OPTIMIZING PROTEIN PURIFICATION FOR RNA-BINDING RECOMBINANT FUSION PROTEINS

Tingyi Li

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Division of Pharmacoengineering and Molecular Pharmaceutics (DPMP) in the Eshelman School of Pharmacy

Chapel Hill
2021

Approved by:

Leaf Huang

Philip Smith

Juliane Nguyen

ABSTRACT

Tingyi Li: Optimizing protein purification for RNA-binding recombinant fusion proteins
(Under the direction of Juliane Nguyen)

There has been great interest in delivering short interfering RNA (siRNA) and microRNA (miRNA) for therapeutic applications. However, the delivery of small RNAs remains challenging due to its inefficient cellular uptake and instability under physiological conditions. Here, we engineered a CXCR4-targeting RNA-protein nanoplex that consists of a CXCR4-targeting single-chain variable fragment (scFv) antibody, which is fused to an RNA-binding protamine peptide (RSQSRSRYYRQRQRSRRRRRRS). To obtain a functional RNA-binding protein, the removal of external nucleic acids is essential. This study aims to optimize the purification process of RNA-binding fusion proteins to free up RNA-binding domains and study if protein/siRNA complexes could successfully protect and deliver siRNA to silence cellular genes. After testing out different nucleic-acid removal methods, the high-salt-wash assisted immobilized metal affinity column purification method showed a great reduction of bound nucleic-acid contaminants. The purified fusion proteins showed a successful complexation with siRNAs and could deliver siRNA to the targeted cells.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Juliane Nguyen, for her guidance, feedback, suggestions, and encouragement throughout my graduate career. I was one of the three students who started the research experience at the University of Buffalo, then transferred to UNC at Chapel Hill. With mine very limited research experience, she was incredibly patient and aiding during the troubleshooting of failed experiments. Dr. Nguyen kept training me to think critically, multitask, and troubleshooting. She was tremendously supportive when I switched to the master's degree program, and guided me throughout the whole graduating process. For all of Dr. Nguyen's guidance, I am incredibly thankful and appreciative.

I would like to thank my thesis committee members, Dr. Leaf Huang and Dr. Philip Smith, for their insightful suggestions and comments throughout my research. I would also like to thank them not only for their academic evaluation of my thesis but also for their warm welcomes as I moved to North Carolina. I'd also like to thank the other professors in the DPMP department for their continued guidance and teachings.

I also want to thank all members of Dr. Nguyen's laboratory, both present, and past members. I've had the great joy working with many undergraduate, Master's students, Ph.D. candidates, and post-docs, such as Dr. Michael B Deci, Dr. Scott Ferguson, Dr. Jinli Wang, Dr. Adam Brown, Dr. Anajli Verma, Emily Bonacquisti, Natalie Smith, Christine Lee, Danni Chen, Jacqueline Gonya, etc. Emily and Natalie in

particular have been tremendously helpful and our friendship has been very supportive and essential in my graduate life.

I would like to express my deepest appreciation to my family and friends for their endless support, and understanding. Their constant encouragement has been invaluable.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

INTRODUCTION...1

MATERIAL AND METHODS..4

RESULTS AND DISCUSSION 11

CONCLUSIONS 24

APPENDIX 26

REFERENCES..29

LIST OF TABLES

Table 1 - PCR primers for cloning CXCR4_scFv and CXCR4_RBM fusion proteins……..…….12

LIST OF FIGURES

Table 1 - PCR primers for cloning CXCR4_scFv and CXCR4_RBM fusion proteins……..……12

Figure 1 - Schematic of cloning for CXCR4_RBM constructs…...……...……12

Figure 2 - EtBr-stained 0.8% agarose gel images of DNA fragments for cloning……..……..…13

Figure 3 -Test digestion image for CXCR4_2P ..…….……13

Figure 4 - The alignment of sequencing results and their corresponding plasmid map……..……..14

Figure 5A - 12% SDS-PAGE of purified CXCR4_scFv, CXCR4_1P, and CXCR4_2P……..…15

Figure 5B - Schematics of CXCR4_1P/miRNA nanoplexes ……..…..…15

Figure 5C - Design of alternative CXCR4_RBM fusion proteins ....…….....15

Figure 6A - RiboGreen assay of CXCR4_1P (Inclusion body) ..…….…….16

Figure 6B - Eluted protein (Inclusion body) loaded on 1% agarose gel stained with EtBr…….16

Figure 6C - RiboGreen assay of CXCR4_1P (Periplasmic extracted) ..……..16

Figure 6D - Eluted protein (Periplasmic extracted) loaded on 1% agarose gel stained with EtBr....…….....16

Figure 7A - EtBr-stained agarose gel images of fractions collected from anion-exchange column ..……...17

Figure 7B - SDS-PAGE of fractions collected from anion-exchange column……..17

Figure 8 - RiboGreen assay result of anion-exchanged-column treated CXCR4 fusion proteins complexed with 20ng (0.14pmol) miRNA duplex ……18

Figure 9A - EtBr-stained agarose gel and SDS-PAGE of fractions collected from high salt wash for CXCR4_only ...……..……19

Figure 9B - EtBr-stained agarose gel and SDS-PAGE of fractions collected from high salt wash for CXCR4_1P…….19

Figure 10 - RiboGreen assay of high-salt treated CXCR4_only and CXCR4_1P complexed with 20ng (0.14pmol) miRNA duplex ...……..……...20

Figure 11 - ELISA-based binding assay against N-terminus of CXCR4 receptor……..……...21

Figure 12 - Luciferase assay result of luciferase-expressing MDA cells treated with different CXCR4 fusion protein/siRNA complexes……..22

Figure 13 - Model of miRNA duplex ducking with different length of protamine sequences ..…….23

Figure 14 - RiboGreen assay of PEI treated CXCR4_only and CXCR4_1P ...…….....27

Figure 15 - Et-Br stained agarose gel image of nuclease treated CXCR4_1P protein……..……..……..28

OPTIMIZING PROTEIN PURIFICATION FOR RNA-BINDING RECOMBINANT FUSION PROTEINS

Introduction

Small RNAs such as small-interfering (siRNA) and microRNA (miRNA) are short (~18 to 30 nucleotides) non-coding RNA molecules that can regulate gene expression by post-transcriptional modification. As important regulators of physiology and development, they play a critical part in cellular processes such as differentiation, proliferation, metabolism.[1].Compelling evidence has illustrated that miRNA expression is dysregulated in human cancer. The miRNA could be either oncogenes or tumor suppressors, which affect cancer hallmarks, including activating invasion or metastasis, inducing angiogenesis, and many other factors[1]. Because small RNAs do not integrate into the genome and can be synthesized easily, the field of small RNAs drug development has expanded considerably. Several small RNA-targeted therapeutics have been recently approved by the FDA, such as patisiran and givosiran, approved in 2018 and 2019 respectively. The first FDA-approved siRNA drug, patisiran is approved for hereditary transthyretin-mediated (hATTR) amyloidosis and works by degrading the messenger RNA transcript for transthyretin. Clinical trials have demonstrated significant improvements in patients' quality of life and consistently slowed neuropathy progression [2]. The second siRNA drug, givosiran was approved for acute intermittent porphyria by inhibiting hepatic delta-aminolevulinic acid synthase 1 (ALAS1) synthesis, thereby decreasing neurotoxic intermediates in this disease[3]. Although the miRNA therapeutics

have not yet translated into FDA-approved candidates for medical intervention, there are on-going phase 1 and phase 2 clinical trials of candidate drugs. For instance, ENGeneIC has designed a bacterially derived minicell containing miRNA mimic, MesomiR-1. This miRNA mimic aims to replace miR-16, a tumor suppressor that is reduced in cases of cancer, including malignant pleural mesothelioma or non-small cell lung cancer [4]. In the first human trial, the drug was reported to suppress tumor transcript miR-16 [5].

Moreover, some small RNAs are also under preclinical and clinical development for replacement therapy or vaccination [6,7]. Most RNA drugs are designed to act intracellularly, but RNA molecules are intrinsically unstable and cannot freely cross cellular membranes due to their anionic charges and large molecular weight. Therefore, new approaches and technologies are needed to tackle these issues by not only protecting cargo from degradation but also helping it to cross the cellular membrane. Among all the existing RNA delivery vehicles, protein-based carriers are highly attractive because of their unique engineerability. The size, charge, binding affinity, and targeting moieties of proteins could be easily manipulated to optimize therapeutic cargos' pharmacological efficacy. For many therapeutic applications such as treatments against cancer, specific cellular uptake is critical to minimize side effects and to enhance therapeutic efficacy. As such, studies have genetically fused short RNA-binding motifs (RBM) to targeting ligands, such as single-chain variable fragments (scFvs) for binding to cellular receptors. For instance, Zang et al. have developed a fusion protein (R3P) consisting of FGFR3-scFv-protamine, which could successfully deliver amplified-in-breast-cancer 1 (AIB1) siRNA, leading to RT112 cell apoptosis

[8]. Additionally, Jiang et al. found that their fusion protein, anti-HER2-scFv-arginine nonamer peptide (e23sFv-9R), showed a significant anti-cancer effect while delivering CXCR4 siRNA [9].

