ARABIAN GULF UNIVERSITY

College of Graduate Studies


Biotechnology Program

# Cloning and Expression of the Human *Interferon-β* Gene from Codon-Optimized Sequences Using the Yeast *Pichia pastoris*



A Thesis Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Medical Biotechnology

**Submitted by**

**Noora Rashed Abdulla Al Senan**

Bachelor of Science in Molecular Biology – Biochemistry, University of Kuwait, State of Kuwait, 2008

**Supervised by**

**Dr. Mohammed Dahmani Fathallah**
Professor of Biotechnology
Arabian Gulf University

**Dr. Sonia Bourguiba-Hachemi**
Assistant Professor of
Molecular and Cellular Biology
Arabian Gulf University

KINGDOM OF BAHRAIN

June 2012 (A.D.)

Sha'ban 1433 (A.H.)

**وَتِلْكَ الْأَمْثَالُ نَضْرِبُهَا لِلنَّاسِ وَمَا يَعْقِلُهَا إِلَّا الْعَالِمُونَ**

**العنكبوت 43**

## *Dedication*

*Deepest love and affection to my parents;*

*Their support and motivation have inspired me to be the best I can be*

## Acknowledgments

First of all, I would like to express my gratitude to Allah, for indulging me with science and knowledge, having faith in his graciousness made me continue till the end of my study, which is the beginning of my science journey. My acknowledgements for the Ministry of Interior, to H.H Sheikh Rashid bin Abdullah Al-Khalifa for allowing me to follow my ambition and granted me a leave to be a full-time researcher.

My sincere appreciation to my supervisors, Professor Mohammed Dahmani Fathallah, Dr. Sonia Bourgiba-Hachemi. Words are not enough to show my gratitude to your guidance and support in my research. You embraced me with your knowledge and allowed me to work and learn from my mistakes. Thank you for your patience and motivation.

I am grateful for the staff of Biotechnology Department in Arabian Gulf University, including Dana Ashour, Ahmed Rasmi Ramadhan, Ashraf Al-Nayal, Nasser A. Haleem and Suhaila Al-Mehri for their support and assistance in many unspoken ways. Special thanks to Dr. Wael Almoslimany and Dr. Mohammed Shahid for their technical advices and for their warm welcome anytime in their offices. I never felt less than a member among the family of biotechnology department. I would like to thank Ali Al-Mahmeed and Maryam Murtadha whom endlessly offered their help and support through different stages of my research.

Many thanks to my colleagues, most profoundly: Dana Ashour, Mariam Marzouq, Fatima Jaffer, Buthaina Salim, Isra Salim and Budor Nasser for their support and motivation in any way possible and for sharing with me their thoughts, ideas and concerns. My appreciation and love beyond measures go to my friend Dalal Al-Awadhi, who had faith in me and in my capabilities and pushed me to do better and move forward. In the end, I would like to show my eternally gratefulness to my beloved family, especially mom and dad for encouraging me to do my best and my truthful gratitude to my husband Mohammed Al-Qallaf and my daughter Amina for moral support and their patience with my unavailability.

**Cloning and Expression of the Human *Interferon-β* Gene from Codon-Optimized Sequences Using the Yeast *Pichia pastoris***

Noora Rashed Abdulla Al Senan

**Supervised by**
Prof. M. Dahmani Fathallah and Dr. Sonia Bourguiba-Hachemi

## Abstract

Multiple Sclerosis (MS) is a serious life-threatening neurodegenerative disorder whose prevalence seems to be increasing in the Arabian Gulf region. The prevalence of MS in Bahrain is 47 per 100,000 people which puts the country in the group of intermediate prevalence of the disease. Whilst there is currently no cure for MS, one of the recent and most successful therapeutic treatments of MS is recombinant human interferon-beta (IFN-β). Recent studies have shown that IFN-β based drugs achieve their beneficial effect on MS progression *via* their anti-inflammatory properties and were shown to be well-tolerated when compared to other drugs. Since most of the bio drugs are currently recombinant proteins, the most commonly used recombinant proteins production systems are namely *Escherichia coli* and its yeast surrogates *Saccharomyces cervisiae* and *Pichia pastoris*.

The work outlined herein consists in using a yeast based expression system and a gene optimization method focused on codon-tuning, to develop a recombinant clone that produce human IFN-β. We used two different codon tuning software's namely GASCO and JCat, to design and synthesize two IFN-β cDNA with optimized codon sequences. We used these two synthetic IFN-β cDNA sequences along with the native sequence to generate recombinant *P. pastoris* strain Mut$^{S}$ KM71 clones, select multiple integrants clones and explore recombinant IFN-β expression. Neither the native sequence, nor GASCO optimized sequence resulted in abundant IFN-β production by the methylotrophic yeast *Pichia pastoris.* The expression level was below the threshold of detection by western blot. However, analytical scale experiments using the recombinant *P. pastoris* clone that integrated the JCat software optimized IFN-β cDNA sequence, generated recombinant human IFN-β proteins of approximately 19 and 33 kDa molecular mass. The specificity of this recombinant IFN-β cDNA was confirmed by western blotting using a specific anti human IFN-β monoclonal antibody. Our data

indicate that we generated a recombinant *P. pastoris* yeast clone which produces secreted and soluble recombinant human IFN-β. Investigation of the biological activity and pilot scale production of this new recombinant human IFN-β could lead to the industrial production of this major bio drug. The strategic aim of mastering bio drugs production technology is to help initiate a biopharmaceutical industry in the region, lower the cost and improve patients` access to this new class of highly efficient drugs. In addition, this will facilitate the development of bio drugs tailored to the genetic make-up of patients from the Arabian Gulf region.



**Keywords:** Multiple sclerosis (MS), gene optimization, GASCO software, JCat software, Interferon-Beta, biopharmaceuticals, Yeast *Pichia pastoris*.

# Table of contents

## List of figures

## List of tables

## List of abbreviations

| | |
|---|---|
| **AOX** | Alcohol Oxidase Genes |
| **APS** | Ammonium Persulfate |
| **ATP** | Adenosine Triphosphate |
| **BBB** | Blood-Brain Barrier |
| **BEDS** | Bicine-NaOH, Ethylene Glycol, Dimethyl sulfoxide and sorbitol solution |
| **BMMY** | Buffered minimum methanol medium |
| **Bp** | Base pair (Nucleotide count) |
| **BSA** | Bovine Serum albumin |
| **$CaCl_2$** | Calcium Chloride |
| **CAI** | Codon Adaptation Index |
| **CARDs** | Caspase recruitment domains |
| **CD** | Cluster designation |
| **cDNA** | Complementary DNA |
| **CFU** | Colony Forming Unit |
| **CHO** | Chinese Hamester Ovaries |
| **CIS** | Clinically Isolated Syndrome |
| **CNS** | Central Nervous System |
| **CpG** | Cytosine-phosphate-Guanine |
| **CSF** | Cerebrospinal fluid |
| **DEPC** | Diethylpyrocarbonate |
| **DMD** | Disease Modifying Drugs |
| **DMSO** | Dimethylsulfoxide |
| **dsRNA** | Double strand RNA |
| **DTT** | Dithiothretol |
| ***E. coli*** | *Escherichia coli* |

| | |
|---|---|
| **EF-Tu** | Elongation factor Tu |
| **eIF-2** | Eukaryotic Translation Initiation Factor-2 |
| **ER** | Endoplasmic Reticulum |
| **FADD** | Fas-associated death domain |
| **FDA** | Food and drug administration |
| **Fop** | Frequency of optimal codons' |
| **GASCO** | Genetic algorithm simulation for codon optimization |
| **GTCA** | **G**raphical **T**ool for **C**odon **A**nalysis |
| **HCl** | Hydrochloric Acid |
| **IFN** | Interferon |
| **IFNAR** | Interferon-alpha receptor |
| **IFNGR** | Interferon-gamma receptor |
| **IFN-α** | Interferon-Alpha |
| **IFN-β** | Interferon-Beta |
| **IFN-γ** | Interferon-Gamma |
| **IFN-λ** | Interferon-Lambda |
| **IFN-ω** | Interferon-Omega |
| **IL** | Interleukins |
| **IRFs** | IFN regulatory factors |
| **ISG** | IFN-stimulated gene |
| **JAK** | Janus-family tyrosine kinases |
| **JCat** | Java codon adaptation tool |
| **KDa** | Kilo Dalton |
| **LB** | Luria-Bertani |
| **$MgCl_2$** | Magnisium Chloride |
| **MHCI** | Major histocompatibility complex I |

| | |
|---|---|
| **MHCII** | Major histocompatibility complex II |
| **mRNA** | Messenger RNA |
| **MS** | Multiple Sclerosis |
| **MW** | Molecular weight |
| **Nc** | Effective number of codons |
| **NK** | Natural Killer Cells |
| **N-terminal** | Nitrogen terminal |
| **O.D** | Optical density |
| **Opt.** | Codon optimized |
| **ORF** | Open reading frame |
| ***P. Pastoris*** | Pichia pastoris |
| **PC-TP** | Phosphatidylcholine transfer protein |
| **PKR** | Protein kinase R |
| **PVDF** | Polyvinylidene fluoride |
| **RAMOS** | Respiration Activity Monitoring system |
| **RCA** | Relative Codon Adaptation |
| **rDNA** | Recombinant DNA |
| **RE** | Restriction Enzyme |
| **REBASE** | The Restriction Enzyme Database |
| **RIP1** | Receptor-interacting protein 1 |
| **RRMS** | relapsing-remitting Multiple Sclerosis |
| **rRNA** | Ribosomal RNA |
| **RSCU** | Relative Synonymous Codon Usage |
| **RT-PCR** | Reverse Transcriptase Polymerase Chain Reaction |
| ***S. cerevisiae*** | Saccharomyces cerevisiae |
| **SDS-PAGE** | Sodium dodecyl sulphate – Polyacrylamide gel electrophoresis |
| **SHP-1** | SH2-containing tyrosine phosphase-1 |

| | |
|---|---|
| **snRNA** | Small nuclear RNA |
| **STAT** | Signal transducer and activator of transcription |
| **TAP** | Transporter associated with antigen processing |
| **TBK1/IKKi** | TANK-binding kinase 1/IκB kinase i |
| **TBS** | Tris buffered Saline |
| **TBST** | Tris buffered Saline, Tween 20 |
| **TCR** | T cell receptor |
| **Th1** | T helper cells |
| **TLR3** | Toll Like Receptor |
| **TRAF6** | TNF receptor–associated factor 6 |
| **TRIF** | Toll/IL-1R homology domain–containing adaptor-inducing IFN-γ |
| **tRNA** | Transfer RNA |
| **VLPs** | Viral Like Particles |
| **VRE** | virus inducible region |
| **YPD** | Yeast peptone dextrose (yeast media) |
| **YPDA** | Yeast peptone dextrose agar |

## Chapter 1: Introduction

Just 50 years after the discovery of the double helix structure of DNA by Watson and Crick (1953), new developments have revolutionized the pharmaceutical industry. A range of newly developed drugs has been produced, based on the applications of biotechnological and gene technological methods (Gellissen, 2005).

### 1.1 Biopharmaceuticals

Most of the drugs are made of proteins, which are essential to all living organisms for structure, regulation of body functions. Protein development in a cell begins with DNA transcribing into RNA, and RNA translating into proteins. Biopharmaceuticals are medical drugs produced using biotechnology. They include proteins (such as antibodies) and nucleic acids (DNA, RNA or antisense oligonucleotides). They can be used for therapeutic or diagnostic purposes and they are characterized by the ability to treat a number of severe diseases that lead to remarkable spread and high marketing sale. Biopharmaceuticals can be produced by means other than direct extraction from a native (non-engineered) biological source, such as recombinant DNA technology (rDNA) and selecting the appropriate expression vector (Walsh, 2003).

Biopharmaceuticals are considered to be a safe substitute from the synthetic drugs. However, due to their long years of development and production, the cost of biopharmaceutical drugs is very expensive (Bos *et al*., 2001) as compared to chemical drugs production. The challenges and concerns of biopharmaceutical production are the eliciting of unexpected and detrimental side effects (i.e. immunogenicity) and lacking the expected efficacy. Some common examples of biopharmaceuticals include Enbrel, Aranesp, Epogen, Neupogen, Gaminex, recombinant human insulin, and human growth hormone, to name a few. Table 1, illustrates some of the biopharmaceuticals and their applications.

**Table 1:** Biopharmaceuticals and their applications (Bos *et al*., 2001)

| Biopharmaceuticals | Applications |
|---|---|
| Erythropoietin | Treatment of anaemia |
| Interferon-α | Treatment of leukaemia |
| Interferon-β | Treatment of multiple sclerosis |
| Monoclonal antibody | Treatment of rheumatoid arthritis |
| Colony stimulating factors | Treatment of neutropenia |

| | |
|---|---|
| Glucocerebrosidase | Treatment of Gaucher's disease |

## 1.2 Interferon family

One of the highly reputable examples of biopharmaceutical is the interferons. The interferons (IFNs) represent an evolutionary conserved family of secreted proteins that participate as extra-cellular messengers in a wide variety of responses, including antiviral, antiproliferative and immunomodulatory and developmental activities that act to maintain homeostasis and in host-defense (Tyring, 1995; Dusheiko, 2003). IFNs belong to the large class of glycoproteins known as cytokines and they were originally named after their ability to "interfere" with viral replication within host cells. They are released by host cells in response to the presence of pathogens such as viruses, bacteria, parasites or tumor cells and allow for communication between cells to trigger the protective defenses of the immune system that eradicate pathogens or tumors. IFNs have other functions for example:

- They activate immune cells, such as natural killer cells and macrophages;
- They increase recognition of infection or tumor cells by up-regulating antigen presentation to T lymphocytes;
- They increase the ability of uninfected host cells to resist a new virus infection.

Certain host symptoms, such as aching muscles and fever, are related to the production of IFNs during infection.

## 1.3 History of interferon

During research to produce a more efficient vaccine for smallpox, Yasu-ichi and Yasuhiko (1954) noticed inhibition of viral growth in an area of rabbit-skin or testis previously inoculated with UV-inactivated virus. They hypothesized that some "viral inhibitory factor" was present in the tissues infected with virus and attempted to isolate and characterize this factor from tissue homogenates. These findings were published in a French journal now known as the “Journal de la Société de Biologie” (Nagano and Kojima, 1954). After Nagano and Kojima separated the viral inhibitory factor from the viral particles using ultracentrifugation, they confirmed its antiviral activity lasted 1 – 4 days and did not result from antibody production; their findings were published in 1958 (Ozato *et al*., 2007). Nagano’s work was never fully appreciated in the scientific community; possibly because it was printed in French, but also because his *in vivo*

system was perhaps too complex to provide clear results in the characterization and purification of interferon (International Society For Interferon And Cytokine Research, 2005). Meanwhile, the British virologist Alick Isaacs and the Swiss researcher Jean

Lindenmann, at the National Institute for Medical Research in London, noticed an interference effect caused by heat-inactivated influenza virus on the growth of live influenza virus in chicken egg chorioallantoic membrane. They published their results, attaining wide recognition, in 1957 (Isaacs and Lindenmann, 1957). In this paper they coined the term "interferon", and today that specific interfering agent is known as a "Type I interferon" (Mergiran, 1980). The majority of the credit for the discovery of the interferon went to Isaacs and Lindenmann (International Society For Interferon And Cytokine Research, 2005). The purification of interferons did not occur until 1978. A series of publications from the laboratories of Sidney Pestka and Alan Waldman between (1978 and 1981) describe the purification of the type I interferons IFN alpha (IFN-α) and IFN-beta (IFN-β) (Pestka, 2007). By the early 1980s, the genes for these interferons were cloned allowing—for the first time—definitive proof that interferons really were responsible for interfering with viral replication (Weissenbach *et al*., 1980; Taniguchi *et al*., 1980). Gene cloning also confirmed that IFN-α was encoded by, not one gene, but a family of related genes (Nagata *et al*., 1980a). The type II IFN which is interferon-gamma (IFN-γ) gene was also isolated around this time (Gray and Goeddel, 1982). IFN was scarce and expensive until 1980 when the interferon gene was inserted into bacteria using recombinant DNA technology, allowing mass cultivation and purification from bacterial cultures or derived from yeast (e.g. Reiferon Retard is the first yeast derived interferon-alpha 2a, Nagata *et al.*, 1980b).

### 1.4 Interferon types

About ten distinct IFNs have been identified in mammals; seven of these have been described for humans. They are typically divided among three human IFN classes based on the type of receptor through which they signal: Type I IFN, Type II IFN, and Type III IFN:

- Interferon type I: All type I IFNs bind to a specific cell surface receptor complex known as the IFN-α receptor (IFNAR) that consists of IFNAR1 and IFNAR2 chains (De Weerd *et al*., 2007). The type I interferons present in humans are IFN-α, IFN-β and interferon-omega (IFN-ω) (Liu, 2005).

- Interferon type II: Binds to IFN-γ receptor (IFNGR) that consists of IFNGR1 and IFNGR2 chains. In humans this is IFN-γ.
- Interferon type III: This type represents interferon lambda (IFN-λ). Signal through a receptor complex consisting of IL10R2 (also called CRF2-4) and IFNLR1 (also called CRF2-12). Acceptance of this classification is less universal than that of type I and type II, and unlike the other two, it was not included in Medical Subject Headings (Vilcek, 2003).

IFNs belonging to all IFN classes are very important for fighting viral infections. IFNs are classified as helical cytokines and are categorized as type I or type II according to their physical and functional characteristics.



### 1.5 Function of interferon

All interferons share several common effects; they are antiviral agents and can fight tumors. As an infected cell dies from a cytolytic virus, viral particles are released that can infect nearby cells.

However, the infected cell can warn neighboring cells of a viral presence by releasing interferon (Tuder and Yun, 2008).

The neighboring cells, in response to interferon, produce large amounts of an enzyme known as protein kinase R (PKR). This enzyme phosphorylates a protein known as eIF-2 in response to new viral infections; eIF-2 is a eukaryotic translation initiation factor that forms an inactive complex with another protein, called eIF2B, to reduce protein synthesis within the cell.

Another cellular enzyme, RNAse L—also induced following PKR activation—destroys RNA within the cells to further reduce protein synthesis of both viral and host genes. Inhibited protein synthesis destroys both the virus and infected host cells (Figure 1, Tuder and Yun, 2008).

Interferons, such as interferon gamma (IFN-γ), directly activate other immune cells, like macrophages and (NK) cells. Production of interferons predominantly occurs in response to viruses and bacteria, and their products. Binding of molecules uniquely found in microbes—viral glycoproteins, viral RNA, bacterial endotoxin (lipopolysaccharide), bacterial flagella, CpG motifs--by pattern recognition receptors, such as membrane bound Toll like receptors or the cytoplasmic receptors RIG-I or MDA5, can trigger release of IFNs (Haller *et al*., 2007). Toll Like Receptor 3 (TLR3) is important for inducing interferon in response to the presence of double-stranded RNA viruses; the ligand for this receptor is double-stranded RNA (dsRNA).

After binding dsRNA, this receptor activates the transcription factors IRF3 and NF-κB, which are important for initiating synthesis of many inflammatory proteins. Release of IFN from cells is also induced by mitogens. Other cytokines, such as interleukin 1, interleukin 2, interleukin-12, tumor necrosis factor and colony-stimulating factor, can also enhance interferon production (Haller *et al*., 2007).

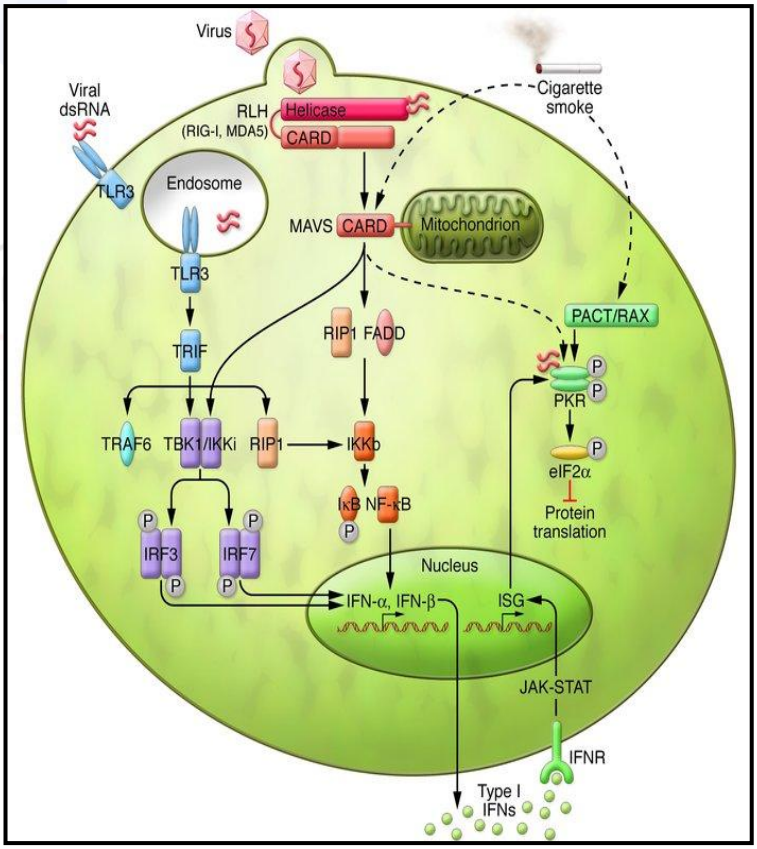



**Figure 1:** Type I IFN production and PKR activation triggered by TLR3 or the RLH system by natural or synthetic dsRNA. Activation of TLR3 by viral dsRNA leads to the recruitment of Toll/IL-1R homology domain–containing adaptor-inducing IFN-γ (TRIF), which activates TNF receptor–associated factor 6 (TRAF6), TANK-binding kinase 1/IκB kinase i (TBK1/IKKi), and receptor-interacting protein 1 (RIP1). TBK1 phosphorylates the IFN regulatory factors (IRFs) 3 and 7. Cytoplasmic dsRNA is also recognized by the RLH system, which interacts with MAVS via caspase recruitment domains (CARDs). MAVS recruits IKK and TBK1, and this converges with TLR3 signaling and leads to NF-κB activation and the induction of type I IFNs (e.g., INF-α and INF-β). MAVS also activates NF-κB through Fas-associated death domain–containing protein (FADD) and RIP1 interaction. The binding of type IIFNs to their receptor (IFNR) causes JAK/STAT-mediated synthesis of IFN-stimulated gene (ISG) products, including PKR. Although it is unclear whether PKR may be directly stimulated by the interaction of MAVS and dsRNA (as indicated by the dashed arrow), the endogenous protein PACT/RAX can also activate PKR independently of dsRNA. Cigarette smoke may (as reflected by dashed arrows) affect lung cell responses to RNA viruses by enhancing MAVS-PKR signaling and therefore trigger alveolar cell death. (Tuder and Yun, 2008).

## 1.6 Interferon-beta

IFN-β gene (777bp) is located in chromosome 9p22 encoding for a globular protein. It consists of 5 helices of 187 amino acids (Figure 2).

It has a calculated MW of 20 kDa, though it often runs on SDS-PAGE gels with an apparent MW closer to 25 kDa due to glycosylation (Arduini *et al*., 1999).

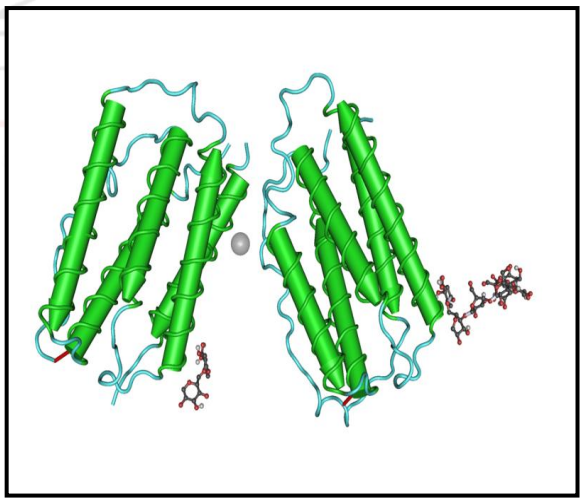


**Figure 2:** Structure of IFN-β (http://www.ncbi.nlm.nih.gov/Structure).

### 1.6.1 IFN-β signal transduction

Cellular response to IFN-β is mediated by the IFN receptor, which is found on many different cell types (Biron, 1998). The receptor is a heterodimer formed by IFNAR1, a 550 residue integral membrane protein, and IFNAR2, a 487 residue integral membrane protein. The extracellular domains of IFNAR1 and 2 are composed of several fibronectin III-like repeats (IFNAR1 has 4 repeats, IFNAR2 has only 2; Arduini *et al*., 1999). In the presence of IFN-β, the two chains assemble into a functional receptor complex (Figure 3), which initiates the signal transduction pathway (Russell-Harde *et al*., 1999).

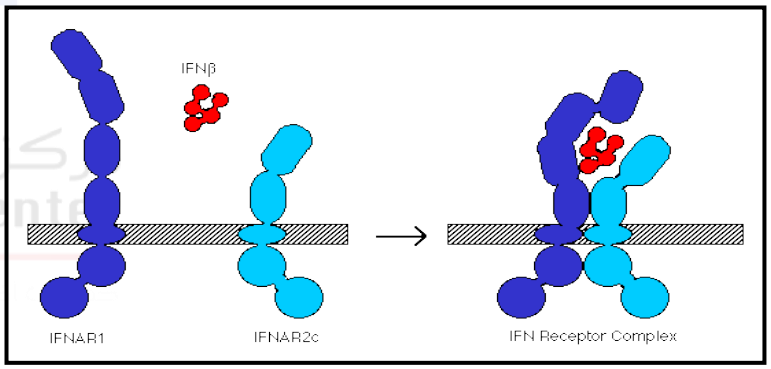



**Figure 3:** Model for interaction of IFN-β and the type I IFN receptor components, IFNAR1 and IFNAR2c, to form a complete receptor complex (Russell-Harde *et al.*, 1999).

Upon assembly of the IFN receptor complex, the intracellular domains of IFNAR1 and IFNAR2 associate with two Janus-family tyrosine kinases, JAK1 and Tyk2, which transphosphorylate themselves and phosphorylate the receptors. The phosphorylated IFNAR1 and IFNAR2 then bind to signal transducer and activator of transcription 1 and 2 (STAT1) and (STAT2). The STAT proteins then dimerize and migrate to the nucleus where they activate transciption of multiple genes (Arduini *et al.*, 1999). Downregulation of the JAK/STAT pathway appears to be effected by the tyrosine phosphatase SH2-containing tyrosine phosphase-1) (SHP-1) (Min *et al.*, 1998).

Interestingly, IFN-α and IFN-β share the same membrane receptor yet elicit different cellular responses. Russell-Harde *et al.* (1999) showed that IFN-β binds to IFNAR1 and IFNAR2 more stable than does IFN-α. The stability of the IFN receptor complex is thus postulated to have an effect on the nature of subsequent signal transduction, allowing IFN-α and IFN-β to elicit different responses via the same receptor.

### 1.7 Clinical applications of IFN-β

### 1.7.1 Multiple sclerosis (MS)

Much clinical research on IFN-β is currently focused on its use as a treatment for multiple sclerosis (MS). Multiple sclerosis (abbreviated to MS, known as *disseminated sclerosis* or *encephalomyelitis disseminata*) is an inflammatory disease in which the fatty myelin sheaths around the axons of the brain and spinal cord are damaged, leading to demyelination and scarring as well as a broad spectrum of signs and symptoms (Compston and Coles, 2008). The patient with MS can suffer almost any neurological symptom or sign. Sexual dysfunction is a significant symptom of multiple sclerosis, affecting 50-90% of men and 40-80% of women (Kessler *et al.*, 2009), bladder and bowel difficulties (Andrews and Husmann, 1997; Henze, 2005; McCombe *et al.*, 2009). Changes in sensation (hypoesthesia and paraesthesia), muscle weakness, muscle spasms, or difficulty in moving; difficulties with coordination and balance (ataxia) (Freeman, 2001); problems in speech (dysarthria) or swallowing (dysphagia) (Merson and Rolnick, 1998), visual problems (nystagmus, optic neuritis, or diplopia) (Kaur and Bennet, 2007), fatigue, acute or chronic pain (Bobholz and Rao, 2003; Pollmann and Feneberg, 2008) Cognitive impairment of varying degrees and emotional symptoms of depression or unstable mood are also common (Kurtzke, 1983; De Seze *et al.*, 2006).

MS affects the ability of nerve cells in the brain and spinal cord to communicate with each other effectively. Nerve cells communicate by sending electrical signals called action potentials down long fibers called axons, which are contained within an insulating substance called myelin. In MS, the body's own immune system attacks and damages the myelin. When myelin is lost, the axons can no longer effectively conduct signal (Compston, 2002). The name *multiple sclerosis* refers to scars (scleroses—better known as plaques or lesions) particularly in the white matter of the brain and spinal cord, which is mainly composed of myelin (Clanet, 2008; Figure 4).

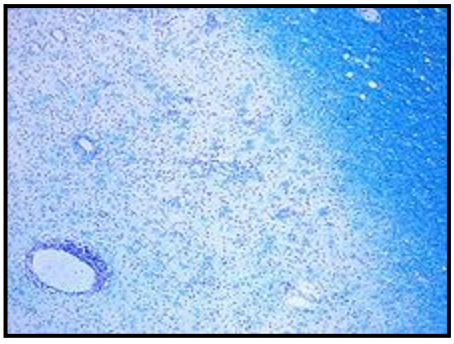


**Figure 4:** Demyelination in MS. On Klüver-Barrera myelin staining, decoloration in the area of the lesion can be appreciated (Original scale 1:100; Clanet, 2008).

#### 1.7.1.1 Classification

Almost any neurological symptom can appear with the disease, and often progresses to physical and cognitive disability (Compston, 2002). MS takes several forms, with new

symptoms occurring either in discrete attacks (relapsing forms) or slowly accumulating over time (progressive forms). Between attacks, symptoms may go away completely, but permanent neurological problems often occur, especially as the disease advances (Lublin and Reingold, 1996). Several subtypes, or patterns of progression, have been described. Subtypes use the past course of the disease in an attempt to predict the future course. They are important not only for prognosis but also for therapeutic decisions. In 1996 the United States National Multiple Sclerosis Society standardized four subtype definitions (Lublin and Reingold, 1996):

1. The relapsing-remitting subtype (Figure 5-1) is characterized by unpredictable relapses followed by periods of months to years of relative quiet (remission) with no new signs of disease activity. Deficits suffered during attacks may either resolve or leave sequelae, the latter being more common as a function of time. This describes the initial course of 80% of individuals with MS (Compston and Coles, 2008). When deficits always resolve between attacks, this is sometimes referred to as benign MS (Pittock and Rodriguez. 2008), although patients will still accrue some degree of disability in the long term (Compston and Coles, 2008). The relapsing-remitting subtype usually begins with a clinically isolated syndrome (CIS). In CIS, a patient has an attack suggestive of demyelination, but does not fulfill the criteria for multiple sclerosis (Miller *et al*., 2005; Compston and Coles, 2008). However only 30 to 70% of persons experiencing CIS later develop MS (Miller *et al*., 2005).

2. The primary progressive subtype (Figure 5-2) describes the approximately 10–15% of individuals who never have remission after their initial MS symptoms (Miller, 2005). It is characterized by progression of disability from onset, with no, or only occasional and minor, remissions and improvements. (Lublin and Reingold, 1996). The age of onset for the primary progressive subtype is later than for the relapsing-remitting, but similar to mean age of progression between the relapsing-remitting and the secondary progressive. In both cases it is around 40 years of age (Miller, 2005).

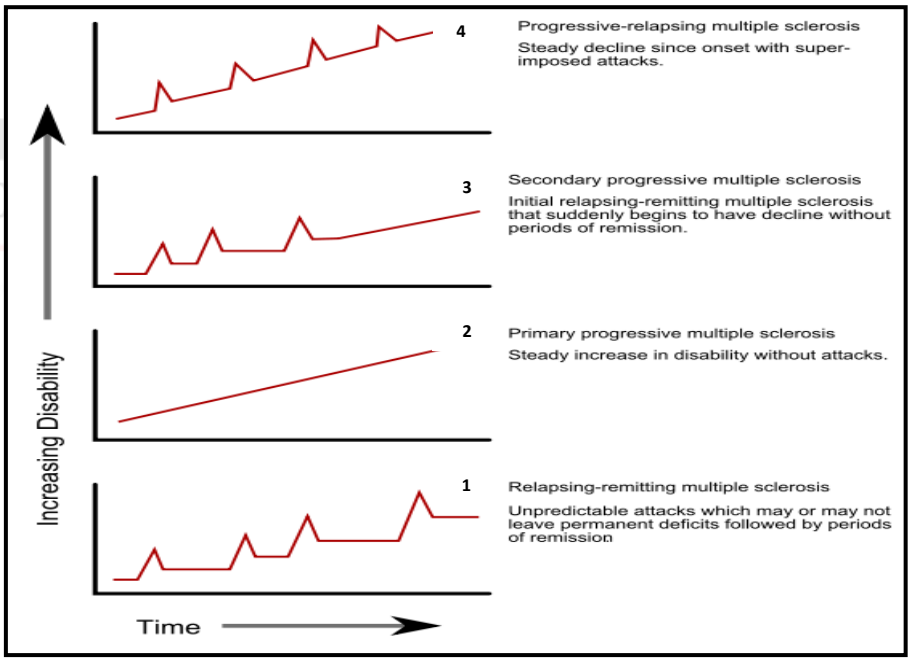



**Figure 5:** Progression types of multiple sclerosis (http://www.ivline.info/2012/03/med-in-small-doses-multiple-sclerosis.html)

3. Secondary progressive MS (sometimes called "galloping MS") (Figure 5-3) describes around 65% of those with an initial relapsing-remitting MS, who then begin to have progressive neurologic decline between acute attacks without any definite periods of remission (Lublin and Reingold, 1996; Compston and Coles, 2008). Occasional relapses and minor remissions may appear (Lublin and Reingold, 1996). The median time between disease onset and conversion from relapsing-remitting to secondary progressive MS is 19 years (Rovaris *et al*, 2006).
4. Progressive relapsing MS (Figure 5-4) describes those individuals who, from onset, have a steady neurologic decline but also suffer clear superimposed attacks. This is the least common of all subtypes (Lublin and Reingold, 1996).

