Generation of a bioactive recombinant protein toolbox of pancreatic-specific transcription factors

A Thesis

Submitted in Partial Fulfilment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Gloria Narayan

Under the supervision of

Dr. Rajkumar P. Thummer

and

Dr. Shirisha Nagotu

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

March, 2023

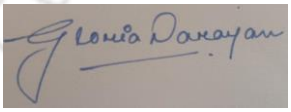
***Dedicated to* Banu**

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering

DECLARATION

I do hereby declare that the content embodied in this thesis entitled "**Generation of a bioactive recombinant protein toolbox of pancreatic-specific transcription factors**" is the result of investigations carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy, under the supervision of **Dr. Rajkumar P. Thummer** and **Dr. Shirisha Nagotu.**

As per the general norms of reporting research findings, due acknowledgments have been made wherever the research findings of other researchers have been cited in this thesis.



Date: 30th March 2023

Gloria Narayan
(166106028)

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering

CERTIFICATE

It is certified that the work described in this thesis entitled "**Generation of a bioactive recombinant protein toolbox of pancreatic-specific transcription factors**", by **Ms. Gloria Narayan** (Roll No. 166106028) for the award of the degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under our supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. This work has not been submitted elsewhere for the award of any degree or diploma.

Dr. Rajkumar P. Thummer
(Thesis Supervisor)

Dr. Shirisha Nagotu
(Thesis Co-supervisor)

Date: 30th March 2023

Acknowledgment

During the course of this research, many aspects of life changed. That said, this thesis would not have seen the light of day without the sincere help and assistance that it has received throughout its journey.

I am privileged to be indebted to the head of the Laboratory for Stem Cell Engineering and Regenerative Medicine (SCERM) and my thesis supervisor **Dr. Rajkumar P. Thummer**, for his constant guidance, support, and encouragement during every step of my work. It was a great privilege to work and study under his guidance. I would also like to thank him for the empathy that he has shown during my tough times. I would also like to express my special thanks of gratitude to my thesis co-supervisor **Dr. Shirisha Nagotu**, for providing me with invaluable guidance throughout the research.

I would also extend my gratitude to the members of my Doctoral committee, Prof. Sachin Kumar and Prof. Manish Kumar of the Department of Biosciences and Bioengineering and Dr. Sunanda Chatterjee of the Department of Chemistry for their valuable suggestions, constant support, and advice which enabled me to improve my work.

I am grateful to the current Head of the Department of Biosciences and Bioengineering, IIT Guwahati, Prof. Rakhi Chaturvedi and other former HODs – Prof. Latha Rangan and Prof. Kannan Pakshirajan for providing me with the departmental facilities for carrying out my research work. I would thank all the technical staff of the department for their help and assistance. I am also grateful to Prof. Sidharth S Ghosh, IIT Guwahati, for providing high-end instrument facilities for my research. I would also like to thank the Ministry of Education, Government of India, for the financial support during my tenure.

My completion of this thesis would not have been accomplished without the assistance and motivation of my hardworking and talented colleagues Dr. H. Krishna Kumar, Dr. Chandrima Dey, Khyati Raina, Dr. Madhuri Thool and Pradeep Kumar S. I also thank OBCAL members and current SCERM members, Atreyee, Naveen and Ronima for maintaining excellent lab environment that has hugely helped me to carry out the experiments. Here, I would like to specially mention Riddhi Bannerjee and Neha Joshifor their assistance with microscopy experiments. My sincere thanks also go to Plaboni Sen and Akriti Agrawal for offering me assistance with experiments.

I owe much more than I could ever express to my parents, Late Mr. Kumar Mukut Narayan and Mrs. Pooravee Narayan, whose support and care have constantly nurtured me in my research journey, reminding me that I must successfully achieve what I had set out to pursue. To say the least, they have shaped the person I am today. I am also blessed to have Reshma Gupta and Kumar Avik Narayan, my loving cousins, for their support, love and care. I am grateful to my hometown, Cooch Behar, where I was educated before starting college. In the schools of this beautiful town, I first realized my penchant for science. To this end, I acknowledge my teachers.

I am happily indebted also to the company of my friends Satakshi Hazra, Anjali Patel, Sinchini Barman, Mongoli and Anita who has been incredibly valuable both as researchers and often as the wind to my sails. It would have been nearly impossible for me to work diligently had it not been for the light-hearted moments I had spent in their company, learning and rejoicing. I am grateful to Triti Nandi and Suchetana Ghose for their constant motivation and inspiration. I am also thankful to Dr. Gaurav Talukder for constantly standing by me through my thick and thin.

Lastly, my warm thanks to my pets (Tojo, Millie, Tuku and Kitto). It had been heart-warming to return to my room after work knowing that the little feline creatures would be patiently awaiting my presence.

Gloria Narayan

Contents

List of Figures

List of Tables

Abbreviations

BeStSel	Beta Structure Selection
DAPI	4′ ,6-diamidino-2-phenylindole
DM	Diabetes mellitus
DMEM	Dulbecco's Modified Eagle medium
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
ER	Endoplasmic reticulum
ESCs	Embryonic stem cells
Far-UV CD	Far ultraviolet Circular Dichroism spectroscopy
FBS	Fetal bovine serum
FoxA2/O1	Forkhead box protein A2/O1
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GCUAT	Graphical Codon Usage Analyzer Tool
GLP1	Glucagon like peptide 1
GLUT4	Glucose transporter 4
GOI	Genes of interest
GRCAT	Genscript Rare Codon Analyzer Tool
GSIS	Glucose-stimulated insulin secretion
Hnf-1β/6	Hepatocyte nuclear factor 1 beta/6
IFNγ	Interferon-gamma
IL-1β	Interleukin-1 beta
IMAC	Immobilized metal ion affinity chromatography
iPSCs	induced pluripotent stem cells
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Insulin receptor
Isl1	Insulin gene enhancer-binding protein islet-1
LB	Luria-Bertini broth
MAFA/B	Musculoaponeurotic fibrosarcoma oncogene family A/B
MAPK	Mitogen-activated protein kinase
mM	Milli molar
MODY4	Maturity onset diabetes of the young

MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NaCl	Sodium chloride
NCBI	National Center for Biotechnology Information
NeuroD1	Neurogenic differentiation 1 isoform 1
NGN3	Neurogenin 3
Ni-NTA	Nickel-nitriloacetic acid
Nkx2.2/6.1	Nirenberg and Kim homeobox factor 2.2/6.1
NLS	Nuclear localization signal
NOD	Nonobese diabetic
OCT4	Octamer binding transcription factor 4
PAX4/6	Paired box 4/6
PB	Phosphate buffer
PBS	Phosphate buffer saline
PC1	Prohormone convertase 1
PC2	Prohormone convertase 2
PC3	Prohormone convertase 3
PDX1	Pancreatic and duodenum homeobox 1
PI	Propidium iodide
P/S	PenStrap
Ptf1a	Pancreas transcription factor 1 subunit alpha
qRT-PCR	Quantitative real-time polymerase chain reaction
RT	Room temperature
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel electrophoresis
SOX2/9	Sex-determining region Y (SRY)-Box Transcription factor 2/9
TIDM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TAT	Transactivator of transcription
TBS	Tris-buffer saline
TNFα	Tumor necrosis factor-alpha

"Generation of a bioactive recombinant protein toolbox of pancreatic-specific transcription factors"

ABSTRACT

The persistent shortage of insulin-producing islet mass or β-cells for transplantation in the ever-growing diabetic population worldwide is a matter of concern. To date, a permanent cure for this medical complication is not available and soon after the establishment of lineage-specific reprogramming, direct β-cell reprogramming became a viable alternative for β-cell regeneration. Direct reprogramming is a straightforward and powerful technique that can provide an unlimited supply of cells by transdifferentiating terminally differentiated cells toward the desired cell type. This approach has been extensively used by multiple groups to reprogram non-β-cells toward insulin-producing β-cells. The β-cell identity has been achieved by various studies via ectopic expression of one or more pancreatic-specific transcription factors in somatic cells, bypassing the pluripotent state. In our present study, we worked on the four most transcription factors, which are critical in the field of pancreatogenesis, and subsequently, in the generation, maturation and maintenance of β-cells. These factors are Pancreatic and duodenum homeobox 1 (PDX1), which is the "master regulator" essential for the proper development of the pancreas, duodenum and antrum. Furthermore, it is an indispensable reprogramming factor for the derivation of human β-cells, and recently, it has been identified as a tumor suppressor protein in gastric cancer. Next is Neurogenin3 (NGN3), which is vital for the development of endocrine cells of the intestine and pancreas. NGN3 is also critical for neural precursor cell determination in the neuroectoderm. Additionally, it is one of the vital transcription factors for deriving human β-cells from specialized somatic cells. Subsequently, we discussed Musculoaponeurotic fibrosarcoma oncogene family A (MAFA), which is a mature β-cell marker and is one of the widely studied transcription factors in the β-cell paradigm. It is also one of the core transcription factors in cell reprogramming cocktails to generate β-cells, thus offering vast potential for cell therapy application for the treatment of diabetes mellitus. Lastly, Paired box 4 (PAX4) is a pivotal transcription factor involved in pancreatogenesis during embryogenesis, and in adults, it plays a crucial role in β-cell proliferation and survival. Additionally, the function of PAX4 as a tumor suppressor protein in human melanomas is also reported. In the current study, the production and purification of the human PDX1, NGN3, MAFA and PAX4 proteins from *Escherichia coli* (*E. coli*) are reported. First, the protein-coding nucleotide sequence of the genes was codon-optimized to enable enhanced protein expression in *E. coli* strain BL21(DE3). The codon-optimized sequences were fused in-frame to three different fusion tags to

enable cell penetration, nuclear translocation, and affinity purification. The gene inserts with the fusion tags was subsequently cloned into an expression vector (pET28a(+)) for heterologous expression in BL21(DE3) cells. A suitable genetic construct and ideal expression conditions were subsequently identified, producing a soluble form of the recombinant fusion proteins. These fusion proteins were then purified to homogeneity (purity>90%) under native conditions, and their secondary structure was retained post-purification. When applied to human cells, this purified protein did not induce cytotoxicity. Further, the cellular uptake and nuclear translocation of these fusion proteins were demonstrated in PANC-1 cells. The bioactivity of these fusion proteins was investigated using RT-qPCR, where these proteins (except PAX4) have been shown to upregulate insulin gene expression. Moreover, various other assays (cell migration, cell proliferation and cell cycle assays) were also performed to demonstrate these proteins also function as tumor suppressors in different cancer cell lines. Prospectively, this recombinant protein toolbox can be utilized for various biological applications to investigate its functionality in cell reprogramming, biological processes, and diseases.

Chapter 1

Introduction and Review of Literature

A brief overview of the chapter

Chapter 1 deals with the overview on diabetes mellitus, its various types, and causes, and provides an estimate of its prevalence worldwide. The chapter also draws the attention to the multiple approaches available for treating diabetes mellitus and the shortcomings associated with each approach. This section further focuses on the current status of direct reprogramming of various terminally differentiated cell types to insulin-producing β-cells and how this approach can benefit the diabetic community. Upon reviewing the available literature, we identified four transcription factors, PDX1, NGN3, PAX4 and MAFA, that have proved to be crucial in the field of β-cell reprogramming. This chapter also elucidates various functions of these selected factors in pancreatic differentiation, β-cell formation and maintenance. This section also focuses on the wide variety of somatic cells reprogrammed into β-cells using these factors in the presence of essential growth factors under defined culture conditions. Finally, the chapter concludes by describing the motivation and rationale behind the work carried out in this thesis toward addressing roadblocks associated with recombinant protein production and conventional direct β-cell reprogramming approaches.

1.1 Introduction

Diabetes mellitus is a chronic metabolic disorder characterized by high blood glucose levels that pose serious damage to the blood vessels, nerves, eyes, heart and kidneys over time. It is also one of the leading causes of death globally, accounting for 536.6 million lives affected worldwide in 2021 (Sun et al. 2022). The 10th edition of the International Diabetes Federation Atlas further projected this number to increase to 783.2 million lives in 2045 if no preventive measures are taken (Sun et al. 2022). The report further highlighted aging of the population as the principal reason contributing to this rising number. Although multiple drugs are available to manage diabetes mellitus effectively, this only provides a temporary solution.

The general approach to replenish β-cells in a diabetic patient is to transplant β-like cells derived via stem cell differentiation or healthy β-cells obtained from cadaveric/brain-dead donors (Merani and Shapiro 2006). Although islet transplantation also emerged as a promising alternative for replacing the lost β-cells, ethical issues, immune rejection, the transmission of the virus from porcine to human (in the case of porcine islet transplantation) and scarcity of cadaveric donors made it unsuccessful (Merani and Shapiro 2006; Pellegrini et al. 2016). When immunosuppressive drugs were used to suppress the immune rejection during transplantation, severe opportunistic infections became another barrier in this approach (Rother and Harlan 2004). For treating Type 1 Diabetes mellitus and Type 2 Diabetes mellitus (later stage), insulin therapy is the only life-saving practical option for this disease to date. The high risk associated with insulin therapy is that it may precipitate acute comatose hypoglycemia if there is no proper control over its dosage and time of treatment. Another drawback of insulin is gastric acid lability; therefore, it must be administered subcutaneously or intramuscularly (Setty et al. 2016). Furthermore, insulin therapy requires multiple insulin shots (after every carbohydrate-rich diet) in a day, thus making it a cumbersome, painful and impractical long-term solution.

Moreover, weight gain during long-term insulin treatment significantly increases cardiovascular risks (Cichosz et al. 2016). These above complications necessitate exploring novel strategies that can cure diabetes mellitus permanently.

Direct reprogramming of various differentiated somatic cells to another cell type bypassing the stem cell state, has proved to be a promising approach in regenerative medicine (Ieda et al. 2010; Colarusso and Zhou 2022). This approach saves us from the arduous and lengthy procedure taken in the case of stem cell-based therapy. Direct reprogramming involves ectopic expression of key transcription factors in terminally differentiated cells in the presence of essential growth factors under defined culture conditions to generate a desired cell type. The thesis aims to establish a recombinant protein toolbox consisting of pancreatic-specific reprogramming transcription factors to address this. Thus, this toolbox will facilitate the safer generation of insulin-producing β-cells with immense clinical applicability.

The field of recombinant protein production has attained widespread popularity due to its enormous applications in therapeutics, diagnostics or understanding basic biological questions and is considered to be a safe and promising alternative compared to viral and gene delivery-based approaches (Sanchez-Garcia et al. 2016; Seo et al. 2017; Serna et al. 2018; Kumar et al. 2019). For these applications, the protein of interest is delivered into a target cell to exert its biological function. Notably, this approach does not manipulate the target cell genome, making it ideal for clinical applications or deriving integration-free, clinical-grade human cells (O'Malley et al. 2009; Sommer and Mostoslavsky 2013; Seo et al. 2017; Kumar et al. 2019). Additionally, this approach provides complete control over the dosage and time of application of a protein of interest to elucidate its specific role in cellular reprogramming (Kumar et al. 2019; Dey et al. 2021a).

Large-scale production of recombinant proteins employs a wide range of expression systems. *Escherichia coli* (*E. coli*) isan obvious choice for large-scale production due to its well-established genetic makeup and knowledge of its physiology, reasonable costs and unparalleled growth rate (Sahdev et al. 2008; Khow and Suntrarachun 2012). However, important criteria must be addressed to maximize protein expression and yield of the protein of interest.

Codon optimization is a critical step to be carried out in the heterologous expression of recombinant proteins, which involves the alteration of the nucleotide sequence of the DNA of one species into the nucleotide sequence of the DNA of another species without changing the amino acid sequence of the protein. Codon optimization facilitates the heterologous production of recombinant proteins, providing several advantages. It allows removal of restriction sites from the gene of interest, enhances mRNA stability or prevents secondary structures, and allows modification in ribosome binding and mRNA degradation sites, etc.

In this study, we have generated a biologically active recombinant toolbox comprising multiple pancreatic-specific transcription factors like Pancreatic and Duodenum transcription factor 1 (PDX1), Neurogenin3 (NGN3), Musculoaponeurotic Fibrosarcoma Oncogene family A (MAFA) and Paired box gene 4 (PAX4) proteins. These proteins can be prospectively used for insulin-producing β-cell reprogramming and various other biological applications.

In this thesis, cloning, expression and purification of the codon-optimized sequences of the above-mentioned pancreatic-specific human recombinant proteins from *E. coli* is reported. We also determined their secondary structure and demonstrated their transduction ability into mammalian cells. To the best of our knowledge, this is the first study to report screening and identification of optimal expression parameters in *E. coli* to express and purify these human proteins to homogeneity under native conditions and give an insight into their specific

secondary structure content. Further, we demonstrated their biological activity using different assays in mammalian cells.

1.2 Literature Review

1.2.1 Pancreas and pancreatic cells

Pancreas is an elongated, flattened abdominal organ at the posterior of the stomach surrounded by the liver, spleen, and small intestine. Knowledge of early events in human pancreas organogenesis, such as the foregut endoderm patterning and pancreatic bud formation, is exceptionally limited due to the scarcity of tissue specimens and ethical concerns. However, studies have shown that human pancreas development initiates at gestational day 26. In the next few days, a dorsal bud is formed, followed by the appearance of two ventral buds (left and right) by gestation day 30 (Pan and Brissova 2014). The left ventral bud gets regressed in the upcoming weeks, and the right ventral bud gets fused to the dorsal pancreatic bud during gestational period 6-7 weeks (Adda et al. 1984; Polak et al. 2000; Piper et al. 2004). The dorsal pancreas gives rise to the majority of the pancreas, including the upper part of the head (caput pancreatitis), the isthmus (isthmus pancreatitis), the body (corpus pancreatitis) and the tail of the pancreas (cauda pancreatitis) whereas the right ventral bud gives rise to the inferior part of the head (Adda et al. 1984; Polak et al. 2000).

The pancreas comprises two types of glands, exocrine and endocrine glands. It is attached to the first section of the small intestine (duodenum) through a duct that helps to secrete the exocrine hormones to the digestive system. The exocrine gland comprises 96-99% of the total pancreas volume and secretes digestive enzymes. Microscopic anatomy of the exocrine pancreas reveals that it comprises small compartments called lobules. Each of these lobules is made of acini. An acinus is a collection of pyramidal acinar cells that funnels secretions to the duodenum through the intercalated duct and duct of Wirsung. The remaining

1-4% of the total volume is the endocrine gland that secretes hormones directly to the bloodstream. This system comprises of Islets of Langerhans or simply islets and secretes four different hormones: glucagon through α-cells, insulin through β-cells, somatostatin and gastrin through δ-cells, and pancreatic polypeptide through PP-cells (Thorel et al. 2010).

β-cells are the predominant cell type in the endocrine pancreas. They are responsible for maintaining the body's glucose levels by synthesizing and secreting insulin and amylin or islet amyloid polypeptide (IAPP) directly into the blood. β-cells rapidly respond to a spike in blood glucose level by secreting stored insulin and amylin and simultaneously producing more.

1.2.2 Insulin biosynthesis, secretion and action

Insulin is initially translated as preproinsulin and is processed to proinsulin form in the rough endoplasmic reticulum (ER), where the signal peptide is cleaved off by signal peptidase (Tokarz et al. 2018). This proinsulin is further folded into two domains, A domain (semi-helical) and B domain (helical), via three disulfide bonds. These two domains remain attached to each other with the help of the C-chain (C-peptide). The complete complex is transported to the Golgi apparatus, where prohormone convertases PC1/3 and PC2 cleave off the C-chain (Rorsman and Renström 2003; Tokarz et al. 2018). This results in the generation of mature insulin, which is reserved as hexameric insulin/Zn^{2+} crystals in the secretory granules in the cytoplasm of β-cells (Rorsman and Renström 2003; Tokarz et al. 2018). This complete procedure is relatively rapid, requiring less than two hours and is efficient (Lameire 2000).

Although many factors control insulin biosynthesis, glucose metabolism is among the most important (Fu et al. 2013). With the increase in glucose levels after diet intake, β-cells speed up the insulin translation, which is partly controlled by dephosphorylation of eukaryotic initiation factor 2a (eIF2a) via protein phosphatase 1. Phosphorylated eIF2a gets rapidly

dephosphorylated with the stimulus of high glucose levels in the blood through a signaling pathway that activates protein phosphatase 1. As a result, new preproinsulin enters ER and burdens it. This causes interaction with ER chaperone BiP and rephosphorylation of pancreatic ER kinase. This rephosphorylation acts as a balance mechanism to switch off insulin translation (Fu et al. 2013).

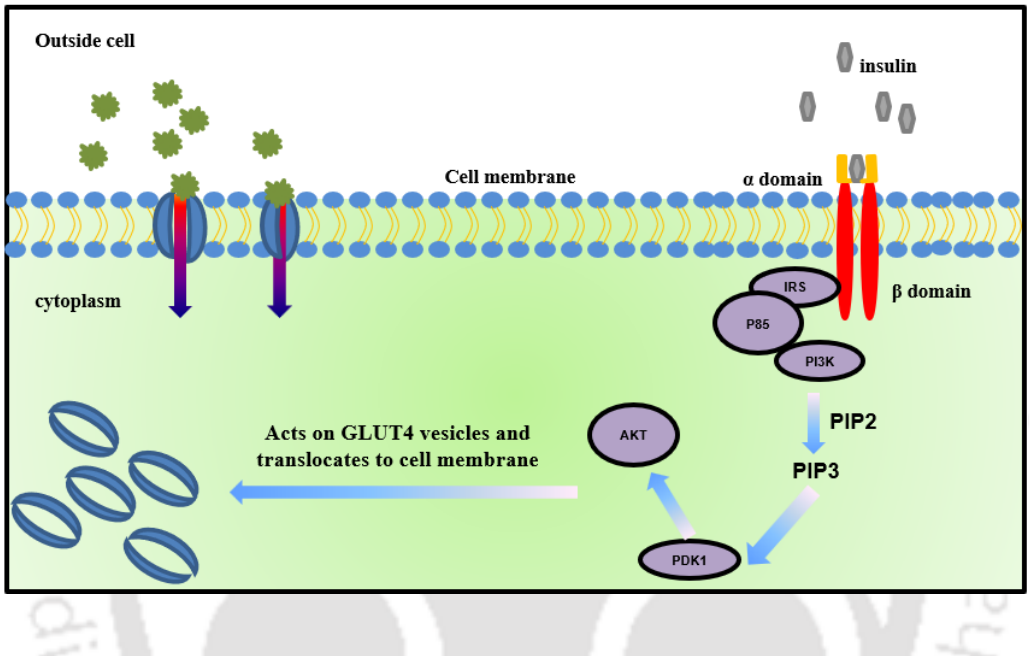


Figure 1.1: Detailed mechanism of insulin binding to its receptor and translocating GLUT4 vesicles to the membrane. As the secreted insulin molecules bind to the α-subunit of insulin receptor (IR), a number of downstream signaling cascade turns on. As a result, the GLUT4 molecules translocate to the membrane facilitating the glucose molecules to enter the cytoplasm.

The secreted insulin binds to the extracellular domain of the insulin receptor that acts as a transducer in the cell membrane to switch on the downstream signaling cascade. The insulin receptor (IR), a transmembrane receptor tyrosine kinase member, is a disulfide-linked heterodimeric protein in the form $(\alpha\beta)_2$. On binding of insulin to the extracellular domain (α domain), it comes closer and brings about conformational changes to its transmembrane

domain (β domain). The intracellular signaling is mediated through IRS-1 and its 20-22 potential phosphorylation sites that bind to various SH2 group-containing proteins, result in the activation of the PI3 kinase pathway. This activation translocates glucose transporter 4 (GLUT4) from the intracellular pool to the cell membrane facilitating glucose intake by the cells (Ottensmeyer et al. 2000). The detailed mechanism is represented in Figure 1.1.

1.2.3 Diabetes Mellitus

Diabetes mellitus (DM) is a group of metabolic disorders characterized by high blood glucose levels due to insufficient insulin secretion, insulin action, or both. These patients have impaired carbohydrate, fat and protein metabolism. Symptoms include polyuria, polydipsia and weight loss. The prolonged hyperglycaemic condition causes retinopathy, neuropathy and nephropathy resulting in renal failure. Moreover, diabetic patients are found to be prone to atherosclerotic cardiovascular and cerebrovascular diseases. In acute cases, a ketoacidosis state may develop, leading to stupor and coma. The global prevalence of DM has been increasing exponentially every year. The statistical data revealed its occurrence to increase from 30 million in 1964 to 536 million in 2021, mainly affecting low- or middle-income countries (Sun et al. 2022). Moreover, its occurrence rate is estimated to increase to 783.2 million by 2045. The vast cases of DM are classified into two major categories:

Type 1 DM (T1DM): T1DM accounts for around 5-10% of diabetic cases. T1DM is signified by the varying rate of destruction of β-cells, from fast destruction (mainly in infants) to slow destruction (in the case of adults). It leads to no or reduced insulin secretion affecting children below 14 years. In most cases, ketoacidosis is the first symptom shown, whereas in other cases, the patients suffer from acute hyperglycemia, making them completely dependent on regular insulin shots. However, prolonged usage of insulin shots is often associated with insulin resistance, weight gain and ketoacidosis (Atkinson et al. 2014).

Type 2 DM (T2DM): T2DM was previously referred to as non-insulin dependent diabetes and is the most common form of diabetes, accounting for 90-95% of the cases encompassing individuals with insulin resistance or lower insulin secretion compared to normal value. In most cases, the patients are prescribed oral anti-diabetic drugs and in later stages, insulin injections may be given to cater to normal insulin levels. Most patients are often obese, and obesity, to some degree, causes insulin resistance. If not obesity, stress level plays a crucial level in the onset of T2DM. Such patients are at an increased risk of getting micro- and macrovascular diseases. However, the specific reasons for the cause are still unknown (Kahn et al. 2014).

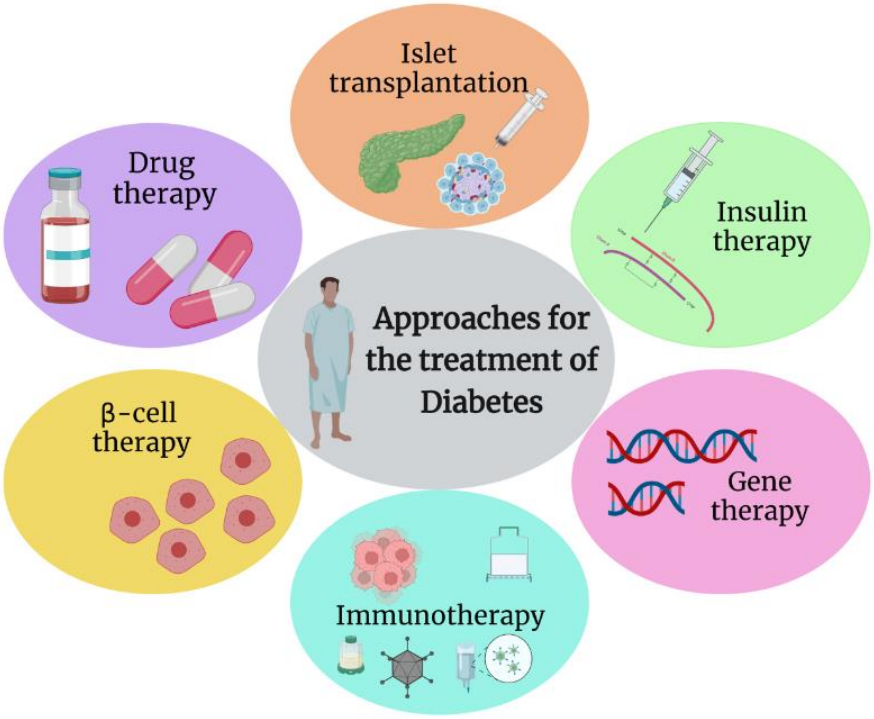


Figure 1.2 A pictorial representation of various approaches for treating diabetes mellitus. This includes islet transplantation, insulin therapy, drug therapy, β-cell therapy, gene therapy, and immunotherapy

To tackle this ever-growing diabetes population, different approaches are undertaken. However, some, like insulin shots or diabetic medications, can only provide temporary relief. A detailed pictorial representation of the existing approaches is provided in Figure 1.2. Therefore, different groups have undertaken different strategies to produce functional β-cells that can cure diabetes permanently. These strategies are further discussed below:

1.2.4 Strategies to generate insulin-producing β-cells

Different groups of researchers have undertaken multiple strategies to generate insulin-producing cells (Figure 1.3). In the recent times, several groups reported the generation of insulin-producing pancreatic β-cells by reprogramming adult stem cells (Oh et al. 2004; Cardinale et al. 2011), progenitors (Hori et al. 2005), somatic cells (Zhu et al. 2016). These cells not only expressed specific β-cell markers but were also capable of reversing hyperglycemia in streptozotocin-induced diabetic mice.

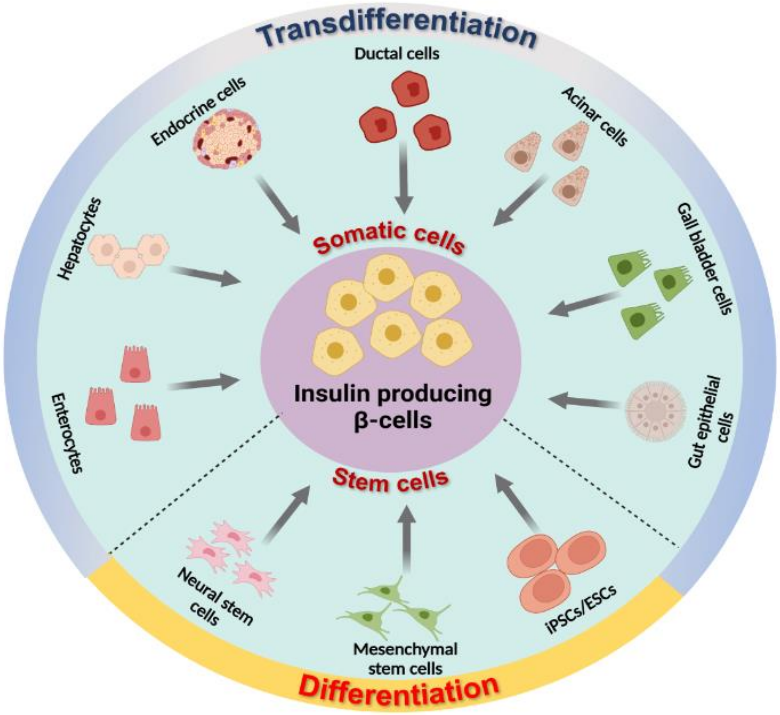


Figure 1.3 Various adult cells that can be reprogrammed into insulin-producing β-cells. A representation of various terminally differentiated somatic cells that can be transdifferentiated to insulin-producing β-cells is depicted. This includes enterocytes, hepatocytes, endocrine cells, ductal cells, acinar cells, gall bladder cells and gut epithelial cells. The stem cells that can be differentiated into insulin-producing β-cells such as neural stem cells, mesenchymal stem cells and iPSCs/ESCs, also represented (iPSCs: induced pluripotent stem cells; ESCs: Embryonic stem cells)

1.2.4.1 Generation of insulin-producing β-cells from induced Pluripotent Stem Cells

The generation of insulin-producing β-cells from induced Pluripotent Stem Cells (iPSCs) offers great promise for treating degenerative diseases or disorders such as DM. To generate these cells, any somatic cell type can be reprogrammed to a pluripotent state by ectopic expression of defined factors. Eventually, these pluripotent stem cells will give rise to any desired cell type. Similarly, somatic cells from a diabetic patient can be reprogrammed to iPSCs, followed by differentiation towards insulin-producing β-cells. These cells can then be transplanted back to the same patient to replace the destroyed (in case of T1DM) and dysfunctional (in case of T2DM) pancreatic β-cells. In essence, iPSCs are non-controversial cell types, providing an unlimited source of pluripotent stem cells allowing its broader application in personalized medicine (Ferreira and Mostajo-Radji 2013).

Pancreas organogenesis in humans and mice demonstrates a series of stages marked by the endogenous expression of several transcription factors, which varies from stage to stage (Pan and Wright 2011; Jennings et al. 2013). To date, numerous studies reveal the differentiation of iPSCs into insulin-producing β-cells *in vitro* in a similar fashion that mimics the embryonic pancreas organogenesis stages (Tateishi et al. 2008; Zhu et al. 2011; Shahjalal

et al. 2014; Millman et al. 2016; Enderami et al. 2018; Yabe et al. 2019). This differentiation process generally includes the sequential treatment of iPSCs with several small molecules and growth factors at multiple differentiation stages, which successfully induces the expression of key pancreatic-specific transcription factors (Tateishi et al. 2008; Zhang et al. 2009; Kunisada et al. 2012). Most iPSCs to β-cell differentiation protocols follow a sequential pancreas developmental stage, which includes the formation of 1) definitive endoderm, 2) primitive gut tube, 3) pancreatic endoderm, 4) pancreatic progenitors/precursors, 5) pancreatic endocrine cells, 6) β-cells (Zhu et al. 2011; Shahjalal et al. 2014; Millman et al. 2016; Pellegrini et al. 2016; Haller et al. 2019; Wang et al. 2019; Yabe et al. 2019). However, generating safer iPSCs devoid of integrated foreign transgenes is considered a prime hurdle that must be addressed. In most cases, iPSCs are generated using viral methods such as γ-retro- and lenti-viral vectors due to their higher efficiency compared to non-viral counterparts. iPSCs generated through viral-based approaches might cause insertional mutagenesis resulting in tumor formation. Integration-free reprogramming approaches like the Sendai virus, microRNAs, synthetic messenger RNAs, small molecules, and recombinant proteins can generate safe iPSCs with potential clinical applicability (Borgohain et al. 2019; Dey et al. 2021a). Moreover, lower efficiency and lengthy protocols often mar its usage. Additionally, the generated β-cells are immature and have limited glucose response (Zhang et al. 2009). These differentiated cells are also observed to be heterologous, containing a mixture of bihormonal (insulin$^+$/glucagon$^+$) cells along with progenitor cells (Vethe et al. 2017). It has also been noted that the reprogrammed cells cannot secrete insulin upon varying glucose concentrations (Tateishi et al. 2008). Hence, further research must be performed to understand the molecular mechanisms that can lead to effective differentiation protocols, resulting in a higher number of functional and homogeneous population of β-cells that can be used in cell therapy.

1.2.4.2 Generation of insulin-producing β-cells from somatic cells

Direct reprogramming of various differentiated somatic cells to another cell type bypassing the stem cell state has proved to be a promising approach in regenerative medicine (Ferber et al. 2000; Ieda et al. 2010). This approach saves us from the arduous and lengthy procedure taken in the case of stem cell-based therapy. Direct reprogramming involves ectopic expression of key transcription factors in terminally differentiated cells in the presence of essential growth factors under defined culture conditions to generate a desired cell type. To date, various adult cell types have been used by different studies to generate functional β-cells for prospective biomedical applications. The most widely used cells are hepatocytes, endocrine-specific cells, acinar cells and ductal cells.

Direct reprogramming of liver cells to β-cells

In embryonic development, the liver and pancreas share a developmental history, and it has been proposed that these organs develop from the same precursor cells (Sumazaki et al. 2004; Cerdá-Esteban et al. 2017; Xiao et al. 2018). These organs contain a variety of different cell types, which are of endodermal origin as they arise from adjacent regions of the anterior endoderm during embryogenesis (Sumazaki et al. 2004; Cerdá-Esteban et al. 2017; Xiao et al. 2018). Therefore, liver cells have been the first choice for direct reprogramming to give rise to insulin-producing β-cells (Figure 1.4). During early 2000, various *in vivo* as well as *in vitro* studies demonstrated the potency of hepatocytes for their efficient conversion to pancreatic cells (Cavelti-Weder et al. 2015a). Studies have also reported that the expression of a single pancreatic-specific gene was sufficient to bring about the direct conversion of hepatocytes to β-cells (Ferber et al. 2000; Horb et al. 2003; Sumazaki et al. 2004; Jarikji et al. 2007). Pdx1 is the most explored transcription factor in this aspect and has been used by various studies to convert hepatocytes to insulin-secreting β-like-cells (Ferber et al. 2000; Zalzman et al. 2003; Horb et al. 2003; Sapir et al. 2005). Forced expression of this homeodomain transcription factor

in mouse liver cells (Ferber et al. 2000), rat enterocytes (Kojima et al. 2002), human hepatocytes (Zalzman et al. 2003; Sapir et al. 2005) and *Xenopus laevis* (Horb et al. 2003; Jarikji et al. 2007) resulted in the upregulation of multiple pancreatic-specific genes like glucagon, pancreatic polypeptide and elastase, including insulin (Zalzman et al. 2003). The reprogrammed cells stored and secreted insulin from defined granules in a glucose-regulated manner. When Pdx1 was retrovirally transduced in mouse liver cells, it activated insulin 1, insulin 2, and PC1/3 genes (Ferber et al. 2000). In addition, it ameliorated streptozotocin-induced diabetes; however, the effect of ectopic Pdx1 on insulin secretion was monitored for only two weeks (Ferber et al. 2000). Pdx1, when tagged to VP16 (a transcriptional activation domain from herpes simplex virus) in transgenic *Xenopus laevis* tadpoles as well as human and murine hepatocytes, converted a specific region of the liver to the endocrine pancreas and reversed streptozotocin-induced hyperglycemia with the substantial increase in blood insulin levels (Horb et al. 2003; Imai et al. 2005; Kaneto et al. 2005). The converted cells secreted either insulin or amylase, which is a key exocrine enzyme (Horb et al. 2003), indicating the formation of both pancreatic endocrine and exocrine cells. Notably, it has been observed that liver-specific gene expression is turned off from the regions converted to the pancreas (Horb et al. 2003). In contrast, Imai and group reported no expression of amylase in the converted cells; however, the cells continued to express albumin, transferrin and hepatocyte markers in insulin-producing cells, suggesting incomplete transdifferentiation of adult murine liver cells (Imai et al. 2005). Moreover, when these groups compared the transdifferentiation capability of Pdx1-VP16 to its unmodified variant in *Xenopus leavis* tadpoles and mammalian cells, it failed to reprogram liver cells to β-cells (Horb et al. 2003; Imai et al. 2005). It was speculated that wild-type Pdx1 required specific coactivators like histone acetyltransferases (originally not expressed in the liver) for transdifferentiation; therefore, the fusion of VP16 to PDX1 helped to overcome the absence of these coactivators to enable transdifferentiation (Horb et al.

