Petri net Modelling to Assess the Performance of the Protein Folding Machinery for Recombinantly Expressed Proteins in *Escherichia coli*

Fairooza Alam

A dissertation submitted in partial fulfilmentof the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Biochemical EngineeringUniversity College London

Mar 20, 2025

I, Fairooza Alam, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The production of complex recombinant proteins in *E. coli* is a foundation mark in bioprocessing. However, roadblocks such as post-translational modifications and inclusion body formation often impede protein expression despite *E. coli's* reputation for achieving rapid growth, cost-effectiveness, and well-characterised genetics. Addressing these hurdles is essential for advancing the efficiency and productivity of bioprocesses.

Formal modelling methods have been adopted in biotechnology to offer alternative solutions to challenges associated with cellular biology. This thesis presents the first model to combine different mechanisms in protein folding machinery in *E. coli* using Petri net, offering a novel approach to overcoming folding-related obstacles. The innovative model sheds light on the critical role of molecular chaperones in protein synthesis, translocation, and disulphide bond formation, which are fundamental processes that correct protein misfolds at the cellular level. By integrating the tricarboxylic acid (TCA) cycle and oxidative phosphorylation, the model also highlights the essential requirements for energy and reducing agents driving protein production.

The fully constructed model comprises 315 unique biological states represented as places and 329 biological reactions represented as transitions, resulting in 896 arcs connecting various biochemical responses to the following biological state. The construction of this Petri net model was a comprehensive process involving a thorough search and analysis of datasets from multiple databases and literature. The discrepancies were carefully evaluated to validate the assumptions made while constructing the model. Using the human growth hormone (HGH) as a model protein, this research revealed a direct correlation between energy availability and the quantity of correctly folded protein produced. Furthermore, the model illustrates the detrimental effects of glucose limitation on protein folding, causing inclusion body formation and generating mis-oxidised proteins.

Test cases evaluated the model's robustness and applicability. One of these test cases was modelling the pathways of signal peptides through the Sec system, which showed fluctuating transition patterns dependent on the coupled or uncoupled pathway. This biological phenomenon was documented for the first time in this work and is useful when determining which signal peptide to use for the expression and correct translocation of a specific protein.

In addition, the model can investigate the utility of the commercial CyDisCo strain in protein production that employs the Tat translocation pathway. Furthermore, the impact of overexpressing critical molecular chaperones such as DnaK/DnaJ and GroEL/ES was explored, and it was shown that even minor changes in the transition rates describing the processes in these pathways could impact the total amount of protein produced.

Lastly, the elimination of DsbA, DsbB, DsbC, and DsbD proteins from the disulphide bond formation process was studied to understand their role and impact on the model. The simulations exposed crucial data gaps and the interdependence between folding and translocation pathways, underscoring its potential as a potent tool for honing strain design. The model simulation showed that when the energy generation rate is slower than the rate of folding, misfolded proteins accumulate, creating a delay in folding.

This model presents complementary and alternative methods to minimise the formation of inclusion bodies and improve the amount of correctly folded protein by removing chaperones and overexpressing chaperones, depending on the protein's nature. The utilisation of this model in strain design to facilitate the successful expression of recombinant proteins in *E. coli* is expected to propel the field of bioprocessing into a more efficient and productive future.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Duygu Dikicioglu. Your expertise, patience, and unwavering support have been instrumental throughout my research journey. Your encouragement and insightful feedback have not only helped me overcome numerous challenges but have also inspired me to pursue my future endeavours with confidence and determination.

I would also like to extend my heartfelt thanks to Professor Gary Lye, whose support and mentorship have been pivotal throughout my PhD. Your vast knowledge and willingness to share your expertise have enriched my academic experience and contributed to my development as a researcher. Thank you for always being available to discuss ideas, provide constructive criticism, and encourage me to push the boundaries of my research.

A significant part of this journey has been the incredible support and motivation from our research group DigiSmartBio . The collaborative environment and shared passion for discovery have made this experience both rewarding and enjoyable.

I am profoundly thankful to my family for their support and love throughout this journey. To my husband, thank you for your endless patience, understanding, and encouragement. Your belief in me has been a constant source of strength and has helped me persevere through the toughest times. To my parents, thank you for instilling in me the value of hard work and perseverance. Your sacrifices and faith in my abilities have been my driving force.

Finally, I would like to express my sincere gratitude to EPSRC for their generous funding and support, which made my dream of becoming a scientist a reality. Their commitment to fostering research and innovation has provided me with the resources and opportunities necessary to pursue my passion and contribute to the scientific community.

Thank you all for being a part of this journey and for helping me reach this significant milestone in my academic career.

Impact Statement

Developing a comprehensive and quantitative model that connects protein synthesis pathways in *Escherichia coli* with translocation, disulphide bond formation pathways, the tricarboxylic acid (TCA) cycle, and oxidative phosphorylation represents a significant advancement in molecular biology and systems biology. This model bridges the understanding of various interconnected cellular processes and provides a versatile tool that can be adapted to any *E. coli* strain and expressed protein. The stoichiometric nature of the model allows for the quantitative comparison of different biological and biotechnological scenarios. The implications of this research are multifaceted, impacting academic research, industrial applications, and potentially therapeutic developments.

This integrated model addresses a critical gap in the understanding of how the process of forming a correctly folded protein is intricately linked with translocation and disulphide bond formation pathways. The model enhances the comprehension of the regulation of the cellular machinery by elucidating the connections between these pathways. This is particularly significant as it offers insights into the routes utilised by *E. coli* under different environmental and physiological conditions. Understanding these interactions at a detailed level allows researchers to decipher the complex web of cellular processes and use this information to design better-suited strains for the expression of a specific protein.

The practical implications of this model are vast. In industrial biotechnology, *E. coli* is a widely used host for the production of recombinant proteins. This model will be a predictive tool to improve protein production processes by identifying bottlenecks and suggesting modifications to improve yield. The ability to modify the model to represent different strains and proteins adds a layer of versatility that is crucial for industrial applications. In addition, the graphical representation of Petri nets makes it more attractive and interactive for scientists with predominantly experimental expertise to implement. Its adaptability ensures that the model remains relevant across various biotechnological processes, making it an asset for companies involved in microbial bioproduction involving protein products.

Furthermore, the model's capacity to illustrate the interaction of molecular chaperones with protein synthesis pathways offers the potential to improve the stability and the extent of correct folding of recombinant proteins. This is particularly beneficial for the pharmaceutical industry, where the correct folding of therapeutic proteins is critical for their efficacy and safety. By leveraging this model, researchers can design superior strategies to enhance protein folding, reduce aggregation, and improve the overall quality of biopharmaceutical products.

This model's broader impact extends to synthetic biology, where designing and constructing biological systems with desired properties is a crucial objective. This model can inform the design of synthetic circuits that require precise control of protein expression. Additionally, the insights gained from this model can influence future research directions, such as engineering *E. coli* strains with enhanced metabolic capabilities for biofuel production.

On a societal level, the model contributes to the understanding of pathways to reduce inclusion bodies and improve protein yield, which can reduce biotherapeutics' product development and manufacturing costs. This will make life-changing treatments available to more people and improve their quality of life.

In summary, the integrated model of protein synthesis pathways in *E. coli* significantly contributes to molecular biology, biotechnology, and systems biology. Its ability to connect various cellular processes offers profound insights into the functioning of *E. coli*, with practical applications that span from improving industrial bioproduction to informing synthetic biology and therapeutic protein development. The versatility and adaptability of this model ensure its continued relevance and utility, paving the way for future innovations and research advancements.

Table of Contents

Table of Figures

List of Tables

Chapter 1

1. Introduction

Biotechnology is a field that involves using biological systems, organisms, or derivatives to develop products or processes that benefit society (Naz, 2015). This has led to the rise of bioprocessing, which employs the natural abilities of living cells to produce various products that notably impact economic and social development (Barragán-Ocaña et al., 2020).

Conventional bioprocessing methods involve three stages: bioprocess development, upstream processing, and downstream purification (Boodhoo et al., 2022). Bioprocess development is a comprehensive term encompassing several small-scale steps, including selecting the strain and cell line, choosing the appropriate media, and selecting the bioreactor and the mode of operation (Tripathi & Shrivastava, 2019a). A standard process flow diagram is shown in Figure 1.1.

Upstream processing involves the cultivation of cells, which includes the expansion of inoculum and the addition of media and bioreactor operations to promote optimal growth of the expression systems. The conditions and media required for this process vary depending on the expression systems and products of interest (Apostolidis et al., 2015; Mollet et al., 2004).

Downstream processing involves purification techniques that separate the desired product from other impurities during manufacturing. The level of purity required for the product depends on the intended application. In some cases, purification is not required, for example, if an enzyme is produced in bulk for industrial purposes (Gao et al., 2019). In other cases, purification is a regulation, for example, the purification of therapeutics, which involves various filtration and chromatography steps (Siew & Zhang, 2021; Spiesberger et al., 2015).

In bioprocessing, reducing costs and improving the yield of target products are crucial. This has led to the need to scale manufacturing after the proof of concept. However, the transfer from small to large scale can pose difficulties for expression systems, which may not perform as optimally at large scale as in small-scale studies. Researchers are developing scale-down simulators to evaluate various strains and environmental conditions that could yield better results during large-scale production.

These simulators replicate the conditions of industrial bioprocessing on a smaller scale, allowing researchers to assess key parameters such as growth rates, product yields, and metabolic pathways under controlled conditions. By conducting these evaluations in a scaled-down environment, researchers can efficiently screen numerous variables and optimise bioprocess parameters before transitioning to costly and time-consuming large-scale production (Drobnjakovic et al., 2023) (Seidel et al., 2023) (Crater & Lievense, 2018; Delvigne et al., 2017).

Mathematical models are indispensable tools in bioprocessing and broader biological research. They provide a quantitative framework to understand better, optimise, and control complex biological systems. In biology, the complexity of living systems requires models to understand intricate interactions and emergent properties. By translating biological processes into mathematical equations, these models simplify complexity, enabling researchers to analyse interactions between genes, proteins, metabolites, and environmental factors (Torres & Santos, 2015).

They offer predictive power, allowing for the simulation of scenarios that help prevent costly failures and optimise processes, such as fermentation conditions or downstream purification steps. In addition, mathematical models are essential for hypothesis testing, enabling researchers to explore theoretical scenarios and design targeted experiments, thereby accelerating discovery (Almquist et al., 2014).

Additionally, mathematical models are essential for cost analysis and economic optimisation. They forecast the financial effects of process parameters and inform resource-efficient approaches. These models also evaluate environmental influences, aiding in the development of sustainable bioprocesses that reduce energy use, waste production, and greenhouse gas emissions (Wang et al., 2014).

Mathematical models offer a comprehensive method that improves efficiency, lowers costs, and guarantees sustainability, from the design of bioprocesses to their scale-up and final environmental assessment. Their capability to integrate data, forecast results, and optimise processes is vital for developing bioprocessing technologies, fostering innovation, and tackling global issues in healthcare, industry, and environmental sustainability.

1.1 Recombinant protein production within biotechnology

Producing recombinant proteins in expression systems comprises multiple stages. Throughout development, numerous choices are made to optimise yields and reduce production costs. Selecting the right strain is crucial to minimise bottlenecks and achieve high yields.

There are many expression systems used in bioprocessing. Bacterial expression systems include *Escherichia coli* (*E. coli)*, which is widely used, and Bacillus subtilis, which has promising secretion pathways (Khow & Suntrarachun, 2012). Yeast expression systems such as Saccharomyces cerevisiae, which is good at performing post-translational modifications, and Pichia pastoris, capable of reaching high cell densities, utilise strong promoters and efficiently secretes the protein of interest into the culture medium, simplifying downstream processing (Baghban et al., 2019). Insect and plant cells can also be used as expression systems for complex proteins that are known to be difficult to express in bacterial expression systems due to their structural, functional, or post-translational requirements which include multiple disulphide bonds (Cox, 2021; Hussain et al., 2017; Nosaki et al., 2021).

Mammalian expression systems, including Chinese Hamster Ovary (CHO) cells, are used widely in bioprocessing. CHO cells are very good at glycosylation as well as being scalable, making them ideal for the expression of protein-based therapeutics (Tihanyi & Nyitray, 2020). Human Embryonic Kidney (HEK) cells are becoming increasingly popular due to their use in viral vector production and gene therapy (Tan et al., 2021). The application of cell-free expression systems is increasing steadily, where pathways are removed in a controller manner to allow for the production of complex proteins (Brookwell et al., 2021).

Certain products, particularly recombinant proteins, may not occur naturally through the expression system. The industry employs various methods, such as restriction enzyme-based cloning, wherein DNA is digested, and the resulting fragments are ligated. (Sun et al., 2023). Recombination-based cloning can generate multiple plasmids through compatible recombinant attachment. SLiCE is a novel cloning method that utilises this feature to conduct gene cloning with high throughput (Zhang et al., 2012). Ligation-independent cloning, which does not utilise restriction enzymes and employs single-stranded complementary ends, was developed three decades ago (Aslanidis & De Jong, 1990).

A plasmid comprises several standard components that enable it to replicate and function within a host expression system. These components include an origin of replication, a promoter, a terminator, a multiple cloning site, and a selectable marker (Gleave, 1992; Maneewannakul et al., 1994).

Once the gene of interest has been cloned, it is inserted into a vector that the expression system will uptake. The choice of promoters is crucial. A common promoter used in *E. coli* is T7, while for mammalian cells, a CMV promoter is utilised (Chaudhary & Lee, 2015; Li et al., 2018). Promoters allow the binding of RNA polymerase and transcription factors required to express a gene (Rist & Kertesz, 1998). The number of plasmids per cell will also determine the gene dosage (Rouches et al., 2022).

Plasmids also carry genes for antibacterial resistance through a selection process in which antibiotics are added to Petri dishes, and the surviving colonies are likely to possess the gene of interest (Bennett, 2008). Plasmid copy number is a fundamental metric during the design step it governs plasmid stability (Friehs, 2004).

Molecular chaperones, which aid the folding process of the recombinant protein, are also expressed in the plasmid (Kyratsous et al., 2009). Inducers are used for some systems where an external trigger, such as IPTG, is required to initiate the expression of the protein. Vector construction in mammalian cells will also require enhancers and polyadenylation signals. These additions have been optimised through the years and are used as a strategy to increase the expression of recombinant proteins (Wang et al., 2022).

In *E. coli,* the vectors are transformed into the cell, using heat shock or chemical methods (Kant et al., 2019). The transformed cells are grown on an agar plate containing antibiotics, which allows only the cells containing the plasmid to grow. Transfection is used for mammalian cells. This can be viral transfections such as lentivirus or non-viral, either chemical or mechanical (Chong et al., 2021).

Optimisation studies are required to ensure that the conditions at which the expression

systems operate will produce the highest yield possible (Packiam et al., 2020)The working temperature must be maintained to prevent the denature of cells and proteins. The pH must be controlled to avoid being too acidic or too basic, which can result in cells growing very slowly or not at all.

The medium must be optimised to ensure that sufficient nutrients are provided to support the growth of the cells and the production of the product of interest. Additionally, the cells will require oxygen for the metabolic reactions necessary for growth. These conditions can be established through small-scale studies to identify the optimal conditions for the expression system (Tripathi & Shrivastava, 2019). Once these conditions are set, large-scale production harvest and purification of the target protein are required before the product is formulated into the desired form.

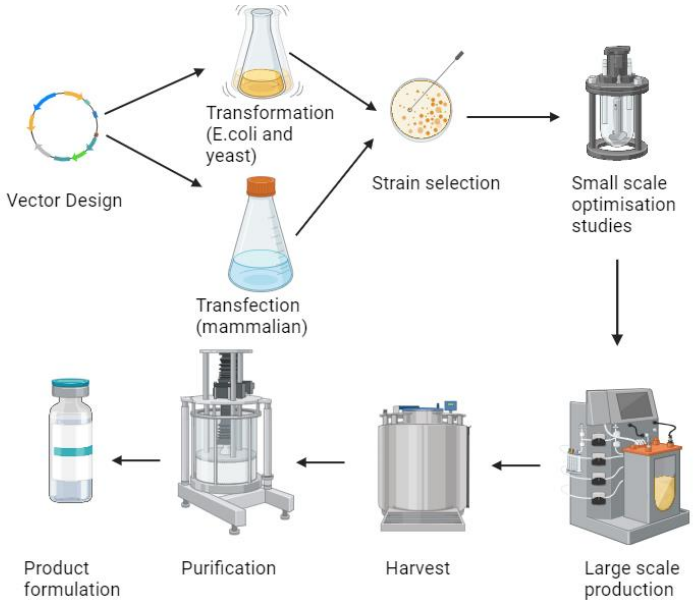


Figure 1. 1 Key Unit Operations in the Bioprocessing of Recombinant Proteins. This schematic illustrates the standard unit operations producing recombinant proteins, from upstream processing to final product formulation. (Biorender drawn image).

Figure 1.1 depicts common unit operations in the bioprocessing of recombinant proteins. However, it is crucial to understand the history of industrial methods used to develop these extensive processes, which have produced many products for everyday use.

1.1.1 Historical milestones in protein synthesis for bioprocessing

The expansion of protein synthesis into large-scale bioprocessing was made possible by the continuous advancement of industrial methods. These efforts included exploring innovative technologies in selection methods, vector design, scale transfer from small to large, and purification techniques; as mentioned previously, optimising parameters is necessary to enhance protein yield. Furthermore, developing analytical methods has played a crucial role in the biotechnology industry. This section will outline key milestones from 1970 to 2000 that have established the foundation for today's recombinant protein-expressing methods across various expression systems. The developments highlighted in Figure 1. 2 mark the progression of bioprocessing.

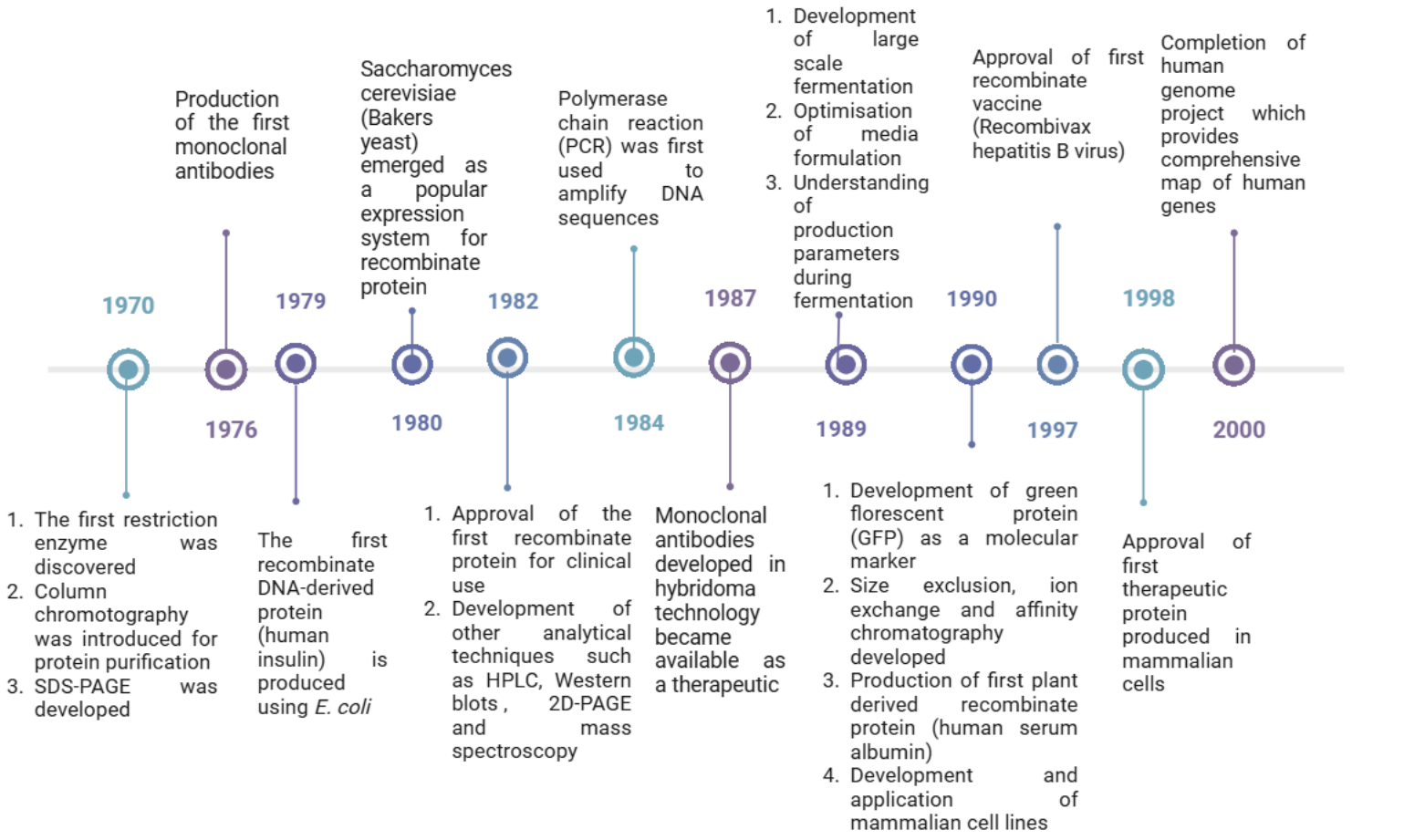


Figure 1. 2 Evolution of Key Biochemical Methods (1970–2000): Foundations of Modern Industrial Standards. This timeline illustrates the development of pivotal biochemical methods between 1970 and 2000, profoundly influencing and shaping the standards used in contemporary industrial bioprocessing. (Di Felice et al., 2019; Gifre et al., 2017; Jay et al., 2019a, 2019b; Jungmeier, 2017; Kratz, 2016; J. K. H. Liu, 2014; Mitra & Tomar, 2021; Napier et al., 2002; Newswire, 2014; Ramachandra Murty, 2011; Ritacco et al., 2018; Singh et al., 2006; Tresse et al., 1998; Villafañez et al., 2019; Wu et al., 2016; Yaish et al., 2013; Zhu et al., 2020).

1.1.2 Current upstream processing methods in bioprocessing

Over the years, several methods have been developed to improve the protein yield during protein production and reduce operation costs. The unit operations that produce and purify a specific recombinant protein depend on the protein itself and its application. This results in long research and development studies to develop a tailored method for that protein of interest (Rosenberg & Cady, 2021).

Bioreactor systems have evolved. In addition to stainless steel bioreactors that are cleaned in place, single-use bioreactor applications have become increasingly popular due to their reduced risk of contamination and shorter cleaning times. However, it is important to note that using single-use bioreactors increases plastic waste, so a detailed cost analysis is necessary before transitioning to this method. (Nogueira et al., 2021; Soerjawinata et al., 2021; J. Yang et al., 2019).

Perfusion bioreactors are used in continuous fermentation, where a filter is attached to the bioreactor to remove 1 to 2 volumes per day. This method allows for high-density cultures and longer operation times, removing waste material and adding fresh nutrients daily. However, scaling up is challenging, requiring more media than other historical methods. Online sensors have allowed data collection during the fermentation process, including temperature and pH, and enable automatic feedback loops. (Born et al., 2023; Nașcu et al., 2022).

Real-time data monitoring has become a crucial aspect of bioprocessing recombinant proteins (Gerzon et al., 2022). It provides process parameters for the unit operation, which must be maintained within a specific range determined by small-scale research and development experiments. This also leads to the creation of key performance indicators in bioprocessing. However, it is important to note that online and offline methods like HPLC, mass spectrometry, and various assays are used to understand the quality of the produced product (Giovanni Campolongo, 2018; Ruhl et al., 2020).

Current methods have been developed to minimise variability in biological systems by utilising advanced data monitoring techniques. Quality by design is another approach that strives to establish reliable methods for regulating upstream processing by employing critical quality attributes. These methods determine when a bioprocess has failed due to deviations to guarantee higher product quality (Jagan et al., 2021; Horst et al., 2021). These current methods have used tools to drive their implementation in bioprocessing, which will be discussed in Section 1.3.

1.1.3 Advancements and limitations in the use of various expression systems within the biotechnology industry

In addition to the methods developed to increase protein yield during bioprocesses, the expression system impacts the quality and quantity of protein produced. Table 1.1 summaries bacterial, yeast and mammalian expression, excluding plant and insect systems, which are not widely used in bioprocessing, covering only 5 to 10% (Diamos et al., 2020).

Table 1. 1 Comparative Analysis of Expression Systems: Key Advantages and Disadvantages of Bacterial, Yeast, and Mammalian Platforms

Expression systems	Advantages	Disadvantages
Bacterial expression systems	• Rapid growth, and cost-effectiveness compared to other expression systems (Maksum et al., 2022). • Engineered strains have been developed to handle post modification (Soares et al., 2019) • Methods developed to improve codon optimisation, fusion tags and vector design (Jayaraj & Smooker, 2014)	• Protein misfolding and aggregation often led to the formation of insoluble inclusion bodies (Kolivand et al., 2020) • Degradation of complex proteins or those requiring specific folding pathways (Jong et al., 2017; McGinness et al., 2006). • Disulphide bond formation is still limited in *E. coil* (Nguyen et al., 2011)
Yeast expression systems	• Good post-translational modifications, including glycosylation, disulphide bond formation, and proteolytic processing	• Regulatory approval is challenging (Sandow et al., 2015; A. J. D. Silva et al., 2022)

	(Azevedo et al., 2020; Zhang et al., 2022) • Pichia pastoris, have robust protein secretion pathways, allowing for the secretion of recombinant proteins (Ito et al., 2022; Liu et al., 2022) • Easy to genetically manipulate, (Huang et al., 2018; Martins-Santana et al., 2018)	• Scaling up is difficult (Mas & Portillo, 2022; Raymond et al., 1998) • Overexpression of heterologous proteins in yeast cells can lead to cellular stress and toxicity (Saini et al., 2018; Unrean et al., 2018)
Mammalian expression system	• Ability to form complex post-translational modifications which means reducing the risk of immunogenicity because of resembles to human cells (Ivanov et al., 2004; Y. Wang & Huang, 2017) • Easy to scale up and produce high cell densities (Büssow, 2015; Gilbertson et al., 2006; Moffat & Sabatini, 2006)	• Higher cost compared to other systems because of specialise media and supplements (Aricescu et al., 2006) • Low protein yield and complex processing conditions (Almo & Love, 2014; Mathias et al., 2020) • Cell line instability (Dorai et al., 2012)

E. coli is a robust expression system which forms 30-50% of the total protein production in many bioprocessing applications. However, the bottlenecks of *E. coli,* listed in Table 1.1, limit its potential to express larger proteins which require post-translational modifications. Due to this, potential strains have been engineered, such as Shuffle and CyDisCo, which aim to improve post-translation pathways in *E. coli* (Li et al., 2019; Matos et al., 2014). Experimental studies have engineered plasmids to overexpress genes, which reduce cellular stress and increase protein production. This includes the overexpression of molecular chaperones which aid protein folding, for example, DnaK (Nagpal et al., 2015; Ramakrishnan et al., 2020).

1.2 Protein Synthesis Pathways in *E. coli* Aided by Molecular Chaperones

Protein synthesis commences with the transcription of DNA into messenger RNA (mRNA). This mRNA is subsequently translated by ribosomes, with the assistance of transfer RNA (tRNA), into a sequence of amino acids, ultimately resulting in the formation of a protein (Baggett et al., 2017). Chaperones like Trigger Factor (TF) and DnaK are important during protein translation to prevent misfolds and aggregation. This section will introduce the chaperones, which are a part of protein synthesis (Agashe et al., 2004).

In addition, after the peptide is formed, multiple modifications are required to obtain the fully folded protein. Translocation to the correct location prevents proteins from misfolding and forming inclusion bodies. Translocation can be through the Sec and Tat translocation system, where their respective chaperones can direct and navigate the protein to the correct location (Browning et al., 2017; Santos et al., 2019). Finally, any post-translational modifications, for example, disulphide bond formation, which also has a unit of proteins to assist the pathways and reduce mis-oxidised protein (Banaś et al., 2020; Kurokawa et al., 2001).

The role of chaperones in protein synthesis and post-translational modifications, which aim to minimise misfolded proteins and repair damage, is connected to understanding how to enhance recombinant protein production in expression systems.

1.2.1 Chaperones in protein synthesis

Proteins are essential molecules that carry out a range of functions within living cells. Proper protein folding is crucial for their functionality, and molecular chaperones are vital in facilitating this process. The heat shock response is a cellular mechanism that controls the levels of chaperones and proteases to maintain an optimal environment for protein folding. Chaperones such as GroEL/ES and DnaK regulate the activity of sigma 32, a transcription factor essential for the heat shock response. GroEL has been shown to bind to sigma 32 and directly influence its transcriptional activity, highlighting the intricate regulatory mechanisms involved in protein folding under stress conditions (Ravitchandirane et al., 2022)

DnaK, another crucial chaperone in bacteria, is part of the Hsp70 system and is essential for protein folding and cellular homeostasis. It plays a central role in the folding of newly synthesised cytosolic proteins. It is involved in the biogenesis of inner membrane proteins and the post-translational targeting of presecretory proteins. DnaK and its co-chaperones DnaJ and GrpE function in protein quality control by promoting the refolding of misfolded proteins and stimulating the proteolytic degradation of irreversibly damaged proteins (Mayer & Bukau, 2005).

In addition to their role in the heat shock response, molecular chaperones assist in protein folding under normal cellular conditions. Co-expression of chaperones like GroEL/GroES alongside over-expressed proteins was shown to enhance the yield of properly folded functional proteins (Nishihara et al., 2000; Soltani & Bundy, 2022). This is particularly important for large and complex proteins prone to misfolding and aggregation. Multidomain proteins, which undergo complex folding pathways, often require assistance from molecular chaperones to achieve their native conformation (Deng et al., 2020; Martínez-Alonso et al., 2010).

A research article presented the overexpression of GroEL- GroES chaperones improved the solubility of recombinantly produced enzymes in *E . coli* called arginine deiminases by 5-fold (Y. Wang & Li, 2014). Another study showed that the overexpression of DnaK/DnaJ chaperones with GroEL- GroES chaperones improved the solubility of recombinant hirudin-PA (Seyed Hosseini Fin et al., 2019).

The interaction between chaperones and recombinant proteins is essential in biotechnology, where efficient production of recombinant proteins is desired. Strategies involving the co-expression of chaperones alongside recombinant proteins were shown to improve solubility and yield, addressing common challenges such as protein aggregation and low productivity (Mamipour et al., 2017).

Chaperones, such as GroEL, are molecular machines that assist in protein folding by rescuing misfolded proteins from kinetic traps. These large oligomeric complexes undergo conformational changes driven by ATP hydrolysis, allowing them to bind, encapsulate, and release substrate proteins in a controlled manner. Understanding the mechanisms of chaperonins is crucial for developing therapeutic strategies targeting protein-folding diseases, as they are implicated in various conditions such as bacterial infection, autoimmune diseases, and protein aggregation disorders (Horwich & Fenton, 2020).

Overall, molecular chaperones play diverse and critical roles in cellular processes, from regulating protein folding under stress conditions to optimising recombinant protein production in biotechnology. Advances in understanding chaperone-mediated protein folding, including the roles of GroEL, GroES, and DnaK, have implications for both basic research and practical applications, offering potential avenues for developing novel therapies and biotechnological innovations.

1.2.2 Chaperones in Sec Translocation

The Sec pathway in bacteria facilitates the movement of unfolded and partially folded protein. This pathway plays a critical role in protein translocation of protein and membrane biology. The pre-proteins contain information in the form of a signal sequence, which directs the protein to the correct location (Crane & Randall, 2017).

Specific components within the Sec translocation make up the pathways depending on the single sequence of the protein. The first component is the Sec YEG, which consists of Sec Y, Sec E and Sec G; these chaperones form a channel allowing the protein to pass through the membrane from the cytoplasm to the periplasm. Although there are three individual components, combining these allows for very stable structures required during translocation (Boy & Koch, 2009).

Sec A is a type of protein with an ATPase function, and it is bound to Sec YEG within the membrane. However, when Sec A is attached to another Sec B molecule, it can utilise the energy from ATP and convert it into mechanical energy, which it uses to move the preprotein to the correct location (Reddy et al., 2019). Sec B is also an ATPase that binds the preprotein to Sec A, forming a complex that transports the peptide to the Sec YEG channel, which is then delivered to the periplasm (Muller, 1999).

The cell shape and lipid composition on the distribution of Sec proteins in bacterial membranes impact the translocation pathways. Furthermore, new techniques such as genomic integration and advanced microscopy have improved the understanding of Sec machinery localisation within cells. This has allowed scientists to understand the components of each part of the translocation system (Lycklama et al., 2012).

Sec translocation is crucial in protein trafficking across bacterial membranes and quality control systems. The Sec machinery mediates the translocation of proteins into and across the cytoplasmic membrane, supported by various quality control systems that ensure proteins remain unfolded during translocation. This provides further prevention methods for transporting aggregated proteins across the membrane (Jiang et al., 2021).

Sec and Tat translocation pathways fulfil different roles, with Sec capable of transporting unfolded proteins and Tat translocating fully folded proteins. The two pathways are viewed as independent because they target different structures. The signal peptide preceding each protein determines its final location, allowing the two systems to function as distinct pathways (Pradel et al., 2009) (Arhar et al., 2021). The Tat system is discussed in section 1.2.3.

In conclusion, the Sec pathway in bacteria is a fundamental component in protein translocation and membrane biology. Its structure, function, regulation, protein targeting, and cooperation with other pathways make it an important biological pathway.

1.2.2.1 Coupled Sec translocation pathway

Coupled translocation is also called co-translational translocation and is one of the two pathways within Sec translocation. It is usually preferred when signal sequences are very hydrophobic and single recognition peptides (SRP) attach to the preprotein whilst it is being translated, hence the name. There are receptors in the membrane, such as FtsY, which interact with the preprotein and SRP to translocate the unfolded protein to the periplasm. Once the preprotein has attached to the Sec YEG completely, the ribosome is released, and the protein is exported to the periplasm (Mermans et al., 2022; Sandhu et al., 2021; Ullers et al., 2007).

1.2.2.2 Uncoupled Sec translocation pathway

The second route the Sec system can employ is uncoupled translocation, also known as post-translational translocation. This pathway is favoured when the signal peptide is less hydrophobic, necessitating energy in the form of ATP to transport the protein correctly. Furthermore, translocation can only be initiated once the protein has been fully translated and the ribosome is no longer attached (Gowrishankar & Harinarayanan, 2004).

Sec B binds to the exposed hydrophobic regions to prevent the unfolded protein from misfolding. The Sec B and protein complex then attach to Sec A, which utilises ATP to transport the unfolded protein into the periplasm. The Sec YEG complex serves as the channel for exporting the protein (Johnson et al., 2020)

1.2.3 Chaperones in Tat Translocation

The twin-arginine translocation (Tat) pathway in *E. coli* provides a unique mechanism for exporting folded proteins across the inner membrane, in contrast to the Sec pathway, which exports proteins in their unfolded form. Before the discovery of Tat translocation, biotechnological applications relied solely on the correct functional performance of the Sec pathway. This limited the export of folded proteins, which are typically complex co-factors such as Dimethyl sulfoxide reductase (Dong et al., 2015; Ray et al., 2003) . Tat's ability to transport folded proteins opened up novel possibilities in biotechnology. Although the Tat pathway in *E. coli* has been extensively studied, its characterisation and understanding of some mechanisms are limited (Matos et al., 2012; Patel et al., 2014).

The Tat pathway depends on specific signal peptides that contain twin arginine motifs, directing proteins to the translocation machinery. Experimental studies characterised 29 putative Tat-specific signal peptides in *E. coli*. They demonstrated that some peptides could be directly exported via the Tat and Sec pathways, while others exhibited Tat

specificity. The charge distribution around the cleavage site influenced export promiscuity, emphasising the importance of signal peptide properties. Additionally, proteins needed to fold in the cytoplasm before Tat-mediated export, suggesting an intrinsic folding quality control mechanism within the Tat pathway (Freudl, 2018; Tullman-Ercek et al., 2007).

Like the Sec system, the Tat translocation pathways require several key protein components. Two such components are Tat A and E. Tat A creates the pore or channel that facilitates translocation and contributes to the pathway's stability. It interacts with Tat B and C to form a stable channel (Lindenstrauß & Brüser, 2009). The proteins Tat B and Tat C have similar functions. They both have an essential function in recognising and binding to folded proteins. It interacts with the signal peptides of the protein to be exported. In addition, Tat B and Tat C are crucial for the assembly and stability of the Tat translocase complex. It may also help regulate the activity and specificity of the Tat pathway (Bolhuis et al., 2001).

The role of Tat E and Tat D is not as well understood as that of Tat A to C. Tat E shares similar responsibilities with Tat A and structural similarities. Tat E may contribute to the flexibility and adaptability of the Tat system, possibly under specific environmental conditions or in response to certain substrates. However, there is limited data to support this (Ize et al., 2003). Tat D protein has a unique function compared to the other Tat proteins. It is located in the periplasm and is responsible for processing and maturing Tat substrates after translocation. Tat D acts as a peptidase or protease by cleaving signal peptides from exported proteins. Moreover, it plays a crucial role in ensuring the quality control of proteins by releasing only properly folded ones into the periplasm (Frain et al., 2019; Matos et al., 2009).

Overall, Tat translocation is an interesting pathway that may present alternative methods of transport that will reduce misfolded protein owning to its unique quality control aspects.

1.2.4 Chaperones in disulphide bond formation

E. coli is a commonly used organism for the production of recombinant proteins. It has well-known genetics and is easy to manipulate. However, one major challenge when producing complex proteins in *E. coli* is the formation of disulphide bonds. Disulphide bonds are essential for stabilising protein structures, ensuring proper folding and functioning of many proteins, that fall in the class of antibodies, enzymes, and growth factors (Berkmen, 2012; Denoncin & Collet, 2013).

Disulphide bond formation occurs in the periplasmic space and is, mediated by the disulphide bonding (Dsb) proteins. DsbA introduces disulphide bonds to substrate proteins, while DsbB regenerates oxidised DsbA. Once the DsbA enzyme forms the disulphide bond, it becomes reduced and dysfunctional until DsbB can transport the electrons to the cytoplasm to maintain the redox balance. The DsbA/DsbB system is coupled with the respiratory chain and uses ubiquinone as a cofactor. DsbA recognises

substrate proteins and transfers a disulphide bond, leading to proper folding. DsbB then oxidises DsbA, completing the cycle (Inaba, 2008; Inaba & Ito, 2008; Yazawa & Furusawa, 2019).

However, when recombinant proteins that may require multiple disulphide bonds are to be expressed, this system cannot function correctly, leading to mis-oxidised protein. *E. coli* can then use two other proteins in the Dsb family. DsbC is an enzyme in *E. coli* that corrects mis-oxidised proteins by rearranging incorrectly formed disulphide bonds. It acts as a proofreading enzyme, ensuring that only correctly folded proteins are released or inserted. The periplasm's redox state regulates DsbC's activity because it requires electrons to mend the mis-oxidised protein (Denoncin et al., 2014; McCarthy et al., 2000).

DsbD is a protein in *E. coli's* inner membrane that recycles DsbC's active site. It transfers electrons via redox-active cysteine pairs arranged in a CXXC motif and is a conduit for electron transfer between the cytoplasmic thioredoxin system and the periplasmic Dsb proteins. DsbD's activity maintains the periplasm's oxidative folding capacity and ensures disulphide bond formation (Mccarthy et al., 2000; Stewart et al., 1999).

Co-expression of chaperones and folding catalysts such as DnaK and DnaJ can improve the formation of disulphide bonds. These molecules help fold substrate proteins and prevent the accumulation of misfolded intermediates. Another method to promote correct disulphide bond formation in target proteins involves using fusion partners or molecular scaffolds. (Kurokawa et al., 2000) (Bhatwa et al., 2021).

Overall, disulphide bond formation pathways have been studied in *E. coli* for many years because improving this step will result in the manufacture of many complex proteins currently being expressed in complicated expression systems. One study demonstrated the expression of single-domain antibodies, which reached yields of 200 mg/L when the strain was engineered to have improved redox balance, leading to improved disulphide bond formation. (Zarschler et al., 2013).

1.3 Introduction to the use of mathematical models in biotechnology

Mathematical models are a representation of the understanding of certain occurrences. In mathematics, every aspect of an event is defined quantitively. This is different from the occurrence of a phenomenon in its respective discipline, such as biology, which will inherently possess variability even within the same colony of cells. The ability to model biological pathways requires the introduction of assumptions to translate biology into a formal language that can then be modelled and analysed (Fischer, 2008; Marion & Scotland, 2008).

Analytical models use equations to find the exact solution to mathematical problems using algebra and calculus. In biological systems, differential equations have been used

to show the rates of reaction as well as the accumulation of certain products (Bernardo et al., 2013; Goel et al., 2006). This is because differential equations can present dynamic situations, for example, the mathematical model for ecological population (Simpson et al., 2024). Statistical methods have been used to analyse the data generated from the biochemical methods throughout the development of bioprocessing. These have also been used to design experiments because they can form relationships between two variables using statistics and understand patterns generated (Hinkelmann et al., 2011).

Computational models, which use algorithms and methods to simulate real-world systems, have proven highly successful in biology because they can form approximations for complex systems. Models have been generated to test hypotheses, gain new insights and understand the interactions between systems, for example, using models to screen potential drug candidates by comparing parameters such as efficiency (J. Sun et al., 2022) (Brodland, 2015). Empirical models are derived from observed data to understand relationships between parameters, for example, in hypothesis testing comparing observed data and predictions from the model (Lema-Perez et al., 2019). Stochastic models may provide the probability of random events occurring, which makes them useful for studying dynamic systems like biological pathways

Metabolic models are representations of cellular metabolism, and many types of metabolic models are used in biotechnology. One type is constraint-based models developed in the 2000s, which require stoichiometric constants. A few techniques are used in bioprocessing, including flux-balanced analysis, elementary analysis, kinetic models, and metabolic flux analysis. (Orth et al., 2010) (Maton et al., 2022).

Recently, multi-scale models have merged. This includes multi-omics data, a powerful method for understanding the complexity of biological pathways. (Hasin et al., 2017; Subramanian et al., 2020). The development of these models started in the early 1990s but was implemented into industrial scenarios in the early 2010s. Section 1.3.1 further details critical milestones in developing mathematical models in bioprocesses.

1.3.1 Historical progress of the use of mathematical models in bioprocessing

Integrating mathematical tools into bioprocessing design and development has been common practice for decades. However, as technology advances, the bioprocess industry is also adopting new methods to streamline and form advanced mathematical representations of biological processes. This section will summarise some modelling approaches used over time in bioprocessing.

Figure 1. 3 outlines the development of modelling methods in bioprocessing. Constant progress has resulted in the development of several tools, which improve recombinant protein manufacturing. The tools used for many years and emerging tools are discussed in section 1.3.2.

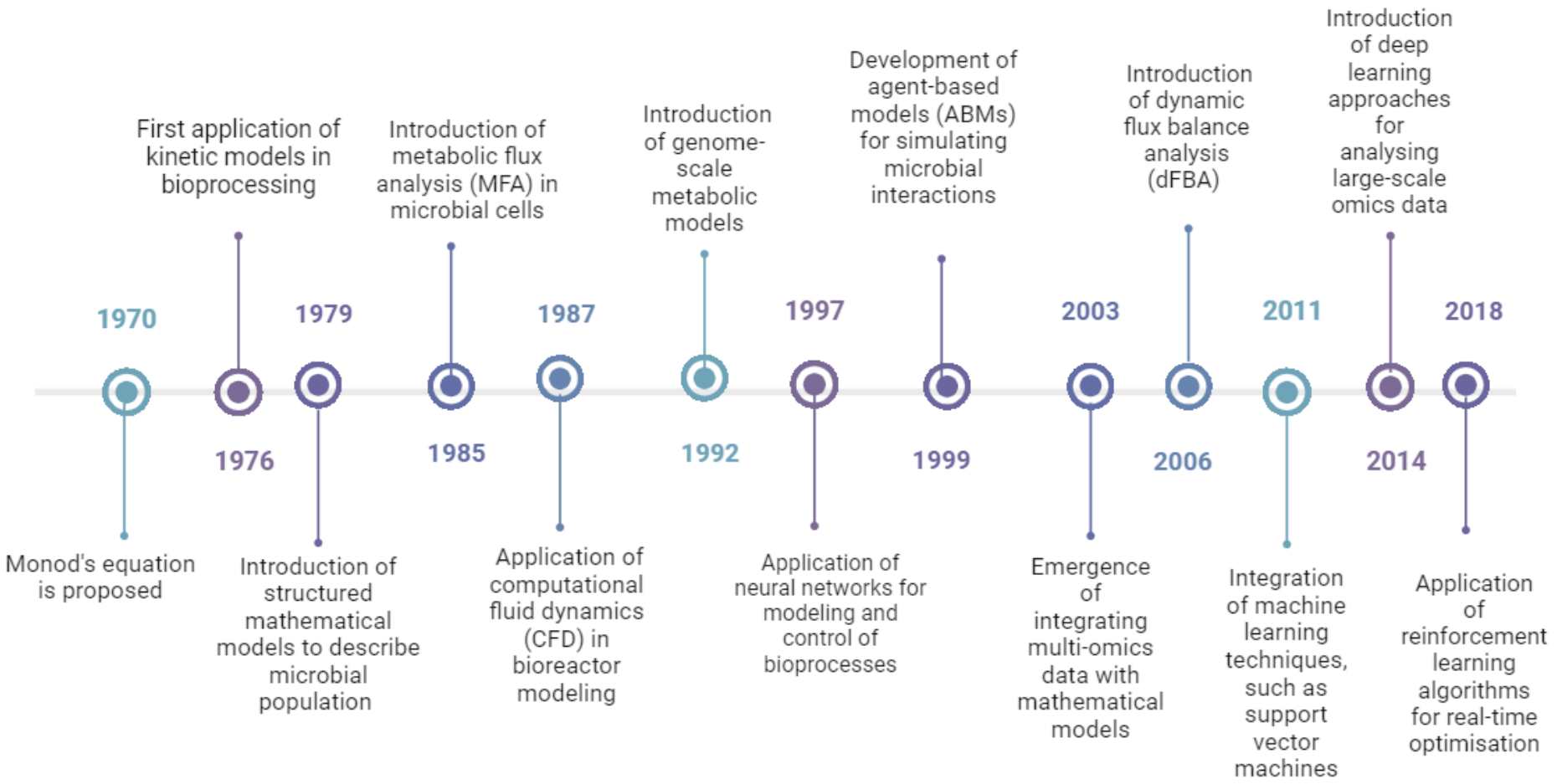


Figure 1. 3 Evolution of Mathematical Models in Bioprocessing (1970–2010s): This timeline depicts the development of mathematical models in bioprocessing from 1970 to the late 2010s, highlighting key milestones that have shaped their application in biological research and industrial processes. The references used to create the timeline (Agharafeie et al., 2023; González-Figueredo et al., 2018; Gorochowski, 2016; Kim et al., 2021; Kova´rova´- et al., 1998; Lee et al., 2006; Maton et al., 2022; Orth et al., 2010; Powers et al., 2019; Simeonidis & Price, 2015; Subramanian et al., 2020; Yuan et al., 2008)

1.3.2 Current trends used for optimisation in bioprocessing

Mathematical tools have transformed bioprocessing by allowing researchers and engineers to enhance processes, elevate product quality, and minimise expenses. These tools are generally classified according to their functionality and usage, highlighting a distinct evolution from fundamental modelling to sophisticated computational and machine-learning techniques. The application of these tools has been placed into categories based on their uses and development over time.

Tools for process simulation and optimisation are crucial in creating and enhancing bioprocessing workflows, such as upstream processing, fermentation, and downstream purification. A notable example is SuperPro Designer, which simulates batch, fed-batch, and continuous processes through differential equations, mass balance, and kinetics. This tool enables users to optimise parameters like temperature and nutrient levels to boost product yield and efficiency (Siletti & Petrides, 2008).

Aspen Plus is another popular tool that illustrates dynamic processes and assesses the effects of changing parameters in real time. It is handy for supervising and optimising bioprocesses, helping to maintain consistent product quality (Aspen Plus, 2000). Aspen DMCplus utilises feedback loops for superior process control, boosting stability and efficiency. This tool allows for real-time modifications, decreasing variability and enhancing overall performance. (Henson, 2008).

Design of Experiments (DoE) tools systematically plan experiments and analyse data to optimise bioprocessing conditions. Tools like JMP and Minitab facilitate the design and analysis of experiments, making it easier to identify optimal conditions for maximising yields and process robustness. Initially developed for agricultural applications, DoE was quickly adopted by the pharmaceutical industry in the early 2000s due to its ability to improve experimental efficiency and reproducibility (Kim & Kalb, 1996) (Christensen, 2015). The importance of DoE has been further recognised by regulatory agencies such as the FDA and EMA, which now emphasise its use in ensuring product quality and consistency (Fukuda et al., 2018) (Heidelberg, 2021).

In a similar timeline, optimisation algorithms are critical for solving complex problems in bioprocessing, such as resource allocation and process design. Techniques like linear, non-linear, and mixed-integer programming are commonly used to optimise bioprocessing scenarios. These methods are often implemented in MATLAB, which provides a versatile platform for solving optimisation problems (Sevella & Bertalan, 2000). Genetic algorithms, which mimic natural selection to find optimal solutions, have also been applied in strain engineering and process optimisation. These algorithms are beneficial for identifying the best genetic modifications to enhance product yields (Katoch et al., 2021).

Bioinformatics tools have transformed DNA and RNA sequences' alignment, analysis, and optimisation. One of the initial and most popular tools, BLAST (Basic Local Alignment Search Tool), is utilised for aligning and comparing DNA or protein sequences. Since its creation in the 1990s, BLAST has established itself as a fundamental resource for sequence similarity searches (McGinnis & Madden, 2004).

More specialised tools like the Burrows-Wheeler Aligner (BWA) use probabilistic methods and machine learning to efficiently align sequences. BWA facilitates high-throughput sequencing analyses for genome alignment workflows (H. Li & Durbin, 2009). Minimap represents the latest development in transcriptome mapping and alignment of long-read sequencing data (H. Li, 2018).

Specialised tools have greatly improved the design and optimisation of genetic constructs for recombinant protein expression. For example, VectorBuilder is a commercial platform that facilitates plasmid design through an easy-to-use interface. Users can select plasmids from a library or customise them with desired promoters, reporter genes, and selection markers (VectorBuilder, 2016).

Codon optimisation tools have gained prominence, particularly after recognising the critical role of codon bias in protein synthesis efficiency. Like those created by Puigbò et al. (2007), online servers make it easier to optimise codon usage for enhanced protein expression. Recently, deep learning models have been employed to optimise codons in *E. coli*, further improving protein synthesis efficiency (Fu et al., 2020).

Integrating machine learning and deep learning into bioprocessing represents a leap forward in this field. These technologies are used more frequently for predictive modelling, real-time monitoring, and optimisation. For example, machine learning algorithms can forecast process outcomes and identify the best conditions, enhancing the efficiency and reliability of bioprocessing (Aspen Plus, 2000).

The evolution of mathematical tools in bioprocessing reflects advancements in technology and methodology. Initially, these tools concentrated on fundamental kinetic models and mass balance equations, establishing a solid foundation for comprehending and optimising bioprocesses. Over time, they have integrated dynamic modelling, optimisation algorithms, and machine learning, permitting more refined and precise control over bioprocess workflows. Including interdisciplinary approaches that combine concepts from mathematics, computer science, and biology has enhanced the capabilities of these tools.

In addition, the shift towards automation and real-time control has improved process stability and efficiency. Concurrently, the emergence of bioinformatics and AI techniques has transformed sequence analysis, plasmid design, and codon optimisation. Adopting tools like DoE and advanced process control within regulatory and industrial frameworks has facilitated their extensive application, further propelling innovation in the field.

1.4 Motivation

The choice of *E. coli* as an expression system for recombinant proteins has many advantages, as described in Section 1.2.1. However, the molecular burden during the production of proteins is a obstacle, resulting in lower yields of correctly folded proteins (Peleg & Unger, 2012). The protein profile comprises unfolded, folded, misfolded proteins, aggregates, inclusion bodies, and mis-oxidised proteins as discussed in Section 1.2. Over the years, research and development have concentrated on optimising protein yield. Common strategies include enhancing vector design and promoter strength, optimising codons, adding fusion tags, fine-tuning fermentation parameters, and overexpressing chaperone systems that support protein folding pathways, as detailed in Section 1.2.1 (Jia & Jeon., 2016) (Pouresmaeil & Azizi-Dargahlou., 2023).

The overexpression of chaperones is an integral part of the protocol of plasmid design because specific chaperones can increase the yield by forty-two fold, as mentioned in Section 3.2.2 (de Marco et al., 2007). However, the chaperone requirement is specific to the type of protein being expressed, and overexpression does not always lead to increased yield. Moreover, burdening the plasmid further with genes that do not contribute to increased yields results in inefficient protein expression systems (F. Silva et al., 2012). Therefore, investigating chaperone systems becomes vital in understanding why there is a forty-two-fold increase in some cases and no difference in protein yield in others.

During plasmid design, central chaperones, such as DnaK and GroEL, are overexpressed because they are known to impact the vitality of the expression system (Schramm et al., 2017). Although chaperones move the yield of recombinant protein and the health of the expression system, a streamlined understanding of how specific chaperones affect the yield of a particular recombinant protein is yet to be explained. In addition, a limited knowledge of these pathways, paired with overexpressing chaperones within the plasmid, has resulted in over fifty per cent of proteins showing no difference in yield (de Marco et al., 2007; Ravitchandirane et al., 2022; Son et al., 2023). This highlights the importance of updating plasmid design protocols to ensure that each design step is one step closer to increased protein yields.

These chaperone systems reverse damage caused by stress by utilising ATP to unfold misfolded proteins, providing a second chance to form correctly folded proteins. Understanding these pathways and how they can be controlled to generate correctly folded proteins within *E. coli* can reform the protocols used during plasmid design.

1.5 Aims and Objectives

This thesis plays a vital role in bioprocessing by tackling issues in recombinant protein production, including protein misfolding, aggregation, and ineffective disulphide bond formation, which can limit yield and functionality. The research examines the precise folding pathways of recombinant proteins in *E. coli*, aiming to streamline the functions of specific chaperones involved in protein folding, translocation, and disulphide bond formation through a flexible mathematical model. This model empowers researchers to simulate different scenarios, such as adding or removing chaperones and modifying folding rates, facilitating quick evaluations of experimental designs and optimising protein yield. Integrating published data with Petri net formalism, this interdisciplinary approach connects biology, engineering, and computational modelling, encouraging innovation and collaboration. The results have wide-ranging implications, from enhancing therapeutic protein production and refining plasmid design to guiding strain engineering.

In addition to practical benefits, this research deepens our understanding of protein folding pathways and chaperone actions while promoting sustainable and cost-effective bioprocessing. This work enriches scientific insight and industrial progress by offering a computational tool to boost protein production efficiency, establishing a benchmark for integrating modelling and experimental biology to tackle intricate biological challenges.

The following objectives were met to achieve this aim.

- **The evaluation of the importance of chaperone-aided folding, including translocation and disulphide bond formation of recombinant proteins in *E. coli*, involved streamlining the available data**. This process involved collecting quantitative and qualitative data while highlighting discrepancies in published data over twenty years.
- **Various methodologies were tested to select a fitting modelling approach based on the collected data**. This process led to the creation of Petri net models because of the different formats of the data pathways. Petri net models can effectively address these gaps within the data sets. The initial attempt used PIPE5, which allowed the designing of Petri nets but could not generate a simulation of the model. Subsequently, Python models were developed where places and transitions could be defined as classes, although adding and deleting elements proved challenging. Finally, Snoopy was deemed adequate for designing and running Petri nets and was identified to be the best-suited for modelling these biological pathways.
- **Identifying the importance of each chaperone group and its connection to recombinant protein yield.** Smaller units were constructed to comprehend each pathway in detail and the relationships between them. This step certified that the various pathways functioned within their system boundaries before being implemented into a larger model.

- **A model protein, Human growth hormone (HGH), was investigated by constructing protein synthesis pathways specific to HGH within *E. coli*.** This included the Sec translocation pathway and the formation of four disulphide bonds. This demonstrated how the model could be adapted to represent various recombinant proteins and highlight the roles of specific chaperones, which could influence protein yield.
- **The model's applicability as a tool within the biotechnology industry was evaluated.** This was achieved by testing various bioprocessing scenarios and comparing the model's results to trends in published literature.

The remainder of the thesis is structured as follows.

Chapter 1 presented the necessary background information detailing the decision process to investigate microbial expression systems and the need for bioprocesses to increase product yields, making them more accessible. It also provided an overview of the techniques for achieving better yields and explored the current mathematical models available to improve protein production. The chaperone groups involved in reversing misfolded and aggregated proteins into unfolded proteins steered the project's motivation.

Chapter 2 explains the methodology used to construct the mathematical model. Creating this thesis's mathematical model involved navigating various software selection and functionality challenges. Petri net models were chosen because they effectively represent intricate biological pathways, even when dealing with incomplete or heterogeneous data sets. Petri nets are adept at filling data gaps through a graphical framework that demands minimal inputs, such as stoichiometry, while capturing the system's essential dynamics. This characteristic made them especially suitable for modelling protein folding pathways, where molecular complexes were designated places and reactions as transitions.

The first model implementation utilised PIPE5, a tool crafted for Petri net design. Although PIPE5 facilitated the design and visualisation of Petri nets, it fell short in simulation capabilities, a crucial aspect of this project. This shortcoming led us to transition to Python, where places and transitions were structured as classes, allowing the model to simulate different scenarios. While Python provided flexibility and access to robust libraries, incorporating and removing elements from the model was labour-intensive and inefficient, emphasising the requirement for a more user-friendly tool.

Ultimately, Snoopy was the most appropriate software for this model structure. Snoopy enabled Petri net design and offered strong simulation capabilities, overcoming the limitations seen in both PIPE5 and Python. Its user-friendly interface and capacity to manage complex biological pathways made it the optimal choice for modelling the protein folding system. Using Snoopy, Petri net models were created to investigate various scenarios, such as the overexpression or elimination of specific chaperones and their effects on protein yield. This iterative software evaluation and selection process highlights the importance of utilising the right tools to meet project objectives and demonstrates the adaptability needed in interdisciplinary research.

Chapter 3 collected over twenty years of available data on these pathways to streamline the chaperone groups and understand the interconnections between protein synthesis, translocation, and disulphide bond formation to be processed for secondary use. This chapter discussed in detail the varying nature of the data available between these pathways and emphasised discrepancies between published information. The literature search made it evident that there were many inconsistencies in the methods used to report data, impacting the quality of data collected. The absence of databases for critical pathways, such as the Sec and Tat and gaps in databases for disulphide bond formation demonstrated the importance of strategically gathering experimental data. Collecting data in a rigorous format to confirm the gaps within the literature was essential when developing the novel mathematical model.

Chapter 4 focused on developing a novel Petri net structure of the pathways defined in Chapter 3. Chapter 4 dealt with the multiple versions of the model established before the working model was finalised. This included formulating smaller models to confirm that the chaperone units could work within their smaller system boundary. The larger structure was developed with a model protein, Human growth hormone (HGH), which had 191 amino acids and four disulphide bonds. HGH was a relatively small protein that utilised the Sec translocation pathway, which was ATP-dependent. The Petri net structure could select these required routes and demonstrate the amount of protein produced at varying conditions. The model's performance was evaluated using energy-limiting scenarios and compared to trends in the literature.

Chapter 5 evaluated the utility of the model. This chapter investigated the tool's novelty by determining what it could reliably predict by modelling common bioprocess problems using the Petri net structure. It included a critical analysis of the tool by testing cases from the literature to determine whether the outcomes produced were consistent or if further tests were required.

Through the test cases, the Sec translocation was shown to be the step that limited the process. A test involving varying hydrophobic signal peptides was conducted to demonstrate their impact on protein yield. Engineered strains, such as CyDisCo, which utilised Tat translocation by reducing the energy barrier of ToA (a potent signal peptide), showed promising results, which would be explored within the model. The formation of disulphide bonds in the cytoplasm affected the cell's redox potential. The role of glutathione was investigated because its deletion could result in an imbalance of redox potential within the periplasm. The deletion of DsbA and DsbC from the model was analysed to understand if this matched the expected literature results. Finally, the overexpression of folded chaperones was investigated to understand the full scope of the tool and its limitations. Five hand-picked cases were presented to establish the application as a proof of concept.

Chapter 6 discussed the novel findings and whether the project's overall aim was met. It also discussed future work to expand the model. Finally, a list of references and an appendix followed the thesis's conclusion.

Chapter 2

2. Methodology

A systematic approach is required to represent the complex challenges highlighted in Chapter 1 through a mathematical structure. This involves defining a methodology for collecting the necessary data and developing a mathematical model using Petri net formalism. This Chapter will focus on the scientific methodologies developed to construct the model suitable for the nature of the data collected.

The constructed model includes six biological mechanisms for folding and transporting the target protein to the periplasm. This encompasses three distinct protein synthesis pathways: translocation pathways such as Sec and Tat, and pathways for disulphide bond formation. While these pathways are interconnected during protein expression in *E. coli*, the approaches used to gather data for each pathway vary, leading to discrepancies in the collected data. Chapter 3 will explore these discrepancies and present assumptions that will standardise the data.

The aim is to develop a model that can simulate the interactions of molecular chaperons during protein synthesis. The model will demonstrate a protein's journey from translation to disulphide bond formation. In addition, the pathways within protein synthesis that aim to reduce misfolded and aggregated proteins will be modelled.

Although these pathways have been linked in the literature using diagrammatic representation, this alone cannot be used as a mathematical model. Therefore, a framework that can gather the rates of each pathway and the ratio between each biological unit to translate these into a mechanistic model is required.

A framework was used to collect the data for each pathway, and an example of how the data is collected for one part of a pathway in the model is shown below.

Step 1: Gather information about the system from literature, which is collected in small chunks to understand a part of the pathway.

- E.g. Tat signal peptides have a lower hydrophobicity compared to Sec-specific signal (Cristó et al., 1999)
- Sec signal sequences are shorter compared to Tat signal sequences (Cristó et al., 1999)
- It has been observed that the overall charge of the c-region plays a crucial role in determining the selectivity of the Tat pathway, as revealed by the analysis of 29 predicted *E. coli* Tat signal peptides (Lee et al., 2006)
- A negative stain was used to discover the initial structures of the Tat translocation pathways. The findings revealed that TatABC forms a complex that facilitates the complete translocation process (Sargent et al., 2001)

- Additional studies revealed that the Tat system comprises four distinct units: TatA, TatB, TatC, and TatE.(Bageshwar et al., 2016)
- Another research study indicates that TatA and E have many similarities in structure, and the choice of chaperone used depends on the protein's charge. (Palmer & Berks, 2012)
- The reactions to transport the protein from chaperone to chaperone require ATP and transport through the membrane requires the proton motive force (Rose et al., 2013).

Step 2: Decide what the model will represent and omit. Identify the gaps with further evidence from the literature of the most used representation.

- For example, Tat E is not used unless Tat A is unable to function at certain conditions, but these conditions have not been defined in literature therefore, including this can lead to illogical assumptions (Blaudeck et al., 2005).

Step 3: Topological representation of the pathway - the inputs and outputs of each step

Step 4: Translated each step into a stoichiometrically- balanced equation.

Step 5: Implement Petri nets. Define the states of the chaperones and the reaction required to move the protein along in the system. Per the steady-state operation assumptions, each branch must be balanced, and nothing can accumulate within the system boundary.

Petri nets present a compelling alternative for modelling complex biological systems, particularly in the context of protein synthesis, due to their unique combination of visual representation, minimal data requirements, and inherent flexibility. Unlike traditional dynamic models such as ordinary differential equations (ODEs) or agent-based approaches, which often require extensive parameterisation and can become unwieldy for large systems, Petri nets utilise a graphical structure that clearly illustrates biological pathways' components and interactions. This visual clarity is especially valuable for modelling intricate processes such as chaperone-assisted protein folding, translocation, and disulphide bond formation.

One of the key strengths of Petri nets is their reliance on stoichiometry rather than detailed kinetic data. This makes them particularly well-suited for modelling protein synthesis pathways, where comprehensive experimental data is often scarce. Obtaining detailed kinetic parameters for every reaction in a biological system is challenging, as it requires extensive experimental work and may not always be feasible.

Traditional models like ODEs depend heavily on such data, which can limit their applicability in systems where information is incomplete or uncertain (Alur & Henzinger, 1999). In contrast, Petri nets focus on the stoichiometric relationships between molecular entities, allowing researchers to capture the essential dynamics of the system even with limited information (Toni et al., 2008). This practical approach enables the exploration of biological phenomena without the need for exhaustive data

collection, making Petri nets a powerful tool for systems where data availability is a constraint.

Another advantage of Petri nets is their flexibility. Researchers can easily modify the model by adding or omitting specific components, such as chaperones or pathways, to test hypotheses and explore different scenarios. This adaptability is crucial for understanding how changes in molecular interactions impact protein folding and aggregation. For example, the model can simulate the effects of overexpressing or deleting specific chaperones, providing insights into their roles in reducing misfolded and aggregated proteins. Such flexibility is often challenging with other modelling approaches, which may require extensive reprogramming or restructuring to accommodate changes (Liu & Heiner, 2013) (Paolo Baldan et al., 2010).

Petri nets also excel in modelling biological systems' concurrent and parallel processes. Protein synthesis involves multiple simultaneous pathways, such as translation, folding, and quality control, which interact dynamically. Petri nets naturally represent these concurrent processes, capturing the synchronisation of various pathways and interactions. This capability allows the model to accurately depict how chaperones assist nascent proteins during translation or how misfolded proteins are targeted for degradation. By encapsulating these dynamics, Petri nets provide a more comprehensive understanding of the multifaceted nature of protein synthesis and quality control.

The systematic methodology of Petri nets further enhances their utility by enabling the seamless integration of existing literature into the model. Parameters can be grounded in empirical data, ensuring the model is credible and computationally efficient (Hardy & Robillard, 2004). This integration of experimental evidence is a critical advantage over traditional modelling methods, which may struggle to incorporate diverse data sources effectively. By leveraging published data, Petri nets provide a robust foundation for simulating biological processes, making them a reliable tool for hypothesis testing and optimisation (Matsuno et al., 2003).

This modelling technique represents a novel and practical approach to unravelling the complexities of biological systems. Their visual framework, efficiency in handling data limitations, adaptability in modelling scenarios, and ability to accurately simulate dynamic interactions distinguish them as a superior choice for studying protein synthesis. By leveraging the strengths of Petri nets, it can illuminate the intricate journey of proteins from translation to final folding, ultimately revealing how molecular chaperones impact folding fidelity and address issues of aggregation in bioprocessing. This innovative method advances the understanding of protein synthesis and paves the way for new insights into protein optimisation in various biological and industrial contexts.

2.1 Petri nets

Carl Adam Petri, a computer scientist, invented Petri nets in 1962 to mathematically and visually represent concurrency, communication, parallel systems, and synchronisation. (Giua & Silva, 2018). Petri nets facilitate the understanding of data flow across various points while also identifying deadlocks and reachability issues. This section of the chapter outlines the components that constitute Petri nets and describes the various types of Petri nets explored based on the necessary modelling structure.

2.1.1 Structure of Petri nets

Petri nets follow a specific structure to represent a process. This structure includes two central nodes, a place and a transition. A place is represented by a circle, which can denote a location. In contrast, a transition is represented by a rectangle or square, representing an event that represents movement from one location to another (Juhh, 1998).

The places and transitions must be connected by arcs and arrows that connect the place to the transition and the transition to a place. The places can be marked with tokens, indicating that tokens are present within them. Tokens can accumulate and deplete; they represent the number of units of that place. For example, if a place is taken as a delivery centre and there are 10 tokens within that place, this shows there are 10 parcels within the delivery centre.

Sometimes, the arcs can weigh two tokens, meaning that the transition cannot occur unless two tokens are transferred from the place before the transition (Hu et al., 2011b). Figure 2.1 demonstrates the basic structure of Petri nets. Figure 2.2 shows how a balanced equation is represented using Petri nets,

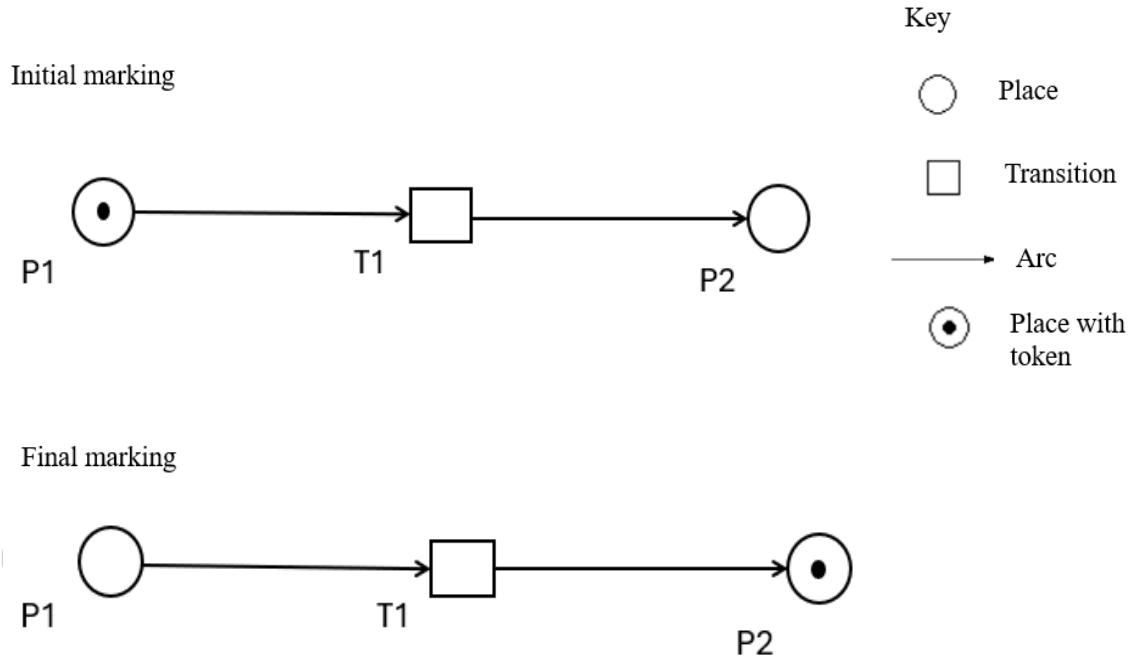


Figure 2. 1 Illustration of Token Firing in a Petri net Model. This figure demonstrates the fundamental mechanism of a Petri net, where a token moves from Place 1 (P1) to Place 2 (P2) through Transition 1 (T1). The final marking is achieved once all pathways have fired, indicating the completion of the process. The absence of weights on the arc signifies a one-to-one stoichiometric ratio between the input and output.

$$Equation\ 1$$

$$2C + 2O_2 = 2CO_2$$

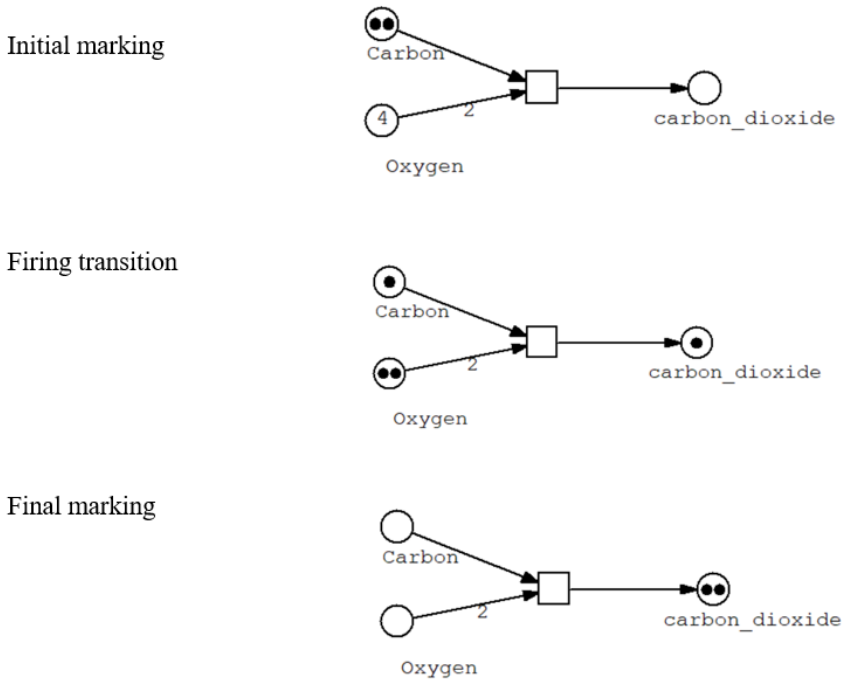


Figure 2. 2 Representation of Balanced Chemical Equations Using Petri nets. This figure illustrates how Petri nets can effectively model balanced chemical equations. Each place corresponds to a chemical element, while each transition represents a reaction that transforms these elements into a compound. The transition models the stoichiometric conversion of reactants (input places) to products (output places), ensuring the equation remains balanced.

Petri nets can also be represented as an incidence matrix. This is possible when.

P: is a set number of places

T: is a set number of transitions

M: is the incidence matrix where P is the number of rows and T is the number of columns.

The incidence matrix shows the tokens added or removed from a place. A token removal is represented by -1, whereas a token addition is represented by +1. If no tokens have been added or removed, the value is 0. The incidence matrix is calculated by subtracting the input matrix from the output matrix. Each matrix analyses the relationship between each transition and a place. In this demonstration, Figure 2.3, representing a Petri net system, will be interpreted as an incidence matrix shown in Table 2.1. For instance, T1 takes input from P1 but has no output for P1. T1 has outputs for P2 and P3 but no input or output for P4 (R. et al., 2005). The tables are displayed below.

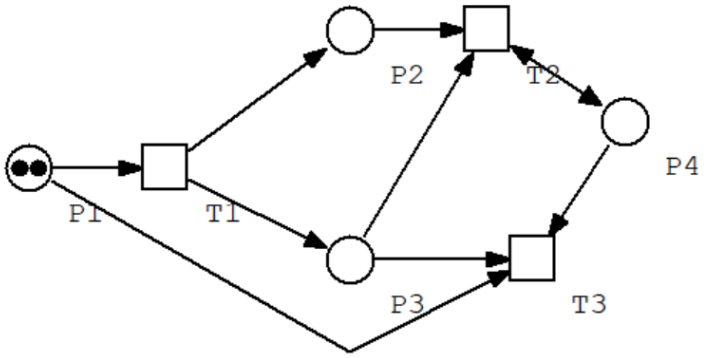


Figure 2. 3 Petri net Structure with Four Places and Three Transitions at Initial Marking. This figure depicts a Petri net structure consisting of four places (P1, P2, P3, P4) and three transitions (T1, T2, T3) at the initial marking stage.

Table 2. 1 The table presents the incidence matrix for the Petri net illustrated in Figure 2.3. The incidence matrix is calculated by subtracting the input matrix (tokens consumed by transitions) from the output matrix (tokens produced by transitions). Each row corresponds to a place (P1, P2, P3, P4), and each column corresponds to a transition (T1, T2, T3). The matrix elements indicate the net change in token distribution across places when transitions fire.

Output	P1	P2	P3	P4
T1	0	1	1	0
T2	0	1	0	1
T3	0	0	0	0

Input	P1	P2	P3	P4
T1	1	0	0	0
T2	0	1	1	0
T3	1	0	1	1

Incident	P1	P2	P3	P4
T1	-1	1	1	0
T2	0	0	-1	1
T3	-1	0	-1	-1

Petri nets are a mathematical modelling tool used to represent complex systems. They are often represented as graphs that come in different structures. State graphs are used when each transition has only one input and output. An event graph represents places where there is only one input and output. A conflict-free graph represents where a place has at most one output and no competition for the tokens with other places. A simple Petri net is a representation where only one conflict is present. These structures can be combined to create hybrid models for complex systems (A. Spiteri Staines, 2011). This research will use a hybrid model to represent the system.

The Petri net formalism has been adopted in biotechnology due to its ability to represent a data flow from one place to another. Petri nets can successfully present biochemical pathways because they can show concourse. In addition, Petri nets can be represented as stochastic, which has been very useful in demonstrating biological processes (Heiner et al., 2008).

Furthermore, Petri nets can represent ordinary differential equations used to model cell growth, protein production, and an increase in metabolite during the fermentation processes of recombinant proteins (Jyotish et al., 2023). Figure 2.4 shows how to construct a Petri net model to represent biological systems and how biosystems can be created by modelling the pathways as Petri nets.

Petri nets have been used in metabolic pathways, for example, in a study on pentose phosphate pathways in *E. coli*, where the Petri net model can present key control points and regulatory mechanisms (Hardy & Robillard, 2004). Another example of Petri nets application in biology is where cell signalling pathways are integrated to decide cell fate (Li et al., 2007). These studies also followed the workflow shown in Figure 2.4.

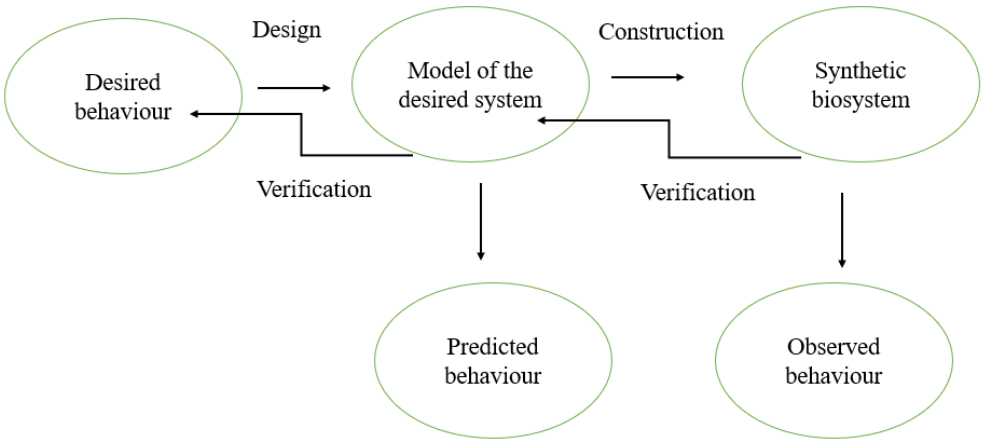


Figure 2. 4 Workflow for Model Construction Prior to Experimental Studies. This diagram outlines the workflow for constructing mathematical models before experimental work, reducing time and resource expenditure. This approach includes model design, construction and verification.

2.1.2 Types of Petri nets

Simple Petri nets consist of places, transitions, and arcs with a simple input-output relationship. They provide a clear graphical representation of the system structure, making them intuitive and easy to interpret. However, they cannot model continuous changes over time or detailed kinetic behaviour, which limits their applicability to complex biological systems. In contrast, differential equations (DEs) use continuous functions to describe the rate of change of system variables over time, making them ideal for modelling systems with continuous dynamics, such as metabolic pathways or gene regulation. While DEs offer precise quantitative descriptions, they require detailed kinetic data and can become computationally intensive for large systems. On the other hand, Simple Petri nets are less detailed but excel in visualising system structure (Cherdal & Mouline, 2018; Finkel & Schnoebelen, 2001).

