CELLULAR AND PROCESS ENGINEERING TO IMPROVE

MAMMALIAN MEMBRANE PROTEIN EXPRESSION

By

Su Xiao

A dissertation is submitted to Johns Hopkins University in conformity with the

requirements for degree of Doctor of Philosophy

Baltimore, Maryland

May 2015

Abstract

Improving the expression level of recombinant mammalian proteins has been pursued for production of commercial biotherapeutics in industry, as well as for biomedical studies in academia, as an adequate supply of correctly folded proteins is a prerequisite for all structure and function studies. Presented in this dissertation are different strategies to improve protein functional expression level, especially for membrane proteins. The model protein is neurotensin receptor 1 (NTSR1), a hard-to-express G protein-coupled receptor (GPCR). GPCRs are integral membrane proteins playing a central role in cell signaling and are targets for most of the medicines sold worldwide. Obtaining adequate functional GPCRs has been a bottleneck in their structure studies because the expression of these proteins from mammalian cells is very low.

The first strategy is the adoption of mammalian inducible expression system. A stable and inducible T-REx-293 cell line overexpressing an engineered rat NTSR1 was constructed. 2.5 million Functional copies of NTSR1 per cell were detected on plasma membrane, which is 167 fold improvement comparing to NTSR1 constitutive expression.

The second strategy is production process development including suspension culture adaptation and induction parameter optimization. A further 3.5 fold improvement was achieved and approximately 1 milligram of purified functional NTSR1 per liter suspension culture was obtained. This was comparable yield to the transient baculovirus-insect cell system.

The third strategy is high throughput miRNA screening. MiRNAs are a novel class of small, non-coding RNAs that can simultaneously silence multiple genes. The

NTSR1-expressing cell line was subjected to human miRNA mimic library screening and nine miRNA mimics were identified to improve functional expression of NTSR1 by as much as 48%. Interestingly, five out of nine identified miRNA mimics were effective in improving the functional expression of other proteins, including luciferase (cytosolic protein), serotonin transporter (membrane protein) and glypican-3 hFc protein (secreted protein). These indicated that the identified miRNAs could have a wide role in enhancing production of proteins with biomedical interest.

As genome-wide siRNA screens has emerged to be a powerful methodology for deducing gene functions in various diseases, we applied this technology on HEK293 cells constitutively expressing luciferase reporter to generate a genome-wide profile for recombinant protein expression process. Up to 362 genes associated with significantly enhanced luciferase expression were discovered and 28 of them were found to be enriched in splicesome pathway. Moreover, top 10 genes leading to greatest improvement of luciferase production were validated and tested with secreted and membrane proteins.

Investigation of these genes/pathways may provide profound information to understanding protein biosynthesis process in mammalian cells. *Oaz1* gene for example, was chosen for further investigation. *Oaz1* encodes ornithine decarboxylase (ODC) antizyme 1, a major negative regulator of ODC and cellular polyamines. In our study, it is found that when antizyme was depleted, ODC enzyme and cellular polyamines levels were up-regulated, leading to enhanced luciferase translation.

Advisors: Michael J. Betenbaugh and Joseph Shiloach

<u>Preface</u>

It has long been pursued by pharmaceutical and biotechnology industry to enhance the quality and quantity of recombinant proteins produced from host cells by traditional (e.g. media and bioprocess conditions optimization) and cell engineering approaches. In academia, the expression of recombinant proteins in general and mammalian proteins in particular, is also at the heart of current medical and structural studies.

This dissertation consists of six chapters and is mainly focused on the improvement of functional expression of a hard-to-express membrane protein-neurotensin receptor (NTSR1).

Chapter 1 provides a review on the ongoing effort to target bottlenecks along transcription, translation, protein processing and secretion pathways, as well as cell growth and survival to improve expression of low yielding proteins in a variety of hosts including bacterial, fungal, insect, and mammalian cells. The contents of Chapter 1 have been published in the *Current Opinion in Structural Biology* journal. Permission for its use was granted by the publisher, Elsvier (license number 3606050366032).

Chapter 2 describes the construction of a stable HEK293 cell line with high-level functional NTSR1 expression and a quantitative comparison with NTSR1 produced from insect cells. The text and illustrations used in Chapter 2 have been published in the *PLOS ONE* journal and *Methods in Molecular Biology* journal. Permission was granted by Springer (license number 3606050761810).

Chapter 3 features the identification of five microRNAs (hsa-miR-22-3p, hsa-miR-22-5p, hsa-miR-18a-5p, hsa-miR-429 and hsa-miR-2110) that can further improve NTSR1 functional expression in HEK293 cells. Their wider application was demonstrated by expression enhancement of other cytosolic, secreted and membrane proteins. The contents of Chapter 3 have been published in the *Biotechnology and Bioengineering* journal. Permission for its reuse was granted by John Wiley and Sons (license number 3606050872125).

Chapter 4 extends the scope to system biology of HEK293 cells and identified 10 significant influential genes for heterologous protein production by genome-wide loss-of-function studies.

Chapter 5 focuses on one of the identified top 10 genes (*oaz1*) and investigated the mechanism of improved protein expression upon *oaz1* gene knockdown. A manuscript is being prepared based on contents of chapter 4 and 5.

Finally, Chapter 6 concludes this dissertation and suggests the future work to extend the efforts.

<u>Acknowledgements</u>

First of all I would like to thank my family, for their encouragement and support in my journey to pursue a doctoral degree. I would like to thank my father, who ignited my dream of becoming an engineer and encouraged me to apply for doctoral education. He has always been an inspiring example of pursuing his dream through thick and thin, and of building his success from ground up. He has always been a strong source of mental support and motivation which carried me through the moments of discouragement and irresolution. I would love to thank my mother, who instilled in me the mind of independence and self-acceptance. She has always been very understanding, supportive and caring along this long journey. Words alone cannot express my deep love and appreciation for her.

I would like to express my sincere appreciation to my advisors, Dr. Michael J. Betenbaugh and Dr. Joseph Shiloach for their wisdom, mentorship and support. Their vision and commitment to research has inspired and will keep motivating me to excel in my scientific research. Their guidance in the past five years helped me to mature professionally and personally. It was an great honor and privilege to get to know them and work with them.

I would also like to thank all my colleagues both at Johns Hopkins University and NIDDK Biotechnology Core Laboratory for their guidance and advices. Especially, I would love to thank Dr. Loc Trinh, Dr. Alejandro Negrete, Dr. Alex Druz, Dr. Bojiao Yin, Dr. Melissa St Amand, Joseph Priola, Amit Kumar and Sarah Inwood for their help

Table of Contents

List of Tables

List of Figures

Hygro cell line.

Chapter 1: Engineering cells to improve recombinant protein expression

Abbreviations

MABs: monoclon al antibodies; CHO: Chinese Hamster Ovary; CMV, Cytomegalovirus; MAR, matrix attachment region; SEAP: secreted alkaline phosphatase; ER: endoplasmic reticulum; SERT: serotonin transporter; SRP: signal recognition particle; TF: trigger factor, ORF, open reading frame; eIF, eukaryotic initiation factor; HSA, human serum albumin; t-PA, tissue-plasminogen activator; NTSR1, neurotensin receptor type I.

1.1 - Summary

Improving the expression level of recombinant mammalian proteins has not only been pursued by biotechnologist for production of commercial biotherapeutics, but has also been at the heart of numerous biomedical studies in academia, as an adequate supply of correctly folded proteins is a prerequisite for all structure and function studies. In industry, there have been consistent efforts made towards process development for recombinant protein therapeutics production, especially on culture medium optimization and feeding strategies. In addition, significant progress has been achieved for the past decades in engineering and development of new cell lines to improve recombinant protein production for structural, biochemical, and commercial applications.

This chapter is a review on recent advances of improving recombinant protein production by engineering a variety of expression host including bacteria, fungi, insect cells and mammalian cells. With increased understanding of the host organism biology,

engineering strategies were developed targeting bottlenecks in transcription, translation, protein processing and secretory pathways, as well as cell growth and survival. A combination of metabolic engineering and synthetic biology has been used to improve the properties of cells for protein production, which has resulted in enhanced yields of multiple protein classes.

1. 2 - Introduction

With hundreds of billion dollar global market[1] and 10-20% annual increases in revenue worldwide[2], protein-based therapeutics have received unprecedented recognition of success and great potential. Since the production of the first approved recombinant protein (insulin) in the early 1980s, there have been continuous efforts to improve recombinant protein productivity and quality, especially for monoclonal antibodies (mABs), which accounts for approximately half of the sales[1]. Process development is very effective to improve productivity, especially the optimization of culture medium and feeding strategy. In industry, up to 10 grams of mAB or Fc fusion proteins can be produced in optimized fed-batch process with chemically-defined media[3]. In parallel, cell line engineering and development is also vital for cell-based production process. Development of cell lines with improved stability and protein productivity, quality, and biosimilarity has been pursued for decades.

Expression of recombinant mammalian proteins is also at the heart of medical and structural studies in academia. Although cloning, expression and production methods are

available for many hosts [4-6], and significant progress has been achieved to produce thousands of recombinant proteins for structural and biochemical studies, it has been challenging to produce many difficult proteins, such as membrane proteins, large protein complexes and post-translationally modified proteins [7, 8].

A critical area is mammalian integral membrane proteins such as receptors, ion channels and transporters which are encoded by 20-40% of all open reading frames (ORFs) in the mammalian genome[9] and are targets of most of the medicines sold worldwide[10]. Even though more than 100,000 structures have been deposited in Protein Data bank, the overexpression of membrane protein remains difficult [8, 11] and only 952 membrane protein structures are available as of April, 2015 (http://blanco.biomol.uci.edu/mpstruc/). Rational attempts to improve membrane protein expression may not lead to expected results as membrane proteins involve particularly complex folding, assembly, and processing pathways, and there is only limited information for the bottlenecks that may reside in the protein production steps, such as transcription, translation, protein folding, secretion and cell viability.

This chapter will summarize the most recent successful cases where the protein productivity was improved through different cellular engineering strategies. Presented here are a variety of host cells, e.g. bacteria, fungi, insect cells and mammalian cells. Different protein types are all included, e.g. secreted, cytosolic and hard-to-express membrane proteins. Cellular engineering strategies reviewed in this chapter were categorized by the biological process engineered, including transcription, translation, protein processing and secretory pathways, as well as cell growth and survival.

3

1.3 - Choice of host cell line

The first step in any strategy to over-produce proteins is the selection of the expression host. Table 1.1 includes 11 biopharmaceutical products selected from GEN's top 25 best-selling drug list from 2013. Seven out of eleven products were produced from mammalian expression system (mostly from Chinese Hamster Ovary cells) and the annual revenue generated using mammalian system took up 74% of overall revenue from the 11 top-selling biopharmaceutical products.

However in academia, especially in crystallography studies, mammalian cells are not widely used. According to statistics from the Protein Data Bank (http://www.rcsb.org) and the 'Membrane Proteins of Known Structure Database' (http://blanco.biomol.uci.edu/mpstruc/), of all the proteins that had their structures determined between 2004 and 2014, 78% were expressed in *Escherichia coli* and only 1.8% in mammalian cells (Fig.1.1A). For the overexpression of membrane proteins, *E. coli* was utilized on average less (61%) and eukaryotic expression systems were used comparatively more (Fig. 1.1B). Notably, there is an increasing trend in the use of more complex eukaryotic hosts (insect and mammalian cells, Fig. 1.1C), which reflects an increase in the number of mammalian membrane proteins being crystallized, particularly G protein-coupled receptors (GPCRs) [12].

1.4 - cell engineering strategies

1.4.1. Optimizing transcription

One of the most important choices in planning a strategy for overexpression of proteins is the type of promoter to use, and it is often the case that the strongest promoter will be the best for producing large amounts of correctly folded protein. Thus the most commonly used promoters are the T7 promoter in *E. coli*, the polyhedrin promoter in the baculovirus expression system and the Cytomegalovirus (CMV) promoter in mammalian cells. If transcription is the rate limiting step in protein production, even after choosing a strong promoter, then increasing further the strength of the promoter may be effective. For example, Quilici *et al*. constructed a strong *CMV* promoter variant through introducing a 200-nucleotide deletion of intron A that increased luciferase expression up to 2 fold in mammalian cells [13]. However, recent studies have shown that increasing the amount of mRNA encoding the protein of interest does not necessarily lead to improved protein production in *E.coli*[14] or insect cells [15]. In these instances, it is possible that the rate limiting step is protein folding, perhaps due to limitations in host cell factors, such as molecular chaperones. Enhancements in protein expression can be achieved through reducing the rate of transcription, either by substituting a strong promoter with a weaker one [14, 15], or by weakening a strong promoter by introducing a point mutation[16]. An alternative approach is to reduce the levels of polymerase in the host cell. For example, the levels of the T7 DNA polymerase expressed in *E. coli* can be modulated by altering the expression levels of the natural inhibitor T7 lysozyme, which is under the control of a tightly regulated inducible promoter, hence fine-tuning the rates of transcription. Wagner *et al*. improved expression of 14 membrane proteins using this methodology [17].

Even when strong promoters are used, host cell factors can result in low rates of transcription. For example, during the construction of stable mammalian cell lines with the gene of interest expressed from the CMV promoter, poor expression could result from epigenetic silencing of the promoter. This can be alleviated by engineering the nuclear matrix attachment region (MAR) [18] or by combining a MAR with a mammalian replication initiation region (IR) [19, 20], consequently improving recombinant protein production in mammalian cell lines.

1.4.2 Enhancing translation

Translation of the gene of interest may also be inhibited by host cell silencing processes during protein production. For example, eukaryotic translation initiation factor 2(eIF2) may become phosphorylated after DNA plasmid transfection or upon virus transduction, which will inhibit translation and thus decrease protein expression. However, viruses have evolved mechanisms to circumvent this. Gantke *et al.* co-expressed the Ebola virus protein 35, which is a viral protein that prevents translational silencing, and increased recombinant protein production by 10-fold [21]. An alternative approach to circumvent translational silencing in insect cells following baculovirus infection is to co-express eIF4E, which resulted in a 2-fold increase in the production of a secreted alkaline phosphatase (SEAP)-EGFP fusion protein (SEFP)[22].

1.4.3 Folding and secretory pathway engineering

Molecular chaperones have been applied to improve protein production in various systems, where they act to preserve nascent proteins in a folding-competent conformation

and prevent aggregation [23]. The most extensively used chaperone systems that have facilitated protein production in *E.coli* are DnaK-DnaJ-GrpE and GroEL-GroES[24, 25]. In insect cells, host protein biosynthesis shuts down as a result of infection by the recombinant baculovirus, which can adversely affect levels of molecular chaperones important for the folding of secreted proteins and membrane proteins in the endoplasmic reticulum (ER), particularly in relation to the high levels of protein synthesis resulting from high mRNA levels produced from the polyhedrin promoter. Hence, co-expression of the membrane-bound molecular chaperone calnexin enhanced the expression of functional serotonin transporter (SERT) by nearly 3 fold [26], and co-expression of the soluble molecular chaperone calreticulin increased secretion of SEAP-EGFP fusion protein in insect cells [22]. Whether a lack of appropriate molecular chaperones in heterologous systems contributes to low levels of functional protein sometimes is difficult to assess. However, overproduction of mammalian calnexin in the yeast *Hansenula polymorpha* did increase production of the truncated glycoprotein of rabies virus[27], suggesting that at least in this case the folding environment in the yeast ER was not optimal for folding large amounts of glycoprotein.

However, co-expression of molecular chaperones is not a panacea and does not often give a 10-fold or more improvement in expression levels. Part of the problem is that overexpression of ER resident chaperones such as calreticulin might burden the ER and activate an unfolded protein response [22]. Another more challenging issue is that molecular chaperones may act in a concerted fashion to promote protein folding in a poorly understood process, suggesting that it may be best to overexpress multiple chaperones simultaneously. However, expression levels will need to be tightly controlled

to prevent overwhelming the cells protein production resource and also the stoichiometry between chaperones will have to be regulated. Another problem associated with engineering the chaperone and secretory pathway is that it can be protein and host specific. For example, co-expression of protein disulfide isomerase increased yields of albumin fusion proteins in the yeast *Pichia pastoris*[28] but did not improve functional SERT expression in insect cells [26]. Similarly, SRP 14 overexpression led to a substantial improvement of IgG production in CHO cells, but the strategy was ineffective in human cell lines producing alkaline phosphatase [29, 30].

An alternative strategy to overexpressing molecular chaperones is to delete endogenous competing chaperones in order to channel the nascent peptide chain to the desired signal recognition particle (SRP) secretory pathway. Indeed, Nannenga *et al.* showed that membrane protein insertion in *E. coli* improved and expression levels increased through eliminating competition between trigger factor (TF) and the SRP for the nascent polypeptide chain [16, 31].

Another strategy to improve secretion is to improve vesicular trafficking from the ER to the cell surface. Co-expression of secretory proteins which modulate vesicle trafficking, such as soluble NSF receptor (SNARE) proteins (SNAP-23 or VAMP8), improved production of SEAP and monoclonal antibodies by 2-3 fold in mammalian CHO-K1 cells [32]. Likewise, overexpression of SNARE-interacting Sec1p and Sly1p proteins improved expression of α-amylase and human insulin precursor in *Saccharomyces cerevisiae*[33]. In addition, the ceramide transfer protein S132A mutant

improved production of tissue-plasminogen activator (t-PA) [34], human serum albumin (HSA) and monoclonal antibodies in CHO [35].

1.4.4 Protein sequence mutagenesis

Mutating the sequence of the protein target can also improve expression levels of the target protein. Sometimes this may be achieved through rational approaches such as analyzing the structure of the protein, as in the D500G mutation of laccase in *E.coli* [36] and the cysteine mutation of coagulation factor VIII [37]. However, in many instances there is insufficient evidence to suggest why a protein does not overexpress, so high-throughput mutagenic strategies can be used. For example, directed evolution coupled with random mutagenesis, followed by screening and selection was used by Sarkar *et al.* to evolve a GPCR, the rat neurotensin receptor type I (NTSR1) in *E. coli*. A mutant with 14 nucleotide substitutions retained the biochemical properties of the wild type receptor together with a 10-fold increase in functional expression and slightly increased thermostability [38]. Similarly, Heggeset *et al.* applied combinatorial mutagenesis and selection based on ampicillin tolerance in *E.coli* to evolve the signal sequence of β-lactamase and improved SEAP production up to 8-fold [39].

In theory, a more elegant and simple strategy would be to use *in vivo* mutagenesis coupled to screening or selection to improve expression. This approach was used by Majors *et al.* to evolve an anti-apoptotic gene *Bcl-x$_L$* in a mammalian expression system by harnessing the somatic hypermutation capability of human Ramos B-cell line. The *Bcl-x$_L$* gene, coupled to the yellow fluorescent protein reporter, was mutated "*in situ*" and

subjected to rounds of staurosporine treatment to identify mutants with reduced apoptosis activation and higher YFP-Bcl-x$_L$ expression levels [40].

1.4.5 Cell proliferation and survival engineering

The delay or prevention of the apoptosis cascade activation has been successful in preventing cell death and improving protein production in CHO cells under stress conditions [41]. Co-expression of the anti-apoptotic protein Bcl-x$_L$ in CHO cells improved the expression of epidermal growth factor receptor, fibroblast growth factor receptor 3 and receptor tyrosine kinases proteins [42]. Knock-out of the genes encoding the pro-apoptotic factors Bax and Bak in a CHO-K1 cell line improved cell viability, reduced levels of transfection-induced apoptosis and led to up to 4 fold higher antibody titers [43]. Similarly, stable inhibition of the pro-apoptotic microRNA mmu-miR-466h-5p in CHO cells delayed the onset of apoptosis, increased the maximum viable cell density and enhanced expression of SEAP [44].

Enhanced cell proliferation represents another potential approach to increase biomass and obtain higher volumetric yield during large scale production processes. For example, a metabolically engineered respiratory strain of *S. cerevisiae* (TM6*) doubled volumetric yield of Fps1 and at least quadrupled the yield of two human GPCRs (A$_{2a}$R and CNR2)[45]. Overexpression of the mammalian target of rapamycin (mTOR) simultaneously improved cell growth, proliferation, viability and specific productivity of antibody, SEAP and secreted α-amylase in CHO cells [46]. Similarly, overexpression of miR-7 in CHO cells enhanced cell proliferation, leading to higher Epo-Fc titer [47]. However, accumulated biomass does not always lead to increased production as

demonstrated by chemical inhibition of autophagy in CHO cells, which led to decreased cell concentration but a 2.8 fold increase in t-PA [48].

1.4.6 Other strategies:

In cases where the heterologous proteins are toxic to the host cells, the presence of inhibitors can protect the host by sequestering proteins and keeping them in an inactive state. For example, co-expression of lysozyme together with its inhibitor Ivy, repressed lysozyme lytic activity in cytoplasm, and, along with transcription enhancement and chaperone co-expression, remarkably improved soluble lysozyme production in *E.coli*[49].

1.4 - Conclusion

Recombinant protein expression has facilitated biochemical and structural studies of thousands of naturally low abundance proteins. Methodologies that improve expression levels can be particularly advantageous for many difficult-to-produce proteins or if the protein is being produced for therapeutic or industrial purposes. To improve expression levels further through cell engineering requires an understanding of both the host organism and the biology of protein expression. In this chapter, we reviewed some recent successful cases that target potential bottlenecks in protein production, using strategies focused on optimizing transcription, translation, engineering the folding and secretory pathways, mutating the target protein sequence, and enhancing cell proliferation and/or survival. Considerable effort has been focused on engineering *E.coli* and yeast

strains, and now there is an expanding effort to engineer insect and mammalian hosts such as HEK293 and CHO cell lines[50, 51], especially for functional expression of mammalian membrane proteins that include particularly complex folding, assembly, and processing pathways [52-54].However, in many instances there is only limited information on the factors that affect expression of any particular protein, so current strategies are often piecemeal and focus on only one or two aspects of the protein production process. A goal for the future is to identify limiting bottlenecks in the protein expression process such as transcription, translation, protein folding, secretion and cell viability and construct robust cell factories through a holistic approach that considers all the bottlenecks and engineer these through an integrative process to enable high-level expression of a wide spectrum of target proteins.

Figures and tables

Fig.1.1. Summary of host cell line usage for production of recombinant proteins in structural studies between 2004 and 2014. (A) Break down of leading host cell choices for the expression of all types of proteins (B) Break down of leading expression organisms for integral membrane protein production. (C) Increasing application of higher eukaryotes (insect and mammalian cells) for recombinant protein production.
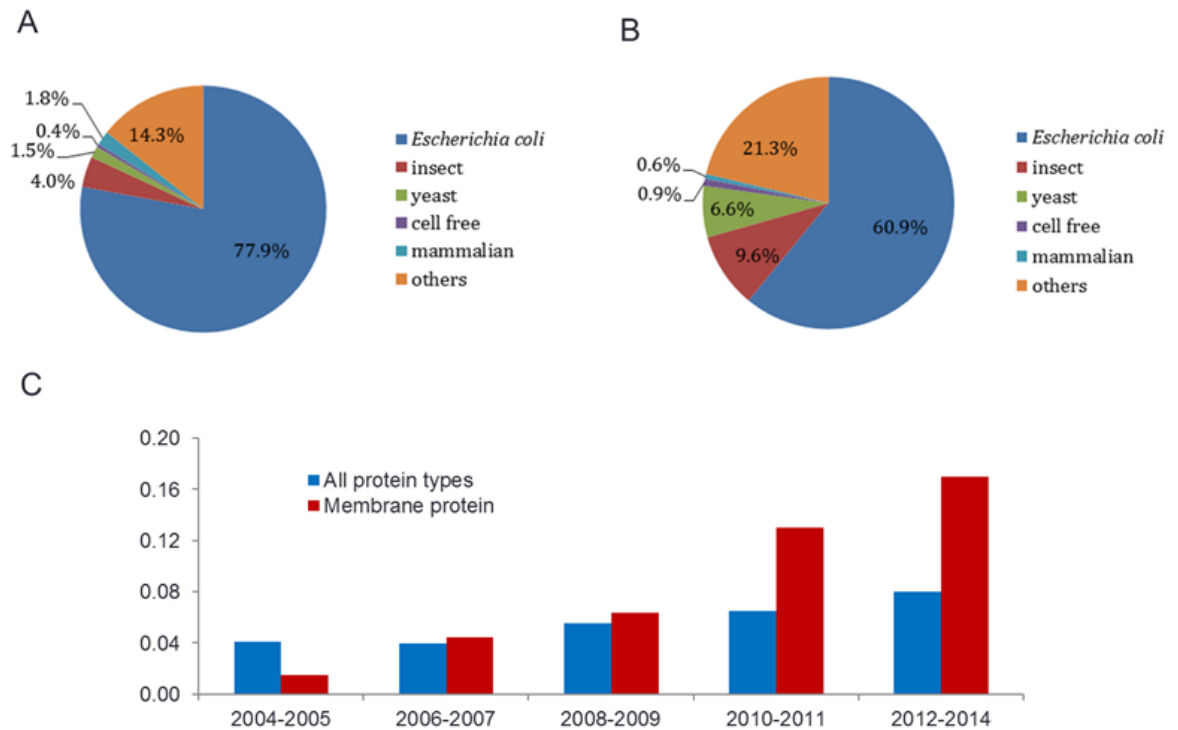
Table.1.1 Top selling biopharmaceutical products in 2013 (EU and US market)

Rank	Product	Revenue US $ (billion)	Manufacturer	Type of molecule	Expression system	Indication
1	Humira	10.66	Abb Vie	Human Mab	mammalian	Rheumatoid arthritis
2	Remicade	8.94	Johnson &Johnson and Merck& Co	Chimeric Mab	mammalian	Rheumatoid arthritis
3	Rituxan	8.92	Roche (Genentech) and Biogen Idec	Chimeric Mab	mammalian	Rheumatoid arthritis
5	Enbrel	8.33	Amgen and Pfizer	Mab fusion protein	mammalian	Rheumatoid arthritis
6	Lantus	7.85	Sanofi	Modified insulin	E.coli	Diabetes
7	Avastin	7.04	Roche	Humanized Mab	mammalian	cancer
8	Herceptin	6.84	Roche	Humanized Mab	Mammalian	Breast cancer
14	Neulasta	4.39	Amgen	PEGylated GCSF	E.coli	Neutropenia
18	Prevnar	3.97	Pfizer	Streptococcus pneumonia vaccine conjugate	Bacterial culture	Prevention of invasive pneumococcal disease
24	NovoLog/ NovoRapid	3.10	Novo Nordisk	Modified insulin	Saccharomyces cerevisiae	Diabetes

Table.1.2. Improvements in Protein Expression Levels for Different Cell Engineering Strategies

Protein	Location	Expression host	Fold increase in protein production	Reference
Strategy 1: optimizing transcription and enhancing translation				
Luciferase	intracellular	CHO-K1, HepG2, HEK-293, COS-7	3	[13]
D-amino acid oxidase	intracellular	*E.coli*	20	[14]
Glutaryl-7-aminocephalosporanic acid acylase	intracellular	*E.coli*	2	[14]
N-carbamyl-D- amino acid amidohydrolase	intracellular	*E.coli*	1.3	[14]
Secreted alkaline phosphatase	extracellular	Insect cells	significant	[15]
Deltarhodopsin	membrane	*E.coli*	5	[16]
Sensory rhodopsin II	membrane	*E.coli*	5	[16]
14 different membrane proteins	membrane	*E.coli*	significant	[17]
Cyclooxygenase-1	extracellular	HEK293T	significant	[20]
Antibody	extracellular	COLO 320DM	>8	[19]
		CHO DG44	>20	[19]
Tumor progression locus 2 complex	intracellular	HEK-293	10	[21]
TBK1	intracellular	HEK-293	n.r.	[21]

Lck	membrane	HEK-293	n.r.	[21]
CD40	membrane	HEK-293	n.r.	[21]
Bcl-2	membrane	HEK-293	n.r.	[21]
SEAP- EGFP fusion protein	extracellular	Insect cells	2	[22]
Strategy 2: Folding and secretory pathway engineering				
Secretory alkaline phosphatase- EGFP fusion protein	extracellular	Insect cells	2	[22]
Human papillomavirus 16 E7 oncoprotein fused to C-terminus of Tobacco mosaic virus coat protein	intracellular	*E.coli*	n.r.	[24]
Aldehyde dehydrogenase 3A1	intracellular	*E.coli*	4.9	[25]
Serotonin transporter	membrane	Insect cells	3	[26]
Glycoprotein of rabies virus (truncated)	extracellular	*H. polymorpha*	n.r.	[27]
ZraS	membrane	*E.coli*	3.6	[31]
Deltarhodopsin	membrane	*E.coli*	3.6	[31]
Sensory rhodopsin II	membrane	*E.coli*	3.4	[31]
SEAP	extracellular	CHO-K1	2	[32]
Antibody	extracellular		3	[32]

16

α-amylase	extracellular	*S. cerevisiae*	1.68	[33]
Insulin precursor	extracellular	*S. cerevisiae*	1.3	[33]
t-PA	extracellular	CHO	1.35	[34]
HSA	extracellular	CHO	1.6	[35]
Antibodies	extracellular	CHO	1.26	[35]
Interleukin-1 receptor antagonist – HSA	extracellular	*P. pastoris*	3.7	[28]
HSA- human growth hormone	extracellular	*P. pastoris*	4	[28]
Antibody	extracellular	CHO	4-7	[29]
Strategy 3: Protein sequence mutagenesis				
Benzenediol- oxygen oxidoreductase	intracellular	*E.coli*	3.14	[36]
Coagulation Factor VIII	extracellular	COS-1	1.3	[37]
		CHO	1.6	[37]
Neurotensin receptor	membrane	*E.coli*	10	[38]
Signal sequence of β-lactamase	intracellular	*E.coli*	5.5	[39]
YFP-Bcl-xL	membrane	Ramos B- cell	n.r.	[40]
		CHO	n.r.	[40]

Strategy 4: Cell proliferation and survival engineering

Epidermal growth factor receptor	membrane	CHO	significant	[42]
Fibroblast growth factor receptor 3	membrane	CHO	significant	[42]
Receptor tyrosine kinases proteins	membrane	CHO	significant	[42]
Antibody	extracellular	CHO-K1	4	[43]
Secreted alkaline phosphatase	extracellular	CHO	1.43	[44]
Glycerol transport facilitator Fps1	membrane	*S. cerevisiae*	2	[45]
A_{2a} adenosine receptor	membrane	*S. cerevisiae*	5	[45]
Cannainoid receptor 2	membrane	*S. cerevisiae*	4.5	[45]
Antibody	extracellular	CHO	4	[46]
SEAP	extracellular	CHO	3-7	[46]
Secreted α- amylase	extracellular	CHO	3-7	[46]
Epo-Fc	extracellular	CHO	n.r.	[47]

<u>Other strategies</u>

lysozyme	intracellular	*E.coli*	3000	[49]

Chapter 2: Stable expression of the neurotensin receptor NTSR1 with T-REx-293 cells and comparison with baculovirus- insect cell system

Abbreviations used:

GPCR, G protein-coupled receptor; NTSR1, neurotensin receptor type 1; NT: neurotensin; HEK, human embryonic kidney; CHO, Chinese hamster ovary; $tetO_2$, tandem tet operator; TetR, tet repressor protein; GnTI, N-acetylglucosamine transferase I; DMEM, Dulbecco's modified Eagle's medium; PBS, phosphate-buffered saline; FBS, fetal bovine serum; DM, *n*-Decyl-β-D-maltopyranoside; DDM, n-dodecyl-β-D-maltopyranoside; CHS, cholesteryl hemisuccinate Tris salt; CHAPS, 3-[(3-cholamidopyropyl) dimethylammonio] -1- propanesulfonate; NaBu, sodium butyrate; CMV, cytomegalusvirus; MOI, multiplicity of infection;

2.1 - Summary

G protein-coupled receptors (GPCRs) are associated with a wide array of diseases and are targets of most of the medicines sold worldwide. Despite their clinical importance, only 62 unique GPCR structures have been determined as of December 2014. The first step for structural studies is to establish the expression of correctly folded, functional receptors in recombinant host cells at quantities to allow subsequent purification and crystallization trials. Baculovirus- insect cell and tetracycline- inducible mammalian cell lines (T-REx-293) are intensively used for GPCR production. Here a stable and inducible T-REx-293 cell line overexpressing an engineered rat neurotensin receptor type 1 (NTSR1) was constructed. With proper clone selection, suspension

culture adaptation and induction parameter optimization, approximately 1 milligram of purified functional NTSR1 per liter suspension culture was obtained. This stable inducible mammalian expression system was also quantitatively compared with the transient baculovirus-insect cell system throughout a milligram-scale expression and purification process. The two systems were comparable on aspects of functional NTSR1 expression level and receptor binding affinity for ligand [^{3}H]NT. However, NTSR1 surface display on T-REx-293 cells determined by radio-ligand binding assays was 2.5 fold higher than that on insect cells. This work demonstrates two approaches for preparing milligram-scale quantities of NTSR1 purified enough for structural studies and provides useful input to users in choosing and optimizing an appropriate expression host for other GPCRs.

2.2 - Introduction

As was discussed in Chapter 1, there have been numerous efforts in improving the expression level of recombinant proteins for production of commercial biotherapeutics as well as for biomedical studies in academia. A critical and difficult area is the expression of mammalian integral membrane proteins such as receptors, ion channels and transporters. GPCRs are a superfamily of integral membrane proteins that have seven transmebrane domains. They play a central role in cell signaling by transmitting extracellular chemical signals across membranes to intracellular effector pathways. Closely associated with a wide array of physiological diseases, these proteins are targets of most of the medicines sold worldwide [10]. As of December 2014, structures of 62 unique GPCRs are available (http://blanco.biomol.uci.edu/mpstruc/listAll/list).

Crystallographic studies of GPCRs are very rewarding because they provide insight into mechanistic aspects of cell signaling at high resolution, thus opening the pathway to ultimately improve drugs and medicines targeting a wide variety of diseases[10]. In the past few years, we started to see an explosion in the field of GPCR structure determination. This exciting progress is based on a tremendous amount of methods development, such as advances in crystallization methods, the development and implementation of the concept of conformational thermostabilization of GPCRs, and the development of microfocus x-ray synchrotron technologies[55]. Yet, the supply of ample amounts of correctly folded receptors is the key prerequisite for successful structural studies.

The goal of this research is to improve expression level of functional neurotensin receptor type 1 (NTSR1), a difficult-to-express GPCR. It's agonist is neurotensin (NT), a 13 amino acid residue peptide that is found in the nervous system and in peripheral tissues [56]. NT displays a wide range of biological activities and plays important roles in Parkinson's disease, in pathogenesis of schizophrenia, in modulation of dopamine neurotransmission, hypothermia, antinociception and in promoting the growth of cancer cells [57-61]. Three neurotensin receptors have been identified. NTSR1 and NTSR2 belong to the class A GPCR family, whereas NTSR3 is a member of the sortilin family with a single transmembrane domain [62-64]. Most of the known effects of NT are mediated through NTSR1[60].

