





"G. d'Annunzio" Università degli studiChieti-Pescara

School of Advanced Studies "G. d'Annunzio"

Dompé farmaceutici S.p.A.

Doctorate of Research in:

BIOMOLECULAR AND PHARMACEUTICAL SCIENCES

CYCLE XXXVIII

"From Neurotrophin Gene Design to Clinical Material Generation"

Department of Pharmacy
Disciplinary Area CHEM/07 A

Ph.D. Student
Dr. ssa Elisa Spaccapaniccia

Coordinator
Prof. Assunta Pandolfi

Tutor
Prof.ssa Ivana Cacciatore

Tutor Dompé

Dr. Franck Martin

Dr. Franca Cattani

Academic years 2022/2025

Contents

1. Introduction 3

2. Neurotrophins 7

2.1. Neurotrophins structures 9

2.2. Neurotrophin Signalling 12

3. Brain Derived Neurotrophic Factor (BDNF) 17

3.1. Gene Structure 19

3.2. Protein Synthesis 19

3.3. Protein Function 22

3.4. Sensory Neurons 26

4. Recombinant Protein Production 28

5. Gene Design 33

5.1. Expression Vector Elements 35

5.2. HOST: *Escherichia coli* 42

5.3. Commercial Host Strain 46

6. Periplasmic Expression 49

6.1. Signal Peptides 53

6.2. Protein Extraction from Periplasm 57

7. Shuffle Expression 61

7.1. Disulphide bridges: Structure and function 62

7.2. Engineering *E. coli* for S-S bonds formation 65

8. Biotech Production Process 69

8.1. Upstream 72

8.2. Downstream 77

9. Materials and Methods 82

9.1. PelB-rhBDNF Gene Design 82

9.2. Shuffle rhProBDNF Gene Design 83
9.3. Transformation and Clone Selection 83
9.4. Upstream: Flask Fermentation Process 86
9.5. Upstream: 1L Fermentation Process 86
9.6. Biomass Harvest 87
9.7. PelB-BDNF extraction from periplasmic compartment. 87
9.8. Benzonase Treatment 88
9.9. rhProBDNF Extraction from Shuffle strain cytoplasm 88
9.10. Turrax treatment 88
9.11. HCIC column 89
9.12. Enzymatic Cleavage 89
9.13. CEX column 90
9.14. HIC 90
9.15. Tangential Flow Filtration 90
10. Results and Discussion 91
10.1. Periplasm gene design, transformation, clone selection and Research Cell Banck (RCB) production 91
10.2. Shuffle gene design, transformation, Clone selection and Research Cell Banck (RCB) production 92
10.3. Fermentation 93
10.4. Biomass Harvest 103
10.5. PelB-BDNF extraction from periplasmic compartment 103
10.6. Benzonase Treatment 108
10.7. rhProBDNF Extraction from Shuffle strain cytoplasm 109
10.8. Turrax treatment 110
10.9. HCIC column 112
10.10. Enzymatic Cleavage and CEX chromatography 113
10.11. HIC column 114

10.12. TFF 115

11. Conclusions 117

12. Perspectives 125

13. Abbreviations 127

14. Bibliography 132

1. Introduction

This PhD program is from "G. d'Annunzio" University of Chieti and performed at Dompé farmaceutici S.p.A. and has been completed in the Dompé R&D Biotech Process Development laboratory of L'Aquila. The team is more precisely involved in developing new production and purification processes, for human therapeutical recombinant proteins of potential pharmaceutical interest, to propose new therapies for rare diseases.

The proposed project, focusing on human neurotrophins, spans from human Brain Derived Neurotrophic Factor (hBDNF) gene design to clinical material generation. In particular, the object of this research starts with the design of a synthetic gene, codifying for the protein of interest and its production in a genetically modified microorganism, down to the setting up of its purification process.

Recombinant proteins production, rather than extraction, permits to obtain high proteins quantities, with good quality attributes to be used preclinically, or clinically, fulfilling all the regulatory requests. So, sufficient quantities of human neurotrophin are only obtainable through Biotech Processes, which use recombinant DNA techniques, and are in general always divided in two phases, the fermentation (upstream) and the purification (downstream).

The Up-Stream phase corresponds to material generation, and more precisely, in our case, to the recombinant protein production in microorganisms and the Down-Stream one to its purification, down to homogeneity and formulation.

Currently, in literature Neurotrophins productions are performed in microorganisms through their accumulation in an insoluble fraction, present in the bacteria cytoplasm, also called inclusion bodies, which then needs to be extracted, refolded and purified to obtain the final soluble and active product.

The subject of this PhD was to explore strategies to obtain the protein in a soluble and active form, directly in the microorganism and try to set-up a straightforward process.

Briefly, the project starts from the selection of an expression system, composed of the most suitable vector and host, then proceed with the setting up of the fermentation process conditions, which should be tuned in function of the protein properties to be expressed. In a second time, the downstream protocol will be established, according to the protein

"biochemistry", and this to obtain a correct final process yield and a suitable protein purity. After those steps, formulation studies are necessaries to have an administrable and stable drug to assess it first in animal model, and then if successful, in clinical studies. For the upstream process the activities developed along these three years encompassed: gene engineering, fermentation set up, biomass harvest and protein extraction. Instead for the Downstream parts activities included isolation and purification of the protein from the Host Cell Protein, endotoxin, DNA and product-related forms, choosing different chromatographic resins to efficiently separate the recombinant neurotrophin from the host cell ones, in a minimum of passages. At the end of the process, rhBDNF is formulated and tested for its biochemical properties.

To do so, two new purification processes have been designed and implemented, from scratch, to exclude inclusion bodies formation, which are due to the complex Neurotrophin three-dimensional structure, requiring the formation of three disulphide bridges, that cannot form into the reducing *Escherichia coli* cytoplasm.

To have a chance to obtain a soluble neurotrophin, it was therefore important to test the possibility to express it in an oxidative cellular compartment, in which disulphide bridges can form, and where the protein can acquire its native, soluble and active conformation.

To get round of this *E. coli* limitation, two strains have been used: one is the BL21 strain and the other one is the Shuffle strain. In the first strategy, the mature recombinant protein is express and transferred, using a leader sequence, into the oxidative periplasm, where disulphide bridges can form and consequently protein has a chance to adopt its native folded and soluble form. This mature protein, theoretically, is almost ready to be purified. The other choose strain was the Shuffle one, and was selected for direct cytoplasmatic expression, as this *E. coli* strain has been engineered to be able to oxidate cysteines and form disulphide bridges. Furthermore, to obtain a good neurotrophin yield the BDNF gene has been cloned, together with its chaperone, which will help precursor refolding.

To start with Upstream experiments, the two different expression vectors have been transformed, and two distinct Research Cell Banks (RCB) produced. Fermentation process has been settled in a battery of eight Minibio fermenters (from Applikon) of 1 litre scale. Fermentation protocols have been settled, dividing the process into two phases: the first one (batch) corresponding to biomass growth, whereas the second one (fed-batch) to the recombinant protein production. This approach was used to always

control dissolved oxygen in the medium, as in its absence, no recombinant protein will be expressed. The biomass growth phase ends, when all glucose present in the medium has been metabolized, at this point the fed batch phase starts, with the controlled addition of new substrate in the fermenter vessel. In-between the two phases, temperature is adjusted, and after one hour the recombinant protein expression is trigged by addition of an inducer, activating the recombinant gene transcription.

During the different runs, several parameters have been optimized (medium and feed compositions, temperature, pH, agitation and aeration, different inducer concentrations and induction time), until the critical control parameters have been defined. Furthermore, those numerous fermentations parameters have been collected and analysed by Artificial Intelligence, to predict final biomass productions. Results have been published in an article with title "A deep learning approach to optimize recombinant protein production in *Escherichia coli* fermentations" [1]. Proceeding with the upstream process development biomass harvest and cells disruption has been investigated, evaluating the presence of different contaminants like DNA, Host Cell Proteins (HCP) and Endotoxins. In both strategies rhBDNF resulted in part soluble and therefore refolded, so two different Downstream processes have been developed to purify the protein of interest.

In the first strategy, a massive DNA presence has been observed in the homogenized material, representing a major obstacle for chromatography, so this strategy has been interrupted, and the project was continued with the more promising Shuffle strategy, where recombinant protein was better expressed. In this case, due to the extractive method, soluble and refolded rhProBDNF was also heavily contaminated by DNA, but following the implementation of a DNA reduction step, it has been possible to proceed with the downstream purification, and a Hydrophobic Charge Induction Chromatography (HCIC) has been implemented as capture column. Once purified on the first column, to obtain the mature BDNF form, remotion of the now useless chaperon was needed, and enzymatic protocol has been developed to optimize the digestion conditions. Several digestion by-products were generated together with the mature BDNF, so two more chromatographic steps have been implemented, to further purify the recombinant mature protein.

Finally, to both concentrate and formulate the protein, tangential flow filtration (TFF), was used. Concentrated and formulated material has been utilized to further understand

the recombinant protein structure and its biological activity. A complete downstream protocol has been developed for the Shuffle system.

So, to summarize, the first strategy with protein production in the *E. coli* periplasm was the most challenging one, as no refolding and enzymatic cleavage were necessaries. Furthermore, in theory this process would have necessitated less chromatographic columns to purify the protein, as compared to the second strategy. A very straightforward protocol, with an easy scaling-up perspective. Unfortunately, process yields were too low to be interesting for industrialization, but it was however important to explore this type of expression strategy.

The second alternative with the elimination of the *in vitro* refolding step, was also consisting in a simpler protocol, than the current process. The initial higher protein yield, than with periplasmic expression, was promising, and propaedeutic to guarantee a more reliable process with a better scalability. Finally, deep analysis of drug quantity, quality and purity revealed that the obtained neurotrophin was unfortunately not complying with regulatory requests in terms of purity and biological activity.

Gene design, protein engineering and alternative processes implementations are always long, an interesting, journeys that have increased our knowledge in recombinant protein productions in *Escherichia coli* and more precisely in rhBDNF properties.

2. Neurotrophins

Neurotrophins, are a family of homologous polypeptide, sharing 50% of amino acid sequence homology [2], similar molecular weight (13.2-15.9 KDa) and isoelectric point (9-10). [3] Those proteins are important regulators of neuronal survival, development, function and plasticity. Moreover, they possess a neuroprotective role after an injury. As suggested by their names, Neurotrophins are secreted proteins, essential for nervous system trophism. They are growth factor, structurally conserved, which have been identified during embryo development, comparing the early stage of nervous system formation, and characterized as playing a crucial role in the development, maintenance and neurone protection.

This family of protein includes Nerve Growth Factor (NGF), Brain-Derived Neurotrophic Factor (BDNF), Neurotrophin-3 (NT-3), and Neurotrophin-4/5 (NT-4/5), each necessary for the survival of specific, but overlapping, neurons populations during development.

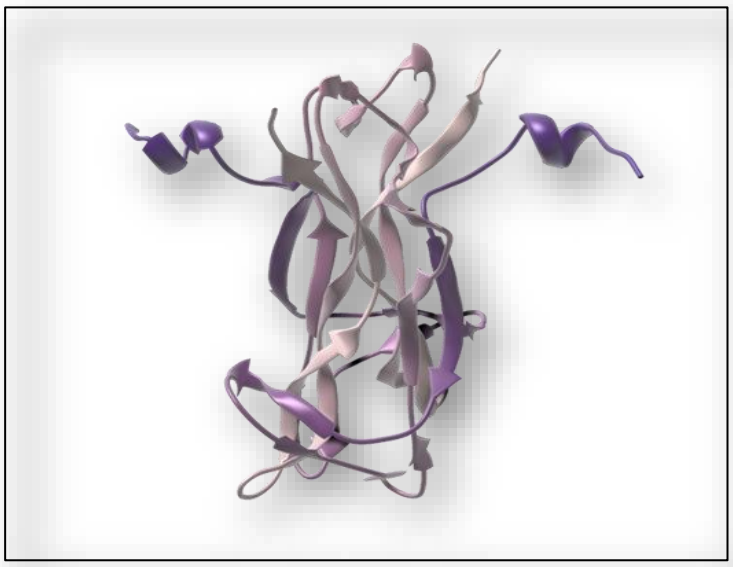


Figure 1-Structure of rhNGF from Protein Data Base. The image is created and used for internal industrial purpose.

Neurotrophins, share structural and functional characteristics, are all derived from an ancestral common gene. Nerve Growth Factor (NGF) has been the first to be discovered

and studied and represent the prototypic member of the neurotrophin family. It was discovered in the 1940s in the male mouse submaxillary glands. [4]

In particular, the first evidence of this neurotrophin prototype, has to be attributed to Rita Levi-Montalcini and Viktor Hamburger, in the 1940s and 1950s, who demonstrated that NGF is a target-derived neuronal survival factor. It has been defined as the "mother" of all the growth factor and becomes an historical achievement [5]. NGF was purified as a factor able to support survival of sympathetic and sensory spinal neurons in culture (R.L. Montalcini, 1987). Subsequent researches led to the purification and characterization of BDNF, NT-3, and NT-4/5, which have similar neurotrophic and neurotropic actions. BDNF is the second on the discovery road map and was extracted (in 1982) from pig brain [6]. Successively, neurotrophins role has been better characterized. Although the number of neurotrophins is small, their actions are enormously subtle and complex, in fact they contribute to established complex pattern of connectivity of the nervous system. The number of neurons in the adult nervous system and the density of innervation are determined by a competitive process, in which a tissue produces growth factors, that support neuronal survival and programmed cell death, which is a prominent feature to limit some proliferative phenomena.

As all polypeptide hormones, neurotrophic factors deliver their messages interacting with cell surface receptors. Neurotrophins exert their biological effect through selective interactions with their receptors and induce dimerization of a family of high affinity tyrosine kinase receptors. Analysis of residues, with high surface accessibility, at the surface of the mouse NGF homodimer show a conservation of this structure between the different member of the neurotrophin family. However, it is impossible to form stable heterodimers between the different members of this family, as they are asymmetric and the N-terminal portion of each member differs from each other. The N-terminal portion is the most critical part for receptor binding and biological activity [7].

The three-dimensional structures of NGF determined by crystallography reveals a tertiary fold dominated by two central pairs of anti-parallel β-strands, that define the elongated shape of molecule. The anti-parallel β-strands are connected by hairpin loops, in which the major variability, among the neurotrophin is registered and explain their receptors specificities [8].

The neuronal population regulated by neurotrophins is growing and includes sensory, sympathetic neurons, hippocampal pyramidal and dentate neurons, motoneurons, retinal ganglion cells, septal and striatal cholinergic neurons, cerebellar Purkinje cells, and neurons of substantia nigra and locus coeruleus.

It has been shown that they are involved in the regulation and prevention of cholinergic neurons degeneration, and in maintaining functional aspects of these neurons. Moreover, they are involved in cellular differentiation, axon guidance, synaptogenesis, synaptic plasticity, modulation of neurotransmitter release, neurite branching and elongation. These countless biological actions, reflect the heterogeneity of the neurological e non neurological tissues able to produce them.

Indeed, in addition to neurons also fibroblasts, myelomonocytic cells, lymphocytes, keratinocytes, and mast cells also produce neurotrophins. Furthermore, neurotrophins are also involved in autoimmunity, allergies, and neurodegenerative diseases, as well as inflammation.

2.1. Neurotrophins structures

All four mammalian neurotrophins (NGF, BDNF, NT-3, and NT-4/5), are encoded by distinct genes and initially synthesized as pre-pro-precursors, called Pre-Pro-Neurotrophins, which are processed by proteolytic cleavages to yield their mature forms.

The Pre-Pro-protein contained a small signal peptide for its exportation to the Golgi, a pro-domain encoding the dedicated neurotrophin chaperone and a mature domain. Successively, to its synthesis the Pro-protein is transported to the Golgi apparatus and the Pro-domain and mature protein interact with high affinity to ease its refolding (chaperone effect). After that, proteins undergo endoproteolytic cleavage, which promotes its targeting to secretory granules, called Dense Core Vesicles (DCVs) [7]. The proteolytic cleavage is predominantly performed by intracellular proteases such as furin [9], and pro-hormone convertases (PC1, PC2 and PC7), but can also occur, once they are secreted, by extracellular proteases, such as plasmin and matrix-metalloproteinases.

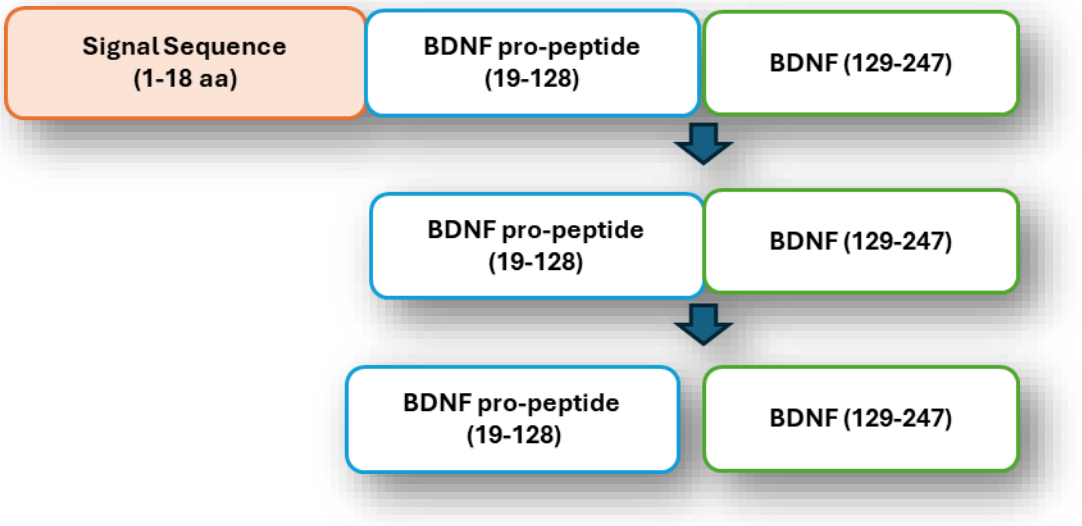


Figure 2-Structure of the entire Pre-Pro-Neurotrophin, and the two forms generated after the two enzymatic cleavages. Ref: H. K. T. M. T. H. Y. I. a. M. K. Koichi Uegaki, "BDNF Binds Its Pro-Peptide with High Affinity and the Common Val66Met Polymorphism Attenuates the Interaction," International Journal of Molecular Sciences, vol. 18, p. 1042, 2017. [29]

Proneurotrophins are composed of an N-terminal prodomain and a C-terminal mature domain and are associated in noncovalent homodimers, via interactions of their mature domains. The length of the entire sequence is around 250 amino acids.

The Propeptide function has been investigated, and its chaperone role demonstrated, as it significantly increases the refolding yield rate with the concomitant disulphide bonds formation, as compared to the mature protein *in vitro* refolding [10].

In mature NGF disulphide bonds formation is facilitated by the covalently attached and specific pro-sequence. This has been elegantly demonstrated by Rattenholl et all. [11]. The refolding yield of the native protein is dramatically enhanced by the presence of the pro-peptide and increases from 0.1-0.4% to 35%, in its presence. Also, the folding kinetics with the pro-peptide is accelerated, compared to mature protein alone.

Moreover, it has been demonstrated that, among the different neurotrophins precursors, only the NT-3 one can acts as chaperon on NGF, highlighting that, even if closely related, this effect is protein specific [2]. All neurotrophins are generated as precursors and then cleaved by proteases to release the mature protein. As neurotrophin prototype, NGF, is secreted as a homodimer precursor and proteolytically cleaved by furin to yield mature NGF. Evidence shows that NGF and ProNGF have distinct functions, in fact the precursor

is more present in the CNS tissue, whereas the mature is barely detectable. And each molecule is able to contact the TrkA receptor, in two distinct sites. The NGF prodomain seems to be able autonomously to induce growth cone collapse *in vitro*, indicating that a well-defined secondary structure contacts the p75 receptor.

Mature neurotrophins are noncovalent homodimers, and each subunit is formed of approximately 118–120 amino acids, and possess a distinctive three-dimensional structure characterized by a cysteine knot, formed by three disulfide bonds. This motif is highly conserved, among all neurotrophins, has a crucial structural role and is essential for dimer stability and receptor binding. Resolution of NGF crystal structure provides information about all neurotrophins and on neurotrophin-receptor interactions.

NGF structure is characterized by three pairs of antiparallel beta strands, arranged along protein axis, linked by hairpin loops at the top and bottom, forming a flat hydrophobic interface, between the two homodimer subunits [12]. Along the axis of symmetry, the amino terminal and the carboxy terminal ends of both chains lie on the same side in the solvent and as they are more flexible, than the rest of the molecule have been poorly solved in the crystal structure. Among all the residues the six cysteines, involved in the disulfide bridge formation, are perfectly conserved suggesting the same shared 3D structure between BDNF and NT-3. If the folded structure is conserved among the neurotrophins, Trk receptors selectivity resides in other regions, and also the binding site of each specific neurotrophin and p75 seems to be overlapping.

The structure of NGF is well characterized and among the domains, the I, II, IV and V are the regions responsible for stabilizing ligand-receptor binding and region II seems to be responsible of receptor activation [8].

The hydrophobic core is involved in the molecule dimerization, exposing the more hydrophilic portion on the external part of the protein.

2.2. Neurotrophin Signalling

In order to exert their biological effects, Neurotrophins bind, in homodimers form, two different cell surface receptors, belonging to different families: one is tropomyosin receptor kinase (Trk) type and the other one the tumour necrosis factor receptor superfamily (TNFR).

The neuronal high-affinity receptors family is composed of TrkA, TrkB and TrkC. Each neurotrophin recognize a specific and unique neurotrophin receptor. Trk receptors display precise interactions with the four NTs family members. NGF binds and activates TrkA, whereas BDNF and NT-4 interact with TrkB and NT-3 with TrkC. NT-3 binds also TrkA and TrkB but with low affinities [4] . The specificity of Trk receptors to mediate biological actions is conferred by specific residues in the NTs and the Trk molecular structures, that lie within the membrane-proximal immunoglobulin-like domain (Fig. 3) (Chao, 2003, Urfer et al., 1998, Wiesmann and de Vos, 2001). Comparison of the TrkA, TrkB and TrkC sequences reveals 66-68% amino acid identity and both in the catalytic and the extracellular domains.

Following NTs binding, the Trk receptors dimerize and are auto phosphorylated on intracellular tyrosine residues, which results in rapid activation of multiple downstream signalling cascades, including the extracellular-regulated kinase, phosphoinositide-3 kinase (PI3K)/Akt (protein kinase B a.k.a. PKB), and phospholipase C γ (PLC γ) pathways. These vital intracellular signalling cascades then activate cell-specific transcription factors, involved in the regulation of host cellular functions. In particular it is known to be critical to neuronal development, including cell survival, axon and dendrite growth, membrane trafficking, glial differentiation and interactions with adjacent neurons (Reichardt, 2006).

Trk receptors exist in multiple isoforms, called truncated, that lack the intracellular tyrosine kinase domain, and can act as dominant-negative modulators by sequestering neurotrophins, rendering less disposable, or forming inactive heterodimers with full-length receptors. In other cases these truncated forms, when activated by neurotrophin ligand, have independent signalling functions, and regulate calcium release in the astrocytes, or modulate synaptic plasticity [13].

On the other hand, the p75 neurotrophin receptor (p75NTR) was first identified in 1973, as the low-affinity nerve growth factor receptor (LNGFR) and all mature neurotrophins,

as well as pro-neurotrophins, bind to it. It is also called, for this reason, the universal or pan-neurotrophic receptor, belonging to the TNF/Fas/CD40 Superfamily [4]. In particular, a double function is modulated by 75NTR, because two pathways are promoted, neuronal survival as well as neuronal death. The p75NTR receptor lacks intrinsic kinase activity, but contains a death domain, in its intracellular region.

When co-expressed, p75NTR modulates Trk receptor ligand specificity and affinity, enhancing up to 100-fold, the selectivity and increasing the binding affinity of TrkA for NGF and TrkB for BDNF. In some other case, p75NTR can restrict TrkA signalling to NGF and TrkB signalling to BDNF, while in its absence, NT-3 can compensate for reduced NGF concentration by activating TrkA [14].

p75NTR also mediates distinct signalling outcomes, including apoptosis, particularly when Trk receptor expression is low or absent, and can activate pathways, such as nuclear factor-kappa B (NF-κB) and Jun kinase. p75NTR is responsible for transducing the signals from BDNF, NGF, NT-3, and NT-4/5 to regulate a broad array of biological activities, such as apoptosis, survival, neurite outgrowth, migration, and cell cycle arrest (Blochl & Blochl, 2007).

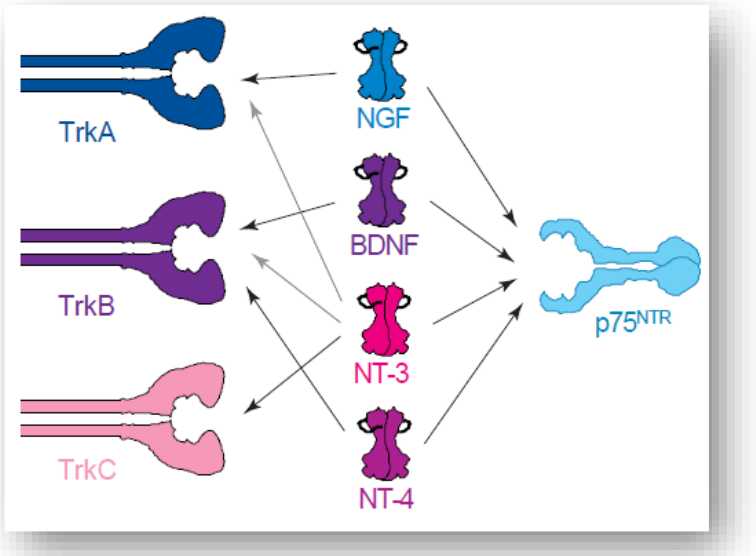


Figure 3- Specificity and crosstalk in neurotrophic factors and receptors. One neurotrophin can bind several receptors and Trk receptors can also bind more than one neurotrophin. All neurotrophins bind the "pan-receptors" p75. **Ref:** C.F: Ibànez, Emerging themes in structural biology of neurotrophic factors, review. TINS Elsevier Volume 21, N°10 1998 **[84]**.

Although most studies reveal that p75 is incapable to activate signalling in absence of Trk receptors, few studies reported an intrinsic activity of the p75 receptor. The death receptor has a cytoplasmic domain able to form an amphiphilic alpha helix. similar to G-protein binding site, present in other receptors.

When a Neurotrophin binds to its tropomyosin-related kinase receptor (Trk), it promotes its dimerization and conserved tyrosine residues autophosphorylation, within the intracellular receptor domain: the final effect is the formation of a docking site for adaptor proteins, containing Src homology 2 (SH2) or creating phosphotyrosine-binding domains, such as Shc, fibroblast growth factor receptor substrate 2 (Frs-2), rAPS, SH2-B, and phospholipase Cγ (PLCγ). These adaptor proteins activate three principal intracellular signalling cascades: the Ras/Raf/MEK/extracellular signal-regulated kinase pathway (mitogen-activated protein kinase, MAPK), the phosphatidylinositol 3-kinase (PI3K)/Akt pathway, and the PLCγ pathway.

The RAS-MAPK cascade leads to Ras activation, which then stimulates Raf, MEK, and ERK kinases. Activated ERK translocates to the nucleus and phosphorylates transcription factors, as cAMP response element binding protein (CREB), promoting gene expression involved in neuronal differentiation, survival, and neurite outgrowth. The second signalling transduction pathway (PI3K/Akt), results in Akt phosphorylation, which inhibits pro-apoptotic proteins, such as BAD and fork head transcription factors, and promotes cell survival, growth, and differentiation.

The third mechanism is the PLCγ pathway, which is characterized by phosphorylation of specific TrkB tyrosine residue (Y785), which recruited phospholipase Cγ (PLCγ), able to hydrolyses phosphatidylinositol (4,5) bisphosphate to generate the second messengers: inositol-1,4,5-triphosphate (PIP3) and diacylglycerol (DAG). These two molecules can acts in two different ways, one starting from IP3 triggering the intracellular calcium release, and the DAG activating protein kinase C (PKC), but both regulating neuronal activity, synaptic plasticity, and genes expression [15].

Signals activated via the p75 neurotrophin receptor (p75NTR), are also distinct, complementary or opposite, from the previous described ones. Death receptor can activate NF-κB, c-Jun N-terminal kinase (JNK), and ceramide production. In particular, NF-κB activation promotes cell survival and dendritic growth, while JNK activation leads to apoptosis, through induction of caspase proteins, a pro-apoptotic factors, such as Bim,

Bid, Bax, and the genome guardian, like p53. Ceramide production, results of the activation of the sphingomyelin hydrolysis, correlated to the cell death and survival, depending on cellular context.

Crosstalk between Trk and p75NTR signalling determines cellular outcomes. Neurotrophin signals are also conveyed via retrograde transport of signalling endosomes. The retrograde signal is associated with transport along the axon to nucleus, of both growth factor and receptors, in response to neurotrophin binding. Upon neurotrophin binding, Trk receptors are internalized into endosomes, which are transported retrogradely, along axons by dynein motors to the cell body, where they continue to signal and regulate gene expression, necessary for neuronal survival and differentiation [16].

During neural development, Neurotrophins are involved in some activities, along the mature nervous system, acting on plasticity and cognitive functions. During embryo development, they regulate neuronal tissue and nervous system associated ones, like glia. The survival and differentiation of neurons, the axonal and dendritic growth, synaptogenesis, and synaptic pruning, with nearly half of neurons generated, undergoing programmed cell death, highlighting the determining role in terms of neuronal population size.

In the mature nervous system, Neurotrophins guarantee the maintenance of neuronal phenotype, modulate synaptic function, and support processes such as learning, memory, neurogenesis and response to injury. For example, an activity includes in synaptic plasticity, long-term potentiation, is critically regulated by BDNF, which is central to learning and memory. BDNF facilitates both early-phase and late-phase of LTP in the hippocampus, acting at pre- and post-synaptic levels. It also regulates the transport and translation of mRNAs in the synapses.

The complex scenario plays by Neurotrophins structure and function, lead us to think that they can play a central role in neurodegenerative pathologies. Neurotrophins are growth factors essential for the differentiation, growth, and dynamic development of neuronal cells. Their absence or deficit in signalling pathways, result in age-related neurodegenerative disorders, such as Alzheimer’s, frontotemporal dementia, Huntington’s, and Parkinson’s diseases.

Neurodegeneration damages, emerge when neurons are unable to respond to oxidative stress, resulting in accumulation of proteins and lipids, and DNA damages. In some of ND diseases, the most important limiting issue is deficit of neurotrophic factors availability, or the loss of their delivery to the brain.

The development of neurotrophins as therapeutic agents has been, up to now, hampered by their recombinant productions and delivery through the blood–brain barrier. All previous clinical trials for peripheral administration, and infusion into the cerebrospinal fluid space, failed to target neurons and often caused adverse effects, such as pain, sympathetic axon sprouting, and weight loss.

Neurotrophin-based therapies offer promises for neuroprotection and neural regeneration, but future research must address challenges such as crossing the blood-brain barrier, site specific delivery methods, and side effects reduction.

3. Brain Derived Neurotrophic Factor (BDNF)

In 1982, Yves-Alain Barde and Hans Thoenen purified a factor from pig brain, that supported survival and growth of embryonic chick sensory neurons, which has been later identified as BDNF. Successively, this protein has been described as "a factor in glioma-conditioned medium, which support the survival of embryonic sensory neurons and can be distinguished from NGF by immunological and functional criteria" [18].

Today BDNF expression is documented in both Central Nervous System (CNS), which represents the major and active neurotrophin synthetized, and in Peripheric Nervous System (PNS), where it is associated with CNS pathologies. It is known as the "neuroprotectant" endogenous factor, for its protective role, after an injury.

Like other neurotrophins, BDNF localization and its expression, varies along the embryo development, childhood and adulthood. It has been reported that, during the 11^{th}-12^{th} days of rat and mouse brain embryogenesis, there is an increment of BDNF expression, that coincide with the onset of neurogenesis and increase of neuronal activity. In the newborn BDNF is the most prominent factor expressed in CNS, but low levels are registered in the peripheric district. In the adulthood high levels of message RNA (mRNA) of BDNF are registered and becomes the most predominant Neurotrophins expressed, in the post-natal brain. The highest diffusion is especially registered in the neocortex, amygdala, hypothalamus and hippocampus. In this area, this upregulation is correlated to the increment of neuronal activity [17].

The protein secretion is spatially and /or temporally regulated and tightly tuned in these neuronal districts. BDNF plays an important role in maintaining the structural integrity and function of neurons, influencing their growth, survival and differentiation during the embryotic development. During embryogenesis, BDNF is highly expressed and TrkB signaling and promotes the differentiation of cortical progenitor cells, and later their differentiation into neurons. Whereas, during development and childhood, a lower presence of BDNF is registered.

Furthermore, in the mature brain and nervous system, and particularly in the adult neurons of hippocampus and striatum of the BDNF promotes neuronal survival, cell differentiation, synaptic plasticity and long-term potentiation. BDNF protein and mRNA BDNF has been identified most in the brain areas, including olfactory bulb, cortex,

hippocampus, basal forebrain, mesencephala's, hypothalamus, brainstem and spinal cord [18].

BDNF is the most distributed and studied neurotrophin in the mammalian brain, because it plays a role in growth, development and plasticity of the glutamatergic and GABAergic synapse. Moreover, through the modulation of the neuronal differentiation, it is involved in the serotoninergic and dopaminergic neurotransmission [19].

Among the cells secreting BDNF, neuronal cells are the most described both neurons and oligodendrocytes. In details, BDNF is expressed by glutamatergic neurons, glial cells, such as astrocytes, isolated from the cortex and hippocampus, but not from the striatum and microglia [20].

In clinical pathologies, BDNF dysregulation is registered in different diseases, such as neuro-degenerative ones (Alzheimer's, Parkinson, Multiple Sclerosis, and Huntington's diseases), obesity, anxiety and depression [21]. Modulating BDNF signalling may present a valid approach to treat neurological and psychiatric disorder. In particular in Alzheimer (AD) e Parkinson (PD) is registered a reduction of BDNF, signalling a neuroprotective role of the molecule [22].

But several reports indicate that other type of cells like platelets, vascular smooth muscle cells, cardiomyocytes, endothelial cells, lymphocytes, macrophages and myocytes can secrete it. This protein is so also involved in other pathologies like cancer, angiogenesis, reduction of liver glucose production and reduced glucose up-take from peripheric tissue. BDNF deficiency is associated with weight increased in rat and human, and its administration can reduce food intake and increases energy expenditure.