Atypical variants of MS with non-standard behavior have been described; these include Devic's disease, Balo concentric sclerosis, Schilder's diffuse sclerosis and Marburg

multiple sclerosis. There is debate on whether they are MS variants or different diseases (Stadelmann and Brück, 2004). Multiple sclerosis also behaves differently in children, taking more time to reach the progressive stage. Nevertheless they still reach it at a lower mean age than adults (Compston and Coles, 2008).

#### 1.7.1.2 Pathophysiology

**Blood-brain barrier (BBB) breakdown**

BBB is a capillary system that should prevent entrance of T cells into the nervous system (Compston, 2002). The blood–brain barrier is normally not permeable to these types of cells, unless triggered by infection or a virus, which decreases the integrity of the tight junctions forming the barrier. When the blood–brain barrier regains its integrity, usually after infection or virus has cleared, the T cells are trapped inside the brain (Compston, 2002).

- **Autoimmunology**

MS is currently believed to be an immune-mediated disorder mediated by a complex interaction of the individual's genetics and as yet unidentified environmental insults. Damage is believed to be caused by the patient's own immune system. The immune system attacks the nervous system, possibly as a result of exposure to a molecule with a similar structure to one of its own (Compston, 2002).

- **Lesions**

MS lesions most commonly involve white matter areas close to the ventricles of the cerebellum, brain stem, basal ganglia and spinal cord; and the optic nerve. The function of white matter cells is to carry signals between grey matter areas, where the processing is done, and the rest of the body. The peripheral nervous system is rarely involved. More specifically, MS destroys oligodendrocytes, the cells responsible for creating and maintaining a fatty layer—known as the myelin sheath—which helps the neurons carry electrical signals.

MS results in a thinning or complete loss of myelin and, as the disease advances, the cutting (transection) of the neuron's extensions or axons. When the myelin is lost, a neuron can no longer effectively conduct electrical signals. (Compston, 2002). A repair process, called remyelination, takes place in early phases of the disease, but the oligodendrocytes cannot completely rebuild the cell's myelin sheath. Repeated attacks lead to successively fewer effective remyelinations, until a scar-like plaque is built up

around the damaged axons (Chari, 2007). Different lesion patterns have been described. (Pittock and Lucchinetti, 2007).

- **Inflammation**

Apart from demyelination, the other pathologic hallmark of the disease is inflammation. According to a strictly immunological explanation of MS, the inflammatory process is caused by T cells, a kind of lymphocyte. Lymphocytes are cells that play an important role in the body's defenses. (Compston, 2002). In MS, T cells gain entry into the brain via the previously described blood–brain barrier. Evidence from animal models also point to a role of B cells in addition to T cells in development of the disease (Iglesias *et al*., 2001). The T cells recognize myelin as foreign and attack it as if it were an invading virus. This triggers inflammatory processes, stimulating other immune cells and soluble factors like cytokines and antibodies. Leaks form in the blood–brain barrier, which in turn cause a number of other damaging effects such as swelling, activation of macrophages, and more activation of cytokines and other destructive proteins. (Compston, 2002). MS is a chronic inflammatory disease characterized by lymphocyte infiltration and demyelination of the central nervous system.

Autoreactive T cells recognizing myelin components such as myelin basic protein (MBP) are thought to contribute to the pathogenesis of MS (Sospedra and Martin, 2005). Although MS is a central nervous system (CNS) disease, myelin reactive T cells exist in an activated state in the peripheral blood and cerebrospinal fluid (CSF) of patients with MS (Zhang *et al*., 1994; Rodney and Gibaldi, 2003).

These activated T lymphocytes and macrophages produce cytokines that modulate the immune response, either positively or negatively. The balance between beneficial and deleterious effects of these cytokines depends on the context of the challenge facing the immune system (Sandra *et al*., 2009).

#### 1.7.1.3 Prevalence

MS is the most common debilitating demyelinating disease that affects young adults (Hafler *et al*, 2005). MS world prevalence ranges between 2 and 150 per 100,000 (Figure 6).

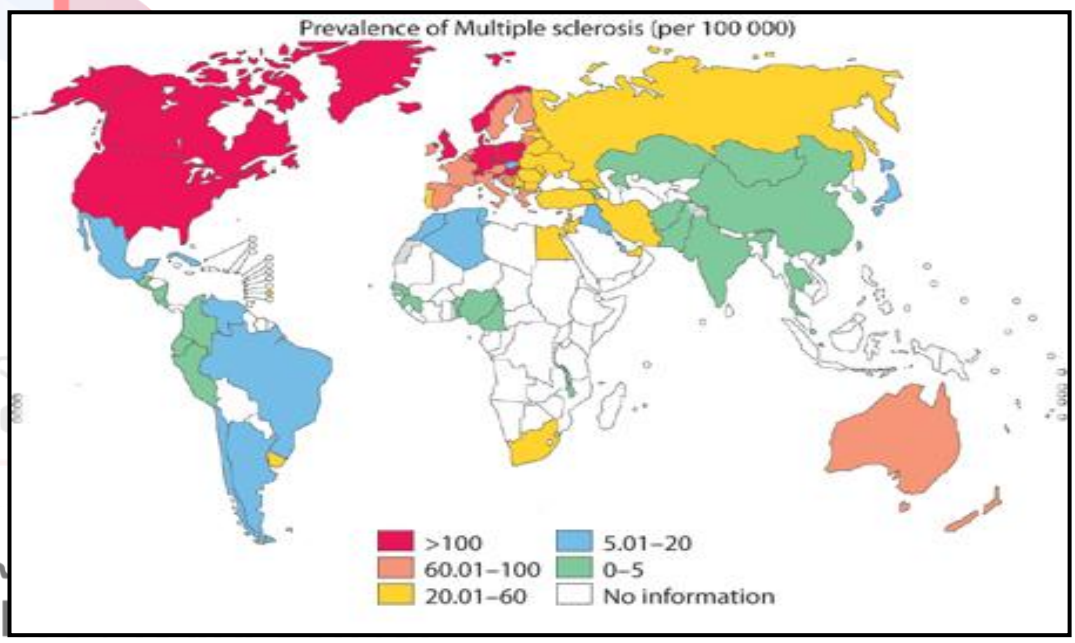



**Figure 6:** Global Prevalence of multiple sclerosis in 2011. (http://www.drbriffa.com/2011/10/20/why-might-shift-workers-be-at-increased-risk-of-multiple-sclerosis/)

There are more than 2. 5 million people affected worldwide and MS is more predominant in females than in male (3:1) (Debouverie *et al*., 2008).

In the Middle East it is 3 per 100,000 (Koch-Hnriksen and Hyllested, 1988; Rosati, 2001; Kobelt *et al.*, 2006).

However, some recent studies in Arabian countries have reported that the MS prevalence may be on the rise, and MS frequency in some parts of the regions may be in the medium range (Hashemilar *et al*., 2011).

Bahrain population (nationals) according to the Kingdom of Bahrain eGoverment portal is equal to 527,433 inhabitants and MS patients according to Bahraini MS Association are 250 patients. So the prevalence of MS in Bahrain is 47.4/100,000 inhabitants, that put Bahrain in the intermediate range [20 to 60 cases per 100,000 inhabitants] with

respect to the world MS prevalence classification which varies from 2 to 150 cases per 100,000 people (Kingdom of Bahrain eGoverment portal, 2012).

#### 1.7.1.4 Management of MS

There is no known cure for multiple sclerosis. Treatments attempt to return function after an attack, prevent new attacks, and prevent disability (Compston, 2002).

MS medications can have adverse effects or be poorly tolerated, and many patients pursue alternative treatments, despite the lack of supporting scientific study.



The prognosis is difficult to predict; it depends on the subtype of the disease, the individual patient's disease characteristics, the initial symptoms and the degree of disability the person experiences as time advances. (Weinshenker, 1994).

Life expectancy of people with MS is 5 to 10 years lower than that of the unaffected population (Compston and Coles, 2008). During symptomatic attacks, administration of high doses of intravenous corticosteroids, such as methylprednisolone, is the routine therapy for acute relapses (Compston and Coles, 2008).

Although generally effective in the short term for relieving symptoms, corticosteroid treatments do not appear to have a significant impact on long-term recovery. Oral and intravenous administrations seem to have similar efficacy (Burton *et al.*, 2009).

Consequences of severe attacks which do not respond to corticosteroids might be treated by plasmapheresis. (Compston and Coles, 2008).

##### 1.7.1.4.1 Disease-modifying treatments

Disease Modifying Drugs – DMDs – are drugs which affect the long-term progression of MS. They target inflammation and are designed to reduce the damage caused by relapses. By doing this, the number and severity of relapses decreases.

Some of these drugs have been found to delay the long term progression of MS and reduce the number of new lesions forming. There are currently six DMDs available as shown in table 2.

**Table 2:** Disease modifying drugs (DMD) used for treatment of MS (http://www.msrc.co.uk/index.cfm/fuseaction/show/pageid/1695**)**

| Name | Trade Name | Type | Delivery Method |
|---|---|---|---|
| **Fingolimod** | Gilneya | Immunomodulator | Orally |
| **Interferon beta-1a** | • Avonex<br>• CinnoVex<br>• ReciGen<br>• Rebif | Inteferon modulator | Avonex: intramascular injections. Once per week<br>Rebif: Subcutaneous injections. Three times a week |
| **Interferon beta-1b** | • Betaseron (Europe)<br>• Betaferon (Japan) | Interferon Modulator | Subcutaneous injections Everyday |
| **Glatiramer Acetate** | Copaxone | non-interferon, non-steroidal immunomodulator | Intravenous injection. Once per day |
| **Mitoxantrone** | Novantrone | immunosuppressant | Intravenous injections. Once every 3 month |
| **Natalizumab** | Tysabri | humanized monoclonal antibody immunomodulator | Intravenous injections. Once every month |

- **Advantages**

All six kinds of medications are modestly effective at decreasing the number of attacks in relapsing-remitting MS (RRMS) while the capacity of interferons and glatiramer acetate is more controversial. Studies of their long-term effects are still lacking (Compston and Coles, 2008; Comi, 2009). It is believed that IFN-β based drugs achieve their beneficial effect on MS progression via their anti-inflammatory properties. Studies have also determined that interferon beta improves the integrity of the blood-brain barrier (BBB), which generally breaks down in MS patients, allowing increasing amounts of undesirable substances to reach the brain. This strengthening of the BBB may be a contributing factor to Interferon-beta's beneficial effects. These studies were carried out *in vitro* and thus may not necessarily work the sae way in people. Comparisons between immunomodulators (excluding mitoxantrone) show that the most effective is natalizumab, both in terms of relapse rate reduction and halting disability progression (Johnson, 2007). Mitoxantrone may be the most effective drug; however, it

is generally not considered as a long-term therapy, as its use is limited by severe secondary effects (Compston and Coles, 2008; Comi, 2009). The earliest clinical presentation of RRMS is the clinically isolated syndrome (CIS). Treatment with interferons during an initial attack can decrease the chance that a patient will develop clinical MS (Compston and Coles, 2008).

Treatment of progressive MS is more difficult than relapsing-remitting MS. Mitoxantrone has shown positive effects in patients with secondary progressive and progressive relapsing courses.

It is moderately effective in reducing the progression of the disease and the frequency of relapses in patients in short-term follow-up (Martinelli *et al*., 2005). No treatment has been proven to modify the course of primary progressive MS. (Leary and Thompson, 2005).

The interferons have a unique place in the history of drug development in that respect that studies in man preceded animal studies (Lublin, 2005). IFN-β was shown to be well-tolerated when compared with other drugs.

The pivotal IFNβ-1b trial was published in 1993 and heralded the start of the therapeutic era in MS and the introduction of IFNβ-1b into the USA – the first therapy proven effective in altering the natural history of relapsing-remitting MS (RRMS). Treatment with recombinant IFN-β shows remarkable beneficial effects in the relapsing-remitting form of MS, leading to a reduction in both the severity and frequency of attacks (Filippini *et al*., 2003; De Jager and Hafler, 2007; Buttmann and Rieckmann, 2007; Sandra *et al*., 2009).

The development of new agents is a long, drawn out, often unsuccessful process, as the number of recent failures illustrates it. However, the long-term safety and efficacy of IFN-β treatment is unquestionable, with over years of clinical experience as evidence IFN- β not only shows high therapeutic activity but also long term safety (McCormack and Scott, 2004; Lublin, 2005).

However, as with many medical treatments, these treatments have several adverse effects.

- **Drawbacks**

Although the DMDs are available to improve the quality life of MS patient, some of the treatments have several adverse effects which can range from flu symptoms to death as shown in Table 3.

**Table 3:** Drawbacks of Disease modifying drugs for MS.

| Treatment | Drawback | References |
|---|---|---|
| **Fingolimod** | Potential fatal infections, bradycardia, skin cancer and, recently, a case of hemorrhaging focal encephalitis, an inflammation of the brain with bleeding. Other common side effects: colds, headache, and fatigue. | • Ascherio *et al.*, 2010 |
| **Natalizumab** | Few cases of skin cancer<br>Some cases of progressive multifocal leukoencephalopathy | • Ghadirian *et al.*, 1998<br>• Compston and Coles, 2008 |
| **Glatiramer acetate** | Most common: Irritation at injection site, lipoatrophy, may develop. Some patients experience a post-injection reaction manifested by flushing, chest tightness, heart palpitations, breathlessness, and anxiety | • Sládková and Kostolanský, 2006<br>• Munari *et al.*, 2004 |
| **Interferons** | Most common: Irritation at injection site, lipoatrophy, may develop. Symptoms similar to influenza. Few cases are liver damage, severe cardiotoxicity, infertility. | • Sládková and Kostolanský, 2006<br>• Tremlett and Oger, 2004 |
| **Mitoxantrone** | Acute myeloid leukemia | • Compston and Coles, 2008<br>• Comi, 2009 |

### 1.8 IFN-β production

In 1993, the FDA approved subcutaneous injections of IFN-β1b for treatment of MS. IFN-β1b is a non-glycosylated form of IFN-β produced by *Escherichia coli (E. coli*) (Arduini *et al.*, 1999); it is marketed as Betaseron. Currently, IFN-β1a (a eukaryotic, glycosylated form obtained from hamsters) is also available under the trade name Avonex. Table 4, shows some of the commercial IFN-β drugs mode of production.

**Table 4:** Some of commercial IFN-β drugs production (Arduini *et al.*, 1999).

| Name | Type | System |
|---|---|---|
| Rebif | IFNβ-1a | Mammalian cells |

| | | |
|---|---|---|
| Avonex | IFNβ-1a | Mammalian cells |
| Betaseron | IFNβ-1b | Modified *E.coli* |

Interferon as a biopharmaceutical drug is manufactured by Chinese hamster ovaries cell (CHO) (Arnold *et al.*, 2009), but it has low yield and very high cost causing increase in the final price (Martin, 1985).

An enhanced IFN-β production was carried out using Chinese hamster ovaries (CHO) cells through elevated osmolality and reduced culture temperature (Han *et al.*, 2009). *Escherichia. coli* (*E. coli)* and the yeast *Pichia pastoris (P. Pastoris)* expression systems are currently the most widely used systems for the large-scale production of various recombinant proteins of medical interest. *P. pastoris* is a species of methylotrophic yeast. *Pichia* is widely used for production using recombinant DNA techniques. Hence it is used in biochemical and genetic research in academia and the biotechnological industry. A number of properties makes *P. pastoris* suited for this task:

- high growth rate
- grow on a simple, inexpensive medium
- grow in either shake flasks or a fermenter, which makes it suitable for both small and large scale production.

*P. pastoris* has two alcohol oxidase genes, *Aox1* and *Aox2*, which have a strongly inducible promoter (Daly and Hearn, 2005). These genes allow *P. pastoris* to use methanol as a carbon and energy source. The *Aox* promoters are induced by methanol and are repressed by e.g. glucose. Usually the gene for the desired protein is introduced under the control of the *Aox1* promoter, which means that protein production can be induced by the addition of methanol. In a popular expression vector, the desired protein is produced as a fusion product to the secretion signal of the α-mating factor from *Saccharomyces cerevisiae* (*S. cerevisiae*) This leads the protein to be secreted into the growth medium, which greatly facilitates subsequent protein purification. There are commercially available plasmids that have these features incorporated (such as the pPICZα vector, Invitrogen).

In standard molecular biology research, the bacterium *E. coli* is the most frequently used organism for the production of recombinant proteins. This is due to *E. coli's* fast growth rate, good protein production rate and undemanding growth conditions. Protein

production in *E. coli* is usually faster than in *P. pastoris* for several reasons: Competent *E. coli* cells can be stored frozen, and thawed immediately before use, whereas *P. pastoris* cells have to be produced immediately before use or else it can be stored for shorter period of time. Expression yields in *P. pastoris* vary between different clones, and usually a large number of clones need to be screened for protein production. Optimal induction times of *Pichia* are usually on the order of days, whereas *E. Coli* usually reaches optimal yields within hours of induction. The major advantage of *P. pastoris* over *E. coli* is that *P. pastoris* is capable of producing disulfide bonds and glycosylations in proteins (Cregg *et al*., 2009). This means that in cases where disulfides are necessary, *E. coli* might produce a misfolded protein that is usually inactive or insoluble (Brondyk, 2009)

The well-studied *S. cerevisiae* is also used as an expression system with similar advantages over *E. coli* as *P. pastoris*. However *P. pastoris* has two main advantages over *S. cerevisiae* in laboratory and industrial settings:

- *P. pastoris*, as mentioned above, is a methylotroph, meaning it can grow with the simple alcohol methanol as its only source of energy — *Pichia* can easily be grown in cell suspension in reasonably strong methanol solutions that would kill most other micro-organisms, a cheap system to set up and maintain.
- *P. pastoris* can grow to very high cell densities, and under ideal conditions can multiply to the point where the cell suspension is practically a paste. As the protein yield from expression in a microbe is roughly equal to the product of the protein produced per cell and the number of cells, this makes *Pichia* of great use when trying to produce large quantities of protein without expensive equipment. (Cregg *et al.*, 2009).

Compared to other expression systems such as S2-cells from *Drosophila melanogaster* or CHO cells, *P. pastoris* usually gives much better yields. Cell lines from multicellular organisms usually require complex rich media, including amino acids, vitamins and growth factors. These media significantly increase the cost of producing recombinant proteins. Additionally, since *P. pastoris* can grow in media containing only one carbon source and one nitrogen source, it is suitable for isotopic labelling applications in e.g. nuclear magnetic resonance spectroscopy (Cregg *et al*., 2009). However, a number of proteins require chaperones for proper folding. Thus, *Pichia* is unable to produce a number of proteins for which the host lacks the appropriate chaperones. In 2006 a research group has managed to create a strain that produces Erythropoietin in its normal

glycosylation form (Hamilton *et al*., 2006). This was achieved by exchanging the enzymes responsible for the fungal type glycosylation, with the mammalian homologs. Thus, the altered glycosylation pattern allowed the protein to be fully functional.

### 1.9 Strategies for optimization of IFN-β production

It is reasonably expected that the effective production of INF-β at high yield will lower the cost of this drug leading to highly economical and therapeutical benefits. Codon optimization is a technique recently used by many scientists to improve the gene expression in living organisms by increasing the translational efficiency of gene of interest (Baev *et al.*, 2001). Optimizing codon for the custom gene design is the best way to increase the functionality of gene. Optimization strategies have been utilized to improve the production of heterologous proteins. The choice of expression vectors and transcriptional promoters are important. The nucleotide sequences encoding the N-terminal region of the protein appear particularly sensitive, both to the presence of rare codons and to the identities of the codons immediately adjacent to the initiation AUG (Baev *et al.*, 2001).

There is also some interplay between translation and mRNA stability which has not been completely de-convoluted, although reduced translational efficiency may be accompanied by a lower mRNA level because decreased ribosomal protection of the mRNA will increase its exposure to endo-RNAses. The structure of the 5' end of the mRNA also has a significant effect, and strategies using short upstream open reading frames for translational coupling of target genes have proved successful in improving the efficiency of expression of some genes. mRNA secondary structure stability as well plays a prominent role in DNA vaccines efficiency. The stability of any secondary structure is quantified as the amount of free energy released or used by forming base pairs. And it's established that the less the free energy the more stable is a secondary structure (Baev *et al.*, 2001).

It should also be noted that efficient translation is necessary but not sufficient to produce a functional protein. The polypeptide chain must fold correctly, in some cases

form appropriate disulphide bonds and even undergo post-translational modifications such as glycosylation. For these processes the absence of the correct redox environment, chaperonins, normal association partners or modifying enzymes will provide additional challenges (Baev *et al.*, 2001).

### 1.9.1 Concept of codon usage bias

Codon usage bias refers to differences in the frequency of occurrence of synonymous codons in coding DNA.

Different organisms often show particular preferences for one of the several codons that encode the same amino acid- that is a greater frequency of one will be found than expected by chance. How such preferences arise is a much debated area of molecular evolution.

However, these synonymous codons (different codons that code for same amino acid) are not used uniformly; in almost every organism that has been studied, there are distinct preferences for particular codons (Andersson *et al.*, 1990). For example, although there are six codons that code for leucine in the universal genetic code, almost half of all leucine codons in the bacterium *E. coli* are CUG (Table 5). This pattern is referred to as "synonymous codon bias" or "synonymous codon usage bias."

**Table 5:** Proportions of leucine and phenyalanine codons encoded by their synonymous codons in four species (Ikemura *et al.*, 1985).

| Amino acid | Codon | *Escherichia coli* | *Saccaromyces cerevisiae* | *Drosophila melanogaster* | *Homo Sapiens* |
|---|---|---|---|---|---|
| | | Bacterium | Yeast | Fruit fly | Human |
| Leucine | TTA | 0.13 | **0.28** | 0.05 | 0.07 |
| | TTG | 0.13 | **0.29** | 0.17 | 0.12 |
| | CTT | 0.11 | 0.13 | 0.10 | 0.13 |
| | CTC | 0.10 | 0.06 | 0.16 | 0.20 |
| | CTA | 0.04 | 0.14 | 0.08 | 0.07 |
| | CTG | **0.49** | 0.10 | **0.44** | **0.41** |
| Phenylalanine | TTT | 0.57 | 0.59 | 0.34 | 0.44 |
| | TTC | 0.43 | 0.41 | 0.66 | 0.56 |

It is generally acknowledged that codon preferences reflect a balance between mutational biases and natural selection for translational optimization. Optimal codons in fast-growing microorganisms, like *E. coli* or *S. cerevisiae*, reflect the composition of their respective genomic tRNA pool. It is thought that optimal codons help to achieve faster translation rates and high accuracy. As a result of these factors, translational selection is expected to be stronger in highly expressed genes, as is indeed the case for the above-mentioned organisms. In other organisms that do not show high growth rates or that present small genomes, codon usage optimization is normally absent, and codon preferences are determined by the characteristic mutational biases seen in that particular genome.

Examples of this are *Homo sapiens* (human) and *Helicobacter pylori*. Organisms that show an intermediate level of codon usage optimization include *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode worm) or *Arabidopsis thaliana* (thale cress).

### 1.9.2 Codon usage bias and gene expression

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but in non-protein coding genes such as ribosomal RNA (rRNA), transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA. The process of gene expression is used by all known life - eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea), possibly induced by viruses - to generate the macromolecular machinery for life. Several steps in the gene expression process may be modulated, including the transcription, RNA splicing, translation, and post-translational modification of a protein. Gene regulation gives the cell control over structure and function, and is the basis for cellular differentiation, morphogenesis and the versatility and adaptability of any organism. Gene regulation may also serve as a substrate for evolutionary change, since control of the timing, location, and amount of gene expression can have a profound effect on the functions (actions) of the gene in a cell or in a multicellular organism (Plotkin and Kudla, 2011).

In genetics, gene expression is the most fundamental level at which the genotype gives rise to the phenotype. The genetic code stored in DNA is "interpreted" by gene expression, and the properties of the expression give rise to the organism's phenotype. Gene expression is achieved by a combination of several mechanisms, the most prominent of which are transcription and translation. Transcription is the process of creating a complementary RNA copy of a sequence of DNA, where translation is the

stage of protein biosynthesis from the mRNA. In order to study the correlation of codon usage bias and gene expression, it is required study to homologous and heterologous expression (Plotkin and Kudla, 2011).

#### 1.9.2.1 Homologous and heterologous expressions

The definition of homologous expression and heterologous expression is mainly determined by the evolutionary distance between the origin of the genes to be produced and the expression host used (Kerrigan *et al.*, 2011). Expression of mammalian genes in

a mammalian cell line is considered to be homologous expression, whereas expression of the mammalian genes in yeast, insect cells, and bacteria hosts are considered to be heterologous expression. In certain cases, there may be a need to produce human proteins in a human cell line to make it strictly homologous. As with protein production of individual proteins, the choice of the expression host for producing protein complexes is influenced by multiple factors, including post-translational modifications, solubility, and stability of each component. Therapeutic vaccine proteins or immunogens such as viral-like particles (VLPs) are routinely expressed in non-mammalian hosts to avoid contamination with mammalian-derived components (Noad and Roy, 2003). The HPV vaccines Cervarix and Gardasil are VLPs of papillomavirus major capsid proteins expressed and assembled in insect T. ni cells via baculovirus and in the yeast *S. cerevisiae*, respectively (Szarewski, 2010).

On the other hand, homologous expression allows expression and assembly of close to native complexes, including patterns of post-translational modifications. It also allows incorporation of potentially unknown factors that may facilitate complex formation and function. However, given the homologous host background, paralogues and orthologues (for example, a human gene is expressed in CHO cells) of the component proteins are likely to be incorporated into the protein complex, thus introducing heterogeneity of the purified protein complex (Szarewski, 2010).

**Heterologous gene expression and the effect of codon usage bias**

Codon bias plays a critical role in heterologous gene expression. However, there is often a disconnect between technological and evolutionary studies of codon bias – a gap that partly reflects genuine differences between endogenous and heterologous situations. In many biotechnological applications, a transgene is massively over-expressed, accounting for up to 30% of the protein mass in cell. As a result, the principles relating

heterologous codon usage to protein levels may differ substantially from the endogenous case (Plotkin and Kudla, 2011). By understanding the concept of codon bias among living organisms and how it affects gene expression, it was mandatory to consider it in the design of heterologous gene expression experiments. In 1977 when Genentech scientists and their academic collaborators produced the first human protein (somatostatin) in a bacterium, only the amino acid sequence of somatostatin was known, so the Genentech group synthesized the 14 codon long somatostatin gene using oligonucleotides instead of cloning it from the human genome. The group designed these oligonucleotides based on three criteria:

**(a)** Codons were adapted to the bacteriophage MS2 codon usage. Not much of the *E. coli* genome DNA sequence was known at the time, but the MS2 phage had just been sequenced and was assumed to provide a good guide to the codons used in highly expressed *E. coli* genes.

**(b)** Eliminating undesirable inter- and intra-molecular pairing of the overlapping oligonucleotides as this would compromise the gene synthesis process.

**(c)** Sequences rich in GC followed by AT rich sequence were avoided, for it's believed it could terminate transcription.

The result was the first production of a functional polypeptide from a synthetic gene (Itakura *et al.*, 1977). Later on, cDNA libraries and PCR replaced *De novo* gene synthesis because of its high costs and effort (Wu *et al.*, 2004). However, after the amplification of gene using either cDNA libraries or PCR and cloning it into an expression vector, often the gene is not expressed or poorly expressed. Much work has been done to improve the expression of cloned genes, including optimization of host growth conditions and the development of new host strains, organisms and cell free systems (Higgins *et al.*, 1999). Despite the advances made to these approaches, the later have skirted a significant underlying problem: the DNA sequence used to encode a protein in one organism is often quite different from the sequence that would be used to encode the same protein in another organism. Those practical approaches proved that the expression of a certain gene differs from its host to various expression vectors which is a hurdle to the production of recombinant protein in heterologous hosts and also to DNA vaccines which is based mainly on producing the antigen protein in the vaccinated heterologous host cell (Gustafsson *et al.*, 2004).

**1.10 Gene optimization**

While the ability to make increasingly long stretches of DNA efficiently and at lower prices is a technological driver of this field, increasing attention is being focused on improving the design of genes for specific purposes. Early in the genome sequencing era, gene synthesis was used as an expensive source of cDNA's that were predicted by genomic or partial cDNA information but were difficult to clone. As higher quality sources of sequence verified cloned cDNA have become available, this practice has become less urgent.

Producing large amounts of protein from gene sequences (or at least the protein coding regions of genes, the open reading frame) found in nature can sometimes prove difficult and is a problem of sufficient impact that scientific conferences have been devoted to the topic (Liszewski, 2010).

### 1.10.1 Concept

Many of the most genes encoding interesting proteins sought by molecular biologist are normally regulated to be expressed in very low amounts in wild type cells. Redesigning these genes offers a means to improve gene expression in many cases. Rewriting the open reading frame is possible because of the degeneracy of the genetic code. Thus, it is possible to change up to about a third of the nucleotides in an open reading frame and still produce the same protein (Welch *et al.*, 2009).

Gene optimization is a logical strategy to improve gene expression in a certain heterologous host cell; to alter the rare codons in the target gene so that they more closely reflect the codon usage of the host without modifying the amino acid sequence of the encoded protein. Such a strategy is referred to as gene optimization, which proved to enhance the production yield, quality and even biological activity of recombinant proteins in heterologous systems, and has been widely implemented in the production of recombinant protein-based drugs and DNA vaccines design. The available number of alternate designs possible for a given protein is astronomical. For a typical protein sequence of 300 amino acids there are over 10150 codon combinations that will encode an identical protein. Using optimization methods such as replacing rarely used codons with more common codons sometimes have dramatic effects. Further optimizations such as removing RNA secondary structures can also be included. At least in the case of *E. coli,* gene expression is maximized by predominantly using codons corresponding to tRNA's that retain amino acid charging during starvation (Welch *et al.*, 2009). Computer programs are written to perform these, and other simultaneous optimizations are used to handle the enormous complexity of the task. A well optimized gene can improve protein expression 2 to 10 fold, and in some cases more than 100 fold

improvements have been reported. Because of the large numbers of nucleotide changes made to the original DNA sequence, the only practical way to create the newly designed genes is to use gene synthesis (Welch *et al.*, 2009).

### 1.10.2 Gene optimization effect on protein production

Experiments suggest that gene optimization affects gene expression at the transcriptional, posttranscriptional and translational level, thus significantly elevating protein levels (Fath *et al.,* 2011).

The majority of optimized genes induced a clear increase in detectable protein levels throughout all protein classes (Gustafsson *et al.*, 2004).

Gene optimization was achieved practically either by site-directed mutagenesis or the re-synthesis of the entire gene (Gustafsson *et al.*, 2004).

The expression in *E.coli* demonstrate dramatic increase in protein production when using the optimized gene versus the native gene. One example is the phosphatidylcholine transfer protein (PC-TP), where the protein production using the native gene sequence was undetectble (i.e. trace level), but an increase that accounts for 10% of cytosolic proteins was detected using the optimized gene sequence.

## 1.11 Gene optimization softwares'

In the field of bioinformatics and computational biology, many algorithms and statistical methods have been proposed and used to analyze codon usage bias (Comeron and Aguade, 1998). Methods such as the 'frequency of optimal codons' (Fop) (Ikemura, 1981), the Relative Codon Adaptation (RCA) (Fox and Erill, 2010) or the 'Codon Adaptation Index' (CAI) (Sharp and Li, 1987) are used to predict gene expression levels, while methods such as the 'effective number of codons' (Nc) and Shannon entropy from information theory are used to measure codon usage uniformity (Peden, 2005).

Multivariate statistical methods, such as correspondence analysis and principal component analysis, are widely used to analyze variations in codon usage among genes. (Suzuki *et al.*, 2008).

There are many computer programs to implement the statistical analyses enumerated above, including CodonW, GCUA, INCA, etc.

Codon optimization has applications in designing synthetic genes. Several software packages are available online for this purpose (Table 6).