Wild-type rat NTSR1 has previously been expressed in *Escherichia coli* fused with maltose-binding protein (MBP), and large-scale purification has been accomplished [65]. Recently, systematic scanning mutagenesis of NTSR1 has been performed using *E.*

coli as the expression host to identify stabilized NTSR1 mutants suitable for crystallization [66, 67]. The structure of stabilized NTSR1 mutant (GW5) with T4 lysozyme replacing most of the third intracellular loop, was determined with receptors transiently expressed in baculovirus-insect cell system in 2012 [67].

In this work, mammalian cell line HEK293 was chosen as the expression host despite it's not a popular choice for structural studies[7], likely as the result of the high cost associated with serum needed for cell culture and the lengthy process of stable cell line construction. However, mammalian cells are ideally suited for efficient expression of functional membrane proteins because of the near-native environment they provide, such as N-glycosylation, a machinery for post-translational modification, molecular chaperones, and a suitable lipid environment [7]. In addition, the continuous production capability is yet another advantage which the baculovirus-insect cell system cannot provide.

As constitutive mammalian expression systems sometimes fail to provide adequate amounts of membrane proteins for structural studies, a tetracycline-inducible expression system [68] was demonstrated to be advantageous for high-level expression of GPCRs in functional form [50-52]. In tetracycline-free medium, mammalian cells are allowed to reach high cell density without the stress from leaky GPCR expression; after addition of tetracycline and thus GPCR production for typically 24 –72 hrs, cells are harvested.

The tetracycline-inducible expression system developed for recombinant mammalian cell expression hosts contains two components of the tetracycline-resistance determinants of gram-negative bacteria: The tandem tet operator sequence (tetO2)[69],

which is positioned upstream of the gene of interest; and the Tet repressor protein (TetR)[70]. In the absence of tetracycline, the TetR homodimer binds with high affinity to the tetO2 sequence downstream of the TATA element of the human cytomegalusvirus (CMV) major immediate-early promoter, thus blocking transcription of the gene of interest (Fig.2.1A). Upon addition of tetracycline, TetR abolishes its association with tetO2, allowing transcription of the gene of interest under the control of the strong CMV promoter [70] (Fig.2.1B).

This system, developed by Yao *et al*.[68], has been commercialized as the T-REx system (Life Technologies) and several tetracycline-inducible mammalian cell lines (HEK293, CHO, Hela, *etc*.) are now available. Those cell lines have been stably transfected with the pcDNA6/TR regulatory vector which leads to high level, constitutive expression of TetR [70]. The gene of interest is cloned into pcDNA™4/TO (or an equivalent plasmid), which contains the complete CMV enhancer-promoter sequence, with the tetO2 operator region starting 10 nucleotides downstream of the last nucleotide of the TATA element [68].

The successful application of the tetracycline-inducible system for GPCR expression was first demonstrated by Reeves *et al*. [51]. A tetracycline-inducible and suspension adaptable HEK293S-TetR cell line successfully overproduced milligram quantities of opsin mutants per liter of culture in bioreactors. Meanwhile, a HEK293S-GnTI⁻-TetR cell line, which lacks the N-acetylglucosaminetransferase I (GnTI) enzyme, was developed and gained increasing popularity for membrane protein production for structural studies [52-54], owing to the homogeneously N-glycosylated proteins produced by this cell line [51].

23

Here we report the expression of NTSR1-GW5-Δi3 in a stable, inducible T-REx-293 cell line and in baculovirus infected insect cells. NTSR1-GW5-Δi3 has six stabilizing mutations[67], truncated N- and C-termini, and parts of the third intracellular loop deleted. We provide a quantitative comparison between the two production hosts regarding aspects of functional NTSR1-GW5-Δi3 expression levels and receptor yield after purification, as well as binding properties and cell surface display of the receptors. The scale-up of NTSR1-GW5-Δi3 production using T-REx-293 suspension cultures in a bioreactor allows the continued production of the receptor suitable for the application of biophysical analyses such as *Nuclear magnetic resonance (*NMR) spectroscopy.

2.3 - Materials and Methods

2.3.1. Materials

The tritiated agonist [^{3}H]NT ([3,11-tyrosyl-3,5-3H(N)]-pyroGlu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu) was purchased from Perkin Elmer. Unlabeled NT was synthesized by the Center for Biologics Evaluation and Research (Food and Drug Administration). The detergents n-dodecyl-β-D-maltopyranoside (DDM), 3-[(3-cholamidopyropyl) dimethylammonio] -1- propanesulfonate (CHAPS) and cholesteryl hemisuccinate Tris salt (CHS) were obtained from Anatrace.

2.3.2. The NTSR1 construct used for expression in insect cells and T-REx-293 cells

The construct NTSR1-GW5-Δi3 consisted of the hemagglutinin signal peptide and the Flag tag [71], followed by the stabilized rat neurotensin receptor NTSR1 (T43-K396 containing the mutations A86L, E166A, G215A, L310A, F358A, V360A) with the

intracellular loop 3 residues G275-E296 deleted [67]. A deca-histidine tag was present at the C-terminus. We refer to this construct in the following as NTSR1. For expression using the baculovirus-insect cell system, NTSR1 was subcloned into the transfer vector pFastBac1 (Invitrogen) thus placing NTSR1 under the control of the strong polyhedrin promoter. For stable expression in T-REx-293 cells, NTSR1 was subcloned into the plasmid pACMV-tetO (a kind gift from Dr. Philip J. Reeves) downstream of the tetracycline-controlled CMV promoter [51] (Fig.2.2).

2.3.3. Transient expression of NTSR1 in the baculovirus-insect cell system

Recombinant baculoviruses were generated using the pFastBac1 transfer plasmid system (Invitrogen). *Trichoplusia ni* cells were infected at a cell density of 0.8-1 million cells/ml with recombinant virus at a multiplicity of infection (MOI) of 5, and the temperature was lowered from 28°C to 21°C. Cells were harvested by centrifugation 48 hours post infection, resuspended in hypotonic buffer (10 mM Hepes pH 7.5, 10 mM $MgCl_2$, 20 mM KCl), flash-frozen in liquid nitrogen and stored at −80°C until use.

2.3.4. Stable expression of NTSR1 in the T-REx-293 system

The T-REx-293 cell line was maintained as an adherent culture in DMEM containing 10% certified FBS and 5 μg/ml blasticidin (Invitrogen). The cells were transfected with the plasmid pACMV-tetO-NTSR1 (Fig.2.2). using Lipofectamine 2000 according to the manufacturer's protocol (Life Technologies). One day after transfection, cells were transferred into fresh DMEM medium containing 800 μg/ml Geneticin (Cellgro) and the medium was replaced every three days. Two weeks later, fourteen cell

clones were separately expanded into two T-flasks each. Cells in one T-flask were harvested during the exponential growth phase and frozen in 10% DMSO for storage. Cells in the other T-flask were induced with 2 µg/ml tetracycline for 24 hrs, after reaching 80% confluency. Cells were then detached from the flask and washed with cold PBS. After adjusting the cell density to around one million cells per ml, protease inhibitors (Roche) were added and the cell suspension was frozen on dry ice in 1ml aliquots. NTSR1 expression levels were determined by [^{3}H]NT binding and the clone with the highest expression level was selected for further experiments.

2.3.5. Adaption of NTSR1-expressing T-REx-293 cells in to suspension culture

Three different medium were tested for suspension culture of NTSR1-expressing T-REx-293 cells: Freestyle^{TM} 293, CD OptiCHO^{TM} (Gibco) and pro293 CD^{TM} Medium (Lonza). The adherent growth medium (DMEM supplemented with 10% certified FBS, 5µg/ml blasticidin and 500µg/ml G418) was gradually replaced with suspension growth medium during subculture in T-flask. With the displacement of the media, increasing amount of viable detached cells can be collected and transferred into shake flask and maintained in suspension growth media supplemented with 1% certified FBS, 5 µg/ml blasticidin and 500 µg/ml G418 for suspension culture. The optimal media giving highest cell density and viability was chosen for further culture.

2.3.6. Growth of T-REx-293 cell line suspension culture in a bioreactor

The suspension-adapted T-REx-293 cells were grown in 5L of CD OptiCHO medium supplemented with 4 mM L-glutamine, 1% certified FBS, 100 unit/ml penicillin,

100 µg/ml streptomycin and 0.1% pluronic F-68 (Gibco), using a 10L glass bioreactor equipped with a pitch blade impeller connected to a Sartorius BDCU controller. The growth parameters were set to 37°C, pH 7, and 30% dissolved oxygen. The latter two parameters were maintained by interactive control delivery of air and CO_2 through direct sparge (up to 10 ml/min). The speed of the impeller was 80 rpm. The cell density at inoculation was 3×10^5 cells per milliliter. On day 5 after inoculation, expression of NTSR1 was induced by addition of 2 µg/ml tetracycline and 2.5 mM sodium butyrate (NaBu). Cells were harvested 36hrs after induction, re-suspended in hypotonic buffer (10 mM Hepes pH 7.5, 10 mM $MgCl_2$, 20 mM KCl), flash-frozen in liquid nitrogen and stored at −80°C until use.

2.3.7. Analytical solubilization of NTSR1

Cell pellets from 10 ml of suspension cultures were suspended in Tris-glycerol-NaCl buffer. Then the detergent DM and CHS was added to give a final buffer composition of 50 mM TrisHCl pH 7.4, 200 mM NaCl, 30% glycerol, 1% DM, and 0.1% CHS in a total volume of 2.5 ml. The samples were placed on a rotating mixer at 4°C for 1 hour. Cell debris and non-solubilized material were removed by ultracentrifugation (TL100 rotor, 60k rpm, 4°C, 30 min in Optima Max bench-top ultracentrifuge, Beckman), and the supernatants containing detergent-solubilized NTSR1 were used to determine the total number of expressed receptors by a detergent-based radio-ligand binding (see below).

2.3.8. Purification of NTSR1 produced in T-REx-293 and insect cells

All buffer volumes relate to 1L of original cell culture. T-REx-293 cells were thawed and the volume was brought to approximately 200 ml with hypotonic buffer. The cells were then re-suspended using a Turrax T-25 (IKA) homogenizer at 8,200 rpm for 2 min. After centrifugation (45Ti rotor, 40,000 rpm, 45 min, 4°C, Optima L90K, Beckman), the membranes were resuspended (Turrax T-25) in approximately 120 ml of high-salt buffer (10 mM Hepes pH 7.5, 1 M NaCl, 10 mM $MgCl_2$, 20 mM KCl) supplemented with DNaseI (final concentration 10 μg/ml) and AEBSF (100 μM), and centrifuged again. The high-salt washes were repeated 4 more times with the DNaseI addition omitted after the 2nd wash. All subsequent steps were performed at 4°C or on ice, and AEBSF (100 μM final concentration) was repeatedly added throughout the procedure. The washed membranes were resuspended in a final volume of 40 ml of buffer (62.5 mM TrisHCl pH 7.4, 625 mM NaCl, 37.5% glycerol) containing 10 μM neurotensin peptide. NTSR1 was extracted by drop-wise addition of 10 ml of a 5% DM / 0.3% CHS solution. After 2.5 hours, the sample was clarified by centrifugation (45Ti rotor, 40,000 rpm, 1 hour, Optima L90K, Beckman), adjusted with imidazole to a final concentration of 20 mM, and then passed through a 0.2 μm filter (Stericup). Next, the sample was loaded at a flow rate of 0.2 ml/min onto 2 ml Talon resin packed into an XK16 column (GE Healthcare) equilibrated with Talon-A buffer (50 mM TrisHCl pH 7.4, 30% glycerol, 500 mM NaCl, 20 mM imidazole, 0.1% DM / 0.01% CHS) containing 1 μM neurotensin peptide. After washing with 29 column volumes of buffer Talon-A, NTSR1 was eluted with Talon-B buffer (50 mM TrisHCl pH 7.4, 30% glycerol, 500 mM NaCl, 200 mM imidazole, 0.1% DM / 0.01% CHS) containing 5 μM neurotensin peptide.

Peak fractions were collected (5 ml) and analyzed. The purification of NTSR1 from insect cells was performed in a similar manner.

2.3.9. Protein analysis and radio-ligand binding assays

The protein content was measured according to the Amido Black method of Schaffner and Weissmann [72] with bovine serum albumin as the standard. Western blot analysis was performed as described [73] using the INDIA HisProbe-HRP reagent (Pierce) and the substrates 3,3'-diaminobenzidine tetrahydrochloride and H_2O_2.

[^{3}H]NT ligand-binding assays with intact cells were carried out in 500 µl of TEBB assay buffer (50 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.1% bovine serum albumin, 40 µg/ml bacitracin) containing 10 nM [^{3}H]NT and about 100,000 cells. After incubation for 2-4 hours on ice, separation of bound from free ligand was achieved by rapid filtration through GF/B glass fiber filters (Whatman) pretreated with polyethylenimine. The amount of radioactivity was quantified by liquid scintillation counting (Beckman LS6500). Non-specific [^{3}H]NT binding of 4160 dpm was subtracted from total binding to calculate the receptor density at the cell surface. The concentration of [^{3}H]NT used in these assays was four-fold above the apparent dissociation constant for membrane-bound NTSR1 [67] to allow high receptor occupancy, but it kept nonspecific [^{3}H]NT binding to a minimum.

Ligand-binding assays with detergent-solubilized receptors were carried out in TEBB assay buffer containing 0.1% DDM, 0.2% CHAPS, 0.04% CHS. For one-point assays, receptors were incubated with 2 nM [^{3}H]NT on ice for 1 hour in a volume of 150

μl. Separation of the receptor-ligand complex from free ligand (100 μl) was achieved by centrifugation-assisted gel filtration using Bio-Spin 30 Tris columns (BioRad), equilibrated with RDB buffer (50 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1% DDM, 0.2% CHAPS, 0.04% CHS). Non-specific [^{3}H]NT binding of 220 dpm was subtracted from total binding, and the amount of specifically bound [^{3}H]NT was then corrected for fractional occupancy (apparent dissociation constants of 0.22 nM and 0.57 nM for receptors produced in HEK-293 cells and insect cells, respectively).

For saturation binding experiments, the [^{3}H]NT concentration was varied from 0.2 nM to 10 nM. Non-specific [^{3}H]NT binding was determined in the presence of 50 μM unlabeled NT. Data were analyzed by nonlinear regression using GraphPad Prism software (version 4, GraphPad Software) and best fit to a one-site binding equation to determine the apparent dissociation constants for NTSR1 produced in insect and HEK-293 cells. Note that the saturation binding experiments using the NTSR1 mutant did not reach equilibrium within the incubation time because of the very slow agonist off-rates determined in a previous study [67]. Individual experiments were conducted as single data points.

2.4 - Results

2.4.1. NTSR1 expression in suspension T-REx 293 cells

For our study, we used the engineered rat neurotensin receptor NTSR1-GW5-Δi3 (here referred to as NTSR1). This construct consists of the hemagglutinin signal peptide and the Flag tag, followed by the conformationally thermostabilized rat neurotensin

receptor NTSR1-GW5 (T43-K396 containing the mutations A86L, E166A, G215A, L310A, F358A, V360A)[67, 74] with the intracellular loop 3 residues G275-E296 deleted. A deca-histidine tag was present at the C-terminus. The NTSR1 DNA was inserted into pACMV-tetO [51] using standard molecular biology techniques. The resulting expression vector pACMV-tetO-NTSR1 (Fig. 2.2) allows tetracycline inducible expression of NTSR1.

The N-terminally truncated rat NTSR1 with 6 stabilizing mutations was stably expressed in the tetracycline-regulated T-REx-293 cell line. A high-expressing clone was selected and adapted step-wisely to suspension culture for scaling-up purposes. This clone grew to a density of 4 million cells /ml in shake flask with viability higher than 95% and a doubling time of 48 hours.

To maximize the production of NTSR1, a preliminary orthogonal array design[75] was carried out investigating three induction parameters: tetracycline concentration (1-4 μg/ml), sodium butyrate (NaBu) concentration (0.5-10 mM) and induction time (24-60 hrs). Initial variance analysis showed negligible impact from higher tetracycline doses, significant effect of NaBu, and considerable cell death with induction time longer than 48hr (data not shown). Therefore, further optimization efforts were focused on the NaBu dose with induction time of 24 or 36 hours. As shown in Fig.2.3, NTSR1 expression was undetectable in the absence of tetracycline, while in the presence of 2 μg/ml tetracycline, 2.5 million plasma membrane localized receptors were produced. The expression of functional NTSR1 improved with increasing NaBu concentrations (0.5-10 mM) and optimal production was achieved by the addition of 2 μg/ml tetracycline and 10 mM sodium butyrate when the viable cell density reached 2

million cells/ml, with harvest 36 hrs later. These optimized conditions resulted in 8.8 million copies of NTSR1 localized at the plasma membrane per cell (Fig. 2.3), a 3.5-fold increase of cell surface expression compared to the induction with tetracycline alone.

Pilot-scale production of NTSR1 was then carried out in a 5L bioreactor. 2μg/ml tetracycline and 2.5 mM NaBu were added to T-REx-293 cells when viable cell density reached 1.5 million cells/ml. Cells were harvested 36 hrs after induction and plasma membrane localized receptors were determined to be 5.6 million per cell with [^{3}H]NT binding assays on intact cells. To determine the total amount of functional NTSR1 (i.e. receptors residing in the endoplasmic reticulum, the Golgi apparatus and plasma membrane), cells were solubilized with detergent and the number of receptors were determined by a detergent-based radio-ligand binding assay. This resulted in 18.2 million receptors per cell and a yield of 1.5 mg of functional NTSR1 per liter culture (Table 2.1).

2.4.2. NTSR1 expression in insect cells

Both Sf9 and *T. ni* cells were tested as insect hosts for NTSR1 expression and all subsequent pilot-scale expression experiments were performed using *T. ni* cells due to their higher viability after infection (we considered cell viability an essential factor as only healthy cells have an intact machinery for insertion and folding of membrane proteins). *T. ni* cells were infected at an MOI of 5 followed by the reduction of temperature to 21°C and cells were harvested at 48 hrs post-infection. NTSR1 was produced at a total number of 17.8 million receptors per cell or 1.7 mg receptor per liter culture as determined by ligand binding assays (Table 2.1).

2.4.3. Purification of NTSR1 from T-REx-293 cells and insect cells

The presence of neurotensin enhance the stability of NTSR1-GW5-Δi3 [67], therefore, all purification steps were conducted in the presence of the agonist peptide. The purification of NTSR1 produced in T-Rex-293 and insect cells was done in one step by immobilized metal affinity chromatography (Talon resin) in the presence of the detergent DM/CHS. The yield of NTSR1 per liter of cell culture, as calculated from the cell density at harvest and [^{3}H]NT binding assays on detergent-solubilized cells, was similar for both expression hosts (Table 2.1). The protein yield of the Talon column eluates was determined by the Amido Black method, as the presence of NT during the purification procedure prevented a radio-ligand binding analysis. Note that because of background contaminants (Fig.2.4), the Amido Black method overestimates the content of purified NTSR1.

2.4.4. Characterization of NTSR1 produced in insect cells and T-Rex-293 cells

To quantify the total amount of NTSR1 and the amount of plasma membrane localized NTSR1, [^{3}H]NT binding assays on detergent-solubilized cell extracts and on intact cells were conducted. The total number of functional NTSR1 per cell, produced in T-Rex-293 cells and insect cells, was similar (Fig. 2.5A). T-Rex-293 cells produced 12.8 million receptors per cell, while insect cells produced 15.8 million receptors per cell. However, the percentage of NTSR1 molecules that had trafficked to the cell surface was 2.5-fold higher in the case of T-Rex-293 cells compared to insect cells (Fig. 2.5B).

Gel electrophoresis of the solubilized protein from T-Rex-293 and insect cells

was performed and analyzed using INDIA HisProbe reagent, which detected all histidine tagged NTSR1. Equivalent amount of correctly folded NTSR1 (as determined by [^{3}H]NT binding assays) from the two hosts were applied and the comparable band intensity indicated comparable total NTSR1 expression (including correctly folded and misfolded ones), thus suggesting similar propensity for receptor misfolding in T-Rex-293 and insect cells (Fig.2.4, lane 4 and 5).

The ligand binding property of the receptors produced in the two hosts were comparable according to saturation binding assays using detergent-solubilized receptors. The dissociation constants for NTSR1 from both hosts were not statistically different (Fig. 2.6).

2.5 - Discussion

Unveiling the protein structure of GPCRs by crystallography will help to elucidate the mechanism of many diseases and enhance potential drug discovery and development. In order to accomplish this, milligram amounts of functional receptors are needed. Seeking an appropriate expression host is vital, as the host can affect the quantity and quality of the starting material [76]. Possible host choices include bacteria [77], yeast [78], baculovirus- insect cell system [79], and mammalian cells[80]. The baculovirus-insect cells system and inducible T-REx- 293 system were intensively investigated in recent years because of their lipid composition, translocation machinery and protein folding capabilities [11]. In this study, NTSR1 served as a model GPCR and the

baculovirus- insect cell and mammalian T-REx- 293 system were compared throughout NTSR1 expression and purification process in a quantitative way.

By using inducible expression strategy, 2.5 million functional copies of NTSR1 per cell were detected on plasma membrane of T-REx cells; this level of expression is 167 fold higher than the NTSR1 constitutively expressed in HEK-293T cells [38]. Optimization of induction parameters led to a further increase in expression level to 8.8 million functional receptors per cell on plasma membrane. Among the three induction parameters tested (tetracycline concentration, NaBu concentration and time length of induction) the concentration of sodium butyrate (NaBu) had the most impact. This compound has been successfully applied together with tetracycline for synergistic induction of many GPCRs expression [81, 82]. NaBu is routinely used at low concentration (1-5mM[83]), possibly because of its cytotoxic effects on cell growth[84]. In the experiments reported here, NaBu was applied at a concentration of 10mM resulting in improved receptor expression. The enhancement of NTSR1 gene transcription by inhibition of histone deacetylase may be a possible explanation. It is also possible that NaBu led to growth arrest of host cells, allowing available energy be channeled to the NTSR1 expression pathway.

Expression levels and binding properties of functional NTSR1 from the T-REx-293 and baculovirus insect cells systems were comparable. For both systems, the yields were approximately 15 million total functional receptors per cell or 1 milligram per liter culture after purification. This indicates that a 10L culture will provide approximately 10mg of high-quality material, a sufficient amount for structural studies. From production

point of view, a significant difference was observed in the surface-presentation of the receptors. In T-REx-293 cells, 2.5 times more NTSR1 were trafficked to the plasma membrane than insect cells as determined by ligand binding assays. The assumption was that higher surface-presentation is correlated with more properly folded receptors which were found to be the case for the serotonin transporter [76]. However, when equal amount of functional NTSR1 from both hosts were analyzed by western blot where INDIA HisProbe reagent would capture histidine tag of functional and non-functional NTSR1, the bands have similar intensity. This indicated comparable NTSR1 folding efficiency in both systems. Also, the fact that NTSR1 from insect cells could crystallize [67] suggested the difference between surface and internal receptors could be subtle. More studies will be needed in order to understand the implications of difference in receptor surface- presentation.

Another difference between the transient expressions in the insect cells and the stable expression in the T-REx-293 cells is the timeframe required for process development and for establishing the expression conditions (Fig. 2.7). The estimated time for expression with baculovirus-insect cells is around 6 weeks while it can take up to 12 weeks to obtain products from the stable T-REx-293 expression system. Based on this time frame the transient expression system is better at an early stage of the research when the GPCR of interest is subject to frequent sequence modifications for identifying constructs which are suitable for crystallization. Whereas stable expression in suspension T-REx-293 cells may be preferable for production of GPCRs of a specified construct due to its superior processing capabilities. The well-established scale-up methods for suspension cultures in a bioreactor will allow for one-step large scale production of

sufficient amounts of receptors for other applications of biophysical techniques such as NMR spectroscopy. Cost-wise, the medium prices for T-REx-293 and insect cells are comparable. It is likely possible that improvement in high density suspension culture of T-REx-293 cells will further reduce cost per milligram of protein.

In conclusion, we generated a suspension stable T-REx-293 cell line capable of expressing 1.5 milligram functional NTSR1 per liter. This cell line was found to be comparable to the transient baculovirus-insect cell expression system in regard to functional NTSR1 expression level and receptor binding properties. It provides a superior receptor surface display of the target proteins which may be advantageous for certain applications.

Contributions from collaborators:

Dr. Reinhard Grisshammer contributed to drafting and revising the manuscript and Jim White carried out ligand binding assays and NTSR1 purification from T-REx-293 cells and insect cells. Both collaborators contributed to conceiving and designing of experiments.

Figures and tables

Fig.2.1. Illustration of the tetracycline inducible expression system. (A) In the absence of tetracycline, TetR binds to tetO2 downstream of the TATA element, thus blocking transcription of the gene of interest. (B) The interaction of tetracyline with TetR causes its conformation to change, resulting in the dissociation of TetR from tetO2, thus initiating transcription.
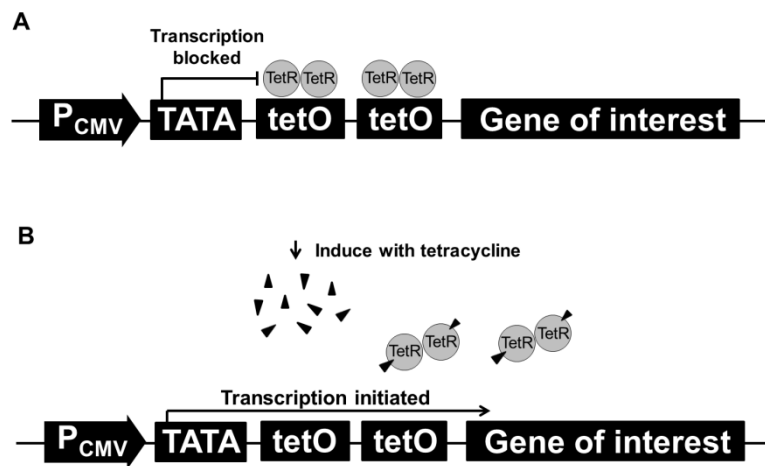
Fig.2.2. Construction of the NTSR1 expression vector. The NTSR1 DNA was inserted into pACMV-tetO using standard molecular biology techniques. The resulting expression vector allows tetracycline inducible expression of NTSR1 under control of CMV promoter.
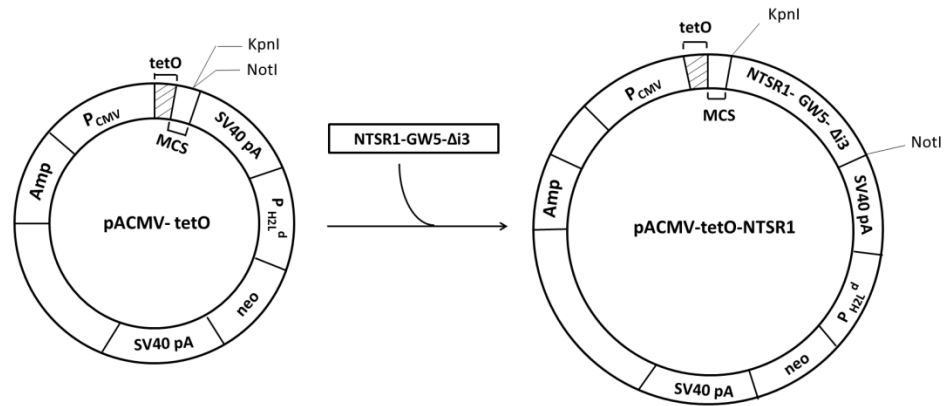
Fig.2.3. Optimization of NTSR1 expression under different induction conditions using a stable T-REx-293 cell line. The data were collected from a selected high-expressing clone. Cells were grown in suspension and were induced in the late exponential growth phase (at a viable cell density of 2 million cells /ml) with tetracycline. The addition of sodium butyrate enhanced expression levels. Intact cells were subjected to [^{3}H]NT binding assay to determine the number of receptors located at the cell-surface.
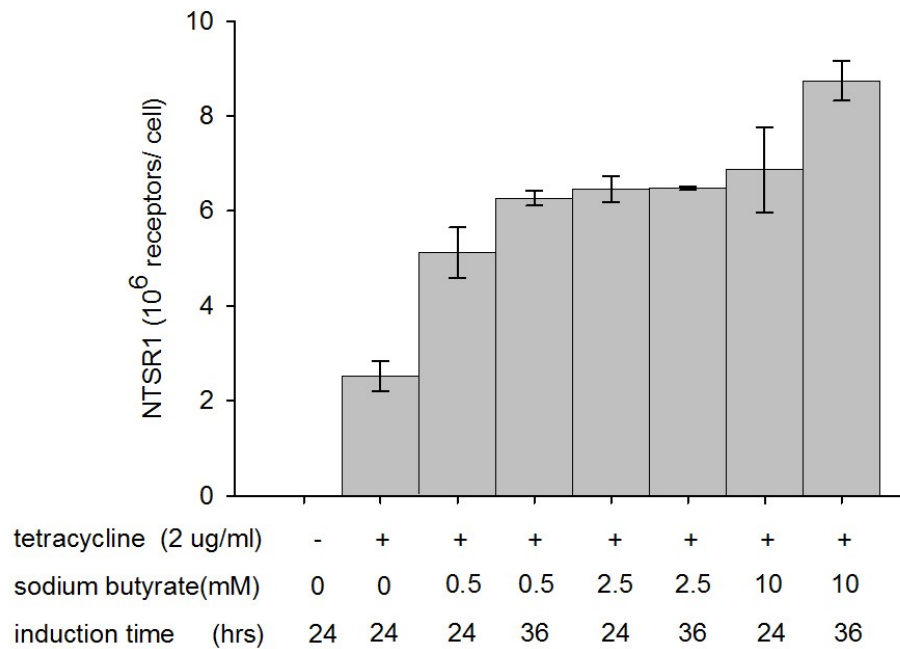
Fig.2.4. Purification of NTSR1. The progress of purification was monitored by SDS-

PAGE (NuPAGE 4–12% Bis-Tris gel, Invitrogen, 1x MES SDS buffer) and SimplyBlue

staining. Lane 1: Novagen Perfect Protein Marker (15–150 kDa); lane 2: Talon eluate of

NTSR1 produced in T-REx-293 cells (3.5μg); lane 3: Talon eluate of NTSR1 produced

in insect cells (6 μg); Western blot analysis of total cell extract was performed using the

HisProbe-HRP reagent recognizing the histidine tag. Lane 4: NTSR1 produced in T-REx-

293 cells (122,000 lyzed cells with 113 ng functional NTSR1); lane 5: NTSR1 produced

in insect cells (110,500 lyzed cells with 107 ng functional NTSR1).
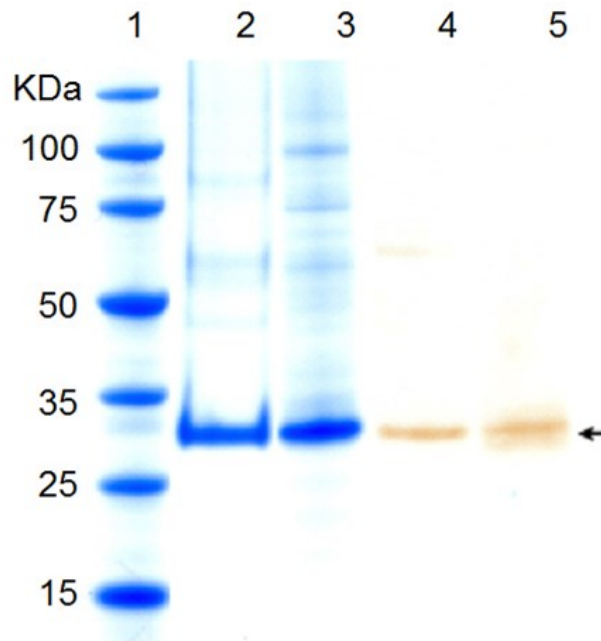
Fig.2.5. Expression of NTSR1 in the transient insect cell system and inducible T-REx-293 system. (A) Total functional NTSR1 numbers were determined by [^{3}H]NT binding assays using detergent solubilized cells (B) Surface localized NTSR1 numbers were determined by [^{3}H]NT binding assays using intact cells and combined with data from (A) to calculate percentage of surface localized NTSR1. (baculovirus insect cell system: 7 independent expression experiments; T-REx-293 system: 3 independent measurements on one 5L expression experiment). The expression of NTSR1 in insect cells was conducted as described in Materials and Methods. Expression of NTSR1 in the T-REx-293 system was induced by the addition of 2 µg/ml tetracycline and 10 mM sodium butyrate, with harvest and analysis 48 hours later.
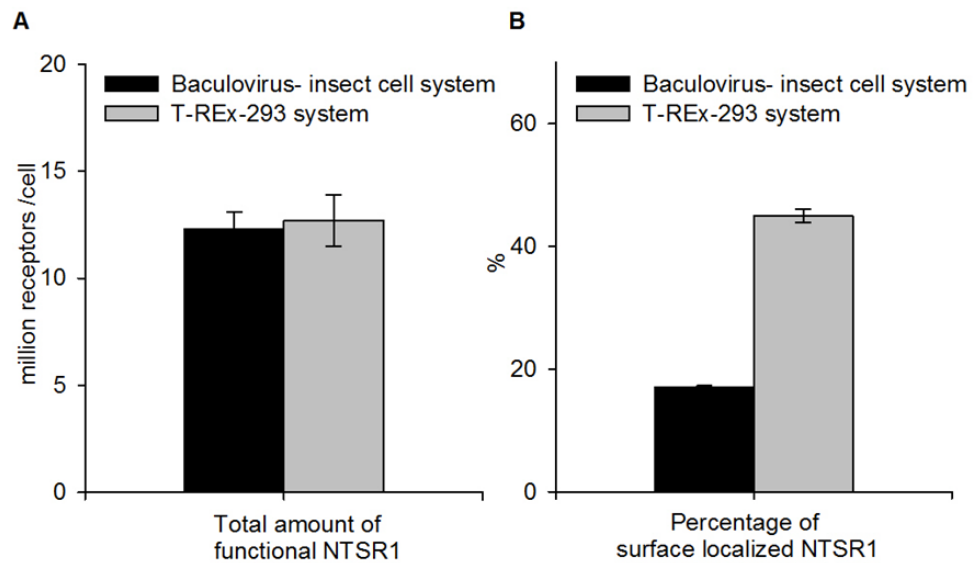
Fig.2.6. [³H]NT saturation binding of NTSR1 expressed in (A) T-REx-293 cells and (B) insect cells. NTSR1 was extracted from membranes with the detergent DM/CHS and subjected to radio-ligand binding analysis. Inset: Scatchard transformation. Representative experiments conducted as single data points are shown. (C) Table summarizing the values of the apparent dissociation constants values for [³H]NT binding. These values were not statistically different (P = 0.2, unpaired two-tailed t-test). Data were collected from three repeated experiments.
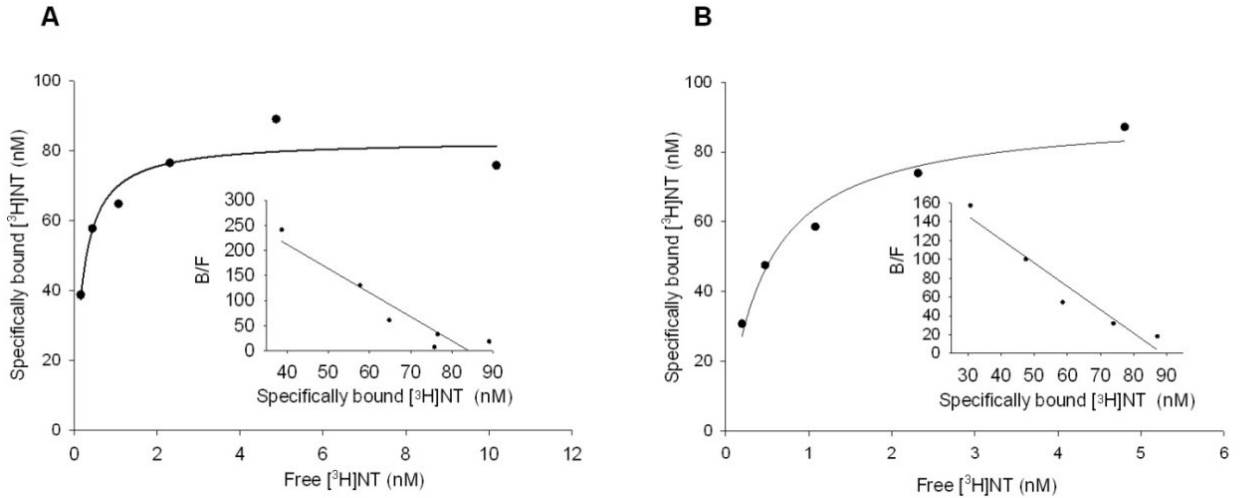
Receptor source	Apparent dissociation constant (nM) (n=3)
T-REx-293 cell	0.22 ± 0.06
Insect cell	0.57 ± 0.22

Fig.2.7. Timeframe for establishment of transient baculovirus- insect cells system and stable expression with inducible T-REx-293 system for GPCR expression.
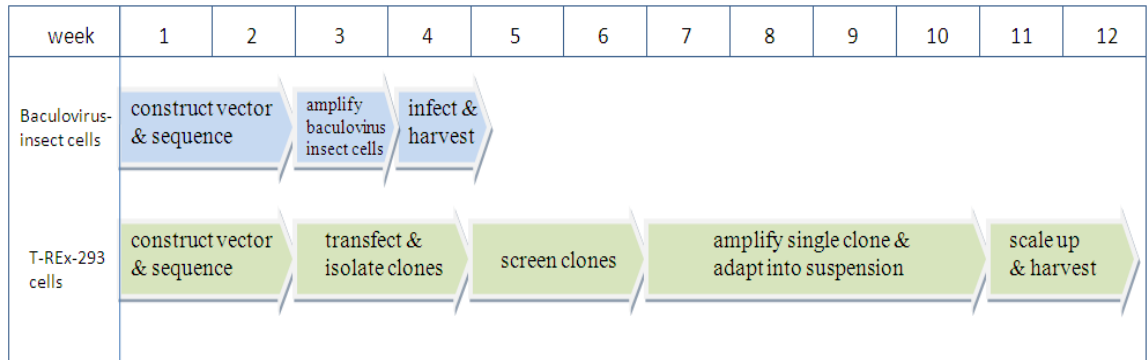
Table.2.1. Purification of NTSR1 from different host cells. Average data for insect cells were from four purification experiments using 1L (3 experiments) or 4L (1 experiment) of cell culture as starting material. Average data for T-REx-293 cells were from two purification experiments using 1L of culture as starting material. All purification procedures were performed in the presence of neurotensin. The theoretical yield was calculated from the cell density at harvest and [^{3}H]NT binding assays on detergent-solubilized cells. The protein yield of the Talon column eluate was determined by the Amido Black method, as the presence of NT during the purification procedure prevented a radio-ligand binding analysis. Because of minor contaminants in the Talon column eluate, the content of NTSR1 is overestimated. [a] average data from three 5L expressions, [b] data from one 5L bioreactor run.

