terminus at the bottom. The structure of the B domains are confidently predicted, while the structure of the linkers and termini cannot be predicted reliably and are likely flexible in solution.

Table 5  
Summary of the results of the structural and functionality analysis of both protein L (PpL) variants. mAb: monoclonal antibody.

	TU Wien PpL	Sigma-Aldrich PpL
Mass [Da]	41816	36038
Structure of single B domain	4 beta sheets + 1 alpha helix	4 beta sheets + 1 alpha helix
B domains	5	4
K <sub>D</sub> [M]	2.93 × 10 <sup>-10</sup>	1.55 × 10 <sup>-9</sup>
Application in ELISA	Detection of mAb possible	Detection of mAb possible

still somewhat low overall recoveries. The structural analysis revealed unstructured protein parts that could be susceptible to cleavage, which would then result in different protein sizes and loss of the purification tag. We believe that further studies focusing on the structure of PpL will help to understand the observed results. Nevertheless, the secondary structure of the produced PpL was similar to the commercial variant with 4B domains. Furthermore, it was shown that the produced PpL is functional and active, without being negatively influenced by the His<sub>6</sub>-tag and the obtained K<sub>D</sub> value was lower compared to the 4B domain

PpL. Additionally, the application in an ELISA detecting Herceptin using a PpL-HRP conjugate was successful.

#### Funding

This research was funded by the Austrian Research Promotion Agency (FFG) (874206). The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program.

#### CRediT authorship contribution statement

SKI and JEB conducted most of the experiments and wrote the draft, MBE supported in the bioprocessing tasks, JLA did the protein model, BDA and RBG performed mass spectrometry, SSC assisted in the ELISA experiments, CAK, ASC and BLE did spectroscopic analyses, OSP initiated and supervised the study, secured funding and corrected the draft.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

Data will be made available on request.

## Acknowledgments

Alfred Gruber GmbH is gratefully thanked for supporting the research and being project partner. This project was further supported by EQ-BOKU VIBT GmbH and the BOKU Core Facility Biomolecular & Cellular Analysis. Furthermore, we thank Karolina Golab for performing the conjugation and ELISA experiments.

## Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jbiotec.2022.10.002.