Recently, BDNF has been found to play a central role in chronic pain conditions, another interesting clinical pathological situation, in which a complex mechanism involve more than one neurotrophin. In this pathology, modulations effects are regulated by the different convolved neurotrophins.

3.1. Gene Structure

In humans BDNF gene is located on chromosome 11 (11p14.1) in a region of 4 Mb. Gene is conserved among different species and surprising possess a complex structure composed of 11 exons in human, 9 in the rat, and 9 different promoters, that explain the different brain regions cell types and spatial and temporal BDNF expression.

The BDNF gene is highly expressed in the CNS. Despite the complexity, ProBDNF gene is only encoded in the IX exon, while the others correspond to the untranslated regions, that regulates the different spatial and temporal responses to various stimuli. Exon IX is present in all human and rat BDNF mRNA. The nine alternative promoters suggest that, they can regulate spatio-temporal BDNF expression and can respond to a great variety of stimuli. The BDNF gene can produce about 34 different transcripts, which are then polyadenylated at two alternative sites, leading to a further population of two different mRNA: one with a short untranslated region at 3' present in the cell soma and the other with a long UTR at 3' present in the dendrites. [22]

For example, promotors I e IV are activated by the glutamate NMDA type receptors in neurons and after depolarization of the neuron's membrane, due to the activation of the L-type Calcium dependent gated channel (L-VGCC). [16]

3.2. Protein Synthesis

The mature monomer protein molecular weight is 12.3 kDa and the factor has some resemblance in terms of molecular weight and pI to Nerve Growth Factor, but "it is clearly a different identities" [6], as it only shares only 55% amino acid sequence similarity and do not respond to Anti-NGF antibodies. Also in its role, as it mostly interacts with different tissues than NGF, such as neurons of central nervous system and retinal neurons.

As indicated in the previous paragraph, BDNF synthesis and maturation is a multistep process, involving several passages. The BDNF protein is highly conserved between rat, pig and human, and the protein has 247 amino acids, it is synthesized and folded in the endoplasmic reticulum as PreProBDNF (32–35 kDa). Upon translocation into the Golgi apparatus, the signal sequence is rapidly cleaved, and the proBDNF isoform (28–32 kDa) is generated. In the presence of the lipid raft associated sorting receptor carboxy peptidase E (CPE), proBDNF is stored in vesicles and transported into activity dependent secretion by post synaptic dendrites [23]. A further cleavage of proBDNF to obtain the mature

isoform (mBDNF, 13 kDa), can be achieved either intracellularly or extracellularly. The intracellular proteolytic proBDNF cleavage can be performed by the subtilisin-kexin family of endoproteases, such as furin, or by convertases, as the plasmin serine protease and matrix metalloproteases 2 and 9 (MMP2 and MMP9), in intracellular vesicles. After those processing, according to species, mature protein contains at latest 119 amino acids, which represent the functional protein form.

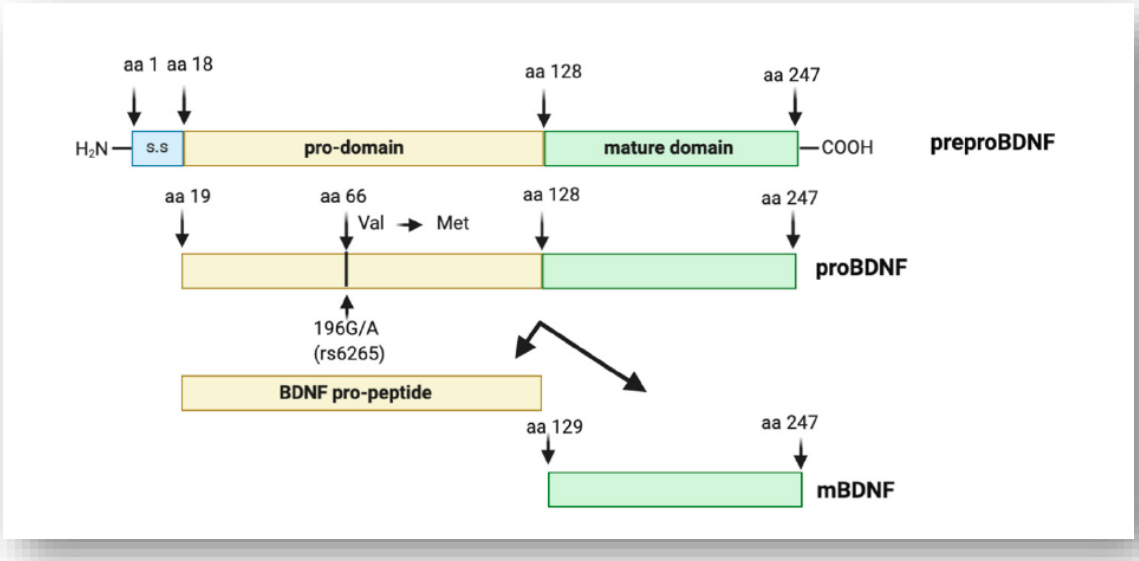


Figure 4 Brain derived neurotrophic factor (BDNF) protein structure. The PreProBDNF consists of three subdomains: signal sequence (s.s), pro-domain, and mature domain. The intra- or extracellular cleavages of PreProBDNF generates functionally active isoforms: BDNF pro-peptide and mature BDNF (mBDNF). Arrowheads indicate the cleavage sites involved in the processing of mature BDNF. Reference: L. C. D. e. all., Neurotrophic Factor BDNF, Physiological Functions and Therapeutic Potential in Depression, Neurodegeneration and Brain Cancer, MDPI, International Journal of Molecular Sciences (IJMS), October 202021(20):7777, DOI:10.3390/ijms21207777 [20]

In its structure BDNF has three disulphide bonds, forming a cysteine-knot, the protein is active in a non-covalent linked homodimer form and stimulates tyrosine kinase receptors, which lead to a series of cascades within the cells, particularly those involved in neurogenesis, differentiation, increase ganglion survival and prevents neuronal degeneration [24].

As describe later, the immature and/or mature forms have different binding properties, and consequently multiple biological functions. The balance between precursor and BDNF depends on particular development stages and regions. Depending on cell type, both species can be constitutively or selectively secreted.

Upon synthesis, BDNF is stored in secretory granules, and depending on the cell type is: 1) delivered to dendrites, to modulate the spines architecture, and thus synaptic plasticity [20], 2) transported along the axon in an anterograde manner to stabilize pre-synaptic site and the growth/branching of the axonal terms [25].



Figure 5- BDNF structure from Protein Data Base.

In physiological and pathological conditions, the situation is further complicated by other modulators, that can exert further positive or negative effects on the neurotrophins. In the striatum BDNF is synthetized in the cortical level and released by descending cortico-striatal pathways and it is regulated by NT4/5 and BDNF itself in a glutamate receptor-dependent manner. TrkB, BDNF and NT4/5 activated the BDNF production in a dose dependent manner, suggesting that this could be a mutual modulation [26].

3.3. Protein Function

Depending on the cell type, BDNF can be secreted in a constitutive or activity-dependent manner, contrary to the other growth factors, which are only constitutively synthetized, most of the BDNF is determined by a regulated pathway, that depends on neuronal activity and Ca^{2+} dependent signalling.

BDNF is the synaptic molecule able to modify synaptic shape, strength and can act as modulator of synaptic plasticity [20]. Moreover, the sequence information in the Propeptide, as the sequence information located in the mature BDNF, participate to the targeting and consequently to the different functions. In non-pathological conditions proBDNF is rapidly converted intracellularly into BDNF, which is stored and release after an excitatory input. For example, in an adult rat brain the ProBDNF concentration is 10-fold less the one of BDNF, confirming that the convertase activity is increased by the excitatory signal.

In neuronal cells, both proBDNF and mBDNF are released following cell membrane depolarization. The balance of proBDNF and mBDNF depends on the brain stage development and interested regions. In the early postnatal period, proBDNF concentration is higher, while in adulthood mBDNF prevails for its neuroprotective role after injury and for regulating synaptic plasticity [19].

Initially, only mature protein was thought to have a biological role, but it has been demonstrated that proBDNF precursor, also elicited opposite effects by binding either to p75 or TrkB receptors. The high affinity receptor for BDNF is represented by TrkB (gp145TrkB WT), which is also recognized by NT4/5. It also exists an isoform, called TrkB gp95TrkB (95 KDa), lacking the tyrosine kinase domain and the low affinity receptor of growth factor (p75NTR) [27]. p75 is implicated in both pro and anti-trophic factor, such as neurite growth and apoptosis.

The truncated TrkB and BDNF are widely expressed in the brain. TrkB receptor has an extracellular domain with 5 repeated glycosylated Ig like domains, a unique transmembrane region and an intracellular domain with tyrosine kinase activity. Receptor activation is quick (2 minutes) and its deactivation is longer (30 minutes) [27].

In details, once secreted BDNF and ProBDNF bind two family sets of receptors kinase. BDNF can binds to TrkB, that induces the activation of different small G protein, leading to neuronal plasticity, neurogenesis and stress resistance. When BDNF binds TrkB, it dimerizes it and receptor auto-phosphorylates on tyrosine residues, to form a docking site for Scr-homology 2-domain containing adaptor protein (SHC) and phospholipase C. While SHC docks the receptor, and binds Grb2 by SOS, RAS activates the RAS/Rap-MAPK, PI3K-AKT and PLC g-PKC, cascade. MARK/ERK is essential for the neurogenesis and promotes survival with two ways: induction of pro-survival genes and inhibition of pro-apoptotic BAD and Bax. The other signalling pathways is the Ras Mitogen Activated protein kinase, that stimulates the expression of anti-apoptotic protein such as BCL2 and CREB. The last pathways represented by PLC -gamma, which activates the inositol triphosphate receptor (IP3R), leading to the release of intracellular calcium that triggers the activation of calmodulin kinase (CamK) and protein kinase C (PKC), and also factor CREB transcription. This last event is rapid and has a short-term effect, which favour cell growth and differentiation, while MAPK and PIK3 generate long term transcriptional effects.

The BDNF-TrkB pathway is known to regulate and coordinate numerous cellular functions such as neuronal survival, neurons growth and differentiation of during development, activity-dependent synaptic plasticity, as well as memory processes and learning in mature CNS. TrkB receptor can sometimes be activated by mechanism called transactivation, in absence of BDNF, but its activation is due to G-protein-coupled receptors (GPRs).

On the other hand, Pro-BDNF, as well as other neurotrophins and pro-neurotrophins, bind to p75NTR receptor, a member of tumour necrosis factor receptor superfamily (TNFR). The binding between these ligands and receptors began in the N-terminal cysteine repeat extracellular domain. Despite the trophic role of BDNF-TrkB activation, its association with p75 can have an opposite effect. In fact, p75 can potentiate and/or reduce TrkB receptor activity, otherwise it can trigger apoptotic cascade when pro-BDNF, binds to it through the sortilin adaptor [19] .

The p75C-terminal cytosolic domain, lacks catalytic activity, but promotes the activation of the Jun-Kinase and NF-KB pathways, and the transcription of multiple genes. The p75/sortilin receptor complex leads to various activators of apoptotic signalling pathways,

such as Jun and caspase 3-9, while Ras homolog gene family member A (RhoA) and the NF-KB can also trigger neuronal survival and growth cone development. The 75NTR plays an important role in survival of specific neuronal population but can also mediate neuronal cells death both in normal and pathological conditions [4].

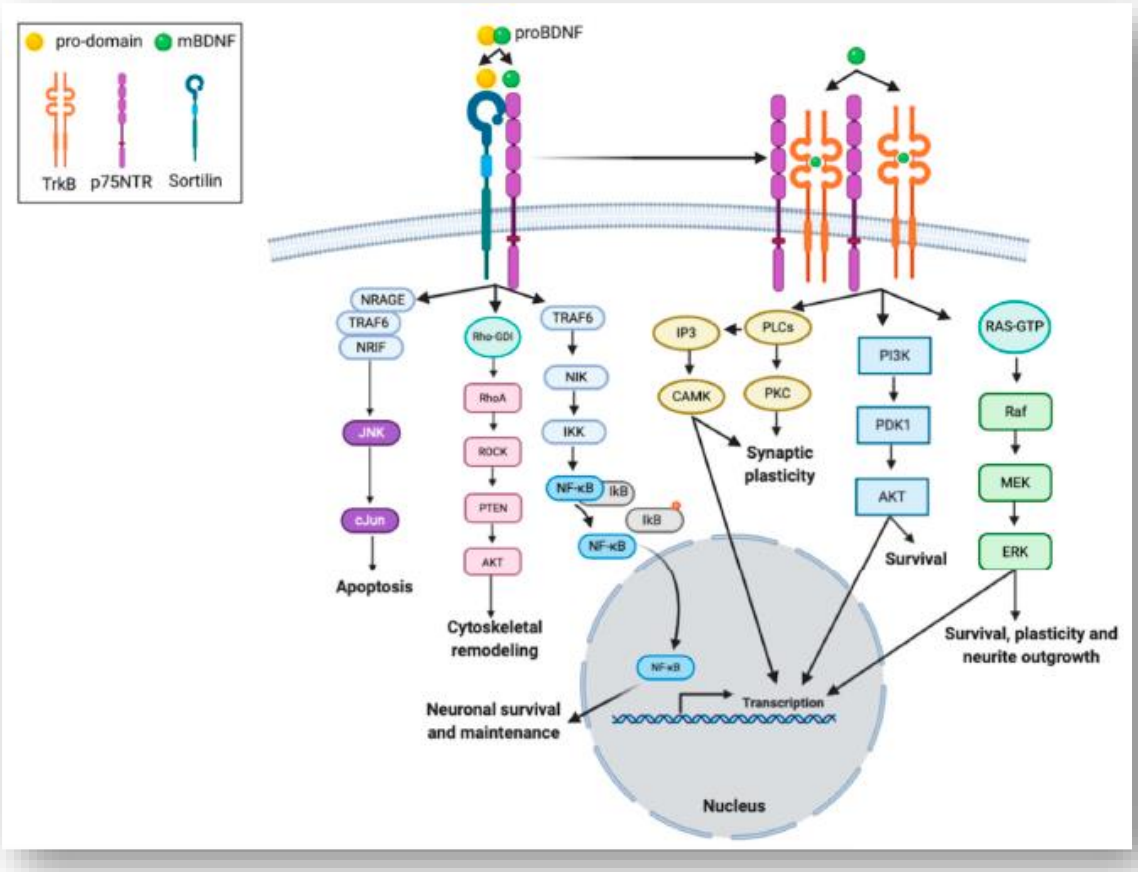


Figure 6- BDNF and ProBDNF Signalling pathways with activation of the two different receptors. Each receptor can promote different activations pathways leading to different final effects. Ref: Neurotrophic Factor BDNF Physiological function and Therapeutic potential in Depression, Neurodegeneration and Brain Cancer. L. Colucci - D'Amato et all. J. Mol. Sci. 2020, 21, 7777 **[20]**

Therefore, ProBDNF and BDNF can differentially activate different receptors and molecular pathway and mediate opposite effects, such as neuronal survival or death. While downregulation of BDNF is linked to brain neurological conditions, pro-BDNF secretion inhibits synapse formation, dendritic arborization and impairs synaptic transmission and plasticity. Different neuronal populations require different types of neurotrophins, which exert long and short-term actions. This indicates temporal and spatial dependent modes of actions of the neurotrophins during development and adulthood.

The p75 ^{NTR} gene is composed of 10 exons and 11 introns. Its domains are extracellular, transmembrane and intracellular. The first one is composed of 4 cysteine repeats, in which cysteines in third and fourth positions are essential for ligands interactions. Intracellular domain acts as death processor, able to cause nerve cells death. But when p75 ^{NTR} is present on Schwann cells, it assures the correct BDNF migration along the peripheral nerves. Moreover, the final effect of BDNF contact with this apoptosis receptor depends on the context, and on the different and numerous pathways, activated in each cell type.

BDNF secretion is controlled by several stimuli like neuronal activity, depolarization, neuropeptides and proinflammatory factors. The calcium-dependent release of BDNF, stored in the dense core vesicles, could be triggered by calcium release from the cytosol, or from its extracellular elevation.

BDNF release from neurons could be pre-synaptic or postsynaptic, but also somatic or dendritic. BDNF is also stored and released from the astrocytes, as for example in the dopaminergic neurons, to protects them during an oxidative stress, contributing to typical effect of the glia as supportive tissue for the neuronal apparatus.

To understand the BDNF mechanism of action, a good system is provided by cerebellum, because here the growth factor is involved in formation and modulation of brain circuit. In the hippocampus, through TrkB binding, BDNF acts as a regulator of the synaptic transmission and long-term potentiation (LTP). Long Term Potentiation is a mechanism of synaptic plasticity (ability of synapses to change strength and efficacy, in response to a stimuli) and it is associated to the formation of new synapses [28]. This phenomenon changes the number and morphology of dendritic spines, the density of post synaptic connections, learning and memory processes. When a L-type voltage-gated calcium channel is activated, BDNF is released from the pre-synaptic vesicles and, in the synaptic cleft, BDNF can simultaneously induce retrograde and anterograde signals. In pre-synaptic space, BDNF through TrkB activates PLC pathway which rise the calcium levels, potentiating BDNF effect.

In the post synaptic way, binding BDNF/TrkB causes the NMDA receptor activation and an increase in calcium concentration. At the same time, number of AMPA receptors on the post synaptic membrane increase. The total effect on the post synaptic neurons is the cytoskeletal rearrangement, in fact BDNF can modify the structure of the spine and can induce growth in dendritic spine and new spine formation.

Therefore, BDNF is now considered as the most critical factor regulating dendrites and axons growth/branching, and its contribution depends on Trk activation. While a positive effect of BDNF/TrkB signalling on the dendritic spine structure and plasticity, ProBDNF and its p75 receptor exert an opposite effect as reduction of spines or immature spines, connecting with an endocytosis of the NMDA and AMPA receptors [29].

Alteration of BDNF expression level represents a shared risk of vulnerability to neurodevelopmental disorders, which are characterized by synaptic defects, and also many neurodegenerative diseases [28].

It is crucial to dose BDNF levels as it could be a possible marker for diagnostic and therapeutic usage, and a better understanding of BDNF signalling mechanisms and neuronal functions, will help to conceive news drugs and innovative strategies for many neurological and non-neurological diseases.

3.4. Sensory Neurons

One of the earliest identified roles of BDNF is to promote the peripheral sensory neurons growth, during brain development [27]. And in a retinal ganglion cell (RGC) from a trauma rat model, it has been observed that continuous application of BDNF for 12 days induced apoptosis and hypothalamus neurogenesis. BDNF is expressed in small and medium sized sensory neurons that also express the NGF receptor, TrkA and calcitonin gene-related peptide (GCRP).

BDNF is synthetized in the dorsal root ganglion (DRG) neurons and transported into the spinal dorsal horn and binds to TrkB on the second order sensory neurons [30]. There it has a neuromodulator role in synaptic transmission and spinal nociceptor. Following a nerve injury and after peripheral inflammation, BDNF mRNA and BDNF protein changed in DRG. There is upregulation of BDNF synthesis in the small and medium size DRG, there is a correlation between the inflammation alarm evoked by NGF, which is augmented in the inflamed tissue and the enhanced retrograde transport of NGF to DRG, that leads to an increase of BDNF. It is reported that also NT-3, and other factors, like cytokines (IL-6) and TNF-α, can increase BDNF levels, and that BDNF itself could have an autocrine and paracrine role [20].

When degeneration fibre occurs, a cascade of cellular and molecular events initiates and ultimate with a plethora of chemokines, cytokines and growth factors being released. In

particular in pain conditions, BDNF increase in DRG by activation of MAPK, but in a pain response not only the neurons are involved. For example ATP-signals, depolarized the membrane and stimulate BDNF production, as a communication molecule between neurons and spinal glia [31].

Changing in BDNF expression in DGR and spinal cord can play an important role in pathological mechanisms of chronic pain and therefore blocking BDNF in sensory neurons, could represent a new approach to treat clinical inflammatory and neuropathic pain, which is a secondary pathological change in first order sensory neurons damage [32].

Even if all neurotrophins show a role in the retinal development and visual system, when injected in the superior colliculus, after injury, BDNF promotes the survival of neonatal retinal ganglion cells. Also, after axotomy in adult rats RGCs survival is promoted by NGF, BDNF or NT 4/5, intraocular injections [33].

4. Recombinant Protein Production

We are observing a rapid growth in the number of approved recombinant products in USA and Europe, and this reflect the evidence that biopharmaceuticals have the potential to become the medicines of the future. Furthermore, extraction and purification of natural products from human is not always a feasible procedure to meets the clinical request and the growing market drug request. This is particularly evident when proteins are weakly or transitory expressed, or for safety reasons such as possible virus contaminations.

The pharmaceutical processes must comply with the regulatory guidelines administered for instance by the European Medicines Agency (EMA) or U.S. Food and Drug Administration (FDA) and harmonized by the International Council of Harmonization of Technical Requirements for Registration of Pharmaceutical for Human use (ICH) and the International Organization for standardization (ISO). The main object of these rules is to guarantee patients safety and drug efficacy [34].

Biotechnology is based on recombinant DNA technology, a very powerful tool for the development of biopharmaceuticals, and allows to produce human therapeutic proteins on large scales. Furthermore, if necessary genetic modifications can also be introduced to improve proteins structures and functions, to better solve some diseases problem.

Initially the use of microorganisms was primarily linked to the human consumables and only from the 20th century cultured cells widespread the usage for medical and industrial aim [35]. Since 1970s, advances in molecular biology and genetic engineering have led to big progress to understand the molecular mechanisms [36], that undergo in several human diseases.

The first use of recombinant technology has been pioneered in 1973 by Boyer, who transformed *E. coli* cells with a plasmid and founded later the Genentech company (San Francisco, California). In 1982, the first treatment based on this new protocol found its application in the production of recombinant human Insulin (Humulin) [36].

After three decades, we assisted in an amplification of the recombinant technology and its application in several pathologies and sectors, and nowadays biomolecules represent the core of the medical industry. More than 110 biotechnology companies are present and involved in discovery, development and commercialization of recombinant drugs.

Analysis of the approved drugs and commercial drug markets reveal that recombinant products have a growing trend and will dominate the market in the next future.

Therapeutics proteins represent the first drug category to be produced in recombinant way, and up to now represent the main approved proteins, ranging from peptides to monoclonal antibodies, each one with specific action and size.

Crossing the first period that could be consider as the "experimental decades", because knowledge about the recombinant technology was minimal and amount of restriction enzymes and expression vectors were limited. This period is represented by several products, and 30% of the market was occupied by interferon and cytokines, first biotech products for defined treatments (inflammatory pathologies, antineoplastic, multiple sclerosis). The two milestones, insulin and growth hormones paved the way for the market ingress to other companies, as biotech productions were becoming easier and to the presence of large markets request, but also generated conflicts for the intellectual properties and market exclusivity.

The main therapeutic area were cardiovascular diseases, antineoplastic, diabetes and endocrinology, anti-infective, autoimmune. Few products were developed for enzyme therapy and ophthalmology. During the period, going from 1990 to 1997, many clinical studies have been dedicated to new biotech products, 19 % of them remained in the monitoring phase, but the other ones have been approved. The therapeutic areas remain the same and, in US, a 35% growth was observed, due to the introduction of 4 categories approved in healing, enzymes, respiratory and bones therapies. From 1998 to 2003 the number of drugs tested in human remains the same one of the previous decades and 18 drugs have been successively approved in four years.

The market during these periods grows from 21 470 million dollars to 32 065 million dollars in 2003. The ten leading products drove market growth in 2001-2003, their sales increased from 12 923 million dollars in 2001 to 18 362 million dollars in 2003.

The growing number of biopharmaceutical products approved, also in the period 2006-2010, was confirmed passing to a total of 58 new drugs, of which 17 were produced in *E. coli*, 32 in mammalian cell lines and 4 in *S. cerevisiae* or transgenic animals.

Recent bioinformatic techniques able to predict expression problems, such as solubility during overexpression, have permitted to choose the best expression systems or to improve some expression limitations [35].

The principal driving factors, that boosted the growing biotech market products, are due to the increasing Research & Development (R&D) investments and clinical activities in biopharmaceuticals, and also to the intensified demand of animal-free recombinant products. In details many companies like Roche, Amgen, Pfizer, and Novo Nordisk have a strong recombinant proteins R &D pipeline and reinvest an important part of their revenues in R&D every year. Furthermore, big companies engaged research collaborations with smaller ones, or Universities, to develop novel drugs. For the second point, there is a growing ethical concern, with the use of animals for antibodies productions, taking also into account that it is a time and cost-consuming procedure [35].

In 2014, up to a quarter of the biopharmaceutical approved by U.S. Food and Drug Administration are produced from *E. coli*, but the number of laboratories that are involved in strain development still remained low. In general, two strains are used: B for DNA manipulation and K12 for protein expression. They share a common genome size and organization [37].

Up to 2019, more than 170 recombinant proteins have been used in the worldwide medicine and, in first attempt, only proteins identical to the natural ones have been produced, then more modified proteins have been generated. The most common techniques often used are mutagenesis, site-directed mutagenesis, domain deletion and/or additions, etc [38].

Other expansion of the recombinant proteins market is due to the higher incidence of new chronic diseases, with the development of dedicated R&D programs, and sometimes favourable funding from the government. The market is also driven by new products development and innovative biotechnology activities. Growing R&D programs and development of new healthcare technologies is further contributing to market expansion. According to a recent report, a huge influx of R&D expenditure and constructive funding is expected in the recombinant protein market. Driven by India and China, facing several untreated diseases, market expand and new solutions for recombinant protein expression are expected to come out.

In 2024, 84 new drugs have been authorized in UE and USA and among them 36 are biotech (43%). These drugs represent, in the last three years, 46% of the global research pipeline. More specifically, in Italy in 2023, clinical trials of biotech authorized drugs, have been 241 in a total of 611, representing around 39% (Reports AIFA on clinical experimentation of 2024).

The current Market Size in 2025 is equal to 3.97 billion USD and the years before was USD 3.52 billion. It is expected that in 2034 market will reach 11.32 billion USD. The Largest Market is still North America, but the fastest growing one is situated in the Asia Pacific Area.

Today, approval and launch of DNA vaccines is driving the market dynamics. Certain biotech therapies are among the fastest growing drug market segments (mAbs sales grew by 52% in 2005); five of the 20 best-selling drugs are large molecules. Biotech increasingly dominates the pipeline (44% of all discovery stage candidates) and has a growing share of drug applications (about one in ten filings) [35].

It is predicted that the enzyme market will growth from 6.6 million USD in 2021 to 9.1 billion by 2026. And the application of recombinant protein strategies is variable, as Antibodies, vaccines, hormones, demonstrating that Recombinant proteins play a significant role in biopharmaceutical market [39].

The recent pandemia gives a boost in this field with a fast and efficient response to a worldwide problem. Several companies have been implicated in developing new therapies and new vaccines to fight the virus.

Comparing the Pharmaceutical productions to the non-pharmaceutical ones, drugs have high costs and long development periods to, around 10-20 years, to cover the numerous clinical trials and the differences in costs with the other industrial productions is 9-folds. The cost of a single clinical trial varies from 5 million dollars, for the small ones, to 350 million dollars for the larger ones. Recent works estimated that bringing to the market a single drug costs 1.4 billion dollars. For example, recombinant products as industrial enzymes (cellulase or glucosidase for bioethanol), do not necessitate clinical trials and final prices are between 10 to 37€ per kg whereas 1 kg of coagulation Factor VIII, to treat haemophilia, has a cost of 9.6 Billion USD [38].

In conclusion, the recombinant proteins market continues to show promising growth prospectives, driven by advancements in technology, increasing research and development activities, and a growing demand for new biopharmaceuticals to cover still unmet medical needs. With such a competitive landscape, stakeholders in this market must stay proactive to tackle opportunities and contribute to the advancement of innovative therapies and solutions in different disease areas. Furthermore, more than 400 million people globally suffer from uncommon genetic illnesses, due to genetic deficiencies, underscoring the critical need for better health awareness, clinical settings, and alternative treatments.

For medicine development, cutting-edge research and drug-targeting mechanisms understanding are key resources.

5. Gene Design

Nowadays the pharmaceutical industry must provide large therapeutical proteins quantities, to face request needs. This is obtained through recombinant DNA technology, which is the only way to manufacture high quantities of human recombinant proteins for human use.

In Dompé we focus on recombinant proteins expression in prokaryotes, and the following chapter will describe the bases of this fast-evolution technology.

In nature, bacteria possess vectors, which are circular DNA also called plasmids, that are used as a horizontally information vehicles, to propagate new functions and resistance capacities [40].

Plasmids are fundamental tools for biotechnology and important agents in bacterial evolution and antibiotics resistances. Understanding the basic biological of plasmid permits to improve their application in those two areas.

In details, genetic material contained in the plasmid is small and has an autonomous structure, it possesses as a matter of fact, all the necessaries components to autonomously replicate within host cells, as they contain key elements like an origin of replication and often selectable markers. They have therefore been the tools of choice to introduce foreign genetic material into a host cell [40].

They have been widely used for genes cloning, or to study gene expression, and have open the doors to gene therapy and recombinant proteins productions. In this last case, expression vectors are the basic tools in biotechnology, for the production of proteins in a “simple” and successful way. One of the more useful conceptual aspect of plasmids is their inherit modularity, both as autonomously replicating genetics elements, as well as the level of their constitutive genetic components [40].

Before vector transformation in the host, it is necessary to study how to encode the gene of interested. Sometimes, for particular reasons, protein sequences can be modified to increase their expression and/or biological functions. Once that the sequence has been engineered, it is possible to proceed with molecular cloning, with the insertion of the DNA sequence of interest into the expression vector.

In *E. coli*, the probability of soluble expression, is inversely proportional to the recombinant protein molecular weight, especially for proteins > 60kDa (Canaves et all., 2004, Goh et al., 2004, Graslund et all., 2008). For example, a large-scale study

examining the protein properties of 95 recombinantly expressed mammalian proteins found that small proteins (average molecular weight of 22.8 kDa) were often expressed as soluble “alone”, while larger ones (average molecular weight of 40.4 kDa) were only soluble expressed when fused to solubility-enhancing tags (Dyson et al., 2004).

In a separate study, small proteins (100 amino acids or less) were soluble expressed in *E. coli* at levels suitable for purification in 47% of the cases, while large proteins (600 to 800 amino acids) were expressed soluble, only 33% of the time (Graslund et al., 2008a). Thus, when using *E. coli* as an expression host, it is typically advantageous to express individual protein domains, as opposed to the full-length protein, whenever possible [41].

As M. Kamionka point out, it is important to define a “Clean” product, and to choose the correct expression vectors among the numerous commercially available ones. To achieve high target protein expression, the choice of an adequate vector is a crucial factor [42].
Efficient protein production is “the” objective of a well-designed expression vector, and this starts by the production of significant amount of stable RNA messenger , which will then be translated into the protein. According to vectors, protein expression can be tightly controlled, and will only be produced, when necessary, in significant quantities.
Playing on different parameters, one can try to avoid recombinant protein precipitation, or favour it to produce inclusion bodies, or avoid abnormal protein localization, to ease protein recovering, at the fermentation process end.

5.1. Expression Vector Elements

An expression vector must include all the necessaries elements for gene expression, which are: an origin of replication, a promoter, a selection marker, the correct translation initiation sequence such as a ribosomal binding site (RBS), the start codon, a termination codon, and a transcription termination sequence.

Obviously, there are differences between prokaryotes and eukaryotes, in the transcription and translation molecular mechanism machinery. For example, prokaryotes expression vectors have a Shine-Dalgarno sequence for ribosomes binding in their translation initiation sites, whereas eukaryotes expression vectors contain the Kozak consensus sequence [43].

It is possible to used different variants of the origins of replication (ori), that guarantee the autonomous replication of the circle DNA in the cytoplasm, to tune the desired vector copy number. This to avoid host stress by excess of replication in the cytoplasm. The most common origin of replication includes: ColE1, pMB1, pSC101, R6K, and p15A.

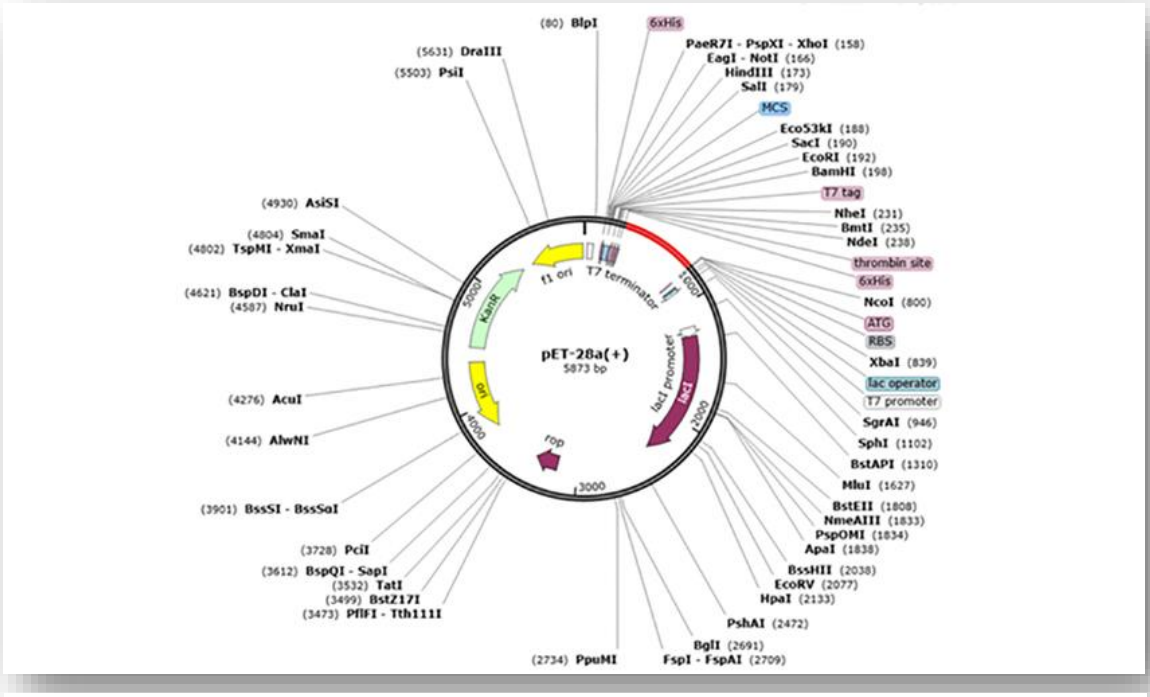


Figure 7- pET 28a Vector Map, with the principal components: ori, resistance, T7 term, LacI promoter, enzymatic restriction site. Ref: Immunoinformatic assisted profiling of West Nile virus proteome to determine immunodominant epitopes for the development of next-generation multi-peptide vaccine, https://doi.org/10.3389/fimmu.2024.1395870.