Host cell	Functional receptor produced		Functional receptor purified (mg/L)
	(10^6/cell)	(mg/L)	
Insect cells	17.8 [a]	1.7 [a]	1.5
T-REx-293 cells	18.2 [b]	1.5 [b]	0.9

Chapter 3: Large-scale miRNA mimic screen for improved functional expression of neurotensin receptor

Abbreviations:

NTSR1, neurotensin receptor type 1; GPCR, G protein-coupled receptor; HEK, human embryonic kidney; CHO, Chinese hamster ovary; ORF, open reading frame; pri-miRNA, primary miRNA; pre-miRNA, precursor miRNA; RISC, RNA-induced silencing complex; 3'-UTR, three prime untranslated region; CMV, cytomegalusvirus; DMEM, Dulbecco's modified Eagle's medium; PBS, phosphate-buffered saline; FBS, fetal bovine serum; GFP, green fluorescent protein; MAD, median absolute deviation; DMSO, Dimethyl sulfoxide; NT, neurotensin; DDM, n-dodecyl-β-D-maltopyranoside; CHS, cholesteryl hemisuccinate Tris salt; CHAPS, 3-[(3-cholamidopyropyl) dimethylammonio] -1- propanesulfonate; GPC3, glypican-3; BSA, bovine serum albumin; MOF, mean of fluorescence.

3.1 - Summary

To explore the possibility of enhancing protein expression using miRNA, a stable T-REx-293 cell line was constructed to over-express the neurotensin receptor type 1 (NTSR1) fused with green fluorescent protein (GFP) to its C-terminal. The cell line was then subjected to high-content image-based human miRNA mimic library screening. Five microRNA mimics: hsa-miR-22-5p, hsa-miR-18a-5p, hsa-miR-22-3p, hsa-miR-429 and hsa-miR-2110 were identified to improve functional expression of NTSR1 by as much as 48%. In parallel, an HEK293 cell line expressing luciferase was also screened with the

same human miRNA mimic library. All five identified microRNA mimics were also found to enhance the luciferase expression up to 239%. Interestingly, all five miRNAs improved glypican-3 hFc fusion protein secretion up to 120%, which indicated that these molecules could have a wide role in enhancing production of proteins with biomedical interest.

3.2 - Introduction

As discussed in chapter 2, the HEK293 cells were demonstrated advantageous for NTSR1 expression in many aspects. They have high functional expression level and surface display ratio of the receptor. They can reach high cell density (4-8 million cells/mL) in bioreactors and have continuous production capability. In addition, the high cost associated with FBS has been brought down by serum-free suspension culture, making the using of the expression system cost-effective. In all, the HEK293 expression system is a good choice for large-scale production of proteins for medical or structural studies, especially for expression of mammalian integral membrane proteins such as receptors, ion channels and transporters [7].

In this chapter, the goal is to use NTSR1 as a model integral membrane protein and to engineer the preferred HEK293 expression system to improve its production capability of heterologous membrane protein. As discussed above and in chapter 1, rational engineering of the host cells may be trial and error due to the lack of thorough understanding in membrane protein biogenesis process. Thus, we chose a bottom-up approach, where host cells were treated by an entire library of engineering tools available to us (miRNAs in this chapter, siRNAs in chapter 4), in a high-throughput

format, for identification of the best sequences that significantly enhanced NTSR1 expression. These top sequences can be investigated in-depth in the future, to decipher the complex membrane protein expression process and provide reference for synthetic biology studies.

The powerful engineering tools used in this chapter are miRNAs. They have been used for engineering cells with desirable properties[85], such as improved protein production capabilities[86] and enhanced anti-apoptotic properties under stress conditions[44, 87]. MiRNAs are a novel class of small, non-coding RNAs that can simultaneously silence multiple genes by binding to their 3'-untranslated regions(3'-UTR)[88]. The miRNA development process starts in the nucleus. Following transcription, the several kb long primary miRNA (pri-miRNA) with a step-loop structure is cleaved by an endoculease III (Drosha) to generate a 70-nucleotide long precursor miRNA (pre-miRNA)[89]. After being transported to the cytoplasm, the stem-loop pre-miRNAs is cleaved by another endoculease III (Dicer), to generate the final mature miRNA duplex with a length of 18-25 nucleotides. The guide strand of the miRNA duplex is then loaded into the RNA-induced silencing complex (RISC) and guides the complex to the complementary 3'-UTR of targeting mRNAs. The degree of complementarity determines the mechanism of the post-transcriptional inhibition function. With high complementarity, target mRNA is degraded via RNA interference process, whereas insufficient complementarity is predictive of a translational repression mechanism.

MiRNAs exhibit a broad spectrum of regulatory effects in eukaryotic cellular processes including cell growth and apoptosis, cell differentiation and metabolism, cancer

development and progression[85]. Their capacity to globally regulate entire gene networks[90] and to introduce no additional translational burden (compared to gene overexpression strategies)[91] makes them particularly advantageous for cell line development.

In this chapter, we explore the ability to improve the receptor expression by applying the powerful miRNAs tool. The NTSR1-GFP-expressing HEK293 cells were subject to a high-throughput image-based screen with a human miRNA mimic library comprising 875 miRNA mimics. Excitingly, 5 top miRNA mimics identified from the screen not only improved NTSR1 functional expression level, but also significantly enhanced production of luciferase (intracellular protein) and an Fc fusion chimeric protein in HEK293 cell lines.

3.3 - Materials and Methods:

3.3.1 Construction of expression plasmid pJMA-NTSR1-GFP

Truncated wild type NTSR1 (T43-K396) was subcloned into the tetracycline inducible plasmid pJMA111(a kind gift from Dr. Christopher G. Tate) replacing the serotonin transporter construct using KpnI and NotI restriction sites. Thus NTSR1 was placed downstream of the tetracycline-controlled CMV promoter and had an eGFP-deca-histidine tag fused to its C-terminal (Fig. 3.1)[92].

3.3.2 Construction of stable NTSR1-GFP-expressing T-REx-293 cell line

The T-REx-293 cell line was maintained as an adherent culture in DMEM containing 10% certified FBS and 5 µg/mL blasticidin (Invitrogen). The cells were

transfected with the plasmid pJMA-NTSR1-GFP using Lipofectamine 2000 according to the manufacturer's protocol (Life Technologies). One day after transfection, cells were transferred into fresh DMEM medium containing 200 µg/mL zeocin (Invitrogen) and the medium was replaced every three days. Two weeks later, ten cell clones were separately expanded into two T-flasks each. Cells in one T-flask were harvested during the exponential growth phase and frozen in 10% DMSO for storage. Cells in the other T-flask were induced with 1 µg/mL tetracycline for 24 hrs, after reaching 80% confluency. Cells were then detached from the flask and washed with cold PBS. After adjusting the cell density to ~1×10^6 cells per mL, protease inhibitors (Roche) were added and the cell suspension was frozen on dry ice in 1mL aliquots. NTSR1 expression levels were determined by [^{3}H]NT binding and the clone with the highest expression level was selected for further experiments. The selected stable T-REx-293-NTSR1-GFP high expressor was then routinely maintained in DMEM containing 10% certified FBS, 5 µg/mL blasticidin and 200 µg/mL zeocin.

3.3.3 High-throughput miRNA screen

T-REx-293-NTSR1-GFP cells were screened with a miRNA mimic library (Qiagen) based on Sanger miRBase 13.0 and consisting of 875 miRNAs mimics. For transfection, 0.8 pmol of each mimic was spotted to 384 well plate wells (Corning) and 20 µL of serum-free DMEM containing 0.1 µL of Lipofectamine RNAiMax (Life Technologies) was then added to each well. This lipid-miRNA mixture was incubated at ambient temperature for 30 min prior to adding 2000 cells in 20 µL of DMEM containing 20% certified FBS (Gibco). Transfected cells were incubated at 37ºC in 5% CO_2 for 72

hours and induced with 1µg/mL tetracycline for 24 hours for NTSR1-GFP expression. Cells were then fixed with 2% paraformaldehyde (Electron Microscopy Sciences), stained with Hoechst 33342 (Life Technologies) for 45 minutes and gently washed with PBS. Plates were imaged with an ImageXpress Micro XL (Molecular Devices). Total cell number and per cell green fluorescence intensity were calculated using MetaXpress software (Molecular Devices) employing the Multi-Wavelength Cell Scoring application module. All screening plates had a full column (16 wells) of Silencer Select Negative Control #2 (Life Technologies) and the median value of each plate's negative control column was used to normalize corresponding sample wells. A full column of positive control siRNA targeting GFP (GFP-22 siRNA, Qiagen) was also used as on-plate reference for transfection efficiency. The median absolute deviation (MAD) - based z-score was calculated for each sample [93].

3.3.4 Validation transfection

Validation transfections were performed in 12-well plates with miScript miRNA mimics (Qiagen, Cat.No. 219600-S0), SilencerSelect Negative Control #2 and with lethal control siRNA (Qiagen AllStars Cell Death Control) served as a control for transfection efficiency. Cells were transfected as described for screening except 0.15 million cells were transfected with 40nM miRNA using 6.25ul Lipofectamine RNAiMax in a total volume of 1mL of media. 72 hours after transfection, cells were induced with 1µg/mL tetracycline. 24 hours later, cells from each well were detached with non-enzymatic cell dissociation buffer (Gibco, Cat. No. 13150-016) and washed twice with cold PBS. Cell densities and viability were determined by trypan blue exclusion using a CEDEX cell

quantification system (Roche, Mannheim, Germany). Based on the counts, cell densities were adjusted to 0.5 million cells/ml with PBS and then subject to flow cytometry analysis. The remaining cells were pelleted and frozen on dry ice for [^{3}H]NT binding assays.

3.3.5 Flow cytometry analysis

Cells harvested from validation transfection step were diluted to 0.2 million cells/ml with cold PBS for flow cytometry analysis. Green fluorescence was measured with Guava Easycyte 5HT and Incyte software (Millipore). The green fluorescence signal and cell gating were adjusted using uninduced T-REx-293-NTSR1-GFP cells, with more than 99.5% of the cells in low fluorescence range (<100). The setting was kept same for acquisition of all cell samples.

3.3.6 Analytical solubilization of NTSR1

The detergents n-Dodecyl-β-D-maltoside (DDM), 3-[(3-cholamidopyropyl) dimethylammonio] -1- propane sulfonate (CHAPS) and cholesteryl hemisuccinate Tris salt (CHS) were obtained from Anatrace. Cell pellets from 2 mL of suspension culture were suspended in Tris-glycerol-NaCl buffer. Then the detergents DDM, CHAPS and CHS were added to give a final buffer composition of 50 mM TrisHCl pH 7.4, 200 mM NaCl, 30%(v/v) glycerol, 1% (w/v) DDM, and 0.6%(w//v) CHAPS and 0.12%(w/v) CHS in a total volume of 0.5 mL. The samples were placed on a rotating mixer at 4°C for 1 hour. Cell debris and non-solubilized material were removed by ultracentrifugation (TL100 rotor, 60k rpm, 4°C, 30 min in Optima Max bench-top ultracentrifuge,

Beckman), and the supernatants containing detergent-solubilized NTSR1 were used to determine the total number of expressed receptors by a detergent-based radio-ligand binding assay.

3.3.7 Ligand binding assay

Tritiated neurotensin agonist [3H]NT ([3,11-tyrosyl-3,5-3H(N)]-pyroGlu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu) was purchased from Perkin Elmer. Ligand-binding assays with detergent-solubilized receptors were carried out in TEBB assay buffer (50mM Tris pH 7.4, 1mM EDTA, 40μg/mL bacitracin, 0.1% BSA) containing 0.1% (w/v) DDM, 0.2% (w/v) CHAPS and 0.04% (w/v) CHS. For one-point assays, receptors were incubated with 2 nM [3H]NT on ice for 1 hour in a volume of 150 μL. The concentration of [3H]NT used was at least 5-fold above the apparent dissociation constants for detergent-solubilized NTSR1 to allow high receptor occupancy. Separation of the receptor-ligand complex from free ligand (100 μL) was achieved by centrifugation-assisted gel filtration using Bio-Spin 30 Tris columns (BioRad), equilibrated with RDB buffer (50 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1% (w/v) DDM, 0.2% (w/v) CHAPS, 0.04% (w/v) CHS). Non-specific [3H]NT binding of 220 dpm was subtracted from total binding to calculate the total amount of receptors in T-REx-293-NTSR1-GFP cells. The number of functional NTSR1 was estimated by specific [3H]NT binding assuming one ligand-binding site per receptor molecule. The number of cells in the assay was derived by cell counting at cell harvest. This approach led to the calculation of the parameter "receptors/cell".

3.3.8 Application with luciferase-expressing cells

HEK-CMV-Luc2-Hygro cell line constitutively expressing luciferase is purchased from Promega. Transfections were performed in 12-well plates with miScript miRNA mimics (Qiagen, Cat.No. 219600-S0), SilencerSelect Negative Control #2 and lethal control siRNA (Qiagen AllStars Cell Death Control) served as a control for transfection efficiency. Cells were transfected in duplicates as described for screening except 0.1 million cells were transfected with 40nM miRNA using 6.25ul Lipofectamine RNAiMax in a total volume of 1mL of media. 72 hours after transfection, 500μL of ONE-Glo™ Reagent (Promega) was added to one set of replicates for luciferase activity quantification and 500μL of CellTiter-Glo™ Reagent (Promega) was added to the second set of replicates for viable cell density measurement. All plates were incubated at room temperature for 20 minutes to stabilize luminescent signal and then measured with SpectraMax i3 plate reader (Molecular Devices). Per cell luciferase production can be calculated from overall luciferase activity and viable cell number.

3.3.9 Application with GPC3-hFc-expressing cells

HEK- GPC3-hFc cell line constitutively secreting glypican-3 hFc-fusion protein (GPC3-hFc) is a kind gift from Dr. Mitchell Ho's group from National Institutes of Health. Cells were grown in DMEM supplemented with 10% FBS in a humidified incubator set at 37ºC and 5% CO_2. Cells were transfected in 12-well plates as described for screening except 0.15 million cells were transfected with 40nM miRNA using 6.25ul Lipofectamine RNAiMax in a total volume of 1mL of media. 7 days after transfection, cell culture supernatant was collected and cleared using centrifuge for GPC3-hFc concentration determination with ELISA and cells were detached and counted by trypan

blue exclusion using a CEDEX cell quantification system (Roche, Mannheim, Germany). Per cell GPC3-hFc production can be calculated from overall GPC3-hFc yield and viable cell number.

3.3.10 ELISA for GPC3-hFc concentration determination

AffiniPure F(ab')$_2$ Fragment Goat Anti-Human IgG (min X Bov, Ms, Rb Sr Prot, Cat. 109-006-170, Jackson Immunology) was used to coat a 96-well plate at 5μg/mL in PBS buffer, 50 μL per well, at 4 °C overnight. After the plate was blocked with 2% BSA in PBS buffer, pre-diluted cell culture supernatant was added, and the plate was incubated at room temperature for 1 h to allow binding to occur. After the plate was washed twice with PBS buffer containing 0.05% Tween 20, Peroxidase-conjugated AffiniPure Goat-anti-uman IgG (Cat. 109-035-098, Jackson Immunology) was added at 1:4000 dilution, 50ul/well. Following incubating at room temperature for 1 h, the plate was washed 4 times and detected with Peroxidase Substrate System (KPL).

3.4 - Results

3.4.1 Construction of inducible T-REx-293-NTSR1-GFP cell line for image-based screen

A stable cell line expressing wild type functional NTSR1-GFP fusion was constructed using the inducible T-REx system[68] by transfecting T-REx-293 cells with the pJMA-NTSR1-GFP plasmid (Fig.3.1). Ten clones were isolated and their neurotensin

receptor expression level upon tetracycline induction was measured by [^{3}H]NT binding assay (data not shown). A high-expressing clone producing 8.4 million receptors per cell was selected for further experiments. As seen in Fig.3.2, the receptors for this clone are located mostly on the plasma membrane as expected.

3.4.2 High-throughput miRNA screen for enhanced NTSR1-GFP expression

To identify miRNAs that improve NTSR1 expression in T-REx-293-NTSR1-GFP cells, the cells were screened with a library comprised of 875 human miRNA mimics. Cells were transiently transfected with mimics in 384-well format for 72 hours followed by tetracycline-induced expression of NTSR1-GFP fusion protein (Fig.3.3A). Twenty four hours after induction, the cells were fixed followed by nuclear staining. Each well was then imaged to obtain total cell number and per cell mean green fluorescent intensity (Fig.3.3B). Sample values were normalized based on the median value of each plate's negative control column. A column of positive control siRNA capable of silencing *gfp* gene was also used as on-plate control for transfection efficiency. GFP-directed siRNA consistently provided a > 80% decrease in green fluorescence intensity. To assess reproducibility, the screen was performed in duplicate, resulting in a correlation coefficient of 0.92 (Fig.3.3C). Furthermore, the screen was completed again in replicate using cells from a different passage. The correlation between the two independent screen was 0.73. The median absolute deviation (MAD) - based z-score[93] was calculated for each sample, and the distribution of miRNA activity is plotted in Fig.3.3D. 40 miRNAs were shown to significantly increase NTSR1-GFP productivity (MAD-based z-score> 2.0. Table 1) in both biological replicates and 26 of them (two thirds of total 40) were selected for follow up analysis. All screen data for the four replicates can be found in

Table S1.

3.4.3 Validation of the selected miRNA candidates by flow cytometry analysis

The expression level of NTRS1-GFP following transient transfection of the cells with the top 26 microRNA was measured by flow cytometry (Fig.3.4). The uninduced cells exhibited basal GFP expression with only 1% of cells exceeding the background fluorescence (10^1) (Fig.3.4A). Following transfection with negative control siRNA (siN.C.) and tetracycline induction, the expression of NTSR1-GFP caused a significant shift in the fluorescence intensity, resulting in a geometric mean of fluorescence (MOF) of 138. A further shift was observed when the cells were transfected with various miRNA mimics followed by tetracycline induction, including miR-129-5p, which led to a MOF of 197. Compared with negative control siRNA, 14 of the 26 miRNAs resulted in an increased MOF. From this group, top 9 miRNAs were selected for further investigation (Fig.3.4B). Following the transfection with the 26 selected miRNAs, a large variance was seen in viable cell density (ranged from 54% to 135%, normalized to negative control) but not in viability (ranged from 84% to 97%) (Fig.3.4C).

3.4.4 [^{3}H]NT binding assay validation for improved functional expression of NTSR1

The effect of the top 9 miRNAs on the functional expression of NTSR1 was also evaluated by measuring the functional activity of the receptor through the binding of labeled neurotensin ([^{3}H]NT). Although all top 9 miRNAs were shown to improve NTSR1-GFP expression based on GFP fluorescence, only 5 of them (miR-22-5p, miR-18a-5p, miR-22-3p, miR-429 and miR-2110) led to improved functional activity levels of NTSR1(Fig.3.5A). Of these, miR-2110-transfected cells expressed 13.8 million functional neurotensin receptors per cell, which was 48% higher than that from siN.C. In

addition, miR-22-5p and miR-22-3p improved functional expression of NTSR1, by 30% and 21% respectively. As seen in Fig.3.5B a number of the top 9 miRNAs had negative effect on cell growth and viability.

3.4.5 MiRNA screen for enhanced luciferase expression

The human mimic miRNA library was also evaluated for its effects on the expression of luciferase in HEK293 cells constitutively expressing luciferase under control of a cytomegalovirus (CMV) promoter. Transfection was performed in duplicate in 384-well format. Seventy two hours post- transfection, one set of plates was assayed for luciferase and the other set was used for viable cell density (Fig.3.6A). Both luciferase activity and viable cell density were normalized to the median value of each plate's negative control column and the luciferase expression per cell was calculated for each miRNA. Though luciferase and NTSR1 screen exhibited a limited correlation (R= 0.31, Fig.3.6B), seven out of nine top hits identified from NTSR1 screen (Fig.3.6C) also significantly improved per cell luciferase productivity on a per cell basis (MAD-based z-score>2.0).

3.4.6 Validation of common hits

The top 9 miRNAs identified from the NTRS1 screen were examined for their effects on luciferase activity in a 12-well plate format. Seven miRNAs improved luciferase activity from 50% to 239% (Fig.3.7A). MiR-892b and miR-22-3p showed the largest effect on luciferase expression with a 239% and 207% improvement respectively. Although these microRNAs inhibited cell growth (Fig.3.7B), the overall production of luciferase from cells transfected with miR-892b and miR-22-3p was still 188% and 127% higher, respectively, than the negative control siN.C. level (Fig.3.7C). Interestingly, both

miR-22-3p and miR-22-5p showed up as top common hits for NTSR1 and luciferase screen.

3.4.7 Application of top common hits on secreted protein

To investigate the impact of top common hits on secreted protein production, the five identified miRNAs(hsa-miR-22-5p, hsa-miR-18a-5p, hsa-miR-22-3p, hsa-miR-429 and hsa-miR-2110) were independently transfected into HEK293 cell line stably expressing secreted hFc-fusion protein: glypican-3 hFc-fusion protein (GPC3-hFc) [94]. All five miRNAs enhanced per cell GPC3-hFc secretion (up to 120% improvement, Fig. 7A), while three miRNAs (hsa-miR-22-5p, hsa-miR-18a-5p and hsa-miR-22-3p) effectively enhanced overall GPC3-hFc (up to 62%, Fig. 3.8C).

3.5 - Discussion:

Integral membrane proteins such as mammalian receptors, ion channels and transporters are vital for medical research. However, obtaining large amounts of functional membrane proteins for medical research, especially structural studies, has been difficult and therefore been a barrier for productive research towards better understanding of their mechanisms and potential medical use [7, 8, 11]. So far, a tetracycline-inducible mammalian expression system [68] has been shown to be an effective method for functional expression of membrane proteins [50, 51, 76]. This inducible system together with optimized production conditions led to a yield of 1 milligram per liter of purified NTSR1 in chapter 2. Compared with well-developed prokaryotic hosts such as *E.coli*, the production of membrane proteins from higher eukaryotic hosts is still in the stage of "trial and error"[95] since engineering tools are limited and the membrane protein synthesis,

insertion, folding and trafficking are not completely understood.

To improve the production of these proteins, a bottom-up screening approach using human miRNA mimics library was implemented to identify candidates that lead to improved expression of the GPCR from the T-Rex-293 cells. This approach has previously proven effective for apoptosis screening[87] and recombinant secreted protein screening[96, 97] in CHO cells. In this study, we developed an image-based high-throughput screening method to detect per cell green fluorescence signal, which is applied as a proxy for the number of molecules of NTSR1 protein expressed per cell. In addition to its high reproducibility (0.92 correlation between technical replicates), this method is cost-effective for protein labeled with fluorescent, as no secondary reagent is needed for protein quantification. . It is also high throughput and high-capacity, as cells are fixed and the screening is not time-sensitive compared to live-cell processes such as flow cytometry. This screen methodology can be applied to other membrane or intracellular protein candidates when the targeted protein is fused with GFP. Although GFP fusion has been widely used for membrane protein overexpression screen and purification in a variety of hosts [98-100], it is possible that the N-terminal GFP fusion may mask signal sequence essential for protein insertion. This may compromise folding or correct localization of the desired membrane protein [101, 102]. C-terminal GFP fusion, on the other hand, is preferable as it is generally better in maintaining the localization and function of the native protein [101] with exceptions when C-terminal contains an essential functional segment [103, 104].

Among the 875 human miRNA mimics tested, 40 mimics consistently led to significant improvement in per cell green fluorescence levels, exhibiting an average

MAD-based-z-score higher than 2.0. Among the top 40 candidates, miR-892b, miR193b-3p and miR-193a-3p share the same seed sequence (ACUGGC), indicating that they may comprise an overlap in target genes. Similarly, miR-129-3p and miR-129-1-3p also share a same seed sequence (AGCCCU).

The activity of two thirds of the 40 mimics was confirmed further by flow cytometry and the 9 mimic candidates contributing to the highest per cell green fluorescence signal were further tested in the [^{3}H]NT binding assay. Five out of the nine mimics showed up to 48% improvement in functional expression of NTSR1. From these five, miR-2110 is a novel miRNA that has been identified but not studied[105]. The other four miRNAs (miR-429, miR-18a-5p, miR-22-5p and miR-22-3p) have been associated with cancer research in which they have exhibited contradicting effects on cell proliferation, cell growth, and protein production depending on the cell type and stage of cell development[86, 106-111]. For example, miR-429, a member of the miR-200 family, was shown to suppress tumor growth in human osteosarcoma[107], while in non-small cell lung cancer (NSCLC), the same miRNA is suggested as a potential target for NSCLC due to its promotion of cell proliferation[106]. miR-18a-5p is part of the miR-17-92 precursor sequence cluster, which is also named Oncomir-1. This miR-17-92 cluster was studied in depth regarding its effect on recombinant EpoFc protein production in CHO cells. Although over-expression of the entire cluster decreased productivity while having no effect on cell growth, the over-expression of miR-17 and miR-92 were shown to increase production[86].

Of the nine miRNAs that enhanced the expression of the NTSRI-GFP fusion protein, four (miR-129-5p, miR-221-5p, miR-892b and miR-639) were not associated

with enhanced binding activity of the agonist in the [^{3}H]NT assay. This may be an indication that NTSR1 could be misfolded in these cells following the enhanced expression. Another observation is that eight of the nine top hits (except miR-129-5p) caused a decrease in the viable cell number. One possible reason for this behavior is that overexpression of NTSR1-GFP could be toxic to the cells. Another possibility is that the introduction of a specific miRNA to the cells is associated with a growth arrest, leading to improved protein production[85, 112]. Since multiple pathways and genes can be targeted by one miRNA, it will be worthwhile to examine which specific genes are down-regulated in these cells and to investigate the mechanism that improved NTSR1 functional expression.

In parallel to the analysis of the miRNA effect on the NTSR1 expression, an HEK293 cell line constitutively expressing luciferase under the control of CMV promoter was subjected to screening of the same miRNA mimics library. This screen showed low degrees of correlation (R= 0.31) with the NTSR1-GFP screen. The low correlation may be the result of the difference between biogenesis process of integral membrane proteins and intracellular soluble proteins; the difference between constitutive expression elements and the inducible expression system; and clonal differences between the two HEK293 cell line used. Despite the overall low correlation between the screens, seven out of nine top miRNAs (except miR-129-5p and miR-639) identified from NTSR1-GFP screen, improved luciferase activity from 50% to 239%. All the final five miRNAs (miR-429, miR-18a-5p, miR-22-5p and miR-22-3p and miR-2110) capable of improving NTSR1 functional expression were also relevant for improving luciferase expression.

These five miRNAs affecting both model proteins were expected to have wider

application for other types of proteins. Therefore, they were tested with HEK293 cell line constitutively secreting an Fc fusion proteins with medical importance [94]. Interestingly, all of the five miRNAs were effective in enhancing per cell Fc fusion protein secretion. However, the overall Fc fusion protein yield varied from 10% decrease to 62% increase, partially depending on the viable cell number after miRNA transfection.

Contributions from collaborators:

Dr. Scott Martin contributed to conceiving and designing of RNAi screen experiments, data analysis and revising the manuscript. Yu-Chi Chen carried out luciferase miRNA mimic screening.