**Table 6:** List of some software available for gene optimization

| Name of Software | Web Address | Server description |
|---|---|---|
| DyNAVacS | http://miracle.igib.res.in/dynavac/ | The program identifies a suitable expression vector and performs codon optimization |
| DNA 2.0 | http://www.dnatwopointo.com/commerce/misc/opt.jsp | Custom gene synthesis including codon optimization for increased protein expression |
| Up gene | http://www.vectorcore.pitt.edu/upgene/upgene.html | A web-based DNA codon optimization algorithm |
| GeneMaker | http://www.blueheronbio.com/genemaker/codon.html | Algorithm matches codon usage of sequence with that of the host |
| Optimizer | http://genomes.urv.es/OPTIMIZER | A codon usage optimization web server for gene expression in bacterial hosts |
| GASCO | http://miracle.igib.res.in/gasco/ | An algorithm for codon optimization |
| JCat | http://www.prodoric.de/JCat | A web-based software for adaptation of codon usage to most prokaryotic and some eukaryotic organisms |

It is noticed that the product genes of optimization using different softwares' varies for the same gene, and that is expected because the softwares' available are based on different algorithms.

Hence, based on different strategies for optimization, some simply replaces the rare codons with the more frequent ones in a chosen host.

Some go further to considering the stability and free energy of the produced messenger RNA (mRNA) from the optimized gene (Menzella, 2011).

Moreover, some of the softwares' consider the secondary structure of the protein, meaning it takes in account the translation process of the mRNA. Some softwares also offer special features when designing DNA vaccines, so they take in account modifying

CpG motifs and di-nucleotides in the vector's backbone; for they play a rule as adjuvant in DNA vaccination, and addition or customizing Kozak sequences.

Some also offer primer selection or designing and restriction mapping for the vector (Harish *et al*., 2006).

### 1.11.1 GASCO [Genetic Algorithm Simulation for Codon Optimization]

GASCO is an algorithm developed for codon optimization using genetic algorithms presented by Satya *et al.* (2003).

The triplet codons for the foreign DNA are selected based on their relative abundance in the host genome. More frequent codons are preferred over others.

Several short and long sequence motifs also affect the protein production from transcribed foreign gene either positively or negatively (by making secondary structures). The sequence motifs are also important to stimulate or suppress the immuno response.

Based on experimental requirements these motifs can be termed as *"desirable"* or *"undesirable".* In short the whole problem can be formulated as *"to select the optimum codons so as to maximize the desirable patterns and minimize the undesirable patterns".* As the above phrase itself indicates that the problem can be tackled as a typical optimization problem.

Satya *et al*. (2003) defined the problem of efficient pattern search to engineer CpG motifs in a sequence and proposed an approach to solve this problem which is implemented in the design of GASCO algorithm.

The algorithm takes a random initial population generated from the standard codon redundancy matrix and calculates the total fitness as a sum of sequence fitness (derived from the codon fitness) and pattern fitness.

Optimum solution is achieved around this fitness function by implementing double cross-over, mutation, elitism operations (by keeping the protein sequence same) and Roulette Wheel selection method (Sandhu *et al.*, 2008).

### 1.11.2 JCat [Java Codon Adaptaion Tool]

The Codon Adaptation Tool (JCAT) presents a simple method to adapt the codon usage to most sequenced prokaryotic organisms and selected eukaryotic organisms.
The codon adaptation plays a major role in cases where foreign genes are expressed in hosts and the codon usage of the host differs from that of the organism where the gene stems from (Sharp *et al.*, 1987).



Unadapted codons in the host can for example lead to a minor expression rate. The adaptation is based on CAI-values proposed by Sharp *et al.* (1987).
The eukaryotic genomes of mouse and human contain different kinds of biases along the chromosomes and the algorithm is not perfectly suited for this problem.

The mean codon usage for a certain organism was derived by summing over all CAI-values of all genes of this organism (except genes This data is also presented in the graphical output of the codon adaptation. As a further option for the codon adaptation the opportunity to avoid rho-independent trancription terminators is provided.

The algorithm for the prediction of these structures is based on a model from Ermolaeva *et al*. (2000). Another feature is the possibility to avoid restriction enzyme binding sites in the adapted DNA. The data for the restriction enzymes was therefore derived from the "The Restriction Enzyme Database" (REBASE). The web interface is easy to understand and the calculations are performed in real time.

Furthermore, the algorithm does not require a definition of highly expressed genes (Grote *et al.*, 2005). Calculations are made in advance with the aid of an algorithm proposed by Carbone *et al.* (2003).

## 1.12 Significance and objectives

### 1.12.1 Study rationale

The current trends in biotechnology for recombinant protein drugs call for gene optimization as an ultimate approach to develop recombinant proteins drugs with good quality and yield in a cost effective way. In addition, mastering the technology for the design and production of bio-drugs is of paramount strategic importance for building a biopharmaceutical industry. The first step toward this end consists in developing genetically engineered microorganisms producing the proteins of interest as recombinant proteins that can be formulated and used as bio-drugs. The rationale behind developing specifically a recombinant yeast clone that produces human IFN-β is as followed:

1. Wide prevalence of MS patients in the Arab population.
2. IFN-β is a treatment of choice available but at high cost.
3. The *Pichia pastoris* expression system can eventually allow the production of recombinant IFN-β locally in a cost effective way. This may improve the excess of MS patients in the Middle East and worldwide to this important drug.

### 1.12.2 Study objectives

#### 1.12.2.1 Overall objective

The overall aim of this work is to use a gene optimization strategy to generate a *Pichia pastoris* recombinant clone expressing human IFN-β gene.

#### 1.12.2.2 Specific objectives

1. Development of a *Pichia pastoris* clone that produces a recombinant human IFN-β from native cDNA.
2. Genetic engineering and development of a *Pichia pastoris* clone that expresses a recombinant human IFN-β from a synthetic cDNA whose sequence is optimized according to the yeast codon-usage bias.

# Chapter 2

## 2.1 Materials

### 2.1.1 Strains and plasmids

- The *E. coli* strain Top10F' (rec A- , end A-) from Invitrogen, (Groningen, the Netherlands) was used as host strain for plasmid preparation experiments. Rec A- represents reduced occurrence of unwanted recombination in cloned DNA; cells UV sensitive, deficient in DNA repair and end A- is for cleaner preparations of DNA and better results in downstream applications due to the elimination of non-specific digestion by endonuclease I.
- *P. pastoris* host strain was the Mut$^S$ KM71H (Genotype: aox1::ARG4). The wild- type ARG4 gene was used to disrupt AOX1, creating KM71H strain.

- Plasmid pPICZαA vector [Invitrogen] (3.6 kb) was used for the expression in *P. pastoris* (Figure 7). The vector can be used to express and secrete recombinant proteins in *P. pastoris*. Recombinant proteins are expressed as fusions to an N-terminal peptide encoding *S. cerevisiae* α-factor secretion signal. The vector allows high level, methanol inducible expression of the gene of interest in *P. pastoris*, and can be used in any *P. pastoris* strain including X-33, SMD1168H, 5' fragment containing the *AOX1* promoter for tightly regulated, methanol induced expression of the gene of interest
- α-factor secretion signal for directing secreted expression of the recombinant protein
- Zeocin resistance gene for selection in both *E. coli* and *Pichia*
- C-terminal peptide containing the *c-myc* epitope and polyhistidine (6xhis) tag for detection and purification of recombinant fusion protein (if desired). Three reading frames to facilitate in-frame cloning with the C-terminal peptide.

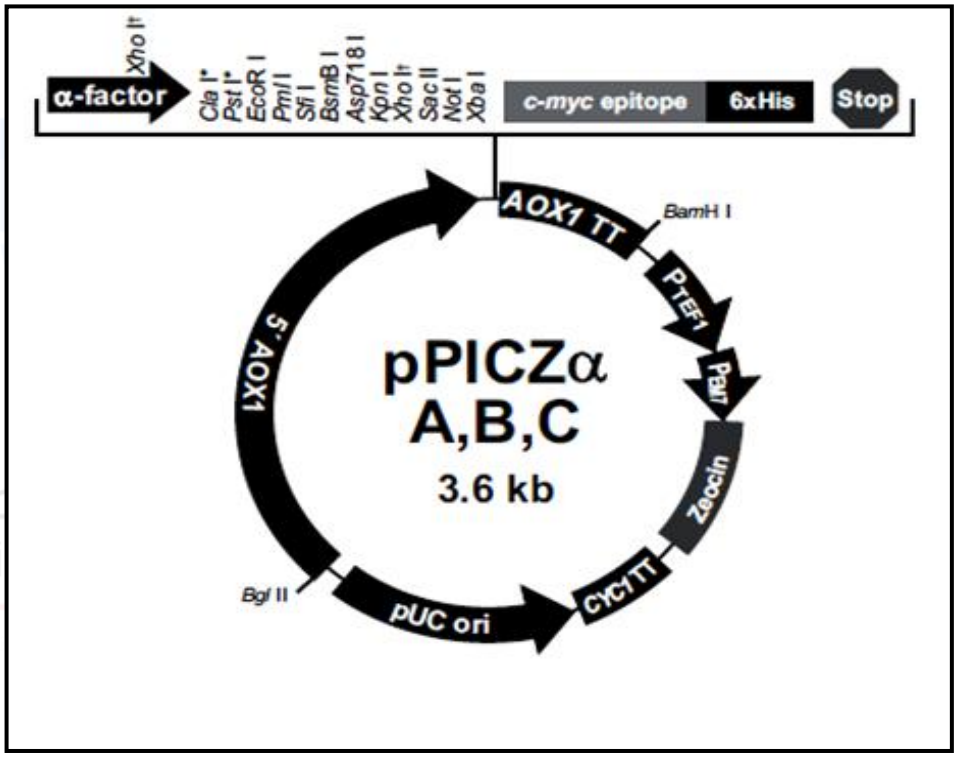



**Figure 7:** Plasmid PICZαA restriction map. It contains the following: **α-factor signal sequence** (941-1207 bases); **multiple cloning site** (1208-1276 bases); **c-myc epitope** (1275-1304); **polyhitidine tag** (1320-1337 bases). (Invitrogen).

### 2.1.2 Instruments

Table 7 represents the instruments used for the study.

**Table 7:** Instrumentation

| Instrument | Company | Model No. |
|---|---|---|
| Shakers | - GallenKamp orbital incubator, UK<br>- Formascientific orbital incubator, UK | -INR200 010V<br>-4535 |
| Nanodrop spectrophotometer | -Eppendorff Biophotometer plus, Germany | 6132-01209 |
| Incubator | - Thermolyne, Australia<br>- Heraeus, UK | - 42000<br>- B6 |
| DNA Engine Tetrad 2 Peltier Thermal Cycler | - Bio-Rad laboratories, Hercules, USA | PTC-0240G |
| Gene Pulser II electroporator | - Bio-Rad laboratories, Hercules, USA | 165-2660 |
| Centrifuge | - Hermle, Germany | Z400K |
| X-ray film processer | - Ecomax, FI-Sales, LLC; Germany | FI-45 |
| Semi-dry blot | - Brenzel BioAnalytik; Germany | UniBlot SC serie |

### 2.1.3 Softwares'

#### 2.1.3.1 Software used in comparing codon-usage procedure

Wild type (WT) cDNA sequences were first analyzed using the **G**raphical **T**ool for **C**odon **A**nalysis [**GTCA,** http://gcua.schoedl.de/] to determine the codons that significantly deviate from the yeast codon-usage.

#### 2.1.3.2 Softwares' for codon optimization

We have used the codon optimization software GASCO (**G**enetic **A**lgorithm **S**imulation for **C**odon **O**ptimization, http://miracle.igib.res.in/gasco/) and JCat (**J**ava **C**odon **A**daptation **T**ool, http://www.jcat.de/) to generate two different codon-optimized sequences compatible with *P. pastoris* codon usage table.



We have chosen the GASCO-optimized IFN-β and JCat-optimized IFN-β sequences for the comparison of protein production to test the correlation between codon usage bias with protein expression by using different codon optimization algorithm softwares'.

#### 2.1.3.3 Software used for DNA sequence alignment

The optimized sequences of IFN-β were aligned along with their WT sequence using EMBOSS (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html).

#### 2.1.3.4 Software used for comparison of protein sequences

The translated DNA sequences: WT, GASCO- and JCat-optimized IFN-β were analyzed and compared (http://web.expasy.org/translate/ and http://fasta.bioch.virginia.edu/).

#### 2.1.3.5 Software used for plasmid restriction pattern prediction

Pattern prediction of the pPICZαA plasmid was done using NEBcutter V2.0 from new England Biolabs (http://tools.neb.com/NEBcutter2/).

#### 2.1.3.6 Software used for primer designing

Primers were designed for the vector pPICZαA (AOX primer) and also for the IFN-β using Primer3 input (www.frodo.wi.mit.edu/) which are used for polymerase chain reaction (PCR) and sequencing. Table 8 illustrates the primer sequences used.

**Table 8:** Primer sequences used for different applications

| DNA Template | Application | Primer Sequence |
|---|---|---|
| **pPICZαA (Flanking region of IFN-β)** | - RT-PCR | **F:** 5' GAAAAGAGAGGCTGAAGCTGA3' **R:** 5' ACGGCGCTATTCAGATCCT 3' |
| **AOX1 region (pPICZαA)** | - Sequencing<br>- Yeast colony PCR | **F:** 5' GACTGGTTCCAATTGACAAG 3' **R:** 5' GCAAATGGCATTCTGACATCC 3' |

### 2.1.4 Restriction enzymes

Restriction enzymes used in this study are listed in table 9.

**Table 9:** Restriction enzymes

| Enzyme | Company | Catalog Number |
|---|---|---|
| *Xho* I | Promega | R6161 |
| *Xba* I | Promega | R6185 |
| *Bst*X I | Promega | R6475 |

### 2.1.5 Kits

- Plasmid purification kit:
  - Qiagen midiprep (Cat. # 12145, Hilden, Germany)
  - Qiagen miniprep (Cat. # 27104, Hilden, Germany)

- High capacity cDNA Reverse Transcriptase Kits: Applied Biosystems (Cat# 4374966, USA)
- GoTaq® Green Master Mix (Cat# M712, Promega, Madison, WI, USA).

## 2.2 Methods

### 2.2.1 Recombinant plasmid optimization and synthesis

IFN-β sequence was optimized to match the *P. pastoris* codon usage using two softwares': GASCO and JCat. Cloning was done by using restriction enzymes (RE) *EcoR* I / *Not* I. The GASCO and Jcat optimized sequences do not have any restriction sites recognized by the above mentioned RE. The restriction sites sequences were included in the optimized sequences construct in the 5' and 3' end for directing cloning orientation. The two constructed optimized sequences were synthesized and cloned in the pPICZαA plasmid (by GeneCust, Dudelange, Luxembourg). Quality control of the synthesized optimized sequences was done by sequencing (GeneCust).



### 2.2.2 Recombinant plasmids amplification

#### 2.2.2.1 Competent cells preparation

*E. coli* competent cells of strain Top10F' were prepared using the calcium chloride method (Dagert and Ehrlich, 1979). *E.coli* TOP10F' strain was streaked on Luria-Bertani (LB) agar plate and incubated at 37°C incubator overnight, and then a single colony from the *E.coli* Top10F' plate was cultured in 5 ml LB broth overnight at 37°C, shaking at 250 rpm. The starting culture 5 ml of *E.coli* Top10F' was then added to 100 ml of LB broth and cultured at 37°C, 250 rpm until it reached an optical density 0.6 at 600 nm. The bacterial cells were then pelleted by centrifugation at 3000 rpm for 10 min at 4°C and the supernatant is discarded. The pellet was then resuspended in 5ml of 100 mM calcium chloride ($CaCl_2$), followed by centrifugation at 3000 rpm for 10 min at 4°C and the supernatant was discarded. The pellet was resuspended with 100 mM $CaCl_2$ and kept overnight on ice. The cells were centrifuged at 3000 rpm for 10 min at 4°C and the supernatant discarded. The pellet was then resuspended with 4:1 $CaCl_2$: Glycerol and frozen at -70°C (aliquot of 200 µl).

#### 2.2.2.2 Competent cells transformation

*E.coli* TOP10F' strain competent cells were transformed with the recombinant plasmids by the heat shock method (Froger and Hall, 2007). 5, 50 and 100 ng of each recombinant plasmid were mixed with 200 µl of *E.coli* TOP10F' competent cells and kept on ice for 15 min, then heat shocked at 42°C for 50 seconds and placed back on ice immediately for 2 min to induce DNA uptake. The cells were then mixed with 900 µl of

LB broth and incubated at 37°C, 100 rpm for 1 hour (shaking incubator) in 5 ml culture tubes. 200 µl of the transformation mix were spread on LB agar plates with Zeocin (25µg/ml) as a selective antibiotic for transformed cells. The plates were incubated overnight at 37°C.

#### 2.2.2.3 Transformation efficiency

Transformation efficiency is the efficiency by which cells can take up extracellular DNA and express genes encoded by it. This is based on the competence of the cells. It can be calculated by dividing the number of successful transformants by the amount of DNA used during a transformation procedure (Hanahan, Jessee and Bloom, 1991). Transformation efficiency is measured in transformants or colony forming unit (cfu) per µg DNA used. It was done by calculating the number of colonies formed on the LB agar plates for The Human IFN-β WT, GASCO-opt and JCat-opt sequences in pPICZαA. In addition to the negative control which is the pPICZαA blank (no insert). Transformation efficiency (transformants/µg) was calculated as followed:

Number of colonies on plate/ng of DNA plated X 1000 ng/µg

#### 2.2.2.4 Amplification

A single colony from the transformation plates was picked for each of the cells containing the GASCO and JCat optimized plasmids, inoculated in 5 ml of LB media with Zeocin (25µg/ml) as well as cells containing IFN-β WT sequence and cells containing pPICZαA without the insert as control. Starter cultures were incubated overnight at 37°C, 250 rpm (shaking incubator). The next day, 1 ml of the overnight culture was inoculated in 100 ml of LB media with Zeocin (25µg/ml) and incubated overnight at 37°C, 250 rpm (Sambrook *et al.*, 1989).

### 2.2.3 Purification and quantification of plasmid DNA

Plasmids were purified using the Qiagen plasmid Midiprep kit (Figure 8) which is based on a modified alkaline lysis procedure. Alkaline lysis is a method used in molecular biology to isolate plasmid DNA from relatively small volumes of bacteria.

Then it is followed by binding of plasmid DNA to Anion-Exchange resin under ppropriate low-salt and pH conditions

RNA, proteins, dyes and low-molecular weight impurities were removed by a medium-salt wash solution.

Plasmid DNA was eluted in a high-salt buffer and then concentrated and desalted by isoprpanol precipitation. The purified plasmids were then resuspended in double distilled water ($ddH_2O$). Plasmid DNA stocks were quantified by measuring their $O.D_{260}$ using the Nanodrop spectrophotometer. The plasmids were stored at 4°C to retain their supercoiled form (www.qiagen.com).

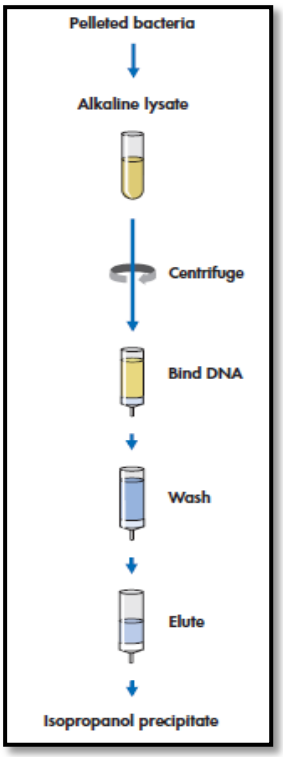



**Figure 8:** Qiagen midiprep flowchart (www.qiagen.com)

### 2.2.4 Recombinant plasmid quality control

The integrity of the purified plasmids was checked on 0.7% agarose gel stained with ethidium bromide and visualized under U.V light. After that plasmid sizing analysis was checked by digestion with *Xho* I and *Xba* I RE. To further investigate the quality of the purified plasmids, three plasmid DNA sequences were analyzed to make sure that they only differ in their IFN-β sequence and not the backbone of the plasmid itself, by using DNA baser tool.

### 2.2.5 Linearization of recombinant plasmids

Linearization of plasmids was done to promote integration prior to transformation into the yeast (host) genome within 5' AOX1 region (www.invitrogen.com). It is mandatory that the insert (IFN-β) does not contain the restriction site of the same enzyme that will be used to linearize the vector. *Bst*X I which has a restriction site on upstream of 5' AOX region in 707 bp of vector was used in this study to linearize pPICZαA (Table 10).

**Table 10**: Linearization reaction components using *Bst*X I

| **Component** | | **Volume (µl)** | | | |
|---|---|---|---|---|---|
| **Master Mix** | Final Concentration | pPICZαA Blank | pPICZαA WT | pPICZαA GASCO | pPICZα A JCat |
| **Nuclease-free water** | -- | 31.7 | 32.7 | 31.9 | 33 |
| **10X buffer D*** | 1X | 4 | 4 | 4 | 4 |
| **BSA (10µg/µl)** | 2.0 | 0.4 | 0.4 | 0.4 | 0.4 |
| ***Bst*XI(12units/µl)** | 20 | 1.7 | 1.7 | 1.7 | 1.7 |
| **DNA Template** | 2µg per reaction | 2.2 | 1.2 | 2 | 0.9 |
| **Volume** | -- | 40 | 40 | 40 | 40 |

* Buffer (D): pH 7.9, 6 mM Tris-HCL, 6 mM $MgCl_2$, NaCl 150 mM 1mM DTT.

10 µg DNA was done in 5 separate tubes; each 2 µg of template DNA. The reaction tubes were incubated in 37°C for 3 hours followed by heat inactivation of the enzyme at 65°C for 20 min.

### 2.2.6 Linearization quality control

As a quality control check, samples were electrophoresed on agarose gel to confirm linearization (0.7 % agarose gel). Only linear bands were selected for yeast transformation which promotes to have homologous recombination into the host genome.

### 2.2.7 Purification of linear recombinant plasmids

Purification of linear plasmids was done by using phenol chloroform method to remove the restriction enzyme and other digest components from the linear plasmids that might affect the upstream processing (Sambrook *et al.*, 1989). An equal volume of phenol: chloroform was added to the DNA samples followed by vortex for 30 seconds and

centrifugation at 8000 rpm for 1 min at room temperature to separate the phases. The upper aqueous phase was transferred to a clean 1.5 ml Eppendorf tube. Phenol wash step was repeated twice. Tris-EDTA (Ethylenediamine tetraacetic acid, TE) buffer of 35 µl was added, vortexed and centrifuged for 5 min at 12000 rpm at room temperature. DNA was precipitated by adding 0.1 volume of 3 M sodium acetate, pH 5.5 and 2 volumes of absolute ethanol (100%) to the samples and incubated at -20°C overnight or for 30 min at -80°C. The precipitated DNA was recovered by centrifugation at 12000 rpm for 15 min at 4°C. DNA pellet was washed with 2 volumes of 70% ethanol and incubated at room temperature for 5-10 min. The samples were centrifuged again for 5 min at 12000 rpm at 4°C, and the DNA pellets were dried from ethanol. The DNA was resuspended in 20 µl deionized distilled water and stored at -20°C for long-term storage.



### 2.2.8 Yeast transformation

#### 2.2.8.1 Yeast competent cells preparation

Yeast strain KM71H was streaked on yeast-peptone-dextrose (YPD) agar plates and kept in 30°C incubator for 2 days (www.invitrogen.com). A single colony was picked and inoculated in 5 ml YPD media and was grown overnight. A sample (1 ml) from the overnight culture was diluted into 50 ml of YPD media till the $O.D_{600}$ reached 0.178 Absorbance Units. The culture was incubated for 6 hours at 30°C till the $O.D_{600}$ reached 0.991 Absorbance Units. The cells were harvested by centrifugation at 500xg for 5 min at room temperature. The pellet was resuspended in 9 ml of ice-cold BEDS solution (bicine-NaOH, ethylene glycol, dimethyl sulfoxide (DMSO) and sorbitol) and supplemented with 1 ml DTT (1M). The cell suspension was incubated at 30°C for 5 min at 100 rpm. The cells were centrifuged at room temperature at 500xg for 5 min and resuspended in 1 ml of BEDS solutions. The competent cells were aliquoted (40µl) and stored at -80°C for transformation (Cereghino *et al*., 2005).

#### 2.2.8.2 Yeast transformation

Linearized DNA of 100 ng and 1µg were used in the transformation mix. DNA of 4µl was mixed with 40 µl of yeast competent cells in 0.2 cm electroporation cuevettes (Cat# 165-2086, Bio-Rad laboratories, Hercules, CA, USA), then incubated for 2 min on ice. Using Gene Pulser® II electroporator ; cuvettes gap, 2.0 mm; charging voltage, 1500 V; resistance 200 Ω ; capacitance, 50 µF.

Immediately after electroporation, the samples were resuspended in 0.5 ml sorbitol (1M) and 0.5 ml YPD, incubated in a 30°C incubator without shaking for 1 hour, and then plated on media containing increasing amount of zeocin (500, 1000 and 2000) µg/ ml for selection of multicopy integrants. Plates were incubated at 30°C for 2 days (www.invitrogen.com).

### 2.2.9 Yeast colony PCR

Yeast colony PCR is applied to identify the clones that are formed in the zeocin YPDA plates whether they contain pPICZαA with IFN-β.

By using a sterile tip, a single colony was picked from each plate of the following plasmids; pPICZαA blank, pPICZαA with WT, GASCO and JCat optimized IFN-β to 30 µl of 0.2% SDS (www.fhcrc.org/labs/hahn/methods/mol_bio_meth/pcr_yeast_colony)

After vortexing for 15 seconds, the samples were incubated in heat-block for 4 min at 90°C which is an important step for DNA extraction and for consistent PCR results. The samples were centrifuged for 1 min at high speed (14000 rpm) and the supernatant containing DNA was then removed to a new tube and stored at -20°C (Table 11).

**Table 11:** Yeast colony PCR experiment reaction mix

| Component | Volume (µl) |
|---|---|
| **5X GoTaq® Green master mix** | 12.5 |
| **5' AOX primer** | 1.25 |
| **3' AOX primer** | 1.25 |
| **25% triton** | 1 |
| **Yeast DNA** | 5 |
| **Water** | 4 |
| **Volume** | 25 |

The PCR conditions were programmed as followed:

| | | | |
|---|---|---|---|
| Initial denaturation | 95°C | | 3 min |
| Denaturation | 95°C | 30 Cycles | 1 min |
| Annealing | 51°C | | 1 min |
| Extension | 72°C | | 2 min |
| Final Extension | 72°C | | 7 min |
| Hold | 10°C | | Forever |

After the completion of PCR, 10 µl of the PCR product was loaded directly into 0.7% agarose gel for assay and the PCR product can be stored in 4°C.



### 2.2.10 Replica plating

After detecting the PCR products using agarose gel electrophoresis, A single positive colony of pPICZαA/IFN-B: (WT, GASCO and JCat optimized) Mut$^S$ KM71H *Pichia pastoris* transformants along with pPICZαA blank were selected by using a sterile tip and grown in YPDA plates supplemented with 100 µg/ml Zeocin.

Each clone represents different DNA and Zeocin concentration used for yeast transformation as shown in Figure 9 (Lederberg and Lederberg, 1952).

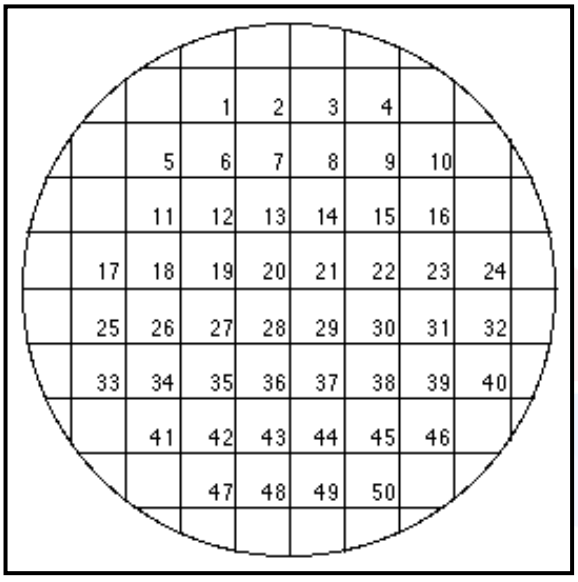



**Figure 9:** Scheme used for replica plating experiment.

### 2.2.11 Recombinant IFN-β production in *P. pastoris* (small scale)

A single colony of 3 different pPICZαA/IFN-B: (WT, GASCO and JCat optimized) Mut$^S$ KM71H *Pichia pastoris* transformants along with pPICZαA blank and *Pichia pastoris* strain as controls from the YPDA plates were grown in 50 ml of YPD medium at 30°C overnight with shaking at 250 rpm (www.invitrogen.com).

The cultures were centrifuged at 1500 x g for 5 min and then pellets were resuspended in 5 ml of minimum methanol medium (BMMY) supplemented with 0.8% glycerol. To maintain induction, methanol was added to the culture medium every 24 h at a final concentration of 1 %.

After each induction, 0.5 ml of culture were centrifuged at 12,000xg for 20 min and the supernatants were stored at -70°C or analyzed immediately by SDS-PAGE (Sommaruga *et al*., 2011)

### 2.2.12 Protein analysis by SDS-PAGE

Sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (1970) using 15% (v/v) polyacrylamide gels in denaturing conditions.

SDS-PAGE maintains polypeptides in a denatured state once they have been treated with strong reducing agents (β-mercaptoethanol) to remove secondary and tertiary structure (e.g. disulfide bonds [S-S] to sulfhydryl groups [SH and SH]) and thus allows separation of proteins by their molecular weight. Sampled proteins become covered in the negatively charged SDS and move to the positively charged electrode through the acrylamide mesh of the gel. Smaller proteins migrate faster through this mesh and the proteins are thus separated according to size (usually measured in kilodaltons, kDa). The concentration of acrylamide determines the resolution of the gel - the greater the acrylamide concentration the better the resolution of lower molecular weight proteins. The lower the acrylamide concentration the better the resolution of higher molecular weight proteins. Proteins travel only in one dimension along the gel. Samples are loaded into wells in the gel. One lane is usually reserved for a marker or ladder, a commercially available mixture of proteins having defined molecular weights, typically stained so as to form visible, colored bands. When voltage is applied along the gel, proteins migrate into it at different speeds. These different rates of advancement (different

electrophoretic mobilities) separate into bands within each lane (Weber and Osborn, 1969).

The electrophoresis took place for about 2 hours (80 Volt) and the proteins were then fixed using a fixing solution (50% methanol, 20% acetic acid, 30% water) for 15 minutes and detected by Coomassie Brilliant Blue R-250 (Cat# 161-0436, Bio-Rad laboratories, Hercules, CA, USA) for 1 hour. Gel was destained using a Coomassie Brilliant Blue R-250 destaining Solution (Cat# 161-0438, Bio-Rad laboratories, Hercules, CA, USA) for overnight (Laemmli ,1970).

### 2.2.13 RNA analysis

#### 2.2.13.1 RNA extraction from yeast cells

RNA was extracted from the yeast pelleted cells by using TRIzol Reagent® (Cat# 15596-026, Invitrogen, Groningen, the Netherlands). TRIzol® Reagent is a monophasic solution of phenol and guanidine isothiocyanate suitable for isolating total RNA from cells and tissues (www.invitrogen.com). During sample homogenization or lysis, TRIZOL® Reagent maintains the integrity of the RNA, while disrupting cells and dissolving cell components to facilitate the immediate and most effective inhibition of RNase activity. 1 ml of TRIzol reagent was added to the yeast cells, and then the homogenate was stored for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes. 0.2 ml of chloroform was added to the samples and mixed vigorously for 15 min. The resulting mixture was stored at room temperature for 10 min and centrifuged at 12000 g for 15 min at 4°C. Following centrifugation, the mixture separated into a lower red-phenol chloroform phase, interphase and the colorless upper aqueous phase. RNA remained exclusively in the aqueous phase, DNA in the interphase and proteins in the organic phase. The aqueous phase was transferred to a fresh tube and precipitated with 500 ml isopropanol. The samples were stored at room temperature for 10 min and centrifuged at 12000 g for 8 min at 4°C. Supernatant was discarded and RNA pellet was washed by vortexing with 200 µl of 75% ethanol (prepared with diethyl pyrocarbonate water, DEPC-water) and subsequent centrifugation at 12000 g for 5 min at 4°C. Ethanol was removed and RNA

pellet was briefly air-dried for 5 min. RNA was resuspended in 20 µl DEPC-water and incubated for 12 min at 55°C to complete RNA dissociation (www.invitrogen.com).

#### 2.2.13.2 RNA quality control

The quality of RNA was observed on 0.7% agarose gel, stained with ethidium bromide and visualized under the U.V light. The purity of RNA extraction was evaluated by ratio $OD_{260/280}$ equivalent to 1.8 using Nanodrop spectrophotometer.