Coloured Petri nets extend simple Petri nets by allowing tokens to carry additional data, such as quantities, values, or states. This enables more detailed and accurate representations of complex systems. Coloured Petri nets are considered high-level models because they identify potential bottlenecks and optimise system performance. However, they still lack the continuous-time dynamics of DEs, which can capture precise kinetic interactions and continuous changes. While DEs provide a continuous and quantitative description of system dynamics, they may not be suitable for systems with discrete or stochastic events. Coloured Petri nets, on the other hand, offer a balance between detail and flexibility, making them helpful in modelling complex interactions without requiring extensive parameterisation (Ahmed & Majid, 2019; Jensen, 1997).

Fuzzy-Coloured Petri nets incorporate kinetic parameters and handle uncertainties, making them suitable for incomplete or uncertain data systems. They bridge the gap between discrete Petri nets and continuous dynamic models by incorporating probabilistic elements similar to stochastic differential equations (SDEs). SDEs extend DEs by incorporating random noise or probabilistic behaviour, making them ideal for modelling systems with inherent uncertainty, such as gene expression or protein interactions. However, SDEs are computationally intensive and require detailed stochastic parameters. On the other hand, Fuzzy-Coloured Petri nets provide a practical alternative for modelling systems with missing or uncertain data, offering a balance between detail and computational efficiency (Assaf et al., 2022).

Timed Petri nets allow transitions to fire at specific times, mimicking the timing of real-world biological processes. This makes them suitable for simulating time-critical processes like metabolic pathways or signal transduction. Timed Petri nets are closer to time-dependent differential equations as they incorporate temporal dynamics. However, they remain discrete-event models, unlike the continuous-time nature of DEs. Time-dependent DEs provide a precise and quantitative description of time-dependent behaviour but require detailed kinetic data and may not capture discrete events effectively. On the other hand, Timed Petri nets offer a more intuitive and flexible approach for modelling time-dependent processes in biological systems (Zuberek, 1991; Lefebvre & Basile, 2018).

Stochastic Petri nets incorporate transition-firing probabilities, making them suitable for modelling random events in biological systems, such as gene expression or protein interactions. They are closely aligned with stochastic differential equations (SDEs), as both account for probabilistic behaviour. However, stochastic Petri nets are discrete-event models, while SDEs are continuous. SDEs provide a precise and quantitative description of stochastic behaviour but are computationally intensive and require detailed stochastic parameters. On the other hand, Stochastic Petri nets offer a more intuitive and flexible approach to modelling systems with inherent randomness, making them a powerful tool for simulating complex biological processes (Goss et al., 2007; Bause & Kritzinger, 2013).

2.1.2.1 Discrete Petri nets

Discrete Petri nets are suitable for data in discrete steps and events where the system state variable does not change until an event happens (Simon et al., 2018). These are frequently used in manufacturing processes as they can effectively demonstrate a simple input and output structure.

Concurrent nets usually use the discrete Petri net formalism because they can represent multiple co-occurring activities. They are also essential for modelling complex systems with many interacting components.

The main characteristic of using a discrete Petri net is that it deals with discrete parameters, which can lead to two possible scenarios. Firstly, parameters can be treated as tokens, representing discrete data units that move between places. Secondly, they can be treated as weights on the arcs, determining the number of tokens that can move across the arc (Giua & Seatzu, 2015).

Discrete Petri nets have various advantages, including their ability to model large and complex systems, simulate and analyse different scenarios, and their flexibility to accommodate changes in the system. Additionally, discrete Petri nets provide a visual representation of the system, which makes it easier for users to understand and communicate the system's behaviour with ease (N. David et al., 2015).

2.1.2.2 Continuous Petri nets

Continuous Petri nets are mathematical modelling tools representing systems with continuously changing parameters. These nets were shown to be helpful to model systems that require continuous fluid levels to operate since the transitions in the model have a rate of change that can adapt to these changes in fluid levels (Wang et al., 2016).

Continuous Petri nets were developed to address the limitations of discrete Petri nets in modelling dynamic scenarios. Discrete Petri nets could not provide a complete system

representation in such scenarios, leading to a new modelling tool that better captures real-world systems.

The structure of autonomous continuous Petri nets is similar to that of discrete Petri nets, with P representing a finite set of places and T representing a finite set of transitions. However, in continuous Petri nets, the evolution rule differs as firing is not restricted to integer amounts. This means that the marking is not forced to be an integer, leading to a new marking that better captures the continuous changes in the system parameters (Mahulea et al., 2005).

Overall, continuous Petri nets are powerful tools for modelling systems with continuously changing parameters. They allow for an accurate representation of real-world systems and can help identify potential issues or bottlenecks in such systems.(Blondin & Esparza, 2022; Recalde et al., 2007).

2.1.2.3 Hybrid Petri nets

A hybrid Petri model combines discrete and continuous elements and provides an interphase between them. The approach allows one to capture the properties of both discrete and continuous models in a unified framework. These models are commonly used in dynamic systems and have proved particularly effective in depicting biological networks.

Hybrid models have been used in various applications, from traffic control systems to chemical reactions. However, they have found a handy application in biological networks. The ability to model both shift pathways and pathways with a continuous data flow has proven helpful in depicting biological pathways (Brinkrolf et al., 2021).

In a hybrid Petri model, the discrete elements are represented as places, while the continuous elements are represented as transitions. Places tend to be described as discrete, while the transition rate is continuous. This way, the model can capture a system's discrete and continuous aspects in a single framework.

Hybrid Petri models are powerful tools for complex systems exhibiting discrete and continuous behaviour. They allow one to capture a system's behaviour at different granularity levels, from macroscopic to microscopic. This makes them particularly useful in biological systems, where the system's behaviour can vary greatly depending on the scale at which it is observed.

In conclusion, hybrid Petri models are powerful modelling approaches which allow one to capture the behaviour of complex systems that exhibit both discrete and continuous behaviour. They have found particular applications in biological networks, where they have proven helpful in depicting biological pathways. By allowing one to capture a system's behaviour at different granularity levels, hybrid Petri models provide a unified framework for modelling complex systems. (Alla, 2007; Herajy et al., 2018).

2.1.3 Structural analysis of Petri nets

If a Petri net has a limit on the number of tokens that can accumulate at any location, then it is known as a bounded Petri net. Structural analysis is the method used to determine whether a Petri net is bounded or unbounded. A bounded Petri net ensures that the system will not experience an uncontrolled growth of tokens in any location. (Lomazova & Popova-Zeugmann, 2015).

A deadlock can happen when no transition is possible, halting the system. Structural analysis techniques are used to study the network's connectivity and arrangement of places and transitions to spot potential deadlocks. One way to analyse deadlocks is to check for cycles in the reachability graph (Hu et al., 2011).

Liveness is a term used to describe a system's capability to trigger a transition by reaching a particular state. Structural analysis of the connectivity of transitions and places can offer valuable insights into the liveness of a Petri net. If specific structural properties are missing, it may indicate potential liveness issues that must be addressed.

Preserving tokens in specific locations or combinations of locations is an integral part of conservation. Analysing the system's structure can help identify conservation relationships based on the net's topology and expressed as algebraic equations. The conservation properties can impact the system's behaviour (Mahulea et al., 2008).

It is important to note that invariants are essential properties that apply to all reachable markings within a Petri net. Structural analysis is a valuable tool for discovering structural invariants, which are conditions that remain unchanged throughout the entire system execution. These invariants offer valuable insights into the behaviour of the network (Schmidt, 2003).

When examining a Petri net, reachability analysis is utilised to establish the feasible states the system can reach from a specific initial state. Concurrently, structural analysis is employed to identify the structure of the reachability graph, which maps out the conceivable state transitions in the Petri net (Czerwinski et al., 2018).

Structural analysis is a technique that aids in identifying a coverability problem in a Petri net. This problem arises when the system can enter states with infinite tokens. Additionally, coverability analysis extends the reachability analysis by allowing certain places to contain infinite tokens (Kleijn et al., 2013).

Concurrency is an inherent aspect of Petri nets. Structural analysis can uncover the relationships between transitions regarding concurrency. By scrutinising concurrent paths and transitions, and can gain valuable insights into the parallelism of the system (Rawson & Rawson, 2022).

2.2 Petri net-tools

The Petri net models have been published in BioModel and will be made public upon the publication of the thesis. The reference number and details of the construed model are provided in Appendix A.

Petri net models can be built using different tools, which are selected based on the model's specific application. Most of these tools have a graphical user interface, which makes them easy to use and helps visualise the Petri net structure.

Some user-friendly tools for creating, editing, and simulating Petri nets can be listed as follows:

- PIPE5: This tool supports discrete-event and continuous Petri nets and provides simulation capabilities (Lladó et al., 2022).
- Snoopy: This graphical editor and simulator allow users to create, edit, and simulate various Petri nets. It has a user-friendly interface (Goos et al., 2012).
- TAPAAL: This tool is designed for modelling, simulating, and verifying timed Petri nets. It supports the analysis of reachability, boundedness, and deadlock-freeness (Goos et al., 2009).
- CPN Tools: This tool is used for modelling, simulation, and analysis of Coloured Petri nets (CPNs). It supports hierarchical and non-hierarchical modelling and offers advanced simulation capabilities (Steinsland et al., 2022).
- WoPeD: This editor and simulator focuses on workflow modelling and allows users to design and simulate Petri nets. It also supports exporting models to various formats(Glory & Sheela, 2012).

Programming languages such as Python and MATLAB provide applicable packages to build Petri nets. In this work, the initial model was created using PIPE5, which displayed the pathway using places and transitions. However, PIPE5 could not successfully simulate the pathways. This included the ability to change the transition rates and the software not being able to run the pathways developed. Therefore, in the next development iteration, the model was constructed using Python, which could simulate the pathways effectively.

However, editing the model and adding new places was difficult in Python. The difficulty lies with the ability to make changes to the model. The transition is set in place, and one change in a place results in changing multiple places in Python mode. However, if biology is defined first and then construed, it can still be a good tool. In the case of this work, many pathways were not defined, and Python making constant changes was not practical when using the Python model.

As a result, the final model presented in Chapter 4 was built using Snoopy. Snoopy displays the pathways using graphical interphase and simulates them as well. Additionally, editing the pathways and varying the rates can be done quickly using Snoopy. Therefore, this tool was selected to construct the final model. The method

development strategy from the preliminary model to the final version is described in section 2.2.1, and the process of adopting and implementing different modelling platforms is described briefly below.

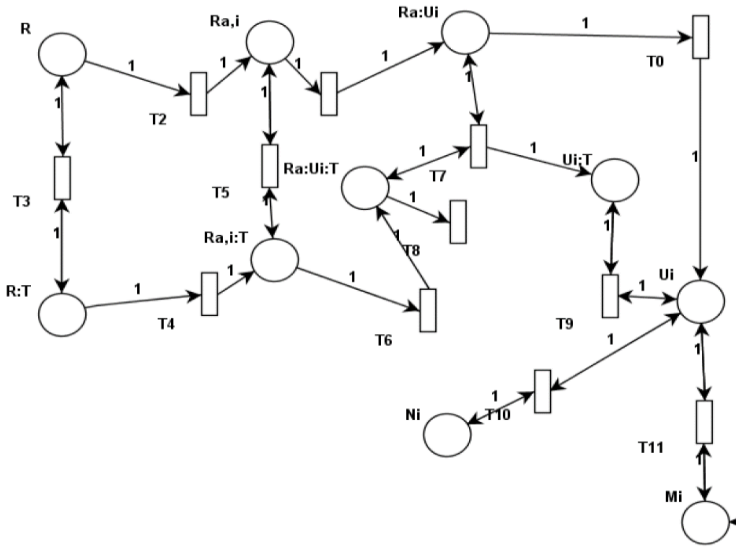
2.2.1 PIPE 5

PIPE 5 was the first tool used to model Petri nets using the data collected, which will be described in Chapter 3. The advantage of using PIPE5 is it has a user-friendly interphase, which is useful when using Petri nets for the first time however, PIPE5 is unable to handle complex Petri nets (Bonet & Lladó, 2007; Dingle & Knottenbelt, 2009).

Table 2.2 displays an example dataset on rates of steps in protein synthesis, which excludes molecular chaperones in its current form (Powers et al., 2012). Figure 2. 5 shows an example of how the tool was used to model the pathways.

Table 2. 2 Transition Rates of Steps During Protein Synthesis Without Additional Chaperones. This table outlines the transition rates for key steps in protein synthesis in the absence of additional chaperones. Where R= Ribosome, T= Trigger Factor, Ra,i = Activated Ribosome, Ui = Unfolded Protein, Mi = Misfolded Protein and Ni = Native Protein

Reactions	Transition value
$R + T \longrightarrow R:T$	0.02 $\mu M^{-1} s^{-1}$
$R:T - T \longrightarrow R$	0.05 $\mu M^{-1} s^{-1}$
$R_{a,i} + T \longrightarrow Ra_{,i}:T$	0.02 $\mu M^{-1} s^{-1}$
$Ra_{,i}:T - T \longrightarrow R_{a,i}$	0.05 $\mu M^{-1} s^{-1}$
$R:T \longrightarrow R_{a,i}:T$	Unknown
$R_{a,i} + U_i \longrightarrow R_a:U_i$	0.25-0.5 s^{-1}
$R_a:U_i - U_i \longrightarrow R_{a,i}$	0.25-0.5 s^{-1}
$R_{a,i}:T + U_i \longrightarrow R_a:U_i:T$	0.25-0.5 s^{-1}
$R_a:U_i:T - R_{a,i} \longrightarrow U_i:T$	10-20s^{-1}
$R_a:U_i - R_a \longrightarrow U_i$	10-20s^{-1}
$U_i + T \longrightarrow U_i:T$	3 $\mu M^{-1} s^{-1}$
$U_i:T - T \longrightarrow U_i$	0.5 $\mu M^{-1} s^{-1}$
$U_i \rightleftharpoons N_i$	Unknown
$M_i \rightleftharpoons U_i \rightleftharpoons N_i$	Unknown



Key
R= Ribosome
T= Trigger Factor
$R_{a,i}$ = Activated Ribosome
U_i = Unfolded Protein
M_i = Misfolded Protein
N_i = Native Protein

Figure 2. 5 Petri net Structure for Protein Synthesis Modelled Using PIPE5 Tool (Before Chaperone Inclusion. This figure illustrates the Petri net structure for modelling protein synthesis, created using the PIPE5 tool before including chaperones. The model captures the key steps of protein synthesis. Each place represents a molecular species or state (e.g., mRNA, ribosomes, unfolded proteins), and each transition represents a biochemical reaction or process (e.g., folding). The arcs connecting places and transitions indicate the flow of tokens, which represent molecules or states.

2.2.2 Python model

Python packages like PNET and Snakes are the most commonly used to construct Petri nets in Python (Chay et al., 2016; Pommereau, 2015). PNET was attempted because it is known to be able to model various forms of Petri nets making it flexible as a tool. However, the integration of models posed problems because of the steep learning curve that came with PNETS. Snakes was a tool used to simplify Petri net development but for this model is was not successful and posed similar issues as PNET .

Figure 2.6 shows an example PNET Python code developed using the data shown in Table 2.2. However, this system makes it difficult to edit and add new information once the model is construed. In addition, the model construed with all the pathways is very complex; therefore, this stepwise approach in PNET is unsuitable. Similarly, with Snakes, the addition of information is problematic, which limits this a tool for the model.

```
# The process
net.add_rules('Ra,i', 'step',
 ['R.a -> Ra,i; 1'])
net.add_rules('RT', 'step',
 ['R.T -> RT; 0.02'])
net.add_rules('Ra,iT', 'step',
 ['Ra,i.T -> Ra,iT; 0.02'])
net.add_rules('Ra,Ui', 'step',
 ['Ra,i.U -> Ra,Ui ;0.25'])
net.add_rules('Ra,Ui,T', 'step',
 ['Ra,i.T.U -> Ra,Ui,T; 0.25'])
net.add_rules('Ui,T', 'step',
 ['Ra.Ui.T/Rai -> UiT; 0.30'])
net.add_rules('Ui', 'step',
['Ra.Ui/Ra -> Ui ; 0.5'])
net.add_rules('Ui,T', 'step',
 ['Ui.T -> Ui,T; 0.30'])
net.add_rules('Ui,T', 'step',
 ['Ra,Ui,T/Rai -> UiT; 0.30'])
net.add_rules('Ui,T', 'step',
 ['Ra,Ui,T/Rai -> UiT; 3'])
net.add_rules('N', 'step',
 ['U.U -> N; 0.5'])
net.add_rules('M', 'step',
 ['N.N -> M; 1'])
def pathway(places):
 place = places['folding']
 rate = place.attributes['rate']
 if rate <= 0.01: return 0.0
 else: return 0.1
# run path
net.simulate(0.5, 0.05, 0.02)
```

```
from snakes.nets import *
x=1
n = PetriNet('First net')
n.add_place(Place('p', [0]))
n.add_transition(Transition('t',Expression('x<5')))
n.add_input('p', 't', Variable('x'))
n.add_output('p', 't', Expression('x+1'))
n = PetriNet('Second net')
n.add_place(Place('p', [0]))
n.add_transition(Transition('t',Expression('x<10')))
n.add_input('p', 't', Variable('x'))
n.add_output('p', 't', Expression('x+2'))
net.place("input").add(Queue(range(10)))
[Substitution(x=Queue([1, 2, 3, 4, 5, 6, 7, 8, 9]))]
```

Figure 2. 6 PNET and Snakes Model on Python for Protein Synthesis Using Data shown in Table 2.2. This figure illustrates the implementation of a PNET and Snakes model in Python to simulate protein synthesis based on the transition rates provided in Table 2.2. The PNET library is used to construct the Petri net structure, while the Snakes library enables dynamic simulation and analysis of the system.

A third approach in Python relies on using specifically defined places and transitions. This successfully models the protein accumulation in the example scenario described above. However, due to the size of the Petri net model to be constructed within the framework of this thesis, encoding of each place and transition individually in the Python framework was suboptimal and prone to error. Another tool was sort after for implementation and analysis of the final model. Figure 2.7 shows the output of the model, denoting the specific instances of simulation that led to failed pathways along with the final number of tokens remaining at each place at the end of the simulation.

```
t24 pathway worked!
  => [2, 2, 5, 1, 2, 2, 2, 2, 4, 3, 2, 2, 3, 1, 4, 2, 1, 0, 2, 3, 2, 3, 2, 3, 2]
t21 ...pathway failed.
t12 pathway worked!
  => [2, 2, 5, 1, 2, 2, 2, 1, 4, 3, 2, 3, 4, 1, 4, 2, 1, 0, 2, 3, 2, 3, 2, 3, 2]
t1 pathway worked!
  => [1, 3, 5, 1, 2, 2, 2, 1, 4, 3, 2, 3, 4, 1, 4, 2, 1, 0, 2, 3, 2, 3, 2, 3, 2]
t19 ...pathway failed.

final [1, 3, 5, 1, 2, 2, 2, 1, 4, 3, 2, 3, 4, 1, 4, 2, 1, 0, 2, 3, 2, 3, 2, 3, 2]
```

Figure 2. 7 Python Model Outcome: Pathways Fired and Token Distribution Across Places. This figure presents the results of the Python-based Petri net model for protein synthesis, showing which pathways have been fired successfully and the token distribution across each place at the end of the simulation.

2.2.3 Snoopy

Snoopy is a valuable tool for analysing, simulating, and graphically representing the Petri net (Rohr et al., 2010). Snoopy offers a comprehensive platform for studying dynamic systems. Its intuitive graphical user interface facilitates the easy creation and visualisation of complex models while integrated simulation capabilities (Heiner et al., 2012)The working model in this thesis was built and analysed in Snoopy. An advantage of Snoopy is the ability to add new pathways and change the rates of each reaction quickly because transitions and places can be accessed easily.

Once the model is simulated, Snoopy can generate graphs that show the number of tokens in each place against simulation time. Figure 2.8 shows the Petri net structure for protein translation, which will be introduced in detail in Chapter 4. Figure 2.8 is shown here to represent how Snoopy can manage larger Petri nets compared to PIPE5 and how this can be done quickly in compassion to PNET or Snakes.

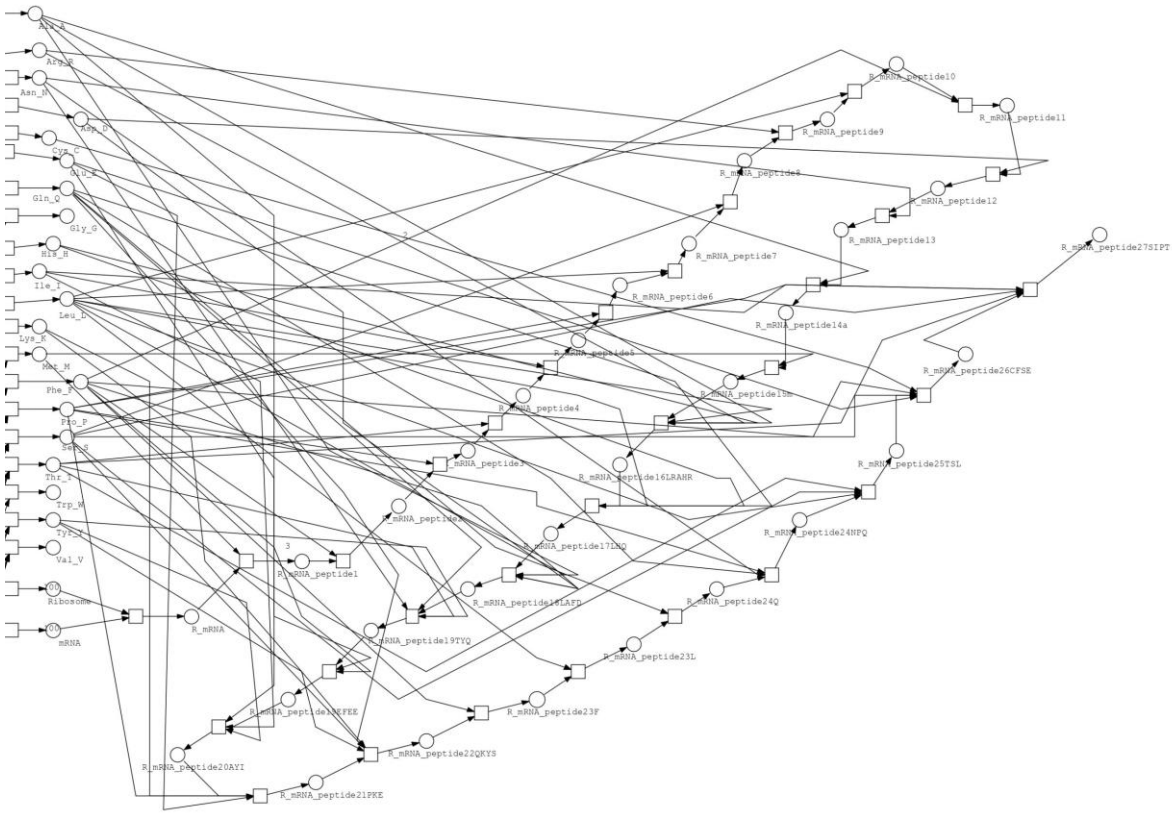


Figure 2. 8 Petri net Model Demonstrating Protein Translation Using Snoopy. This figure illustrates a Petri net model created in Snoopy to simulate the translation of a protein.

Chapter 3

3. Data collection for secondary use in model construction and validation

Constructing a backbone model for protein folding pathways in *E. coli* is an immensely intricate task, primarily due to the variability and heterogeneity of data from diverse publications and experiments. This complexity necessitated a meticulous data collection and integration process to identify the model's most reliable and representative inputs.

The challenges inherent in this process—ranging from reconciling conflicting data to addressing gaps in pathway annotations—underscore the depth and complexity of the research. A rigorous approach to data collection was essential to ensure the model's accuracy and robustness, highlighting the critical role of high-quality data in systems biology modelling.

Chapter 3 focuses on the data acquisition process used to understand protein folding pathways in *E. coli* and to develop the initial equations that form the foundation of the final Petri net model, which will be presented in Chapter 4. This chapter emphasises the role of molecular chaperones in protein folding pathways and their importance in maintaining proteostasis.

It also addresses the limitations of available data, particularly for processes such as translocation and disulphide bond formation, while identifying key gaps in the current knowledge. By summarising the literature and databases used, this chapter provides a comprehensive overview of the data collected to explore these pathways. It underscores the critical role of chaperones in protein folding.

The chapter also discusses the data sources used for model construction, including scientific journals, databases, and online repositories. Given that the model is constructed using secondary data, a systematic approach was essential to ensure the accuracy and relevance of the information collected.

The term equations vary depending on the dataset and the biological process being modelled. For instance, in the case of molecular chaperones such as DnaK and DnaJ, well-defined biochemical equations with known rate constants are available, allowing for precise modelling of their interactions and kinetics. However, specific reaction rates are often unknown for processes like Sec and Tat translocation. Instead, ratios of molecular chaperones to unit proteins are used to approximate their behaviour.

As a result, some equations cannot be strictly classified as traditional biochemical reactions but are better described as hybrid equations that incorporate biochemical principles and empirical ratios. The variability in data types and sources inevitably

impacts the results generated, necessitating careful curation. The data collected in this chapter is presented as mathematical equations to illustrate the flow of information and the importance of each component in the model. Chapter 4 will detail the development of the final model, integrating all sub-models.

3.1 Protein synthesis data

Molecular chaperones play a critical role in protein folding pathways in *E. coli*, ensuring the proper folding of nascent polypeptides, preventing misfolding and aggregation, and maintaining proteostasis—the balance of protein synthesis, folding, and degradation. This section focuses on the data acquired to understand these pathways and construct the initial equations to form the basis of the final Petri net model presented in Chapter 4. The discussion will summarise key literature and databases used to explore these pathways, particularly emphasising the role of chaperones in protein folding.

The chaperones investigated in this study are part of the heat shock response proteins in *E. coli*, which are upregulated under stress conditions such as elevated temperatures or oxidative stress. These chaperones also maintain proteostasis, ensuring proteins are correctly folded and functional.

Key chaperone systems in *E. coli* include:

- DnaK/DnaJ/GrpE (Hsp70 system): This system assists in folding nascent polypeptides and prevents aggregation by binding to exposed hydrophobic regions (Bukau et al., 2006).
- GroEL/GroES (chaperonin system): This system provides a protected environment for the folding of larger proteins that cannot be handled by the Hsp70 system (Figueiredo et al., 2004).
- ClpB, in particular, plays a critical role in protein disaggregation, working in concert with the DnaK system to resolubilise and refold aggregated proteins (Doyle et al., 2013).

The Petri net simulates how chaperone activity affects protein folding efficiency and misfolding rates by modelling these processes. This offers critical insights into proteostasis mechanisms and their consequences for cellular function.

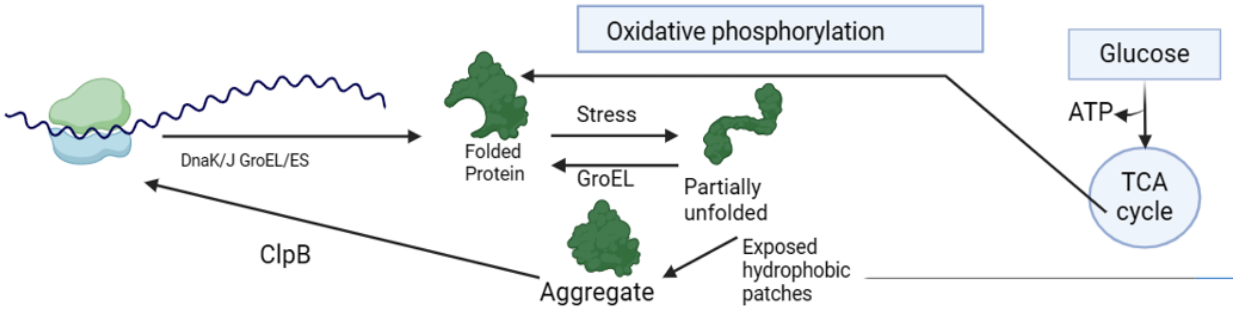


Figure 3. 1 Role of Molecular Chaperones in Protein Folding Pathways. Folding a chain of amino acids into a fully functional protein involves several key chaperone systems. DnaK and DnaJ assist in the initial folding of nascent polypeptides, preventing misfolding and aggregation. If a protein becomes partially unfolded, the GroEL/GroES chaperonin system facilitates its refolding by providing a protected environment. In cases where proteins aggregate, ClpB disaggregates and unfolds the misfolded proteins, giving the polypeptide chain a second chance to achieve its functional conformation. Each step requires ATP hydrolysis to drive the conformational changes necessary for proper protein folding. (Image created using BioRender.)

3.1.1 Data collection from literature for molecular chaperones during folding

Although the data from FoldEco provides a foundation for constructing the required mathematical equations, it is insufficient to fully understand the importance of each chaperone in protein folding pathways. DnaK, for instance, is a central player in these pathways and is essential for numerous reactions and transition steps. The absence of or mutations in chaperones like DnaK can severely impair cellular growth and the production of functional proteins, underscoring their critical role in proteostasis (Aguilar-Rodríguez et al., 2016; Grindle et al., 2021; Killelea et al., 2023).

Given the importance of chaperones, the data collection process for this part of the model is crucial to its utility as a predictive tool. To ensure accuracy, data has been collected in two complementary ways: (1) a summary of literature sources, as shown in Table 3.3, which provides qualitative insights into the roles of chaperones and reaction rates, and (2) quantitative data extracted from databases, as summarised in Table 3.4.

The qualitative data from the literature is cross-checked with multiple sources to identify recurring findings and new information. In contrast, the quantitative data from databases is used to parameterise the model. By comparing the equations derived from databases with the qualitative insights from literature, the pathways are constructed to reflect the best available knowledge.

This dual approach ensures that the model is not reliant on a single dataset but is cross-validated across multiple sources, streamlining the information and formulating a robust simulation of these pathways.

Table 3. 3 Summary and Critical Assessment of Literature on the Roles of Chaperones in Protein Folding Pathways

Type of Literature and Source	Summary of relevant information	New information	Similarities with other references	Reference
Report – Cell reports	Overexpression of the entire proteostasis network benefited all three test proteins. However, the effect of upregulating individual chaperones or the major degradation enzyme, Lon, varied for proteins with different biophysical properties.	Proteostasis network is a consequence of concerted action by the Hsp70 system (DnaK/DnaJ/GrpE), the Hsp60 system (GroEL/ GroES), and Lon	The folding fate of a protein in vivo is determined by the interplay between its folding energy landscape and the actions of the proteostasis network.	(Cho et al., 2015)
Research paper – International Journal Of Biochemistry And Cell Biology	Testing the overexpression of chaperones to show the impact on total yield	The observation in this paper highlights that GroEL and GroES chaperones can assist in the folding of multiple substrate proteins simultaneously when over-expressed	Folding of aggregation-prone recombinant proteins through co-expression of chaperones GroEL and GroES	(Goyal & Chaudhuri, 2015)
Research paper – PLOS	Gold nanoparticles exhibit a chaperone-like function in vitro, and the effect of citrate-coated	Role of DnaK in alleviating nanoparticle-induced stress in *E. coli* and highlighting this	Heat shock proteins (Hsps) are conserved molecules whose	(Makumire et al., 2015)

	gold nanoparticles on the growth of *E. coli* was investigated by deleting the DnaK gene.	gene's importance in reducing stress.	primary role is facilitating the folding of other proteins (chaperone function). Hsp70 (DnaK in prokaryotes) is one of the most prominent molecular chaperones.	
Research paper – Journal of Microbiology	First, to show that the chaperones GroEL-ES work well in *Rhodococcus* and simultaneously possess protein-folding assistance.	38% increase in yield.	Three combinations of molecular chaperones from *Escherichia coli* (i.e., DnaK-DnaJ-GrpEGroEL-GroES, GroEL-GroES, and DnaK-DnaJ-GrpE) were overproduced in *E. coli* BL2.	(Tian et al., 2015)
Research paper– PLOS One	Two stable conformations of cHSP70 with Substrate Binding Domain in open	Identification of the chaperon binding cavity, which explains the underlying mechanism	This new information aids in understanding how proteins bind to the	(S. Gupta et al., 2015)

	and closed states were obtained.	involved during heat and cold stress conditions.	chaperones to be folded correctly	
Research paper – Biotechnology and Biotechnological	The recombinant cells which showed inducible expression of DnaK2 were observed to have improved viability compared to the wild-type strain. While the temperature cannot be implemented into the model the information in this study is very useful because the transition rates can be changed to proxy this scenario.	The growth of the wild-type cell culture decreased to 64% of that of the recombinant cell culture following a transient (15 min) heat shock at 45 °C, after which the culture was subsequently transferred back to 37 °C.	Recombinant cells were shown to be able to grow faster in these conditions. I was aiming to show varying rates for increased temp as chaperones will respond but presenting this mathematically without real concentrations is problematic	(Gaber et al., 2015)
Research paper – Journal Of Biochemistry And Cell Biology	Glutathionylation down-regulates the functions of DnaK under oxidising conditions, and this down-regulation may facilitate the release ofσ32 from its interaction with DnaK, thus triggering the heat shock response.	Oxidative stress results in temporary inactivation of DnaK due to depletion of cellular ATP and thiol modifications such as glutathionylation until normal cellular ATP levels and a reducing environment are restored.	Although poorly understood, this study can show a relationship between oxidative stress and the heat shock response.	(Zhang et al., 2016)

Research paper – Microbiological Research	HP0795 is a constitutively expressed Trigger Factor (TF) - like protein.	The presence of HP0795 reduced protein aggregation caused by deficiencies of chaperones in these strains.	Trigger factor (TF) is one of the multiple bacterial chaperone proteins interacting with nascent peptides, facilitating their folding in bacteria.	(Fan et al., 2016)
Research paper – Nature Structural and Molecular Biology	Characterised the kinetic mechanism of substrate folding by the small ATP-independent chaperone Spy from *Escherichia coli.*	This is a potentially widespread mechanism whereby ATP-independent chaperones assist in protein refolding.	Chaperones use cycles of ATP binding and hydrolysis to assist in protein folding.	(Stull et al., 2016)

Table 3. 4 Molecular Chaperone Pathways: Known and Unknown Parameters from the FoldEco Database

Step in DnaK/DnaK pathway	**Equation**	**Reaction rate (s^{-1})**
One	Trigger factor + Unfolded protein $\longrightarrow$ Unfolded_T	0.02
Two	Unfolded protein $\longrightarrow$ Misfolded protein	Unknown
Three	Misfolded + DnaJ + ATP $\longrightarrow$ Misfolded_J +ADP +Pi	0.01

Four	DnaK + ATP ⟶ DnaK_Active	0.02
Five	DnaK_Active + Misfolded_J ⟶ Dnak_Active_Misfolded_j2	0.01
Six	DnaK_Active_Misfolded_j2 + ATP ⟶ new_active_DnaK_unfolded + ADP + Pi + DnaJ	0.02
Seven	new_active_DnaK_unfolded + GrpE ⟶ DnaK + unfolded +GrpE	0.02
Eight	Unfolded protein + ATP ⟶ ADP +Pi + Folded Protein	0.02
Step in GroEL/ES pathway	**Equations**	**Reaction rate (s^{-1})**
One	ATP + GroEL ⟶ GroEL_ATP	0.02
Two	Misfolded protein + GroEL_ATP ⟶ GroEL_ATP_Mis	0.05
Three	GroEL_ATP_Mis + GroES ⟶ GroEL_ES_ATP_Mis	0.05
Four	GroEL_ES_ATP_Mis ⟶ GroEL_ES_ADP_unfolded + Pi	0.05
Five	GroEL_ES_ADP_Un + ATP ⟶ GroEL_ES_ADP_Native + ADP + Pi	0.05
Six	GroEL_ES _Native + ATP ⟶ GroEL + GroES + ADP + Pi + Native protein	0.05
Steps in the Aggregation/ ClpB pathway	**Equation**	**Reaction rate (s^{-1})**
One	Misfolded protein ⟶ Agg_1	Unknown

Two	Agg_1 + Misfolded protein $\longrightarrow$ Agg_2	0.05
Three	Agg_2 + Misfolded protein $\longrightarrow$ Agg_3	0.05
Four	Agg_3 + Misfolded protein $\longrightarrow$ Agg_4	0.05
Five	ATP_bound_DnaK + Agg_1 $\longrightarrow$ Agg_1_ATP_K_1	0.05
Six	ATP_bound_K + Agg_2 $\longrightarrow$ Agg_1_ATP_K + unfolded protein	0.05
Seven	ATP_bound_K + Agg_3 $\longrightarrow$ Agg_1_ATP_K + 2 unfolded protein	0.05
Eight	ATP_bound_K + Agg_4 $\longrightarrow$ Agg_1_ATP_K + 3 unfolded protein	0.05
Nine	ClpB + Agg_1 + ATP $\longrightarrow$ unfolded protein + ADP +Pi	1

[1]

[1] Unfolded_T is a complex of unfolded protein and trigger factor
Misfolded_J is a complex of misfolded protein and DnaJ chaperone
DnaK_Active is a complex of DnaK chaperone and ATP
new_active_DnaK_unfolded when DnaK chaperone is attached to newly unfolded protein
GroEL_ATP_Mis when GroEL chaperone, ATP and misfolded protein form a complex
GroEL_ES_ATP_Mis when GroEL and GroES chaperone form a complex with misfolded protein and ATP
Agg_1 is two misfolded protein, Agg_2 is 2 units of Agg_1, Agg_3 is 2 units of Agg_2 and Agg_4 is 2 units of Agg_3

As reported in databases and publications, the equations above are easy to follow. However, they do not follow stoichiometry, and there is no understanding of the system boundary, which would be essential to building a formal model. Chapter 4 expands the above data to generate the Petri nets, where further adjustments to build the model will be explained in detail.

3.1.2 Discrepancies in the available chaperone folding data

The data presents missing information that requires further experimental work to understand. Many unknown rates, especially going directly from unfolded to misfolded, are not reported; these gaps will impact the final model. The ratio between the chaperones is not well understood. Although it is known, GroES chaperone forms a cage structure which requires 7 units of GroEL chaperone for every unit of protein, (Kalyani et al., 2016).

In addition, the application of ATP-independent chaperones has been shown to require DnaK and other heat shock proteins to refold proteins. The research in this thesis will focus on ATP-dependent chaperones because more datasets are available for these. However, it is acknowledged that there are other chaperones, such as IbpA and IbpB, which play a role in preventing aggregation (Schiffrin et al., 2016).

Overall, the literature data presents the importance of each chaperone, and well-established databases are available to construct an initial model for these pathways using Petri nets.

3.2 Tricarboxylic acid (TCA) cycle and oxidative phosphorylation data

TCA and oxidative phosphorylation are two critical pathways that drive the chaperone systems. Therefore, these were modelled with Petri nets as an energy source for the various pathways. Many databases have detailed maps of the TCA cycle available, making understanding the information flow relatively easy (Zhou et al., 2024).

The TCA cycle is a clearly defined metabolic pathway crucial for cellular respiration and energy generation. Accurately translating this pathway into a Petri net requires careful attention to the specific rules and assumptions representing its biochemical reactions. While Petri nets are adequate for modelling complex systems, they might not be the best choice for well-established pathways like the TCA cycle, especially when knowledge gaps are minimal, as other, more detailed modelling formalisms better suit

the TCA data. This section reviews the rules for converting the TCA cycle into a Petri net and examines the drawbacks of this approach.

To model the TCA cycle using Petri nets, the following rules and steps must be applied:

- Each metabolite (e.g., citrate, isocitrate) represents a place in the Petri net.
- Enzymes (e.g., citrate synthase) are represented as transitions, facilitating the conversion of substrates into products.
- The kinetic parameters of each enzymatic reaction (e.g., Michaelis-Menten constants) must be incorporated into the transitions. These parameters can be derived from experimental data or literature (Almaas et al., 2004).
- Feedback inhibition and activation mechanisms (e.g., ATP inhibition of citrate synthase) must be modelled as arcs in the Petri net. This accurately captures regulatory interactions (Reznik et al., 2013).
- Reversible reactions (e.g., the interconversion of isocitrate and α-ketoglutarate) require bidirectional arcs or separate transitions for the forward and reverse reactions.

The pathways translated in this chapter are from the KEGG database. The TCA cycle and oxidative phosphorylation pathways are well understood and, therefore, have limited gaps, which will be an issue for this modelling method. However, during data collection, it was evident that the data for the same pathways in EcoCyc are not the same as shown in KEGG.

The formula for Isocitrate is $C_6H_8O_7$ in KEGG, which is the correct form, but it is $C_6H_5O_7$ in EcoCyc, which may be a misprint. Therefore, the pathways were only taken from KEGG for the TCA cycle and oxidative phosphorylation. (Sharma et al., 2012). The rate is given as a range in the literature, and the middle value is used for the model, as shown in Table 3.5.

A simulation was conducted to determine whether the range of the highest and lowest reaction rates in the TCA cycle impacts the production of ATP and reducing agents, which are critical for protein folding pathways. The results in Appendix C show that higher TCA cycle rates produce more ATP and deplete glucose faster. However, varying the rates from low to high did not affect the amount of folded protein produced, indicating that the TCA cycle is not a limiting factor in the protein folding pathway. Instead, the bottleneck likely lies elsewhere in the system.

These findings suggest that, under normal conditions, the TCA cycle operates within a range that supports sufficient energy production for protein folding without becoming a bottleneck. Therefore, setting the TCA cycle rates to average values (as listed in Table 3.5) is appropriate for this study. While energy production is important for protein folding, the simulation demonstrates that the TCA cycle's role is adequately captured using average rates. This allows the focus to remain on understanding the interplay between chaperone activity and protein folding efficiency.

Table 3. 5 Step-by-step equations of the TCA cycle and oxidative phosphorylation in *E. coli,* a range of rates given for each step from literature (Bennett et al., 2009)

Steps in the TCA cycle pathways	**Equation**	**Rate range mM/s**	**Rate used in Model mM/s**
1	Glucose + ADP + Pi ⟶ phosphoenolpyruvate (PEP) + ATP + NADH + H20	0.10-0.20	0.15
2	PEP to pyruvate ⟶ oxaloacetate	0.05-0.20	0.10
3	Pyruvate + Co A ⟶ Acetyl CoA	0.10-0.50	0.25
4	Pyruvate ⟶ Lactate	0.05-0.20	0.10
5	Oxaloacetate + Acetyl CoA ⟶ Citrate + CoA	0.1-0.2	0.15
6	Citrate ⟶ Aconitate ⟶ Isocitrate ⟶ 2-oxoglutarate	0.05-0.50	0.25
7	2-oxoglutarate + Ubiquinone ⟶ Succinyl-CoA + ubiquinol	0.05-0.20	0.10
8	Succinyl- CoA + ADP +Pi ⟶ ATP + CoA + Succinate	0.05-0.20	0.10
9	Succinate ⟶ fumarate ⟶ malate ⟶ oxalacetate	0.05-0.2	0.10
Steps in Oxidative phosphorylation	**Equation**	**Kinetic parameters**	**Kinetic parameters**
Complex 1	Ubiquinol + NADH ⟶ NAD^+ + H^+ + Ubiquinone	0.0.5-0.20	0.10

Complex 2	Succinate dehydrogenase, which is step 8 above	0.0.5-0.20	0.10
Complex 3	Cytochrome bc1 complex where Ubiquinol becomes Ubiquinone whilst H^+ is pumped out	0.0.5-0.20	0.10
Complex 4	Cytochrome c oxidase to transfer electrons where $H^+ + O_2 \longrightarrow H_2O$	0.0.5-0.20	0.10
Complex 5	ATP synthase where $\longrightarrow$ ADP + Pi (when H^+ is pumped across membrane) ATP	0.0.5-0.20	0.10

3.3 Sec translocation data

The availability of databases with comprehensive information on translocation pathways for *E. coli* is limited. EcoCyc is one of these databases; however, this alone does not provide enough information for the model. There is limited information about kinetic rates; however, rations between the Sec translocation chaperones and the protein of interest have been reported. The turnover rate for Sec translocation has been reported as 1-100 per second per molecule of protein reported in the BRENDA database. However, this is a very wide range. The importance of accurate data in this process cannot be overstated, as it forms the foundation of the model's reliability and is responsible for ensuring its accuracy, thereby facilitating a deeper understanding of translocation pathways.

Visual aids are crucial in enhancing the understanding of complex concepts like translocation pathways and the subsequent construction of mathematical equations. Figure 3.2 was constructed based on the information acquired from literature, and it depicts the formation of amino acid chains transported through a Sec translocation channel (Ohta et al., 2016; Tan et al., 2022). Depending on the hydrophobicity of the protein chain, two possible routes exist. Lower hydrophobic chains necessitate ATP from the TCA cycle to transport the unfolded protein to the periplasm. On the other hand, highly hydrophobic chains can travel without ATP but may require the assistance of specific chaperones (J. et al. et al., 2014).

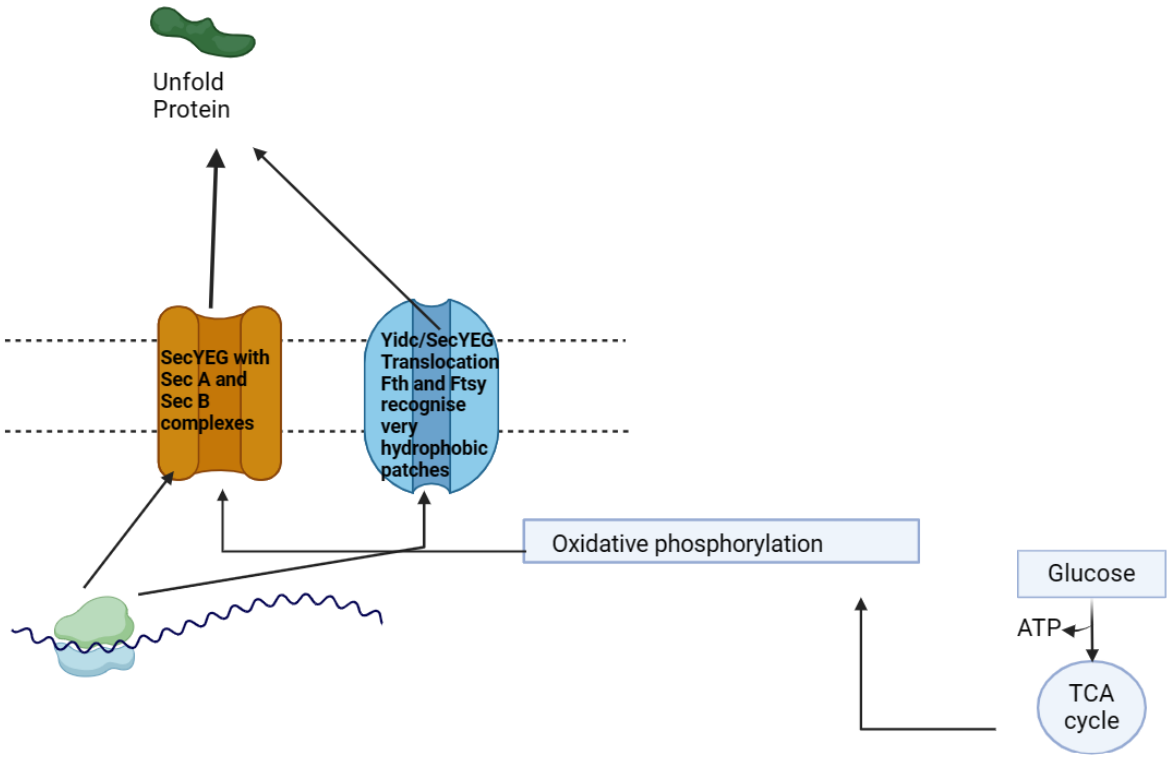


Figure 3. 2 Pathway of Sec Translocation in *E. coli*.The Sec translocation pathway facilitates protein transport across the cytoplasmic membrane. Hydrophobic regions in the amino acid sequence require ATP, generated by the TCA cycle, to drive translocation. In contrast, highly hydrophilic protein chains can translocate without ATP, relying on the YidC translocase and single recognition peptides. (Image created using BioRender.)

3.3.1 Collection of data on Sec translocation from the literature

The data collected for the Sec translocation system focuses on understanding the relationship between the Sec chaperones and the protein of interest (De Waelheyns et al., 2015). Moreover, *E. coli's* ability to adapt its protein translocation machinery impacts the quality and quantity of protein produced. (Karyolaimos et al., 2020).

Existing research proposed using Sec pathways for drug targeting and applied the translocation pathways to solve complex problems faced within biotechnology (Crowther et al., 2015; Mordkovich et al., 2015). An example of a study includes Sec-containing peptides that can be engineered to have specific targeting capabilities which improves drug delivery (Troman, 2021). The table below summarises some quantitative and qualitative results used to characterise the Sec translocation pathways and how this can be represented as mathematical equations.

Table 3. 6 Summary of information used to construct the Sec translocation pathways in *E. coli*.

Type of Literature and Source	Summary of relevant information	New information	Similarities with other references	Reference
Research paper - JMB	A new method called forced profile analysis (FPA) tracks protein folding in co-translational translocation.	The first fifteen residues do not play a role in correctly folding the protein but impact the journey to the periplasm.	A new method to show most unfolded proteins utilise SecYEG channel	(Sandhu et al., 2021)
Research paper - Journal of Biological Chemistry	PpiD and YfgM are inner membrane proteins which promote protein translocation Presents the understanding that the relationship of the components of Sec translocation is limited	PpiD participates in the periplasmic chaperone network and functions mainly to facilitate the early periplasmic folding of newly translocated proteins SDS-PAGE is used mainly to detect the presence of these chaperones. The ratios of values of these chaperones are not mentioned	(SecY, SecE and SecG) create a narrow channel for membrane translocation. Sec is a passive channel, an essential ATPase motor (SecA) proton motive force–driven motor the Sec can interact with the membrane protein, YidC, to integrate membrane proteins	(Miyazaki et al., 2022)

Scientific Reports – Nature Research	Study investigating mutations introduced to signal peptide loops and the ability of these mutants to translocate protein.	Mutations in the loop domain forming the eyelet can disrupt transport selectivity and thereby interfere with bacterial viability	A breakdown of the structure of the translocation system Proofreading ability to prevent defective signal peptides through the channel Highlighting little is known of these components	(Jung et al., 2020)
Research paper– Nature Communications	Understanding the structure of the complexes formed by electron cryo-microscopy during translocation to understand the pathway taken by the protein during translocation	A chain of events occurs where the resting state of SecYEG, the primed, the early, and the final state are inserted to allow protein translocation.	SecA is required for protein direction and required ATP. The ratios of the chaperones and protein seem to be one-to-one. Values for rates have not been stated, but an affinity value of 1 nM is used.	(Jomaa et al., 2017)
Research paper – JMB	The essentiality of the molecular chaperone DnaK for cytoplasmic stability.	Stabilising effects of DnaK can be exploited to increase the expression and translocation of Tat substrates under conditions where the substrate production	There is a connection between folding chaperones and translocation chaperones.	(Pérez-Rodríguez et al., 2007)

		level exceeds the capacity of the Tat translocase.		
Report - Molecular Microbiology	SecD, SecF, and YajC form another heterotrimeric complex that can be associated with the SecYEG complex.	SecD and SecF are present in low abundance in *E. coli* (less than 30 copies per cell). Although these chaperones are present, they are used to the low numbers.	SecA serves both as a receptor for preproteins and as an ATP-driven molecular motor that directs the movement of a translocating preprotein. Upon binding of SecA and a preprotein, the trimeric SecYEG complex assembles into a tetrameric channel.	(Nouwen & Driessen, 2002)
Research paper – Journal of Biological Chemistry	Depletion of YidC in *E. coli* leads to a specific defect in the functional assembly of primary energy-transducing complexes.	Rre-CyoA is a substrate of a novel pathway that involves both SecYEG and YidC.	20% of the Escherichia coli proteome encodes inner membrane proteins. Most of these proteins insert into the membrane via the Sec translocase.	(Du Plessis et al., 2006)
Report - JMB	Some membrane proteins require the membrane protein YidC. ATP synthase (F0a) which strictly depends on Ffh,	SDS-PAGE analysis and reaction assays were used to demonstrate the	This pathway is not as ATP taxing in comparison to Sec A	(Kol et al., 2009)

	SecYEG and YidC for its membrane insertion but is independent of the proton motive force.	importance of YidC to allow for translocation	Further data on values and rations are not shown in this report	
Research paper - Molecular Biology of the Cell	An early study on the varying routes of Sec translocation, which are independent of each other. It demonstrates the presence of different targeting routes.	Simultaneously deficient in SRP/SR, SecA, and SecB, membrane integration of mannitol permease and SecY was strictly dependent on the presence of the bacterial SRP/SR, whereas SecA, SecB, and H, all of which were found to be required for the efficient translocation of pOmpA under identical experimental conditions, were without any influence on the integration of the two membrane proteins	Translocation pore consists of SecY/SecE/SecG, whereas SecD/SecF and YajC improves the efficiency of translocation and is often called ancillary protein. SecA, which possesses ATPase activity, is also the motor for the subsequent translocation	(Koch et al., 1999)
Research paper – Molecular Cell	Conserved motifs mediate interaction in	This paper can show interaction rations, such as the binding of	Posttranslational translocation	(Huber et al., 2011)

	SecA and ribosomal protein L23, partial disruption of this interaction in vivo by introducing mutations into the genes encoding SecA or L23 affect translocation efficiency. Even in the presence of mutations, SecA can interact with pre-proteins to translocate, although it is not as effective	Sec A to a ribosome in a 1 to 1 ratio.	the machinery consists minimally of two components: the SecYEG translocon and the ATPase SecA. SecYEG forms an hour glass-shaped hydrophilic pore in the cytoplasmic membrane through which proteins must pass in an unfolded conformation in transit across the membrane	
Research paper – Molecular biology of the cell	The cargo is signalled by a finger loop system within *E. coli*, where the single reionisation peptide can start the translocation process when ATP is not required.	The experimental work carried out in this study shows that the finger loop is essential for efficient cargo processes within *E. coli*. Although varying mutations showed some differences in the concentration of SRP, the translocation process continued.	It is shown that the signal recognition peptide binds to FtsY. Once these complexes break, the chaperones are recycled, keeping them abundant for the pathways.	(Ariosa et al., 2013)

Report – EcoSal Plus	A detailed overview of the Sec system within *E. coli*	The report has given the translocation molecules' molarity (1 cell is 1 nM). Sec Y 0.3 µM Sec E 0.4 µM Sec G 1 µM Sec D 0.02 µM Sec F 0.03 µM YidC 2.5µM Sec A 4µM Sec B 0.020 µM SRP 0.1 µM	The ratio of SEC Y, E, and G is 1:1:1. These chaperones are abundant in the cell. Sec A is almost ten times the number of SecYEG.	(Crane & Randall, 2017)
Research paper – Applied and Environmental Microbiology	The co-translational targeting of the single-chain variable antibody fragment BL1 saturates the capacity of the Sec, leading to impaired translocation of secretory proteins and protein misfolding/aggregation in the cytoplasm	This research also recognised that the pathway potential has not been thoroughly investigated. It aimed to study the targeting and secretion of proteins into the periplasm, which could be used to design strategies to improve recombinant protein production yields.	The chaperones used for folding, such as DnaK, and catalysts used for disulphide bond formation, such as DsbA, strongly connect with Sec translation as further post-translation steps are required. This allows connecting the pathways to understand the impact on the protein produced.	(Ytterberg et al., 2019)

Research paper – Nature Communications	Kinetically dissect Sec-dependent translocation by monitoring translocation of a folded substrate protein with tuneable stability at high time resolution. That substrate unfolding constitutes the rate-limiting step during translocation	The translocon motor SecA has similarly been suggested to convert chemical energy from ATP hydrolysis into mechanical work that might help to unravel folded substrate proteins.	SecA undergoes repeated cycles of ATP binding, hydrolysis, and ADP and Pi release, which converts the chemical energy from ATP hydrolysis into mechanical work that drives translocation, which is poorly understood.	(R. Gupta et al., 2020)
Research paper– Journal of Bacteriology	The concentration of SecYEG is higher than that of the precursor; most active translocons are used only once due to limitations caused by an insufficient number of precursors for subsequent cycles. However, when precursors are sufficiently available, the active translocons can undergo more than one translocation cycle. Thus, resetting would be a candidate for the rate-limiting step if the translocon needs to be reset for each cycle.	Wild-type SecYEG for the precursors that undergo single and multiple cycles differed by 2- to 6-fold, with the higher rate constant observed for single turnovers.	SecB, a small cytosolic chaperone, captures precursors before they acquire a stable structure and delivers them to SecA.	(Mao et al., 2020)

Table 3.6 demonstrates that the availability of quantitative data on Sec translocation is minimal. Research papers explicitly state that this pathway is not fully understood, and the lack of research in this area is evident. Nonetheless, this gives more reason to demonstrate that via modelling these pathways using the available qualitative and quantitative data would assist deriving novel hypothesis to expand the research to increase translocation efficiency for various recombinant proteins.

To support this claim, Chapter 5 test case 1, explores the execution of translocation using uncoupled and coupled translocation systems using varying signal peptides. The study could successfully identify and explore new patterns that were not previously mentioned or investigated in the literature and consequently add to the understanding of the Sec system and the differences between the two routes within the Sec translocation pathway.

3.3.2 Discrepancies in the available Sec translocation data

EcoCyc provides a summary of the roles of the chaperones and their molecular weights, which helps to understand the material and information flow through the pathway and the size of the chaperones involved in the process. However, this alone is insufficient to develop a Petri net structure of the pathways. In addition, the importance of the two varying pathways within the Sec translocation is not highlighted.

The data from the literature gathered in Table 3.6 summarises the effort to understand the difference between these pathways and the role of each chaperone involved. However, even in the literature, there are discrepancies. The dynamics of the chaperones have been well established; Sec Y, E, and G form a complex within the membrane to allow for transportation. The cycle of opening and resting has also been documented well. However, the affinity for SecA being less than YidC is poorly understood. This interaction can be attractive for the model in understanding whether the requirement of ATP results in a preference for YidC. (Lee Upton et al., 2023; Öjemalm et al., 2013; Petriman et al., 2018; Robson et al., 2009; Roussel & White, 2020; Swaving et al., 1999).

In addition, the interaction of SecB appears as key for the preprotein to avoid folding. This chaperone works with DnaK to reduce the protein's premature folding and translocate unfolded protein to the periplasm. Sec B leads to a cascade effect, which forms the complex with Sec A and then Sec Y to allow for translocation. (Collier et al., 1988; Diao et al., 2012; Ha et al., 2004; Hartl et al., 1990; Sala et al., 2014; Sapriel et al., 2003; Watanabe & Blobel, 1989). Although some understanding of environmental factors impacting these pathways exists, the literature has been limited in providing data for this.

Another understudied area relates to the Sec D and F chaperones, which have been labelled as auxiliary chaperones, but the reason behind this is unclear. The limited contractions seem to be a possible reason, but this alone cannot justify the reason to categorise them as auxiliary. (Economou et al., 1995; Findik et al., 2018; Hand et al., 2006; S. Liu et al., 2016; Taura et al., 1993; Young Jae et al., 1994).

Although SDS-PAGE and other assays can be employed to confirm the existence of chaperones, the details of their mechanisms of action are poorly understood. This is particularly true when attempting to comprehend how the chaperones coordinate with one another. It is difficult to quantify the affinity for each chaperone and the affinity for pre-proteins, which creates a knowledge gap when discussing these pathways. (Baars, 2007; Cranford-Smith et al., 2020).

The substantial gap identified during data collection makes quantifying and forming a systematic relationship between the chaperones along the pathway difficult. There are several reasons, such as technical limitations related to measuring and reporting styles and the impact of different experimental methods on providing comparable results due to differences in the experimental or measurement conditions. Although these concepts are straightforward to perceive in biological narrative form, it is challenging to represent them as equations with limited data on parameters and mechanisms. Reducing these discrepancies requires a conscious effort of interdisciplinary research where data is shared using standards everyone can follow.

3.3.3 Equations for the Sec Translocation Pathway in *E. coli*

The available data can form stoichiometric equations because the ratio between the chaperones and their molarities is available, as reported in Table 3.6. The qualitative data allow us to understand the interaction and cycles through this pathway, which will be used to form equations to model them as a Petri net structure later.

Assumptions

- The ratio between Sec Y, E and G is one-to-one
- The ratio of Sec A is ten times SecYEG. However, the efficiency decreases as ATP is limited
- YidC is in abundance, and SRP can maintain efficiency in varying conditions
- The role of Sec D and F is not well defined, and it is present in low quantities; therefore, it will not be included in equations
- Once chaperones are used, they are recycled; therefore, concentration does not change under the quasi-steady-state assumption

Table 3. 7 List of equations constructed for the Sec translocation pathways within *E. coli*

Steps in the Sec translocation pathway	**Chaperones involved**	**Consecutive equations of the pathway**
Inner membrane proteins	YidC	1. Ribosome + amino acid ⟶ Ribosome_unfolded protein 2. Ribosome_unfolded protein + YidC ⟶ (cytoplasm) Complex with YidC + Ribosome 3. Complex with YidC + inner membrane ⟶ pore Unfolded inner membrane protein + YidC (cytoplasm)
Highly hydrophilic protein chain	SRP, YidC and SecYEG	1. Ribosome + amino acid ⟶ Ribosome_unfolded protein 2. Ribosome_unfolded protein + SRP ⟶ (cytoplasm) + YidC (cytoplasm) Complex with YidC and SRP + Ribosome 3. Complex with YidC and SRP + SecYEG ⟶ Complex SecYEG + YidC (cytoplasm) + SRP (cytoplasm) 4. Complex SecYEG ⟶ Unfolded protein in periplasm + SecYEG
Low hydrophilic protein chain	Sec B, Sec A, SecYEG	1. Ribosome + amino acid ⟶ Ribosome + unfolded protein 2. Unfolded protein + SecB ⟶ Complex with Sec B 3. Complex with Sec B/ Sec A⟶ Complex with Sec B/Sec A 4. Complex with Sec B and Sec A + 5. SecYEG ⟶ Complex with Sec A/SecYEG + Sec B 6. Complex with Sec A/SecYEG +ATP ⟶ Complex with SecYEG + ADP +Pi + Sec A 7. Complex with SecYEG + ATP⟶ Sec YEG + ADP + Pi + unfolded protein in periplasm

In addition to these equations, the relative ratios of the Sec chaperones from the literature are as follows: 1:1:2:10:0.1:5:0.5 for Sec Y, Sec E, Sec G, Sec A, Sec B YidC, and SRP,

respectively (Crane & Randall, 2017). The collected data yielded a starting point to construct the Petri net models of the Sec translocation pathways.

3.4 Tat translocation data

Tat translocation was discovered in the late 1990s and is not as well documented as Sec translocation (Santini et al., 1998; Wexler et al., 2000). However, interest in Tat translocation has increased because of its ability to transport fully folded proteins. Its chaperones are an additional checkpoint to ensure that misfolded proteins are not exported. (Stanley et al., 2001). Only a few databases similar to the ones available for Sec translocation are available for Tat translocation.

Figure 3.3 provides a diagrammatic representation of the Tat translocation pathway in *E. coli*. It demonstrates the formation of fully folded protein within the cytoplasm, which is then translocated using ATP into the periplasm.

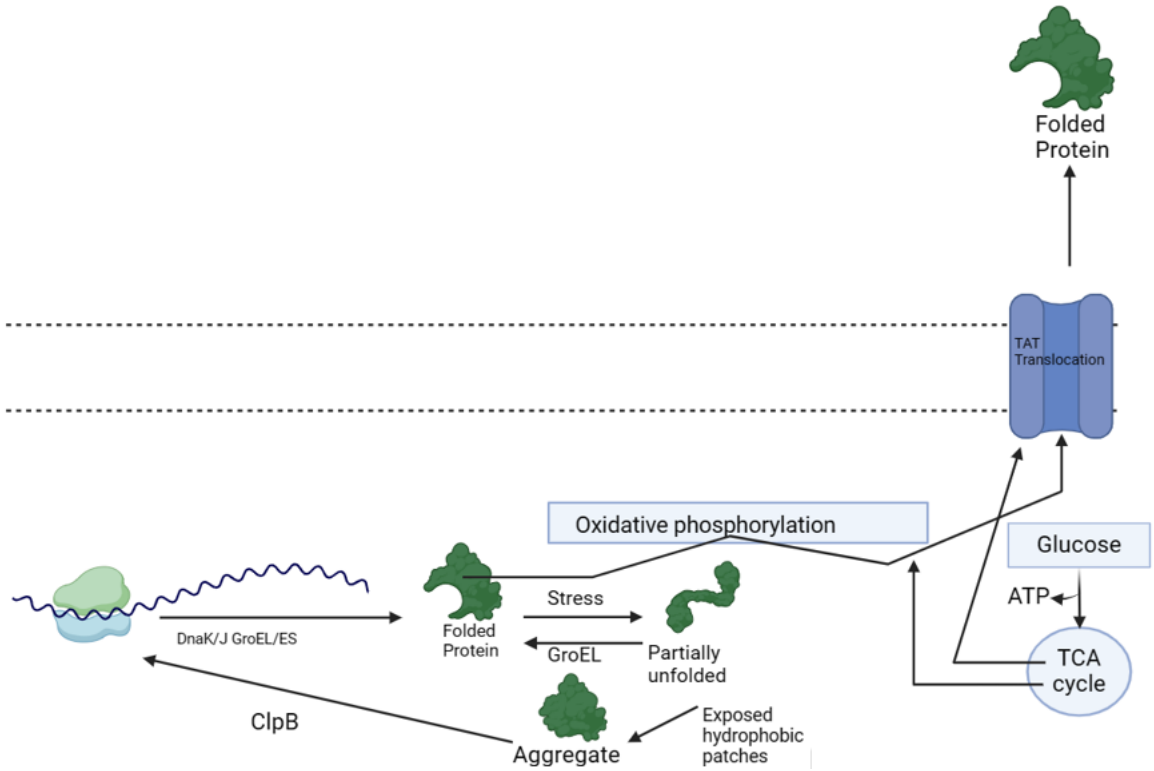


Figure 3. 3 Pathway of Tat Translocation in *E. coli*. The Tat translocation system forms a pore in the cytoplasmic membrane, utilising the proton motive force to drive protein translocation. Unlike other systems, the Tat pore exclusively transports fully folded proteins across the membrane into the periplasm. (Image created using BioRender.)

3.4.1 Collection of data from the literature for Tat translocation

Transport using the Tat pathway requires the assembly of Tat chaperones into the inner membrane to allow the formation of a channel to export folded protein. These folded proteins cannot be transported with the Sec translocation system, which makes the role of Tat very important when moving folded proteins in *E. coli* (P. A. Lee et al., 2006). Although it is understood that the Tat proteins align in a specific order of Tat A and Tat B/C complex, the functional presence of other Tat chaperones, such as Tat D and Tat E, is not well understood.

In addition, Tat translocation is used as a proofreading mechanism to ensure misfolded and unfolded proteins are not exported to the periplasm. (Delisa et al., 2002). Although this leads to low titres, engineered strains such as "TatExpress" have been developed to have higher yields by improving folding abilities within the cytoplasm. (Guerrero Montero et al., 2019). Moreover, the signal peptide for Tat systems, in particular, is derived from pre-trimethylamine N-oxide (TMAO) reductase (TorA), which, together with its bound cofactor, is bound to *E. coli's* periplasm (Blaudeck et al., 2001).

Tat systems are also present in plant chloroplast and are better understood for the transportation of minerals within plants. (Oertel et al., 2015)The ratio between the Tat proteins requires further investigation within E. coli but was reported to be one-to-one within plants. (Rose et al., 2013) (Zinecker et al., 2020). An alternative understanding is that the ratio varies depending on the nature of the protein to be transported, but there is not enough evidence to confirm this. In addition, Tat components require the proton motive force to move to the correct locations (Robinson et al., 2011; Sutherland et al., 2018). The data collected in this section will focus on the Tat system within bacterial cells, specifically *E. coli.*

Table 3. 8 Summary of information used to construct the Tat translocation pathway data in *E. coli*.

Type of Literature and Source	Summary of relevant information	New information	Similarities with other references	Reference
Research paper – Journal of Biological Chemistry	TatC is required for the interaction of TatA with TatB, suggesting that TatA may interact with the complex via binding to TatC.	Tat B and C are in a 1 to 1 ratio. Tat A is present in varying levels dependent on the size and hydrophobicity of the protein required to export. However, it tends to be a 1 to 1 ratio with Tat B and C.	Tat ABC required a protein motive force to translocate the folded protein.	(Bolhuis et al., 2001)
Research paper – Journal of Biological Chemistry	TatE is 50% less abundant than Tat A. However, it has inefficiently replaced Tat A's role.	TatE associates with active translocases and forms direct contacts with TatA, TatB, and TatC.	TatE is a small protein of 67 amino acids (7 kDa), 53% identical to TatA's.	(Eimer et al., 2015)
Research paper – Biochimica et Biophysica Acta – Biomembranes	Demonstrating the role of Redox Enzyme Maturation Proteins (REMPs) chaperones in targeting respiratory enzymes to the Tat system.	REMPs are system-specific chaperones which play roles in the maturation of Tat-	Tat B and C have regions in which proteins can bind. This is similar to a	(Kuzniatsova et al., 2016)

	Different REMPs may have different mechanisms for this task	dependent respiratory enzymes.	cascade where Tat B has two regions, and Tat C has four.	
Research paper – Biochemical Journal	Electron microscopy with immunogold labeling and array tomography to investigate the Tat system. SDS-PAGE analysis reveals that Tat proteins are more concentrated within the membrane compared to the cytoplasm. This approach contrasts with conventional methods, such as fluorescence tagging and 3D distribution, which often yield inaccurate results	TatA protein forms linear clusters in the inner membrane of *E. coli* upon increased expression of TatBC, indicating that TatA organisation in the inner membrane changes in response to changes in Tat subunit stoichiometry.	Tat A initiates the pore-forming elements. Proteins organise in the inner membrane to allow Tat substrates to pass through while maintaining membrane integrity, a process that is still not well understood.	(S. M. Smith et al., 2017)
Research paper – Journal of Biological Chemistry	Analytical techniques like LC-MS and MS to demonstrate how single peptides dock to TatC, forming a hydrophilic cavity in the membrane. This cavity requires the proton motive force for the protein to translocate to the periplasm	TatC, involving both its N- and C-tails. This suggests that two distant sections of TatC work together as the primary docking site for	TatB and TatC form circular substrate–receptor complexes with a central binding cavity	(Blümmel et al., 2017)

		twin-arginine signal peptides.	for twin-arginine containing signal peptides. TatA oligomers, as well as the thereby triggered translocation event, require the proton-motive force (PMF) as the sole energy source	
Research paper – PNAS	The Tat signal peptide triggers the targeting and assembly of Tat translocation.	TorA identified mutations that could restore some level of Tat transport to fusions with export defective twin arginine substitutions.	Proteins are targeted to TatBC by signal peptides containing an essential pair of arginine residues	(Huang et al., 2017)
Research paper – Advanced biomedical research	Overexpression of folding chaperones DnaK/GrpeE and using Tat translocation to residue misfolding protein.	TorA could transport more expressed proteins to the periplasmic	DnaK and DnaJ prevent protein misfolding.	(Bagherinejad et al., 2018)

	However, the choice of signal peptide can impact the expression pattern.	space compared to SufI. However, a considerable amount of human growth hormone remained in the cytoplasm and could not be transported outside the cytoplasmic space	DnaK binds to the TAT leader sequence, while DnaJ stimulates ATP hydrolysis. GrpE acts as a nucleotide exchange factor. Together, they regulate normal and stress-related functions.	
Research paper – Microbiological Research	Tat pathway is essential for physiological stability, and its dysfunction leads to severe changes in *E. coli* cells.	The loss of Tat secretion impairs the outer membrane integrity and leads to decreased cell growth. Revealed activation of several stress responses and experimentally confirmed the dependence of specific proteins	That system can transport folded proteins and discriminate against misfolded proteins. This exciting feature implies the existence of a so-called proofreading and a quality control mechanism	(Dolata et al., 2019)

		on the Tat system for export.	that recognises correctly folded proteins. However, the quality system is not well understood. TatA and TatE are more than 50% identical.	
Research paper – Microbiological research	Changes occur in response to the production of correctly folded and misfolded proteins. Additionally, a unique stress response occurs when proteins remain unfolded. Specifically, when disulphide bonds form in the cytoplasm of *E. coli*, a process called CyDisCo triggers an oxidative stress response that increases the production of redox regulation system proteins.	Tat-dependent secretion leads to increased tatA expression upon protein induction, but post-induction analysis shows decreased tatA and tatB expression, consistent with lowered TatA and TatB protein levels.	Tat signal peptide actively regulates the amount of Tat components being produced in a feed back loop to prevent an unnecessarily high concentration of Tat substrates in the periplasm	(Guerrero Montero et al., 2019)

Research paper – PLOS	Tat A, B and C are required for the aerobic growth of bacterial cells	TatB, besides TatA and TatC, is strictly required for unimpaired aerobic growth. In addition, TatB was also found to be essential for the secretion of a heterologous Tat-dependent model protein into the *C. glutamicum* culture supernatant	This shows the importance of the Tat proteins within *E. coli* and the cell membrane integrity shown in other research. This information is critical to incorporate within the model because host cell proteins require the Tat system.	(Oertel et al., 2015)
Research paper – eLife	Combine sequence co-evolution analysis, molecular simulations, and experimentation to define the interactions between the Tat proteins of *E. coli* at molecular-level resolution.	A new schematic is proposed where Tat B is present with Tat C; the pore is closed when the signal peptide binds, the affinity for Tat B reduces, and the incoming Tat A binds to form a complex, and the protein is released. However, there is	This study shows that Tat A and C are equal amounts, and Tat B is half the amount. It is also sandwiched between two Tat C molecules.	(Alcock et al., 2016)

		still a lack of understanding of the steps before release.		
Research paper – Frontiers	The signal sequence of TorA did not facilitate targeting but instead led to the deposition of fusion proteins in insoluble aggregates.	Inclusion body formation was inhibited by co-overexpressing the TorD and DnaK chaperones with amino acid substitutions	TorA is a signal sequence that typically directs protein export via the Tat pathway in *E. coli*	(Jong et al., 2020)
	TatA and TatB transmembrane helices (TMHs) can be seen as a compromise, enabling the membrane to achieve the necessary thinness for toroidal pore formation while maintaining a degree of hydrophobic mismatch to allow reversible pore formation.	Short TMHs cause a hydrophobic mismatch between Tat subunits and the membrane bilayer, although the functional significance of this mismatch is unclear.	TatA and TatB possess short transmembrane alpha helices, both of which are only 15 residues long in *E. coli.*	(Hao et al., 2022)

3.4.2 Discrepancies in the available Tat translocation data

One key distinction between the Tat and other translocation systems is protein composition. The Tat translocation process involves the core components TatA, TatB, and TatC, which form the translocation pore and facilitate substrate recognition. However, the roles of TatD and TatE remain poorly defined, with limited evidence supporting their involvement in the translocation process (Barnett et al., 2008; Pop et al., 2002).

This observation reflects a recurring theme in translocation systems: the presence of accessory proteins with unclear or redundant roles. The coexistence of the Sec and Tat translocation pathways within the cytoplasm has long been an area of interest, yet it remains poorly understood.

While studies have characterised the individual strengths of each pathway, including their substrate specificity and proofreading capabilities, the mechanisms by which these systems function concurrently are still unclear. The signal peptide plays a critical role in directing proteins to the appropriate pathway, but the initiation of this process and the coordination between the two systems remain unresolved (Choi & Lee, 2004; Erkut, 2021; Kolkman et al., 2008; Pradel et al., 2009; Sutherland et al., 2018).

The Tat system is unique in its reliance on the proton motive force (PMF) for protein translocation, leading to its classification as an ATP-independent system in some literature. However, ATP is indirectly required during protein synthesis to ensure that substrates are fully folded before export through the Tat pathway.

Additionally, the proofreading chaperone TorD has been shown to require GTP hydrolysis to initiate translocation. While some studies suggest that ATP and GTP can be used interchangeably for this step, others indicate a strong preference for GTP due to its higher affinity (Bageshwar et al., 2021; Bhuwan et al., 2016; Brüser et al., 2003; Dubini & Sargent, 2003; Grahl et al., 2012; Hatzixanthis et al., 2005). This discrepancy in the literature necessitates careful assumptions when constructing models of the Tat system.

Despite these uncertainties, the Tat system is a fascinating research subject due to its critical role in cellular physiology. Mutation and deletion studies have demonstrated that the Tat system is essential for cell integrity and stress response, with its absence leading to impaired cell growth and viability (Liu et al., 2020; Yang et al., 2010). However, the underlying reasons for these effects remain poorly understood, highlighting a gap in our knowledge of Tat system function and regulation.

3.4.3 Equations for Tat Translocation in *E. coli*

The data collected for Tat translocation is sufficient to understand how the folded protein forms part of a complex within the membrane and then, via Tat C, gets translocated to the periplasm. Even though many intermediate steps of this pathway have not been reported explicitly or were reported not to be understood adequately, elucidating the mechanism by which the Tat proteins come together to form the complex would be an invaluable contribution to the model towards establishing an understanding of the flow of information further.

Summary of assumptions made in model construction based on the available literature:

- Tat A and C are in a one-to-one ratio
- Tat B is in a 1:2 ratio with Tat C
- Tat D and E are auxiliary proteins, which are present in small quantities
- TorD is used to proofread the folded protein
- TorA binds to Tat C, and the affinity to Tat B reduces the exposure of a pore in the membrane
- Tat A binds to Tat C, and the single peptide interacts with the membrane
- Proton motive force (PMF) is required to move folded protein to the periplasm

Table 3. 9 List of equations constructed for the Sec translocation pathways within *E. coli*

Steps in the Tat translocation pathway	**Chaperones and proteins**	**Equations**
Fully folded protein	TorD, TorA, Tat A, Tat B and Tat C	1. Tat B +Tat C $\longrightarrow$ Tat BC 2. Folded protein with TorA +TorD $\longrightarrow$ Folded protein complex with TorD 3. Folded protein complex with TorD + TatBC $\longrightarrow$ Folded protein complex with TorD and Tat C +Tat B 4. Folded protein complex with TorD and Tat C + Tat A $\longrightarrow$ Folded protein complex with Tat A and Tat C +TorA (in membrane) 5. Folded protein complex with Tat A and Tat C + PMF $\longrightarrow$ Folded protein (in periplasm) + Tat A +Tat C

3.5 Disulphide bond formation data

Disulphide bond formation has been a bottleneck of protein expression in *E. coli* for decades (Pratama et al., 2021)Therefore, the number of available databases is subsequently higher than those available for the translocation pathways. Although the formation of disulphide bonds within proteins to increase their stability was understood as early as the 1950s, it took three decades to characterise the disulphide bond formation and identify a family of enzymes called the Dsb proteins that played a role in this pathway (Jaswal et al., 2020) (Guilhot et al., 1995). The Sec translocation system and disulphide bond formation are linked because the unfolded protein must be directed to the Dsb proteins to form a correctly folded active protein. (Hajihassan et al., 2024; Maleki & Hajihassan, 2021). Where the Tat translocation received fully folded protein to be transported to the periplasm.

Disulphide bonds form when the thiol groups of two cysteines form a covalent bond. This requires the enzyme DsbA, which completes an oxidation reaction with the protein to form the bond. Engineered cells where the redox balance is maintained with the cytoplasm by maintaining glutathione can complete disulphide bond formation within the cytoplasm. However, wild-type *E. coli* can only complete this process within the periplasm, so electrons are transferred across the Dsb proteins to ensure correct protein isomerisation. (Stewart et al., 1998). Recombinant proteins are often mis-oxidised; therefore, DsbC and DsbD are required to correct the isomerisation, as shown in Figure 3.4 (Hemmis et al., 2011).

