

Review

Advances in *Escherichia coli*-Based Therapeutic Protein Expression: Mammalian Conversion, Continuous Manufacturing, and Cell-Free Production

Sarfaraz K. Niazi ^{1,*}  and Matthias Magoola ²¹ College of Pharmacy, University of Illinois, Chicago, IL 60012, USA² DEI Biopharma, Kampala P.O. Box 35854, Uganda; dei@deigroupinternational.com

* Correspondence: niazi@niazi.com; Tel.: +1-312-297-0000

Abstract: Therapeutic proteins treat many acute and chronic diseases that were until recently considered untreatable. However, their high development cost keeps them out of reach of most patients around the world. One plausible solution to lower-cost manufacturing is to adopt newer technologies like using *Escherichia coli* to express larger molecules, including full-length antibodies, generally relegated to Chinese Hamster Ovary (CHO) cells, adopt continuous manufacturing, and convert the manufacturing to cell-free synthesis. The advantages of using *E. coli* include a shorter production cycle, little risk of viral contamination, cell host stability, and a highly reproducible post-translational modification.

Keywords: biosimilars; *Escherichia coli*; recombinant proteins; monoclonal antibody; bispecific antibody; inclusion body; continuous manufacturing (CM); cell-free protein synthesis (CFPS)



Citation: Niazi, S.K.; Magoola, M. Advances in *Escherichia coli*-Based Therapeutic Protein Expression: Mammalian Conversion, Continuous Manufacturing, and Cell-Free Production. *Biologics* **2023**, *3*, 380–401. <https://doi.org/10.3390/biologics3040021>

Academic Editor: Alex Loukas

Received: 29 September 2023

Revised: 26 October 2023

Accepted: 20 November 2023

Published: 29 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Therapeutic proteins represent a diverse class of drugs first made accessible as a recombinant DNA (rDNA) insulin in 1982 [1]. There are now 266 such approved proteins [2], comprising a wide range of products with unique mechanisms of action and size, ranging from peptides to monoclonal antibodies (Table 1).

It is noteworthy that the European Medicines Agency (EMA) lists peptides as proteins, unlike the US Food and Drug Administration (FDA) [3]. The *E. coli*-expressed products, including hormones, cytokines, enzymes, antibody fragments, and shorter-than-full-length antibodies, came into clinical use long before the use of Chinese Hamster Ovary (CHO) cells (Table 1). Some proteins still extracted from tissues can also be manufactured using *E. coli*. Examples include Alpha-1 antitrypsin, Antithrombin III, Botulinum toxin, C1 inhibitor, Fibrinogen, Heparin, Hirudin, Snake venom proteins, Streptokinase, Thrombin, and Urokinase. Figure 1 shows that most proteins expressed in *E. coli* are of lower molecular weight since most of the more complex and higher molecular weight monoclonal antibodies (mAbs) are expressed in CHO cells.

The FDA-approved therapeutic proteins expressed in *E. coli* are primarily of molecular weight of less than 32 kDa; it is difficult to produce proteins higher than 100 kDa in *E. coli* as it places excessive cell host load that prevents correct protein folding while maintaining adequate expression levels [4].

The value of *E. coli* will become more apparent when we start using it to express larger molecules such as the mAbs. The first study on the expression of full-length (FL) immunoglobulins (FL-IgGs) in *E. coli* was published in 2002. Later, in 2020, a modular system-based synthetic biology technique was applied to knock down gene expression using short regulatory ribonucleic acids (RNAs) with cetuximab as a target FL-IgG for enhancing expression, reaching up to 200 mg/L [5].

Table 1. FDA-approved therapeutic proteins as of July 2023. (<https://drugs.ncats.io/>, accessed on 10 July 2023).

Listed Protein Class	Number
Monoclonal Antibody	94
Hormone	10
Enzyme	8
Monoclonal Antibody Conjugate	8
Cytokine	4
Bispecific Antibody	3
Coagulation Factor	3
Growth Factor	3
Peptide	3
Carrier Protein	1
Enzyme Inhibitor	1
Fab	1
Fusion Proteins	1
Single-Domain Antibody	1
Toxin	1

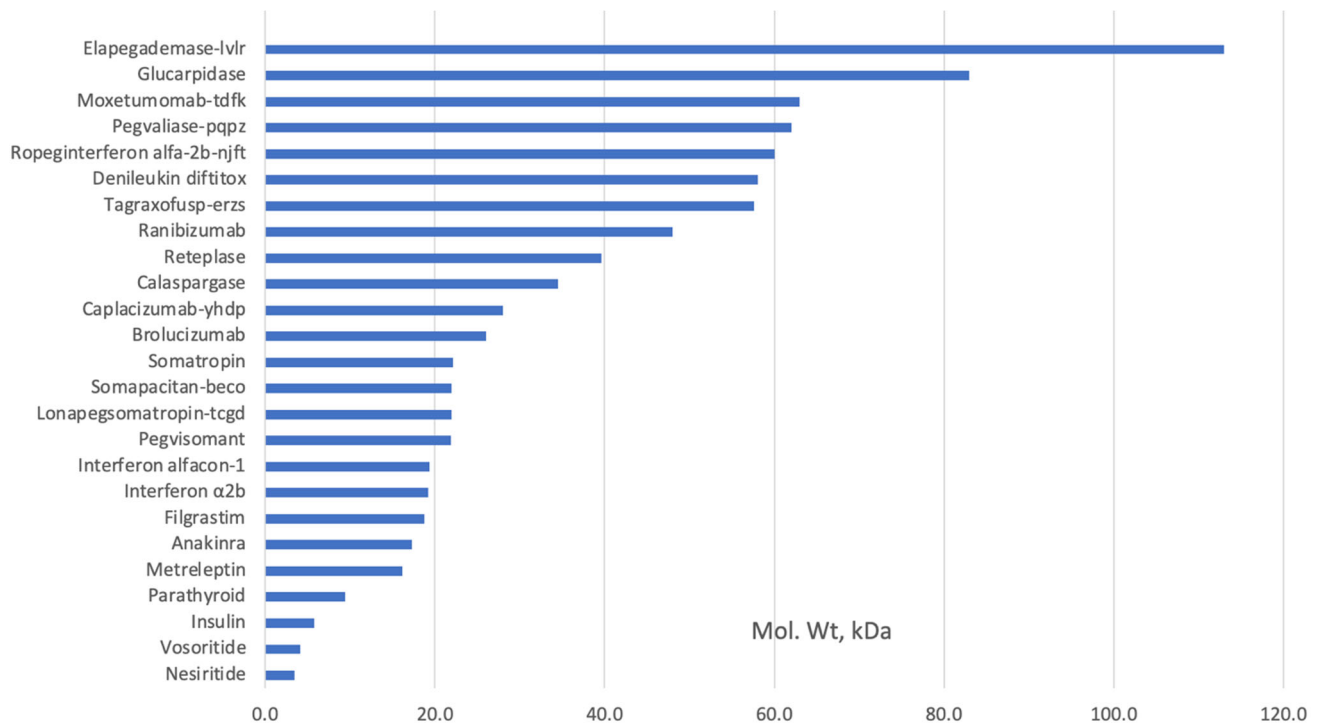


Figure 1. Examples of FDA-approved recombinant proteins produced in *E. coli* (www.fda.hhs.gov, accessed on 10 July 2023).

Combining *E. coli* with emerging technologies like bioinformatics, novel methods for genetic manipulation to force *E. coli* to secrete heterologous proteins, and managing post-translational modification offer many new opportunities, including *E. coli*-based continuous manufacturing and cell-free synthesis, that can significantly reduce the cost of development and manufacturing, as well enhance product safety.

2. Background

Recombinant DNA products involve the fusion of DNA from distinct species, followed by introducing the resulting hybrid DNA into a host cell, typically a bacterium or mammalian cell, to express the desired protein. The pioneering development of this molecular chimera dates back to 1972 [6,7] when researchers affiliated with the University of California, San Francisco, and Stanford University accomplished this technique. The United States patent for the invention was granted to Stanley Cohen from Stanford University and Herbert Boyer from the University of California, San Francisco (UCSF) in 1980. In 1976, Boyer played a crucial role in establishing Genentech, Inc. These patents have been licensed to more than 500 licensees and yielded royalties exceeding USD 250 million for Stanford and UCSF [8].

Since the FDA approved the first recombinant protein for therapeutic purposes in 1982, *E. coli* has remained a prominent organism for producing recombinant proteins despite the availability of many newer expression systems. The utilization of microbial expression systems, particularly *E. coli*, continues to offer a more straightforward and cost-effective approach for producing even heterologous recombinant proteins compared to mammalian cell culture and other systems. *E. coli* presents several notable benefits in genetic manipulation, growth conditions, high product yields, product purity, absence of viral contamination, and many more.

3. Advantages

Using *E. coli* as an expression system offers several advantages.

- It is the most well-understood expression system. The genome of *Escherichia coli* strain K-12 MG1655, which is the most studied and best-characterized strain, has been fully sequenced and annotated. It was first completely sequenced in 1997, and the annotation and analysis have been continually updated since then as our understanding of genomics and the biology of *E. coli* has advanced [9]. This knowledge base is critical in its utilization as a robust expression system.
- Numerous prokaryotic genes [10] are expressed in operons [11], where a solitary promoter leads to the synthesis of multiple proteins from a single mRNA molecule, which has a ribosome binding site (RBS) preceding the beginning AUG codon of each protein. This enables the simultaneous production [12] of subunits that assemble into complexes or the simultaneous expression of auxiliary components that may be necessary for the protein to attain its native shape.
- Straightforward genetic manipulation, tolerant host, T7 RNA polymerase (RNAP) fused with deaminases, and compartmentalization by creating artificial organelles and modifications to bring in precise PTMs and other protein characters [13–15].
- Simpler scale-up compared to eukaryotic systems, including mechanical cell disruption, which is less variable than eukaryotic cells, requires gentler lysis to preserve more fragile organelles and structures [16].
- Avoiding virus contamination risk. Proteins synthesized in mammalian cell lines, the host cells possess multiple copies of endogenous retrovirus-like sequences, which subsequently generate retrovirus-like particles (RVLPs) together with the target protein. While RVLPs are commonly regarded as dysfunctional, certain instances have demonstrated their ability to infect cell lines that are not of rodent origin. Exogenous viral contamination resulting from raw materials or persons is also possible; however, such concerns are not relevant in the context of *E. coli*-based expression systems [17].
- Low-cost growth medium, fast cellular proliferation, uncomplicated fermentation procedures, no viral contaminants in the final product, and high product yields [18].

4. Challenges

Choosing between these systems requires careful consideration of the specific protein and its intended application.

