ORTA DOĞU TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ ŞABLONU ONAY FORMU
THESIS TEMPLATE CONFIRMATION FORM

1. Şablonda verilen yerleşim ve boşluklar değiştirilmemelidir.
2. **Jüri tarihi** Başlık Sayfası, İmza Sayfası, Abstract ve Öz'de ilgili yerlere yazılmalıdır.
3. İmza sayfasında jüri üyelerinin unvanları doğru olarak yazılmalıdır. Tüm imzalar **mavi pilot kalemle** atılmalıdır.
4. **Disiplinlerarası** programlarda görevlendirilen öğretim üyeleri için jüri üyeleri kısmında tam zamanlı olarak çalıştıkları anabilim dalı başkanlığının ismi yazılmalıdır. Örneğin: bir öğretim üyesi Biyoteknoloji programında görev yapıyor ve biyoloji bölümünde tam zamanlı çalışıyorsa, İmza sayfasına biyoloji bölümü yazılmalıdır. İstisnai olarak, disiplinler arası program başkanı ve tez danışmanı için disiplinlerarası program adı yazılmalıdır.
5. Tezin **son sayfasının sayfa** numarası Abstract ve Öz'de ilgili yerlere yazılmalıdır.
6. Bütün chapterlar, referanslar, ekler ve CV sağ sayfada başlamalıdır. Bunun için **kesmeler** kullanılmıştır. **Kesmelerin kayması** fazladan boş sayfaların oluşmasına sebep olabilir. Bu gibi durumlarda paragraf (¶) işaretine tıklayarak kesmeleri görünür hale getirin ve yerlerini **kontrol edin**.
7. Figürler ve tablolar kenar boşluklarına taşmamalıdır.
8. Şablonda yorum olarak eklenen uyarılar dikkatle okunmalı ve uygulanmalıdır.
9. Tez yazdırılmadan önce PDF olarak kaydedilmelidir. Şablonda yorum olarak eklenen uyarılar PDF dokümanında yer almamalıdır.
10. **Bu form aracılığıyla oluşturulan PDF dosyası arkalı-önlü baskı alınarak tek bir spiralli cilt haline getirilmelidir.**
11. Spiralli hale getirilen tez taslağınızdaki ilgili alanları imzalandıktan sonra, Tez Juri Atama Formu ile birlikte bölüm sekreterliğine teslim edilmelidir.
12. Tez taslaklarının kontrol işlemleri tamamlandığında, bu durum öğrencilere METU uzantılı öğrenci e-posta adresleri aracılığıyla duyurulacaktır.
13. Tez yazım süreci ile ilgili herhangi bir sıkıntı yaşarsanız, **Sıkça Sorulan Sorular (SSS)** sayfamızı ziyaret ederek yaşadığınız sıkıntıyla ilgili bir çözüm bulabilirsiniz.

1. Do not change the spacing and placement in the template.
2. Write **defense date** to the related places given on Title page, Approval page, Abstract and Öz.
3. Write the titles of the examining committee members correctly on Approval Page. **Blue ink** must be used for all signatures.
4. For faculty members working in **interdisciplinary programs**, the name of the department that they work full-time should be written on the Approval page. For example, if a faculty member staffs in the biotechnology program and works full-time in the biology department, the department of biology should be written on the approval page. Exceptionally, for the interdisciplinary program chair and your thesis supervisor, the interdisciplinary program name should be written.
5. Write **the page number of the last page** in the related places given on Abstract and Öz pages.
6. All chapters, references, appendices and CV must be started on the right page. **Section Breaks** were used for this. **Change in the placement** of section breaks can result in extra blank pages. In such cases, make the section breaks visible by clicking paragraph (¶) mark and **check their position.**
7. All figures and tables must be given inside the page. Nothing must appear in the margins.
8. All the warnings given on the comments section through the thesis template must be read and applied.
9. Save your thesis as pdf and Disable all the comments before taking the printout.
10. **Print two-sided the PDF file that you have created through this form and make a single spiral bound.**
11. Once you have signed the relevant fields in your thesis draft that you spiraled, submit it to the department secretary together with your Thesis Jury Assignment Form.
12. This will be announced to the students via their METU students e-mail addresses when the control of the thesis drafts has been completed.
13. If you have any problems with the thesis writing process, you may visit our **Frequently Asked Questions (FAQ)** page and find a solution to your problem.

☒ **Yukarıda bulunan tüm maddeleri okudum, anladım ve kabul ediyorum. / I have read, understand and accept all of the items above.**

Name : Nooraldeen
Surname : Ayoub
E-Mail : e200006@metu.edu.tr
Date : 03/02/2021
Signature : _____

SYSTEMATIC EVOLUTION OF NOVEL 2′F-PY RNA APTAMERS TARGETING THE MEMBRANE PROTEIN L-ARGININE/AGMATINE ANTIPORTER PURIFIED IN MILD DETERGENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

NOORALDEEN AYOUB

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOCHEMISTRY

FEBRUARY 2021

Approval of the thesis:

SYSTEMATIC EVOLUTION OF NOVEL 2′F-PY RNA APTAMERS TARGETING THE MEMBRANE PROTEIN L-ARGININE/AGMATINE ANTIPORTER PURIFIED IN MILD DETERGENT

submitted by **NOORALDEEN AYOUB** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Biochemistry, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Assoc. Prof. Dr. Yeşim Soyer
Head of the Department, **Biochemistry** _____

Assoc. Prof. Dr. Çağdaş Devrim Son
Supervisor, **Biochemistry, METU** _____

Dr. Müslüm İlgü
Co-Supervisor, **Aptalogic Inc.** _____

Examining Committee Members:

Prof. Dr. Ayşe Gül Gözen
Biological Sciences, METU _____

Assoc. Prof. Dr. Çağdaş Devrim Son
Biochemistry, METU _____

Assoc. Prof. Dr. Yeşim Soyer
Biochemistry, METU _____

Assoc. Prof. Dr. Özgül Persil Çetinkol
Chemistry, METU _____

Assoc. Prof. Dr. Eda Çelik Akdur
Chemical Engineering, Hacettepe University _____

Date: 15.02.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Nooraldeen Ayoub

Signature :

ABSTRACT

SYSTEMATIC EVOLUTION OF NOVEL 2′F-PY RNA APTAMERS TARGETING THE MEMBRANE PROTEIN L-ARGININE/AGMATINE ANTIPORTER PURIFIED IN MILD DETERGENT

Ayoub, Nooraldeen
Master of Science, Biochemistry
Supervisor: Assoc. Prof. Dr. Çağdaş Devrim Son
Co-Supervisor: Dr. Müslüm İlgü

February 2021, 257 pages

Bacterial microorganisms have developed numerous and diverse systems to handle potentially detrimental acidic conditions in their external environment. In particular, some pathogenic and other nonpathogenic enteric bacteria have a number of exceptional and well-organized acid resistance (AR) mechanisms that work together to counter intracellular acidification and damage and enable survival under the extreme acidic conditions of the mammalian stomach. The common *Escherichia coli* with both its virulent and benign strains is particularly remarkable considering its adaptation to neutral pH. In these bacteria, the challenge of a low external pH is efficiently mitigated by highly potent proton consumption AR systems that consist of two generic components: a cytoplasmic pyrodoxal-5′-phoshpate (PLP)-dependent amino acid decarboxylase and an inner plasma membrane amino acid antiporter. Both, decarboxylation of a substrate amino acid into a product and CO_2 in a proton-dependent PLP-catalyzed manner and antiporter-mediated exchange of the internally generated decarboxylation products with the externally located substrate amino acid maintain the system running and aid in restoring pH homeostasis of the cell. One of these systems is the arginine-dependent acid resistance (ADAR) system constituted

by acid-induced arginine decarboxylase (AdiA) and arginine-agmatine antiporter (AdiC).

The impressive competence of AR systems and especially those used by infectious, disease-causing enteric bacteria, comes at a big cost—the well-being and health of humans. Henceforth, in this work, for the first time, we raised five candidate 2′FY RNA candidate aptamers (noorA, noorB, noorC, AdiC19 and AdiC72) against the ADAR system's integral membrane protein (MP), AdiC. Briefly, AdiC with a 10× Histidine tag on the C-terminal was recombinantly overexpressed using the convenient prokaryotic expression system in BL21(DE3) pLysS cell line and was extracted by solubilizing using n-dodecyl-β-D-maltopyranoside (DDM) and purified by metal affinity chromatography in the same detergent. After confirming the purity by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis), recombinant AdiC was subjected to 8 rounds of conventional Systematic Evolution of Ligands by EXponential enrichment (SELEX) for the *in vitro* selection of 2′F-Pyrimidine-modified candidate RNA aptamers. Oligonucleotide sequences from the final library of selection were cloned and identified by Sanger sequencing. Three sequences (noorA noorB and noorC) were found abundant among 100 clones and thus chosen as the basis for subsequent informatics analyses. Using a computational approach with comparative analyses of primary and optimal (MFE) 2D secondary structures predicted by *RNAfold* and *Kinefold* programs, we were able to identify three conserved motifs (one of which we annotated as a putative local binding domain and the other two as stabilizers) in noorA, B and C and also in two more sequences (AdiC19 and AdiC72) identified in a different phylogenetic family. We also defined a consensus secondary structure to which these five proto-aptamer conform. Subsequent prediction-driven (*in-silico*) mutational deletions of the stabilizer motifs revealed a global impact on several of the predicted MFE structures in maintaining the substructure that harbors the putative binding motif.

The five aptamer candidates generated in this work may have prospects as versatile agents applied in areas of aptamer-based diagnostics and therapeutics where

they may be used to assess bacterial infections in patients (e.g. by aptasensors) or as antimicrobial pharmaceutical drugs or drug delivery agents for these infections. At the level of basic scientific research, these aptamers may also be useful in further understanding the structural features and biochemical functions of AdiC. More broadly, given the scarcity of SELEX trials against purified full-length MPs over the past ~20 years, this thesis presents itself as one of the proof-of-principle studies that bring attention to the usability of purified full-length MPs for the selection of aptamers after MP reconstitution in detergent micelles. This advocacy is in line with the advantages provided by conventional protein-SELEX over cell-SELEX and utility of soluble intra- and extracellular MP domains used as targets. At the end of this thesis, we direct the spotlight on a special opportunity available for complex MPs such as receptors, transporters, and channels that may often be inaccessible for aptamer targeting due to multi-passing or deep integration to and intimacy with their membranes.

Keywords: AdiC, DDM, SELEX, Aptamer, Secondary Structure Prediction

ÖZ

HAFİF DETERJANDA SAFLAŞTIRILMIŞ L-ARJİNİN/AGMATİN ANTİPORT MEMBRAN PROTEİNİNİ HEDEFLEYEN YENİ 2′F-PY RNA APTAMERLERİNİN SİSTEMATİK EVRİMİ

Ayoub, Nooraldeen
Yüksek Lisans, Biyokimya
Tez Yöneticisi: Doç. Dr. Çağdaş Devrim Son
Ortak Tez Yöneticisi: Dr. Müslüm İlgü

Şubat 2021, 257 sayfa

Bakteriyel mikroorganizmalar, dış ortamlarındaki potansiyel olarak zararlı asidik koşulların üstesinden gelebilmek için çok sayıda ve çeşitli sistemler geliştirmiştir. Özellikle, bazı patojenik ve diğer patojenik olmayan enterik bakteriler, hücre içi asidifikasyona ve memeli midesinin aşırı asidik koşulları altında hayatta kalmayı mümkün kılan hasara karşı koyabilmek için birlikte çalışan bir dizi istisnai ve iyi organize edilmiş asit direnci (AR) mekanizmasına sahiptir. Hem virülan hem de iyi huylu suşları ile ortak Escherichia coli, nötr pH'a adaptasyonu göz önüne alındığında özellikle dikkat çekicidir. Bu bakterilerde, düşük dış pH baskısı, bir sitoplazmik pirodoksal-5′-fosfat (PLP) bağımlı amino asit dekarboksilaz ve bir iç plazma membran amino asidi antiporter jenerik bileşenden oluşan oldukça güçlü proton tüketimli AR sistemleri aracılığıyla etkili bir şekilde hafifletilir. Bir substrat amino asidinin dekarboksilasyonu sonrasında bir ürüne ve CO2'ye dönüştürülmesi protona bağımlı PLP katalizli bir şekilde elde edilir. Daha sonra, dahili olarak üretilen dekarboksilasyon ürününün hem antiporter aracılı değişimi hem de harici olarak yerleştirilmiş substrat amino asidi, sistemin çalışmasını sağlar ve hücrenin pH

homeostazının geri kazanılmasına yardımcı olur. Bu sistemlerden biri, AdiA ve AdiC tarafından oluşturulan arginin bağımlı asit direnci (ADAR) sistemidir: sırasıyla indüklenebilir arginin dekarboksilaz enzimi ve arginin / agmatin antiportörü.

AR sistemlerinin etkileyici yeterliliği ve özellikle bulaşıcı, hastalığa neden olan enterik bakteriler tarafından kullanılanlar, büyük bir maliyete- insanların refahına ve sağlığına – sebep olmaktadır. Bundan böyle, bu çalışmada, ve ilk kez, ADAR sisteminin integral membran proteini (MP) AdiC'ye karşı beş 2'FY RNA aday aptamerini (noorA, noorB, noorC, AdiC19 ve AdiC72) seçilimini gerçekleştirdik. Kısaca, AdiC, BL21 (DE3) pLysS hücre hattında uygun prokaryotik ekspresyon sistemi kullanılarak rekombinant olarak eksprese edildi ve his-etiketli AdiC, daha sonra n-dodesil-β-D-maltopiranosid (DDM) kullanılarak çözündürme yoluyla ekstrakte edildi ve aynı deterjanda metal afinite kromatografisiyle saflaştırıldı. SDS-PAGE (sodyum dodesil sülfat poliakrilamid jel elektroforezi) ile varlığı onaylandıktan sonra, saf rekombinant AdiC, 2'F-Pirimidin ile modifiye edilmiş aday RNA aptamerlerinin in vitro seçimi için Eksponansiyel zenginleştirme (SELEX) ile 8 tur geleneksel Ligandların Sistematik Evrimine tabi tutuldu. Son olarak elde edilen seçim kütüphanesinden oligonükleotid sekansları klonlandı ve Sanger sekanslaması ile sekansları belirlendi. Üç dizi (noorA noorB ve noorC) 100 klon arasında bol miktarda bulundu ve bu nedenle sonraki bilişim analizleri için temel olarak seçildi. RNAfold ve Kinefold programları tarafından tahmin edilen birincil ve optimal (MFE) 2D ikincil yapıların karşılaştırmalı analizleri ile hesaplamalı bir yaklaşım kullanarak, üç korunmuş motifi belirleyebildik (bunlardan biri varsayılan yerel bağlanma alanı ve diğer ikisi de stabilizatör olarak açıklandı). Bu motifler noorA, B ve C'de ve farklı bir filogenetik ailede iki tane daha tanımlanmış dizide (AdiC19 ve AdiC72) belirlendi. Ayrıca, bu beş proto-aptamerin uyduğu bir konsensüs ikincil yapı tanımladık. Stabilizatör motiflerinin sonraki tahmin odaklı (in-siliko) mutasyonel delesyonları, varsayılan bağlanma motifini barındıran alt yapının korunması üzerindeki tahmini MFE yapılarının birçoğunda küresel etkilerini ortaya çıkardı.

Bu çalışmada elde edilen beş aptamer adayı, hastalardaki bakteriyel enfeksiyonları değerlendirmek için (örneğin aptasensörler tarafından) veya antimikrobiyal farmasötik ilaçlar veya ilaç verme ajanları olarak kullanılabilecekleri aptamer bazlı teşhis ve tedavi alanlarında uygulama için çok yönlü ajanlar olarak kullanılma potansiyeline sahiptir. Temel bilimsel araştırma düzeyinde, bu aptamerler, AdiC'nin yapısal özelliklerini ve biyokimyasal işlevlerini daha iyi anlamak için de yararlı olabilir. Daha genel olarak, son 20 yılda saflaştırılmış tam uzunluktaki proteinlere karşı SELEX denemelerinin azlığı göz önüne alındığında, bu tez kendini, saflaştırılmış tam uzunluktaki aptamerlerin kullanılabilirliğine dikkat çeken ilk çalışmalardan biri olarak sunulmaktadır. Bu yöntem, geleneksel protein-SELEX'in hücre-SELEX'e göre sağladığı avantajlarla ve çözülebilir hücre içi ve hücre dışı MP alanlarının hedef olarak kullanımıyla uyumludur. Bu tezin sonunda, çok geçişli veya derin entegrasyon ve membranlar içindeki yakınlık nedeniyle aptamer hedeflemesi için genellikle erişilemeyen reseptörler, taşıyıcılar ve kanallar gibi karmaşık MP'ler için mevcut özel bir fırsat sunmaktadır.

Anahtar Kelimeler: AdiC, DDM, SELEX, Aptamer, İkincil Yapıların Tahmini

To the woman who sacrificed her life to raise me—my mother, Nazik. To the man who believed in me—my father, Fathi. To the ambitious, loving, and supportive best friend—my brother, Mohammed.

ACKNOWLEDGMENTS

Before, amidst and after all else, all praise and glorification be ultimately to The Most Supreme, The Almighty, and The Absolute Truth–Allah. In essence, all my gratitude goes to Him in a destined-to-fail attempt to appreciate His uncountable blessings and unreckonable mercy, for He is truly The One Who guides and The One Who provides. Indeed, all opportunities are distributed fairly by Him as He is The Most Knowledgeable and The All-Wise.

I would like to expand my gratitude to warmly accommodate anyone that aided by any degree in the completion of this research and this thesis. I want to start by Dr. Müslüm İlgü. To put simply, I am grateful to him as he has embodied a strong force that has pushed me towards improvement at different levels. At the level of work, I am deeply thankful for both his theoretical and experimental teachings. He has really taken us by the hand and introduced and guided us through the new field of aptamers. He nurtured our growth in this interesting and exciting field. His lab meetings were of great value for the overall progress and growth of all members of our laboratory as well. On a personal level, I specially want to thank him as I have learnt from him not to get carried away with the vastness of science and to stay focused on the task at hand without setting perfection as the central goal. Atop those deeds and virtues, I want to thank him for his patience as we learn, make mistakes, write, and grow as young scientists. Müslüm hoca was a source of encouragement in this journey. He tries to provide the right stimuli for growth of his students in science. He has also tried to help us excel at mental/emotional levels to optimize our performance inside and outside the lab. I would like to wrap up my appreciation by saying that he was not only an advisor but also a kind/generous friend who can reach to and touch the hearts of all students.

Next, I would like to express a very genuine and special happiness I feel after having met the younger brother of Müslüm hoca, Dr. Hüseyin İlgü. He visited us in Ankara from Switzerland in the fall of 2019 and his period of stay was such a blessed and delightful time. Although his duration of stay was for only a few weeks, it was very fruitful and productive. We have also instantly become friends! I am thankful to have learnt the purification of recombinant membrane proteins with him and thanks to him, my interest in this intriguing class of proteins has grown exponentially! Also, thanks to him I did my own purification trials confidently after he had travelled back to Switzerland. I am grateful for meeting such

a unique and friendly person. His period of stay was an engaging experience not only experimentally but also interpersonally. I am also delighted with all the conversations and debates we had on all the different topics during our free and even work time. I sincerely wish upon him, his family, and his new daughter, Ada, endless blessings, health, and joy. Appreciation extends to encompass Prof. Dr. Dimitrios Fotiadis from the University of Bern and members of his lab for their kind and valuable contribution in the production of AdiC.

Next is my lab partners. I am grateful for all the help and support provided during our studies and for the kindness and happy times we spent. I am thankful especially to Nilüfer Kara, the closest companion and partner in this distressing journey. We've been through many ups and downs. We laughed, teared, fought and came a long way to finally arrive at this point and emerge as winners and not quitters. Despite all the pain we faced, Nilüfer has truly proven herself to be a genuine soul with a sincere and loving nature that was very eye-opening at times and had truly left me speechless. Appreciation continues with dear Meriç Öztürk. Special thanks to this kind, hard-working and helpful man as this lab would not progress easily without his efforts. I am thankful for all the discussions we had. All insights, scientific knowledge and general information that pushed this work towards completion are all appreciated. I also thank him for saving a significant amount of my time by handling many of my English-Turkish and Turkish-English translations. Accomplishment of these needed tasks would have been time-consuming and near-impossible without him given the pressure/time constraints. Finally, I would also like to acknowledge his very funny comments as they sincerely were a good source of laughter. Next is Funda Ardıç; our lively lab partner. In her presence, the lab is more joyful and fun. I also want to acknowledge her high sense of responsibility and organization and her generosity and all the help and support she had provided. Rezzan Fazlıoğlu; the first master's student to graduate from our lab. I am very thankful for your guidance with isothermal titration calorimetry during the first days in the lab and I certainly wish you all the best in your PhD studies and your future undertakings. Finally, I would like to mention the undergraduate attendees of İlgü lab. It was a pleasure to have people such as Rojbin, Batuhan, Mustafa, and Dilara attending our lab. Your presence was light-hearted and guiding you in the lab was also a valuable learning process.

Gratitude would not be complete without recognizing the efforts and hard work of all the members of the department of biological sciences including all assistants and staff.

A special place in the heart holds all of the great faculty members for their restlessness and academic efforts with all students. I am also very thankful to the jury members for their patience as we set up the committee and for their constructive criticisms and suggestions during the thesis defense. Special thanks to Assoc. Prof. Dr. Çağdaş Devrim Son for being the helpful, supportive, and simply, the go-to professor that he is. I am happy with all the discussions he had had with me no matter how divergent they were. His valuable guidance and his personality are both truly heartfelt and appreciated. I am also deeply thankful for his acceptance to be our supervisor at such a short notice. Finally, I would like to take this chance to admiringly acknowledge Prof. Dr. Mesut Muyan for being a truly exceptional professor and amazing person. I simply see him as a true hero of all the students, and he will never be forgotten. From the department of biochemistry, I want to thank Assoc. Prof. Dr. Yeşim Soyer, the department's head. She has been a great and an extremely helpful and supportive person. Her organized meetings were also highly valuable and enjoyable.

As they say, the best is saved for the last. My mother; the true source of light in my life. A complete written book would not come close to conveying your value to me. No matter what I do, I will never be able to repay my indebtedness to you. You are a true warrior and a source of inspiration to all mothers around the world. You are the most affectionate and loving person I know. Despite my many unbearable mistakes, you've always unconditionally loved and supported me like nobody ever does. You are the beauty to my life, and I cannot imagine my life without you. I see my world through you. Your actions inspire me. You reset me on the correct path when I lose vision at my darkest times and drift away from the right track. You help me disregard the unnecessary things in life and focus on the big picture and the brighter aspects of things. Mum, indeed, you are truly unmatched, and I gladly want to dedicate my life to making you happier and happier. May Allah help me make this wish a reality and may he bless you eternally and fulfill all your wishes and dreams. Addressing the dearest people to my heart must inevitably have my dear and only brother, Mohammed, mentioned. My dear Mohammed, you are the greatest and most loving brother one could ever have. I'm glad that today, we are closer than ever. Always, my heart prays for you. I wish you all the best from the bottoms of my heart and may you achieve all your ambitions and dreams. I promise you that we will always walk this life hand-in-hand and make it together through any hardship that dares to cross our way. In the end, may Allah reunite us and our mother again one day with our father in the highest rank of heaven.

PREFACE

“Taught man that which he knew not” (*The Holy Qur’an*. 96:5, The Clot. Translated by Sahih International).

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

2D	2-Dimensional
3D	3-Dimensional
ABS	Absorbance
AC	Affinity Chromatography
ADAR	Arginine-Dependent Acid Resistance
AdiC	L-arginine/agmatine Antiporter
AGE	Agarose Gel Electrophoresis
APC	Amino acid/Polyamine/organoCation Superfamily
AptaBiD	Aptamer-facilitated Biomarker Discovery
AR	Acid Resistance
BSA	Bovine Serum Albumin
bp(s)	Basepair(s)
CO	Cytoplasmic-Open
Cryo-EM	Cryogenic Electron Microscpy
DDM	n-Dodecyl-β-D-maltopyranoside
DNA	Deoxyribonucleic Acid
DP	Dynamic Programming
ds	Double-stranded
EB	Elution Buffer

ECM	Extracellular Matrix
ER	Endoplasmic Reticulum
FACS	Fluorescence-Activated Cell Sorter/Sorting
FDA	U.S. Food and Drug Administration
FT	Flow-Through
GDAR	Glutamic acid-Dependent Acid Resistance
GFP	Green Fluorescent Protein
GPCRs	G Protein-Coupled Receptors
IB(s)	Inclusion Body(s)
IVT	*In-vitro* Transcription
K_d	Dissociation Constant
LDAR	Lysine-Dependent Acid Resistance
MFE	Minimum Free Energy
MFS	Major Facilitator Superfamily
MP(s)	Membrane Protein(s)
MTs	Membrane Transporters
MQ-water	Milli-Q® water
MW	Molecular Weight
MWCO	Molecular Weight Cut-off
NGS	Next-Generation Sequencing
NMR	Nuclear Magnetic Resonance
ODAR	Ornithine-Dependent Acid Resistance

OM	n-Octyl-β-D-maltopyranoside
PA	Polyacrylamide
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDL	Protein-Detergent-Lipid
PLP	Pyrodoxal-5′-Phosphate
PO	Periplasmic-Open
PTM(s)	Post-Translational Modification/s
RBC	Red Blood Cell
RNA	Ribonucleic Acid
RNAi	RNA Interference
RT	Room Temperature
rt	Reverse Transcription
SCFAs	Short-Chain Fatty Acids
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size-Exclusion Chromatography
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SLCs	Solute Carriers
ss	Single-stranded
TM	Transmembrane
UPR	Unfolded Protein Response
WB	Wash Buffer

CHAPTER 1

INTRODUCTION

1.1 Thesis Organization

This thesis is organized into five chapters with subdivisions termed here as units (e.g. 1.1), sections (e.g. 1.1.1) and subsections (e.g. 1.1.1.1). Chapter 1 describes the general layout and organization of the thesis. Chapter 2 surveys the scientific literature for the definition of aptamers, and their various development (selection) techniques against membrane proteins (MPs) and their applications to these MPs. Special focus is given to amino acid transporters and especially the bacterial MP, AdiC. Additionally, chapter 2 covers an area of research that is distinct but resonates in synchrony and equal importance with aptamer research. This area is the recombinant production and purification of membrane proteins as is indeed an important prerequisite for successful aptamer selections by conventional protein-SELEX. Chapter 2 also addresses the biological importance of RNA folding and the importance of its understanding as well as the available approaches to achieve this goal. Chapter 3 addresses the materials and methods used in this work. Chapter 4 lays out the logic behind the methodological approaches used here and presents the findings of this research and their discussion. Finally, chapter 5 concludes this thesis by integrating its findings and highlighting its contribution to both the network of scientific research and to “real-life” applications. In post-references appendices (A to O), the thesis compiles extra material including tables and also more methods and results related to this work.

1.2 Aims and Objectives of this Work

This work aims to select nuclease-resistant (2'F-PY) RNA aptamers that target the bacterial amino acid transporter protein, L-arginine/agmatine antiporter, also known as AdiC, used as a model MP in this study. AdiC was overexpressed, extracted, and purified in the mild detergent, DDM. RNA aptamers are to be selected by SELEX for purified AdiC and the identified proto-aptamers are to be analyzed *in-silico* at their primary and secondary structure levels and for their phylogenetic relationship. Finally, conclusions are drawn from the research findings.

1.3 Research Significance and Value

This research supports the relatively neglected concept of aptamer selection against membrane proteins reconstituted as a part of pure, soluble protein-detergent-lipid complexes and discusses the merits of this selection compared to methods such as cell-based selection. Future research and development of the identified 2'F-modified RNA aptamers in this study may take them into diagnostic and therapeutic applications where they can be used to target AdiC in infectious bacterial agents. Alternatively, these aptamers may participate as versatile molecular tools in building the growing wealth of scientific knowledge about the structural/functional features of AdiC and subsequently, its eukaryotic homologs that can be important in various human pathologies such as cancer.

Parts or the whole of this work may be published in accredited journals as a research and/or a review article.

CHAPTER 2

LITERATURE REVIEW

2.1 APTAMERS

2.1.1 What are Aptamers?

Understanding and exploiting target-ligand binding interactions are basis for different areas of biological sciences and their applications (Spill et al., 2016). Aptamers are short synthetic (unnatural) nucleic acid (DNA or RNA) or peptide molecules that form highly structured architectures and can bind to their targets with high affinity (pM-to-nM range equilibrium dissociation constant, K_d) and specificity (Afanasyeva, Nagao, & Mizuguchi, 2019; Catuogno & Esposito, 2017; J. Wang & Li, 2011). The length of an oligonucleotide aptamer can typically range from 40 to 180 nucleotides while that of a peptide aptamer can range from 10 to 30 amino acid residues (Santosh & Yadava, 2014).

Aptamer properties such as nuclease stability are determined by the structure it adopts (usually globular) and the affinity as well as the selectivity of the aptamer to its target are determined by surface complementarity to the target molecule; hence, their name, which comes from the latin word “aptus” meaning to fit (Janas & Janas, 2011; Santosh & Yadava, 2014). Aptamer targets range from small chemical and biological molecules (e.g. metal ions, organic dyes, amino acids, drugs, antibiotics, base analogs etc.) to large proteins (e.g. growth factors, enzymes, immunoglobulins, gene regulatory factors, cell surface receptors etc.) and protein complexes, supramolecular structures (e.g. liposomes with nano-meter size), intact viral particles, and whole cells including cancerous and pathogenic cells. Traditionally,

purified soluble proteins are the most common targets in SELEX (Catuogno & Esposito, 2017) and are followed by small molecules and then cells (Dunn, Jimenez, & Chaput, 2017). Aptamers have shown to discriminate even between two enantiomers or proteins that differ by a few amino acids (Sola et al., 2020). Although some aptamers exist in nature such as riboswitches and ribozymes and have important biological functions that include catalyzing reactions, gene expression control, and mediating cellular responses and protein synthesis, most of them are generated in vitro and are raised against a specific target (Dunn et al., 2017).

Nucleic acid aptamers, making up the vast majority of developed aptamers (Spill et al., 2016), are isolated from large, random-sequence, combinatorial libraries (typically > 1 trillion distinct sequences) in a selection process known as *in vitro* selection or SELEX (systematic evolution of ligands by exponential enrichment) (Dunn et al., 2017; Takahashi, 2018). This is a unique and directed process that follows Darwinian evolution principles by mimicking natural selection as it iteratively generates enriched populations of ligands. This is achieved under a defined set of conditions and by selection and amplification of ligands that have a desired property (e.g. binding to a specific target) and then, separation from ligands that are weakly or not exhibiting that property. Aptamers and SELEX were described 30 years ago concurrently by three different laboratories (Shigdar et al., 2013). Ellington et al. generated dye-binding RNA molecules (Ellington & Szostak, 1990) while Robertson and Joyce described the generation of specific DNA-cleaving RNA enzymes (Robertson & Joyce, 1990), and Tuerk and Gold targeted a bacteriophage T4 DNA polymerase by RNA ligands (Tuerk & Gold, 1990). The idea of utilizing nucleic acid molecules as a tool in areas such as biology and medicine was not a new idea in 1990 (Shigdar et al., 2013). Antisense oligonucleotides (ASOs), for example, were first proposed as tools for gene silencing in 1977; their binding action is based on Watson-Crick base pairing rules. However, the ingenious of the SELEX method (different from that for the generation of ASOs and RNAi) lied in the introduction of a new class of nucleic acid molecules, aptamers, that are isolated to interact with

their targets via an “induced-fit” or “adaptive” mechanism guided by their complex, three-dimensionally folded shapes.

2.1.2 A Quick Overview on the Applications of Aptamers

Aptamers as a distinct class of molecules have important contributions to practical applications across a wide range of scientific disciplines (Dunn et al., 2017). These application areas can be classified into five major technological groups. **1)** Scientific tools; which involves areas such as gene regulation, nanotechnology, affinity chromatography and non-clinical sensors, **2)** clinical reagents; involving areas of therapeutics, diagnostics, drug delivery systems, and clinical biosensors, **3)** environmental sensors; in which aptamers are used as reagents for food and water analysis, **4)** informatics, and **5)** biophysical discovery. The vast majority of published articles (~90%) belong to the first three technological categories.

Aptamers are strongly reminiscent of antibodies as both molecules can be used as affinity reagents (Catuogno & Esposito, 2017; Cibiel, Dupont, & Ducongé, 2011; Dunn et al., 2017). In fact, aptamers are also called “chemical antibodies” and even have advantages that surpass those of antibodies and other used ligands (e.g. small drug molecules and peptide-based ligands such as antibody fragments, affibodies, and ankyrin repeat proteins) in the areas of research, diagnostics and therapeutics. Compared to antibodies, aptamers can be produced quickly and cheaply on larger scales more (Dunn et al., 2017; Janas & Janas, 2011). As the process of aptamer production is chemical rather than biological, the problem of viral or bacterial contamination known to occur in antibody manufacturing is eliminated. The expensive and labor-intensive biological systems needed for antibody production are not required either. Moreover, the batch-to-batch variability seen in antibody production and which hinders reproducing data by researchers is reduced. Additionally, aptamers as therapeutic agents are less immunogenic than proteins and their small size (< 30 KDa), compared to antibodies (~150 KDa), allows them to access biological areas inaccessible to antibodies. Unlike for antibodies and small

peptides, chemical modification of aptamers is a straightforward process and an attractive method to control and extend their renal clearance and half-life. Therapeutic aptamers can be rapidly deactivated by antidotes which are antisense oligonucleotides designed to base-pair with the aptamer's binding domain. Although aptamers can unfold when stored for prolonged periods at ambient temperature, aptamers can refold back into their functional state simply by a heating-cooling procedure carried out in an appropriate buffer. In addition to thermal stability, aptamers are stable at a wide range of pH and storage conditions and are not sensitive to organic solvents (Janas & Janas, 2011). Also they lower their shipping costs compared to antibodies by overcoming the cold-chain problem (Dunn et al., 2017). Aptamers can be engineered into ligand-responsive devices (sensors and other genetically controlled elements) and can be integrated with nucleic-acid based systems such as amplification systems, DNA nanotechnology, or DNA computing (Dunn et al., 2017; Ilgu & Nilsen-Hamilton, 2016).

In the area of therapeutics, the first and only FDA-approved aptamer-based pharmaceutical drug is the anti-human VEGF (vascular endothelial growth factor) aptamer (Regina Stoltenburg, Reinemann, & Strehlitz, 2007; T. Wang, Chen, Larcher, Barrero, & Veedu, 2019). Its PEGylated form called Pegaptanib was used as the active therapeutic component of the drug developed for the treatment of wet age-related macular degeneration. The drug, Macugen® (Pegaptanib sodium injection), from Pfizer Inc./OSI Pharmaceuticals was approved in the USA (in December 2004) and Europe (in January 2006). In the area of clinical diagnostics, the giant market of immunological diagnostics is also witnessing competition as different groups are developing and standardizing the use of aptamer-based test kits by replacing antibody-based diagnostic platforms. One example is the enzyme-linked oligonucleotide assay (ELONA) similar to ELISA technology. In the area of *in vivo* imaging, applied aptamers are also developed for diseases such as cancers (M. Liu et al., 2020; Shi et al., 2011).

2.1.3 Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

SELEX is the premier framework for the discovery and development of high-affinity aptamers from an initial combinatorial library consisting of 10^{12}–10^{16} random oligo sequences (Ilgu & Nilsen-Hamilton, 2016; Spill et al., 2016). The process is a cyclical one that involves repeated rounds of binding, partitioning and amplification (Catuogno & Esposito, 2017). Ligands are incubated with the target of interest and those with the better binding ability are amplified and survive to the next round (Spill et al., 2016). Aptamer-target interactions include hydrogen bonds, salt bridges, van der Waals forces, as well as hydrophobic and electrostatic interactions (Ilgu & Nilsen-Hamilton, 2016; Ilgu, Yan, Khounlo, Lamm, & Nilsen-Hamilton, 2019; Tan et al., 2016). Separation of the bound from unbound ligands is facilitated by immobilization of the target on a substrate material. Through several rounds of selection, the initial library is reduced to high-affinity aptamers enriched in an exponential manner (Santosh & Yadava, 2014). SELEX in particular is applicable with nucleic acids due to the convenience provided for intermittent amplifications of the selected ligands by *rt*-PCR or PCR. Partitioning techniques for the separation of aptamers from non-binders involves affinity columns and tags, size fractionation columns, nitrocellulose filters, hydrophobic plates, flow cytometry, polyacrylamide gel electrophoresis, capillary electrophoresis, magnetic beads, surface plasmon resonance, and microfluidic systems (Ilgu & Nilsen-Hamilton, 2016; J. Wang & Li, 2011).

Thousands of aptamers were selected over the past 30 years since the introduction of SELEX (Zhuo et al., 2017). Despite the great achievements in selection, modification, and applications, few aptamer-based products have been successfully translated into clinical and industrial use. This prominent lag is due to several major reasons such as the fact that aptamers compete with conventionally accepted antibodies in a common niche (Sola et al., 2020). Furthermore, a large phase clinical trial involving an anticoagulant aptamer held great expectations until

toxicity appeared in a small number of patients (Ganson et al., 2016). This likely caused discouragement in launching more trials for aptamer-based technologies in therapeutics. Nevertheless, it is important to note that the toxicity observed was due to the pre-existence of anti-PEG antibodies as the anticoagulant aptamer was PEG-modified to enhance its half-life.

Major intrinsic reasons for the lag of aptamers are: first, the SELEX process is still time- and labor-consuming (takes weeks to months) and the hit rates of generating aptamer candidates are low; second, most aptamers are selected *in vitro* and whether they function *in vivo* needs further elucidation (Sola et al., 2020; T. Wang, Chen, et al., 2019; Zhuo et al., 2017). To accelerate selection and enhance success rates, methods that had modified conventional SELEX were developed and used (e.g. counter/negative SELEX, cell-based and *in vivo*-based selections, capillary electrophoresis SELEX, microfluidic SELEX, high-throughput sequencing SELEX etc.). Today, some groups have reduced SELEX time from months to few hours (T. Wang, Chen, et al., 2019). Moreover, there is promise for predicting aptamer structures using computer-supported assays before even conducting SELEX (Ahirwar et al., 2016). Unfortunately, most studies have used SELEX as a “black box” instrument and paid attention more to aptamer characterization and their development for applications, and less to investigating the SELEX process itself (Komarova & Kuznetsov, 2019). Despite its simple concept, the SELEX procedure is in fact plagued by uncertainty (Spill et al., 2016). Beyond using the properties of selected aptamers (e.g. affinity and specificity) and the labor efforts invested in selection (indicated primarily by the number of rounds) to assess the efficiency of the SELEX procedure, studying and optimizing aspects such as library design (including understanding the structural features and affinity distribution of the library), selection conditions (e.g. target amount, incubation conditions, and separation method), amplification, and pool conditioning is crucial to enhance aptamer selection efficiency (Komarova & Kuznetsov, 2019; Spill et al., 2016; T. Wang, Chen, et al., 2019)

To improve the functions of *in vitro*-selected aptamers, post-SELEX modifications are typically done using several developed strategies for optimization (Dunn et al., 2017; Shigdar et al., 2013; J. Wang & Li, 2011). Most modifications aim to enhance biological stability (e.g. for use in human serum) of the selected aptamers and they typically involve the artificial addition of chemical functional groups that confer nuclease-resistance by protecting from hydrolytic attacks. Modifications can also help reduce clearance from the body by the kidney and therefore increase circulation time in the serum. Optimizing pharmacokinetic and pharmacodynamic properties of aptamers can increase the circulating half-life of the aptamers from minutes to day or a week. Although a goal pursued less commonly due to certain prediction difficulties, post-SELEX aptamer modifications can also be done to substantially enhance the binding affinity of aptamers. Because post-SELEX modifications may risk weakening the specific interaction between aptamer and its target, pools of chemically modified oligonucleotides are alternatively used in the selection process. Chemical modifications of aptamers can involve the use of 2'-fluoro or amino pyrimidines, O-methyl nucleotides, polyethylene glycol (PEG), phosphorothioates, cholesterol, organic molecular drugs, spiegelmers, and nanomaterials (J. Wang & Li, 2011). Another type of post-SELEX optimization is the "rational" truncation of aptamers (Catuogno & Esposito, 2017). This is done to synthesize selected aptamers efficiently and cost-effectively since long aptamers (> 60–70 nucleotides) are difficult to synthesize and costly to manufacture. The alternative of selection with shorter oligos, on the other hand, may not be effective due to reduced library complexity/diversity.

2.1.4 RNA Folding, Structural Stability, and Kinship with Function

Although RNAs have always been known as the intermediate biomolecules in the central dogma of biology where they carry the genetic message from DNA to proteins, the new paradigm of RNAs in molecular genetics takes us to a deeper understanding and advanced implications in which RNA molecules are shown to

have truly exceptional capacities for conformational flexibility and functional versatility (Janas & Janas, 2011; Langdon, Petke, & Lorenz, 2018). 75% of the DNA in humans is transcribed into RNA but less than 3% is translated into protein. A positive correlation has been observed between the ratio of noncoding to coding RNAs and the level of developmental complexity of the organism (Mathews, Moss, & Turner, 2010). This reflects the functional diversity of RNAs including their regulatory and catalytic functions. It has been established that RNAs have intriguing interactions with metabolites, proteins and other RNAs through their primary sequences, secondary structures, and the higher-order tertiary and quaternary structures (Weeks, 2015). However, beyond all the many levels at which information is encoded, RNA continues to surprise and excite researchers with the ability of its different structural organizations to operate at vast scales simultaneously. For example, long viral RNAs and mRNAs encode for proteins and form long-range interactions that span hundreds/thousands of nucleotides which are important in juxtaposing critical regulatory elements. Such elements that functionally interact with other proteins and small or large RNAs through specific binding sites are also structurally modulated (for enhancement or suppression) by both small- and large-scale modulatory actions such as post-transcriptional modifications (e.g. methylation) and high-order changes (secondary and tertiary) in structural motifs, respectively. All in all, the specific structures that RNA molecules adopt are important for their biological activities. Therefore, structural biology studies of RNA secondary and tertiary structures have associated RNA structural changes with human diseases such as neurodegenerative conditions and cancer genesis (Andrzejewska, Zawadzka, & Pachulska-Wieczorek, 2020; Holbrook, 2005).

In addition to its roles in gene expression such as acting as the intermediary carrier of genetic information to the ribosome and the recruitment of the correct amino acids to the translation site, through their specific structural folds, RNA molecules themselves are important regulators of the process of gene expression (Andrzejewska et al., 2020). Many of the cellular processes and mechanisms including transcription and post-transcriptional processing, the translation and

folding of proteins, cellular localization, and RNA turnover (stability and decay) are controlled by RNAs through their adopted specific structures. *In vitro* studies were very successful in determining many of the secondary structures of RNA molecules since the majority of information on a secondary structure is encoded within the RNA sequence. That is, Watson-Crick base pairs (paired by hydrogen bonds) and unpaired loop-forming bases describe the secondary structure of an RNA molecule (Afanasyeva et al., 2019). Stacking of these base pairs results in formation of the RNA's three-dimensional (3D) structure driven by the formed scaffold (Singh, Hanson, Paliwal, & Zhou, 2019). *In vitro* experimental methods used routinely for the inference of base pairs are mainly one-dimensional or multi-dimensional probing approaches that use enzymes, chemicals, mutations, and cross-linking techniques coupled with various reading methods. However, high-resolution data at the level of a single base pair requires the 3D structures solved by X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (cryo-EM). Unfortunately, these experimental methods do not easily conform to the nature of RNA molecules and that is, being flexible, dynamic and often, large in size (e.g. pre-mRNAs) and with multiple conformations (Andrews & Moss, 2019; Antunes, Jorge, Caffarena, & Passetti, 2018). However, the secondary structure of RNAs is much more accessible and can provide valuable structural/functional information.

For a long time and aside from *in vitro* experiments (Washietl, Hofacker, Stadler, & Kellis, 2012) our body of knowledge on the secondary structures of RNAs was built primarily based on *in-silico* methods that calculate from a primary sequence, the most thermodynamically favorable structure or predict consensus structures conserved in homologous RNA sequences (Andrews & Moss, 2019; Andrzejewska et al., 2020). This is largely the case due to limitations in probing experiments (e.g. incomplete structural data) as well as the difficulties and time-consumption associated with solving 3D RNA structures by the traditionally known experimental techniques, thus making computational methods an attractive, time-saving and cost-effective alternative (Afanasyeva et al., 2019; Singh et al., 2019). Less than 0.01% of over 14 million secondary structures of noncoding RNAs found

in RNAcentral are experimentally determined structures. In the Protein Data Bank (PDB), by November 2017, 0.9% of all structures deposited were RNA structures (Antunes et al., 2018). By further examination through the computational prediction pipeline, we see that this approach is also desirable for the reconstitution of RNA 3D structures from the predicted secondary structures, and the simulation of RNA-target complex structures (e.g. aptamer-protein complexes) by docking or any other method for RNA-target complex structure prediction (Afanasyeva et al., 2019). Nonetheless, tertiary structure prediction is still considered to be in its infancy and is limited to very small molecules (Laing & Schlick, 2011; Langdon et al., 2018; Shapiro, Yingling, Kasprzak, & Bindewald, 2007).

In this thesis, we are mainly concerned with the first step of the prediction pipeline and that is predicting secondary structures from their primary sequences and particularly for RNA-based aptamers. Although most of the modeling tools used currently in prediction were designed for RNA sequences, these tools can also be preliminary applied for DNA sequences and with acceptable accuracies (Afanasyeva et al., 2019). Interestingly, more recent efforts have aimed at developing a new tool to predict nucleic acid secondary structures by extraction of these structures from the 3D ones given as PDB files (Zok et al., 2018).

Unlike the prediction of secondary structure by comparative analyses of homologous sequences where conserved base pairs are revealed, folding algorithms that commonly use thermodynamic, statistical, or probabilistic scoring functions are the most common approach (Rivas, 2013; Rodriguez & Cortes-Mancera, 2013; Singh et al., 2019). Although the former approach can be more accurate, it requires a significant number of available homologs and alignment expertise. In the latter approach (folding algorithms), the structure of the query RNA is divided into substructures such as loops and stems according to the nearest-neighbor model. Then, dynamic programming (DP) algorithms, most commonly, are used to locate the global minimum or probabilistic structures from these substructures. The scoring parameters used for each substructure can be either obtained experimentally (e.g. *RNAfold*, *RNAstructure*, and *RNAshapes*) or by machine learning (e.g. *CONTRAfold*,

CentroidFold, *ContextFold*). It is noteworthy that machine learning is rarely used to directly predict secondary structure of a single RNA sequence as a low number of nonredundant RNA structures is typically available. Historically, *Mfold* developed by Zuker et al. in the 1980s has been the first approach based on finding the minimum free energy (MFE) structure using energy parameters obtained from thermodynamic experiments (Afanasyeva et al., 2019; Zuker, 2003). Other MFE-based approaches and further modifications done to *Mfold* have later incorporated implementation of the partition function which provided base pair probabilities. This had led to a substantial improvement in prediction accuracies. Later on, the approach in which prediction depended on analyzing the ensemble of all possible solutions with the centroid estimator had emerged (used by *CentroidFold*, for example).

For approaches that use the laws of thermodynamics to find the lowest free energy structure (i.e. those that have the most negative free energy), meta-heuristic algorithms are also used in addition to DP algorithms (Mohsen, Khader, & Ramachandram, 2009). These can have superior performance to DP and particularly for complex energy minimization problems where an increased number of variables result in an exponential increase in the number of recursive function evaluations. Meta-heuristic algorithms can also provide both suboptimal and optimal structures since assuming that the correct structure to have is the lowest energy structure is often not true (5–10% energy from the MFE of a query is usually that of the native structure). Meta-heuristic algorithms include Genetic Algorithm (*RnaPredict*), Simulated Annealing (*SARNA-Predict*), Particle Swarm Optimization (*HelixPSO*), and Harmony Search (HS) algorithm (*HSRNAFold*). In a comparison between *RNAfold* and *HSRNAFold*, both are said to have good prediction accuracies for short RNA sequences. However, for tested RNA sequences between 117 and 945 nucleotides (especially those above 500), *HSRNAFold* had clearly excelled in prediction accuracy when all predicted structures are compared to the native structures. The meta-heuristic HS algorithm is a stochastic iterative optimization algorithm that was applied to many different problems such as engineering and vehicle transportation problems. The algorithm was in fact inspired by musical

improvisation where orchestral musicians try to find a fantastic harmony with the best combination of states to produce that fantastic and "optimized" harmony.

Although methods of computational prediction are still developed and improved, they are often ineffective for long and structurally complex RNAs and they usually don't consider the intimate interplay between an RNA molecule and the *in vivo* environment by which cellular/physiological factors interact with and affect RNA folding (Andrzejewska et al., 2020). These factors include but are surely not limited to RNA modifications, RNA-binding proteins, the translational machinery, and the strict control of RNA stability and degradation. Many of the results from *in vitro* studies were also inconsistent with the RNA behavior *in vivo* (Ding et al., 2014; Mortimer, Kidwell, & Doudna, 2014; Smola et al., 2016). However, advancements in methods of enzymatic/chemical probing of RNA structures and incorporation of the experimental data into the folding algorithms has improved the accuracy of computational predictions of RNA structures significantly and thus, mitigated the inconsistencies that were prevalent when determined RNA structures are compared with those in the *in vivo* environment (Andrzejewska et al., 2020). Moreover, during the last decade, the development of newer and more sophisticated methods for the measurement of RNA structures inside living cells has also been revolutionary for structural and functional RNA studies. Coupling RNA probing with next-generation sequencing (NGS) and advanced bioinformatical tools has not only allowed effectively studying RNA structures both *in vitro* and *in vivo*, but has also accelerated the results up to high-throughput levels whereby thousands of heterogeneous RNAs and even whole transcriptomes, all in the cell's complex environment, are progressively studied (Ritchey et al., 2017; Spitale et al., 2015; Zubradt et al., 2016). Despite all, today, there are still large gaps in our understanding of RNA folding influenced by the *in vivo* environment and these gaps must be sealed if the functional mechanisms of these RNAs are to be understood. Unfortunately, the performance of current folding-based prediction algorithms in fact has been in stagnation for over a decade (Singh et al., 2019). The overall precision determined by the fraction of correctly predicted base pairs (bps) among all bps seems to have hit a "performance

ceiling" of ~80%. This is in part because in addition to ignoring the *in vivo* conditions, all available secondary structure prediction methodologies do not take into account some or all of the tertiary-interactions-induced bps such as lone (unstacked) bps, pseudoknots (non-nested bps), and noncanonical bps (not AU, GC and GU), as well as triplet interactions. While some methods such as *pknotsRG*, *Probknot*, *IPknot*, and *Knotty* and others like *MC-Fold*, *MC-Fold-DP*, and *CycleFold* can predict the RNA secondary structures with pseudoknots (e.g. interactions between loop nucleotides or a free strand and stems) and noncanonical bps, respectively, none of them can simultaneously predict both. This is without mentioning lone bps, triplets and even the complex and diverse G-quadruplexes that are found more often in DNA molecules and work to enhance the thermodynamic stability of nucleic acid structures (Afanasyeva et al., 2019; Davydova et al., 2020).

2.2 APTAMERS AND MEMBRANE PROTEINS

2.2.1 Applications of Aptamers to Membrane Proteins

MPs are carefully embedded into lipid bilayers by strong hydrophobic interactions and have water-soluble protrusions extending onto the intra- and extracellular matrix (ECM) (Ilgü et al., 2014). They undertake highly critical cellular roles such as signal transductions, energy production, cell-to-cell communication and regulation of incoming and outgoing materials moving across plasma membrane barriers (Mus-Veteau, Demange, & Zito, 2014). Although, MPs comprise 23% of the human proteome, given their important roles in many diseases, these proteins are an enormous majority that represents over 60% of pharmacologically attractive targets intended for drug discovery, development and delivery (Errey & Fiez-Vandal, 2020). Cell-surface proteins, unlike intracellular proteins, are highly accessible drug targets making them attractive for clinical purposes (Dua, Kim, & Lee, 2011). Alterations in cell-surface proteins have been linked to a large number of human diseases. Changes in expression levels, localizations, structure (e.g.

truncations, mis-assembly), protein aggregation, or loss of function are common in disease. Therefore, all these MP abnormalities represent potentially targetable disease biomarkers. Monoclonal antibodies have tremendously contributed to drug targeting of surface MPs. However, their limitations in various important areas also led largely to the high demand of new and specific ligands against cell-surface MPs (Cibiel et al., 2011). Refer to **section 2.1.2** that presents an overview on how of aptamers are compared to antibodies.

Using MPs (whether located on cell surfaces or purified) in aptamer selection focused strongly on MPs from mammalian sources as well as infectious pathogens like viruses, bacteria, fungi, and parasites (Cibiel et al., 2011; Dunn et al., 2017; Mallikaratchy, 2017). For mammalian cells, most selections have targeted cancerous over non-cancerous cells. Aptamers for MPs can potentially be used in many different areas of applications. For example, these aptamers can be used as new drugs that inhibit or activate their targets involved in a specific disease condition. Alternatively, these aptamers can be used as *in vitro* and *in vivo* delivery agents by conjugation to packages such as drugs, siRNAs, miRNAs, nanoparticles, and contrast agents for imaging applications. In the areas of diagnostics and monitoring, aptamers can be potent probes against cell-surface targets in the fields of cancer, infectious diseases, food safety, and bioterrorism. In basic research, aptamers can be used in areas such as cell-surface biomarker discovery and cell phenotyping especially due to the limited success with immunization of many MPs and their presentation to antibodies in a native conformation.

2.2.2 Decades of R&D in SELEX Methodology to Accessing Membrane Proteins for Successful Aptamer Selections

Ever since the race to selecting aptamers against protein targets has begun, MPs have fallen behind and struggled to catch up with their soluble cousins when the number of aptamers being generated are compared (Dua et al., 2011; Errey & Fiez-Vandal, 2020; Janas & Janas, 2011; S. P. Ohuchi, Ohtsu, & Nakamura, 2006;

Pandey, Shin, Patterson, Liu, & Rainey, 2016; Carine Pestourie et al., 2006). Surely, this circumstance cannot imply the unimportance of MPs or targeting them by SELEX for aptamer selection. Raising novel aptamers against this class of proteins has shown to have a significant value in diagnostic and therapeutic areas as discussed previously. The articulated historic struggle of MPs compared to water-soluble proteins with successful aptamer selection is similar to that of and is as unique as the problems faced commonly with solving 3D molecular structures of MPs by X-ray crystallography compared to solving those for soluble proteins. These struggles have manifested simply due to the many work challenges that have long been associated, and thus, are now customary to the production of this fascinating and enigmatic class of proteins; perhaps, most commonly, their complex, laborious, time-consuming, and expensive production process carried out using crude biological samples or samples of genetic engineering and recombinant DNA technologies (Goto, Tsukakoshi, & Ikebukuro, 2017; S. P. Ohuchi et al., 2006; S. M. Smith, 2017). To select aptamers against any purified protein of interest, producing that protein in sufficient quantities and a stable, correctly folded and preferably functional form is an indisputable prerequisite for selection and unfortunately, is the greatest bottleneck of all protein production processes (Catuogno & Esposito, 2017; Pandey et al., 2016; Shamah, Healy, & Cload, 2008). Unlike soluble proteins, MPs have been imposing unparalleled difficulties and very unique challenges to the process of their production due to several of their intrinsic properties: **1)** their low abundance in natural sources (and especially in mammalian MPs) which consequently forces researchers to go through the protocols of recombinant overexpression and purification that are quite troublesome on their own in many aspects; **2)** their inherent structural complexity augmented by their range of molecular flexibility, all of which are a function of many factors such as various post-translational modifications (PTMs) and also the need, in many cases, for the co-presence with other residents of the lipid bilayer membrane to attain optimal functional conformity; **3)** the difficulty to retain their structure throughout and after purification procedures as they are molecularly unstable (e.g. proneness to aggregation due to their highly hydrophobic nature) and chemically

fragile (e.g. proneness to degradation by proteases) when removed from their natural hydrophobic milieu and solubilized by certain solubilizing agents (e.g. detergents, organic solvents and chaotropic agents) (Dua et al., 2011; Goto et al., 2017; Ilgü et al., 2014; Rasmussen et al., 2007; Shamah et al., 2008; S. M. Smith, 2017; C. G. Tate, 2001).

Recombinant protein production has clearly relieved many of the difficulties with purifying MPs from crude extracts. These include sample heterogeneity, labor-intensive purification, and small amount of MP of interest in natural sources just to name a few. Many impressive accomplishments have been made over the past 20 years in the development of novel recombinant MPs production techniques that in turn have narrowed the gap between MPs and soluble proteins and have sped up advancements such as uncovering many of the molecular structures of MPs by X-ray diffraction (XRD), NMR spectroscopy and cryo-EM (García-Nafría & Tate, 2020; Pandey et al., 2016; Rosano & Ceccarelli, 2014; Routledge et al., 2016). Nevertheless, producing satisfactory amount of recombinant MPs of interest by overexpression and purification in soluble, stabilized, well-folded and functional states, remains a task that can, in many cases, be notoriously daunting to achieve (Dilworth et al., 2018; Gul, Linares, Ho, & Poolman, 2014). In general, the trial-and-error and the wide spectrum and different dimensions in which a typical recombinant proteins production process can generally fail, all exist and are to be expected (C. G. Tate & Grisshammer, 1995). However, more than that, as it is a hallmark of any rising technology and its applications, oftentimes novel challenges, limitations and drawbacks always emerge. For example, compared to water-soluble proteins, our understanding of the nature of MPs such as G protein-coupled receptors (GPCRs), serine/threonine kinase receptors and ion channels (S. Ohuchi, 2012; S. P. Ohuchi et al., 2006), and their structure-function relationships is lacking significantly (Gul et al., 2014). This can be viewed as a direct result of the intricacy of their biogenesis pathways which translates into the novel difficulties presented by these proteins to the attempts of their production. To put matters more into perspective and from a different point of view, only a fraction of human proteins has been recombinantly

expressed and purified (Goto et al., 2017), and the majority of these are known to be the soluble ones. Less than 1% of proteins recorded in PDB are structures of MPs (Almeida, Preto, Koukos, Bonvin, & Moreira, 2017). Without a doubt, the difficulty in producing and handling MPs can be a reason for such statistics given their special and intricate nature. However, in many other cases, undertaking the stereotypical means to obtaining these proteins provides a large share of the problem. Many of these methods can also be associated with expensive and time-consuming protocol optimizations. An example to these issues is the conventional use of detergent-based purification methods in which lipid bilayer systems are disrupted, leading MPs to be stripped off their natural lipidic environment that they use to maintain their delicate folding and structure-dependent functionalities (Lee et al., 2016; Rothnie, 2016). Exercising such dramatic attempts to purify MPs may result lastly in hindering the success rates of selection of useful and relevant aptamers as will be elaborated in the next subsection (Catuogno & Esposito, 2017; Takahashi, 2018). Indeed, many mammalian MPs are unstable when extracted into aqueous detergent solutions and finding optimal detergent and buffer conditions (for MP homogeneity, functionality, stability, and possibly crystallization) is a common bottleneck (Mus-Veteau et al., 2014). Despite any shortcomings, detergent-based purification methods have been indispensable for the purification of MPs and are of an undeniable value. Moreover, many other tools and techniques are undergoing development in order to find the best conditions to obtaining MPs. Altogether, recombinant protein production technologies will surely remain indefinitely as an attractive and an ever-evolving tool in our arsenal and used to propel MPs towards faster, easier and more successful production of functional proteins (Dilworth et al., 2018; Rosano & Ceccarelli, 2014). Continuous progress and growth in the field of recombinant production of functional MPs (expression and purification) directly prompts commencement of traditional "protein-SELEX" trials so that fruitful selections yielding better, and more effective aptamers are achieved. Of course, this is possible only when adequate amounts of authentically purified target is acquired priorly; i.e., the purified MP successfully maintains a structural and functional resemblance that is close to the native protein

(Catuogno & Esposito, 2017; Shamah et al., 2008). Refer to **section 2.2.3** for a more in-depth review on the production of recombinant MPs.

2.2.2.1 Selection of Aptamers against Purified Full-length Forms of Membrane Proteins

The majority of aptamers are known to be selected against purified soluble forms of protein targets (Catuogno & Esposito, 2017; Shamah et al., 2008; Zhou & Rossi, 2014). Due to its simplicity and ability to be conducted under well-controlled conditions, conventional protein-SELEX remains to be a widely used, if not the most widely used method. In one statistic, approximately 74% of SELEX publications from recent years have described the use of the standard method of selection (T. Wang, Yin, et al., 2019). Aptamers are a highly specific class of molecules with the ability to discriminate between closely related targets (Sola et al., 2020). A very common bottleneck that emerges when protein-SELEX is performed against a purified MP is that the aptamer selected against that MP may or may not interact well with the same MP from the native expression; i.e. recognition of the naturally expressed MP while residing in the embrace of its native host (Dua et al., 2011; Shamah et al., 2008; S. M. Smith, 2017). In the latter scenario, the aptamer is essentially rendered irrelevant (binding-wise) to the native MP and subsequently, irrelevant to any intended applications involving binding to native protein structures. This discrepancy can be chiefly attributed to a structural/conformational divergence of the purified selection target from the unpurified naturally localized target. This problem can also persist even if a virtually “perfect” purification procedure was accomplished wherein the MP of interest is solubilized, stabilized and obtained in a well-folded form that bears no alterations to its structure and retains the natural protein form found when membrane-embedded and in the physiological environment of the expressing organism. Such a discrepancy is simply a consequence of the well-known and recurrently encountered lack of a clear-cut, total superimposition shared between biotechnological recombinant protein production methods, and how nature

manufactures and architects its own proteins. In other words, an overexpressed and "perfectly" purified MP can still adopt a different structure compared to its natural nonrecombinant analog due to a myriad of reasons (Bernaudat et al., 2011; Sola et al., 2020). In practice, isolation of a well-conformed and stabilized MP (whether it being recombinant or otherwise) fairing a native-like structure can pose a great challenge and, indeed, requires great expertise and skills to permit proceeding to downstream biochemical and biophysical assays, structural determinations and to any further applications such as drug design (Bill et al., 2011; Bill & von der Haar, 2015) or, in our case, raising novel aptamers that are relevant and effective for the final intended future applications.

Albeit a seeming lack of synchrony with unwanted ramifications—especially for projects in aptamers selection—lives constantly between recombinant and natural MPs due to imperfections in the road to their production and purification, it is without a doubt that such a gap is remediable by tools from classical bioprocess optimization as well as the exciting and promising methods of systems and synthetic biology for protein and cellular re-engineering (Palomares, Estrada-Mondaca, & Ramírez, 2004; S. Xiao, Shiloach, & Betenbaugh, 2014). Deficiency in knowledge on all factors affecting a satisfactory overexpression and purification of a given MP (especially those with complex folding, oligomeric assembly, and processing pathways) is a main source of disparity. Therefore, bypassing all such roadblocks will lead to the ultimate goal of recombinant products generated in high yields, desirable structures/conformations and biologically active forms (Bernaudat et al., 2011; Peleg & Unger, 2012).

Despite the popularity of conventional protein-SELEX, a limited number of reports in the current literature address the use of purified full-length MPs in selection experiments (Cibiel et al., 2011; Takahashi, 2018). This limitedness is most prominent with multi-pass MPs compared to single-pass ones (Zhou & Rossi, 2014). Moreover, in addition to the challenges of purifying solubilized MPs, many MPs are functional in the native conformation or when in multiprotein complexes. Despite all challenges, a number of successful aptamer selections were reported for purified full-

length MPs and will be reviewed along this chapter. It wasn't until the early 2000's that SELEX experiments using full-length MPs was conducted. In 2002, RNA aptamers were generated against the rat G-protein-coupled receptor (GPCR) which binds to a tridecapeptide neurotensin (NT). The receptor, called neurotensin type I (NTS-1), was recombinantly expressed in *E. coli* as a maltose-binding protein (MBP) fusion with a C-terminal affinity tag and was then stabilized in detergent micelles (Daniels, Sohal, Rees, & Grisshammer, 2002; Janas & Janas, 2011). The fusion protein, MBP-NTS-1, was solubilized in a mixture comprised of three detergents (CHAPS, CHS and LM) to obtain protein-detergent micelles. Purification of the receptor was carried out using the C-terminal affinity tag followed by an NT column. The functional protein was obtained after cleavage of the N-terminal MBP moiety and the C-terminal tag in protease-sensitive delineation sites. A 2′F-modified RNA library with a 40-nucleotide random region was incubated with functional NTS-1 receptor adsorbed hydrophobically to paramagnetic polystyrene beads. Elution to obtain aptamers was done using NT. After 7 rounds of enrichment, candidates were cloned and sequenced. Aptamers binding to membrane-inserted MBP-NTS-1 in *E. coli* had a K_d value of 4–20 nM. More than that, one of the aptamers was shown to bind specifically to NTS-1 expressed in live Chinese hamster ovary (CHO) cells. **Table 2.1** summarizes a few examples from the history of the selections of high-affinity aptamers against purified full-length target MPs.

Table 2.1 Examples of purified full-length MPs targeted for aptamer selection [references: (Ababneh et al., 2013; J. Liu et al., 2010; Moreno et al., 2003; Ryul & Lee, 2013)]**.**

Target MP	***Native Source***	***Pool Chemistry***	***Selection Rounds***	***Partitioning Method***	***Binds Native MP***
CD44	Mammalian	2′F-Py RNA	11	GST Magnetic Beads	Yes
HBsAg	Viral	RNA	8	Membrane Filtration	Yes
OmpC	Bacterial	2′F-Py RNA	5	Precipitation	Yes

Table 2.1 (Cont'd)

KMP-11	Parasitic	DNA	10	AC	Binds target from total cell lysate

2.2.2.2 Selection of Aptamers against Purified Ectodomains of Membrane Proteins

Purification of the extra- or intracellularly localized domains of MPs is often a less challenging task compared to that of full-length MPs (Cibiel et al., 2011). Indeed, most aptamer selections were carried out against the former two in contrast to the latter. Because these cell surface MP protrusions are soluble, their hydrophilicity makes them more likely to conserve their folding during purification and hence, maintain their 3D structure. The first successful selection of aptamers targeting an MP was in fact attempted in the year 1996 by NeXtar Pharmaceuticals Inc. against the extracellular domain of L-selectin—a calcium dependent cell surface molecule used by leukocytes for cellular adhesion to endothelial cells through specific cell surface carbohydrates. An L-selectin ectodomain-IgG2 Fc C-terminal domain chimera was prepared (named L-selectin receptor globulin or LS-Rg) and immobilized on protein A-sepharose resins (Koenig et al., 1996). 14 rounds of SELEX were performed on the purified immobilized chimeric molecule and using a nuclease-stabilized 2′-aminopyrimidine (2′NH_2-Py) RNA library. Elution of the bound candidates was done with 5mM EDTA solution. The identified aptamers had a high affinity at 4 and 22°C but very low affinity at the physiological temperature 37°C. Newer SELEX trials were performed using a DNA library to obtain aptamers that exhibit thermal stability and compatibility with physiological conditions for *in vivo* applications.