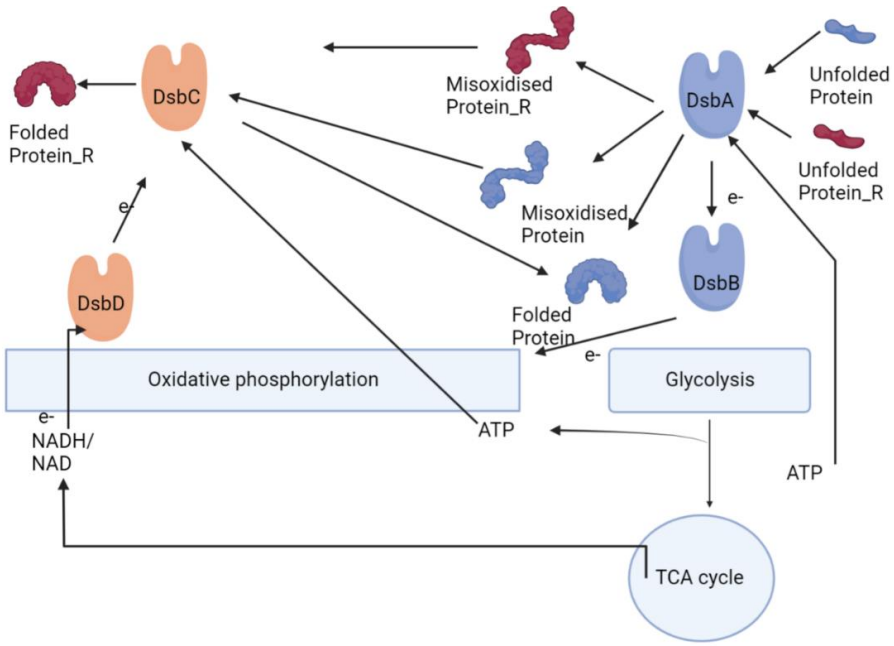


Figure 3. 4 The pathways for disulphide bond formation in *E. coli* identify the four main enzymes involved. Whenever ATP is required, it is supplemented from the TCA cycle. Unfolded native protein is isomerised by DsbA, which transports electrons to DsbB within the membrane. While most native proteins can fold, some are mis-oxidised, necessitating DsbC for correction. Additionally, electrons from DsbD are required to facilitate this process. DsbD receives electrons carried by NADH. The mis-oxidised native protein can now rectify itself. Recombinant proteins with more than two disulphide bonds are often mis-oxidised and require DsbC to correct the isomerisation (Biorender drawn image)

3.5.1 Data collection from literature and databases for disulphide bond formation

Disulphide bond formation requires the oxidation and reduction of Dsb proteins to carry out the bond formation and recycle the Dsb proteins for the next protein. The thioredoxin fold within DsbA allows for hydrophobic patches around the active site, which allow proteins to bind and form a complex. When DsbA is reduced, the protein is released and recycled by gaining electrons from DsbB. (Jonda et al., 1999; Zapun et al., 1993). The flow of electrons is controlled via the reducing potential, where DsbA has a higher reducing potential than other Dsb proteins. (Manta et al., 2019).

DsbC is of greater importance when non-conservative disulphide bonds, such as recombinant proteins, are present. Because of this, DsbC is considered the most versatile disulphide protein; the thioredoxin fold is similar to DsbA, allowing binding to proteins to correct the isomerisation. The separation between the two pathways of DsbA/DsbB and DsbC/DsbD is vital to maintaining the correct redox portion within the periplasm for efficient disulphide bond formation. (Kpadeh et al., 2015; Nakamoto & Bardwell, 2004; Shevchik et al., 1994). Another protein is DsbG, which has been shown to work similarly with DsbC but not as efficiently. (Van Straaten et al., 1998); the purpose of DsbG is not well understood in the literature. (Manta et al., 2019).

Table 3.10 below summarises data collected from literature and databases to construct the pathway in as much detail as possible.

Table 3. 10 Summary of literature and information from databases for disulphide bonds in *E. coli*

Type of Literature and Source	Summary of relevant information	New information	Similarities with other references	Reference
Research paper – Angewandte Chemi	Biophysical screening was used to identify fragments that bind to DsbA from *E. coli*. Elaboration of one such fragment resulted in the development of compounds inhibiting DsbA activity in vitro.	The absence of DsbA renders many proteins unfolded, making the bacteria avirulent.	Crystal structures of inhibitors bound to DsbA indicate that they bind adjacent to the active site.	(Adams et al., 2015)
Research paper – Antioxidants & Redox Signalling	Developed a simple blue/white screen that can detect disulphide bond isomerisation in vivo, using a mutant alkaline phosphatase	Differences in their substrate specificity may be due to differences in their CXXC active site pocket	Disulphide bond isomerases, such as DsbC of *E. coli* can recognise mis-oxidised proteins and shuffle the disulphide bonds of the substrate protein into their native folded state. DsbG is a periplasmic oxidoreductase that shares 27% amino acid identity to DsbC	(Chatelle et al., 2015)

Research paper– Nature Chemical Biology	DsbB was inhibited in understanding the impact of DsbB defects on the disulphide bond formation pathways.	Inactivation of the pathway for disulphide bond formation disrupts the function of multiple virulence factors.	The flow of electrons from the disulphide bonds is transported to DsbA, which must transfer to DsbB to maintain the redox within the membrane. Quinones transport electrons from DsbB to the terminal electron acceptors.	(Landeta et al., 2015)
Research paper– Nature Chemical Biology	Solubilised DsbB variants and an export-defective copy of DsbA were coexpressed in *E. coli* cells, creating artificial oxidation pathways. This facilitated efficient de novo disulphide-bond formation in substrate proteins, dependent on DsbA and quinone.	DsbB can oxidise ~40% of DsbA in a 1:1 stoichiometric reaction	DsbB is a transmembrane enzyme that catalyses the reoxidation of the periplasmic oxidase DsbA by ubiquinone.	(Mizrachi et al., 2017)
Research paper – Analytical Chemistry	Measured previously uncharacterised disulphide bond redox chemistry in *E. coli*	Electron transfer rates (both rate constants ≥ 4000 s–1) and quantify the energetics of the cysteine disulphide redox-reaction	Unfolded protein + $2e^-$ + $2H^+$ = folded protein	(Adamson et al., 2017)
Research paper –	These findings indicate that bacteria likely form	Oxidoreductases (DsbE or DsbF) may catalyse protein disulphide bond	In Escherichia coli, the DsbA and DsbB enzymes promote	(Ke et al., 2018)

Journal of Bacteriology	disulphide bonds in the cell envelope.	formation. However, there is no direct evidence of VKOR interactions with DsbA, DsbE, or DsbF.	disulphide bond formation. Other bacteria, including the Actinobacteria, use instead of DsbB the enzyme vitamin K epoxide reductase (VKOR)	
Research paper – ACS Omega	DsbA oxidation occurs independently of UQ and is greatly affected by the shuffling of electrons among the four cysteine residues in DsbB, regardless of pH kinetics.	DsbA's oxidation by DsbB is energetically unfavourable but is promoted by UQ's oxidising power, which binds with DsbB. Recent findings show this reaction can occur without UQ, using UQ-deficient DsbB.	A (DsbA), which contains two reactive cysteines. DsbA is reoxidised by a membrane protein, disulphide-bond formation protein B (DsbB), which has four catalytic cysteines.	(Yazawa & Furusawa, 2019)
Research paper– BMC Biotechnology	Changed the translation efficiencies of five native *E. coli* proteins, DsbA, DsbB, Skp, SppA, and DegP, by modifying the strength of their ribosome binding sites	Single-gene mutant strains showed improved periplasmic production of at least one of the recombinant proteins	This study confirmed the pathways of translocation combined with disulphide bond formation. However, ratios were not discussed.	(Gawin et al., 2020)
Research paper – Molecular Microbiology	ToxS, to understand the mechanisms of ToxR transcription factor activation. This is a way to	ToxR were strongly dependent on the oxidoreductases DsbA/DsbC	ToxR-ToxR PPIs are increased in response to ToxR operators, the	(Lembke et al., 2020)

	understand protein-protein interaction.		co-activator ToxS and bile salts	
Brenda database [Access dates August 2023 – June 2024]	This database collects biochemical information on enzymes.	The protein containing L-cysteines binds to DsbA, forming a disulphide bond. The reduced DsbA binds to quinone. DsbB binds to quinone, recycles DsbA, and produces quinol. DsbD is an inner membrane protein in Gram-negative bacteria that transfers electrons from cytoplasmic thioredoxin to the periplasmic substrate proteins DsbC, DsbG, and CcmG. This process reduces disulphide bonds in the target proteins, converting them to dithiols.	The database and literature pathway for Dsb proteins are similar.	(Lee et al., 2021; Smith et al., 2018; Yazawa & Furusawa, 2019)

The table above highlights vital information to consider when constructing the model. The work summarised presents studies primarily focused on engineering *E. coli* strains to enhance their ability to form disulphide bonds.

Different strains have been engineered to tackle the bottlenecks of post-translational steps in *E. coli*. One example is the expression of the tPA protein, which has seventeen disulphide bonds that tend to unfold and form misfolded proteins. Furthermore, engineering the cell to have autoinduction has been shown to improve the yield of soluble products by fifty per cent. (Long et al., 2015) (Seras-Franzoso et al., 2012).

Another study of interest is that of the CyDisCo strain, which can form disulphide bonds within the cytoplasm so that the Tat translocation pathway can be employed utilising its natural proofreading ability (Alanen et al., 2015). Owing to its particular capabilities across multiple pathways, this strain will be further investigated as part of a case study in Chapter 5.

3.5.2 Discrepancies between data in the literature for disulphide bond formation

Although the scientific understanding of disulphide bonds is better than other pathways in *E. coli*, such as the translocation pathways described above, many gaps still exist in the literature that increase the complexity of using the available information in constructing a formal model of the pathway. The requirement of a redox state within the periplasm is well understood, and how the Dsb proteins interact with reducing agents such as NADH to maintain this is well-studied (Christensen et al., 2019; Østergaard et al., 2001). However, the new strains developed to facilitate enhanced disulphide bond formation use the cytoplasm to foster disulphide bonds. Although this is innovative, the impact of this modification on the performance and efficiency of other pathways that function within the cytoplasm has not been fully understood (Lobstein et al., 2016).

Although the stoichiometric ratio between the protein within which the disulphide bond will be formed and the Dsb proteins is one-to-one, the folding kinetics of the protein of interest and the Dsb proteins is not entirely understood and documented. The formation of a disulphide bond requires two cysteines, which form a bond within the DsbA protein, during which electrons are removed. DsbA can recognise the unfolded protein because of energy potentials; however, a detailed understanding of the substrate specificity is not yet understood. (Jalomo-Khayrova et al., 2018; Malojčić et al., 2014).

The interactions between DsbA /DsbB and DsbC/DsbD are reported to be facilitated to maintain the required redox environment, which drives disulphide bond formation. However, an in-depth inquiry into their precise roles leads to identifying some discrepancies in the available literature; an example of this is observed in the role of DsbA, which is required to fold proteins. However, there are genes identified to have the ability to be upregulated to reduce the impact of this defect. (Jurado et al., 2002;

Kong & Guo, 2014; Meehan et al., 2017). Furthermore, the role of DsbG is not fully understood.

During a stress response, the disulphide bond formation pathways get modified to accommodate the response, especially in oxidative stress. The impact of this change depends on the number of electrons available for the oxidation and reduction of Dsb proteins. An interesting regulation action was reported to trigger heat shock proteins when non-native proteins are expressed (Müller et al., 2013). This notion is of particular interest as it aligns with the core investigation that forms the aim of the thesis and will be further explored in a test case in Chapter 5.

The role of oxidoreductases and glutaredoxins in disulphide bond formation needs to be better understood. Studies have shown that they maintain the redox balance in the periplasm, as expected. (Fabianek et al., 2000; Prinz et al., 1997). However, the response of anaerobic conditions and oxidative stress on these enzymes and the role of the Dsb proteins must be clarified. The lack of sufficient data prevents the stress response from being modelled correctly and the role of these enzymes from being elucidated during the stress response (Eser et al., 2009; Feeney et al., 2012; Nordstrand et al., 1999; Prinz et al., 1997).

3.5.3 Equations for disulphide bond formation in *E. coli*

The disulphide bond formation proteins follow specific events when forming bonds and managing the periplasm's redox balance. To model these pathways, assumptions are made based on data drawn from the literature and databases.

Assumptions

- Stoichiometry for DsbA, DsbB and the protein of interest is one-to-one-to-one
- electrons are transferred from the protein to DsbA, which then transfers these to DsbB
- Recombinant proteins mis-oxidise and require DsbC/DsbD to reverse this; the reaction for this also takes place with a one-to-one stochiometric ratio
- Electrons are transferred to DsbD from NADH
- The redox balance within the periplasm is always maintained constant

Table 3. 11 List of equations for disulphide bond formation pathway in *E. coli*

Steps in the disulphide bond formation pathways	**Equations**
1	Unfolded protein + DsbA $\longrightarrow$ DsbA_Reduced + folded protein
2	DsbB + DsbA_R $\longrightarrow$ ComplexAB_Reduced
3	Ubiquinol + ComplexAB_Reduced $\longrightarrow$ H_2O + ubiquinone + DsbB + DsbA
4	Mis-oxidised protein + DsbC $\longrightarrow$ folded protein + DsbC_Reduced
5	DsbC_R + DsbD $\longrightarrow$ DsbC + DsbD_Reduced
6	NADH + DsbD_Reduced $\longrightarrow$ DsbD + NAD^+ H^+

3.6 Summary

This chapter highlights the critical roles of molecular chaperones in *E. coli*, particularly DnaK and GroEL/GroES, in maintaining protein conformation and cellular proteostasis. These chaperones are essential for ensuring proper protein folding, preventing aggregation, and responding to cellular stress.

However, despite advancements in understanding their mechanisms, gaps in data on reaction rates, stoichiometry, and kinetic parameters pose substantial challenges to developing precise and predictive models. For example, while data from KEGG Pathways on the TCA cycle and oxidative phosphorylation provide valuable insights, they are often incomplete or inconsistent, requiring careful integration and interpretation.

The Sec translocation pathway further exemplifies the complexity of modelling protein transport systems. Limited comprehensive data on substrate specificity, energy requirements, and interaction dynamics make it challenging to construct accurate models. Similarly, while unique in its ability to transport fully folded proteins, the Tat translocation pathway presents challenges due to its reliance on the proton motive force and the poorly defined roles of accessory proteins like TatD and TatE. Addressing these

complexities will require interdisciplinary research, improved data sharing, and the development of novel experimental and computational approaches.

Disulphide bond formation represents another persistent challenge, particularly in recombinant protein expression in *E. coli*. While the roles of DsbA, DsbB, DsbC, and DsbD are well-documented, issues such as inefficient bond formation, mis-oxidation, and the folding of proteins with multiple disulphide bonds remain unresolved. Gaps in our understanding of pathway efficiencies and regulatory mechanisms compound these challenges. Refining models of disulphide bond formation and leveraging new technologies, such as redox-sensitive reporters and high-throughput screening, could improve protein expression and stability in biotechnological applications.

An important contribution of this thesis is tackling the difficulty of gathering and integrating data across different biological scales and systems. Compiling information from various sources, ranging from kinetic parameters and reaction rates to qualitative insights from literature, presents inherent complexities and inconsistencies. Data regarding chaperone activity and translocation pathways frequently arise from diverse experimental conditions, complicating the reconciliation of findings. Reaction rates and stoichiometric coefficients for the TCA cycle and oxidative phosphorylation differ across studies, necessitating meticulous validation and normalisation.

Furthermore, the Sec and Tat translocation systems are commonly examined separately, limiting the understanding of their coordination and interaction in physiological contexts. These issues highlight the necessity for a systematic and interdisciplinary approach to data collection and model development. This thesis seeks to close the gaps by merging experimental data with computational modelling and systems biology methodologies, ultimately enhancing our understanding of protein folding and translocation pathways.

This chapter emphasises the critical need to address data inconsistencies and improve model accuracy to deepen our comprehension of cellular processes. By integrating data from scientific literature, databases, and experimental research, alongside the use of computational tools, this thesis establishes a framework for investigating protein folding and translocation pathways. The challenges presented here further illustrate the intricate nature of this work and its potential to foster innovation in both fundamental and applied research.

Chapter 4

4. Process and evaluation of the constructed sub-models

This thesis explores the theory and applications of Petri nets for modelling complex biological systems, specifically focusing on protein synthesis and folding pathways in *E. coli*. By developing detailed sub-models of these pathways, this work utilises the mathematical principles of Petri nets to simulate biological processes, yielding insights into recombinant protein production and the complex interactions of molecular chaperones. Integrating computational modelling with experimental data presents a strong framework for understanding protein folding dynamics and its regulation in varying cellular conditions.

This chapter presents a thorough overview of the methodology used to create the Petri net model and an in-depth analysis of the simulation results. The model captures the essential characteristics of protein folding pathways, highlighting the roles of key chaperones such as DnaK/DnaJ, GroEL/GroES, and ClpB, along with the interactions between energy production, translocation, and disulphide bond formation. The simulation findings are compared to trends and data described in the literature, facilitating an assessment of the model's accuracy and its capability to address critical questions related to chaperone interactions and protein folding efficiency.

This chapter examines the impact of overexpressed chaperones on total protein accumulation within the model. This analysis illuminates experimentally observed trends, such as the decline in protein yield in the absence or dysfunction of specific chaperones. For instance, studies show that the lack of DnaK or GroEL can markedly reduce the amount of correctly folded proteins, depending on their function within the folding pathway.

The data utilised to parameterise the model are derived from an extensive review of the literature and database analyses detailed in Chapter 3. In Chapter 5, the model will undergo further testing against additional datasets from the literature to ensure its robustness and generalisability. This chapter aims to illustrate how chaperones' roles have been accurately simulated and to compare these results with anticipated trends in the literature. This strategy emphasises the models' potential as a predictive tool for improving recombinant protein production and elucidating the molecular mechanisms of protein folding.

This research marks an advancement in studying protein folding pathways by bridging computational modelling with experimental biology. The findings from this work hold implications for biotechnology, synthetic biology, and therapeutic development, providing novel strategies for enhancing protein expression and stability in industrial and medical settings.

4.1 Proposed model

Constructing a model of the *E. coli* protein machinery using Petri nets is an intriguing and complex journey. It involves several vital steps, each contributing to developing a comprehensive representation of the protein synthesis pathways. The process begins with clearly defining the problem and system to be modelled, followed by the abstraction of essential components and interactions.

Subsequently, the model's structure and behaviour are specified using Petri net formalism, incorporating places, transitions, arcs, and initial markings. Constructing proposed models using Petri nets is not just a theoretical exercise (Heiner & Koch, 2004); it has insightful implications for system analysis and design across diverse domains. By capturing the dynamics of complex systems in a structured and formal manner, Petri nets enable a deeper understanding of system behaviour (Breitling et al., 2008). This chapter will evaluate these capabilities by modelling the impact of specific chaperone groups on the total protein produced.

4.1.1 Transforming a biological model of protein processing pathways into the formal Petri net model

Figure 4.1 depicts the pathways for protein synthesis in *E. coli* through a biological framework. The diagram clearly outlines the stages of a polypeptide becoming a fully folded protein. In contrast, representing this process mathematically requires considering biological interactions that yield products; each must include input, output, and a clearly defined system boundary. Figure 4.2 offers this mathematical representation, where certain assumptions have been made within the system boundary to clarify the information from Chapter 3 for each sub-model.

The detailed distinction between the straightforward biological representation and the rigorous mathematical interpretation is sizable. In biology, translocation might be indicated by an arrow pointing to the membrane. However, the mathematical model raises questions about how this biological complex transitions from the cytoplasm to the periplasm and the necessary factors that facilitate or enable this movement. Section 4.1.2 explains how the components of the Petri net are developed as smaller sub-models before the entire pathway is interconnected, as illustrated in Figure 4.2.

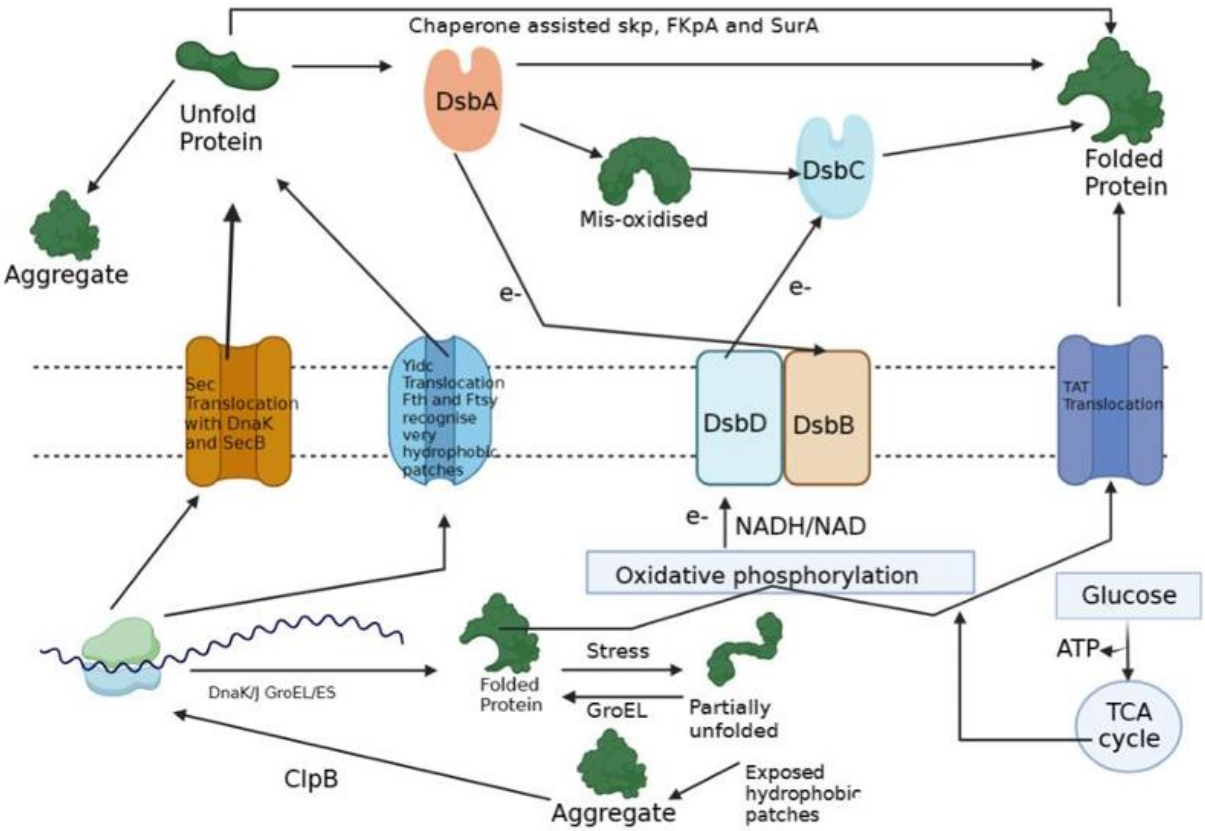


Figure 4. 1 Diagrammatic Representation of Protein Synthesis, Translocation, Energy Production, and Disulphide Bond Formation Pathways in *E. coli.* The figure illustrates the interplay between molecular chaperones (e.g., DnaK, GroEL/GroES), translocation systems (Sec and Tat pathways), energy production (TCA cycle and oxidative phosphorylation), and disulide bond formation (DsbA, DsbB, DsbC, DsbD). These pathways work together to ensure proper protein folding, transport, and stability. (Image created using BioRender.)

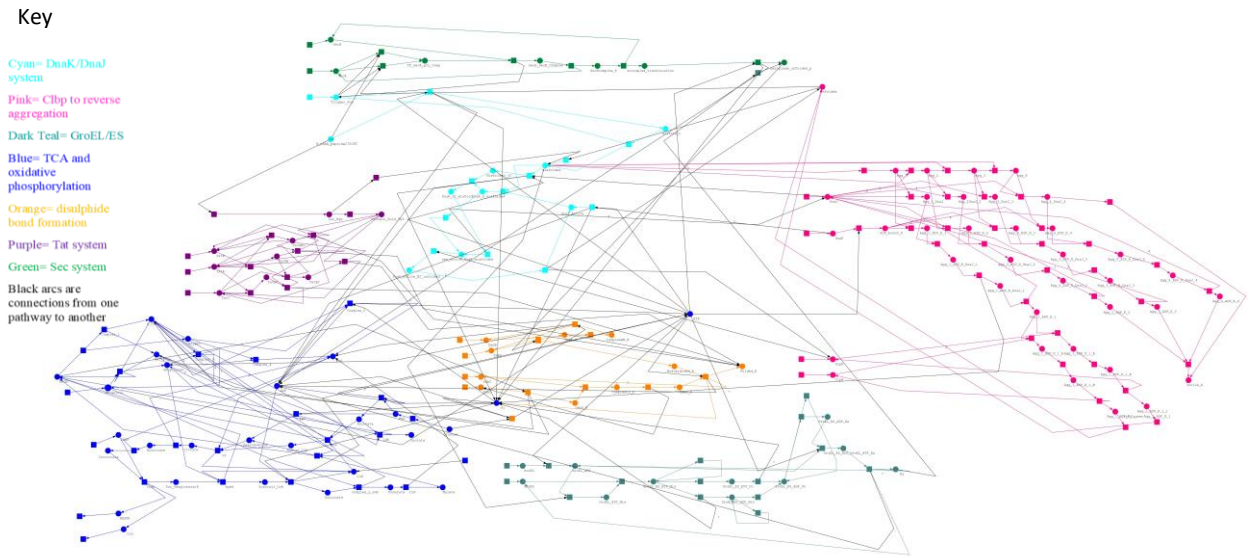


Figure 4. 2 The Petri net model for protein synthesis, Sec and Tat translocation, TCA, oxidative formation and disulphide bond formation. Sec translocation, Tat translocation, TCA, disulphide bond formation, GroEL/ES facilitated pathways, DnaK/DnaJ pathways and ClpB aggregation. Due to the size of the image, the resolution of this image may be reduced. To view the image in full resolution, please click here.

4.1.2 Protein processing is organised into smaller units depending on the nature of the pathway

The protein synthesis pathways in *E. coli* are categorised based on the roles and targets of the chaperones involved. These pathways operate concurrently given the dynamic intracellular environment, making Petri nets an ideal modelling framework.

Figure 4.2 offers an overview of the comprehensive Petri net model for the biological processes depicted in Figure 4.1. Due to the complexity of these pathways, they are divided into smaller sub-sections, each representing distinct biological phenomena. This division allows for detailed analysis and characterisation of each pathway, resulting in a well-defined, integrated model.

Constructing a Petri net model for biological systems involves defining its core components: places, transitions, arcs, and tokens. However, this must be achieved accurately to represent the underlying biological processes.

In protein folding pathways, places in a Petri net symbolise various states and components: unfolded proteins, misfolded proteins, folded proteins, chaperone-substrate complexes, and energy molecules like ATP and NADH. A place labelled Unfolded protein indicates the initial state of a newly synthesised polypeptide. Conversely, a place labelled GroEL/ES complex signifies the chaperonin system facilitating protein folding. This structured representation allows a clearer understanding of the biological processes involved in protein folding (Heiner et al., 2008) (Jong, 2008).

Transitions are important for representing biological systems; they show events that change a system's state, like biochemical reactions and conformational changes. They affect protein dynamics and stability during folding pathways, ensuring proper conformation and function, including chaperone binding, folding, and disulphide bond formation. Reaction rates are assigned to transitions to simulate the system's dynamics.

For instance, DnaK chaperone binding helps prevent protein aggregation and promotes folding during stress. Transitions highlight the dynamic nature of protein interactions and the complex events governing cellular functions. These are crucial for understanding molecular biology and biochemistry, particularly in protein folding and chaperone roles (Koch, 2014).

A Petri net's Arcs define token flow, like molecules or energy, through a system. In protein folding pathways, arcs specify each transition's inputs and outputs. For example, an arc from an unfolded protein to DnaK binding shows that the unfolded protein is essential for binding. This illustrates how transitions occur in biological systems, enabling effective modelling and analysis of complex biochemical processes (Heiner et al., 2008).

Tokens represent the quantities or concentrations of entities in the system, such as the number of protein molecules or the amount of ATP. In the context of protein folding pathways. Tokens are used to track the population of molecules in each place.

After setting up the components, the Petri net model is simulated to examine how the system behaves in different scenarios. For example, the model can assess the effects of chaperone knockout or energy restriction on protein folding efficiency.

Section 4.2 focuses on decomposing the Petri net model into individual pathways to provide a detailed understanding of the underlying processes. By isolating specific pathways such as chaperone-assisted folding, translocation, and disulphide bond formation, this section elucidates their contributions to overall protein production and folding efficiency. Additionally, the section examines how altering specific parameters (e.g., chaperone concentrations) can influence protein production, offering insights into potential strategies for optimising recombinant protein yields.

A key aspect of this analysis is the distinction between the model's default and tested conditions. The model is considered default when the transition rates are set to the parameters established in Chapter 3, derived from experimental data and literature. These default rates serve as the baseline for comparison.

To analyse the impact of varying rates, specific transition rates are increased or decreased by a defined percentage. This allows for systematically evaluating how parameter changes affect the simulation outcomes, such as protein folding efficiency and accumulation.

4.2 Protein synthesis

Protein synthesis in *E. coli* involves multiple steps and numerous interactions between chaperones, such as DnaK and DnaJ (Hupp et al., 1994). Petri nets can effectively model this intricate mechanism by providing a structured and formal representation of the system's dynamics.

Although protein synthesis can be depicted as a single pathway from translation to fully folded protein, various chaperones interact with the protein throughout this process, each playing a specific role. The Trigger Factor interacts with a polypeptide as it emerges from the ribosome. DnaK and DnaJ are essential for proper protein folding and address misfolding and aggregation.

In contrast, GroEL and GroES chaperones focus on refolding misfolded proteins by forming a cavity with seven GroEL chaperone units, within which the misfolded protein is correctly folded. Additionally, ClpB chaperones can reverse aggregation when fewer than four units of misfolded protein are aggregated; larger aggregates lead to inclusion body formation.

4.2.1 Protein folding by DnaK and DnaJ chaperones

This section of Chapter 4 explores the interactions between the chaperones DnaK and DnaJ during protein synthesis, emphasising their roles in the folding pathways and how these interactions are simulated within the model. Additionally, the role of the chaperone GrpE in converting misfolded proteins back into their unfolded form by promoting the release of ADP from DnaK will be discussed. This process offers another opportunity for the protein to fold correctly.

The model's initial step is the translation of the protein sequence, represented as P_sequence as shown in Figure 4.4. This step is kept generic to maintain model flexibility; it can be replaced with any protein of interest, as demonstrated in Chapter 5. Trigger Factor (TF), typically attached to the ribosome, binds to the emerging polypeptide sequence to prevent premature folding, which helps reduce misfolds and aggregation. Translation often occurs faster than proper folding, leading to aggregates. TF provides a protective environment to minimise the rate of misfolding. However, some proteins still misfold while attached to TF, represented as misfold_t in the model.

DnaJ detects misfolded proteins and facilitates the binding of DnaK chaperone. DnaK, utilising its ATPase activity, forms a complex with DnaJ chaperone and the misfolded protein to convert into the unfolded form. Once unfolded, the protein must be released from DnaK and DnaJ chaperones, which requires the assistance of GrpE, an ATPase. The unfolded protein contains hydrophobic patches that can lead to misfolding if exposed for too long. DnaK can bind to these unfolded proteins and assist in forming the correctly folded protein structure.

The impact of the DnaK chaperone will be further explored later in this chapter. Figure 4.3 shows the preliminary model, demonstrating the collected data as a Petri net model constructed in PIPE5. The expanded final model, simulated in Snoopy, corrects some initial assumptions and ensures each branch is a balanced equation, as illustrated in Figure 4.4.

Each sub-section begins with a preliminary model as the first step in understanding how these biological pathways can be represented using Petri nets. Initially, these models were unable to simulate the pathways effectively. Therefore, additional information was incorporated into the Petri nets, leading to the development of the final model, which was successfully simulated in Snoopy.

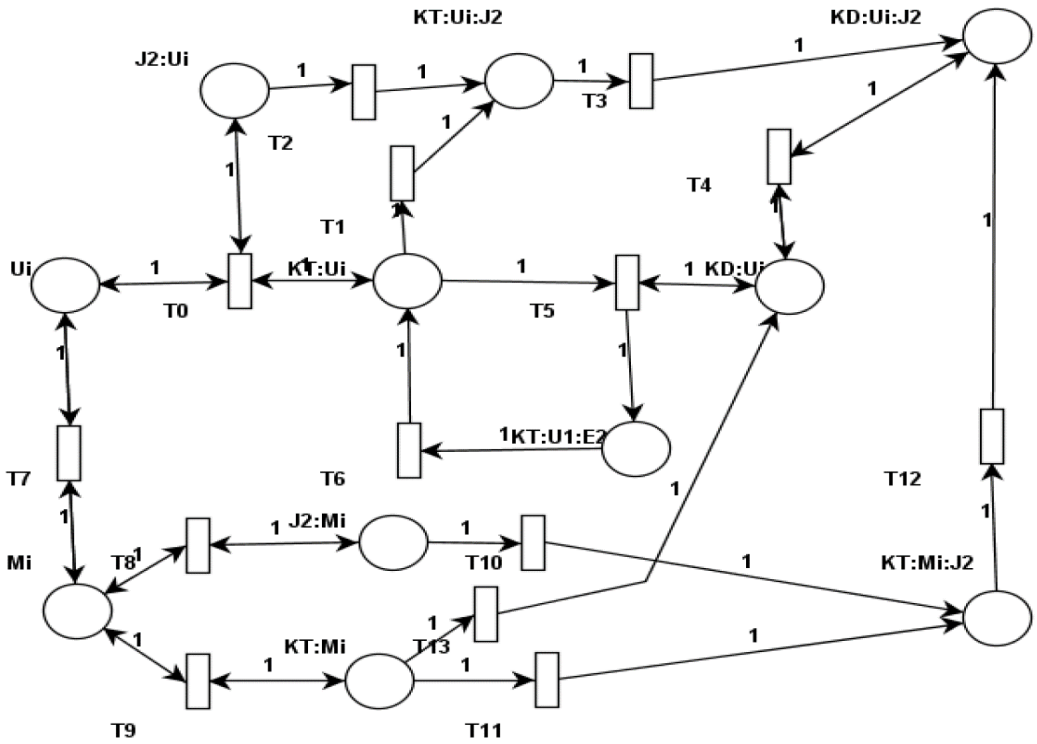


Figure 4. 3 Initial Petri net Model of DnaK-Mediated Protein Folding. The model illustrates the interaction of DnaK with unfolded protein (Ui) to facilitate the formation of folded protein. Ui is unfolded protein, Mi is misfolded protein, J2 is DnaJ, K is DnaK, T is ATP, D is ADP, and E is GrpE.

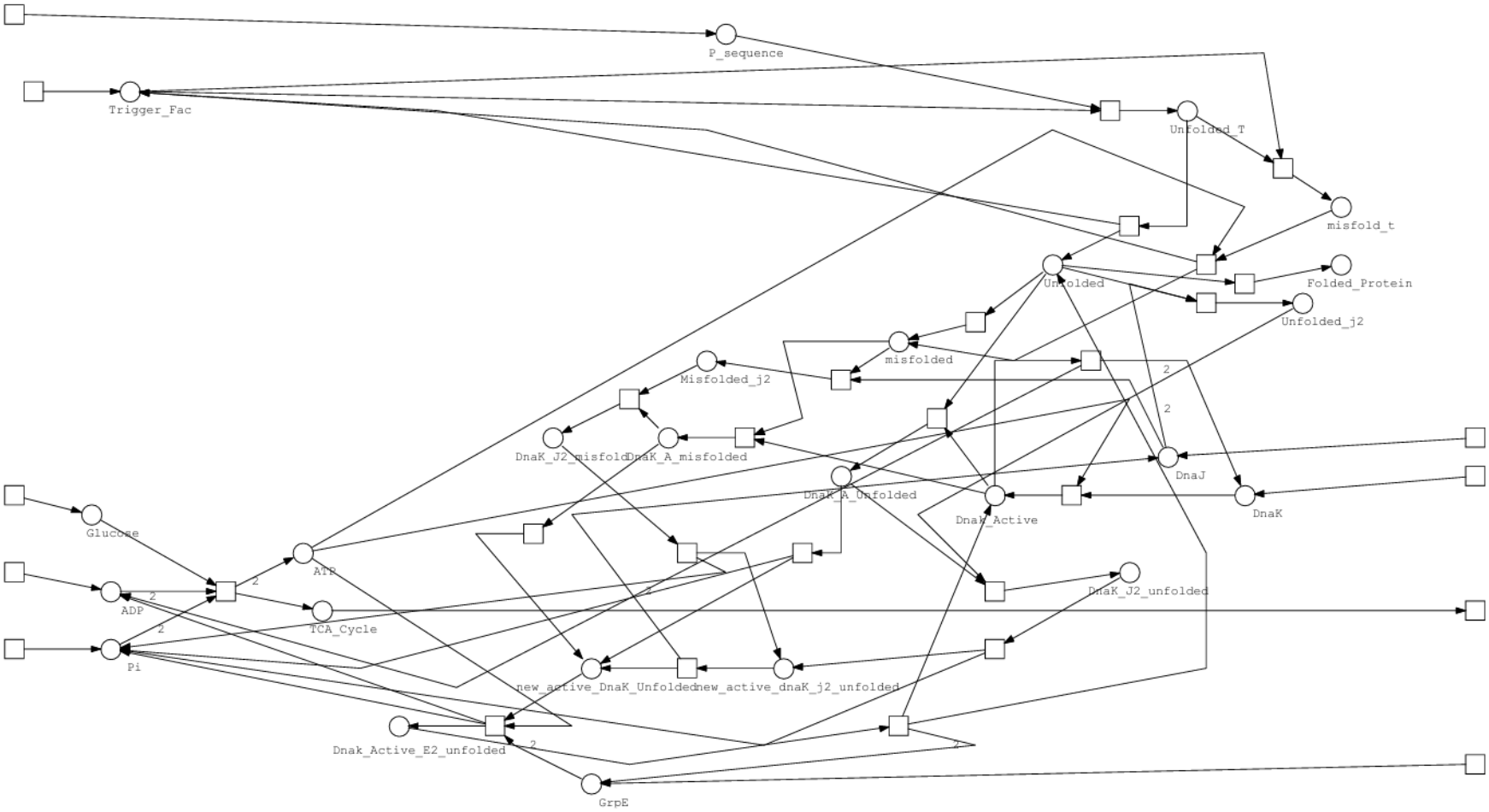


Figure 4. 4 Final Model of Protein Synthesis Interaction with DnaK, DnaJ, and GrpE. The model illustrates the interaction of DnaK, DnaJ, and GrpE in the protein synthesis and folding pathways. In this version, the TCA cycle is represented as a single place for simplicity, providing a simplified view of energy production. The integrated model will expand this representation, where all components are merged to provide a comprehensive system view.

It is difficult to find a complete study that tests all the simulations that can be tested using the model. However, it is crucial to benchmark it with expected results from the literature to know if the model represents the biological pathways studies well. These scenarios can help better understand the role of chaperones.

A comprehensive study conducted by Hartl et al. meticulously explored the intricate roles of DnaK and Trigger Factor (TF) in the protein folding processes. This investigation critically analysed how variations in chaperone activity impact the equilibrium between folded, misfolded and unfolded protein states. Notably, the researchers observed a increase in the unfolded to folded proteins ratio upon DnaK or Trigger Factor activity reduction.

Furthermore, the overexpression of DnaK and Trigger Factor markedly enhanced the yield of correctly folded proteins, concurrently diminishing the presence of unfolded proteins. The findings underscored the detrimental effects of elevated translation rates on protein misfolding, particularly when chaperone capacity was compromised.

Additionally, the study elucidated that excessive chaperone activity could paradoxically lead to a staggering 100-fold escalation in misfolded protein levels during periods of heightened translation as the chaperone systems became saturated by the overwhelming influx of nascent polypeptides (Hartl et al., 2011).

A separate study by Dukan et al. demonstrated that increased translation rates elevate the presence of misfolded proteins. This increase occurs because the cellular folding machinery, including chaperones like DnaK and GroEL/ES, becomes overwhelmed by the influx of nascent polypeptides. Under high translation rates, misfolded proteins can accumulate up to 100-fold compared to customary conditions, aligning with the observation that excessive translation rates increase protein misfolding (Dukan et al., 2000).

Conversely, when translation rates are reduced, the study observed a decline in the ratio of the unfolded-to-folded protein. A slower rate of protein synthesis provides chaperones and folding machinery with more time to process each polypeptide, thereby diminishing the likelihood of misfolding. In this scenario, the number of misfolded proteins remained relatively constant, as misfolding is mainly driven by intrinsic protein properties rather than translation rate alone. These observations are very similar to the simulation results obtained by the model.

Figure 4.5 illustrates the impact on the total amount of protein produced when the translation rate is increased or decreased by 20% to simulate the above studies. Figure 4.6 shows the absence of DnaK and DnaJ chaperones on the amount of folded protein.

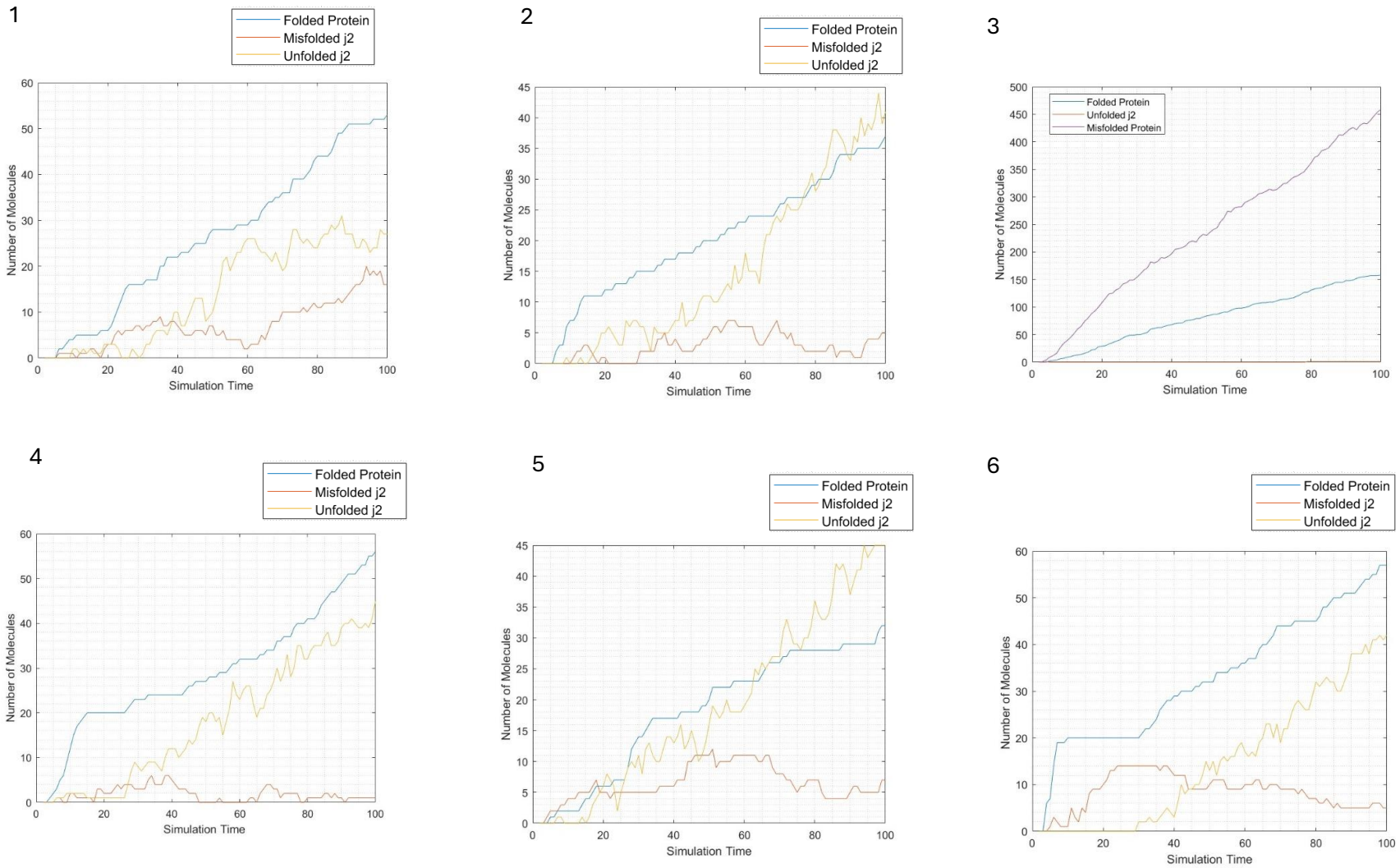


Figure 4. 5 Simulation Results of Protein Folding Pathways Under Varied Conditions. Graph 1: Ratios of folded, unfolded, and misfolded proteins when the system operates at constant rates. Graph 2: Impact of a 20% decrease in translation rate, showing a corresponding 20% reduction in folded protein. Graph 3: Impact of a 20% increase in translation rate, resulting in higher levels of both folded and misfolded proteins. Graph 4: Distribution of molecules when 100 input molecules of the protein sequence are processed. Graph 5: Effect of a 20% decrease in Trigger Factor (TF) rate, leading to increased unfolded proteins. Graph 6: Effect of a 20% increase in Trigger Factor (TF) rate, resulting in higher levels of folded proteins.

The simulation results reveal the influence of translation rate and Trigger Factor (TF) activity on protein folding efficiency and misfolding.

A 20% reduction in translation rate leads to a 20% decrease in the unfolded-to-folded protein ratio, while the number of misfolded proteins remains unchanged (Graph 1). This indicates that slower translation rates give the folding machinery more time to properly process nascent polypeptides, thereby reducing unfolded proteins.

Conversely, a 20% increase in translation rate results in a 100-fold rise in misfolded proteins, with the number of folded proteins remaining stable. The absence of accumulation of unfolded proteins suggests that higher translation rates overwhelm the folding machinery, leading to misfolding. These findings demonstrate the negative impact of excessively high translation rates on protein folding.

Reducing TF activity by 20% results in a more accumulation of unfolded proteins relative to folded ones (Graph 5), aligning with TF's established role in co-translational folding. Conversely, a 20% increase in TF activity leads to more correctly folded proteins (Graph 6), highlighting TF's importance in enhancing folding efficiency.

Interestingly, the quantity of misfolded proteins remains consistent in both scenarios, indicating that misfolding is primarily driven by intrinsic protein properties rather than TF activity alone.

Overall, the results underscore the necessity of balancing translation rates with chaperone capacity to optimise protein production and minimise misfolding. Excessive translation rates strain the folding machinery, while reduced TF activity decreases folding efficiency, resulting in more unfolded proteins. This aligns with research, such as Dukan et al. (2000) and Kramer et al. (2004), which emphasise the roles of translation rate and TF in protein folding dynamics.

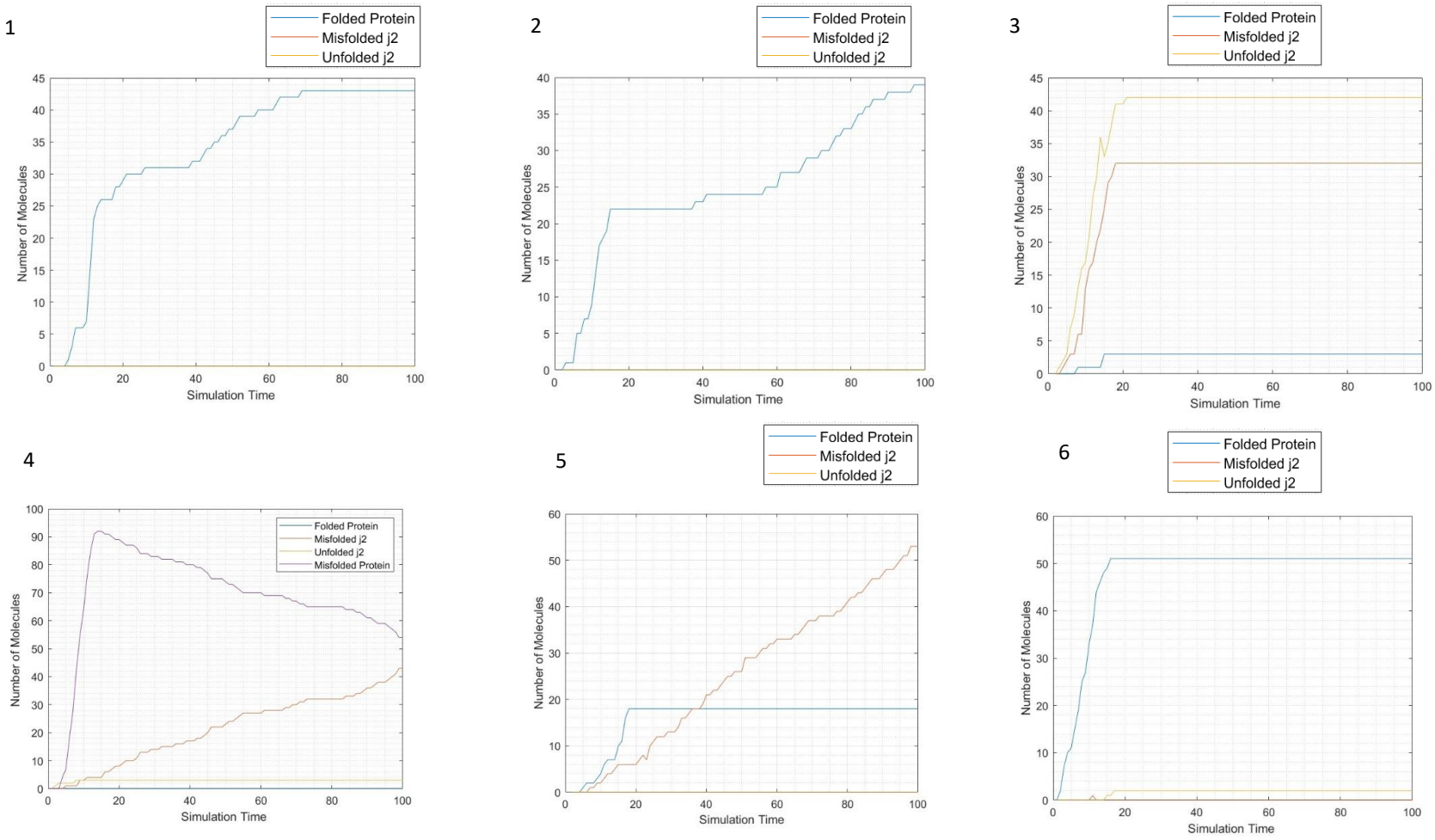


Figure 4. 6 Impact of DnaJ and DnaK Activity on Protein Folding Pathways. Graph 1: Accumulation of folded protein in the absence of DnaJ. Graph 2: Introduction of DnaJ at a 20% decreased rate, showing its impact on protein folding. Graph 3: Introduction of DnaJ at a 20% increased rate, resulting in the accumulation of misfolded and unfolded proteins bound to DnaJ. Graph 4: Deletion of DnaK, leading to the accumulation of folded and misfolded proteins. Graph 5: 20% decrease in DnaK activity, showing its effect on protein folding. Graph 6: 20% increase in DnaK activity, demonstrating its influence on protein folding efficiency.

Figure 4.6 provides a detailed examination of the role of chaperone proteins DnaK and DnaJ in protein folding, presenting the outcomes of simulations under various conditions. The figure explores the effects of deleting DnaK and DnaJ and modulating their activity by ±20%, achieved by altering the transition rates associated with these chaperones.

In Graph 1, the deletion of DnaJ reveals that protein folding can still occur in its absence; however, no complexes are observed with unfolded or misfolded proteins, implying an inactivation of pathways involved in the refolding of misfolded proteins. When DnaJ activity is decreased by 20%, there is a notable reduction in the quantity of folded proteins, coupled with an absence of complexes involving DnaJ and unfolded or misfolded proteins. Conversely, an increase in DnaJ activity by 20% leads to a rise in molecules corresponding to misfolded j2 and unfolded j2, suggesting an activation of refolding pathways. However, this also results in a decrease in the amount of folded protein, indicating that elevated DnaJ activity may adversely affect the efficiency of the folding process.

Graph 4 illustrates the consequences of DnaK deletion, resulting in a complete lack of folded protein accumulation and a concomitant increase in misfolded protein and complexes associated with DnaJ. This underscores the essential role of DnaK in promoting proper protein folding. A 20% reduction in DnaK activity results in only 20 molecules of the folded protein and an increase in misfolded_j2. In contrast, a 20% increase in DnaK activity yields a 60% rise in the folded protein, with no concurrent accumulation of misfolded or unfolded proteins. These findings demonstrate that DnaK is pivotal in facilitating correct protein folding while minimising misfolding.

The analysis further highlights the critical function of DnaJ in processing unfolded proteins, thereby providing an opportunity for the formation of correctly folded proteins. The model suggests that while DnaK primarily influences the pathways leading to correctly folded proteins, DnaJ is essential for activating the refolding pathways necessary for addressing misfolded proteins. Together, DnaK and DnaJ synergistically ensure the proper folding and refolding of proteins, maintaining cellular proteostasis.

4.2.2 Role of ClpB chaperones in unfolding of aggregated proteins

Under conditions of reduced refolding pathway activity, misfolded proteins can aggregate, leading to the formation of inclusion bodies. To mitigate this, *E. coli* employs alternative pathways facilitated by chaperones such as ClpB. These chaperones utilise ATP to convert aggregated proteins back into unfolded states. Additionally, ClpA and ClpX degrade aggregates that resist unfolding. This thesis focuses exclusively on chaperones capable of reversing damage, excluding discussions on ClpA and ClpX pathways.

Figure 4.7 presents the initial model where aggregated proteins interact with DnaJ or DnaK. These chaperones form complexes with aggregates, employing ATP to unfold

them into their constituent proteins. When aggregates exceed four units, DnaK and DnaJ cooperate, necessitating increased ATP for successful unfolding. Aggregates that cannot unfold beyond four units lead to the formation of inclusion bodies.

The pathway from misfolded proteins to unfolded forms is more intricate than depicted in Figure 4.7. Thus, the model expanded to encompass multiple cycles requiring ATP sourced from unfolded proteins. Figure 4.8 depicts the final model structure for this pathway. Due to its complexity, the screenshot may obscure specific places and transition names. Nevertheless, it underscores the progression from initial to final models, highlighting the model's comprehensive scope.

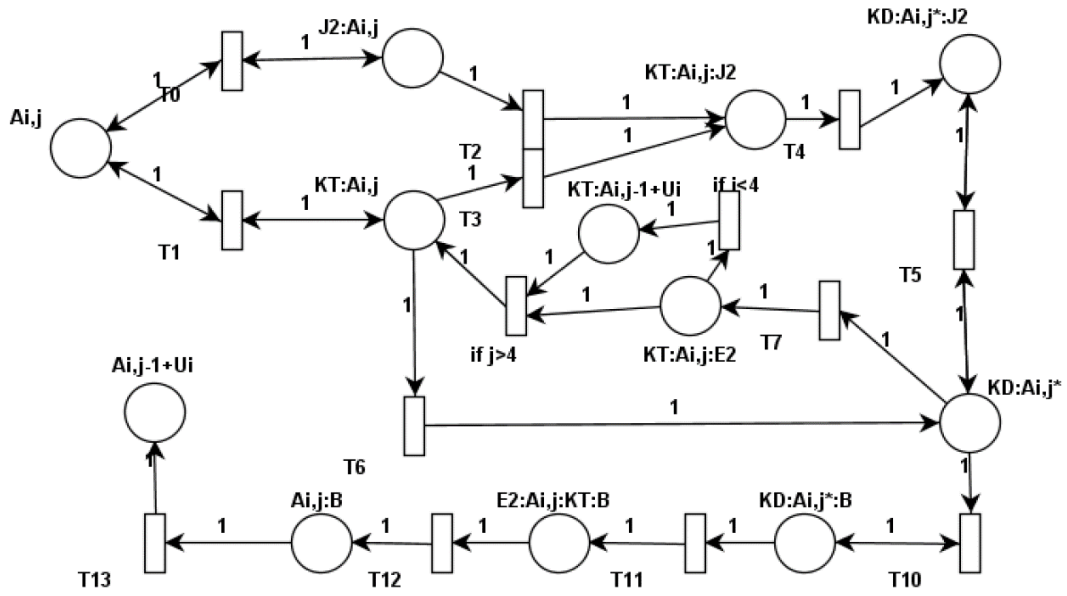


Figure 4. 7 Initial Petri net Model for Aggregated Protein (Ai.j). The model illustrates the interaction of key chaperones and co-chaperones in the disaggregation and refolding of aggregated proteins. Components include Ai.j: Aggregated protein. J2: DnaJ co-chaperone. KT: DnaK bound to ATP. KD: DnaK bound to ADP. E2: GrpE nucleotide exchange factor. B: ClpB.

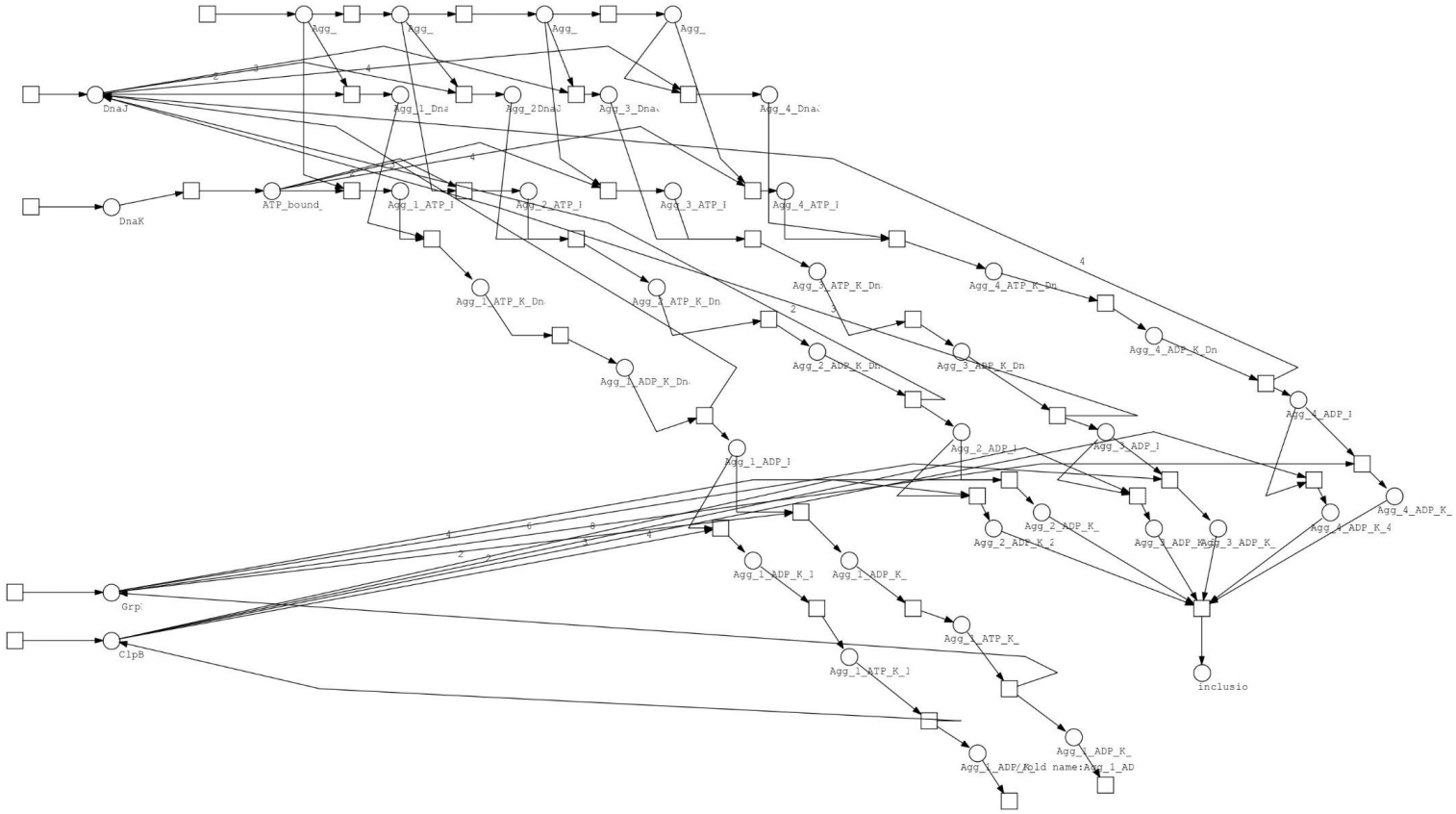


Figure 4. 8 Final Petri net Model for Aggregated Protein Disaggregation. The model illustrates the ClpB-mediated disaggregation of aggregated proteins into unfolded proteins, which can then be refolded into their functional state. The model captures the multi-step process of reversing protein aggregates, highlighting the role of ClpB and associated chaperone systems in resolving protein misfolding and aggregation.

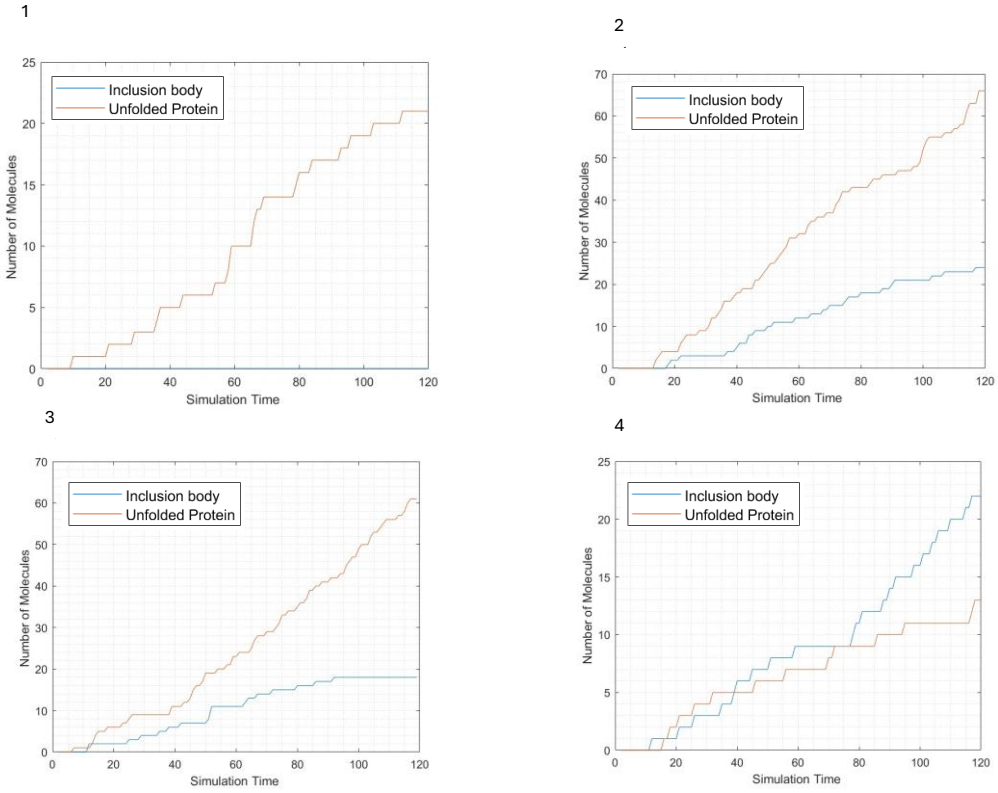


Figure 4. 9 Simulation Results of ClpB Chaperone Activity. Graph 1: Default scenario with ClpB and translation rate set at baseline levels, resulting in 21 unfolded protein molecules and no inclusion body formation. Graph 2: Increased translation rate by 20%, leading to 65 molecules of unfolded protein and 24 molecules of inclusion bodies, demonstrating the impact of elevated protein synthesis on misfolding and aggregation. Graph 3: Increased ClpB rate by 20%, resulting in 61 molecules of unfolded protein and 18 molecules of inclusion bodies, highlighting ClpB's role in reducing aggregation and enhancing protein solubility. Graph 4: Decreased ClpB rate by 20%, resulting in 13 molecules of unfolded protein and 22 molecules of inclusion bodies, underscoring the importance of ClpB in preventing protein aggregation.

The study conducted by Mogk. A explored the reversion of aggregation using ClpB and DnaK. The study demonstrated that ClpB is critical in resolving protein aggregates, working in concert with the DnaK/DnaJ/GrpE system. ClpB uses ATP hydrolysis to disentangle aggregated proteins, converting them into unfolded proteins that other chaperones can refold (Mogk, 1999). This aligns with the model's assumption that ClpB is essential for reversing aggregates into unfolded proteins.

The study highlighted that ClpB's disaggregation activity highly depends on the presence of DnaK/DnaJ, which prepares aggregated proteins for ClpB-mediated unfolding. This supports the model's assumption of an abundant presence of DnaK and DnaJ for ClpB analysis.

Figure 4.9 shows similar trends to those seen in this study. Under default conditions, the model simulates the pathways as they are, resulting in 21 unfolded protein molecules. This baseline scenario represents the system's equilibrium, where ClpB and other chaperones work together to balance protein folding and disaggregation.

When the translation rate increases by 20%, the model shows a rise in unfolded proteins and the formation of 24 molecules of inclusion bodies. This demonstrates that high

translation rates overwhelm the cellular folding machinery, increasing misfolding and aggregation. The results align with experimental studies showing that excessive protein synthesis exacerbates protein aggregation (Dukan et al., 2000).

Increasing the ClpB chaperone rate results in a 10% decrease in inclusion body formation and a corresponding increase in unfolded proteins. This highlights ClpB's role in resolving protein aggregates and converting them into unfolded proteins that other chaperones can refold. These findings are consistent with studies showing that ClpB activity reduces aggregation and enhances protein solubility (Mogk et al., 1999) (Barnett et al., 2005).

Decreasing the ClpB chaperone rate increases inclusion body formation and reduces unfolded proteins. This underscores the importance of ClpB in preventing protein aggregation and maintaining a pool of unfolded proteins for refolding. The results align with experimental evidence showing reduced ClpB activity impairs disaggregation and increases cellular stress (Barnett et al., 2005).

4.2.3 Role of GroEL and GroES chaperones in the unfolding of misfolded proteins

The GroEL and GroES chaperone system collaborates to create a protective environment essential for protein folding. This system necessitates the assembly of seven units of GroEL chaperones to form a cavity where misfolded proteins undergo correction to achieve their folded state. This structural arrangement is critical for converting unfolded polypeptides into properly folded proteins and rectifying misfolded proteins. However, a constraint lies in the energy demand; each cavity formation requires seven units of ATP (Goyal & Chaudhuri, 2015; Kim, 2008; Tian et al., 2015). Unlike other chaperones discussed, no preliminary model exists for this specific set, with only the final model depicted in Figure 4.10.

This section investigates the impact of GroEL and GroES chaperones on the accumulation of folded and misfolded proteins by varying the chaperone rates. The focus is solely on understanding their role without additional considerations. Limitations such as energy requirements will be explored in Chapter 5.

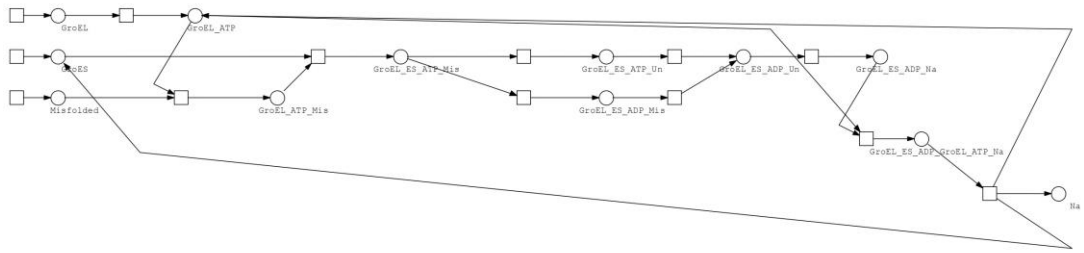


Figure 4. 10 Petri net Model of the GroEL/GroES Chaperonin System. The model represents the GroEL/GroES chaperonin system, where misfolded proteins are unfolded within the GroEL cavity, refolded into their correct conformation, and subsequently released as functional proteins. The transitions in the Petri net correspond to key steps in the chaperonin-assisted folding cycle.

Figure 4.11 illustrates the results of the GroEL/GroES chaperonin system, demonstrating the critical role these chaperones play in protein folding and preventing misfolding. The simulations highlight the importance of maintaining balanced activity levels of GroEL and GroES to enhance protein folding efficiency.

Under default transition rates, the model shows 89 molecules of folded protein and 12 molecules of misfolded protein, establishing a baseline for comparison. This reflects the system's equilibrium, where GroEL/GroES effectively fold most proteins while a small fraction misfolds.

A 20% decrease in the translation rate results in a 9-fold reduction in folded protein, with no accumulation of misfolded proteins. This suggests that lower translation rates reduce the substrate load on the chaperonin system, minimising misfolding and limiting overall protein production.

Increasing the GroEL chaperone rate by 20% prevents the accumulation of misfolded proteins and produces 90 folded protein molecules. This demonstrates that higher GroEL activity enhances folding efficiency, consistent with its role in providing a protected environment for protein refolding (Horwich et al., 2006).

Decreasing the GroEL chaperone rate by 20% does not lead to the accumulation of folded protein but results in 85 molecules of misfolded protein. This underscores GroEL's essential role in preventing misfolding and maintaining proteostasis.

Graph 5 shows a 20% decrease in GroES activity, resulting in 20 misfolded proteins and 8 folded protein molecules. This highlights GroES's role in facilitating GroEL's function.

Graph 6 shows a 20% increase in GroES activity, yielding 85 molecules of folded protein and 5 molecules of misfolded protein. This demonstrates that the best GroES activity enhances folding efficiency.

The results emphasise that GroEL and GroES must operate synergistically at balanced rates to achieve high levels of folded protein and minimise misfolding. This aligns with experimental studies showing that GroEL and GroES work together to encapsulate and refold misfolded proteins (Saibil et al., 2013).

The model effectively captures the role of the GroEL/GroES chaperonin system in protein folding and misfolding prevention. By simulating varying activity levels of GroEL and GroES, the model provides valuable insights into the mechanisms underlying protein folding efficiency and the importance of maintaining balanced chaperone activity.

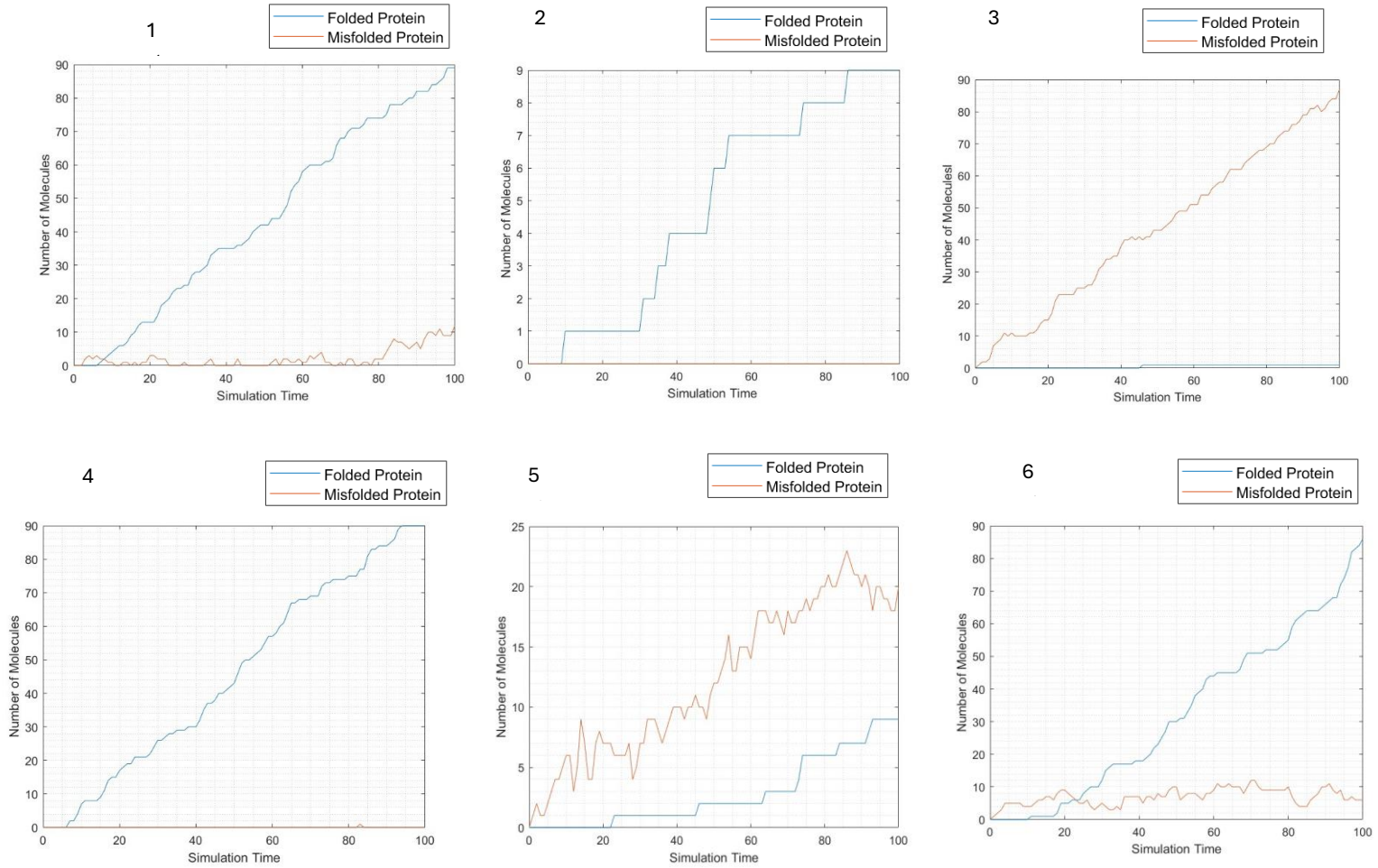


Figure 4. 11 Impact of Translation and Chaperone Rates on Folded and Misfolded Protein Accumulation. Graph 1: Folded and misfolded protein accumulation under default model conditions. Graph 2: Effect of increasing the translation rate by 20%. Graph 3: Impact of increasing the GroEL chaperone rate by 20%. Graph 4: Effect of decreasing the GroEL chaperone rate by 20%. Graph 5: Impact of decreasing the GroES chaperone rate by 20%. Graph 6: Effect of increasing the GroES chaperone rate by 20%.

4.2.4 Model performance evaluation with data collected from the FoldEco database

The data extracted from the tables in Chapter 3 have been integrated into the model, consolidating the protein synthesis pathways to evaluate the production of folded protein. This integration facilitates the identification of potential bottlenecks and allows for the adjustment of rates to improve folded protein production. While the preceding sections emphasised the importance of individual chaperone groups, combining these pathways proves instrumental in uncovering limitations at specific stages, which impact overall protein production dynamics. Figure 4.13 visually synthesises these integrated pathways; however, the specific place and transition names are not discernible due to the model's complexity and scale.

The three outputs of interest are the extent of folded protein, inclusion body formation, and misfolded protein accumulation. In Figure 4.12, Graph 1 displays the default outputs: 115 molecules of folded protein, 40 molecules of inclusion bodies, and no misfolded protein accumulation. In Graph 2, when the rate of GroEL activity is decreased by 30%, 36 molecules of inclusion bodies accumulate, 17 molecules of folded protein are produced, and no misfolded protein accumulates.

Graph 3 shows the outcomes when GroEL chaperone activity is decreased, and DnaK activity is increased by 20%, resulting in 72 molecules of the folded protein, 55 molecules of misfolded protein, and no accumulation of inclusion bodies. In Graph 4, when the transition rate increases while maintaining low GroEL chaperone activity, the model yields 85 molecules of folded protein, 30 molecules of inclusion bodies, and 15 molecules of misfolded protein. These results underscore the crucial roles of DnaK and GroEL in enhancing the amount of folded protein, highlighting their impact on the overall folding process and the potential trade-offs with misfolding and inclusion body formation.

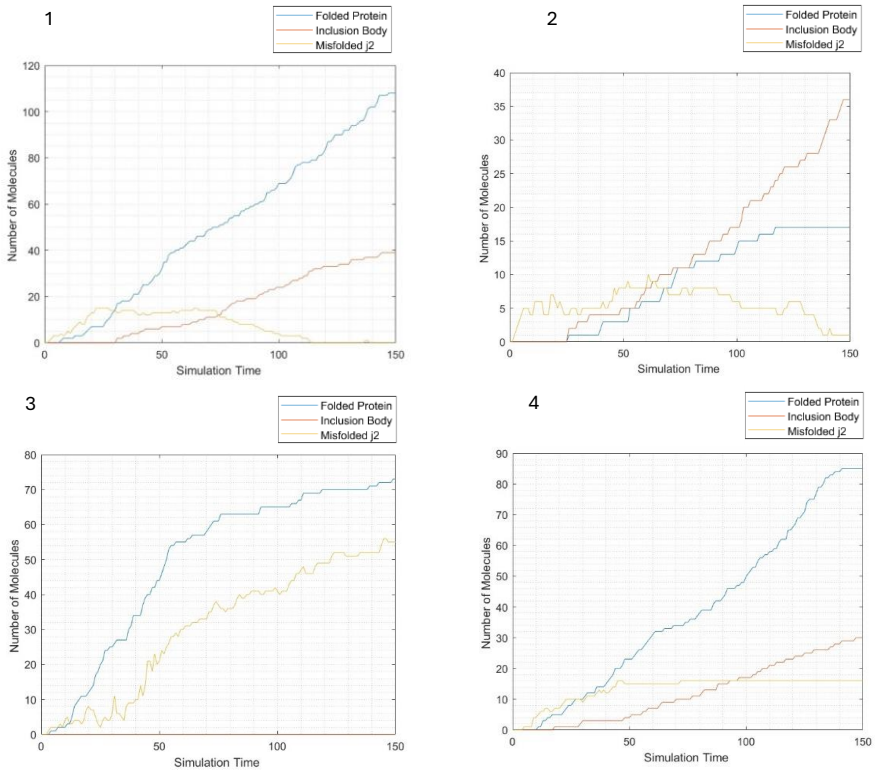


Figure 4. 12 Impact of Chaperone Activity on Protein Synthesis and Folding Efficiency. The figure illustrates the effects of varying GroEL and DnaK chaperone activity on protein synthesis and folding efficiency. Graphs 1–4 depict scenarios including default conditions, decreased GroEL activity, increased DnaK activity, and imbalanced synthesis rates with low GroEL activity, respectively. These results highlight the critical roles of chaperones in maintaining proteostasis and protein production.