Generally, control of replication is referred as "relaxed" or "stringent", depending on whether the *ori* is positively regulated by RNA or proteins, respectively.

High copy number plasmids guarantee a high number of plasmid, but represent a high metabolic burden for the host. Plasmid copy number can vary from 20 copies per cell with pMB1 *ori*, while pUC will maintain as much as 700 copies per cell. A low copy number plasmid permits to reduce the number of replications, the most commonly used are pBR322 and pET, that give between ~15-20 plasmid copies. Therefore, the *ori* choice depends not only on how many plasmid copies are desired, but also compatibility with the host that one intends to use and, if more than one plasmid is contemporaneously used, plasmid's compatibilities are also to be considered.

Indeed, plasmids with the same *ori* are incompatibles, because they will compete for the same machinery, creating an unstable and unpredictable environment, in which one of the two will be lost after a certain number of generations. The table below reports the most common replication origins and their main characteristics [44].

Common Vectors	**Copy Number**	***Ori***	**Control**
pUC	~500-700	pMB1 (derivative)	Relaxed
pBR322	~15-20	pMB1	Relaxed
pET	~15-20	pBR322	Relaxed
pGEX	~15-20	pBR322	Relaxed
pColE1	~15-20	ColE1	Relaxed
pR6K	~15-20	R6K	Stringent
pACYC	~10	p15A	Relaxed
pSC101	~5	pSC101	Stringent
pBluescript	~300-500	ColE1 (derivative)	Relaxed
pGEM	~300-500	pUC	Relaxed

Table 1- Common vector used in the recombinant protein production. Table described the main characteristics.

There are other factors that affects plasmid copy number: the insert length and its toxicity, *E. coli* strain, growth conditions and culture inoculum. Culture conditions are important for the expression, so the dissolved oxygen amount, temperature, culture volume, antibiotic, and medium can all affect copy number. Also, some origins of replication are sensitive to temperature, indeed with low ones and for the same plasmid, duplication time is slow down and plasmid copy will be low, and with higher temperatures, duplication is doubling or decupling, giving higher plasmid copy number.

The second element necessary for the vector design is the promoter, a DNA sequence that initiates the transcription and therefore controls expression of the cloned gene. Normally selected promoters are inducible, meaning that protein synthesis is spatially and temporally defined by an inducer, properly added when transcription is required. The most used inducer is Isopropyl β-D-1-thiogalactopyranoside (IPTG), a non-hydrolysable galactose analogue. In some rare cases gene expression can be constitutive (protein is constantly expressed) [43].

The table below reports the main promoters used.

Promoter	**Primarily used for**	**Description**	**Expression**	**Additional considerations**
T7	*in vitro* transcription/ general expression	Promoter from T7 bacteriophage	Constitutive, but requires T7 RNA polymerase.	Also used for *in vitro* transcriptions
T7lac	High gene expression levels	Promoter from T7 bacteriophage plus lac operators.	Negligible basal expression when not induced. Requires T7 RNA polymerase, controlled by lac operator. Induced by IPTG.	Commonly found in pET vectors. Very tightly regulated by the lac operators. Good for modulating gene expression through varied inducer concentrations.
araBAD	General expression	Promoter of the arabinose metabolic operon.	Inducible by arabinose and catabolically repressed in presence of glucose or by competitive binding of the anti-inducer fucose.	Weak promotor. Commonly found in pBAD vectors. Good for rapid regulation and low basal expression. Not well-suited for modulating gene expression through varied inducer concentrations.
Trp	High gene expression levels	Promoter from *E. coli* tryptophan operon	Repressible	Gets turned off with high levels of cellular tryptophan.
Lac	General expression	Promoter from lac operon	Constitutive in the absence of lac repressor (lacI or	Leaky promoter with somewhat weak expression. $lacI^q$ mutation

			lacI^{q}). Induce by IPTG or lactose.	increases expression of the repressor 10x. Good for modulating gene expression through varied inducer concentrations.
Ptac	General expression	Hybrid promoter of lac and trp.	Regulated like the lac promoter	Contains -35 region from trpB and -10 region from lac. Very tight regulation. Good for modulating gene expression through varied inducer concentrations. Generally better expression than lac alone.
pL	High gene expression levels	Promoter from bacteriophage lambda.	Can be temperature regulated	Often paired with the temperature sensitive cI857 repressor.
T3	in vitro transcription/ general expression	Promoter from T3 bacteriophage	Constitutive, but requires T3 RNA polymerase	Used for *in vitro* transcription.

Table 2- The table shows principal promoters used in the recombinant protein's productions. Combined with the main characteristic, are also reported the usage and troubleshooting according to the protein expression optimization. D. M. F. a. R. Page, "Strategies to Optimize Protein Expression in *E. coli*," *Current Protocols in Protein Science,* 2014. **[41]**

In summary, the promotor choice mainly depends on host metabolic load, recombinant protein toxicity and the fermentation design.

Due to its intrinsic properties the lac promoter has been the first to be used in molecular biology, but it was "weak" and leaky. It was therefore combined with the Trp one in the Tac construct resulting better repressed and 10 times more potent. This promotor is

repressed by the Lac I gene product (repressor) which is present, as a single copy, on the bacteria chromosome, but can also be cloned in the expression vector giving therefore a higher repressor concentration in the bacteria. This effect can be ulteriorly amplified using the Lac I^q repressor version further increasing its concentration.

At the end, if one has the lac I^q gene variant present on the expression vector, together with the Tac promoter, this later has a fine-tuned transcription, with no basal expression, which can be turned on upon IPTG addition. This system has been considered for decades as the most efficient in *Escherichia coli* and only recently substituted by the more performant T7 one.

The T7 promotor, prevalent in the pET vectors, is frequently employed because the gene of interest is placed, under the transcriptional control of the T7 RNA polymerase, from the T7 bacteriophage. In this system, the vector needs to be transformed in an *Escherichia coli* strain, bearing in its chromosome the T7 polymerase, under the control of lacUV5 promoter inducible by lactose or IPTG. This system is particularly performant because the T7 promotor is, up to now, the most efficient promotor in *Escherichia coli* and possess the faculty to turn off the other host cell promotors, favouring therefore the recombinant gene transcription [45].

PBAD promoter derives from the arabinose operon and possesses positive and negative regulators and has the advantage to use arabinose which is cheaper than the IPTG. Finally, pL promoter can be modulated by pH or temperature [21].

In summary an optimal choice consists in the selection of a tightly controlled strong promoter, with low basal expression level.

The selection marker is another essential vector element, it is first used after transformation to select the transformed cells, and then to maintain a selective pressure during fermentation to have a homogeneous population, where each cell maintains the expression vector and can produce the recombinant protein. Typically, they encode antibiotic resistance genes, like kanamycin, streptomycin, ampicillin, chloramphenicol, etc.

In prokaryotes, just downstream of the promotor, one will find a translation initiation sequence named ribosomal binding site (RBS) or Shine-Dalgarno, that initiates the translation process. RBS sequences have been shown to modulate the expression of

heterologous genes, and their optimization can lead to higher products yields. Translation initiation is affected not only by the consensus Shine-Dalgarno sequence, but also by the nucleotides upstream of the initiation codon, the secondary structures of the ribosomal binding and of mRNA can also prevent ribosome binding [46].

In *E. coli* the translation starting codon is always an ATG, codifying for a methionine, and the termination codons, or Stop codons, identify the translation end. In *E. coli*, termination codons are UAA, UAG and UGA. But those codons are leaky, and it has been shown that substituting UGA with the more efficient UAA triplet, results in a 2-fold increase in the protein expression level, in some genes two consecutive stop codons can be found [43].

Moreover, transcription terminators stabilized the mRNA at its 3' end by creating a stem loop structure.

The protein sequence inserted in the vector is nowadays not limited to its cDNA and can be the natural protein sequence, or its precursor or can directly encode the mature one, including some point mutations, with the intend to improve the expression [42].

Also, elements to address the recombinant protein synthesis, in a specific cellular compartment, can be added.

Other strategies to improve protein expression are applied during vector design to enhance for example mRNA stability, codon usage, addition of expression tags, addition of chaperons. As in *Escherichia coli* transcription and translation are two simultaneous processes a stable mRNA is a plus for efficient protein synthesis.

If during gene design at the ORF 5' region between -35 to -40 AT nucleotides are replaced by GC, to improve stability properties, preventing the secondary structure of mRNA. This mechanism prevents also the effective initiation of translation.

Most of the time, for heterologous proteins expression, codon usage in the host is different, than the one of the original organisms, phenomenon known as codon bias. To overcome this problem, gene codons are optimized to render them more compatible with the new host synthesis machinery.

Codon optimizations consist in substituting the rare codons, with more abundant ones in the host, without altering the final protein amino acid sequence. Often the rare codons substitution leads to an improved ribosomes activity and protein synthesis. This is particularly the case for human arginine codons when the protein is express in *Escherichia coli*.

In another strategy to solve the codon bias problem, some plasmids have developed and encodes special tRNAs (pRARE). This type of tRNA are able to recognize in a define host the rare codon sequences.

Expression tag are sequences that guarantee an improve stability, solubility and in some cases ease protein recovery. The most frequent sequences are: His6, trx, MBP, SUMO, GST, which can be added at the N- o C- protein terminus, or even inside the coding sequence. Some are bigger than other, like GST and MBP and this factor can influence the protein refolding, but also compromise the function, activity and toxicity of the protein.

After protein recovery tags can be removed with specific proteases able to cleave off the TAG and for example TEV protease, thrombin, factor Xa, Enterokinase are the most used enzymes. His TAG is frequently used and consist in the addition to the recombinant protein, of five to six histidine. Then a nickel -affinity column can be used to purify the engineered protein [21].

SUMO is the last developed Tag and is used to solubilize difficult to express proteins. When fused to the target protein it can improves its solubility up to 5-50 times and once that the dedicated protease removes the TAG, mature and active protein is obtained.

Chaperons are molecules that helps proteins to refold, in some cases, to increase their concentrations in the host, they can be directly cloned in the expression vector containing the heterologous gene. The protein misfolding risk increases proportionally with the recombinant protein size, and with proteins naturally prone to aggregation. Consequently, chaperones like GroES/GroEL and DnaK-DnaJ-GrpE are sometimes overexpressed in the host, to assist folding.

5.2. HOST: *Escherichia coli*

For recombinant protein production, among all available expression systems, *Escherichia coli,* remain the first choice, however a variety of hosts, including: yeast, mammalian cell lines, filamentous fungi, transgenic animals, insect cells and plants, can also be used.

Escherichia coli attractivity relies on several advantages offered by this prokaryotic host: easy genetic manipulations, well-characterized genome, accessibility of versatile plasmid vectors, availability of a large collection of already engineered strains, fermentation cost-effectiveness and the possibility to obtain very high expression levels. As reported by Zi-Xu Zhang et all. [39], *Escherichia coli* will remain "the brightest star among recombinant protein production host".

Escherichia coli is the most well appreciated expression prokaryotic system, it has been entirely sequenced in 1997, making it the best characterized host, and being continuously used in a large number of processes has accumulated in abundant quantity of available literature easing new processes implementations, render it a very attractive and robust expression system [42]. The simple and efficient genetic architecture is organized on the operons, a construct which is constituted by a single promoter to synthetize multiples proteins from a single mRNA, in which for each protein a ribosome binding site (RBS), precedes the starting AUG codon [47].

Moreover, laboratory manipulations are very simple, and cultivations happens in a low-cost growth medium, rendering fermentation straightforward. For some strains duplication time can be 20 minutes, and high biomass quantities can be obtained by fermentation.

Compared to Eukaryotic systems, both fermentation and cell disruption are procedures easy to scale up, even if harsher mechanical procedures are needed as compared to the gentle lysis used for Eukaryotic cells, where it is necessary to preserve their fragile organelles to obtain a clean product.

However, *Escherichia coli,* as Prokaryotic in general cannot perform most of the post transductional modifications, since they do not have specialized cell compartments, which are present in the Eukaryotic ones [42]. This represents the principal drawbacks of *E. coli* utilization, in the recombinant protein production field, because often the post

translational modifications are essential for the protein folding and the protein functionality.

Moreover, in addition to the inability to perform many of the post-translational modifications, *E. coli* also lacks secretion mechanisms, to actively release protein into the culture medium, and cannot form disulfide bonds in its reductive cytoplasm, those two points are however subjects to further optimization and strain(s) engineering [48].

Due to *E. coli* limited genetic repertoire, the protein folding and post transductional modifications failures often occurs, when a recombinant protein is produced in this organism.

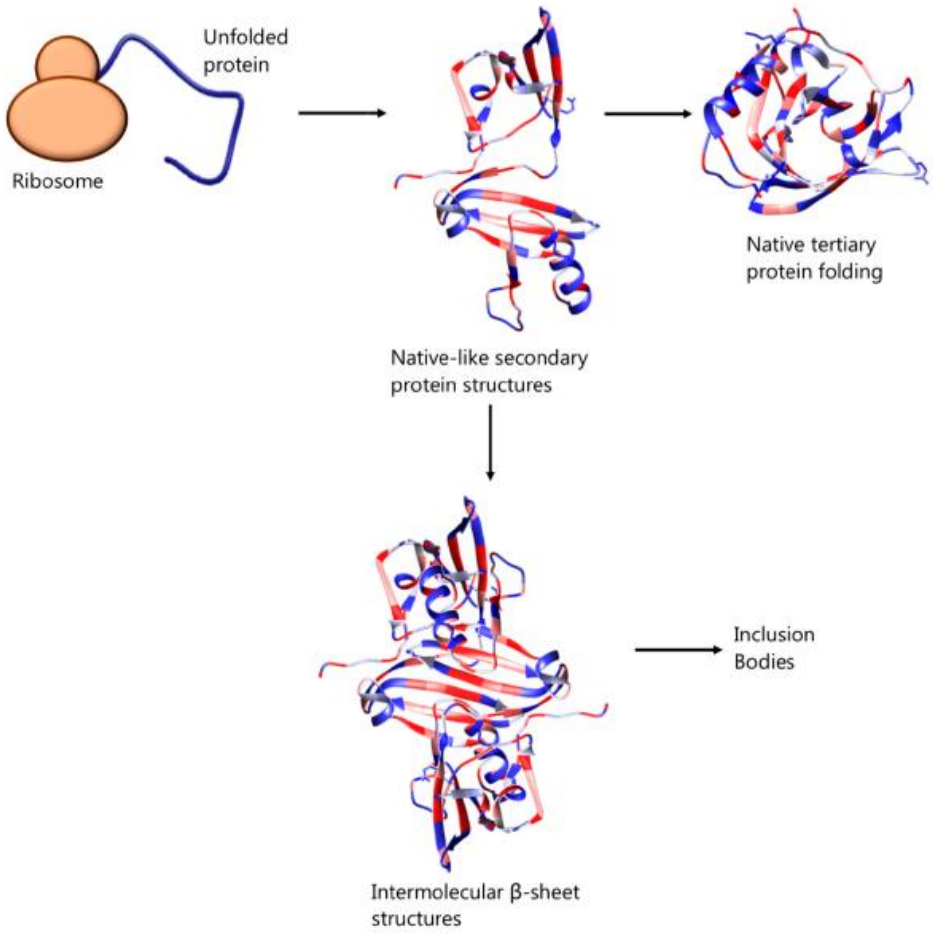


Figure 8- Graphic representation of the Inclusion bodies structure. Ref: V. P. a. F. H. Falcone, Polyionic Tags as Enhancers of Protein Solubility in Recombinant Protein Expression, Microorganism 2018 **[87]**.

As *Escherichia coli* cannot perform the post transductional modifications and together with the reduced redox state, provided by thioredoxin reductase (TrxB) and glutaredoxin reductase system (gor), which prevent the proper formation of disulphide bonds in recombinant proteins, heterologous proteins often accumulate in inclusion bodies.

In details, inclusion bodies are structures composed of misfolded, or partially unfolded, insoluble proteins. In the past, they have been considered as the result of unspecific hydrophobic interactions, among intermediates or partially folded proteins, actually they are considered as ordered dynamic structures [49]. In bacterial inclusion bodies (IBs), Fourier transform infrared microscopy (FTIM) has revealed the presence of native-like secondary structures and intermolecular β-sheets structures, that looks like amyloid fibrils.

Even though IBs are biologically inactive, high production yield and versatility to express different proteins are two advantages associated with their formation [50]. They are stable protein aggregates, resistant to temperature and longtime storage. During the downstream processing, IBs can be easily isolated after cell disruption, and the resultant IB slurry can be frozen and stored for several months, providing manufacturing flexibility. Based on the above considerations, recombinant proteins productions in inclusion bodies can be an acceptable passage for some difficult to express proteins.

IBs refolding, to obtain an active protein, always represents a challenge, as it requires considerable optimization for each new target protein. It is necessary to develop an IBs re-solubilization solution, using chaotropic agents, and to set-up a refolding buffer and adequate conditions to reach the protein native three-dimensional structure [51].

To design an *in vitro* protocol for protein refolding, it is firstly necessary to understand if once refolded the protein can be soluble. In general, the solubilization procedure requires strong denaturants, such as 6M guanidine-HCl or 6-8 M urea, the first chaotropic agent being stronger than the second one [49].

In the case of cystine containing proteins, disulphide bonds are present and these bonds need to be reduced as they are usually miss-formed and therefore reduce solubility, so a reducing agent like 1,4-Dithiothreitol (DTT), dithioerythritol (DTE), Glutathione (GSH), cysteine, cysteamine or 2-mercaptoethanol needs to be used. These low molecular weight thiol reagents reduce the S-S bonds, by thiol -disulphide exchange. When a reducing agent is used, solubilization is usually performed under mildly alkaline conditions.

Refolding protocols are usually performed by dilution or dialysis. For scaling up reasons a rapid dilution of the denatured protein, in the refolding buffer, is the preferred technique. At the same time critical parameters like protein concentration, pH, temperature, ionic strength and kind of ions (Hofmeister series), protecting or chaotropic agents, and low molecular weight folding enhancers, are also taken in considerations.

Temperature commonly applied, range from 0°C to 40°C, with respect to pH, the isoelectric conditions should be avoided in denaturation/renaturation experiments. Furthermore, pH conditions must be optimized within the range of pH stability of a given protein. The Time required to complete renaturation may be comprised between hours and days. There are a lot of "protein enhancers" that can be added to the refolding buffer to increase the refolding yield; among them L-Arginine has been demonstrated to have a positive effect *in vitro*, but sometimes its role is enigmatic.

Other enhancers like Tris, Guanidine Hydrochloride, Lauryl maltoside, Triton X-100, phospholipids, glycerol, metal ions, ammonium sulphate or polyethylene glycol can be used.

The non-ionic, as well as ionic or zwitterionic detergents, have been found to also have positive effects in the folding process; these micelles of polar and non-polar moieties can transiently interact with various sticky sites of intermediates in the folding pathways. Applying detergents with cyclodextrins has been described to exhibit a chaperone like action.

5.3. Commercial Host Strain

The host strains are often divided into six classes: protein-free protease (BL21, BL21 DE, BL21Star), toxic proteins expressing strains (BL21 plysS, BL21 pLysE), strains suitable for expressing protein with S-S bonds (Origami and Shuffle), strains designed to overexpressed t-rare RNAs codons (Rosetta e BL21 codon-Plus), strains capable to produce membrane proteins (Lemo21 (DE3), C41(DE3), C43(DE3)), inducible proteins expression strains and finally strains able to express aggregation-prone proteins (Arctic Express DE3) [46].

In general, BL21 DE3 is the most used strain because it has several advantageous characteristics like the absence of the two OmpT and lon proteases, a copy of T7 RNA polymerase gene under the control of the lac promoter, a fast growth and a higher robustness as compared to the K strain. Moreover, this strain guarantees a higher mRNA stability due to mutation of RnaseE gene.

Table 3- The main commercial strains used for recombinant protein production. In the table are also listed all the main characteristics and the genotype. Ref: Dana M. Francis and Rebecca Page, Strategies to Optimize protein Expression in *E.coli*, Current Protocol in Protein Sciences, 2010. **[41]**

Strain	Resistance	Key Features	Genotype	Use
BL21 (DE3)	Kanamycin	IPTG-inducible strain containing T7 RNA polymerase (DE3)	F$^-$ *omp*T *lon hsd*SB (rB- mB) *gal dcm* (DE3)	General protein expression
BL21 (DE3) pLysS*	Chloramphenicol (pLysS)	pLysS expresses T7 lysozyme to reduce basal expression levels. Expression vector cannot have p15A origin of replication	F- *omp*T *lon hsd*SB (rB- mB) *gal dcm* (DE3) pLysS (CamR)	Expression of toxic proteins
BL21 (DE3) pLysE*	Chloramphenicol (pLysE)	pLysE has higher T7 lysozyme expression than pLysS. Expression vector cannot have p15A origin of replication	F- *omp*T *lon hsd*SB (rB- mB) *gal dcm* (DE3) pLysE (CamR)	Expression of toxic proteins
BL21 star (DE3)	Kanamycin	Lacks functional RNaseE which results in longer transcript half-life	F- *omp*T *lon hsd*SB (rB- mB) *gal dcm rne131* (DE3)	General expression; not recommended for toxic proteins

BL21-A1	Tetracycline	Arabinose-inducible expression of T7 RNAP; IPTG may still be required for expression	F- *omp*T *lon hsd*SB (rB- mB) *gal dcm ara*B::T7RNAP-*tet*A	General protein expression
BLR (DE3)	Tetracycline	RecA-deficient; for plasmids with repetitive sequences.	F- *omp*T *lon hsd*SB (rB- mB) *gal dcm* (DE3) Δ(srl-recA) 306::Tn*10* (Tet^R)	Expression of unstable proteins
HMS174 (DE3)**	Rifampicin	RecA-deficient; allows for cloning and expression in same strain	F- recA1 hsdR(rK12- mK12+) (DE3) (RifR)	Expression of unstable proteins
Tuner (DE3)	Kanamycin	Contains mutated lac permease which allows for linear control of expression	F- *omp*T *lon hsd*SB (rB- mB) *gal dcm lacY1* (DE3)	Expression of toxic or insoluble proteins
Origami 2 (DE3) **	Streptomycin and Tetracycline	Contains highly active trx reductase and GSH reductase to facilitate proper folding; may increase multimer formation	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F′[lac+ lacIq pro] (DE3) gor522::Tn10 trxB (StrR, TetR)	Expression of insoluble proteins
Rosetta2 (DE3)*	Chloramphenicol (pRARE)	Good for “universal” translation; contains 7 additional tRNAs for rare codons not normally used in *E. coli.* Expression vector cannot have p15A origin of replication	F^- ompT hsdSB(rB^- mB^-) gal dcm (DE3) pRARE2 (CamR)	Expression of eukaryotic proteins
Lemo21 (DE3)*	Chloramphenicol (pLemo)	Rhamnose-tuneable T7 RNAP expression alleviates inclusion body formation. Expression vector cannot have p15A origin of replication	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo (CamR)	Expression of toxic, insoluble, or membrane proteins
m15 pREP4*, **	Kanamycin (pREP4)	Cis-repression of the *E. coli* T5 promoter (found on vectors such as pQE), inducible under IPTG (lac repressor on the pREP4 plasmid). Expression vector cannot have p15A origin of replication	F-, Φ80ΔlacM15, thi, lac-, mtl-, recA+, KmR	Expression of toxic proteins

As in the current neurotrophins production processes, those recombinant proteins are expressed, in an insoluble form in inclusion bodies, in the *Escherichia coli* cytoplasm, it was important to explore further strategies to try to express them in a soluble form still in this microorganism. Indeed, their productions have been hampered by their complex three-dimensional structures based on the formation of three disulphide bridges forming a cysteine-knot motif which represent the core structure in the mature protein.

We therefore wanted to explore conditions in which disulphide bonds can be formed in *Escherichia coli*, corresponding to the first mandatory parameter, to obtain the protein in its native conformation. One possibility fulfilling this requisite consist in the neurotrophin expression outside of the bacteria cytoplasm, which is a naturally reducing compartment, or by engineering the strain to have an oxidative cytoplasm.

6. Periplasmic Expression

Recombinant protein secretion in the medium, compared to intracellular expression strategies, represents a great advantage for their recovery [52]. But *E. coli* usually does not secrete proteins into the culture medium; recombinant products are often once synthetized, located inside the cell, and in the worst case, if the protein is huge, toxic, poorly soluble or represent an important burden for the host, they are accumulated in the cytoplasm as inclusion bodies. Furthermore, if the protein necessitates particular structures, like disulphide bonds formation, their export in an oxidative environment, like periplasm, permits to obtain them with a correct folding and therefore soluble [48].

Periplasmic expression is a valuable compromise strategy that exploit two natural *E. coli* advantages: transmembrane transport mechanism and its oxidative environment. Indeed, using the natural mechanism of protein transport and cleavage of its signal peptide, permits to directly obtain the mature protein in this cellular compartment. As the periplasmic space possess an oxidative environment, disulphide bridges can form and, in the case of BDNF, protein can acquire its correct folding and consequently its biological activity. Moreover, periplasm compartment protects the neosynthesised proteins from the proteases, as only 7 of the 25 *Escherichia coli* known proteases are present in this compartment (4-8%) [53].

The first step to address recombinant protein expression to the periplasm, pass through the expression vector design. However, no standard procedure exists, as every protein represent a special case, but simple shared mechanisms have been described in literature, to address recombinant protein expression in this compartment.

Briefly it is first important to distinguish Gram-Negative from Gram-Positive membranes structures. Gram-positive Bacteria, as *Bacillus subtilis*, only possess a single membrane; whereas Gram-negative ones are surrounded by two membranes, the outer one (OM) and the inner one (IM), separated by a space called periplasm [54].

The cytoplasmic Inner Membrane is composed of bilayer phospholipid and α-helical integral proteins and in the outer leaflet lipoproteins are anchored. The peptidoglycan layer, also called murein, is composed of glycan strands, cross-linked to peptides which provides cells strength and shape. This peptide glycans are immersed into the periplasmic space. The Outer Membrane is a barrier and protect cells from the outside, and is

composed of an asymmetric bilayer, in which the inner leaflet consist in phospholipids and the outer leaflets presents lipopolysaccharides that have an endotoxins role, and causes the immune response in human [48] .

This is a result of a structural evolution of prokaryotic microorganisms, some adaption mechanism to solve the problem of toxic exclusion. Indeed, some enzymes are toxic in the cytoplasm (phosphorylase, etc) and are stored in this delimited compartment [42].

The periplasm evolved as the first extracytosolic space to provide efficient and diverse mechanisms of protein oxidation, folding and "quality control" of their 3D structures [54], with high homology to the eukaryotic plasmic reticulum. It looks like a gel-like structure, and represent the 30% of total cell volume, with a width in electron spectrometry equal to 50nm [55]. It is filled of peptidoglycans and with large pores, and the protein present move much slowly, then in the cytoplasm, and the protein remain in contact with the other periplasm proteins or with membrane components.

Protein secretion consists in a translocation of the molecules from the inside of the cell to the exterior, and the secretome is defined as the different proteins involved in the transport mechanism. There are several proteins involved in one-ways translocation, that is a transport from the cytoplasm directly into the medium, and other proteins that are involved into two-steps mechanism (cytoplasm-periplasm, periplasm-outer), passing firstly into the IM and then to the OM [48]. This last one is the major studied for its utilization in the periplasmic expression strategy.

Inner Membrane has a complex structure which is involved in specialized mechanisms for proteins translocation. Among them, there are two major mechanisms of translocation: TAT and Sec. TAT pathway can transport the complete folded proteins and the Sec pathway the partially folded or unfolded proteins, in this later case folding can occur in a co-translational or in a post-translational manner.

The first translocation system is named TAT or Twin -Arginine Translocation, due to the arginine pair presence in the signal peptide of the TAT substrates. In Gram negative bacteria TAT translocase is composed of three proteins Tat A, B and C. Tat pathways is slower and less efficient for the lower number of translocons, but protein refolded in cytoplasm prevent aggregation and has advantages to be more stable.

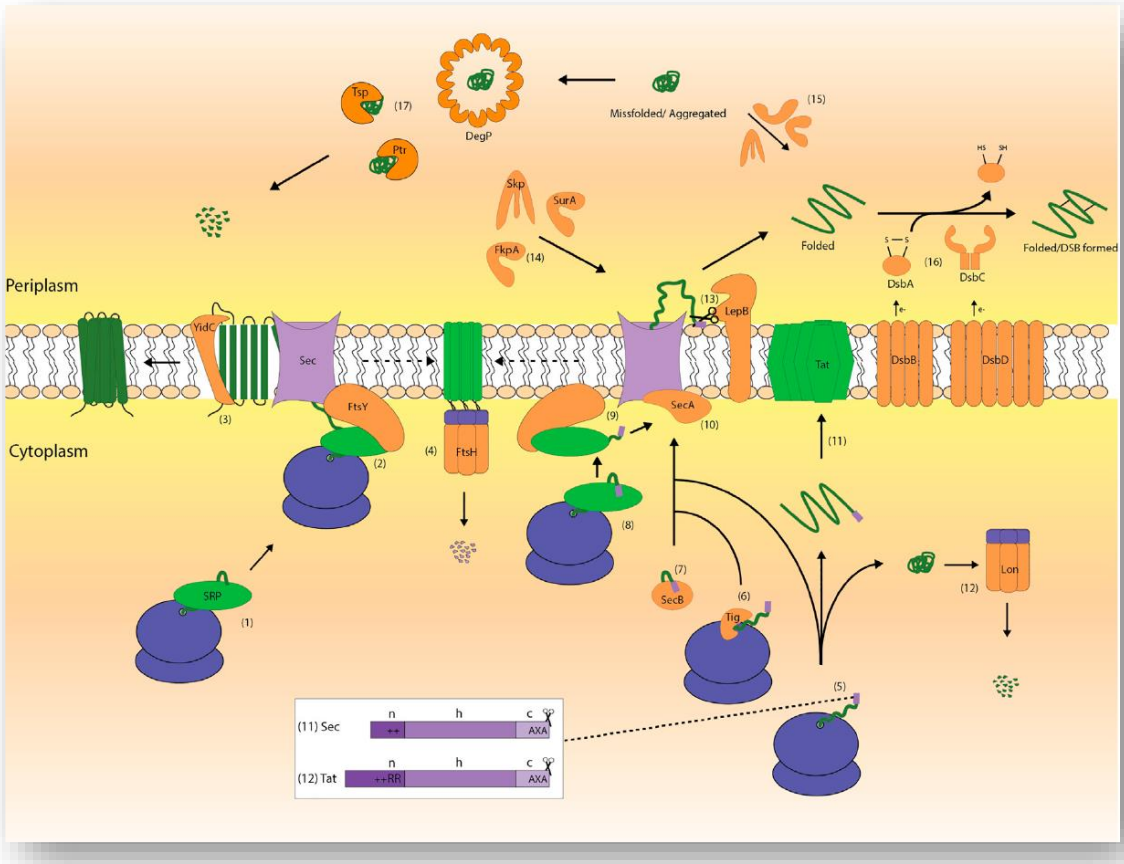


Figure 9- Complex structure of the Inner Membrane. Presence of channel and transmembrane proteins involved in the translocation mechanism. The two different system called SEC e TAT pathways guarantee a strictly control of folded and not still folded proteins. Picture from review of Alexandros Karyolaimos and Jan-Willem de Gier, entitle: "Strategies To Enhance Periplasmic Recombinant Protein Production Yield in *Escherichia coli*, Frontiers in Bioengineering and Biotechnology", published in December 2021, Volume 9. **[56]**

To meliorate TAT translocation sometimes *E. coli* can be modified with translocation systems of other microorganisms like *B. subtilis, Salmonella etc*, and sometimes other chaperones are added to improve translocation efficiency [56].

In the Sec co-translational pathway, a highly hydrophobic signal peptide is recognized by a signal recognition particle (SRP), during protein transduction. Most of the secreted proteins are transported with this way and the Sec translocase is composed of SecYEG complex, an heterotrimeric channel that pass through the membrane and the Sec A protein, which is a ATP-dependent motors [48]. The nascent sequence is targeted to the FtsY a protein membrane, that transfer the ribosome nascent chain complex (RNC) to the translocation pore named Sec YEG, which is composed of membrane proteins. RNC binds upon the cytoplasmic side of the pore and ribosome synthesis and translocation are both achieved in the pore.

In the other type of Sec Translocation, the precursors has a lower hydrophobic signal peptide, so protein are fully synthetized and release by ribosome, and after transduction by ribosome, protein interacts with a protein called PiP (post translationally interacting protein) such as specific target chaperone SecB, the general chaperones GroEL e GroES, Dna K- Dna-J GrpE or the soluble Sec A, which are able to prevent protein from aggregation, to facilitate its export [57]. Protein is transported to the translocase pore and push through the SecYEG pore by the ATPase component Sec A. SecD and SecF exerts also a pushing effect on the translocation from the extracytosolic side of the membrane. At the end the signal peptide is removed by a specific signal peptidase and the mature protein is released into the periplasm, with the authentic N-terminal of the mature protein [58].

Since translocation mechanism is one of the major energy consuming processes, many studies have been conducted on the identification of energy source. The ATP, as energetic font, in the periplasm is absent because the IM is impermeable to most nucleotides, and it has been individuated that the ATP transport is mediated by specific apparatus. It has been individuated a membrane component that has a nucleotide binding sequence, another with the histidine transport system and the last cytoplasmic ATP binding and hydrolysis site [55].

6.1. Signal Peptides

Recombinant proteins, to be expressed in the periplasmic space, are fused to a N-terminal signal peptide, which consents to the protein to be secreted out of the bacterial cytoplasm in the periplasmic space. Type of signal peptide define the secretion pathway used [59]. Upon signal peptide removal, the protein then corresponds to its mature sequence. Moreover, the protein, present in mature form and protected from the intracellular proteases, avails of the oxidative and chaperone enriched environment can assume its native 3-D structure.

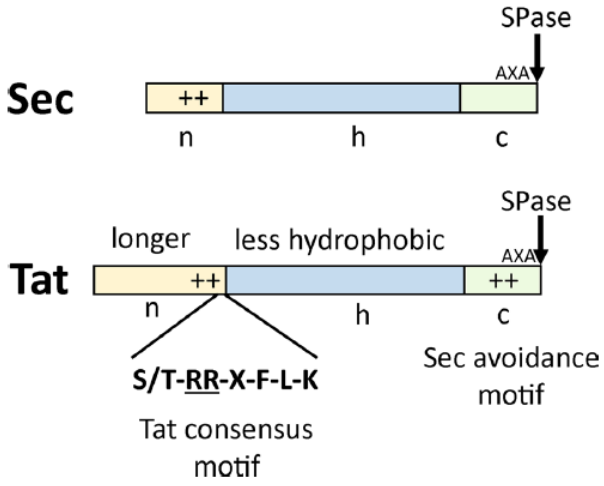


Figure 10- Differences in length and composition of the two categories of signal peptide. Sec signals differ from the Tat signal peptides in each component, length and amino acid compositions. Ref: Roland Freudl, Signal peptides for recombinant protein secretion in bacterial expression system, Microbial Cell Factories (2018) 17:52 **[61]**.