Another example of early aptamers selected against ectodomains of MPs is on the prostate-specific membrane antigen (PSMA) (Janas & Janas, 2011; Lupold,

Hicke, Lin, & Coffey, 2002). The experiments, carried out in 2002, demonstrated for the first time, the cell-specific activity of the aptamers. PSMA is a glycosylated MP overexpressed in human prostate cancer cells. Expression of xPSM (the extracellular 706-amino acid domain of PSMA) was carried out in sf-9 insect cells using a recombinant Baculovirus system. The system is capable of mammalian protein modifications and the size of the purified product suggests glycosylation. The native conformation of the product (xPSM fusion protein) was confirmed by enzymatic activity. xPSM was then bound to magnetic beads and selection was performed using an RNA pool with sequences containing 40-nucleotide random regions and flanked by constant regions. Partitioning from the target was done magnetically and the bound RNAs were then reverse-transcribed and PCR-amplified. 6 selection rounds were performed and during *in vitro* transcription, 2′-fluoro-pyrimidines were integrated for enhanced nuclease stability. The selected aptamers suppressed xPSM enzymatic activity with an inhibition constant (K_i) of 2-10 nM and a 15-nucleotide truncation of one of the aptamers did not affect its binding capacity. A truncated aptamer, as seen by fluorescence microscopy, was also capable of specifically recognizing PSMA-expressing LNCaP human prostate cancer cells but not PSMA-lacking PC-3 human prostate cancer cells.

As it has been introduced earlier, an aptamer selected against a purified MP may not necessarily bind the same target expressed on the cell surface most commonly as a result of differences in conformation/modification states and/or the need for localization and interaction with other cell surface molecular components. This binding inconsistency also stretches out to include aptamers for purified extracellular fragments of the proteins of interest targeted during SELEX. In 2009, a study was published by a group showing the inability of aptamers selected against the un-glycosylated epidermal growth factor receptor variant III (EGFRvIII) ectodomain expressed in and purified from *Escherichia coli* system to bind the membrane-bound full-length glycosylated form of the protein expressed in a eukaryotic system (Yingmiao Liu et al., 2009). EGFRvIII is a heavily glycosylated protein and this undoubtedly overthrows the conformation of the purified

ectodomain when it is compared to that of the full-length EGFRvIII. Since selected aptamers have high affinities due to rigorous conformational complementarity with the target, any change in the target's structure will significantly shrivel their affinities. This incidence demonstrates a potential for the irrelevance of aptamers to the native MPs following their selection against purified representatives. Moreover, it is also possible that abandoning a prokaryotic expression system and inclusion of correct PTM patterns on an ectodomain purified from a eukaryotic system may still not guarantee the selection of aptamers that bind the native form (Cibiel et al., 2011). **Table 2.2** outlines more examples of purified MP ectodomains targeted for high-affinity aptamer selection.

Table 2.2 Examples of MPs with ectodomains purified and targeted for aptamer selection [References: (Barfod, Persson, & Lindh, 2009; Chauveau et al., 2007; C. H. B. Chen, Chernis, Hoang, & Landgraf, 2003; Sung, Kayhan, Ben-Yedidia, & Arnon, 2004)]**.**

Target MP	***Native Source***	***Pool Chemistry***	***Selection Rounds***	***Partitioning Method***	***Binds Native MP***
VCAM-1	Mammalian	2′F-Py RNA	12	AC	No
HER3	Mammalian	RNA	15	Filtration/Gel-Shift Assay	Yes
Hemagglutinin (HA)	Viral	DNA	3	AC	Yes
PfEMP1	Parasitic	2′F-Py RNA	8	AC	Yes

As described earlier, MPs are not easy to purify and to obtain in stabilized and functional forms (Takahashi, 2018). Moreover, the majority of aptamers selected against purified protein targets since the invention of SELEX have failed to recognize their targets in the native environment (Kaur, 2018; Mallikaratchy, 2017). Such typical SELEX-associated issues drove scientists to develop and use newer and more advanced SELEX techniques that had also, made the selection process much

faster, more robust and cost-effective, and led to the development of more relevant and effective aptamers (i.e. aptamers that bind their MP partners at their native states) with higher applicative potential (M. Chen et al., 2016). Novel methods of selection that have evolved from classical SELEX can be vast and overwhelming to explore in details (Takahashi, 2018; T. Wang, Chen, et al., 2019). Most have conserved the core of the traditional SELEX methodology; while others are astoundingly divergent from and add much to the original method and accordingly, portray how far SELEX technology has come to fulfill new purposes and to overcome unique challenges. The novelty in SELEX methods is comprehensible by the specific designations given to them. These designations may reveal insights into the type of sample targeted, certain technicalities in implementation of the method, or simply a generic description of the approach undertaken to reach the end goal for which they were tailored. In addition to the progress in devising new selection techniques to expand the range of targets, advancements in synthetic sciences, sequencing, bioinformatics and the use of various specialized equipment have contributed greatly to boosting breakthroughs in the SELEX technology. Some of the remarkable advances in SELEX to discover aptamers particularly for MPs and without any prior need for their purification, will be explored in the forthcoming subsections of **section 2.2.2**. Finally, it is crucial to point out that the current stance of modern SELEX technologies remains limited in its various ways and at different levels. For example, there is a significant difficulty in targeting complex MPs entirely embedded in membrane bilayers (e.g. GPCRs, channels and transporters) for aptamer discovery. This limitation highlights in particular the significance of efforts to robustly express and even functionally purify and stabilize these MPs as a ubiquitous and conserved approach that is valuable to expanding the repertoire of targetable MPs and selected aptamers. Such lines of efforts were exceptionally demonstrated by a series of experiments performed recently on beta2-adrenoreceptor (β_2AR), a well-characterized cell surface receptor and the first cloned ligand-binding GPCR. Identification of RNA aptamers for β_2AR-detergent micelles was followed by showing competence for allosteric stabilization and modulation of β_2AR and its activity by these aptamers (Kahsai et al., 2016).

2.2.2.3 Selection of Aptamers against Membrane Proteins in Membrane Preparations

To generate aptamers against MPs residing in a more natural environment, SELEX can be done with extracted/ lipidic membrane preparations [e.g. plasma membrane fragments (see **Table 2.3**), vesicles (Ulrich et al., 1998), exosomes (Murakami, Zhao, Yamasaki, & Miyagishi, 2017), and liposomes (Takahashi, 2018), etc.]. In many cases, the purified membrane sample is a heterogeneous mixture of many potential aptamer targets and hence, is called a complex-target sample. Such a target sets the ground for multiple parallel selections that are later analyzed and resolved (Shamah et al., 2008). Accordingly, a purified membrane sample homing ideally a single MP, is a prerequisite for a defined selection experiment as it is directed against a specific target. SELEX performed against any sample pertaining to a complex mixture of many potential targets is generally dubbed as "complex-target SELEX". This technique, on the other hand, grants aptamer selection without any prior knowledge about the identity and characteristics of the target/s. Due to the heterogeneity of the sample at hand, selection commonly entails involving some type of "deconvolving" strategy. When no molecular-level knowledge on the targets present in the complex mixture sample is available, i.e. the targets' identities are unknown and undefined, employing a "deconvolving" strategy helps to reveal the binding partners (molecular targets) of the selected aptamers, and/or simply to obtain aptamers that can differentiate between different cells and tissues. Conversely, when the target is known and specifically defined, deconvolution aids in faster and better targeting that is directed to the desired MP in the complex sample. The incorporation of deconvolution strategies can wildly alter the outcome of a certain SELEX experiment. The importance of these techniques will be elaborated further throughout the remainder of this literature review in **section 2.2.2**.

Some of the earliest complex targets that were used for the selection of high-affinity nucleic acid ligands are intact viral particles (Pan et al., 1995) and 30S ribosomes (Ringquist et al., 1995). However, selection against MPs embedded

within membrane preparations was attempted for the first time by using human red blood cell (RBC) ghosts (Mallikaratchy, 2017; Morris, Jensen, Julin, Weil, & Gold, 1998). The success of this iconic SELEX was largely due to the possession of RBC ghosts a reasonable number of targets that remained constant and with unperturbed folding states throughout selection. This led eventually to the isolation of high-affinity ssDNA aptamers. Unlike conventional SELEX, rounds of selection against the complex target, RBC ghosts, were unusually high—25 rounds—to obtain a pool that converged better towards the dominant family of sequences. By photoaffinity cross-linking, SELEX against RBC ghosts was shown to have simultaneously generated aptamers for multiple target MPs and with binding affinities comparable to those obtained from previous selections that have used the same targets in purified forms. The Diversity of potential targets in the heterogeneous sample imposes the disadvantage of lack of knowledge about the corresponding target of each aptamer. That is, the individual target of each unique sequence is unknown, initially at least. To overcome this limitation, the study incorporated, for the first time, a secondary selection scheme, “deconvolution-SELEX”, which facilitated further evolution of the enriched round 25 pool to rapidly produce pools with a clear bias to their individual protein targets. Briefly, 4 rounds of this selection were done where crosslinking products between the pool and RBC ghosts analyzed by SDS-PAGE and autoradiography were transferred to a nitrocellulose filter. Un-crosslinked ssDNAs remain in the gel and four varying bands on the filter were excised and had the associated ssDNA molecules amplified by PCR to be cross-linked to the ghosts again. In conclusion, Morris et al. have successfully used RBC ghosts, a complex mixture sample of multiple targets, as a comprehensive model target system to evolve high-affinity nucleic acid ligands by complex-target SELEX. The selection technique represents a powerful tool to dissecting and comparing complex biological systems such as whole cells, tissues and clinical samples such as serum and urine from different patients.

If SELEX is to be carried out to target an initially known and predefined MP in a complex-target sample, a secondary selection scheme, “counter-SELEX” (also

commonly referred to as negative or subtractive selection), is a standard pool-deconvolving strategy that is commonly used (Shamah et al., 2008; Regina Stoltenburg et al., 2007; Takahashi, 2018). For a pool evolving against a complex-target sample, a counterselection scheme involves using a second complex-target mixture lacking only the desired MP target for that secondary selection and by the same pool. Consequently, undesirable co-selections can be counteracted by the elimination of all the target-nonspecific sequences obtained from the negative selection. Since the MP of interest is only present in the primary sample and not in the negative selection sample, selected aptamers in the final enriched pool have enhanced specificity to the MP of interest and to the primary sample. Counter-SELEX may be used even if a target cell surface MP is undefined. Such a procedure yields cell- or tissue-specific aptamers that discriminate against other related cells or tissues. Counterselection hence, is a strategy that allows—but not always (Cibiel et al., 2014; T. Wang, Chen, et al., 2019)—aptamer pools to deconvolve faster and converge better and specifically towards an MP of interest or more generally, the primary complex-target sample. Counterselections were performed much earlier than when complex mixtures were targets of selection. It was, and still is in various studies, used to reduce the parallel selection of nonspecific aptamers that bind unwanted structures such as immobilization matrices or other molecules that are related structurally to the target molecule. Another effective deconvolution strategy that can be used to steer SELEX towards a more specific selection when the target is known integrates the use of known high-affinity ligands of the target of interest during selection rounds to specifically elute off the corresponding desired aptamer candidates (after removal of the unbound sequences) (Cibiel et al., 2011). Employment of such a technique has resulted in very potent and competitive aptamers compared to the known ligand(s) of a targeted MP. **Table 2.3** summarizes some selection reports for MPs purified within extracted membranes and **Table 2.4**, elaborates on the deconvolution approaches used to facilitate a more efficient enrichment of the selected aptamers against these MPs.

Table 2.3 Different selection examples against MPs harbored within membrane preparations [References: (Y Cui, Ulrich, & Hess, 2004; Yang Cui, Rajasethupathy, & Hess, 2004; Z. Huang, Pei, Jayaseelan, Shi, & Niu, 2007; Joshi et al., 2009)]**.**

Membrane Extract	***Target MP***	***Pool Chemistry***	***Selection Rounds***	***Partitioning Method***
Rat Brain Membranes	$GABA_A$ Receptor	2′F-Py RNA	12	Membrane Filtration
***T. californica* Electric Organ**	Nicotinic ACh Receptor	2′F-Py RNA	9	Filtration/Gel-shift Assay
Transfected HEK293S Cells	GluR2Qflip AMPA Receptor	RNA	14	ND
***Salmonella enterica* serovar Typhimurium**	Targets were initially unknown	DNA	7	Membrane Filtration

Table 2.4 A summary of the deconvoluting strategies used to efficiently enrich for aptamers against the MP targets purified in membrane preparations and exemplified in Table 2.3.

Membrane Extract	***Comments***
Rat Brain Membranes	Membrane fragments of $GABA_A$ receptor were prepared from forebrain tissue of CO_2-euthanized rats. RNAs that were bound were displaced from the binding site of interest by picrotoxin, a well-known convulsant and non-competitive inhibitor of $GABA_A$ receptors.
T. californica Electric Organ	Membrane fragments of nicotinic Acetylcholine Receptor (nAChR) were purified from nAChR-rich *T. californica* electric organ. Displacement of selected RNAs from the binding site was done by 1-phenyl-cyclohexylpiperidine (PCP), a nAChR non-competitive inhibitor that binds stronger than cocaine.
Transfected Human Embryonic Kidney (HEK)293S Cells	SV40 large T-antigen (TAg) was co-transfected for maximized cell surface density of recombinant GluR2 receptor, allowing better oligo targeting by exposure and therefore, higher chance of selection against GluR2. Bound RNAs were eluted by NBQX, a high-affinity competitive inhibitor of GluR2. Negative selection with HEK293 cells lacking the target was done to minimize unwanted binding oligos.
Salmonella enterica serovar Typhimurium	A specific MP target was not defined prior to the selection procedure. Counterselection was done against *E. coli* crude lysate of outer membrane proteins and lipopolysaccharides (LPS) to obtain *Salmonella*-specific aptamers. Analyses by South-Western blotting and mass spectrometry identified three MP targets of aptamers.

2.2.2.4 Selection of Aptamers against Whole Living and/or Functional Biological Systems

Complex-target SELEX has bolstered exponential acceleration in the number of novel aptamers being selected for MP targets (Mallikaratchy, 2017). This was not achieved only by fostering non-purification-based SELEX against MPs in their native states, but also by **1)** permitting higher degrees of freedom that favor continuity of selection efforts, for the most part, irrespective of the availability of knowledge on the identity or functionality of MP targets, and **2)** allowing simultaneous screening for multiple aptamers against multiple targets in a given complex sample (Berezovski, Lechmann, Musheev, Mak, & Krylov, 2008; Blank, Weinschenk, Priemer, & Schluesener, 2001; Mallikaratchy, 2017). Together, these factors contributed to the currently witnessed gross rapidity in aptamer discovery by complex-target SELEX. The strength of complex-target SELEX and the outcomes of its use (hundreds of aptamer selections) were paved theoretically by mathematical models confirming the validity of the idea of simultaneous generation of ligands against multiple targets in a given complex sample (Levine & Nilsen-Hamilton, 2007; Vant-Hull, Payano-Baez, Davis, & Gold, 1998) and experimentally by selection trials with samples such as RBC ghosts thereby confirming the theoretical results and presenting for the first time, a new horizon of possibilities for the selection of aptamers against MPs of interest whilst in their physiological environment and without much of regard to the level of familiarity with these proteins (Hicke et al., 2001; Morris et al., 1998; S. Ohuchi, 2012).

The versatility of complex-target SELEX as a selection technique kept on affirming with each study published. It has continued to acquire improvements and modifications gradually. Designations that describe the process more specifically have emerged too. For example, “live cell-SELEX” (or simply, cell-SELEX) is a famous and well-established complex-target selection method possessing a nomenclature that is more descriptive of the type of sample being targeted for selection (i.e. whole living cells). The introduction of this technique was

revolutionary because it simplified the SELEX method and allowed more laboratories to extensively target cells using aptamer-based probes (Mallikaratchy, 2017). Since its introduction, the use of cell-SELEX has been on exponential growth for the identification of novel aptamers. **Figure 2.1** outlines a typical cell-SELEX scheme. The term cell-SELEX itself was coined first in an article published by late 2001 describing aptamers against tenascin-C (TN-C), a large extracellular matrix hexameric glycoprotein that is known to overexpress during processes of tissue remodeling such as embryogenesis, angiogenesis, atherosclerosis and wound healing, and by tumor cells to support their own growth (tumorigenesis) (Hicke et al., 2001). A hybrid selection scheme (expounded later) was carried out against both purified recombinant TN-C and TN-C-expressing U251 glioblastoma cells. Earlier studies, although published later than 2001, have selected an aptamer GBI-10 that was identified to bind a 250 KDa target—TN-C; however, the aptamer showed considerably weakened affinity to its target at physiological temperature (37ºC) (Daniels, Chen, Hicke, Swiderek, & Gold, 2003). There, the procedure was referred to as tumor cell-SELEX. Although the original use of the term "cell-SELEX" can be traced back to the 2001 article, targeting whole biological systems for aptamer selection had already been in conduct in several preceding studies such as the selection of RNA aptamers against an unidentified 42 KDa protein located specifically within the flagellar pocket of live African trypanosomes—unicellular protozoan parasites (Matthias Homann & Göringer, 1999), the selection of DNA aptamers that bind to the surface of anthrax spores (Bruno & Kiel, 1999) and the selection of ribonuclease-stabilized RNA aptamers that bind the initially unidentified B and H glycoproteins of the infectious human cytomegalovirus in which the inhibition of viral infection in tissue cultures was shown (J Wang, Jiang, & Liu, 2000). The aptamers selected in the latter study have converged to predominantly target the most abundant and exposed proteins (B and H), whereas, interestingly, in the former study, the selection had not favored the VSG protein, the most abundant protein on the surface of *Trypanosoma brucei*. Moreover, only a few types of

aptamers were generated given the significant number of targetable polypeptides on the surface of the parasite; suggestive of the presence of dominant epitopes.

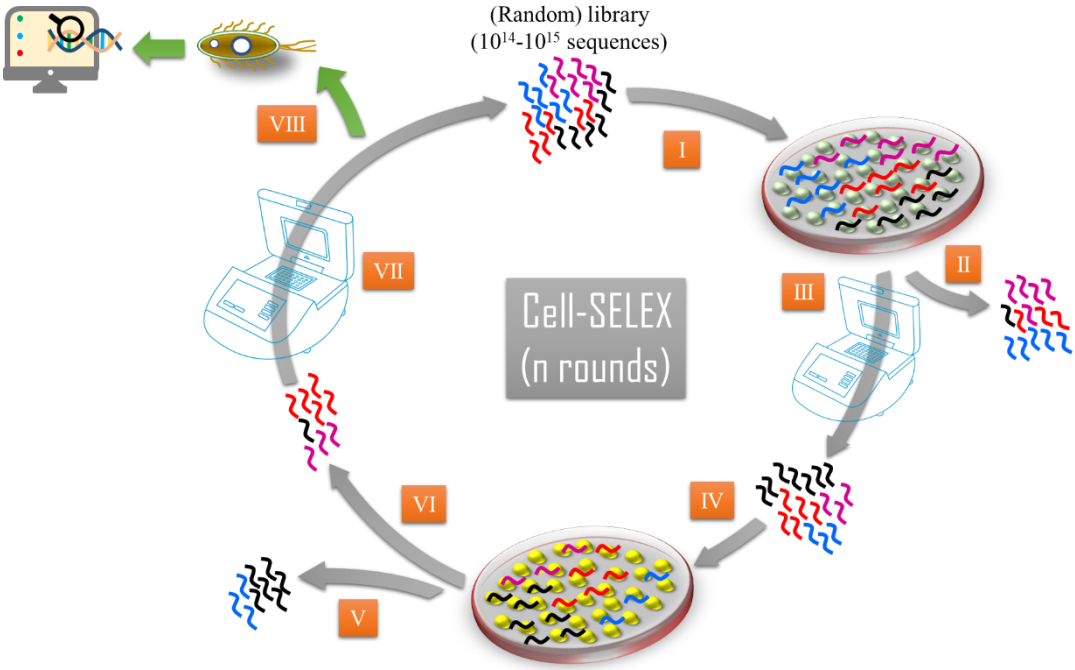


Figure 2.1. A typical cell-SELEX scheme. Step I, a nucleic acid initial library is incubated with whole cells for the positive selection of oligos. **Step II**, unbound ligands are partitioned (discarded) from the unbound ones (collected). **Step III**, bound ligands are then amplified by PCR or *rt*-PCR if RNA oligos are used in the library. **Step IV**, the positively selected and PCR-amplified oligo pool is then incubated with a different cell type for the counter selection against nonspecific binders. **Step V**, bound ligands are discarded and, **Step VI**, unbound ones are collected, **Step VII**, amplified, and used in the next cell-SELEX round. In the final round of selection, the enriched aptamer pool is typically cloned, sequenced, and analyzed for binding affinities, **Step VIII**.

Cell-SELEX method targeting whole living mammalian cells was attempted for the first time in early 2001 (Cibiel et al., 2011) against Adenovirus-12 SV40-transformed YPEN-1 rat endothelial cell line, a pathologic complex target sample (Blank et al., 2001). High affinity fluorescent single-stranded DNA ligands (aptamers) were selected to act as a histological probe for the detection and staining of neoangiogenic micro-vessels of rat experimental glioblastoma, a highly

vascularized and lethal brain tumor. However, like selections against MPs in the context of membrane preparations, cell-SELEX against the transformed YPEN-1 cells confronted by a lot of targetable cell surface components that may be shared with other cell types. Therefore, the study employed counterselection prior to every round and against the N9 microglial cell line (a population of brain monocytes) to minimize co-selection to other cell types (particularly macrophages) and obtain aptamers that better suited to differentiate YPEN-1 endothelial cells. Partitioning of the bound from unbound sequences is commonly done by gentle plate washing, for adherent cells (Cibiel et al., 2011). However, since cultured YPEN-1 cells have a phenotype that is characterized by non-adherence, partitioning was done by centrifugation, a separation method that is known and used for such cell types. A deconvolution-SELEX approach was used where individual aptamers against YPEN-1 cells were analyzed quantitatively for fluorescence and further analyzed on cryostat tissue sections of C6 rat brain glioblastoma tumors. The deconvolution protocol used had allowed for the evolution of discriminatory binding of the aptamers to favor neoangiogenic micro-vessels over normal rat brain vasculature including peritumoral areas. In this study, all SELEX procedures were monitored and analyzed by fluorescence-based methods, flow cytometry and fluorescence microscopy. The molecular target facilitating histological discrimination between the two differentiated states of normal and cancerous tissue types by the selected aptamer was identified for the first time (Janas & Janas, 2011) by peptide mass fingerprinting using mass spectrometry (following ligand-mediated magnetic DNA affinity purification). The target was identified to be the rat homolog of mouse pigpen, an endothelial protein shown to be highly upregulated in tumor micro-vessels of experimental rat brain glioblastoma.

The entirety of the molecular targets described earlier in **subsection 2.2.2.4** has had novel aptamers selected without any prior knowledge about the target identity. Although identifying unknown aptamer targets can be a difficult task and one that requires significant additional efforts (Cibiel et al., 2011), complex-target SELEX by using whole biological entities and against unidentified molecular targets

was means for opening the promising path of “Aptamer-facilitated Biomarker Discovery” or simply, “AptaBiD” (Berezovski et al., 2008). AptaBiD is a comprehensive technological avenue with extensive research potential and prospects of application. Different protocols have been used to isolate and identify the protein targets of selected aptamers (Mallikaratchy, Zumrut, & Ara, 2015). While the systematic approach of aptamer generation and target isolation and identification used in AptaBiD studies allows the discovery of novel and specific biomarkers, the real power of complex-target selection and specially cell-SELEX, is demonstrated by its ability to generate a panel of aptamers that can, unlike conventionally available methods, distinguish different subpopulations of the same cell type (Sola et al., 2020). This has been shown elegantly by (Blank et al., 2001) in the last example involving the pigpen aptamer and by many others (Jia et al., 2016; Y. Kim et al., 2013; Sefah et al., 2013; Shangguan, Cao, Li, & Tan, 2007; Shangguan et al., 2006; Yang et al., 2014). A specific cell phenotype rather than a single molecular biomarker is identified by uncovering differential expressions between different but very closely related biological systems (Catuogno & Esposito, 2017; Cibiel et al., 2011). This type of selection is sometimes referred to as differential-SELEX and such molecular profiling opens doors for cell phenotyping. These molecular fingerprints are useful in areas such as dissecting and understanding biological systems and processes that occur within such as development and differentiation, or in distinguishing between normal and diseased states of cells which translates directly into diagnostic and therapeutic applications. Refer to **sections 2.1.2** and **2.2.1** for the general applications of aptamers as well as their applications specifically to MPs, respectively.

Applying the Live-cell SELEX method does not intrinsically impose the notion of aptamer selection against exclusively unknown and highly expressed cell surface targets. Similar to the examples of complex-target SELEX directed against predefined MP targets in membrane preparations given in **Table 2.3** and **Table 2.4** (except for *Salmonella*), selection against a predefined target localized in the natural hydrophobic environment of the lipid bilayer of a whole living cell is feasible with

live cell-SELEX. In fact, cell-SELEX, more commonly, involves recombinant overexpression in a cell line of choice the MP of interest that is specifically intended for selection (Dua et al., 2011). By the exogenous high expression of the that MP, the ground is set for a SELEX procedure that is very likely to target the protein of interest. After growth and proliferation in cell culture, the living cells harboring the recombinant MP are used as a positive selection target by incubation with the naïve library. Counter (or negative) selection is of course critical to minimize all unwanted selections. This form of cell-SELEX, developed and termed "target expressed on cell surface-SELEX" (or TECS-SELEX), was precisely devised to bypass the difficulties and shortcomings associated with aptamer selection against purified cell surface proteins (S. P. Ohuchi et al., 2006). TECS-SELEX was reported for the first time in a selection of RNA aptamers against transforming growth factor-β (TGF- β) type III receptor (TbRIII) that is recombinantly expressed and displayed on the surface of Chinese hamster ovary (CHO) cells. The counter selection was carried out using parental mock CHO cells. One of the aptamers selected was identified to have a 1 nM dissociation constant and it also competed with TGF-β during in vitro cell surface binding assays. An article published in 2009 employed the same principle to select against live mammalian cell line CT26 overexpressing the envelope glycoprotein E2 of hepatitis C virus (HCV), the molecular target of selection (F. Chen, Hu, Li, Chen, & Zhang, 2009). In the article, the approach was termed "alive cell surface-SELEX" (CS-SELEX). In a different study (Kang, Huh, Kim, & Lee, 2009), cell-SELEX was performed against a human breast cancer line, SK-BR-3, known to overexpress the human epidermal growth factor receptor 2 (HER2). The study presented a novel strategy for counterselection where cells of the same breast cancer cell line were treated with HER2-targeting small interfering RNA (siRNA). Using the siRNA-based method, an opportunity is presented to perform counter-SELEX against the same cell line suppressed for the target of interest and thus, making counterselection more ideal by avoiding the use of unrelated cell lines. In the same study, a counterselection using the HER2-negative breast cancer cell line MD-AMB-231 was performed in parallel; both counterselections resulted in aptamers

that are highly specific for HER2-positive cell lines (SK-BR-3 and NIH-3T3). This indeed confirms the effectiveness and broad applicability of the siRNA-based counter-SELEX method.

When exploring aptamers against complex-target samples [e.g. whole living cells, fixed cells, tissue samples and cell lysates (Kaur, 2018)], multiple cons present themselves during the selection process (Dua et al., 2011; Trujillo, Majumder, Gonzalez, Moaddel, & Ulrich, 2007). For instance, a high selection rounds number may be required to compensate for the impairment in pool deconvolution rate resulting from the complexity of the sample at hand. This was reported in the pioneering SELEX investigation against RBC ghosts (Morris et al., 1998). Due to an increased background binding in such samples, a slow pool convergence is observed. Hence, increasing selection rounds tries in principle to minimize sequence complexity in the final pool and to maximize the odds of having well-converged (highly specific) aptamers. It is noteworthy that even after a high number of selection rounds, generated aptamers may still have an affinity for other targets as Morris et al. have observed. Another disadvantage is having a compromised affinity of the selected aptamers against a target of interest due to the lack of knowledge about its concentration in the complex mixture. The target, having an unknown concentration in the sample, does not necessarily represent the dominant molecular epitope and in turn, this causes it to be disfavored for the selection of high-affinity ligands that bind other predominant targets. Although the ectopic expression, as discussed earlier in TECS-SELEX, can be a worthy solution to this problem, the induction of other MPs may become a complication upon overexpression of the MP of interest. These induced MPs can override the presented target pool with their dominance. Counter-SELEX is an indispensable tool and a well-established method of choice that can alleviate by a large degree many negative effects that harm aptamer affinity and specificity (T. Wang, Chen, et al., 2019). However, to expand the arsenal of selection procedure enhancers, a two-stage selection method known as hybrid- or crossover-SELEX was developed first by Hicke et al. in 2001 during the TN-C aptamer selection trials. Next to counterselection, crossover-SELEX has also proven itself to

be an effective and powerful strategy to resolve the issues mentioned above and in enhancing the efficiency of aptamer screening (C. Pestourie, Tavitian, & Duconge, 2005). Using crossover-SELEX, aptamer affinity and specificity are maximized with less rounds of iterative selection, thereby reducing the time and cost of aptamer production (Trujillo et al., 2007). This improvisational method synergizes both cell-SELEX and traditional protein-SELEX to reap the combined benefit of the individual methods; hence the designation, "crossover-SELEX". Throughout the literature, crossover-SELEX can be seen executed by performing cell- and protein-SELEX sequentially and in either of the orders. Crossover-SELEX and counter-SELEX can be complemented with each other for a more efficient selection (Carine Pestourie et al., 2006). Perhaps it is clear by now that crossovers require knowledge on the identity of the MP of interest and cannot be used with unidentified targets (Dua et al., 2011). To implement crossover-SELEX, Hicke's lab has subjected the TN-C-expressing U251 glioblastoma cells to 9 rounds of cell-SELEX followed by only 2 selection rounds against the purified protein. By factoring in 2 additional rounds of protein-SELEX with the purified form of TN-C, the aptamer pool is further enriched to obtain rare high-affinity aptamers in the pool (M. Chen et al., 2016). Using crossover selection, the investigators managed to isolate two high-affinity aptamers with a 50-fold increase in the K_d value observed after the additional 2 rounds. A different group also demonstrated their use of crossover-SELEX (Carine Pestourie et al., 2006). A 7th round aptamer pool obtained by selection against PC12 cells expressing a mutant transmembrane receptor tyrosine kinase RET^{C634Y}, PC12/MEN2A, and also by a counterselection against PC12 cells, was chosen due to its enrichment with sequences against cell surface targets and without the reduction in sequence diversity. The pool was fed into a crossover selection, called S4, involving the purified recombinant extracellular domain of RET^{C634Y}. However, unlike the crossover selection employed for TN-C by Hicke et al., S4 was presumably incapable of identifying conserved epitopes between the purified and the membrane-embedded version of RET^{C634Y}. That is, although a very rapid evolution of the population was shown by RFLP analysis just after 2 rounds, most of the

sequences have predominantly enriched against the purified protein and had low affinities towards the native form. Only two aptamers from this population had shown good affinities to PC12/MEN2A cells and both of which still had affinity values lower than those identified for aptamers obtained by plain cell-SELEX trials (S1 and S2). In conclusion, the examples addressed above bring to light dependency of the outcome of a crossover-SELEX experiment on an array of elements that may include the conformational identity of the protein of interest in its native and purified form as well as all certain aspects of the experimental plot and its results (e.g. features of generated aptamer families). **Table 2.5** outlines reports that have effectively used crossover-SELEX (also referred to as reverse crossover-SELEX when selection begins with a purified target and not a complex target). Reports' references: (Boltz et al., 2011; Soldevilla et al., 2016; Wilner et al., 2012; G. Zhu et al., 2017)

Table 2.5 Reports effectually applying crossover-SELEX [References: (Boltz et al., 2011; Soldevilla et al., 2016; Wilner et al., 2012; G. Zhu et al., 2017)]**.**

Stage-1 Selection Target	***Stage-2 Selection Target***	***Stage-1 Selection Pool Chosen***	***Stage-2 Selection Rounds***	***Comments***
CD16α Receptor	CD16α-positive Jurkat cells	3rd and 5th (pre-enriched pools)	6 rounds (for each pool)	Out of over 25 CD16α-specific aptamers, two were CD16α-positive cells-specific. One of them was identified without stage-2 selection and was found in the 9th round pool of protein-SELEX.
Human Transferrin Receptor (hTfR) CD71	TfR-expressing HeLa cells	4th	1 round	Round 5 aptamer population displayed a 10-fold better binding to cells. siRNA-containing nucleic acid lipid particles (known as SNALPs) were functionalized by a minimized aptamer candidate for internalization and gene silencing.
Multidrug Resistant-associated Protein 1 (MRP1) Ectodomain	MRP1-expressing H69AR tumor cells	10th	1 round	MRP1Apt, a high-affinity aptamer (50 nM), was selected and used to engineer a bispecific MRP1-CD28 aptamer for the delivery of CD28 costimulatory signals to tumor-infiltrating lymphocytes following binding to MRP1-expressing tumors.

Table 2.5 (Cont'd)

Human Epidermal growth factor Receptor 2 (HER2) Ectodomain	HER2-expressing SKOV3 tumor cells	8th	7 rounds	Enhanced binding was observed during pool enrichment. Based on homology and frequency, 7 out of 2 million sequences were chosen for downstream studies. Affinity and specificity to HER2-postive cells were confirmed. Two were identified for their robust HER2-specific cancer *in vivo* imaging potential.

2.2.2.5 Perpetuating the Innovation in Complex-target SELEX Methodology to Targeting Membrane Proteins

The emergence of a variety of modified cell-based aptamer screening methods greatly expanded the range of targetable cell surface residents and provided new dimensions for aptamer selection with higher success rates (M. Chen et al., 2016). Notably, cell-based SELEX has found extensive utilization especially in the area of oncology as it has become the first method of choice to generate aptamers against cancer cells' surface biomarkers for applications in diagnostics and therapeutics. Limitations of conventional protein-SELEX are defeated by supporting aptamer selection against important MPs while in the native setting of the cell's surface and adopting correct PTMs and 3D conformations, allowing selection of biomedically relevant aptamers that recognize targets in their functional states. Moreover, the screening procedure does not even require prior knowledge on the potential target(s) and can generate aptamers for multiple targets simultaneously (Cibiel et al., 2014), thus, realizing applications in AptaBiD. These merits take us to the next one and that is bypassing tedious and difficult MP purifications in pure, high-quality and satisfactory amounts before engaging in selection. Despite all the advantages and potential of cell-SELEX (and its developed variants), the methods

are not without pitfalls and limitations. The most obvious is the requirement to maintain all necessary cell lines throughout the selection procedure (Mallikaratchy, 2017). Moreover, due to the complexity of cell-based selection methods, aptamers can be lost during the procedure. Other times, a certain protein cannot even be targeted altogether (Goto et al., 2017). In another scenario, unlike protein-SELEX which generates high-affinity aptamers very fast due to singularity and purity of the target, heterogeneity of the cell-based targets and the components of the membrane can lead us to include a higher number of selection rounds (Kaur, 2018) that can go as far as 35 rounds (Darmostuk, Rimpelova, Gbelcova, & Ruml, 2015) to arrive at aptamers with high affinities and specificities. Consequently, this highly risks, for example, the introduction of enzymatic amplification reaction bias (Zhou & Rossi, 2014). Such additional rounds not only elevate costs but also make selection lengthier, more laborious and time-consuming (M. Chen et al., 2016). Other additional series of experimentation such as counter- and crossover selection are important to increase selection efficiency of and to evolve competent aptamers. However, these can also pose similar disadvantages. Rarely expressed MPs represent another issue as they are difficult to target for selection (Kaur, 2018) and while utilization of the cell-SELEX variant TECS-SELEX can overcome this problem, expression profiles and target abundance can be altered (Mercier, Dontenwill, & Choulier, 2017) as it was pointed earlier in **subsection 2.2.2.4**. However, it also possible that overexpressed gene products may not affect selection against the desired target because of the high efficiency of recombinant expression (T. Wang, Chen, et al., 2019). Following the selection of novel aptamers, target identification is a crucial post-cell SELEX endeavor and a problem that must be tackled. Unfortunately, uncovering identities of the aptamer targets is still a complex process and we are yet to become efficient in it (Yan et al., 2019). Only a few aptamer targets have been identified in comparison to the hundreds of aptamers that were selected by live cell-SELEX (Mallikaratchy, 2017). This goes back to the inherent challenges put forth by MPs against their isolation, identification and characterization (Kaur, 2018); isolation of MPs has been addressed at the start of **section 2.2.2**. Nonetheless,

improvements are continuously made; namely in methods for MPs isolation and adaptation of mass spectrometry to their hydrophobicity. It may be inferred by now that cell-based aptamer selection generally has a higher degree of complexity (M. Chen et al., 2016; Kaur, 2018) and requires a certain level of technical expertise in addition to being expensive and time-consuming (Zhang, Lai, & Juhas, 2019). Throughout the remainder of this discussion, interesting advancements in aptamer screening technology are introduced, each of which is developed to add new and unique advantages to and overcome limitations of the classical cell- and protein-based selection methods.

Targeting Cell-Surface Carbohydrates

One specific difficulty that is faced with selection against cell surface carbohydrates is the presence of negatively charged moieties (Cibiel et al., 2011). A very limited number of aptamers have been selected against these important cell surface biomarkers and the difficulty arises due to the charge repulsion that can occur between the negative charges often present in glycans and those present in the targeting ligands. The first successful anti-carbohydrate aptamer was developed to target the MP-associated tetra-saccharide, sialyl Lewis X (sLe^X), a ligand of selectin proteins that plays critical roles in inflammatory cellular adhesion and metastasis in different cancers (Jeong, Eom, Kim, Lee, & Yu, 2001). The selected RNA aptamer had an affinity to its target in the sub-nanomolar range, a specificity that extends to include sLe^X-related sugars but not dissimilar sugars such as lactose, and an inhibitory effect on sLe^X-mediated adhesion of HL60 cells to E- and P-selectins. To select against negatively charged carbohydrates, a novel strategy was adopted in 2004 where DNA oligonucleotides modified with a charged cationic functional group (a protonated-amino group at the C5 position of thymidine residues) were used to enhance binding to sialyllactose that bears an anionic carboxyl group (Masud, Kuwahara, Ozaki, & Sawai, 2004). The selected modified DNA aptamer had a dissociation constant of 4.9 μM.

FACS-SELEX

As seen earlier, a variety of specialized devices such as flow cytometers and mass spectrometers are adopted and used effectively during cell-based aptamer discovery; they have become well integrated into the process of selection (Yan et al., 2019). One of the most recognized issues with classical cell-SELEX that impeded implementation of the technique, is the compromised membrane integrity or death of cells—specifically for cells in suspension—during the separation of the bound from unbound species by centrifugation (Raddatz et al., 2008). Damaged cells show ligand binding that is sequence-independent, strong and nonspecific, and can also unselectively internalize such ligands. Centrifugation cannot discriminate between vital and dead cells. Consequently, nonspecific ligands are fed into successive SELEX rounds due to inefficient separation of target-bound and -unbound species making the enrichment of high affinity and specificity sequences progress poorly (slow pool convergence) due to contamination with nonspecifically binding dead cells. Furthermore, with a higher round number, the pool highly risks drifting off to favor enrichment of the unwanted nonspecific aptamer population over target-specific aptamers and eventually, failing of the SELEX experiment. Ideal separation procedures ensure efficient separation of the target-bound species from unbound and nonspecific binders without disrupting target-ligand complexes or favoring isolation of nonspecific ligands. To this end, Raddatz et al. have developed a novel and sophisticated approach by integrating into cell-SELEX, an advanced flow cytometry technique called fluorescence-activated cell sorting (FACS) that can separate subpopulations of cells from other subpopulations in composite cell mixtures. The selection method, given the name, "FACS-SELEX", uses a digital high-speed fluorescence-activated cell sorter device to separate vital and dead Burkitt lymphoma B cells based on their light-scattering characteristics after incubation of a DNA library with the composite cell suspension. 10 rounds of FACS-based selection were enough to isolate C10, a high-affinity aptamer (K_d = 49.6 nM) with enhanced binding and distinguishability for vital $CD19^+$ Burkitt lymphoma B cells over primary B cells. Conversely, 20 rounds of classic centrifuge-based selection with the tumor

cells yielded no enhanced binding to vital cells. 6 additional FACS-based selections with the obtained nonspecific library yielded a library that binds to vital Burkitt lymphoma cells, demonstrating the method's ability to eliminate false positives. The state-of-the-art method is, therefore, high-throughput, more efficient, and effective (but cost-intensive) and can also include both positive and counter selections in one round by sorting out the negative and positive cells leading to higher efficiency (Mallikaratchy, 2017). A step-by-step detailed protocol for FACS-SELEX was later published by the group (Mayer et al., 2010). Reports employing the method include aptamers isolated against epithelial cell adhesion molecule (EpCAM), a transmembrane glycoprotein detected in most adenocarcinomas and cancer stem cells (J. W. Kim et al., 2014), and aptamers recognizing white mature adipocytes without preadipocytes, primary brown adipocytes and other common cell lines (E. Y. Kim et al., 2014). In another article, a strategy was established by Meltem et al. to eliminate dead cells and reduce the enrichment of unspecific sequences contaminating selection rounds (Avci-Adali, Metzger, Perle, Ziemer, & Wendel, 2010). It was suggested that dead cells in a cell population detached into suspension could be removed prior to incubation of that suspension with the random library. In cell-SELEX experiments, the strategy enables enrichment of specific aptamer sequences as target cells have high viability. Also, during counterselections, loss of target-binding sequences that would have favored nonspecific binding to co-present dead (negative) cells is avoided. Therefore, this strategy can greatly complement and enhance FACS-SELEX as the removal of dead cells before incubation of the cell suspension with the library avoids loss of target-binding sequences by cell sorting. FACS procedure itself, however, may not be the best method to remove dead cells (by staining damaged cells for elimination) because these dead cells are continuously generated during sorting due to the pressure and shearing force of the procedure that damage viable cells and thus produce new dead cells persistent in the collected cell suspension. The gentler method devised by Meltem et al. for the removal of non-viable cells involves the optimized detachment of adherent cells by 2 mM EDTA, centrifugation at low speed to discard the majority of these cells and use of dead cell

removal microbeads for magnetic depletion of remaining dead cells. The proportion of dead cells was reduced by this method down to 5.2%. A different strategy suggested to decrease numbers of unspecific sequences during selection is to run a counterselection against dead cells, a simple and cost-effective approach (Darmostuk et al., 2015). Apart from cell suspensions, directly using attached cells for cell-SELEX is a viable option due to its ease and efficiency as reported by many articles (T. Wang, Chen, et al., 2019). By using attached cells, weak and dead cells are eliminated during stringent washing, cell damage during cell suspension preparation is eliminated, and nonspecific binding of the library to incubation containers is minimized.

Tissue Slide-Based SELEX

The concept of targeting complex mixtures for nucleic acid ligand selection was patented by the U.S. under the designation, “tissue SELEX” and comprised a wide range of targetable biological entities (Cibiel et al., 2011; Li et al., 2009). However, for over a decade following its utility by Morris et al. in 1998, “tissue SELEX” employed particularly in generating tumor-specific aptamers had used for selection mainly cultured cancer cell lines and separated cells. Moreover, such samples are susceptible to the *in vitro* variables which are not always easy to adjust towards mimicking the physiological conditions (Sola et al., 2020). Tumor pathogenesis, for example, is influenced by many factors such as the tumor microenvironment of blood vessels and constituents of the ECM. Thus, even cell-SELEX using *in vitro* cultures can undermine the significance of *in vivo* conditions and consequently, due to the high specificity of aptamers, leading to the evolution of ligands that fail to recognize and reflect the “real” biomarkers of cancer (or any other targeted tissue). To resolve this problem, Li et al. developed the novel approach termed “tissue slide-based SELEX” which uses clinical specimen sections that are more representative of the underlying disease. Pathological neoplastic tissues of breast cancer were screened *in situ* for aptamer selection after tumor tissue excision from breast infiltrating ductal carcinoma patients and fixation as paraffin-embedded tissue slides. For counterselections, clinical specimens were also taken from adjacent

normal tissue of the same patient and slides were prepared. After 12 rounds of tissue slide-based SELEX using paraffin tissue sections, a high-affinity aptamer, BC15, emerged and is shown to recognize different types of breast cancer in clinical tissue sections and breast cancer cell lines. The aptamer target was identified to be the subcellular heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1), a well-studied protein that plays an important role in breast cancer tumorigenesis. Preserving tissue samples in their diseased state in slide-based preparations can greatly support the use of aptamer probes for tumor imaging, pathological diagnosis, therapy and revealing molecular mechanisms that bring about the diseases under investigation. Tissue slide-based SELEX can also be used to isolate aptamers for all fractions of the tissue such as the ECM, membrane components and intracellular targets. However, the main disadvantage of this technique is the possible denaturation of protein targets during tissue-slicing, immobilization and chemical staining (Goto et al., 2017). A recent article has utilized the same method for identification of an aptamer that can perform specific target recognition of papillary thyroid carcinoma (PTC) in cell lines, tissue slides and tumor-bearing mice through an unidentified cytoplasmic target (Zhong et al., 2019).

IP-SELEX

It has been observed earlier strategies are generally used to drive selection towards a known and specific target of interest in a complex-target sample. These strategies include exploiting TECS-SELEX (S. P. Ohuchi et al., 2006) where cell lines known to already overexpress the target of interest are used, counterselections where the target molecule is favored over other shared components, and/or using known high-affinity ligands of the target to elute the relevant and desired aptamers from that target. Another strategy developed to facilitate selection against a predefined target couples SELEX with an immunoprecipitation (IP) step in the novel "IP-SELEX" method (Chang et al., 2009). The method was used first against Toll-like receptor 2 (TLR2) to isolate high-affinity functional DNA aptamers that can suppress nuclear factor NF-κB activity and significantly impede cytokine secretion. IP-SELEX has a number of advantages over cell-based and conventional protein-

based SELEX methods. It is more efficient and can generate aptamers after a low number of rounds (7 rounds by Chang et al.). Also, except for the first round in which the random oligonucleotide library is incubated with intact cells to ensure reaction only with the extracellular domain of the intended MP and not its tagged intracellular domain, IP-SELEX does not require maintenance of a cell culture during selection in every round. Instead, starting from the 2nd round, the library is incubated with cell lysates. In the IP-SELEX paper, recombinant Fc fragment-fused TLR2 was transiently expressed in HEK293T cells. After incubation with the library and completion of the binding reaction, intact cells of the 1st round are lysed by a mild surfactant such as Nonidet P-40 or Triton X-100. Cell lysates are then used in IP of TLR2-Fc/aptamer complexes by protein A or antibody-coated beads; other constituents of the lysate serve for counterselection. Elution of the complexes from the beads was carried out using citric acid (pH 4) and the aptamers were PCR-amplified and used in the next cell culture-less round. With IP-SELEX, there is no prior need for costly and cumbersome protein purification as is the case with protein-SELEX; only transient expression of the target is required. However, it is important to note that MPs that require strong or even denaturing detergent treatments for solubilization might require a different approach. Nonetheless, IP-SELEX, unlike conventional protein-SELEX where solid-phase immobilization of the target on beads or polymer membranes is carried out, features liquid-state hybridization, which adds to the efficiency of selection. The method has been used to isolate an anti-CD8 aptamer against cytotoxic T lymphocytes (CTLs) to inhibit immune disorders (Mercier et al., 2017; C. W. Wang et al., 2013).

In vivo-SELEX

The vision of selection carried out inside biological systems has been around before the complex-target SELEX performed by Morris et al. in 1998. The first published report regarding this type of selection showed the transfection of an infectious library of HIV-1 DNA genomes with random mutations into CD4+ T cells and selection of replication-competent viruses after multiple rounds of retroviral replication (Berkhout & Klaver, 1993). In another report, an approach was

established to select the prevalent from the intracellular exon enhancer sequences—RNA-processing signals that function during exon recognition for pre-mRNA splicing—in cultured vertebrate cells (Coulter, Landree, & Cooper, 1997). Whether purified targets, subcellular preparations or cultured whole cells represent an accurate replica model of the *in vivo* is a question that is brought up constantly (Mi et al., 2010). "*In vivo*-SELEX" was developed by Mi et al. to select aptamers in a completely physiological environment as aptamer binding and applicability depend on the conformations of targets which is in turn, conditioned by the environment they are in. Tumor-bearing mouse models of intrahepatic colorectal cancer metastases were injected intravenously with a large library of nuclease-resistant (2′-fluoropyrimidine-modified) random RNA oligonucleotides. The liver tumors were surgically harvested, followed by extraction and amplification of the injected RNA molecules. The RNA pool was then re-injected for a new round of selection. 14 rounds of *in vivo* selection yielded an aptamer, RNA 14-16, with more than a threefold higher affinity (K_d = 30.8 nM) for aggregate CT26 tumor proteins compared to the initial library (K_d = 97.2 nM). By tumor fluorescence staining, Cy-3 labelled RNA 14-16 was able to exclusively localize to colon carcinoma *in vivo* and therefore, selectively recognize it apart from normal colon tissue, to which it had an affinity of only 2.2 µM. By running gel analyses (gel-shift assay and SDS-PAGE), peptide-mass fingerprinting and MS/MS peptide fragment ion matching, a 70 KDa tumor-specific protein was identified to be, surprisingly, the intracellularly localized mouse p68 RNA helicase, known to be deregulated and overexpressed in colorectal tumors. Therefore, RNA 14-16, like the aptamer selected for nucleolin (an RNA helicase) (Soundararajan, Chen, Spicer, Courtenay-Luck, & Fernandes, 2008), can readily gain access into cancerous cells. Furthermore, similar to the nucleolin aptamer, RNA 14-16 was shown to interfere with its target's function as it binds and effectively inhibits the ATPase activity of p68. Such aptamers are strongly applicable in diagnostics, treatments and cargo delivery to the tissues of interest. During *in vivo*-SELEX (also called "live animal-SELEX"), nonbinding sequences are simply cleared from the circulation by the kidneys and counter selection is intrinsically

included in the procedure (Sola et al., 2020). Nonspecific aptamers bind to nontarget organs after injection and can be analyzed via NGS to be withdrawn as nonspecific. It is however, also feasible to introduce counterselections by injecting the library into healthy mice after organ harvest from which isolation and amplification of the positively selected sequences can be done (Zhang et al., 2019). Consequently, disease-specific aptamers are enriched to have high affinity and specificity to the target tissue. Aptamer screening *in vivo* is attractive but has several major complications including the need for larger, more complex and unbiased libraries, the limited ability of polymerases to amplify modified oligos, the criticality of the organ harvest time due to the clearance of injected aptamers from the body, and the requirement for longer periods when running *in vivo*-SELEX in medium-to-large-sized animals as a result of circulation of the injected library in the bloodstream. Other drawbacks include the procedure's high invasiveness, the increased costs due to extraction, and the significant safety risks (Goto et al., 2017; Sola et al., 2020; Yan et al., 2019). More recent applications of *in vivo*-SELEX involve blood-brain barrier (BBB)-penetrating aptamers that were selected in mice (Cheng, Chen, Lennox, Behlke, & Davidson, 2013), bone-targeting thioaptamers selected from mice with prostate cancer bone metastasis (L. Chen et al., 2019), thioaptamers with affinity to bone marrow endothelium in a murine model of lymphoma with bone marrow involvement (Mai et al., 2018), and PEGylated aptamers isolated in xenograft mouse models of non-small-cell lung cancer cells NCI-H460 (H. Wang et al., 2018) and hormone refractory prostate cancer (Civit et al., 2019).

Cell-Internalization SELEX

Cellular uptake and cytosolic translocation of oligonucleotide aptamers are hindered by their size and charge (Sola et al., 2020). However, aptamers remain small molecules and this allows them to access protein epitopes that might not be accessible to bulky antibodies (Dua et al., 2011). Smaller size also provides them with a better capacity for internalization but only when bound to a molecule that is internalized by the cell or to a surface receptor for which the aptamer acts as a ligand. This feature allows aptamers to be used as bifunctional molecules for, in addition to

ligand recognition, the delivery of payloads such as drugs and siRNAs into the cell. Many internalizing aptamers have been discovered; one of the earliest aptamers recognized to be taken up by cells belongs to the study on selection with African trypanosomes where the aptamer 2-16 was found to be taken up by receptor-mediated endocytosis upon which intracellular vesicular trafficking delivers the RNA molecules to the lysosome (M. Homann & Göringer, 2001; Matthias Homann & Göringer, 1999). Following examples on internalized selected aptamers and their promising biomedical applications included the development of the unique selection strategy against human lymphoma B cells for ssDNA aptamers with an enhanced uptake and immune-activation of primary human chronic lymphocytic leukemia B cells (C. C. N. Wu et al., 2003), the bioconjugation of PSMA aptamers to nanoparticles (Farokhzad et al., 2004) and siRNAs (Chu, Twu, Ellington, & Levy, 2006) and the subsequent uptake of bioconjugates by the prostate LNCaP epithelial cells for therapeutic deliveries, the selection of tRNA derivatives for their import into isolated yeast and human mitochondria for the treatment of mitochondrial disorders (Kolesnikova et al., 2010), and many more (Gourronc, Rockey, Thiel, Giangrande, & Klingelhutz, 2013; Y. Z. Huang et al., 2012; Mende et al., 2007; Z. Xiao, Shangguan, Cao, Fang, & Tan, 2008). By 2012, the term “cell-internalization SELEX” was introduced in a research article describing novel internalizing RNA ligands that can deliver therapeutic siRNAs to HER2-expressing breast cancer cells (K. W. Thiel et al., 2012). Later, a detailed protocol of the novel SELEX method was published wherein, briefly, incubation of the combinatorial library with the target cells in every round is followed by discarding both bound and unbound species and recovery of the internalizing sequences by cell lysis (W. H. Thiel et al., 2015). However, some drawbacks of the application of cell-internalization SELEX are the complexity and labor-intensive isolation process of the internalized aptamers (especially when counterselections are incorporated) which can themselves be lost by this method (Mallikaratchy, 2017). Likewise, studying intracellularly transported aptamers is time-consuming (Kaur, 2018). These disadvantages limit the use of cell-internalization SELEX. More recent publications exemplified cell-internalization

SELEX to enrich for rapidly internalizing aptamers targeting insulin receptor (Iaboni et al., 2016), lysosomal-associated membrane protein 1 (LAMP-1) (Ranches et al., 2017), skeletal muscle cells (Philippou et al., 2018), and B cell maturation antigen (BCMA) (Catuogno et al., 2019).

3D cell-SELEX

Standard two-dimensional (2D) or monolayer cell culture has been a tremendously essential tool in the development of modern biology and its wide range of applications (Haisler et al., 2013; Kapałczyńska et al., 2018). However, this tool is well-recognized for its inadequacy in many cases to recreate the *in vivo* states of cells and the microenvironment in which they naturally thrive. Altered cellular states and environments pose a great obstacle for various goals such as understanding cell biology (cellular vitality, growth, proliferation, differentiation, morphology, communication, signaling, responses to stimuli, gene and protein expression profiles, disease mechanisms etc.), performing drug screens, and the development of tissue engineering. To this end, *in vitro* 3D culture techniques have emerged as tools that provide better mimicry of *in vivo* cells and their physiological environment. 3D cell cultures were combined with aptamer selection in a novel method, "3D cell-SELEX", to improve biomarker discovery for human prostate cancer and consequently, improve diagnosis, prognosis and therapy (Souza et al., 2016). Spheroid cells of aggressive PC-3 prostate cancer cell line (and RWPE-1, non-tumor cell line) were formed in a 3D culture by the magnetic levitation method (MLM) that levitates incubated cells after applying a magnetic field. Nine rounds of selection (1st round with the negative cell line and 2nd–9th round with the positive cell line) with the spheroidal cells led to the selection of eight RNA aptamers with high affinity to a membrane protein target (K_d in the nanomolar range) and with specificity to prostate tumor cells. In conclusion, 3D cell-SELEX is a useful method because it improves the arrangement of cell surface proteins, the ECM and their exposure to the aptamer library. However, the method still needs improvement as it is time-consuming and needs technical laboratory skills to establish 3D cell cultures (Kaur, 2018).

Ligand-guided Selection

Later through the midyear of 2016, the cell-SELEX variant, "ligand-guided selection" (LIGS), was introduced for the selection of highly specific aptamers and without pre- or post-SELEX sample manipulations (H. E. Zumrut, Ara, Fraile, Maio, & Mallikaratchy, 2016). Using LIGS approach, pre-determined epitopes of cell-surface MPs are targeted for the selection of increasingly specific aptamers by using, in the partitioning step, an excess of a known secondary cognate ligand (e.g. antibody, growth factor, neurotransmitter, hormone, enzyme substrate etc.) that strongly binds to that epitope of interest. Binding of the stronger ligand is used to effectively outcompete and displace epitope-specific aptamers obtained originally from partially evolved cell-SELEX libraries. LIGS was introduced in a selection of aptamers against membrane-bound immunoglobulin M (mIgM) known to be expressed in Burkitt's lymphoma Ramos cells. After the partial pre-enrichment of an aptamer library against Ramos cells in cell-SELEX for 13 rounds, an anti-mIgM antibody model ligand was used as a secondary competitor during LIGS for the elution of a subpopulation of that library with high specificity for the antibody's target. Basically, LIGS is a way to enhance separation efficiency in SELEX by competition for specific epitopes as the strategy of choice (H. E. Zumrut, Ara, Maio, et al., 2016). However, the method is limited to known surface markers with known high-affinity ligands. For this reason, it finds limited use in biomarker discovery (Mallikaratchy, 2017). Additionally, the affinity of the selected aptamer is unlikely to surpass that of the original ligand (Sola et al., 2020). More follow-up studies on LIGS can be found here (H. E. Zumrut & Mallikaratchy, 2020; H. Zumrut et al., 2020).

Isogenic cell-SELEX

"Isogenic cell-SELEX" or in short, "Icell-SELEX", is yet another selection strategy that has been developed to promote the use of isogenic cell lines for cell-based selection and counterselection to generate aptamers with high affinity and specificity to a predefined cell-surface MP target (Takahashi, Sakota, & Nakamura,

2016). Cell lines known to overexpress an MP target of interest (for selection) or TECS-SELEX employing artificial overexpression, both of which combined with the use of appropriate mock cells (for counterselection), has been the widely used valid and feasible method for the selection of aptamers against specific cell surface targets. However, most cell-SELEX studies use non-isogenic mock cells (sometimes from non-human origins) in pairs with the selected cells. This can cause inefficiency in targeting the defined surface MP due to the background expression of the same protein in mock cells, or to the heterogeneity of cell surface proteins between mock cells and selection cells. Icell-SELEX solves this problem by constructing pairs of isogenic cell lines to be used in the experiment of selection and counterselection. Cells that are used for positive selection are transfected with a suitable vector to overexpress the target cell surface protein and those used for counter selection are manipulated by gene silencing to deplete the same MP target. To use this method, Takahashi et al. have targeted the ubiquitously expressed transmembrane receptor integrin alpha V (ITGAV) (Takahashi et al., 2016). Stable HEK293 cells were established to overexpress and downregulate ITGAV by an expression plasmid encoding human ITGAV cDNAs and microRNA-mediated silencing, respectively. A hundred-fold difference in ITGAV expression between the two isogenic cell line has allowed isolation of anti-ITGAV aptamers easily and efficiently. Icell-SELEX is therefore a robust method to target a broad range of cell surface targets including pharmaceutically challenging proteins such as channels, transporters and GPCRs. One limitation of Icell-SELEX, however, is the need to determine the dispensability of the target protein to the cell. Targets unimportant for cell growth or cell physiology can be heavily depleted by u6 promoter-driven shRNA or even eliminated all together by CRISPR/Cas (or TALEN etc.) gene knockout systems. Targets important for the cell growth and physiology (such as ITGAV) on the other hand must be downregulated by designing the depletion system carefully.

Possibilities for novel SELEX methods can be viewed as virtually infinite and thus it is a cumbersome task to cover all SELEX variants undertook to date. At any rate, most important is that selection technologies have improved dramatically

over the past 30 years and these efforts have progressively aimed to push towards higher efficiencies, time- and energy conservation, cost effectiveness, and expansion of the current coverage of aptamer targets with aptamers that have even stronger affinities. Delivering novelty in modifications of the SELEX methodology is brought forth in various aspects including improvements in designs of selection libraries and engineering of polymerases, devising new and better target preparation and aptamer screening strategies, and incorporation of modern and promising and technologies such as chromatography, cell sorting systems, microfluidic systems, nanotech, emulsion and droplet digital PCR, high-throughput sequencing, and bioinformatics (Aquino-Jarquin & Toscano-Garibay, 2011; Takahashi, 2018; T. Wang, Chen, et al., 2019). Trustfully, all undergoing R&D in SELEX technologies since presentation of the method in 1990 will project the field of aptamers into a future where the attractive features of these molecules and their production process are seen translated into rich databases and diverse applications that will deeply impact many industries (Darmostuk et al., 2015). cell-based selections particularly are still in their early stages and much research and work await before we achieve extensive applicative success (M. Chen et al., 2016). For this, modern technologies and newly devised methods must find efficient applicability to and integration with cell-based selections (Catuogno & Esposito, 2017). SELEX is a conservative yet comprehensive method, the aptamer discovery potential of which spans a very wide variety of targets (T. Wang, Chen, et al., 2019). Consequently, it follows that deep understanding of the aptamer selection case under investigation is necessary. Personalized SELEX protocols in which the approaches used in each step of SELEX and post-SELEX are carefully designed will help researchers overcome many of the obstacles currently faced by aptamer development. To end, **appendix A** is supplementary to the subjects addressed throughout **section 2.2.2**. It chronologically compiles in table form some more interesting SELEX method modifications that make SELEX more convenient and powerful, and defeat many of its previously known pitfalls. In connection to the overall theme of this literature review and the topic of my thesis, it is important to

bring to attention that many of the SELEX methods listed in **appendix A** are surely adoptable—if not already adopted—into selection endeavors targeting MPs.

2.2.3 Production of Recombinant Membrane Proteins

As seen earlier, successful isolation of an MP of interest is the essential first step and a prerequisite to applying classical protein-SELEX against that MP. *Escherichia coli* has been the workhorse of recombinant protein manufacturing and the most favored expression host through the decades-spanning history of recombinant technology (Bill, 2014). Today, *E. coli* contributes to the multi-billion-dollar global business of production of recombinant FDA-approved biopharmaceuticals (e.g. antibodies, vaccines and many other recombinant proteins) by hosting an expression of 30% of pharmaceutical protein products. For anticancer drugs alone, 69% are produced in this bacterium (Sanchez-Garcia et al., 2016). Beyond biopharmaceutical products, statistics published in 2014 on data available in the PDB and 'Membrane Proteins of Known 3D Structure' database between 2004 and 2013 showed that 78% of recombinant proteins used in structural studies were expressed in *E. coli*; favoring *E. coli* as an overexpression system is no different when it comes to the expression MPs where the percentage still hits 61% and is accompanied by an appreciable increase in MP expression in eukaryotic yeast and insect systems (S. Xiao et al., 2014). In addition to being economical, *E. coli*, as a host for protein overproduction has many merits that make it an experimentally potent and versatile one (Bernaudat et al., 2011). Nonetheless, any chosen expression system bears its pros and cons. To broadly illustrate, unlike overexpression studies on prokaryotic MPs in bacterial systems, earlier studies on many mammalian MPs using the same systems showed functional expression of a much smaller collection of MPs and with many other MPs requiring insect or mammalian cell systems to achieve functional and high-level expression (Reinhard Grisshammer, 2006; C. G. Tate, 2001). On the other hand, mammalian MPs that have succeeded to overexpress in bacteria (e.g. GPCRs) show, at their best, expression levels that are still lower than

many of the prokaryotic MPs and thus require the growth of tens of liters of cells cultures to obtain 1-2 mg of purified MP for studies such as structural determinations (C. G. Tate, 2001). Although *E. coli* has become a well-established and a very mature platform—aged over 45 years (Cohen, 2013)—for the expression of foreign genes, it is still far from being an ideal expression system (Sørensen, 2010). Challenges dwell and continue to frustrate especially with difficult-to-express proteins (large proteins and MPs) (Rosano & Ceccarelli, 2014) and specifically eukaryotic MPs by virtue of their intricate synthesis that is also a rich source of novel recombinant production bottlenecks (Reinhard Grisshammer, 2006; He, Wang, & Yan, 2014).

While solving prokaryotic recombinant MP structures was prosperous, elucidation of recombinant MP structures from eukaryotic sources had boomed only about 14 years ago. The first of the eukaryotic MPs (rat voltage-dependent potassium ion channel, Kv1.2, and rabbit calcium-ATPase, SERCA1a, both produced recombinantly in yeast) had their solved structures published in 2005 while the first of the prokaryotic ones (MscL and KcsA, both produced recombinantly in *E. coli*) had their solved structures published in 1998 (Dilworth et al., 2018). Indeed, the minority of MPs available on PDB are native to eukaryotic species and are outnumbered by their prokaryotic peers mostly represented by bacterial-specific MPs (Dilworth et al., 2018; He et al., 2014). Few MPs can be found abundantly in their native membranes for their extraction and biophysical and biochemical characterization. Recombinant proteins usually need to be obtained in high concentrations especially for structural determination and thus, it is of no surprise that the earliest MPs that had their crystallographic structures solved were endogenously abundant and stable. These are represented by mammalian and bacterial rhodopsins, aquaporins, ATPases, photosynthetic and respiratory complexes, reaction centers and light-harvesting proteins (Bill et al., 2011). Currently, it is known that the majority of medically and pharmaceutically relevant MPs are found in native tissues in very low concentrations and hence, recombinant overproduction systems come into play (Mus-Veteau et al., 2014). As per statistics published in 2018 and obtained from Stephen White's database, at least a third (31%)

of MPs with coordinate files deposited in PDB are recombinantly produced (Dilworth et al., 2018).

