MINI-REVIEW

Advances in monitoring and control of refolding kinetics combining PAT and modeling

Jan Niklas Pauk[1,2] · Janani Raju Palanisamy[1] · Julian Kager[1] · Krisztina Koczka[3] · Gerald Berghammer[3] · Christoph Herwig[1] (iD) · Lukas Veiter[1,2]

Abstract

Overexpression of recombinant proteins in *Escherichia coli* results in misfolded and non-active protein aggregates in the cytoplasm, so-called inclusion bodies (IB). In recent years, a change in the mindset regarding IBs could be observed: IBs are no longer considered an unwanted waste product, but a valid alternative to produce a product with high yield, purity, and stability in short process times. However, solubilization of IBs and subsequent refolding is necessary to obtain a correctly folded and active product. This protein refolding process is a crucial downstream unit operation—commonly done as a dilution in batch or fed-batch mode. Drawbacks of the state-of-the-art include the following: the large volume of buffers and capacities of refolding tanks, issues with uniform mixing, challenging analytics at low protein concentrations, reaction kinetics in non-usable aggregates, and generally low re-folding yields. There is no generic platform procedure available and a lack of robust control strategies. The introduction of Quality by Design (QbD) is the method-of-choice to provide a controlled and reproducible refolding environment. However, reliable online monitoring techniques to describe the refolding kinetics in real-time are scarce. In our view, only monitoring and control of re-folding kinetics can ensure a productive, scalable, and versatile platform technology for re-folding processes. For this review, we screened the current literature for a combination of online process analytical technology (PAT) and modeling techniques to ensure a controlled refolding process. Based on our research, we propose an integrated approach based on the idea that all aspects that cannot be monitored directly are estimated via digital twins and used in real-time for process control.

Jan Niklas Pauk and Janani Raju Palanisamy contributed equally to this work.

✉ Christoph Herwig
  christoph.herwig@tuwien.ac.at

1   Research Area Biochemical Engineering, Institute of Chemical, Environmental and Bioscience Engineering, Vienna University of Technology, Gumpendorferstrasse 1a/166, 1060 Vienna, Austria

2   Competence Center CHASE GmbH, Altenbergerstraße 69, 4040 Linz, Austria

3   Bilfinger Industrietechnik Salzburg GmbH, Mooslackengasse 17, 1190 Vienna, Austria

Introduction

Overexpression of proteins in *Escherichia coli* results in inclusion bodies (IBs) which generally are biologically inactive and need to be processed to yield an active, final product. However, this disadvantage has to be considered in view of several benefits of IBs like high purity, simple separation from cells and stability against mechanical, thermal, and proteolytic stress (Humer and Spadiut 2018). The native structure of proteins is recovered from IBs via solubilization followed by the refolding process. Solubilization is performed with denaturants in the presence of reducing agents. After the IBs have

been solubilized, protein refolding can be achieved by decreasing the amount of denaturant and providing a suitable environment for the protein to refold into its native structure (Jungbauer and Kaar 2007). This is a critical step toward the efficient recovery of proteins (Yamaguchi and Miyazaki 2014). During refolding, the denatured solubilized protein first forms its secondary structure as part of a self-folding process, which further leads to a final stable native tertiary or quaternary structure, depending on the protein. However, transient intermediates formed during the process may engage in nonspecific intermolecular interactions—primarily due to their exposed hydrophobic surfaces—resulting in aggregates and a major loss of yield (Mayer and Buchner 2004; Su et al. 2011). For simplification, a reduced view of this refolding process is useful consisting of only three distinct protein forms: solubilized protein S as the starting material for the refolding process, native protein N as the final product in a correct, and active form and aggregated protein A consisting of all folded but not active proteins including folding intermediates. These three defined protein forms will be referenced over the course of this review.

Numerous refolding methods are available such as dilution, dialysis, diafiltration, on-column refolding, or refolding by high hydrostatic pressure (Jungbauer and Kaar 2007; Middelberg 2002; Singh et al. 2015; Yamaguchi and Miyazaki 2014; Qoronfleh et al. 2007). Other approaches add additives to the refolding buffer to increase the refolding yield such as arginine or polyethylene glycol (Cleland et al. 1992; Kudou et al. 2011; Rathore et al. 2013). A recent and detailed summary of appropriate solubilization and refolding procedures can be found in the publication by Singhvi et al. (Singhvi et al. 2020). All these techniques aim to achieve a higher refolding efficiency while avoiding aggregate formation. Dilution is the traditional approach most widely applied on an industrial level (Jungbauer and Kaar 2007; Singh et al. 2015). In commercial applications, dialysis is rather time-consuming as it depends on the slow diffusion of ions and molecules. In addition, this may result in aggregate formation due to prolonged protein exposure at medium denaturant concentration (Cabrita and Bottomley 2004; Tsumoto et al. 2003). Diafiltration is shown to be an efficient system for protein refolding, not only to achieve higher yields than refolding in batch mode, but by reducing buffer consumption as well (Ryś et al. 2015a). However, fouling of membranes by aggregated proteins poses a problem (De Bernardez Clark 1998; Ryś et al. 2015a). The applied pressure in the high-pressure refolding method dissociates the existing protein aggregates and prevents the formation of further aggregates while requiring only small concentrations of chaotropic substances. Thus, this method can combine protein solubilization and refolding in one operation and is in most cases not restricted to low protein concentrations (Qoronfleh et al. 2007).

Despite the advantages of the aforementioned alternatives, the industry seems to be reluctant to abandon the extensively studied and well-established dilution method for protein refolding and thus it is also the main focus of this review. As part of this procedure, the solubilized IBs are mixed with refolding buffer at large volumes typically resulting in a 10- to 50-fold dilution of denaturants and a final protein concentration of 1–100 μg/mL (Pathak et al. 2016). From an industrial perspective, the method's simplicity, suitability for screening of additives, and easy implementation at various scales (Mirhosseini et al. 2019; Su et al. 2011) are advantageous. Also, refolding kinetics can be influenced independently of other effects, unlike during on-column refolding or diafiltration, which is an important aspect in view of the current industrial trend of increased process control. However, further processing of the refolded proteins requires concentration steps since the handling of low protein concentrations and high volumes is difficult and time intensive (Rathore et al. 2013; Singh and Panda 2005; Su et al. 2011). The dilution can be performed quickly and efficiently in two ways. As for the "dilution by batch mode," the solubilized IBs can be added directly to the refolding buffer as a single batch, which allows the denaturants and solubilized proteins to be diluted in a short time. However, this approach poses a higher risk of aggregation and misfolding due to the inefficient mixing at large reaction volumes and the formation of protein concentration centers while the solubilized protein is forced to reach its native structure quickly. The more effective alternative, "dilution by fed-batch mode," is to slowly dilute the solubilized IBs by gradually adding them to the refolding mixture either dropwise, in a pulsed way or as a constant, fed-batch-like addition (Cabrita and Bottomley 2004; Katoh and Katoh 2000). This method provides the denatured proteins with enough time for folding and thus prevents the forming of aggregations in the early folding pathway (De Bernardez Clark 1998).

For an efficient refolding process with the dilution by fed-batch mode, several parameters need to be considered such as denaturant concentration (De Bernardez Clark 1998; Dong et al. 2004; Tsumoto et al. 2003), protein concentration, mixing intensity, reaction temperature, and buffer components (Anselment et al. 2010; Eiberle and Jungbauer 2010) as well as dissolved oxygen and redox potential for proteins with disulfide bonds (Pizarro et al. 2009). Furthermore, the process needs to be adequately monitored to achieve good product quality. Consequently, possible measurement techniques are needed to assure process monitoring and product quality, as stated by the authorities in the PAT and QbD approaches (Food and Drug Administration 2004). The supporting use of modeling and simulation techniques in dilution refolding is regarded as a potent method to overcome the previously mentioned limitations of the process as well as limited available measurements (Dong et al. 2004; Jungbauer and Kaar 2007; Kiefhaber et al. 1991) and to generate platform

knowledge applicable for a range of IB products (Humer and Spadiut 2018). Full-state estimation in the form of a digital twin, e.g., through Kalman or particle filters, allows the control of the process in real time and thus would satisfy the industrial need for product quality assurance and process optimization besides providing a fault detection system.

In this review, we focus on the measurement, modeling, monitoring, and control of protein refolding processes to achieve optimal model-based control. The first section "Analytical methods of refolding processes" focuses on the timely assessment of refolding yield during dilution refolding, and "Models of refolding processes" describes current refolding models in the literature; in "Monitoring of refolding processes" and "Control of refolding processes" we will describe computational methods to control refolding kinetics in real time using modeling techniques including soft sensors and model-based control. Finally, we summarize our findings under "Proposal for a model-based refolding control strategy" including a detailed proposal for controlled kinetics of dilution refolding combining PAT and modeling according to the industrial need in agreement with the QbD guidelines.

Analytical methods of refolding processes

To successfully monitor refolding kinetics in a dilution refolding process, three distinct refolding species need to be analyzed timely and with high specificity: solubilized IBs, refolded protein, and protein aggregates. In the following, several analytical methods suitable for the monitoring of these aspects are discussed. Each method is applicable from the perspective of dilution refolding processes. For each technique, the general measurement procedure and the potential gain of information are covered, corresponding advantages and disadvantages are characterized as well. Finally, a summary of the recommended use and potential benefits of these methods is given.

Circular dichroism

Circular dichroism (CD) is used for the analysis of protein folding on the basis that folded and unfolded proteins show different spectra (Greenfield 2006). To study a dilution refolding process, a stopped-flow instrument with high-mixing capability can be combined with CD for online spectral analysis. As part of this method, the protein sample is first mixed with a denaturant to make it unfold, which is followed by a dilution step with the required buffer. Thus, with a low concentration of denaturant in the system, refolding may start and can be monitored by CD (Clarke 2012). In general, secondary structure information of proteins is obtained from the far UV range of their spectrum, as amide bonds absorb light at 180–250 nm. Additionally, the aromatic amino acids and

disulfide bonds provide information on the tertiary structure in the near UV ranges (250–320 nm) (Lindon et al. 2016).

The optimal protein concentration of a sample for CD analysis ranges from 0.2 to 1 mg/mL, depending on the path length of the applied cell that is normally between 0.01 and 0.05 cm (Greenfield 2006). The sample preparation involves a filtering step at 0.2 μm to obtain a homogenous solution that is free from scattering particles. Additionally, it is generally recommended to avoid a high concentration of chloride ions in sample buffers since they absorb strongly at wavelengths below 195 nm (Pelton and McLean 2000).

CD is a highly sensitive, robust, and non-destructive method suitable for the study of protein folding. Analysis can be carried out in a wide range of solvent environments with very small quantities of liquid (Kelly et al. 2005; Micsonai et al. 2015). However, optimal protein concentration ranges are rather high for refolding purposes (0.2 to 1.0 mg/mL) and a feedback time within minutes might be too long for process control, depending on data analysis of different spectra (Clarke 2012). Also, the nature of the transient refolding intermediates cannot be predicted completely using this technique.