2003). Furthermore, cells treated with Pdx1 along with Activin A in serum-free conditions showed upregulation of β-cell markers like NeuroD and Nkx2.2 and downregulation of liver and other pancreatic-specific markers like pancreatic polypeptide and glucagon (Zalzman et al. 2005a). Other pivotal β-cell transcription factors like NeuroD1 and Ngn3, along with Pdx1-VP16, have been shown to improve insulin biosynthesis and significantly ameliorate glucose tolerance (Kaneto et al. 2005). To further investigate, other groups also transdifferentiated liver cells to pancreatic endocrine and exocrine cells by ectopically overexpressing Pdx1 and observed the upregulation of endocrine-specific hormones like insulin, glucagon, somatostatin, and pancreatic polypeptide (Miyatsuka et al. 2003; Ber et al. 2003). In a study, Ber and colleagues suggested that Pdx1 induces its expression and confers the "long-lasting" effect to the converted cells (Ber et al. 2003). However, a wide repertoire of pancreatic genes like glucagon, somatostatin, etc., were activated during the process and the mice transplanted with generated β-like cells suffered from hyperbilirubinemia (Miyatsuka et al. 2003; Ber et al. 2003). In 2004, Sumazaki and co-workers converted cells of the biliary system (epithelium) to pancreatic tissue in *Hes1*-deficient mice (Sumazaki et al. 2004). The reason to choose *Hes1*-deficient mice was that the *Hes1* gene is known to encode for Hes1 protein, which negatively regulates endodermal endocrine differentiation by repressing Ngn3 during biliary organogenesis (Sumazaki et al. 2004). The group noticed that biliary epithelium in *Hes1*$^{-/-}$ mice expressed Ngn3 and the cells differentiated into endocrine and exocrine cells forming acini or islet-like structures (Sumazaki et al. 2004). Similar to Pdx1, Ptfa also can convert early liver buds to pancreatic cells when tagged to VP16 in *Xenopus laevis* tadpoles (Jarikji et al. 2007). It was observed that wild-type Ptf1a protein failed to convert liver cells to pancreatic cells; however, it succeeded in converting stomach and duodenum cells to insulin-producing cells (Jarikji et al. 2007). It is noteworthy that Ptf1a-VP16 treated liver cells mimicked acinar cells with abundant expression of only amylase enzyme and no detection of insulin or somatostatin,

suggesting its role in promoting acinar fate (Jarikji et al. 2007). On the other hand, ectopic overexpression of wild-type Ptf1a in stomach and duodenum cells promoted both exocrine as well as endocrine fate (Jarikji et al. 2007). Furthermore, three-dimensional culture conditions not only improved the viability of the reprogrammed cells but also promoted the expansion of the starting cell type (Nagaya et al. 2009). When intrahepatic biliary epithelial adult mouse cells were cultured in a two-dimensional culture condition, the cells died within two weeks, whereas cells cultured in a three-dimensional condition could sustain cell growth and promote expansion with improved cell viability (Nagaya et al. 2009). These cells were positive for mature β-cell markers like Ins1, Ins2, and PC1 and PC2 (Nagaya et al. 2009). However, faint expression of amylase was also observed in the differentiated cells (Nagaya et al. 2009), indicating the presence of pancreatic exocrine cells in the transdifferentiated cells. It is also worth mentioning that only a few cells released insulin after seven days of transduction of the Ad-Pdx1-VP16 adenoviral vector and the inclusion of multiple growth factors and small molecules like nicotinamide, Activin A, betacellulin, hepatocyte growth factor, and exendin-4 could not improve the differentiation efficiency significantly (Nagaya et al. 2009). Furthermore, a different group claimed that overexpression of Ngn3 using a helper-dependent adenoviral vector along with betacellulin is sufficient to switch cellular fate from hepatic-progenitor cells to glucose-responsive insulin-producing cells (Yechoor et al. 2009). However, these cells showed abnormal transcriptional cascade and eventually lost the capability of producing insulin due to failure in expressing other β-cell markers like Pax4, Nkx2.2 and Nkx6.1. A combination of Ngn3 along with Pdx1 and MafA secreted 10-fold higher insulin protein and prolonged the glucose responsiveness for up to four months in streptozotocin-treated NOD-SCID mice, proving the greater effectiveness of the combination compared to a single factor (Ackermann et al. 2016). However, the converted cells failed to express insulin exclusively and stained positive for other pancreatic markers like glucagon and somatostatin,

proving mixed phenotypes. The cells also showed lower levels of *Nkx2.2* and *Nkx6.1* (Ackermann et al. 2016), which are crucial for β-cell development, proliferation and maintenance (Doyle and Sussel 2007; Aigha and Abdelalim 2020). The same group also reprogrammed mouse embryonic liver cells, and their findings were in line with the previous results (Yang et al. 2013). However, the insulin content in these cells was as low as 3% of the mature β-cells and lacked glucose responsiveness showing the immature phenotype of the reprogrammed cells (Yang et al. 2013). In the following year, Berneman-Zeitouni and the group shed light on the temporal and hierarchical control of the transcription factors imparting on the reprogramming process (Berneman-Zeitouni et al. 2014). The study explored the role of three pivotal transcription factors, PDX1, PAX4 and MAFA, which control three stages of pancreatic development in the human liver to pancreatic transdifferentiation (Berneman-Zeitouni et al. 2014). The combined effect of these factors substantially increased the number of insulin-positive cells. Moreover, the reprogrammed cells displayed mature β-cell characteristics like glucose-stimulated insulin secretion (Berneman-Zeitouni et al. 2014). The aforementioned studies used viral gene delivery methods, which can result in random viral integration causing insertional mutagenesis and tumorigenicity (Borgohain et al. 2019; Haridhasapavalan et al. 2019; Dey et al. 2021a). Although these studies elucidated the transcriptional landscape in pancreatogenesis, the reprogrammed cells hugely failed for clinical applications and, thus, require a non-genetic approach. In 2017, Yang and group used a hydrodynamic-based gene delivery method to introduce Pdx1, Ngn3 and MafA into mouse hepatocytes and reprogrammed them into insulin-producing cells (Cavelti-Weder et al. 2015a). Multiple hydrodynamic injections successfully reversed hyperglycaemic conditions in diabetic mice along with the expression of multiple pancreatic-specific developmental genes like *Nkx6.1, NeuroD1, Pax4* and *Isl1* (Cavelti-Weder et al. 2015a). Thus, multiple studies have proved that liver cells or hepatocytes are a preferable option in direct reprogramming to β-cells.

However, the safest method for transdifferentiation remains in question and requires further detailed investigation.

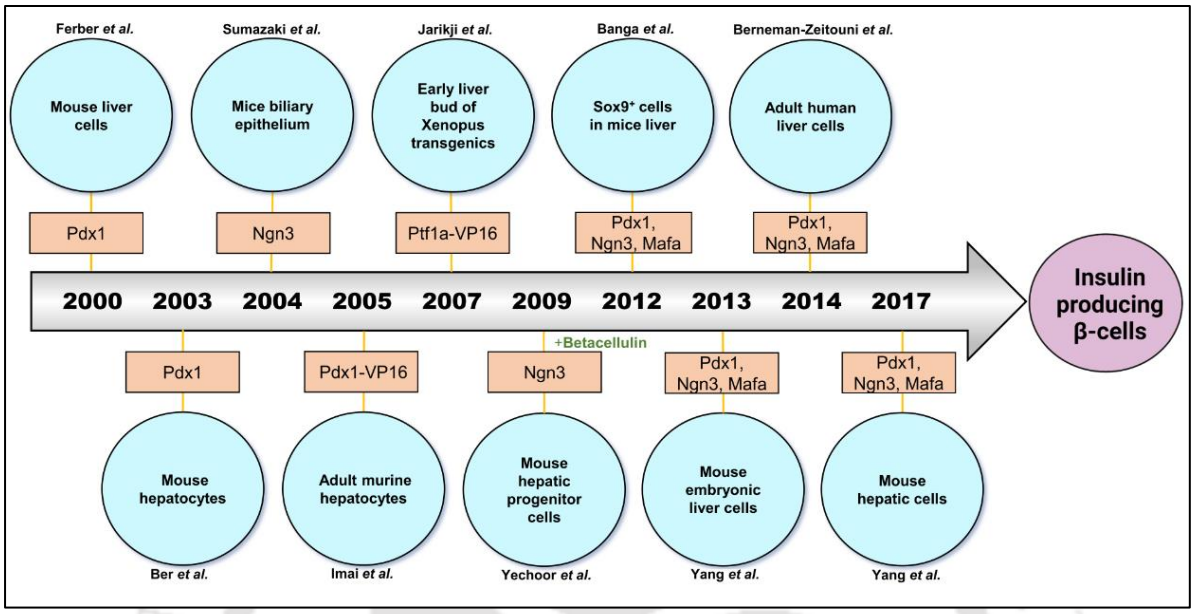


Figure 1.4 A timeline depicting different cells from the embryonic or adult liver that can be reprogrammed into insulin-producing β-cells using various reprogramming factors

Direct reprogramming of endocrine cells to β-cells

Endocrine pancreatic cells are arranged in a small cluster of cells called islets of Langerhans or simply islets. Each cluster/islet comprises majorly of β-cells (around 90%), which secretes insulin. It also contains α-cells, δ-cells, ε-cells and PP-cells (around 10%) that secretes glucagon, somatostatin, ghrelin and pancreatic polypeptide, respectively. Insulin secreted by β-cells helps to maintain normoglycemic conditions. Several studies have used endocrine cells as the starting cell source (Figure 1.5) (Thorel et al., 2010; Bramswig et al., 2013; Chera et al., 2014; Collombat et al., 2009; Sangan et al., 2015; Yang et al., 2011), primarily because of the developmental similarities. Among the various endocrine cell types available, α-cells are the most widely used due to multiple reasons such as (i) their inherent plasticity, (ii) sharing developmental similarity with β-cells that enables reprogramming, (iii) α-cell hyperplasia

observed in diabetic patients, which in turn, provides an abundant source for reprogramming, and (iv) easily accessible α-cell genome showed by Assay for Transposase-Accessible Chromatin with high-throughput sequencing, making these cells amenable for transdifferentiation (Ye et al. 2015; Ackermann et al. 2016; Chakravarthy et al. 2017; Xiao et al. 2018). The plasticity of α-cells was demonstrated by Thorel and co-workers through cell lineage-tracing experiments *in vivo* (Thorel et al. 2010). This study showed that α-cells got converted to β-cells spontaneously upon near-total β-cell ablation and insulin treatment, demonstrating the occurrence of inter-endocrine cell plasticity (Thorel et al., 2010). This study speculated that a near-total ablation of β-cells released some form of signals, which resulted in the physiological need to replenish β-cells that allowed its regeneration from α-cells (Thorel et al., 2010). This study also observed that the regeneration efficiency hugely varied between 32 and 81%, even when the ablation degree was similar (Thorel et al., 2010). Later the same group reported that the conversion of α-cells to β-cells was age-independent (Chera et al. 2014). The study showed that senescence did not alter α-cell plasticity, and these cells continued to convert to β-cells from puberty to adulthood upon near-total β-cell loss. However, before the rise of puberty, no detectable conversion of α-cells to β-cells was observed, and it was in juveniles that δ-cells got converted to β-cells to replenish the lost cells, with a remarkable decrease in somatostatin levels post-conversion (Chera et al. 2014). This was demonstrated by cell lineage tracing experiments, which revealed that regenerated insulin^{+} cells were from dedifferentiated δ-cells (Chera et al. 2014). However, deeper insights into the transcriptional network and epigenetic regulation are required to understand the underlying mechanisms involved in converting multiple endocrine cell lineages, such as α-cells and δ-cells to β-cells.

Various studies have focused on the forced expression of a particular key β-cell transcription factor and elucidated how each factor bestows β-cell phenotype to non-β-cells. One of the key transcription factors, Pax4, is known to favor β-cell fate and identity during

endocrine morphogenesis (Mellado-Gil et al. 2016). Thus, Collambat and co-workers explored the potentiality of Pax4 in converting mouse pancreatic progenitor cells and adult α-cells to insulin-producing β-cells (Collombat et al. 2009). From this study, it was worth noticing that post-partum, an increase in the islet size was observed upon ectopic Pax4 overexpression, and this increase was primarily due to insulin$^+$/Pax4$^+$ cells (Collombat et al. 2009). Markedly, the number of glucagon$^+$ cells was reduced by 77%, while the population of other islet-specific cell types remained unaltered (Collombat et al. 2009). In addition, there was an increase in islet size and the number of insulin-producing cells, suggesting the role of Pax4 in continuously converting glucagon$^+$ cells to β-cell phenotype (Collombat et al. 2009). Also, it was highlighted that the reactivation of Ngn3 upon Pax4 overexpression is crucial to address the hypoglucagonemia condition, which activated various compensatory mechanisms, provoking progenitor cells to convert first to α-cells and then to β-cells. These converted β-cells are functional at an early age and repopulate the islets of diabetic mice and could maintain normoglycaemia (Collombat et al. 2009). Another common and remarkable transcription factor that has been employed to convert different endocrine-specific cell lineages to β-cells is Pdx1. Pdx1 is the 'master regulator' in β-cell maturation, proliferation, and function as its homozygous mutation led to hyperglycaemic condition (Gannon et al. 2001; Gao et al. 2014). When Pdx1 expression was enforced in endocrine-committed Ngn3$^+$ cells during embryonic development, a slight increase in the number of β-cells was observed with the subsequent decrease in α-cells, while postnatally, this conversion continued at a higher rate, leading to the complete absence of α-cells. This study revealed the single-handed role played by Pdx1 as a potent reprogramming factor. Pdx1, when combined with MafA, successfully reversed diabetes in alloxan-treated NOD/SCID mice within two weeks, restoring euglycemia (Xiao et al., 2018). Moreover, significant improvement in the glucose response was also observed in mice, along with a substantial increase in β-cell mass (Xiao et al., 2018). Sangan et al. explored the role of

the HNF4α transcription factor in transdifferentiation and demonstrated that this factor could also induce β-cell phenotype in α-cells (Sangan et al. 2015). When HNF4α was ectopically expressed in αTC1-9 cells, a commonly used murine α-cell line, it converted α-cells to β-cells by suppressing glucagon expression (Sangan et al. 2015). The reprogrammed cells were of increased size and positive for β-cell-specific markers like insulin, C-peptide, GLUT2, glucokinase and Pax4, enabling the cells to secrete insulin in a glucose-regulated manner (Sangan et al. 2015). However, the conversion was incomplete as the cells lacked the expression of key factors like Pdx1 (Sangan et al. 2015). Moreover, ectopic expression of Pax4 alone in αTC1-9 cells showed no effect on gene expression *in vitro*, which contradicted the previous *in vivo* findings by Collombat and group (Collombat et al. 2009; Sangan et al. 2015), indicating that additional factors are essential for the transdifferentiation of α-cells to β-cells. Glucagon-like peptide 1 (GLP1) is one such potential factor that reversed hyperglycemia in diabetic animal models and restored the β-cell population (Zalzman et al. 2003). This study demonstrated that GLP1, when infused in streptozotocin-induced rat pancreatic islets, promoted the endogenous neogenesis of β-cells, possibly by regulating GLP1 receptor and its downstream signaling pathway (PI3K/AKT/FOXO1) (Zalzman et al. 2003). This regulation resulted in enhanced Pdx1 and MafA mRNA expression and reduced MafB levels (Zalzman et al. 2003), suggesting that GLP1 promoted β-cell function and suppressed α-cell identity.

Apart from pancreatic-specific genes, epigenetic markers also play a pivotal role in reprogramming α-cells into β-cells. Bramswig and group elucidated the epigenetic landscape involved in α-cells and exocrine cell transition to β-cells (Bramswig et al. 2013). The group discovered that α-cells displayed multiple bivalent marks on the development genes and showed similar bivalent modification profiles to ESCs. This similarity indicates that α-cells have a plastic epigenomic state, explaining the relative ease of reprogramming α-cells to β-cells (Bramswig et al. 2013).

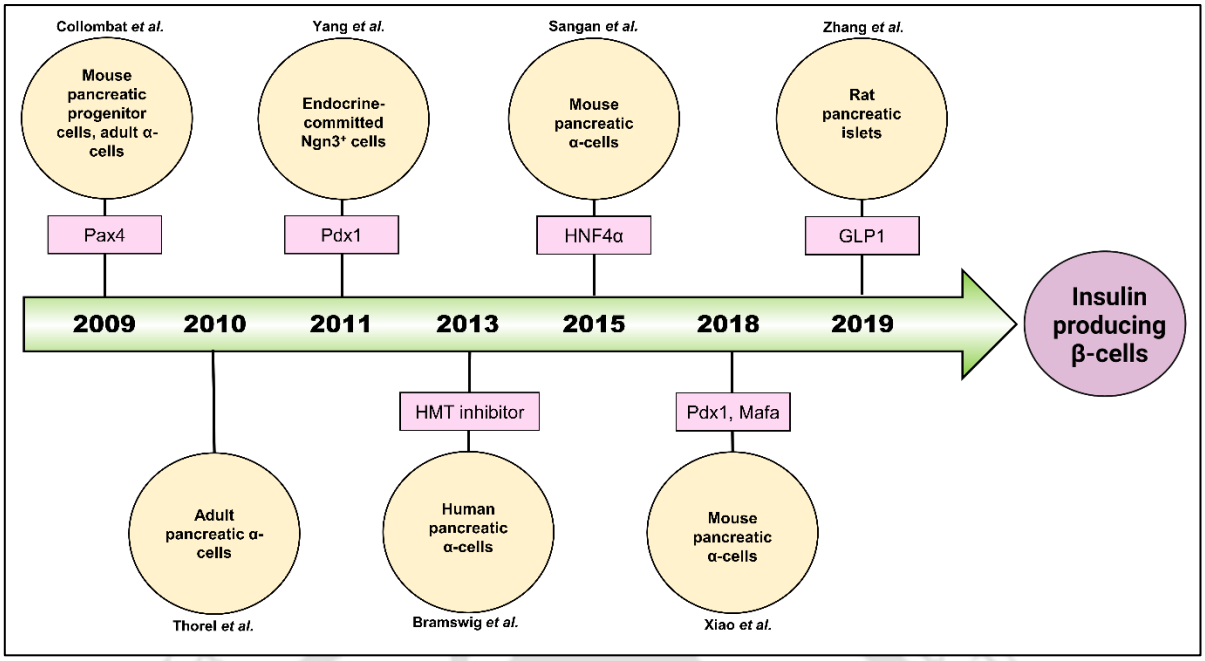


Figure 1.5 A timeline depicting endocrine cells from different sources (both mouse and human) that can be reprogrammed to insulin-producing β-cells using various reprogramming factors. (HMT: histone methyltransferase; HNF4α: hepatocyte nuclear factor 4 alpha)

Direct reprogramming of exocrine cells to β-cells

Acinar cells

The exocrine portion of the pancreas comprises acinar and ductal cells that contribute to 95% of the total mass of the pancreas (Das et al. 2014). It also shares developmental similarities with the endocrine pancreas, which makes these cells a preferred choice for dedifferentiation into β-cells. It has been observed that exocrine cells, such as acinar cells, can turn on endocrine machinery when cultured in specific media conditions *in vitro* (Figure 1.6) (Baeyens et al. 2005; Minami et al. 2005). This conversion was shown by merely treating acinar cells with Activin A and betacellulin to generate insulin-producing cells (Mashima 1996). However, the cells also stained positive for the pancreatic polypeptide, indicating incomplete transdifferentiation of acinar cells. Later, two groups attempted to convert pancreatic exocrine

cells to insulin-producing cells (Baeyens et al. 2005; Minami et al. 2005). These groups used rat or mouse exocrine cells and treated the cells with different growth factors and small molecules. The resulting cells were identical to endogenous β-cells with insulin-secretory granules (Baeyens et al. 2005; Minami et al. 2005). Furthermore, the cells showed glucose responsiveness, maintained normoglycaemia, and stained positive for key β-cell markers like C-peptide, GLUT2, etc. (Baeyens et al. 2005; Minami et al. 2005). These findings were further validated using cell lineage tracing experiments, which revealed that amylase^{+}/elastase^{+} acinar cells activated epidermal growth factor signaling that favored the transition (Minami et al. 2005). Although these studies claimed to have generated mature β-cells, some reprogrammed colonies stained positive for pre-endocrine or early endocrine markers like protein gene product 9.5. This proved that the cells belonged to a transitional state where it was insulin and protein gene product 9.5 double positives (Minami et al. 2005). Therefore, subsequent studies modified the dedifferentiation protocols to include key β-cell-specific transcription factors for efficient direct conversion.

In 2008, Zhou and group screened 9 (Pdx1, Ngn3, NeuroD1, Nkx2.2, Nkx6.1, Pax4, Pax6, Isl1 and MafA) transcription factors that were reported to have a pivotal role in reprogramming exocrine cells to β-cells (Zhou et al. 2008). Among these factors, Ngn3, Pdx1 and MafA showed the most promising results (Zhou et al. 2008). The differentiated cells closely resembled the endogenous β-cells in size and ultrastructure. Moreover, the cells ameliorated hyperglycemia, remodeled local vasculature, and expressed mature β-cell markers (Zhou et al. 2008). However, the number of induced β-cells declined over the next two months and became undetectable after seven months of transduction (Li et al. 2014). So, it was speculated that the reprogramming efficiency could be improved if the factors could be introduced in a single polycistronic construct to induce efficient direct lineage conversion. Using this strategy, the researchers observed enhanced insulin production that ameliorated

diabetes for up to 13 months (Li et al. 2014). A similar approach was used to generate β-cells from acinar cells using a polycistronic vector containing Pdx1, Ngn3 and MafA and performing their *in vivo* transduction in adult mice (Cavelti-Weder et al., 2017). Using the same set of transcription factors, Akinci and co-workers delivered these factors through an adenoviral approach into rat pancreatic acinar cells (AR42j-B13 cell line) (Akinci et al. 2012). The study observed clear expression of *Ins1* and *Ins2* genes at the RNA level, and the transduced cells stained positive for insulin and C-peptide at the protein level and could cure diabetes in streptozotocin-induced mice (Akinci et al. 2012). However, the major limitation was that the cells were not glucose-responsive and secreted insulin in an unregulated manner (Akinci et al. 2012). Furthermore, it was shown that the inhibition of epithelial-to-mesenchymal transition along with the addition of Rho-associated kinase inhibitor, Y27632, and TGFβ inhibitor reprogrammed human exocrine cells to behave like β-cells (Lima et al. 2013). When dedifferentiated *in vitro*, the target cells tend to grow in monolayers of mesenchymal cells, facilitating lineage conversion of acinar cells to β-cells. The group further demonstrated that the transdifferentiated cells secreted insulin upon glucose stimulation and could normalize blood glucose levels in the streptozotocin-diabetic NOD/SCID mice model (Lima et al. 2013). All these studies thus proved the capability of human as well as murine exocrine cells to convert to β-cells and secrete insulin in a glucose-responsive manner. However, the major shortcoming in these studies was the use of viral methods to bring about the transition. Thus, to avoid the possibility of tumorigenic transformation associated with integrating viral methods and enhancing the safety of the generated β-cells, Koblas and group attempted to reprogram rat pancreatic exocrine cells to insulin-producing β-cells using synthetic mRNAs of the core reprogramming factors, Pdx1, Ngn3 and MafA (Koblas et al. 2016). When mRNAs of these transcription factors were introduced into AR42J cells, the treated cells could process proinsulin to insulin and its byproduct, C-peptide (Koblas et al. 2016). Additionally, the

reprogrammed cells showed the upregulation of multiple key β-cell genes like *Insulin, Sur1, Kir6.2, Pcsk1 and Pcsk2*, along with limited insulin secretion similar to immature β-cells (Koblas et al. 2016). The study further demonstrated that pretreatment of cells with 5-Aza-2'-deoxycytidine, a DNA-hypomethylating agent, followed by transfection with mRNAs improved the reprogramming efficiency from 3.5% to 14.3% (Koblas et al. 2016). Moreover, only 5-Aza-2'-deoxycytidine pre-treated cells could secrete insulin successfully in the presence of high glucose. However, despite the improved reprogramming efficiency, the insulin content was significantly lower than the rat pancreatic islets used as a control (Koblas et al. 2016). The study speculated that epigenetic modulators and additional transcription factors like Nkx6.1, Pax6 and Isl1 that could positively impact insulin gene expression might be required to attain a mature β-cell phenotype (Koblas et al. 2016). Therefore, further optimization and addition of factors are needed to reprogram human or murine pancreatic exocrine cells to β-cells.

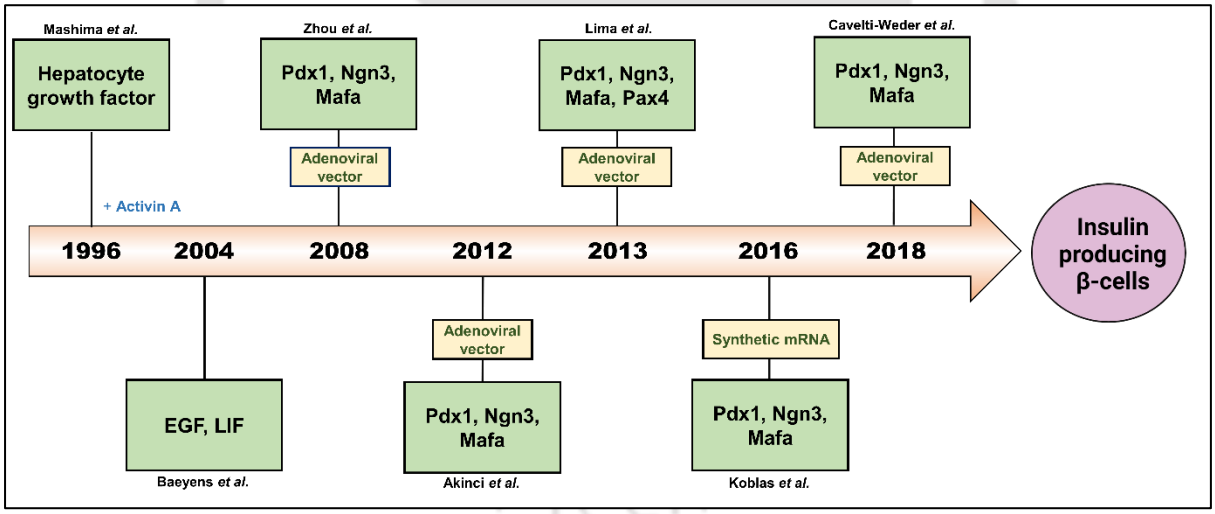


Figure 1.6 A timeline indicating different factors that can be used for reprogramming acinar cells to insulin-producing β-cells. Growth factors like hepatocyte growth factor, epidermal growth factor, leukemia inhibitory factor, and reprogramming factors like Pdx1, Ngn3, and MafA are mostly used. Adenoviral transduction is the most common gene delivery

method used by different studies. Other methods used for the delivery of reprogramming factors are also mentioned.

Ductal cells

It has been proved that ductal cells have the capacity for endocrine differentiation due to the presence of ductal stem cells in the fetus as well as in neonates (Solar et al. 2009). This capability to expand and differentiate is expected to continue even after the perinatal period. Under certain severe conditions, such as pancreatectomy in rats, the cells have been observed to grow and multiply. This opens up a whole new avenue to study islet neoformation involving ductal cells. Different groups have isolated human and mouse ductal cells to test the hypothesis and attempted to reprogram them to hormone-secreting islet cells (Figure 1.7). In early 2000, Bonner-Weir and colleagues cultured pancreatic tissue from eight individuals (Bonner-Weir et al. 2000). The cells were grown in monolayers and overlaid with a film of matrigel. This alteration helped the cells acquire cyst-like structures, eventually leading to islet-like buds. Subsequent culturing of these cells increased the overall insulin content by 10-15-fold (Bonner-Weir et al. 2000). However, the islet cells stained double positive for cytokeratin-19 (ductal marker) and insulin, indicating the immature nature of the transdifferentiated cells (Bonner-Weir et al. 2000).

In the subsequent years, multiple studies have explored the potential of pancreatic ductal cells to differentiate into endocrine and exocrine cells upon deliberate pancreatic injury or pancreatic ductal ligation in both neonatal and adult animals (Inada et al. 2008; Xu et al. 2008; Li et al. 2010; Van de Casteele et al. 2013). The studies proved that, during pancreatic ductal ligation, β-cell progenitors which resided in the ductal lining got activated and compensated for the β-cell loss. These cells were $Ngn3^{+}$ and gave rise to all islet cell types, including proliferative glucose-responsive cells both *ex vivo* and *in situ* (Xu et al. 2008; Van

de Casteele et al. 2013). Interestingly, the conversion of non-β to β-cells improved at higher levels of Ngn3 (Xu et al. 2008; Van de Casteele et al. 2013). It was suggested that multipotent progenitor cells come to play autonomously to elevate β-cell mass rather than self-duplicating the existing β-cells (Xu et al. 2008; Van de Casteele et al. 2013). Moreover, the ablation of $Ngn3^{+}$ cells led to the depletion of β-cell expansion, suggesting the role of $Ngn3^{+}$ $insulin^{-}$ cells in β-cell proliferation (Van de Casteele et al. 2013). These observations were supported by different groups where the ductal cells were specifically marked using the Cre-loxP system under the human carbonic anhydrous II promoter (Inada et al. 2008). These carbonic anhydrous II^{+} cells gave rise to both new islets and acini cells (Inada et al. 2008). Additionally, the expression of Hnf6 was originally lost in mature ductal cells; however, during pancreatic ductal ligation, its expression reappeared in proliferative ductules, which further differentiated into pancreatic lobes (Li et al. 2010). Upon ligation, the cells also started expressing MafA-like mature β-cells, further supporting that pancreatic ductal cells serve as the starting cell source for mature insulin-producing β-cells (Li et al. 2010). In addition, treating ductal cells with gastrin and epidermal growth factor further led to the rise in β-cell number upon β-cell ablation (Rooman and Bouwens 2004). Glp1, in conjunction with Pdx1, also reprogrammed immortalized pancreatic epithelial cells of a ductal origin to insulin-producing cells (Koizumi et al. 2005).

The duct-to-endocrine shift can also be enhanced by inhibiting Delta-Notch signaling and Fbw7, the recognition receptor of the SCF FBW7 E3 ubiquitin ligase, and by co-expressing Myt1 (Sancho et al. 2014; Valdez et al. 2016). In line with this, the loss of Fbw7 resulted in the stabilization of Ngn3 and promoted the endocrine developmental differentiation program (Sancho et al. 2014). Apart from these, activation of MAPK and Akt signaling by overexpression of preadipocyte factor 1 directed human pancreatic ductal cells toward β-like-cell fate by improving glucose homeostasis and facilitating insulin secretion (Rhee et al. 2016).

The study reported that MAPK signaling increased insulin secretion by nucleocytoplasmic translocation of two factors, namely FOXO1 and PDX1 (Rhee et al. 2016). Further, it was reported that Akt signaling activated its downstream targets like Ras-related GTP-binding protein 43, a GTPase-activating protein, which enhanced glucose-stimulated insulin secretion (Rhee et al. 2016). In addition, inflammatory cytokines like TNFα, IL-1β and IFNγ also enabled ductal-to-endocrine cell reprogramming via STAT-dependent NGN3 activation in human ductal epithelial cells, PANC-1 (Valdez et al. 2016). Similarly, treating PANC-1 cells with andrographolide, a chemical compound extracted from *Andrographis paniculata*, induced pancreatic ductal cell differentiation to insulin-producing cells by stimulating PDX1 expression at both mRNA as well as protein levels (Zhang et al. 2020). However, the plasticity of these ductal cells remains a question since different groups reported that the differentiation capacity of these cells is restrained after birth and postnatally (Solar et al. 2009; Kopp et al. 2011). Also, during pancreatic ductal ligation, the ductal epithelium had an insignificant contribution to the acinar and endocrine cell population due to its restricted plasticity (Solar et al. 2009; Kopp et al. 2011). The study supported its findings using lineage-tracing experiments, which revealed that treatment with gastrin and epidermal growth factor failed to reprogram ductal cells to β-cells during pancreatic ductal ligation (Solar et al. 2009). Also, $Sox9^{+}$ ductal cells had an insignificant contribution towards β-cell neogenesis in the early postnatal stage and no contribution in adulthood (Kopp et al. 2011). Nevertheless, exogenous overexpression of pancreatic-specific key transcription factors, PDX1, Ngn3 and MafA, has been reported to augment the reprogramming efficiency of pancreatic ductal cells to insulin-producing cells by expressing β-cell-specific genes and also corrected hyperglycemia in the diabetic mouse model (Van de Casteele et al. 2013; Miyashita et al. 2014; Yamada et al. 2015). Sequential administration of these defined factors in the required dosage can further improve the reprogramming efficiency (Miyashita et al. 2014). It has been observed that excessive

expression of MafA along with Pdx1 and Ngn3 resulted in an inhibitory effect in the reprogramming process (Miyashita et al. 2014). However, overexpression of Pdx1 along with Ngn3 and MafA enhanced the reprogramming efficiency (Miyashita et al. 2014). These findings established the importance of the dose of application in the reprogramming paradigm.

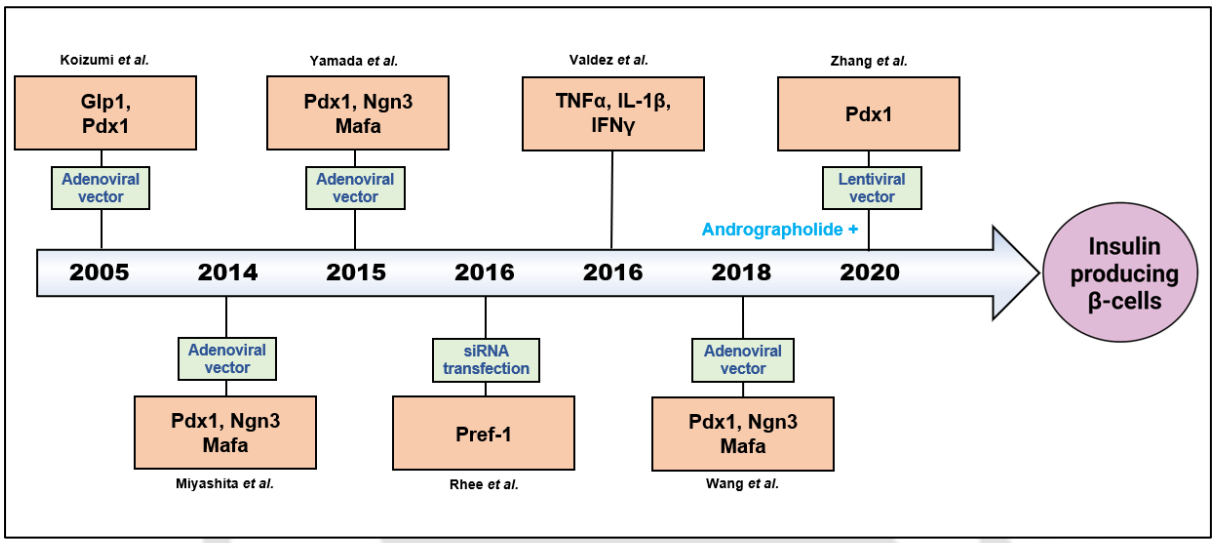


Figure 1.7 A timeline indicating different factors that can be used for reprogramming ductal cells to insulin-producing β-cells. Proinflammatory cytokines such as TNFα, IL-1β, IFN-γ, and reprogramming factors such as Glp1, Pdx1, Ngn3, MafA, and Pref-1 are commonly used. Adenoviral transduction is the most common gene delivery method used by different studies. Other methods used for the delivery of reprogramming factors are also mentioned (TNFα: Tumor necrosis factor alpha, IFN-γ: Interferon-gamma, IL-1β: Interleukin 1β).

1.2.5 Role of pancreatic-specific transcription factors in β-cell formation and maintenance

With the emergence of alternative means to treat the budding number of diabetic patients, a clear understanding of the underlying molecular mechanism involving the various β-cell-specific transcription factors in its formation, expansion and maintenance is required. Transcription factors are those regulatory proteins that play a pivotal role in the pancreas

development paradigm and islet differentiation/formation. However, most of the knowledge acquired is derived from studies carried out in rodent models and genetically modified mouse; the distinction between rodent pancreatic development from humans are worth mentioning.

PDX1

Pancreatic and duodenum homeobox 1 (or Pdx1) is a homeodomain transcription factor that has important roles in pancreatic development and is considered to be the "master regulator" in β-cell formation (Conrad et al. 2014). This factor comprises primarily two domains: the transactivation domain and the homeobox domain. It is reported to play a substantial role in pancreatic development during embryogenesis, and in due course of development, its expression is confined to β-cells (Krapp et al. 1996; Oliver-krasinski et al. 2009; Gao et al. 2014). However, the level of PDX1 differs between mature and pre-mature β-cells in rodents or human islets, and this heterogeneity is required for islet dynamics and glucose-stimulated insulin secretion (GSIS) (Nasteska et al. 2021; Benninger and Kravets 2022). The study further reported that proportionally higher levels of Pdx1 and MafA lead to reduced insulin secretion leading to defects in metabolism and ionic fluxes. In contrast, lower levels contribute to transcriptionally immature cells that hamper islet function and insulin secretion (Nasteska et al. 2021). However, the expression level of PDX1 does not remain constant during embryonic development. In the human fetal pancreas, the first peak in its expression is observed during the 8^{th}-12^{th} week of development and is considered the first transition (Lyttle et al. 2008). The expression level decreases by the 16^{th} week of embryonic development, facilitating the formation of glucagon-producing α-cells. The next peak is observed around the 19^{th}-21^{st} week of embryonic development, creating a second transition (Lyttle et al. 2008). Conditional removal of *the Pdx1* gene in developing β-cells resulted in hyperglycaemic conditions within days due to decreased number of $insulin^{+}$ cells (Gao et al. 2014). This proved that PDX1 directly regulates the insulin gene, and thus it was formerly known as Insulin Promoter Factor

1 (IPF1) (Wang et al. 2016). Further, the importance of *the PDX1* gene in human pancreatogenesis was demonstrated by different groups, which reported that homozygous mutation due to single nucleotide deletion on codon 63 of the *PDX1* gene caused pancreatic agenesis (Ahlgren et al. 1996; Wang et al. 2018b). Moreover, a heterozygous mutation in the gene encoding for *Pdx1* is reported to cause T2DM and MODY4 (Oliver-krasinski et al. 2009; Sachdeva et al. 2009; Wang et al. 2018b). Pdx1 is also reported to be indispensable for β-cell expansion in diet-induced insulin-resistant mouse models, and its deficiency makes the β-cells susceptible to ER stress-associated apoptosis (Sachdeva et al. 2009).