2.2.3.1 Membrane Proteins in Light of the New Approach to Production

Generally, recombinant MPs overexpress functionally at low levels due to a variety of reasons that commonly include their imposition of deleterious effects on their hosts, and/or the formation of inclusion bodies (IBs) (Claassens et al., 2017; Pandey et al., 2016). Many bacterial studies have shown that the expression of MPs tends to be far more cytotoxic in comparison to the same level of expression of soluble cytosolic proteins (H. M. Jensen, Eng, Chubukov, Herbert, & Mukhopadhyay, 2017). MPs purification and analysis can also be challenging due to their highly hydrophobic nature and propensity to aggregate (Plucinsky, Root, & Glover, 2018). Even amongst themselves, recombinant MPs seem to exhibit an "express-ability" signature that ranges from the easy- to difficult-to-express (C. G. Tate, 2001). The final expression signature of MPs is contributed to by a whopping array of host-dependent factors that determine the competence of all systems in the host to handle overexpression efficiently (Reinhard Grisshammer, 2006; Gul et al., 2014; H. M. Jensen et al., 2017; Snijder & Hakulinen, 2016; C. G. Tate, 2001). Any fault in each of these factors can potentially obstruct successful overexpression. These include the abundance of amino acids available for use in protein synthesis, any mismatches in codon usage and tRNA abundance, general rates of stability and degradation (mRNA/protein turnover) in a given host, upregulation and/or downregulation of certain genes to which low expression of the target recombinant MP has been correlated, and the competence of systems related to membrane insertion and correct protein folding [here, efficiency is dictated by factors such as: the dependence on specific molecular chaperones/foldases and their abundances; the availability of necessary PTMs and endoplasmic reticulum (ER) and post-ER protein quality control systems; synchrony between the rates of translation, membrane insertion and protein folding; structural and functional variations in the insertion

apparatus in different host types; direct and non-direct bilayer-protein interactions governed by physico-chemical properties such as membrane lateral pressure, curvature, tension, fluidity and also by membrane lipid compositions; capacity of the insertion apparatus before its saturation; host tolerance for overexpression of a recombinant MP before activation of cellular stress responses such as the UPR (unfolded protein response) in ERs of eukaryotes as well as other proteolytic systems in prokaryotes; etc.]. Levels of expression of MPs can vary largely even among closely related cell types and not only between pro- and eukaryotic cell systems (Bill et al., 2011; Hassaine et al., 2006). At the level of individual MPs, some MPs insert and fold significantly better than others (Reinhard Grisshammer, 2006) even when they are homologous to each other (Wagner, Bader, Drew, & de Gier, 2006), thereby highlighting the effect of the intrinsic characteristics of an MP of interest (its identity) on its expression levels (C. G. Tate, 2001). The length of a hydrophobic segment and presence of charged residues in the sequence of an MP of interest, for example, can affect the decision of the translocon complex on whether to translocate the protein into the ER lumen or into the membrane. Interestingly, earlier studies have suggested that intrinsic properties of an MP such as the number of transmembrane domains, protein size, hydrophobicity and codon usage do not affect MP expression levels (C. G. Tate, 2001; Wagner et al., 2006). However, conflicting studies have suggested the importance of at least a portion of these factors to the expression levels. These conflicting results may be resolved by examining different experimental methods used to monitor the localization, quantity and quality of the MPs overexpressed. With all the aforementioned factors, it is very plausible to think that incomplete understanding of the mechanistic events giving rise to the variations in heterologous expression levels of MPs in different and even within the same host system(s) significantly impedes prediction efforts to come up with the best strategy for MPs overexpression. Such lack of understanding has left us mostly with the traditional, time-consuming, and trial-and-error- and optimization-based approach to achieve enhanced recombinant MP production (Bill et al., 2011; de Marco, 2009; Reinhard Grisshammer, 2006). While screening different hosts, expression vectors

(for optimal promoter, selection marker, gene construct, fusion tag, copy number etc.) and culture conditions (for optimal temperature, pH, aeration, growth medium, inducer strength etc.), and employment of more targeted strategies such as using modified hosts that, for example, co-express chaperones, maintain respiratory metabolism, or have favorable abilities acquired by gene deletions (through genetic selections of mutants) in proteases, secretion pathways, or other genome-wide disruptions, are all critical aspects constituting the basis of our approach to optimizing recombinant production (Bill et al., 2011; H. M. Jensen et al., 2017); understanding the biology of MPs overexpression as well as the mechanisms of synthesis will undoubtedly promote tackling overproduction in a more efficient and systematic manner (Bill & von der Haar, 2015). Principally, it is inadequate and arbitrary-like to vary external parameters in order to find the right conditions and then use reported expression levels as evidence for future optimizations (Bill et al., 2011; de Marco, 2009). While high throughput micro-expression trials integrating robotic systems allow a single person to test more than a thousand culture condition combinations within a single week (Rosano & Ceccarelli, 2014), confronting the enormous expression space with more rational, targeted and hypothesis-driven experimental designs potentially produces a narrowed-down and more concise list of choices for the combination of success to achieving maximized functional overexpression yields and with reduced costs (Reinhard Grisshammer, 2006; Gul et al., 2014; Gustafsson et al., 2012). Bioinformatical analyses and comparisons of experimental data that are based on a deeper understanding of overexpression will be important for more confident extrapolations of annotated overexpression data to predict the best methodologies for future endeavors in recombinant production and for freedom from anecdotal evidence (de Marco, 2009). Ultimately, there is simply no single universal strategy that can be used to produce all recombinant proteins (Bill, 2014). Moreover, like discovery-driven optimization strategies, hypothesis/prediction-driven strategies will still need to be tailored for each individual MP of interest (Gustafsson et al., 2012). However, the latter has proven

to be very promising and much more efficient way to find the best approach to a protein overexpression problem.

2.2.3.2 Common Expression Host Systems

After establishing the protein of interest, the choice of expression system (although in most cases, it is *E. coli*) is the first step that paints the workflow for protein production (Rosano & Ceccarelli, 2014; S. Xiao et al., 2014). It defines all methods and technology that are used from molecular tools to equipment and reagents. This choice is commonly based on (in addition to affordability) an assessment for the expression system's abilities to provide the MP of interest with proper synthesis, targeting, insertion and folding (all governed by appropriate cellular machineries such as the translocon complex, molecular chaperones and PTMs), and healthy intracellular environments (e.g. the lipid membrane with its components and physico-chemical properties) (Pandey et al., 2016; Wagner et al., 2006).

E. coli is the most widely used expression host for recombinant proteins due to **1)** its fast growth kinetics (doubling time is 20 minutes in optimal conditions) which reduces the time to obtain a protein of interest, **2)** its capability to grow to high densities which translates into high protein production, **3)** its inexpensive growth media which can also be manipulated without significant losses in yield, **4)** the extensive knowledge available on its genetics, physiology and metabolism which subsequently enabled many intelligent genetic manipulations, and **5)** the abundance of expression vectors (Pandey et al., 2016). However, as introduced earlier, *E. coli*, does not seem to uphold well its glory with recombinant eukaryotic MPs production as it does with prokaryotic ones (Dilworth et al., 2018). Although some notable exceptions of eukaryotic MPs exist especially from more recent years, these proteins generally require eukaryotic expression systems (He et al., 2014). To illustrate, by March 2014, only 4 out of a grand total of 466 MP structures were identified as the structures for eukaryotic MPs produced in *E. coli*. As an expression host, *E. coli*

lacks the essential lipids, proper chaperones, and the capability to perform eukaryotic PTMs all of which are needed for eukaryotic MPs. Furthermore, prokaryotes generally also have a different codon bias and therefore a weak preference for eukaryotic codons. This causes inefficient expression of proteins native to eukaryotes. Prokaryotes are also known to have a much faster rate of protein synthesis and folding in comparison to eukaryotes; this jeopardizes the correct synthesis and integrity of functional eukaryotic proteins. Production of mammalian MPs in *E. coli* usually requires a lot of time and effort to achieve desirable functional levels (Bill et al., 2011). Nevertheless, overexpression in *E. coli* is undergoing continuous improvements in order to allow such a valuable host to adapt to recombinant expression of eukaryotic MPs (Rosano, Morales, & Ceccarelli, 2019). For example, recent efforts have engineered defined glycosylation pathways in *E. coli*. Anyhow, the lack of glycosylation pathways in *E. coli* does not disqualify it as a host for functional expression of eukaryotic MPs as seen with the human CB2 receptor (Wagner et al., 2006). Ultimately, however, irrespective of the ongoing great and creative advances to optimize *E. coli*, this host may still not be the ideal choice to extensively produce eukaryotic MPs (He et al., 2014).

Lower eukaryotes and particularly yeast, like prokaryotic *E. coli*, can grow rapidly, densely and cheaply; have well-known genetics and can be genetically manipulated with many advanced tools (Bill, 2014; Pandey et al., 2016). Yeast cells, however, prevails in their ability to perform a remarkable collection of eukaryotic PTMs that include proteolytic processing, phosphorylation, acylation, prenylation, and *O*- and *N*-linked glycosylation, all of which can be essential for insertion, proper protein folding, and activity. Over 1500 species of yeast are known and a large variety of strains are available (Routledge et al., 2016). However, yeast strains display among each other some variations in glycosylation patterns (Bill, 2014). They also have a propensity to hyper-glycosylate recombinant proteins. On the other hand, other strains can have lower degrees of hyper-glycosylation as they have been engineered to yield more uniform and humanized glycosylation patterns. Unlike, *E. coli*, yeast systems are able to cope with expression of large proteins that size of

which can go up to 160 KDa (50-60 KDa for *E. coli*). Such coping is possibly due to efficient ribosomal translation (Dilworth et al., 2018). Right after *E. coli*, and with at least 7% of all recombinant MPs produced, *Pichia pastoris* and *Saccharomyces cerevisiae* are the major contributors from the microbial world to recombinant MPs production. These popular yeast species were used almost exclusively to produce eukaryotic MPs such as human GPCRs and ion channels.

The baculovirus-infected insect cell expression system was reported to be the dominant system for the recombinant production of most of structurally characterized eukaryotic MPs (He et al., 2014). Insect cell systems also account for approximately 85% of deposited GPCR structures (Errey & Fiez-Vandal, 2020). A wide range of insect cells is available commercially for the expression of MPs. These include *Spodoptera frugiperda* Sf_9, Sf_{21}, High Five, and *Drosophila* Schneider S2 cells. The insect system is regarded as safe, accurate and convenient to scale up (He et al., 2014). In addition to fewer truncated proteins, it also provides codon usage and PTMs that better resemble those in higher eukaryotes and by that, better expression levels and proper protein folding can be permitted (Bernaudat et al., 2011; Pandey et al., 2016). However, the system is more expensive than bacteria and yeast to employ and conversely, cheaper than mammalian systems. Also, in some cases, heterogeneous PTMs (glycosylation) can be generated and complex type N-glycans can be missing which lead to severely defected protein functionality (Errey & Fiez-Vandal, 2020). Significant efforts thus were directed towards the humanization of insect cells glycosylation machinery (e.g. SweetBac cells). Another issue of these cells involves differences in the environment of their lipid membrane compared to that of mammalian cells.

Finally, mammalian expression systems seem to be in principle, the best of choices to produce eukaryotic MPs in terms of supplying the full package of sophisticated and neat processing mechanisms (Bernaudat et al., 2011; Pandey et al., 2016). Some mammalian MPs, indeed, are especially impossible to obtain with prokaryotes or even lower eukaryotes and require very specific cellular settings on levels that range from correct translation to authentic folding, modification,

membrane insertion and protein activity (Errey & Fiez-Vandal, 2020). Overexpression at high levels, however, may often produce incomplete or heterogeneous glycosylation. Typically, overexpression levels in mammalian systems are reasonable but are lower compared to the systems mentioned earlier (Pandey et al., 2016). Similar to those mammalian MPs overexpressed in and purified from bacteria, many of them expressed in mammalian (and even insect) systems can also require liters of culture volumes to overcome the low expression which, unlike in bacterial hosts, is tremendously time-consuming and costly to achieve (He et al., 2014; C. G. Tate, 2001). Also, scalability in itself can become a problem with mammalian cell cultures (Errey & Fiez-Vandal, 2020). With non-adherent cells such as HEK293, yields may be poor as cells struggle to grow in suspension. On the other hand, obtaining large yields from anchoring cells can be costly and impractical. Recombinant protein production using mammalian cell lines is achievable transiently or stably and various mammalian cell lines are available. Transient expression is widely used and can be relatively quick and easy to utilize. Switching from transient to stable transfection is a feasible option to acquire high-level expression and reproducibility (Chelikani, Reeves, Rajbhandary, & Khorana, 2006). Even though a constant source of recombinant protein is obtained with stable mammalian systems, the system itself can be time-consuming to generate (Errey & Fiez-Vandal, 2020). Remarkably, recombinant MPs may exclusively require certain engineered cell lines to produce successfully. For instance, a rhodopsin's mutant form could not be expressed constitutively likely due to its toxicity but tetracycline-inducible HEK293S-TetR cell line allowed crystallization of the protein (Reeves, Kim, & Khorana, 2002).

Cell-free expression systems, also known as *in vitro* translation systems, have been developed to bypass the typical problems of conventional cell-based systems involving their complexity and sensitivity to the overexpression of proteins (responses to toxicity) and to label proteins for structural NMR and X-ray studies (Pandey et al., 2016; Schlegel, Hjelm, Baumgarten, Vikström, & De Gier, 2014; Wagner et al., 2006). These *in vitro* systems have gained popularity in recent years

for the expression of MPs. This system is an open one allowing *in vitro* protein synthesis by translation machineries that are obtained from cell lysates of different organisms. It also allows manipulation of protein synthesis reactions with ease. Available cell-free systems are based on cellular components sourced from organisms ranging from prokaryotic (e.g. *E. coli*) to eukaryotic ones (e.g. insect, plant, animal and human cells). Currently, *E. coli*-based cell-free expression medium is the most widely and successfully used to express both prokaryotic and eukaryotic MPs. These MPs have been produced either as precipitates that were later solubilized by detergents, or in soluble forms in the presence of certain scaffolds such as detergents lipids and nanolipoprotein particles. Various *E. coli*-based cell-free systems have been developed such as the Protein synthesis Using Recombinant Elements (PURE) system which comprises a minimal set of purified elements used in the translation reaction. This system was used to produce small amounts of membrane-inserted MP but unfortunately, is expensive and difficult to scale-up. However, PURE system was recently optimized to rapidly produce high amounts of functional MPs by the addition of liposomes to the reaction mixture. The Cytomim system is another *E. coli*-based cell-free system in which inverted inner membrane vesicles, also from *E. coli*, are added along. Several *E. coli*-based cell-free systems have shown to produce high expression yields of MPs (in the range of milligrams per milliliter of reaction volume. Furthermore, a significant number of MPs produced from these systems are also shown to be fully functional. Cell-free systems produce MPs in a very slow fashion compared to biogenesis *in vivo*; this allows these MPs to fold more properly. MP-GFP fusions are also still usable in cell-free systems to monitor production. Interestingly, MPs can properly assemble with scaffolds and in the absence of many of the factors that facilitate their membrane integration *in vivo*. Taken together, cell-free expression systems are a serious and attractive alternative platform with great potential for the production of MPs even for structural studies.

Success in overexpression of an MP of interest can never be guaranteed with any of the aforementioned host expression systems and even with mammalian cell lines (Reinhard Grisshammer, 2006). To exemplify, a systematic study involving a

batch of 101 closely related GPCRs that were overexpressed in different mammalian cell lines found that expression levels were varying largely for a particular cell line (Hassaine et al., 2006). Furthermore, from BHK-21 cells alone, 21 of 95 receptors showed negative signals.

It has been prominently said that the best host for overexpression of a certain MP is the native one or the most evolutionarily related host [i.e. homologous overexpression is far better than heterologous overexpression to produce functional MPs (C. G. Tate, 2001)]; this justifies the difficulty in overexpressing mammalian MPs in *E. coli* in order achieve success in functional expression (Bill et al., 2011; Schlegel et al., 2014). The feasibility of this generalization is refuted by a counterargument which points out that in many cases, such an approach to selecting suitable expression hosts is not reliable (Gul et al., 2014; H. M. Jensen et al., 2017; Shaw & Miroux, 2003). A recipe for failure of MP overexpression is generated in an ultimatum possibly due to an "additively annihilative" effect emerging as a result of the nature of the methodologies of recombinant protein production wherein high amounts [micrograms to gram scale (Snijder & Hakulinen, 2016)] are required for the different applications, and the nature of recombinant MPs themselves wherein they generally express at low levels.

2.2.3.3 Resolving Bottlenecks in Overexpression of Membrane Proteins

Maximizing recombinant protein overexpression is essentially a multi-dimensional optimization problem (Gustafsson et al., 2012). Unfortunately, the choice of an appropriate expression host is largely empirical as it is difficult to predict with high accuracy the most suitable host for the goal of high-level and functional overexpression of an MP of interest (Bernaudat et al., 2011). The same can be said when defining candidate plasmids and cell culture conditions (Rosano & Ceccarelli, 2014). Conjointly, arriving at optimal expression for a recombinant MP (mammalian or non-mammalian) via choices from these parameters is frequently accompanied by several bottlenecks that hinder desirable and functional protein

yields (Bernaudat et al., 2011; Gul et al., 2014; Pandey et al., 2016; Snijder & Hakulinen, 2016). As seen earlier, functional expression of recombinant MPs tends to occur at low levels due to a broad variety of reasons. Commonly, these proteins disclose some toxic nature (whether in a biologically functional or non-functional form) upon their host by various possible interactions with its metabolic networks. When overexpressed, MPs are also prone to unfolding/misfolding and can form aggregates resulting in the formation of dense inclusion bodies. Another problem lies in the availability of membrane space; recombinant MPs are often produced in hosts bearing limited membranous capacity. Consequently, these hosts are rendered unable to accommodate a constantly synthesized MP. Hydrophobic mismatches of recombinant MPs with the host's membrane are also another possibility. Improper insertion into the target membranes thus may lead these MPs towards non-functional conformations. Another issue that can be inflicted by MPs is their ability to destabilize cellular membranes. On the brighter side of the subject matter, many different troubleshooting techniques and optimizations are available to be used for the resolution of overexpression issues.

Cultivation Conditions Adjustment and Monitoring

Firstly, although often overlooked, manipulation of culture conditions is the easiest way to remedy poor recombinant expression; one way to manipulate among the culturing conditions is temperature and is in particular not overlooked (Rosano et al., 2019; Snijder & Hakulinen, 2016). In *E. coli*, a reduction from 37°C to 18-20°C often yields higher levels of functional MP. Lowering the culture temperature induces various cellular changes such as reduced growth rate, changes in RNA, reduced translational speed, induction of cold-shock proteins, and alterations in membrane compositions. Remarkably, equilibrium is struck between protein synthesis and correct protein folding at lower temperatures. Albeit at such temperatures overexpression can be lower, membrane-inserted MPs on the other hand are observed to be in larger amounts (Wagner et al., 2006). It is important to mention that optimization techniques are applied in fed-batch cultures and in shake-flask cultures do not necessarily align perfectly (Rosano et al., 2019). In large

bioreactors, for example, temperature control elevates cost exponentially. Whereas, parameters such as aeration and glucose feeding rate are of an utmost importance for growth control and are monitored carefully in bioreactors for optimized yields. These factors however are rarely considered at lab-scale cultures and one major reason for that is the need for specialized monitoring hardware. Oxygen limitation, especially with large cell densities, can trigger the expression of over 200 genes in attempt to allow the cell to adjust its metabolic activities to the available oxygen amount, consequently affecting optimal growth over long culture periods (Rosano & Ceccarelli, 2014). The easiest way to increase the availability of oxygen is by increasing the shaking speed; 400-450 rpm and 350-400 rpm are optimal speeds for regular flasks and baffled flasks, respectively. Vigorous shaking can generate foaming which decreases oxygen transfer. Although antifoaming agents are suggested, they may affect the growth and recombinant protein production of several microorganisms. In line with the great importance of such factors for obtaining higher culture densities and the best protein production, technological advances in carbon source supply, oxygen availability, and shake flask/microplate-adapted monitoring devices have shown promising results and will lead to wider adoption of such technologies (Rosano et al., 2019).

Moving on to a different adaptable cultivation parameter, nutrient availability also has a strong influence on growth rates and protein expression machineries (Routledge et al., 2016). Growth media such as LB and M9 standard media are the most common for protein production in *E. coli* including MPs (Snijder & Hakulinen, 2016). Although less common, richer media such as 2TY (tryptone-yeast), TB (terrific broth) and NPS have also been used and work better for a higher protein harvest simply by increasing culture biomass (no alteration of expression per cell). Standard media, however, can still be supplemented with higher amounts of peptone/yeast extract and divalent cations ($MgSO_4$) to boost cell growth and cell densities and compensate for the scarce amounts of carbon sources and divalent cations (Rosano & Ceccarelli, 2014). In yeast cultures, specific improvements in yields have been observed by additives such as dimethyl sulphoxide (DMSO),

glycerol, histidine and even specific ligands of the overexpressed protein (Routledge et al., 2016). Complex media are rich in nutrients and can provide higher yields. However, they allow little control over the cells' metabolic state in culture (Rosano et al., 2019). Although not cheap (Dilworth et al., 2018), autoinduction media are also gaining popularity for both soluble and MPs as they have shown their superiority; possibly due to a chance for adaptation of the growing bacterial cells to gradual increments in protein biogenesis (Snijder & Hakulinen, 2016). With autoinduction media, higher bacterial densities can be achieved, induction time point is highly reproducible, and culture manipulation or stoppage for timely addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) is no longer needed (Rosano & Ceccarelli, 2014; Rosano et al., 2019). These media contain at least two carbon sources: typically, glucose and lactose (glycerol is also added to obtain higher yields). Consumption and depletion of the favored carbon source (glucose) during exponential growth is then followed by initiation of consumption of the secondary carbon source (lactose). The latter process—facilitated by activation of lactose permease—initially inactivated by the presence of glucose (that is, catabolite repression), initiates protein overexpression in *lac* promoter-based systems that operate by the inducer lactose (Rosano & Ceccarelli, 2014). A different autoinduction platform known as SILEX (*S*elf-*I*nducib*L*e *EX*pression) was discovered serendipitously in 2016 to work by the removal of the *lac* promoter repressor LacI in *E. coli* by the production of human Hsp70 that interferes with the enzymatic function of endogenous glyceraldehyde 3-phosphate dehydrogenase (Briand et al., 2016). Eventual LacI removal by this system facilitates overexpression of the gene of interest.

Plasmid Design: Gene and Protein Engineering

Along with the importance of an optimal expression hosts and culturing conditions, optimized expression plasmids are also crucial to the attainment of high bioproduct yields (Rosano et al., 2019). A well-designed expression vector entails numerous and important optimizations to its different components such as the promoter, translation initiation signal, selection maker and replication origin.

Construct design also needs considerable attention if success is to be ensured, a process that has its fair share of trial-and-error and cycles of design and evaluation (Snijder & Hakulinen, 2016). Nevertheless, there has existed from researchers a reluctance to alter protein sequences in fear of destabilization of the protein or disruption of its function (Rosano et al., 2019). However, protein variants have had improved recombinant overproduction as well as retainment of structure and function. MP constructs should be designed for transcription into mRNAs, translation into corresponding polypeptide chains, and translocation into the membrane and forming membrane-inserted functional MPs (Snijder & Hakulinen, 2016). The recent years have witnessed a substantial replacement of traditional molecular biology work with many commercial DNA synthesis services and proprietary algorithms. These are used in construct optimizations for codon usage, GC-content, sequence motifs and RNA secondary structure. Generally, biosystems engineering has been driven mainly by trial-and-error methods and usually with relatively little sampling of potential variable that could affect protein production (Gustafsson et al., 2012). Hence, the largely unexplored realm of modern engineering approaches applied to biological systems and particularly the field of genes engineering, was reviewed elegantly and extensively by Gustafsson et al. The synthetic biology and systematic bioengineering approaches described by the authors are also applicable in areas such as plasmid engineering, genome engineering, and pathway engineering.

Bias in codon usage is one of the earliest recognized issues affecting the expression of proteins and has become evident with the increased knowledge on genome sequences. It arises as a result of the significant difference in the frequency of synonymous codons between the foreign coding DNA sequence and the expression host system which leads to the depletion of low-abundance tRNAs (Rosano & Ceccarelli, 2014). Consequently, this may cause translational errors such as amino acid misincorporations and/or polypeptide truncations by ribosomal stalling and subsequent detachment, all of which can affect the levels and/or activity of the overexpressed heterologous protein. Plenty of free online tools are available

to check for codon bias and the presence of rare codons in a gene of interest for a given host such as *E. coli*. To fix codon bias, two approaches are commonly undertaken. The first is codon optimization where the rare codons in the foreign DNA sequence are modified to mirror codon usage of the host without alterations in the identities of the amino acids. This can be carried out by site-directed silent mutagenesis or by re-synthesizing the whole or parts of the gene. Each of these methods is not without its drawbacks. Silent mutagenesis is a cumbersome and expensive process; especially when many recombinant proteins are involved. On the other hand, synthesizing gene designs has its own difficulties where the best sequence is chosen out of a vast number of possible codon combinations. A strategy known as the "one amino acid-one codon" presents the simplest way because it encodes all instances of a given amino acid in the gene of interest by the most abundant codon in the host. More advanced algorithms exist and use different parameters for codon optimizations such as codon context and codon harmonization. Such tools are found on free webservers or in standalone software products. It is important to address the fact that switching codons to optimize their usage by the host can be tricky since it may negatively impact expression levels by affecting a variety of frameworks and processes. Such influences for example can lead to the formation of RNA secondary structures, alterations in the open reading frame (ORF) decoding speed, and the creation of certain undesired motifs or structures such as repeats, RNase sites, and Shine-Dalgarno-like sequences (Claassens et al., 2017; Rosano & Ceccarelli, 2014). The latter for example can be a cause for translational pausing due to the occurrence of certain codon combinations. However, it may also be advantageous as it can facilitate proper protein folding. Silent substitutions in the ORF and the interdependencies between the coding sequence and the expressed protein have been the least understood areas (Gustafsson et al., 2012). MP biogenesis in particular has been shown to be a codon-sensitive process (Snijder & Hakulinen, 2016). Single synonymous codon substitutions can affect different schemes in the cell producing an MP (e.g. mRNA stability and structure, translational initiation and elongation, translocation, and protein folding). Optimization of MP constructs for

codon usage has indeed produced contradicting data with either positive or negative effects on yields of functionally overexpressed heterologous MPs. This reflects target-specific effects and/or variations in the optimization algorithms. When optimizing the genetic code against codon bias to obtain higher MP overexpression, overlooking effects of the coding sequence identity on protein biogenesis and its machineries can result in failure of MP overexpression (Gustafsson et al., 2012; Nørholm et al., 2012). Unrelated to plasmid design is the second approach to resolve codon usage bias. Briefly, the method involves modifying host systems to increase their availability of underrepresented tRNAs (Claassens et al., 2017; Rosano & Ceccarelli, 2014). *E. coli* strains such as the Rosetta (DE3) strains were developed to overexpress tRNA species for codons normally rare in *E. coli* thus allowing expression of eukaryotic proteins. Such strains harbor plasmids that encode for genes of the extra copies of desired tRNA. Although the use of such strains can increase protein production levels, their use may sometimes lead to protein aggregation due to, for instance, overriding translational pausing.

In addition to the possible corrections performed on coding sequences to bypass codon bias, the protein sequence may also be modified by truncations and certain mutations (by systematic, evolutionary or even computational methods) not only to elevate the expression levels of functional recombinant protein but also to enhance its stability and homogeneity when purified and to make it suitable for any intended downstream studies and applications while keeping biological/pharmacological functionalities of the protein in check (Snijder & Hakulinen, 2016). Structural studies demand the most with respect to quantity, purity, stability and homogeneity of the protein of interest. Moreover, mobile regions may need to be removed for crystallization, domains may need to be separated in order to keep the total size manageable for NMR studies, and stable folding soluble domains (e.g. T4L) can be strategically introduced into the exposed loops of MPs to enhance protein stability and crystal contact formation. The latter method has been used with GPCRs as depicted here (Chun et al., 2012). N- and C- termini modifications have been known for a long time to enhance protein production yields

(Schlegel et al., 2014; Wagner et al., 2006). Many of the prokaryotic and eukaryotic MPs have N-terminal tails that need to be translocated across membranes; these usually show strong preference for the cytoplasmic space. The translocation of N-terminal tails is dependent on its ability to maintain a translocation-competent structure, the number of positively charged residues in the tail region and the 'strength' of the first transmembrane segment (i.e. the charge difference, the length, and the overall hydrophobicity of the reverse-signal anchor). Functional overexpression of the yeast mitochondrial carrier AAC2 (ADP/ATP exchanger) in *L. lactis*, for example, can be increased by N-terminus shortening or by swapping the tail with a shorter one from the isoform AAC3 (Monné, Chan, Slotboom, & Kunji, 2005). Several GPCRs such as the human cannabinoid, opioid, dopamine and serotonin receptors have also demonstrated an increase in expression in eukaryotic systems by engineering a signal peptide sequence to their N-termini (Wagner et al., 2006). Fusing signal peptides allows targeting and translocation of the long N-terminal of the receptor. In yeast expression systems, for example, the membrane can be targeted using native yeast signal sequences (Pandey et al., 2016). Random and direct mutagenesis of genes is also used to isolate MP variants with better production characteristics (Schlegel et al., 2014). This was shown in an *in vitro* evolutionary screen of random mutant libraries of nine MP genes and using a detergent-adapted colony filtration blot method (Molina et al., 2008). One cycle of directed evolution was enough to obtain a significant improvement in yields for five of the nine MPs. In one of the five mutant MPs (the human microsomal glutathione S-transferase 2), it was confirmed that mutagenesis had not affected the enzymatic activity of the protein.

Fusion tags such as GFP, MBP, GST, NusA, SUMO and Mistic are some of the most powerful molecular tools to resolve various problems with the production of recombinant proteins. Besides enhancing protein solubility and facilitating affinity for chromatographic purification methods and antibody detection, these tags can improve the yield of difficult-to-express proteins (including MPs) by increasing their expression (Bernaudat et al., 2011; Butt, Edavettal, Hall, & Mattern, 2005;

Pandey et al., 2016). Reports using tags with MPs include the rat neurotensin receptor and the human adenosine A_{2a} receptor, both which expressed in *E. coli* and fused from their N-termini to MBP (R. Grisshammer, Duckworth, & Henderson, 1993; Weiß & Grisshammer, 2002). A more recent study used two short hydrophilic bacterial proteins (YaiN and YbeL) as N- and/or C-terminal fusions with 14 pro- and eukaryotic MPs for expression in *E. coli* (Leviatan, Sawada, Moriyama, & Nelson, 2010). Moderate to high expression was achieved for most MPs and 5 reconstituted MPs also elicited transport activity comparable to the native proteins. In a counter example, fusion of the N-terminus of the serotonin transporter (SERT) to the C-terminus of GST has yielded reduced functional expression of SERT in insect cells (C. G. Tate, 2001). Using tagged chimeric proteins requires careful design and planning especially if multiple tags are used together at once (Pandey et al., 2016). Considerations must be given to their metabolic effect on the host, their interaction with the target protein, and other downstream optimizations such as finding the optimal purification protocol and tag (and linker region) removal strategy.

The size and location of a given tag may have an influence on expression and purification properties (Mohanty & Wiener, 2004; Snijder & Hakulinen, 2016). For example, compared to a hexa-His tag, a deca-His tag is more suitable for lowly expressed MPs because it improves enhanced capturing and allows more stringent washing. However, for a tetrameric *E. coli* aquaporin (AqpZ), a deca-His tag resulted in lower yield compared to a hexa-His tag. Yields may also be affected by locational preferences for the attached tag (i.e. biases for either N- or C-terminal tags). For example, C-terminal tagging is useful when the N-terminal end contains a signal peptide attachment for protein secretion (Rosano & Ceccarelli, 2014). At any rate, large extra-cytoplasmic C-terminal tails translocation is unlikely to be as problematic as that of N-terminal tails and the attachment of almost any fusion partner seems reasonable (Wagner et al., 2006).

Tags, and most commonly, green fluorescent protein (GFP), are easily utilized to monitor and quantitatively assess MP folding (Bill et al., 2011). Correct folding of the target MP is reported robustly by it facilitating correct folding in the

fusion partner (forming a GFP barrel) followed by fluorescence as well as Sodium dodecyl sulfate (SDS)-resistance of the GFP. This technique is very useful in selecting hosts that display the best MPs functional expression. Host Membranes can be targeted for MPs localization also by using fusion tags (Bernaudat et al., 2011; Schlegel et al., 2014). For example, the 13 KDa Mistic was successfully used in *E. coli* to improve the overexpression of different MPs as an N-terminal tag and to directly target membrane insertion as it spontaneously associates with the inner membrane of *E. coli* without any need for recognition by the Sec-translocon machinery. Fusion with Mistic also showed boosted MP production in the single-membraned *L. Lactis* (Song, In, Lim, & Rahim, 2017). The outer membrane of *E. coli* which is commonly used for surface display of soluble proteins, was targeted for the localization of MPs (Pandey et al., 2016).

Protein Toxicity Control

Maintenance and processing of expression plasmids place a considerable amount of metabolic stress onto the host cell (Routledge et al., 2016). Inevitably, whichever choice is made about the expression vector and optimal culture conditions, it is important to test for stability of the plasmid/host strain combination in the medium over time before scaling-up protein production. (Baneyx, 1999; Dilworth et al., 2018; Palomares et al., 2004). However, establishing the right host, plasmid and cultivation parameters still do not guarantee a satisfactory production process and the protein may remain undetected by sensitive techniques such as western blotting (Rosano & Ceccarelli, 2014). For this matter, the protein of interest may well be triggering cytotoxicity and interfering with normal homeostatic and proliferative processes of the cell. This is observed as a slower growth rate, lower final densities and increased cell death. The subject of expression of highly toxic genes in *E. coli* has been reviewed in more details (Saida, Uzan, Odaert, & Bontems, 2006). If the recombinant MP is toxic before induction, one option is to control basal (also called leaky) expression to attain normal growth until the time of induction (Rosano & Ceccarelli, 2014). For *lac*-based promoters, expression of lacI from $lacI^{Q}$ (has a mutant *lacI* gene promoter) allows tighter transcriptional control and

repression of the gene of interest by a 10-fold higher lacI expression compared to expression from *lacI* gene. With *lac*-based promoters, addition of 0.1-1% w/v glucose or using glucose as an energy source in the defined medium can also provide a means to achieve tighter control. The T5 promoter recognized by *E. coli* RNA polymerase uses two *lac* operators for increased control on the promoter.

The T7 promoter is designed in the pET series (pMB1 ori; medium copy number; Novagen) and is recognized by the phage T7 RNA polymerase (T7RNAP). It is one of the strongest and most widely used promoters for recombinant protein expression. The expression system is used most with the famous *E. coli* strain, BL21(DE3) and its derivatives (Dilworth et al., 2018). In a T7-based system, the highly active T7RNAP introduced through another plasmid or into the bacterial genome (more common) in the λDE3 prophage, is under control of the *lac*UV5 promoter, a mutant of the *lac* promoter that partially desensitizes the cell to catabolite repression by allowing some expression in the presence of glucose (Rosano & Ceccarelli, 2014). The T7-based expression system is hence inducible by *lac* promoter inducers and repressible by $lacI^{Q}$. High stringency of inhibition of the T7 promoter (and consequently, inhibition of transcription initiation and control of leaky expression) is also achievable by T7 lysozyme co-expression, an inhibitor of T7RNAP by direct protein-protein interaction provided in the compatible companion plasmids, pLysS or pLysE. Upon induction, the amount of inhibitor is titrated by the high generation of T7RNAP. “Free” T7RNAPs are thus able to engage in transcription of the recombinant gene of interest. Yet a third level of control that can be added is constructing a hybrid T7/*lac* promoter that contains an inserted *lac*O operator present downstream of the T7 promoter. A different and attractive approach to eliminating toxic basal expressions of proteins of interest is by using promoters that rely strongly on positive control for expression induction. By this approach, lower background expressions are expected. The $araP_{BAD}$ promoter available in pBAD vectors is an example; the dual repressor/activator function of AraC protein is determined by the inducer arabinose. In the absence of the inducer, AraC binds to DNA sites to form protein-DNA complexes that form loops preventing RNAP

promoter binding. In the presence of the inducer, AraC switches to become an activator of transcription. To demonstrate the use of transcriptional control by reducing basal expression and reduce the toxicity and improve the quality of human GPCRs overexpression in *E. coli*, a research group has inserted the *lac* operator at the +1 position of an existing *phoA* promoter and created a hybrid *phac* promoter (H. S. Kim et al., 2012). This system requires both phosphate starvation and *lac* induction by the inducer IPTG and suppression of possible upstream cryptic promoters was done by introducing the λ t_o transcriptional terminator upstream of the *phac* promoter, resulting in the *tphac* promoter.

In addition to promoters, origin of replication controlling the plasmid copy number is also one of the important elements found in a vector (Gomes, Monteiro, & Mergulhão, 2020). High-copy number plasmids were shown to impose high metabolic loads that may impair cell growth and cause plasmid instability. Thus, use of low-copy number plasmids has been suggested to control protein toxicity from leaky expression (Rosano & Ceccarelli, 2014).

Using autoinduction media can be another useful way to resolve toxic leaky expressions that hinders cell growth prior to IPTG addition (Dilworth et al., 2018). Two improvable parameters that are also important and worth looking at but rarely discussed for higher protein yields are proper preparation of the starting preculture and setting the induction time (Rosano & Ceccarelli, 2014).

In lab setting, if not detrimental to cell growth and plasmid stability, leaky expression alone and without using any inducer can be a successful strategy and sufficient to attain functional MP levels (Dilworth et al., 2018). This has been demonstrated with at least two MPs. If induction is to be carried out, optimizing inducer concentration for high-level MP expression and insertion is an critical factor to be considered irrespective of the induction method used or the promoter present in the system (Snijder & Hakulinen, 2016).

IPTG is one of the widely used synthetic inducers for genes under control of the *lac* operator; it has the ability to induce strong overexpression to produce a

recombinant protein that can comprise up to 50% of the total cellular protein content (Gomes et al., 2020). Although advantageous to use, IPTG can be very toxic to the cell and can affect its growth by imposing metabolic burdens as it redirects the cell's resources and machineries to fuel energetically-demanding processes; that is, rapid generation of high amounts of a non-essential foreign protein and to replicate and maintain the plasmid vectors (Dvorak et al., 2015).

Burdens on the cell do not stop there; they are also associated with importing IPTG into the cell by *lacY* where it can promote its own cytotoxic effects. Cellular stress can also be exacerbated due to toxic cross-talks of the overexpressed foreign protein with the host's metabolic networks by toxicity from itself, or its substrates, intermediates or products. "The Walker *E. coli* strains", C41(DE3) and C43(DE3), and Lemo21(DE3) were developed to tolerate toxic MPs by reducing the high transcription rate of T7RNAP and to help attenuating insertion machinery overload and chaperones overuse (Claassens et al., 2017; Snijder & Hakulinen, 2016). These strains were developed in 1996 through a screen designed to isolate derivatives of BL21(DE3) with enhanced MPs overproduction ability (Rosano & Ceccarelli, 2014). Until 2008, it was unknown why MPs overexpression in these strains is hardly toxic and often results in high yields (Wagner et al., 2008). It was then discovered that two mutations in the -10 region of the *lac*UV5 promoter prevented cell death during MPs overexpression by reverting the promoter back to its weaker wild-type form, causing dampened T7RNAP activity. The Lemo21(DE3) strain also developed by Wagner et al. was engineered based on the physiological, proteomic and genetic studies done on the Walker strains and is characterized by the principle of tunability which is described later. Toxicity of a foreign protein in many cases is dependent on reaching and exceeding a certain threshold of tolerance of the host (Rosano & Ceccarelli, 2014). If a protein of interest is toxic after IPTG induction, mitigating the negative effects on cell growth becomes critical and can be achieved by optimizing IPTG concentration (Dvorak et al., 2015; Snijder & Hakulinen, 2016).

Switching to IPTG's natural analog, lactose, was also shown to dramatically reduce the stress on the cell. If the non-hydrolysable analog (IPTG) is used, it is

recommended to start with low concentrations as it can be sufficient to attain full induction. In doing so, the number of parameters screened is also simultaneously reduced. Howbeit, optimal concentration of this inducer (and virtually, any other inducer) is ultimately case-specific. However, reproducibility of dose-dependent IPTG induction is low because of the high heterogeneity in active IPTG transport into cells (Rosano & Ceccarelli, 2014). In other words, the level of protein expression is not predictable from IPTG concentration due to significant variance in availability and activity of lac permease among cell. Per this issue, a BL21 derivative, The Tuner™ (DE3) strain from Novagen, was developed to emphasize permease-independent IPTG entry and therefore, due to a mutation in lac permease giving a $LacY^{-}$ population, gain controlled, concentration-dependent and homogeneous induction. Similarly, another strain was constructed with a derivative ($lacO^{c}$) of the wild-type *lac* operator for the purpose of uniform entry of lactose. IPTG is highly potent even at low concentrations; for this reason recombinant protein levels can be difficult to fine-tune for optimized overexpression (Rosano et al., 2019). In such cases, Lemo21(DE3) strain emerges as a useful option (Rosano & Ceccarelli, 2014). Tunability in this strain stems from the $rhaP_{BAD}$ promoter that controls the *lysY* gene encoding for the production of T7 lysozyme. As concentration of the inducing sugar L-rhamnose reaches towards the higher end during a dose-screening trial, T7 lysozyme increases and T7RNAP is suppressed thereby declining production of the protein of interest. Other tunable (non-constitutive) promoters such as *araBAD* may be rewarding to examine to manipulate RNA levels by adjusting inducer concentrations (Rosano et al., 2019). Promoters are not only induced by chemical cues to initiate transcription but are also inducible by physical signals such as temperature (e.g. upshifts and downshifts in temperature induce protein expression in the pL and cspA promoters, respectively) and light (e.g. blue-light inducible T7RNAPs known as “opto-T7RNAPs”), both of which can be very attractive for industrial-scale fermentation (Rosano & Ceccarelli, 2014; Rosano et al., 2019). As far as plasmid vectors are concerned, it is a common misconception that a higher gene dosage equals to more protein production; this is not always the case (Gomes

et al., 2020) as demonstrated by the improved MP production in strain EXP-Rv1337-4 that has a reduced plasmid copy number (Snijder & Hakulinen, 2016). Hence, again, in addition to slowing down expression rate to a level that does not overload the translocation machinery, decreasing the plasmid copy number can also be powerful to improve yields of proteins that are toxic after induction (Rosano & Ceccarelli, 2014). Reductions in temperatures before IPTG addition have also shown to not only prevent the formation of IBs and improve MPs solubility but also to effectively reduce toxicity of recombinant MPs (Dilworth et al., 2018). Another solution (which can sometimes be the only solution) to handle heterologous toxic proteins in the cell is to drive them out of the cellular environment by targeted secretion to the periplasm or the medium (Rosano & Ceccarelli, 2014).

Valuable Alternative Bacterial Hosts

Although *E.coli* is by far the most popular prokaryotic host for the production of recombinant proteins including MPs, other bacteria can have superior properties for the overproduction of some of these proteins (Schlegel et al., 2014). For instance, such hosts may have lower translational rates, different (extended) chaperone repertoires as well as differences in the lipid composition of their membranes.

Lactococcus lactis is a non-pathogenic, non-invasive Gram-positive food-grade lactic acid bacterium and a widely used microbial cell factory system for industrial protein production (Cano-Garrido et al., 2014, 2016). A major advantage of this host is not favoring the formation of IBs although more recent studies have broken this generic rule. *L. lactis* has proven itself an outstanding expression system for both prokaryotic and eukaryotic MPs including plant as well as rat and human MPs which were in some cases expressed with higher functional yields than in typical *E. coli* and yeast systems (Schlegel et al., 2014; Song et al., 2017). Having a single layer of cell membrane, *L. lactis* also is a great candidate for functional studies of ligands and inhibitors (King, Boes, & Kunji, 2015). Because of the industrial value of this organism, its physiology has been studied in great detail (Schlegel et al., 2014). It is genetically accessible and a variety of plasmids with both high and low

copy numbers and inducible promoters have been developed. One of its numerous merits is having a lower doubling rate (~1 doubling/hour) which can be advantageous with the production of "difficult folders". *L. lactis* also shows lack of excessive proteolytic activity, and has a capacity for co-expression from different plasmids, and great strain engineering potential to enhance the production of MPs.

Bacillus subtilis is a Gram-positive bacterium that is widely used in the industry for the production of secretory proteins. *B. subtilis* physiology has also been extensively investigated. The organism is genetically accessible, and various expression systems are available. Although MPs biogenesis in *B. subtilis* is a largely unstudied area of research, this organism is expected to have great prospects for the production of these proteins as per the available studies and evidence (Zweers et al., 2008). *B. subtilis* has been investigated for the overproduction of essential model MPs from itself as well as *Staphylococcus aureus* (Zweers, Wiegert, & Van Dijl, 2009). The study shows that membrane-associated stress-responsive systems (σ^W regulon and the CssRS two-component regulatory system) are a major bottleneck for the overexpression of MPs in *B. subtilis* and that the removal of at least one of these dispensable systems results in a dramatic improvement in MP yields.

Lipid Membrane-Membrane Protein Incompatibilities

Incompatibility of generated MPs with the properties of a membrane lipid bilayer can be a significant bottleneck when overexpressing those MPs in a heterologous fashion; i.e. in an expression host other than the native one (Wagner et al., 2006). Such differences can significantly hamper proper MP insertion, folding and functioning. However, these incompatibilities are not necessarily a dead end for heterologous overexpression. Lipid composition incompatibilities, for instance, can be bypassed by adding certain lipids during and/or after protein purification to restore functionality of the overexpressed MP. In another study, it was shown that coupling of overexpression of the membrane-anchored AAA+ protease FtsH to the production of four different class I GPCRs (CB1, CB2, BR2 and NKR1) in *E. coli* greatly enhances their expression yield as well as the amount of detergent-solubilized and

isolated recombinant CB1 and BR2 (Link, Skretas, Strauch, Chari, & Georgiou, 2008). It has been suggested that the rationale for these FtsH-mediated improvements in GPCRs overexpression are due to changes in the lipid composition of the cytoplasmic membrane. Yeast strains were engineered to have humanized membranes by synthesizing cholesterol rather than the native sterol, ergosterol (Routledge et al., 2016). Humanization of yeast cells does not impair cell viability; however, growth rates and cell densities can somewhat be affected.

The membrane space available for use to accommodate the overexpressed MPs is another source for bottlenecks (Wagner et al., 2006). Co-expression of the *E. coli* b-subunit of ATP-synthase is known to trigger intracellular membrane proliferation. C41 and C43 strains isolated by Miroux and Walker also provide concomitant formation of membranes by supporting improved expression of the ATP-synthase b-subunit (Snijder & Hakulinen, 2016). *Rhodobacter sphaeroides*, a purple non-sulfur photosynthetic bacterium, was also exploited to overcome membrane space shortage (Bernaudat et al., 2011). This bacterium immensely amplifies its production of intracytoplasmic membranes (chromatophores) to accommodate the large amounts of photosystems being synthesized in response to light and/or reduced oxygen tension. Recombinant MPs were produced for structural studies using this approach. Procedures and instrumentation for handling *R. sphaeroides* do not differ significantly from *E. coli* and thus, it can be easily established and put for use in any laboratory. Another naturally optimal system for the high biogenesis of MPs and accommodating membrane can be found in the well-known fruit fly *Drosophila melanogaster* (Eroglu, Cronet, Panneels, Beaufils, & Sinning, 2002). The photoreceptor cells of *D. melanogaster* have extensive stacks of specialized membranes with high levels of naturally present rhodopsins. In transgenic flies, these cells have been used successfully to overexpress a number of GPCRs including the *D. melanogaster*, metabotropic glutamate receptor, which was later purified and functionally reconstituted in liposomes strictly composed of ergosterol.

Inclusion Bodies (IBs)

Protein folding is a complex process that can become deregulated for a variety of reasons such as cellular stress to which the cell is capable of responding in many different ways (Palomares et al., 2004). Introducing a foreign gene into a host such as *E. coli* is usually associated with loss of the spatio-temporal control of expression of that gene (Rosano & Ceccarelli, 2014). The unnaturally high levels of expression allow for a higher probability of interactions between similar hydrophobic stretches in the polypeptide due to their high concentration. This, and variations in the expression host's microenvironment compared to the native host (e.g. pH, osmolarity, redox potential, cofactors, and folding and PTM mechanisms) lead to protein instability and buildup of the protein aggregates known as inclusion bodies. These aggregates have been observed throughout the different expression hosts discussed earlier in **subsection 2.2.3.2** (Palomares et al., 2004), and are known to occur frequently in bacteria and especially with non-*E. coli* targets (Dilworth et al., 2018). Different methods are constantly under suggestion and development to allow proper recombinant protein folding under overexpression conditions and to eliminate the aggregation problem (Rosano & Ceccarelli, 2014). In the yeast *S. cerevisiae*, it was shown that deployment of the UPR by the cell as a result of failure of the ER's folding capacity to cope with the lethal accumulation of unfolded (and therefore, aggregating) proteins can be exploited as means to maximize overexpression of functional MPs by monitoring induction of the UPR through a reporter gene (Reinhard Grisshammer, 2006). In another work, the *E. coli* Sm^P strain known as CH184 was successfully used to attain enhanced folding of a diverse set of eukaryotic proteins by reduction in translational speeds (Siller, DeZwaan, Anderson, Freeman, & Barral, 2010). The bacterial strain harbors ribosomes with mutated S12 proteins; in the absence of streptomycin, a "hyperaccurate" phenotype is displayed in comparison to the wild-type where translation rate slows down to ~5 amino acids *per* second and accuracy of amino acid incorporation increases by ~20-fold. Such a mechanism may be efficient to eliminate the toxicity and translocational loads imposed by eukaryotic MPs (Snijder & Hakulinen, 2016).

Molecular chaperones lie at the heart of protein quality control where they assist newly synthesized proteins reach their final structure (Rosano & Ceccarelli, 2014). Other specialized chaperones (e.g. ClpB) can disassemble IBs of unfolded polypeptides. Folding of MPs in bacterial cytoplasmic and eukaryotic ER membranes is known to be mediated by generic integral membrane chaperones (Wagner et al., 2006). The soluble domains of MPs may also require cytoplasmic and periplasmic or luminal chaperones. All in all, these folding mediators may lack the capacity to support the folding of overexpressed MPs due to their insufficient numbers. In other words, elevated expression levels and crowding of a recombinant protein can saturate and overwhelm quality control mechanisms. Some MPs such as some yeast transporters and a few rhodopsins, on the other hand, require specific chaperones that may be absent altogether in overexpressing hosts. Termination of protein expression by addition of fresh medium supplemented with protein synthesis inhibitors (and by inducer removal), or by tuning concentration of the inducer are intuitive solutions to allow recruitment of chaperones to support the folding of nascent recombinant proteins (Rosano & Ceccarelli, 2014). Co-expression of molecular chaperones is also a very widely used strategy to inhibit the formation IBs given the roles of these proteins in mediation of folding. Adjustment of co-expression conditions, however, is mostly empirical and can be a tedious process (Wagner et al., 2006). The *E. coli* Mg^{2+} transporter CorA and the eukaryotic cocaine-senstive serotonin transporter (SERT) are two early examples of MPs that have had an increase in functional overexpression after their co-expression with the DnaK/DnaJ chaperone system and the ER folding catalyst calnexin, respectively (Y. Chen, Song, Sui, & Wang, 2003; Christopher G Tate, Whiteley, & Betenbaugh, 1999). If plasmid systems for chaperones co-expression are unavailable commercially, an alternative option is to induce the natural chaperones network by adding benzyl alcohol or by heat shock (Rosano & Ceccarelli, 2014); the latter however rarely recommended. Supplementing culture media with osmolytes (also known as chemical chaperones) such as proline, glycine-betaine, and trehalose may be a viable option for *in vivo* folding. Specific cofactors such as different metal ions

and polypeptide cofactors may also be required for the folding pathways leading to the correct final conformation and for the stabilization of the folded protein. Therefore, such cofactors are also supplemented into growth media.

As seen earlier in **subsection 2.2.3.3**, slowing down growth rate of the expression host in culture (most commonly by reducing temperature) slows the rate of protein production and concentration in the cell thereby giving time for proper folding of nascent proteins. Lowering the temperature in particular, disfavors aggregation because the hydrophobic interactions have a dependency on higher temperatures. A report in 2013 demonstrated successful expression with reduced IBs formation at a temperature as low as 4°C for 72 hours. In addition to the risk of lower protein yields due to slowed-down growth and synthesis rates at lower temperatures, the chaperone network may also not be as efficient as it could at low temperatures. Consequently, they affect protein folding processes. Such issues can be circumvented by using engineered hosts that possess cold-adapted folding mediators giving the host an enhanced ability to flourish at low temperatures. The ArcticExpress™ strain (B line; Stratagene), for example, harbors the cold-adapted chaperonin Cpn60 and co-chaperonin Cpn10 from the *Oleispira antarctica*, a psychrophilic bacterium; both allow thriving and high refolding activities at 4–12°C. Other remedies to inclusion bodies include changing vector components (e.g. promoter or signaling sequences), hosts, or other culturing conditions (e.g. pH or medium composition) (Wingfield, 2015).

Disulfide bonds are essential for many recombinant proteins to achieve biologically active 3D conformations. Failure in formation of disulfide bond can lead to misfolding and aggregation into IBs. To evade aggregation of disulfide bond-dependent proteins in *E. coli*, targeting to the periplasmic space is an implementable approach as it aids reaching to the folded state by promoting stability of folding intermediates by formation of the disulfide bridges (de Marco, 2009). Cysteine oxidation in *E. coli* rarely occurs in the cytoplasmic space but takes place in the periplasm where disulfide exchange reactions are catalyzed by a myriad of enzymes (Rosano & Ceccarelli, 2014). Improved yields have been realized after directed

localization of aggregate-prone and hydrophobic recombinant MPs to the periplasm or the expression medium by fusion tags (Pandey et al., 2016). Normally, the cytoplasmic environment has a more negative redox potential maintained by different redox enzyme systems (Rosano & Ceccarelli, 2014). Nevertheless, engineered *E. coli* such as the Origami (Novagen) and SHuffle (NEB) strains possess mutations that allow maintenance of an oxidative environment in the cytoplasm to favor formation of disulfide bonds and stabilize the proteins.

A fundamentally different strategy for recovery of recombinant MPs—though generally avoided by many researchers (Peleg & Unger, 2012)—is by consenting to the formation of inclusion bodies (IBs) of an aggregate-prone MP of interest. Promoting IB formation can also be achieved by fusion protein constructs (insoluble proteinaceous fusion tag and protein of interest). With IB formation, the protein of interest is shielded considerably from proteolytic degradation (and therefore concentrates in high amounts) and the cell is also protected from any potential toxic and detrimental interferences of the foreign protein with its functions. IBs themselves are usually not toxic to the cells (Dilworth et al., 2018). Although their formation simplifies purification of the protein of interest, albeit the protein is obtained in a denatured/aggregated form (Wingfield, 2015), Purification of functionally overexpressed MPs is usually preferred over IB purification because attempting to obtain authentically folded protein from an IB can be time-consuming and requires customized, protein-specific *in vitro* refolding protocols after the imposition of denaturation conditions on the protein aggregate (Baumgarten et al., 2017; Peleg & Unger, 2012). Despite the tremendous efforts in this area, few MPs were successfully refolded using this method. These include several GPCRs that have been refolded into functional states (Dilworth et al., 2018). Overall, in contrast to β-barrel MPs which can be readily isolated and refolded from IBs, helical bundle MPs specifically are difficult to isolate and refold as such attempts are usually unsuccessful (Schlegel et al., 2014). Another major drawback with refolding IB proteins is the reductions observed with recovery (Khow & Suntrarachun, 2012; Wingfield, 2015). Conclusively, there is simply no guarantee that *in vitro* refolding

of a protein of interest will generate high amounts of biologically active product (Khow & Suntrarachun, 2012). However, in some cases and after failure of functional overexpression, this approach may be the only option to produce a recombinant protein of interest (Rosano et al., 2019).

Protein Inactivity

Obtaining high amount of non-aggregated protein of interest is questionable in its reliability as an indicator of the quality of that protein (Rosano & Ceccarelli, 2014). A protein is considered to be of bad quality if it does not exhibit its specific activity. In such cases, although the protein is present in a stable conformation, it may not have folded completely wherein the exact architecture of the active site is not formed to perform its activity. In a recent study, two different prokaryotic multidrug ABC transporters were produced in *E. coli* strains and although strong expression was achieved in the membranes of several strains, drastic differences were observed in the functionality of these proteins (Mathieu et al., 2019). Moreover, mild detergents have extracted mainly active transporters whereas harsh detergents solubilized the transporters irrespective of their functionality. The study presents a clear falsification of the general assumption that MPs inserted into membranes and extracted by detergents are properly folded and functional. Also, it was shown that the widely used GFP fusion assay where fluorescence is used to attest to the folding and quality of MPs of interest, is not a reliable strategy *per se*. Thus, caution must be taken with this assay to avoid false-positives. Conclusions from this study were in line with those from an earlier one that studied folding of eukaryotic MPs produced in both baculovirus-infected insect cells and stable mammalian cell lines (Thomas & Tate, 2014). Modifications in the cultivation conditions such as lowering the culture temperature or the addition of small molecules or prosthetic groups needed by the overexpressed protein to acquire its final folded form, can both result in a significant improvement in protein quality and functionality (Rosano & Ceccarelli, 2014). Co-expression of post-translational machineries such as chaperones or translocation machineries are also shown to facilitate protein folding

and improve conformational and protein quality (He et al., 2014; Nannenga & Baneyx, 2011; Rosano & Ceccarelli, 2014).

Non-functional proteins—especially eukaryotic ones produced in *E. coli*—are in many cases obtained due to the lack of correct PTMs such as glycosylation or disulfide bridges (Bernaudat et al., 2011; Peleg & Unger, 2012). Glycosylation is considered to be the most critical PTM controlling protein quality as indicated by the documents from regulatory agencies around the world such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) (Amann, Schmieder, Faustrup Kildegaard, Borth, & Andersen, 2019). A traditional solution to the lack of an essential PTM is to simply utilize a different overexpression host system that is better suited to carry out the desired PTM task. PTMs such as glycosylation are also not only determined by a specific species or cell type but are also affected by the culture conditions. On the other hand, several sophisticated genetic engineering toolboxes are now available and are promising for the various pro- and eukaryotic expression platforms. They can be used to improve host cell lines or engineer them to provide a specific lacking PTM that is required for a certain bioproduct. Classical engineering strategies rely on the overexpression or disruption of individual proteins and enzymes in pathways known to mediate naturally occurring PTMs. However, more targeted and fine-tuned genetic engineering incorporate the use of molecular tools such as non-coding RNAs (shRNA, microRNA and siRNA) and the genome-targeting system called clustered regularly interspaced short palindromic repeats (CRISPR) (Amann et al., 2019). A prerequisite to implementation of such cell line editing approaches is an extensive knowledge on genomic sequences as well as the expression and function of endogenous proteins including tracking their omics blueprints at all cellular levels: the genome, epigenome, transcriptome, proteome and reactome. A sample material overexpressed and purified is also of bad quality if it exists in heterogeneity (Wagner et al., 2006); and although various approaches have been used to prevent PTM-based as well as proteolytic degradation-based heterogeneity, combining targeted genetic engineering methods and data from functional studies with *in-silico* tools now allow

for the control of optimal expression for a homogeneous and high quality protein processing, and even the engineering of entire cellular pathways (Amann et al., 2019). For *E. coli* in particular, even though PTMs have been obtained through engineered strains, tools for PTMs in this host remain limited and more research is needed in this area (Amann et al., 2019; Khow & Suntrarachun, 2012; Rosano et al., 2019).

In another scenario which results in bad protein quality and lack of activity, a toxic protein overexpressed in the host may cause genetic alterations in the expression vector leading to the loss of activity of the protein as well as permitting survival and proliferation of the host (Rosano & Ceccarelli, 2014). Purification followed by DNA sequencing of the plasmid before and after induction can detect the plasmid's structural instability as point mutations, deletions, insertions, or rearrangements. Circumventing this issue can take place by using strains that ensure plasmid stability (e.g. *recA*$^{-}$) or by introduction of the foreign gene into the host before each expression round.

Protein Degradation

Protein folding and proteolytic degradation are intimately linked (Baneyx, 1999). Cellularly integrated catabolic systems are an efficient way to conserve cell resources as proteins that are improperly folded or damaged beyond repair are recycled back into their constituent amino acids. Expression hosts deficient in key proteolytic systems can produce recombinant proteins with extended lifetimes, allowing their accumulation (Rosano et al., 2019). Degradation of the protein of interest can also be alleviated by manipulating culture conditions, using exogenous protease inhibitors, secretion to the periplasm or the medium, co-expression of post-translational machineries, engineering proteolytic cleavage sites to stabilize the protein, inducing IBs formation, or changing the expression system altogether (Wingfield, 2015).

2.3 AMINO ACID TRANSPORTER MEMBRANE PROTEINS

2.3.1 Conceptual Background

The regulated movement of substances such as nutrients, wastes, ions, as well as certain drugs across biological membranes is an area of cellular activity that is governed by a subset of MPs known generally as transporter proteins (Bianchi, van't Klooster, Ruiz, & Poolman, 2019; Hediger et al., 2004). These membrane-integrated proteins, often called the gatekeepers of cells, control a process that is essential to the healthy life of any living cell. For instance, amino acid transporters (AATs) play critical roles in regulating diverse cellular activities such as gene expression, protein synthesis, neurotransmission, energy metabolism, redox balance, signal transduction pathways, cell survival and growth, and cell volume regulation (Kandasamy, Gyimesi, Kanai, & Hediger, 2018; Schweikhard & Ziegler, 2012). Membrane transport systems are encoded by around 10% of the human genome (César-Razquin et al., 2015) and they can be classified broadly into channels (also known as pores or porins) and transporters (also known as carriers, porters, or permeases) (Bianchi et al., 2019; Bosshart & Fotiadis, 2019; Hediger et al., 2004). Channels undertake a conformation that is double-opened on both sides of the membrane during transport of a substrate and they catalyze highly specific movements down their concentration gradient (downhill) by the simplest process and driving force known as facilitated/passive diffusion. Secondary transporters represent one of two common categories of membrane transporters (MTs), the other being primary active transporters that use energy from various possible sources (e.g. chemical, electrical or light) to drive the movement of substances against their concentration gradient (uphill) across the membrane. The hydrolysis of ATP is a well-known process used by cell transporters such as ATP-binding cassette (ABC) transporters and ion pumps (ATPases) to drive the primary active transport of a variety of different substances. MPs in the category of secondary transporters are subdivided into uniporters, symporters and antiporters. Uniporters are driven by the energy stored in their

substrate's concentration gradient and thus facilitate the movement of solutes down their concentration gradient; whereas symporters and antiporters move their substrate against its concentration gradient using energy from the cotransport of another solute in a downhill flux and hence, are called secondary active transporters. In other words, these transporters fascinatingly use gradients generated across the membrane by primary energy sources and couple them to the active transport of their specific substrates. The differences between uniporters, symporters and antiporters can be addressed generally based on the number of substrates shuttled per one translocation cycle (one or two substrates), the directionality of transport (in terms of the co-transported solute and/or concentration gradient), and the type of substrate(s) shuttled. Beyond channels and primary and secondary transporters, other types of special membrane transport systems exist. For instance, tertiary active transporters use gradients generated by secondary active transporters and can generate their own gradients (Lin, Yee, Kim, & Giacomini, 2015). Another type of MTs couples the transport of a substrate to its chemical modification to release a chemically modified substrate (Bosshart & Fotiadis, 2019). To generalize, transporters have a relatively lower turnover rate ($\sim 10^2$–10^4 substrates/sec) compared to channels ($\sim 10^7$–10^8 substrates/sec).

The major facilitator superfamily (MFS) is the largest evolutionarily related and most diverse superfamily of secondary transporters. The MFS harbors more than 80 distinct families and can be found throughout all kingdoms of life. In bacterial genomes, MTs are encoded for by a significant portion of the genome (16%). In *E. coli*, over 500 different MTs have been annotated and 25% of them are members of the MFS. Irrespective of differences in the primary protein sequence, all MFS MPs adopt a common structural fold known as the MFS fold (Vishwakarma, Banerjee, Pasrija, Prasad, & Lynn, 2018). The amino acid/polyamine/organocation (APC) superfamily holds the status of being the second largest superfamily of secondary transporters (Wong et al., 2012). Currently, the APC superfamily is made up of 18 families and its members operate mainly as solute:cation symporters or as solute:solute antiporters (Bosshart & Fotiadis, 2019). Despite sharing a relatively

low level of sequence identity, families from this superfamily have a common and a unique protein fold—the LeuT fold. Like the MFS, the APC superfamily is found across all phyla from prokaryotes to higher eukaryotes (total of ∼250 members) and its members have polypeptide chains that vary in lengths between 350 and 850 amino acid residues (Reig et al., 2007).The APC superfamily was originally recognized to be essential, nonexclusively, in the transport of amino acids and/or their derivatives (Jack, Paulsen, & Saier, 2000). However, it is now established that its members transport a variety of many other solutes such as sugars, nucleosides/-bases, inorganic sulfates, osmoprotectants, and neurotransmitters (Wong et al., 2012). Although members of the MFS and the APC superfamily exhibit different topologies, both superfamilies appear to have risen from an ancestral two α-helical transmembrane (TM) segments hairpin structure that had undergone intragenic triplication followed by loss of one of the TM segments in the APC superfamily, thereby giving the repeats characteristic of these two superfamilies (Wong et al., 2012).