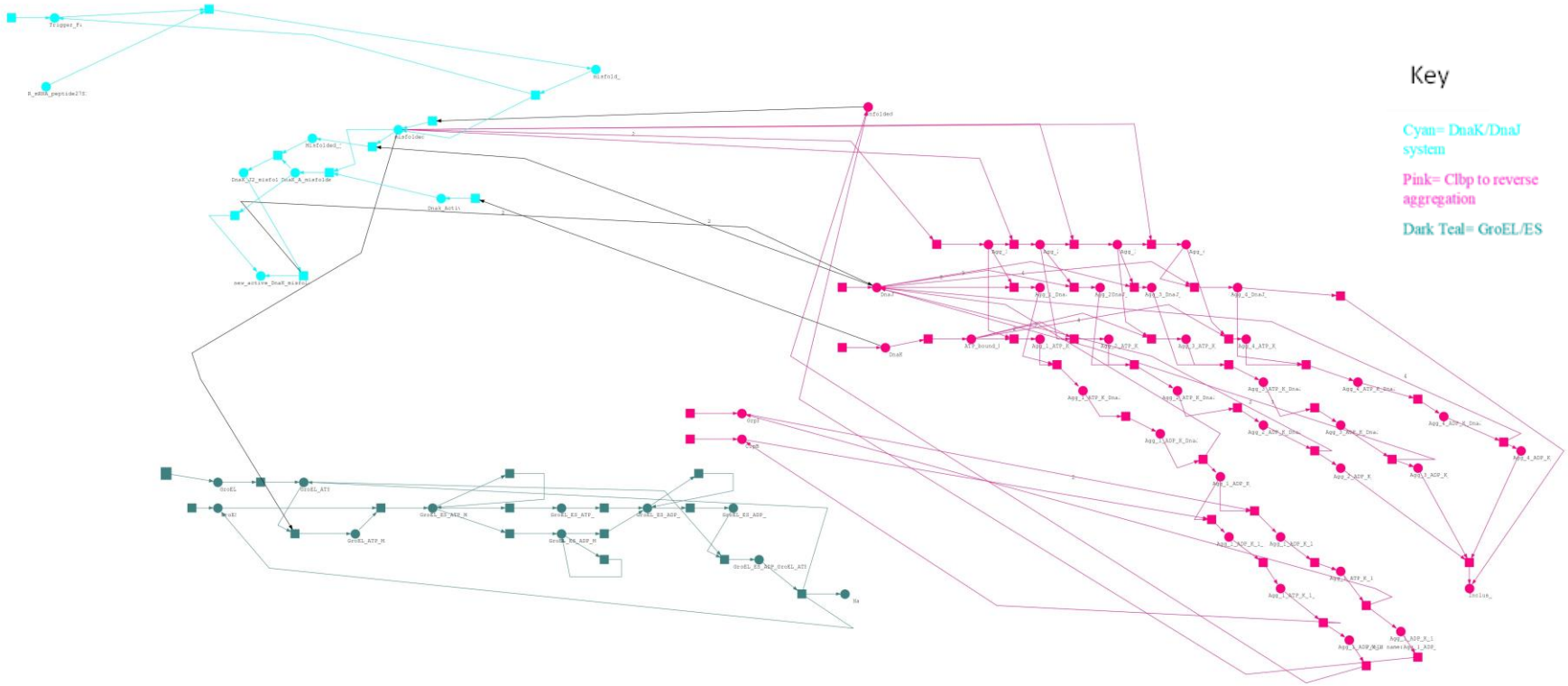


Figure 4. 13 Diagrammatic Representation of Protein Folding Pathways Combined with Chaperone Groups in *E. coli*.

4.3 TCA cycle and oxidative phosphorylation

The Tricarboxylic Acid (TCA) cycle is a crucial metabolic pathway in aerobic organisms. It plays a central role in the oxidative breakdown of acetyl-CoA derived from carbohydrates, fats, and proteins, leading to the generation of ATP and NADH. Modelling the TCA cycle using Petri nets provides a robust framework for understanding its dynamic behaviour and the interplay of its constituent reactions.

In a Petri net model of the TCA cycle, key metabolites such as citrate, isocitrate, α-ketoglutarate, succinyl-CoA, succinate, fumarate, and malate are represented as "places" in the Petri net. The transitions in the Petri net represent the enzymatic reactions that interconvert these metabolites. For example, the conversion of citrate to isocitrate by the enzyme aconitase is described as a transition in the Petri net, with appropriate input and output arcs connecting the places corresponding to citrate and isocitrate, (De Mets et al., 2019).

The TCA cycle involves a series of redox reactions that generate NADH, a crucial electron carrier in the electron transport chain. (Goel et al., 2023). The Petri net model representing the TCA cycle effectively captures these redox reactions, appropriately incorporating the electron carrier molecules. Figure 4.14 presents the Petri net model constructed for the TCA cycle.

Additionally, the Petri net model of the TCA cycle accommodates the coupling of the cycle with oxidative phosphorylation. The generation of ATP through substrate-level phosphorylation and the production of NADH, which influence the electron transport chain, is represented in detail.

In conclusion, modelling the TCA cycle as a Petri net offers a systematic and visual representation of its intricate biochemical processes and regulatory mechanisms. This approach facilitates the in-depth analysis of the cycle's behaviour and integration with other metabolic pathways, thereby contributing to a better understanding of cellular metabolism. Consequently, it serves as a platform for the protein machinery to source energy and reduce power requirements. Within the integrated model, it provides reducing agents for disulphide bond formation. To exploit these relationships more effectively, these pathways were paired with the protein synthesis pathways to understand the impact of protein synthesis on cellular energy requirements.

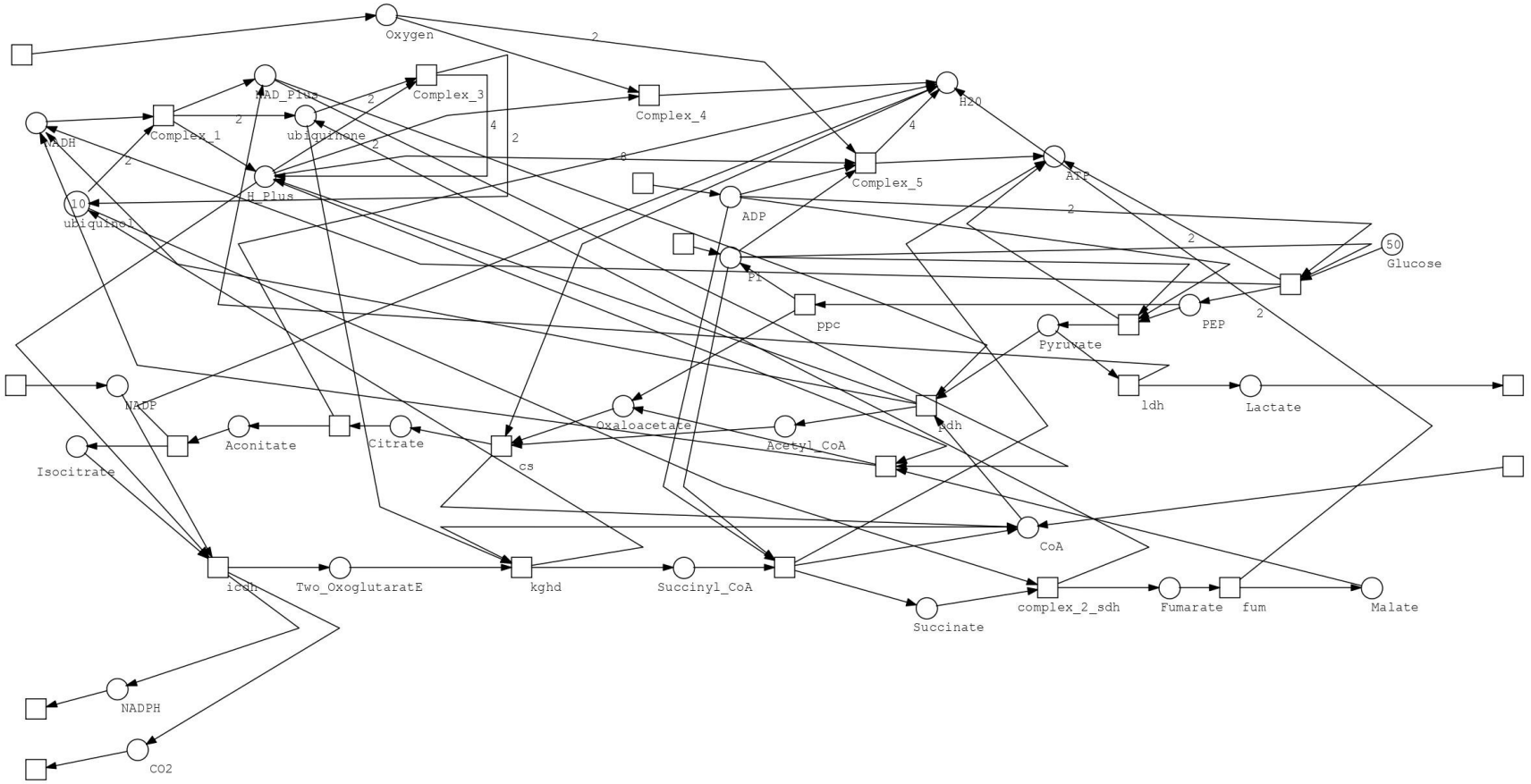


Figure 4. 14 Petri net Model of the TCA Cycle and Oxidative Phosphorylation. The model represents the TCA cycle and oxidative phosphorylation pathways, where transitions are labelled with known biochemical reactions. Each transition corresponds to a specific enzymatic step, capturing the flow of metabolites and energy production.

4.3.1 Modelling of the *E. coli* glycolytic pathway

Once the model was constructed, confirming that key metabolites such as ATP, H^+, and NAD^+ are produced when glucose is used as the energy source was essential. Glucose serves as an input for the TCA cycle and glycolysis, driving the production of ATP through oxidative phosphorylation and generating the proton motive force (PMF) required for cellular energy transduction.

Although glucose is continuously supplied into the system boundary in this simulation to investigate other potential limiting factors, this scenario is not entirely realistic. In vivo, glucose availability is often limited and subject to regulatory mechanisms. This limitation will be addressed in future model iterations to reflect physiological conditions better.

Figure 4.15 illustrates the simulation results under these conditions: When glucose is provided without limitation, and the model is run for 100 seconds (simulation time), 142 molecules of ATP are produced. This confirms that the TCA cycle and associated pathways are functioning as expected, generating the energy currency required for cellular processes.

Additionally, 160 molecules of H^+ are produced under default conditions, which is essential for establishing the proton motive force needed for ATP synthesis and other energy-dependent processes.

These results validate the model's ability to simulate energy production and highlight the importance of glucose as a substrate for the TCA cycle. However, the unrealistic assumption of unlimited glucose availability underscores the need for future refinements to incorporate regulatory mechanisms and substrate limitations.

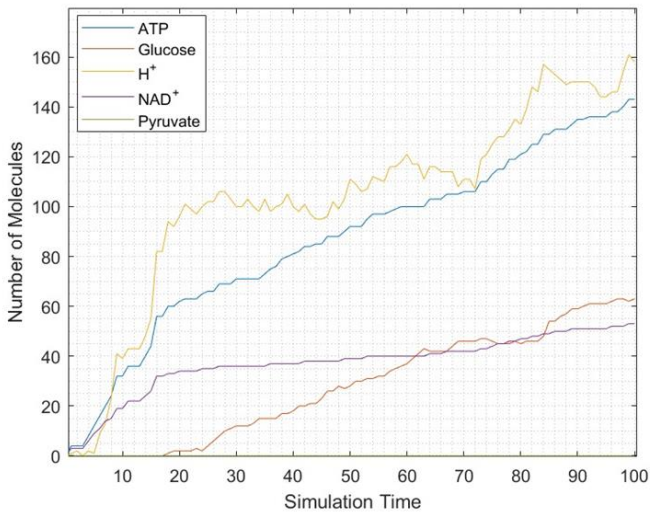


Figure 4. 15 Output Variables from the TCA Cycle and Oxidative Phosphorylation Petri net Model. The model simulates the production of key metabolites—ATP, H^+, and NAD^+—using glucose as the energy source.

4.3.2 Evaluating the model's performance by limiting oxygen availability

Lactate production during fermentation in *E. coli* is a metabolic indicator that reflects the cellular response to environmental conditions and nutrient availability. Under anaerobic conditions, *E. coli* can undergo mixed-acid fermentation, leading to the production of lactate as a means of regenerating NAD^+ from NADH and maintaining redox balance.

Furthermore, lactate production in *E. coli* is relevant to biotechnology and bioprocess engineering. *E. coli* is widely used in bioreactor systems to produce various compounds, and understanding the factors influencing lactate production can inform strategies for optimising metabolic pathways and enhancing productivity.

Figure 4.16 portrays the impact of reduced oxygen on lactate production. However, lactate production takes place simultaneously with protein synthesis in dynamic cultures, so this model must be connected to the protein synthesis pathways to understand the impact of reduced oxygen.

Graph 1 presents the default setting where the TCA cycle is not integrated with protein synthesis. Under these conditions, the accumulation of ATP remains at one hundred forty molecules. Glucose is fed into the system. It is not being used up because it has not been attached to any folding pathways at this stage, and some lactate molecules are recorded as twenty. In Graph 2, when oxygen production is reduced by twenty per cent, one hundred twenty molecules of ATP are produced, and twenty lactate molecules are produced again. Glucose is fed into the system and is not used up. This indicates that a decrease in oxygen availability alone, when it is not linked to protein synthesis, does not impact the lactate level, which is expected because lactate is produced during fermentation. This test is indicating the pathways are operating as intended.

However, in Graph 3, when protein synthesis is integrated and oxygen is limited, lactate accumulates to seventy molecules, and ATP and glucose are depleted. In Graph 4, when the oxygen rate is reduced by twenty per cent, lactate accumulates to ninety molecules, and ATP and glucose depletion are observed. This demonstrates that when integrating protein synthesis, the limitations of the TCA cycle are better understood, particularly under conditions of reduced oxygen availability.

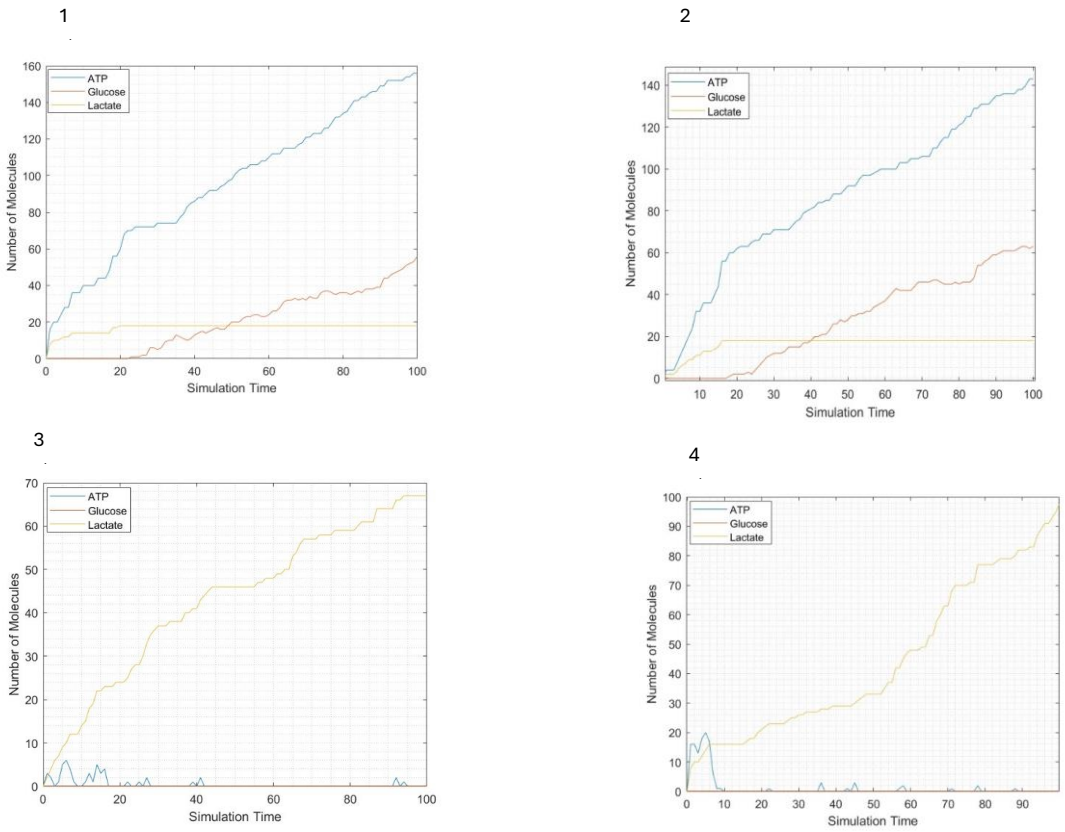


Figure 4. 16 Impact of Oxygen Availability on the TCA Cycle and Protein Synthesis Pathways. Graph 1: Default oxygen rate, showing baseline metabolite production when the TCA cycle is not integrated with protein synthesis pathways. Graph 2: 20% reduction in oxygen rate, illustrating the effects of oxygen limitation on the TCA cycle in isolation. Graph 3: TCA cycle integrated with protein synthesis pathways at default oxygen rates, demonstrating the interplay between energy production and protein folding. Graph 4: TCA cycle integrated with protein synthesis pathways with a 20% decrease in oxygen rate, highlighting the impact of oxygen limitation on protein production.

4.3.3 Evaluating the model's performance by limiting available glucose

The glucose test is pivotal in validating the tricarboxylic acid (TCA) cycle and oxidative phosphorylation model because glucose is the essential energy substrate fuelling these metabolic pathways. This test ensures that the model accurately replicates the production of crucial metabolites, namely ATP, H^+, and NAD^+, and confirms their in cellular energy metabolism and redox balance. By systematically examining the effects of glucose availability on these metabolites, the test offers profound insights into the intricate relationship between energy production and protein synthesis pathways.

Regarding simulation results shown in Figure 4.17, Graphs 1 and 2 depict the TCA cycle in isolation from protein synthesis pathways. Under default conditions, as illustrated in Graph 1, the model demonstrates the accumulation of 70 molecules of glucose, 140 molecules of H^+, 140 molecules of ATP, and 50 molecules of NAD^+. These findings substantiate the model's efficacy in simulating energy production under baseline conditions. Conversely, Graph 2, which illustrates the impact of a 20% increase in

glucose production rate, reveals a striking enhancement in glucose availability to 900 molecules, thereby underscoring the model's responsiveness to variations in substrate concentrations.

Further analysis is presented in Graphs 3 and 4, which integrate protein synthesis pathways with the TCA cycle. In Graph 3, where glucose availability is set at 100 molecules, the incorporation of folded protein pathways results in the depletion of glucose and the production of 130 molecules of folded protein. The total consumption of glucose results in a cessation of protein production, highlighting the critical dependency of protein synthesis on energy substrate availability.

Graph 4 demonstrates the scenario of limited glucose availability, where only 50 glucose molecules are present. In this case, the system generates 60 molecules of folded protein, illustrating a direct positive correlation between glucose availability and protein production. This observation underscores the notion that glucose acts as a limiting factor in the synthesis of proteins.

Key insights from these simulations reveal that glucose availability profoundly influences the interplay between energy production and protein synthesis. The TCA cycle and oxidative phosphorylation are integral in providing the ATP and NADH necessary for essential processes such as protein folding and the activity of molecular chaperones. Furthermore, by integrating protein synthesis pathways, the model adeptly captures the realistic restrictions imposed by glucose availability, thus enhancing the accuracy of its representation of cellular metabolism.

Additionally, the positive correlation between glucose consumption and the production of folded proteins aligns with experimental studies in the literature. Notably, Buttgereit and Brand (1995) demonstrated that ATP production is intricately coupled with cellular energy demands, including those required for protein synthesis. Their findings substantiate the model's key conclusion that glucose availability is intrinsically linked to protein production efficiency (Buttgereit & Brand, 1995).

Similarly, Vander Heiden et al. (2009) underscored the of glucose metabolism in facilitating biosynthetic processes, including protein synthesis, by generating ATP and NADH. Their work corroborates the model's assertion that glucose is a limiting factor in protein production (Vander Heiden et al., 2009) (Tao et al., 1999).

In summary, the glucose test is necessary to understand the interactions within the TCA cycle and oxidative phosphorylation, confirming glucose's vital role as an energy source and a critical driver of protein synthesis within cellular metabolism.

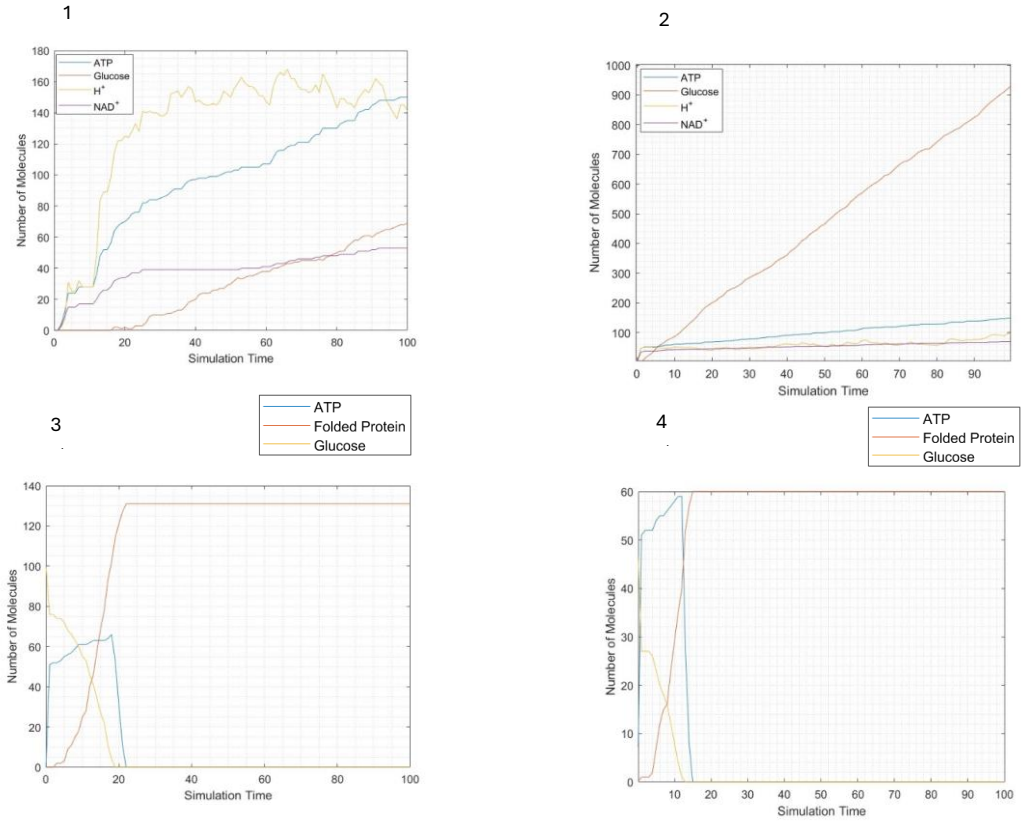


Figure 4. 17 Impact of Glucose Availability on the TCA Cycle and Protein Synthesis Pathways. Graph 1 presents the default setting of the non-integrated model, showing baseline metabolite production when the TCA cycle operates independently of protein synthesis pathways. Graph 2 illustrates the effect of a 20% increase in glucose production rate, demonstrating how elevated glucose availability impacts metabolite accumulation. Graph 3 shows the integration of the TCA cycle with protein synthesis pathways, with glucose availability limited to 100 molecules, highlighting the dependency of protein production on energy supply. Graph 4 introduces a limited glucose supply of 50 molecules, reducing protein production and demonstrating the positive correlation between glucose availability and folded protein yield.

4.4 Sec translocation

The Sec translocation pathway is the central transport system between the cytoplasm and periplasm within *E. coli*. Sections 4.4.1 and 4.4.2 model two possible pathways, identified as coupled and uncoupled translocation (Gowrishankar & Harinarayanan, 2004; Johnson et al., 2020). This section will explain the step-by-step formulation of the Petri net model for both pathways and highlight the differences. In section 4.4.3, the models will be evaluated using graphs generated from the simulation with a range of translocation rates.

The pathway's activity relies on the correct action taken by the signal peptide outlined in Chapter 3, Section 3.3.1; integration of the pathway with DnaK establishes connections between translocation and protein synthesis pathways. This section will develop a system boundary via the Sec translocation pathway and demonstrate the close interactions between protein synthesis and disulphide bond formation. Chapter 5 will investigate human growth hormone as the model protein.

The preliminary model to demonstrate translocation presented a combined view of the Tat and Sec pathways; however, this integrated model then needed to be modified and expanded to accommodate for the different types of data available for the two pathways, and consequently, a decision was made to develop them into separate models. Figure 4.18 presents the initial model where T denotes the trigger factor, Ui, Ni, MF and MA denote the unfolded protein, the native protein, the misfolded protein, and the aggregated protein, respectively. The other nomenclature including the chaperone names remained the same

4.4.1 Modelling of coupled Sec translocation

The data shows that the route taken for each protein depends on the type of signal peptide used in the process. Inner membrane proteins use coupled translocation, also called co-translational translocation, because the ribosome remains attached to the protein. When the protein chain forms, it is translocated to its correct location (Crossley et al., 2024).

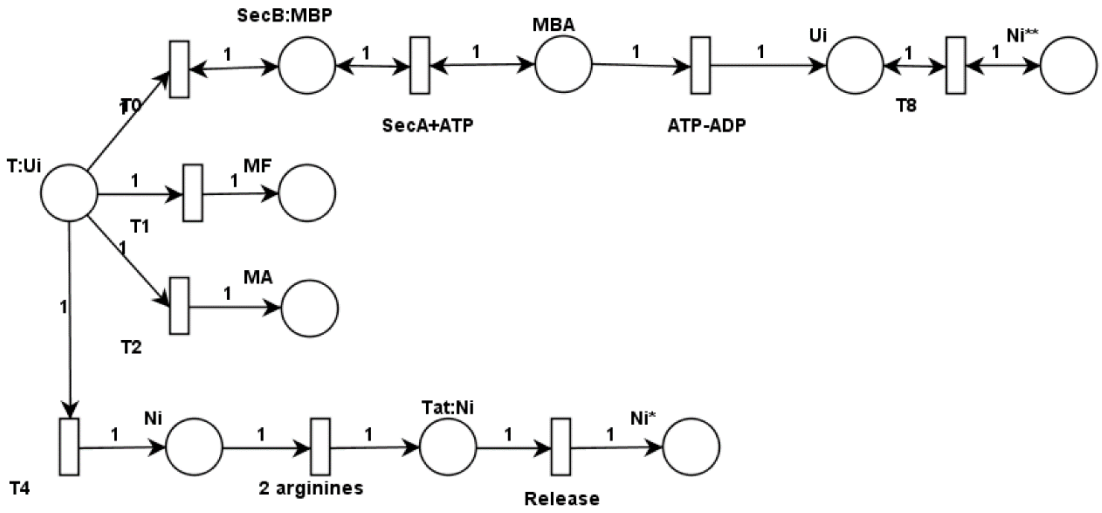


Figure 4. 18 Initial Petri net model of translocation combining Sec and Tat translocation pathways.

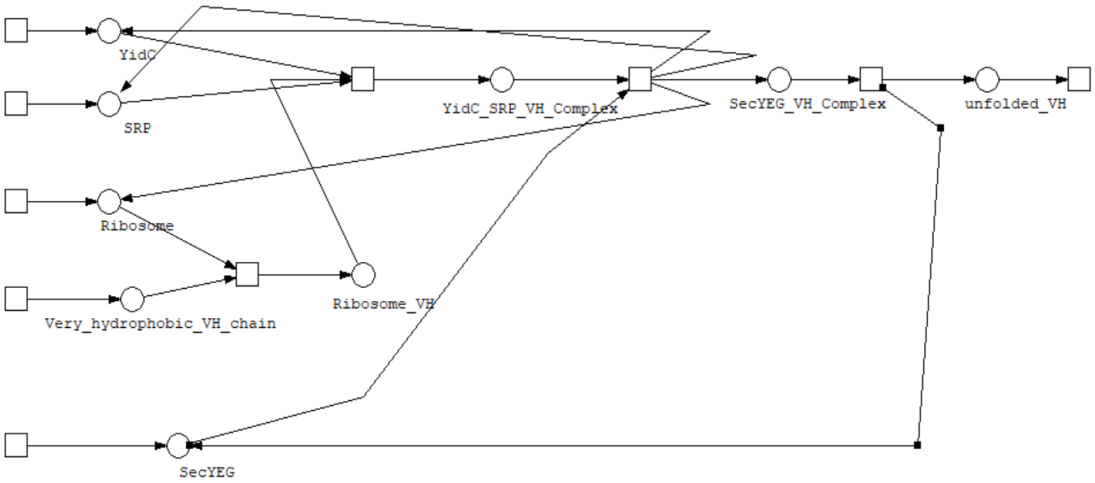


Figure 4. 19 The Petri net pathway for co-translational translocation. The names of some places have been shortened and abbreviated for visual ease. Two examples are seen in Very_hydrophobic_VH_chain, a protein chain with a very hydrophobic signal peptide, and Unfolded_VH, an unfolded protein with a hydrophobic signal peptide. This is also useful for distinguishing the highly hydrophobic chains from those of low hydrophobic chains.

Figure 4.19 shows that as the amino acid bonds form the protein chain within the ribosome, the protein is translocated within the cytoplasm. Signal recognition proteins (SRP) can detect the protein chain with a highly hydrophobic single peptide, which results in a quick translocation pathway that does not require ATP hydrolysis, (Jomaa et al., 2016, 2017; Petriman et al., 2018). However, the proton motive force is needed to allow movement. When the model is integrated with the other components, the representation of this biological phenomenon will be included in Chapter 5.

Once SRP and the protein chain form a complex chaperone, YidC is required to transfer the protein chain within the membrane, which will attach to SecYEG. SecYEG has formed a channel between the cytoplasm and the periplasm of *E. coli*. In the presence of YidC, the protein can integrate within the membrane through the channel and be exported to the periplasm, where further modifications will occur (Samuelson et al., 2000; Yi et al., 2004)This is only half the Sec translation model. The other half is uncoupled Sec translocation, which will be modelled together in section 4.4.3.

4.4.2 Modelling of uncoupled Sec translocation

During the transportation of soluble protein translocation, proteins follow the uncoupled translocation pathway, which is also called post-translational translocation, where the protein chain is detached from the ribosome. These pathways require more chaperone interaction than coupled Sec translocation, requiring ATP hydrolysis and proton motive force to allow for movement within the cytoplasm (Diao et al., 2012; Hartl et al., 1990).

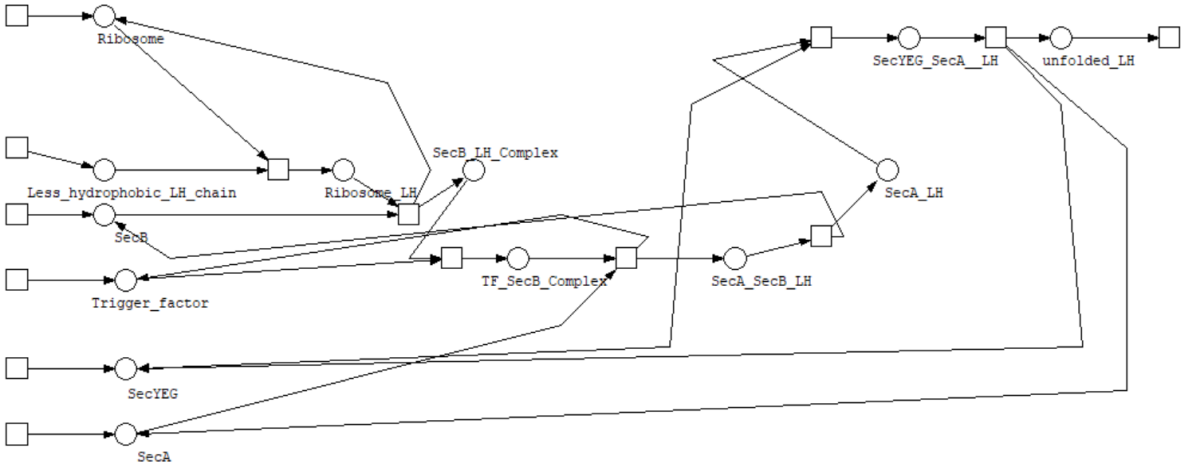


Figure 4. 20 Petri net model for uncoupled translocation in Sec translocation pathways. Less_hydrophobic_LH chain refers to a protein chain with a less hydrophobic signal peptide. The trigger factor is shortened to TF within the model. Unfolded_LH refers to the unfolded protein with a low hydrophobic signal peptide, which will be translated with a different pathway.

There are distinct differences between coupled and uncoupled translocation. The most crucial difference is the chaperones' contribution to protecting the protein chain from aggregating. Once the ribosome covalently bonds the amino acids into a peptide chain, Sec B binds to the preprotein to ensure it does not aggregate.

The trigger factor is another chaperone whose goal is to reduce the protein aggregation as it waits for the next step. A complex is formed between the trigger factor, recombinant protein and Sec B, which is recognised by Sec A. The trigger factor is released and recycled once Sec A binds to the unfolded protein. Sec A binds to SecYEG to open a pore for the unfolded protein to be released into the periplasm (Koch et al., 1999; Ullers et al., 2004). The chaperones used are all recycled and maintain their concentration during normal conditions.

4.4.3 Evaluating the model by testing different rates of translocation for a range of signal peptides

Figure 4.21 shows the final Sec translocation model, combining the coupled and uncoupled pathways. Because the signal peptide differs, two different outputs represent the translocated protein. One of these pathways will be selected when a model protein is to be chosen for simulations, depending on the nature of that protein.

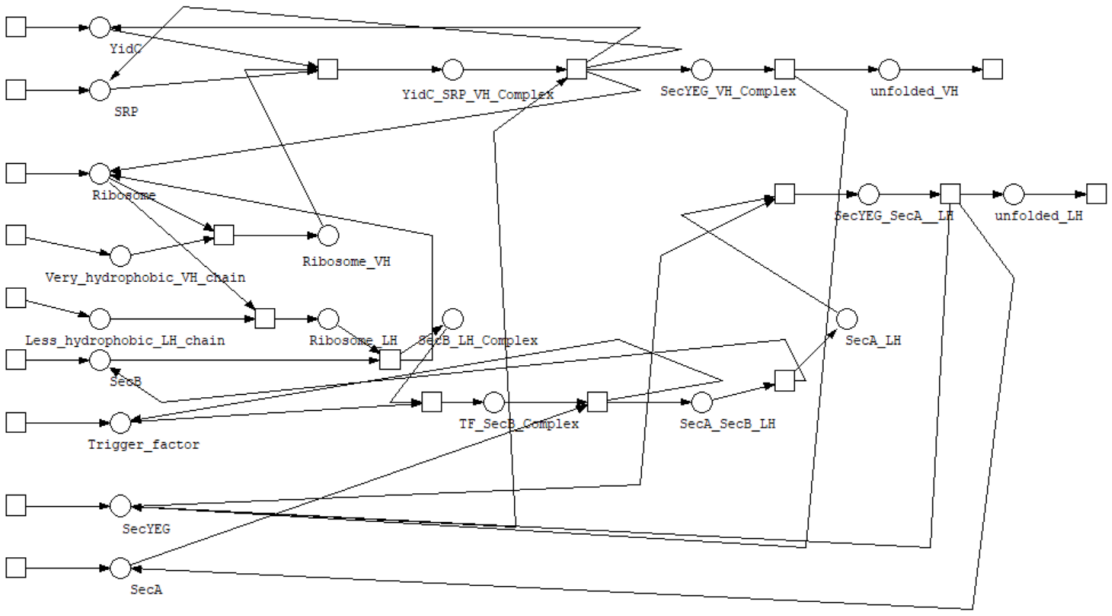


Figure 4. 21 Final Petri net model of the *E. coli*. Sec translocation pathways.

To better understand these pathways, the model is tested under a range of conditions. This will include changing the transport rate into the periplasm, which will model varying strengths of the signal peptide. Chapter 5 will present a case study that explores the model's behaviour based on the selection of different signal peptides. This section will explore only a few set ranges for a limited number of signal peptides to improve the understanding of the translocation pathways as represented in the integrated model.

Figure 4.22 examines the impact of varying transport rates from the cytoplasm to the periplasm, imposing limitations on molecule availability. The results reveal key trends in protein translocation efficiency, particularly for proteins with different hydrophobicity levels.

In the default conditions illustrated in Graph 1, unfolded proteins steadily increase for both high-hydrophobicity (VH) and low-hydrophobicity (LH) pathways. Notably, the quantity of unfolded VH proteins surpasses that of LH proteins by 20%, indicating the challenge of translocating hydrophobic proteins. This observation is consistent with research suggesting that such proteins typically demand more incredible energy and chaperone support for their translocation (Driessen & Nouwen, 2008).

The transport rate for VH proteins was reduced by 10%, leading to a 20-molecule decrease in the total number of unfolded VH proteins, while the number of unfolded LH proteins increased. This indicates a competitive dynamic for the transport channel, allowing LH proteins to benefit from the slower rate of VH translocation. This pattern aligns with experimental findings showing that transport systems favour less hydrophobic proteins when resources are scarce (Tsirigotaki et al., 2017).

Decreasing the transport rate for LH proteins by 10% reduces 20 unfolded LH proteins. However, this adjustment does not affect the translocation of VH proteins. This highlights that the coupled pathway used for membrane proteins is more efficient and is only impacted when its rate is lowered. This finding is expected, given that membrane proteins play a crucial role in cell growth and are prioritised over other host cell proteins. This study emphasises the importance of understanding protein transport dynamics in cellular processes (Daley et al., 2005).

When the initial molecule supply is limited to 100 molecules, the pathways halt once the molecules are exhausted. The unfolded protein formation rate is initially higher, as the model does not need time to create molecules but levels off as the supply is depleted. This reflects the dependency of protein translocation on substrate availability, a trend observed in studies of nutrient-limited bacterial systems (Browning et al., 2010).

The simulation results align closely with experimental studies, demonstrating that hydrophobicity, transport rates, and molecule availability impact protein translocation efficiency.

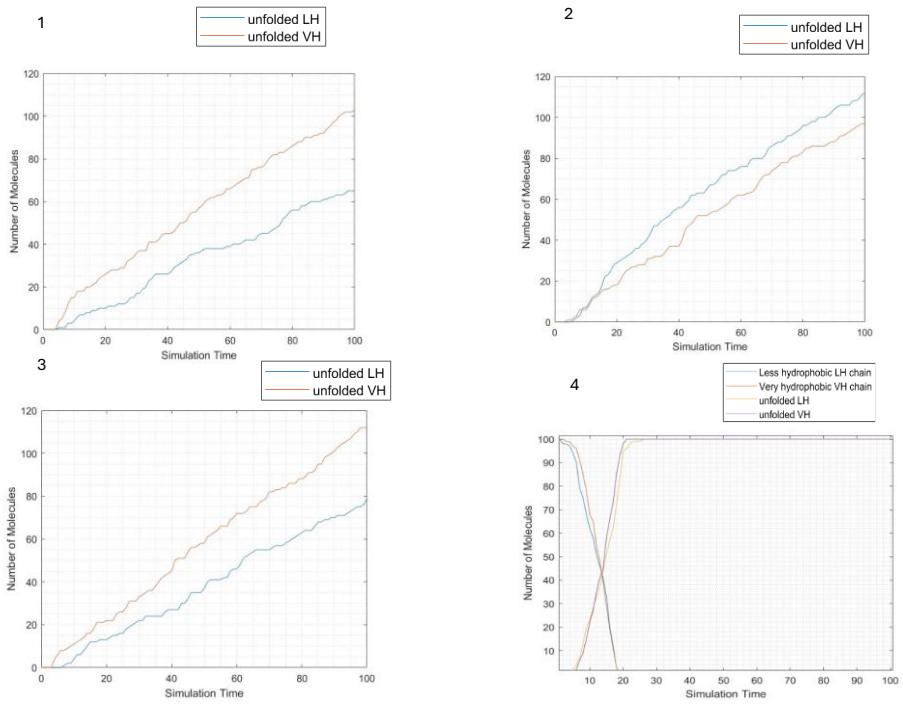


Figure 4. 22 Impact of Transport Rates and Molecule Availability on Protein Translocation. Graph 1: Default settings, showing the accumulation of unfolded_VH (very hydrophobic) and unfolded_LH (less hydrophobic) proteins under baseline conditions. Graph 2: 10% decrease in translocation rate for unfolded_VH, reducing its accumulation and increasing unfolded_LH due to competition for transport channels. Graph 3: 10% decrease in translocation rate for unfolded_LH, reducing its accumulation without significantly affecting unfolded_VH, indicating the prioritisation of the coupled pathway for membrane proteins. Graph 4: Initial molecule supply limited to 100 molecules, demonstrating the dependency of protein translocation on substrate availability and the eventual halting of pathways once molecules are depleted.

The simulation employs Petri nets to model the SecYEG-mediated protein translocation process, focusing on the transport rate and other parameters influencing the total number of molecules generated. SecYEG serves as the central transport channel for both the post-translational and co-translational pathways. The simulation results suggest that the coupled translocation pathway is favoured under normal conditions. However, to validate this observation, the functionality of SecYEG was systematically reduced by limiting its molecular count and recycling capacity. This approach imposes constraints on the model, allowing the identification of which pathway remains functional under limiting conditions and the quantification of the differences compared to scenarios without such limitations.

Figure 4.23 illustrates the impact of SecYEG availability on the model's output. Initially, each pathway was assigned a molecule count of one hundred to track protein accumulation and identify potential bottlenecks. The results reveal that when SecYEG is limited to 200 molecules, unfolded protein chains from both pathways accumulate in the cytoplasm as complexes with their respective chaperones, unable to traverse the membrane. This accumulation becomes more pronounced when SecYEG is further restricted to 100 molecules, indicating a severe bottleneck in protein translocation.

Interestingly, the simulation demonstrates a recurring pattern where highly hydrophobic proteins are transported more efficiently than less hydrophobic ones, even under SecYEG-limiting conditions. This suggests that protein hydrophobicity is critical in determining translocation efficiency, potentially due to differences in chaperone binding affinity or interaction with the SecYEG channel.

The simulation findings align with experimental studies that highlight the critical role of SecYEG in protein translocation and its limitations under stress conditions. For instance, studies by Dalal et al. (2012) and Tsukazaki et al. (2008) have demonstrated that SecYEG capacity is a limiting factor in protein export, particularly during high protein synthesis. Additionally, the observed preference for hydrophobic proteins is consistent with experimental evidence showing that hydrophobic signal sequences enhance translocation efficiency (Zhang & Shan, 2014).

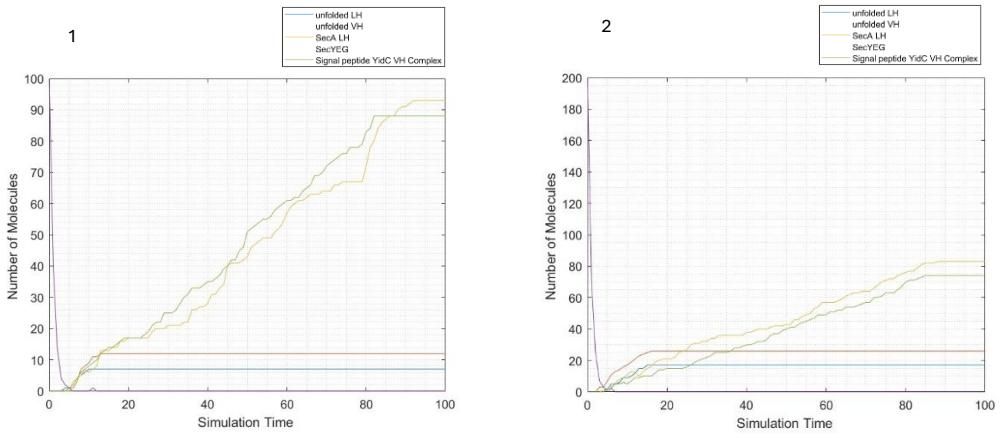


Figure 4. 23 Limited Capacity of SecYEG Translocon Restricts Unfolded Protein Transport, Leading to Cytoplasmic Accumulation. The SecYEG translocon is constrained to handle 200 molecules under normal conditions and 100 molecules under stress conditions. This limited capacity restricts the number of unfolded proteins transported to the periplasm, resulting in their accumulation within the cytoplasm. This bottleneck highlights the critical role of SecYEG in protein translocation.

The model reveals how alterations in these pathways affect the interactions between chaperones and the effectiveness of protein transport. The simulation shows that the rate of the Sec pathways can be adjusted to examine these interactions thoroughly.

One key finding from the model is that hydrophobic proteins are transported more effectively through the Sec pathways. Even under constraints like fewer SecYEG molecules, hydrophobic proteins still achieve higher export levels than less hydrophobic ones. This indicates that hydrophobicity is essential for efficient protein transport across the membrane.

The simulation underscores the importance of SecYEG availability and protein hydrophobicity in determining the efficiency of protein translocation. By comparing these results to experimental data, we validate the model's predictive capabilities and highlight key factors influencing protein transport across the membrane.

4.5 Tat translocation

Section 4.5.1 shows the initial model for Tat translocation. Expansion of the model was required to understand each chaperone's role in the transportation of proteins. Tat translocation can only export fully folded proteins, and the TorA and TorD act as checkpoints to ensure that only folded proteins reach the periplasm. One input variable for this sub-model is a fully folded protein attached to TorA to find TorD and indicate the pathways. This is very different from the input of Sec translocation, an unfolded protein; therefore, they have been separated.

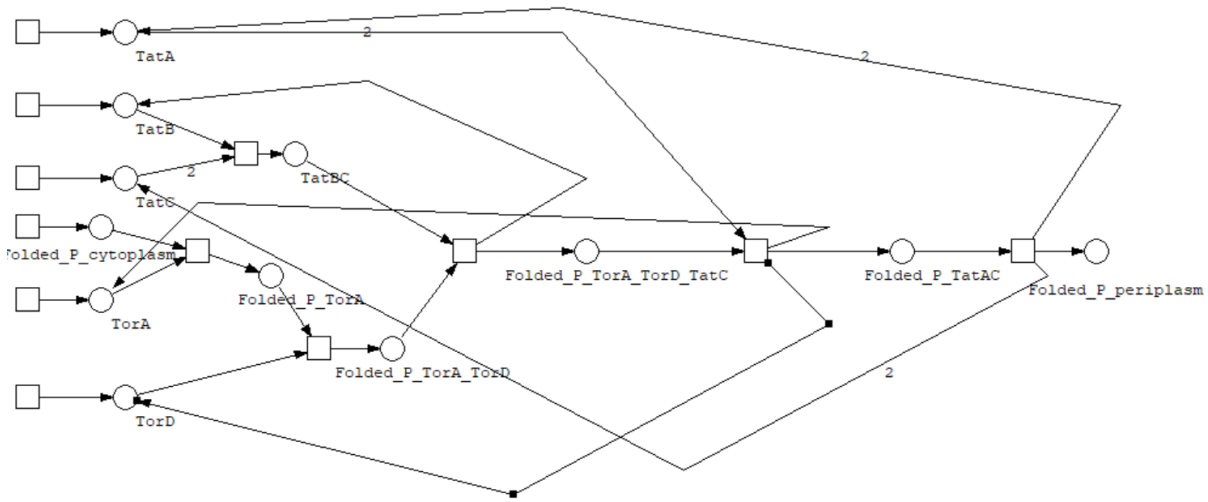


Figure 4. 24 Tat translocation pathways within *E. coli*, the folded protein is transported to the periplasm by several interactions with Tat A, B and C. The chaperone names, as denoted in the model, remain the same as in the literature: Folded_P_cytoplasm refers to folded protein in the cytoplasm, and Folded_P_periplasm refers to folded protein exported to the periplasm.

The Tat system transports fully folded proteins across the bacterial inner membrane into the periplasmic space. The components Tat A, Tat B, and Tat C are assumed to be abundant within the system boundary. The Tat B and Tat C proteins recognise the signal peptide attached to the protein, which is often bound by the chaperone TorD, indicating it is ready for transport (Bageshwar et al., 2021).

Tat C is the system's central receptor and binds to Tat B to initiate the translocation pathway. If the protein is correctly folded, it is detected by TorD, which specifically binds to the protein's signal peptide, typically TorA. Once the Tat BC complex recognises the TorD-protein complex, the Tat system is primed for translocation.

Tat B is an adaptor chaperone that navigates the folded protein to Tat C and facilitates translocation to the membrane. Tat A then oligomerises to form a channel and acquires the necessary receptors to bind to Tat C, forming the Tat AC complex. Upon complex formation, TorD and the signal peptide (e.g., TorA) detach from the protein.

A proton motive force (PMF) is required to drive the protein's translocation from the Tat AC complex into the periplasm. This energy dependency is crucial for the successful transport of folded proteins across the inner membrane (Barnett et al., 2008; Bolhuis et al., 2001; Habersetzer et al., 2017; Oertel et al., 2015; Smith et al., 2017).

This section will investigate the impact of these chaperones on the total amount of protein transported. In addition, the rate of each chaperone will be manipulated to understand what impacts the Tat system the most. The proton motive force has not been modelled here but will be integrated later with all the sub-models.

4.5.1 Modelling the role of Tat A, Tat B and Tat C transmembrane proteins in translocation pathways

The Tat translocation pathways reuse the Tat chaperone components; therefore, they are referred to as being recycled (Palmer & Berks, 2012)The model presents the Tat components working together to transport the complete folded protein to the periplasm.

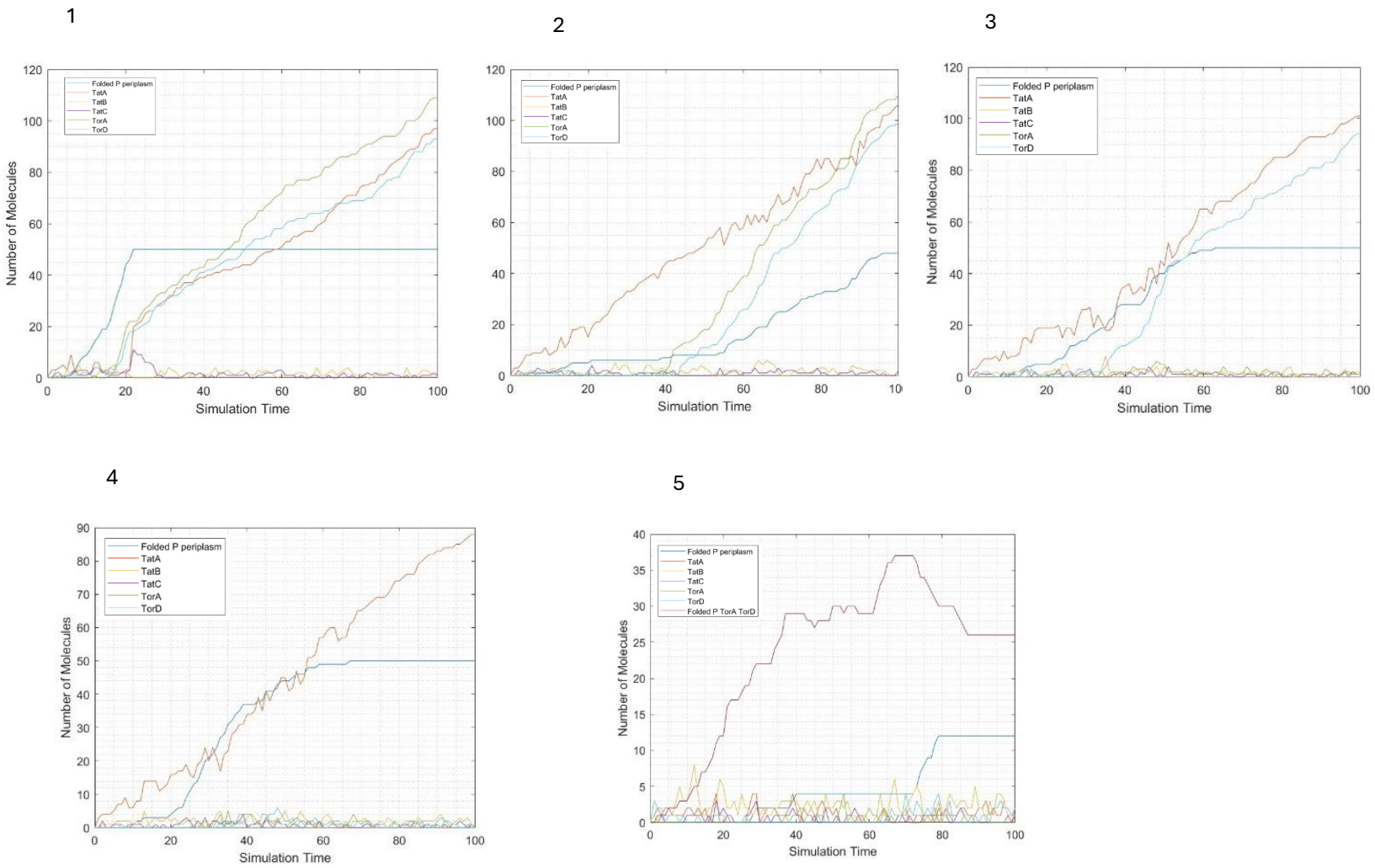


Figure 4. 25 Tat Translocon Component Limitations on Protein Translocation Efficiency. Graph 1: Model outputs under default conditions are the comparative analysis baseline. Graph 2: Impact of reducing the recycling rate of TatB on the model's output. Graph 3: Effect of reducing the recycling rate of TatC on the model's behaviour. Graph 4: Influence of reduced availability of TorA on the system's performance. Graph 5: Consequences of reducing the availability of TorD on the model's dynamics. Graph 6: Impact of decreasing the recycling rate of TatA on the overall system.

This test evaluates the Tat pathways by sequentially reducing the recycling ability of each component to examine the effects on translocation. Figure 4.25 illustrates how reducing the recycling rates of the TatA, TatB, and TatC chaperones affects the functionality of TorA and TorD. We observe translocation differences as each chaperone's functionality decreases. Previous simulation analysis focuses on the amount of protein produced. This data set will focus on the time taken for varying rates of recycling Tat to use for the next unit of protein.

Graph 1 depicts the effects of TatB's slower recycling, showing a steep translocation rate curve that flattens at 20 simulation time units. In contrast, Graph 2 presents the impact of reduced TatC recycling, resulting in a gentler curve that flattens at 100 simulation time units, indicating an effect on the transport of folded proteins to the periplasm.

Graph 3 illustrates a 20% reduction in TorA, leading to translocation in 52 simulation time units notably faster than Graph 2 but slower than Graph 1. This suggests that TorA influences translocation less than TatC but more than TatB. Graph 4 demonstrates a similar 20% reduction in TorD, resulting in slower translocation, with a flatline at 64 simulation time units but still a lesser impact than TatB.

When TatA is reduced, no folded proteins reach the membrane, accumulating protein in the cytoplasm. This study underscores TatA's essential role in efficient translocation within the Tat system.

4.5.2 Evaluating the model performance by testing different translocation rates

A study focusing on Tat's translation rate will be evaluated using the model to perform increased and reduced translocation rates. This will include transport from the cytoplasm to the membrane, the rate of protein recognition, and the rate of forming complexes between chaperones. The recycling of chaperones will be set at default to evaluate the impact of only varying rates of TAT translocation.

Three variables will be tested at varying rates, tabulated below. The input to the mode is the rates of the Tat chaperones, which are changing. The output is the amount of folded protein transported to the periplasm. The number of molecules for the transported protein will be used to understand the impact. A comment section is also present if the molecule's generation differs for the transfer rates. The purpose will be to understand the effect of rates; therefore, a set molecule number will not be used for this test; instead, an abundance of protein is available.

Table 4. 1 Mass action rate for TorD and amount of protein transported to the periplasm

Mass action (AU)	Amount of folded protein (AU)	Comments
0.0001	24	As expected, the general pattern is prolonged and does not form a smooth curve.
0.01	85	The curve is not smooth but has a pattern similar to graph 1 in Figure 4.25 above.
1	110	The graph is smoother
10	117	Similar to the previous value, it suggests other limitations could exist.
100	98	This shows that increasing the rate while all the other rates are the same leads to lower amounts of transported protein.

Table 4. 2 Rate of folded protein production and the impact on protein transport.

Mass action (AU)	Amount of folded protein (AU)	Comments
0.0001	0	No protein transport curve
0.01	0	No protein transport curve
1	87	Smooth curve demonstrating transportation of protein
10	970	It's a smooth curve, indicating that the amount of exported

		protein will increase if the protein production rate increases.
100	1355	This is a smooth exponential graph. This is the first indication using the model that the protein production rate impacts the amount that can be transported. It is also worth mentioning that 800 times more protein is within the cytoplasm than in the periplasm, which shows the transport limitations at very high rates.

The results shown in Table 2 are exciting but must be investigated further. One way to approach this is to use the same rates for the Tat chaperones, such as Tat A. In the investigation above, Tat A impacts the transport pathways more than the other chaperones. However, when the rate increases to 100, as above, I get thousands of folded protein molecules transported. Table 3 will explore the rate variation from 0.0001 to 100 for Tat A and note the amount of protein transported.

Table 4. 3 Testing a range of mass action rates for Tat A

Mass action (AU)	Amount of folded protein AU)	Comments
0.0001	0	No protein transport
0.01	0	No protein transport
1	99	A steady accumulation of molecules indicates that protein transportation is smooth at this rate.
10	104	A similar trend to the above mass action at 1 AU

100	95	There is a slight decrease as the rate increases. This shows that TatA is not the limiting factor; protein generation impacts the amount transported.

The final table will not test a viable method like the tables above but will vary the rate of the final step of Tat translocation. This will also confirm if protein production is the limiting factor because the rate is exceptionally high. However, the same quantity is transported; it depends on how many units the simulation can handle regardless of increased rates.

Table 4. 4 Range of rates of mass action for the final transport of folded protein

Mass action (AU)	Amount of folded protein (AU)	Comments
0.0001	0	No protein transport curve
0.01	17	A step pattern was formed, demonstrating that transportation is possible at this rate but not smoothly.
1	94	A steady and smooth curve similar to other curves when the rate is at 1
10	91	These results are very similar to the above rate in this data set. This again indicates that protein transport is limited to the amount present in the cytoplasm.
100	97	Very smooth and steady transposition of protein. Similar to 1 and 10 for this dataset.

The results presented in this section conclude that the translocation rate does not impact the amount transported after a threshold. However, if the protein production rate increases, the amount of protein transported increases by tenfold. The amount of protein left in the cytoplasm highlights the inefficiency of increased rates. For high rates, such as 100 mass transfer for protein folding, the amount of protein left in the cytoplasm was 800-fold the amount left when operating at a rate of 1. Although more protein is transported, the system could be more efficient, and the balance between the chaperones and protein production must be maintained to ensure protein does not remain within the cytoplasm.

The data presented in this section shows how this model is used to understand where the protein is being held up and improve the relationship between chaperones and protein to increase transport efficiency in *E. coli*. The model can provide new information on the impact of varying protein production rates on the transport ability of *E. coli*, which will be helpful when designing a vector for a recombinant protein that requires Tat translocation. A summary of the plots is shown in Appendix B.

4.6 Disulphide bond formation

Disulphide bond formation presents a challenge in recombinant protein expression in *E. coli* (Berkmen, 2012). Addressing this bottleneck necessitates integrating protein synthesis and transport pathways with disulphide bond formation pathways to elucidate their interactions and identify critical parameters affecting disulphide bond formation efficiency (de Marco, 2009; Lee & Davey, 2017).

Unfolded native proteins enter the periplasm, where DsbA facilitates the incorporation of two electrons from the protein. Subsequently, DsbA must be regenerated to resume its function, ensuring efficient disulphide bond formation (Kurokawa et al., 2000). In cases where the turnover rate of DsbA is slower than the transport rate, mis-oxidised native proteins may result in a condition corrected by DsbC, which also requires reduction for electron removal.

DsbB plays a crucial role in maintaining electron flow by accepting electrons from DsbA and transferring them to ubiquinone. DsbD complements this process by acquiring electrons from the electron transport chain's complex one, subsequently channelling them to DsbC to correct mis-oxidations (Kurokawa et al., 2000).

Similar processes form disulphide bonds in recombinant proteins, albeit with higher mis-oxidation rates than native proteins. This heightened complexity arises because *E. coli* faces more challenges in processing recombinant proteins for disulphide bond formation (Chung et al., 2000; Krupp et al., 2001; Stirnimann et al., 2005).

Figure 4.26 illustrates the early model's focus on electron movement, which is represented by Dsb proteins, which are crucial for disulphide bond formation proteins that require both reduction and oxidation processes to form bonds. The model initially included conditions for oxidation within the periplasm, depicted with sink and source directions. However, these conceptual representations lacked mathematical rigour and were later omitted in the refined model to ensure biological fidelity.

Furthermore, Figure 4.26 did not account for the competition from native proteins, such as DsbA and DsbB, which play essential roles in disulphide bond formation (Adams et al., 2015; Guilhot et al., 1995; Inaba, 2008; Inaba et al., 2006). These factors have since been integrated into the updated model to comprehensively represent disulphide bond formation dynamics.

Figure 4.27 shows the improved and updated Petri net model, which presents the pathways to native and recombinant proteins. The redox potential must be maintained within the periplasm, which is achieved by the electron transport chain. Although ATP is not directly required for disulphide bond formation, the rate at which electrons are transferred from oxidative phosphorylation impacts the recycling rate of DsbA, DsbB, DsbC, and DsbD.

This section aims to deepen the understanding of how disulphide bond formation pathways interact with protein synthesis and transport processes, shedding light on critical factors influencing protein folding and functionality in recombinant systems.

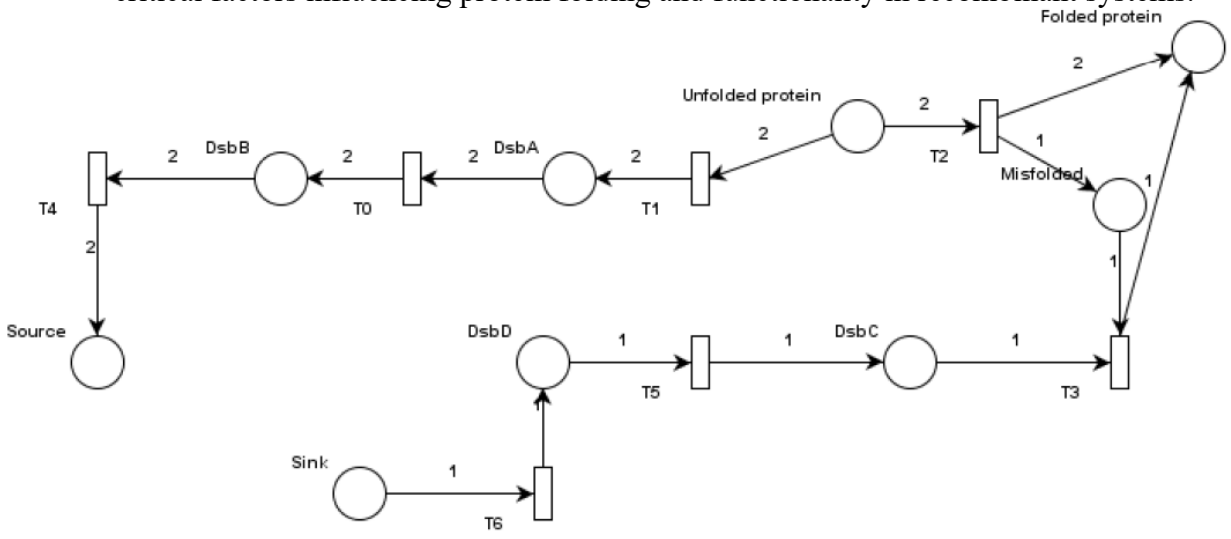


Figure 4. 26 Initial model developed for disulphide bond formation in *E. coli* where electrons are taken from an electron sink and retuned to electron source. The nomenclature for disulphide bond formation proteins is the same as found in the literature; some uniquely named places include Unfolded_Pro_N, which is an unfolded native protein, Unfolded_Pro_R is unfolded recombinant protein, misoxidised_N is a mix-oxidised native protein, misoxidised_R is mis-oxidised recombinant protein, Folded_N is folded native protein and finally Folded_R is folded recombinant protein.

The model illustrates that DsbA and DsbB function together initially to form disulphide bonds, with DsbC and DsbD coming into play to correct mis-oxidised proteins. These distinct roles will be explored further in sections 4.6.2 and 4.6.3.

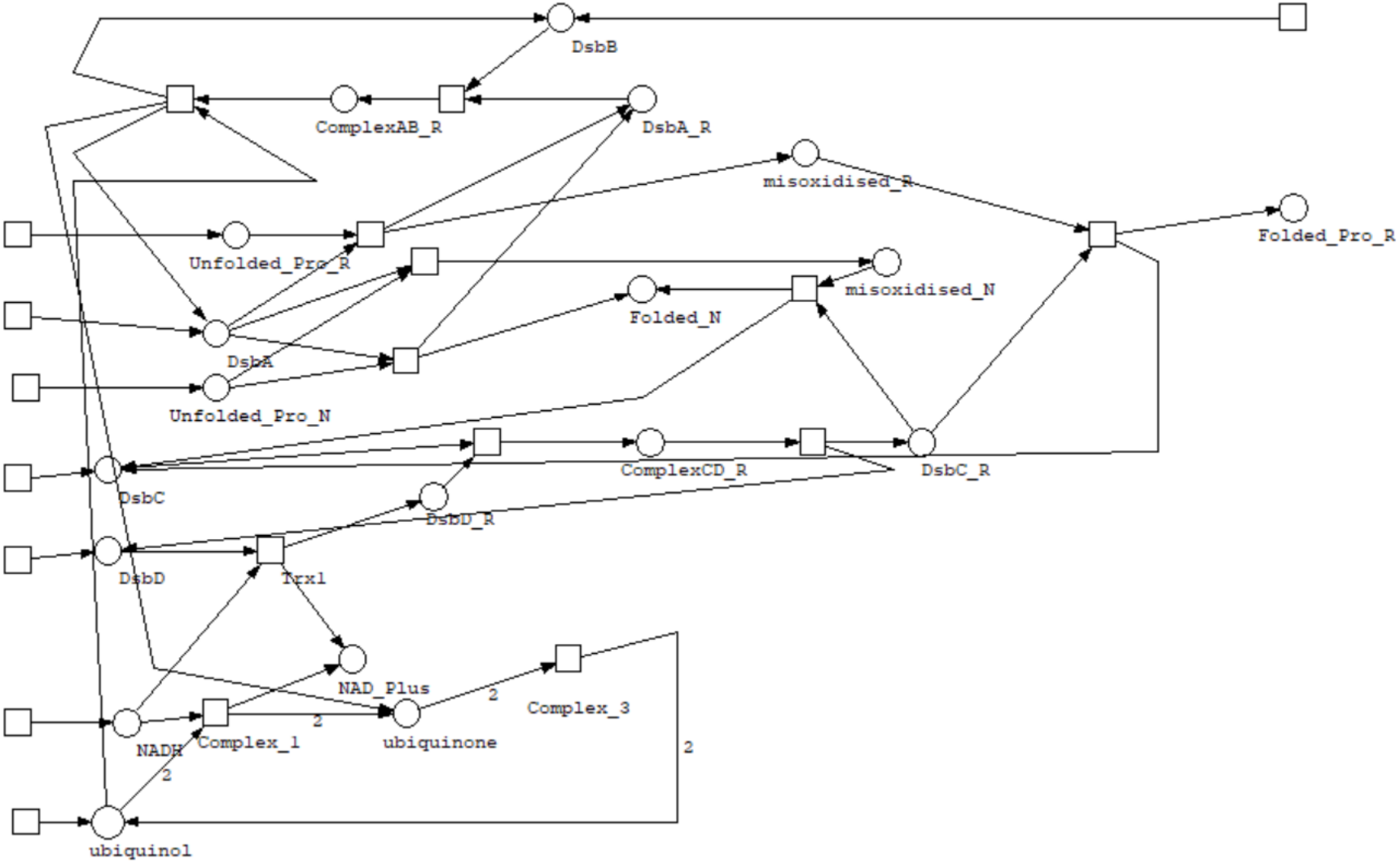


Figure 4. 27 Disulphide bond formation pathway in *E. coli* modelled using Petri net formalism.

4.6.1 Modelling the role of DsbA-DsbB catalyst in forming disulphide bonds

The constructed model provides insights into the role of disulphide bond formation. DsbA and DsbB influence protein production. This section highlights the necessity of reducing the turnover rates of these catalysts to assess their impact on the quantity of proteins synthesised. According to Figure 4.28, the amount of protein produced under default model conditions without any modifications to the transition rates serves as a baseline for our analysis.

Graph 1 illustrates the limitations evident in recombinant protein production. Reduced efficiency is likely attributed to initial competition for DsbA. In addition, recombinant proteins are prone to forming mis-oxidised products, suggesting that mechanisms aimed at mitigating mis-oxidation are less effective for recombinant proteins than their native counterparts.

Graph 2 presents the effects of implementing restrictions to monitor the accumulation of molecules within the model. Despite introducing these restrictions, the general trend persists, exhibiting higher native protein production relative to recombinant proteins. However, a notable observation in Graph 2 is the narrowing gap between folded native proteins and mis-oxidised native proteins. This shift may stem from the system's reallocation of resources, as it no longer prioritises the generation of new molecules.

Consequently, material flows through pathways speeds up, resulting in a more balanced count of folded and mis-oxidised native proteins. Conversely, the gap between folded and mis-oxidized proteins remains constant, demonstrating the persistent challenges associated with recombinant protein folding and oxidation.

The model underscores the pivotal roles of DsbA and DsbB in enhancing the efficiency of protein production, particularly regarding the balance between folding and mis-oxidation of native and recombinant proteins. The observed patterns underscore the significance of catalyst availability and turnover rates in protein production pathways. These insights pave the way for further research into strategies to increase recombinant protein yields and minimise mis-oxidation.

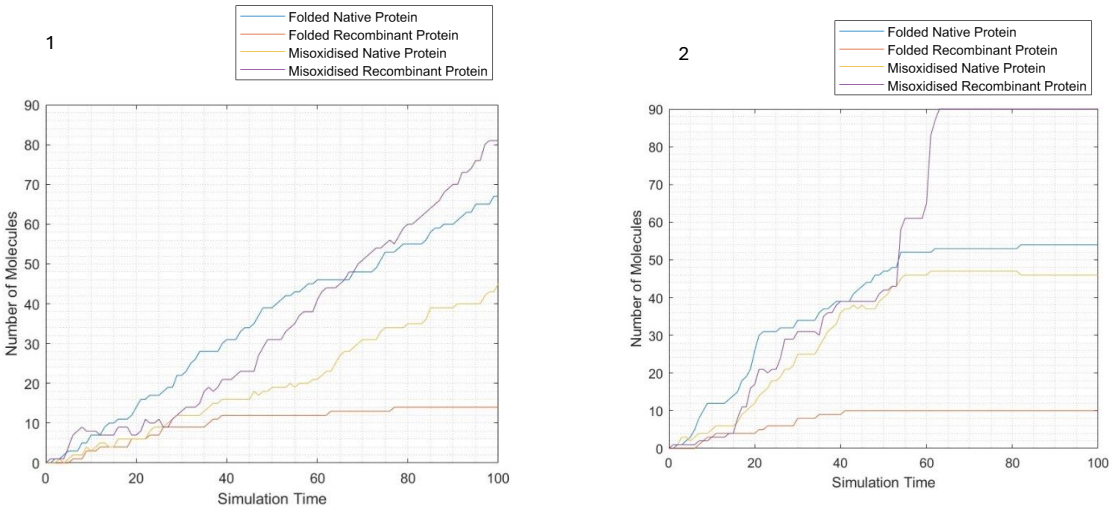


Figure 4. 28 Molecule accumulation profiles for native and recombinant protein production in the disulphide bond formation pathways of *E. coli*. The analysis includes correctly oxidised and mis-oxidised forms of native and recombinant proteins to identify specific points of molecule accumulation within the pathways.

Figure 4.29 explores the roles of DsbA and DsbB in protein production. Increasing DsbA and DsbB recycling rates does not change the accumulation of native or mis-oxidised recombinant proteins in the model, mirroring wild-type *E. coli* pathways that are unsuitable for recombinant protein production. Consequently, recombinant protein yields remain relatively low. Notably, native protein production decreases by about 10% without chaperone recycling, underscoring the importance of recycling for protein folding efficiency.

These findings establish a foundational understanding of pathway interactions affecting protein production. Further investigations with varying rates and conditions will identify specific limiting factors and improve pathways for higher recombinant protein yields.

Gleiter and Bardwell (2008) examined DsbA and DsbB's roles in disulphide bond formation in *E. coli*. They identified DsbA as the primary oxidase introducing disulphide bonds, while DsbB re-oxidises DsbA to sustain its activity. Their study indicated that dysfunction in DsbA or DsbB hinders protein folding and increases mis-oxidised protein accumulation, consistent with our model's patterns.

Similarly, Kadokura et al. (2003) investigated the interplay between DsbA and DsbB in recombinant protein production. They found that proteins with complex disulphide patterns are more susceptible to mis-oxidation due to inefficient folding pathways in wild-type *E. coli*. Their findings support the simulation results of inherent limitations in recombinant protein production within non-engineered systems.

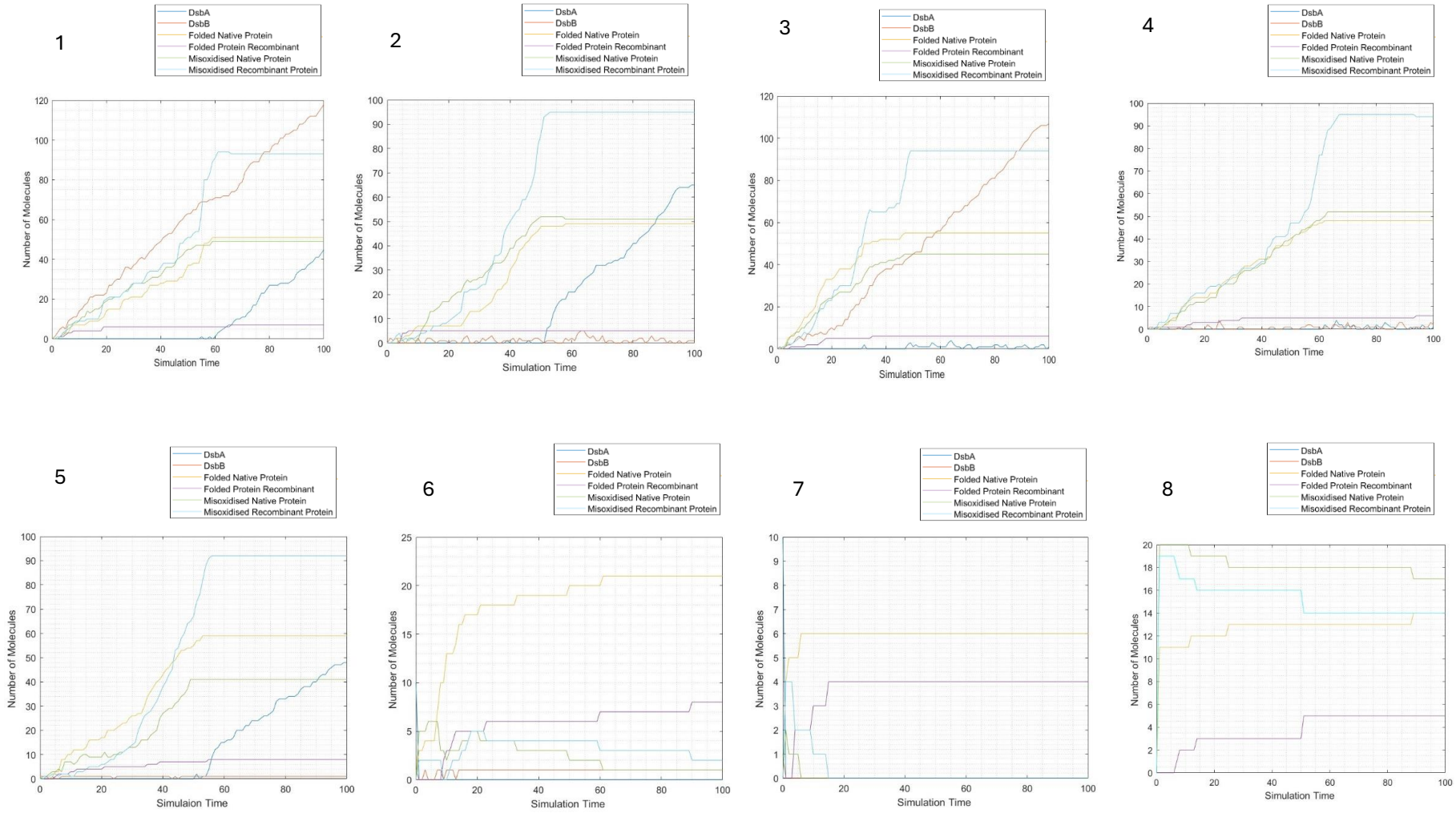
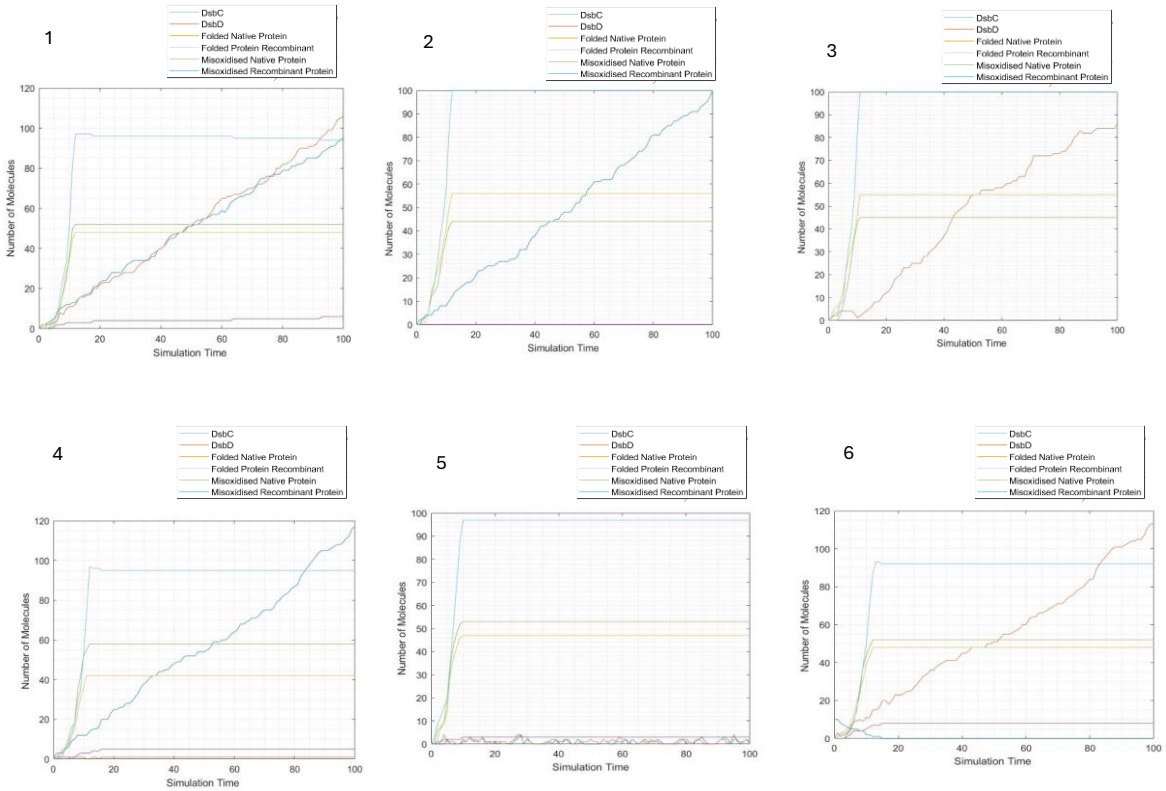


Figure 4. 29 Impact of DsbA and DsbB Recycling Rates and Quantities on Molecule Accumulation in the Disulphide Bond Formation Pathway of E. coli. Graph 1: Model output with DsbA and DsbB recycled at default rates. Graph 2: Impact of increasing the recycling rate of DsbB Graph 3: Effect of increasing the recycling rate of DsbA Graph 4: Combined increase in DsbA and DsbB recycling rates, showing no molecule accumulation. Graph 5: Operation of a single unit of DsbB. Graph 6: Operation of ten units of DsbA. Graph 7: Operation of ten units of DsbA and one unit of DsbB without recycling. Graph 8: Model behaviour with DsbA set to fifty molecules and DsbB deleted.