Figures and tables

Fig.3.1. Plasmid map for pJMA-NTSR1-GFP.
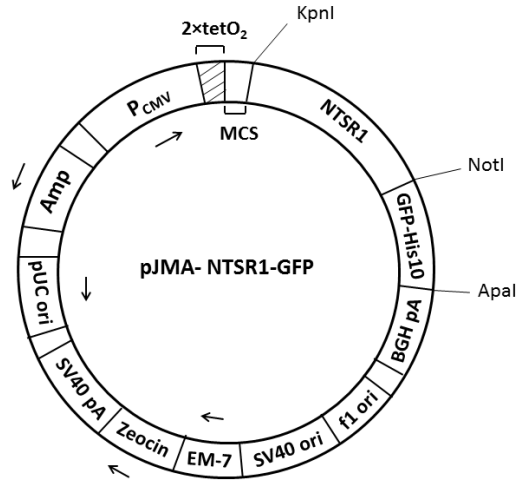
Fig.3.2. Confocal microscopy of tetracycline-induced T-REx-293-NTSR1-GFP cells with

NTSR1-GFP fusion protein located on plasma membrane.
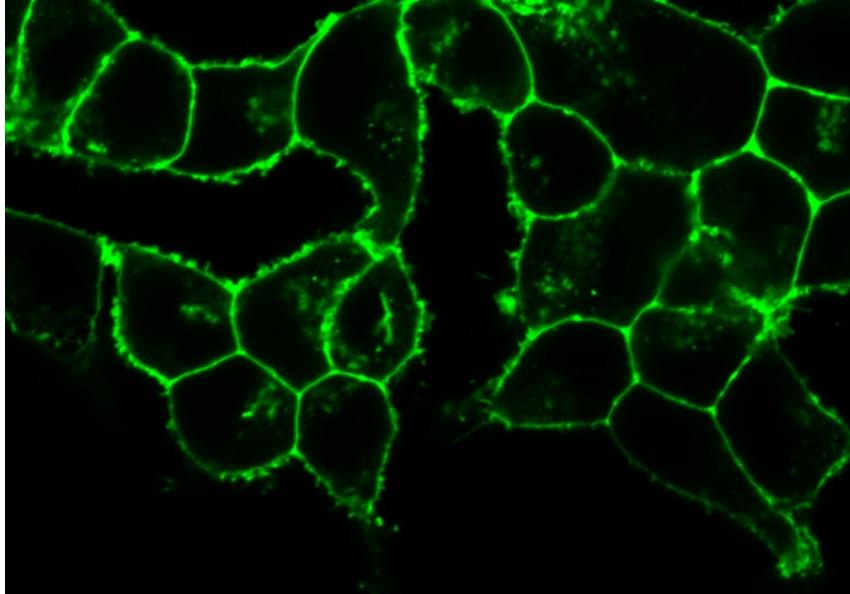
Fig.3.3. MiRNA screen with stable T-REx-293-NTSR1-GFP cell line. (A) Workflow of the screen. (B) Nuclei staining and GFP expression was captured by ImageXpress (C) Correlation plot of replicates from the miRNA library screen. (D) Distribution of miRNA mimics activity on improved NTSR1 expression; top hits (MAD based z-score>2.0) are highlighted.
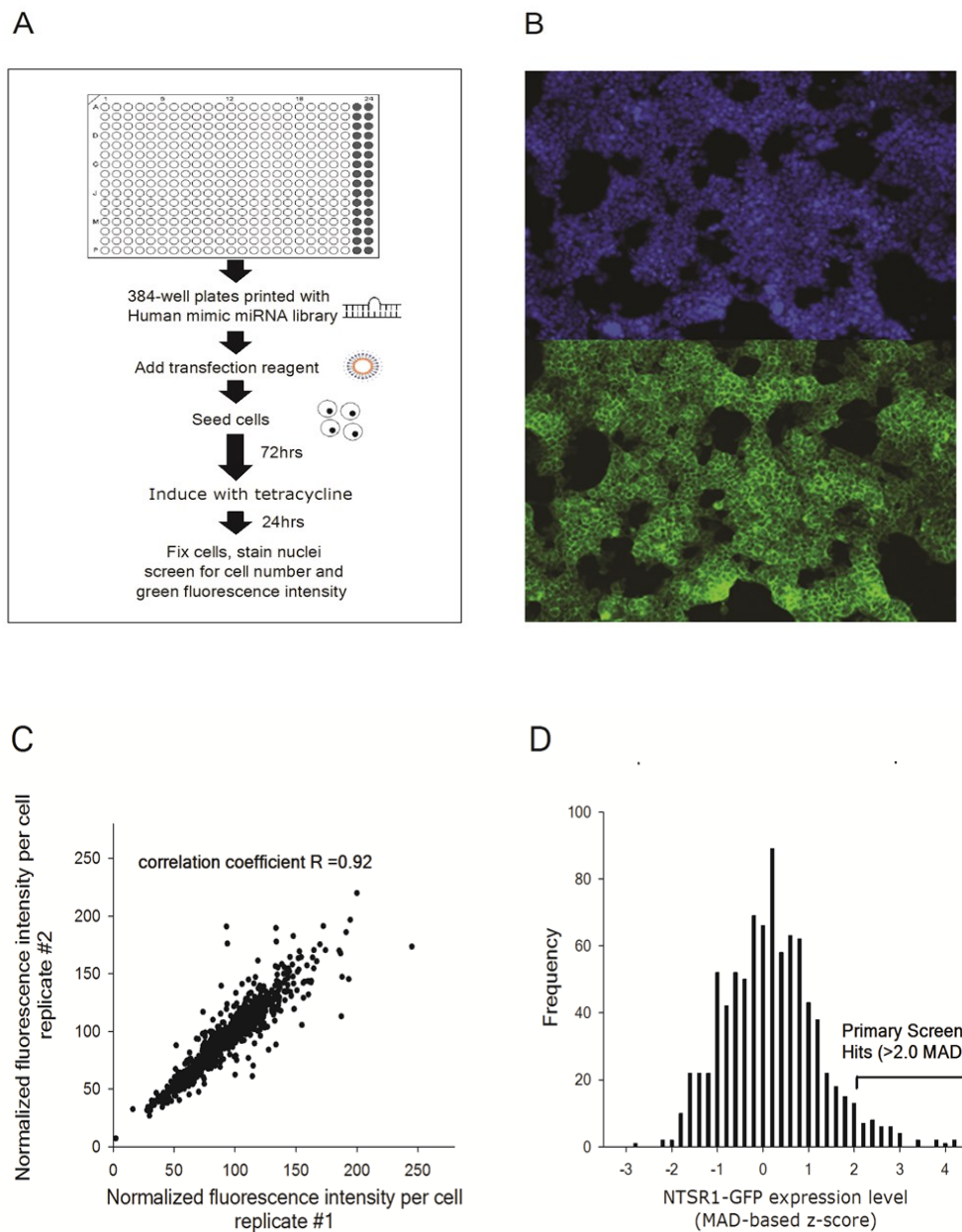
Fig.3.4. Flow cytometry analysis on T-REx-293-NTSR1-GFP cells transfected with 26 miRNAs selected from those MAD-based z-score>2.0. (A) Fluorescence histogram of un-induced cells (grey), induced cells transfected with negative control siRNA siN.C. (dash line) and induced cells transfected with miR-129-5p (solid line). (B) MOF from each sample was normalized to the negative control (siN.C). Top 9 miRNAs are indicated. (C) Normalized viable cell density and viability of cells transfected with 26 miRNA hits. Three biological samples were collected for each transfection experiment. Error bars represent SEM (standard error of the mean).
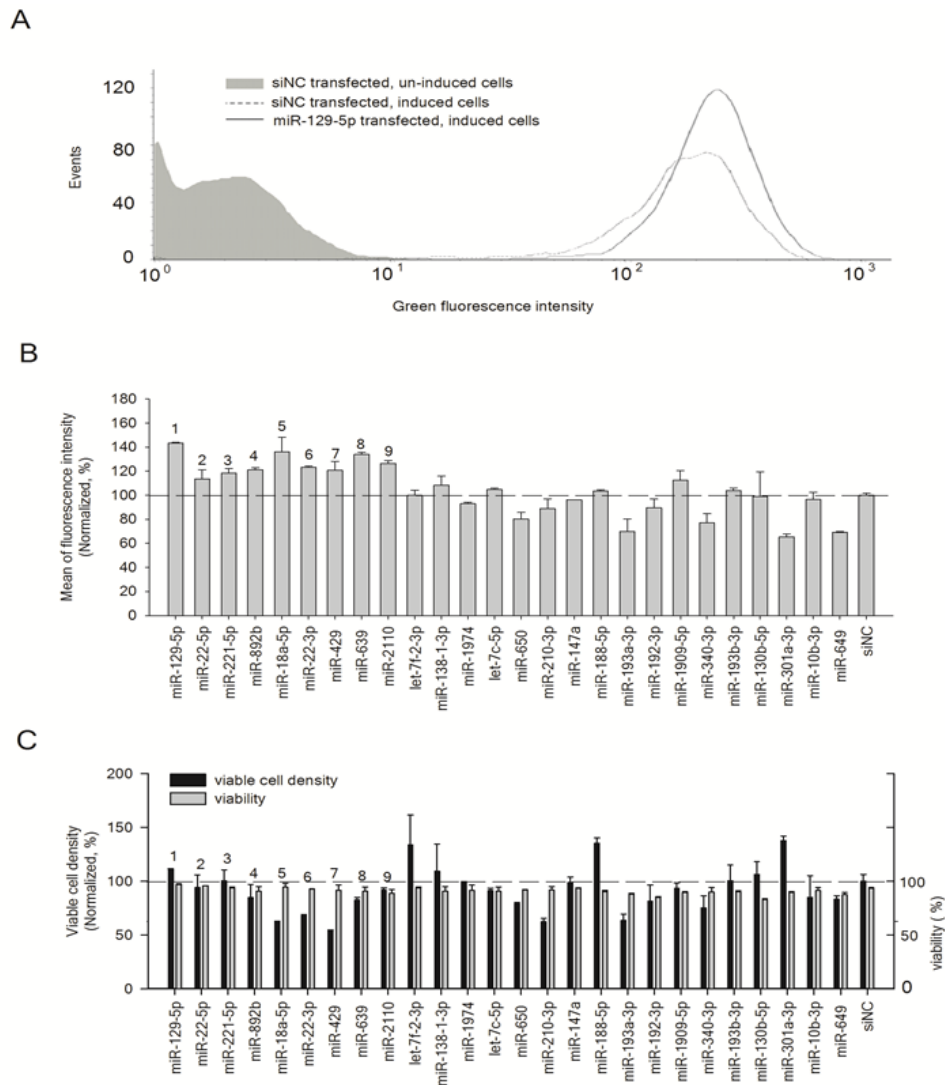
Fig.3.5. Validation of improved functional expression of NTSR1 with [³H]NT binding assay. (A) Functional NTSR1 numbers were determined by [³H]NT binding assays using detergent solubilized cells. (B) Cells were counted at harvest and normalized to the control (siN.C.). Two independent experiments were carried out with different passages of T-REx-293-NTSR1-GFP cells, and each independent experiment was tested in duplicate. Error bars indicate SEM.
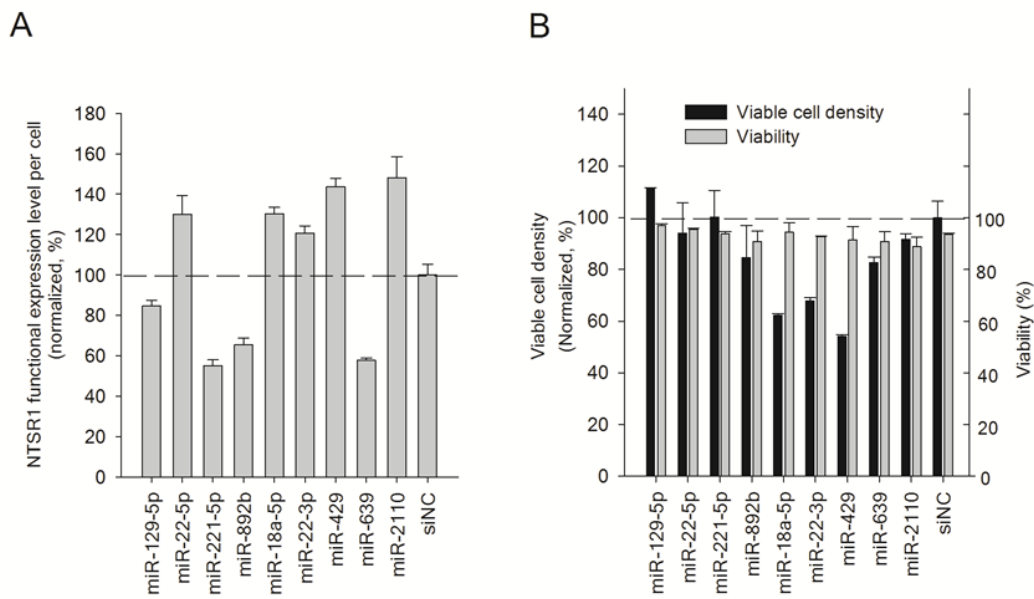
Fig.3.6. MiRNA screen with stable HEK-CMV-Luc2-Hygro cell line. (A) Workflow of the screen. (B) Correlation plot of screen result from luciferase screen and NTRS1-GFP screen. (C) Top common hits from miRNA library screen with NTSR1 and luciferase as target protein.
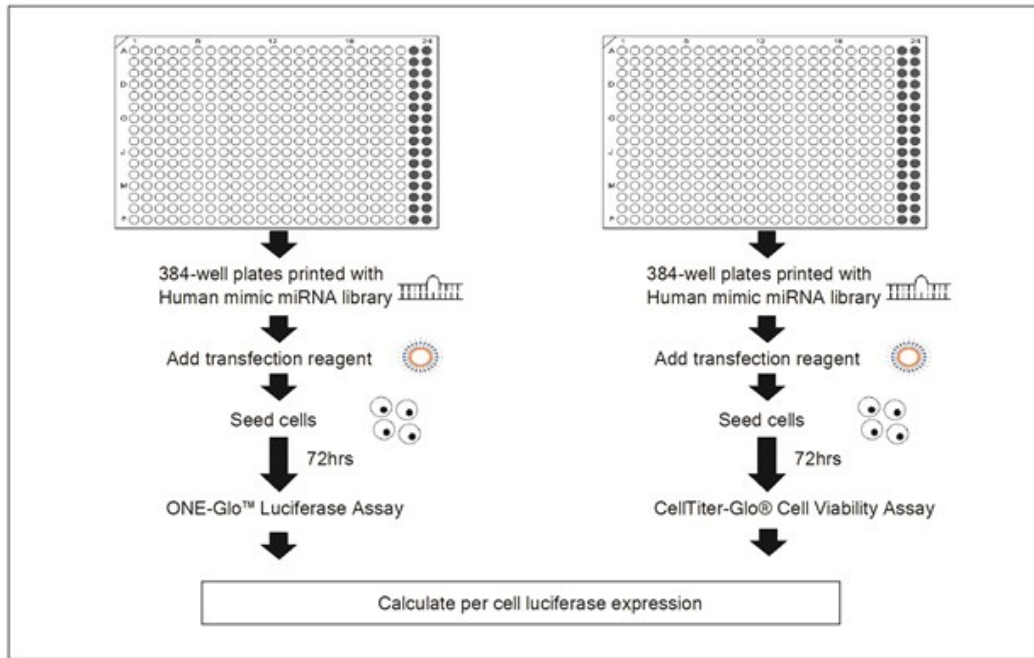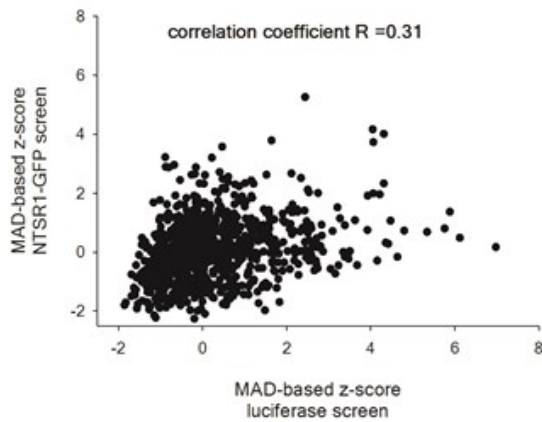
C

miRNAs (hsa-)	Mature miRNA sequence
miR-22-5p	AGUUCUUCAGUGGCAAGCUUUA
miR-221-5p	ACCUGGCAUACAAUGUAGAUUU
miR-892b	CACUGGCUCCUUUCUGGGUAGA
miR-18a-5p	UAAGGUGCAUCUAGUGCAGAUAG
miR-22-3p	AAGCUGCCAGUUGAAGAACUGU
miR-429	UAAUACUGUCUGGUAAAACCGU
miR-2110	UUGGGGAAACGGCCGCUGAGUG

Fig.3.7. Validation of improved luciferase activity. HEK-CMV-Luc2-Hygro cells were transfected in 12-well plates with the top 9 miRNAs in duplicate. 72 hours post transfection, one replicate was used for luciferase measurement and the other one was subject to cell counting. (A) Per cell luciferase activity was determined by ONE-Glo luciferase assay and viable cell density. (B) Viable cell density and (C) Overall luciferase production were normalized to the negative control (siN.C.). For each biological sample, the measurement was done in duplicates. The experiment was performed twice with different passages of cells. Error bars indicate SEM.
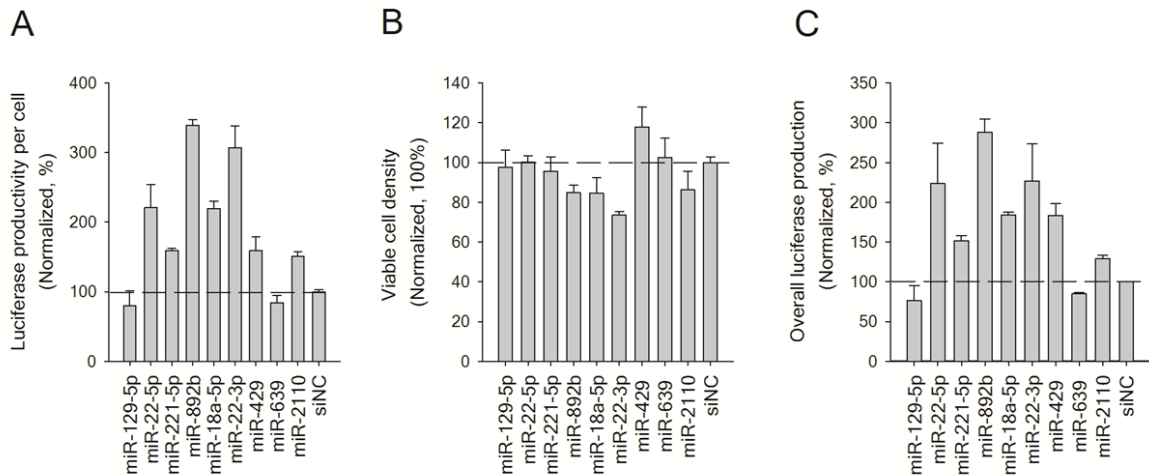
Fig.3.8. Improved glypican-3(GPC3) hFc-fusion protein secretion by the five top miRNAs. (A) Per cell GPC3-hFc secretion was determined by ELISA and viable cell density. (B) Viable cell density and (C) Overall GPC3-hFc production were normalized to the negative control (siN.C.).The experiment was performed twice with different passages of cells. For each biological sample, the measurement was done in triplicates. Error bars indicate SEM.
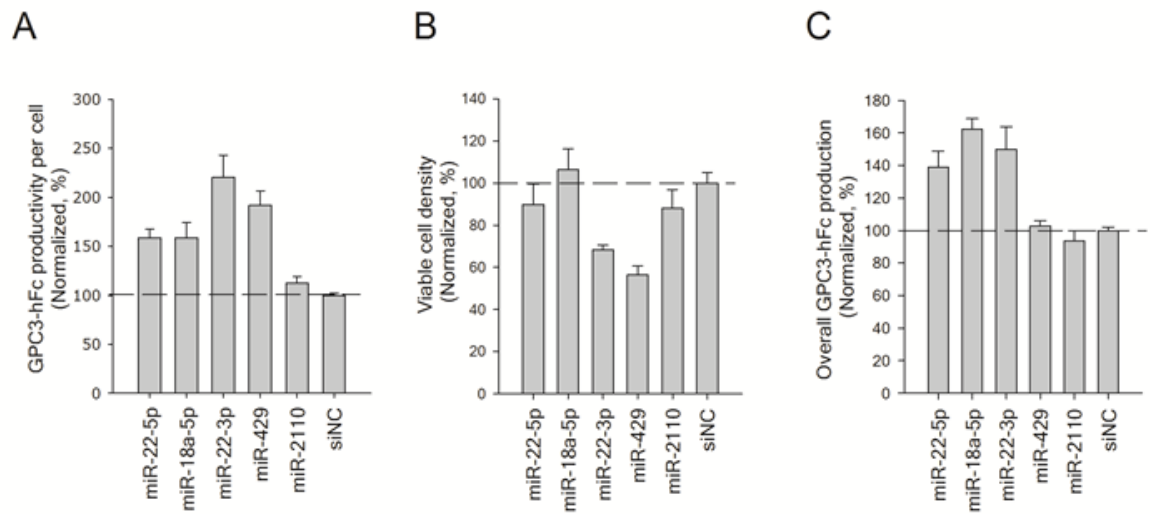
Table.3.1. Top hits from human miRNA mimics screen based on per cell green fluorescence intensity (MAD-based z-score>2.0)

Human miR ID (hsa-)	Variant	Mature miRNA sequence	MAD-based z-score	Signal relative to negative control (%)
miR-221	5p	ACCUGGCAUACAAUGUAGAUUU	5.3	248
miR-429	-	UAAUACUGUCUGGUAAAACCGU	4.2	212
miR-22	5p	AGUUCUUCAGUGGCAAGCUUUA	4.0	215
miR-892b	-	CACUGGCUCCUUUCUGGGUAGA	3.7	201
miR-1974	-	UGGUUGUAGUCCGUGCGAGAAUA	3.6	201
miR-210	3p	CUGUGCGUGUGACAGCGGCUGA	3.2	183
let-7f-2	3p	CUAUACAGUCUACUGUCUUUCC	3.0	178
miR-130b	5p	ACUCUUUCCCUGUUGCACUAC	2.9	178
miR-188	5p	CAUCCCUUGCAUGGUGGAGGG	2.9	177
miR-301a	3p	CAGUGCAAUAGUAUUGUCAAAGC	2.9	176
miR-129	5p	CUUUUUGCGGUCUGGGCUUGC	2.7	172
miR-147a	-	GUGUGUGGAAAUGCUUCUGC	2.6	168
let-7c	5p	UGAGGUAGUAGGUUGUAUGGUU	2.6	168
miR-1909	5p	UGAGUGCCGGUGCCUGCCCUG	2.6	169
miR-138-1	3p	GCUACUUCACAACACCAGGGCC	2.5	167
miR-193b	3p	AACUGGCCCUCAAAGUCCCGCU	2.5	166
miR-650	-	AGGAGGCAGCGCUCUCAGGAC	2.5	163
miR-639	-	AUCGCUGCGGUUGCGAGCGCUGU	2.4	165

miR-10b	3p	ACAGAUUCGAUUCUAGGGGAAU	2.4	162
miR-2110	-	UUGGGGAAACGGCCGCUGAGUG	2.3	160
miR-22	3p	AAGCUGCCAGUUGAAGAACUGU	2.3	158
miR-193a	3p	AACUGGCCUACAAAGUCCCAGU	2.3	156
miR-340	3p	UCCGUCUCAGUUACUUUAUAGC	2.3	159
miR-649	-	AAACCUGUGUUGUUCAAGAGUC	2.0	150
miR-18a	5p	UAAGGUGCAUCUAGUGCAGAUAG	2.0	149
miR-192	3p	CUGCCAAUUCCAUAGGUCACAG	2.0	148

All mature miRNAs variants are shown: 5'-end pre-miRNA derived (marked 5p) and 3'-end pre-miRNA derived (marked 3p)

Chapter 4: High-throughput genome-wide siRNA screen identifies important genes for improved heterologous protein expression

Abbreviations:

RNAi, RNA interference; siRNA, small interfering RNAs; RISC, RNA-induced silencing complex; FBS, fetal bovine serum; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAD, median absolute deviation; RSA, redundant siRNA activity; NTSR1, neurotensin receptor type I; SERT, serotonin transporter; GFP, green fluorescent protein; RNP, ribonucleoprotein; snRNA, small nuclear RNA; hnRNP, heterogeneous nuclear ribonucleoprotein.

4.1 - Summary

Large-scale RNA interference has been developed and utilized as a revolutionary tool in deducing gene functions in many diseases. However, the genome-wide loss-of-function data is very limited for heterologous protein production process. In this chapter, HEK293 cells constitutively expressing luciferase reporter were subject to a genome-wide siRNA screen. Among the 21,585 genes that were individually knocked down, 56 genes were selected for validation screen and top 10 genes leading to the greatest improvement of luciferase production were identified. Interestingly, from KEGG pathway analysis, genes significantly improving luciferase expression were found to be highly enriched in splicesome pathway. In addition, the effects of top 10 genes on secreted and membrane proteins were investigated and the co-transfection of different top 10 siRNAs was also explored.

4.2 - Introduction

RNA interference (RNAi), firstly discovered as a natural biological process of eukaryotic cells to protect the genome against foreign nucleic acids [113, 114], has been developed and utilized as a revolutionary tool in deducing gene functions and in combating genetic defects, viral diseases, autoimmune disorders, and cancers [115]. Following the discovery of double strand RNAs in *Caenorhabditis elegans*[116], chemically synthesized small interfering RNAs (siRNA) were shown to efficiently silence endogenous genes in mammalian cells [117, 118], providing the foundation for developing RNAi applications. SiRNAs are 21-25 nucleotide double strand RNA fragments with symmetric 2-nucleotides 3'-end overhangs [119]. The guide strand of siRNA can be incorporated into RNA-induced silencing complex (RISC), which brings about sequence-specific degradation of the homologous single stranded mRNAs[120].

In recent years, large-scale genetic screens were made possible by the availability of genome-wide siRNA library as well as the development of sophisticated new instrumentation and bioinformatics approaches for data analysis [121, 122]. They have been used to interrogate the biological functions of specific genes and pathways in various diseases[123] and important biological process including signal transduction, cell aging or death, cell or organelle organization, protein localization and responses of host cells to pathogens[124-128]. However, the genome-wide loss-of-function data is very limited for heterologous protein production [124, 129], an important process being intensively investigated by pharmaceutical and biotechnology industry using genetic or metabolic engineering approaches.

75

In this work, we performed a genome-wide siRNA screen to identify genes involved in recombinant protein production using *Photinus pyralis* (firefly) luciferase as a reporter protein. Luciferase is a 62kDa large multidomain enzyme that has been widely used as a reporter in cell biology[130]. As one of the first described large, multidomain proteins, luciferase has been intensively used to study co-translational folding of polypeptide and the role of chaperones in protein folding [131, 132]. Here, with high-throughput format, 21,585 genes were individually knocked down with three different siRNAs in HEK-CMV-Luc2-Hygro cells constitutively expressing firefly luciferase. The results from end-point viable cell number and luciferase activity measurement were cooperated into genome-wide loss-of-function data. Statistical data analysis was executed followed by validation screen, where top 10 genes leading to greatest improvement of luciferase production were confirmed. In this chapter, a brief overview of these top 10 genes was provided and their effects on secreted and membrane proteins were also investigated.

4.3 – Materials and Methods:

4.3.1. High-throughput genome-wide screen for luciferase expression

HEK-CMV-Luc2-Hygro cells constitutively expressing *P. pyralis* luciferase (Progema) were maintained in DMEM containing 10% fetal bovine serum (FBS) in a humidified incubator set at 37ºC and 5% CO_2. The library for primary screen is the Silencer® Select Human genome siRNA library (Ambion), which targets 21,585 human genes with 3 siRNAs per gene. Each siRNA is arrayed in an individual well. The

transfection was done in duplicates: 0.8 pmol of each siRNA was spotted to 384 well plate wells (Corning) and 20 μL of serum-free DMEM containing 0.15 μL of Lipofectamine RNAiMax (Life Technologies) was then added to each well. This lipid-siRNA mixture was incubated at ambient temperature for 30 min prior to adding 4000 cells in 20 μL of DMEM containing 20% FBS (Gibco). After incubating transfected cells at 37ºC in 5% CO_2 for 72 hours, 20 μL of ONE-Glo™ Reagent(Promega) was added to one set of replicates for luciferase activity quantification and 20 μL of CellTiter-Glo™ Reagent(Promega) was added to the second set of replicates for viable cell density measurement. All plates were incubated at room temperature for 20 minutes to stabilize luminescent signal and the signal was then measured with PerkinElmer Envision 2104 Multilabel plate reader. All plates had a full column (16 wells) of Silencer Select Negative Control #2 (Life Technologies) for data normalization and a full column of AllStars Hs Cell Death Control siRNA was also used as on-plate reference for transfection efficiency.

4.3.2. Statistical analysis of primary screen data and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis

For each plate, the median value of negative control column was set as 100% and was used to normalize corresponding sample wells. The luciferase activity and viable cell density was exported as % of negative control and the median absolute deviation (MAD) - based z-score was calculated for each sample [93]. The redundant siRNA activity (RSA) analysis was performed as described[133] to rank candidate genes that enhance luciferase activity. Briefly, luciferase activity data for each sample was subject to

iterative hypergeometric test and p-values indicating the statistical significance of all wells targeting a single gene distributed toward the top ranking slots were generated for each gene. All genes with $p < 0.005$ were imported for KEGG pathway analysis. For the pathway enrichment map, red color indicates lower p-value ($p<0.001$) thus higher rank of the genes; while pink color denotes lower rank ($0.001<p<0.005$).

4.3.3. Validation screen

For each of the top 54 genes selected from the primary screen, 3 independent Silencer® siRNAs (Ambion) with different sequences from the primary screen were seeded into 384-well plates. The transfection and assay process is the same as the primary genome-wide screen.

4.3.4 Validation with GPC3-hFc-expressing cells

HEK- GPC3-hFc cell line constitutively secreting glypican-3 hFc-fusion protein (GPC3-hFc)[94] is a kind gift from Dr. Mitchell Ho's group from National Institutes of Health. Cells were grown in DMEM supplemented with 10% FBS in a humidified incubator set at 37ºC and 5% CO_2.

Validation transfections were performed in 12-well plates with Silencer or SilencerSelect siRNAs and SilencerSelect Negative Control #2 (Ambion). AllStars Cell Death Control (Qiagen) was used as lethal control for transfection efficiency. Cells were transfected in 12-well plates as described except 0.15 million cells were transfected with 40nM siRNA using 3.75ul Lipofectamine RNAiMax in a total volume of 1mL of media.

5 days after transfection, cell culture supernatant was collected and cleared using centrifuge for GPC3-hFc concentration determination with ELISA and cells were detached and counted by trypan blue exclusion using a CEDEX cell quantification system (Roche, Mannheim, Germany).

4.3.5. ELISA for GPC3-hFc concentration determination

AffiniPure F(ab')$_2$ Fragment Goat Anti-Human IgG (min X Bov, Ms, Rb Sr Prot, Cat. 109-006-170, Jackson Immunology) was used to coat a 96-well plate at 5µg/mL in PBS buffer, 50 µL per well, at 4 °C overnight. After the plate was blocked with 2% BSA in PBS buffer, pre-diluted cell culture supernatant was added, and the plate was incubated at room temperature for 1 h to allow binding to occur. After the plate was washed twice with PBS buffer containing 0.05% Tween 20, Peroxidase-conjugated AffiniPure Goat-anti-uman IgG (Cat. 109-035-098, Jackson Immunology) was added at 1:4000 dilution, 50ul/well. Following incubating at room temperature for 1 hr, the plate was washed 4 times and detected with Peroxidase Substrate System (KPL).

4.3.6. Validation with cells producing membrane proteins in fusion with GFP

HEK293 cell lines producing neurotensin receptor (NTSR1)-GFP fusion protein and serotonin transporter (SERT)-GFP fusion protein were independently tested with top 10 siRNAs. The stable SERT-GFP-expressing cell line is a kind gift from Dr. Chris Tate [92](Andréll & Tate unpublished), and NTSR1-GFP-expressing T-REx-293 cell line was previously constructed. Both cell lines were maintained as an adherent culture in DMEM

containing 10% certified FBS and 5 µg/mL blasticidin and 200 µg/mL zeocin (Invitrogen).

Cells were transfected as described except 0.15 million cells were transfected with 40nM siRNA using 6.25ul Lipofectamine RNAiMax in a total volume of 1mL of media. 72 hours after transfection, cells were induced with 1µg/mL tetracycline. 24 hours later, cells from each well were detached with non-enzymatic cell dissociation buffer (Gibco, Cat. No. 13150-016) and washed twice with cold PBS. Cell densities and viability were determined by trypan blue exclusion using a CEDEX cell quantification system (Roche, Mannheim, Germany). Based on the counts, cell densities were adjusted to 0.5 million cells/ml with PBS and then subject to flow cytometry analysis.

4.3.7. Flow cytometry analysis

Cells harvested from validation transfection step were diluted to 0.2 million cells/ml with cold PBS for flow cytometry analysis. Green fluorescence was measured with Guava Easycyte 5HT and Incyte software (Millipore). The green fluorescence signal and cell gating were adjusted using uninduced T-REx-293-NTSR1-GFP cells, with more than 99.5% of the cells in low fluorescence range (<100). The setting was kept same for acquisition of all cell samples.

4.3.8. Co-transfection with two of top 10 siRNAs

HEK-CMV-Luc2-Hygro cells (Progema) were transfected in quadruplicate in 96-well plates as described except 2.4×10^4 cells were transfected using 0. 9µl Lipofectamine RNAiMax in a total volume of 120µl media. Each well contains two of top 10 siRNAs,

each with final concentration of 40nM. Luciferase activity was measured by ONE-Glo™ assay and viable cell density was measured by CellTiter-Glo™ assay. Wells containing 80nM of Silencer Select Negative Control #2 siRNA were used to normalize all sample wells.

4.4 - Results

4.4.1. Primary genome-wide screen identifies a number of genes that enhance CMV driven luciferase expression in HEK293 cells.

We conducted a human genome siRNA screen to identify genes and pathways associated with CMV driven luciferase activity. Screening employed a siRNA library targeting 21,585 human genes with 3 separately arrayed siRNAs per gene in HEK-CMV-Luc2-Hygro cells. For the screen, the transfection was done in duplicates: one set of plates was used for overall luciferase yield measurement and the other set was used for viable cell number determination, from which the per cell luciferase expression can be calculated (Fig. 4.1A). Based on overall luciferase yield, the distribution of siRNA activity is illustrated in the histogram (Fig.4.1B). 1,778 siRNAs were able to significantly enhance luciferase expression (MAD-based z-score>3, or 40% to 178% increase than negative control), and they were categorized as 'strong enhancers'. From the 1,778 siRNAs, we identified 56 genes, which were targeted by at least two independent siRNAs from this 'strong enhancer group' and they were subject to validation screen.

The overall luciferase yield could be resulted from either improved viable cell number, or improved per cell expression, or both. We found that 11,207(17.3%) siRNAs

improved per cell luciferase expression by more than 20% (Fig. 1C quadrant I&II), while only 254(0.4%) siRNAs achieved more than 20% enhancement in viable cell number (Fig. 4.1C quadrant I&IV). Surprisingly, only 2 siRNAs were capable of improving both per cell luciferase expression and viable cell density by more than 20%.

4.4.2. Validation screen for top 10 gene confirmation

To further investigate many of the top candidates, additional three siRNAs were tested for the 56 'strong enhancer' genes selected from 4.4.1 (Fig. 1B). The primary screen and validation screen data for the 56 genes were combined and top 10 genes (Table 4.1) were identified based on the criteria that least 3 out of 6 siRNAs tested had led to MAD-based z-scores higher than 3.0. Viable cell number was also taken into consideration to remove candidates with significant toxicity. The overall luciferase yield, per cell luciferase yield and viable cell numbers associated with the 6 siRNAs targeting one of the top 10 genes were summarized in Supplemental Table 4.S1. The median value of the overall luciferase yield for each gene was calculated from the 6 siRNAs (Table 4.1). In summary, the median value of overall luciferase yield was improved by 24% to 72% than negative control cells and the median of MAD-based z-scores range from 2.13 to 4.55.

Interestingly, 4 out of top 10 genes (*ints1, ints2, hnrnpc* and *prpf19*) are involved in mRNA splicing process. They all encode important proteins for splicesome formation, e.g. integrator complex, heterogeneous nuclear ribonucleoprotein and pre-mRNA processing factor 19. The rest of top genes encode proteins covered a wide span of biological functions, including cell growth and division, signal transduction, apoptosis,

regulation of cellular polyamine concentration, protein translation and folding (Table 4.1).

4.4.3. KEGG pathway analysis identifies splicesome as a statistically significantly enriched pathway for top genes

To identify the pathways where the top candidate genes are enriched in, the redundant siRNA activity (RSA) analysis [133] was performed to rank genes enhancing luciferase activity. Then 362 genes with p <0.005 from RSA analysis (Supplemental Table.4.S2) were imported for KEGG pathway analysis. Among the 362 genes, 28 were involved in splicesome pathway (p=6.4^{-23}, Fig. 4.2). The genes with higher rank (p<0.001, red) encode for important splicesome components, e.g. small nuclear ribonucleoprotein (Sm), splicing factor 3A (SF3a), heterogeneous nuclear ribonucleoproteins (hnRNPs), pre-mRNA-processing factor 19 (Prp19), etc.(Fig. 4.2).

4.4.4. Test of top 10 genes with secreted and membrane proteins

In order to investigate if the knock-down of identified genes will contribute to similar improvement in HEK293 cell lines expressing other proteins, three cell lines were tested. HEK- GPC3-hFc cell line can constitutively secrete glypican-3 hFc-fusion protein (GPC3-hFc) [94] and was tested in this study as a representative of antibody secreting cell lines. T-REx-293-NTSR1-GFP cell line has been constructed previously for production of functional neurotensin receptor type I (NTSR1) in chapter 3. T-REx-293-SERT-GFP [92](Andréll & Tate unpublished) is an inducible cell line for high level expression of serotonin transporter (SERT), a hard-to-express 12 transmembrane domain

protein. Both NTRS1 and SERT were fused with GFP to their C-terminal, allowing proximal protein quantification with flow cytometry. As shown in Fig4.3, the top 10 siRNAs demonstrated various effects on secreted and membrane proteins. Notably, the knocking down of *ints1, hnrhpc, oaz1* and *ppp2r1a* consistently improved the expression of all proteins tested. However, the knocking down of *ints1* and *hnrnpc* led to significantly reduced viable cell number, indicating these genes may be essential for cell survival or cell growth. Silencing of *oaz1* and *ppp2r1a* gene showed minimal negative effects on viable cell number.