## References

- Akbar, C.K., Ramer, G., Žbik, M., Trajnerowicz, A., Pawluczky, J., Schwaighofer, A., Lendl, B., 2020. The next generation of IR spectroscopy: EQ-QCL-Based Mid-IR transmission spectroscopy of proteins with balanced detection. *Anal. Chem.* 92, 9901–9907.
- Andrescu, S., Magearu, V., Lougarre, A., Pournier, D., Marty, J.L., 2001. Immobilization of enzymes on screen-printed sensors via an histidine tag. Application to the detection of pesticides using modified cholinesterase. *Anal. Lett.* 34, 539–540.
- Barth, A., 2007. Infrared spectroscopy of proteins. *Biochim. Et Biophys. Acta (BBA) - Bioenergy* 1767, 1073–1101.
- Bhatwa, A., Wang, W., Hassan, Y.I., Abraham, N., Li, X.-Z., Zhou, T., 2021. Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Front. Bioeng. Biotechnol.* 9, 65.
- Björck, L., 1988. Protein L: A novel bacterial cell wall protein with affinity for Ig L chains. *J. Immunol.* 140, 1194–1197.
- Bobinski, R.C., 2000. Immunoprecipitation of serum albumin with protein A-sepharose: a biochemistry laboratory experiment. *J. Chem. Educ.* 77, 1460.
- Bornhorst, J.A., Palke, J.J., 2000. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* 325, 245–254.
- Cytiva, (https://www.cytivalifesciences.com/en/at) (accessed 10 August 2021a).
- Cytiva, (2021b) Tips for successful ion exchange chromatography.
- Delias, M.P., Li, J., Rao, G., Weigand, W.A., Bentley, W.E., 1999. Monitoring GFP-operon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. *Biotechnol. Bioeng.* 65, 54–64.
- Douzi, B., 2017. Protein-protein interactions: surface plasmon resonance. In: Journet, L., Cascales, E. (Eds.), *Bacterial Protein Secretion Systems: Methods and Protocols*. Springer, New York, New York, NY, pp. 257–275.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romero-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodensteiner, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.
- Kastern, W., Sjöbring, U., Björck, L., 1992. Structure of peptostreptococcal protein L and identification of a repeated immunoglobulin light chain-binding domain. *J. Biol. Chem.* 267, 12820–12825.
- Kittler, S., Kopp, J., Voelenturf, P.G., Spadiut, O., Delvigne, F., Herwig, C., Slouka, C., 2020. The lazarus *Escherichia coli* effect: recovery of productivity on glycerol/lactose mixed feed in continuous biomanufacturing. *Front. Bioeng. Biotechnol.* 8, 993.
- Kittler, S., Beskanga, M., Ebner, J., Spadiut, O., 2021. Protein L—More Than Just an Affinity Ligand. *Processes* 9.
- Kopp, J., Slouka, C., Ulonska, S., Kager, J., Fricke, J., Spadiut, O., Herwig, C., 2017. Impact of Glycerol as Carbon Source onto Specific Sugar and Inducer Uptake Rates and Inclusion Body Productivity in *E. coli* BL21(DE3). *Bioeng. (Basel, Switz.)* 5, 1.
- Kopp, J., Zauner, P.B., Pell, A., Hausjell, J., Humer, D., Ebner, J., Herwig, C., Spadiut, O., Slouka, C., Pell, R., 2020. Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. *J. Pharm. Biomed. Anal.* 188, 113412.
- Kröger, D., Liley, M., Schiweck, W., Skerra, A., Vogel, H., 1999. Immobilization of histidine-tagged proteins on gold surfaces using chelator thioalkanes. *Biosens. Bioelectron.* 14, 155–161.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- Ley, C., Holtmann, D., Mangold, K.-M., Schrader, J., 2011. Immobilization of histidine-tagged proteins on electrodes. *Colloids Surf. B: Biointerfaces* 88, 539–551.
- Molin So Pau, -, Nygren, H., Nygren, H., Pau, -, Dolonius, L., Dolonius, L., 1978a. A new method for the study of glutaraldehyde-induced crosslinking properties in proteins with special reference to the reaction with amino groups. *J. Histochem. Cytochem.: Off. J. Histochem. Soc.* 26.
- Moscetti, I., Cannistraro, S., Bizzarri, A.R., 2017. Surface plasmon resonance sensing of biorecognition interactions within the tumor suppressor p53 network. *Sensors* 17, 2680.
- Nilson, B.H., Solomon, A., Björck, L., Akerström, B., 1992. Protein L from *Peptostreptococcus magnus* binds to the kappa light chain variable domain. *J. Biol. Chem.* 267, 2234–2239.
- Nilson, B.H., Löfdberg, L., Kastern, W., Björck, L., Akerström, B., 1993. Purification of antibodies using protein L-binding framework structures in the light chain variable domain. *J. Immunol. Methods* 164, 33–40.
- Nygren, H., Pau, -, Hansson, H.A., Hansson, H.A., 1981. Conjugation of horseradish peroxidase to staphylococcal protein A with benzoinquinone, glutaraldehyde, or periodate as cross-linking reagents. The journal of histochemistry and cytochemistry: official journal of the Histochemistry. *J. Histochem. Cytochem.: Off. J. Histochem. Soc.* 29, 266–270.
- PROTEIN L, HIS, (https://www.prospector.com/protein-l/his) (accessed 02.07.2019).
- Rodrigo, G., Gruvegård, M., Van Altine, J.M., 2015. Antibody fragments and their purification by protein L affinity chromatography. *Antibodies* 4.
- Rosenthal, M.E., Rojzman Ad Pau, -, Frank, E., Frank, E., 2012. *Finagoldia magna* (formerly *Peptostreptococcus magnus*): an overlooked etiology for toxic shock syndrome? *Med. Hypotheses* 79, 138–140.
- Schwaighofer, A., Lendl, B., 2020. Chapter 3 - quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In: Ozaki, Y., Baranska, M., Lednev, I.K., Wood, B.R. (Eds.), *Vibrational Spectroscopy in Protein Research*. Academic Press, pp. 59–88.
- Schwaighofer, A., Montemurro, M., Freitag, S., Kristament, C., Culzoni, M.J., Lendl, B., 2018. Beyond fourier transform infrared spectroscopy: external cavity quantum cascade laser-based mid-infrared transmission spectroscopy of proteins in the amide I and amide II region. *Anal. Chem.* 90, 7072–7079.
- Shen, H., Schmuck, M., Pilz, I., Gilkes, N.R., Kilburn, D.G., Miller, R.C., Warren, R.A., 1991. Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity. *J. Biol. Chem.* 266, 11335–11340.
- Shen, M., Rustling, J., Dixit, C.K., 2017. Site-selective orientated immobilization of antibodies and conjugates for immunodiagnosis development. *Methods* 116, 95–111.
- Slouka, C., Kopp, J., Hutwimmer, S., Strahammer, M., Strohm, D., Eitenberger, E., Schwaighofer, A., Herwig, C., 2018. Custom made inclusion bodies: impact of classical process parameters and physiological parameters on inclusion body quality attributes. *Microb. Cell Factor.* 17, 148–149.
- So Pau, M., Nygren, H., Nygren, H., Pau, -, Dolonius, L., Dolonius, L., Pau, -, Hansson, H.A., Hansson, H.A., 1978b. A kinetic study of the reaction between glutaraldehyde and horseradish peroxidase. *J. Histochem. Cytochem.: Off. J. Histochem. Soc.* 26.
- Sparks, R.P., Jenkins, J.L., Pratti, R., 2019. Use of surface plasmon resonance (SPR) to determine binding affinities and kinetic parameters between components important in fusion machinery. *Methods Mol. Biol.* 1860, 199–210.
- Sultana, A., Lee, J.E., 2015. Measuring protein-protein and protein-nucleic acid interactions by bilayer interferometry. *Curr. Protoc. Protein Sci.* 79, 19.25.11–19.25.26.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444.
- Wikström, M., Drakenberg, T., Forsen, S., Sjöbring, U., Björck, L., 1994. Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. *Biochemistry* 33, 1401–1407.
- Zheng, Z., Chinnaiyamy, N., Morgan, R.A., 2012. Protein L: a novel reagent for the detection of Chimeric Antigen Receptor (CAR) expression by flow cytometry. *J. Transl. Med.* 10, 29.