The signal peptide is a short amino-terminal part of the exported precursor proteins and direct the protein to one of the two export bacteria systems. The signal peptide is not unique and universal, and its efficiency will vary according to the nature of the recombinant protein [60].

Signal peptides vary in length and amino acid sequences, but in general Sec signal peptides have a tripartite structure, composed of a positively charge amino terminal region (n region), a hydrophobic core (h-region) and a polar carboxyl-terminal domain (C-term), that contains the signal peptidase recognition site (Ala-X-Ala) [60]. All three domains contribute to efficiently export the protein *in vivo*, in fact changing even only one of the specific characteristics, translocation capacity can be deeply reduced.

Also, the TAT signal sequences have a tripartite structure, but they are distinct by a conserved amino acid consensus motif S/T R-R-X-F-L-K, where X is often a polar amino acid residue [61].

Moreover, they have a distinctive codon usage, which includes the AAA (Lys) codon, in second position and a high frequence of non-optima codons; this strategy is used to slow down the folding rate in the cytoplasm and to ease translocation through the periplasm inner membrane. A requisite for correct leader sequence cleavage is that the signal peptidase finds at position -3 and -1 a in the C-region amino acids, that have neutral side chains.

Signal peptide choice is not obvious, and it is common to screen signal peptide libraries to select the best one giving high level of protein synthesis and secretion [62]. If for one protein, a signal peptide can give high level of protein synthesis and secretion, for another recombinant protein the same sequence will lead to low-level of protein production and/or secretion, if any [63]. For a new protein identifying the best available leader sequence, is a time-consuming and an expensive procedure, but some studies evidenced that some leader sequences are more processive than the others, and that some of them could be genetically modified to increase processivity [52]. Signal peptides have a significant impact on the translation initiation, a rate limitation step in the protein transfer.

As indicated in the gene design paragraph, the Translational Initiation Region (TIR) is composed of 30 nucleotides, spanning from the Shine-Dalgarno sequence down to the fifth protein codon (first ribosomal footprint) and, when a signal peptide is fused with protein, TIR is formed by fusing the vector 5' UTR with the signal peptide [60].

The usage of a specific signal peptide can reduce protein thermodynamic stabilization and activity, moreover, can affect the phenomena of protein aggregation. Signal peptides can affect the amount of protein of each precursor which is synthetized.

Accumulation of protein translocated into the periplasmic space, increase the osmotic pressure, rendering the OM leaky and prone to disruption.

Signal sequences	Amino acid sequences
PelB (pectate lyase B) from *Erwinia carotovora*	**MKY**LLPTAAAGLLLLAAQPAMA
OmpA (outer-membrane protein A)	**MKKT**AIAIAVALAGFATVAQA
StII (heat-stable enterotoxin 2)	**MKK**NIAFLLASMFVFSIATNAYA
Endoxylanase from *Bacillus* sp.	**MFKFKKK**FLVGLTAAFMSISMFSATASA
PhoA (alkaline phosphatase)	**MKQS**TIALALLPLLFTPVTKA
OmpF (outer-membrane protein F)	**MMKR**NILAVIVPALLVAGTANA
PhoE (outer-membrane pore protein E)	**MKKS**TLALVVMGIVASASVQA
MalE (maltose-binding protein)	**MKIKT**GARILALSALTTMMFSASALA
OmpC (outer-membrane protein C)	**MKVK**VLSLLVPALLVAGAANA
Lpp (murein lipoprotein)	**MKATK**LVLGAVILGSTLLAG
LamB (λ receptor protein)	**MMITLRK**LPLAVAVAAGVMSAQAMA
OmpT (protease VII)	**MRAK**LLGIVLTTPIAISSFA
LTB (heat-labile enterotoxin subunit B)	**MNKVK**CYVLFTALLSSLYAHG

Figure 11- Table with the most common signal sequences used for periplasmic expression. The amino acid sequences report the main leader sequence characteristics: N terminal in bold and C-term underlined. **Ref:** J. H. Choi & S. Y. Lee, Secretory and extracellular production of recombinant proteins using Escherichia coli, Applied Microbiology and Biotechnology, 2004, 64, 625-635. **[64]**

For *Escherichia coli*, some homologous and heterologous signal sequences have been individuated for the Sec pathways. The homologous are LamB, MalE, OmpA, OmpC, OmpF, Omp T, DsbA, SfmC, TolB, TorT and PhoA. Heterologous are: G1 from *B. pertussis*, SpA from *Streptococcus*, Pel B from *E. caratovora*, gene III from M13 bacteriophage [65]. Among the signal peptide for the TAT mechanism, 29 homologues signal sequences are known to work in *E. coli*, and the mostly used are the CueO and the Tor A. For TAT heterologous signal sequences are organophosphorus hydrolase (OPH) from *flavobacterium*, Cel9B from *R. albus*, Dag A from *S. coelicolor*.

Also, tom increase protein refolding efficiency in the periplasm some chaperones are necessary, they are involved in different stages of translocation and protein structuration. As describe before SecB is a chaperone use during translocation and SecD plays a role in the periplasm to release the translocated protein. SKP is another protein that seems to have characteristic similar to SecA. Many other “folding catalysts” have been individuated and play a role in the protein folding in the periplasm.

As we described in the Shuffle paragraph, DsbA is a protein involved in the disulfide bond formation, coupled with a DsbB which restore the first protein functionality. This mechanism plays a role also in the periplasm folding. Other cytosolic chaperones

represented by GroE, DnaK and CiPA, B and C, and this last family members are also present in the periplasmic space [71].

These complex environments do not always guarantee a perfect folding and translocation mechanism, so some studies are focused on individuating the fate of the unfolded protein. The unfolded protein seems to be controlled by a heat shocked protein (HSP), which co-localized with the mismatch's one. This HSP resist to temperature and act as protease or degradase. The one present in the periplasm has been individuated as the DegP. The action of the periplasmic proteases is potentiated by the present of DegP, and their genes are often deleted like OmpT and OmpP. Sometimes the precipitation of unfolded protein in the periplasm generates inclusion bodies, but this structure is less recognizable for their strict association with the membrane and are called "Membrane associated", or periplasmically precipitates forms[71].

Limitations linked to the periplasmic expression are essentially three: periplasmic space is limited, the secretion apparatus could be overloaded and, at last, protein that could be naturally folded in the periplasm should have small number of disulfide bridges and simple disulfide patterns, because often Dsb system is in troubles when protein contains large numbers of disulfide bridges or complex patterns.

6.2. Protein Extraction from Periplasm

The tricky point of the periplasmic expression is not only the balance between the production capacity and the efficiency of the transport system, but also resid in the possibility to selectively extract the recombinant protein from this compartment. This to obtain a much purer extract, compared to whole cell lysis, where HCP, LAL and DNA are released [57].

The extraction is facilitated because only the outer membrane must be disrupted, and the periplasm contain minimum number of proteins and components, reducing purification steps. To the best of our knowledge, no generic protocol for selective membranes disruption has been up to now developed. Many researchers have unsuccessfully tried to set-up a specific protocol for periplasmic proteins extraction, but selective Outer Membrane permeabilization without affecting the Inner one integrity, is not an easy task [58].

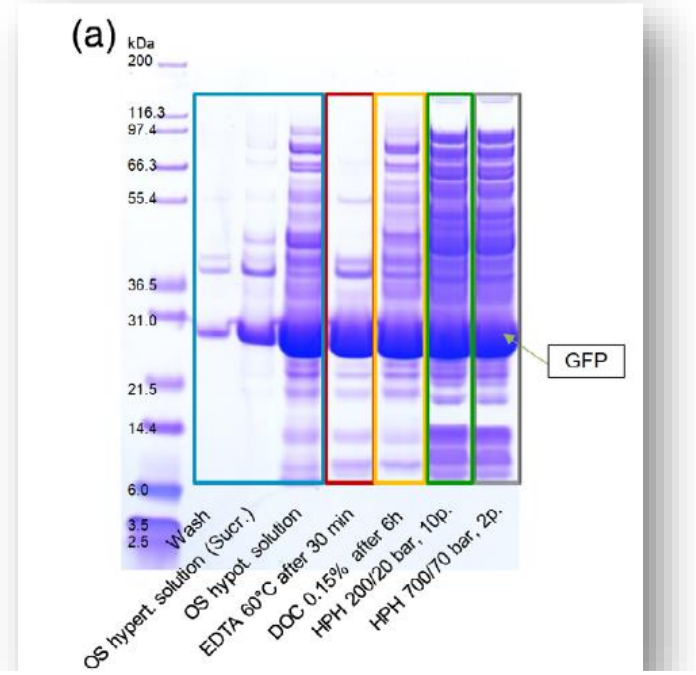


Figure 12- SDS-PAGE shows different extraction treatments and the protein of interest Green Fluorescence Protein (GFP) with the impurities. As it is possible to observe the procedure with EDTA and temperature is the best one, according to the low presence of other bands. The harder procedure like Homogenization (HPH) both with high and low pressure, recover material massively contaminated by HCP. **Ref:** Schimek et all. Extraction of recombinant periplasmic proteins under industrially relevant process conditions: selectively and yield strongly depend on protein titer and methodology. Biotechnology Progress 1 April 2020, DOI:10.1002. **[58]**

With the purpose of destabilizing OM, a variety of methods have been proposed based on physical forces (shear, heat, or osmotic pressure) or through the chemical removal of membranes components.

The selective release of periplasmic proteins, during fermentation or after biomass harvest, combining different treatments, must be performed without lysing the cells. This is to keep contaminants like HCP, DNA and endotoxin, together with the biomass, and not with the extracted product [58].

So, it is a tight balance between membranes selective disruption and extracted recombinant protein yield. Often mechanical, physical and chemical treatments are used. Mechanical treatments are usually sonication, and homogenizers, which can be based on high pressure systems or high-sharing forces (Turrax); physical treatments intend high temperature applied to the sample, osmotic choc; and chemical treatments using permeabilizing agents, such as guanidine, Triton-X, Glycerol, urea, Polyethyleneimines (PEI), glycine, etc. In literature those strategies have been described, and compared, in several articles, but often with contrasting conclusions. Sometimes the treatment is applied along the fermentation processes or at its end, according to the procedure efficacy and intensity [53].

Starting from the analysis of the mechanical methods, the most common used is the high-pressure homogenization, a fast procedure but leading to total cell disruption, without any selectivity. Usage of some instruments, like Turrax, with high shearing forces, and more adapt for eukaryotic cells, is not selective either. The current industrial standard procedure for cell disruption is high-pressure homogenization (HPH), which is efficient, easy to develop, and straightforward to scale up. Number of cycles, or pressure level can be adjusted but again selectivity is not contemplated.

Then the physical methods like freeze-and-thaw treatments can be successful, as the sole cell disintegration method, or to enhance the performance of another method. Also, temperature treatment during fermentation, or at its end just before harvest, seems not to induce Outer Membrane fragmentation, but can be really efficient to release some small and highly soluble proteins present in the periplasm.

Osmotic shock (OS) procedure relies on cells transfer from high osmotic buffers into solutions, having very low solute concentration. This causes a periplasmatic wash, with

solutions coming from cells cytoplasm and passing through the periplasm, rinsing it and extracting therefore small and highly soluble proteins. Osmotic Shock methods are widely used in laboratory scale for “selective” periplasm disruption, but it is not a scalable procedure, due to the very high buffers volumes required.

Another laboratory protocols use permeabilization agents and a short description list of the mostly used procedures is reported in the next paragraph [66]. The agents are chosen according to the removability, toxicity and price, in addition to their selectivity [53] .

EDTA is the most often used reagent, this metal-chelating agent, binds divalent cations (mainly Mg^{2+} and Ca^{2+}), which are present in the membranes [67]. Its addition consequently leads to a rapid permeability change, and to an OM asymmetry, that provokes its disruption. EDTA addition provokes 30–50% lipopolysaccharide component loss, whereas the cell structure morphology remains intact.

Periplasmic release can also be induced by different chemicals agents, like detergents, amino acids, polyols, chaotropic salts, and solvents [68]. Utilization of these reagents is a balance between cell lysis, presence of contaminants (HCP, DNA and Debris) and scalability in industrial plants.

Sodium deoxycholate (DOC) at a concentration of 0.1% can be highly efficient and selective for some periplasmic proteins and can result in low levels of host-cell proteins (HCP). In literature glycol and surfactants treatments are also described. Other treatments, such as glycine or chloroform can be used to permeabilize the cells.

Lysozyme addition increases efficiency but documentation of its efficient remotion in the downstream process is then necessary.

Non-destructive permeabilization methods, with selective periplasmic extraction, often corresponds to periplasmic protein enrichment but not to real selectivity. Those treatments can also be combined. For example, heat treatment combined with the addition of a permeabilization agent, like EDTA releases nearly 30% of the periplasmic content, while keeping cell lysis at a minimum.

Furthermore, the combination of EDTA/heat are also potentially applicable on a large scale. However, the use of other chemicals is restricted due to toxicity and environmental issues, and applications must be evaluated on a case-by-case basis.

Another combined procedure is represented by 350 mM TRIS treatment before applying a mild heat treatment at 20°C which is gradually increased to 38°C. In this case around 30% of periplasmic proteins are released and only 5% of the cells lysed. Compared to the trigger leakiness induced with only lysis buffer, at the same Tris concentration, there is no increase of efficacy [58].

Evaluation of the best extraction strategy must not only be based on the amount of recovered recombinant protein but also on the contaminants present. This is a difficult task, as some components will be deleterious in the downstream process. A study reveals that DNA, endotoxin and Host Cell Protein contents are different according to the strategies used. For example, HPH releases the highest protein concentration compared to the other methods, but HCP contamination is superior to with OS, EDTA e DOC. Endotoxin level is always elevated independently of the applied method because this type of contaminant is in the OM, so only methods that did not cause external membrane disruption can prevent endotoxin release. Therefore, osmotic shock seems to be the best procedure for DNA reduction but only applicable on lab scale. EDTA extraction seems to be a valid extraction method because cell integrity can be preserved with low contaminants liberation, but this protocol only releases part of the periplasmic content depending on the recombinant protein size and solubility.

In conclusion, in the literature a variety of different strategies to selectively release the recombinant protein from the periplasmic space are described, but when results are carefully compared, none of those are selective [57] and always induce some cell lysis, furthermore yields are low and Good Manufacturing Practices (GMP) scaling up uncertain.

7. Shuffle Expression

Most human proteins are characterized by the presence of post translational modifications and disulfide bridges, so Biotechnologists need to deeply understand how and where these enzymatic reactions take place in *Escherichia coli*.

Analysis of the human genome revealed that 30% of the proteins are predicted to be targeted to the endoplasmic reticulum (ER), where disulfide bonds can form [74]. These modifications are essential for refolding and functionality of the human protein [76]. Therefore, when a new process is designed to express heterologous recombinant proteins in *E. coli*, it is important to take into consideration disulfide bonds limitation at prior. Two traditional strategies are used for the Disulphide bonds formation: the extraction of inclusion bodies and re-naturation *in vitro* or translocation in the oxidative periplasm.

In the first case, the success of the refolding procedure is protein dependent, and the scale up is also complicated. In the case of periplasmic expression three limitations can occur during the production: overload of secretion apparatus, periplasmic volume insufficient for protein localization and to fold in the periplasm protein must have a simple disulphide pattern [76].

Production of disulphide bonded proteins within the model organism *E. coli* offers a fast, genetically versatile platform with a multitude of options, for improving the yield of protein production and purification. The modified cytoplasm has enough space to encompass two of the three limitations offers by the periplasmic expression.

The power of *E. coli* genetics permitted to develop various genetically engineered strains resulting in novel expression strategies, especially the Shuffle one which is capable to produce multi disulfide bonded proteins in its cytoplasm.

7.1. Disulphide bridges: Structure and function

Protein disulfide bridges are formed by an oxidative reaction, in which the thiol group of the two cysteines are linked by a covalent bond. The disulfide bonds are formed co-translationally or post-translationally by the oxidation of the thiol group of cysteines. This bind occurs during folding and the pair of cysteine involved could be close to each other in the same chain or between two polypeptides, resulting in intra or inter-chain disulfide bonds. This type of link is however reactive and easily reversible, they also isomerize by thiol:disulfide exchange reactions, spontaneously or catalyzed by an enzyme. In presence of an electron donor, such as dithiothreitol (DTT) or disulfide bond reductase as Thioredoxin, disulfide bonds are ready reduced back to their thiol state.

The S-S structure gives more stability to protein and contributes to achieving the 3-dimensional structure required for full protein activity. Many of the enzymes that play a role in S-S formation have been identified, as well as have been understood the mechanisms that undergo for the bridge's breaking and shuffling.

As previously indicated, these mechanisms occur in organelles, as endoplasmic reticulum, in eukaryotic cells, or in the Prokaryotes periplasmic space. In Eukaryotes the formation of disulfide bonds isomerase (PDI) catalyses both the oxidation and the isomerization of disulfide bonds, and the reduced PDI is reactivated by Ero1a, protein membrane. In prokaryotes, the formation of disulfide bonds is compartmentalized in periplasm and in the *E. coli* cytoplasm, disulfide bridges are reduced by two systems: thioredoxins (trxA/trxC), kept in the reduced state by the thioredoxin reductase trxB; and the reduced glutathione (GSH), kept in the reduced state by the action of glutathione reductase (gor). Whereas in the periplasmic space disulfide bonds are oxidized by DsbA system.

The Dsb system, composed of two pathways: one DSbA-DsbB is responsible for the oxidation and DsbC-DsbD is responsible for the reduction and isomerization of wrong mis-oxidized forms. In brief, the Dsb A-Dsb B pair form disulfides during protein translocation, while Dsb C -Dsb D pair use electrons to transfer through the inner membrane to isomerize mismatched disulfide bridges. DsBA is an oxidase and has been the first enzyme to be discovered, successively others have been identified.

The interaction between DsbA and the cysteines, in a reductive state, can occur post-translationally or co-translationally on the protein. In the first case, the protein is already refolded and in the second one, the bond is created when the protein acquires its secondary structure. DsbA has no substrate specificity and simply reacts with the first encountered cysteine of the polypeptide chain and creates a disulfide bond with the next one in the primary protein sequence. So DsbA disulfide bonds catalyzation is a linear oxidative folding mechanism. The DsbA active state is restored by DsbB, an inner membrane protein, that transfer electrons from the respiratory chain [71].

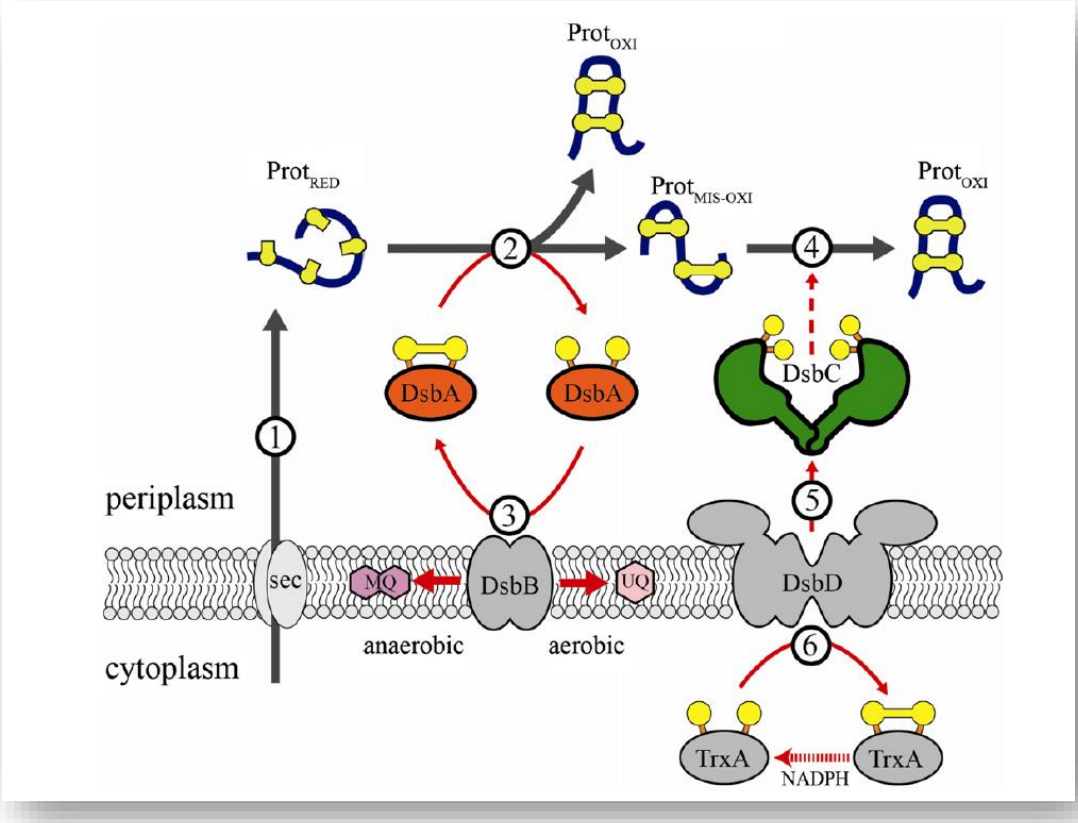


Figure 13- Periplasmic disulfide bond formation picture from review [72]. It is illustrated the Sec translocation of protein in the periplasm combined with the different stages of OX and Red and the energy request for disulphide bridge formation.

DsbA, belongs to the Txr fold superfamily, and is a soluble monomeric protein, translocated along the by Sec pathways in the periplasm. The main characteristic of DsbA is the presence of the Cys-X-X-Cys motif and, among other proteins belonging to the same family, it is the one with the highest redox potential [69].

The DsbA active site promotes its binding to unfolded substrates through hydrophobic interactions. Neither DsbA, nor its partner DsbB, are essential for *E. coli* aerobic growth. DsbA catalytic Cystines 30 and 33, initially forming a disulfide bonds become reduced during the catalysis with the substrate, and will be reoxidated by DsbB, which has an oxidase activity and is capable to transfer the electron pair extracted from DsbA to the electron chain via quinones to ubiquinone during aerobic growth or to menaquinone in absence of oxygen [70].

Among the periplasmic proteins involved in the S-S formation, DsbC is involved in mismatched disulfides isomerization and protects the unpaired cysteines. As previously indicated, the catalyzation of the disulfide bonds formation is a linear operation between two closed cysteine, without specificity. The isomerase operates to correct the errors generated by DsbA.

DsbC rearranges the disulfide bonds, promoting a correct final folded protein. If proteins present distant cysteines in their primary sequences, the DsbC role of is more important in protein structure organization. So amino acids distance defines the dependence of DsbC action, during S-S bridges formation. The isomerization procedure seems not to result in any net gain or loss of electrons and at the end of the reaction Dsb C remains in its reduced form. But DsbD may maintain Dsb C in its reduced state, to perform its action through electron-neutral isomerization. Among the Dsb proteins, DsbC is the most versatile member because it is not only involved in the isomerization, but it has been indicated that it can efficiently reduce protein like thioredoxin *in vitro*. It has been identified that the DsbA/B oxidative and DsbC/D reductive pathways could not be separated within the periplasmic compartment and that DsbA needs to be maintained in an oxidized state and DsbC in its reduced one.

Another element to considered in the periplasm redox system is the glutathione (GSH). This tripeptide is exported in the periplasm, and seems to improve Dsb C and D actions, but inhibit DsbA oxidase activity.

Knowledge of the different pathways involved in disulphide bridge formation illustrated above, has permitted the engineering of *Escherichia coli* strains able to form those structures directly in their cytoplasms opening the door to easier heterologous disulphide containing proteins expression.

7.2. Engineering *E. coli* for S-S bonds formation

High yields productions of soluble and active disulfide-bonded proteins in *E. coli* is a challenge, as usually overexpression of heterologous protein in its cytoplasm leads to IBs formation and that protein translocation into the periplasm, for the S-S bond formation, is not an easy task.

The major biopharmaceutical products are represented by monoclonal antibodies, which for example have numerous disulfide bridges, essential for their structures and functions. Most of them are produced in Eukaryotic (mainly mouse) hybridoma, but the major cause of heterogeneity is represented by disulfide mismatches. Antibodies have anyway complex structures and there is an increasing necessity to define a better solution to solve this problem.

Currently, several expression systems are available to express complex multi-disulfide-bonded proteins, as eukaryotic expression ones (Chinese Hamster Ovary, CHO), yeast or insect cells, but these systems are limited by their slow growth and expensive media [71]. For most applications, prokaryotic expression remains the most attractive (see *E. coli* paragraph), but it is still necessary to develop, or test, new genetic tools optimized for disulfide bonds formation.

The first group who developed a strain, competent for the S-S formation, was Beckwith and co-workers, which studied *E. coli* redox pathways and the involved enzymes [71]. Then, they proceed with the *E. coli* strain engineering, to obtain a mutant capable of promoting disulfide bonds formation directly in its cytoplasm. Two are the reductive pathways that render not friendly the cytoplasm environments to the disulfide bonds formation: the glutaredoxin and the thioredoxin pathways. Moreover, the presence of the thiol-disulfide redox buffer (up to 5mM of reduced glutathione) and also the cytosolic pool of NADPH+ supplies the reducing potential for the two pathways.

In 1999, the *E. coli* trxB gor suppressor has been “the” initial strain, for starting the engineering of this special *Escherichia coli*. Engineering by knocking out the trxB and gor pathways, resulting in decoupling the connection between the cytosolic pool of NADP+ and the disulfide bond reductases. But selection of suppressors of trxB and gor lethality resulted in conversion of the peroxidase AhpC to the disulfide bond reductase AhpC*.

The reduction of GrxA by AhpC* results in sufficient reducing capacity of the cells to permit their growth, while the thioredoxin remain in their oxidated state. The oxidated trxA cannot catalyze the formation of disulfide bonds in protein that required it for the S-S formation. The name of this engineered strain with all three mutations is Origami and is commercialized by Novagen [71].

Furthermore, in the Shuffle cells, cytoplasmic DsbC co-expression has been added to improve protein folding. The DsbC gene was cloned, under the relatively strong and highly regulated rRNA promoter. The addition of DsbC is important because all non-consecutive disulfide bonded proteins and all non-cytoplasmic proteins required the corrective capacity of DsbC the be correctly folded. Shuffle strain represents the first example of Dsb system engineered for industrial applications. *In vitro* DsbC seems to be more effective than only isomerization as it can also assist protein folding, in a redox independent manner, but this still needs to be confirmed *in vivo*.

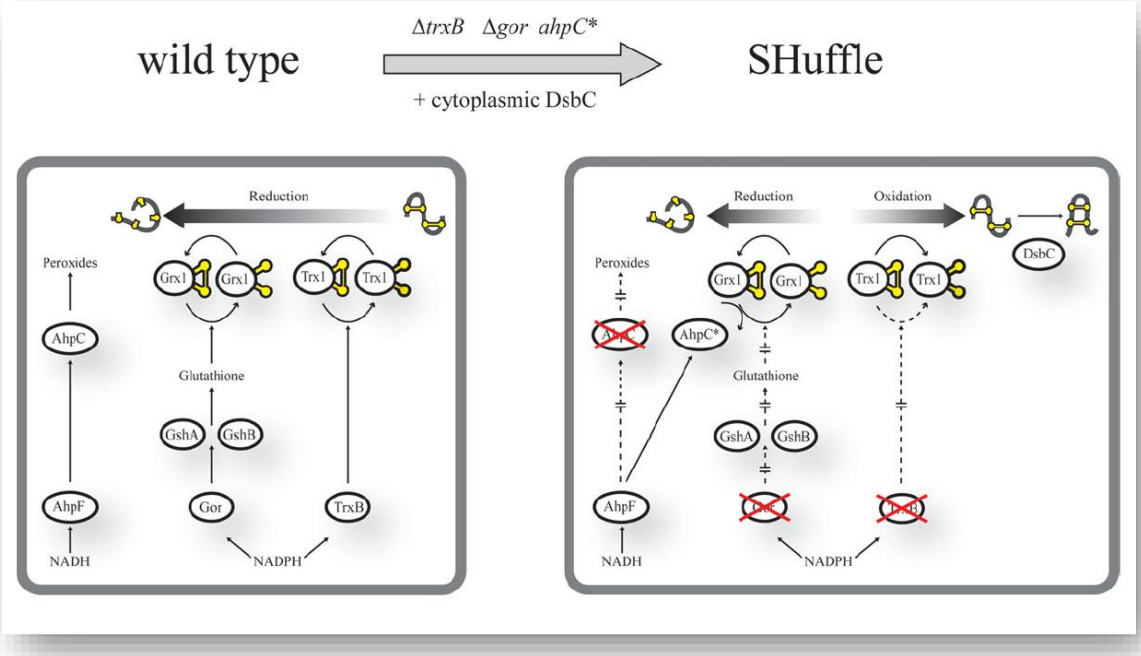


Figure 14- Pathways of WT cytoplasm vs Shuffle cytoplasm. As it can be observed, the deletion of trxB, gor and ahpC gene and the addition of DsbC gene, give a cytoplasmic redox pathway in the Shuffle strain. **Ref:** G.Ren, Na Ke and Mehmet Berkmen, Use of the Shuffle Strain in the Production of Proteins, Current Protocols in Protein Science, 5.26.1-5.26.21, August 2016. **[71]**.

Disulfide protein studies have also permitted to develop treatments, for clinically relevant strains such as *Mycobacterium, Pseudomonas and Helicobacter,* to design new specific enzymes inhibitors and block in those strains the enzymes catalyzing the formation S-S bridges, to compensate for the worldwide emerging antibiotic resistance.

Some strains have modified DsbG, which has 45 % homology with DsbC, and *in vitro* has a chaperone effect together with its isomerization role, even if its primary enzymatic activity is a reductase, restoring cysteines that have been oxidized. Due to the differences in specificities, DsbG when highly expressed, helps DsbC when the folding activity fails.

To enhance their activities, or to generate fusion partners, some modifications have also been introduced in the oxidoreductase structure. Some expression vectors permit to fuse the gene of interest with DsbA (pET 39), DsbC (pET 40) or TrxA (pET 32) and are commercialized by Novagen.

Another pathway to maintain a reducing cytoplasm is the glutaredoxin one. Glutathione reductase reduces two enzymes, GshA and GshB which are in turn responsible for synthesizing reduced glutathione (GSH). Glutathione is a small tripeptide whose primary role is to act as the major component of the cytoplasmic redox buffer. Depending on the GSH concentration, it could be in its reduced form, or in its oxidated one GSSG. GSH can reduce disulphide bonded proteins.

In literature, one can find examples of Shuffle strain usage as host for the expression of many recombinant proteins of interest. Most of the principles described above for recombinant proteins expression are still applicable to Shuffle, but some further crucial parameters have been foresighted. The first choice is between the B or K12 engineered versions of the shuffle strategy, then which culture temperature will be used, as the strain is sensible to 42°C and above [74].

Moreover, if the protein to be expressed represents a high metabolic burden for the host, because it contains a several S-S bridges, it is suggested to use temperatures inferior to 37°C. Suggested IPTG concentration and induction time are respectively 0.1mM and 5 hours at maximum or overnight according to the temperature used [70].

It is also suggested that in order to increase recombinant protein refolding yield addition of a chaperone (GroEL/ES, DnaK/J) to assist folding is advised. In the constant oxidative stress, in which the Shuffle strain survives, cells are more sensitive to the media dissolved

oxygen levels. So, it is important to monitor this parameter together with the small redox-active molecules present in the complex rich medium (glutathione, cysteine, cystine from yeast extract) which could act as buffer in order to reach high biomass and protein productions yield [72].

Productions of high yields of correctly disulfide bonded proteins in *Escherichia coli* remains a challenge, even if there is a growing demand for new expression strains, few of them have been developed. Thus, new *E. coli* strains, such as Shuffle or Origami, have been engineered to improve the production and folding of disulfide bonded active proteins in their cytoplasms.

8. Biotech Production Process

The advancement of genetic engineering technology, research in protein drug achieved significant breakthroughs. Recombinant proteins have also economical values, corresponding to their pharmaceutical growing request and market sales relevance.

But, recombinant proteins production is not an easy task, the genetic construct design and expression system definition, represent only the project starting points. The process of recombinant protein production is a structured workflow composed of Upstream and Downstream phases. The Upstream aims to produce high cell density cultures (HCDC), able to express the maximum of recombinant protein and the downstream aim is to specifically isolate "this" protein of interest, from the contaminants and host cell ones and to formulate it [73].

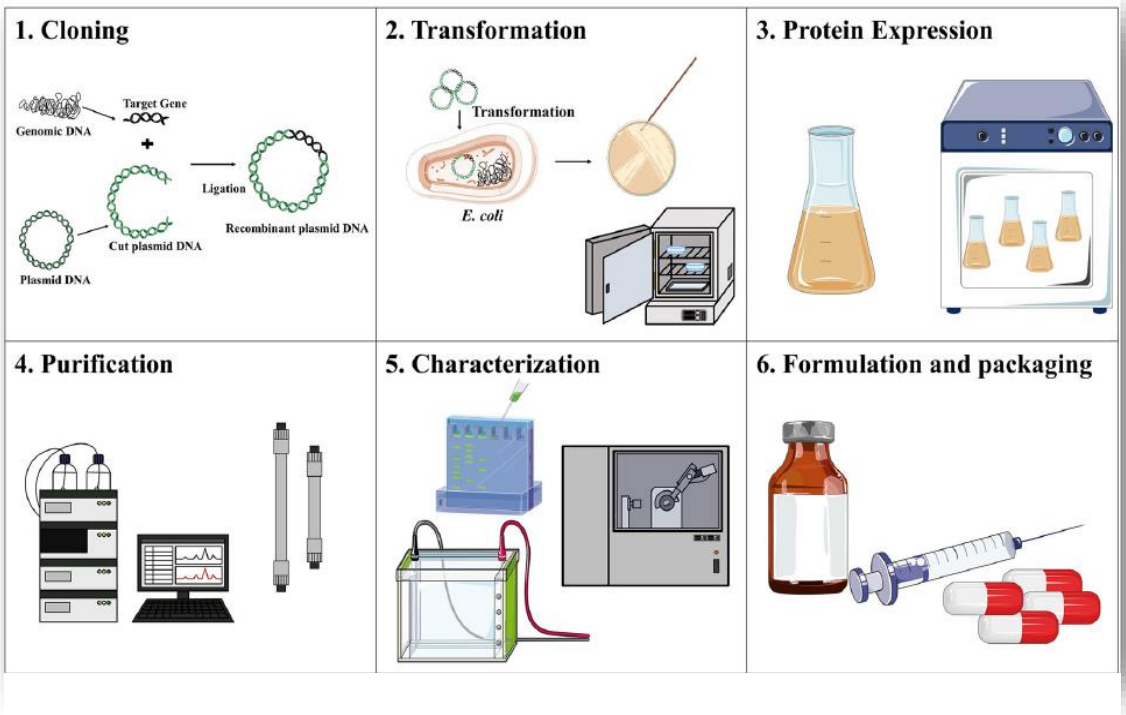


Figure 15- Recombinant Production Process: in the first lane of the figure are reported all the Upstream Passages, and in the second one all the downstream purification steps. **Ref:** Ibrahim Incir and Ozlem Kaplan E. coli in the production of biopharmaceuticals. Wiley, DOI: 10.1002/bab. 2664. **[36]**

Recombinant protein production relies on the upstream process, which consist in the development of a particular cell line, able to overexpress the recombinant protein and to develop the best culture protocol in bioreactor, a close perfectly controlled system.