In addition, our lab has also engineered a first-generation CXCR4-targeting protein/miRNA nanoplex system, which is composed of CXCR4-targeting scFvs fused with a short protamine peptide sequence RSRYYRQRQRSRRRRRRS for miRNA delivery (hereinafter referred to as CXCR4_RBM). The protamine peptide is commonly used for fusion to scFv for small RNA delivery [10]. Because of the electrostatic interaction between the negatively charged RNA and positively charged protamine sequence, the CXCR4_RBM fusion protein and miRNA will self-assemble into nanoplexes that serve as protective vehicles for delivery of therapeutic miRNA (Fig. 5B). Our published results have shown that the CXCR4_RBM fusion proteins are able to repolarize macrophages to suppress tumors through the CXCR4 blockade, and the co-delivery of M1-polarizing miRNAs using nanoplexes enhanced the outcome [10]. However, the improvement was not statistically significant. Previous studies published by others were lacking critical controls and failed to include non-specific siRNA as a comparison. Thus, the observed gene silencing could be partially due to toxic effects or off-target effects. We hypothesized that 1) the RNA-binding region may be limited or occupied due to unwanted binding with bacterial nucleic acids; 2) the complexation efficiency and RNA-loading efficiency is insufficient due to the short RBM region or low charge density.

We selected *Escherichia coli (E. coli)* as our biological protein factory because of the existence of well-established protocols, its rapid growth, and high-level protein productions. Theoretically, the RNA-binding proteins could bind non-specifically to host

RNA or DNA during overexpression. In order to isolate a functional RNA-binding protein, complete purification of the protein is essential; otherwise, the contaminants will interfere in subsequent biological assays and characterization of the protein-RNA complexes. We implemented several approaches to solve the contamination issue, including ultracentrifugation, polyethylenimeine (PEI) treatment, anion exchange columns, and high-salt washes.

Materials and Methods

Cloning of recombinant protein-coding plasmids

We have identified an scFv sequence that specifically binds to CXCR4, and this protein was cloned into a pet21a vector (Novagen, Burlington, MA). This clone was then used as the vector for cloning the CXCR4-1P (shown in Fig.1). The pET21a was digested with restriction enzymes, NdeI and XhoI (NEB) 4 hours at 37 ^{o}C. The CXCR4 insert was amplified using PCR with forward and reverse primers shown in Table.1. Briefly, The PCR reaction mixture contained forward and reverse primers added to a final concentration of 0.5 µM concentration, 10 ng of template DNA, 5 µL of 5X Q5 reaction buffer, and 0.25 µL of Q5 polymerase (New England Biolabs [NEB], Ipswich, MA). The PCR reaction was activated at 94°C for 1 minute followed by 25 cycles at 94°C for 1 minute, 69°C for 1 minute, and 72°C for 1 minute. The PCR product was then purified and digested with the same restriction enzyme pair, NdeI and XhoI. Both digested PCR and vectors were prepared with gel loading dye (NEB) and then loaded onto 0.8% agarose gel. The gel electrophoresis was performed at 120 V for 20 minutes using a Bio-Rad PowerPac (Hercules, CA) and the gel was imaged on the Bio-Rad ChemiDox imager. The linearized DNA band was excised and purified using the

QIAquick Gel Extraction Kit (Qiagen) as described by the manufacturer. A 1:10 vector to insert molar ratio ligation reaction was prepared with a T7 DNA ligase reaction kit (NEB). The ligation reaction was then incubated at 16 ^{o}C overnight. Total 2 ul of the ligation reaction was transformed into 25 ul of chemically competent cells, NEB-5α (NEB), through heat-shock and then plated on LB+Ampcilin agar plates for overnight incubation at 37 ^{o}C. On the following day, the colonies grown on the agar plates were selected for plasmid isolation. All the plasmid was isolated using Plasmid Mini Kit (Norcross, GA) following the manufacturer's instructions. DNA concentrations were measured using NanoDrop 2000 (Thermofisher Scientific, Waltham, MA). The sequences for the clone were confirmed by the Genwiz sequencing company. For the cloning of CXCR4-1P, CXCR4_2P, the same procedure was performed with different restriction enzyme pairs shown in Figure1.

Inclusion body isolation

The CXCR4-scFv cloned into the pET21a vector (Novagen, Burlington, MA) was transformed into BL-21 (DE3) cells (Lucigen, Middleton, WT). The BL21 cells producing the CXCR4_scFv or CXCR4_RBM fusion proteins were initially grown overnight at 37 °C in LB broth with 100 μg/mL ampicillin, shaking at 250 rpm. Then, the overnight culture was diluted in a 1:100 ratio in a 1 L culture media bottle and incubated at 37 °C until reached an optical density at 600 nm (OD600) of 0.6. The induction of protein expression was performed with IPTG at a final concentration of 0.2 mM at 24 °C for overnight. The cell pellet was resuspended in lysis buffer containing 50 mM Tris, 100 mM KCl, 2 M urea and one tablet of Pierce protease and phosphatase inhibitor mini tablets, EDTA-free (ThermoFisher Scientific, Waltham, MD) at pH 8.5. The cell suspension was centrifuged

at 9000 x *g* for 5 min at 4 °C, and the pellet was collected. After four lysis washes, the pellet was resuspended in solubilization buffer containing 50 mM Tris, 100 mM KCl, 8 M guanidine, and 10 mM β-mercaptoethanol at pH 8.5. The solution was incubated at 4°C overnight and then centrifuged at 9000 x *g* for 15 min at 4 °C. The supernatant was then purified with TALON cobalt resin (Clontech, Mountain View, CA) and diluted into refolding buffer (50 mM Tris, 500 mM NaCl, 400 mM sucrose, 3 mM reduced glutathione, 0.3 mM oxidized glutathione, 0.5% Triton X-100, 10% glycerol, and 10 mM imidazole at pH 8.5). The solution was stirred at 200 rpm at 4 °C for 24 h and then purified with TALON cobalt beads (Clontech, Mountain View, CA) by immobilized metal affinity chromatography (IMAC). After incubation of the supernatant with cobalt resin for 1 h at 4 ^{o}C, an imidazole gradient method was used to collect protein at elution buffer containing 13 mM sodium phosphate monobasic, anhydrous, 73 mM sodium phosphate dibasic, anhydrous, 300 mM NaCl, 8 M urea and 10 mM imidazole at pH 7.4. The protein in elution buffer was then diluted by 1:10 into refolding buffer containing 50 mM Tris, 500 mM NaCl, 400 mM sucrose, 3 mM reduced glutathione, 0.3 mM oxidized glutathione, 0.5 % Triton X-100, 10% glycerol, and 10 mM imidazole at pH 8.5. The solution was stirred at 250 rpm at 4 ^{o}C for 48 hours. After 48 hours in refolding buffer, the protein was dialyzed against 1x PBS with 10% glycerol (pH 7.4) using SnakeSkin™ Dialysis Tubing, 10 K MWCO, 35 mm (ThermoFisher Scientific, Waltham, MA). Finally, the protein was concentrated in Amicon® ultra protein concentrator, 10 kDa, 15 ml (EMD Millipore Corporation, Danvers, MA). Protein expression for CXCR4-1P, CXCR4_2P was performed as described above for CXCR4-scFv.

Periplasmic extraction method

The protein was induced as previously described in the inclusion body method. Cell pellets were collected by centrifugation at 9000 rpm for 10 minutes at 4 ºC. For 1L of culture, resuspending cell pellet with 25ml of ice-cold buffer A (50 mM Tris-HCl, 1mM EDTA, 20% sucrose, protease inhibitor tablet, at pH 7.2). The resuspension was incubated on ice for 30min~1hr with gentle shaking, and then was centrifuged at 10,000 xg for 20min. The supernatant was collected. Then, resuspending cell pellet with 25ml of buffer B (cold 5 mM MgSO4 with protease inhibitor), also incubate for 30min on ice. After centrifuging at 10,000 xg for 20 minutes, retain the 2nd supernatant and combined it with 1st supernatant. The protein-containing supernatant was then dialyzed against 1x PBS with 10% glycerol to remove EDTA. The dialyzed solution was then purified by cobalt-bead-based IMAC chromatography as previously described.

Purification and identification

Protein samples were mixed with 4x Laemmli SDS sample reducing buffer (Alfa Aesar, Haverhill, MA) and then denatured at 95 ºC for 10 minutes. After denaturation, protein samples and PageRuler^{TM} plus prestained protein ladder (Thermofisher Scientific, Waltham, MA) were loaded into 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The SDS-PAGE was run at 180 V for 1.5 h and then stained with Coomassie Brilliant Blue R-250 staining solution (Bio-Rad, Hercules, CA) for 1 h. After staining, the SDS-PAGE was destained by the destain solution containing 50% of methanol, 40% distilled water, and 10% glacial acetic acid overnight. The next day, the SDS-PAGE was further destained by distilled water in a rocker for 30 minutes and then the SDS-PAGE was imaged using eh Bio-Rad ChemiDoc Imager. Proteins were also

stained with ethidium bromide following electrophoresis on 1% agarose gel to detect nucleic acid contaminants.