4.6.2 Modelling the role of DsbC-DsbD catalyst to correct mis-oxidation of proteins

DsbC and DsbD are essential catalysts influencing protein production in the model. They play a key role in minimising mis-oxidised proteins by promoting disulphide bond isomerisation. When both chaperones are actively recycled, the model experiences low recombinant protein production, accumulating 100 molecules of mis-oxidised protein. Notably, the amounts of native and mis-oxidised native proteins are nearly equal under these conditions.

No folded recombinant protein is produced in the absence of DsbC and DsbD, consistent with expectations. However, recombinant protein production increases to ten molecules when DsbD is present. While this highlights their relationship, it still does not offer insights into enhancing recombinant protein yield, which remains a vital goal of the model.



*F*igure 4. 30 Impact of DsbC and DsbD Catalysts on Protein Production in the Disulphide Bond Formation Pathway. Graph 1: DsbC and DsbD recycled at their default rates. Graph 2: DsbD deleted while DsbC is present in abundance. Graph 3: DsbC was deleted while DsbD was present in abundance. Graph 4: One functional channel of DsbD operating alongside an abundance of DsbC. Graph 5: Both DsbC and DsbD are limited in quantity. Graph 6: Ten molecules of DsbC and one functional channel of DsbD.

The interplay between DsbC and DsbD catalysts in protein production is explored through various scenarios depicted in Figure 4.30. These scenarios provide insights into how variations in the availability and functionality of DsbC and DsbD influence protein folding and disulphide bond formation.

Graph 1 represents the baseline scenario, where DsbC and DsbD are recycled at their default rates, as reported in the literature. This is a reference point for comparing the effects of altering their availability

Graph 2 illustrates the impact of deleting DsbD while maintaining an abundance of DsbC. The results show a reduction in folded protein production, highlighting the critical role of DsbD in supporting DsbC's function.

Graph 3 demonstrates the effect of deleting DsbC while maintaining an abundance of DsbD. This scenario also results in reduced protein production, underscoring the importance of DsbC in correcting mis-oxidised proteins and ensuring proper folding.

Graph 4 shows the model with one functional channel of DsbD and an abundance of DsbC. This scenario reveals that even limited DsbD functionality can support protein production when DsbC is abundant, suggesting a cooperative relationship between the two catalysts.

Graph 5 explores the scenario where both DsbC and DsbD are limited. The results demonstrate a decline in protein production, emphasising their essential and interdependent roles in disulphide bond formation and protein folding.

Graph 6 examines the scenario with ten molecules of DsbC and one functional channel of DsbD. This configuration shows improved protein production compared to scenarios with limited DsbC or DsbD, indicating that a balance between the two catalysts is crucial for optimal yields.

The results highlight the cooperative roles of DsbC and DsbD in ensuring proper disulphide bond formation and protein folding. While DsbD impacts the amount of folded protein, a balanced availability of both catalysts is essential for achieving higher yields. These findings align with studies by Rietsch et al. (1997) and Katzen and Beckwith (2000), which demonstrated the interdependence of DsbC and DsbD in maintaining redox homeostasis and supporting protein folding in *E. coli*.

4.6.3 Evaluating model performance by varying the rates of disulphide bond formation proteins

The collected data offers insights into how disulphide bond formation proteins interact with native and recombinant proteins. However, this investigation did not examine the impact of changes in transition rates on the process. Figure 4.31 illustrates the alterations in the rates of DsbA-DsbB catalysis, while Figure 4.32 similarly depicts the changes in

the rates of catalysis for DsbC and DsbD. The simulation may be able to give better insights into the role of these catalysts.

Figure 4.31 shows DsbA across varying transition rates; mis-oxidised recombinant protein levels remained constant, while folded native and mis-oxidised native protein quantities were nearly identical. A slight increase in folded recombinant protein was observed at a mass transfer rate of 100 arbitrary units (AU), suggesting that higher DsbA activity may marginally improve recombinant protein yields.

DsbB shows similar trends, with no noticeable variation in protein production across transition rates. However, at lower rates (Graphs 5 and 6), a small accumulation of folded recombinant protein was noted, indicating that slower rates may enhance folding efficiency. This was observed in other simulation parts and discussed in section 4.2.

Figure 4.32 presents DsbC with rates similar to those of Figure 4.31. No noticeable differences were observed across the tested transition rates, suggesting that DsbC's role in isomerising mis-oxidised proteins is less sensitive to rate variations under the modelled conditions.

DsbD presented the most compelling finding in Graph 8, where increasing the rate to 100 AU resulted in a 100-fold increase in recombinant protein production. This dramatic improvement underscores the critical role of DsbD in electron transfer, which is essential for repairing mis-oxidised proteins and supporting efficient protein folding. Concurrently, native protein production also increased at this rate, highlighting DsbD's broader impact on protein homeostasis.

The findings of this model are consistent with established literature. However, locating a study that explicitly investigates these rates proves to be challenging. Consequently, general trends are linked to the simulation and compared with studies concerning Dsb proteins. For example, research conducted by Gleiter and Bardwell (2008) and Kadokura et al. (2003) has shown that DsbA and DsbB are crucial for forming disulphide bonds. In contrast, DsbC and DsbD are vital for correcting mis-oxidized proteins and maintaining redox balance. The observation made by the model that DsbD enhances recombinant protein production at elevated rates agrees with experimental evidence, which indicates that efficient electron transfer is essential for better protein folding (Rietsch et al., 1997).

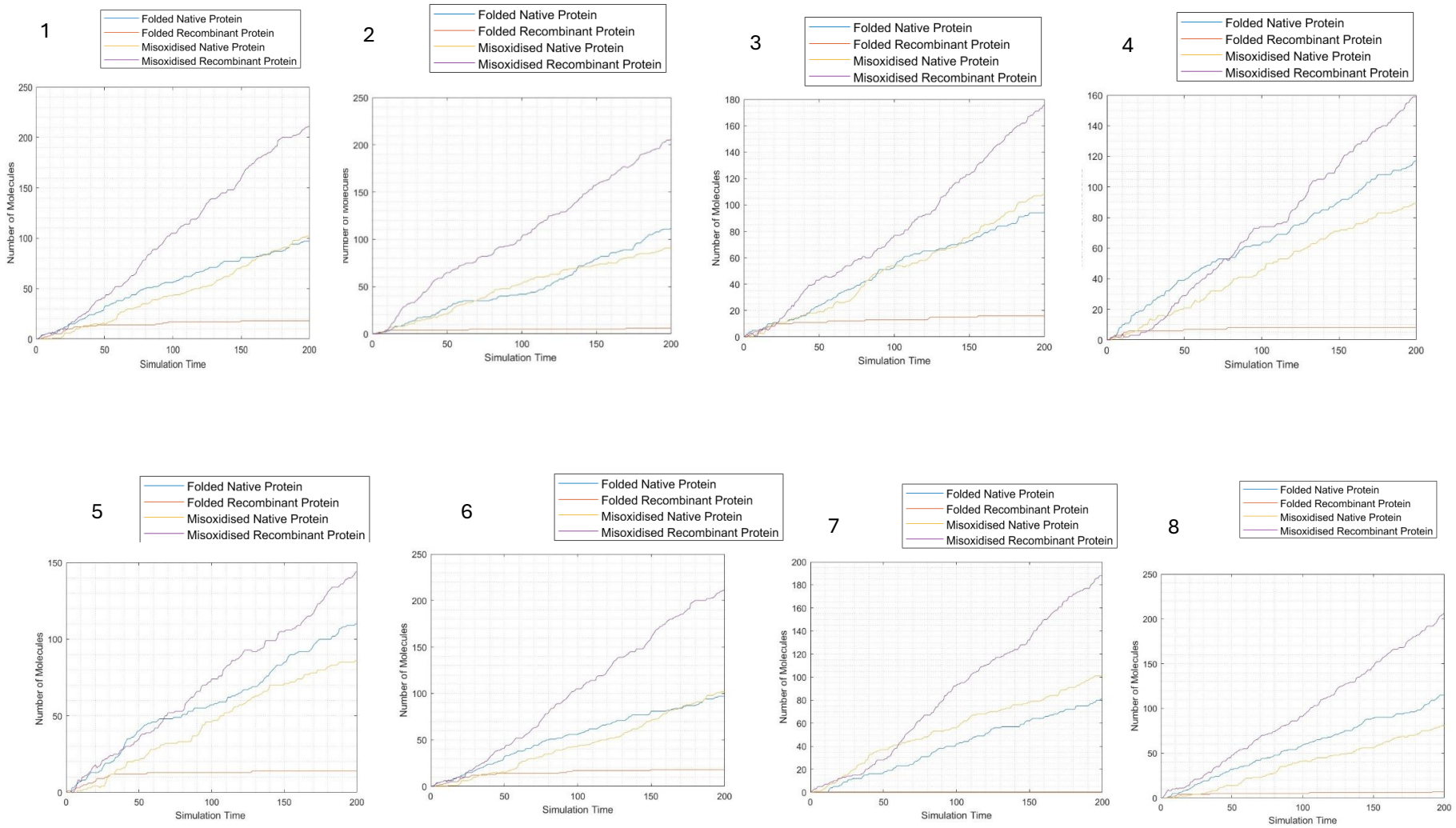


Figure 4. 31 Exploration of Mass Transfer Rates for DsbA and DsbB in the Disulphide Bond Formation Pathway. The mass transfer rates for DsbA and DsbB are systematically varied across eight graphs: Graphs 1–4: Mass transfer rates for DsbA at 0.0001 AU, 0.01 AU, 1 AU, and 100 AU, respectively. Graphs 5–8: Mass transfer rates for DsbB at 0.0001 AU, 0.01 AU, 1 AU, and 100 AU, respectively.

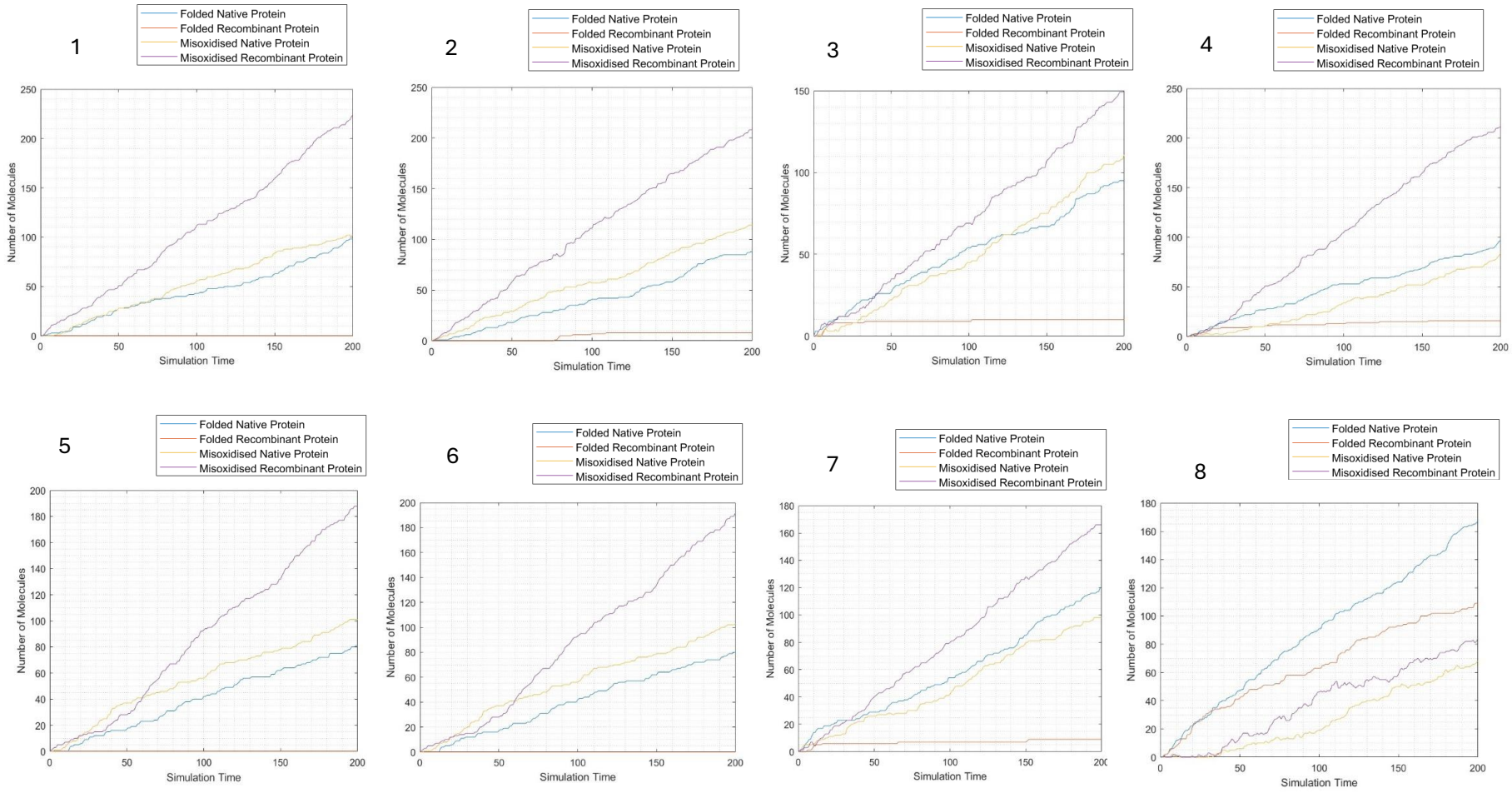


Figure 4. 32 Exploration of Mass Transfer Rates for DsbC in the Disulphide Bond Formation Pathway. The mass transfer rates for DsbC are systematically varied across eight graphs: Graphs 1–4: Mass transfer rates for DsbC at 0.0001 AU, 0.01 AU, 1 AU, and 100 AU, respectively. Graphs 5–8: Mass transfer rates for DsbC at 0.0001 AU, 0.01 AU, 1 AU, and 100 AU, respectively.

4.7 Summary

The work presented in this chapter shows how Petri net models were developed to simulate key processes in protein synthesis, translocation, and disulphide bond formation in *E. coli*. The initial models provided a foundation, with subsequent iterations refining and expanding upon them to create a comprehensive working model. By breaking down the model into smaller segments, new insights into how each pathway impacts the overall protein production were formally explored.

The data presented in the chapter led to specific conclusions that are valuable for decision-making in the design of recombinant protein expression in *E. coli*. Notably, the model highlighted the role of chaperones during protein synthesis, particularly emphasising the effectiveness of DnaK in increasing the yield of correctly folded recombinant proteins when overexpressed. In contrast, when overexpressed, chaperones like ClpB did not show a substantial increase in folded protein yield.

The analysis showed that while the rate of protein translocation impacts the amount of protein transported to some extent, beyond a certain threshold, the protein concentration in the cytoplasm becomes the limiting factor. Increasing the protein production rate resulted in a ten-fold increase in transported proteins. However, proteins remained in the cytoplasm, underscoring the need for a balanced approach between chaperone availability and protein production to ensure efficient transport.

Furthermore, investigating the disulphide bond formation pathways provided insights into balancing redox potential within the periplasm, which is crucial for proper disulphide bond formation. Notably, an increase in the rate of DsbD was found to enhance the amount of folded recombinant proteins, highlighting the importance of redox balance in these processes.

Overall, the Petri net model proved effective in assessing the performance of protein machinery in *E. coli*. It successfully identified critical chaperones for protein folding and simulated trends observed in existing literature, thereby enhancing understanding and potentially guiding future research and industrial applications in protein production.

Chapter 5

5. Application and testing

After developing a working model, it is essential to investigate test cases to evaluate its utility. Test Case-Based Testing (TCBT) is the most common approach to conducting this analysis. In this approach, tests are planned and executed, and the outcomes of these tests are compared to expected results. TCBT consists of five steps (Reddy et al., 2012).

Step 1: Planning the test, where the scope is defined for testing.

Step 2: Monitoring the test results, including tracking variables that have changed in the model and understanding the reasons behind these changes.

Step 3: Defining the criteria by which the test is considered complete. For instance, if the energy requirement is limited and protein production has reached a plateau, the test will be concluded at this point.

Step 4: Evaluate whether the model's results align with expected trends from the literature. If they do, the test has passed; if not, it has failed.

Step 5: Upon the test passes, the results are logged, and the actual and predicted outcomes are compared. If the test fails, the reason for failure is documented.

The test cases presented in the chapter can be effectively analysed and explored quantitatively using Petri nets. Below is an explanation of how each test case can be modelled and analysed with Petri nets, emphasising the quantitative insights that can be gained. Each test case is bench-matched with data from the literature, which was not used to construct the model. This will highlight the strengths and weaknesses of the model when compared to literature data.

Test Case 1 examines how the hydrophobicity of signal peptides affects the selection of translocation pathways. The objective is to explore the influence of hydrophobicity on choosing between coupled and uncoupled translocation pathways. In the Petri net structure, places represent various protein states (such as unfolded protein, protein with a hydrophobic signal peptide, and protein with a hydrophilic signal peptide). At the same time, transitions denote translocation pathways, with rates dependent on the signal peptide's hydrophobicity. Tokens symbolise the quantity of protein involved.

The quantitative analysis within the Petri net framework involves simulating the model at varying hydrophobicity levels and tracking the movement of tokens in each pathway. It calculates the proportion of proteins translocated via Sec based on hydrophobicity, identifies the threshold for hydrophobicity, determines the selection of pathways, and assesses its effect on translocation efficiency.

Test Case 2 explores the formation of disulphide bonds in the cytoplasm and its impact on redox balance. This case aims to assess the ability of proteins to form these bonds

and evaluate their effect on redox balance. The study employs Petri net modelling, where places represent various molecular species such as unfolded proteins, oxidised proteins, reduced cytoplasm, and oxidised cytoplasm. Transitions signify disulphide bond formation, redox reactions, and protein translocation, while tokens symbolise proteins. Quantitative analysis involves simulating the model to monitor the accumulation of oxidised proteins and the shifts in redox balance. Additionally, the impact of Tat translocation on redox balance is measured by comparing token distributions in scenarios with and without the Tat pathway activity. Overall, this analysis seeks to quantify how disulphide bond formation in the cytoplasm affects redox balance and to identify conditions that may mitigate any negative consequences.

Test Case 3 investigates the redox potential in the periplasm, mainly focusing on the effects of dysfunctional reducing agents such as DsbC and DsbD. This study aims to understand how these non-functional agents affect the redox potential and contribute to the accumulation of mis-oxidised proteins in the periplasm.

Petri net implementation places represent various molecular species, including mis-oxidised proteins, functional and dysfunctional forms of DsbC and DsbD, and the oxidised periplasm. Transitions within this model illustrate key processes such as oxidation, reduction, and protein folding, while tokens symbolise the proteins and redox agents involved.

For quantitative analysis, the model is simulated under varying degrees of DsbC and DsbD functionality, measuring the accumulation of mis-oxidised proteins and any resultant changes in the periplasmic redox potential. This analysis will calculate the efficiency of protein folding and the formation of disulphide bonds across different conditions. Ultimately, this research aims to quantify the detrimental impact of reducing agent dysfunction on periplasmic redox balance and the overall quality of proteins.

In Test Case 4, the focus is on assessing the impact of decreased functionality of DsbA and DsbC on the disulphide bond formation process. The study employs a Petri net model where places symbolise various molecular species, including unfolded, oxidised, and functional and dysfunctional forms of DsbA and DsbC. The transitions in this model illustrate the processes of disulphide bond creation, oxidation, and reduction. Tokens within the model stand for proteins. To conduct a quantitative analysis, the model will be simulated under different scenarios of DsbA and DsbC functionality to evaluate the rate at which disulphide bonds form and the rate at which misfolded proteins accumulate. Additionally, the analysis will involve calculating the ratio of correctly folded proteins in varying states of functionality. The ultimate goal is to quantify the connection between the functionality of DsbA and DsbC and the efficiency of disulphide bond formation.

Test Case 5 investigates how increasing the levels of molecular chaperones affects the yield and quality of proteins. The model simulates the interactions between various molecular species, including unfolded, folded, and misfolded proteins and chaperones. In this simulation, protein folding, misfolding, and chaperone rates are represented as transitions, while the amounts of each protein type are depicted as tokens. Simulations with different levels of chaperone expression are run, and how this influences the ratio of correctly folded proteins to misfolded ones is observed. The goal is to identify the

best level of chaperone overexpression that maximises protein yield and quality, providing insights into the optimal conditions for protein production.

Understanding the distinction between a test case and a scenario was essential when searching for challenges to test the model. A test case was a detailed investigation that collected data to verify that the model operated correctly. It included the steps mentioned above and specified the test's priority level. Meanwhile, a scenario involves modifying a single variable to observe its impact on the overall model Da Silva Michael (2022). Chapter 4 provides examples of scenario analysis, where the model is deconstructed and analysed a parameter at a time for model validation.

The model protein used is human growth hormone (HGH) because it is one of the most documented recombinant proteins. This allows one to compare the model results with expected results from the literature. However, data availability is not even across the test cases, and the impact of these limitations is highlighted in each test case as necessary.

The translation process is also a limiting factor within the model, which must be accounted for. Petri net models work as a model of the flow of information, and assuming a large availability of translated peptides may lead to inaccurate interpretations if not evaluated appropriately. To avoid such discrepancies, the translation process is represented in this model by adding each amino acid into a peptide sequence after another, considering the *E. coli* specific codon availability.

In *E. coli*, translation relies on the availability of tRNAs and codons. Codons are triplet sequences in mRNA that specify amino acids—the genetic code degenerates, with multiple codons encoding the same amino acid. tRNAs are small RNA molecules with anticodons that pair with codons and are attached with corresponding amino acids by aminoacyl-tRNA synthetases (Kohno et al., 2024). *E. coli* has about 86 tRNA genes, reflecting codon usage bias and ensuring translational efficiency and fidelity(Samant et al., 2014).

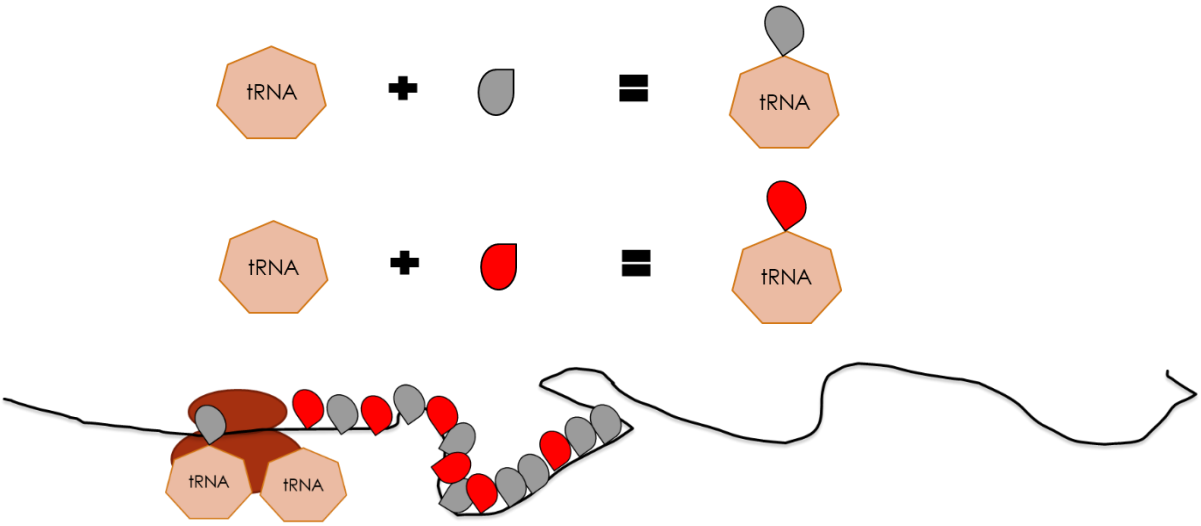


Figure 5. 1 Diagram Portraying Protein Translation in *E. coli* with Varying Codons. This figure illustrates the process of protein translation in *E. coli*, highlighting the impact of varying codons on translation efficiency and protein yield.

The number of codons and the composition were also considered when simulating this model expressing the recombinant protein of interest. The human growth hormone is 191 amino acids long and is a relevantly small protein consisting of four disulphide bonds (Roytrakul et al., 2001).

Figure 5. 2 represents the Petri net model, including the translation of HGH integrated with the protein synthesis pathways, translation pathways and disulphide bond formation, which require energy in the form of ATP from the TCA cycle and oxidative phosphorylation. This model will form the basis of the test cases, and modifications will be made according to the test case's requirements.

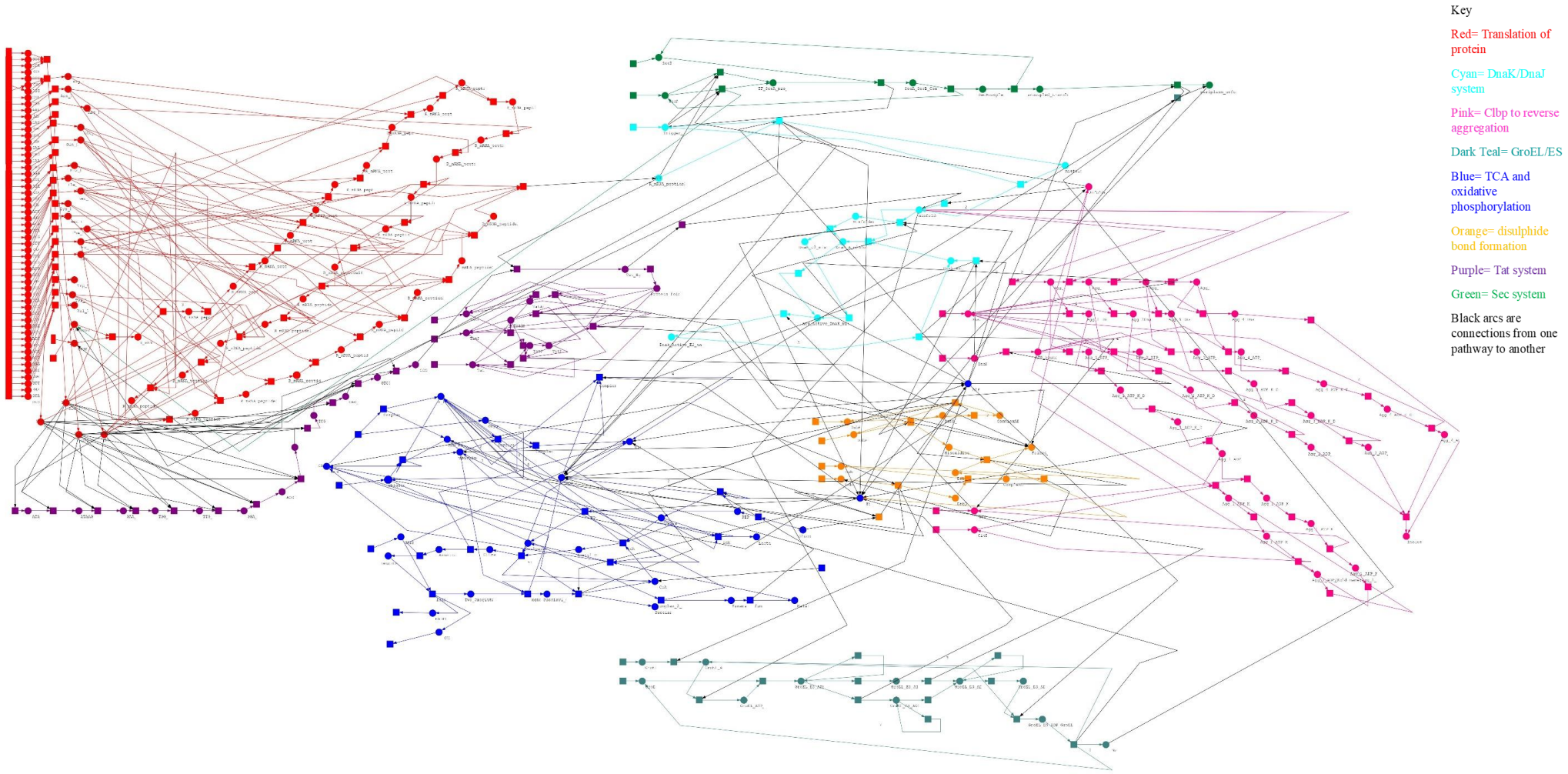


Figure 5. 2 Petri net Model of Human Growth Hormone (HGH) Translation Integrated with Protein Folding, Misfolding, and Aggregation Pathways. The model captures the entire production of HGH in a microbial expression system, such as *E. coli*, from mRNA translation to the formation of functional or misfolded proteins. Due to the size of the image, the resolution of this image may be reduced. To view the image in full resolution, please click here.

5.1 Test Case 1: Impact of hydrophobicity of Sec signal peptide on the total number of proteins produced

5.1.1 Introduction to Test Case 1

The ability to transport a protein to the correct location is influenced by the overall hydrophobicity of a single peptide (Zanen et al., 2005). This understanding has practical implications, as it reveals two possible routes for Sec translocation: via signal recognition proteins (SRP) for highly hydrophobic signal peptides and chaperone-assisted activity with the consumption of ATP as an energy source to allow for translocation. Signal peptides play an essential role in the early stages of protein folding. Their influence on the solubility of the target protein renders signal peptides a crucial aspect of expression system decisions (Singh et al., 2013).

A signal peptide comprises the first 20-25 amino acids in a peptide chain; the combination of amino acids impacts the characteristics of the signal peptide. Depending on the hydrophobicity of these initial amino acids, the route taken by the protein can change (Brockmeier et al., 2006). For highly hydrophobic protein chains, protein channels such as SecYEG are utilised for translocation to the membrane without needing ATP. However, if the first 20 amino acids have low hydrophobicity, ATP is required to facilitate the translocation of the protein (Papanikou et al., 2007).

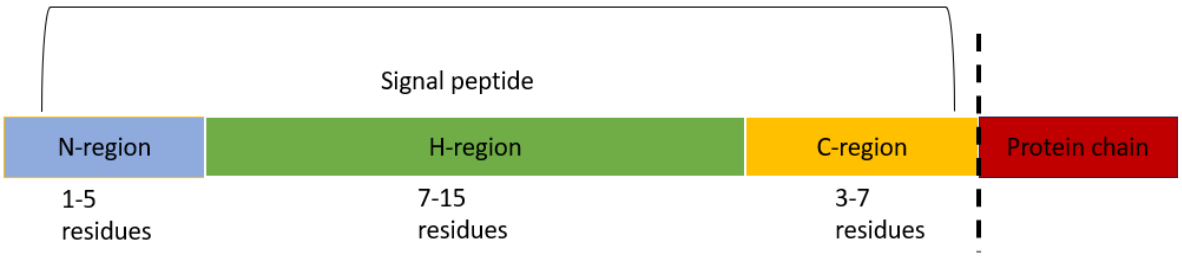


Figure 5. 3 Impact of Amino Acid Regions in Signal Peptides on Hydrophobicity and Cellular Translocation Pathways. This figure illustrates how specific regions of amino acids within signal peptides influence their hydrophobicity and determine the translocation pathway they follow within the cell. Signal peptides are critical for directing proteins to their correct cellular compartments, and their hydrophobicity plays a key role in pathway selection,

The N region (called n terminal in some instances) determines the net charge of the peptide, the H region governs hydrophobicity, and the C region influences the adoption of beta conformation in its secondary structure (Zanen et al., 2005). The dashed line in Figure 5. 3 marks the cleavage point where the protein is released into the periplasm. The choice of amino acids in the signal peptide determines its hydrophobic or

hydrophilic nature. The hydrophobicity of a signal peptide is crucial for successfully transporting a protein to the correct destination. Generally, low hydrophobic signal peptides necessitate more ATP for this process, reducing the cell's efficiency.

Cleavage percentage plays a crucial role in successfully delivering proteins into the periplasm. If the signal peptide fails to cleave off the unfolded protein, it can lead to misfolded protein. Once the signal peptide reaches its destination, it detaches from the unfolded protein, allowing the protein to move into the periplasmic space (Han et al., 2017). Assessing the peptide’s solubility within the inner membrane is essential to prevent it from entering the periplasm (Plessis et al., 2011) .

Hydrophobicity determines the signal peptide's translocation route. Highly hydrophobic peptides can travel to the periplasm through an ATP-independent route, while less hydrophobic peptides require ATP-dependent SecA and SecB.

5.1.2 Collection of data on signal peptides and modifications introduced to the model

This experiment aims to study the impact of modifying the hydrophobicity of a signal peptide on the quantity of folded protein produced. The focus is on adjusting the cleavage percentage and the solubility of the signal peptide to determine their effect on the translocation pathway and the amount of correctly folded protein.

The assumptions for this test case are as follows:

- Cleavage percentage is directly linked to the rate at which the protein transitions from the cytoplasm to the periplasm.
- Codons for constructing the signal peptides are readily available and integrated into the system boundaries.
- The signal peptide's translocation is limited to the periplasm, and changes in signal peptides do not affect Tat.
- The current model uses L-asparaginase II as the signal peptide (Zamani et al., 2016).

Petri net modifications for highly hydrophobic signal peptides are shown on the nets and explained in a Chapter 3 scenario. The hydrophobicity for each tested signal peptide depends on its specific amino acid sequence.

5.1.3 Results and Discussion

5.1.3.1 Simulations for the high and low hydrophobic signal peptides

In this study, the input parameters for test cases concerning glucose and ATP availability remain consistent across all scenarios. However, the glucose depletion rate varies depending on the energy demands associated with protein production. Notably, highly hydrophobic signal peptides do not require ATP, allowing for the reallocation of this resource to other pathways within the model. In contrast, signal peptides with lower hydrophobicity exhibit a higher ATP demand, reducing ATP availability for other integrated pathways. This, in turn, has a consequential impact on overall protein production.

The collected data is presented in Table 5.1. Figures 5. 4, 5. 5, 5. 6 and 5. 7, are assigned a number as indicated in Table 1, and these numbers are used as references throughout the test case.

Figure 5. 4 shows the different transition rates corresponding to the hydrophobicity data in Table 1. Graph 1 represents the signal peptide with the highest hydrophobicity, which displays the highest transition rate during the translocation process. Additionally, Graph 6 illustrates the ATP-independent routes with the lowest hydrophobicity within this group of signal peptides, shown in the lightest shade.

Figure 5. 5 shows the different translocation rates for the signal peptides with the least hydrophobicity. These peptides require ATP to complete the translocation process. The light pink colour in the graph represents the least hydrophobic peptide from the second group. In contrast, the dark purple colour represents a single peptide that is more hydrophobic but still requires ATP for translocation. The order of the translation rates follows the information presented in Table 1.

Figures 5. 5 and 5. 7 display graphs that were generated using different signal peptides. The figures were numbered in the same order as Figures 5. 4, and 5. 6. There are apparent differences between the graphs generated by different Sec translocation pathways. Graphs 1-6 represent cases that can produce 50% more folded protein than those presented in Graphs 7-12 because of the additional ATP-dependent steps involved when low hydrophobicity signal peptides are used. The ATP-dependent steps become a limiting factor that results in faster glucose depletion and, consequently, faster de novo ATP production in cases represented in Graphs 7-12 compared to those shown in Graphs 1-6.

Table 5. 1 Scenarios Evaluated in Test Case 1: Impact of Signal Peptide Hydrophobicity on Translocation Pathway Selection. This table presents the 12 scenarios evaluated in Test Case 1, which investigates the effect of signal peptide hydrophobicity on the selection of translocation pathways. The scenarios are divided into two groups based on the hydrophobicity of the signal peptides: Highly Hydrophobic Signal Peptides (Scenarios 1-6): Hydrophobicity levels greater than 100 kJ/mol (derived from the Kyte-Doolittle scale). These peptides are more likely to favour the Sec translocation pathway, which transports unfolded proteins. Less Hydrophobic Signal Peptides (Scenarios 7-12): Hydrophobicity levels less than 100 kJ/mol (derived from the Kyte-Doolittle scale). These peptides are more likely to favour the Tat translocation pathway, which transports folded proteins.

Signal Peptide	Associated graph number	Amino Acid Sequence	Cleavage percentage	Signal peptide solubility (µM)	Hydrophobicity (kJ/mol)	Reference
flgI (H)	1	MVIKFLSALILLLVTTAAQA	97%	0.806	185.500	(Ahmadi et al., 2023) (Nishikino et al., 2023)
OmpC (H)	2	MKVKVLSLLVPALLVAGAANA	96%	0.797	171.900	(Ahmadi et al., 2023) (Jung et al., 2020) (Xu & Lee, 1999)
NPPC (H)	3	MHLSQLLACALLLTLLSLRPSEA	98%	0.737	165.600	(Ahmadi et al., 2023) (Low et al., 2013)
MepA (H)	4	MNKTAIALLALLASSVSLA	95%	0.833	164.740	(Ahmadi et al., 2023) (Marcyjaniak et al., 2004)
CysP (H)	5	MAVNLLKKNSLALVASLLLAGHVQA	99%	0.765	164.000	(Ahmadi et al., 2023) (Mansilla & De Mendoza, 2000) (Hryniewicz et al., 1990)
DsbA (H)	6	MKKIWLALAGLVLAFSASA	94%	0.700	144.210	(Ahmadi et al., 2023)(Durrani et al., 2019)

BcsB (L)	7	MKRKLFWICAVAMGMSAFPSFMTQA	89%	0.874	58.800	(Ahmadi et al., 2023) (Acheson et al., 2021)
DsbC (L)	8	MKKGFMLFTLLAAFSGFAQA	99%	0.836	78.500	(Ahmadi et al., 2023) (Grasso et al., 2023)
ZraP (L)	9	MKRNTKIALVMMALSAMAMGSTSAFA	95%	0.834	79.230	(Ahmadi et al., 2023)
AnsB (L)	10	MEFFKKTALAALVMGFSGAALA	96%	0.846	93.640	(Ahmadi et al., 2023) (Jennings & Beacham, 1990)
AraF (L)	11	MHKFTKALAAIGLAAVMSQSAMA	98%	0.876	93.910	(Ahmadi et al., 2023) (Acheson et al., 2021)
AmpC (L)	12	MFKTTLCALLITASCSTFA	63%	0.783	97.980	(Ahmadi et al., 2023) (Nelson & Elisha, 1999)

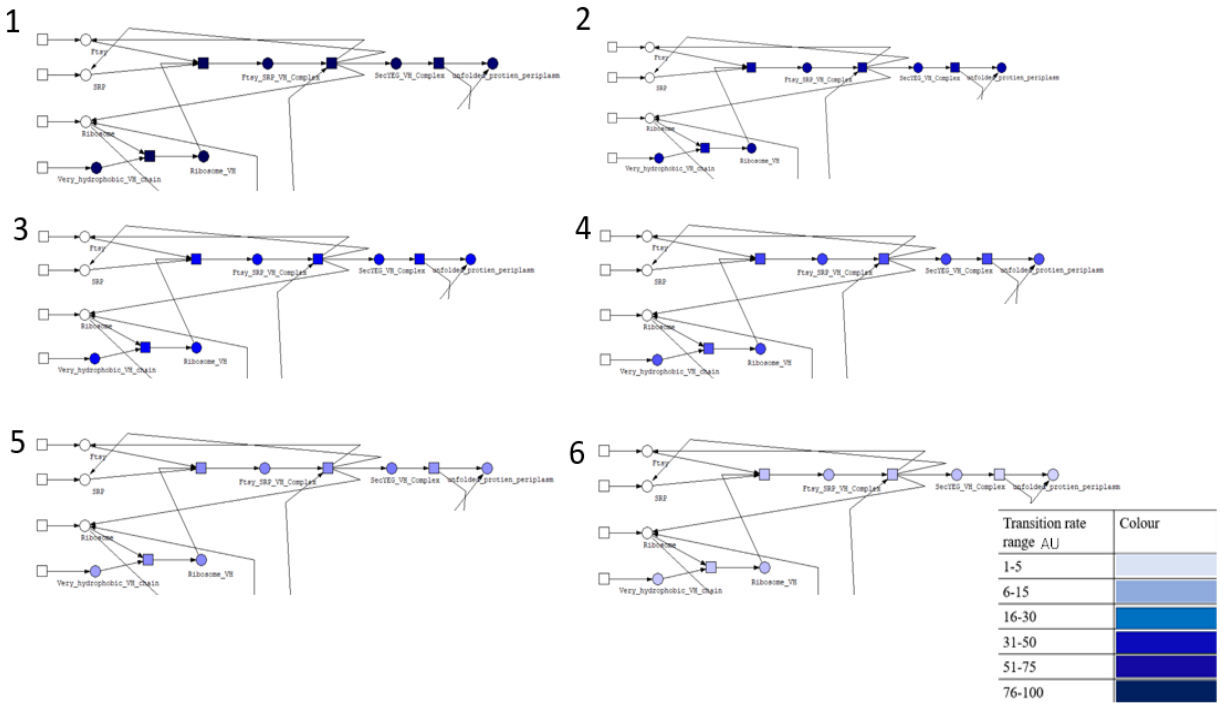


Figure 5. 4 Transition Rates for Signal Peptides with High Hydrophobicity: Impact of Slow and Fast Rates on Translocation Efficiency. The transition rates represent the speed at which signal peptides are recognised and processed by the translocation machinery, influencing the overall efficiency of protein secretion

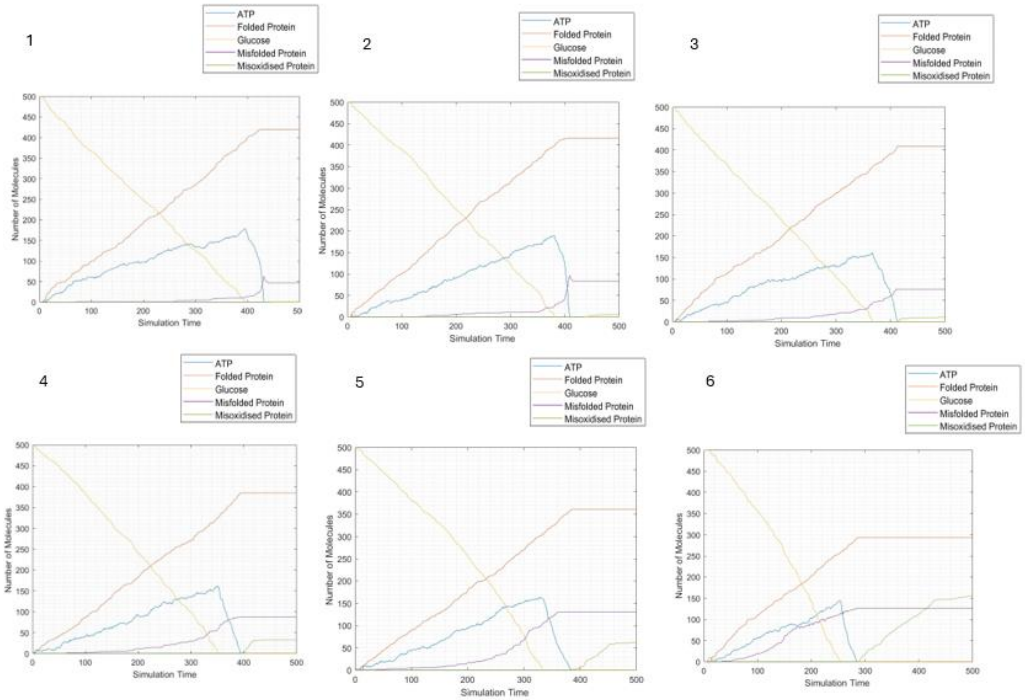


Figure 5. 5 Glucose Consumption, ATP Utilisation, and Misfolded Protein Production for Highly Hydrophobic Signal Peptides (Graphs 1-6). This figure presents Graphs 1-6, which depict the relationship between glucose consumption, ATP utilisation, and misfolded protein production for six highly hydrophobic signal peptides. The graphs illustrate how variations in hydrophobicity and energy metabolism influence protein folding efficiency and the accumulation of misfolded proteins.

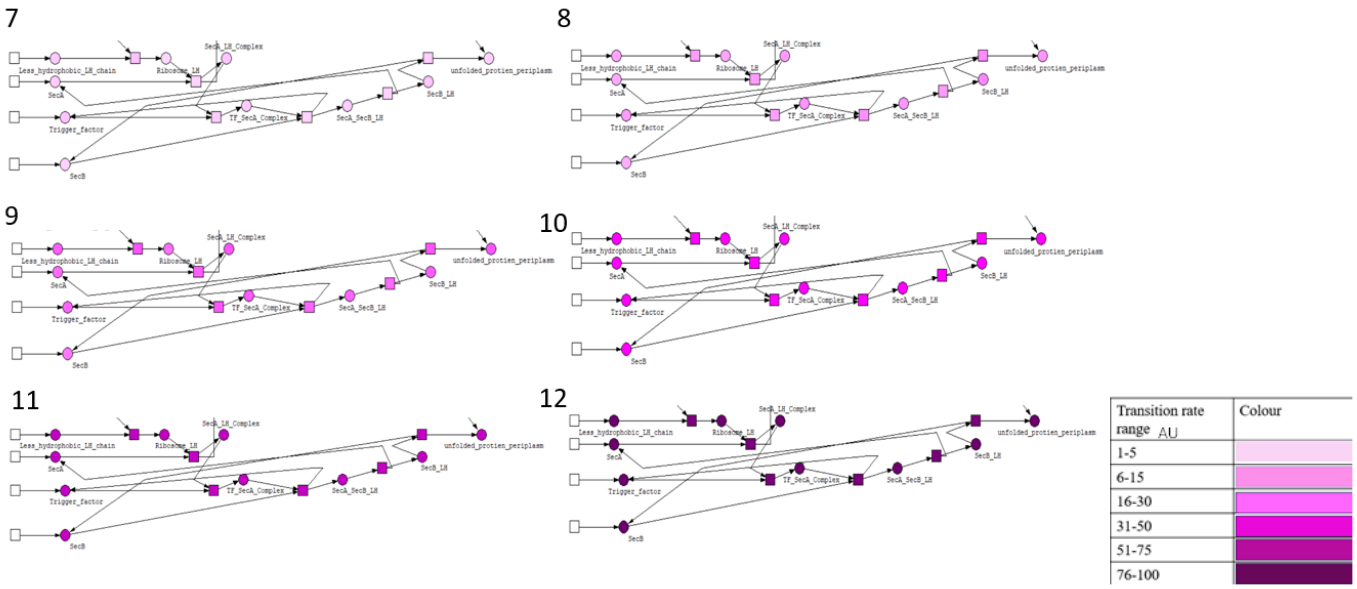


Figure 5. 6 Transition Rates for Signal Peptides with Low Hydrophobicity: Impact of Slow and Fast Rates on Translocation Efficiency. This figure illustrates the transition rates assigned to signal peptides with low hydrophobicity. The transition rates represent the speed at which signal peptides are recognised and processed by the translocation machinery, influencing the overall efficiency of protein secretion.

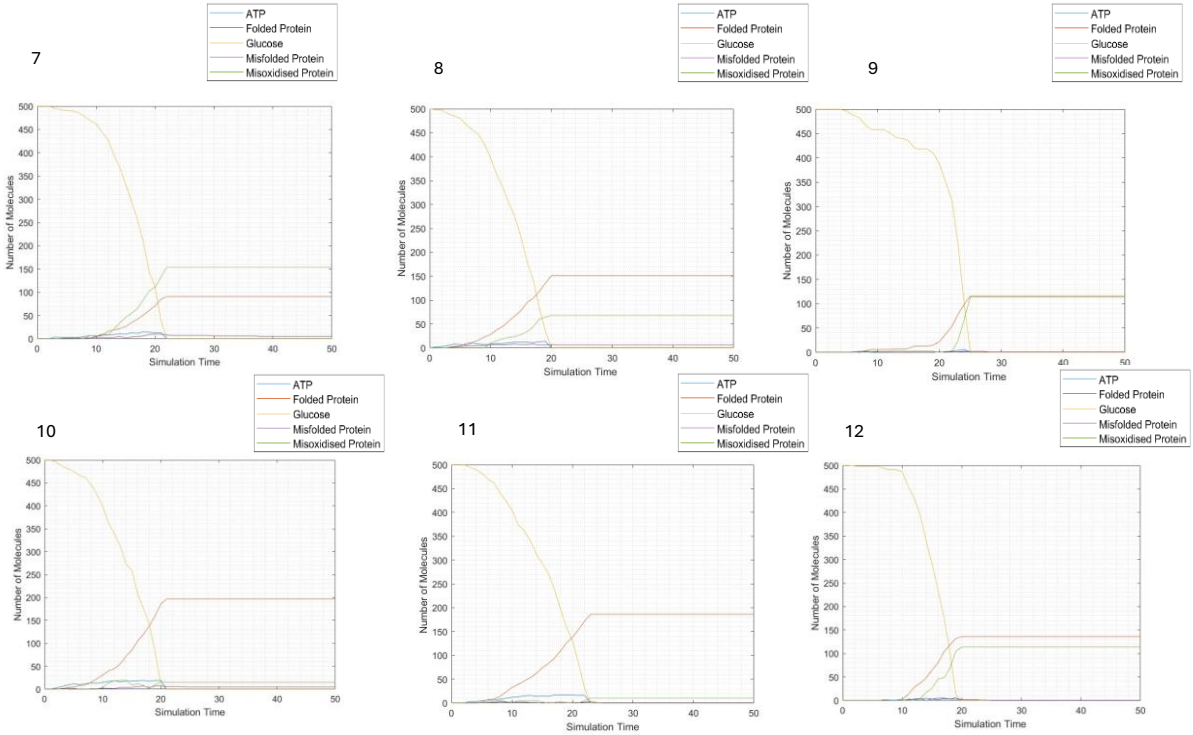


Figure 5. 7 Glucose Consumption, ATP Utilisation, and Protein Production for Low Hydrophobicity Signal Peptides (Graphs 7-12). This figure presents Graphs 7-12, which depict the dynamics of glucose consumption, ATP utilisation, and protein production for signal peptides with low hydrophobicity. The graphs illustrate how energy demands and signal peptide properties influence protein production efficiency and the accumulation of mis-oxidised proteins.

5.1.3.2 Evaluation of simulation outcomes

This is a detailed comparative analysis of the performance of various signal peptides in protein translocation, as depicted in Graphs 1-12. The analysis focuses on the relationship between signal peptide properties (e.g., hydrophobicity, solubility, cleavage efficiency) and their impact on protein folding, misfolding, and mis-oxidisation. The results are contextualised within the broader literature to highlight the implications for recombinant protein production in microbial systems.

Graph 1 (flgI) demonstrates the efficiency of highly hydrophobic signal peptides, producing 400 molecules of folded protein, 50 molecules of misfolded protein, and no mis-oxidised protein. The flgI signal peptide does not require ATP for translocation, utilising the signal recognition particle (SRP) pathway, which is highly efficient for hydrophobic peptides (Karamyshev & Johnson, 2005). The absence of mis-oxidised proteins indicates effective redox balance and proper disulphide bond formation.

In Graph 2 (OmpC), the amount of folded protein is similar to flgI, but there is 42% more misfolded protein. The slightly lower solubility of OmpC results in slower translocation, leading to increased misfolding. Graph 3 (NPPC) shows 11% less folded protein than flgI and OmpC, along with 42% more misfolded protein. The reduced efficiency of NPPC highlights the importance of signal peptide solubility and hydrophobicity in determining translocation success (Auclair et al., 2012).

Graph 4 (MepA) produces 400 molecules of folded protein but shows double the misfolded protein and 25% more mis-oxidised protein compared to flgI. Glucose depletion is 14% faster, indicating higher energy demands for translocation and folding. Graph 5 (CysP) exhibits an increase in mis-oxidised protein compared to MepA, suggesting challenges in maintaining redox balance during translocation.

Graph 6 (DsbA) shows more than double the mis-oxidised protein compared to CysP. DsbA is effective for proteins requiring thioredoxin, which maintains redox balance in the periplasm. However, the presence of 4 disulphide bonds in human growth hormone (HGH) complicates folding, necessitating further investigation into disulphide bond formation pathways (Test Cases 3 and 4).

Graph 7 (BcsB) produces 50% more mis-oxidised protein than folded protein, with rapid glucose depletion. BcsB is used in type 4 translocation pathways, which are less efficient for periplasmic targeting. In contrast, Graph 8 (DsbC) produces 50% more folded protein and 75% less mis-oxidised protein compared to BcsB. DsbC's higher hydrophobicity and post-translational requirements make it more efficient for protein folding.

Graph 9 (ZraP) shows a 50% increase in mis-oxidised protein and a 25% decrease in folded protein compared to DsbC. ZraP's suitability for zinc-stress conditions limits its efficiency in standard translocation pathways. Graph 10 (AnsB) produces results similar to DsbC, with slightly higher hydrophobicity improving folding efficiency.

Graph 11 (AraF) produces the same amount of folded protein as AnsB but with no mis-oxidised protein. AraF's high solubility and cleavage efficiency make it ideal for reducing oxidative stress and improving translocation success (Nilsson et al., 1993). Finally, Graph 12 (AmpC) produces results comparable to ZraP, with mis-oxidised protein due to its reliance on ATP-dependent Sec translocation.

The simulation results demonstrate that the choice of signal peptide impacts protein translocation efficiency, folding, and energy metabolism. Highly hydrophobic peptides (e.g., flgI, OmpC) are more efficient for SRP-mediated translocation, requiring no ATP and producing fewer misfolded proteins. In contrast, low hydrophobicity peptides (e.g., BcsB, ZraP) rely on ATP-dependent pathways, leading to higher energy demands and increased misfolding.

Energy metabolism plays a critical role in protein folding. Rapid glucose depletion in low hydrophobicity pathways (e.g., BcsB, MepA) highlights the trade-off between protein production rates and energy availability. Efficient energy utilisation is essential for maintaining redox balance and minimizing mis-oxidised protein production.

Signal peptide properties such as solubility and cleavage efficiency also influence protein yield and quality. Peptides with high solubility and cleavage efficiency (e.g., AraF) produce more folded proteins with fewer mis-oxidised byproducts, making them ideal for recombinant protein production (Auclair et al., 2012) (Nilsson et al., 1993).

Finally, the presence of multiple disulphide bonds (e.g., in HGH) complicates folding and requires redox balance and chaperone activity. This underscores the need for further investigation into disulphide bond formation pathways, as explored in Test Cases 3 and 4.

Table 5. 2 Summary of Outcomes for Test Case 1: Impact of Signal Peptide Hydrophobicity on Protein Translocation and Folding Efficiency

Graph number /Signal Peptide	Model pass/fail	Reference
1 flgI (H)	Pass	(Y. Zhou et al., 2014)
2 OmpC (H)	Pass	(Kojima & Nikaido, 2014) (Jung et al., 2020b)
3 NPPC (H)	Pass	(Voulhoux et al., 2001)
4 MepA (H)	Pass. However, due to limited literature, the actual impact of this signal peptide is not well documented. The model can show that the amount of protein is impacted as the hydrophobicity is reduced.	(Allen et al., 2022)
5 CysP (H)	Pass	(M.-J. Han & Lee, 2006)
6 DsbA (H)	Pass	(Schierle et al., 2003)
7 BcsB (L)	Pass	(Abidi et al., 2022)
8 DsbC (L)	Pass	(Berkmen et al., 2005)
9 ZraP (L)	This single peptide is better represented when a recombined protein is affected by Zinc stress. The model can show similar trends to other low hydrophobic signal peptides and has passed.	(Weiner & Li, 2008)
10 AnsB (L)	Pass	(Huber et al., 2005)
11 AraF (L)	Pass	(Arts et al., 2015)

12 AmpC (L)	Pass	(Kaderabkova et al., 2022)

Table 5. 2 provides a comprehensive account of how each signal peptide influences protein production in response to its extent of hydrophobicity. Highly hydrophobic signal peptides exhibit a range of trends compared to the available literature. In contrast, low hydrophobic signal peptides demonstrate how rapid ATP consumption results in lower yields than their highly hydrophobic counterparts.

The data presented above highlights hydrophobicity's impact on the amount of folded recombinant protein produced. This variation is attributed to the differences in signal peptide chains and their interaction with various chaperones. The model generated another insightful feature, i.e., the transition rate profile for the translocation steps. The pattern generated by this feature and the values can indicate whether the signal peptide has low or high hydrophobicity based only on the information provided as the target protein sequence. This is shown below in Figures 5. 8 and 5. 9.

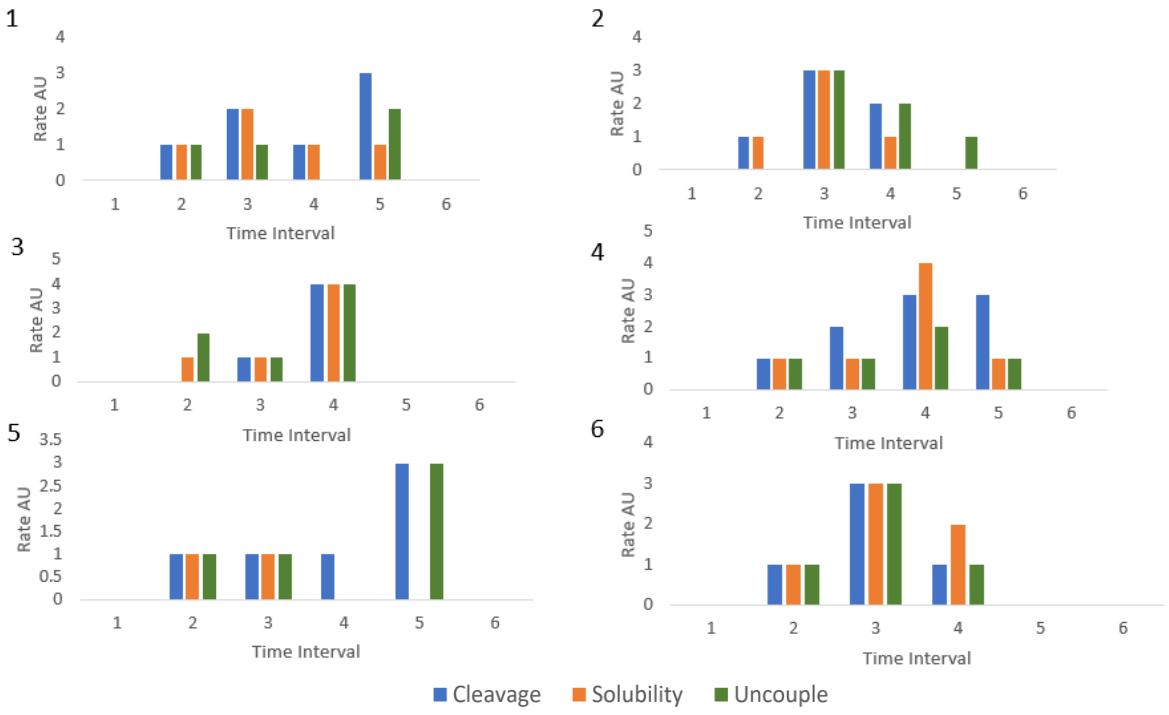


Figure 5. 8 The ratio of Transition Rate to Simulation Time for Highly Hydrophobic Signal Peptides. This figure presents the ratio of transition rate to simulation time for the first six highly hydrophobic signal peptides (flgI, OmpC, NPPC, MepA, CysP, and DsbA) across six intervals. The translocation rate varies during the simulation, reflecting the dynamic nature of protein translocation and folding processes. The cleavage, solubility, and uncoupled translocation rates have been analysed to understand the relationship between overall translocation rates and the amount of protein generated.

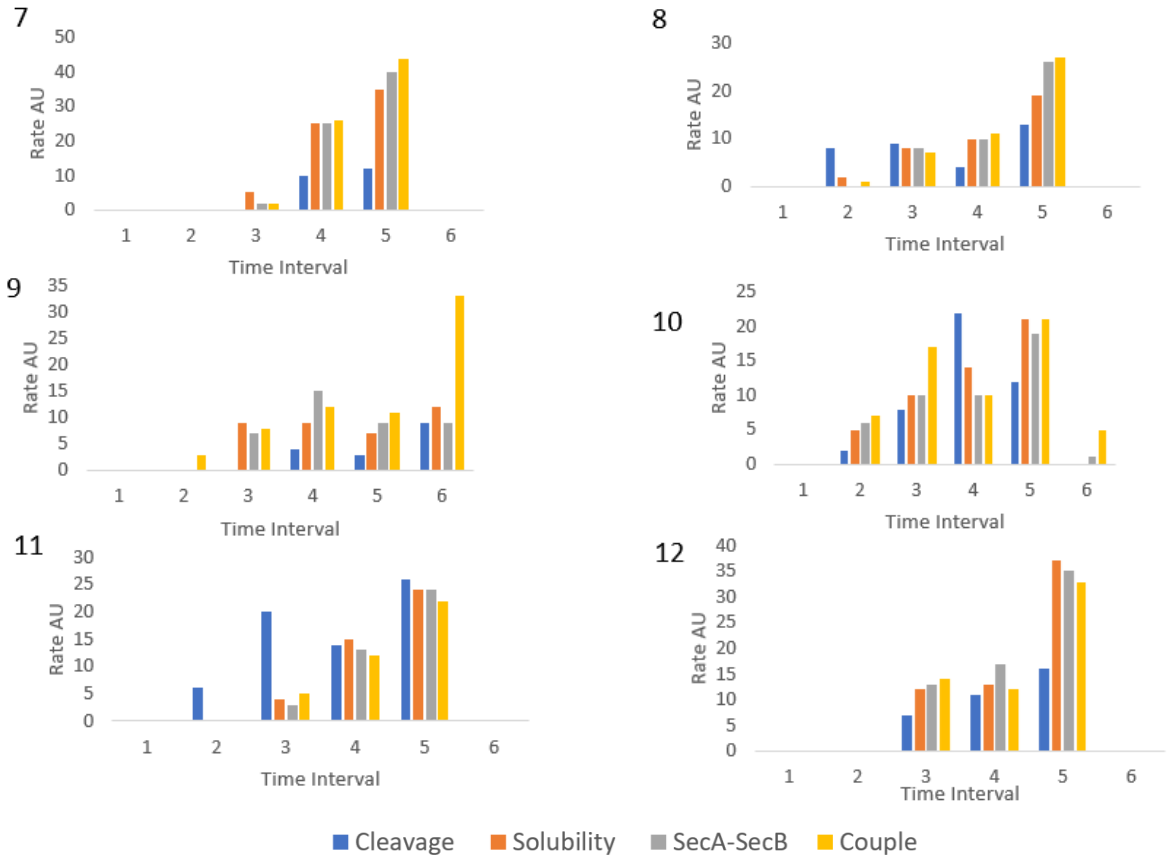


Figure 5. 9 Ratio of Transition Rate to Simulation Time for Less Hydrophobic Signal Peptides. This figure presents the ratio of transition rate to simulation time for six less hydrophobic signal peptides (BcsB, DsbC, ZraP, AnsB, AraF, and AmpC) across six intervals. The translocation rate varies during the simulation, reflecting the dynamic nature of protein translocation and folding processes in ATP-dependent pathways. The rates of cleavage, solubility, SecA-SecB complex activity, and coupled translocation have been analysed to understand the relationship between overall translocation rates and the amount of protein generated..

These figures illustrate the transition rates and simulation times for protein translocation pathways involving highly hydrophobic (Figure 5.8) and less hydrophobic (Figure 5.9) signal peptides. The analysis highlights the impact of hydrophobicity, energy requirements, and additional translocation steps on protein production efficiency and the accumulation of mis-oxidised proteins.

Highly hydrophobic signal peptides require longer simulation times but result in higher protein production. This is because these peptides utilise the SRP-mediated translocation pathway, which does not require ATP and is highly efficient for hydrophobic sequences. In Graph 1, interval 1,

the rates of cleavage, solubility, and uncoupling are equal, indicating a balanced translocation process. However, as the simulation progresses, cleavage and solubility rates increase, revealing these as bottlenecks in later intervals. The transition rates in Figure 5.8 are 10 times faster than those in Figure 5.9, reflecting the efficiency of highly hydrophobic peptides in SRP-mediated translocation.

Less hydrophobic signal peptides rely on the SecA-SecB complex, which requires ATP for translocation. This additional step leads to faster ATP depletion and shorter simulation times, limiting protein production due to insufficient energy. The transition rates in Figure 5.9 are slower than those in Figure 5.8, reflecting the energy demands of the SecA-SecB pathway. Graphs 8, 10, and 11 exhibit lower transition rates but produce fewer mis-oxidised proteins, highlighting the trade-off between translocation speed and protein quality. Lower transition rates correlate with reduced production of mis-oxidised proteins, as slower translocation allows for more efficient folding and quality control.

These figures provide a quantitative comparison of translocation pathways for high and low hydrophobicity signal peptides, highlighting the trade-offs between protein yield, energy requirements, and protein quality. The insights gained from this analysis are critical for improving signal peptide selection and recombinant protein production in microbial systems. By understanding the factors that influence translocation efficiency, researchers can design more effective strategies for enhancing protein yield and quality in biotechnological applications.

To further analyse the trends, a paired t-test can be performed by pooling the transition rates for the translocation pathways of high and low hydrophobicity signal peptides. A threshold value of 0.01 will be used to determine statistical significance. This analysis will provide quantitative insights into the differences in translocation efficiency and protein yield between the two pathways.

Table 5. 3 The average translocation for Graphs 1 to 12 with the calculations for a paired t-test

Number of pairs	Average translocation rate (AU) (Graph 1- 6)	Average translocation rate (AU) (Graph 7-12)	Difference
1	1.5	21.5	20.00
2	1.5	10.1	9.00
3	2.3	12.5	10.2
4	1.8	11.1	9.30
5	1.4	14.4	13.00

6	1.7	18.3	16.60
Mean difference			12.95
The standard deviation of the difference			7.14
Degrees of freedom			5.00
t value			6.99
P value			0.00092

The p-value is less than the critical value, which rejects the null hypothesis and confirms that the two data sets are significantly different. This is beneficial knowledge when testing proteins where the translocation pathways are not yet known, but the sequence is given. The model will be able to demonstrate which pathway the protein of interest will take, which can aid the design step during plasmid construction.

Overall, this test case examines how the hydrophobicity of signal peptides affects protein translocation across the *E. coli* membrane. The results show that using the SRP pathway, highly hydrophobic signal peptides, like flgI, OmpC, and NPPC, translocate proteins more efficiently than less hydrophobic signal peptides and without ATP consumption. Factors such as solubility and cleavage efficiency also influence translocation effectiveness; low solubility or poor cleavage can lead to slower translocation and more misfolded proteins.

The study further explores various signal peptides (e.g., MepA, CysP, DsbA, BcsB, DsbC, ZraP, AnsB, AraF, AmpC) using the Petri net model. It finds that highly hydrophobic peptides generally produce more protein and fewer mis-oxidised proteins. A paired t-test confirms a significant difference in translocation rates between highly and less hydrophobic peptides.

These findings underscore the importance of signal peptide hydrophobicity in protein translocation and offer insights for improving protein expression in biotechnological applications.

5.2 Test Case 2: Role of cytosolic disulphide bond formation on the extent of protein mis-oxidation in the CyDisCo strain

5.2.1 Introduction to Test Case 2

There has been an effort to engineer strains capable of forming disulphide bonds within the cytoplasm to enhance disulphide bond formation within *E. coli*. Test case 3 aims to study the redox potential and its importance in constructing disulphide bonds within *E. coli*, mainly when the protein is not native to *E. coli*. This test case will involve an examination of a commercial strain, implementing pathways within the Petri net, and focusing on the necessary adjustments to enhance the number of folded proteins.

Cytoplasmic Disulphide Bond Formation (CyDisCo) is an engineered strain used commercially because it can form disulphide bonds within the cytoplasm and transport protein to the periplasm using the Tat translocation pathways. This strain has undergone many modifications and tests to reduce the number of inclusion body formations and increase the amount of folded protein. (Tungekar & Ruddock, 2023). This included conjugating DsbA protein with sulfhydryl oxidase and mutating glutathione reductase to regulate the redox potential. (Gaciarz et al., 2017).

The Petri net model simulates the formation of disulphide bonds within the cytoplasm and the transport of the folded protein using the Tat pathway. The model can generate results to demonstrate the impact of this modification and conduct several scenarios to test for methods to increase protein yield. The model protein used for this test case will be human growth hormone, which has 4 disulphide bonds, as in the previous test case (Alanen et al., 2015).

This protein can be simulated within the model, and the pathways can be tested under different conditions to understand what impacts the amount of folded protein. Reducing the number of inclusion bodies and mis-oxidised protein is essential in addition to obtaining folded protein (Castillo-Corujo, Saaranen, et al., 2024). This test case will simulate the impact of folding disulphide bonds within the cytoplasm and test methods for reducing the mis-oxidised protein formed.

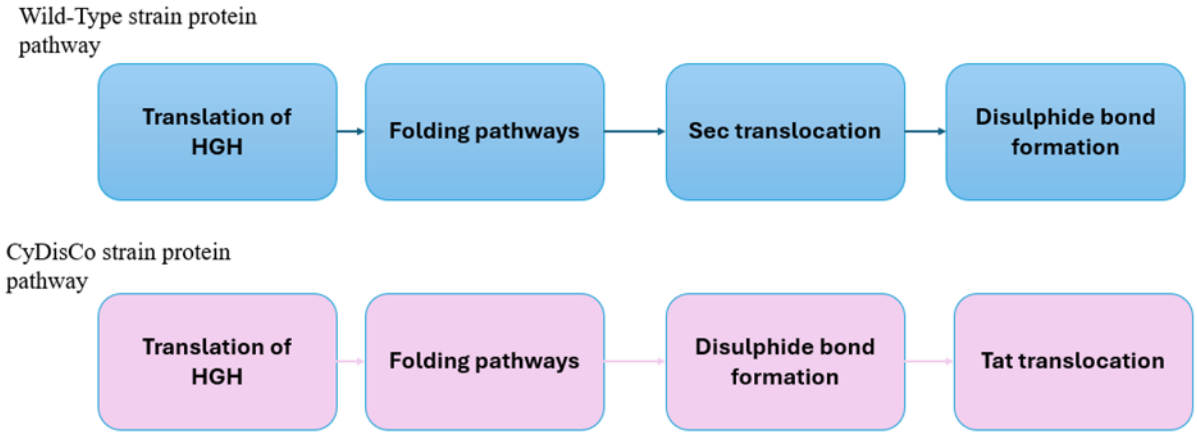


Figure 5. 10 Flow Diagram Comparing Protein Pathways in CyDisCo Strain and Wild-Type Strain. This figure presents a flow diagram illustrating the differences in protein pathways between the CyDisCo and wild-type strains. The CyDisCo strain, engineered for efficient disulphide bond formation in the cytoplasm, demonstrates enhanced protein folding and reduced mis-oxidised protein production compared to the wild-type strain, which relies on periplasmic disulphide bond formation.

The CyDisCo strain and wild-type *E. coli* differ in their ability to form disulphide bonds in the cytoplasm. In wild-type *E. coli*, the cytoplasm's reducing environment prevents disulphide bond formation, typically occurring in the periplasm. The CyDisCo strain, however, is genetically engineered to enable disulphide bond formation in the cytoplasm by expressing proteins like sulfhydryl oxidase and disulphide isomerase (Matos et al., 2014). Whereas wild-type *E. coli* is suitable for expressing proteins that do not require disulphide bonds (A. Zhou et al., 2015).

5.2.2 Collection of data on disulphide bond formation using CyDisCo strain and modifications to the model

This study collects data regarding disulphide bond formation in proteins using the CyDisCo strain of *E. coli*. Employing this genetically engineered strain can achieve efficient disulphide bond formation in the cytoplasm, which is typically challenging in wild-type *E. coli*. Additionally, the model is modified to accurately reflect the unique oxidative environment and folding pathways present in the CyDisCo system, enhancing the understanding of disulphide bond formation and its impact on protein functionality

The data used to construct the model is from a comparative study highlighting chaperones' upregulation during protein production in CyDisCo. (Guerrero Montero et al., 2019). It is important to note that the number of studies on CyDisCo is limited compared to Sec translocation, which means conclusions must be drawn cautiously. Due to data limitations, this test case aims to show how any strain can be modelled with the required modifications. CyDisCo was modelled to gather information on how the Petri net model responds to

disulphide bonds within the cytoplasm. Modelling this strain aims to demonstrate how the Petri net model can be altered for different recombinant proteins and strains, providing information to aid experiment designs.

The assumptions for this test case are as follows:

1. DsbA is linked with sulfhydryl oxidase
2. Redox potential is balanced within the cytoplasm
3. Glucose will be fed into the system boundary to understand the impact of disulphide bonds within cytoplasm as opposed to energy limitations
4. Signal peptide for Tat is TorA (Arauzo-Aguilera et al., 2023)

The Petri net model was modified by removing Sec translocation pathways and transferring folded recombinant protein within the cytoplasm, which is transported to the periplasm using the Tat translocation pathways. The green square marks the end of the Petri net, where Folded_TAT marks folded protein transported to the periplasm using Tat translocation.

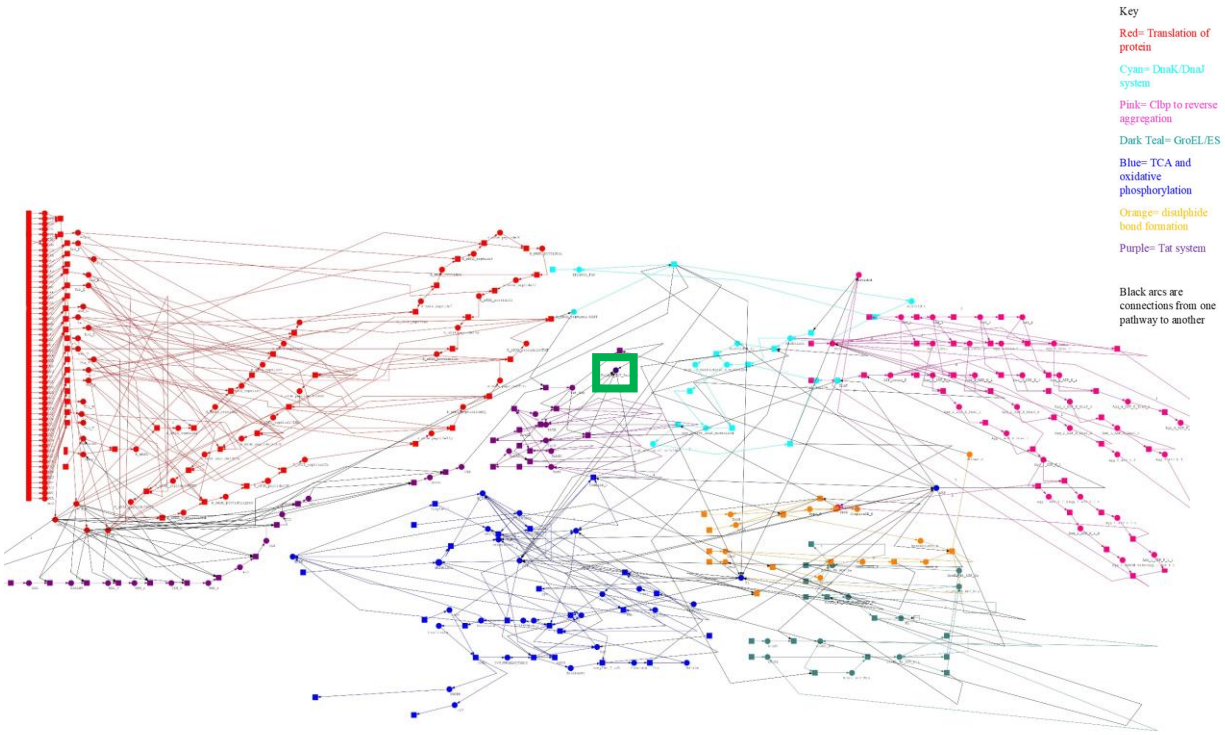


Figure 5. 11 Petri net Model Construction for CyDisCo System. This figure depicts the Petri net model constructed to simulate the CyDisCo system, which is engineered for efficient disulphide bond formation in the cytoplasm. The model integrates multiple pathways involved in protein translation, folding, misfolding, aggregation, and translocation, with each pathway represented by a distinct colour scheme. The green box marks the end of the model, indicating successful protein translocation to the periplasm. Due to the size of the image, the resolution of this image may be reduced. To view the image in full resolution, please click here.

5.2.3 Results and Discussion

The production of recombinant proteins, such as human growth hormone (HGH), in microbial systems like *E. coli* often faces challenges related to protein folding, disulphide bond formation, and redox balance. The CyDisCo strain has been engineered to address these challenges by enabling efficient disulphide bond formation in the cytoplasm. However, achieving optimal protein yield and quality requires a detailed understanding of the underlying biological processes and the ability to predict and mitigate bottlenecks.

This test case analyses the simulation results of HGH production in the CyDisCo strain against current literature, highlighting the role of Petri nets in identifying bottlenecks and improving protein yield. The discussion integrates findings from experimental studies within the literature and computational modelling to underscore the importance of Petri nets in strain development and bioprocess optimisation.

The Petri net model simulating HGH production in the CyDisCo strain revealed that the number of mis-oxidised proteins is almost three times the amount of folded protein (Figure 5. 12). This outcome is consistent with the expected challenges of maintaining redox balance in the cytoplasm, where the absence of proper redox regulation leads to inefficient disulphide bond formation and increased misfolding. Additionally, the model predicts a rise in inclusion body formation as folded proteins that cannot be transported accumulate and aggregate (Prahlad et al., 2021). These results align with experimental observations, where improper redox balance and inefficient translocation often lead to protein aggregation and reduced yields.

Experimental studies on the CyDisCo strain have focused on enhancing its ability to form disulphide bonds to increase the yield of correctly folded proteins. These studies have identified key folding and translocation pathway steps that can be optimised to improve strain performance. For example, introducing thioredoxin and glutathione pathways has enhanced the cytoplasm's redox balance and disulphide bond formation (Gaciarz et al., 2017). The Petri net model complements these experimental efforts by predicting which parameters are likely to reduce bottlenecks, such as an increased rate of DnaK before any experimental work is conducted.

By adjusting parameters such as chaperone activity, redox balance, and translocation rates, the model can predict the impact of these changes on protein yield and quality. This approach is particularly advantageous for strain development, as it allows researchers to focus experimental efforts on the most critical bottlenecks, reducing the time and resources required for strain optimisation (Gaciarz et al., 2017).