Fluorescence spectroscopy

In fluorescence spectroscopy, the fluorescence intensity measured as a function of a wavelength is recorded during each stage of the refolding process to predict the conformations and structure of proteins (Printz and Friess 2012). Offline analysis can be conducted in a stopped-flow system where the reactants are swiftly mixed in a cuvette, while the change in fluorescence intensity is monitored over time (Ladokhin 2009; Lew et al. 1997; Qin and Pyle 1997).

Protein folding studies by fluorescence spectroscopy are based on the monitoring of fluorescence originating from fluorophores in the sample. Once the protein folds to form its tertiary structure, some of the fluorophores become covered in the inner hydrophobic environment resulting in high quantum yield and hence large fluorescence intensity. In comparison, in a partially folded or unfolded state, these compounds are exposed to the hydrophilic environment of the solvent leading to low fluorescence intensity. Hydrophobic interactions thus help to determine the conformation, solubility, or aggregation properties of a protein (Lamba et al. 2009). In case of intrinsic fluorescence, the signal originates from the naturally occurring aromatic amino acid residues tryptophan and tyrosine present in the sample (Moore-Kelly et al. 2019). Tryptophan and tyrosine are usually buried within the protein core in the native folded state and only get exposed to the hydrophilic environment of the solvent during a partially folded or unfolded state (Lakowicz 1988). However, when such intrinsic fluorophores are not present in the sample, an extrinsic fluorescence signal can be generated by the covalent

attachment of dyes to the protein in order to be able to monitor subtle changes in hydrophobicity. Dyes such as ANS (1-anilinonaphthalene-8-sulfonate) and Nile red have been used in refolding experiments for monitoring differences in surface hydrophobicity and for the detection of aggregates (Hawe et al. 2008; Pathak et al. 2016; Sutter et al. 2007).

Fluorescence spectroscopy is a sensitive method offering rapid data acquisition and sample analysis even at sub-nanomolar concentrations with a low feedback time within seconds (Jazaj et al. 2019; Ladokhin 2009). An optimal performance of the measurements is largely dependent on the selection of a suitable excitation wavelength according to the fluorophore. Optimal emission wavelengths need to be evaluated by a wavelength scan between 310 and 450 nm for tryptophan fluorescence, 400–600 nm for ANS, and 565–750 nm for Nile red (Hawe et al. 2008; Lamba et al. 2009). Drawbacks of this method include intrinsic fluorophores being limited to proteins containing tryptophan and tyrosine residues, while extrinsic fluorophores may alter not only the stability of the proteins but also the folding kinetics due to their covalent attachment (Pathak et al. 2016). Therefore, the use of extrinsic fluorophores might interfere with the refolding process itself making its use for process control potentially problematic.

Infrared spectroscopy

Fourier-transform infrared spectroscopy (FTIR) can be applied to predict the secondary structure of proteins by analyzing which wavelengths of radiation in the infrared region of the spectrum are absorbed by the sample (Tatulian 2019). Thereby, Fourier transformation enables the decomposition of a detector obtained time-domain spectrum into its constituent frequency domain spectra that can be easily interpreted (Griffiths 1983). With the inline use of a fiber active attenuated total reflection (ATR) probe, online analysis is possible during dilution refolding processes with high sensitivity and in real time. Offline monitoring is performed by manual sampling and analyzing of samples in a cuvette (Walther et al. 2014).

FTIR can illuminate the structural related changes during the refolding process and can be used to determine when to terminate the process to avoid protein loss due to aggregation. In comparison to CD, salt solutions are not problematic and turbid samples can be analyzed (Gregoire et al. 2012; Pelton and McLean 2000; Walther et al. 2014). Further advantages include low feedback time (within seconds) using online sample analysis, which resolves reproducibility problems that arise from offline sample preparation by other methods. Additionally, water can be used for solvent preparation since it gets eliminated as background noise in the resulting spectra. The minimum sample concentration for this method is 0.01 g/L (Humer and Spadiut 2018). The downside of ATR-FTIR is the

high sensitivity of probes to vibrations (Schuttlefield and Grassian 2008) and interference with common solubilization buffer components like urea or guanidinium chloride (Hauser 2013), as well as IR absorbance of water in the same range as proteins (Pathak et al. 2016).

Raman spectroscopy

Raman spectroscopy is used to analyze changes in the secondary structure of proteins (Brewster et al. 2013) and the formation of disulfide bonds (Wang et al. 2016). Traditional Raman spectroscopy is used to study protein solubilization (Brewster et al. 2013) and aggregation (Dolui et al. 2020) at protein concentrations of 1 g/L (Wen 2007). The method is very sensitive to small conformational changes (Brewster et al. 2013), is non-destructive, needs low to no sample preparation, and is insensitive to water (Bunaciu et al. 2015). However, solubilization and refolding buffer components show on the spectra and have to be subtracted by performing a reference analysis of the buffers (Brewster et al. 2013), especially with changing concentrations due to the refolding with dilution by fed-batch mode. Therefore, it might be applied for real-time monitoring of refolding processes.

Innovative alterations such as the surface-enhanced resonance Raman spectroscopy (SERS) enable measurements of concentrations as low as 0.08 g/L at feedback times of 5 min (Eryilmaz et al. 2017), whereas time-resolved resonance Raman spectroscopy (TR RR) allows the monitoring of solubilization and refolding of proteins at time resolutions of down to 100 µs (Buhrke and Hildebrandt 2020). However, both techniques are not suitable for real-time monitoring of a protein refolding process, because SERS requires the analyte to be adsorbed onto metal particles (Eryilmaz et al. 2017) and TR RR requires chromophores such as aromatic amino acids or cofactors (Buhrke and Hildebrandt 2020). Thus, the TR RR analysis is restricted to specific proteins and often reflects only local parts of these proteins (Buhrke and Hildebrandt 2020).

Nuclear magnetic resonance spectroscopy

In protein analysis, nuclear magnetic resonance spectroscopy (NMR) provides information on the molecular and geometric composition of covalent bonds, while additionally enabling assessment of non-covalent bonds between neighboring atoms (Wüthrich 2003). In refolding, this technique has been used to determine secondary structure in inclusion bodies (Umetsu et al. 2004), the structure of transient protein configurations during refolding (Dyson and Wright 2004), and more generally the formation of a protein's native, tertiary structure over time (Humer and Spadiut 2018; Ogura et al. 2013). NMR spectra of refolded protein samples at different refolding times in the form of chemical shifts may indicate completion of

refolding processes after formation of native protein structures (Pathak et al. 2016).

NMR is not suitable for routine real-time monitoring of refolding processes, as it requires protein purification and concentration prior to measurement (Ogura et al. 2013); the minimum sample concentration is rather high in the range of 0.1–3 mM (Kelly et al. 2005); additionally, the method is limited to protein sizes below 40–70 kDa (Frueh et al. 2013).

Light scattering techniques

Dynamic light scattering (DLS) is a measurement technique applicable in real time during refolding processes to monitor protein refolding and aggregation. Fluctuations in light scattering are detected to determine the hydrodynamic size of the particles (Yu et al. 2013). The radius of the protein is proportional to its folding state; thus, the smallest radius of the protein represents its completely folded form, while the largest corresponds to the unfolded state or denatured state. DLS detects aggregates based on the fact that the size of native proteins and aggregates varies considerably (Amin et al. 2014; Yasuda et al. 1998; Yu et al. 2013). A cuvette-based method can be used in an online capacity combining the measuring cell with automated sampling, similar to online flow cytometry approaches (Veiter and Herwig 2019). As an alternative to cuvette-based measurements, an integrated fiber optic probe can also be used. This approach has enabled online analysis of samples (Dhadwal et al. 1993).

With DLS, a sample range of 0.1–50 mg/mL can be estimated at a range of diameters from 1–2 nm to 3–5 μm. DLS is a fast and non-destructive technique, where quantification can be completed within a minute of sampling and samples can be re-used. This method has advantages over spectroscopic techniques, as it does not require a correlation between secondary and tertiary structures as in CD spectroscopy and it can be applied to all proteins, as there is no need for intrinsic and extrinsic fluorophores as in fluorescence spectroscopy (Dhadwal et al. 1993; Yu et al. 2013).

A drawback of this analysis is poor resolution due to a limited differentiation between particle species. Furthermore, the method is sensitive to the presence of particles in the refolding buffer, but this can be addressed by filtering the samples. Overall, DLS is a qualitative tool and not a quantitative one; however, Yu et al. showed a good correlation between DLS and SEC aggregation data which enabled quantitative statements (Den Engelsman et al. 2011; Yu et al. 2013).

Another powerful light scattering–based technique is multi-angle light scattering (MALS) which increases the robustness of the measurement by measuring the scattered light at multiple angles simultaneously and, thus, preventing to omit populations present in the sample (Naiim et al. 2015). This method can be used in combination with size exclusion chromatography or more recently ion exchange chromatography (Amartely et al. 2018) enabling the determination of molar masses of peaks separated by the chromatographic steps. Protein shape, aggregation, and oligomerization can be characterized as described by Machuca and Roujeinikova (2017) and Hemmig et al. (2005).

Reversed-phase HPLC

During the refolding process, chromatographic techniques can be used for the separation and quantification of folded and unfolded protein species to identify the current extent of refolding. Online applicability of this technique is possible using automated sampling and sample processing. For this purpose, sampling and potentially dilution need to be performed in a modular PAT system with a connected HPLC (Veiter and Herwig 2019). Samples need to be collected at different time points and analyzed based on the hydrophobicity of the respective proteins by reversed-phase HPLC (RP-HPLC) with an UV detector. As reduced proteins are completely unfolded resulting in an open structure, their hydrophobicity will be greater compared to that of native and oxidized (incorrect disulfide bondings) proteins (Cho et al. 2001; Choi et al. 2005; Pathak et al. 2016). If oxidized impurities occur, they typically display lower hydrophobicity than the native protein due to the differences in the disulfide linkages (Pathak et al. 2016; Rathore et al. 2013). However, different states of proteins not containing disulfide bonds cannot be separated effectively, because high temperatures of the method can destabilize the protein structure and consequently lead to unfolding of all protein states during the measurement (Pathak et al. 2016).

A robust RP-HPLC method is a powerful protein quantification system, with a limit of quantification (LOQ) lower than 10 μg/mL. The minimum sample concentration is 0.3 g/L assuming a sample volume of 2 μL. However, RP-HPLC is a time-consuming technique that can take 20–80 min for the high-resolution analysis of a sample. Consequently, a timely depiction of refolding kinetics can be problematic. Furthermore, overloading should be avoided by applying less than 1 mg per 1 mL of the column to have a better resolution of the resulting peaks (Humer and Spadiut 2018; Lindon et al. 2016; Živančev et al. 2015). Moreover, higher temperature during analysis improves recovery and plays a key role in selectivity and resolution, but it might induce aggregation of proteins during analysis (Hussain et al. 2019).

A major drawback of this method is the relatively large feedback time (several minutes) even when using a sampling device, such that other measurements are needed as information sources concerning monitoring and control of the process.

Size-exclusion HPLC

Further processing of samples using size-exclusion HPLC (SEC-HPLC) facilitates the monitoring of aggregates during

refolding. Thereby, a timely depiction of the changes in aggregate kinetics is possible (Choi et al. 2005; Pathak et al. 2016). SEC-HPLC separation is based on size, enabling the isolation of oligomeric aggregates. The native, correctly folded proteins have a more compact shape and size; therefore, they are distinguishable from unfolded and partially folded proteins (Cowan et al. 2008; Davidson 2008). Due to the flexibility and reproducibility of the process, SEC-HPLC is considered the standard process for measuring the aggregation of proteins (Hong et al. 2012).