RNA Complexation Assay

The Quant-iT^{TM} Ribogreen®RNA Assay kit purchased from ThermoFisher Scientific (Waltham, MA) was used to assay RNA complexation. A fixed amount, 20ng of 22-bp miRNA duplex mimic was complexed with the following molar ratios of each protein: 1:0, 1:10, 1:20, 1:30, 1:40, and 1:50 at room temperature for 30 minutes. And then, 100 ul of Quant-iT^{TM} Ribogreen reagent solution (1:200 dilution) was added to each well in a Corning® 96 well black polystyrene microplate (Sigma-Aldrich, St.Louis, MO). After incubation at room temperature for 5 minutes in the dark, fluorescence was measured at an excitation wavelength of 480 nm and an emission wavelength of 520 nm by using a SpectraMax i3 Plate Reader.

Anion exchange column

The pH of the protein samples was increased to 12, so that the pH is greater than the isoelectric point of the CXCR4-1P (isoelectric point =9.7). At such alkalized conditions, fusion proteins are no longer positively charged, leading to the dissociation of nucleic acid contaminants. The protein sample was then loaded onto columns packed with POROS AEX resins (Thermo scientific) and incubated for 30 min. Packing of columns was performed according to specifications by the manufacturer. The more negatively charged nucleic acids would bind to resin more tightly than the CXCR4-1P fusion proteins. Buffers containing increased NaCl concentrations were used as washing buffers, from 200 mM, 500 mM, 800 mM to 1 mM. Protein was eluted with a

low concentration buffer, whereas the contaminants were eluted with a high concentration buffer.

High-salt-wash assisted IMAC purification

After proteins were applied to a His-trap cobalt bead packed column, the flowthrough was discarded by gravity flow. The beads were then incubated with 1.5 M NaCl buffer for 20 minutes. For a 1 ml of bead slurry, 50 ml buffer was used. The concentration of NaCl was then brought down to 200mM by sequential gradient washes. The protein of interest was then eluted using elution buffer (50 mM sodium phosphate, 200 mM sodium chloride, 250mM imidazole, pH 7.4).

ELISA-based fusion protein binding affinity assay

The binding affinity of the CXCR4 scFv and CXCR4_1P was assessed with an ELISA. The 96-well plate was first coated with neutravidin and then coated with the biotinylated CXCR4 N-terminal peptide. The purified fusion proteins were serially diluted and added to the 96-well plate. Anti-FLAG-HRP conjugate (Sigma Aldrich) was added to bind fusion proteins as the primary antibody. 100 µL of Ultra-TMB substrate (Thermo Fisher Scientific) was added to quantify the amount of bound fusion proteins. The reaction was then quenched with 12.5 M H_2SO_4 aqueous solution. Readings were taken using a plate reader with the wavelength set to 450 nm.

Luciferase knockdown assay

MDA-MB-231 breast cancer cells stably expressing firefly luciferase were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), and 5ug/ml hygromycin. On day 1, cells were plated

in 96-well plates at a density of 10,000/well. CXCR4-1P fusion proteins were complexed with 0.14 pmol of either anti-luciferase siRNA or non-specific siRNA at 1:10 1:20, 1:40, and 1:50 (protein to siRNA) molar ratios. The CXCR4_RBM/siRNA nanoplexes or lipofectamine/siRNA formulation were added to each well containing 200 µl of medium supplemented with 10% FBS. After 24 h incubation, the medium was replaced, and the cells were allowed to grow for an additional 24 h. Non-specific siRNA was used as a negative control to account for cytotoxic effects or off-target effects. On day 3, cells were washed with PBS and 100 µl of Brightglo reagent (Promega) was added to each well and incubated for 15 min. Luciferase expression was measured with the plate reader. Luciferase expression was expressed as a percentage of naked siRNA treated cells. All experiments were performed in triplicate and are presented as mean ± sd.

Molecular docking

The structure of protamine peptides was predicted by the I-TASSER server using iterative template-based fragment assembly simulations. The structure of miRNA duplex was predicted by SimRNA web server, which utilizes the Monte Carlo method to sample the conformational space, and relies on a statistical potential to describe the folding interaction. Docking simulations were performed using the HDOCK server, in which an improved shape-based pairwise scoring function has been used (available at http://hdock.phys.hust.edu.cn/). The docking server uses both sequences and structures as input for proteins, and uses structure input for DNAs/RNAs. The binding model with the best docking score was selected.

Results and discussion

Cloning of alternative CXCR4_RBM protein constructs

Our CXCR4-RBM fusion proteins were constructed based on a CXCR4-targeting scFv that we identified using phage display. In previously published data, we illustrated that the CXCR4-scFv bind to both human and mouse CXCR4+ cells with high affinity [10]. The short protamine sequence was fused to the C terminus of CXCR4_scFv protein.

We first started with inclusion body isolation for our protein expression and then observed significant nucleic acid contaminants. In order to solve this problem, we implemented another isolation method: periplasmic extraction. This method requires a signaling peptide to allow bacteria to export proteins into periplasmic space. In general, signaling sequences are rich in hydrophobic amino acids, such as alanine, valine, and leucine, which is an essential feature for the secretion of the proteins into the periplasm of E. coli [12]. Several studies have used ompA (MKKTAIAIAVALAGFATVAQA) as the signaling peptide for their scFv proteins and showed successful isolation of functional proteins from the periplasm [13,14].

Therefore, we incorporated the ompA peptide sequence at the N terminus of the CXCR4-scFv, and the new ompA_CXCR4-scFv DNA fragment was cloned into the pET21a vector (Novagen, Burlington, MA). Sequential cloning was performed for ompA_CXCR4_2P. The schematic of cloning is shown in Figure 1.

The images of EtBr-stained 0.8% agarose gel of CXCR4 were selected as representative images to illustrate the cloning process shown in Figure 2. The two images shown in Figure 2 demonstrate the successful digestion of the vector and PCR insert

fragments. After ligation and transformations, multiple colonies were picked for test digestion. The agarose gel image of the digested CXCR4-2P was selected as the representative image for an illustration, as shown in Figure 3. The molecular length of one protamine insert was about 85 bp, and the size of two protamine inserts would be 170 bp. In Figure 3, the released fragments (red boxed) were the two-protamine inserts, indicating potentially working candidates of plasmids. The plasmid sequencing was performed for further confirmation of correct cloning. The alignment of sequencing results with the plasmid map of all constructs is shown in Figure 4.

RBM	Forward 5'-3'	Reverse 5'-3'
CXCR4	5'-TCTATCATATGATGAAAAAAACGGCAATC-3'	5'-ATCGTCTCGAGACCAGAACCAAGCTTCCG-3'
1P	5'-TCAGTAGAATTCCGTTCTCAAAGCCGCAGCCG-3'	5'-TACAAAGCTTGTAATGCCTAGGGCTGCGACGACG-3'
2P	5'-CGTCAGAAGCTTCGTTCTCAAAGCCGCAG-3'	5'-CATGAACTCGAGGCTGCGACGACGACGA-3'