- Proteins overexpressed in *E. coli* may form insoluble aggregates known as inclusion bodies, requiring specific solubilization and refolding steps, adding complexity to the purification process compared to eukaryotic cells [19]; additional purification steps if inclusion bodies are formed. A frequently encountered challenge is the formation of inclusion bodies—insoluble aggregates of misfolded proteins. Several tactics have been developed to address this. Incorporating solubility-enhancing fusion tags, such as SUMO or maltose-binding protein (MBP), has proven to enhance the solubility of certain target proteins [20]. Additionally, co-expressing the protein of interest with molecular chaperones can help in its proper folding, making inclusion body formation less likely [21]. Like adjusting the temperature or the IPTG concentration, fine-tuning expression conditions can also modulate protein synthesis rates and improve solubility [22]. Even if inclusion bodies form, there is a workaround: the proteins can be solubilized with denaturants and then refolded, salvaging the protein for further use [23]. These adaptive strategies emphasize the versatility and adaptability of the *E. coli* expression system, showcasing the myriad tools researchers have at their disposal to optimize protein production.
- *E. coli* lacks the machinery for many eukaryotic PTMs, such as glycosylation, which may affect protein stability, folding, and activity [24].
- Unlike eukaryotic systems, *E. coli* produces endotoxin contamination from its lipopolysaccharide, which must be removed during purification [25].
- The toxicity of overexpressed proteins to *E. coli* often forces the expression of toxic protein fragments or domains retaining essential functions [26]. One strategy involves using signal sequences attached to the protein's N-terminus, directing the protein's export to the periplasm, and decreasing cytoplasmic accumulation, thereby reducing potential toxicity [27].
- Regulating the expression through weak promoters or controlled induction can temper any adverse impacts on the host cells. This requires codon optimization to enhance translation efficiency [28].
- Expressing monoclonal antibodies (mAbs) in *Escherichia coli* (*E. coli*) presents multiple challenges, stemming primarily from the intricacy of these proteins. One of the main hurdles is ensuring the proper folding of mAbs, especially since they possess multiple domains. *E. coli* often struggles to correctly fold such large eukaryotic proteins, especially when they have multiple disulfide bonds. Furthermore, bacteria lack the machinery for certain post-translational modifications like glycosylation, which are vital for the function of mAbs. This absence can compromise the mAb's efficacy [29]. The reducing environment of the *E. coli* cytoplasm also makes disulfide bond formation problematic, while protein degradation can occur if the expressed proteins are unstable or perceived as foreign. Several strategies can be employed to counter these challenges. One approach is the expression of single-chain variable fragments (scFvs), which comprise the variable regions of the mAb's heavy and light chains connected by a short peptide linker. Researchers can also leverage specialized *E. coli* strains designed for disulfide bond formation in the cytoplasm, such as SHuffle strains [30]. Directing mAbs or scFv expression to the periplasmic space of *E. coli*, which is more oxidizing than the cytoplasm, can also encourage proper disulfide bond formation. Adjustments in expression conditions, co-expression with molecular chaperones, and codon optimization for *E. coli* are additional strategies to improve yields [31]. The ability of bispecific antibodies (BsAbs) [32] to effectively target two entities concurrently enhances the practicality of antibody-based treatments. Genentech has successfully devised a periplasmic expression system in *Escherichia coli*, known as the BsAb expression system. This system utilizes either the Knobs-into-Holes (KiH) [33] technology or Fc domain HC heterodimerization [34]. Genentech has made significant advancements in the production process of bispecific antibodies (BsAbs), including two distinct heavy chains (HCs) and two distinct light chains (LCs). These improvements have been

achieved by utilizing either a two-culture or a coculture strategy in *Escherichia coli* (*E. coli*) systems [35].

5. Bioinformatics Applications

Choosing whether *E. coli* is a suitable system and whether its adoption will fulfill the goal of reduced-cost manufacturing requires a systematic process (Figure 2) that commences by considering several factors, such as potential splice variants, signal sequences, transmembrane helices, and post-translational modifications observed in the native protein. Protein databases, such as UniProt [36], are valuable initial bioinformatics resources, albeit they remain suggestive, not definitive. For example, knowing what can be produced in *E. coli* and what cannot be is critical knowledge; cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) is an obligate dimer and requires N-glycosylation of Asn78 and Asn110 for dimerization [37], and this PTM cannot be made in *E. coli*. This will need a synthetic biology method since the only eukaryotic-like PTMs that *E. coli* can produce is disulfide bond formation in the periplasm [38].

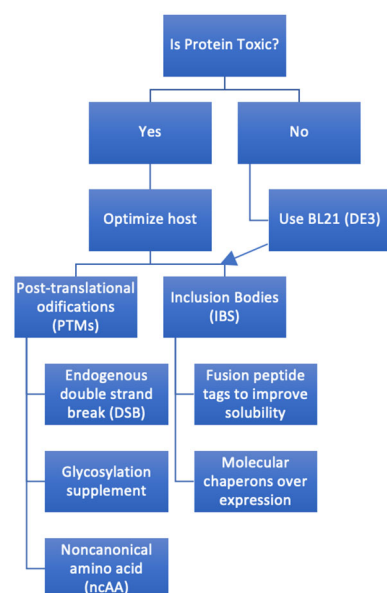


Figure 2. A systematic process of creating a plan to express a full-length antibody in *E. coli*. Arrow indicates optimized host follow on.

Bioinformatics methodologies, such as the software JPRED 4 [39], facilitate the investigation of domain boundaries and the prediction of regions of intrinsically disordered proteins (IDPs). Failure to express a construct that is insufficient in length and lacks a crucial component within a certain domain, such as a β -strand, should be anticipated. Conversely, attempting to express an excessively lengthy construct encompassing flexible regions susceptible to proteolysis will likely lead to either heterogeneity or the removal of a purification tag. Proteins characterized by a significant proportion of intrinsically disordered regions (IDRs) often provide challenges during production due to their inherent susceptibility to degradation. Proteins characterized by a significant proportion of intrinsically disordered regions (IDRs) often provide challenges in their production due to their inherent susceptibility to degradation. However, these may become structured upon interaction with other molecules, forming complexes such as acetyltransferase (ACTR) and nuclear coactivator binding domain (NCBD) [40] that can help as partners, making a stable and soluble protein.

Expanding the scope of the use of *E. coli* systems is now based on several well-defined prospects:

- Exploiting the use of bioinformatics tools to determine the biophysical characteristics of the protein [41]. It is a complex process that involves various computational

methods. These methods utilize algorithms and statistical models to analyze the protein's primary sequence, infer its three-dimensional structure, and predict its interactions and functions.

- Sequence analysis involves comparing the amino acid sequence of a protein with known sequences in databases to identify conserved domains, motifs, or families [42];
- Structure prediction includes methods like homology modeling, ab initio modeling, and threading to predict a protein's three-dimensional (3D) structure based on its sequence [43];
- Functional prediction identifies the biological role of a protein by assessing its structural and sequential features, often in conjunction with known protein–protein interactions and pathway analyses [44];
- Molecular dynamics simulations and related techniques are used to study the movement and interactions of proteins, providing insight into their behavior in the cellular environment [45];
- Specific bioinformatics tools are designed to predict sites in proteins likely to undergo post-translational modifications (PTMs) such as phosphorylation or glycosylation [46];
- Predicting how proteins interact with other proteins or ligands can be achieved through docking simulations and other modeling techniques [47].
- Accurate delineation [48]:
 - Identifying the boundaries of protein domains is essential for understanding the function and evolution of proteins [49];
 - Signal sequences are crucial for the targeting of proteins to specific cellular locations. Identifying these sequences helps in understanding the transportation and localization of proteins [50];
 - Transmembrane regions anchor proteins in membranes, playing essential roles in cellular communication, signaling, and transport. Accurate prediction of these regions aids in understanding membrane protein structure and function [51];
 - Identifying obligate oligomeric complexes is essential for understanding protein–protein interactions and the assembly of multi-protein complexes [52];
 - Identification of PTMs is vital for understanding protein regulation and signaling [53].

Optimization of genetic and translation variables encompasses various elements, including codon use, the characteristics and placement of the ribosome binding site, and disparities in translation rates between prokaryotes and eukaryotes [54].

6. Gene Cloning and Design

Gene cloning typically involves selecting a purification method, such as affinity chromatography, which utilizes the inherent characteristics of the protein. This can be achieved through immobilized ligand or substrate mimic chromatography, using compounds like Cibacron Blue F3GA [55] or cyclic peptide-based ligands [56]. Alternatively, a purification tag, such as a maltose-binding protein (MBP)-tag, glutathione-S-transferase (GST)-tag, or commonly a hexahistidine tag (his-tag), can be added to facilitate purification. Immobilized metal affinity chromatography (IMAC) [57] is commonly employed. If a protein with similar characteristics is present, its attributes can be utilized to assess the feasibility of adding a tag to the N- and C-terminus. Alternatively, one might utilize structure prediction software such as Phyre 2 [58]. Although N-terminal histidine tags are highly valuable and extensively employed, they can introduce heterogeneity in the final product due to varying (phospho)glucosylation occurring at the N-terminus [59].

After the design of the protein construct, gene design starts to yield maximum expression that depends much on cellular homeostasis or keeping a delicate balance within the

cell. When a high-copy number plasmid is employed with a robust promoter, it consistently leads to a reduced protein yield [60]. This is attributed to the excessive allocation of cellular resources towards synthesizing plasmid DNA and mRNA. Consequently, the abundance of mRNA exceeds the capacity of the translation machinery, resulting in suboptimal protein production. Toxic effects of overexpressed recombinant proteins on *E. coli* cells can be anticipated to avoid these processes [61].

Transcriptome analysis can identify and remove the genes in charge of the cellular stress response. The number of growth-essential genes' down-regulated expression is reduced when cell surface receptor (CSR) is blocked [62].

The prevailing strategy involves the integration of genes into the bacterial chromosome to circumvent the issue of plasmid loss during extensive fermentation processes. However, despite their drawbacks, plasmids can be employed in their original form because they are more expeditious and cost-effective. The selection of plasmids for protein production is determined by their copy quantity, which is contingent upon the plasmid's origin of replication, promoter, and selection marker. The optimization of cellular resources allocated to protein production is contingent upon achieving an appropriate equilibrium between plasmid copy number and promoter strength, with consideration for the specific media conditions.

The field of synthetic biology has witnessed notable progress in developing growth-decoupled recombinant protein production. This has been achieved using the co-expression of Gp2, a peptide generated from a bacteriophage, which acts as an inhibitor of RNA polymerase in *Escherichia coli*. This methodology facilitated the regulation of metabolic resources, ensuring their exclusive allocation towards synthesizing the intended protein.

In addition to the plasmid, the origin of the gene is a crucial factor. Traditionally, the gene has been obtained directly from the original organism, typically by using a cDNA library acquired via reverse transcription polymerase chain reaction (RT-PCR) from a pool of messenger RNA (mRNA) to circumvent the inclusion of introns. Although the process can exhibit rapidity, cost-effectiveness, and efficiency, it can also lead to challenges associated with disparities in translation initiation and codon utilization between prokaryotic and eukaryotic organisms.

Due to a significant decrease in pricing, the cost of synthesizing a gene artificially has become lower than the combined expenses of labor and materials involved in cloning a gene from a complementary DNA (cDNA) library. Synthetic genes can also alleviate the potentially harmful consequences of another dissimilarity in protein translation rates between eukaryotes and prokaryotes [63]. In prokaryotic organisms like *Escherichia coli* (*E. coli*), a coupling exists between the transcription and translation rates [64]. Specifically, transcription occurs at a rate of 50 nucleotides, whereas translation occurs at 16 amino acids.