4.4.5. Co-transfection with top 10 siRNAs

To investigate the combinatorial effect of top siRNAs, HEK-CMV-Luc2-Hygro cells were transfected with two of top 10 siRNAs, each siRNA with final concentration of 40nM. Per cell luciferase production upon co-transfection was calculated from overall luciferase activity and viable cell density and was summarized in Fig. 4.4. Cells transfected with 80nM of negative control siRNA were used for normalization and their per cell luciferase expression level was set as 100. Depending on the genes of choice, the combinatorial effect can be positive or negative. For example, 80nM of siRNA against *hnrnpc* and *ppp1r2a* independently led to luciferase level of 299 and 251. When 40nM of both siRNAs were co-transfected while keeping the overall siRNA concentration at 80nM, the luciferase level went up to 435. The negative effects can be observed with the co-transfection of siRNAs for *casp8ap2* and *eef1b2*. They each independently led to luciferase level of 121 and 170 but the combination brought luciferase expression down to 94.

We chose the best combination from binary co-transfection (*hnrnpc* and *ppp1r2a*) and added a third siRNA to explore further improvement. Maintaining the overall siRNA concentration at 80nM, 27nM of each siRNAs were co-transfected and the best combination is *ints1* with *hnrnpc* and *ppp1r2a*. Altogether, they led to luciferase level of 463 and the addition of a fourth siRNA didn't improve the expression level further (data not shown).

4.5 - Discussion:

Genome-wide RNAi screening has emerged to be a revolutionary and powerful tool for interrogating gene functions and for target discovery in various diseases [124-128]. However, it has been scarcely used to identify targets for enhancement of recombinant protein production process. The screening in CHO cells was limited to miRNA library [87, 96, 97], as a proper synthetic hamster siRNA library is still lacking. To our knowledge, there have been two publications investigating protein secretion using human whole-genome siRNA libraries [124, 129]. However, both were focusing on disruption of secretory pathway with siRNA-mediated gene depletion as their goal was to identify genes required for protein secretion process. To the contrary, we identified genes that need to be down-regulated for improved recombinant protein production. Besides, our screening is not limited to secretion process. Indeed, the identified gene targets covered a variety of functions, including DNA replication, mRNA splicing, translation, apoptosis mediation, cell growth arrest, etc. (Table 4.1).

In contrast to miRNA screening from chapter 3, which requires intensive follow-up work on gene targets identification through microarray or next-gen sequencing studies, siRNA screening gives direct correlation between the arrayed siRNAs and gene targets. This allows faster gene indentification and construction of stable knock-out cell lines for industry.

In this work, in order to aid in the discovery of new genes that are potential engineering targets for improved production of biopharmaceuticals in mammalian expression system, a HEK293 reporter cell line expressing luciferase reporter was subject to interference with 64,755 siRNAs targeting 21,585 human genes. 1,778 siRNAs strongly improved luciferase expression by having MAD-based z-core>3, which corresponded to 2.7% of the library. In order to exclude the 'false positives' introduced by off-target effects, gene hits were only considered 'true positive' if 2 or more single siRNAs targeting this gene passed the MAD-based z-core>3 requirement. 54 genes were selected with this stringent requirement and were subject to validation screen with 3 more siRNAs for each gene. Finally, data generated from 6 siRNAs for each of the 54 genes (3 siRNAs from primary screen and 3 siRNAs from validation screen) were combined for the selection of top 10 genes. The gene is selected if 3 or more siRNAs yield >3 MAD-based z-scores. This high statistical significance also corresponds to biologically relevance, which is 40% increase in luciferase activity.

To identify cellular functions or biological pathways correlated with improved luciferase expression, KEGG pathway analysis was done to discover important pathways where top gene hits were enriched. The top genes list and their statistical significance

were generated from the RSA analysis, a probability-based approach to circumvent off-target effects by analyzing the collective behavior of all wells targeting the same gene[133]. While $p < 0.05$ in RSA analysis indicates statistical significance of a single gene being distributed toward the top ranking slots, we chose a more stringent cutoff, $p < 0.005$. Then, 362 genes (supplemental table.4.S2) were selected based on this criteria and were used for KEGG pathway analysis. Surprisingly, 28 out of 362 genes were involved in splicesome, a multimegadalton ribonucleoprotein (RNP) complex for pre-mRNA splicing [134]. The 28 genes encode many important splicesome components such as splicing factors (sf3a3, sf3b2, sf3b3, sf3b4) and small nuclear ribonucleoproteins (snrnp200, snrpb, snrpd2, snrpd3, snrpe, snrpf, snrpg). This finding is counter-intuitive as the knock-down of these genes should theoretically compromise mRNA-splicing and could negatively regulate recombinant protein production.

In good agreement with primary screen pathway analysis, splicesome was also highlighted by validation screen result (Table 4.1). Four (*ints1, ints2, hnrnpc* and *prpf19*) out of top ten genes were associated with splicesome pathway. *Ints1* and *ints2* encode two subunits of integrator complex, which indicates the involvement of this complex in influencing protein overexpression. Indeed, the integrator complex contains at least 12 proteins in humans [135] and based on primary screen data, knocking down of *ints1, ints2, ints3, ints4, ints5, ints8* and *ints12* all resulted in significantly improved luciferase reporter expression (Supplemental Table.4.S3). Integrator complex mediates small nuclear RNAs (snRNAs) 3' processing [135] and snRNAs are essential for the removal of introns, proper expression of histone mRNA and biosynthesis of ribosomal RNA[136]. Thus the interference of the complex is found to disrupt pre-mRNA processing [135].

The knocking-down of *ints1, ints5, ints6, ints9* and *ints11* has been reported. RNA interference of *ints1, ints5* or *ints11* resulted in the accumulation of misprocessed snRNA and ultimately splicing defects in pre-mRNAs for multiple genes [135, 137]. *Ints6* is presumed to be tumor-suppressor gene as overexpression of *ints6* in a prostate cancer cells reduces colony formation and causes a cell-cycle arrest[138]. RNAi-mediated down-regulation of *ints11* leads to G1 arrest in human cells [139]. It is not clear why the down-regulation of integrator complex genes, which would presumably inhibit snRNA biosynthesis, leads to higher recombinant protein production. It could be resulted from splicing defects in specific genes, who are negative regulators of protein expression. Another possibility is that the depletion of integrator complex leads to growth arrest, thus allowing more energy and resources to be channeled to protein production process [85, 112].

The other two genes in splicesome pathway are *hnrnpc* and *prpf19*. They encode Heterogeneous Nuclear Ribonucleoprotein C (C1/C2) and Pre-mRNA Processing Factor 19 respectively and are both associated with pre-mRNA splicing. Heterogeneous Nuclear Ribonucleoprotein C is a core component of heterogeneous nuclear ribonucleoproteins (hnRNPs) and is one of the most abundant proteins in the nucleus[140]. It's well known that it's involved in alternative splicing of pre-mRNA[141] and stabilization of mRNA[142]. However, its exact role in splicing regulation remained contradictory and unresolved[143]. Pre-mRNA Processing Factor 19 is a component of the prp19p-associated complex (or NTC), which activates splicesome for pre-mRNA splicing [144]. It is also found to be essential for cell survival and damage DNA repair [145]. *Prpf19* up-regulation has been demonstrated to expand life span of human endothelial cells while its

depletion led to reduced resistance to apoptosis [146, 147]. In our luciferase primary and validation screen, the siRNA-mediated knocking-down of *prpf19* gene didn't lead to significant cell death.

The rest of top genes encode proteins covering a wide range of biological functions.

Casp8ap2 encodes caspase 8 associated protein 2, which is highly similar to the mouse apoptotic protein FLASH. It is shown to regulate caspase 8-induced activation of NF-kappa-B and it's required for S-phase progression and histone gene transcription [148, 149].

Oaz1 encodes ornithine decarboxylase antizyme 1, which down-regulates cellular polyamine level by binding to and inhibiting ornithine decarboxylase [150]. This gene will be investigated in detail in the next chapter.

Ppp2r1a encodes an alpha isoform of the constant regulatory subunit A (PR 65) of protein phosphatase 2A(PP2A) and it serves as a scaffolding molecule[151]. PP2A is an important regulator in cell proliferation, signal transduction and apoptosis and it is implicated in the negative control of cell growth and division [152].

Chaf1a encodes subunit A of chromatin assembly factor 1 (CAF-1), which assembles histone octamers onto replicating DNA [153]. CAF-1 is a histone chaperone that plays important roles in chromatin restoration after DNA synthesis, cell cycle progression, heterochromatin maintenance and asymmetric cell division[154].

Cct2 encodes beta subunit of cytosolic chaperonin-containing t-complex polypeptide-1(CCT), which is a molecular chaperone important for folding of actin, tublulin and numerous other proteins eukaryotic cytosol[155]. Depletion of CCT were

associated with growth arrest and perturbation of actin-based cell motility in mammalian cells and it is believed to be required for cell cycle progression and cytoskeletal organization[156].

Eef1b2 encodes beta subunit of eukaryotic translation elongation factor 1. The protein is a guanine nucleotide exchange factor involved in recruitment of aminoacyl-tRNAs onto the ribosome, which is the first step of elongation[157].

To test the effect of silencing top genes for expression of other proteins, one secreted protein(GPC3-hFc) and two hard-to-express membrane proteins (neurotensin receptor type I and serotonin transporter) were investigated. For each cell line, at least 6 out of 10 siRNAs tested improved the recombinant protein production. Notably, 4 genes were always improving protein production in all three cell lines: *ints1, hnrhpc, oaz1* and *ppp2r1a*. It is found that *ints1* or *hnrnpc* depleted cells consistently have growth disadvantages (Fig. 4.3, viable cell number), indicating they could be essential genes. Indeed, targeted disruption of *ints1* in mouse embryos results in growth arrest followed by apoptotic cell death[158]. Taken together, *oaz1* gene was chosen for follow-up studies in the next chapter since its knock-down was associated with minimal cell growth disadvantage.

Contributions from collaborators:

Dr. Scott Martin contributed to conceiving and designing of RNAi screen experiments and data analysis. Yu-Chi Chen carried out luciferase siRNA primary screening.

Figures and tables

Fig. 4.1. Genome-wide human siRNA library screen with HEK-CMV-luc2-Hygro cell line. (A) Workflow of the screen; (B) Distribution of siRNA activity on improved overall luciferase expression; top hits (MAD-based z-score>3.0) are highlighted; (C) For each siRNA sample, the relative per cell luciferase yield is plotted against relative viable cell number. The 20% increase cutoffs are highlighted and they divide the entire population into four quadrants (I, II, III and IV).
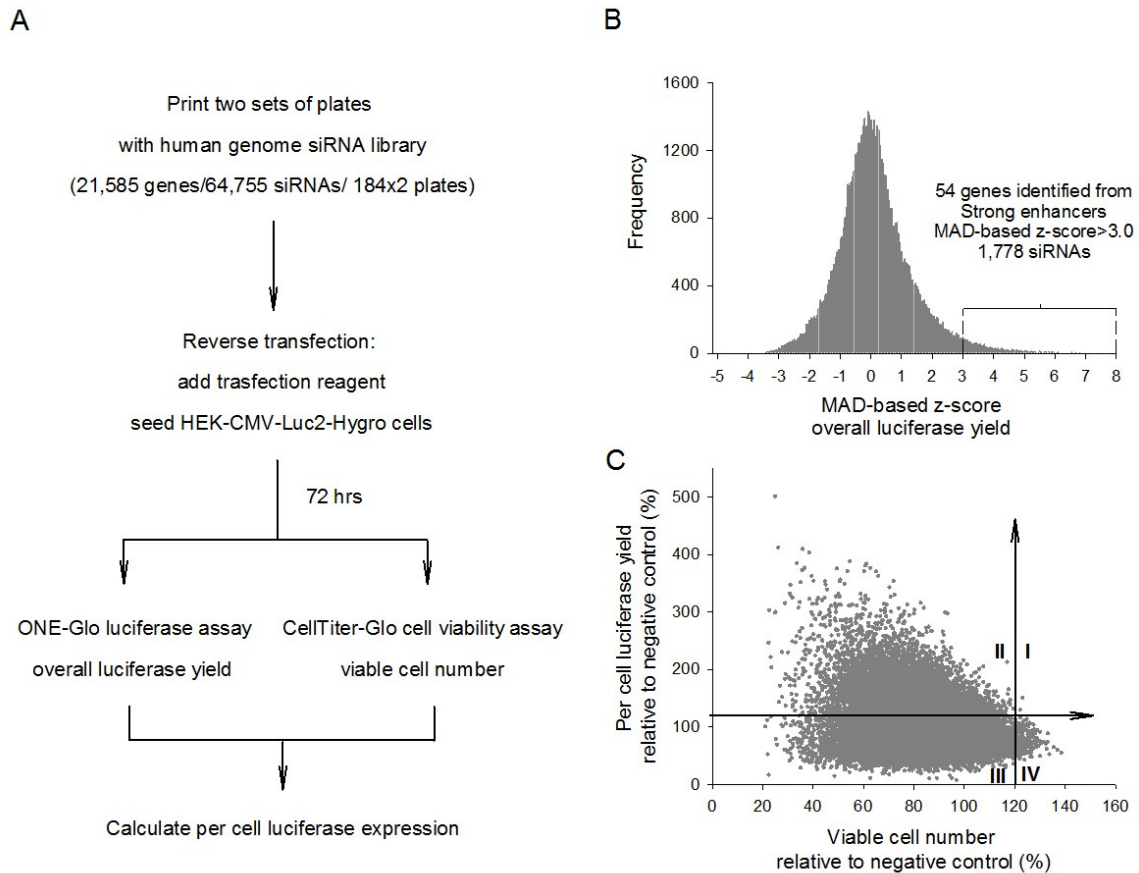
Fig.4.2. Illustration of splicesome pathway enriched with primary screen top hits. All 362 genes significantly improved luciferase expression (p<0.005 in RSA analysis) were imported for KEGG pathway mapping. 28 genes were found in splicesome pathway and were highlighted in red or pink in the map. Red color indicates genes with p<0.001thus higher rank; while pink color denotes lower rank (0.001<p<0.005).
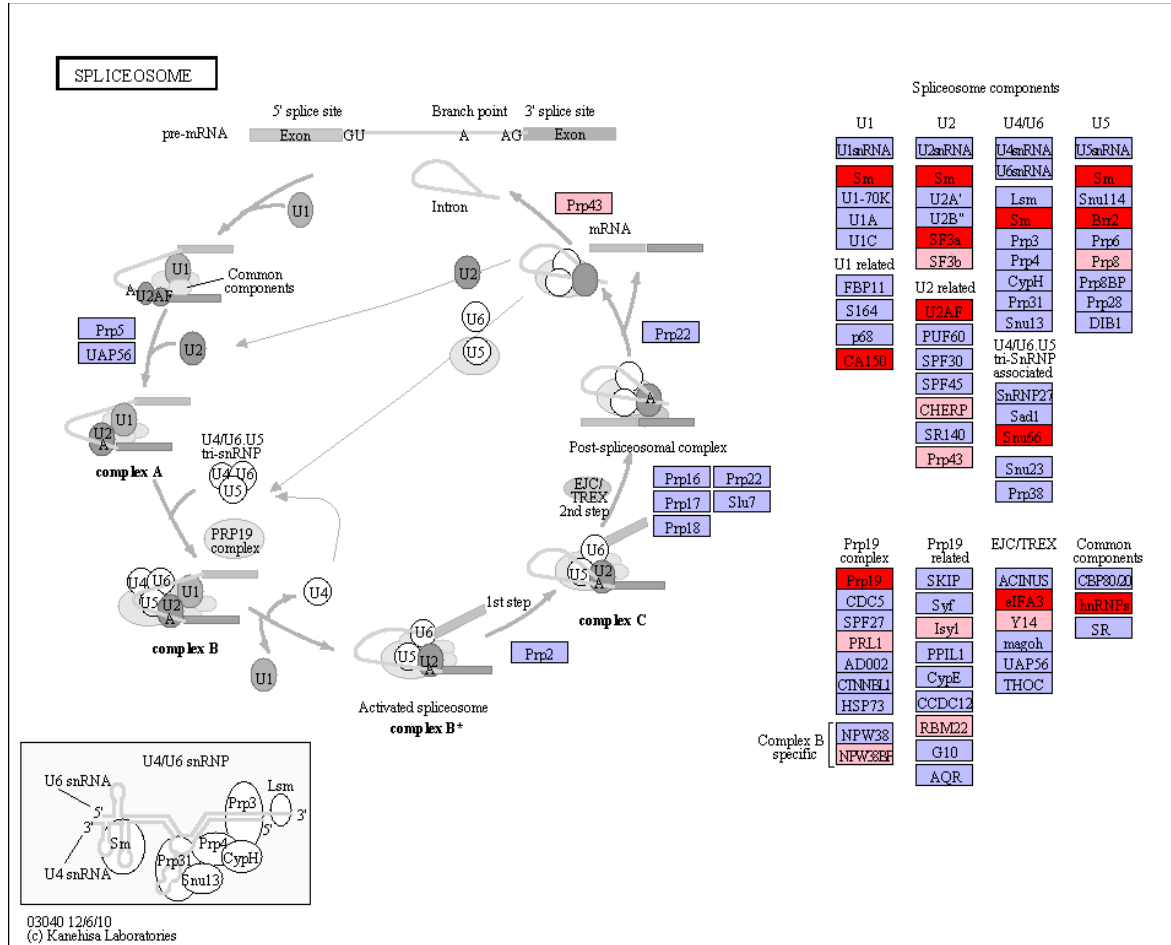
Fig.4.3. Test of top 10 siRNAs with (A) luciferase, (B) GPC3-hFc, (C) NTSR1-GFP and (D) SERT-GFP expressing HEK cell line. All protein expression was normalized to cells tranfectecd with the negative control (siN.C.). The experiment was performed twice with different passages of cells. For each biological sample, the measurement was done in duplicates. Error bars indicate SEM.
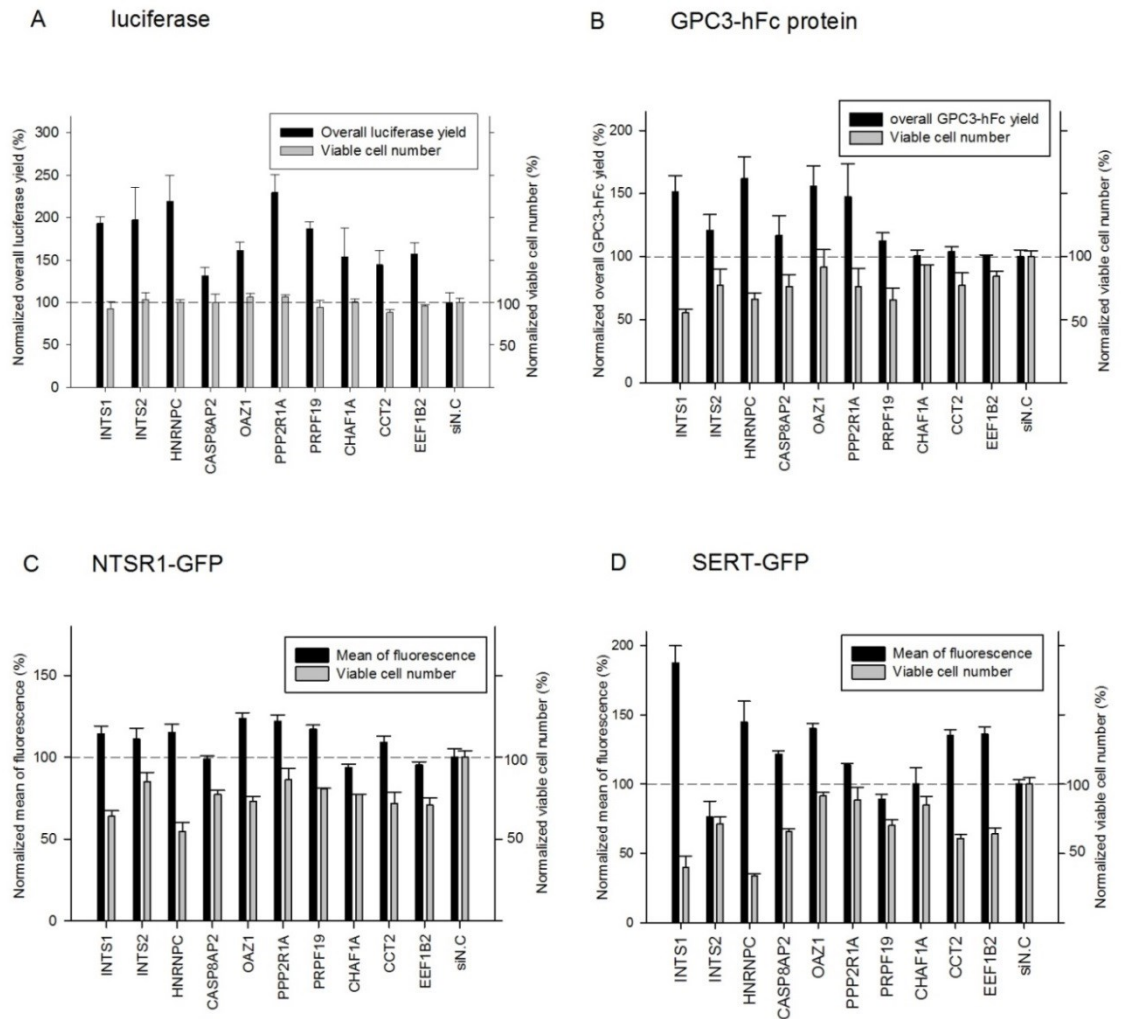
Fig.4.4. Co-transfection of binary siRNA mixtures from top 10 siRNAs. HEK-CMV-Luc2-Hygro cells were transfected with two of top 10 siRNAs, each with final concentration of 40nM. Luciferase level of cells transfected negative control siRNA(siN.C.) was set at 100 and all co-transfection result was normalized. Subscripts indicate SEM. The experiment was performed twice with different passages of cells. For each biological sample, the measurement was done in duplicates.

	INTS1	INTS2	HNRNPC	CASP8AP2	OAZ1	PPP1R2A	PRPF19	CHAF1A	CCT2	EEF1B2	
INTS1	$186_{\pm8}$	$174_{\pm22}$	$346_{\pm31}$	$153_{\pm16}$	$209_{\pm26}$	$253_{\pm21}$	$189_{\pm35}$	$185_{\pm36}$	$149_{\pm12}$	$152_{\pm16}$	INTS1
INTS2		$159_{\pm26}$	$342_{\pm45}$	$144_{\pm24}$	$194_{\pm28}$	$294_{\pm29}$	$202_{\pm31}$	$130_{\pm14}$	$124_{\pm16}$	$144_{\pm14}$	INTS2
HNRNPC			$299_{\pm30}$	$260_{\pm33}$	$304_{\pm37}$	$435_{\pm13}$	$327_{\pm23}$	$296_{\pm39}$	$210_{\pm41}$	$260_{\pm23}$	HNRNPC
CASP8AP2				$121_{\pm14}$	$149_{\pm25}$	$236_{\pm23}$	$146_{\pm15}$	$111_{\pm9}$	$136_{\pm16}$	$94_{\pm10}$	CASP8AP2
OAZ1					$151_{\pm18}$	$311_{\pm14}$	$172_{\pm4}$	$152_{\pm8}$	$159_{\pm3}$	$106_{\pm9}$	OAZ1
PPP1R2A						$251_{\pm35}$	$285_{\pm41}$	$222_{\pm27}$	$197_{\pm20}$	$204_{\pm27}$	PPP1R2A
PRPF19							$179_{\pm13}$	$151_{\pm11}$	$132_{\pm19}$	$123_{\pm19}$	PRPF19
CHAF1A								$120_{\pm10}$	$134_{\pm15}$	$105_{\pm17}$	CHAF1A
CCT2									$127_{\pm16}$	$154_{\pm14}$	CCT2
EEF1B2										$170_{\pm11}$	EEF1B2

Table.4.1. Confirmed top 10 genes with 3 or more siRNAs yielding MAD-based z-score>3

Gene	Description	Overall luciferase yield (%)[*,†]	MAD-based z-score[*]	Function
INTS1	Integrator Complex Subunit 1	172	4.55	3'- end processing of small nuclear RNAs U1 and U2
INTS2	Integrator Complex Subunit 2	165	4.17	3'- end processing of small nuclear RNAs U1 and U2
HNRNPC	Heterogeneous Nuclear Ribonucleoprotein	163	4.10	Influencing pre-mRNA processing and other aspects of mRNA metabolism and transport
CASP8AP2	Caspase 8 Associated Protein 2	156	3.70	Regulation of CASP8 in FAS-mediated apoptosis
OAZ1	Ornithine Decarboxylase Antizyme	153	3.57	Inhibiting ornithine decarboxylase and inactivating the polyamine uptake transporter
PPP2R1A	Protein Phosphatase 2, Regulatory Subunit A, Alpha	153	3.56	Serving as a scaffold for Protein Phosphatase 2 assembly, essential for signal transduction pathways
PRPF19	Pre-mRNA Processing Factor 19	147	3.27	Spliceosome assembly and activating pre-mRNA splicing
CHAF1A	Chromatin Assembly Factor 1, Subunit A	138	2.80	mediating chromatin assembly in DNA replication and DNA repair
CCT2	Chaperonin Containing TCP1, Subunit 2 (Beta)	126	2.23	Chaperonin-mediated protein folding of actin, tubulin and other proteins
EEF1B2	Eukaryotic Translation Elongation Factor 1 Beta 2	124	2.13	exchanging GDP bound to EF-1-α to GTP during the transfer of aminoacylated tRNAs to the ribosome

[*] All values are medians of result from 6 siRNAs(3 siRNAs in primary screen and 3 siRNAs in validation screen) targeting a top gene.

[†] Values are normalized to negative control siN.C. transfected cells (set as 100%).

Table.4.S1. Effects of knocking down of top 10 genes with six different siRNAs from primary and validation screens.

Gene	SiRNA sequence	Overall luciferase yield		Per cell luciferase yield		Viable cell number	
		Fold change(%)	z-score	Fold change(%)	z-score	Fold change(%)	z-score
INTS1	GCAUGAGCAAACUCCUCCAtt	200.62	5.97	238.61	2.58	100.63	0.8
	GUUCAUCCAUAAGUACAUUtt	198.35	5.86	153.44	-1.1	83.13	-0.56
	GGGUUUGUCGCUGGUGCUUtt	197.22	5.8	120	-2.55	94.71	0.34
	AGAUCUUUGUCAAGGUGUAtt	147.69	3.31	199.36	4.75	96.26	0.46
	GCAGGUCCUCUAUACCGCAtt	127.12	2.27	208.24	5.13	105.93	1.21
	CGCCUCCAUCAACUUCAAGtt	74.22	-0.39	76.63	-0.56	96.85	0.5
INTS2	GCGUAUUUUGAGAGUACUUtt	235.6	7.73	185.52	0.28	94.32	0.31
	GCACCCGAAUUGUGGAAGAtt	176.86	4.78	156.98	-0.95	105.67	1.19
	GACAUUGGAUCAUACUAAAtt	164.94	4.18	165.71	-0.57	88.91	-0.11
	GCAGCUUAGGCAUAAACUUtt	164.78	4.17	249.79	6.92	108.47	1.4
	GGCGAAUGCUCCUGACUAAtt	163.08	4.08	167.38	3.36	103.89	1.05
	GCAUGGAUCCUGAUGUACAtt	115.82	1.71	151.92	2.7	69.89	-1.58
HNRNPC	GAUGAAGAAUGAUAAGUCAtt	177.94	4.83	184.2	0.23	96.6	0.48
	ACACUCUUGUGGUCAAGAAtt	168.09	4.33	227.74	2.11	73.81	-1.28
	GCAGGUGUGAAACGAUCUGtt	167.64	4.31	163.5	-0.67	91.28	0.07
	CAACGGGACUAUUAUGAUAtt	159.23	3.89	183.65	4.07	97.39	0.55
	GGCAAUCUCAUUUAGUUGAtt	149.57	3.4	156.71	2.9	95.45	0.39
	GGCAAUCUUUUCGAAGUAUtt	128.44	2.34	121.86	1.4	105.4	1.17
CASP8AP2	CCCUGUUCAUUAUAAGUCUtt	216.11	6.75	192.64	4.46	79.25	-0.86
	CCAACAAGGAAGACGAAAAtt	191.23	5.5	83.35	-0.27	67.32	-1.78
	GGAUAUUGGAGGCUAGUCAtt	169.46	4.4	82.15	-0.32	87.97	-0.18

	GGAUAUUGGAGGCUAGUCAtt	141.57	3	272.68	4.05	89.68	-0.05
	GGCUCACUGGACAUAUACGtt	83.8	0.1	284.06	4.54	100.53	0.79
	GGCAACAUAUAAUGAUUUGtt	82.52	0.03	157.87	-0.91	100.45	0.78
OAZ1	CCGUAGACUCGCUCAUCUCtt	174.47	4.65	204.23	4.96	85.43	-0.38
	GCUAACUUAUUCUACUCCGtt	171.15	4.49	154.72	2.82	110.62	1.57
	GCCUUGCUCCGAACCUUCAtt	161.2	3.99	90.31	0.04	94.84	0.35
	GAUUAUCCUUGUACUUUGAtt	144.54	3.15	169.97	-0.39	101.9	0.89
	GGCUGAAUGUAACAGAGGAtt	127.7	2.3	141.84	-1.6	94.98	0.36
	GGGAAUAGUCAGAGGGAUCtt	92.81	0.55	134.45	-1.92	102.77	0.96
PPP2R1A	GGUCAAAGAGUUCUGUGAAtt	208.07	6.34	161.61	-0.75	103.63	1.03
	CUUCGACAGUACUUCCGGAtt	168.11	4.34	254.01	3.24	104.02	1.06
	GAACAGCUGGGAACCUUCAtt	154.04	3.63	157.24	-0.94	60.65	-2.3
	GGAGUUCUUUGAUGAGAAAtt	151.51	3.5	200.79	4.81	96.36	0.46
	GGCGGAACUUCGACAGUACtt	143.16	3.08	135.32	1.98	105.8	1.2
	GGACCCGAAGUGAGCUUCUtt	84.88	0.15	87.22	-0.1	97.31	0.54
PRPF19	GGUAAAGUCACUGAUCUUUtt	195.28	5.7	191.25	4.4	102.11	0.91
	GCUCAUCGACAUCAAAGUUtt	165	4.18	130.64	1.78	97.01	0.52
	GCGCAAGCUUAAGAACUUUtt	161.12	3.98	96.79	0.32	94.58	0.33
	GGUCACCAGCGUGGUGUUUtt	132.76	2.56	170.09	-0.38	101.62	0.87
	CAACUUUGAGGUAAAGUCAtt	128.76	2.36	170.35	-0.37	95.58	0.41
	GGCCAUACCAAGAAGGUCAtt	92.93	0.56	134.71	-1.91	96.01	0.44
CHAF1A	GCCUGAAUCUUGUCCCAAAtt	188.25	5.35	228.48	2.14	82.39	-0.62
	CGAAACUUGUCAACGGGAAtt	187.37	5.3	179.21	0.01	104.55	1.1
	GAAGAAGACUCUGUACUCAtt	143.57	3.1	215.85	1.59	66.51	-1.85
	CCGACUCAAUUCCUGUGUAtt	131.57	2.5	144.01	2.36	91.36	0.08
	GCAGCUCAAGUUACGUGCAtt	111.6	1.49	121.47	1.38	91.87	0.12

	GCCGAUGACAUGUCAGACGtt	67.27	-0.73	65.03	-1.06	103.44	1.01
	CAUUGGUGUUGACAAUCCAtt	161.46	4	98.81	0.4	85.05	-0.41
	GUUGCAAACUUAUCGAGGAtt	147.74	3.31	107.88	0.79	79.09	-0.87
CCT2	CUCUUAUGGUAACCAAUGAtt	146.72	3.26	77.6	-0.51	91.56	0.09
	GGGUUCAAGAUGAUGAAGUtt	105.94	1.21	189.85	0.47	107.22	1.31
	GGCAUGGACAAAAUUCUUCtt	103.66	1.09	186.79	0.34	96.09	0.44
	GGGAAGCAGAAUCUUUAAUtt	81.36	-0.03	160.25	-0.81	104.85	1.12
	GGAAGAACGUCUUGCACAAtt	156.47	3.75	149.19	2.58	85.47	-0.38
	AGAAAGCUUUGGGCAAAUAtt	151.19	3.48	92.16	0.12	75.57	-1.14
EEF1B2	GGAGUGAAGAAAGCUUUGGtt	144.92	3.17	84.55	-0.21	97.14	0.53
	GAUAAAGUUGGAACAGAUAtt	103.4	1.08	183.07	0.18	110.85	1.59
	GGAAGUGGAGCUACAGAUAtt	98.4	0.83	200.07	0.91	106.78	1.27
	GGAAAGUGAAGAAGCAAAGtt	82.29	0.02	93.28	-3.7	97.32	0.54

Table.4.S2. Top ranking genes from RSA analysis

Gene	RSA analysis rank	RSA analysis p-value
SNRPB	1	2.18E-12
INTS2	2	3.79E-10
HNRNPC	3	1.04E-08
PPP2R1A	4	4.85E-08
DKFZP586J0619	5	1.17E-07
INTS1	6	1.17E-07
U2AF1	7	6.96E-07
SF3B3	8	1.33E-06
SNRPF	9	1.51E-06
PRPF19	10	1.69E-06
OAZ1	11	2.27E-06
SF3B4	12	2.32E-06
SART1	13	5.35E-06
CASP8AP2	14	6.01E-06
SNRPD2	15	6.22E-06
KAT5	16	2.05E-05
LOC642861	17	4.62E-05
CHAF1A	18	4.92E-05
CNOT1	19	5.15E-05
EEFSEC	20	6.22E-05
RDBP	21	7.29E-05
CCT7	22	7.77E-05
CSNK2B	23	8.48E-05
LOC391322	24	9.24E-05
USPL1	25	9.82E-05
LOC731069	26	1.02E-04
SNRNP200	27	1.14E-04
PNMA2	28	1.39E-04
MZF1	29	1.45E-04
CCT2	30	1.61E-04
SNRPE	31	1.80E-04
ACADVL	32	1.85E-04
LOC389722	33	1.89E-04
TCERG1	34	1.94E-04
APOBEC3H	35	2.02E-04
EEF1B2	36	2.07E-04
STAT6	37	2.51E-04
CSE1L	38	2.54E-04
LOC728268	39	2.58E-04
LOC119358	40	2.65E-04
TYW3	41	2.77E-04
LOC729227	42	3.13E-04
CCT4	43	3.17E-04
DNALI1	44	3.70E-04
ALG3	45	4.16E-04
ACAD8	46	4.55E-04
C20orf165	47	4.62E-04
TH1L	48	4.93E-04
FERMT1	49	5.08E-04
CDCA7	50	5.29E-04
CD37	51	5.54E-04