In order to maintain optimal resolution and sensitivity, an automated sampling device—as previously mentioned—needs to control the sample volume, ideally as 5–10% of the total volume of the column (Cowan et al. 2008). Using an UV detector, lower wavelengths (214 or 220 nm) are optimal for the high sensitivity measurement of proteins present in low concentrations, while higher wavelengths (280 nm) enable the linear range detection of major species. With the dual-wavelength detection method, two wavelength ranges can be obtained. The wavelength ratio, which is the ratio of absorbance from two wavelengths, helps in the high sensitivity determination of the aggregate percentage (Hong et al. 2012; Printz and Friess 2012). Next to a UV detector, a fluorescence detector may also be used to enhance the selectivity and sensitivity of the method, and in general to further facilitate the measurement and quantification of protein content (Hong et al. 2012).

SEC-HPLC is a robust and sensitive analytical technique. Because of its high reproducibility and flexibility, it is a common approach for quantitative analysis of proteins (Amin et al. 2014; Hong et al. 2012). Potential drawbacks of the method include non-ideal interactions between large molecules and column packing materials, which might negatively affect the retention time, recovery, and peak shape (Hong et al. 2012). Similar to RP-HPLC, this method has long feedback times and is therefore problematic for monitoring and control of the process.

Summarizing the use of the presented analytical methods

Since the refolding of proteins is on a timescale of minutes to hours, real-time monitoring tools are preferred (Glassey et al. 2011). Some of the methods discussed in this section provide information on key aspects of the refolding process itself; Table 1 summarizes the characteristics of these techniques including information on concentration ranges, feedback time, and unwanted interaction with buffer components. CD and fluorescence spectroscopy are complementary spectral analysis techniques. With CD spectroscopy, the conformational changes of proteins during the refolding process can be followed, while fluorescence spectroscopy detects changes of the aromatic residues in the protein backbone (Reed et al.

2014). FTIR spectroscopy can resolve the limitations of CD spectroscopy regarding turbid and high-salt samples and has advantages over fluorescence spectroscopy as additional sample preparations can be avoided. Furthermore, FTIR can be applied for online monitoring of refolding processes (Walther et al. 2014).

To quantify the amount of protein and aggregates during refolding, SEC-HPLC and RP-HPLC are commonly used and robust techniques. Analysis can be carried out at multiple sampling points in near real time if an automated sampling device is available (Veiter and Herwig 2019). Also, DLS can be used to evaluate the formation of high-molecular-weight aggregates, but it is mainly used for qualitative measurements (Amin et al. 2014; Den Engelsman et al. 2011). FTIR and fluorescence spectroscopy can also be applied for aggregate monitoring, but the former method is not very sensitive at low aggregate concentrations and the latter might alter the behavior of the sample protein in case the binding of extrinsic fluorophores is necessary (Sutter et al. 2007; Walther et al. 2014).

There is no universal tool that can be applied to understand each aspect of the refolding process. To address this issue, a number of researchers employed a selection of techniques: Umetsu et al. studied secondary structure formation and tertiary structure by CD spectroscopy, explored folding by fluorescence spectroscopy using the shift in tryptophan emission, and performed FTIR for structural analysis of aggregated materials (Umetsu et al. 2003); Vincentelli et al. conducted CD spectroscopy to investigate protein folding and DLS for protein aggregation (Vincentelli et al. 2004); Cowan et al. used CD spectroscopy and SEC-HPLC to analyze protein refolding, and to study the protein's multimeric state (Cowan et al. 2008); Pathak et al. performed RP-HPLC for disulfide linkage analysis accompanied by SEC-HPLC for the study of aggregates and carried out CD spectroscopy for the investigation of secondary structure (Pathak et al. 2016). To address the problem of long feedback times when using RP-HPLC, Pathak et al. (2016) also performed zeta potential analysis: this technique may be used to monitor initial refolding stages involving primary structure, thereby enabling a relatively low analysis time within minutes.

If the monitoring of refolding kinetics is possible, critical process parameters (CPPs) can be defined. Since different analytical methods differ not only in robustness, sensitivity, and accuracy but also in real-time measurement capabilities, the additional use of modeling is a promising possibility to enhance and align several methods. Some of the monitoring techniques described do not feature sufficiently low feedback times for process control. However, these techniques nevertheless provide process knowledge that can be used in modeling approaches which will be discussed in section "Models of refolding processes." For instance, solubilized IBs and native protein could be quantified offline through SEC-HPLC

Table 1 Summary of the analytical methods discussed in this section. Online analysis is defined as automated sampling connected to the process followed by timely evaluation. Offline analysis is defined as manual sampling, typically followed by discontinuous sample preparation, measurement, and evaluation

Techniques	Type of analysis	Target measurement	Concentration range	Advantages	Disadvantages	References
CD	Online Offline	Folded/unfolded protein and intermediates: secondary structure, tertiary structure	Optimal range: 0.2 to 1.0 mg/mL Far UV-CD: 0.25 mg/mL Near UV-CD: 2.5 mg/mL	• Highly sensitive • Reliable • Accurate • Non-destructive • Wide range of solvent environments can be used	• No prediction of transient intermediates • No information on residues • High protein concentration needed, problematic for refolding processes • Feedback/data analysis time (several minutes) of CD spectra might be too long for process control • Interference with buffers with high ionic strength	(Clarke 2012) (Greenfield 2006) (Kelly et al. 2005) (Lindon et al. 2016) (Micsonai et al. 2015) (Pathak et al. 2016)
Fluorescence spectroscopy	Online Offline	Folded/unfolded protein: tertiary structure, partially folded or unfolded state	0.015 mg/mL depending on protein size	• Sensitive • Performed at sub-nanomolar concentrations • Rapid data acquisition • High signal-to-noise ratio • Low feedback time (seconds), applicable for process control	• Intrinsic fluorophores are limited to proteins containing tryptophan and tyrosine residues • Extrinsic fluorophores alter stability of the proteins and affect kinetics of protein folding due to covalent attachment to the protein	(Hawe et al. 2008) (Ladokhin 2009) (Lamba et al. 2009) (Lew et al. 1997) (Pathak et al. 2016)
FTIR	Online Offline	Folded/unfolded protein: structural related changes during refolding	> 0.01 mg/mL	• Use of salt buffers possible • Analyses of turbid samples • Direct sample analysis resolving reproducibility issues • Low feedback time (seconds) applicable for process control	• Probes sensitive to vibrations • Interaction with common solubilization buffer compounds like urea and guanidinium chloride • Changing background buffer signals due to dilution refolding by fed-batch mode	(Pathak et al. 2016) (Pelton and McLean 2000) (Tatulian 2019) (Walther et al. 2014)
Raman spectroscopy	Online Offline	Folded/unfolded protein: secondary structure, disulfide bond formation/tertiary structure	> 1 mg/mL SERS: > 0.08 mg/mL	• Non-destructive • No sample preparation • Insensitive to water • Sensitive to small conformational changes • Very low feedback times (TR RR)	• Interaction with common solubilization buffer compounds like urea and guanidium hydrochloride • Changing background buffer signals due to dilution refolding by fed-batch mode • Feedback times of few minutes • TR RR restricted to proteins containing chromophores	(Brewster et al. 2013) (Buhrke and Hildebrandt 2020) (Bunaciu et al. 2015) (Eryilmaz et al. 2017) (Wang et al. 2016) (Wen 2007)
NMR	Offline	Formation of native, tertiary structure over time	> 1 g/L 0.1–5 mM	• Information on tertiary protein structure	• Protein purification and concentration necessary • Limited to protein sizes below 50 kDa	(Frueh et al. 2013) (Kelly et al. 2005) (Pathak et al. 2016)
DLS MALS	Offline Online	Protein refolding and aggregation: radius of protein is proportional to its folding state, size of aggregates and native protein varies considerably	> 0.05 mg/mL	• Nondestructive • Does not require correlation between secondary and tertiary structure • No need for intrinsic and extrinsic fluorophores • Low feedback time (seconds) applicable for process control	• Primarily a qualitative method • Complicated • Poor resolution • Sensitive to particle presence in refolding buffer	(Amartely et al. 2018) (Amin et al. 2014) (Dhadwal et al. 1993) (Hemmig et al. 2005) (Machuca and Roujeinikova 2017) (Yasuda et al. 1998) (Yu et al. 2013)
RP-HPLC	Online Offline	Solubilized protein and native protein, loss through aggregation can be estimated	> 0.01 mg/mL	• Robust • High resolution	• High temperature induces aggregation of proteins	(Cho et al. 2001) (Choi et al. 2005) (Pathak et al. 2016)

Table 1 (continued)

Techniques	Type of analysis	Target measurement	Concentration range	Advantages	Disadvantages	References
				• Quantification of solubilized, native and misfolded protein states if the protein contains disulfide bonds	• Only quantification of soluble protein species possible if the protein does not contain disulfide bonds • Feedback time (several minutes) problematic for use in process control • Unwanted effects with components of refolding buffer possible (e.g., PEG)	(Rathore et al. 2013)
SEC-HPLC	Online Offline	Aggregates and native protein	> 0.01 mg/mL	• High reproducibility • Sensitive • Robust	• Non-ideal interactions between large molecules and column packing materials • Feedback time (several minutes) problematic for use in process control	(Hong et al. 2012) (Pathak et al. 2016) (Printz and Friess 2012)

and RP-HPLC for model refinement, while the refolding process itself can be monitored in real-time through detecting native protein via FTIR and aggregation through DLS or combinations of the aforementioned methods. Another problem arises for the spectroscopic methods with changing reference spectra due to the feeding of the dilution by fed-batch mode. Therefore, updated measurements of the reference spectrum are necessary.

Models of refolding processes

A model describes the change of a dynamical system over time. Therefore, it enables predictions of the system behavior into the future with given initial conditions. In the case of refolding processes, a model can be utilized to estimate the targeted product forms from other measurements; hence, it functions as a state observer providing additional indirect measurements of otherwise unmeasurable states. Mechanistic models describe a dynamical system with one or a set of differential equations. They are derived from physical first principles such as mass, momentum, heat, or energy balances (Gernaey et al. 2010; Wechselberger et al. 2013). A white box model completely describes a process with first principles and known parameters (Sohlberg and Jacobsen 2008); thus, for this type of model, no process data is necessary. However, if parameters or the underlying differential equations are unknown, data is needed to develop a model. Unknown model parameters in a protein refolding process are the reaction kinetics, which must be identified by fitting the process model on experimental data.

In the following, representative models for protein refolding and techniques to analyze their applicability are described.