6.1. Ribosomes

In 1987 [65], a modified ribosome system was developed to facilitate the production of the proteins in *E. coli* through modifications made to the Shine–Dalgarno (SD) sequence of the mRNA and the corresponding anti-SD sequence of the 16S ribosomal RNA (rRNA). Other alternative ribosome systems can be utilized, including the orthogonal riboswitch system [66], the RiboTite system, and the Ribo-T system [67,68]. The riboswitch system facilitates the adjustable co-expression of several genes in a dose-dependent manner in response to tiny synthetic chemicals. On the other hand, the RiboTite system, an extension of the riboswitch technology, has demonstrated the ability to synchronize protein translation rates with protein release. The Ribo-T system utilizes a modified hybrid rRNA that combines small and large subunit rRNA sequences. This modified rRNA is connected into a single translating unit using short RNA linkers that form covalent bonds between the subunits. The functionality of the orthogonal ribosome-mRNA system has been demonstrated in sustaining bacterial growth in the absence of wild-type ribosomes. Furthermore, a recent study has documented the development of an enhanced tethered version of this system [69].

- The characteristics and location of the ribosome binding site (RBS) and the disparities in translation rates observed in prokaryotic and eukaryotic organisms [70]. The ribosome binding site (RBS) plays a crucial role in the translation initiation. The sequence and position of a gene relative to the initiation codon can influence the translation efficiency. Customizing the RBS to the host organism might enhance the efficiency of translating the desired protein [71];
- Correct use of the strain and media to optimize production, though with many limitations [72]. The optimization of production in *E. coli* strains through proper selection of the strain and media is a common strategy in biotechnology but comes with certain limitations;
- Optimization in *E. coli* can vary widely depending on the protein or other manufactured product. Selecting the right strain of *E. coli*, determining the optimal temperature, and choosing the appropriate culture media are crucial considerations for recombinant protein expression.

The presence of secondary structural components in mRNA might obstruct ribosome binding, resulting in hindered translation and various limits in the translational process [73]. Eukaryotic ribosomes exhibit a binding affinity towards the cap located at the 5' terminus of the mRNA molecule. Subsequently, they traverse along the mRNA until they commence translation at the initial AUG codon, preceded by a Kozak sequence. In contrast, prokaryotic ribosomes engage with a specific region on the mRNA called the Shine–Dalgarno sequence or ribosome binding site. The ribosome binding sites (RBS) typically consist of 5–13 base pairs [74] upstream of the beginning AUG codon, with an ideal spacing of 5–6 base pairs [75]. These RBS sequences complement the 3' end of the 16S ribosomal RNA. The nucleotide sequence AGGAGGU [76] is seen in *Escherichia coli*. The presence of a separate ribosome binding site (RBS) in eukaryotic protein production in *Escherichia coli* (*E. coli*) leads to two unique outcomes. Before beginning the AUG codon, a ribosome binding site (RBS) must be present. This phenomenon may be observed within the plasmid region external to the multi-cloning site. However, it is imperative to exercise caution to ensure that the distance is appropriate and that the translation process does not inadvertently introduce more AUG trinucleotides.

Furthermore, it is essential that this specific nucleotide sequence does not occur inside the gene of interest. An internal ribosome binding site (RBS) can have two outcomes: it can lead to the production of a second protein if there is an AUG codon at the appropriate distance from it, or it can cause translation stalling as a ribosome binds to this site and obstructs translation. Therefore, special consideration is provided to the choice of codons for Gly-Gly pairs (excluding GGA-GGU), Arg-Arg pairs (excluding AGG-AGG), and sequences surrounding Glu (GAG), including Glu-Glu pairs (GAG-GAG). *Escherichia coli* (*E. coli*) exhibits infrequent utilization of AGG and GGA codons. Therefore, it is crucial to exercise caution while optimizing codons to prevent the occurrence of internal ribosome binding sites (RBS) associated with sequences around glutamic acid (Q/K/E-E or E-V).

6.2. Promoter

The significant functional sections close to PT7 include the −35/−10 region, translation initiation region (TIR), operator sequence, and replicon of the TpET plasmid. Numerous functional areas close to the PT7, the pET plasmid's core region, control the basal expression level before induction and the proper transcription rate following induction.

By maximizing transcription or translation levels, the T7 RNAP objective is attained. The lacUV5 promoter (PlacUV5), a strongly inducible promoter that is activated by the amino acid isopropyl-beta-d-thiogalactopyranoside (IPTG), controls this process [77], and the P_{lacUV5} is independent of recombinant product, which makes it leakier than P_{lac} [78]. Three inducible promoters—ParaBAD [79], PrhaBAD, and Ptet—are appropriate for toxin-protein fermentation that lasts a long time. PrhaBAD and Ptet, however, more strictly control T7 RNAP transcription, giving additional expression possibilities for various recom-

binant products—especially dangerous proteins [80]. When the lac repressor gene (*lacI*) is altered, leaky expression is decreased by improving the ability to inhibit proteins [81].

- To create the promoter variation *lac1G*, the promoter *lacUV5* and *lac* were joined again. (G was substituted for A at position +1) [82];
- The expression of T7 RNA polymerase (RNAP) is effectively regulated to prevent leakage by the presence of a mutant form of the Lac repressor protein (*LacI*), specifically the V192F variant. This mutant variant cannot bind to isopropyl β -D-1-thiogalactopyranoside (IPTG), hence preventing its activation. Consequently, the mutant *LacI* dynamically governs the levels of transcripts produced by T7 RNAP [83];
- Building a T7 RNAP RBS library quickly involves using the base editor and CRISPR/Cas9 to screen potential expression hosts [84];
- The ability of T7 RNA polymerase to bind to the PT7 promoter was impaired due to a specific amino acid substitution (A102D), resulting in an alteration in the rate of RNA production. The T7 RNA polymerase (T7 RNAP) was fragmented into two segments and co-expressed with a light-responsive dimerization domain, exhibiting functional behavior upon exposure to blue light [85].

6.3. Codons

The expression level of the *ColE1* plasmid replication-associated gene can be regulated by utilizing CRISPRi and the inducible promoter *Ptet* [86].

The distribution of codon usage is not uniform throughout the available codons, and there is significant variance in the degree of codon usage bias observed among different organisms. Using codons exhibits substantial variation across other microorganisms and is associated with corresponding transfer RNA (tRNA) quantities [87].

mRNA, which contains multiple rare codons, can exhibit translation stalling and degradation [88]. Bioinformatic approaches can examine codon usage issues, e.g., Graphical Codon Usage Analyzer [89]. One method to prevent this problem is to overexpress the rare tRNAs [90], such as from *pLysSRARE* [91,92]. The usual approach is using synthetic genes that can be codon optimized for the expression host while avoiding internal RBS, internal restriction sites, and factors that influence mRNA structure and stability [93,94].

6.4. Protein Folding

Translation rates in eukaryotes are comparatively slower, typically occurring at approximately three amino acids per second. The process of protein folding has co-evolved with translation rates, resulting in a situation where the translation rate [95] of a eukaryotic protein expressed in *E. coli* may exceed the folding rate. This poses a challenge, particularly for multi-domain proteins. However, this challenge can be addressed through various strategies, such as adjusting the translation rate, harmonizing codon usage [96], or intentionally inducing ribosome stalling by incorporating rarer codons at domain boundaries.

When the host cell cannot handle the rate or volume of recombinant products being expressed, many proteins will misfold and cluster, eventually creating IBs and obstructing the expression. The primary reasons for the synthesis of IBs are limited post-translational modifications (PTMs) capacity and folding efficiency, which are of the utmost importance for increasing the functional activity of recombinant products [97].

To ensure the proper folding and functionality of antibodies with disulfide linkages, it is necessary to expose the individual antibody chains to the oxidizing conditions present in the bacterial periplasm. In addition, it should be noted that the periplasmic space serves as a habitat for specific proteins known as chaperonins and disulfide isomerases, which play a crucial role in correctly folding newly synthesized proteins [98]. A leader sequence (*PelB*, *OmpA*, *PhoA*) drives the antibody to the oxidizing periplasm for periplasmic expression [99]. After being expressed, the antibody is extracted from the periplasmic region by osmotic shock. Yields obtained from shaking flask cultures have been documented to range from 0.1 mg/L to 100 mg/L, while using fermenters has demonstrated the potential to achieve yields as high as 2 g/L [100]. Utilizing specific *E. coli* strains that offer an oxidizing

environment in the cytoplasm is an additional choice; typically, it comprises mutations of the enzymes, glutathione oxidoreductases, and thioredoxin reductases [101].

Choosing the right molecular chaperones, such as GroES/GroEL, DnaK-DnaJ-GrpE, and co-expression, for overexpression to increase folding efficiency [102].

7. Enhanced Efficiency

7.1. Solubilization

Despite careful selection of domain boundaries and solubilization tags, not all eukaryotic proteins can achieve proper folding in *Escherichia coli*, which possesses a diverse array of molecular chaperones (e.g., GroEL/ES, DnaK, Skp) and ten peptidyl cis-trans prolyl isomerases. Issues related to protein folding can be attributed to various factors, including translation rates, the formation of disulfide bonds through oxidative folding, the presence of essential post-translational modifications that *E. coli* cannot perform, the existence of buried prosthetic groups that wild-type *E. coli* cannot synthesize, or, in rare cases, the involvement of specialized folding factors. For instance, when attempting to express a hyperthermophilic α -amylase from *Pyrococcus furiosus* (a hyperthermophilic archaeum) in *E. coli*, it was found crucial to co-express small heat shock proteins (sHSP) or chaperonins (HSP60) from the same *P. furiosus* to facilitate proper folding. The folding and assembly of multi-subunit proteins continue to provide challenges in achieving proper folding and assembly of these proteins [103].

One effective approach for enhancing the solubility of recombinant products involves the utilization of peptide tags. Commonly tagged proteins include maltose-binding protein (MBP), glutathione S-transferase (GST), carbohydrate-binding module (CBM), thioredoxin, and NusA. Notably, novel CBM66 promotes the solubilization of several recombinant products and raises production titer. The NEXT tag [104], low-molecular-weight protamine [105], and 6HFh8 [106] are a few examples of tags smaller than recombinant proteins.

Fusion proteins, which assist in purifying, can incorporate solubilization tags. These tags, typically consisting of tiny, highly soluble, and stable proteins, facilitate the final product's solubilization and folding intermediates. Suppose a eukaryotic protein possesses a quantity of N-glycans exceeding one per every 100 amino acids. In that case, using a solubilization tag becomes necessary to facilitate the production of such protein in a soluble state within *Escherichia coli*. Solubilization tags, such as MBP (an affinity purification tag), thioredoxins, Sumo, or Fh8, are employed to maintain an optimal equilibrium for attaining soluble proteins. It is crucial to strike a balance, as excessive solubilization may impede proper protein folding. This step almost always requires a lot of trial and error. Different low-molecular-weight protein tags helped to solubilize and increase the yield of other RPs, only requiring fusion expression with recombinant proteins.