TEX13B	52	5.64E-04		DUSP26	79	8.90E-04
LOC340602	53	5.82E-04		LOC731312	80	9.24E-04
ACSF2	54	5.98E-04		MGRN1	81	9.31E-04
CYLD	55	6.06E-04		LOC728310	82	9.64E-04
LOC340113	56	6.11E-04		MKL2	83	9.70E-04
LOC649259	57	6.33E-04		LOC652595	84	9.70E-04
KCNJ10	58	6.39E-04		NLRC3	85	9.80E-04
IFIT1L	59	6.39E-04		LOC728073	86	1.03E-03
C20orf106	60	6.47E-04		LOC731523	87	1.04E-03
EIF4A3	61	6.56E-04		CNTF	88	1.06E-03
SNRPD3	62	6.59E-04		TST	89	1.06E-03
ACTL6A	63	6.69E-04		OR10P1	90	1.11E-03
SNRPG	64	6.70E-04		PCGF1	91	1.11E-03
ODF2L	65	6.74E-04		SNRPEL1	92	1.11E-03
LOC729316	66	6.75E-04		ANKRD6	93	1.15E-03
LOC727974	67	6.77E-04		FEZF1	94	1.18E-03
LOC645602	68	6.93E-04		STAB1	95	1.18E-03
LOC728232	69	7.14E-04		FLRT1	96	1.20E-03
ARCN1	70	7.25E-04		FLJ43582	97	1.21E-03
PSMD4	71	7.37E-04		PLRG1	98	1.24E-03
ZNF195	72	7.39E-04		RNF138	99	1.24E-03
LOC441282	73	7.85E-04		EFCAB4A	100	1.25E-03
ZFAND2A	74	7.86E-04		GPI	101	1.26E-03
SF3A3	75	8.13E-04		SNX9	102	1.29E-03
C22orf26	76	8.15E-04		RAB31	103	1.33E-03
NFYA	77	8.32E-04		IL17REL	104	1.34E-03
FAM102A	78	8.48E-04		LOC647174	105	1.37E-03

RPTN	106	1.38E-03	L3MBTL4	133	1.77E-03
FXR2	107	1.39E-03	RNF7	134	1.80E-03
COPB2	108	1.39E-03	LOC729658	135	1.80E-03
LOC642656	109	1.40E-03	DHX15	136	1.80E-03
KIAA0947	110	1.40E-03	LOC641796	137	1.80E-03
LOC641845	111	1.42E-03	LOC643916	138	1.81E-03
PPP1R13L	112	1.43E-03	RGS8	139	1.85E-03
DMAP1	113	1.44E-03	RIF1	140	1.88E-03
NUDT21	114	1.48E-03	FLJ40039	141	1.89E-03
LOC728292	115	1.50E-03	LOC646823	142	1.89E-03
ZNF621	116	1.52E-03	GTF2A2	143	1.93E-03
LOC728086	117	1.56E-03	SNX33	144	1.94E-03
NRN1	118	1.57E-03	NKX3-2	145	1.96E-03
KIAA0999	119	1.57E-03	RAB1B	146	1.99E-03
PPP2R4	120	1.57E-03	hCG_1817208	147	2.01E-03
MPHOSPH8	121	1.58E-03	TMEM180	148	2.04E-03
NTF3	122	1.60E-03	OR4K2	149	2.08E-03
LOC650689	123	1.61E-03	SUPT6H	150	2.08E-03
LOC390956	124	1.62E-03	MORC2	151	2.12E-03
TRIM72	125	1.66E-03	PDZD3	152	2.12E-03
PELI3	126	1.69E-03	SMAD3	153	2.15E-03
LOC729017	127	1.71E-03	NMNAT2	154	2.15E-03
L3MBTL2	128	1.71E-03	METTL2B	155	2.17E-03
SP8	129	1.72E-03	C10orf91	156	2.17E-03
PRPF8	130	1.73E-03	FLJ36144	157	2.18E-03
LOC653303	131	1.73E-03	C11orf84	158	2.21E-03
ISOC1	132	1.75E-03	CHERP	159	2.22E-03

SCIN	160	2.22E-03		UPF1	187	2.49E-03
SDHC	161	2.23E-03		SCGB1D2	188	2.51E-03
LOC728977	162	2.23E-03		SNAPC5	189	2.52E-03
ICA1L	163	2.25E-03		G3BP1	190	2.53E-03
VAV2	164	2.26E-03		MIZF	191	2.54E-03
DNAJB12	165	2.26E-03		GPR109B	192	2.55E-03
VSIG8	166	2.28E-03		PLEKHA6	193	2.55E-03
VARS	167	2.30E-03		ZBTB39	194	2.58E-03
CCL23	168	2.30E-03		WDR68	195	2.58E-03
ANKRD1	169	2.31E-03		ZCWPW2	196	2.59E-03
LOC286411	170	2.35E-03		ZNF562	197	2.62E-03
PAQR7	171	2.35E-03		MFSD11	198	2.63E-03
RBM22	172	2.35E-03		MUS81	199	2.63E-03
TFCP2L1	173	2.35E-03		LOC730259	200	2.64E-03
LOC646463	174	2.36E-03		POU5F1	201	2.65E-03
NPAT	175	2.37E-03		TDRD10	202	2.66E-03
ZNF596	176	2.37E-03		CSPG5	203	2.66E-03
C12orf43	177	2.38E-03		CHD1L	204	2.68E-03
ZNF213	178	2.40E-03		KDELR2	205	2.72E-03
ZNF207	179	2.40E-03		TPM3	206	2.73E-03
AMN	180	2.40E-03		WHSC2	207	2.73E-03
SLC4A1	181	2.44E-03		SFRS1	208	2.74E-03
C1orf140	182	2.45E-03		MARK2	209	2.77E-03
SFRS7	183	2.45E-03		C1orf146	210	2.77E-03
OR4C13	184	2.49E-03		MACROD1	211	2.78E-03
CLEC4M	185	2.49E-03		OPN1SW	212	2.82E-03
SULT2A1	186	2.49E-03		ETS1	213	2.83E-03

PPP6C	214	2.84E-03	FAM122B	241	3.23E-03
IL28B	215	2.84E-03	LOC647591	242	3.23E-03
ECHDC3	216	2.86E-03	TACC2	243	3.26E-03
WNT2	217	2.89E-03	KRT71	244	3.26E-03
YTHDF2	218	2.89E-03	TBP	245	3.27E-03
HES4	219	2.91E-03	INTS3	246	3.27E-03
PNPLA6	220	2.91E-03	C14orf80	247	3.28E-03
VMO1	221	2.92E-03	KCNE4	248	3.28E-03
U2AF2	222	2.95E-03	LOC731292	249	3.31E-03
RABGAP1L	223	2.98E-03	CMIP	250	3.31E-03
ITPRIPL1	224	3.02E-03	ZNF519	251	3.32E-03
ALDH3A2	225	3.03E-03	HNRPDL	252	3.33E-03
C20orf144	226	3.05E-03	ISY1	253	3.34E-03
LOC730203	227	3.06E-03	CRYBA2	254	3.35E-03
LOC401577	228	3.08E-03	4-Sep	255	3.35E-03
CDC20B	229	3.09E-03	SF3B2	256	3.35E-03
LOC646051	230	3.10E-03	LOC644403	257	3.37E-03
TNFRSF13C	231	3.11E-03	LOC728677	258	3.37E-03
SULT1A2	232	3.13E-03	ABCB8	259	3.37E-03
GSTZ1	233	3.13E-03	HOXC13	260	3.41E-03
DKFZp781N1041	234	3.13E-03	LOC644961	261	3.42E-03
LOC729619	235	3.14E-03	FOXN1	262	3.42E-03
WFDC2	236	3.14E-03	INTS8	263	3.45E-03
LOC727874	237	3.15E-03	LOC645914	264	3.46E-03
LOC644093	238	3.18E-03	BEST4	265	3.46E-03
TNF	239	3.19E-03	LOC729904	266	3.51E-03
SPRR3	240	3.20E-03	LOC390335	267	3.52E-03

GPR162	268	3.55E-03		LOC653458	295	4.01E-03
C7orf62	269	3.57E-03		SLC16A12	296	4.02E-03
USP49	270	3.60E-03		FLJ44653	297	4.04E-03
UPK1B	271	3.60E-03		SLC15A5	298	4.06E-03
FLJ30403	272	3.64E-03		MGC39606	299	4.07E-03
MRPL24	273	3.65E-03		KCNE3	300	4.10E-03
HSPE1	274	3.67E-03		DDOST	301	4.11E-03
ACTA1	275	3.67E-03		PIP5K1A	302	4.11E-03
KPNB1	276	3.67E-03		LOC643749	303	4.12E-03
AP3M2	277	3.69E-03		COBRA1	304	4.12E-03
CCDC132	278	3.71E-03		C1orf114	305	4.15E-03
3-Mar	279	3.74E-03		ANKRD28	306	4.15E-03
TRAF4	280	3.78E-03		NRXN3	307	4.16E-03
NHP2	281	3.78E-03		MEGF11	308	4.18E-03
BRP44L	282	3.79E-03		RBM8A	309	4.18E-03
VAV1	283	3.80E-03		TSPAN14	310	4.20E-03
LOC732312	284	3.83E-03		C17orf80	311	4.20E-03
CHAT	285	3.88E-03		ASF1A	312	4.25E-03
SAMD11	286	3.88E-03		FLJ11286	313	4.29E-03
PMVK	287	3.90E-03		MS4A4E	314	4.30E-03
ZNF780B	288	3.92E-03		LOC644592	315	4.31E-03
hCG_1657980	289	3.92E-03		MGC42105	316	4.33E-03
PEPD	290	3.93E-03		NUCKS1	317	4.34E-03
CHAF1B	291	3.94E-03		ZNF709	318	4.34E-03
CES7	292	3.97E-03		PROM2	319	4.35E-03
ADRBK2	293	3.97E-03		PABPC1	320	4.36E-03
LOC647855	294	3.98E-03		TLK2	321	4.37E-03

SMAP2	322	4.38E-03	COL29A1	349	4.80E-03
ROMO1	323	4.38E-03	NRAP	350	4.81E-03
KCTD15	324	4.39E-03	DCP1A	351	4.82E-03
WBP11	325	4.41E-03	NEURL2	352	4.84E-03
ADORA2A	326	4.42E-03	THRA	353	4.85E-03
LOC645852	327	4.43E-03	UNC45A	354	4.88E-03
TTC19	328	4.44E-03	LOC388553	355	4.89E-03
SLC11A1	329	4.46E-03	MGC16121	356	4.89E-03
COPA	330	4.52E-03	ZNF534	357	4.91E-03
CD79A	331	4.53E-03	LOC728030	358	4.91E-03
GEMIN8	332	4.53E-03	RBM39	359	4.91E-03
LOC387927	333	4.55E-03	LOC729594	360	4.94E-03
LOC158376	334	4.57E-03	LOC642277	361	4.97E-03
CTBP1	335	4.61E-03	SDR39U1	362	4.98E-03
TMEM174	336	4.61E-03			
GDPD3	337	4.66E-03			
UBE2NL	338	4.67E-03			
PPP2CA	339	4.69E-03			
LOC389365	340	4.71E-03			
TCEAL4	341	4.71E-03			
TLN1	342	4.74E-03			
LOC138652	343	4.76E-03			
ERMP1	344	4.77E-03			
JTV1	345	4.77E-03			
EP400	346	4.78E-03			
OCRL	347	4.79E-03			
SLC25A18	348	4.80E-03			

Table.4.S3. Summary of knocking down of different integrator complex subunits

Gene	SiRNA sequence	Overall luciferase yield		Viable cell number	
		Fold change (%)	z-score	Fold change (%)	z-score
INTS1	GCAUGAGCAAACUCCUCCAtt	200.62	5.97	100.63	0.8
	GUUCAUCCAUAAGUACAUUtt	198.35	5.86	83.13	-0.56
	GGGUUUGUCGCUGGUGCUUtt	197.22	5.8	94.71	0.34
	AGAUCUUUGUCAAGGUGUAtt	147.69	3.31	96.26	0.46
	GCAGGUCCUCUAUACCGCAtt	127.12	2.27	105.93	1.21
	CGCCUCCAUCAACUUCAAGtt	74.22	-0.39	96.85	0.5
INTS2	GCGUAUUUUGAGAGUACUUtt	235.6	7.73	94.32	0.31
	GCACCCGAAUUGUGGAAGAtt	176.86	4.78	105.67	1.19
	GACAUUGGAUCAUACUAAAtt	164.94	4.18	88.91	-0.11
	GCAGCUUAGGCAUAAACUUtt	164.78	4.17	108.47	1.4
	GGCGAAUGCUCCUGACUAAtt	163.08	4.08	103.89	1.05
	GCAUGGAUCCUGAUGUACAtt	115.82	1.71	69.89	-1.58
INTS3	GAGUUGCUAUGACAAUGCAtt	123.34	2.08	89.05	-0.1
	GGUGCGAUUUGGUCAACAAtt	118.51	1.84	93.06	0.2
	CCUGGUUAUGUUUCGAAAAtt	106.78	1.25	108.42	1.39
INTS4	CUAUCUUCGCUGUCAACUAtt	134.61	2.65	82.06	-0.64
	CAGCGAAACAGAUUAUGGAtt	126.76	2.25	84.39	-0.46
	CGUCUCAUGGUGUAAGAAAtt	73.11	-0.44	72.38	-1.39
INTS5	GGAGGGAUUUGGUCAGUUUtt	137.68	2.8	96.29	0.46
	GGUGCAUGCAGGGACAUUAtt	113.5	1.58	92.51	0.16
	GUACAUUAAUGGACAUCUAtt	91.86	0.5	87.47	-0.22
INTS6	UGCUGGUCACUUUCGAAGAtt	97.63	0.79	106.46	1.24
	GGCUGAAGGACUUACGACUtt	94.78	0.64	105.68	1.18
	GGACAGCUUUUGAUUUAUUtt	85.27	0.17	93.2	0.22
INTS7	GGAUGGACUGUAUACCGUAtt	180.96	4.98	82.14	-0.63

	CCAGGACUCUUCCGCAAAAtt	110.06	1.41	109.63	1.49
	CCAGUAGACUUGAAGCUAAtt	88.21	0.31	89.3	0
INTS8	CCAUAGAUGAGAAGCGGUUtt	188.15	5.34	93.47	0.24
	CCACUCGUUUUGACUAUUAtt	136.39	2.74	91.32	0.075
	GACCAGGUAAUAAAACGAAtt	95.02	0.66	89.19	-0.08
INTS9	GGAUCAUCCAGUCUCAUUAtt	115.21	1.67	95.94	0.43
	CAUGCAAUGUGCUCAAAUUtt	106.67	1.24	68.11	-1.72
	GCAAAGUCCUGAAGCCUUUtt	84.71	0.14	76.16	-1.09
INTS10	GGAGGAACCCUCGAAAGUAtt	100.32	0.92	94.1	0.29
	GGAGAUAUUUUGCAUAGAAtt	96.92	0.75	85.37	-0.38
	CAACCAUGAUGUUCGAUUAtt	94.63	0.64	114.27	1.85
INTS12	GUAGCAAGGAUUUACCUAUtt	116.43	1.73	98.09	0.59
	CCUCUAACCUUGGGUAAAAtt	112.69	1.54	98.82	0.66
	GCUCAAUGCUAUGAAGCGAtt	42.69	-1.97	78.59	-0.91

Chapter 5: knockdown of ornithine decarboxylase antizyme1 causes increased polyamine accumulation and improved luciferase translation in HEK293 cells

Abbreviations used:

ODC, ornithine decarboxylase; OAZ, ornithine decarboxylase antizyme; AZI, antizyme inhibitor; HEK, human embryonic kidney; CMV, cytomegalusvirus; DMEM, Dulbecco's modified Eagle's medium; FBS, fetal bovine serum; PBS, phosphate-buffered saline; TCA, trichloroacetic acid; EDTA, ethylenediaminetetraacetic acid; DMFO, difluoromethylornithine.

5.1 - Summary

Polyamines are essential molecules intimately involved in the regulation of cell proliferation, transformation and tumorigenesis and their homeostasis is tightly maintained at multiple levels. Ornithine decarboxylase (ODC) is the rate-limiting enzyme in polyamine biosynthesis and is probably the most highly regulated protein in mammalian cells. *Oaz1* gene encodes ODC antizyme, a major negative regulator of ODC and polyamines. In this chapter, we investigated the mechanisms why luciferase expression is enhanced in antizyme depleted cells. By comparing ODC, antizyme and cellular polyamines concentrations between *oaz1*-knockdown and negative control cells, it is found that when antizyme was depleted, ODC enzyme and cellular polyamines levels were up-regulated, leading to enhanced luciferase translation.

5.2 - Introduction

Polyamines (putrescine, spermidine, and spermine) are essential polycations that affect many biochemical processes [159, 160]. They interact with negatively charged molecules such as DNA, RNA, protein and phospholipids [160, 161], exerting a wide range of effects on replication, transcription, translation and post-translational modification processes as well as on membrane stability[161]. Even though the exact physiological functions and the precise mechanisms of polyamines in mammalian cells have remained unclear, it has been shown that polyamines are largely involved in regulation of cellular proliferation, transformation, differentiation, apoptosis, and tumorigenesis [162, 163]. Elevated polyamine accumulation has been associated with cell proliferation and transformation while decreased polyamines level inhibits cellular growth, migration, and embryonic development [164]. As dysregulation of cellular polyamine is associated with various cancer pathology, polyamine pathways have been explored intensively as targets for cancer chemotherapy and chemoprevention[163, 165].

Ornithine decarboxylase (ODC), spermidine synthase and spermine synthase are three major enzymes in the polyamine biosynthesis pathway and ODC is the first and rate-limiting enzyme (Fig.5.S1). Firstly, ornithine is decarboxylated by ODC to form putrescine. Then, catalyzed by spermidine synthase, putrescine receives the amino propylic group from decarboxylated s-adenosylmethionine (dcSAM) to form spermidine. Spermine is formed by obtaining one more amino propylic group from dcSAM (Fig. 5.S1).

Cellular polyamines are highly regulated by antizyme (OAZ) and antizyme inhibitor (AZI)-mediated regulation of ODC (Fig.5.S2). Antizymes are naturally occurring negative regulators of ODC[166]. They can inhibit ODC activity by forming heterodimers with ODC [167] and can lead to fast ubiquitination –independent ODC degradation by 26S proteasome[168]. ODC has a half-life of one to two hours in the absence of antizyme and the presence of antizyme reduces the half-life to minutes [169]. Antizyme levels are very low in most tissues [170], but they are rapidly increased in response to elevating polyamine levels via a mechanism that is not completely understood[171]. Antizyme inhibitors can inactivate antizyme by binding to antizymes with higher affinity than ODC, thus allowing release of ODC from antizyme inhibition. [172].

In humans, antizymes comprise of a family of at least 3 members[173]. Antizyme 1 is distributed in all tissues. It promotes ODC degradation via the 26S proteasome [168], inhibits polyamine uptake into the cell and stimulates polyamine export [174, 175]. Mammalian antizyme 2 is also distributed in all tissues but is less abundantly expressed [176]. It lacks ODC degradation capability but can inhibit polyamine uptake [177]. Antizyme 3 is only expressed during spermatogenesis in testis tissues [178]. Antizyme 1 is most intensively studied due to its effective regulation on cellular polyamine levels and it is directly referred to as "antizyme". It has been shown that antizyme mRNA level does not change with cellular polyamine concentration[179]; but polyamine level influences antizyme expression on translational level. Elevated polyamine concentration induces a +1 ribosomal frameshift, which is needed for functional antizyme expression [180]. As a result, functional antizymes capable of binding ODC are rapidly produced [171].

In chapter 4, we have identified *oaz1* gene as a top candidate to knock-down for enhanced luciferase expression. As *oaz1* gene encodes ODC antizyme 1, it is speculated that the oaz1 gene knock down may lead to elevated cellular ODC and polyamine concentrations, resulting in improved protein production. In this chapter, ODC, antizyme and cellular polyamines concentrations are compared between *oaz1*-knockdown and negative control cells. The results showed that in antizyme depleted cells, ODC enzyme and cellular polyamines levels were up-regulated, leading to enhanced luciferase translation.

5.3 - Materials and Methods

5.3.1. Cell culture and transfection

Silencer siRNA for *oaz1* gene (Catalog number: AM51331, assay ID: 46078) and Silencer Select Negative Control #2 were purchased from Life Technologies. AllStars Hs Cell Death Control was purchased from Qiagen. HEK-CMV-Luc2-Hygro cells constitutively expressing *P. pyralis* luciferase (Progema) were grown in DMEM supplemented with 10% FBS in a humidified incubator set at 37ºC and 5% CO_2.

The transfection was done in 6-well plate format: 0.12 nmol of each siRNA was added to each well (Corning) and 1.5mL of serum-free DMEM containing 11.25 μL of Lipofectamine RNAiMax (Life Technologies) was then added to each well. This lipid-siRNA mixture was incubated at ambient temperature for 30 min prior to adding 2×10^5 cells in 1.5mL of DMEM containing 20% FBS (Gibco). The transfected cells were

incubatied at 37ºC in 5% CO_2 and were harvested at 24hours, 48hours, 72 hours and 96 hours. Cells transfected with Silencer Select Negative Control siRNA were used for data normalization and cells transfected with AllStars Hs Cell Death Control siRNA was used as transfection efficiency control.

5.3.2. Isolation of RNA and real-time qRT-PCR

Cells were trypsinized from 6-well plates, washed with PBS twice and cell pellets were flash frozen on dry ice and then stored at -80˚C until extraction. RNA was extracted from the HEK-CMV-Luc2-Hygro cell pellet using the RNeasy kit (Qiagen) and then treated with DNase using TURBO DNA-free™ Kit (life technologies). CDNA was generated from the RNA using the Maxima Frist Strand cDNA Synthesis Kit for qRT-PCR (Thermo Scientific). The *real-time qPCR* was done using *Fast SYBR® Green* Master Mix (*life technologies*) in 7900 HT Fast Real Time PCR System (*Applied Biosystems*). The $2^{-\Delta\Delta Ct}$ method was used for relative expression analysis[181] with *gapdh* as reference gene. Cells transfected with negative control siRNA and harvested at 24hr was set as calibrator. Primer used for each gene: *luc* (Promega), 5'-TCACGAAGGTGTACATGCTTTGG-3' and 5'-GATCCTCAACGTGCAAAAGAAGC-3'; *odc1*, 5'-TAAAGGAACAGACGGGCTCT-3' and 5'-CCATAGACGCCATCATTCAC-3'; *oaz1*: 5'- GGAACCGTAGACTCGCTCAT-3' and 5'-TCGGAGTGAGCGTTTATTTG-3'; *gapdh*: 5'- CATCAATGGAAATCCCATCA-3' and 5'- TTCTCCATGGTGGTGAAGAC-3'.

5.3.3. Luciferase concentration and viable cell number measurement

Overall luciferase activity was determined by ONE-Glo™ Reagent (Promega) and the number of viable cells was determined with CellTiter-Glo™ Reagent (Promega). Upon addition of assay reagent, all plates were incubated at room temperature for 20 minutes to stabilize luminescent signal before reading with SpectraMax i3 plate reader (Molecular Devices). Cells tranfected with Silencer Select Negative Control siRNA was used for normalization. Per cell luciferase production was calculated from overall luciferase activity and viable cell number.

5.3.4. Western Blotting

Transfected cells were lysed in buffer containing 50 mm Tris-HCl, pH 7.4, 5 mM EDTA, 150 mM NaCl, 1% Nonidet P-40, and protease inhibitor mixture. Proteins (~20 μg) were separated by SDS-PAGE (4–12% gel) in MES buffer and transferred to 0.2-μm nitrocellulose membrane for immunodetection using mouse anti-ODC (Sigma, catalog number O1136) and mouse anti-β-actin (BD biosciences, catalog number 612657) primary antibodies and HRP conjugated anti-mouse secondary antibodies (abCAM, catalog number ab20043). Signals were detected with an ECL Plus chemiluminescence reagent.

5.3.5. Cellular polyamine concentration measurement

Cells in six-well plates were washed with PBS twice, harvested, and precipitated with 0.1 mL cold 10% (vol/vol) trichloroacetic acid (TCA). A total of 50μL of the TCA supernatant was used for polyamine analysis by ion exchange chromatographic system as described [182]. TCA precipitates were used for protein determination as above.

5.4 - Results

5.4.1. Effective knocking-down of *oaz1* gene with siRNA

Of the three mammalian antizyme genes, *oaz1, oaz2,* and *oaz3* encode three antizyme isoforms. We tested several siRNAs designed against three oaz genes (Supplemental Table 5.S1). Five of six *oaz1* siRNA significantly enhanced luciferase expression. In contrast, none of the siRNAs against *oaz2* or *oaz3* caused an increase in luciferase expression, probably due to the fact that antizyme 2 is a minor form and that antizyme 3 is expressed only in testis.

Among the six siRNAs targeting *oaz1* gene, one was chosen for further studies (siOAZ1: GCUAACUUAUUCUACUCCGtt), as it caused significant increase in luciferase without reducing cell viability. In order to examine the efficacies of antizyme 1 knockdown using siOAZ1, qRT-PCR was executed to determine *oaz1* mRNA level (Fig. 5.1). In the first 72 hours, the relative expression of *oaz1* mRNA was less than 3% upon siRNA transfection compare to negative control siRNA transfected cells, confirming good silencing effect of the siRNA. The 96 hour sample showed 7% relative expression level.

5.4.2. The improved luciferase expression is not due to improved transcription of *luc* gene

To investigate if the enhanced luciferase expression is resulted from transcription or translation step, the mRNA level of *luc* gene was also determined. While cells transfected with siOAZ1 had significantly higher production of luciferase (Fig. 5.2A),

their *luc* mRNA levels remained consistently comparable to negative control cells throughout the 96 hour period (Fig. 5.2B). In addition, there is significant cell growth improvement in the first 48 hours upon siOAZ1 transfection (Fig. 5.2A).

5.4.3. Ornithine decarboxylase is over-expressed in *oaz1*-depleted cells

As antizyme 1 inhibit ornithine decarboxylase (ODC) and render it for degradation, ODC protein and mRNA levels in siOAZ- and siN.C.- transfected cells were also determined. Western blot showed limited ODC enzyme level in un-transfected and siN.C. transfected cells. Upon silencing of *oaz1* gene, ODC level was significantly elevated, from 48 to 96 hours post siOAZ1transfection (Fig. 5.3A). However, the elevated ODC protein level is not due to enhanced *odc1* gene transcription. To the contrary, qRT-PCR demonstrated consistent reduction of *odc1* mRNA levels after silencing of *oaz1* gene (Fig. 5.3B).

5.4.4. Cellular putrescine concentration is highly up-regulated in *oaz1*-depleted cells

To investigate if the knocking-down of *oaz1* gene influence cellular polyamine concentration, the levels of three major polyamines, putriscine, spermidine and spermine were determined in siOAZ1- and siN.C- transfected cells. As shown in Fig.5.4, polyamines concentrations were normalized with total protein and presented as nmol/mg total protein. In *oaz1*-depleted cells, putriscine concentration is 10 fold as high as that in negative control cells. Spermidine concentration was also up-regulated, but to a lesser degree of up to one-fold increase in the first 72 hours. Spermine, the last product down the polyamine pathway, was reduced upon *oaz1* depletion.

5.4.5. Exogenous addition of polyamine into culture media improves luciferase expression

To further verify whether elevated polyamine concentration can lead to improved luciferase expression, different titers of putriscine, spermidine and spermine were independently added into cell culture media and their impacts on luciferase expression level and viable cell number were determined. With putrescine addition, up to 40% increase of luciferase expression was observed (at 100μM) and the cell growth was enhanced up to 10% (at 50 μM). Higher concentrations could not generate further improvements (Fig.5.5.A). For spermidine addition, the best effects were reached at a lower titer: a 36% increase in luciferase expression was obtained at 20 μM spermidine while a 24% increase in cell growth was achieved with 10 μM spermidine. In addition, reduction in both luciferase expression and viable cell number were observed with >100 μM spermidine (Fig.5.5.B). The addition of spermine is not beneficial when the titer is >20 μM. At a lower concentration (10 μM), a 16% increase in luciferase expression was observed (Fig. 5.5.C).

5.5 - Discussion

In chapter 4, we identified *oaz1* gene encoding ornithine decarboxylase antizyme 1 to be a promising target for detailed investigation, as the siRNA-mediated knocking-down of *oaz1* gene can consistently improve expression of multiple model proteins without significantly reducing cell viability. This provides highlights on the importance of cellular polyamines in protein synthesis and cell growth.

Polyamines are essential molecules, intimately involved in the regulation of cell proliferation, transformation and tumorigenesis [183]. Dysregulation of polyamine metabolism has been implicated in the pathogenicity of several human diseases, including cancer [183]. As polycations, they specifically bind to nucleic acids, proteins and phospholipids and regulate the stability, synthesis and activities of the macromolecules [150, 161, 184]. The most critical cellular function of polyamines appears to be the promotion of translation, because depletion of cellular spermidine and spermine by overexpression of a key polyamine catabolic enzyme, spermidine/spermine N1-acetyltransferase 1 (SSAT1) leads to a total suppression of protein synthesis without inhibition of synthesis of DNA and RNA [185]. The current data also suggest that, in a reciprocal situation, increased cellular polyamines can enhance reporter protein synthesis without increasing their transcription. In spite of abundant evidence for the critical function of polyamines, ODC and antizyme in cellular transformation and proliferation [186-188], their mechanism of action and the sequence of events are poorly understood at the molecular level.

In this chapter, we firstly examined the efficacies of antizyme 1 knockdown using siOAZ1. It is confirmed from real time qPCR results that this siRNA effectively down-regulated *oaz1* mRNA by 97% and the depletion extended to 96 hours post transfection (Fig. 5.2A). Upon antizyme depletion, luciferase expression was significantly improved from 48hrs to 96hrs but luciferase mRNA levels were slightly reduced compared to cells transfected with non-complementary siRNA (siN.C.). This finding indicates that the enhanced expression of luciferase occurred without an increase in transcription of

luciferase gene and that antizyme depletion enhanced luciferase expression at the level of translation.

As expected from the known function of antizyme in down-regulation of ODC by inducing its proteosomal degradation, ODC enzyme level was found increased in antizyme-depleted cells (Fig.5.3A). As a highly regulated protein with an extremely short half-life, ODC level fluctuates depending on cellular proliferative status, cellular polyamine level and external stimulus. The significantly higher ODC level in *oaz1* knockdown cells indicates the stabilization of ODC protein upon depletion of antizyme 1. Interestingly, ODC mRNA levels were significantly reduced upon siOAZ1 tranfection in a time-dependent manner. Judging from the nucleotide sequence of ODC mRNA, siOAZ1 is not predicted to bind to ODC mRNA to cause its degradation. Thus a decrease in ODC mRNA may be an indirect effect of *oaz1* knockdown.

Consistent with the ODC enzyme level, cellular putrescine level was sharply increased by 24 hrs of transfection with siOAZ1. Spermidine level also increased, but spermine level was lower than in the control siRNA-transfected cells. The sum of three polyamines was higher in OAZ1 knockdown cells, mainly due to a large increase in putrescine.

It has been reported that addition of polyamines in serum free medium increased production of recombinant proteins in mammalian cells [189] and polyamines can be included as components of commercial serum-free medium used for recombinant protein production. Our data is in accordance with these findings. It was found that exogenous addition of 50-100µM putrescine led up to 40% increase of luciferase expression and up

to 10% increase of cell growth. The best effects with spermidine and spermine were observed at a much lower titer, 20 μM and 10 μM respectively. Higher concentration led to significant proliferation inhibition, probably due to ruminant plasma amine oxidase which generates toxic acrolein[190].

Taken together, our studies suggest that in *oaz1* knocking-down cells, ODC enzyme and cellular polyamines levels were found to be up-regulated, leading to enhanced translation of luciferase and other reporter proteins. This is the first report to our knowledge to have improved recombinant protein production by engineering ODC enzyme or its antizyme.

ODC and antizymes were mostly investigated for their impact on cell proliferation and transformation, especially in cancer research. For example, overexpression of ODC caused enhancement of proliferation and transformation of immortalized mouse fibroblasts NIH3T3 cells [191, 192]. In transgenic mice overexpressing ODC under a keratin promoter, tumor incidence increased after a variety of stimuli, including chemical carcinogens, UV radiation, and an activated Ras, whereas treatment with ODC inhibitor difluoromethylornithine (DFMO) prevented tumor development [193, 194] . In contrast to ODC, lots of research confirmed the anti-proliferative effect of antizyme 1 and suggest a role for antizyme as a potential tumor suppressor[195]. For example, in several human oral cancer cell lines, the expression level of the *oaz1* gene was down-regulated [196, 197]. Overexpression of antizyme in different cell culture models abolishes ODC activity, suppresses cellular levels of polyamines, inhibited the cell proliferation rate, induced G0/G1 arrest and led to

119

apoptosis [164, 198-201]. *In vivo*, overexpression of an *oaz1* mutant clone under a keratin promoter reduced tumor formation in mouse skin, suggesting its tumor suppressive effects [202].

In addition to inhibition of ODC, antizymes also inhibit transport/uptake of polyamines [174, 203]. Since ODC and antizyme have opposite effects on polyamine levels, they would have opposite effects on reporter gene expression. Indeed, ODC siRNA reduced expression of luciferase (Supplemental Table 5.S2). In this regard, it would be interesting to determine whether overexpression of ODC would stimulate reporter protein production as antizyme knockdown does. This will provide interesting insights on the translational regulation of gene expression by polyamines in mammalian cells.

Contributions from collaborators:

Dr. Myung Hee Park contributed to drafting of results and discussion part and Dr. Swati Mandal carried out western blot for ODC, cellular polyamine concentration measurement and corresponding data analysis.