Studies conducted in rodent models showed that PDX1 also regulates several other key transcription factors like Ngn3, MafA, FoxA2, Neurod1 and Hnf-1β (Oliver-krasinski et al. 2009). For example, it binds to the *Ngn3* enhancer region and interacts with HNF-6 to activate the enhancer (Oliver-krasinski et al. 2009). It undergoes mutual cooperative binding with NEUROD1 through direct or indirect protein-protein interaction, forming a transcriptional activation complex, which efficiently activates the insulin gene promoter (Glick et al. 2000). PDX1 also participates with FOXA2 in a cooperative binding, facilitating postnatal β-cell maturation (Bastidas-Ponce et al. 2017). Thus, PDX1 cross-talks with other multiple factors during pancreatic development, and postnatally, it plays a special role in β-cell formation. Due to this characteristic, it has always been considered an indispensable factor in cellular reprogramming (Ferber et al. 2000; Ber et al. 2003). This approach of transdifferentiating terminally differentiated somatic cells to β-cells is an upcoming field to treat diabetes and understand its underlying machinery involving the transcription factors. For example, mouse hepatocytes and α-cells have been stated to transdifferentiate into insulin-secreting β-cells by overexpression of the Pdx1 factor only (Ferber et al. 2000). In humans, Zhang and the group demonstrated the reprogramming of ductal PANC-1 cells to functional β-cells by inducing PDX1 overexpression using an andrographolide (Zhang et al. 2020). Apart from these studies,

Pdx1 is also reported to act with different transcription factors like Ngn3, Pax4, MafA, and so forth to transdifferentiate non-β-cells to functional β-cells (Lima et al. 2013). An exciting study has also reported that sequential ON/OFF of Pdx1 with Ngn3 and MafA can drive iPSCs towards β-cell fate (Saxena et al. 2016). Moreover, it is observed that Pdx1, along with Nkx6.1, can enhance reprogramming efficiency to insulin-secreting β-cells (Walczak et al. 2016). Another study observed that mere overexpression of *Pdx1* using the adenoviral method in iPSCs could successfully induce insulin-producing capability within two weeks of transduction (Kim et al. 2020). Notably, PDX1 is one of the pioneer factors that was studied in this prospect, where its overexpression activated insulin production in the liver, endocrine and exocrine cells and ameliorated drug-induced hyperglycemia in mice (Ferber et al. 2000; Zhang et al. 2020).

NGN3

Neurogenin3 (Neurog3 or Ngn3) is one of the crucial transcription factors that has been extensively studied with respect to its contribution towards endocrine-lineage commitment and β-cells specification. NGN3 is encoded by the *NEUROG3* gene, which provides a specific region for binding of other key transcription factors like PDX1 (Oliver-krasinski et al. 2009). It belongs to the family of basic helix-loop-helix and consists of two domains, transactivation domain and basic helix-loop-helix domains, that act as a DNA binding domain. Its expression is first detected during late embryogenesis corresponding to the 7^{th}-8^{th} week post-conception and the expression level peaks around the 10^{th}-14^{th} week with the surfacing of fetal β-cells. However, there is a sharp decline by the 18^{th} week, and its expression is completely lost by the 35^{th}-41^{st} week (Salisbury et al. 2022).

Different studies are carried out to investigate the role of NGN3 in regulating the transcriptional network in β-cell specification. For example, Ngn3 is reported to bind to the E1 and E3 boxes situated in the promoter region of *Neurod1* and activate the early stages of islet

cell differentiation (Huang et al. 2000). Ngn3, along with Hnf1, acts synergistically and binds to the transcription initiation site of the *Pax4* gene (a short sequence located ~2 kb upstream), which in turn, initiates the formation of β-cells and δ-cells (Smith et al. 2003). NGN3 also opposes *SOX9*, which represses ductal fate by promoting endocrine-lineage formation (Shih et al. 2012). Microarray analysis was carried out to understand better the function of Ngn3 in human fetal pancreas development. The Ingenuity Pathway Analysis system revealed 12 pancreatic-specific genes that acted as direct targets for Ngn3 (Lyttle et al. 2008). Of these, 10 genes like *INS, SST, IAPP, MYT1L, NKX2.2, NEUROD1, ISL1, PAX6, MAFB* and Glucagon, were upregulated upon Ngn3 induction, while the other 2 genes, *MNX1* and *SNAI2* are downregulated (Lyttle et al. 2008).

Homozygous mutation in *Ngn3* failed to produce endocrine cells, and the mice died postnatally from diabetes, which suggested that it is indispensable for initiating endocrine progenitor cell formation (Gradwohl et al. 2000). Moreover, the expression of many pancreas-specific factors like Isl1, Pax4, Pax6 and NeuroD1 was also lost (Gradwohl et al. 2000). However, in humans, few β-cells are still formed in the absence of NGN3, suggesting the presence of an Ngn3-independent pathway (Pinney et al. 2011). Overexpression of Ngn3 is associated with reduced endocrine mass (Apelqvist et al. 1999). Ectopic overexpression drives the cells towards premature differentiation, depleting the multipotent progenitor cell population prior to their expansion and reducing overall endocrine cell mass (Apelqvist et al. 1999). These data suggest the tight regulation in the expression profile of Ngn3 during development.

An important question that remained unanswered was whether NGN3 specifically acted only during embryogenesis or had an equally important role in the adult pancreas. To answer this, Xu and co-workers demonstrated that Ngn3 activation is needed to restore islet cells in partial duct ligated adult mouse pancreas. These findings further suggest that multipotent

progenitor remains in the adult pancreas and that, upon injury, differentiates to hormone-producing β-cells (Xu et al. 2008).

Ngn3 is also considered a critical factor in reprogramming studies and has been reported to facilitate the differentiation of iPSCs/adult stem cells to β-cells (Saxena et al. 2016; Gao et al. 2022). Saxena and co-workers verified that Ngn3 and other key factors in a sequential ON/OFF manner could reprogram iPSCs to β-cells (Saxena et al. 2016). These cells secreted insulin, were glucose-sensitive, and expressed mature β-cell markers like MafA, Pdx1, NeuroD1 and Pax4 (Saxena et al. 2016). When adenovirus-mediated gene transfer of *Ngn3* along with *Pdx1* and *Nkx2.2* were carried out in canine adipose-derived mesenchymal stem cells, the reprogrammed cells were driven toward insulin-producing β-cell fate and regulated blood glucose levels (Gao et al. 2022). Furthermore, Ngn3 plays a crucial role in the direct cell reprogramming (transdifferentiation) paradigm. Expression of Ngn3 and other key factors like Pdx1 and MafA have result in transdifferentiation of hepatocytes and exocrine cells to functional β-cells in both human and murine models (Yang et al. 2017; Wang et al. 2018b). In 2009, Yehchoor and group showed that overexpression of only Ngn3 along with betacellulin transdifferentiated liver cells to functional β-cells without other core transcription factors (Yechoor et al. 2009). Thus, Ngn3 proved to be an important candidate in the endocrine-lineage specification and β-cell differentiation during pancreatic organogenesis and plays an important role in cellular reprogramming.

MAFA

MafA or musculoaponeurotic fibrosarcoma oncogene family A is a RIPE3b1 binding transcriptional activator that belongs to the large basic leucine zipper family of transcription factors (Nishizawa et al. 1989). The expression of MafA is exclusively observed in β-cells; however, MafA is also reported to induce insulin gene expression in α-cells (Kataoka et al.

2002; Matsuoka et al. 2004). MafA expression is first detected in mice at E13.5, while in humans, MafA is not detected until the 21st week, and its expression gradually increases postnatally (Zhu et al. 2017). MafA was first identified as a β-cell-specific transcription factor that binds to the insulin gene (C1 element) (Olbrot et al. 2002). Subsequently, it is reported that MafA interacts with NeuroD1 and Pdx1 and binds to Maf-responsive elements (MAREs) through a DNA-consensus sequence (TGCTGAC(G)TCAGCA) and activates insulin gene expression in rat islets (Kataoka et al. 2002; Zhao et al. 2005; Lu et al. 2012). Multiple MARE sites like MARE1/2/3 is also recognized in rat and human Insulin promoter. In the rat, MafA binds to MARE2 and MARE3, while in humans, MafA binds to MARE1 in the Insulin promoter (Nishimura et al. 2015; Liang et al. 2022). Zhang and the group demonstrated that *MafA* double-negative mutant mice develop diabetes and impaired GSIS (Zhang et al., 2005). However, MafA is dispensable in terms of β-cell formation and insulin expression since insulin^{+} cells are present before MafA expression. This indicated that MafA acts downstream in the β-cell transcriptional cascade. However, the deletion of *MafA* results in the formation of a few insulin^{+} cells that causes the onset of DM (Matsuoka et al., 2004; Zhang et al., 2005). Patients with T2DM show compromised MAFA expression and therefore have poor GSIS properties compared to normal islets leading to the diabetic condition (Guo et al., 2013; Liang et al., 2022). Moreover, MafA is recently reported to act on protein phosphatase 1 inhibitory protein 1A, which regulates GLP1R-mediated GSIS (Cataldo et al. 2021). In the case of T1DM patients and non-obese diabetic mice, a mutation in the *MafA* gene is reported to have reduced insulin expression (Noso et al. 2010, 2013).

MafA is an important factor used in transdifferentiation and de-differentiation studies of reprogramming non-β-cells into insulin^{+} cells when used accompanying other regulatory factors. E.g., the transdifferentiation of acinar cells to insulin-producing β-like-cells was facilitated by combining Pdx1 and Ngn3 with MafA via adenoviral vectors (Saxena et al., 2016;

Zhu et al., 2017). MafA is further reported to prevent diabetes in streptozotocin-treated NOD mice by secreting insulin for up to four months (Ackermann et al. 2016).

Growth factors/small molecules like N-acetyl cysteine, AXL inhibitor (R428), thyroid hormone and Alk5 inhibitor II can induce the transcript level of MafA (Rezania et al. 2014; Zhu et al. 2017). Using exome sequencing, a heterozygous mutation produces a variant of the β-cell-enriched MAFA [p.Ser64Phe (S64F)] is identified, which caused familial insulinomatosis and diabetes by impairing phosphorylation within the transactivation domain of MafA (by this, the transactivation potential was increased). This massively increases the stability of MafA protein under the influence of low and high concentrations of glucose in β-cell lines. This property reflected both oncogenic capacity and the key role of MAFA in islet β-cell activity (Iacovazzo et al. 2018). Long noncoding RNA, Meg3, is reported to epigenetically regulate MafA expression by binding to EZH2, which is a methyltransferase belonging to the polycomb repressive complex-2, and inhibiting *Rad21*, *Smc3* or *Sin3α* promoting *MafA* gene expression (Wang et al., 2018). Further work on MafA molecular mechanism can provide a detailed understanding of its contribution towards DM prevention and treatment.

PAX4

Paired box 4 (Pax4) is crucial for islet cell fate determination of β-cells during embryogenesis, and in adulthood, it contributes to β-cell proliferation and maintenance (Brun et al., 2004; Lorenzo et al., 2017). The quantitative real-time (qRT)-Polymerase Chain Reaction (PCR) analysis from whole fetal pancreatic mRNA revealed that Pax4 expression in humans was observed from the 9th week onwards of gestation (Jeon et al. 2009; Conrad et al. 2014). Brun and colleagues demonstrated that adenoviral-mediated Pax4 overexpression increased β-cell proliferation when mitogens such as Activin A and betacellulin were used in rat pancreatic

islets (Brun et al. 2004). On the other hand, the introduction of mutation in *Pax4* caused a gradual decrease in insulin$^+$ cells and finally resulted in DM (Brun et al. 2004). Moreover, some additional attributes of Pax4 in maintaining β-cell survival are to maintain ER homeostasis and immunomodulation (Lorenzo et al. 2017). Pax4 is also reported to function as a repressor as it restricts the expression of both glucagon as well as insulin, although a limited effect of inhibition is observed in β-cell lines, indicating cell type specificity (Smith et al. 1999). Recently, it was reported that complete deletion of the *PAX4* gene in isogenic human models caused the iPSC-derived β-cells to de-repress the α-cell-specific genes (Lau et al., 2022). This Pax4-dependent decrease in insulin is observed in prolonged *in vivo* expression of Pax4, which increases PDX1$^+$ β-cells with very low insulin levels. Furthermore, Pax4 can also inhibit MAFA and GLUT-2 (β-cell markers). This suggests Pax4 promotes an immature state in β-cells (Lorenzo et al. 2017). *Pax4* is a DM-susceptible gene since several mutations (with the main mutation in its two DBDs, the paired-box and homeodomain) and polymorphisms are present in it, resulting in T1DM, T2DM and MODY (Lorenzo et al. 2015, 2017; Martin-Montalvo et al. 2017). A missense mutation in the *Pax4* gene (R121W-a single amino acid substitution in the paired domain) is associated with T2DM in the Japanese population, a rare form of T2DM, which is ketosis-prone diabetes caused due to mutations (R133W or R37W) of *Pax4* (Sosa-Pineda 2004). In some European populations, mutation of *Pax4* corresponds to T1DM. Additionally, genome-wide association studies in Asian populations for polymorphism associated with T2DM have shown that two single nucleotide polymorphisms exist in the intergenic regions close to Pax4, which corresponds to T2DM (Cho et al. 2012; Ma et al. 2013; Lorenzo et al. 2017). The etiology of the disease in any of these mutations is the same with β-cell dysfunction and aberrant insulin secretion in response to glucose, which is a consequence of loss of functional insulin-producing β-cell mass in islets (Sosa-Pineda 2004; Lorenzo et al. 2015). Recently, Lau and the group reported that even an altered coding variant p.Arg192His

in the *PAX4* gene, is associated with a higher risk of T2DM due to decreased β-cell function in the East Asian population (Lau et al., 2022). Ectopic expression of Pax4 in αTC1.9 cells results in increased insulin expression and the corresponding decrease in glucagon expression (Zhang et al., 2016). These cells also show GSIS, thus, providing evidence of Pax4-induced α- to functional β-cell transdifferentiation (Zhang et al., 2016), which is valuable in producing potential therapeutic for diabetes treatment. Collombat and group overexpressed only Pax4 in mouse pancreatic progenitor and α-cells, which were initially insulin$^-$ cells. Due to this overexpression, these cells started behaving like β-cells and producing insulin (Collombat et al. 2009). This finding additionally supports the major role played by Pax4 in transdifferentiation. Pax4 and added factors like Pbx1, Pdx1, Rfx3, Ngn3 and MafA have recently reprogrammed adipose mesenchymal stem cells into β-like-cells (Dai et al. 2022). The study also demonstrated reprogrammed cells could relieve diabetes in model dogs (Dai et al. 2022). Moreover, misexpression of Pax4 in gastrointestinal δ-cells can turn the cells into insulin-secreting cells (Garrido-Utrilla et al. 2022). This further proves the single-handed nature of Pax4 in the cellular reprogramming paradigm.

1.2.6 Recombinant proteins

The field of recombinant protein production has attained widespread popularity due to the enormous applications of recombinant proteins in therapeutics, diagnostics or understanding basic biological questions. These recombinant proteins can be produced *ex-vivo* from diverse biological systems such as bacteria, yeast, insect cell, and mammalian cell systems (Assenberg et al. 2013; Sanchez-Garcia et al. 2016). The mammalian expression system is a suitable choice for proteins that require appropriate post-translational modifications for their function. For other proteins, the high production time, costs, slow growth kinetics, and complex nutritional requirements associated with mammalian systems render them less preferred than bacterial

systems (Walsh and Jefferis 2006). Therefore, *E. coli* became an obvious choice for large-scale production of such proteins due to their well-established genetic makeup and knowledge of its physiology, reasonable costs and unparalleled growth rate (Sahdev et al. 2008; Khow and Suntrarachun 2012). Post-expression, these recombinant proteins are purified using a wide range of purification tags. Poly-histidine tags are the most widely used tag for purification since they are inexpensive and do not alter the proteins' characteristics (Young et al. 2012; Wood 2014).

The production of recombinant proteins for therapeutic applications or protein-based cell reprogramming is considered to be a safe and promising alternative compared to viral and gene delivery-based approaches (Sanchez-Garcia et al. 2016; Serna et al. 2018; Borgohain et al. 2019). For these applications, the protein of interest is delivered into a target cell to exert its biological function. Notably, this approach does not manipulate the target cell genome, making it ideal for clinical applications or deriving integration-free, clinical-grade human cells (Borgohain et al. 2019). Additionally, this approach provides complete control over the dosage and time of application of a protein of interest to elucidate its specific role in the cellular reprogramming process (Borgohain et al. 2019). However, various impediments are associated with this heterologous expression, namely codon usage bias, gene product toxicity, mRNA stability, protein degradation, protein insolubility, protein misfolding, and protein inactivity (Rosano and Ceccarelli 2014). All of these mentioned bottlenecks can hugely affect the overall yield and biological activity of the purified proteins. Thus, in our present work, we worked on addressing these limitations and aimed to generate a cell- and nuclear-permeant bioactive human recombinant protein toolbox comprising PDX1, NGN3, PAX4 and MAFA proteins, which can prospectively be used for β-cell reprogramming and various other biological applications.

1.3 Scope and significance of the study

To date, the ever-growing diabetic community relies on temporary solutions like administering insulin shots or anti-diabetic drugs to combat high blood glucose levels. The current need for a safe and permanent solution to the diabetic condition has motivated us to carry out the present work. In this perspective, direct reprogramming is a straightforward and powerful technique that can provide an unlimited supply of cells by transdifferentiating terminally differentiated cells towards the desired cell type. Multiple groups have extensively used this approach to reprogram non-β-cells towards insulin-producing β-cells. The β-cell identity has been achieved by various studies via ectopic expression of one or more pancreatic-specific transcription factors in somatic cells, bypassing the pluripotent state.

In early 2000, Ferber and co-workers were the first group to demonstrate the induction of only one pancreatic-specific transcription factor, Pdx1, can bestow β-cell features to liver cells. The treated cells expressed insulin, which is exclusive to β-cells and, once transplanted in non-obese diabetes mice, ameliorated hyperglycemia. From then onwards, multiple groups have demonstrated the key role played by the different transcription factors, single-handedly or in combination to confer β-cell attributes to different terminally differentiated cells like α-cells, exocrine cells, cells of intestinal origin, etc. Among the varied range of pancreatic-specific transcription factors, the most widely used cocktail of factors consists of PDX1, NGN3, MAFA and PAX4. However, most of these studies applied integrative approaches for reprogramming that marred the clinical applicability of the reprogrammed cells. This limitation has broadened the scope of developing a non-integrative-based approach to derive integration-free cells in β-cell reprogramming. The application of transcription factors in the form of recombinant proteins can address this shortcoming. This further motivated us to design a simple, cost-effective and systematic strategy to generate the reprogramming factors in the

form of recombinant proteins that can further potentially replace their genetic and viral counterparts.

We aim to generate a biologically active recombinant protein toolbox comprising the key and widely used pancreatic-specific transcription factors that will aid in direct reprogramming of differentiated somatic cells to insulin-producing β-cells. The purified proteins generated in this study maintained their secondary structure and retained their bioactivity. Furthermore, these recombinant proteins can be directly applied to the cells of interest and can successfully translocate to the mammalian cell nucleus. In addition, when subjected to different cell lines, these proteins brought about various changes. In the near future, this toolbox encompassing the pivotal factors will open the way for the generation of integration-free autologous patient-specific functional β-cells. The reprogrammed cells generated through this approach can be directly transplanted into diabetic patients and can treat diabetes permanently.

Based on these findings and necessities, we have designed the following objectives.

1.4 Objectives

1. Gene cloning and optimization of expression parameters to achieve maximal soluble expression of pancreatic-specific transcription factors
2. Purification and biophysical analysis of pancreatic-specific transcription factors
3. Demonstration of cell penetration, nuclear localization and biological activity of pancreatic-specific transcription factors

Chapter 2

Gene cloning and optimization of expression parameters to achieve maximal soluble expression of pancreatic-specific transcription factors

A brief overview of the chapter

The field of recombinant protein production has attained widespread popularity due to the enormous applications of recombinant proteins in therapeutics, diagnostics, or understanding basic biological questions. Several factors, like choice of expression systems, codon optimization, optimal expression parameters, etc., hamper efficient production and are needed to be addressed to increase the expression and yield of the protein while retaining its biological activity. Here, we retrieved the full-length protein-coding sequence from NCBI's RefSeq database and performed codon optimization of the selected pancreatic-specific transcription factors to replace the rare codons with codons that are more abundant in the genes of the host organism (here *E. coli*). Further, we added fusion tags consisting of 8x histidine residues (His or H) for affinity purification, Transactivator of transcription (TAT) for cellular penetration and nuclear localization signal (NLS) for nuclear translocation to 5'- and 3'-terminal of the protein-coding DNA sequence. Next, the constructs were cloned into the pET28a(+) vector and transformed and expressed in the BL21(DE3) strain of *E.coli*. Post-cloning, we screened the optimal expression parameters for each of these factors for maximal soluble expression. Based on our findings, we selected N-terminal tagged PDX1, C-terminal tagged NGN3, C-terminal tagged MAFA and both C- and N-terminal tagged PAX4. Interestingly, our results signified that the optimal expression conditions for each transcription factor vary based on the nature of the transcription factor. We thus conclude the importance of identifying the optimal expression conditions and influence of fusion tags at either end of the coding sequence.

2.1 Materials and Methods

2.1.1 Plasmid constructs

The full-length protein-coding sequence of the selected four transcription factors was obtained from the National Center for Biotechnology Information (NCBI's) reference sequence (RefSeq) database (Table 2.1). Next, to remove codon biasness, these full-length protein-coding sequences were codon-optimized using ThermoFisher Scientific GeneOptimizer online tool (https://www.thermofisher.com/in/en/home/lifescience/cloning/genesynthesis/geneoptimizer.html). Then the codon-optimized sequences were analyzed using two independent online tools, namely Graphical Codon Usage Analyser Tool (GCUAT) 2.0 (https://gcua.schoedl.de/) and Genscript Rare Codon Analysis Tool (GRCAT) (https://www.genscript.com/tools/rare-codon-analysis) online softwares (Figure 2.1). Thereafter, this codon-optimized gene sequence was tagged either before the start codon or before the stop codon with codon-optimized fusion tags [polyhistidine-tag (octahistidine; (H)), Transactivator of Transcription (TAT; (T)), and nuclear localization sequence (NLS; (N))] to create HTN-*GOI* and *GOI*-NTH, respectively. These fusion tags were also flanked with restriction sites for their easy removal. These customized gene inserts (HTN-*GOI* and *GOI*-NTH) were purchased from Genscript Biotech Corporation, China and were next cloned into the pET28a(+) protein expression vector (Novagen) to generate pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH constructs. The resulting pET28a(+) vectors harboring fusion genes were evaluated using restriction digestion and DNA sequencing (Eurofins Genomics India Pvt. Ltd. Bengaluru, Karnataka, India) using standard T7 promoter (5′- TAATACGACTCACTATAGGG-3′) and terminator (5′-GCTAGTTATTGCTCAGCGG-3′) primers.

Table 2.1 List of genes and their respective RefSeq accession numbers used in this study.

Gene Name	RefSeq Accession number	Length of coding sequence (in bp)
PDX1	NM_000209.3	1035
NGN3	NM_020999.3	828
MAFA	NM_201589.3	1245
PAX4	NM_002093.2	1236

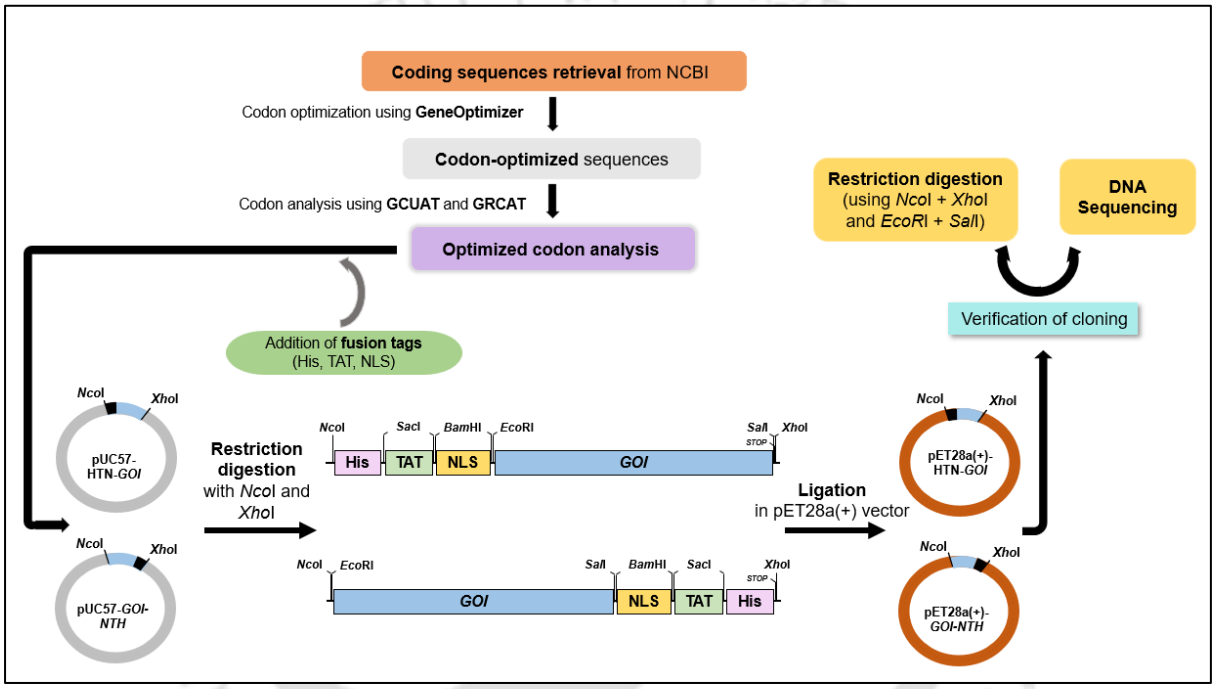


Figure 2.1 Schematic showing the cloning strategy employed in this study. Full-length protein-coding sequences of the transcription factors were retrieved and optimized using GeneOptimizer before cloning in pET28a(+) protein expression vectors using *Nco*I and *Xho*I restriction sites. Cloning was confirmed using restriction digestion and DNA sequencing. GCUAT: Graphical Codon Usage Analyzer Tool; GRCAT: Genscript Rare Codon Analyzer Tool.

2.1.2 Identification of optimal induction parameters for maximal soluble expression

Codon optimization is not the only criterion to obtain efficient heterologous expression of human proteins in *E. coli*. Numerous studies have demonstrated that the identification of optimal expression parameters is also critical to achieving maximal expression of biologically active recombinant proteins from *E. coli* in a soluble form (Ryan and Henehan 2013; Liu et al. 2018). Therefore, to prevent protein aggregation and achieve maximal soluble expression of fusion proteins in *E. coli*, the genetic constructs were transformed into *E. coli* and screened for inducer concentration (Isopropyl β-D-1-thiogalactopyranoside (IPTG)), induction cell density, post-induction incubation time and placement of fusion tags at the N- or C-terminal ends of the protein of interest.

Inducer concentration

In order to identify the optimal inducer concentration, *E. coli* BL21(DE3) was transformed with recombinant plasmids and a single bacterial colony was grown overnight in Luria-Bertini (LB) broth supplemented with 50 µg/ml kanamycin (HiMedia). Overnight grown culture was used as primary inoculum for secondary cultures, which were grown until the OD_{600} reached ~0.5 and then induced with different concentrations (0.05, 0.1, 0.25, and 0.5 mM) of IPTG (Sisco Research Laboratories) for 2 h at 37 ˚C with constant shaking at 180 rpm (Orbital incubator shaker, IKON instruments, India). Post incubation, the cells were harvested at 10000 rpm, at 4 ˚C for 5 minutes and cell pellets were stored at -80 ˚C for future use.

Optical cell density

For optimizing cell density at the time of induction, secondary cultures were induced with optimal IPTG concentration at different OD_{600} (~0.5, ~1.0, and ~1.5) and incubated for 2 h at shaking conditions at 37 ˚C. Post incubation, the cells were harvested at 10000 rpm, at 4 ˚C for 5 minutes and cell pellets were stored at -80 ˚C for future use.

Induction time

To identify the optimal post-induction incubation time, the cultures were induced at the optimal cell density with the optimal inducer concentration at 37 ˚C with constant shaking for 8 h, and samples were collected at 2, 4, and 8 h for analysis. Post incubation, the cells were harvested at 10000 rpm, at 4 ˚C for 5 minutes and cell pellets were stored at -80 ˚C for future use.

2.1.3 Ultrasonication

The harvested cell pellets were resuspended using a pre-chilled resuspension buffer which constitutes the total cell lysate (Table 2.2). The lysate was next sonicated until clear (4 seconds ON, 26 seconds OFF, 33% amplitude) and subjected to centrifugation at 8000 rpm, 4 ˚C for 30 minutes to separate the soluble fraction from the insoluble pellet fraction. Post-centrifugation, the supernatant was carefully separated and incubated overnight with nickel-nitriloacetic acid (Ni-NTA) resin at 4 ˚C at shaking conditions.

Table 2.2 Summary of the composition of the resuspension buffer used in this study.

Ingredients	Resuspension buffer			
	PDX1	NGN3	MAFA	PAX4
PB (mM)	20	20	20	20
NaCl (mM)	150	300	300	600
Imidazole (mM)	20	20	20	20
Glycerol (%)	20	20	20	20
pH (at RT)	8.0	8.5	7.0	8.5

PB, phosphate buffer; RT, room temperature

2.1.4 Sodium Dodecyl Sulphate-Polyacrylamide Gel electrophoresis (SDS-PAGE) and western blotting

The total protein concentrations were determined using Bradford assay [Bradford reagent (Bio-Rad)] (Bradford 1976) using bovine serum albumin (BioRad) as a standard and measured with a multi-plate reader (Multiskan GO, Thermo Scientific). Sonicated protein samples were diluted using 4X Laemmli buffer and 10% SDS to a final concentration of 1X and 2%, respectively. The protein samples were boiled at 95 ˚C for 10 minutes and resolved in 12% SDS-PAGE gel along with a protein ladder (Precision Plus Protein™ Dual Color (Bio-Rad)) to assess the molecular weight of the protein of interest.

For Coomassie staining, the SDS-PAGE gels were first fixed with a fixing solution containing 50% ethanol (v/v) and 10% acetic acid (v/v) (Merck Millipore) in deionized water for 1 h and then washed overnight using wash buffer containing 50% methanol (v/v) (Merck Millipore) and 10% acetic acid (v/v) in deionized water. The next day, the gels were stained with staining solution [50% methanol (v/v),10% acetic acid (v/v) and 0.4% (w/v) Coomassie brilliant blue G-250 (Merck Millipore)] for 20 minutes. The gels were destained using wash buffer and imaged in the molecular imager (ChemiDoc™ XRS+) equipped with Image Lab™ software (Bio-Rad, California, USA).

For Western blotting, the samples were resolved by SDS-PAGE gel and transferred to a nitrocellulose membrane (Bio-Rad) by transfer buffer containing 20% methanol (v/v) in 1X transfer buffer using Pierce Power Blotter XL System (Thermo Scientific™, Bremen, Germany) at constant voltage (20 V) for 25 minutes. Post-transfer, the membranes were stained with Ponceau S staining (HiMedia) for protein visualization. The stain was washed off using 1X Tris-buffered saline (TBS) and blocked with 5% (w/v) fat-free skimmed milk containing 0.1% Tween-20 (v/v) (Invitrogen) in 1X TBS (TBST) for 1 h to limit non-specific binding. Next, the blocking solution was washed for 10 minutes with 1X TBST and the membrane was incubated with primary antibody (anti-His antibody (1:5000; BioBharati; BB-AB0010))

overnight at shaking condition at 4 ˚C. Post incubation, the membranes were washed thrice with 1X TBST, 5 minutes incubation each in shaking condition and further incubated with horseradish Peroxide-conjugated secondary antibody (anti-Rabbit antibody (1:5000; BioBharati; Cat # BB-SAB01C)) for 1 h in shaking condition at room temperature. All primary [were diluted with 5% (w/v) bovine serum albumin (HiMedia) in TBST] and secondary [were diluted with 5% (w/v) fat-free milk in TBST] antibodies used, and their respective dilutions are listed in Appendix 1. Finally, Western blots were developed in the presence of a chemiluminescence substrate (Bio-Rad) and were captured and documented by the molecular imager (ChemiDoc™ XRS+) installed with Image Lab™ software (Bio-Rad).

2.1.5 Agarose gel electrophoresis

The DNA samples were mixed with 6X loading buffer (Thermo Scientific) and then resolved on 0.8-1% agarose gel, prepared in 1X Tris-EDTA buffer (HiMedia). The resolved agarose gel was visualized under ultraviolet light, and the image was recorded in the molecular imager.

2.2 Results and discussion

2.2.1 Codon optimization of pancreatic-specific human transcription factors for expression in *E. coli*

Heterologous expression of human genes in *E. coli* presents multiple constraints that lead to compromised overall protein expression (Burgess-Brown et al. 2008; Maertens et al. 2010). Hence, to achieve high heterologous expression of fusion proteins in *E. coli*, the full-length protein-coding gene sequences were codon-optimized using GeneOptimizer (Thermo Fisher Scientific) online tool. This tool substitutes codons present in a non-optimized sequence with the most preferred codons (based on the selection of the desired organism) without modifying the protein sequence. Further, codon-optimized and non-optimized sequences were

validated using GenScript Rare Codon Analysis Tool (GRCAT; Figure 2.2) and Graphical Codon Usage Analyser Tool 2.0 (GCUAT; Figure 2.3) online softwares.

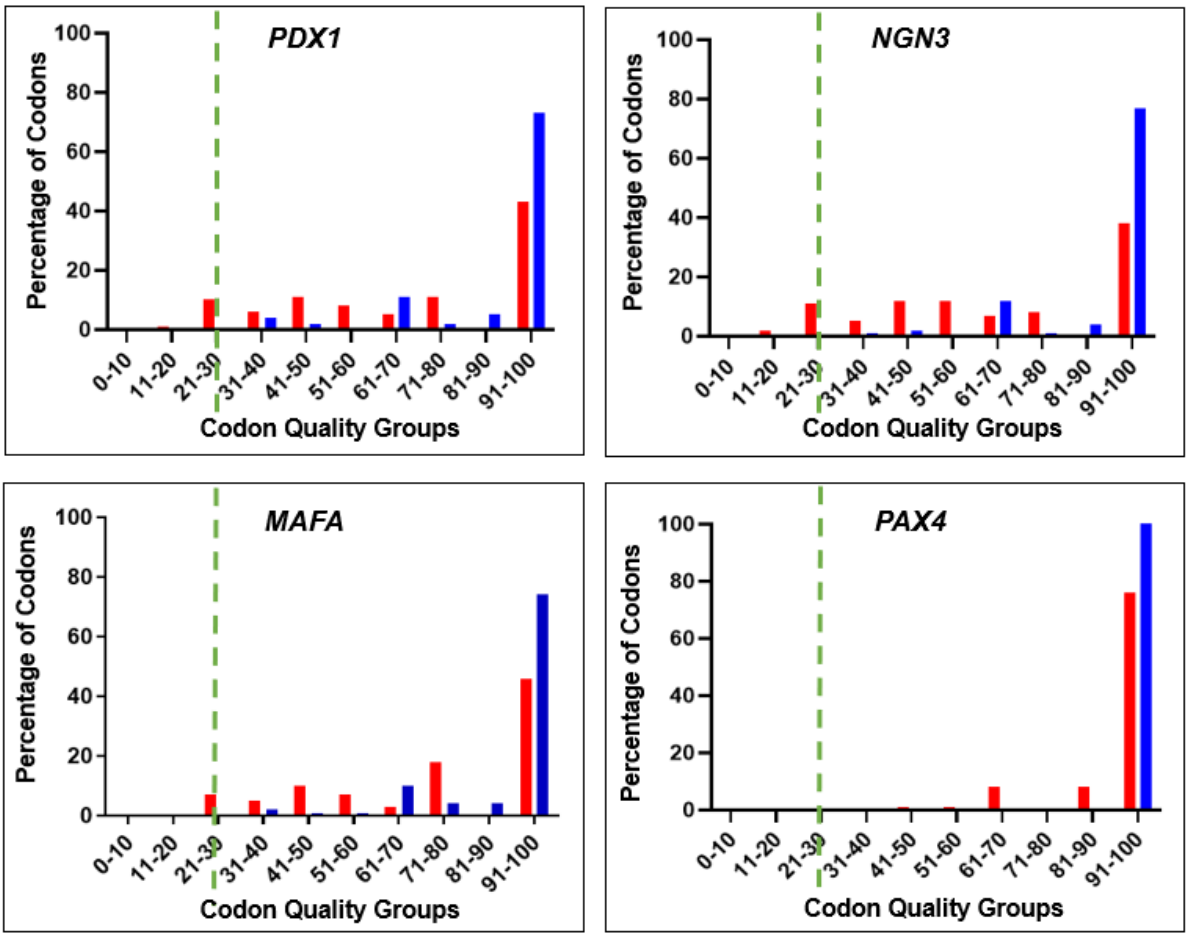


Figure 2.2 Evaluation of codon-optimized over non-optimized gene sequences of pancreatic-specific transcription factors using the GRCAT. The graph shows the percentage distribution of codons that belong to a specific quality group. Codons with the most frequently used codon for a given amino acid in *E. coli* are assigned a value of 100. The codons distributed in codon quality groups with a value ≤ 30 are used less frequently by that organism (*threshold is shown as a dotted line*) and are more likely to impede the expression. The presence of red bars (i.e., non-optimized) with values ≤ 30 in the graph signified that it had codons that could affect its expression in *E. coli*. The absence of blue bars (i.e., codon-optimized) with

values ≤ 30 in the graph signified that it lacked codons that could affect its expression in *E. coli*.

Using the GRCAT, it was found that 7-14% of the codons in the non-optimized sequence had a codon usage frequency of ≤30%, which could hamper the expression of these genes in *E. coli* (Figure 2.2). Using the GCUAT, multiple codons have been found to have a relative adaptiveness value of ≤30% in the non-optimized sequence that might affect the expression of human *PDX1* [Figure 2.3 (shown in magenta)] and other human genes of this study (*NGN3*, *MAFA* and *PAX4*) in *E. coli*. These codons and other codons that might impact the expression were replaced with the most favored synonymous codons using the GeneOptimizer tool to enhance gene expression (Figure 2.2).

In addition, an improved codon adaptation index was observed in codon-optimized over a non-optimized sequence (Table 2.3). The codon adaptation index is ideally between 0.8 and 1.0 for efficient gene expression in the desired organism, whereas an index less than 0.8 will result in inefficient gene expression. Remarkably, the codon adaptation index, a parameter generally attributed to translational efficiency, increased for codon-optimized gene sequences in all the four studied factors (Table 2.3). This implied that the codon optimization of *PDX1, NGN3, MAFA and PAX4* gene sequences would favour its expression in *E. coli*. This crucial step removed all the undesirable elements to increase the expression of desired proteins in *E. coli*.