Protein solubility instantly alters when inclusion bodies form, leading to protein clumping. Though produced in a soluble form, the proteins can also end up as insoluble inclusion bodies that must be redissolved to refold into an active functional structure, primarily in the reducing environment of cytoplasm to improve their solubility. Insolubility or instability of *E. coli* proteins is managed by proper strain selection, the target of expression area, and post-translational modification [107]. Additionally, the localization of the protein to the periplasmic space can be achieved by introducing an N-terminal periplasmic signal sequence. Although the process of native disulfide production is often inflexible, it is possible for the Sec secretion system and the folding apparatus in the periplasm to get overrun with relative ease.

7.2. Disulfide Bond

The predominant challenge encountered is the development of native disulfide bonds. There exist three distinct approaches to address this issue. Initially, it is plausible for the protein to undergo the formation of misfolded or unfolded protein aggregates, commonly referred to as inclusion bodies. Aggregates within inclusion bodies caused by the cytoplasmic production of antibody fragments (Fab and scFv) in *E. coli* frequently result in

significant protein loss throughout the recovery process [108]. Removing cysteine residues from the recombinant antibody sequences is one way to enhance soluble expression. One choice is the periplasmic expression, although yields can be a problem. Misfolding, inadequate solubility, and host burden are other causes of IB formation.

It is possible to employ a genetically modified variant that eliminates pathways responsible for decreasing disulfide bonds inside the cytoplasm or introduces catalysts that facilitate oxidative folding [109]. Integrating the twin-arginine translocation (TAT)–secretion system can facilitate the exportation of adequately folded proteins to the periplasmic space. TatABC membrane protein overexpression improves the TAT translocation mechanism when the signal peptide TorA fusion RPs exploit the TAT translocation pathway [110].

Since disulfide bond (DSB) formation is an oxidative process, it occurs in the periplasm of *E. coli* rather than the cytoplasm, which is a reductive environment [111–115]. This calls for the localization and translocation of the protein to the correct location for the modification [116]. If proteins require a disulfide bond, then periplasm will be suitable because of its oxidizing properties. The reductive cytoplasmic milieu of a *gor*/*trxB* strain becomes oxidative when the normal reduction pathway is blocked, which promotes the production of DSBs [117]. Based on this idea, Novagen [118] created Origami, the first commercial DSB-forming *E. coli* strain. A host dubbed CyDisCo was also created to produce recombinant molecules with a high disulfide bond (DSB) by overproducing the human cell enzyme disulfide bond isomerase and sulfhydryl oxidase from yeast mitochondria [119] using different sulfhydryl oxidases, inversion, or the periplasmic transmembrane disulfide bond-forming enzyme DsbB are additional strategies [120].

The development begins by isolating and sequencing the light and heavy chains that appear in the variable chains from hybridoma or B-cells. These sequences are then introduced in *E. coli* after cloning into a plasmid. Additional requirements can involve adding constant regions for the heavy and light chains. The constructs include a signal peptide sequence for periplasmic localization and a tag for easier purification. The *E. coli* is transformed with the plasmid containing the antibody sequences to induce expression with an inducer like isopropyl-beta-d-thiogalactopyranoside (IPTG) if the promoter used is lac. Following the induction, the protein, including antibodies, can be extracted from the periplasmic space, wherein *E. coli* often secretes recombinant proteins that are further purified using an affinity column that targets the tag in the construct or other purification strategies based on the properties of the antibody. It is then tested for functionality using techniques like ELISA, flow cytometry, or Western blotting to verify that the expressed antibody can bind its antigen.

Yeast mitochondrial sulfhydryl oxidase and human cell disulfide bond isomerase have been overexpressed to produce the CyDisCo host [121].

7.3. Post-Translational Modifications

PTMs, which include acetylation, glycosylation, and phosphorylation, influence the functional activity of recombinant products [122]. While the functional activities increase by joining monosaccharides, oligosaccharides, or polysaccharides to proteins, glycosylation is the most prevalent and complex PTM [123]. Because *Campylobacter jejuni*-related genes must be added, *E. coli* must be glycoengineered from the bottom up because it lacks a natural mechanism for glycosylation [124]. The organism under investigation was able to facilitate glycoprotein synthesis by utilizing the first N-glycosylation expression system. In the last two decades, a considerable quantity of N/O-glycoproteins derived from *E. coli* or cell-free extracts have been generated. These advancements encompass the identification and orthogonality of glycosyltransferase substrates, investigations into the roles of diverse glycosylases, and enhancements in the host environment, metabolic pathways, and culture conditions [125–129]. These methods are used to create several products, such as the recombinant vaccine exotoxin A, the therapeutic protein O-glycosylated interferon-alpha2b [130], and N-glycosylated mannose3-N-acetylglucosamine2 [131]. Serine recombinase expression

increases, while the replication starts gene (oriC) is knocked down during late development stages, inhibiting host growth [132].

Recent research indicates that aglycosylated antibodies may have similar characteristics and functions, raising the possibility that glycosylation may not be required [133]. Despite minor biophysical changes in the melting temperatures (T_m) of gFc and aFc [124], the orientation of the CH2 domain of aFc in solution was only very slightly affected, according to a small angle X-ray scattering research. This observation is supported by the lack of discernible changes in the *E. coli*-produced antibodies' "drug-like" properties; one such reported example is the antibodies made using *E. coli* against COVID-19 that are not glycosylated yet proven as potent as other types of antibodies.

Methodologies in synthetic biology have been employed to facilitate the production of various PTMs within the cytoplasmic environment. One such example is the synthesis of mucin-type O-glycosylation in *Escherichia coli*. Combining oligosaccharide production pathways and ApNGT overexpression for cytoplasmic N-glycosylation assists in cytoplasmic glycosylation [134].

It is noteworthy that *E. coli*'s cytoplasm harbors methionine aminopeptidase, an enzyme capable of eliminating the initiating methionine residue based on the nature of the subsequent amino acids. Specifically, amino acids such as serine, alanine, cysteine, proline, or glycine at position P1' are preferred, while proline at position P2' inhibits this process [135]. Furthermore, additional amino acids can be included in this list by implementing engineered systems. Additionally, this phenomenon is interconnected with the N-end rule, which governs the process of protein degradation inside a cellular environment. In *Escherichia coli*, proteins with an N-terminal residue of Arginine (Arg), Lysine (Lys), Leucine (Leu), Phenylalanine (Phe), Tyrosine (Tyr), or Tryptophan (Trp) are susceptible to fast degradation. However, the extent of degradation is contingent upon the specific characteristics of the N-terminal residue and the subsequent amino acids.

7.4. Strain and Media

Growth conditions and media optimization [136] influence growth conditions and medium culture composition on recombinant protein production. Different strains of *E. coli* are designed to overcome specific challenges in protein expression, such as protein toxicity, formation of inclusion bodies, or codon usage. For example, BL21(DE3) and its derivatives BL21(DE3) are common strains used for recombinant protein production [137]. Rosetta is another strain that helps express proteins from organisms with rare codon usage [138].

To obtain optimal expression, the temperature of the upstream process should be normalized since the optimal temperature for *E. coli* growth is usually 37 °C [139], but lower temperatures, 30 °C or 25 °C, may promote proper folding and solubility of the expressed proteins [140].

The choice of media depends on the required growth rate, the need for specific selection markers, or the type of protein being expressed. The LB Media is the most used for protein expression [141]. Specialized media (e.g., M9) can be customized for specific needs [142].

The chromosomal integration of six tRNA genes with low abundance, namely in the BL21(DE3) strain, is carried out to enable their production in the presence of a ribosomal manipulator [143]. After developing an appropriate protein expression construct, the subsequent stage involves protein expression, wherein the utilization of the *E. coli* system offers notable advantages. *Escherichia coli* is a bacterial species that exhibits significant genetic diversity, with approximately 20% of the genome being shared among all strains. The categorization of strains can be broadly classified into four sub-groupings: K-12 strains, B-strains, and the C and W strains, which are distinguished depending on their first isolation. Recombinant protein manufacturing commonly involves the utilization of K-12 and B-strains in the context of K-12 and B-strain bacteria. Specific proteins notably depend on strain, frequently with unexplained underlying causes. Consequently, subjecting any novel protein to testing in both a K-12 and a B-strain has become customary. In a similar vein, there exists a diverse range of media options that can be categorized into two main

groups: rich media, characterized by the inclusion of yeast extract and/or other mixed sources of peptides like tryptone, and chemically defined or minimum media, typically consisting of only 1–3 carbon sources and a singular nitrogen source.

7.5. Fermentation Conditions

Moreover, it is possible to manipulate the fermentation conditions, including the temperature of the culture after induction. This factor can significantly impact the production of properly folded proteins. The effect is attributed to the alteration in relative hydrophobicity at different temperatures and the reduced protein translation rate. This adjustment is necessary to ensure the folding machinery can handle the protein load without being overwhelmed. Recombinant protein preparations are also designed to exclude lipopolysaccharides (LPS) as potential contaminants.

7.6. Purification

Protein purification methods may vary considerably depending on the host organism used for protein expression, including *Escherichia coli* (*E. coli*), a commonly used prokaryotic host, or various eukaryotic and recombinant cells.

The affinity tag removal following protein purification can be achieved through proteolysis, wherein enzymes with broad specificity, such as trypsin, are employed. This enzymatic process is utilized to eliminate an N-terminal tag and the C-peptide from insulin derivatives, depending on the intended application of the protein [144].

The tags can be removed through more specific proteases such as Tobacco Etch Virus (TEV) (consensus site ENLYFQ↓G/S) and Factor Xa (consensus site IE/DGR). It is worth noting that proteases' efficacy may depend on their source, such as recombinant bovine Factor Xa has different specificity than recombinant human Factor Xa [145,146]; consulting the MEROPS database will help decide on the choice of proteases [147]. By their specificity, proteins frequently exclude one or more amino acids from the cleavage site. To access buried cleavage sites, proteases necessitate the placement of the cleavage site within a flexible linker region, typically rich in glycine or serine. Consequently, this process leads to adding additional residues to the mature protein. The presence of proteases can degrade recombinant proteins, thereby diminishing yields and compromising stability.

8. Cell-Free Protein Synthesis System (CFPS)

Proteins can also be produced in a cell-free system, an in vitro process wherein even difficult-to-express molecules can be produced by manipulating the reaction media, which may not be possible when using cellular culture [148]. The basis of this technology is based on engaging ribosomes in vitro instead of in vivo in the cells to translate proteins of any type. CFPS systems provide a controlled environment for protein synthesis and offer several advantages, including rapid production, flexibility, and the ability to incorporate non-natural amino acids or post-translational modifications. The vitro technology utilizes the extracts of *E. coli* or other cell sources such as rabbit reticulocytes, insect cells, or wheat germ from commercial sources, making cell-free synthesis a substantially less capital-intensive and a practical choice [149]. The most widely used source remains *E. coli* [150].

Traditionally, cell-free protein synthesis has involved separate steps for transcription and translation. However, efforts have been made to develop systems that allow for the simultaneous coupling of transcription and translation, resulting in more efficient and streamlined protein synthesis processes [151]. CFPS systems require an energy source, such as adenosine triphosphate (ATP), and amino acids for protein synthesis. ATP can be generated by adding an energy-regenerating system, such as creatine phosphate and creatine kinase. A mixture of natural or non-natural amino acids is supplemented to support protein production [152].