Figures and tables

**Fig. 5.S1.** Biosynthetic pathway of polyamines (putrescine, spermidine and spermine). Ornithine decarboxylase is the first and rate-limiting enzyme converting ornithine into putrescine.
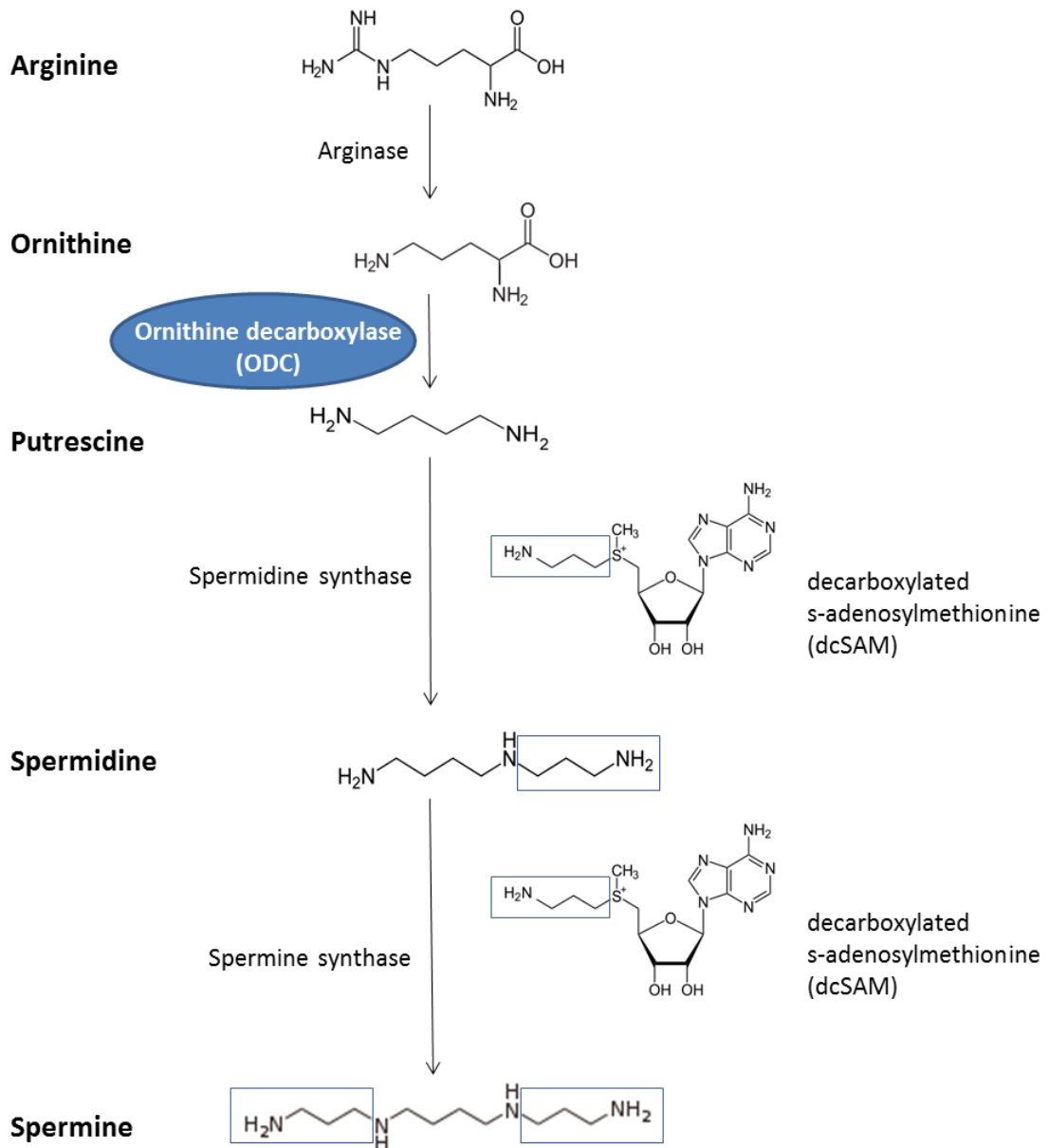
Fig.5.S2. Schematic diagram showing antizyme (OAZ) and antizyme inhibitor (AZI)-mediated regulation of ornithine decarboxylase (ODC). Figure adapted from reference [195], under license number 3603681212301.
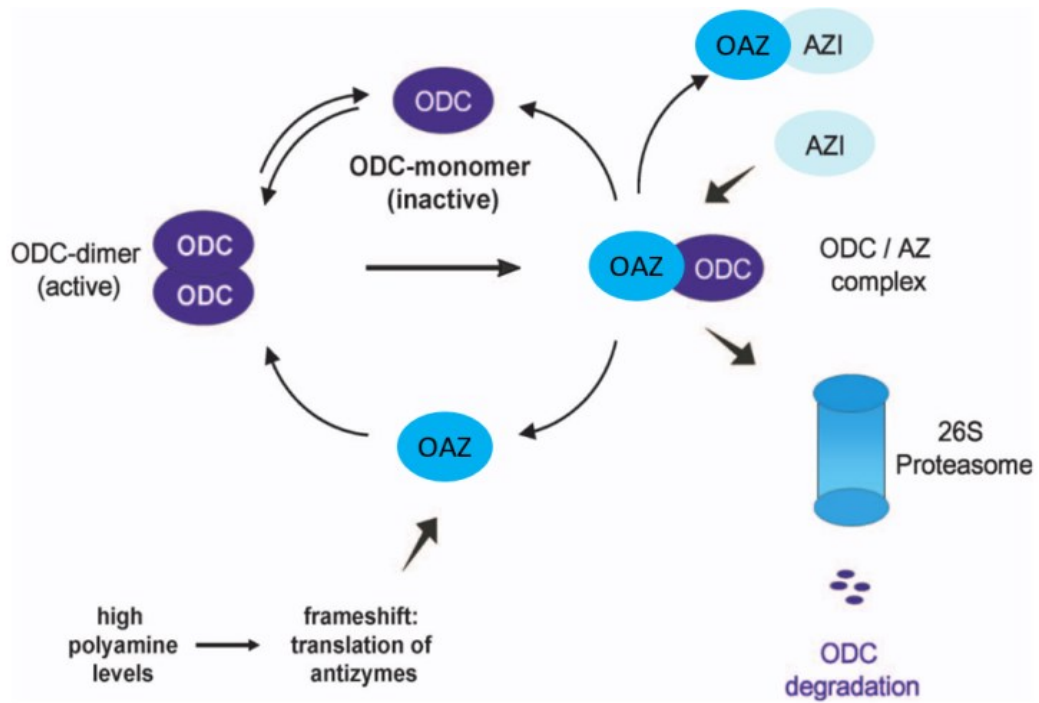
Fig. 5.1. Relative expression of *oaz1* gene in cells transfected with siRNA targeting *oaz1* (siOAZ1) and in cells transfected with negative control siRNA (siN.C.). The relative changes in genes expression were compared at 24, 48, 72 and 96 hours. Transfection was done with two different passages of cells and each biological sample was measured in triplicates. Error bars represent SEM.
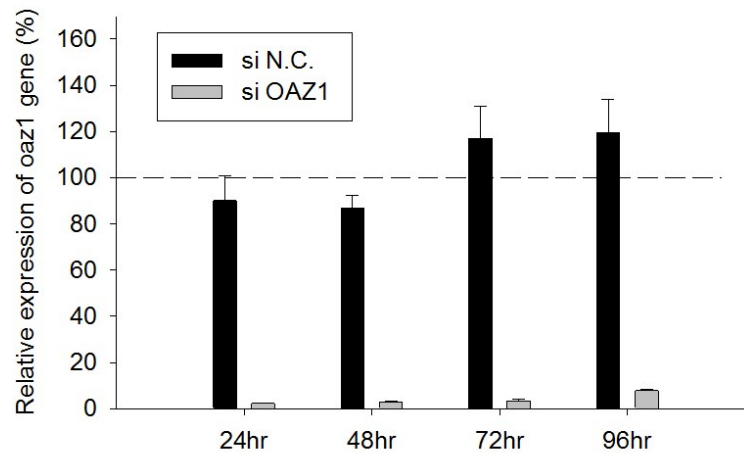
**Fig. 5.2.** Luciferase protein expression was enhanced (A) but not the relative transcription of *luc* gene (B) in antizyme depleted cells. The relative changes in luciferase protein/gene expression and viable cell number were compared to cells transfected with negative control siRNA (siN.C.). Transfection was done with two different passages of cells and each biological sample was measured in triplicates. Error bars represent SEM.
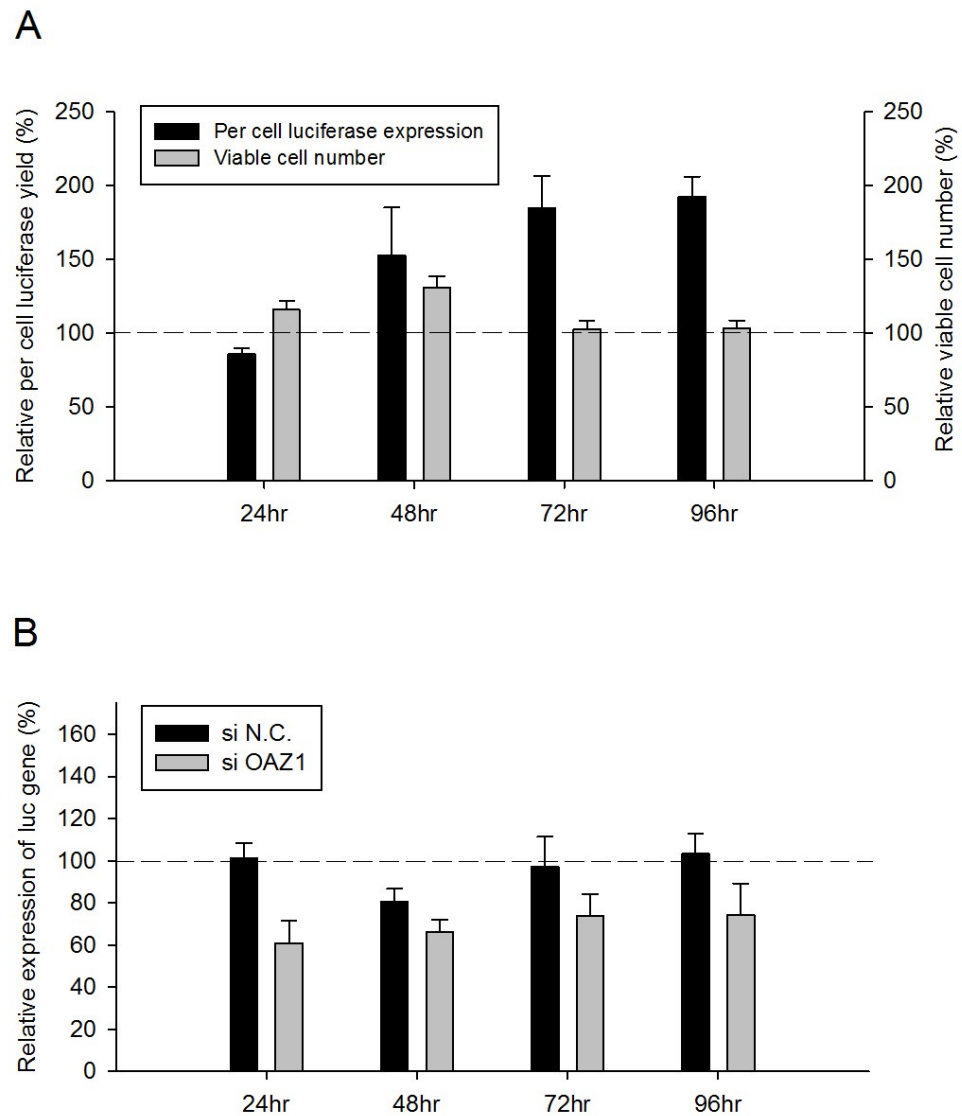
Fig. 5.3. Elevated ODC enzyme concentration (A) and reduced *odc* mRNA level (B) were detected in cells transfected with siOAZ1. Transfection was done with two different passages of cells and each biological sample was measured in triplicates. Error bars represent SEM.

A

B
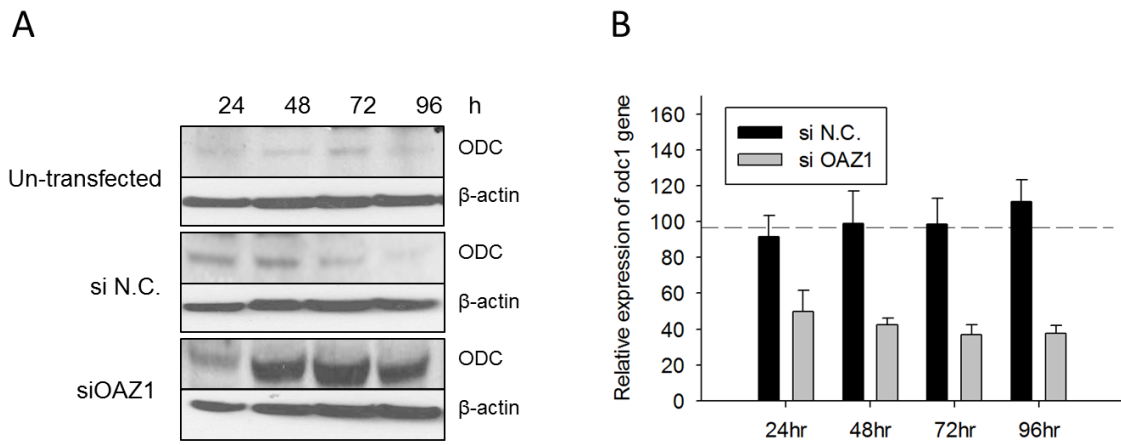
Fig. 5.4. Cellular (A) putrescine, (B) spermidine and (C) spermine concentration in *oaz1* depleted and negative control cells. Polyamines concentrations were normalized with total protein and presented as nmol/mg total protein.

A   Putrescine

B   Spermidine

C   Spermine

Fig. 5.5. The effect of exogenous addition of polyamines on luciferase expression and cell growth. Polyamine addition was done with two different passages of cells and each biological sample was measured in triplicates. Error bars represent SEM.

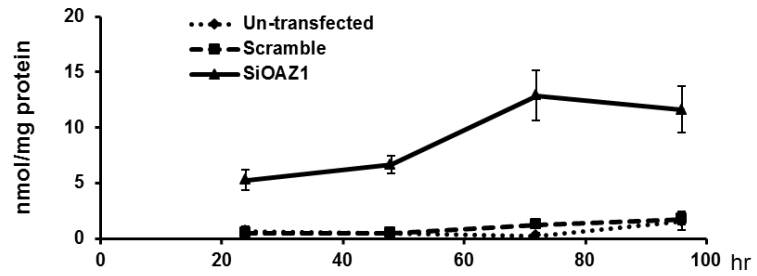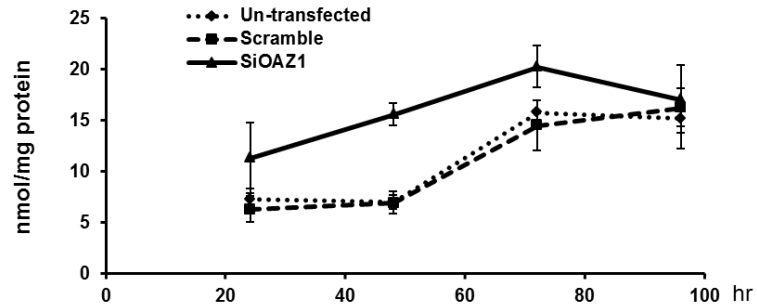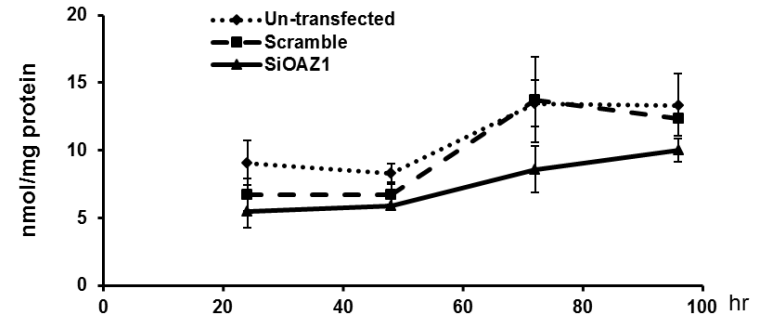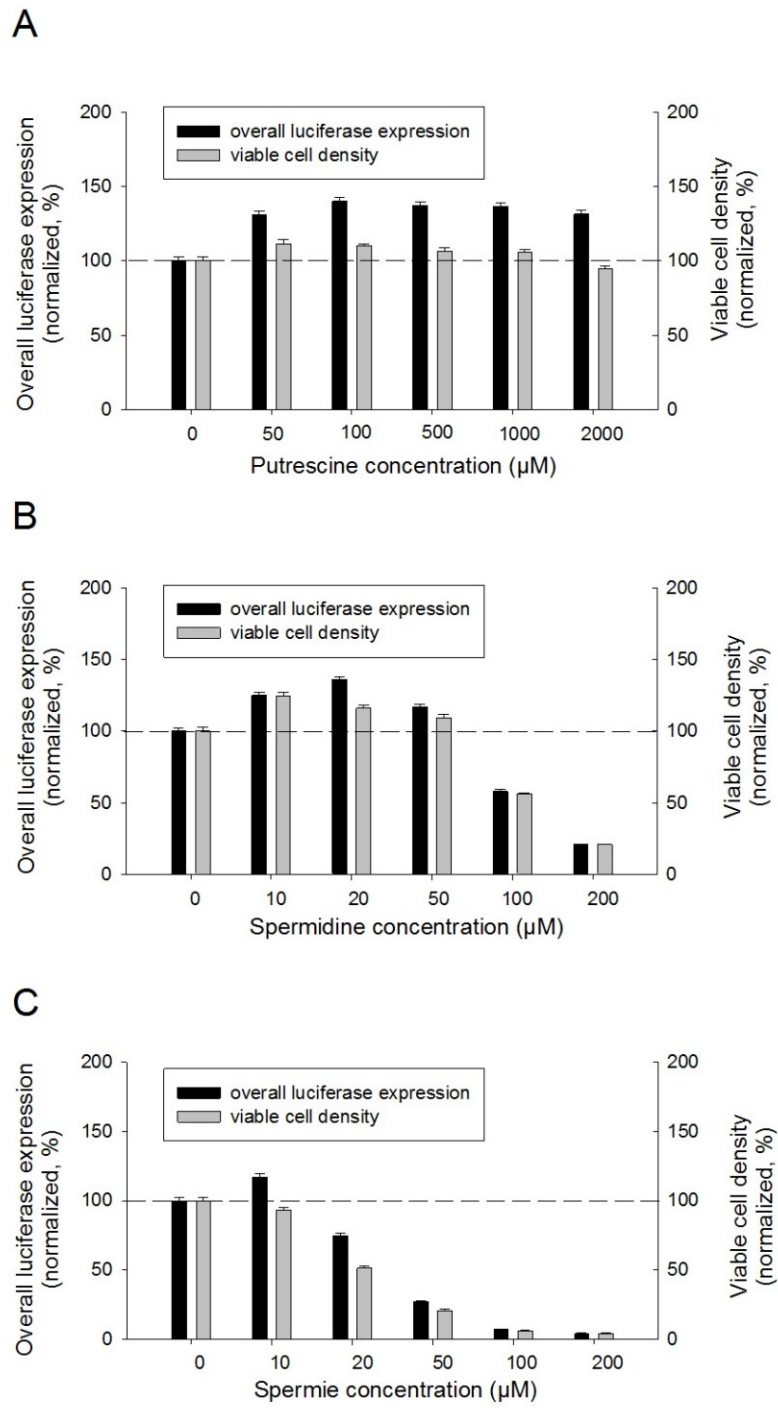**Table.5.S1.** The effects of antizyme genes knocking-down with different siRNAs.

gene	siRNA sequence	Luciferase activity (%)*	Viable cell density (%)*
oaz1	GCCUUGCUCCGAACCUUCAtt	161.2	94.8
oaz1	GAUUAUCCUUGUACUUUGAtt	144.5	101.9
oaz1	GGCUGAAUGUAACAGAGGAtt	127.7	95.0
oaz1	CCGUAGACUCGCUCAUCUCtt	174.5	85.4
oaz1	GCUAACUUAUUCUACUCCGtt	171.1	110.6
oaz1	GGGAAUAGUCAGAGGGAUCtt	92.8	102.8
oaz2	ACAUCGUCCACUUCCAGUAtt	97.4	96.3
oaz2	GGACCUCCCUGUGAAUGAUtt	95.4	86.0
oaz2	CAGAUGGAUUAUUAGCUGAtt	94.9	105.4
oaz3	CCGGGAAAGUUUGACUGCAtt	101.6	75.8
oaz3	CCACGACCAGCUUAAAGAAtt	90.5	95.8
oaz3	GACUUUCACUUCCGCCUUAtt	74.3	87.7

*values are normalized with cells transfected with negative control siRNA (siN.C.) . The

siOAZ1 selected for further investigation was highlighted in box.

Table.5.S2. The effects of *odc1* gene knocking-down with different siRNAs.

siRNA sequence	Luciferase activity (%)*	Viable cell density (%)*
GCUUGCAGUUAAUAUCAUUtt	28.4	60.8
GCAUGUAUCUGCUUGAUAUtt	20.0	50.7
GAUGACUUUUGAUAGUGAAtt	18.0	56.1

*values are normalized with cells transfected with negative control siRNA (siN.C.) .

Chapter 6: Conclusion and future work

6.1 – Final Remarks

The goal of this study is to develop strategies to improve the functional expression of mammalian membrane proteins. Neurotensin receptor type I (NTSR1) was utilized as a model protein, as it's a hard-to-express G protein-coupled receptor (GPCR) and its structure was not solved. We proposed that by engineering of the host organism and production process, we would be able to produce adequate amounts of purified receptors, to aid the crystallography studies of NTSR1. In addition, the successful strategies can be transferred to improving the production of other difficult proteins.

This dissertation demonstrated three effective strategies, including the adoption of mammalian inducible expression system, production process development and high-throughput RNA interference screening.

The utilization of human embryonic kidney (HEK) cell line harnessed the near-native environment of mammalian cells for protein translation, modification, folding and trafficking, thus contributed to the ideal quality of the membrane proteins produced. The adoption of the tetracycline inducible system allowed the external switching-on of protein production process, making it possible to manipulate temporally the event of cell growth and toxic protein production. This first strategy remarkably improved NTSR1 production comparing to constitutive expression approach.

The process development strategy in this dissertation included suspension culture adaptation for high-density cell culture and induction parameter optimization. The nature

130

of this strategy is fine-tuning of the production procedures to push the NTSR1 production to the extent of host organism's limit.

Once the host organism's capacity has been reached, further cell line engineering will be necessary to expand the host organism's limit. This dissertation focused on a bottom-up high throughput screening approach, to prove that this strategy can efficiently enhance membrane protein production and can lead to identification of limiting factors that need to be reserve-engineered.

Indeed, the polyamine studies in chapter 5 served as a proof of concept that the genome-scale loss-of-function data is valuable for understanding the protein expression process and can lead to exciting findings and applications. For example, even though the polyamine biosynthesis pathway and its regulation have been studied for decades, there were no reports on engineering any factors in this pathway for the purpose of enhancing protein production. The screen result shed light on this pathway, leading to targeted investigation followed by fast and efficient application of pre-existing biological research results.

6.2 – Future work

Continuation and expansion of the work detailed in this dissertation is currently underway. Firstly, five miRNAs (hsa-miR-22-5p, hsa-miR-18a-5p, hsa-miR-22-3p, hsa-miR-429 and hsa-miR-2110) have been identified in chapter 3 and the identification of target genes of these miRNAs are in progress. Combined with our whole-genome screen

data, it will contribute to our understanding of the mechanism behind improvement of protein expression. Secondly, a HEK293 cell line stably over-expresing hsa-miR-22 is being constructed and tested. This experiment will allow us to investigate stable inhibition and combinatorial effects of miRNAs (hsa-miR-22 is composed of hsa-miR-22-5p and hsa-miR-22-3p arms). Thirdly, CRISPR cell lines knocking out each of identified top 10 genes (*ints1, ints2, hnrnpc, casp8ap2, oaz1, ppp2r1a, prpf19, cct2, chaf1a, eef1b2*) will be constructed to investigate if the gene deletion can intensify the effects. This will also serve as a proof of concept for instustrial application of screen hits. Finally, as discussed in chapter 5, as ODC and antizyme has opposite effects in cellular polyamine concentration, it would be interesting to determine whether overexpression of ODC would stimulate reporter protein production as antizyme knockdown does. The comparison of the two system is especially interesting because antizyme not only inhibit ODC but also regulate polyamine trasport /updake.

In addition, there are many more interesting project candidates derived from this study.

First of all, as large-scale transient gene expression with HEK293 cells has been widely applied in contract manufacturing organizations (CMO) for fast preparation of secreted medical proteins, is it possible to establish a similar strategy for membrane protein production? Such expression strategy will greatly benefit crystallography society as it will generate correctly-folded proteins while keeping a fast turn-over rate.

Secondly, as all of the work in this dissertation is based on loss-of-function studies, our dataset is not complete. It will be very interesting to carry out gain-of-function studies with the same reporter protein and same human genome library [204].

With a two-sided story, we will be able to understand better the positive or negative involvment of different genes in protein biosynthesis and cell growth.

Thirdly, from production point of view, it will be more beneficial to execute genome-scale siRNA screen with Chinese Hamster Ovary cells (CHO). Since the hamster siRNA library is still lacking, murine library could be used instead. Or top hits from this HEK293 screen can be tested with CHO cells.

Additionally, the splicesome pathway has been high-lighted in chapter 4 and the clear involvment of integrator complex is very intriguing. It'll be worthwhile to look into luciferase mRNA level upon the transfection of these siRNAs, to investigate if the altered splicesome pathway changed luciferase transcription level. Specifically for integrator complex, it'll be interesting to check U1/U2 snRNA level.

Apart from genes involved in mRNA splicing process and polyamine pathway, there are many others worth further investigation(casp8ap2, ppp2r1a, cct2, chaf1a, eef1b2) for mechanistic studies. The knocking down of *ppp2r1a* for example, can enhance the expression of cytosolic, secreted and membrane proteins tested, and also introduced minimal cell growth disadvantage. As an important serine/threonine phosphatase with diverse cellular functions, protein phosphatase 2 (PP2) has been intensively studied for years. It will be interesting to deciper how this global regulator can be engineered for a host cell line with improved protein expression capability.

References

1.   Dimitrov, D.S., *Therapeutic proteins.* Methods Mol Biol, 2012. **899**: p. 1-26.
2.   Zhu, J., *Mammalian cell protein expression for biopharmaceutical production.* Biotechnol Adv, 2011. **30**(5): p. 1158-70.
3.   Huang, Y.M., et al., *Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment.* Biotechnol Prog, 2010. **26**(5): p. 1400-10.
4.   Francis, D.M. and R. Page, *Strategies to optimize protein expression in E. coli.* Curr Protoc Protein Sci, 2011. **Chapter 5**: p. Unit 5 24 1-29.
5.   Jarvis, D.L., *Baculovirus-insect cell expression systems.* Methods Enzymol, 2009. **463**: p. 191-222.
6.   Zhu, J., *Mammalian cell protein expression for biopharmaceutical production.* Biotechnol Adv, 2012. **30**(5): p. 1158-70.
7.   Andrell, J. and C.G. Tate, *Overexpression of membrane proteins in mammalian cells for structural studies.* Mol Membr Biol, 2013. **30**(1): p. 52-63.
8.   Bill, R.M., et al., *Overcoming barriers to membrane protein structure determination.* Nat Biotechnol, 2011. **29**(4): p. 335-40.
9.   Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms.* Protein Sci, 1998. **7**(4): p. 1029-38.
10.   Pierce, K.L., R.T. Premont, and R.J. Lefkowitz, *Seven-transmembrane receptors.* Nat Rev Mol Cell Biol, 2002. **3**(9): p. 639-50.
11.   Tate, C.G., *Overexpression of mammalian integral membrane proteins for structural studies.* FEBS Lett, 2001. **504**(3): p. 94-8.
12.   Venkatakrishnan, A.J., et al., *Molecular signatures of G-protein-coupled receptors.* Nature, 2013. **494**(7436): p. 185-94.
13.   Quilici, L.S., et al., *A minimal cytomegalovirus intron A variant can improve transgene expression in different mammalian cell lines.* Biotechnol Lett, 2013. **35**(1): p. 21-7.
14.   Hou, J., et al., *High activity expression of D-amino acid oxidase in Escherichia coli by the protein expression rate optimization.* Protein Expr Purif, 2013. **88**(1): p. 120-6.
15.   Lin, C.H. and D.L. Jarvis, *Utility of temporally distinct baculovirus promoters for constitutive and baculovirus-inducible transgene expression in transformed insect cells.* J Biotechnol, 2013. **165**(1): p. 11-7.
16.   Nannenga, B.L. and F. Baneyx, *Enhanced expression of membrane proteins in E. coli with a P(BAD) promoter mutant: synergies with chaperone pathway engineering strategies.* Microb Cell Fact, 2011. **10**: p. 105.
17.   Wagner, S., et al., *Tuning Escherichia coli for membrane protein overexpression.* Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14371-6.
18.   Grandjean, M., et al., *High-level transgene expression by homologous recombination-mediated gene transfer.* Nucleic Acids Res, 2011. **39**(15): p. e104.
19.   Araki, Y., et al., *Efficient recombinant production in mammalian cells using a novel IR/MAR gene amplification method.* PLoS One, 2012. **7**(7): p. e41787.
20.   Yoshimura, H., et al., *High levels of human recombinant cyclooxygenase-1 expression in mammalian cells using a novel gene amplification method.* Protein Expr Purif, 2011. **80**(1): p. 41-6.

21. Gantke, T., et al., *Ebola virus VP35 induces high-level production of recombinant TPL-2-ABIN-2-NF-kappaB1 p105 complex in co-transfected HEK-293 cells.* Biochem J, 2013. **452**(2): p. 359-65.

22. Teng, C.Y., et al., *Enhanced protein secretion from insect cells by co-expression of the chaperone calreticulin and translation initiation factor eIF4E.* Mol Biotechnol, 2012. **54**(1): p. 68-78.

23. Hartl, F.U. and M. Hayer-Hartl, *Molecular chaperones in the cytosol: from nascent chain to folded protein.* Science, 2002. **295**(5561): p. 1852-8.

24. Folwarczna, J., et al., *Efficient expression of Human papillomavirus 16 E7 oncoprotein fused to C-terminus of Tobacco mosaic virus (TMV) coat protein using molecular chaperones in Escherichia coli.* Protein Expr Purif, 2012. **85**(1): p. 152-7.

25. Voulgaridou, G.P., et al., *Efficient E. coli expression strategies for production of soluble human crystallin ALDH3A1.* PLoS One, 2013. **8**(2): p. e56582.

26. Tate, C.G., E. Whiteley, and M.J. Betenbaugh, *Molecular chaperones stimulate the functional expression of the cocaine-sensitive serotonin transporter.* J Biol Chem, 1999. **274**(25): p. 17551-8.

27. Qian, W., et al., *Secretion of truncated recombinant rabies virus glycoprotein with preserved antigenic properties using a co-expression system in Hansenula polymorpha.* J Microbiol, 2013. **51**(2): p. 234-40.

28. Shen, Q., et al., *The effect of gene copy number and co-expression of chaperone on production of albumin fusion proteins in Pichia pastoris.* Appl Microbiol Biotechnol, 2012. **96**(3): p. 763-72.

29. Le Fourn, V., et al., *CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion.* Metab Eng, 2012.

30. Lakkaraju, A.K., et al., *SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites.* Cell, 2008. **133**(3): p. 440-51.

31. Nannenga, B.L. and F. Baneyx, *Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli.* Protein Sci, 2011. **20**(8): p. 1411-20.

32. Peng, R.W., E. Abellan, and M. Fussenegger, *Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells.* Biotechnol Bioeng, 2010. **108**(3): p. 611-20.

33. Hou, J., et al., *Engineering of vesicle trafficking improves heterologous protein secretion in Saccharomyces cerevisiae.* Metab Eng, 2012. **14**(2): p. 120-7.

34. Rahimpour, A., et al., *Engineering the Cellular Protein Secretory Pathway for Enhancement of Recombinant Tissue Plasminogen Activator Expression in Chinese Hamster Ovary Cells: Effects of CERT and XBP1s Genes.* J Microbiol Biotechnol, 2013. **23**(8): p. 1116-22.

35. Florin, L., et al., *Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells.* J Biotechnol, 2009. **141**(1-2): p. 84-90.

36. Nasoohi, N., et al., *Enhancement of catalysis and functional expression of a bacterial laccase by single amino acid replacement.* Int J Biol Macromol, 2013. **60**: p. 56-61.

37. Selvaraj, S.R., et al., *Bioengineering of coagulation factor VIII for efficient expression through elimination of a dispensable disulfide loop.* J Thromb Haemost, 2012. **10**(1): p. 107-15.

38. Sarkar, C.A., et al., *Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 14808-13.

39.   Heggeset, T.M., et al., *Combinatorial mutagenesis and selection of improved signal sequences and their application for high-level production of translocated heterologous proteins in Escherichia coli.* Appl Environ Microbiol, 2012. **79**(2): p. 559-68.

40.   Majors, B.S., et al., *Directed evolution of mammalian anti-apoptosis proteins by somatic hypermutation.* Protein Eng Des Sel, 2011. **25**(1): p. 27-38.

41.   Majors, B.S., M.J. Betenbaugh, and G.G. Chiang, *Links between metabolism and apoptosis in mammalian cells: applications for anti-apoptosis engineering.* Metab Eng, 2007. **9**(4): p. 317-26.

42.   Ohsfeldt, E., et al., *Increased expression of the integral membrane proteins EGFR and FGFR3 in anti-apoptotic Chinese hamster ovary cell lines.* Biotechnol Appl Biochem, 2012. **59**(3): p. 155-62.

43.   Macaraeg, N.F., D.E. Reilly, and A.W. Wong, *Use of an anti-apoptotic CHO cell line for transient gene expression.* Biotechnol Prog, 2013. **29**(4): p. 1050-8.

44.   Druz, A., et al., *Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells.* Metab Eng, 2013. **16**: p. 87-94.

45.   Ferndahl, C., et al., *Increasing cell biomass in Saccharomyces cerevisiae increases recombinant protein yield: the use of a respiratory strain as a microbial cell factory.* Microb Cell Fact, 2010. **9**: p. 47.

46.   Dreesen, I.A. and M. Fussenegger, *Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells.* Biotechnol Bioeng, 2010. **108**(4): p. 853-66.

47.   Jadhav, V., et al., *A screening method to assess biological effects of microRNA overexpression in Chinese hamster ovary cells.* Biotechnol Bioeng, 2012. **109**(6): p. 1376-85.

48.   Jardon, M.A., et al., *Inhibition of glutamine-dependent autophagy increases t-PA production in CHO cell fed-batch processes.* Biotechnol Bioeng, 2012. **109**(5): p. 1228-38.

49.   Lamppa, J.W., S.A. Tanyos, and K.E. Griswold, *Engineering Escherichia coli for soluble expression and single step purification of active human lysozyme.* J Biotechnol, 2012. **164**(1): p. 1-8.

50.   Dostalova, Z., et al., *High-level expression and purification of Cys-loop ligand-gated ion channels in a tetracycline-inducible stable mammalian cell line: GABAA and serotonin receptors.* Protein Sci, 2010. **19**(9): p. 1728-38.

51.   Reeves, P.J., J.M. Kim, and H.G. Khorana, *Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13413-8.

52.   Standfuss, J., et al., *The structural basis of agonist-induced activation in constitutively active rhodopsin.* Nature, 2011. **471**(7340): p. 656-60.

53.   Deupi, X., et al., *Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II.* Proc Natl Acad Sci U S A, 2012. **109**(1): p. 119-24.

54.   Gruswitz, F., et al., *Function of human Rh based on structure of RhCG at 2.1 A.* Proc Natl Acad Sci U S A, 2010. **107**(21): p. 9638-43.

55.   Grisshammer, R., *Why we need many more G protein-coupled receptor structures.* Expert Rev Proteomics, 2013. **10**(1): p. 1-3.