Codon optimization of human genes for expression in *E. coli* is crucial as codon bias is considered as one of the critical barriers in heterologous protein expression (Lithwick and Margalit 2003; Gustafsson et al. 2004; Liu et al. 2018). This is because of the variable preference of codons for a particular amino acid in different organisms. Codon bias leads to inefficient translation due to the limited availability of specific tRNAs in the host organism for the codons present in the sequence of the gene of interest. The concern with codon bias can be

alleviated by codon optimization or co-expressing genes that encode the rare tRNAs in *E. coli* (Burgess-Brown et al. 2008).

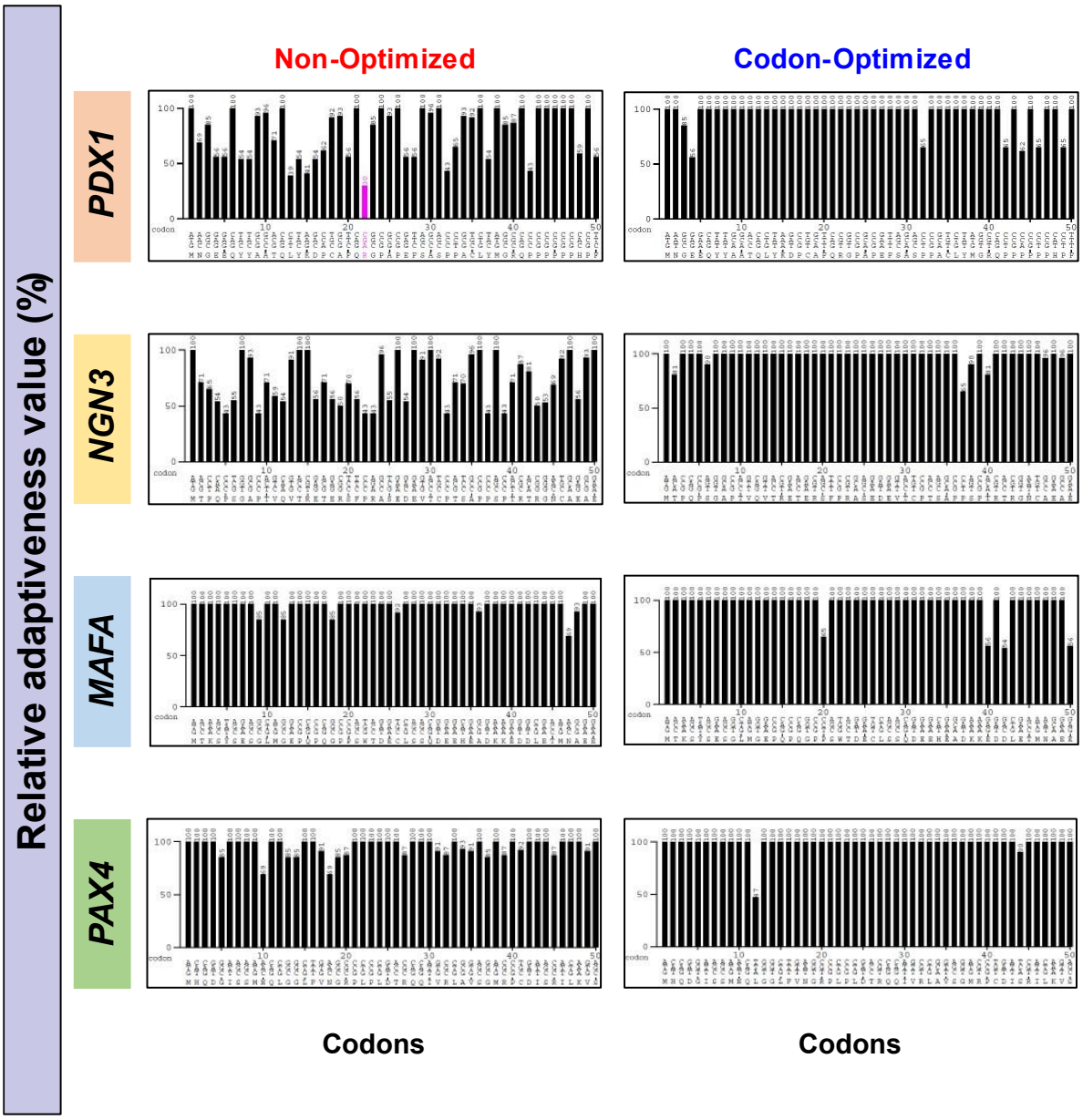


Figure 2.3 Evaluation of codon-optimized over non-optimized gene sequences of pancreatic-specific transcription factors using the GCUAT. The graph shows the codons of non-optimized and codon-optimized coding sequences against the relative adaptiveness value. Codons with a relative adaptiveness value ≤ 30% (shown by the arrow in magenta) are

very likely to impede the expression in *E. coli*. The absence of magenta bars in the codon-optimized graph shows that it was devoid of inhibitory codons.

Overall, the codon optimization strategy had a more positive impact compared to the tRNA overexpression strategy on heterologous (human) gene expression in *E. coli* (Burgess-Brown et al. 2008; Maertens et al. 2010). Two multi-gene studies have demonstrated that codon optimization of human genes improved protein expression and production in *E. coli* (Gustafsson et al. 2004; Burgess-Brown et al. 2008). This is mainly due to the elimination of codon bias, destabilizing RNA elements, RNA secondary structures, intragenic poly(A)-sites, cryptic splice-sites, direct repeats, high GC content, and so forth (Maertens et al. 2010; Liu et al. 2018). Therefore, codon optimization was performed to improve gene expression and translational efficiency.

Table 2.3 Parametric GenScript Analysis summary for non-optimized and codon-optimized sequences

Genes	Parameter	Non-Optimized	Codon-Optimized	Ideal range
PDX1	Codon Adaption Index (CAI)	0.64	0.87	0.8-1.0
	GC content	71.22%	55.99%	30%-70%
	Codon Frequency Distribution (CFD)	11%	0%	<30%
NGN3	Codon Adaption Index (CAI)	0.61	0.91	0.8-1.0
	GC content	69.24%	50.09%	30%-70%
	Codon Frequency Distribution (CFD)	13%	0%	<30%
MAFA	Codon Adaption Index (CAI)	0.70	0.89	0.8-1.0
	GC content	73.73%	54.25%	30%-70%
	Codon Frequency Distribution (CFD)	7%	0%	<30%
PAX4	Codon Adaption Index (CAI)	0.59	0.91	0.8-1.0
	GC content	59.66%	54.17%	30%-70%

Codon Frequency Distribution (CFD)	14%	0%	<30%

2.2.2 Cloning of codon-optimized transcription factors with fusion tags into a protein expression vector

The codon-optimized coding sequences of the transcription factors were fused to a set of tags, namely, the Octa-His (H) tag to enable affinity chromatography-based purification, a protein transduction domain TAT (T) for intracellular, and NLS (N) for intranuclear delivery. All three tags were either placed before the start codon to generate HTN-*GOI* or before the stop codon to generate *GOI*-NTH (Figure 2.4). The customized gene inserts (HTN-*GOI* and *GOI*-NTH) obtained in pUC57 plasmid were excised out using restriction endonucleases, *Nco*I and *Xho*I, and cloned into the pET28a(+) protein expression vector between *Nco*I and *Xho*I restriction sites. This gene was placed under the transcriptional control of a tightly regulated strong T7 promoter. The obtained plasmids pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH were confirmed using restriction digestion analysis (Figure 2.5). The fidelity of the cloned gene sequence was confirmed with DNA sequencing using standard T7 promoter [#69348-3 (Novagen); 5′-TAATACGACTCACTATAGGG-3’] and terminator [#69337-3 (Novagen); 5′-GCTAGTTATTGCTCAGCGG-3’] primers.

The influence of the position of these tags on protein expression was determined in this study. This strategy enabled the identification and selection of the gene construct with fusion tags that had no adverse effect on expression in *E. coli*. In addition, the fusion of tags such as TAT and NLS can serve as interesting molecular tools to deliver a protein of interest at the target site in mammalian cells. Earlier studies have asserted the importance of the position of tags which can influence the expression, solubility and stability of proteins expressed in *E.*

coli (Braun et al. 2002; Bosnali and Edenhofer 2008; Haridhasapavalan et al. 2020). A similar strategy like ours was employed by various groups to facilitate efficient delivery of the transcription factors, namely NANOG, SOX2, and OCT4, to their target site into mammalian cells (Bosnali and Edenhofer 2008; Thier et al. 2011, 2012; Peitz et al. 2014; Münst et al. 2016).

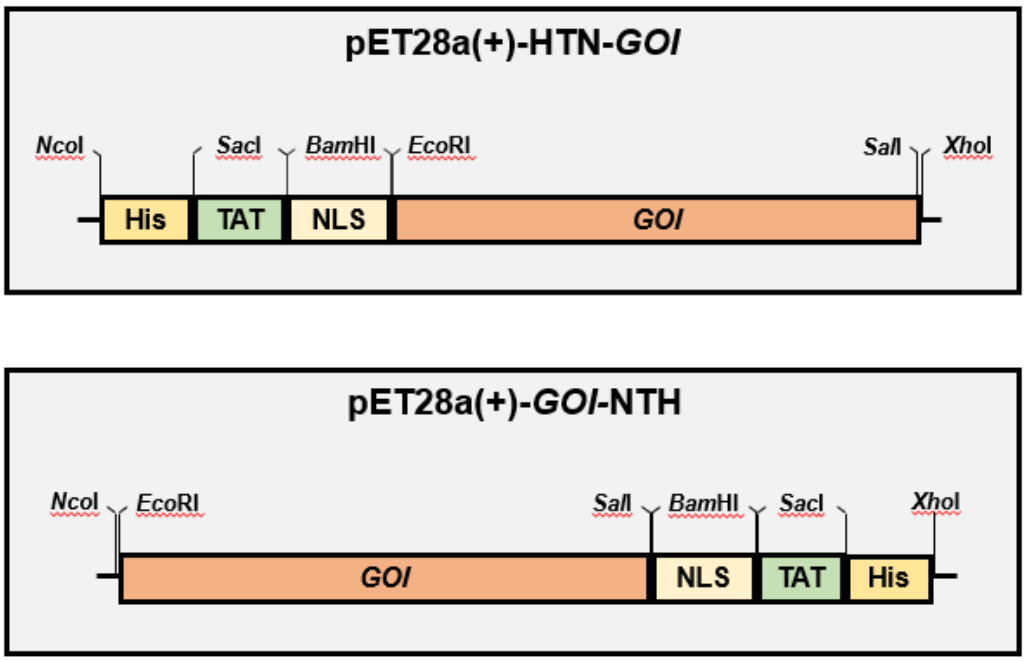


Figure 2.4 Schematic illustrations of fusion gene inserts [HTN-*GOI* (above) and *GOI*-NTH (below)] (not drawn to scale). *GOIs* were fused to His-tag for affinity chromatography, TAT to enable cell penetration, and NLS to facilitate nuclear translocation in mammalian cells. His (H): Histidine (8X); TAT (T): transactivator of transcription; NLS (N): nuclear localization sequence.

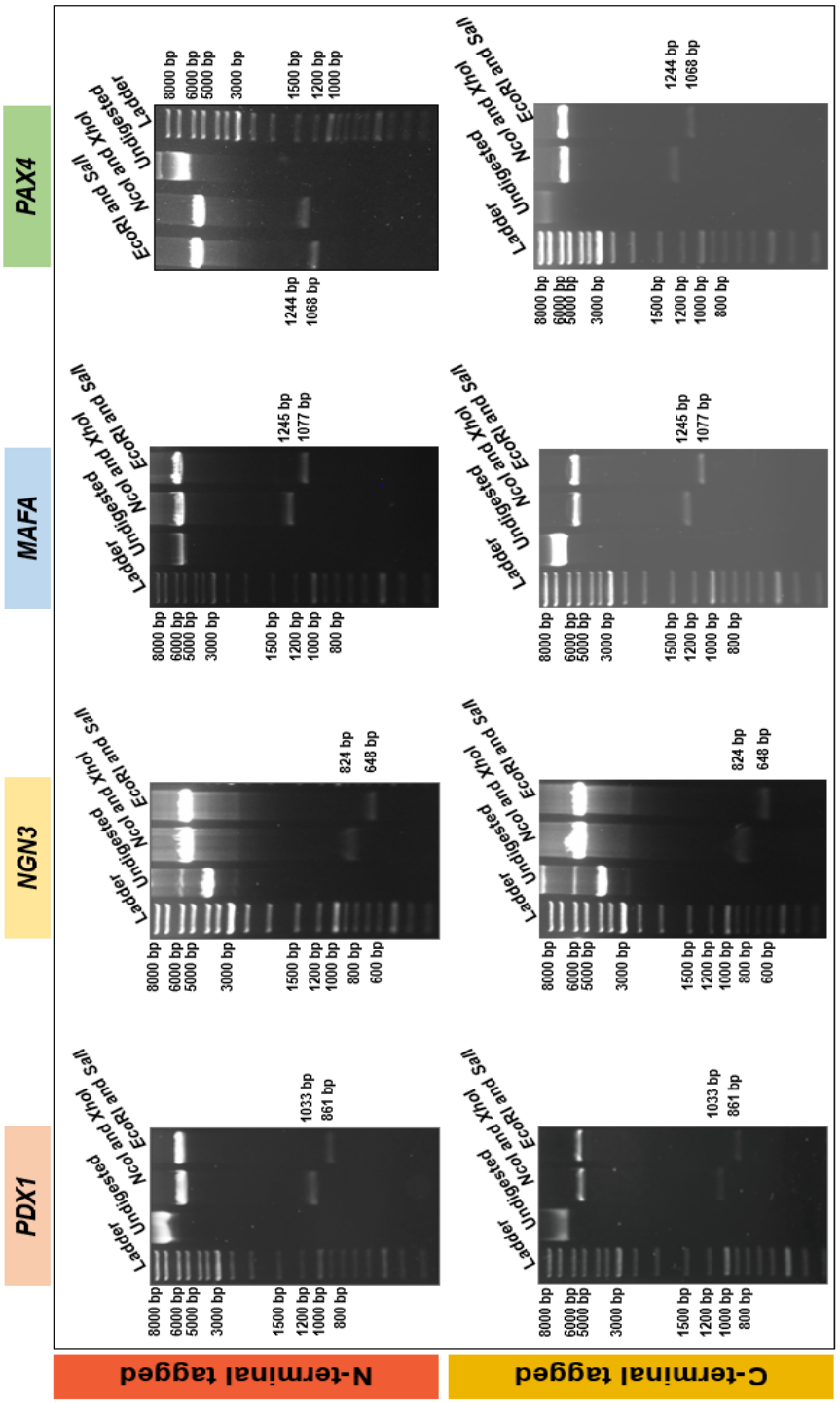


Figure 2.5 Confirmation of cloning of the gene inserts into the protein expression vector. Restriction digestion analysis was performed to confirm the successful cloning of HTN-*GOI* and *GOI*-NTH inserts into the protein expression vector. The synthetic gene inserts were cloned in the protein expression vector (pET28a(+)) using restriction endonucleases, *Nco*I, and *Xho*I. The resulting plasmids, pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH, were then confirmed by restriction digestion using various restriction enzymes. NLS (N): nuclear localization signal/sequence; TAT (T): transactivator of transcription; His tag (H): poly-histidine (8X).

2.2.3 Screening of genetic constructs and expression conditions for maximal protein production in soluble form from *E. coli*

Producing soluble proteins in *E. coli* is still a major bottleneck to obtaining bioactive proteins with high yield. Various reports have mentioned the importance of optimal recombinant protein expression parameters for maximal expression and production of bioactive proteins from *E. coli* (San-Miguel et al. 2013; Liu et al. 2018). Among the optimized parameters, inducer concentration, in this case, IPTG, was first screened for all the generated genetic constructs in *E. coli*. Interestingly, we observed that a minimum of 0.05 mM of IPTG was sufficient for maximal protein expression in the case of HTN-NGN3 and NGN3-NTH. The expression profile remained unaltered even after increasing the inducer concentration up to 0.5 mM. HTN-PDX1 and PDX1-NTH required a higher concentration of inducer for maximal expression; thus, 0.1 mM of IPTG was used. However, recombinant MAFA and PAX4 fusion proteins required an even higher concentration of inducer, and so, in these cases, 0.25 mM IPTG was used. Remarkably, we observed that different factors required varied inducer concentrations for optimal expression in a heterologous system (Table 2.4).

The next parameter that was screened was optimal bacterial density because bacterial gene expression depends not only on specific regulatory mechanisms but also on bacterial growth because important global parameters such as the abundance of RNA polymerases and ribosomes are all growth-rate dependent. Understanding these global effects is necessary to quantitatively understand gene regulation (Klumpp et al. 2009). Studies have reported that bacterial system overexpresses recombinant proteins when induced at log phase, i.e., ~0.3-0.8 (García-Fraga et al. 2015), and for this reason, the recombinant fusion proteins were induced at various bacterial densities (~0.5, ~1.0 and ~1.5 OD_{600}). The maximal expression study was carried out using immunoblotting and SDS-PAGE. In line with the previous findings by different groups, our results demonstrated the maximal expression for each of these recombinant proteins was when induced at the log phase of OD_{600} of ~0.5. Moreover, induction

of bacterial culture at late log phase (~1.0-1.5 OD_{600}) did not affect/alter protein expression level for all four factors.

Next, the incubation time post-induction was screened for all the recombinant fusion proteins. The cultures were grown to OD_{600} of ~0.5 and induced with screened inducer concentration. Next, induced cultures were incubated at 37 °C for various time points. Collected samples were analyzed using SDS-PAGE and immunoblotting. The results demonstrated that 2 h incubation was sufficient for maximal protein expression for all the constructs. However, only HTN-PDX1 required 4 h for its maximal expression. As a result, a 4 h incubation time was used for HTN-PDX1 for further studies.

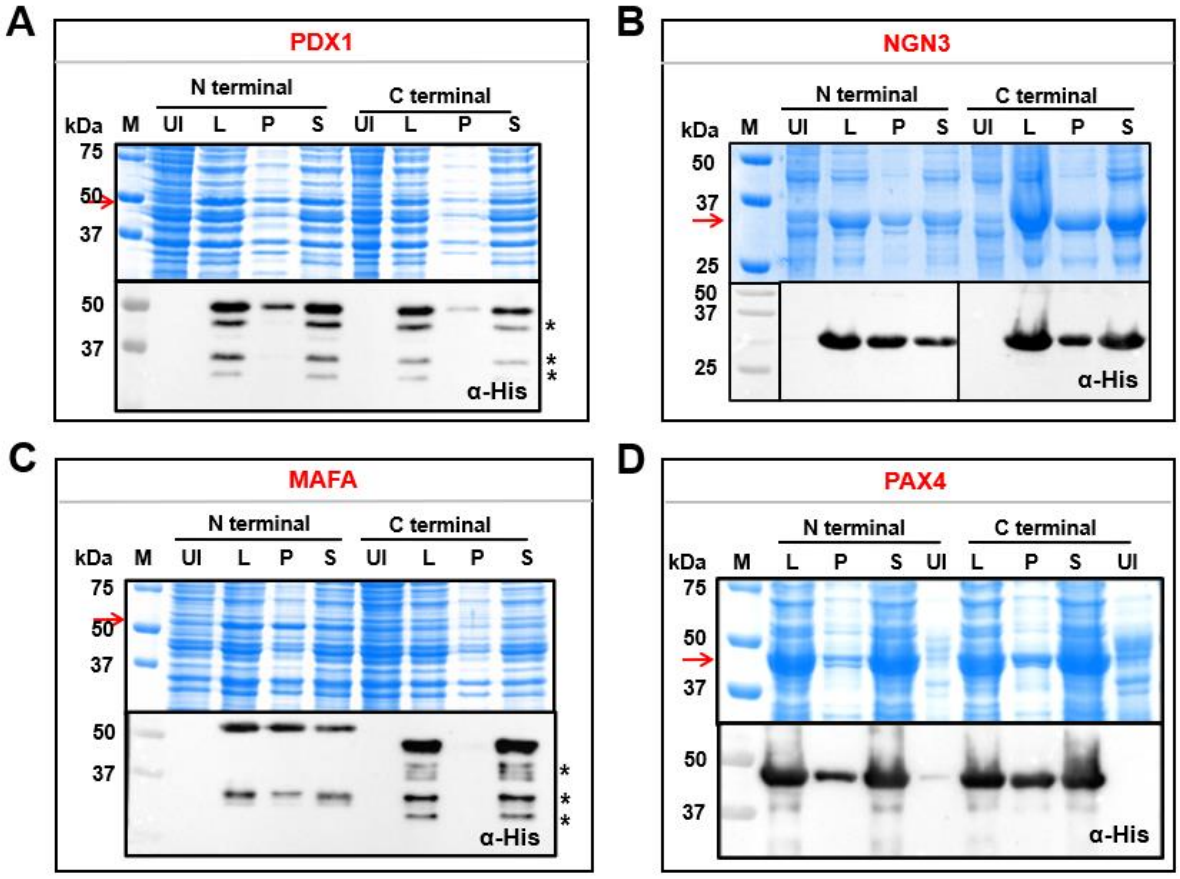


Figure 2.6 Identification of optimal induction parameters (genetic construct and temperature) to achieve maximal soluble expression of the recombinant human fusion proteins in soluble form. *E. coli* BL21(DE3) strain was transformed with pET28a(+) vectors

harbouring the fusion gene inserts and the expression of recombinant proteins was induced at optimal cell density with an optimal concentration of IPTG at 37 °C. Subsequently, the cells were harvested and lysed to obtain the total cell lysate (L) fraction and were then centrifuged to obtain a soluble/supernatant (S) cell fraction and insoluble/pellet (P). Protein samples were normalized to 20 µg/lane for L fractions, and equal volumes for P and S fractions corresponding to the respective L fraction were used. The protein samples were separated on 12% SDS-PAGE gels and either stained using Coomassie brilliant blue (top) or performed Western blotting using histidine antibody (bottom) (A) PDX1 fusion proteins, (B) NGN3 fusion proteins, (C) MAFA fusion proteins, (D) PAX4 fusion proteins. M, Marker; UI, uninduced (total cell lysate); L, cell lysate; P, pellet/insoluble cell fraction; S, supernatant/soluble fraction; kDa, kilodaltons; α-His, anti-histidine antibody; n=4. *Truncations of the fusion protein.

Using the identified parameters, we compared the expression and solubility profiles of two genetic constructs, HTN-PDX1 and PDX1-NTH. The results indicated that the expression and solubility profile of HTN-PDX1 was higher compared to PDX1-NTH (Figure 2.6A). Uninduced control showed no leaky expression of the PDX1 fusion protein (Figure 2.6A). Also, truncated fragments of PDX1 protein were observed, as shown in Western blotting (Figure 2.6A). This was irrespective of the various parameters and the genetic constructs (Figure 2.6A) analyzed. These truncated products may be due to proteolysis at specific sites in some protein molecules during expression (Ryan and Henehan 2013), or due to the presence of intragenic sequences mimicking *E. coli* ribosomal entry sites (Maertens et al. 2010), or due to cleavage at Asp-Pro bonds because of heating of protein samples. Including protease inhibitors in the solutions or using different *E. coli* strains or media did not decrease proteolysis. Although faint truncated fragments of PDX1 protein were observed, most of it was still retained as a full-length protein (Figure 2.6A). Due to high soluble expression in the case of HTN-PDX1, this gene construct was chosen for further experiments (Table 2.4).

For NGN3, in addition to the above-mentioned expression parameters, we have also screened induction media [Terrific broth (TB) or Luria Bertini broth (LB)]. The screening test revealed that optimal expression and production of the NGN3 fusion protein was achieved at a mid-log-phase cell density of OD_{600} ~ 0.5, IPTG inducer concentration of 0.05 mM, and a post-induction incubation time of 2 h (Table 2.4). The culture of bacteria in both TB and LB gave high NGN3 expression; the former was slightly higher among the two (Table 2.4). These results imply that a higher amount of inducer concentration, late growth phase, prolonged post-induction incubation time, and different media had no significant effect on the expression of NGN3 fusion proteins. Determination of these expression parameters was critical to achieving maximal expression and production of NGN3 fusion proteins. Additionally, the expression and maximal solubility profile of recombinant HTN-NGN3 and NGN3-NTH fusion proteins induced in TB at 37 °C were examined (Figure 2.6B). Interestingly, SDS-PAGE and immunoblotting analysis demonstrated a high amount of protein production profile of NGN3-NTH, whereas the production profile of HTN-NGN3 was predominant in the insoluble cell fraction or pellet (Figure 2.6B). Moreover, the effect of inducer concentration on protein expression was validated by the absence of leaky expression in uninduced controls (Figure 2.6B). Hence, based on the above-mentioned protein production analysis, we selected NGN3-NTH induced at 37 °C for subsequent analysis. To the best of our knowledge, our study is the first to report optimal expression parameters for obtaining maximal protein production of the recombinant NGN3 fusion protein, facilitating its native purification.

Table 2.4 Summary of the optimal expression conditions to obtain maximal soluble expression of human recombinant fusion proteins in *E. coli*

	Optimal expression parameters

Recombinant Proteins	Inducer concentration (mM)	Induction cell density (OD_{600})	Induction temperature (°C)	Induction time (hours)
HTN-PDX1	0.1	~0.5	37	2
NGN3-NTH	0.05	~0.5	37	2
MAFA-NTH	0.25	~0.5	37	2
HTN-PAX4 and **PAX4-NTH**	0.25	~0.5	37	2

Lastly, for recombinant MAFA and PAX4, the induction of transformed *E. coli* BL21(DE3) cultures at 0.5 induction cell density, 0.25 mM inducer concentration, 2 h post-induction incubation time, and 37 °C induction temperature gave a maximal expression of fusion proteins. Interestingly, for HTN-PAX4 and PAX4-NTH, SDS-PAGE and immunoblot analysis revealed an equivalent soluble expression profile of these recombinant proteins with predominant expression in the soluble fraction (Figure 2.6D). However, truncations were also observed in the expression of both HTN-MAFA and MAFA-NTH recombinant proteins (Figure 2.6C). This could be in response to the proteolysis of a few MAFA protein molecules during expression and solubilization (Ryan and Henehan 2013). Additionally, in the case of HTN-MAFA, the immunoblotting results revealed the presence of a high amount of recombinant protein in the pellet/insoluble fraction; thus, MAFA-NTH was used for further studies. Moreover, this analysis also demonstrated no leaky expression as uninduced cultures were used as a control and no protein expression was observed in control (Figure 2.6C). Thus, our results suggest maximal expression of recombinant MAFA-NTH, HTN-PAX4 and PAX4-NTH fusion proteins in the soluble form at 37 °C using the above-mentioned expression conditions.

All these results signify that the position of fusion tags at the N- or C- terminal can largely affect the expression and solubility profile of recombinant NGN3 fusion proteins. This was in agreement with the studies that made similar observations (Braun et al. 2002; Bosnali and Edenhofer 2008; Haridhasapavalan et al. 2020; Thool et al. 2021).

2.3 Conclusion

To conclude, this chapter proves the importance of codon optimization and identification of optimal expression parameters for maximal heterologous expression of these pancreatic-specific transcription factors in the form of recombinant fusion proteins. In addition, this chapter also shows the effect of the position of fusion tags with respect to the quantity and solubility of our protein of interest. Here, we report that the fusion tags at the C-terminal end of PDX1 recombinant protein compromise the quality when compared to their N-terminal fused counterparts. However, in the case of NGN3 and MAFA, we observed higher protein expression when the tags were placed at the C-terminal. But, the expression profile of recombinant PAX4 fusion proteins remained unaffected with the position of fusion tags. From these data, we can infer that the induction parameters and the position of fusion tags play an important role in the recombinant protein expression in terms of quality, quantity and solubility.

Chapter 3

Purification and biophysical analysis of pancreatic-specific transcription factors

A brief overview of the chapter

Heterologous protein expression and purification are simple as well as inexpensive means for large-scale protein production for research or commercial use. The proteins can be purified under native conditions (without using any denaturants) from the soluble fraction or under denaturing conditions (using mild to severe concentrations of detergents, mostly urea) from inclusion bodies. Proteins purified under native conditions do not require refolding, which makes the process simple and straightforward, whereas the use of denaturants during purification completely unfolds the protein and, as a result, requires refolding methods to refold to the native structure. Therefore, we have purified pancreatic-specific transcription factors, PDX1, NGN3, PAX4 and MAFA, in the form of recombinant proteins from *E. coli* under native conditions using the Immobilized metal affinity chromatography (IMAC) method. Bacterial cultures were grown in large volumes and induced with optimal expression parameters identified for each protein in the earlier chapter for maximizing protein production and purified using IMAC. SDS-PAGE and Western blotting using anti-His antibody as well as protein-specific antibodies were performed to confirm the identity of the protein. Post-purification, the overall yield of the purified recombinant proteins was determined using Bradford reagent and quantification using ImageJ software. We also performed far UV circular dichroism (CD) spectroscopy after desalting the protein to demonstrate that purified proteins maintained their secondary structure. The estimation of the secondary structure content of each of these proteins was analyzed using BeStSel online tool.

3.1 Materials and methods

3.1.1 Immobilized metal ion affinity chromatography

Immobilized metal ion affinity chromatography (IMAC) was carried out for each fusion protein to obtain purified recombinant proteins. Briefly, proteins of interest were expressed and were induced in large culture volumes (1.2 L) with the identified optimal expression conditions (Table 2.4; Chapter 2). Post-induction, bacterial cells were harvested and stored until further use. Next, the harvested cell pellets were resuspended in a pre-chilled resuspension buffer (Table 3.1) and lysed by ultrasonication at an amplitude of 33% (Pulse ON: 5 secs and OFF: 25 secs) using Vibracell^{TM} VCX-130 cell disruptor Sonics and Materials Inc., Newtown, CT, USA on ice for 30-40 minutes. The cell suspension was clarified by centrifugation at 11,000 rpm, 4 °C for 30 minutes.

3.1.2 Purification under native conditions

Once soluble protein expression parameters were verified for each of the factors, recombinant fusion proteins were expressed and induced on a large scale through the induction of 1.2 L of cultures. After that, cells were centrifuged, weighed and resuspended in a lysis buffer and disrupted in an ice-cold environment by ultrasonication to produce total cell lysate. The lysed product was fractionated by centrifugation to obtain a soluble supernatant fraction. Further, the fraction was incubated with Ni-NTA resin at 4 °C for 12 h with continuous shaking to facilitate protein binding to the resin. Following binding, the supernatant was drained out of the column (Bio-Rad) as a flow-through to remove any unbound undesirable bacterial proteins. The columns were washed with 5 column volumes of wash buffers 1, 2 and 3. Subsequently, bound fusion proteins were eluted using an elution buffer. All wash and elution buffer compositions are listed in Table 3.1. At various stages, protein wash and elution samples were collected to

analyze the fusion proteins on 12% SDS-PAGE and immunoblotting. Further, these fusion proteins were loaded onto PD10 columns (GE Healthcare) for desalting and then the buffer was exchanged against glycerol buffer (20% phosphate buffer along with 5% glycerol)/DMEM cell culture media. Lastly, the desalted protein was stored at –80 °C for future use.

Table 3.1 Summary of purification buffers used for each transcription factor.

Protein	Ingredients	Resuspension buffer	Wash buffers			Elution buffer
			1	2	3	
PDX1	PB (mM)	20	20	20	20	20
	NaCl (mM)	150	150	150	150	150
	Imidazole (mM)	20	50	100	150	500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	8.1	8.1	8.1	8.1	8.1
NGN3	PB (mM)	20	20	20	20	20
	NaCl (mM)	300	300	300	300	300
	Imidazole (mM)	20	50	100	150	500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	8.5	8.5	8.5	8.5	8.5
MAFA	PB (mM)	20	20	20		20
	NaCl (mM)	600	600	600		600
	Imidazole (mM)	20	200	400		1500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	7.5	7.5	7.5		7.5
PAX4	PB (mM)	20	20	20	20	20
	NaCl (mM)	600	600	600	600	600
	Imidazole (mM)	20	50	100	150	500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	8.5	8.5	8.5	8.5	8.5

PB, phosphate buffer; RT, room temperature

3.1.3 Far-ultraviolet Circular Dichroism spectroscopy

To determine if the secondary structure was still retained post-purification and estimate the composition of the secondary structure of purified fusion proteins, far ultraviolet Circular Dichroism spectroscopy (Far-UV CD) was performed using a J-815/J-1500 spectropolarimeter (Jasco, Japan) equipped with a thermoelectric cooling-based temperature control unit. Far-UV CD spectra of the protein were recorded as an average of a minimum of 5 accumulations from wavelength 280–150 nm in a 0.1 cm path length quartz cuvette at a scan rate of 100 nm/minute with a data integration time of 1 second. Before performing Far-UV CD, the protein was buffer exchanged (to remove imidazole) and desalted (to remove any background noise caused due to sodium chloride) post-purification and used for Circular Dichroism in 20 mM phosphate buffer. The final Circular Dichroism spectra obtained (after background subtraction) were analyzed using an *in silico* Beta Structure Selection (BeStSel) online tool, the information about which is described elsewhere (Micsonai et al. 2015, 2018).

3.1.4 SDS-PAGE and Western blotting

The SDS-PAGE and Western blotting were performed as described in the previous chapter (see Chapter 2, 2.1.4 section for details). All primary and secondary antibodies used and their respective concentrations are listed in Appendix 1.

3.2 Results and discussion

3.2.1 Purification of pancreatic-specific recombinant fusion proteins

Native purification using the soluble fraction of recombinant fusion proteins was performed using IMAC facilitated by nickel-nitrilotriacetic acid (Ni-NTA) charged purification column. These techniques will provide these recombinant factors with a better possibility of maintaining native folding conformation since several studies have reported that purification of recombinant proteins under native conditions helps them to maintain their secondary structure and thus, can be biologically active (Bosnali and Edenhofer 2008; Thier et al. 2011, 2012; Peitz

et al. 2014; Münst et al. 2016; Dey et al. 2021b, 2022; Thool et al. 2021). Moreover, the purified proteins should be devoid of any bacterial proteins or else these contaminating proteins could show negative effects. Therefore, we next purified these fusion proteins employing the IMAC protein purification technique from the soluble fraction (under native conditions). IMAC is based on the interaction between residues of a His-tagged protein and immobilized metal ions (in the current study, Ni^{2+}-NTA) (Bornhorst, J.A., Falke 2010). Ni^{2+} is extensively used for IMAC purification to obtain a high yield. Thus, the recombinant fusion proteins were first induced in large 1.2 L culture volumes and expressed with the identified optimal conditions (see Chapter 1, Table 2.4 for details), clarified to separate insoluble protein molecules, and then purified under native conditions from the obtained soluble cell fractions using Ni–NTA affinity chromatography, as shown in Figure 3.1.

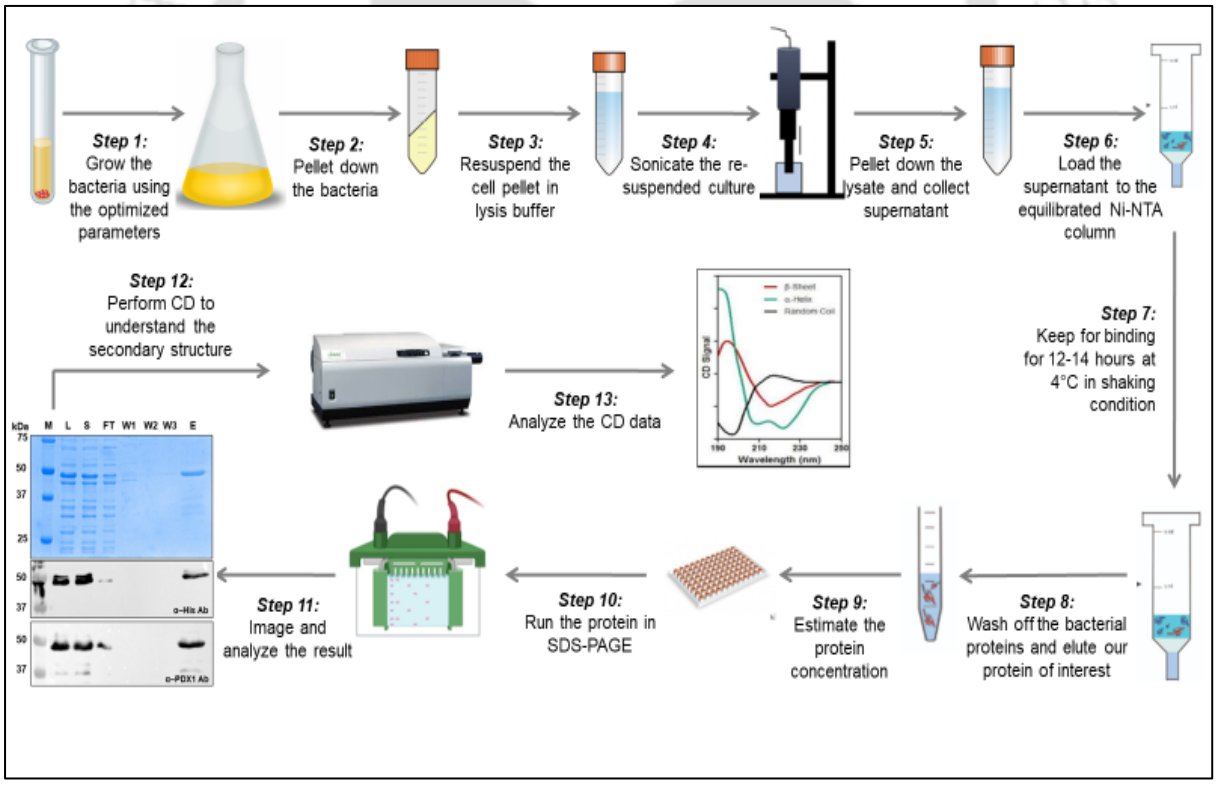


Figure 3.1 Schematic representation of the overall purification experimental strategy.

3.2.1.1 Purification of HTN-PDX1 fusion protein

The HTN-PDX1 fusion protein was purified using IMAC under native conditions (from the soluble cell fraction). Thus, the HTN-PDX1 fusion protein was induced with identified parameters **(Chapter 2, Table 2.4 for details)** and purified using IMAC under native conditions. The purity of the HTN-PDX1 fusion protein was confirmed by SDS-PAGE (Figure 3.2A; *top*), and the fusion protein was detected by Western blotting using an anti-His antibody (Figure 3.2A; *middle*). A band corresponding to human HTN-PDX1 fusion protein was observed at ~48 kDa (Figure 3.2A; *bottom*). Interestingly, no major visible truncations were observed when the expression was carried out at a large scale (Figure 3.2), unlike those observed in Figure 2.6A. On the other hand, faint truncated fragments were observed during large scale purification. These truncated fragments could be due to reasons mentioned earlier in Section 3.2.1.2. The yield of the purified HTN-PDX1 fusion protein was around 1.31 mg/L (Figure 3.2B). The elution profile of the purified HTN-PDX1 is shown in Figure 3.2C and the elution fraction analysis corresponding to the same is provided in Figure 3.2D. Thus, the PDX1 fusion protein was purified under native conditions from the soluble cell fraction.