Table 1. PCR primers for cloning CXCR4_scFv and CXCR4_RBM fusion proteins

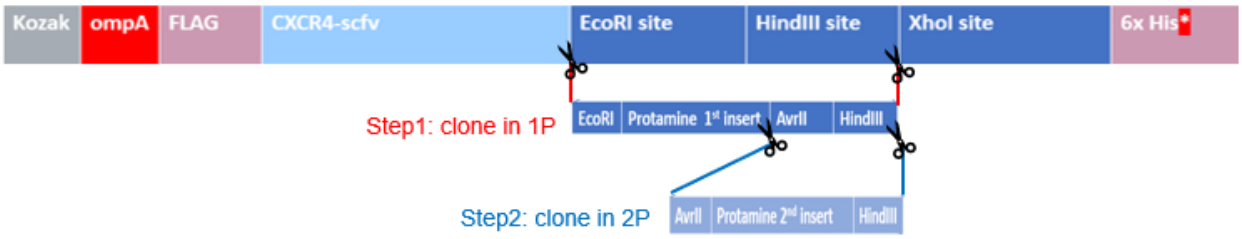

Figure 1. Schematic of cloning for CXCR4_RBM constructs

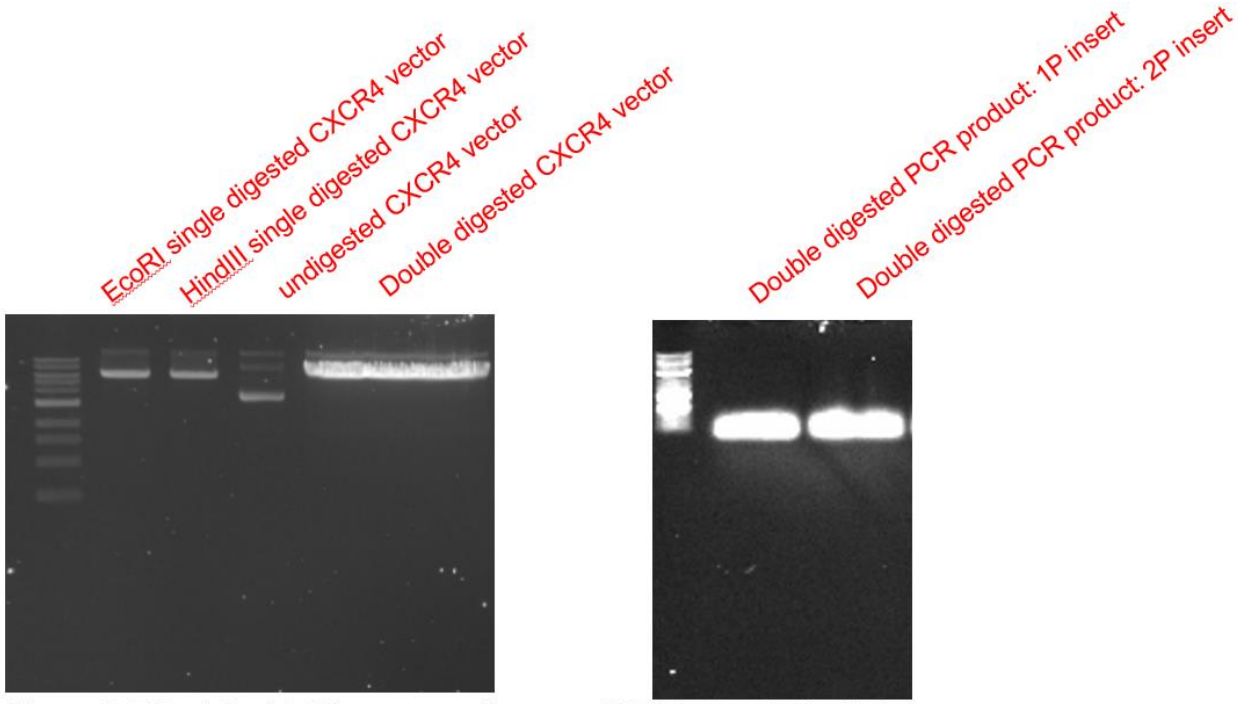

Figure 2. EtBr-stained 0.8% agarose gel images of DNA fragments for cloning

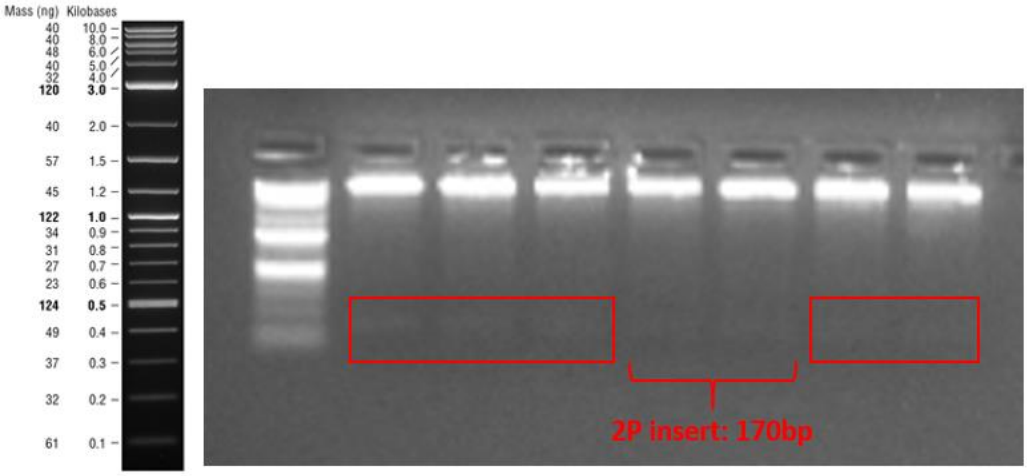

Figure 3. Test digestion image for CXCR4_2P

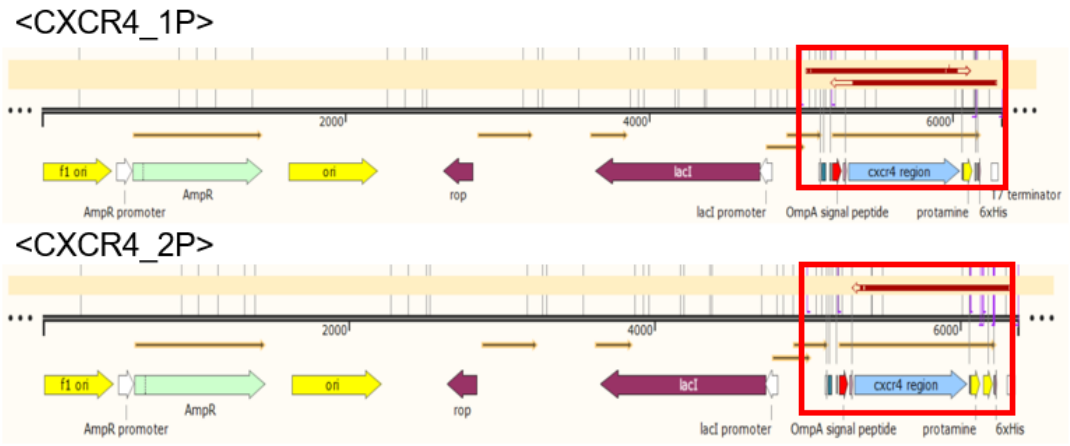

Figure 4. The alignment of sequencing results and their corresponding plasmid map

Protein expression and purification

Plasmids encoding for CXCR4, CXCR4_1P, CXCR4_2P were transformed into a BL21 strain for protein expression. CXCR4 (MW = 32 kDa), CXCR4_1P (MW = 35 kDa), CXCR4_2P (MW = 38kDa) were isolated from E. coli bacterial culture using either inclusion body isolation method (Fig. 5A) or periplasmic extraction (SDS-PAGE images are not shown). Because the CXCR4 fusion protein construct contains 6x His tag, the protein of interest can be purified using an immobilized metal affinity column. The single band at the correct molecular weight position on the SDS-PAGE image indicated the successful isolation of the protein of interest. The fusion proteins with one or two RBMs (CXCR4_1P, and CXCR4_2P) displayed a slightly higher shift in migration due to an increase in molecular weight. The design of the fusion proteins construct was shown in Figure 5C.

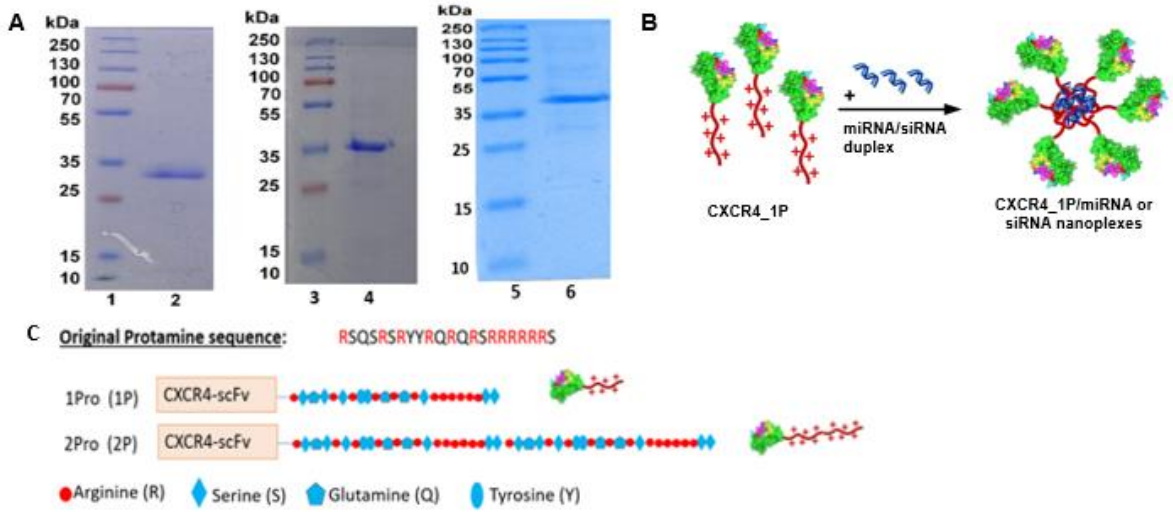

Figure 5. (A) 12% SDS-PAGE of purified CXCR4-scFv (MW approx. 32.3 kDa), CXCR4_1P (MW approx. 35kDa), CXCR4_2P (MW approx. 38 kDa): ladder (lane 1,3,5), CXCR4-scFv (lane 2), CXCR4_1P (lane 4), CXCR4_2P (lane 6); **(B)** Schematics of CXCR4_1P/miRNA nanoplexes; **(C)** Design of alternative CXCR4_RBM fusion proteins, including CXCR4_1P, and CXCR4_2P.