The mechanism behind the folding of proteins is of great interest and many research groups tried to describe it using a variety of models (Cleland et al. 1992; Dong et al. 2004; Jungbauer and Kaar 2007; Kiefhaber et al. 1991; Zettlmeissl et al. 1979; Ryś et al. 2015a). The general structure of these models resembles one of the three classical models shown by Dill and Chan (Dill and Chan 1997), namely

the off-pathway model,

$$I \leftrightarrow S \leftrightarrow N \tag{1}$$

the on-pathway model,

$$S \leftrightarrow I \leftrightarrow N \tag{2}$$

and the sequential model,

$$S \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow \cdots \leftrightarrow N \tag{3}$$

where S corresponds to the solubilized protein, I to the folding intermediates, and N represents the native protein. However, an important aspect missing in these models is the fraction of aggregated proteins, which results from the solubilized protein or either of the folding intermediates (Kiefhaber et al. 1991). Hevehan and de Bernardez Clark described a simplified model with an on-pathway folding intermediate and off-pathway aggregation (Hevehan and de Bernardez Clark 1997) as depicted in Fig. 1.

The reaction of the solubilized protein to the folding intermediates is considered to happen immediately, simplifying the model to the off-pathway model with the two reaction rates k_r and k_a (Hevehan and de Bernardez Clark 1997). The reaction of correct folding follows a first-order reaction while the aggregation follows a reaction with an order of two or higher (Dong et al. 2004; Hevehan and de Bernardez Clark 1997; Kiefhaber et al. 1991; Zettlmeissl et al. 1979), or first order for low protein concentrations (Pan et al. 2015). In a simple dilution by batch mode refolding process, Kiefhaber et al. described the concentration of the solubilized and native protein with a second-order aggregation reaction as follows (Kiefhaber et al. 1991),

$$\frac{dc_{SL}}{dt} = -\left(k_r \cdot c_{SL} + k_a \cdot c_{SL}^2\right) \tag{4}$$

$$\frac{dc_{NL}}{dt} = k_r \cdot c_{SL} \tag{5}$$

Following this scheme, an equation for the aggregated protein can be described as,

$$\frac{dc_{AL}}{dt} = k_a \cdot c_{SL}^2 \tag{6}$$

with:

c_{SL} concentration of solubilized protein in the refolding vessel [g L^{-1}]

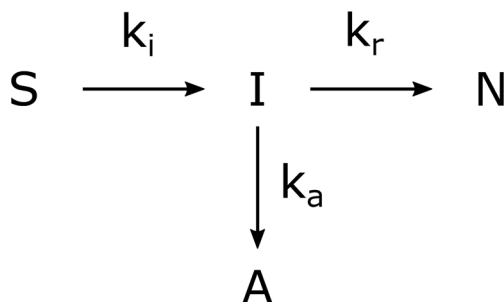c_{NL} concentration of native protein in the refolding vessel [g L^{-1}]

Fig. 1 Model depicting a refolding reaction scheme with on-pathway folding intermediate and off-pathway aggregation (Hevehan and de Bernardez Clark 1997). S corresponds to the solubilized protein, I stands for the folding intermediates, A for the protein aggregates and N represents the native protein. k signifies the corresponding reaction rate

c_{AL} concentration of aggregated protein in the refolding vessel [g L^{-1}]

k_r reaction rate for refolding [h^{-1}]

k_a reaction rate for aggregation [L g^{-1} h^{-1}]

The refolding yield is defined as the percentage of the native protein to the total protein in the system. Due to the higher order aggregation constant, the reaction favors aggregation with higher protein concentration, thus decreasing the refolding yield. To counter the aggregation, refolding processes are diluted; however, strong dilutions are economically not feasible in large-scale production (Hevehan and de Bernardez Clark 1997). By employing analytical techniques described in "Analytical methods of refolding processes," solubilized IBs can be determined prior to refolding through RP-HPLC while the refolding process can be monitored in real time through detecting native protein via FTIR and aggregation through DLS.

With the introduction of the dilution by fed-batch mode refolding, the existing equations are extended and the volume and denaturant concentration are included in the model (Dong et al. 2004). Due to the feeding of solubilized protein from the reservoir, the volume of the refolding vessel,

$$\frac{dV_L}{dt} = F_R \tag{7}$$

and the concentration of the denaturant,

$$\frac{dc_{DL}}{dt} = \frac{F_R}{V_L} \cdot c_{DR} - \frac{F_R}{V_L} \cdot c_{DL} \tag{8}$$

change over time. Assuming the reservoir consists solely of protein in the solubilized form, Eq. 4 extends with the incoming feed and dilution,

$$\frac{dc_{SL}}{dt} = -\left(k_r \cdot c_{SL} + k_a \cdot c_{SL}^2\right) + \frac{F_R}{V_L} \cdot c_{SR} - \frac{F_R}{V_L} \cdot c_{SL} \tag{9}$$

Eqs. 5 and 6 extend with dilution,

$$\frac{dc_{NL}}{dt} = k_r \cdot c_{SL} - \frac{F_R}{V_L} \cdot c_{NL} \tag{10}$$

$$\frac{dc_{AL}}{dt} = k_a \cdot c_{SL}^2 - \frac{F_R}{V_L} \cdot c_{AL} \tag{11}$$

with:

V_L volume of the refolding vessel [L]

F_R feed rate [L h^{-1}]

c_{DL} concentration of the denaturant in the refolding vessel [g L^{-1}]

c_{DR} concentration of the denaturant in the reservoir [mol L^{-1}]

c_{SR} concentration of solubilized protein in the reservoir [g L^{-1}]

The refolding and aggregation rates are functions of the denaturant concentration as described by Hevehan and de Bernardez Clark,

$$k_i(t) = a_i \cdot (1 + c_{DL}(t))^{b_i} \qquad (12)$$

with $i = r,a$ for refolding and aggregation respectively and two modeling constants a and b (Hevehan and de Bernardez Clark 1997).

The total protein concentration in the refolding vessel,

$$c_{PL} = \frac{F_R \cdot t}{V_L} \cdot c_{SR} \qquad (13)$$

is dependent on the feed rate into the reactor and the concentration of solubilized protein in the reservoir and can therefore easily be calculated, assuming that no protein is present in the refolding vessel at start and the reservoir contains solubilized protein only. The refolding of proteins is on a timescale of minutes to hours (Glassey et al. 2011), if the kinetics are known the measurement frequency of corresponding PAT systems can be estimated.

Besides, other models for protein refolding have been described as well. Ryś et al. set up a more complex model including a state M for misfolded protein with forward and reverse reaction due to disulfide bond reshuffling (Ryś et al. 2015a, b). Moreover, further adaptation of the process model may be necessary by including additional feeds such as that of an oxidizing agent (Fazeli et al. 2011) or of a cofactor needed for protein maturation (Rogers et al. 2000) to enhance the refolding process. The control of the dO_2 and redox potential becomes important when proteins containing disulfide bonds are processed. An oxidizing agent is necessary for the formation of disulfide bonds (De Bernardez Clark 2001) and since the reaction rates are dependent on the concentration of oxidizer and reducer (De Bernardez Clark 1998), they need to be added to the process model if they are not otherwise controlled to a fixed value (Dong et al. 2004).

Good modeling practice guidelines (Van Waveren et al. 1999; Nopens 2018) provide important steps throughout the model development to end up with an application-oriented model, such as the workflow described by Daume et al. (2020). After the formulation of equations for the reaction scheme and reaction kinetics, the model parameters, such as the folding and aggregation constants, are estimated by fitting the model to experimental data (Daume et al. 2020; Villaverde 2019). The predictive power of the model is usually determined by performing a normalized root mean square error calculation with the model simulation and experimental data.

Since the generation of a fitting process model is extremely difficult, especially in biological applications, hybrid models may offer a solution. These types of models combine the already-generated process knowledge in the form of mechanistic models with data-driven techniques such as artificial neural networks to compensate for simplifications made in the mechanistic model or to model a part of the process where mechanistic relations are unknown (Sohlberg and Jacobsen 2008).

Monitoring of refolding processes

To monitor biological reactions, direct measurements are often hard to obtain. However, monitoring of a biological process is not solely dependent on direct measurements, but rather a combination of available direct measurements, indirect measurements, and state estimation. Direct measurements of the formation of native protein can be directly measured in real time as described in "Analytical methods of refolding processes." For example, RP-HPLC is used to differentiate folding variants for disulfide bond containing proteins (Pathak et al. 2016), changes in the secondary structure can be analyzed with CD spectroscopy (Pathak et al. 2016), and FTIR spectroscopy (Walther et al. 2014; Pathak et al. 2016). Tertiary or quaternary protein structure can be investigated by NMR and extrinsic fluorescence (Pathak et al. 2016) and the surface charge of proteins by analysis of the zeta-potential (Pathak et al. 2016). Indirect measurements are available from soft sensors, which calculate an estimate of an otherwise unmeasurable system state by processing a measurable signal with an underlying estimation algorithm in real time (Kadlec et al. 2009; Luttmann et al. 2012). Pizarro et al. for example were able to monitor a protein refolding process with measurements of dissolved oxygen and redox potential for a disulfide bond containing protein (Pizarro et al. 2009).

Complementation of these direct and indirect measurements is realized by state estimation as a model-based approach that uses the process model to estimate the process states from system inputs and system outputs. Input for a protein refolding process in case of dilution by fed-batch mode can be the feeding rate of solubilized protein, while key performance indicators (KPIs) such as yield and space-time-yield of the refolding reaction can be the system outputs. Both KPIs can be determined online by soft sensors using the total protein concentration in the refolding vessel if the concentration of the native protein can be directly measured or accurately estimated. Simple examples of successful process estimation are based on mass, energy, or elemental balancing. Considering the law of mass conservation, unknown states can be directly calculated from others.

For more complex system descriptions, subject to non-linear reaction kinetics and internal dynamics, the most probable internal state can be estimated with non-linear Bayesian filters, such as the extended (Julier and Uhlmann 1997) and the unscented (Wan and van der Merwe 2000) Kalman filter or the particle filter (Arulampalam et al. 2002). Although some system states can be measured directly, the

implementation of a state observer can be advantageous toward feedback or predictive control applications. Directly measured signals are prone to measurement errors and substantial white noise and subject to outliers, which hamper a direct use of these signals for control. A state observer may remove these artifacts in a probabilistic way, without losing the information content of the measurement.

To our knowledge, there is no state observer described specifically for the use of protein refolding. However, the aforementioned techniques are applicable, as shown in different chemical and biochemical processes (Shen et al. 2006; Sun et al. 2008; Wang et al. 2010). Furthermore, interested readers are referred to Mohd et al. for a summary of different state observers and how to design them (Mohd Ali et al. 2015).

Control of refolding processes

Through the control of CPPs, KPIs can be held in their optimal range resulting in sufficient product quality. As a prerequisite, the CPPs (e.g., pH, temperature, agitation speed, dissolved oxygen (dO$_2$), redox potential, and feeding rate) and KPIs (e.g., refolding yield, space-time yield) need to be monitored through direct measurements or soft sensors to enable their control (Kadlec et al. 2009). The concept of controllability states that a process is controllable, if every state of the system can be modified to any arbitrary value by the system inputs in finite time (Ogata 1997).

Recent advances show two approaches to achieve control of protein refolding processes. Hebbi et al. are using statistical process control and batch evolution modeling spanning from the buffer generation, over solubilization until refolding with online measurements of redox potential, pH, and temperature. These parameters are commonly measured and hence there are many probes available. Monitoring of these CPPs is sufficient to control the process in the established design space (Hebbi et al. 2019). Another approach suggests measuring the changes in the secondary structure of proteins during folding via an inline ATR-FTIR. This monitoring enables controlled termination of the refolding process instead of termination after a given reaction time, therefore preventing additional aggregation (Walther et al. 2014).