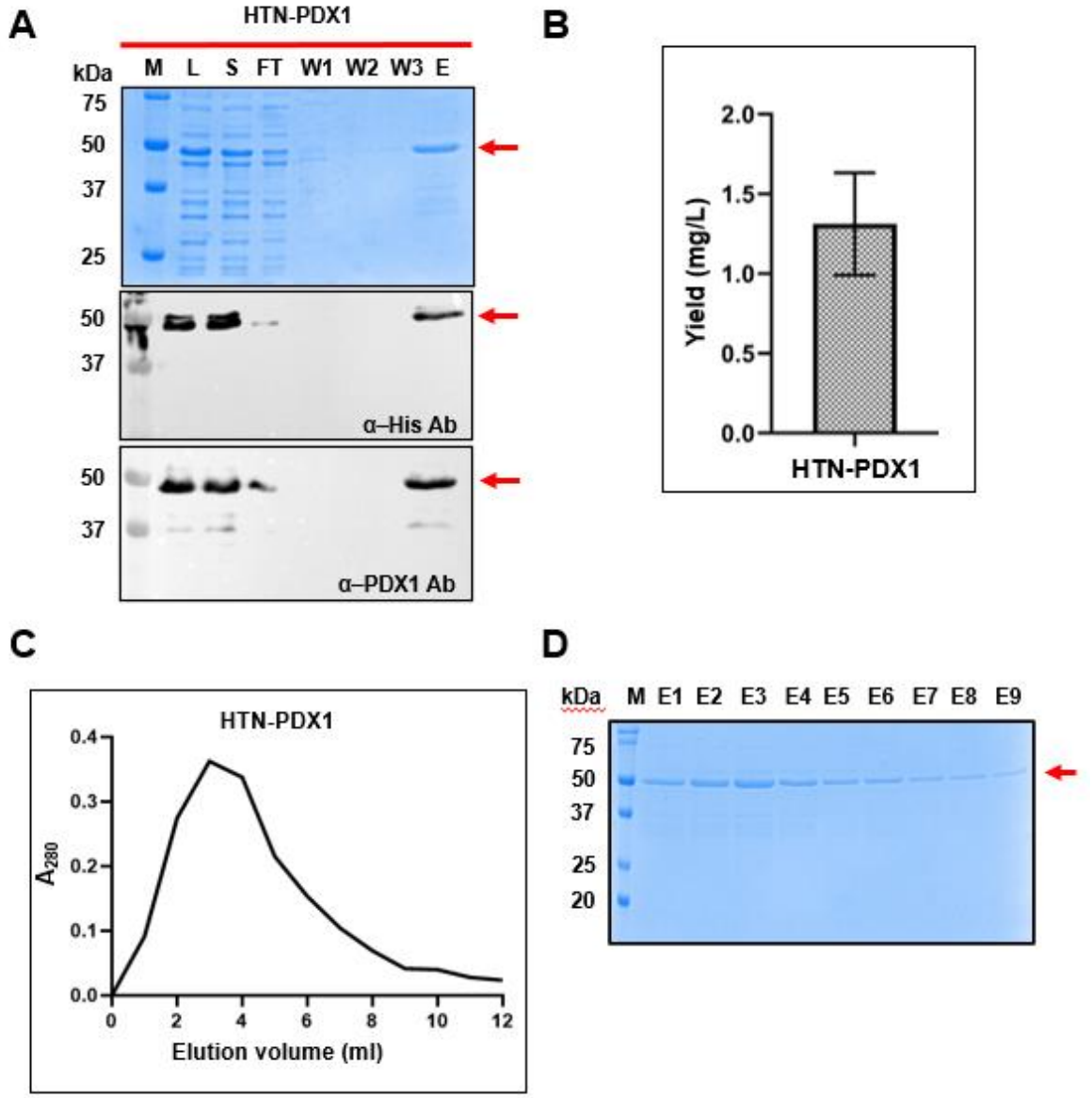


Figure 3.2 Purification of recombinant HTN-PDX1 fusion protein under native conditions. **(A)** BL21(DE3) cells harbouring pET28a(+)-*HTN-PDX1* were induced with 0.1 mM IPTG at OD_{600} ~0.5 and incubated at 37 °C for 2 h. Subsequently, the cells (2 g wet weight of *E. coli* BL21(DE3) cell pellet) were harvested, lysed by ultrasonication and centrifuged to obtain soluble supernatant fraction. From the supernatant fraction, the expressed proteins were purified under native conditions. Samples collected during different stages of purification were boiled at 95 °C for 10 minutes and separated on 12% SDS-PAGE gels with normalized loading and either stained with Coomassie Brilliant Blue G-250 (*top*) and Western blotting was

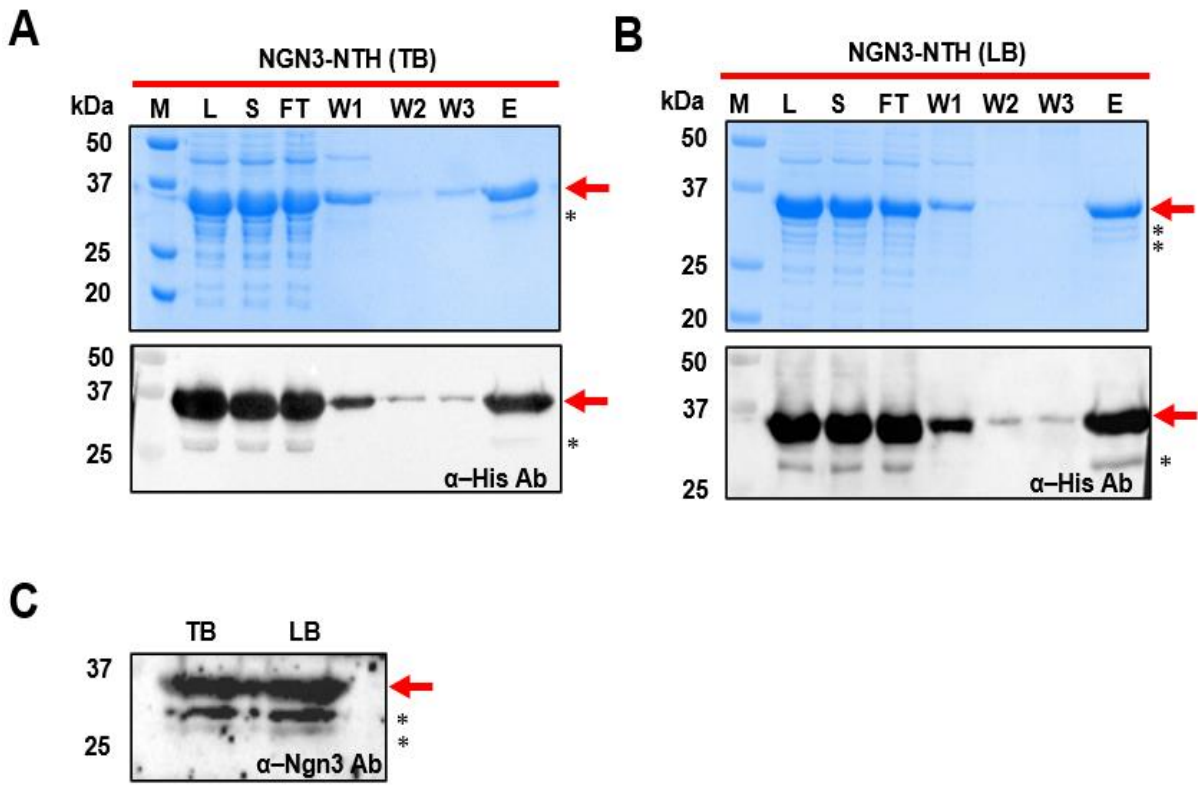
performed with anti-Histidine (α-His) antibody (*middle*) and anti-PDX1 (α-PDX1) antibody (*bottom*) **(B)** The bar graph shows the final yield of purified HTN-PDX1 purified protein under native conditions, (n=3) **(C)** Elution profile of HTN-PDX1 protein purification under native conditions. **(D)** Elution fraction analysis of human PDX1 fusion protein. Elution fractions collected during the purification of HTN-PDX1 protein were resolved with equal loading volume by SDS-PAGE and then resolved proteins were stained using Coomassie brilliant blue. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; F, flowthrough fraction; W, wash fractions; E, eluted fraction; UI, uninduced (total cell lysate)

3.2.1.2 Purification of NGN3-NTH fusion protein

Codon optimization, identification of optimal expression conditions, and suitable gene constructs facilitated the high expression of the recombinant NGN3 fusion protein in *E. coli* in a soluble form. Therefore, we purified the NGN3-NTH employing the IMAC protein purification technique from the soluble fraction (under native conditions). NGN3-NTH was induced in a TB medium with the parameters identified **(Chapter 2, Table 2.4 for details),** which resulted in very high NGN3 expression, and purified it using IMAC. The high purity of the NGN3-NTH fusion protein was determined by Coomassie staining of SDS-PAGE gels (Figure 3.3A; *top* and Figure 3.3D), and the protein was detected by Histidine antibody (Figure 3.3A; *bottom*). We also induced NGN3 expression in the LB medium with the parameters identified since the expression in the LB medium was also high and purified it using IMAC. The high purity of the NGN3-NTH (induced in LB medium) was determined by Coomassie staining of SDS-PAGE gels (Figure 3.3B; *top* and Figure 3.3E), and the protein was detected by Histidine antibody (Figure 3.3B; *bottom*). Further, the purified fusion protein (induced in TB and LB media) was confirmed with an anti-NGN3 antibody (Figure 3.3C). Surprisingly, we also observed faint truncated fragments of NGN3 protein (Figure 3.3A and 3.3B (*top*);

3.3D; 3.3E), which was detected by Histidine antibody (Figure 3.3A and 3.3B; *bottom*) and also by NGN3 antibody (Figure 3.3C). These truncated fragments could be due to reasons mentioned earlier in Section 3.2.1.2.

Also, a major amount of expressed protein was lost in the unbound (flow-through) fraction [Figure 3.3A and 3.3B (*top and bottom*)], probably due to the overloading of the soluble fraction on the purification column, which might have resulted in inefficient binding of the NGN3 fusion protein to the resin or low volume of resin used (1:1; for 2 g of bacterial pellet, 2 mL of resin was used). Further improvements in the protocol can improve the yield of the purified protein. A band corresponding to the NGN3-NTH protein was observed at ~34 kDa (Figure 3.3A and B). The calculated molecular weight of NGN3-NTH protein is ~30 kDa.



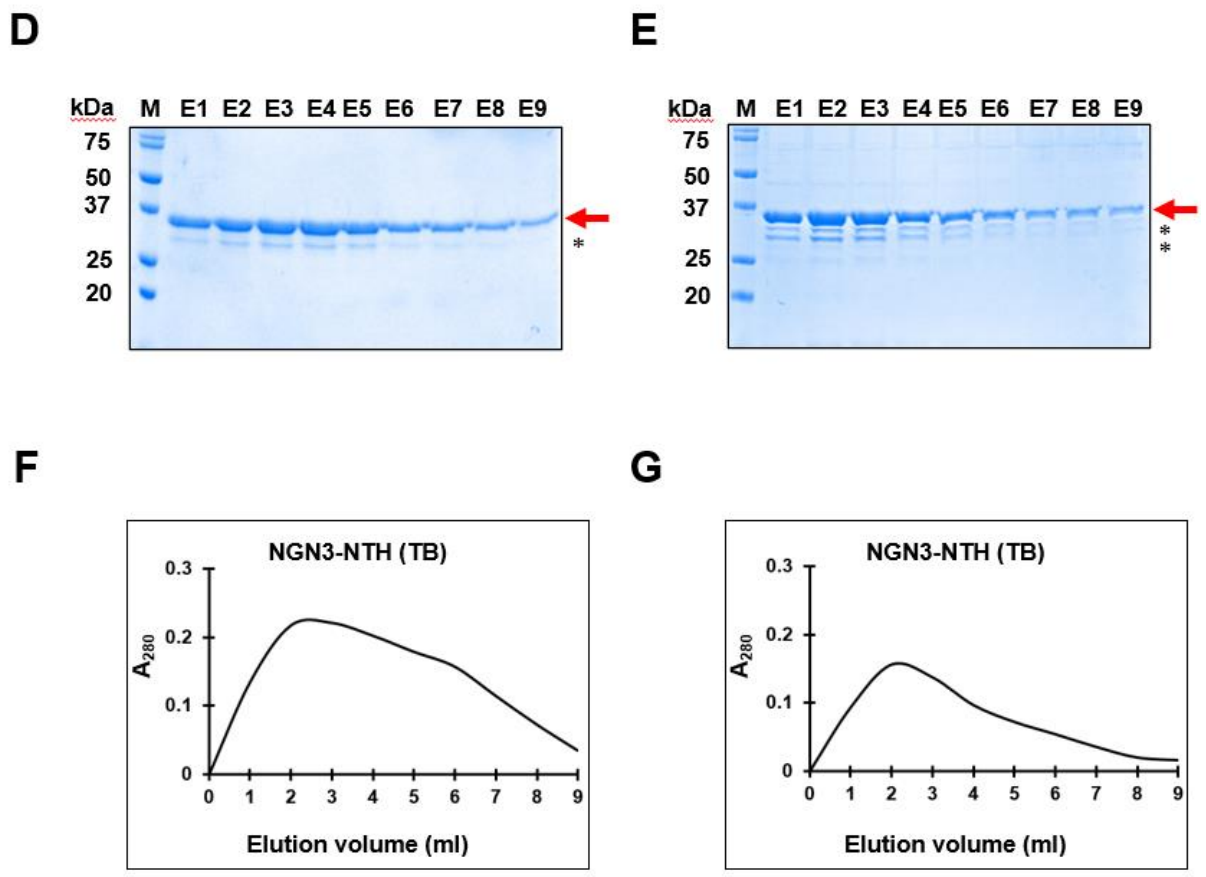


Figure 3.3 Purification of recombinant NGN3-NTH fusion protein (induced in TB and LB) under native conditions. The recombinant human NGN3-NTH fusion was induced in TB **(A)** and LB **(B)**, expressed and purified using Ni–NTA affinity chromatography under native conditions. The purification samples were resolved on 12% SDS-PAGE gels and then stained using Coomassie brilliant blue (A and B; top) and performed Western blotting using histidine antibody (A and B; bottom). **(C)** Western blotting of pooled eluate fractions (1–9) using the anti-Ngn3 antibody of purified NGN3 fusion protein induced in TB and LB. *Truncations of NGN3 fusion protein. TB terrific broth, LB lysogeny broth, M protein marker (kDa), L total cell lysate, S supernatant/soluble cell fraction, FT flow-through fraction, W (1–3) wash buffer (1–3), E eluate fractions, Ab antibody. Elution fraction analysis of human NGN3-NTH fusion protein (induced in TB and LB). **(D)** and **(E)** Elution fractions collected during purification of NGN3-NTH protein [induced in TB **(D)** and LB **(E)**] were resolved by SDS-PAGE (with equal

loading volume) and the proteins resolved were stained using Coomassie brilliant blue (top). *Truncations of NGN3 fusion protein. M protein marker (kDa), E1-9 elution 1–9. **(F)** and **(G)** Elution profile [plot of Absorbance (A_{280}) versus elution volume (ml)] of affinity purified NGN3- NTH fusion protein [**F** (induced in TB) and **G** (induced in LB)]

The purity of NGN3-NTH induced in TB and LB was >90% (calculated using ImageJ software). Recombinant NGN3-NTH induced in TB had a better yield (~12 mg from 0.6 L culture) compared to induced in LB medium (~6 mg from 0.6 L culture). The elution profile of affinity-purified NGN3 fusion proteins induced in TB and LB is shown in Figure 3.3F (for TB) and Figure 3.3G (for LB). Thus, we have established the recombinant human NGN3 fusion protein purification to homogeneity under native conditions.

3.2.1.3 Purification of MAFA-NTH fusion protein

We next purified MAFA-NTH using affinity chromatography from the soluble fraction. For this, MAFA-NTH was induced in 1.2 L of TB culture media, supplemented with 50% glycerol, for 2 h at 37 °C with 0.25 mM of IPTG at shaking condition. Post induction, the culture was clarified using centrifugation, and native purification was carried out using IMAC. Samples were collected at every purification step for further analysis. The purified MAFA-NTH protein was detected using Histidine as well as MAFA antibodies. The calculated molecular weight of the protein, along with the tags, is 44.5 kDa, and we observed our protein of interest at ~48 kDa. Interestingly, we also observed multiple truncated protein bands along with the full-length protein, which could be due to reasons mentioned earlier in Section 3.2.1.2. Moreover, little loss of protein was also observed in the flow-through or unbound fraction, which could be due to the overloading of the supernatant fraction to the purification column resulting in inefficient binding to the resin (for ~6 g of bacterial wet pellet, the resin used is ~3 mL (2:1)).

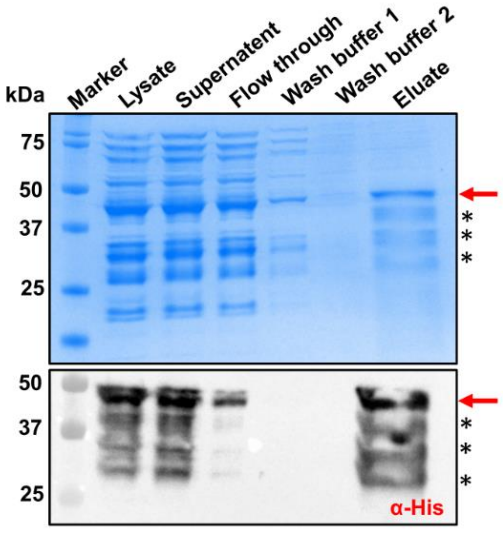


Figure 3.4 Purification of recombinant MAFA-NTH fusion protein under native conditions. MAFA-NTH native purification. kDa, kilodalton; α-His, anti-His Antibody. *Truncations of MAFA-NTH fusion protein

3.2.1.4 Purification of HTN-PAX4 and PAX4-NTH fusion proteins

We also purified the HTN-PAX4 and PAX4-NTH recombinant proteins employing the IMAC protein purification technique under native conditions. Purification under native conditions results in the production of recombinant protein with native-like secondary structure, retained protein folding, and biological activity (Wingfield 2015). HTN-PAX4 and PAX4-NTH were induced in 1.2 L of TB culture medium with the above-mentioned optimal expression conditions and purified using IMAC. Samples were collected at every purification step for analysis, and the purity was determined by Coomassie staining of SDS-PAGE gels (Figure 3.5A; *top* and 3.5B; *top*). The protein was detected by Histidine antibody (Figure 3.5A and B; *middle*) and Pax4 antibody (Figure 3.5A and B; *bottom*). The calculated molecular weight of HTN-PAX4 and PAX4-NTH is ~45.8 kDa. We observed our band of interest at ~45 kDa in the elution fraction. Although faint truncated fragments (~40 kDa) in both HTN-PAX4 and PAX4-NTH fusion proteins (more dominant in the case of HTN-PAX4) were observed below our band of interest as detected by Pax4 antibody, most of the protein was retained as a full-length

protein (Figure 3.5A and B; *bottom*). These faint truncated fragments could be due to reasons mentioned earlier in Section 3.2.1.2. However, the majority of expressed protein was lost in flow-through (unbound) fraction, probably due to overloading of the soluble fraction on the purification column resulting in inefficient binding of recombinant PAX4 fusion proteins to the resin (2:1; for ~7 g of bacterial pellet, 3.5 mL of resin was used). An increase in the resin volume along with further improvements in the protocol, can enhance the overall yield of the purified protein. HTN-PAX4 showed better (~3.15 mg from 1.2 L culture) yield compared to PAX4-NTH (~0.83 mg from 1.2 L culture) as detected by Coomassie and immunoblot analysis (Figure 3.5A and B). The elution profile of PAX4-NTH and HTN-PAX4 fusion proteins induced in TB is shown in Figure 3.5C (for HTN-PAX4) and Figure 3.5D (for PAX4-NTH). Thus, we have demonstrated the purification of recombinant human PAX4 fusion proteins to homogeneity under native conditions.

The purification strategy established for all proteins is simple, economical, reproducible, and produced recombinant fusion proteins with good yield. This strategy avoids protein purification from inclusion bodies; the latter is cumbersome, time-consuming, and expensive and requires protein extraction using strong denaturants and, subsequently, the refolding of the misfolded protein (Baneyx and Mujacic 2004). Native purification of recombinant proteins provides a greater chance of retaining the native-like conformation and biological activity (Wingfield 2015).

3.2.2 Secondary structure determination of purified proteins

Neither the crystal structure nor the secondary structure content of these recombinant proteins has been reported to date. Therefore, the secondary structure of the purified recombinant proteins was determined by performing far ultraviolet (UV) Circular Dichroism (CD)

spectroscopy. The far UV CD spectroscopic method is a widely used method to understand the folding features of proteins where the secondary structure is not determined (Kelly et al. 2005; Micsonai et al. 2015). The far UV CD spectrum depends directly on the protein folding conformation, and its characteristic shape and magnitude indicate different secondary structure content, namely α-helix, β-sheet, turn and random coil (Greenfield 2006). The CD spectrum representing α-helix shows negative peaks at 222 and 208 nm and a positive peak of about 193 nm (Greenfield 2006). Likewise, the β-sheet has a negative peak at 218 nm and a positive peak at 195 nm, while the random coil has a positive peak at 210 nm and a negative peak at 195 nm (Greenfield 2006).

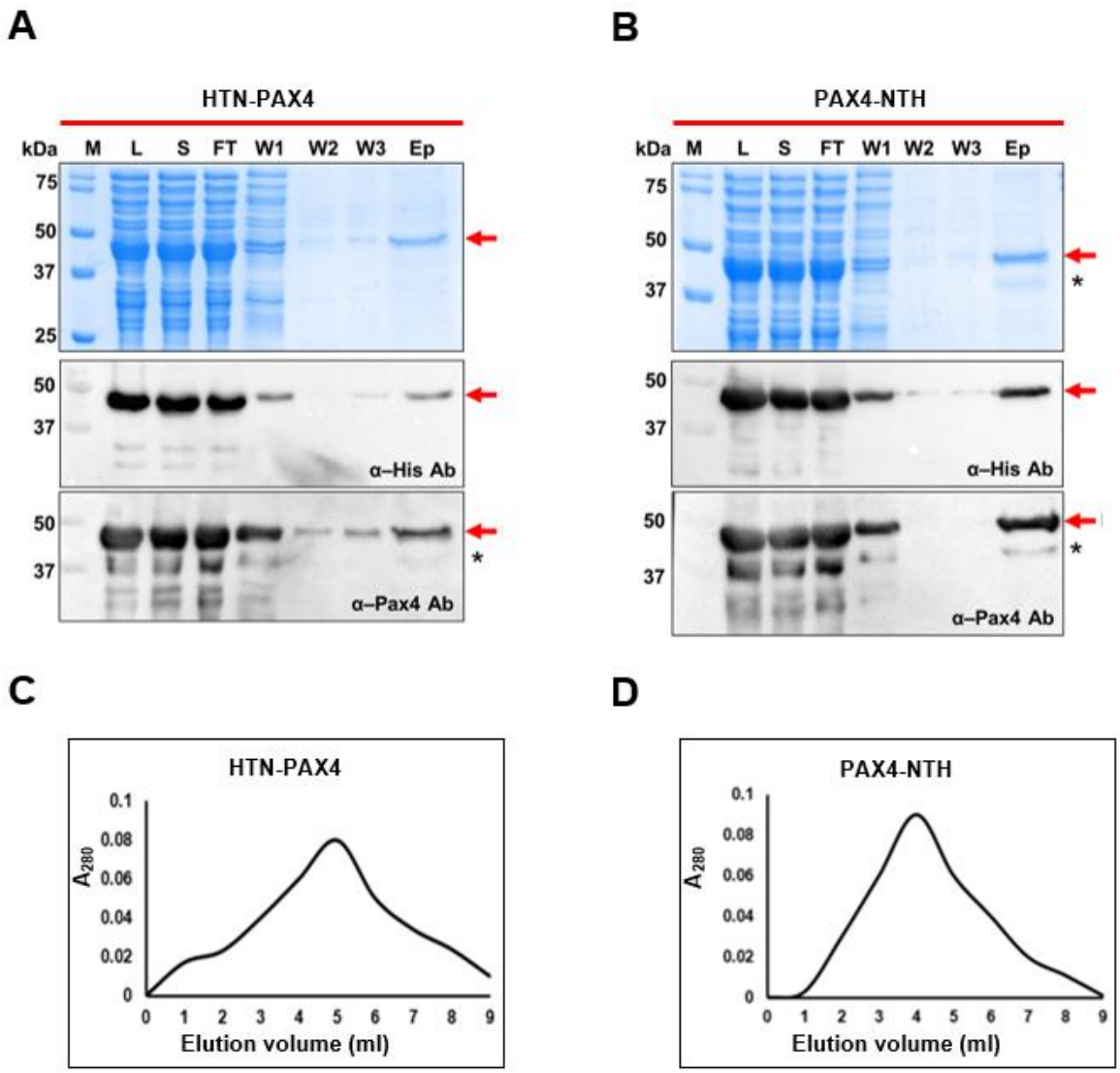


Figure 3.5 Purification of recombinant HTN-PAX4 and PAX4-NTH fusion proteins under native conditions. The recombinant human PAX4 fusion proteins were induced, expressed and purified using Ni– NTA affinity chromatography under native conditions. The purification samples were resolved on 12% SDS-PAGE gels and then stained using Coomassie brilliant blue (**A** and **B**; top) and performed Western blotting using histidine antibody (**A** and **B**; middle) and PAX4 antibody (**A** and **B**; bottom). **(C)** and **(D)** Elution profile of HTN-PAX4 and PAX4-NTH, respectively [plot of Absorbance (A_{280}) versus elution volume (ml)] of affinity purified PAX4 fusion proteins.

From the CD spectra plotted using BeStSel (Figure 3.6A), it is evident that the recombinant HTN-PDX1 protein has maintained its secondary structure post-purification and desalting. The far UV CD data were further examined using *in silico* BeStSel tool to quantify the secondary structure content (Figure 3.7A) (Micsonai et al. 2018). BeStSel is a recently developed improvised online tool that estimates the secondary structure content using the characteristic CD spectrum of the target protein (Micsonai et al. 2018). The estimated secondary structure of the purified HTN-PDX1 protein indicates that it predominantly comprises of random coils (~47%) with an almost equal amount of α-helices (~19%) and β-sheets (~20%) along with a significant amount of turns (~14%) (Figure 3.7A). This data established that the recombinant HTN-PDX1 fusion protein had maintained its secondary structure and it shows great promise of being bioactive.

We next sought to study the secondary structure of NGN3 protein using far UV CD spectroscopic technique as the crystal structure of partial or full-length human NGN3 protein has not been reported to date. The far UV CD spectra of recombinant NGN3-NTH protein induced in TB and LB appeared similar (Figure 3.6B). Further, the secondary structure was quantified using the BeStSel tool. The CD spectra (plotted using BeStSel result) of recombinant

NGN3-NTH protein (induced in TB medium) showed that its secondary structure majorly comprised of α-helices (~36%) and random coils (~43%) with a substantial amount of β-sheets (~7%) and turns (~14%) (Figure 3.7B). Similarly, the CD spectra (plotted using BeStSel result) of recombinant NGN3-NTH protein (induced in LB medium) showed that its secondary structure majorly comprised of α-helices (~40%) and random coils (~40%) with a substantial amount of β-sheets (~8%) and turns (~12%) (Figure 3.7C). Although the spectra appeared similar, minor variability was observed upon the quantification of the secondary structure content. These results established that the purified NGN3 fusion proteins had retained their secondary structure post-purification and, most presumably, are bioactive. However, due to the non-availability of the 3D structure of human NGN3 protein, the spectra and the secondary structure content determined in this study may differ from the native protein. In addition, the presence of fusion tags may also alter the structure of the NGN3 protein.

Post-purification MAFA-NTH, we planned to determine whether the purified protein maintained its secondary structure. To understand this, we desalted and buffer exchanged the purified protein and performed far UV CD spectroscopy (Figure 3.6C). Further, the secondary structure was quantified using the BeStSel tool (Figure 3.7D). The results showed that the secondary structure content of purified HTN-MAFA contained predominantly random coils (~48%) and an almost equal amount of α-helices (~20%) and β-sheets (~18%) and the amount of turns is ~14% (Figure 3.7D). The result confirmed that the purified recombinant MAFA-NTH maintained a secondary structure post-native purification and desalting. However, the complete crystal structure of human MAFA protein is not available and, thus, cannot be compared to our results.

Next, the retention of secondary structure conformation of full-length human PAX4 protein was determined using far UV CD. To achieve this, the purified HTN-PAX4 and PAX4-NTH fusion proteins were subjected to CD analysis after desalting and buffer exchange. Thus,

the secondary structural conformation of purified recombinant PAX4-NTH and HTN-PAX4 proteins was evaluated using far UV CD (Figure 3.6D). Subsequently, BeStSel online server was used to quantify the secondary structure content from data obtained from the CD spectra. The quantification of CD spectra of recombinant PAX4-NTH protein suggests that this protein contains the majority of the random coil (~38%), then α-helices (~32%), β-sheets (~19%), and turns (~11%) (Figure 3.7E). Conversely, the quantification of CD spectra HTN-PAX4 protein suggests that this protein contains the majority of α-helices (~38%), then random coils (~37%), β-sheets (~18%), and turn (~7%) (Figure 3.7F).

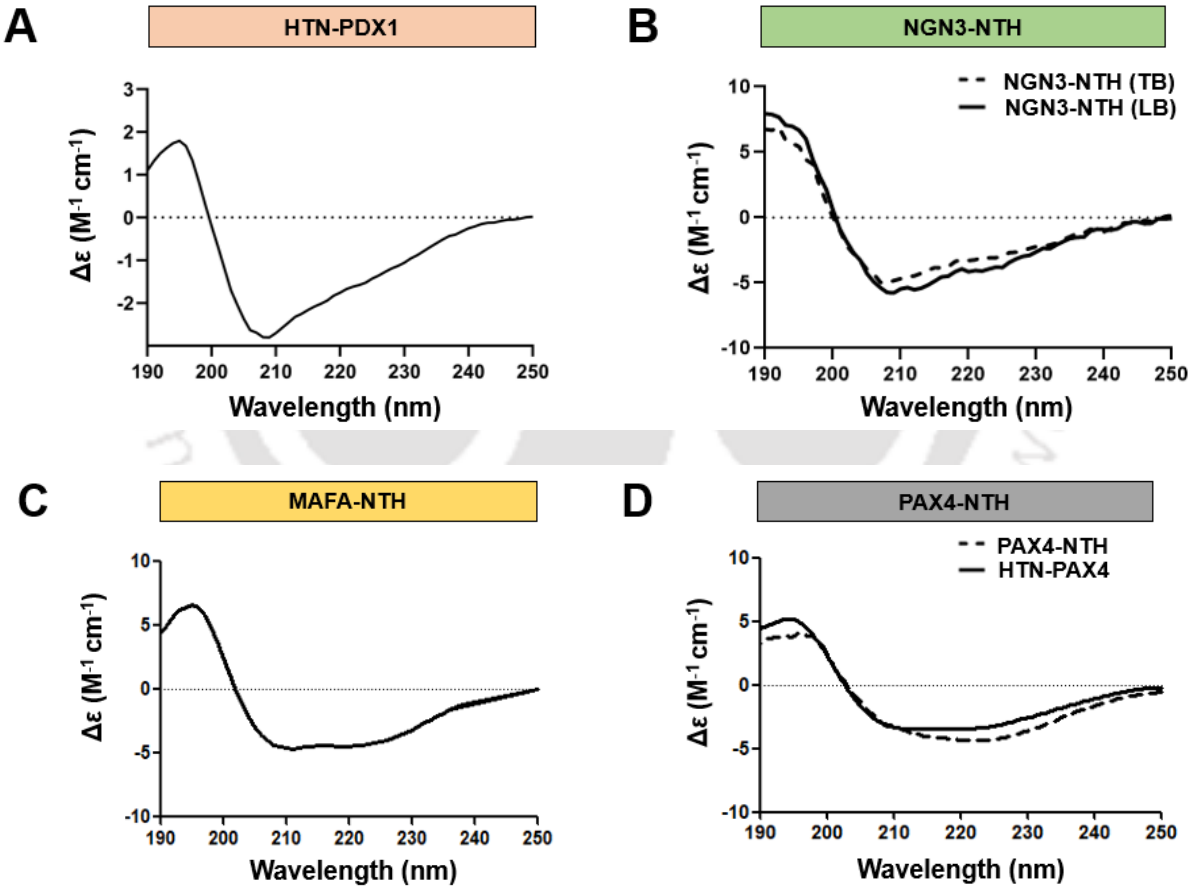


Figure 3.6 Secondary structure determination of purified recombinant fusion proteins using far-UV CD spectroscopy. The purified recombinant fusion proteins were desalted and then analyzed for their secondary structures with CD spectroscopy. The obtained

measurements were analyzed with the BeStSel web server, and the spectra were plotted with wavelength (nm) and delta epsilon (M^{-1} cm^{-1}) on the x-axis and y-axis, respectively.

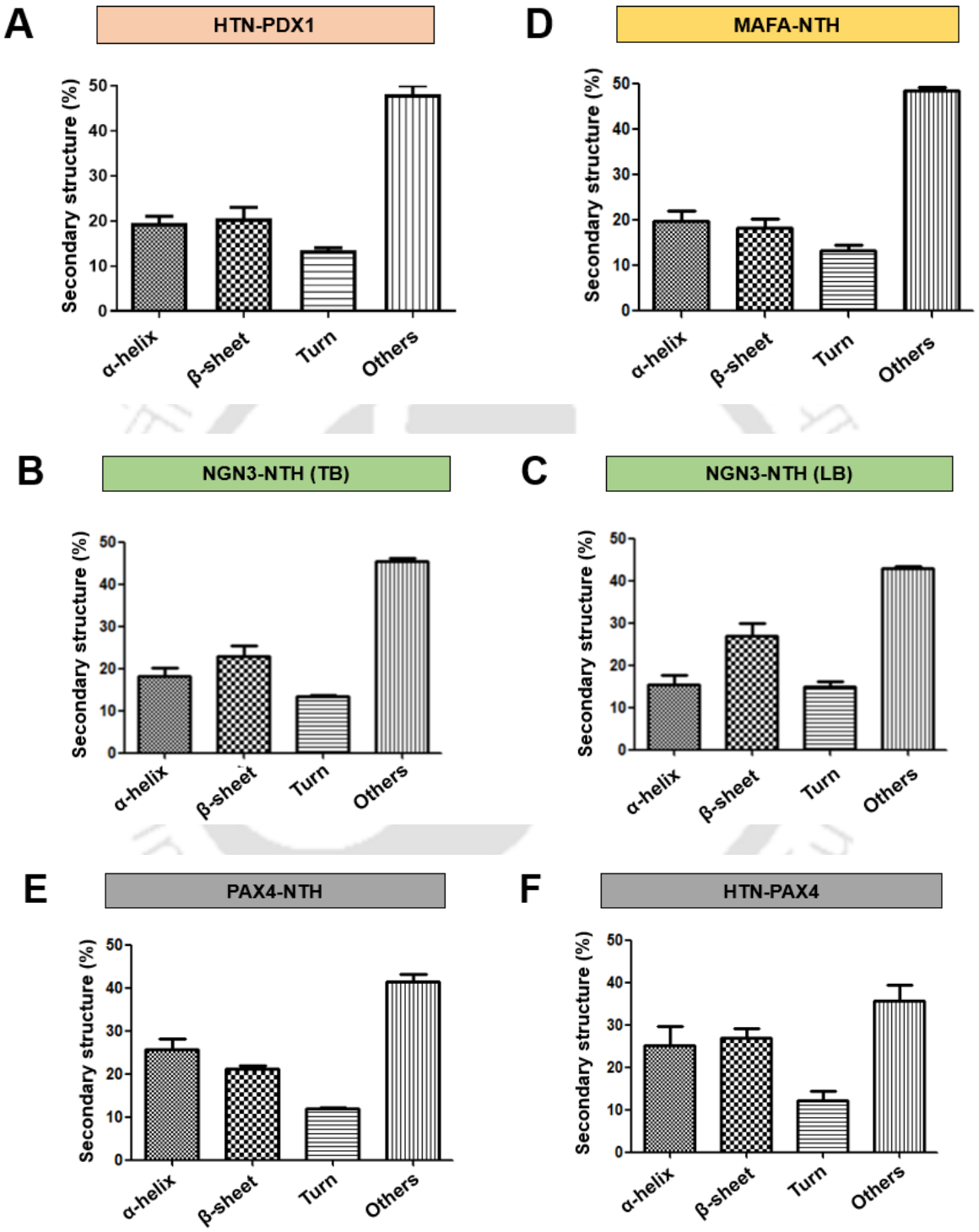


Figure 3.7 Summary of the quantified secondary structures of the purified recombinant fusion proteins from their spectra using BeStSel server. The purified recombinant fusion

proteins were desalted and then analyzed for their secondary structures with CD spectroscopy. The CD spectra were analyzed with the BeStSel web server, and the resulting quantitative secondary structure composition (α-helices, β-sheets, turns, and others) was represented using the Bar graphs. The quantitative data shown are mean ± SEM (n≥4).

Although there are minor differences in the amount of secondary structure content for both the fusion proteins, the exact amount of secondary structure in PAX4 protein needs to be further validated with its crystal structure, which is unavailable to date. These differences could be due to the presence of an additional slightly intense truncated band observed in the case of HTN-PAX4 compared to PAX4-NTH and/or minor differences in the protein concentrations of these two proteins. Thus, these results account for the fact that purified recombinant PAX4-NTH and HTN-PAX4 proteins have a retained secondary structure post-purification and are plausible to be biologically active.