As previously described these processes often involve genetic engineering, microbial culture knowledge, identification of the best host cells and their production capabilities. Progress in protein expression system has enabled the synthesis of many desired protein or peptide, and the advancements empowered the creation of unprecedent molecules, that diverge from the natural ones, holding a potential in enhancing drug efficiency.

Upstream process also includes extensive development procedures and optimization to maximize protein yield, purity, and quality. The choice of the culture strategy has a significant influence on the success of protein expression and in easier downstream processing [50].

Downstream process is a crucial phase in biologics manufacturing, because once recombinantly produced proteins needs to be purified to homogeneity and formulated into a final product, passing through several steps, ensuring production of high-quality biopharmaceuticals ready for clinical trials, and once approved for patients' therapies.

Purification consists of all the procedures employed to purify the drug substance, gradually removing impurities and contaminants, to meet stringent pre-established quality standards. Techniques involved in this process include chromatography, filtration, and centrifugation. Formulation is the downstream last phase, determining final product composition with excipients to enhance its stability, solubility, and shelf-life. Final formulation buffer composition is established on the biologic molecule's properties to render them stable along storage and easy to handle for patient.

According to the regulatory authorities, all production processes must be validated, and the active pharmaceutical product needs to be deeply characterized, using validated analytical methods, before commercialization. Also, after introduction on the market the pharmacovigilance will assure continuous drug monitoring.
So, during the fermentation and the purification process development, critical process parameters (CPP) must be individuated, and validated, to define the process, scale it up in a GMP authorized plant.

Moreover, sampling plan must be defined and agreed with the regulatory authorities, to constantly monitor criticalities, during process and keep it under control.

During GMP manufacturing, the analytical Quality Control department follows the production process, analysing the identified and critical In Process Control samples (IPCs) and, at the end of the process will analyse the Active Pharmaceutical Ingredients (API) bulk with the complete preestablish analytical panel, to allow the Qualified Person (QP) to release it. The analytical panel certifies the purity, identity and activity (potency) of the produced protein. The results must comply with the pre-defined acceptance criteria, which permits to release the protein batch in accordance with the predefined range approved and/or imposed by the Regulatory authorities (ICH, GMP, CFR21, etc.) [74].

8.1. Upstream

Genetically Modified Microorganisms (MOGM) is a generic term, that indicates a genetic modification of the expression system. In our specific case it is composed of a host, often *E. coli* strain, already heavily engineered, and transformed with an expression vector bearing a DNA fragment codifying for a human protein.

Fermentation is the heart of the Upstream, which represent the initial phase of biopharmaceutical productions. Then harvest and washes represent the last upstream process passages. In the last 20 years many progresses have been achieved in optimization of upstream processing (USP), leading to high cell densities and high product titters. It has been identified that Upstream titters depend mostly on biological limits (e.g., cell line or media optimization) and can be optimized to decrease cost of goods.

Pharmaceutical fermentations, also called precision fermentations, are complex processes, that take place in a Bioreactor. All the procedure must be conducted in sterile conditions, from vessel preparation to the cell's inoculum. Along the fermentation process all parameters need to be controlled, to have a successful protein production and to avoid batches variability and deviations from the standard batch record.

The main fermenter-controlled parameters are temperature, pH, Dissolved Oxygen (DO), nutrient concentration, stirrer, air, etc., They are precisely controlled and adjusted according to the microorganism's metabolism. The goals of fermentation development are high volumetric productivity, high selectivity and low raw materials costs. To do so, the strategy is often to reach high cell densities, try to maintain high growth rates to assure high levels of product accumulation.

Media preparation, that represent the first step in upstream process, is not an obvious procedure. The culture media must provide all the necessary nutrients for cell growth and product formation. Culture medium is usually composed of carbohydrates as carbon source, amino acids, vitamins, minerals and growth factors, in the case of Eukaryote cells [75]. During the fed batch phase, the selection of nutrients in the "feed" is critical because it can negatively influence the metabolic pathways fluxes and consequently afflict the maximum attainable cell concentration, the specific productivity of the recombinant proteins and induce the formation of by-products [39].

Media sterilization is also crucial to prevent contaminations. In stainless-steel fermenters sterilization is performed in place (SIP) and in single-used or glass ones, is performed by filtration or by autoclaving the medium before its transfer into the bioreactor [76].

The next fermentation step corresponds to cell inoculation (RCB, MCB or WCB). Inoculum volume will be determined to have the exact concentration of cells required to seed the fermenter. This step often requires one or more pre-culture passages, to reach a sufficient number of viable cells. In the bioreactors, cells are cultivated under controlled conditions to maximize their growth. There are different kinds of fermentation processes: continuous, batch, fed batch, etc. All culture processes below described take place in so-called fermenters, also known as bioreactors. These are specially developed vessels, in glass, steel or single-use plastic bags, in which optimal mixing conditions are provided for microorganism growth. The batch, fed-batch, and continuous or perfusion modes of cultivation are used to produce recombinant protein-based biopharmaceuticals.

In a batch fermentation, nutrients are all provided in the initial medium. In a fed-batch process, the first part of the culture corresponds to a batch fermentation then further nutrients are fed during cultivation, in certain mode and at a controlled rate. In perfusion cultures, the medium is circulated through the growing fermenter, and exchanged through a membrane, to allow simultaneous exhausted medium removal and new nutrients supply. Instead, in a continuous or chemostat bioreactor culture, when the medium is fed in, the same amount is pumped out and contains the product that is continuously harvested in the outcoming solution. If the desired rate of dilution is less than the cells growth rate, this growth needs to be controlled using a turbidostatic or chemostat culture. However, if the rate of dilution is more than the cells growth rate, cells need to be returned to the bioreactor [39].

The fermentation process begins with the assembly of raw materials in the culture medium, typically sugar serves as nutrient for recombinant microorganisms growth and heterologous protein production.

The process starts from inoculation, with sometimes an adaptation period (lag) of the microorganism to the culture conditions, then the growth phase, in which the microorganisms exponentially duplicate and metabolize the raw materials.

The exponential phase ends when one medium component, as for example sugar or oxygen, become limiting. The first fermentation phase is usually dedicated to biomass growth, then the fed-batch one to the recombinant protein production by addition of an inducer into the medium, to promotes its gene transcription. When microorganism starts to produce the protein of interest it's metabolic load is so high that it stops growing and biomass reaches the stationary phase. The fermentation key parameters must be meticulously monitored and adjusted to optimize cell growth and productivity. Various nutrient feeding strategies have been developed for high cell density cultures in *E. coli* such as constant, linear, or exponential feedings.

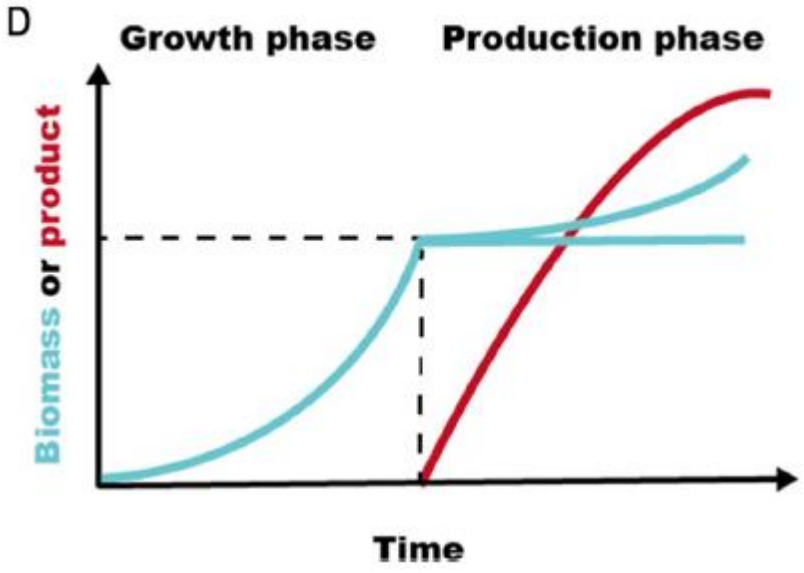


Figure 16- Description of growth phase and production phase. That graph shows the uncoupling of the fermentation phases, and the clear division between biomass growth and protein production. Ref: Zi-Xu Zhang, fang-Tong Nong, Yu-Zhou Wang, Chun-Xiao Yan, Yang Gu, Ping Song and Xiao-Man Sun, Strategies for efficient production of recombinant proteins in *Escherichia coli*: Alleviating the host burden and enhancing protein activity, Microbial Cell Factories 2022. **[56]**

The exponential feeding strategy compensates the still growing number of cells present in the fermenter and allows, in most of the cases, maximal growth at a constant specific rate driven by the glucose addition one.

Substrate (glucose) is given in a limited amount and, due to the high consumption by the microorganism its concentration in the vessel is always close to zero with minimal fluctuation in the culture during feeding, without perturbing the carbon metabolism [77].

Other feedback controls are pH-stat [78] and DO-stat, in which when the principal carbon source is depleted, pH and DO values changes, and trigger new medium addition into the fermenter. These strategies prevent the microorganism from nutrients overfeeding. As well, an optimal post-induction feeding strategy is important for the efficient production of recombinant protein to always maintain a good medium oxygenation.

The recombinant protein productivity equally varies with the *E. coli* host and type of recombinant protein. It has been for example reported that the *E. coli K12* strain produce more acetic acid, rather than the *E. coli* B strain, so this characteristic must be considered when High Cell Density Culture (HCDC) will be required. In general, the maximum of HCDC is obtained with *E. coli* when growth and production phases are separated [56].

In fed-batch phase for example, the induction is delayed obtaining a maximum of biomass before turning on recombinant protein production. This is especially implemented when the recombinant product is toxic for the host. Sometimes high inducer concentrations could be a growth inhibitor. So, optimal induction time and inducer concentration are important to increase the overall recombinant protein production. The most suitable time for induction is around the mid-log phase when cells density is already important, and that one has sufficient remaining growth window to induce the culture, before lysis. There is a correlation between cell density and protein synthesis.

In general, cultivation conditions are developed, using the traditional “one factor at a time” (OFAT) approach. This is essential to optimize culture conditions for a specific cell line expressing a specific protein in specific conditions. Up to now several programs as Minitab, using the Design of Experiments (DoE) approach can reduces the number of laboratory trials to be performed, identifying the most critical parameters to be investigated [65]. Also, artificial intelligence, can predict the fermentation outcome, according to mathematics formulae, and establish a correlation between growth curve and final protein yield. Advanced bioreactors are equipped with automated systems to maintain the fermentation conditions, ensuring consistent and reproducible results.

To develop a more efficient fermentation process using *E. coli* it is necessary to deeply understand the cellular response and physiological changes. The metabolomic and transcriptome knowledge provide information on how metabolic pathways and regulatory circuits could be optimized to increase recombinant protein production. For example, it has been observed that enzymes, involved in the amino acid synthesis, are down regulated when HCDC is reached, explaining why in some cases the recombinant protein productivity can be impaired.

At the end of the fermentation process, biomass is harvested, and medium separated from the cells. Harvesting is the last upstream process procedure, characterized by the cells collection. Harvesting can be done through centrifugation, filtration or sedimentation, according to the protein localization and the equipment present in the plant. The method choice depends on the type of cells and product nature.

Harvesting efficiency directly impacts final product yield and quality, making it a critical step for the next downstream process. However, several types of centrifuges are available according to the volume to be processed, batch centrifuges for lab scale and continuous one for the industrial scale. Obviously recovered material quality will depend on the type of instrument use and batch centrifuges will always give a better separation.

Impurities derived from the cell culture can trigger potential adverse reaction or allergic responses. Immune responses to these impurities can result in the loss of tolerance to their endogenous counterparts, leading to serious adverse events. The ICH Q6B (Test procedures and acceptance criteria for biotechnological/biological products) guidelines set out the principles for product quality specifications.

8.2. Downstream

Biological macromolecules, such as proteins, constitute an important class of products in the food, biotechnology, pharmaceutical, and cosmetics industries. The growing interested on biotech drugs development, led to research and development of new efficient and rapid protein purification methods.

Methods have been elaborated, and more techniques developed, but protein purification costs remain extremely high, and constitute more than half of the overall production costs. Optimization of experimental purification conditions have become critical for the manufacturing industry, and containing production costs while satisfying all regulatory requirements, represents a challenge. Purification of high protein quantities, in the lowest number of passages, remains the best compromise between purity, yield and costs.

Protein structure and function are the main critical factors to consider, to define the most appropriate purification strategy. The chemical, biochemical or biological activity of a recombinant protein are determined by specific domains, within the protein, which are formed during the folding into its secondary, tertiary, and quaternary structures.

Protein folding is the mechanism by which proteins acquired their higher-order structures (HOS), referred to its essential correct three-dimensional shape and perform its function. Another characteristic to consider along the purification, is the protein solubility, which is influenced by numerous factors, including its overall size and the elements present at the N- and C-terminal ends. If protein has N- and C-terminal tags, detection and purification are facilitated; in these cases, immunohistochemistry or affinity chromatography could be used to isolate the protein of interest.

Researchers have now numerous methods, reagents, and tools available to increase purification yields and reduce the volumes of industrial-scale biotechnology and pharmaceutical productions [79].

In details, purification corresponds to all the downstream steps, involved in protein isolation and contaminants removal, deriving from the host (host cell proteins, endotoxins, DNA) and the product-related forms. General rules are to apply a different technique at every passage, exploiting different protein characteristic (dimension, charge, hydrophobicity, ligand-binding), minimizing the manipulation steps, additives, and purification passages.

If at each step product purity increases, on the other hand there is a product loss, due to the purification technique. Simple, one-step separation protocols are ideal situations but rarely encountered and will hardly comply with all the regulatory requests.

The most common methods used in downstream purification are centrifugation, precipitation, chromatography, and filtration. Centrifugation can separate proteins according to their solubility, size and density, according to the different sedimentation time. It is a non-specific procedure and is often perform in the initial steps of purification process.

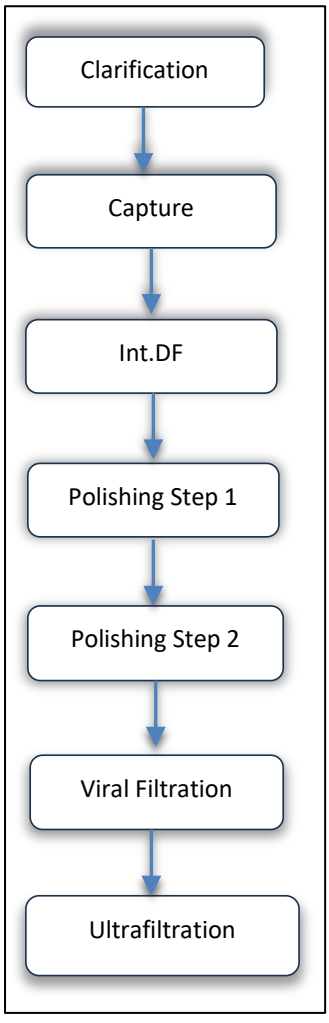


Figure 17- Flowchart of the principal downstream purification steps that usually have been performed to maximize the final product purity and safety.

Filtration can eliminate more significant contaminants, using filters with different cut off size. Depth filters are unit composed of porous filtration medium able to retain particles throughout the medium, rather than just on the surface of the medium. Depth filtration is characterized by a multi-layer stratification of filter units and is used to capture solid contaminants from the liquid phase. Filters are used for clarification and before other purification passages, as before columns loading. Clarification in general is the first step in which cells and cells debris are removed from the culture broth.

Precipitation is another technique, mainly used in lab scale, but not easily scalable and do not have a high resolution. By introducing a chemical substance into the mixture, precipitation separate the protein from the solution.

The second step in general is the capture and the main techniques used are: Isolation, concentration and volume reducing. Then the Intermediate purification which consist in the removal of bulk impurities and main proteins contaminants. The last step is the polishing in which trace impurities are reduced or eliminated, closed to the remotion f protein aggregates and closely related contaminants.

The most effective, specific and advanced technique is chromatography. Chromatography corresponds to the separation of proteins from other components, based on their chemical and physical characteristics. Using a packed column with a defined resin it is possible to separate on the solid phase the protein of interested, which binds to the resin, from the liquid phase and unbound proteins or contaminants, that passe through the bed column. Proteins can be purified, according to their characteristics, with different chromatographic methods, including ion exchange, size exclusion, hydrophobic interactions and affinity chromatography, etc [80].

Hydrophobic interaction chromatography (HIC) separates or purifies proteins, based on differences in their hydrophobic surfaces. It uses a reversible interaction between the biomolecule and the hydrophobic ligand of a HIC resin.

Multimodal or mixed-mode chromatography (MM) purifies proteins in a capture or polishing step. It uses resins functionalized with ligands capable of multiple interactions, which is useful to purify target molecules without a known specificity.

Size-exclusion chromatography (SEC), known as gel filtration, used a gel medium to discriminate proteins on their size. In SEC molecules don’t bind to the chromatography resin. Due to the limited amount of sample that can be loaded and corresponding to a maximum of 5% of the column volume, this technique is not used in industrial processes.

Affinity chromatography (AC) separates proteins using a strong, but reversible interaction, between the protein and a specific ligand. Affinity chromatography resins rely on binding interactions between a ligand immobilized on a resin and its binding partner. Because the interaction is so specific, high purity can be achieved in a single passage. It is often used for Antibodies purification on protein A, and with tagged proteins as for example His tagged proteins with IMAC (Immobilized metal-ion affinity chromatography) resin.

Ion exchange chromatography (IEX) separates molecules based on their overall surface charges. Separation is based on the reversible interactions, between a charged protein and a chromatographic resin, that has an opposite charge.

Column binding could also involve protein ionic force, so cation exchange column (CEX) or anion exchange column (AEX) could be used. In the CEX column the resin is a negatively charged polymer material with acidic groups (like sulfonic or carboxylic) and can binds and exchanges positively charged ions (cations) from a solution. In anion exchange chromatography, the resin is positively charged and attracts and binds negatively charged ions (anions) from a solution, exchanging them with other anions on the resin. These resins have functional groups (like quaternary amines) that provide the positive charge, allowing them to "trap" unwanted anions and release desirable ones, functioning as a solid-phase ion exchanger.

It is important to consider that, when a column is packed in house, its efficiency can be tested and expressed as the number of theoretical plates per meter chromatography bed (N) or as H (height equivalent to a theoretical plate, HETP), which is the bed length (L) divided by the plate number. Column efficiency is related to the broadening that can occur on a column and can be calculated from the expression. Generally, a good Height value is about two to three times the average particle diameter of the medium being packed.

Concentration is another strategy often combined with buffer exchange. In Tangential Flow Filtration (TFF) a multilayer cassette with a precise cut off, usually five time lower than the protein size, is used. A tangential flow is applied, and contaminants smaller than the cut-off will pass through, whereas protein will be concentrated, in the same way buffer can be exchanged to prepare the next purification step [81].

The last purification step is the protein formulation, in a buffer able to guarantee drug product stability and compatibility with patient administration. In this step, the parameters such as pH, excipients composition and drug concentration will be accurately set.

The desire protein quality, production process size, and purification degree will all influence the selection of the purification techniques.
Generally, during a downstream process, a combination of various chromatographic techniques is used, and in many cases, the first step will be an affinity chromatography, depending on the presence or not of an affinity tag, or an affinity resin.
If a protease cleavage site has been included, between the affinity tag and the protein of interest, this specific protease can be used to remove the affinity tag either immediately after the affinity chromatography column or later in the purification process. To increase purity, a second chromatographic passage such as ion exchange chromatography or hydrophobic interactions chromatography can be used. As a final polishing step, a size exclusion chromatography is usually performed, on the lab scale, and will also remove oligomerized variants of the protein(s) of interest. However, as described above this technique is not scalable and never used in industrial processes [81].
A standard purification protocol does not exist and each new protein, according to its own biochemical structure and activity, will necessitate a specific set up of its downstream protocol, which will be calibrated on its characteristics, project scalability, regulatory requirements and GMP plant facility equipment.

Recombinant protein has an economic value, based on their growing request and the market sales relevance. Producing a recombinant protein is not an easy task, the genetic construct design and expression system definition, represent only a starting point. The process of recombinant protein production is a structured workflow composed of Upstream and Downstream phases. The Upstream aim is to produce a high cell density culture (HCDC), able to express the recombinant protein and the downstream aim is to isolate the protein of interest, from the contaminants and formulate it.

9. Materials and Methods

9.1. PelB-rhBDNF Gene Design

For periplasmic expression, a commercially available pET28b+ vector has been used to clone the expression cassette. A synthetic gene has been ordered to Eurofins, in which the mature neurotrophin protein has been fused, in frame with the PelB leader sequence.

The XbaI (TCTAGA) and XhoI (CTCGAG) restriction enzymes sites have been inserted in the gene sequence. XbaI was a unique restriction site in the vector sequence situated in position 335, upstream of the gene of interest, whereas XhoI was the last polylinker restriction enzyme. By doing so, the His tag present in the vector has been deleted and the resulting expressed neurotrophin do not contain any tag.

A synthetic gene has been synthetized with codons adapted for its expression in *Escherichia coli*. The resulting protein has therefore the following sequence:

MKYLLPTAAA GLLLLAAQPA MAHSDPARRG ELSVCDSISE WVTAADKKTA
VDMSGGTVTV LEKVPVSKGQ LKQYFYETKC NPMGYTKEGC RGIDKRHWNS
QCRTTQSYVR ALTMDSKKRI GWRFIRIDTS CVCTLTIKRG R

Figure 18- Protein sequence of rhPelB-BDNF. In blue is reported the PelB sequence and in black the BDNF one.

With in blue the PelB leader and in black mature human BDNF.

The pelB leader sequence consists of 22 N-terminal amino acid residues and derived from the pectate lyase of *Erwinia caratovora*. This leader sequence fused together with the mature neuroprotein consent its transfer into the *E. coli* periplasmic space.

The construct was synthetized by Eurofins with animal free safety level, and successive steps have been conducted in Dompé. Once the vectors have been supplied, transformation protocol, following the supplier instructions, has been performed to obtain the recombinant clones (see below).

9.2. Shuffle rhProBDNF Gene Design

For the Shuffle expression strategy, the synthetic gene corresponding to the Proneurotrophin, has been synthetized, and inserted in the pET 28b+ expression vector (Novagen, code#69865).

The gene of interested has been insert into the vector in the NcoI and XhoI enzyme restriction sites, with the ATG included in the NcoI restriction site sequence corresponding to the methionine triplet codon.

The construct was synthetized by Eurofins with animal free safety level, and successive steps have been conducted in the Dompé. Once the vectors have been supplied, transformations have been performed, following the competent cells supplier instructions.

9.3. Transformation and Clone Selection

Two different hosts have been chosen according to the two designed strategies to obtain the best possible recombinant protein expression yield, propaedeutic to a possible future industrial process.

For the periplasmic expression, the selected host was the *Escherichia coli BL21(* New England Biolabs, code# C2527H) strain, a microorganism, with genetic modifications, possessing the following characteristics: T7 RNA Polymerase inserted in the DE3 prophage (*λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5)*, deficient in proteases Lon and OmpT, resistant to phage T1 (fhuA2). Enhanced BL21 derivative, does not restrict methylated DNA ($McrA^-$, $McrBC^-$, $EcoBr^-m^-$, Mrr^-).

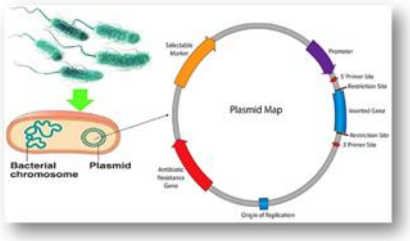
For the Shuffle Expression the *Escherichia coli* Shuffle (New England Biolabs, code #C3026J) strain has been used. Chemically competent *E. coli* cells, engineered to form disulfide bonds in the cytoplasm have been used.

They are suitable for T7 promoter driven protein expression and constitutively expresses a chromosomal copy of the disulfide bond isomerase DsbC, which promotes the correction of mis-oxidized proteins into their correct form. The cytoplasmic DsbC is also a chaperone that can assist the folding of proteins that do not require disulfide bonds. They express a chromosomal copy of the T7 RNA polymerase tightly controlled by $lacI^q$ and allows expression of potentially toxic proteins. They are also resistant to phage T1(fhuA2), and to the antibiotics: Nitrofurantoin, Streptomycin, Spectinomycin.

After transformation, four well isolated independent colonies have been selected and, to assure their clonalities, three selection passages have been performed on LB agar plates (Merck, code#1.10283), containing 50µg/mL of kanamycin (Sigma, code #K1637).
Then, four different Research Cleaned Cell Banks (RCB) have been produced, and after analytical characterization, to confirm their cell identities and purities, tested for the level of recombinant protein expression by SDS-PAGE and Western Blot. Among them, the best expressing clone has been chosen to proceed with project. A cryo-preservative chemical reagent has been added and RCB stored at -70°C until use.

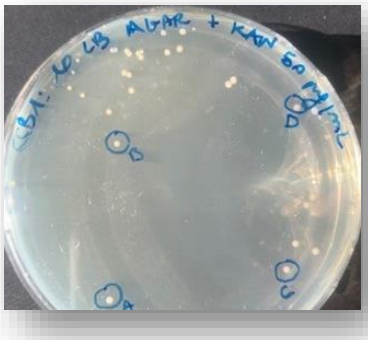
Table 4- Process of Research Cell Bank production. Strating with a selection of the expression system, composed of host and vector containing the protein of interest. After transformation to assure clonality three selection passages have been performed on solid medium. At the end of selection strategy, a liquid culture has been performed to amplify the seeded clone, growth was stopped by dropping the temperature to 4°C, and a cryopreservant added. Cells have been stored at -70°C until further use.

Transformation
E. coli strain HOST + Vector with neurotrophin gene.



↓

1° SELECTION
4 well isolated colonies have been re-streaked on agar plates



↓

2° SELECTION
4 well isolated colonies have been re-streaked on agar plates

↓

3° SELECTION
4 well isolated colonies have been re-streaked on agar plates

↓

CCB ON Pre-Culture Synchronization
4 Pre-Culture Flasks with Complex Medium Growth until 4-5 OD_{600nm}



↓

RCB Culture and Vialing
Pre-Culture Flasks with Complex Medium
Growth until 4-5 OD_{600nm} + 20% glycerol



9.4. Upstream: Flask Fermentation Process

The purpose of the cultivations is to obtain high cell densities, and to produce the maximum of the target protein to ease the successive downstream process.

At laboratory scale, to seed a 1-liter fermenter, a 125 ml Flask containing 25 mL of culture medium (1/5 of its volume) is inoculated with 52 µL of RCB cells. The culture medium is composed of a modified Terrific Broth which has the following composition per liter: 24 g Yeast Extract (Merck, code#1.13885), 12 g soy peptone (Merck, code#1.11932), 4.8 g potassium di-hydrogen orthophosphate (Merck, cose#P6696), 2.2 g di-potassium hydrogen orthophosphate (Merck, code#P5504) and 12 g glucose (Merck#code1.37048). TB medium was supplemented with 50 µg/mL kanamycin (Sigma, code #K1637). Medium was slightly modified to assure the perfect growth conditions.

The medium composition is identical for all fermentations. After overnight shake flask growth, 20 ml of pre-culture have been transferred in the Minibio fermenter filled with 700 mL of medium.

9.5. Upstream: 1L Fermentation Process

To scout fermentation conditions, a battery of eight Minibio fermenters (from Applikon) working in parallel, present in the R&D Dompé Biotech process development laboratory, has been used.

Bioreactors containing 700 mL of TB medium supplemented 50 mg/L kanamycin, were inoculated with 20 mL of this overnight seed. Each Applikon MiniBio was driven by an Applikon my-Control unit and connected to a central computer, containing the Lucullus software, registering all the fermentation parameters.

Each fermenter has pH, dissolved oxygen and temperature probes, whereas optical density values were estimated through connected biomass probes and in parallel OD_{600nm} values were hourly manually measured.

The actuators present can restore the parameters imposed in the My-Control units, through pumps activations or by acting on air, pure oxygen addition and on stirrer speed.

Batch fermentation conditions are temperature 30°C, pH 7.0, Dissolved oxygen (DO) at 50%, condenser at 100% and the dissolved oxygen value is automatically maintained by a sequential cascade of agitation (stirrer) between 500 and 1500 rpm, aeration between 0.5-1,5 L/min of compressed air and up to 1 L/min pure oxygen in cascade.

Agitation was provided by two axially mounted six-bladed Rushton turbines, with equal distance between them. The pH was controlled at 7.0 using 10% Orthophosphoric Acid (Merck, code# 1.00563) and 25% Ammonium solution (Merck, code#1.05422), supplied respectively by pump 1 and pump 2, with variable speed according to the PID. Antifoam 204 (Sigma, code#8311), diluted 1 to 10, in culture media was automatically added, by pump 3, when triggered by the level control probe 1. This conductivity probe is mounted 5 cm below the fermenter head. During the fed-batch phase, temperature has been changed to 20°C or 37°C according to the fermentation strategies, and 200 mL of feeding medium (TB medium), have been dispensed at a fixed value of 2% of pump 4 speed, equal to 200µL/min, for cultures run overnight at 20°C or with an exponential rate from 2% to 10% maximum pump speed in 4 hours for cultures run at 37°C.

9.6. Biomass Harvest

At the end of the fermentation process, biomass is harvested by batch centrifugation (ThermoFisher, Model #Lynx 4000) at 8500 rpm, 4°C and 15 minutes.

Cells were resuspended in lysis buffer composed of 2M Tris-HCl (Merck #PHG200), 20mM EDTA (Merck, code#1.08421), pH 8.0. The biomass to lysis buffer ratio was 0.5 mL of Lysis buffer for 1g of cellular pellet.

In some fermentations, for PelB construct, before biomass harvest, the temperature effect has been tested to try to release the protein present in the periplasm into the medium and in parallel to obtain a higher purified material, by thermically denaturing host cell proteins.

In the same way, this treatment has also been performed on the culture medium putatively containing the recombinant protein to induce denaturation of the HCP.

9.7. PelB-BDNF extraction from periplasmic compartment.

High Pressure Homogenizer (HPH) with Panda instruments (Niro Soavi, Model Panda type NS1001L2K) have been used to disrupt the cells. Four cycles at 790 bar pressure have been performed. Cell disruption is evaluated by spectrophotometer, measuring the reduction of OD_{600nm}.

In some experiments, other treatments have been tested to selective disrupt the outer membrane (OM) such as temperature, acidification, salts and ammonium sulphate additions. Temperature applied were 50°C for 1 hour. Salt treatment has been performed with Sodium Chloride (Merck, code#1.37017), at different concentration. Ammonium sulphate (Sigma, code#1952) is added to sample with a 30% W/V ratio.
After homogenization, or selective treatments, lysed cells have been centrifugated to separate soluble from insoluble material.
Localization of the soluble and refolded rhBDNF has been tested by SDS-PAGE and Western blot.

9.8. Benzonase Treatment

To reduce the DNA presence, representing the major contaminant after cell disruption, an enzymatic treatment has been performed, using Benzonase Nuclease (Merck, #E1014). According to the supplier, Benzonase have been used in the following buffer: 20 mM Tris-HCl (Merck, code #PHG200), 2 mM Magnesium chloride (Sigma, code #M8266), 20 mM Sodium Chloride (Merck, code #1.13017) at pH 8.0.
Different incubation times have been tested: 24, 48 and 72 hours. The combination of benzonase with temperature has also been tested. DNA fragmentation has been evaluated by running the samples in an agarose gel.

9.9. rhProBDNF Extraction from Shuffle strain cytoplasm

High Pressure Homogenizer (HPH) with Panda instruments (Niro Soavi, Model Panda, type NS1001L2K) have been used to disrupt the cells. Four cycles at 790 bar pressure have been performed. Cell disruption was evaluated by spectrophotometer, measuring the OD_{600nm} reduction. After homogenization, lysed cells have been centrifugated and supernatant recovered. Localization of the protein of interest has been confirmed by SDS-PAGE and Western blot.

9.10. Turrax treatment

To break down DNA fragments present in the homogenized supernatant, a mechanical procedure with Turrax (Ika, Model Werke T25 13000 rpm) was performed. Solution has been incubated, in ice, and treated for 2, 4 and 6 minutes. Turrax mixing induces share forces which mechanically breaks DNA into smaller fragments.
Reduction in DNA length, for each treatment time, has been evaluated on Agarose gel.

9.11. HCIC column

Shuffle rhProBDNF soluble material was loaded onto a hydrophobic charge-induction chromatography (MEP Hypercell PALL, code#12035-040) at pH 8.0.

For this several columns of diameters ranging from 1 to 5 centimetres of diameter but all of 20 cm bed, have been packed. The column was washed with lysis buffer, then equilibrated with 10 mM trometamol (Merck, code #1.08386) and eluted with 1 M urea (Merck, code #15604), 50 mM citric acid (Merck, code#1.00243) at pH 4.0 until.

This type of resin can bind proteins at basic pH and desorbs them, with a simple drop of pH.



Figure 19- Chromatograph present in Dompé laboratories with main size and type on column packed used along Processes development.

Load material together with the collected fractions and flow through were characterized by RP-HPLC.

9.12. Enzymatic Cleavage

As the soluble and therefore refolded neurotrophin precursor, extracted from the Shuffle strain, still has the now useless Propeptide (chaperon), that needs to be removed to obtain the mature BDNF, trypsin was used to cleave it.

Enzymatic digestion was directly performed for 16 h on the HCIC eluted peak, just by adjusting its pH to 6.0. Trypsin (Roche, code #06369880103) was added with an enzyme to protein ratio of 1 to 25.000 and reaction incubated at room temperature.