#### 2.2.13.3 RT-PCR

##### 2.2.13.3.1 cDNA synthesis

By using high capacity cDNA Reverse Transcription Kit (Cat# 4374966, Applied Biosystems, USA) as shown in table 12. It is used to synthesize single-stranded cDNA from total RNA using reverse transcriptase. cDNA can be obtained both from prokaryotes and eukaryotes. RT-PCR is mainly used for gene expression studies. It determines whether the gene of interest is expressed or not, and the level of its expression.

**Table 12:** Components used for cDNA synthesis.

| | **Volume (µl)/ Reaction kit** | |
|---|---|---|
| **Component** | With RNase Inhibitor | Without RNase Inhibitor |
| **10X RT Buffer** | 2.0 | 2.0 |
| **25X dNTP Mix (100 mM)** | 0.8 | 0.8 |
| **10X RT Random Primers** | 2.0 | 2.0 |
| **MultiScribe Reverse Transciptase** | 1.0 | 1.0 |
| **RNase Inhibitor** | 1.0 | - |
| **Nuclease-Free Water** | 3.2 | 4.2 |
| **Volume** | 10.0 | 10.0 |

It uses the RT random primer scheme for initiating cDNA synthesis. Random primers ensure that the first strand synthesis occurs efficiently with all species of RNA molecular present, including mRNA and rRNA. 1 µg of RNA sample was used (10µl).

As a negative control, a total of master mix was added excluding the reverse transcriptase enzyme, to ensure that we do not have genomic DNA contamination. Another negative control consisting of water instead of RNA sample.

The PCR conditions were programmed as followed (www.appliedbiosystems.com):

| | | |
|---|---|---|
| Step 1 | 25°C | 10 min |
| Step 2 | 37°C | 120 min |
| Step 3 | 85°C | 5 min |
| Step 4 | 10°C | Forever |

**2.2.13.3.2 PCR**

PCR (DNA Engine Tetrad 2 Peltier Thermal Cycler, Bio-Rad) was done for the cDNA samples by using IFN-β primers. Table 13 represents the components used for PCR.

**Table 13:** PCR component

| Component | Final concentration | Volume (µl) |
|---|---|---|
| **5X GoTaq Buffer** | 1X | 5 |
| **dNTP 2mM** | 0.2 mM | 2.5 |
| **$MgCl_2$ 25mM** | 4mM | 4 |
| **5' IFN-β primer 100µM** | 0.5µM | 1.25 |
| **3' IFN-β primer 100µM** | 0.5µM | 1.25 |
| **GoTaq Enzyme** | 2 u / reaction | 0.4 |
| **cDNA template 100ng/µl** | 20ng/ reaction | 5 |
| **Water** | - | 5.6 |
| **Volume** | - | 25 |

The PCR conditions were programmed as followed (http://www.k-state.edu/hermanlab/protocols/StandardPCRConditions.html):

| | | | |
|---|---|---|---|
| Initial denaturation | 95°C | | 3 min |
| Denaturation | 95°C | 33 Cycles | 1 min |
| Annealing | 60°C | | 1 min |
| Extension | 72°C | | 2 min |
| Final Extension | 72°C | | 7 min |
| Hold | 10°C | | Forever |

As a quality control check, 0.7% agarose gel was done stained with ethidium bromide and visualized under U.V light.

### 2.2.14 Western blotting and immunodetection

After the detection of the proteins with denaturing SDS-PAGE gels, proteins were first concentrated using microconcentrators (centricon-10, Product # 4205, Amicon; USA). Concentration is achieved by ultrafiltration of solution through the concentrator membrane, with convective removal of low-molecular-weight solutes and solvent. The material emerges as filtrate and is collected in the concentrator's filtrate cup. The driving force for filtration is provided by centrifugation at 1000-5000 x g and the concentrate is recovered by capping the sample reservoir with the retentate cup, inverting the device, then centrifuging at 300-1000 x g (Amicon, 1985).

The concentrated proteins were transferred from gels to a Polyvinylidene fluoride (PVDF) membrane (Amersham Hybond-P, Product # RPN2020F, GE Healthcare; UK) by using a semi-dry blot apparatus (UniBlot SC serie, Brenzel BioAnalytik; Germany), which was followed by immunodetection. The primary method for transferring the proteins is called electroblotting and uses an electric current to pull proteins from the gel into the PVDF or nitrocellulose membrane.

The proteins move from within the gel onto the membrane while maintaining the organization they had within the gel. As a result of "blotting" process, the proteins are exposed on a thin surface layer for detection. Both varieties of membrane are chosen for their non-specific protein binding properties (i.e. binds all proteins equally well). Protein binding is based upon hydrophobic interactions, as well as ionic interactions between the membrane and protein. Nitrocellulose membranes are cheaper than PVDF, but are far more fragile and do not stand up well to repeat probings. The uniformity and overall effectiveness of transfer of protein from the gel to the membrane can be checked by staining the membrane with Coomassie Brilliant Blue or Ponceau S dyes. Ponceau S is the more common of the two, due to its higher sensitivity and water solubility, the latter making it easier to subsequently destain and probe the membrane (Corley, 2005).

The membrane was incubated for 1 h in Tris-buffered saline (TBS) containing 1% (w/v) bovine serum albumin (BSA, Cat#SC-2323, Chem Cruz, Santa Cruz Biotechnology, INC; CA, USA) and 0.05% Tween 20 (USB, Cat# 20605; Cleveland, USA).

Monoclonal Interferon-beta antibody (produced by inoculating mouse with human interferon beta, Mybiosource.com, Cat# MBS 140078; USA) was diluted to 1/200 antibody dilution with TBS containing 0.1% (w/v) BSA and 0.05% Tween 20.
The membrane was incubated for overnight at 4ºC with gentle rocking. Then it was washed for 5 times each 5 min duration with TBS containing 0.05% Tween 20 and one time without Tween 20 for 5 min.

After that the secondary antibody was applied which was "donkey anti-mouse IgG Horse raddish peroxidase" (Santa Cruz, Cat# SC 2306, D1511, Santa Cruz Biotechnology, INC; CA, USA), diluted into 1:4000 with TBS containing 0.1% (w/v) BSA and 0.05% Tween 20 and it was incubated for 1 hour at room temperature with gentle rocking.

The membrane was washed for 5 times each 5 min duration with TBS containing 0.05% Tween 20 and one time for 5 min with only TBS solution. Finally, in order to detect membrane – bound secondary antibody, 0.5 µl of solution A (luminal solution) and B (peroxide solution) of ECL prime western Blotting Detection Reagent (GE Healthcare, Amersham, Lot# 4641456; UK) were mixed together and applied to the membrane in dark room for 5 min. X-ray film processer (Ecomax, FI-45; FI-Sales, LLC; Germany) was used to detect the bands.

# Chapter 3

## 3.1 Results

As described in the aim of the study the main purpose of this work was to generate a recombinant yeast *P. pastoris* strain (Mut$^S$ KM71H) that produces human IFN-β from codon optimized cDNA sequences.

The experimental work was carried out in two steps:

1. Gene optimization, cloning of the human IFNβ cDNA into the PICZαA vector and transformation of the yeast *P. pastoris* by homologous recombination.
2. Production of native and optimized recombinant human IFNβ protein by the yeast *P.pastoris*.



### 3.1.1 Gene optimization

#### 3.1.1.1 *In-silico* sequence analysis

The purpose of redesigning IFNβ was to create a synthetic gene that would be expressed at high level in yeast cells. We first analyzed The human wild-type (WT) IFN-β cDNA sequence, GASCO- and Jcat- IFN-β-opt. cDNA sequences was also analyzed using the "graphical codon usage analyzer" software (GCUA) for codon usage frequencies versus yeast codon usage bias. Table 14 shows the *IFN-β* gene sequence used in our study.

**Table 14:** *IFN-β* gene sequence used in the study.

| Protein | Gene sequence length (nt) | GeneBank Accession Number | Putative Host |
|---|---|---|---|
| **Interferon Beta** | 561 | NP_002167.1 | Yeast |

Analysis of the WT-IFN-β gene revealed that 14.17% deviates significantly from the yeast *P. pastoris* codon usage table. Analysis of the optimized sequences generated using GASCO and JCat optimization softwares' showed 14.88% and 32.09% of the codons respectively of their codons that significantly deviate from the yeast *P. pastoris*.

The analysis of each triplet position for the WT-IFN-β versus the *P. pastoris* codon usage table revealed the positions of the codons which are most likely to be optimized to the desired expression system, *P. pastoris* (Figures 10-12).

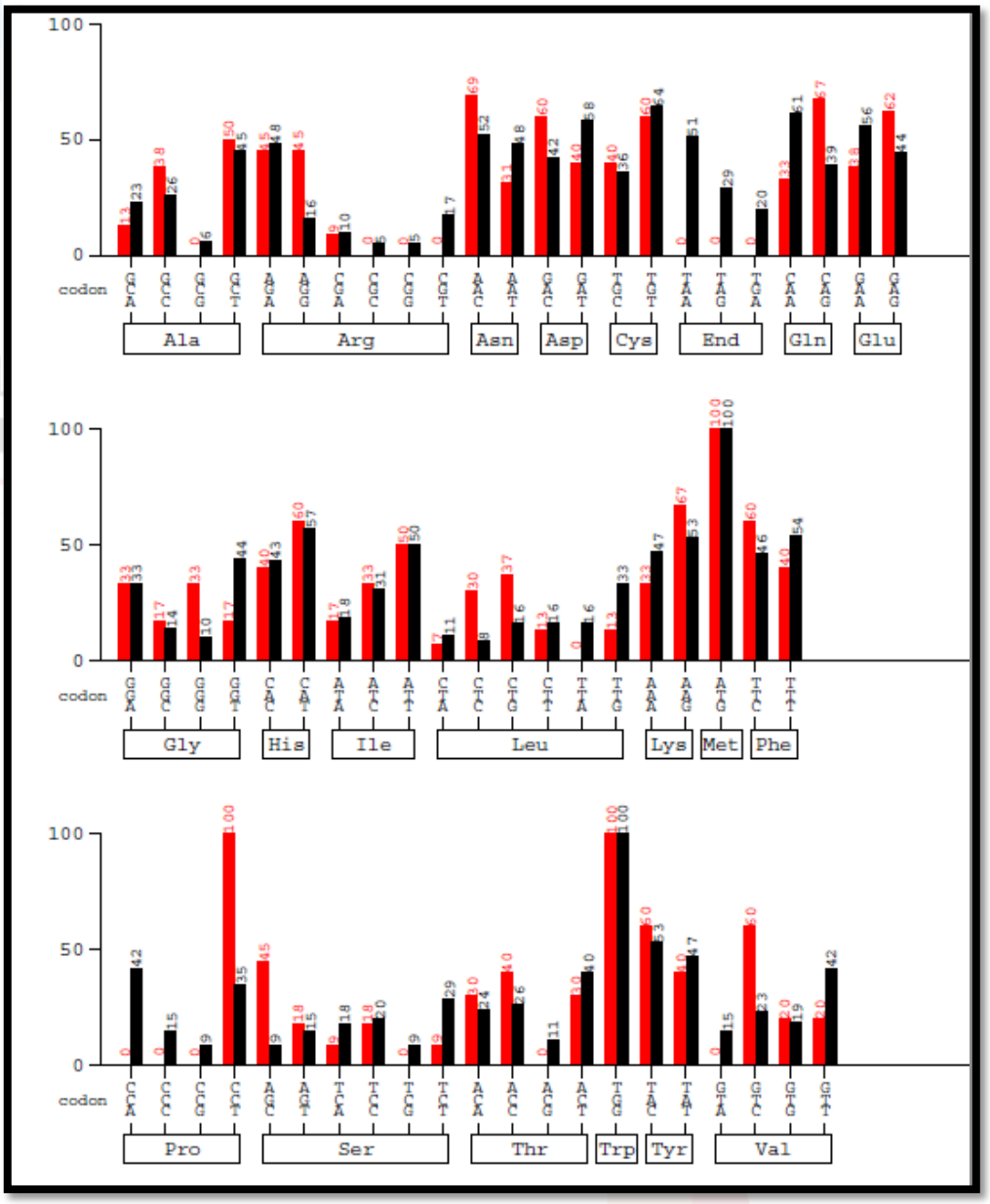



**Figure 10:** Codon analysis of the WT-IFN-β gene sequence using "graphical codon-usage analyzer" software. Triplet codons of the gene analyzed versus yeast (*Pichia pastoris*) codon usage table by GCUA software. **Red**: Wild-type IFN-B; **black**: *Pichia pastoris*. 14.17% is the mean difference between both codon usage tables.

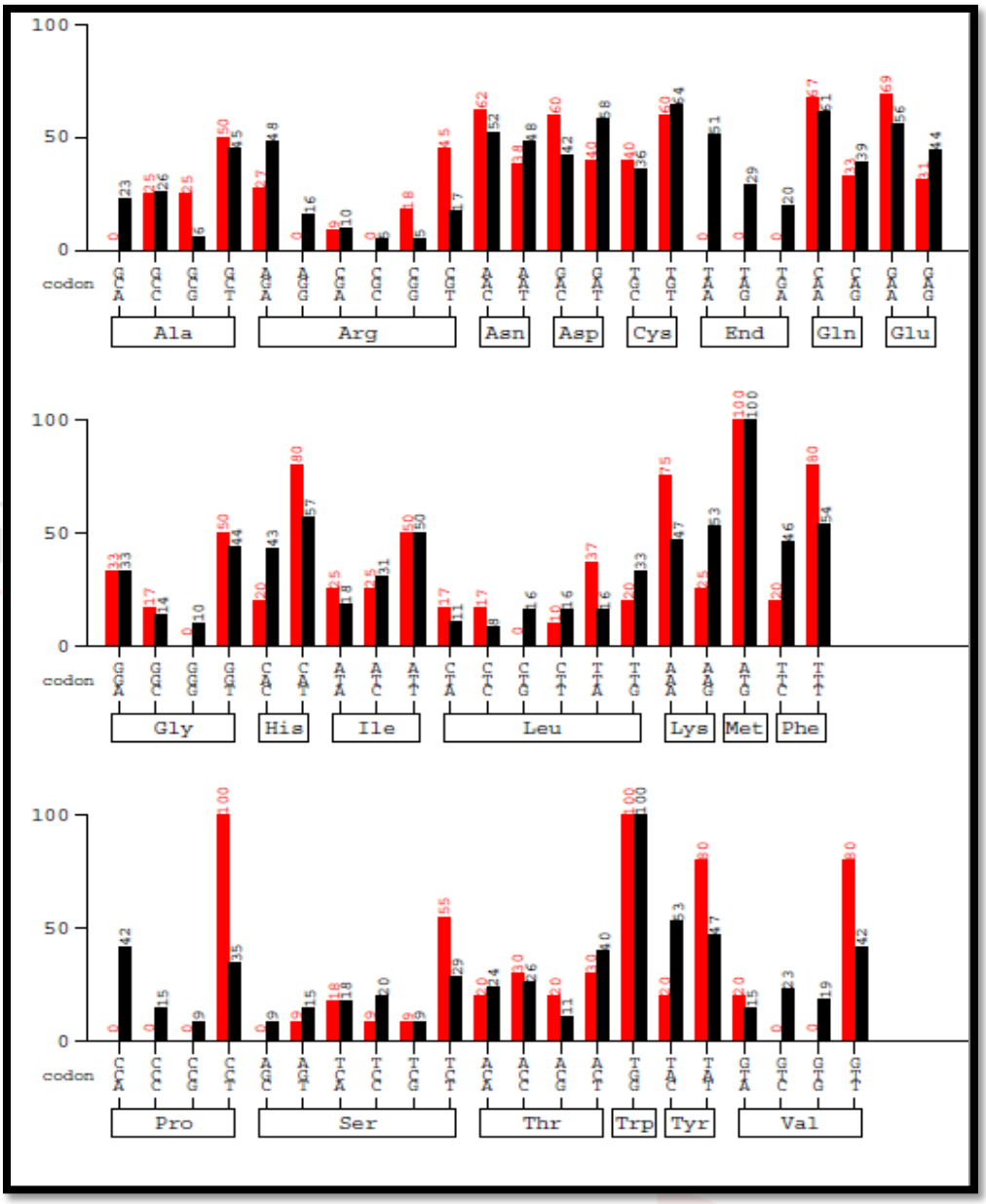



**Figure 11:** Codon analysis of the GASCO-IFN-β opt. gene sequence using "graphical codon-usage analyzer" software. Triplet codons of the gene analyzed versus yeast (*Pichia pastoris*) codon usage table by GCUA software. **Red**: GASCO- IFN-β opt.; **black**: *Pichia pastoris*. 14.88% is the mean difference between both codon usage tables.

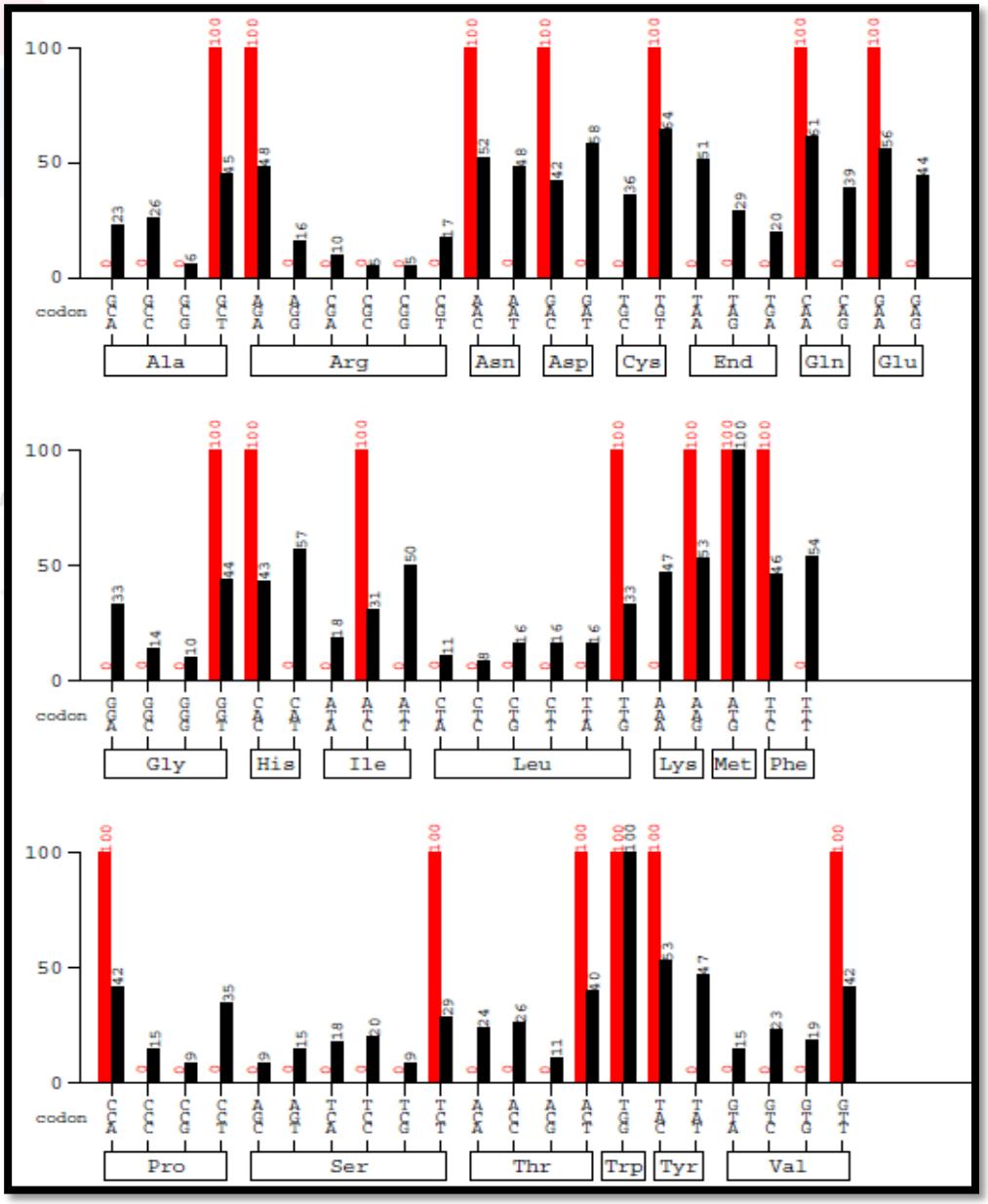



**Figure 12:** Codon analysis of the JCat-IFN-β opt. gene sequence using "graphical codon-usage analyzer" software. Triplet codons of the gene analyzed versus yeast (*Pichia pastoris*) codon usage table by GCUA software. **Red**: JCat opt. IFN-B; **black**: *Pichia pastoris*. 32.09% is the mean difference between both codon usage tables.

In figures 13-15, the red bars show the codons rarely used by *P.pastoris*, less than 10%. The blue bars indicate the codons that are used at 20% or less. A threshold of 20% and 10% for codon usage indicate that the abundance of cognate tRNA in the putative host cell is less than 20% which affects the translation rate of the sequence.

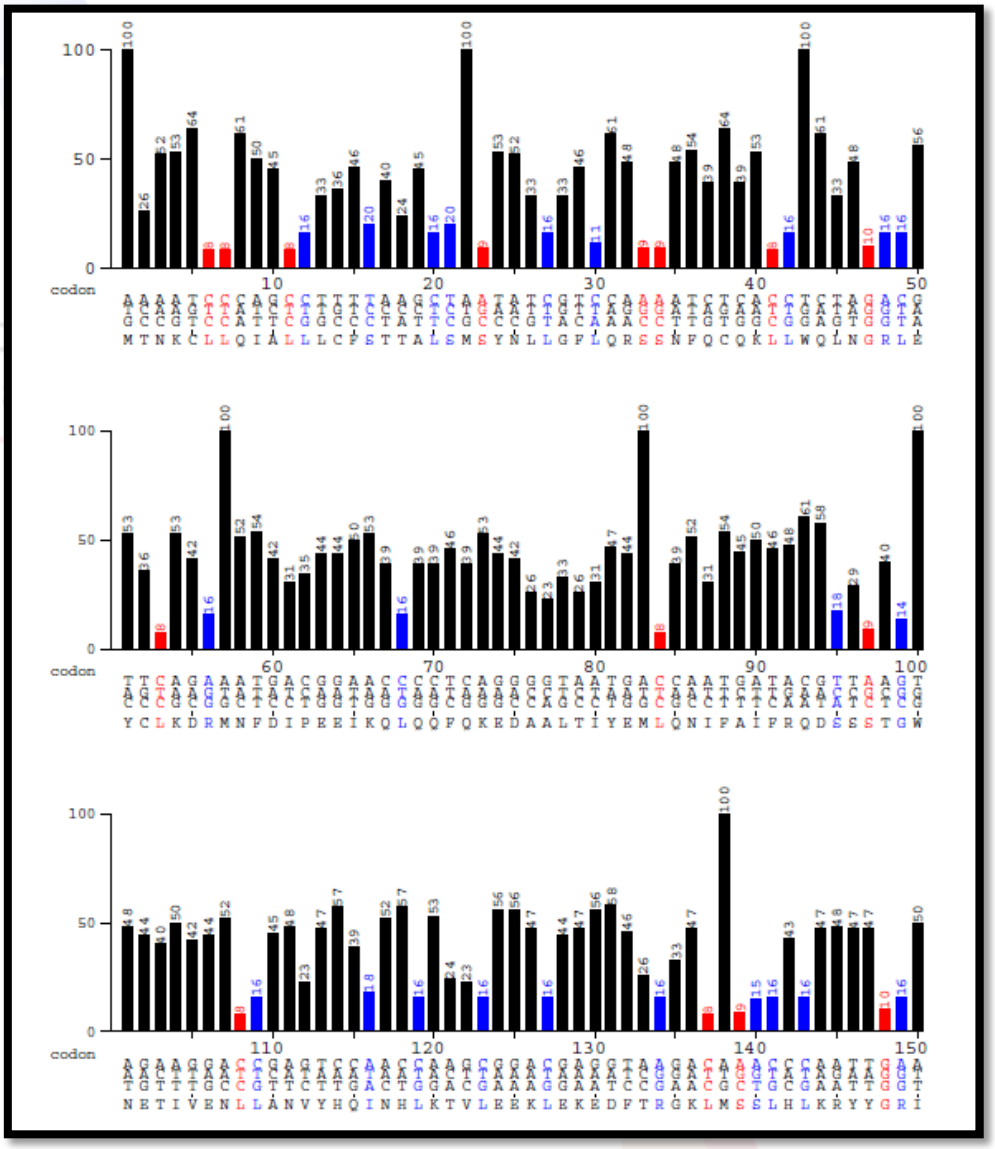


**Figure 13:** WT-IFN-β gene sequence frequency analysis. Triplet codons position of the gene analyzed versus yeast (*Pichia pastoris*) codon usage table by GCUA software. **Red** = **10%** frequency threshold, **Blue** = **20%** frequency threshold.

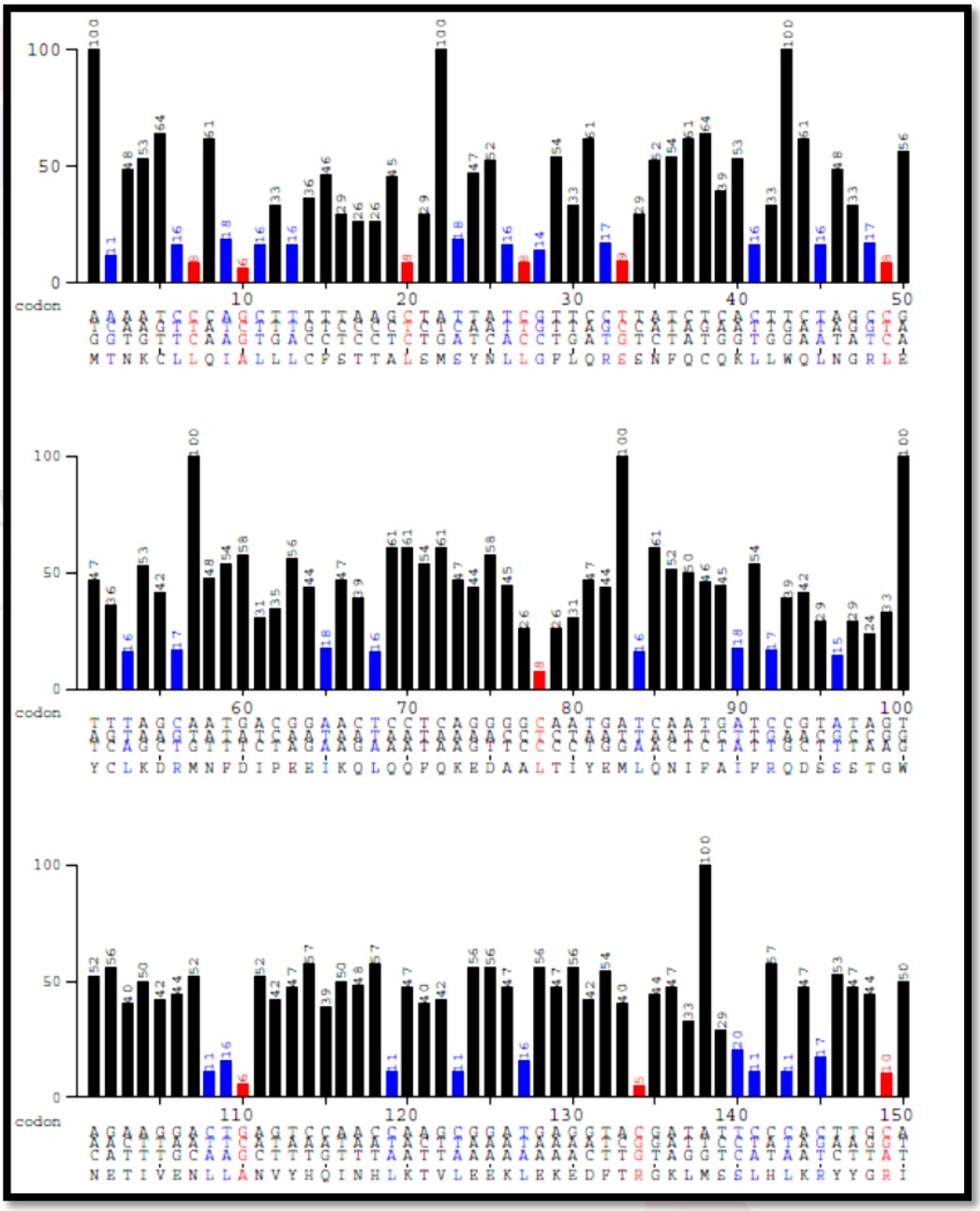



**Figure 14:** GASCO-IFN-β opt. gene sequence frequency analysis. Triplet codons position of the gene analyzed versus yeast (*Pichia pastoris*) codon usage table by GCUA software. **Red** = **10%** frequency threshold, **Blue** = **20%** frequency threshold.

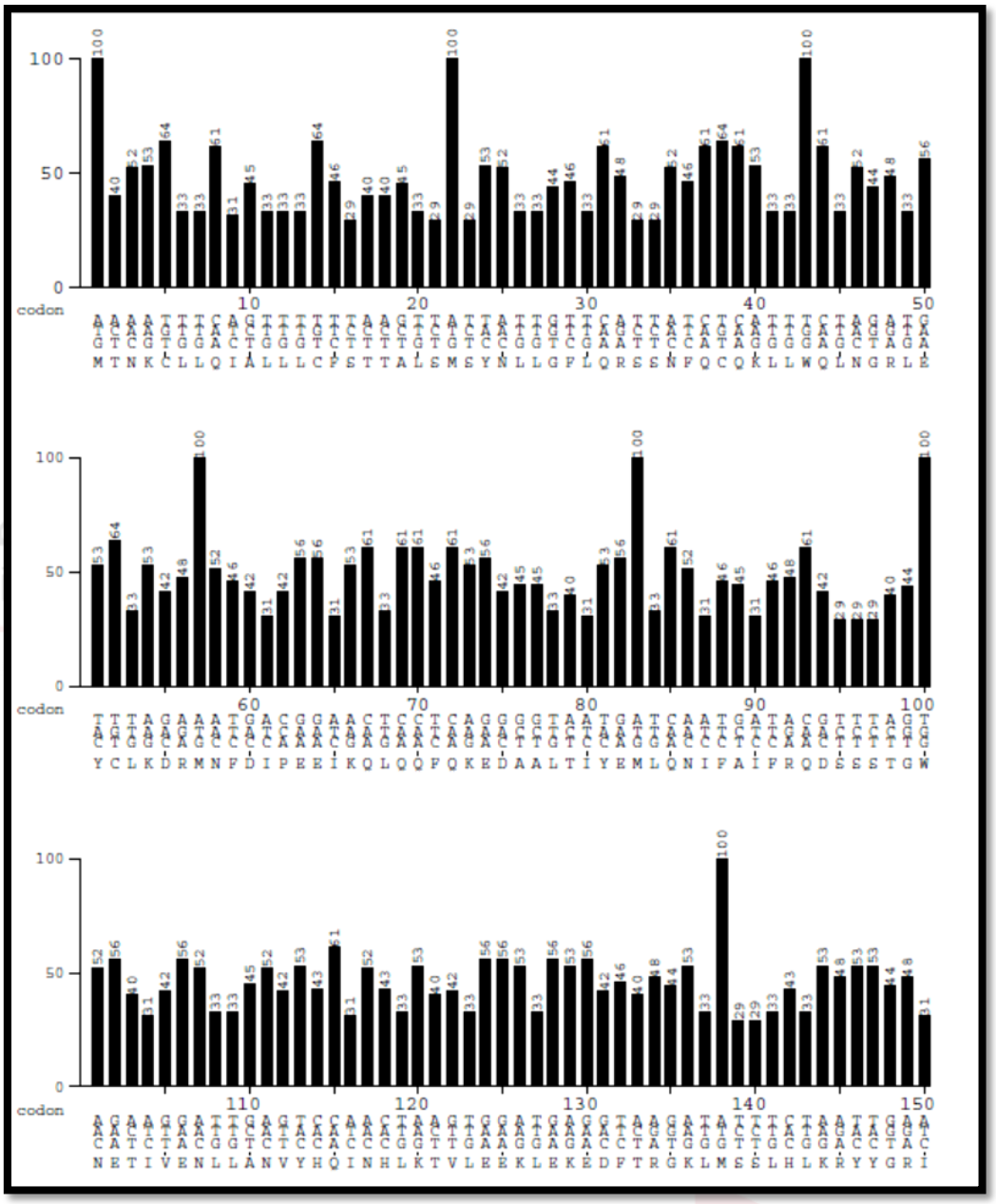


**Figure 15:** JCat-IFN-β opt. gene sequence frequency analysis. Triplet codons position of the gene analyzed versus yeast (*Pichia pastoris*) codon usage table by GCUA software. **Black:** All values are above 20% frequency threshold.

### 3.1.1.2 Comparison of WT-, GASCO and JCat- optimized IFN-β cDNA sequences

The nucleotide sequence of the native IFN-β gene (WT-IFN-β) was modified using GASCO and JCat softwares' by replacing the less preferred codons with the respective highly preferred degenerate codon, so as to match the yeast codon usage without altering the amino acid sequence (Figure 16).

