Establishment of a bioactive recombinant protein toolbox for the prospective generation of integration-free cardiomyocytes and other biological applications

A Thesis

Submitted in Partial Fulfilment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Krishna Kumar Haridhasapavalan

Under the supervision of

Dr. Rajkumar P. Thummer

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

August, 2022

Dedicated to my Appa & Amma

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

DECLARATION

I do hereby declare that the content embodied in this thesis entitled "**Establishment of a bioactive recombinant protein toolbox for the prospective generation of integration-free cardiomyocytes and other biological applications**" 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 degree of Doctor of Philosophy, under the supervision of **Dr. Rajkumar P. Thummer**.

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: 5th August 2022

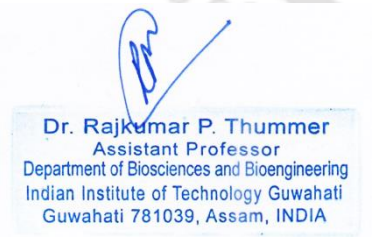
Krishna Kumar Haridhasapavalan

(156106036)

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering

CERTIFICATE

It is certified that the work described in this thesis entitled "**Establishment of a bioactive recombinant protein toolbox for the prospective generation of integration-free cardiomyocytes and other biological applications**", by **Mr. Krishna Kumar Haridhasapavalan** (Roll No. 156106036) for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my 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)

Date: 5th August 2022

Acknowledgement

I feel privileged to jot down my thankfulness towards all the lovely souls who have contributed directly or indirectly to achieving this milestone in my life. I am blessed to acknowledge each one of them for sparing their precious time and extending their valuable and constant support during my entire graduate research journey.

First and foremost, my deepest gratitude goes out to my thesis supervisor **Dr. Rajkumar P. Thummer,** the head of the Laboratory for Stem Cell Engineering and Regenerative Medicine (SCERM), for providing the wonderful opportunity to work under his guidance and thereby to be a part of this esteemed institution. His faith and confidence in me, even when I did not have self-confidence in the tough times, motivated me to accomplish the most challenging tasks and face science as an enthusiastic independent researcher. Through electrifying scientific discussions, agreements and disagreements, successes and failures, he made me understand and think about science from more than one outcome and perspective. Being his first student, I got his unconditional care and affection, consistent support along with the experience I gained under his supervision in establishing a fully functional laboratory from the terrace, which molded me to think critically, design and execute things systematically, and mainly to focus and work under most difficult/hard circumstances. The most encouraging and turning point in my graduate journey was when he said, "do not worry about what is not in your control; rather, focus on what is in your control" changed the pandemic loss into the most productive manuscript writing time. Besides being a great academician, I have known him as one of the humblest persons ever, standing beside me through all my ups and downs. His invaluable advice and the things he taught me will undoubtedly help me in my future career as well as to evolve as a progressive and responsible citizen.

My heartfelt gratitude towards **Dr. Shrisha Nagotu** of the Department of Biosciences and Bioengineering for her immense support and kind words, especially during my difficult times, and for her encouragement and unbiased affection in every stage of this journey.

I would like to extend my gratitude to the members of my Doctoral committee, **Prof. Bithiah G. Jaganathan** and **Prof. Sachin Kumar** of the Department of Biosciences and Bioengineering and **Dr. Sunanda Chatterjee** of the Department of Chemistry for their

valuable suggestions, constant support and encouragement, and advice which enabled me to improve my work.

I would like to acknowledge and thank both **Indian** and **Foreign reviewers** for appreciating my thesis work and for their valuable suggestions/comments to improvise my thesis.

I would also like to extend my gratitude to my viva-voce board members, **Prof. Gopinath Packirisamy** of the Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, **Prof. Bithiah G. Jaganathan** and **Prof. Nitin Chaudhary** of the Department of Biosciences and Bioengineering and **Dr. Krishna Pada Bhabak** of the Department of Chemistry, Indian Institute of Technology Guwahati for their support, motivation and providing an opportunity to prove my understanding and thinking capability against the critical questions raised by them.

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**, **Prof. Kannan Pakshiranjan** and **Prof. V Venkata Dasu** 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** and **Dr. Shrisha Nagotu** of the Department of Biosciences and Bioengineering for providing high-end instrument facilities, without which the thesis would be incomplete. I would also like to thank the Ministry of Education, Government of India, for the financial support during my tenure.

It is my pleasure to thank all the members of the SCERM lab who gave me a wonderful, workable, interesting and warm environment in the lab and also bore my crazy activities. I am beholden to my lab mates cum friends-like family, **Ms. Chandrima Dey** and **Ms. Gloria Narayan**, who stood by my side through all my ups and downs, starting from converting the terrace to a fully functional lab to till date and counting on. My journey would be incomplete without these two-crazy people for their moral support, unconditional love and affection at my lowest and cheered louder at my highest. Next, I would like to acknowledge **Mr. Manash Borgohain**, **Ms. Chandrima Dey**, **Ms. Khyati Raina** and **Mr. Pradeep Kumar S** for the fruitful scientific arguments/discussions and for their support in designing experiments, optimizing protocols, data interpretation and for troubleshooting

each and every technical problem I/we encountered. I am also thankful to my hardworking and talented cardiac team members, **Mr. Pradeep Kumar S**, **Mr. Sujal Harsh Ranjan** and **Mr. Nayan Jyoti Das;** without them, my thesis would not have been completed. My enthusiastic thanks go out to all my juniors, **Ms. Khyati Raina**, **Mr. Pradeep Kumar S**, **Ms. Akriti Agrawal**, **Mr. Sujal Harsh Ranjan**, **Mr. Nayan Jyoti Das**, **Ms. Atreyee Borthakur**, **Mr. Bitan Saha**, **Ms. Madhuri Thool** and **Ms. Ronima K R** for their affectionate treatment and valuable assistance throughout my work. I want to thank all the OBCAL lab members (**Mr. Nayan**, **Ms. Rachayeeta**, **Mr. Riddhi**, **Ms. Neha**, **Mr. Sahay**, **Mr. Terrence**, **Ms. Suchetana**, **Ms. Isha**, and **Ms. Tanveera**) for being supportive and cooperative throughout this journey.

My long journey would be boring without my friends, **Dr. Vinoth**, **Dr. Muthuvel**, **Dr. Anil**, **Dr. Abhishek**, **Dr. Nivedhita**, **Mr. Kamlesh**, **Ms. Sahaya**, **Dr. Kuldeep** and others, who made my life crazy, fun filling, providing more precious memories to cherish and definitely made me mad at most of the times, but supported me in the critical situations, whenever I needed before I approach them. I must also acknowledge all my school and college friends (**Ms. Ivy**, **Mr. Aravind**, **Mrs. Anu**, **Mr. John**, **Ms. Geetha** and others) and Ph.D. batch mates for their love, encouragement and support.

Most importantly, I would like to thank my appa (dad) **Arulnithi K. Haridhasapavalan** and my amma (mom) **Smt. H. Velammal** for their patience, never-ending blessings and unconditional love and support to carry out my future goals with full freedom. Their numerous personal sacrifices have enabled me to reach this juncture in life. Next, I would like to thank my annas (brothers), **Dr. H. Kesava Kumar** and **Mr. H. Arun Kumar,** and annis (sisters-in-law), **Mrs. B. Bavithra** and **Mrs. T. Deepa,** for being my inspiration, motivation and biggest support in all my achievements. Without these people, nothing could be more fun and fulfilling. My special thanks to my special niece, Kutty **Ms. K. Vidhusana**, for being one of the best blessings and extending our family. Simple thanks are not enough to convey my deep respect and gratitude toward my family.

Finally, I would like to thank the Lord, the *Almighty Nature* who governs everything, for showering his grace at every stage of my life.

Signing off

Krishna Kumar Haridhasapavalan

Contents

Abstract

Cardiovascular diseases (CVDs) have been among the leading life-threatening diseases worldwide for over a decade. Heart failure is the primary concern of CVDs as it contributes majorly to global mortality. In most of these diseases, cardiomyocytes, the functional unit of the heart, are affected. The very low or nil regeneration capability of heart tissue, specifically cardiomyocytes, is the main limitation of heart regeneration. In the field of cardiology, balancing this loss of cardiomyocytes was highly challenging, even in the modern medical world. Current clinical therapies (excluding heart transplantation) can only prevent further cardiac tissue damage or assist affected heart functionality. In this prospect, cell-based therapies appear to be a promising and effective way to repair damaged/injured cardiac tissue. Soon after the discovery of cell reprogramming, direct conversion or direct reprogramming of somatic cells to functional induced cardiomyocytes bypassing the pluripotent state offers a promising alternative in cell therapies. However, the major problem with the generation of induced cardiomyocytes using traditional integrative approaches is genomic instability, which reduces the clinical applicability of induced cells. In order to overcome this limitation, several other methods are emerging to fulfill the needs of clinically relevant approaches for direct cardiac reprogramming, and protein-based cellular reprogramming has been reported as the safest approach among other available non-integrative approaches. Although recombinant protein-based therapeutics or cellular reprogramming is a safer and more promising alternative, there are several challenges associated with the heterologous production of these recombinant proteins. Therefore, we aim to establish a recombinant protein toolbox consisting of key cardiac reprogramming transcription factors (GATA4, MEF2C, TBX5, ETS2, MESP1 and HAND2) by addressing those associated problems for the prospective generation of integration-free functional cardiomyocytes. The recombinant proteins generated in this research by employing simple and systematic strategies will have native-like folding conformations, thereby retaining their biological activity. To accomplish this, we first individually cloned their codon-optimized full-length protein-coding nucleotide sequences fused to a nuclear localization signal/sequence, a cell-penetrating peptide, and a poly-Histidine tag into the protein expression vector and expressed in the bacterial system (*E. coli* strain BL21(DE3)). Subsequently, we screened and identified the optimal expression parameters to obtain this recombinant fusion protein in soluble form from *E. coli* and examined the effect of tagging

fusion tags with these recombinant proteins at two different terminals. Overall our results demonstrated the importance of identifying the optimal expression conditions and the influence of the position of fusion tags on these recombinant protein expressions in terms of quality and quantity. As proof, we reported that the fusion tags at the C-terminal end of TBX5, ETS2 and HAND2 proteins compromise the quality when compared to their N-terminal counterparts. Also, C-terminal fused MEF2C and TBX5 and N-terminal fused MESP1 proteins have significantly higher soluble expression, unlike their respective counterparts. Additionally, we found MESP1 to be an ion-sensitive protein, as we observed salt-dependent aggregation of this fusion protein while screening the optimal soluble expression parameters. We then established the one-step homogeneous purifications of these recombinant proteins. Further, we showed that purification under native conditions and an appropriate salt concentration resulted in retaining the native-like folding conformation of these proteins. In addition, the generated recombinant fusion proteins can be directly delivered to the mammalian cells mediated through the trans-activator of transcription (a cell-penetrating peptide) and can be localized into the nucleus mediated through nuclear localization signal/sequence without any addition of transduction reagents. Interestingly, we did not find any substantial differences in the cell penetration ability; however, we observed some considerable differences in the nuclear translocation efficiency of these proteins between different cell lines. Notably, we showed the positive regulation of the purified ETS2 and MESP1 fusion proteins as well as the negative regulation of purified GATA4 fusion protein in mammalian breast cancer cells. Likewise, we also demonstrated the tumor suppressive role of the recombinant TBX5 fusion protein in its pure form in colon cancer cells. Moreover, the generated recombinant HAND2 and MEF2C fusion proteins showed angiogenic and anti-angiogenic potential, respectively, in the *ex vivo* chicken embryo model. Using a reporter system, we confirmed the synergistic effect of purified MEF2C and HAND2 fusion proteins in the activation of alpha-Myosin Heavy Chain, thus suggesting these factors may contribute to the maturation of cardiomyocytes. In the near future, this established recombinant protein toolbox will pave the way to generate transgene-free patient-specific autologous and functional cardiomyocytes with no involvement of genetic manipulation. These cells will have a wide range of applications in the field of therapeutics, disease modeling, and drug screening. These purified bioactive recombinant proteins can potentially be a safe and effective molecular tool in the cardiac reprogramming process and other biological applications. As mentioned, some of the

proteins have shown tumor-suppressing or oncogenic activity, which can be studied further to explore their application in cancer biology.

List of Figures

List of Tables

Abbreviations

CVDs	Cardiovascular diseases
WHO	World health organization
GATA4	GATA binding protein 4
MEF2C	Myocyte enhancer factor 2C
TBX5	T-box transcription factor 5
ETS2	Avian erythroblastosis virus E26 oncogene homolog-2
MESP1	Mesoderm-specific transcription factor 1
HAND2	Heart and neural crest derivative expressed 2
E. coli	*Escherichia coli*
c-kit	Tyrosine protein kinase
Sca-1	Stem cell antigen-1
ISL1	Insulin gene enhancer binding protein 1
PDGFRα	Platelet-derived growth factor receptor-alpha
Lin^-	Lineage- negative
miR	Micro RNA
ESCs	Embryonic stem cells
iPSCs	Induced pluripotent stem cells
BMP	Bone morphogenic protein
OCT4	Octamer binding transcription factor 4
SOX2	SRY-box transcription factor 2
KLF4	Kruppel-like factor 4
c-MYC	Myc protooncogene
JAK	Janus kinase
Snai1	Snail family transcriptional repressor 1
MYOCD	Myocardin
NKX2.5	NK2 Homeobox 5
TGF- β	Transforming growth factor β
ESRRG	Estrogen related receptor gamma
ZFPM2	Zinc finger protein, FOG family member 2
9C	9 small molecules
Dmap1	DNA methyltransferase 1-associated protein 1
ARF	Alternative reading frame

GOI	Gene of interest
RefSeq	Reference sequence
HIS	Octahistidine
TAT	HIV-Transactivator of transcription
NLS	Nuclear localisation signal/sequence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
NaCl	Sodium chloride
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TBS	Tris-buffered saline
TBST	Tris-buffered saline tween-20
EDTA	Ethylenediaminetetraacetic acid
CAI	Codon adaptation index
IMAC	Immobilized metal ion affinity chromatography
Ni-NTA	Nickel-nitrilotriacetic acid
MS	Mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectroscopy
UV	Ultra violet
CD	Circular dichroism
NANOG	Nanog homeobox
PDX1	Pancreatic and duodenal homeobox 1
GLIS1	GLIS family zinc finger 1
NGN3	Neurogenin 3
BeStSel	Beta Structure Selection
MALDI-TOF	Matrix-assisted laser desorption ionization-time of flight
DMEM	Dulbecco's modified eagle medium
FBS	Fetal bovine serum
P/S	Penicillin-streptomycin
CO_2	Carbon dioxide
HFF	Human foreskin fibroblasts
HUVEC	Human umbilical vein endothelial cells
PBS	Phosphate buffered saline
DAPI	4′,6-diamidino-2-phenylindole

VEGF	Vascular endothelial growth factor
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
cDNA	complementary deoxyribonucleic acid
CAM	Chorioallantoic membrane
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
MTSS1	Metastasis suppressor 1
MTA2	Metastasis associated protein 1 family member 2
α-MHC	alpha-Myosin heavy chain
eGFP	enhanced green fluorescence protein
$CaCl_2$	Calcium chloride
HBS	HEPES buffered saline
NF-κB	Nuclear factor kappa B
PARP1	Poly(ADP-ribose)polymerase 1
ANP	Atrial naturetic peptide

Chapter 1

Introduction and Review of Literature

Brief Overview of the Chapter

Chapter 1 consists of a brief overview of cardiovascular diseases and the impact of coronary artery disease (which leads to myocardial infarction), one of the leading life-threatening diseases worldwide. It highlights the limitations of current therapeutic approaches toward myocardial infarction. The chapter also describes the current status and progress in the area of direct cardiac reprogramming, a promising approach to generate autologous cardiomyocytes to repair the damaged heart tissue due to myocardial infarction. Based on the critical review of the literature, we have selected six key transcription factors, namely GATA4, MEF2C, TBX5, ETS2, MESP1 and HAND2, which play a crucial role in cardiac reprogramming. The chapter further summarizes the overview of these six transcription factors stating their importance in cardiac reprogramming and their role in other physiological and pathological processes, including cancer. Following, the importance of generating these six transcription factors in the form of recombinant proteins was described. This chapter also focuses on the challenges associated with recombinant protein production from the heterologous system and the advantage of using recombinant proteins over their genetic counterpart in cellular reprogramming. Finally, the chapter concludes by describing the motivation and rationale behind the work carried out in this thesis towards addressing roadblocks associated with recombinant protein production and conventional direct cardiac reprogramming approaches.

1.1 Introduction

In today's world of advancing health research, cardiovascular diseases (CVDs) still remain one of the leading causes of mortality and morbidity. The World Health Organization (WHO) in 2019 reported nearly 32% of all deceased (i.e., 17.7 million) annually have succumbed to CVDs, which is expected to surpass 23.6 million by the next decade (Kaptoge et al. 2019). The primary CVD of concern is coronary heart disease (also known as ischemic heart disease), which accounts for 14.4% of these cases, closely followed by cerebrovascular disease, accounting for 11.2%. In India, the office of the Registrar General has reported that nearly 26% of mortality and morbidity has been contributed by CVDs in which coronary heart disease and cerebrovascular disease are the major contributors (Prabhakaran et al. 2016). The major causes of CVDs are the use of tobacco, alcohol consumption, stress, poor/unhealthy diet, sedentary lifestyle, etc., in our daily life.

Coronary heart disease is the main cause of myocardial infarction, commonly known as heart attack, in which the functional cardiomyocytes are affected. In a healthy human heart, the average left ventricle has roughly 4 billion cardiomyocytes, while a post-infarct heart has a myocyte shortage of about 1 billion (Murry et al. 2006). Loss of cardiomyocytes in this disease is irreversible due to the very low regenerative ability of the heart. Due to the minimal capability of cardiomyocyte regeneration, the cardiac cells undergo scar tissue formation comprising fibroblasts as a reparative response to the irreversible loss, ultimately culminating in chronic heart failure in the long run. A common therapeutic approach includes pharmacotherapy, which is mainly focused on limiting disease progression instead of repairing and restoring healthy tissue and function. Therefore, the limited efficacy of this current

treatment has generated interest in considering other viable and long-lasting therapeutic strategies.

From this aspect, a cell-based therapy appears to be a promising alternative to meet the requirements in eradicating cardiovascular-related diseases such as coronary heart disease. In injury and tissue degeneration cases, loss of functional cells is a major setback where cell therapy can provide a viable alternative by restoring the depleted functional heart cells. In such cases, allogeneic cell transplantation can bring a possible long-term solution, but it is associated with high risk of immune rejection and also makes the patient highly susceptible to other infectious diseases due to prolonged exposure to immuno-suppressants. The above-mentioned problem can be addressed by the direct reprogramming of somatic cells of the patient to generate autologous functional cardiomyocytes and transplanting these cells into the same patient. Various studies have reported the generation of functional cardiomyocytes by delivering a cocktail of transcription factors in fibroblasts (Yamakawa and Ieda 2021). These studies mostly employed viral (retroviral and lentiviral) gene delivery approaches, which are robust and efficient. But these methods have the potential to integrate into the host cell genome and have the susceptibility to form tumors, thus restricting their applicability in patient-specific therapeutic approaches. To address this, the thesis aims to establish a recombinant protein toolbox consisting of cardiac reprogramming transcription factors. This toolbox will thus serve as a safer alternative than commonly used traditional viral methods in cardiac reprogramming for the prospective derivation of integration-free functional cardiomyocytes from human fibroblasts.

The production of recombinant proteins has revolutionized the biotechnology field. So far, numerous biologically active recombinant proteins have reached the market in a short

period for various applications, such as therapeutics, diagnostics, research and understanding fundamental biological questions. Recombinant proteins are also considered to be one of the safest approaches for cell reprogramming and cancer therapy, where the protein of interest is delivered to the target cell to perform the desired function (O'Malley et al. 2009; Sommer and Mostoslavsky 2013; Nezafat et al. 2015; Kintzing et al. 2016; Serna et al. 2018; Borgohain et al. 2019; Dey et al. 2021a). In addition, recombinant proteins allow precise control over time and dosage of application of the protein of interest into the target cells, which will help researchers to identify its biological role (Borgohain et al. 2019; Dey et al. 2021a). Importantly, the recombinant protein-based approach does not modify or alter the genome of the target cells, which is ideal for generating genetically stable cells for regenerative medicine (O'Malley et al. 2009; Sommer and Mostoslavsky 2013; Borgohain et al. 2019; Dey et al. 2021a). These human therapeutic proteins are produced in the recombinant form, either from prokaryotic or eukaryotic expression systems (Borgohain et al. 2019; Dey et al. 2021a). The bacterial system is generally chosen to produce human recombinant proteins in large quantities with simple purification techniques in a cost-effective manner (Khow and Suntrarachun 2012). However, generating highly pure and stable recombinant proteins remains challenging due to several challenges, namely codon bias, gene product toxicity, mRNA stability, low protein expression, protein degradation by host cell proteases, and *in vitro* solubility and stability (Wingfield 2015). In this study, we worked on circumventing these limitations and aimed to generate a cell- and nuclear-permeant bioactive human recombinant protein toolbox comprising GATA binding protein 4 (GATA4), Myocyte enhancer factor 2C (MEF2C), T-box transcription factor 5 (TBX5), Avian erythroblastosis virus E26 oncogene homolog 2 (ETS2), Mesoderm-specific transcription factor 1 (MESP1) and Heart and neural crest derivative expressed 2 (HAND2)

proteins, which can prospectively be used for cardiac reprogramming and various biological applications.

In this thesis, we report codon optimization, cloning, expression and purification of the above-mentioned human recombinant proteins from *Escherichia coli* (*E. coli*), then determine their secondary structure and validate their transduction ability. 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 demonstrate their biological activity using different assays in mammalian cells.

1.2. Literature Review

1.2.1 Human Heart

The heart is the first organ to form during embryogenesis which pumps and circulates blood, thereby providing oxygen and nutrients to all parts of the body and contributing to the removal of metabolic wastes. It is a cone-shaped midline, valvular, muscular pump and the size of a fist. In adult humans, it weighs around 300 g and is located in the middle mediastinum of the thorax (Whitaker 2010). The heart consists of different types of cells that are responsible for the biochemical, structural, mechanical and electrical properties of the functional heart. The myocardium, which is the muscular layer of the heart consists of atrial and ventricular cardiomyocytes. Cardiomyocytes are the functional unit of the heart in which these cells are majorly responsible for the conduction system. The epicardium of the heart is made up of cardiac fibroblasts and precursors of the coronary vasculature (Xin et al. 2013). Cardiac fibroblasts play a crucial role in remodeling, i.e., changes in the myocardial organization that

helps in the adaptation of the heart to chemical/mechanical/electrical signaling changes (Brower et al. 2002; Stewart Jr et al. 2003). During this remodeling, cardiac fibroblasts either secrete or degrade the extracellular matrix based on the requirements (Chancey et al. 2002; Raffetto and Khalil 2008). Moreover, more than 50% of the heart comprise of cardiac fibroblasts (Xin et al. 2013). The endocardium comprises of endothelial cells and thereby the interior lining of blood vessels and cardiac valves. The conduction system (i.e., the pacemaker cells and Purkinje fibers) of the heart is contributed by the specialized cardiomyocytes called nodal cardiomyocytes that generate and conduct electrical impulses. These specialized cells are present in the sinoatrial node in the right atrium, which initiates heart contraction. The generated electrical impulses from the sinoatrial node are then conducted to the ventricles by the atrioventricular node (part of the conduction system) located between the upper and lower chambers of the heart (Xin et al. 2013). The unique property of the atrioventricular node is that it stimulates its slower conduction, termed decremental conduction (Patterson and Scherlag 2002).

1.2.2 Cardiovascular Diseases (CVDs)

CVDs continue to be the main cause of morbidity and mortality in the modern world, in which the structure and function of the heart and blood vessels are affected. According to WHO, CVDs include mostly the diseases of heart and blood vessels and also vascular diseases of brain (Mendis et al. 2011). Although ~90% of CVDs are preventable with medications, exercise, a healthy diet and avoidance of tobacco and alcohol (McGill Jr et al. 2008), an increase in mortality has been observed in recent times, which might be due to inadequate preventive measures taken against the same (Mendis et al. 2011).

The most prevalent CVDs are ischemic heart disease (46%) and cerebrovascular disease (34%) (Prabhakaran et al. 2016). Out of the variety of CVDs, the ones like ischemic heart disease, cardiomyopathies and arrhythmias mostly affect the functionality of cardiomyocytes (Mendis et al. 2011), mainly due to apoptosis and necrosis in the cardiac tissue. Most of these diseases are due to the loss of regenerative capacity of the host tissue by the remaining myocytes and the consequent weakening of the diseased heart over time. The loss of cardiomyocytes leads to the formation of scar tissue by the spontaneous division and migration of fibroblasts over the damaged area, which in turn, results in improper contraction. This myocardial growth transition gives rise to terminally differentiated cardiomyocytes (adult) that are characterized by bi-nucleated cells with the arrested cell cycle. Naturally, the human heart has a limited capacity to regenerate cardiomyocytes, as indicated by lasting scar tissue following myocardial infarction, and ultimately culminates in chronic heart failure in the long run.

1.2.3 Strategies to generate cardiomyocytes from different cells

Since the late 19th century, the search for effective ways to treat infarcted hearts has increased remarkably. In this regard, cell-based therapy for cardiac regeneration appears to be a promising alternative to achieve cardiac repair. In the near future, cell therapy by transplanting autologous functional cardiomyocytes into diseased patients could be a potential solution to control current epidemic rates of heart failure by providing human cardiomyocytes to support heart regeneration. Various strategies have been developed to date to make an attempt to generate cardiomyocytes from different cell types (Figure 1.1), which are explained below:

1.2.3.1 Cardiomyocytes derived from cardiac progenitor cells

Cardiac progenitor cells (also sometimes called cardiac stem cells) are a heterogeneous group of endogenous multipotent progenitor cells in the heart and are identified by the expression of various markers such as tyrosine kinase receptors (c-Kit) and/or stem cell antigen-1 (Sca-1) or the insulin gene enhancer binding protein 1 (ISL-1) or platelet-derived growth factor receptor-alpha (PDGFRα) or the ability to grow into cardiospheres (Beltrami et al. 2003; Oh et al. 2003; Messina et al. 2004; Laugwitz et al. 2005; Chong et al. 2011; Chong et al. 2014; Amini et al. 2017). In the year 2003, Beltrami et al. first demonstrated the differentiation ability of lineage-negative (Lin^-) c-kit^+ cells, isolated from the adult rat heart, towards cardiomyocytes, smooth muscle and endothelial cells (Beltrami et al. 2003). Soon after, Schneider and colleagues established the isolation of the Sca-1^+ subpopulation from adult mouse hearts and showed *in*

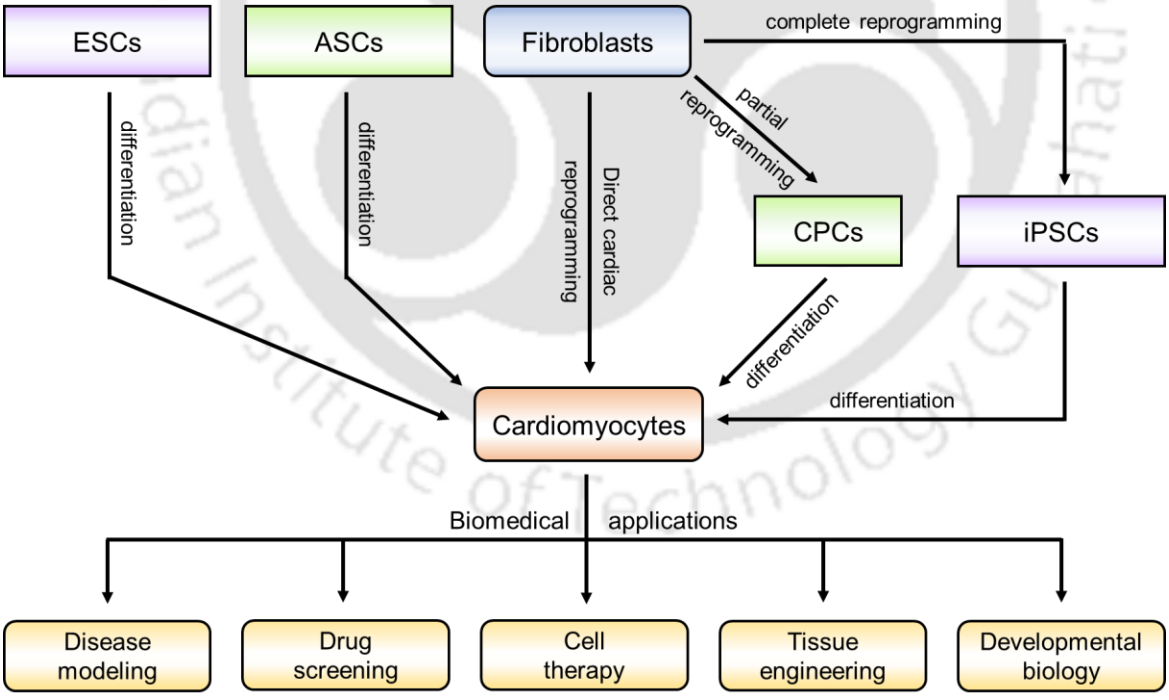


Figure 1.1 Schematic of different strategies to generate cardiomyocytes from different sources and the biomedical applications of the generated cells.

vitro differentiation of these cells into cardiomyocytes in the presence of 5-azacytidine, a DNA demethylating agent (Oh et al. 2003). Notably, the authors also demonstrated the *in vivo* differentiation of Sca-1^+ cells to cardiomyocytes after intravenous injection in mice (Oh et al. 2003). Further investigation of these cells led to the isolation of Sca-1^+ and c-kit^+ subpopulation of cardiosphere-forming cells from both mouse and human hearts (Messina et al. 2004). Interestingly, after cardiosphere formation, spontaneously beating mouse cells were reported without co-culturing with neonatal rat cardiomyocytes, unlike the human cardiospheres, which require co-culturing of rat neonatal cells (Messina et al. 2004). Another subpopulation expressing Isl1 was also discovered in rat, mouse and human hearts as clusters in atria and single cells in ventricles (Laugwitz et al. 2005). These cells lack Sca-1 and c-kit expression, and when co-cultured with rat neonatal cardiomyocytes, it differentiated into cardiomyocytes. Furthermore, studies have also demonstrated the derivation of cardiomyocytes from the PDGFRα^+ subpopulation of cells from mouse and human hearts (Chong et al. 2011; Chong et al. 2014; Le and Chong 2016). Interestingly, Raghunathan et al. demonstrated the conversion of human adipogenic mesenchymal stem cells-derived cardiac progenitor cells into pacemaker-like cells, a specialized cardiomyocyte, through the ectopic expression of SHOX2, HCN2 and TBX5 transcription factors (Raghunathan et al. 2020). In general, these cardiac progenitor cells are in an inactivated or quiescent state under normal physiological conditions. In this state, cardiac progenitor cells do not contribute to the regeneration of cardiomyocytes; however, upon cardiac injury, these progenitor cells can get activated, and subsequently differentiate into cardiomyocytes (Le and Chong 2016).

Despite the identification of these different cardiac progenitor cell populations, their physiological and pathophysiological functions are not entirely understood (Amini et al. 2017).

Notably, these cells induce unfavorable or regenerative effects upon exogenously delivered within the injured heart (Le and Chong 2016). Moreover, the molecular mechanisms behind these effects still remains unclear (Le and Chong 2016). Thus, the applications of these progenitor cells are limited to the regeneration of cardiac tissue after injury.

1.2.3.2 Cardiomyocytes derived from adult stem cells

Adult stem cells such as mesenchymal stem cells is another cell source that can be used to differentiate them into functional cardiomyocytes. Several studies on hematopoietic stem cells were inconclusive with respect to cardiac fate determination (Orlic et al. 2001; Balsam et al. 2004; Kawada et al. 2004; Murry et al. 2004). Therefore, researchers focused on non-hematopoietic stem cells, such as mesenchymal stem cells, as a primary source to obtain cardiomyocytes. In this regard, adult mouse bone marrow-derived non-hematopoietic stem cells developed features of cardiomyocytes when being treated with 5-azacytidine (Makino et al. 1999). Similarly, studies reported the formation of cardiomyocytes from mouse mesenchymal stem cells in the presence of 5-azacytidine (Hattan et al. 2005; Antonitsis et al. 2007) and also by injecting into the mouse embryos (Jiang et al. 2002). However, 5-azacytidine has been reported to induce carcinogenicity by introducing mutations in the somatic cell genome (Alagesan and Griffin 2014), thus serving as a roadblock to the therapeutic applications of differentiated cells. Therefore, Shen et al. focused on the downstream targets of 5-azacytidine and found the significant upregulation of miR-1-2 during differentiation (Shen et al. 2017). The authors demonstrated that mimics of miR-1-2 promoted the differentiation of bone marrow-derived mesenchymal stem cells into cardiomyocytes by activating the Wnt/β-catenin signaling pathway (Shen et al. 2017). Similarly, miR-1 has been reported to induce the

differentiation of mesenchymal stem cells into myocardial cells only in a specific medium, i.e., serum-free cardiomyogenic medium containing 10 nM 5-azacytidine (Zhao et al. 2016).

Alternatively, Shim et al. obtained human cardiomyocyte-like cells from adult bone marrow stem cells by treating them with a low concentration (10^{-9} M) of dexamethasone (corticosteroid) (Shim et al. 2004). On the other hand, differentiation of bone marrow-derived clonal subpopulation by co-culturing method showed phenotypes of a heterogeneous populations of cells comprising cardiomyocytes, endothelial cells and smooth muscle cells (Yoon et al. 2006). Likewise, Cai et al. employed the same co-culturing method with minor modifications (1:10 instead of 1:4 ratio) to differentiate bone marrow-derived mesenchymal stem cells into cardiomyocytes (Cai et al. 2012). Apart from 5-azacytidine and miRNAs, several growth factors/cytokines, microenvironment, Caveolin-1, Vanilloid receptor-1 and Histone deacetylase 1 were reported to induce cardiac differentiation of mesenchymal stem cells (Guo et al. 2018b). Moreover, these mesenchymal stem cells (along with fresh bone marrow) promoted the activation of angiogenesis, inhibition of fibrosis and decrease in apoptosis to restore heart function in the infarcted swine model (Pak et al. 2003). Among the various sources of mesenchymal stem cells like the umbilical cord, adipose tissue, placenta, hair follicle, skeletal muscle, etc., adipose tissue serves as an easily obtainable source compared to the invasive process of bone marrow aspiration. Kakkar et al. underscore the merits associated with the induction of adipose tissue-derived stem cells with TGF-β1, which is non-toxic and a more efficient cardiac inducer compared to 5-azacytidine (Kakkar et al. 2019). Another study focused on human amniotic fluid-derived mesenchymal stem cells, which were effectively differentiated into the cardiomyogenic lineage upon treatment with 10 μM 5-azacytidine and 20% human platelet lysate (Markmee et al. 2020).

Ramesh et al. summarized the use of various biological and chemical inducers that enable the cardiac differentiation of adult stem cells into cardiomyocytes (Ramesh et al. 2021). However, direct transplantation of these cells is limited due to low differentiation or success rate *in vivo*, and these do not entirely reciprocate the functional and morphological characteristics of cardiomyocytes (Guo et al. 2018b).

1.2.3.3 Cardiomyocytes derived from pluripotent stem cells

In 1999, Guan et al. reported differentiation of undifferentiated Embryonic Stem Cells (ESCs) into cardiomyocytes, skeletal muscle, neuronal, epithelial and vascular smooth muscle cells (Guan et al. 1999). In this study, Guan et al. concluded that the differentiation of ESCs towards cardiomyocytes was influenced by cell density, medium and its supplements, type of cells and time of seeding cells. Contrastingly, Kehat et al. demonstrated cell density independent differentiation of human ESCs towards cardiomyocytes with similar structural and functional properties of early stage cardiomyocytes (Kehat et al. 2001). In this study, the authors performed differentiation by forming embryoid bodies and then seeded these embryoid bodies in 0.1% gelatin-coated petri-dishes. They observed the first embryonic bodies with rhythmically contracting areas on day 4 (Kehat et al. 2001), which is two days earlier than the previous study (Guan et al. 1999). However, most of these studies used two different culture conditions for ESCs maintenance and differentiation. Interestingly, Denning and colleagues developed a common culture condition for the maintenance of ESCs, and subsequently demonstrated efficient differentiation of these cells towards cardiomyocytes (Denning et al. 2003).

Soon after the astounding discovery of induced Pluripotent Stem Cells (iPSCs) from mouse fibroblasts (Takahashi and Yamanaka 2006), the same group generated human iPSCs

from adult human fibroblasts and differentiated into cardiac and other cells as an evidence of pluripotency (Takahashi et al. 2007). For the directed differentiation, Yamanaka et al. followed the previously reported protocol (Laflamme et al. 2007) in which human ESCs were treated with Activin A and Bone Morphogenetic Protein (BMP) 4 to form beating cardiomyocytes. Further, induction with Activin A and BMP4 enhanced the generation of cardiomyocytes, however, there was a lot of variability between cell lines and experiments (Paige et al. 2010). Using small molecules, Lian et al. fine-tuned the Wnt/β-catenin signaling pathway to generate cardiomyocytes robustly from multiple pluripotent cells. In this study, the authors showed that Wnt signaling activation is crucial for mesoderm formation from pluripotent cells whereas its inhibition was crucial for the sequential differentiation of these cells into cardiomyocytes (Lian et al. 2012). This was further fine-tuned by Kadari et al., which reported the formation of three different phases, namely cardiovascular induction, cardiac specification and cardiomyocyte enrichment (Kadari et al. 2015). In the first phase, authors used CHIR99021 and BMP4 to stimulate cardiac induction, and then used Wnt inhibitor, XAV939, to induce cardiac specification (Kadari et al. 2015). In order to reduce line-to-line variability, they performed lactate enrichment to obtain pure populations of cardiomyocytes with a very high efficiency (Kadari et al. 2015).

Ou et al. described a protocol wherein co-culturing of iPSCs with neonatal cardiomyocytes resulted in cardiomyocytes expressing cardiac-specific genes like Mef2c, cTnT and MLC-2V (Ou et al. 2016), indicating efficient differentiation and enhanced proliferation ability upon co-culture. Studies also suggest the administration of electrical stimulations in human iPSC or cardiosphere-derived cells to achieve functionally mature cardiomyocytes (Ma et al. 2018; Nazari et al. 2020). These stimulations could mimic the native

property of synchronous contractions of the heart and aid in attaining a functionally mature state in shorter time duration (Ma et al. 2018; Nazari et al. 2020). Funakoshi et al. provided insights into the advantage of generating human PSC-derived mature cardiomyocytes that have enhanced contraction force, mitochondrial oxidative property, and improved sarcomere structure (Funakoshi et al. 2021). Transplantation of these mature cells compared to the functionally immature ones generated better grafts, suggesting the prior manipulation of the cells before the transplant could help mitigate arrhythmias and serve as a safer therapy (Funakoshi et al. 2021). Exciting opportunities exist in the field of PSC-derived cardiomyocytes, but not without a series of hurdles preventing its use in the clinical setting. Some of them include the fetal-like immature phenotype of generated cardiomyocytes, variability in the cardiac subtype, the risk of arrhythmogenesis upon transplant, teratoma formation, and immune rejection, to name a few.

1.2.3.4 Direct cardiac reprogramming strategies to derive cardiomyocytes

Direct cardiac reprogramming of mouse fibroblasts

Several studies have generated induced cardiomyocytes by directly reprogramming somatic cells using different integrative and non-integrative (Figure 1.2) approaches (Ieda et al. 2010; Fu et al. 2013; Nam et al. 2013; Wada et al. 2013; Muraoka et al. 2014; Wang et al. 2014; Lee et al. 2015; Wang et al. 2015; Cao et al. 2016; Miyamoto et al. 2018; Paoletti et al. 2020; Yamakawa and Ieda 2021). The very first study to generate fibroblast-derived induced cardiomyocytes was reported in 2010 (Ieda et al. 2010). The authors selected 13 potential cardiogenic factors critical for survival and cardiogenesis in the embryo from the previously reported microarray analysis data of variable expression patterns observed between myocytes and non-myocytes cells (Ieda et al. 2009). Additionally, Mesp1 was also included due to its

cardiac reprogramming ability in *Xenopus laevis* (David et al. 2008). Of these 14 factors, this study identified a combination of three transcription factors, namely GATA4, MEF2C and TBX5 (referred to as GMT), sufficient to reprogram mouse cardiac/dermal fibroblasts to induced cardiomyocytes (Ieda et al. 2010), bypassing the pluripotent and cardiac progenitor stem cell state. These induced cells exhibited almost similar gene expression pattern and electrophysiology and contracted spontaneously as native cardiomyocytes. However, the efficiency of the beating cells was very low, and the majority of the cell population was only partially reprogrammed. The reason behind this inefficient reprogramming is that the likelihood of transducing a single cell with all three factors is low and the imbalanced stoichiometric expression of these factors (Lee et al. 2015; Wang et al. 2015). Relatively high levels of MEF2C protein expression compared to GATA4 and TBX5 improved reprogramming efficiency and the quality of induced cardiomyocytes (Wang et al. 2015). Of the two isomeric forms of MEF2C, viz Mi2 and Mi4, the former, in combination with GATA4 and TBX5, reprogrammed mouse embryonic fibroblasts more efficiently than the latter (Wang et al. 2015). This might be the possible reason for the discrepancy in reprogramming efficiency reported by different groups (Yamakawa and Ieda 2021). Further, the addition of HAND2 to this reprogramming cocktail enhanced the efficiency, irrespective of the stoichiometry using either cocktail of retroviral-GMT vectors or retroviral single polycistronic-MGT vector (Song et al. 2012; Nam et al. 2013; Zhang et al. 2019; Wang et al. 2020a). These studies thus demonstrated the requirement of HAND2 in the GMT cocktail (GMT + HAND2) in the direct cardiac reprogramming of mouse fibroblasts into functional cardiomyocytes.

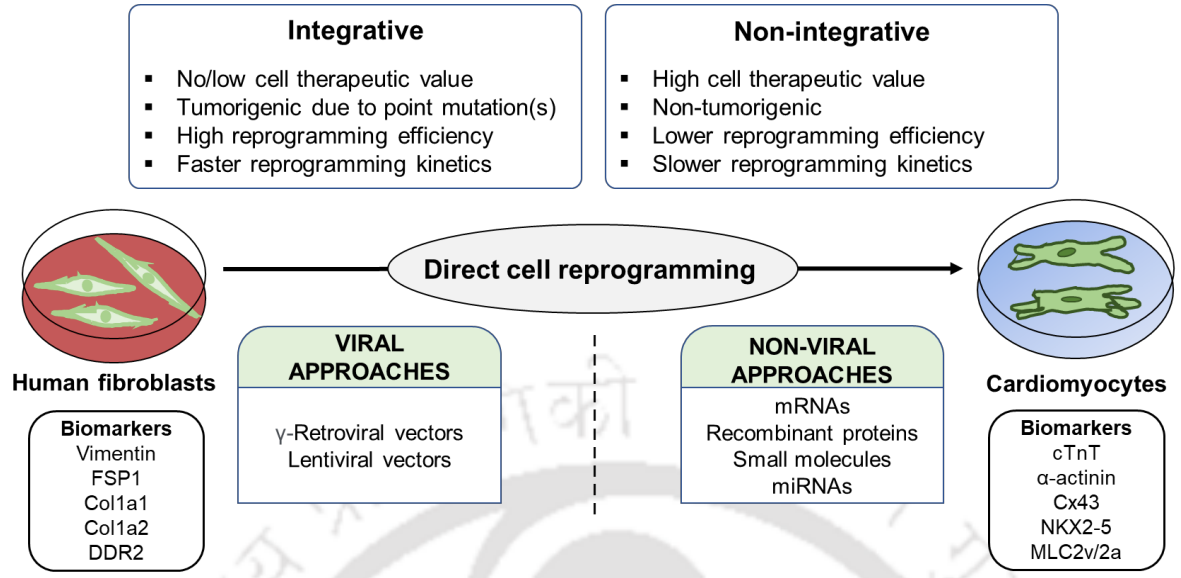


Figure 1.2 Schematic illustration of different direct cardiac reprogramming approaches. FSP1, fibroblast specific protein 1; Col1a1, collagen alpha-1(I) chain; Col1a2, collagen alpha-2(I) chain; DDR2, discoidin domain receptor 2; cTnT, cardiac Troponin T; Cx43, connexin 43; NKX2.5, NK2 Homeobox 5; MLC2v/2a, myosin light chain-2 cardiac ventricular/atrial isoforms.

Using the Yamanaka factors (OCT4, SOX2, KLF4 and c-MYC), Efe et al. partially reprogrammed mouse fibroblasts, instead of inducing full pluripotency, and then derived cardiomyocytes by diverting them towards cardiac lineage with specific media conditions (Efe et al. 2011). The first spontaneous beating was observed after 11 days using this approach (Efe et al. 2011), compared to 4-5 weeks in the first study (Ieda et al. 2010). The addition of ascorbic acid (Vitamin C) to the Yamanaka factors enhances the derivation of cardiomyocytes from mouse fibroblasts through a partial pluripotent reprogramming strategy (Talkhabi et al. 2015).With a different set of transcription factors (TBX5, MEF2C and MYOCD), cardiomyocyte-like cells were obtained by time-dependent conversion of mouse embryonic fibroblasts through the lentiviral expression of these three factors, upregulating a broader spectrum of cardiac genes (Protze et al. 2012), compared to the combination used by Ieda et

al. (Ieda et al. 2010). Similarly, screening a combination of transcription factors to establish a minimally efficient reprogramming cocktail, Hirai et al. reported the use of MYOCD with MEF2C and GATA6 to generate smooth muscle resembling cells with characteristic cardiac marker genes and reduced fibroblast-specific gene expression (Hirai et al. 2018).