3.3 Conclusion

This chapter dealt with the one-step purification of pancreatic-specific transcription factors in the form of recombinant proteins. The strategy used in purifying these proteins is simple, economical and reproducible. Furthermore, these proteins are purified from soluble fraction, which avoids the usage of denaturing agents like a high concentration of urea. As a result, the protocols do not require any refolding procedures that are relatively expensive and cumbersome. However, the overall yield of some of the purified proteins (like HTN-PAX4 and PAX4-NTH) is low, and further refinement in the purification protocol is required. Also, we observed that the proteins maintained their secondary structure post-desalting and buffer exchange. The secondary structure content of these proteins was estimated using BestSel. The analysis revealed that all the purified proteins consisted of predominantly random coils. In the

case of HTN-PDX1, MAFA-NTH and HTN-PAX4, there are almost equal amounts of α-helices, but PAX4-NTH had a slight less amount of α-helices. However, in the case of NGN3-NTH, there is a higher amount of β-sheets compared to α-helices. Therefore, these purified recombinant fusion proteins can replace its viral and genetic forms to generate genetically stable β-cells for the treatment of diabetes mellitus and can also be used to investigate their biological roles in cancer and other cellular processes. These proteins fusion proteins can be delivered to mammalian cells to carry out multiple cellular-based studies.

Chapter 4

Determination of cell penetration, nuclear localization and biological activity of pancreatic specific transcription factors

A brief overview of the chapter

Far-UV CD spectroscopy showed that the purified proteins maintained their secondary structure; however, to determine the biological activity of these proteins, we applied them to various mammalian cell lines. First, we studied the stability of these proteins at standard cell culture conditions over a period of 72 h. The immunoblotting results demonstrated that the proteins were stable for more than 48 h and could be used for further studies. Next, the transduction ability of these recombinant proteins into the nucleus of mammalian cells was studied. The immunocytochemistry results demonstrated that the recombinant fusion proteins successfully penetrated the cell membrane and translocated to the nucleus of human pancreatic duct cancer cells, PANC-1. These findings were further validated using sub-cellular fractionation assay. PDX1 and PAX4 are reported to act as tumor suppressor proteins in human gastric cancer (AGS cells) and melanoma (A375 cells) cell lines, respectively. In order to understand the role of these recombinant proteins as tumor suppressor proteins, we carried out migration ability using scratch assay, cell proliferation assay and cell cycle analysis (using flow cytometry). The results demonstrated that cells treated with fusion proteins (either HTN-PDX1 or HTN-PAX4/PAX4-NTH) reduced the migration ability of AGS cells (in case of HTN-PDX1) and A375 cells (in case of HTN-PAX4/PAX4-NTH). Furthermore, cell cycle analysis of the HTN-PAX4 or PAX4-NTH proteins treated cells showed sub G0/G1 arrest. Next, the effect of these recombinant proteins on the induction of insulin mRNA was studied. PANC-1 cells, when treated with varying concentrations of fusion proteins (HTN-PDX1, NGN3-NTH, and MAFA-NTH) for 5 days, could successfully induce insulin mRNA expression. Thus, we conclude that these recombinant fusion proteins are biologically active and, therefore, can be used for future reprogramming studies and investigate their function in other biological processes.

4.1 Materials and methods

4.1.1 Stability of fusion proteins at cell culture conditions

Protein stability assays are used to evaluate the shelf life/longevity of a protein in solution and thus, the stability of purified recombinant proteins was determined. Recombinant proteins from glycerol stock were diluted with protein transduction media containing high-glucose Dulbecco's Modified Eagle medium (DMEM; Invitrogen), which is supplemented with 5% fetal bovine serum (FBS; Invitrogen) and 1% penicillin-streptomycin solution (P/S; Invitrogen) to the final concentration of 200 nM. The prepared transduction media was next aliquoted in different vials and incubated for a varied period under standard cell culture conditions (37 °C, 5% CO_2, humidified condition). After completion of the incubation period, samples were collected, centrifuged at 8000 rpm for 30 minutes at 4 °C, and stored at -20 °C for further use. These clarified samples were analyzed using Western blotting against an anti-histidine antibody.

4.1.2 Cell culture

Human pancreatic ductal cancer (PANC-1) cells, human melanoma (A375) cells and human gastric cancer (AGS) cells were procured from National Centre for Cell Science, Pune, India. PANC-1 and A375 cells were cultured in high-glucose DMEM supplemented with 10% FBS and 1% P/S. AGS cells were cultured in Ham's F12K media supplemented with 10% FBS and 1% P/S. All the cell lines were incubated under standard cell culture conditions, 37 °C, 5% CO_2, humidified condition and passaged using Trypsin-EDTA (Invitrogen) at 70-80% confluency in the ratio 1:4.

4.1.3 Protein transduction, immunofluorescence and microscopy

The PANC-1 cells were adjusted to 5×10^5 cells/2 ml and seeded in 6-well culture dishes. Cells were grown till 40–50% confluency at 37 °C with 5% CO_2 in a humidified atmosphere. The medium was then replaced with filter sterile protein transduction medium (2% FBS, 1% P/S, and 200 nM of purified recombinant protein or glycerol buffer (vehicle) in DMEM) and re-incubated for 12 h. After the incubation, cells were washed with PBS and fixed with 2% paraformaldehyde (Merck Millipore). Cells were then permeabilized by treating with 0.1% Triton™ X-100 (Sigma-Aldrich) in PBS for 10 minutes. Next, cells were blocked with a blocking solution (0.5% bovine serum albumin, 0.15% glycine in PBS) for 1 h at room temperature. The primary antibody was then added and incubated for 2 h at room temperature in a moist chamber. Cells were then washed three times with PBS and incubated with a secondary antibody for 1 h in a moist chamber at room temperature. After incubation, cells were washed three times with PBS, and the nucleus was counterstained with 4′ ,6-diamidino-2-phenylindole (DAPI) dye (1:10000; Sigma-Aldrich) for 10 minutes. Excess DAPI staining was removed with PBS wash. Fluorescently labeled dye and secondary antibodies were excited at 345 nm (DAPI) and 555 nm (RFP) or 489 nm (GFP) and visualized under an inverted fluorescent microscope (IX83 Olympus, Japan) equipped with a DP80 CCD camera. Samples were illuminated using a pE-300 white Cool LED light source, and image stacks were acquired using the 20x/0.45NA objective at 1.57 μm intervals. Images were analyzed by Cell Sens dimension (Olympus) and Image J software.

4.1.4 Sub-cellular fractionation

The PANC-1 cells were adjusted to 1×10^6 cells and seeded in T-25 culture flasks, and incubated with 200 nM of recombinant protein or vehicle control for overnight under standard cell culture conditions. The next day, cells were washed with ice-cold PBS and SF buffer containing 20 mM HEPES at pH 7.4, N-(2-Hydroxyethyl) piperazine-N (2-ethane sulphonic acid) (SRL), 250 mM sucrose (SRL), 10 mM KCl (HiMedia), 1.5 mM $MgCl_2$ (Magnesium

Chloride; SRL), 1 mM EDTA (Ethylenediaminetetraacetic Acid; SRL), 1 mM EGTA (Ethylene Glycol-O-O-bis-(2-Aminoethyl) N, N, N, N-Tetraacetic Acid; SRL), 1 mM DTT (Dithiothreitol, SRL) and Protease Inhibitors cocktail (Thermo Scientific)] was added to the treated as well as control flasks. Next, the cells were scrapped with a cell scraper, and cell lysates were collected in a 1.5 mL microcentrifuge tube. Lysates were then agitated at 50 rpm at 4 °C for 30 minutes and then centrifuged at 700 x g at 4 °C for 5 minutes. Both supernatant and pellet fractions were stored for the next step. The supernatant was further centrifuged at 10,000 x g at 4 °C for 10 minutes, and the supernatant or cytoplasmic fraction was collected. The remaining cell pellet was washed with SF buffer again and centrifuged at 700 x g 4 °C for 10 minutes. The pellet was resuspended in Nuclear Lysis buffer (NL) containing 50 mM Tris HCl at pH 8.0, 150 mM NaCl (HiMedia), 1% Nonidet P-40 (Merck), 0.5% sodium deoxycholate (SRL), 0.1% SDS (SRL), Protease Inhibitor cocktail and 10% glycerol, and agitated at 50 rpm at 4 °C for 15 minutes. The lysates were further sonicated at 4 °C for 3 cycles (2 second ON and 3 second OFF) under 30% amplitude. This is the nuclear fraction which was collected and analyzed along with the cytoplasmic fraction by Western blotting.

4.1.5 *In vitro* scratch assay

Human melanoma cells, A375, and human gastric cancer cells, AGS, were adjusted to 1×10^5 cells/ml in growth medium and seeded in a 12-well plate, and grown till the confluency reached around 90-95%. The monolayers were scratched with a 10 µL sterile pipette tip, and the spent medium was removed, followed by rinsing with PBS twice to remove dead cells and debris. Next, the scratched monolayers of cells were treated with recombinant proteins for two days in a protein transduction medium at standard cell culture conditions. Glycerol buffer containing 20 mM phosphate buffer in standard cancer cell growth media was used as a negative control. Images were captured at regular intervals using an inverted bright field microscope (ZOE

Fluorescent Cell Imager, Bio-Rad, California, USA) at 20 × magnification, and the migration rate was analysed using ImageJ software. The migration percentage was calculated using the following formula.

$$Migration\ (\%) = \frac{(initial\ area - final\ area)}{initial\ area} \text{ X } 100$$

4.1.6 Cell proliferation

Cell proliferation assay was carried out to observe the tumor suppression capability of purified recombinant proteins on A375 and AGS cells which were procured from NCCS, Pune. Cells were adjusted at $1x10^5$ cells/ml and seeded in 6 wells of a 12-well plate. Post-adherence, cells were treated with 400 nM of recombinant fusion proteins. The experiment was performed in triplicates. After every 48 h of treatment, the cell number of each well was counted and replated at the original seeding density.

4.1.7 Cell cycle analysis using Propidium Iodide (PI) staining following flow cytometry

Measurement of cellular DNA content and the analysis of the cell cycle can be performed by flow cytometry. In addition to determining the relative cellular DNA content, flow cytometry also identifies the cell distribution during the various phases of the cell cycle. Four distinct phases could be recognized in a proliferating cell population: The G0/G1-, S- (DNA synthesis phase), G2- and M-phase (mitosis). However, G2- and M-phase, which both have identical DNA content, could not be distinguished based on their differences in DNA content. A375 cells were seeded at a density of $3x10^5$ cells/well in 6 well plates in cancer cell growth media. Post adherence, cells were treated with 400 nM of HTN-PAX4 and PAX4-NTH protein for the next 6 days, replenishing with fresh protein media after 24 h. Cells were dislodged using 0.1% Trypsin/EDTA after completion of treatment and fixed with ice-cold 70% ethanol overnight. Post fixation, cells were washed with ice-cold PBS to remove ethanol, treated with RNase, and

incubated at 37 ºC for 30 minutes. Next, the cells were stained with PI (50 µg/ml) and incubated in the dark for further 30 minutes following flow cytometry analysis in BD FACS Calibur flow cytometer.

4.1.8 Determination of the apoptotic and/or necrotic cells

Detection and differentiation among the apoptotic and necrotic cellular populations were performed using Annexin V-FITC/PI-based assay. The Annexin V-FITC/PI apoptotic detection kit obtained from Invitrogen was utilized to determine the early-apoptotic, late-apoptotic and necrotic cells using flow cytometry analysis. Briefly, 0.5 x 10^5 AGS cells/well were plated in a 6-well plate and were subsequently treated for 6 days with 400 nM of HTN-PDX1. After 6 days, cells were trypsinized, washed with chilled PBS, and incubated with 5 µl of Annexin V-FITC and 3 µl of PI, mixed in 200 µl of 1X-Annexin V binding buffer. Following incubation, the cells were washed carefully thrice with chilled PBS, and the cells were analyzed using a CytoFlex flow cytometer (Beckman Coulter). Moreover, the fluorescence data were collected in the green (FL1) and red (FL2) channels. Conclusively, the data analysis and fluorescence compensation were performed using the CytExpert software.

4.1.9 RNA isolation, cDNA synthesis and quantitative RT-PCR (RT-qPCR)

The total RNA was extracted using the TRIzol reagent (Invitrogen) from a 6 cm dish and dissolved in RNase-free water. Genomic DNA was removed by digestion with DNAseI (Invitrogen). Subsequently, the isolated RNA was quantified by spectrophotometry. 1 µg of RNA was used for the synthesis of cDNA using a commercially available iScript cDNA Synthesis Kit (BIO-RAD). Subsequently, the expression of genes using real-time-quantitative polymerase chain reaction (RT-qPCR) was measured. qPCR was executed for 40 cycles; the protocol included denaturation at 95 °C for 1 minute; amplification at 95 °C for 15 s, and 58 °C for 15 s; elongation at 72 °C for 45 s; final extension at 72 °C for 10 minutes. The thermocycler

used in RT-qPCR was QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific). RT-qPCRs were performed in triplicates and normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in the same run. Data were analyzed using the $2^{-\Delta\Delta CT}$ method. The sequence of primers used in this assay was specific for human insulin [forward primer (F) 5′-TCACACCTGGTGGAAGCTC-3′, reverse primer (R) 5′-ACAATGCCACGCTTCTGC-3′] and GAPDH [forward primer (F) 5′-GTCTCCTCTGACTTC AACAGCG-3′, reverse primer 5′-ACCACCCTGTTGCTG TAGCCAA-3′]

4.1.10 Western blotting

The Western blotting was performed as described earlier (**Chapter 2, Materials and Methods section** for details). All antibodies used and their respective concentrations are listed in the Appendix.

4.1.11 Statistical analysis

The cell culture experimental data obtained were analyzed by t-test or one-way ANOVA using GraphPad Prism 8 software. Values are expressed as mean ± SE. $P<0.05$ was considered significant.

4.2 Results and discussion

4.2.1 Stability of fusion proteins at standard cell culture conditions

Protein stability assays are used to evaluate the shelf life/longevity of a protein in solution (here cancer cell media) and the stability of purified recombinant proteins (Bosnali and Edenhofer 2008; Borgohain et al. 2019). Thus, this assay was carried out to determine the stability of the purified recombinant proteins in transduction media containing 200 nM of purified fusion proteins (400 nM for HTN-PAX4 and PAX4-NTH) in DMEM supplemented with 5% FBS and 1% P/S. The prepared media was next aliquoted and incubated at standard cell culture

conditions (37 ºC, 5% CO_2) for 24, 48, 72 h. Samples collected at each interval were centrifuged, and the supernatant containing the native protein was analyzed by immunoblotting using an anti-His antibody (Figure 4.1). The result clearly demonstrated that all the fusion proteins except NGN3-NTH induced in LB media remained stable in cell culture growth media up to 72 h (Figure 4.1), whereas NGN3-NTH fusion protein induced in LB media remained stable for 48 h. Cancer cell media (without the recombinant protein) were also incubated along with protein transduction media. This served as a negative control, and as expected, we did not observe any signal/bands in the control samples. This assay attests that the purified recombinant proteins are stable in cell culture media and hence, can be used in further studies to demonstrate various biological activities of the recombinant proteins.

4.2.2 Protein transduction, immunofluorescence and microscopy

Further, the sub-cellular and sub-nuclear delivery of the purified recombinant fusion proteins into PANC-1 cells was examined. In order to study the transduction ability of these purified proteins, PANC-1 cells were incubated with fusion proteins (200 nM for HTN-PDX1, NGN3-NTH and MAFA-NTH and 400 nM for HTN-PAX4 and PAX4-NTH) or glycerol buffer (vehicle control) for 12 h and washed with PBS to remove extracellularly bound protein. The transduced cells were then visualized and analyzed using fluorescence microscopy. Microscopy results confirmed that the vehicle control did not lead to any false positive signal during analysis (Figure 4.2). The TAT-mediated protein transduction resulted in the efficient transduction of all of the fusion proteins into PANC-1 cells (Figure 4.2).

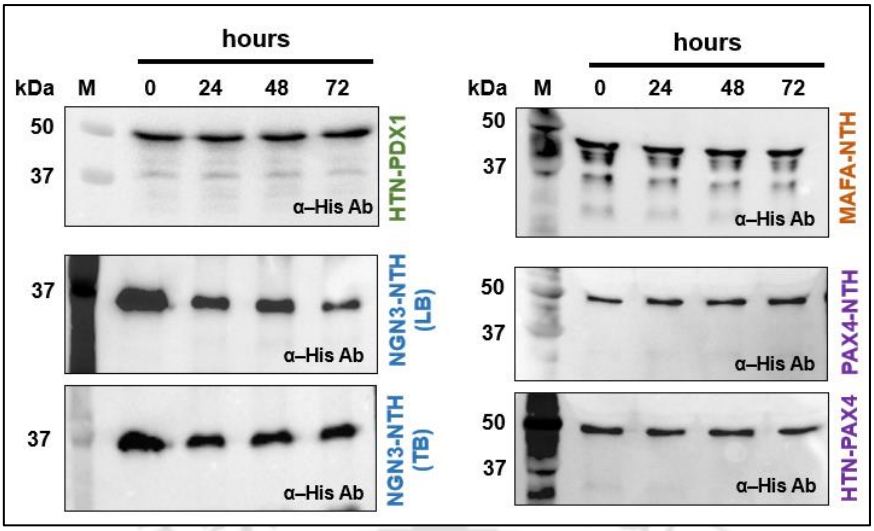


Figure 4.1 Stability of purified recombinant fusion proteins at standard cell culture conditions. The stability of purified recombinant proteins was analyzed by diluting the protein glycerol stock in the protein transduction medium at standard cell culture conditions for 72 h. Samples collected at respective time points were analyzed using Western blotting.

Interestingly, almost all the cells showed successful uptake of purified proteins into their cytoplasm as well as to the nucleus (Figure 4.2). Being transcription factors, translocation of PDX1, NGN3, MAFA and PAX4 fusion proteins to the nucleus is critical for its biological activity. Thus, our strategy using TAT and NLS fused to these recombinant proteins resulted in successful sub-cellular and sub-nuclear delivery into the PANC-1 cells. The fusion of these two tags (TAT and NLS) with the recombinant protein can deliver the protein of interest to the target site to perform its biological function. Moreover, we did not require a protein transduction reagent to deliver fusion proteins into cells efficiently. Other studies have also reported similar results using TAT-mediated cell penetration and NLS-mediated nuclear translocation for other stem cell-specific transcription factors to derive desired cells for biological applications (Bosnali and Edenhofer 2008; Thier et al. 2011, 2012; Peitz et al. 2014). These studies have also reported that the presence of these fusion tags did not affect the

biological activity of the protein of interest. This biological tool can be used to derive integration-free cells and for other biological applications, avoiding the serious concerns associated with plasmid or viral integration into the host cell chromosomes (Borgohain et al. 2019; Kumar et al. 2019).

4.2.3 Sub-cellular fractionation

Further, we characterized the distribution of recombinant fusion proteins in the cytosolic and nuclear fractions of PANC-1 cells by performing subcellular fractionation. PANC-1 cells do not express any of these proteins endogenously. GAPDH and Histone H3 were used as internal controls to confirm cytosolic and nuclear fractions, respectively. In the case of HTN-PDX1, HTN-PAX4 and PAX4-NTH, these human recombinant fusion proteins fractionated exclusively/majorly in the nuclear fraction, which is required to perform its function (Figure 4.3A, D). For NGN3-NTH, the results suggest that the NGN3-NTH fusion proteins were also observed in the nuclear fraction of the PANC-1 cells (Figure 4.3B). However, a few background bands (shown by # in Figure 4.3B) were also detected at different sizes, which were also observed in the vehicle control. NGN3-NTH fusion proteins were detected at ~36 kDa, similar to the expected size. In the case of HTN-MAFA, the majority of the fusion protein was observed in the nuclear fraction; however, unlike other recombinant proteins, a minute amount of protein was also present in the cytosolic fraction (Figure 4.3C).

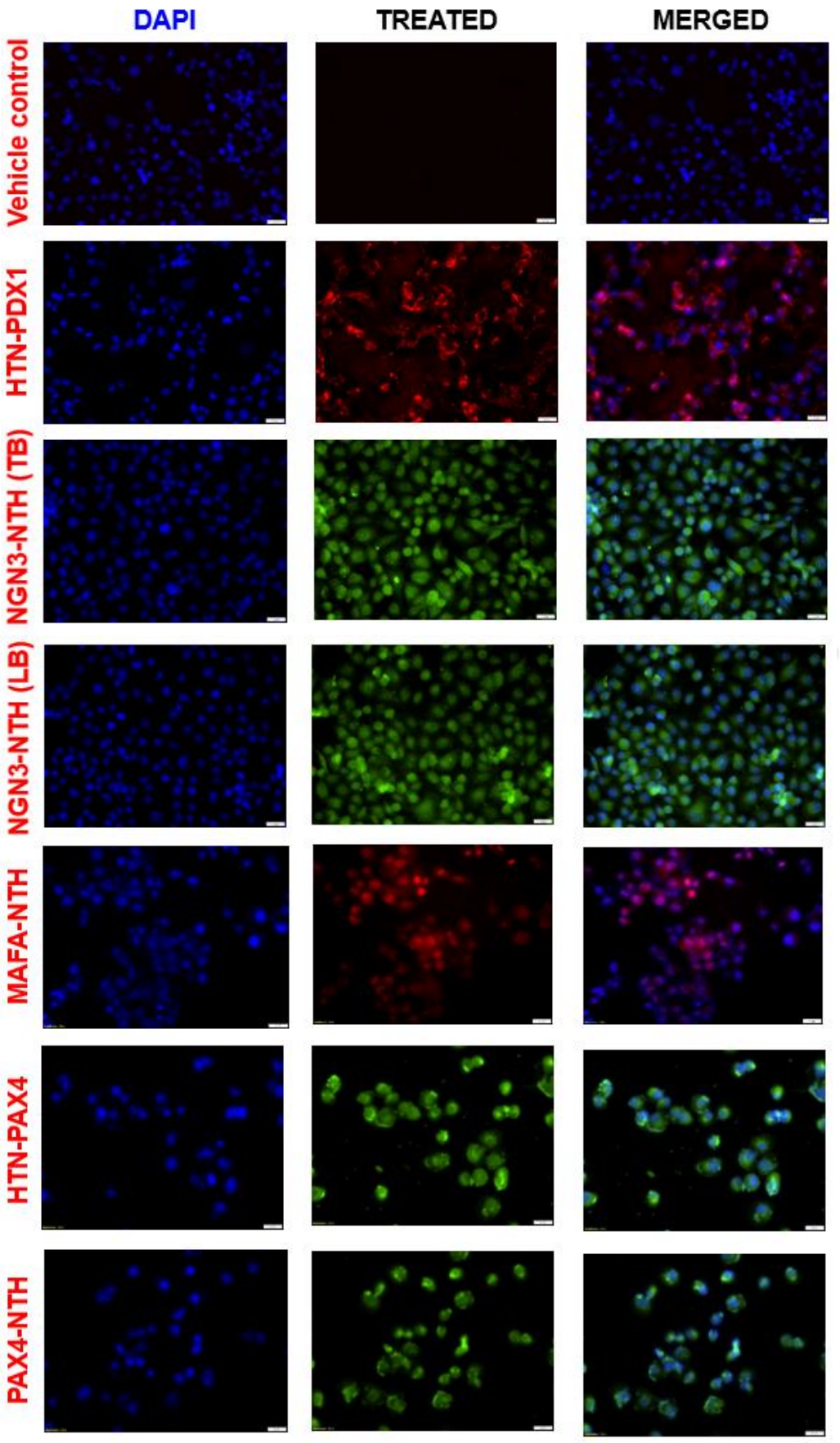

DAPI
TREATED
MERGED
Vehicle control
HTN-PDX1
NGN3-NTH (TB)
NGN3-NTH (LB)
MAFA-NTH
HTN-PAX4
PAX4-NTH

Figure 4.2 Transduction ability of purified recombinant human fusion proteins in a human pancreatic cancer cell line. PANC-1 cells (5×10^5 cells/well) were plated and grown to 50% confluency, and purified proteins (200 nM) or an equal volume of glycerol buffer (vehicle control) was added to the cells and incubated for 12 h under standard cell culture conditions. After the incubation, cells were washed with PBS, fixed, permeabilized, and incubated with primary antibody followed by secondary antibody. Nuclei were counterstained with DAPI. Cellular uptake and nuclear translocation ability were analyzed using an inverted fluorescence microscope.

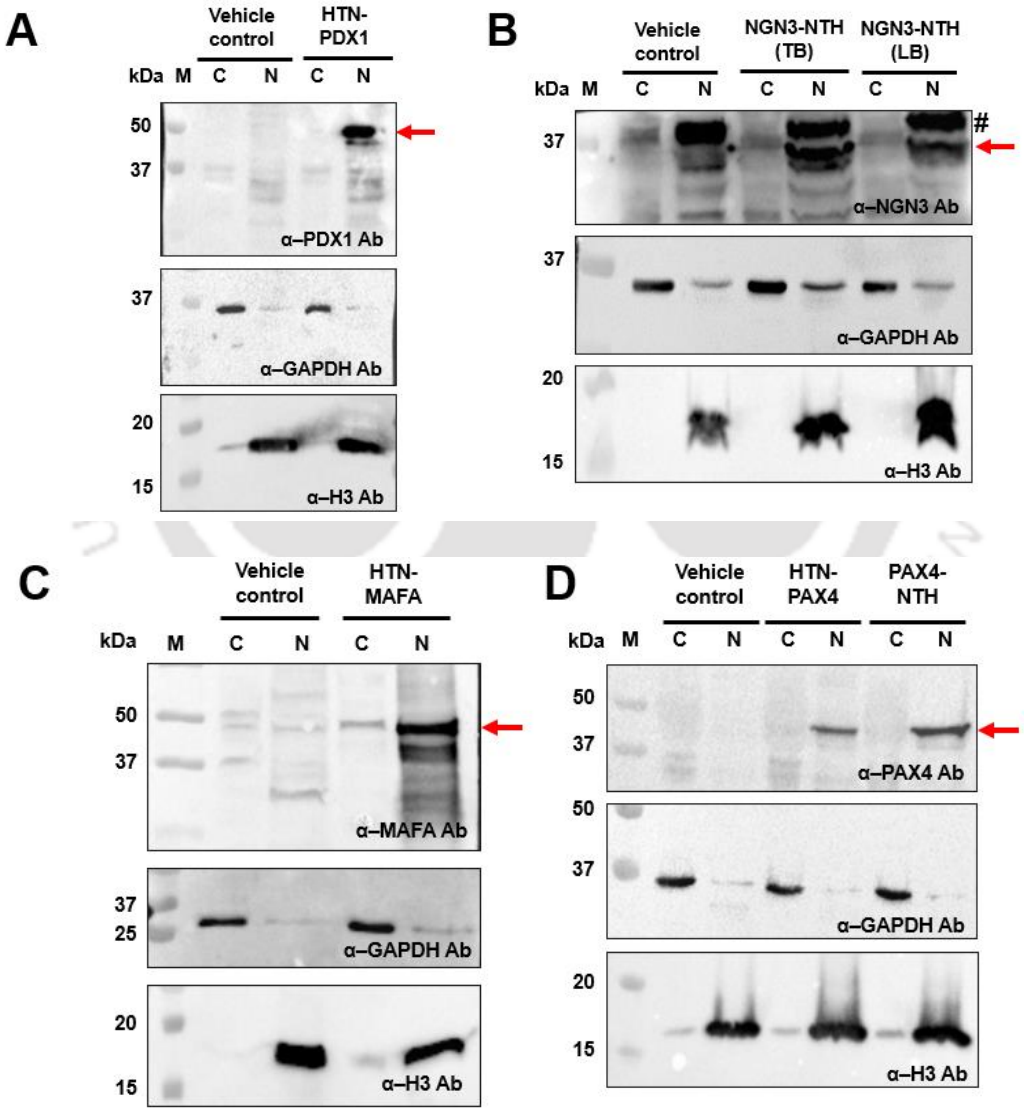


Figure 4.3 Determination of localization of recombinant fusion proteins using subcellular fractionation assay. The representative Western blotting demonstrates the cytoplasmic (C) and nuclear (N) fractions of recombinant fusion protein-treated or untreated (Vehicle control) PANC-1 cells. The internal control antibodies used were against GAPDH, and Histone H3, detecting bands at ~36 kDa and ~17 kDa, respectively. M protein marker (kDa), C cytosolic fraction, N nuclear fraction

4.2.4 Effect of NGN3 fusion protein on cell viability

Recombinant proteins applied to mammalian cells should not induce cytotoxicity to study their biological role. Therefore, the effect of purified NGN3 fusion protein transduction on the viability of PANC-1 cells was investigated. Cells were cultured with different concentrations (100, 200, and 400 nM) of NGN3-NTH protein for 24 h, and the cytotoxic effect produced by the protein was analyzed by MTT assay. The results indicate that the viability of PANC-1 cells was not affected by NGN3-NTH protein transduction (Figure 4.4). Thus, confirming that purified NGN3 fusion protein did not induce cell cytotoxicity and is safe for cell culture analysis. Since all three (100, 200, and 400 nM) concentrations tested were found to be non-toxic, 200 nM of NGN3-NTH fusion protein was used for subsequent experiments.

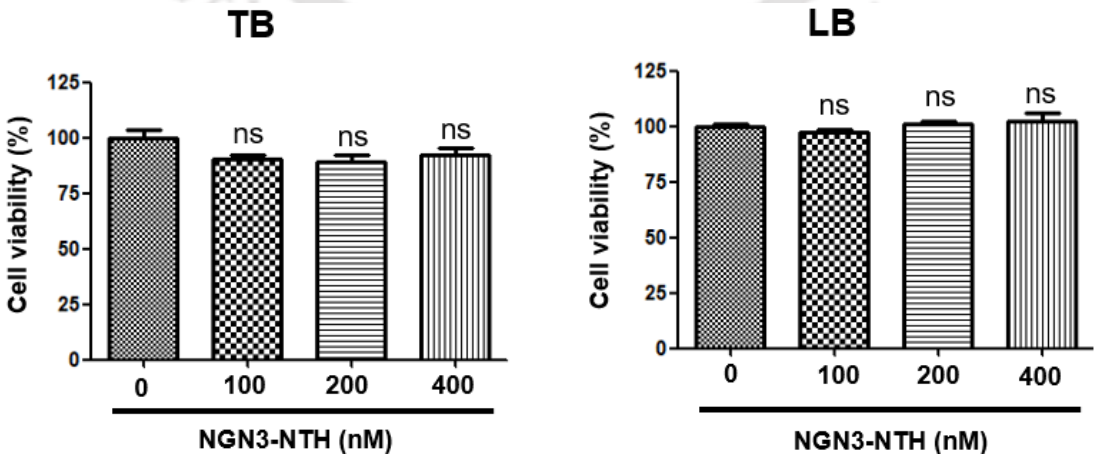


Figure 4.4 Effect of NGN3-NTH fusion protein transduction in PANC-1 cells. 8000 cells/well in a 96-well cell culture plate were cultured under cell culture conditions. Cells were then treated with different concentrations of recombinant NGN3-NTH protein or vehicle control (20% glycerol in phosphate buffer) for 24 h. After the completion of treatment, MTT was added to the cells for 2 h, and subsequently, DMSO was added to dissolve the crystals. Next, the absorbance was measured at 570 nm using a multi-plate reader to calculate the cell viability (%). A bar graph was plotted with the calculated values, and statistical tests (one-way ANOVA) were carried out using GraphPad Prism 5 software. ns: not significant

4.2.5 Functional assessment of purified recombinant fusion proteins

In the previous chapter (**Chapter 3**), we showed that the purified fusion proteins had retained their secondary structure, and in this chapter, we demonstrated their cell penetration and nuclear translocation ability. Next, we carried out multiple experiments to further examine their functional activity. Therefore, to validate the biological activity of these fusion proteins, we investigated their effect on mammalian cells.

4.2.5.1 Effect of purified HTN-PDX1 fusion proteins on mammalian cells

Reduction in cell migration of AGS cells upon application of HTN-PDX1 fusion protein

An earlier study reported PDX1 to function as a putative tumor suppressor protein in human gastric cancer cells by reducing tumor growth (Ma et al. 2008). The study transfected PDX1 into multiple gastric cancer cell lines and performed cell proliferation, migration, colony formation assays, etc., to demonstrate their findings. Therefore, to understand the effect of its recombinant form on human gastric cancer (AGS) cells *in vitro*, we also carried out a cell migration (scratch) assay, where we treated AGS either with 200 nM of the PDX1 fusion protein or glycerol buffer (negative control) for 48 hours. The experiments were performed in

triplicates for both the vehicle control and the treated cells. A significant decrease (~20%) in the migration ability of the PDX1 protein-treated cells post-24 hours was observed (Figure 4.5A). Moreover, treatment with PDX1 fusion protein for an additional 24 hours further reduced the migration ability of the treated cells significantly when compared to untreated cells (vehicle control). Microscopic images of the same are provided in Figure 4.5B. A similar observation was also reported in an earlier study in gastric cancer cells (Ma et al. 2008). In the earlier study, Ma and the group demonstrated that the migration ability of PDX1 transfected SGC7901 cells was reduced over 24 hours of treatment (Ma et al. 2008), which is in congruence with our findings.

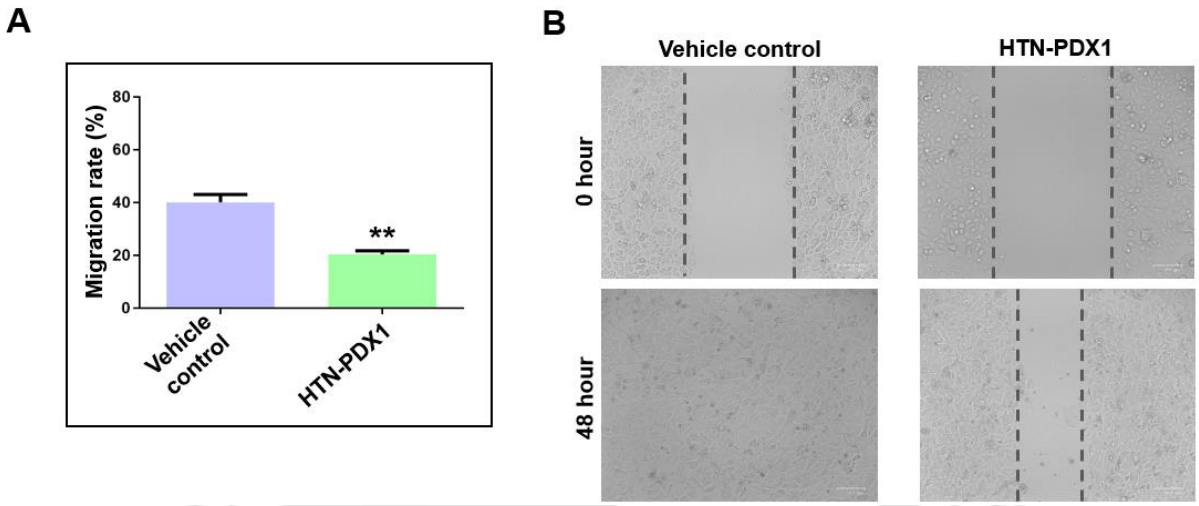


Figure 4.5 Assessment of migration ability of HTN-PDX1 fusion protein-treated AGS cells. (A) Graphical representation of the treated cells compared to its counterparts (Vehicle control) **(B)** Scratch assay to determine reduced migration rate in HTN-PDX1 treated AGS cells. Scale bar, 100 μm. n = 3. **P<0.001.

Suppression of cell proliferation upon recombinant HTN-PDX1 treatment on AGS cells

Apart from affecting cell migration ability, PDX1 was also reported to suppress the proliferation of gastric cancer cells (Ma et al. 2008). Hence, a cell proliferation assay was

carried out using AGS cells to understand the effect of the PDX1 fusion protein on the propagation of human gastric cancer cells *in vitro*. AGS cells were seeded in a 12-well plate (100,000 cells/well) and post adherence; cells were treated with protein transduction media containing 200 nM of either PDX1 fusion protein or glycerol buffer (negative control). After 48 hours of treatment, cells were counted using a hemocytometer and replated at the original seeding density with fresh protein culture media. The cumulative cell number was plotted against the number of days of treatment. The results clearly demonstrated that the cell proliferation rate in protein-treated cells was significantly (P<0.05) slower than the untreated (vehicle control) cells on the 6th day of treatment (Figure 4.6). Similar results were observed by Ma and co-workers when they observed a negative correlation between PDX1 and Ki-67, which is a commonly used cell proliferation marker, in gastric cancer cells upon 72 hours of treatment (Ma et al. 2008). This signifies that the recombinant version of the PDX1 fusion protein is equally potent in decreasing the cell proliferation rate in gastric cancer cells.

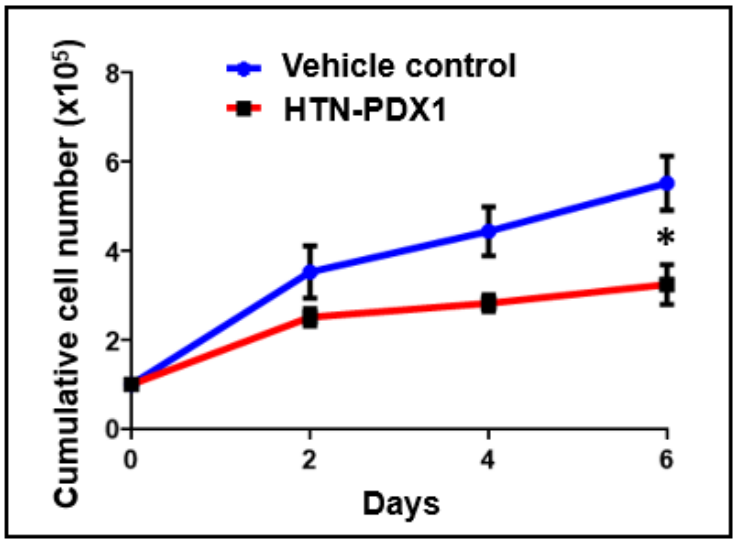


Figure 4.6 Effect of HTN-PDX1 fusion protein on the proliferation rate of AGS cells. Cell proliferation growth curves for AGS cells in protein transduction media or 5% control media at standard cell culture conditions for 6 days (P<0.05). A line graph was plotted with the

calculated values, and statistical tests (non-parametric t-test) were carried out using GraphPad Prism 5 software. n=3. *P<0.05.

Induction of G2/M phase arrest and determination of apoptosis upon application of HTN-PDX1 fusion protein on AGS cells

Cell cycle arrest results in the inhibition of cell growth. In order to investigate whether the application of the recombinant form of PDX1 can induce cell cycle arrest, we treated AGS cells for 3 days with 200 nM of the PDX1 fusion protein and analyzed the cell cycle profile using PI staining. We observed a 10% increase in the number of G2/M phase-arrested cells with PDX1 fusion protein-treated cells (Figure 4.7A). Various studies have reported that blocking the G2/M phase at cell cycle progression correlated with DNA damage (Shimizu et al. 1998; Chung et al. 2017). Subsequently, treatment with PDX1 fusion protein was continued for further 3 days and flow cytometry analysis was carried out to determine the percentage of apoptotic cells. Upon 6 days of treatment, we observed a substantial increase in the number of apoptotic cells in comparison to the control untreated cells, which is clearly depicted in Figure 4.8B, C. Interestingly, no significant changes were observed in the population of necrotic cells, following treatment with PDX1 fusion protein. Altogether, this study suggests the plausible role of PDX1 fusion protein in inducing apoptosis in gastric cancer cells.