## Inclusion body-Prozessierung

Von der Wurzel ins Labor:  
Meerrettichperoxidase produziert in *E. coli*

JULIAN EBNER, DIANA HUMER, OLIVER SPADIUT  
IBD-GROUP; INSTITUT FÜR VERFAHRENSTECHNIK, UMWELTECHNIK  
UND TECHN. BIOWISSENSCHAFTEN, TU WIEN

**The enzyme Horseradish Peroxidase (HRP) is omnipresent in modern biotechnology. Although promising for therapeutic purposes, no suitable production process for this enzyme has been available until now. Medical applications require the enzyme to be highly pure, homogeneous and well-defined. We have developed an efficient production process for recombinant HRP from *Escherichia coli* inclusion bodies. With this strategy we are able to provide active, highly pure and non-glycosylated enzyme at competitive yields.**

DOI: 10.1007/s12268-021-1660-y  
© Die Autorinnen und Autoren 2021

■ Meerrettichperoxidase (*horseradish peroxidase*, HRP) ist ein in der Meerrettichwurzel vorkommendes Enzym aus der Klasse der Peroxidasen. Generell spielen Peroxidasen eine wichtige Rolle in der Pflanzenzelle, z. B. beim Zellwandmetabolismus, als Reaktion auf oxidativen Stress oder beim Wachstum von Wurzeln und Trieben [1]. Dennoch konnte die physiologische Funktion der HRP bis heute nicht vollständig geklärt werden. Hingänglich bekannt ist allerdings, dass HRP eine große Bandbreite aromatischer Verbindungen oxidiert. Diese Fähigkeit macht man sich seit Jahrzehnten in Forschung und Industrie zunutze, wo HRP vor allem als „Reporterenzym“ verwendet wird.

Bereits 1810 beschrieb Louis Antoine Planché das Auftreten einer intensiv blauen Färbung, sobald ein Stück Meerrettichwurzel mit Guajak-Tinktur benetzt wurde. Der Grund hierfür ist die im Wurzelgewebe vorhandene HRP, welche die Oxidation verschiedener in der Tinktur enthaltener Phenolverbindungen zu radikalischen Produkten katalysiert. Wie bereits bei der Guajak-Tinktur beobachtet, absorbieren viele Reaktionsprodukte der HRP sichtbares Licht.