Comparison of two isolation methods

When bacteria cells are expressing proteins, produced proteins can bind to bacteria cellular nucleic acids, impairing the complexation efficiency to therapeutic miRNAs or siRNAs. Besides the bound nucleic acids, free cellular nucleic acids will also be released if the isolation method requires cell lysis (e.g., inclusion body isolation). As shown in Figure 6B, proteins isolated using the inclusion body method showed substantial nucleic acid contamination on the agarose gel. When complexing the proteins to the miRNA, the RiboGreen assay showed an increase in fluorescent signal, which mainly comes from the free nucleic acid contamination present in the protein sample, as shown in Figure 6A. On the other hand, when proteins were isolated from the periplasmic space, they ended up with much fewer contaminants. Periplasmic extraction has several advantages over inclusion body isolation, such as less protease activity and the production of more stable and correctly folded proteins [15]. However, periplasmic extracted proteins resulted in a much lower yield compared to proteins isolated from the inclusion

body. For a liter of culture media, the inclusion body could generate about 200~500 ug of fusion proteins, whereas periplasmic extraction only generates roughly about 30 ugs. Moreover, this isolation method did not completely resolve the contamination issue. As shown in Figures 6C and 6D, the agarose gel image and RiboGreen assay illustrated that there are nucleic acid contaminants in proteins that are periplasmically extracted. The fluorescent signal magnitudes in Figures 6A and 6C were different because they were measured on two different plate readers, so a direct comparison is not possible in this case. However, when comparing Figures 6B and 6D, it is clear that periplasmic extraction resulted in much fewer nucleic acid contaminants in the eluted protein sample.

The association of CXCR4_RBM proteins with the contaminating bacterial nucleic acid would significantly limit the binding of therapeutic small RNAs. Therefore, to recover functional RNA-binding CXCR4_RBM proteins, not only do we need to remove the free nucleic acids, but also dissociate the bound nucleic acids from the protein. Therefore, additional purification steps are required for the CXCR4_RBM fusion protein.

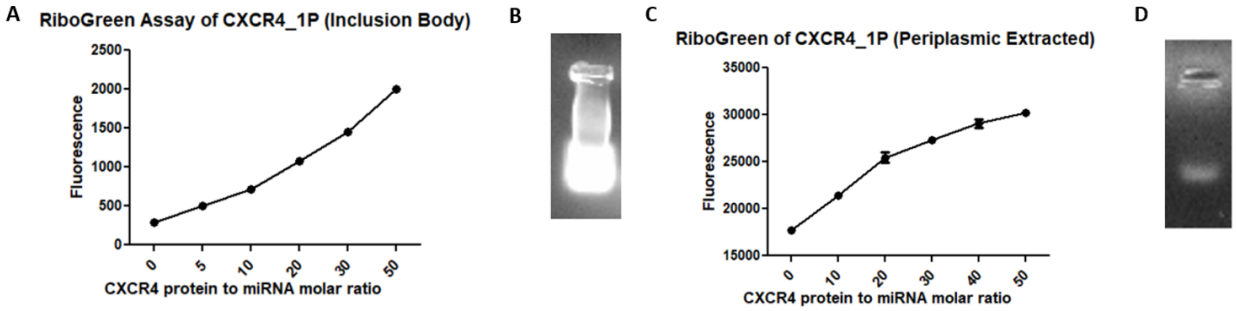

Figure 6. **(A)** RiboGreen assay of CXCR4_1P (Inclusion body) complexed with 20ng (0.14pmol) miRNA duplex; **(B)** Eluted protein (Inclusion body) loaded on 1% agarose gel stained with EtBr; **(C)** RiboGreen assay of CXCR4_1P (Periplasmic extracted) complexed with 20ng (0.14pmol) miRNA duplex; **(D)** Eluted protein (Periplasmic extracted) loaded on 1% agarose gel stained with EtBr.

Anion exchange column

We first assessed if the nucleic acid contaminants can be removed from the CXCR4_1P using anion exchange purification. By increasing the pH to 12, the bound nucleic acids are expected to dissociate from the protein, because due to its isoelectric point of 9.7, the proteins are no longer positively charged at high pH. The more negatively charged nucleic acids bound to the resin more tightly than the CXCR4-1P fusion proteins, and thus the protein eluted at lower concentration buffer, whereas the contaminants eluted at high concentration buffer (1M NaCl) shown in Figure 7. Based on the RiboGreen assay result (Figure 8), CXCR4-1P without anion-exchange treatment had an increased fluorescent signal (red line), indicating the original protein has nucleic acid contamination. While the CXCR4-1P was treated with the anion-exchange column, it (blue line) showed a certain level of complexation with miRNA duplex, but the maximum complexation was only about 20~30% at the highest protein to miRNA molar ratio.

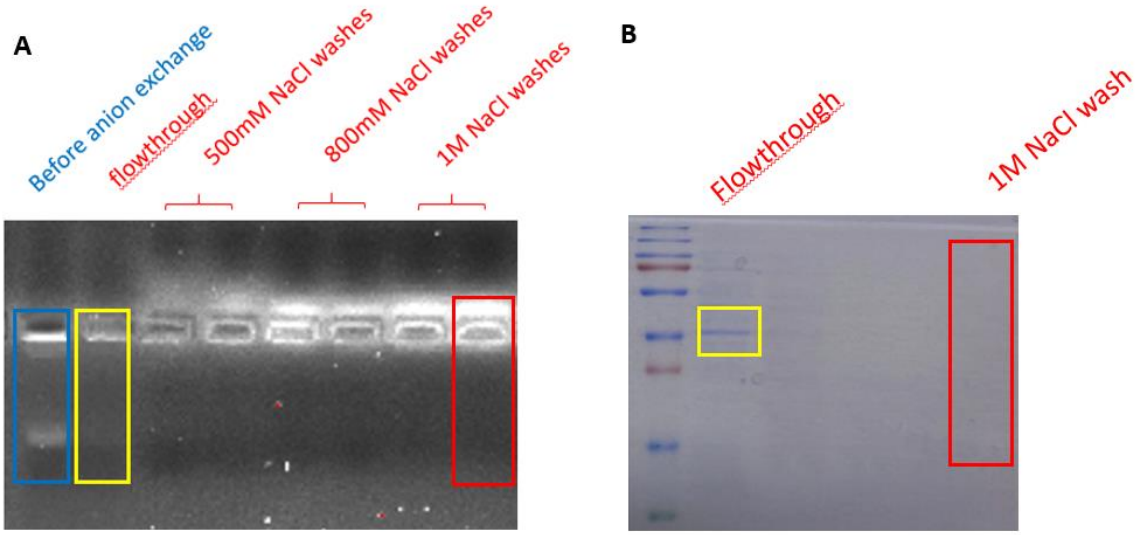

Figure 7. **(A)** EtBr-stained agarose gel images of fractions collected from anion-exchange column; **(B)** SDS-PAGE of fractions collected from anion-exchange column

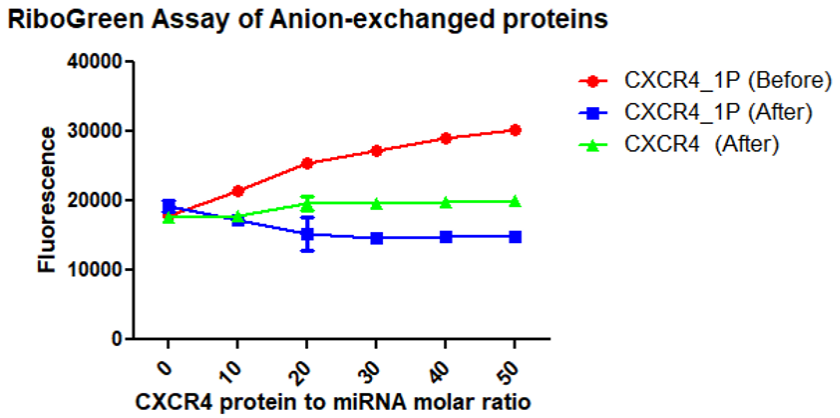

Figure 8. RiboGreen assay result of anion-exchanged-column treated CXCR4 fusion proteins complexed with 20ng (0.14pmol) miRNA duplex.

High-salt-wash assisted IMAC purification

It is well known that protein-RNA interactions are largely electrostatic in nature and thus are sensitive to high salt concentrations. Since the dissociation of nucleic acids from proteins is dependent on salt concentration, high salt (1.5M NaCl) with extensive washing steps during affinity purification could theoretically remove the dissociated contaminants from the protein. After capturing the protein of interest by cobalt beads, the beads were then incubated with 1.5 M NaCl buffer to mediate nucleic acid dissociation. At such a high salt concentration, the ions will weaken the binding between the fusion protein and the impurity by interrupting the electrostatic interactions. The dissociated nucleic acids were washed off the column, leaving pure protein bound to cobalt beads, which were then eluted. From the agarose gel image (Fig 9), we can clearly see that nucleic acid contaminants were removed successfully by the high-salt washes, while the protein of interest eluted at the final step without significant loss.