Hybrid cell-free systems combine the advantages of different cell extracts or components from various organisms to create tailored environments for protein synthesis. By

leveraging the strengths of different sources, these hybrid systems can overcome limitations and expand the range of proteins that can be synthesized [153].

CFPS systems can be modified to incorporate non-natural amino acids into synthesized proteins, enabling the production of proteins with enhanced properties or novel functionalities. This is achieved by supplementing the reaction mixture with non-natural amino acids and using orthogonal translation systems that selectively incorporate these amino acids at specific codons [154].

Cell-free protein synthesis can be optimized by adjusting various parameters, such as reaction conditions, concentrations of components, and supplementation with specific factors. Optimization strategies include using different energy-regenerating systems, additives, or chaperones to enhance protein yield, folding, and functionality. Scale-up of CFPS can be achieved by increasing the reaction volume or employing high-throughput microscale platforms [155].

While CFPS primarily focuses on protein synthesis, efforts have been made to introduce post-translational modifications (PTMs) into cell-free systems. PTMs such as phosphorylation, glycosylation, and acetylation can be incorporated into proteins by adding the required enzymes or co-factors to the reaction mixture. This allows the production of proteins with functional PTMs for downstream applications [156].

Membrane proteins play crucial roles in various cellular processes and are challenging to produce using traditional expression systems. Cell-free protein synthesis offers a promising approach to producing membrane proteins, allowing for the incorporation of necessary membrane components and the utilization of specialized cell extracts [157].

DNA templates encoding the desired proteins are prepared and added to the cell extracts. These templates can be obtained through gene synthesis, PCR amplification, or isolation from natural sources. They typically contain an intense promoter sequence to drive transcription and a ribosome binding site (RBS) for efficient translation initiation [158].

Once the cell-free reaction mixture is prepared, it is incubated at an appropriate temperature, typically around 30–37 °C, to facilitate protein synthesis. The synthesized proteins can be harvested and purified using various techniques, including affinity chromatography, filtration, or precipitation [159].

Various platforms and technologies have been developed for cell-free protein synthesis, each offering unique advantages and capabilities. These include microfluidic, droplet-based, and cell-free systems integrated with nanomaterials or synthetic biology components [160].

The *E. coli*-based expression of proteins can also be adapted for novel technologies, particularly the cell-free synthesis system (CFPS) and continuous manufacturing (CM).

The FDA has not yet approved any biological drug manufactured using cell-free synthesis, but it is anticipated that biosimilars will adopt this technology since using a different technology is allowed in producing biosimilars; the cost factors will drive this decision.

9. Continuous Manufacturing (CM)

Continuous manufacturing (CM) of chemical and biological products has long been a goal to optimize the cost of manufacturing; however, the Current Good Manufacturing Practice (cGMP) compliance issues had pushed it back until March 2023, when the FDA released the first guideline to advise how to develop and adopt continuous manufacturing, particularly the biological products [161] (Figure 3). Continuous manufacturing requires a perfusion system, and it can be designed to use *E. coli*, which will be a better choice over the CHO cells due to a much shorter batch cycle, generally a few hours compared to weeks for the CHO cells. In *E. coli*, the proteins can be directed to the cytoplasm or periplasm or secreted directly into culture media, offering several choices on routing the recombinant protein, exploiting the features of each cellular compartment and the protein produced.

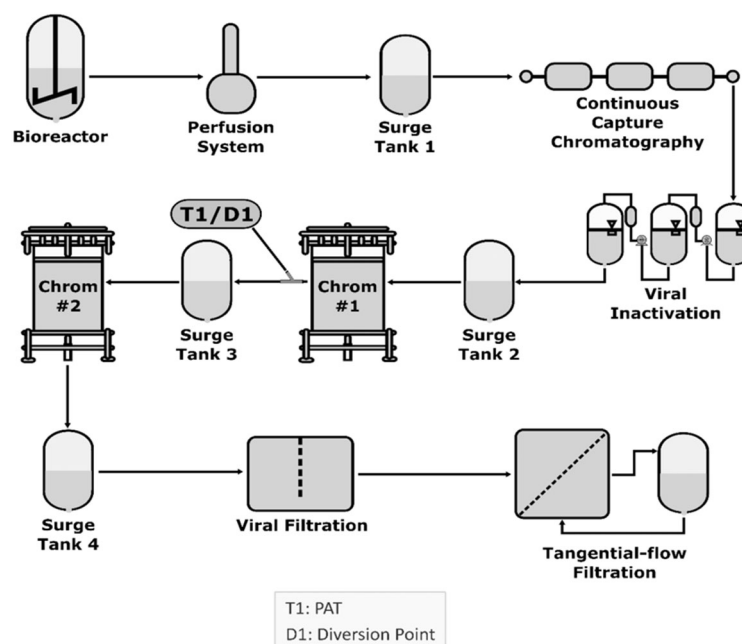


Figure 3. Continuous manufacturing system flow (as shown by the arrows) for therapeutic proteins as de-fined by the FDA. (Source: FDA.)

The quest for a CM process has been in the works for several years [162]. While the FDA has yet to approve a biological product manufactured in a continuous system, it anticipates much interest. As a result, in March 2023, the FDA released its first guidance on CM [163] addressing the scientific and regulatory issues, including the eCTD filing structure, which arose during the designing, installation, operation, and lifecycle management of CM for chemical and biological drugs. This guideline has opened the path to continuous systems over batch systems that will significantly impact the development and production cost, stability of proteins, and a significant reduction in the size of the bioreactors. Figure 2 shows a manufacturing flow in a CM for a therapeutic protein. The FDA also identifies other guidelines that control CM. The recombinant protein technology batch process is the industry standard. However, proteins can also be produced in a vessel, from which the yield is continuously removed, provided the protein is secreted into the culture medium. It helps to improve the yield of labile proteins and prevents inconsistent PTMs while maintaining cells at higher viabilities and critical factors [164]. Besides the material costs, the reduced testing also adds significant savings.

10. Conclusions

The dawn of recombinant engineering came with *E. coli* expression systems. However, it was overshadowed by eukaryote systems that proved more suitable to express more complex and larger recombinant proteins. The *E. coli* system is re-emerging for all proteins, including antibodies and soon the full-length ones, as a cost-effective option since the technologies have evolved that make it possible to engineer *E. coli* to express any protein, bringing down the production cost due to its shorter manufacturing cycle, avoidance of viral contamination, and provide the base material to bring CFPS to reality. *E. coli* can be designed for continuous manufacturing, a feat not possible a few years ago. Most limitations in the *E. coli* system have now been well-resolved because of deep knowledge about the *E. coli*; many technologies, including the updated BL21(DE3) genome, novel engineering, AI-based molecular docking, sensitive analytics, accurate optimization and fusing several engineering and mathematical tools; and many innovations that were not available to reference product when it was developed, likely 25 years ago [165]. The suggestions made in this paper should encourage both the developers of new biologics and biosimilars to consider adopting these improved and more recent technologies to help

bring down the cost of development and goods. There is a bright future for therapeutic proteins arriving, thanks to humanity's oldest friend, *E. coli*.

Author Contributions: Conceptualization, writing S.K.N.; investigation M.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors confirm that no funding was received to prepare this manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: Matthias Magoola is the CEO of the company DEI Biopharma, Uganda. Sarfaraz K. Niazi reports no conflict of interest. The company DEI Biopharma has no role in this paper.

References

- Landgraf, W.; Sandow, J. Recombinant Human Insulins—Clinical Efficacy and Safety in Diabetes Therapy. *Eur. Endocrinol.* **2016**, *12*, 12–17. [CrossRef]
- Approved Protein Drugs in the US and EU. Available online: <https://drugs.ncats.io/> (accessed on 10 July 2023).
- Dimitrov, D.S. Therapeutic proteins. In *Methods in Molecular Biology*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 899, pp. 1–26.
- Zhang, Z.X.; Nong, F.T.; Wang, Y.Z.; Yan, C.-X.; Gu, Y.; Song, P.; Sun, X.-M. Strategies for efficient production of recombinant proteins in *Escherichia coli*: Alleviating the host burden and enhancing protein activity. *Microb. Cell Fact.* **2022**, *21*, 191. [CrossRef]
- Zhang, J.; Zhao, Y.; Cao, Y.; Yu, Z.; Wang, G.; Li, Y.; Ye, X.; Li, C.; Lin, X.; Song, H. Synthetic sRNA-based engineering of *Escherichia coli* for enhanced production of full-length immunoglobulin G. *Biotechnol. J.* **2020**, *15*, e1900363. [CrossRef] [PubMed]
- Jackson, D.A.; Symons, R.H.; Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 2904–2909. [CrossRef]
- Cohen, S.N.; Chang, A.C.; Boyer, H.W.; Helling, R.B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3240–3244. [CrossRef] [PubMed]
- Feldman, M.P.; Colaianni, A.; Liu, K. Lessons from the Commercialization of the Cohen-Boyer Patents: The Stanford University Licensing Program. Handbook of Best Practices. In *Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices*; MIHR: Oxford, UK; PIPRA: Davis, CA, USA, 2007. Available online: <https://maryannfeldman.web> (accessed on 10 July 2023).
- Blattner, F.R.; Plunkett, G., 3rd; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462. [CrossRef] [PubMed]
- Kozak, M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **1983**, *47*, 1–45. [CrossRef] [PubMed]
- Jacob, F.; Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **1961**, *3*, 318–356. [CrossRef] [PubMed]
- Terpe, K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 211–222. [CrossRef]
- Wang, Y.; Liu, M.; Wei, Q.; Wu, W.; He, Y.; Gao, J.; Zhou, R.; Jiang, L.; Qu, J.; Xia, J. Phase-Separated Multienzyme Compartmentalization for Terpene Biosynthesis in a Prokaryote. *Angew. Chem. Int. Ed.* **2022**, *8*, 61–69.
- Wei, S.-P.; Qian, Z.-G.; Hu, C.-F.; Pan, F.; Chen, M.-T.; Lee, S.Y.; Xia, X.-X. Formation and functionalization of membraneless compartments in *Escherichia coli*. *Nat. Chem. Biol.* **2020**, *16*, 1143–1148. [CrossRef] [PubMed]
- McElwain, L.; Phair, K.; Kealey, C.; Brady, D. Current trends in biopharmaceuticals production in *Escherichia coli*. *Biotechnol. Lett.* **2022**, *44*, 917–931. [CrossRef] [PubMed]
- Mueller, M.; Grauschopf, U.; Maier, T.; Glockshuber, R.; Ban, N. The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **2009**, *459*, 726–730. [CrossRef]
- Available online: <https://www.fda.gov/drugs/regulatory-science-action/impact-continuous-manufacturing-processes-viral-safety-therapeutic-proteins> (accessed on 10 July 2023).
- Tripathi, N.K.; Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front. Bioeng. Biotechnol.* **2019**, *7*, 420. [CrossRef]
- Fahnert, B.; Lilie, H.; Neubauer, P. Inclusion bodies: Formation and utilisation. *Adv. Biochem. Eng./Biotechnol.* **2004**, *89*, 93–142. [PubMed]
- Butt, T.R.; Edavettal, S.C.; Hall, J.P.; Mattern, M.R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **2005**, *43*, 1–9. [CrossRef] [PubMed]
- Thomas, J.G.; Baneyx, F. Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing Heat-shock proteins. *J. Biol. Chem.* **1996**, *271*, 11141–11147. [CrossRef]