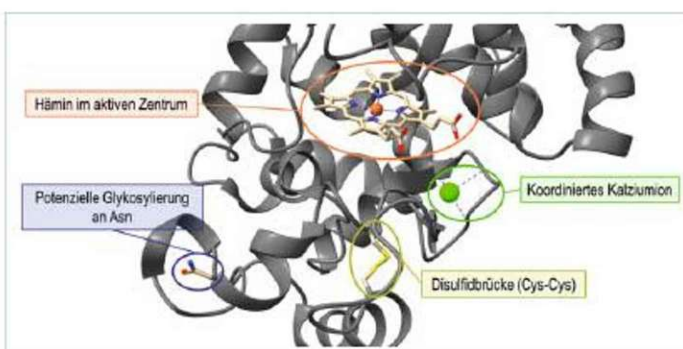
Dadurch ergibt sich eine einfach zugängliche Messgröße, die direkt mit der Menge an HRP im Reaktionsgemisch korreliert. An Nukleinsäuren, Antikörper oder andere Pro-

teine konjugiert, kann HRP so für Immunoassays, Mikroarrays und medizinisch-diagnostische Tests verwendet werden. Dies ermöglicht die Detektion und Mengenbestimmung verschiedenster Moleküle, Proteine und Virenpartikel [2]. In den letzten Jahren ist mit der spezifischen Krebsimmuntherapie (*targeted cancer treatment*) eine weitere Anwendungsmöglichkeit für HRP in den Fokus gerückt. Diese neue Therapieform bietet eine vielversprechende, nebenwirkungs-

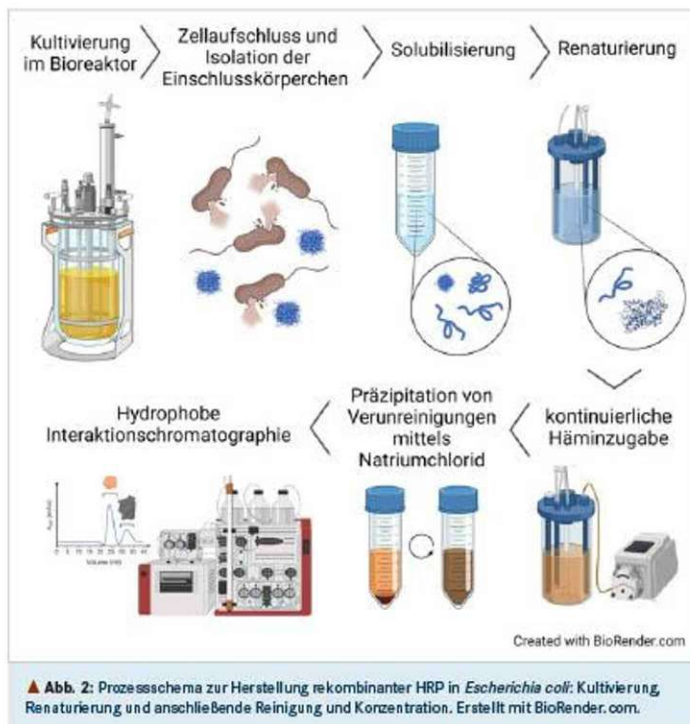
arme Alternative zu herkömmlichen Strahlen- oder Chemotherapien [3, 4].

Gerade für diese Anwendung ist die großtechnische Gewinnung durch Isolation aus der Meerrettichwurzel ein Problem, da das Enzym aufgrund seiner Zuckerketten (Glykosylierung) schnell aus dem Körper ausgeschieden wird oder gar eine Immunantwort beim Menschen auslöst [5]. Weitere Nachteile dieser Produktionsstrategie sind niedrige Ausbeuten und eine heterogene Mischung verschiedener Isoenzyme der HRP. Isoenzyme besitzen unterschiedliche biochemische Eigenschaften und sind damit für medizinische Anwendungen ungeeignet, da ein hochreines Enzym mit einer definierten Zusammensetzung benötigt wird. Aus diesem Grund ist die Entwicklung einer rekombinanten, biotechnologischen Herstellungsmethode für die HRP sehr erstrebenswert. Dazu wird die genetische Information des genau definierten Produkts gezielt in einen Wirtsorganismus eingebracht. Etablierte Wirtsorganismen sind dabei Säugetierzellen, Hefen oder das Bakterium *Escherichia coli*.