Both of these approaches offer tools to perform controlled refolding processes. However, the control in these approaches is rather data-driven than knowledge-driven. Hence, the generation of a process model is favorable. A refolding process can be controlled through its model if the model parameters describing the refolding kinetics (Eq. 12) are identifiable and every state is controllable. Model-based control techniques rely on the estimation of the system states to compare estimated with measured values and consequently drive the process toward the defined reference using appropriate control actions (Brosilow and Joseph 2002). Inversion of the model (Ferrarin

et al. 2001; Kager et al. 2020) generates a control law using the feed rate of the dilution by fed-batch mode refolding as the control input to steer the refolding productivity to a constant value. A prominent model-based control technique is the model predictive control (MPC). At each time step, the actual state of the system is the initial condition for an online optimization process of state predictions based on the process model to gain the optimal system output (Grüne and Pannek 2017; Kaiser et al. 2018). The difference between the predicted output and the reference is optimized by minimizing a cost function. The user's control strategy regarding a given protein is reflected in the cost function, where a weight is associated with both KPIs to favor one of them. Additionally, the optimization process can be subject to constraints such as physical limits for pump rates or vessel volumes (Grüne and Pannek 2017; Kaiser et al. 2018). Limitations to time-dependent changes of the feed rate as control input are important as well since they prevent an alternation between slack and high speeds.

Since MPC, in combination with soft sensors and state observers, enables the control of non-measurable process parameters, it shows better control behavior compared to other control strategies, for example elemental balance control or classical PID control (Kager et al. 2020; Ulonska et al. 2018). The prediction into the future results in faster transition behavior and better reference tracking and generates profound process knowledge (Ulonska et al. 2018). As biochemical systems are very sensitive to small changes, MPC is desirable, because the prediction of control actions in advance can prevent critical overshoots, compared to classical PID control (Kager et al. 2020).

Proposal for a model-based refolding control strategy

With the industry gradually turning to the production of medically relevant products in the form of inclusion bodies (Humer and Spadiut 2018), there is also an increasing need for the improvement of the corresponding downstream processing methodologies that normally account for 50–80% of the manufacturing costs (Rathore et al. 2013). According to an economical assessment of various refolding strategies (Freydell et al. 2011), major cost-drivers on an industrial scale are large buffer volumes with expensive additives, requiring huge vessels and yet resulting in non-competitive yields of 15–25% of the total protein (Zhang et al. 2009) due to low recovery rates. Although other methods, such as ultrafiltration and on-column refolding, show promising results on a laboratory scale, dilution refolding is still the industrial method of choice. It is a simple, flexible, and already widely established technique with huge potential for optimization through process knowledge and control (Humer and Spadiut 2018; Linke

et al. 2014; Singh et al. 2015; Vallejo and Rinas 2004). From an economical and practical point of view, increasing overall product yields of an established methodology with minimal investment is in general more favorable than the implementation of completely new techniques. Additionally, thanks to the ongoing transformation of pharmaceutical manufacturing along the QbD principles, a shift from the empirical refolding methodologies toward mechanistic knowledge-based approaches has begun. In accordance with this, several studies have been conducted pointing to the advantages of PAT in terms of increased control of refolding processes leading to better efficiencies (Hebbi et al. 2019; Humer and Spadiut 2018; Pizarro et al. 2009; Walther et al. 2014). Furthermore, IB downstream processing could become more economic by implementing continuous refolding strategies (Pan et al. 2014; Wellhoefer et al. 2014). However, to obtain the complete picture of the underlying correlations in a refolding process and establish a thorough understanding of the kinetics, an integrated approach by introducing MPC could offer a solution.

Since MPC is a versatile and useful tool for the control of a process with non-linear dynamics, we suggest the application of this technique for a protein refolding process together with the dilution by fed-batch mode method. To control the refolding process with MPC, the system needs to be modeled in such a way that every state is observable and controllable and every model parameter is identifiable from available measurements. Hence, a mechanistic process model needs to be generated, the identifiability of the model parameters must be checked, and the model parameters be estimated as described by Daume et al. (2020) and Deppe et al. (2020). We propose that such a model must be generated with a state error of below 10–15%. Furthermore, the model needs to represent the process dynamics accurately and all model parameters need to be identifiable from experimental data (Daume et al. 2020). A systematic approach for parameter estimation is presented by Brun et al., including a classification of relative parameter uncertainty to simplify this time-consuming process (Brun et al. 2002). The model described by Dong et al. (2004) can be used as the starting reaction scheme and reaction kinetics can be taken from Hevehan and Bernardez Clark (1997).

In the following, we present a proposal on a refolding control strategy: we suggest the use of two parameters, refolding yield and refolding space-time-yield, as inputs for the calculation of the optimal control strategy of the feeding rate using MPC. Both parameters can be calculated online via soft sensors. To compute the yield, the native protein concentration as well as the maximal total protein concentration in the refolding vessel must be known. The refolding yield,

$$\text{Refolding yield} = \frac{c_{NL}}{c_{PL}} \qquad (14)$$

is the quotient of native protein concentration in the refolding vessel over the total protein concentration in the refolding vessel. The refolding space-time-yield,

$$\text{Refolding space−time−yield} = \frac{c_{NL}}{t} \qquad (15)$$

is the amount of native protein which is produced per volume of the refolding vessel and per process time.

Sensors and soft sensors can measure enough of the internal states of the process such that the process is completely observable and feed them to the controller (Fig. 2). After state estimation, the KPIs are calculated and the optimizer computes the optimal control inputs for the refolding process by minimizing the error between the estimated KPIs and their respective reference signal using a cost function and possibly fulfilling constraints. The cost function allows the user to weigh a KPI over another by setting these weights process-specific according to the targeted product. If the upstream process for example is expensive, higher yields are favorable; however, if the major bottleneck of the whole production process is the refolding, higher space-time-yields may be preferred.

The tradeoff of these two parameters is important, because focusing on a single KPI would generate misleading results. Very slow dilution rates could lead to a yield close to 100%; however, the necessary time and therefore the space-time-yield is uneconomical. Very fast feed rates on the other hand show higher productivity but decreased yield. Figure 3 visualizes the tradeoff between yield and space-time-yield based on simulations with the model and model parameters from Dong et al. (Dong et al. 2004). The solubilized protein is fed into the refolding vessel from a reservoir until it is depleted. The refolding process is continued as dilution by batch mode, during which the remaining solubilized protein folds. A comparison of varying final total protein concentration ($c_{PL,end}$) with constant feed rate versus varying feed rates with identical final total protein concentration shows that higher yields are reached when the concentration of solubilized protein in the refolding vessel at each time step is smaller. This is achieved through either lower final total protein concentration or lower feeding rates. To further emphasize the relation between yield and space-time-yield, multiple refolding simulations by dilution in fed-batch mode with feeding rates from 0.015 to 10 mL min^{-1} are used to visualize this behavior in a Pareto plot (Fig. 4), where the maximal yield is plotted against the mean space-time-yield during the fed-batch phase.

To transform this proposal into a working refolding process with monitoring and control, the defined model needs to be parametrized and validated to ensure its predictive power and applicability. Parametrization is usually performed by applying the weighted residual sum of squares WRSS, and model validation is performed by calculating and evaluating the goodness of fit (R^2) and the normalized root mean square error
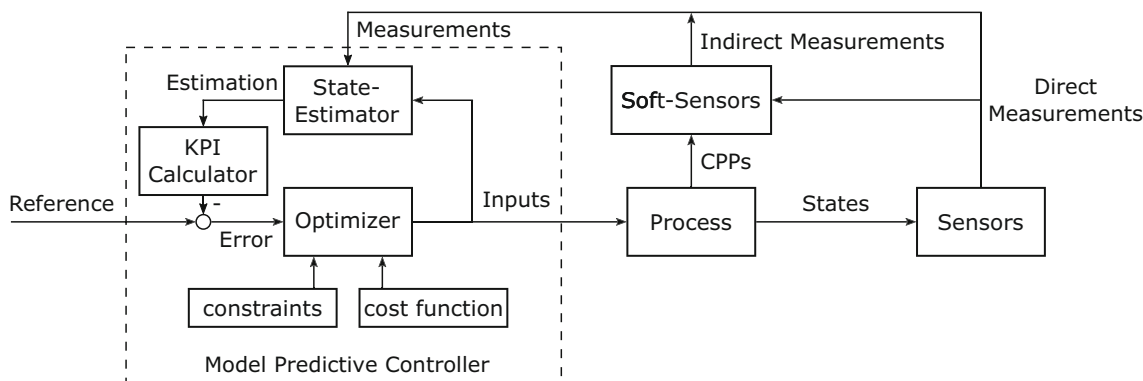
Fig. 2 Control strategy scheme. Combined direct and indirect measurements of CPPs and states of the process are used for state estimation together with computed control inputs. The KPI yield and space-time-yield are computed from the state estimate and their error compared to the reference signal is minimized by the optimizer of the model predictive controller according to the cost function and fulfilling constraints to achieve an optimal control input for the process