Delivering miRNAs in an integration-free manner, Jayawardena et al. established the derivation of cells having characteristics of cardiomyocytes via transient expression of muscle-specific miR-1, miR-133, miR-208 and miR-499 in mouse cardiac fibroblasts by a single transfection (Jayawardena et al. 2012). The authors also reported further enhancement of cardiomyocyte-like cells by the addition of Janus kinase (JAK) inhibitor I, which is believed to induce expression of cardiac ion channels as well as enhance α-MHC (Jayawardena et al. 2012). Following this, Muraoka et al. included miR-133 in the GMT reprogramming cocktail and demonstrated a 7-fold increase in the reprogramming efficiency and improved kinetics of <12 days compared to 30 days, compared to GMT alone in mouse embryonic fibroblasts (Muraoka et al. 2014). Apart from mouse embryonic fibroblasts, Muraoka et al. showed enhanced cardiac reprogramming in adult mouse and human cardiac fibroblasts with the inclusion of miR-133 in GMT or GMT with MESP1 and MYOCD reprogramming factors (Muraoka et al. 2014). Notably, miR-133 promoted cardiac reprogramming by suppressing Snail family transcriptional repressor 1 (Snai1), a key molecular roadblock of cardiac reprogramming and a master regulator of epithelial-to-mesenchymal transition, that facilitates the repression of fibroblast gene expression (Muraoka et al. 2014). Similarly, the addition of miR-1 and miR-133 to the GMT + HAND2 and A83-01 (a TGF-β signaling inhibitor) cocktail resulted in a maximum number of spontaneously contracting cells from mouse embryonic fibroblasts (Zhao et al. 2015).

Using a calcium indicator GCaMP, Addis et al. constructed a reporter system driven by cardiac-specific troponin-T promoter and demonstrated that NKX2.5 in combination with GMT + HAND2 (GMT + HAND2 + NKX2.5) is the most effective cocktail in cardiac reprogramming, resulting in 50-fold increase in cardiomyocyte formation efficiency compared to GMT factors (Addis et al. 2013). This GMT + HAND2 + NKX2.5 combination in the presence of SB432542, a potent small molecule that inhibits TGF-β signaling pathway, further improved the reprogramming of mouse embryonic and adult fibroblasts up to 5-fold (Ifkovits et al. 2014). Similarly, Zhao et al. showed that the suppression of pro-fibrotic signaling by TGF-β or Rho-associated kinase inhibitors, remarkably enhanced the efficiency by up to 60% and dramatically enhanced the kinetics of cardiac reprogramming of mouse embryonic fibroblasts (Zhao et al. 2015).

Although these conventional methods using retroviral and lentiviral vectors are efficient and robust, they result in random viral integration in the host cell genome, leading to insertional mutagenesis and tumorigenicity (Zhao et al. 2015; Borgohain et al. 2019; Haridhasapavalan et al. 2019; Dey et al. 2021a). To circumvent these issues, non-integrative strategies of reprogramming are developed with minimal or no genetic modifications (Zhao et al. 2015; Borgohain et al. 2019; Haridhasapavalan et al. 2019; Dey et al. 2021a). In an effort to develop clinically suitable direct reprogramming strategies, Wang et al. established a defined cocktail of small molecules (SB431542, CHIR99021, Parnate and Forskolin) that, along with a single pluripotency factor, OCT4, reprogrammed mouse fibroblasts to ventricular-like cardiomyocytes through a cardiac progenitor state (Wang et al. 2014). Following this, two different groups reported the generation of mouse cardiomyocytes using different combinations of small molecules to generate these cells in an integration-free manner (Fu et

al. 2015; Park et al. 2015). Remarkably, Fu et al. reprogrammed mouse fibroblasts using only a cocktail of small molecules (without any transcription factors), namely CHIR99021, RepSox, Forskolin, valproic acid, Parnate and TTNPB, into chemical-induced cardiomyocyte-like cells pass through a cardiac progenitor state and bypassing pluripotent state (Fu et al. 2015). Interestingly, the same cocktail has been used previously by the same group to generate chemical-iPSCs (Hou et al. 2013); however, the culture condition appears to play a crucial role in generating desired cells. Similarly, Park et al., with a different combination of small molecules (Forskolin, A-8301, SC1, CHIR99021, and BayK 8644), converted mouse fibroblasts into cardiomyocyte-like cells (Park et al. 2015) passing through a progenitor state. In both these small molecule-based studies, the reprogramming efficiency of mouse tail tip fibroblasts was very low compared to embryonic fibroblasts (Fu et al. 2015; Park et al. 2015). Alternatively, a novel integration-free approach employing multiple mRNA transfections of GMT for just two weeks reprogrammed mouse cardiac fibroblasts directly to cardiomyocytes (Lee et al. 2015). This study demonstrated that the expression of cardiomyocyte-specific marker genes is dependent on the stoichiometric ratio of GMT mRNAs (Lee et al. 2015). However, in all these studies, reprogramming efficiency was compromised significantly compared to integrating viral vectors.

Considering the risk factors associated with integrating vectors as well as the requirement of transgene expression for a limited duration to bring about reprogramming, the focus now shifted toward finding a more clinically relevant set of non-integrating vectors for direct cardiac reprogramming. Keeping this in mind, several studies reported the use of integration-free adenoviral or adeno-associated viral or Sendai viral vectors to safely and effectively generate induced cardiomyocytes from mouse fibroblasts (Mathison et al. 2017;

Miyamoto et al. 2018; Yoo et al. 2018; Isomi et al. 2021). One such approach involved the administration of adenoviral-GMT vectors in post-infarct rats, successfully bringing about the transdifferentiation of fibroblasts into cardiomyocytes as well as ameliorating the cardiac condition (Mathison et al. 2017). This strategy proved as effective as the use of lentiviral-GMT vectors, generating ~6% cardiac troponin-T expressing cells with an elevated ejection fraction compared to the control, further suggesting that the transient expression of reprogramming factors using adenoviral vectors could be sufficient in bringing about cardiac reprogramming (Mathison et al. 2017). In an effort to accomplish *in vivo* direct cardiac reprogramming, Kisby et al. utilized direct adenoviral vector injection to transiently induce the expression of pluripotency factors (Yamanaka factors) in healthy and post-infarct mouse hearts (Kisby et al. 2021). This resulted in partial reprogramming and therefore was insufficient to regenerate the post-infarct heart. Remarkably, Yoo et al. demonstrated an *in vivo* chimeric approach using an adeno-associated viral vectors encoding the combination of GMT along with thymosin β4 (an antifibrotic angiogenic protein), that facilitated regeneration in the injured heart by promoting upregulation of cardiac-specific genes as well as a gradual downregulation of fibrosis-specific genes (Yoo et al. 2018). In the near future, the use of thymosin β4 will serve as a probable clinical aid in cardiovascular regeneration (Yoo et al. 2018).

In another effort to bring about reprogramming using Sendai viral-GMT vectors, Ieda and colleagues reported the efficient generation of cardiomyocytes with a 100-fold enhanced beating compared to the use of retroviral-GMT from mouse cardiac fibroblasts in both *in vitro* and *in vivo* (Miyamoto et al. 2018). As early as 7-days post-transduction, Sendai virus-mediated *in vivo* reprogramming resulted in the generation of cardiomyocytes, which enhanced cardiac function in the post-infarct heart after 4 weeks of transplantation in immunodeficient

mice (Miyamoto et al. 2018). The same group reproduced Sendai virus-mediated *in vivo* reprogramming in immunocompetent mice and showed that the generated cardiomyocytes not only managed to exist in post-infarct heart conditions for up to 4 weeks but also reduced fibrosis by collagen suppression, thereby improving cardiac function for at least 12 weeks (Isomi et al. 2021). This study reported a reprogramming efficiency of 2.5% after 4 weeks of viral transduction (Isomi et al. 2021), whereas the previous study showed an efficiency of 1.5% after 1 week of viral transduction (Miyamoto et al. 2018).

Direct cardiac reprogramming of human fibroblasts

In contrast to the murine cardiac reprogramming, several studies have demonstrated that GMT cocktail is insufficient to reprogram human fibroblasts to cardiomyocytes (Fu et al. 2013; Nam et al. 2013; Wada et al. 2013; Muraoka et al. 2014; Singh et al. 2016; Yamakawa and Ieda 2021). These results concluded that additional factors may be required to induce cardiac reprogramming in human somatic cells. Therefore, Nam et al. included MYOCD and muscle-specific miRNAs (miR-1 and miR-133) in combination with GMT to reprogram human fibroblasts into cardiomyocytes (Nam et al. 2013). The authors reported a heterogenous population of reprogrammed cells with varying levels of cardiac gene expression, and only a small subset of cells showing spontaneous contractility. Interestingly, another study reported that MESP1, along with the ETS2 transcription factor, reprogrammed human dermal fibroblasts to cardiac progenitor cells via lentiviral-based approach; however, the recombinant protein forms of the same transcription factors when co-transduced in human dermal fibroblasts resulted in the derivation of immature cardiomyocytes (Islas et al. 2012). Subsequently, the inclusion of MESP1 and ESRRG/NR3B3 (a nuclear receptor that plays a critical role in mitochondrial biogenesis) to GMT cocktail reprogrammed human fibroblasts

derived from different sources such as ESCs, fetal heart and neonatal skin (Fu et al. 2013). This study showed phenotypic shift and induced global cardiac gene-expression in the starting cell type, albeit at a low frequency (Fu et al. 2013). Fu et al. further showed enhanced reprogramming with the addition of MYOCD and ZFPM2 (FOG2), but the efficiency still remained low (Fu et al. 2013). Notably, they also reported the importance of TGF-β signaling in the cardiac reprogramming of human cells. Similarly, by adding MESP1 and MYOCD to the reprogramming cocktail (GMT + MESP1 + MYOCD), Wada et al. reported the successful generation of cardiomyocyte-like cells from human fibroblasts (Wada et al. 2013). In order to further understand the time course of induction, Wada et al. constructed doxycycline-inducible lentiviral vectors and found that the induction of GMT + MESP1 + MYOCD expression for 2 weeks is sufficient for the stable conversion of human fibroblasts to cardiomyocytes (Wada et al. 2013). These studies demonstrated that, unlike mouse cells, reprogramming human cells is time-consuming and inefficient (Fu et al. 2013; Nam et al. 2013; Wada et al. 2013). Further addition of miR-133 to the human cardiac fibroblasts overexpressing GMT + MESP1 + MYOCD improved cardiac reprogramming efficiency as well as the kinetics by directly suppressing Snai1 (Muraoka et al. 2014). This can be correlated with the Nam et al. study (Nam et al. 2013), which collectively suggests that suppressing Snail is essential for efficient cardiac reprogramming of human cells.

Using a different combination of transcription factors and miRNA, Singh et al. demonstrated that human cardiac fibroblasts, upon treatment with GMT, HAND2 and MYOCD along with miR-590, resulted in cardiac troponin-T expressing cells that exhibited spontaneous contractions. Particularly, the addition of miR-590 serves as a suitable alternative to enhance the reprogramming efficiency by suppression of Specificity Protein 1 (Singh et al.

2016), a zinc finger protein that regulates fibrosis genes (Verrecchia et al. 2001). Another study demonstrated transdifferentiation of human dermal fibroblasts towards cardiac cell lineage using the same Nam et al. cocktail along with NKX2.5 transcription factor (Christoforou et al. 2017). Further, it has been shown that the efficiency can be increased either by the inhibition of Janus kinase 1 or Glycogen synthase kinase 3 or by the addition of NRG1, a protein that promotes cardiomyocyte proliferation and myocardial regeneration (Christoforou et al. 2017).

Remarkably, in an effort to generate integration-free cardiomyocytes, Ding and colleagues reported chemically induced cardiomyocytes that exhibited uniform contractility as a result of treating human fibroblasts with a defined cocktail of 9 small molecules (9C), including those molecules that have been reported to downregulate fibroblasts-specific genes (Cao et al. 2016). These 9C-treated fibroblasts have the ability to repair infarcted mouse hearts upon transplantation. Using a combinatorial approach with a defined set of small molecules and GMT + HAND2 , Singh et al. demonstrated the reprogramming of human fibroblasts towards cardiomyocytes, exhibiting calcium transients and spontaneous beating in co-culture conditions (Singh et al. 2020). Using another non-integrative approach that included miR-1, miR-133, miR-208 and miR-499 (miR-combo), the authors were able to generate immature cells expressing cardiac troponin-T and exhibiting spontaneous calcium oscillation after 15- and 30-days post-transfection, respectively (Paoletti et al. 2020). Similarly, another group used non-integrative Sendai viral vectors to deliver GMT + MESP1 + MYOCD [as mentioned in (Wada et al. 2013)] or GMT + MESP1 + MYOCD/miR-133 [as mentioned in (Muraoka et al. 2014)] in human cardiac fibroblasts, resulting in the generation of cardiac troponin-T^+ cells (Miyamoto et al. 2018). Interestingly, three-fold higher cardiac troponin-T^+ cells were reported with Sendai viral vectors-GMT + MESP1 + MYOCD in combination with miR-133 compared

to the one without miR-133 (Miyamoto et al. 2018). Further, the authors demonstrated synchronous beating of cardiac troponin-T^+ cells from human cardiac fibroblasts transduced with Sendai viral vectors-GMT + MESP1 + MYOCD + miR-133 when co-cultured with neonatal rat cardiomyocytes. To summarize, human cardiac reprogramming is generally less efficient and has slower kinetics than murine cardiac reprogramming (Chen et al. 2017). Therefore, further refinements are required to enhance the efficiency and kinetics with a robust, clinically applicable strategy.

Alternatively, Yu et al. employed a transgene-free CRISPR-based approach to knockout Dmap1 (DNA methyltransferase 1-associated protein 1) and revealed its significance in regulating the reprogramming process (Yu et al. 2019). This novel therapeutic intervention modulates the process by reducing promoter methylation and elevating the expression of Nkx2-5 by upto 11% (Yu et al. 2019). Another study demonstrates the use of lineage reprogramming to generate induced cardiac progenitor cells via CRISPR/Cas9-mediated transcription activation (Wang et al. 2020b). Targeting endogenous cardiac factor combination of GMT + HAND2, Wang et al. successfully reprogrammed human fibroblasts into cardiac progenitor cells that have the potential to differentiate further into induced cardiomyocytes (Wang et al. 2020b). The use of such novel technologies, apart from the conventional reprogramming methods, may serve as a potential approach to patient-specific cardiac cell therapy.

These studies clearly demonstrate that fibroblasts can be directly reprogrammed towards the cardiac lineage and lay the foundation for future refinements with respect to the inclusion of other essential factors in the reprogramming cocktails, improving reprogramming efficiency and understanding the underlying mechanisms of the cardiac reprogramming

process. Further extensive studies are required to explore their *in vivo* reprogramming potential for regenerative therapies.

1.2.3.5 Role of cardiac reprogramming transcription factors in cardiac fate determination

Direct reprogramming or direct conversion of adult somatic cells to functional cardiomyocytes offers a promising alternative in cell therapies. Several groups have generated cardiomyocytes by directly reprogramming mouse and human fibroblasts by ectopic expression of various cardiac reprogramming transcription factors using different strategies as described in the previous section. Based on these studies, we have selected GMT, as they are the core-cardiac reprogramming transcription factors. This GMT combination is essential and sufficient in murine cardiac reprogramming, whereas it is essential but insufficient in human cardiac reprogramming (Chen et al. 2017). Therefore, we have selected two more factors, namely ETS2 and MESP1, that are reported to reprogram human dermal fibroblasts to cardiac progenitor cells (Islas et al. 2012). In addition, ETS2 has been reported to stimulate the expression of core cardiac-specific transcription factors (Islas et al. 2012), while MESP1 reported to be expressed in cardiac progenitor cells and also programs nascent mesoderm toward a cardiovascular cell fate (Bondue et al. 2008). Lastly, we have selected HAND2 as it plays an important role in the maturation of the cardiomyocytes by inducing the expression of cardiac troponin-T and tropomyosin (Chen et al. 2017). The functional mechanisms of selected (GATA4, MEF2C, TBX5, ETS2, MESP1 and HAND2) cardiac-specific transcription factors are discussed below.

GATA4

GATA4 belongs to the family of GATA transcription factors that has a pair of highly conserved zinc-finger domains, which recognizes the core WGATAR (W: A/T; R: A/G) motif present in the promoters and cis-regulatory elements of numerous target genes (Arceci et al. 1993). During mouse embryogenesis, GATA4 is expressed in the primitive endoderm (hypoblast), extra-embryonic endoderm, and in cells associated with cardiac and gonadal development (Arceci et al. 1993; Heikinheimo et al. 1994). In addition, it is also crucial for the induction of definitive endoderm, and therefore essential for the formation of endodermal lineages (Viger et al. 2008). The deletion of the *Gata4* gene in mice resulted in early embryonic lethality with severe developmental defects (Kuo et al. 1997; Molkentin et al. 1997; Rojas et al. 2008). These mice failed to form a proper heart tube and showed defects in ventral morphogenesis (Kuo et al. 1997; Molkentin et al. 1997), indicating that it is a critical factor during the early stage of embryonic development. It is also a crucial factor for the survival and proliferation of both embryo and adult cardiomyocytes and vital for proper cardiac development (Rojas et al. 2008), as its loss perturbed cardiomyocyte formation and resulted in acardia in mice (Zhao et al. 2008). Besides, mutations in the human *GATA4* gene showed cardiac septal defects and dilated cardiomyopathy (Li et al. 2014). Recently, the role of GATA4 has been implicated in the development of the heart, neural crest and craniofacial skeleton (bone and teeth) in mice (Guo et al. 2018a). All these studies implicate GATA4 as a crucial factor for proper endoderm, gonadal, neural crest, and cardiac development.

In addition, various studies have identified GATA4 as a core reprogramming factor in the cocktail of transcription factors to generate functional cardiomyocytes from mouse and human fibroblasts using different gene delivery approaches (Ieda et al. 2010; Nam et al. 2013).

These studies have demonstrated that GATA4 is an indispensable cardiac reprogramming factor for reprogramming non-myocyte cells to derive functional cardiomyocytes. Notably, GATA4 was identified as a tumor suppressor protein in different types of cancers (Gong et al. 2018; Gao et al. 2019; Han et al. 2019; Xiang et al. 2019). This novel function of GATA4 has improved its prospects to be used as a potential anti-cancer agent for the treatment of various cancers. In the current study, we aimed to produce pure GATA4 recombinant protein, which can further be used for cell reprogramming, as an anti-cancer agent, and for various biological applications.

MEF2C

MEF2C is a muscle-specific factor belonging to the MADS box family transcription factors. It has a DNA binding domain called MADS box domain, which binds to A/T rich regulatory regions and regulates several cardiac and skeletal muscle genes (Edmondson et al., 1994; Subramanian and Nadal-Ginard, 1996). It collaborates with the MyoD family of basic helix-loop-helix transcription factors to play a vital role in skeletal muscle development (Edmondson et al., 1994; Potthoff et al., 2007), especially in sarcomere integrity (Potthoff et al., 2007). Primarily, MEF2C is reported to be expressed in cardiac and skeletal muscle, spleen, and brain (Martin et al., 1993). During early embryogenesis, MEF2C is expressed in myocardiogenic precursor cells, somites, endochondral cartilage and early cardiac cells (Arnold et al., 2007; Edmondson et al., 1994; Subramanian and Nadal-Ginard, 1996).

In addition, MEF2C is vital for anterior heart field development. It has been reported that MEF2C is activated by ISL1 and a cardiac-specific transcription factor, GATA4, by binding to its enhancer, aiding in the development of right ventricle (Dodou et al., 2004). Also, the homozygous inactivation of *MEF2C* leads to embryonic lethality on E9.5 with severe

cardiovascular development defects (Lin et al., 1998). Nevertheless, conditional knockout of *MEF2C* in developing bone lead to the death of the neonate within a week. All the bones in the neonatal body failed to ossify, illustrating the importance of MEF2C in bone development (Arnold et al., 2007). Furthermore, MEF2C have a prominent role in chondrocyte hypertrophy and vascularization (Arnold et al., 2007).

Apart from muscle and bone development, the expression of MEF2C is crucial for craniofacial development (Verzi et al., 2007). Conditional knockout of *MEF2C* in neural crest leads to the death of the offsprings within an hour of their birth due to defective tongue positioning leading to constricted airway, with severe defects in craniofacial structures (Verzi et al., 2007). Also, MEF2C assists in the differentiation of neural stem and progenitor cells whose conditional knockout in the latter lead to abnormal electrophysiology and behavioural defects (Li et al., 2008). As mentioned earlier, MEF2C plays a critical role in cardiogenesis and hence, it is one of the core factors used in direct cardiac reprogramming (Fu et al., 2013; Ieda et al., 2010; Wada et al., 2013).

TBX5

The *TBX5* gene is a transcription factor that codes for the protein named TBX5. TBX5 (518 amino acids) belongs to the T-box family transcription factor, which has a highly conserved TBX domain located between amino acid residues 56 and 236 (Steimle and Moskowitz 2017). It is expressed in the epicardium, myocardium and endocardium of the embryonic and adult heart (Hatcher et al. 2000). Deletion of *TBX5* gene results in embryonic lethality at day 10.5 (E10.5) due to arrest in heart development at E9.5 (Bruneau et al. 2001) and results in outgrowth defects of forelimb buds (Agarwal et al. 2003; Rallis et al. 2003). Pathogenic mutations or variants in *TBX5* gene cause Holt-Oram syndrome, characterized by variable

cardiac abnormalities and upper limb malformations (Patterson et al. 2020). TBX5 activates genes responsible for differentiation and development of embryonic heart (in humans). It is particularly essential for the formation of a wall (septum) separating the right and left sides of the developing heart (Takeuchi et al. 2003; Boogerd and Evans 2016). It is also crucial for the conduction of electrical system that controls the synchronized contractions of the heart chambers (Steimle and Moskowitz 2017). In addition, TBX5 is crucial factor in the reprogramming cocktail to derive cardiomyocytes from mouse and human fibroblasts (Ieda et al. 2010; Islas et al. 2012; Nam et al. 2013; Wada et al. 2013). All these studies highlight the importance of TBX5 in forelimb and heart development, and direct cardiac reprogramming. However, the exact mechanism how TBX5 regulates gene expression during development and cell reprogramming to perform these critical functions is still under investigation. A simple and convenient recombinant tool with methodological advancement can help advance research to understand TBX5 function and regulation.

ETS2

ETS2 belongs to the ETS family of transcription factors and has a specific winged helix-turn-helix DNA binding domain called the ETS domain (Donaldson et al. 1996). This domain is composed of three α-helices and four-stranded anti-parallel β-sheets scaffold that recognizes the DNA sequence 5'-GGA(A/T)-3' in gene promoters to regulate their transcription (Donaldson et al. 1996). Targeted deletion of the ETS domain in this transcription factor resulted in embryonic lethality in mice (Yamamoto et al. 1998). ETS2 is critical for the key events of embryonic development such as the formation of posterior character in the epiblast (Georgiades and Rossant 2006), anterior-posterior patterning (Georgiades and Rossant 2006), initiation of mesoderm formation (Georgiades and Rossant 2006), and trophoblast formation

(Yamamoto et al. 1998; Georgiades and Rossant 2006). In addition, the optimum level of ETS2 is essential for the survival of cardiomyocytes (Sheydina et al. 2012), and regulates the core cardiac transcription factors to efficiently generate cardiac progenitors (Islas et al. 2012). Furthermore, ETS2 is one of the core factors essential for reprogramming fibroblasts into induced trophoblast stem cells (Kidder 2020). Apart from these, ETS2 is known to exhibit both tumor promoting and suppressive effects in different types of carcinomas (Seth and Watson 2005; Kabbout et al. 2013; Liu and Kang 2017; Fry and Inoue 2018). Further, it has been reported that ETS2 has higher affinity towards mutant p53 to promote metastasis in different types of cancer (Do et al. 2012; Liu and Kang 2017; Fry and Inoue 2018).

MESP1

The *MESP1* gene maps on to human chromosome 15q26.1, which contains two exons (807 nucleotide transcript) and together encodes for a basic helix-loop-helix transcription factor with 268 amino acids referred to as MESP1. It was identified as the earliest marker of the cardiovascular lineage and a crucial factor in regulating heart morphogenesis (Saga et al. 1999; Saga et al. 2000; Liu 2017). Specifically, it binds to the canonical E-box motif (CACGTG) and activates mesendoderm lineage-specific genes (Soibam et al. 2015). During embryonic development, the transient expression of MESP1 was primarily reported in the heart, cranial and intersomitic vessels, dorsal aorta, and the amnion contiguous to the closing foregut (Liang et al. 2015). It is a critical regulator of cardiovascular differentiation in mammalian pluripotent stem cells and cardiovascular development in vertebrates (David et al. 2008; Lindsley et al. 2008). Besides cardiac development, MESP1 promotes epithelial-mesenchymal transition (Lindsley et al. 2008; Tandon et al. 2019), regulates cell migration (Chiapparo et al. 2016), and also promotes hematopoietic and skeletal myogenic development in a context-dependent

manner (Chan et al. 2013; Liang et al. 2015). *Mesp1* knockout mice develop cardiac malformation leading to cardia bifida (Saga et al. 1999; Liang et al. 2015; Liu 2017), and further resulting in embryonic lethality (Bondue and Blanpain 2010). Thus, MESP1 is a crucial factor for early embryonic development and lineage specification. In the cell reprogramming process, MESP1 is a key member in the reprogramming cocktail to generate functional cardiomyocytes from adult somatic cells (Islas et al. 2012; Fu et al. 2013; Wada et al. 2013; Muraoka et al. 2014; Liu 2017; Miyamoto et al. 2018). Notably, viral-mediated co-expression of MESP1 and ETS2 in human fibroblasts activated the cardiac-specific program to give rise to cardiac progenitors and were differentiated into mature cardiomyocytes (Islas et al. 2012). Recently, it has also been reported that aberrant expression of MESP1 promoted tumorigenesis in lung cells (Tandon et al. 2019). The tumorigenic tendency of lung cancer-derived cells was reported to be dependent on MESP1 expression for their proliferation, colony formation, and subcutaneous tumor formation (Tandon et al. 2019). This attribute was corroborated by the loss of alternative reading frame (ARF) tumor suppressor protein in murine fibroblasts after ectopic expression of MESP1 (Tandon et al. 2019).

HAND2

HAND2 is a member of the basic helix-loop-helix family of transcription factors that have a consensus DNA binding sequence 5'-CANNTG-3' (Srivastava et al. 1997). Transcript of HAND2 gene is 2780 bp long, containing two exons. The coding sequence is 654 bp in length which translates to the protein of 218 amino acids. HAND2 is highly expressed in maternal decidua (Cross et al. 1995) and adult heart, liver, and testes (Cross et al. 1995). The knockout of the *HAND2* gene in mice leads to defects in ventricle formation, eventually leading to embryonic lethality (Srivastava et al. 1997; Liu et al. 2009). A recent study reported that

HAND2 regulated genes involved in the atrioventricular canal and cardiac valve development (Laurent et al. 2017). Besides, HAND2 governs the development of epicardium, vascularization and angiogenesis, second heart field development and survival, and cardiac neural crest development (George and Firulli 2019). All these studies indicate the role of HAND2 in cardiac development, both in neural crest-derived and mesoderm-derived structures.

Apart from these, HAND2 plays a critical role in the development of other tissues during mouse embryogenesis. It is known to have a significant role in limb and branchial arch development (Liu et al. 2009; Osterwalder et al. 2014). HAND2 regulates the anterior-posterior polarity of limb bud through a chain of downstream transcriptional regulators (Osterwalder et al. 2014). It also aids in craniofacial development (Liu et al. 2009), and is essential for developing the sympathetic nervous system in humans, especially noradrenergic neurons (Hendershot et al. 2008).

In addition, HAND2 is crucial for the formation and maturation of cardiomyocytes (Srivastava et al. 1997). Therefore, it is also regarded as one of the most crucial cardiac reprogramming factors to derive functional cardiomyocytes. Several studies have used the genetic form of HAND2 in their reprogramming cocktail, and its inclusion resulted in higher efficiency than the original GATA4, MEF2C and TBX5 combination (Song et al. 2012; Addis et al. 2013; Nam et al. 2013). Recently, the role of HAND2 in altering the chromatin accessibility and gene expression in fibroblasts to convert them to cardiac pacemaker-like cells has also been reported (Fernandez-Perez et al. 2019).

1.2.4. Recombinant proteins

Recombinant protein production has transformed the field of biotechnology, and to date, numerous recombinant proteins have been commercially used for various biomedical and research applications (Nezafat et al. 2015). The role of recombinant proteins in therapeutics has been indispensable for the past four decades and will continue to be so. Production of recombinant proteins using different host organisms, such as algae (Specht et al. 2010), bacteria (Overton 2014), yeast (Mattanovich et al. 2012), insects (McCarroll and King 1997), and mammalian cells (Wurm and Bernard 1999) has so far proven to be a complex but an effective process (Chen 2012). The most commonly used expression host system for recombinant protein production is the bacterial system, especially *E. coli* strains, due to easy handling and maintenance, well-studied genetics, well-understood cell machinery, high protein yield, and so forth (Huang et al. 2012; Baeshen et al. 2015; Al-Hejin et al. 2019). Once expressed, these proteins are purified using a wide range of purification tags. The most widely used tag for purification is the poly-histidine tag since it is inexpensive and does not alter the characteristics of the proteins (Young et al. 2012; Wood 2014).

Several studies have reported that recombinant proteins to be safe for cell reprogramming and cancer therapeutics (O'Malley et al. 2009; Sommer and Mostoslavsky 2013; Nezafat et al. 2015; Kintzing et al. 2016; Dey et al. 2017; Borgohain et al. 2019; Dey et al. 2021a). The introduction of recombinant proteins into mammalian cells has been proven to be an effective alternative since it does not integrate and alter the genome, and also manipulation of cell fate can be done in a time- and dosage-dependent manner (O'Malley et al. 2009; Sommer and Mostoslavsky 2013; Dey et al. 2017; Borgohain et al. 2019; Dey et al. 2021a).

Although recombinant protein-based therapeutics or cellular reprogramming is a safer and promising alternative, there are certain challenges associated with the heterologous production of these recombinant proteins (Dey et al. 2017). First, the problem lies with codon usage bias (Dey et al. 2017). This problem can be overcome by performing codon optimization, suitable for expression in the desired heterologous system (Burgess-Brown et al. 2008; Maertens et al. 2010). Next, the *in vitro* protein solubility and proper native folding conformation can be a major complication due to variations in physiological conditions (Dey et al. 2017). Lastly, obtaining these proteins routinely in a pure and stable form is highly challenging (Dey et al. 2017). All of these complications can severely affect the final yield as well as the biological activity of recombinant proteins. In this study, we worked on circumventing these limitations and aimed to generate a cell- and nuclear-permeant bioactive human recombinant protein toolbox comprising GATA4, MEF2C, TBX5, ETS2, MESP1 and HAND2 proteins, which can prospectively be used for cardiac reprogramming and various biological applications.

1.3 Scope and significance of the study

The current need for a safer alternative to repair injured or damaged heart tissue due to the limitations of current clinical therapies (excluding heart transplantation), which can only prevent further cardiac tissue damage or assist affected heart functionality, motivated us to conduct the present study. In this perspective, the direct cardiac reprogramming of adult somatic cells to functional cardiomyocytes is reported to be one of the promising alternatives for cardiac repair. In 2010, a remarkable study demonstrated the reprogramming of fibroblasts, which plays an important role in scar tissue formation in an injured heart, to induced cardiomyocytes using a cocktail of three transcription factors (GATA4, MEF2C and TBX5),

bypassing the pluripotent state. Following this, several studies reprogrammed fibroblasts to induced cardiomyocytes using different combinations of transcription factors. However, almost all these studies used integration-based reprogramming approaches to derive induced cardiomyocytes. The major problem with the generation of induced cardiomyocytes using integrating approaches is limiting its application for clinical therapies. Another issue is the low reprogramming efficiency and slow kinetics due to the presence of various reprogramming barriers that prevent the efficient derivation of induced cardiomyocytes. These limitations have broadened the scope of developing a non-integrative-based approach to derive integration-free cells in cardiac reprogramming. Recombinant protein-based cellular reprogramming has been reported as the safest non-integrative approach among the other approaches. This further opened scope for us to design a simple and systematic strategy to generate the reprogramming transcription factors in the form of recombinant proteins that can further potentially replace their genetic and viral counterparts in the cardiac reprogramming of fibroblasts to induce a cardiac transcriptional profile in an integration-free manner.

We aim here to establish a recombinant protein toolbox consisting of key cardiac reprogramming transcription factors for the prospective generation of transgene-free functional cardiomyocytes. The recombinant proteins generated in this study by employing simple and systematic strategies will have native-like folding conformations, thereby retaining their biological activity. In addition, the generated recombinant proteins can be directly delivered to the mammalian cells and can be localized into the nucleus without any addition of transduction reagents. In the near future, this established recombinant protein toolbox will pave the way to generate transgene-free patient-specific autologous and functional cardiomyocytes with no involvement of genetic manipulation. Since functional cardiomyocytes will be

transplanted back to the same patient from whom fibroblasts will be isolated for direct reprogramming, there are no chances of immune rejections. Further, this established protein toolbox allows for precise control over time and dosage of application to understand the stage-specific role of individual factors in the reprogramming paradigm. Our aim was to generate a recombinant protein toolbox using systematic strategies with simple, cost-effective and high reproducibility.

Motivated by the current scope and necessities in the field, we had designed the following three objectives in this thesis work.

1.4 Objectives

1. Gene cloning and identification of optimal expression parameters to achieve maximal soluble expression of recombinant proteins.
2. Purification, biochemical and biophysical analysis of recombinant proteins.
3. Demonstration of cell penetration, nuclear translocation and biological activity of recombinant proteins.

Chapter 2

Gene cloning and identification of optimal expression parameters to achieve maximal soluble expression of recombinant proteins

Brief Overview of the Chapter

Obtaining maximal soluble expression of recombinant protein in their native form is highly challenging but pivotal in heterologous protein production. In this context, optimizing the expression parameters is the key to the successful soluble expression of recombinant proteins. Here, we demonstrated the construction of recombinant plasmids bearing the six selected human transcription factors having a critical role in cardiac reprogramming, cancer and other biological processes, and identified the optimal expression conditions for their heterologous expression in *E. coli*. To achieve this, we first retrieved their full-length protein-coding sequences from NCBI's RefSeq database and then performed codon optimization. The codon-optimized sequences were further validated using the *in silico* tools and fused with a set of tags for cell penetration, nuclear translocation, and affinity purification either at the 5' or the 3'end. We then synthetically synthesized these gene inserts and cloned them individually into the pET28a(+) protein expression vector. The resulting plasmids were then individually transformed and expressed in the *E. coli* strain, BL21(DE3). In this chapter, we next focused on identifying the optimal expression conditions for the maximal soluble expression of these transcription factors. Notably, based on the protein solubility, quality and expression level, we have selected appropriate GATA4, MEF2C, TBX5, ETS2, MESP1 and HAND2 genetic constructs for further experiments. Interestingly, our results signified that the optimal expression conditions for each transcription factor varied based on their nature. We thus conclude the importance of identifying the optimal expression conditions and influence of the position of fusion tags on these recombinant protein expressions in terms of quality and quantity.

2.1 Materials and Methods

2.1.1 Plasmid constructs

The pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH gene constructs: The full length coding sequence of selected six human gene-of-interest (*GOI)* were retrieved from the NCBI reference sequence (RefSeq) database (Table 2.1). These sequences were codon-optimized for the heterologous expression in *E. coli* using ThermoFisher Scientific GeneOptimizer online tool (https://www.thermofisher.com/in/en/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis/geneoptimizer.html). The codon-optimized sequences were validated using Graphical Codon Usage Analyser 2.0 (http://gcua.schoedl.de) and Genscript Rare Codon Analysis (https://www.genscript.com/tools/rare-codon-analysis) online tools (Figure 2.1). Subsequently, the codon-optimized sequences were tagged with codon-optimized nucleotide sequences of poly-histidine-tag [His (H); octahistidine], HIV-Trans-Activator of Transcription [TAT (T); a short peptide sequence and also called a cell-penetrating peptide], and nuclear localization signal/sequence [NLS (N)]. All of these tags were flanked by restriction sites for easy removal of any individual component that might affect the protein functionality. This customized HTN-*GOI* and *GOI*-NTH inserts were then gene synthesized with *Nco*I (at 5' end) and *Xho*I (at 3' end) restriction sites and obtained from GenScript Biotech Corporation, Nanjing, China. The gene inserts were then cloned into the protein expression vector pET28a(+) (Novagen, Merck Millipore) using restriction endonucleases *Nco*I and *Xho*I (Fermentas). The resulting plasmids, pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH, were verified by restriction digestion analysis and then by DNA sequencing (Eurofins Genomics India Pvt. Ltd. Bengaluru, Karnataka, India) using standard T7 promoter (5′-

TAATACGACTCACTATAGGG-3′) and terminator (5′-GCTAGTTATTGCTCAGCGG-3′) primers.

2.1.2 Identification of optimal inducer concentration for maximal heterologous expression of recombinant fusion proteins in *E. coli*

To identify the optimal expression parameters, *E. coli* BL21(DE3) host cells were individually transformed with constructed recombinant pET28a(+) plasmids and grown overnight in Luria-Bertani broth (HiMedia) supplemented with 50 µg/mL of kanamycin (HiMedia) and used as an inoculum for the protein expression analysis. Secondary cultures were incubated at 37 °C with continuous shaking at 180 rpm (Orbital incubator shaker, IKON instruments, India) until the OD_{600} reached ~0.5. The expression of recombinant fusion proteins was then induced with different inducer concentrations (0.05, 0.1, 0.25 and 0.5 mM) of Isopropyl β-D-1-thiogalactopyranoside (IPTG) (HiMedia) and incubated for the next 2 hours at 37 °C. After the incubation, cells were harvested and stored at –80 °C until further use.

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)
GATA4	NM_002052.4	1329
MEF2C	NM_001193350.1	1422
TBX5	NM_000192.3	1557
ETS2	NM_005239.5	1410
MESP1	NM_018670.3	807
HAND2	NM_021973.2	654

2.1.3 Identification of optimal cell density for maximal heterologous expression of recombinant fusion proteins in *E. coli*

In order to identify the optimal cell density of *E. coli* BL21(DE3) cells at the time of induction, secondary cultures at different cell densities (~0.5, ~1.0 and ~1.5 OD_{600}) were induced for the recombinant fusion protein expression with the optimal concentration of IPTG for 2 hours at 37 °C with continuous shaking at 180 rpm. After the incubation, cells were harvested and stored at –80 °C until further use.

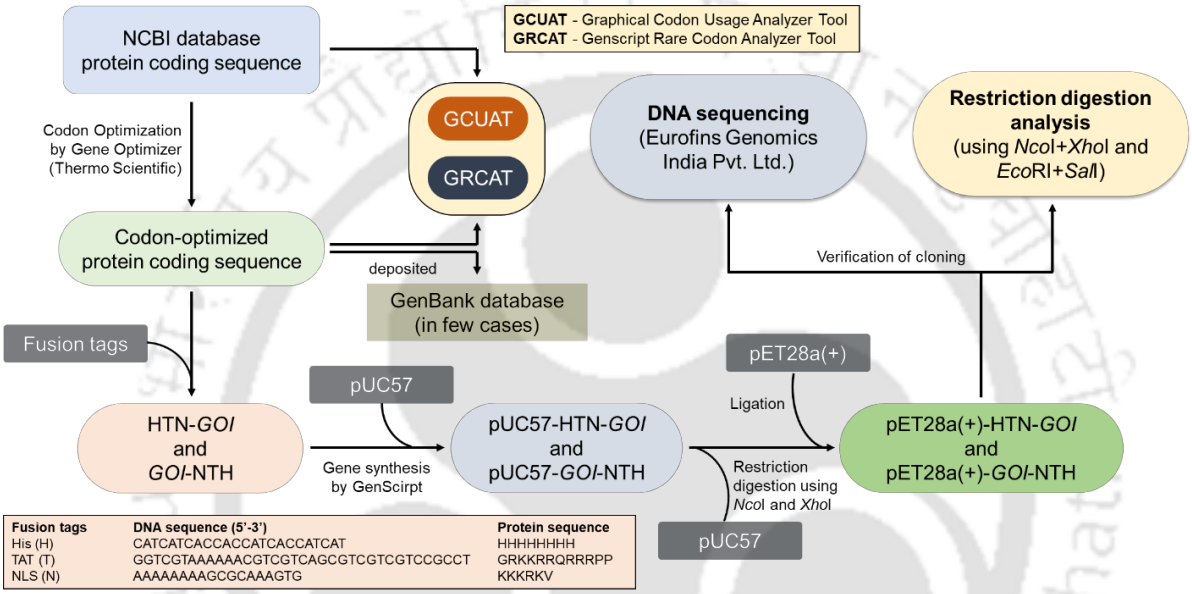


Figure 2.1 Schematic overview of the workflow of codon optimization and cloning methodology.

2.1.4 Identification of optimal induction temperature for maximal heterologous expression of recombinant fusion proteins in *E. coli*

For identifying the optimal induction temperature, the secondary cultures with optimal cell density were induced for the recombinant fusion protein expression at different temperatures (37 °C for 2 hours or 18 °C for 24 hours with continuous shaking at 180 rpm) with the optimal concentration of IPTG. After the incubation, cells were harvested and stored at –80 °C until further use.

2.1.5 Identification of optimal induction time for maximal heterologous expression of recombinant fusion proteins in *E. coli*

The induction time was optimized by inducing the expression of recombinant fusion protein at the optimal cell density with the optimal inducer concentration at 37 °C (for 8 hours) or 18 °C (for 48 hours) with continuous shaking at 180 rpm, and samples were collected every 2 hours (for 37 °C) or 12 hours (for 18 °C) for analysis. After the incubation, cells were harvested and stored at –80 °C until further use.

2.1.6 Identification of optimal salt concentration in resuspension buffer to obtain soluble recombinant fusion proteins from *E. coli*

The expression of recombinant fusion protein was induced in the secondary cultures with the identified optimal expression parameters. After induction, cells were harvested and resuspended in pre-chilled resuspension buffer (Table 2.2) consisting of sodium phosphate buffer (HiMedia), imidazole (HiMedia), and glycerol (HiMedia) with different NaCl (HiMedia) concentrations (0, 150, 300, 450, 600, 750, 900 and 1200 mM) and then lysed by ultrasonication using Vibracell™ VCX-130 cell disruptor (Sonics and Materials Inc., Newtown, CT, USA) on ice. Sonicated samples were centrifuged to obtain insoluble and soluble cell fractions and then stored at –20 °C until further use.

2.1.7 Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blotting

The harvested cell pellets were resuspended in pre-chilled resuspension buffer (Table 2.2) and further lysed by ultrasonication on ice. Protein concentrations were determined using the Bradford assay (Bradford reagent (Bio-Rad)) using bovine serum albumin (Bio-Rad) as a

standard (Bradford 1976) and measured with a multiplate reader (Multiskan GO, Thermo Scientific). Protein samples were mixed with 4X Laemmli buffer and 10% sodium dodecyl sulfate (SDS) to achieve final concentrations of 1X and 2%, respectively, and then resolved on 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel. Along with protein samples, a broad range of pre-stained Precision Plus Protein™ Dual Color (Bio-Rad) protein standards was also resolved as a molecular size control.

Table 2.2 Summary of the composition of resuspension buffers used.

Ingredients	Resuspension Buffer					
	GATA4	MEF2C	TBX5	ETS2	MESP1	HAND2
PB (mM)	20	20	20	20	20	20
NaCl (mM)	300	300	500	150	150	150
Imidazole (mM)	20	20	20	20	20	20
Glycerol (%)	20	20	20	20	20	20
pH (at RT)	8.0	7.6	7.2	7.5	8.5	8.0

PB, sodium phosphate buffer; RT, room temperature

For Coomassie staining, the resolved SDS-PAGE gel was stained with staining solution containing 50% (v/v) methanol (Merck Millipore), 10% (v/v) acetic acid (Merck Millipore) and 0.4% (w/v) Coomassie Brilliant Blue G-250 (Merck Millipore) in deionized water. The stained gel was then destained with 50% (v/v) methanol and 10% (v/v) acetic acid in deionized water and the image was recorded in the molecular imager (ChemiDoc™ XRS+) equipped with Image Lab™ software (Bio-Rad, California, USA).

For Western blotting, the proteins were resolved on 12% SDS-PAGE gel and the resolved samples were transferred onto the nitrocellulose membrane (Bio-Rad) using Pierce

Power Blotter XL System (Thermo Scientific™, Bremen, Germany). After confirming the transfer with Ponceau S staining (HiMedia), the membrane with transferred proteins was washed (destained) and blocked with 5% (w/v) fat-free milk in Tris-buffered saline (TBS) (HiMedia) with 0.1% (v/v) Tween-20 (TBST) (Invitrogen) for 1-2 hours at room temperature to reduce non-specific binding, followed by a wash with TBST for 10 minutes. The membrane was then incubated with primary antibody overnight at 4 °C and washed thrice with TBST for 5 minutes each, followed by respective horseradish peroxidase-conjugated secondary antibody incubation for 1 hour at room temperature. After washing thrice with TBST, immunoblot was developed in the presence of a chemiluminescence substrate (Bio-Rad), and the image was recorded in the molecular imager. 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 the Appendix 1.