Dennoch existiert derzeit keine rekombinante Herstellungsmethode, welche kommerziell mit jener aus der Pflanze konkurrieren kann, obwohl z. B. Bakterien als Wirts-



▲ Abb. 1: Schematische Darstellung der wichtigsten Strukturmerkmale der Meerrettichperoxidase (*horseradish peroxidase*, HRP). In Summe enthält eine funktionsfähige Enzymeinheit ein Hämin, zwei Calciumionen, vier Disulfidbindungen und neun potenzielle Glykosylierungsstellen. Darstellung mittels UCSF Chimera [7].



organismen den Vorteil einer schnellen und kostengünstigen Kultivierung bieten [6]. Der Grund hierfür findet sich in der besonderen Struktur der HRP, welche beispielhaft in **Abbildung 1** gezeigt ist. HRP ist ein Metalloenzym, welches einen organischen Eisenkomplex, die Häm-Gruppe, im aktiven Zentrum trägt. Zudem benötigt das funktionsfähige Enzym zwei Calciumionen und vier Disulfidbindungen für strukturelle Stabilität. In der Pflanze werden zusätzlich noch an mehreren definierten Aminosäuren Zuckerketten angehängt, die Glykosylierung. Wird HRP nun, wie oben erwähnt, in *E. coli* produziert, so können manche posttranslationalen Proteinmodifikationen, wie z. B. die Glykosylierung, nicht durchgeführt werden. Dadurch kann sich das produzierte Protein in der Bakterienzelle nicht richtig falten und es bilden sich unlösliche Proteinaggregate. Diese Aggregate lagern sich in weiterer Folge zu Einschlusskörperchen (*inclusion bodies*) zusammen.

Somit erscheint die Herstellung in *E. coli* auf den ersten Blick als ungeeignet. Jedoch besteht die Möglichkeit, die unlöslichen Proteinaggregate *in vitro* in ihre native Faltung zu überführen. Dabei werden die Aggregate

in einem ersten Schritt denaturiert und dadurch in Lösung gebracht (Solubilisierung). Im Anschluss wird die chemische Umgebung so verändert, dass die einzelnen Proteinketten wieder in ihre native Konformation überführt werden (Faltung) [8]. Aufgrund der komplexen Struktur konnten während der *in vitro*-Faltung bisher allerdings nur niedrige Ausbeuten erreicht werden. Unter Berücksichtigung der besonderen Anforderungen dieses Enzyms haben wir kürzlich einen gesamtheitlichen Herstellungsprozess für die HRP entwickelt (Patent Nr. EP20165131.2) [9]. Die einzelnen Prozessschritte von der Fermentation bis zur Proteinreinigung sind in **Abbildung 2** gezeigt und können in drei Abschnitte eingeteilt werden:

1. Kultivierung von *E. coli*, Zellaufschluss und Isolation der *inclusion bodies* (Produktionsprozess)
2. Solubilisierung, *in vitro*-Faltung und Aktivierung durch Zugabe von Hämin (Renaturierungsprozess)
3. Abtrennung von Verunreinigungen und Aufkonzentrierung der HRP (Reinigungsprozess)

Im ersten Schritt werden rekombinante *E. coli*-Zellen in einem Bioreaktor kultiviert. Dies ermöglicht die Produktion von HRP in Form von *inclusion bodies* unter kontrollierten Bedingungen. Danach erfolgt der Zellaufschluss, um die in der Zelle lokalisierten und unlöslichen *inclusion bodies* freizusetzen. Diese werden anschließend mittels Zentrifugation von löslichen Verunreinigungen getrennt. Im zweiten Schritt, dem Renaturierungsprozess, wird nun aus den isolierten *inclusion bodies* wieder funktionsfähige und richtig gefaltete HRP gewonnen. Dazu wird das Protein während der Solubilisierung in eine lösliche Form überführt, wobei alle Disulfidbindungen vollständig geöffnet werden. Dies geschieht im Allgemeinen durch die Zugabe eines Reduktionsmittels. Im nächsten Schritt, der *in vitro*-Faltung, müssen die reduzierten Cysteine mithilfe eines Oxidationsmittels wieder zu Disulfidbindungen überführt werden [10]. In diesem Fall ist die Ausbeute der Faltungsreaktion nicht nur von der gewählten Konzentration des Oxidationsmittels abhängig, sondern zeigt auch eine Interaktion mit der gewählten Konzentration des Reduktionsmittels während der Solubilisierung. Um das funktionsfähige Enzym zu bilden, benötigt HRP außerdem eine Häm-Gruppe im aktiven Zentrum. Die Zugabe des Hämins zu Beginn der *in vitro*-Faltung resultiert jedoch in einer stark verringerten Ausbeute, da Fehlfaltungen der HRP gefördert werden. Aus diesem Grund wird Hämin im von uns entwickelten Prozess am Ende der *in vitro*-Faltung kontinuierlich über mehrere Stunden zugesetzt. Die dadurch erreichte geringe, aber konstante Hämin-Konzentration fördert den Einbau des Hämins ins aktive Zentrum und führt zu einer signifikanten Steigerung der Ausbeute. Trotz dieser Verbesserungen liegt ein Teil des Hämins am Ende des Prozesses ungebunden – und damit als Verunreinigung – vor. Daher erfolgt im letzten Schritt die Reinigung der korrekt gefalteten HRP. Hierfür wird zuerst überschüssiges Hämin und nicht korrekt gefaltete HRP mittels Salzfällung abgetrennt. Die noch in Lösung befindliche aktive HRP wird dann durch eine präparative Chromatographie (hydrophobe Interaktionschromatographie) von den verbleibenden Verunreinigungen getrennt und konzentriert.