The analysis presented in Figures 5.12, 5.13, and 5.14 demonstrates the utility of Petri nets in understanding and optimising disulphide bond formation and protein production in the

CyDisCo strain. By aligning computational modelling with experimental approaches, the model provides a powerful tool for predicting and addressing bottlenecks, ultimately improving protein yield and quality. This integrated approach highlights the importance of combining computational and experimental methods to advance recombinant protein production in microbial systems.

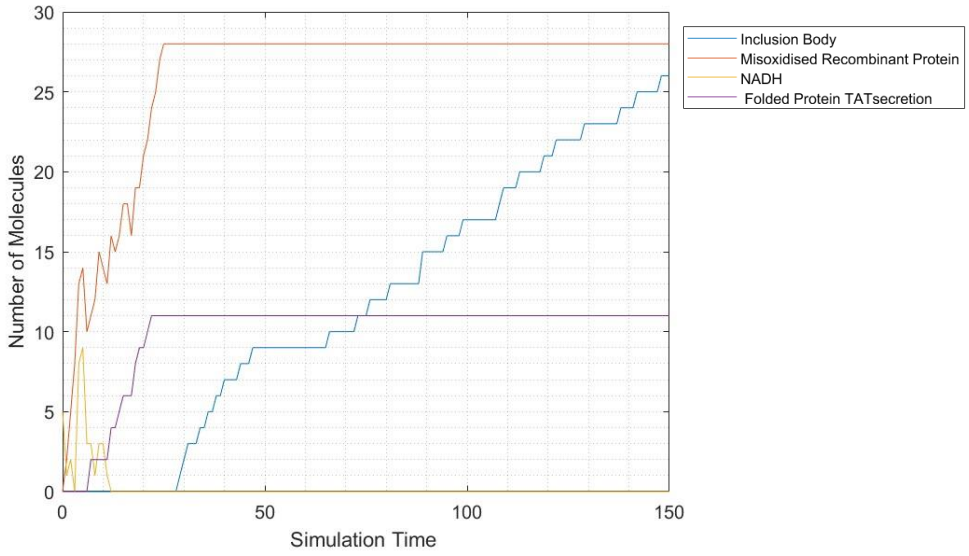


Figure 5. 12 Simulation of Disulphide Bond Formation in the Cytoplasm Using NADH as a Reducing Agent. This graph presents the results of modelling cytoplasmic disulphide bond formation in a microbial system, where NADH serves as the reducing agent. The simulation demonstrates the production of 11 molecules of folded protein with a Tat signal peptide, alongside 28 molecules of mis-oxidised protein and 26 molecules of inclusion bodies. The units for protein quantities are arbitrary, reflecting relative trends rather than absolute values.

Under normal conditions, the simulation shows that mis-oxidised protein levels are 60% higher than folded protein levels, primarily due to the inability of DsbC and DsbD to reduce unfolded proteins effectively. However, when the isomerisation rate of DsbC and DsbD is increased, the number of folded proteins is reduced to 18 molecules, and mis-oxidised protein accumulation is reduced. These results align with experimental data, demonstrating that upregulating DsbC and DsbD enhances protein folding efficiency and reduces mis-oxidised protein accumulation (Kadokura et al., 2003).

The simulation highlights the critical role of these proteins in maintaining redox balance and proper protein folding, providing a predictive framework for optimising their activity in microbial systems. By demonstrating the impact of upregulating DsbC and DsbD, the model guides experimental efforts to enhance protein production and reduce misfolding, ultimately improving the efficiency of biotechnological processes.

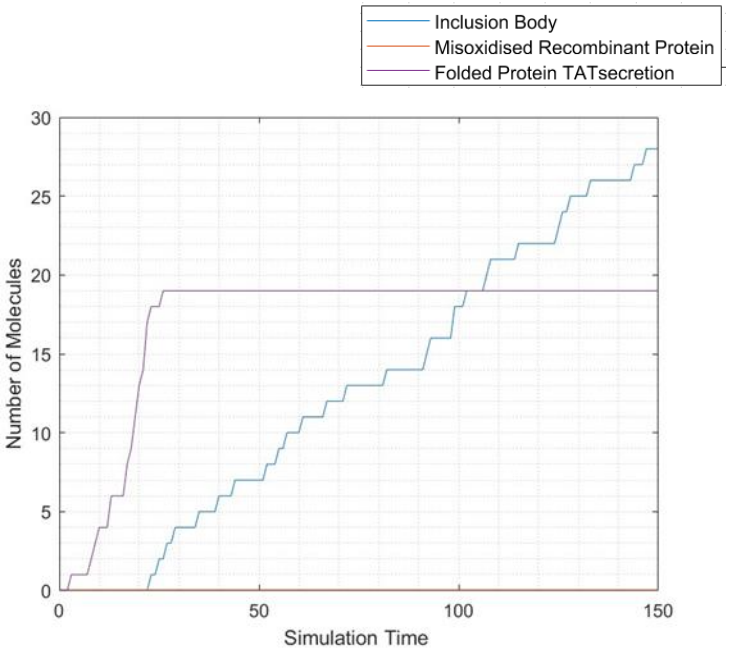


Figure 5. 13 Impact of DsbC and DsbD Upregulation on Protein Folding and Mis-oxidised Protein Accumulation. This figure illustrates the impact of upregulating DsbC and DsbD proteins on protein folding efficiency and the accumulation of mis-oxidised proteins.

The simulation demonstrates that while increasing the rate of the ClpB chaperone does not reduce inclusion body formation, upregulating the DnaK chaperone effectively prevents accumulation and increases folded protein production by 33%. These results align with experimental studies showing that DnaK is critical in preventing protein aggregation and promoting proper folding (Mogk et al., 1999).

The model's ability to predict the impact of specific parameters, such as chaperone activity, on protein yield underscores its utility as a predictive tool for strain development. By identifying key factors influencing protein production, the model enables researchers to design targeted experiments, reducing the time and resources required for strain development. This approach enhances our understanding of protein folding dynamics and provides a framework for improving recombinant protein production in microbial systems.

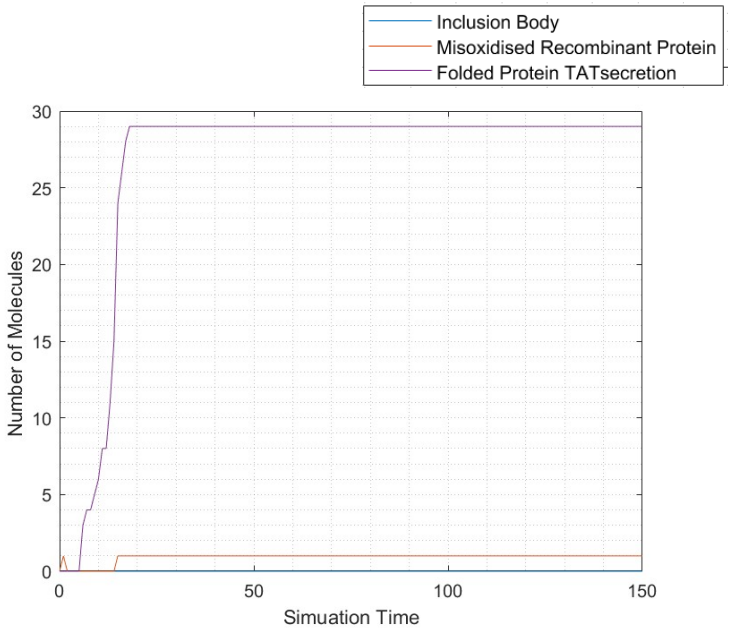


Figure 5. 14 Impact of Chaperone Activity on Inclusion Body Formation and Protein Folding in the CyDisCo Strain The graph shows 29 molecules of folded protein, 1 molecule of mis-oxidised protein and no inclusion body formation. The units are arbitrary.

5.2.3.1 Graphs showing the impact of changing redox potential

The redox potential must be maintained to allow the Dsb proteins to gain and lose electrons when forming a disulphide bond to fold the protein correctly. The electron carrier within the model is NADH, and the amount of NADH impacts the total protein yield. The rate of Dsb proteins also affects the amount of folded protein produced within the model. The data generated from the model is presented in Figure 5. 15. This test will only look at DsbA, DsbC, and DsbD because the redox balance is assumed to be corrected using these proteins, and DsbB will not impact the test case. Meanwhile, the amount of DsbA, DsbC, and DsbD present impacts the total quantity of protein folded accumulated within the model.

Graph 1 illustrates a 10% increase in DsbA protein levels, while the other Dsb proteins remain unaltered. This condition results in a tenfold increase in mis-oxidised proteins. This outcome is anticipated, as the limitation in the isomerisation rate caused by insufficient DsbC activity leads to a higher proportion of unfolded proteins becoming mis-oxidised. These findings suggest that a balanced expression of Dsb proteins is critical to maximising protein yield and minimising mis-oxidation.

In Graph 2, DsbC levels are increased by 10%, with other Dsb proteins not overexpressed. DsbC plays a pivotal role in correcting mis-oxidised proteins, a role presented in the graph, which shows that with elevated DsbC and the presence of a reducing agent like NADH, no

mis-oxidised proteins are detected. However, the total number of correctly folded protein molecules is limited to 150, indicating that DsbA has become the limiting factor under these conditions. This demonstrates that increasing DsbC reduces mis-oxidation but does not necessarily enhance overall protein folding efficiency if DsbA is not proportionally increased.

Graph 3 depicts the simultaneous 10% increase in DsbC and DsbD levels, eliminating mis-oxidised proteins but reducing the quantity of correctly folded proteins to 43 molecules. This suggests a potential imbalance, where an excess of DsbC and DsbD, without adequate DsbA, leads to suboptimal protein folding outcomes.

Graph 4 continues this observation, showing no mis-oxidised proteins and a reduced amount of folded proteins compared to Graph 2. This implies that even with DsbD overexpression, the protein folding pathway may become constrained, potentially due to the requirement for DsbD to acquire two electrons to reduce disulphide bonds within folded proteins.

Graph 5, where none of the Dsb proteins are overexpressed, shows an accumulation of 93 mis-oxidised proteins and 78 molecules of folded protein. Conversely, in Graph 6, where all Dsb proteins are overexpressed, there is an almost tenfold increase in folded protein molecules, with no accumulation of mis-oxidised protein. This indicates that a balanced overexpression of DsbA, DsbC, and DsbD is necessary to enhance protein yield effectively in this strain.

The impact of NADH as a reducing agent is also critical, as shown in Graph 7. Without NADH, there are no correctly folded proteins and an excess of over 500 molecules of mis-oxidised protein. Conversely, in Graph 8, a 50% increase in reducing power via NADH leads to a substantial increase in folded protein molecules, with no mis-oxidised proteins detected.

These results underscore the importance of maintaining redox balance and the availability of reducing agents like NADH for efficient protein production. The model highlights that while increased DsbA activity is crucial for achieving high protein yields, it must be carefully balanced with DsbC and DsbD levels to prevent the accumulation of mis-oxidised proteins. This balance is essential for optimising the folding and yield of proteins in engineered strains.

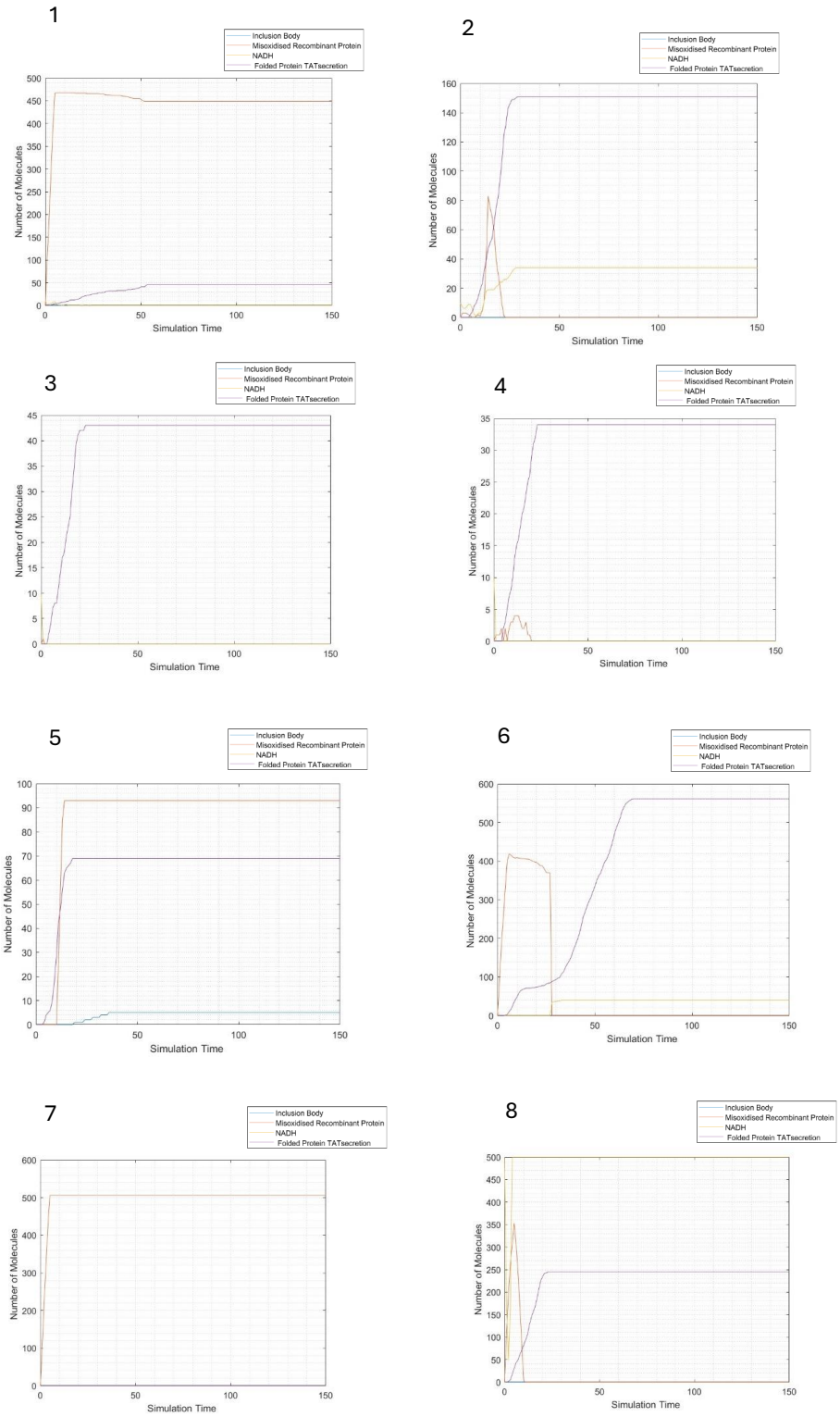


Figure 5. 15 Sensitivity of Folded Protein Production to Redox Potential and Dsb Protein Activity. This figure presents Graphs 1 to 8, analysing redox potential and its effect on protein folding using Petri nets. Graphs 1 to 6 focus on folded protein production sensitivity to Dsb proteins (DsbA, DsbC, DsbD). Specifically, Graphs 1-4 increase the rates of DsbA, DsbC, DsbC and DsbD, and DsbD, respectively, while Graph 5 shows no overexpression, and Graph 6 increases all three proteins. Results indicate that optimal folding occurs with all three Dsb proteins upregulated. Graphs 7 and 8 assess the influence of NADH on protein folding, showing that NADH enhances redox balance, reduces mis-oxidised proteins, and increases folded protein yield. Protein quantities are presented in arbitrary units.

5.2.3.2 Graphs showing the impact of overexpressing molecular chaperones on protein production in response to cytosolic disulphide bond formation

Each parameter within the model was systematically tested to evaluate the impact of cytosolic disulphide bond formation on protein production. The primary molecular chaperones considered in this analysis include Trigger Factor (TF), DnaK, DnaJ, GroEL/ES chaperones, and ClpB. While the effects of Dsb proteins were explored in previous sections, this analysis maintains redox balance and focuses on the influence of overexpressing molecular chaperones on protein folding outcomes.

Figure 5. 16 presents Graphs 1 through 8, which depict the effects of increasing chaperone expression rates on protein folding. In Graph 1, a 20% increase in DnaK chaperone activity results in no inclusion body formation, 52 molecules of correctly folded protein, and 5 molecules of mis-oxidised protein. This outcome highlights the role of DnaK chaperones in maintaining protein folding efficiency and preventing aggregation.

Graph 2 illustrates the consequences of a 20% increase in DnaJ expression without corresponding overexpression of DnaK. This condition leads to the accumulation of 12 molecules in inclusion bodies and a 50% reduction in the number of folded proteins, although no mis-oxidised protein is detected. These results emphasise the centrality of DnaK chaperones in mitigating inclusion body formation, likely due to their role in substrate release and refolding, which DnaJ alone cannot compensate for.

In Graph 3, a 20% increase in ClpB expression results in the formation of 20 inclusion bodies, 32 molecules of folded protein, and no mis-oxidised protein. Although ClpB overexpression produces more folded protein than DnaJ overexpression, it also doubles the number of inclusion bodies, indicating a complex interaction between ClpB and the aggregation-prone intermediates.

Graph 4 explores the simultaneous overexpression of ClpB and DnaJ, which reduces inclusion body formation and decreases the amount of folded protein. While the model predicts this outcome, it is not well-documented in the literature, highlighting a potential area for further experimental investigation. Notably, while the overexpression of these chaperones did not increase total protein yield, it considerably reduced inclusion body formation by 75%, suggesting potential synergies that warrant deeper exploration.

Graph 5 reveals that when Trigger Factor (TF) alone is overexpressed by 20%, inclusion body formation occurs, even without other chaperone overexpression. Chapter 4 details TF's role in protecting nascent polypeptides; however, overexpressing TF beyond the translation rate may lead to aggregation, as observed here. The lack of literature on the effects of TF overexpression beyond the translation rate of a target protein suggests a novel area for further study, particularly to validate the interactions predicted by the model.

In Graph 6, co-overexpression of DnaK and TF by 20% results in the absence of inclusion bodies and a reduction in the number of folded proteins to 7 molecules. This outcome indicates that a delicate balance between these chaperones is essential for improving protein yield, suggesting that their combined overexpression may require fine-tuning to enhance protein production effectively.

GroEL/ES chaperones, which were critical for folding misfolded proteins, were also examined. Graph 7 shows that overexpressing GroEL chaperones alone results in no inclusion body accumulation and the production of 20 molecules of folded protein, comparable to those observed in Graphs 2 and 5. However, when both GroEL and GroES chaperones are overexpressed, as shown in Graph 8, the amount of folded protein doubles to 40 molecules, with no inclusion of body formation and no mis-oxidised protein.

These findings demonstrate the model's capability to elucidate the relationship between chaperone overexpression and protein folding outcomes. While eliminating inclusion bodies is a positive result, the reduced amounts of folded protein observed in specific scenarios suggest complex interactions between chaperones that require further experimental investigation to fully understand and improve the conditions for enhanced protein yield.

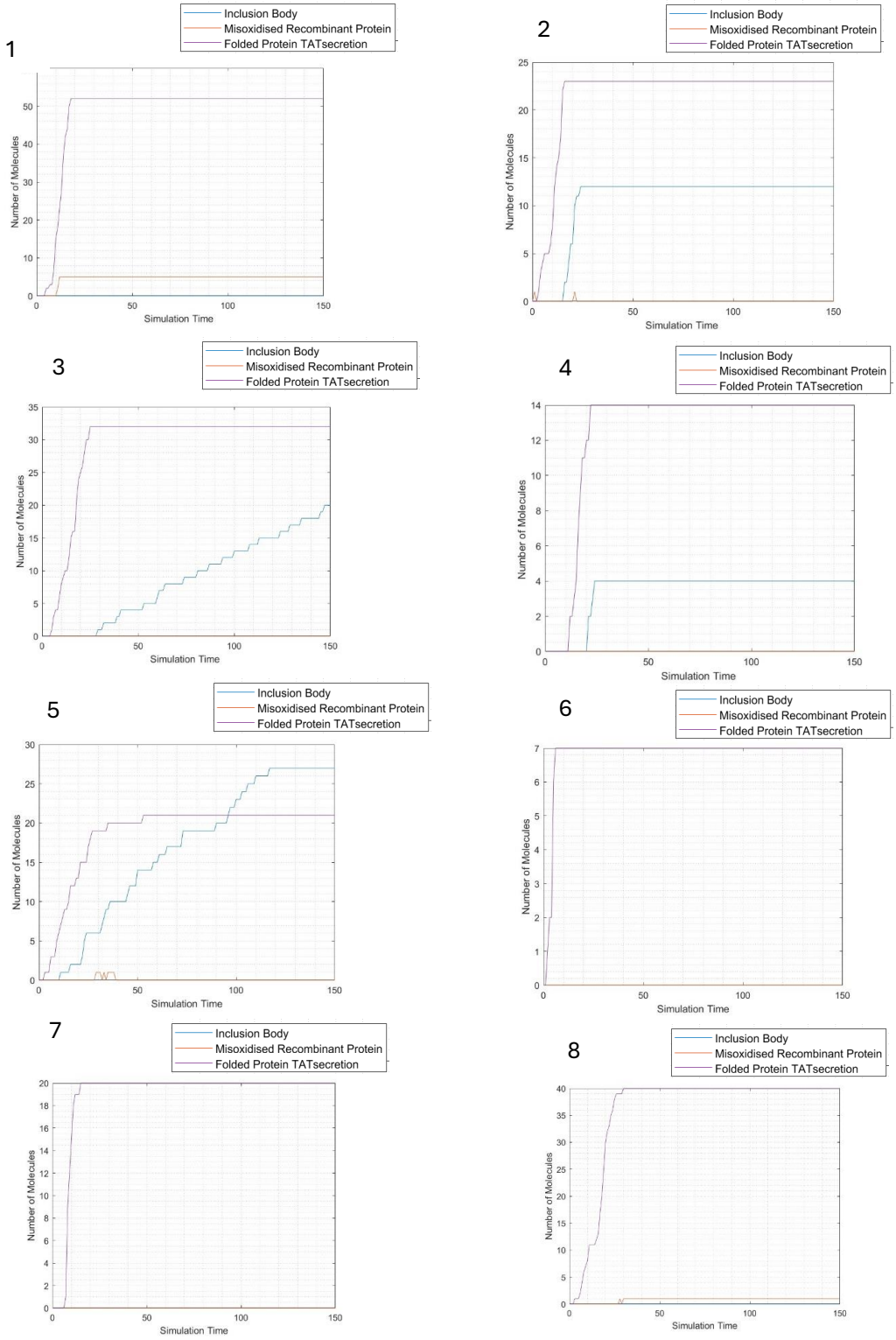


Figure 5. 16 Impact of Chaperone Overexpression on Protein Folding and Misfolding. This figure presents Graphs 1 to 8, generated by the Petri net model, to analyse the impact of chaperone overexpression on protein folding and misfolding. Each graph represents a specific scenario of chaperone upregulation: Graph 1: DnaK expression increased by 20%. Graph 2: DnaJ expression increased by 20%, with DnaK at standard levels. Graph 3: ClpB expression increased by 20%, with all other chaperones at standard levels. Graph 4: ClpB and DnaJ expression increased by 20%.Graph 5: Trigger factor expression increased by 20%. Graph 6: DnaK expression increased by 20%, with the Trigger factor also increased by 20%.Graph 7: GroEL expression increased by 20%.Graph 8: GroEL and GroES expression increased by 20%.The units are arbitrary.

5.2.3.3 Limitations on data availability

CyDisCo technology, designed to facilitate disulphide bond formation in the cytoplasm of *E. coli*, presents several challenges due to limited data in key areas. The success rate of protein expression using CyDisCo varies, with factors such as protein size potentially influencing outcomes. (Castillo-Corujo, Uchida, et al., 2024).

The current model identifies crucial parameters and interactions requiring further experimental validation, such as the effect of Trigger Factor (TF) and DnaK on protein yield. These insights are valuable for guiding experimental design and prioritising data collection efforts. Researchers can conduct more efficient and informative experiments by focusing on these critical interactions, ultimately refining the model to enhance its accuracy and predictive capabilities.

The Petri net model has proven robust in predicting outcomes that are aligned with experimental data and literature. It effectively identifies key parameters affecting protein yield and quality, making it a valuable tool for experimental design and enhancing *E. coli* protein production. The model excels in simulating various conditions and pinpointing critical factors in protein folding pathways. However, it has also highlighted areas needing further investigation, such as the unexplained reduction in folded protein when specific chaperones are combined. This indicates the need for additional research to understand chaperone interactions and redox balance effects better.

5.3 Test case 3: Role of glutathione in the redox potential of the periplasm

5.3.1 Introduction to Test Case 3

It is important to note the crucial role of glutathione (GSH) in maintaining the redox potential within the periplasm. The intricate balance of redox reactions within the periplasm is essential for the cell's survival. Test case 2 demonstrated how engineered strains can effectively maintain redox balance with the cytoplasm.

Another popular commercial strain is called OrigamiTM 2. This strain was designed to allow the mutation of thioredoxin reductase (trxB) and glutathione reductase (gor) genes and improve disulphide bond formation. (Karim et al., 2023; Rauniyar et al., 2022). This test case will examine the role of GSH and how changing its availability impacts the redox balance, subsequently affecting the amount of protein produced.

Electron transfer is fundamental to this redox balance, particularly in reducing disulphide bond proteins. These proteins must undergo a precise sequence of electron transfers to ensure the correct formation of disulphide bonds. Notably, the relationship between glutathione and proteins such as DsbC and DsbD is pivotal in this process. DsbC and DsbD proteins play a critical role in rectifying improperly formed disulphide bonds, necessitating the transfer of electrons from the cytoplasm to correct these errors. (Xi et al., 2023).

This upcoming test case explores the impact of various amounts of GSH, a thiol-containing molecule. The objective is to assess whether the model can simulate the expected outcomes of the redox potential within the periplasm. Defining and discussing the anticipated outcomes in this section can comprehensively evaluate the model's capacity to represent the intricate redox potential within the periplasm.

5.3.2 Glutathione dependence of electron generation for DsbC-DsbD complex

The glutathione dependence of electron generation for the DsbC-DsbD protein complex is fundamental in understanding bacterial protein folding. This system relies on glutathione to transfer electrons, enabling the formation of disulphide bonds crucial for protein stability. Glutathione plays a pivotal role by serving as a primary source of reducing equivalents in bacteria, and its interaction with the DsbC-DsbD complex is essential for the efficient transfer of electrons, ultimately contributing to the correct folding and functionality of bacterial proteins. (Pittman et al., 2005). Understanding this process provides insights into the redox machinery of *E. coli*, which can improve disulphide bond formation and allow the expression of complex proteins.

Before testing the model with reduced rates, it is essential to have a reference point to compare and use data confirming that the trends generated by the model are close to the expected results. A 50% reduction in available glutathione can impact the electron generation for the DsbC-DsbD system in several ways.

Glutathione reductase (Gor) maintains glutathione in its reduced form (GSH), which is crucial for a balanced redox environment. (Allocati et al., 2009; Masip et al., 2006). With 50% less glutathione, the cytoplasm's redox state becomes more oxidised, impairing overall redox homeostasis. (Sies, 1999).

As DsbC activity declines due to an insufficient electron supply, correcting improperly formed disulphide bonds in periplasmic proteins becomes less efficient. This leads to an accumulation of misfolded proteins, which affects cell viability and function.

The study interpreting the trends also considers the impact of DsbA and DsbB proteins when glutathione activity is decreased. (Knoke et al., 2023). This is because the Dsb proteins

operate cyclically between a reduced and oxidative state; the balance between all the Dsb proteins is required for disulphide bond formation. (Andersen et al., 1997; Norambuena et al., 2012).

An 80% reduction in available glutathione severely disrupts the electron transfer mechanism critical for the DsbA-DsbB, DsbC-DsbD system in *E. coli*. This reduction leads to a highly oxidised cytoplasmic environment and impaired thioredoxin, DsbB, and DsbD function, reducing DsbA and DsbC activity.

The consequent accumulation of misfolded proteins triggers intense stress responses and can severely compromise cellular viability. This highlights the essential role of glutathione in maintaining cellular redox homeostasis and the proper functioning of protein folding mechanisms in prokaryotic cells. Understanding these impacts is crucial for comprehending the broader implications of redox imbalances in bacterial physiology. The envelope stress response will not be investigated but will be due to decreased glutathione availability.

The unavailability of glutathione has catastrophic effects on the electron transfer mechanism critical for the Dsb pathways in *E. coli*. The resultant highly oxidised cytoplasmic environment prevents the proper reduction of thioredoxin. This leads to a buildup of misfolded proteins and a severe cellular stress response, likely resulting in cell death. (Ferguson & Booth, 1998).

5.3.3 Results and Discussion

The model was modified to include glutathione reductase (Gor), which is required to reduce TrxA and transfer the electrons to the DsbB and DsbD. In addition, the availability of GSH will be reduced to 50% and 80%. Furthermore, this test case will explore the overexpression of Dsb proteins by 50% for the 80% reduction to test if there are ways to improve folded protein amount when glutathione reductase works at reduced functionality.

5.3.3.1 Difference in glutathione availability

The availability of glutathione influences total protein accumulation and the levels of mis-oxidised proteins within the model. In *E. coli,* glutathione concentration is 1 -10 mM, with 10 mM considered abundant. (Smirnova et al., 2000; Umezawa et al., 2017). In Figure 5.17 the control condition, depicted in Graph 1, an abundance of glutathione is simulated by its external addition to the system boundary. This results in the accumulation of 105 molecules of correctly folded protein with no detectable mis-oxidised proteins, establishing a baseline for optimal conditions.

In Graph 2, the glutathione supply is reduced by 50% to simulate the impact of diminished functionality. Under these conditions, the amount of folded protein decreases to 81

molecules. Despite this reduction, there is still no accumulation of mis-oxidised proteins, suggesting a moderate decrease in glutathione availability does not immediately compromise protein quality.

Graph 3 explores a further 80% reduction in glutathione availability compared to the control. The model shows 102 folded protein molecules are still produced, accompanied by 25 molecules of mis-oxidised protein accumulation. This indicates that reduced glutathione levels do not impact protein folding. This can be justified because electrons are recycled within the system; therefore, some availability is sufficient for protein to fold. However, the risk of mis-oxidation increases.

Graph 4 represents a scenario with no available glutathione, resulting in a complete halt in the production of folded proteins and the accumulation of 131 molecules of mis-oxidised protein. This contrast highlights the critical role of glutathione in maintaining protein folding integrity.

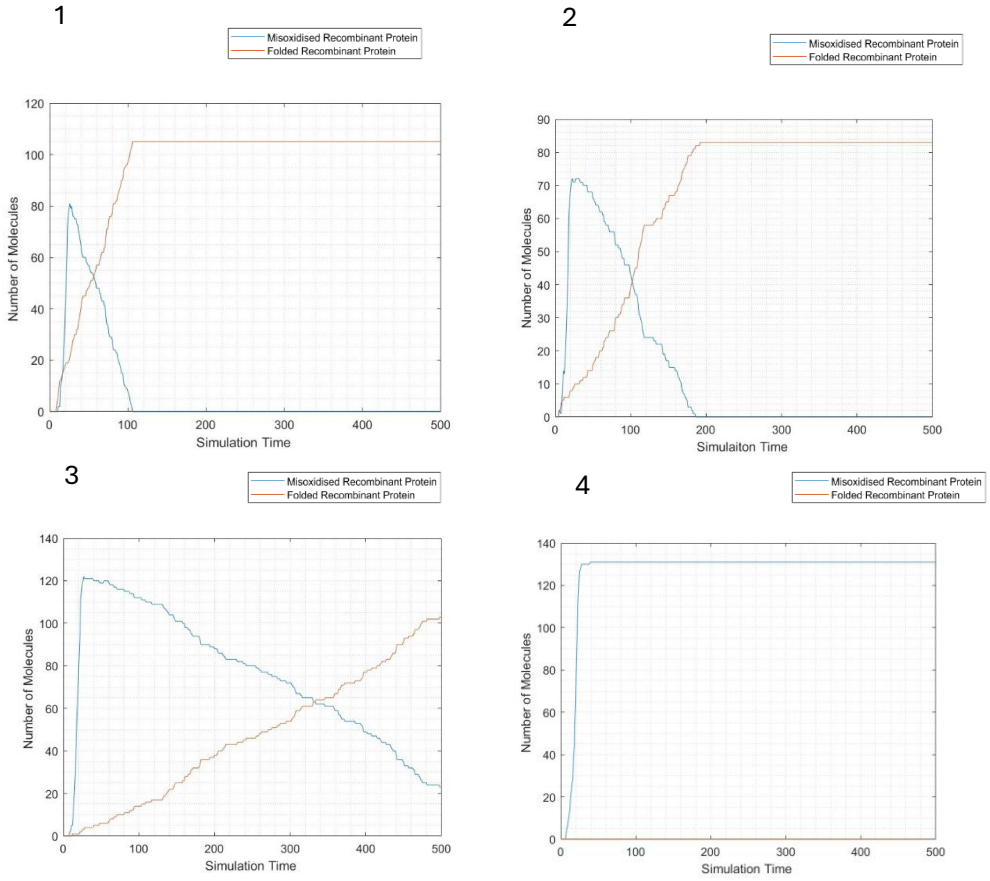


Figure 5. 17 Impact of Glutathione Availability on Folded and Mis-oxidised Protein Production. This figure presents the impact of glutathione availability on the production of folded protein and mis-oxidised protein in a microbial expression system. Graph 1: Glutathione is available in abundance. Graph 2: Glutathione availability is reduced by 50%, Graph 3: Glutathione availability is reduced by 80%. Graph 4: No glutathione is available.

The model demonstrated the impact of reducing glutathione availability and presented its role in forming disulphide bonds by providing the required electrons. An advantage of the Petri net model is the ability to overexpress chaperones to explore how

protein production can be improved when abundant glutathione is unavailable. Figure 5. 18 shows the overexpression of Dsb proteins when glutathione is decreased by 80%. The results are tabulated to compare the amount of folded and mis-oxidised proteins produced.

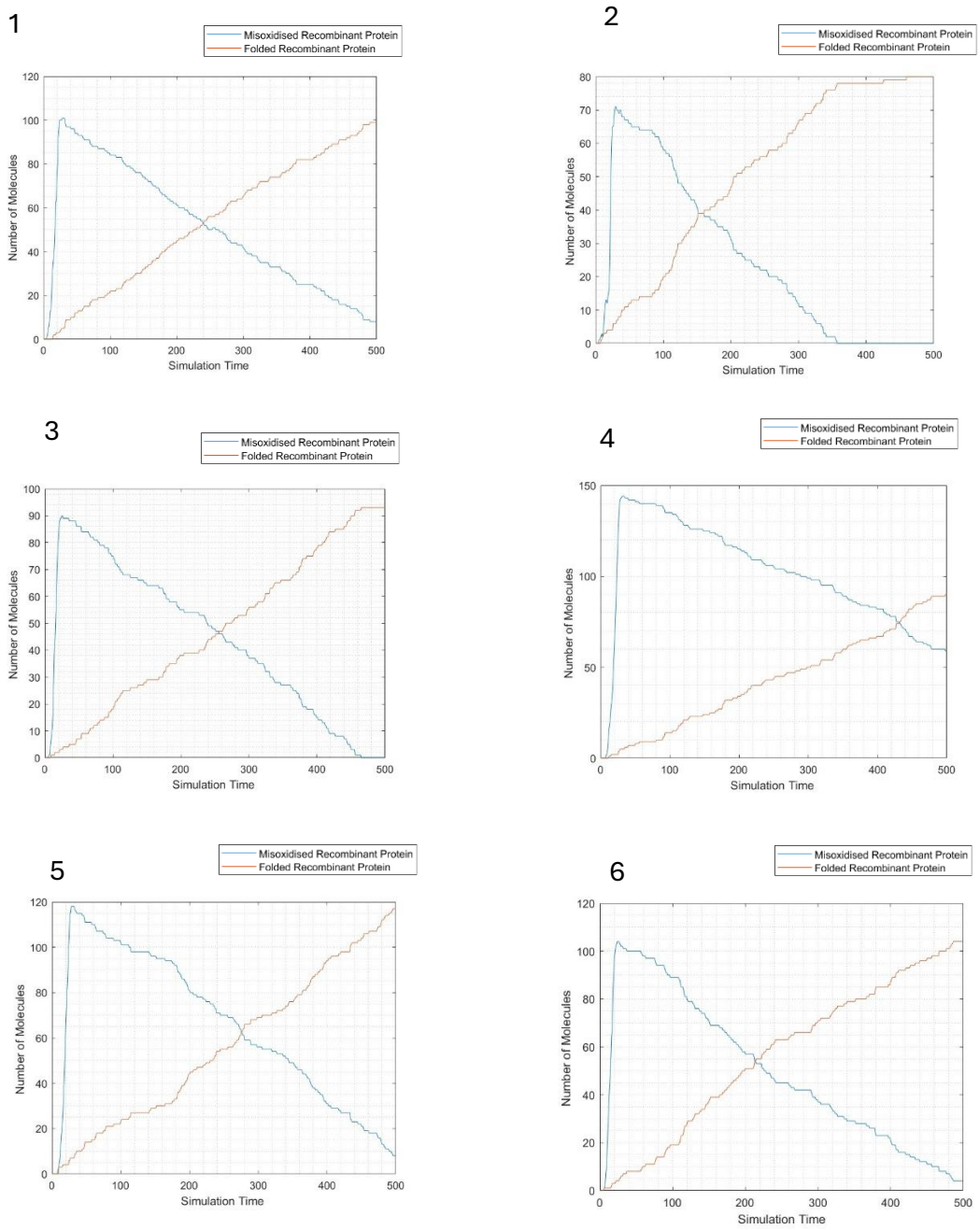


Figure 5. 18 Graphs presenting the impact of overexpressing Dsb proteins when glutathione is decreased by 80%. Graph 1 shows DsbA is overexpressed by 50%; Graph 2 shows DsbB is overexpressed by 50%, and Graph 3 shows both DsbA and DsbB are overexpressed by 50%. Graph 4 shows DsbC is overexpressed by 50%, Graph 5 shows DsbD overexpressed by 50% and Graph 6 shows both DsbC and DsbD overexpressed.

Table 5. 4 Summary of Percentage Changes in Folded and Mis-oxidised Protein Levels for Different Levels of Dsb Protein Overexpression

Graph number and conditions	Number of molecules for folded recombinant protein	Number of molecules for mis-oxidised recombinant protein	Percentage % increased /decreased (+/-) for folded protein	Percentage % increased /decreased (+/-) for mis-oxidised protein
1- DsbA	100	10	0	-67
2- DsbB	80	0	-20	-100
3- DsbA/DsbB	93	0	-5	-100
4- DsbC	90	60	0	+100
5- DsbD	115	5	+20	-50
6- DsbC/DsbD	105	5	0	-50

The results presented in Table 5.4 indicate that DsbD plays a crucial role in maintaining redox balance when glutathione is limited. The overexpression of DsbA and DsbC and combining DsbC with DsbD did not affect the overall protein yield. However, when DsbB was overexpressed, both alone and in combination with DsbA, a reduction in the amount of folded protein was observed compared to the overexpression of DsbA alone.

Notably, there is a 100% increase in mis-oxidised proteins when DsbC is overexpressed, which shows that while overexpression accelerates disulphide bond formation, it may outpace the capacity of other redox proteins, leading to an accumulation of improperly oxidised proteins. Conversely, overexpression of DsbD alone resulted in a 50% reduction in mis-oxidised proteins, underscoring its essential role in ensuring correct disulphide bond formation. Additionally, no mis-oxidised proteins were detected when DsbB was

overexpressed. Overexpression of DsbA alone led to a 67% decrease in mis-oxidised proteins, further emphasising its role in promoting protein folding.

5.3.3.2 Critical evaluation of results

The research findings demonstrate the impact of varying glutathione levels on redox regulation and protein folding in *E. coli*. At 80% glutathione availability, there is a noticeable slowdown in folded protein formation alongside the accumulation of misfolded proteins, indicating partial impairment in redox homeostasis and electron transfer processes. The gradual decline continues at 50% availability, with partial functionality of redox systems and thioredoxin reduction. Complete depletion of glutathione at 0% results in the failure of folded protein formation and a collapse in the protein folding machinery dependent on the DsbC-DsbD system.

The model emphasises Glutathione's essential role in maintaining a reduced cytoplasmic environment necessary for redox-balanced pathways such as disulphide bond formation. It is noted that even at 50% availability, certain cellular functions can still be supported, aligning with the observed partial function of redox systems and thioredoxin reduction. The heavy reliance of the thioredoxin system on glutathione for maintaining its reduced form is also highlighted, consistent with the observed impaired electron transfer to DsbD at 80% glutathione availability. (Smirnova et al., 2015). Furthermore, the complete collapse of the protein folding machinery at 0% glutathione aligns with literature confirming glutathione's critical role in redox-dependent processes. (Ferguson & Booth, 1998).

In summary, the model's analysis of redox dynamics within the *E. coli* periplasm highlights the essential roles of DsbA, DsbB, DsbC, and DsbD in maintaining redox balance and proper protein folding. The findings underscore glutathione's crucial role as a reducing agent in supporting electron transfer and disulphide bond formation, which is vital for cellular viability. The model provides valuable insights into bacterial redox biology by demonstrating the impact of reduced glutathione availability.

Using Petri nets for quantitative analysis of redox regulation and protein folding in *E. coli* allows for a detailed and dynamic understanding of how varying glutathione levels impact cellular processes. The model has simulated different scenarios, providing insights into the critical thresholds and dependencies within the redox system. This approach aligns well with the experimental findings and offers a robust framework for further exploration of bacterial redox biology.

5.4 Test case 4: The impact of the misfunctioning of DsbA and DsbC proteins

5.4.1 Introduction to Test Case 4

This test case will demonstrate the malfunctioning of DsbA and DsbC when HGH is expressed in *E. coli*. The data collected in Chapter 4 did not focus on a specific recombinant protein; therefore, the impact of varying the extent of overexpression of Dsb proteins on the amount of recombinant protein has not been demonstrated using the model. Although reducing agents such as glutathione to maintain the redox balance was highlighted in test case 3, the present test case will specifically focus on DsbA and DsbC, which are present in the periplasm and form complexes with the protein to form the disulphide bond.

The redox balance will be a limiting factor when focusing on the disulphide bond formation pathways. However, for this test case, the redox balance will be maintained by providing an excess reducing agent, allowing the isolation and testing of the impact of DsbA and DsbC as individual proteins.

This test case aims to compare the model's response to expected results. The roles of DsbA and DsbC are interesting since studies have shown that co-expressing DsbA and DsbC in plasmid design improved the yield of difficult-to-express proteins. (Dariushnejad et al., 2019; Rodriguez et al., 2017). The roles of DsbA and DsbC have been studied by deleting these genes to understand their impact on the cell. (Berkmen et al., 2005). Here, the cellular responses to reduced functionality will also be investigated in addition to deletion. This will provide a comprehensive understanding of the roles of DsbA and DsbC within the protein machinery.

5.4.2 The impact of reduced functionality of disulphide bond formation proteins

Disulphide bonds are crucial for numerous proteins' proper folding, stability, and functionality. In *E. coli*, DsbA and DsbC are essential for forming and isomerising these bonds. DsbA, a thiol-disulphide oxidoreductase, introduces disulphide bonds into newly synthesised proteins within the periplasm. When DsbA functionality is compromised, it accumulates unfolded or mis-oxidised proteins with incorrect disulphide bonds, impacting

protein synthesis and folding, particularly for recombinant proteins like HGH (Karshikoff et al., 2013).

DsbC functions as a disulphide isomerase, correcting incorrectly formed disulphide bonds in proteins. Insufficient DsbC activity means that proteins with improper disulphide bonds cannot be refolded appropriately, resulting in an accumulation of mis-oxidised proteins. This reduces functional protein yield and increases protein aggregation risk (Hiniker & Bardwell, 2004).

Existing literature indicates that a reduction in the functionality of disulphide bond-forming proteins, such as DsbA, can lead to a number of proteins remaining mis-oxidised due to incorrect disulphide bond formation (Kouwen et al., 2007; Winter et al., 2000). This impacts the folding and stability of many periplasmic and secreted proteins, which rely on proper disulphide bond formation to achieve their functional conformations (Kobayashi et al., 1997).

Deleting DsbA in *E. coli* impairs protein folding, accumulating mis-oxidised proteins and reducing the efficiency of recombinant protein production (Önder et al., 2008). Similarly, a 50% reduction in DsbC functionality compromises the ability to correct mis-oxidised proteins, resulting in lower yields of desired recombinant proteins due to increased degradation of misfolded products by cellular quality control mechanisms (Dartigalongue et al., 2001). The permanent loss of DsbC functionality in *E. coli* has many implications for protein folding (Vertommen et al., 2008). Without functional DsbC, proteins relying on disulphide bonds are likely to mis-oxidise. This reduces the efficiency of recombinant protein production and impacts overall cell viability and the functionality of proteins necessary for various cellular processes. (Kim et al., 2008). This test case will delete DsbC from the model to investigate the impact on folded and mis-oxidised proteins.

5.4.3 Results and Discussion

The model will test the reported impact of reduced functionality of Dsb proteins. The focus will be on DsbA and Dsb because they interact with the protein of interest during disulphide bond formation. Therefore, the impact of misfunctioning DsbA and DsbC may give new insights into disulphide bond formation in *E. coli*.

5.4.3.1 New data introduced into the Petri net model

This section will outline the approach for establishing the data set used in the Petri net model for test case 4. The data set will include quantitative and qualitative information relevant to test case 4 and input data, as shown in Table 5. 5. It will be structured to align with the requirements of the Petri net model, ensuring that it adequately represents the system under test.

Table 5. 5 Kinetic constants for DsbA and DsbC with 6 different scenarios. (Joly & Swartz, 1997)

Scenario	**K_{cat} (min^{-1})**
DsbA	0.2
DsbC	0.2
DsbA – decreased by 50%	0.1
DsbC – decreased by 50%	0.1
DsbA – non-functional	0.0
DsbC – non-functional	0.0

5.4.3.2 Evaluation of DsbA and DsbC Functionality

This section presents an analysis of simulation results that elucidate the impact of reduced functionality in DsbA and DsbC on the production of correctly folded proteins. It provides insight into how alterations in disulphide bond formation enzymes affect protein synthesis and folding.

Figure 5.19, Graph 1 illustrates the baseline scenario with DsbA and DsbC operating at a kinetic rate of 0.2 min^{-1}, representing typical conditions for HGH production in *E. coli*. Under these conditions, the model predicts the production of 65 molecules of folded proteins, 12 molecules of inclusion bodies, and 3 molecules of mis-oxidised proteins.

Graph 2 shows the effects of a 50% reduction in DsbC functionality, decreasing its kinetic rate to 0.1 min^{-1}. Despite the maintenance of folded protein production at 63 molecules, there is a notable increase in mis-oxidised proteins to 58 molecules and a reduction in inclusion bodies to 5 molecules. This result highlights DsbC's critical role in preventing mis-oxidation, as reduced DsbC functionality allows more unfolded proteins to exit the cytoplasm. However, they form disulphide bonds less efficiently, leading to increased mis-oxidation.

Graph 3 examines the impact of a 50% reduction in DsbA functionality while DsbC remains at 0.2 min^{-1}. Here, folded protein production decreases to 48 molecules, and 40 molecules of mis-oxidised proteins are produced. Notably, there is no accumulation of inclusion bodies, indicating that reduced DsbA functionality does not lead to protein aggregation but affects the efficiency of disulphide bond formation, thereby impacting overall protein folding.

Graph 4 simulates the scenario of non-functional DsbC, resulting in a dramatic increase in mis-oxidised proteins to 141 molecules, nearly a 40-fold increase compared to the scenario shown in Graph 1. Folded protein production is impossible in these conditions, underscoring DsbC's essential role in ensuring proper protein folding and stability.

Graph 5 addresses the permanent malfunction of DsbA, which results in no folded protein production. The absence of mis-oxidised proteins indicates that without functional DsbA, the cell's ability to form disulphide bonds is entirely compromised, leading to an accumulation of unfolded proteins in the periplasm, as depicted in Graph 6, where 150 molecules of unfolded protein accumulate.

These analyses highlight the crucial interplay between DsbA and DsbC in maintaining protein quality control in *E. coli*. They highlight the severe consequences of diminished or absent functionality of these enzymes on protein folding pathways, presenting their importance in biotechnological and industrial applications that depend on efficient recombinant protein production. Understanding these dynamics is essential for developing strategies to improve protein folding processes and enhance cellular productivity in microbial systems.

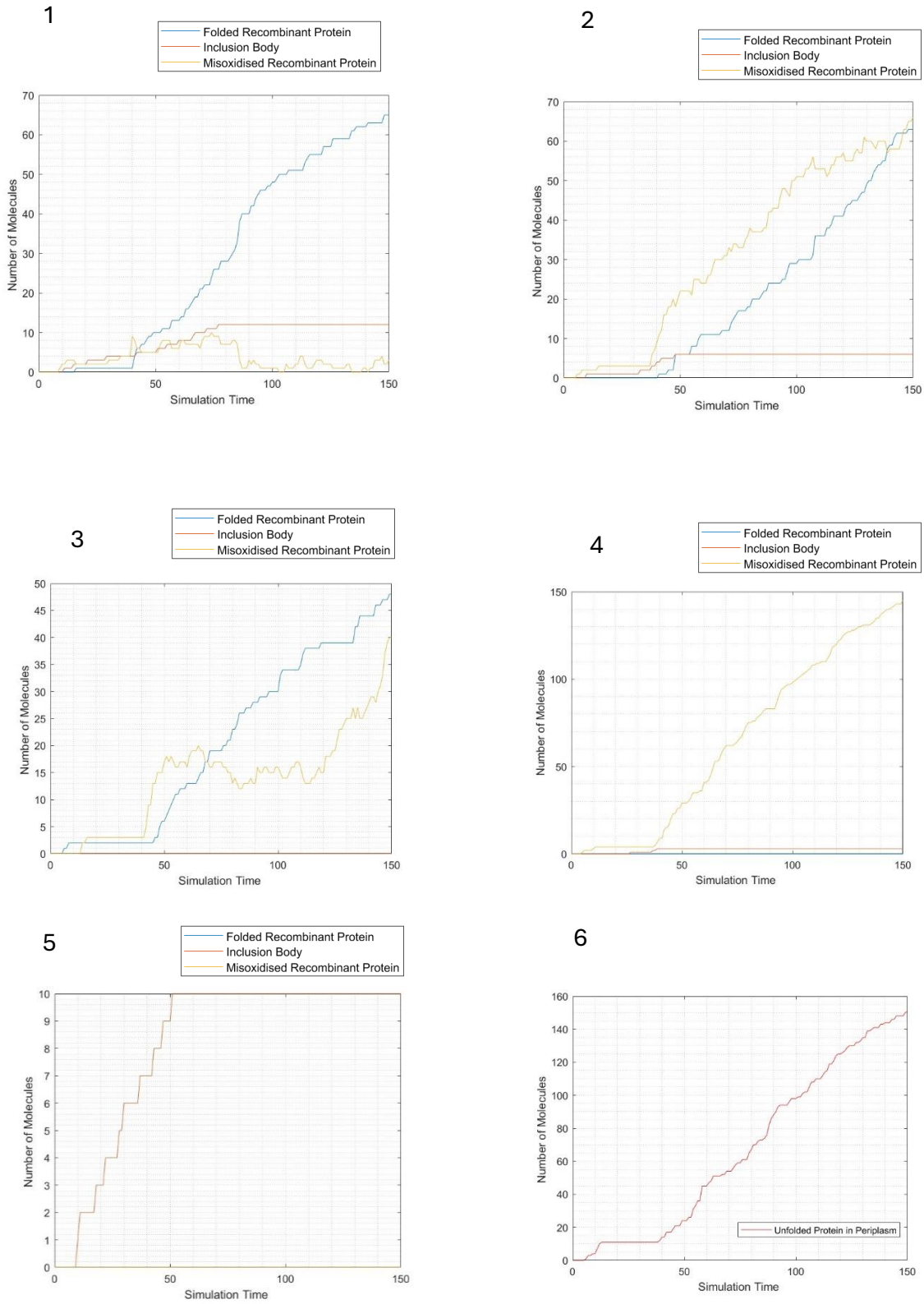


Figure 5. 19 Simulation Results from the Petri net Model of Redox Regulation and Protein Folding in *E. coli*. Graph 1 shows the normal function of DsbA and DsbC under optimal glutathione availability. Graph 2 illustrates the system behaviour when the rate of DsbC is reduced by 50%. Graph 3 depicts the effects of a 50% reduction in the rate of DsbA. Graph 4 represents the deletion of DsbC, leading to impaired protein folding. Graph 5 shows the consequences of DsbA deletion, highlighting its critical role in disulphide bond formation. Graph 6 demonstrates the accumulation of unfolded proteins in the periplasm under conditions of severe redox imbalance. The results highlight the essential roles of DsbA and DsbC in maintaining redox homeostasis and proper protein folding.

This test case highlights the model's strength in demonstrating the role of Dsb proteins in influencing the production of correctly folded proteins. However, it also reveals a limitation due to the scarcity of suitable data for this type of modelling, particularly when compared to other test cases. Despite this, the model provides valuable insights by identifying areas needing further investigation and generating new research questions for experimental support.

The Petri net model offers valuable insights into the dynamics of protein folding and redox regulation in *E. coli*. Simulating various scenarios demonstrates the critical roles of DsbA and DsbC in maintaining protein quality control. While the model provides a robust framework for understanding these processes, its predictions must be corroborated with additional experimental data to refine its accuracy and applicability. This modelling and experimental validation integration is essential for advancing biotechnological applications and improving protein production systems.

5.5 Test case 5: Investigating the overexpression of folding chaperones by decreasing the rate of protein synthesis

5.5.1 Introduction to Test Case 5

The Petri net model was constructed to understand the interactions of molecular chaperones and protein folding. The model protein was used to demonstrate the impact of overexpressing these molecular chaperones on HGH. However, this test case will highlight the relationship between the molecular chaperones and folding outcomes. The test case will investigate the impact of reduced translation rates and how these same chaperone networks can function to improve the amount of protein produced (Chen et al., 2002).

This test case will explore the overexpression of DnaK, DnaJ, GrpE, GroEL and GroES chaperones. The studies used to compare the results from the literature have used strains, which overexpresses one chaperone to understand the impact on the amount of production produced. However, some groups of chaperones will not function unless other groups are present in the right quantities. This is the case for GroEL and GroES chaperones, which required Dnak/DnaJ chaperones to improve the protein amounts. (Haeger et al., 2023; Wagner et al., 2007).

This test case is designed to evaluate the model's capacity to identify chaperones that influence the yield of correctly folded proteins and to assess the effects of chaperone overexpression. Specifically, it aims to determine whether overexpressing chaperones leads to increased protein production without the formation of inclusion bodies or mis-oxidised proteins. Previous research has shown that overexpressing specific chaperones can result in a 42-fold increase in protein yield, though in some instances, overexpression has not affected protein levels. This test will systematically examine various chaperone combinations to assess their collective impact on protein production and folding efficiency.

5.5.2 Overexpression of chaperones and reduced rate of translation

Molecular chaperones such as DnaK are overexpressed to assist folding. The first scenario will increase the folding rate by 50%, and this will be compared with the control scenario, where no rates on any pathways were increased, and no chaperones were overexpressed. In addition, DnaK, DnaJ, GrpE, GroEL, and GroES chaperones will be expressed and compared to the control. This test will investigate the impact of increasing the folding rate and whether overexpressing chaperones can produce more molecules of folded protein than increasing the folding rate, which is more complicated in practice. A second test will explore the overexpressing of multiple chaperones during protein translation, which will be reduced by 20%, 50%, and 80%. This scenario aims to understand if groups of overexpressed chaperons make a difference in the amount of folded protein produced when coupled with slowed-down translation.

Unfolded HGH protein is secreted into the periplasmic space, where disulphide bonds form to produce the folded recombinant protein. To enhance protein production, the ideal scenario would be to increase the folding rate, which can be achieved in this model; however, the practical equivalent is more complex (Nguyen et al., 2014). Overexpressing chaperones presents a straightforward method to enhance protein production in practice. The aim of increasing the protein folding rate is to compare this with the overexpression of chaperones to understand how chaperones can improve the production of folded protein. To simulate this, the transport rate to the periplasm and disulphide bond formation will be increased by 50% in the model.

Cellular homeostasis and proteostasis are crucial for the survival and functionality of all living organisms. The protein folding, modification, and degradation processes are tightly regulated within the cell. One of the primary strategies cells utilise to maintain proteostasis under stressful conditions is the overexpression of chaperones (Fauvet et al., 2019; Zuily et al., 2022). Chaperones are specialised proteins that assist in adequately folding and refolding nascent and stress-damaged proteins. Heat shock proteins (HSPs) are a well-known family of chaperones that are upregulated in response to stress, such as elevated temperatures or oxidative stress. HSP70, for instance, has been extensively studied for its

role in mitigating protein misfolding and aggregation, which can otherwise lead to cellular dysfunction and diseases. The upregulation of HSPs, including Hsp70, highlights their critical role in maintaining protein quality control under stress conditions (Morano, 2007).

Cells employ multiple strategies to manage protein folding stress. In addition to chaperone overexpression, cells utilise translation attenuation to reduce the influx of nascent proteins requiring folding. This decreases the burden on protein quality control systems. Translation attenuation is crucial during stress conditions when the cellular machinery is overwhelmed with misfolded proteins. The integrated stress response is a mechanism that temporarily reduces global protein synthesis while selectively translating stress-responsive proteins to aid recovery and adaptation. These strategies collectively help cells cope with protein folding stress.(Holcik & Sonenberg, 2005).

5.5.3 Results and Discussion

This section will present the data generated by the model, focusing on understanding the impact of overexpressing individual chaperones on the amount of protein produced. Figure 5. 20 illustrates the overexpression of each chaperone individually, providing insights that will inform the selection of chaperone combinations for testing the reduced transaction rate.

The data collected from the literature presents an expected trend as the translation rate decreases; the amount of protein generated from the model will also decrease when the protein folding pathways are maintained.

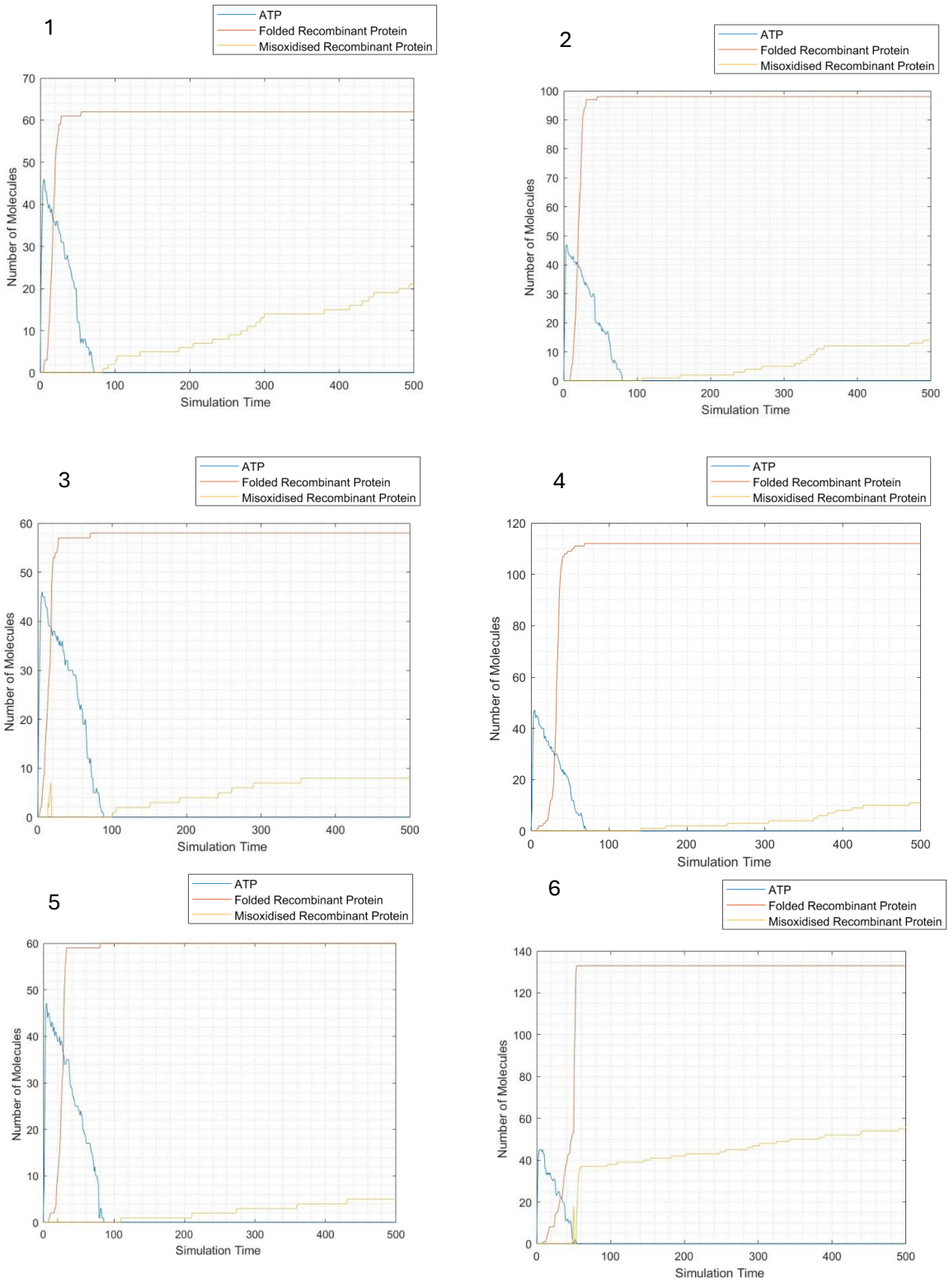


Figure 5. 20 Simulation Results of Protein Folding Dynamics Under Varied Cellular Conditions. Graph 1 illustrates the baseline scenario, depicting the production of folded proteins, ATP depletion, and mis-oxidized proteins under standard physiological conditions. Graph 2 demonstrates the effects of a 50% increase in the folded protein production rate, highlighting changes in system behaviour. Graph 3 examines the impact of 50% overexpression of Trigger factor, a key chaperone involved in co-translational folding. Graph 4 explores the consequences of 50% overexpression of DnaK, a central component of the Hsp70 chaperone system. Graph 5 analyses the effects of 50% overexpression of DnaJ, a co-chaperone that regulates DnaK activity. Graph 6 evaluates the system response to 50% overexpression of the GroEL-GroES chaperonin complex, which facilitates post-translational protein folding.

In Figure 5. 20, Graph 1 shows the results when the model is set at standard conditions, which means there are no increased rates or overexpressed molecular chaperones. 62 molecules of folded protein and 21 molecules of mis-oxidised protein are produced. The protein folding rate is increased by 50% within the model, which increases the number of folded protein molecules to 98, and the mis-oxidised protein is decreased to 14 molecules in Graph 2. This test case aims to highlight the impact of overexpressing chaperones. Although increasing the folding rate has shown improvements in the number of folded proteins produced by 67%, this test case aims to understand the impact of overexpressing molecular chaperones and if this impacts the amount of folded protein produced.

Graph 3 was generated, where only the TF chaperone is overexpressed. This results in no change in the amount of folded protein molecules. Meanwhile, the mis-oxidised protein decreases by 80%. Graph 4 presents a 50% overexpression of DnaK, where the protein amount increases to 110, and the mis-oxidised protein is at 10 molecules. Compared to Graph 1, there is an 83% increase in folded proteins and a 40% decrease in mis-oxidised protein.

Graph 5 presents an overexpression of DnaJ by 50%, which does not impact the amount of protein generated compared to Graph 1. However, mis-oxidised protein is reduced by 80%. Graph 6 presents the overexpression of GroEL and GroES chaperones, increasing the number of folded protein molecules to 130 and the mis-oxidised protein to 55 molecules. This overexpression increased the folded and mis-oxidised proteins by two-fold compared to the data in Graph 1.

The data collected in Figure 5. 20 demonstrates which chaperons impact the amount of folded protein. DnaK and GroEL/ES chaperone systems can double the amount of folded protein.

Figure 5. 21 will focus on combining chaperones and individually testing DnaK and GroEL/ES chaperones for each scenario.

In Figure 5. 21, protein translation is reduced by 20%. Graph 1 shows that 70 molecules of folded protein are generated, 15 molecules of misfolded protein are generated, and 13 molecules of mis-oxidised protein are generated. Graph 2 shows the overexpression of DnaK only, where the number of molecules for folded protein increases to 120. In contrast, the number of mis-oxidised proteins and misfolded proteins has reduced to 0 and 22 molecules, respectively.

The data presented in Figures 5. 21, 5. 22 and 5.23 shows the chaperones that were overexpressed for each scenario and their impact on the total amount of folded protein and misfolded protein produced. The results do not follow a set pattern; one chaperone always produces the most protein when overexpressed. This is also expected in a biological scenario, where scenarios are varied, and the requirement to improve the yield is generally not the same.

The first scenario presents a 20% reduction of protein translation, which results in 70 molecules of folded protein, 20 molecules of misfolded protein, and 18 molecules of mis-oxidised protein. When the DnaK chaperone is overexpressed, the protein amount improves by 71%, whereas the amount of misfolded protein is reduced to 0 molecules. However, there is an increase in mix-oxidised protein by 22%.

When DnaK and DnaJ are overexpressed, the amount of folded protein improves by 42%, less than DnaK. This suggests that an increase in DnaJ expression will not improve the folded protein amount for this scenario. In addition, the amount of misfolded protein was reduced by 25%, which is lower than DnaK expression alone.

When DnaK, DnaJ, and GrpE chaperones are overexpressed, the folded protein amount improves by only 7%, and the amount of misfolded protein does not change in this scenario. However, the amount of mis-oxidised protein is reduced by 44%. This shows that combining DnaK, DnaJ, and GrpE chaperones cannot improve the amount of folded protein produced, and overexpressing many chaperones can lead to inefficient systems.

GroEL and GroES chaperons are the second most common chaperone set to be overexpressed to improve protein yields. (Gill et al., 2000; Kovács et al., 2001; Widersten, 1998)To understand the influence of GroEL and GroES chaperones on the amount of protein produced, this group of chaperones was overexpressed alone and in a group with all the other chaperones. When GroEL and GroES are overexpressed alone, the folded protein amount improves to 71%, similar to DnaK overexpression.

In addition, the amount of mis-oxidised protein was reduced by 44%, whereas the overexpression of DnaK could not reduce the amount of mis-oxidised protein. When GroEL and GroES chaperone is overexpressed with DnaK, DnaJ, and GrpE, the protein produced improves to 71%. However, mis-oxidised protein increases by 11%.

The model shows that overexpressing a number of chaperones is not beneficial to the production system in this scenario, and overexpressing GroEL and GroES chaperones is sufficient.

In the following scenario, where 50% of protein translation is reduced, the impact of chaperone systems on improving the amount of folded protein is very different from the first scenario. Where 45 molecules of folded protein are produced, 15 molecules of misfolded protein and 20 molecules of mis-oxidised protein. When DnaK is overexpressed, the amount of folded protein increases by 22%, the amount of misfolded protein reduces by 67%, and the amount of mis-oxidised protein reduced by 25%. However, when DnaK is overexpressed with DnaJ, the folded protein improves further to 33%. This shows that DnaJ can be more influential in improving the folded protein amount. In addition, the mis-oxidised protein amount was reduced to 35%. However, the misfolded protein is reduced to only 33%.

When GrpE is overexpressed with DnaK and DnaJ, the amount of folded protein increases to 55%, a decrease of 50% of mis-oxidised protein, and no accumulation of misfolded

protein. This shows that overexpressing these three chaperones will improve the mound of folded protein and misfolded protein for this particular scenario.

However, when GroEL and GroES chaperones are overexpressed with DnaK, DnaJ and GrpE, the amount of folded protein only improves to 29%, and the amount of mis-oxides doubles. At the same time, misfolded protein did not accumulate. This presents overexpression of all the chaperones when protein transition is reduced to 50%, which is insufficient for improved protein expression.

When GroEL and GroES chaperones are overexpressed only for this scenario, the amount of protein improves by only 11%. The literature has shown that overexpressing GroEL and GroES at all times has not improved the protein yield. (Martínez-Alonso et al., 2010). This shows that the model can present which chaperone group will improve the amount of folded protein before any experimental work, which will aid the design of experiments.

In the final scenario, protein translation is reduced by 80%, resulting in 14 molecules of folded protein, 7 molecules of misfolded protein, and 18 molecules of mis-oxidised protein. When DnaK is overexpressed, folded protein increases by 79%, and misfolded protein decreases by 87%. However, the amount of mis-oxidised protein does not change.

When DnaJ is expressed with DnaK, the folded protein amount only improves by 36%, but the amount of mis-oxidised protein is reduced by 56% and the misfolded protein by 71%. Although the amount of folded protein has not improved, the reduced mis-oxidised protein is desirable. When GrpE is overexpressed with DnaK and DnaJ, the amount of folded protein improves by 29% only, while the amount of misfolded and mis-oxidised proteins does not improve compared to DnaJ and DnaK overexpression only.

However, when all the chaperones are overexpressed, the folded protein amount improves by 79% compared to the control and the amount of mis-oxidised protein reduces by 44%. In contrast, there is no accumulation of misfolded protein. When GroEL and GroES chaperones are overexpressed, only the folded protein amount improves by 7%. Therefore, in this scenario, the overexpression of all 5 champions results in improved protein folded amounts.

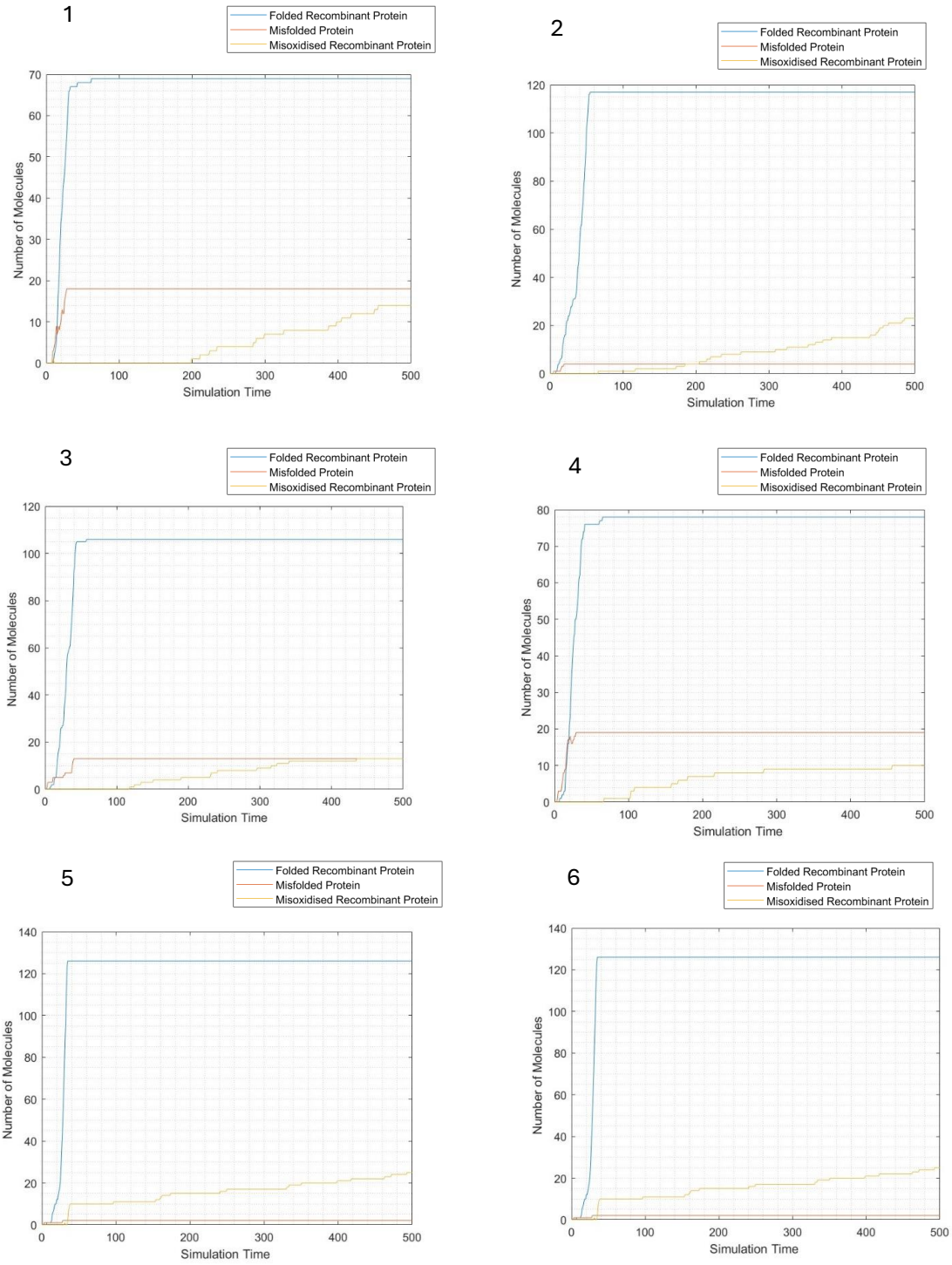


Figure 5. 21 Simulation Results of Protein Folding Dynamics Under Reduced Translation Rates and Chaperone Overexpression. The Petri net model simulates the effects of a 20% reduction in protein translation rates on protein folding efficiency. Graph 1 represents the control scenario, where no chaperones are overexpressed, establishing a baseline for comparison. Graph 2 demonstrates the impact of overexpressing DnaK. Graph 3 examines the combined overexpression of DnaK and DnaJ, highlighting their synergistic role in protein folding. Graph 4 explores the effects of overexpressing the DnaK/DnaJ/GrpE chaperone system, which enhances the ATPase activity of DnaK. Graph 5 evaluates the system response to the overexpression of the DnaK/DnaJ/GrpE complex alongside the GroEL/ES chaperonin system, representing a comprehensive chaperone network. Graph 6 focuses on the overexpression of GroEL/ES chaperonins only, illustrating their independent contribution to protein folding.

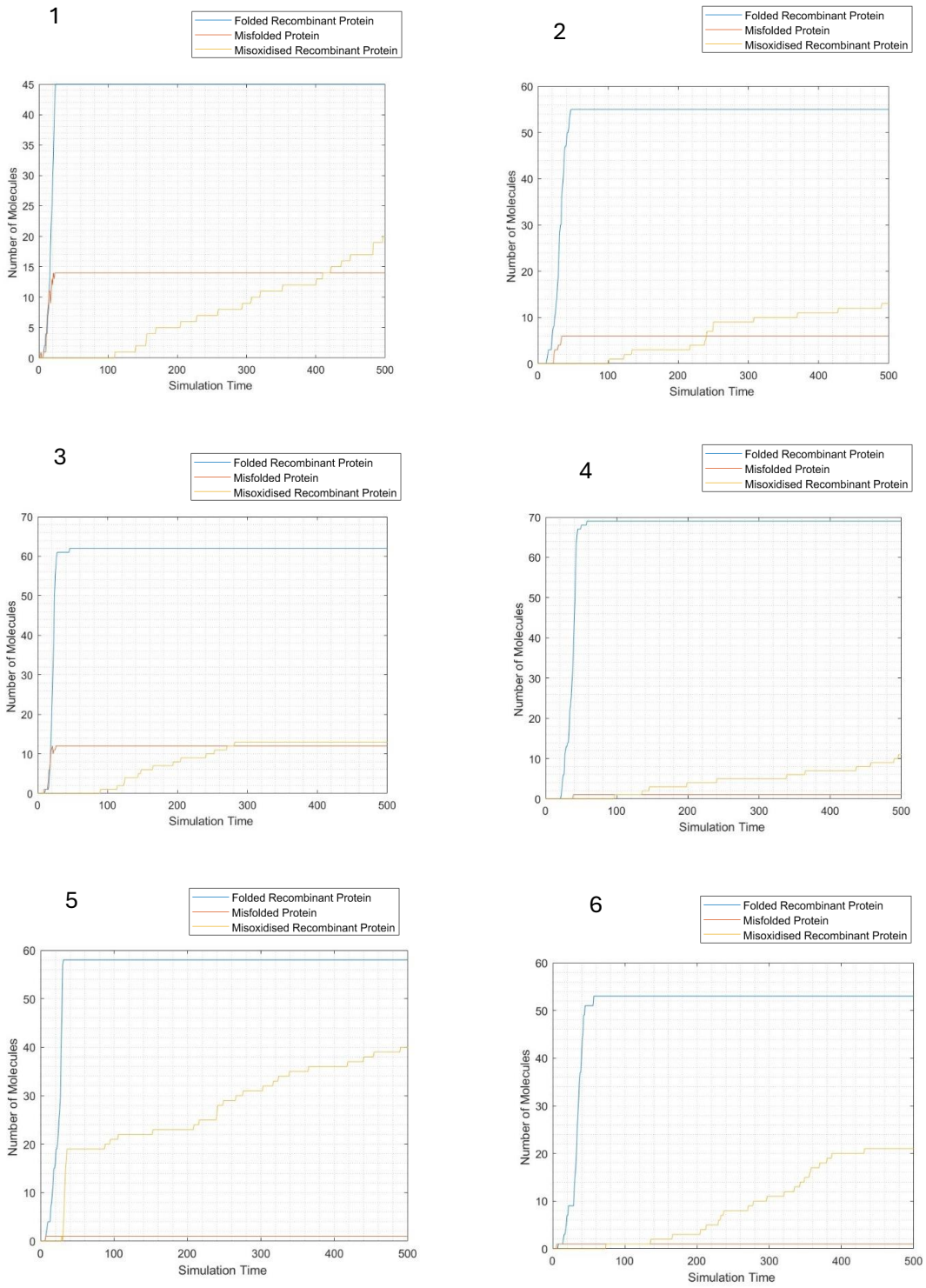


Figure 5. 22 Simulation Results of Protein Folding Dynamics Under Reduced Translation Rates and Chaperone Overexpression. The Petri net model simulates the effects of a 50% reduction in protein translation rates on protein folding efficiency. Graph 1 represents the control scenario, where no chaperones are overexpressed, establishing a baseline for comparison. Graph 2 demonstrates the impact of overexpressing DnaK. Graph 3 examines the combined overexpression of DnaK and DnaJ, highlighting their cooperative role in substrate binding and processing. Graph 4 explores the effects of overexpressing the DnaK/DnaJ/GrpE chaperone system, which enhances the ATPase cycle of DnaK. Graph 5 evaluates the system response to the overexpression of the DnaK/DnaJ/GrpE complex alongside the GroEL/ES chaperonin system, representing a comprehensive chaperone network. Graph 6 focuses on the overexpression of GroEL/ES chaperonins, illustrating their independent contribution to protein folding under stress conditions.