By far, the undisputedly largest superfamily of MTs—larger than the classes of ABC and ATPases transporters—is the solute carrier (SLC) superfamily engaged in the secondary (facilitative/active) transport (influx/efflux) of a plethora of small molecules that range from nutrients (e.g. sugars, amino acids, vitamins, minerals etc.) to electrolytes, signaling molecules and certain drugs across membranes of cells and organelles (Bai, Moraes, & Reithmeier, 2017; César-Razquin et al., 2015; Garibsingh & Schlessinger, 2019; Hediger, Clémençon, Burrier, & Bruford, 2013; Kandasamy et al., 2018). **Table 2.6** outlines the types of proteins classified as part of the SLC superfamily and those that are not. Among all human MPs, the size of the SLC superfamily comes only second after the family of GPCRs, the largest family of MPs encoded by the human genome. In humans, the SLC superfamily encompasses over 420 annotated members found expressed ubiquitously throughout the whole body. Due to the lack of sequence homology and structural similarity between families of the SLC superfamily, all SLC genes have been classified solely based on the function of the corresponding protein as a solute transporter (Rives,

Javitch, & Wickenden, 2017). Additionally, by definition, a specific SLC is assigned to a specific SLC family if it shares a sequence identity of 20-25% with other members of that family (Hediger et al., 2004). However, few exceptions can exist as the SLC51 family in which its members, SLC51A and SLC51B, are not related by any sequence similarity (Hediger et al., 2013). Even though the SLC series was developed originally for human genes, SLCs can now be identified throughout prokaryotic and eukaryotic species (Bai et al., 2017). The prokaryotic homologs of AAT SLCs, for instance, participate in important physiological processes such as maintaining homeostasis and responding to environmental stressors such as acidity and osmotic stress (Schweikhard & Ziegler, 2012). As classified based on sequence identity by the human gene nomenclature committee, AATs are found throughout various families of the SLC superfamily (Kandasamy et al., 2018). 11 out of a total of 65 different SLC families contain AATs that in turn add up to at least 66 known members. AATs of the SLC superfamily show vast structural heterogeneity and diverse translocation and energy harnessing mechanisms. For example, AAT members of different SLC families and even whole families of SLCs cluster into diverse structural classes or folds where the MFS and the APC structural superfamilies represent the most common clusters. Most SLCs operate as symporters and antiporters and use various energy coupling mechanisms to drive their transport activity, while a few others demonstrate channel-like properties (Bai et al., 2017). SLCs, including those involved in amino acids transport, are implicated heavily in playing significant roles in human health and when dysfunctional, the onset of a wide array of diseases that range from central nervous system diseases to metabolic disorders, diabetes and cancer (Garibsingh & Schlessinger, 2019; Kandasamy et al., 2018). Despite that, unlike other “star” families such as GPCRs and protein kinases that have comparable sizes, physiological importance and pharmacological relevance to the SLC superfamily, SLCs have long been understudied and minimally targeted with very few tool compounds and clinically approved drugs. This is due to a multitude of reasons such as the challenging technical issues to obtain biological reagents (e.g. antibodies) and to express, purify and detect these difficult and

complex MPs (César-Razquin et al., 2015). These translate into challenges in biochemical, biophysical and structural characterizations of SLCs. However, efforts have begun to pour towards SLCs in recent years in areas such as experimental determination of molecular structures and development of drugs.

Table 2.6 An outline of major membrane transport systems and their classification as SLC proteins or non-SLC proteins (Hediger et al., 2004; Kandasamy et al., 2018)**.**

SLC Proteins	Non-SLC Proteins
✓ Coupled transporters ✓ Passive/facilitated transporters ✓ Exchangers (antiporters)	✓ ATP-dependent transporters (e.g. ABC transporters and ion pumps) ✓ Channels (e.g. ion channels and aquaporins) ✓ Ionotropic receptors ✓ Subunits of transporters and channels ✓ Auxiliary/regulatory transport proteins

2.3.2 L-arginine/agmatine Antiporter: A Hero of Extreme Acid Resistance, and More

Natural evolution has allowed microorganisms to adapt optimally to grow and carry out cellular functions in their environment but has allowed many of them to survive and adapt to minor stressor changes in their environment such as declines in the pH by enabling acid tolerance (Guan & Liu, 2020). Looking at orally ingested enteric bacteria, these on the other hand are confronted with lethal bodily attacks from the high acidity of the stomach ($1.5 < pH < 3.0$) while on their journey towards breaching the gastric barrier and gaining access into the intestines of the host organism (S. Gong, Richard, & Foster, 2003; Kanjee & Houry, 2013). Although most microbial species such as *Vibrio cholerae* are incapable of handling such extreme stress and therefore rely on massive insults in the order of billions of cells in hopes that a handful of survivors fulfill the mission of infection, other species are remarkably well-equipped with highly effective and sophisticated acid resistance

(AR) mechanisms that enable survival in such extreme and inhospitable environments and thus help in dramatic reductions of the dose of cells needed for infection down to a very small number (hundreds or less). The family known as *Enterobacteriaceae* delivers some of the most encountered and virulent pathogens including strains of *E. coli*, *Shigella*, *Salmonella*, and *Yersinia*. (Iyer, Williams, & Miller, 2003). Extreme AR in these infectious species, as well as the naturally colonizing benign ones, allows them to withstand acidic shocks for up to several hours, thus buying time until safe passage from the harsh environment of the stomach and into the gut. However, it is still important that although a neutral/basic pH defines the intestinal environment, high fermentation rates driven by the richness of carbohydrates and lack of oxygen in the intestines produces neutral short-chain fatty acids (SCFAs) that cross through the cell membranes of pathogens, dissociate in the cytoplasm, and result in a pH-dropping effect that the cells are compelled to manage again (Kanjee & Houry, 2013).

Mechanisms of AR in bacteria are many and diverse (Guo et al., 2019; Kanjee & Houry, 2013). They operate differently at different values of the pH scale and in different microbial species. AR mechanisms have complex genetic regulations that are slowly but steadily being studied over the years (Aquino et al., 2017). Although our knowledge on gene expression control of these systems is evolving, large areas remain unexplored and await to be uncovered. Furthermore, with help from the field of synthetic biology, their importance is seen to have great exploitability and application in a variety of areas of biotechnology such as industrial bioprocesses and environmental bioremediations (Yuping Liu, Tang, Lin, & Xu, 2015). AR mechanisms can generally be classified into passive and active AR systems (Kanjee & Houry, 2013). The former is dependent upon the buffering capacity of amino acids, proteins, polyamines and phosphates. The latter on the other hand is grouped broadly according to their cellular adaptation level into physiological, metabolic, and proton-consuming systems. **Table 2.7** briefly provides a few insights into the functional aspects of these active AR systems operating at each of these levels.

Table 2.7 Different levels of adaptation for the active AR systems in a bacterial cell and some insights into the functional aspects of the systems present in each level (Guan & Liu, 2020; Kanjee & Houry, 2013)**.**

Level	Insights into Active AR Systems
Physiological	- Reduction of proton influx by modifications in membrane composition and by blocking of outer membrane porins (OMPs). - pH-dependent Periplasmic/cytoplasmic chaperones and other proteins are available to manage acid-induced damage by protecting cellular constituents such as the DNA and other functional proteins.
Metabolic	- Stress response systems (e.g. periplasmic and oxidative responses) are induced/upregulated when growing in mild acidic stress and during either aerobic or anerobic conditions. - Transport and metabolism of secondary carbon sources that produce less acids when metabolized compared to glucose. - Coupling of Proton efflux to energy generation via components of the electron transport Chain (NDH-I, NDH-II, SDH, and cytochromes) during aerobic growth in mild acid stress.
Proton consumption	- Catalysis of the production of hydrogen gas from cytoplasmic protons by formate hydrogen lyase (FHL) complex under anaerobic conditions. - AR systems involving PLP-dependent amino acid decarboxylases and inner membrane amino acid antiporters. - Production of ammonia by the proton-neutralizing urease system.

Proton consumption-based AR systems consists of an interesting subset of systems, each comprised of a PLP and amino acid-dependent decarboxylase that catalyzes a cytoplasmic reaction involving proton-dependent decarboxylation of a substrate amino acid into a product and CO_2. An inner membrane antiporter working in coordination to facilitate export of the cytoplasmic decarboxylation product and

simultaneous import of the periplasmic substrate amino acid. By keeping the described system under this cyclical-like operation, intracellular acidification and damage are alleviated by the acid-activated, decarboxylase-mediated proton consumption and the accompanying efflux and influx of decarboxylation products and substrates. With respect to this amino acid-dependent AR mechanism, four distinct decarboxylase/antiporter systems are currently known and can be found in both Gram-positive and Gram-negative bacteria (Guo et al., 2019); summarized in **Table 2.8**. GDAR and ADAR systems mediate robust protection against extreme acid stress conditions while LDAR and ODAR systems operate most efficiently under mildly acidic stress conditions (Kanjee & Houry, 2013). Such a trend in efficiency at a given pH is attributed to the pH-dependent optimal enzyme activity of each of the decarboxylases in the systems. A sharp decrease in enzyme activity (GDAR > ADAR > LDAR >> ODAR) is observed as the pH increases. Also, the extreme AR systems (GDAR and ADAR) have been implicated in the enhancement of survival of cells during exposure to SCFAs.

Table 2.8 Amino acid-dependent decarboxylase/antiporter proton consumption AR systems and their components (Kanjee & Houry, 2013)**.**

System	Decarboxylase component	Antiporter Component
Glutamic acid-dependent acid resistance (GDAR) system	Inducible glutamic acid decarboxylases, GadA/GadB	Glutamate/γ-aminobutyric acid antiporter, GadC
Arginine-dependent acid resistance (ADAR) system	Inducible arginine decarboxylase, AdiA	Arginine/agmatine antiporter, AdiC
Lysine-dependent acid resistance (LDAR) system	Inducible lysine decarboxylase, LdcI	Lysine/cadaverine antiporter, CadB
Ornithine-dependent acid resistance (ODAR) system	Inducible ornithine decarboxylase, SpeF	Ornithine/putrescine antiporter, PotE

AdiC is an integral transmembrane protein with 12 membrane-spanning α-helices and is found in the inner plasma membrane of bacteria such as *E. coli* and *Salmonella enterica* serovar Typhimurium (95% similarity to *E. coli* AdiC). It is a member of the basic amino acid/polyamine (APA) family of transporters (Jack et al., 2000; Kanjee, 2012) and is a prokaryotic homolog in the SLC7 family based on the SLC series nomenclature and classification (Fotiadis, Kanai, & Palacín, 2013; Ilgü et al., 2016). AdiC is made up of 445 amino acids and has N and C-termini found in the cytoplasmic region of the cell and a polypeptide chain with a theoretical molecular weight (MW) of 46.8 KDa (Bosshart et al., 2008; Fang, Kolmakova-Partensky, & Miller, 2007). Research on AdiC has enriched our understanding at various levels such as molecular genetics, protein biochemistry, structure, and transport functions (Ilgü et al., 2014, 2016). AdiC is the antiporter component and the pumping heart of the ADAR system involved in the arginine-dependent extreme AR response (Kanjee & Houry, 2013). The ADAR system is activated maximally at low pH, under anaerobic growth conditions, and in complex medium. Briefly, the inducible decarboxylase enzyme known as AdiA, catalyzes, in a PLP-dependent reaction that consumes acidifying protons leaking from the extracellular milieu into the cytoplasm, α-decarboxylation of L-arginine present inside the cell to produce agmatine and CO_2. Agmatine acts as a vehicle that carries one proton that is exported by AdiC using the free energy of decarboxylation, while co-importing from the extracellular environment, and in a one-to-one exchange stoichiometry, L-arginine, the substrate amino acid of the next decarboxylation reaction (Fang et al., 2007). Here, L-arginine acts as a new and an empty carrier for another intracellular proton. A schematic diagram that describes the decarboxylation/transport theme of the ADAR system is shown in **Figure 2.2**. In this system, the concentration gradient of agmatine across the bacterial membrane drives the uptake of L-arginine into the cytoplasm (Bosshart & Fotiadis, 2019). However, on the other hand, transport of unequally charged AdiC substrates generates an outwardly directed positive current and thus, an inside negative voltage that the cell relieves by certain precautionary mechanisms that function to prevent protons leakage and compromised membrane

integrity caused by hyperpolarization (Fang et al., 2007; Foster, 2004). In addition to the established knowledge that the ADAR system is activated under extreme acid conditions, the transport activity of AdiC itself was also proven later by its 3D structures, to be pH-dependent with high activity at lower pH (Kanjee & Houry, 2013). Additionally, these values were found to be different at each side of the membrane and indeed, are in good agreement with the internal and external pH values that are known to induce an extreme acid response.

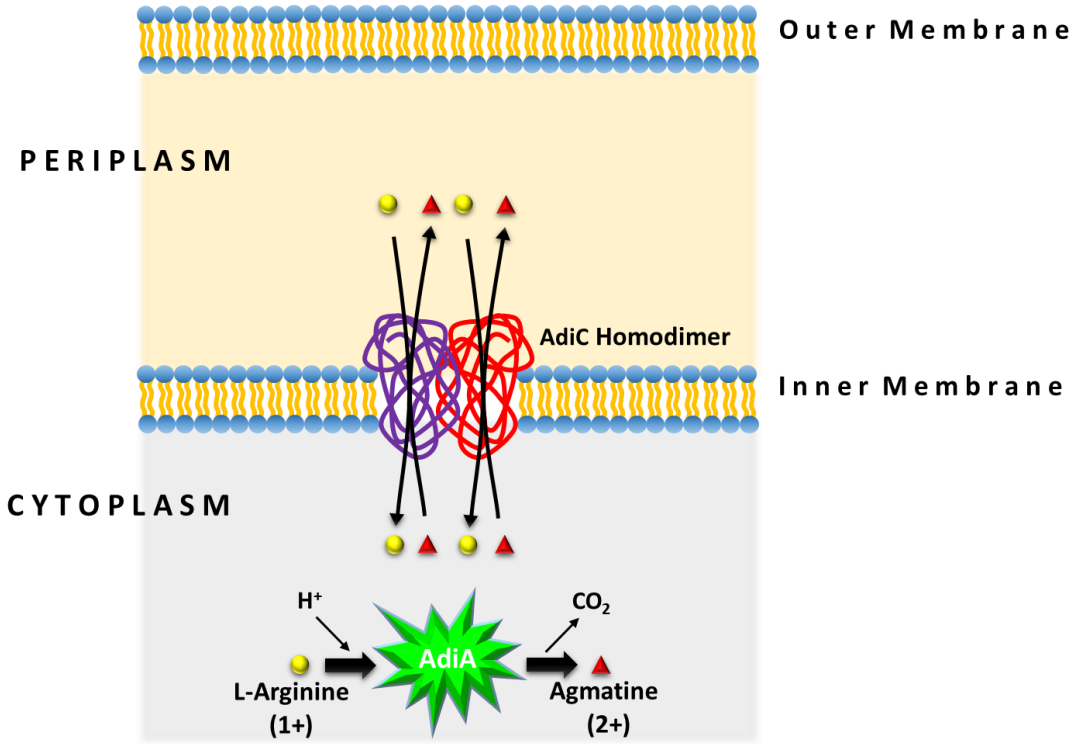


Figure 2.2. A synoptic diagram illustrating the general scheme of operation of the ADAR system (Kanjee & Houry, 2013)**.** AdiC is shown in its homodimeric form integrated into the inner plasma membrane. AdiC Monomers are in different colors (purple and red) only to emphasize the MPs' oligomerization state. The ADAR system confers AR to its host by eliminating cytoplasmic protons through their consumption in the AdiA-mediated decarboxylation of L-arginine (yellow spheres) into agmatine (red triangles) and CO_2. In the grand scheme of the ADAR system, AdiC acts as a "virtual proton pump" by continuously and coordinately expelling generated Agmatine (charged as 2+) while simultaneously admitting Arginine (charged as 1+) to be used up in the next proton-consuming decarboxylation reaction.

Although the role of amino acid decarboxylases in relieving the stress of an internal acidic pH was suggested back in 1946 (Gale, 2006), discovery of the role of AdiC (known earlier as yjdE and defined as a putative MP with 12 TM helices and a hypothetical transport function) as the antiporter of the arginine-dependent process of extreme AR has not been made until 2003, almost 60 years later (S. Gong et al., 2003; Iyer et al., 2003). Preliminary biochemical and functional characterizations of AdiC were carried out later by Fang et al. and have set a stage for further investigations such as ones concerned with the structural and mechanistic features of AdiC (Fang et al., 2007). Some of the important findings in their study included that AdiC: is a homodimer in detergent micelles (decyl maltoside) and lipid membranes by glutaraldehyde crosslinking experiments, does not transport the D-configured stereoisomer of arginine (only the L-configured form), and has an electrogenic transport mechanism by virtue of the unequal electrical charges featured on its transported substrates arginine and agmatine.

As stated earlier, regulation of these AR systems is highly complex (Aquino et al., 2017; Foster, 2004). Among bacteria with amino acid-dependent extreme AR systems, *E. coli*, perhaps unsurprisingly, represents the species with the most characterized and studied systems. At a transcriptional level, the main regulatory elements have been determined. Though not elucidated as profoundly as the GDAR system, regulation of the ADAR system was shown to occur through AdiY (a member of the XylS/AraC family of transcriptional regulators) as the primary regulator where, only when overexpressed, it positively controls expression of genes, *adiA*, encoding for the enzyme arginine decarboxylase, and *adiC*, encoding for AdiC antiporter (Foster, 2004; Stim-Herndon, Flores, & Bennett, 1996). Deletions in chromosomal *adiY* gene, however, showed no effect on the transcription of *adiA* and *adiC* (S. Gong et al., 2003). These three genes are clustered in the *adi* locus on the chromosome one after the other as shown in **Figure 2.3.** However, they are not thought to form an operon (Kanjee, 2012).

CysB, the major transcription factor (TF) of the cysteine biosynthetic pathway, is also another proven transcriptional regulator that works directly with

AdiY as an activator of *adiA* and *adiC* genes (Foster, 2004; S. Gong et al., 2003). Mutants of CysB were found clearly defective in ADAR. CysB is suggested to act as a sensor for conditions that activate ADAR (anaerobic, low pH, and rich medium conditions). The histone-like protein HU has also been suggested in positive regulation of AdiA and AdiC levels (Bi, Sun, Fukamachi, Saito, & Kobayashi, 2009) whereas the transcriptional regulator of the LDAR system, CadC, has been shown to downregulate *adiA* expression (Casalino et al., 2010). Finally, the master and global regulator of at least 250 genes, the nucleoid-associated H-NS protein, lies near the top of the hierarchy of the complex regulatory network of extreme AR where it indirectly controls repression of the GDAR, ADAR and LDAR systems via the network's central regulator, the RcsB-P/GadE complex (Krin, Danchin, & Soutourina, 2010). H-NS is also found to directly modulate different AR genes such as *adiY* and *cadC* by binding to their promoter regions. Clearly, it is undeniable that *E. coli* has evolved to manage a vast number of regulatory proteins for complex control of a seemingly simple strategy to survive low pH (Foster, 2004). However, such extensive regulatory networks point towards the elegant integrity of extreme AR systems not only amongst themselves but also with many other different networks for survival in alkaline conditions and involved in central aspects of cell physiology and metabolism (Foster, 2004; Guo et al., 2019; Yohannes, Barnhart, & Slonczewski, 2004). Indeed, for example, a recent work has demonstrated a coordinated interconnection between regulatory networks of AR systems and the broader cellular metabolomic pathways of carbon and nitrogen (Aquino et al., 2017).

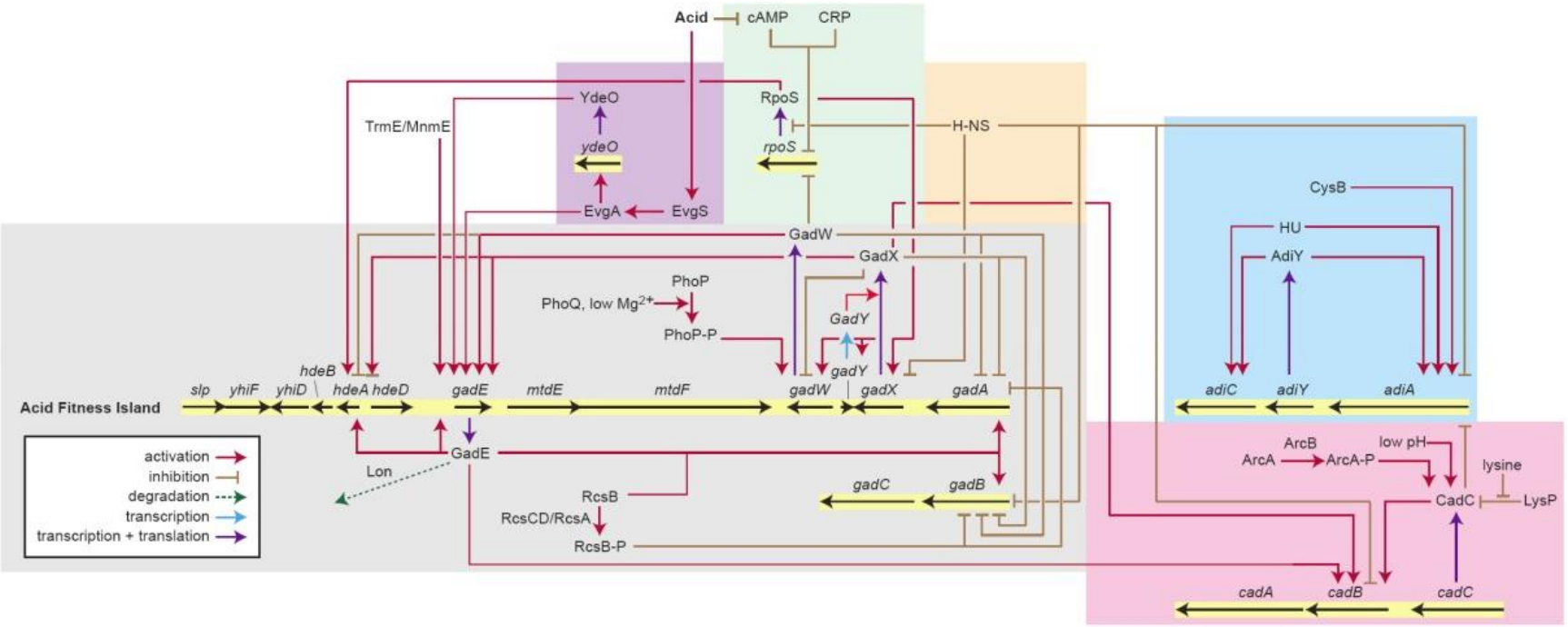


Figure 2.3. The Regulatory circuits of AR systems in *E. coli* (Kanjee, 2012)**.** Collections of sub-circuitries controlling different AR systems in *E. coli* are shown in boxes of different colors where the GDAR, ADAR and LDAR systems are shown in grey, cyan and pink boxes, respectively. The global regulatory circuits of RpoS, H-NS, and EvgA/S are shown in pale green, pale orange and purple boxes, respectively. Intertwining between regulation of the bio-circuitries is also emphasized and can be seen. Black arrows with yellow backgrounds portray the chromosomal organization of the different AR genes (gene names are shown in italic) in an AR system. The key for the different bioprocesses occurring (activation, inhibition, degradation, transcription, and translation) is shown in the grey box. (This figure is courtesy of Dr. Usheer Kanjee).

AdiC, and the other three amino acid antiporters (GadC, CadB and PotE) of extreme AR are known members of the APC superfamily, the 2nd largest superfamily of secondary transporters with members that share a relatively low level of sequence identity but a common structural fold (Krammer & Prévost, 2019). The sequence identities for the four AR antiporters range between 20% and 38%. The first 10 TM helices of the AR antiporters are known to form the protein core with the first and the second five helices being related by a pseudo-twofold rotational symmetry, a feature that is known as a 5+5 inverted repeat and is conserved in other transporter families such as the Na^+-coupled symporters with members like LeuT, BetP and MhlP (Kanjee & Houry, 2013). TM11 and TM12 on the other hand are known to be involved in the homodimerization of the antiporter proteins. Functionality of a homodimer does not require crosstalk between each of its subunits which are completely capable of independent transport as it is shown in the scheme in **Figure 2.2**. So far, AdiC has been the best studied antiporter biochemically, structurally and functionally (Krammer & Prévost, 2019).

Valuable insights into substrate binding and transport mechanisms were revealed by its high-resolution X-ray crystal structures as well as the mutagenesis, binding and transport experiments (Fang et al., 2009; Gao et al., 2009, 2010; Ilgü et al., 2016; Kowalczyk et al., 2011). These efforts in turn helped identifying substrate selectivity of AdiC and accelerating acquisition of valuable knowledge on different eukaryotic APC superfamily transporters (Krammer & Prévost, 2019). An overview of the currently available 3D crystal structures of AdiC elucidated in three different conformations is given in **Table 2.9**. AdiC, as are almost all transporters, is proposed to alternate between two major conformations: periplasmic-open (PO) and cytoplasmic-open (CO) conformation (Kanjee & Houry, 2013; Krammer & Prévost, 2019). In each of these conformations, a substrate-binding site is exposed to bind a substrate near the corresponding leaflet of the membrane bilayer. For the periplasmic-open conformation of AdiC, three distinct gates were proposed to control arginine binding and its translocation: a proximal, middle, and a distal gate. Substitution of amino acid residues in these gates alters/destroys the transport

function of AdiC. After stabilized binding of a substrate (Arg or Agm) for the subsequent translocation, a series of conformational changes are thought to occur including transitioning through occluded states where the substrate is buried deep in the transporter. In particular, to arrive from a PO conformation and finally to a CO conformation, two models were proposed in which the movement action of the TM helices is described after they were grouped in both models into two functional units: the bundle (or gate) domain (TMs 1, 2, 6 and 7) and the core domain (TMs 3, 4, 8 and 9). For the other TM helices (5, 10, 11, and 12), both models agree on the absence of any significant movements. In the first model, it is suggested that the pseudosymmetry of the 5+5 inverted repeat is a protein movement constraint, and that both functional units pivot around the central pseudosymmetry axis (Kowalczyk et al., 2011). In the second model, the core domain remains fixed while the gate domain carries out major rigid-body movements (Ma et al., 2012).

Table 2.9 An overview of the current 3D crystal structures of AdiC elucidated in its three different conformational states (Krammer & Prévost, 2019)**.**

Structure / Criterion	PO, substrate-free[Δ*⁂]	PO, substrate-bound[⁋]	Occluded, substrate-bound[†⁂]
PDB code	3LRB; 3NCY; 5J4I	3OB6	3L1L; 5J4N
Resolution (Å)	3.6; 3.2; 2.2	3.0	3.0; 2.9
Oligomeric state	Dimer; Dimer; Dimer	Dimer	Monomer; Dimer
Substrate	–; –; –	Arginine	Arginine; Agmatine
Mutations	–; –; –	N101A	N22A; –
pH	7;8;7	8.5	8; 7
Organism	*E. coli*; *S. typhimurium*; *E. coli*	*E. coli*	*E. coli*; *E. coli*

Structures' references: (Fang et al., 2009[*]; Gao et al., 2009[Δ], 2010[†]; Ilgü et al., 2016[⁂]; Kowalczyk et al., 2011[⁋]). A unique symbol is attributed to each of these references. The order of information in a table cell is given with respect to the order of symbol(s) shown with each of the three structures.

As seen earlier, substrate uptake activity of AdiC is pH-dependent and exhibits two different pH sensitivities at each side. Experiments on AdiC-oriented

liposomes have shown that differences in the activity profile on each side suggest the presence of at least two pH sensors in the antiporter (M. F. Tsai, Fang, & Miller, 2012). Glu208, a residue of the distal gate, and Tyr74, present on the cytoplasmic side, were both suggested to function as the aforementioned AdiC pH sensors (Gao et al., 2009; S. Wang, Yan, Zhang, Chu, & Shi, 2014). The protonation state of Glu208 was implicated to play a significant role in the arginine selectivity of AdiC more towards singly protonated carboxyl group of arginine (Arg^{1+}) than doubly protonated (Arg^{2+}). This selectivity in the periplasm is necessary to prevent futile transport cycles where Arg^{2+} is permitted in and Agm^{2+} is expelled, leading to a defected "virtual proton pumping" property that is useless for AR (Krammer, Gibbons, Roos, & Prévost, 2018). A net charge-based mechanism of arginine selectivity by AdiC, rather than a mechanism by direct recognition of protonation states, is supported by strong evidence from different sources including computational studies that involve MD simulations and molecular and quantum mechanics calculations (Krammer & Prévost, 2019).

Compared to the large variety of the eukaryotic members of the APC superfamily of transporters, their atomic resolution 3D structures have been scarce. For example, until recently, the only eukaryotic transporters structures solved for monoamine transporters of the neurotransmitter Na^+ symporter family belong to the drosophila dopamine transporter and the human SERT. Building up atomic-level structural knowledge on AdiC (and to a lesser extent, on GadC) has significantly sped up construction of important models of different eukaryotic amino acid APC transporters. This has allowed evolution of our understanding of these MPs function-, regulatory- and transport-wise. The valuable insights obtained at these levels were acquired by the structures of AdiC (and GadC) used as templates to build computer-assisted structural models of these eukaryotic amino acid APC transporters, and by extensive mutagenesis experimental procedures. Using AdiC as a model, at least ten different eukaryotic transporters have been investigated including three human transporters and members of the SLC7 family—LAT1, LAT2 and xCT.

CHAPTER 3

MATERIALS AND METHODS

3.1 Materials

Ultrapure water with a resistivity of 18.2 ΩM × cm (referred to hereafter simply as MQ-water) has been obtained from the Milli-Q® Synthesis A10 water purification system (Millipore SAS, Molsheim, France). The MQ-water was autoclaved for 20 min at 121 °C and used where indicated. Membrane homogenate aliquots were thankfully prepared and supplied to our laboratory by Dr. Hüseyin İlgü (Universität Bern, Switzerland). **Table 3.1** shows a subset of the materials used throughout this work by listing most of the chemical agents that were used and also their supplying company. The rest of the chemicals are noted throughout the chapter.

Table 3.1 A list of most chemicals used in this work and their suppliers.

Chemical	*Supplier*
Tris-HCl	Sigma-Aldrich, MO, USA
NaCl	Merck KGaA, Darmstadt, Germany
Glycerol	Aklar Kimya, Ankara, Turkey
DDM	Affymetrix, CA, USA
L-Histidine monohydrochloride monohydrate	Merck KGaA, Darmstadt, Germany
Imidazole	Sigma-Aldrich, Steinheim, Germany
Acrylamide/bis-acrylamide (29:1 solution)	nzytech, Lisboa, Portugal
Tris-Base	Sigma-Aldrich, MO, USA
SDS	nzytech, Lisboa, Portugal
APS	Bio-Rad, CA, USA
TEMED	Bio-Rad, CA, USA
2-Propanol	Riedel-de Haën, Seelze, Germany
Glycine	Merck KGaA, Darmstadt, Germany
Bromophenol blue	Merck KGaA, Darmstadt, Germany
β-ME	Merck KGaA, Darmstadt, Germany
Glacial acetic acid	EmirKimya, Ankara, Turkey
Methanol	Sigma-Aldrich, Steinheim, Germany
Coomassie® Brilliant Blue R-250	Merck KGaA, Darmstadt, Germany
Absolute ethanol	Merck KGaA, Darmstadt, Germany
PBS	nzytech, Lisboa, Portugal
Bradford reagent	Sigma-Aldrich, MO, USA
EDTA	EmirKimya, Ankara, Turkey
Ethidium bromide	AppliChem, Darmstadt, Germany
Boric acid	Merck KGaA, Darmstadt, Germany
Urea	Sigma-Aldrich, Steinheim, Germany
Sodium acetate trihydrate	Merck KGaA, Darmstadt, Germany
$RuCl_3$	Merck KGaA, Darmstadt, Germany
KAc	AppliChem, Darmstadt, Germany
$CaCl_2$	Merck KGaA, Darmstadt, Germany
MOPS	Merck KGaA, Darmstadt, Germany
KOH	Merck KGaA, Darmstadt, Germany
HCl	Sigma-Aldrich, Steinheim, Germany
Pure NaOH pellets	Merck KGaA, Darmstadt, Germany
Molecular Grade Water (DNase/RNase free) (dH_2O)	A.B.T.™ Laboratory Industry

3.2 AdiC Cloning, Overexpression, and Membrane Preparation

The amino acid transporter, AdiC, was cloned as described previously (Ilgü et al., 2014, 2016) into the pZUDF21 vector. Briefly, genomic DNA was extracted from *E. coli* cells strain XL1-Blue collected from bacterial liquid cell culture. The open reading frame encoding for AdiC was amplified by polymerase chain reaction (PCR) using the forward primer 5′-AAAAAAGCTTATGTCTTCGGATGCTGATGCTC-3′ and the reverse primer 5′-AAAACTCGAGATCTTTGCTTATTGGTGCATC -3′. PCR products were digested by HindIII and XhoI restriction enzymes and ligated into the vector to generate the construct pZUDF21-*adiC* with *adiC* gene encoding for a recombinant AdiC protein. The constructed vector encodes for a C-terminal amino acid tail (LELEVLFQGPVDHHHHHHHHHH) of AdiC comprising a Prescission (human rhinovirus 3C) protease cleavage site followed by a 10× His tag. The DNA construct was finally verified by sequencing and successfully transformed into BL21(DE3) pLysS, an *E. coli* strain that is commonly used to relieve toxic leaky expressions such as those from the overexpression of MPs (Wagner et al., 2008). AdiC was then overexpressed and purified as described earlier (Ilgü et al., 2014, 2016).

To overexpress AdiC in successfully transformed BL21(DE3) pLysS, the cells were grown at 37 °C, in Luria-Bertani (LB) medium supplemented with 0.1 mg/mL ampicillin and by shaking in an orbital shaker at 180 rpm. When an OD_{600} of 0.5–0.6 was reached, protein expression was induced using 0.3 mM IPTG and incubation was continued for 4 h at 37 °C. The cells were then harvested by centrifugation at 10,000×g for 10 min at 4 °C, and the pelleted cells were resuspended in Lysis buffer [20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol]. The cell suspension was stored at −20 °C and later thawed when needed for further use in membrane preparations.

For membrane preparation, the resuspended thawed cells were incubated for 5 min with bovine pancreas DNase I and then lysed using Microfluidizer M-110P (Microfluidics, Newton, MA) at 16,000 psi (five passages). Cell debris as well as

high-MW matter were separated by low-spin centrifugation at 12,000×g for 20 min at 4 °C. The supernatant was collected and used in ultracentrifugation at 150,000×g for 1.5 h at 4 °C to harvest the cell membranes into the pellet. The formed pellet was then resuspended and homogenized in a small volume of Lysis buffer. For instance, 15 mL final volume for membranes obtained from a 10 L cell culture. The membrane solution was aliquoted into 1.85 mL fractions (1–2 mL, typically) corresponding to 1 L cell culture. Finally, aliquots were frozen in liquid nitrogen, and were stored at −80 °C until needed for further use.

3.3 Purification of AdiC

To purify AdiC protein, two of the frozen aliquots were thawed by hand, briefly vortexed, and placed immediately in an ice bath. The samples were solubilized in 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 1.5% (w/v) DDM for 2 h at 4 °C on a slow rotational shaker. Chemicals were added consecutively to the membrane samples and the total volume of a single solubilization reaction mixture adds up to 7 mL. Detergent (DDM) is added lastly after dilution of the mixture by water and vortexing. Alternatively, if a pre-prepared solubilization buffer (SB) is added to the membrane sample, DDM is not included in SB and is added directly to the sample after the addition of SB and vortexing. After solubilization, the samples were transferred into centrifugation tubes and prepared for ultracentrifugation at 34,000 rpm for 20 min at 4 °C. Following ultracentrifugation, each of the two supernatants was transferred quickly into a corresponding 50 mL Falcon tube before the pellet comprised of membrane debris is released back again into solution. 10 µL of supernatant and the pellet resuspended in 7 mL of MQ-water were stored for later use in SDS-PAGE analysis. In the Falcon tubes, the supernatants were each completed to a final volume of 15 mL (~twofold dilution) by **purification buffer 5** (PB-5) containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 5 mM L-histidine, 0.04% (w/v) DDM to reduce association of non-specific binders to the affinity beads used in the next step. Each

sample was then added into a new corresponding 50 mL Falcon tube with 1 mL bed volume of TALON® Metal Affinity Resin (Takara, CA, USA) pre-equilibrated with PB-5 (see **appendix B** for pre-equilibration protocol). Finally, samples were incubated overnight at 4 °C on a slowly rotating shaker.

For metal affinity chromatography, AdiC-bound beads were transferred each into a corresponding column-based setup fitted with a plunging syringe (see **appendix C1**) to manually assist flow by gravity through the filtered columns (Promega). The rate of flow was allowed to be fast during the prewash step where the flow-through (FT) is passed through the filter from the loaded samples but was maintained slow during wash steps. Per one sample duplicate, the beads were washed twice with 5 mL of PB-5, and a third time with 5 mL of **purification buffer 6.5** (PB-6.5) containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 6.5 mM L-histidine, 0.04% (w/v) DDM. In a post-wash step, 3 mL of **wash buffer** (WB) containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 0.04% (w/v) DDM was added to the beads to get rid of L-Histidine. To recover a purer AdiC, all buffers (PB-5, PB-6.5 and WB) can be incubated with the beads for 10 min before filtration. During filtrations, the beads were not allowed to dry out, so that the protein of interest is not harmed. However, the beads were allowed to partially dry out when filtering out WB so that the elimination of L-histidine is maximized (filtration was stopped before bubbles of detergent began emerging from the tip of the column). 1 mL samples have been collected at every filtration including 10 µL from FT for later use in SDS-PAGE analysis. To detach bead-bound AdiC, 450 µL of an elution buffer (EB) containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 400 mM imidazole, 0.04% (w/v) DDM was added to the beads after the lower segment of the column was carefully cleaved off. The obtained miniature column was parafilm-sealed, gently inverted, and incubated for 10 min at 4 °C on a rotating shaker. AdiC was then eluted out by fitting the mini column into a 2 mL Eppendorf tube (see **appendix C2**) and spinning the Eppendorf-column assemblage in a bench-top microcentrifuge at 1000 rpm for 1 min. For short-term storage, purified AdiC,

now retrieved in the 2 mL Eppendorf tube, was stored at 4 °C until needed for further use.

3.4 Relative Protein Quantification

Rough measurements of the concentration of the purified AdiC sample were taken using BioDrop µLite Spectrophotometer (Biochrom Ltd, Cambridge, UK) at A_{280}.

3.5 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Briefly, to prepare a polyacrylamide (PA) gel and run an SDS-PAGE experiment, Hoefer's SE245 dual gel caster and SE250 mini vertical protein electrophoresis unit were used, respectively. Clean glass plates (from Bio-Rad or Amersham Biosciences, USA) with integrated or separated spacers were fitted into the gel caster and leakage was tested using MQ-water. After ensuring no leakage, water was discarded, and the plates were dried with Whatman paper. A 12% separating gel solution was prepared in a 50 mL Falcon tube using the recipe indicated in **appendix D1**; the solution makes two gels. Following addition of the acrylamide polymerization catalyst, N,N,N',N'-tetramethylethylene-diamene (TEMED), 4-5 mL of the 12% gel solution was immediately casted between each plate with care not to form bubbles when ejecting the solution from the pipette. The gels were allowed to solidify efficiently by the addition of a 1 mL layer of 100% isopropanol topping each of the casted gel solutions. After ~15 min into gel solidification, isopropanol was discarded, and residual isopropanol was absorbed carefully with a tissue paper. A 5% stacking gel solution was prepared in a 15 mL Falcon tube using the recipe shown in **appendix D2**. After TEMED addition, 1-2 mL from the 5% gel solution was casted immediately onto each of the two polymerized 12% gels. Plastic combs were then fitted into each gel to allow well

formation. After ~15 min, the combs were removed from the now-solidified stacking gels, and the two gel-plates sandwiches were clamped onto the electrophoretic apparatus using binding clips. The system was then filled up with 1× gel running buffer (recipe on **appendix E1**).

For SDS-PAGE sample preparation, 10 µL was taken from each of the eluted AdiC sample duplicates (and from each of wash 1, 2 and 3) and added into a new 1.5 mL Eppendorf tube to dilute a pre-added 2.5 µL of 5× loading buffer (recipe on **appendix E2**) down to 1× concentration. 5 µL from each of the supernatant, resuspended pellet, and FT was also taken and used for the same purpose while accounting for the new 5× loading buffer volume (1.25 µL). After mixing by pipetting, samples may be heated in a heat block for optimized protein denaturation; although in this work, heating at 95°C for 2 min did not produce any variation in the running of heated versus nonheated AdiC samples and the purity of the obtained bands (data not shown). Assisted by a sample loading guide, 5 µL from PageRuler™ Plus Prestained Protein Ladder (Fermentas) and the dyed samples were loaded into the wells. The electrophoretic unit was sealed and connected properly to a power supply (Fisher Scientific) providing a constant current of 70 mA for ~60 min (for two gels). As soon as the dye ran off the gel, power was disconnected, the unit was opened, and the gels were carefully collected from the plates and placed in a plastic container. Then, the gels were submerged in standard Coomassie staining solution (recipe on **appendix E3**), heated up for 30 sec in a microwave at a power of 550 W, and placed at 4 °C in a rotating shaker for overnight staining. Next day, the staining solution was poured away from the gels, and the de-staining solution (recipe on **appendix E4**) was then added. The gels were allowed to de-stain for 2 h at room temperature (RT) and without shaking. Finally, the gels were analyzed, photographed, and safely discarded.

3.6 Increasing AdiC Concentration

To concentrate purified AdiC in the elution samples, Amicon® Ultra-4 Centrifugal Filter Unit (Merck KGaA, Darmstadt, Germany) with a molecular weight cut-off (MWCO) of 50,000 was used. Firstly, the storage fluid in the filter device (stored at 4 °C) was discarded and the sample reservoir was filled up with WB. The conical centrifuge tube enclosing the capped filter device was placed into a fixed-angle centrifuge rotor and spun at 1000 ×g for 10 min at 4 °C to rinse the ultrafiltration membrane by at least 4 mL of the added buffer. After centrifugation, the ultrafiltrate was discarded and washing was repeated two more times. After the membrane was washed and enriched with WB, both duplicate AdiC elution samples were combined into one by loading them into the sample reservoir of the filter device which was then capped and centrifuged at 1000 ×g for 15 min at 4 °C. The protein sample was concentrated to a total volume of 250 µL and collected back from the filter device and into an Eppendorf tube using a suitable pipette. The membrane was then washed again three times with water and stored in 20% ethanol at 4 °C for future use.

3.7 Desalting/Buffer Exchange

Zeba™ Spin Desalting Column, 7K MWCO, 2 mL (Thermo Scientific) was used to desalt AdiC and relocate the protein into 1× phosphate-buffered saline (PBS) to which 0.04% DDM was included. PBS was prepared according to the manufacturer's instructions (dissolve 99.3 g of PBS in 1 L for 10× solution). For desalting/buffer exchange procedure, firstly, the column, stored priorly at 4 °C, was placed inside a 15 mL conical collection tube and centrifuged at 1000 ×g for 2 min at 4 °C to discard the storage fluid used to maintain the size-exclusion chromatography resin. All centrifugations in this protocol were done under the same centrifugation parameters and the column was capped loosely each time. After discarding the storage fluid, the column was washed thrice by adding 1 mL of MQ-

water at each wash accompanied by the respective centrifugation step. The column was then preequilibrated by washing thrice using, at every wash, 1 mL of the desired buffer (1× PBS + 0.04% DDM) followed by centrifugation. After equilibration, the column was placed into a new 15 mL tube to collect the protein sample in the new buffer. The 250 µL AdiC sample was loaded into the column and then centrifuged. After collection of the protein in 1× PBS + 0.04% DDM buffer, the column was washed again with 1 mL of MQ-water (three times) by centrifugation. The beads were then stored in 2 mL of 20% ethanol at 4 °C.

3.8 Quantification of AdiC

For a more accurate determination of AdiC concentration, a typical Bradford assay was conducted. As shown in the table in **appendix F1**, serial dilutions of a protein standard, Bovine Serum Albumin (BSA) (Sigma-Aldrich, MO, USA), were prepared to perform a microplate assay using the compatible 1× PBS + 0.04% DDM as a diluent (DDM is not expected to cause any interferences due to its low concentration in the assay). Bradford reagent (Linear range = 1–1400 µg/mL) was left at to RT to warm up from 4 °C. 5 µL from each of the prepared BSA standard dilutions was loaded in duplicates into the microplate wells. 5 µL was also loaded in duplicates from each of three dilutions (2×, 5× and 10×) of the original unknown AdiC sample. The final microplate layout is shown in **appendix F2**. To each of these wells, 250 µL of gently mixed Bradford reagent was added. The microplate was carefully shaken in a rotational motion on a smooth flat surface and samples were allowed to incubate for 10 min at RT in the microplate wells. Spectrophotometric absorbance readings were taken at 595 nm (A_{595}) using the plate reader Multiskan GO (Thermo Scientific) and the software program SkanIt RE 4.1. Lastly, the data were then tabulated, a standard curve was plotted, and an accurate concentration of AdiC protein was calculated.

3.9 Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

3.9.1 Synthesis and Preparation of the Combinatorial SELEX Library and Primers

We used a single-stranded (ss) DNA (complexity ~10^{15} sequences) that was synthesized by Integrated DNA Technologies™ (IDT) to form the generic oligonucleotide called oligo 487D: 5′-GCC TGT TGT GAG CCT CCT GTC GAA-$[N]_{53}$-TTG AGC GTT TAT TCT TGT CTC CC-3′. The 100-mer oligo was designed with a central randomized region of 53 nucleotides (N stands for any randomly inserted nucleotide from an equimolar mixture of A, C, T and G) flanked by the 5′ and 3′ primer hybridization constant regions that have 24 and 23 nucleotides, respectively, forming a rich random ssDNA library. Here, we converted this pool of ss-oligos in an Arktik™ thermal cycler block (Thermo Scientific, Vantaa, Finland) into a double-stranded (ds) DNA library by extension of an annealing forward primer called oligonucleotide 484 (IDT) having the sequence 5′-<u>TAA TAC GAC TCA CTA TA</u>**G** GGA GAC AAG AAT AAA CGC TCA A-3′ and using WizPure™ Taq DNA polymerase Kit (Wizbiosolutions, Seongnam, South Korea). In this *Taq*-extension reaction, the nucleobase in bold (guanine) and the downstream nucleobases of oligo 484 anneal specifically to the 3′ constant region of oligo pool (and in the opposite direction) leaving a 5′ overhang formed by the underlined sequence. *Taq* pol then extends both oligo 484 and 487D in their 3′ directions by incorporating the complementary dNTPs in the medium into the extending DNA biopolymer to form the dsDNA pool with sequences that are now 117 nucleotides in length. **Figure 3.1** illustrates the *Taq*-extension scheme of oligo 487D taking place in the thermal cycler. **Table 3.2** shows (per one PCR tube) the separate components of the mixture in the tube used in this non-amplifying polymerization reaction. For convenience, a master mix was prepared (excluding *Taq* pol) for 21 tubes. The thermal cycler temperature settings used to catalyze the extension reaction are also shown in **Table 3.2**.

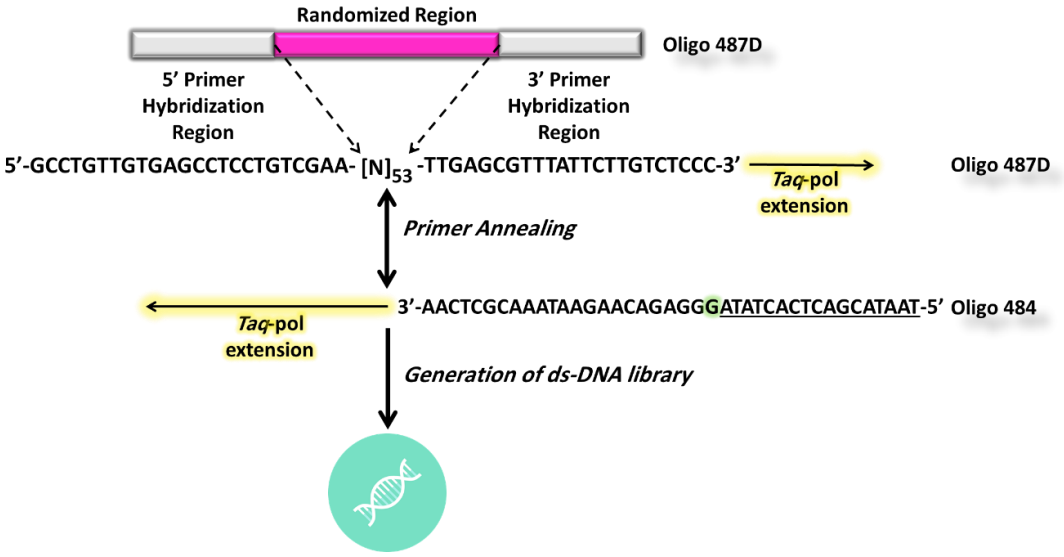


Figure 3.1. A mechanistic illustration of the extension reaction of oligo 487D taking place in the thermal cycler. After annealing of oligo 487D and oligo 484, *Taq* DNA polymerase extends them towards their 3′ directions using the opposite antiparallel strand as the template for extension. The final product is a dsDNA random library with sequences of 117 nucleotides (no amplification) with T7 promoter sequence underlined.

Table 3.2 Extension reaction mixture components and their amounts, and the temperatures used in the extension program (stock concentrations and working volumes are shown)**.**

Reaction Component	***Volume (µL)***
Oligo 487D (100 µM)	0.835
Oligo 484 (100 µM)	1.670
dNTPs (2.5 mM)	10.00
10× PCR buffer	5.000
dH_2O	32.50
Taq DNA polymerase	1.000
Total volume/PCR tube	**50.00**

Table 3.2 (Cont'd)

Temperature (°C)	*Duration (min)*
95 (Denaturation)	05
65 (Annealing)	10
72 (Extension)	90
04 (Stop reaction)	05–∞

3.9.2 Agarose Gel Electrophoresis (AGE)

We analyzed the samples obtained from the thermal cycler with AGE. For one AGE run, 40 mL of 1, 1.5 or 2% agarose gel was prepared. For 1% gel, 0.4 g of agarose powder (BioShop® Canada Inc.) was mixed with 40 mL of 1× Tris-acetate-EDTA (TAE) buffer (recipe in **appendix G1**) in a conical flask. The solution was swirled and heated in a microwave (550 W) for 1 min to dissolve agarose completely. 4 µL ethidium bromide was added immediately into the solution and swirled. The solution was then gently poured into a suitable tray with a comb in place. The agarose solution was left for ∼20 min to solidify at RT. After solidification, the comb was gently removed, and the gel was transferred to Hoefer's HE33 mini horizontal submarine unit filled up with 1× TAE buffer until the gel was fully submerged. The DNA samples were then mixed with a 6× loading dye (Wizbiosolutions, Seongnam, South Korea) at the right proportions to achieve a dilution down to 1×. For example, a 10 µL DNA sample is mixed with 2 µL of 6× loading dye. The DNA samples were then loaded carefully into the wells and 5 µL of 100 bp Plus DNA ladder (TransGen Biotech, Beijing, China) was also loaded as the standard. The electrophoretic was then sealed properly and connected to a constant power supply of 100 volts until the dye had run through 75–80% of the gel (30–40 min). After running, the gel was collected from the unit and was taken to be UV-scanned by Quantum-ST4 (Vilber Lourmat, France) imaging system to analyze the results. Finally, an image was captured, and the gel was safely discarded in a biological waste container.

3.9.3 DNA Purification, Quantification, and Concentration

DNA products of all *Taq* pol reactions were purified from the enzymatic reaction mixture by QIAquick PCR Purification Kit (Qiagen, MD, USA). First, reaction mixtures were transferred into 1.5 mL Eppendorf tubes such that each tube contains 100 µL sample (50+50 PCR reaction). 500 µL of buffer PB was then added onto each of the 100 µL samples and the mixtures were briefly vortexed and spun down. After that, each of the sample mixtures was applied into a corresponding QIAquick spin column placed in a 2 mL collection tube to bind the DNA. The samples were then centrifuged at 13,000 rpm for 60 sec in a bench-top centrifuge and the FT was discarded (primers are filtered out due to the 100-bps MWCO of the silica membrane). The columns were placed back into the collection tubes and 700 µL of buffer PE was added into each column to wash. The columns were then centrifuged again at 13,000 rpm for 120 sec followed by discarding the FT. After placing the columns back into the collection tubes, an additional centrifugation at 13,000 rpm was done for 120 sec to remove the residual buffer. To elute bound DNA, the columns were first placed into clean 1.5 mL Eppendorf tubes and then 35–40 µL of prewarmed dH_2O (for 2 min at 42 °C) was slowly and carefully added to the center of the filter without direct contact. The columns were left to stand for 2 min before eluting the purified DNA by centrifugation at 13,000 rpm for 60 sec into the fresh Eppendorf tubes. We did not discard the spin column unless a satisfactory amount of DNA was obtained, and so no second elution is needed (when needed, second elution from the columns was done consecutively by a single volume of water and not in parallel by replicated volumes). To quantify the DNA, the DNA samples were combined in a single tube and A_{260} measurements were taken using the BioDrop using dH_2O as a blank. The concentration of our purified DNA samples was increased by vacuum centrifugation using Maxi Dry Lyo Freeze-dyer (Heto) and BioDrop measurements were taken again.

3.9.4 In vitro Transcription (IVT)

The libraries used in this SELEX were 2′-fluoro pyrimidine containing 100-mer RNA libraries that we generated before each round of selection by *in vitro* transcribing the purified 117-mer dsDNA library of the preceding round with Apt-Get 2′F T7 Transcription Kit (Roboklon GmbH, Berlin, Germany). The modified T7 RNA polymerase binds to its recognition promoter sequence (the underlined sequence in oligo 484 shown in **Figure 3.1** and in the text in **section 3.9.1**) found in the dsDNA library sequences. Production of the 2′F-modified 100-mer RNA transcripts is then promoted, and the complementary DNA strand is used as the template of transcription. Transcription begins at the guanine residue indicated with a green glow in **Figure 3.1** and bold text in **section 3.9.1**; 2′-fluoro CTP, 2′-fluoro UTP, and nonmodified ATP and GTP are incorporated where appropriate. We carried out IVT for each round in PCR tubes and at 42.5 °C provided by a water bath. The reaction components present in a single tube are shown in **Table 3.3**. The DNA template quantity in each PCR tube dictates the time duration of the IVT reaction in that tube. Therefore, to determine the amount of DNA template (in µg) in a PCR tube at a given round, we first determined the total quantity of DNA template that will be used in IVT of that round as the total volume of DNA sample available. 10–12 µL of DNA was reserved for use in NGS of the pools later if needed. Since the total IVT template volume is known and was divided equally at known volumes between the PCR tubes (we mostly used 3–7 tubes during this SELEX), the quantity of DNA template (in µg) added into each PCR tube can be obtained using the concentration of the original DNA sample measured by the BioDrop. The DNA quantities in each tube during all rounds ranged between 0.1 and 1.0 µg and consequently the time course of the reaction was shortened when the DNA quantity per tube was high and was lengthened when the DNA quantity was low. The shortest and longest time durations were 3 and 16 hours, respectively.

Table 3.3 Components of the IVT reaction mixture and their amounts per PCR tube.

Reaction Component	***Amount***
5× Reaction buffer	5.0 µL
2′-F Py NTP mix (25 mM each)	1.5 µL
DNA template	0.1–1.0 µg
Apt-Get 2′F-T7 RNA polymerase	0.5 µL
dH_2O	Based on DNA template volume
Total volume/PCR tube	**25 µL**

3.9.5 Native- and Urea-PAGE

Unlike agarose gels that were run after each selection round, 12% PAGE gels were run occasionally throughout SELEX and to confirm the transcribed RNA pools as bands visualized on the native/urea PA gels. For the recipes of 12% PAGE gels for running RNA samples in denaturing (urea) or nondenaturing (native) conditions, refer to **appendix G2**. Also, for the equipment used and the overall preparatory procedure, refer to **unit 3.5** (SDS-PAGE). Briefly, 6 mL of 12% RNA gel solutions were prepared for a single gel and were casted quickly but carefully between the plates after the addition of TEMED. The comb was fitted between the plates to allow formation of the wells (no isopropanol or stacking gels are added here). After solidifying the gel (∼30 min), the comb was removed, and the gel was transferred to the running apparatus that was then filled up with 1× Tris-borate-EDTA (TBE) buffer (recipe on **appendix G3**). RNA samples were prepared in the same way as agarose samples (see **section 3.9.2**). Both the samples and the DNA marker (used in AGE) were loaded into the wells and for one gel, a constant current of 30 mA was applied for ∼30 min or until the dye had run through most of the gel. After running, the gel was collected, placed in a suitable container and submerged in 1× TBE buffer solution with 2–4 µL of ethidium bromide. The container was then covered fully with an aluminum foil due to the light-sensitive nature of ethidium bromide and was

placed in a slow shaker for 10 minutes at RT. Finally, the gel was collected for scanning and documentation using Quantum-ST4 system.

3.9.6 Purification and Resuspension of the 2'F-Py-modified RNA Library

We purified the transcribed 2′-fluoro-modified RNA library from the IVT reaction mixture by first degrading the template DNA molecules in the mixture. This was done by adding 0.5 µL of DNase I (GeneAll® Biotechnology, Seoul, Korea) to each PCR tube and incubating at 37.5 °C for 30 minutes in a water bath. After DNA degradation, the samples were transferred into 1.5 mL Eppendorf tubes such that each tube contains IVT sample from 3–4 PCR tubes (i.e. 75–100 µL). The volumes in the Eppendorf tubes were completed to 150 µL by MQ-water for a more efficient purification process (by precipitation). Sodium acetate (NaOAc) 2.86 M, pH 5.3 at 24 °C (recipe on **appendix G4**) was added to each tube at one-tenth of the sample volume (i.e. 15 µL added) and the tubes were then vortexed, spun down, and allowed to sit for 2 min. 100% isopropanol was added next at a 1:1 ratio with the sample in each tube (i.e. 150 µL added). The samples were vortexed, spun down, and placed in −80 °C overnight. Following overnight precipitation, the samples were taken and immediately centrifuged at 13,000 rpm for 30 minutes (at RT). After centrifugation, the supernatant was quickly poured away before it diffuses back into the solution but gently enough not to lose the isoelectric RNA pellet by its effortless mobility. 500 µL of 70% of cold ethanol (kept at 4 °C) was then added to each pellet for washing. Centrifugation was immediately done at 13,000 rpm for 15–30 minutes. The supernatant was discarded in the same manner and the pellets were left to air-dry for 3–5 hours by carefully placing the tubes in a semi-horizontal position with the mouths of the tubes pointing below the horizontal plane to optimize drying by the circular flow of air. After drying, the pellets were briefly spun down and resuspended in a 100 µL MQ-water volume to avoid excessive dilution of the pure RNA sample. The resuspension process went briefly as follows: water was added to a dried RNA pellet and the sample was properly vortexed and spun down and then placed for 2

min in a water bath at 42.5 °C followed by a brief vortexing and spinning-down again. This RNA solution was then used to resuspend another RNA pellet. 25 µL of MQ-water was occasionally used to wash the empty tubes to collect residual RNA. After all RNA is resuspended and collected in a single tube, the concentration of the sample is measured using the BioDrop spectrophotometer.

3.9.7 In vitro Selection of Aptamers

Our approach to mediating oligo-target binding was to incubate the participating molecules together in a free solution environment in a 1.5 mL Eppendorf tube. The SELEX described here is an 8-rounds process and the selection program that we used to simulate evolutionary selection pressures and competition between oligos for binding to the purified AdiC protein is described mathematically in **Table 3.4**. Before incubation of the RNA pool with the target, the volumes of all the components of the solution in which the binding reaction takes place were calculated and added accordingly into the 1.5 mL Eppendorf tube; those are the RNA pool, the purified AdiC protein, and PBS. Here, DDM was not part of PBS buffer because its absence is not anticipated to cause major detrimental changes to the target's structure since its concentration is already low and the incubation time is relatively short. **Appendix G5** shows the details of the calculations needed to execute the SELEX plan shown in **Table 3.4**.

Table 3.4 The mathematical aspect of our SELEX program. The increase in the stringency of selection is a function of 3 major variable factors that we altered by their controlled reductions after each round: 1) the RNA pool's concentration, 2) the target's concentration, and 3) the time allowed for binding between the RNA pool and the target.

Selection Round	Oligo (µM) : Target (µM)	$\frac{\text{Oligo } (\mu M)}{\text{Target } (\mu M)}$	Binding Rxn Volume (µL)	Incubation Time (min)
1	20.0 : 0.250	80.00	100	60
2	16.0 : 0.180	88.89	100	50
3	12.8 : 0.120	106.7	150	40

Table 3.4 (Cont'd)

4	10.0 : 0.090	111.1	150	30
5	8.00 : 0.070	114.3	150	30
6	6.00 : 0.045	133.3	150	20
7	4.00 : 0.028	142.9	150	20
8	2.00 : 0.013	153.8	150	15

Before preparing the binding reaction tube, the RNA pool was first refolded into its most stable structure while in PBS. For this, the working volumes of the pool and PBS as well as the volume of the MQ-water (see **appendix G5**) where added into one Eppendorf tube, briefly vortexed and spun down, and then placed in a dry bath at 95 °C for 2 minutes in a heat block and the block was left for 1 hour at RT to cool down and for refolding to take place. After 1 hour, the tube was briefly vortexed and spun down to wash down the droplets condensed on its inside walls. Next, the AdiC sample was taken out and allowed to warm up spontaneously to RT from 4 °C. The working volume of AdiC was then added into the tube with the refolded pool of oligos and the solution was gently mixed by pipetting. Incubation was done at RT for the pre-determined time duration of that particular selection round.

After the end of every incubation term, the binding mixture was collected and added onto the center of a nitrocellulose membrane placed on top of a Büchner flask with a porous sintered plate at its mouth and connected to a suction vacuum pump. Before addition of the sample onto the membrane, the center of the membrane was prewetted with 1 mL of 1×PBS solution at the area of sample addition. To wash down unbound/weakly binding oligos, the membrane-bound oligo-target complexes were washed stringently with slow addition of 10–15 mL of 1× PBS. The first 1 mL taken was used to collect residual sample from the Eppendorf tube. After washing, the membrane was collected into a sterile 15 mL Falcon tube and 0.5 mL of 6 M urea solution was added directly on the sample area to dissociate the selected oligos from the target by denaturation with the chaotropic agent. The solution is then vortexed

well, diluted to 1 M by 2.5 mL MQ-water, and then vortexed again. To maximize denaturation and release from AdiC, the tube was incubated at 50–55 °C and vortexed every 15 min.

3.9.8 Purification and Resuspension of the Selected Oligos

To purify the selected oligos from their target and the denaturing urea solution, six fresh 1.5 mL Eppendorf tubes were prepared and loaded each with 65 µL of 2.86 M NaOAc (pH 5.3 at 24 °C). 650 µL aliquots were taken from the RNA-AdiC-urea solution and added into the tubes. At tube #5, less than 400 µL would be loaded due to material that remains in the membrane and on the walls of the Falcon tube. To collect RNA material left on the membrane, 150 µL of 6 M urea was added into the membrane followed vortexing and dilution to 1 M with 750 µL MQ-water. Finally, the solution was distributed between tube #5 and 6 and any material remaining in the membrane was squeezed out as well. Next, the Eppendorf tubes were vortexed, spun down, and left to stand for 2 min before adding 650 µL of 100% isopropanol into the aliquots. The samples were vortexed and spun down again, and then incubated at −80 °C overnight. For the rest of the extraction protocol and resuspension of the purified RNA pellets, please refer back to **section 3.9.6**. Notable, features distinct to the protocol in this section include the invisibility of the RNA pellet due to its constituents being selected RNAs and not RNAs amplified by transcription. The low amount of RNA also makes it undetectable by BioDrop measurements, taking us directly to the next steps without having to take A_{260} readings.

3.9.9 Reverse Transcription (rt)

To feed the selected 2′F-modified RNA oligos into the next round of selection, they first needed to be expanded into a large library of oligos by amplification of the sequences. The oligos were first reverse-transcribed into DNA

in a thermal cycler using WizScript™ cDNA Synthesis Kit (High Capacity) (Wizbiosolutions, Seongnam, South Korea). Primer-annealing and the subsequent *rt* reaction were done separately in a two-step procedure. For one PCR tube, the volumes of the components used in these two steps are shown in **Table 3.5** and **Table 3.6**, respectively. The thermal cycler settings for the two steps are shown in these tables, respectively, too. The primer used in annealing to the RNA strands is the reverse primer called oligo 485 with the sequence: 5′-GCC TGT TGT GAG CCT CCT GTC GAA-3′. For one reaction, our 2× master mix provides 2 µL of 10× reaction buffer, 0.5 µL 20× dNTP mix, 0.25 µL RNase inhibitor, and 6.75 µL dH_2O. All the RNA sample was used and converted into cDNA in this procedure. Before running in the thermal cycler, the samples in the PCR tubes were briefly vortexed and spun down to dismiss any bubbles.

Table 3.5 Tube components for oligo 485 annealing and their amounts, and the temperatures used for the procedure (stock concentration and working volumes are shown)**.**

Component	***Volume (µL)***
Oligo 485 (200 µM)	1.00
Selected-RNA template	5.00
dH_2O	4.00
Total volume/PCR tube	**10.0**
Temperature (°C)	***Duration (min)***
65	05
04	05–∞

Table 3.6 Tube components for the *rt* reaction and their amounts, and the temperatures used for the reaction (stock concentration and working volumes are shown; Total volume/PCR tube is the accumulative volume as of RNA-oligo 485 annealing)**.**

Component	***Volume (µL)***
2× master mix	9.50
WizScript™ RTase	0.50
Total volume/PCR tube	**20.0**
Temperature (°C)	**Duration (min)**
65	60
85	05
04	05–∞

3.9.10 Polymerase Chain Reaction (PCR)

After converting our selected RNA oligos into their analogous cDNA by *rt*, the cDNA pool was exponentially amplified by PCR for the next selection round. The DNA samples in the PCR tubes from the preceding *rt* procedure were loaded with the materials needed for amplifying as shown in **Table 3.7,** where the temperatures used for the amplification reaction are also shown. We used WizPure™ PCR 2× Master (Wizbiosolutions, Seongnam, South Korea) which contains a mix of Taq DNA polymerase, $MgCl_2$, dNTPs, enhancer and stabilizer. After properly adding the amplification materials to the PCR tubes, the tubes were briefly vortexed and spun down and finally loaded into the thermal cycler. The number of PCR cycles used after each round of selection is summarized in **Table 3.8**. After thermal cycling, the product DNA amplicons were visualized by AGE and then were purified by Qiagen's PCR purification kit, concentrated and quantified (see **section 3.9.3**) Finally, they were made ready for the next round of selection after their IVT (see **section 3.9.4**). The A_{260} concentration measurements of the DNA pools of every

round were used in designing the IVT reactions; these measurements are given in **Table 4.2**.