(NRMSE) of model simulations and experimental data (Daume et al. 2020; Deppe et al. 2020). Additionally, other methods such as Akaike information matrix (AIC) or Bayesian information criteria (BIC) give useful information about the risk of overfitting (the model fits the training data but fails with further datasets) and underfitting (approximate model is too simple to accurately predict the reaction kinetics), while incorporating the simplicity of the model in their evaluations (Deppe et al. 2020). For both parameter estimation and model validation, multiple data sets are necessary to prevent overfitting and to achieve a robust and applicable process model. Since the kinetics of the refolding process are generally understood, the model parameters can be estimated from few datasets. However, the addition of kinetic dependencies on oxidizing agent and cofactor as well as misfolded proteins and folding intermediates as states, the number of necessary datasets can rise sharply if the correct model structure is unclear.
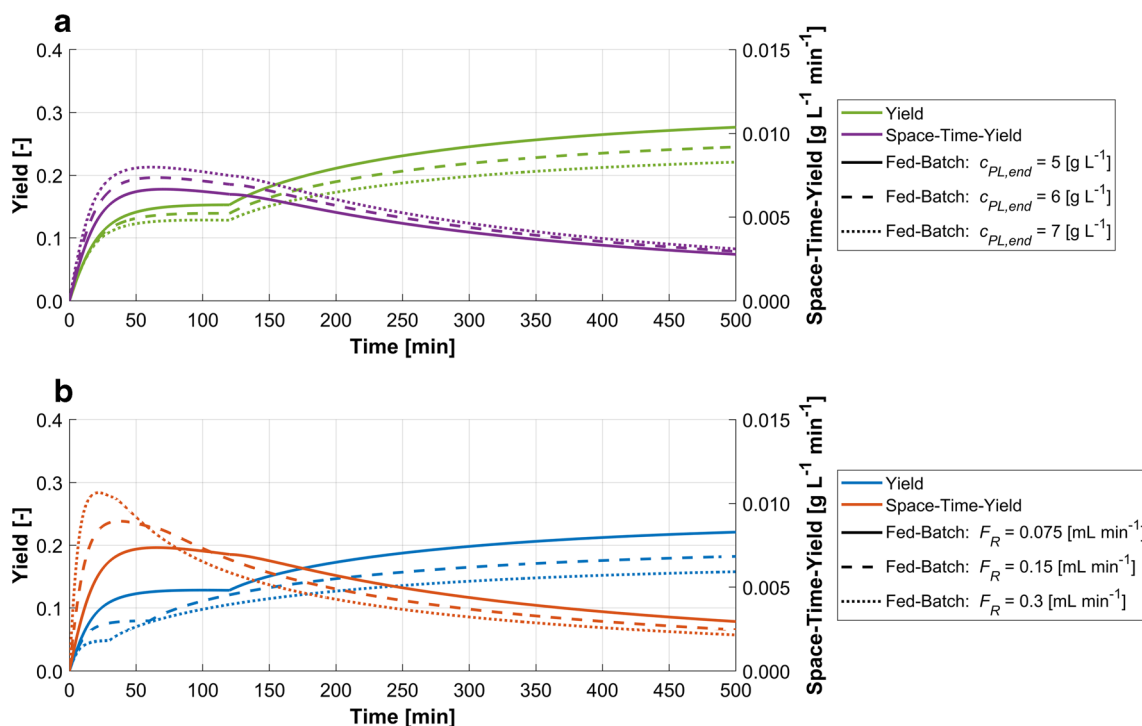
Fig. 3 Simulations of the refolding process as dilution by fed-batch mode. After depletion of the reservoir, the refolding is continued as batch mode. **a** Influence of final total protein concentration on yield (green) and space-time-yield (purple) with constant feed rate during the fed batch. Lowest final total protein concentration (solid line) results in the highest yield and lowest space-time-yield and vice versa. **b** Influence of feed rates on yield (blue) and space-time-yield (red) with constant final total protein concentration. Lowest feed rate (solid line) results in the highest yield and lowest space-time-yield and vice versa
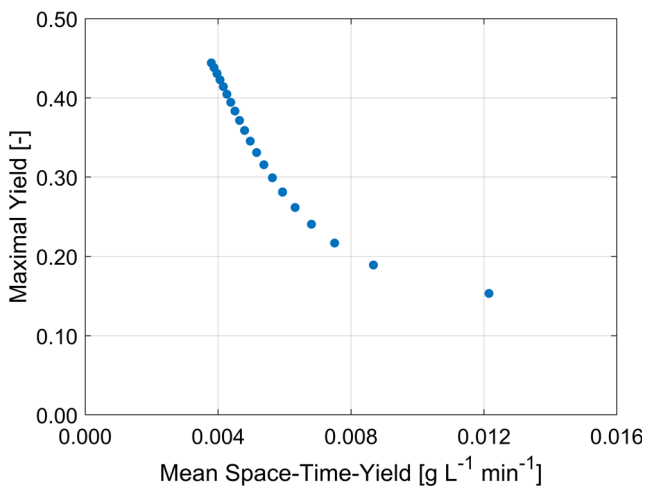
Fig. 4 Pareto plot to illustrate the tradeoff of the maximal refolding yield against the mean space-time-yield during the fed-batch phase with 20 feed rates between 0.015 and 10 mL min^{-1} and otherwise equal conditions

Conclusion

Although biopharmaceutical manufacturing has been transforming along the QbD principles in the recent years, industrial refolding is still conducted based on empirically established procedures instead of sound process knowledge. The reason behind this lies within the complex nature of these processes requiring cross-disciplinary methodology and knowledge to identify the key correlations. Until these black box processes are resolved, considerable efforts need to be made in terms of refolding process development that is costly, requires time, and yet results in non-scalable and strictly product-specific techniques with low yields.

However, the generation of mechanistic process knowledge on refolding kinetics could establish universally applicable methodologies within product families. To achieve this, real-time monitoring of relevant process variables with online sensors and soft sensors is indispensable. The analytical methods described in this review could provide a good basis for this; however, as previously discussed, the monitoring of refolding kinetics is still very challenging due to low protein concentrations. Additionally, not every process variable is directly measurable and related analytical methods have certain limitations. To meet these challenges, a modeling approach can be applied in addition to analytical techniques to estimate the non-measurable process variables in real time, thereby illuminating the complete picture of the process. Furthermore, when all relevant variables and their correlations are available, more complex model systems can be established to predict the process progression. Based on these, optimal and adaptive control trajectories are possible to achieve more efficient IB recovery processes, better product quality, and higher yields. Furthermore, in accordance with the current trends in the industry aiming at the digital transformation of

manufacturing processes, an adequate software environment is needed to support this endeavor. Such a software needs to support the real-time integration and calculation of complex models, based on the acquisition of current process data and subsequently, it should be able to provide the optimized output trajectories for the manufacturing system to realize closed-loop control. Eventually, not only refolding processes can be steered, but as a next step, the complete IB downstream processing chain could be optimized.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Amartely H, Avraham O, Friedler A, Livnah O, Lebendiker M (2018) Coupling multi angle light scattering to ion exchange chromatography (IEX-MALS) for protein characterization. Sci Rep 8(1):6907. https://doi.org/10.1038/s41598-018-25246-6

Amin S, Barnett GV, Pathak JA, Roberts CJ, Sarangapani PS (2014) Protein aggregation, particle formation, characterization & rheology. Curr Opin Colloid Interface Sci 19(5):438–449. https://doi.org/10.1016/j.cocis.2014.10.002

Anselment B, Baerend D, Mey E, Buchner J, Weuster-Botz D, Haslbeck M (2010) Experimental optimization of protein refolding with a

genetic algorithm. Protein Sci 19(11):2085–2095. https://doi.org/10.1002/pro.488

Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE T Signal Proces 50(2):174–188. https://doi.org/10.1109/78.978374

Brewster VL, Ashton L, Goodacre R (2013) Monitoring guanidinium-induced structural changes in ribonuclease proteins using Raman spectroscopy and 2D correlation analysis. Anal Chem 85(7):3570–3575. https://doi.org/10.1021/ac303265q

Brosilow, C., Joseph, B. (2002). Techniques of model-based control. Prentice-Hall international series in the physical and chemical engineering sciences. Upper Saddle River, N.J: Prentice Hall.

Brun R, Kühni M, Siegrist H, Gujer W, Reichert P (2002) Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets. Water Res 36(16):4113–4127. https://doi.org/10.1016/S0043-1354(02)00104-5

Buhrke D, Hildebrandt P (2020) Probing structure and reaction dynamics of proteins using time-resolved resonance Raman spectroscopy. Chem Rev 120(7):3577–3630. https://doi.org/10.1021/acs.chemrev.9b00429

Bunaciu AA, Aboul-Enein HY, Hoang VD (2015) Raman spectroscopy for protein analysis. Appl Spectrosc Rev 50(5):377–386. https://doi.org/10.1080/05704928.2014.990463

Cabrita LD, Bottomley SP (2004) Protein expression and refolding—a practical guide to getting the most out of inclusion bodies. Biotechnol. Annu. Rev. 31

Cho TH, Ahn SJ, Lee EK (2001) Refolding of protein inclusion bodies directly from E. coli homogenate using expanded bed adsorption chromatography. Bioseparation 10(4-5):189–196. https://doi.org/10.1023/a:1016305603569

Choi WC, Kim MY, Suh CW, Lee EK (2005) Solid-phase refolding of inclusion body protein in a packed and expanded bed adsorption chromatography. Process Biochem 40(5):1967–1972. https://doi.org/10.1016/j.procbio.2004.07.012

Clarke DT (2012) Circular dichroism in protein folding studies. Curr Protoc Protein Sci 70(1):28–23. https://doi.org/10.1002/0471140864.ps2803s70

Cleland JL, Hedgepeth C, Wang DI (1992) Polyethylene glycol enhanced refolding of bovine carbonic anhydrase B. Reaction stoichiometry and refolding model. J Biol Chem 267(19):13327–13334

Cowan RH, Davies RA, Pinheiro TTJ (2008) A screening system for the identification of refolding conditions for a model protein kinase, p38alpha. Anal Biochem 376(1):25–38. https://doi.org/10.1016/j.ab.2008.01.036

Daume S, Kofler S, Kager J, Kroll P, Herwig C (2020) Generic workflow for the setup of mechanistic process models. In: Pörtner R (ed) Animal Cell Biotechnology: Methods and Protocols. Springer US, New York, NY, pp 189–211. https://doi.org/10.1007/978-1-0716-0191-4_11

Davidson, K. A. (2008). Protein refolding via immobilisation on crystal surfaces (PhD Thesis). University of Glasgow. Retrieved from http://theses.gla.ac.uk/345/1/2008Davidson1PhD.pdf

De Bernardez Clark E (1998) Refolding of recombinant proteins. Curr Opin Biotech 9(2):157–163. https://doi.org/10.1016/S0958-1669(98)80109-2

De Bernardez Clark E (2001) Protein refolding for industrial processes. Curr Opin Biotech 12(2):202–207. https://doi.org/10.1016/s0958-1669(00)00200-7

Den Engelsman J, Garidel P, Smulders R, Koll H, Smith B, Bassarab S, Seidl A, Hainzl O, Jiskoot W (2011) Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. Pharm Res 28(4):920–933. https://doi.org/10.1007/s11095-010-0297-1

Deppe, S., Frahm, B., Hass, V. C., Hernández Rodríguez, T., Kuchemüller, K. B., Möller, J., Pörtner, R. (2020). Estimation of process model parameters. Methods in Molecular Biology (Clifton, N.J.), 2095, 213–234. doi:https://doi.org/10.1007/978-1-0716-0191-4_12

Dhadwal HS, Khan RR, Suh K (1993) Integrated fiber optic probe for dynamic light scattering. Appl Opt 32(21):3901–3904. https://doi.org/10.1364/AO.32.003901

Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4(1):10–19. https://doi.org/10.1038/nsb0197-10

Dolui S, Mondal A, Roy A, Pal U, Das S, Saha A, Maiti NC (2020) Order, disorder, and reorder state of lysozyme: aggregation mechanism by Raman spectroscopy. J Phys Chem B 124(1):50–60. https://doi.org/10.1021/acs.jpcb.9b09139

Dong X-Y, Shi G-Q, Li W, Sun Y (2004) Modeling and simulation of fed-batch protein refolding process. Biotechnol Prog 20(4):1213–1219. https://doi.org/10.1021/bp0499597

Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104(8):3607–3622. https://doi.org/10.1021/cr030403s

Eiberle MK, Jungbauer A (2010) Technical refolding of proteins: do we have freedom to operate? Biotechnol J 5(6):547–559. https://doi.org/10.1002/biot.201000001

Eryilmaz M, Zengin A, Boyaci IH, Tamer U (2017) Rapid quantification of total protein with surface-enhanced Raman spectroscopy with o - phthalaldehyde. J Raman Spectrosc 48(5):653–658. https://doi.org/10.1002/jrs.5099

Fazeli A, Shojaosadati SA, Fazeli MR, Ilka H (2011) Effect of parallel feeding of oxidizing agent and protein on fed-batch refolding process of recombinant interferon beta-1b. Process Biochem 46(3):796–800. https://doi.org/10.1016/j.procbio.2010.11.007

Ferrarin M, Palazzo F, Riener R, Quintern J (2001) Model-based control of FES-induced single joint movements. IEEE Trans Neural Syst Rehabil Eng 9(3):245–257. https://doi.org/10.1109/7333.948452

Food and Drug Administration. (2004). Guidance for industry, PAT-A framework for innovative pharmaceutical development, manufacturing and quality assurance. https://www.fda.gov/media/71012/download.