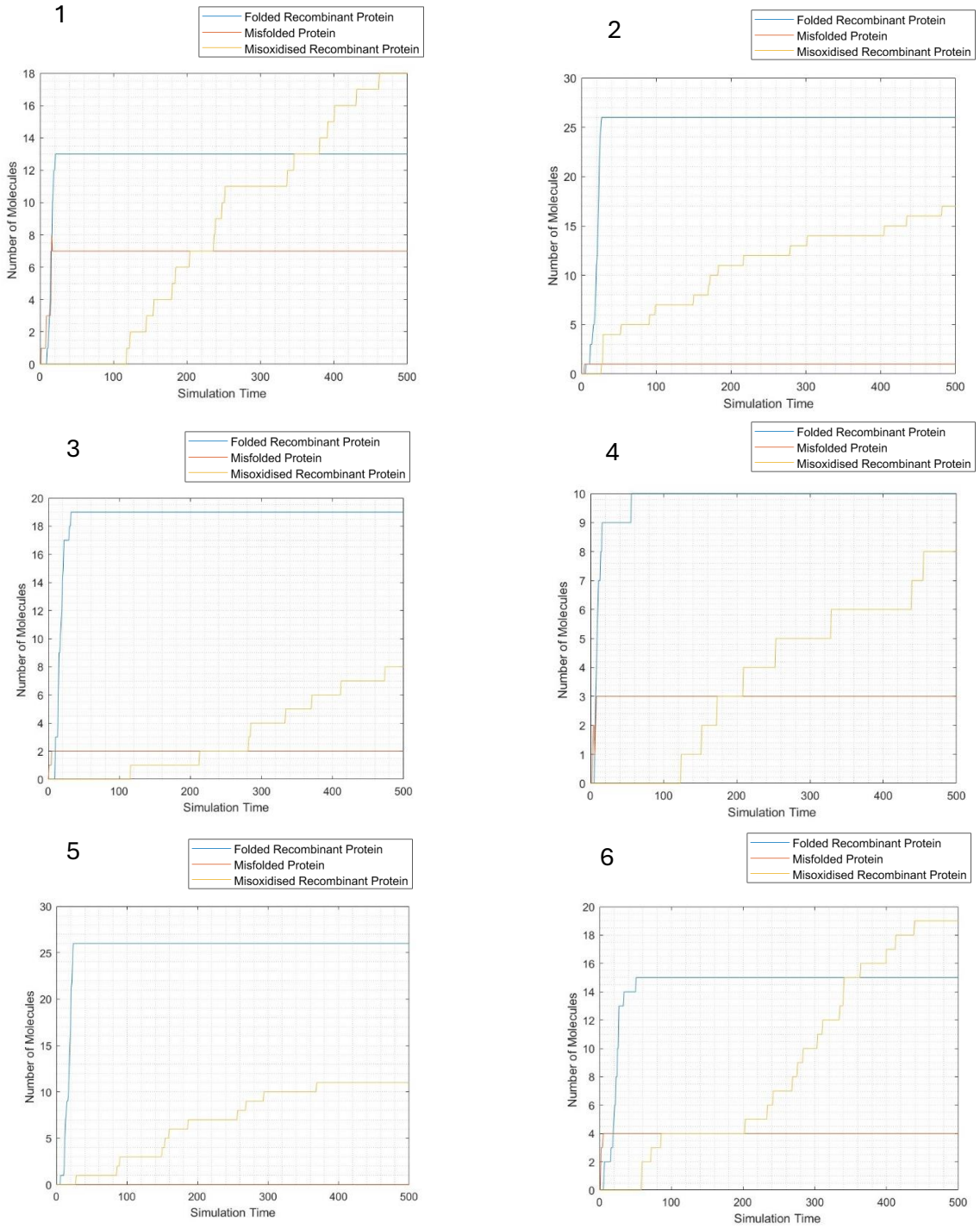


Figure 5. 23 Simulation Results of Protein Folding Dynamics Under Reduced Translation Rates and Chaperone Overexpression. The Petri net model simulates the effects of an 80% reduction in protein translation rates on protein folding efficiency. Graph 1 represents the control scenario, where no chaperones are overexpressed, establishing a baseline for comparison. Graph 2 demonstrates the impact of overexpressing DnaK. Graph 3 examines the combined overexpression of DnaK and DnaJ, highlighting their cooperative role in substrate binding and processing. Graph 4 explores the effects of overexpressing the DnaK/DnaJ/GrpE chaperone system, which enhances the ATPase cycle of DnaK. Graph 5 evaluates the system response to the overexpression of the DnaK/DnaJ/GrpE complex alongside the GroEL/ES chaperonin system, representing a comprehensive chaperone network. Graph 6 focuses on the overexpression of GroEL/ES chaperonins, illustrating their independent contribution to protein folding under stress conditions.

5.5.3.1 Data collected from the model to present the impact of reduced rate of translation and overexpression of chaperones

Tables 5.6, 5.7, and 5.8 present the data generated by the Petri net model for reduced translation rates of 20%, 50%, and 80%, respectively, under conditions where specific groups of chaperones are overexpressed. The percentage increase or decrease in folded protein production is tabulated for each scenario, highlighting the impact of chaperone overexpression as translation rates are progressively reduced. These results demonstrate the Petri net model's ability to simulate complex protein folding dynamics and provide quantitative insights into the roles of chaperones under varying cellular conditions.

Table 5.6 presents a 71% increase in folded protein is observed when DnaK is overexpressed alone, when GroEL/ES chaperonins are overexpressed alone, or when the DnaK/DnaJ/GrpE/GroEL/ES chaperone system is overexpressed together. However, overexpression of DnaK and DnaJ together results in only a 7% increase in folded protein. Notably, while DnaK overexpression increases folded protein by 71%, it also leads to a 22% increase in mis-oxidised proteins, highlighting a trade-off between folding efficiency and oxidative stress.

Table 5.7 shows that overexpression of the DnaK/DnaJ/GrpE system results in a 55% increase in folded protein. In contrast, overexpression of GroEL/ES alone yields only an 11% increase in folded protein. This demonstrates that the DnaK/DnaJ/GrpE system is more effective than GroEL/ES alone under moderate translation stress.

Table 5.8 shows that when DnaK is overexpressed alone or when the DnaK/DnaJ/GrpE/GroEL/ES system is overexpressed together, a 79% increase in folded protein is observed. However, overexpression of GroEL/ES alone results in only a 7% increase in folded protein. This indicates that DnaK is critical in maintaining protein folding under severe translation stress, while GroEL/ES alone is less effective.

The Petri net model simulates the interplay between chaperone systems and protein folding efficiency under varying translation rates. It demonstrates that not all chaperone combinations improve folded protein production equally, and the optimal chaperone system depends on the specific cellular conditions.

For example, Table 5.6 shows that while DnaK overexpression increases folded protein, it also elevates mis-oxidised proteins, suggesting a need to balance folding efficiency with oxidative stress management.

The model’s ability to simulate these scenarios provides valuable insights for biotechnology decision-making, such as designing expression vectors and selecting target chaperones for recombinant protein production. The model’s quantitative and dynamic nature allows for evaluating multiple scenarios, including the effects of chaperone overexpression and

translation rate reductions. It identifies the most favourable chaperone combinations for specific conditions, such as the effectiveness of the DnaK/DnaJ/GrpE system at a 50% translation rate reduction and the critical role of DnaK under severe translation stress (80% reduction). The model also highlights trade-offs, such as increased mis-oxidised proteins under certain conditions, which can guide experimental design.

The Petri net model provides a robust framework for analysing protein folding dynamics under stress conditions. The model identifies key chaperone systems for specific scenarios and highlights potential trade-offs by simulating the effects of chaperone overexpression and reduced translation rates. These insights are invaluable for biotechnological applications, such as improving recombinant protein production and designing cellular systems for enhanced protein folding efficiency. The model's ability to generate quantitative data under diverse conditions underscores its utility for theoretical analysis and practical decision-making.

Table 5. 6 Data collected from simulation when protein translation is reduced by 20%

Graph number for 20% reduction of translation and overexpression of chaperones by 50%	**Number of molecules for folded protein**	**Number of molecules for misfolded protein**	**Number of molecules for mis-oxidised protein**	**Percentage % increased /decreased (+/-) from Graph 1 for folded protein**	**Percentage % increased/ decreased (+/-) from Graph 1 for misfolded protein**	**Percentage % increased /decreased (+/-) from Graph 1 for mis-oxidised protein**
1–No overexpression	70	18	14	-	-	-
2 – DnaK	118	0	25	+71	-100	+22
3 – Dnak/DnaJ	105	13	13	+42	-25	+11
4 –DnaK/DnaJ /GrpE	78	19	10	+7	0	-44
5- DnaK/DnaJ GrpE/GroEL/ES	126	0	25	+71	-100	+11
6- GroEL/ES	126	0	10	+71	-100	-44

Table 5. Data collected from simulation when protein translation is reduced by 50%

Graph number for 50% reduction of translation and overexpression of chaperones by 50%	**Number of molecules for folded protein**	**Number of molecules for misfolded protein**	**Number of molecules for mis-oxidised protein**	**Percentage % increased /decreased (+/-) from Graph 1 for folded protein**	**Percentage % increased/ decreased (+/-) from Graph 1 for misfolded protein**	**Percentage % increased /decreased (+/-) from Graph 1 for mis-oxidised protein**
1–No overexpression	45	14	20	-	-	-
2 – DnaK	55	6	13	+22	-67	-25
3 – Dnak/DnaJ	62	12	13	+33	-33	-35

4 –DnaK/DnaJ /GrpE	70	0	10	+55	-100	-50
5- DnaK/DnaJ GrpE/GroEL/ES	58	0	40	+29	-100	+100
6- GroEL/ES	53	0	23	+11	-100	+15

Table 5. 7 Data collected from simulation when protein translation is reduced by 80%

Graph number for 80% reduction of translation and overexpression of chaperones by 50%	**Number of molecules for folded protein**	**Number of molecules for misfolded protein**	**Number of molecules for mis-oxidised protein**	**Percentage % increased /decreased (+/-) from Graph 1 for folded protein**	**Percentage % increased/ decreased (+/-) from Graph 1 for misfolded protein**	**Percentage % increased /decreased (+/-) from Graph 1 for mis-oxidised protein**
1–No overexpression	13	7	18	-	-	-
2 – DnaK	26	1	17	+79	-87	0
3 – Dnak/DnaJ	19	2	8	+36	-71	-56
4 –DnaK/DnaJ /GrpE	10	3	8	+29	-57	-56
5- DnaK/DnaJ GrpE/GroEL/ES	26	0	11	+79	-100	-44
6- GroEL/ES	15	4	19	+7	-43	+5

5.6 Summary

The model has tested various cases that highlight different parts of the model. Test case 1 analyses how signal peptide hydrophobicity affects protein translocation in *E. coli*. Highly hydrophobic signal peptides lead to more efficient translocation, resulting in successful protein translocation without consuming ATP. The study also explores the outcomes of various signal peptides on protein translocation pathways and provides valuable insights for protein translocation studies and plasmid design for recombinant protein production. The model's impact is evident in its ability to predict the translocation pathway of a protein of interest based on its sequence, which can aid in plasmid construction.

In test case 2, the model's capability to simulate the impact of cytosolic disulphide bond formation on protein production is instrumental in aiding strain development decisions even before conducting experiments. Using the model to assess the potential effects of altering chaperone rates and other parameters, researchers can predict how different modifications may influence protein yield, quality, and inclusion body formation in specific *E. coli* strains. This predictive ability equips researchers and biomanufacturers with valuable insights that can guide decisions related to strain development, ultimately streamlining the process and contributing to more efficient and effective recombinant protein production.

In test case 3, the role of glutathione in maintaining redox potential within the periplasm and its impact on protein production was investigated. The model's strength lies in its comprehensive exploration of the relationship between glutathione, redox balance, and protein folding mechanisms.

Test case 4 focuses on the impact of malfunctioning DsbA and DsbC on the expression of human growth hormone (Hgh) in *E. coli*. It emphasises the importance of maintaining redox balance and provides a detailed understanding of the roles of DsbA and DsbC. The model's strength lies in predicting the response to reduced functionality or deletion of DsbA and DsbC, providing insights into protein folding, stability, and overall cellular health *in E. coli.*

Test case 5 evaluates the impact of overexpressing different chaperones on the production of folded and misfolded proteins. The model's strength is its ability to predict the influence of specific chaperone groups on the production of folded proteins, which can potentially aid in designing experiments. Additionally, it highlights the intricate and non-linear relationships between chaperone overexpression and protein yields, providing crucial information for protein production systems in *E. coli.*

Chapter 6

6. Conclusion and future work

Biotechnology advancements have optimised *E. coli* for recombinant protein production, yet challenges persist, particularly in achieving high yields of correctly folded proteins. Key issues include bottlenecks in disulphide bond formation and the efficiency of translocation pathways. The development of mathematical models has made notable progress in addressing these challenges. Predictive tools are now utilised from early-stage developments of the protein of interest to purification.

In plasmid design, overexpression of molecular chaperones, translocation chaperones, and Dsb proteins are frequently employed to enhance product yield. However, this approach does not consistently result in improved yields. This research presents a novel application of a Petri net framework for integrating biochemical pathways within *E. coli*, aimed at elucidating the interactions between these pathways. This innovative framework models critical pathways, including protein synthesis, translocation, and disulphide bond formation, providing a comprehensive in silico tool to address key challenges in the bioprocessing industry. By simulating biological scenarios, the Petri net model offers a detailed understanding of molecular chaperones and proteins of interest, thereby supporting the design of more targeted and effective experimental strategies.

Chapter 3 comprehensively evaluates the existing literature and identifies gaps in current knowledge. It highlights missing information in databases related to the Sec and Tat translocation pathways and various chaperone networks, noting that some are more thoroughly documented. For example, the DnaK chaperone network is relatively well-understood compared to the ClpB chaperone network, which remains less explored. During this phase, the investigation was guided by speculative concepts, positing that integrating these pathways into a cohesive model might yield new insights. This chapter underscores the gaps and challenges in practices. It establishes a foundational understanding and emphasises the essential need for an integrative model to enhance comprehension of these complex processes.

Chapter 4 presents the development of the model that elucidates the interactions between folded proteins and molecular chaperones. This advancement represents progress in biotechnology, particularly for enhancing recombinant protein production in *E. coli*. The

model highlights DnaK as a central chaperone, aligning with established empirical research. Moreover, the model's capability to predict outcomes based on varying chaperone expression levels provides a strategic advantage for designing more efficient protein production systems. This predictive ability enables more informed decisions in chaperone systems to progress and improve the amount of folded protein.

Expanding the model to include the translation of human growth hormone (HGH) as an example protein recombinantly expressed in *E. coli* presents the model's versatility and practical applications. This approach allows for the simulation of various scenarios and the interpretation of their outcomes, providing valuable insights into current pathways and suggesting improvements in protein production techniques. Nonetheless, the observation that increased chaperone levels do not always enhance protein yields highlights a crucial gap in understanding chaperone dynamics. This underlines the need for a thorough investigation into how chaperones interact and influence protein production, which remains inadequately explored in the literature. The model has tested these interactions, revealing new insights into the interplay of molecular chaperone networks.

Chapter 5 effectively showcases the model's strengths by providing new insights and practical applications. The chapter details test cases that modify different model components to evaluate their impact on the production of folded and misfolded proteins. While increasing the amount of correctly folded protein is desirable, minimising misfolded protein is equally crucial for maintaining cellular health.

The chapter explores Sec translocation pathways to map the routes taken by a protein when the signal peptide sequence is either hydrophobic or hydrophilic. It also investigates how modifications of disulphide bond formation pathways, such as those found in the CyDisCo commercial strain, result in elevated levels of inclusion bodies and mis-oxidised proteins within the model. The chapter further demonstrates that overexpressing chaperones like DnaK and DsbC can mitigate these issues, potentially enhancing the yield of correctly folded proteins.

These findings underscore the model's utility in strain engineering, revealing its capability to reduce the number of experimental trials needed to improve strains. The model can provide actionable insights and guide practical improvements in protein production. This will reduce the development cost, making these potential therapeutics more accessible.

Despite its strengths, the model has inherent limitations, as is familiar with all mathematical models. Its accuracy is contingent upon the assumptions and parameters used, and thus, predictions may not always capture real-world complexities. Due to data limitations, the model may simplify intricate interactions between proteins and chaperones, potentially overlooking meaningful interactions crucial for the overall system. These simplifications can result in inaccuracies or incomplete representations of biological processes. Ultimately, the model's effectiveness is limited by the quality and scope of the available data.

The model facilitates the identification of critical bottlenecks and inefficiencies within the pathways, allowing for targeted interventions. Furthermore, the model's ability to simulate

various scenarios and predict the effects of altering parameters provides valuable insights for improving process efficiency. This has led to enhanced strategies for recombinant protein production, ultimately advancing bioprocessing techniques and contributing to more effective and scalable production systems.

It has advanced bioprocessing by offering a structured framework for analysing complex pathways involved in protein production. It visually represents interactions and transitions across various pathways within *E. coli*. This detailed exploration facilitates a deeper understanding of these processes and helps identify critical points within the path. Additionally, the graphical interface of Petri nets provides a unique advantage, as it makes the model accessible to scientists without extensive modelling backgrounds, thereby broadening its utility and fostering collaboration across disciplines.

6.1 Future Work

Although the Petri net model is highly accessible due to its intuitive visual representation and user-friendly tools, these features alone are insufficient for rigorous quantitative analysis. A robust strategy must be implemented to maximise the model's utility, combining optimisation techniques, statistical validation, and systematic sensitivity analysis. For instance, optimisation algorithms like linear programming can identify the optimal combination of chaperones and translation rates to maximise folded protein production (Berthomieu & Diaz, 1991).

Statistical metrics such as root mean square error and correlation coefficients should be used to compare the model's predictions with experimental data, ensuring accuracy and reliability. Additionally, sensitivity analysis can identify which parameters impact the system, guiding further model refinement (Tang et al., 2007).

These requirements for quantitative analysis naturally lead to several avenues for future work. Experimental validation is essential to confirm the model's predictions, particularly for scenarios involving chaperone overexpression and reduced translation rates (Cajo et al., 2006). In addition, integrating omics data (e.g., transcriptomic, proteomic, and metabolomic data) can refine the model's parameters and improve its predictive capabilities (Griffin et al., 2002).

Although Petri nets show some resilience to incomplete data, as demonstrated in this thesis, their efficacy is influenced by the modelling strategies employed. Techniques such as parameter estimation, sensitivity analysis, and hybrid modelling can help alleviate the effects of data gaps. Nevertheless, challenges persist in maintaining robustness across various scales. The system boundary was established in the Petri nets utilised for this thesis, and assumptions were made to construct the model as discussed; however, if the model is

broadened to incorporate other metabolic pathways, the methodology will need to adapt, and data gaps may result in inaccuracies, particularly for finely detailed models. Future developments in data integration, uncertainty quantification, and hybrid methodologies will be crucial for improving the robustness of Petri net models when confronted with incomplete or fragmented data.

Future developments include:

1. Incorporating experimental data into the Petri net model will enable the validation and calibration of its predictions. For instance, experimental data on chaperone expression levels and protein folding efficiency can be used to validate and adjust the model's assumptions and outputs.

2. By integrating multi-omics data, the model can better predict how changes at one molecular level affect other levels, offering a more accurate representation of cellular behaviour.

3. Integrating dynamic experimental data that reflect how cells respond to environmental changes, such as variations in temperature, pH, and nutrient availability. This will allow the model to simulate real-time responses and optimise conditions for protein production. Experimental data can provide insights into how these environmental factors influence the behaviour of chaperones and the efficiency of translocation pathways.

4. Integrating whole-cell multi-omics data—into the Petri net model will provide a holistic view of cellular processes. This approach will enable the model to account for complex interactions between molecular layers and their impact on protein production.

By focusing on these advancements, the Petri net model can be further refined to address specific challenges and opportunities in *E. coli* protein production, enhancing its utility and effectiveness for research and industrial applications.

Another intriguing direction is integrating Petri nets with other modelling formalisms to represent biological pathways comprehensively. Combining Petri nets with data-driven models offers a promising strategy for improving the precision and relevance of computational models in systems biology.

Petri nets effectively capture biological systems' topology and dynamic interactions through their graphical and mathematical frameworks. At the same time, data-driven models can utilise extensive experimental data to enhance predictions and reveal hidden patterns. Merging these two approaches could harness the benefits of both paradigms, combining the structural clarity of Petri nets with the predictive capabilities of data-driven techniques.

Although some studies have begun to explore this hybrid approach, it remains a developing field. For instance, hybrid Petri nets have integrated continuous data (such as gene expression levels) with discrete events (like protein interactions), allowing for more realistic simulations of complex biological systems (Wingender, 2010). Likewise, machine learning methods have been employed to fine-tune the parameters of Petri net models, aligning them more closely with experimental data (Carvalho et al., 2018). However, the complete potential of integrating Petri nets with data-driven models has yet to be fully realised, especially in protein folding and chaperone-mediated processes.

References

Staines, A. S. (2011). Rewriting Petri nets as Directed Graphs. INTERNATIONAL JOURNAL of COMPUTERS (Vol. 5, Issue 2).

Abidi, W., Torres-Sánchez, L., Siroy, A., & Krasteva, P. V. (2022). Weaving of bacterial cellulose by the Bcs secretion systems. In FEMS Microbiology Reviews (Vol. 46, Issue 2). Oxford University Press. https://doi.org/10.1093/femsre/fuab051

Acheson, J. F., Ho, R., Goularte, N. F., Cegelski, L., & Zimmer, J. (2021). Molecular organization of the E. coli cellulose synthase macrocomplex. Nature Structural and Molecular Biology, 28(3), 310–318. https://doi.org/10.1038/s41594-021-00569-7

Adams, L. A., Sharma, P., Mohanty, B., Ilyichova, O. V., Mulcair, M. D., Williams, M. L., Gleeson, E. C., Totsika, M., Doak, B. C., Caria, S., Rimmer, K., Horne, J., Shouldice, S. R., Vazirani, M., Headey, S. J., Plumb, B. R., Martin, J. L., Heras, B., Simpson, J. S., & Scanlon, M. J. (2015). Application of fragment-based screening to the design of inhibitors of Escherichia coli DsbA. Angewandte Chemie - International Edition, 54(7). https://doi.org/10.1002/anie.201410341

Adamson, H., Robinson, M., Bond, P. S., Soboh, B., Gillow, K., Simonov, A. N., Elton, D. M., Bond, A. M., Sawers, R. G., Gavaghan, D. J., & Parkin, A. (2017). Analysis of HypD Disulphide Redox Chemistry via Optimization of Fourier Transformed ac Voltammetric Data. Analytical Chemistry, 89(3). https://doi.org/10.1021/acs.analchem.6b03589

Agashe, V. R., Guha, S., Chang, H. C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F. U., & Barral, J. M. (2004). Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. Cell, 117(2). https://doi.org/10.1016/S0092-8674(04)00299-5

Agharafeie, R., Ramos, J. R. C., Mendes, J. M., & Oliveira, R. (2023). From Shallow to Deep Bioprocess Hybrid Modeling: Advances and Future Perspectives. In Fermentation (Vol. 9, Issue 10). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/fermentation9100922

Aguilar-Rodríguez, J., Sabater-Muñoz, B., Montagud-Martínez, R., Berlanga, V., Alvarez-Ponce, D., Wagner, A., & Fares, M. A. (2016). The molecular chaperone dnak is a source of mutational robustness. Genome Biology and Evolution, 8(9). https://doi.org/10.1093/gbe/evw176

Ahmed, F. D., & Majid, M. A. (2019). Towards agent-based petri net decision making modelling for cloud service composition: A literature survey. In Journal of network and Computer Applications (Vol. 130, pp. 14–38). Academic Press. https://doi.org/10.1016/j.jnca.2018.12.001

Alanen, H. I., Walker, K. L., Lourdes Velez Suberbie, M., Matos, C. F. R. O., Bönisch, S., Freedman, R. B., Keshavarz-Moore, E., Ruddock, L. W., & Robinson, C. (2015). Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the Escherichia coli Tat pathway without

prior disulphide bond formation. Biochimica et Biophysica Acta - Molecular Cell Research, 1853(3). https://doi.org/10.1016/j.bbamcr.2014.12.027

Alcock, F., Stansfeld, P. J., Basit, H., Habersetzer, J., Baker, M. A., Palmer, T., Wallace, M. I., & Berks, B. C. (2016). Assembling the tat protein translocase. ELife, 5(DECEMBER2016). https://doi.org/10.7554/eLife.20718.001

Alla, H. (2007). Modeling and Analysis using Hybrid Petri nets.

Allen, W. J., Corey, R. A., Watkins, D. W., Oliveira, A. S. F., Hards, K., Cook, G. M., & Collinson, I. (2022). Rate-limiting transport of positively charged arginine residues through the Sec-machinery is integral to the mechanism of protein secretion. In eLife (Vol. 11). eLife Sciences Publications Ltd. https://doi.org/10.7554/eLife.77586

Allocati, N., Federici, L., Masulli, M., & Di Ilio, C. (2009). Glutathione transferases in bacteria. In FEBS Journal (Vol. 276, Issue 1). https://doi.org/10.1111/j.1742-4658.2008.06743.x

Almo, S. C., & Love, J. D. (2014). Better and faster: Improvements and optimization for mammalian recombinant protein production. In Current Opinion in Structural Biology (Vol. 26, Issue 1). https://doi.org/10.1016/j.sbi.2014.03.006

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., & Jirstrand, M. (2014). Kinetic models in industrial biotechnology – Improving cell factory performance. Metabolic Engineering, 24, 38–60. https://doi.org/10.1016/j.ymben.2014.03.007

Alur, R., & Henzinger, T. A. (1999). Formal Methods in System Design, 15(1), 7–48. https://doi.org/10.1023/a:1008739929481

Andersen, C. L., Matthey-Dupraz, A., Missiakas, D., & Raina, S. (1997). A new Escherichia coli gene, dsbG, encodes a periplasmic protein involved in disulphide bond formation, required for recycling DsbA/DsbB and DsbC redox proteins. Molecular Microbiology, 26(1). https://doi.org/10.1046/j.1365-2958.1997.5581925.x

Apostolidis, P. A., Tseng, A., Koziol, M. E., Betenbaugh, M. J., & Chiang, B. (2015). Investigation of low viability in sparged bioreactor CHO cell cultures points to variability in the Pluronic F-68 shear protecting component of cell culture media. Biochemical Engineering Journal, 98, 10–17. https://doi.org/10.1016/j.bej.2015.01.013

Arauzo-Aguilera, K., Saaranen, M. J., Robinson, C., & Ruddock, L. W. (2023). Highly efficient export of a disulphide-bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in Escherichia coli. MicrobiologyOpen, 12(2). https://doi.org/10.1002/mbo3.1350

Arhar, T., Shkedi, A., Nadel, C. M., & Gestwicki, J. E. (2021). The interactions of molecular chaperones with client proteins: Why are they so weak? In Journal of Biological Chemistry (Vol. 297, Issue 5). American Society for Biochemistry and Molecular Biology Inc. https://doi.org/10.1016/j.jbc.2021.101282

Aricescu, A. R., Lu, W., & Jones, E. Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallographica Section D: Biological Crystallography, 62(10). https://doi.org/10.1107/S0907444906029799

Ariosa, A. R., Duncan, S. S., Saraogi, I., Lu, X., Brown, A., Phillips, G. J., & Shan, S. O. (2013). Fingerloop activates cargo delivery and unloading during cotranslational protein targeting. Molecular Biology of the Cell, 24(2). https://doi.org/10.1091/mbc.E12-06-0434

Arts, I. S., Gennaris, A., & Collet, J. F. (2015). Reducing systems protecting the bacterial cell envelope from oxidative damage. In FEBS Letters (Vol. 589, Issue 14, pp. 1559–1568). Elsevier. https://doi.org/10.1016/j.febslet.2015.04.057

Aslanidis, C., & De Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-POR). In Nucleic Acids Research (Vol. 18, Issue 20).

Aspen Technology, Inc. (2000, February 20). Aspen Plus. AspenTech. https://web.ist.utl.pt/ist11038/acad/Aspen/AspUserGuide10.pdf

Assaf, G., Heiner, M., & Liu, F. (2022). Coloured fuzzy Petri nets for modelling and analysing membrane systems. BioSystems, 212. https://doi.org/10.1016/j.biosystems.2021.104592

Azevedo, C., Desfougères, Y., Jiramongkol, Y., Partington, H., Trakansuebkul, S., Singh, J., Steck, N., Jessen, H. J., & Saiardi, A. (2020). Development of a yeast model to study the contribution of vacuolar polyphosphate metabolism to lysine polyphosphorylation. In Journal of Biological Chemistry (Vol. 295, Issue 6). https://doi.org/10.1074/jbc.RA119.011680

Baars, L. (2007). Protein targeting, translocation and insertion in Escherichia coli Proteomic analysis of substrate-pathway relationships.

Bageshwar, U. K., DattaGupta, A., & Musser, S. M. (2021). Influence of the TorD signal peptide chaperone on Tat-dependent protein translocation. PLoS ONE, 16(9 September). https://doi.org/10.1371/journal.pone.0256715

Bageshwar, U. K., VerPlank, L., Baker, D., Dong, W., Hamsanathan, S., Whitaker, N., Sacchettini, J. C., & Musser, S. M. (2016). High throughput screen for Escherichia coli twin arginine translocation (Tat) inhibitors. PLoS ONE, 11(2). https://doi.org/10.1371/journal.pone.0149659

Baggett, N. E., Zhang, Y., & Gross, C. A. (2017). Global analysis of translation termination in E. coli. PLoS Genetics, 13(3). https://doi.org/10.1371/journal.pgen.1006676

Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A. A., Hoseinpoor, R., Rahbarnia, L., & Aria, M. (2019). Yeast Expression Systems: Overview and Recent Advances. In Molecular Biotechnology (Vol. 61, Issue 5, pp. 365–384). Humana Press Inc. https://doi.org/10.1007/s12033-019-00164-8

Bagherinejad, M., Sadeghi, H.-M., Abedi, D., Moazen, F., & Rabbani, M. (2018). Effect of Twine-arginine Translocation-signaling Fusion System and Chaperones Co-expression on Secretory Expression of Somatropin. Advanced Biomedical Research, 7(1). https://doi.org/10.4103/abr.abr_273_16

Banaś, A. M., Bocian-Ostrzycka, K. M., Plichta, M., Dunin-Horkawicz, S., Ludwiczak, J., Płaczkiewicz, J., & Jagusztyn-Krynicka, E. K. (2020). C8J-1298, a bifunctional thiol oxidoreductase of Campylobacter jejuni, affects Dsb (disulphide bond) network functioning. PLoS ONE, 15(3). https://doi.org/10.1371/journal.pone.0230366

Barnett, M. E., Nagy, M., Kedzierska, S., & Zolkiewski, M. (2005). The amino-terminal domain of ClpB supports binding to strongly aggregated proteins. The Journal of Biological Chemistry, 280(41), 34940–34945. https://doi.org/10.1074/jbc.M505653200

Barnett, J. P., Eijlander, R. T., Kuipers, O. P., & Robinson, C. (2008). A minimal tat system from a gram-positive organism: A bifunctional TatA subunit participates in discrete TatAC and TatA complexes. Journal of Biological Chemistry, 283(5). https://doi.org/10.1074/jbc.M708134200

Barragán-Ocaña, A., Silva-Borjas, P., Olmos-Peña, S., & Polanco-Olguín, M. (2020). Biotechnology and bioprocesses: Their contribution to sustainability. Processes, 8(4). https://doi.org/10.3390/pr8040436

Bause, F., & Kritzinger, P. (2013). Stochastic Petri nets-An Introduction to the Theory. https://www.researchgate.net/publication/258705139

Berkmen, M. (2012). Production of disulphide-bonded proteins in Escherichia coli. In Protein Expression and Purification (Vol. 82, Issue 1, pp. 240–251). https://doi.org/10.1016/j.pep.2011.10.009

Berkmen, M., Boyd, D., & Beckwith, J. (2005). The nonconsecutive disulphide bond of Escherichia coli phytase (AppA) renders it dependent on the protein-disulphide isomerase, DsbC. Journal of Biological Chemistry, 280(12), 11387–11394. https://doi.org/10.1074/jbc.M411774200

Bernardo, M., Vink, E., Pierro, A., & Wiklicky, H. (2013). Formal Methods for Dynamical Systems 13th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2013, Bertinoro, Italy, June 17-22, 2013. Advanced Lectures. Berlin, Heidelberg Springer.

Berthomieu, B., & Diaz, M. (1991). Modeling and verification of time dependent systems using time Petri nets. IEEE Transactions on Software Engineering, 17(3), 259–273. https://doi.org/10.1109/32.75415

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. In Frontiers in Bioengineering and Biotechnology (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2021.630551

Bhuwan, M., Arora, N., Sharma, A., Khubaib, M., Pandey, S., Chaudhuri, T. K., Hasnain, S. E., & Ehtesham, N. Z. (2016). Interaction of Mycobacterium tuberculosis virulence factor RipA with chaperone MoxR1 is required for transport through the TAT secretion system. MBio, 7(2). https://doi.org/10.1128/mBio.02259-15

Blaudeck, N., Kreutzenbeck, P., Müller, M., Sprenger, G. A., & Freudl, R. (2005). Isolation and characterization of bifunctional Escherichia coli TatA mutant proteins that allow efficient Tat-dependent protein translocation in the absence of TatB. Journal of Biological Chemistry, 280(5). https://doi.org/10.1074/jbc.M411210200

Blaudeck, N., Sprenger, G. A., Freudl, R., & Wiegert, T. (2001). Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation. Journal of Bacteriology, 183(2), 604–610. https://doi.org/10.1128/JB.183.2.604-610.2001

Blondin, M., & Esparza, J. (2022). Separators in Continuous Petri nets. http://arxiv.org/abs/2209.02767

Blümmel, A. S., Drepper, F., Knapp, B., Eimer, E., Warscheid, B., Müller, M., & Fröbel, J. (2017). Structural features of the TatC membrane protein that determine docking and insertion of a twin-arginine signal peptide. Journal of Biological Chemistry, 292(52). https://doi.org/10.1074/jbc.M117.812560

Bolhuis, A., Mathers, J. E., Thomas, J. D., Claire, M., Barrett, L., & Robinson, C. (2001). TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli. Journal of Biological Chemistry, 276(23), 20213–20219. https://doi.org/10.1074/jbc.M100682200

Bonet, P., & Lladó, C. (2007). PIPE v2. 5: A Petri net tool for performance modelling. Proc. of 23rd Latin American Conference on Informatics (CLEI 2007).

Boodhoo, K. V. K., Flickinger, M. C., Woodley, J. M., & Emanuelsson, E. A. C. (2022). Bioprocess intensification: A route to efficient and sustainable biocatalytic transformations for the future. In Chemical Engineering and Processing - Process Intensification (Vol. 172). Elsevier B.V. https://doi.org/10.1016/j.cep.2022.108793

Born, G., Plantier, E., Nannini, G., Caimi, A., Mazzoleni, A., Asnaghi, M. A., Muraro, M. G., Scherberich, A., Martin, I., & García-García, A. (2023). Mini- and macro-scale direct perfusion bioreactors with optimized flow for engineering 3D tissues. Biotechnology Journal, 18(2). https://doi.org/10.1002/biot.202200405

Boy, D., & Koch, H.-G. (2009). Visualization of Distinct Entities of the SecYEG Translocon during Translocation and Integration of Bacterial Proteins. Molecular Biology of the Cell, 20, 1804–1815. https://doi.org/10.1091/mbc.E08

Brinkrolf, C., Ochel, L., & Hofestädt, R. (2021). VANESA: An open-source hybrid functional Petri net modeling and simulation environment in systems biology. BioSystems, 210. https://doi.org/10.1016/j.biosystems.2021.104531

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., & Eggert, T. (2006). Systematic Screening of All Signal Peptides from Bacillus subtilis: A Powerful Strategy in Optimizing Heterologous Protein Secretion in Gram-positive Bacteria. Journal of Molecular Biology, 362(3), 393–402. https://doi.org/10.1016/j.jmb.2006.07.034

Brodland, G. W. (2015). How computational models can help unlock biological systems. In Seminars in Cell and Developmental Biology (Vols. 47–48, pp. 62–73). Academic Press. https://doi.org/10.1016/j.semcdb.2015.07.001

Brookwell, A., Oza, J. P., & Caschera, F. (2021). Biotechnology Applications of Cell-Free Expression Systems. In Life (Vol. 11, Issue 12). MDPI. https://doi.org/10.3390/life11121367

Browning, D. F., Richards, K. L., Peswani, A. R., Roobol, J., Busby, S. J. W., & Robinson, C. (2017). Escherichia coli "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. Biotechnology and Bioengineering, 114(12). https://doi.org/10.1002/bit.26434

Bukau, B., Weissman, J., & Horwich, A. (2006). Molecular Chaperones and Protein Quality Control. Cell, 125(3), 443–451. https://doi.org/10.1016/j.cell.2006.04.014

Brüser, T., Yano, T., Brune, D. C., & Daldal, F. (2003). Membrane targeting of a folded and cofactor-containing protein. European Journal of Biochemistry, 270(6). https://doi.org/10.1046/j.1432-1033.2003.03481.x

Büssow, K. (2015). Stable mammalian producer cell lines for structural biology. In Current Opinion in Structural Biology (Vol. 32). https://doi.org/10.1016/j.sbi.2015.03.002

Buttgereit, F., & Brand, M. D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. Biochemical Journal, 312(1), 163–167. https://doi.org/10.1042/bj3120163

Cajo, G. C., Horne, B. E., Kelley, W. L., Schwager, F., Georgopoulos, C., & Genevaux, P. (2006). The Role of the DIF Motif of the DnaJ (Hsp40) Co-chaperone in the Regulation of the DnaK (Hsp70) Chaperone Cycle. Journal of Biological Chemistry, 281(18), 12436–12444. https://doi.org/10.1074/jbc.m511192200

Carvalho, R. V., Verbeek, F. J., & Coelho, C. J. (2018). Bio-modeling Using Petri nets: A Computational Approach. Computational Biology, 3–26. https://doi.org/10.1007/978-3-319-74974-7_1

Castillo-Corujo, A., Saaranen, M. J., & Ruddock, L. W. (2024). Cytoplasmic production of Fabs in chemically defined media in fed-batch fermentation. Protein Expression and Purification, 215. https://doi.org/10.1016/j.pep.2023.106404

Castillo-Corujo, A., Uchida, Y., Saaranen, M. J., & Ruddock, L. W. (2024). Escherichia coli Cytoplasmic Expression of Disulphide-Bonded Proteins: Side-by-Side Comparison between Two Competing Strategies. Journal of Microbiology and Biotechnology, 34(5), 1126–1134. https://doi.org/10.4014/jmb.2311.11025

Chatelle, C., Kraemer, S., Ren, G., Chmura, H., Marechal, N., Boyd, D., Roggemans, C., Ke, N., Riggs, P., Bardwell, J., & Berkmen, M. (2015). Converting a sulfenic acid reductase into a disulphide bond isomerase. Antioxidants and Redox Signaling, 23(12). https://doi.org/10.1089/ars.2014.6235

Chaudhary, A. K., & Lee, E. Y. (2015). Tightly regulated and high level expression vector construction for Escherichia coli BL21 (DE3). Journal of Industrial and Engineering Chemistry, 31, 367–373. https://doi.org/10.1016/j.jiec.2015.07.011

Chay, Z. E., Goh, B. F., & Ling, M. H. T. (2016). Pnet : A Python Library for Petri net Modeling and Simulation. ACSIJ Advances in Computer Science: An International Journal, 5(4).

Chen, J., Acton, T. B., Basu, S. K., Montelione, G. T., & Inouye, M. (2002). Enhancement of the solubility of proteins overexpressed in Escherichia coli by heat shock. Journal of Molecular Microbiology and Biotechnology, 4(6).

Cherdal, S., & Mouline, S. (2018). Modelling and simulation of biochemical processes using Petri nets. Processes, 6(8). https://doi.org/10.3390/pr6080097

Cho, Y., Zhang, X., Pobre, K. F. R., Liu, Y., Powers, D. L., Kelly, J. W., Gierasch, L. M., & Powers, E. T. (2015). Individual and collective contributions of chaperoning and degradation to protein homeostasis in E. coli. Cell Reports, 11(2), 321–333. https://doi.org/10.1016/j.celrep.2015.03.018

Choi, J. H., & Lee, S. Y. (2004). Secretory and extracellular production of recombinant proteins using Escherichia coli. In Applied Microbiology and Biotechnology (Vol. 64, Issue 5). https://doi.org/10.1007/s00253-004-1559-9

Chong, Z. X., Yeap, S. K., & Ho, W. Y. (2021). Transfection types, methods and strategies: A technical review. In PeerJ (Vol. 9). PeerJ Inc. https://doi.org/10.7717/peerj.11165

Christensen, R. (2015). Analysis of variance, design, and regression: Linear modeling for unbalanced data, second edition. In Analysis of Variance, Design, and Regression: Linear Modeling for Unbalanced Data, Second Edition.

Christensen, S., Halili, M. A., Strange, N., Petit, G. A., Huston, W. M., Martin, J. L., & McMahon, R. M. (2019). Oxidoreductase disulphide bond proteins DsbA and DsbB form an active redox pair in Chlamydia trachomatis, a bacterium with disulphide dependent infection and development. PLoS ONE, 14(9). https://doi.org/10.1371/journal.pone.0222595

Collier, D. N., Bankaitis, V. A., Weiss, J. B., & Bassford, P. J. (1988). The antifolding activity of SecB promotes the export of the E. coli maltose-binding protein. Cell, 53(2). https://doi.org/10.1016/0092-8674(88)90389-3

Cox, M. M. J. (2021). Innovations in the Insect Cell Expression System for Industrial Recombinant Vaccine Antigen Production. In Vaccines (Vol. 9, Issue 12). MDPI. https://doi.org/10.3390/VACCINES9121504

Crane, J. M., & Randall, L. L. (2017). The Sec System: Protein Export in Escherichia coli . EcoSal Plus, 7(2). https://doi.org/10.1128/ecosalplus.esp-0002-2017

Cranford-Smith, T., Jamshad, M., Jeeves, M., Chandler, R. A., Yule, J., Robinson, A., Alam, F., Dunne, K. A., Aponte Angarita, E. H., Alanazi, M., Carter, C., Henderson, I. R., Lovett, J. E., Winn, P., Knowles, T., & Huber, D. (2020). Iron is a ligand of SecA-like metal-binding domains in vivo. Journal of Biological Chemistry, 295(21). https://doi.org/10.1074/jbc.RA120.012611

Crater, J. S., & Lievense, J. C. (2018). Scale-up of industrial microbial processes. In FEMS Microbiology Letters (Vol. 365, Issue 13). Oxford University Press. https://doi.org/10.1093/femsle/fny138

Cristó, S., De Gier, J.-W., Nielsen, H., & Von Heijne, G. (1999). Competition between Sec-and TAT-dependent protein translocation in Escherichia coli. In The EMBO Journal (Vol. 18, Issue 11).

Crossley, J. A., Allen, W. J., Watkins, D. W., Sabir, T., Radford, S. E., Tuma, R., Collinson, I., & Fessl, T. (2024). Dynamic coupling of fast channel gating with slow ATP-turnover underpins protein transport through the Sec translocon. EMBO Journal, 43(1). https://doi.org/10.1038/s44318-023-00004-1

Czerwinski, W., Lasota, S., Lazic, R., Leroux, J., & Mazowiecki, F. (2018). The Reachability Problem for Petri nets is Not Elementary. http://arxiv.org/abs/1809.07115

Da Silva, M. (2022). What is the Difference Between a Test Case and a Test Scenario? Uilicious.com. https://uilicious.com/blog/difference-between-test-case-test-scenario/

Dalal, K., Chan, C. S., Sligar, S. G., & Duong, F. (2012). Two copies of the SecY channel and acidic lipids are necessary to activate the SecA translocation ATPase. Proceedings of the National Academy of Sciences, 109(11), 4104–4109. https://doi.org/10.1073/pnas.1117783109

Daley, D. O. (2005). Global Topology Analysis of the Escherichia coli Inner Membrane Proteome. Science, 308(5726), 1321–1323. https://doi.org/10.1126/science.1109730

Dariushnejad, H., Farajnia, S., Zarghami, N., Aria, M., & Tanomand, A. (2019). Effect of DnaK/DnaJ/GrpE and DsbC Chaperons on Periplasmic Expression of Fab Antibody by E. coli SEC Pathway. International Journal of Peptide Research and Therapeutics, 25(1). https://doi.org/10.1007/s10989-017-9637-x

David, N., Jard, C., Lime, D., Henri Roux, O., & Roux, O. H. (2015). Discrete Parameters in Petri nets. Application and Theory of Petri nets and Concurrency. 137–156. https://doi.org/10.1007/978-3-319-19488-2_7ï

David, R., & Alla, H. (Hassane). (2005). Discrete, continuous, and hybrid Petri nets. Springer.

Davis, M. W., & Jorgensen, E. M. (2022). ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program. Frontiers in Bioinformatics, 2. https://doi.org/10.3389/fbinf.2022.818619

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., & Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. BMC Biotechnology, 7. https://doi.org/10.1186/1472-6750-7-32

De Mets, F., Van Melderen, L., & Gottesman, S. (2019). Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. Proceedings of the National Academy of Sciences of the United States of America, 116(3). https://doi.org/10.1073/pnas.1815288116

De Waelheyns, E., Segers, K., Sardis, M. F., Anné, J., Nicolaes, G. A. F., & Economou, A. (2015). Identification of small-molecule inhibitors against SecA by structure-based virtual ligand screening. Journal of Antibiotics, 68(11). https://doi.org/10.1038/ja.2015.53

Delisa, M. P., Tullman, D., Georgiou, G., & Beckwith, J. (2002). Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. www.pnas.orgcgidoi10.1073pnas.0937838100

Delvigne, F., Takors, R., Mudde, R., van Gulik, W., & Noorman, H. (2017). Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. Microbial Biotechnology, 10(5), 1267–1274. https://doi.org/10.1111/1751-7915.12803

Deng, J., Li, J., Ma, M., Zhao, P., Ming, F., Lu, Z., Shi, J., Fan, Q., Liang, Q., Jia, J., Li, J., Zhang, S., & Zhang, L. (2020). Co-expressing GroEL-GroES, Ssa1-Sis1 and Bip-PDI chaperones for enhanced intracellular production and partial-wall breaking improved stability of porcine growth hormone. Microbial Cell Factories, 19(1). https://doi.org/10.1186/s12934-020-01304-5

Denoncin, K., & Collet, J. F. (2013). Disulphide bond formation in the bacterial periplasm: Major achievements and challenges ahead. In Antioxidants and Redox Signaling (Vol. 19, Issue 1, pp. 63–71). Mary Ann Liebert Inc. https://doi.org/10.1089/ars.2012.4864

Denoncin, K., Vertommen, D., Arts, I. S., Goemans, C. V., Rahuel-Clermont, S., Messens, J., & Collet, J. F. (2014). A new role for escherichia coli dsbc protein in protection against oxidative stress. Journal of Biological Chemistry, 289(18), 12356–12364. https://doi.org/10.1074/jbc.M114.554055

Diamos, A. G., Hunter, J. G. L., Pardhe, M. D., Rosenthal, S. H., Sun, H., Foster, B. C., DiPalma, M. P., Chen, Q., & Mason, H. S. (2020). High Level Production of Monoclonal Antibodies Using an Optimized Plant Expression System. Frontiers in Bioengineering and Biotechnology, 7. https://doi.org/10.3389/fbioe.2019.00472

Diao, L., Dong, Q., Xu, Z., Yang, S., Zhou, J., & Freudl, R. (2012). Functional implementation of the posttranslational SecB-SecA protein-targeting pathway in Bacillus subtilis. Applied and Environmental Microbiology, 78(3). https://doi.org/10.1128/AEM.07209-11

Dingle, N. J., & Knottenbelt, W. J. (2009). Automated Customer-Centric Performance Analysis of Generalised Stochastic Petri nets Using Tagged Tokens. Electronic Notes in Theoretical Computer Science, 232(C). https://doi.org/10.1016/j.entcs.2009.02.051

Dolata, K. M., Montero, I. G., Miller, W., Sievers, S., Sura, T., Wolff, C., Schlüter, R., Riedel, K., & Robinson, C. (2019). Far-reaching cellular consequences of tat deletion in Escherichia coli revealed by comprehensive proteome analyses. Microbiological Research, 218. https://doi.org/10.1016/j.micres.2018.10.008

Dong, Z., Zhang, J., Du, G., Chen, J., Li, H., & Lee, B. (2015). Periplasmic Export of Bile Salt Hydrolase in Escherichia coli by the Twin-Arginine Signal Peptides. Applied Biochemistry and Biotechnology, 177(2). https://doi.org/10.1007/s12010-015-1755-3

Dorai, H., Corisdeo, S., Ellis, D., Kinney, C., Chomo, M., Hawley-Nelson, P., Moore, G., Betenbaugh, M. J., & Ganguly, S. (2012). Early prediction of instability of chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. Biotechnology and Bioengineering, 109(4). https://doi.org/10.1002/bit.24367

Doyle, S. M., Genest, O., & Wickner, S. (2013). Protein rescue from aggregates by powerful molecular chaperone machines. Nature Reviews Molecular Cell Biology, 14(10), 617–629. https://doi.org/10.1038/nrm3660

Driessen, A. J. M., & Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. Annual Review of Biochemistry, 77, 643–667. https://doi.org/10.1146/annurev.biochem.77.061606.160747

Drobnjakovic, M., Hart, R., Kulvatunyou, B., Ivezic, N., & Srinivasan, V. (2023). Current challenges and recent advances on the path towards continuous biomanufacturing. In Biotechnology Progress (Vol. 39, Issue 6). https://doi.org/10.1002/btpr.3378

Du Plessis, D. J. F., Nouwen, N., & Driessen, A. J. M. (2006). Subunit a of cytochrome o oxidase requires both YidC and SecYEG for membrane insertion. Journal of Biological Chemistry, 281(18). https://doi.org/10.1074/jbc.M600048200

Du Plessis, D. J. F., Nouwen, N., & Driessen, A. J. M. (2011). The Sec translocase. In Biochimica et Biophysica Acta - Biomembranes (Vol. 1808, Issue 3, pp. 851–865). Elsevier B.V. https://doi.org/10.1016/j.bbamem.2010.08.016

Dubini, A., & Sargent, F. (2003). Assembly of Tat-dependent [NiFe] hydrogenases: Identification of precursor-binding accessory proteins. FEBS Letters, 549(1–3). https://doi.org/10.1016/S0014-5793(03)00802-0

Dukan, S., Farewell, A., Ballesteros, M., François Taddéi, Radman, M., & Nyström, T. (2000). Protein oxidation in response to increased transcriptional or translational errors. Proceedings of the National Academy of Sciences of the United States of America, 97(11), 5746–5749. https://doi.org/10.1073/pnas.100422497

Durrani, F. G., Gul, R., Mirza, M. U., Kaderbhai, N. N., Froeyen, M., & Saleem, M. (2019). Mutagenesis of dsbass is crucial for the signal recognition particle mechanism in escherichia coli: Insights from molecular dynamics simulations. Biomolecules, 9(4). https://doi.org/10.3390/biom9040133

Economou, A., Pogliano, J. A., Beckwith, J., Oliver, D. B., & Wickner, W. (1995). SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. Cell, 83(7). https://doi.org/10.1016/0092-8674(95)90143-4

Eimer, E., Fröbel, J., Blümmel, A. S., & Müller, M. (2015). TatE as a regular constituent of bacterial twin-arginine protein translocases. Journal of Biological Chemistry, 290(49). https://doi.org/10.1074/jbc.M115.696005

Erkut, E. (2021). Bacterial signal peptides: structure, optimization, and applications. Eureka, 6(1). https://doi.org/10.29173/eureka28759

Eser, M., Masip, L., Kadokura, H., Georgiou, G., & Beckwith, J. (2009). Disulphide bond formation by exported glutaredoxin indicates glutathione's presence in the E. coli periplasm. Proceedings of the National Academy of Sciences of the United States of America, 106(5). https://doi.org/10.1073/pnas.0812596106

Fabianek, R. A., Hennecke, H., & Thöny-Meyer, L. (2000). Periplasmic protein thiol:disulphide oxidoreductases of Escherichia coli. In FEMS Microbiology Reviews (Vol. 24, Issue 3). https://doi.org/10.1016/S0168-6445(00)00028-0

Fan, D., Zhou, Q., Liu, C., & Zhang, J. (2016). Functional characterization of the Helicobacter pylori chaperone protein HP0795. Microbiological Research, 193. https://doi.org/10.1016/j.micres.2016.08.012

Fauvet, B., Finka, A., Castanié-Cornet, M.-P., Cirinesi, A.-M., Genevaux, P., Quadroni, M., & Goloubinoff, P. (2019). Bacterial Hsp90 mediates the degradation of aggregation-prone Hsp70-Hsp40 substrates preferentially by HslUV proteolysis. BioRxiv.

Feeney, M. A., Ke, N., & Beckwith, J. (2012). Mutations at several loci cause increased expression of ribonucleotide reductase in Escherichia coli. Journal of Bacteriology, 194(6). https://doi.org/10.1128/JB.05989-11

Ferguson, G. P., & Booth, I. R. (1998). Importance of glutathione for growth and survival of Escherichia coli cells: Detoxification of methylglyoxal and maintenance of intracellular K+. Journal of Bacteriology, 180(16). https://doi.org/10.1128/jb.180.16.4314-4318.1998

Figueiredo, L., Klunker, D., Ang, D., Naylor, D. J., Kerner, M. J., Georgopoulos, C., Hartl, F. Ulrich., & Hayer-Hartl, M. (2004). Functional Characterization of an Archaeal GroEL/GroES Chaperonin System. Journal of Biological Chemistry, 279(2), 1090–1099. https://doi.org/10.1074/jbc.m310914200

Findik, B. T., Smith, V. F., & Randall, L. L. (2018). Penetration into membrane of amino-terminal region of SecA when associated with SecYEG in active complexes. Protein Science, 27(3). https://doi.org/10.1002/pro.3362

Finkel, A., & Schnoebelen, P. (2001). Well-structured transition systems everywhere! In Theoretical Computer Science (Vol. 256). www.elsevier.com/locate/tcs

Fischer, H. P. (2008). Mathematical Modeling of Complex Biological Systems From Parts Lists to Understanding Systems Behavior. http://teaching.anhb.uwa.edu.au/mb140/

Frain, K. M., Robinson, C., & van Dijl, J. M. (2019). Transport of Folded Proteins by the Tat System. In Protein Journal (Vol. 38, Issue 4, pp. 377–388). Springer Science and Business Media, LLC. https://doi.org/10.1007/s10930-019-09859-y

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. In Microbial Cell Factories (Vol. 17, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12934-018-0901-3

Friehs, K. (2004). Plasmid copy number and plasmid stability. In Advances in biochemical engineering/biotechnology (Vol. 86). https://doi.org/10.1007/b12440

Fu, H., Liang, Y., Zhong, X., Pan, Z. L., Huang, L., Zhang, H. L., Xu, Y., Zhou, W., & Liu, Z. (2020). Codon optimization with deep learning to enhance protein expression. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-74091-z

Fukuda, I. M., Pinto, C. F. F., Moreira, C. D. S., Saviano, A. M., & Lourenço, F. R. (2018). Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD). In Brazilian Journal of Pharmaceutical Sciences (Vol. 54, Issue Special Issue). Faculdade de Ciencias Farmaceuticas (Biblioteca). https://doi.org/10.1590/s2175-97902018000001006

Gaber, A., Hassan, M. M., & El-Awady, M. A. (2015). The overproduction of synechocystis sp. PCC 6803 heat-shock protein (Sll0170) protects escherichia coli against high-temperature stress. Biotechnology and Biotechnological Equipment, 29(6). https://doi.org/10.1080/13102818.2015.1074056

Gaciarz, A., Khatri, N. K., Velez-Suberbie, M. L., Saaranen, M. J., Uchida, Y., Keshavarz-Moore, E., & Ruddock, L. W. (2017). Efficient soluble expression of disulphide bonded proteins in the cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal media. Microbial Cell Factories, 16(1). https://doi.org/10.1186/s12934-017-0721-x

Gao, S., Holkar, A., & Srivastava, S. (2019). Protein-polyelectrolyte complexes and micellar assemblies. In Polymers (Vol. 11, Issue 7). https://doi.org/10.3390/polym11071097

Gawin, A., Ertesvåg, H., Hansen, S. A. H., Malmo, J., & Brautaset, T. (2020). Translational regulation of periplasmic folding assistants and proteases as a valuable strategy to improve production of translocated recombinant proteins in Escherichia coli. BMC Biotechnology, 20(1). https://doi.org/10.1186/s12896-020-00615-0

Gerzon, G., Sheng, Y., & Kirkitadze, M. (2022). Process Analytical Technologies – Advances in bioprocess integration and future perspectives. Journal of Pharmaceutical and Biomedical Analysis, 207. https://doi.org/10.1016/j.jpba.2021.114379

Gilbertson, J. A., Sen, A., Behie, L. A., & Kallos, M. S. (2006). Scaled-up production of mammalian neural precursor cell aggregates in computer-controlled suspension bioreactors. Biotechnology and Bioengineering, 94(4). https://doi.org/10.1002/bit.20900

Giovanni Campolongo. (2018). Biopharma PAT - Quality Attributes, Critical Process Parameters & Key Performance Indicators at the Bioreactor. Hamilton White Paper, May.

Giua, A., & Seatzu, C. (2015). Petri nets for the control of discrete event systems. Software and Systems Modeling, 14(2), 693–701. https://doi.org/10.1007/s10270-014-0425-1

Giua, A., & Silva, M. (2018). Petri nets and Automatic Control: A historical perspective. Annual Reviews in Control, 45, 223–239. https://doi.org/10.1016/j.arcontrol.2018.04.006

Gleave, A. P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Molecular Biology, 20(6). https://doi.org/10.1007/BF00028910

Gleiter, S., & Bardwell, J. C. A. (2008). Disulphide bond formation in the periplasm of Escherichia coli. Molecular Microbiology, 68(4), 835–844. https://doi.org/10.1111/j.1365-2958.2008.06207.x

Glory, P. P., & Sheela, E. J. (2012). WORKFLOW NET MODELING USING WoPeD FOR THE EFFECTS OF INHIBITORS IN THE SYNTHESIS AND PROCESSING OF N-LINKED GLYCOPROTEIN International Journal of Systems Biology. 3(1). http://www.bioinfo.in/contents.php?id=29

Goel, G., Chou, I.-C., & Voit, E. O. (2006). Biological Systems Modeling and Analysis: A Biomolecular Technique of the Twenty-first Century. In Journal of Biomolecular Techniques (Vol. 17, Issue 4).

Goel, N., Srivastav, S., Patel, A., Shirsath, A., Panda, T. R., Patra, M., Feist, A. M., & Anand, A. (2023). TCA cycle tailoring facilitates optimal growth of proton-pumping NADH dehydrogenase-dependent Escherichia coli . Microbiology Spectrum, 11(6). https://doi.org/10.1128/spectrum.02225-23

Goloubinoff, P., Mogk, A., Anat Zvi, Toshifumi Tomoyasu, & Bernd Bukau. (1999). Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. 96(24), 13732–13737. https://doi.org/10.1073/pnas.96.24.13732

González-Figueredo, C., Flores-Estrella, A., & Rojas-Rejón, O. A. (2018). Fermentation: Metabolism, Kinetic Models, and Bioprocessing. www.intechopen.com

Goos, G., Hartmanis, J., Van, J., Board, L. E., Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Kobsa, A., Mattern, F., Zurich, E., Mitchell, J. C., Naor, M., Nierstrasz, O., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., & Weikum, G. (2009). Automated Technology for Verification and Analysis.

Goos, G., Hartmanis, J., Van, J., Board, L. E., Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Kobsa, A., Mattern, F., Zurich, E., Mitchell, J. C., Naor, M., Nierstrasz, O., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., & Weikum, G. (2012). LNCS 7347 - Application and Theory of Petri nets. In Lecture Notes in Computer Science.

Goos, G., Hartmanis, J., Van, J., Board, L. E., Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Zurich, E., Mitchell, J. C., Naor, M., Nierstrasz, O., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., & Weikum, G. (2007). LNCS 4486 - Formal Methods for Performance Evaluation.

Gorochowski, T. E. (2016). Agent-based modelling in synthetic biology. Essays in Biochemistry, 60(4), 325–336. https://doi.org/10.1042/EBC20160037

Gowrishankar, J., & Harinarayanan, R. (2004). Why is transcription coupled to translation in bacteria? In Molecular Microbiology (Vol. 54, Issue 3, pp. 598–603). https://doi.org/10.1111/j.1365-2958.2004.04289.x

Goyal, M., & Chaudhuri, T. K. (2015). GroEL-GroES assisted folding of multiple recombinant proteins simultaneously over-expressed in Escherichia coli. International Journal of Biochemistry and Cell Biology, 64, 277–286. https://doi.org/10.1016/j.biocel.2015.04.018

Grahl, S., Maillard, J., Spronk, C. A. E. M., Vuister, G. W., & Sargent, F. (2012). Overlapping transport and chaperone-binding functions within a bacterial twin-arginine signal peptide. Molecular Microbiology, 83(6). https://doi.org/10.1111/j.1365-2958.2012.08005.x

Grasso, S., Dabene, V., Hendriks, M. M. W. B., Zwartjens, P., Pellaux, R., Held, M., Panke, S., van Dijl, J. M., Meyer, A., & van Rij, T. (2023). Signal Peptide Efficiency: From High-Throughput Data to Prediction and Explanation. ACS Synthetic Biology, 12(2), 390–404. https://doi.org/10.1021/acssynbio.2c00328

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., & Aebersold, R. (2002). Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels inSaccharomyces cerevisiae. Molecular & Cellular Proteomics, 1(4), 323–333. https://doi.org/10.1074/mcp.m200001-mcp200

Grindle, M. P., Carter, B., Alao, J. P., Connors, K., Tehver, R., & Kravats, A. N. (2021). Structural communication between the e. Coli chaperones dnak and hsp90. International Journal of Molecular Sciences, 22(4). https://doi.org/10.3390/ijms22042200

Guerrero Montero, I., Dolata, K. M., Schlüter, R., Malherbe, G., Sievers, S., Zühlke, D., Sura, T., Dave, E., Riedel, K., & Robinson, C. (2019). Comparative proteome analysis in an Escherichia coli CyDisCo strain identifies stress responses related to protein production, oxidative stress and accumulation of misfolded protein. Microbial Cell Factories, 18(1), 1–15. https://doi.org/10.1186/s12934-019-1071-7

Guerrero Montero, I., Richards, K. L., Jawara, C., Browning, D. F., Peswani, A. R., Labrit, M., Allen, M., Aubry, C., Davé, E., Humphreys, D. P., Busby, S. J. W., & Robinson, C. (2019). Escherichia coli “TatExpress” strains export several g/L human growth hormone to the periplasm by the Tat pathway. Biotechnology and Bioengineering, 116(12). https://doi.org/10.1002/bit.27147

Guilhot, C., Jander, G., Martin, N. L., & Beckwith, J. (1995). Evidence that the pathway of disulphide bond formation in Escherichia coli involves interactions between the cysteines of DsbB and DsbA. Proceedings of the National Academy of Sciences of the United States of America, 92(21). https://doi.org/10.1073/pnas.92.21.9895

Gupta, R., Toptygin, D., & Kaiser, C. M. (2020). The SecA motor generates mechanical force during protein translocation. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-17561-2

Gupta, S., Rao, A. R., Varadwaj, P. K., De, S., & Mohapatra, T. (2015). Extrapolation of inter domain communications and substrate binding cavity of camel HSP70 1A: A molecular modeling and dynamics simulation study. PLoS ONE, 10(8). https://doi.org/10.1371/journal.pone.0136630

Ha, S. C., Lee, T. H., Cha, S. S., & Kim, K. K. (2004). Functional identification of the SecB homologue in Methanococcus jannaschii and direct interaction of SecB with trigger factor. Biochemical and Biophysical Research Communications, 315(4). https://doi.org/10.1016/j.bbrc.2004.02.002

Haeger, G., Wirges, J., Tanzmann, N., Oyen, S., Jolmes, T., Jaeger, K. E., Schörken, U., Bongaerts, J., & Siegert, P. (2023). Chaperone assisted recombinant expression of a mycobacterial aminoacylase in Vibrio

natriegens and Escherichia coli capable of N-lauroyl-L-amino acid synthesis. Microbial Cell Factories, 22(1). https://doi.org/10.1186/s12934-023-02079-1

Hajihassan, Z., Yazdi, M., Fadaie, A., & Akbarsemnani, N. (2024). Comparison of the efficiency of the Sec and Tat secretory pathways in the secretion of recombinant neurturin protein using de novo designed signal peptides. Preparative Biochemistry and Biotechnology. https://doi.org/10.1080/10826068.2024.2331203

Han, M.-J., & Lee, S. Y. (2006). The Escherichia coli Proteome: Past, Present, and Future Prospects . Microbiology and Molecular Biology Reviews, 70(2), 362–439. https://doi.org/10.1128/mmbr.00036-05

Han, S. J., Machhi, S., Berge, M., Xi, G., Linke, T., & Schoner, R. (2017). Novel signal peptides improve the secretion of recombinant Staphylococcus aureus Alpha toxinH35L in Escherichia coli. AMB Express, 7(1). https://doi.org/10.1186/s13568-017-0394-1

Hand, N. J., Klein, R., Laskewitz, A., & Pohlschröder, M. (2006). Archaeal and bacterial SecD and SecF homologs exhibit striking structural and functional conservation. Journal of Bacteriology, 188(4). https://doi.org/10.1128/JB.188.4.1251-1259.2006

Hao, B., Zhou, W., & Theg, S. M. (2022). Hydrophobic mismatch is a key factor in protein transport across lipid bilayer membranes via the Tat pathway. Journal of Biological Chemistry, 298(7). https://doi.org/10.1016/j.jbc.2022.101991

Hardy, S., & Robillard, P. N. (2004). Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. Journal of Bioinformatics and Computational Biology, 2(4). https://doi.org/10.1142/S0219720004000764

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature, 475(7356), 324–332. https://doi.org/10.1038/nature10317

Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P., & Wickner, W. (1990). The binding cascade of SecB to SecA to SecY E mediates preprotein targeting to the E. coli plasma membrane. Cell, 63(2). https://doi.org/10.1016/0092-8674(90)90160-G

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. In Genome Biology (Vol. 18, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13059-017-1215-1

Hatzixanthis, K., Clarke, T. A., Oubrie, A., Richardson, D. J., Turner, R. J., & Sargent, F. (2005). Signal peptide-chaperone interactions on the twin-arginine protein transport pathway. Proceedings of the National Academy of Sciences of the United States of America, 102(24). https://doi.org/10.1073/pnas.0500737102

Heidelberg, C. (2021). Highlights DoE by Hand Calculations: Effects and Interactions Acquaintance with Minitab Basic Statistical tools for Interpretation of DoE Output Are the Factors Significant? Full Factorial DoE Experiments with Minitab Screening Design Experiments with Minitab Optimisation with Response Surface Methodology. http://www.minitab.

Heiner, M., Gilbert, D., Donaldson, R., Bernardo, M., Degano, P., & Zavattaro, G. (2008). Petri nets for Systems and Synthetic Biology The advantages of using Petri nets as a kind of umbrella formalism are seen in the following. In LNCS (Vol. 5016). Springer-Verlag.

Heiner, M., Herajy, M., Liu, F., Rohr, C., & Schwarick, M. (2012). Snoopy - A unifying Petri net tool. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7347 LNCS. https://doi.org/10.1007/978-3-642-31131-4_22

Hemmis, C. W., Berkmen, M., Eser, M., & Schildbach, J. F. (2011). TrbB from Conjugative Plasmid F is a Structurally distinct disulphide isomerase that requires DsbD for redox state maintenance. Journal of Bacteriology, 193(18). https://doi.org/10.1128/JB.00351-11

Henson, M. A. (2008). Real-Time Implementation of Nonlinear Predictive Control.

Herajy, M., Liu, F., Rohr, C., & Heiner, M. (2018). Coloured Hybrid Petri nets: An adaptable modelling approach for multi-scale biological networks. Computational Biology and Chemistry, 76, 87–100. https://doi.org/10.1016/j.compbiolchem.2018.05.023

Hinkelmann, F., Brandon, M., Guang, B., McNeill, R., Blekherman, G., Veliz-Cuba, A., & Laubenbacher, R. (2011). ADAM: Analysis of Discrete Models of Biological Systems Using Computer Algebra. BMC Bioinformatics, 12. https://doi.org/10.1186/1471-2105-12-295

Hinkelmann, K., & Kempthorne, O. (1962). Design and Analysis of Experiments Advanced Experimental Design. In Design and Analysis of Experiments (Vol. 2).

Hobman, J. L., Penn, C. W., & Pallen, M. J. (2007). Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully? Molecular Microbiology, 64(4), 881–885. https://doi.org/10.1111/j.1365-2958.2007.05710.x

Horwich, A. L., & Fenton, W. A. (2020). Chaperonin-assisted protein folding: a chronologue. In Quarterly Reviews of Biophysics (Vol. 53). Cambridge University Press. https://doi.org/10.1017/S0033583519000143

Horwich, A. L., Farr, G. W., & Fenton, W. A. (2006). GroEL−GroES-Mediated Protein Folding. Chemical Reviews, 106(5), 1917–1930. https://doi.org/10.1021/cr040435v

Hryniewicz, M., Sirko, A., Pa0ucha, A., Bock, A., Hulanicka1, D., Hryniewicz, (M, Sirko, A., & Hulanicka, D. (1990). Sulfate and Thiosulfate Transport in Escherichia coli K-12: Identification of a Gene Encoding a Novel Protein Involved in Thiosulfate Binding “cysA” locus. In JOURNAL OF BACTERIOLOGY. https://journals.asm.org/journal/jb

Hu, W., Zhu, Y., & Lei, J. (2011a). The Detection and Prevention of Deadlock in Petri nets. Physics Procedia, 22, 656–659. https://doi.org/10.1016/j.phpro.2011.11.102

Huang, M., Wang, G., Qin, J., Petranovic, D., & Nielsen, J. (2018). Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production. Proceedings of the National Academy of Sciences of the United States of America, 115(47). https://doi.org/10.1073/pnas.1809921115

Huang, Q., Alcock, F., Kneuper, H., Deme, J. C., Rollauer, S. E., Lea, S. M., Berks, B. C., & Palmer, T. (2017). A signal sequence suppressor mutant that stabilizes an assembled state of the twin arginine translocase. Proceedings of the National Academy of Sciences of the United States of America, 114(10). https://doi.org/10.1073/pnas.1615056114

Huber, D., Boyd, D., Xia, Y., Olma, M. H., Gerstein, M., & Beckwith, J. (2005). Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation. Journal of Bacteriology, 187(9), 2983–2991. https://doi.org/10.1128/JB.187.9.2983-2991.2005

Huber, D., Rajagopalan, N., Preissler, S., Rocco, M. A., Merz, F., Kramer, G., & Bukau, B. (2011). SecA Interacts with Ribosomes in Order to Facilitate Posttranslational Translocation in Bacteria. Molecular Cell, 41(3). https://doi.org/10.1016/j.molcel.2010.12.028

Humphreys, D. P., Sehdev, M., Chapman, A. P., Ganesh, R., Smith, B. J., King, L. M., Glover, D. J., Reeks, D. G., & Stephens, P. E. (2000). High-level periplasmic expression in Escherichia coli using a eukaryotic signal peptide: Importance of codon usage at the 5′ end of the coding sequence. Protein Expression and Purification, 20(2). https://doi.org/10.1006/prep.2000.1286

Hupp, T. R., Keasling, J. D., Cooper, S., & Kaguni, J. M. (1994). Synthesis of DnaK protein during the division cycle of Escherichia coli. Research in Microbiology, 145(2). https://doi.org/10.1016/0923-2508(94)90003-5

Hussain, H., Fisher, D. I., Abbott, W. M., Roth, R. G., & Dickson, A. J. (2017). Use of a protein engineering strategy to overcome limitations in the production of “Difficult to Express” recombinant proteins. Biotechnology and Bioengineering, 114(10). https://doi.org/10.1002/bit.26358

Inaba, K. (2008). Protein disulphide bond generation in Escherichia coli DsbB-DsbA. Journal of Synchrotron Radiation, 15(3), 199–201. https://doi.org/10.1107/S090904950706061X

Inaba, K., & Ito, K. (2008). Structure and mechanisms of the DsbB-DsbA disulphide bond generation machine. In Biochimica et Biophysica Acta - Molecular Cell Research (Vol. 1783, Issue 4, pp. 520–529). https://doi.org/10.1016/j.bbamcr.2007.11.006

Ito, Y., Ishigami, M., Hashiba, N., Nakamura, Y., Terai, G., Hasunuma, T., Ishii, J., & Kondo, A. (2022). Avoiding entry into intracellular protein degradation pathways by signal mutations increases protein secretion in Pichia pastoris. Microbial Biotechnology, 15(9). https://doi.org/10.1111/1751-7915.14061

Ivanov, S. S., Chung, A. S., Yuan, Z. L., Guan, Y. J., Sachs, K. V., Reichner, J. S., & Chin, Y. E. (2004). Antibodies immobilized as arrays to profile protein post-translational modifications in mammalian cells. Molecular and Cellular Proteomics, 3(8). https://doi.org/10.1074/mcp.M300130-MCP200

Ize, B., Stanley, N. R., Buchanan, G., & Palmer, T. (2003). Role of the Escherichia coli Tat pathway in outer membrane integrity. Molecular Microbiology, 48(5), 1183–1193. https://doi.org/10.1046/j.1365-2958.2003.03504.x

Jagan, B. G. V. S., Narasimha Murthy, P., Mahapatra, A. K., & Patra, R. K. (2021). Quality by design (QBD): Principles, underlying concepts, and regulatory prospects. In Thai Journal of Pharmaceutical Sciences (Vol. 45, Issue 1).

Jalomo-Khayrova, E., Mares, R. E., Muñoz, P. L. A., Meléndez-López, S. G., Rivero, I. A., & Ramos, M. A. (2018). Soluble expression of an amebic cysteine protease in the cytoplasm of Escherichia coli SHuffle Express cells and purification of active enzyme. BMC Biotechnology, 18(1). https://doi.org/10.1186/s12896-018-0429-y

Jaswal, K., Shrivastava, M., Roy, D., Agrawal, S., & Chaba, R. (2020). Metabolism of long-chain fatty acids affects disulphide bond formation in escherichia coli and activates envelope stress response pathways as a combat strategy. PLoS Genetics, 16(10). https://doi.org/10.1371/journal.pgen.1009081

Jayaraj, R., & Smooker, P. M. (2014). So you Need a Protein - A Guide to the Production of Recombinant Proteins. The Open Veterinary Science Journal, 3(1). https://doi.org/10.2174/1874318809003010028

Jennings, M. P., & Beacham, I. R. (1990). Analysis of the Escherichia coli Gene Encoding L-Asparaginase II, ansB, and Its Regulation by Cyclic AMP Receptor and FNR Proteins. In JOURNAL OF BACTERIOLOGY (Vol. 172, Issue 3). https://journals.asm.org/journal/jb

Jensen, K. (1997). A Brief Introduction to Coloured Petri nets.

Jia, B., & Jeon, C. O. (2016). High-throughput recombinant protein expression in Escherichia coli: Current status and future perspectives. In Open Biology (Vol. 6, Issue 8). Royal Society of London. https://doi.org/10.1098/rsob.160196

Jiang, C., Wynne, M., & Huber, D. (2021). How Quality Control Systems AID Sec-Dependent Protein Translocation. In Frontiers in Molecular Biosciences (Vol. 8). Frontiers Media S.A. https://doi.org/10.3389/fmolb.2021.669376

Johnson, G. E., Lalanne, J. B., Peters, M. L., & Li, G. W. (2020). Functionally uncoupled transcription–translation in Bacillus subtilis. Nature, 585(7823), 124–128. https://doi.org/10.1038/s41586-020-2638-5

Jomaa, A., Boehringer, D., Leibundgut, M., & Ban, N. (2016). Structures of the E. coli translating ribosome with SRP and its receptor and with the translocon. Nature Communications, 7. https://doi.org/10.1038/ncomms10471

Jomaa, A., Fu, Y. H. H., Boehringer, D., Leibundgut, M., Shan, S. O., & Ban, N. (2017). Structure of the quaternary complex between SRP, SR, and translocon bound to the translating ribosome. Nature Communications, 8. https://doi.org/10.1038/ncomms15470

Jonda, S., Huber-Wunderlich, M., Glockshuber, R., & Mössner, E. (1999). Complementation of DsbA deficiency with secreted thioredoxin variants reveals the crucial role of an efficient dithiol oxidant for catalyzed protein folding in the bacterial periplasm. EMBO Journal, 18(12). https://doi.org/10.1093/emboj/18.12.3271

Jong, H. de. (2008). Computational Methods in Systems Biology. In Lecture notes in computer science. Springer Science+Business Media. https://doi.org/10.1007/978-3-540-88562-7

Jong, W. S. P., ten Hagen-Jongman, C. M., Vikström, D., Dontje, W., Abdallah, A. M., de Gier, J. W., Bitter, W., & Luirink, J. (2020). Mutagenesis-Based Characterization and Improvement of a Novel Inclusion Body Tag. Frontiers in Bioengineering and Biotechnology, 7. https://doi.org/10.3389/fbioe.2019.00442

Jong, W. S. P., Vikström, D., Houben, D., Berg van Saparoea, H. B., Gier, J. W., & Luirink, J. (2017). Application of an E. coli signal sequence as a versatile inclusion body tag. Microbial Cell Factories, 16(1). https://doi.org/10.1186/s12934-017-0662-4

Juhh, G. (1998). The essence of Petri nets and transition systems through Abelian groups. In Electronic Notes in Theoretical Computer Science (Vol. 18).