2.1.8 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 six human genes (transcription factors) for their expression in *E. coli*

Codon optimization is one of the critical criteria for efficient heterologous expression of human proteins in *E. coli*. It involves synonymous codon substitution to improve the gene expression and its translational efficiency (Gustafsson et al. 2004; Burgess-Brown et al. 2008; Rosano and

Ceccarelli 2014). To determine whether codon optimization is required for the heterologous expression of selected six human transcription factors in *E. coli*, we first analyzed its full-length coding sequences using two different *in silico* tools Graphical Codon Usage Analyser and GenScript Rare Codon Analysis. From the analysis using the Graphical Codon Usage Analyser tool, nearly 7 (in the case of *HAND2* sequence) and >10 codons (in other gene sequences) with a relative adaptiveness value of ≤30% were identified to hamper their expression in *E. coli* (Figure 2.2; (*left*; *shown by magenta bars*)). The relative adaptiveness value calculated for each codon is the ratio of the frequency of the used codon to that of the most abundant codon for the same amino acid multiplied by 100. Similarly, ≥7% of the codons with codon usage frequency of ≤30 (shown by dotted threshold line) were identified to affect their heterologous expression using the GenScript Rare Codon Analysis tool (Figure 2.3 (*red*); Table 2.3). These codons are determined as low-frequency rare codons and were present in a small number (7-12%; ideal value: of ≤30%) in the selected six transcription factors. However, these rare codons may reduce translational efficiency of the gene of interest as reported earlier (Gustafsson et al. 2004; Burgess-Brown et al. 2008; Rosano and Ceccarelli 2014). As both the independent tools identified rare codons that could hamper the expression of these human transcription factors in *E. coli*, their full-length protein-coding sequences were codon-optimized using an *in silico* online tool, GeneOptimizer (Thermo Fisher Scientific).

Using Graphical Codon Usage Analyser and GenScript Rare Codon Analysis *in silico* tools, the codon-optimized gene sequences were evaluated for their heterologous expression in *E. coli* and compared with their non-optimized sequences. The validation results revealed that the identified rare codons (having relative adaptiveness value ≤30%) as well as other likely codons (having relative adaptiveness value between 30-100%) that could affect the expression

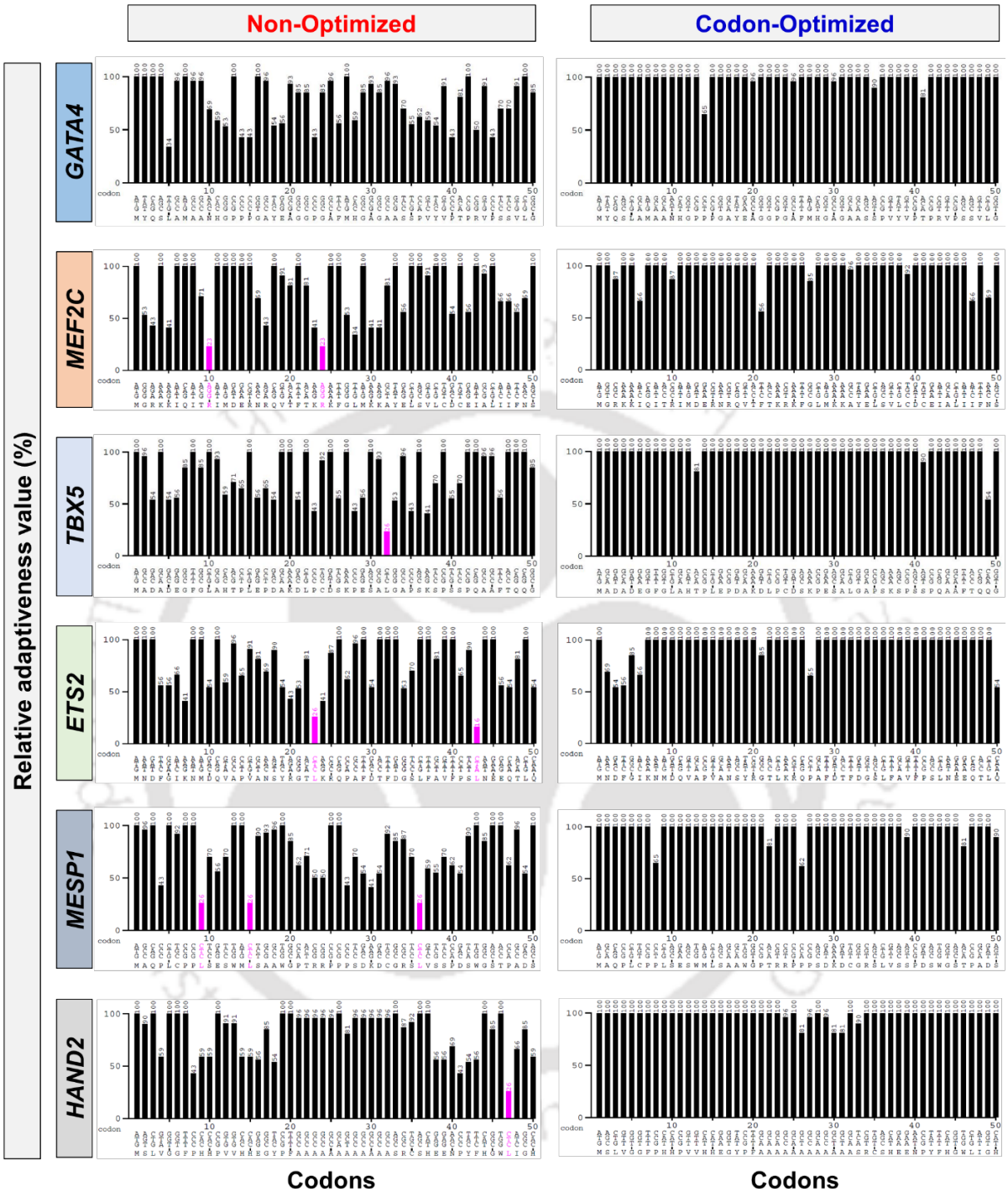


Figure 2.2 Validation and comparison of codon-optimized over non-optimized gene sequences for their heterologous expression in *E. coli* using Graphical codon usage analyzer tool. The bar graphs shown here are only for the first 50 codons of non-optimized (*left*) and

codon-optimized (*right*) sequences of each gene as the representation. This analysis depicts codon quality as a measure of relative adaptiveness value. All the codons were analyzed by the relative adaptiveness value; therefore, the value >30% means that those codons should be suitable for expression in *E. coli*.

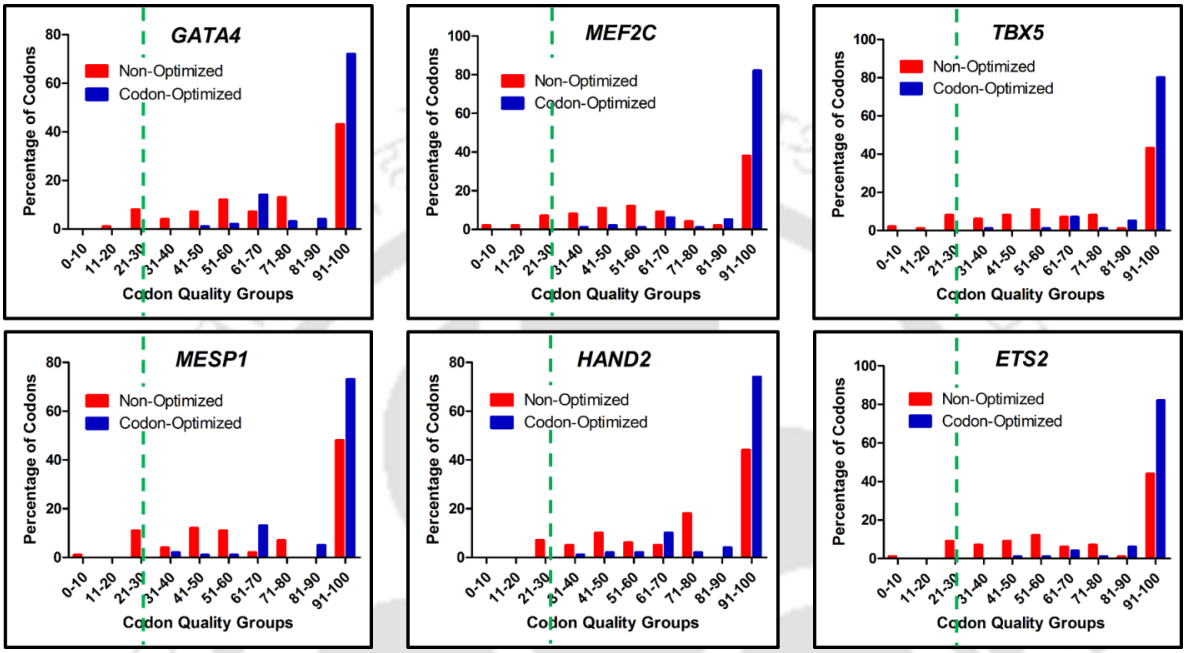


Figure 2.3 Validation and comparison of codon-optimized over non-optimized gene sequences for their heterologous expression in *E. coli* using GenScript rare-codon analysis tool. The bar graph shows the percentage distribution of codons in computed codon quality groups for the non-optimized (*red*) and codon-optimized (*blue*) gene sequences. Codons with the highest usage frequency for a given amino acid in the desired expression organism are assigned a value of 100. Codons with values ≤30 are likely to hamper the expression in the desired expression organism.

of these human transcription factors in *E. coli* were substituted with the most preferred synonymous codons to improve their heterologous expressions (Figure 2.2; (*right*)) and Figure 2.3; (*blue*)). Additionally, codon optimization resulted in an increased codon adaptation index (CAI) value to ≥0.88 for the codon-optimized sequences from <0.7 of their non-optimized

sequences (Table 2.3). CAI is the relative adaptiveness of the codon usage of a protein-coding gene towards the codon usage of highly expressed (reference set) genes. The CAI value in the range of 0.8–1 is considered optimal for expressing the desired gene in the heterologous system, whereas CAI <0.8 hampers its expression. These results confirmed that the codon-optimized gene sequences of these six transcription factors were devoid of rare codons that could affect their expression, and therefore, favor their heterologous expression in *E. coli*.

2.2.2 Cloning of codon-optimized human genes with fusion tags into the protein expression vector

The codon-optimized gene sequences of the selected six human transcription factors were fused with tags (fusion tags were also codon-optimized for the expression in *E. coli*), either before the start codon (to generate HTN-*GOI*) or at the end of the coding sequence of the gene of interest (to generate *GOI*-NTH) as shown in Figure 2.4. The reason to have fusion tags at either the 5′ or the 3′ end is that its position can influence expression level, solubility and stability of human proteins expressed in *E. coli* (Braun et al. 2002; Bosnali and Edenhofer 2008). The three fusion tags used are His (H) tag for affinity chromatography, TAT (T) to enable cell penetration, and NLS (N) to facilitate subnuclear localization. A similar approach was employed in earlier studies to enable efficient cell and nuclear delivery of transcription factors in mammalian cells (Bosnali and Edenhofer 2008; Peitz et al. 2014; Münst et al. 2016).

To express these six human transcription factors in *E. coli*, HTN-*GOI* and *GOI*-NTH inserts were cloned under the control of a strong inducible T7 promoter of the protein expression vector, pET28a(+). These gene inserts were cloned between *Nco*I and *Xho*I sites using restriction endonucleases, *Nco*I and *Xho*I, to remove the poly-histidine tag (present before the start codon) and thrombin cleavage site sequence present in the pET28a(+) vector.

Table 2.3 Parametric GenScript rare codon analysis summary for non-optimized and codon-optimized protein-coding gene sequences.

Genes	Parameters	Non-Optimized	Codon-Optimized	Ideal range
	Codon Adaptation Index (CAI)	0.68	0.89	0.8-1.0
GATA4	GC content	69.41%	55.95%	30-70%
	Codon Frequency Distribution (CFD)	9%	0%	<30%
	Codon Adaptation Index (CAI)	0.60	0.92	0.8-1.0
MEF2C	GC content	50.77%	50.37%	30-70%
	Codon Frequency Distribution (CFD)	11%	0%	<30%
	Codon Adaptation Index (CAI)	0.63	0.91	0.8-1.0
TBX5	GC content	55.39%	49.18%	30-70%
	Codon Frequency Distribution (CFD)	11%	0%	<30%
	Codon Adaptation Index (CAI)	0.65	0.93	0.8-1.0
ETS2	GC content	52.79%	47.35%	30-70%
	Codon Frequency Distribution (CFD)	10%	0%	<30%
	Codon Adaptation Index (CAI)	0.64	0.88	0.8-1.0
MESP1	GC content	75.30%	59.64%	30-70%
	Codon Frequency Distribution (CFD)	12%	0%	<30%
	Codon Adaptation Index (CAI)	0.69	0.89	0.8-1.0
HAND2	GC content	66.52%	48.90%	30-70%
	Codon Frequency Distribution (CFD)	7%	0%	<30%

The resulting plasmids (pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH) were primarily verified by restriction digestion analysis (Figure 2.5). The empty vector, pET28a(+) only, was also included in the analysis as a control to confirm the absence of gene of interest (*data not*

shown). Further, these gene constructs, HTN-*GOI* and *GOI*-NTH, were confirmed via DNA sequencing from both the ends using T7 promoter and T7 terminator primers. The sequencing results of the cloned inserts confirmed the fidelity of the gene sequence and fusion tags.

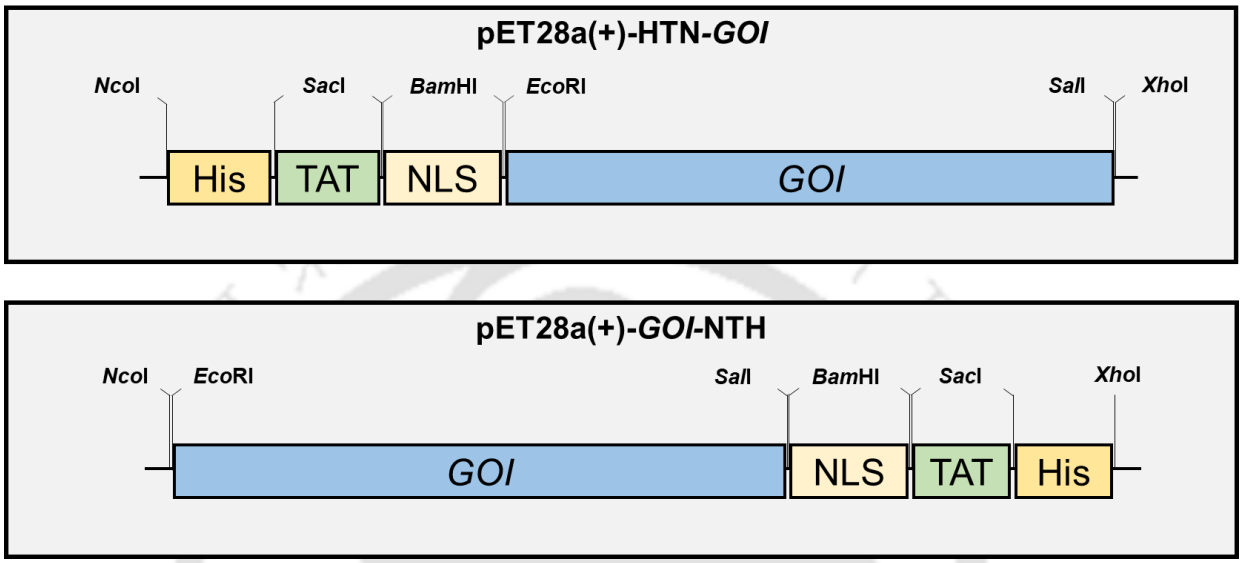


Figure 2.4 Schematic diagram of HTN-*GOI* and *GOI*-NTH inserts (not drawn to scale). The *GOI* was fused either at the 5′ or at the 3′ end with nucleotide sequences of NLS (N) and TAT (T) to facilitate nuclear translocation and cell penetration in mammalian cells, respectively, followed by His (H) tag for affinity chromatography-based purification.

2.2.3 Identification of optimal induction conditions for the heterologous expression of six human recombinant fusion proteins

One of the most critical factors in the production of recombinant proteins is the selection of an appropriate host expression system. In this study, the widely used bacterium *E. coli* was used as an expression host. This organism has a high transformation efficiency, fast growth rate, well-understood genetics, and cost-effective protein production (Burgess-Brown et al. 2008).

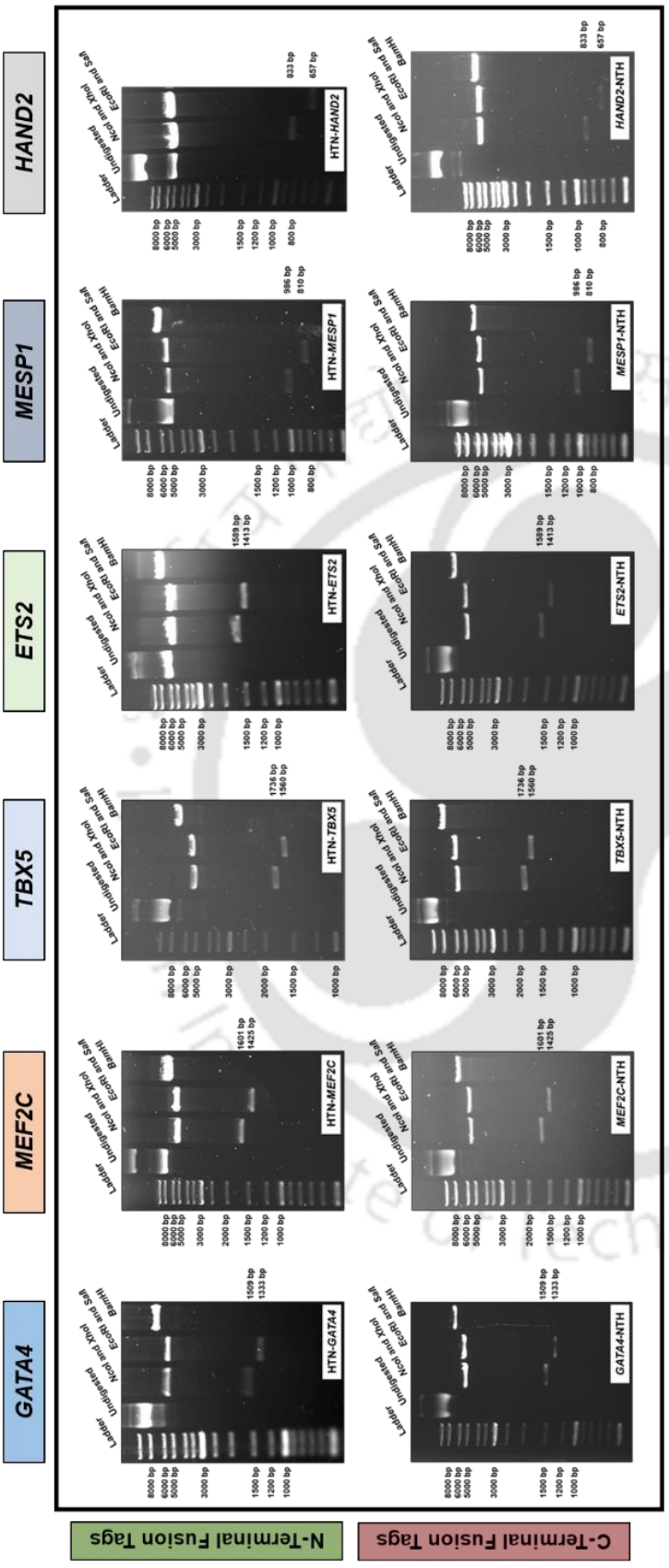


Figure 2.5 Confirmation of cloning of the gene inserts into the protein expression vector. Restriction digestion analysis was performed to confirm the successful cloning of HTN-*GOI* and *GOI*-NTH inserts into the protein expression vector. The synthetic gene inserts were cloned in the protein expression 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)

This expression host is commonly used for human proteins for which post-translational modifications are not essential for their bioactivity (Lili et al. 2006; Stefan et al. 2018; Bhat et al. 2020). Several studies reported that even in the absence of post-translational modifications, the human transcription factors in the form of recombinant proteins had retained their biological activity (Bosnali and Edenhofer 2008; Thier et al. 2010; Thier et al. 2012; Peitz et al. 2014; Münst et al. 2016; Dey et al. 2021b; Narayan et al. 2021b; Dey et al. 2022). Specifically, the commonly used *E. coli* BL21(DE3) strain was used, which is devoid of *lon* and *OmpT* proteases, allowing a high level of heterologous protein expression without any degradation.

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

Recombinant Proteins	Optimal expression parameters			
	Inducer concentration (mM)	Induction cell density (OD_{600})	Induction temperature (°C)	Induction time (hours)
rhHTN-GATA4 and rhGATA4-NTH	0.25	~0.5	37	2
rhMEF2C-NTH	0.05	~0.5	37	2
			18	24
rhTBX5-NTH	0.05	~0.5	37	4
rhHTN-ETS2	0.1	~0.5	37	2
rhHTN-MESP1	0.05	~0.5	18	24
rhHTN-HAND2	0.05	~0.5	37	2
			18	24

Note: Minimum of 2-3 different cultures from individual colonies were analyzed using SDS-PAGE and Western blotting.

Numerous studies have demonstrated the importance of identifying the optimal expression parameters for obtaining a high yield of biologically active recombinant proteins in a soluble form (Galloway et al. 2003; Ou et al. 2004; Sørensen and Mortensen 2005; Rabhi-Essafi et al. 2007; San-Miguel et al. 2013; García-Fraga et al. 2015). Based on these studies, we have first investigated the effect of inducer concentration (IPTG) on the expression of these six recombinant fusion proteins (rhHTN-GOI or rhGOI-NTH) in *E. coli*. Interestingly, we observed the maximal expression in most of these fusion proteins with the minimal concentration (0.05 mM) of IPTG (Table 2.4). Notably, in the case of rhGATA4 and rhETS2, the maximal expression was observed with slightly higher concentrations of IPTG compared to the rest of the factors. However, the results further revealed that the expression level of each factor varied with different inducer concentrations (for the induction of T7 RNA polymerase-mediated expression of the protein), and it reached the maximum when induced with the appropriate IPTG concentration (Table 2.4).

Next, the effect of cell density at the time of induction on the expression of these recombinant human fusion proteins were investigated. Various studies have reported that the bacterial system overexpresses heterologous proteins when induced at the log phase, i.e., 0.3-0.8 OD_{600} (García-Fraga et al. 2015). In contrast, Gavidia and colleagues reported that the induction at an early log phase (~0.1 OD_{600}) resulted in ~ three-fold higher soluble expression, unlike inducing at the log phase, ~0.6 OD_{600} (San-Miguel et al. 2013). Apart from these, few reports state that high-density induction enhances the solubility of heterologous mammalian proteins under controlled conditions (Galloway et al. 2003; Ou et al. 2004). Hence, identifying the optimal cell density at the time of induction is crucial for the maximal expression of recombinant proteins. In this study, from the SDS-PAGE analysis, the maximal expression of

these six transcription factors was observed when induced during the log phase, ~0.5 OD_{600} (Table 2.4), compared to the late log phase (~1 or ~1.5 OD_{600}). Interestingly, this is the only induction parameter where the optimal value is the same for all the six different recombinant proteins. However, the expression pattern with respect to different cell densities varies significantly between these proteins. Overall, at the point of induction, cell density plays a significant role in expressing these exogenous proteins in *E. coli*.

Recombinant proteins expressed in the bacterial system either end up in a properly folded soluble form or in an insoluble form consisting of inclusion bodies. Purifying proteins from inclusion bodies involves the usage of strong detergents to solubilize them and then requires refolding, which is a complex process (Baneyx and Mujacic 2004). Therefore, obtaining maximal soluble expression of these recombinant fusion proteins were given utmost importance. Accordingly, we investigated the influence on the solubility of these fusion proteins (rhHTN-GOI and rhGOI-NTH) using SDS-PAGE and Western blotting based on the terminal at which the tags were fused. To analyze the solubility, we first examined the amount of fusion proteins induced at 37 °C in the total cell lysate, and insoluble pellet, and soluble supernatant fractions. Interestingly, the GATA4 fusion proteins were found in both the pellet and supernatant fractions of the cell lysates (Figure 2.6A). Other studies reported that the reduction in temperature during induction enhanced the solubility of the protein of interest (Sørensen and Mortensen 2005; Huang et al. 2012; San-Miguel et al. 2013). Based on these studies and to improve solubility, the effect of temperature on the rhHTN-GATA4 and rhGATA4-NTH protein solubility was examined. In both cases, the reduction in the induction temperature to 18 °C did not enhance the solubility; instead, a decrease in the expression of the GATA4 fusion proteins (Figure 2.6A) was observed. Therefore, the optimal temperature

to obtain the maximal soluble expression of the recombinant GATA4 fusion proteins was found to be 37 °C.

In the case of MEF2C fusion proteins, the presence of fusion tags at the N-terminal of the rhMEF2C protein hindered its solubility, in contrast to its C-terminal counterpart (Figure 2.6B). However, a similar overall expression profile of these fusion proteins was observed with all (rhHTN-MEF2C) or majority (rhMEF2C-NTH) of the protein molecules lying in the insoluble pellet fraction. We assume that this might be the effect of misfolding or aggregation (in the case of folded ones in inclusion bodies) of the target fusion proteins. We then hypothesize that reducing the induction temperature will assist in the enhanced soluble expression of rhMEF2C fusion proteins. But reducing the induction temperature to 18 °C did not help in enhancing the soluble expression of rhHTN-MEF2C, contrastingly, its expression was completely hindered (Figure 2.6B; *left*). Interestingly, enhanced soluble expression was observed in the case of the rhMEF2C-NTH fusion protein on the reduction of induction temperature to 18 °C (Figure 2.6B; *right*). However, the overall expression of rhMEF2C-NTH fusion protein declined at 18 °C in contrast to 37 °C. Taken together, these results conclude that the position of fusion tags and induction temperature significantly impact the overall and soluble expression of rhMEF2C fusion proteins. Inferring from the above results, rhMEF2C-NTH protein-induced either at 37 or 18 °C were used for further experiments.

Next, we examined the amount of rhTBX5 fusion protein induced at 37 °C in the total cell lysate, and the insoluble pellet, and soluble supernatant fractions. As shown in Figure 2.6C (*left*), the overall expression of N-terminal tagged rhTBX5 fusion protein in the total cell lysate fraction was lower than that of C-terminal tagged rhTBX5; however, we did not observe any expression in the supernatant fraction of N-terminal tagged fusion proteins, unlike C-terminal

tagged fusion proteins (low amount). This finding was very likely to be due to the misfolding of recombinant proteins or the aggregation of folded proteins in inclusion bodies. We then assessed whether decreasing the induction temperature promoted the solubility of the rhHTN-TBX5 protein. Even after decreasing the induction temperature to 18 °C, we did not observe any soluble expression of the rhHTN-TBX5 protein; instead, its expression completely declined (Figure 2.6C; *left*). Interestingly, we observed a significant shift of rhTBX5-NTH protein molecules from insoluble to soluble cell fractions with induction at 18 °C (Figure 2.6C, *right*). However, the overall expression of rhTBX5-NTH declined when expressed at 18 °C compared to 37 °C. Moreover, we observed truncated protein fragments of rhTBX5-NTH in both induction temperatures, in contrast to the results for rhHTN-TBX5. Our results further indicated that the tagging pattern and induction temperature influenced the soluble expression of the heterologous rhTBX5 fusion protein. From the immunoblotting analysis, the intensity of the truncated protein fragments of rhTBX5-NTH induced at 18 °C was higher than or comparable to that of the full-length protein (Figure 2.6C; *bottom*). On the basis of these results, we proceeded with using rhTBX5-NTH protein induced at 37 °C for further experiments.

In the case of rhETS2 fusion proteins, the majority of protein molecules were found in the insoluble cell fraction when induced at 37 °C. We then compared the soluble expression of rhETS2 fusion proteins when induced at 18 °C with that of 37 °C. Although in both cases, expression at 18 °C was lower than that of 37 °C (Figure 2.6D), reduction in temperature did help in ETS2 fusion protein solubility (Figure 2.6D). From the Western blot analysis (Figure 2.6D; *right bottom*), it is evident that there are truncated fragments of rhETS2-NTH protein irrespective of induction temperatures, which might compromise the concentration of the full-

length protein as well as its purity in the later stages. However, in the case of rhHTN-ETS2, no visible truncations were detected in the Western blotting of the soluble fraction at 37 °C (Figure 2.6D; *left bottom*). These results indicate that the terminal at which the tags were fused played an essential role in rhETS2 fusion protein expression in terms of quantity and quality. Hence, we chose N-terminal fused rhETS2 (rhHTN-ETS2) fusion protein induced at 37 °C for further experiments.

Next, the solubility of rhMESP1 fusion proteins (rhHTN-MESP1 and rhMESP1-NTH) when induced at 37 °C was analyzed. The SDS-PAGE and Western blotting analysis revealed that most of these rhMESP1 fusion protein molecules (induced at 37 °C) were in the pellet or insoluble cell fraction, independent of the terminal in which the fusion tags were fused (Figure 2.6E). Notably, several other studies have demonstrated that reducing the induction temperature improved recombinant protein folding and solubility by preventing its aggregation into the insoluble/pellet fraction (Vasina and Baneyx 1997; Sørensen and Mortensen 2005; Huang et al. 2012; San-Miguel et al. 2013). Therefore, the solubility of these recombinant fusion proteins was analyzed when induced at 18 °C. Consistent with the earlier reported studies (Vasina and Baneyx 1997; Sørensen and Mortensen 2005; Huang et al. 2012; San-Miguel et al. 2013), we observed a significant shift in the recombinant protein solubility when the induction temperature was reduced to 18 °C for rhHTN-MESP1 but the same was not observed when induced at 37 °C (Figure 2.6E; *left*). Approximately 50% of the expressed rhHTN-MESP1 recombinant protein was found in the supernatant or soluble cell fraction when induced at 18 °C (Figure 2.6E; *left*). However, induction at 18 °C did not favor any soluble expression of rhMESP1-NTH; instead, its expression was inhibited (Figure 2.6E; *right*). These results indicate that the induction temperature and position of fusion tags play a crucial role in

achieving the soluble expression of the protein of interest. Thus, we selected rhHTN-MESP1 recombinant protein induced at 18 °C for further expression and purification analysis based on this soluble expression analysis.

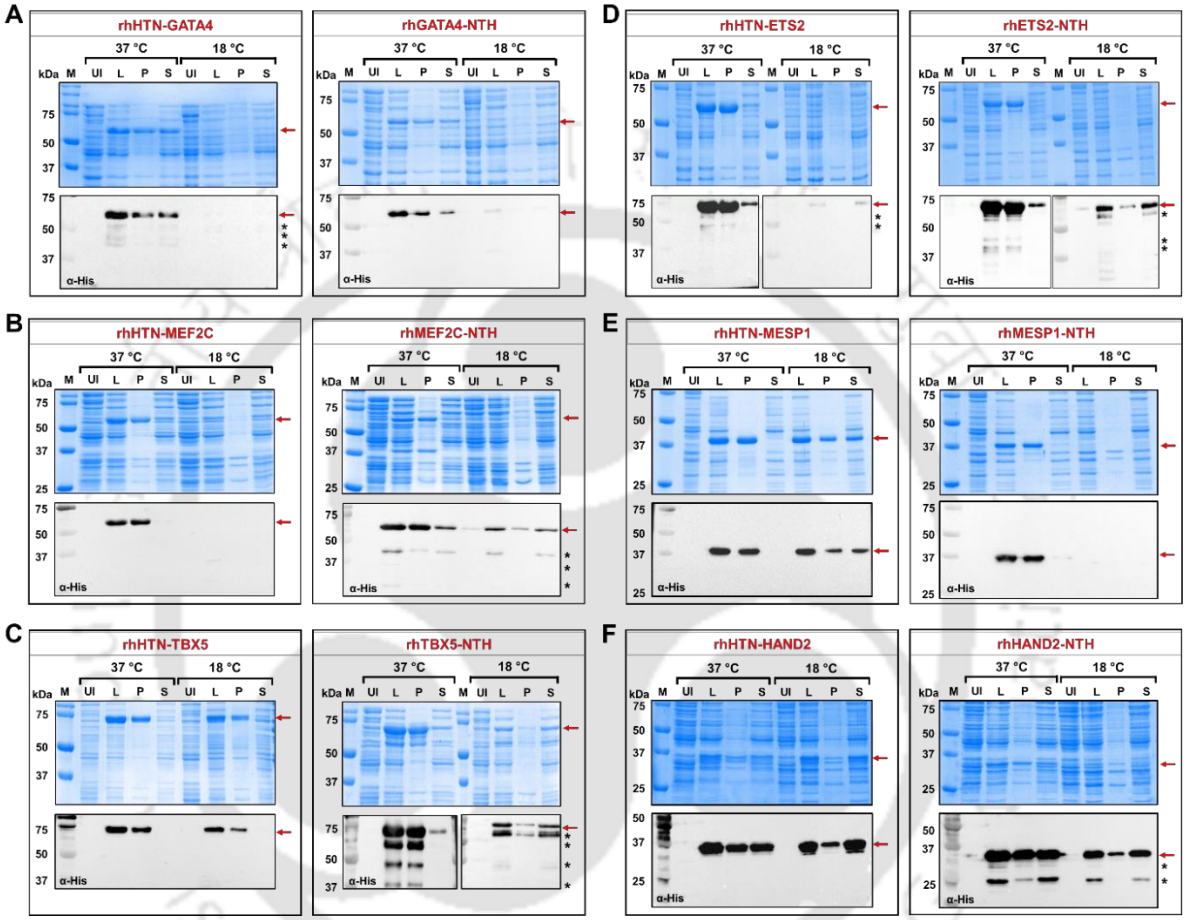


Figure 2.6 Identification of optimal induction temperature to obtain the recombinant human fusion proteins in soluble form. *E. coli* BL21(DE3) competent cells were transformed with pET28a(+) vectors harboring the fusion gene inserts and the expression of recombinant proteins were induced at optimal cell density with an optimal concentration of IPTG at two different temperatures: 37 °C (IPTG induction for 2 h) and 18 °C (IPTG induction for 24 h). Harvested cells were lysed to obtain the total cell lysate (L) fraction and were then centrifuged to obtain a pellet/insoluble (P) and a soluble (S) cell fraction. These fractions (L, P, and S) were run on SDS-PAGE

with normalized protein loading concentration for L fractions. Equal volume corresponding to the respective L fraction was loaded for the P and S fractions. Resolved gels were then either stained with Coomassie Brilliant Blue (*top*) or performed Western blotting with His antibody (*bottom*). (**A**) rhGATA4 fusion proteins (n=3), (**B**) rhMEF2C fusion proteins (n=5), (**C**) rhTBX5 fusion proteins (n=7), (**D**) rhETS2 fusion proteins (n=3), (**E**) rhMESP1 fusion proteins (n=3), (**F**) rhHAND2 fusion proteins (n=3). M, protein marker; UI, uninduced (total cell lysate); L, cell lysate; P, pellet (insoluble) fraction; S, supernatant (soluble) fraction.

Subsequently, we examined the soluble expression of rhHAND2 fusion proteins induced at 37 °C, using SDS-PAGE and Western blotting. The results revealed that in both the cases (rhHTN-HAND2 and rhHAND2-NTH), nearly half of the overall expressed rhHAND2 protein molecules were found in pellet/insoluble cell fractions and rest in the supernatant/soluble cell fractions (Figure 2.6F). Therefore, we further investigated whether the reduction in temperature enhances the solubility of the rhHAND2 fusion proteins. In both cases, induction at 18 °C considerably improved the solubility of the rhHAND2 fusion proteins (Figure 2.6F). However, a decline in the overall expression only in rhHAND2-NTH fusion protein was observed when induced at 18 °C compared to 37 °C. The immunoblotting analysis further represents the truncated protein fragments of rhHAND2-NTH protein in both the induction temperatures, unlike rhHTN-HAND2 (Figure 2.6F; *right bottom*). Importantly, these truncations might compromise the full-length rhHAND2 fusion protein quality and purity at the final stages. Thus, based on this analysis, fusion tags at the N-terminal end of rhHAND2 (rhHTN-HAND2) induced at 37 or 18 °C, were selected for further experiments.

Among these recombinant fusion proteins, soluble expression analysis using Western blotting of rhTBX5 (both N- and C-terminal fused), rhETS2-NTH and rhHAND2-NTH showed compromised expression quality as multiple truncations were observed in these cases.

The reason for these truncated fragments could be due to various possible reasons, such as (i) intragenic sequences that mimic *E. coli* ribosomal entry sites present within the protein-coding sequence (Maertens et al. 2010), (ii) proteolysis of some protein molecules at specific sensitive sites during expression (Ryan and Henehan 2013), (iii) protein cleavage at Asp-Pro bonds because of overheating of protein samples (Kurien and Scofield 2012). Out of these possible reasons, the last one is least likely in this study. For example, the protein samples of both rhHTN-HAND2 and rhHAND2-NTH were treated similarly, and visible truncations were observed only in the case of rhHAND2-NTH. In contrast, no such truncations were observed with its counterpart. Moreover, these observations signified the importance of induction temperature and the terminal at which the fusion tags were coupled in the production of quality recombinant proteins. Earlier study also reported similar observations with different transcription factors (Bosnali and Edenhofer 2008) .

We next sought to examine the effect of postinduction incubation time on the maximal soluble expression of these recombinant fusion proteins. Various studies have also reported the importance of postinduction incubation time on the recombinant protein yield (Wong et al. 1998; Rosano and Ceccarelli 2014; Malik et al. 2016). For this analysis, BL21(DE3) competent *E. coli* cells transformed with respective plasmids were induced at 37 °C for 8 hours or 18 °C for 48 hours, and samples were collected at different time intervals. As summarized in Table 2.4, the expression level of these factors other than rhTBX5 reached the maximum within 2 hours of induction at 37 °C and remained unchanged up to 8 hours. In the case of rhTBX5, the maximum expression was observed after 4 hours of induction at 37 °C and remained unchanged up to 8 hours. This result signified that the recombinant fusion proteins were stable in *E. coli*. However, in the case of rhMEF2C-NTH, rhHTN-MESP1 and rhHTN-HAND2

proteins, the optimal postinduction incubation time was identified as 12 hours when induced at 18 °C. Our results, thus, specify that the postinduction incubation time significantly contributed towards the maximal expression of recombinant proteins. The overall identified optimal conditions for the maximal expression of these six recombinant fusion proteins are summarized in Table 2.4.

Interestingly, in the case of rhHTN-MESP1 protein, we observed that higher salt concentration (>300 mM) resulted in the precipitation of this fusion protein, whereas low salt concentration (≤300mM) resulted in an increased amount of protein in the soluble cell fraction (Figure 2.7). This result indicated that the NaCl concentration should be ≤300 mM in the resuspension buffer to prevent salt-dependent aggregation of rhMESP1 fusion protein. A similar salt-dependent aggregation effect was also observed for ferritin protein expressed in *E. coli* (Sun et al. 2016), indicating an ion-sensitive protein. This observation implies MESP1 to be sensitive to the ionic strength of the environment. Thus, this highlighted the importance of identification of optimal salt concentration to obtain the maximal rhMESP1 fusion protein in soluble form to achieve native purification.

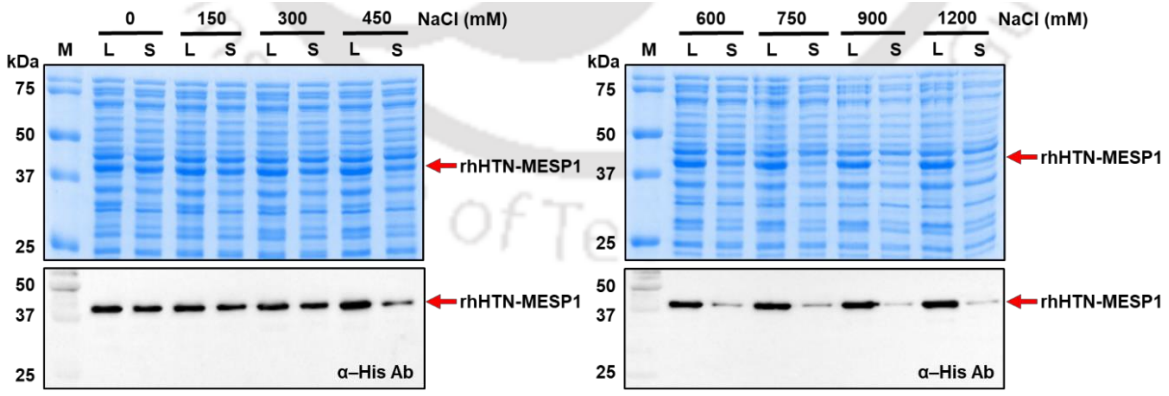


Figure 2.7 Effect of salt concentration on the solubility of rhHTN-MESP1 protein in resuspension buffer. Transformed BL21(DE3) competent *E. coli* cells with pET28a(+) vector

harboring *MESP1* fusion gene were induced under optimal conditions with continuous shaking to determine the effect of salt concentration in resuspension buffer on rhHTN-MESP1 protein solubility. Harvested cells were resuspended in resuspension buffer with varying NaCl concentration and then lysed using sonication to obtain total cell lysate (L) fraction. The lysates were then centrifuged to obtain a soluble supernatant (S) fraction. These fractions (L and S) were run on SDS-PAGE with a normalized protein loading concentration of 10 µg/well for L fractions. Equal volume corresponding to the respective L fraction was loaded for S fractions. Resolved gels were then either stained with Coomassie Brilliant Blue (*top*) or performed Western blotting with His antibody (*bottom*). M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; (n=3)

Taken together, our results demonstrated the influence of the position of fusion tags and expression conditions in the heterologous expression of recombinant proteins. Thus, based on these results, we have identified the appropriate genetic constructs suitable for the soluble expression of these fusion proteins under the identified optimal conditions.

2.3 Conclusions

In this chapter, we have shown the importance of codon optimization and identification of optimal expression parameters for the successful heterologous expression of these six transcription factors in the form of recombinant fusion proteins with enhanced solubility. In addition, we have demonstrated that the terminal at which the tags are fused plays a crucial role in protein expression in terms of quality, quantity and solubility. As proof, we reported that the fusion tags at the C-terminal end of rhTBX5, rhETS2 and rhHAND2 proteins compromise the quality when compared to their N-terminal fused counterparts. Also, we observed that rhMEF2C-NTH, rhTBX5-NTH and rhHTN-MESP1 proteins have significantly higher soluble expression compared to their respective counterparts. Moreover, when inducing

the rhMEF2C-NTH, rhETS2-NTH, rhHTN-MESP1 and rhHAND2 fusion proteins at low temperature (18 °C), we observed a shift from aggregated insoluble to a soluble form. In fact, induction at 18 °C significantly enhanced the solubility (more protein molecules in soluble fractions than pellet fractions) of rhETS2-NTH and rhHAND2 fusion proteins. However, inducing rhHTN-MEF2C, rhHTN-TBX5, rhHTN-ETS2 and rhMESP1-NTH proteins at low temperature (18 °C) does not favor any soluble expression; instead, their overall expression declined. Similarly, in the case of rhGATA4 fusion proteins, lowering the induction temperature hampers the expression significantly. Thus, low-temperature induction prevents aggregation and enhances solubility based on the nature of the protein. 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, biochemical and biophysical analysis of recombinant proteins

Brief Overview of the Chapter

Recombinant proteins purified under native conditions often produce bioactive molecules with native-like folding conformation. Therefore, in this chapter, we report a one-step homogeneous purification of the six selected human transcription factors in the form of recombinant proteins under native conditions from *E. coli*. To achieve homogeneous purification, we have performed immobilized metal ion affinity chromatography from the clarified total cell lysate (soluble cell fraction). In the case of the GATA4 fusion protein, we initially purified it using partial denaturation-based affinity chromatography. Although we have purified it with maximum purity, we could not refold it to its native-like folding conformation even after performing various refolding methods. Therefore, we expressed the GATA4 fusion protein in large volumes and then purified it under native conditions using affinity chromatography. Notably, this purified recombinant fusion protein retained its native-like folding conformation. Similarly, we purified all the other transcription factors only under native conditions. We further confirmed their identity using Western blotting with protein-specific antibodies. Notably, we have demonstrated that these purified recombinant fusion proteins had retained their secondary structure post-purification using far UV circular dichroism spectroscopy. We also found that the salt concentration in buffers plays an important role in the purification process. For instance, in the case of the MESP1 fusion protein, purification only in the absence of salt retained its secondary structure. This chapter concludes that purification under native conditions and an appropriate salt concentration (based on the nature of protein) resulted in retaining the native-like folding conformation of these recombinant fusion proteins.

3.1 Materials and Methods

3.1.1 Immobilized Metal Ion Affinity Chromatography

Immobilized metal ion affinity chromatography (IMAC) was performed to obtain purified recombinant proteins. Briefly, protein expression was induced in large culture volumes (1.2 L) with the identified optimal expression conditions (Table 2.4; Chapter 2). The harvested cell pellets were resuspended in 40 mL of 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™ VCX-130 cell disruptor Sonics and Materials Inc., Newtown, CT, USA on ice for 30-60 minutes. The cell suspension was then clarified by centrifugation at 11,000 rpm, 4 °C for 30 minutes.

3.1.1.1 Purification under native conditions

The clarified soluble fraction was incubated with nickel-nitrilotriacetic acid (Ni-NTA) resin at 4 °C for 8-12 hours with continuous shaking. After incubation, the clarified fraction with Ni-NTA was loaded onto the purification column (Bio-Rad), equilibrated with resuspension buffer, and subsequently, the flow-through was drained out. Consequently, the column was washed with wash buffer 1 (50-100 mL) with incubation at 4 °C for 5 minutes. Subsequently, the column was washed with wash buffer 2 and 3 sequentially. After the wash buffers were drained out completely, the bound proteins were eluted with elution buffer (10-20 mL). The eluted proteins were collected and stored at 4 °C. Samples were also collected at different stages for analysis. Then the samples were run on SDS-PAGE and confirmed with Western blotting. The respective buffer compositions are listed in Table 3.1.

Table 3.1 List of purification buffers used for each transcription factor and their composition.