22. Vasina, J.A.; Baneyx, F. Recombinant protein expression at low temperatures under the transcriptional control of the major *Escherichia coli* cold shock promoter cspA. *Appl. Environ. Microbiol.* **1996**, *62*, 1444–1447. [[CrossRef](#)]
23. Singh, S.M.; Panda, A.K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **2005**, *99*, 303–310. [[CrossRef](#)]
24. Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **2006**, *24*, 1241–1252. [[CrossRef](#)]
25. Magalhães, P.O.; Lopes, A.M.; Mazzola, P.G.; Rangel-Yagui, C.; Penna, T.C.V.; Pessoa, A., Jr. Methods of endotoxin removal from biological preparations: A review. *J. Pharm. Pharm. Sci.* **2007**, *10*, 388–404.
26. Baneyx, F.; Mujacic, M. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **2004**, *22*, 1399–1408. [[CrossRef](#)]
27. Mergulhão, F.J.; Summers, D.K.; Monteiro, G.A. Recombinant protein secretion in *Escherichia coli*. *Biotechnol. Adv.* **2005**, *23*, 177–202. [[CrossRef](#)]
28. Gustafsson, C.; Govindarajan, S.; Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **2004**, *22*, 346–353. [[CrossRef](#)] [[PubMed](#)]
29. Wurm, F.M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **2004**, *22*, 1393–1398. [[CrossRef](#)] [[PubMed](#)]
30. Lobstein, J.; Emrich, C.A.; Jeans, C.; Faulkner, M.; Riggs, P.; Berkmen, M. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **2012**, *11*, 56. [[CrossRef](#)] [[PubMed](#)]
31. Skretas, G.; Georgiou, G. Simple genetic selection protocol for isolation of overexpressed genes that enhance accumulation of membrane-integrated human G protein-coupled receptors in *Escherichia coli*. *Appl. Environ. Microbiol.* **2009**, *75*, 3853–3863. [[CrossRef](#)]
32. Wang, S.; Chen, K.; Lei, Q.; Ma, P.; Yuan, A.Q.; Zhao, Y.; Jiang, Y.; Fang, H.; Xing, S.; Fang, Y.; et al. The state of the art of bispecific antibodies for treating human malignancies. *EMBO Mol. Med.* **2021**, *13*, e14291. [[CrossRef](#)] [[PubMed](#)]
33. Atwell, S.; Ridgway, J.B.; Wells, J.A.; Carter, P. Stable heterodimers from remodeling the domain interface of a homodimer using a phage display library. *J. Mol. Biol.* **1997**, *270*, 26–35. [[CrossRef](#)] [[PubMed](#)]
34. Merchant, M.; Ma, X.; Maun, H.R.; Zheng, Z.; Peng, J.; Romero, M.; Huang, A.; Yang, N.Y.; Nishimura, M.; Greve, J.; et al. Monovalent antibody design and mechanism of action of Onartuzumab, a MET antagonist with anti-tumor activity as a therapeutic agent. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E2987–E2996. [[CrossRef](#)]
35. Spiess, C.; Bevers, J.; Jackman, J.; Chiang, N.; Nakamura, G.; Dillon, M.; Liu, H.; Molina, P.; Elliott, J.M.; Shatz, W.; et al. Development of a human IgG4 bispecific antibody for dual targeting of interleukin-4 (IL-4) and interleukin-13 (IL-13) cytokines. *J. Biol. Chem.* **2013**, *288*, 26583–26593. [[CrossRef](#)]
36. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)] [[PubMed](#)]
37. Darlington, P.J.; Kirchhof, M.G.; Criado, G.; Sondhi, J.; Madrenas, J. Hierarchical Regulation of CTLA-4 Dimer-Based Lattice Formation and Its Biological Relevance for T Cell Inactivation. *J. Immunol.* **2005**, *175*, 996–1004. [[CrossRef](#)] [[PubMed](#)]
38. Manta, B.; Boyd, D.; Berkmen, M. Disulfide Bond Formation in the Periplasm of *Escherichia coli*. *EcoSal Plus* **2019**, *8*, 10–1128. [[CrossRef](#)] [[PubMed](#)]
39. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **2015**, *43*, W389–W394. [[CrossRef](#)] [[PubMed](#)]
40. Demarest, S.J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H.J.; Evans, R.M.; Wright, P.E. Mutual synergistic folding in recruitment of cbp/p300 by p160 nuclear receptor coactivators. *Nature* **2002**, *415*, 549–553. [[CrossRef](#)]
41. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A.W.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [[CrossRef](#)]
42. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
43. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **1999**, *12*, 85–94. [[CrossRef](#)]
44. Radivojac, P.; Clark, W.T.; Oron, T.R.; Schnoes, A.M.; Wittkop, T.; Sokolov, A.; Graim, K.; Funk, C.; Verspoor, K.; Ben-Hur, A.; et al. A large-scale evaluation of computational protein function prediction. *Nat. Methods* **2013**, *10*, 221–227. [[CrossRef](#)]
45. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652. [[CrossRef](#)] [[PubMed](#)]
46. Blom, N.; Gammeltoft, S.; Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999**, *294*, 1351–1362. [[CrossRef](#)]
47. Vakser, I.A. Protein docking for low-resolution structures. *Protein Eng.* **1995**, *8*, 371–377. [[CrossRef](#)] [[PubMed](#)]
48. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 5th ed.; Garland Science: New York, NY, USA, 2008; ISBN 978-0-8153-4105-5.
49. Sillitoe, I.; Lewis, T.E.; Cuff, A.; Das, S.; Ashford, P.; Dawson, N.L.; Furnham, N.; Laskowski, R.A.; Lee, D.; Lees, J.G.; et al. CATH: Comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* **2015**, *43*, D376–D381. [[CrossRef](#)] [[PubMed](#)]
50. von Heijne, G. The signal peptide. *J. Membr. Biol.* **1990**, *115*, 195–201. [[CrossRef](#)] [[PubMed](#)]

51. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)] [[PubMed](#)]
52. Levy, E.D.; Teichmann, S.A. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **2013**, *117*, 25–51.
53. Walsh, C.T.; Garneau-Tsodikova, S.; Gatto, G.J., Jr. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **2005**, *44*, 7342–7372. [[CrossRef](#)]
54. Kudla, G.; Murray, A.W.; Tollervey, D.; Plotkin, J.B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **2009**, *324*, 255–258. [[CrossRef](#)]
55. Subramanian, S.; Ross, P.D. Dye-ligand affinity chromatography: The interaction of cibacron blue f3GA® with proteins and enzyme. *Crit. Rev. Biochem. Mol. Biol.* **1984**, *16*, 169–205. [[CrossRef](#)]
56. Kish, W.S.; Roach, M.K.; Sachi, H.; Naik, A.D.; Menegatti, S.; Carbonell, R.G. Purification of human erythropoietin by affinity chromatography using cyclic peptide ligands. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2018**, *1085*, 1–12. [[CrossRef](#)] [[PubMed](#)]
57. Young, C.L.; Britton, Z.T.; Robinson, A.S. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **2012**, *7*, 620–634. [[CrossRef](#)] [[PubMed](#)]
58. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)] [[PubMed](#)]
59. Geoghegan, K.F.; Dixon, H.B.F.; Rosner, P.J.; Hoth, L.R.; Lanzetti, A.J.; Borzilleri, K.A.; Marr, E.S.; Pezzullo, L.H.; Martin, L.B.; LeMotte, P.K.; et al. Spontaneous α -N-6-phosphogluconoylation of a “His tag” in *Escherichia coli*: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **1999**, *267*, 169–184. [[CrossRef](#)] [[PubMed](#)]
60. Wood, W.N.; Smith, K.D.; Ream, J.A.; Lewis, L.K. Enhancing yields of low and single copy number plasmid DNAs from *Escherichia coli* cells. *J. Microbiol. Methods* **2017**, *133*, 46–51. [[CrossRef](#)] [[PubMed](#)]
61. Jeong, K.J.; Lee, S.Y. High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification. *Appl. Environ. Microbiol.* **1999**, *65*, 3027–3032. [[CrossRef](#)]
62. Sharma, A.K.; Shukla, E.; Janoti, D.S.; Mukherjee, K.J.; Shiloach, J. A novel knock out strategy to enhance recombinant protein expression in *Escherichia coli*. *Microb. Cell Fact.* **2020**, *19*, 148. [[CrossRef](#)]
63. Ganoza, M.C.; Kiel, M.C.; Aoki, H. Evolutionary conservation of reactions in translation. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 460–485. [[CrossRef](#)]
64. Irastorza-Olaziregi, M.; Amster-Choder, O. Coupled Transcription-Translation in Prokaryotes: An Old Couple with New Surprises. *Front. Microbiol.* **2021**, *11*, 624830. [[CrossRef](#)]
65. Hui, A.; De Boer, H.A. Specialized ribosome system: Preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4762–4766. [[CrossRef](#)]
66. Dixon, N.; Robinson, C.J.; Geerlings, T.; Duncan, J.N.; Drummond, S.P.; Micklefield, J. Orthogonal Riboswitches for Tuneable Coexpression in Bacteria. *Angew. Chem. Int. Ed.* **2012**, *51*, 3620–3624. [[CrossRef](#)] [[PubMed](#)]
67. Orelle, C.; Carlson, E.D.; Szal, T.; Florin, T.; Jewett, M.C.; Mankin, A.S. Protein synthesis by ribosomes with tethered subunits. *Nature* **2015**, *524*, 119–124. [[CrossRef](#)] [[PubMed](#)]
68. Morra, R.; Shankar, J.; Robinson, C.J.; Halliwell, S.; Butler, L.; Upton, M.; Hay, S.; Micklefield, J.; Dixon, N. Dual transcriptional-Translational cascade permits cellular level tuneable expression control. *Nucleic Acids Res.* **2016**, *44*, 21. [[CrossRef](#)] [[PubMed](#)]
69. Carlson, E.D.; d’Aquino, A.E.; Kim, D.S.; Fulk, E.M.; Hoang, K.; Szal, T.; Mankin, A.S.; Jewett, M.C. Engineered ribosomes with tethered subunits for expanding biological function. *Nat. Commun.* **2019**, *10*, 3920. [[CrossRef](#)] [[PubMed](#)]
70. Sharp, P.M.; Li, W.H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [[CrossRef](#)]
71. Salis, H.M.; Mirsky, E.A.; Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **2009**, *27*, 946–950. [[CrossRef](#)]
72. Rosano, G.L.; Ceccarelli, E.A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **2014**, *5*, 172. [[CrossRef](#)]
73. Zhang, G.; Darst, S.A. Structure of the *Escherichia coli* RNA polymerase α subunit amino-terminal domain. *Science* **1998**, *281*, 262–266. [[CrossRef](#)]
74. Chen, H.; Bjerknes, M.; Kumar, R.; Jay, E. Determination of the optimal aligned spacing between the shine-dalgarno sequence and the translation initiation codon of *Escherichia coli* m RNAs. *Nucleic Acids Res.* **1994**, *22*, 4953–4957. [[CrossRef](#)]
75. Shepard, H.M.; Yelverton, E.; Goeddel, D.V. Increased Synthesis in *E. coli* of Fibroblast and Leukocyte Interferons Through Alterations in Ribosome Binding Sites. *DNA* **1982**, *1*, 125–131.
76. Shine, J.; Dalgarno, L. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 1342–1346. [[CrossRef](#)]
77. Jeong, H.; Barbe, V.; Lee, C.H.; Vallenet, D.; Yu, D.S.; Choi, S.-H.; Couloux, A.; Lee, S.-W.; Yoon, S.H.; Cattolico, L. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* **2009**, *394*, 644–652. [[CrossRef](#)]
78. Du, F.; Liu, Y.-Q.; Xu, Y.S.; Li, Z.J.; Wang, Y.Z.; Zhang, Z.X.; Sun, X.M. Regulating the T7 RNA polymerase expression in *E. coli* BL21(DE3) to provide more host options for recombinant protein production. *Microb. Cell Fact.* **2021**, *20*, 189. [[CrossRef](#)]