**Wild-Type IFN-B cDNA**

ATGACCAACAAGTGTCTCCTCCAAATTGCTCTCCTGTTGTGCTTCTCCACTACAGCTCTTTCCATGAGC
TACAACTTGCTTGGATTCCTACAAAGAAGCAGCAATTTTCAGTGTCAGAAGCTCCTGTGGCAATTGAAT
GGGAGGCTTGAATACTGCCTCAAGGACAGGATGAACTTTGACATCCCTGAGGAGATTAAGCAGCTGCAG
CAGTTCCAGAAGGAGGACGCCGCATTGACCATCTATGAGATGCTCCAGAACATCTTTGCTATTTTCAGA
CAAGATTCATCTAGCACTGGCTGGAATGAGACTATTGTTGAGAACCTCCTGGCTAATGTCTATCATCAG
ATAAACCATCTGAAGACAGTCCTGGAAGAAAAACTGGAGAAAGAAGATTTCACCAGGGGAAAACTCATG
AGCAGTCTGCACCTGAAAAGATATTATGGGAGGATTCTGCATTACCTGAAGGCCAAGGAGTACAGTCAC
TGTGCCTGGACCATAGTCAGAGTGGAAATCCTAAGGAACTTTTACTTCATTAACAGACTTACAGGTTAC
CTCCGAAAC

**GASCO opt. IFN-B cDNA**

ATGACGAATAAGTGTCTTCTCCAAATAGCGCTTTTGTTATGCTTCTCTACCACCGCTCTCTCTATGTCA
TATAACTTACTCGGCTTTTTGCAACGTTCGTCTAACTTTCAATGTCAGAAGCTTTTGTGGCAATTAAAT
GGACGTCTCGAATATTGCTTAAAGGACCGTATGAATTTTGATATCCCTGAAGAGATAAAACAGTTACAA
CAATTTCAAAAAGAGGATGCTGCCCTCACCATCTATGAGATGTTACAAAACATTTTCGCTATATTTCGT
CAGGACTCTAGTTCTACAGGATGGAACGAAACTATTGTTGAGAACCTATTAGCGAACGTTTATCATCAG
ATTAATCATCTAAAAACTGTTCTAGAAGAAAAATTAGAAAAAGAAGACTTTACTCGGGGTAAATTGATG
TCTTCCCTACATCTAAAACGTTACTATGGTCGAATTTTACACTATTTAAAAGCCAAAGAATATTCACAT
TGTGCTTGGACGATCGTAAGAGTTGAAATTTTGAGAAACTTTTACTTTATTAATCGGTTAACAGGTTAT
TTGAGAAAC

**JCat opt. IFN-B cDNA**

ATGACTAACAAGTGTTTGTTGCAAATCGCTTTGTTGTTGTGTTTCTCTACTACTGCTTTGTCTATGTCT
TACAACTTGTTGGGTTTCTTGCAAAGATCTTCTAACTTCCAATGTCAAAAGTTGTTGTGGCAATTGAAC
GGTAGATTGGAATACTGTTTGAAGGACAGAATGAACTTCGACATCCCAGAAGAAATCAAGCAATTGCAA
CAATTCCAAAAGGAAGACGCTGCTTTGACTATCTACGAAATGTTGCAAAACATCTTCGCTATCTTCAGA
CAAGACTCTTCTTCTACTGGTTGGAACGAAACTATCGTTGAAAACTTGTTGGCTAACGTTTACCACCAA
ATCAACCACTTGAAGACTGTTTTGGAAGAAAAGTTGGAAAAGGAAGACTTCACTAGAGGTAAGTTGATG
TCTTCTTTGCACTTGAAGAGATACTACGGTAGAATCTTGCACTACTTGAAGGCTAAGGAATACTCTCAC
TGTGCTTGGACTATCGTTAGAGTTGAAATCTTGAGAAACTTCTACTTCATCAACAGATTGACTGGTTAC
TTGAGAAAC

**Figure 16:** WT-, GASCO- and JCat-IFN-β cDNA optimized sequences.

Native IFN-β cDNA, GASCO- and JCat-IFN-β-opt. cDNA sequences were analyzed for the GC content (GC%), AT% at the 5' tail and mRNA free energy. As shown in Table 15 comparison of the three sequences of IFN-β revealed differences in the GC%. it is interesting to note that JCat- IFN-β opt. sequence has 38 % GC content as compared to native sequence which had 45 % GC content (7% decrease from the native sequence, GASCO: 5% decrease from the native sequence). However, a slight decrease from the native sequence in AT% count (in the 5' tail) in the JCat- IFN-β opt.

sequence was detected. The free energy is a parameter reflecting the stability of mRNA secondary structure. Here again, JCat-opt. sequence exhibits a more stable mRNA structure than the GASCO opt. sequence.

**Table 15:** IFN-β cDNA sequences comparisons and analysis.

| GeneBank Accesion number (native) | cDNA Sequence | GC% | AT%5' tail (First 6 codons) | ΔG mRNA KCal/mol | Similarity to native (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | Native | GASCO | JCat |
| **NP_002167.1** | **IFN Beta** | **Yeast** | **Yeast** | **Yeast** | | | |
| **Native** | | 45 | 56 | -174.76 | **100** | 74.0 | 74.7 |
| **GASCO** | | 40 | 72 | -134.97 | 74.0 | **100** | 73.1 |
| **Jcat** | | 38 | 47 | -154.38 | 74.7 | 73.1 | **100** |

### 3.1.2 Construction of the expression vectors

The *in silico* modified IFN-β gene as well as the native IFN-β sequence were custom synthesized at GeneCust, Luxemburg, with 5' *Eco*RI and 3' *Not*I as flanking restriction sites.

In order to produce human IFNβ protein in yeast, PICZαA plasmid was used for recombinant vector construction in which different cDNA sequences of IFN-β (Wild-type and optimized sequences) were inserted.

### 3.1.2.1 DNA sequencing

The final synthetic sequences were verified by sequencing which showed 100% sequence similarity with *in silico* modified gene sequences (Figure 17) (Complete sequence alignment is shown in the annex section 9).

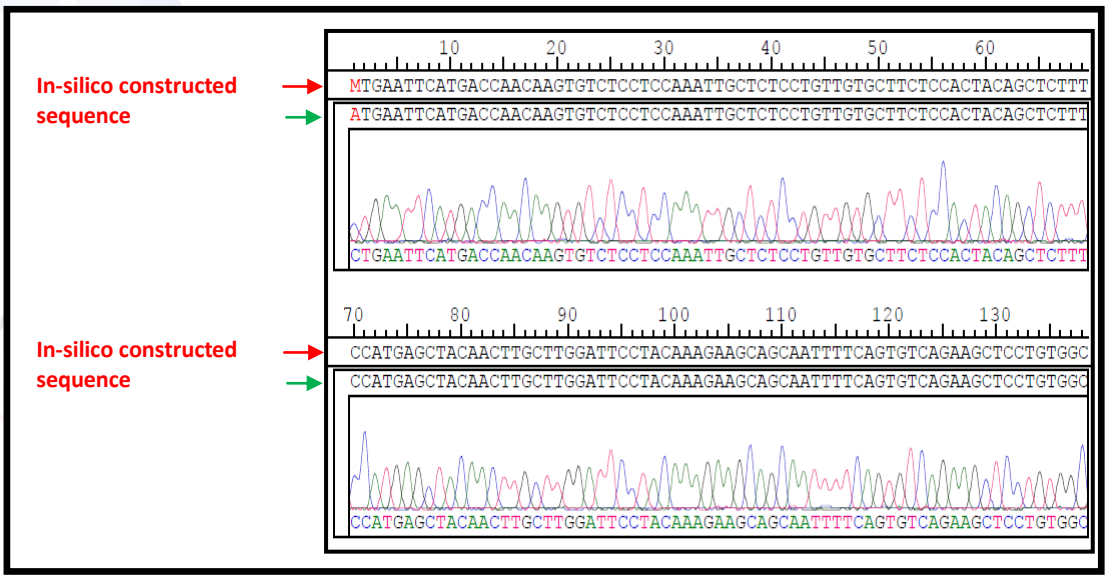



**Figure 17:** Illustration of alignment of wild-type IFN-β *in-silico* sequence with the synthesized wild-type IFN-β by GeneCust.

### 3.1.2.2 Plasmids structure and sizing

The preparation of uncut plasmids DNA appear more in supercoiled or covalently closed-circular DNA (ccc) [over 90%] that is fully intact plasmid with both strands uncut and with a twist built in, resulting in a compact form (Figure 18).

Supercoiled plasmid has certain shape in the gel as it interacts with more molecules of stain such as ethidium bromide, tending to make it more fluorescent than other DNA forms. Furthermore, the band of the WT, GASCO and JCat pPICZαA/IFNβ are about 600 bp larger than the pPICZαA Blank, indicating that they contain an insert.

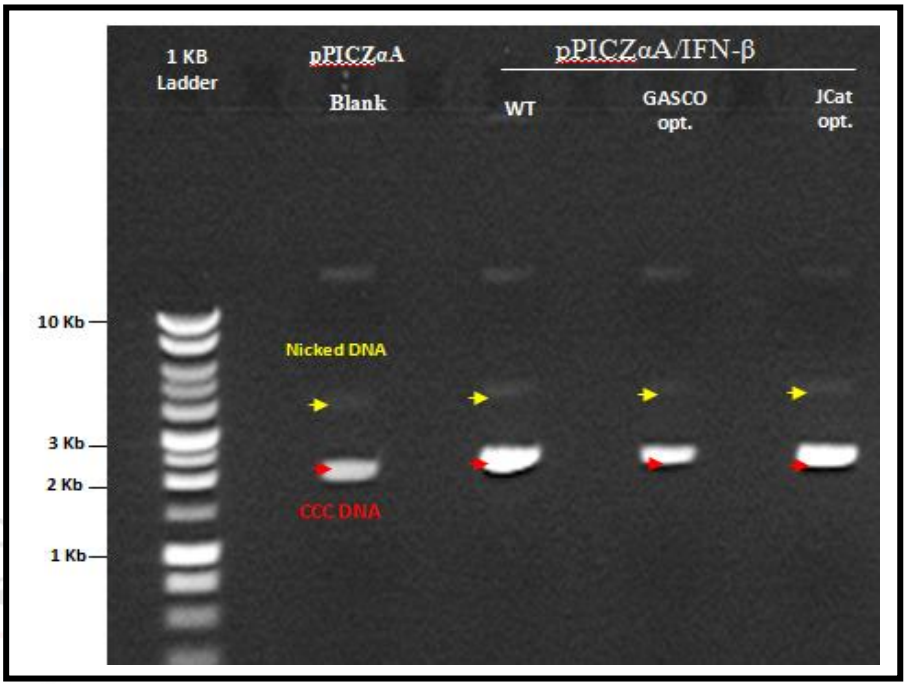



**Figure 18:** Quality Control of plasmid preparations. Evaluation of the ratio PICZαA plasmids ccc versus the other forms is carried out through the visual estimation of the Fluoresecence due to Ethidium Bromide bound to the plasmid DNA. Agarose gel 0.7%, DNA ladder 1Kb, pPICZαA blank (without insert), pPICZαA/IFN-β plasmid with: native (WT) IFN-β sequence, Gasco-IFN-β optimized (opt.) sequence, JCat-IFN-β optimized (opt.) sequence.

Plasmids were propagated in *E. coli* Top10F' strain. To confirm the integration of insert, double digestion of the expression vector was carried out using *Xho*I and *Xba*I restriction enzymes as those enzymes are located in the flanking region of the insert. The results presented in Figure 19 show that digested pPICZαA WT and JCat-IFN-β opt. plasmids have the same 2 bands, a 580 bp band which is the IFNβ insert and the upper band (labeled with a star) indicate the rest of the plasmid after digestion. However, digestion of pPICZαA GASCO-IFN-β opt.

yielded two small bands: 370 bp and 200 bp due to the presence of *Xba* I restriction site in the IFN-β optimized sequence, in addition to the upper band indicating the rest of the plasmid after digestion. No band of size 580 bp was observed after digestion of pPICZαA blank plasmid.

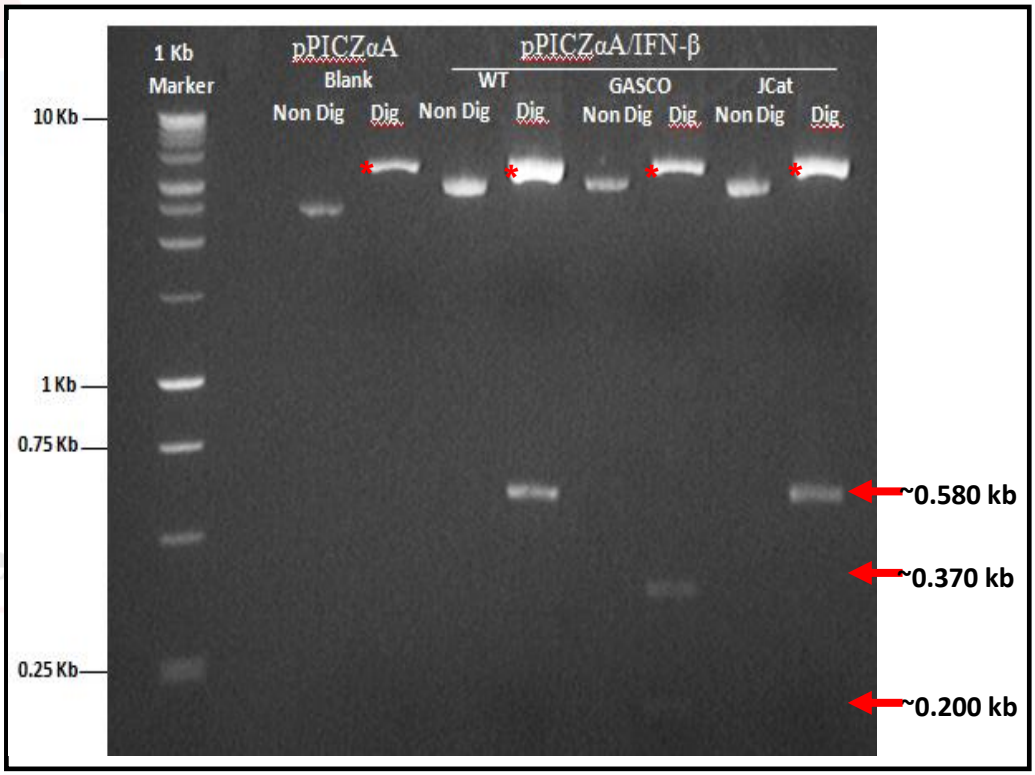



**Figure 19:** Restriction analysis to confirm integration of IFN-β insert into PICZαA plasmid. Double digestion of plamids containing the insert IFN-β to show the integrity of the insert, non dig: Non digested, dig; digested, Blank; plasmid without insert, WT; Wildtype, Gasco; Gasco-IFN-β opt. sequence, JCat: JCat-IFN-β opt. sequence. Agarose gel 0.7%, DNA ladder 1Kb.

### 3.1.3 Transformation efficiency

Transformation efficiency is calculated as a function of plasmid DNA concentration in order to ensure the high quality of the plasmids prior to yeast transformation. Table 16 illustrates the transformation efficiency from bacterial Top10F' *E.coli* LB agar plates supplemented with Zeocin as selection marker. The results showed that the Jcat and GASCO opt. IFN-β has the highest and lowest transformation efficiency, respectively. The calculation was done in a web application (http://www.sciencegateway.org/tools/transform.htm).

**Table 16:** Transformation efficiency.

| Plasmids | Colonies | CFU/ml | Transformation Efficiency (Transformant/ µg DNA) |
|---|---|---|---|
| **pPICZαA/Wild-type IFN-β** | 888 | 44,400 | $2.44 \times 10^5$ |
| **pPICZαA/GASCO-IFN-β opt.** | 172 | 8,600 | $4.73 \times 10^4$ |
| **pPICZαA/JCat-IFN-β opt.** | 1604 | 80,200 | $4.41 \times 10^5$ |

### 3.1.4 Yeast transformation

Prior to linearization and yeast transformation, plasmids: pPICZαA WT, GASCO- and JCat-IFN-β opt. and pPICZαA blank were precipitated using sodium acetate and ethanol to increase the quantity and the quality of DNA.

As shown in figure 20, more than 90% of the precipitated plasmids are in supercoiled (ccc) form.

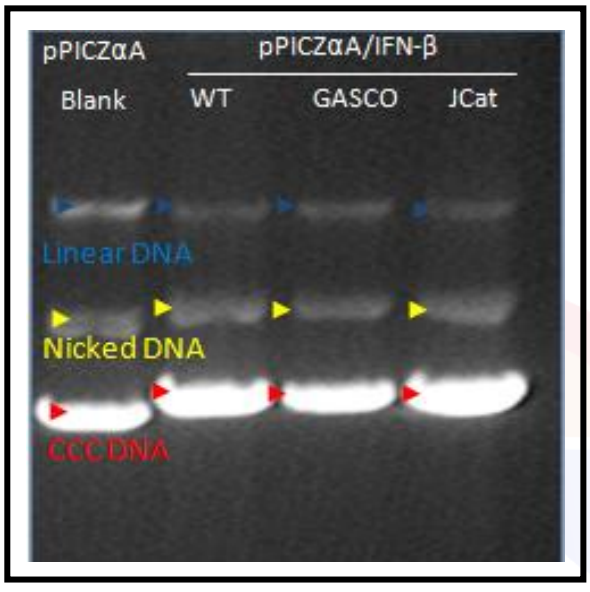



**Figure 20:** Quality Control of plasmid after precipitation step. Agarose gel (0.7%) stained with ethidium bromide and visualized under U.V. **>90%** of plasmid PICZαA is under the ccc form.

### 3.1.4.1 Linearization of plasmids

Plasmids were linearized with *Bst*X I in order to promote integration into the yeast genome. *Bst*X I restriction enzyme does not have a restriction site in the cDNA IFN-β sequences and has a restriction site in the 5'AOX1 region (707 bp upstream of plasmid). Linearization efficiency was checked by agarose gel electrophoresis (Figure 21).

Samples containing one band of expected size (4.0 kb) were selected for yeast transformation. Digestion is incomplete when the reaction yields more than one band.

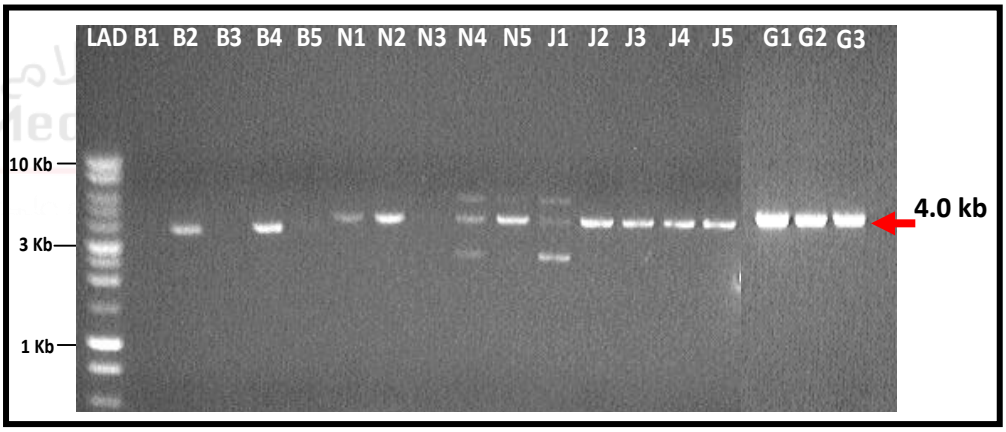



**Figure 21:** Quality Control of plasmid linearization. Agarose gel stained with ethidium bromide, **1 Kb LAD:** 1 kb DNA ladder, **B1-B5:** pPICZαA blank**, N1-N5:** pPICZαA Wildtype (Native) IFN-β**, J1-J5:** pPICZαA/JCat-IFN-β opt., **G1-G3:** pPICZαA/GASCO-IFN-β opt.

Prior to yeast electroporation, linearized plasmids were purified by phenol/chloroform to enhance the quality of DNA and the efficiency of yeast transformation.

The bands in PICZαA/IFNβ WT, GASCO and JCat are linearized and contain an insert about 600 bp (Total size 4.0 Kb, Figure 22).

The band formed in the line PICZαA Blank is linear but having a smaller size than the bands of 4.0 Kb, indicating the absence of the insert.

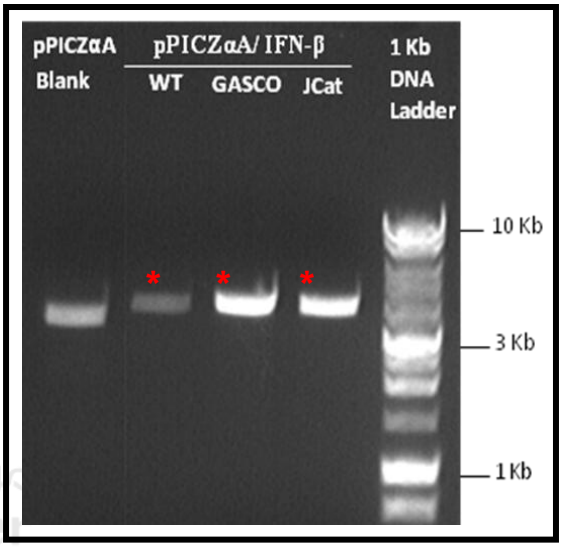



**Figure 22:** Quality Control of linear plasmid after phenol/chloroform purification. Agarose gel stained with ethidium bromide. 100% of plasmids PICZαA are in the linear form (* **4.0 Kb**).

### 3.1.5 Yeast colony PCR

The number of Zeocin-resistant colonies that were generated by electroporation transformation with Zeocin resistant-based plasmids decreased as the concentration of Zeocin increased.

For selection of Zeocin resistant transformants, we have plated the transformation mix on YPDA plates supplemented with Zeocin at 500, 1000 and 2000 µg/ml.
A subpopulation of total Zeocin-resistant transformats is hyperresistant to Zeocin (colonies appearing on plate containing Zeocin at 2000 µg/ml) and likely to represent multiycopy events. Representative colonies that arose on each of these different Zeocin concentrations were screened for plasmid insert by colony PCR.
Primers were designed to be specific for the AOX region of the plasmid, so the PCR product is an approximate of 1000 bp for the plasmids with IFN-β insert and around 600 bp for the empty pPICZαA (blank).

Each figure (23-25) represents 10 clones of different plasmids with corresponding features (amount of DNA used for yeast transformation and Zeocin concentration added to YPDA plate).

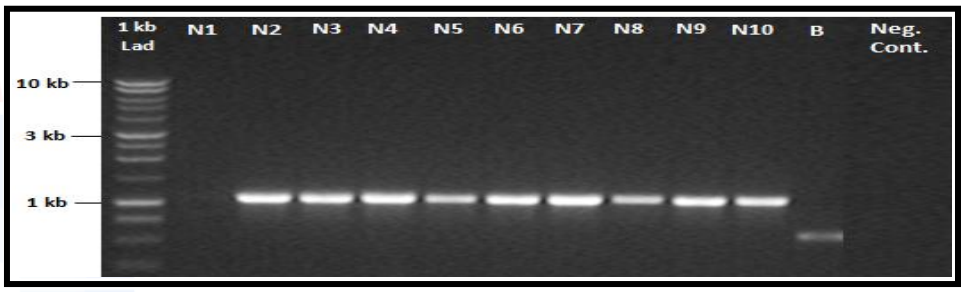



**Figure 23:** Yeast colony PCR for pPICZαA/Native IFN-β transformants. **1 Kb Lad:** 1 kb DNA ladder, **N1-N6:** 1µg DNA 1000 µg/ml zeocin. **N7-N10:** 100ng DNA 500 µg/ml zeocin. **B:** pPICZαA blank (without insert), **Neg cont**.: Reaction mix without DNA template. 0.7% agarose gel stained with ethedium bromide.

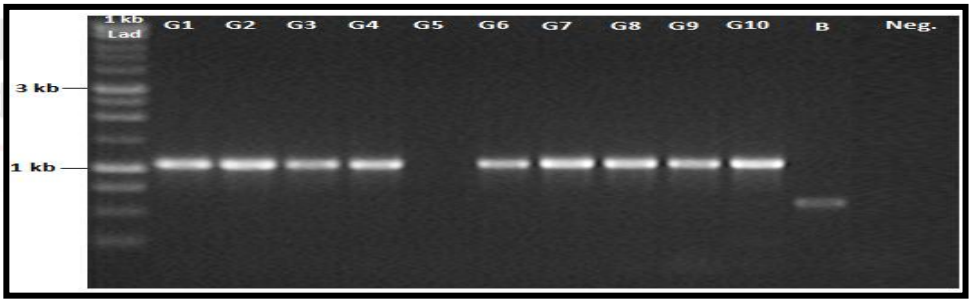



**Figure 24:** Yeast colony PCR for pPICZαA/GASCO-IFN-β opt. transformants. **1 Kb Lad:** 1 kb DNA ladder, **G1-G4:** 100 ng DNA 500 µg/ml zeocin. **G5-G8:** 1 µg DNA 500 µg/ml zeocin. **G9-G10:** 1 µg DNA 1000 µg/ml zeocin, **B:** pPICZαA blank (without insert), **Neg cont**.: Reaction mix without DNA template. 0.7% agarose gel stained with ethedium bromide.

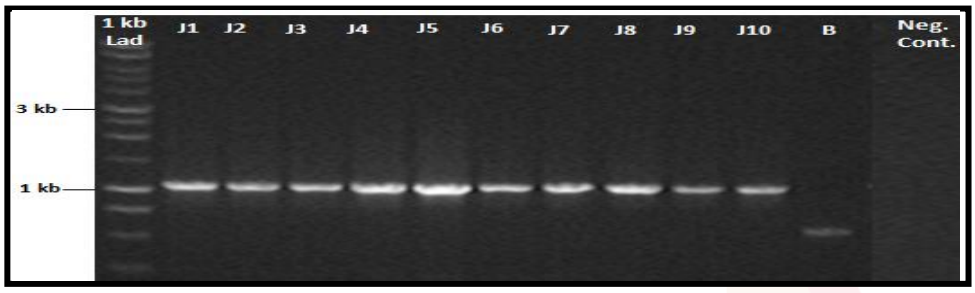



**Figure 25:** Yeast colony PCR for pPICZαA/JCat-IFN-β opt. transformants. **1 Kb Lad:** 1 kb DNA ladder, **J1-J2:** 1 µg DNA 1000 µg/ml zeocin. **J3-J4:** 100 ng DNA 1000 µg/ml zeocin. **J5-J6:** 1 µg DNA 2000 µg/ml zeocin, **J7-J8:** 100 ng 500 µg/ml zeocin **J9-J10:** 1 µg DNA 500 µg/ml zeocin **B:** pPICZαA blank (without insert), **Neg cont**.: Reaction mix without DNA sample. 0.7% agarose gel stained with ethedium bromide.

### 3.1.6 Replica plating

After choosing the positive clones showing the correct integration of *IFN-β* gene in the yeast genome, a number of clones were plated together in YPDA plates supplemented with 100 µg/ml zeocin (Figure 26).

Each clone was originated from a specific DNA quantity and zeocin concentration in YPDA plates.

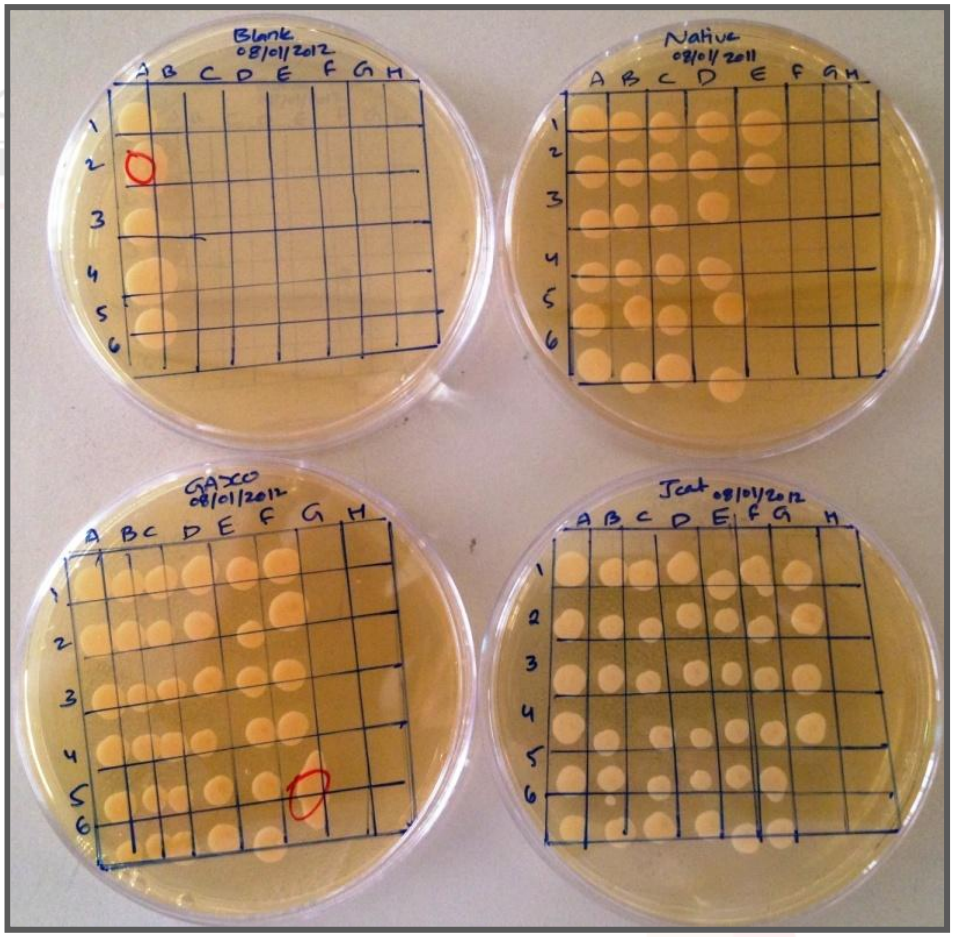



**Figure 26:** Replica plate of four different plasmids PICZαA (WT, GASCO- and JCat-IFN-β optimized and blank plasmid) in YPDA plates supplemented with 100 µg/ml zeocin.

### 3.1.7 Production of human IFN-β in *P. pastoris*

Three plasmids (pPICZαA/Native IFN-β, pPICZαA/Gasco-IFN-β opt. and pPICZαA/JCat-IFN-β opt.) were constructed in order to express the human *IFN-β* gene in *P. pastoris*. The expression of *IFN-β* gene was placed under the control of the *P. pastoris AOXI* promoter and the *S. cervisiae* α-factor secretion signal targeted the protein to the secretion pathway.

Culture supernatants from yeast carrying hyperresistant expression plasmids were analyzed by SDS/PAGE (15%) to evaluate the efficiency of recombinant IFN-β production in *P. pastoris.*

In Figures 27-31, comparing the pattern of protein production to the non induced clones we observed in the analyzed clones (C1, C2 and C3) from each construct (Native, Gasco- and JCat-IFN-β opt.) the appearance of two main bands with molecular masses of ~33 and ~25 kDa. However, Native (C2, Figure 27) and JCat (C1, Figure 29 and C3, Figure 31) clones exhibit a third band which have a molecular mass of 19 kDa.

This band appears after extended incubation (from 72 to 120h induction). Likewise, after extended incubation, some clones (Figures 30 and 31) exhibited an additional band with molecular mass of 23 kDa.

The size of the band 33 kDa is in agreement with that predicted for the glycosylated fusion protein (IFN-β + α-factor secretion signal).

Conversely, the band with molecular mass of 19 kDa is in accordance with that predicted for the non glycosylated IFN-β (166 aa). The increase in molecular mass from 19 to 25 kDa (and 23 kDa) would be the result of glycosylation.

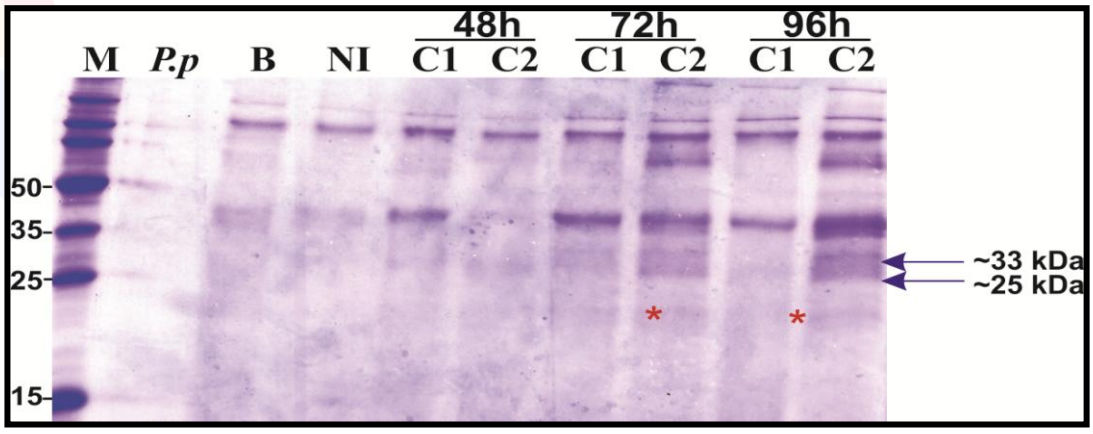



**Figure 27:** SDS-PAGE analysis of time course of recombinant IFN-β production in *P. pastoris* strain ($Mut^S$ KM71H). Culture supernant from *P. pastoris* carrying pPICZαA/ **Native** IFN-β clones (C1 and C2) after different timing of methanol induction in small scale production of protein. **M:** Molecular mass marker (kDa), ***P. p:*** *Pichia* strain, **B:** pPICZαA without insert, **NI:** Non induced clone at 96 h. **C1 and C2:** pPICZαA/Native IFN-β clones grown on YPDA plate supplemented with 1000 µg/ml Zeocin . Red star (*) represents bands of size ~19 kDa.