56.   Carraway, R. and S.E. Leeman, *The isolation of a new hypotensive peptide, neurotensin, from bovine hypothalami.* J Biol Chem, 1973. **248**(19): p. 6854-61.

57.   Bissette, G., et al., *Hypothermia and intolerance to cold induced by intracisternal administration of the hypothalamic peptide neurotensin.* Nature, 1976. **262**(5569): p. 607-9.

58.   Carraway, R.E. and A.M. Plona, *Involvement of neurotensin in cancer growth: evidence, mechanisms and development of diagnostic tools.* Peptides, 2006. **27**(10): p. 2445-60.

59.   Griebel, G. and F. Holsboer, *Neuropeptide receptor ligands as drugs for psychiatric diseases: the end of the beginning?* Nat Rev Drug Discov, 2012. **11**(6): p. 462-78.

60.   Kitabgi, P., *Targeting neurotensin receptors with agonists and antagonists for therapeutic purposes.* Curr Opin Drug Discov Devel, 2002. **5**(5): p. 764-76.

61.   Schimpff, R.M., et al., *Increased plasma neurotensin concentrations in patients with Parkinson's disease.* J Neurol Neurosurg Psychiatry, 2001. **70**(6): p. 784-6.

62.   Chalon, P., et al., *Molecular cloning of a levocabastine-sensitive neurotensin binding site.* FEBS Lett, 1996. **386**(2-3): p. 91-4.

63.   Mazella, J., *Sortilin/neurotensin receptor-3: a new tool to investigate neurotensin signaling and cellular trafficking?* Cell Signal, 2001. **13**(1): p. 1-6.

64.   Tanaka, K., M. Masu, and S. Nakanishi, *Structure and functional expression of the cloned rat neurotensin receptor.* Neuron, 1990. **4**(6): p. 847-54.

65.   White, J.F., et al., *Automated large-scale purification of a G protein-coupled receptor for neurotensin.* FEBS Lett, 2004. **564**(3): p. 289-93.

66.   Shibata, Y., et al., *Thermostabilization of the neurotensin receptor NTS1.* J Mol Biol, 2009. **390**(2): p. 262-77.

67.   White, J.F., et al., *Structure of the agonist-bound neurotensin receptor.* Nature, 2012. **490**(7421): p. 508-13.

68.   Yao, F., et al., *Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells.* Hum Gene Ther, 1998. **9**(13): p. 1939-50.

69.   Hillen, W. and C. Berens, *Mechanisms underlying expression of Tn10 encoded tetracycline resistance.* Annu Rev Microbiol, 1994. **48**: p. 345-69.

70.   Postle, K., T.T. Nguyen, and K.P. Bertrand, *Nucleotide sequence of the repressor gene of the TN10 tetracycline resistance determinant.* Nucleic Acids Res, 1984. **12**(12): p. 4849-63.

71.   Guan, X.M., T.S. Kobilka, and B.K. Kobilka, *Enhancement of membrane insertion and function in a type IIIb membrane protein following introduction of a cleavable signal peptide.* J Biol Chem, 1992. **267**(31): p. 21995-8.

72.   Schaffner, W. and C. Weissmann, *A rapid, sensitive, and specific method for the determination of protein in dilute solution.* Anal Biochem, 1973. **56**(2): p. 502-14.

73.   White, J.F. and R. Grisshammer, *Stability of the neurotensin receptor NTS1 free in detergent solution and immobilized to affinity resin.* PLoS One, 2010. **5**(9): p. e12579.

74.   Shibata, Y., et al., *Optimising the combination of thermostabilising mutations in the neurotensin receptor for structure determination.* Biochim Biophys Acta, 2013. **1828**(4): p. 1293-301.

75.   Hedayat, A., N.J.A. Sloane, and J. Stufken, *Orthogonal arrays : theory and applications.* Springer series in statistics. 1999, New York: Springer. xxii, 416 p.

76.   Tate, C.G., et al., *Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter.* Biochim Biophys Acta, 2003. **1610**(1): p. 141-53.

77. Grisshammer, R., R. Duckworth, and R. Henderson, *Expression of a rat neurotensin receptor in Escherichia coli.* Biochem J, 1993. **295 (Pt 2)**: p. 571-6.

78. O'Malley, M.A., et al., *High-level expression in Saccharomyces cerevisiae enables isolation and spectroscopic characterization of functional human adenosine A2a receptor.* J Struct Biol, 2007. **159**(2): p. 166-78.

79. Kobilka, B.K., *Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor.* Anal Biochem, 1995. **231**(1): p. 269-71.

80. Andrell, J. and C.G. Tate, *Overexpression of membrane proteins in mammalian cells for structural studies.* Mol Membr Biol, 2012. **30**(1): p. 52-63.

81. Reeves, P.J., R.L. Thurmond, and H.G. Khorana, *Structure and function in rhodopsin: high level expression of a synthetic bovine opsin gene and its mutants in stable mammalian cell lines.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11487-92.

82. Cook, B.L., et al., *Study of a synthetic human olfactory receptor 17-4: expression and purification from an inducible mammalian cell line.* PLoS One, 2008. **3**(8): p. e2920.

83. Corin, K., et al., *Structure and function analyses of the purified GPCR human vomeronasal type 1 receptor 1.* Sci Rep, 2011. **1**: p. 172.

84. Kim, J.S., et al., *High-level scu-PA production by butyrate-treated serum-free culture of recombinant CHO cell line.* Biotechnol Prog, 2004. **20**(6): p. 1788-96.

85. Jadhav, V., et al., *CHO microRNA engineering is growing up: recent successes and future challenges.* Biotechnol Adv, 2013. **31**(8): p. 1501-13.

86. Jadhav, V., et al., *Stable overexpression of miR-17 enhances recombinant protein production of CHO cells.* J Biotechnol, 2014. **175**: p. 38-44.

87. Druz, A., et al., *Large-scale screening identifies a novel microRNA, miR-15a-3p, which induces apoptosis in human cancer cell lines.* RNA Biol, 2013. **10**(2): p. 287-300.

88. Filipowicz, W., S.N. Bhattacharyya, and N. Sonenberg, *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?* Nat Rev Genet, 2008. **9**(2): p. 102-14.

89. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.

90. Hobert, O., *Gene regulation by transcription factors and microRNAs.* Science, 2008. **319**(5871): p. 1785-6.

91. Hackl, M., N. Borth, and J. Grillari, *miRNAs--pathway engineering of CHO cell factories that avoids translational burdening.* Trends Biotechnol, 2012. **30**(8): p. 405-6.

92. Abdul-Hussein, S., J. Andrell, and C.G. Tate, *Thermostabilisation of the serotonin transporter in a cocaine-bound conformation.* J Mol Biol, 2013. **425**(12): p. 2198-207.

93. Chung, N., et al., *Median absolute deviation to improve hit selection for genome-scale RNAi screens.* J Biomol Screen, 2008. **13**(2): p. 149-58.

94. Feng, M., et al., *Therapeutically targeting glypican-3 via a conformation-specific single-domain antibody in hepatocellular carcinoma.* Proc Natl Acad Sci U S A, 2013. **110**(12): p. E1083-91.

95. Wagner, S., et al., *Rationalizing membrane protein overexpression.* Trends Biotechnol, 2006. **24**(8): p. 364-71.

96. Fischer, S., et al., *A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells.* Biotechnol J, 2014.

97. Strotbek, M., et al., *Stable microRNA expression enhances therapeutic antibody productivity of Chinese hamster ovary cells.* Metab Eng, 2013. **20**: p. 157-66.

98.   Drew, D., et al., *Optimization of membrane protein overexpression and purification using GFP fusions.* Nat Methods, 2006. **3**(4): p. 303-13.

99.   Newstead, S., et al., *High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 13936-41.

100.   Thomas, J.A. and C.G. Tate, *Quality Control in Eukaryotic Membrane Protein Overproduction.* J Mol Biol, 2014. **426**(24): p. 4139-4154.

101.   Palmer, E. and T. Freeman, *Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays.* Comp Funct Genomics, 2004. **5**(4): p. 342-53.

102.   Wiemann, S., et al., *Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs.* Genome Res, 2001. **11**(3): p. 422-35.

103.   Huh, W.K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**(6959): p. 686-91.

104.   Sato, K., M. Sato, and A. Nakano, *Rer1p, a retrieval receptor for endoplasmic reticulum membrane proteins, is dynamically localized to the Golgi apparatus by coatomer.* J Cell Biol, 2001. **152**(5): p. 935-44.

105.   Zhu, J.Y., et al., *Identification of novel Epstein-Barr virus microRNA genes from nasopharyngeal carcinomas.* J Virol, 2009. **83**(7): p. 3333-41.

106.   Lang, Y., et al., *MicroRNA-429 induces tumorigenesis of human non-small cell lung cancer cells and targets multiple tumor suppressor genes.* Biochem Biophys Res Commun, 2014. **450**(1): p. 154-9.

107.   Liu, X., et al., *Tumor-Suppressing Effects of miR-429 on Human Osteosarcoma.* Cell Biochem Biophys, 2014. **70**(1): p. 215-24.

108.   Song, S.J., et al., *The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation.* Cell Stem Cell, 2013. **13**(1): p. 87-101.

109.   Song, Y., et al., *MiR-18a regulates the proliferation, migration and invasion of human glioblastoma cell by targeting neogenin.* Exp Cell Res, 2014. **324**(1): p. 54-64.

110.   Takakura, S., et al., *Oncogenic role of miR-17-92 cluster in anaplastic thyroid cancer cells.* Cancer Sci, 2008. **99**(6): p. 1147-54.

111.   Xiong, J., Q. Du, and Z. Liang, *Tumor-suppressive microRNA-22 inhibits the transcription of E-box-containing c-Myc target genes by silencing c-Myc binding protein.* Oncogene, 2010. **29**(35): p. 4980-8.

112.   Barron, N., et al., *Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7.* J Biotechnol, 2011. **151**(2): p. 204-11.

113.   Napoli, C., C. Lemieux, and R. Jorgensen, *Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans.* Plant Cell, 1990. **2**(4): p. 279-289.

114.   van der Krol, A.R., et al., *Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression.* Plant Cell, 1990. **2**(4): p. 291-9.

115.   Aagaard, L. and J.J. Rossi, *RNAi therapeutics: principles, prospects and challenges.* Adv Drug Deliv Rev, 2007. **59**(2-3): p. 75-86.

116.   Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.

117. Elbashir, S.M., et al., *Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.* Nature, 2001. **411**(6836): p. 494-8.

118. Elbashir, S.M., W. Lendeckel, and T. Tuschl, *RNA interference is mediated by 21- and 22-nucleotide RNAs.* Genes Dev, 2001. **15**(2): p. 188-200.

119. Hamilton, A.J. and D.C. Baulcombe, *A species of small antisense RNA in posttranscriptional gene silencing in plants.* Science, 1999. **286**(5441): p. 950-2.

120. Jinek, M. and J.A. Doudna, *A three-dimensional view of the molecular machinery of RNA interference.* Nature, 2009. **457**(7228): p. 405-12.

121. Conrad, C. and D.W. Gerlich, *Automated microscopy for high-content RNAi screening.* J Cell Biol, 2010. **188**(4): p. 453-61.

122. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

123. Seyhan, A.A. and T.E. Rya, *RNAi screening for the discovery of novel modulators of human disease.* Curr Pharm Biotechnol, 2010. **11**(7): p. 735-56.

124. Bard, F., et al., *Functional genomics reveals genes involved in protein secretion and Golgi organization.* Nature, 2006. **439**(7076): p. 604-7.

125. Brognard, J. and T. Hunter, *Protein kinase signaling networks in cancer.* Curr Opin Genet Dev, 2011. **21**(1): p. 4-11.

126. Cherry, S., *Genomic RNAi screening in Drosophila S2 cells: what have we learned about host-pathogen interactions?* Curr Opin Microbiol, 2008. **11**(3): p. 262-70.

127. Ni, Z. and S.S. Lee, *RNAi screens to identify components of gene networks that modulate aging in Caenorhabditis elegans.* Brief Funct Genomics. **9**(1): p. 53-64.

128. Orvedahl, A., et al., *Image-based genome-wide siRNA screen identifies selective autophagy factors.* Nature, 2011. **480**(7375): p. 113-7.

129. Simpson, J.C., et al., *Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway.* Nat Cell Biol, 2012. **14**(7): p. 764-74.

130. Gould, S.J. and S. Subramani, *Firefly luciferase as a tool in molecular and cell biology.* Anal Biochem, 1988. **175**(1): p. 5-13.

131. Scholl, Z.N., W. Yang, and P.E. Marszalek, *Chaperones rescue luciferase folding by separating its domains.* J Biol Chem, 2014. **289**(41): p. 28607-18.

132. Svetlov, M.S., et al., *Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family.* Protein Sci, 2006. **15**(2): p. 242-7.

133. Konig, R., et al., *A probability-based approach for the analysis of large-scale RNAi screens.* Nat Methods, 2007. **4**(10): p. 847-9.

134. Will, C.L. and R. Luhrmann, *Spliceosome structure and function.* Cold Spring Harb Perspect Biol, 2011. **3**(7).

135. Baillat, D., et al., *Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II.* Cell, 2005. **123**(2): p. 265-76.

136. Chen, J. and E.J. Wagner, *snRNA 3' end formation: the dawn of the Integrator complex.* Biochem Soc Trans, 2010. **38**(4): p. 1082-7.

137. Tao, S., Y. Cai, and K. Sampath, *The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling.* Development, 2009. **136**(16): p. 2757-65.

138. Filleur, S., et al., *INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling.* Cancer Cell Int, 2009. **9**: p. 28.

139. Dominski, Z., et al., *A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100.* Mol Cell Biol, 2005. **25**(4): p. 1489-500.

140. Beyer, A.L., et al., *Identification and characterization of the packaging proteins of core 40S hnRNP particles.* Cell, 1977. **11**(1): p. 127-38.

141. Choi, Y.D., et al., *Heterogeneous nuclear ribonucleoproteins: role in RNA splicing.* Science, 1986. **231**(4745): p. 1534-9.

142. Rajagopalan, L.E., et al., *hnRNP C increases amyloid precursor protein (APP) production by stabilizing APP mRNA.* Nucleic Acids Res, 1998. **26**(14): p. 3418-23.

143. Konig, J., et al., *iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution.* Nat Struct Mol Biol, 2010. **17**(7): p. 909-15.

144. Ohi, M.D., et al., *Structural and functional analysis of essential pre-mRNA splicing factor Prp19p.* Mol Cell Biol, 2005. **25**(1): p. 451-60.

145. Beck, B.D., et al., *Human Pso4 is a metnase (SETMAR)-binding partner that regulates metnase function in DNA repair.* J Biol Chem, 2008. **283**(14): p. 9023-30.

146. Voglauer, R., et al., *SNEV overexpression extends the life span of human endothelial cells.* Exp Cell Res, 2006. **312**(6): p. 746-59.

147. Dellago, H., et al., *ATM-dependent phosphorylation of SNEVhPrp19/hPso4 is involved in extending cellular life span and suppression of apoptosis.* Aging (Albany NY), 2012. **4**(4): p. 290-304.

148. Krieghoff, E., K. Milovic-Holm, and T.G. Hofmann, *FLASH meets nuclear bodies: CD95 receptor signals via a nuclear pathway.* Cell Cycle, 2007. **6**(7): p. 771-5.

149. Barcaroli, D., et al., *FLASH is required for histone transcription and S-phase progression.* Proc Natl Acad Sci U S A, 2006. **103**(40): p. 14808-12.

150. Tabor, C.W. and H. Tabor, *Polyamines.* Annu Rev Biochem, 1984. **53**: p. 749-90.

151. Shi, Y., *Assembly and structure of protein phosphatase 2A.* Sci China C Life Sci, 2009. **52**(2): p. 135-46.

152. Seshacharyulu, P., et al., *Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer.* Cancer Lett, 2013. **335**(1): p. 9-18.

153. Kaufman, P.D., et al., *The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication.* Cell, 1995. **81**(7): p. 1105-14.

154. Yu, Z., et al., *Histone chaperone CAF-1: essential roles in multi-cellular organism development.* Cell Mol Life Sci, 2014. **72**(2): p. 327-37.

155. Kubota, H., *Function and regulation of cytosolic molecular chaperone CCT.* Vitam Horm, 2002. **65**: p. 313-31.

156. Brackley, K.I. and J. Grantham, *Activities of the chaperonin containing TCP-1 (CCT): implications for cell cycle progression and cytoskeletal organisation.* Cell Stress Chaperones, 2009. **14**(1): p. 23-31.

157. Le Sourd, F., et al., *eEF1B: At the dawn of the 21st century.* Biochim Biophys Acta, 2006. **1759**(1-2): p. 13-31.

158. Hata, T. and M. Nakayama, *Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death.* Biochim Biophys Acta, 2007. **1773**(7): p. 1039-51.

159. Heby, O. and L. Persson, *Molecular genetics of polyamine synthesis in eukaryotic cells.* Trends Biochem Sci, 1990. **15**(4): p. 153-8.

160. Igarashi, K. and K. Kashiwagi, *Polyamines: mysterious modulators of cellular functions.* Biochem Biophys Res Commun, 2000. **271**(3): p. 559-64.

161. Igarashi, K. and K. Kashiwagi, *Modulation of cellular function by polyamines.* Int J Biochem Cell Biol, 2010. **42**(1): p. 39-51.

162. Park, J.H., et al., *Reversal of the deoxyhypusine synthesis reaction. Generation of spermidine or homospermidine from deoxyhypusine by deoxyhypusine synthase.* J Biol Chem, 2003. **278**(35): p. 32683-91.

163. Hawel, L., 3rd and C.V. Byus, *A streamlined method for the isolation and quantitation of nanomole levels of exported polyamines in cell culture media.* Anal Biochem, 2002. **311**(2): p. 127-32.

164. Iwata, S., et al., *Anti-tumor activity of antizyme which targets the ornithine decarboxylase (ODC) required for cell growth and transformation.* Oncogene, 1999. **18**(1): p. 165-72.

165. Matsui, I. and A.E. Pegg, *Effect of inhibitors of protein synthesis on rat liver spermidine N-acetyltransferase.* Biochim Biophys Acta, 1981. **675**(3-4): p. 373-8.

166. Hayashi, S., Y. Murakami, and S. Matsufuji, *Ornithine decarboxylase antizyme: a novel type of regulatory protein.* Trends Biochem Sci, 1996. **21**(1): p. 27-30.

167. Fong, W.F., J.S. Heller, and E.S. Canellakis, *The appearance of an ornithine decarboxylase inhibitory protein upon the addition of putrescine to cell cultures.* Biochim Biophys Acta, 1976. **428**(2): p. 456-65.

168. Murakami, Y., et al., *Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination.* Nature, 1992. **360**(6404): p. 597-9.

169. Coffino, P., *Regulation of cellular polyamines by antizyme.* Nat Rev Mol Cell Biol, 2001. **2**(3): p. 188-94.

170. Kitani, T. and H. Fujisawa, *Purification and some properties of a protein inhibitor (antizyme) of ornithine decarboxylase from rat liver.* J Biol Chem, 1984. **259**(16): p. 10036-40.

171. Perez-Leal, O. and S. Merali, *Regulation of polyamine metabolism by translational control.* Amino Acids, 2012. **42**(2-3): p. 611-7.

172. Mangold, U. and E. Leberer, *Regulation of all members of the antizyme family by antizyme inhibitor.* Biochem J, 2005. **385**(Pt 1): p. 21-8.

173. Ivanov, I.P., R.F. Gesteland, and J.F. Atkins, *Antizyme expression: a subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit.* Nucleic Acids Res, 2000. **28**(17): p. 3185-96.

174. Sakata, K., K. Kashiwagi, and K. Igarashi, *Properties of a polyamine transporter regulated by antizyme.* Biochem J, 2000. **347 Pt 1**: p. 297-303.

175. Suzuki, T., et al., *Antizyme protects against abnormal accumulation and toxicity of polyamines in ornithine decarboxylase-overproducing cells.* Proc Natl Acad Sci U S A, 1994. **91**(19): p. 8930-4.

176. Ivanov, I.P., R.F. Gesteland, and J.F. Atkins, *A second mammalian antizyme: conservation of programmed ribosomal frameshifting.* Genomics, 1998. **52**(2): p. 119-29.

177. Zhu, C., D.W. Lang, and P. Coffino, *Antizyme2 is a negative regulator of ornithine decarboxylase and polyamine transport.* J Biol Chem, 1999. **274**(37): p. 26425-30.

178. Ivanov, I.P., et al., *Discovery of a spermatogenesis stage-specific ornithine decarboxylase antizyme: antizyme 3.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4808-13.

179. Hayashi, S., *Antizyme-dependent degradation of ornithine decarboxylase.* Essays Biochem, 1995. **30**: p. 37-47.

180. Matsufuji, S., et al., *Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme.* Cell, 1995. **80**(1): p. 51-60.

181. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

182. Folk, J.E., et al., *Polyamines as physiological substrates for transglutaminases.* J Biol Chem, 1980. **255**(8): p. 3695-700.

183. Gerner, E.W. and F.L. Meyskens, Jr., *Polyamines and cancer: old molecules, new understanding.* Nat Rev Cancer, 2004. **4**(10): p. 781-92.

184. Pegg, A.E. and R.A. Casero, Jr., *Current status of the polyamine research field.* Methods Mol Biol, 2011. **720**: p. 3-35.

185. Mandal, S., et al., *Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells.* Proc Natl Acad Sci U S A, 2013. **110**(6): p. 2169-74.

186. Haddox, M.K., B.E. Magun, and D.H. Russell, *Ornithine decarboxylase induction during B1 progression of normal and Rous sarcoma virus-transformed cells.* Cancer Res, 1980. **40**(3): p. 604-8.

187. Holtta, E., M. Auvinen, and L.C. Andersson, *Polyamines are essential for cell transformation by pp60v-src: delineation of molecular events relevant for the transformed phenotype.* J Cell Biol, 1993. **122**(4): p. 903-14.

188. Holtta, E., L. Sistonen, and K. Alitalo, *The mechanisms of ornithine decarboxylase deregulation in c-Ha-ras oncogene-transformed NIH 3T3 cells.* J Biol Chem, 1988. **263**(9): p. 4500-7.

189. Miyazaki, Y., et al., *Spermine enhances IgM productivity of human-human hybridoma HB4C5 cells and human peripheral blood lymphocytes.* Cytotechnology, 1998. **26**(2): p. 111-8.

190. Sakata, K., et al., *Acrolein produced from polyamines as one of the uraemic toxins.* Biochem Soc Trans, 2003. **31**(2): p. 371-4.

191. Auvinen, M., et al., *Ornithine decarboxylase activity is critical for cell transformation.* Nature, 1992. **360**(6402): p. 355-8.

192. Moshier, J.A., et al., *Transformation of NIH/3T3 cells by ornithine decarboxylase overexpression.* Cancer Res, 1993. **53**(11): p. 2618-22.

193. Shantz, L.M. and V.A. Levin, *Regulation of ornithine decarboxylase during oncogenic transformation: mechanisms and therapeutic potential.* Amino Acids, 2007. **33**(2): p. 213-23.

194. Gilmour, S.K., *Polyamines and nonmelanoma skin cancer.* Toxicol Appl Pharmacol, 2007. **224**(3): p. 249-56.

195. Mangold, U., *The antizyme family: polyamines and beyond.* IUBMB Life, 2005. **57**(10): p. 671-6.

196. Tsuji, T., et al., *Ornithine decarboxylase antizyme upregulates DNA-dependent protein kinase and enhances the nonhomologous end-joining repair of DNA double-strand breaks in human oral cancer cells.* Biochemistry, 2007. **46**(31): p. 8920-32.

197. Yamamoto, D., et al., *Ornithine decarboxylase antizyme induces hypomethylation of genome DNA and histone H3 lysine 9 dimethylation (H3K9me2) in human oral cancer cell line.* PLoS One, 2010. **5**(9): p. e12554.

198. Koike, C., D.T. Chao, and B.R. Zetter, *Sensitivity to polyamine-induced growth arrest correlates with antizyme induction in prostate carcinoma cells.* Cancer Res, 1999. **59**(24): p. 6109-12.

199. Murakami, Y., et al., *Forced expression of antizyme abolishes ornithine decarboxylase activity, suppresses cellular levels of polyamines and inhibits cell growth.* Biochem J, 1994. **304 (Pt 1)**: p. 183-7.

200. Wang, X. and L. Jiang, *Effects of ornithine decarboxylase antizyme 1 on the proliferation and differentiation of human oral cancer cells.* Int J Mol Med, 2014. **34**(6): p. 1606-12.

201. Ivanov, I.P., et al., *Conservation of polyamine regulation by translational frameshifting from yeast to mammals.* EMBO J, 2000. **19**(8): p. 1907-17.

202. Feith, D.J., et al., *Tumor suppressor activity of ODC antizyme in MEK-driven skin tumorigenesis.* Carcinogenesis, 2006. **27**(5): p. 1090-8.

203. Mitchell, J.L., et al., *Feedback repression of polyamine transport is mediated by antizyme in mammalian tissue-culture cells.* Biochem J, 1994. **299 (Pt 1)**: p. 19-22.

204. Yang, X., et al., *A public genome-scale lentiviral expression library of human ORFs.* Nat Methods, 2011. **8**(8): p. 659-61.

Su Xiao

suxiao.728@gmail.com | +1 (410) 949-7333 | 10504 Grove Ridge Pl, Rockville, MD 20852

Education

Ph.D. Chemical and Biomolecular Engineering, **Johns Hopkins University**   GPA 3.8   May,2015 (expected)

   Joint program with **National Institutes of Health (NIH)**

   Advisors: Michael J. Betenbaugh and Joseph Shiloach

   Thesis title: cellular and process engineering to improve mammalian membrane protein expression

B.S. Chemical Engineering, **Tsinghua University**, China   GPA 3.3   July, 2010

Professional experience

Jan. 2011-present: pre-doctoral research fellow, NIDDK Biotechnology Core Laboratory, National Institutes of Health, Bethesda, MD.

Research

• 10L-scale production and purification of hard-to-express membrane protein

 * Designed and constructed vectors for inducible neurotensin receptor (NTSR1) expression in HEK293 cells

 * Constructed stable HEK293 cell clones for inducible NTSR1 expression

 * Selected production clones using flow cytometry and ligand-binding assay

 * Adapted production clones into high density serum-free suspension culture

 * Designed and optimized NTSR1 manufacturing method

 * Operated 10L bioreactor followed by immobilized metal ion affinity chromatography for manufacturing of NTSR1

 * Determined the quality and quantity of NTSR1 product

 * Quantitatively compared NTSR1 production from HEK293 and baculovirus-insect cell expression system

• High-throughput microRNA screen for improved recombinant protein expression

 * Developed and optimized a high-content imaged-based assay for RNAi screening with GFP tagged proteins.

* Fine-tuned workflow and solved technical difficulties in the automated screening procedure

 * Screened human microRNA library with NTSR1-GFP expressing cells

 * Proposed to combine NTSR1-GFP data with luciferase data to identify common hits improving protein produciton

 * Verified the effectiveness of identified common hits (microRNAs) with other secreted and membrane proteins

 * Constructed stable microRNA over-expressing cell line for enhanced phenotype

• High-throughput whole-genome siRNA screen for improved cell growth and recombinant protein expression

 * Fine-tuned workflow and solved technical difficulties for whole-genome siRNA screening

 * Screened human whole-genome siRNA library with NTSR1-GFP expressing cells

 * Statistically analyzed large-scale whole genome data for both NTSR1-GFP and luciferase screens

 * Executed pathway analysis to obtain biological interpretation of the screen result

 * Identified and validated top molecules (siRNAs) for improved cell growth and protein expression

 * Verified the effectiveness of identified top molecules (siRNAs) with other secreted and membrane proteins

 * Investigated the mechanistic explanation for improved protein overexpression

 * Designed CRISPR array screening for construction of stable cell line knocking out identified genes

• Transient overexpression and purification of Respiratory syncytial virus (RSV) fusion protein

 * Investigated four expression vectors for optimal secreted expression of RSV F0 protein.

 * Optimized transient transfection parameters and produced RSV F0 protein accordingly.

 * Purified RSV F0 protein with immobilized metal ion affinity chromatography.

• Rotavirus amplification in Madin-Darby canine kidney (MDCK) cells

 * Amplified four strains of rotavirus in MA104 cells and demined rotavirus titer in cell culture lysate

 * Adapted rotavirus to MDCK cells

 * Determined virus titer change during adaptation using classical plaque assay and ELISA test

Collaborative/ supervisory/ technical work

 * Collaborated with students, postdocs, staff scientists and technicians from both domestic and international locations. Collaborative effort includes setting up regular meetings, transferring of cell lines and reagents, coordinating and travelling to different sites to used shared facilities, data sharing and

editing manuscripts, etc.

 * Served as a teaching assistant for bioseparation course (90 undergraduate students) for 2 semesters. Work includes delivery lectures, holding office hours to answer questions, establishing and trouble-shooting new equipment for lab, organizing students for lab practicing, etc.

 * Supervised two graduate students and two undergraduate students on cell culture, cell engineering, virus amplification, data analysis and thesis writing.

 * In charge of preventive maintenance and calibration of analytical instruments.

 * Maintained updated inventory of analytical instrument, lab equipment, consumables and cell lines.

 *Searched for state-of-the-art analytical instruments for purchasing recommendations.

 * Created bacterial and mammalian cell banks

June- Sept. 2014: managerial internship at Center for Drug Evaluation (CDE), China Food and Drug Administration, Beijing, China.

 * Investigated Good Review Practices (GRP) and Quality Management System(QMS)

 * Proposed to CDE on implementing ISO 9000 QMS with Six Sigma or Lean methodologies for continuous improvement

 * Led a group of interns to search, translate and compile USFDA QMS/GRP documents

 * Interviewed and worked with consulting firms on implementing QMS/GRP in CDE

October 2008- June 2010: Undergraduate research at Department of Chemical Engineering, Tsinghua University, Beijing, China

 * Produced violacein through microbial fermentation, extraction and purification.

 * Characterized and analyzed violacein product.

 * Introduced the entire violacein synthetic pathway from *Duganella sp*. B2 into *C. freundii* to achieve high efficient biosynthesis of violacein.

 * Explored possible impact factors on efficient biosynthesis of violacein by integrating the information on transcription/translation level of violacein biosynthesis gene cluster, fermentation characteristics and intracellular pH.

July-Sept. 2009: summer training program at Technical University of Denmark, Novozymes and Nordic Sugar, Denmark

 * Received training on large-scale unit operation, including liquid-liquid extraction, thermal-fluid heat

exchange, fluid bed powder processing and filtration.

 * Studied the bioprocess of enzyme and sugar production and inspected equipment in each section.

Publications

Xiao S, Chen YC, Betenbaugh MJ, Martin SE, Shiloach J: MiRNA screen for improved functional expression of

neurotensin receptor. *Biotechnology and Bioengineering* 2015.

Xiao S, Shiloach J, Grisshammer R: Construction of recombinant HEK293 cell lines for the expression of the

neurotensin receptor NTSR1. *Methods in Molecular Biology* 2015, 1272:51-64.

Xiao S, Betenbaugh MJ, Shiloach J: Engineering Cells to Improve Protein Expression. *Current Opinion in Structural Biology*. 2014 Apr 3; 26C:32-38.

Xiao S, White JF, Betenbaugh MJ, Grisshammer R, Shiloach J: Transient and stable expression of the neurotensin receptor NTS1: a comparison of the baculovirus-insect cell and the T-REx-293 expression systems. *PLoS One* 2013, 8:e63679.

Jiang PX, Zhang RP, Wang HS, **Xiao S**, Yang C, Xing XH: Outer space mutagenesis of violacein-producing strain by spacecraft Shenzhou-7 and screening of the mutants with high violacein productivity. *Journal of Chemical Industry and Engineering* 2010, 61(2): 461-461.

Scientific meetings/ presentations

Xiao S, Chen YC, While JF, Betenbaugh MJ, Martin SE, Grisshammer R, Shiloach J, 'Engineering towards improved functional expression of neurotensin receptor'. 11th Annual Graduate Student Research Symposium, Bethesda, US, Jan. 2015 (poster, travel award winner)

Xiao S, Chen YC, Betenbaugh MJ, Martin SE, Shiloach J, 'Systems and Synthetic Biotechnology Platform for Characterizing and Designing HEK293 Cells'. IBC's 10th Annual Cell Line Development & Engineering conference, Berkeley, CA, US, Sept. 2014 (Oral)

Xiao S, White JF, Betenbaugh MJ, Grisshammer R, Shiloach J, 'High throughput RNA interference for improved functional expression of neurotensin receptor'. Cell Culture Engineering XIV, Quebec city, Canada, May 2014 (poster)

Xiao S, White JF, Betenbaugh MJ, Grisshammer R, Shiloach J, 'Transient and stable expression of the neurotensin receptor NTSR1: a comparison of the baculovirus-insect cell and the T-REx-293 expression systems'. 10th Annual NIDDK Conference, Bethesda, MD, US, Apr. 2013 (poster)

Xiao S, White JF, Betenbaugh MJ, Grisshammer R, Shiloach J, 'Transient and stable expression of the neurotensin receptor NTSR1: a comparison of the baculovirus-insect cell and the T-REx-293 expression systems'. Delaware Membrane Protein Symposium, Newark, DE, US, May. 2012 (poster)

Awards and Honors

 * 2015 NIH Graduate Student Research Award (4%), National Institutes of Health, 2015

 * 1st Prize in National Competition of Chemical Process Design for College Students (2%), Chemical Industry and Engineering Society of *China,* 2009

 * Scholarship for Scientific Innovation, Tsinghua University, 2009

 * 1st Prize in "Challenge Cup" Scientific and Technological Contest (2.5%), Tsinghua University, 2009

 * *1st Prize in Chemical Product Design Contest (3%), Tsinghua University, 2008*

Skills

Laboratory techniques

* Mammalian cell culture (CHO, MDCK, HEK293, Vero, MA104), insect cells (sf9), yeast and bacteria.

* Gene and primer design, plasmid DNA cloning, bacterial transformation, transient and stable mammalian transfection, virus infection

* Gene overexpression, siRNA and shRNA knock-down, CRISPR knockout

* DNA, RNA extraction and purification, DNA and RNA gel electrophoresis, cDNA synthesis, qRT-PCR

* Protein expression and IMAC purification, SDS-PAGE, western blotting, ELISA, BCA, Bradford protein assay, ligand binding assay, other enzymatic assays (SEAP, luciferase, etc.)

* metabolites analysis with YSI and HPLC

* Flow cytometry, fluorescent microscopy, confocal microscopy, ImageXpress Micro.

* Virus amplification, plaque assay, TCID50 assay, virus isolation and identification

* Large-scale fermentation, process optimization, design of experiment

* high-throughput RNAi and compound screening, genome scale data analysis and pathway analysis

Personal skills

Excellent problem-solving, organizational and interpersonal skills, able to multi-task, work independently or as a part of a team, good time management and **prioritization** skills, creative, fast-learning, able to identify existing bottleneck and implement new protocols.

computer

Proficient OFFICE, ORIGIN, ENDNOTE, MATLAB, ASPEN Plus, IPA, DNASTAR, PATBASE