79. Khlebnikov, A.; Risa, Ø.; Skaug, T.; Carrier, T.A.; Keasling, J.D. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* **2000**, *182*, 7029–7034. [\[CrossRef\]](#)
80. Lutz, R.; Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **1997**, *25*, 1203–1210. [\[CrossRef\]](#)
81. Mueller, K.L.; Simon, J.D.; Elf, J. Design, construction, and implementation of a fully repressible bistable genetic switch in *E. coli*. *Nucleic Acids Res.* **2019**, *47*, 6307–6317.
82. Sun, X.M.; Zhang, Z.X.; Wang, L.R.; Wang, J.G.; Liang, Y.; Yang, H.F.; Tao, R.S.; Jiang, Y.; Yang, J.J.; Yang, S. Downregulation of T7 RNA polymerase transcription enhances pET-based recombinant protein production in *Escherichia coli* BL21(DE3) by suppressing autolysis. *Biotechnol. Bioeng.* **2021**, *118*, 153–163. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Kim, S.K.; Lee, D.-H.; Kim, O.C.; Kim, J.F.; Yoon, S.H. Tunable control of an *Escherichia coli* expression system for the overproduction of membrane proteins by titrated expression of a mutant lac repressor. *ACS Synth. Biol.* **2017**, *6*, 1766–1773. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Li, Z.J.; Zhang, Z.X.; Xu, Y.; Shi, T.Q.; Ye, C.; Sun, X.M.; Huang, H. CRISPR-Based Construction of a BL21 (DE3)-derived variant strain library to rapidly improve recombinant protein production. *ACS Synth. Biol.* **2022**, *11*, 343–352. [\[CrossRef\]](#)
85. Baumschlager, A.; Aoki, S.K.; Khammash, M. Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control. *ACS Synth. Biol.* **2017**, *6*, 2157–2167. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Rouches, M.V.; Xu, Y.; Cortes, L.B.G.; Lambert, G. A plasmid system with tunable copy number. *Nat. Commun.* **2022**, *13*, 3908. [\[CrossRef\]](#)
87. Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **1981**, *146*, 1–21. [\[CrossRef\]](#)
88. Boël, G.; Letso, R.; Neely, H.; Price, W.N.; Wong, K.H.; Su, M.; Luff, J.D.; Valecha, M.; Everett, J.K.; Acton, T.B.; et al. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* **2016**, *529*, 358–363. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Fuhrmann, M.; Hausherr, A.; Ferbitz, L.; Schödl, T.; Heitzer, M.; Hegemann, P. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol. Biol.* **2004**, *55*, 869–881. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Kleber-Janke, T.; Becker, W.M. Use of modified BL21(DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage. *Protein Expr. Purif.* **2000**, *19*, 419–424. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Novy, R.; Drott, D.; Yaeger, K.; Mierendorf, R. Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations* **2001**, *12*, 1–3.
92. Komar, A.A. The Yin and Yang of codon usage. *Hum. Mol. Genet.* **2016**, *25*, R77–R85. [\[CrossRef\]](#)
93. Chemla, Y.; Peeri, M.; Heltberg, M.L.; Eichler, J.; Jensen, M.H.; Tuller, T.; Alfonta, L. A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon. *Nat. Commun.* **2020**, *11*, 4827. [\[CrossRef\]](#)
94. Lenz, G.; Doron-Faigenboim, A.; Ron, E.Z.; Tuller, T.; Gophna, U. Sequence Features of *E. coli* mRNAs Affect Their Degradation. *PLoS ONE* **2011**, *6*, e28544.
95. Siller, E.; DeZwaan, D.C.; Anderson, J.F.; Freeman, B.C.; Barral, J.M. Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. *J. Mol. Biol.* **2010**, *396*, 1310–1318. [\[CrossRef\]](#)
96. Angov, E.; Hillier, C.J.; Kincaid, R.L.; Lyon, J.A. Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host. *PLoS ONE* **2008**, *3*, e2189. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Gasser, B.; Saloheimo, M.; Rinas, U.; Dragosits, M.; Rodríguez-Carmona, E.; Baumann, K.; Giuliani, M.; Parrilli, E.; Branduardi, P.; Lang, C.; et al. Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview. *Microb. Cell Fact.* **2008**, *7*, 11. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Goemans, C.; Denoncin, K.; Collet, J.F. Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta (BBA)-Mol. Cell Res.* **2014**, *1843*, 1517–1528. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Skerra, A.; Pluckthun, A. Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* **1988**, *240*, 1038–1041. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Chen, C.; Snedecor, B.; Nishihara, J.C.; Joly, J.C.; McFarland, N.; Andersen, D.C.; Battersby, J.E.; Champion, K.M. High-level accumulation of a recombinant antibody fragment in the periplasm of *Escherichia coli* requires a triple-mutant (degP prc spr) host strain. *Biotechnol. Bioeng.* **2004**, *85*, 463–474. [\[CrossRef\]](#) [\[PubMed\]](#)
101. Carter, P.; Kelley, R.F.; Rodrigues, M.L.; Snedecor, B.; Covarrubias, M.; Velligan, M.D.; Wong, W.L.T.; Rowland, A.M.; Kotts, C.E.; Carver, M.E.; et al. High level *Escherichia coli* expression and production of a bivalent humanized antibody fragment. *Biotechnology* **1992**, *10*, 163–167. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Huang, M.N.; Lu, X.Y.; Zong, H.; Bin, Z.G.; Shen, W. Bioproduction of trans-10, cis-12-Conjugated Linoleic Acid by a Highly Soluble and Conveniently Extracted Linoleic Acid Isomerase and an Extracellularly Expressed Lipase from Recombinant *Escherichia coli* Strains. *J. Microbiol. Biotechnol.* **2018**, *28*, 739–747. [\[CrossRef\]](#)
103. Guzmán, L.M.; Belin, D.; Carson, M.J.; Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **1995**, *177*, 4121–4130. [\[CrossRef\]](#)
104. Jo, B.H. An intrinsically disordered peptide tag that confers an unusual solubility to aggregation-prone proteins. *Appl. Environ. Microbiol.* **2022**, *88*, e00097–22. [\[CrossRef\]](#)
105. Choi, S.W.; Pangeni, R.; Jung, D.H.; Kim, S.J.; Park, J.W. Construction and characterization of cell-penetrating peptide-fused fibroblast growth factor and vascular endothelial growth factor for an enhanced percutaneous delivery system. *J. Nanosci. Nanotechnol.* **2018**, *18*, 842–847. [\[CrossRef\]](#)

106. Kim, Y.S.; Lee, H.-J.; Han, M.-H.; Yoon, N.-K.; Kim, Y.-C.; Ahn, J. Effective production of human growth factors in *Escherichia coli* by fusing with small protein 6HFh8. *Microb. Cell Fact.* **2021**, *20*, 9. [CrossRef]
107. Chowdhury, T.; Chien, P.; Ebrahim, S.; Sauer, R.T.; Baker, T.A. Versatile modes of peptide recognition by the ClpX N domain mediate alternative adaptor-binding specificities in different bacterial species. *Protein Sci.* **2010**, *19*, 242–254. [CrossRef] [PubMed]
108. Cabilly, S.; Riggs, A.D.; Pande, H.; Shively, J.E.; Holmes, W.E.; Rey, M.; Perry, L.J.; Wetzel, R.; Heyneker, H.L. Generation of antibody activity from immunoglobulin polypeptide chains produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 3273–3277. [CrossRef] [PubMed]
109. Saaranen, M.J.; Ruddock, L.W. Applications of catalyzed cytoplasmic disulfide bond formation. *Biochem. Soc. Trans.* **2019**, *47*, 1223–1231. [CrossRef]
110. 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. *Escherichia coli* “TatExpress” strains export several g/L human growth hormone to the periplasm by the Tat pathway. *Biotechnol. Bioeng.* **2019**, *116*, 3282–3291.
111. de Marco, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.* **2009**, *8*, 26. [CrossRef] [PubMed]
112. Karyolaimos, A.; de Gier, J.W. Strategies to Enhance Periplasmic Recombinant Protein Production Yields in *Escherichia coli*. *Front Bioeng Biotechnol.* **2021**, *9*, 797334. [CrossRef] [PubMed]
113. Kipriyanov, S.M.; Little, M. Affinity purification of tagged recombinant proteins using immobilized single chain Fv fragments. *Anal Biochem.* **1997**, *244*, 189–191. [CrossRef]
114. Menon, V.; Thomas, R.; Ghale, A.R.; Reinhard, C.; Pruszk, J. Flow cytometry protocols for surface and intracellular antigen analyses of neural cell types. *J. Vis. Exp.* **2014**, *18*, 52241.
115. Brinkmann, U.; Mattes, R.E. High-level expression of a human immunoglobulin fragment Fab in *Escherichia coli*. *Gene* **1989**, *85*, 517–521.
116. Ma, Y.; Lee, C.J.; Park, J.S. Strategies for Optimizing the Production of Proteins and Peptides with Multiple Disulfide Bonds. *Antibiotics* **2020**, *9*, 541. [CrossRef] [PubMed]
117. Derman, A.I.; Prinz, W.A.; Belin, D.; Beckwith, J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **1993**, *262*, 1744–1747. [CrossRef] [PubMed]
118. Available online: https://www.emdmillipore.com/US/en/product/Origami-BDE3-Competent-Cells-Novagen,EMD_BIO-70837 (accessed on 10 July 2023).
119. Zhang, W.; Zheng, W.; Mao, M.; Yang, Y. Highly efficient folding of multidisulfide proteins in superoxidizing *Escherichia coli* cytoplasm. *Biotechnol. Bioeng.* **2014**, *111*, 2520–2527. [CrossRef] [PubMed]
120. Hatahet, F.; Ruddock, L.W. Topological plasticity of enzymes involved in disulfide bond formation allows catalysis in either the periplasm or the cytoplasm. *J. Mol. Biol.* **2013**, *425*, 3268–3276. [CrossRef] [PubMed]
121. Sohail, A.A.; Gaikwad, M.; Khadka, P.; Saaranen, M.J.; Ruddock, L.W. Production of extracellular matrix proteins in the cytoplasm of *E. coli*: Making giants in tiny factories. *Int. J. Mol. Sci.* **2020**, *21*, 688. [CrossRef]
122. Lapteva, Y.S.; Vologzhanikova, A.A.; Sokolov, A.S.; Ismailov, R.G.; Uversky, V.N.; Permyakov, S.E. In Vitro N-Terminal Acetylation of Bacterially Expressed Parvalbumins by N-Terminal Acetyltransferases from *Escherichia coli*. *Appl. Biochem. Biotechnol.* **2021**, *193*, 1365–1378. [CrossRef] [PubMed]
123. Eichler, J.; Koomey, M. Sweet new roles for protein glycosylation in prokaryotes. *Trends Microbiol.* **2017**, *25*, 662–672. [CrossRef]
124. Harding, C.M.; Feldman, M.F. Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in *E. coli*. *Glycobiology* **2019**, *29*, 519–529. [CrossRef]
125. Wacker, M.; Linton, D.; Hitchen, P.G.; Nita-Lazar, M.; Haslam, S.M.; North, S.J.; Panico, M.; Morris, H.R.; Dell, A.; Wren, B.W. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **2002**, *298*, 1790–1793. [CrossRef]
126. Silverman, J.M.; Imperiali, B. Bacterial N-glycosylation efficiency is dependent on the structural context of target sequence. *J. Biol. Chem.* **2016**, *291*, 22001–22010. [CrossRef]
127. Ollis, A.A.; Zhang, S.; Fisher, A.C.; DeLisa, M.P. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat. Chem. Biol.* **2014**, *10*, 816–822. [CrossRef]
128. Kightlinger, W.; Warfel, K.F.; DeLisa, M.P.; Jewett, M.C. Synthetic glycobiology: Parts, systems, and applications. *ACS Synth. Biol.* **2020**, *9*, 1534–1562. [CrossRef] [PubMed]
129. Pratama, F.; Linton, D.; Dixon, N. Genetic and process engineering strategies for enhanced recombinant N-glycoprotein production in bacteria. *Microb. Cell Fact.* **2021**, *20*, 198. [CrossRef]
130. Valderrama-Rincon, J.D.; Fisher, A.C.; Merritt, J.H.; Fan, Y.-Y.; Reading, C.A.; Chhiba, K.; Heiss, C.; Azadi, P.; Aebi, M.; DeLisa, M.P. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat. Chem. Biol.* **2012**, *8*, 434–436. [CrossRef] [PubMed]
131. Du, T.; Buenbrazo, N.; Kell, L.; Rahmani, S.; Sim, L.; Withers, S.G.; DeFrees, S.; Wakarchuk, W. A bacterial expression platform for production of therapeutic proteins containing human-like O-linked glycans. *Cell Chem. Biol.* **2019**, *26*, 203–212. [CrossRef] [PubMed]
132. Kasari, M.; Kasari, V.; Kärmas, M.; Jöers, A. Decoupling growth and production by removing the origin of replication from a bacterial chromosome. *ACS Synth. Biol.* **2022**, *11*, 2610–2622. [CrossRef] [PubMed]