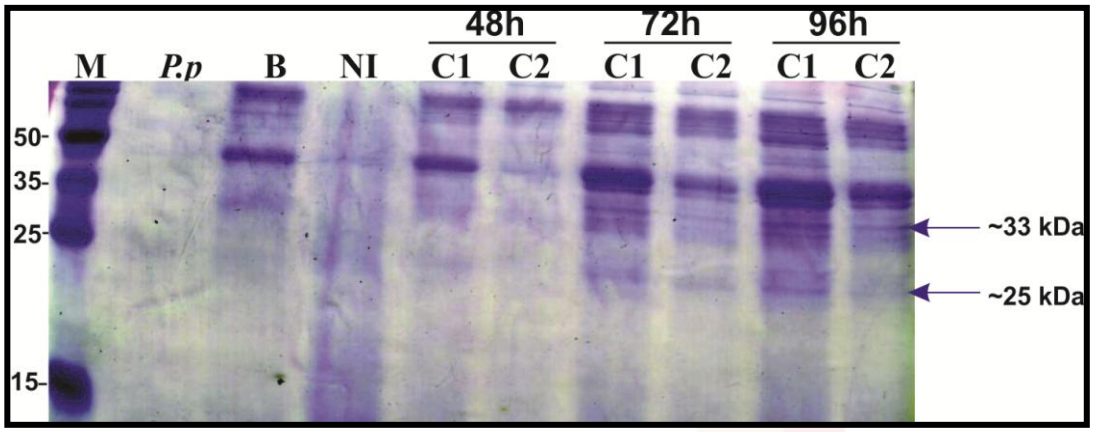



**Figure 28:** SDS-PAGE analysis of time course of recombinant IFN-β production in *P. pastoris* strain ($Mut^S$ KM71H). Culture supernant from *P. pastoris* carrying pPICZαA/ **GASCO** opt. IFN-β clones (C1 and C2) after different timing of methanol induction in small scale production of protein**. M:** Molecular mass marker (kDa), ***P.p:*** *Pichia* strain, **B:** pPICZαA without insert, **NI:** Non induced clone at 96 h. **C1 and C2:** pPICZαA/GASCO opt. IFN-β clones grown on YPDA plate supplemented with 1000 µg/ml Zeocin.

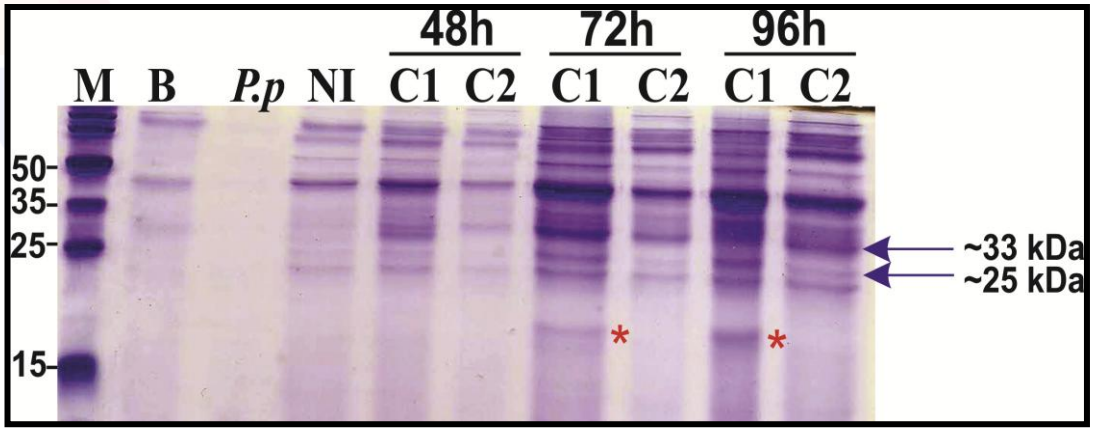



**Figure 29:** SDS-PAGE analysis of time course of recombinant IFN-β production in *P. pastoris* strain (Mut$^S$ KM71H). Culture supernant from *P. pastoris* carrying pPICZαA/ **JCat** opt. IFN-β clones (C1 and C2) after different timing of methanol induction in small scale production of protein**. M:** Molecular mass marker (kDa), **B:** pPICZαA without insert, ***P.p:*** *Pichia* strain, **NI:** Non induced clone at 96 h. **C1 and C2:** pPICZαA/JCat opt. IFN-β clones grown on YPDA plate supplemented with 2000 µg/ml Zeocin. Red Star (*) represents band of size ~19 kDa.

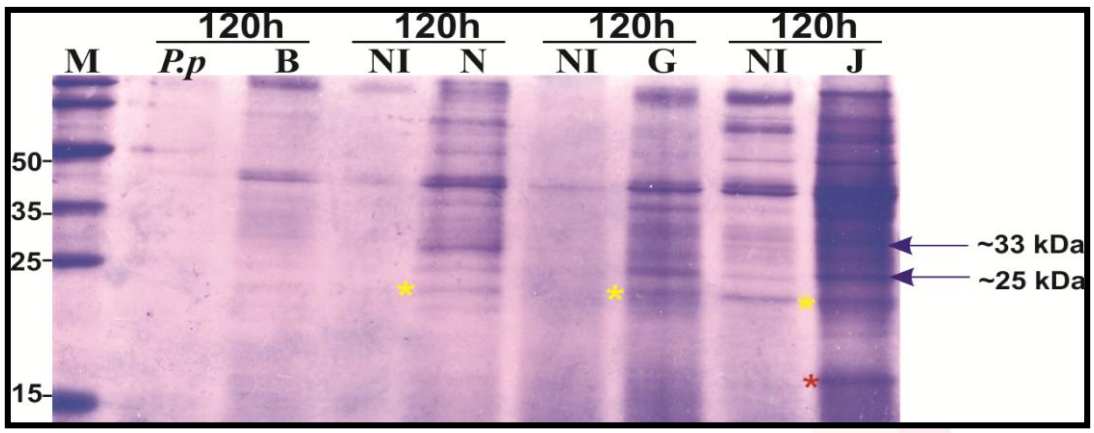



**Figure 30:** SDS-PAGE analysis of Culture supernant from *P. pastoris* carrying pPICZαA/ blank, Native **(N, C2)**, Gasco opt. **(G, C1)** and JCat opt. **(J, C1)** IFN-β clones after 120h of methanol induction in small scale production of protein. **M:** Molecular mass marker (kDa), ***P.p:*** *Pichia* strain, **B:** *P. pastoris* containing empty pPICZαA, **NI:** non induced culture for each transformant. Yellow star (*) represents band of size ~23 kDa. Red Star (*) represents band of size ~19 kDa.

**Figure 31:** SDS-PAGE analysis of human IFN-β. 15% SDS-PAGE of 20 µl sample of culture supernatant from a *P. pastoris* strain (Mut$^S$ KM71H) expressing human IFN-β Native (**N, C3**), GASCO- (**G, C3**) and JCat- (**J, C3**) optimized sequence. Cells were induced in BMMY media (supplemented with 0.8% glycerol) for 48, 72, 96. Lane **M** contains molecular mass markers (kDa), Yellow star (*) represents band of size ~23 kDa. Red Star (*) represents band of size ~19 kDa.

The highest production of recombinant IFN-β protein represented by 33, 25 and 19 kDa bands was obtained after 96h of induction with methanol.

### 3.1.8 Analysis of *IFN-β* gene expression in *P. pastoris* by RT-PCR

To investigate whether the poor IFN-β protein production, especially in Native and GASCO transformed yeast, was caused by low levels or absence of transcript, mRNA analysis experiment using RT-PCR was performed.

Figure 32 shows the presence of IFN-β transcripts in *P. pastoris* strain (Mut$^S$ KM71H) expressing human IFN-β Native (**N, C3**), GASCO- (**G, C3**) and JCat- (**J, C3**) optimized sequence. The integration of *IFN-β* native gene and optimized sequence was successful and already shown by colony PCR. Here we confirm that the transcription of *IFN-β* native gene and optimized sequence into mRNA occurs normally in yeast *P. pastoris*.

A visually distinct band at 600 bp, corresponding to the size of IFN-β sequence, was amplified by RT-PCR from total RNA extracted from yeast cells transformed with Native, GASCO- and JCat-IFN-β opt. plasmids. All of the bands are not of equal intensity. JCat-IFN-β opt. has the most intense band, followed by GASCO-IFN-β opt. and Native-IFN-β. no band of 600 bp was observed in cells transformed with the empty pPICZαA (Blank).

**Figure 32:** Detection of *IFN-β* gene expression by RT-PCR in cells of *P. pastoris* transformed with the referred plasmids. **1 Kb Lad:** 1 kb DNA ladder, **pPICZαA Blank:** yeast transformed with empty pPICZαA, **pPICZαA/ IFN-β Native, Gasco, JCat:** *P.pastoris* carrying pPICZαA/ Native, GASCO- and JCat- IFN-β opt. plasmids**,** **Pos.1 Cont. IFN-β:** PGEX4T-1 expressing human IFN-β in bacteria, **Neg. Cont. PCR:** PCR mix without cDNA template, **Neg. Cont. RT:** product of RT step without reverse transcriptase enzyme. 0.7% agarose gel stained with ethedium bromide.

### 3.1.9 Characterization of the recombinant IFN-β secreted by *P. pastoris*

Based on visual comparison of band intensity of SDS-PAGE Coomassie brilliant blue stained we selected the highest producer clones showing the bands with molecular masses of 33, 25 and 19 kDa to further characterize the recombinant IFN-β secreted by *P. pastoris*.

Western blot analysis was carried out with the recombinant human IFN-β protein produced from Native cDNA (clone C2) and from the optimized sequences GASCO (clone C1) and JCat (clone C1) in culture supernatants after 96h induction. Prior to conduct western blot analysis samples were concentrated to increase the amount of protein up to the detection level (Figure 33A).

On western blot analysis using a monoclonal antibody against human IFN-β, only sample from *P. pastoris* strain (Mut$^{S}$ KM71H) expressing human IFN-β JCat optimized

sequence (**J, C1**) revealed the presence of a major band with a molecular mass of ~33 kDa and a minor band of ~19 kDa (Figure 33B).

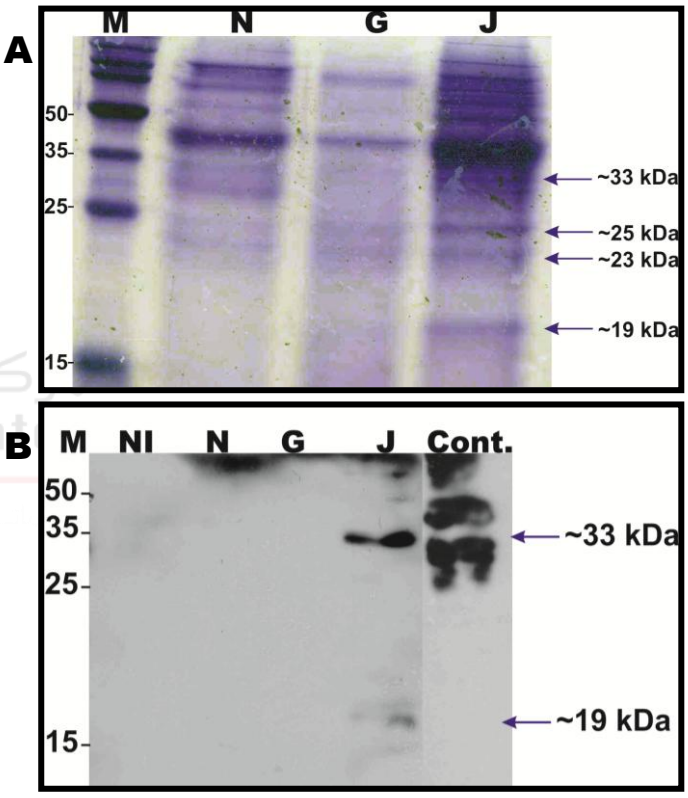



**Figure 33:** SDS-PAGE and Western blot analysis of recombinant human IFN-β protein expression in yeast *P. pastoris*. **A**. Concentrated culture supernatant from cultures of yeast carrying expression plasmids were seperated by SDS-PAGE (15%) and stained with Coomassie brillant blue. **B**. Concentrated culture supernatants (25 µl for all samples) were run on gel for blotting and probed with a monoclonal anti-human IFN-β antibody. **M:** Molecular mass marker (kDa), **NI:** non induced culture, **N**: Native **(**C2**),** **G:** Gasco opt. **(**C1**)** and **J**: JCat opt. **(**C1**)** IFN-β clones after 96h of methanol induction in small scale production of protein. **Cont.**: protein extract from *E. coli* expressing native recombinant human IFN-β protein (fusion protein GST-IFN-β) used as positive control for the antibody.

## 4.1 Discussion

Recent developments with respect to the *P. pastoris* system have had an impact not only on the expression levels that can be achieved, but also the quality of heterologous proteins (Benabdesselem *et al.*, 2007). INF-β has been already produced by *E. coli* as an unfolded protein and need refolding after expression (Utsumi *et al.*, 1986; Arnold *et al.*, 2009).

The most indicated surrogate system for the production of INF-β is the yeast *P. pastoris*. It has not been widely used so far for the production of INF-β although *P. pastoris* has a lot of advantages. Among these advantages is the economic and efficient protein production especially with high yield and the availability of strong inducible promoter. *P. pastoris* processes protein in a similar way to mammalian cells. It can carry out efficient protein folding and uses low cost growth media (Cregg *et al.*, 1993). *P. pastoris* has been used efficiently for high yield expression of INF-α (Liu *et al.*, 2001), bee hyaluronidase (Reitinger *et al.*, 2008), and also hepatitis B surface antigen (Vassileva *et al.*, 2001).

Recently, Rabhi & Fathallah (Patent number WO 2007/099462A2) achieved high level secretion of biologically active human IFN-α2b in the recombinant yeast *P. pastoris*. This result will facilitate the therapeutic study of hIFN-α2b as well as the mechanistic study of hIFN-α2b. This strategy has potential for the large scale production of biologically active rhIFN-B1 (Shi *et al.*, 2007).

The aim of this project was to use a gene optimization approach to develop a recombinant *Pichia pastoris* clone that produces soluble recombinant human interferon-beta (IFN-β) that can be ultimately used for the development of an IFN-β as bio-drug for the treatment of Arabian Gulf patients with multiple sclerosis. Constructing the optimal expression vector and the use of a simple eukaryotic system made it possible to produce high levels of stable, functional protein is the first step in this process. Also, constructing human IFN-β expression in the methylotrophic yeast *P. pastoris* strain Mut$^S$ KM71, was designed, introduced in the yeast and the expression levels assessed.

There are many genetic and physiological factors influencing the expression of target recombinant proteins in *P. pastoris* and many studies have outlined specific features

enabling an increase in expression level of a particular protein of interest. It seems universally accepted that there is no one solution applicable to the production of all, or even the majority of recombinant proteins, and the reasons behind the low expression level of some proteins remain unclear. In attempts to express human *IFN-β* in the methylotrophic yeast *P. pastoris*, codon optimization of the *IFN-β* gene was done to overcome any potential bottle neck caused by unavailability of rare tRNA molecules in the host system.

It has been well documented that different organisms show a preference for different subpopulations of the 61 possible codons, with the genes coding for abundant proteins using these major codons almost exclusively throughout their sequences. This concept of codon bias, where the codons utilized to encode a protein in one organism are substantially different from those used to code for the same protein in a different species, has long been recognized to have a profound impact on the expression of heterologous proteins (Grantham *et al.*, 1980a; Gouy and Gautier, 1982).

In this study, prior to cloning of the human IFN-β cDNA and its expression in yeast *P. pastoris*, cDNA codon-tuning was performed to assess whether codon optimization of the *IFN-β* gene would affect the level of recombinant protein expression by *P. pastoris* strain Mut$^{S}$ KM71H. An innovative gene optimization strategy was followed, which was developed in our laboratory by Salim *et al*., 2011 (submitted paper) for human heterologous gene expression and high yield production of recombinant protein, which can assist in biodrugs manufacturing in an innovative system.

First, the IFN-β cDNA sequence was analyzed using the nearest neighbor analysis "graphical codon usage analyzer" to assess the divergence of the codon usage from the *Pichia pastoris* preferred codon usage pattern. Then the codon optimization software GASCO (**G**enetic **A**lgorithm **S**imulation for **C**odon **O**ptimization, http://miracle.igib.res.in/gasco/) and JCat (**J**ava **C**odon **A**daptation **T**ool, http://www.jcat.de/) were used to generate two different codon-optimized sequences compatible with *P. pastoris* codon usage table.

Entering the wild-type IFN-β sequence into the Graphical Codon Usage Analyser program (GCUA) indicated that 14.17% of its codons are not the preferred for the yeast

*P. pastoris.* GASCO- and JCat- IFN-β opt. cDNA analysis showed 14.88% and 32.09% respectively of their codons that significantly deviate from the yeast *P. pastoris*.

Apparently, these results contradict with the basic concept of codon optimization for superior adaptability to the host cell codon usage table. However, when we take into account the main concept of the gene design and integrating parameters like GC% content, AT% in the 5'tail (first six codon), mRNA free energy, these results become coherent and show that JCat can incorporate rare codons instead for highly preferable ones for the advantage of higher expression by managing translation elongation [JCat opt. sequence: AT% = 47%, GASCO opt. sequence: AT% =72 and Native sequence: AT%= 56%], mRNA secondary structure stability (moderate free energy) and proper protein folding. The *in-silico* analysis that we performed on the wild-type, GASCO- and JCat-optimied sequences has given an insight of what to expect when carry out the recombinant protein production in yeast cells.

Synthesis of human IFN-β codon-optimized cDNA sequences was performed by GeneCust (Dudelange, Luxembourg). The synthetic 561 bp sequences were designed flanked by 5' *EcoR* I and 3 *Not* I sites, and were supplied as a cloned fragment contained within the pPICZαA vector. Sequencing analysis confirmed that pPICZαA contained the desired synthesized genes sequences cloned into the vector polylinker.
Optimized constructs were propagated through the transfer of the plasmids DNA to the *E. coli* Top10F' strain (recA- , endA-). This strain was chosen because of its genotype that includes the endA1 and the recA1 mutations. These mutations respectively prevent plasmid loss through endonuclease activity and plasmid DNA recombination. Top10F' *E. coli* are provided at a transformation efficiency of 1 x $10^9$ cfu/µg supercoiled DNA and are ideal for high-efficiency cloning and plasmid propagation. They allow stable replication of high-copy number plasmids (www.invitrogen.com).

The plasmids were produced from the corresponding recombinant *E. coli* strain, and quantified the preparations by spectrophotometry. The quality of those plasmids was controlled in agarose gel, which showed a satisfactory plasmids' purification indicated by the presence of the majority of each plasmid sample (around 90%) in a supercoiled form.

The methylotrophic yeast *Pichia pastoris* has become one of the leading eukaryotic expression systems for general laboratory use due to its ease of cultivation and the availability of a range of commercially available expression vectors. Indeed, strain and vector systems have been developed, which allow both intra- and extracellular protein production under the control of various promoters, the most common being the methanol-induced *alcohol oxidase 1* (*Aox1*) promoter (Tschopp *et al.*, 1987).

One of the key advantages of *P. pastoris* for recombinant protein production lies in the ability of this yeast to secrete proteins to the culture medium under the agency of the *Saccharomyces cerevisiae* α-factor secretion signal peptide. Furthermore, strains with a slow methanol utilization phenotype ($Mut^S$), which do not grow rapidly in the presence of methanol, may increase the stability of recombinant proteins that are difficult to synthesize, slow to fold, or must undergo other complex posttranslational modifications (Daly and Hearn, 2005).

Prior to transformation of *P. pastoris* strain $Mut^S$ KM71, linearization of plasmids with *Bst*X I favors recombination at the *Aox1* locus, as *Bst*X I has a restriction site in the 5'*Aox1* region. All Constructs: pPICZαA blank (empty plasmid), pPICZαA with native IFN-β sequence, GASCO- and JCat-IFN-β optimized sequence, sequences verified mutation free, were successfully introduced into the *P. pastoris* for expression and secretion of the resultant recombinant protein (human IFN-β). A variety of different clones for both codon optimized and native IFN-β were produced in an attempt to develop a clone that expresses soluble recombinant human IFN-β.

*P. pastoris* is capable of integrating multiple copies of transforming DNA into genomic sites of sequence homology. As successful integration of multiple copies of the gene of interest may increase the level of protein produced, determining which clones have multiple integrants is highly desirable. This was achieved by selection on YPDA plates containing increasing concentrations of Zeocin (selection marker). A subpopulation of total Zeocin-resistant transformants is hyperresistant to Zeocin (colonies appearing on plate containing Zeocin at 2000 µg/ml) and likely to represent multiycopy events.

Representative colonies that arise on each of these different Zeocin concentrations are screened for plasmid insert by colony PCR. In this project, 3 clones from each construct were selected based on their hyper resistance to Zeocin.

pPICZαA/Native and GASCO opt. IFN-β C1, C2 and C3 clones were grown in YPDA plate supplemented with 1000 µg/ml Zeocin and JCat-IFN-β opt. C1, C2 and C3 clones were grown in YPDA plate supplemented with 2000 µg/ml Zeocin.

To analyze expression of the recombinant human IFN-β, supernatant samples were assayed every 24 h post-induction over a time course of 120 h from selected clone's cultures grown at 30ºC.

SDS-PAGE followed by Coomassie Brilliant blue analysis of crude culture supernatants fractions shows a putative IFN-β specific bands of approximately 33 and 25 kDa present in all induced tested clones and are absent at all time points of the empty vector negative control and the non-induced cultures for each clones (Figures 27-31 in results section). Furthermore, Native (C2, Figure 27) and JCat (C1, Figure 29 and C3, Figure 31) clones exhibit a third band which have a molecular mass of 19 kDa. This band appears after extended incubation (from 72 to 120h induction). Likewise, after extended incubation, some clones (Figures 30 and 31) exhibit an additional band with molecular mass of 23 kDa.

Although this induced, secreted protein appears specific to strains containing IFN-β constructs, it consistently shows only low to extremely low levels (Native clone especially) of expression and is presumably beneath the detection level of immunoblot analysis when using anti- IFN-β antibody as the primary antibody for detection. However, there was an appreciable increase in expression from JCat codon optimized; Zeocin hyper resistant clones (especially clone C1).

It is worth mentioning that in this study selection for increased Zeocin resistance does not conclusively indicate that colonies contain multiple copies of the vector. The actual copy number could be determined by quantitative dot blot analysis or Southern blot techniques.

Prior to transformation, the entire open reading frame (ORF) of all constructs were sequenced to control for their integrity. Growth of transformants on appropriate media then showed successful integration had occurred for all these clones.

To confirm this successful integration, extensive PCR analysis (described in results section) was conducted. However, the negligible levels of recombinant protein detected

especially in clones carrying pPICZαA/Native IFN-β and pPICZαA/GASCO-IFN-β opt. necessitated analysis of expression at the RNA level. *P. pastoris* has been reported to produce truncated transcripts resulting in low or no protein expression, this is usually ascribed to RNAs containing AT-rich stretches or genes with codon biases very different from those of the organism in which the recombinant protein is being expressed (Scorer *et al.*, 1993; Sinclair and Choy, 2002).

Although GASCO codon optimised IFN-β transcript showed a higher AT% in the 5' tail than all the other sequences, it was still necessary to establish the point at which expression was being blocked. To ensure that our gene of interest was being actively transcribed, total RNA extracted from cell pellets of *P. pastoris* transformed by pPICZαA blank, pPICZαA/Native IFN-β, pPICZαA GASCO- and JCat-IFN-β opt plasmids. DNase treated and then oligo dT reverse transcribed to generate a cDNA template. Primers chosen for amplification lie within the extreme 3' end of the transcript, thus ensuring that only when complete transcription of the gene has successfully occurred. The presence of IFN-β transcripts, even though at different concentration, confirms that the *P. pastoris* expression system is effectively transcribing the integrated DNA. Thus, the observed low levels of recombinant protein expression indicate translational or posttranslational difficulties.

As western blot analysis on crude supernatant samples of clones carrying pPICZαA/Native IFN-β and pPICZαA/GASCO- and JCat-IFN-β opt. vector failed to elicit bands, western blot analysis on concentrated samples was undertaken. Samples concentrated from culture medium taken at 96 h post-induction were assessed, time point at which the highest expression level of recombinant human IFN-β was observed. Immunoblot analysis detected a clear intensive band of approximately 33 kDa and a second band less intensive of approximately 19 kDa, these bands appeared only in clones carrying pPICZαA/ JCat-IFN-β opt. This pattern of expression confirmed the bands seen in SDS-PAGE stained with Coomassie brilliant blue.

As found in literature (Runkel, Meier, Repinsky, *et al.*, 1998), the IFN-β size is around 19 kDa. The 33 kDa band represents the hybrid form of glycosylated protein, which means it contains the α-factor (~10 kDa). The 25 kDa band refers to the glycosylated protein after cleavage from the α-factor.

The impact of upstream sequences in the pPICZαA vector, such as the non native *S. cerevisiae* α-factor secretion signal, on IFN-β expression has not been determined. Furthermore, the lack of an apparent protein homologue to IFN-β in *P. pastoris* may mean the specific chaperone proteins needed for correct folding may also be absent in this yeast, resulting in degradation of the misfolded protein.

It has been suggested that the maximum level of protein secretion is ultimately determined by the protein folding capacity of the ER (Parekh and Wittrup, 1997). Secretory proteins enter the ER in an unfolded state and are modified and folded to acquire their functional conformations by ER-resident chaperones. Improperly folded proteins are then transported back to the cytosol and degraded by the ubiquitin-proteasome machinery to prevent toxification by the accumulation of aberrant proteins (Hirsch *et al.*, 2004; Romisch, 2005). High level synthesis of recombinant proteins could mean the components of the folding machinery become rate limiting, leading to inefficient folding of heterologous proteins (Inan *et al.*, 2006).

Incorrect posttranslational modification or lack of necessary chaperones for a specific recombinant protein could also give rise to incorrect folding of recombinant proteins, thus eliciting the cellular stress response leading to ER-associated degradation and resulting in the observed lack of secreted heterologous protein (Inan *et al.*, 2006).

In this study, to express IFN-β in methylotrophic yeast system many approaches were tried: selection of a $Mut^S$ KM71H strain of *P. pastoris* to more closely imitate the posttranslational modifications possible in mammals, codon optimisation of *IFN-β* gene to overcome any potential bottle neck caused by unavailability of rare tRNA molecules in the host system, assessing the impact of the alcohol oxidase pathways in terms of $Mut^S$ phenotypes, the affect of multiple integration events.

Preliminary experiments utilizing JCat-IFN-β opt protein have yielded promising results. Western blot assay indicates IFN-β protein of the expected approximately 33 and 19 kDa size has been produced.

These findings show that gene optimization by JCat software improves, the production of recombinant human IFN-β protein by yeast *P. pastoris.* Further optimization procedures at the level of the production process can be carried out to improve the yield.

When a gene is synthesized, it is generally modified from the natural version. These modifications are made to simplify subsequent manipulations (adding or eliminating restriction sites, for example), but also for a much more significant reason: natural genes are often poorly expressed in heterologous hosts, even when the expression system is related to the organism from which the gene originated.

The current trends in biotechnology for recombinant protein drugs call for gene optimization as an ultimate approach to enhance the yield and quality of recombinant product.

Several strategies have been developed to increase the production level of recombinant proteins in different expression host systems mainly *E. coli* (Makrides, 1980; Sorensen and Mortensen, 2005) and *P. pastoris* (Sreekrishna, 1997; Cereghino & Cregg, 2000; Daly & Hearn, 2005; Benabdesselem *et al*, 2007).

Design of an 'optimal' gene (which is defined as one in which the codon choices do not limit expression in addition to other factors) requires a thorough understanding of the interaction of the gene sequence with the expression environment and specification of the desired goal (expression level, solubility, localization of expressed protein, etc., Figure 34).

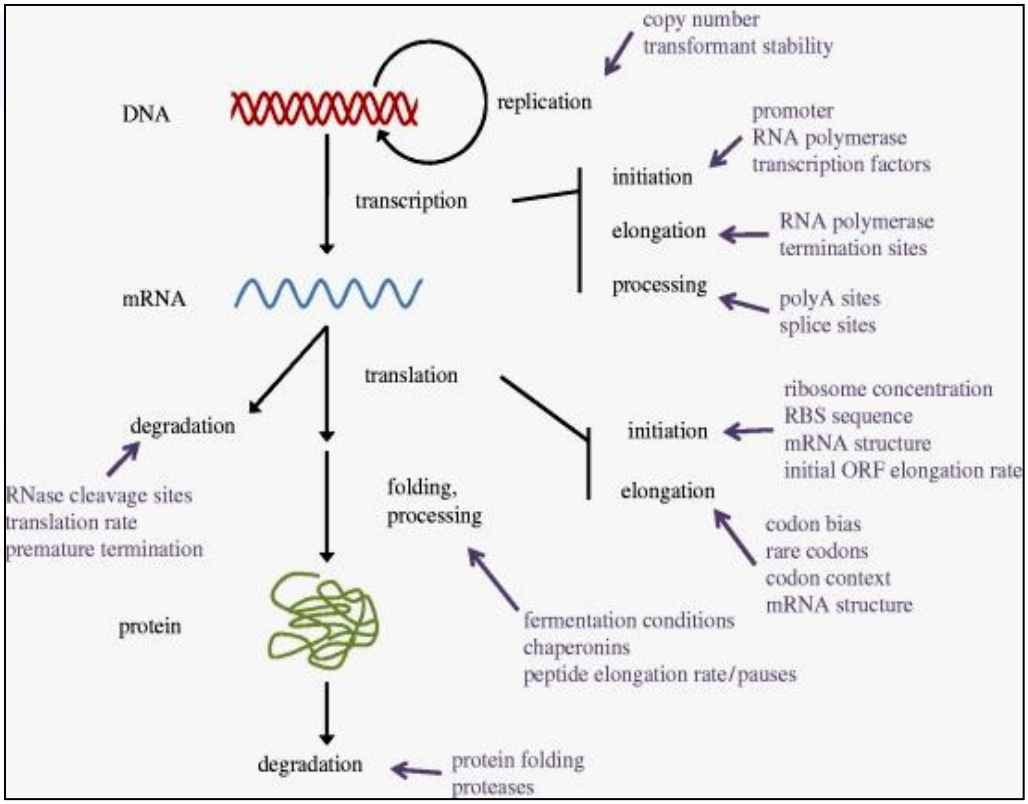



**Figure 34:** Factors influencing protein expression. Several factors that act along the path of expression from DNA to mRNA to protein are shown, any of which could be altered by or could affect the impact of gene design. **RBS**: ribosome-binding site (Welch *et al.*, 2009)

A reasonable number of published works exists in which significant changes in expression are found in genes that have been re-synthesized according to simple design rules (Gustafsson *et al.*, 2004; Wu *et al.*, 2007). This is encouraging because it suggests that the optimization problem can be reduced to a manageable number of variables to describe a gene sequence. What has not yet emerged is the identity of the most important sequence variables or their contributions to protein expression.Part of the problem is one of variable definition—we know categorically what is important but not exactly what, how and/or when. It is in line with this approach that our lab developed a 10-step gene optimization protocol that calls for the generation of three optimized sequences at minimum using different codon-tuning softwares' and various sequence analysis and *in silico* mRNA secondary structure prediction tools. This method allows the design of the best optimized gene version and cuts down the trial and error time of bio-drugs development phase.

## 4.2 Conclusions

Taking into account the results obtained in this study, the following have been concluded:

- *P. pastoris* recombinant clone that secretes soluble human Interferon β has been developed in ten steps.
- Gene optimization for IFN-β by different softwares' using different algorithms results in different sequences encoding the same protein
- Gene optimization has a positive effect on the expression of IFN-β in yeast *P. pastoris* system. Optimization of *IFN-β* sequence with JCat's software improves the expression of IFN-β protein. However, expression of GASCO optimized sequence was below the threshold of detection by western blot.

## 4.3 Recommendations

- Improve the gene optimization protocol for a better expression in the yeast *P.pastoris* for optimization of human IFN-β sequence.
- Using the humanized *P. pastoris* strain pGlycoSwitchM5 will improve the glycolysation and the folding of the protein.
- Adjustment of culturing conditions (pH , Temperature, and media components) in order to improve the protein expression in the yeast system.
- Upscale production of IFN-β and purification using His GraviTrap Gravity-flow Purification.
- Analyzing the biological activity of IFN-β to study the effectiveness of protein.