Table 3.7 PCR tube components and amounts, and the temperatures used for cycling (stock concentration and working volumes are shown; Total volume/PCR tube is the accumulative volume as of RNA-oligo 485 annealing)**.**

Component	***Volume (µL)***
100 µM Oligo 484 (Forward Primer)	05.0
200 µM Oligo 485 (Reverse Primer)	2.50
Wizpure™ PCR 2× Master	27.5
Total volume/PCR tube	**55.0**
Temperature (°C)	***Duration***
93 (Initial Denaturation)	05 min
93 (Denaturation)	30 sec
65 (Annealing)	01 min
72 (Extension)	01 min
72 (Final Extension)	10 min
04 (Stop)	05–∞ min

Table 3.8 The number of PCR cycles used after each selection round.

SELEX ROUND	***1st***	***2nd***	***3rd***	***4th***	***5th***	***6th***	***7th***	***8th***
PCR Cycles	15	17	17	17	15	13	13	14

3.10 Bacterial Transformation and Aptamer Cloning

After 8 iterative selection rounds, the final enriched dsDNA pool (amplified and purified) was cloned into pJET1.2/blunt cloning vector (genetic map in **appendix H**) using the CloneJET™ PCR Cloning Kit (Thermo Scientific, Vilnius. Lithuania). The sticky-end cloning protocol was used in which a proprietary

thermostable DNA blunting polymerase clears away any 5′ and 3′ overhangs in the to-be-cloned purified DNA fragments. In the ligation reaction, T4 DNA ligase was used to bring together the blunted DNA fragments into the linear pJET1.2/blunt vector. The blunting and ligation reaction mixtures are shown in **Table 3.9**. Before the onset of the reactions, the mixtures were briefly vortexed and spun down and then incubated at 70 °C (heat block) and RT, respectively, both for 5 minutes.

Table 3.9 Components of the blunting and ligation reaction (stock concentration and working volumes are shown; Total volume/PCR tube is the accumulative volume as of the blunting rxn)**.**

Blunting Reaction	
Component	***Volume (µL)***
2× Reaction buffer	10.0
8th pool DNA aptamers	1.00
DNA blunting enzyme	1.00
dH_2O	6.00
Total volume/PCR tube	**18.0**
Ligation Reaction	
Component	***Volume (µL)***
pJET1.2/blunt Vector (50 ng/µL)	1.00
T4 DNA ligase	1.00
Total volume/PCR tube	**20.0**

The final mixture was used directly in the heat-shock transformation of competent BL21(DE3) cells (for competent cells preparation, see **appendix I**). Briefly, the ligation mixture was added to 120 µL of competent BL21 cells and the mixture was incubated in ice for 40 min. Then, the tube mixture was placed in 42 °C (water bath) for 30–45 sec and then back in ice for 10 min. 880 µL of LB medium was then added and the bacterial cells were cultivated at 37 °C for 1.5 hour in an incubator shaking at 180 rpm. Cultivation was continued on four agar culture plates

supplemented each with ampicillin (100 µg/mL) by spreading 250 µL of the transformed bacteria on the agar and incubation overnight at 37 °C while the plates are placed upside-down to avoid the formation of smeared colonies. Rather than by blue/white screening, cells that have taken up recombinant pJET1.2/blunt vector were selected positively on the ampicillin plates by disruption of the lethal gene in the vector through the inserted aptamer.

3.11 Plasmid Isolation, and Amplification and Sequencing of the Clones

Since each individual colony on the plate represents a proto-aptamer, 100 colonies that respectively and through the cloning plasmid, has taken up 100 candidate proto-aptamers, were hand-picked randomly with an inoculation wire loop. Each colony was inoculated into 5 mL of ampicillin-supplemented LB medium and the colonies were cultivated overnight at 37 °C in the liquid medium with shaking in an incubator. After growth, the recombinant plasmids harboring the DNA sequences of the candidate proto-aptamers were extracted and purified from the bacterial cultures using NucleoSpin® Plasmid purification technology (Macherey-Nagel, Düren, Germany). The protocol for high-copy plasmid DNA isolation from *E. coli* was used. Briefly, the cultivated bacterial cells were pelleted in a standard bench-top microcentrifuge at 11,000 × g for 30 seconds and then, as much of the supernatant as possible was discarded. 250 µL of Buffer A1 was added and the cells were properly resuspended (by vortexing or pipetting). To lyse the bacterial cells, 250 µL of Buffer A2 was added and the mixture was gently mixed inversions (6–8 times) and not by vortexing to avoid genomic DNA shearing. The mixture was then incubated at RT for 5 minutes (or until the lysate appears clear). Next, 300 µL of neutralizing Buffer A3 was added and mixed thoroughly with the lysate by inversions (6–8 times) until the blue color turns completely colorless. The lysate was then clarified by centrifugation(s) at 11,000 × g for 5–10 minutes at RT and the clear supernatant was decanted into a NucleoSpin® column placed in a 2 mL-collection tube. The plasmid DNA was then bound to the silica membrane by centrifuging the

column for 1 minute at 11,000 × g; the FT was discarded, and the column was returned back into the collection tube. The membrane and its bound plasmid DNA were washed with 600 µL of Buffer A4 (with ethanol) by centrifugation for 1 min at 11,000 × g; the FT was discarded, and the column was placed into the collection tube. The silica was dried by an additional spinning for 2 min at 11,000 × g. Finally, the column was placed into a clean 1.5 mL microcentrifuge tube, and the plasmid DNA was eluted by adding 50 µL of Buffer AE, incubating for 1 minute at RT, and centrifugation at 11,000 × g. The purified plasmid samples were finally supplied to INTERGEN Genetics and Rare Diseases Diagnosis Research and Application Center (Ankara, Turkey) for PCR amplification of the cloned aptamer inserts and their Sanger sequencing. The plasmids were supplied together with pJET1.2 forward and reverse sequencing primers: 5′-CGACTCACTATAGGGAGAGCGGC-3′ and 5′-AAGAACATCGATTTTCCATGGCAG-3′, respectively (both primers are provided in the cloning kit).

3.12 Preliminary Bioinformatical Analyses of the Aptamer Sequences

3.12.1 Refinement of the Raw Sequencing Data

After receiving the sequencing results, the raw sequencing data were converted from DNA into RNA and were then filtered and narrowed down from 100 to 38 proto-aptamer sequences. The eliminated sequences (62 sequences) were a combination of short-length sequences (< 79 nucleotides), sequences that had altered constant regions (base deletions, substitutions etc.), sequences with the constant regions appearing in the middle of those sequences in addition to flanking the 5′ and 3′ ends (i.e. duplications), low quality data (e.g. multiple nucleotide peaks), and finally, sequences with tandem repeats (e.g. > 10 consecutive uracil residues).

3.12.2 In silico Prediction of Aptamer Secondary Structures

The lowest energy (minimum free energy (MFE); most thermodynamically stable) 2D secondary structures of the sequenced RNA aptamer candidates were predicted by submitting their primary nucleobase sequences to *RNAfold* (http://rna.tbi.univie.ac.at/) and *Kinefold* (http://kinefold.curie.fr/) webservers and running their folding algorithms. For *RNAfold* (version 2.4.13) predictions (Gruber, Lorenz, Bernhart, Neuböck, & Hofacker, 2008; Lorenz et al., 2011), no constrained folding parameters were appointed to the folding algorithm. Both basic and advanced folding options were kept at default (i.e. avoid isolated base pairs; allow dangling energies on both sides of a helix in any case; and calculate the MFE structure at 37 °C). The RNA energy parameters used in MFE calculation are those described in the Turner model (Mathews et al., 2004). The quantitative prediction results for the thermodynamic ensemble were adopted. MFE structure drawings encoding base-pair probabilities were viewed in and adopted from forna; the color scheme of all structures was chosen based on the sequence. For *KineFold* predictions (Xayaphoummine, Bucher, & Isambert, 2005), each structure were generated through both of the available stochastic folding simulation types: co-transcriptional folding in which bases were allowed to add every 3 milliseconds (average T7 phage RNA polymerase transcription rate), and renaturation folding in which instantaneous cooling from 99 to 37 °C at 1 M NaCl is simulated. The simulated molecular time was kept as the default. Pseudoknot formations and non-crossing entanglements were allowed. No helices were traced or forced. Most simulations were run in immediate mode and occasionally in batch mode when the suggested molecular folding time was not fully simulated within the lapse of the immediate job. Each sequence was run 2-3 times with a different random seed in each run to confirm no variations in the stochastic folding process. In the few cases where different structures had emerged in different runs, the structure with the lower free energy was taken.

3.12.3 Multiple Sequence Alignment

Multiple sequence alignments were performed using Clustal Omega (Ω), a new alignment program provided as a web service (https://www.ebi.ac.uk/Tools/msa/clustalo/) by the European Bioinformatics Institute EMBL-EBI (Madeira et al., 2019). Fasta format of the sequences was used as input. The alignment parameters were not changed, and the output was obtained in ClustalW format with character counts. Other results adopted from this alignment include a phylogenetic tree and the percent identities.

CHAPTER 4

RESULTS AND DISCUSSION

The use of purified full-length MPs as targets in SELEX technology is often seen as an approach to aptamer selection that is to be avoided mainly due to the many challenges associated with acquiring these proteins and also for the many advantages offered by the other approaches. The body of publications available in the current literature is undeniably limited in describing the approach to selection where purified full-length MPs are used. The vast majority of MP-specific aptamers were selected using either soluble domains of MPs or whole live cells (Cibiel et al., 2011; Dua et al., 2011; Mercier et al., 2017). This is mainly because soluble MP domains permit, by being hydrophilic in nature, bypassing the need for the clumsy and troublesome purifications of full-length MPs; especially if the protein had not been expressed and purified before. For live cells on the other hand, these permit bypassing purification procedures altogether and even the need for awareness about the identity of targets. While these are highly desirable and advantageous properties, MPs purified and subsequently targeted for selection as full-length proteins provide advantages that may not be achieved by SELEX trials where ectodomains or live cells/organisms are used as baits. These include the simplicity and ease of selection (less procedural technicalities and experimental complexities and uncertainties), fast aptamer evolution due to minimal background binding (i.e. target is pure and not complex as a cell surface therefore making SELEX highly specific), and the specificity of the selected aptamers for the native structure and conformation of the binding site if the protein is purified right.

Herein, we have recombinantly expressed the MP L-arginine/agmatine antiporter also known as AdiC and responsible for extreme bacterial AR. We then purified and reconstituted AdiC in detergent micelles of the mild nonionic detergent n-dodecyl-β-D-maltopyranoside (DDM). To our knowledge, this dissertation serves

as the first account to report the use of SELEX for the generation of nucleic acid aptamers against AdiC. Moreover, surveying the literature allowed locating only one article in which aptamers were selected against an MP (glycine receptor, GlyRαI) overexpressed and then purified in DDM (Shalaly et al., 2015). Testing these RNA aptamers demonstrated positive modulation of membrane potential and Cl^- currents. Another article reported the modulatory effect of RNA aptamers selected for β2AR purified also in DDM but later reconstituted in maltose-neopentyl glycol (MNG) micelles prior to selection to obtain a more stable and conformationally active MP receptor (Kahsai et al., 2016).

In the light of the background information above, our work here consequently supports the notion of the feasibility of selection of nucleic acid aptamers against MPs purified and solubilized by reconstitution in detergent micelles.

4.1 Purified AdiC: A Prerequisite for Aptamer Selection

4.1.1 Establishing a Pure, Stable and Abundant SELEX Target

Based on the multitudinous and diverse research articles available in the literature in which selection experiments are done, a SELEX target that is pure and present uniformly in a stable structural state and defined amounts at each round of selection leads quickly to the emergence of aptamers that can bind well and even specifically to that target. To that end, and by implementing a previously described MP purification protocol used to purify overexpressed AdiC prior to its biochemical, biophysical and structural characterizations (Ilgü et al., 2014, 2016), we have isolated AdiC from provided cell membrane preparations using DDM in the extraction and solubilization processes, and the highly selective cobalt-charged TALON affinity resin to purify and collect AdiC at high purity by column-based affinity chromatography. Previous data on elution profiles from a SEC analysis carried out for recombinantly expressed AdiC purified by the method adopted here showed that AdiC was stable in DDM and also in 8 other detergents with no

indication for a major aggregation problem (Ilgü et al., 2014). In addition, modeling of a protein-detergent-lipid (PDL) ternary complex by bioinformatical tools and based on gained experimental data on the detergent binding capacities and phospholipid content of the purified AdiC PDL complexes have shown that the data indeed can be reasonably accommodated around the generated model. **Figure 4.1** shows the model built for a ternary complex of AdiC protein solubilized and purified in 1.5% and 0.04% DDM, respectively.

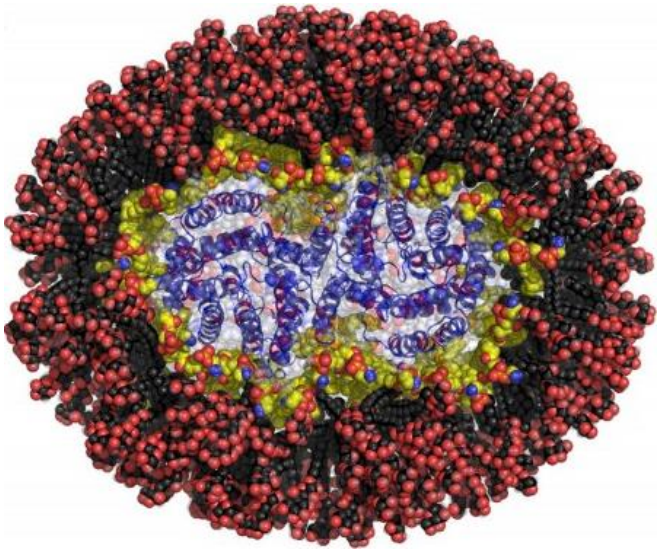


Figure 4.1. A computer-generated model of a PDL complex of AdiC. The model is based on the dimeric structure of AdiC (PDB: 3LRB) and the experimental data that describes the number of detergent and lipid molecules bound in the complex purified in DDM. 257 and 52 molecules of detergent (DDM) and lipid (POPE) are shown, respectively. The AdiC dimer is shown as a blue cartoon and transparent spheres. Phospholipid and detergent molecules, and oxygen and nitrogen atoms are shown as yellow and black spheres, and red and blue spheres, respectively. The figure is reprinted from Biophysical Journal, 106, İlgü et al., Variation of the Detergent-Binding Capacity and Phospholipid Content of Membrane Proteins When Purified in Different Detergents, 1660–1670, Copyright (2014), with permission from Elsevier.

To initially confirm the presence of AdiC protein in our eluate samples, we carried out a rough assessment of the concentration at A_{280} using a BioDrop machine blanked with EB. The concentration reading was 4.3 mg/mL of AdiC protein per one of the duplicates. After combining the duplicate samples into one tube, the

concentration of AdiC in the sample was raised by a centrifuge-driven ultrafiltration procedure that we performed using a MWCO concentrator: a filter device with a specialized membrane and contained in a suitable 15 mL conical tube. The ultrafiltration process concentrates the protein of interest efficiently in the sample solution loaded into the concentrator by reducing the solution's total volume without reducing the amount of protein material in that solution. Centrifugal spinning of the concentrator device with the loaded sample facilitates rapid passage of low-MW entities (e.g. water molecules) through the membrane, and allows high-MW entities (50 KDa and above)—here, AdiC—to be retained as they are unable to cross the membrane due to hindrance by their large molecular size. Virtually, all AdiC molecules are in the end recovered in a smaller solution volume thereby raising the protein's concentration in the sample.

4.1.2 Ensuring Longevity of the Selection Target for the 8-rounds SELEX Procedure

Being in EB which contains potentially "harmful" imidazole molecules, the AdiC protein sample, after concentration, was buffer-exchanged into 1× PBS buffer supplemented with the detergent DDM at 0.04% (w/v), a much more amiable buffer. PBS is commonly used in mimicking physiological environments due to its isotonicity and pH that resemble that of the human body. Although in nature, AdiC is subjected to wide ranges of pH values when considered as a functional component of the bacterial ADAR system, here, we relocated our purified AdiC in pH ~7.4 which is provided by the buffer PBS, a pH value that is consistent with those used in obtaining the currently available AdiC 3D crystal structures (see **Table 2.9**) and accordingly, one that makes PBS the buffer chosen to be the medium for the oligo-target binding reaction that will take place in our SELEX experiment. The additive DDM in this buffer was included at a concentration of 0.04% (w/v) in order to maintain its steady presence with AdiC PDL complexes as they transfer from EB to PBS. Maintaining DDM consistency (0.04%) minimizes any potential risks of

undesirable and unexpected effects on the equilibrium of the AdiC structure with its bound detergent and phospholipid molecules; in other words, stability of the purified PDL complexes is ensured. Instead of dialyzing the protein sample for the desalting and exchange of buffers, a more rapid and time-saving, centrifuge-driven procedure was performed wherein a spin desalting column with proprietary high-performance size-exclusion chromatography resin was used. During spinning with the loaded protein sample, the column highly retains all undesired small molecules (in our case, EB constituents such as imidazole and Tris-HCl) and allows larger macromolecules above 7 KDa (i.e. AdiC) to pass easily through the resin to emerge down and be recovered in the collection tube in the desired buffer used in equilibration of the column prior to sample loading; i.e. 1× PBS buffer with 0.04% DDM. After concentrating our purified AdiC sample and relocating it into DDM-supplemented PBS, the protein concentration was reassessed by the BioDrop device, this time blanked with 1× PBS + 0.04% DDM. The displayed concentration was found at a higher value—4.65 mg/mL. It is important to bring to attention the argument that this value not only shows an increase from the apparent concentration of the initial AdiC elution samples (4.3 mg/mL) but may also be considered a somewhat stronger proof for the actual presence of protein material in the solution compared to the proof from measuring the elution samples. This is because the components of the buffer-exchanged AdiC sample, namely, PBS and DDM, also present at their relatively low concentrations, represent a group of molecules that are less effective in the interference with the A_{280} measurements when compared to the highly UV light-absorbing components of EB present at high concentrations (400 mM imidazole and 10% (v/v) glycerol) in the eluates. Indeed, it was shown that irrespective of whether blanking is carried out prior to A_{280} measurements, buffers that absorb highly at this wavelength (or any other test wavelength) affect protein concentration estimations negatively by limiting the amount of light that is available for analyte measurement (Thermo Scientific, 2019). Moreover, blanking with such a high-absorbing buffer may not even compensate fully for absorbance of the buffer itself (Beringer, Ash, &

Page, 2011). As far DDM is concerned for the buffer-exchanged sample, this detergent absorbs light maximally at λ = 230 nm.

4.1.3 Direct Confirmation of the Presence and Purity of the SELEX Target

For the most confident declaration of success in purifying recombinant AdiC, the MP was run through in an SDS-PAGE and its presence was confirmed visually on the PA gel. **Figure 4.2** shows an image of a SDS-PAGE with single and sharp AdiC bands indicating the high purity of the protein after metal AC. Refer to **appendix J1** for an additional SDS-PA gel showing the different samples collected throughout the AdiC purification procedure and electrophoresed along with the elution samples shown in **Figure 4.2**.

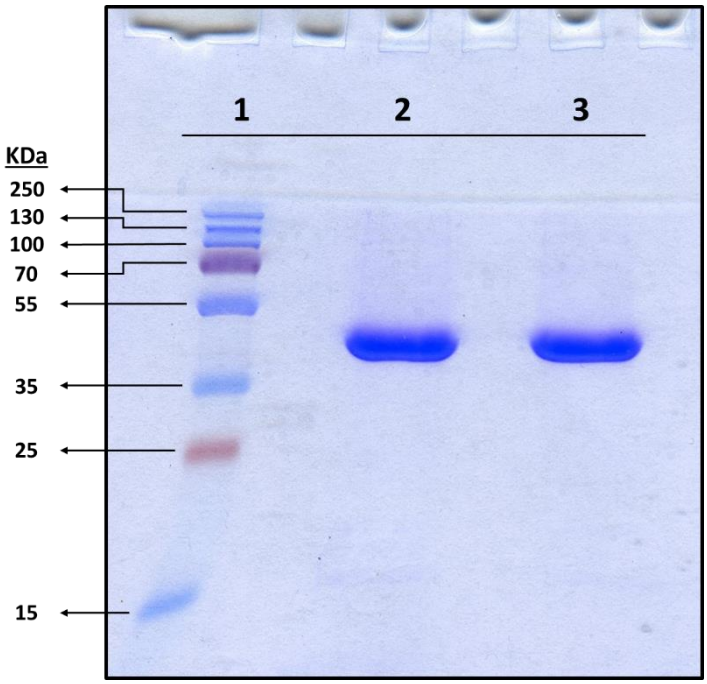


Figure 4.2. 12% SDS-PA gel from an SDS-PAGE for the duplicate AdiC eluate samples (prior to the MWCO concentrator procedure)**.** The gel is stained by standard staining protocol (Coomassie Brilliant Blue R-250 staining). Lane 2 and 3 show sharp single bands of pure recombinant AdiC monomers and lane 1 shows the reference protein ladder (see **appendix J2** for the commercially distributed image of the migration pattern of this ladder).

Since we have purified AdiC as a recombinant protein with an additional C-terminal tail of amino acids that forms a protease cleavage site and a deca-His tag, the calculated MW of monomeric AdiC elevates to ~50.9 KDa (Ilgü et al., 2014). This theoretical MW value is corroborated experimentally by the qualitative results on the gel in **Figure 4.2** as the AdiC bands (lanes 2 and 3), when compared to the protein ladder, are seen closer to the 55 KDa checkpoint (and thus, closer to the 50.9 KDa) than to the 35 KDa checkpoint. Moreover, in conjunction with these observations, it becomes interesting that unlike the SDS-PAGE for the purified recombinant AdiC published in the supporting material of the work by İlgü et al. in 2014, the same AdiC purified in this work did not run significantly faster than the calculated MW (neither have their c-terminal tail cleaved). In that study, AdiC was determined to run until the 37 KDa mark on the gel. In general, such discrepancies in electrophoretic migration may intuitively be attributed to differences in the adopted SDS-PAGE methodology/materials. To exemplify, variations can exist between the components that make up the SDS-PAGE sample (salts, organic solvents, detergents etc.) or exist in the sample heating temperature and time whereby in the latter case, proteins denatured inadequately may, for instance, run faster than expected. This exact observation can also be seen with nonreduced proteins due to disulfide bridges resulting in a less linearized and a more compact protein structure. The biochemical laboratory technique, SDS-PAGE, is a treasured universal and mainstream method that has been used countlessly and for a long time in the high-resolution separation of proteins and the determination of their molecular sizes (Shirai et al., 2008). However, SDS-PAGE determines the molecular sizes of proteins relatively and not absolutely. Acquisition of information about MWs from an SDS-PAGE analysis depends primarily on its working principle where a certain amount of SDS molecules bind to a fully denatured protein in the sample in proportion to its molecular size to form the densely negatively charged SDS-protein complexes that run in theory, only based on the MW of the protein. For many of the investigated proteins, this ratio is very similar, and for each protein, it resembles the ratio of SDS binding to the protein marker used in MW calibration. This in turn

allows determination of relativistic MWs. Therefore, it is sensible that any factor (e.g. experimental) that can distort the presumed typical SDS-protein binding profile and skews the "golden" charge-to-mass ratio is likely to cause anomalous running (faster or slower) of the proteins in the electric field on the SDS-PAGE. In addition to the above-mentioned factors that can affect the expected gel mobility profile, intrinsic aspects of the protein under investigation can also strongly affect running; these include risks of truncations or proteolytic attacks, the presence (or absence) of PTMs or other artificial protein modifications (e.g. fluorescent labels), and the protein's amino acid composition, properties of the amino acids (charged, hydrophobic, isoelectric points etc.) and also how they're adjacently ordered in the primary structure. All these intrinsic factors are strong determinants of the protein's structure, shape and folding and of its interaction with other protein or nonprotein species in the SDS-PAGE sample's environment and in turn this determines the degree of SDS loading to the protein which ultimately dictates the observed migration pattern. SDS-protein interactions are difficult to model and predict. Perhaps more intriguing is that the seemingly aberrant "gel-shifting phenomenon" that is addressed here is in fact rather common for MPs running in an SDS-PAGE experiment (Rath, Glibowicka, Nadeau, Chen, & Deber, 2009). A compelling number of helical MPs were reported in the literature to have experienced gel shifts that portray polypeptides in certain cases almost up to 50% larger or smaller than expected MW. Particularly, for AdiC, aberrant running was reported in different publications where it had run faster than the expected known mass of ~46 KDa (Fang et al., 2007; S. Gong et al., 2003). MPs were shown to load in certain cases at least twofold higher amounts of SDS than the detergent saturation value of globular proteins (1.5–2 g SDS/g protein). In their study, Rath et al. have revealed valuable insights by modeling the strong correlational relationship between gel shifts and detergent loading capacity and helicity for a convenient system of a library of helical hairpin MPs derived from TM segments of the human cystic fibrosis transmembrane conductance regulator (CFTR). Indeed, they have concluded that altered detergent binding explains anomalous SDS-PAGE migration of MPs. Their results also

suggested the intimate relationship between altered SDS loading and the conformation as well as the hydropathy of the hairpin MP, given the expectation that SDS aggregates to hydrophobic residues. Another protein-derived effect that alters SDS binding to MPs is the tendency of the protein to form higher oligomeric states that can even be induced under SDS-PAGE conditions only. Resistance for detergent binding in these states may be seen as a lower-than-expected band on the gel. Also, considering that the study by Rath et al. showed a detergent loading on the hairpins of up to 10 g SDS/ g protein, it then becomes feasible that MPs running in SDS-PAGE are in many other cases incompletely denatured, hence, the anomalous running. Additionally, it is worthwhile to consider sample heating as a process that may not fully denature specifically MPs; or conversely, a process that may drive them to their aggregation.

4.1.4 Accurate Assessment of the Abundance of the SELEX Target

SELEX is a profoundly dynamic process the outcome of which is shifted by the slightest of pressures and forces acting through a myriad of parameters and exerting their effects on the process of selection. For instance, high-affinity ligands that are present even as single copies in a pool of billions of ligands were found to be capable of influencing the population dynamics (Spill et al., 2016). Due to the fundamental feature of SELEX where binding equilibrium exists between aptamers and targets, the many uncertainties of SELEX can be tuned for an optimized selection process and improved outcomes using computational and mathematical modeling (Ilgu & Nilsen-Hamilton, 2016; Jinpeng Wang, Rudzinski, Gong, Soh, & Atzberger, 2012). As such, these studies can take place at various levels including selection stringency and the impact of counter selections, pool and target concentrations, and also non-specific background binding, at each round. The effect of target concentration has been the focus of many modeling studies. Almost all of them have pointed towards the existence of an optimal target concentration and often provided the equations needed for its calculation (Komarova & Kuznetsov, 2019). High target

concentrations are known to disfavor the evolution of aptamers with have high affinities to the presented target and on the other side of the spectrum, target scarcity hinders selection as equilibria between nonspecific ligands and targets take over and become augmented and more apparent (Spill et al., 2016). In the former case, families of ligands are part of a very deep library while in the latter case, ligand families are lost as the target's concentration falls below their K_d value for the binding site (Morris et al., 1998).

Although the aim of our work here does not necessitate finding that optimal target concentration, it was still necessary—in addition to affirming the purity and stability of our SELEX target (AdiC)—to determine accurately the target's concentration in our sample to ensure that the protein sample as a whole is abundant for distribution upon the multi-round selection scheme we have adopted; that is, abundant enough for sampling each round for incubation with the oligonucleotide pool at the mathematically defined amounts featuring controlled target reductions (see **Table 3.4**). Although the concentration of AdiC was measured by evaluation in the BioDrop, the inherent crudeness of that analysis dictated resort to finding a more exact value. As pointed out earlier, errors in optical density measurements of a protein can arise due to the presence of highly-absorbing materials accompanying the protein. However, we also expect the measurement obtained even after the buffer exchange for AdiC (4.65 mg/mL) to be a misrepresentation of its true concentration as the real molar extinction coefficient of AdiC was not accounted for when taking these measurements. For this reason, we have set to find the concentration of AdiC through a classical colorimetric Bradford assay. Nevertheless, the previous BioDrop readings remain a fair indicator of the presence of protein material as well as of the relative increase in its concentration after treatment with the MWCO concentrator.

The Bradford assay is a popular and sensitive protein quantification assay described originally by Dr. Marion Bradford in 1976 in which Coomassie Brilliant Blue G-250 dye binds to a protein in a sample allowing determination of the protein's concentration based on the respective amount of dye that binds and the subsequent color change (brown to blue) and absorbance shift to 595 nm (Johnson, 2012).

Accurately quantifying a protein by the Bradford assay depends on the availability of specific dye-binding amino acid residues in the primary sequence of the protein and at sufficient frequencies. Strongly binding amino acid residues are arginine, lysine, histidine, and the aromatic ones: tyrosine, tryptophan and phenylalanine. Summed up, these amino acids make up 88 out of 467 amino acids of the structure of our recombinant AdiC protein, or ~20%. However, because the standard protein BSA is known to exhibit a strong dye response and has a greater number of dye-binding amino acids (69 more residues), there still exists a chance for deviation of the AdiC concentration determined by the assay from its true value and particularly by underestimation. Although for SELEX, deviation by underestimation means that selection will now take a different path with different dynamics and towards a different outcome that may be better or worse, as pointed out in the previous paragraph, our goal here is not to study the effect of variation of SELEX parameters on the outcome of SELEX but is to obtain an accurate determination of the AdiC sample's concentration. A precisely determined concentration of the protein material (AdiC) allows assessing whether this sample is abundant enough and qualifies to impose the full course of the evolutionary program used (see **Table 3.4**). Otherwise, a low amount of target may force alterations to the program such as reducing the working sample taken each round or reducing the total number of rounds. Also, an accurately determined concentration ensures that the results of our selection trial are attributed to the features defining the adopted selection pressures program and that the program is imposed exactly as mathematically devised.

As seen in **appendix F1**, a series of six dilutions have been prepared for the protein standard BSA using DDM-containing PBS along with a 7th tube containing only that buffer to account for the blank measurement. Three dilutions (2×, 5× and 10×) were also prepared from the purified AdiC original sample to serve as test samples for which the concentration is to be determined in the assay. Preparation and use of diluted AdiC in the Bradford assay ensures the use of test samples with unknown concentrations that still reside within the linear range of the BSA-based assay and therefore provide a safe and reliable extrapolation tool towards the exact

concentration of AdiC in the original undiluted AdiC sample. Moreover, these dilutions weaken the effect of any substances that are co-present with the protein and thus can interfere with the assay. For this reason, DDM is not expected to distort the results of our assay. A_{595} measurements for all of the duplicate samples in the microplate wells (dilutions of BSA and AdiC) are shown in **Table 4.1**. The table also lists absorbance (ABS) averages calculated from the individual ABS values for each pair of duplicates and it shows those averages after the elimination of background absorbances by arithmetically subtracting the average ABS of the blank duplicates. From a biophysical point of view, those "zeroed" averages represent the quantity of light absorbed only by the Coomassie dye that had bound to certain amino acid residues in BSA and AdiC. In other words, those averages do not implicitly carry any ABS contributions from possible light-absorbing substances present in the well samples other than the protein material. Hence, each value is a good reflection of the concentration of the protein located in a corresponding well as ABS from all other nonprotein matter is arithmetically omitted. Since the concentration of BSA in each of its dilutions is known, an absorbance-versus-concentration graph for the prepared standard BSA dilution samples was constructed and is shown in **Figure 4.3**. Based on the datapoints plotted in this standard curve, a best-fit line was generated.

Table 4.1 Spectrophotometric A_{595} readings of all Bradford Assay microplate well samples and the calculated ABS averages pre- and post-zeroing.

Absorbances / Well Source	ABS Duplicates 1	ABS Duplicates 2	ABS Averaged	Zeroed ABS Averages
Standard Tube 1	1.523	1.502	1.5125	1.0120
Standard Tube 2	1.260	1.291	1.2755	0.7750
Standard Tube 3	1.069	1.124	1.0965	0.5960
Standard Tube 4	0.883	0.877	0.8800	0.3795
Standard Tube 5	0.631	0.628	0.6295	0.1290
Standard Tube 6	0.578	0.583	0.5805	0.0800
Tube 7 (Blank)	0.507	0.494	0.5005	0.0000
AdiC (2× dilution)	0.623	0.601	0.6120	0.1115

Table 4.1 (Cont'd)

AdiC (5× dilution)	0.566	0.544	0.555	0.0545
AdiC (10× dilution)	0.528	0.517	0.5225	0.0220

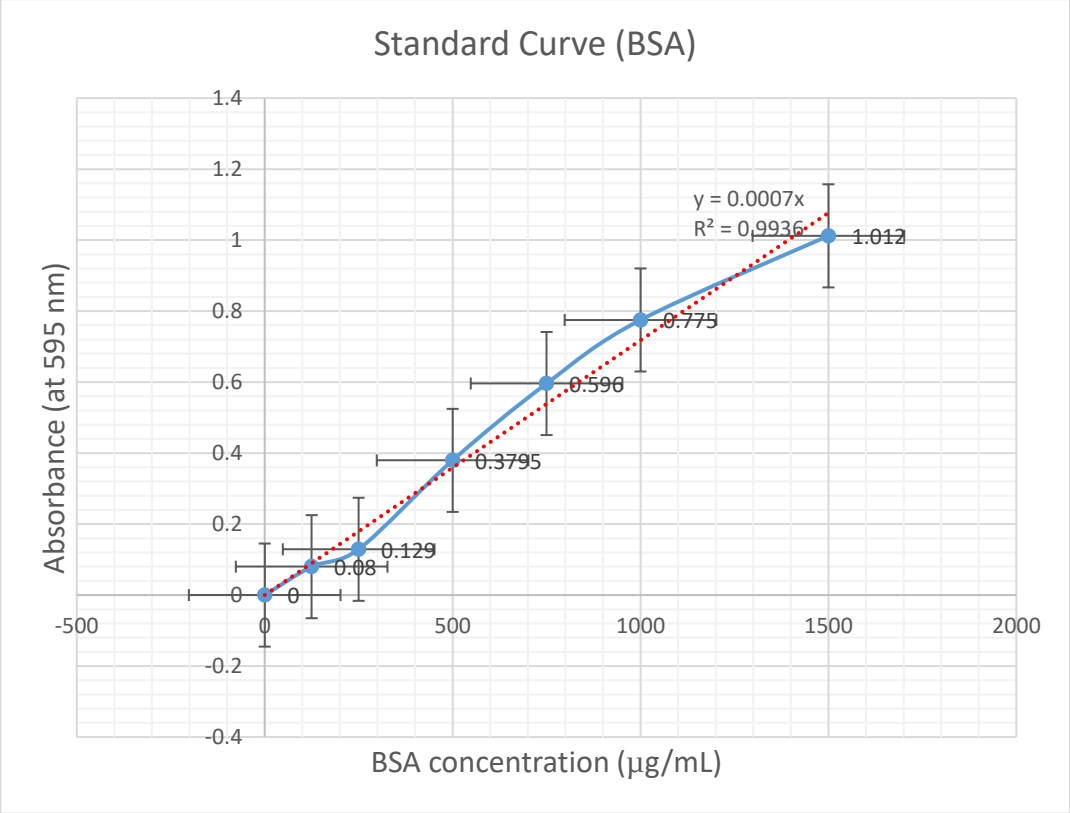


Figure 4.3. Standard curve (BSA). The continuous blue curve is charted by plotting the zeroed absorbance averages of the six standard BSA samples against the corresponding concentration of each of these samples prepared by serial dilutions. The discontinuous red line is an automatically generated best-fit line based on the plotted datapoints. The coefficient of determination (R^2) of the datapoints is ~0.99 and the equation of the best-fit line is y = 0.0007x; thus, its slope is 0.0007.

Using the slope of the best-fit line in **Figure 4.3** and the zeroed ABS averages of the three AdiC dilution samples shown in **Table 4.1**, the concentration of AdiC in each of these dilutions is calculated to be as follows:

- [AdiC $_{2\times \text{ dilution}}$] = = 159.286 µg/mL = 0.16 mg/mL
  - Multiplication by 2 for 2× dilution factor: [AdiC] = 0.32 mg/mL

- [AdiC $_{5\times \text{ dilution}}$] = = 77.857 µg/mL = 0.08 mg/mL
    - Multiplication by 5 for 5× dilution factor: [AdiC] = 0.40 mg/mL

- [AdiC $_{10\times \text{ dilution}}$] = = 31.429 µg/mL = 0.03 mg/mL
    - Multiplication by 10 for 10× dilution factor: [AdiC] = 0.30 mg/mL

Multiplying each of the dilution samples' concentrations by the respective dilution factor reveals the concentration of AdiC protein in the original undiluted source sample: 0.32 mg/mL by 2× dilution, 0.4 mg/mL by 5× dilution, and 0.3 mg/mL by 10× dilution. Averaging out the three numbers gives an average concentration for purified, concentrated and buffer-exchanged AdiC that is **∼0.34 mg/mL**; a value that is in wild disagreement with that obtained from the BioDrop (4.65 mg/mL) which is nearly 14 times larger! Since BSA has dye-binding amino acid residues that are more than those of AdiC by ∼1.8 times, we deem the Bradford assay to be the more accurate evaluation of the concentration of our protein. Furthermore, at this concentration (0.34 mg/mL) and at this volume (250 µL, after concentrator treatment), it is safe to say that our sample is abundant enough to carry out all the AdiC-consuming analyses and procedures. For SELEX specifically, the total amount of protein material consumed to execute the selection program with its specific and unique stringency level (see **Table 3.4**) is only ∼80 µL out of 250 µL, as calculated and shown in **appendix G5**. To expand on a different aspect of this discussion, it can be inferred from the above calculations that the absorbances and therefrom, the concentrations of the dilution samples of AdiC lie within the range of linearity of the Bradford assay (1–1400 µg/mL) and precisely at its lower end. Therefore, assuming that an underestimation of these concentrations has taken place, the actual concentrations of the dilution samples are likely to remain resident also within the linear range even with a higher confidence. Furthermore, being close to the lower end also means that *a priori*, any underestimations that may further arise (in addition to those from a dissimilar amount of light-absorbing amino acids) in the calculated concentrations of these dilution samples as a result of being too close to the upper bound of the linear range are confidently eliminated. If that were to occur,

the risk for the actual concentrations being outside the upper bound becomes higher and so in addition to the underestimation coming from the dye's variable binding efficiency assumed significant between the standard and the unknown protein, further underestimation arises due to saturation of the assay and the breaking of Beer-Lambert's law. Henceforth, unless the samples are diluted even further and new readings are taken, more unreliable ABS measurements and calculations will unfold. Ultimately, this translates into SELEX dynamics that do not reflect the selection program and in in which competition between the ligands for the presented target may be weakened and consequently, rising of high-affinity aptamers is impeded.

4.1.5 Purified AdiC, Post-selection

As discussed earlier, it is important for the selection process that the target remains stable and exhibits no structural changes as each round is successively brought about. The *in vitro* SELEX procedure that we carried out has spanned a time period of ∼2.5 months, a period that is enough for the purified AdiC protein to potentially degrade in the sample. Between each sampling of AdiC for each selection round, we stored our protein at 4 °C. In this section, we show by SDS-PAGE analysis that AdiC had impressively remained intact in the source protein sample throughout this time period, indicating the great stability of this bacterial protein. The post-SELEX SDS-PAGE is shown in **Figure 4.4** and in fact, this gel was run ∼1 month after the end of selection. The amount of AdiC protein run in the gels before SELEX and after SELEX (**Figure 4.2** and **4.4**) is the same—10 µL. Although very thin ambiguous bands with high MWs have appeared at the AdiC lane in the newer gel, these insignificant bands may simply be reflecting the recurrent running anomalies of MPs in SDS-PAGE such as their oligomerization during preparation for the SDS-PAGE experiment.

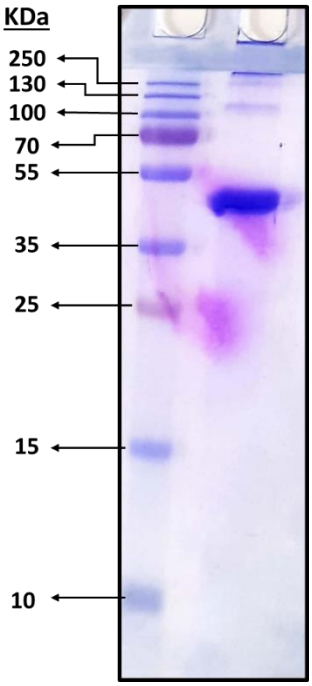


Figure 4.4. Post-SELEX AdiC SDS-PAGE. The SDS-PA gel above was run ∼1 month after the end of selection and it shows that the SELEX target, AdiC (right lane), has resisted degradation and remained intact well through the selection procedure which itself had lasted ∼2.5 months. Despite the high-MW thin ambiguous bands present in the AdiC lane, when compared to the protein ladder (left lane), it is safe to say that close to 95% of the SDS-PAGE sample (10 µL AdiC) appears as a band that is at the same position as in **Figure 4.2** and as a band that is also at the same level of high purity and fullness in the gel.

4.2 Selection of 2′F-Py-modified RNAs Allured by Purified AdiC Protein

İlgü et al. demonstrated among a number of other important findings, the reliability of size-exclusion chromatography (SEC) as a method for the determination of molecular masses of purified MP ternary complexes (Ilgü et al., 2014). They showed that the purified ternary complexes of the MP AdiC formed by solubilization using the mild alkylmaltoside detergent, DDM, had an apparent mass of 257 KDa by SEC. Although this value may or may not be seen very deviant from the theoretical value they calculated (273 KDa), when viewed among the 21 other experimental combinations of MPs (AdiC, LacY, and UT) and commonly used

detergents (alkylmaltosides and Cymals), it can be seen that 16 KDa is the largest deviation among these combinations and that SEC is still a reliable method to closely determine the masses of ternary complexes. In this work, we used their theoretically calculated MW of AdiC ternary complexes and the concentration of our purified AdiC that we determined by the Bradford assay (0.34 mg/mL) to calculate the molarity of our AdiC sample as 1.25 µM; the calculations are given in **appendix G5**. The appendix also gives the molarity calculations for the RNA pool at every round based on the MW of the 100-nucleotide RNA chain and its concentration in the pool of that particular round determined by the BioDrop spectrophotometer. These determinations are important to set up the incubation trial at every round for the oligo-target binding reaction as the mathematically defined SELEX program (**Table 3.4**) characterized by increased stringency of selection at every round is followed (Levine & Nilsen-Hamilton, 2007; Shubham et al., 2018). The ligands that were chosen for the selection procedure were 2′-F-modified RNAs due to their superior resistance to degradation by nucleases compared to RNAs that have hydroxyl groups attached to the carbon-2 position of ribose in the nucleotides.

Our approach to selection overall involved high molar ratios of oligo-to-protein at each round. In the 1st round, the ratio was as high as 80:1; that is, for every nanomole of protein, 80 nanomoles of RNA are co-present. From there, we have increased the selection pressure on the RNA oligos round by round by decreasing the amount of AdiC available for binding in the solution and also by decreasing the time of incubation. On average, the concentration of the target (AdiC) was reduced each round by 33.7%. Additionally, on average, the incubation time was reduced by 6.4 min each round. By the final 8th round of selection, the molar ratio has climbed up to almost 154:1 which is almost double that of the first round. Another important factor that contributed to the shifts in the equilibrium of the selection pressure each round is the reduction in the concentration of the RNA pool. On average, the pool was reduced each round by 27.2%. Although this global reduction in the population of the binding oligos serves to open a better chance for the ligands with stronger interactions to find their target, it can however also reduce the pressure of selection

for the “top-notch” aptamers with the highest affinities and reduce the rate of reduction in pool complexity. Nevertheless, it is strongly arguable that the very high RNA-to-AdiC molar ratio that we began selection with and have propagated through each round, may work effectively to counter or even overtake the effect elicited by a hampered stringency of selection.

Prior to the two-step *rt*-PCR that was done at the end of every round for the selected RNA pools and prior to the confirmation of amplification of a selected pool by AGE and the A_{260} measurements, it was necessary to establish that RNAs were in fact successfully selected at that round. For this, *rt*-PCR trials were done prior to *rt* and amplifying the whole selected RNA sample. In these trials, only a single PCR tube was prepared rather than many tubes among which all the presumably selected RNA sample would be distributed. Subsequent trial AGE gels were then run for a sample from the trial *rt*-PCR tube to confirm that selection had occurred successfully at that round by observing the amplified cDNA band on the gel. Evidence from these gels was used as basis to proceed with *rt* and PCR for the whole, now-confirmed selected RNA pool. This approach has at least three important advantages: **1)** and the most obvious, ensuring functionality of the kits used (cDNA synthesis and PCR kits) without compromising the whole sample of selected oligos, **2)** providing a chance to optimize the *rt*-PCR protocol such as that provided to minimize the number of PCR cycles as a higher cycles number increases the risk of introducing *Taq*-amplification bias that can lead to the loss of valuable ligands as selection progresses, and **3)** allowing optimization of kit economy as the functional kits will not be used with a chance that RNA ligands may have failed to select at that round. Thus, *rt*-PCR trials help clear away such uncertainties and attain these advantages. This trial approach was also used before the 1st selection round during both extension and the subsequent IVT in order to initially observe the expected outcomes of these reactions (extension of the initial ss-DNA pool oligo 487D, and IVT of this extended pool) instead of feeding the whole ssDNA library and resulting dsDNA library into extension and IVT, respectively. Nevertheless, trial IVTs specifically were not done for each of the SELEX rounds; transcription of the whole selected DNA pools was proceeded-with

directly at every round. To visualize the results from the pre-selection IVT and from the IVTs done throughout the whole SELEX procedure, native-/urea-PAGE gels were run for the yielded RNA pools. **Figure 4.5** shows an agarose gel (**4.5A**) and a native PAGE (**4.5B**) displaying the initial random dsDNA library synthesized by extension of oligo 487D and the RNA library subsequently synthesized, respectively (both run before the 1st selection round). For the agarose gel, the 117-basepairs DNA band shown is a concentrated sample (15 µL) hence its thickness. To get a better resolution on the position of this DNA sample on the gel, we run another gel with a dilute sample and as expected, a much thinner band was observed just above the 100-bps position (gel image not available).

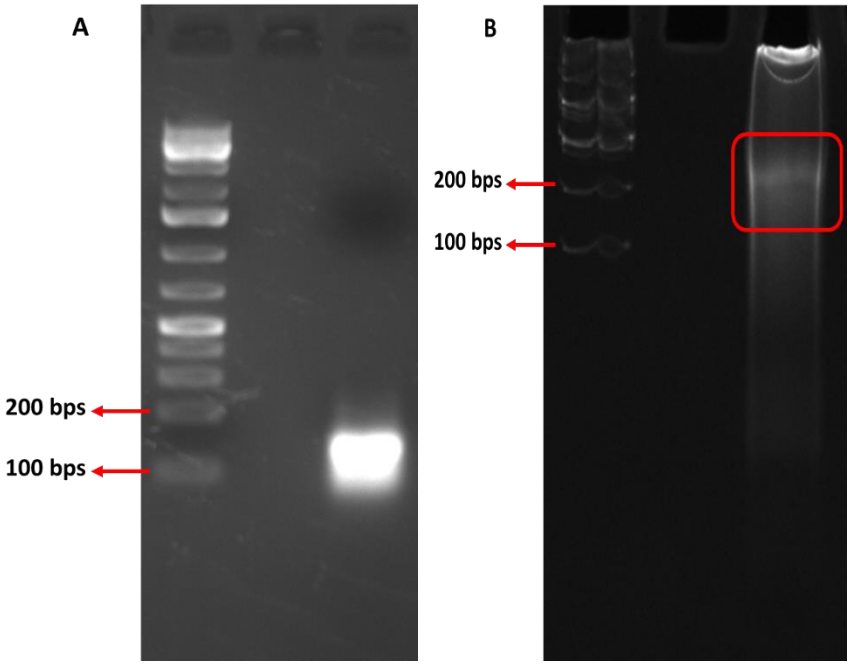


Figure 4.5. Pre-selection 1% agarose and 12% native PA gels for the initial random libraries of dsDNA and RNA, respectively. A) AGE for a 15 µL sample from the extension reaction on a 1% agarose gel reveals as expected, a dsDNA band near the 117-bps position. **B)** native-PAGE for a 1.5 µL sample (purified) from the initial RNA library on a 12% PAGE confirms the RNA as a band shown in the round-cornered red rectangle however, the DNA ladder on the left lane does not seem to be able to predict the size of the RNA strands (known to be 100 bps) in this particular case of native-PAGE. The manufacturer of the DNA ladder has recommended against running the ladder in PAGE which likely explains the aberrant

running seen above. This discrepancy is probably due to factors such as the ladder's double-stranded nature and the nondenatured structure of the RNA sample (native gel). Find the commercially distributed image of the run of the DNA ladder in **appendix K**.

Agarose gels are summarized in **Figure 4.6** and each represents the pool of selected ligands (reverse-transcribed and *Taq*-amplified) for the respective SELEX round. The cDNAs on these gels are post-trial samples obtained from one randomly picked PCR tube from the batch of sample tubes. For the trial agarose gels, refer to **appendix L**. The number of PCR cycles used to amplify most of the post-trial DNA pools at each SELEX round (**Table 3.8**) was lower than that used to amplify the trial DNA pools at those rounds. As mentioned earlier, this serves as an optimization that reduces the risks for amplification bias. We based the choice of the number of cycles used at a specific round (**Table 3.8**) on the intensity of the band from the its trial *rt*-PCR. All PCR cycle number minimizations were intuitively determined such that the bands are minimally visible in the post-trial agarose gels (variations in the intensity of the trial bands reflect the variation in the PCR cycles for the post-trial DNA amplification products). Additionally, to reduce the average number of PCR cycles used during SELEX even further, as of round 3, we increased the binding reaction's total volume from 100 µL to 150 µL (**Table 3.4**) so that more ligands are captured by the bait thus reducing the number of PCR cycles needed to amplify them. For the rest of the IVT-synthesized RNA pools used during selection, bands were visualized on PAGE gels as shown in **Figure 4.7**. However, unlike AGE gels that were run for all the SELEX pools, RNA gels were run only for pools of round 3, 6 and 7 in addition to that used for round 1 and shown in **Figure 4.5** (**B**). BioDrop concentration readings (µg/mL) for the RNA and DNA (after vacuum centrifugation) pools of all SELEX round are given in **Table 4.2**. Measurements were taken after purification of the RNA and DNA pools (see **sections 3.9.3** and **3.9.6**) and their resuspension in distilled H_2O which was used to blank the spectrophotometer. For the random initial DNA pool extended from oligo 487D, its concentration was measured as 65 µg/mL after extension.

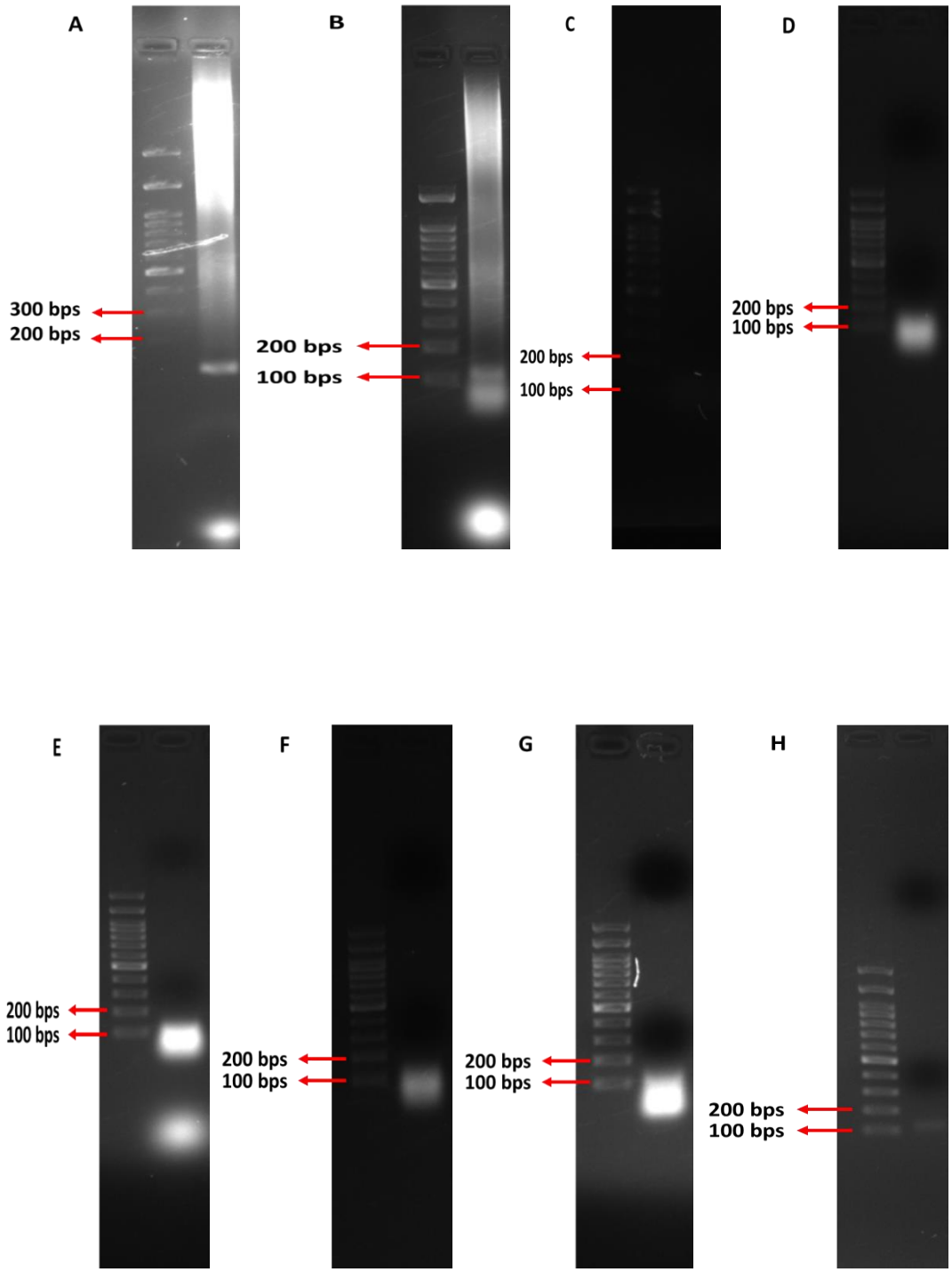


Figure 4.6. Post-trial agarose gels showing 117-bps cDNA bands run after *rt*-PCR for the selected ligand pools of every SELEX round. A) Round 1; 1.5% gel. **B)** Round 2; 1.5% gel. **C)** Round 3; 1% gel. **D)** Round 4; 1% gel. **E)** Round 5; 1% gel. **F)** Round 6; 1% gel. **G)** Round 7; 1% gel. **H)** Round 8; 1% gel. The smears seen on the sample lane of gels

A and **B** are random hexamers present in the master mix of a different cDNA synthesis kit used; the kit was replaced by round 3. Some of the gels shown have recognizable bands that are lower than 100 bps. These represent the primers and primer in from the *rt*-PCR mixture; they were reduced in later rounds by reducing primers concentration in the two-step *rt*-PCR procedure. Despite the questionable presence of a cDNA band in gel **C** due to the very low visibility, the cDNA band was indeed confirmed. After all, a non-existent DNA pool in the 3rd SELEX round is not grounds for propagating selection towards the consecutive rounds.

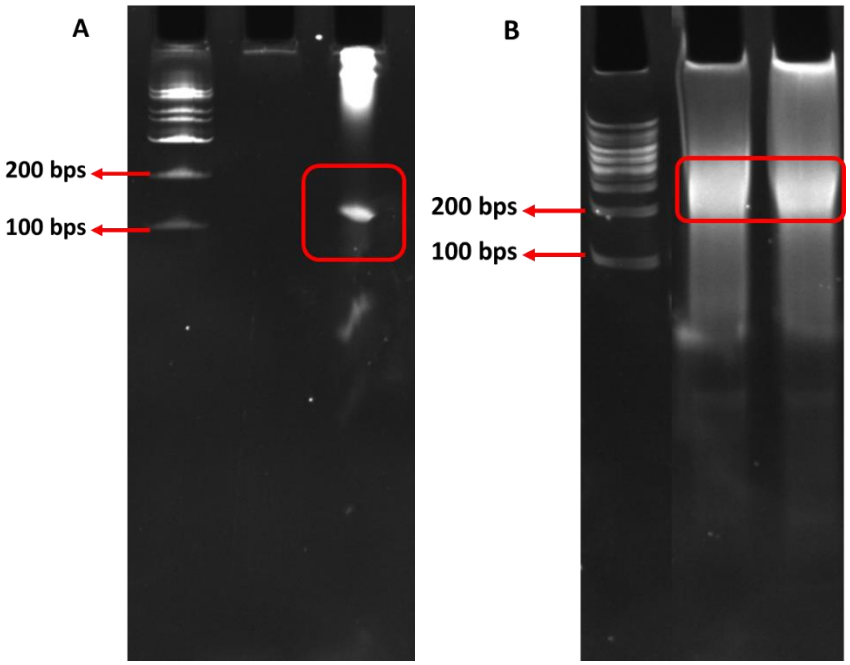


Figure 4.7. PAGE analyses for the RNA pools for rounds 3, 6 and 7 of SELEX. A) Denaturing urea-PAGE run for a sample taken from the RNA pool (unpurified) of round 3. The denaturant, urea, enforcing its effect on the ds-DNA ladder and the RNA sample seem in this case to have left us with a gel from which the size of our RNA sample is approximately predictable (100 bps). **B)** In this gel, non-denaturing native-PAGE was run for the RNA pools (purified) of round 6 and 7 (from left to right). Again, a nondenaturing environment resulted in an anomalous gel pattern from which the sizes of the RNA oligos cannot be predicted however, their presence can be visualized and confirmed. Although, the DNA ladder, round 6 pool, and round 7 pool were all run at the same time and in this very same gel, the image in gel **B** was prepared by generating two separate images from the photographed gel: one with the ladder lane alone and one with the two RNA pools lanes. The two images were then joined into one (gel **B**) while maintaining the relative positions

of the RNA bands to the ladder. This was done to exclude out additional samples that have been run in the lanes between the DNA ladder and the RNA pools and therefore to provide a simpler, cleaner and relevant image.

Table 4.2 Concentrations of the RNA (used in selection) and amplified DNA (resulting from selection) pools at every SELEX round measured at A_{260} by the BioDrop spectrophotometer.

SELEX Round	*RNA Pool (µg/mL)*	*DNA Pool (µg/mL)*
Round 1	2644.0	70.00
Round 2	2322.0	237.0
Round 3	2485.0	20.00
Round 4	1900.0	25.00
Round 5	1400.0	128.0
Round 6	1531.0	101.0
Round 7	920.00	70.00
Round 8	994.00	126.0

It was discussed in **unit 4.1** that in order to efficiently evolve aptamers by protein-SELEX to bind strongly and specifically to a purified MP, this target must be obtained at high purity, in sufficient amounts and also in stable and monodisperse structures. Perhaps, it isn't surprising that, these parameters are similarly required for the crystallization of an MP of interest (Ilgü et al., 2014, 2016). However, one parameter that can uniquely affect SELEX and its progression is that purified MPs are typically part of PDL complexes. Given the fact that these MPs are purified with a solubilizing lipid-detergent belt that surrounds their hydrophobic lipid-accessible surface, aptamers may bind to that area of these ternary complexes and not the actual protein surface. The MP ternary complex that we purified in this work and which is formed by AdiC homodimer complexed to DDM and phospholipid molecules is shown in **Figure 4.1**. In our SELEX procedure, we assume that aptamers are not likely to select against the lipid-detergent belt of our purified AdiC and we base the validity of our assumption on two arguments: first, aptamers generally have affinities

that are weaker for small molecules like DDM (K_d in the micromolar range) compared to those for macromolecules like proteins (K_d in the nanomolar range) and second, unlike for the hydrophilic and charged surfaces of the cytoplasmic and periplasmic domains of AdiC, the nonionic and only-hydrophilic heads of DDM in the belt encircling AdiC are unlikely to have the same strong molecular interactions due to their aforementioned electrical stance. Moreover, in a more practical sense, it is likely that a significant amount of any oligos that could bind to the lipid-detergent belt will dissociate from the complexes after capturing on the nitrocellulose filter due to the mechanical force inflicted on these PDL complexes by stringent washing with the hydrophilic PBS solution. On the other hand, for the oligos that may have bound to free DDM molecules, these simply cross the membrane as they are unable to attach to it like proteins. From a SELEX point of view, their very low abundance serves to make them a weakly favorable selection target in the aptamer evolution process.

4.3 Preliminary Bioinformatical Analyses of the Aptamer Sequences

4.3.1 The full Collection of Aptamer Candidates and the Tackling Strategy

After 8 selection rounds, the enriched pool was cloned, and 100 clones were chosen randomly and sequenced. The refined list of sequenced clones (aptamer candidates) can be found in **Table 4.3**. The table contains 38 proto-aptamers obtained after cleanup of the raw data from abnormal sequences (see Material and Methods) that have probably resulted due to errors in the mechanism of cloning and amplification as well as that of the Sanger sequencing process. By inspecting the refined list, we found that 3 different sequences are present in more than a single copy; that is, each of three different clone sequences was isolated more than once during random colony picking. These 3 clones are represented by AdiC1 and AdiC88 that are identical, AdiC30, AdiC57 and AdiC60 that are themselves identical, and finally, AdiC87 and AdiC90 that are identical as well. Usually, the isolation of clones multiple times is an indicator that the SELEX protocol is approaching its final round

(Morris et al., 1998). That said, we obtained a total of 34 unique proto-aptamer sequences and the three sequences that made multiple appearances are taken as the most enriched among all. Together, these three are likely the dominant species of the 8th round enriched aptamer pool and they collectively make up 7% of the analyzed clones.

Table 4.3 Refined list of the sequences of AdiC aptamer candidates (38 sequences) selected and cloned in this study. The core region of each proto-aptamer enriched during SELEX is shaded with a light orange color. The non-shaded 5′ and 3′ sequences are constant regions used in PCR-amplification of these proto-aptamers during SELEX.