Protein	Ingredients	Resuspension buffer	Wash buffers 1	Wash buffers 2	Wash buffers 3	Elution buffer
GATA4	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)	8.0	8.0	8.0	8.0	8.0
MEF2C	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)	7.6	7.6	7.6	7.6	7.6
TBX5	PB (mM)	20	20	20	20	20
	NaCl (mM)	500	500	500	500	500
	Imidazole (mM)	20	50	100	150	500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	7.2	7.2	7.2	7.2	7.2
ETS2	PB (mM)	20	20	20	20	20
	NaCl (mM)	500	500	500	500	500
	Imidazole (mM)	20	50	100	150	500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	7.5	7.5	7.5	7.5	7.5
MESP1	PB (mM)	20	20	20	20	20
	NaCl (mM)	0/150	0/150	0/150	0/150	0/150
	Imidazole (mM)	20	50	100	150	500
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	8.5	8.5	8.5	8.5	8.5
HAND2	PB (mM)	20	20	20	20	20
	NaCl (mM)	150	150	150	150	150
	Imidazole (mM)	20	50	100	150	300
	Glycerol (%)	20	-	-	-	20
	pH (at RT)	8.0	8.0	8.0	8.0	8.0

PB, sodium phosphate buffer; RT, room temperature

3.1.1.2 Purification under mild denaturation conditions

The clarified soluble fraction was diluted accordingly with denaturation buffer containing 8 M urea to the final concentration of 0 or 2 or 4 M urea and incubated under shaking conditions overnight at 4 °C prior to loading on to the purification column. This soluble fraction was incubated with Ni-NTA resin at 4 °C for 8-12 hours with continuous shaking. After incubation, the clarified fraction with Ni-NTA was loaded onto the purification column (Bio-Rad). Remaining procedures were carried out similar to native purification; however, the buffer compositions were different from the native purification (Table 3.2). The purification samples were then resolved and analyzed using 12% SDS-PAGE and Western blotting.

Table 3.2 Summary of buffers used for purification under mild denaturation conditions and their composition.

Ingredients	Resuspension buffer	Denaturation buffer	Wash buffers			Elution buffer
			1	2	3	
PB (mM)	20	160	80	80	80	80
NaCl (mM)	300	500	500	500	500	500
Imidazole (mM)	20	20	50	100	150	500
Glycerol (%)	20	20	20	20	20	20
Urea (M)	-	8	0/2/4	0/2/4	0/2/4	0/2/4
pH (at RT)	8.0	8.0	8.0	8.0	8.0	8.0

PB, sodium phosphate buffer; RT, room temperature

The purified proteins (under both the conditions) were further dialyzed using snake-skin dialysis tubing (HiMedia) or buffer exchanged using PD10 columns (GE Healthcare) against sterile 20 mM sodium phosphate buffer (HiMedia) with (glycerol buffer) or without 20% glycerol or against Dulbecco's modified eagle medium (DMEM) depending on the

experimental requirement. The dialyzed or buffer exchanged purified proteins were stored at –80˚C until further use.

3.1.2 Mass spectrometry

3.1.2.1 In-gel digestion

The purified recombinant fusion protein was run on SDS-PAGE gel and stained with staining solution containing 50% (v/v) methanol, 10% (v/v) acetic acid and 0.4% (w/v) Coomassie Brilliant Blue G-250 in deionized water. The desired band was excised and destained with 40 mM ammonium bicarbonate (HiMedia) in a 40% (v/v) acetonitrile solution (Merck Millipore). The destained gel was then treated with reduction solution [5 mM dithiothreitol (Sigma-Aldrich) in 40 mM ammonium bicarbonate] for 30 minutes at 60 ºC, followed by alkylation solution [20 mM iodoacetamide (Sigma-Aldrich) in 40 mM ammonium bicarbonate] for 10 minutes at room temperature (dark). The excised gel slice was dehydrated by adding 100% (v/v) acetonitrile and then digested with trypsin (Promega). After overnight digestion, the peptides were extracted using extraction buffer [5% formic acid (Merck Millipore) and 40% acetonitrile (Merck Millipore)] and were vacuum-dried using SpeedVac and stored at -80 °C.

3.1.2.2 Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

The dried peptides were reconstituted in 0.1% formic acid (Merck Millipore) and analyzed using Q Exactive™ (Thermo Scientific™) mass spectrometer coupled with the Proxeon Easy nLC system (Thermo Scientific™). For peptide enrichment, protein fragments were passed on to an Acclaim™ PepMap™ trap column (Michrom Biosciences Inc.). Peptides were separated on an analytical column employing a linear gradient of 7–30% acetonitrile for 80 minutes. MS and MS/MS scan acquisitions were executed in the quadrupole Orbitrap mass analyzer.

3.1.2.3 Data analysis

The MS data were searched against NCBI HsRefSeq81 (human protein database; version 81) and analyzed using the Mascot search engine (Matrix Science, London, UK; version 2.2.0) and SequestHT program and Proteome Discoverer software (Thermo Fisher Scientific; version 1.4.0.288).

3.1.3 Far-ultraviolet Circular Dichroism spectroscopy

The secondary structure of the purified and desalted recombinant fusion proteins (rhGOI-NTH or rhHTN-GOI) were determined using far ultraviolet (UV) circular dichroism (CD) spectra from 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 minimum of 5 accumulations from wavelength 260–190 nm in a 0.1 cm path length quartz cuvette at a scan rate of 100 nm/min with a data integration time of 1 second. From the sample spectrum, background noise due to the sodium phosphate buffer was subtracted, and the final spectrum was analyzed and quantified using *in silico* Beta Structure Selection (BeStSel) online tool (http://bestsel.elte.hu/index.php) to estimate the secondary structure of the purified protein. Detailed information about the BeStSel algorithm is described elsewhere (Micsonai et al. 2015; Micsonai et al. 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 1, Materials and Methods section for details**). All primary and secondary antibodies used and their respective concentrations are listed in the Appendix 1.

3.2 Results and discussion

3.2.1 Purification of recombinant fusion proteins

To retain the native folding conformation of these recombinant human fusion proteins (rhGOI-NTH and rhHTN-GOI), we first achieved its soluble expression and then performed one-step affinity chromatography-based purification under native conditions. Several studies demonstrated that recombinant proteins purified under native conditions retain their folding conformation, and often also retain their functional activity (Bosnali and Edenhofer 2008; Stock et al. 2010; Thier et al. 2010; Thier et al. 2012; Peitz et al. 2014; Münst et al. 2016; Dey et al. 2021b; Narayan et al. 2021a; Narayan et al. 2021b; Thool et al. 2021; Dey et al. 2022). Also, the recombinant proteins should be purified to homogeneity, or else the application of impure proteins (with bacterial contaminant proteins) on the target mammalian cells could have detrimental effects (Araki et al. 2012). This is critical in producing therapeutic or reprogramming proteins as these bacterial contaminant proteins might evoke an immune response or can induce undesired effects when applied to target mammalian cells (Araki et al. 2012). To achieve this, we employed IMAC technique under native conditions to purify these poly-histidine tag-fused recombinant human fusion proteins in their native form. IMAC is a widely used purification technique that depends on the interactions between the poly-histidine residues and charged transition metal ions such as Ni^{2+} in this study, immobilized on a matrix such as nitrilotriacetic acid in this study. Thus, the recombinant fusion proteins were first induced in large 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.

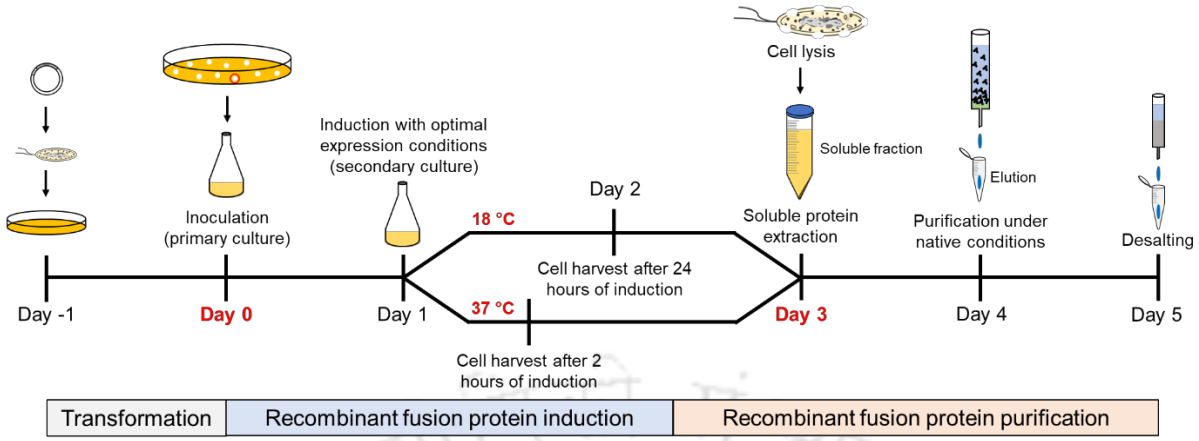


Figure 3.1 Timeline and schematic representation of the overall purification experimental strategy.

3.2.1.1 Purification of rhGATA4 fusion proteins

Based on the soluble expression analysis, both the fusion proteins (rhGATA4-NTH and rhHTN-GATA4) were expressed at the identified parameters (Table 2.4; Chapter 2) and purified to homogeneity. The high purity of the rhGATA4 fusion proteins was confirmed by SDS-PAGE analysis (Figure 3.2A and 3.2B), which was not reported in the earlier study (Li et al. 2015). In fact, no biochemical data showing the purification of human GATA4 protein were reported (Li et al. 2015). A band of ~55 kDa corresponding to rhGATA4 fusion proteins was observed when the purified fractions were run on 12% SDS-PAGE gel (Figure 3.2A and 3.2C). The elution profile of this rhGATA4-NTH fusion protein purification is shown in Figure 3.2B. Although overall expression of the rhHTN-GATA4 protein was higher than rhGATA4-NTH, the yield of rhHTN-GATA4 (0.38 mg/L) was very low (Figure 3.2D) compared to its counterpart (0.98 mg/L). The rhGATA4 fusion protein purification data are summarized in Table 3.3.

Table 3.3 Purification summary of recombinant fusion proteins.

Protein	Steps	Total protein (mg)[b]	Target protein (mg)[c]	Yield (%)	Purity (%)[d]
rhGATA4-NTH	Crude lysate [a]	102.44	15.84	100	15.46
	Cleared lysate (soluble)	92.14	11.36	71.74	12.33
	IMAC (pooled)	0.16	0.15	0.95	94.4
rhMEF2C-NTH	Crude lysate [a]	97.45	20.31	100	20.84
	Cleared lysate (soluble)	83.59	13.62	67.05	16.29
	IMAC (pooled)	0.24	0.23	1.1	92.88
rhTBX5-NTH	Crude lysate [a]	110.76	23.20	100	20.95
	Cleared lysate (soluble)	97.40	15.88	68.42	16.31
	IMAC (pooled)	0.33	0.31	1.99	93.67
rhHTN-ETS2	Crude lysate [a]	116.83	25.34	100	21.74
	Cleared lysate (soluble)	102.89	18.19	71.62	17.68
	IMAC (pooled)	0.26	0.24	0.94	92.06
rhHTN-MESP1	Crude lysate [a]	90.84	12.96	100	14.27
	Cleared lysate (soluble)	84.17	12.04	92.92	14.31
	IMAC (pooled)	0.63	0.61	4.69	96.46
rhHTN-HAND2 (induced at 37 °C)	Crude lysate [a]	58.89	10.40	100	17.67
	Cleared lysate (soluble)	55.30	9.88	95	17.87
	IMAC (pooled)	0.88	0.87	4.13	99.01
rhHTN-HAND2 (induced at 18 °C)	Crude lysate [a]	54.65	9.13	100	16.70
	Cleared lysate (soluble)	50.76	7.65	83.79	15.07
	IMAC (pooled)	0.62	0.60	5.37	97.92

[a] From 1 g wet weight of induced *E. coli* BL21(DE3) cell pellet (from 0.15-0.35 L of culture).

[b] Protein concentration determined by Bradford assay using Bovine Serum Albumin as a standard protein.

[c] Determined from total protein concentration and purity.

[d] Purity determined from SDS-PAGE analysis using ImageJ software.

Based on the Western blotting analysis (Figure 3.2A and 3.2C; *middle* and *bottom*), loss of rhGATA4 protein was detected in the flow-through/unbound fractions of both the fusion proteins. The main reason behind this low yield of rhHTN-GATA4 was the loss of the majority of protein molecules in flow-through fractions during purification. These observations indicated that the poly-histidine tag could have got buried during protein folding in these protein molecules. For this reason, many protein molecules did not efficiently bind to the Ni-NTA resin and were eventually lost in the flow-through fraction. To prevent the loss, purification of rhGATA4-NTH protein under mild denaturing conditions was performed by exposing the poly-histidine tag with 2 or 4 M urea. The SDS-PAGE and Western blotting analysis confirmed that rhGATA4-NTH was purified successfully, and based on the intensity, an increase in protein yield in 4 M urea treated soluble fraction was observed compared to 0 and 2 M treated soluble fraction (Figure 3.2E). These results confirm that the loss of the majority of the GATA4 fusion protein molecules in flow-through fractions could be due to the unexposed poly-histidine tags (Figure 3.2F). Also, the terminal at which the fusion tags were fused with GATA4 plays a crucial role in protein expression (in terms of low or high) and purification (in terms of protein yield). Furthermore, the identity of this purified rhGATA4 fusion protein was confirmed by Western blotting using GATA4 protein-specific antibody (Figure 3.2A; *bottom*). In this study, we report for the first time the biochemical data showing the purification of human GATA4 protein to homogeneity under native conditions. Thus, we have demonstrated the purification of rhGATA4 fusion proteins to homogeneity, and from these results, the rhGATA4-NTH protein purified under native or mild denaturation conditions will be used for further analysis.

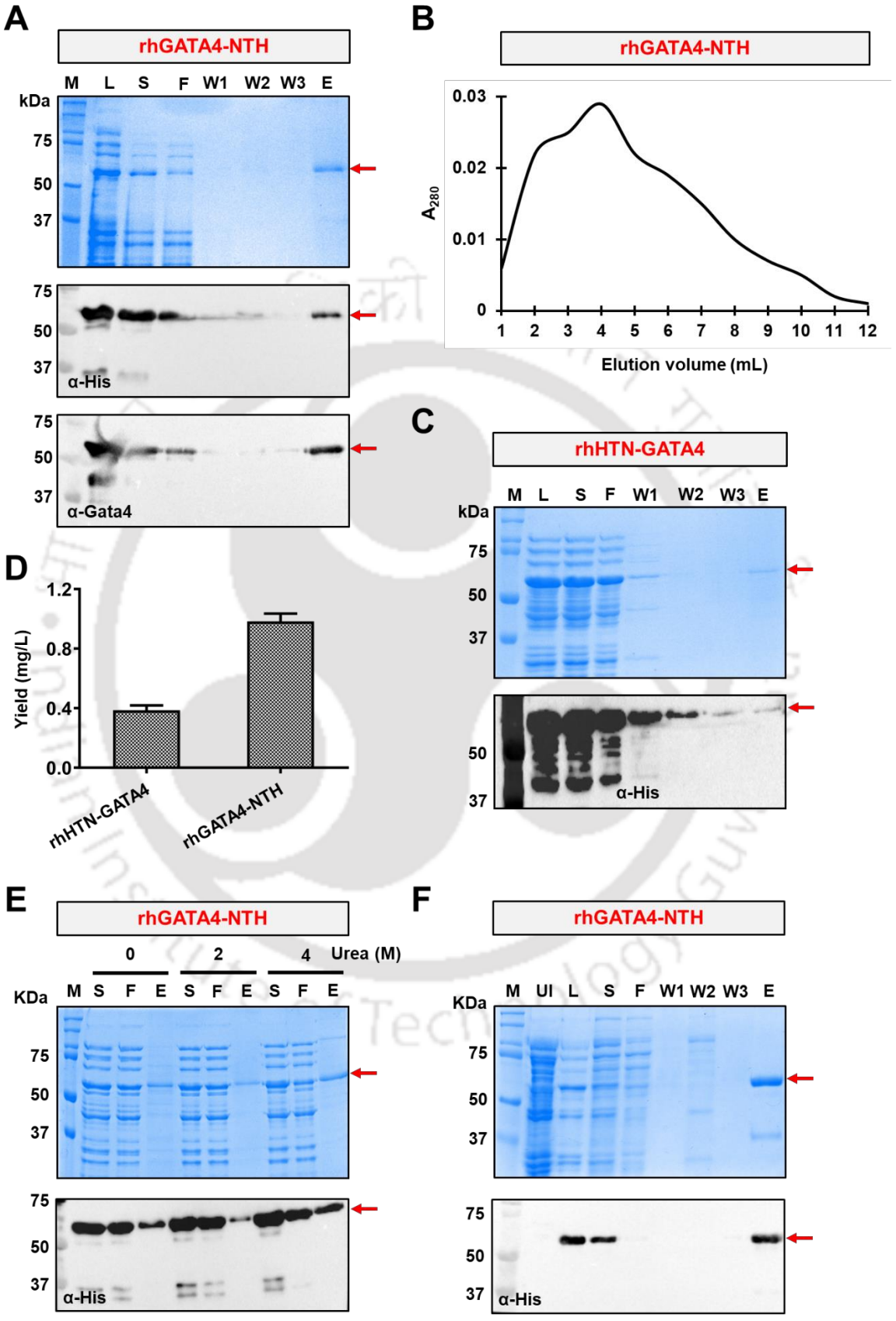


Figure 3.2 SDS-PAGE and Western blotting analysis of rhGATA4 fusion protein purification under native or mild denaturation conditions. (**A**) rhGATA4-NTH protein purification under native conditions. (**B**) Elution profile of rhGATA4-NTH protein purification under native conditions. (**C**) rhHTN-GATA4 protein purification under native conditions. (**D**) The bar graph shows the final yield of purified rhHTN-GATA4 and rhGATA4-NTH proteins under native conditions, (n=3). (**E**) Identification of the minimal denaturant concentration required for the maximum purification yield of rhGATA4-NTH. (**F**) rhGATA4-NTH protein purification under mild denaturation (with 4 M urea) conditions. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; F, flow-through fraction; W, wash fraction; E, eluted fraction; UI, uninduced (total cell lysate).

3.2.1.2 Purification of rhMEF2C-NTH protein

From the soluble expression analysis, we selected a C-terminally fused genetic construct of *MEF2C* and induced its expression at 37 °C for 2 hours or 18 °C for 24 hours, then purified it to homogeneity. The SDS-PAGE analysis clearly showed a single high-intensity band at ~58 kDa corresponding to the full-length rhMEF2C-NTH protein, signifying the homogeneously purified proteins in the elution fractions irrespective of induction temperature [Figure 3.3A; *top* (for 37 °C) and data not shown (for 18 °C)]. This purification data thus implies that the purified fusion protein is free of other proteins from the host expression system. Western blotting with MEF2C protein-specific antibody revealed the loss of rhMEF2C-NTH protein in the flow-through fractions (Figure 3.3A; *middle* and *bottom*), which might be probably due to overloading of protein molecules or low resin volume in the affinity column. The purification profile (only elution) of this rhMEF2C-fusion protein (induced at 37 °C) is shown in Figure 3.3B. Moreover, Western blotting with Histidine and MEF2C protein-specific antibodies confirmed the identity of this purified rhMEF2C fusion protein (Figure 3.3A; *bottom*).

Interestingly, the final yield obtained was around ~1.27 (induced at 37 °C) and 0.51 (induced at 18 °C) mg/L for purified rhMEF2C-NTH proteins (Figure 3.3C). The rhMEF2C-NTH protein purification data are summarized in Table 3.3. Notably, this is the first study to report the biochemical data showing the purification of full-length human MEF2C protein to homogeneity under native conditions. Thus, we demonstrate the homogeneous purification of rhMEF2C fusion protein irrespective of the induction temperature under native conditions from *E. coli*.

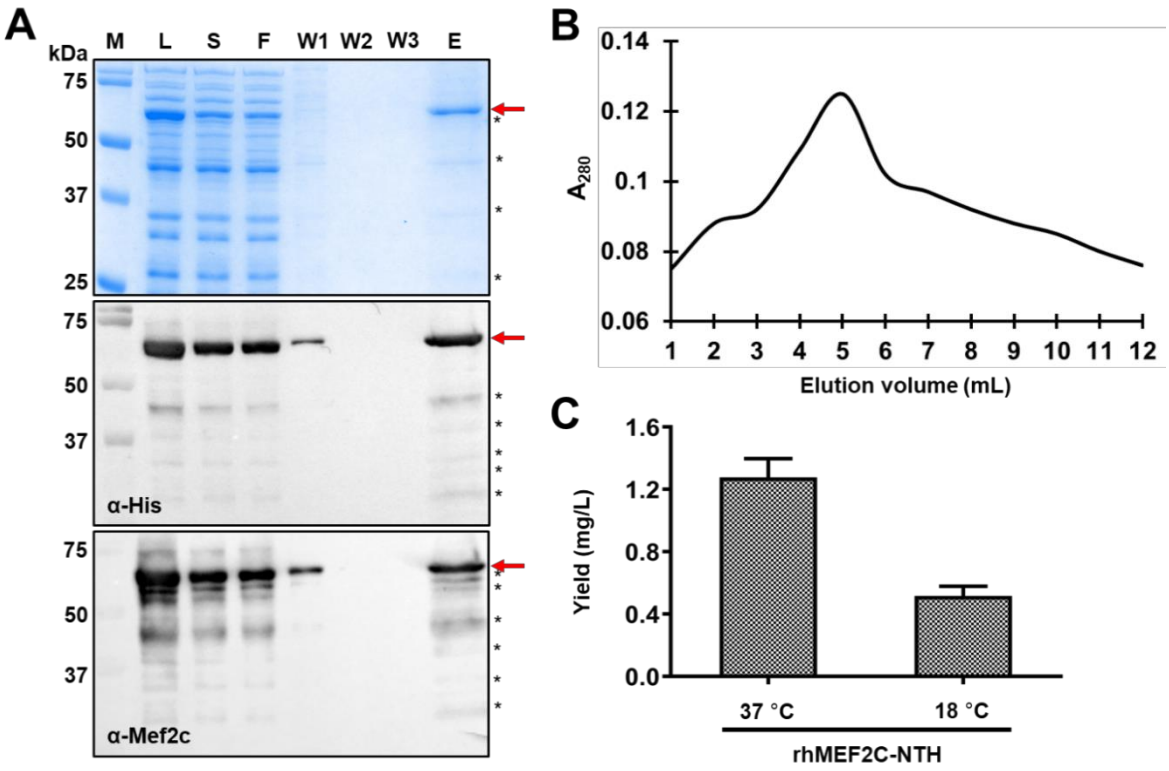


Figure 3.3 SDS-PAGE and Western blotting analysis of rhMEF2C-NTH protein purification under native conditions. (**A**) SDS-PAGE and Western blotting analysis of rhMEF2C-NTH protein purification. (**B**) Elution profile of rhMEF2C-NTH protein purification. (**C**) The bar graph shows the final yield of purified rhMEF2C-NTH protein-induced at 37 and 18 °C, (n=3). M, protein

marker; L, cell lysate; S, supernatant (soluble) fraction; F, flow-through fraction; W, wash fraction; E, eluted fraction; *truncated fragments of rhMEF2C-NTH protein.

3.2.1.3 Purification of rhTBX5-NTH protein

After having successfully expressed rhTBX5 fusion protein in soluble form (although in low amounts), we next sought to purify this protein with one-step affinity chromatography-based purification. Notably, the purification data revealed that the elution fraction was free of bacterial proteins (Figure 3.4A). At ~64 kDa, a single high-intensity band corresponding to full-length rhTBX5-NTH protein-induced at 37 °C was observed. The purity of the rhTBX5-NTH protein was >90% in the elution fraction, as quantified with ImageJ (Table 3.3). Western blotting with Histidine (Figure 3.4A; *middle*) and Tbx5 (Figure 3.4A; *bottom*) antibodies, we detected the loss of this fusion protein in the unbound fraction, probably because of the low resin volume or overloading of protein–molecules on the affinity column. Furthermore, the identity of this purified rhTBX5-NTH protein was verified by Western blotting with an antibody specific to Tbx5 (Figure 3.4A; *bottom*). The elution profile of this rhTBX5-NTH protein purification is shown in Figure 3.4B. Notably, with the optimized induction parameters and one-step purification, we obtained a final yield of ~2 mg/L of purified rhTBX5-NTH protein. The rhTBX5-NTH protein purification data are summarized in Table 3.3. Importantly, this is the first study to report the purification of full-length human TBX5 protein to homogeneity under native conditions. Thus, we have demonstrated the one-step homogeneous purification of rhTBX5 fusion protein under native conditions from *E. coli*.

3.2.1.4 Purification of rhHTN-ETS2 protein

We also sought to purify rhHTN-ETS2 protein under native conditions. It is apparent from the SDS-PAGE analysis that rhHTN-ETS2 has been purified without any bacterial proteins (Figure 3.5A; *top*). The purified rhETS2 fusion protein was further characterized by Western

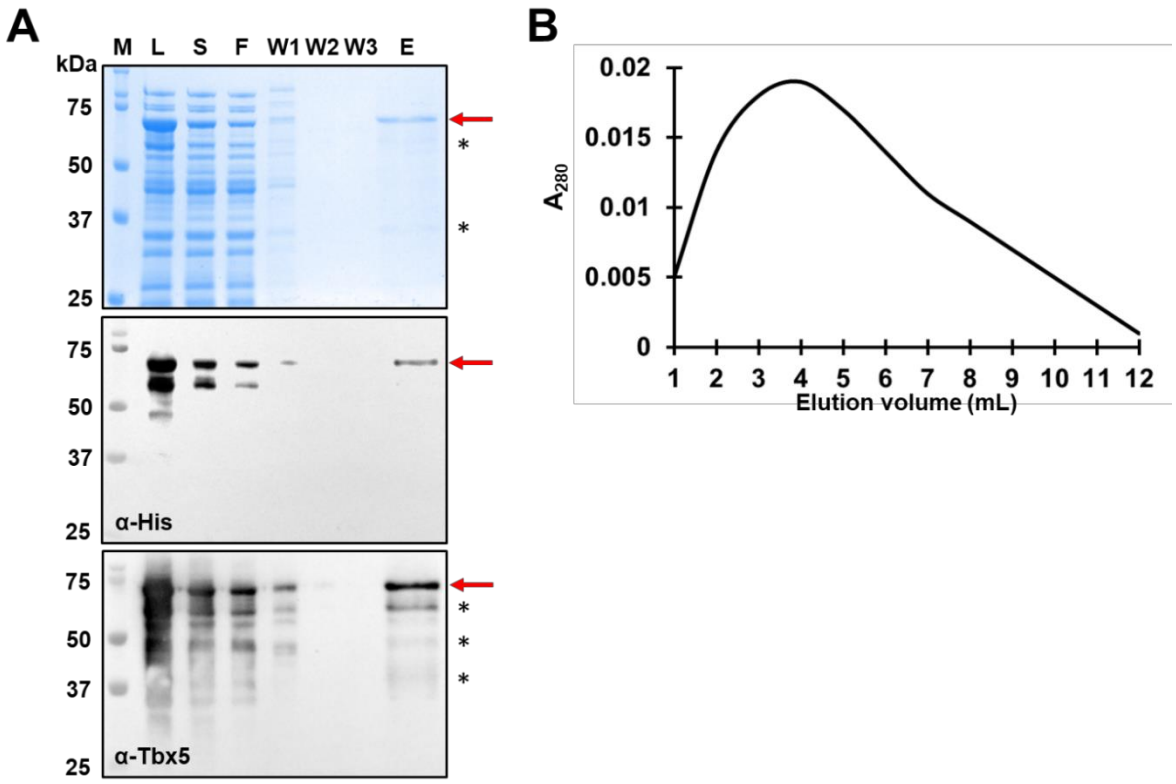


Figure 3.4 SDS-PAGE and Western blotting analysis of rhTBX5-NTH protein purification under native conditions. (**A**) SDS-PAGE and Western blotting analysis of rhTBX5-NTH protein purification. (**B**) Elution profile of rhTBX5-NTH protein purification. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; F, flow-through fraction; W, wash fraction; E, eluted fraction; *truncated fragments of rhTBX5-NTH protein.

blotting with Histidine (Figure 3.5A; *middle*) and Ets2 (Figure 3.5A; *bottom*) antibodies, and from this result, we confirmed the identity of this purified recombinant protein. Further, Western blotting analysis revealed the loss of rhHTN-ETS2 protein molecules in the flow-

through fractions (Figure 3.5A; *middle* and *bottom*), which might be probably due to the overloading of protein molecules or low resin volume in the affinity column. The elution profile of this rhHTN-ETS2 protein is shown in Figure 3.5B. The rhHTN-ETS2 protein purification data are summarized in Table 3.3, and the final yield of this purified fusion protein was around ~1.65 mg/L. This is the first study to demonstrate the generation of highly pure rhETS2 protein with fusion tags that can allow cell penetration and nuclear translocation into mammalian cells. An earlier study that reported the generation of human ETS2 protein from *E. coli* could not obtain a highly pure protein as additional bands were also observed in the SDS-PAGE gel (Islas et al. 2012). Thus, the one-step purification of rhETS2 fusion protein under native conditions from soluble fractions was demonstrated.

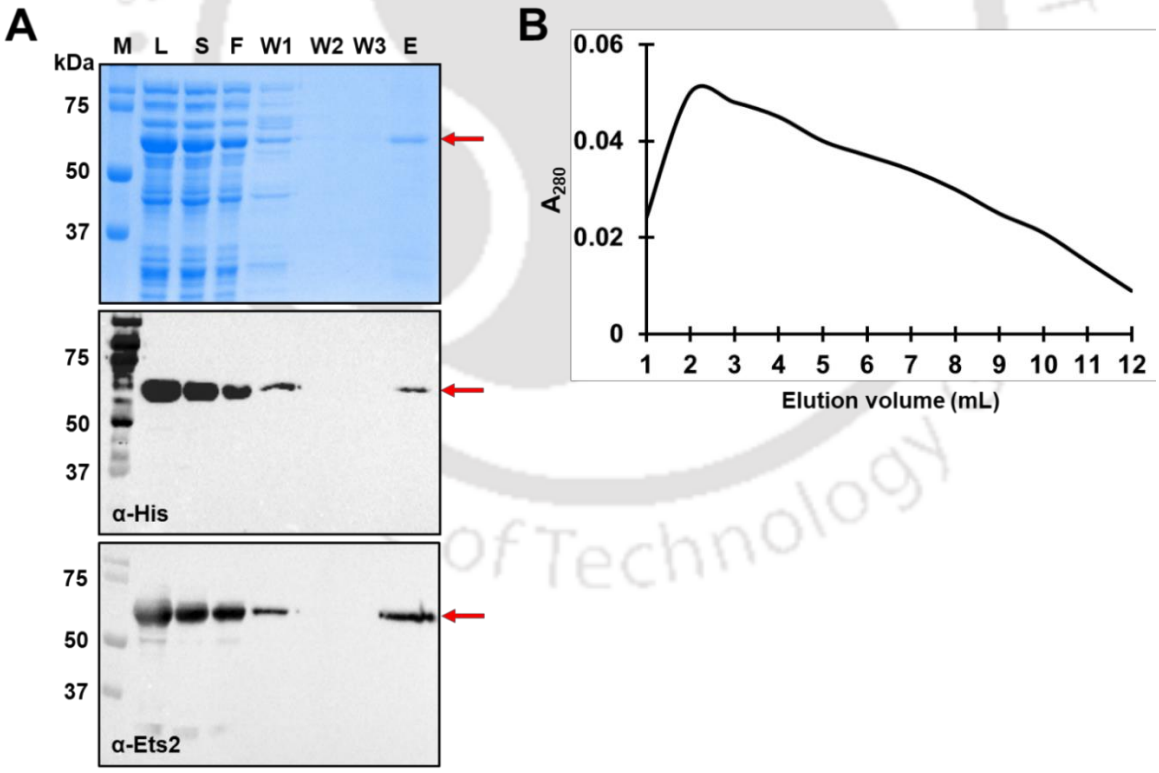


Figure 3.5 SDS-PAGE and Western blotting analysis of rhHTN-ETS2 protein purification under native conditions. (**A**) SDS-PAGE and Western blotting analysis of rhHTN-

ETS2 protein purification. (**B**) Elution profile of rhHTN-ETS2 protein purification. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; F, flow-through fraction; W, wash fraction; E, eluted fraction.

3.2.1.5 Purification of rhHTN-MESP1 protein

As identification of optimal conditions facilitated the expression of the rhHTN-MESP1 protein in a soluble form (**see Chapter 1, Results and Discussion section for details**), we next purified this fusion protein under native conditions with two different salt concentrations that gave soluble protein. First, purification was carried out in the absence of salt (0 mM NaCl) and then in the presence of salt (150 mM NaCl) in all the purification buffers. The results show that rhHTN-MESP1 recombinant protein has been purified to homogeneity without any bacterial contaminants in both conditions (Figure 3.6A and 3.6B). Further, the fusion protein was detected by Western blotting using Histidine antibody (Figure 3. 6A and 3.6B; *middle*). Moreover, our results showed that increasing the salt concentration in the purification buffers decreased the purified rhMESP1 fusion protein yield (Figure 3.6A and 3.6B). The purification data are summarized in Table 3.3, and the protein yield of purified rhHTN-MESP1 without salt was around ~4.85 mg/L, whereas, with 150 mM salt concentration, it was around ~0.92 mg/L. This could be due to the promotion of unfavorable interaction between salt and protein, resulting in salt-dependent aggregation during the purification procedure (Tsumoto et al. 2007), thus affecting the protein yield. Further, the identity of the purified fusion protein was confirmed using Western blotting with Mesp1 antibody (Figure 3. 6A and 3.6B; *bottom*). A band corresponding to human MESP1 fusion protein was observed at ~39 kDa (Figure 3. 6A and 3.6B). The calculated molecular weight of the rhMESP1 fusion protein is 35.5 kDa. The elution profile of one-step affinity-purified rhMESP1 fusion proteins without salt is shown in

Figure 3.6C. Notably, this is the first study to establish the generation of highly pure full-length rhMESP1 fusion protein free of bacterial proteins, and the fusion tags can allow cell penetration and nuclear translocation into mammalian cells. An earlier study that reported the generation of human MESP1 protein from *E. coli* could not obtain a highly pure protein as additional bands were also observed in the SDS-PAGE gel (Islas et al. 2012). Thus, we demonstrate the homogeneous purification of rhMESP1 fusion protein under native conditions from *E. coli.*

3.2.1.6 Purification of rhHTN-HAND2 protein

To purify rhHTN-HAND2 protein from soluble cell fraction, we first investigated the effect of imidazole concentration on the elution of this fusion protein, induced at 37 °C. As shown in Figure 3.7A, this recombinant fusion protein started eluting with 200 mM of imidazole, and the maximum amount was eluted with 250-350 mM of imidazole. Importantly, no bacterial proteins were observed in any of the elution. The imidazole gradient elution profile of affinity-purified rhHAND2 fusion protein, induced at 37 °C, is shown in Figure 3.7B. Thus, for the one-step purification of the rhHAND2 fusion protein, induced at two different temperatures, we used a maximum of 150 mM of imidazole during washing and 300 mM of imidazole for eluting the rhHAND2 fusion protein.

From the SDS-PAGE and Western blotting analysis, it is clear that the rhHTN-HAND2 protein, induced at two different temperatures, 37 and 18 °C, has been purified without bacterial contaminants (Figure 3.8). A single band (~34 kDa) in the eluted fractions was observed on SDS-PAGE, indicating the high purity of rhHAND2 fusion protein (Figure 3.8; *top*). However, loss of HAND2 fusion protein was detected in the Western blotting analysis (Figure 3.8) in the flow-through/unbound fraction. This could probably be due to the

overloading of the soluble fraction on the purification column or the low volume of resin used for purification. Further, this fusion protein identity was established by Western blotting with

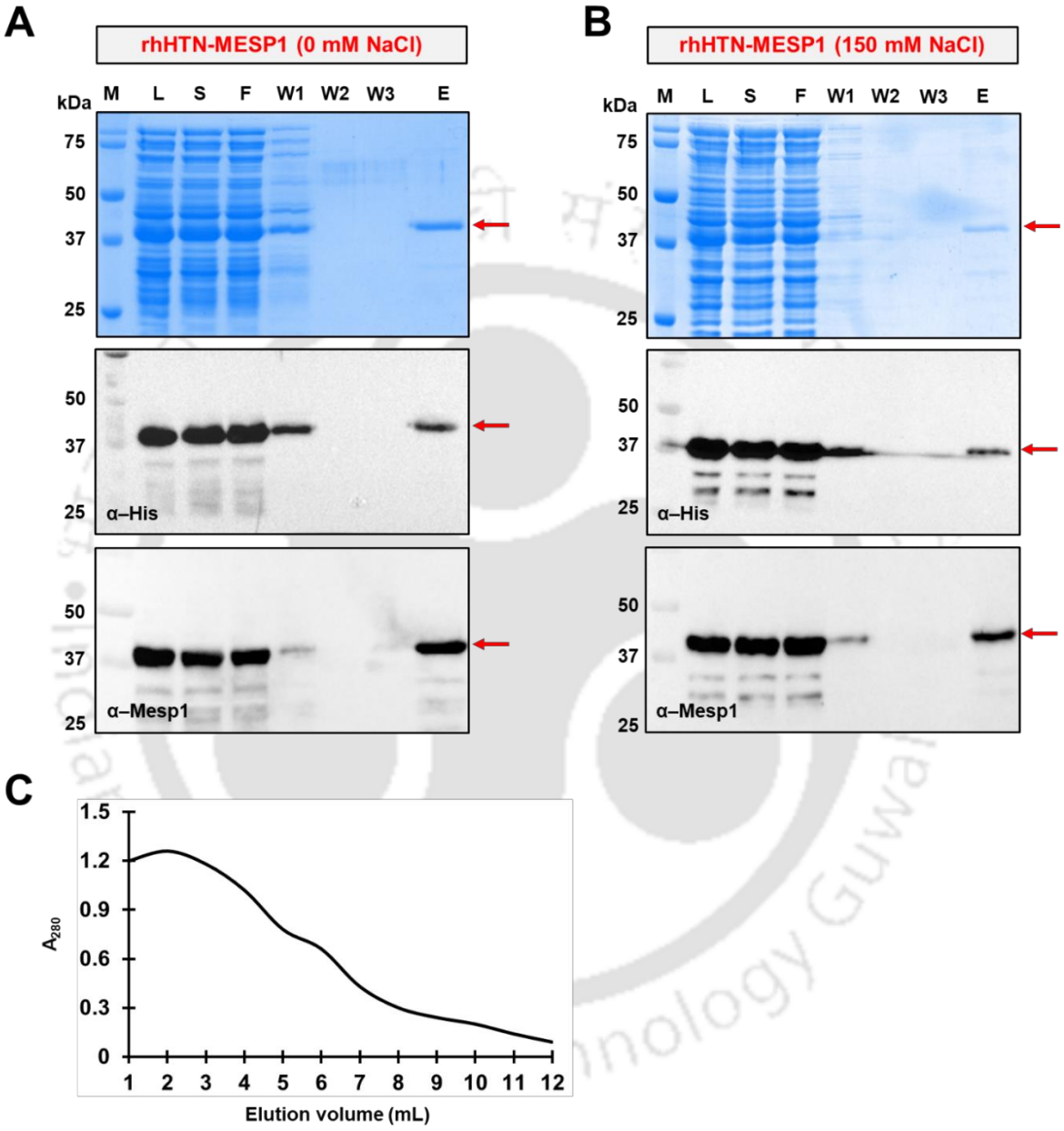


Figure 3.6 SDS-PAGE and Western blotting analysis of rhHTN-MESP1 protein purification under native conditions. SDS-PAGE and Western blotting analysis of rhHTN-MESP1 protein purification (**A**) without salt and (**B**) with salt (150 mM NaCl). (**C**) Elution profile of

rhHTN-MESP1 protein purification without salt. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; F, flow-through fraction; W, wash fraction; E, eluted fraction.

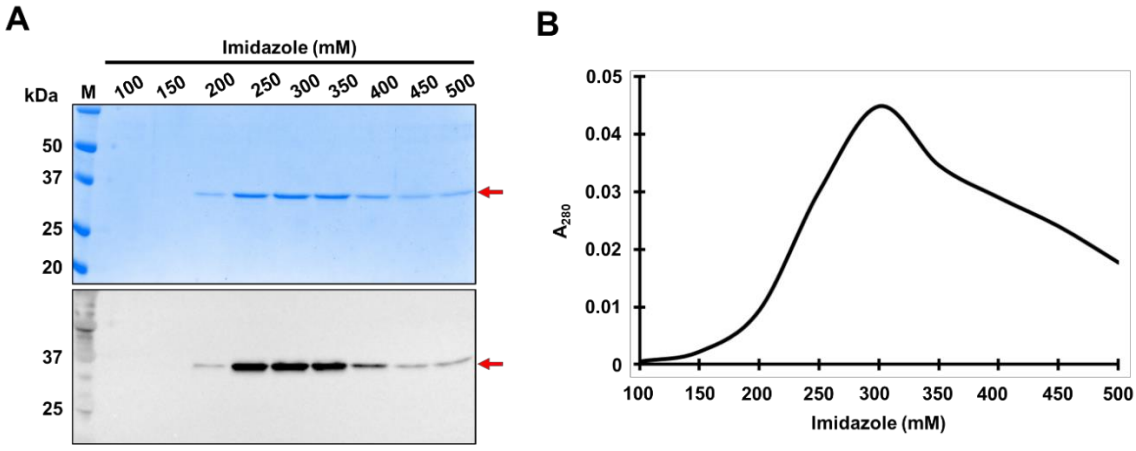


Figure 3.7 Effect of imidazole concentration on the amount of rhHAND2 fusion protein eluted. (**A**) SDS-PAGE and Western blotting analysis of rhHTN-HAND2 protein purification under native conditions. During purification, the imidazole concentration in the elution buffer was varied to see its effect on the amount of protein eluted. As mentioned in the figure, different concentrations of imidazole were used in the elution buffer, and the protein was eluted. (**B**) The absorbance of the protein eluted using different concentrations of imidazole in elution buffer was measured at 280 nm, and a graph was plotted with imidazole concentration on the x-axis against absorbance at 280 nm on the y-axis. M, protein marker.

the Hand2 antibody (Figure 3.8; *bottom*). Interestingly, the purified rhHTN-HAND2 protein was eluted efficiently only from the second elution fraction (Figure 3.8), signifying that the interactions between the protein molecules and Ni-NTA were stronger and required more elution with an elution buffer to weaken these interactions before eluting. The elution profile of one-step affinity-purified rhHAND2 fusion proteins induced at 37 °C and 18 °C is shown in Figure 3.8C and Figure 3.8D, respectively. The purification data are summarized in Table 3.3, and the final yield of the purified rhHAND2 fusion proteins was around 3.38 and 2.73

mg/L when induced at 37 and 18 °C, respectively. Remarkably, this is the first study to demonstrate the one-step purification of full-length rhHAND2 protein to homogeneity under native conditions. Thus, the homogeneous purification of rhHAND2 fusion protein under native conditions from soluble fractions was demonstrated, irrespective of the induction temperatures.

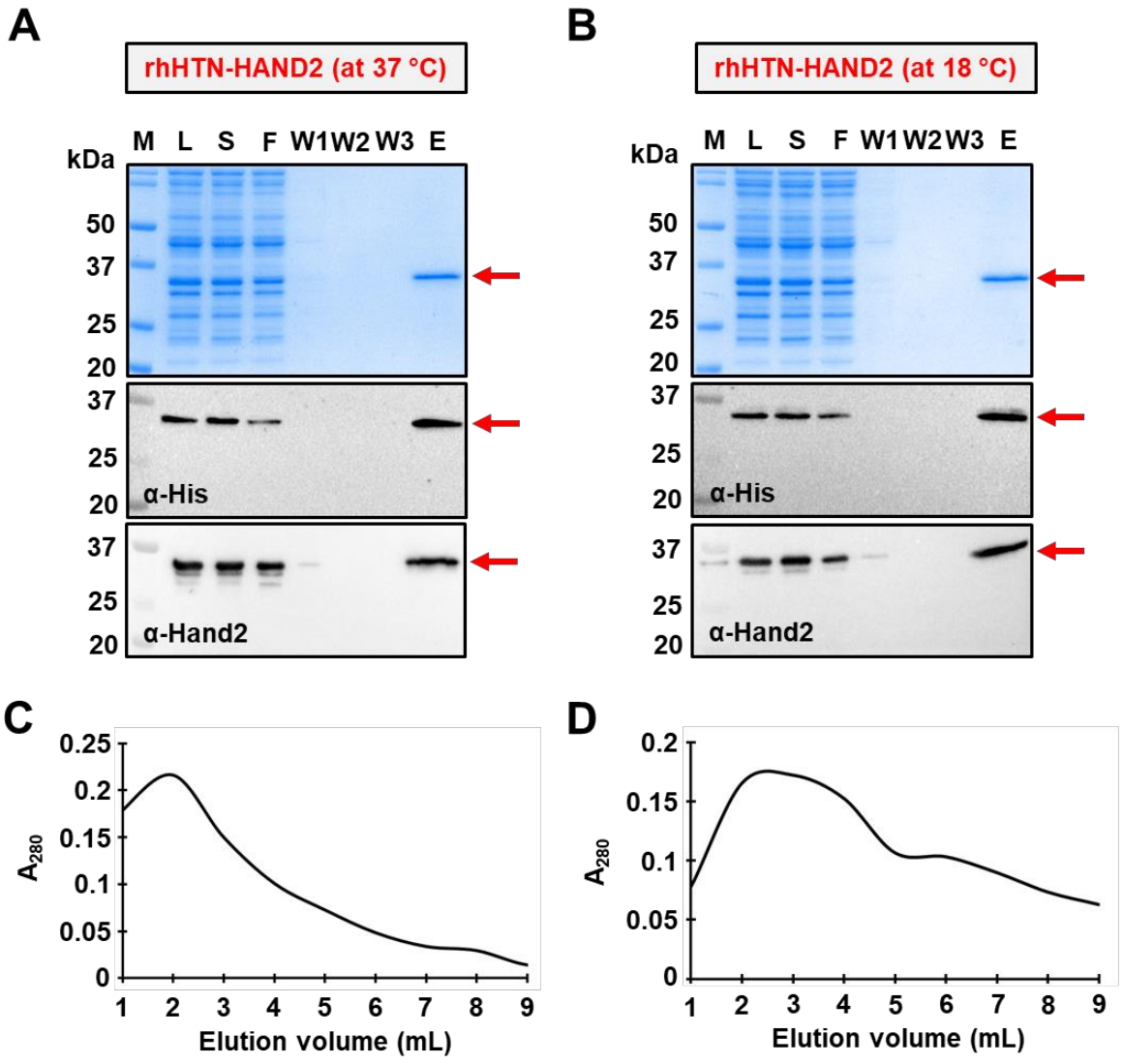


Figure 3.8 SDS-PAGE and Western blotting analysis of rhHTN-HAND2 protein purification under native conditions. SDS-PAGE and Western blotting analysis of the

purification of rhHTN-HAND2 protein (**A**) induced at 37 °C and (**B**) 18 °C. Elution profile of the purification of rhHTN-HAND2 protein (**C**) induced at 37 °C and (**D**) 18 °C. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; F, flow-through fraction; W, wash fraction; E, eluted fraction.