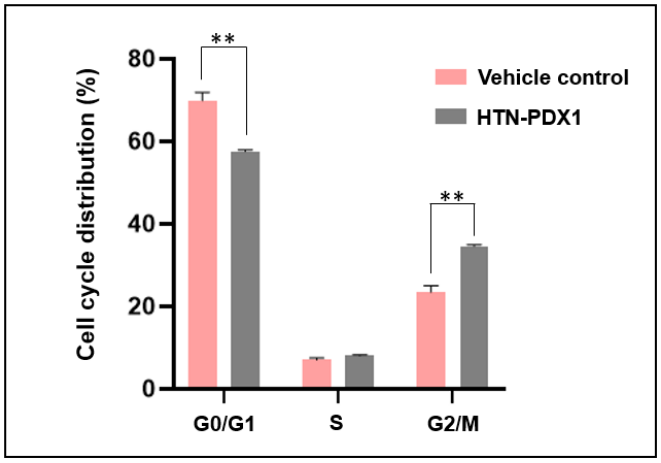


Figure 4.7 Effect of HTN-PDX1 fusion protein on the cell cycle of AGS cells. The G2/M phase population was significantly increased in PDX1-treated cells ($P<0.05$). A bar graph was plotted with the calculated values, and statistical tests (one-way ANOVA) were carried out using GraphPad Prism 5 software. *: significant.

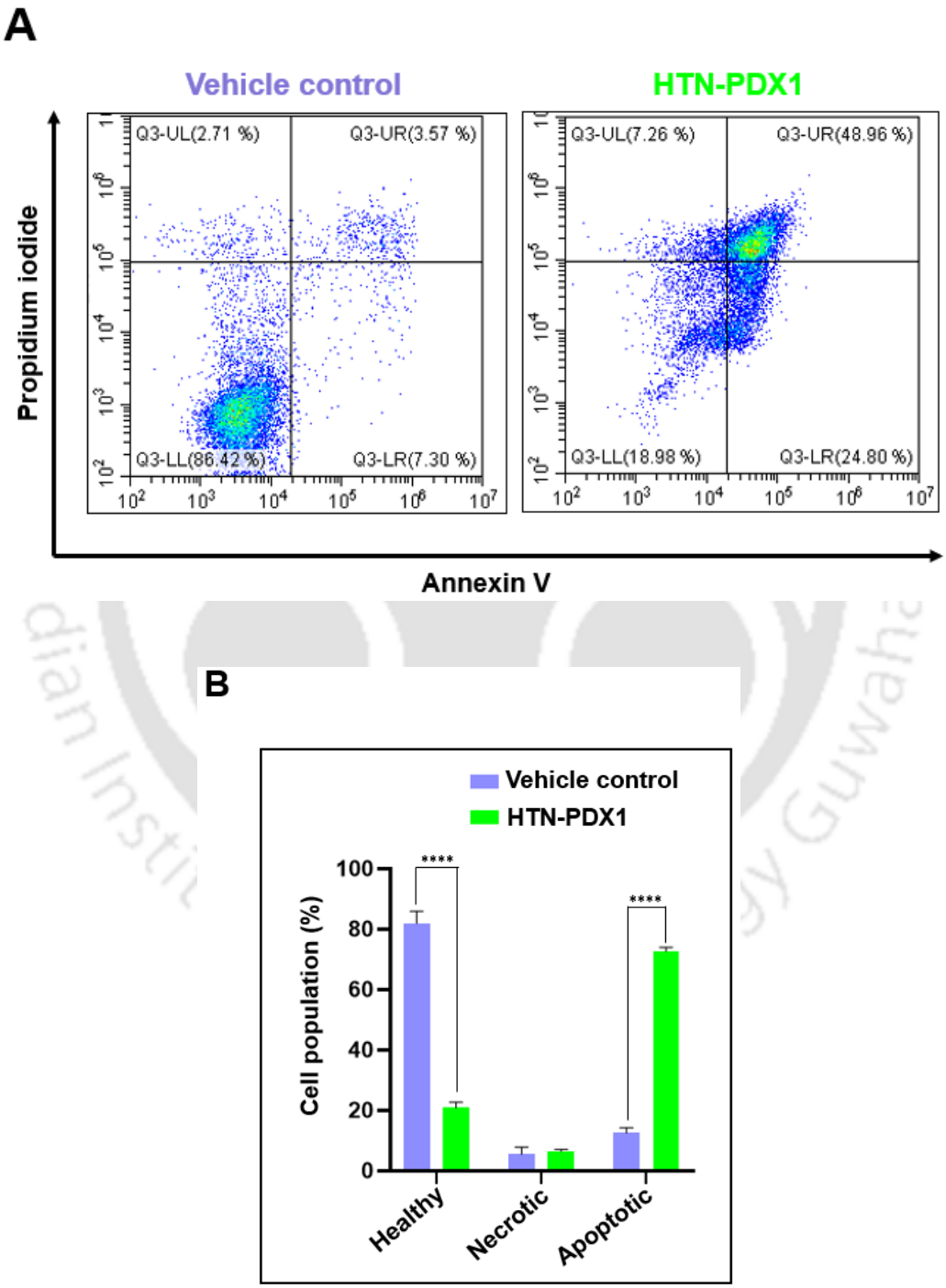


Figure 4.8 Determination of apoptosis in HTN-PDX1 fusion protein-treated AGS cells (A) Pictorial representation of apoptotic cell **(B)** Quantitative cell cycle distribution data of **(A).** The quantitative data shown are mean ± SEM (n = 3). ****$P < 0.0001$.

Induction of insulin gene expression in PANC-1 cells upon application of HTN-PDX1 fusion protein

PDX1 was reported to bind directly to the insulin promoter and upregulate insulin gene expression (Wang et al. 2018b). Therefore, to determine whether our recombinant version of PDX1 protein can also induce insulin gene expression, we transduced PANC-1 cells with PDX1 fusion protein (200 and 400 nM) for consecutive 5 days and analyzed the fold change in insulin mRNA using real-time PCR. PANC-1 cells were chosen for this analysis as these cells do not express insulin (Noguchi et al. 2006). A four-fold change in the insulin mRNA level in the PDX1 fusion protein-treated cells was observed in this study (Figure 4.9). To further identify whether increasing the concentration of protein would result in higher insulin expression, we also treated cells with 400 nM of PDX1 fusion protein for 5 days. Interestingly, we did not observe any substantial difference in the insulin mRNA expression upon increasing the concentration of protein (Figure 4.9). A previous study by Noguchi et al., 2006, also demonstrated that PDX1 is capable of inducing insulin mRNA expression in both PANC-1 and HN-5 cells, which is a mouse monoclonal stem cell line, when transfected adenovirally for 7 days (Noguchi et al. 2006). They reported a three-fold increase in insulin mRNA expression, similar to our findings. This corroborated that the recombinant PDX1 fusion protein is biologically active and is at par with its viral counterpart.

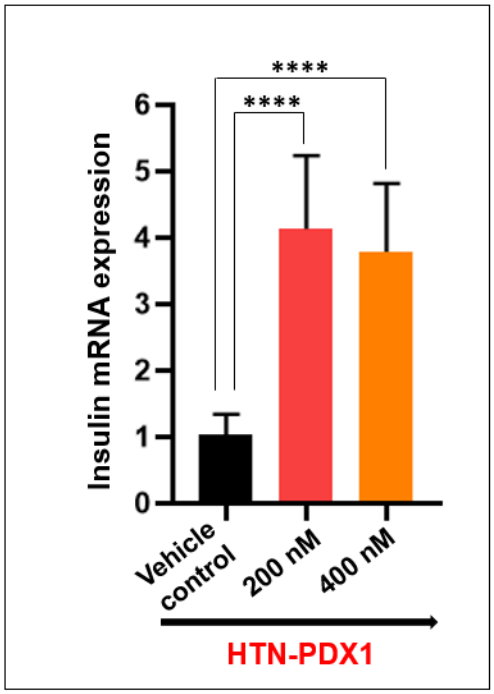


Figure 4.9 Effect of HTN-PDX1 fusion protein in the activation of insulin gene expression in PANC-1 cells. Induction of insulin gene is observed upon HTN-PDX1 treatment. Quantitative statistics are shown as the mean ± SEM (n = 3). ****$P < 0.0001$.

4.2.5.2 Effect of purified NGN3-NTH fusion proteins on human pancreatic ductal cells

Exogenously expressed NGN3-NTH induced insulin expression in PANC1 cells

Next, we sought to demonstrate the biological activity of NGN3-NTH fusion proteins *in vitro*. To accomplish this, the PANC-1 cells were transduced with NGN3-NTH fusion proteins obtained from both LB and TB media. As reported earlier, the PANC-1 cells do not have endogenous insulin expression (Noguchi et al. 2006). Also, the insulin gene is a downstream target of NGN3 during human fetal endocrine development (Lyttle et al. 2008). Thus, the nuclear uptake of NGN3-NTH protein in PANC-1 cells should induce insulin gene expression. To verify this, we transduced the PANC-1 cells with 200 nM NGN3-NTH fusion protein for 72 h and analyzed the insulin mRNA levels by RT-qPCR. We observed that the treatment of the recombinant protein obtained from TB and LB media induced the insulin mRNA levels in the PANC-1 cells by ~three-fold and four-fold, respectively. (Figure 4.10). Noguchi and group

showed similar results when the NGN3 was ectopically expressed in the PANC-1 cells (Noguchi et al. 2006). Additionally, previous studies have reported that the ectopic expression of NGN3 in human ductal cell lines not only induced insulin gene expression but also played a crucial role in the transdifferentiation of ductal cells into insulin-expressing cells (Heremans 2003; Gasa et al. 2004). Our study provides similar results of early signs of conversion of human pancreatic ductal cells into insulin-expressing cells. Thus, it elucidates the biological activity of the purified NGN3-NTH fusion proteins in PANC-1 cells. However, further detailed experiments are required to confirm insulin expression at the protein level and elucidate its function in β-cell generation.

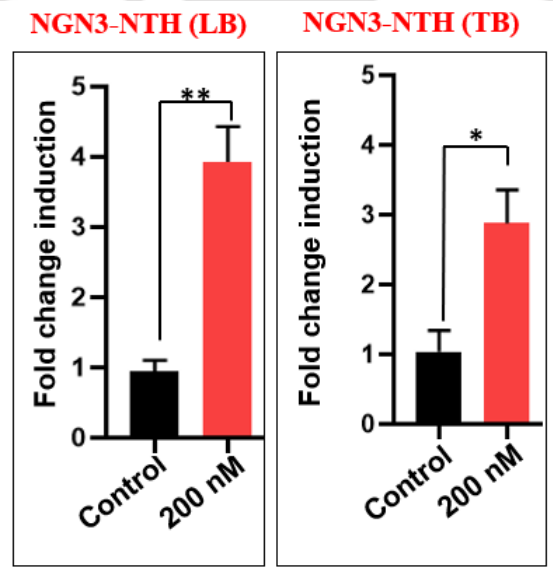


Figure 4.10 Effect of NGN3-NTH fusion protein in the activation of insulin gene expression in PANC-1 cells. The PANC-1 cells were transduced with NGN3-NTH fusion protein obtained from TB and LB media and the levels of insulin mRNA were measured using RT-qPCR. Cells treated with vehicle (glycerol buffer) were taken as a control. (n = 3). *P** < 0.05, *P*** < 0.01.

4.2.5.3 Effect of the recombinant MAFA-NTH protein on insulin gene expression

We carried out a similar experiment to demonstrate the biological activity of MAFA-NTH recombinant protein on mammalian cells. To achieve this, the MAFA-NTH fusion protein was transduced into PANC-1 cells. It was previously reported that PANC-1 cells do not have endogenous insulin expression (Noguchi et al. 2006), and MAFA directly affects insulin gene expression by binding to the MARE element present on the insulin gene (Kataoka et al. 2002; El Khattabi and Sharma 2015). As a result, treatment of cells with MAFA-NTH recombinant protein should upregulate insulin expression. To corroborate this, PANC-1 cells were treated with 200 nM of purified recombinant MAFA-NTH for 5 days and the fold change in insulin mRNA was analyzed by RT-qPCR. We observed approximately a three-fold change in the insulin mRNA level in the treated cells compared to their counterparts (Figure 4.11). A similar experiment was carried out by Kataoka and group where they demonstrated that MAFA alone could strongly activate insulin gene expression in NIH3T3 cells, and our findings are in line with the previous findings (Kataoka et al. 2002). A study conducted by Lu and co-workers reported the purification of recombinant mouse MAFA protein from the BL21(DE3) strain. However, the protein failed to efficiently translocate to the nucleus, so the activation of downstream target genes was significantly impaired (Lu et al. 2012). This could be due to the lack of addition of NLS along with the full-length coding sequence. Thus, we have successfully purified a bioactive form of recombinant MAFA from *E. coli* from soluble fractions under native conditions.

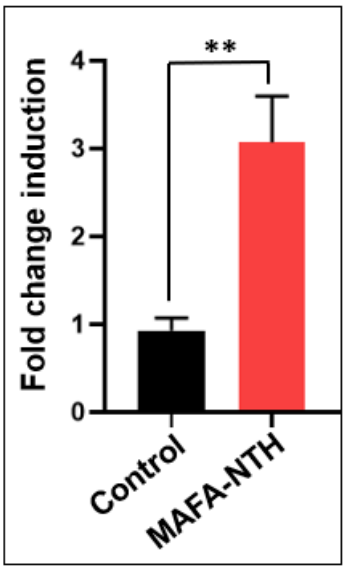


Figure 4.11 Effect of MAFA-NTH fusion protein in the activation of insulin gene expression in PANC-1 cells. Induction of insulin gene is observed upon MAFA-NTH treatment. Quantitative statistics are shown as the mean ± SEM (n = 3). *P*** < 0.01.

4.2.5.4 Effect of purified PAX4 fusion proteins on human melanoma cells

Decreased cell proliferation in recombinant HTN-PAX4 and PAX4-NTH-treated human melanoma cells

PAX4 protein is reported to act as a tumor suppressor protein for human melanoma cells by reducing tumor growth (Hata et al. 2008). The study further reported that this retardation in tumor progression was due to decreased DNA synthesis and arrest at the G0/G1 phase (Hata et al. 2008). Therefore, to understand the effect of the purified recombinant HTN-PAX4 and PAX4-NTH proteins on the propagation of A375 human melanoma cells, *in vitro*, we carried out a cell proliferation assay, where A375 cells at a seeding density of 100,000 cells/well were seeded in a 12-well plate. The experiments were carried out in triplicates. Post-adherence, growth media was replaced with protein transduction media containing 400 nM of either HTN-PAX4 or PAX4-NTH recombinant proteins along with 5% FBS and 1% P/S and incubated for

the next 48 hours. For the control wells, cells were maintained at low serum (5% FBS) containing cancer growth media. After 48 hours, cells from the treated as well as control wells were counted using a hemocytometer and replated with renewed protein transduction media (in case of treated wells) and low-serum media (in case of control wells) at the original seeding density. The cumulative number of cells was plotted against days of treatment. The result clearly demonstrated that the cell proliferation rate in protein-treated cells was comparatively slower than in the control cells. The difference becomes evident on the 6th day of treatment (Figure 4.12). Similar results were observed when MeWo cells, which are of malignant melanoma origin, when these cells were overexpressed with Pax4 (Hata et al. 2008). In accordance with Hata and coworkers, there is a decrease in cell proliferation rate in A375 cells after treatment with purified HTN-PAX4 and PAX4-NTH fusion proteins.

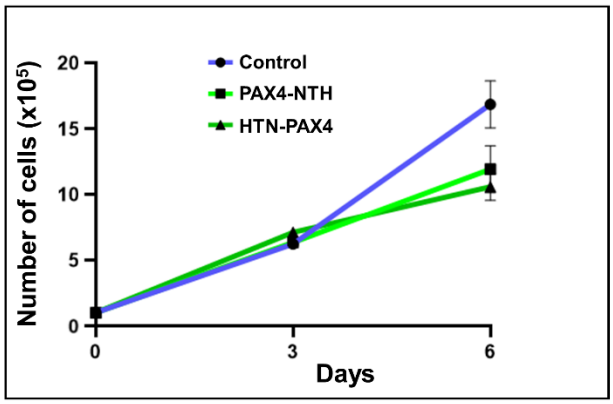


Figure 4.12 Effect of HTN-PAX4 and PAX4-NTH on the proliferation of A375 cells. Cell proliferation growth curves for A375 cells in protein transduction media or 5% control media at standard cell culture conditions for 6 days. A line graph was plotted with the calculated values, and statistical tests (non-parametric t-test) were carried out using GraphPad Prism 5 software.

Effect of the purified recombinant PAX4 fusion proteins on migration rate of human melanoma cells

In the same study, Hata and group reported the over-expression of human PAX4 protein on the migration rate of treated cells and reported that ectopic PAX4 expression in human melanoma cells slowed down the migration rate as compared to the control (Hata et al. 2008). Based on this, we also analyzed the migration rate in A375 cells upon treatment with 400 nM of PAX4 fusion proteins using migration assay. The migration rate of A375 cells was significantly slower in the presence of PAX4 recombinant proteins compared to its counterpart ($p < 0.05$) (Figure 4.13). After 48 hours of treatment, transduced cells showed an overall migration rate of around 35%, whereas the control cells displayed around 70% migration. The migration rate was calculated as stated in an earlier study (Grada et al. 2017; Dey et al. 2021b).

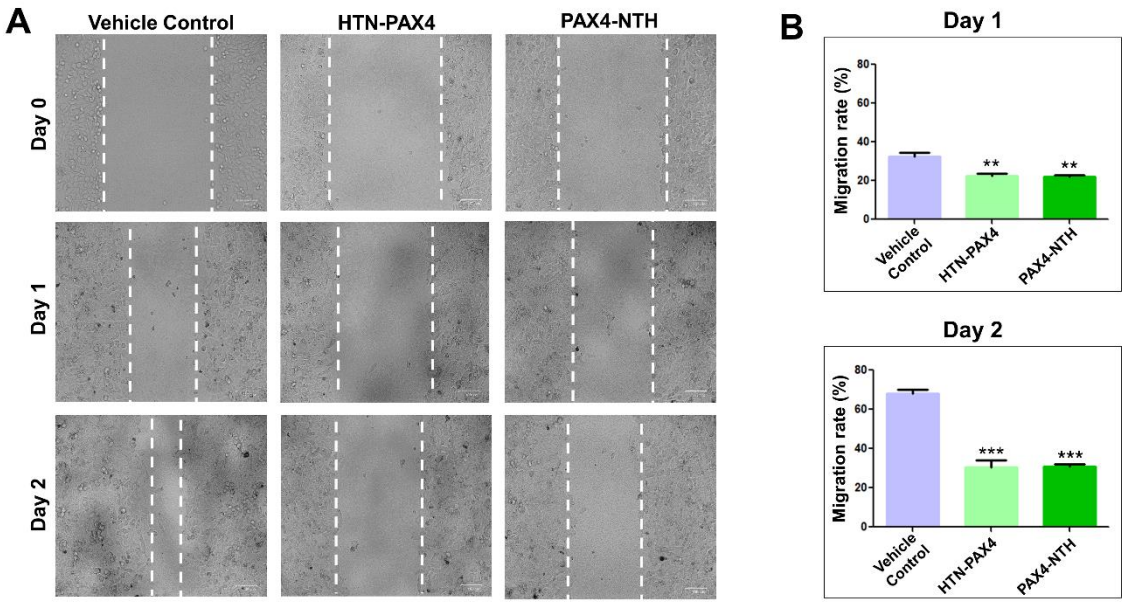


Figure 4.13 Effect of HTN-PAX4 and PAX4-NTH on migration ability of A375 cells. (A) Scratch assay to demonstrate slower migration in the case of PAX4-treated A375 cells. Scale bar, 100 μm **(B)** Graphical representation of the HTN-PAX4 or PAX4-NTH treated cells v/s untreated negative controls. n = 3. **P<0.01, ***P<0.001.

Decreased DNA synthesis and cell cycle arrest due to recombinant PAX4 treatment

To study the effect of PAX4 on cell progression and survival, the cell cycle was monitored after 72 hours of treatment with 400 nm of either of the PAX4 fusion proteins. A cell cycle arrest at G0/G1 phase was noted, whereas the G0/G1 phase fraction was extended, while there was a significant reduction in the S or G2/M phase fraction, as depicted in Figure 4.14. The cells arrested in the G0/G1 phase not only prevented the entry of the cells to the S-phase but also halted or slowed down DNA replication. In general, the increase in G0/G1 phase signifies the induction of apoptosis, a protective cellular mechanism that ensures homeostasis of the host cells, ultimately leading to cell death. Moreover, according to the reports of Hata et al. 2008, the association of PAX4 with retardation in tumor development and malignant progression was ensured in human melanoma *in vitro* (Hata et al. 2008). PAX4 diminished the DNA synthesis and resulted in the G0/G1 phase cell cycle arrest, which is at par with our current findings in A375 human melanoma cells.

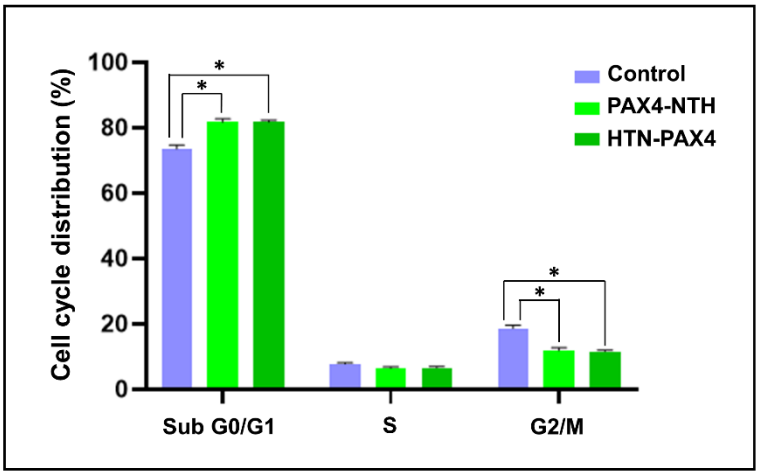


Figure 4.14 Effect of HTN-PAX4 and PAX4-NTH on cell cycle of A375 cells. Quantitative cell cycle distribution of PAX4 treated cells compared to its untreated counterpart. n=3. *P<0.05

4.3 Conclusion

We conclude that all four purified pancreatic-specific fusion proteins retained their stability at standard cell culture conditions for at least 48 h. Next, we demonstrated the internalization capacity of these proteins using immunofluorescence and subcellular fractionation. The microscopic images from immunofluorescence studies showed that the proteins had successfully penetrated the cell membrane and translocated to the nucleus. The findings were further validated using subcellular fractionation studies against protein-specific antibodies. The results again proved that the proteins have internalization capacity in the PANC-1 cells. Some of these purified recombinant proteins (PDX1 and PAX4) are reported to act as tumor suppressors when applied in viral forms. So, we planned to examine if the recombinant forms are equally capable of acting as tumor suppressors. In order to investigate the same, we carried out multiple cellular-based studies on cancer cells, like cell migration ability, cell proliferation rate, cell cycle analysis, etc. When gastric cancer cells (AGS) were treated with 200 nM of HTN-PDX1, the cell migration and proliferation rates were significantly reduced. This could be due to DNA damage, as proved by the cell cycle analysis data, which showed a G2/M phase arrest in the treated cells compared to its untreated counterparts. Additionally, HTN-PDX1 was further determined to induce apoptosis in AGS cells when treated for consecutive 6 days. Similarly, when human melanoma cells (A375) were treated with HTN-PAX4 and PAX4-NTH, proliferation and migration rates were decreased with a G2/M phase arrest. Again, these purified proteins were also reported to induce insulin gene expression in PANC-1 cells. We carried out RT-qPCR to study the same and observed at least a three-fold rise in insulin mRNA expression when treated for 5 days. Thus, we conclude that these purified proteins are biologically active and can provide high therapeutic value in the treatment of diabetes since these are generated using integration-free methods. Moreover, these proteins can also be used as tumor suppressors in the treatment of gastric as well as skin cancer.

Chapter 5

Overall conclusions and future perspective

In our present work in this thesis, we have successfully generated a biologically active pancreatic-specific recombinant protein toolbox. This toolbox can be used for reprogramming various cell sources like terminally differentiated somatic cells and differentiating embryonic stem cells or induced pluripotent stem cells to produce insulin-producing, functional pancreatic β-cells. These reprogrammed cells, generated using the toolbox, can give rise to patient-specific β-cells that can be transplanted into diabetic patients to treat diabetes mellitus. Apart from this, the toolbox thus generated is simple and cost-effective. Moreover, screening of different induction parameters like inducer concentration, bacterial density, induction time, induction temperature, along with the position of fusion tags to maximize the expression level of these recombinant proteins, has hugely proved beneficial for attaining high yield after native purification. Besides, these purified recombinant fusion proteins retained their secondary structure and could also transduce to mammalian cells and localized to the nucleus. We further report the recombinant proteins are biologically active using mammalian cell-based studies. The recombinant version of PDX1 and PAX4 acts as a tumor suppressor in the case of human gastric cancer and human melanoma cell lines where the protein-treated cells showed reduced migration rate and cell proliferation compared to untreated control cells. Furthermore, these pancreatic-specific factors are reported to induce insulin mRNA expression in PANC-1 cells. In order to investigate the effect of the recombinant versions with respect to insulin expression, we treated PANC-1 cells with various concentrations of recombinant proteins. The result demonstrated that PDX1, NGN3 and MAFA-treated cells showed significantly higher insulin mRNA expression, which further proved the biological activity of these purified recombinant

proteins. To conclude, the established pancreatic-specific recombinant protein toolbox can be used to directly reprogram somatic cells to generate directly transplantable insulin-producing β-cells to treat diabetes, which will provide a range of applications in the field of disease modeling and drug development. Moreover, these recombinant proteins are integration-free and, as a result, can yield clinical grade β-cells. Additionally, as mentioned earlier, some of the recombinant proteins, like PDX1 and PAX4, have the tumor-suppressing capability, and this can be studied further to explore their application in the field of cancer biology and understand their mechanism of action in different cancers.

Highlights

1. Determined the appropriate genetic construct and expression conditions for each transcription factor.

2. Establishment of homogeneous purification of these recombinant fusion proteins under native conditions.

3. Purified fusion proteins are stable at cell conditions and have cell penetration and nuclear translocation ability.

4. The purified fusion proteins are biologically active, and the fusion tags did not affect their functionality.

The future application of this recombinant toolbox lies in understanding the complete role played by these pancreatic-specific factors individually or in combination, not only in the formation of β-cells but also with respect to tumor suppression in multiple mammalian cells. Moreover, this toolbox has the potential to replace their genetic versions in the field of cellular reprogramming to generate patient-specific, transplantable β-cells to treat diabetes mellitus.

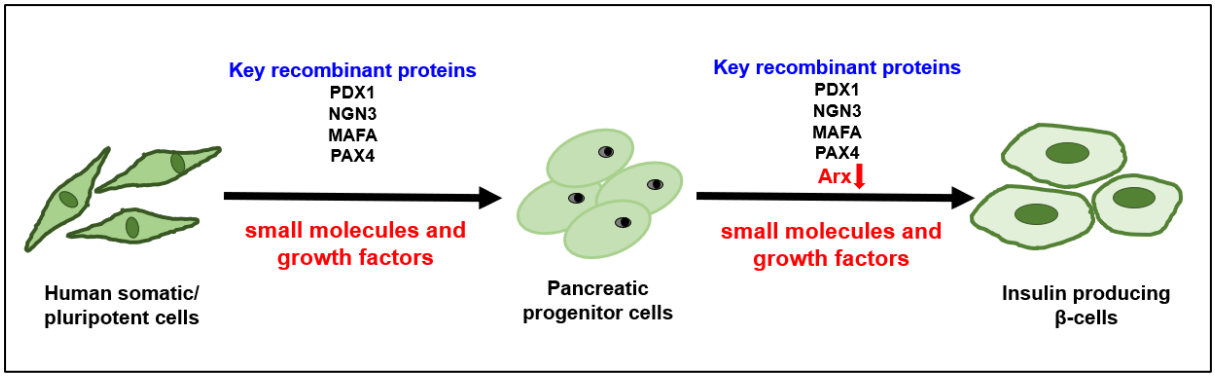


Figure 5.1 Pictorial representation of prospective future applications of the generated recombinant protein toolbox in this study.

List of antibodies used in this study

Antibodies	**Dilutions**			**Company**	**Catalog no.**
	Western blotting	Immunostaining	Subcellular fractionation		
Anti-His	1:5000	N/A	N/A	BioBharati	BB-AB0010
Anti-PDX1	1:1000	1:300	1:1000	Santa Cruz	SC-390792
Anti-NGN3	1:500	N/A	N/A	Santa Cruz	SC-376607
Anti-NGN3	N/A	1:100	1:500	Novus Biologicals	NBP1-47878
Anti-MAFA	1:1000	1:100	1:500	Cell Signalling Technology	D2Z6N
Anti-PAX4	1:250	N/A	1:250	Santa Cruz	SC-376607
Anti-PAX4	N/A	1:50	N/A	R&D systems	AF2614
Anti-H3	N/A	N/A	1:5000	BioBharati	BB-AB0055
Anti-GAPDH	N/A	N/A	1:5000	BioBharati	BB-AB0060

Anti-Rabbit IgG, HRP-conjugated	1:5000	N/A	1:5000	Invitrogen	31460
Anti-Mouse IgG, HRP-conjugated	1:5000	N/A	1:5000	Invitrogen	31430
Anti-Goat IgG, Alexa Flour	N/A	1:1000	N/A	Invitrogen	A11055
Anti-Goat IgG, HRP-conjugated	N/A	1:1000	N/A	Invitrogen	A11037

References

Ackermann AM, Wang Z, Schug J, et al (2016) Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol Metab 5:233–244. https://doi.org/10.1016/j.molmet.2016.01.002

Adda G, Hannoun L, Loygue J (1984) Development of the human pancreas: variations and pathology. A tentative classification. Anat Clin 5:275–283. https://doi.org/10.1007/BF01798752

Ahlgren U, Jonsson J, Edlund H (1996) The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. Development 122:1409–1416. https://doi.org/10.1242/dev.122.5.1409

Aigha II, Abdelalim EM (2020) NKX6.1 transcription factor: a crucial regulator of pancreatic β cell development, identity, and proliferation. Stem Cell Res Ther 11:1–14. https://doi.org/10.1186/s13287-020-01977-0

Akinci E, Banga A, Greder L V., et al (2012) Reprogramming of pancreatic exocrine cells towards a beta (β) cell character using Pdx1, Ngn3 and MafA. Biochem J 442:539–550. https://doi.org/10.1042/BJ20111678

Apelqvist Å, Li H, Sommer L, et al (1999) Notch signalling controls pancreatic cell differentiation. Nature 400:877–881. https://doi.org/10.1038/23716

Assenberg R, Wan PT, Geisse S, Mayr LM (2013) Advances in recombinant protein expression for use in pharmaceutical research. Curr Opin Struct Biol 23:393–402. https://doi.org/10.1016/j.sbi.2013.03.008

Atkinson MA, Eisenbarth GS, Michels AW (2014) Type 1 diabetes. Lancet 383:69–82. https://doi.org/10.1016/S0140-6736(13)60591-7

Baeyens L, De Breuck S, Lardon J, et al (2005) *In vitro* generation of insulin-producing beta cells from adult exocrine pancreatic cells. Diabetologia 48:49–57. https://doi.org/10.1007/s00125-004-1606-1

Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*.

Nat Biotechnol 22:1399–1407. https://doi.org/10.1038/nbt1029

Bastidas-Ponce A, Roscioni SS, Burtscher I, et al (2017) Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic β-cells. Mol Metab 6:524–534. https://doi.org/10.1016/j.molmet.2017.03.007

Benninger RKP, Kravets V (2022) The physiological role of β-cell heterogeneity in pancreatic islet function. Nat Rev Endocrinol 18:9–22. https://doi.org/10.1038/s41574-021-00568-0

Ber I, Shternhall K, Perl S, et al (2003) Functional, persistent, and extended liver to pancreas transdifferentiation. J Biol Chem 278:31950–31957. https://doi.org/10.1074/jbc.M303127200

Berneman-Zeitouni D, Molakandov K, Elgart M, et al (2014) The temporal and hierarchical control of transcription factors-induced liver to pancreas transdifferentiation. PLoS One 9:1–11. https://doi.org/10.1371/journal.pone.0087812

Bonner-Weir S, Taneja M, Weir GC, et al (2000) *In vitro* cultivation of human islets from expanded ductal tissue. Proc Natl Acad Sci U S A 97:7999–8004. https://doi.org/10.1073/pnas.97.14.7999

Borgohain MP, Haridhasapavalan KK, Dey C, et al (2019) An insight into DNA-free reprogramming approaches to generate integration-free induced pluripotent stem cells for prospective biomedical applications. Stem Cell Rev Reports 15:286–313 https://doi.org/10.1007/s12015-018-9861-6

Bornhorst, J.A., Falke J. (2010) Purification of proteins using polyhistidine affinity tags. Methods Enzym 2000:245–254. https://doi.org/10.1016/j.immuni.2010.12.017

Bosnali M, Edenhofer F (2008) Generation of transducible versions of transcription factors Oct4 and Sox2. Biol Chem 389:851–861. https://doi.org/10.1515/BC.2008.106

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254 https://doi.org/10.1016/0003-2697(76)90527-3

Bramswig NC, Everett LJ, Schug J, et al (2013) Epigenomic plasticity enables human pancreatic α to β cell reprogramming. J Clin Invest 123:1275–1284. https://doi.org/10.1172/JCI66514

Braun P, Hu Y, Shen B, et al (2002) Proteome-scale purification of human proteins from bacteria. Proc Natl Acad Sci U S A 99:2654–2659. https://doi.org/10.1073/pnas.042684199

Brun T, Franklin I, St.-Onge L, et al (2004) The diabetes-linked transcription factor PAX4 promotes β-cell proliferation and survival in rat and human islets. J Cell Biol 167:1123–1135. https://doi.org/10.1083/jcb.200405148

Burgess-Brown NA, Sharma S, Sobott F, et al (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expr Purif 59:94–102. https://doi.org/10.1016/j.pep.2008.01.008

Cardinale V, Wang Y, Carpino G, et al (2011) Multipotent stem/progenitor cells in human biliary tree give rise to hepatocytes, cholangiocytes, and pancreatic islets. Hepatology 54:2159–2172. https://doi.org/10.1002/hep.24590

Cataldo LR, Vishnu N, Singh T, et al (2021) The MafA-target gene PPP1R1A regulates GLP1R-mediated amplification of glucose-stimulated insulin secretion in β-cells. Metabolism 118:154734. https://doi.org/10.1016/j.metabol.2021.154734

Cavelti-Weder C, Li W, Zumsteg A, et al (2015a) Direct Reprogramming for pancreatic beta-cells using key developmental genes. Curr Pathobiol Rep 3:57–65. https://doi.org/10.1007/s40139-015-0068-0

Cerdá-Esteban N, Naumann H, Ruzittu S, et al (2017) Stepwise reprogramming of liver cells to a pancreas progenitor state by the transcriptional regulator Tgif2. Nat Commun 8:. https://doi.org/10.1038/ncomms14127

Chakravarthy H, Gu X, Enge M, et al (2017) Converting adult pancreatic islet a cells into b cells by targeting both Dnmt1 and Arx. Cell Metab. 25:622–634 https://doi.org/10.1016/j.cmet.2017.01.009

Chera S, Baronnier D, Ghila L, et al (2014) Diabetes recovery by age-dependent conversion of pancreatic δ-cells into insulin producers. Nature 514:503–507. https://doi.org/10.1038/nature13633

Cho YS, Lee JY, Park KS, Nho CW (2012) Genetics of type 2 diabetes in east asian populations. Curr Diab Rep 12:686–696. https://doi.org/10.1007/s11892-012-0326-z

Chung T, Lin S, Su J, et al (2017) Sinularin induces DNA damage , G2 / M phase arrest , and apoptosis in human hepatocellular carcinoma cells. BMC Complement Altern Med 1–8. https://doi.org/10.1186/s12906-017-1583-9

Cichosz SL, Johansen MD, Hejlesen O (2016) Toward big data analytics: Review of predictive models in management of diabetes and its complications. J Diabetes Sci Technol 10:27–34. https://doi.org/10.1177/1932296815611680

Colarusso JL, Zhou Q (2022) Direct reprogramming of different cell lineages into pancreatic β-like cells. Cell Reprogram 24:252–258. https://doi.org/10.1089/cell.2022.0048

Collombat P, Xu X, Ravassard P, et al (2009) The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into α and subsequently β cells. Cell 138:449–462. https://doi.org/10.1016/j.cell.2009.05.035

Conrad E, Stein R, Hunter CS (2014) Revealing transcription factors during human pancreatic b cell development. Trends Endocrinol Metab 25:407–414. https://doi.org/10.1016/j.tem.2014.03.013

Dai P, Qi G, Xu H, et al (2022) Reprogramming adipose mesenchymal stem cells into islet β-cells for the treatment of canine diabetes mellitus. Stem Cell Res Ther 13:1–14. https://doi.org/10.1186/s13287-022-03020-w

Das SLM, Kennedy JIC, Murphy R, et al (2014) Relationship between the exocrine and endocrine pancreas after acute pancreatitis. World J Gastroenterol 20:17196–17205. https://doi.org/10.3748/wjg.v20.i45.17196

Dey C, Raina K, Haridhasapavalan KK, et al (2021a) An overview of reprogramming approaches to derive integration-free induced pluripotent stem cells for prospective biomedical applications. Recent Adv iPSC Technol 231–287 https://doi.org/10.1016/B978-0-12-822231-7.00011-4

Dey C, Thool M, Bhattacharyya S, et al (2021b) Generation of biologically active recombinant human OCT4 protein from *E. coli*. 3 Biotech 11:1–16. https://doi.org/10.1007/s13205-021-02758-z

Dey C, Venkatesan V, Thummer RP (2022) Identification of optimal expression parameters

and purification of a codon-optimized human GLIS1 transcription factor from *Escherichia coli*. Mol Biotechnol 64:42–56. https://doi.org/10.1007/s12033-021-00390-z

Doyle MJ, Sussel L (2007) Nkx2.2 regulates β-cell function in the mature islet. Diabetes 56:1999–2007. https://doi.org/10.2337/db06-1766