## References


Amicon. (1985). Operating *Instructions: Centricon™ Microconcentrators For Small Volume Concentration*. a GRACE company; USA.

Andersson, S. G., Kurland C. G. (1990). Codon preferences in free-living microorganisims . *Microbial Rev.*, **54** (2):198-210.

Andrews, K. L. and Husmann, D. A. (1997). Bladder dysfunction and management in multiple sclerosis. *Mayo Clin. Proc.*, **72** (12): 1176–83.

Applied Biosystems (2012). Retrieved on 10 April from <www.appliedbiosystems.com>

Arduini, R.M., Strauch, K.L., Runkel, L.A., Carlson, M.M., Hronowski, X., Foley, S.F., Young, C.N., Cheng, W., Hochman, P.S., Baker, D.P. (1999). Characterization of a soluble ternary complex formed between human interferon-β-1a and its receptor chains. *Protein Science* **8**: 1867-1877.

Arnold, L, Demain, Preeti Vaishnav (2009) .production of recombinant proteins by microbes and higher organisms. *Biotechnology advances* **27:** 297-306.

Ascherio, A, Munger, K.L., Simon, K.C. (2010). Vitamin D and multiple sclerosis. *Lancet Neurol* **9** (6): 599–612

Baev, D, Lil, X.W., Edgerton, M. (2001). Genetically engineered human salivary histatin genes are functional in Candida albicans: development of a new system for studying histatin candidacidal activity, MICROBIOL-SGM **147**, pp. 3323-3334.

Benabdesselem, C, Barbouche, M.R., Jarboui, M.A., Dellagi, K, Fathallah, D.M. (2007). High Level Expression of Recombinant *Mycobacterium Tuberculosis* Culture Filtrate Protein CFP32 in *Pichia Pastoris*. *Mol. Biotech*. **35**, 41-49.

Biron, C. (1998). Role of early cytokines, including alpha and beta interferons (IFN-β), in innate and adaptive immune responses to viral infections. *Seminars in Immunology* **10**: 383-390.

Bobholz, J. A. and Rao, S. M. (2003) Cognitive dysfunction in multiple sclerosis: a review of recent developments. *Curr. Opin. Neurol*. ***16** (3): 283–8.*

Bos, G, Verrijk, R, Franssen, O, Besemer, J, Hennink, W.E. & Crommelin, D.JA. (2001) Hydrogels for Controlled Release of Pharmaceutical Proteins. *Pharm Tech Europe*. **13**:64-74.

Brondyk, W.H. (2009). Selecting an appropriate method for expressing a recombinant protein. *Meth. Enzymol.* Methods in Enzymology **463**: 131–47.

Burton, J.M., O'Connor, P.W., Hohol, M, Beyene, J. (2009). Oral versus intravenous steroids for treatment of relapses in multiple sclerosis. *Cochrane Database Syst Rev* **(3)**: CD006921

Buttmann, M and Rieckmann, P. (2007). Interferon-beta1b in multiple sclerosis. *Expert Rev Neurother*.**7**:227–239 Review.

Carbone, A, Zinoveyv, A, Kepes, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics*. 1; **19** (16):2005-2015.

Cereghino, G.P.L., Sunga, A.J., Cereghino, J.L., Cregg, J.M. (2001). Expression of foreign genes in the yeast *Pichia pastoris*. *In Genetic Engineering: Principles and Methods* Setlow JK (ed.). Kluwer Academic/ Plenum: London; 157-169.

Cereghino, J.L. and Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, FEMS. *Microbiol. Reviews*. **24**:45-66.

Cereghino, J.L., Wong, W.W, Xiong, S., Giang, W., Luong, L.T., Vu, J., Johnson. S.D., and Cereghino, G.P. (2005). Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *BioTechniques*. **38**:44-48.

Chari, D.M. (2007). Remyelination in multiple sclerosis. *Int. Rev. Neurobiol.*. International Review of Neurobiology **79**: 589–620.

Clanet, M. (2008). Jean-Martin Charcot. 1825 to 1893 (PDF). *Int MS J* **15** (2): 59–61

Codon Usage Database. (2007) Retrieved 23 February 2012 from <http://www.kazusa.or.jp/codon/>

Comeron, J.M., Aguadé, M. (1998). An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol.* **47** (3): 268–74.

Comi, G. (2009). Treatment of multiple sclerosis: role of natalizumab. *Neurol. Sci.*. 30 **Suppl 2** (S2): S155–8.

Compston, A and Coles, A. (2002). Multiple sclerosis. *Lancet* **359** (9313): 1221–31

Compston, A and Coles, A. (2008). Multiple sclerosis. *Lancet* **372** (9648): 1502–17

Corley, R.B. (2005). *A Guide to Methods in the Biomedical Sciences.* Springer; USA.

Cregg, J. M., Vedvick, T. S., Raschke, W. C. (1993). Recent advances in the expression of foreign genes in Pichia pastoris. *Bio/Technology* **11**, 905-910.

Cregg, J.M., Cereghino, L., Shi, J., Higgins, D.R. (2000). Recombinant protein expression in *Pichia pastoris. Mol Biotech.* **16l** 23-52

Cregg, J.M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., Chappell, T. (2009). Expression in the yeast Pichia pastoris. *Meth. Enzymol.*. Methods in Enzymology **463**: 169–89

Dagert, M.; Ehrlich, S. (1979). "Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells". *Gene* **6** (1): 23–28

Daly R, Hearn MT (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition* **18**:119-138.

De Jager, P. L. and Hafler, D. A (2007). New therapeutic approaches for multiple sclerosis. *Annu Rev Med* **58**:417–432 Review.

De Seze, J., Zephir, H., Hautecoeur, P., Mackowiak, A., Cabaret, M. and Vermersch, P. (2006) Pathologic laughing and intractable hiccups can occur early in multiple sclerosis. *Neurology **67** (9): 1684–86.*

De Weerd, N.A., Samarajiwa, S.A., Hertzog, P.J. (2007). Type I interferon receptors: biochemistry and biological functions. *J Biol Chem* **282** (28): 20053–20057.

Debouverie, M., Pittion-Vouyovitch, S., Louis, S., Guillemin. F. (2008). "Natural history of multiple sclerosis in a population-based cohort". *Eur. J. Neurol.* **15**: 916.

Dr. Briffa. Why might shift-workers be at increased risk of multiple scelerosis? (2011). Retrieved on 23 Februrary 2012 from <http://www.drbriffa.com/2011/10/20/why-might-shift-workers-be-at-increased-risk-of-multiple-sclerosis/>

Dusheiko, G.M. (2003). *Interferon α: biology, pharmacology and Therapy for Chronic Viral Hepatitis.* In: Thomsonm A.W., Lotze, M.T. (Eds.). The Cytokine Handbook, Academic Press, Amsterdam, pp. 1233-1254.

E-Government of Kingdom of Bahrain, about Bahrain. Available at: http://www.bahrain.bh/pubportal/wps/portal/ Accessed April 2012.

EMBL-EBI. Emboss Needle, Pairwise Sequence Alignment (2012) Retrieved on 10 April 2012 from <http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html>

Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O., Salzberg, S.L. (200). Prediction of transcription terminators in bacterial genomes. *J Mol Biol.* **301** (1):27-33.

ExPASy. Translate (2011). Retrieved on 10 April 2012 from <http://web.expasy.org/translate/>

Fath, S, Bauer, A.P., Liss, M., Spriestersbach, A., Maertens, B., *et al.* (2011). Multiparameter RNA and Codon Optimization: A Standardized Tool to Assess and Enhance Autologous Mammalian Gene Expression. *PLoS ONE* **6**(3): e17596. doi:10.1371/journal.pone.0017596.

Filippini, G., Munari, L., Incorvaia, B., Polman, C., Dámico, R., Rice, G.P. (2003). Interferons in relapsing remitting multiple sclerosis: *a systematic review. Lancet*. **361**: 545–552.

Fox, J.M., Erill, I. (2010). Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Res.* **17** (3): 185–96.

Freeman. J.A. (2001). Improving mobility and functional independence in persons with multiple sclerosis. *J. Neurol.* **248** (4): 255–259.

Froger, A. and Hall, J.E. (2007). Transformation of Plasmid DNA into E. coli Using the Heat Shock Method, J Vis Exp, (6): 253.

GASCO - Institute of Genomics and Integrative Biology (2012). Retrieved on 10 April 2012 from <http://miracle.igib.res.in/gasco/>

GCUA: Graphical Codon Usage Analyzer. (2006). Retrieved on 10 April 2012 from <http://gcua.schoedl.de>

Gellissen, G. (2005). *Production of recombinant proteins: Novel mictobial and eukaryotic expression systems.* Wiley-vch; USA

Ghadirian, P., Jain, M., Ducic, S., Shatenstein, .B, Morisset, R. (1998). Nutritional factors in the aetiology of multiple sclerosis: a case-control study in Montreal, Canada. *Int J Epidemiol*. **27** (5): 845–52

Gouy, M., Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Research **10**:7055-7074.

Grantham, R.G. and Gouy, M. (1980a). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type, *Nucleic. Acids. Res*. **8**:1893-1912.

Gray, P.W. and Goeddel, D.V. (August 1982). "Structure of the human immune interferon gene". *Nature* **298** (5877): 859–63.

Gray, P.W., Goeddel, D.V. (1982). Structure of the human immune interferon gene. *Nature* **298** (5877): 859–63.

Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D.C., Jahn, D. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. **1**;33 (Web Server issue):W526-31.

Gustafsson, C., Govindarajan, S., Minshull, J. (2004). Codon bias and heteologous protein expression. *Trends in biotechnology* , **22**(7):346-353.

Hafler, D. A., Slavik, J.M., Anderson, D. E., O'Connor, K. C., De Jager, P., Baecher, C,. Allan. (2005). Multiple sclerosis. *Immunol Rev*;204:208–**31** Review

Haller, O., Kochs, G., Weber, F. (2007). Interferon, Mx, and viral countermeasures. *Cytokine Growth Factor Rev.* **18** (5–6): 425–33.

Hamilton, S.R., Davidson, R.C., Sethuraman, N., *et al.* (2006). Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313** (5792): 1441–3**.**

Han, Y. K., Koo, T. Y. and Lee, G. M. (2009), Enhanced interferon-β production by CHO cells through elevated osmolality and reduced culture temperature. *Biotechnology Progress*, **25**: 1440–1447.

Hanahan, D.; Jessee, J.; Bloom, F. R. (1991). "Plasmid transformation of Escherichia coli and other bacteria". *Methods in enzymology*, **204**: 63–113.

Harish, N.**,** Gupta, R., Agarwal, P., Scaria, V., Pillai, B. (2006). DyNAVacS: an integrative tool for optimized DNA vaccine design. *Nucleic Acids Research*, **1**;34 (Web Server issue):W264

Hashemilar, M., Ouskui, D.S., Farhoudi, M., Ayromlou, H., Asadollahi, A. (2011). Multiple sclerosis in East Azerbaijan, North West Iran. *Neurology Asia*, **16** (2): 127-131.

Henze. T. (2005). Managing specific symptoms in people with multiple sclerosis *(PDF). Int MS J* **12** *(2): 60–8.*

Higgins, S, Hames, B. D. (1999). *Protein Expression: A practical aprroch.* Oxford: Oxford University press.

Hirsch C, Jarosch E, Sommer T, Wolf DH (2004). Endoplasmic reticulum-associated protein degradation--one model fits all? *Biochimica and Biophysica Acta* **1695**:215-223.

Hutchinson Cancer Research Center. Yeast PCR Colony (2005). Retrieved on 10 April 2012 from <www.fhcrc.org/labs/hahn/methods/mol_bio_meth/pcr_yeast_colony.html>

Iglesias, A., Bauer, J., Litzenburger, T., Schubart, A., Linington, C. (2001). T- and B-cell responses to myelin oligodendrocyte glycoprotein in experimental autoimmune encephalomyelitis and multiple sclerosis. *Glia* **36** (2): 220–34

Ikemura, T. (1981). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol*. **151** (3): 389–409

Ikemura, T. (1985). Codon Usage and tRNA Content in Unicellular and Multicellular Organisms. *Molecular Biology and Evolution*, **2**(1)13–34

Inan M, Aryasomayajula D, Sinha J, Meagher MM (2006). Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. *Biotechnology and Bioengineering.* **93**:771-778.

*International Society For Interferon And Cytokine Research*, October 2005 Volume 12, No. 3.

Invitrogen (2010). *pPICZ A, B, and C; Pichia expression vectors for selection on Zeocin™ and purification of recombinant proteins*. Retrieved 23 February 2012 from <www.invitrogen.com>

Isaacs, A., Lindenmann, J. (1957). Virus interference. I. The interferon. *Proc. R. Soc. Lond., B, Biol. Sci.* **147** (927): 258–67.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., Boyer, H.W. (1977). Expression in Eschericia coli of a chemically synthesized gene for the hormone somatostatin. *Science* , **198**:1056-1063.

IVLine. Med in Small doses – multiple sclerosis (2012). Retrieved on 23 February 2012 from <http://www.ivline.info/2012/03/med-in-small-doses-multiple-sclerosis.html>

JCat. Java Codon Adaptation Tool (2008). Retrieved on 10 April 2012 from <http://www.jcat.de>

Johnson, K.P. (2007). Control of multiple sclerosis relapses with immunomodulating agents. *J. Neurol. Sci.* **256** (Suppl 1): S23–8.

Kansas State University. Standard PCR conditions (2011). Retrieved on 10 April 2012 from <http://www.k-state.edu/hermanlab/protocols/StandardPCRConditions.html>

Kaur, P. and Bennett, J. L (2007). Optic neuritis and the neuro-ophthalmology of multiple sclerosis. *Int. Rev. Neurobiol*. **79**: 633–63.

Kessler, T. M., Fowler, C. J., Panicker, J. N. (2009). Sexual dysfunction in multiple sclerosis, *Expert Rev Neurother.* (3):341-50Kobelt, G., Berg J., Lindgren P., Jonsson, B.

(2006). Costs and quality of life in multiple sclerosis in Europe: method of assessment and analysis. *Eur J Health Econ* **7** (Suppl 2):S5–S13.

Koch-Henriksen, N. and Hyllested, K. (1988). Epidemiology of multiple sclerosis: incidence and prevalence rates in Denmark 1948–64 based on the Danish Multiple Sclerosis Registry. *Acta Neurol Scand* **78**:369–380.

Kurtzke, J. F. (1983) Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* **33** (11): 1444–52.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** (5259): 680–685.

Leary, S.M. and Thompson, A.J. (2005). Primary progressive multiple sclerosis: current and future treatment options. *CNS drugs* **19** (5): 369–76.

Lederberg, J. and Lederberg, E.M. (1952) Replica plating and indirect selection of bacterial mutants. *J Bacteriol*. **63**: 399–406

Liszewski, K. (2010), New Tools Facilitate Protein Expression, *Genetic Engineering & Biotechnology News*, **30** (9): 1, 40–41.

Liu, Y.J. (2005). IPC: professional type 1 interferon-producing cells and plasmacytoid dendritic cell precursors. *Annu Rev Immunol* **23**: 275–306.

Lublin, F.D. (2005). History of modern multiple sclerosis therapy *J Neurol* **252** [Suppl 3].

Lublin, F.D., Reingold, S.C. (1996). Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* **46** (4): 907–11

Makrides, S.C. (1980) Strategies for achieving high-level expression of genes in *Escherichia coli, Microbiol. Rev*. **60** 512-538.

Martin, J. (1985). Page Expression of amplified human beta interferon genes using heavy metal induction in Chinese hamster ovary cells *Gene,***37** (1-3):139-44.

Martinelli Boneschi, F., Rovaris, M., Capra, R., Comi, G. (2005). Martinelli Boneschi, Filippo. ed. Mitoxantrone for multiple sclerosis. *Cochrane database of systematic reviews (Online)* (4): CD002127

Massachusetts Institute of Technology**.** Primer 3 Input (2009). Retrieved on 10 April 2012 from <www.frodo.wi.mit.edu/>

McCombe, P. A., Gordon, T. B., Jackson, M. W. (2009). Bladder dysfunction in multiple sclerosis, *Expert Rev Neurother.* (3):331-40.

McCormack, P. L. and Scott, L. J. (2004). Spotlight on Interferon-beta-1b in relapsing-remitting and secondary progressive multiple sclerosis. *BioDrugs.* **18**(5):343-47.

Memorial University. Faculty of Science (Biology) (2011) Retreived 23 February 2012 from <http://www.mun.ca/biology/>

Menzella, H.G. (2011). Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli***.** *Microbial Cell Factories,* **10**:15

Mergiran, T.C. (1980). *Worldbook Science Year.*

Merson, R. M. and Rolnick, M. I. (1998) . Speech-language pathology and dysphagia in multiple sclerosis. *Phys Med Rehabil Clin N Am* ***9*** *(3): 631–41.*

Miller, D., Barkhof, F., Montalban, X., Thompson, A., Filippi, M. (May 2005). Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis. *Lancet Neurol* **4** (5): 281–8.

Min, W., Pober, J.S., Johnson, D.R. (1998). Inteferon induction of TAP1: The phosphatase SHP-1 regulates crossover between the IFN-α/β and the IFN-γ signal-transduction pathways. Circulation Research 83: 815-823.

Multiple Sclerosis International Federation. Retrieved 23 February 2012 from <http://www.msif.org/en/ >

Multiple Sclerosis Resource Center. Disease modifying drugs. (2012). Retrieved 10 April 2012 from <http://www.msrc.co.uk/index.cfm/fuseaction/show/pageid/1695>

Munari, L., Lovati, R., Boiko, A. (2004). Munari, Luca M. ed. Therapy with glatiramer acetate for multiple sclerosis. *Cochrane database of systematic reviews (Online)* (1): CD004678.

Nagano, Y. and Kojima, Y. (1954). Pouvoir immunisant du virus vaccinal inactivé par des rayons ultraviolets (in French). *C. R. Seances Soc. Biol. Fil.* **148** (19–20): 1700–2.

Nagata, S., Mantei, N., Weissmann, C. (1980a). The structure of one of the eight or more distinct chromosomal genes for human interferon-alpha. *Nature* **287** (5781): 401–8

Nagata, S., Taira, H., Hall, A., *et al.* (1980b). Synthesis in *E. coli* of a polypeptide with human leukocyte interferon activity. *Nature* **284** (5754): 316–20.

NCBI (2012). NCBI Structure Group. Retrieved 23 February from <http://www.ncbi.nlm.nih.gov/Structure>

New England BioLabs. NEBcutter V2.0. (2012) Retrieved on 10 April 2012 frm <http://tools.neb.com/NEBcutter2/>

Noad. R. and P. Roy (2003). Virus-like particles as immunogens, Trends Microbiol. **11** 438–444.

Ozato. K., Uno, K., Iwakura, Y. (2007). Another road to interferon: Yasuichi Nagano's journey. *J. Interferon Cytokine Res.* **27** (5): 349–52.

Parekh RN, Wittrup KD (1997) Expression level tuning for optimal heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnology Progress*.**13**:117-122.

Peden, J. (2005). Codon usage indices. *Correspondence Analysis of Codon Usage*. SourceForge. http://codonw.sourceforge.net/Indices.html. Retrieved 2012-02-23.

Pestka, S. (2007). "The interferons: 50 years after their discovery, there is much more to learn". *J. Biol. Chem.* **282** (28): 20047–51.

Pittock, S.J. and Lucchinetti, C.F. (2007). The pathology of MS: new insights and potential clinical applications. *Neurologist*., **13** (2): 45–56.

Pittock, S.J. and Rodriguez, M. (2008). Benign multiple sclerosis: a distinct clinical entity with therapeutic implications. *Curr. Top. Microbiol. Immunol.*, **318**: 1–17

Plotkin, J.B. and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics,* **12**, 32-42.

Pollmann, W. and Feneberg W. *(2008).* Current management of pain associated with multiple sclerosis. *CNS Drugs **22** (4*): 291–324.

QIAGEN. Sample & Assay Technologies (2012) Retrieved on 10 April 2012 from <www.qiagen.com>

Reitinger, S., Boroviak, T. and Laschober, G.T. (2008). High yield recombinant expresson of the extremephile ezyme bee hyaluronidase in *Pichia Pastoris. Protein Expression And Purification.* **57**: 226-233.

Rodney, J.Y., Milo Gibaldi. (2003). *Biotechnology and Biopharmaceutical.* A John Wiley & Sons, Inc., New Jersey, pp. 161-188.

Romisch K (2005). Endoplasmic reticulum-associated degradation. *Annual Review of Cell and Developmental Biology.* **21**:435-456.

Rosati, G. (2001). The prevalence of multiple sclerosis in the world: an update. *Neurol. Sci.* **22** (2): 117–39.

Rovaris, M., Confavreux, C., Furlan, R., Kappos, L., Comi, G., Filippi, M. (2006). Secondary progressive multiple sclerosis: current knowledge and future challenges. *Lancet Neurol* **5** (4): 343–54.

Runkel, L., Meier, W., Repinsky, R.B, *et al*. (1998). Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta). Pharm Res., 15(**4**): 641-649.

Russell-Harde, D., Wagner, T.C., Perez, H.D., Croze, E. (1999). Formation of a uniquely stable type I interferon receptor complex by interferon β is dependent upon particular interactions between interferon β and its receptor and independent of tyrosine phosphorylation. *Biochemical and Biophysical Research Communications* **255**: 539-544.

Sambrook J, Fritsch E.F., Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sandhu, K.S., Pandey, S., Maiti, S. and Pillai, B. (2008). GASCO: Genetic Algorithm Simulation for Codon Optimization. *In Silico Biology,* **8**, 0016

Sandra, R., Mirandola, A., Dannie, E. M., Hallal, Alessandro, S., Farias, Elain, C., Oliveira, Carlos, Brandão, O., Heloisa, H., Ruocco, P., Benito, Damasceno, Leonilda, M. B. (2009). Santos Interferon-beta modifies the peripheral blood cell cytokine secretion in patients with multiple sclerosis. *International Immunopharmacology* **9:** 824–830.

Satya, R. V., Mukherji, M. and Ranga, U. (2003). A pattern matching algorithm for codon optimization and CpG motif-engineering in DNA expression Vectors. *In*: Proceedings of Computational System Bioinformatics (CSB'03), IEEE, pp. 294-305.

Science Gateway. Tools: Bacteria Transformation Efficiency Calculator. Retrieved on 10 April 2012 from <http://www.sciencegateway.org/tools/transform.html>.

Scorer, C.A., Buckholz, R. G., Clare, J.J., Romanos, M.A. (1993). The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast *Pichia pastoris*. *Gene* **136**:111-119.

Sharp, P.M., Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15** (3): 1281–95.

Shi, L., D. Wang, et al. (2007). Efficient expression and purification of human interferon alpha2b in the methylotrophic yeast *Pichia pastoris*. *Protein Expression and Purification* **54**(2): 220-226.

Sinclair, G., Choy, F.Y. (2002). Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. *Protein Expression and Purification* **26**:96-105.

Sládková, T., Kostolanský, F. (2006). The role of cytokines in the immune response to influenza A virus infection. *Acta Virol.* **50** (3): 151–62

Sommaruga, S., Lombardi, A., Salvade, A., Mazzuchelli, A., Corsi, F., Galeffi, P., Tortora, P. and Prosperi, D. (2011). Highly efficient production of anti-HER2 scFv

antibody variant for targeting breast cancer cells. Appl Microbial *Biotechnol*. **91**:613-612

Sorensen, P. And Mortensen, K.K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. J.Biotechnol. **115**:113-128.

Sospedra, M. and Martin, R. (2005). Immunology of multiple sclerosis. *Annu Rev Immunol* **23**: 683–747 Review.

Sreekrishna, K., Brankamp, R.G., Kropp, K.E., Blankenshio, D.T., Tsay, J.T., Smith, P.L., Wierschke, J.D., Subramaniam, A., Birkenberger, L.A. (1997). Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*, *Gene*. **190** 55-62.

Stadelmann, C., Brück, W. (2004). Lessons from the neuropathology of atypical forms of multiple sclerosis. *Neurol. Sci.* **25** (Suppl 4): S319–22

Suzuki, H., Brown, C.J., Forney, L.J., Top, E.M. (2008). Comparison of correspondence analysis methods for synonymous codon usage in bacteria. *DNA Res.* **15** (6): 357–65.

Szarewski, A. (2010). HPV vaccine: Cervarix, Expert. *Opin. Biol. Ther*, **10** 477–487

Taniguchi, T, Fujii-Kuriyama, Y, Muramatsu, M (1980). Molecular cloning of human interferon cDNA. *Proc. Natl. Acad. Sci. U.S.A.* **77** (7): 4003–6.

Todd, S. and Naylor, S.L (1992). New chromosomal mapping assignments for argininosuccinate synthetase pseudogene 1, interferon-beta 3 gene, and the diazepam binding inhibitor gene. Somat. Cell. Mol. Genet. **18**: 381-385.

Tremlett, H., Oger, J. (2004). Hepatic injury, liver monitoring and the beta-interferons for multiple sclerosis. *J. Neurol.* **251** (11): 1297–303.

Tschopp J.F, Brust P.F, Cregg J.M, Stillman C.A, Gingeras T.R (1987). Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. *Nucleic Acids Res.* **15**:3859–3876.

Tuder, R.M. and Yun, J.H. (2008). It takes two to tango: cigarette smoke partners with viruses to promote emphysema. *J Clin Invest*, **118** (8): 2689–2693.

Tyring, S.K. (1995). Interferons: Biochemistry and Mechanisms of Action. Am. J. *Obstet. Gynecol*. **172**, 1350-1353.

Uchijima, M., Yochida, A., Nagata, T., Koide, Y. (1998). Optimization of codon usage of plasmid DNA vaccine is required of the effective MHC class I- restricted T cell response against an intacellular bacterium, *J. Immun*. **166**:5594-5599.

University of Virgina, Fasta Sequence Comparison (2012) Retrieved on 10 April 2012 from < http://fasta.bioch.virginia.edu/>

Utsumi, Shojiro Yamazaki, Kazuo Hosoi,Hirohiko Shimizu, Ken Kawaguchi and Fuyuhiko Inagaki (1986). Conformations of Fibroblast and *E. coli*-Derived Recombinant Human Interferon-ßs as Studied by Nuclear Magnetic Resonance and Circular Dichroism *J. Biochem*, Vol. **99**, No. 5 1533-1535.

Vassileva, A., Dipiti, A., Chugh, Khanna, N. (2001). Expression of hepatitis B surface antigen in methylotrophic yeast *Pichia pastoris* using GAP promoter. *Journal of Biotechnology*, **88**:21–35.

Vilcek, J. (2003). Novel interferons. *Nat. Immunol.* **4** (1): 8–9.

Walsh, G. (2003). *Biopharmaceuticals: Biochemistry and Biotechnology, Second Edition*. John Wiley & Sons Ltd.

Weber K and Osborn, M. (1969). The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem.* **244** (16): 4406–4412.

Weinshenker, B.G. (1994). Natural history of multiple sclerosis. *Ann. Neurol.* **36** (Suppl): S6–11

Weissenbach, J., Chernajovsky, Y., Zeevi, M., *et al.* (1980). Two interferon mRNAs in human fibroblasts: in vitro translation and *Escherichia coli* cloning studies. *Proc. Natl. Acad. Sci. U.S.A.* **77** (12): 7152–6

Welch, M, Govindarajan, M, Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., Gustafsson, C. (2009). Design Parameters to Control Synthetic Gene Expression in *Escherichia coli*. *PLoS ONE* **4** (9): e7002.

Wu, G., Dress, L., Freeland, S.J.(2007). Optimal encoding rules for synthetic genes: the need for a community effort. *Mol. Syst. Biol.* **3**:134. Epub

Zhang, J., MarkovicPlese, S., Lacet, B., Raus, J., Weiner H.L., Hafler D.A. (1994). Increased frequency of interleukin 2-responsive T cells specific for myelin basic protein and proteolipid protein in peripheral blood and cerebrospinal fluid of patients with multiple sclerosis. *J Exp Med*. **179**:973–984.

Zolotukhin, S., Potter M., Hausworth, W.W., Guyand, J., Muzyczka, N.A. (1996). Humanized green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *J. Virol.* **70**:4646-4654.

## Appendices

### 1. Luria-Bertani broth

| | | |
|---|---|---|
| LB broth | BioLabs, 12106-05 | 25g |
| $dH_2O$ | | 1000 ml |

Autoclave, store at 4°C for use within 2 weeks.

### 2. Luria-Bertani Agar

| | | |
|---|---|---|
| LB agar | BioLabs, 12107-05 | 40g |
| $dH_2O$ | | 1000 ml |

Autoclave, store at 4°C for use within 2 weeks.

### 3. 100mM Calcium Chloride (CaCl2)

| | | |
|---|---|---|
| $CaCl_2$ | Fluka, 21114 | 1.47 g |
| $dH_2O$ | | 100 ml |

Filtered using 0.2µm filter and stored at 4°C.

### 4. 0.5 M Ethylenediaminetetraacetic acid (EDTA)

| | | |
|---|---|---|
| EDTA disodium | BioBasic, 18941316 | 93.05 g |
| $dH_2O$ | | 500 ml |

pH is adjusted to 8.0. stored at room temperature

**5. 50X Tris-acetate-EDTA (TAE) Buffer**

| | | |
|---|---|---|
| Tris base | Amresco, 2137B34 | 242 g |
| $dH_2O$ | | 750 ml |
| glacial acid acid | Sigma, 19967 | 57.1 ml |
| 0.5 M EDTA | | 100 ml |

pH is adjusted to 8.5 and the volume is then adjusted to final volume of 1 L.

**6. 1X Tris-acetate-EDTA (TAE) Buffer**

Dilution of the 50XTAE, 1:50 for final concentration of 1XTAE.

**7. 0.7% Agarose Gel**

| | |
|---|---|
| agarose (Promega, V3125) | 0.7 g |
| 1XTAE buffer | 100ml |
| Ethidium bromide | 0.5gr/ml |

The gel is casted immediately

**8. YPD media**

| | |
|---|---|
| YPD media(Clontech, 630409) | 50.0 g |
| water | 1 liter |

Autoclave, store at 4°C for use within 2 weeks.

**9. YPDA media**

| | |
|---|---|
| YPD agar media (Clontech, 630410) | 70.0 g |
| water | 1 liter |

Autoclave, store at 4°C for use within 2 weeks.

**10. 1 M Sorbitol**

| | |
|---|---|
| D-Sorbitol (Aldrich, 240850) | 182.17 g |
| water | 1 liter |

Filtered by 0.2µm filter and stored at 4°C

**11. 1 M Ethylene glycol**

| | |
|---|---|
| Ethylene glycol (Fluka, 03750) | 62.07 g |
| water | 1 liter |

**12. 1 M DMSO**

DMSO (SIGMA, D8418)

Store at room temperature

**13. 10% Ammonium peroxodisulphate (Ammonium perseulphate)**

| | |
|---|---|
| Ammonium persulphate (AnalaR, 100323W) | 1 g |
| water | 10 ml |

Aliqouted in 1.5 ml eppendorf tubes and stored at -20°C

**14. 10x YNB (13.4% Yeast nitrogen base with ammonium sulphate)**

| | |
|---|---|
| YNB (Fluka, 70161) | 34 g |
| Ammonium sulphate (Fluka, 09980) | 100 g |
| water | 1 L |

Filtered by 0.2µm filter and stored at 4°C

**15. 500x Biotin (0.02% biotin)**

| | |
|---|---|
| Biotin | 20 mg |
| water | 100 ml |

Filtered by 0.2µm filter and stored at 4°C, shelf life one year

**16. 1 M Di-potassium phosphate ($K_2HPO_4$)**

| | |
|---|---|
| $K_2HPO_4$ (Fluka, 60355) | 174.18 g |
| water | 1 liter |

Autoclave, store at room temperature for use within 1 year

**17. 1 M Di Hydrogen potassium phosphate ($KH_2PO_4$)**

| | |
|---|---|
| $KH_2PO_4$ (Fluka, 60220) | 136.08 g |
| water | 1 liter |

**18. 1 M Di Hydrogen potassium phosphate ($KH_2PO_4$)**

| | |
|---|---|
| 1 M $K_2HPO_4$ | 132 ml |
| 1 M $KH_2PO_4$ | 868 ml |

Adjust pH to 6.0 with KOH. Autoclave, store at 4° for use within 1 year

**19. 10x methanol (5%)**

| | |
|---|---|
| Methanol (Fluka, 65543) | 5.0 ml |
| water | 95. ml |

Filtered by 0.2µm filter and stored at 4°C, shelf life is 2 months

**20. BMMY (0.8% Glycerol)**

| | |
|---|---|
| 1% yeast Extract | 1.0 g |
| 2% Bactopeptone (BD, 211677) | 2.0 g |
| 100 mM potassium phosphate | 10 ml |
| 1.34% YNB | 10 ml |
| 0.00004% Biotin | 40 µl |
| 0.5% Methanol | 10 ml |
| 0.8% Glycerol | 800 µl |

Autoclave, store at 4°C for use within 2 weeks.