Several other reprogramming transcription factors, namely, OCT4, SOX2, NANOG, PDX1, GLIS1 and NGN3 were generated using similar strategies from *E. coli* in the form of recombinant proteins (Bosnali and Edenhofer 2008; Stock et al. 2010; Thier et al. 2010; Thier et al. 2012; Peitz et al. 2014; Münst et al. 2016; Dey et al. 2021b; Narayan et al. 2021a; Narayan et al. 2021b; Thool et al. 2021). These studies have also used NLS and protein transduction domains to facilitate recombinant protein delivery to target sites in mammalian cells. Most of these studies have reported that the presence of these fusion tags after the delivery of the protein of interest into the cells (via TAT) and nucleus (via NLS) did not affect the biological activity of the protein (Bosnali and Edenhofer 2008; Stock et al. 2010; Thier et al. 2010; Thier et al. 2012; Peitz et al. 2014; Münst et al. 2016; Dey et al. 2021b; Narayan et al. 2021a). Thus, the cytosolic and nuclear delivery of these six purified transcription factors in the form of recombinant proteins can be achieved via NLS and TAT fusion without the need for any additional transduction reagents.

3.2.2 Secondary structure determination of these purified recombinant fusion proteins

The experimental data showing the secondary structure content of these six human proteins are not reported to date. To determine the secondary structure conformation of these recombinant fusion proteins, we first dialyzed (in the case of purification under mild denaturation conditions) or buffer exchanged (in the case of purification under native conditions) and/or desalted these purified recombinant fusion proteins to evade background noise due to NaCl, and then performed the far-UV CD spectroscopy. This spectroscopic technique is used to

estimate the secondary structure content of proteins purified from cells/tissues (Greenfield 2006), particularly for proteins whose secondary structure content is not known, such as GATA4, MEF2C, TBX5, ETS2, MESP1, HAND2 and so forth. The characteristic shape and magnitude of the far-UV CD spectrum represent different secondary structures, namely α-helix, β-sheet, turn, and other structural components that include 3_{10}-helix, π-helix, β-bridge, bend, loop/irregular and invisible regions of the structure (Kelly et al. 2005; Greenfield 2006). Notably, α-helix has negative peaks at 222 nm and 208 nm and a positive peak at 193 nm, while the β-sheet has a negative peak at 218 nm and a positive peak at 195 nm (Greenfield 2006). Likewise, the random coil shows a positive peak at 210 nm and a negative peak of about 195 nm (Greenfield 2006). The obtained far-UV CD data was further quantified and analyzed using a recently developed online tool, BeStSel (Micsonai et al. 2015; Micsonai et al. 2018). It is a widely used *in silico* tool to predict the fold recognition and secondary structure content of the protein of interest from its CD spectrum (Micsonai et al. 2015; Micsonai et al. 2018).

From the CD spectrum plotted using BeStSel results (Figure 3.9), it is evident that the rhGATA4 fusion proteins purified under native conditions have maintained their secondary structure. Notably, the characteristic shape and magnitude of the CD spectra of both fusion proteins (rhHTN-GATA4 and rhGATA4-NTH) were similar (Figure 3.9), signifying that the secondary structure of the human GATA4 protein was independent of the terminal at which the tags were fused. Further, the estimated secondary structure of the purified rhGATA4-NTH protein under native conditions revealed that this protein predominantly has α-helices and other structural components in its secondary structure (Figure 3.10). The estimated secondary structure of this purified rhGATA4-NTH protein indicates that it constitutes ~32% α-helices, ~16% β-sheets, ~34% other structural components, and ~18% turns (Figure 3.10). Similarly,

the purified rhGATA4-NTH protein under mild denaturing conditions was dialyzed against 20 mM sodium phosphate buffer (to refold) and/or buffer exchanged and then analyzed using far-UV CD spectroscopy. From the CD spectra, it is clear that the secondary structure was not regained after several attempts of refolding approaches (Figure 3.10). To summarize these results, the rhGATA4 fusion proteins purified under native conditions, without undergoing any denaturation treatment, retained their secondary structure after purification.

We next sought to determine the secondary structure of rhMEF2C fusion protein and to understand the influence of induction temperature on its secondary structure folding conformations. BeStSel analysis of the obtained far-UV CD spectra showed that the purified rhMEF2C fusion protein retained its secondary structure irrespective of induction temperature (Figure 3.9). Additionally, quantification of the spectra revealed that the purified rhMEF2C fusion protein (induced at 37 or 18 °C) majorly consists of ~41% α-helices, ~24% turns, and ~22% other structural components, along with a small number of β-sheets (~13%) (Figure 3.10). From these results, we can infer that the purified rhMEF2C fusion protein retains its secondary structure, and the induction temperature does not influence its secondary structure.

Next, we determined the secondary structure of rhTBX5 fusion protein purified under native conditions. The plotted far-UV CD spectra (Figure 3.9) indicated that the purified rhTBX5-NTH proteins have proper secondary structure folding conformation after purification. Moreover, the CD spectrum analysis showed that this fusion protein consisted primarily of α-helices, β-sheets, and other structural components, along with a small number of turns (Figure 3.10). The estimated secondary structure of this purified rhTBX5-NTH protein indicates that it constitutes ~26% α-helices, ~20% β-sheets, ~40% other structural components,

and ~14% turns (Figure 3.10). This result indicated that the purified rhTBX5-NTH protein retained its secondary structure after purification under native conditions.

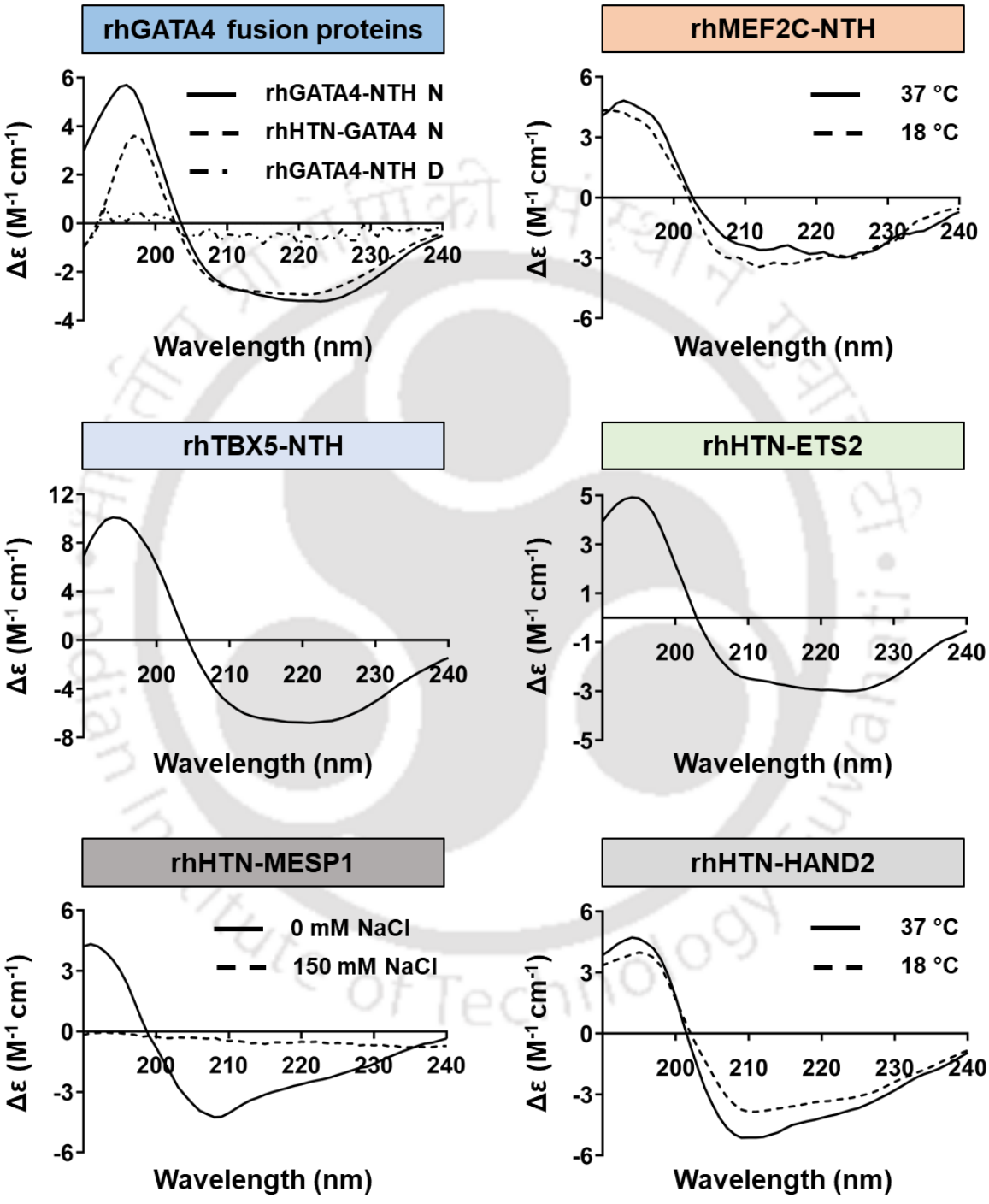


Figure 3.9 Secondary structure determination of purified recombinant fusion proteins with far-UV CD spectroscopy. The purified recombinant fusion proteins were dialyzed (only for rhGATA4-NTH purification under mild denaturation conditions) and/or desalted and then analyzed for their secondary structures 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. N, purification under native conditions; D, purification under mild denaturation conditions.

After having successfully purified rhHTN-ETS2 and rhHTN-MESP1 fusion proteins under native conditions, we next sought to determine their secondary structures using far-UV CD spectroscopy. The obtained far-UV CD spectra (plotted using BeStSel result) of rhHTN-ETS2 protein (Figure 3.9) confirmed that it had retained its secondary structure. Our predicted secondary structure (Figure 3.10) comprised predominantly of β-sheets (~28%) and other structural components (~44%) with a significant contribution of α-helices (~14%) and turns (~14%). Further, BeStSel analysis confirmed that the purified rhHTN-MESP1 protein in the absence of salt had maintained its secondary structure (Figure 3.9), and it majorly comprised of α-helices (~33.8%) and other structural components (~34%) with a significant contribution from β-sheets (~19%) and turns (~13%) (Figure 3.10). Next, we investigated the effect of salt (150 mM NaCl) on the folding conformations of the rhHTN-MESP1 protein. Notably, the presence of salt in the purification buffers compromises the folding conformations of rhMESP1 fusion protein (Figure 3.9), indicating again that the human MESP1 is an ion-sensitive protein. This might be due to the influence of unfavorable interaction between salt and protein, resulting in salt-dependent aggregation during the purification procedure (Tsumoto et al. 2007), thus affecting the secondary structure of the protein. Overall our results confirmed that the purified

rhETS2 and rhMESP1 (in the absence of salt) fusion proteins had retained their secondary structures after purification and are likely to be biologically active.

From the CD spectra of purified rhHTN-HAND2 protein, induced at two different temperatures (37 °C and 18 °C) (Figure 3.9), it is evident that HAND2 protein has maintained

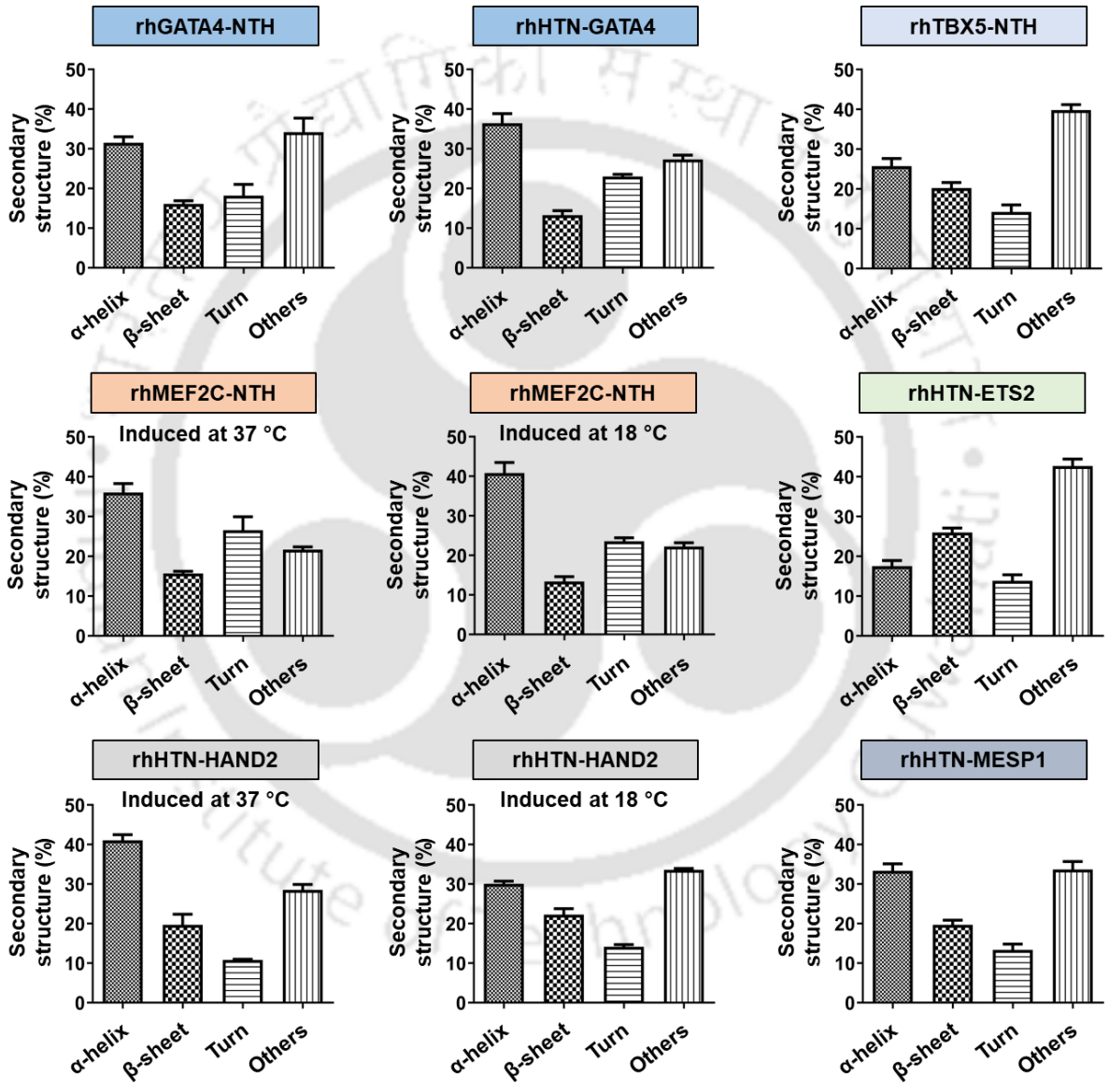


Figure 3.10 Summary of the quantified secondary structures of the purified recombinant fusion proteins from their spectra with the BeStSel server. The purified recombinant fusion

proteins were dialyzed (only for rhGATA4-NTH purification under mild denaturation conditions) and/or 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≥3).

its secondary structure. Moreover, the level of similarity between both spectra indicates that the secondary structure is consistent irrespective of the induction temperature. Subsequently,from the CD spectra, different secondary structures were quantified using BeStSel online tool. The secondary structure content of rhHAND2 fusion protein, induced at two different temperatures, 37 °C and 18 °C (Figure 3.10), revealed that the purified protein comprises majorly of α-helices and other structures. From the results (Figure 3.10), it is apparent that ~41% of α-helices and ~28% of other structures were observed when the induction temperature was 37 °C, whereas ~30% of α-helices and ~34% of other structures were observed when the induction temperature was 18 °C. Furthermore, there is a significant contribution of β-sheets [~20% (induced at 37 °C) and ~22% (induced at 18 °C)] and turns [~11% (induced at 37 °C) and ~14% (induced at 18 °C)] to the secondary structure of rhHAND2 fusion protein. Although the spectra appeared similar, variability in the secondary structure content was observed upon the quantification using BeStSel analysis. This variability is likely due to the induction and purification of rhHTN-HAND2 protein at different temperatures, which might have resulted in differences in their folding. Also, the possible minor differences in the protein concentrations (37 °C v/s 18 °C) may also be the reason for this variability.

In addition, the secondary structure analysis of bovine serum albumin was also performed as a control (data not shown), and the data suggested that the protein was predominantly α-helical, as reported in earlier studies (Reed et al. 1975; Lu et al. 2015; Manjushree and Revanasiddappa 2018). To date, the crystal structure of these six full-length human proteins is unknown; therefore, the similarity between their native structures and the secondary structures reported in this study is difficult to validate. Further comparison can be made only once the crystal structures of these full-length human proteins are available. Moreover, the fusion tags used in this study might have influenced the secondary structures of these full-length fusion proteins, thus accounting for deviations, if any, from their native structures. Overall, our secondary structure analysis indicated that these purified recombinant fusion proteins under native conditions retained their secondary structures after purification.

3.2.3 Mass spectrometry analysis of purified rhGATA4-NTH and rhHTN-HAND2 fusion proteins

The purified rhGATA4-NTH protein was primarily characterized using LC-MS/MS. LC-MS/MS is a widely used powerful analytical technique to identify and quantify peptides or the proteome of an organism (Karpievitch et al. 2010). In the present study, the Orbitrap LC-MS/MS technique was utilized. This powerful technique is used to confirm the identity of the purified recombinant proteins (Karpievitch et al. 2010). Therefore, the purified rhGATA4-NTH protein was trypsin digested to produce small peptides, and the resulting peptides were then separated, fragmented, ionized, and analyzed using Orbitrap mass spectrometry. The MS/MS study of digested peptides generated a match with human GATA4 protein, and from that, twelve unique peptide sequences (including two pairs of overlapping sequences) were identified (Figure 3.11A) to be a part of human GATA4 (Accession no.: NP_002043.2) using

Mascot, SequestHT algorithm and Proteome Discoverer software. Identified twelve unique peptide sequences are shown in Figure 3.11A. The annotated spectra of one of these twelve unique peptides (HPNLDMFDDFSEGR) are shown in Figure 3.11B. Thus, with the mass spectrometry and Western blotting analysis (Figure 3.2A; *bottom,* **see section 3.2.1.1 Purification of rhGATA4 fusion proteins for details**), the identity of the rhGATA4 protein was confirmed.

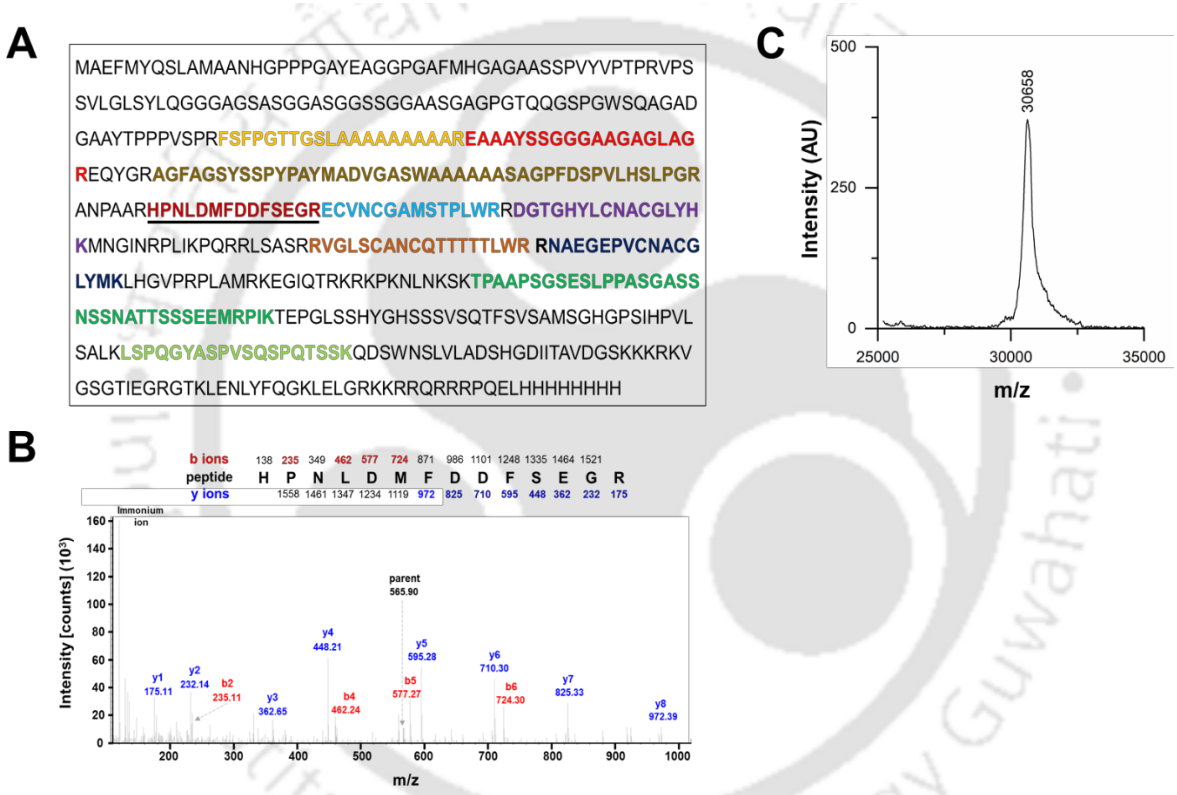


Figure 3.11 Characterization of the purified recombinant GATA4 and HAND2 fusion proteins. (**A)** The desired rhGATA4-NTH protein band was excised from the Coomassie-stained polyacrylamide gel and then destained, processed, and followed by trypsin digestion. The resulting peptide fragments were then analyzed using mass spectrometry, and the identified 12 unique peptide sequences were highlighted in the full-length amino acid sequence of the rhGATA4-NTH protein. (**B)** From the identified peptide sequences, annotated spectra have been plotted for one of the unique peptides, HPNLDMFDDFSEGR. (**C)** The purified and concentrated rhHTN-HAND2 protein was

diluted two times in α-Cyano-4-hydroxycinnamic acid matrix solution and then analyzed using MALDI-TOF mass spectrometry. The MALDI-TOF analysis is depicted using a graph with a mass per charge (m/z) ratio on the x-axis and intensity (AU) on the y-axis.

The purified rhHTN-HAND2 protein was desalted in deionized Milli-Q water using a PD-10 column and concentrated fivefold using a three kDa MWCO protein concentrator (Thermo scientific). The concentrated rhHAND2 fusion protein was diluted two times in α-Cyano-4-hydroxycinnamic acid matrix solution. The molecular weight was accurately determined using the Bruker Autoflex speed matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometer and found to be 30.6 kDa (Figure 3.11C). The monoisotopic molecular weight of the rhHTN-HAND2 fusion protein 30.687805 kDa was calculated from the amino acid sequence using Peptide Mass Calculator online server (https://www.peptidesynthetics.co.uk/tools/). The difference in the observed and calculated molecular weight remains well within the calibration error of the instrument. Thus, with the mass spectrometry, the molecular weight of the purified rhHTN-HAND2 protein was obtained and compared with the theoretical molecular weight of this protein.

3.3 Conclusions

This chapter demonstrates one-step homogeneous purification of these six recombinant fusion proteins under native conditions. The established protocols are simple, economical, and reproducible. In addition, this protocol involves purification from soluble cell lysate fraction to obtain a native protein and precludes purification from inclusion bodies; the latter is expensive, cumbersome, time-consuming, and requires the refolding of the denatured protein. One aspect where the purification procedure still requires further attention is to improve the yield of these fusion proteins as a good amount of protein molecules are lost in the flow-

through fraction. Although we improved the yield of rhGATA4 fusion protein by purifying this protein under mild denaturation conditions, we were not able to refold this fusion protein to native-like folding conformations. Markedly, we also showed the impact of the terminal at which the fusion tags were fused on the expression, purification and final yield of the rhGATA4 fusion proteins. Moreover, we have demonstrated the influence of induction temperature on the final yield of the purified rhMEF2C fusion protein, unlike rhHAND2 fusion protein. However, the difference in the secondary structure of the rhHAND2 fusion protein induced at two different temperatures was observed but not in the case of rhMEF2C fusion protein. In the previous chapter, we reported that optimal salt concentration is important for the enhanced solubility of the rhMESP1 fusion protein; here, we found that the presence of salt influence the final yield of the purified rhMESP1 protein and compromises its secondary structure. Our results thus suggest that the rhMESP1 is most likely to be an ion-sensitive protein. Furthermore, our secondary structure determination revealed that rhGATA4, rhMEF2C, rhMESP1, and rhHAND2 (induced at 37 °C) fusion proteins predominantly have α-helical secondary structures, whereas rhETS2 have predominantly β-sheets. In addition, rhTBX5 and rhHAND2 (induced at 18 °C) have a predominant mixture of both α-helical and β-sheets in their secondary structures. Prospectively, these purified recombinant proteins can be used to derive integration-free functional cardiomyocytes by circumventing the concerns that involve deleterious rearrangements due to viral or plasmid integration into the host chromosomes and also investigate its molecular role in cancer (Seth and Watson 2005; Xu et al. 2008; Islas et al. 2012; Kabbout et al. 2013). In addition, these proteins can be useful for further structural analysis, identifying potential novel interaction partners, and investigating their molecular functions in different cell types.

Chapter 4

Demonstration of cell penetration, nuclear translocation and biological activity of recombinant proteins

Brief Overview of the Chapter

In Chapter 3, we have demonstrated that the purified recombinant fusion proteins (GATA4, MEF2C, TBX5, ETS2, MESP1 and HAND2) have retained their secondary structure but further examination was required to validate their functional activity. Therefore, to determine the biological activity of these purified fusion proteins, we first confirmed their cell penetration and nuclear translocation ability by applying these proteins to human cell lines. Notably, we did not find any major differences in the cell penetration ability, however, we observed some differences in the nuclear translocation efficiency of these proteins between different cell lines. We then examined the role of the GATA4, TBX5, ETS2 and MESP1 fusion proteins in cancer. We showed the tumor suppressor role of GATA4 fusion protein and the oncogenic role of the ETS2 and MESP1 fusion proteins in mammalian breast cancer cells. We also demonstrated the tumor suppressor role of the TBX5 fusion protein in human colon cancer cells. Next, to determine the biological activity of other fusion proteins, namely MEF2C and HAND2, we investigated their role in angiogenesis. Notably, we showed that MEF2C and HAND2 fusion proteins act as negative and positive regulators of angiogenesis, respectively, using an *ex vivo* chicken embryo model. In the context of cardiac reprogramming, we demonstrated the synergistic activation of the α-MHC gene, a critical cardiac marker, by the MEF2C and HAND2 fusion proteins. This chapter thus concludes that the generated recombinant fusion proteins have cell penetration and nuclear translocation ability and are biologically active.

4.1 Materials and Methods

4.1.1 Stability analysis of purified recombinant fusion proteins

Glycerol stock of the purified recombinant fusion proteins were diluted with the transduction medium consisting of high-glucose Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 5% fetal bovine serum (FBS; Invitrogen) and 1% penicillin-streptomycin solution (P/S; Invitrogen) to the final concentration of 400 nM (protein transduction medium). The prepared protein transduction medium was then aliquoted in a 1.5 mL microcentrifuge tube and incubated at standard cell culture conditions (37 °C in a humidified atmosphere containing 5% CO_2 in a CO_2 incubator) for 48 hours. Samples were collected at different time points and stored immediately at -20 °C until further use. After collecting the 48 hours incubated sample, all the samples were then clarified by centrifugation at 10,000-14,000 rpm at 4 °C for 10 minutes. The clarified samples were then analyzed by Western blotting using a protein-specific antibody.

4.1.2 Cell culture

HeLa, SW620 and H9C2 cell lines were obtained from the National Centre for Cell Science, Pune, India. HeLa and SW620 cells were cultured in a complete growth medium containing 10% FBS and 1% P/S in high-glucose DMEM under standard cell culture conditions. H9C2 cells were cultured on 0.1% gelatin-coated (HiMedia) dishes in a complete growth medium containing 10% FBS and 1% P/S in low-glucose DMEM under standard cell culture conditions. Human foreskin fibroblasts (HFFs; BJ (ATCC® CRL-2522™) were obtained from the American Type Culture Collection, USA, and cultured on 0.1% gelatin-coated dishes in high-glucose DMEM supplemented with 10% FBS, 1% P/S, and 1X non-essential amino acids

(Invitrogen) under standard cell culture conditions. HiFi™ human umbilical vein endothelial cells (HUVEC) were procured from HiMedia Labs, India, and cultured on 0.5% gelatin-coated dishes in HiEndoXL™ endothelial cell expansion medium (HiMedia) under standard cell culture conditions. All the cell lines were passaged with trypsin-EDTA (Invitrogen) at ~70-80% confluency in the ratio of 1:4.

4.1.3 Protein transduction, immunocytochemistry and microscopy

Cells were plated onto 24-well culture plates at a density of 1×10^5 cells/well (for HeLa and SW620 cell lines) or 0.5×10^5 cells/well (for HFFs) in their respective growth medium. After cells had adhered to the plates, the medium was replaced with a protein transduction medium containing respective growth medium with 5% FBS (instead of 10%) supplemented with 400 nM recombinant fusion protein or an equal volume of glycerol buffer and incubated under cell culture conditions for 6 hours. After incubation, cells were washed twice with phosphate buffer saline (PBS) and fixed with 4% paraformaldehyde (Merck Millipore). Fixed cells were then permeabilized by treating with 0.1% Triton™ X-100 (Sigma-Aldrich) in PBS for 15 minutes. After permeabilization, cells were blocked with a blocking solution (0.5% bovine serum albumin, 0.15% glycine in PBS) for an hour at room temperature. The primary antibody was added and incubated overnight at 4 °C in a moist chamber. Cells were washed thrice with PBS and then incubated with the corresponding secondary antibody for an hour in a moist chamber at room temperature. After incubation, cells were washed three times with PBS and then stained with 4′,6-diamidino-2-phenylindole (DAPI) (1:15,000; Sigma-Aldrich) or Hoechst 33342 (1:10,000; Invitrogen) for 5 minutes. Excess staining was removed with PBS and then visualized under an inverted fluorescent microscope (IX83, Olympus, Japan) equipped with a DP80 CCD camera. Samples were illuminated using a pE-300 white CoolLED light source,

and image stacks were acquired using the 20x/0.45NA objective at 2 µm intervals. Images were analyzed by CellSens dimension (Olympus) and Image J software.

4.1.4 *In vitro* scratch assay

Cells were seeded at 1×10^5 cells/well in a 12-well culture plate in an appropriate medium and grown until 95% confluence was achieved. Then, a sterile pipette tip (20 µL) was used to make a scratch in the confluent monolayers and the spent media was discarded. Subsequently, scratched monolayers were rinsed with sterile PBS and then treated with 400 nM recombinant fusion protein or an equivalent volume of glycerol buffer (vehicle control) or 20 nM vascular endothelial growth factor [VEGF (used as a positive control)] in an appropriate medium containing 5% FBS for 24-48 hours. Images were captured at 20x magnification using an inverted brightfield microscope (ZOE Fluorescent Cell Imager, Bio-Rad) at different time intervals. The images were exported in .tif file format for the quantification of the scratched area using ImageJ (1.48v) software. The migration percentage was calculated using the following formula.

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

4.1.5 Cell cycle analysis

Cells (2.5×10^5 cells/well) were seeded onto 12-well cell culture plates and grown in a complete growth medium. After 70% confluency was reached, the medium was replaced with a protein transduction medium (containing purified fusion protein or glycerol buffer), and cells were incubated for 72–80 hours (with the medium replenished every 24 hours with fresh protein transduction medium). Cells were harvested, fixed in 70% ice-cold ethanol, and treated with RNase (100 µg/mL) to remove RNA contamination. After RNase treatment, the cells were

stained with propidium iodide (50 ng/mL) for 30 minutes. At least 3×10^4 cells were analyzed with flow cytometry (BD Biosciences), and the cell distribution was analyzed with FCS Express 5 software.

4.1.6 Colony formation assay

SW620 cells were diluted in a complete growth medium, and 400 cells/well were plated in 12-well cell culture plates and incubated under standard cell culture conditions. After 48 hours of incubation, the medium was replaced with a protein transduction medium containing 5% FBS supplemented with rhTBX5-NTH or glycerol buffer and incubated for 10–12 days (with the medium replenished every 24 hours). Surviving colonies were counted in Image J software after being fixed with methanol/acetic acid (3:1) and stained with 5% crystal violet. Experiments were performed in triplicates.

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

A total of 5×10^5 SW620 cells/well were plated onto six-well culture plates and grown to 70% confluency. Subsequently, the cells were incubated in protein transduction medium (2% serum supplemented with rhTBX5-NTH or glycerol buffer) for 72–80 hours (with medium replenished every 24 hours). Total RNA was isolated from the control and protein transduced SW620 cells with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Total RNA (~1 μg) was reverse transcribed with iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. Template cDNA was amplified with gene-specific primers (Table 4.1) with an AriaMx Real-Time PCR System (Agilent). Reactions contained iTaq Universal SYBR Green Supermix (2X concentration). Each sample was

analyzed in triplicate, and GAPDH served as an internal control. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method, with normalization to GAPDH as a reference gene.

Table 4.1 Primers used for RT-qPCR assay

Genes	Primer Sequences (5' – 3')	AT (°C)	PS (bp)
MTSS1	F - CCATCATCAGCGACATGAAG R - TGTGTTGGTGGCCATGTCAG	57	152
MTA2	F - TCTCACAGACCGGCAGATCG R - TGGCCTTAGCCAGGTCGTAG	63	192
GAPDH	F - GTCTCCTCTGACTTCAACAGCG R - ACCACCCTGTTGCTGTAGCCAA	57	131

AT, annealing temperature; PS, product size; F, forward; R, reverse; bp, base pair

4.1.8 Chicken Chorioallantoic Membrane (CAM) assay

The CAM assay was performed as described previously (D et al. 2016; Seetaraman Amritha et al. 2020) with slight modifications. The 3-4 days old embryonated chicken eggs were directly obtained from the local chicken egg hatching unit. Chicken eggshells were gently cut open at the top like a window to expose the CAM layer. Paper discs soaked in the purified recombinant fusion protein (400 nM) or glycerol buffer (equal volume) or VEGF (200 nM) were placed directly on the blood vessel (one disc/egg CAM layer) and sealed with an adhesive tape. Sealed eggs were then incubated for 12 hours at 37 °C with 60% humidity. The exposed blood vessels were visualized under LCD Digital Stereomicroscope (2X) equipped with a 2 MP camera, and images were captured at a particular interval of time, followed by documentation with the help of Adobe Photoshop CC 2019 software.

4.1.9 Lentivirus production

Second-generation lentiviral vectors were packaged in HEK293T cells using the calcium phosphate transfection method. Briefly, 6×10^6 HEK293T cells were seeded in a 100 mm gelatin-coated dish and then grown to 65-75% confluence in the complete growth medium. Once the culture attained the desired confluence, the medium was replaced with a transfection medium containing advanced DMEM supplemented with 2% FBS, 1% P/S, and 1X non-essential amino acids, 1-2 hours prior to transfection. The calcium phosphate transfection mixture was then prepared by mixing 12 µg lentiviral backbone plasmid [α-MHC-eGFP (Addgene plasmid: 21229) or FUdeltaGW-rtTA (Addgene plasmid: 19780) or tetO-MEF2C (Addgene plasmid: 46031)], 7.7 µg psPAX2 (Addgene plasmid: 12260), 4.3 µg pMD2.G (Addgene plasmid: 12259), 125 mM $CaCl_2$, and makeup with tissue-culture grade water with 2X HBS buffer (1:1), and incubated for 15 minutes at room temperature. This mixture was then added to the cells dropwise in the presence of 25 µM chloroquine. After 5-6 hours of transfection, the medium was replaced with a virus production medium containing advanced DMEM supplemented with 5% FBS, 1x P/S, and 1x NEAA. The medium was then replaced with a fresh medium after 12-16 hours post-transfection. The lentiviral supernatant was then harvested at 24, 36 and 48 hours of incubation, pooled, concentrated (centrifugation at 14000 rpm for 2 hours), and aliquoted for storage at -80 °C.

4.1.10 alpha-Myosin heavy chain (α-MHC) reporter assay

H9C2 cells were seeded in a gelatin-coated 12-well culture plate at 1 x 10^5 cells/well and grown until 60-70% confluence was achieved. Cells were then transduced with α-MHC-eGFP (~5 x 10^5 IFU) and FUdeltaGW-rtTA (~5 x 10^5 IFU) lentiviral vectors (Addis et al. 2013). On the next day (D-1), cells were infected with tetO-MEF2C (~5 x 10^5 IFU). Cells were transduced

only in the presence of 5 µg/mL of polybrene. On the following day (D0), cells were reseeded in the ratio of 1:3 in a fresh gelatin-coated 12-well culture plate. On Day 1, the spent medium was replaced with a 5% FBS growth medium supplemented with 2 µg/mL of doxycycline (Sigma). The next day (D2), cells were washed with PBS and then treated with a protein transduction medium containing 400 nM of rhHAND2 fusion protein or an equivalent volume of vehicle control supplemented with 2 µg/mL of doxycycline. The medium was renewed every alternative day. After the treatment, images were captured in the green channel at 20x magnification using an inverted fluorescent microscope (ZOE Fluorescent Cell Imager, Bio-Rad). For flow cytometry analysis, cells were harvested from culture dishes and washed with PBS, followed by fixation with 4% paraformaldehyde for 15 minutes. Fixed cells were washed and resuspended in PBS and then analyzed using a BD FACS Calibur Flow Cytometer (BD Biosciences) with FlowJo software.

4.1.11 Western blotting

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

4.1.12 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 and transduction ability of purified recombinant fusion proteins

To test the cell permeability of the purified recombinant fusion proteins, we first examined their stability under standard cell culture conditions. The solubility and stability of recombinant proteins under cell culture conditions play a crucial role in developing efficient and robust protein transduction system (Bosnali and Edenhofer 2008; Thier et al. 2010; Dey et al. 2021a). Thus, *in vitro* cell culture, solubility and stability were analyzed for the purified recombinant fusion proteins using Western blotting with protein-specific antibodies. From the result (Figure 4.1A), it is clear that 400 nM rhGATA4-NTH protein was soluble and stable under standard cell culture conditions after 24 hours of incubation. However, after 48 hours, the stability of rhGATA4-NTH protein considerably decreased compared to 0 and 24 hours of incubation. This signifies that rhGATA4 fusion protein was stable for up to 24 hours at standard cell culture conditions. In the case of rhMEF2C-NTH protein, the purified version was stable for up to 48 hours of incubation under cell culture conditions (Figure 4.1A). Though this fusion protein was stable, a gradual decrease in the stability was observed over time.

Subsequently, we analyzed rhTBX5-NTH fusion protein stability under cell culture conditions, and the results revealed that this fusion protein was partially stable (~25%, quantified with ImageJ) after 24 and 48 hours of incubation (Figure 4.1A and 4.1B). Therefore, we further analyzed its stability every 6 hours for up to 24 hours and found that the rhTBX5-NTH protein was stable (>50–75%, quantified with ImageJ) for 12–18 hours under standard cell culture conditions (Figure 4.1C). A gradual decrease in the stability of rhTBX5-NTH protein over time was observed. Similarly, the rhHTN-ETS2 fusion protein was substantially

stable for up to 24 hours and partially stable after 48 hours of incubation under cell culture conditions (Figure 4.1A).

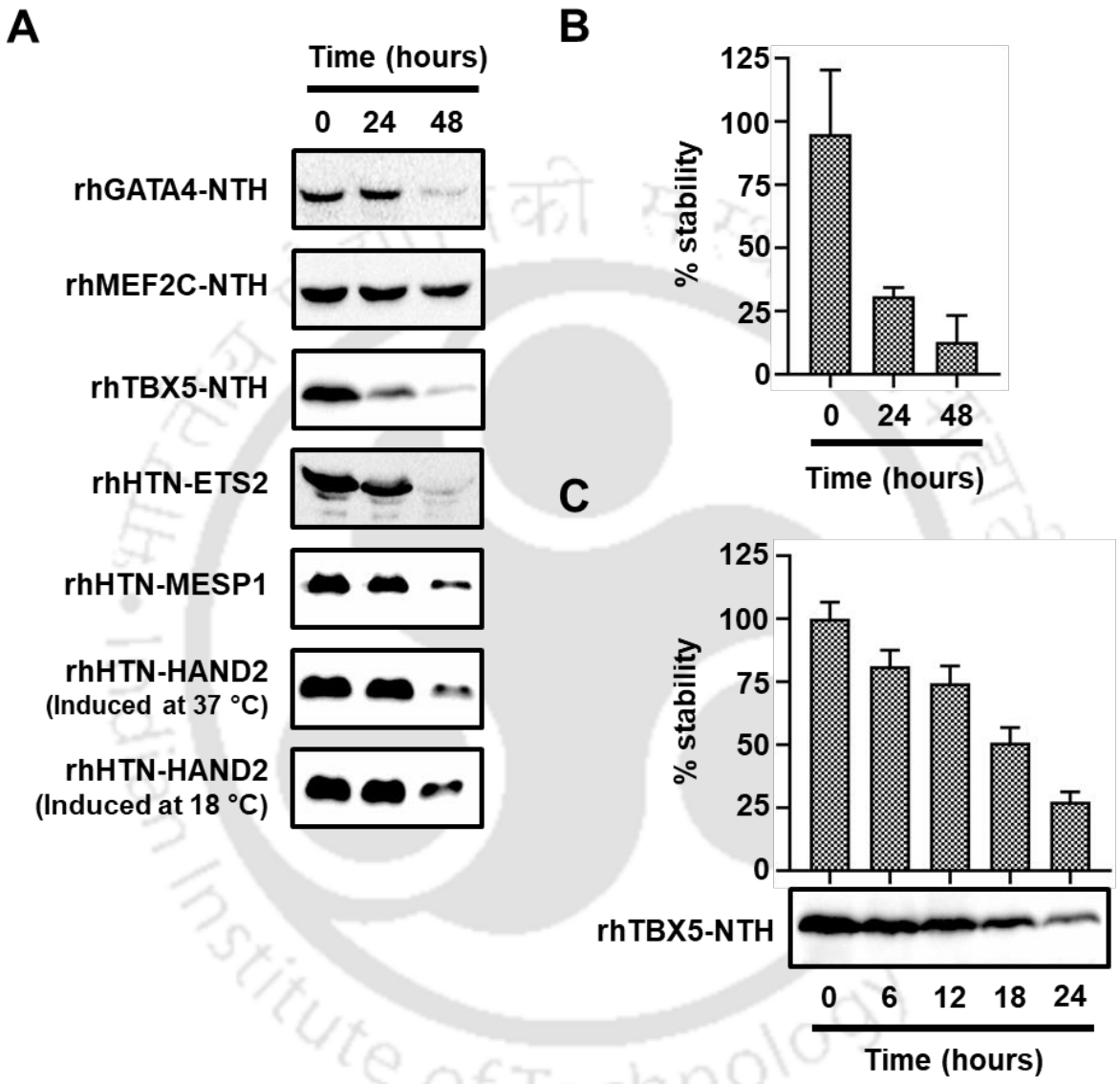


Figure 4.1 Stability analysis of recombinant fusion proteins under standard cell culture conditions. Purified fusion proteins were diluted in transduction medium (low serum) and incubated under standard cell culture conditions for 48 hours. (**A**) Stability analysis was performed every 24 hours up to 48 hours under cell culture conditions (n≥3). (**B**) Quantitative analysis of rhTBX5-NTH protein stability (from **A**). (**C**) Stability analysis of rhTBX5-NTH protein, which was performed every 6 hours

up to 24 hours under cell culture conditions [*top*; quantitative analysis and *bottom*; Western blotting analysis (n=2)]. The quantitative data shown are mean ± SE.

We next sought to examine the stability of rhHTN-MESP1 protein under standard cell culture conditions. Western blotting analysis showed that after 24 hours of incubation, this purified rhHTN-MESP1 protein was soluble and stable, and the intensity was much similar to that of the 0^{th}-hour sample (Figure 4.1A). However, the stability of this purified fusion protein decreased considerably after 24 hours of incubation. Likewise, the rhHTN-HAND2 fusion protein was stable for up to 24 hours of incubation, irrespective of the induction temperature (Figure 4.1A). After 48 hours, a faint band was observed in the Western blotting analysis, indicating this fusion protein is partially stable over time.

Protein-based reprogramming approaches generally require several rounds of protein transduction, with one round for 4–24 hours per day (Bosnali and Edenhofer 2008; Islas et al. 2012; Borgohain et al. 2019; Dey et al. 2021a). Specifically, in direct cardiac reprogramming, 4 hours of protein treatment per day is sufficient to induce a cardiac transcriptional profile in human fibroblasts (Islas et al. 2012). In this aspect, our purified recombinant fusion proteins passed the minimum criteria required for reprogramming in terms of solubility and stability under cell culture conditions. Overall these purified proteins are soluble and stable for up to at least 24 hours. Thus, our results confirmed that these purified recombinant fusion proteins had suitable solubility and stability under cell culture conditions for cellular reprogramming and other cell culture analysis.