133. Leabman, M.K.; Meng, Y.G.; Kelley, R.F.; DeForge, L.E.; Cowan, K.J.; Iyer, S. Effects of altered FcγR binding on antibody pharmacokinetics in cynomolgus monkeys. *mAbs* **2013**, *5*, 896–903. [\[CrossRef\]](#)
134. Tytgat, H.L.; Lin, C.-W.; Levasseur, M.D.; Tomek, M.B.; Rutschmann, C.; Mock, J.; Liebscher, N.; Terasaka, N.; Azuma, Y.; Wetter, M. Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. *Nat. Commun.* **2019**, *10*, 5403. [\[CrossRef\]](#)
135. Gallwitz, M.; Enoksson, M.; Thorpe, M.; Hellman, L. The extended cleavage specificity of human thrombin. *PLoS ONE* **2012**, *7*, e31756. [\[CrossRef\]](#)
136. Gasser, B.; Mattanovich, D. Antibiotic resistance marker genes in *Saccharomyces cerevisiae*: An overview on different mechanisms and new approaches for the elimination. *FEMS Yeast Res.* **2004**, *4*, 235–245.
137. Studier, F.W.; Moffatt, B.A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **1986**, *189*, 113–130. [\[CrossRef\]](#)
138. Miroux, B.; Walker, J.E. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **1996**, *260*, 289–298. [\[CrossRef\]](#) [\[PubMed\]](#)
139. Inada, T.; Kimata, K.; Aiba, H. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: Challenge to the cAMP model. *Genes Cells* **1996**, *1*, 293–301. [\[CrossRef\]](#) [\[PubMed\]](#)
140. San-Miguel, T.; Pérez-Bermúdez, P.; Gavidia, I. Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *Springerplus* **2013**, *2*, 89. [\[PubMed\]](#)
141. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **1951**, *62*, 293–300. [\[CrossRef\]](#) [\[PubMed\]](#)
142. Neidhardt, F.C.; Bloch, P.L.; Smith, D.F. Culture medium for enterobacteria. *J. Bacteriol.* **1974**, *119*, 736–747. [\[CrossRef\]](#)
143. Lipinski, Z.; Verniyk, V.; Farago, N.; Sari, T.; Puskas, L.G.; Blattner, F.R.; Posfai, G.; Gyorfy, Z. Enhancing the translational capacity of *E. coli* by resolving the codon bias. *ACS Synth. Biol.* **2018**, *7*, 2656–2664. [\[CrossRef\]](#) [\[PubMed\]](#)
144. Castellanos-Serra, L.R.; Hardy, E.; Ubieta, R.; Vispo, N.S.; Fernandez, C.; Besada, V.; Falcon, V.; Gonzalez, M.; Santos, A.; Perez, G.; et al. Expression and folding of an interleukin-2-proinsulin fusion protein and its conversion into insulin by a single step enzymatic removal of the C-peptide and the N-terminal fused sequence. *FEBS Lett.* **1996**, *378*, 171–176. [\[CrossRef\]](#)
145. Ludeman, J.P.; Pike, R.N.; Bromfield, K.M.; Duggan, P.J.; Cianci, J.; Le Bonniec, B.; Whisstock, J.C.; Bottomley, S.P. Determination of the P'1, P'2 and P'3 subsite-specificity of factor Xa. *Int. J. Biochem. Cell Biol.* **2003**, *35*, 221–225. [\[CrossRef\]](#)
146. Bianchini, E.P.; Louvain, V.B.; Marque, P.E.; Juliano, M.A.; Juliano, L.; Le Bonniec, B.F. Mapping of the catalytic groove preferences of factor Xa reveals an inadequate selectivity for its macromolecule substrates. *J. Biol. Chem.* **2002**, *277*, 20527–20534. [\[CrossRef\]](#)
147. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [\[CrossRef\]](#) [\[PubMed\]](#)
148. Smolskaya, S.; Logashina, Y.A.; Andreev, Y.A. *Escherichia coli* Extract-Based Cell-Free Expression System as an Alternative for Difficult-to-Obtain Protein Biosynthesis. *Int. J. Mol. Sci.* **2020**, *21*, 928. [\[CrossRef\]](#)
149. Chong, S. Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Curr. Protoc. Mol. Biol.* **2013**, *108*, 16.30.1–16.30.11. [\[CrossRef\]](#) [\[PubMed\]](#)
150. Spirin, A.S.; Baranov, V.I.; Ryabova, L.A.; Ovodov, S.; Alakhov, Y.B. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **1988**, *242*, 1162–1164. [\[CrossRef\]](#)
151. Wang, S.; Majumder, S.; Emery, N.J.; Liu, A.P. Simultaneous monitoring of transcription and translation in mammalian cell-free expression in bulk and in cell-sized droplets. *Synth. Biol.* **2018**, *3*, ysy005. [\[CrossRef\]](#) [\[PubMed\]](#)
152. Shin, J.; Noireaux, V. An *E. coli* cell-free expression toolbox: Application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* **2012**, *1*, 29–41.
153. Kwon, Y.C.; Jewett, M.C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* **2015**, *5*, 8663. [\[CrossRef\]](#)
154. Ge, X.; Luo, D.; Xu, J. Cell-free protein expression under macromolecular crowding conditions. *PLoS ONE* **2011**, *6*, e28707. [\[CrossRef\]](#)
155. Ho, D.; Quake, S.R.; McCabe, E.R.B.; Chng, W.J.; Chow, E.K.; Ding, X.; Gelb, B.D.; Ginsburg, G.S.; Hassenstab, J.; Ho, C.M.; et al. Enabling Technologies for Personalized and Precision Medicine. *Trends Biotechnol.* **2020**, *38*, 497–518. [\[CrossRef\]](#)
156. Caschera, F.; Noireaux, V. Integration of biological parts toward the synthesis of a minimal cell. *Curr. Opin. Chem. Biol.* **2014**, *22*, 85–91. [\[CrossRef\]](#)
157. Klammt, C.; Lohr, F.; Schafer, B.; Haase, W.; Dotsch, V.; Ruterjans, H.; Glaubitz, C.; Bernhard, F. High-level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **2006**, *273*, 720–734. [\[CrossRef\]](#) [\[PubMed\]](#)
158. Zhang, Y.; Forster, A.C. Synthesis and applications of DNA constructs for cell-free expression systems. *Nucleic Acids Res.* **2006**, *34*, e9.
159. Jewett, M.C.; Swartz, J.R. Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol Bioeng.* **2004**, *86*, 19–26. [\[CrossRef\]](#)
160. Pardee, K.; Slomovic, S.; Nguyen, P.Q.; Lee, J.W.; Donghia, N.; Burrill, D.; Ferrante, T.; McSorley, F.R.; Furuta, Y.; Vernet, A.; et al. Portable, on-demand biomolecular manufacturing. *Cell* **2016**, *167*, 248–259. [\[CrossRef\]](#)

161. National Academies of Sciences, Engineering, and Medicine; Division on Earth and Life Studies; Board on Chemical Sciences and Technology. *Continuous Manufacturing for the Modernization of Pharmaceutical Production: Proceedings of a Workshop*; National Academies Press: Washington, DC, USA, 2019.
162. FDA. Q13 Continuous Manufacturing of Drug Substances and Drug Products. Available online: <https://www.fda.gov/media/165775/download> (accessed on 19 July 2023).
163. Bielser, J.M.; Wolf, M.; Souquet, J.; Broly, H.; Morbidelli, M. Perfusion mammalian cell culture for recombinant protein manufacturing—A critical review. *Biotechnol. Adv.* **2018**, *36*, 1328–1340. [[CrossRef](#)]
164. Kim, S.; Jeong, H.; Kim, E.-Y.; Kim, J.F.; Lee, S.Y.; Yoon, S.H. Genomic and transcriptomic landscape of *Escherichia coli* BL21(DE3). *Nucleic Acids Res.* **2017**, *45*, 5285–5293. [[CrossRef](#)] [[PubMed](#)]
165. Packiam, K.A.R.; Ramanan, R.N.; Ooi, C.W.; Krishnaswamy, L.; Tey, B.T. Stepwise optimization of recombinant protein production in *Escherichia coli* utilizing computational and experimental approaches. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 3253–3266. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.