Freydell EJ, van der Wielen LAM, Eppink MHM, Ottens M (2011) Techno-economic evaluation of an inclusion body solubilization and recombinant protein refolding process. Biotechnol Prog 27(5):1315–1328. https://doi.org/10.1002/btpr.652

Frueh DP, Goodrich AC, Mishra SH, Nichols SR (2013) Nmr methods for structural studies of large monomeric and multimeric proteins. Curr Opin Struct Biol 23(5):734–739. https://doi.org/10.1016/j.sbi.2013.06.016

Gernaey KV, Lantz AE, Tufvesson P, Woodley JM, Sin G (2010) Application of mechanistic models to fermentation and biocatalysis for next-generation processes. Trends Biotechnol 28(7):346–354. https://doi.org/10.1016/j.tibtech.2010.03.006

Glassey J, Gernaey KV, Clemens C, Schulz TW, Oliveira R, Striedner G, Mandenius C-F (2011) Process analytical technology (PAT) for biopharmaceuticals. Biotechnol J 6(4):369–377. https://doi.org/10.1002/biot.201000356

Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1(6):2876–2890. https://doi.org/10.1038/nprot.2006.202

Gregoire S, Irwin J, Kwon I (2012) Techniques for monitoring protein misfolding and aggregation in vitro and in living cells. Korean J Chem Eng 29(6):693–702. https://doi.org/10.1007/s11814-012-0060-x

Griffiths PR (1983) Fourier transform infrared spectrometry. Science 222(4621):297–302. https://doi.org/10.1126/science.6623077

Grüne, L., Pannek, J. (2017). Nonlinear model predictive control. In Nonlinear Model Predictive Control: Theory and Algorithms (pp. 45–69). Cham: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-46024-6_3

Hauser K (2013) Infrared spectroscopy of protein folding, misfolding and aggregation. In: Roberts GCK (ed) Encyclopedia of Biophys. Springer Berlin, Berlin, pp 1089–1095. https://doi.org/10.1007/978-3-642-16712-6_115

Hawe A, Sutter M, Jiskoot W (2008) Extrinsic fluorescent dyes as tools for protein characterization. Pharm Res 25(7):1487–1499. https://doi.org/10.1007/s11095-007-9516-9

Hebbi V, Thakur G, Rathore AS (2019) Process analytical technology implementation for protein refolding: Gcsf as a case study. Biotechnol Bioeng 116(5):1039–1052. https://doi.org/10.1002/bit.26900

Hemmig R, Johann C, Ramage P (2005) Asymmetric flow field-flow fractionation (AF4) with multi-angle light scattering (MALS) for high-throughput protein refolding. LC GC Europe 18(10):532–538

Hevehan DL, de Bernardez Clark E (1997) Oxidative renaturation of lysozyme at high concentrations. Biotechnol Bioeng 54(3):221–230. https://doi.org/10.1002/(SICI)1097-0290(19970505)54:3<221::AID-BIT3>3.0.CO;2-H

Hong P, Koza S, Bouvier ESP (2012) Size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. J Liq Chromatogr Relat Technol 35(20):2923–2950. https://doi.org/10.1080/10826076.2012.743724

Humer D, Spadiut O (2018) Wanted: more monitoring and control during inclusion body processing. World J Microb Biot 34(11):158. https://doi.org/10.1007/s11274-018-2541-5

Hussain MT, Forbes N, Perrie Y (2019) Comparative analysis of protein quantification methods for the rapid determination of protein loading in liposomal formulations. Pharmaceutics 11(1):39. https://doi.org/10.3390/pharmaceutics11010039

Jazaj D, Ghadami SA, Bemporad F, Chiti F (2019) Probing conformational changes of monomeric transthyretin with second derivative fluorescence. Sci Rep 9(1):10988. https://doi.org/10.1038/s41598-019-47230-4

Julier, S. J., Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems. In Signal processing, sensor fusion, and target recognition VI (Vol. 3068, pp. 182-193). International Society for Optics and Photonics. doi:https://doi.org/10.1117/12.280797

Jungbauer A, Kaar W (2007) Current status of technical protein refolding. J Biotechnol 128(3):587–596. https://doi.org/10.1016/j.jbiotec.2006.12.004

Kadlec P, Gabrys B, Strandt S (2009) Data-driven soft sensors in the process industry. Comput Chem Eng 33(4):795–814. https://doi.org/10.1016/j.compchemeng.2008.12.012

Kager J, Tuveri A, Ulonska S, Kroll P, Herwig C (2020) Experimental verification and comparison of model predictive, PID and model inversion control in a *Penicillium chrysogenum* fed-batch process. Process Biochem 90:1–11. https://doi.org/10.1016/j.procbio.2019.11.023

Kaiser E, Kutz JN, Brunton SL (2018) Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc Math Phys Eng 474(2219):20180335. https://doi.org/10.1098/rspa.2018.0335

Katoh S, Katoh Y (2000) Continuous refolding of lysozyme with fed-batch addition of denatured protein solution. Process Biochem 35(10):1119–1124. https://doi.org/10.1016/S0032-9592(00)00145-X

Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta 1751(2):119–139. https://doi.org/10.1016/j.bbapap.2005.06.005

Kiefhaber T, Rudolph R, Kohler HH, Buchner J (1991) Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation. Nat Biotechnol 9(9):825–829. https://doi.org/10.1038/nbt0991-825

Kudou M, Yumioka R, Ejima D, Arakawa T, Tsumoto K (2011) A novel protein refolding system using lauroyl-l-glutamate as a solubilizing detergent and arginine as a folding assisting agent. Protein Expr Purif 75(1):46–54. https://doi.org/10.1016/j.pep.2010.08.011

Ladokhin, A. S. (2009). Fluorescence spectroscopy in thermodynamic and kinetic analysis of pH-dependent membrane protein insertion. In G. K. Ackers, J. M. Holt, & M. L. Johnson (Eds.), Methods in Enzymology: v. 466. Biothermodynamics: Part B (Vol. 466, pp. 19–42). San Diego, Calif: Academic Press/Elsevier. doi:https://doi.org/10.1016/S0076-6879(09)66002-X

Lakowicz JR (1988) Principles of frequency-domain fluorescence spectroscopy and applications to cell membranes. In: Hilderson HJ (ed) Fluorescence Studies on Biological Membranes. Springer US, Boston, MA, pp 89–126. https://doi.org/10.1007/978-1-4613-9359-7_3

Lamba J, Paul S, Hasija V, Aggarwal R, Chaudhuri TK (2009) Monitoring protein folding and unfolding pathways through surface hydrophobicity changes using fluorescence and circular dichroism spectroscopy. Biochemistry (Mosc) 74(4):393–398. https://doi.org/10.1134/s0006297909040063

Lew J, Taylor SS, Adams JA (1997) Identification of a partially rate-determining step in the catalytic mechanism of cAMP-dependent protein kinase: a transient kinetic study using stopped-flow fluorescence spectroscopy. Biochemistry 36(22):6717–6724. https://doi.org/10.1021/bi963164u

Lindon JC, Trantner GE, Koppenaal DW (2016) Encyclopedia of spectroscopy and spectrometry. Elsevier AP Academic Press, Amsterdam, Boston, Heidelberg

Linke T, Aspelund MT, Thompson C, Xi G, Fulton A, Wendeler M, Pabst TM, Wang X, Wang WK, Ram K, Hunter AK (2014) Development and scale-up of a commercial fed batch refolding process for an anti-CD22 two chain immunotoxin. Biotechnol Prog 30(6):1380–1389. https://doi.org/10.1002/btpr.1983

Luttmann R, Bracewell DG, Cornelissen G, Gernaey KV, Glassey J, Hass VC, Kaiser C, Preusse C, Striedner G, Mandenius C-F (2012) Soft sensors in bioprocessing: a status report and recommendations. Biotechnol J 7(8):1040–1048. https://doi.org/10.1002/biot.201100506

Machuca, M. A., Roujeinikova, A. (2017). Method for efficient refolding and purification of chemoreceptor ligand binding domain. JoVE. Advance online publication. doi:https://doi.org/10.3791/57092

Mayer, M., Buchner, J. (2004). Refolding of inclusion body proteins. In Molecular Diagnosis of Infectious Diseases (pp. 239–254). Humana Press. doi:https://doi.org/10.1385/1-59259-679-7:239

Micsonai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M, Kardos J (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U. S. A. 112(24):E3095–E3103. https://doi.org/10.1073/pnas.1500851112

Middelberg APJ (2002) Preparative protein refolding. Trends Biotechnol 20(10):437–443. https://doi.org/10.1016/s0167-7799(02)02047-4

Mirhosseini SA, Latifi AM, Mahmoodzadeh Hosseini H, Seidmoradi R, Aghamollaei H, Farnoosh G (2019) The efficient solubilization and refolding of recombinant organophosphorus hydrolyses inclusion bodies produced in *Escherichia coli*. Journal of Appl Biotechnol Rep 6(1):20–25. https://doi.org/10.29252/JABR.06.01.04

Mohd Ali J, Ha Hoang N, Hussain MA, Dochain D (2015) Review and classification of recent observers applied in chemical process systems. Comp Chem Eng 76:27–41. https://doi.org/10.1016/j.compchemeng.2015.01.019

Moore-Kelly C, Welsh J, Rodger A, Dafforn TR, Thomas ORT (2019) Automated high-throughput capillary circular dichroism and intrinsic fluorescence spectroscopy for rapid determination of protein structure. Anal Chem 91(21):13794–13802. https://doi.org/10.1021/acs.analchem.9b03259

Naiim M, Boualem A, Ferre C, Jabloun M, Jalocha A, Ravier P (2015) Multiangle dynamic light scattering for the improvement of multi-modal particle size distribution measurements. Soft Matter 11(1): 28–32. https://doi.org/10.1039/c4sm01995d

Nopens, I. (2018). Good modelling practice for process engineering: pitfalls and requirements to develop fit for purpose models. In A. Friedl, J. J. Klemeš, P. S. Varbanov, S. Radl, & T. Wallek (Eds.), Computer-aided chemical engineering: Vol. 43. 28th European Symposium on Computer Aided Process Engineering: Part A (Vol. 43, pp. 293–294). Amsterdam: Elsevier. doi:https://doi.org/10.1016/B978-0-444-64235-6.50054-1

Ogata, K. (1997). Modern control engineering (5th ed.): Prentice Hall.