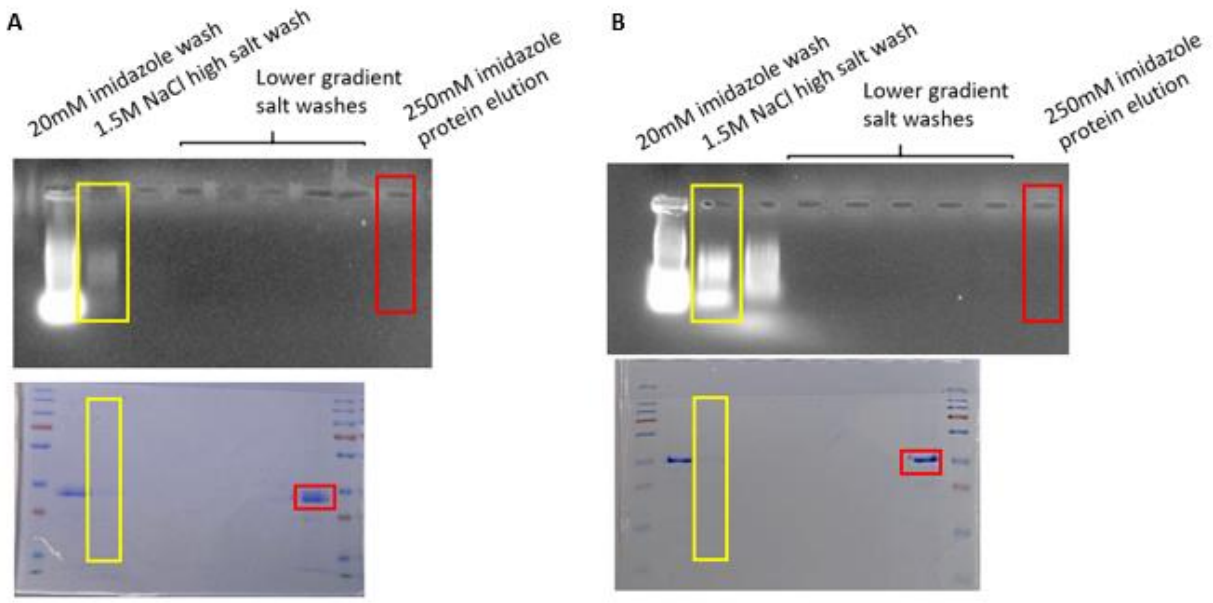

Figure 9. (**A**) EtBr-stained agarose gel (top) and SDS-PAGE (bottom) of fractions collected from high salt wash for CXCR4_only; (**B**) EtBr-stained agarose gel (top) and SDS-PAGE (bottom) of fractions collected from high salt wash for CXCR4_1P

After complexing these proteins with 20 ng miRNA, the RiboGreen assay (Fig. 10) showed that CXCR4_1P complexed about 50% of miRNA at as low as 1:10 pmol molar ratio (protein to miRNA), and then gradually plateaued out. The maximum complexation efficiency is about 55%, which was an improvement compared to all previous approaches. However, the maximum complexation efficiency is still not 100%. One possible explanation would be that the complexation between CXCR4_1P and miRNA was not sufficiently tight, and the RiboGreen dye molecule was still able to diffuse into the complex, intercalate with the RNA, and generate a fluorescent signal[16]. Therefore, there is a need to further improve the RNA binding/loading efficiency. The different CXCR4_RBM constructs were designed to characterize the length of RBM regarding its ability to complex and functionally deliver small RNAs. To figure out the maximum complexation, a single-stranded miRNA could be incorporated as a better control. Since the RiboGreen dye will not bind to the single-stranded nucleic acids, the

fluorescence signal generated from this control group could be considered as the theoretical maximum.

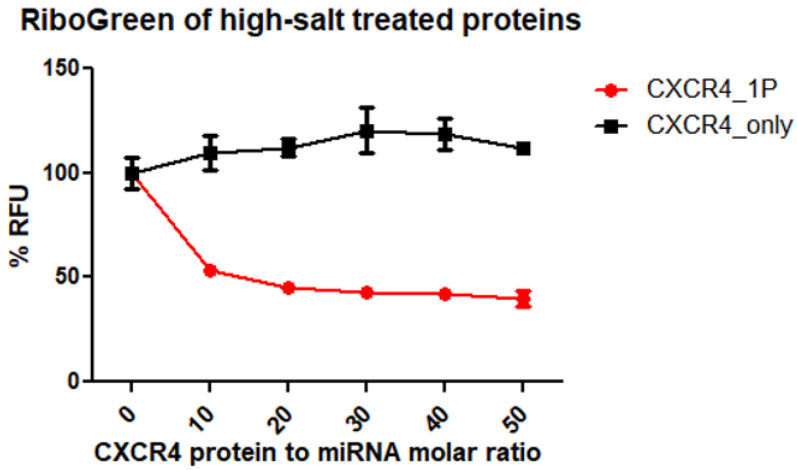

Figure 10. RiboGreen assay of high-salt treated CXCR4_only and CXCR4_1P complexed with 20ng (0.14pmol) miRNA duplex.

ELISA-based binding assay

To confirm that the purified proteins are still able to bind and target the CXCR4 receptor, an ELISA-based binding assay was performed. In Figure 11, the ELISA result showed successful binding of CXCR4-1P fusion proteins to the N-terminus of the CXCR4 receptor, with a K_D value of 104 ± 61 nM. The CXCR4-scFv without RBM is also bound to the N-terminus of the CXCR4 receptor, with a K_D value of 25 ± 5.9 nM.

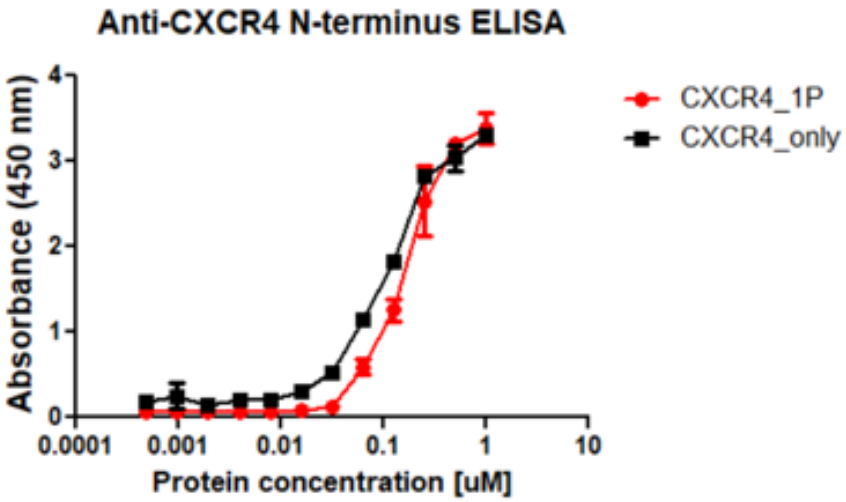

Figure 11. ELISA-based binding assay against N-terminus of CXCR4 receptor

Luciferase assay

To assess if the CXCR4_1P can deliver siRNA into cells, we performed a luciferase gene silencing assay. Luciferase-expressing MDA-MB-231 cells were treated with either CXCR4/siRNA or CXCR4-1P/siRNA complexes with different protein to siRNA molar ratios. The non-specific siRNA duplex (NS_siRNA) was complexed with fusion proteins as a negative control to account for cytotoxicity or off-target effects. A fixed amount of siRNA, 20 ng (0.14 pmol) was used in all experiments, and all the data were expressed as a percentage of control (naked siRNA treated cells). Based on Figure 12A, when cells were treated with CXCR4/siRNA complexes, the luminescence was not significantly affected. When cells were treated with CXCR4_1P/anti-Luc siRNA, the gene was knocked down about 10%~15% starting at 1:20 (protein:siRNA) molar ratio, indicating successful cellular delivery of siRNA (Fig. 12B). At 1:50 (protein:siRNA) molar ratio, the luminescence was decreased compared to the control group. The knockdown efficiency was low, which could be due to 1) low siRNA dosage (0.14 pmol),

2) low CXCR4_1P complexation efficiency, or 3) insufficient endosomal escape. Further experiments are needed to find the optimal dosage of siRNA to achieve an adequate luciferase knockdown with minor or no cytotoxicity, and also additional optimization of the RBM is needed. Another possible reason could be the endosomal entrapment, which not only prevents siRNA from reaching the cytosol but also degrades siRNA. Strategies aimed to overcome this issue includes lipid conjugation, fusogenic peptides, and chemical agents (e.g. chloroquine)[17,18]. One study conducted by Wickline et.al, has demonstrated that a melittin-based strategy could disrupt the membrane and enable safe and effective endosomolysis[19]. For future studies, we could incorporate endosomal escape agents, such as melittin, to enhance the endosomal escape of the CXCR4_siRNA complexes.