Name	*Sequence (5′–3′)*
AdiC1	GGGAGACAAGAAUAAACGCUCAAGUGUCUGUUCAUGAUUGCAUUUUUAUAUUAUAUUUUCACCUUGGUGUUUUUAAUUCGACAGGAGGCUCACAACAGGC
AdiC3	GGGAGACAAGAAUAAACGCUCAAUGGGCAUUUGUUCCUUUUUGUAAUUCAUUUCGCGCGCUUGGAAGAUCUGUUAUUUCGACAGGAGGCUCACAACAGGC
AdiC5	GGGAGACAAGAAUAAACGCUCAAUCUUUGUUGCACAUGUUUUAACUAUUGGUCUAACCAUAAUUCUUAUUUAUAUUUUCGACAGGAGGCUCACAACAGGC
AdiC8	GGGAGACAAGAAUAAACGCUCAAUUUUUCCUGUCAUUAUUUCUGUAUACCUAGCAUUUUUUUAUUAUUUAGAUUCUUCGACAGGAGGCUCACAACAGGC
AdiC9	GGGAGACAAGAAUAAACGCUCAAUAUAACUCGCUUUUCUUUCUUAUUUCUUUUUUCACCCAUGCUGUUAUUGGUCUUCGACAGGAGGCUCACAACAGGC
AdiC10	GGGAGACAAGAAUAAACGCUCAAUUUUUUUCAGUUUAUCUCUUUGUCAUUUAUAAUUUUCUAUUAUUCAUACUAUUUUUCGACAGGAGGCUCACAACAGGC
AdiC12	GGGAGACAAGAAUAAACGCUCAAUUCAUUUGUCCUUUAUUAUUUUAAACCUAAAUUUUACUUAUUAAUCUCAGGUCUUCGACAGGAGGCUCACAACAGGC
AdiC18	GGGAGACAAGAAUAAACGCUCAAGUUUAGCUGGCGCAUUUAUUAUACUCCGACUGUAGAUACUGAUAUCGCGUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC19	GGGAGACAAGAAUAAACGCUCAACCUGUGUAUUCUCUUUAUACUCUCUUUAUUAAUUUUUGGUAUUUAUUUUUGAUUUCGACAGGAGGCUCACAACAGGC
AdiC25	GGGAGACAAGAAUAAACGCUCAAUAAUGUUUUCCUCUUUUUCUUGUUUUAUACCUUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC26	GGGAGACAAGAAUAAACGCUCAAUAUCUAUUUUUUAGUUAUUAUAUCAUAUUUUUUAGAGUUCUAAUUUAAUUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC30	GGGAGACAAGAAUAAACGCUCAACGAUUGUGAUGCCUCUUUUUUAUUUUAUUUGUGUUGUUUAUAUUGGGAUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC33	GGGAGACAAGAAUAAACGCUCAACCUGUUAUUCAAUUCAUACAAAUAAUUUCUUAUUAUUUUUGCUGUUUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC36	GGGAGACAAGAAUAAACGCUCAAGCGCCUUGCAUCUUUACGCCGUCCUAUUUUCUUAUAUUUUACGAUUCUUUUUUUCGACAGGAGGCUCACAACAGGC
AdiC41	GGGAGACAAGAAUAAACGCUCAAGUGCCGCUGUGAGUGUGUACUGCAAACUUUUUUAUUAUUAUUUGACUAUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC43	GGGAGACAAGAAUAAACGCUCAAUCCUACAUCACAUUAAUUACUAUUUAUGAUUUUUUUAUGUUUUUACCUCAUUAAUUCGACAGGAGGCUCACAACAGGC

Table 4.3 (Cont'd)

AdiC46	GGGAGACAAGAAUAAACGCUCAACAUAUCUAAUUUUAUCUCUCUUAUUUAC AUUUAUUUAAAGCGAAUUUAUUUUUUCGACAGGAGGCUCACAACAGGC
AdiC53	GGGAGACAAGAAUAAACGCUCAAGUACGUGCUACUGUUUAGACCUCCGGUU UGUCGGCAAUGUGCUGGCGUAAAAUUUCGACAGGAGGCUCACAACAGGC
AdiC56	GGGAGACAAGAAUAAACGCUCAAUGCUCUCUAUAUUAAUAUUUAUCAUUUA UUAUUUAAUCUAUAUUAUUUUGUGUUUCGACAGGAGGCUCACAACAGGC
AdiC57	GGGAGACAAGAAUAAACGCUCAACGAUUGUGAUGCCUCUUUUUUAUUUUAU UUGUGUUGUUUAUAUUGGGAUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC58	GGGAGACAAGAAUAAACGCUCAAUAUUUCAACCCUUUCUUAUUUCCUAUUA CGCUUUCUUUUAAGCUUUUUGGCAUUUCGACAGGAGGCUCACAACAGGC
AdiC59	GGGAGACAAGAAUAAACGCUCAAUUUUCUCUUUAUUUAAUAUUUAUACAAU CAUUUUCUUAAACUUUCCACAACAUUUCGACAGGAGGCUCACAACAGGC
AdiC60	GGGAGACAAGAAUAAACGCUCAACGAUUGUGAUGCCUCUUUUUUAUUUUAU UUGUGUUGUUUAUAUUGGGAUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC69	GGGAGACAAGAAUAAACGCUCAAUUGUGAUAUUCUUUCUGAUUCGUAUUCC ACCGUUUUGAAUUUUUGAUGAUUUUUUCGACAGGAGGCUCACAACAGGC
AdiC71	GGGAGACAAGAAUAAACGCUCAACUCCGUUAUUUACUAUUUAUAAUUAUAC AUUUUGUUUACUGGAUUUUAUCUUAUUCGACAGGAGGCUCACAACAGGC
AdiC72	GGGAGACAAGAAUAAACGCUCAACUCUAACUGUGAUUUCACUUUCUCUUUA UCAUUCUUGAUUUUCUCUUUGGAGCUUCGACAGGAGGCUCACAACAGGC
AdiC73	GGGAGACAAGAAUAAACGCUCAAUUUCCUCAAUUAAUUUUAUACUGAUUAG CUGCUGCUUAUUUAAUUUUUAUACAUUUCGACAGGAGGCUCACAACAGGC
AdiC75	GGGAGACAAGAAUAAACGCUCAACCCAUUCAUUUCUUUCUUCUUUUCUUUA UCUUAUUUCGACAGGAGGCUCACAACAGGC
AdiC79	GGGAGACAAGAAUAAACGCUCAAUUUAAUCUUUUCAUGUUACCACAGAGUC UCUCAUUUAUUGCCAUACCUUUUAUUUCGACAGGAGGCUCACAACAGGC
AdiC80	GGGAGACAAGAAUAAACGCUCAACCGUUAUUUCACAUUUCUUUAUCUUUUA UCCUUUCCUUUUUUUUUAUUUUUUCGACAGGAGGCUCACAACAGGC
AdiC83	GGGAGACAAGAAUAAACGCUCAAUUCUUUUACCUUUAUUUAUUAAUUUACU AUUUACAUUCUUUGCAGUGUUUAUUUCGACAGGAGGCUCACAACAGGC
AdiC85	GGGAGACAAGAAUAAACGCUCAACCCAUAUUUUUACCAUACAAAUUCAUUU UACUUCUUUGAUCAUAAUUUUUUAUUUCGACAGGAGGCUCACAACAGGC
AdiC87	GGGAGACAAGAAUAAACGCUCAAUGAGCCCCUAUUUUUCUACCAACAACUU GGUAUUUUAUUUUAUAUAUUUUCAAUUCGACAGGAGGCUCACAACAGGC
AdiC88	GGGAGACAAGAAUAAACGCUCAAGUGUCUGUUCAUGAUUGCAUUUUUAUAU UAUAUUUUCACCUUGGUGUUUUUAAUUCGACAGGAGGCUCACAACAGGC
AdiC90	GGGAGACAAGAAUAAACGCUCAAUGAGCCCCUAUUUUUCUACCAACAACUU GGUAUUUUAUUUUAUAUAUUUUCAAUUCGACAGGAGGCUCACAACAGGC

Table 4.3 (Cont'd)

AdiC92	GGGAGACAAGAAUAAACGCUCAAUUUUGCUUAAAUACUUUUUCUUACCUUCAUUAUACUUUCGUGUGCCAGCUAGUUUCGACAGGAGGCUCACAACAGGC
AdiC95	GGGAGACAAGAAUAAACGCUCAAUCCUCCAAAAUCUUUGUACAGUCUAUUUACUUAUUUAUUAUUUAUUAAAUUUUAUUCGACAGGAGGCUCACAACAGGC
AdiC99	GGGAGACAAGAAUAAACGCUCAAGCCGAGUACUUCUGUCGGUUUACAUUAUUAAUUUUUCGCGCAUGCACUUUUUUUCGACAGGAGGCUCACAACAGGC

For simplicity of discussion, the three proto-aptamers each of which had manifested as multiple clones will be designated as noorA (manifested in AdiC1 and AdiC88), noorB (manifested in AdiC30, AdiC57 and AdiC60) and noorC (manifested in AdiC87 and AdiC90). Due to their dominant frequency among the 100 analyzed clones and probably in the final SELEX round pool as well, these three aptamer candidates were chosen for further examination to unravel their structural features that had led them to form effective aptamer-target binding interactions and to their dominance among the thousands of other sequences co-present in the final pool. Commonly, investigations done to determine the binding motif of an aptamer follow experimental methods such as radioactive labeling, enzymatic footprinting, partial hydrolysis, or truncated DNA *in vitro* transcription (Bing, Yang, Mei, Cao, & Shangguan, 2010). However, these techniques are often complicated, expensive and/or time-consuming. Fortunately, highly efficient and yielding analytical methods such as massively parallel microarray-based analyses were developed and used for the purpose of motif discovery. Here, we follow a fully *in-silico*-based approach along with presenting biologically-relevant deductive arguments to infer enrichment of potential AdiC (purified)-interactive binding motifs in our final selected pool of 2′FY-modified RNA aptamers. The approach followed here draws in from comparative sequence analytics and computing theoretical 2D MFE secondary structures by energy minimization programs. Approaches involving computational structure prediction were undertaken previously by other groups for their advantages including bypassing complicated and time-consuming experimental methods and they were successful in deducing/confirming aptamer binding motifs

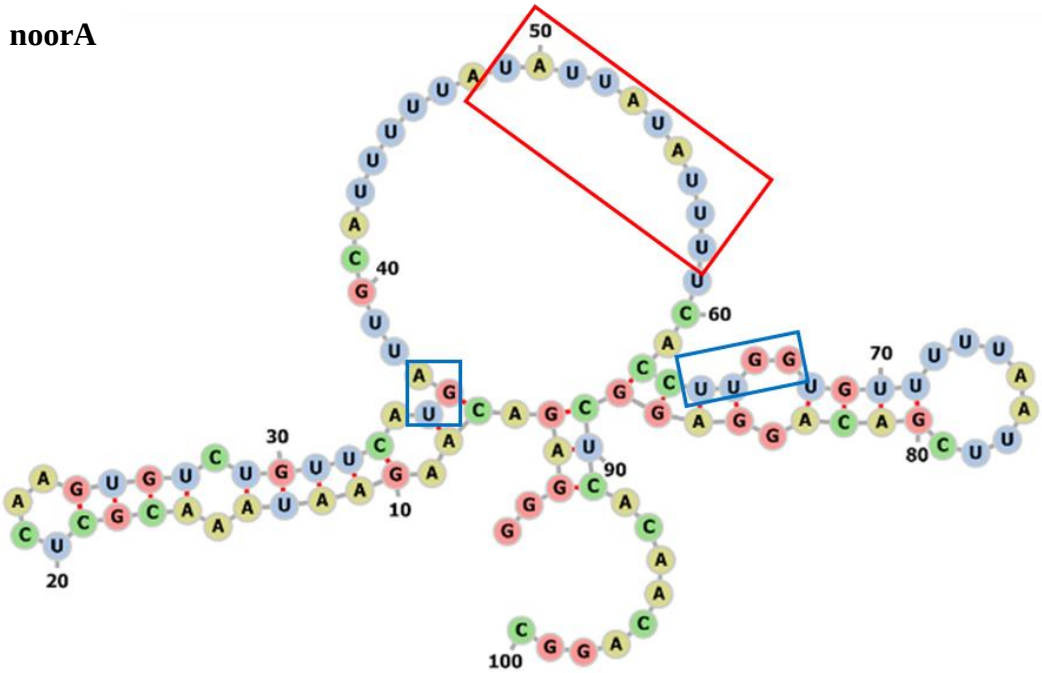
(Bing et al., 2010; Yang Cui et al., 2004; Dubey, Baker, Romeo, & Babitzke, 2005; Jang, Lee, Yeo, Jeong, & Kim, 2008; Mathews et al., 2010). However, generation of robust data from comparative approaches requires deep conservation of structure and a wealth of different sequences (Andrews & Moss, 2019).

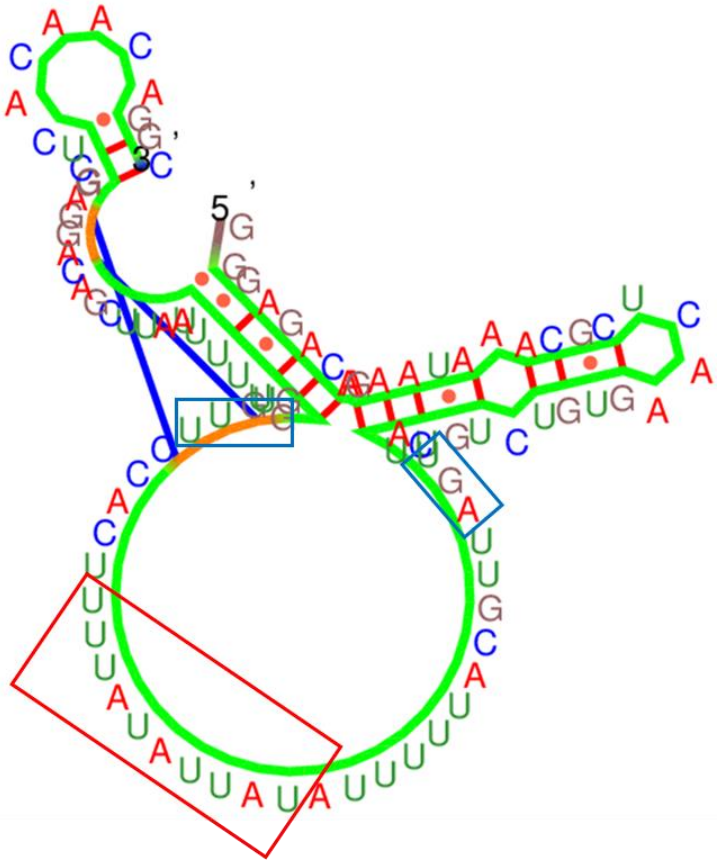
4.3.2 SELEX-evolved Conserved Regions in the Three Most Abundant Proto-aptamers

In our analysis of proto-aptamers noorA, B and C, we computationally modeled the global 2D secondary structures of their RNA sequences using the webserver-based *in-silico* tools, *RNAfold* and *Kinefold*. *RNAfold* (Gruber et al., 2008; Lorenz et al., 2011) ranks among the top three cited RNA secondary structure programs; the other two being *Mfold* and *RNAstructure* (Andrews & Moss, 2019; Reuter & Mathews, 2010; Zuker, 2003). The performance of these programs is fairly similar. Generated *RNAfold* models are known to follow a thermodynamic scoring scheme parameterized by fitting to the experimental data (Rodriguez & Cortes-Mancera, 2013). The scoring scheme's folding algorithm for *RNAfold* (typically, for all scoring schemes, a DP algorithm but implemented in different ways) employs the technique called Cocke-Younger-Kasami (CYK) algorithm to calculate the highest-probability structure. *Kinefold* models on the other hand follow a probabilistic scoring scheme with parameterization based on training by maximum likelihood methods. The folding algorithm used by *Kinefold* predicts secondary structures using the γ-centroid estimator developed to use posterior decoding methods on the base-pairing probability matrix. *Kinefold*'s algorithm kinetically simulates stochastic folding paths of nucleic acids on second-to-minute molecular time scales and these folding paths are simulated at the level of helix formation and dissociation based on seminal experimental results (Andrews & Moss, 2019; Xayaphoummine et al., 2005). Moreover, *Kinefold* is capable of efficiently predicting pseudoknots and topologically entangled helices (i.e. knots) based on simple geometrical and topological constraints. **Figure 4.8** shows the predicted optimal (MFE) 2D

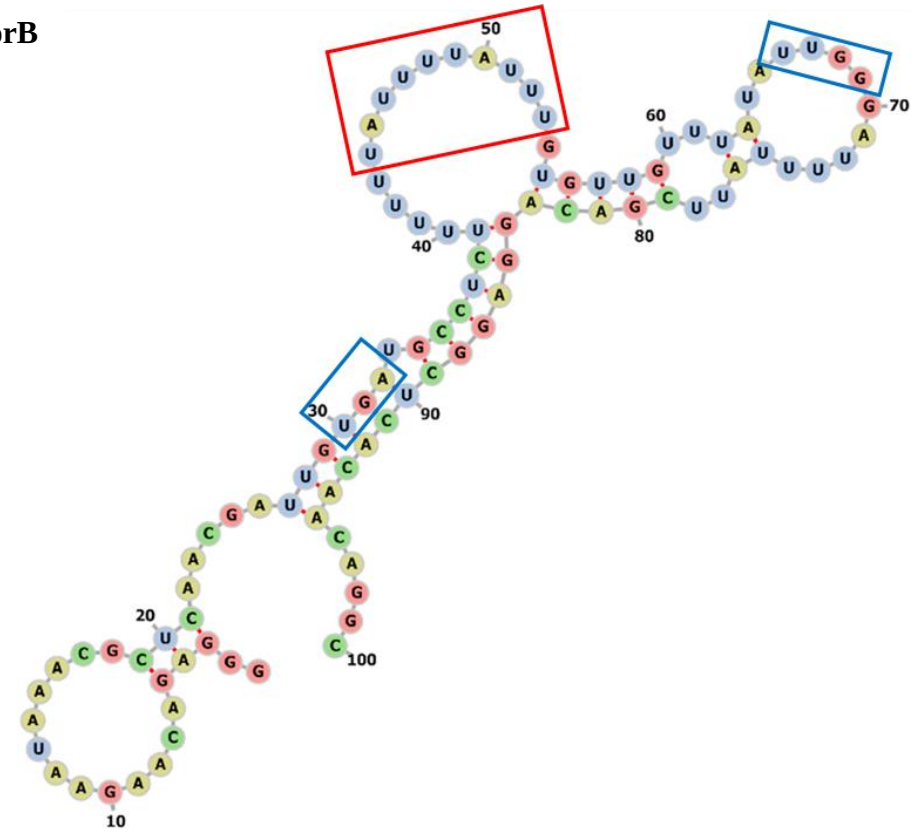
secondary structures of noorA, noorB and noorC generated once using *RNAfold* and a second time by *Kinefold*. For the *Kinefold* models, only structures obtained by renaturation folding simulation are given as they are identical to those obtained by the co-transcriptional folding simulation.

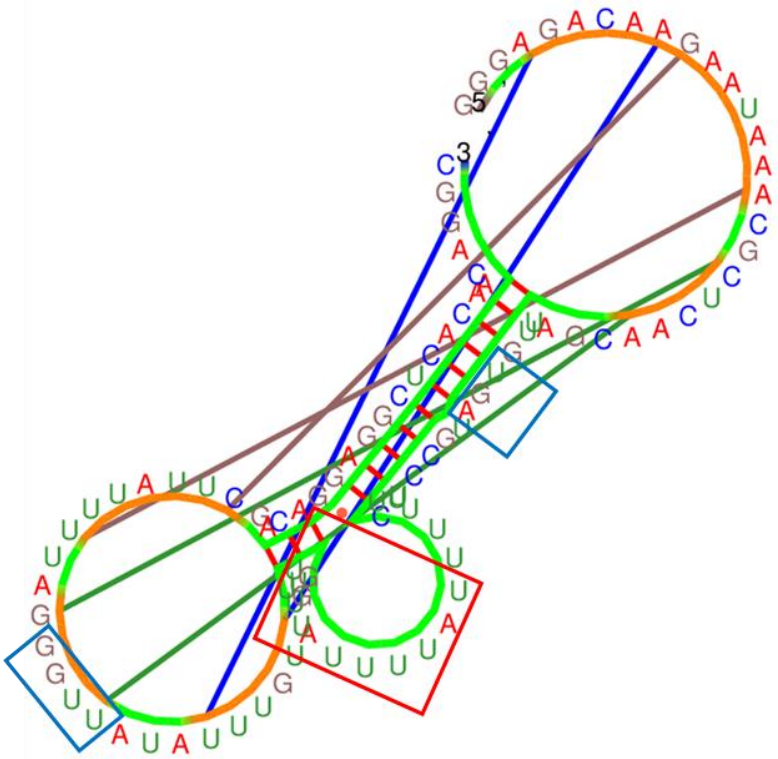
noorA



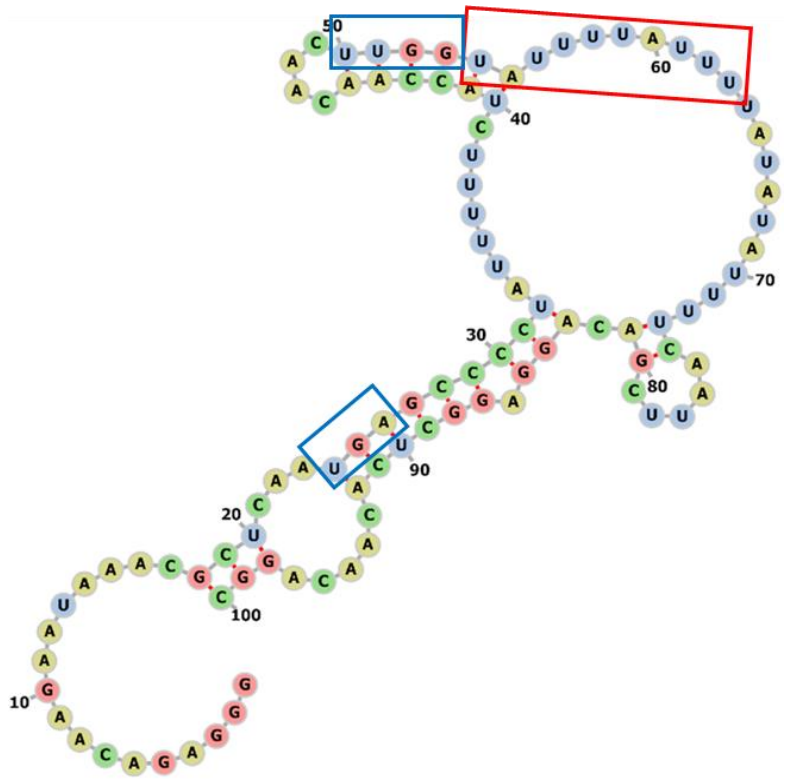


noorB





noorC



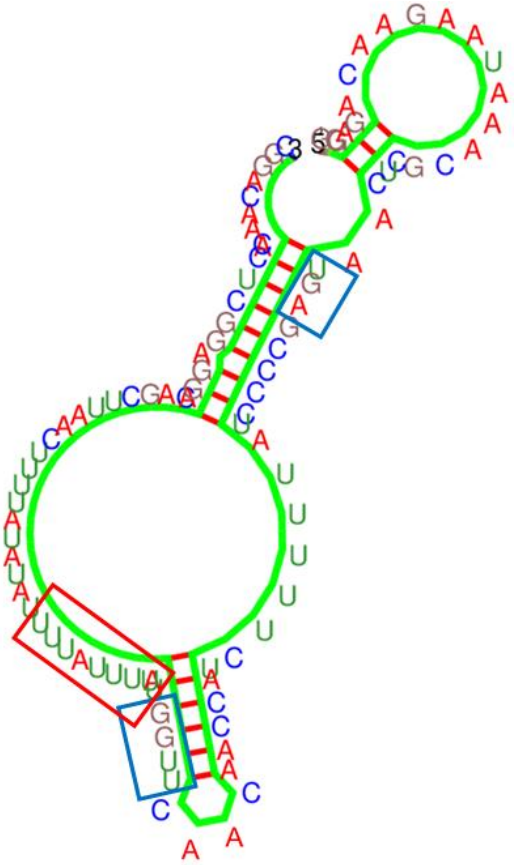


Figure 4.8. 2D MFE secondary structures of noorA, noorB and noorC predicted by *RNAfold* and *Kinefold*. The top and bottom structures at each page were generated by *RNAfold* and *Kinefold*, respectively, and the proto-aptamer to which each pair belongs is indicated in bold at the top-left corner of the page. The 10-nucleotides sequences shown in red boxes are evolutionary conserved sequences and they add up to the consensus: 5′-UAUUWUAUUU-3′ (W = U or A). The shorter sequences (UGA and UUGG) in blue boxes are conserved sequences as well. A straight line that binds two stretches of nucleobases in a given sequence is a pseudo-knotted region as predicted by *Kinefold*. No topologically entangled helices (knots) were predicted by *Kinefold* (when found, are indicated by the base pairs drawn with the same color as the phosphate-ribose backbone). Dotted base pairs in contrast to those connected by a straight line in *Kinefold* models represent the non-Watson-Crick base pair GU. Stochastic folding simulation: renaturation folding.

By examining the six global secondary structures generated, it can be seen that a consensus secondary structure emerges and is describable as a multi-way looped junction. Despite this defined consensus, the six MFE structures vary slightly between each other. For example, the junction in 2 of the 6 structures is a bulged loop and not an entirely closed loop (sealed by base pairing). These structures are of noorA and noorB and both are predicted by *Kinefold*. Bulged loops are one of the most common secondary structure formations in RNA; they were investigated numerously and are known to be versatile architectural motifs and also mediators of target recognition and binding (e.g. to proteins) and are known to be significant in many biological processes such as intron splicing feedback recognition and tertiary folding (Crowther et al., 2017; Hermann & Patel, 2000). The third *Kinefold* structure (noorC) as well as the remaining three structures (that are predicted by *RNAfold*), conform to a subsidiary consensus secondary structure defined by a closed-loop junction instead of the generalized “looped junction”. Other variations arise from the stems of the structures. For example, by counting (in addition to noorB *RNAfold* and noorC *Kinefold*) the stems before and after a bulge (*Kinefold* structures of noorA and B) as two stems and not one stem, the consensus secondary structure becomes a two-way looped junction with 2 of the 6 structures (noorA and C; *RNAfold* versions) deviating from this consensus by forming a three-stem junction. Otherwise, if the

pre- and post-bulge stems are considered as a single stem (which is unlikely to be the case due to the significant size of the bulges), then, one-stem, two-stem and three-stem structures become distributed equally (in doubles) among the six predicted secondary structures. Regardless of their number around the loop, these stems (or at least a portion of them) may play a significant role in the final stabilization of the main central loop. This is especially the case when considering the fact that local stem areas that are in close proximity to a given loop-stem intersection are GC-rich. In addition to all the aforementioned structural variations that may or may not be relevant to discern the efficiency of target binding (by dictating the stability and accessibility properties of the aptamer), we found that all six RNA proto-aptamers share in their randomized core region an evolutionarily conserved (by virtue of the SELEX process) deca-nucleotide RNA sequence that may be a strong candidate to serve as the binding motif that constitutes (from an aptamer's perspective) the AdiC-aptamer binding interface. Through multiple sequence alignment of noorA, noorB and noorC, the consensus of this conserved putative motif is revealed to be the uracil stretch: 5′-UAUUWUAUUU-3′ (W = 66.67% uracil and 33.33% adenine, among the three proto-aptamers). This sequence is located about midway of the core region (and the full proto-aptamer) and, perhaps interestingly, in the central looped junctions (from which the multiple stems branch off) of the six secondary structures in **Figure 4.8**. The localization of this putative deca-nucleotide binding motif in the central loops of the secondary structures of noorA, B and C is interesting as it is atypical to the paradigm of having motifs located in the loops of hairpin substructures. More than that, localization of this conserved sequence in the looped junctions supports its annotation as a putative AdiC binding motif further when this localization is viewed by the notion that the surrounding stems may be strongly needed for the stabilization of this loop and hence, stabilization and presentation of this recognition motif to its target. These hypotheses are testable in different ways: for example, targeted mutational analyses can be done to experimentally determine the indispensability of the discovered motif for binding activity as done in previous studies (Dollins et al., 2008; Dubey et al., 2005; Mende et al., 2007). Also, controlled 5′- and 3′-end

truncations of the candidate aptamers can be designed to assess different aspects such as conservation of the global secondary structure or the loop junction as well as assessing any changes in the dissociation constant (K_d value) for the target molecule (Bing et al., 2010; Rockey et al., 2011). The alignment of the three abundant proto-aptamers is shown below in ClustalW format and the conserved sequences are shown inside a red box in both the aligned sequences as well as the secondary structures in **Figure 4.8** (asterisks in the alignment below represent a 100% identical nucleobases for all sequences):

```
CLUSTAL O(1.2.4) multiple sequence alignment

noorA   GGGAGACAAGAAUAAACGCUCAAGUGUCUGUU-----CAUGAUUGCAUUUUUAUAUUAUA 55
noorB   GGGAGACAAGAAUAAACGCUCAACGA-----U-----UGUGAUGCCUCUUUUUUAUUUUA 50
noorC   GGGAGACAAGAAUAAACGCUCAAUGAGCCCCUAUUUUUCUACCAACAACUUGGUAUUUUA 60
        ************************        *       *     *   **  **** **

noorA   UUU-------UCACCUUGGUGUUUUUAAUUCGACAGGAGGCUCACAACAGGC 100
noorB   UUUGUGUUGUUUAUAUUGGGA--UUUUAUUCGACAGGAGGCUCACAACAGGC 100
noorC   UUU--------UAU----AUAUUUUCAAUUCGACAGGAGGCUCACAACAGGC 100
        ***         *          **  *************************
```

Despite us defining a potential aptamer binding motif by comparative sequence and structural analysis, it is possible that this discovered motif may require fine-tuning by shifting the red box upstream or downstream of the loops and/or adjusting its extent especially that the loop junctions of the three sequences are basically uracil-rich regions that are broken at different positions in that substructure mostly by adenine residues. Better definition of the true borders of the motif and accurately determining its extent may also be achieved by mutational analyses or by investigating the aptamer-target complex and the formed binding interface. In line with the idea that multiple binding motifs are definable, we set to explore the possibility for other conserved sequences located at other regions outside the U-rich central loop junctions in the three proto-aptamer sequences being investigated. Because the quality and the accuracy of an output sequence alignment can always be a matter under question especially when developing a conservational relationship between the primary sequence and higher order structures (Andrews & Moss, 2019;

Chatzou et al., 2016; Mathews et al., 2010), we decided to realign noorA, B and C in Clustal Ω to try to uncover the new conserved regions. Our strategy to the new alignment was to simply realign the core region of the proto-aptamers by eliminating the flanking constant regions from the alignment. The result of this alignment is shown below:

```
CLUSTAL O(1.2.4) multiple sequence alignment

noorC   -----------UGAGCCCCUAUUUUUCUACCA----------ACAACUUGGUAUUUUAUU 39
noorA   GUGUCUGUUCAUGAUUGCAUUUUUAUAUUAUAUUU-------UCACCUUGGUG------- 46
noorB   -----CGAUUGUGAUGCCUCUUUUUUAUUUUAUUUGUGUUGUUUAUAUUGGGA------- 48
                  ***   *   *** * *   *            *  ****

noorC   UUAUAUAUUUUCAA 53
noorA   -------UUUUUAA 53
noorB   -------U-UUUA- 53
               * ** *
```

As it can be seen, alignment of only the SELEX-evolved core regions revealed two additional invariant sequences that are 100% conserved in noorA, B and C; these are 5′-UGA-3′ and 5′-UUGG-3′ and are both indicated on the alignment and the secondary structures (**Figure 4.8**) by a blue box. The two motifs (not necessarily binding motifs) seem to indirectly flank the conserved binding motif introduced earlier, with UGA from the 5′ side and UUGG from 3′ side; this however is not the cases for noorC as both UGA and UUGG are located before the binding motif region (5′ side). Nonetheless, UUGG may be able to flank the binding motif of noorC from the 5′ side if the binding motif was found to require readjustment towards somewhere in the region between the 32 and 40 nucleotide positions. After all, a new derivative binding motif seems feasible when looking at the new alignment above (see the asterisks in the blue oval shape). For noorA, UGA is located at around the base of the loop junction near the first branching stem. For noorB and C, UGA is located fully on the first stem and towards its midsection. On the other hand, for UUGG, noorA seems to have it also around the base of loop near the second stem. For noorB, UUGG is located fully on the loop of the second stem, and for noorC, UUGG can be found on the second stem just before the predicted binding motif

(UAUUWUAUUU). At this point and given their locations, it may be suitable to annotate both UGA and UUGG as stabilizing agents of the loop junctions of noorA, B and C in light of the overall structural-functional model we are trying to build herein for these proto-aptamers. To justify this novel putative annotation, for UUGG, for example, the sequence seems to participate in pseudoknot formation in both noorA and noorB as it is seen from their *Kinefold* predictions. However, in noorC, no pseudoknots were predicted and UUGG may be participating in loop-junction stabilization by promoting the formation of the second stem in that region. Pseudoknots can actually be found frequently in natural non-coding RNAs and they have many diverse and important functions including RNA structure stabilization, direct target binding, gene regulation (e.g. riboswitches) and even catalytic activity (e.g. ribozymes) (Peselis & Serganov, 2014; Staple & Butcher, 2005). For the UGA motif, although a loop-stabilization function is appointable, at this point, it is unknown whether UGA may provide significant contribution to stabilization of the loop given its small size (3 nucleotides) and its significant distance from the actual loop junction (for noorB and noorC). Moreover, for noorA, it is unknown whether UGA constitutes the point of intersection between the loop and the stem or it fully resides as part of the loop (*RNAfold* vs. *Kinefold* model). This may be important to determine how exactly UGA stabilizes the loop (if the stabilizer annotation was valid). For instance, if the latter case was correct, UGA may be an important stabilizer of the bulge and the bending angle between the stems (Luebke & Tinoco, 1996) In any case and unlike for noorB and noorC, the element of distance is at least eliminated for the motif (UGA) in noorA as it is found near the base of the loop. Answers to questions like if UGA and UUGG are true loop-stabilizing motifs and whether UAUUWUAUUU is a true binding motif are approachable (as described earlier) in many different ways ranging from experimental methods to analyzing the primary and secondary structure conservation of these motifs comparatively in the context of the complete collection of proto-aptamer sequences (**Table 4.3**), or by a combination of these approaches. This way, a higher statistical sense for the

magnitude of the selection pressure imposed on these three motifs can be gained, as elaborated in the upcoming sections.

4.3.3 Construction of a Family Tree and Phylogenetic Analysis of the Aptamer Families

After the isolation of aptamer clones and elucidating their primary sequences, these aptamers are typically grouped into aptamer families/classes/clusters based on the homology (similarity) of their primary sequences (Bing et al., 2010; Davis, 1998; Iaboni et al., 2016). Alternatively, aptamers have also been grouped primarily based on the occurrence of certain consensus sequences (Dubey et al., 2005; Matthias Homann & Göringer, 1999; Jang et al., 2008) or even based on secondary structure formations (Y. Z. Huang et al., 2012). In other approaches, aptamers were grouped based on their affinities (Jinpeng Wang et al., 2012) or their functional effect (e.g. inhibitory versus non-inhibitory) on their target (Yang Cui et al., 2004), or were ranked and subsequently chosen for further study based on their abundance (our approach with noorA, B and C) and/or their enrichment profile throughout SELEX (Civit et al., 2019; Kahsai et al., 2016; Shangguan et al., 2006; Shubham et al., 2018; Zhong et al., 2019).

Here, instead of blindly analyzing the 34 unique proto-aptamer sequences (**Table 4.3**) against each other by directly comparing their individual 2D secondary structures (our approach with noorA, B and C due to a small sample size), we approached their analysis by first creating an “ancestral” relationship established based on sequence homology and using the multiple sequence alignment algorithm Clustal Ω. This relationship is expected to help us expand from the three abundant proto-aptamers and outwards to the rest of the aptamer candidates by homology-based, phylogeny-directed guidance. This is because it is often assumed that closely related sequences forming an aptamer family by sequence homology share the same binding secondary substructure and thus should bind the same site/target, albeit, binding of those same-family members can occur with different affinities possibly

due to a few critical nucleotides that can vary in the binding motifs of the aptamers under comparison (Bing et al., 2010; Shangguan, Tang, Mallikaratchy, Xiao, & Tan, 2007). Although the "different-family-different-structure" (and therefore, different target) assumption had been notably verified in certain studies, other studies on the other hand obtained a consensus secondary substructure that is common to all cloned sequences irrespective of their family grouping (by sequence homology) (Bing et al., 2010; Shangguan, Tang, et al., 2007). In light of the above insights, our 34 aptamer candidates (including noorA, B and C) were aligned against each other and grouped into three major aptamer families (I, II and III) based on their clustering in the phylogenetic tree. The tree and the family groupings are shown in **Figure 4.9**. Family I was divided into two subfamilies: IA which has 12 members and IB which has 10 members. In contrast, family II and III have 7 and 5 members, respectively. The alignment of these 34 aptamer candidate sequences is shown in **appendix M** in ClustalW format. From the phylogenetic tree in **Figure 4.9**, it can be seen that noorA and noorC are members of family II while noorB is a member of family III. This simple distinction indicates that the families we have defined do not necessarily have completely distinct secondary structures and that these families can overlap by their members (e.g. noorA and noorB) conserving the same structure (we have shown earlier that noorA and B, as well as noorC, follow a prominent consensus structure: a multi-way looped junction). In fact, based on the percent identity matrix that had resulted from the alignment shown in **appendix M**, noorA and noorB share a homology (76.84%) that is higher than that shared between noorA and noorC (71.43%) (noorB:noorC = 70%). Thus, members clustering into distinct families is not necessarily an absolute indicator of neither different structures nor weaker base homologies (at least, for the case of noorA, B and C in our tree). Interestingly, our observation that the same secondary structure (the multi-way looped junction) exists in two separate families matches with the common view that structure tends to be more conserved than sequence (Capriotti & Marti-Renom, 2010; Mathews et al., 2010) and may be considered as an indicator that the process of SELEX conducted here was indeed approaching its final round.

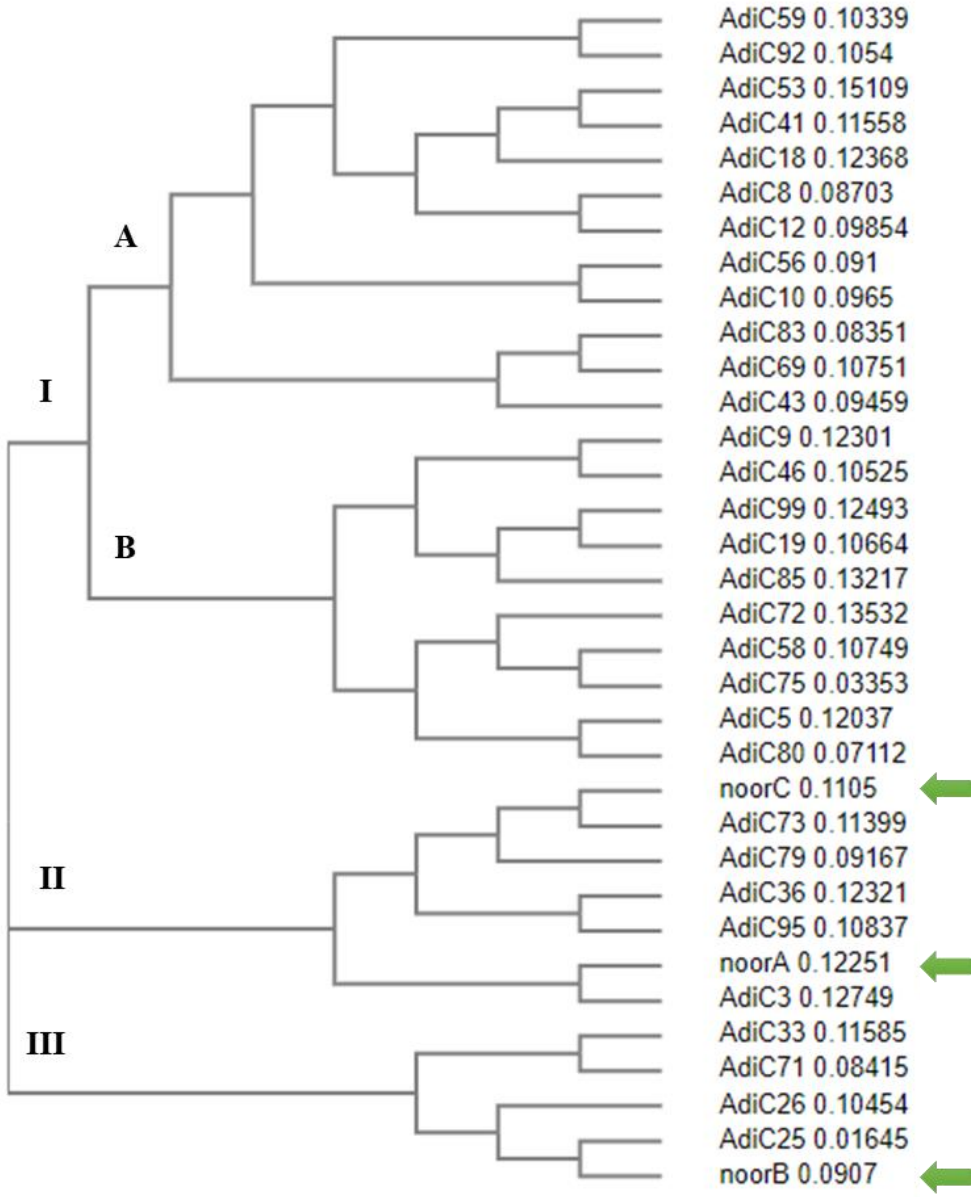


Figure 4.9. A Phylogenetic Tree Revealing the SELEX-derived Evolutionary Relationships Between the 34 Unique Proto-aptamer Sequences we Isolated. The "ancestry" seen between the 34 sequences is built based on the homology (multiple sequence alignment by Clustal Ω) between these sequences. Three aptamer families were defined based on this ancestry. The flanking constant regions of the proto-aptamers were included with the sequences in the input form so that they are incorporated during clustering and calculation of the final tree. noorA, noorB and noorC are shown with a green arrow. The scores next to each proto-aptamer name is a way to indicate evolutionary distance between the sequences.

It is intuitive to think that all members of a family share the same secondary structures and motifs due to the high similarity of their sequences (regardless of whether similarities with members from other families exist). However, for families of a tree that is built solely based on sequence homology, it may still be unwise to quickly infer higher-order structures conservation for all members of a given family in the tree. This is because a structure that is conserved between two homologous sequences in the same family begins to diverge significantly into very different structures after a certain point and it is generally accepted that the point of divergence is 60% sequence identity (Andrews & Moss, 2019; Capriotti & Marti-Renom, 2010). In other words, for identities < 60%, the relationship between sequence and structure conservation weakens progressively. Hence, for any two sequences with low identity, structure conservation is highly likely to be lost even if they are in the same family. For our case of 34 proto-aptamers for which a homology-based phylogenetic relationship was built, it could be reliable to assume that at least for families II and III, conservation of a stable loop substructure with the consensus binding domain defined earlier (UAUUWUAUUU) is likely to be high between members of these two families. This assumption is supported by the fact that noorA, B and C (which highly conserve the multi-way looped junction among themselves) are distributed well apart from each other and are not clustered into one family as it is seen from the tree. This phylogenetic distancing (determined by the high number of branch points) allows establishing a notional zone of high conservation defined by the consensus structure shared between these three proto-aptamers (the central loop harboring the binding motif and having stabilizing stems). Within this zone, other members of family II and III are "likely" to reside given their phylogenetic proximity to at least one of the three proto-aptamers. More importantly, given the % identities matrix formed between noorA, B and C (all identities 70% and above), the "phylogenetic zone" of high structure conservation defined by these proto-aptamers now becomes more credible and inhabitants of this zone (more likely to be members of family II and III) are also likely to have highly similar sequences (> 60% identity) in addition to conservation of base pairing (Capriotti & Marti-Renom, 2010). A comprehensive

and direct identity-based approach can be used to assess the overall trend of structure conservation between the tree families and/or the overall level of conservation in the tree. The approach is to determine the level of complexity between the repertoire of 34 sequences by examining the sequence homology between them and/or examining the two pairs of sequences representing the lowest and highest identities. From the percent identity matrix, we were able to determine that the lower and upper bounds for sequence similarity are 60.44% (AdiC95 and AdiC53) and 89.29% (AdiC25 and noorB), respectively. Since these values are above 60%, we expect a significant number of all the structures to conserve at least a loop substructure (e.g. hairpin) that contains the binding domain defined earlier (a U-stretch broken with A residues). Such a substructure must be of similar size to that of the consensus (both loop-wise and stem-wise). By extrapolating the arguments put forth above, we also hypothesize that we are generally more likely to experience less conserved and more divergent structures (do not conform closely to the multi-way loop junction consensus) with members from family I and especially the subfamily IA given its phylogenetic tree distance from families II and III. In addition, subfamily IA is more likely to constitute those sequences that share homologies closer to 60% with families II and III. However, as discussed earlier, these statements are not necessarily the case for all members of a given distant family.

To validate the strings of analyses and logical arguments above, we examined the level of secondary structure conservation for all sequences and we provide this here in the form of **Table 4.4** that shows for each family, the number of structures predicted (by at least one prediction programs) to conserve at least a stem-loop substructure with both the predefined binding motif and a size that resemble the multi-way loop junction. This information is presented in table format due to the sheer number of MFE structures predicted. *Kinefold* predictions with renaturation and co-transcriptional folding yielded the same structure for all of the sequences. From the table, indeed the phylogenetically distant family I has more sequences that tend towards not conserving a desired secondary structure and a binding motif, with subfamily IA having the least number of members (compared to its number of

members) that resemble the defined minimal secondary structure or the consensus. Additionally, the zone of structural conservation created by the proto-aptamers noorA, B and C can be seen as 100% of the members from family II and III resemble the consensus. In the end, we note again that even if structures were conserved, similar binding affinity may not be a reality as subtle nucleobase variations can make huge differences for these affinities.

Table 4.4 A count of the number of secondary structures predicted for each aptamer family to conserve at least a stem-loop substructure with the putative binding domain and a similar size to the consensus. A "predicted" designation is given if a structure fits the described criteria as predicted by a given program. A "–" is given when a predicted structure differs from the desired criteria. Structures predicted by both programs are counted as one structure. The level of conservation for one family is given in percentage of the number of predicted structures divided by the total number of members in that family. noorA, B and C are not counted as the comparison is made against the consensus which is defined by them. Proto-aptamers with superscripted "a" and "b" have UGA and UUGG motifs, respectively. The number in brackets next to each "a" and "b" is the number of times the motif was repeated in the sequence.

Family	***Proto-aptamer***	***RNAfold***	***Kinefold***	***Conserved Structures Per Family***
IA	**AdiC59**	Predicted	Predicted	**58%**
	AdiC92	Predicted	Predicted	
	AdiC53	–	–	
	AdiC41 $^{a(2)}$	Predicted	Predicted	
	AdiC18 $^{a(1)}$	–	–	
	AdiC8	Predicted	–	
	AdiC12	–	Predicted	
	AdiC56	–	–	
	AdiC10	Predicted	Predicted	
	AdiC83	–	–	
	AdiC69 $^{a(5)}$	–	Predicted	
	AdiC43 $^{a(1)}$	–	–	
IB	**AdiC9** $^{b(1)}$	–	–	**70%**
	AdiC46	–	Predicted	
	AdiC99	Predicted	Predicted	
	AdiC19 $^{a(1)b(1)}$	Predicted	Predicted	
	AdiC85 $^{a(1)}$	–	–	
	AdiC72 $^{a(2)b(1)}$	Predicted	Predicted	

Table 4.4 (Cont'd)

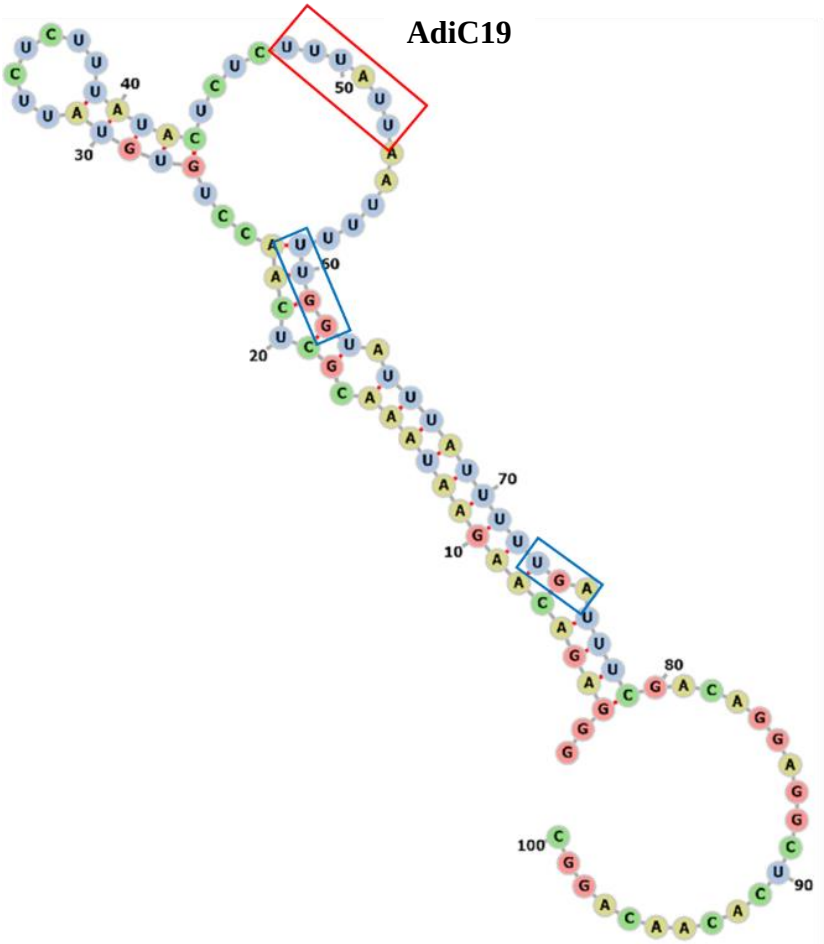
	AdiC58 [b(1)]	Predicted	Predicted	
	AdiC75	–	–	
	AdiC5 [b(1)]	Predicted	Predicted	
	AdiC80	Predicted	Predicted	
II	**noorC**			**100%**
	AdiC73 [a(1)]	Predicted	Predicted	
	AdiC79	–	Predicted	
	AdiC36	Predicted	Predicted	
	AdiC95	Predicted	Predicted	
	noorA			
	AdiC3 [b(1)]	–	–	
III	**AdiC33**	Predicted	–	**100%**
	AdiC71	–	Predicted	
	AdiC26	–	Predicted	
	AdiC25	–	Predicted	
	noorB			

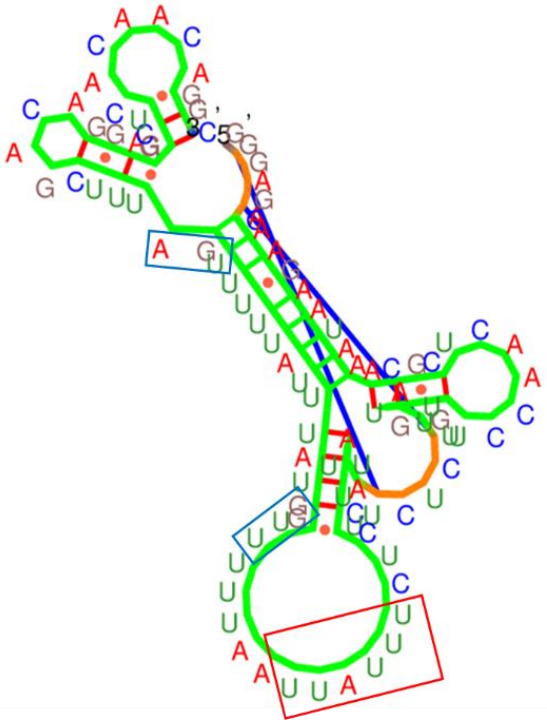
4.3.4 Identification of Singletons with Similar Architecture to the Abundant Candidates

In addition to noorA, B and C, we set to determine which other single-copy proto-aptamers from the 34 evolved and isolated candidates had all the three domains (UGA, UAUUWUAUUU, and UUGG) and also a predicted MFE structure that conforms closely to the criteria structure defined in **Table 4.4** (i.e. mutual non-exclusivity). Distribution of UGA and UUGG among the 34 sequences is shown in **Table 4.4**. Two new candidates were found—AdiC19 and AdiC72 to fulfill the requirements despite being in a family (family IB) that is other than II and III. Both of them were found to have a stemmed loop structure containing a potential binding

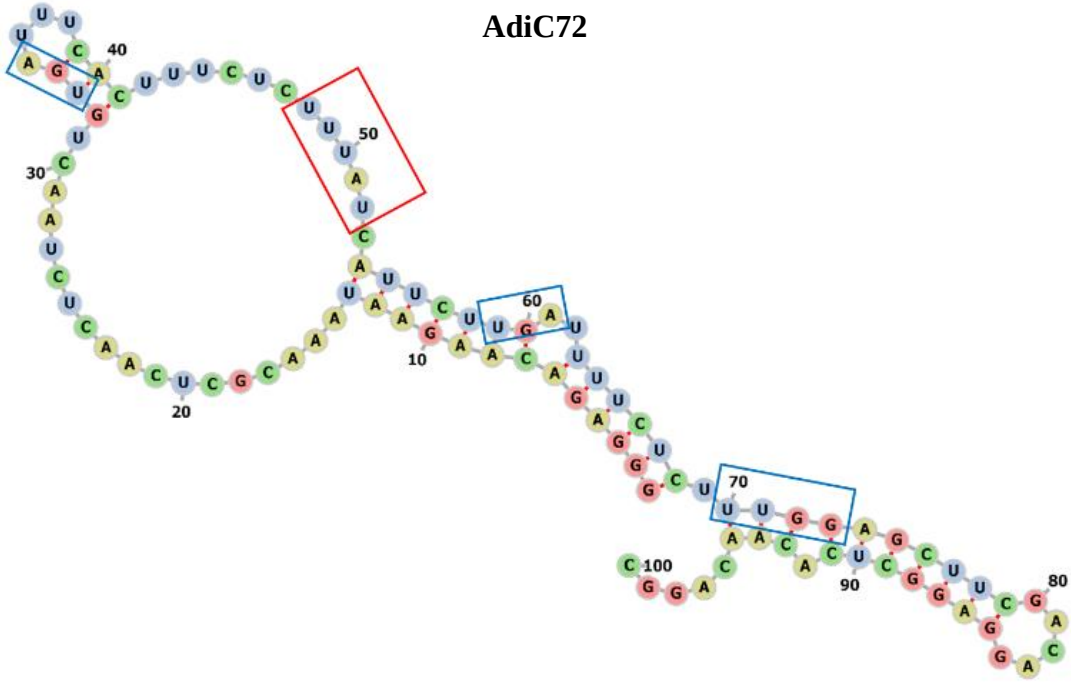
U-stretch motif as well as the UGA and UUGG domains participating in stem formation in both proto-aptamers. AdiC19 has a sequence homology with the previous three proto-aptamers of > 70–71% and AdiC72 comes at a 67% < homology < 69%. **Figure 4.1** shows the *RNAfold* and *Kinefold* models of the secondary structures of AdiC19 and AdiC72. The *RNAfold* model of AdiC19 highly resembles that of noorB. In addition, interestingly, the *Kinefold* model of AdiC19 is also very divergent from its *RNAfold* model just as the *Kinefold* model of noorB is divergent from its *RNAfold* model. The reason for this can very well be due to the disparate capability *Kinefold* and *RNAfold* to predict pseudo-knotted structures. Indeed, pseudoknots are molecular constructions that are unfortunately difficult to predict most of the time by the common DP methods (Andrews & Moss, 2019; Shapiro et al., 2007). This is because they violate the tree topology of pseudoknot-free structures that is convenient for DP by requiring a graph topology. Although pseudoknots are typically forbidden in DP methods, topological constraints may be placed on them to aid their prediction. Heuristic approaches are also more common for the prediction of non-nested pseudoknots. Another reason for the extreme structural disparity between the AdiC19 models is that *Kinefold* seems to have predicted actual topological knots at the central stem of the structure. AdiC72, on the other hand, highly resembles noorC and no additional topologies were predicted by the program. Moreover, it is notable that AdiC72 seems to have in addition to a UUGG domain, two UGA domains that indirectly flank the binding domain in stems and not one. Finally, it is notable that if the criteria to identification of these singletons was less strict, more candidate singletons can be harvested. For example, it may be worth it to explore AdiC25 which has the highest sequence homology with noorB (89.29%) among any other sequences pair and a predicted structure that conforms to the consensus (by *Kinefold*) despite lacking UGA and UUGG motifs. In addition, since all the structures in **Table 4.4** were predicted at 37°C, it may be interesting to explore the possible secondary structures generated from prediction at the temperature of selection—RT (~22°C).

AdiC19





AdiC72



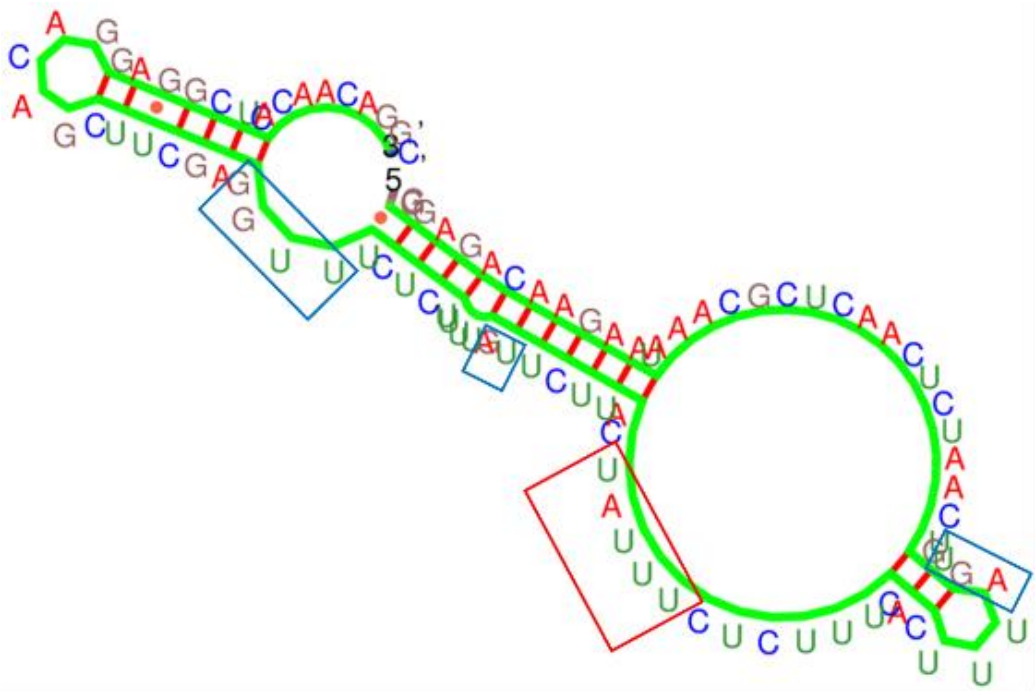


Figure 4.10. 2D MFE secondary structures of AdiC19 and AdiC72 predicted by *RNAfold* and *Kinefold*. The left- and right-hand side structures were built by *RNAfold* and *Kinefold*, respectively, and the proto-aptamer to which two adjacent structures belongs is indicated in bold above that pair. Possible binding motifs are shown inside a red box. UGA and UUGG are shown inside a blue box. AdiC19 (*Kinefold*) is pseudo-knotted and also seems to have entangled helices at the central stem by the green base pairs. folding simulation (*Kinefold*): renaturation folding.

4.3.5 Prediction-driven Mutational Deletions and Significance of the Discovered UGA and UUGG domains

We systematically conducted *in-silico* prediction-driven mutational deletions of motifs UGA and UUGG to assess the effect of their deletion on compromising the secondary structures of the five candidate proto-aptamers (noorA, noorB, noorC, AdiC19, and AdiC72) and on the display of the binding motif in a loop substructure. The strategy to this was to eliminate UGA and UUGG sequences one at a time (not together) and then predict the MFE structure from the "mutated" sequence using both *RNAfold* and *Kinefold*. For all *Kinefold* predictions, renaturation and co-transcriptional folding yielded the same secondary structures. Although most UGA deletions had no significant effect on the overall structures of the five proto-aptamers (for both *RNAfold* and *Kinefold* models) and on their presentation of the binding motif, several notable observations do exist. For example, the mutated *kinefold* MFE model of noorB had the same overall the structure as that of the original model but with the bulged loop completely gone. Further, the structure became even more heavily pseudo-knotted and new topological knots have appeared as well. However, the second MFE mutated structure shown by *Kinefold* (Gibbs free energy change difference between the two mutated structures = 0.1 Kcal/mol) returns to a structure that is almost similar to that of the nonmutated one (ΔG is 3.2 Kcal/mol higher for the second mutated MFE structure compared to the original structure in **Figure 4.8**) and with the looped bulge restored too. No such significant change was seen for the RNAfold comparisons of UGA deletion perhaps, again, due to the difference in

pseudoknot prediction ability. Another example is that of AdiC19; interestingly, the deletion of UGA had resulted in complete annihilation of the two-way looped junction (*RNAfold* model) and now, residence of the binding motif in the in an open loop that is unlikely to form binding interactions. This observation is interesting given the large distance of this tri-nucleotide domain from the binding domain. Conversely, for the model predicted by *Kinefold*, the mutated structure seems to have preserved the binding motif in the hairpin structure but reverted the rest of the structure into a different base pairing pattern; the free energy of the new mutated structure increased by 2.3 Kcal/mol. For AdiC72, although UGA deletion (5′ side) caused no change in display of the binding domain in either models, the structure however converted to one that is very similar to that of noorB and AdiC19 in the *RNAfold* model and eliminated the small stem and formed a new pseudoknot in its position in the *Kinefold* model. The deletion of UUGG motif seems to have more prominent effects on the structures of the five proto-aptamers than UGA deletions do. For example, the looped junction of noorA harboring its binding motif in both RNAfold and Kinefold models broke into an open loop. In noorC, UUGG deletion in both models resulted in a significant expansion of the loop junction by incorporation of stem nucleotides to form a large loop that is connected only to the main stem. Due to its increased size, the loop may have a diminished binding affinity. In AdiC19, the deletion of UUGG (*RNAfold* model) had the same effect of its UGA deletion on the binding motif and the structure. However, in the *Kinefold* model, like deletion of UGA, the structure takes a different fold and maintains the binding domain this time in a bulged loop and not as a hairpin substructure. UUGG deletion resulted in significant changes in the models of noorB as well. In *RNAfold*, the structure becomes more linearized with longer stems and the binding motif now becomes distributed between two small adjacent internal loops and a 3′ stem. In *Kinefold*, the structure is also linear, but the binding motif is now part of an internal loop that is slightly larger than the bulging loop seen in the original structure. Despite the various effects described due to the UGA and UUGG motif deletions from the five aptamer candidates in two different prediction programs, one common effect

that unites the deletions is that almost all structures had a kind of decrease in their predicted stabilities (ΔG increased).

4.3.6 Are the Optimal (MFE) Secondary Structures Reliable?

DP approaches which are known to thermodynamically optimize RNAs by energy minimizations to achieve folding of the RNA in question into its optimum MFE secondary structure are by far the most dominant approach to prediction (Andrews & Moss, 2019). However, in addition to the many inherent assumptions and limitations of DP methods that limit the accuracies of prediction results, it is certainly an open question whether the predicted optimal structures do represent the real/native structure. This question is especially relevant for RNAs that function in the *in vivo* environment where their folding is influenced by multiple factors that determine their final optimal or suboptimal (and therefore, biologically relevant) structure. Nevertheless, methods to assessment and improvement of the modeled structure do exist (e.g. partition functions, comparative analyses and constraining with experimental data).

At the present time, it is unknown whether the structures predicted in this study represent a molecular reality. However, reliability of these structures may be spoken of in two aspects. First, if at the molecular level, the five proto-aptamers selected in this work truly embrace a minimized structure dictated only by its thermodynamic forces guiding its folding, then the question of reliability boils down to factors such as the kinetic effects of folding, non-nearest-neighbor effects (e.g. 3D pseudoknots) and energetic contribution of noncanonical base pairs. Because it is generally known that prediction accuracies of RNAs under 700 nucleotides identifies ∼70% of bps correctly and that it drops rapidly for RNAs > 700 nucleotides, it may be reasonable to expect that the structures predicted for our ∼100-nucleotides RNAs in this study are accurate and that the intrinsic effects of RNA folding dynamics described above can be undermined. Especially, this is true when we use a program that permits formation and visualization of pseudoknots.

The second aspect to addressing the reliability of the secondary structures predicted in this work is if they were introduced into an *in vivo* environment or to co-function with a cellular system. Although it is unknown whether the aptamer candidates selected here will assume different structures when placed in a biological environment due to factors such as molecular crowding or intermolecular interactions (another example of non-nearest-neighbor effects), again, the small size of these candidates may be their safe haven from external factors as thermodynamic equilibrium dominates. The predicted stabilities of the five aptamer candidates are summarized in **appendix N**.

All in all, the fact that we were able to identify a recurrent structural pattern (three motifs and a consensus secondary structure) in five *in vitro*-selected aptamer candidates using comparative analyses of specifically optimal MFE structures (and primary sequences as well) points at the success of the systematic evolution process (SELEX) carried out here and more importantly, says something that may be worth hearing about the true structural nature of these five RNA molecules.

CHAPTER 5

CONCLUSIONS, FUTURE WORK, & OUTLOOKS

An extensive review of the current literature has allowed us to conclude that over the past ~20 years and since the first full-length protein target (the GPCR, NTS-1) had been overexpressed, purified stably in detergent micelles, and used during *in vitro* aptamer selection (Daniels et al., 2002), selections targeting purified full-length MPs have been reported scarcely. This is more pronounced especially for complex MPs such as channels, transporters and GPCRs mainly due to the difficulties in overexpressing and purifying them in desirable states (abundant, stable, and preferably, functional) to be fed later into protein-SELEX. Consequently, most MP-specific aptamers were selected through the use of live cells and also MP soluble domains as the baits of the selection process. Use of these targets in SELEX has provided revolutionary benefits in addition to circumventing the issues of obtaining full-length MPs. In cell-SELEX, the target protein can even be identified pre-SELEX, post-SELEX, or unidentified at all (in the latter, selection targets the general phenotypic blueprint of the cell). Despite all benefits, full-length MPs cannot be dispensed completely from the repertoire of SELEX targets because in certain cases, they can prove to be superior in different ways ranging from their quality as targets to impacting the SELEX procedure and its outcome as well. For example, whether soluble domains truly represent full-length proteins is always a question that dwells around and is one that must be addressed on a case-by-case basis. Also, cell-based selections have a risk of being not very feasible or reliable. During such scenarios *full-length*-MP SELEX may come as a surpassingly valid and powerful option. More than that, in the light of the rapidly transpiring innovations in SELEX methodology where complex materials and instrumentations are currently being incorporated (see **subsection 2.2.2.5** and **appendix A**), the defining principles of classical SELEX in which a polypeptide that is isolated at high purity from heterogeneity is used as a

target to evolve high-affinity aptamers by a smaller number of rounds, have especially remained appealing for and used by the scientific community (T. Wang, Yin, et al., 2019) Thus, preference for these principles is worth all efforts supporting development of the necessary tools to improve the recombinant expression and purification of MPs that have high value for the pipeline of aptamer research and biomedical applications. After all, the prevalence of publications on conventional SELEX where soluble proteins (including ectodomains) and many other nonprotein baits are targeted for selection only testifies to the popularity and so, the versatility and value of classical SELEX.

For the work herein, we have used for the first time pure PDL complexes of DDM-purified AdiC protein as a selection target in a protein-SELEX. From three AdiC aptamer families evolved, we were able to identify using only computational methods, five aptamer candidates that share high sequence homology as well as motif sequences and structural features. More candidates with varying features can certainly be harvested from the list of proto-aptamer. This thesis does not only support *in-silico*-based methods to aptamer identification but is also a proof-of-concept study that advocates for the use (and benefits) of purified PDL complexes of a protein of interest for targeting by aptamer libraries despite the difficulties that may be encountered during obtaining that protein in the first place. In addition to the benefits of *full-length*-protein SELEX, when viewed together with the work of (Ilgü et al., 2014), this thesis highlights a unique opportunity for complex MPs such as those found entirely embedded in biological membranes or highly insoluble (Zhou & Rossi, 2014). Since these attractive and crucial MPs may be inaccessible for the selection of aptamers while integrated in the bilayer of a virus or on the surface of a cell, they can be made more accessible after purification by good choice of the detergent and by carefully tuning its concentration until a desirable degree of de-lipidation is achieved. This is demonstrated by contrasting the DDM-purified AdiC PDL complex model we showed in **Figure 4.1** with a model built for another AdiC PDL complex this time purified in the detergent n-octyl-β-D-maltopyranoside (OM) (**see appendix O)**.

For the five RNA aptamer candidates identified in this study, we believe that these may have potential for use not only in scientific research to understanding and contributing to the knowledge on the structural/functional biology of AdiC growing over the past years but also be used in “real-life” applications in diagnostics and therapeutics. Actualizing these outlooks certainly requires further experimentation. For instance, questions on the binding affinities (dissociation constant, K_d) of these candidate aptamers may be explored *in vitro* qualitatively (e.g. pull-down assays or gel-shift assays) or quantitatively (e.g. SPR [surface plasmon resonance] or ITC [isothermal titration calorimetry]). Modeling the 3D tertiary structures of the aptamers and performing molecular docking simulations to understand the aptamer-AdiC interactions and to reveal the site of binding in the formed complex (intra- or extracellular side of AdiC) are feasible paths of experimentation as well. In bacterial culture-based assays, fluorescence microscopy can be used to assess aptamer binding to AdiC in its natural membrane environment for diagnostic applications. Functional effects such as cell growth inhibition may be studied for therapeutic applications. Moreover, as the aptamers candidates are 2’F-modified, their biostability is testable (Shangguan, Tang, et al., 2007). Additionally, truncations to determine the minimal folding/binding aptamer can be done; such efforts may even enhance the binding affinity of aptamers and increase the yield and lower the cost of aptamer synthesis (Bing et al., 2010; Rockey et al., 2011). Truncations may also reduce susceptibility to random degradations and ease downstream studies as well.