Cell penetration and nuclear translocation is also a critical criterion for transcription factor functionality, as its major role lies inside the nucleus of the cell. Therefore, we next sought to determine whether these purified and stable proteins have subcellular and subnuclear

localization ability. The human cells were treated with these purified proteins or glycerol buffer (as negative and vehicle controls) for 6 hours and immunocytochemistry and fluorescence microscopy were performed. Human fibroblasts and HeLa cells were selected for the study because of their differing protein transduction efficiency (Weill et al. 2008) and also, they lack endogenous expression of these transcription factors. Fluorescence microscopy and flow cytometry have indicated that the protein transduction efficiency varies from 60% in human fibroblasts (primary) to >75% in HeLa cells (Weill et al., 2008). Additionally, fibroblasts are widely used for cell reprogramming experiments (Borgohain et al. 2019; Dey et al. 2021a) and HeLa cells are widely used for protein transduction analysis experiments (Dey et al. 2021b). At first, the characterization of human BJ fibroblasts (HFFs) with immunocytochemistry was performed, and the result showed that these cells express Vimentin, a fibroblast-specific marker (Figure 4.2).

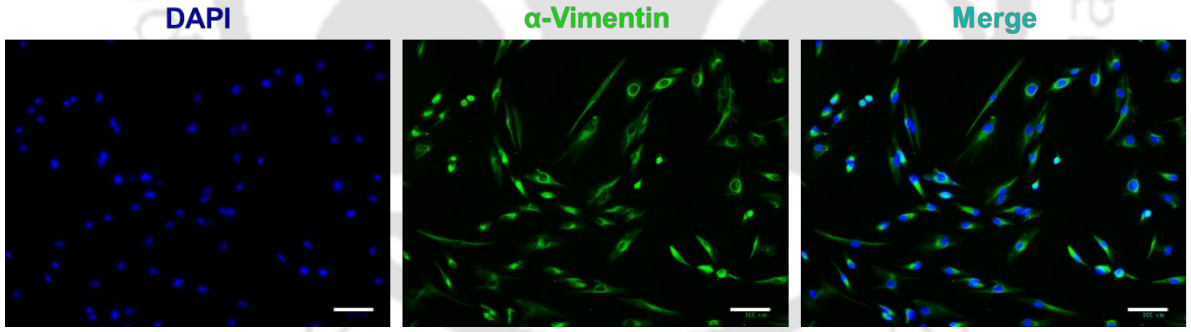


Figure 4.2 Characterization of BJ fibroblasts using the Vimentin antibody. Nuclear staining was carried out with a DAPI solution. Scale bar: 100 μm.

In order to study the transduction ability of rhGATA4-NTH protein into human cells, we first confirmed that the glycerol buffer (vehicle control) did not stimulate the expression of GATA4 in HeLa cells and HFFs or did not lead to any false positive signal during analysis (Figure 4.3; *Vehicle control panels*). The TAT-mediated protein transduction resulted in

efficient transduction of rhGATA4 fusion protein into human cells (Figure 4.3). Interestingly, almost all the cells showed successful uptake of purified rhGATA4-NTH protein into their cytoplasm. Of these, majority of the cells showed NLS-mediated nuclear localization of rhGATA4 fusion protein (Figure 4.3).

In the case of rhMEF2C-NTH protein, the fluorescence microscopy analysis revealed that the TAT and NLS fused purified rhMEF2C fusion protein transduced into the HeLa cells and HFFs and translocated to their nucleus (Figure 4.4). Further, the microscopy results showed differences in the nuclear translocation of this rhMEF2C fusion protein between these two human cell lines. As shown in the vehicle control panels, our results confirmed that these cell lines did not express the MEF2C transcription factor and further ruled out the possibility of glycerol buffer mediated endogenous expression of MEF2C or false positive signals during analysis (Figure 4.4; *Vehicle control panels*). Thus, our fusion strategy successfully mediated the purified rhMEF2C-NTH protein to cross the subcellular and subnuclear regions of the mammalian cells.

With immunocytochemistry, we confirmed that the HeLa, HFF and SW620 cells did not express TBX5 endogenously as shown in the vehicle control panels (Figure 4.5). The results demonstrated that this pure rhTBX5-NTH protein penetrated the cell membrane and also translocated into the nucleus in these cell lines, without a need for any additional transduction reagent (Figure 4.5). Also, we did not observe any difference in rhTBX5-NTH protein transduction efficiency in these cell lines. Thus, in HeLa, HFF and SW620 cells, the generated recombinant protein showed efficient cell penetration and nuclear translocation ability, mediated by TAT and NLS. Similarly, we confirmed the absence of endogenous expression of ETS2 in HeLa cells using immunocytochemistry, as shown in the vehicle control

panel (Figure 4.6). Microscopy results revealed that the TAT and NLS fused purified rhETS2 fusion protein transduced the HeLa cells and then translocated to its nucleus (Figure 4.6).

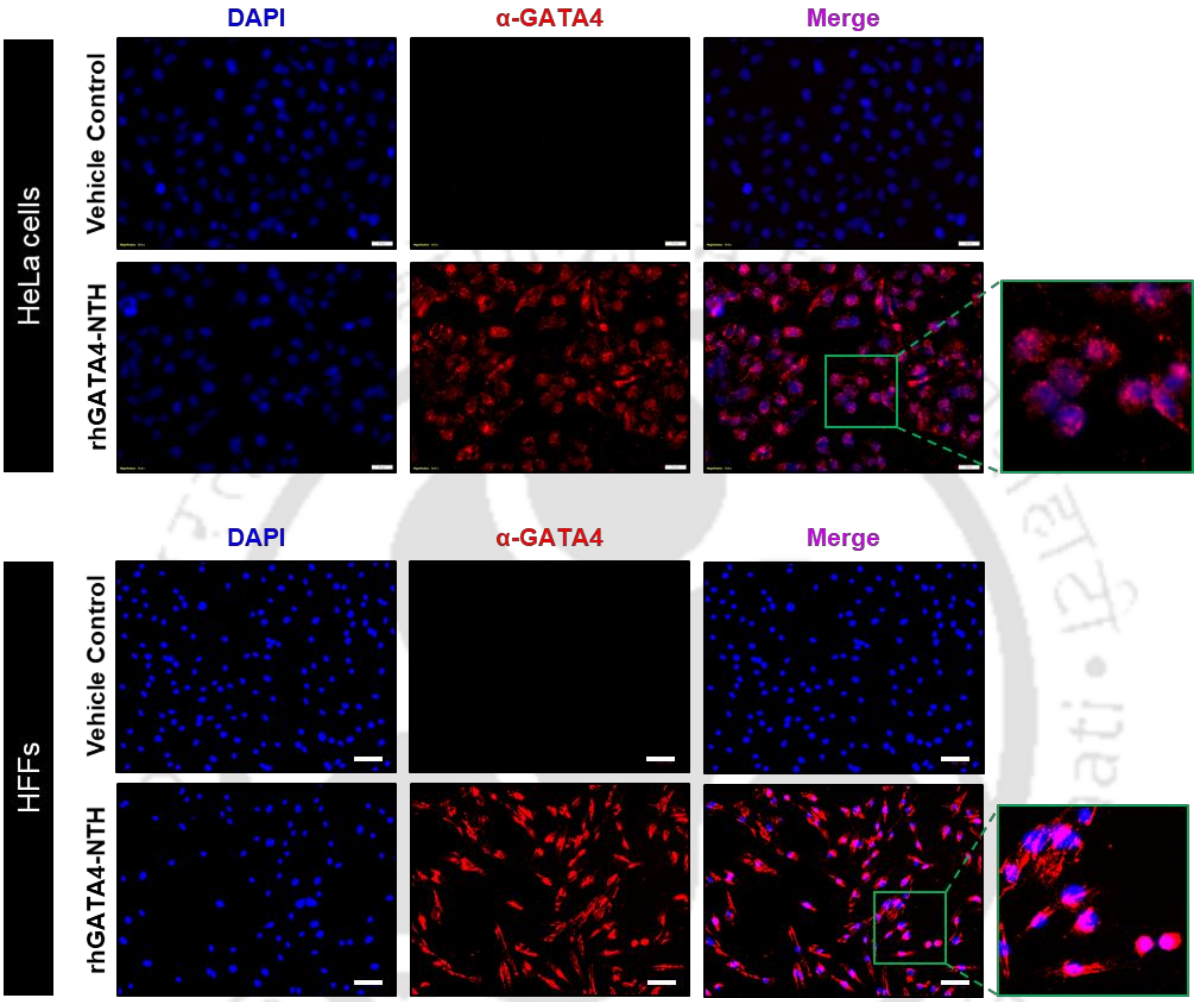


Figure 4.3 Subcellular and subnuclear delivery of purified rhGATA4-NTH protein into human cells. HeLa and HFF cells were exposed to 400 nM of purified rhGATA4-NTH protein for 4-6 hours under standard cell culture conditions. After protein transduction, cells were fixed, permeabilized, blocked, and then stained with an α-GATA4 antibody. Transduced cells were detected with an Alexa Fluor® 594 conjugated α-mouse secondary antibody. Nuclei were stained with DAPI, and images were taken at 20x magnification. All images were taken with identical camera settings. Scale bar: 50 µm for HeLa cells and 100 µm for HFFs (n=3).

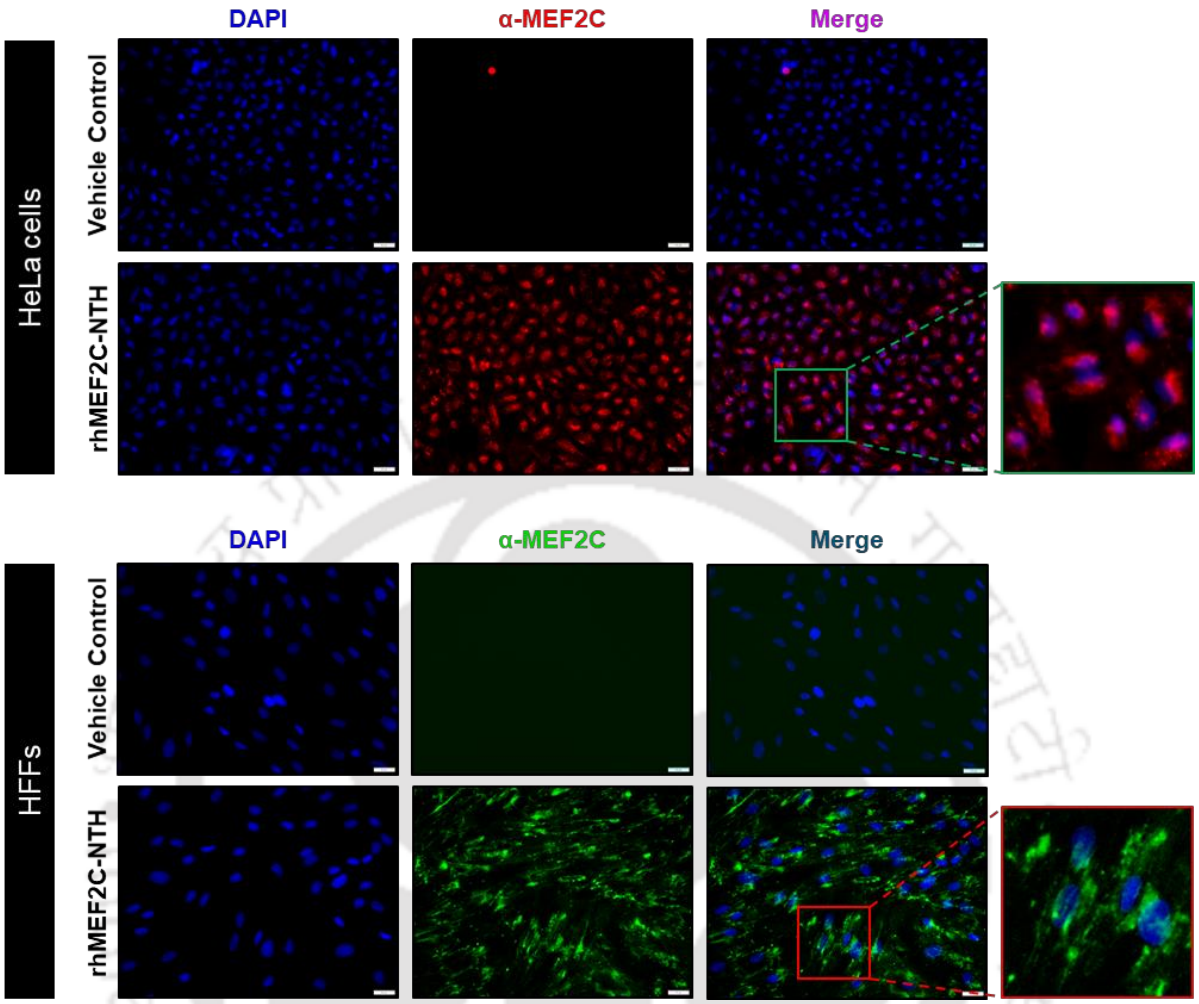


Figure 4.4 Subcellular and subnuclear delivery of purified rhMEF2C-NTH protein into human cells. HeLa and HFF cells were exposed to 400 nM of purified rhMEF2C-NTH protein for 4-6 hours under standard cell culture conditions. After protein transduction, cells were fixed, permeabilized, blocked, and then stained with an α-Mef2c antibody. Transduced cells were detected with an Alexa Fluor® 594/488 conjugated α-mouse secondary antibody. Nuclei were stained with DAPI, and images were taken at 20x magnification. All images were taken with identical camera settings. Scale bar: 50 μm (n=3).

The NLS-mediated nuclear translocation resulted in efficient delivery of rhETS2 fusion protein into the nucleus of the cells. Thus, our fusion strategy validated the cell penetration and nuclear translocation ability of this purified recombinant fusion protein.

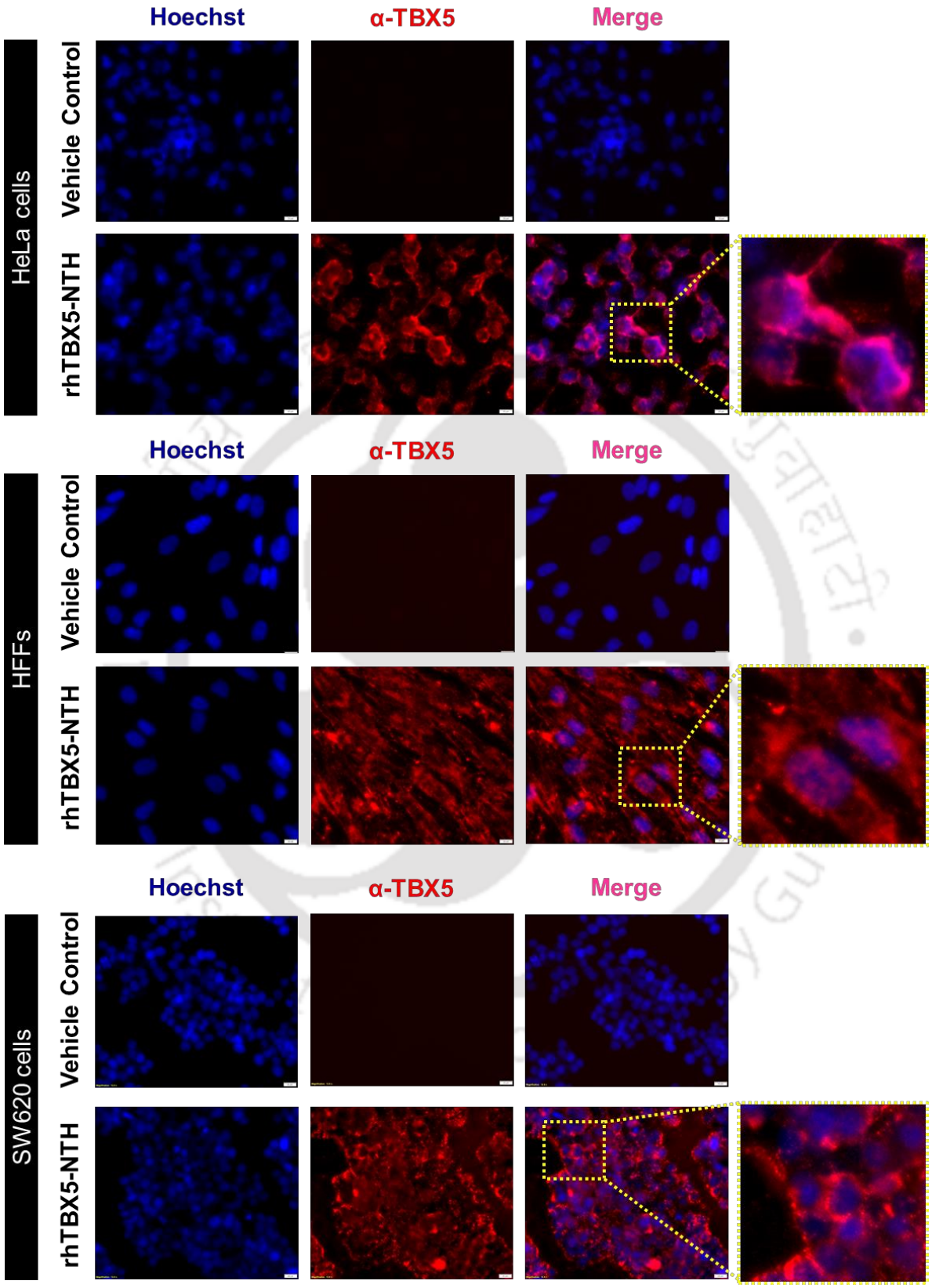

Hoechst
α-TBX5
Merge
HeLa cells
Vehicle Control
rhTBX5-NTH
Hoechst
α-TBX5
Merge
HFFs
Vehicle Control
rhTBX5-NTH
Hoechst
α-TBX5
Merge
SW620 cells
Vehicle Control
rhTBX5-NTH

Figure 4.5 Subcellular and subnuclear delivery of purified rhTBX5-NTH protein into human cells. HeLa, HFF, and SW620 cells were treated with 400 nM of purified rhTBX5- NTH protein for 4-6 hours, then examined with immunostaining (stained with α-Tbx5 and detected with Alexa Fluor® 594 conjugated α-mouse secondary antibodies) under a fluorescence microscope. Nuclei were stained with Hoechst, and images were taken at 20x magnification. All images were taken with identical camera settings. Scale bar: 20 μm (n = 2).

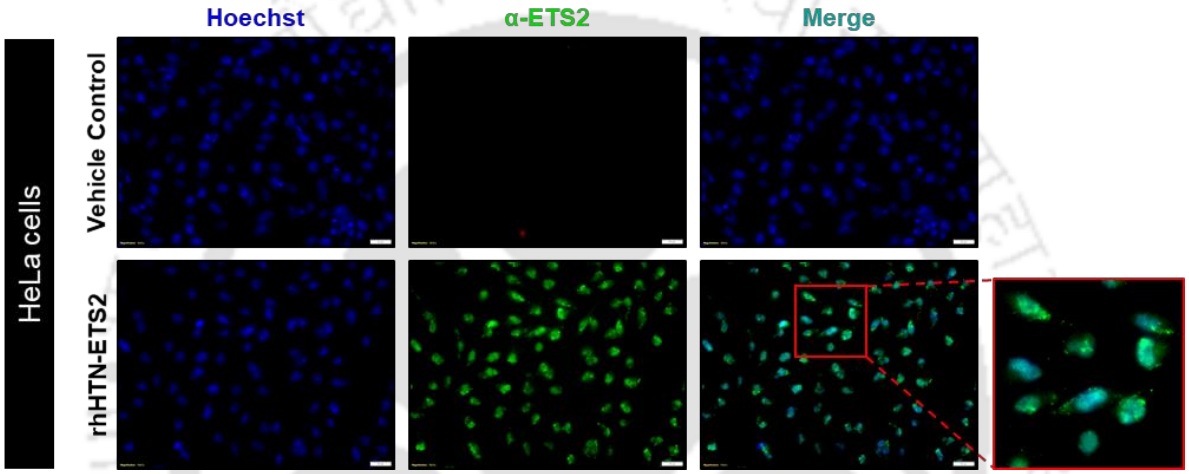


Figure 4.6 Subcellular and subnuclear delivery of purified rhHTN-ETS2 protein into human cells. HeLa cells were treated with 400 nM of purified rhHTN-ETS2 protein for 4-6 hours, then examined with immunostaining (stained with α-Ets2 and detected with Alexa Fluor® 488 conjugated α-mouse secondary antibodies) under a fluorescence microscope. Nuclei were stained with Hoechst, and images were taken at 20x magnification. All images were taken with identical camera settings. Scale bar: 50 μm (n = 2).

We next investigated the transduction ability of purified rhMESP1 fusion protein in human cells. Firstly, we confirmed that HeLa and HFF cells does not express MESP1 protein endogenously as shown in Figure 4.7 (*Vehicle control panels*). These cells were then exposed with rhHTN-MESP1 protein for 4 hours, and the fluorescence microscopy analysis after immunostaining revealed the efficient transduction and translocation of this purified fusion

protein to the target site of the cell (Figure 4.7). Similarly, as shown in Figure 4.8, the

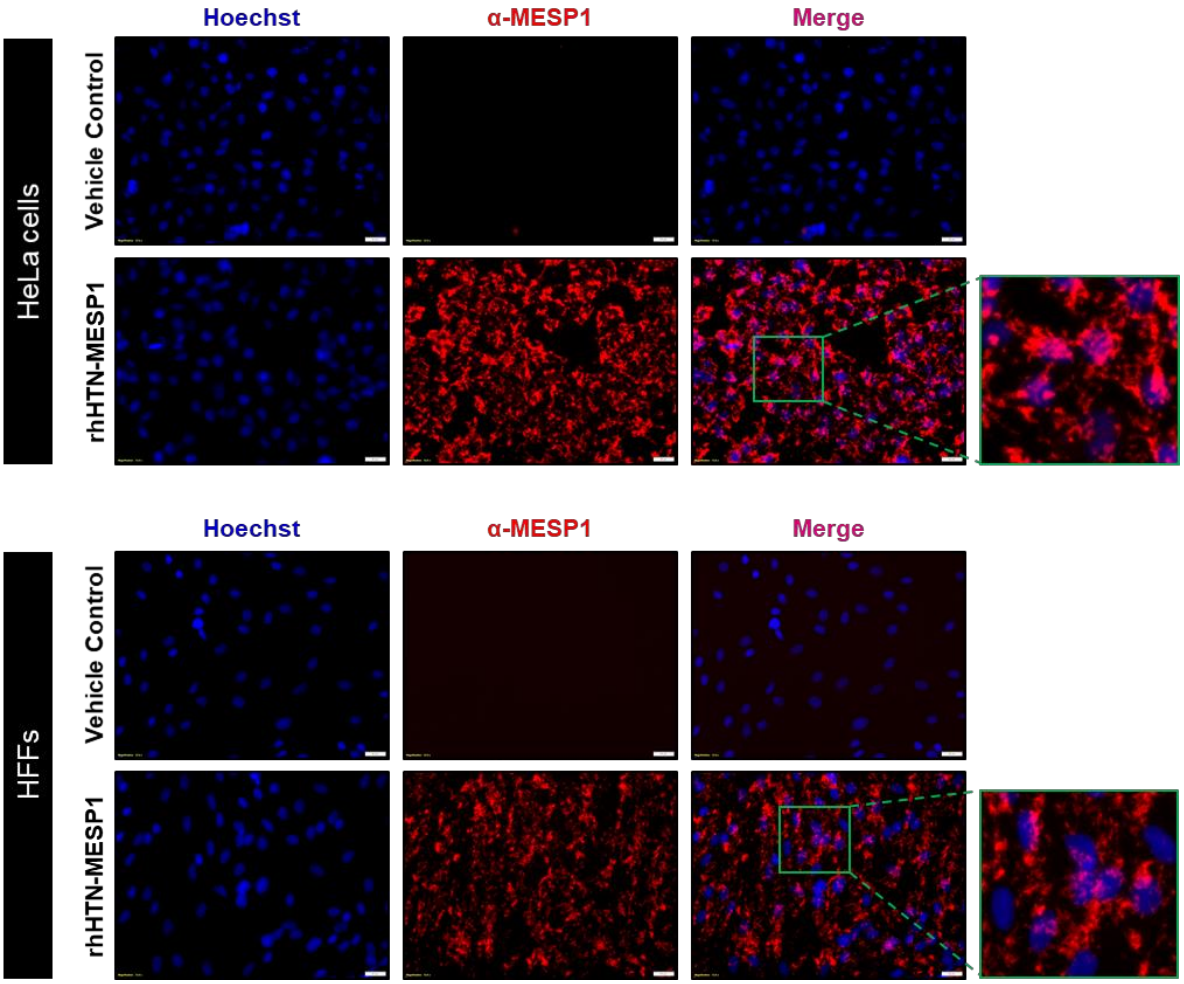


Figure 4.7 Subcellular and subnuclear delivery of purified rhHTN-MESP1 protein into human cells. HeLa and HFF cells were treated with 400 nM of purified rhHTN-MESP1 protein for 4-6 hours, then examined with immunostaining (stained with α-MESP1 and detected with Alexa Fluor® 594 conjugated α-mouse secondary antibodies) under a fluorescence microscope. Nuclei were stained with Hoechst, and images were taken at 20x magnification. All images were taken with identical camera settings. Scale bar: 50 μm (n = 2).

transduction of rhHAND2 fusion proteins across the subcellular and subnuclear regions of both cell lines was observed irrespective of their induction temperature. The microscopy results further confirmed the absence of endogenous expression of HAND2 in HeLa cells and HFFs

as shown in the vehicle control panel, and this also signified that the glycerol buffer did not

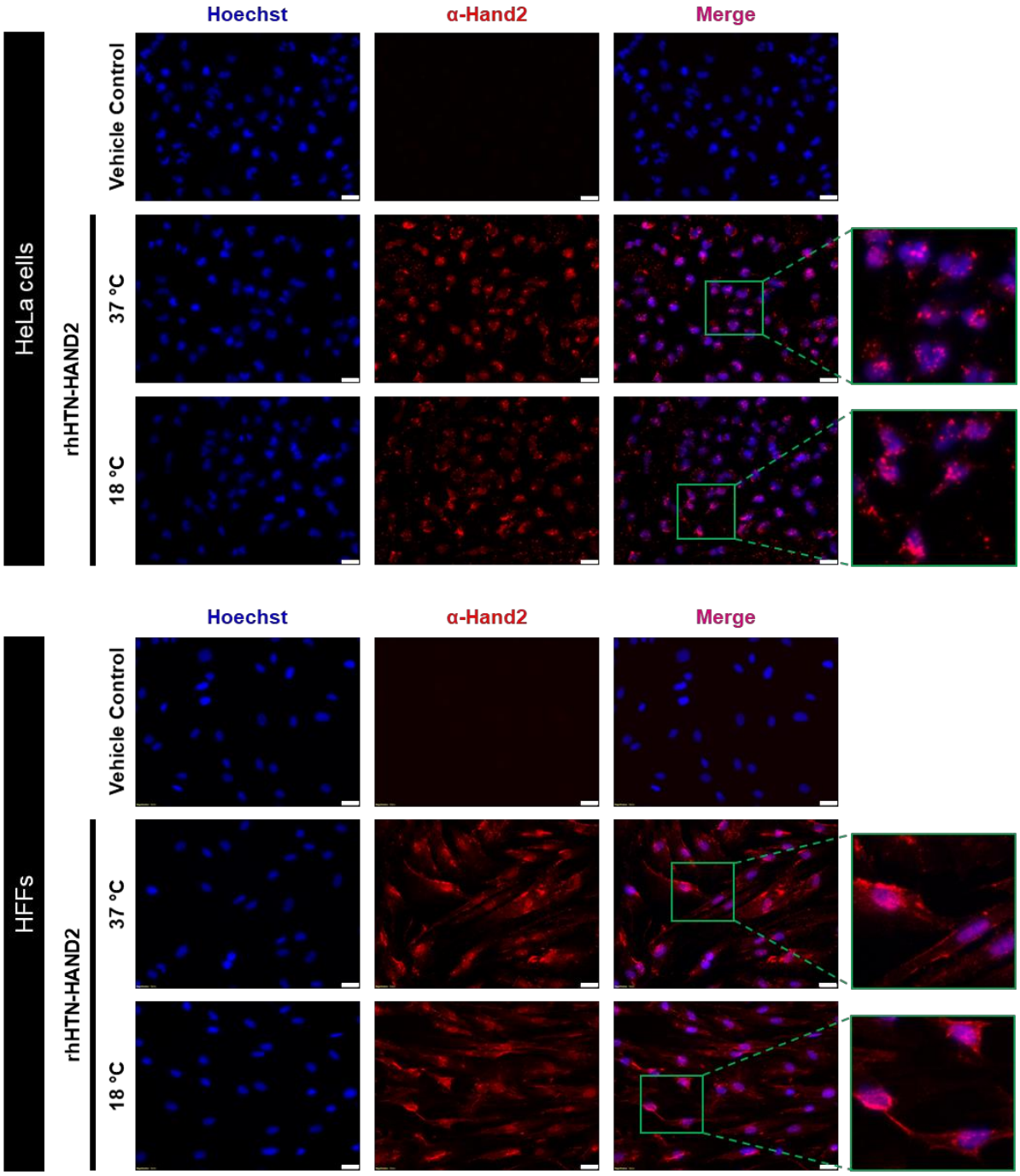


Figure 4.8 Subcellular and subnuclear delivery of purified rhHTN-HAND2 protein into human cells. HeLa and HFF cells were exposed to 400 nM of purified rhHTN-HAND2 protein for 4-6 hours under standard cell culture conditions. After protein transduction, cells were fixed,

permeabilized, blocked, and then stained with an α-Hand2 antibody. Transduced cells were detected with an Alexa Fluor® 594 conjugated α-rabbit secondary antibody. Nuclei were stained with Hoechst, and images were taken at 20x magnification. All images were taken with identical camera settings. Scale bar: 50 μm for HeLa cells and 100 μm for HFFs (n=3).

trigger the expression of HAND2 or did not lead to any false positive signal during analysis (Figure 4.8; *Vehicle control panels*). Thus, both the purified rhHTN-MESP1 and rhHTN-HAND2 (irrespective of the induction temperature in which they were expressed) fusion proteins showed cell penetration (mediated by TAT) and nuclear translocation (mediated by NLS) abilities, indicating the functionality of these tags post-purification.

Our strategy of coupling these transcription factors with the protein transduction domain (TAT) and NLS facilitated their delivery into the mammalian cell target site without any requirement of protein transduction reagent. As transcription factors, they have their major role inside the nucleus; therefore, their nuclear entry is one of the most critical aspects of their functionality. Similar fusion strategies were employed in previous studies for the efficient subcellular and subnuclear delivery of reprogramming factors in the form of recombinant proteins in mammalian cells (Bosnali and Edenhofer 2008; Stock et al. 2010; Thier et al. 2010; Thier et al. 2012; Peitz et al. 2014; Münst et al. 2016; Dey et al. 2021b; Narayan et al. 2021a; Narayan et al. 2021b). Indeed, in the majority of these studies, the fusion tags do not affect the functionality of the desired proteins to which they were fused (Bosnali and Edenhofer 2008; Stock et al. 2010; Thier et al. 2010; Thier et al. 2012; Peitz et al. 2014; Münst et al. 2016; Dey et al. 2021b; Narayan et al. 2021b) , and the biological activity of these proteins was comparable to their genetic counterparts (Stock et al. 2010; Thier et al. 2010; Thier et al. 2012). Henceforth, these generated recombinant fusion proteins can replace their genetic form to

derive integration-free cells and also for other biomedical applications. Thus, eliminating the problems associated with the plasmid or viral-based approaches (Islas et al. 2012; Borgohain et al. 2019; Haridhasapavalan et al. 2019; Dey et al. 2021a).

4.2.2 Functional assessment of purified recombinant fusion proteins

Although these purified fusion proteins had retained their secondary structure and demonstrated cell penetration and nuclear translocation, further examination was required to corroborate their functional activity. Therefore, to validate the biological activity of these fusion proteins, we investigated their effect on the mammalian cells.

4.2.2.1 Effects of rhGATA4, rhETS2, and rhMESP1 fusion proteins on breast cancer cells

To examine whether the purified rhGATA4, rhETS2 and rhMESP1 fusion proteins are biologically active, we first analyzed their effect on the migration ability of breast cancer cells. For this analysis, we performed the most commonly used *in vitro* scratch assay (Dey et al. 2021b; Dey et al. 2022) to understand the impact of these purified fusion proteins on MDA-MB-231 breast cancer cell migration. Previously, a study demonstrated that overexpression of GATA4 decreased the proliferation, viability, migration and invasion capacities of breast cancer cells (Han et al. 2019). In congruence with the previous report, we also observed significantly ($P = 0.0075$) reduced migration potential of MDA-MB-231 cells upon treatment with rhGATA4-NTH protein (Figure 4.9A and 4.9B). Contrastingly, rhHTN-ETS2 and rhHTN-MESP1 protein significantly ($P<0.0001$ and 0.0053, respectively) increased the migration potential of breast cancer cells (Figure 4.9A and 4.9B). In point of fact, after 48 hours of incubation, rhETS2 and rhMESP1 fusion proteins treated MDA-MB-231 cells migrated >80 and >65% of the scratched area, respectively, whereas only <30 and <50%

scratched area was covered upon treatment with rhGATA4 fusion protein and vehicle control treated cells, respectively (Figure 4.9A and 4.9B). These results correlate with those from a

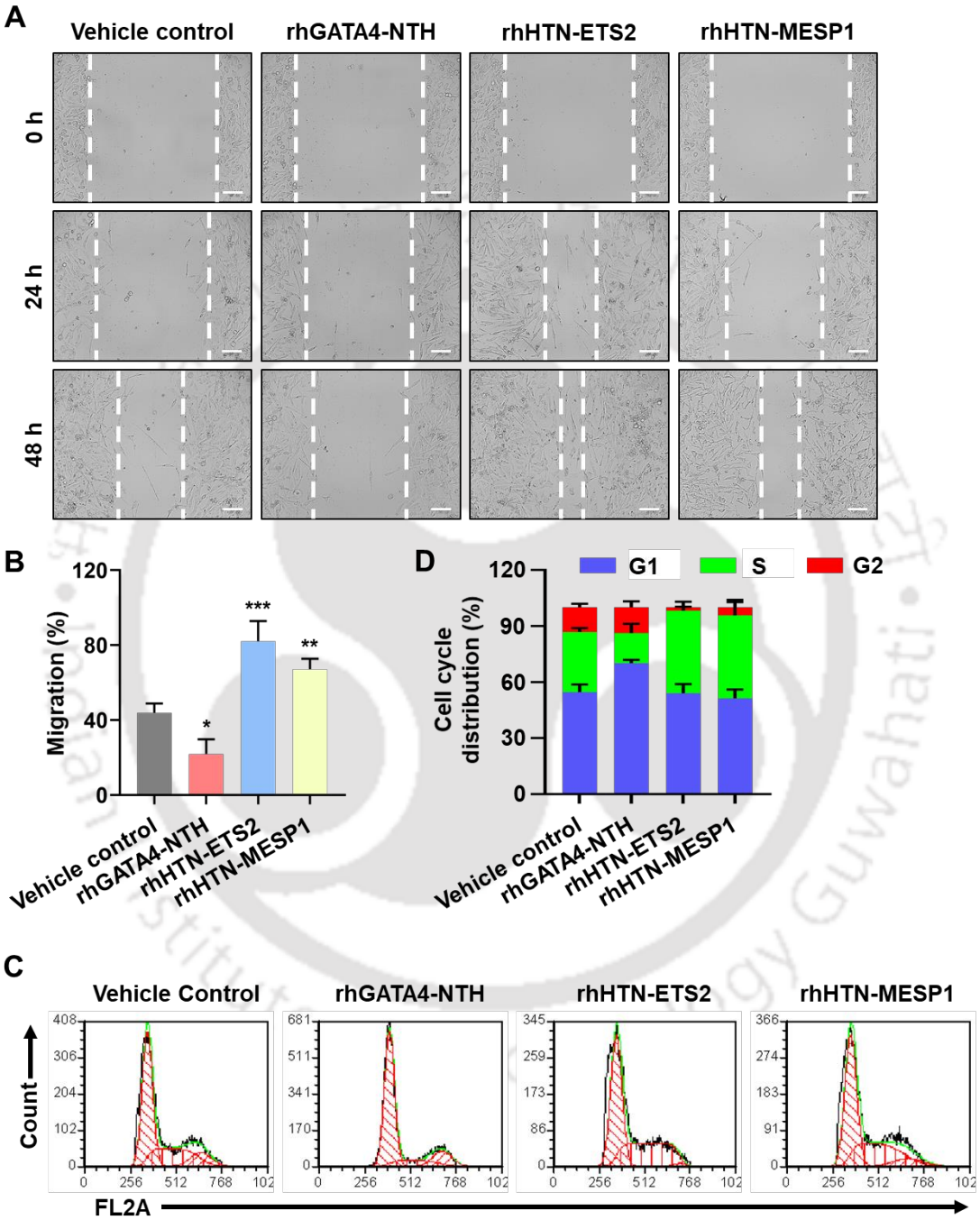


Figure 4.9 Effect of purified rhGATA4, rhETS2 and rhMESP1 fusion proteins on breast cancer cells. (**A**) Microscopy analysis of the migration potential of MDA-MB-231 cells upon

treatment with these recombinant fusion proteins or vehicle control. The scale bar indicates 100 μm. (**B**) The quantitative representation of (**A**). (**C**) Representative cell cycle profiles of these fusion proteins transduced and non-transduced (vehicle control) cells. (**D**) Quantitative cell cycle distribution data of (**C**). The quantitative data shown are mean ± SEM (n = 3). *$P < 0.05$; **$P < 0.001$, ***$P < 0.0001$.

previous report in which ETS2 has been found to contribute to the survival and telomerase activity of breast cancer cells (Dwyer and Liu 2010). Apart from breast cancer, the oncogenic role of ETS2 has been reported in prostate and blood cancers (Carbone et al. 2004; Ge et al. 2008). Similarly, MESP1 plays a crucial role in the initiation, proliferation and progression of lung cancer and was found as a novel biomarker in lung cancer cells (Tandon et al. 2019; Wang et al. 2020c).

We next asked whether these purified fusion proteins alter the cell cycle of the MDA-MB-231 cells. Flow cytometry analysis revealed that rhGATA4 fusion protein decreased the S phase and increased the G1 and G2 phases of the cell cycle upon transduction in the breast cancer cells (Figure 4.9C and 4.9D). These results were consistent with the previously published studies on breast and lung cancers (Gao et al. 2019; Han et al. 2019). On the contrary, rhETS2 and rhMESP1 fusion proteins increased the S phase and decreased the G1 and G2 phases compared to the control cells (Figure 4.9C and 4.9D), in accordance to an earlier study with ectopic overexpression on MESP1 in lung cancer cells (Wang et al. 2020c). Thus, we confirmed that these fusion proteins regulate the cell cycle machinery of breast cancer cells.

The role of GATA4 in cancer is contradictory, and it depends on the type and subtype of cancer. In breast and lung cancer, GATA4 has been identified as a tumor suppressor, as it regulates NF-κB (Han et al. 2019) and TGF-β (Gao et al. 2019) signaling pathways,

respectively. It also decreased the epithelial-mesenchymal transition in breast cancer cells (Han et al. 2019). On the other hand, ETS2 and MESP1 transcription factors were identified as potent oncogenes in acute megakaryocytic leukemia (Ge et al. 2008) and lung cancer cells (Tandon et al. 2019; Wang et al. 2020c), respectively. MESP1 prevented apoptosis in lung cancer cells through the caspase3/poly(ADP-ribose)polymerase 1 (PARP1) signaling pathway (Wang et al. 2020c). Its overexpression induced oncogene stress in the presence of ARF, a potent tumor suppressor gene, and increased cell proliferation and progression in the absence of ARF (Tandon et al. 2019). It also regulated epithelial-mesenchymal transition in lung cancer cells (Wang et al. 2020c), unlike the effect of GATA4 in lung cancer (Gao et al. 2019). Based on these studies, we can suggest that rhGATA4 fusion protein reduced breast cancer cell migration and induced cell cycle arrest at G1 phase, most likely by decreasing epithelial-mesenchymal transition and through the NF-κB signaling pathway. On the other hand, rhETS2 fusion protein enhanced the migration of MDA-MB-231 cells and altered the cell cycle, most likely by activating the telomerase activity. Likewise, the rhMESP1 fusion protein effect on MDA-MB-231 cells might be due to an increase in the epithelial-mesenchymal transition and cooperation with the loss of ARF. Our results and previous studies propose these transcription factors as novel biomarkers and therapeutic targets for various cancers (Ge et al. 2008; Gao et al. 2019; Han et al. 2019; Tandon et al. 2019; Wang et al. 2020c). Thus, these results confirm that these purified fusion proteins are biologically active.

4.2.2.2 Effects of rhMEF2Cand rhHAND2 fusion proteins on endothelial cell migration

Previously, studies have reported that MEF2C (Xu et al. 2012; Sturtzel et al. 2014) and HAND2 (VanDusen et al. 2014; George and Firulli 2019) transcription factors play a crucial role in regulating angiogenesis. Following this, we then assessed the functionality of these two

factors in their recombinant form on angiogenesis. Endothelial cell migration is one of the important processes of angiogenesis. Therefore, we first performed an *in vitro* scratch migration assay, a widely used assay to study the migration-inducing/inhibiting potential of any specific biological molecules, including recombinant proteins (D et al. 2016; Seetaraman Amritha et al. 2020). In that aspect, we observed that these two fusion proteins, upon transduction, exhibited opposite effects on endothelial cell migration. As shown in Figure 4.10A and 4.10B, we observed significantly (P = 0.0142) reduced migration potential of rhMEF2C-NTH protein transduced HUVECs compared with the non-transduced control cells. However, significantly (P < 0.0001 for protein-induced at 37 °C and P = 0.0013 for protein-induced at 18 °C) enhanced migration potential of rhHTN-HAND2 protein transduced HUVECs was observed compared to the control cells (Figure 4.10A and 4.10B). Similarly, HUVECs migration was significantly (P < 0.0001) increased in the presence of VEGF (Figure 4.10A and 4.10B), which was used as a positive control. After 12 hours of treatment, rhHAND2 fusion protein transduced HUVECs (irrespective of the induction temperature) and VEGF treated positive control cells migrated to >75% and >80% scratched area, respectively, whereas in rhMEF2C fusion protein transduced and non-transduced vehicle control cells only <35% and <50% area was covered, respectively (Figure 4.10A and 4.10B). Interestingly, rhHTN-HAND2 protein-induced at 37 °C migrated significantly (P = 0.0097) faster than that induced at 18 °C. The rhHTN-HAND2 protein-induced at 37 °C promoted endothelial cell migration similar (no significant difference) to that of VEGF treated ones (Figure 4.10B). Notably, MEF2C influence only migration and does not affect the proliferation of endothelial cells (Sturtzel et al. 2014). On the other hand, studies have reported the enhanced migration of endothelial cells by *E. coli*-derived recombinant human dermatopontin (Seetaraman Amritha

et al. 2020) and Brugia malayi asparaginyl-tRNA synthetase (D et al. 2016) proteins. These data thus confirm that our purified rhMEF2C and rhHAND2 fusion proteins significantly alter the migration potential of endothelial cells.

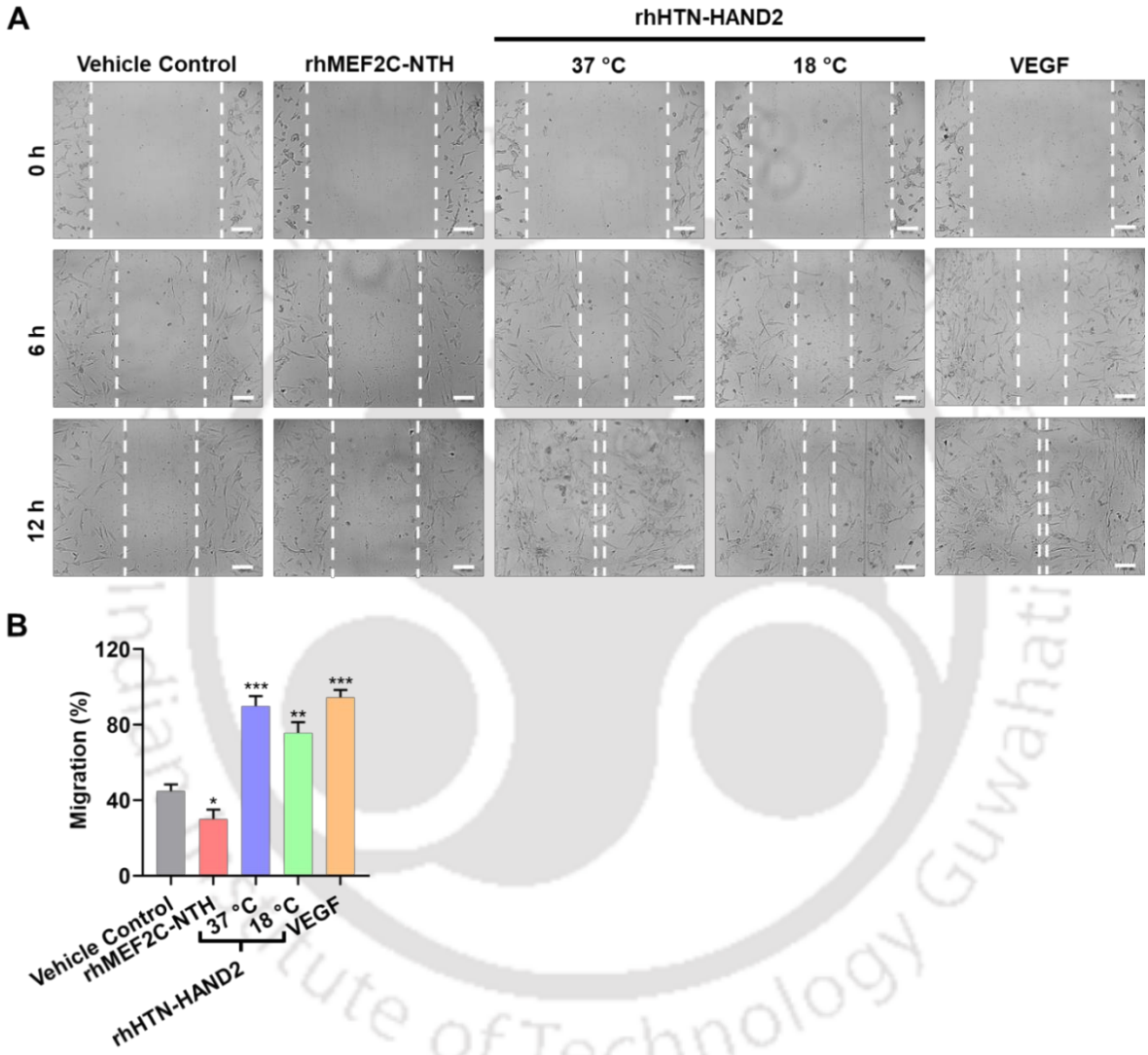


Figure 4.10 Effect of purified rhMEF2C and rhHAND2 fusion proteins on the migration of endothelial cells. (**A**) Microscopy analysis of the migration-inducing potential of rhMEF2C-NTH and rhHTN-HAND2 proteins transduced, non-transduced, and VEGF induced HUVECs. Scale bar: 100 µm. (**B**) Quantitative data of the migration in (**A**). The quantitative data shown are mean ± SEM (n = 3). **P*<0.05; ***P*<0.001, ****P*<0.0001.