Jung, S., Bader, V., Natriashvili, A., Koch, H. G., Winklhofer, K. F., & Tatzelt, J. (2020a). SecY-mediated quality control prevents the translocation of non-gated porins. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-73185-y

Jung, S., Bader, V., Natriashvili, A., Koch, H. G., Winklhofer, K. F., & Tatzelt, J. (2020b). SecY-mediated quality control prevents the translocation of non-gated porins. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-73185-y

Junne, T., Kocik, L., & Spiess, M. (2010). The Hydrophobic Core of the Sec61 Translocon Defines the Hydrophobicity Threshold for Membrane Integration. Molecular Biology of the Cell, 21, 1662–1670. https://doi.org/10.1091/mbc.E10

Jurado, P., Ritz, D., Beckwith, J., De Lorenzo, V., & Fernández, L. A. (2002). Production of functional single-chain Fv antibodies in the cytoplasm of Escherichia coli. Journal of Molecular Biology, 320(1). https://doi.org/10.1016/S0022-2836(02)00405-9

Jyotish, N. K., Singh, L. K., & Kumar, C. (2023). Performance measurement of safety-critical systems based on ordinary differential equations and Petri nets: A case study of nuclear power plant. Nuclear Engineering and Technology, 55(3), 861–869. https://doi.org/10.1016/j.net.2022.11.015

Kaderabkova, N., Bharathwaj, M., Furniss, R. C. D., Gonzalez, D., Palmer, T., & Mavridou, D. A. I. (2022). The biogenesis of β-lactamase enzymes. In Microbiology (United Kingdom) (Vol. 168, Issue 8). Microbiology Society. https://doi.org/10.1099/mic.0.001217

Kadokura, H., Katzen, F., & Beckwith, J. (2003). Protein Disulfide Bond Formation in Prokaryotes. *Annual Review of Biochemistry*, *72*(1), 111–135. https://doi.org/10.1146/annurev.biochem.72.121801.161459

Kalyani, D. C., Munk, L., Mikkelsen, J. D., & Meyer, A. S. (2016). Molecular and biochemical characterization of a new thermostable bacterial laccase from Meiothermus ruber DSM 1279. RSC Advances, 6(5). https://doi.org/10.1039/c5ra24374b

Kant, R., Bhatt, G., Patel, V. K., Ganguli, A., Singh, D., Nayak, M., Mishra, K., Gupta, A., Gangopadhyay, K., Gangopadhyay, S., Ramanathan, G., & Bhattacharya, S. (2019). Synchronized Electromechanical Shock Wave-Induced Bacterial Transformation. ACS Omega, 4(5), 8512–8521. https://doi.org/10.1021/acsomega.9b00202

Karamyshev, A. L., & Johnson, A. E. (2005). Selective SecA association with signal sequences in ribosome-bound nascent chains: A potential role for SecA in ribosome targeting to the bacterial membrane. Journal of Biological Chemistry, 280(45). https://doi.org/10.1074/jbc.M509100200

Karim, J. A., Lambert, N. A., & Pioszak, A. A. (2023). Time- and cost-efficient bacterial expression and purification of potato apyrase. Protein Expression and Purification, 203. https://doi.org/10.1016/j.pep.2022.106215

Karyolaimos, A., Dolata, K. M., Antelo-Varela, M., Mestre Borras, A., Elfageih, R., Sievers, S., Becher, D., Riedel, K., & de Gier, J. W. (2020). Escherichia coli Can Adapt Its Protein Translocation Machinery for Enhanced Periplasmic Recombinant Protein Production. Frontiers in Bioengineering and Biotechnology, 7. https://doi.org/10.3389/fbioe.2019.00465

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications, 80(5), 8091–8126. https://doi.org/10.1007/s11042-020-10139-6

Katzen, F., & Beckwith, J. (2000). Detection and characterization of a protein disulfide bond isomerase in Escherichia coli. Journal of Biological Chemistry, 275(47), 37254–37260. https://doi.org/10.1074/jbc.M004702200

Ke, N., Landeta, C., Wang, X., Boyd, D., Eser, M., & Beckwith, J. (2018). Identification of the thioredoxin partner of vitamin K epoxide reductase in mycobacterial disulphide bond formation. Journal of Bacteriology, 200(16). https://doi.org/10.1128/JB.00137-18

Khow, O., & Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pacific Journal of Tropical Biomedicine, 2(2), 159–162. https://doi.org/10.1016/S2221-1691(11)60213-X

Killelea, T., Dimude, J. U., He, L., Stewart, A. L., Kemm, F. E., Radovčić, M., Ivančić-Baće, I., Rudolph, C. J., & Bolt, E. L. (2023). Cas1-Cas2 physically and functionally interacts with DnaK to modulate CRISPR Adaptation. Nucleic Acids Research, 51(13). https://doi.org/10.1093/nar/gkad473

Kim, J. S., & Kalb, J. W. (1996). Design of experiments: an overview and application example. Medical Device and Diagnostic Industry, 18(3).

Kim, Y. J., Neelamegam, R., Heo, M. A., Edwardraja, S., Paik, H. J., & Lee, S. G. (2008). Improving the productivity of single-chain Fv antibody against c-Met by rearranging the order of its variable domains. Journal of Microbiology and Biotechnology, 18(6).

Kim, Y. M., Petzold, C. J., Kerkhoven, E. J., & Baker, S. E. (2021). Editorial: Multi-Omics Technologies for Optimizing Synthetic Biomanufacturing. In Frontiers in Bioengineering and Biotechnology (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2021.818010

Kleijn, H. C. M., Spieksma, F. M., & Van De Nes, M. (2013). Coverability and Extended Petri nets. https://www.universiteitleiden.nl/binaries/content/assets/science/mi/scripties/bachvandenes.pdf

Knoke, L. R., Zimmermann, J., Lupilov, N., Schneider, J. F., Celebi, B., Morgan, B., & Leichert, L. I. (2023). The role of glutathione in periplasmic redox homeostasis and oxidative protein folding in Escherichia coli. Redox Biology, 64. https://doi.org/10.1016/j.redox.2023.102800

Koch, H. G., Hengelage, T., Neumann-Haefelin, C., MacFarlane, J., Hoffschulte, H. K., Schimz, K. L., Mechler, B., & Müller, M. (1999). In vitro studies with purified components reveal signal recognition particle (SRP) and SecA/SecB as constituents of two independent protein- targeting pathways of Escherichia coli. Molecular Biology of the Cell, 10(7). https://doi.org/10.1091/mbc.10.7.2163

Koch, I. (2014). Petri nets in systems biology. Software & Systems Modeling, 14(2), 703–710. https://doi.org/10.1007/s10270-014-0421-5

Kojima, S., & Nikaido, H. (2014). High salt concentrations increase permeability through ompc channels of escherichia coli. Journal of Biological Chemistry, 289(38), 26464–26473. https://doi.org/10.1074/jbc.M114.585869

Kol, S., Majczak, W., Heerlien, R., van der Berg, J. P., Nouwen, N., & Driessen, A. J. M. (2009). Subunit a of the F1F0 ATP Synthase Requires YidC and SecYEG for Membrane Insertion. Journal of Molecular Biology, 390(5). https://doi.org/10.1016/j.jmb.2009.05.074

Kolivand, S., Nazari, M., Modarressi, M. H., Najafabadi, M. R. H., Hemati, A., Ghafouri-Fard, S., & Motevaseli, E. (2020). Optimized protocol for soluble prokaryotic expression, purification and refolding of the human inhibin α subunit, a cysteine rich peptide chain. Human Antibodies, 28(2). https://doi.org/10.3233/HAB-190399

Kolkman, M. A. B., Van Der Ploeg, R., Bertels, M., Van Dijk, M., Van Der Laan, J., Van Dijl, J. M., & Ferrari, E. (2008). The twin-arginine signal peptide of Bacillus subtilis YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: Secretion of active subtilisin via the B. subtilis Tat pathway. Applied and Environmental Microbiology, 74(24). https://doi.org/10.1128/AEM.01401-08

Kong, B., & Guo, G. L. (2014). Soluble expression of disulphide bond containing proteins FGF15 and FGF19 in the cytoplasm of Escherichia coli. PLoS ONE, 9(1). https://doi.org/10.1371/journal.pone.0085890

Kouwen, T. R. H. M., Van Der Goot, A., Dorenbos, R., Winter, T., Antelmann, H., Plaisier, M. C., Quax, W. J., Van Dijl, J. M., & Dubois, J. Y. F. (2007). Thiol-disulphide oxidoreductase modules in the low-GC Gram-positive bacteria. Molecular Microbiology, 64(4). https://doi.org/10.1111/j.1365-2958.2007.05707.x

Kova´rova´-, K., Kova´rova, K., Kova´rova´-Kovar, K., & Egli, T. (1998). Growth Kinetics of Suspended Microbial Cells: From Single-Substrate-Controlled Growth to Mixed-Substrate Kinetics. In MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS (Vol. 62, Issue 3).

Kpadeh, Z. Z., Day, S. R., Mills, B. W., & Hoffman, P. S. (2015). Legionella pneumophila utilizes a single-player disulphide-bond oxidoreductase system to manage disulphide bond formation and isomerization. Molecular Microbiology, 95(6). https://doi.org/10.1111/mmi.12914

Kurokawa, Y., Yanagi, H., & Yura, T. (2000). Overexpression of Protein Disulphide Isomerase DsbC Stabilizes Multiple-Disulphide-Bonded Recombinant Protein Produced and Transported to the Periplasm in Escherichia coli. In APPLIED AND ENVIRONMENTAL MICROBIOLOGY (Vol. 66, Issue 9).

Kurokawa, Y., Yanagi, H., & Yura, T. (2001). Overproduction of Bacterial Protein Disulphide Isomerase (DsbC) and Its Modulator (DsbD) Markedly Enhances Periplasmic Production of Human Nerve Growth Factor in Escherichia coli. Journal of Biological Chemistry, 276(17). https://doi.org/10.1074/jbc.M100132200

Kuzniatsova, L., Winstone, T. M. L., & Turner, R. J. (2016). Identification of protein-protein interactions between the TatB and TatC subunits of the twin-arginine translocase system and respiratory enzyme specific chaperones. Biochimica et Biophysica Acta - Biomembranes, 1858(4). https://doi.org/10.1016/j.bbamem.2016.01.025

Kyratsous, C. A., Silverstein, S. J., DeLong, C. R., & Panagiotidis, C. A. (2009). Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in Escherichia coli. Gene, 440(1–2), 9–15. https://doi.org/10.1016/j.gene.2009.03.011

Landeta, C., Blazyk, J. L., Hatahet, F., Meehan, B. M., Eser, M., Myrick, A., Bronstain, L., Minami, S., Arnold, H., Ke, N., Rubin, E. J., Furie, B. C., Furie, B., Beckwith, J., Dutton, R., & Boyd, D. (2015). Compounds

targeting disulphide bond forming enzyme DsbB of Gram-negative bacteria. Nature Chemical Biology, 11(4). https://doi.org/10.1038/nchembio.1752

Lee, C. A., & Saier, M. H. (1983). Mannitol-specific Enzyme II of the Bacterial Phosphotransferase System. The Journal of Biological Chemistry, 258(17).

Lee, J. M., Gianchandani, E. P., & Papin, J. A. (2006). Flux balance analysis in the era of metabolomics. In Briefings in Bioinformatics (Vol. 7, Issue 2, pp. 140–150). https://doi.org/10.1093/bib/bbl007

Lee, P. A., Tullman-Ercek, D., & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. In Annual Review of Microbiology (Vol. 60, pp. 373–395). https://doi.org/10.1146/annurev.micro.60.080805.142212

Lee, S. F., Li, L., Jalal, N., & Halperin, S. A. (2021). Identification of a thiol-disulphide oxidoreductase (sdba) catalyzing disulphide bond formation in the superantigen spea in streptococcus pyogenes. Journal of Bacteriology, 203(17). https://doi.org/10.1128/JB.00153-21

Lee Upton, S., Tay, J. W., Schwartz, D. K., & Sousa, M. C. (2023). Similarly slow diffusion of BAM and SecYEG complexes in live E. coli cells observed with 3D spt-PALM. Biophysical Journal, 122(22). https://doi.org/10.1016/j.bpj.2023.10.017

Lefebvre, D., & Basile, F. (2018). Design of control sequences for timed Petri nets based on tree encoding. 51(7), 218–223. https://doi.org/10.1016/j.ifacol.2018.06.304

Lema-Perez, L., Muñoz-Tamayo, R., Garcia-Tirado, J., & Alvarez, H. (2019). On parameter interpretability of phenomenological-based semiphysical models in biology. Informatics in Medicine Unlocked, 15. https://doi.org/10.1016/j.imu.2019.02.002

Lembke, M., Höfler, T., Walter, A. N., Tutz, S., Fengler, V., Schild, S., & Reidl, J. (2020). Host stimuli and operator binding sites controlling protein interactions between virulence master regulator ToxR and ToxS in Vibrio cholerae. Molecular Microbiology, 114(2). https://doi.org/10.1111/mmi.14510

Li, C., Ge, Q. W., Nakata, M., Matsuno, H., & Miyano, S. (2007). Modelling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets. Journal of Biosciences, 32(1). https://doi.org/10.1007/s12038-007-0011-6

Li, D., Ji, F., Huang, C., & Jia, L. (2019). High expression achievement of active and robust Anti-β2 microglobulin Nanobodies via E.coli Hosts Selection. Molecules, 24(16). https://doi.org/10.3390/molecules24162860

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Y. mei, Tian, Z. wei, Xu, D. hua, Wang, X. yin, & Wang, T. yun. (2018). Construction strategies for developing expression vectors for recombinant monoclonal antibody production in CHO cells. In Molecular Biology Reports (Vol. 45, Issue 6, pp. 2907–2912). Springer netherlands. https://doi.org/10.1007/s11033-018-4351-0

Lindenstrauß, U., & Brüser, T. (2009). Tat transport of linker-containing proteins in Escherichia coli. FEMS Microbiology Letters, 295(1), 135–140. https://doi.org/10.1111/j.1574-6968.2009.01600.x

Liu, C., Gong, J. S., Su, C., Li, H., Li, H., Rao, Z. M., Xu, Z. H., & Shi, J. S. (2022). Pathway engineering facilitates efficient protein expression in Pichia pastoris. In Applied Microbiology and Biotechnology (Vol. 106, Issue 18). https://doi.org/10.1007/s00253-022-12139-y

Liu, F., & Heiner, M. (2013). Modeling membrane systems using colored stochastic Petri nets. Natural Computing, 12(4), 617–629. https://doi.org/10.1007/s11047-013-9367-8

Liu, J., Yin, F., Liu, T., Li, S., Tan, C., Li, L., Zhou, R., & Huang, Q. (2020). The Tat system and its dependent cell division proteins are critical for virulence of extra-intestinal pathogenic Escherichia coli. Virulence, 11(1), 1279–1292. https://doi.org/10.1080/21505594.2020.1817709

Liu, S., Skory, C., Qureshi, N., & Hughes, S. (2016). The yajC gene from Lactobacillus buchneri and Escherichia coli and its role in ethanol tolerance. Journal of Industrial Microbiology and Biotechnology, 43(4). https://doi.org/10.1007/s10295-015-1730-6

Liu, Y., Li, X., Luo, J., Su, T., Si, M., & Chen, C. (2021). A novel mycothiol-dependent thiol–disulphide reductase in Corynebacterium glutamicum involving oxidative stress resistance. 3 Biotech, 11(8). https://doi.org/10.1007/s13205-021-02896-4

Lladó, C. M., & Lladó, L. (2022). PIPE 2.7 overview A Petri net tool for performance modeling and evaluation.

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2016). Erratum to: SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulphide bonded proteins in its cytoplasm. Microbial Cell Factories, 15(1), 1–16. https://doi.org/10.1186/s12934-016-0512-9

Lomazova, I. A., & Popova-Zeugmann, L. (2015). Controlling Petri net Behavior Using Time Constraints. ResearchGate. https://doi.org/10.13140/RG.2.1.5172.8086

Long, X., Gou, Y., Luo, M., Zhang, S., Zhang, H., Bai, L., Wu, S., He, Q., Chen, K., Huang, A., Zhou, J., & Wang, D. (2015). Soluble expression, purification, and characterization of active recombinant human tissue plasminogen activator by auto-induction in E. coli. BMC Biotechnology, 15(1). https://doi.org/10.1186/s12896-015-0127-y

Low, K. O., Mahadi, N. M., & Illias, R. M. (2013). Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. In Applied Microbiology and Biotechnology (Vol. 97, Issue 9, pp. 3811–3826). Springer Verlag. https://doi.org/10.1007/s00253-013-4831-z

Lycklama a Nijeholt, J. A., & Driessen, A. J. M. (2012). The bacterial Sec-translocase: Structure and mechanism. In Philosophical Transactions of the Royal Society B: Biological Sciences (Vol. 367, Issue 1592, pp. 1016–1028). Royal Society. https://doi.org/10.1098/rstb.2011.0201

Mahulea, C., Ramírez-Treviño, A., Recalde, L., & Silva, M. (2008). Steady-state control reference and token conservation laws in continuous petri net systems. IEEE Transactions on Automation Science and Engineering, 5(2), 307–320. https://doi.org/10.1109/TASE.2007.893504

Mahulea, C., Recalde, L., & Silva, M. (2005). OPTIMAL OBSERVABILITY FOR CONTINUOUS PETRI NETS.

Maksum, I. P., Utami, D. F., Nurhakim, E. A., Yosua, Sriwidodo, Yusuf, M., Fadhlillah, M., & Haryanto, R. A. (2022). Overexpression of soluble recombinant Thermus thermophilus (Tth) DNA polymerase in Escherichia coli BL21(DE3) using an MBP fusion tag as a solubility enhancer. Journal of Applied Pharmaceutical Science, 12(9). https://doi.org/10.7324/JAPS.2022.120903

Makumire, S., Revaprasadu, N., & Shonhai, A. (2015). DnaK protein alleviates toxicity induced by citrate-coated gold nanoparticles in escherichia coli. PLoS ONE, 10(4). https://doi.org/10.1371/journal.pone.0121243

Maleki, M., & Hajihassan, Z. (2021). De Novo Designing a Novel Signal Peptide for Secretion of Neurturin to the Periplasmic Space of Escherichia coli. Applied Biochemistry and Microbiology, 57. https://doi.org/10.1134/S0003683821100057

Malojčić, G., Owen, R. L., & Glockshuber, R. (2014). Structural and mechanistic insights into the PAPS-independent sulfotransfer catalyzed by bacterial aryl sulfotransferase and the role of the DsbL/DsbI system in its folding. Biochemistry, 53(11). https://doi.org/10.1021/bi401725j

Mamipour, M., Yousefi, M., & Hasanzadeh, M. (2017). An overview on molecular chaperones enhancing solubility of expressed recombinant proteins with correct folding. In International Journal of Biological Macromolecules (Vol. 102, pp. 367–375). Elsevier B.V. https://doi.org/10.1016/j.ijbiomac.2017.04.025

Maneewannakul, S., Maneewannakul, K., & Ippen-Ihler, K. (1994). The pKSM710 Vector Cassette Provides Tightly Regulated lac and T7lac Promoters and Strategies for Manipulating N-Terminal Protein Sequences. Plasmid, 31(3). https://doi.org/10.1006/plas.1994.1032

Mansilla, M. C., & De Mendoza, D. (2000). The Bacillus subtilis cysP gene encodes a novel sulphate permease related to the inorganic phosphate transporter (Pit) family. In Microbiology (Vol. 146).

Manta, B., Boyd, D., & Berkmen, M. (2019). Disulphide Bond Formation in the Periplasm of Escherichia coli . EcoSal Plus, 8(2). https://doi.org/10.1128/ecosalplus.esp-0012-2018

Mao, C., Bariya, P., Suo, Y., & Randall, L. L. (2020). Comparison of Single and Multiple Turnovers of SecYEG in Escherichia coli. Journal of Bacteriology, 202(24). https://doi.org/10.1128/JB.00462-20

Marcyjaniak, M., Odintsov, S. G., Sabala, I., & Bochtler, M. (2004). Peptidoglycan amidase MepA is a LAS metallopeptidase. Journal of Biological Chemistry, 279(42), 43982–43989. https://doi.org/10.1074/jbc.M406735200

Marion, G., & Scotland, S. (2008). An Introduction to Mathematical Modelling.

Martínez-Alonso, M., García-Fruitós, E., Ferrer-Miralles, N., Rinas, U., & Villaverde, A. (2010). Side effects of chaperone gene co-expression in recombinant protein production. http://www.microbialcellfactories.com/content/9/1/64

Martins-Santana, L., Nora, L. C., Sanches-Medeiros, A., Lovate, G. L., Murilo, M. H., & Silva-Rocha, R. (2018). Systems and synthetic biology approaches to engineer fungi for fine chemical production. In Frontiers in Bioengineering and Biotechnology (Vol. 6, Issue OCT). https://doi.org/10.3389/fbioe.2018.00117

Mas, A., & Portillo, M. C. (2022). Strategies for microbiological control of the alcoholic fermentation in wines by exploiting the microbial terroir complexity: A mini-review. In International Journal of Food Microbiology (Vol. 367). https://doi.org/10.1016/j.ijfoodmicro.2022.109592

Masip, L., Veeravalli, K., & Georgiou, G. (2006). The many faces of glutathione in bacteria. In Antioxidants and Redox Signaling (Vol. 8, Issues 5–6). https://doi.org/10.1089/ars.2006.8.753

Mathias, S., Wippermann, A., Raab, N., Zeh, N., Handrick, R., Gorr, I., Schulz, P., Fischer, S., Gamer, M., & Otte, K. (2020). Unraveling what makes a monoclonal antibody difficult-to-express: From intracellular accumulation to incomplete folding and degradation via ERAD. Biotechnology and Bioengineering, 117(1). https://doi.org/10.1002/bit.27196

Maton, M., Bogaerts, P., & Vande Wouwer, A. (2022). Hybrid Dynamic Models of Bioprocesses Based on Elementary Flux Modes and Multilayer Perceptrons. Processes, 10(10). https://doi.org/10.3390/pr10102084

Matos, C. F. R. O., Branston, S. D., Albiniak, A., Dhanoya, A., Freedman, R. B., Keshavarz-Moore, E., & Robinson, C. (2012). ARTICLE High-Yield Export of a Native Heterologous Protein to the Periplasm by the Tat Translocation Pathway in Escherichia coli. Biotechnol. Bioeng, 109, 2533–2542. https://doi.org/10.1002/bit.24535/abstract

Matos, C. F. R. O., Di Cola, A., & Robinson, C. (2009). TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in Escherichia coli. EMBO Reports, 10(5), 474–479. https://doi.org/10.1038/embor.2009.34

Matos, C. F. R. O., Robinson, C., Alanen, H. I., Prus, P., Uchida, Y., Ruddock, L. W., Freedman, R. B., & Keshavarz-Moore, E. (2014). Efficient export of prefolded, disulphide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. Biotechnology Progress, 30(2). https://doi.org/10.1002/btpr.1858

Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M., & Miyano, S. (2003). Biopathways representation and simulation on hybrid functional Petri net. In Silico Biology, 3(3), 389–404. https://pubmed.ncbi.nlm.nih.gov/12954096/

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. In Cellular and Molecular Life Sciences (Vol. 62, Issue 6, pp. 670–684). https://doi.org/10.1007/s00018-004-4464-6

Mccarthy, A. A., Haebel, P. W., Törrönen, A., Rybin, V., Baker, E. N., & Metcalf, P. (2000). Crystal structure of the protein disulphide bond isomerase, DsbC, from Escherichia coli. http://structbio.nature.com

McGinness, K. E., Baker, T. A., & Sauer, R. T. (2006). Engineering Controllable Protein Degradation. Molecular Cell, 22(5). https://doi.org/10.1016/j.molcel.2006.04.027

McGinnis, S., & Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research, 32(WEB SERVER ISS.). https://doi.org/10.1093/nar/gkh435

Meehan, B. M., Landeta, C., Boyd, D., & Beckwith, J. (2017). The disulphide bond formation pathway is essential for anaerobic growth of Escherichia coli. Journal of Bacteriology, 199(16). https://doi.org/10.1128/JB.00120-17

Mermans, D., Nicolaus, F., Fleisch, K., & Von Heijne, G. (2022). Cotranslational folding and assembly of the dimeric Escherichia coli inner membrane protein EmrE. https://doi.org/10.1073/pnas

Miyazaki, R., Ai, M., Tanaka, N., Suzuki, T., Dhomae, N., Tsukazaki, T., Akiyama, Y., & Mori, H. (2022). Inner membrane YfgM–PpiD heterodimer acts as a functional unit that associates with the SecY/E/G translocon and promotes protein translocation. Journal of Biological Chemistry, 298(11). https://doi.org/10.1016/j.jbc.2022.102572

Mizrachi, D., Robinson, M. P., Ren, G., Ke, N., Berkmen, M., & Delisa, M. P. (2017). A water-soluble DsbB variant that catalyzes disulphide-bond formation in vivo. In Nature Chemical Biology (Vol. 13, Issue 9). https://doi.org/10.1038/nchembio.2409

Moffat, J., & Sabatini, D. M. (2006). Building mammalian signalling pathways with RNAi screens. In Nature Reviews Molecular Cell Biology (Vol. 7, Issue 3). https://doi.org/10.1038/nrm1860

Mollet, M., Ma, N., Zhao, Y., Brodkey, R., Taticek, R., & Chalmers, J. J. (2004). Bioprocess equipment: Characterization of energy dissipation rate and its potential to damage cells. Biotechnology Progress, 20(5), 1437–1448. https://doi.org/10.1021/bp0498488

Mogk, A. (1999). Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. The EMBO Journal, 18(24), 6934–6949. https://doi.org/10.1093/emboj/18.24.6934

Müller, A., Hoffmann, J. H., Meyer, H. E., Narberhaus, F., Jakob, U., & Leichert, L. I. (2013). Nonnative disulphide bond formation activates the σ32-dependent heat shock response in escherichia coli. Journal of Bacteriology, 195(12), 2807–2816. https://doi.org/10.1128/JB.00127-13

Muller, J. P. (1999). Effects of pre-protein overexpression on SecB synthesis in Escherichia coli . FEMS Microbiology Letters, 176(1), 219–227. https://doi.org/10.1111/j.1574-6968.1999.tb13665.x

Nagpal, S., Tiwari, S., Mapa, K., & Thukral, L. (2015). Decoding Structural Properties of a Partially Unfolded Protein Substrate: En Route to Chaperone Binding. PLoS Computational Biology, 11(9). https://doi.org/10.1371/journal.pcbi.1004496

Nakamoto, H., & Bardwell, J. C. A. (2004). Catalysis of disulphide bond formation and isomerization in the Escherichia coli periplasm. In Biochimica et Biophysica Acta - Molecular Cell Research (Vol. 1694, Issues 1-3 SPEC.ISS.). https://doi.org/10.1016/j.bbamcr.2004.02.012

Nașcu, I., Sebastia-Saez, D., Chen, T., Nașcu, I., & Du, W. (2022). Global sensitivity analysis for a perfusion bioreactor based on CFD modelling. Computers and Chemical Engineering, 163. https://doi.org/10.1016/j.compchemeng.2022.107829

Naz, Z. (2015). Introduction to Biotechnology. University of Punjab . https://doi.org/10.13140/RG.2.1.3517.8968

Nelson, E. C., & Elisha, B. G. (1999). Molecular Basis of AmpC Hyperproduction in Clinical Isolates of Escherichia coli (Vol. 43, Issue 4). https://journals.asm.org/journal/aac

Nguyen, M. T., Koo, B. K., Thi Vu, T. T., Song, J. A., Chong, S. H., Jeong, B., Ryu, H. B., Moh, S. H., & Choe, H. (2014). Prokaryotic soluble overexpression and purification of bioactive human growth hormone by fusion to thioredoxin, maltose binding protein, and protein disulphide isomerase. PLoS ONE, 9(3). https://doi.org/10.1371/journal.pone.0089038

Nguyen, V. D., Hatahet, F., Salo, K. E. H., Enlund, E., Zhang, C., & Ruddock, L. W. (2011). Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulphide bond containing proteins expressed in the cytoplasm of E.coli. Microbial Cell Factories, 10. https://doi.org/10.1186/1475-2859-10-1

Nishihara, K., Kanemori, M., Yanagi, H., & Yura, T. (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. Applied and Environmental Microbiology, 66(3). https://doi.org/10.1128/AEM.66.3.884-889.2000

Nishikino, T., Hijikata, A., Kojima, S., Shirai, T., Kainosho, M., Homma, M., & Miyanoiri, Y. (2023). Changes in the hydrophobic network of the FliGMC domain induce rotational switching of the flagellar motor. IScience, 26(8). https://doi.org/10.1016/j.isci.2023.107320

Nogueira, D. E. S., Cabral, J. M. S., & Rodrigues, C. A. V. (2021). Single-use bioreactors for human pluripotent and adult stem cells: Towards regenerative medicine applications. In Bioengineering (Vol. 8, Issue 5). https://doi.org/10.3390/BIOENGINEERING8050068

Norambuena, J., Flores, R., Cárdenas, J. P., Quatrini, R., Chávez, R., & Levicán, G. (2012). Thiol/Disulphide System Plays a Crucial Role in Redox Protection in the Acidophilic Iron-Oxidizing Bacterium Leptospirillum ferriphilum. PLoS ONE, 7(9). https://doi.org/10.1371/journal.pone.0044576

Nordstrand, K., Åslund, F., Holmgren, A., Otting, G., & Berndt, K. D. (1999). NMR structure of Escherichia coli glutaredoxin 3-glutathione mixed disulphide complex: Implications for the enzymatic mechanism. Journal of Molecular Biology, 286(2). https://doi.org/10.1006/jmbi.1998.2444

Nosaki, S., Hoshikawa, K., Ezura, H., & Miura, K. (2021). Transient protein expression systems in plants and their applications. In Plant Biotechnology (Vol. 38, Issue 3, pp. 297–304). Japanese Society for Plant Cell and Molecular Biology. https://doi.org/10.5511/plantbiotechnology.21.0610a

Nouwen, N., & Driessen, A. J. M. (2002). SecDFyajC forms a heterotetrameric complex with YidC. Molecular Microbiology, 44(5), 1397–1405. https://doi.org/10.1046/j.1365-2958.2002.02972.x

Oertel, D., Schmitz, S., & Freudl, R. (2015). A TatABC-type Tat translocase Is required for unimpaired aerobic growth of Corynebacterium glutamicum ATCC13032. PLoS ONE, 10(4). https://doi.org/10.1371/journal.pone.0123413

Ohta, N., Kato, Y., Watanabe, H., Mori, H., & Matsuura, T. (2016). In vitro membrane protein synthesis inside Sec translocon-reconstituted cell-sized liposomes. Scientific Reports, 6. https://doi.org/10.1038/srep36466

Öjemalm, K., Botelho, S. C., Stüdle, C., & Von Heijne, G. (2013). Quantitative analysis of SecYEG-mediated insertion of transmembrane α-helices into the bacterial inner membrane. Journal of Molecular Biology, 425(15). https://doi.org/10.1016/j.jmb.2013.04.025

Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? In Nature Biotechnology (Vol. 28, Issue 3, pp. 245–248). https://doi.org/10.1038/nbt.1614

Østergaard, H., Henriksen, A., Hansen, F. G., & Winther, J. R. (2001). Shedding light on disulphide bond formation: Engineering a redox switch in green fluorescent protein. EMBO Journal, 20(21). https://doi.org/10.1093/emboj/20.21.5853

Packiam, K. A. R., Ramanan, R. N., Ooi, C. W., Krishnaswamy, L., & Tey, B. T. (2020). Stepwise optimization of recombinant protein production in Escherichia coli utilizing computational and experimental approaches. In Applied Microbiology and Biotechnology (Vol. 104, Issue 8, pp. 3253–3266). Springer. https://doi.org/10.1007/s00253-020-10454-w

Palmer, T., & Berks, B. C. (2012). The twin-arginine translocation (Tat) protein export pathway. In Nature Reviews Microbiology (Vol. 10, Issue 7, pp. 483–496). https://doi.org/10.1038/nrmicro2814

Papanikou, E., Karamanou, S., & Economou, A. (2007). Bacterial protein secretion through the translocase nanomachine. In Nature Reviews Microbiology (Vol. 5, Issue 11, pp. 839–851). https://doi.org/10.1038/nrmicro1771

Paolo Baldan, Cocco, N., Marin, A., & Simeoni, M. (2010). Petri nets for modelling metabolic pathways: a survey. Natural Computing, 9(4), 955–989. https://doi.org/10.1007/s11047-010-9180-6

Patel, R., Smith, S. M., & Robinson, C. (2014). Protein transport by the bacterial Tat pathway. In Biochimica et Biophysica Acta - Molecular Cell Research (Vol. 1843, Issue 8, pp. 1620–1628). Elsevier. https://doi.org/10.1016/j.bbamcr.2014.02.013

Peleg, Y., & Unger, T. (2012). Resolving bottlenecks for recombinant protein expression in E. coli. Methods in Molecular Biology, 800, 173–186. https://doi.org/10.1007/978-1-61779-349-3_12

Pérez-Rodríguez, R., Fisher, A. C., Perlmutter, J. D., Hicks, M. G., Chanal, A., Santini, C. L., Wu, L. F., Palmer, T., & DeLisa, M. P. (2007). An Essential Role for the DnaK Molecular Chaperone in Stabilizing Over-expressed Substrate Proteins of the Bacterial Twin-arginine Translocation Pathway. Journal of Molecular Biology, 367(3). https://doi.org/10.1016/j.jmb.2007.01.027

Petriman, N. A., Jauß, B., Hufnagel, A., Franz, L., Sachelaru, I., Drepper, F., Warscheid, B., & Koch, H. G. (2018). The interaction network of the YidC insertase with the SecYEG translocon, SRP and the SRP receptor FtsY. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-017-19019-w

Pittman, M. S., Robinson, H. C., & Poole, R. K. (2005). A bacterial glutathione transporter (Escherichia coli CydDC) exports reductant to the periplasm. Journal of Biological Chemistry, 280(37). https://doi.org/10.1074/jbc.M503075200

Pommereau, F. (2015). SNAKES: A flexible high-level Petri nets library. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9115. https://doi.org/10.1007/978-3-319-19488-2_13

Pop, O., Martin, U., Abel, C., & Müller, J. P. (2002). The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of Bacillus subtilis form an autonomous tat translocation system. Journal of Biological Chemistry, 277(5). https://doi.org/10.1074/jbc.M110829200

Pouresmaeil, M., & Azizi-Dargahlou, S. (2023). Factors involved in heterologous expression of proteins in E. coli host. Archives of Microbiology, 205(5). https://doi.org/10.1007/s00203-023-03541-9

Powers, D. N., Wang, Y., Fratz-Berilla, E. J., Velugula-Yellela, S. R., Chavez, B., Angart, P., Trunfio, N., Yoon, S., & Agarabi, C. (2019). Real-time quantification and supplementation of bioreactor amino acids to prolong culture time and maintain antibody product quality. Biotechnology Progress, 35(6). https://doi.org/10.1002/btpr.2894

Powers, E. T., Powers, D. L., & Gierasch, L. M. (2012). FoldEco: A Model for Proteostasis in E. coli. Cell Reports, 1(3). https://doi.org/10.1016/j.celrep.2012.02.011

Pradel, N., Delmas, J., Wu, L. F., Santini, C. L., & Bonnet, R. (2009). Sec- and Tat-dependent translocation of β-lactamases across the Escherichia coli inner membrane. Antimicrobial Agents and Chemotherapy, 53(1). https://doi.org/10.1128/AAC.00642-08

Prahlad, J., Struble, L. R., Lutz, W. E., Wallin, S. A., Khurana, S., Schnaubelt, A., Broadhurst, M. J., Bayles, K. W., & Borgstahl, G. E. O. (2021). CyDisCo production of functional recombinant SARS-CoV-2 spike receptor binding domain. Protein Science, 30(9). https://doi.org/10.1002/pro.4152

Pratama, F., Linton, D., & Dixon, N. (2021). Genetic and process engineering strategies for enhanced recombinant N-glycoprotein production in bacteria. Microbial Cell Factories, 20(1). https://doi.org/10.1186/s12934-021-01689-x

Prinz, W. A., Åslund, F., Holmgren, A., & Beckwith, J. (1997). The role of the thioredoxin and glutaredoxin pathways in reducing protein disulphide bonds in the Escherichia coli cytoplasm. Journal of Biological Chemistry, 272(25). https://doi.org/10.1074/jbc.272.25.15661

Puigbò, P., Guzmán, E., Romeu, A., & Garcia-Vallvé, S. (2007). OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. Nucleic Acids Research, 35(SUPPL.2). https://doi.org/10.1093/nar/gkm219

Ramakrishnan, R., Houben, B., Kreft, Ł., Botzki, A., Schymkowitz, J., & Rousseau, F. (2020). Protein Homeostasis Database: Protein quality control in E.coli. Bioinformatics, 36(3). https://doi.org/10.1093/bioinformatics/btz628

Rauniyar, K., Akhondzadeh, S., Gąciarz, A., Künnapuu, J., & Jeltsch, M. (2022). Bioactive VEGF-C from E. coli. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-22960-0

Ravitchandirane, G., Bandhu, S., & Chaudhuri, T. K. (2022). Multimodal approaches for the improvement of the cellular folding of a recombinant iron regulatory protein in E. coli. Microbial Cell Factories, 21(1). https://doi.org/10.1186/s12934-022-01749-w

Rawson, M., & Rawson, M. (2022). Petri nets for Concurrent Programming. http://arxiv.org/abs/2208.02900

Ray, N., Oates, J., Turner, R. J., & Robinson, C. (2003). DmsD is required for the biogenesis of DMSO reductase in Escherichia coli but not for the interaction of the DmsA signal peptide with the Tat apparatus. FEBS Letters, 534(1–3). https://doi.org/10.1016/S0014-5793(02)03839-5

Raymond, F., Rolland, D., Gauthier, M., & Jolivet, M. (1998). Purification of a recombinant protein expressed in yeast: Optimization of analytical and preparative chromatography. Journal of Chromatography B: Biomedical Applications, 706(1). https://doi.org/10.1016/S0378-4347(97)00512-4

Recalde, L., Haddad, S., & Silva, M. (2007). Continuous Petri nets: Expressive Power and Decidability Issues.

Reddy, R., Syed, P., & Irtaza, A. (2012). Software Testing: A Comparative Study Model Based Testing VS Test Case Based Testing. www.bth.se/com

Reddy Sanganna Gari, R., Chattrakun, K., Marsh, B. P., Mao, C., Chada, N., Randall, L. L., & King, G. M. (2019). Direct visualization of the E. coli Sec translocase engaging precursor proteins in lipid bilayers. In Sci. Adv (Vol. 5). https://www.science.org

Rietsch, A., Bessette, P., Georgiou, G., & Beckwith, J. (1997). Reduction of the periplasmic disulfide bond isomerase, DsbC, occurs by passage of electrons from cytoplasmic thioredoxin. Journal of Bacteriology, 179(21), 6602–6608. https://doi.org/10.1128/jb.179.21.6602-6608.1997

Rist, M., & Kertesz, M. A. (1998). Construction of improved plasmid vectors for promoter characterization in Pseudomonas aeruginosa and other Gram-negative bacteria . FEMS Microbiology Letters, 169(1), 179–183. https://doi.org/10.1111/j.1574-6968.1998.tb13315.x

Robinson, C., Matos, C. F. R. O., Beck, D., Ren, C., Lawrence, J., Vasisht, N., & Mendel, S. (2011). Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. In Biochimica et Biophysica Acta - Biomembranes (Vol. 1808, Issue 3). https://doi.org/10.1016/j.bbamem.2010.11.023

Robson, A., Carr, B., Sessions, R. B., & Collinson, I. (2009). Synthetic peptides identify a second periplasmic site for the plug of the SecYEG protein translocation complex. FEBS Letters, 583(1). https://doi.org/10.1016/j.febslet.2008.12.003

Rodriguez, C., Nam, D. H., Kruchowy, E., & Ge, X. (2017). Efficient Antibody Assembly in E. coli Periplasm by Disulphide Bond Folding Factor Co-expression and Culture Optimization. Applied Biochemistry and Biotechnology, 183(2). https://doi.org/10.1007/s12010-017-2502-8

Rohr, C., Marwan, W., & Heiner, M. (2010). Snoopy-a unifying Petri net framework to investigate biomolecular networks. Bioinformatics, 26(7). https://doi.org/10.1093/bioinformatics/btq050

Rose, P., Fröbel, J., Graumann, P. L., & Müller, M. (2013). Substrate-Dependent Assembly of the Tat Translocase as Observed in Live Escherichia coli Cells. PLoS ONE, 8(8). https://doi.org/10.1371/journal.pone.0069488

Rosenberg, J. N., & Cady, N. C. (2021). Surveilling cellular vital signs: toward label-free biosensors and real-time viability assays for bioprocessing. In Current Opinion in Biotechnology (Vol. 71). https://doi.org/10.1016/j.copbio.2021.07.004

Rouches, M. V., Xu, Y., Cortes, L. B. G., & Lambert, G. (2022). A plasmid system with tunable copy number. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-31422-0

Roussel, G., & White, S. H. (2020). Binding of SecA ATPase monomers and dimers to lipid vesicles. Biochimica et Biophysica Acta - Biomembranes, 1862(2). https://doi.org/10.1016/j.bbamem.2019.183112

Roytrakul, S., Eurwilaichitr, L., Suprasongsin, C., & Panyim, S. (2001). A Rapid and Simple Method for Construction and Expression of a Synthetic Human Growth Hormone Gene in Escherichia coli. Journal of Biochemistry and Molecular Biology, 34(6).

Ruhl, S., de Almeida, N., Carpio, M., Rupprecht, J., Greller, G., & Matuszcyk, J.-C. (2020). A Rapid, Low-Risk Approach for Process Transfer of Biologics from Development to Manufacturing Scale. BioProcess International, 18(5).

Saibil, H. R., Fenton, W. A., Clare, D. K., & Horwich, A. L. (2013). Structure and Allostery of the Chaperonin GroEL. Journal of Molecular Biology, 425(9), 1476–1487. https://doi.org/10.1016/j.jmb.2012.11.028

Saini, P., Beniwal, A., Kokkiligadda, A., & Vij, S. (2018). Response and tolerance of yeast to changing environmental stress during ethanol fermentation. In Process Biochemistry (Vol. 72). https://doi.org/10.1016/j.procbio.2018.07.001

Sala, A., Bordes, P., & Genevaux, P. (2014). Multitasking SecB chaperones in bacteria. In Frontiers in Microbiology (Vol. 5, Issue DEC). https://doi.org/10.3389/fmicb.2014.00666

Samuelson, J. C., Chen, M., Jiang, F., Möller, I., Wiedmann, M., Kuhn, A., Phillips, G. J., & Dalbey, R. E. (2000). YidC mediates membrane protein insertion in bacteria. Nature, 406(6796). https://doi.org/10.1038/35020586

Sandhu, H., Hedman, R., Cymer, F., Kudva, R., Ismail, N., & von Heijne, G. (2021). Cotranslational Translocation and Folding of a Periplasmic Protein Domain in Escherichia coli. Journal of Molecular Biology, 433(15). https://doi.org/10.1016/j.jmb.2021.167047

Sandow, J., Landgraf, W., Becker, R., & Seipke, G. (2015). Equivalent recombinant human insulin preparations and their place in therapy. European Endocrinology, 11(1). https://doi.org/10.17925/ee.2015.11.01.10

Santini, C. L., Ize, B., Chanal, A., Müller, M., Giordano, G., & Wu, L. F. (1998). A novel Sec-independent periplasmic protein translocation pathway in Escherichia coli. EMBO Journal, 17(1), 101–112. https://doi.org/10.1093/emboj/17.1.101

Santos, B. D., Morones-Ramirez, J. R., Balderas-Renteria, I., Casillas-Vega, N. G., Galbraith, D. W., & Zarate, X. (2019). Optimizing Periplasmic Expression in Escherichia coli for the Production of Recombinant Proteins Tagged with the Small Metal-Binding Protein SmbP. Molecular Biotechnology, 61(6). https://doi.org/10.1007/s12033-019-00176-4

Sapriel, G., Wandersman, C., & Delepelaire, P. (2003). The SecB chaperone is bifunctional in Serratia marcescens: SecB is involved in the Sec pathway and required for HasA secretion by the ABC transporter. Journal of Bacteriology, 185(1). https://doi.org/10.1128/JB.185.1.80-88.2003

Sargent, F., Gohlke, U., De Leeuw, E., Stanley, N. R., Palmer, T., Saibil, H. R., & Berks, B. C. (2001). Purified components of the Escherichia coli Tat protein transport system form a double-layered ring structure. European Journal of Biochemistry, 268(12), 3361–3367. https://doi.org/10.1046/j.1432-1327.2001.02263.x

Schierle, C. F., Berkmen, M., Huber, D., Kumamoto, C., Boyd, D., & Beckwith, J. (2003). The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the Escherichia coli periplasm

via the signal recognition particle pathway. Journal of Bacteriology, 185(19), 5706–5713. https://doi.org/10.1128/JB.185.19.5706-5713.2003

Schiffrin, B., Calabrese, A. N., Devine, P. W. A., Harris, S. A., Ashcroft, A. E., Brockwell, D. J., & Radford, S. E. (2016). Skp is a multivalent chaperone of outer-membrane proteins. Nature Structural and Molecular Biology, 23(9). https://doi.org/10.1038/nsmb.3266

Schmidt, K. (2003). Using Petri net Invariants in State Space Construction. https://link.springer.com/content/pdf/10.1007/3-540-36577-x_35.pdf

Schramm, F. D., Heinrich, K., Thüring, M., Bernhardt, J., & Jonas, K. (2017). An essential regulatory function of the DnaK chaperone dictates the decision between proliferation and maintenance in Caulobacter crescentus. PLoS Genetics, 13(12). https://doi.org/10.1371/journal.pgen.1007148

Seidel, S., Mozaffari, F., Maschke, R. W., Kraume, M., Eibl-Schindler, R., & Eibl, D. (2023). Automated Shape and Process Parameter Optimization for Scaling Up Geometrically Non-Similar Bioreactors. Processes, 11(9). https://doi.org/10.3390/pr11092703

Seras-Franzoso, J., Affentranger, R., Ferrer-Navarro, M., Daura, X., Villaverde, A., & García-Fruitósa, E. (2012). Disulphide bond formation and activation of Escherichia coli β-galactosidase under oxidizing conditions. Applied and Environmental Microbiology, 78(7). https://doi.org/10.1128/AEM.06923-11

Sevella, B., & Bertalan, G. (2000). Development of a MATLAB based bioprocess simulation tool.

Seyed Hosseini Fin, N. A., Barshan-tashnizi, M., Sajjadi, S. M., Asgari, S., Mohajerani, N., & Mirzahoseini, H. (2019). The effects of overexpression of cytoplasmic chaperones on secretory production of hirudin-PA in E. coli. Protein Expression and Purification, 157. https://doi.org/10.1016/j.pep.2019.01.011

Sharma, P., Hellingwerf, K. J., Mattos, M. J. T., & Bekker, M. (2012). Uncoupling of substrate-level phosphorylation in escherichia coli: During glucose-limited growth. Applied and Environmental Microbiology, 78(19). https://doi.org/10.1128/AEM.01507-12

Shevchik, V. E., Condemine, G., & Robert-Baudouy, J. (1994). Characterization of DsbC, a periplasmic protein of Erwinia chrysanthemi and Escherichia coli with disulphide isomerase activity. EMBO Journal, 13(8). https://doi.org/10.1002/j.1460-2075.1994.tb06470.x

Sies, H. (1999). Glutathione and its role in cellular functions. Free Radical Biology and Medicine, 27(9–10). https://doi.org/10.1016/S0891-5849(99)00177-X

Siew, Y. Y., & Zhang, W. (2021). Downstream processing of recombinant human insulin and its analogues production from E. coli inclusion bodies. In Bioresources and Bioprocessing (Vol. 8, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s40643-021-00419-w

Siletti, C., & Petrides, D. (2008). Superpro Designer: An Interactive Software Tool for Designing and Evaluating Integrated Chemical, Biochemical, and Environmental Processes. http://www.intelligen.com

Silva, A. J. D., Rocha, C. K. da S., & de Freitas, A. C. (2022). Standardization and Key Aspects of the Development of Whole Yeast Cell Vaccines. In Pharmaceutics (Vol. 14, Issue 12). https://doi.org/10.3390/pharmaceutics14122792

Silva, F., Queiroz, J. A., & Domingues, F. C. (2012). Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. In Biotechnology Advances (Vol. 30, Issue 3). https://doi.org/10.1016/j.biotechadv.2011.12.005

Simeonidis, E., & Price, N. D. (2015). Genome-scale modeling for metabolic engineering. In Journal of Industrial Microbiology and Biotechnology (Vol. 42, Issue 3, pp. 327–338). Springer Verlag. https://doi.org/10.1007/s10295-014-1576-3

Simon, E., Oyekan, J., Hutabarat, W., Tiwari, A., & Turner, C. J. (2018). Adapting petri nets to des: Stochastic modelling of manufacturing systems. International Journal of Simulation Modelling, 17(1), 5–17. https://doi.org/10.2507/IJSIMM17(1)403

Simpson, M. J., Murphy, R. J., & Maclaren, O. J. (2024). Modelling count data with partial differential equation models in biology. Journal of Theoretical Biology, 580. https://doi.org/10.1016/j.jtbi.2024.111732

Singh, P., Sharma, L., Kulothungan, S. R., Adkar, B. V., Prajapati, R. S., Ali, P. S. S., Krishnan, B., & Varadarajan, R. (2013). Effect of Signal Peptide on Stability and Folding of Escherichia coli Thioredoxin. PLoS ONE, 8(5). https://doi.org/10.1371/journal.pone.0063442

Smirnova, G. V., Muzyka, N. G., Glukhovchenko, M. N., & Oktyabrsky, O. N. (2000). Effects of menadione and hydrogen peroxide on glutathione status in growing Escherichia coli. Free Radical Biology and Medicine, 28(7). https://doi.org/10.1016/S0891-5849(99)00256-7

Smirnova, G. V., Muzyka, N. G., Ushakov, V. Y., Tyulenev, A. V., & Oktyabrsky, O. N. (2015). Extracellular superoxide provokes glutathione efflux from Escherichia coli cells. Research in Microbiology, 166(8). https://doi.org/10.1016/j.resmic.2015.07.007

Smith, R. P., Whitten, A. E., Paxman, J. J., Kahler, C. M., Scanlon, M. J., & Heras, B. (2018). Production, biophysical characterization and initial crystallization studies of the N- and C-terminal domains of DsbD, an essential enzyme in Neisseria meningitidis. Acta Crystallographica Section F: Structural Biology Communications, 74(1). https://doi.org/10.1107/S2053230X17017800

Smith, S. M., Yarwood, A., Fleck, R. A., Robinson, C., & Smith, C. J. (2017). TatA complexes exhibit a marked change in organisation in response to expression of the TatBC complex. Biochemical Journal, 474(9). https://doi.org/10.1042/BCJ20160952

Soares, A., Azevedo, A., Gomes, L. C., & Mergulhão, F. J. (2019). Recombinant protein expression in biofilms. In AIMS Microbiology (Vol. 5, Issue 3). https://doi.org/10.3934/microbiol.2019.3.232

Soerjawinata, W., Schlegel, K., Fuchs, N., Schüffler, A., Schirmeister, T., Ulber, R., & Kampeis, P. (2021). Applicability of a single-use bioreactor compared to a glass bioreactor for the fermentation of filamentous fungi and evaluation of the reproducibility of growth in pellet form. Engineering in Life Sciences, 21(5). https://doi.org/10.1002/elsc.202000069

Soltani, M., & Bundy, B. C. (2022). Streamlining cell-free protein synthesis biosensors for use in human fluids: In situ RNase inhibitor production during extract preparation. Biochemical Engineering Journal, 177. https://doi.org/10.1016/j.bej.2021.108158

Son, A., Cabral, V. H., Huang, Z., Litberg, T. J., & Horowitz, S. (2023). G-quadruplexes rescuing protein folding. Proceedings of the National Academy of Sciences of the United States of America, 120(20). https://doi.org/10.1073/pnas.2216308120

Spiesberger, K., Paulfranz, F., Egger, A., Reiser, J., Vogl, C., Rudolf-Scholik, J., Mayrhofer, C., Grosse-Hovest, L., & Brem, G. (2015). Large-scale purification of r28m: A Bispecific Scfv antibody targeting human melanoma produced in transgenic cattle. PLoS ONE, 10(10). https://doi.org/10.1371/journal.pone.0140471

Stanley, N. R., Findlay, K., Berks, B. C., & Palmer, T. (2001). Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. Journal of Bacteriology, 183(1). https://doi.org/10.1128/JB.183.1.139-144.2001

Steinsland, V., Kristensen, L. M., & Zhang, S. (2022). Modelling and Validation of Power Electronics Converter Systems using Coloured Petri nets. http://arxiv.org/abs/2212.06781

Stewart, E. J., Åslund, F., & Beckwith, J. (1998). Disulphide bond formation in the Escherichia coli cytoplasm: An in vivo role reversal for the thioredoxins. EMBO Journal, 17(19), 5543–5550. https://doi.org/10.1093/emboj/17.19.5543

Stewart, E. J., Katzen, F., & Beckwith, J. (1999). Six conserved cysteines of the membrane protein DsbD are required for the transfer of electrons from the cytoplasm to the periplasm of Escherichia coli. In The EMBO Journal (Vol. 18, Issue 21).

Stull, F., Koldewey, P., Humes, J. R., Radford, S. E., & Bardwell, J. C. A. (2016). Substrate protein folds while it is bound to the ATP-independent chaperone Spy. Nature Structural and Molecular Biology, 23(1). https://doi.org/10.1038/nsmb.3133

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. In Bioinformatics and Biology Insights (Vol. 14). SAGE Publications Inc. https://doi.org/10.1177/1177932219899051

Sun, J., Wen, M., Wang, H., Ruan, Y., Yang, Q., Kang, X., Zhang, H., Zhang, Z., & Lu, H. (2022). Prediction of drug-likeness using graph convolutional attention network. Bioinformatics, 38(23). https://doi.org/10.1093/bioinformatics/btac676

Sun, M., Gao, A. X., Liu, X., Yang, Y., Ledesma-Amaro, R., & Bai, Z. (2023). High-throughput process development from gene cloning to protein production. In Microbial Cell Factories (Vol. 22, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12934-023-02184-1

Sutherland, G. A., Grayson, K. J., Adams, N. B. P., Mermans, D. M. J., Jones, A. S., Robertson, A. J., Auman, D. B., Brindley, A. A., Sterpone, F., Tuffery, P., Derreumaux, P., Leslie Dutton, P., Robinson, C., Hitchcock, A., & Neil Hunter, C. (2018). Probing the quality control mechanism of the Escherichia coli twin-arginine translocase with folding variants of a de novo– designed heme protein. Journal of Biological Chemistry, 293(18). https://doi.org/10.1074/jbc.RA117.000880

Swaving, J., Van Wely, K. H. M., & Driessen, A. J. M. (1999). Preprotein translocation by a hybrid translocase composed of Escherichia coli and Bacillus subtilis subunits. Journal of Bacteriology, 181(22). https://doi.org/10.1128/jb.181.22.7021-7027.1999

Tan, E., Chin, C. S. H., Lim, Z. F. S., & Ng, S. K. (2021). HEK293 Cell Line as a Platform to Produce Recombinant Proteins and Viral Vectors. In Frontiers in Bioengineering and Biotechnology (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2021.796991

Tan, Y., Henehan, G. T., Kinsella, G. K., & Ryan, B. J. (2022). Extracellular secretion of a cutinase with polyester-degrading potential by E. coli using a novel signal peptide from Amycolatopsis mediterranei. World Journal of Microbiology and Biotechnology, 38(4). https://doi.org/10.1007/s11274-022-03246-z

Tang, Y., Reed, P., Wagener, T., & van Werkhoven, K. (2007). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Sciences, 11(2), 793–817. https://doi.org/10.5194/hess-11-793-2007

Tao, H., Bausch, C., Richmond, C., Blattner, F. R., & Conway, T. (1999). Functional Genomics: Expression Analysis ofEscherichia coli Growing on Minimal and Rich Media. Journal of Bacteriology, 181(20), 6425–6440. https://doi.org/10.1128/jb.181.20.6425-6440.1999

Taura, T., Baba, T., Akiyama, Y., & Ito, K. (1993). Determinants of the quantity of the stable SecY complex in the Escherichia coli cell. Journal of Bacteriology, 175(24). https://doi.org/10.1128/jb.175.24.7771-7775.1993

Ter Horst, J. P., Turimella, S. L., Metsers, F., & Zwiers, A. (2021). Implementation of Quality by Design (QbD) Principles in Regulatory Dossiers of Medicinal Products in the European Union (EU) Between 2014 and 2019. Therapeutic Innovation and Regulatory Science, 55(3). https://doi.org/10.1007/s43441-020-00254-9

Tian, Y., Chen, J., Yu, H., & Shen, Z. (2015). Overproduction of the escherichia coli chaperones GroEL-GroES in rhodococcus ruber improves the activity and stability of cell catalysts harboring a nitrile hydratase. Journal of Microbiology and Biotechnology, 26(2), 337–346. https://doi.org/10.4014/jmb.1509.09084

Tihanyi, B., & Nyitray, L. (2020). Recent advances in CHO cell line development for recombinant protein production. In Drug Discovery Today: Technologies (Vol. 38, pp. 25–34). Elsevier Ltd. https://doi.org/10.1016/j.ddtec.2021.02.003

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. H. (2008). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface, 6(31), 187–202. https://doi.org/10.1098/rsif.2008.0172

Torres, N. V., & Santos, G. (2015). The (Mathematical) Modeling Process in Biosciences. Frontiers in Genetics, 6. https://doi.org/10.3389/fgene.2015.00354

Tripathi, N. K., & Shrivastava, A. (2019a). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. In Frontiers in Bioengineering and Biotechnology (Vol. 7). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2019.00420

Tripathi, N. K., & Shrivastava, A. (2019b). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. In Frontiers in Bioengineering and Biotechnology (Vol. 7). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2019.00420

Troman, L. (2021). Pushing the Envelope: The Mysterious Journey Through the Bacterial Secretory Machinery, and Beyond. In Frontiers in Microbiology (Vol. 12). https://doi.org/10.3389/fmicb.2021.782900

Tsirigotaki, A., De Geyter, J., Šoštaric´, N., Economou, A., & Karamanou, S. (2016). Protein export through the bacterial Sec pathway. Nature Reviews Microbiology, 15(1), 21–36. https://doi.org/10.1038/nrmicro.2016.161

Tsukazaki, T., Mori, H., Fukai, S., Ishitani, R., Mori, T., Dohmae, N., Perederina, A., Sugita, Y., Vassylyev, D. G., Ito, K., & Nureki, O. (2008). Conformational transition of Sec machinery inferred from bacterial SecYE structures. Nature, 455(7215), 988–991. https://doi.org/10.1038/nature07421

Tullman-Ercek, D., DeLisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., & Georgiou, G. (2007). Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. Journal of Biological Chemistry, 282(11), 8309–8316. https://doi.org/10.1074/jbc.M610507200

Tungekar, A. A., & Ruddock, L. W. (2023). Design of an alternate antibody fragment format that can be produced in the cytoplasm of Escherichia coli. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-41525-3

Ullers, R. S., Genevaux, P., & Luirink, J. (2007). Cotranslational Protein Targeting in Escherichia coli. In Enzymes (Vol. 25, Issue C, p. 3). Academic Press. https://doi.org/10.1016/S1874-6047(07)25001-2

Ullers, R. S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C., & Genevaux, P. (2004). SecB is a bona fide generalized chaperone in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 101(20). https://doi.org/10.1073/pnas.0402398101

Umezawa, K., Yoshida, M., Kamiya, M., Yamasoba, T., & Urano, Y. (2017). Rational design of reversible fluorescent probes for live-cell imaging and quantification of fast glutathione dynamics. Nature Chemistry, 9(3). https://doi.org/10.1038/nchem.2648

Unrean, P., Gätgens, J., Klein, B., Noack, S., & Champreda, V. (2018). Elucidating cellular mechanisms of Saccharomyces cerevisiae tolerant to combined lignocellulosic-derived inhibitors using high-Throughput phenotyping and multiomics analyses. FEMS Yeast Research, 18(8). https://doi.org/10.1093/femsyr/foy106

Van Straaten, M., Missiakas, D., Raina, S., & Darby, N. J. (1998). The functional properties of DsbG, a thiol-disulphide oxidoreductase from the periplasm of Escherichia coli. FEBS Letters, 428(3). https://doi.org/10.1016/S0014-5793(98)00539-0

Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science, 324(5930), 1029–1033. https://doi.org/10.1126/science.1160809

VectorBuilder. (2016). User-inserted region Lentivirus region Bacterial region. www.vectorbuilder.com

Vertommen, D., Depuydt, M., Pan, J., Leverrier, P., Knoops, L., Szikora, J. P., Messens, J., Bardwell, J. C. A., & Collet, J. F. (2008). The disulphide isomerase DsbC cooperates with the oxidase DsbA in a DsbD-independent manner. Molecular Microbiology, 67(2). https://doi.org/10.1111/j.1365-2958.2007.06030.x

Voulhoux, R., Ball, G., Ize, B., Vasil, M. L., Lazdunski, A., Wu, L. F., & Filloux, A. (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. EMBO Journal, 20(23), 6735–6741. https://doi.org/10.1093/emboj/20.23.6735

Wagner, S., Baarst, L., Ytterberg, A. J., Klussmerer, A., Wagner, C. S., Nord, O., Nygren, P. Å., Van Wijks, K. J., & De Gier, J. W. (2007). Consequences of membrane protein overexpression in Escherichia coli. Molecular and Cellular Proteomics, 6(9). https://doi.org/10.1074/mcp.M600431-MCP200

Wang, L., Agyemang, S. A., Amini, H., & Shahbazi, A. (2014). Mathematical modeling of production and biorefinery of energy crops. Renewable and Sustainable Energy Reviews, 43, 530–544. https://doi.org/10.1016/j.rser.2014.11.008

Wang, L., Mahulea, C., Júlvez, J., & Silva, M. (2016). Control of continuous Petri nets using ON/OFF based method.

Wang, X. Y., Du, Q. J., Zhang, W. L., Xu, D. H., Zhang, X., Jia, Y. L., & Wang, T. Y. (2022). Enhanced Transgene Expression by Optimization of Poly A in Transfected CHO Cells. Frontiers in Bioengineering and Biotechnology, 10. https://doi.org/10.3389/fbioe.2022.722722

Wang, Y., & Huang, S. (2017). Golgi structure formation, function, and post-translational modifications in mammalian cells. In F1000Research (Vol. 6). https://doi.org/10.12688/f1000research.11900.1

Wang, Y., & Li, Y. Z. (2014). Cultivation to improve in vivo solubility of overexpressed arginine deiminases in Escherichia coli and the enzyme characteristics. BMC Biotechnology, 14. https://doi.org/10.1186/1472-6750-14-53

Watanabe, M., & Blobel, G. (1989). SecB functions as a cytosolic signal recognition factor for protein export in E. coli. Cell, 58(4). https://doi.org/10.1016/0092-8674(89)90104-9

Weiner, J. H., & Li, L. (2008). Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis. In Biochimica et Biophysica Acta - Biomembranes (Vol. 1778, Issue 9, pp. 1698–1713). https://doi.org/10.1016/j.bbamem.2007.07.020

Wexler, M., Sargent, F., Jack, R. L., Stanley, N. R., Bogsch, E. G., Robinson, C., Berks, B. C., & Palmer, T. (2000). TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in Sec-Independent protein export. Journal of Biological Chemistry, 275(22). https://doi.org/10.1074/jbc.M000800200

Wingender, E. (2010). Petri net Applications in Molecular Biology. In Silico Biology, 10(1,2), 1–4. https://doi.org/10.3233/isb-2010-0420

Winter, J., Neubauer, P., Glockshuber, R., & Rudolph, R. (2000). Increased production of human proinsulin in the periplasmic space of Escherichia coli by fusion to DsbA. Journal of Biotechnology, 84(2). https://doi.org/10.1016/S0168-1656(00)00356-4

Xi, Y., Xu, H., Zhan, T., Qin, Y., Fan, F., & Zhang, X. (2023). Metabolic engineering of the acid-tolerant yeast Pichia kudriavzevii for efficient L-malic acid production at low pH. Metabolic Engineering, 75. https://doi.org/10.1016/j.ymben.2022.12.007

Xu, Z., & Lee, S. Y. (1999). Display of Polyhistidine Peptides on the Escherichia coli Cell Surface by Using Outer Membrane Protein C as an Anchoring Motif. In APPLIED AND ENVIRONMENTAL MICROBIOLOGY (Vol. 65, Issue 11). https://journals.asm.org/journal/aem

Yang, C., Song, C., Freudl, R., Mulchandani, A., & Qiao, C. (2010). Twin-arginine translocation of methyl parathion hydrolase in bacillus subtilis. Environmental Science and Technology, 44(19). https://doi.org/10.1021/es100860k

Yang, J., Guertin, P., Jia, G., Lv, Z., Yang, H., & Ju, D. (2019). Large-scale microcarrier culture of HEK293T cells and Vero cells in single-use bioreactors. AMB Express, 9(1). https://doi.org/10.1186/s13568-019-0794-5

Yazawa, K., & Furusawa, H. (2019). Entropy-Driven Mechanisms between Disulphide-Bond Formation Protein A (DsbA) and B (DsbB) in Escherichia coli. ACS Omega, 4(5), 8341–8349. https://doi.org/10.1021/acsomega.9b00474

Yi, L., Celebi, N., Chen, M., & Dalbey, R. E. (2004). Sec/SRP requirements and energetics of membrane insertion of subunits a, b, and c of the Escherichia coli F1F0 ATP synthase. Journal of Biological Chemistry, 279(38). https://doi.org/10.1074/jbc.M405490200

Young Jae, K., Rajapandi, T., & Oliver, D. (1994). SecA protein is exposed to the periplasmic surface of the E. coli inner membrane in its active state. Cell, 78(5). https://doi.org/10.1016/S0092-8674(94)90602-5

Ytterberg, A. J., Zubarev, R. A., & Baumgarten, T. (2019). Posttranslational targeting of a recombinant protein promotes its efficient secretion into the Escherichia coli periplasm. Applied and Environmental Microbiology, 85(13). https://doi.org/10.1128/AEM.00671-19

Yuan, J., Bennett, B. D., & Rabinowitz, J. D. (2008). Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nature Protocols, 3(8), 1328–1340. https://doi.org/10.1038/nprot.2008.131

Zamani, M., Nezafat, N., & Ghasemi, Y. (2016). Evaluation of recombinant human growth hormone secretion in E. coli using the L-asparaginase II signal peptide. Avicenna Journal of Medical Biotechnology, 8(4).

Zanen, G., Houben, E. N. G., Meima, R., Tjalsma, H., Jongbloed, J. D. H., Westers, H., Oudega, B., Luirink, J., Van Dijl, J. M., & Quax, W. J. (2005). Signal peptide hydrophobicity is critical for early stages in protein export by Bacillus subtilis. FEBS Journal, 272(18), 4617–4630. https://doi.org/10.1111/j.1742-4658.2005.04777.x

Zapun, A., Creighton, T. E., & Bardwell, J. C. A. (1993). The Reactive and Destabilizing Disulphide Bond of DsbA, a Protein Required for Protein Disulphide Bond Formation in Vivo. Biochemistry, 32(19). https://doi.org/10.1021/bi00070a016

Zarschler, K., Witecy, S., Kapplusch, F., Foerster, C., & Stephan, H. (2013). High-yield production of functional soluble single-domain antibodies in the cytoplasm of Escherichia coli. Microbial Cell Factories, 12(1). https://doi.org/10.1186/1475-2859-12-97

Zhang, H., Yang, J., Wu, S., Gong, W., Chen, C., & Perrett, S. (2016). Glutathionylation of the bacterial Hsp70 chaperone dnak provides a link between oxidative stress and the heat shock response. Journal of Biological Chemistry, 291(13). https://doi.org/10.1074/jbc.M115.673608

Zhang, Y., Werling, U., & Edelmann, W. (2012). SLiCE: A novel bacterial cell extract-based DNA cloning method. Nucleic Acids Research, 40(8). https://doi.org/10.1093/nar/gkr1288

Zhang, Y., Zhao, J., Jiang, Y., Zhao, D., Wang, C., & Li, M. (2022). Research progress in yeast metabolic engineering to produce natural product flavors. In China Surfactant Detergent and Cosmetics (Vol. 52, Issue 6). https://doi.org/10.3969/j.issn.1001-1803.2022.06.011

Zhang, X., & Shan, S. (2014). Fidelity of Cotranslational Protein Targeting by the Signal Recognition Particle. Annual Review of Biophysics, 43(1), 381–408. https://doi.org/10.1146/annurev-biophys-051013-022653

Zhou, A., Kang, T. M., Yuan, J., Beppler, C., Nguyen, C., Mao, Z., Nguyen, M. Q., Yeh, P., & Miller, J. H. (2015). Synergistic interactions of vancomycin with different antibiotics against escherichia coli: Trimethoprim and nitrofurantoin display strong synergies with vancomycin against wild-type E. coli. Antimicrobial Agents and Chemotherapy, 59(1). https://doi.org/10.1128/AAC.03502-14

Zhou, H., Zhang, Y., Long, C. P., Xia, X., Xue, Y., Ma, Y., Antoniewicz, M. R., Tao, Y., & Lin, B. (2024). A citric acid cycle-deficient Escherichia coli as an efficient chassis for aerobic fermentations. Nature Communications, 15(1). https://doi.org/10.1038/s41467-024-46655-4

Zhou, Y., Ueda, T., & Müller, M. (2014). Signal recognition particle and SecA cooperate during export of secretory proteins with highly hydrophobic signal sequences. PLoS ONE, 9(4). https://doi.org/10.1371/journal.pone.0092994

Zinecker, S., Jakob, M., & Klösgen, R. B. (2020). Functional reconstitution of TatB into the thylakoidal Tat translocase. Biochimica et Biophysica Acta - Molecular Cell Research, 1867(2). https://doi.org/10.1016/j.bbamcr.2019.118606

ZUBEREK W. M. (1991). TIMED PETRI NETS DEFINITIONS, PROPERTIES, AND APPLICATIONS. Memorial University of Newfoundland, St. John's, Canada A]C 5S7 .

Zuily, L., Lahrach, N., Fassler, R., Genest, O., Faller, P., Sénèque, O., Denis, Y., Castanié-Cornet, M. P., Genevaux, P., Jakob, U., Reichmann, D., Giudici-Orticoni, M. T., & Ilbert, M. (2022). Copper Induces Protein Aggregation, a Toxic Process Compensated by Molecular Chaperones. MBio, 13(2). https://doi.org/10.1128/mbio.03251-21

Appendix A

Table A. 1 Petri net models published in BioModel

Name of model	**BioModel reference number**
Petri net model with human growth hormone as model protein using Snoopy	MODEL2407150001
Python models of protein synthesis, Sec, Tat and disulphide bond formation pathways in *E. coli*	MODEL2407190001
Petri net model of DnaK and DnaJ chaperones using Snoopy	MODEL2407190003.
Petri net model of ClpB chaperone using Snoopy	MODEL2407190002
Petri net model of GroEL and GroES chaperones using Snoopy	MODEL2407190004.
Petri net model of Sec translocation using Snoopy	MODEL2407190005.
Petri net model of Tat translocation using Snoopy	MODEL2407190006.
Petri net model of disulphide bond formation using Snoopy	MODEL2407190007.
Petri net model of TCA cycle and oxidative phosphorylation using Snoopy	MODEL2407190008.

Appendix B

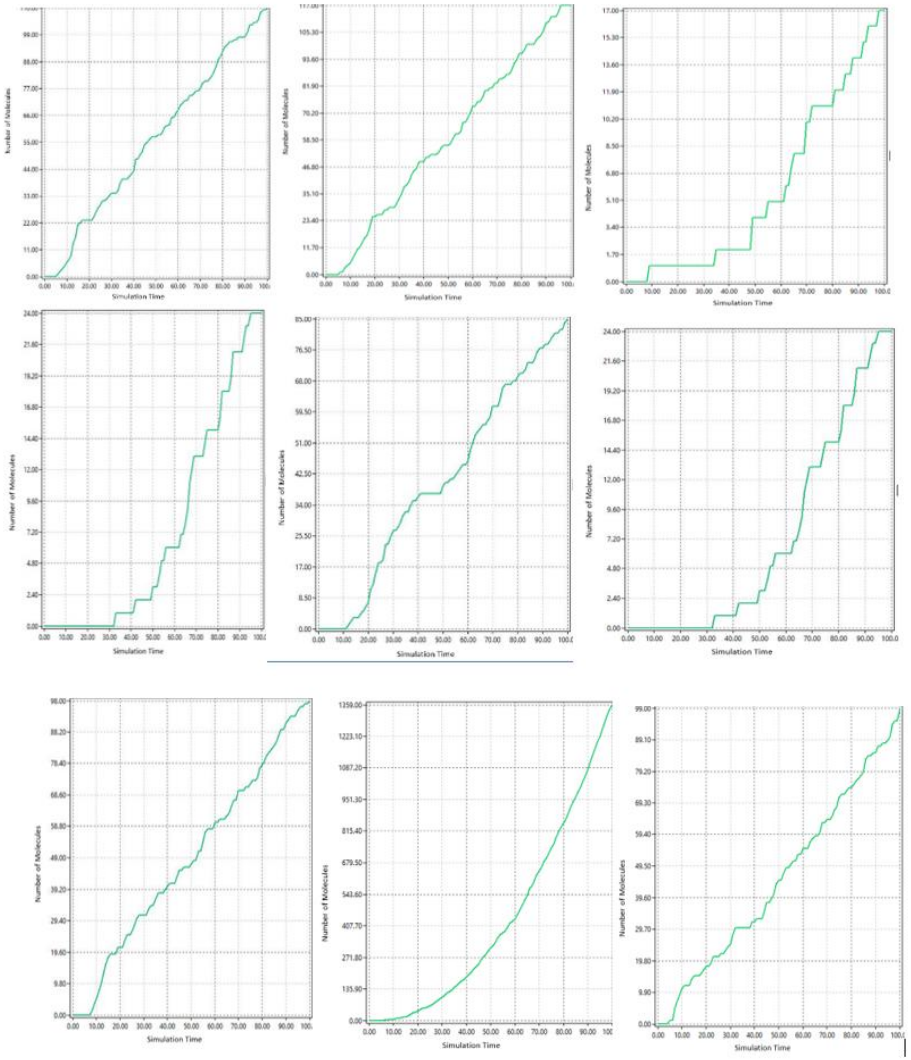


Figure B. 1 Summary of plots for section 4.5.2 to demonstrate the curves achieved at different rates of translocation using the Tat pathways. The scenarios where no protein is generated is not included in this Figure more information can be found on tables 4.1, 4.2 4.3 and 4.4.

Appendix C

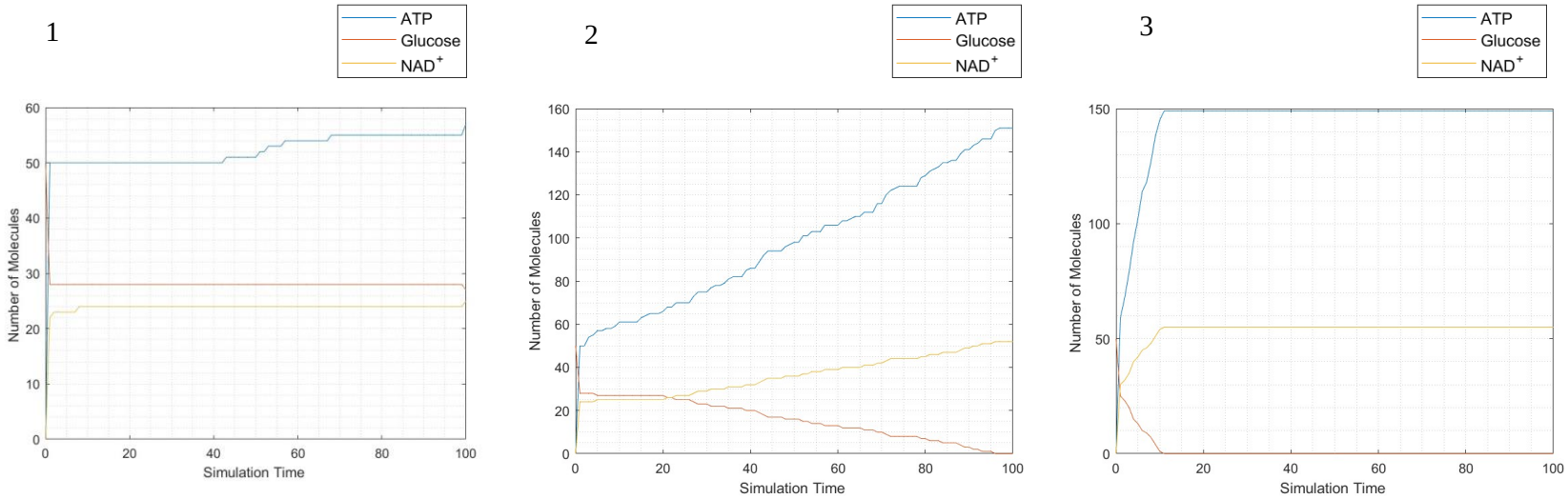


Figure C 1 TCA Cycle Simulation for Low, Middle, and High Reaction Rates. Graph 1 illustrates the low-rate scenario, where glucose is not completely depleted, and ATP production is limited. Graph 2 represents the middle-rate scenario, showing gradual glucose depletion and the production of approximately 150 molecules of ATP. Graph 3 depicts the high-rate scenario, characterised by rapid glucose depletion and the production of 150 molecules of ATP. These results demonstrate that increased TCA cycle rates impact the amount of ATP and reducing agents generated within the simulation, highlighting the dynamic relationship between metabolic activity and energy availability.

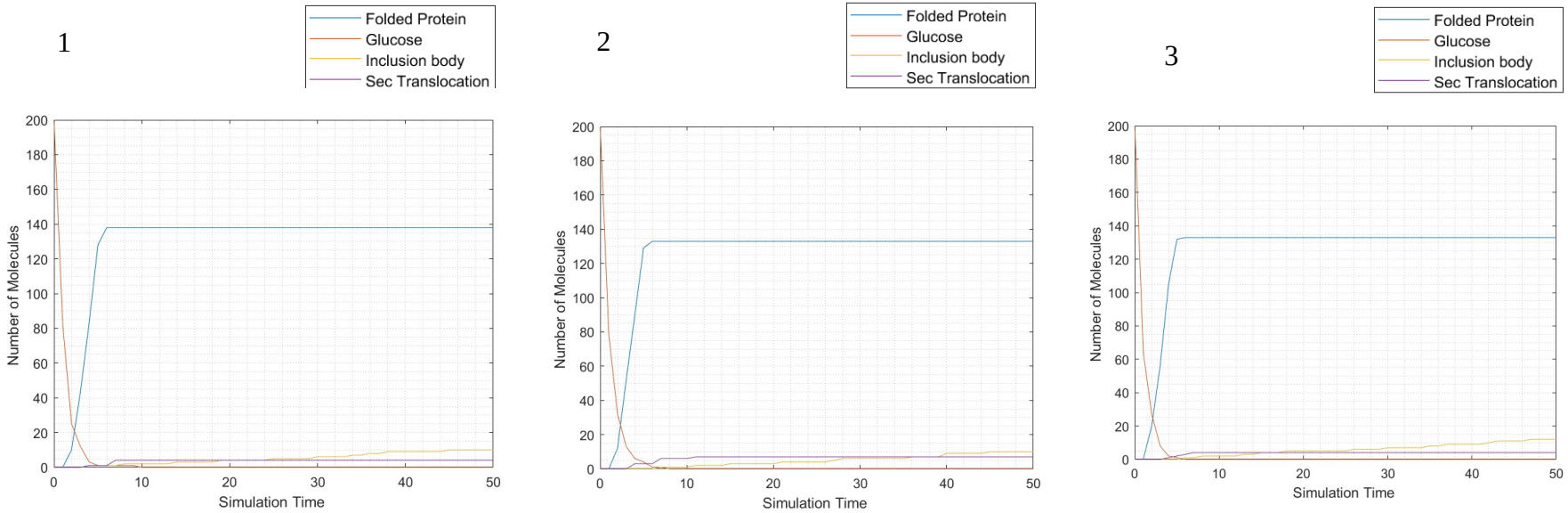


Figure C 2 Simulation Results for Low, Middle, and High TCA Cycle Rates. Graph 1 shows the low-rate scenario, with 139 molecules of folded protein, 10 molecules of inclusion bodies, and no accumulation in the Sec translocation pathway. Graph 2 represents the middle-rate scenario, with 138 molecules of folded protein, 10 molecules of inclusion bodies, and 7 molecules accumulating in the Sec translocation pathway. Graph 3 depicts the high-rate scenario, with 138 molecules of folded protein, 10 molecules of inclusion bodies, and no accumulation in the Sec translocation pathway. These results indicate that varying TCA cycle rates have minimal impact on protein production, suggesting that other factors, such as chaperone availability or translocation efficiency, may be limiting the system.