Durch die integrierte Prozessentwicklung und die Berücksichtigung der speziellen Anforderungen wird eine Ausbeute von 960 Milligramm reiner (> 99 %), nicht glykosylierter HRP pro Liter Fermentationsbrühe



erreicht. Dies ist eine deutliche Erhöhung zu den bekannten Werten aus der Literatur (ca. 20 mg/L) [11]. Wie exemplarisch für die HRP gezeigt wurde, gelingt es mithilfe moderner Methoden der biotechnologischen Prozessentwicklung, selbst für komplexe Proteine effiziente Produktionsstrategien zu entwickeln. Dadurch werden sich auch in Zukunft herausfordernde und spannende neue Anwendungsgebiete für die rekombinante Proteinproduktion in *E. coli* ergeben.

#### Danksagung

Wir bedanken uns für die Unterstützung dieser Arbeit durch den Fonds zur Förderung der wissenschaftlichen Forschung (FWF, Projektnummer P30872-B26) und dem Stipendium der Monatshefte für Chemie (MoChem). Wir danken den Entwicklern der Programme UCSF Chimera und BioRender, welche für die Erstellung der Abb. 1 bzw. Abb. 2 verwendet wurden.

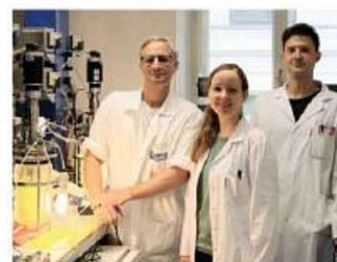
#### Literatur

- [1] Velich NC (2004) Horseradish peroxidase: a modern view of a classic enzyme. *Phytochemistry* 65: 249–259

- [2] Pandey VP, Awarthi M, Singh S et al. (2017) A comprehensive review on function and application of plant peroxidases. *Biochem Anal Biochem* 6: 308
- [3] Folkes LK, Greco O, Dachs GU et al. (2002) 5-Fluorouracil-3-acetic acid: a prodrug activated by a peroxidase with potential for use in targeted cancer therapy. *Biochem Pharmacol* 63: 265–272
- [4] Bonifert G, Folkes L, Gmeliner C et al. (2016) Recombinant horseradish peroxidase variants for targeted cancer treatment. *Cancer Med* 5: 1194–203
- [5] Brooks SA (2004) Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system. *Mol Biotechnol* 28: 241–256
- [6] Kraizer FW, Gledner A (2015) An updated view on horseradish peroxidases: recombinant production and biotechnological applications. *Appl Microbiol Biotechnol* 99: 1611–1625
- [7] Pettersen EF, Goddard TD, Huang CC et al. (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612
- [8] Eggenreich B, Willim M, Wurm DJ et al. (2016) Production strategies for active heme-containing peroxidases from *E. coli* inclusion bodies – a review. *Biotechnol Rep (Amst)* 10: 75–83
- [9] Humer D, Ebner J, Spadiut O (2020) Scalable high-performance production of recombinant horseradish peroxidase from *E. coli* inclusion bodies. *Int J Mol Sci* 21: 4625
- [10] Lange C, Rudolph R (2008) Oxidative Protein Folding in vitro. *BioSpektrum* 4: 376–378
- [11] Gazaryan IG, Chubar TA, Ignatenko OV et al. (1999) Tryptophanless recombinant horseradish peroxidase: stability and catalytic properties. *Biochem Biophys Res Commun* 262: 297–301

Funding note: Open Access funding enabled and organized by TU Wien (TUW). Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Verbreitung, Weiterleitung und Reproduktion in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und

angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.