After tryptic digestion, pH was drop to 5.0 to stop the reaction. Trypsin Enzymatic digestion was monitored by RP-UPLC analysis and the parameter to follow was the percentage the hyper digested variant. Digestion was stopped when its peak area was corresponding to 20% of the mature BDNF one.

9.13. CEX column

The Cation Exchange Column (CEX) column containing Sepharose HP (high performance) resin (Cytiva, code #71-1087-03) was equilibrated with enzymatic cleavage buffer but at pH 5.0.

After material loading, the column was first washed with 1 M urea (Merck, code #15604), and 50 mM citric acid (Merck, code#1.00243), pH 6.5, and the protein has been successively eluted isocratically in a buffer with a higher salt concentration. Loaded material together with the collected fractions and flow through have been analysed by RP-UPLC.

9.14. HIC

CEX pool was subjected to salt addition and then loaded on a Phenyl Sepharose Fast Flow resin (Cytiva, #71-1065-03), pre-equilibrated with the same high salt concentration.

After washing, the protein was eluted by dropping salt concentration, in a gradient manner. Loaded material together with the collected fractions and flow through have been analytically characterized by RPUPLC.

9.15. Tangential Flow Filtration

A cassette (Pall, code# 4503685434) was used to concentrate the pooled fractions to the desired concentration. The solution was diafiltrated with 6 volumes of final formulation buffer at pH 7.0.

The concentration and final formulation of the protein have been followed by UV_{280nm} and for the batches destined to the pre-clinical studies, the complete analytical panel was performed.

10.Results and Discussion

10.1. Periplasm gene design, transformation, clone selection and Research Cell Banck (RCB) production

A gene has been designed to fuse the PelB leader sequence in frame with hBDNF. A synthetic gene f, with optimized triplets, has been synthetized and cloned by Eurofins in pET28 b+ expression vector.

The construct has been transformed into *E. coli* BL21 strain and three selection passages have been performed.

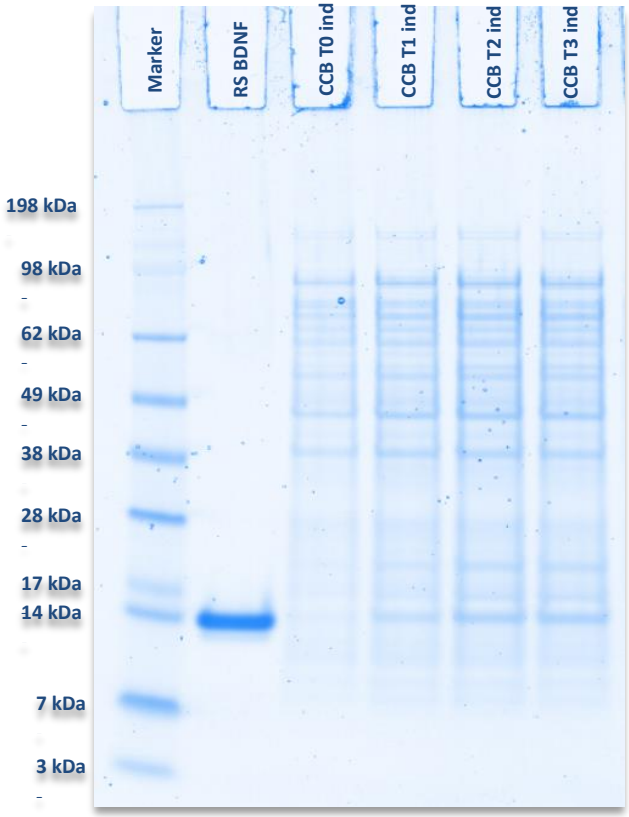


Figure 20- This SDS-PAGE reports the Markers, with a panel of different molecular weight proteins and the rhBDNF reference standard witch present a band at 14 kDa; In the three induction time points one can observe a band running at the same molecular weight of the RS one.

Among the different characterized clones expressing the recombinant protein, the best expressing one has been chosen according to SDS-PAGE and Western Blot.

This clone represents the research cell bank (RCB), which has been stored at −70 °C until use. It has been characterized and used for the development of the best fermentation conditions.

10.2. Shuffle gene design, transformation, Clone selection and Research Cell Banck (RCB) production

The synthetic gene of SH_rhProBDNF, with optimized triplets and substitution of the natural human leader sequence with a methionine residue, has been designed by Dompé and synthetized and cloned by Eurofins in pET28 b+ expression vector, downstream of the T7 promotor.

The construct has been transformed into Shuffle *E. coli* strain, and three selection passages have been performed. After clonality test, the best clone, chosen according to SDS-PAGE and Western Blot, was used to generate the research cell bank (RCB), which has been stored at −70 °C until use.

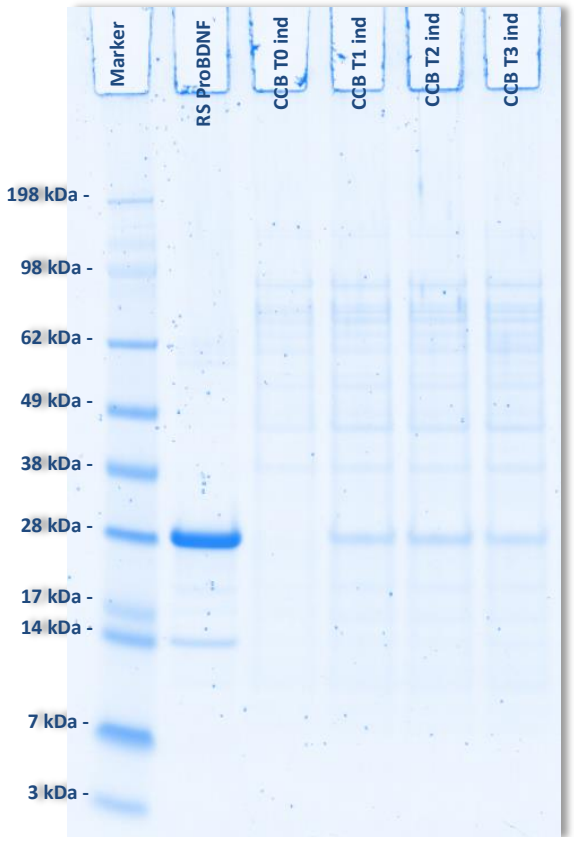


Figure 21- In this SDS-PAGE are reported the Marker, whit different molecular weight proteins and the rhProBDNF reference standard, which present a band at 28KDa; In the three induction time points is possible to see a band appearing after induction with a comparable molecular weight to the Reference Standard one.

This clone has been used for the development of the best fermentation conditions.

10.3. Fermentation

For each mutant, the selected clone has been used to seed the overnight flask culture, which, in turn, has been used the next day to inoculate 1L fermenter.

The Complex culture broth corresponding to a modified TB medium, was dispensed in the flask for the overnight pre-culture and the identified conditions are 200 rpm, 28°C and around 15 hours incubation time, with a starting inoculum of 52 µL of RCB.

The day after 20 ml of the overnight culture have been transferred into 1L Minibio, using a sterile syringe to inoculate the fermenter, through the dedicated septum. Initial experiments have been performed in Minibio fermenters of 1-litre capacity (Applikon), where up to 8 different conditions could be tested in parallel (figure 22).



Figure 22- Battery of eight 1L Minibio Fermenter, equipments present in Dompé laboratories.

Minibio are supplied with 6 probes: temperature, dissolved oxygen, pH, biomass, level 1, corresponding to foam and level 2 to vessel filling. For each fermentation thirteen parameters have been followed during the cultures: dissolved oxygen, temperature, pH,

stirrer, 4 pumps (acid, base, foam, Feed), biomass growth (probes and offline determination), air, oxygen and foam.

All these parameters have been tested to define the general fermentation kinetic. The standard cultivation is divided in two phases: the first one (batch) corresponds to microorganism growth, whereas the second one (fed-batch) to the recombinant protein production. This approach was used to always control dissolved oxygen (DO) in the medium. In fermentation, oxygen is used by microorganisms to oxidize substrate and is therefore strictly necessary for a good metabolism and recombinant protein production. In the batch phase the culture growth is limited by temperature (30°C), whereas in the second one by substrate supply (glucose). The medium during the batch phase is composed of modified TB medium, and along the fed-batch phase 200 mL of feed are added, composed of glucose and yeast extract.

In the first part of the fermentation biomass grow until all glucose initially present in the medium has been metabolized, at this point the fed-batch phase starts, with the controlled addition of new substrate in the fermenter vessel. At the end of the batch phase, as glucose has been depleted, the microorganism changes its metabolism. This is perceptible in the pH switch (figure 23) or the oxygen spike (figure 25), as a consequence of the metabolism change. Between the two phases, the temperature is switched to 37°C or to 20°C, according to the desired strategy (Periplasmic or cytoplasmic protein expression).



Figure 23- Applikon software able to register all the parameters, and also all the actuators, pumps and probes values. The figure shows one of the software pages, where the probes trends are displayed. Here in particular the pH probe trends is reported and the moment in which the pH switch from 6.9 to 7.1, indicating the *E. coli* metabolism change, in-between the two fermentation phases.

One hour after metabolism switch, the recombinant protein expression is triggered, by addition of an inducer that activates the recombinant gene transcription.

In our case IPTG is used as non-metabolizable inducer and turns on the transcription of the T7 polymerase, which will in turn transcribes the recombinant gene. Once turn on the expression of the recombinant gene will continue until fermentation ends.

During the different runs several parameters have been optimized: medium and feed compositions, temperature, pH, agitation and aeration. Furthermore, different concentrations of inducer and induction time have been tested to find the best compromise between them. The table below reports the main conducted experiments:

Table 5 - List of the main fermentation conditions tested in 1L Minibio fermenter. Table reports the critical parameters individuated and varied. For the Fed-Batch three temperatures: 20°C, 30°C and 37°C were tested. The IPTG concentration has been varied from 0.1, 0.5 and 3 mM. The induction time assessed between 3 hours and overnight. Combining three parameters at time, the best fermentation conditions has been individuated.

Fermentations	Fed - batch TEMP (°C)	IPTG concentration	Induction time	Fermentations post treatments
I	20	0,1 mM	ON	Harvest and homogenization
II	20	0,5 mM	ON	Harvest and homogenization
III	37	3 mM	3h	Harvest and homogenization
IV	30	3 mM	ON	Harvest and homogenization
V	20	0,1 mM	ON	50°C for 1 h
VI	30	0,1 mM	ON	50°C for 1 h
VII	37	3 mM	3h	50°C for 1 h

During fermentation, culture growth has been hourly monitored, using spectrometer evaluation, measuring the OD_{600nm} values. Meanwhile, in parallel, biomass probes have been installed and tested to punctually measure culture growth.

Furthermore, these numerous fermentations parameters have been collected and analysed by Artificial Intelligence to predict final biomass productions. Results have been published in an article with title "A deep learning approach to optimize recombinant protein production in *Escherichia coli* fermentations" [1].

Among the different experiments, it becomes clear that slowing down the fed-batch phase temperature at 20°C, ease the recombinant protein transport into the periplasm, and in the Shuffle system facilitates the formation of the disulfide bridges. Furthermore, it was observed that a higher concentration of inducer, increases recombinant protein production.

So, a standard fermentation is defined and divided in two phases. The first one at 30 °C, knowns as batch phase, where the microorganism has its maximum growing velocity, and rapidly duplicates in an exponential manner. During this phase the DO% trigger is imposed at 50% and a loop for the air, stirrer and oxygen implemented.

Then, once the carbon source is exhausted, a second phase at lower temperature (20°C) or higher temperature (37°C) have been tested. In this phase the limitation factor is the substrate feeding rate, which is supplied by adding fresh medium in the fermenter. In this manner, during fermentation phases, first controlled by suboptimal temperature, then by limited substrate supply, the culture medium is always oxygenated, and cells never go into anaerobiosis, a condition needed to have an efficient recombinant protein production. One hour after fed-batch starting, the recombinant protein production is triggered by isopropyl-β-D-1-tiogalattopiranoside (IPTG), addition in the medium. Fermentation is then continued for a further three hours to obtain high recombinant protein accumulation, or prolonged, in the case of lower temperature condition, for additional 16 hours (overnight), to guarantee periplasmic transport (periplasm Expression) or protein folding (Shuffle expression).

Minibio fermenters are supplied with the Lucullus program (Applikon) able to impose parameters and register raw data every second. This program can summarise the data in different graphs as for example the on reported below:

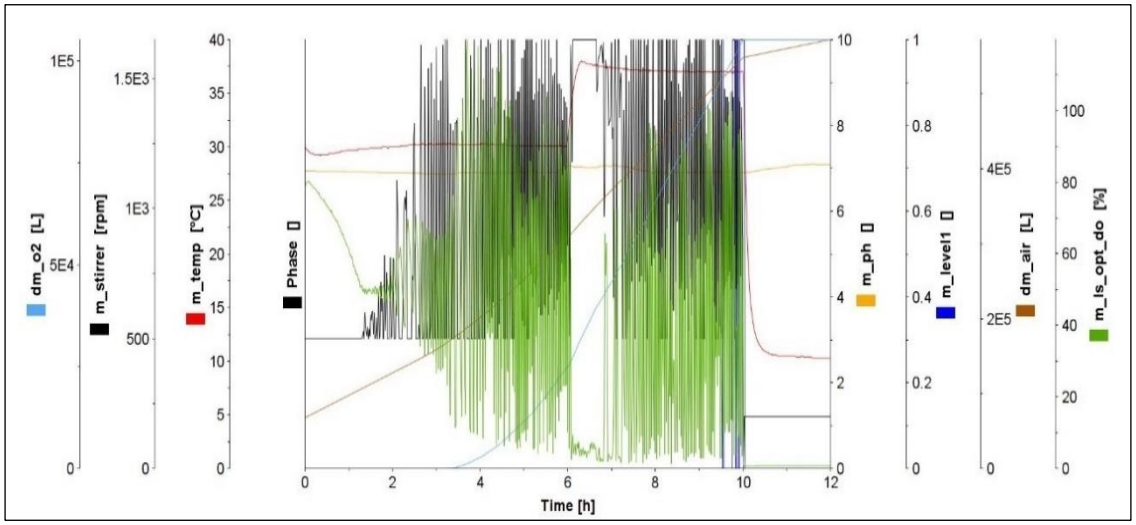


Figure 24- Lucullus Image for the Fermentation. Lucullus is a software able to register and programmed the Minibio fermenters. Here a typical fermentation graph is reported, in which parameters are highlighted with different colours and each axis are delimited with the parameter's ranges. The stirrer values is reported in black, in green the DO values and in the legend is indicated as LS_OPT_DO; in yellow the pH, in Red the temperature. The actuators curves are in blue the antifoam, in light blue the oxygen supplied and in brown the air supplied.

As we can see, the DO value evolved rapidly during the fermentation and varied around the fixed value of 50% and the oscillations, registered by the program, demonstrates how *E. coli* "breath", or in other word that the oxygen consummation is proportionally correlated to the substrate presence in the medium.



Figure 25- Applikon software able to register all the parameters, and all the actuators, pumps and probes values. The figure show one of the software page, where the probe trends are shown. Here in particular is reportede the pH probe trends and the moment in which the pH switch from 6.9 to 7.1, indicating the *E. coli* metabolism change, corrisponding to the two fermentation phases.

At the beginning glucose is the primary source of nutriment, then yeast extract represents the only possible substrate. After roughly five hours of culture, the number of bacteria in the fermenter starts to be consistent, and the dissolved oxygen value becomes too low, and despite the supply of compressed air, oxygen and agitation increase it is impossible to maintain enough oxygen in the culture, it is therefore necessary to limit cell growth by another method.

At the same time, pH reflecting the metabolism change, increases from 6.9 to 7.1 confirming the host needs. Contemporaneously due to the substrate lack a DO spike, also corroborating the metabolism switch, is observed.

At this precise moment, as it is impossible to supply enough oxygen to the fermenter, it is time to shift to the second fermentation phase, where glucose is limitedly added in the culture to slow down its growth.

According to the strategy, the culture temperature was shifted from 30°C to 37°C or to 20°C. At the same time feed has been supplied at sufficient rate to allows culture to grow aerobically. When 20°C temperature was applied along the overnight culture, the percentage of feed was supplied at constat and low percentage (2% of the pump capacity, that is equal to 0.2 mL/min); when the temperature was set to 37°C, the feeding supply is increased along the fed-batch phase, passing from 2% to 10% of the maximum pump speed. After one hour from the pH switch, induction has been performed manually, adding the IPTG inductor at 3mM concentration and at this point culture starts to produce the recombinant protein. During daily fermentations, three hours of induction phase are performed, instead during overnight fermentation the induction phase is longer, and last around 16 hours. Protein production during the induction phase have been evaluated and no heavy differences have been registered in terms of protein yield, between the two fermentation conditions. Sample have been collected along induction phase, four samples are collected for the daily fermentation and two or three for the overnight one.

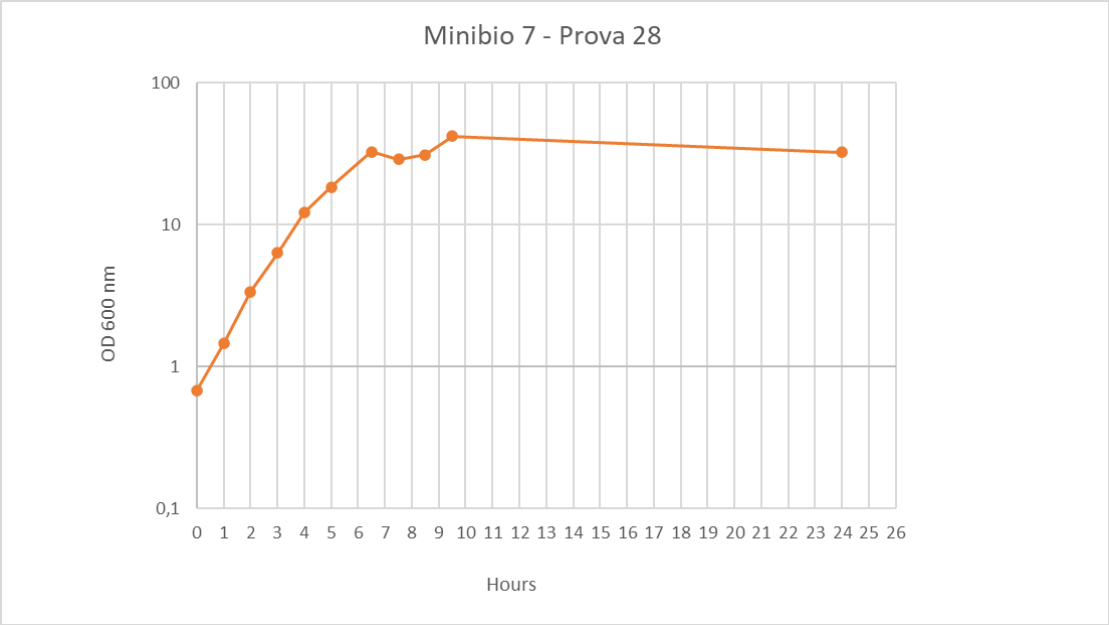


Figure 26- Growth curve corresponding to a 1L Minibio of PelB-rhBDNF Fermentation. The X axes reports fermentation time and the Y axes are reports sampled OD_{600nm} values.

During fermentation, OD600nm has been monitored, and it has been observed that OD600nm values are directly correlated to biomass and protein production. The figure 25 reports a typical fermentation growth curve of BDNF periplasmic expression, obtained in the following culture conditions: 30°C along the batch phase, 20°C along the overnight Fed-Batch phase and 3 mM IPTG concentration for induction.

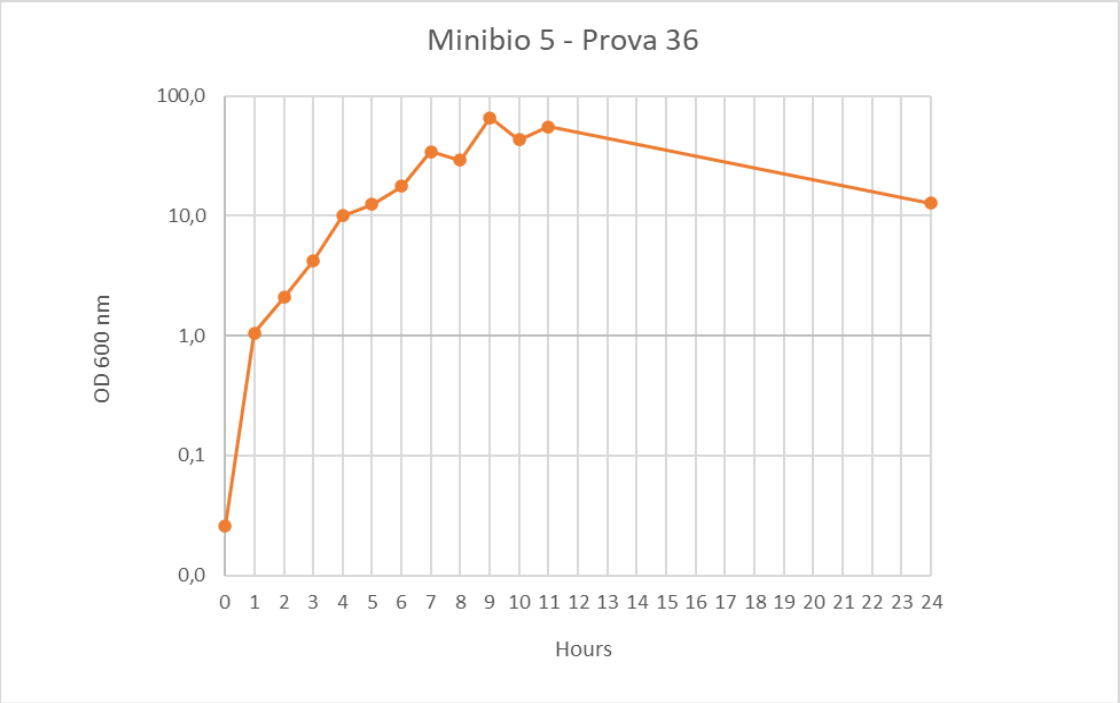


Figure 27- Growth curve corresponding to a 1L Minibio of Shuffle Fermentation. The X axes reports fermentation time and on the Y axes reports OD_{600nm} values. Dots indicated the value corresponding to the sampling plan and corrected for the dilution apply on the spectrophotometer.

For the Shuffle, a typical fermentation growth curve is also reported in the above graph (Fig.27). The curve in the figure below is obtained with 30°C for batch phase, loop for the DO at 50% imposed condition, 20°C along the overnight Fed-Batch phase and 3 mM IPTG concentration for induction. Protein production identity and purity have been evaluated, by SDS-PAGE and Western Blot, by the analytical R&D department (Figure29).

The image below shows the typical expression pattern profile for the periplasmic expression. Note the band present in the sample T24 (IPC after 16 hours of induction) corresponding to the mature BDNF, according to hypothesis that adding the heterologous PelB sequence is possible to transport in the periplasm, and the translocation mechanism removes the leader sequence. In the same gel are present also some samples tested to individuate protein localization. (Figure 28).

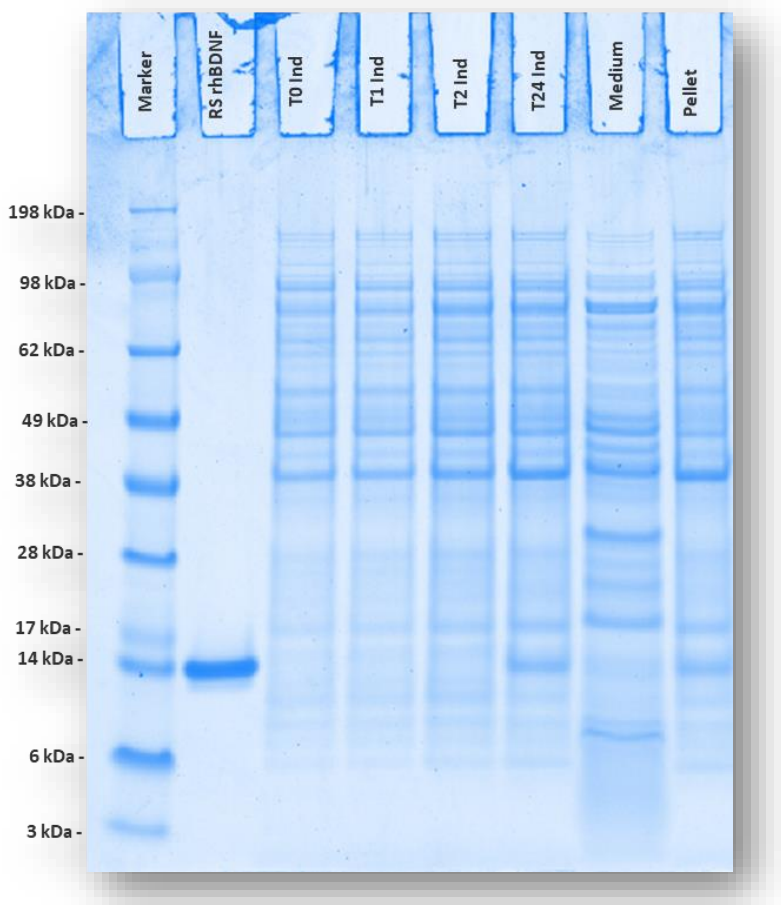


Figure 28- The SDS-PAGE dedicated to mature soluble and refolded protein expression and localization, after PelB fermentation process. Molecular weights markers and the reference standard indicated the presence of a Protein corresponding to the rhBDNF molecular weight (14 kDa). After 24 hours the protein presence is evident. No protein presence is reported in the medium, and part of it accumulates as insoluble in the pellet.

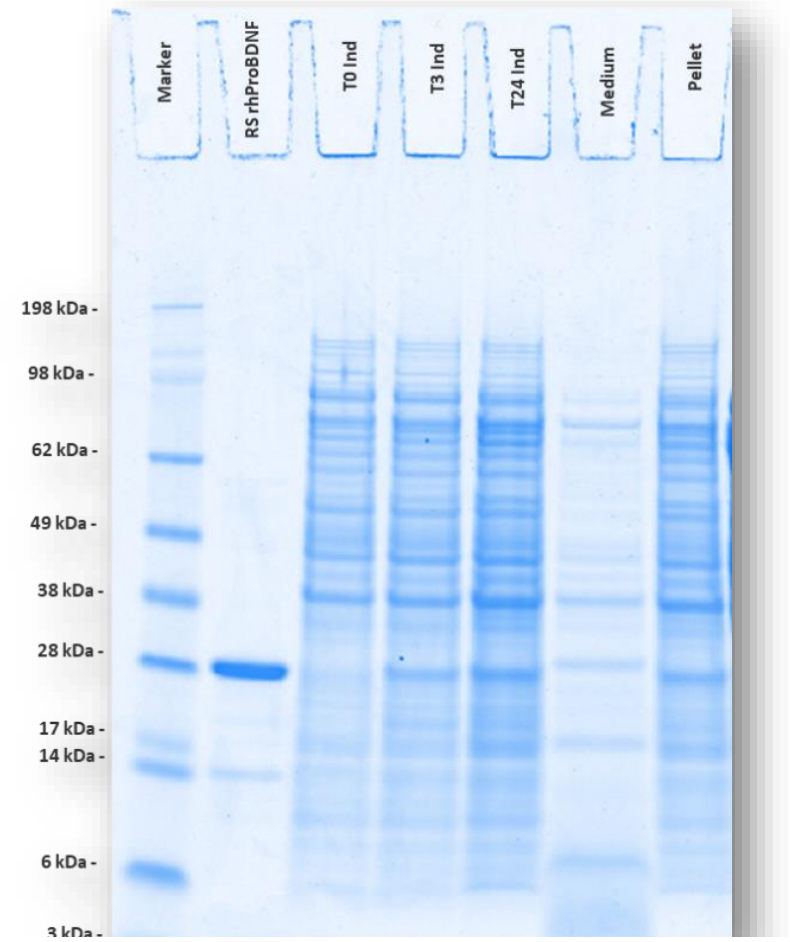


Figure 29- The SDS-PAGE dedicated to protein expression and localization. Molecular weights markers and the reference standard indicated the presence of a Protein corresponding to the rhProBDNF molecular weight (28 kDa). After three hours of induction the recombinant protein can be observed and is more evident after 24 hours. No recombinant protein is present in the medium, and part can be evidenced in the insoluble fraction.

At this point culture has been scaled up in a 20 liters fermenter and process yield confirmed for both strategies tested (results do not showed).

10.4. Biomass Harvest

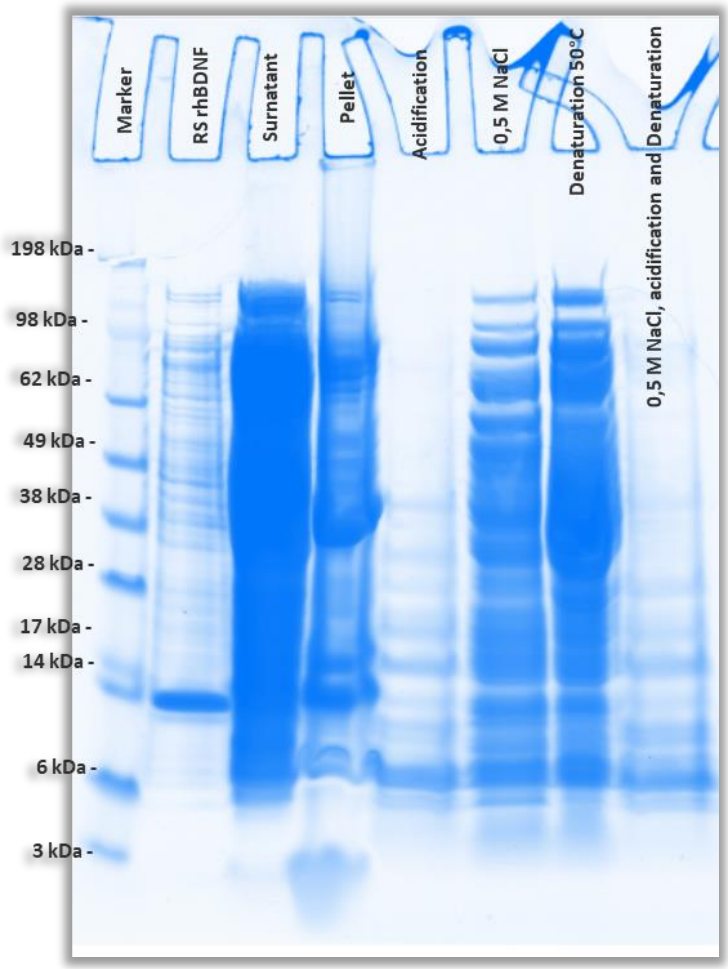
In two experiments a temperature treatment was applied in the fermenter vessel and using 50°C which should in theory permit to first destabilize the membranes and "extract" the proteins from the periplasmic space, if they are small and soluble, and at the same time denaturate and precipitate the Host Cell Protein (HCP), which are not resistant to thermic inactivation. But, after analysis, no recombinant protein was obtained with those treatments. We can conclude that the mature BDNF is not "naturally" present in the culture media, and that the 50°C temperature treatment does not release it from the periplasmic space.

At the end of the culture, 0.9 L of fermentation has been harvested by batch centrifugation at 8500 rpm, 4°C and 15 minutes. After centrifugation, the biomass pellet was resuspended in a lysis buffer (ratio 0.5 mL Lysis buffer to 1g biomass).

10.5. PelB-BDNF extraction from periplasmic compartment

As after analysis of the whole cell extract of the pelB-BDNF clone, a protein corresponding to mature BDNF was present in the soluble material (see below), confirming that this strategy would have correctly expressed the protein in the periplasmic space. In a first attempt, selective disruption of the outer membrane (OM) has been tested.

Figure 30 - After homogenization, on the soluble protein fraction three different treatments have been tested: acidification, Salt addition and thermic denaturation. The combined effect of the three treatments has also been tested. As reported in the gel picture absence of mature BDNF reveals the inefficacy of all the treatments.



The supernatant has been analyzed as is, or acidified, or after salt addition, or following treatment at 50°C for 1 hour or simultaneously acidifying it in presence of salt and treated at 50°C. Acidification has been performed with HCl 37%, until pH reaches 4.0. Denaturation consists in a 1-hour treatment at 50°C . and salt (NaCl) was added at different concentrations. After centrifugation, a protein of the expected molecular weight was observed in the pellet (insoluble fraction) of the sample treated with salt, and in the one incubated at 50°C.

All other treatments are too drastic to maintain protein integrity as revealed by SDS-PAGE analysis (figure 30).

The temperature treatment might have been too aggressive and have, in the best case, pulled BDNF down together with the other denatured proteins or directly denatured the neurotrophin which had precipitated (figure 31).

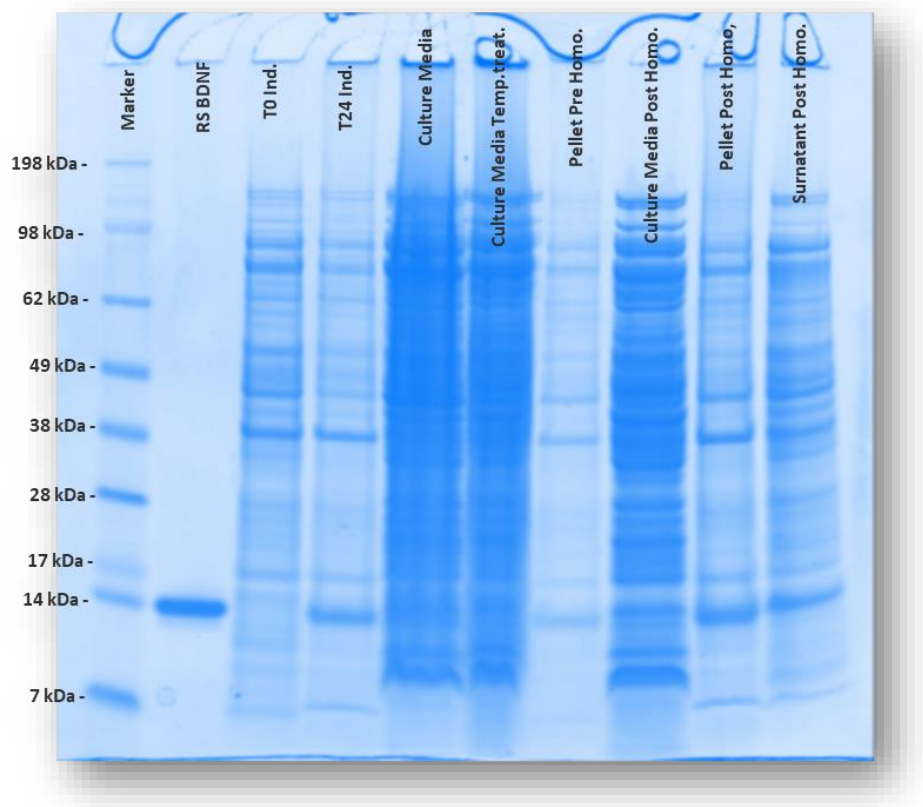


Figure 31- The Protein presence is evaluated by SDS-PAGE. the image report, together with the induction point and Culture medium, the soluble and insoluble fractions of mature BDNF present in the samples collecting along the homogenized procedure.