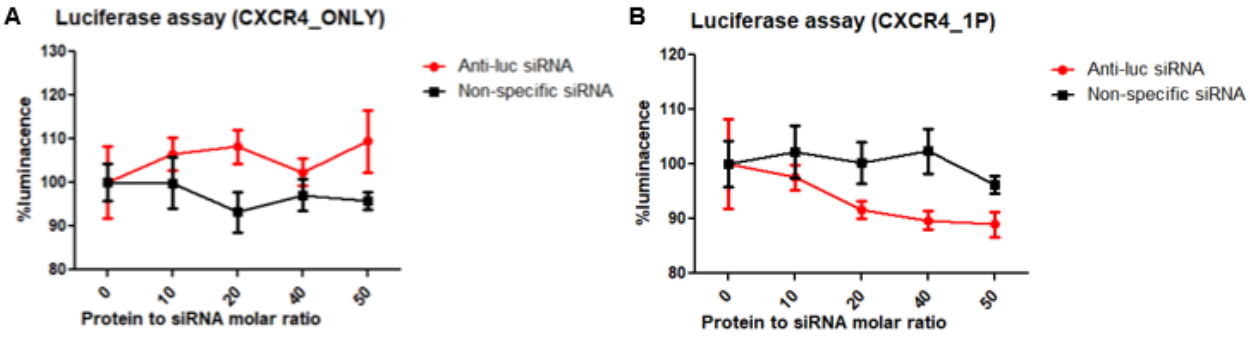

Figure 12. Luciferase assay result of luciferase-expressing MDA cells treated with (**A**) CXCR4_only complexed with anti-luc siRNA or non-specific siRNA. (**B**) CXCR4_1P complexed with anti-luc siRNA or non-specific siRNA. Data are normalized with cells treated with naked siRNA. Data are presented as mean ± sd (n=3).

Computational molecular docking of RBM/miRNA_Duplex complexes

The CXCR4-1P contains a single short fragment of the RNA-binding element of the protamine (RSQSRSRYYRQRQRSRRRRRRS). The single protamine peptide is only 7 nm in length if we assume it adopts an alpha-helix confirmation, while the overall length of the small RNA duplex (22 bp) is about 13 nm in length. By molecular docking,

we can visually observe the size difference between the small RNA duplex and the protamine sequences. Figure 13 shows that one protamine sequence only binds the RNA duplex partially, while the 3-protamine sequence displays more contacts with the RNA duplex. One study conducted by Dr. Jordina Guillén-Boixet showed that the extent of protein/RNA condensation could be controlled by the valence of arginine residues in the RNA-binding regions. His study concluded that the higher the number of arginine residues, the better the complexation of the RNA [20]. The longer RBM with higher charge density may improve protein/RNA complexation and thus provide more protection from degradation.

Due to the complexity of molecular docking methodologies and algorithms, the following docking was simulated in an extreme case, where only one molecule of protamine was complexed with one molecule of miRNA duplex. The simulation only illustrated the protamine peptide, but not CXCR4_RBM fusion proteins. With additional CXCR4-scFv fused with the protamine, the acutual contacting region would be alternated. The molecular docking was only aimed to visualize the length difference, further advanced simulation is needed for a more accurate prediction of the binding molar ratio between the fusion protein and miRNA duplex.

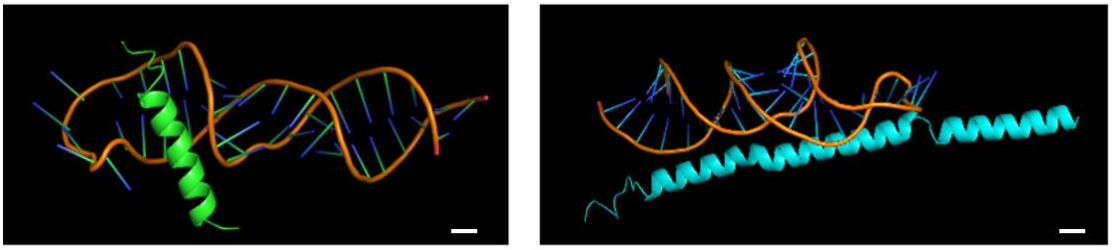

Figure 13. Model of miRNA duplex ducking with different length of protamine sequences. (Scale bar:1nm)

Conclusions

CXCR4-RBM fusion proteins that are overexpressed in *E. coli* tend to bind endogenous RNAs with high affinity due to their RNA-binding domain (e.g., protamine sequence) featuring a positively charged stretch of arginines. Nucleic acid contaminations are an inherent problem shared among most RNA-binding and positively charged protein purification[8,21]. The contaminating nucleic acids occupied the RNA-binding domain, which limited binding and complexation between therapeutic RNA and the fusion proteins. Different isolation methods could improve the nucleic acid contaminants. Unlike inclusion body isolation, the periplasmic extraction could prevent most endogenous nucleic acids released from the bacterial cell. However, the low protein yield became another limiting step for downstream experiments.

Studies have reported a variety of nucleic-acid removal approaches such as ethanol precipitation from crude cell extracts [22,23], heparin-sepharose affinity chromatography [24], precipitation with the polycation PEI [21], and so on. In this study, we tested several nucleic acid removal methods, such as the anion exchange column. PEI precipitation, and others (see in Appendix). The anion exchange method could remove partially bound nucleic acids but failed to economically produce large amounts of functional proteins required for structural and other biophysical studies. The extreme pH conditions, additional incubation time, and washing steps may denature the protein, making this method unpredictable and unreliable without further structure-specific characterization.

As an alternative route, high-salt washes were implemented during the IMAC affinity column purification. Under physiological conditions, at pH 7.4, the CXCR4-1P

(isoelectric point = 9.71) and nucleic acids are fully ionized. Because the electrostatic interactions contribute to the stability of protein/nucleic acid complexes, by adding additional salt, we neutralized the electrostatic interactions and thus dissociated nucleic acids from the proteins. As a result, the high-salt wash method successfully improved the removal of most bound nucleic acids from the protein, leaving relatively pure functional RNA-binding CXCR4 fusion proteins.

The purified fusion proteins retained their shape and ability to bind the specific receptor. They have been shown to deliver siRNA successfully as well, albeit efficiency was low. Future studies will be focusing on structural optimization of the RBMs for enhanced small RNA delivery. Possible strategies include increasing RBM length, the position of charged amino acids, and incorporating aromatic amino acids, such as tyrosine, that can effectively mediate the π-π interactions with the RNA backbone and enhanced nucleic acid/RBM interactions. Additionally, there may be a need to incorporate endosomal escape agents, such as melittin, into the scFv-RBM fusion proteins for enhanced gene silencing. These studies could provide a framework for the design of efficient RBMs that can be combined with other materials used for nucleic acid delivery.

APPENDIX: UNSUCCESSFUL BACTERIAL-CONTAMINATION-REMOVAL APPROACHES

I. Polyethyleneimine (PEI) precipitation

The protein was isolated following the inclusion body isolation. After eluting proteins from the IMAC column, the fractions containing the protein of interest were collected, pooled, and the ionic strength of the solution was increased to 1 M sodium chloride. PEI was added to the final concentration of 0.5~0.8% (w/v) and the solution was incubated for 1 h. The resulting suspension was centrifuged at 15000*g* for 20 min to remove PEI-nucleic acid complexes. The protein was recovered from the excess of PEI present in the supernatant by precipitation with 75% ammonium sulfate. After overnight incubation, this suspension was centrifuged, and the pellet containing the protein was resuspended in binding buffer (50 mM sodium phosphate, pH 7.4, 200 mM sodium chloride, 1 mM DTT, 0.02% sodium azide), applied onto a cobalt-beads-packed column, and purified with the elution buffer. Protein samples were concentrated using an Amicon Ultra-15 (Millipore) with a 10 kDa MWCO membrane. Residual traces of imidazole were removed by dialyzing the eluate at 4 °C against PBS buffer containing 10% glycerol.

Protein isolated using this purification method was then complexed with the miRNA duplex for RiboGreen assay. Comparing to the increased curve shown in Figure 6A, the flat curve of CXCR4_1P in Figure 14 indicated a successful removal of free unbound nucleic acids in the protein sample. However, the PEI precipitation method was not able to remove the bound contaminants from the protein, leaving the CXCR4_1P with no complexation activity. Another problem of the PEI precipitation method is that any trace amount of PEI that remained in the protein could dominate the

RNA complexation, and thus interfering with the downstream miRNA complexation assay.