Oliver Spadiut, Diana Humer und Julian Ebner (v. l. n. r.)

#### Korrespondenzadresse:

Prof. Dr. Oliver Spadiut  
IBD-Group  
Institut für Verfahrenstechnik, Umwelttechnik  
und Techn. Biowissenschaften  
TU Wien  
Gumpendorferstraße 1a  
A-1060 Wien  
[oliver.spadiut@tuwien.ac.at](mailto:oliver.spadiut@tuwien.ac.at)

# Hier steht eine Anzeige.



## JULIAN EBNER



e [REDACTED]



[REDACTED]



Vienna, Austria



julian-ebner

### PROFESSIONAL EXPERIENCE

**Senior Industry Relations Mng.** | 01/2024 - now  
**Industry Relations Manager** | 07/2023 - 01/2024  
 Funding Support & Industry Relations, TU Wien

Business development for R&D & innovation activities of TU Wien

- Primary contact point and stakeholder management for industry
- Strategic development of TU Wien's internal and external relations (coordinated projects within a matrix structure)
- Coordinated 6 strategic cooperations from introduction to ideation
- Managed 200+ industry request from SMEs and mid-caps
- Organised 15+ networking events
- Set up 5 partnerships between TU Wien and event organisers

**Project Assistant** | 12/2019 - 01/2023

**PhD Researcher** | 12/2019 - now

Alfred Gruber GmbH / TU Wien

Realised Novel Quality by Design strategies for industrial biopharmaceutical processes

- Planned, executed and managed 6 projects with industry leaders (e.g. Boehringer Ingelheim)
- Collaborated with 3 complementary academic research groups
- Developed production and analysis strategies (work flows and analytical tools)
- Supervised 3 BSc thesis and 3 MSc thesis student
- (Co-)authored 1 patent, 9 papers, 3 book chapters, and presented at 2 conferences

**Work exchange** | 05/2022 - 10/2022

EnginZyme AB - Sweden

Designed and executed research activities to meet industry project goals

- Defined project milestones and timelines, based on techno-economic models
- Data analysis (MODDE and ELN/LIMS)

**IAESTE Internship exchange** | 08/2018 - 09/2018

MUST - Mongolia

Food Research Center - Method development for the determination of vitamins in medicinal plants.

### MOTIVATION

Driven by technological challenges and innovation, I pursued an applied PhD in close collaboration with industry. I tackled parallel projects with my structured, results-driven, and meticulous attitude. With the motivation to bring solutions to market, I am currently bridging the gap between academia and industry by establishing strong partnerships. I thrive when strategising with the bigger picture in mind, building connections and realizing the full potential of people around me. In a team, I value open communication, loyalty, and social engagement.

### EDUCATION

**TU Wien** | 04/2017 - 02/2021

MSc in Technical Chemistry

Biotechnology and Bioanalytics

### SKILLS AND COMPETENCES

#### Competences:

- Collaboration in diverse and international teams
- Internal and external cooperations
- Project management
- Process Design
- Negotiation with Industry Partners
- Structured and strategic decision making
- Analytical and solution-oriented

#### Skills:

- Miro, ClickUp, Trello, Teams, Slack, etc.
- Customer relationship management (EPAM)
- LinkedIn presence and content creation

#### Workshops and trainings:

- Project Management for Chemical and Process Engineering
- Training course Technology Exploitation
- Effective writing for web editors (Typo3)
- Workshop: AI for Adobe Express
- Machine learning for medical informatics

#### Extracurricular activities:

- Jury Member for Innovation Funding of the Vienna Business Agency

#### Peer reviewer for scientific papers:

- Microbial Cell Factories (ISSN 1475-2859)
- Nature Scientific Reports (ISSN 2045-2322)

#### Languages:

- English (full professional)
- German (native)



## PATENTS

**Ebner, J., Humer, D., Spadiut, O.,** Patent No. EP20165131.2, Methods for producing heme peroxidases

## SCIENTIFIC PUBLICATIONS

**Ebner, J.; Sedlmayr, V.; Klausser, R. (2023).** High Pressure Homogenization for Inclusion Body Isolation. Inclusion Bodies. Methods in Molecular Biology, vol 2617. [https://doi.org/10.1007/978-1-0716-2930-7\\_9](https://doi.org/10.1007/978-1-0716-2930-7_9)