4.2.2.3 Effects of rhMEF2Cand rhHAND2 fusion proteins on angiogenesis

We next sought to understand whether the purified rhMEF2Cand rhHAND2 fusion proteins can regulate angiogenesis as we obtained positive results in the preliminary experiments using endothelial cells. In order to study the angiogenic/anti-angiogenic potential of these fusion proteins, we performed a chicken CAM assay. The chick embryo model is a widely used *ex vivo* model to study the angiogenic/anti-angiogenic potential of various biological molecules including recombinant proteins (D et al. 2016; Seetaraman Amritha et al. 2020). For the slow diffusion of recombinant protein, we soaked the filter paper discs in the protein solution and gently placed them on the CAMs. At the end of 12 hours of incubation, we did not observe any sprouting blood vessels in the rhMEF2C-NTH protein diffused CAM, whereas increased neovascularization was observed in the rhHTN-HAND2 protein diffused CAM (Figure 4.11). Interestingly, as early as after 5-6 hours of incubation, induced sprouting of small capillaries from the pre-existing blood vessels was observed in this rhHAND2 fusion protein and VEGF-treated (positive control) CAMs compared to the vehicle control. No substantial difference was observed between this rhHAND2 fusion protein (irrespective of the induction temperature) and VEGF-treated CAMs (Figure 4.11). Contrastingly, even after 15 hours of incubation, no sign of sprouting blood vessels was observed in the rhMEF2C fusion protein diffused CAM. This inhibition of sprouting angiogenesis might be due to decreased migration and tube formation potential of endothelial cells mediated by the MEF2C transcription factor (Sturtzel et al. 2014). Consistent with this, we also observed the migration inhibition potential of purified rhMEF2C fusion protein on endothelial cells (Figure 4.10A and 4.10B). Although studies suggest that MEF2C binds to the Notch intracellular domain in cardiomyocytes (Wilson-Rawls et al. 1999; Pallavi et al. 2012), it does not regulate angiogenesis through the Notch signaling pathway

(Sturtzel et al. 2014), unlike HAND2 (VanDusen et al. 2014). Notably, MEF2C induced the expression of alpha-2-macroglobulin in endothelial cells, which then negatively regulated angiogenesis (Sturtzel et al. 2014). These results imply that the purified rhMEF2C-NTH protein inhibited the sprouting angiogenesis. This might be mediated through the induction of alpha-2-macroglobulin, which in turn inhibited the endothelial cells migration and tube formation potential.

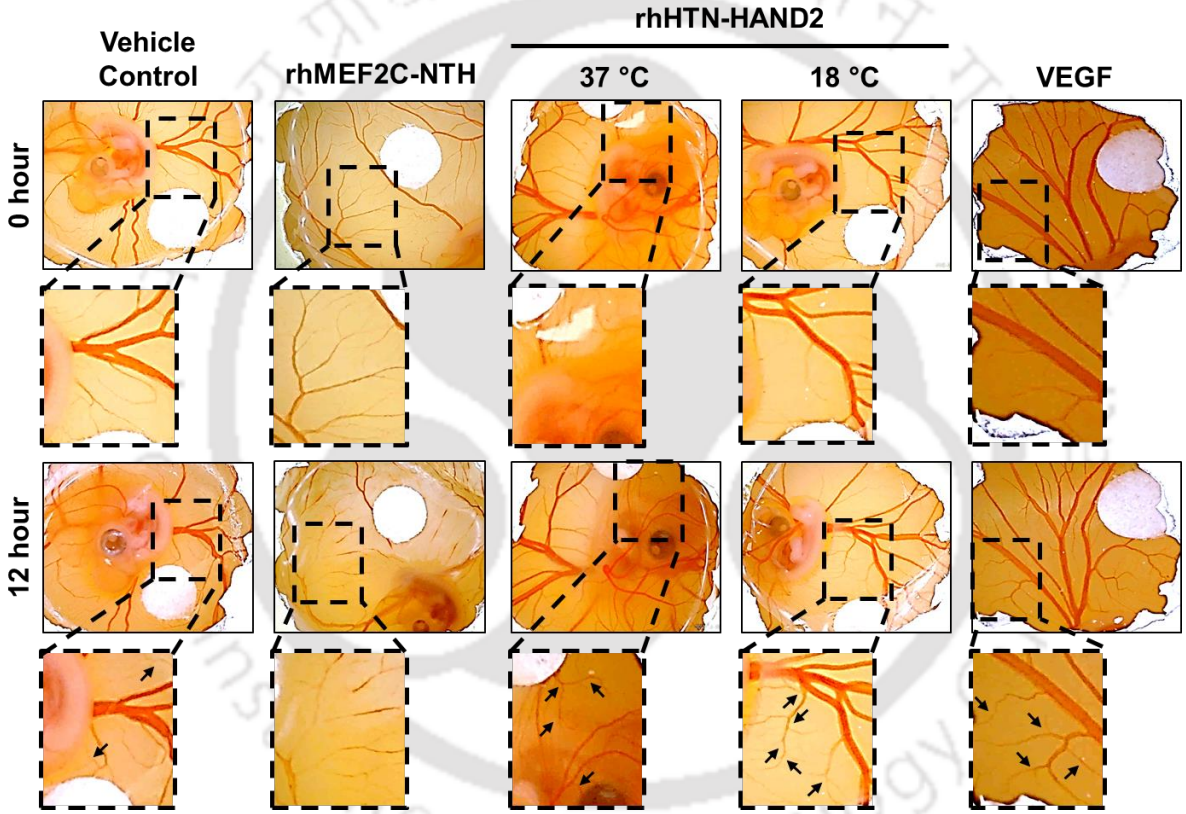


Figure 4.11 Effect of the purified rhMEF2C and rhHAND2 fusion proteins on angiogenesis in *ex vivo* chicken CAM model. Embryonated chicken eggs 3-4 days old were gently cut open on the top and treated with purified rhMEF2C-NTH or rhHTN-HAND2 protein (induced either at 37 or 18 °C) or vehicle control or VEGF (positive control) for 12 hours at 37 °C. Macroscopic images of CAM were captured before and after the incubation. All images were taken with identical camera settings. Images are representative of three different experiments.

Since HAND2 has been reported to promote angiogenesis (VanDusen et al. 2014; George and Firulli 2019), the effect of its pure recombinant fusion protein version in the neovascularization was investigated, and the results showed that the rhHTN-HAND2 protein promoted the migration of endothelial cells and induced neovascularization by sprouting angiogenesis. The endothelial cell migration and angiogenic potential of this purified fusion protein were comparable to the VEGF, a crucial inducer of angiogenesis. Similar observations of neovascularization were reported, induced by *E. coli*-derived recombinant asparaginyl-tRNA synthetase (D et al. 2016) and dermatopontin (Seetaraman Amritha et al. 2020) proteins. It is reported that HAND2 regulated angiogenesis through the Notch signaling pathway (VanDusen et al. 2014). Thus, the rhHAND2 fusion protein generated by us might also induce neovascularization through the Notch signaling pathway. However, further detailed studies are required to confirm the same.

4.2.2.4 Synergistic activation of an α-MHC gene by rhMEF2C and rhHAND2 fusion proteins

To further examine the transcriptional activity of these rhMEF2C and rhHAND2 fusion proteins, we selected one of their downstream targets, a cardiomyocyte-specific α-MHC gene, and performed a reporter assay using the α-MHC promoter-driven eGFP reporter system in cardiomyoblast cells. It has been reported that HAND2, along with another cardiac restricted transcription factor, MEF2C, synergistically activated α-MHC and atrial natriuretic peptide (ANP) genes (Zang et al. 2004a). We first transduced the cardiomyoblast, H9C2, cells with lentivirus α-MHC-eGFP reporter and then treated them with either purified rhMEF2C or rhHAND2 fusion protein (Figure 4.12A). Even after 3-4 rounds of protein transduction, we did not observe any GFP expression in H9C2 cells when treated only with one of the two fusion

proteins (Figure 4.12B). This result implied the possible reason that these fusion proteins cannot activate the α-MHC individually or were transcriptionally inactive proteins. Therefore, to eliminate the latter part, we infected the α-MHC-eGFP transduced H9C2 cells with FUdeltaGW-rtTA and inducible tetO-MEF2C lentiviral vectors. After 2-3 days of post-transduction, we confirmed the exogenous expression of MEF2C only upon induction with doxycycline in H9C2 cells (Figure 4.12C). Although the exogenous expression of MEF2C was confirmed, the α-MHC promoter-driven eGFP expression was not observed even after 6-8 days of doxycycline induction (Figure 4.12B). Hence, it is confirmed that MEF2C transcription factor alone is not sufficient to activate α-MHC gene.

Next, to study the synergistic role of MEF2C and HAND2 factor in the activation of α-MHC, we transduced the H9C2 cells with lentiviral vectors (α-MHC-eGFP/FUdeltaGW-rtTA/tetO-MEF2C) followed by treatment with rhHTN-HAND2 protein along with doxycycline (Figure 4.12A). Remarkably, the GFP^+ cells were observed in purified rhHAND2 fusion protein (induced either at 37 or 18 °C) and doxycycline-treated samples (Figure 4.12B). In fact, as early as day 4, expression of GFP was observed due to the transcriptional activation of its α-MHC promoter by rhHTN-HAND2 and MEF2C (doxycycline-induced expression in H9C2 cells) proteins. Flow cytometry analyses showed that around 32% of the population were GFP^+, whereas <1% of GFP^+ cells were observed in the respective control conditions (Figure 4.12D and 4.12E). Our results thus signify that along with MEF2C, our purified rhHAND2 fusion protein, independent of the induction temperature, synergistically activated and induced the expression of α-MHC promoter-driven eGFP, thereby confirming its biological activity.

We next focused on investigating the transcriptional activity of purified rhMEF2C fusion protein. To examine this activity, we replaced the doxycycline-induced MEF2C

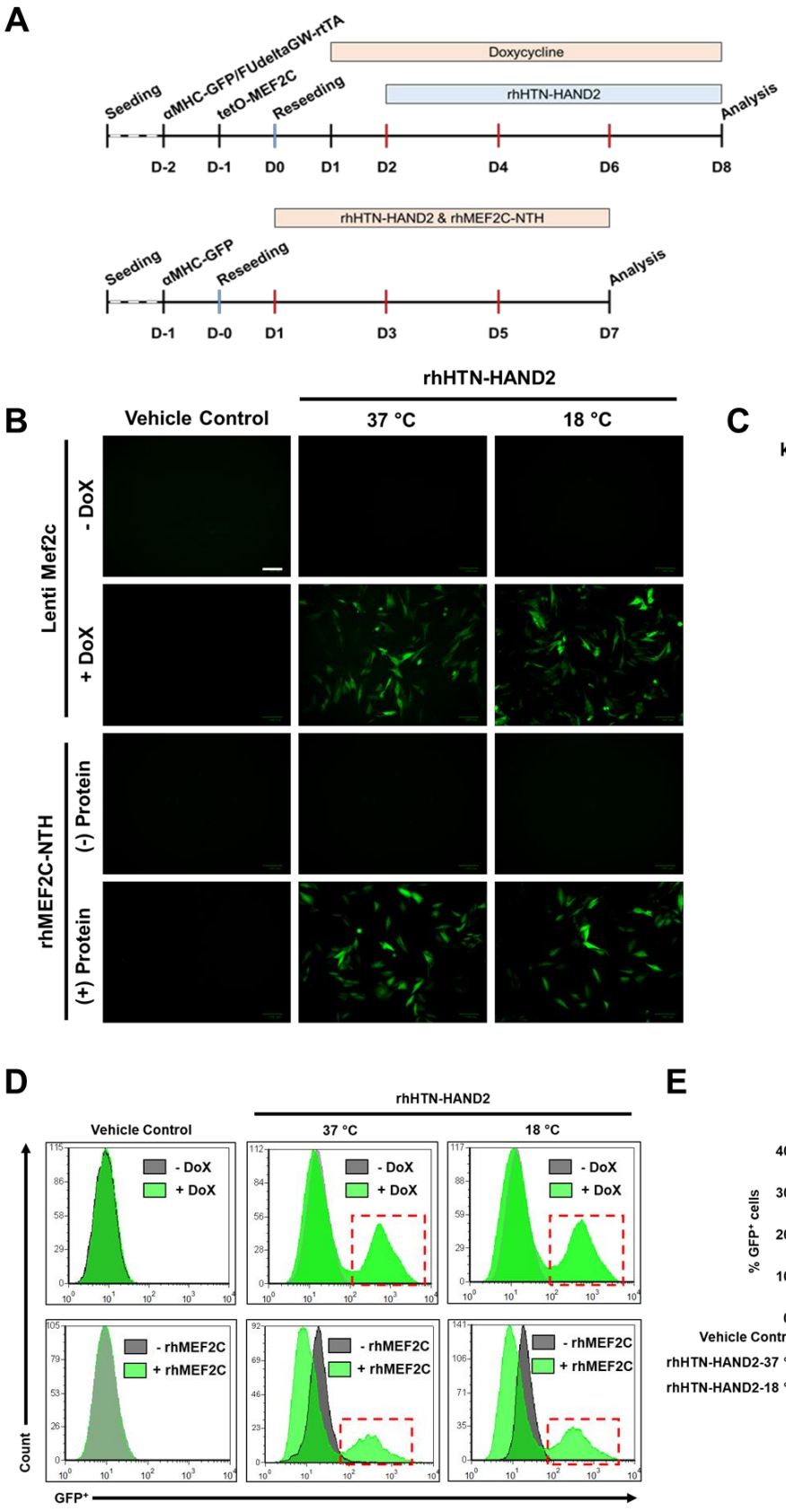


Figure 4.12 Effect of the purified rhMEF2C and rhHAND2 fusion proteins on synergistic activation of the α-MHC promoter. (**A)** Doxycycline-induced expression analysis of tetO-MEF2C transduced H9C2 cells. (**B)** Schematic representative of the strategy employed to analyze the transcriptional activity of the purified rhMEF2C-NTH and rhHTN-HAND2 proteins. (**C)** Firstly, H9C2 cells were transduced with α-MHC-eGFP, FUdeltaGW-rtTA, and tetO-MEF2C viral vectors. Subsequently, they were treated with doxycycline or purified rhMEF2C-NTH and rhHTN-HAND2 protein (induced either at 37 or 18 °C) or vehicle control. After 7-8 days, images were captured and analyzed. All images were taken with identical camera settings. Scale bar: 100 μm. (**D)** Flow cytometry analysis of GFP^+ cells. (**E)** Quantitative representation of **(F)**. The quantitative data shown are mean ± SEM (n = 4). Dox: doxycycline.

lentivectors with purified rhMEF2C-NTH fusion protein and then analyzed the GFP expression (Figure 4.12A). Interestingly, induction of GFP expression was observed when the H9C2 cells were treated with rhMEF2C-NTH protein along with purified rhHTN-HAND2 protein (Figure 4.12B). Further analyses with the flow cytometry revealed that around 22% of the H9C2 cells were GFP^+ when treated with either fusion proteins, while <1% GFP^+ cells were observed when treated with only rhMEF2C or rhHAND2 fusion protein (Figure 4.12D and 4.12E). These results were in line with the previous report that demonstrated the synergistic activation of an α-MHC gene (~5 fold) by MEF2C and HAND2 transcription factors in H9C2 cells using a luciferase-based reporter system (Zang et al. 2004a).

Cardiac restricted transcription factors interact with each other to regulate various target genes. For example, GATA4 recruits MEF2C to regulate the ANP transcriptional activity (Morin et al. 2000). Similarly, MEF2C recruited HAND2 to the α-MHC and ANP promoters and regulated transcription through protein-protein interaction (Zang et al. 2004a).

Studies demonstrated that MEF2C protein physically interacts with HAND2 protein both *in vitro* and *in vivo* (Zang et al. 2004a; Zang et al. 2004b). Thus, these purified fusion proteins (rhMEF2C and rhHAND2) physically interact with each other and forms a protein complex to regulate the transcription of an α-MHC gene. Physiologically, HAND2 has been identified in both homodimer and heterodimer forms. HAND2 forms heterodimers with the ubiquitously expressed protein (majorly bHLH proteins) to enhance its transcriptional activation (Dai et al. 2002; Thattaliyath et al. 2002). Taken together, our results thus signify transcriptional synergy of the rhMEF2C fusion protein with rhHAND2 fusion protein in the activation of the α-MHC promoter.

4.2.2.5 Effect of rhTBX5-NTH on colon cancer cells

To determine the biological activity of the purified cell-permeant rhTBX5-NTH protein, we first examined its effect on the clonogenic potential of SW620 cancer cells. Clonogenic assays have been widely used to analyze the effects of desired/target proteins, such as the effects of tumor suppressor genes WT1 and pRB including TBX5 on growth/tumor suppression (Guan et al. 1998; Zhang et al. 1999; He et al. 2002; Yu et al. 2010). Thus, we performed clonogenic assays to understand whether our purified rhTBX5-NTH protein might have any tumor suppressor ability in SW620 colon cancer cells. We selected this cell line because it has been used to demonstrate the tumor suppressor ability of TBX5 (Yu et al. 2010). We observed significantly lower (P = 0.0006) colony numbers of cells transduced with rhTBX5-NTH proteins than control cells after 12 rounds of transduction (Figure 4.13A and 4.13B). This observation was similar to those in previous studies demonstrating the effects of TBX5 (in genetic form) on the clonogenic potential of human osteosarcoma cells (He et al. 2002) and

colon cancer cells (Yu et al. 2010). Thus, the rhTBX5-NTH fusion protein inhibited the clonogenic potential of SW620 cells.

We next sought to understand whether this inhibition of clonogenic potential in SW620 cells might have been due to cell cycle arrest. After protein transduction, flow cytometry analysis (Figure 4.13C and 4.13D) indicated absence of the S-phase, which was present in the control cells (Figure 4.13E). These results correlate with those from a previous report in which decreased cell proliferation rates or cell cycle arrest have been found to strongly contribute to suppressed colony-formation potential (He et al. 2002). Given that TBX5 has been found to induce apoptosis (He et al. 2002; Yu et al. 2010), we performed Western blot analysis to determine whether rhTBX5-NTH protein transduced cells might undergo apoptosis. The results showed a decrease in procaspase-3 (Figure 4.13F; *top*) and the presence of cleaved PARP (Figure 4.13F; *middle*) in protein treated cell lysate compared to control cells. Notably, the quantitative analysis of immunoblots indicated a significant decrease ($P = 0.022$) in the expression of procaspase-3 in rhTBX5-NTH protein transduced cells compared with control cells, normalized to β-actin expression (Figure 4.13G). These findings confirmed that TBX5 played a role in decreasing the clonogenic potential of SW620 colon cancer cells by inhibiting the S-phase and inducing apoptosis.

Colon cancer cells are highly metastatic in nature, and TBX5, as a tumor suppressor, has been reported to play a role in regulating anti-metastatic and pro-metastatic genes, such as MTSS1 (metastasis suppressor 1) and MTA2 (metastasis associated protein 1 family member 2), respectively (Yu et al. 2010). Interestingly, ectopic expression of TBX5 has been reported to increase the expression of the MTSS1 gene and decrease the expression of the MTA2 gene in SW620 cells, thus leading to the inhibition of its metastatic potential (Yu et al. 2010). To

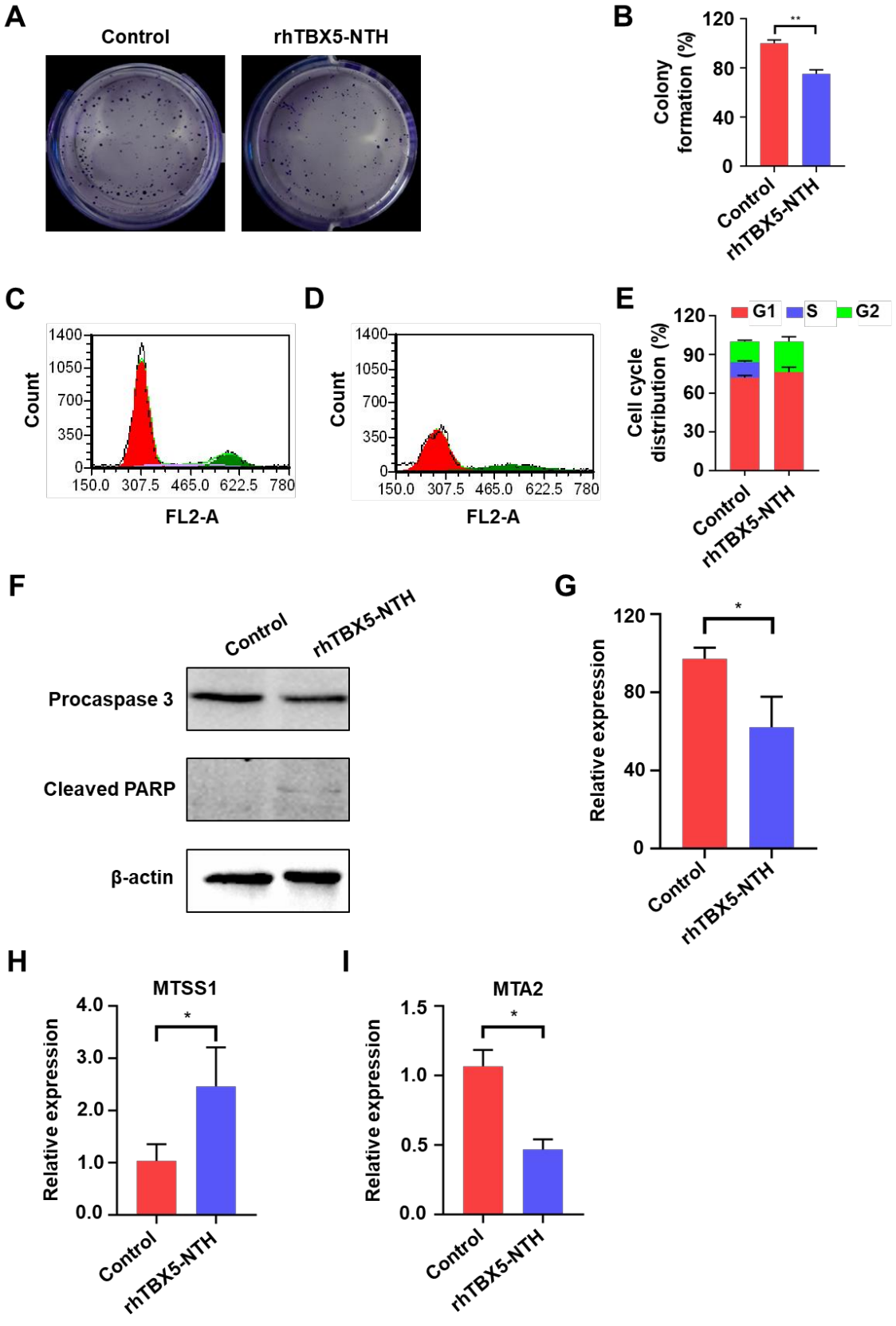


Figure 4.13 Tumor suppressor ability of rhTBX5-NTH protein in SW620 cancer cells. (**A**) Effect of rhTBX5-NTH protein on the clonogenic potential of SW620 cells. (**B**) Quantitative analysis of the clonogenic potential of rhTBX5-NTH protein transduced and non-transduced (control) SW620 cells. (**C**) and (**D**) Representative cell cycle profiles of rhTBX5-NTH protein transduced and non-transduced (control) SW620 cells, respectively. (**E**) Quantitative cell cycle distribution data. (**F**) Immunoblot analysis of apoptosis-associated protein expression in rhTBX5-NTH protein transduced and non-transduced (control) SW620 cells. (**G**) Quantitative data on procaspase-3 in (**F**). (**H**) and (**I**) Relative expression of pro-metastatic (MTA2) and anti-metastatic (MTSS1) genes in control and rhTBX5-NTH transduced SW620 cells, analyzed by RT-qPCR. The quantitative data shown are mean ± SEM (n = 3). *$P < 0.05$; **$P < 0.001$, ***$P < 0.0001$.

determine the role of our protein in the expression of these genes, we performed RT-qPCR, which indicated a significant upregulation (~2.4 fold; P=0.038) of MTSS1 and a significant downregulation (~2.3 fold; $P = 0.0017$) of MTA2 after protein treatment (Figure 9H and 9I). Thus, the increase in MTSS1 and decrease in MTA2 gene expression mediated by TBX5 protein correlated with the diminished clonogenic potential of SW620 cells, thus indicating an interdependency of the metastatic potential of cancer cells, and the ability to form colonies and the proliferation capacity.

4.3 Conclusion

Here, we have demonstrated that the six purified recombinant fusion proteins were stable under standard cell culture conditions. Among these purified fusion proteins, rhMEF2C-NTH was found to be highly stable even after 48 hours of incubation, and rhTBX5-NTH was partially stable after 18 hours of incubation. The remaining fusion proteins were stable for at least 24 hours of incubation under cell culture conditions. We then confirmed the cell penetration and

nuclear translocation ability of these fusion proteins without the need of a protein transduction reagent. Interestingly, we observed differences in the nuclear translocation between HeLa cells and HFFs, which signified that this translocation ability depends on the cell type used for the protein transduction. Next, we demonstrated the effect of purified rhTBX5 fusion proteins as a potent tumor suppressor on human colon cancer cells, followed by rhGATA4 on human breast cancer cells. Furthermore, rhETS2 and rhMESP1 promoted cell migration and acted as oncogenes in human breast cancer cells. Moreover, the generated rhMEF2C and rhHAND2 fusion proteins exhibited opposite roles in regulating angiogenesis. Remarkably, these two fusion proteins synergistically activated α-MHC and indicated their importance in the maturation of cardiomyocytes. Further, these results signify that our fusion strategy employed in this study does not affect the functionality of the fusion proteins. Prospectively, these purified bioactive recombinant fusion proteins can potentially be safe and effective molecular tools in the direct cardiac reprogramming process to generate functional cardiomyocytes in an integration-free manner and other biological applications such as determination of their 3D structure, identification of possible novel interaction partners, and elucidating their biological role in healthy and diseased mammalian cells.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 5

Overall conclusions and future perspective

From the extensive work carried out in this thesis, we conclude that we have su4ccessfully generated six biologically active cardiac reprogramming transcription factors in the form of recombinant proteins that can be utilized for various biological applications. To achieve this, we codon-optimized the gene sequences to attain the enhanced production of these recombinant proteins and employed a fusion strategy to enable cell penetration and nuclear translocation. However, gene optimization alone is insufficient to obtain the maximum soluble expression of these recombinant fusion proteins from the bacterial system, as shown by the results of this study. Notably, we have established the one-step homogeneous purification of these proteins. The established approach is simple, cost-effective, highly reproducible and involves purification under native conditions. However, further refinements are required in enhancing the yield of these purified fusion proteins from *E. coli*. We have identified the optimal expression conditions for each transcription factor and showed that the position of fusion tags had a significant impact on their expression, purification and yield. Besides, our highly pure recombinant versions of transcription factors had retained their folding conformation, are stable under cell culture conditions, and were able to transduce into the cells and localize into the nucleus. Notably, we have demonstrated that induction temperature and salt concentration highly contribute to the folding conformation of the recombinant proteins. In fact, our results suggest that MESP1 is more likely to be an ion-sensitive protein as the presence of salt influences its solubility and secondary structure conformation. Further, we have demonstrated that these purified recombinant proteins are biologically active using different assays. Here, we report that GATA4 and TBX5 fusion proteins act as tumor

suppressors, whereas ETS2 and MESP1 had an oncogenic effect in different types of cancer. We also showed that MEF2C and HAND2 played an essential role in regulating angiogenesis, and these fusion proteins were able to activate the expression of the α-MHC gene synergistically. Notably, our results confirmed that the fusion strategy employed in this study does not affect the functionality of these recombinant fusion proteins. Overall, the established cardiac-specific recombinant protein toolbox can be used to directly reprogram somatic cells to generate integration-free cardiomyocytes, which will have a range of applications in the field of disease modeling and drug development. Since recombinant proteins do not integrate in and alter the genome, the resulting myocytes will be integration-free and, therefore, can be used for cell therapy applications as well. Moreover, these purified fusion proteins can also aid in elucidating their molecular role in various diseases and cellular processes. As mentioned, some of the proteins have shown tumor-suppressing or oncogenic activity, which 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. Identified the appropriate genetic construct and expression conditions for each transcription factor.
2. One-step homogeneous purification of these recombinant fusion proteins were established under native conditions.
3. Purified fusion proteins are stable under cell conditions and has cell penetration and nuclear translocation ability.

4. Purified fusion proteins are biologically active, and the fusion strategy did not affect their functionality.

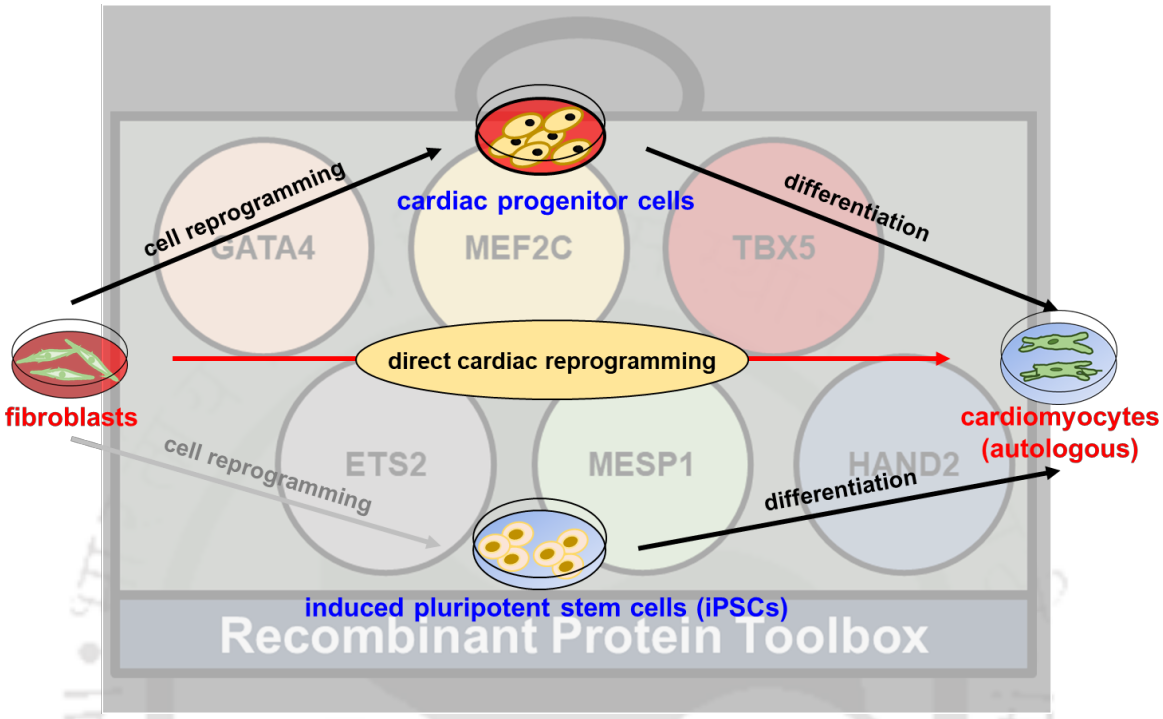


Figure 5.1 Illustration of prospective future applications of recombinant protein toolbox.

The future applications of this recombinant protein toolbox lie in understanding and elucidating the detailed role of these factors individually or in combination in various cellular environments. For example, implementing this toolbox can replace their genetic version and can be used to reprogram (i) somatic cells to cardiomyocytes via direct reprogramming approach, bypassing the pluripotent stage, (ii) reprogram somatic cells to cardiac progenitor cells and their subsequent differentiation into cardiomyocytes, and (iii) apply the components of this toolbox individually or in different combinations to differentiate iPSCs into cardiomyocytes efficiently (Figure 5.1). Further, this established protein toolbox can be employed to determine the appropriate stoichiometry of the reprogramming factors to enhance the cardiac reprogramming efficiency and kinetics. It also provides the sliding window

approach to understand the stage-specific role of each factor in the reprogramming paradigm. In addition, the components of this toolbox can be used for understanding their detailed mechanism of action in cancer biology.

List of antibodies used in this study

Antibodies	Dilutions		Source	Identifier
	Western blotting	Immunostaining		
Anti-His	1:5000	N/A	BioBharati	Cat# BB-AB0010
Anti-Gata4	1:1000	1:300	Santa Cruz Biotechnology	Cat#SC-25310
Anti-Mef2c	1:2000	1:100	Invitrogen	Cat#MA5-25477
Anti-Tbx5	1:2000	1:50	Santa Cruz Biotechnology	Cat#SC-515536
Anti-Ets2	1:1000	1:50	Santa Cruz Biotechnology	Cat#SC-374509
Anti-Ets2	N/A	1:100	Novus Biologicals	Cat#NBP3-04749
Anti-Mesp1	1:500	1:200	Santa Cruz Biotechnology	Cat#SC-130461
Anti-Hand2	1:2000	1:40	Invitrogen	Cat#PA5-35186
Anti-β-Actin	1:5000	N/A	BioBharati	Cat# BB-AB0024
Anti-Rabbit IgG, HRP-conjugated	1:5000	N/A	Invitrogen	Cat#31460
Anti-Mouse IgG, HRP-conjugated	1:5000	N/A	Invitrogen	Cat#31430
Alexa Fluor 488 goat anti rabbit IgG	N/A	1:2000	Invitrogen	Cat#A11034
Alexa Fluor 488 goat anti-mouse IgG	N/A	1:2000	Invitrogen	Cat#A11029
Alexa Fluor 594 goat anti-rabbit IgG	N/A	1:500	Invitrogen	Cat#A11037
Alexa Fluor 594 goat anti-mouse IgG	N/A	1:1000	Invitrogen	Cat#A11032

N/A, Not Applicable

References

Addis RC., Ifkovits JL., Pinto F., Kellam LD., Esteso P., Rentschler S., Christoforou N., Epstein JA., Gearhart JD 2013 Optimization of direct fibroblast reprogramming to cardiomyocytes using calcium activity as a functional measure of success. Journal of molecular cellular cardiology 60 97-106

Agarwal P., Wylie JN., Galceran J., Arkhitko O., Li C., Deng C., Grosschedl R., Bruneau BG 2003 Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. Development 130 623-633

Al-Hejin AM., Bora RS., Ahmed MMM 2019 Plasmids for optimizing expression of recombinant proteins in *E. coli*. Plasmid 13

Alagesan S., Griffin MD 2014 Autologous and allogeneic mesenchymal stem cells in organ transplantation: what do we know about their safety and efficacy? Current Opinion in Organ Transplantation 19 (1): 65-72

Amini H., Rezaie J., Vosoughi A., Rahbarghazi R., Nouri M 2017 Cardiac progenitor cells application in cardiovascular disease. Journal of Cardiovascular Thoracic Research 9 (3): 127

Antonitsis P., Ioannidou-Papagiannaki E., Kaidoglou A., Papakonstantinou C 2007 In vitro cardiomyogenic differentiation of adult human bone marrow mesenchymal stem cells. The role of 5-azacytidine. Interactive cardiovascular thoracic surgery 6 (5): 593-597

Araki Y., Hamafuji T., Noguchi C., Shimizu N 2012 Efficient recombinant production in mammalian cells using a novel IR/MAR gene amplification method. PLoS One 7 (7): e41787

Arceci RJ., King A., Simon MC., Orkin SH., Wilson DB 1993 Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. Molecular and cellular biology 13 (4): 2235-2246

Baeshen MN., Al-Hejin AM., Bora RS., Ahmed MM., Ramadan HA., Saini KS., Baeshen NA., Redwan EM 2015 Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. Journal of Microbiology and Biotechnology 25 (7): 953-962

Balsam LB., Wagers AJ., Christensen JL., Kofidis T., Weissman IL., Robbins RC 2004 Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium. Nature 428 (6983): 668-673

Baneyx F., Mujacic M 2004 Recombinant protein folding and misfolding in *Escherichia coli*. Nature biotechnology 22 (11): 1399-1408

Beltrami AP., Barlucchi L., Torella D., Baker M., Limana F., Chimenti S., Kasahara H., Rota M., Musso E., Urbanek K 2003 Adult cardiac stem cells are multipotent and support myocardial regeneration. Cell 114 (6): 763-776

Bhat EA., Sajjad N., Sabir JS., Kamli MR., Hakeem KR., Rather IA., Bahieldin A 2020 Molecular cloning, expression, overproduction and characterization of human TRAIP Leucine zipper protein. Saudi Journal of Biological Sciences 27 (6): 1562-1565

Bondue A., Blanpain C 2010 Mesp1: a key regulator of cardiovascular lineage commitment. Circulation research 107 (12): 1414-1427

Bondue A., Lapouge G., Paulissen C., Semeraro C., Iacovino M., Kyba M., Blanpain C 2008 Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell 3 (1): 69-84

Boogerd CJ., Evans SM 2016 TBX5 and NuRD divide the heart. Developmental cell 36 (3): 242-244

Borgohain MP., Haridhasapavalan KK., Dey C., Adhikari P., Thummer RP 2019 An insight into DNA-free reprogramming approaches to generate integration-free induced pluripotent stem cells for prospective biomedical applications. Stem cell reviews and reports 15 (2): 286-313

Bosnali M., Edenhofer F 2008 Generation of transducible versions of transcription factors Oct4 and Sox2. Biol Chem 389 (7): 851–861

Bradford MM 1976 A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry 72 (1-2): 248-254

Braun P., Hu Y., Shen B., Halleck A., Koundinya M., Harlow E., LaBaer J 2002 Proteome-scale purification of human proteins from bacteria. Proceedings of the National Academy of Sciences 99 (5): 2654-2659

Brower GL., Chancey AL., Thanigaraj S., Matsubara BB., Janicki JS 2002 Cause and effect relationship between myocardial mast cell number and matrix metalloproteinase activity. American Journal of Physiology-Heart Circulatory Physiology 283 (2): H518-H525

Bruneau BG., Nemer G., Schmitt JP., Charron F., Robitaille L., Caron S., Conner DA., Gessler M., Nemer M., Seidman CE 2001 A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell 106 (6): 709-721

Burgess-Brown NA., Sharma S., Sobott F., Loenarz C., Oppermann U., Gileadi O 2008 Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expression and Purification 59 (1): 94-102

Cai B., Li J., Wang J., Luo X., Ai J., Liu Y., Wang N., Liang H., Zhang M., Chen N 2012 microRNA-124 regulates cardiomyocyte differentiation of bone marrow-derived mesenchymal stem cells via targeting STAT3 signaling. Stem Cells 30 (8): 1746-1755

Cao N., Huang Y., Zheng J., Spencer CI., Zhang Y., Fu J-D., Nie B., Xie M., Zhang M., Wang H 2016 Conversion of human fibroblasts into functional cardiomyocytes by small molecules. Science 352 (6290): 1216-1220

Carbone GM., Napoli S., Valentini A., Cavalli F., Watson DK., Catapano CV 2004 Triplex DNA-mediated downregulation of Ets2 expression results in growth inhibition and apoptosis in human prostate cancer cells. Nucleic Acids Research 32 (14): 4358-4367

Chan SS-K., Shi X., Toyama A., Arpke RW., Dandapat A., Iacovino M., Kang J., Le G., Hagen HR., Garry DJ 2013 Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. Cell stem cell 12 (5): 587-601

Chancey AL., Brower GL., Janicki JS 2002 Cardiac mast cell-mediated activation of gelatinase and alteration of ventricular diastolic function. American Journal of Physiology-Heart Circulatory Physiology 282 (6): H2152-H2158

Chen R 2012 Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnology Advances 30 (5): 1102-1107

Chen Y., Yang Z., Zhao Z-A., Shen Z 2017 Direct reprogramming of fibroblasts into cardiomyocytes. Stem Cell Research Therapy 8 (1): 1-8

Chiapparo G., Lin X., Lescroart F., Chabab S., Paulissen C., Pitisci L., Bondue A., Blanpain C 2016 Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. Journal of Cell Biology 213 (4): 463-477

Chong JJ., Chandrakanthan V., Xaymardan M., Asli NS., Li J., Ahmed I., Heffernan C., Menon MK., Scarlett CJ., Rashidianfar A 2011 Adult cardiac-resident MSC-like stem cells with a proepicardial origin. Cell Stem Cell 9 (6): 527-540

Chong JJ., Forte E., Harvey RP 2014 Developmental origins and lineage descendants of endogenous adult cardiac progenitor cells. Stem Cell Research 13 (3): 592-614

Christoforou N., Chakraborty S., Kirkton RD., Adler AF., Addis RC., Leong KW 2017 Core transcription factors, microRNAs, and small molecules drive transdifferentiation of human fibroblasts towards the cardiac cell lineage. Scientific Reports 7 (1): 1-15

D JJ., Dhanraj M., Solaiappan S., Sivanesan S., Kron M., Dhanasekaran A 2016 Brugia malayi asparaginyl-tRNA synthetase stimulates endothelial cell proliferation, vasodilation and angiogenesis. PloS one 11 (1): e0146132

Dai Y-S., Cserjesi P., Markham BE., Molkentin JD 2002 The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. Journal of Biological Chemistry 277 (27): 24390-24398

David R., Brenner C., Stieber J., Schwarz F., Brunner S., Vollmer M., Mentele E., Müller-Höcker J., Kitajima S., Lickert H 2008 MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. Nature cell biology 10 (3): 338-345

Denning C., Allegrucci C., Priddle H., Barbadillo-Muñoz MD., Anderson D., Self T., Smith NM., Parkin T., Young LE 2003 Common culture conditions for maintenance and cardiomyocyte differentiation of the human embryonic stem cell lines, BG01 and HUES-7. International Journal of Developmental Biology 50 (1): 27-37

Dey C., Narayan G., Krishna Kumar H., Borgohain M., Lenka N 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. Life Sciences Group

Dey C., Raina K., Haridhasapavalan KK., Thool M., Sundaravadivelu PK., Adhikari P., Gogoi R., Thummer RP 2021a An overview of reprogramming approaches to derive integration-free induced pluripotent stem cells for prospective biomedical applications. Recent Advances in iPSC Technology 231-287

Dey C., Thool M., Bhattacharyya S., Sudhagar S., Thummer RP 2021b Generation of biologically active recombinant human OCT4 protein from *E. coli*. 3 Biotech 11 (5): 1-16

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*. Molecular Biotechnology 64 (1): 42-56

Do PM., Varanasi L., Fan S., Li C., Kubacka I., Newman V., Chauhan K., Daniels SR., Boccetta M., Garrett MR 2012 Mutant p53 cooperates with ETS2 to promote etoposide resistance. Genes & development 26 (8): 830-845

Donaldson LW., Petersen JM., Graves BJ., McIntosh LP 1996 Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. The EMBO journal 15 (1): 125-134

Dwyer J., Liu J-P 2010 Ets2 transcription factor, telomerase activity and breast cancer. Clinical experimental pharmacology physiology 37 (1): 83-87

Efe JA., Hilcove S., Kim J., Zhou H., Ouyang K., Wang G., Chen J., Ding S 2011 Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. Nature Cell Biology 13 (3): 215-222

Fry EA., Inoue K 2018 Aberrant expression of ETS1 and ETS2 proteins in cancer. Cancer reports and reviews 2 (3):

Fu J-D., Stone NR., Liu L., Spencer CI., Qian L., Hayashi Y., Delgado-Olguin P., Ding S., Bruneau BG., Srivastava D 2013 Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. Stem cell reports 1 (3): 235-247

Fu Y., Huang C., Xu X., Gu H., Ye Y., Jiang C., Qiu Z., Xie X 2015 Direct reprogramming of mouse fibroblasts into cardiomyocytes with chemical cocktails. Cell Research 25 (9): 1013-1024

Funakoshi S., Fernandes I., Mastikhina O., Wilkinson D., Tran T., Dhahri W., Mazine A., Yang D., Burnett B., Lee J 2021 Generation of mature compact ventricular cardiomyocytes from human pluripotent stem cells. Nature Communications 12 (1): 1-23

Galloway CA., Sowden MP., Smith HC 2003 Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. Biotechniques 34 (3): 524-530

Gao L., Hu Y., Tian Y., Fan Z., Wang K., Li H., Zhou Q., Zeng G., Hu X., Yu L 2019 Lung cancer deficient in the tumor suppressor GATA4 is sensitive to TGFBR1 inhibition. Nature Communications 10 (1): 1-15

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 biosystems engineering 38 (12): 2477-2486