As those treatments to specifically extract proteins from the periplasmic space failed, and to set-up a GMP scalable protocol, it has then been decided to change strategy and to homogenize the harvested biomass, to release the soluble refolded rhBDNF, from the

microorganism periplasm. Biomass homogenization was performed on a further pellet and a nice protein pattern obtained.

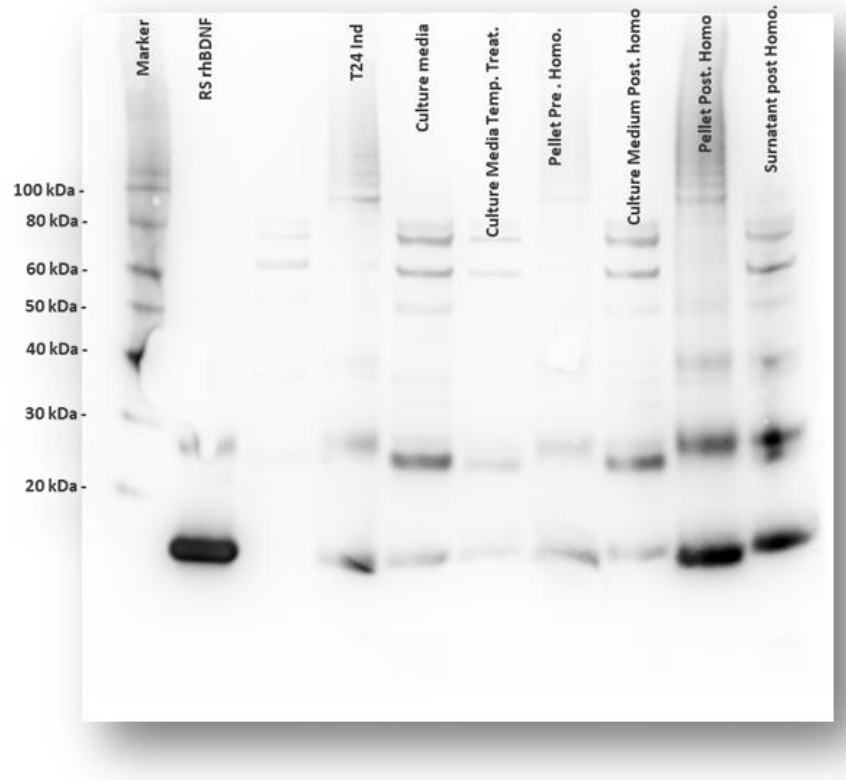


Figure 32- The Protein identity was confirmed by Western Blot. The membrane after development shows the protein presence at 14 kDa in sample called T24 and supernatant post homogenization, which represent the starting downstream material. As is possible to note that other bands are present, indicating that the monoclonal antibodies recognize the mature protein also in the insoluble portion but the bands with upper molecular weight could correspond to a low specificity of the Second AB, along the WB analysis.

High Pressure Homogenizer (HPH) from Panda instruments (Niro Soavi, Model Panda type NS1001L2K) has therefore been used to disrupt the cells, applying four cycles at 790 bars. Cell disruption was monitored by spectrophotometer, and 75% signal reduction has been observed, confirming treatment efficiency.



Figure 33-Homogenizer Panda by Niro Soavi present in the Dompé. The instrument is used to apply the high pressure while cooling the sample with an circuit of refrigerated water. Sample homogeneity is guaranteed by the magnetic stirrer bar and the Magnetic stirrer plate. Before and after the sample homogenization a cleaning procedure has been developed to sanitize the instrument between two different samples.

However, even after 4 cycles of homogenization the samples are still heavily contaminated with host cell DNA. Lysed cells have been centrifuged, and as expected the recombinant protein was localized both in soluble and insoluble fractions. The presence of soluble and refolded rhBDNF has been confirmed by SDS-PAGE and Western blot and the results are reported in the pictures below (figure 31 and 32 respectively).

10.6. Benzonase Treatment

After cells disruption, to separate the soluble and refolded protein from the insoluble part, another centrifugation has been applied on the homogenised material. As anticipated supernatant is heavily contaminated by host DNA, which is incompatible with the use of chromatographic columns to purify the protein.

Since, with those elevated concentrations, its remotion is impossible by a simple filtration on a positively charged filter (or column), it was decided to fractionate it by Benzonase enzymatic treatment which can degrade it down to trinucleotides fragments.

For this benzonase were added to the extract, with a buffer containing $MgCl_2$ as recommended by the supplier. Several experiments have been conducted, changing incubation time, from overnight to overweened, and at different temperatures (RT and 30°C) as indicated by supplier.

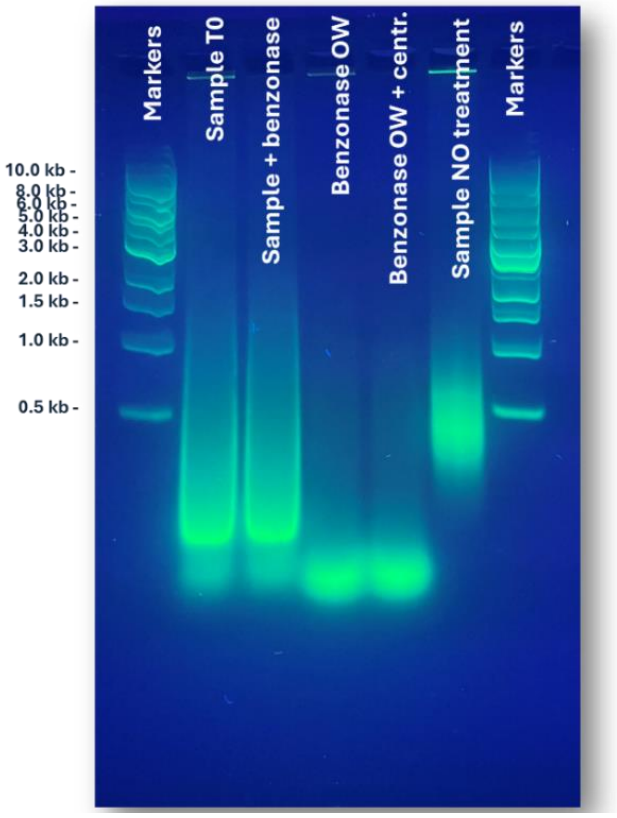


Figure 34- Agarose Gel Electrophoresis. DNA markers within 10 000-0.5 KB range. Supernatant after homogenization has been analysed before and after benzonase treatment. Here is reported the most effective treatment after Over-weekend incubation.

Results show that benzonase can digest DNA, and the best DNA reduction is obtained after 72 hours of incubation with the enzyme as it can be observed in the agarose gel above. But digestion buffer composition and long incubation time are incompatible with the BDNF protein solubility and stability.

Another possible regulatory problem, related to the long incubation time in this rich "medium" is the possibility that an adventitious contamination can occur, if sample is kept at room temperature for 72 hours. Moreover, a further regulatory request is the demonstration of DNase and DNA removal at the end of the Downstream purification.

10.7. rhProBDNF Extraction from Shuffle strain cytoplasm

Also, in this case biomass has been treated with High Pressure Homogenizer (HPH) to disrupt the cells. Four cycles at 790 bars of pressure have been performed. Cell disruption was monitored by spectrophotometer. After homogenization, lysed cells have been centrifugated to separate the insoluble contaminants from the recombinant protein. This procedure is not selective, but efficiently disrupt the cells, and has a straightforward protocol for scaling-up.

Localization of the protein of interest has been checked by SDS-PAGE and revealed by Western blot. As can be observed in the following SDS-PAGE rhProBDNF is visible at 36 kDa where the band starts to appear after three hours of induction and is more intense after 24 hours of fermentation (Figure 35).

Also, in this case protein localization has been investigated, and as it possible to see the recombinant protein is not released in the medium. But there is good evidence of its presence in the cytoplasm as soluble and refolded, as observable in the lane corresponding to the post homogenization supernatant.

Also, the pellet post homogenization contains a band corresponding to ProBDNF indicating that not all the synthetized protein has been refolded in the shuffle strain oxidant cytoplasm. All those considerations have been confirmed in the following Western blot.

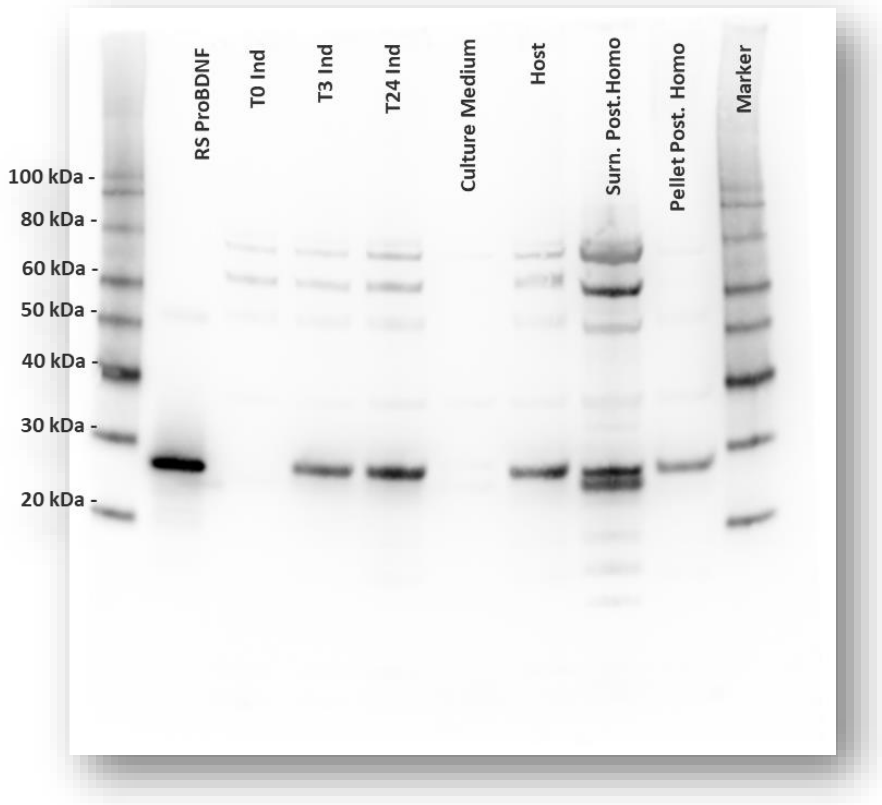


Figure 35- Assessment of Protein production evaluated for the Shuffle strategy by SDS-PAGE. The samples at T3 and at T24 hours show a high recombinant protein presence with a band at 28 kDa. Also, after homogenizer procedure sample have been tested on the same gel and it is possible to note that supernatant post homogenization and pellet post homogenization samples report the recombinant protein presence. Other bands are present in the Supernatant post homogenization, indicating that Host cell protein are also efficiently released by this treatment.

10.8. Turrax treatment

To fragment the DNA present in the supernatant after homogenization, a mechanical procedure with Turrax has been implemented. This system is a high-speed blending device, manufactured by IKA, which creates vortexes in the solution, introducing high shear forces which break down DNA into smaller pieces, compatible with chromatography.

Samples have been incubated in ice and treated with Turrax for 2, 4 and 6 minutes.

This reduction is sufficient to avoid column clogging and high backpressure. Moreover, the treatment is rapid as only 6 minutes at maximum are required. Whit this short treatment at low temperature protein is more stable in the Lysis buffer, respect to the benzonase digestion one.

The results are reported in the agarose gel on the left where DNA size has been assessed. As it can be observed DNA is still present and the fragmentation is however mild but sufficient for our scope. In the different samples run on this gel, the smears at the side of gel wells decrease in function of the treatment time. This is not really evident but once again sufficient to proceed with chromatography.

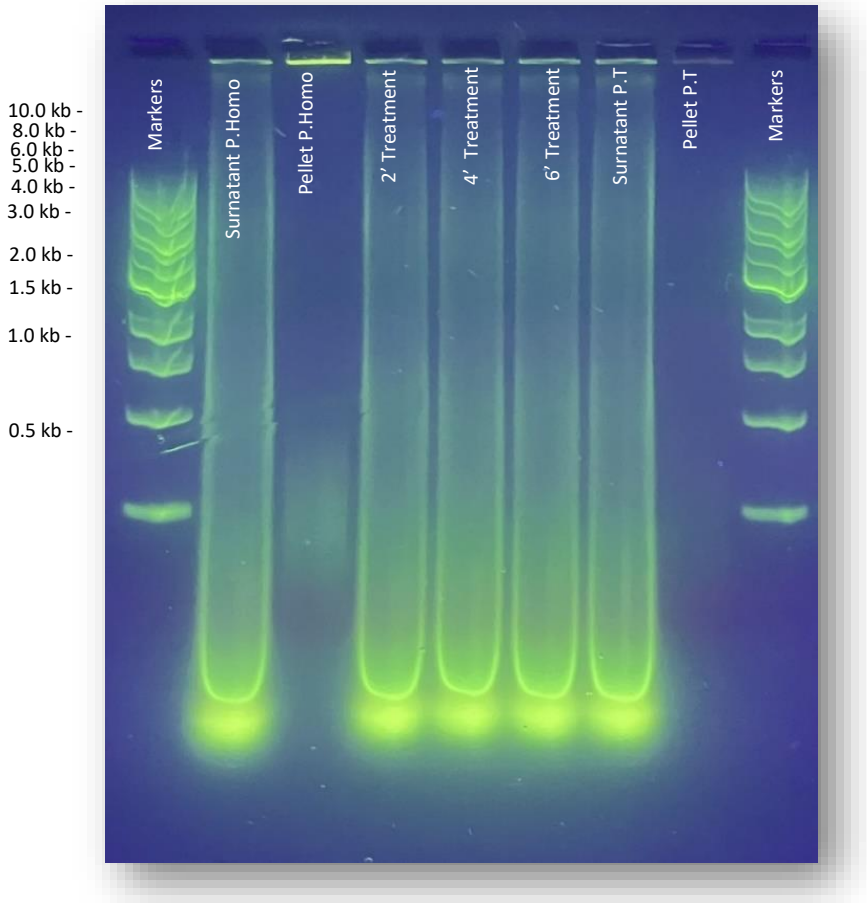


Figure 36- Agarose Gel Electrophoresis after Turrax treatment. Samples collected after each passage have been analysed, as it is possible to note is that reduction of DNA smear is more evident with the longest treatment time.

This mechanical treatment better fulfils regulatory requirements, and do not introduce exogenous contaminants in the treated solution as compared to benzonase that needs to be removed after use and no adventitious contamination can be foreseen.

At this point it was possible to directly load the solution onto a Hydrophobic Charge Induction Chromatography (HCIC) column.

10.9. HCIC column

After extraction and Turrax treatment, solution was filtered to eliminate aggregates and loaded on the first chromatographic column. A hydrophobic charge-induction chromatography (HCIC) has been chosen as capture step. HCIC is a mixed-mode chromatography technique, that uses hydrophobic interactions for binding and charge repulsion for elution.

It separates proteins by first using hydrophobic interactions at basic pH, to elicit ProBDNF binding to the column, and then elute it with a pH drop, creating electrostatic repulsion desorbing and eluting the neurotrophin. This technique is perfect for capturing the highly hydrophobic ProBDNF precursor, which is almost the only protein present in the mixture able to bind to the resin in those conditions. Among the different HCIC resins, the MEP one has been selected; this resin is characterized by a cellulose matrix linked to 4-mercapto-ethyl-pyridine (4-MEP) and the average of the bead diameter is 80-100 µm, allowing high flux and low column backpressure, two well-suited characteristics for a capture step. Due to its high hydrophobic nature, the soluble and refolded rhProBDNF obtained from Shuffle fermentation can efficiently binds to the column.

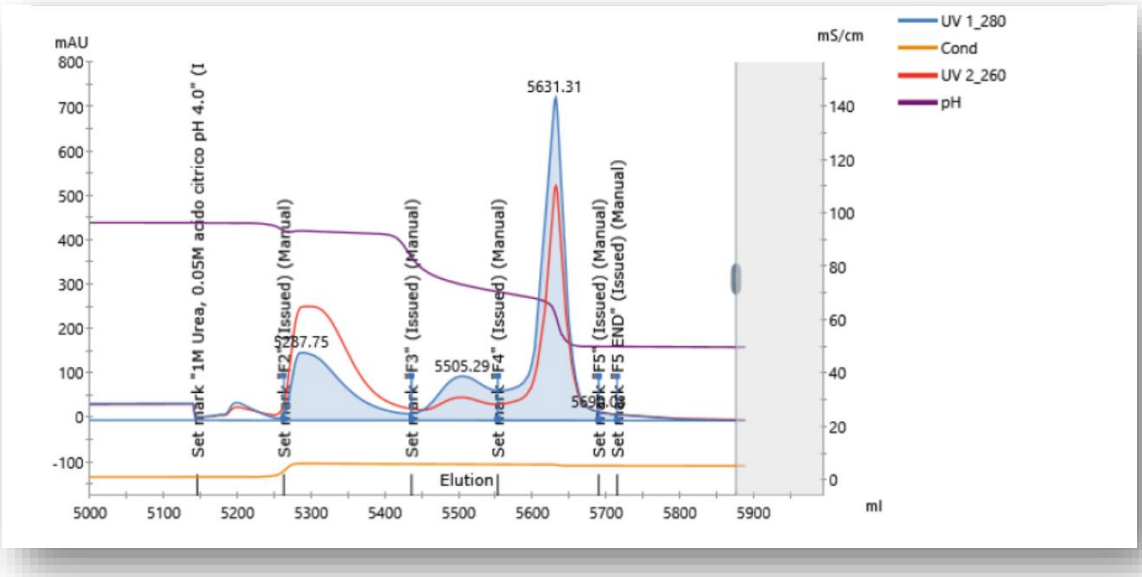


Figure 37- HCIC Chromatogram. The X axis reports the Elution volume, along with the sample fractions. The Y axes report the mAU, corresponding to the protein absorbance. The curve obtained has a good profile and the peaks are well separated. rhProBDNF is present in the highest peak.

Then column is washed and re-equilibrated with the Lysis buffer. Elution is then achieved with a buffer containing 1 M urea and citric acid at pH 4, able to detach the protein with the acid pH. The rhProBDNF protein binds to the resin and is eluted by dropping the pH from 8 to 4 where, acquiring a charge, the resin pyridine ring (4-MEP) desorbs the protein. Those conditions facilitate the following process phase, in fact the protein enzymatic digestion, which removes the now useless Propeptide, which has completed its chaperone function. According to the analytical IPC results, the main peak observed in the chromatograph is rhProBDNF and the reversed-phase RP-UPLC analysis, reveal a purity around 80%.

10.10. Enzymatic Cleavage and CEX chromatography

Once purified, the following process steps have been implemented to obtain the mature form, and to convert rhProBDNF precursor into rhBDNF. *In vivo*, this conversion is performed intracellularly by the furin protease or, extracellularly, with tolloid/bone morphogenetic protein-1 (BMP-1), which recognize and cleave the $R^{108}V^{109}R^{110}R^{111}$ sequence present in the pro-region carboxy terminus.

As the natural enzyme is not commercially available in GMP grade, and among the different available proteases present on the market, an alternative is represented by Trypsin. Trypsin is a well-known protease, commercially available in different quality grades, which specifically cleaves after arginine or lysine residues and as in the BDNF precursor the last pro-region residue is an arginine, trypsin is putatively compatible with the removal of the neurotrophin pro-sequence. So, starting from the fraction eluted from MEP column, rhProBDNF sample was adjusted to pH 6.00 and digested with trypsin for 16 h at room temperature, using an enzyme to protein ratio equal to 1:25 000 [82].

The RP-HPLC analysis revealed that almost 50% of the species present in the enzymatic digestion were compatible with the mature BDNF, whereas other forms were corresponding to a hyper digested variant, lacking the first six amino acids of the mature BDNF neurotrophin and to the hypo digested forms (incomplete pro-region digestion) together with several peptides corresponding to peptides coming from the proregion hydrolyse. All these data have been further confirmed by LC-MS mass spectrometry. Enzymatic reaction was stopped by dropping the pH to 5, one of the easiest ways to inactivate trypsin.

The downstream process continued with a second chromatographic column, a cation exchange chromatography (CEX). For this step, a column packed with sulphopropyl (SP)

Sepharose resin, pre-equilibrated at pH 5, was loaded with the digested material, washed, and finally eluted with a salt step. Increasing the amount of salt, the digested material elutes in three different peaks: the first peak, mainly corresponding to the elution of the hypo digested material; the central portion, containing the mature BDNF; the last peak, corresponding to the hyper digested one. All fractions have been analysed by RP-HPLC.

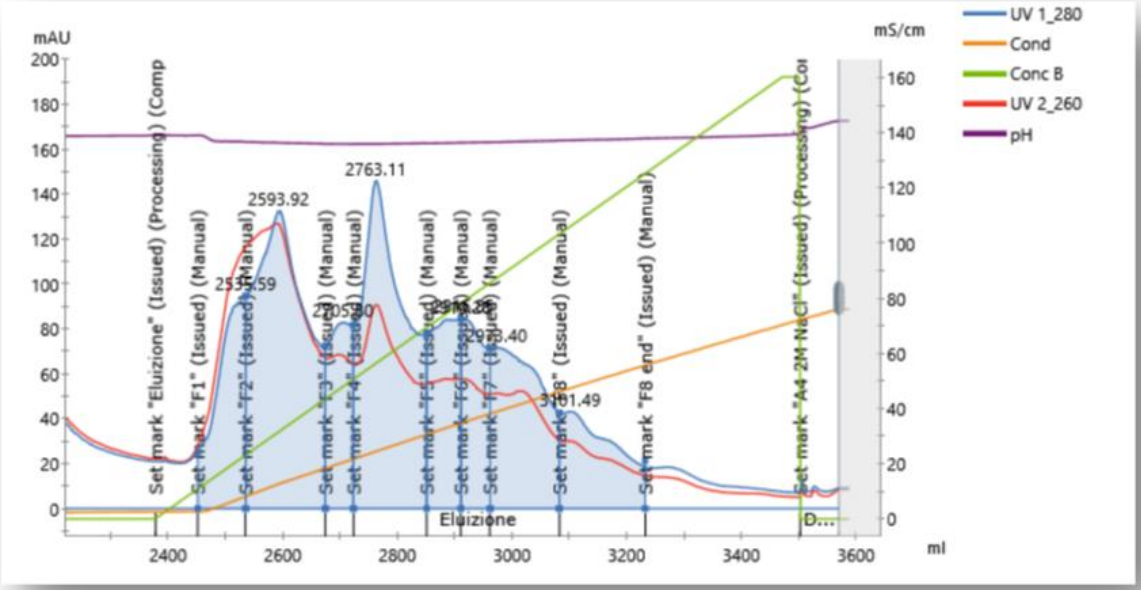


Figure 38- Chromatogram obtained after Cation Exchanged Chromatography. Elution is performed adding salt with a gradient from no salt to 2M Sodium Chloride. Separation is sufficient, and recombinant protein is present in the second and highest peak.

In this chromatography, rhBDNF was present in the central portion, but only in marginal quantities and heavily contaminated by variants, that represented up to 50% of the eluted material.

10.11. HIC column

After having eluted the protein with a moderated quantity of salt in the previous column, we chose to use a hydrophobic one to perform the next step, as it is easier to add salt to the loaded sample than to remove it.

Hydrophobic chromatography is based on the protein capacity to bind the resin in high salt and to elute it by dropping its concentration. Among the different resins, Phenyl Sepharose Fast Flow Low Substitution (PS FF LS) has been selected, as other resins are

too hydrophobic to release the protein, as for example the Phenyl Sepharose FF high substitution (data not shown), or not enough hydrophobic to capture it.

Therefore, after several trials to find the best conditions, salt has been added to the CEX pooled fractions to reach 2 M NaCl to load the sample on the Phenyl Sepharose column, this to separate the protein variants, according to their hydrophobicities.

After loading and washing the protein was eluted by a drop of salt able to separate the rhBDNF in its dimeric form, from the aggregates and other more hydrophobic variants, or contaminant proteins that elute, in low salt at the end of the run.

After chromatography the eluted fractions have been analysed by the RP-UPLC. At this step, the recovered rhBDNF protein showed roughly 80% of purity.

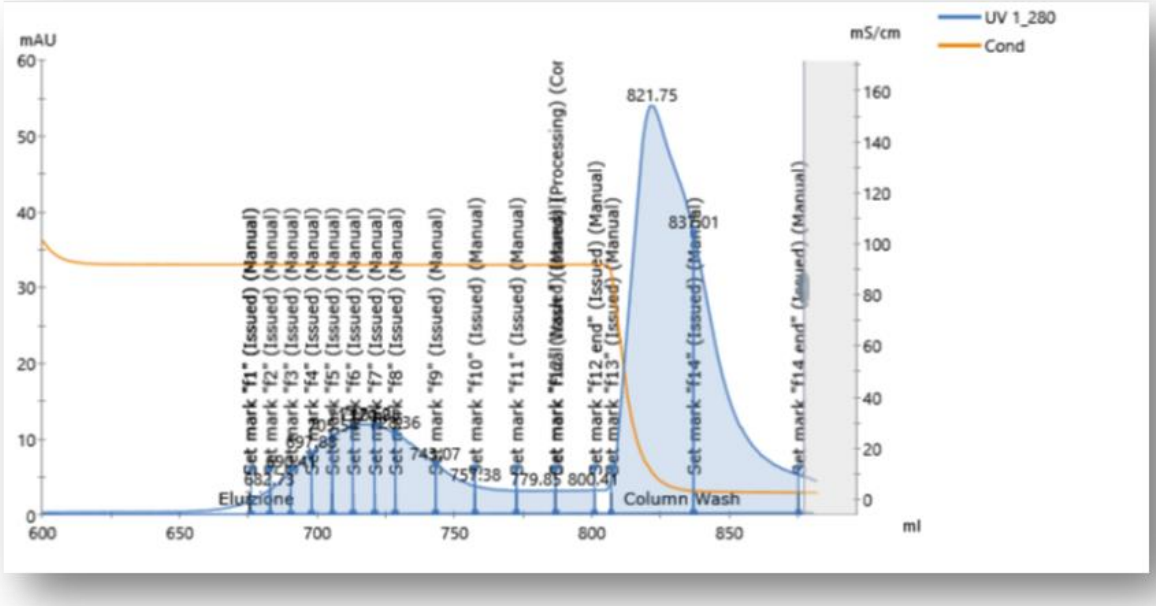


Figure 39- HIC Chromatogram. Pooled fractions from CEX added with salt, have been loaded on the Phenyl Sepharose FF low substitution to increase protein purity and to avert the other contaminants. From F1 to F 12 the protein is present, and the other peaks represent the protein aggregation.

10.12. TFF

Tangential Flow Filtration (TFF) is a method used for the separation, purification, and concentration of biomolecules. It works with a feed solution flowing in parallel to the membrane surface, which helps in reducing clogging and increases its efficiency. It is particularly useful for difficult to filtrate feeds with high solid content, as compared to traditional filtration methods.

TFF systems are available in various designs, including hollow fibers modules, and are routinely used in bioprocessing. In our case a cassette with a cut-off of 3KDa has been used, the correct size to retain rhBDNF and let the other contaminants pass through. Tangential Flow Filtration corresponds both to formulation and concentration passages, in which sample concentration and buffer exchange is performed. Six diafiltration volumes have been performed for exchanging the buffer, and homogeneity of the product has been assessed with the predefined analytical panel.

11.Conclusions

Neurotrophins are a family of closely related proteins composed of the NGF (Nerve Growth Factor), BDNF (brain-derived neurotrophic factor), NT-3 (neurotrophin-3), and NT-4/5 (neurotrophin 4-5), which are involved in survival, development and function of neurons in the central and peripheral nervous systems. They are also involved in neuroprotection, both in physiological and pathological state.

The recombinant protein has been produced to explore its potential as pharmaceutical drug, for still unmet clinical needs. BDNF is implicated in several diseases of nervous development and maintenance but is also implicated in non-neuronal diseases.

Setting up a recombinant expression technology for a robust production process to further scale it up into an industrial GMP plant, was very challenging.

The Ph.D. project has the ambitious aim to avoid inclusion bodies formation, during the production process, to directly obtain a soluble and refolded rhBDNF in the *Escherichia coli* strain, ready for the downstream process [1].

The standard neurotrophin protocol passing through inclusion bodies formation has represent the skeleton for the development of new production strategies and the team know-how have eased the develop of different approaches put in place to reach the goal.

As previously described, BDNF has a complex three-dimensional structure with three imbedded disulfide bridges, its production as a recombinant protein can be achieved both in Eukaryotic and Prokaryotic expression systems. In Eukaryotes, the protein can be produced in a soluble form, but in small quantities, whereas, in Prokaryotes, it can be obtained in large amounts, but insoluble [21].

Starting from the rhNGF canonical process , it is well known that if BDNF is produced in prokaryotes, such as *E. coli,* its structure is not correctly folded in the microorganism cytoplasm, and the recombinant protein, accumulates in an insoluble form, in inclusion bodies [83]. Once harvested the biomass needs to be disrupted and the recovered inclusion bodies must be denatured and refolded *in vitro*, before starting the downstream purification process.

So, the aim of this thesis project was to identify other strategies, able to directly produce soluble and refolded neurotrophins from E*scherichia coli* fermentations, side-stepping

the *in vitro* time-consuming refolding step. Furthermore, this downstream step requires elevated quantities of buffer(s) which normally corresponds to 3-10 times the fermentation volume and needs to be cooled to 4-6 °C to increase process efficiency. Those protocols are difficult to handle, require big "clean" tanks and represent a substantial part of the downstream costs aggravated by plant occupancy [36].

The optimization of secretory recombinant protein production in *Escherichia coli* is of a great interest for research, as well as industrial settings. A multitude of optimization approaches are available, genetic engineering approaches, optimization of production process, as well as downstream processing after cell harvesting, to obtain the protein of interest in the correct compartment.

As earlier described, gene design and protein engineering have permitted to elaborate two different constructs, to explore two different strategies, to obtain a soluble neurotrophin, the first one, with a leader sequence added upstream of mature BDNF to drive its expression in the periplasmic space, and the second corresponding to the complete proneurotrophin to be transformed in the special *Escherichia coli* Shuffle strain.

Theoretically, with both strategies, we should obtain soluble and refolded protein.

In this PhD thesis, fermentation conditions and purification strategies tested, have been described. Process development starts with the expression vector design and transformation into the selected host, to produce the Research Cell Banks.

After careful RCBs characterization to check the correct recombinant protein expression, those banks have been used to seed all the fermentations to generate material, this passage has been implemented to introduce "batch to batch" consistency. Setting up an efficient fermentation protocol has taken many runs and a long time, because several experimental conditions have been tested, until fermentation was optimized to obtain a correct balance between biomass production and recombinant protein expression. A typical fermentation is divided in a batch and fed-batch phases.

The best yields have been obtained with a temperature of 20°C, during the fed-batch phase corresponding to recombinant protein expression. It consents to the host to slow down its growth and correlated protein synthesis, avoiding saturation of the Sec translocation mechanism, in the periplasmic expression. This is also true for the Shuffle strain, where low temperature gives time to the protein to refold into the cytoplasm.

This also validates the expression vector choice. Indeed, for this type of protein production a low copy number plasmid has been chosen. In effect, the pET vectors have a plasmid copy number below 10 copies per cell, whereas pUC derivatives can reach numbers as high as 300 to 600 copies per genome. This low copy number favour limited recombinant protein expression, together with a high plasmid stability, two important assets for efficient soluble recombinant protein expression.

With the periplasmic expression strategy, the mature BDNF protein is found soluble and refolded in this compartment because the leader sequence has directly transferred its production into this oxidative environment, allowing therefore disulfide bridges formation.

Notwithstanding our efforts, no rhBDNF has been translocated from the periplasmic space and detected directly in the medium. The protein is relatively small as a dimer has a molecular weight of 26 kDa, but it may be too hydrophobic, or not sufficiently soluble, to be passively “exported” into the medium.

Secretory production of proteins has significant advantages, compared to the conventional cytosolic or even periplasmic ones, both on laboratory and on large scales. Secretion into the medium would have been more advantageous, than the one into the periplasm, because protein would have been recovered directly from the clean culture medium, without any need of outer membrane permeabilization. Moreover, isolation and purification would have been simplified resulting in much lower process costs.

Whereas in the second strategy exploiting the special Shuffle strain, with its provided enzymatic kit, disulfide bridges can form directly into the cytoplasm compartment, which in our case represents the first step toward recombinant protein refolding.

These strategies avoid the tedious solubilization and refolding procedures; moreover, with the periplasmic translocation strategy and corresponding gene engineering the protein is also directly synthesized in its mature form, so enzymatic cleavage is not necessary and rhBDNF is ready to be chromatographed. Whereas with the Shuffle expression system, the protein is also soluble and refolded, but an enzymatic cleavage is still necessary to remove the Propeptide, corresponding to the BDNF dedicated chaperon, to obtain the mature and active protein.

As expected, in both strategies we can find some soluble Neurotrophin in the targeted cellular compartment and this is a positive and encouraging result, as we can expect that if the protein is soluble, it means that it has been refolded and should also be active.

This is somewhat confirmed by the observation that in those compartments we can see some insoluble BDNF, which can be separated from the refolded one by centrifugation. This means that if the protein does not refold correctly, we will obtain an insoluble and certainly inactive protein. The same phenomenon can be observed during *in vitro* refolding where, even in presence of solubilizing agents, as for example an elevated concentration of arginine, most of the material precipitate and only the refolded and biologically active part of it remains in solution.