**Ebner, J.; Humer, D.; Sedlmayr, V. (2023).** Unit Operation-Spanning Investigation of the Redox System. Inclusion Bodies. Methods in Molecular Biology, vol 2617. [https://doi.org/10.1007/978-1-0716-2930-7\\_11](https://doi.org/10.1007/978-1-0716-2930-7_11)

**Humer, D.; Ebner, J. (2023).** The Purification of Heme Peroxidases from Escherichia coli Inclusion Bodies: A Success Story Shown by the Example of Horseradish Peroxidase. Inclusion Bodies. Methods in Molecular Biology, vol 2617. [https://doi.org/10.1007/978-1-0716-2930-7\\_16](https://doi.org/10.1007/978-1-0716-2930-7_16)

**Ebner, J.; Kittler, S.; Besleaga, M.; Larsbrink, J.; Darnhofer, B.; Birner-Gruenberger, R.; Schobesberger, S.; Akhgar, C. K.; Schwaighofer, A.; Lendl, B.; Spadiut, O. (2022).** Recombinant Protein L: Production, Purification and Characterization of a Universal Binding Ligand. J Biotechnol, 359, 108-115. doi:10.1016/j.jbiotec.2022.10.002

**Ebner, J.; Akhgar, C. K.; Alcaraz, M. R.; Kopp, J.; Goicoechea, H.; Spadiut, O.; Schwaighofer, A.; Lendl, B. (2022)** Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography. Anal Chem, 94(32), 11192-11200. doi:10.1021/acs.analchem.2c01542

**Ebner, J.; Akhgar, C. K.; Spadiut, O.; Schwaighofer, A.; Lendl, B. (2022)** QCL-IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography. Anal Chem, 94(14), 5583-5590. doi:10.1021/acs.analchem.1c05191

**Ebner, J.; Humer, D.; Spadiut, O. (2021)** Von der Wurzel ins Labor: Meerrettichperoxidase produziert in E. coli. Biospektrum 27, 773-775. <https://doi.org/10.1007/s12268-021-1660-y>

**Ebner, J.; Humer, D.; Klausser, R.; Rubus, V.; Pell, R.; Spadiut, O.; Kopp, J. (2021)** At-Line Reversed Phase Liquid Chromatography for In-Process Monitoring of Inclusion Body Solubilization. Bioengineering, 8, 78. doi.org/10.3390/bioengineering8060078

**Kittler, S.; Besleaga, M.; Ebner, J.; Spadiut, O. (2021)** Protein L—More Than Just an Affinity Ligand. Processes 2021,9, 874. doi.org/10.3390/pr9050874

**Kastenhofer J., Cataldo A.L., Ebner J., Sedlmayr V.L., Jungbauer A., Spadiut O. (2021)** Economic and ecological benefits of a leaky E. coli strain for downstream processing: a case study for staphylococcal protein A. Journal of Chemical Technology & Biotechnology, doi: 10.1002/jctb.6691

**Ebner, J., Humer, D., Spadiut, O., (2020)** Scalable high-performance production of recombinant horseradish peroxidase from E. coli inclusion bodies. International Journal of Molecular Sciences 21, 1-20. doi: 10.3390/ijms21134625

**Kopp, J., Zauner, F., Pell, A., Hausjell, J., Humer, D., Ebner, J., Herwig, C., Spadiut, O., Slouka, C., Pell, R., (2020)** Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. Journal of Pharmaceutical and Biomedical Analysis 188, doi: 10.1016/j.jpba.2020.113412

## CONFERENCES

**03/2022: Oral Presentation** at 7th International BioProScale Symposium: "Development of a scalable production for recombinant horseradish peroxidase"

**03/2022 Scientific Poster** at 7th International BioProScale Symposium: "Small scale mechanical cell disruption: A workflow to screen for ideal disruption conditions for recombinantly produced proteins in E. coli"  
Kittler, S.; Ebner, J.; Kopp, J.; Spadiut, O.; 2nd place in Best Poster Award

**02/2022 Conference Paper and Poster** at Proceedings of SPIE – The International Society for Optical Engineering: "Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography"  
Akhgar, C.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B.; DOI: 10.1117/12.2609419 and DOI: 10.1117/12.2609419

**11/2021 Oral Presentation** at International Conference on Analytical and Bioanalytical Methods: "At-line monitoring principles via reversed phase: A case-study for inclusion body solubilization"

**04/2019 Scientific Poster** at BioProcess International Europe 2019: "Optimization of a horseradish peroxidase inclusion body process in a refolding reactor vessel"  
Humer, D.; Ebner, J.; Spadiut O.;