**INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI**
**SHORT ABSTRACT OF THESIS**

| | | |
|---|---|---|
| Name of the Student | : | **GLORIA NARAYAN** |
| Roll Number | : | **166106028** |
| Programme of Study | : | **Ph.D.** |

Thesis Title: **Generation of a bioactive recombinant protein toolbox of pancreatic-specific transcription factors**

| | | |
|---|---|---|
| Name of Thesis Supervisor(s) | : | **Dr. Rajkumar P. Thummer**<br>**Dr. Shirisha Nagotu** |
| Thesis Submitted to the **Department**/ ~~Center~~ | : | **Biosciences and Bioengineering** |
| Date of completion of Thesis Viva-Voce Exam | : | **5th June 2023** |
| Key words for description of Thesis Work | : | Diabetes mellitus, protein expression, protein purification, recombinant proteins, PDX1, NGN3, MAFA, PAX4, cancer, β-cells, mammalian cell culture |

## SHORT 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 in 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 were 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.