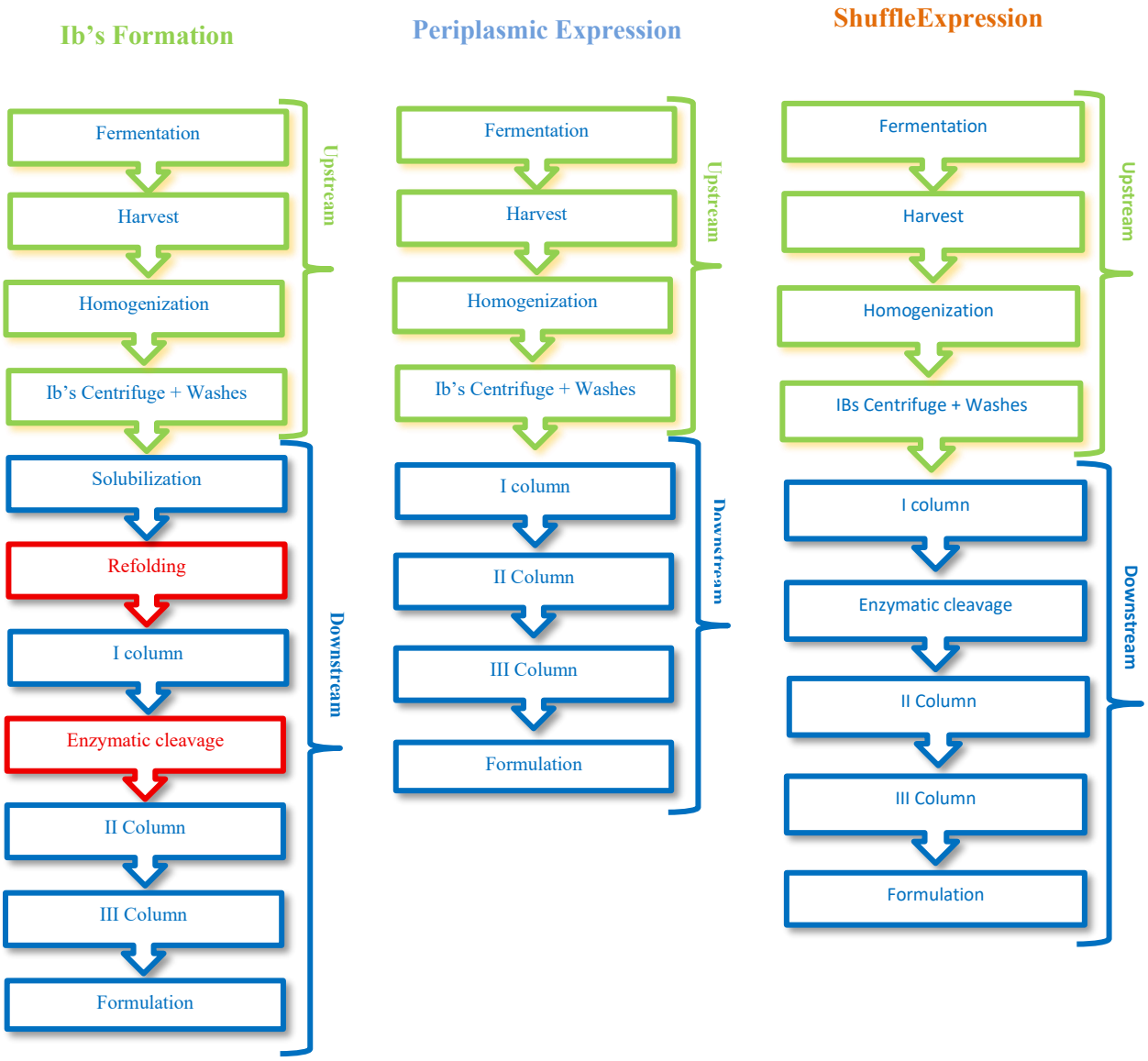


Figure 40- Flowcharts of the different processes. In red are highlighted the critical phases and in the others are reported the processes developed for this project.

The next process step was the protein extraction, but no selective pulling out strategy has been individuated and even if contaminated by the presence of DNA, HCP and endotoxin the High-Pressure Homogenization remains the only straightforward and scalable protocol. As theoretically hypothesized, after homogenization, analytical results confirmed the presence of the soluble and refolded protein in both strategies, and this material has represented the best starting one for protein purification.

But after several infructuous attempts, to enzymatically remove DNA by benzonase, the periplasm strategy was abandoned, because the Shuffle one with higher production yields

seems to be more promising. Secretion represents a limit and selective disruption of the outer membrane another one, for product recovery. Outer membrane selective and efficient permeabilization agents still needs to be identified and tested, understand and characterized to be used in a possible industrial processes.

As the goal of this research was to set-up an eventual scalable process, we have intentionally limited our investigations to techniques fulfilling regulatory requirements and scalability.

Another point to consider is that when the periplasmic expression vector was designed it has been engineered in such a way to directly express the neurotrophin in its mature form, that is rhBDNF lacking its dedicated chaperon encoded in the Propeptide part. Therefore, the minor protein expression yield in the periplasm, as compared to the cytoplasm, may be in part due to the Sec exportation limitation but also to the limited protein refolding, due to the dedicated chaperon absence. This phenomenon is also valid for the *in vitro* refolding as observed by Rattenholl et all [2], which have reported that *in vitro* 25% of ProNGF can be refolded whereas this yield falls to 1% for NGF.

Even if less downstream passages have been eliminated in the Shuffle strategy, after homogenization part of the recombinant protein was released as soluble whereas the insoluble material was eliminated by centrifugation. Furthermore, the mechanical treatment was sufficient to reduce DNA to smaller fragments, consenting the successive chromatography.

Turrax treatment was more efficient than the benzonase one, since it can rapidly reduce DNA fragments size without inducing protein precipitation as enzymatic treatment does. So, the column backpressure remains sufficiently low, and the recombinant protein has been purified on the first chromatographic column.

Due to its high hydrophobicity a good chromatographic separation has been obtained for this step, and ProBDNF was sufficiently purified to proceed with the enzymatic digestion. Most of DNA, HCP and LAL have been eliminated after this first passage.

After enzymatic digestion, two additional columns have been implemented and chromatographic conditions optimized to improve product purity and guarantee drug quality. The CEX and HIC have been developed and characterized, and at the end of the

process, final formulation and concentration have performed with TFF to produce an API bulk.

Analytical characterization has been conducted by the R&D Analytical department and batches have been released according to the R&D analytical monograph. Even if identity and purity in chemical assays have been confirmed, the biological test revealed a lower protein activity as compared to the reference standard protein obtained with the "classical" protocol starting from inclusion bodies.

This is probably due to the presence of DNA and HCP contaminants still present in the sample. LAL was not dosed as the preparation was performed in the R&D laboratories where even if we try to minimize this type of contamination by sanitizing columns and chromatographs before using them, its complete elimination is impossible as non depyrogenated bottles are used to prepare chromatographic buffers.

Limited rhBDNF activity can also result from disulfide bridges mispairing as the Shuffle strain possess an enzymatic kit which form disulfide bridges from two consecutive cysteines present in the primary protein sequence. This reaction is catalysed by DsbC enzyme expressed in the Shuffle cytoplasm, together with gor and trxB, to render the cytoplasm non reducing. Whereas in our case the neurotrophin possess a cysteine Knot composed of three imbedded disulfide bridges corresponding to a very complex situation [84]. So may be only a portion of the expressed soluble protein may have reached its correct folding and complex three-dimensional structure [69].

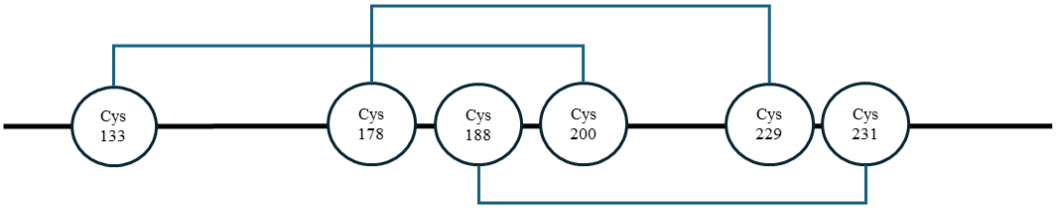


Figure 41- Schematic representation of the Cysteine positions and their pairing in the disulphide bridges. To reach its correct folding three disulphide bonds needs to be formed among non-consecutive cysteine residues, representing a complicate task for the Shuffle system.

Comparing the different yields obtain from the “classical” process, in which downstream process starts from inclusion bodies refolding, the ones obtained with the periplasmic and the Shuffle strategies are lower.

Inclusion bodies are aggregated form of overexpressed insoluble heterologous proteins, that corresponds to a pre-purification step, so representing an advantage for protein refolding. Indeed, after inclusion bodies preparation the recombinant ProBDNF is obtained with high purity and during refolding, performed in a diluted solution, no other protein interferes giving a high process yield. Instead, in the shuffle case, the protein must refold directly into the microorganism cytoplasm rich in host cell proteins.

However, defining a mild procedure to both solubilized and refolded the protein from IBs is a time-consuming procedure.

Many different strategies have been used to enhance periplasmic protein production yield in *E. coli*, using omics approach, may help to identify critical parameters that can positively and also negatively affect this strategy.

To conclude, we can say that even if the two individuated strategies to directly express BDNF, as a soluble protein in the Shuffle cytoplasm and in the periplasm have been successful, its further purification was more problematic. If the protein can effectively be produced, quantities are limited as compared to the inclusion bodies strategy, and due to its complex three-dimensional structure, the protein poorly refolds. In the periplasmatic compartment refolding is further hampered by the absence of the dedicated chaperon, whereas in the shuffle cytoplasm by the high host cell proteins concentration.

12.Perspectives

This PhD project was driven by the industrial vision, and the objective was to try to develop a faster and easier production process, to directly recover soluble and refolded rhBDNF, which was representing an important challenge. Building each passage of the process, with the aim to fill the still unmet clinical needs was very exciting, keeping in mind the noble-minded ethical aim to solve patient illness or to cure a disease for which a drug was still missing.

Collaborations with a wonderful team with huge knowledge on neurotrophins and with great technological expertise, has been much boosting and at the same time working in a research centre, equipped with new laboratories and state of the art equipment, have created the perfect environment to come out of myself and reach the perfect balance between science and technique expertise.

I make particular attention to the possibility to scale-up the process in the future, so when new experiments were designed, or new strategies developed, it was always necessary to keep in mind their implementation to the next step with high industrial scale volumes. So, it was important that each procedure could be as simple as possible, feasible and implementable in the plant.

I have had therefore the opportunity to deeply collaborate with the Dompé manufacturing teams, to understand timing and their facilities capacities and physical or regulatory limits. Further interactions have been developed with the intellectual property team, which has also represented an educational path. This team is responsible for the “idea” and company know how protections and elaborate all the documents, that a private company must produce, to defend and protect their discovery coming from the R&D department.

The results obtained along this new production process development reveal that, the two strategies achieved low protein production yields, far away from our initial aspiration. Even with its low production yield direct soluble neurotrophin production represents a successful result, as it increases the neurotrophin package knowledge and indicates the alternative process flow paths, that can be envisaged.

Furthermore, it was important to understand the scientific bases that have hampered an efficient production and purification of the target protein.

The principal contaminants are HCP and DNA are still too high to allow pharmacological administration of the purified protein even in preclinical animal models. Moreover, compared to refence standard, this material has a lower biological activity in the potency assay, lower than expected and probably due to the presence of some product-related variants, most probably composed of scrambled disulphide bridges, that cannot be discriminated from the correct neurotrophin by the developed downstream process. Complete characterization reveals that less than 80% of activity has been registered and contaminants may also influence long term product storage and stability. Now that project critical factors have been identified, further tuning needs to be put in place to try to proceed in this direction.

Taking advantage of those results and now having identified project limits, some new mutants maybe be designed, in future, to tackle the refolding and precursor digestion steps, which still represents the process bottleneck. So, research is not always a win or lose strategy, even negative results improve and increase the knowledge luggage.

In particular, this journey has been the occasion to step into gene design and recombinant proteins productions, enriched personal and professional knowledge and competences, and represents an important project part as “take home message”.

13.Abbreviations

°C	Celsius
µg	micrograms
µL	microliters
4-MEP	4-mercapto-ethyl-pyridine
aa	amino acids
AD	Alzheimer Disease
AEX	Anion Exchange Chromatography
AF	Affinity Chromatography
API	Active Pharmaceutical Ingredient
BDNF	Brain-derived neurotrophic factor
CEX	Cation Exchange Chromatography
CIP	Cleaning In Place
CNS	Central Nervous System
CPP	Critical Process Parameters
Cys	Cysteine
DCVs	Dense Core Vesicles
DNA	Deoxyribonucleic Acid
DNase	DNA protease
DO	Dissolved Oxygen
DGR	Dorsal Ganglion Root
DoE	Design of Experiment
EDG	enzymatic digestion
EDTA	ethylenediaminetetraacetic acid.
EMA	European Medicine Agency
ER	Endoplasmic Reticulum
FDA	Food and Drug Administration

FF	Fast Flow
g	grams
GF	Growth Flask Culture
GFP	Green Fluorescence Protein
GMP	Good manufacturing practices
GPRs	G-Protein Coupled Receptors
H	hours
HCDC	High Cell Density Culture
HCIC	Hydrophobic charge-induction chromatography
HCP	Host Cell Protein
HD	Huntington Disease
HEPT	Height Equivalent to a Theoretical Plate
HIC	Hydrophobic Interaction Chromatography
HOS	Higher Order structure
HP	High Performance
HPH	High Pressure Homogenization
HSP	Heat Shock Protein
Ib's	Inclusion Bodies
ICH	International Council for Harmonisation
IEX	Ion Exchange Chromatography
IM	Inner Membrane
IMAC	Immobilized metal-ion affinity chromatography
IPC	In Process Control
IPTG	Isopropil-β-D-1-tiogalattopiranoside
ISO	International Standard Operation
Kb	Kilobases
kDa	Kilodalton

LAL	Limulus Amebocyte Lysate
LC–MS	Liquid chromatography–mass spectrometry
LNGFR	Low Nerve Growth Factor Receptor
LS	Low substitution
M	Molar
mAbs	monoclonal Antibodies
mAU	milli Absorbance Ultraviolet
Mb	Megabasis
MCB	Master Cell Bank
mg	milligrams
MK	Marker
mL	milliliter
mM	millimolar
MM	Mixed mode
MOGM	Microorganisms Genetically Modified
mRNA	messenger ribonucleic acid
MW	Molecular Weight
NDs	Neurological/Neurodegenerative Diseases
NGF	Nerve growth factor
NT-3	Nneurotrophin-3
NT-4,5	Neurotrophin 4-5
NTs	Neurotrophins
OD	Optical density
OFAT	One factor at time
OM	Outer Membrane
ON	Overnight
Ori	Origin of Replication

OS	Osmotic Shock
PC	Preculture
PD	Parkinson Disease
pI	Isoelectric point
ProBDNF	Brain Derided Neurotrophic Factor precursor
PSN	Peripheral System Nervous
QP	Qualified Person
R&D	Research and Development
RBS	Ribosome Binding Site
RCB	Research Cell Bank
rh	recombinant human
rhBDNF	recombinant Human Brain Derived Neurotrophic Factor
RGC	Retinal Ganglion Cells
rhNGF	recombinant human Nerve growth Factor
rhProBDNF	recombinant Human Brain Derived Neurotrophic Factor precursor
RNA	ribonucleic acid
RP	Recombinant protein
RP-HPLC	Reversed-Phase High Performance Liquid Chromatography
rpm	rotations per minutes
RP-UPLC	Reversed-Phase Ultra Performance Liquid Chromatography
RS	Reference standard
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SH	Shuffle
SIP	Sanitization In Place
SP	Sulphopropyl
S-S	disulphide bridges

ss	signal sequence
T ind	Time of Induction
TB	Terrific Broth
TIR	Translation Initiator Receptor
TNF	Tumor Necrosis Factor
TFF	Tangential flow filtration
Trk	Tropomyosin-related kinase
UTR	Untranslated Region
WB	Western Blot
WCB	Working Cell Bank
µm	micrometers

14. Bibliography

[1] B. a. all., "A deep learnng approach to optimize recombinant protein production in Escherichia coli fermentation," *Fermentation ,* 2023.

[2] M. K. P. M. R. R. E. A. Hauburger, "Oxidative folding of Nerve growth factor can be mediated by the pro-peptide of Neurotrophin-3," *FEBS Letters 581,* pp. 4159-4164, 2007.

[3] H. F. F. S. P. J. A. S. M. N. S. a. R. M.] S. J. Mowla, "Biosynthesis and Post Translational Processing of the Precursor to brain derived neurotrophic Factor," *The Journal of Biological Chemistry,,* vol. 276, no. 16, pp. 12660-12666, 2001.

[4] W. J. F. a. L. A. Greene, "Neurotrophins Signalling via Trks and p75," *Experimental Cell Research,* vol. 253, pp. 131-142, 1999.

[5] S. M. a. Y.-A. B. S. Ateaque, "Neurotrophin signalling in the human nervous system," *Frontiers in Molecular Neuroscience,* 2023.

[6] D. E. a. H. T. Yves-Alain Barde, "Purification of a new neurotrophic factor from mammalian brain," *The EMBO Journal,* vol. 1, no. 5, pp. 549-553, 1982.

[7] P. T. P. a. N. H. W. Bai Lu, "The Ying and Yang of Neurotrophin Action," *Neuroscience,* vol. 6, 2005.

[8] M. B. a. all., "Specificity in Trk Receptor: Neurotrophins Interactions: The Crystal structure of TrkB-d5 in Complex with Neurotrophin-4/5," *Elsevier Science,* vol. 9, pp. 1191-1199, 2001.

[9] G. Thomas, "Furin at the cutting edge: from protein traffic to embryogenesis and disease," *NIH Nat Rev Mol Cell Biol,* vol. 3, pp. 753-766, 2002.

[10] H. T. F. P. F. G. L. M. A. K. D. L. F. L. G. P. V. k. A. C. a. A. P. Robert Yan, "The structure of the Pro-domain of mouse proNGF in contact with the NGF Domain.," *Cell Press,* vol. 27, pp. 79-89, 2019.

[11] R. e. all., "The Prosequence facilitates folding of uman nerve growth factor from Escherichia coli inclusion bodies," *Eur.J.Bioche, ,* vol. 268, pp. 3296-3303, 2001.

[12] R. R. J. T. S. J. P. T. A. J. L. C. G. K. R. H. T. B. J. A. M. a. D. A. Y. L.O. Narhi, "Comparison of the Biophysical Characteristic of Human Brain -Derived Neurotrophic Factor, Neurotrophin-3 and Nerve Growth Factor," *The Journal of Biological Chemistry,* vol. 268, no. 18, pp. 13309-13317, 1993.

[13] I.-C. I. M. E. V.-P. F. G.-S. E.-E. F.M.Torres -Cruz, "NT-4/5 antagonizes the BDNF modulation of cosrticostriatal transmission: Role of TrkB.T1 receptor," *Wiley CNS Neuroscience and Therapeutics ,* pp. 621-631, 2018.

[14] E. Hernàndez-Echeagaray, “The role of TrkB-T1 receptor in the neurotrophin-4/5 antagonism of brain-derived neurotrophic factor on corticostriatal synaptic transmission,” *Neural Regeneration Research,,* 2020.

[15] M. Bothwell, “Functional Interaction of Neurotrophin and Neurotrophin receptors,” *Annual Review Neurosciences,* pp. 18:223-53, 1995.

[16] R. B. a. A. Konnerth, “Neurotrophin-Mediated Rapid Signaling in the Central Nervous System: mechanism and Function,” *Physiology,* vol. 20, pp. 70-78, 2005.

[17] E. H. a. L. F. Reichardt, “ Neurotrophins: Roles in Neuronal Development and Function,” *NIH Public Access Ann. Rev. Neurosci.,* vol. 24, pp. 677-736, 2001.

[18] P. a. T. T. A.E. West, “Neurotrophin: transcription and translation,” *Handbook of Experimental Pharmacology,* vol. 220, no. Neurotrophic Factors.

[19] N. V.-A. A. C.-L. a. R. D. Carlos Hernàndez-del Cano, “Neurotrophins and Their Receptors: BDNF’s role in GABAergic Neurodevelopment and Disease.,” *International Journal of Molecular Sciences,* vol. 25, p. 8312, 2024.

[20] L. C. D. e. all., “Neurotrophin factor BDNF, physiological Functions and Therapeutic Potential in Depressio Neurodegeneration and Brain Cancer,” *International Journal of Molecular Sciences,* vol. 21, p. 7777, 2020.

[21] G. L. R. a. E. A. Ceccarelli, “Recombinant protein Expression in Escherichia coli: advances and challenges.,” *Frontiers in mIcrobiology, ,* 2014.

[22] S. I. P. R. P. A. G. S. a. C. R. B. G. Schirò, “A brief Overview on BDNF-Trk Pathways in the Nervous System: A potential Biomarker or Possible Treatment of Multiple Sclerosis,” *Frontiers in Neruology,* vol. 13, 2022.

[23] S. B. a. U. N. Das, “Brain-derived neurotrophic factor and its clinical implications,” *State of art paper,* 2014.

[24] c. S. B. J. W. D. F. H. K. a. L. B. J.J.S. Treanor, “Heterodimeric Neurotrophins Induce Phosphorylation of Trk receptors and Promotes Neuronal Differentiation in PC12 Cells,,” *The Journal of Biological Chemistry,* vol. 270, no. 39, pp. 23104-23110, 1995.

[25] P. L. R. M. F. a. S. C. S. Camuso, “Pleiotropic effects of BDNF on cerebellum and hippocampus: Implications for neurodevelopmental disorders,” *Neurobiology of Disease,* vol. 163, p. 105606, 2022.

[26] L. B. B. F. R. F. A. S. J. W. M. F. R. M. L. a. G. Y. P. C. Maisonpierre, “NT-3, BDNF and NGF in the Developing Rat Nervus System: Parallel as well as Reciprocal Patterns Of Expression,” *Neurons,* vol. 5, pp. 501-509, 1990.

[27] S. B. e. all, “Brain-derived neurotrophic factor and its clinical implications,” *Arch. Med. Sci.,* vol. 11, no. 6, pp. 1164-1178, 2015.

[28] J. C. A. a. R. Deogracias, “Mechanism Controlling the Expression and Secretion of BDNF,” *Biomolecules,* vol. 13, p. 789, 2023.

[29] H. K. T. M. t. H. Y. I. a. M. K. Koichi Uegaki, “BDNF Binds Its Pro-Peptide with High Affinity and the Common Val66Met Polymorphism Attenuates the Interaction,” *International Journal of Molecular Sciences,* vol. 18, p. 1042, 2017.

[30] J. E. S. M. G. A. L. D. C. Seyed M.S. Ghazi-Nouri, “Expression and localisation of BDNF, NT-4 and TrkB in proliferative Vitreopathy,” *Elsevier, Experimental Eye Research,* vol. 86, pp. 819-827, 2008.

[31] K. O. a. K. Noguchi, “ BDNF in sensory and chronic pain,” *Neuroscience Research,* vol. 55, pp. 1-10, 2006.

[32] a. a. I. H. T. Ho, “A novel Peptide Interfering With ProBDNF-Sortilin Interaction Alleviates Chronic Inflammatory Pain,” *Theranostics,* vol. 9, no. 6, pp. 1651-1665, 2019.

[33] T. H. H. M. A. Q. K. N. K. Y. S. O. K. T. a. L. F. P. C. Harada, “Role of Neurotrophin-4/5 in Neural Cell Death during Retinal Development and Ischemic Retinal Injury,” *nvestigative Ophthalmology & Visual Science,* vol. 46, no. 2, 2005.

[34] M. C. M. M. B. C. B. L.Buscajoni, “Refolding in the modern biopharmaceutical industry,” *Biotechnology Advances,* vol. 61, 2022.

[35] A. K. a. J. M. Reichert, “Recombinant protein therapeutics-success rates, market trends and values to 2010,” *Nature Biotechnology,* vol. 22, no. 12, 2004.

[36] I. I. a. O. Kaplan, “Escherichia coli in the production of biopharmaceuticals,” *Biotechnol Appl Biochem,* vol. 72, pp. 528-541, 2025.

[37] N. K. a. M. B. Guoping Ren, “ Use of the Shuffle Strain in Production of Proteins,” *Current Protocols in Protein Sciences,* vol. 85, pp. 5.26.1-5.56.21, 2016.

[38] J. P. a. F. M. Wurm, “Recombinant Proteins for Industrial versus Pharmaceutical Purpose: A review of Process and Pricing,” *Process ,* vol. 7, no. 476, 2019.

[39] F.-T. N. Y.-Z. W. C.-X. Y. Y. G. P. S. a. X.-M. S. Zi-Xu Zhang, “Strategies for efficient production of recombinant protein in Escherichia coli: alleviating the host burden and enhancing protein activity,” *Microbial Cell Factories ,* 2022.

[40] M. L. R. C. .. a. S. L. C. S. Chen, “ Direct and convenient measurement of plasmid stability in lab and clinical isolated of E. coli,” *Scientific Reports ,* vol. 7, 2017.

[41] D. M. F. a. R. Page, “Strategies to Optimize Protein Expression in E.coli,” *Current Protocols in Protein Science,* 2014.

[42] M. Kamionka, “Engineering of Therapeutic Protein Production in E. coli,” *Current Pharmaceutical Biotechnology,* vol. 12, pp. 268-274, 2011.

[43] A.-H. R. M. A. H. R. K. N. B. a. E. R. M.N.Baeshen, " Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives," *J. Microbiol. biotechnol ,* vol. 25, no. 7, pp. 953-962, 2015.

[44] S. C. Markrides, "Strategies for Achieving High-Level Expression of Genes in Escherichia coli," *Microbiological Reviews,* pp. 512-538, 1996.

[45] K. A. M. W. Z. a. D. D. P.J. Shilling, "P.J. Shilling, K.Mirzadeh, A.J.Cumming, M. Widesheim, Z.Kock, and D.O. Daley, Improved design for pET expression plasmids increase protein production yield in Escherichia coli," *Communication Biology,* vol. 3, p. 214, 2020.

[46] I. I. a. O. Kaplan, "Escherichia coli as a versatile cell factory: advances and challenges in recombinant protein production," *Protein Expression and Purification 219 (2024), 116463. ,* vol. 219, 2024.

[47] J. E. N. C. G. J. M. a. S. G. A.Villalobos, "Gene Designer: a synthetic biology tool for constructing artificial DNA segments," *BMC Bioinformatics,* vol. 7, p. 285, 2006.

[48] J. M. K. F. G.R.M. Kleiner-Grote, "G.R.M. Kleiner-Grote, Joe M.Risse, K. Friehs, Secretion of recombinant proteins from E. coli Engeneering in life science 2018,18, 532-550.," *Engeneering in life science,* vol. 18, pp. 532-550, 2018.

[49] V. A. K. U. S. S. a. A. K. P. A.Singh, "Protein recovery from Inclusion bodies of Escherichia coli using mild solubilization process," *Microbial Cell Factories,* vol. 14, no. 41, 2015.

[50] A. a. P.] M.Krause, " The fed-batch principle for the molecular biology lab: controlled nutrient diets in ready-made media improve production of recombinant proteins in Escherichia coli," *Microbial Cell Factories,* vol. 15, p. 110, 2016.

[51] S. K. J. P. a. S. K.Kachhawaha, "Bioprocessing of recombinant proteins from Escherichia coli inclusion bodies: Insight from structure-function relationship for novel applications," *Preparative Biocheistry &Biotecnology,* 2022.

[52] C. X. R. W. W. J. J. W. . Su, "A novel strategy for enhancing extracellular secretion of recombinant proteins in Escherichia coli," *Appl.Microbiol. Biotechnol. ,* 2013.

[53] C. T. C. O. D. J.Wurm, "How to trigger periplasmic release in recombinant Escherichia coli: a comparative analysis, 2017,17, 215-222.," *Engeneering in life science,* vol. 17, pp. 215-222, 2017.

[54] N. R. S. Samuel I.Miller, "The gran-negative bacterial periplasm: size metters," *Plos Biology,* 2018.

[55] C. W. a. A. Pluckthun, "C. Wulfing and Andreas Pluckthun, Protein folding in the periplasm of Escherichia coli, Molecular Microbiology (1994), 12 (5), 685-692.," *Molecular Microbiology,* vol. 12, no. 5, pp. 685-692, 1994.

[56] A. K. a. J. W. d. G. Gier, “Strategies to Ehnance Periplasmic Recombinant Protein Production Yield in Escherichia coli,” *Frontiers in Bioengineering and Biotechnology,* vol. 9, 2021.

[57] L. S. S. K. B. A. R. P. B. R. V. P.Singh, “ Effect of signal Peptide on stability and folding of Escherichia coli Thioredoxin,” *Plos One ,* 2013.

[58] E. C. G. S. C. M.-P. R. C.Schimek, “Extraction of recombinant periplasmic proteins under industrially relevant process conditions: selectivity and yield strongly depend on protein titer and methodology,” *Biotechnology Progress,* 2020.

[59] I. P. T. S. Y. R. T. R. M. Y. M. F. L. D. T. R. R. A. S.A.F. Kusuma, “Real-time monitoring of Rhamnose induction effect on the expression of mpt64 gene fused with pelB signal peptide in Escherichia coli BL21 (DE3),” *Jour of Advanced Pharmaceutical Technology & Research,* 2020.

[60] P. S. R. A. C. H. M. M. N. a. D. K.Mirzadeh, “Increase production of periplasmic proteins in Escherichia coli by directed evolution of the translation initiation region,” *Microb Cell Fact,* vol. 19, no. 8, 2020.

[61] R. Freudl, “Signal peptides for recombinant protein secretion in bacterial expression system,” *Microbial Cell Factories ,* vol. 17, no. 52, 2018.

[62] S. G. W. A. H. J. C. D. S. a. T. O. T.S. Castineiras, “Development of a generic B-lactamase screening system for improved signal peptides for periplasmic targeting of recombinant protein in Escherichia coli,” *Scientific Reports ,* vol. 8, 2018.

[63] R. Freudl, “Signal peptidefor recombinnat protein secretion in bacterial expression system,” *Microbial Cell Factories ,* 2018.

[64] J. H. C. &. S. Y. Lee, “Secretory and extracellular production of recombinnat proteins using E.coli.,” *Applied Microbiology and Biotechnology ,* vol. 64, pp. 625-635, 2004.

[65] O. T. I. M. Kasli, “ Use of a design of experiments approach to optimize production of a recombinant antibody fragment in the periplasm of Escherichia coli: selection of signal peptide and optimal growth conditions,” *AMB Express ,* vol. 9, no. 5, 2019.

[66] M. Vaara, “Agents That increase the permeability of the other membrane,” *Microbiological reviwews,* pp. 395-411, 1992.

[67] S. Y. a. Y. F. T. Katsu, “ Increases in permeability of Escherichia coli outer membrane induced by polycations,” *Elsevier Science ,* vol. 166, no. 1, 1984.

[68] Q. W. S. J. Y. Z. G. Z. Y. M. L. Huang, “Improved extracellular expression and high-cell-density fed-batch fermentation of chitosanase from Aspergillus Fumigatus in Escherichia coli,” *Bioporcess Biosyst Eng,* 2016.

[69] M. Berkmen, “Production of disulfide-bonded proteins in Escherichia coli,” *Protein Expression and purification ,* vol. 82, pp. 240-251, 2012.

[70] d. B. a. M. B. Bruno Mata, “ Disulfide Bond formation in the Periplasm of Escherichia coli,” *EcoSalPlus ,* vol. 10, 2019.

[71] C. A. E. C. J. M. F. P. R. a. M. B. J. Lobstein, “ Shuffle, a novel Escherichia coli protein expression strain capable of correctly refolding disulfide bonded proteins in its cytoplasm,” *Microbial Cell Factories ,* vol. 11, no. 56, 2021.

[72] Y. U. M. J. S. a. L. W. R. Angel Castillo-Corujo, “Escherichia coli Cytoplasmic Expression of Disulfide Bonded Proteins: Side-by-Side Comparison between Two Competing Strategies,” *J.Microbiol.Biotechnol,* vol. 34, no. 5, pp. 1126-1134, 2024.

[73] S. K. N. a. M. Magoola, “Advances in Escherichia coli-Based Therapeutic Protein Expression: Mammalian Conversion, Continuous Manufacturing, And Cell-Free Production,” *Biologics,* vol. 3, pp. 380-401, 2023.

[74] v. O. S. A. Pekarsky, “The impact of technical failures during cultivation of an inclusion body process,” *Bioprocess and Biosystem Engineering,* vol. 42, pp. 1611-1624, 2019.

[75] F. M. r. A. E. A. E. D. M. a. A. A. M.Aghaeepoor, “M.Aghaeepoor, F.Kobarfard, M. reza Akbari Eidgahi, A.Optimization o culture media for extracellular expression of streptokinase in Escherichia coli using response surface methodology in combination with Plackett-Burman Design,” *Tropical Journal of Pharmaceutical research ,* vol. 16, pp. 2567-2576, 2017.

[76] C. D. S. J. O. a. C. H. J. Kopp, “Inclusion Body Bead Size in E.coli Controlled by physiological Feeding,” *Microorganism,* vol. 6, no. 116, 2018.

[77] P. H. a. J. A. M.Akesson, “ Avoiding Acetate Accumulation in Escherichia coli cultures using Feedback Control of glucose feeding,” *Wiley,* 2000.

[78] S. L. S. L. Y. C. H. N. C.] B.S. Kim, “ High Cell Density fed-batch cultivation of Escherichia coli using exponential feeding combined with pH-stat,” *Bioprocess Biosyst Eng,* vol. 26, pp. 147-150, 2004.

[79] D. P. i. t. B. I. M. B. V. 2. 2. Monahar Kalyanpur, “Downstream Processing in the Biotechnology Industry,” *Molecular Biotechnology,* vol. 22, 2002.

[80] J. T. a. Y. L. Shuo Tang, “Challenge and Solution for the downstream purification of therapeutic proteins,” *Antibody Therapeutics ,* vol. 7, no. 1, pp. 1-12, 2024.

[81] F. K.-B. O. T. R.] Laura A. Palomares, “ndustrial Recombinant Protein Production,” *Biotechnology,* vol. 5.

[82] M. a. L. E.Vandermarliere, “Getting intimate with trypsin, the leading protease in proteomics,” *Mass Spectronomy Review,* vol. 32, pp. 453-465, 2013.

[83] N. G.-M. E.-F. a. A. M.Martìnez-Alonso, “Learning about protein solubility from bacterial inclusion bodies,” *Microbial Cell factories,* vol. 8, no. 4, 2009.

[84] C. F. Ibanez, “Emerging themes in structural biology of neurotrophic factors,” *TINS,* vol. 21, no. 10, 1998.

[85] L. e. all., “A p75 neurotrophin receptor-sparing nerve growth factor protects retinal ganglion cells from neurodegeneration by targeting microglia,,” *British Journal of Pharmacology ,* pp. 1-30, 2024.

[86] S. M. a. Y.-A. B. Sarah Ateaque, “Neurotrophin signalling in the human nervous system,” *Frontiers in Molecular science,* 2023.

[87] V. P. a. F. H. Falcone, “Polyionic Tags as Henhancers of Protein Solubility in Recombinant Protein Expression,” *Microorganisms,* vol. 6, no. 47, 2018.