**21. BEDS**

| | |
|---|---|
| 10 mM Bicin | 0.163 g |
| Ehylene glycol | 3 ml |
| DMSO | 5 ml |
| 1 M sorbitol | 18.2 g |

pH adjusted to 8.3 with NaOH. Filtered by 0.2µm filter and stored at 4°C.



**22. 1 M Dithiothretol (DTT)**

| | |
|---|---|
| DTT (Fluka, 43817) | 154 g |
| water | 1 liter |

Filtered by 0.2µm filter, aliquoted and stored at -20°C

**23. DEPC Water**

| | |
|---|---|
| 0.1 % DEPC | 0.1 g |
| water | 100 ml |

The solution was mixed for about 1 hour before autoclaving.

**24. Separating gel buffer 4x**

| | |
|---|---|
| 1.5 M Tris (Amresco, 2137B34) | 18.2 g |
| water | 100 ml |

pH is adjusted to 8.8 with HCl.

**25. Stacking gel buffer 4x**

| | |
|---|---|
| 0.5 M Tris | 6.1 g |
| water | 100 ml |

pH is adjusted to 6.8 with HCl

**26. Sample buffer 5x**

| | |
|---|---|
| 10x SDS | 1.0 g |
| 10 mM DTT | 0.1 ml |
| 20% Glycerol (Fluka,49780) | 2.0 g |
| 0.2 M Tris-HCl pH 6.8 | 0.242 g |
| 0.05% Bromophenol blue (BPB) | 0.005 g |
| 6M Urea (Fluka, 51465) | 3.6 g |
| Water | 10 ml |

Aliqouted 500 µl and stored at -20ºC.

**27. Running buffer 10x**

| | |
|---|---|
| Tris (Amresco, 2137B34) | 30 g |
| Glycine (Fluka, 50052) | 144 g |
| SDS (Fluka, 71729) | 10 g |
| water | 1 liter |

**28. Acrylamide-bis solution**

| | |
|---|---|
| Acrylmide (Bio-rad, 161-0101) | 37.5 g |
| N,N'-methylene bis acrylmide (Bio-rad, 161-0200) | 1.0 g |
| water | 128 ml |

Filtered by 0.2µm filter and stored at 4ºC for less than 3 months.

**29. 15% SDS-PAGE separating gel ( Use immediately to case the gel)**

| | |
|---|---|
| 30% Acrylamide – bis-solution | 5 ml |
| Separating buffer 4x | 2.50 ml |
| Water | 2.46 |
| 10% SDS | 0.1 ml |
| Temed (Bio-rad, 161-0800) | 30 µl |
| 10% APS | 0.1 ml |



**30. 4% SDS-PAGE stacking gel**

| | |
|---|---|
| 30% Acrylamide – bis-solution | 0.668 ml |
| Stacking buffer 4x | 1.25 ml |
| Water | 3.01 |
| 10% SDS | 0.1 ml |
| Temed (Bio-rad, 161-0800) | 30 µl |
| 10% APS | 0.1 µl |

Use Immediately to cast the gel.

**31. Continuous Transfer Buffer**

| | |
|---|---|
| 50 mM Tris (Amresco, 2137B34) | 0.668 ml |
| 200 mM Glycine (Fluka, 50052) | 1.25 ml |
| Water | 1 L |

**32. Tris buffered Saline (10X TBS)**

| | |
|---|---|
| 1500 mM NaCl (Fluka, 71380) | 87.66 g |
| 50 mM Tris.HCl (Promega, H5123) | 78.8 g |
| Water | 1 L |

Adjust the pH to 7.4 and complete the volume to 1 L $H_2O$

**33. Tris buffered Saline (1X TBS)**

| | |
|---|---|
| 150 mM NaCl (Fluka, 71380) | 8.8 g |
| 50 mM Tris.HCl (Promega, H5123) | 7.9 g |
| Water | 1 L |

Adjust the pH to 7.4 and complete the volume to 1 L $H_2O$.

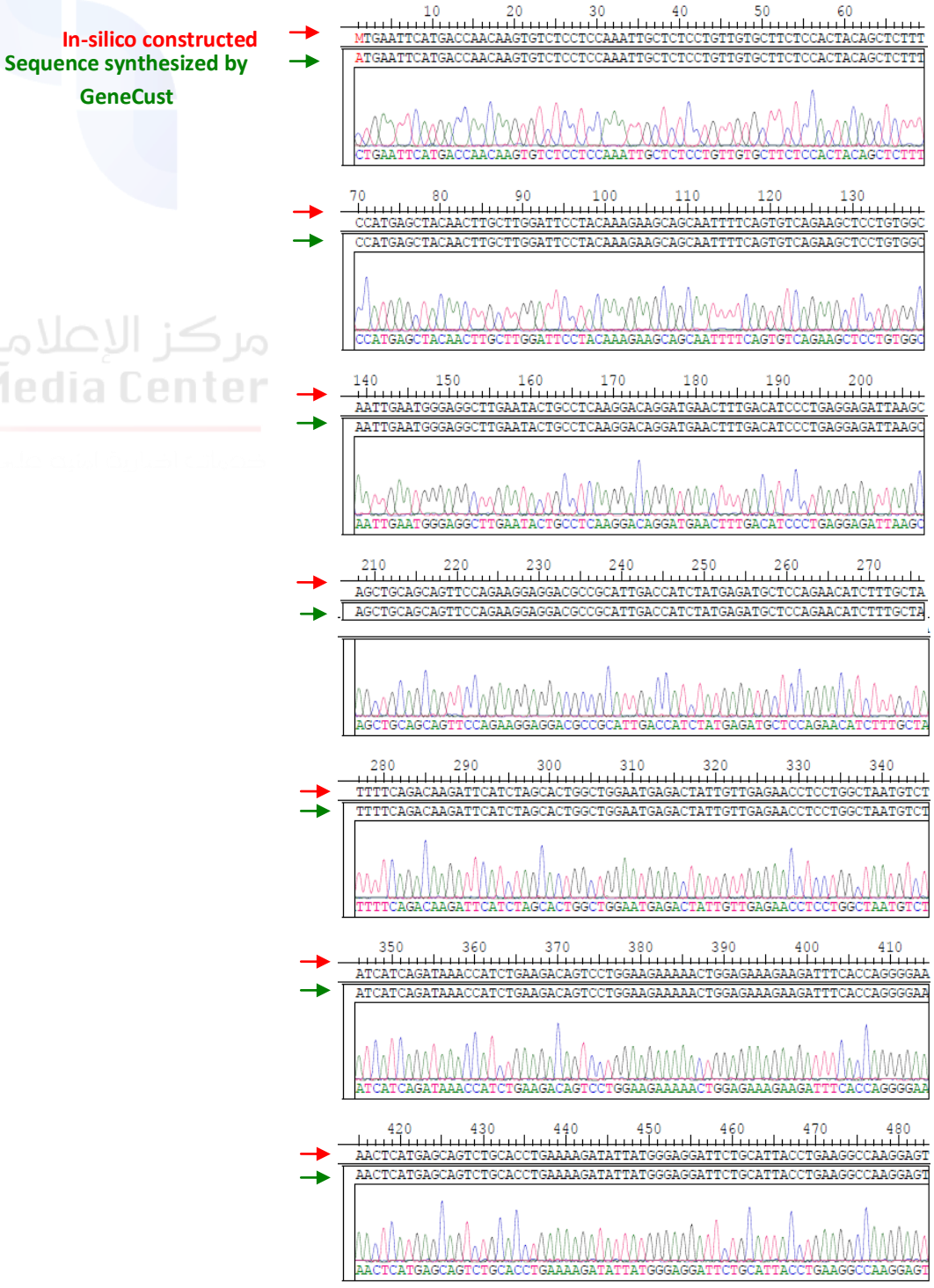
**34. Tris buffered Saline, Tween 20 (1X TBST)**

| | |
|---|---|
| 150 mM NaCl (Fluka, 71380) | 8.8 g |
| 50 mM Tris.HCl (Promega, H5123) | 7.9 g |
| 0.05% Tween 20 (USB, 20605) | 0.5 ml |
| Water | 1 L |

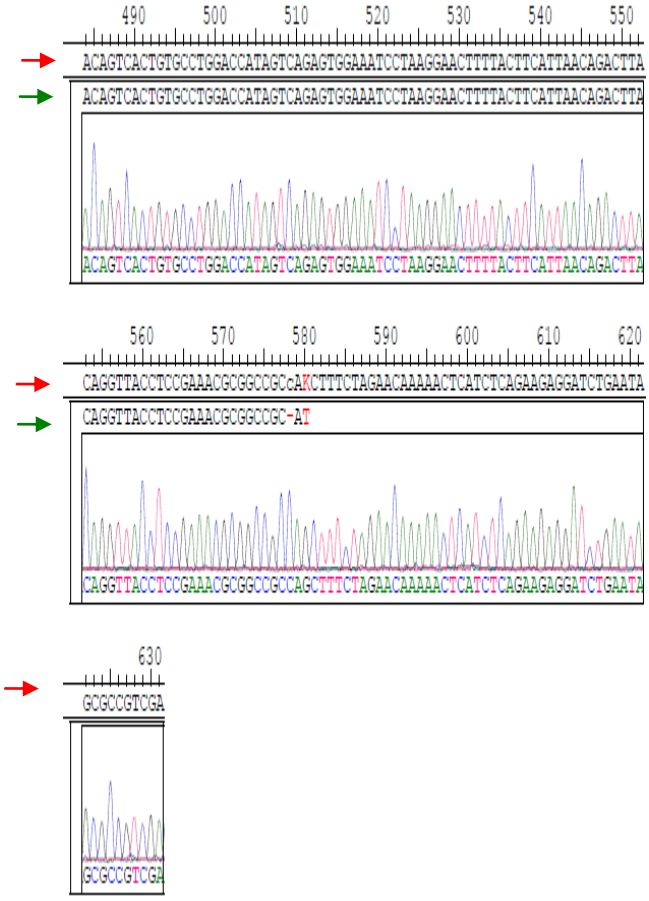
Adjust the pH to 7.4 and complete the volume to 1 L $H_2O$

# Annex

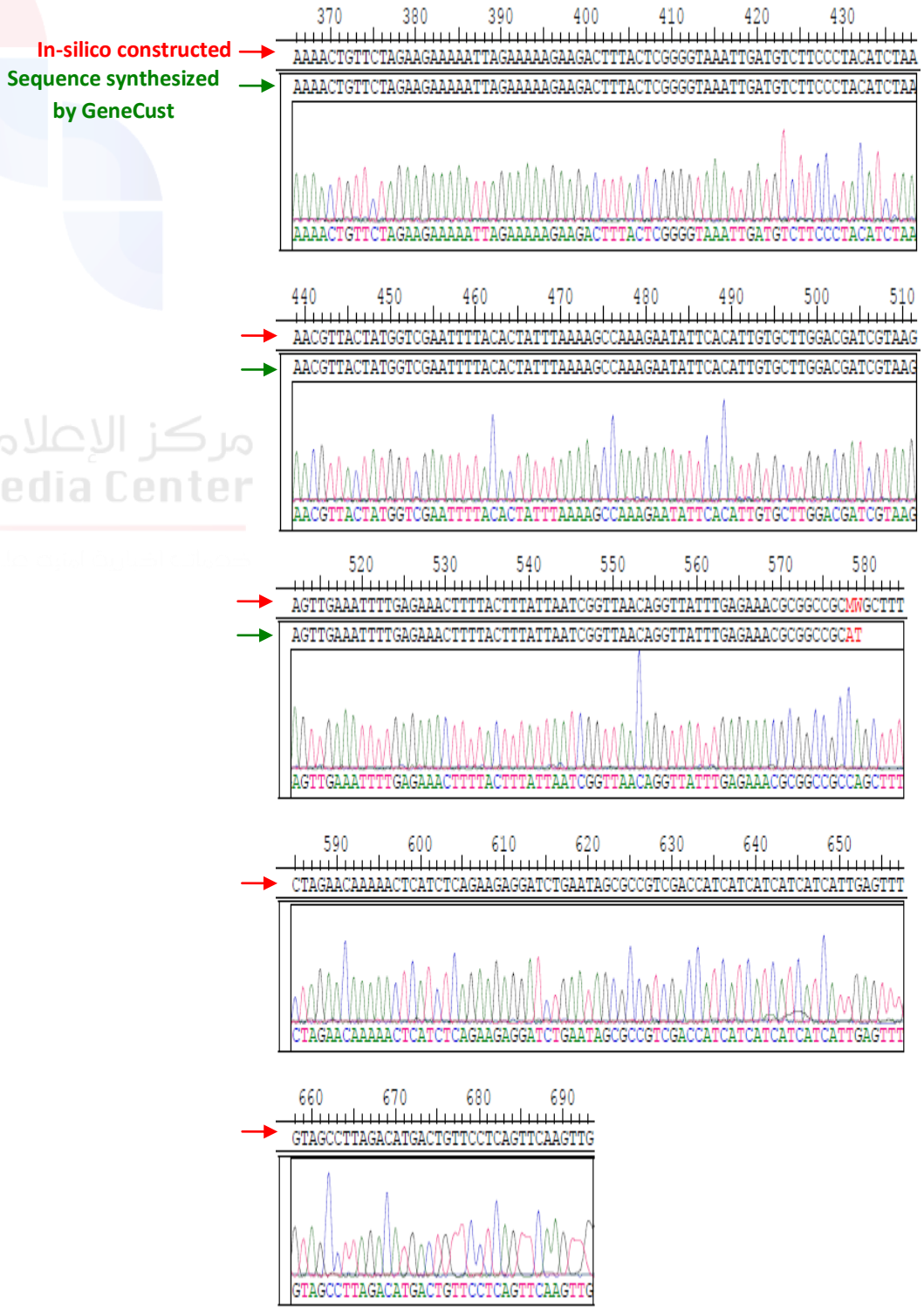
## 1. Sequencing results of wildtype (native) IFN-β from GeneCust



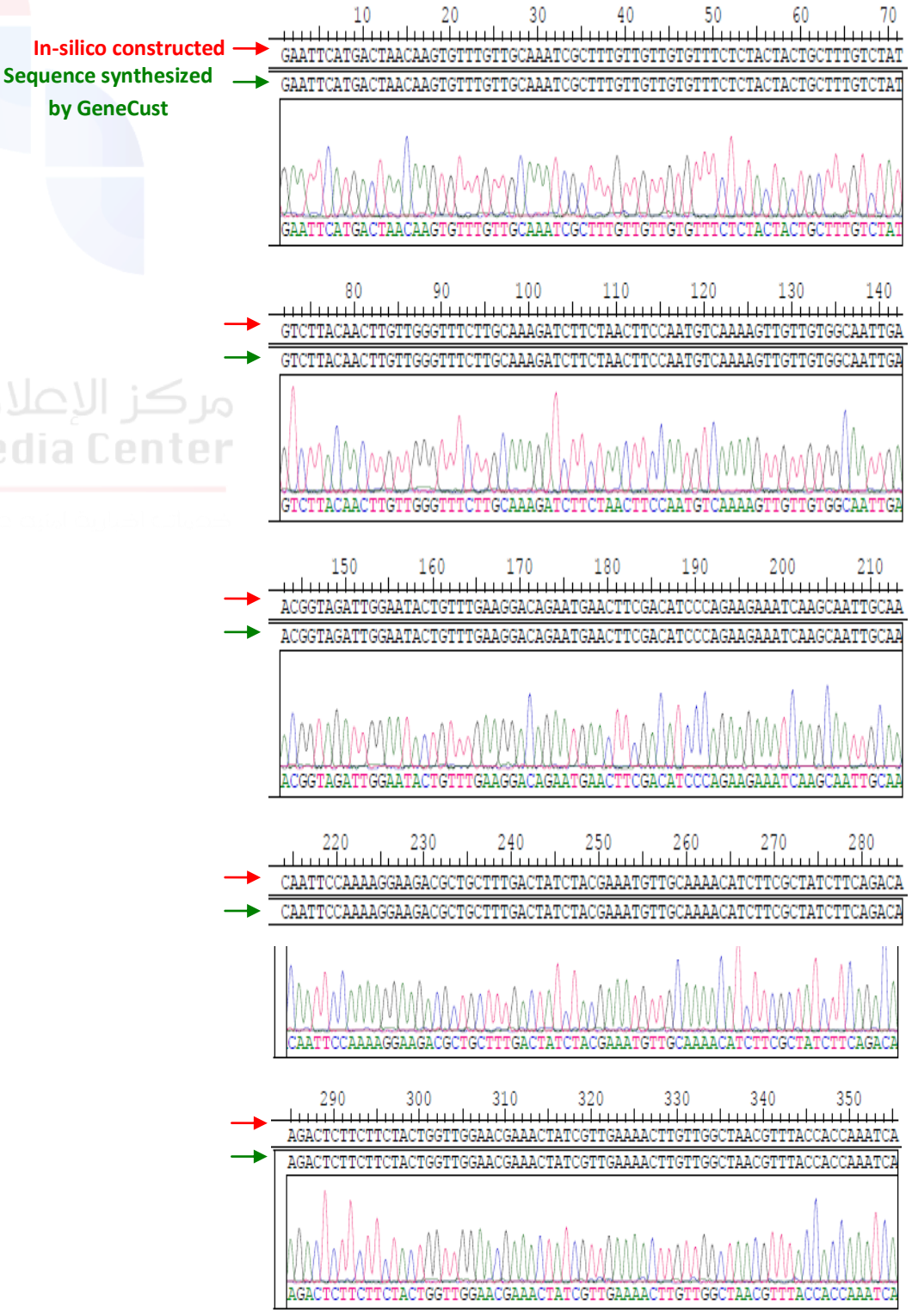
In-silico constructed Sequence synthesized by GeneCust



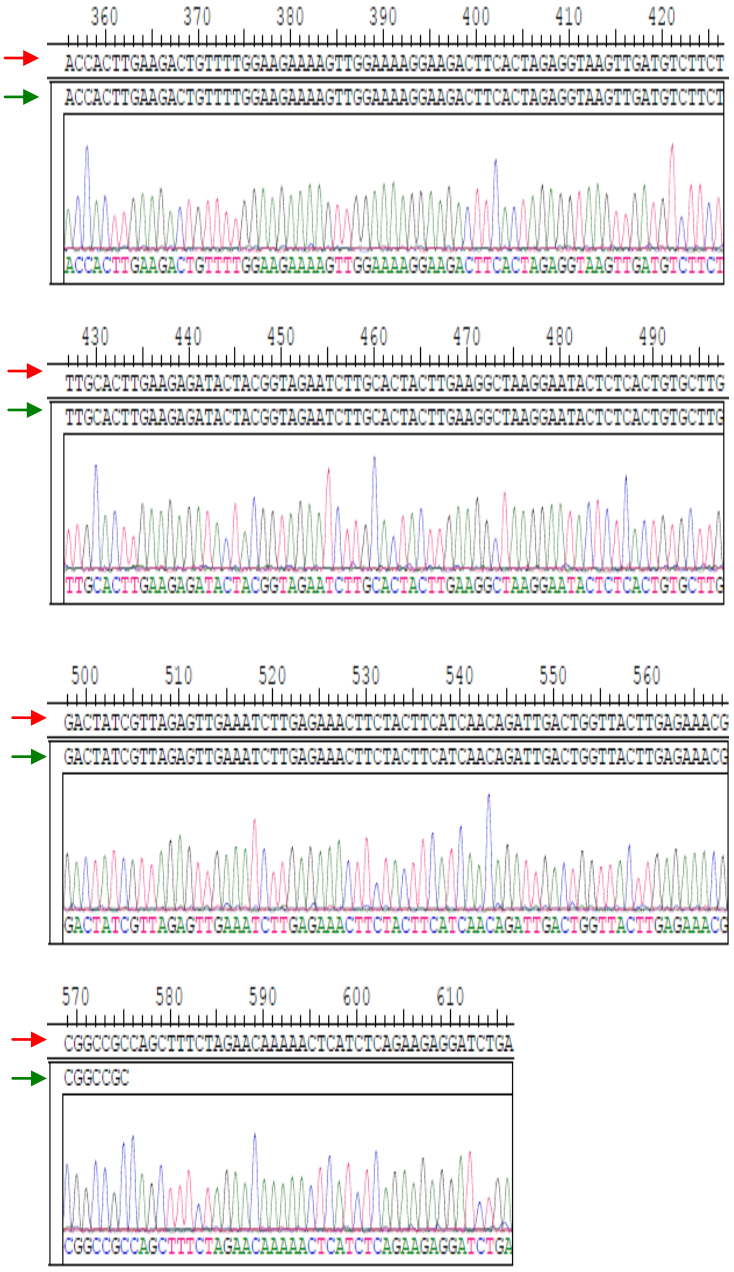
## 2. Sequencing results of GASCO optimized IFN-β from GeneCust



## 3. Sequencing results of JCat optimized IFN-β from GeneCust



In-silico constructed
Sequence synthesized by
GeneCust



## 4. Sequence alignment: Native IFNβ vs. GASCO-optimized IFNβ (Yeast)

```
EMBOSS_001         1 ATGACCAACAAGTGTCTCCTCCAAATTGCTCTCCTGTTGTGCTTCTCCAC     50
                     ||||||||.||.|||||..|.||.||.||||||||.||.||.||.||.||
EMBOSS_001         1 atgaccaataaatgtcttttacagatcgctctcctcttatgtttttcgac     50

EMBOSS_001        51 TACAGCTCTTTCCATGAGCTACAACTTGCTTGGATTCCTACAAAGAAGCA    100
                     .||.||.||.||.|||||.||.||.|||||.||.|||||||||.|....|
EMBOSS_001        51 caccgccctctcgatgagttataatttgctcggcttcctacaacgttcga    100

EMBOSS_001       101 GCAATTTTCAGTGTCAGAAGCTCCTGTGGCAATTGAATGGGAGGCTT-GA    149
                     |.|||||.||.||.||.||.|||||.|||||..|.||| |||.|.|| ||
EMBOSS_001       101 gtaatttccaatgccaaaaactcctctggcagctcaat-ggacgtttgga    149

EMBOSS_001       150 ATACTGCCTCAAGGACAGGATGAACTTTGACATCCCTGAGGAGATTAAGC    199
                     .|||||..|.||.||..|.||||||||.||.||||||||.||||||||.|
EMBOSS_001       150 gtactgtttaaaagatcgtatgaacttcgatatccctgaagagattaaac    199

EMBOSS_001       200 AGCTGCAGCAGTTCCAGAAGGAGGACGCCGCATTGACCATCTATGAGATG    249
                     |..|.||||||||||||||||.||.||.||..|.||||||||||.|||
EMBOSS_001       200 aattacagcagttccagaaggaagatgcagcgctcaccatctatgaaatg    249

EMBOSS_001       250 CTCCAGAACATCTTTGCTATTTTCAGACAAGATTCATCTAGCACTGGCTG    299
                     .|.||.|||||.|||||.||.|||.|.||.|||||.||....||.|||||
EMBOSS_001       250 ttacaaaacatttttgcaatcttccgtcaggattcctcgtctaccggctg    299

EMBOSS_001       300 GAATGAGACTATTGTTGAGAACCTCCTGGCTAATGTCTATCATCAGATAA    349
                     |||||||||.||.||||||||||||||.||.||.||.|||||.|||||.|
EMBOSS_001       300 gaatgagaccatcgttgagaacctccttgcgaacgtttatcaccagatta    349

EMBOSS_001       350 ACCATCTGAAGACAGTCCTGGAAGAAAAACTGGAGAAAGAAGATTTCACC    399
                     ||||.||.||.||.||.||.||||||||||.||.||||||||||.|||
EMBOSS_001       350 accacctaaaaaccgttcttgaagaaaaactcgaaaaagaagatttacc    399

EMBOSS_001       400 AGGGGAAAACTCATGAGCAGTCTGCACCTGAAAAGATATTATGGGAGGAT    449
                     .|.||.||..|.|||...|||.|.|||||.|||.|.|||||||..|.||
EMBOSS_001       400 cgtggtaagttaatgtctagtttacacctcaaacgttattatggtcgtat    449

EMBOSS_001       450 TCTGCATTACCTGAAGGCCAAGGAGTACAGTCACTGTGCCTGGACCATAG    499
                     ..|||||||..|.||.||.||.||.|||..|||.||.||||||||||.|
EMBOSS_001       450 attgcattatttaaaagcgaaagaatactctcattgcgcctggaccatcg    499

EMBOSS_001       500 TCAGAGTGGAAATCCTAAGGAACTTTTACTTCATTAACAGACTTACAGGT    549
                     |..|.||.|||||.||.|.||.|||||.||.||.||..|..|.||.|||
EMBOSS_001       500 tgcgtgtagaaatcttacgtaatttttatttttatcaatcggttgaccggt    549
EMBOSS_001       550 TACCTCCGAAAC    561
                     |||.|.||.|||
EMBOSS_001       550 tacttacgcaac    561
```

## 5. Sequence alignment: Native IFNβ vs. JCat-optimized IFNβ (Yeast)

```
EMBOSS_001         1 ATGACCAACAAGTGTCTCCTCCAAATTGCTCTCCTGTTGTGCTTCTCCAC     50
                     |||||.|||||||||.|..|.|||||.|||.|..|||||||.|||||.||
EMBOSS_001         1 ATGACTAACAAGTGTTTGTTGCAAATCGCTTTGTTGTTGTGTTTCTCTAC     50

EMBOSS_001        51 TACAGCTCTTTCCATGAGCTACAACTTGCTTGGATTCCTACAAAGAAGCA    100
                     |||.|||.|.||.|||...|||||||||.|.||.|||.|.|||||    |
EMBOSS_001        51 TACTGCTTTGTCTATGTCTTACAACTTGTTGGGTTTCTTGCAAAG----A     96

EMBOSS_001       101 GC----AATTTTCAGTGTCAGAAGCTCCTGTGGCAATTGAATGGGAGGCT    146
                     .|    ||.||.||.|||||.|||.|..||||||||||||.||.||..|
EMBOSS_001        97 TCTTCTAACTTCCAATGTCAAAAGTTGTTGTGGCAATTGAACGGTAGATT    146

EMBOSS_001       147 TGAATACTGCCTCAAGGACAGGATGAACTTTGACATCCCTGAGGAGATTA    196
                     .||||||||..|.||||||||.||||||||.||||||||.||.||.||.|
EMBOSS_001       147 GGAATACTGTTTGAAGGACAGAATGAACTTCGACATCCCAGAAGAAATCA    196

EMBOSS_001       197 AGCAGCTGCAGCAGTTCCAGAAGGAGGACGCCGCATTGACCATCTATGAG    246
                     ||||..||||.||.|||||.|||||.|||||.||.|||||.|||||.||.
EMBOSS_001       197 AGCAATTGCAACAATTCCAAAAGGAAGACGCTGCTTTGACTATCTACGAA    246

EMBOSS_001       247 ATGCTCCAGAACATCTTTGCTATTTTCAGACAAGATTCATCTAGCACTGG    296
                     |||.|.||.||||||||.|||||.||||||||||.||.|||...|||||
EMBOSS_001       247 ATGTTGCAAAACATCTTCGCTATCTTCAGACAAGACTCTTCTTCTACTGG    296

EMBOSS_001       297 CTGGAATGAGACTATTGTTGAGAACCTCCTGGCTAATGTCTATCATCAGA    346
                     .|||||.||.|||||.|||||.|||.|..|||||||.||.||.||.||.|
EMBOSS_001       297 TTGGAACGAAACTATCGTTGAAAACTTGTTGGCTAACGTTTACCACCAAA    346

EMBOSS_001       347 TAAACCATCTGAAGACAGTCCTGGAAGAAAAACTGGAGAAAGAAGATTTC    396
                     |.||||||.|||||||.||..||||||||||..||||.||.||||||.|||
EMBOSS_001       347 TCAACCACTTGAAGACTGTTTTGGAAGAAAAGTTGGAAAAGGAAGACTTC    396

EMBOSS_001       397 ACCAGGGGAAAACTCATGAGCAGTC-----TGCACCTGAAAAGATATTAT    441
                     ||.||.||.||..|.||     |||     |||||.||||.||||||.||.
EMBOSS_001       397 ACTAGAGGTAAGTTGAT-----GTCTTCTTTGCACTTGAAGAGATACTAC    441

EMBOSS_001       442 GGGAGGATTCTGCATTACCTGAAGGCCAAGGAGTACAGTCACTGTGCCTG    491
                     ||.||.||..||||.|||.|||||||.||||||.|||..|||||||||.||
EMBOSS_001       442 GGTAGAATCTTGCACTACTTGAAGGCTAAGGAATACTCTCACTGTGCTTG    491

EMBOSS_001       492 GACCATAGTCAGAGTGGAAATCCTAAGGAACTTTTACTTCATTAACAGAC    541
                     |||.||.||.|||||.||||||.|.||.|||||.||||||||.||||||.
EMBOSS_001       492 GACTATCGTTAGAGTTGAAATCTTGAGAAACTTCTACTTCATCAACAGAT    541

EMBOSS_001       542 TTACAGGTTACCTCCGAAAC    561
                     |.||.||||||.|..|||||
EMBOSS_001       542 TGACTGGTTACTTGAGAAAC    561
```

## 6. Sequence alignment: GASCO-optimized IFNβ vs. JCat-optimized IFNβ (Yeast)

```
EMBOSS_001         1 ATGACTAACAAGTGTTTGTTGCAAATCGCTTTGTTGTTGTGTTTCTCTAC     50
                     |||||.||.||.|||.|.||.||.||||||.|..|.||.|||||.||.||
EMBOSS_001         1 atgaccaataaatgtcttttacagatcgctctcctcttatgtttttcgac     50

EMBOSS_001        51 TACTGCTTTGTCTATGTCTTACAACTTGTTGGGTTTCTTGCAAAGATCTT    100
                     .||.||..|.||.|||..|||.||.|||.|.||.|||.|.|||.|.||..
EMBOSS_001        51 caccgccctctcgatgagttataatttgctcggcttcctacaacgttcga    100

EMBOSS_001       101 CTAACTTCCAATGTCAAAAGTTGTTGTGGCAATTGAACGGTAGATTGGAA    150
                     .|||.||||||||.||||||..|..|.||||||..|.||.||..|.|||||.
EMBOSS_001       101 gtaatttccaatgccaaaaactcctctggcagctcaatggacgtttggag    150

EMBOSS_001       151 TACTGTTTGAAGGACAGAATGAACTTCGACATCCCAGAAGAAATCAAGCA    200
                     |||||||.||.||..|.||||||||||.|||||.|||||.||.||.||
EMBOSS_001       151 tactgtttaaaagatcgtatgaacttcgatatccctgaagagattaaaca    200

EMBOSS_001       201 ATTGCAACAATTCCAAAAGGAAGACGCTGCTTTGACTATCTACGAAATGT    250
                     |||.||.||.|||||.||||||||.||.||..|.||.|||||.|||||||
EMBOSS_001       201 attacagcagttccagaaggaagatgcagcgctcaccatctatgaaatgt    250

EMBOSS_001       251 TGCAAAACATCTTCGCTATCTTCAGACAAGACTCTTCTTCTACTGGTTGG    300
                     |.||||||||.||.||.||||||.|.||.||.||.||.|||||.||.|||
EMBOSS_001       251 tacaaaacatttttgcaatcttccgtcaggattcctcgtctaccggctgg    300

EMBOSS_001       301 AACGAAACTATCGTTGAAAACTTGTTGGCTAACGTTTACCACCAAATCAA    350
                     ||.||.||.||||||||.|||.|..|.||.||||||||.|||||.||.||
EMBOSS_001       301 aatgagaccatcgttgagaacctccttgcgaacgtttatcaccagattaa    350

EMBOSS_001       351 CCACTTGAAGACTGTTTTGGAAGAAAAGTTGGAAAAGGAAGACTTCACTA    400
                     ||||.|.||.||.|||.|.||||||||..|.|||||.|||||.||.||..
EMBOSS_001       351 ccacctaaaaaccgttcttgaagaaaaactcgaaaaagaagattttaccc    400

EMBOSS_001       401 GAGGTAAGTTGATGTCTTCTTTGCACTTGAAGAGATACTACGGTAGAATC    450
                     |.||||||||.||||||..|||.|||.|.||..|.||.||.|||.|.||.
EMBOSS_001       401 gtggtaagttaatgtctagtttacacctcaaacgttattatggtcgtata    450

EMBOSS_001       451 TTGCACTACTTGAAGGCTAAGGAATACTCTCACTGTGCTTGGACTATCGT    500
                     |||||.||.||.||.||.||.||||||||||||.||.||.||||||.|||||
EMBOSS_001       451 ttgcattatttaaaagcgaaagaatactctcattgcgcctggaccatcgt    500

EMBOSS_001       501 TAGAGTTGAAATCTTGAGAAACTTCTACTTCATCAACAGATTGACTGGTT    550
                     ..|.||.||||||||..|.|||.||.||.||.||||||..|.||||||||
EMBOSS_001       501 gcgtgtagaaatcttacgtaatttttatttttatcaatcggttgaccggtt    550

EMBOSS_001       551 ACTTGAGAAAC    561
                     |||||..|.|||
EMBOSS_001       551 acttacgcaac    561
```