Ge Y., LaFiura KM., Dombkowski AA., Chen Q., Payton SG., Buck SA., Salagrama S., Diakiw AE., Matherly LH., Taub JW 2008 The role of the proto-oncogene ETS2 in acute megakaryocytic leukemia biology and therapy. Leukemia 22 (3): 521-529

George RM., Firulli AB 2019 Hand factors in cardiac development. The Anatomical Record 302 (1): 101-107

Georgiades P., Rossant J 2006 Ets2 is necessary in trophoblast for normal embryonic anteroposterior axis development. Development 133 (6): 1059-1068

Gong Y., Zhang L., Zhang A., Chen X., Gao P., Zeng Q 2018 GATA4 inhibits cell differentiation and proliferation in pancreatic cancer. PloS one 13 (8):

Greenfield NJ 2006 Using circular dichroism spectra to estimate protein secondary structure. Nature protocols 1 (6): 2876-2890

Guan K., Rohwedel J., Wobus AM 1999 Embryonic stem cell differentiation models: cardiogenesis, myogenesis, neurogenesis, epithelial and vascular smooth muscle cell differentiation in vitro. Cytotechnology 30 (1): 211-226

Guan L-S., Liu J-j., Xu Y-H., Wang Z-Y 1998 A point mutation within exon 5 of the WT1 gene of a sporadic unilateral Wilms' tumor alters gene function. Cancer research 58 (18): 4180-4184

Guo S., Zhang Y., Zhou T., Wang D., Weng Y., Chen Q., Ma J., Li Y-p., Wang L 2018a GATA4 as a novel regulator involved in the development of the neural crest and craniofacial skeleton via Barx1. Cell Death & Differentiation 25 (11): 1996-2009

Guo X., Bai Y., Zhang L., Zhang B., Zagidullin N., Carvalho K., Du Z., Cai B 2018b Cardiomyocyte differentiation of mesenchymal stem cells from bone marrow: new regulators and its implications. Stem Cell Research Therapy 9 (1): 1-12

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

Han X., Tang J., Chen T., Ren G 2019 Restoration of GATA4 expression impedes breast cancer progression by transcriptional repression of ReLA and inhibition of NF-κB signaling. Journal of Cellular Biochemistry 120 (1): 917-927

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

Hatcher CJ., Goldstein MM., Mah CS., Delia CS., Basson CT 2000 Identification and localization of TBX5 transcription factor during human cardiac morphogenesis. Developmental dynamics: an official publication of the American Association of Anatomists 219 (1): 90-95

Hattan N., Kawaguchi H., Ando K., Kuwabara E., Fujita J., Murata M., Suematsu M., Mori H., Fukuda K 2005 Purified cardiomyocytes from bone marrow mesenchymal stem cells produce stable intracardiac grafts in mice. Cardiovascular Research 65 (2): 334-344

He M-L., Chen Y., Peng Y., Jin D., Du D., Wu J., Lu P., Lin MC., Kung H-F 2002 Induction of apoptosis and inhibition of cell growth by developmental regulator hTBX5. Biochemical biophysical research communications 297 (2): 185-192

Heikinheimo M., Scandrett JM., Wilson DB 1994 Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. Developmental biology 164 (2): 361-373

Hirai H., Yang B., Garcia-Barrio MT., Rom O., Ma PX., Zhang J., Chen YE 2018 Direct Reprogramming of Fibroblasts Into Smooth Muscle-Like Cells With Defined Transcription Factors—Brief Report. Arteriosclerosis, Thrombosis, Vascular Biology 38 (9): 2191-2197

Huang C-J., Lin H., Yang X 2012 Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. Journal of Industrial Microbiology and Biotechnology 39 (3): 383-399

Ieda M., Fu J-D., Delgado-Olguin P., Vedantham V., Hayashi Y., Bruneau BG., Srivastava D 2010 Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142 (3): 375-386

Ieda M., Tsuchihashi T., Ivey KN., Ross RS., Hong T-T., Shaw RM., Srivastava D 2009 Cardiac fibroblasts regulate myocardial proliferation through β1 integrin signaling. Developmental Cell 16 (2): 233-244

Ifkovits JL., Addis RC., Epstein JA., Gearhart JD 2014 Inhibition of TGFβ signaling increases direct conversion of fibroblasts to induced cardiomyocytes. PloS one 9 (2): e89678

Islas JF., Liu Y., Weng K-C., Robertson MJ., Zhang S., Prejusa A., Harger J., Tikhomirova D., Chopra M., Iyer D 2012 Transcription factors ETS2 and MESP1 transdifferentiate human dermal fibroblasts into cardiac progenitors. Proceedings of the National Academy of Sciences 109 (32): 13016-13021

Isomi M., Sadahiro T., Fujita R., Abe Y., Yamada Y., Akiyama T., Mizukami H., Shu T., Fukuda K., Ieda M 2021 Direct reprogramming with Sendai virus vectors repaired infarct hearts at the chronic stage. Biochemical Biophysical Research Communications 560 87-92

Jayawardena TM., Egemnazarov B., Finch EA., Zhang L., Payne JA., Pandya K., Zhang Z., Rosenberg P., Mirotsou M., Dzau VJ 2012 MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes. Circulation Research 110 (11): 1465-1473

Jiang Y., Jahagirdar BN., Reinhardt RL., Schwartz RE., Keene CD., Ortiz-Gonzalez XR., Reyes M., Lenvik T., Lund T., Blackstad M 2002 Pluripotency of mesenchymal stem cells derived from adult marrow. Nature 418 (6893): 41-49

Kabbout M., Garcia MM., Fujimoto J., Liu DD., Woods D., Chow C-W., Mendoza G., Momin AA., James BP., Solis L 2013 Ets2 mediated tumor suppressive function and met oncogene inhibition in human non–small cell lung cancer. Clinical cancer research 19 (13): 3383-3395

Kadari A., Mekala S., Wagner N., Malan D., Köth J., Doll K., Stappert L., Eckert D., Peitz M., Matthes J 2015 Robust generation of cardiomyocytes from human iPS cells requires precise modulation of BMP and WNT signaling. Stem Cell Reviews and Reports 11 (4): 560-569

Kakkar A., Nandy SB., Gupta S., Bharagava B., Airan B., Mohanty S 2019 Adipose tissue derived mesenchymal stem cells are better respondents to TGFβ1 for in vitro generation of cardiomyocyte-like cells. Molecular Cellular Biochemistry 460 (1): 53-66

Kaptoge S., Pennells L., De Bacquer D., Cooney MT., Kavousi M., Stevens G., Riley LM., Savin S., Khan T., Altay S 2019 World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. The Lancet Global Health 7 (10): e1332-e1345

Karpievitch YV., Polpitiya AD., Anderson GA., Smith RD., Dabney AR 2010 Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects. The annals of applied statistics 4 (4): 1797

Kawada H., Fujita J., Kinjo K., Matsuzaki Y., Tsuma M., Miyatake H., Muguruma Y., Tsuboi K., Itabashi Y., Ikeda Y 2004 Nonhematopoietic mesenchymal stem cells can be mobilized and differentiate into cardiomyocytes after myocardial infarction. Blood 104 (12): 3581-3587

Kehat I., Kenyagin-Karsenti D., Snir M., Segev H., Amit M., Gepstein A., Livne E., Binah O., Itskovitz-Eldor J., Gepstein L 2001 Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. The Journal of Clinical Investigation 108 (3): 407-414

Kelly SM., Jess TJ., Price NC 2005 How to study proteins by circular dichroism. Biochimica et Biophysica Acta -Proteins and Proteomics 1751 (2): 119-139

Khow O., Suntrarachun S 2012 Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pacific journal of tropical biomedicine 2 (2): 159-162

Kidder BL,2020 Direct Reprogramming of Mouse Embryonic Fibroblasts to Induced Trophoblast Stem Cells. Stem Cell Transcriptional Networks. Springer, pp 285-292

Kintzing JR., Interrante MVF., Cochran JR 2016 Emerging strategies for developing next-generation protein therapeutics for cancer treatment. Trends in pharmacological sciences 37 (12): 993-1008

Kisby T., de Lázaro I., Fisch S., Cartwright EJ., Cossu G., Kostarelos K 2021 Adenoviral Mediated Delivery of OSKM Factors Induces Partial Reprogramming of Mouse Cardiac Cells In Vivo. Advanced Therapeutics 4 (2): 2000141

Kuo CT., Morrisey EE., Anandappa R., Sigrist K., Lu MM., Parmacek MS., Soudais C., Leiden JM 1997 GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes & development 11 (8): 1048-1060

Kurien BT., Scofield RH,2012 Common artifacts and mistakes made in electrophoresis. Protein Electrophoresis. Springer, pp 633-640

Laflamme MA., Chen KY., Naumova AV., Muskheli V., Fugate JA., Dupras SK., Reinecke H., Xu C., Hassanipour M., Police S 2007 Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nature Biotechnology 25 (9): 1015-1024

Laugwitz K-L., Moretti A., Lam J., Gruber P., Chen Y., Woodard S., Lin L-Z., Cai C-L., Lu MM., Reth M 2005 Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. Nature 433 (7026): 647-653

Le T., Chong J 2016 Cardiac progenitor cells for heart repair. Cell death discovery 2 (1): 1-4

Lee K., Yu P., Lingampalli N., Kim HJ., Tang R., Murthy N 2015 Peptide-enhanced mRNA transfection in cultured mouse cardiac fibroblasts and direct reprogramming towards cardiomyocyte-like cells. International Journal of Nanomedicine 10 1841

Li J., Liu W-D., Yang Z-L., Yuan F., Xu L., Li R-G., Yang Y-Q 2014 Prevalence and spectrum of GATA4 mutations associated with sporadic dilated cardiomyopathy. Gene 548 (2): 174-181

Li X-H., Li Q., Jiang L., Deng C., Liu Z., Fu Y., Zhang M., Tan H., Feng Y., Shan Z 2015 Generation of functional human cardiac progenitor cells by high-efficiency protein transduction. Stem cells translational medicine 4 (12): 1415-1424

Lian X., Hsiao C., Wilson G., Zhu K., Hazeltine LB., Azarin SM., Raval KK., Zhang J., Kamp TJ., Palecek SP 2012 Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proceedings of the National Academy of Sciences 109 (27): E1848-E1857

Liang Q., Xu C., Chen X., Li X., Lu C., Zhou P., Yin L., Qian R., Chen S., Ling Z 2015 The roles of Mesp family proteins: functional diversity and redundancy in differentiation of pluripotent stem cells and mammalian mesodermal development. Protein & cell 6 (8): 553-561

Lili W., Chaozhan W., Xindu G 2006 Expression, renaturation and simultaneous purification of recombinant human stem cell factor in *Escherichia coli*. Biotechnology letters 28 (13): 993-997

Lindsley RC., Gill JG., Murphy TL., Langer EM., Cai M., Mashayekhi M., Wang W., Niwa N., Nerbonne JM., Kyba M 2008 Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell stem cell 3 (1): 55-68

Liu DD., Kang Y 2017 Ets2 anchors the prometastatic function of mutant p53 in osteosarcoma. Genes & development 31 (18): 1823-1824

Liu Y 2017 Earlier and broader roles of Mesp1 in cardiovascular development. Cellular and Molecular Life Sciences 74 (11): 1969-1983

Lu R., Li W-W., Katzir A., Raichlin Y., Yu H-Q., Mizaikoff B 2015 Probing the secondary structure of bovine serum albumin during heat-induced denaturation using mid-infrared fiberoptic sensors. Analyst 140 (3): 765-770

Ma R., Liang J., Huang W., Guo L., Cai W., Wang L., Paul C., Yang H-T., Kim HW., Wang Y 2018 Electrical stimulation enhances cardiac differentiation of human induced pluripotent stem cells for myocardial infarction therapy. Antioxidants redox signaling 28 (5): 371-384

Maertens B., Spriestersbach A., von Groll U., Roth U., Kubicek J., Gerrits M., Graf M., Liss M., Daubert D., Wagner R 2010 Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. Protein Science 19 (7): 1312-1326

Makino S., Fukuda K., Miyoshi S., Konishi F., Kodama H., Pan J., Sano M., Takahashi T., Hori S., Abe H 1999 Cardiomyocytes can be generated from marrow stromal cells in vitro. The Journal of clinical investigation 103 (5): 697-705

Markmee R., Aungsuchawan S., Tancharoen W., Narakornsak S., Pothacharoen P 2020 Differentiation of cardiomyocyte-like cells from human amniotic fluid mesenchymal stem cells by combined induction with human platelet lysate and 5-azacytidine. Heliyon 6 (9): e04844

Malik A., Alsenaidy AM., Elrobh M., Khan W., Alanazi MS., Bazzi MD 2016 Optimization of expression and purification of HSPA6 protein from Camelus dromedarius in *E. coli*. Saudi Journal of Biological Sciences 23 (3): 410-419

Manjushree M., Revanasiddappa HD 2018 A diversified spectrometric and molecular docking technique to biophysical study of interaction between bovine serum albumin and sodium salt of risedronic acid, a bisphosphonate for skeletal disorders. Bioinorganic chemistry applications 2018

Mathison M., Singh VP., Chiuchiolo MJ., Sanagasetti D., Mao Y., Patel VB., Yang J., Kaminsky SM., Crystal RG., Rosengart TK 2017 In situ reprogramming to transdifferentiate fibroblasts into cardiomyocytes using adenoviral vectors: Implications for clinical myocardial regeneration. The Journal of Thoracic Cardiovascular Surgery 153 (2): 329-339. e3

Mattanovich D., Branduardi P., Dato L., Gasser B., Sauer M., Porro D 2012 Recombinant protein production in yeasts. Recombinant gene expression 329-358

McCarroll L., King LA 1997 Stable insect cell cultures for recombinant protein production. Current Opinion in Biotechnology 8 (5): 590-594

McGill Jr HC., McMahan CA., Gidding SS 2008 Preventing heart disease in the 21st century: implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study. Circulation 117 (9): 1216-1227

Mendis S., Puska P., Norrving B (2011) Global atlas on cardiovascular disease prevention and control. World Health Organization

Messina E., De Angelis L., Frati G., Morrone S., Chimenti S., Fiordaliso F., Salio M., Battaglia M., Latronico MV., Coletta M 2004 Isolation and expansion of adult cardiac stem cells from human and murine heart. Circulation Research 95 (9): 911-921

Micsonai A., Wien F., Bulyáki É., Kun J., Moussong É., Lee Y-H., Goto Y., Réfrégiers M., Kardos J 2018 BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic acids research 46 (W1): W315-W322

Micsonai A., Wien F., Kernya L., Lee Y-H., Goto Y., Réfrégiers M., Kardos J 2015 Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proceedings of the National Academy of Sciences 112 (24): E3095-E3103

Miyamoto K., Akiyama M., Tamura F., Isomi M., Yamakawa H., Sadahiro T., Muraoka N., Kojima H., Haginiwa S., Kurotsu S 2018 Direct in vivo reprogramming with Sendai virus vectors improves cardiac function after myocardial infarction. Cell Stem Cell 22 (1): 91-103. e5

Molkentin JD., Lin Q., Duncan SA., Olson EN 1997 Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes & development 11 (8): 1061-1072

Morin S., Charron F., Robitaille L., Nemer M 2000 GATA-dependent recruitment of MEF2 proteins to target promoters. The EMBO journal 19 (9): 2046-2055

Münst B., Thier MC., Winnemöller D., Helfen M., Thummer RP., Edenhofer F 2016 Nanog induces suppression of senescence through downregulation of p27KIP1 expression. Journal of cell science 129 (5): 912-920

Muraoka N., Yamakawa H., Miyamoto K., Sadahiro T., Umei T., Isomi M., Nakashima H., Akiyama M., Wada R., Inagawa K 2014 MiR-133 promotes cardiac reprogramming by directly repressing Snai1 and silencing fibroblast signatures. The EMBO journal 33 (14): 1565-1581

Murry CE., Reinecke H., Pabon LM 2006 Regeneration gaps: observations on stem cells and cardiac repair. Journal of the American College of Cardiology 47 (9): 1777-1785

Murry CE., Soonpaa MH., Reinecke H., Nakajima H., Nakajima HO., Rubart M., Pasumarthi K., Ismail Virag J., Bartelmez SH., Poppa V 2004 Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts. Nature 428 (6983): 664-668

Nam Y-J., Song K., Luo X., Daniel E., Lambeth K., West K., Hill JA., DiMaio JM., Baker LA., Bassel-Duby R 2013 Reprogramming of human fibroblasts toward a cardiac fate. Proceedings of the National Academy of Sciences 110 (14): 5588-5593

Narayan G., Agrawal A., Joshi N., Gogoi R., Nagotu S., Thummer RP 2021a Protein Production and Purification of a Codon-Optimized Human NGN3 Transcription Factor from *E. coli*. The Protein Journal 1-16

Narayan G., Sundaravadivelu PK., Agrawal A., Gogoi R., Nagotu S., Thummer RP 2021b Soluble expression, purification, and secondary structure determination of human PDX1 transcription factor. Protein Expression and Purification 180 105807

Nazari H., Kehtari M., Rad I., Ashtari B., Joghataei MT 2020 Electrical stimulation induces differentiation of human cardiosphere-derived cells (hCDCs) to committed cardiomyocyte. Molecular Cellular Biochemistry 470 (1): 29-39

Nezafat N., Sadraeian M., Rahbar MR., Khoshnoud MJ., Mohkam M., Gholami A., Banihashemi M., Ghasemi Y 2015 Production of a novel multi-epitope peptide vaccine for cancer immunotherapy in TC-1 tumor-bearing mice. Biologicals 43 (1): 11-17

O'Malley J., Woltjen K., Kaji KJCoib 2009 New strategies to generate induced pluripotent stem cells. 20 (5): 516-521

Oh H., Bradfute SB., Gallardo TD., Nakamura T., Gaussin V., Mishina Y., Pocius J., Michael LH., Behringer RR., Garry DJ 2003 Cardiac progenitor cells from adult myocardium: homing, differentiation, and fusion after infarction. Proceedings of the National Academy of Sciences 100 (21): 12313-12318

Orlic D., Kajstura J., Chimenti S., Limana F., Jakoniuk I., Quaini F., Nadal-Ginard B., Bodine DM., Leri A., Anversa P 2001 Mobilized bone marrow cells repair the infarcted heart, improving function and survival. Proceedings of the National Academy of Sciences 98 (18): 10344-10349

Ou J., Wang L., Ding X., Du J., Zhang Y., Chen H., Xu A 2004 Stationary phase protein overproduction is a fundamental capability of *Escherichia coli*. Biochemical biophysical research communications 314 (1): 174-180

Ou D., Wang Q., Huang Y., Zeng D., Wei T., Ding L., Li X., Zheng Q., Jin Y 2016 Co-culture with neonatal cardiomyocytes enhances the proliferation of iPSC-derived cardiomyocytes via FAK/JNK signaling. BMC Developmental Biology 16 (1): 1-12

Overton TW 2014 Recombinant protein production in bacterial hosts. Drug discovery today 19 (5): 590-601

Paige SL., Osugi T., Afanasiev OK., Pabon L., Reinecke H., Murry CE 2010 Endogenous Wnt/β-catenin signaling is required for cardiac differentiation in human embryonic stem cells. PloS one 5 (6): e11134

Pak HN., Qayyum M., Kim DT., Hamabe A., Miyauchi Y., Lill MC., Frantzen M., Takizawa K., Chen LS., Fishbein MC 2003 Mesenchymal stem cell injection induces cardiac nerve sprouting and increased tenascin expression in a Swine model of myocardial infarction. Journal of Cardiovascular Electrophysiology 14 (8): 841-848

Pallavi S., Ho DM., Hicks C., Miele L., Artavanis-Tsakonas S 2012 Notch and Mef2 synergize to promote proliferation and metastasis through JNK signal activation in Drosophila. The EMBO journal 31 (13): 2895-2907

Paoletti C., Divieto C., Tarricone G., Di Meglio F., Nurzynska D., Chiono V 2020 MicroRNA-mediated direct reprogramming of human adult fibroblasts toward cardiac phenotype. Frontiers in Bioengineering Biotechnology 8 529

Park G., Yoon BS., Kim YS., Choi S-C., Moon J-H., Kwon S., Hwang J., Yun W., Kim J-H., Park C-Y 2015 Conversion of mouse fibroblasts into cardiomyocyte-like cells using small molecule treatments. Biomaterials 54 201-212

Patterson E., Scherlag BJ 2002 Decremental conduction in the posterior and anterior AV nodal inputs. Journal of interventional cardiac electrophysiology 7 (2): 137-148

Patterson J., Coats C., McGowan R 2020 Familial dilated cardiomyopathy associated with pathogenic TBX5 variants: Expanding the cardiac phenotype associated with Holt–Oram syndrome. American Journal of Medical Genetics Part A 182 (7): 1725-1734

Peitz M., Münst B., Thummer RP., Helfen M., Edenhofer F 2014 Cell-permeant recombinant Nanog protein promotes pluripotency by inhibiting endodermal specification. Stem cell research 12 (3): 680-689

Prabhakaran D., Jeemon P., Roy A 2016 Cardiovascular diseases in India: current epidemiology and future directions. Circulation 133 (16): 1605-1620

Protze S., Khattak S., Poulet C., Lindemann D., Tanaka EM., Ravens U 2012 A new approach to transcription factor screening for reprogramming of fibroblasts to cardiomyocyte-like cells. Journal of Molecular Cellular Cardiology 53 (3): 323-332

Rabhi-Essafi I., Sadok A., Khalaf N., Fathallah DMJPE, Design., Selection 2007 A strategy for high-level expression of soluble and functional human interferon α as a GST-fusion protein in *E. coli*. 20 (5): 201-209

Raffetto JD., Khalil RA 2008 Matrix metalloproteinases and their inhibitors in vascular remodeling and vascular disease. Biochemical pharmacology 75 (2): 346-359

Raghunathan S., Islas JF., Mistretta B., Iyer D., Shi L., Gunaratne PH., Ko G., Schwartz RJ., McConnell BK 2020 Conversion of human cardiac progenitor cells into cardiac pacemaker-like cells. Journal of Molecular and Cellular Cardiology 138 12-22

Rallis C., Bruneau BG., Del Buono J., Seidman CE., Seidman J., Nissim S., Tabin CJ., Logan MP 2003 Tbx5 is required for forelimb bud formation and continued outgrowth. Development 130 2741–2751

Ramesh S., Govarthanan K., Ostrovidov S., Zhang H., Hu Q., Camci-Unal G., Verma RS., Ramalingam M 2021 Cardiac differentiation of mesenchymal stem cells: impact of biological and chemical inducers. Stem Cell Reviews and Reports 17 (4): 1343-1361

Reed RG., Feldhoff RC., Clute O., Peters Jr T 1975 Fragments of bovine serum albumin produced by limited proteolysis. Conformation and ligand binding. Biochemistry 14 (21): 4578-4583

Rojas A., Kong SW., Agarwal P., Gilliss B., Pu WT., Black BL 2008 GATA4 is a direct transcriptional activator of cyclin D2 and Cdk4 and is required for cardiomyocyte proliferation in anterior heart field-derived myocardium. Molecular and cellular biology 28 (17): 5420-5431

Rosano GL., Ceccarelli EA 2014 Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers in microbiology 5 172

Ryan BJ., Henehan GT 2013 Overview of approaches to preventing and avoiding proteolysis during expression and purification of proteins. Current protocols in protein science 71 (1): 5.25. 1-5.25. 7

Saga Y., Kitajima S., Miyagawa-Tomita S 2000 Mesp1 expression is the earliest sign of cardiovascular development. Trends in cardiovascular medicine 10 (8): 345-352

Saga Y., Miyagawa-Tomita S., Takagi A., Kitajima S., Miyazaki J., Inoue T 1999 MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 126 (15): 3437-3447

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): 1-4

Seetaraman Amritha TM., Mahajan S., Subramaniam K., Chandramohan Y., Dhanasekaran A 2020 Cloning, expression and purification of recombinant dermatopontin in *Escherichia coli*. PloS one 15 (11): e0242798

Serna N., Sánchez-García L., Unzueta U., Díaz R., Vázquez E., Mangues R., Villaverde A 2018 Protein-based therapeutic killing for cancer therapies. Trends in biotechnology 36 (3): 318-335

Seth A., Watson DK 2005 ETS transcription factors and their emerging roles in human cancer. European journal of cancer 41 (16): 2462-2478

Shen X., Pan B., Zhou H., Liu L., Lv T., Zhu J., Huang X., Tian J 2017 Differentiation of mesenchymal stem cells into cardiomyocytes is regulated by miRNA-1-2 via WNT signaling pathway. Journal of Biomedical Science 24 (1): 1-8

Sheydina A., Volkova M., Jiang L., Juhasz O., Zhang J., Tae HJ., Perino MG., Wang M., Zhu Y., Lakatta EG 2012 Linkage of cardiac gene expression profiles and ETS2 with lifespan variability in rats. Aging cell 11 (2): 350-359

Shim WS., Jiang S., Wong P., Tan J., Chua YL., Tan YS., Sin YK., Lim CH., Chua T., Teh MJB 2004 Ex vivo differentiation of human adult bone marrow stem cells into cardiomyocyte-like cells. Biochemical Biophysical Research Communications 324 (2): 481-488

Singh VP., Mathison M., Patel V., Sanagasetti D., Gibson BW., Yang J., Rosengart TK 2016 MiR-590 promotes transdifferentiation of porcine and human fibroblasts toward a cardiomyocyte-like fate by directly repressing specificity protein 1. Journal of the American Heart Association 5 (11): e003922

Singh VP., Pinnamaneni JP., Pugazenthi A., Sanagasetti D., Mathison M., Wang K., Yang J., Rosengart TK 2020 Enhanced generation of induced cardiomyocytes using a small-molecule cocktail to overcome barriers to cardiac cellular reprogramming. Journal of the American Heart Association 9 (12): e015686

Soibam B., Benham A., Kim J., Weng KC., Yang L., Xu X., Robertson M., Azares A., Cooney AJ., Schwartz RJ 2015 Genome-Wide Identification of MESP1 Targets Demonstrates Primary Regulation Over Mesendoderm Gene Activity. Stem Cells 33 (11): 3254-3265

Sommer CA., Mostoslavsky G 2013 The evolving field of induced pluripotency: recent progress and future challenges. Journal of cellular physiology 228 (2): 267-275

Song K., Nam Y-J., Luo X., Qi X., Tan W., Huang GN., Acharya A., Smith CL., Tallquist MD., Neilson EG 2012 Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 485 (7400): 599-604

Sørensen HP., Mortensen KK 2005 Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial cell factories 4 (1): 1-8

Specht E., Miyake-Stoner S., Mayfield S 2010 Micro-algae come of age as a platform for recombinant protein production. Biotechnology letters 32 (10): 1373-1383

Stefan A., Calonghi N., Schipani F., Dal Piaz F., Sartor G., Hochkoeppler A 2018 Purification of active recombinant human histone deacetylase 1 (HDAC1) overexpressed in *Escherichia coli*. Biotechnology letters 40 (9-10): 1355-1363

Steimle J., Moskowitz I 2017 TBX5: a key regulator of heart development. Current topics in developmental biology 122 195-221

Stewart Jr JA., Wei C-C., Brower GL., Rynders PE., Hankes GH., Dillon AR., Lucchesi PA., Janicki JS., Dell'Italia L 2003 Cardiac mast cell-and chymase-mediated matrix metalloproteinase activity and left ventricular remodeling in mitral regurgitation in the dog. Journal of molecular cellular cardiology 35 (3): 311-319

Stock K., Nolden L., Edenhofer F., Quandel T., Brüstle O 2010 Transcription factor-based modulation of neural stem cell differentiation using direct protein transduction. Cellular and Molecular Life Sciences 67 (14): 2439-2449

Sturtzel C., Testori J., Schweighofer B., Bilban M., Hofer E 2014 The transcription factor MEF2C negatively controls angiogenic sprouting of endothelial cells depending on oxygen. PloS One 9 (7): e101521

Sun W., Jiao C., Xiao Y., Wang L., Yu C., Liu J., Yu Y., Wang L 2016 Salt-dependent aggregation and assembly of *E coli*-expressed ferritin. Dose-Response 14 (1): 1559325816632102

Takahashi K., Tanabe K., Ohnuki M., Narita M., Ichisaka T., Tomoda K., Yamanaka S 2007 Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131 (5): 861-872

Takahashi K., Yamanaka S 2006 Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126 (4): 663-676

Takeuchi JK., Ohgi M., Koshiba-Takeuchi K., Shiratori H., Sakaki I., Ogura K., Saijoh Y., Ogura T 2003 Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis. Development 130 5953-5964

Talkhabi M., Pahlavan S., Aghdami N., Baharvand H 2015 Ascorbic acid promotes the direct conversion of mouse fibroblasts into beating cardiomyocytes. Biochemical Biophysical Research Communications 463 (4): 699-705

Tandon N., Goller K., Wang F., Soibam B., Gagea M., Jain AK., Schwartz RJ., Liu Y 2019 Aberrant expression of embryonic mesendoderm factor MESP1 promotes tumorigenesis. EBioMedicine 50 55-66

Thattaliyath BD., Firulli BA., Firulli AB 2002 The basic-helix-loop-helix transcription factor HAND2 directly regulates transcription of the atrial naturetic peptide gene. Journal of molecular cellular cardiology 34 (10): 1335-1344

Thier M., Münst B., Edenhofer F 2010 Exploring refined conditions for reprogramming cells by recombinant Oct4 protein. International Journal of Developmental Biology 54 (11-12): 1713-1721

Thier M., Münst B., Mielke S., Edenhofer F 2012 Cellular reprogramming employing recombinant sox2 protein. Stem cells international 2012

Thool M., Dey C., Bhattacharyya S., Sudhagar S., Thummer RP 2021 Generation of a recombinant stem cell-specific human SOX2 protein from *Escherichia coli* under native conditions. Molecular Biotechnology 63 (4): 327-338

Tsumoto K., Ejima D., Senczuk AM., Kita Y., Arakawa T 2007 Effects of salts on protein–surface interactions: applications for column chromatography. Journal of pharmaceutical sciences 96 (7): 1677-1690

VanDusen NJ., Casanovas J., Vincentz JW., Firulli BA., Osterwalder M., Lopez-Rios J., Zeller R., Zhou B., Grego-Bessa J., De La Pompa JL 2014 Hand2 is an essential regulator for two Notch-dependent functions within the embryonic endocardium. Cell reports 9 (6): 2071-2083

Vasina JA., Baneyx F 1997 Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the *Escherichia coli* cspAandtacPromoter Systems. Protein Expression and Purification 9 (2): 211-218

Verrecchia F., Mauviel A., Rossert J 2001 Blocking sp1 transcription factor broadly inhibits extracellular matrix gene expression in vitro and in vivo: implications for the treatment of tissue fibrosis. Journal of Investigative Dermatology 116 (5): 755-763

Viger RS., Guittot SM., Anttonen M., Wilson DB., Heikinheimo M 2008 Role of the GATA family of transcription factors in endocrine development, function, and disease. Molecular endocrinology 22 (4): 781-798

Wada R., Muraoka N., Inagawa K., Yamakawa H., Miyamoto K., Sadahiro T., Umei T., Kaneda R., Suzuki T., Kamiya K 2013 Induction of human cardiomyocyte-like cells from fibroblasts by defined factors. Proceedings of the National Academy of Sciences 110 (31): 12667-12672

Wang H., Cao N., Spencer CI., Nie B., Ma T., Xu T., Zhang Y., Wang X., Srivastava D., Ding S 2014 Small molecules enable cardiac reprogramming of mouse fibroblasts with a single factor, Oct4. Cell Reports 6 (5): 951-960

Wang L., Huang P., Near D., Ravi K., Xu Y., Liu J., Qian L 2020a Isoform specific effects of Mef2C during direct cardiac reprogramming. Cells 9 (2): 268

Wang J., Jiang X., Zhao L., Zuo S., Chen X., Zhang L., Lin Z., Zhao X., Qin Y., Zhou X 2020b Lineage reprogramming of fibroblasts into induced cardiac progenitor cells by CRISPR/Cas9-based transcriptional activators. Acta Pharmaceutica Sinica B 10 (2): 313-326

Wang L., Liu Z., Yin C., Asfour H., Chen O., Li Y., Bursac N., Liu J., Qian L 2015 Stoichiometry of Gata4, Mef2c, and Tbx5 influences the efficiency and quality of induced cardiac myocyte reprogramming. Circulation Research 116 (2): 237-244

Wang L., Yang C., Li F., Mu D., Ran P., Shen H., Li W., Ma J., Wu J., Yang X 2020c High levels of MESP1 expression in non-small cell lung cancer can facilitate cell proliferation, metastasis and suppresses cell apoptosis. Translational Cancer Research 9 (10): 5956

Weill CO., Biri S., Adib A., Erbacher P 2008 A practical approach for intracellular protein delivery. Cytotechnology 56 (1): 41-48

Whitaker RH 2010 Anatomy of the heart. Medicine 38 (7): 333-335

Wilson-Rawls J., Molkentin JD., Black BL., Olson EN 1999 Activated notch inhibits myogenic activity of the MADS-Box transcription factor myocyte enhancer factor 2C. Molecular Cellular Biology 19 (4): 2853-2862

Wingfield PT 2015 Overview of the purification of recombinant proteins. Current protocols in protein science 80 (1): 6.1. 1-6.1. 35

Wong HH., Kim YC., Lee SY., Chang HN 1998 Effect of post-induction nutrient feeding strategies on the production of bioadhesive protein in *Escherichia coli*. Biotechnology Bioengineering 60 (3): 271-276

Wood DW 2014 New trends and affinity tag designs for recombinant protein purification. Current Opinion in Structural Biology 26 54-61

Wurm F., Bernard A 1999 Large-scale transient expression in mammalian cells for recombinant protein production. Current Opinion in Biotechnology 10 (2): 156-159

Xiang Q., Zhou D., He X., Fan J., Tang J., Qiu Z., Zhang Y., Qiu J., Xu Y., Lai G 2019 The zinc finger protein GATA4 induces mesenchymal-to-epithelial transition and cellular senescence through the nuclear factor-κB pathway in hepatocellular carcinoma. Journal of gastroenterology and hepatology 34 (12): 2196-2205

Xin M., Olson EN., Bassel-Duby R 2013 Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. Nature reviews Molecular cell biology 14 (8): 529-541

Xu D., Dwyer J., Li H., Duan W., Liu J-P 2008 Ets2 maintains hTERT gene expression and breast cancer cell proliferation by interacting with c-Myc. Journal of biological chemistry 283 (35): 23567-23580

Xu Z., Gong J., Maiti D., Vong L., Wu L., Schwarz JJ., Duh EJ 2012 MEF2C ablation in endothelial cells reduces retinal vessel loss and suppresses pathologic retinal neovascularization in oxygen-induced retinopathy. The American Journal of Pathology 180 (6): 2548-2560

Yamakawa H., Ieda M 2021 Cardiac regeneration by direct reprogramming in this decade and beyond. Inflammation Regeneration 41 (1): 1-10

Yamamoto H., Flannery ML., Kupriyanov S., Pearce J., McKercher SR., Henkel GW., Maki RA., Werb Z., Oshima RG 1998 Defective trophoblast function in mice with a targeted mutation of Ets2. Genes & development 12 (9): 1315-1326

Yoo SY., Jeong S-N., Kang J-I., Lee S-W 2018 Chimeric adeno-associated virus-mediated cardiovascular reprogramming for ischemic heart disease. ACS omega 3 (5): 5918-5925

Yoon BS., Yoo SJ., Lee JE., You S., Lee HT., Yoon HS 2006 Enhanced differentiation of human embryonic stem cells into cardiomyocytes by combining hanging drop culture and 5-azacytidine treatment. Differentiation 74 (4): 149-159

Young CL., Britton ZT., Robinson AS 2012 Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnology Journal 7 (5): 620-634

Yu J., Ma X., Cheung K., Li X., Tian L., Wang S., Wu C., Wu WKK., He M., Wang M 2010 Epigenetic inactivation of T-box transcription factor 5, a novel tumor suppressor gene, is associated with colon cancer. Oncogene 29 (49): 6464-6474

Yu JS., Palano G., Lim C., Moggio A., Drowley L., Plowright AT., Bohlooly-Y M., Rosen BS., Hansson EM., Wang Q-D 2019 CRISPR-knockout screen identifies Dmap1 as a regulator of chemically induced reprogramming and differentiation of cardiac progenitors. Stem Cells 37 (7): 958-972

Zang M-X., Li Y., Wang H., Wang J-B., Jia H-T 2004a Cooperative interaction between the basic helix-loop-helix transcription factor dHAND and myocyte enhancer factor 2C regulates myocardial gene expression. Journal of Biological Chemistry 279 (52): 54258-54263

Zang MX., Li Y., Xue LX., Jia HT., Jing H 2004b Cooperative activation of atrial naturetic peptide promoter by dHAND and MEF2C. Journal of cellular biochemistry 93 (6): 1255-1266

Zhang HS., Postigo AA., Dean DC 1999 Active transcriptional repression by the Rb–E2F complex mediates G1 arrest triggered by p16INK4a, TGFβ, and contact inhibition. Cell 97 (1): 53-61

Zhang Z., Zhang W., Nam Y-J 2019 Stoichiometric optimization of Gata4, Hand2, Mef2c, and Tbx5 expression for contractile cardiomyocyte reprogramming. Scientific reports 9 (1): 1-10

Zhao R., Watt AJ., Battle MA., Li J., Bondow BJ., Duncan SA 2008 Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. Developmental biology 317 (2): 614-619

Zhao X-L., Yang B., Ma L-N., Dong Y-H 2016 MicroRNA-1 effectively induces differentiation of myocardial cells from mouse bone marrow mesenchymal stem cells. Artificial Cells, Nanomedicine, and Biotechnology 44 (7): 1665-1670

Zhao Y., Londono P., Cao Y., Sharpe EJ., Proenza C., O'Rourke R., Jones KL., Jeong MY., Walker LA., Buttrick PM 2015 High-efficiency reprogramming of fibroblasts into cardiomyocytes requires suppression of pro-fibrotic signalling. Nature Communications 6 (1): 1-15

List of publications

Publications from Ph.D. Thesis

Research articles

1. **Haridhasapavalan, K. K.**, Sundaravadivelu, P. K., Joshi, N., Das, N. J., Mohapatra, A., Voorkara, U., Kaveeshwar, V., Thummer, R. P. (**2022**). Generation of a recombinant version of a biologically active cell-permeant human HAND2 transcription factor from *E. coli*. *Scientific Reports*. 12(1), 1-16 (Springer Nature). (IF: 4.996).
2. **Haridhasapavalan, K. K.**, Sundaravadivelu, P. K., Voorkara, U., Kaveeshwar, V., Thummer, R. P. (**2022**). Generation of the recombinant version of a bioactive human MEF2C transcription factor from *E. coli*. *Healthcare Research and Related Technologies - Proceedings of NERC 2022* (Springer Nature).
3. **Haridhasapavalan, K. K.**, Das, N. J., Thummer, R. P. (**2022**). Generation of transducible version of a bioactive recombinant human TBX5 transcription factor from *E. coli*. *Current Research in Biotechnology*. 4(2022), 66-77 (Elsevier).
4. **Haridhasapavalan, K. K.**, Sundaravadivelu, P. K., Bhattacharyya, S., Ranjan, S. H., Raina, K., Thummer, R. P. (**2021**). Generation of cell-permeant recombinant human transcription factor GATA4 from *E. coli*. *Bioprocess and Biosystems Engineering*. 44(6), 1131-46 (Springer Nature). (IF: 3.434).
5. **Haridhasapavalan, K. K.***, Ranjan, S. H.*, Bhattacharyya, S., Thummer, R. P. (**2021**). Soluble expression, purification, and secondary structure determination of human MESP1 transcription factor. *Applied Microbiology and Biotechnology*. 105(6), 2363-76. (Springer Nature). *Equal contribution. (IF: 5.560).
6. **Haridhasapavalan, K. K.***, Sundaravadivelu, P. K.*, Thummer, R. P. (**2020**). Codon Optimization, Cloning, expression, purification, and secondary structure determination of human ETS2 transcription factor. *Molecular Biotechnology*, 62(10), 485-494 (Springer Nature). *Equal contribution. (IF: 2.860).

Review articles / Book chapters

7. **Haridhasapavalan, K. K.**, Borthakur, A., Thummer, R. P. (**2022**). Direct cardiac reprogramming: Current status and future prospects. *Cell Biology and Translational Medicine. Series: Advances in Experimental Medicine and Biology*. (Springer Nature). (Accepted). (IF: 3.650).
8. Borgohain, M. P.*, Narayan, G.*, **Haridhasapavalan, K. K.**, Dey, C., Thummer, R. P. (**2018**). Maximizing expression and yield of human recombinant proteins from bacterial cell factories for biomedical applications. In Advances in Microbial Biotechnology (pp. 447-486). Apple Academic Press (Book Chapter). *Equal contribution.
9. **Haridhasapavalan, K. K.**, Borthakur, A., Thummer, R. P. An overview of non-integrative direct cardiac reprogramming approaches for prospective biomedical applications. (Under preparation).

Publications from other collaborative research

Research articles

1. Bag, S. S., Bora, A., Golder, A., Raina, K., **Haridhasapavalan, K. K.**, Thummer, R. P. (**2022**). Gelatin-Pva-Agnps triad composite as wound healing hydrogel with wounded skin surface protective efficiency. *SSRN Electronic Journal*. (Elsevier).

Review articles / Book chapters

2. Dey, C.*, Raina, K.*, Thool, M.*, Adhikari, P., **Haridhasapavalan, K. K.**, Sundaravadivelu, P. K., Venkatesan, V., Sudhagar S., Gogoi, R., Thummer, R. P. (**2022**) Auxiliary pluripotency-associated genes and their contributions in the generation of induced pluripotent stem cells (Book Chapter). Volume 12. Elsevier: Academic Press. ISBN: 9780323900591. *Equal contribution.
3. Dey, C.*, Raina, K.*, **Haridhasapavalan, K. K.***, Thool, M.*, Sundaravadivelu, P. K., Adhikari, P., Gogoi, R., Thummer, R. P. (**2021**) An overview of reprogramming approaches to derive integration-free induced pluripotent stem cells for prospective biomedical applications. Recent Advances in iPSC Technology, Volume 5. Advances in Stem Cell Biology. Elsevier: Academic Press. ISBN: 9780128222317 (Book Chapter). *Equal contribution.
4. **Haridhasapavalan, K. K.***, Raina, K.*, Dey, C.*, Adhikari, P., Thummer, R. P. (**2020**). An Insight into reprogramming barriers to iPSC generation. Stem Cell Reviews and Reports, 16(1), 56-81 (Springer Nature).*Equal contribution. (IF: 6.692).
5. Borgohain, M. P.*, **Haridhasapavalan, K. K.***, Dey, C., Adhikari, P., Thummer, R. P. (**2019**). An insight into DNA-free reprogramming approaches to generate integration-free induced pluripotent stem cells for prospective biomedical applications. Stem Cell Reviews and Reports, 15(2), 286-313 (Springer Nature).*Equal contribution. (IF: 6.692).
6. **Haridhasapavalan, K. K.***, Borgohain, M. P.*, Dey, C., Saha, B., Narayan, G., Kumar, S., Thummer, R. P. (**2019**). An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene, 686, 146-159 (Elsevier). *Equal contribution. (IF: 3.913)
7. Saha, B., **Haridhasapavalan, K. K.**, Borgohain, M. P., Thummer, R. P. (**2018**). Prospective applications of induced pluripotent stem cells in military medicine. Medical Journal Armed Forces India, 74(4), 313-320.
8. Dey, C.*, Narayan, G.*, **Haridhasapavalan, K. K.***, Borgohain, M. P.*, Lenka, N., Thummer, R. P. (**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. Life Sciences Group. *Equal contribution.

Conferences / workshop

1. **Haridhasapavalan, K. K.**, Sundaravadivelu, P. K., Voorkara, U., Kaveeshwar, V., Thummer, R. P. (**2022**). Generation of the recombinant version of a bioactive human MEF2C 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 (**Oral presentation**).

2. **Haridhasapavalan, K. K.**, Thummer, R. P. Production of Cell-permeant and Bioactive Recombinant Human MEF2C and HAND2 Transcription Factors from *E. coli*. Student Indian Peptide Symposium 2022 (Virtual SYMPOSIUM) organized by, Indian Institute of Technology Bombay from 31st March – 1st April, 2022 (**Oral presentation**).
3. **Haridhasapavalan, K. K.**, Thummer, R. P. Heterologous expression, purification, characterization and intracellular delivery of recombinant human GATA4. 1st Departmental Retreat (Biotech Express) organized by Departmrnt of Biosciences and Bioengineering, Indian Institute of Technology Guwahati on 21st December, 2019 (**Poster presentation**).
4. **Haridhasapavalan, K. K.**, Ranjan, S. H, Thummer, R. P. Generation of Transducible Recombinant Cardiac-Specific Transcription Factor GATA4. Research Conclave'19 organized by Students' Academic Board, Indian Institute of technology Guwahati from 14-17th March, 2019 (**Poster presentation**).
5. **Haridhasapavalan, K. K.** Participated in five-day workshop on "**ZE5 and Droplet Digital PCR – QX200**", organized by Bio-Rad Laboratories and Indian Institute of technology Guwahati from 30th October – 3rd November, 2017.
6. **Haridhasapavalan, K. K.** 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.
7. **Haridhasapavalan, K. K.**, Dey, C., Thummer, R. P. Direct reprogramming of human fibroblasts to functional cardiomyocytes using non-genetic approaches for cell therapy. Research Conclave'17 organized by Students' Academic Board, Indian Institute of Technology Guwahati from 16-19th March, 2017 (**Poster presentation**).

Awards

1. **Augmenting Writing Skills for Articulating Research (AWSAR) Award 2021**: Awarded for the best scientific story writing titled "Love failure may or may not be fixed, but heart failure can definitely be fixed", which includes a prize of ₹10,000/- along with a Certificate of Appreciation. The story will be published later in the book "AWSAR Awarded Popular Science Articles by Scientists for the People- 2021" by the Department of Science and Technology, Government of India. https://www.awsar-dst.in/results-2021.