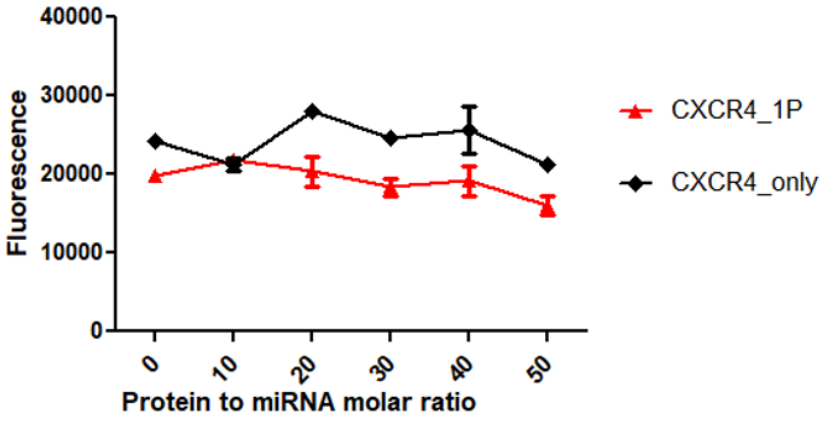

Figure 14. RiboGreen assay of PEI treated CXCR4_only and CXCR4_1P

II. Nuclease enzyme treatment

Following the inclusion body isolation method, the bacteria cell pellet was resuspended in lysis buffer containing 25 units/ml benzonase nuclease (EMD Millipore). Comparing to the bacterial cell suspension in normal lysis buffer, with additional benzonase nuclease, the viscosity was significantly reduced, indicating successful digestion of free nucleic acids contaminants. However, the eluted proteins still contain nucleic acid contaminants, shown in Figure 15. One explanation could be that the nuclease was unable to digest the bound nucleic acids because they were shielded or protected by the proteins. Additional pre-treatment is required to dissociate the bound nucleic acid before enzyme digestion. More importantly, this method requires successful removal of nuclease enzyme, which could digest miRNA of interest in the downstream assays. An alternative approach would be implementing an immobilized nuclease-packed column chromatography, avoiding nuclease residue in the final protein elution.

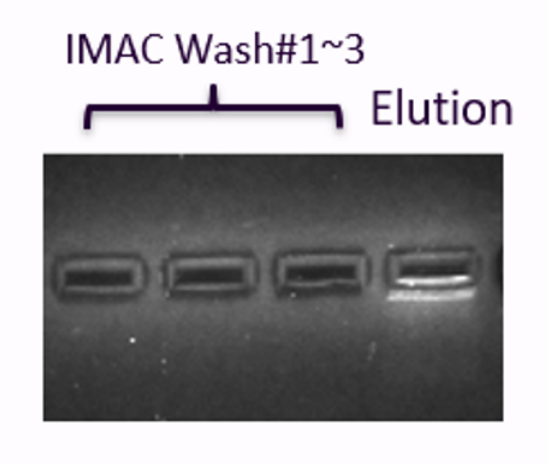

Figure 15. Et-Br stained agarose gel image of fractions collected during IMAC column purification for nuclease treated CXCR4_1P protein

REFERENCES

1. Zhang C. Novel functions for small RNA molecules. *Curr Opin Mol Ther*. 2009;11(6):641-651. /pmc/articles/PMC3593927/?report=abstract. Accessed December 21, 2020.
2. Kristen A V., Ajroud-Driss S, Conceição I, Gorevic P, Kyriakides T, Obici L. Patisiran, an RNAi therapeutic for the treatment of hereditary transthyretin-mediated amyloidosis. *Neurodegener Dis Manag*. 2019;9(1):5-23. doi:10.2217/nmt-2018-0033
3. Sardh E, Harper P, Balwani M, et al. Phase 1 Trial of an RNA Interference Therapy for Acute Intermittent Porphyria. *N Engl J Med*. 2019;380(6):549-558. doi:10.1056/nejmoa1807838
4. De Gooijer CJ, Baas P, Burgers JA. Current chemotherapy strategies in malignant pleural mesothelioma. *Transl Lung Cancer Res*. 2018;7(5):574-583. doi:10.21037/tlcr.2018.04.10
5. van Zandwijk N, Pavlakis N, Kao SC, et al. Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *Lancet Oncol*. 2017;18(10):1386-1396. doi:10.1016/S1470-2045(17)30621-6
6. Sahin U, Karikó K, Türeci Ö. MRNA-based therapeutics-developing a new class of drugs. *Nat Rev Drug Discov*. 2014;13(10):759-780. doi:10.1038/nrd4278
7. Lieberman J. Tapping the RNA world for therapeutics. *Nat Struct Mol Biol*. 2018;25(5):357-364. doi:10.1038/s41594-018-0054-4
8. Zang C, Liu Z, Yang K, et al. Expression, purification and biological effect of a novel single chain Fv antibody and protamine fusion protein for the targeted delivery of siRNAs to FGFR3 positive cancer cells. *Electron J Biotechnol*. 2017;28:14-19. doi:10.1016/j.ejbt.2017.05.009
9. Jiang K, Li J, Yin J, et al. Targeted delivery of CXCR4-siRNA by scFv for HER2+ breast cancer therapy. *Biomaterials*. 2015;59:77-87. doi:10.1016/j.biomaterials.2015.04.030
10. Deci MB, Liu M, Gonya J, et al. Carrier-Free CXCR4-Targeted Nanoplexes Designed for Polarizing Macrophages to Suppress Tumor Growth. *Cell Mol Bioeng*. 2019;12(5):375-388. doi:10.1007/s12195-019-00589-w
11. O'Day E, Le MTN, Imai S, et al. An RNA-binding protein, Lin28, recognizes and remodels G-quartets in the MicroRNAs (miRNAs) and mRNAs it regulates. *J Biol Chem*. 2015;290(29):17909-17922. doi:10.1074/jbc.M115.665521
12. Juibari AD, Ramezani S, Rezadoust MH. Bioinformatics analysis of various signal peptides for periplasmic expression of parathyroid hormone in E.coli. *J Med Life*. 2019;12(2):184-191. doi:10.25122/jml-2018-0049

13. Ayala M, Balint R, Fernández-de-Cossío L, Canaán-Haden JW, Larrick J, Gavilondo J. Variable region sequence modulates periplasmic export of a single-chain Fv antibody fragment in Escherichia coli. *undefined*. 1995.

14. Karyolaimos A, Ampah-Korsah H, Hillenaar T, et al. Enhancing Recombinant Protein Yields in the E. coli Periplasm by Combining Signal Peptide and Production Rate Screening. *Front Microbiol*. 2019;10:1511. doi:10.3389/fmicb.2019.01511

15. Cornelis P. Expressing genes in different Escherichia coli compartments. *Curr Opin Biotechnol*. 2000;11(5):450-454. doi:10.1016/S0958-1669(00)00131-2

16. Shoute LCT, Loppnow GR. Characterization of the binding interactions between EvaGreen dye and dsDNA. *Phys Chem Chem Phys*. 2018;20(7):4772-4780. doi:10.1039/c7cp06058k

17. Erazo-Oliveras A, Muthukrishnan N, Baker R, Wang TY, Pellois JP. Improving the endosomal escape of cell-penetrating peptides and their cargos: Strategies and challenges. *Pharmaceuticals*. 2012;5(11):1177-1209. doi:10.3390/ph5111177

18. Pei D, Buyanova M. Overcoming Endosomal Entrapment in Drug Delivery. *Bioconjug Chem*. 2019;30(2):273-283. doi:10.1021/acs.bioconjchem.8b00778

19. Hou KK, Pan H, Schlesinger PH, Wickline SA. A role for peptides in overcoming endosomal entrapment in siRNA delivery - A focus on melittin. *Biotechnol Adv*. 2015;33(6):931-940. doi:10.1016/j.biotechadv.2015.05.005

20. Guillén-Boixet J, Kopach A, Holehouse AS, et al. RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell*. 2020;181(2):346-361.e17. doi:10.1016/j.cell.2020.03.049

21. Orsini MJ, Thakur AN, Andrews WW, Hammarskjold ML, Rekosh D. Expression and purification of the HIV type 1 Rev protein produced in Escherichia coli and its use in the generation of monoclonal antibodies. *AIDS Res Hum Retroviruses*. 1995;11(8):945-953. doi:10.1089/aid.1995.11.945

22. Malim MH, Tiley LS, McCarn DF, Rusche JR, Hauber J, Cullen BR. HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence. *Cell*. 1990;60(4):675-683. doi:10.1016/0092-8674(90)90670-A

23. Nalin CM, Purcell RD, Antelman D, et al. Purification and characterization of recombinant Rev protein of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*. 1990;87(19):7593-7597. doi:10.1073/pnas.87.19.7593

24. Heaphy S, Finch JT, Gait MJ, Karn J, Singh M. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich “bubble” located within the rev-responsive region of viral mRNAs. *Proc Natl Acad Sci U S A*. 1991;88(16):7366-7370. doi:10.1073/pnas.88.16.7366