El Khattabi I, Sharma A (2015) Proper activation of MafA is required for optimal differentiation and maturation of pancreatic β-cells. Best Pract & Res Clin Endocrinol & Metab 29:821–831. https://doi.org/10.1016/j.beem.2015.09.006

Enderami SE, Kehtari M, Abazari MF, et al (2018) Generation of insulin-producing cells from human induced pluripotent stem cells on PLLA/PVA nanofiber scaffold. Artif Cells, Nanomedicine Biotechnol 46:1062–1069. https://doi.org/10.1080/21691401.2018.1443466

Ferber S, Halkin A, Cohen H, et al (2000) Pancreatic and duodenal homeobox gene 1 induces expression of insulin genes in liver and ameliorates streptozotocin-induced hyperglycemia. Nat Med 6:568–572. https://doi.org/10.1038/75050

Ferreira LMR, Mostajo-Radji MA (2013) How induced pluripotent stem cells are redefining personalized medicine. Gene 520:1–6. https://doi.org/10.1016/j.gene.2013.02.037

Fu Z, R Gilbert E, Liu D (2013) Regulation of insulin synthesis and secretion and pancreatic Beta-cell dysfunction in diabetes. Curr Diabetes Rev 9:25–53. https://doi.org/10.2174/157339913804143225

Gannon M, Gamer LW, Wright CVE (2001) Regulatory regions driving developmental and tissue-specific expression of the essential pancreatic gene pdx1. Dev Biol 238:185–201. https://doi.org/10.1006/dbio.2001.0359

Gao D, Dai P, Fan Z, et al (2022) The roles of different multigene combinations of Pdx1, Ngn3, Sox9, Pax4, and Nkx2.2 in the reprogramming of canine ADSCs into IPCs. Cell Transplant 31:. https://doi.org/10.1177/09636897221081483

Gao T, Mckenna B, Li C, et al (2014) Pdx1 maintains b cell identity and function by repressing an a cell program. 259–271. https://doi.org/10.1016/j.cmet.2013.12.002

García-Fraga B, da Silva AF, López-Seijas J, Sieiro C (2015) Optimized expression conditions

for enhancing production of two recombinant chitinolytic enzymes from different prokaryote domains. Bioprocess Biosyst Eng 38:2477–2486. https://doi.org/10.1007/s00449-015-1485-5

Garrido-Utrilla A, Ayachi C, Friano ME, et al (2022) Conversion of gastrointestinal somatostatin-expressing D cells into insulin-producing beta-like cells upon Pax4 misexpression. Front Endocrinol (Lausanne) 13:. https://doi.org/10.3389/fendo.2022.861922

Gasa R, Mrejen C, Leachman N, et al (2004) Proendocrine genes coordinate the pancreatic islet differentiation program *in vitro*. Proc Natl Acad Sci 101:13245–13250 https://doi.org/10.1073/pnas.0405301101

Glick E, Leshkowitz D, Walker MD (2000) Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter. J Biol Chem 275:2199–2204. https://doi.org/10.1074/jbc.275.3.2199

Grada A, Otero-Vinas M, Prieto-Castrillo F, et al (2017) Research techniques made simple: analysis of collective cell migration using the wound healing assay. J Invest Dermatol 137:e11–e16. https://doi.org/10.1016/j.jid.2016.11.020

Gradwohl G, Dierich A, LeMeur M, Guillemot F (2000) neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci 97:1607–1611

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

Guo S, Dai C, Guo M, et al (2013) Inactivation of specific β cell transcription factors in type 2 diabetes. J Clin Invest 123:3305–3316. https://doi.org/10.1172/JCI65390

Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346–353. https://doi.org/10.1016/j.tibtech.2004.04.006

Haller C, Piccand J, De Franceschi F, et al (2019) Macroencapsulated human iPSC-derived pancreatic progenitors protect against STZ-induced hyperglycemia in mice. Stem Cell Reports 12:787–800. https://doi.org/10.1016/j.stemcr.2019.02.002

Haridhasapavalan KK, Borgohain MP, Dey C, et al (2019) An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene 686:146–

159. https://doi.org/10.1016/j.gene.2018.11.069

Haridhasapavalan KK, Sundaravadivelu PK, Thummer RP (2020) Codon optimization, cloning, expression, purification, and secondary structure determination of human ETS2 transcription factor. Mol Biotechnol 62:485–494. https://doi.org/10.1007/s12033-020-00266-8

Hata S, Hamada J-I, Maeda K, et al (2008) PAX4 has the potential to function as a tumor suppressor in human melanoma. Int J Oncol 33:1065–1071 https://doi.org/10.3892/ijo_00000095

Heremans Y (2003) Modulation of transcription factor expression induces adult human pancreatic exocrine-endocrine transdifferentiation and protects rat beta-cells against cytokine-induced apoptosis

Horb ME, Shen C, Tosh D, Slack JM. (2003) Experimental conversion of liver to pancreas. Curr Biol 13:105–115. https://doi.org/10.1016/S0960-9822(02)01434-3

Hori Y, Gu X, Xie X, Kim SK (2005) Differentiation of insulin-producing cells from human neural progenitor cells. PLoS Med 2:0347–0356. https://doi.org/10.1371/journal.pmed.0020103

Huang H-P, Liu M, El-Hodiri HM, et al (2000) Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3 . Mol Cell Biol 20:3292–3307. https://doi.org/10.1128/mcb.20.9.3292-3307.2000

Iacovazzo D, Flanagan SE, Walker E, et al (2018) MAFA missense mutation causes familial insulinomatosis and diabetes mellitus. Proc Natl Acad Sci U S A 115:1027–1032. https://doi.org/10.1073/pnas.1712262115

Ieda M, Fu JD, Delgado-Olguin P, et al (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142:375–386. https://doi.org/10.1016/j.cell.2010.07.002

Imai J, Katagiri H, Yamada T, et al (2005) Constitutively active PDX1 induced efficient insulin production in adult murine liver. Biochem Biophys Res Commun 326:402–409. https://doi.org/10.1016/j.bbrc.2004.11.047

Inada A, Nienaber C, Katsuta H, et al (2008) Carbonic anhydrase II-positive pancreatic cells are progenitors for both endocrine and exocrine pancreas after birth. Proc Natl Acad Sci U S A 105:19915–19919. https://doi.org/10.1073/pnas.0805803105

Jarikji ZH, Vanamala S, Beck CW, et al (2007) Differential ability of Ptf1a and Ptf1a-VP16 to convert stomach, duodenum and liver to pancreas. Dev Biol 304:786–799. https://doi.org/10.1016/j.ydbio.2007.01.027

Jennings RE, Berry AA, Kirkwood-Wilson R, et al (2013) Development of the human pancreas from foregut to endocrine commitment. Diabetes 62:3514–3522. https://doi.org/10.2337/db12-1479

Jeon J, Correa-Medina M, Ricordi C, et al (2009) Endocrine cell clustering during human pancreas development. J Histochem Cytochem 57:811–824. https://doi.org/10.1369/jhc.2009.953307

Kahn SE, Cooper ME, Del Prato S (2014) Pathophysiology and treatment of type 2 diabetes: Perspectives on the past, present, and future. Lancet 383:1068–1083. https://doi.org/10.1016/S0140-6736(13)62154-6

Kaneto H, Nakatani Y, Miyatsuka T, et al (2005) PDX-1/VP16 fusion protein, together with NeuroD or Ngn3, markedly induces insulin gene transcription and ameliorates glucose tolerance. Diabetes 54:1009–1022. https://doi.org/10.2337/diabetes.54.4.1009

Kataoka K, Han S iee, Shioda S, et al (2002) MafA is a glucose-regulated and pancreatic β-cell-specific transcriptional activator for the insulin gene. J Biol Chem 277:49903–49910. https://doi.org/10.1074/jbc.M206796200

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

Khow O, Suntrarachun S (2012) Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pac J Trop Biomed 2:159–162 https://doi.org/10.1016/S2221-1691(11)60213-X

Kim MJ, Lee EY, You YH, et al (2020) Generation of iPSC-derived insulin-producing cells

from patients with type 1 and type 2 diabetes compared with healthy control. Stem Cell Res 48:101958. https://doi.org/10.1016/j.scr.2020.101958

Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. Cell 139:1366–1375. https://doi.org/10.1016/j.cell.2009.12.001

Koblas T, Leontovyc I, Loukotova S, et al (2016) Reprogramming of pancreatic exocrine cells AR42J into insulin-producing cells using mRNAs for Pdx1, Ngn3, and MafA transcription factors. Mol Ther - Nucleic Acids 5:e320. https://doi.org/10.1038/mtna.2016.33

Koizumi M, Doi R, Fujimoto K, et al (2005) Pancreatic epithelial cells can be converted into insulin-producing cells by GLP-1 in conjunction with virus-mediated gene transfer of pdx-1. Surgery 138:125–133. https://doi.org/10.1016/j.surg.2005.06.008

Kojima H, Nakamura T, Fujita Y, et al (2002) Combined expression of pancreatic duodenal homeobox 1 and islet factor 1 induces immature enterocytes to produce insulin. Diabetes 51:1398–1408. https://doi.org/10.2337/diabetes.51.5.1398

Kopp JL, Dubois CL, Schaffer AE, et al (2011) Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. Development 138:653–665. https://doi.org/10.1242/dev.056499

Krapp A, Knöfler M, Frutiger S, et al (1996) The p48 DNA-binding subunit of transcription factor PTF1 is a new exocrine pancreas-specific basic helix-loop-helix protein. EMBO J 15:4317–4329. https://doi.org/10.1002/j.1460-2075.1996.tb00806.x

Krentz N, Hui Lau H, Abaitua F, et al (2022) 249-LB: PAX4 Loss of function alters human endocrine cell development and influences diabetes risk. Diabetes 71:1–83. https://doi.org/10.2337/db22-249-lb

Kumar HK, Borgohain MP, Dey C, Saha B (2019) An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene 686:146–159. https://doi.org/10.1016/j.gene.2018.11.069

Kunisada Y, Tsubooka-Yamazoe N, Shoji M, Hosoya M (2012) Small molecules induce efficient differentiation into insulin-producing cells from human induced pluripotent stem cells. Stem Cell Res 8:274–284. https://doi.org/10.1016/j.scr.2011.10.002

Lameire N (2000) Pathogenesis and treatment of diabetic nephropathy. EDTNA-ERCA J 26:8–10. https://doi.org/10.1111/j.1755-6686.2000.tb00082.x

Li W, Cavelti-Weder C, Zhang Y, et al (2014) Long-term persistence and development of induced pancreatic beta cells generated by lineage conversion of acinar cells. Nat Biotechnol 32:1223–1230. https://doi.org/10.1038/nbt.3082

Li WC, Rukstalis JM, Nishimura W, et al (2010) Activation of pancreatic-duct-derived progenitor cells during pancreas regeneration in adult rats. J Cell Sci 123:2792–2802. https://doi.org/10.1242/jcs.065268

Liang J, Chirikjian M, Pajvani UB, Bartolomé A (2022) MafA regulation in β-cells: from transcriptional to post-translational mechanisms. Biomolecules 12:535. https://doi.org/10.3390/biom12040535

Lima MJ, Muir KR, Docherty HM, et al (2013) Suppression of epithelial-to-Mesenchymal transitioning enhances ex vivo reprogramming of human exocrine pancreatic tissue toward functional insulin-Producing β-Like cells. Diabetes 62:2821–2833. https://doi.org/10.2337/db12-1256

Lithwick G, Margalit H (2003) Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Res 13:2665–2673. https://doi.org/10.1101/gr.1485203

Liu B, Kong Q, Zhang D, Yan L (2018) Codon optimization significantly enhanced the expression of human 37-kDa iLRP in *Escherichia coli*. 3 Biotech 8:1–7. https://doi.org/10.1007/s13205-018-1234-y

Lorenzo PI, Fuente-Martín E, Brun T, et al (2015) PAX4 defines an expandable β-cell subpopulation in the adult pancreatic islet. Sci Rep 5:1–14. https://doi.org/10.1038/srep15672

Lorenzo PI, Juárez-Vicente F, Cobo-Vuilleumier N, et al (2017) The diabetes-linked transcription factor PAX4: from gene to functional consequences. Gene. 8(3):101 https://doi.org/10.3390/genes8030101

Lu X, Guanga GP, Wan C, Rose RB (2012) A novel DNA binding mechanism for maf basic region-leucine zipper factors inferred from a MafA-DNA complex structure and binding specificities. Biochemistry 51:9706–9717. https://doi.org/10.1021/bi301248j

Lyttle BM, Li J, Krishnamurthy M, Fellows F (2008) Transcription factor expression in the developing human fetal endocrine pancreas. 1:1169–1180. https://doi.org/10.1007/s00125-008-1006-z

Ma J, Chen M, Wang J, et al (2008) Pancreatic duodenal homeobox-1 (PDX1) functions as a tumor suppressor in gastric cancer. 29:1327–1333. https://doi.org/10.1093/carcin/bgn112

Ma RCW, Hu C, Tam CH, et al (2013) Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4. Diabetologia 56:1291–1305. https://doi.org/10.1007/s00125-013-2874-4

Maertens B, Spriestersbach A, Von Groll U, et al (2010) Gene optimization mechanisms: A multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. Protein Sci 19:1312–1326. https://doi.org/10.1002/pro.408

Martin-Montalvo A, Lorenzo PI, López-Noriega L, Gauthier BR (2017) Targeting pancreatic expressed PAX genes for the treatment of diabetes mellitus and pancreatic neuroendocrine tumors. Expert Opin Ther Targets 21:77–89. https://doi.org/10.1080/14728222.2017.1257000

Mashima H (1996) Formation of insulin-producing cells from pancreatic acinar AR42J cells by hepatocyte growth factor. Endocrinology 137:3969–3976. https://doi.org/10.1210/en.137.9.3969

Matsuoka TA, Artner I, Henderson E, et al (2004) The MafA transcription factor appears to be responsible for tissue-specific expression of insulin. Proc Natl Acad Sci U S A 101:2930–2933. https://doi.org/10.1073/pnas.0306233101

Mellado-Gil JM, Jiménez-Moreno CM, Martin-Montalvo A, et al (2016) PAX4 preserves endoplasmic reticulum integrity preventing beta cell degeneration in a mouse model of type 1 diabetes mellitus. Diabetologia 59:755–765. https://doi.org/10.1007/s00125-016-3864-0

Merani S, Shapiro AMJ (2006) Current status of pancreatic islet transplantation. Clin Sci 110:611–625 https://doi.org/10.1042/CS20050342

Micsonai A, Wien F, Bulyáki É, et al (2018) BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic Acids Res 46:W315–W322. https://doi.org/10.1093/nar/gky497

Micsonai A, Wien F, Kernya L, et al (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U S A 112:E3095–E3103. https://doi.org/10.1073/pnas.1500851112

Millman JR, Xie C, Van Dervort A, et al (2016) Generation of stem cell-derived β-cells from patients with type 1 diabetes. Nat Commun 7:. https://doi.org/10.1038/ncomms11463

Minami K, Okuno M, Miyawaki K, et al (2005) Lineage tracing and characterization of insulin-secreting cells generated from adult pancreatic acinar cells. Proc Natl Acad Sci U S A 102:15116–15121. https://doi.org/10.1073/pnas.0507567102

Miyashita K, Miyatsuka T, Matsuoka TA, et al (2014) Sequential introduction and dosage balance of defined transcription factors affect reprogramming efficiency from pancreatic duct cells into insulin-producing cells. Biochem Biophys Res Commun 444:514–519. https://doi.org/10.1016/j.bbrc.2014.01.083

Miyatsuka T, Kaneto H, Kajimoto Y, et al (2003) Ectopically expressed PDX-1 in liver initiates endocrine and exocrine pancreas differentiation but causes dysmorphogenesis. Biochem Biophys Res Commun 310:1017–1025. https://doi.org/10.1016/j.bbrc.2003.09.108

Münst B, Thier MC, Winnemöller D, et al (2016) Nanog induces suppression of senescence through downregulation of p27KIP1 expression. J Cell Sci 129:912–920. https://doi.org/10.1242/jcs.167932

Nagaya M, Katsuta H, Kaneto H, et al (2009) Adult mouse intrahepatic biliary epithelial cells induced *in vitro* to become insulin-producing cells. J Endocrinol 201:37–47. https://doi.org/10.1677/JOE-08-0482

Nasteska D, Fine NHF, Ashford FB, et al (2021) PDX1LOW MAFALOW β-cells contribute to islet function and insulin release. Nat Commun 12:. https://doi.org/10.1038/s41467-020-20632-z

Nishimura W, Takahashi S, Yasuda K (2015) MafA is critical for maintenance of the mature beta cell phenotype in mice. Diabetologia 58:566–574. https://doi.org/10.1007/s00125-014-3464-9

Nishizawa M, Kataoka K, Goto N, et al (1989) v-maf, a viral oncogene that encodes a" leucine

zipper" motif. Proc Natl Acad Sci 86:7711–7715 https://doi.org/10.1073/pnas.86.20.7711

Noguchi H, Xu G, Matsumoto S, et al (2006) Induction of pancreatic stem/progenitor cells into insulin-producing cells by adenoviral-mediated gene transfer technology. Cell Transplant 15:929–938. https://doi.org/10.3727/000000006783981431

Noso S, Babaya N, Hiromine Y, et al (2013) Mafa knockout NOD mice are protected from type 1 diabetes despite accelerated infiltration of lymphocytes into islets. In: DIABETOLOGIA. pp S63--S63

Noso S, Kataoka K, Kawabata Y, et al (2010) Insulin transactivator MafA regulates intrathymic expression of insulin and affects susceptibility to type 1 diabetes. Diabetes 59:2579–2587. https://doi.org/10.2337/db10-0476

O'Malley J, Woltjen K, Kaji K (2009) New strategies to generate induced pluripotent stem cells. Curr Opin Biotechnol 20:516–521. https://doi.org/10.1016/j.copbio.2009.09.005

Oh SH, Muzzonigro TM, Bae SH, et al (2004) Adult bone marrow-derived cells trans differentiating into insulin-producing cells for the treatment of type I diabetes. Lab Investig 84:607–617. https://doi.org/10.1038/labinvest.3700074

Olbrot M, Rud J, Moss LG, Sharma A (2002) Identification of β-cell-specific insulin gene transcription factor RIPE3b1 as mammalian MafA. Proc Natl Acad Sci 99:6737–6742

Oliver-krasinski JM, Kaestner KH, Doris A, et al (2009) The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice Find the latest version : The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice. https://doi.org/10.1172/JCI37028.1888

Ottensmeyer FP, Beniac DR, Luo RZT, Yip CC (2000) Mechanism of transmembrane signaling: Insulin binding and the insulin receptor. Biochemistry 39:12103–12112. https://doi.org/10.1021/bi0015921

Pan FC, Brissova M (2014) Pancreas development in humans. Curr Opin Endocrinol Diabetes Obes 21:77 https://doi.org/10.1097/MED.0000000000000047

Pan FC, Wright C (2011) Pancreas organogenesis: from bud to plexus to gland. Dev Dyn 240:530–565 https://doi.org/10.1002/dvdy.22584

Peitz M, Münst B, Thummer RP, et al (2014) Cell-permeant recombinant Nanog protein promotes pluripotency by inhibiting endodermal specification. Stem Cell Res 12:680–689. https://doi.org/10.1016/j.scr.2014.02.006

Pellegrini S, Cantarelli E, Sordi V, et al (2016) The state of the art of islet transplantation and cell therapy in type 1 diabetes. Acta Diabetol 53:683–691 https://doi.org/10.1007/s00592-016-0847-z

Pinney SE, Oliver-Krasinski J, Ernst L, et al (2011) Neonatal diabetes and congenital malabsorptive diarrhea attributable to a novel mutation in the human neurogenin-3 gene coding sequence. J Clin Endocrinol Metab 96:1960–1965. https://doi.org/10.1210/jc.2011-0029

Piper K, Brickwood S, Turnpenny LW, et al (2004) Beta cell differentiation during early human pancreas development. J Endocrinol 181:11–24

Polak M, Bouchareb-Banaei L, Scharfmann R, Czernichow P (2000) Early pattern of differentiation in the human pancreas. Diabetes 49:225–232. https://doi.org/10.2337/diabetes.49.2.225

Rezania A, Bruin JE, Arora P, et al (2014) Reversal of diabetes with insulin-producing cells derived *in vitro* from human pluripotent stem cells. Nat Biotechnol 32:1121–1133. https://doi.org/10.1038/nbt.3033

Rhee M, Lee SH, Kim JW, et al (2016) Preadipocyte factor 1 induces pancreatic ductal cell differentiation into insulin-producing cells. Sci Rep 6:1–11. https://doi.org/10.1038/srep23960

Rooman I, Bouwens L (2004) Combined gastrin and epidermal growth factor treatment induces islet regeneration and restores normoglycaemia in C57BI6/J mice treated with alloxan. Diabetologia 47:259–265. https://doi.org/10.1007/s00125-003-1287-1

Rorsman P, Renström E (2003) Insulin granule dynamics in pancreatic beta cells. Diabetologia 46:1029–1045. https://doi.org/10.1007/s00125-003-1153-1

Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. 5:1–17. https://doi.org/10.3389/fmicb.2014.00172

Rother KI, Harlan DM (2004) Challenges facing islet transplantation for the treatment of type 1 diabetes mellitus. J Clin Invest 114:877–883. https://doi.org/10.1172/JCI200423235

Ryan BJ, Henehan GT (2013) Overview of approaches to preventing and avoiding proteolysis during expression and purification of proteins. Curr Protoc protein Sci 71:5–25 https://doi.org/10.1002/0471140864.ps0525s71

Sachdeva MM, Claiborn KC, Khoo C, et al (2009) Pdx1 (MODY4) regulates pancreatic beta cell susceptibility to ER stress. Proc Natl Acad Sci U S A 106:19090–19095. https://doi.org/10.1073/pnas.0904849106

Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. Mol Cell Biochem 307:249–264. https://doi.org/10.1007/s11010-007-9603-6

Salisbury RJ, Blaylock J, Berry AA, et al (2022) The window period of NEUROGENIN3 during human gestation The window period of NEUROGENIN3 during human gestation. 2014:. https://doi.org/10.4161/19382014.2014.954436

San-Miguel T, Pérez-Bermúdez P, Gavidia I (2013) Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. Springerplus 2:1–4. https://doi.org/10.1186/2193-1801-2-89

Sanchez-Garcia L, Martín L, Mangues R, et al (2016) Recombinant pharmaceuticals from microbial cells: A 2015 update. Microb Cell Fact 15:1–7. https://doi.org/10.1186/s12934-016-0437-3

Sancho R, Gruber R, Gu G, Behrens A (2014) Loss of Fbw7 reprograms adult pancreatic ductal cells into α, δ, and β cells. Cell Stem Cell 15:139–153

Sangan CB, Jover R, Heimberg H, et al (2015) Molecular and cellular endocrinology *In vitro* reprogramming of pancreatic alpha cells towards a beta cell phenotype following ectopic HNF4α expression. Mol Cell Endocrinol 399:50–59. https://doi.org/10.1016/j.mce.2014.09.009

Sapir T, Shternhall K, Meivar-Levy I, et al (2005) Cell-replacement therapy for diabetes: Generating functional insulin-producing tissue from adult human liver cells. Proc Natl Acad Sci U S A 102:7964–7969. https://doi.org/10.1073/pnas.0405277102

Saxena P, Heng BC, Bai P, et al (2016) A programmable synthetic lineage-control network

that differentiates human IPSCs into glucose-sensitive insulin-secreting beta-like cells. Nat Commun 7:1–14. https://doi.org/10.1038/ncomms11247

Seo BJ, Hong YJ, Do JT (2017) Cellular reprogramming using protein and cell-penetrating peptides. Int J Mol Sci 18:. https://doi.org/10.3390/ijms18030552

Serna N, Sánchez-García L, Unzueta U, et al (2018) Protein-Based Therapeutic Killing for Cancer Therapies. Trends Biotechnol 36:318–335. https://doi.org/10.1016/j.tibtech.2017.11.007

Setty SG, Crasto W, Jarvis J, et al (2016) New insulins and newer insulin regimens: a review of their role in improving glycaemic control in patients with diabetes. Postgrad Med J 92:152–164 https://doi.org/10.1136/postgradmedj-2015-133716

Shahjalal HM, Shiraki N, Sakano D, et al (2014) Generation of insulin-producing β-like cells from human iPS cells in a defined and completely xeno-free culture system. J Mol Cell Biol 6:394–408. https://doi.org/10.1093/jmcb/mju029

Shih HP, Kopp JL, Sandhu M, et al (2012) A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. Dev 139:2488–2499. https://doi.org/10.1242/dev.078634

Shimizu A, Nishida J, Ueoka Y, et al (1998) CyclinG contributes to G2/M arrest of cells in response to DNA damage. Biochem Biophys Res Commun 242:529–533 https://doi.org/10.1006/bbrc.1997.8004

Smith SB, Ee HC, Conners JR, German MS (1999) Paired-Homeodomain Transcription Factor PAX4 Acts as a Transcriptional Repressor in Early Pancreatic Development. Mol Cell Biol 19:8272–8280. https://doi.org/10.1128/mcb.19.12.8272

Smith SB, Gasa R, Watada H, et al (2003) Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of Pax4. J Biol Chem 278:38254–38259. https://doi.org/10.1074/jbc.M302229200

Solar M, Cardalda C, Houbracken I, et al (2009) Pancreatic Exocrine Duct Cells Give Rise to Insulin-Producing β Cells during Embryogenesis but Not after Birth. Dev Cell 17:849–860. https://doi.org/10.1016/j.devcel.2009.11.003

Sommer CA, Mostoslavsky G (2013) The evolving field of induced pluripotency: recent progress and future challenges. J Cell Physiol 228:267–275 https://doi.org/10.1002/jcp.24155

Sosa-Pineda B (2004) The gene Pax4 is an essential regulator of pancreatic β-cell development. Mol Cells 18:289–294

Sumazaki R, Shiojiri N, Isoyama S, et al (2004) Conversion of biliary system to pancreatic tissue in Hes1-deficient mice. Nat Genet 36:83–87. https://doi.org/10.1038/ng1273

Sun H, Saeedi P, Karuranga S, et al (2022) IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 183:109119. https://doi.org/10.1016/j.diabres.2021.109119

Tateishi K, He J, Taranova O, et al (2008) Generation of insulin-secreting islet-like clusters from human skin fibroblasts. J Biol Chem 283:31601–31607. https://doi.org/10.1074/jbc.M806597200

Thier M, Münst B, Edenhofer F (2011) Exploring refined conditions for reprogramming cells by recombinant Oct4 protein. Int J Dev Biol 54:1713–1721. https://doi.org/10.1387/ijdb.103193mt

Thier M, Münst B, Mielke S, Edenhofer F (2012) Cellular reprogramming employing recombinant Sox2 protein. Stem Cells Int 2012:. https://doi.org/10.1155/2012/549846

Thool M, Dey C, Bhattacharyya S, et al (2021) Generation of a recombinant stem cell-specific human SOX2 protein from *Escherichia coli* under native conditions. Mol Biotechnol 63:327–338. https://doi.org/10.1007/s12033-021-00305-y

Thorel F, Népote V, Avril I, et al (2010) Conversion of adult pancreatic α-cells to B-cells after extreme B-cell loss. Nature 464:1149–1154 https://doi.org/10.1038/nature08894

Tokarz VL, MacDonald PE, Klip A (2018) The cell biology of systemic insulin function. J Cell Biol 217:2273–2289. https://doi.org/10.1083/jcb.201802095

Valdez IA, Dirice E, Gupta MK, et al (2016) Proinflammatory cytokines induce endocrine differentiation in pancreatic ductal cells via STAT3-dependent NGN3 activation. Cell Rep 15:460–470. https://doi.org/10.1016/j.celrep.2016.03.036

Van de Casteele M, Leuckx G, Baeyens L, et al (2013) Neurogenin 3+ cells contribute to β-cell neogenesis and proliferation in injured adult mouse pancreas. Cell Death Dis 4:e523–e523. https://doi.org/10.1038/cddis.2013.52

Vethe H, Bjørlykke Y, Ghila LM, et al (2017) Probing the missing mature β-cell proteomic landscape in differentiating patient iPSC-derived cells. Sci Rep 7:1–14. https://doi.org/10.1038/s41598-017-04979-w

Walczak MP, Drozd AM, Stoczynska-Fidelus E, et al (2016) Directed differentiation of human iPSC into insulin producing cells is improved by induced expression of PDX1 and NKX6.1 factors in IPC progenitors. J Transl Med 14:1–16. https://doi.org/10.1186/s12967-016-1097-0

Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 24:1241–1252 https://doi.org/10.1038/nbt1252

Wang N, Zhu Y, Xie M, et al (2018a) Long Noncoding RNA Meg3 Regulates Mafa Expression in Mouse Beta Cells by Inactivating Rad21, Smc3 or Sin3α. Cell Physiol Biochem 45:2031–2043. https://doi.org/10.1159/000487983

Wang Q, Donelan W, Ye H, et al (2019) Real-time observation of pancreatic beta cell differentiation from human induced pluripotent stem cells. Am J Transl Res 11:3490–3504

Wang W, Shi Q, Guo T, et al (2016) PDX1 and ISL1 differentially coordinate with epigenetic modifications to regulate insulin gene expression in varied glucose concentrations. Mol Cell Endocrinol 428:38–48. https://doi.org/10.1016/j.mce.2016.03.019

Wang X, Sterr M, Burtscher I, et al (2018b) Genome-wide analysis of PDX1 target genes in human pancreatic progenitors. Mol Metab 9:57–68. https://doi.org/10.1016/j.molmet.2018.01.011

Wingfield PT (2015) Overview of the purification of recombinant proteins. Curr Protoc protein Sci 80:1–6 https://doi.org/10.1002/0471140864.ps0601s80

Wood DW (2014) New trends and affinity tag designs for recombinant protein purification. Curr Opin Struct Biol 26:54–61. https://doi.org/10.1016/j.sbi.2014.04.006

Xiao X, Guo P, Shiota C, et al (2018) Endogenous reprogramming of alpha cells into beta cells, induced by viral gene therapy, reverses autoimmune diabetes. Cell Stem Cell 22:78-90.e4.

https://doi.org/10.1016/j.stem.2017.11.020

Xu X, D'Hoker J, Stangé G, et al (2008) β cells can be generated from endogenous progenitors in injured adult mouse pancreas. Cell 132:197–207. https://doi.org/10.1016/j.cell.2007.12.015

Yabe SG, Fukuda S, Nishida J, et al (2019) Induction of functional islet-like cells from human iPS cells by suspension culture. Regen Ther 10:69–76. https://doi.org/10.1016/j.reth.2018.11.003

Yamada T, Cavelti-Weder C, Caballero F, et al (2015) Reprogramming mouse cells with a pancreatic duct phenotype to insulin-producing β-like cells. Endocrinology 156:2029–2038. https://doi.org/10.1210/en.2014-1987

Yang X-F, Ren L-W, Yang L, et al (2017) *In vivo* direct reprogramming of liver cells to insulin producing cells by virus-free overexpression of defined factors. Endocr J 64:291–302. https://doi.org/10.1507/endocrj.EJ16-0463

Yang Y, Akinci E, Dutton JR, et al (2013) Stage specific reprogramming of mouse embryo liver cells to a beta cell-like phenotype. Mech Dev 130:602–612. https://doi.org/10.1016/j.mod.2013.08.002

Ye L, Robertson MA, Hesselson D, et al (2015) Glucagon is essential for alpha cell transdifferentiation and beta cell neogenesis. Dev 142:1407–1417. https://doi.org/10.1242/dev.117911

Yechoor V, Liu V, Espiritu C, et al (2009) Neurogenin3 is sufficient for transdetermination of hepatic progenitor cells into neo-islets *in vivo* but not transdifferentiation of hepatocytes. Dev Cell 16:358–373. https://doi.org/10.1016/j.devcel.2009.01.012

Young CL, Britton ZT, Robinson AS (2012) Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnol J 7:620–634. https://doi.org/10.1002/biot.201100155

Zalzman M, Anker-Kitai L, Efrat S (2005a) Differentiation of human liver-derived, insulin-producing cells toward the β-cell phenotype. Diabetes 54:2568–2575. https://doi.org/10.2337/diabetes.54.9.2568

Zalzman M, Gupta S, Giri RK, et al (2003) Reversal of hyperglycemia in mice by using human

expandable insulin-producing cells differentiated from fetal liver progenitor cells. Proc Natl Acad Sci U S A 100:7253–7258. https://doi.org/10.1073/pnas.1136854100

Zhang C, Moriguchi T, Kajihara M, et al (2005) MafA is a key regulator of glucose-stimulated insulin secretion. Mol Cell Biol 25:4969–4976. https://doi.org/10.1128/mcb.25.12.4969-4976.2005

Zhang D, Jiang W, Liu M, et al (2009) Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. Cell Res 19:429–438. https://doi.org/10.1038/cr.2009.28

Zhang S, Huang F, Tian W, et al (2020) Andrographolide promotes pancreatic duct cells differentiation into insulin-producing cells by targeting PDX-1. Biochem Pharmacol 174:113785. https://doi.org/10.1016/j.bcp.2019.113785

Zhang Y, Fava GE, Wang H, et al (2016) PAX4 gene transfer induces α-to-β cell phenotypic conversion and confers therapeutic benefits for diabetes treatment. Mol Ther 24:251–260

Zhao L, Guo M, Matsuoka TA, et al (2005) The islet β cell-enriched MafA activator is a key regulator of insulin gene transcription. J Biol Chem 280:11887–11894. https://doi.org/10.1074/jbc.M409475200

Zhou Q, Brown J, Kanarek A, et al (2008) *In vivo* reprogramming of adult pancreatic exocrine cells to β-cells. Nature 455:627–632 https://doi.org/10.1038/nature07314

Zhu FF, Zhang PB, Zhang DH, et al (2011) Generation of pancreatic insulin-producing cells from rhesus monkey induced pluripotent stem cells. Diabetologia 54:2325–2336. https://doi.org/10.1007/s00125-011-2246-x

Zhu S, Russ HA, Wang X, et al (2016) Human pancreatic beta-like cells converted from fibroblasts. Nat Commun 7:1–13. https://doi.org/10.1038/ncomms10080

Zhu Y, Liu Q, Zhou Z, Ikeda Y (2017) PDX1, Neurogenin-3, and MAFA: Critical transcription regulators for beta cell development and regeneration. Stem Cell Res Ther 8:1–7. https://doi.org/10.1186/s13287-017-0694-z

List of publications

Publications from Ph.D. thesis

Research articles accepted / under revision / to be submitted

1. **Narayan G**, Sundaravadivelu PK, Agrawal A, Gogoi R, Nagotu S, Thummer RP (**2021**) Soluble expression, purification, and secondary structure determination of human PDX1 transcription factor. Protein Expression and Purification, 180: 105807. https://doi.org/10.1016/j.pep.2020.105807 (Elsevier).

2. **Narayan G**, Agrawal A, Gogoi R, Nagotu S, Thummer RP (**2021**) Protein production and purification of a codon-optimized human NGN3 transcription factor from *E. coli*. Protein Journal, 40(6):891-906. https://doi.org/10.1007/s10930-021-10020-x (Springer Nature).

3. **Narayan G**, Nagotu S, Thummer RP (**2023**) Heterologous expression, purification and biological activity determination of codon-optimized human MAFA transcription factor from *E. coli*. Healthcare Research and Related Technologies - Proceedings of NERC 2022 (Springer Nature)

4. **Narayan G**, Agrawal A, Sen P, Nagotu S, Thummer RP (**2023**) Production of bioactive recombinant human PAX4 protein from *E. coli*. Protein Journal (Springer Nature) (under revision)

5. **Narayan G**, Sen P, Nagotu S, Thummer RP (**2023**) Biological activity of recombinant Pancreatic and Duodenum homeobox 1 produced in *Escherichia coli*. *Manuscript submitted*

Review articles / Book chapters

6. Borgohain MP*, **Narayan G***, Krishna Kumar H, Dey C, Thummer RP (**2018**) Maximizing expression and yield of human recombinant proteins from bacterial cell factories for biomedical applications. Advances in Microbial Biotechnology: Current Trends and Future Prospects. Apple Academic Press. ISBN: 9781771886673. *Equal contribution.

7. Agrawal A*, **Narayan G***, Gogoi R, Thummer RP (**2021**) Recent advances in the generation of β-cells from induced pluripotent stem cells as a potential cure for diabetes mellitus. In: Turksen K. (eds) Cell Biology and Translational Medicine. Advances in Experimental Medicine and Biology (Springer Nature). *Equal contribution.

8. **Narayan G**, Ronima KR, Thummer RP (**2022**) Direct reprogramming of somatic cells into induced β-cells: An overview. Cham: Springer International Publishing, 1410: 171-189. https://doi.org/10.1007/5584_2022_756

9. **Narayan G**, Ronima KR, Agrawal A, Thummer, RP (**2023**) An insight into vital genes responsible for β-cell formation In: Turksen K. (eds) Cell Biology and Translational Medicine. Advances in Experimental Medicine and Biology (Springer Nature). *In Press*

Publications from collaborative work

Review articles / Book chapters

10. Dey C*, **Narayan G***, Krishna Kumar H*, Borgohain MP*, Lenka N, Thummer RP (**2017**) Cell-Penetrating peptides as a tool to deliver biologically active recombinant proteins to generate transgene-free induced pluripotent stem cells. Stud Stem Cells Res Ther 3(1): 006-015. *Equal contribution. DOI: 10.17352/sscrt.000011

11. Haridhasapavalan KK*, Borgohain MP*, Dey C, Saha B, **Narayan G**, Kumar S, Thummer RP (**2019**) An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene. 686:146-159. https://doi.org/10.1016/j.gene.2018.11.069 (Elsevier). *Equal contribution

Conferences / workshop

1. **Narayan G**. Participated in two-day hands on "**Training program on Confocal Laser Scanning Microscopy and its Applications**", organized by Guwahati Biotech Park under Bioprospecting Facility of Guwahati Biotech Park Incubation Centre in association with National Institute of Pharmaceutical Education and Research Guwahati during 2nd and 3rd November, **2017**.

2. **Narayan G**. Participated in workshop on human induced pluripotent stem cells (iPSCs) organized by Centre for Stem Cell Research (CSCR) from 18th to 23rd February, **2019**.

3. **Narayan G.** Participated as scholar at 1st Departmental Retreat (Biotech Express) organized by Departmrnt of Biosciences and Bioengineering, Indian Institute of Technology Guwahati on 21st December, **2019**.

4. **Narayan G.**, Agrawal A., Thummer RP. Heterologous expression and purification of codon-optimized human PAX4 transcription factor from *E. coli*. Healthcare Research and Related Technologies; North-East Research Conclave (NERC) 2022 organized by Indian Institute of Technology Guwahati from 20th–22nd May, **2022** (Poster presentation).