Ogura K, Kobashigawa Y, Saio T, Kumeta H, Torikai S, Inagaki F (2013) Practical applications of hydrostatic pressure to refold proteins from inclusion bodies for NMR structural studies. Protein Eng Des Sel 26(6):409–416. https://doi.org/10.1093/protein/gzt012

Pan S, Zelger M, Jungbauer A, Hahn R (2014) Integrated continuous dissolution, refolding and tag removal of fusion proteins from inclusion bodies in a tubular reactor. J Biotechnol 185:39–50. https://doi.org/10.1016/j.jbiotec.2014.06.010

Pan S, Odabas N, Sissolak B, Imendörffer M, Zelger M, Jungbauer A, Hahn R (2015) Engineering batch and pulse refolding with transition of aggregation kinetics: an investigation using green fluorescent protein (GFP). Chem Eng Sci 131:91–100. https://doi.org/10.1016/j.ces.2015.03.054

Pathak M, Dixit S, Muthukumar S, Rathore AS (2016) Analytical characterization of in vitro refolding in the quality by design paradigm: refolding of recombinant human granulocyte colony stimulating factor. J Pharm Biomed Anal 126:124–131. https://doi.org/10.1016/j.jpba.2016.05.001

Pelton JT, McLean LR (2000) Spectroscopic methods for analysis of protein secondary structure. Anal Biochem 277(2):167–176. https://doi.org/10.1006/abio.1999.4320

Pizarro SA, Dinges R, Adams R, Sanchez A, Winter C (2009) Biomanufacturing process analytical technology (PAT) application for downstream processing: using dissolved oxygen as an indicator of product quality for a protein refolding reaction. Biotechnol Bioeng 104(2):340–351. https://doi.org/10.1002/bit.22382

Printz M, Friess W (2012) Simultaneous detection and analysis of protein aggregation and protein unfolding by size exclusion chromatography with post column addition of the fluorescent dye BisANS. J Pharm Sci 101(2):826–837. https://doi.org/10.1002/jps.22808

Qin PZ, Pyle AM (1997) Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg2+ ions in the tertiary structure. Biochemistry 36(16):4718–4730. https://doi.org/10.1021/bi962665c

Qoronfleh MW, Hesterberg LK, Seefeldt MB (2007) Confronting high-throughput protein refolding using high pressure and solution screens. Protein Expr Purif 55(2):209–224. https://doi.org/10.1016/j.pep.2007.05.014

Rathore AS, Bade P, Joshi V, Pathak M, Pattanayek SK (2013) Refolding of biotech therapeutic proteins expressed in bacteria: review. J Chem Technol Biotechnol 88(10):1794–1806. https://doi.org/10.1002/jctb.4152

Reed CJ, Bushnell S, Evilia C (2014) Circular dichroism and fluorescence spectroscopy of cysteinyl-tRNA synthetase from Halobacterium salinarum ssp. Nrc-1 demonstrates that group I cations are particularly effective in providing structure and stability to this halophilic protein. PloS One 9(3):e89452. https://doi.org/10.1371/journal.pone.0089452

Rogers MS, Baron AJ, McPherson MJ, Knowles PF, Dooley DM (2000) Galactose oxidase pro-sequence cleavage and cofactor assembly are self-processing reactions. J Am Chem Soc 122(5):990–991. https://doi.org/10.1021/ja993385y

Ryś S, Muca R, Kołodziej M, Piątkowski W, Dürauer A, Jungbauer A, Antos D (2015a) Design and optimization of protein refolding with crossflow ultrafiltration. Chem Eng Sci 130:290–300. https://doi.org/10.1016/j.ces.2015.03.035

Ryś S, Piątkowski W, Antos D (2015b) Predictions of matrix-assisted refolding of α-lactalbumin: process efficiency versus batch dilution method. Eng Life Sci 15(1):140–151. https://doi.org/10.1002/elsc.201400105

Schuttlefield JD, Grassian VH (2008) ATR–FTIR spectroscopy in the undergraduate chemistry laboratory. Part I: Fundamentals and Examples. J Chem Educ 85(2):279. https://doi.org/10.1021/ed085p279

Shen H, Nelson G, Kennedy S, Nelson D, Johnson J, Spiller D, White MRH, Kell DB (2006) Automatic tracking of biological cells and compartments using particle filters and active contours. Chemom Intell Lab Syst 82(1-2):276–282. https://doi.org/10.1016/j.chemolab.2005.07.007

Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 99(4):303–310. https://doi.org/10.1263/jbb.99.303

Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK (2015) Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Factories 14(1):41. https://doi.org/10.1186/s12934-015-0222-8

Singhvi P, Saneja A, Srichandan S, Panda AK (2020) Bacterial inclusion bodies: a treasure trove of bioactive proteins. Trends Biotechnol 38(5):474–486. https://doi.org/10.1016/j.tibtech.2019.12.011

Sohlberg B, Jacobsen EW (2008) Grey Box Modelling – branches and experiences. IFAC Proc Volumes 41(2):11415–11420. https://doi.org/10.3182/20080706-5-kr-1001.01934

Su Z, Lu D, Liu Z (2011) Refolding of inclusion body proteins from E. coli. Methods Biochem Anal 54:319–338. https://doi.org/10.1002/9780470939932.ch13

Sun X, Jin L, Xiong M (2008) Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks. PloS One 3(11):e3758. https://doi.org/10.1371/journal.pone.0003758

Sutter M, Oliveira S, Sanders NN, Lucas B, van Hoek A, Hink MA, Visser AJWG, de Smedt SC, Hennink WE, Jiskoot W (2007) Sensitive spectroscopic detection of large and denatured protein aggregates in solution by use of the fluorescent dye Nile red. J Fluoresc 17(2):181–192. https://doi.org/10.1007/s10895-007-0156-6

Tatulian SA (2019) FTIR analysis of proteins and protein–membrane interactions. In: Kleinschmidt JH (ed) Lipid-Protein Interactions: Methods and Protocols. Springer New York, New York, NY, pp 281–325. https://doi.org/10.1007/978-1-4939-9512-7_13

Tsumoto K, Ejima D, Kumagai I, Arakawa T (2003) Practical considerations in refolding proteins from inclusion bodies. Protein Expr Purif 28(1):1–8. https://doi.org/10.1016/S1046-5928(02)00641-1

Ulonska S, Waldschitz D, Kager J, Herwig C (2018) Model predictive control in comparison to elemental balance control in an E. coli fed-batch. Chem Eng Sci 191:459–467. https://doi.org/10.1016/j.ces.2018.06.074

Umetsu M, Tsumoto K, Hara M, Ashish K, Goda S, Adschiri T, Kumagai I (2003) How additives influence the refolding of immunoglobulin-folded proteins in a stepwise dialysis system. Spectroscopic evidence for highly efficient refolding of a single-chain Fv fragment. The J Biol Chem 278(11):8979–8987. https://doi.org/10.1074/jbc.M212247200

Umetsu M, Tsumoto K, Ashish K, Nitta S, Tanaka Y, Adschiri T, Kumagai I (2004) Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins. FEBS Lett 557(1-3):49–56. https://doi.org/10.1016/S0014-5793(03)01441-8

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. https://doi.org/10.1186/1475-2859-3-11

Van Waveren, H., Groot, S., Scholten, H., Van Geer, F. C., Wösten, H., Koeze, R., Noort, J. J. (1999). Good Modelling Practice Handbook.

Veiter L, Herwig C (2019) The filamentous fungus *Penicillium chrysogenum* analysed via flow cytometry-a fast and statistically sound insight into morphology and viability. Appl Microbiol Biotechnol 103(16):6725–6735. https://doi.org/10.1007/s00253-019-09943-4

Villaverde AF (2019) Observability and structural identifiability of nonlinear biological systems. Complexity 2019:1–12. https://doi.org/10.1155/2019/8497093

Vincentelli R, Canaan S, Campanacci V, Valencia C, Maurin D, Frassinetti F, Scappucini-Calvo L, Bourne Y, Cambillau C, Bignon C (2004) High-throughput automated refolding screening of inclusion bodies. Protein Sci 13(10):2782–2792. https://doi.org/10.1110/ps.04806004

Walther C, Mayer S, Jungbauer A, Dürauer A (2014) Getting ready for PAT: Scale up and inline monitoring of protein refolding of Npro fusion proteins. Process Biochem 49(7):1113–1121. https://doi.org/10.1016/j.procbio.2014.03.022

Wan EA, van der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: Haykin SS (ed) The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium: As-SPCC : October 1-4, 2000, Chateau Lake Louise, Lake Louise, Alberta, Canada. IEEE, Piscataway, NJ, pp 153–158. https://doi.org/10.1109/ASSPCC.2000.882463

Wang J, Zhao L, Yu T (2010) On-line estimation in fed-batch fermentation process using state space model and unscented Kalman filter. Chin J Chem Eng 18(2):258–264. https://doi.org/10.1016/S1004-9541(08)60351-1

Wang C-H, Huang C-C, Lin L-L, Chen W (2016) The effect of disulfide bonds on protein folding, unfolding, and misfolding investigated by FT-Raman spectroscopy. J Raman Spectrosc 47(8):940–947. https://doi.org/10.1002/jrs.4935

Wechselberger P, Sagmeister P, Herwig C (2013) Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. Bioproc Biosystems Eng 36(9):1205–1218. https://doi.org/10.1007/s00449-012-0848-4

Wellhoefer M, Sprinzl W, Hahn R, Jungbauer A (2014) Continuous processing of recombinant proteins: integration of refolding and purification using simulated moving bed size-exclusion chromatography with buffer recycling. J Chromatogr A 1337:48–56. https://doi.org/10.1016/j.chroma.2014.02.016

Wen Z-Q (2007) Raman spectroscopy of protein pharmaceuticals. J Pharm Sci 96(11):2861–2878. https://doi.org/10.1002/jps.20895

Wüthrich K (2003) Nmr studies of structure and function of biological macromolecules. Biosci Rep 23(4):119–168. https://doi.org/10.1023/b:bire.0000007689.27614.7c

Yamaguchi H, Miyazaki M (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules 4(1):235–251. https://doi.org/10.3390/biom4010235

Yasuda M, Murakami Y, Sowa A, Ogino H, Ishikawa H (1998) Effect of additives on refolding of a denatured protein. Biotechnol Prog 14(4):601–606. https://doi.org/10.1021/bp9800438

Yu Z, Reid JC, Yang Y-P (2013) Utilizing dynamic light scattering as a process analytical technology for protein folding and aggregation monitoring in vaccine manufacturing. J Pharm Sci 102(12):4284–4290. https://doi.org/10.1002/jps.23746

Zettlmeissl G, Rudolph R, Jaenicke R (1979) Reconstitution of lactic dehydrogenase. Noncovalent aggregation vs. Reactivation. 1. Physical properties and kinetics of aggregation. Biochemistry 18(25):5567–5571. https://doi.org/10.1021/bi00592a007

Zhang T, Xu X, Shen L, Feng Y, Yang Z, Shen Y, Wang J, Jin W, Wang X (2009) Modeling of protein refolding from inclusion bodies. Acta Biochim Biophys Sin 41(12):1044–1052. https://doi.org/10.1093/abbs/gmp098

Živančev D, Horvat D, Torbica A, Belović M, Šimić G, Magdić D, Đukić N (2015) Benefits and limitations of lab-on-a-chip method over reversed-phase high-performance liquid chromatography method in gluten proteins evaluation. J Chem 2015:1–9. https://doi.org/10.1155/2015/430328