REFERENCES

Ababneh, N., Alshaer, W., Allozi, O., Mahafzah, A., El-Khateeb, M., Hillaireau, H., … Ismail, S. (2013). In vitro selection of modified RNA aptamers against CD44 cancer stem cell marker. *Nucleic Acid Therapeutics*, *23*(6), 401–407. https://doi.org/10.1089/nat.2013.0423

Afanasyeva, A., Nagao, C., & Mizuguchi, K. (2019). Prediction of the secondary structure of short DNA aptamers. *Biophysics and Physicobiology*, *16*(0), 287–294. https://doi.org/10.2142/biophysico.16.0_287

Ahirwar, R., Nahar, S., Aggarwal, S., Ramachandran, S., Maiti, S., & Nahar, P. (2016). In silico selection of an aptamer to estrogen receptor alpha using computational docking employing estrogen response elements as aptamer-alike molecules. *Scientific Reports*, *6*(1), 1–11. https://doi.org/10.1038/srep21285

Almeida, J. G., Preto, A. J., Koukos, P. I., Bonvin, A. M. J. J., & Moreira, I. S. (2017). Membrane proteins structures: A review on computational modeling tools. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1859*(10), 2021–2039. https://doi.org/10.1016/J.BBAMEM.2017.07.008

Amann, T., Schmieder, V., Faustrup Kildegaard, H., Borth, N., & Andersen, M. R. (2019, October 1). Genetic engineering approaches to improve posttranslational modification of biopharmaceuticals in different production platforms. *Biotechnology and Bioengineering*, Vol. 116, pp. 2778–2796. https://doi.org/10.1002/bit.27101

Andrews, R. J., & Moss, W. N. (2019). Computational approaches for the discovery of splicing regulatory RNA structures. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1862*(11–12), 194380. https://doi.org/10.1016/j.bbagrm.2019.04.007

Andrzejewska, A., Zawadzka, M., & Pachulska-Wieczorek, K. (2020). On the way to understanding the interplay between the rna structure and functions in cells: A genome-wide perspective. *International Journal of Molecular Sciences*, *21*(18), 1–15. https://doi.org/10.3390/ijms21186770

Antunes, D., Jorge, N. A. N., Caffarena, E. R., & Passetti, F. (2018). Using RNA sequence and structure for the prediction of riboswitch aptamer: A comprehensive review of available software and tools. *Frontiers in Genetics*, *8*(JAN), 231. https://doi.org/10.3389/fgene.2017.00231

Aquino-Jarquin, G., & Toscano-Garibay, J. D. (2011). RNA aptamer evolution: Two decades of

selection. *International Journal of Molecular Sciences*, Vol. 12, pp. 9155–9171. https://doi.org/10.3390/ijms12129155

Aquino, P., Honda, B., Jaini, S., Lyubetskaya, A., Hosur, K., Chiu, J. G., … Galagan, J. E. (2017). Coordinated regulation of acid resistance in Escherichia coli. *BMC Systems Biology*, *11*(1). https://doi.org/10.1186/s12918-016-0376-y

Asai, R., Nishimura, S. I., Aita, T., & Takahashi, K. (2004). In Vitro Selection of DNA Aptamers on Chips Using a Method for Generating Point Mutations. *Analytical Letters*, *37*(4), 645–656. https://doi.org/10.1081/AL-120029741

Avci-Adali, M., Metzger, M., Perle, N., Ziemer, G., & Wendel, H. P. (2010). Pitfalls of cell-systematic evolution of ligands by exponential enrichment (SELEX): Existing dead cells during in vitro selection anticipate the enrichment of specific aptamers. *Oligonucleotides*, *20*(6), 317–323. https://doi.org/10.1089/oli.2010.0253

Bai, X., Moraes, T. F., & Reithmeier, R. A. F. (2017). Structural biology of solute carrier (SLC) membrane transport proteins. *Molecular Membrane Biology*, *34*(1–2), 1–32. https://doi.org/10.1080/09687688.2018.1448123

Baneyx, F. (1999, October 1). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, Vol. 10, pp. 411–421. https://doi.org/10.1016/S0958-1669(99)00003-8

Barfod, A., Persson, T., & Lindh, J. (2009). In vitro selection of RNA aptamers against a conserved region of the Plasmodium falciparum erythrocyte membrane protein 1. *Parasitology Research*, *105*(6), 1557–1566. https://doi.org/10.1007/s00436-009-1583-x

Baumgarten, T., Schlegel, S., Wagner, S., Löw, M., Eriksson, J., Bonde, I., … De Gier, J. W. (2017). Isolation and characterization of the E. coli membrane protein production strain Mutant56(DE3). *Scientific Reports*, *7*(1), 45089. https://doi.org/10.1038/srep45089

Berezovski, M., Musheev, M., Drabovich, A., & Krylov, S. N. (2006). Non-SELEX selection of aptamers. *Journal of the American Chemical Society*, *128*(5), 1410–1411. https://doi.org/10.1021/ja056943j

Berezovski, M. V, Lechmann, M., Musheev, M. U., Mak, T. W., & Krylov, S. N. (2008). Aptamer-facilitated biomarker discovery (AptaBiD). *Journal of the American Chemical Society*, *130*(28), 9137–9143. https://doi.org/10.1021/ja801951p

Beringer, R. W., Ash, D. L., & Page, A. F. (2011). Influence of Buffer on Choice of Protein Quantification Method. In *Thermo Scientific*. Retrieved from https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/T112-Influence-of-

Buffer-on-Choice-of-Protein-Quantification-Method.pdf

Berkhout, B., & Klaver, B. (1993). In vivo selection of randomly mutated retroviral genomes. *Nucleic Acids Research*. https://doi.org/10.1093/nar/21.22.5020

Bernaudat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., … Rolland, N. (2011). Heterologous expression of membrane proteins: Choosing the appropriate host. *PLoS ONE*, *6*(12), e29191. https://doi.org/10.1371/journal.pone.0029191

Bi, H., Sun, L., Fukamachi, T., Saito, H., & Kobayashi, H. (2009). HU participates in expression of a specific set of genes required for growth and survival at acidic pH in Escherichia coli. *Current Microbiology*, *58*(5), 443–448. https://doi.org/10.1007/s00284-008-9340-4

Bianchi, F., van't Klooster, J. S., Ruiz, S. J., & Poolman, B. (2019). Regulation of Amino Acid Transport in Saccharomyces cerevisiae . *Microbiology and Molecular Biology Reviews*, *83*(4). https://doi.org/10.1128/mmbr.00024-19

Bill, R. M. (2014). Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. *Frontiers in Microbiology*, *5*(MAR), 85. https://doi.org/10.3389/fmicb.2014.00085

Bill, R. M., Henderson, P. J. F., Iwata, S., Kunji, E. R. S., Michel, H., Neutze, R., … Vogel, H. (2011). Overcoming barriers to membrane protein structure determination. *Nature Biotechnology*, *29*(4), 335–340. https://doi.org/10.1038/nbt.1833

Bill, R. M., & von der Haar, T. (2015, June 1). Hijacked then lost in translation: The plight of the recombinant host cell in membrane protein structural biology projects. *Current Opinion in Structural Biology*, Vol. 32, pp. 147–155. https://doi.org/10.1016/j.sbi.2015.04.003

Bing, T., Yang, X., Mei, H., Cao, Z., & Shangguan, D. (2010). Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories. *Bioorganic and Medicinal Chemistry*, *18*(5), 1798–1805. https://doi.org/10.1016/j.bmc.2010.01.054

Blank, M., Weinschenk, T., Priemer, M., & Schluesener, H. (2001). Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels: Selective targeting of endothelial regulatory protein pigpen. *Journal of Biological Chemistry*, *276*(19), 16464–16468. https://doi.org/10.1074/jbc.M100347200

Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., & Toole, J. J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, *355*(6360), 564–566. https://doi.org/10.1038/355564a0

Boltz, A., Piater, B., Toleikis, L., Guenther, R., Kolmar, H., & Hock, B. (2011). Bi-specific aptamers mediating tumor cell lysis. *Journal of Biological Chemistry*, *286*(24), 21896–21905. https://doi.org/10.1074/jbc.M111.238261

Bosshart, P. D., Casagrande, F., Frederix, P. L. T. M., Ratera, M., Bippes, C. A., Müller, D. J., … Fotiadis, D. (2008). High-throughput single-molecule force spectroscopy for membrane proteins. *Nanotechnology*, *19*(38), 384014. https://doi.org/10.1088/0957-4484/19/38/384014

Bosshart, P. D., & Fotiadis, D. (2019). Secondary Active Transporters. In *Subcellular Biochemistry* (Vol. 92, pp. 275–299). https://doi.org/10.1007/978-3-030-18768-2_9

Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., … Neiers, F. (2016). A self-inducible heterologous protein expression system in Escherichia coli. *Scientific Reports*, *6*. https://doi.org/10.1038/srep33037

Bruno, J. G. (1997). In vitro selection of DNA to chloroaromatics using magnetic microbead-based affinity separation and fluorescence detection. *Biochemical and Biophysical Research Communications*, *234*(1), 117–120. https://doi.org/10.1006/bbrc.1997.6517

Bruno, J. G., & Kiel, J. L. (1999). In vitro selection of DNA aptamers to anthrax spores with electrochemiluminescence detection. *Biosensors and Bioelectronics*, *14*(5), 457–464. https://doi.org/10.1016/S0956-5663(99)00028-7

Burke, D. H., & Willis, J. H. (1998). Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. *Rna*, *4*(9), 1165–1175. https://doi.org/10.1017/S1355838298980542

Butt, T. R., Edavettal, S. C., Hall, J. P., & Mattern, M. R. (2005, September). SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*, Vol. 43, pp. 1–9. https://doi.org/10.1016/j.pep.2005.03.016

Cano-Garrido, O., Rueda, F. L., Sànchez-García, L., Ruiz-Ávila, L., Bosser, R., Villaverde, A., & García-Fruitós, E. (2014). Expanding the recombinant protein quality in lactococcus lactis. *Microbial Cell Factories*, *13*(1), 167. https://doi.org/10.1186/s12934-014-0167-3

Cano-Garrido, O., Sánchez-Chardi, A., Parés, S., Giró, I., Tatkiewicz, W. I., Ferrer-Miralles, N., … Garcia-Fruitós, E. (2016). Functional protein-based nanomaterial produced in microorganisms recognized as safe: A new platform for biotechnology. *Acta Biomaterialia*, *43*, 230–239. https://doi.org/10.1016/j.actbio.2016.07.038

Capriotti, E., & Marti-Renom, M. A. (2010). Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC Bioinformatics*, *11*(1), 322.

https://doi.org/10.1186/1471-2105-11-322

Casalino, M., Prosseda, G., Barbagallo, M., Iacobino, A., Ceccarini, P., Carmela Latella, M., … Colonna, B. (2010). Interference of the CadC regulator in the arginine-dependent acid resistance system of Shigella and enteroinvasive E. coli. *International Journal of Medical Microbiology*, *300*(5), 289–295. https://doi.org/10.1016/j.ijmm.2009.10.008

Cassiday, L. A., & Maher, L. J. (2003). Yeast genetic selections to optimize RNA decoys for transcription factor NF-κB. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 3930–3935. https://doi.org/10.1073/pnas.0736013100

Catuogno, S., Di Martino, M. T., Nuzzo, S., Esposito, C. L., Tassone, P., & de Franciscis, V. (2019). An Anti-BCMA RNA Aptamer for miRNA Intracellular Delivery. *Molecular Therapy - Nucleic Acids*, *18*, 981–990. https://doi.org/10.1016/j.omtn.2019.10.021

Catuogno, S., & Esposito, C. L. (2017, September 1). Aptamer cell-based selection: Overview and advances. *Biomedicines*, Vol. 5, p. 49. https://doi.org/10.3390/biomedicines5030049

César-Razquin, A., Snijder, B., Frappier-Brinton, T., Isserlin, R., Gyimesi, G., Bai, X., … Superti-Furga, G. (2015). A Call for Systematic Research on Solute Carriers. *Cell*, *162*(3), 478–487. https://doi.org/10.1016/j.cell.2015.07.022

Chang, Y.-C., Kao, W.-C., Wang, W.-Y., Wang, W.-Y., Yang, R.-B., & Peck, K. (2009). Identification and characterization of oligonucleotides that inhibit Toll-like receptor 2-associated immune responses. *The FASEB Journal*, *23*(9), 3078–3088. https://doi.org/10.1096/fj.09-129312

Chatzou, M., Magis, C., Chang, J. M., Kemena, C., Bussotti, G., Erb, I., & Notredame, C. (2016, November 1). Multiple sequence alignment modeling: Methods and applications. *Briefings in Bioinformatics*, Vol. 17, pp. 1009–1023. https://doi.org/10.1093/BIB/BBV099

Chauveau, F., Aissouni, Y., Hamm, J., Boutin, H., Libri, D., Ducongé, F., & Tavitian, B. (2007). Binding of an aptamer to the N-terminal fragment of VCAM-1. *Bioorganic and Medicinal Chemistry Letters*, *17*(22), 6119–6122. https://doi.org/10.1016/j.bmcl.2007.09.046

Chelikani, P., Reeves, P. J., Rajbhandary, U. L., & Khorana, H. G. (2006). The synthesis and high-level expression of a β 2 -adrenergic receptor gene in a tetracycline-inducible stable mammalian cell line . *Protein Science*, *15*(6), 1433–1440. https://doi.org/10.1110/ps.062080006

Chen, C. H. B., Chernis, G. A., Hoang, V. Q., & Landgraf, R. (2003). Inhibition of heregulin signaling by an aptamer that preferentially binds to the oligomeric form of human epidermal

growth factor receptor-3. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9226–9231. https://doi.org/10.1073/pnas.1332660100

Chen, F., Hu, Y., Li, D., Chen, H., & Zhang, X. L. (2009). CS-SELEX generates high-affinity ssDNA aptamers as molecular probes for hepatitis C virus envelope glycoprotein E2. *PLoS ONE*, *4*(12). https://doi.org/10.1371/journal.pone.0008142

Chen, L., He, W., Jiang, H., Wu, L., Xiong, W., Li, B., … Qian, Y. (2019). In vivo SELEX of bone targeting aptamer in prostate cancer bone metastasis model. *International Journal of Nanomedicine*, *14*, 149–159. https://doi.org/10.2147/IJN.S188003

Chen, M., Yu, Y., Jiang, F., Zhou, J., Li, Y., Liang, C., … Zhang, G. (2016). Development of cell-SELEX technology and its application in cancer diagnosis and therapy. *International Journal of Molecular Sciences*, Vol. 17, p. 2079. https://doi.org/10.3390/ijms17122079

Chen, Y., Song, J., Sui, S. F., & Wang, D. N. (2003). DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in Escherichia coli. *Protein Expression and Purification*, *32*(2), 221–231. https://doi.org/10.1016/S1046-5928(03)00233-X

Cheng, C., Chen, Y. H., Lennox, K. A., Behlke, M. A., & Davidson, B. L. (2013). In vivo SELEX for identification of brain-penetrating aptamers. *Molecular Therapy - Nucleic Acids*, *2*, e67. https://doi.org/10.1038/mtna.2012.59

Cho, M., Oh, S. S., Nie, J., Stewart, R., Eisenstein, M., Chambers, J., … Soh, H. T. (2013). Quantitative selection and parallel characterization of aptamers. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(46), 18460–18465. https://doi.org/10.1073/pnas.1315866110

Cho, M., Xiao, Y., Nie, J., Stewart, R., Csordas, A. T., Oh, S. S., … Soh, H. T. (2010). Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(35), 15373–15378. https://doi.org/10.1073/pnas.1009331107

Chu, T. C., Twu, K. Y., Ellington, A. D., & Levy, M. (2006). Aptamer mediated siRNA delivery. *Nucleic Acids Research*, *34*(10). https://doi.org/10.1093/nar/gkl388

Chun, E., Thompson, A. A., Liu, W., Roth, C. B., Griffith, M. T., Katritch, V., … Stevens, R. C. (2012). Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure*, *20*(6), 967–976. https://doi.org/10.1016/j.str.2012.04.010

Cibiel, A., Dupont, D. M., & Ducongé, F. (2011). Methods to identify aptamers against cell surface

biomarkers. *Pharmaceuticals*, Vol. 4, pp. 1216–1235. https://doi.org/10.3390/ph4091216

Cibiel, A., Nguyen Quang, N., Gombert, K., Thézé, B., Garofalakis, A., & Ducongé, F. (2014). From Ugly Duckling to Swan: Unexpected Identification from Cell-SELEX of an Anti-Annexin A2 Aptamer Targeting Tumors. *PLoS ONE*, *9*(1), e87002. https://doi.org/10.1371/journal.pone.0087002

Civit, L., Theodorou, I., Frey, F., Weber, H., Lingnau, A., Gröber, C., … Mayer, G. (2019). Targeting hormone refractory prostate cancer by in vivo selected DNA libraries in an orthotopic xenograft mouse model. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41460-2

Claassens, N. J., Siliakus, M. F., Spaans, S. K., Creutzburg, S. C. A., Nijsse, B., Schaap, P. J., … van der Oost, J. (2017). Improving heterologous membrane protein production in Escherichia coli by combining transcriptional tuning and codon usage algorithms. *PLOS ONE*, *12*(9), e0184355. https://doi.org/10.1371/journal.pone.0184355

Cohen, S. N. (2013). DNA cloning: A personal view after 40 years. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 110, pp. 15521–15529. https://doi.org/10.1073/pnas.1313397110

Coulter, L. R., Landree, M. A., & Cooper, T. A. (1997). Identification of a new class of exonic splicing enhancers by in vivo selection. *Molecular and Cellular Biology*, *17*(4), 2143–2150. https://doi.org/10.1128/mcb.17.4.2143

Crowther, C. V., Jones, L. E., Morelli, J. N., Mastrogiacomo, E. M., Porterfield, C., Kent, J. L., & Serra, M. J. (2017). Influence of two bulge loops on the stability of RNA duplexes. *RNA*, *23*(2), 217–228. https://doi.org/10.1261/rna.056168.116

Cui, Y, Ulrich, H., & Hess, G. P. (2004). Selection of 2′-fluoro-modified RNA aptamers for alleviation of cocaine and MK-801 inhibition of the nicotinic acetylcholine receptor. *Journal of Membrane Biology*, *202*(3), 137–149. https://doi.org/10.1007/s00232-004-0725-4

Cui, Yang, Rajasethupathy, P., & Hess, G. P. (2004). Selection of stable RNA molecules that can regulate the channel-opening equilibrium of the membrane-bound γ-aminobutyric acid receptor. *Biochemistry*, *43*(51), 16442–16449. https://doi.org/10.1021/bi048667b

Daniels, D. A., Chen, H., Hicke, B. J., Swiderek, K. M., & Gold, L. (2003). A tenascin-C aptamer identified by tumor cell SELEX: Systematic evolution of ligands by exponential enrichment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15416–15421. https://doi.org/10.1073/pnas.2136683100

Daniels, D. A., Sohal, A. K., Rees, S., & Grisshammer, R. (2002). Generation of RNA aptamers to the G-protein-coupled receptor for neurotensin, nts-1. *Analytical Biochemistry*, *305*(2), 214–226. https://doi.org/10.1006/abio.2002.5663

Darmostuk, M., Rimpelova, S., Gbelcova, H., & Ruml, T. (2015). Current approaches in SELEX: An update to aptamer selection technology. *Biotechnology Advances*, *33*(6), 1141–1161. https://doi.org/10.1016/j.biotechadv.2015.02.008

Davis, K. (1998). Staining of cell surface human CD4 with 2'-F-pyrimidine-containing RNA aptamers for flow cytometry. *Nucleic Acids Research*, *26*(17), 3915–3924. https://doi.org/10.1093/nar/26.17.3915

Davydova, A. S., Timoshenko, V. V., Lomzov, A. A., Pyshnyi, D. V., Venyaminova, A. G., & Vorobyeva, M. A. (2020). G-quadruplex 2′-F-modified RNA aptamers targeting hemoglobin: Structure studies and colorimetric assays. *Analytical Biochemistry*, *611*, 113886. https://doi.org/10.1016/j.ab.2020.113886

de Marco, A. (2009, May 14). Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microbial Cell Factories*, Vol. 8, p. 26. https://doi.org/10.1186/1475-2859-8-26

Dilworth, M. V., Piel, M. S., Bettaney, K. E., Ma, P., Luo, J., Sharples, D., … Bill, R. M. (2018). Microbial expression systems for membrane proteins. *Methods*, *147*, 3–39. https://doi.org/10.1016/j.ymeth.2018.04.009

Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, *505*(7485), 696–700. https://doi.org/10.1038/nature12756

Dobbelstein, M., & Shenk, T. (1995). In vitro selection of RNA ligands for the ribosomal L22 protein associated with Epstein-Barr virus-expressed RNA by using randomized and cDNA-derived RNA libraries. *Journal of Virology*, *69*(12), 8027–8034. https://doi.org/10.1128/jvi.69.12.8027-8034.1995

Dollins, C. M., Nair, S., Boczkowski, D., Lee, J., Layzer, J. M., Gilboa, E., & Sullenger, B. A. (2008). Assembling OX40 Aptamers on a Molecular Scaffold to Create a Receptor-Activating Aptamer. *Chemistry and Biology*, *15*(7), 675–682. https://doi.org/10.1016/j.chembiol.2008.05.016

Dua, P., Kim, S., & Lee, D. ki. (2011). Nucleic acid aptamers targeting cell-surface proteins. *Methods*, Vol. 54, pp. 215–225. https://doi.org/10.1016/j.ymeth.2011.02.002

Dubey, A. K., Baker, C. S., Romeo, T., & Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *Rna*, *11*(10), 1579–1587. https://doi.org/10.1261/rna.2990205

Dunn, M. R., Jimenez, R. M., & Chaput, J. C. (2017). Analysis of aptamer discovery and technology. *Nature Reviews Chemistry*, *1*(10). https://doi.org/10.1038/s41570-017-0076

Dvorak, P., Chrast, L., Nikel, P. I., Fedr, R., Soucek, K., Sedlackova, M., … Damborsky, J. (2015). Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. *Microb Cell Fact*, *14*, 201. https://doi.org/10.1186/s12934-015-0393-3

Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818–822. https://doi.org/10.1038/346818a0

Eroglu, Ç., Cronet, P., Panneels, V., Beaufils, P., & Sinning, I. (2002). Functional reconstitution of purified metabotropic glutamate receptor expressed in the fly eye. *EMBO Reports*, *3*(5), 491–496. https://doi.org/10.1093/embo-reports/kvf088

Errey, J. C., & Fiez-Vandal, C. (2020). Production of membrane proteins in industry: The example of GPCRs. *Protein Expression and Purification*, *169*, 105569. https://doi.org/10.1016/j.pep.2020.105569

Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., … Miller, C. (2009). Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature*, *460*(7258), 1040–1043. https://doi.org/10.1038/nature08201

Fang, Y., Kolmakova-Partensky, L., & Miller, C. (2007). A bacterial arginine-agmatine exchange transporter involved in extreme acid resistance. *Journal of Biological Chemistry*, *282*(1), 176–182. https://doi.org/10.1074/jbc.M610075200

Farokhzad, O. C., Jon, S., Khademhosseini, A., Tran, T. N. T., LaVan, D. A., & Langer, R. (2004). Nanoparticle-aptamer bioconjugates: A new approach for targeting prostate cancer cells. *Cancer Research*, *64*(21), 7668–7672. https://doi.org/10.1158/0008-5472.CAN-04-2550

Foster, J. W. (2004). Escherichia coli acid resistance: Tales of an amateur acidophile. *Nature Reviews Microbiology*, *2*(11), 898–907. https://doi.org/10.1038/nrmicro1021

Fotiadis, D., Kanai, Y., & Palacín, M. (2013). The SLC3 and SLC7 families of amino acid transporters. *Molecular Aspects of Medicine*, *34*(2–3), 139–158. https://doi.org/10.1016/j.mam.2012.10.007

Gale, E. F. (2006). The Bacterial Amino Acid Decarboxylases. In *Advances in Enzymology and Related Areas of Molecular Biology* (Vol. 6, pp. 1–32). https://doi.org/10.1002/9780470122518.ch1

Ganson, N. J., Povsic, T. J., Sullenger, B. A., Alexander, J. H., Zelenkofske, S. L., Sailstad, J. M., … Hershfield, M. S. (2016). Pre-existing anti-polyethylene glycol antibody linked to first-exposure allergic reactions to pegnivacogin, a PEGylated RNA aptamer. *Journal of Allergy and Clinical Immunology*, *137*(5), 1610-1613.e7. https://doi.org/10.1016/j.jaci.2015.10.034

Gao, X., Lu, F., Zhou, L., Dang, S., Sun, L., Li, X., … Shi, Y. (2009). Structure and mechanism of an amino acid antiporter. *Science*, *324*(5934), 1565–1568. https://doi.org/10.1126/science.1173654

Gao, X., Zhou, L., Jiao, X., Lu, F., Yan, C., Zeng, X., … Shi, Y. (2010). Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature*, *463*(7282), 828–832. https://doi.org/10.1038/nature08741

García-Nafría, J., & Tate, C. G. (2020). Cryo-electron microscopy: Moving beyond X-ray crystal structures for drug receptors and drug development. *Annual Review of Pharmacology and Toxicology*, *60*(1), 51–71. https://doi.org/10.1146/annurev-pharmtox-010919-023545

Garibsingh, R. A. A., & Schlessinger, A. (2019). Advances and Challenges in Rational Drug Design for SLCs. *Trends in Pharmacological Sciences*, *40*(10), 790–800. https://doi.org/10.1016/j.tips.2019.08.006

Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., … Zichi, D. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE*, *5*(12). https://doi.org/10.1371/journal.pone.0015004

Gold, L., Javornik, B., Pagratis, N., & Shtatland, T. (2000). *TRUNCATION SELEX METHOD (WO/2000/056930)*.

Gomes, L., Monteiro, G., & Mergulhão, F. (2020). The impact of IPTG induction on plasmid stability and heterologous protein expression by escherichia coli biofilms. *International Journal of Molecular Sciences*, *21*(2). https://doi.org/10.3390/ijms21020576

Gong, Q., Wang, J., Ahmad, K. M., Csordas, A. T., Zhou, J., Nie, J., … Soh, H. T. (2012). Selection strategy to generate aptamer pairs that bind to distinct sites on protein targets. *Analytical Chemistry*, *84*(12), 5365–5371. https://doi.org/10.1021/ac300873p

Gong, S., Richard, H., & Foster, J. W. (2003). YjdE (AdiC) is the arginine:Agmatine antiporter essential for arginine-dependent acid resistance in Escherichia coli. *Journal of Bacteriology*,

185(15), 4402–4409. https://doi.org/10.1128/JB.185.15.4402-4409.2003

Goto, S., Tsukakoshi, K., & Ikebukuro, K. (2017). Development of aptamers against unpurified proteins. *Biotechnology and Bioengineering*, Vol. 114, pp. 2706–2716. https://doi.org/10.1002/bit.26389

Gourronc, F. A., Rockey, W. M., Thiel, W. H., Giangrande, P. H., & Klingelhutz, A. J. (2013). Identification of RNA aptamers that internalize into HPV-16 E6/E7 transformed tonsillar epithelial cells. *Virology*, *446*(1–2), 325–333. https://doi.org/10.1016/j.virol.2013.08.015

Grisshammer, R., Duckworth, R., & Henderson, R. (1993). Expression of a rat neurotensin receptor in Escherichia coli. *Biochemical Journal*, *295*(2), 571–576. https://doi.org/10.1042/bj2950571

Grisshammer, Reinhard. (2006, August). Understanding recombinant expression of membrane proteins. *Current Opinion in Biotechnology*, Vol. 17, pp. 337–340. https://doi.org/10.1016/j.copbio.2006.06.001

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Research*, *36*(Web Server issue), W70–W74. https://doi.org/10.1093/nar/gkn188

Guan, N., & Liu, L. (2020). Microbial response to acid stress: mechanisms and applications. *Applied Microbiology and Biotechnology*, *104*(1), 51–65. https://doi.org/10.1007/s00253-019-10226-1

Gul, N., Linares, D. M., Ho, F. Y., & Poolman, B. (2014). Evolved escherichia coli strains for amplified, functional expression of membrane proteins. *Journal of Molecular Biology*, *426*(1), 136–149. https://doi.org/10.1016/j.jmb.2013.09.009

Guo, J., Ma, Z., Gao, J., Zhao, J., Wei, L., Liu, J., & Xu, N. (2019). Recent advances of pH homeostasis mechanisms in Corynebacterium glutamicum. *World Journal of Microbiology and Biotechnology*, *35*(12). https://doi.org/10.1007/s11274-019-2770-2

Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., & Welch, M. (2012). Engineering genes for predictable protein expression. *Protein Expression and Purification*, *83*(1), 37–46. https://doi.org/10.1016/j.pep.2012.02.013

Haisler, W. L., Timm, D. M., Gage, J. A., Tseng, H., Killian, T. C., & Souza, G. R. (2013). Three-dimensional cell culturing by magnetic levitation. *Nature Protocols*, *8*(10), 1940–1949. https://doi.org/10.1038/nprot.2013.125

Hassaine, G., Wagner, R., Kempf, J., Cherouati, N., Hassaine, N., Prual, C., … Lundstrom, K.

(2006). Semliki Forest virus vectors for overexpression of 101 G protein-coupled receptors in mammalian host cells. *Protein Expression and Purification*, *45*(2), 343–351. https://doi.org/10.1016/J.PEP.2005.06.007

He, Y., Wang, K., & Yan, N. (2014). The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein & Cell*, *5*(9), 658–672. https://doi.org/10.1007/s13238-014-0086-4

Hediger, M. A., Clémençon, B., Burrier, R. E., & Bruford, E. A. (2013). The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Molecular Aspects of Medicine*, *34*(2–3), 95–107. https://doi.org/10.1016/j.mam.2012.12.009

Hediger, M. A., Romero, M. F., Peng, J. Bin, Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflugers Archiv European Journal of Physiology*, *447*(5), 465–468. https://doi.org/10.1007/s00424-003-1192-y

Hermann, T., & Patel, D. J. (2000). RNA bulges as architectural and recognition motifs. *Structure*, *8*(3), 47–54. https://doi.org/10.1016/S0969-2126(00)00110-6

Hicke, B. J., Marion, C., Chang, Y. F., Gould, T., Lynott, C. K., Parma, D., … Warren, S. (2001). Tenascin-C Aptamers Are Generated Using Tumor Cells and Purified Protein. *Journal of Biological Chemistry*, *276*(52), 48644–48654. https://doi.org/10.1074/jbc.M104651200

Holbrook, S. R. (2005). RNA structure: The long and the short of it. *Current Opinion in Structural Biology*, *15*(3 SPEC. ISS.), 302–308. https://doi.org/10.1016/j.sbi.2005.04.005

Homann, M., & Göringer, H. U. (2001). Uptake and intracellular transport of RNA aptamers in African trypanosomes suggest therapeutic “Piggy-Back” approach. *Bioorganic and Medicinal Chemistry*, *9*(10), 2571–2580. https://doi.org/10.1016/S0968-0896(01)00032-3

Homann, Matthias, & Göringer, H. U. (1999). Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Research*, *27*(9), 2006–2014. https://doi.org/10.1093/nar/27.9.2006

Hong, K. L., Yancey, K., Battistella, L., Williams, R. M., Hickey, K. M., Bostick, C. D., … Sooter, L. J. (2015). Selection of Single-Stranded DNA Molecular Recognition Elements against Exotoxin A Using a Novel Decoy-SELEX Method and Sensitive Detection of Exotoxin A in Human Serum. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/417641

Huang, C. J., Lin, H. I., Shiesh, S. C., & Lee, G. Bin. (2010). Integrated microfluidic system for rapid screening of CRP aptamers utilizing systematic evolution of ligands by exponential

enrichment (SELEX). *Biosensors and Bioelectronics*, *25*(7), 1761–1766. https://doi.org/10.1016/j.bios.2009.12.029

Huang, Y. Z., Hernandez, F. J., Gu, B., Stockdale, K. R., Nanapaneni, K., Scheetz, T. E., … McNamara, J. O. (2012). RNA aptamer-based functional ligands of the neurotrophin receptor, TrkB. *Molecular Pharmacology*, *82*(4), 623–635. https://doi.org/10.1124/mol.112.078220

Iaboni, M., Fontanella, R., Rienzo, A., Capuozzo, M., Nuzzo, S., Santamaria, G., … Esposito, C. L. (2016). Targeting Insulin Receptor with a Novel Internalizing Aptamer. *Molecular Therapy - Nucleic Acids*, *5*, e365. https://doi.org/10.1038/mtna.2016.73

Ilgü, H., Jeckelmann, J. M., Gachet, M. S., Boggavarapu, R., Ucurum, Z., Gertsch, J., & Fotiadis, D. (2014). Variation of the detergent-binding capacity and phospholipid content of membrane proteins when purified in different detergents. *Biophysical Journal*, *106*(8), 1660–1670. https://doi.org/10.1016/j.bpj.2014.02.024

Ilgü, H., Jeckelmanna, J. M., Gapsys, V., Ucurum, Z., De Grootc, B. L., & Fotiadis, D. (2016). Insights into the molecular basis for substrate binding and specificity of the wild-type L-arginine/agmatine antiporter AdiC. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10358–10363. https://doi.org/10.1073/pnas.1605442113

Ilgu, M., & Nilsen-Hamilton, M. (2016). Aptamers in analytics. *Analyst*, Vol. 141, pp. 1551–1558. https://doi.org/10.1039/c5an01824b

Ilgu, M., Yan, S., Khounlo, R. M., Lamm, M. H., & Nilsen-Hamilton, M. (2019). Common secondary and tertiary structural features of aptamer–ligand interaction shared by RNA aptamers with different primary sequences. *Molecules*, *24*(24). https://doi.org/10.3390/molecules24244535

Iyer, R., Williams, C., & Miller, C. (2003). Arginine-Agmatine Antiporter in Extreme Acid Resistance in Escherichia coli. *Journal of Bacteriology*, *185*(22), 6556–6561. https://doi.org/10.1128/JB.185.22.6556-6561.2003

Jack, D. L., Paulsen, I. T., & Saier, J. (2000). The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology*, *146*(8), 1797–1814. https://doi.org/10.1099/00221287-146-8-1797

Jacoby, K., Lambert, A. R., & Scharenberg, A. M. (2017). Characterization of homing endonuclease binding and cleavage specificities using yeast surface display SELEX (YSD-SELEX). *Nucleic Acids Research*, *45*(3), e11. https://doi.org/10.1093/nar/gkw864

Jacoby, K., & Scharenberg, A. M. (2014). Homing endonuclease target determination using SELEX

adapted for yeast surface display. In *Methods in Molecular Biology* (Vol. 1123, pp. 165–190). https://doi.org/10.1007/978-1-62703-968-0_13

Janas, T., & Janas, T. (2011). The selection of aptamers specific for membrane molecular targets. *Cellular and Molecular Biology Letters*, Vol. 16, pp. 25–39. https://doi.org/10.2478/s11658-010-0023-3

Jang, K. J., Lee, N. R., Yeo, W. S., Jeong, Y. J., & Kim, D. E. (2008). Isolation of inhibitory RNA aptamers against severe acute respiratory syndrome (SARS) coronavirus NTPase/Helicase. *Biochemical and Biophysical Research Communications*, *366*(3), 738–744. https://doi.org/10.1016/j.bbrc.2007.12.020

Jensen, H. M., Eng, T., Chubukov, V., Herbert, R. A., & Mukhopadhyay, A. (2017). Improving membrane protein expression and function using genomic edits. *Scientific Reports*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-12901-7

Jensen, K. B., Atkinson, B. L., Willis, M. C., Koch, T. H., & Gold, L. (1995). Using in vitro selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(26), 12220–12224. https://doi.org/10.1073/pnas.92.26.12220

Jeong, S., Eom, T. Y., Kim, S. J., Lee, S. W., & Yu, J. (2001). In vitro selection of the RNA Aptamer against the Sialyl Lewis X and its inhibition of the cell adhesion. *Biochemical and Biophysical Research Communications*, *281*(1), 237–243. https://doi.org/10.1006/bbrc.2001.4327

Jhaveri, S., Rajendran, M., & Ellington, A. D. (2000). In vitro selection of signaling aptamers. *Nature Biotechnology*, *18*(12), 1293–1297. https://doi.org/10.1038/82414

Jia, W., Ren, C., Wang, L., Zhu, B., Jia, W., Gao, M., … Liu, W. (2016). CD109 is identified as a potential nasopharyngeal carcinoma biomarker using aptamer selected by cell-SELEX. *Oncotarget*, *7*(34), 55328–55342. https://doi.org/10.18632/oncotarget.10530

Johnson, M. (2012). Protein Quantitation. *Materials and Methods*, *2*. https://doi.org/10.13070/mm.en.2.115

Joshi, R., Janagama, H., Dwivedi, H. P., Senthil Kumar, T. M. A., Jaykus, L. A., Schefers, J., & Sreevatsan, S. (2009). Selection, characterization, and application of DNA aptamers for the capture and detection of Salmonella enterica serovars. *Molecular and Cellular Probes*, *23*(1), 20–28. https://doi.org/10.1016/j.mcp.2008.10.006

Kahsai, A. W., Wisler, J. W., Lee, J., Ahn, S., Cahill III, T. J., Dennison, S. M., … Lefkowitz, R. J.

(2016). Conformationally selective RNA aptamers allosterically modulate the β2-adrenoceptor. *Nature Chemical Biology*, *12*(9), 709–716. https://doi.org/10.1038/nchembio.2126

Kandasamy, P., Gyimesi, G., Kanai, Y., & Hediger, M. A. (2018). Amino acid transporters revisited: New views in health and disease. *Trends in Biochemical Sciences*, *43*(10), 752–789. https://doi.org/10.1016/j.tibs.2018.05.003

Kang, H. S., Huh, Y. M., Kim, S., & Lee, D. ki. (2009). Isolation of RNA aptamers targeting HER-2-overexpressing breast cancer cells using cell-SELEX. *Bulletin of the Korean Chemical Society*, *30*(8), 1827–1831. https://doi.org/10.5012/bkcs.2009.30.8.1827

Kanjee, U. (2012). *Structural and Functional Studies on the Escherichia coli Inducible Lysine Decarboxylase : Linking the Acid Stress and Stringent Responses Structural and Functional Studies on the Escherichia coli Inducible Lysine Decarboxylase : Linking the Acid Stress a* (University of Toronto). Retrieved from http://hdl.handle.net/1807/32786

Kanjee, U., & Houry, W. A. (2013). Mechanisms of acid resistance in escherichia coli. *Annual Review of Microbiology*, *67*, 65–81. https://doi.org/10.1146/annurev-micro-092412-155708

Kapałczyńska, M., Kolenda, T., Przybyła, W., Zajączkowska, M., Teresiak, A., Filas, V., … Lamperska, K. (2018). 2D and 3D cell cultures – a comparison of different types of cancer cell cultures. *Archives of Medical Science*, *14*(4), 910–919. https://doi.org/10.5114/aoms.2016.63743

Kaur, H. (2018). Recent developments in cell-SELEX technology for aptamer selection. *Biochimica et Biophysica Acta - General Subjects*, Vol. 1862, pp. 2323–2329. https://doi.org/10.1016/j.bbagen.2018.07.029

Kawakami, J., Imanaka, H., Yokota, Y., & Sugimoto, N. (2000). In vitro selection of aptamers that act with Zn2+. *Journal of Inorganic Biochemistry*, *82*(1–4), 197–206. https://doi.org/10.1016/S0162-0134(00)00158-6

Khati, M., Schüman, M., Ibrahim, J., Sattentau, Q., Gordon, S., & James, W. (2003). Neutralization of Infectivity of Diverse R5 Clinical Isolates of Human Immunodeficiency Virus Type 1 by gp120-Binding 2′F-RNA Aptamers. *Journal of Virology*, *77*(23), 12692–12698. https://doi.org/10.1128/jvi.77.23.12692-12698.2003

Khow, O., & Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine*, *2*(2), 159–162. https://doi.org/10.1016/S2221-1691(11)60213-X

Kim, E. Y., Kim, J. W., Kim, W. K., Han, B. S., Park, S. G., Chung, B. H., … Bae, K.-H. (2014). Selection of Aptamers for Mature White Adipocytes by Cell SELEX Using Flow Cytometry. *PLoS ONE*, *9*(5), e97747. https://doi.org/10.1371/journal.pone.0097747

Kim, H. S., Ernst, J. A., Brown, C., Bostrom, J., Fuh, G., & Lee, C. V. (2012). Translation Levels Control Multi-Spanning Membrane Protein Expression. *PLoS ONE*, *7*(4), 35844. https://doi.org/10.1371/journal.pone.0035844

Kim, J. W., Kim, E. Y., Kim, S. Y., Byun, S. K., Lee, D., Oh, K. J., … Bae, K. H. (2014). Identification of DNA aptamers toward epithelial cell adhesion molecule via Cell-SELEX. *Molecules and Cells*, *37*(10), 742–746. https://doi.org/10.14348/molcells.2014.0208

Kim, Y., Wu, Q., Hamerlik, P., Hitomi, M., Sloan, A. E., Barnett, G. H., … Rich, J. N. (2013). Aptamer identification of brain tumor-initiating cells. *Cancer Research*, *73*(15), 4923–4936. https://doi.org/10.1158/0008-5472.CAN-12-4556

King, M. S., Boes, C., & Kunji, E. R. S. (2015). Membrane protein expression in Lactococcus lactis. In *Methods in Enzymology* (Vol. 556, pp. 77–97). https://doi.org/10.1016/bs.mie.2014.12.009

Klußmann, S., Nolte, A., Bald, R., Erdmann, V. A., & Fürste, J. P. (1996). Mirror-image RNA that binds D-Adenosine. *Nature Biotechnology*, *14*(9), 1112–1115. https://doi.org/10.1038/nbt0996-1112

Koenig, A., Dan O'Connel, M., Jennings, S., Hicke, B., Man, H. L., Fitzwater, T., … Varki, A. (1996). Calcium-dependent oligonucleotide antagonists specific for l-selectin. *Journal of Investigative Medicine*, *44*(3), 5883–5887.

Kolesnikova, O., Kazakova, H., Comte, C., Steinberg, S., Kamenski, P., Martin, R. P., … Entelis, N. (2010). Selection of RNA aptamers imported into yeast and human mitochondria. *RNA*, *16*(5), 926–941. https://doi.org/10.1261/rna.1914110

Komarova, N., & Kuznetsov, A. (2019). Inside the black box: What makes Selex better? *Molecules*, *24*(19). https://doi.org/10.3390/molecules24193598

Kopylov, A. M., & Spiridonova, V. A. (2000). Combinatorial chemistry of nucleic acids: SELEX. *Molecular Biology*, *34*(6), 940–954. https://doi.org/10.1023/A:1026696330245

Kowalczyk, L., Ratera, M., Paladino, A., Bartoccioni, P., Errasti-Murugarren, E., Valencia, E., … Palacín, M. (2011). Molecular basis of substrate-induced permeation by an amino acid antiporter. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(10), 3935–3940. https://doi.org/10.1073/pnas.1018081108

Krammer, E. M., Gibbons, A., Roos, G., & Prévost, M. (2018). Molecular mechanism of substrate selectivity of the arginine-agmatine Antiporter AdiC. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-33963-1

Krammer, E. M., & Prévost, M. (2019). Function and Regulation of Acid Resistance Antiporters. *Journal of Membrane Biology*, *252*(4–5), 465–481. https://doi.org/10.1007/s00232-019-00073-6

Krin, E., Danchin, A., & Soutourina, O. (2010). Decrypting the H-NS-dependent regulatory cascade of acid stress resistance in Escherichia coli. *BMC Microbiology*, *10*. https://doi.org/10.1186/1471-2180-10-273

Kubiczek, D., Raber, H., Bodenberger, N., Oswald, T., Sahan, M., Mayer, D., … Rosenau, F. (2020). The Diversity of a Polyclonal FluCell-SELEX Library Outperforms Individual Aptamers as Emerging Diagnostic Tools for the Identification of Carbapenem Resistant Pseudomonas aeruginosa. *Chemistry - A European Journal*, *26*(64), 14536–14545. https://doi.org/10.1002/chem.202000213

Lai, J. C., & Hong, C. Y. (2014). Magnetic-assisted rapid aptamer selection (MARAS) for generating high-affinity DNA aptamer using rotating magnetic fields. *ACS Combinatorial Science*, *16*(7), 321–327. https://doi.org/10.1021/co5000272

Laing, C., & Schlick, T. (2011). Computational approaches to RNA structure prediction, analysis, and design. *Current Opinion in Structural Biology*, *21*(3), 306–318. https://doi.org/10.1016/j.sbi.2011.03.015

Langdon, W. B., Petke, J., & Lorenz, R. (2018). Evolving better RNAfold structure prediction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10781 LNCS*, 220–236. https://doi.org/10.1007/978-3-319-77553-1_14

Lao, Y. H., Chiang, H. Y., Yang, D. K., Peck, K., & Chen, L. C. (2014). Selection of aptamers targeting the sialic acid receptor of hemagglutinin by epitope-specific SELEX. *Chemical Communications*, *50*(63), 8719–8722. https://doi.org/10.1039/c4cc03116d

Layzer, J. M., & Sullenger, B. A. (2007). Simultaneous generation of aptamers to multiple gamma-carboxyglutamic acid proteins from a focused aptamer library using DeSELEX and convergent selection. *Oligonucleotides*, *17*(1), 1–11. https://doi.org/10.1089/oli.2006.0059

Lecocq, S., Spinella, K., Dubois, B., Lista, S., Hampel, H., & Penner, G. (2018). Aptamers as biomarkers for neurological disorders. Proof of concept in transgenic mice. *PLoS ONE*, *13*(1).

https://doi.org/10.1371/journal.pone.0190212

Lee, S. C., Knowles, T. J., Postis, V. L. G., Jamshad, M., Parslow, R. A., Lin, Y. P., … Dafforn, T. R. (2016). A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols*, *11*(7), 1149–1162. https://doi.org/10.1038/nprot.2016.070

Leviatan, S., Sawada, K., Moriyama, Y., & Nelson, N. (2010). Combinatorial method for overexpression of membrane proteins in Escherichia coli. *Journal of Biological Chemistry*, *285*(31), 23548–23556. https://doi.org/10.1074/jbc.M110.125492

Levine, H. A., & Nilsen-Hamilton, M. (2007). A mathematical analysis of SELEX. *Computational Biology and Chemistry*, *31*(1), 11–35. https://doi.org/10.1016/j.compbiolchem.2006.10.002

Li, S., Xu, H., Ding, H., Huang, Y., Cao, X., Yang, G., … Shao, N. (2009). Identification of an aptamer targeting hnRNP A1 by tissue slide-based SELEX. *Journal of Pathology*, *218*(3), 327–336. https://doi.org/10.1002/path.2543

Lin, L., Yee, S. W., Kim, R. B., & Giacomini, K. M. (2015). SLC transporters as therapeutic targets: Emerging opportunities. *Nature Reviews Drug Discovery*, *14*(8), 543–560. https://doi.org/10.1038/nrd4626

Link, A. J., Skretas, G., Strauch, E.-M., Chari, N. S., & Georgiou, G. (2008). Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli. *Protein Science*, *17*(10), 1857–1863. https://doi.org/10.1110/ps.035980.108

Liu, J., Yang, Y., Hu, B., Ma, Z. Y., Huang, H. P., Yu, Y., … Yang, D. L. (2010). Development of HBsAg-binding aptamers that bind HepG2.2.15 cells via HBV surface antigen. *Virologica Sinica*, *25*(1), 27–35. https://doi.org/10.1007/s12250-010-3091-7

Liu, M., Zhang, B., Li, Z., Wang, Z., Li, S., Liu, H., … He, N. (2020). Precise discrimination of Luminal A breast cancer subtype using an aptamer: In vitro and in vivo. *Nanoscale*, *12*(38), 19689–19701. https://doi.org/10.1039/d0nr03324c

Liu, Yingmiao, Kuan, C.-T., Mi, J., Zhang, X., Clary, B. M., Bigner, D. D., & Sullenger, B. A. (2009). Aptamers selected against the unglycosylated EGFRvIII ectodomain and delivered intracellularly reduce membrane-bound EGFRvIII and induce apoptosis. *Biol. Chem*, *390*, 137–144. https://doi.org/10.1515/BC.2009.022

Liu, Yuping, Tang, H., Lin, Z., & Xu, P. (2015). Mechanisms of acid tolerance in bacteria and prospects in biotechnology and bioremediation. *Biotechnology Advances*, *33*(7), 1484–1492. https://doi.org/10.1016/j.biotechadv.2015.06.001

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 26. https://doi.org/10.1186/1748-7188-6-26

Lou, X., Qian, J., Xiao, Y., Viel, L., Gerdon, A. E., Lagally, E. T., … Soh, H. T. (2009). Micromagnetic selection of aptamers in microfluidic channels. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 2989–2994. https://doi.org/10.1073/pnas.0813135106

Luebke, K. J., & Tinoco, I. (1996). Sequence effects on RNA bulge-induced helix bending and a conserved five-nucleotide bulge from the group I introns. *Biochemistry*, *35*(36), 11677–11684. https://doi.org/10.1021/bi960914r

Lupold, S. E., Hicke, B. J., Lin, Y., & Coffey, D. S. (2002). Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Research*, *62*(14), 4029–4033.

Ma, D., Lu, P., Yan, C., Fan, C., Yin, P., Wang, J., & Shi, Y. (2012). Structure and mechanism of a glutamate-GABA antiporter. *Nature*, *483*(7391), 632–636. https://doi.org/10.1038/nature10917

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., … Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268

Mai, J., Li, X., Zhang, G., Huang, Y., Xu, R., Shen, Q., … Shen, H. (2018). DNA Thioaptamer with Homing Specificity to Lymphoma Bone Marrow Involvement. *Molecular Pharmaceutics*, *15*(5), 1814–1825. https://doi.org/10.1021/acs.molpharmaceut.7b01169

Mallikaratchy, P. (2017). Evolution of complex target SELEX to identify aptamers against mammalian cell-surface antigens. *Molecules*, *22*(2), 215. https://doi.org/10.3390/molecules22020215

Mallikaratchy, P., Zumrut, H., & Ara, N. (2015). Discovery of Biomarkers Using Aptamers Evolved in Cell-SELEX Method. In *Aptamers Selected by Cell-SELEX for Theranostics* (pp. 265–299). https://doi.org/10.1007/978-3-662-46226-3_12

Martell, R. E., Nevins, J. R., & Sullenger, B. A. (2002). Optimizing aptamer activity for gene therapy applications using expression cassette SELEX. *Molecular Therapy*, *6*(1), 30–34. https://doi.org/10.1006/mthe.2002.0624

Masud, M. M., Kuwahara, M., Ozaki, H., & Sawai, H. (2004). Sialyllactose-binding modified DNA

aptamer bearing additional functionality by SELEX. *Bioorganic and Medicinal Chemistry*, *12*(5), 1111–1120. https://doi.org/10.1016/j.bmc.2003.12.009

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7287–7292. https://doi.org/10.1073/pnas.0401799101

Mathews, D. H., Moss, W. N., & Turner, D. H. (2010, December 1). Folding and finding RNA secondary structure. *Cold Spring Harbor Perspectives in Biology*, Vol. 2, p. a003665. https://doi.org/10.1101/cshperspect.a003665

Mathieu, K., Javed, W., Vallet, S., Lesterlin, C., Candusso, M. P., Ding, F., … Orelle, C. (2019). Functionality of membrane proteins overexpressed and purified from E. coli is highly dependent upon the strain. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-39382-0

Mayer, G., Ahmed, M. S. L., Dolf, A., Endl, E., Knolle, P. A., & Famulok, M. (2010). Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures. *Nature Protocols*, *5*(12), 1993–2004. https://doi.org/10.1038/nprot.2010.163

Mende, M., Hopert, A., Wünsche, W., Overhoff, M., Detzer, A., Börngen, K., … Sczakiel, G. (2007). A hexanucleotide selected for increased cellular uptake in cis contains a highly active CpG-motif in human B cells and primary peripheral blood mononuclear cells. *Immunology*, *120*(2), 261–272. https://doi.org/10.1111/j.1365-2567.2006.02497.x

Mendonsa, S. D., & Bowser, M. T. (2004a). In Vitro Evolution of Functional DNA Using Capillary Electrophoresis. *Journal of the American Chemical Society*, *126*(1), 20–21. https://doi.org/10.1021/ja037832s

Mendonsa, S. D., & Bowser, M. T. (2004b). In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. *Analytical Chemistry*, *76*(18), 5387–5392. https://doi.org/10.1021/ac049857v

Mercier, M. C., Dontenwill, M., & Choulier, L. (2017). Selection of nucleic acid aptamers targeting tumor cell-surface protein biomarkers. *Cancers*, Vol. 9, p. 69. https://doi.org/10.3390/cancers9060069

Mi, J., Liu, Y., Rabbani, Z. N., Yang, Z., Urban, J. H., Sullenger, B. A., & Clary, B. M. (2010). In vivo selection of tumor-targeting RNA motifs. *Nature Chemical Biology*, *6*(1), 22–24. https://doi.org/10.1038/nchembio.277

Misono, T. S., & Kumar, P. K. R. (2005). Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. *Analytical Biochemistry*, *342*(2), 312–317. https://doi.org/10.1016/j.ab.2005.04.013

Mohanty, A. K., & Wiener, M. C. (2004). Membrane protein expression and production: Effects of polyhistidine tag length and position. *Protein Expression and Purification*, *33*(2), 311–325. https://doi.org/10.1016/j.pep.2003.10.010

Mohsen, A. M., Khader, A. T., & Ramachandram, D. (2009). Comparison of HSRNAFold and RNAFold algorithms for RNA secondary structure prediction. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 1–6. https://doi.org/10.1109/TENCON.2009.5395933

Molina, D. M., Cornvik, T., Eshaghi, S., Haeggström, J. Z., Nordlund, P., & Sabet, M. I. (2008). Engineering membrane protein overproduction in Escherichia coli. *Protein Science*, *17*(4), 673–680. https://doi.org/10.1110/ps.073242508

Monné, M., Chan, K. W., Slotboom, D.-J., & Kunji, E. R. S. (2005). Functional expression of eukaryotic membrane proteins in Lactococcus lactis. *Protein Science*, *14*(12), 3048–3056. https://doi.org/10.1110/ps.051689905

Moreno, M., Rincón, E., Piñeiro, D., Fernández, G., Domingo, A., Jiménez-Ruíz, A., … González, V. M. (2003). Selection of aptamers against KMP-11 using colloidal gold during the SELEX process. *Biochemical and Biophysical Research Communications*, *308*(2), 214–218. https://doi.org/10.1016/S0006-291X(03)01352-4

Morris, K. N., Jensen, K. B., Julin, C. M., Weil, M., & Gold, L. (1998). High affinity ligands from in vitro selection: Complex targets. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(6), 2902–2907. https://doi.org/10.1073/pnas.95.6.2902

Mortimer, S. A., Kidwell, M. A., & Doudna, J. A. (2014). Insights into RNA structure and function from genome-wide studies. *Nature Reviews Genetics*, *15*(7), 469–479. https://doi.org/10.1038/nrg3681

Mu, Q., Annapragada, A., Srivastava, M., Li, X., Wu, J., Thiviyanathan, V., … Vigneswaran, N. (2016). Conjugate-SELEX: A High-throughput Screening of Thioaptamer-liposomal Nanoparticle Conjugates for Targeted Intracellular Delivery of Anticancer Drugs. *Molecular Therapy - Nucleic Acids*, *5*, e382. https://doi.org/10.1038/mtna.2016.81

Murakami, K., Zhao, J., Yamasaki, K., & Miyagishi, M. (2017). Biochemical and structural features of extracellular vesicle-binding RNA aptamers. *Biomedical Reports*, *6*(6), 615–626.

https://doi.org/10.3892/br.2017.899

Mus-Veteau, I., Demange, P., & Zito, F. (2014). Membrane protein production for structural analysis. In *Membrane Proteins Production for Structural Analysis* (pp. 3–44). https://doi.org/10.1007/978-1-4939-0662-8_1

Nannenga, B. L., & Baneyx, F. (2011). Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. *Protein Science*, *20*(8), 1411–1420. https://doi.org/10.1002/pro.669

Nitsche, A., Kurth, A., Dunkhorst, A., Pänke, O., Sielaff, H., Junge, W., … Kage, A. (2007). One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. *BMC Biotechnology*, *7*. https://doi.org/10.1186/1472-6750-7-48

Nørholm, M. H. H., Light, S., Virkki, M. T. I., Elofsson, A., Von Heijne, G., & Daley, D. O. (2012). Manipulating the genetic code for membrane protein production: What have we learnt so far? *Biochimica et Biophysica Acta - Biomembranes*, *1818*(4), 1091–1096. https://doi.org/10.1016/j.bbamem.2011.08.018

Ohuchi, S. (2012). Cell-Selex technology. *BioResearch Open Access*, Vol. 1, pp. 265–272. https://doi.org/10.1089/biores.2012.0253

Ohuchi, S. P., Ohtsu, T., & Nakamura, Y. (2006). Selection of RNA aptamers against recombinant transforming growth factor-β type III receptor displayed on cell surface. *Biochimie*, *88*(7), 897–904. https://doi.org/10.1016/j.biochi.2006.02.004

Ouellet, E., Foley, J. H., Conway, E. M., & Haynes, C. (2015). Hi-Fi SELEX: A high-fidelity digital-PCR based therapeutic aptamer discovery platform. *Biotechnology and Bioengineering*, *112*(8), 1506–1522. https://doi.org/10.1002/bit.25581

Palomares, L. A., Estrada-Mondaca, S., & Ramírez, O. T. (2004). Production of recombinant proteins: challenges and solutions. *Methods in Molecular Biology (Clifton, N.J.)*, Vol. 267, pp. 15–52. https://doi.org/10.1385/1-59259-774-2:015

Pan, W., Craven, R. C., Qiu, Q., Wilson, C. B., Wills, J. W., Golovine, S., & Wang, J. F. (1995). Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(25), 11509–11513. https://doi.org/10.1073/pnas.92.25.11509

Pan, W., Xin, P., & Clawson, G. A. (2008). Minimal primer and primer-free SELEX protocols for selection of aptamers from random DNA libraries. *BioTechniques*, *44*(3), 351–360. https://doi.org/10.2144/000112689

Pandey, A., Shin, K., Patterson, R. E., Liu, X. Q., & Rainey, J. K. (2016). Current strategies for protein production and purification enabling membrane protein structural biology. *Biochemistry and Cell Biology*, Vol. 94, pp. 507–527. https://doi.org/10.1139/bcb-2015-0143

Park, J. W., Tatavarty, R., Kim, D. W., Jung, H. T., & Gu, M. B. (2012). Immobilization-free screening of aptamers assisted by graphene oxide. *Chemical Communications*, *48*(15), 2071–2073. https://doi.org/10.1039/c2cc16473f

Park, S. M., Ahn, J. Y., Jo, M., Lee, D. K., Lis, J. T., Craighead, H. G., & Kim, S. (2009). Selection and elution of aptamers using nanoporous sol-gel arrays with integrated microheaters. *Lab on a Chip*, *9*(9), 1206–1212. https://doi.org/10.1039/b814993c

Peleg, Y., & Unger, T. (2012). Resolving bottlenecks for recombinant protein expression in E. coli. *Methods in Molecular Biology*, *800*, 173–186. https://doi.org/10.1007/978-1-61779-349-3_12

Peng, L., Stephens, B. J., Bonin, K., Cubicciotti, R., & Guthold, M. (2007). A combined atomic force/fluorescence microscopy technique to select aptamers in a single cycle from a small pool of random oligonucleotides. *Microscopy Research and Technique*, *70*(4), 372–381. https://doi.org/10.1002/jemt.20421

Peselis, A., & Serganov, A. (2014). Structure and function of pseudoknots involved in gene expression control. *Wiley Interdisciplinary Reviews. RNA*, *5*(6), 803–822. https://doi.org/10.1002/wrna.1247

Pestourie, C., Tavitian, B., & Duconge, F. (2005). Aptamers against extracellular targets for in vivo applications. *Biochimie*, *87*(9–10), 921–930. https://doi.org/10.1016/j.biochi.2005.04.013

Pestourie, Carine, Cerchia, L., Gombert, K., Aissouni, Y., Boulay, J., De Franciscis, V., … Ducongé, F. (2006). Comparison of different strategies to select aptamers against a transmembrane protein target. *Oligonucleotides*, *16*(4), 323–335. https://doi.org/10.1089/oli.2006.16.323

Philippou, S., Mastroyiannopoulos, N. P., Makrides, N., Lederer, C. W., Kleanthous, M., & Phylactou, L. A. (2018). Selection and Identification of Skeletal-Muscle-Targeted RNA Aptamers. *Molecular Therapy: Nucleic Acid*, *10*, 199–214. https://doi.org/10.1016/j.omtn.2017.12.004

Plucinsky, S. M., Root, K. T., & Glover, K. J. (2018). Efficient solubilization and purification of highly insoluble membrane proteins expressed as inclusion bodies using perfluorooctanoic acid. *Protein Expression and Purification*, *143*, 34–37. https://doi.org/10.1016/j.pep.2017.10.012

Raddatz, M. S. L., Dolf, A., Endl, E., Knolle, P., Famulok, M., & Mayer, G. (2008). Enrichment of cell-targeting and population-specific aptamers by fluorescence-activated cell sorting. *Angewandte Chemie - International Edition*, *47*(28), 5190–5193. https://doi.org/10.1002/anie.200800216

Rajendran, M., & Ellington, A. D. (2003). In vitro selection of molecular beacons. *Nucleic Acids Research*, *31*(19), 5700–5713. https://doi.org/10.1093/nar/gkg764

Ranches, G., Lukasser, M., Schramek, H., Ploner, A., Stasyk, T., Mayer, G., … Hüttenhofer, A. (2017). In Vitro Selection of Cell-Internalizing DNA Aptamers in a Model System of Inflammatory Kidney Disease. *Molecular Therapy - Nucleic Acids*, *8*, 198–210. https://doi.org/10.1016/j.omtn.2017.06.018

Rasmussen, S. G. F., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., … Kobilka, B. K. (2007). Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature*, *450*(7168), 383–387. https://doi.org/10.1038/nature06325

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., & Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(6), 1760–1765. https://doi.org/10.1073/pnas.0813167106

Reeves, P. J., Kim, J. M., & Khorana, H. G. (2002). Structure and function in rhodopsin: A tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(21), 13413–13418. https://doi.org/10.1073/pnas.212519199

Reig, N., Del Rio, C., Casagrande, F., Ratera, M., Gelpí, J. L., Torrents, D., … Palacín, M. (2007). Functional and structural characterization of the first prokaryotic member of the L-amino acid transporter (LAT) family: A model for APC transporters. *Journal of Biological Chemistry*, *282*(18), 13270–13281. https://doi.org/10.1074/jbc.M610695200

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-129

Ringquist, S., Jones, T., Snyder, E. E., Gibson, T., Boni, I., & Gold, L. (1995). High-Affinity RNA Ligands to Escherichia coli Ribosomes and Ribosomal Protein S1: Comparison of Natural and Unnatural Binding Sites. *Biochemistry*, *34*(11), 3640–3648. https://doi.org/10.1021/bi00011a019

Ritchey, L. E., Su, Z., Tang, Y., Tack, D. C., Assmann, S. M., & Bevilacqua, P. C. (2017).

Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Research*, *45*(14), e135. https://doi.org/10.1093/nar/gkx533

Rivas, E. (2013). The four ingredients of single-sequence RNA secondary structure prediction. A unifying perspective. *RNA Biology*, *10*(7), 1185–1196. https://doi.org/10.4161/rna.24971

Rives, M. L., Javitch, J. A., & Wickenden, A. D. (2017). Potentiating SLC transporter activity: Emerging drug discovery opportunities. *Biochemical Pharmacology*, *135*, 1–11. https://doi.org/10.1016/j.bcp.2017.02.010

Robertson, D. L., & Joyce, G. F. (1990). Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, *344*(6265), 467–468. https://doi.org/10.1038/344467a0

Rockey, W. M., Hernandez, F. J., Huang, S. Y., Cao, S., Howell, C. A., Thomas, G. S., … Giangrande, P. H. (2011). Rational truncation of an RNA aptamer to prostate-specific membrane antigen using computational structural modeling. *Nucleic Acid Therapeutics*, *21*(5), 299–314. https://doi.org/10.1089/nat.2011.0313

Rodriguez, L. F. R., & Cortes-Mancera, F. M. (2013). Evaluation of nine bioinformatic platforms using a viral model for RNA secondary structure prediction. *Symposium of Signals, Images and Artificial Vision - 2013, STSIVA 2013*, 1–5. https://doi.org/10.1109/STSIVA.2013.6644925

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, Vol. 5. https://doi.org/10.3389/fmicb.2014.00172

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019, August 1). New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Science*, Vol. 28, pp. 1412–1422. https://doi.org/10.1002/pro.3668

Rothnie, A. J. (2016). Detergent-free membrane protein purification. In *Methods in Molecular Biology* (Vol. 1432, pp. 261–267). https://doi.org/10.1007/978-1-4939-3637-3_16

Roulet, E., Busso, S., Camargo, A. A., Simpson, A. J. G., Mermod, N., & Bucher, P. (2002). High-throughput SELEX-SAGE method for quantitative modeling of transcription-factor binding sites. *Nature Biotechnology*, *20*(8), 831–835. https://doi.org/10.1038/nbt718

Routledge, S. J., Mikaliunaite, L., Patel, A., Clare, M., Cartwright, S. P., Bawa, Z., … Bill, R. M. (2016). The synthesis of recombinant membrane proteins in yeast for structural studies. *Methods*, *95*, 26–37. https://doi.org/10.1016/j.ymeth.2015.09.027

Ryul, S., & Lee, S.-W. (2013). In Vitro Selection of RNA Aptamer Specific to Salmonella Typhimurium. *J. Microbiol. Biotechnol*, *23*(6), 878–884. https://doi.org/10.4014/jmb.1212.12033

Saida, F., Uzan, M., Odaert, B., & Bontems, F. (2006). Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. *Current Protein and Peptide Science*, *7*(1), 47–56. https://doi.org/10.2174/138920306775474095

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., & Villaverde, A. (2016). Recombinant pharmaceuticals from microbial cells: A 2015 update. *Microbial Cell Factories*, *15*(1). https://doi.org/10.1186/s12934-016-0437-3

Santosh, B., & Yadava, P. K. (2014). Nucleic acid aptamers: Research tools in disease diagnostics and therapeutics. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/540451

Schlegel, S., Hjelm, A., Baumgarten, T., Vikström, D., & De Gier, J. W. (2014, August 1). Bacterial-based membrane protein production. *Biochimica et Biophysica Acta - Molecular Cell Research*, Vol. 1843, pp. 1739–1749. https://doi.org/10.1016/j.bbamcr.2013.10.023

Schweikhard, E. S., & Ziegler, C. M. (2012). Amino acid secondary transporters: Toward a common transport mechanism. In *Current Topics in Membranes* (Vol. 70, pp. 1–28). https://doi.org/10.1016/B978-0-12-394316-3.00001-6

Scientific, T. (2019). Blanking with high absorbing buffers such as RIPA negatively affects Protein A280 measurements. In *Protein quantification and qualification: Application compendium* (pp. 3–5). Retrieved from https://assets.thermofisher.com/TFS-Assets/MSD/Application-Notes/T144-protein-quantification-qualification-ebook.pdf

Sedighian, H., Halabian, R., Amani, J., Heiat, M., Amin, M., & Fooladi, A. A. I. (2018). Staggered Target SELEX, a novel approach to isolate non-cross-reactive aptamer for detection of SEA by apta-qPCR. *Journal of Biotechnology*, *286*, 45–55. https://doi.org/10.1016/j.jbiotec.2018.09.006

Sefah, K., Bae, K.-M., Phillips, J. A., Siemann, D. W., Su, Z., McClellan, S., … Tan, W. (2013). Cell-based selection provides novel molecular probes for cancer stem cells. *International Journal of Cancer*, *132*(11), 2578–2588. https://doi.org/10.1002/ijc.27936

Sefah, K., Yang, Z., Bradley, K. M., Hoshika, S., Jiménez, E., Zhang, L., … Benner, S. A. (2014). In vitro selection with artificial expanded genetic information systems. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1449–1454. https://doi.org/10.1073/pnas.1311778111

Shalaly, N. D., Aneiros, E., Blank, M., Mueller, J., Nyman, E., Blind, M., … Sandberg, K. (2015). Positive modulation of the glycine receptor by means of glycine receptor-binding aptamers. *Journal of Biomolecular Screening*, *20*(9), 1112–1123. https://doi.org/10.1177/1087057115590575

Shamah, S. M., Healy, J. M., & Cload, S. T. (2008). Complex target SELEX. *Accounts of Chemical Research*, *41*(1), 130–138. https://doi.org/10.1021/ar700142z

Shangguan, D., Cao, Z. C., Li, Y., & Tan, W. (2007). Aptamers Evolved from Cultured Cancer Cells Reveal Molecular Differences of Cancer Cells in Patient Samples. *Clinical Chemistry*, *53*(6), 1153–1155. https://doi.org/10.1373/clinchem.2006.083246

Shangguan, D., Li, Y., Tang, Z., Cao, Z. C., Chen, H. W., Mallikaratchy, P., … Tan, W. (2006). Aptamers evolved from live cells as effective molecular probes for cancer study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 11838–11843. https://doi.org/10.1073/pnas.0602615103

Shangguan, D., Tang, Z., Mallikaratchy, P., Xiao, Z., & Tan, W. (2007). Optimization and modifications of aptamers selected from live cancer cell lines. *ChemBioChem*, *8*(6), 603–606. https://doi.org/10.1002/cbic.200600532

Shapiro, B. A., Yingling, Y. G., Kasprzak, W., & Bindewald, E. (2007). Bridging the gap in RNA structure prediction. *Current Opinion in Structural Biology*, *17*(2), 157–165. https://doi.org/10.1016/j.sbi.2007.03.001

Shaw, A. Z., & Miroux, B. (2003). A general approach for heterologous membrane protein expression in Escherichia coli: the uncoupling protein, UCP1, as an example. *Methods in Molecular Biology (Clifton, N.J.)*, *228*, 23–35. https://doi.org/10.1385/1-59259-400-x:23

Shi, H., He, X., Wang, K., Wu, X., Ye, X., Guo, Q., … Zhou, B. (2011). Activatable aptamer probe for contrast-enhanced in vivo cancer imaging based on cell membrane protein-triggered conformation alteration. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(10), 3900–3905. https://doi.org/10.1073/pnas.1016197108

Shigdar, S., Macdonald, J., O'Connor, M., Wang, T., Xiang, D., Shamaileh, H. Al, … Duan, W. (2013). Aptamers as theranostic agents: Modifications, serum stability and functionalisation. *Sensors (Switzerland)*, *13*(10), 13624–13637. https://doi.org/10.3390/s131013624

Shirai, A., Matsuyama, A., Yashiroda, Y., Hashimoto, A., Kawamura, Y., Arai, R., … Yoshida, M. (2008). Global analysis of gel mobility of proteins and its use in target identification. *Journal of Biological Chemistry*, *283*(16), 10745–10752. https://doi.org/10.1074/jbc.M709211200

Shubham, S., Hoinka, J., Banerjee, S., Swanson, E., Dillard, J. A., Lennemann, N. J., … Nilsen-Hamilton, M. (2018). A 2′FY-RNA Motif Defines an Aptamer for Ebolavirus Secreted Protein. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-30590-8

Siller, E., DeZwaan, D. C., Anderson, J. F., Freeman, B. C., & Barral, J. M. (2010). Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. *Journal of Molecular Biology*, *396*(5), 1310–1318. https://doi.org/10.1016/j.jmb.2009.12.042

Singh, J., Hanson, J., Paliwal, K., & Zhou, Y. (2019). RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13395-9

Smith, D., Kirschenheuter, G. P., Charlton, J., Guidot, D. M., & Repine, J. E. (1995). In vitro selection of RNA-based irreversible inhibitors of human neutrophil elastase. *Chemistry and Biology*, *2*(11), 741–750. https://doi.org/10.1016/1074-5521(95)90102-7

Smith, J. D., & Gold, L. (2002). *Transcription-free SELEX (Patent No US 6,387,620)*.

Smith, S. M. (2017). Strategies for the purification of membrane proteins. In *Methods in Molecular Biology* (Vol. 1485, pp. 389–400). https://doi.org/10.1007/978-1-4939-6412-3_21

Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., Keene, J. D., … Weeks, K. M. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10322–10327. https://doi.org/10.1073/pnas.1600008113

Snijder, H. J. A., & Hakulinen, J. (2016). Membrane protein production in E. coli for applications in drug discovery. In *Advances in Experimental Medicine and Biology* (Vol. 896, pp. 59–77). https://doi.org/10.1007/978-3-319-27216-0_5

Sola, M., Menon, A. P., Moreno, B., Meraviglia-Crivelli, D., Soldevilla, M. M., Cartón-García, F., & Pastor, F. (2020). Aptamers Against Live Targets: Is In Vivo SELEX Finally Coming to the Edge? *Molecular Therapy - Nucleic Acids*, *21*, 192–204. https://doi.org/10.1016/j.omtn.2020.05.025

Soldevilla, M. M., Villanueva, H., Casares, N., Lasarte, J. J., Bendandi, M., Inoges, S., … Pastor, F. (2016). MRP1-CD28 bi-specific oligonucleotide aptamers: Target costimulation to drug-resistant melanoma cancer stem cells. *Oncotarget*, *7*(17), 23182–23196. https://doi.org/10.18632/oncotarget.8095

Song, A. A. L., In, L. L. A., Lim, S. H. E., & Rahim, R. A. (2017, April 4). A review on

Lactococcus lactis: From food to factory. *Microbial Cell Factories*, Vol. 16, pp. 1–15. https://doi.org/10.1186/s12934-017-0669-x

Sørensen, H. P. (2010). Towards universal systems for recombinant gene expression. *Microbial Cell Factories*, *9*. https://doi.org/10.1186/1475-2859-9-27

Soundararajan, S., Chen, W., Spicer, E. K., Courtenay-Luck, N., & Fernandes, D. J. (2008). The nucleolin targeting aptamer AS1411 destabilizes Bcl-2 messenger RNA in human breast cancer cells. *Cancer Research*, *68*(7), 2358–2365. https://doi.org/10.1158/0008-5472.CAN-07-5723

Souza, A. G., Marangoni, K., Fujimura, P. T., Alves, P. T., Silva, M. J., Bastos, V. A. F., … Goulart, V. A. (2016). 3D Cell-SELEX: Development of RNA aptamers as molecular probes for PC-3 tumor cell line. *Experimental Cell Research*, *341*(2), 147–156. https://doi.org/10.1016/j.yexcr.2016.01.015

Spill, F., Weinstein, Z. B., Shemirani, A. I., Ho, N., Desai, D., & Zaman, M. H. (2016). Controlling uncertainty in aptamer selection. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(43), 12076–12081. https://doi.org/10.1073/pnas.1605086113

Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J. W., … Chang, H. Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, *519*(7544), 486–490. https://doi.org/10.1038/nature14263

Staple, D. W., & Butcher, S. E. (2005). Pseudoknots: RNA structures with diverse functions. *PLoS Biology*, *3*(6), 0956–0959. https://doi.org/10.1371/journal.pbio.0030213

Stim-Herndon, K. P., Flores, T. M., & Bennett, G. N. (1996). Molecular characterization of adiY, a regulatory gene which affects expression of the biodegradative acid-induced arginine decarboxylase gene (adiA) of Escherichia coli. *Microbiology*, *142*(5), 1311–1320. https://doi.org/10.1099/13500872-142-5-1311

Stoltenburg, R., Reinemann, C., & Strehlitz, B. (2005). FluMag-SELEX as an advantageous method for DNA aptamer selection. *Analytical and Bioanalytical Chemistry*, *383*(1), 83–91. https://doi.org/10.1007/s00216-005-3388-9

Stoltenburg, Regina, Nikolaus, N., & Strehlitz, B. (2012). Capture-SELEX: Selection of DNA aptamers for aminoglycoside antibiotics. *Journal of Analytical Methods in Chemistry*, *1*(1). https://doi.org/10.1155/2012/415697

Stoltenburg, Regina, Reinemann, C., & Strehlitz, B. (2007, October). SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering*, Vol. 24, pp.

381–403. https://doi.org/10.1016/j.bioeng.2007.06.001

Sung, H. J., Kayhan, B., Ben-Yedidia, T., & Arnon, R. (2004). A DNA aptamer prevents influenza infection by blocking the receptor binding region of the viral hemagglutinin. *Journal of Biological Chemistry*, *279*(46), 48410–48419. https://doi.org/10.1074/jbc.M409059200

Szeto, K., Latulippe, D. R., Ozer, A., Pagano, J. M., White, B. S., Shalloway, D., … Craighead, H. G. (2013). RAPID-SELEX for RNA aptamers. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0082667

Takahashi, M. (2018). Aptamers targeting cell surface proteins. *Biochimie*, *145*, 63–72. https://doi.org/10.1016/j.biochi.2017.11.019

Takahashi, M., Sakota, E., & Nakamura, Y. (2016). The efficient cell-SELEX strategy, Icell-SELEX, using isogenic cell lines for selection and counter-selection to generate RNA aptamers to cell surface proteins. *Biochimie*, *131*, 77–84. https://doi.org/10.1016/j.biochi.2016.09.018

Tan, S. Y., Acquah, C., Sidhu, A., Ongkudon, C. M., Yon, L. S., & Danquah, M. K. (2016). SELEX Modifications and Bioanalytical Techniques for Aptamer–Target Binding Characterization. *Critical Reviews in Analytical Chemistry*, *46*(6), 521–537. https://doi.org/10.1080/10408347.2016.1157014

Tate, C. G. (2001). Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Letters*, *504*(3), 94–98. https://doi.org/10.1016/S0014-5793(01)02711-9

Tate, C. G., & Grisshammer, R. (1995). Overexpression of Integral Membrane Proteins for Structural Studies. *Quarterly Reviews of Biophysics*, *28*(3), 315–422. https://doi.org/10.1017/S0033583500003504

Tate, Christopher G, Whiteley, E., & Betenbaugh, M. J. (1999). Molecular chaperones stimulate the functional expression of the cocaine- sensitive serotonin transporter. *Journal of Biological Chemistry*, *274*(25), 17551–17558. https://doi.org/10.1074/jbc.274.25.17551

Thiel, K. W., Hernandez, L. I., Dassie, J. P., Thiel, W. H., Liu, X., Stockdale, K. R., … Giangrande, P. H. (2012). Delivery of chemo-sensitizing siRNAs to HER2+-breast cancer cells using RNA aptamers. *Nucleic Acids Research*, *40*(13), 6319–6337. https://doi.org/10.1093/nar/gks294

Thiel, W. H., Thiel, K. W., Flenker, K. S., Bair, T., Dupuy, A. J., Mc namara, J. O., … H. Giangrande, P. (2015). Cell-internalization SELEX: method for identifying cell- internalizing RNA aptamers for delivering siRNAs to target cells. *Methods in Molecular Biology*, *1218*, 187–199. https://doi.org/10.1007/978-1-4939-1538-5_11

Thomas, J., & Tate, C. G. (2014). Quality control in eukaryotic membrane protein overproduction. *Journal of Molecular Biology*, *426*(24), 4139–4154. https://doi.org/10.1016/j.jmb.2014.10.012

Tok, J. B. H., & Fischer, N. O. (2008). Single microbead SELEX for efficient ssDNA aptamer generation against botulinum neurotoxin. *Chemical Communications*, (16), 1883–1885. https://doi.org/10.1039/b717936g

Tolle, F., Brändle, G. M., Matzner, D., & Mayer, G. (2015). A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angewandte Chemie - International Edition*, *54*(37), 10971–10974. https://doi.org/10.1002/anie.201503652

Trujillo, C. A., Majumder, P., Gonzalez, F. A., Moaddel, R., & Ulrich, H. (2007). Immobilized P2X2 purinergic receptor stationary phase for chromatographic determination of pharmacological properties and drug screening. *Journal of Pharmaceutical and Biomedical Analysis*, *44*(3 SPEC. ISS.), 701–710. https://doi.org/10.1016/j.jpba.2007.03.006

Tsai, M. F., Fang, Y., & Miller, C. (2012). Sided functions of an arginine-agmatine antiporter oriented in liposomes. *Biochemistry*, *51*(8), 1577–1585. https://doi.org/10.1021/bi201897t

Tsai, R. Y. L., & Reed, R. R. (1998). Identification of DNA Recognition Sequences and Protein Interaction Domains of the Multiple-Zn-Finger Protein Roaz. *Molecular and Cellular Biology*, *18*(11), 6447–6456. https://doi.org/10.1128/mcb.18.11.6447

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*(4968), 505–510. https://doi.org/10.1126/science.2200121

Ulrich, H., Ippolito, J. E., Pagán, O. R., Eterović, V. A., Hann, R. M., Shi, H., … Hess, G. P. (1998). In vitro selection of RNA molecules that displace cocaine from the membrane-bound nicotinic acetylcholine receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14051–14056. https://doi.org/10.1073/pnas.95.24.14051

Vant-Hull, B., Payano-Baez, A., Davis, R. H., & Gold, L. (1998). The mathematics of SELEX against complex targets. *Journal of Molecular Biology*, *278*(3), 579–597. https://doi.org/10.1006/jmbi.1998.1727

Vater, A., Jarosch, F., Buchner, K., & Klussmann, S. (2003). Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX. *Nucleic Acids Research*, *31*(21). https://doi.org/10.1093/nar/gng130

Vishwakarma, P., Banerjee, A., Pasrija, R., Prasad, R., & Lynn, A. M. (2018). Phylogenetic and

conservation analyses of MFS transporters. *3 Biotech*, *8*(11). https://doi.org/10.1007/s13205-018-1476-8

Wagner, S., Bader, M. L., Drew, D., & de Gier, J. W. (2006). Rationalizing membrane protein overexpression. *Trends in Biotechnology*, Vol. 24, pp. 364–371. https://doi.org/10.1016/j.tibtech.2006.06.008

Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., … De Gier, J. W. (2008). Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14371–14376. https://doi.org/10.1073/pnas.0804090105

Wang, C. W., Chung, W. H., Cheng, Y. F., Ying, N. W., Peck, K., Chen, Y. T., & Hung, S. I. (2013). A new nucleic acid-based agent inhibits cytotoxic T lymphocyte-mediated immune disorders. *Journal of Allergy and Clinical Immunology*, *132*(3), 713-722.e11. https://doi.org/10.1016/j.jaci.2013.04.036

Wang, Hanlu, Zhang, Y., Yang, H., Qin, M., Ding, X., Liu, R., & Jiang, Y. (2018). In Vivo SELEX of an Inhibitory NSCLC-Specific RNA Aptamer from PEGylated RNA Library. *Molecular Therapy - Nucleic Acids*, *10*, 187–198. https://doi.org/10.1016/j.omtn.2017.12.003

Wang, Hongyu, Li, X., Volk, D. E., Lokesh, G. L. R., Elizondo-Riojas, M. A., Li, L., … Gorenstein, D. G. (2016). Morph-X-select: Morphology-based tissue aptamer selection for ovarian cancer biomarker discovery. *BioTechniques*, *61*(5), 249–259. https://doi.org/10.2144/000114473

Wang, J., & Li, G. (2011). Aptamers Against Cell Surface Receptors: Selection, Modification and Application. *Current Medicinal Chemistry*, *18*(27), 4107–4116. https://doi.org/10.2174/092986711797189628

Wang, J, Jiang, H., & Liu, F. (2000). In vitro selection of novel RNA ligands that bind human cytomegalovirus and block viral infection. *RNA (New York, N.Y.)*, *6*(4), 571–583. https://doi.org/10.1017/s1355838200992215

Wang, Jinpeng, Gong, Q., Maheshwari, N., Eisenstein, M., Arcila, M. L., Kosik, K. S., & Soh, H. T. (2014). Particle display: A quantitative screening method for generating high-affinity aptamers. *Angewandte Chemie - International Edition*, *53*(19), 4796–4801. https://doi.org/10.1002/anie.201309334

Wang, Jinpeng, Rudzinski, J. F., Gong, Q., Soh, H. T., & Atzberger, P. J. (2012). Influence of Target Concentration and Background Binding on In Vitro Selection of Affinity Reagents.

PLoS ONE, *7*(8), e43940. https://doi.org/10.1371/journal.pone.0043940

Wang, S., Yan, R., Zhang, X., Chu, Q., & Shi, Y. (2014). Molecular mechanism of pH-dependent substrate transport by an arginine-agmatine antiporter. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(35), 12734–12739. https://doi.org/10.1073/pnas.1414093111

Wang, T., Chen, C., Larcher, L. M., Barrero, R. A., & Veedu, R. N. (2019). Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances*, Vol. 37, pp. 28–50. https://doi.org/10.1016/j.biotechadv.2018.11.001

Wang, T., Yin, W., AlShamaileh, H., Zhang, Y., Tran, P. H. L., Nguyen, T. N. G., … Duan, W. (2019). A Detailed Protein-SELEX Protocol Allowing Visual Assessments of Individual Steps for a High Success Rate. *Human Gene Therapy Methods*, *30*(1), 1–16. https://doi.org/10.1089/hgtb.2018.237

Washietl, S., Hofacker, I. L., Stadler, P. F., & Kellis, M. (2012). RNA folding with soft constraints: Reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Research*, *40*(10), 4261–4272. https://doi.org/10.1093/nar/gks009

Waybrant, B., Pearce, T. R., Wang, P., Sreevatsan, S., & Kokkoli, E. (2012). Development and characterization of an aptamer binding ligand of fractalkine using domain targeted SELEX. *Chemical Communications*, *48*(80), 10043–10045. https://doi.org/10.1039/c2cc34217k

Weeks, K. M. (2015). Review toward all RNA structures, concisely. *Biopolymers*, *103*(8), 438–448. https://doi.org/10.1002/bip.22601

Weiß, H. M., & Grisshammer, R. (2002). Purification and characterization of the human adenosine A2a receptor functionally expressed in Escherichia coli. *European Journal of Biochemistry*, *269*(1), 82–92. https://doi.org/10.1046/j.0014-2956.2002.02618.x

Wen, J. Der, & Gray, D. M. (2004). Selection of genomic sequences that bind tightly to Ff gene 5 protein: primer-free genomic SELEX. *Nucleic Acids Research*, *32*(22). https://doi.org/10.1093/nar/gnh179

White, R., Rusconi, C., Scardino, E., Wolberg, A., Lawson, J., Hoffman, M., & Sullenger, B. (2001). Generation of species cross-reactive aptamers using “toggle” SELEX. *Molecular Therapy*, *4*(6), 567–573. https://doi.org/10.1006/mthe.2001.0495

Wilner, S. E., Wengerter, B., Maier, K., Borba Magalhães, M. D. L., Del Amo, D. S., Pai, S., … Levy, M. (2012). An RNA alternative to human transferrin: A new tool for targeting human

cells. *Molecular Therapy - Nucleic Acids*, *1*(5), e21. https://doi.org/10.1038/mtna.2012.14

Wingfield, P. T. (2015). Overview of the purification of recombinant proteins. *Current Protocols in Protein Science*, *2015*, 6.1.1-6.1.35. https://doi.org/10.1002/0471140864.ps0601s80

Wong, F. H., Chen, J. S., Reddy, V., Day, J. L., Shlykov, M. A., Wakabayashi, S. T., & Saier, M. H. (2012). The amino acid-polyamine-organocation superfamily. *Journal of Molecular Microbiology and Biotechnology*, *22*(2), 105–113. https://doi.org/10.1159/000338542

Wu, C. C. N., Castro, J. E., Motta, M., Cottam, H. B., Kyburz, D., Kipps, T. J., … Carson, D. A. (2003). Selection of Oligonucleotide Aptamers with Enhanced Uptake and Activation of Human Leukemia B Cells. *Human Gene Therapy*, *14*(9), 849–860. https://doi.org/10.1089/104303403765701141

Wu, L., & Curran, J. F. (1999). An allosteric synthetic DNA. *Nucleic Acids Research*, *27*(6), 1512–1516. https://doi.org/10.1093/nar/27.6.1512

Xayaphoummine, A., Bucher, T., & Isambert, H. (2005). Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Research*, *33*(SUPPL. 2), W605. https://doi.org/10.1093/nar/gki447

Xiao, S., Shiloach, J., & Betenbaugh, M. J. (2014). Engineering cells to improve protein expression. *Current Opinion in Structural Biology*, Vol. 26, pp. 32–38. https://doi.org/10.1016/j.sbi.2014.03.005

Xiao, Z., Shangguan, D., Cao, Z., Fang, X., & Tan, W. (2008). Cell-specific internalization study of an aptamer from whole cell selection. *Chemistry - A European Journal*, *14*(6), 1769–1775. https://doi.org/10.1002/chem.200701330

Yan, J., Xiong, H., Cai, S., Wen, N., He, Q., Liu, Y., … Liu, Z. (2019). Advances in aptamer screening technologies. *Talanta*, *200*, 124–144. https://doi.org/10.1016/j.talanta.2019.03.015

Yang, M., Jiang, G., Li, W., Qiu, K., Zhang, M., Carter, C. M., … Li, Y. (2014). Developing aptamer probes for acute myelogenous leukemia detection and surface protein biomarker discovery. *Journal of Hematology and Oncology*, *7*(1), 5. https://doi.org/10.1186/1756-8722-7-5

Yohannes, E., Barnhart, D. M., & Slonczewski, J. L. (2004). pH-Dependent Catabolic Protein Expression during Anaerobic Growth of Escherichia coli K-12. *Journal of Bacteriology*, *186*(1), 192–199. https://doi.org/10.1128/JB.186.1.192-199.2004

Yu, F., Li, H., Sun, W., Zhao, Y., Xu, D., & He, F. (2019). Selection of aptamers against

Lactoferrin based on silver enhanced and fluorescence-activated cell sorting. *Talanta*, *193*, 110–117. https://doi.org/10.1016/j.talanta.2018.09.063

Zhang, Y., Lai, B. S., & Juhas, M. (2019). Recent advances in aptamer discovery and applications. *Molecules*, Vol. 24. https://doi.org/10.3390/molecules24050941

Zhong, W., Pu, Y., Tan, W., Liu, J., Liao, J., Liu, B., … Liu, H. (2019). Identification and Application of an Aptamer Targeting Papillary Thyroid Carcinoma Using Tissue-SELEX. *Analytical Chemistry*, *91*(13), 8289–8297. https://doi.org/10.1021/acs.analchem.9b01000

Zhou, J., & Rossi, J. J. (2014, June 17). Cell-type-specific, aptamer-functionalized agents for targeted disease therapy. *Molecular Therapy - Nucleic Acids*, Vol. 3. https://doi.org/10.1038/mtna.2014.21

Zhu, G., Zhang, H., Jacobson, O., Wang, Z., Chen, H., Yang, X., … Chen, X. (2017). Combinatorial Screening of DNA Aptamers for Molecular Imaging of HER2 in Cancer. *Bioconjugate Chemistry*, *28*(4), 1068–1075. https://doi.org/10.1021/acs.bioconjchem.6b00746

Zhu, Z., Song, Y., Li, C., Zou, Y., Zhu, L., An, Y., & Yang, C. J. (2014). Monoclonal surface display SELEX for simple, rapid, efficient, and cost-effective aptamer enrichment and identification. *Analytical Chemistry*, *86*(12), 5881–5888. https://doi.org/10.1021/ac501423g

Zhuo, Z., Yu, Y., Wang, M., Li, J., Zhang, Z., Liu, J., … Zhang, B. (2017). Recent advances in SELEX technology and aptamer applications in biomedicine. *International Journal of Molecular Sciences*, *18*(10). https://doi.org/10.3390/ijms18102142

Zok, T., Antczak, M., Zurkowski, M., Popenda, M., Blazewicz, J., Adamiak, R. W., & Szachniuk, M. (2018). RNApdbee 2.0: Multifunctional tool for RNA structure annotation. *Nucleic Acids Research*, *46*(W1), W30–W35. https://doi.org/10.1093/nar/gky314

Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S., & Rouskin, S. (2016). DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nature Methods*, *14*(1), 75–82. https://doi.org/10.1038/nmeth.4057

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595

Zulkeflee Sabri, M., Azzar Abdul Hamid, A., Mariam Sayed Hitam, S., & Zulkhairi Abdul Rahim, M. (2019). In-Silico Selection of Aptamer: A Review on the Revolutionary Approach to Understand the Aptamer Design and Interaction Through Computational Chemistry. *Materials Today: Proceedings*, *19*, 1572–1581. https://doi.org/10.1016/j.matpr.2019.11.185

Zumrut, H. E., Ara, M. N., Fraile, M., Maio, G., & Mallikaratchy, P. (2016). Ligand-guided selection of target-specific aptamers: A screening technology for identifying specific aptamers against cell-surface proteins. *Nucleic Acid Therapeutics*, *26*(3), 190–198. https://doi.org/10.1089/nat.2016.0611

Zumrut, H. E., Ara, M. N., Maio, G. E., Van, N. A., Batool, S., & Mallikaratchy, P. R. (2016). Ligand-guided selection of aptamers against T-cell Receptor-cluster of differentiation 3 (TCR-CD3) expressed on Jurkat.E6 cells. *Analytical Biochemistry*, *512*, 1–7. https://doi.org/10.1016/j.ab.2016.08.007

Zumrut, H. E., & Mallikaratchy, P. R. (2020). Ligand Guided Selection (LIGS) of Artificial Nucleic Acid Ligands against Cell Surface Targets. *ACS Applied Bio Materials*, *3*(5), 2545–2552. https://doi.org/10.1021/acsabm.9b00938

Zumrut, H., Yang, Z., Williams, N., Arizala, J., Batool, S., Benner, S. A., & Mallikaratchy, P. (2020). Ligand-Guided Selection with Artificially Expanded Genetic Information Systems against TCR-CD3ϵ. *Biochemistry*, *59*(4), 552–562. https://doi.org/10.1021/acs.biochem.9b00919

Zweers, J. C., Barák, I., Becher, D., Driessen, A. J. M., Hecker, M., Kontinen, V. P., … van Dijl, J. M. (2008). Towards the development of Bacillus subtilis as a cell factory for membrane proteins and protein complexes. *Microbial Cell Factories*, *7*(1), 1–20. https://doi.org/10.1186/1475-2859-7-10

Zweers, J. C., Wiegert, T., & Van Dijl, J. M. (2009). Stress-responsive systems set specific limits to the overproduction of membrane proteins in Bacillus subtilis. *Applied and Environmental Microbiology*, *75*(23), 7356–7364. https://doi.org/10.1128/AEM.01560-09

Zykovich, A., Korf, I., & Segal, D. J. (2009). Bind-n-Seq: High-throughput analysis of in vitro protein DNA interactions using massively parallel sequencing. *Nucleic Acids Research*, *37*(22). https://doi.org/10.1093/nar/gkp802

APPENDICES

A. Chronological Compilation of More Selection Methods

SELEX Method Variant	Novelty
Invention of SELEX	The *in vitro* selection method was devised concurrently by three independent laboratories with RNA (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990). Shortly after, DNA was investigated as an aptamer for purified human thrombin (Bock, Griffin, Latham, Vermaas, & Toole, 1992).
Blended SELEX	An inhibitor of human neutrophil elastase was coupled (blended) to a SELEX library and *in vitro* selection had led to enhancement of the inhibitor (D. Smith, Kirschenheuter, Charlton, Guidot, & Repine, 1995).
(photo)crosslinking SELEX/ photoSELEX/ covalent SELEX	Selection against Rev of HIV-1 using a library substituted with a photoreactive chromophore. RNA sequences that bind and covalently photocrosslink to Rev protein were isolated upon irradiation with long-wavelength UV laser light (K. B. Jensen, Atkinson, Willis, Koch, & Gold, 1995; Kopylov & Spiridonova, 2000).
Genomic SELEX/ cDNA-SELEX	Elucidiation of meaningful biological molecular interactions by SELEX that uses libraries (and targets) generated from the organism's genome (Aquino-Jarquin & Toscano-Garibay, 2011; Dobbelstein & Shenk, 1995; Regina Stoltenburg et al., 2007).
Spiegelmer® Technology or mirror-image SELEX	Highly serum-stable mirror-image RNA (L-RNA) ligands were selected against D-adenosine with 9000-fold greater affinity for D-adenosine compared to the L form (Klußmann, Nolte, Bald, Erdmann, & Fürste, 1996).
Magnetic bead (MB)-based SELEX/ Mag-SELEX	DNA ligands were selected against chloroaromatic chemicals (TCA and PCP) and affinity separation was done using magnetic microbeads (Bruno, 1997; Yan et al., 2019).
Chimeric SELEX	Sequence recombination between previously selected aptamers. Reselection of the chimerae yielded bi-functional aptamers that bind multiple targets (Burke & Willis, 1998).
EMSA-SELEX	Selection using electrophoretic mobility shift assay (EMSA) against the zinc finger protein Roaz purified as a fusion protein with GST (Goto et al., 2017; R. Y. L. Tsai & Reed, 1998).
Multi-stage SELEX	Selection of DNA aptamers against Cibacron blue or cholic acid and isolation of allosterically active aptamers (L. Wu & Curran, 1999).

Appendix A (Cont'd)

Indirect SELEX	Selection of Zn^{2+}-dependent aptamers that bind tightly to purified HIV-1 Tat only in the presence of zinc ions (Kawakami, Imanaka, Yokota, & Sugimoto, 2000).
Truncation SELEX	Participation of fixed sequences in selection is minimized or eliminated; Patented (Gold, Javornik, Pagratis, & Shtatland, 2000).
Signaling aptamers/ Molecular beacons	Selection of aptamers that transduce a signal (e.g. fluorescent) by structure change for applications in biosensors (Jhaveri, Rajendran, & Ellington, 2000; Rajendran & Ellington, 2003).
Toggle-SELEX	RNA aptamers were selected to bind both human and procine thrombin with high affinity by "toggling" the protein target between human and porcine thrombin at each alternating SELEX round (White et al., 2001).
Transcription-free SELEX	Instead of using transcription, nucleic acid ligands are assembled by annealing synthetic nucleic acid fragments to templates and ligating the fragments together. Thus, SELEX using nucleic acids with more diverse chemistries is achieved than was allowed by RNA polymerase; Patented (J. D. Smith & Gold, 2002).
Expression Casette SELEX	Reselection of RNA aptamers after their insertion into a tRNA expression cassette yielded tRNA-aptamer chimeras that had retained functionality and allowed high-level expression in mammalian cells (Martell, Nevins, & Sullenger, 2002).
SELEX-SAGE	SELEX and serial analysis of gene expression (SAGE) protocols are combined together with bioinformatically driven methods to generate accurate, quantitative models of the binding-site of transcription factors (Roulet et al., 2002).
SPR-SELEX	Selection of 2′F-RNA aptamers against the gp120 of HIV-1 using an SPR system (Khati et al., 2003). SPR was explored further for monitoring and enrichment aptamers against hemagglutinin (HA) of human influenza virus and for consensus sequence identification (Misono & Kumar, 2005).
Tailored-SELEX	A selection strategy that identifies short high-affinity aptamers/spiegelmers. It uses customized primers/adapters added by ligation before and removed within the amplification process (Vater, Jarosch, Buchner, & Klussmann, 2003).
Yeast Genetic Selection	Yeast genetic selections were used to optimize the *in vivo* binding and activity of an RNA aptamer previously selected against transciption factor NF- κB (Cassiday & Maher, 2003).

Appendix A (Cont'd)

Capillary Electrophoresis (CE)-SELEX	Capillary electrophoresis technique was integrated with the SELEX process and the procedure yielded high-affinity aptamers against immunoglobulin E only after a few number of rounds (Mendonsa & Bowser, 2004a, 2004b).
Primer-free (PF)/ minimal primer (MP) genomic SELEX	Primer-annealing regions are removed/reduced before selection, and regenerated back before amplification of selected genomic fragments to eliminate selection artifcats of genomic libraries (Pan, Xin, & Clawson, 2008; Wen & Gray, 2004).
On-Chip Selection	A novel method that used on-chip selection combined with a method for point mutations rapidly selected resorufin-binding DNA aptamers (Asai, Nishimura, Aita, & Takahashi, 2004).
FluMag-SELEX	The SELEX method was modified by using fluorescent labels to quantify DNA, and using magnetic beads for target immobilization (R. Stoltenburg, Reinemann, & Strehlitz, 2005).
Non-SELEX/ NECEEM-based non-SELEX	A selection process involving repetitive partitioning steps with no amplification. Partitioning is done using a highly efficient affinity method called NECEEM (non-equilibrium capillary electrophoresis of equilibrium mixtures) (Berezovski, Musheev, Drabovich, & Krylov, 2006).
MonoLEX	A single-round selection method produced high-affinity DNA aptamers against Vaccinia virus used as a model for complex target. MonoLEX comprises a single affinity chromatography step, physical segmentation of affinity resin, and PCR amplification of bound aptamers (Nitsche et al., 2007).
NanoSelection®/ nanoManipulator atomic force microscopy (nM-AFM) selection	Selection method that isolates aptamers in a single cycle by utilizing a combined atomic force microscope (AFM) /fluorescence microscope and a small copy-number PCR (Peng, Stephens, Bonin, Cubicciotti, & Guthold, 2007). Refer to this review for more one-round SELEX methods (Darmostuk et al., 2015).
DeSELEX and Convergent (or focused library) selection	A deconvoluting selection strategy that allows isolation of aptamers for multiple protein targets in a complex mixture at different concentrations (abundant and less abundant proteins). DeSELEX redirects selection to the less abundant targets and convergent selection reveals their aptamers (Layzer & Sullenger, 2007).
Single microbead SELEX	A single target-conjugated microbead is used in selection instead of the traditional use of thousands of beads. A ssDNA library was exposed to a single micrbead to select high affinity aptamers against botulinim neurotoxin (Tok & Fischer, 2008).

Appendix A (Cont'd)

Sol-gel SELEX	Chip-based selection method where sol-gel protein arrays used in a microfluidic system are used for efficient selection of RNA aptamers against multiple targets molecules (S. M. Park et al., 2009).
Microfluidic (M)-SELEX	Selection using microfluidic systems that integrate multiple functions like sample preparation, reaction, separation and detection, all on a single chip, to achieve efficient selection (Aquino-Jarquin & Toscano-Garibay, 2011; Cho et al., 2010; C. J. Huang, Lin, Shiesh, & Lee, 2010; Lou et al., 2009).
High-throughput SELEX	Newer variants of SELEX-SAGE wherein multiplexed and massively parallel sequencing (high-throughput sequencing) technologies are incorporated into the selection process together with bioinformatical analyses for extremely efficient and informative selections (Darmostuk et al., 2015; Yan et al., 2019; Zykovich, Korf, & Segal, 2009).
SOMAmer SELEX	A new aptamer class known as Slow Off-rate Modified Aptamers (SOMAmers) enabled selection of high-affinity aptamers against difficult-to-target proteins and allowed devlopment of a novel highly-multiplexed assay for high-performance proteomic studies (Gold et al., 2010).
Domain targeted SELEX	High-affinity aptamers were selected to specifically target the chemokine domain of the cell-surface protein fractalkine over other nontarget domains. The modified SELEX procedure however, uses denaturation of the target domain (Waybrant, Pearce, Wang, Sreevatsan, & Kokkoli, 2012).
GO-SELEX	An immobilization-free SELEX method used graphene oxide (GO) for simple and efficient separation of target-bound and unbound ssDNA by the difference in adsorption to GO (J. W. Park, Tatavarty, Kim, Jung, & Gu, 2012).
MAI-SELEX	A multivalent aptamer isolation (MAI) SELEX technique used to select pairs of aptamers where each aptamer in a pair can recognize different binding sites on a single target (Q. Gong et al., 2012).
Capture-SELEX	The method bypasses the difficulty in target immobilization of small molecules by immobilizing the SELEX library instead. This was dhown by selecting a DNA aptamer for the aminoglycoside antibiotic kanamycin A (Regina Stoltenburg, Nikolaus, & Strehlitz, 2012).
RAPID-SELEX	RNA Aptamer Isolation via Dual-cycles (RAPID)-SELEX is a method combining the efficiency of non-SELEX and robustness of conventional SELEX. RAPID-SELEX simplifies selection by skipping unnecessary amplification steps (Szeto et al., 2013).

Appendix A (Cont'd)

Quantiative Parallel Aptamer Selection System (QPASS)	An avenue for high-throughput aptamer discovery and characterization. QPASS integrates microfluidic selection and next-generation sequencing with aptamer arrays synthesized in situ, thereby enabling accelerated screening of thousands of aptamers simultaneously (Cho et al., 2013).
Yeast Surface Display (YSD)-SELEX	Yeast surface display and SELEX technologies were combined for the easy and high-throughput identification of oligos binding to a displayed protein. Precisely, cleavable protein-binding DNAs were identified *de novo* for displayed homing endonucleases (Jacoby, Lambert, & Scharenberg, 2017; Jacoby & Scharenberg, 2014).
Particle Display selection/ Monoclonal Surface Display (MSD)-SELEX/ Aptamer Particles (APs)-SELEX	An aptamer library is transformed into a library of aptamer particles by emulsion PCR (ePCR) with each particle displaying many copies of a distinct sequence. High-affinity aptamers can be identified by FACS (Jinpeng Wang et al., 2014; Z. Zhu et al., 2014).
AEGIS-SELEX	Artificially expanded genetic information systems (AEGISs) oligonucleotides are used in a SELEX experiment and a high-affinity AEGIS aptamer (containing the standard G, A, C, and T, and the nonstandard AEGIS nucleotides P and Z) was produced (Sefah et al., 2014).
ES-SELEX	Epitope-specific (ES) selection was demonstrated with HA from the influenza virus used as a model protein. The sialic acid receptor (SAR) of native HA (not denatured) was targeted and SAR-specific rather than HA-specific aptamers were selected by eluting oligos with a SAR-binding known, specific ligand competitor (Lao, Chiang, Yang, Peck, & Chen, 2014).
MARAS	Magnetic-assisted rapid aptamer selection (MARAS) uses magnetic beads and a rotating magnetic field applied externally to provide the competitive mechanism for selection. The binding affinities of selected aptamers can be varied by varying the magnetic field frequency (Lai & Hong, 2014).
Hi-Fidelity (Hi-Fi) SELEX	A platform for efficient aptamer discovery. It safeguards the functional diversity of a library by using fixed-rergion blocking elements and it efficiently excludes nonspecific aptamers. Droplet-digital PCR is used for high-fidelity amplification of selected candidates by elimination of amplification artifcacts (Ouellet, Foley, Conway, & Haynes, 2015).
Click-SELEX	Selection is performed using libraries modified by click chemistry to access a wider spectrum of targets and aptamer functionalities (Tolle, Brändle, Matzner, & Mayer, 2015).

Appendix A (Cont'd)

Decoy-SELEX	A method that emphasizes on the selection against multiple negative targets. In its first report, the method was used to select a ssDNA molecular recognition element specific to Exotoxin A. Negative selection against bovine serum albumin, Cholera toxin, streptavidin, and biotin were carried out (Hong et al., 2015).
Conjugate-SELEX	An improvement over traditional cell-SELEX whereby targeted drug delivery vehicles are identified from libraries of nanoparticle-aptamer conjugates instead of typical free aptamer libraries. Aptamers facilitating efficient internalization of liposomal nanoparticles with payloads were identified (Mu et al., 2016).
Automated SELEX	A general term that describes automation of the selection process (e.g. binding, partitioning, elution, amplification, conditioning etc.) by various means such as robotic and microfluidic systems (Regina Stoltenburg et al., 2007).
In silico Selection	The collective of computational methods to aptamer design and screening such as molecular docking and molecular modeling simulations and the approaches to the prediction of 3D structures and thermodynamic parameters and elucidation of aptamer-target interactions (Darmostuk et al., 2015; Zulkeflee Sabri, Azzar Abdul Hamid, Mariam Sayed Hitam, & Zulkhairi Abdul Rahim, 2019).
Morph-X-Select	A variant of tissue SELEX where tissue sections from patients are used to select high-affinity aptamers by laser microdissections of specific regions based on a morphological assessment. The method overcomes typical tissue heterogeneity and selects aptamers systematically and accurately (Hongyu Wang et al., 2016).
FRELEX	A free-selection platform developed by NeoVentures Biotechnology Inc. (WO 2017035666 A1) that does not require target or library immobilization, making it suitable for true free-free selection against fully exposed targets like constituents of bodily fluids (Lecocq et al., 2018).
Staggered target (ST)-SELEX	ST-SELEX allows isolation of highly specific aptamers that do not exhibit cross-reactivity beteween highly similar target proteins like staphylococcal enterotoxins (Sedighian et al., 2018).
AgFACS-SELEX (AgNP-assisted FACS SELEX)	A selection method combining the use of silver decahedral nanoparticles (AgNPs) and FACS. The method is sensitive and efficient as it reports advantages such as enhanced fluorescence intensities by AgNP and selection of non-self-hybridized species (Yu et al., 2019).

Appendix A (Cont'd)

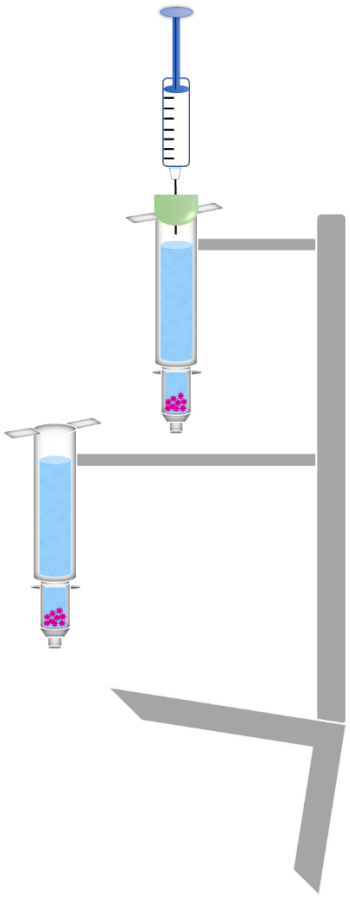
FluCell-SELEX	A derivative of FluMag-SELEX where fluorescently labelled polyclonal aptamer libraries can be generated by cell selections and have superior properties to single enriched aptamers (Kubiczek et al., 2020).

B. Pre-equilibration of Talon® Metal Affinity Resin

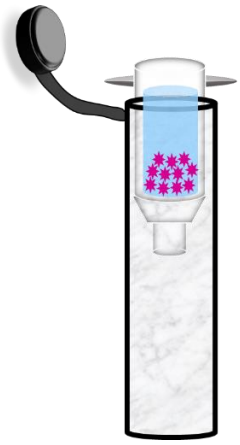
Two batches of 1 mL of beads were loaded into two affinity chromatography filtered columns. The beads storage solution was discarded completely leaving the filtered beads. The dried-out beads were then washed and filtered three times with 600 µL of MQ-water and were then washed and filtered three times with 600 µL of PB-5. Finally, at least 600 µL of PB-5 was added again to the beads that were then carefully collected with a suitable pipette into a clean 50 mL Falcon tube. The two columns were thoroughly washed with excess water and stored to be used again later in AdiC metal affinity purification.

C. Metal Affinity Chromatography Setups

1) Column-Format Setup for AdiC Purification:



2) 2 mL Eppendorf-Promega Column Formation for AdiC Elution by Microcentrifugation:



D. SDS-PAGE Gel Recipes

1) 12% Separating Gel Preparation (two gels):
   - 4 mL acrylamide/bis-acrylamide (29:1 solution)
   - 2.5 mL 1.5 M Tris-Base, pH 8.8
   - 3.3 mL MQ-water
   - 0.1 mL 10% SDS
   - 0.1 mL 10% APS
   - 0.008 mL TEMED

2) 5% Stacking Gel Preparation (two gels):
   - 0.85 mL acrylamide/bis-acrylamide (29:1 solution)
   - 0.625 mL 1 M Tris-HCl, pH 6.8
   - 3.4 mL MQ-water
   - 0.05 mL 10% SDS
   - 0.05 mL 10% APS
   - 0.005 mL TEMED

E. SDS-PAGE Buffers/Solutions Recipes

1) 10× Gel Running Buffer Preparation pH 8.3, 1 L:
   - 30 g Tris-base
   - 144 g Glycine
   - 10 g SDS
   - 1000 mL MQ-water

2) 5× Sample Loading Buffer Preparation:
   - 250 mM Tris-HCl, pH 6.8
   - 10% SDS
   - 50% Glycerol
   - 0.25% Bromophenol Blue
   - 5% β-mercaptoethanol (β-ME)
     - Alternatively, 500 mM dithiothreitol (DTT)

3) Staining Solution Preparation, 1 L:
   - Add 100 mL of glacial acetic acid to 500 mL of MQ-water.
   - Add 400 mL of 100% methanol and mix.
   - Add 1 g of Coomassie R-250 dye and mix by rotational shaking (mixing overnight yields better dye solubility).
   - Filter any insoluble particles and store buffer at RT.

4) De-staining Solution Preparation, 1 L:
   - Add 100 mL of glacial acetic acid to 700 mL of MQ-water.
   - Add 200 mL of methanol and mix rotationally.
   - Store buffer at RT.

F. Bradford Protein Assay

1) Serial Dilutions of the Protein Standard (BSA):

Tube #	Standard's Volume (µL)	Standard's Source	Diluent's Volume (µL)	Final [Standard] in µg/mL
1	30	2 mg/mL stock	10	1,500
2	20	2 mg/mL stock	20	1,000
3	30	Tube 2	10	750
4	20	Tube 3	10	500
5	20	Tube 4	20	250
6	20	Tube 5	20	125
7 (blank)	-	-	20	0

2) Microplate Layout and Sources of the BSA Standard and the AdiC Sample:

Duplicates 1	**Duplicates 2**		**Duplicates 1**	**Duplicates 2**		
Tube 1	Tube 1		AdiC (2× dil)	AdiC (2× dil)		
Tube 2	Tube 2		AdiC (5× dil)	AdiC (5× dil)		
Tube 3	Tube 3		AdiC (10× dil)	AdiC (10× dil)		
Tube 4	Tube 4					
Tube 5	Tube 5					
Tube 6	Tube 6					
Tube 7	Tube 7					

G. SELEX: Supplementary Material

1) TAE Buffer Recipe:
   - To prepare 500 mL of 0.5 M EDTA stock solution:
     - Dissolve 93.05 g of EDTA (372.24 g/mol) completely in 400 mL of MQ-water by adjusting the pH with sodium hydroxide (NaOH) to 8.0.
     - Top up to 500 mL, filter-sterilize and autoclave.
   - To prepare 50× TAE buffer stock solution:
     - Dissolve 242 g of Tris-base in 700 mL of MQ-water.
     - Add 57.1 mL of glacial acetic acid.
     - Add 100 mL of 0.5 M EDTA (pH 8.0).
     - Complete to 1 L. pH of the final solution should be 8.5. Store at RT.
   - To prepare 1× TAE buffer working solution:
     - Dilute 20 mL of 50× TAE buffer by adding MQ-water until a final volume of 1000 mL.

2) Recipes to Preparing a Native- or Urea-PAGE 12% PA Gel:
   - To prepare a 100 mL of 12% stock gel solution:
     - Add 40 mL acrylamide/bis-acrylamide (29:1 solution) in a Duran bottle.
     - Add 20 mL of 5× TBE buffer and complete to 100 mL.
     - The above instructions are for a Native-PAGE gel. For a Urea-PAGE gel, dissolve 50 g of urea in the solution.
     - Store the solution(s) at 4°C.
   - T prepare a 6 mL of 12% working gel (denaturing or nondenaturing):
     - Take 6 mL of gel solution from the prepared stock.
     - Add 40 µL of 10% APS.
     - Add 4 µL of TEMED.
   - 5× TBE Buffer Recipe:
     - Dissolve 54 g of Tris-base in 700 mL of MQ-water.
     - Dissolve 27.5 g of Boric acid in the solution.
     - Add 20 mL of 0.5 M EDTA (pH 8.0).

- Complete to 1 L. pH of the final solution should be 8.3. Store at RT.

3) To prepare 1× TBE Buffer Working Solution:
   - Dilute 200 mL of 5× TBE buffer by adding MQ-water until a final volume of 1 L.

4) NaOAc 2.86 M, pH 5.3 (at 24 °C), 200 mL Recipe:
   - Dissolve 77.84 g of NaOAc trihydrate in 170 mL of MQ-water.
   - Adjust the pH to 5.3 by adding glacial acetic acid.
   - Allow the solution to cool overnight and then readjust the pH to 5.3.
   - Top up the solution to the final volume (200 mL).
   - Filter-sterilize and autoclave.

5) Volume Calculations Used for the Oligo-target Incubation at Each Round and to Execute our SELEX Program Shown in **Table 3.4**.

Selection Round	Calculations
1	*1)* *RNA pool:* Concentration (BioDrop datum): 2644 µg/mL Concentration in µM: = = M = 80.1 µM Volume taken: **V1 = 25 µL** *2)* *AdiC:* Concentration (Bradford assay): 0.34 mg/mL Concentration in µM: = = M = 1.25 µM Volume taken:

V1 = 20 µL

3) *2× PBS:*

V1 = 40 µL

4) Add **15 µL** MQ-water to complete the volume to 100 µL.

2

1) *RNA pool:*

Concentration (BioDrop datum): 2322 µg/mL

Concentration in µM:

= M = 70.4 µM

Volume taken:

V1 = 22.72 µL

2) *AdiC:*

Volume taken:

V1 = 14.4 µL

3) *2× PBS:*

V1 = 42.8 µL

4) Add **20.1 µL** MQ-water to complete the volume to 100 µL.

3

1) *RNA pool:*

Concentration (BioDrop datum): 2485 µg/mL

Concentration in µM:

= M = 75.3 µM

Volume taken:

V1 = 25.5 µL

2) *AdiC:*

Volume taken:

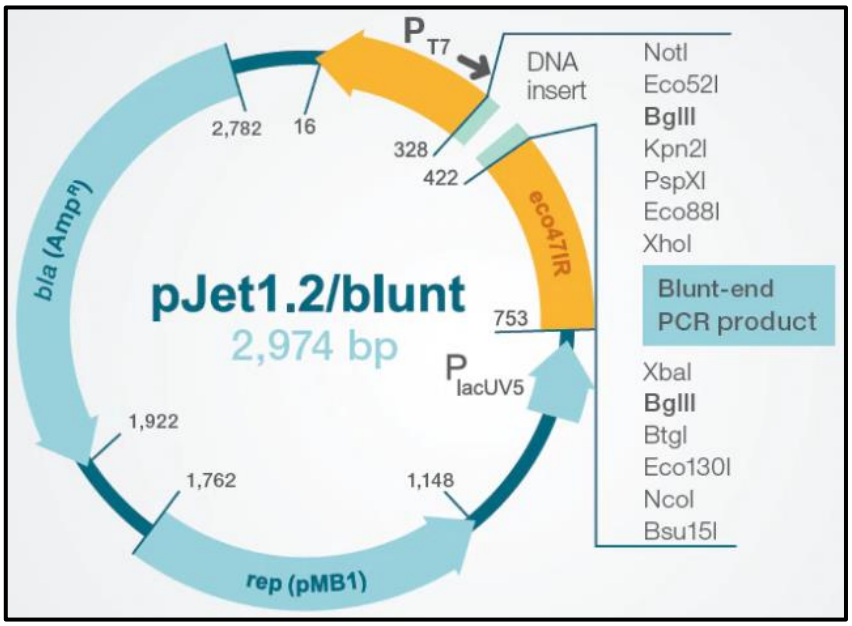
		V1 = 14.4 µL
	3)	*2× PBS:*
		V1 = 67.8 µL
	4)	Add **42.3 µL** MQ-water to complete the volume to 100 µL.
4	*1)*	*RNA pool:* Concentration (BioDrop datum): 1900 µg/mL Concentration in µM: = M = 57.6 µM Volume taken:
		V1 = 26.04 µL
	2)	*AdiC:* Volume taken:
		V1 = 10.8 µL
	3)	*2× PBS:*
		V1 = 69.6 µL
	4)	Add **43.56 µL** MQ-water to complete the volume to 100 µL.
5	*1)*	*RNA pool:* Concentration (BioDrop datum): 1400 µg/mL Concentration in µM: = M = 42.4 µM Volume taken:
		V1 = 28.3 µL
	2)	*AdiC:* Volume taken:
		V1 = 8.4 µL
	3)	*2× PBS:*

	V1 = 70.8 µL *4)* Add **42.5 µL** MQ-water to complete the volume to 100 µL.
6	*1)* *RNA pool:* Concentration (BioDrop datum): 1531 µg/mL Concentration in µM: = M = 46.4 µM Volume taken: **V1 = 19.4 µL** *2)* *AdiC:* Volume taken: **V1 = 5.4 µL** *3)* *2× PBS:* **V1 = 72.3 µL** *4)* Add **52.9 µL** MQ-water to complete the volume to 100 µL.
7	*1)* *RNA pool:* Concentration (BioDrop datum): 920 µg/mL Concentration in µM: = M = 27.9 µM Volume taken: **V1 = 21.51 µL** *2)* *AdiC:* Volume taken: **V1 = 3.36 µL** *3)* *2× PBS:* **V1 = 73.32 µL**

	4)	Add **51.81 µL** MQ-water to complete the volume to 100 µL.
8	*1)*	*<u>RNA pool:</u>* Concentration (BioDrop datum): 994 µg/mL Concentration in µM: = M = 30.1 µM Volume taken: **V1 = 9.97 µL**
	2)	*<u>AdiC:</u>* Volume taken: **V1 = 1.56 µL**
	3)	*<u>2× PBS:</u>* **V1 = 74.22 µL**
	4)	Add **64.25 µL** MQ-water to complete the volume to 100 µL.

H. pJET 1.2/blunt Cloning Vector Map

- The vector contains a pMB1-plasmid replicon and confers ampicillin resistance by the β-lactamase gene. It also contains the lethal gene *eco47IR* the expression of which is controlled by *lac*UV5 promoter and is disrupted by ligation of the insert into the multiple cloning site. The vector also contains a T7 promoter for transcription of the cloned DNA fragment.



I. Competent Bacterial Cells Preparation

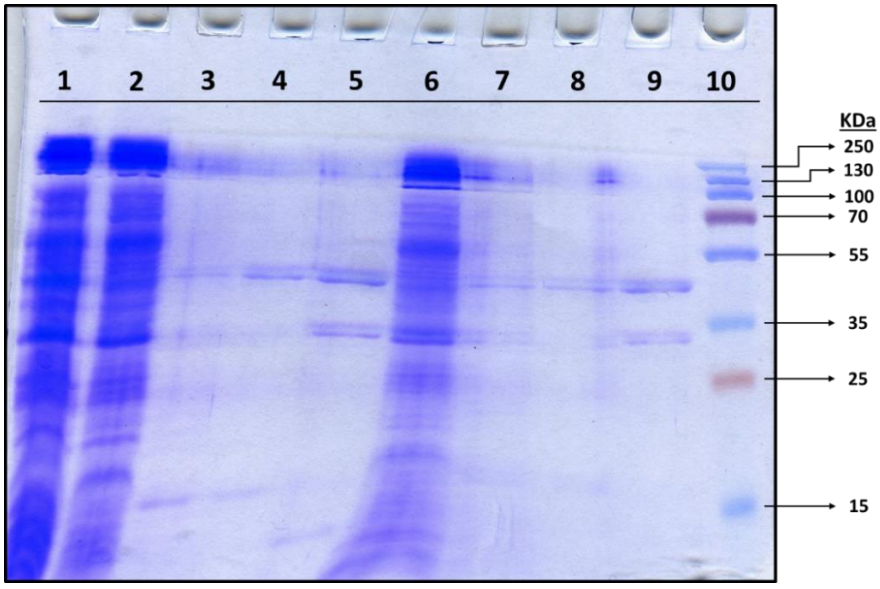
- Protocol:
  - Inoculate 3 mL of LB medium with a single picked colony of BL21(DE3) and grow overnight at 37 °C with shaking at 180 rpm.
  - Inoculate 200 mL of LB with the prepared bacterial seed and grow at 37 °C with shaking at 180 rpm until an OD_{600} of 0.4–0.5 is reached.
  - Incubate the culture on ice for 15 min and then harvest it by centrifugation at 3500 rpm for 5 min at 4 °C followed by carefully removing all of the supernatant.
  - Gently resuspend the pellet in 20 mL of ice-cold buffer 1 (recipe below) by swirling and then incubate the resuspended cells in ice for 15 min.
  - Centrifuge the cells again at 3500 rpm for 5 min at 4 °C and then carefully remove all of the supernatant.
  - Gently resuspend the cells in 8 mL of ice-cold buffer 2 (recipe below) and incubate in ice for 15–30 min.
  - Aliquot (100–120 µL) the cells into microcentrifuge tubes and then freeze in liquid nitrogen or at −80 °C.

- Buffer 1 Recipe:
  - $RuCl_3$ 100 mM, KAc 30 mM, $CaCl_2$ 10 mM, glycerol 15%.
  - Adjust buffer pH to 5.8 with dilute acetic acid; filter-sterilize.

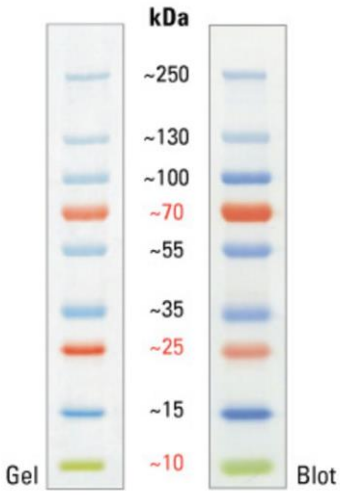
- Buffer 2 Recipe:
  - $CaCl_2$ 75 mM, $RuCl_3$ 10 mM, MOPS 10 mM, glycerol 15%.
  - Adjust buffer pH to 6.5 with 0.2 M KOH; filter-sterilize.

J. SDS-PAGE: Supplementary Material

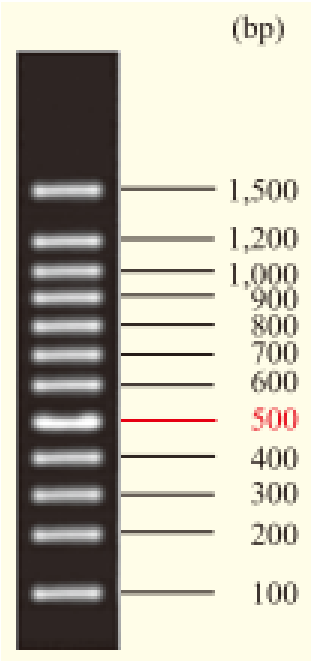
1) Electrophoretic run of the SDS-PAGE containing the various samples collected during the purification of Recombinant AdiC (Lane 1: Supernatant; Lane 2: FT; Lane 3: Wash 1; Lane 4: Wash 2; Lane 5: Wash 3; Lane 6: FT; Lane 7: Wash 1; Lane 8: Wash 2; Lane 9: Wash 3; Lane 10: Protein ladder):



2) Commercial image of the band profile of PageRuler™ Plus Prestained Protein Ladder on a 4-20% Tis-glycine SDS-PAGE gel and a subsequently blotted membrane:

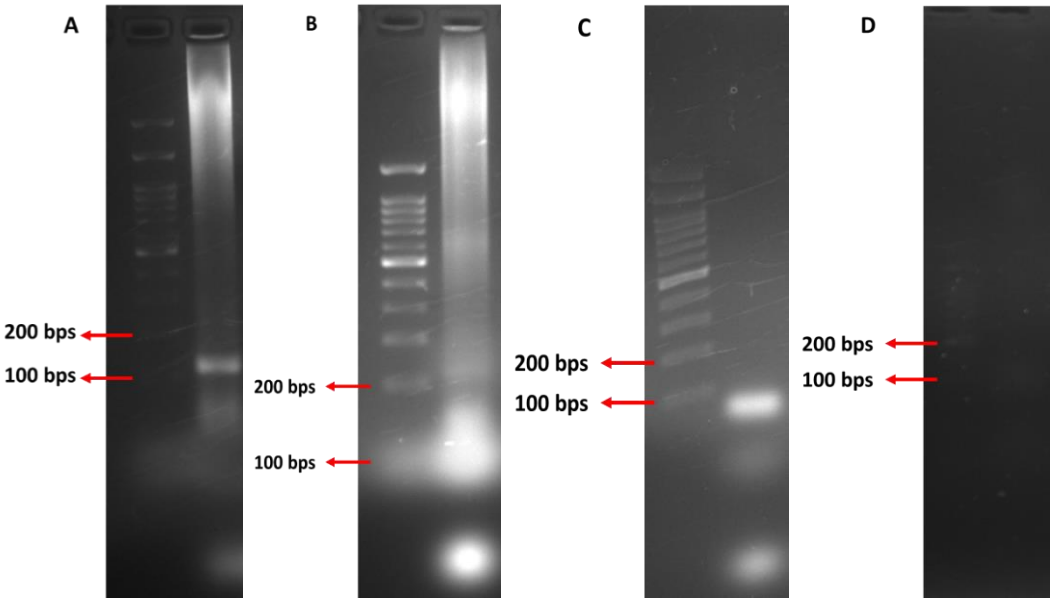


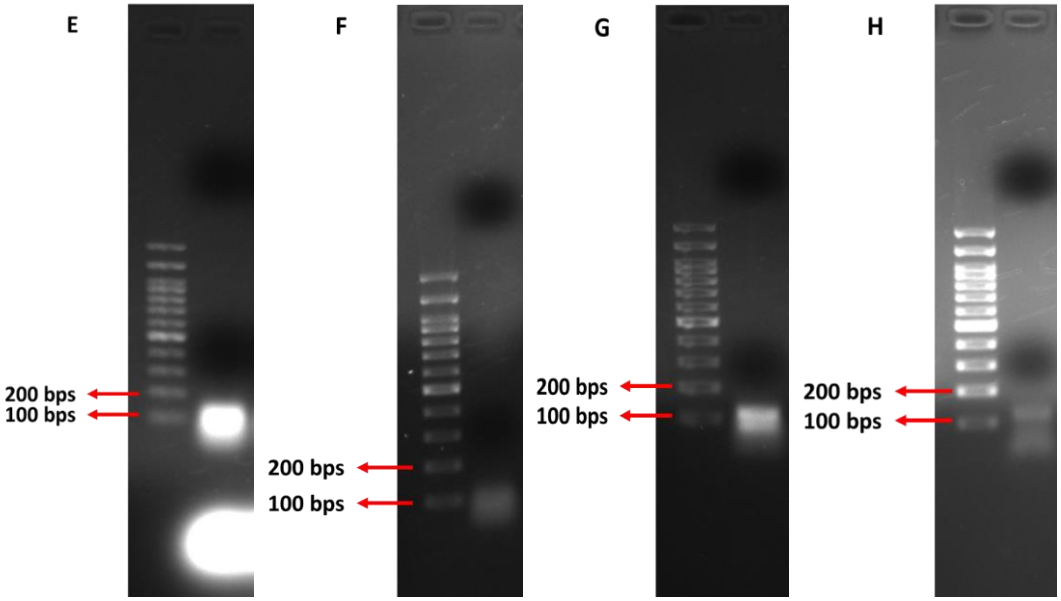
K. Commercial Image of the Band Profile of 100 bp Plus DNA ladder on a 2.0% TAE Agarose Gel Stained with GelStain (Loading Volume: 5 µL).



L. Trial Agarose Gels with 117-bps cDNAs Obtained After *rt*-PCR for the Selected Ligand Pools of Every SELEX Round

- **A)** Round 1; 2% gel. **B)** Round 2; 1.5% gel. **C)** Round 3; 1.5% gel. **D)** Round 4; 1% gel. **E)** Round 5; 1% gel. **F)** Round 6; 1% gel. **G)** Round 7; 1% gel. **H)** Round 8; 1% gel. The smears seen on the sample lane on gels **A** and **B** are random hexamers present in the master mix of a different cDNA synthesis kit; the kit was replaced by round 3. Some of the gels shown have distinguishable bands that are lower than 100 bps. These represent primers and primer dimers from the rt-PCR mixture; they were diminished in later rounds by reducing primers concentration in the two-step rt-PCR procedure. Despite the questionable presence of a cDNA band in gel **D** due to the very low visibility, the cDNA band was confirmed; after all, a non-existent DNA pool in the 3rd SELEX round is not grounds for propagating selection towards consecutive rounds.




E
200 bps
100 bps
F
200 bps
100 bps
G
200 bps
100 bps
H
200 bps
100 bps

M. Multiple Sequence Alignment (Clustal Ω) of the 34 Unique Proto-aptamer Clones

```
CLUSTAL O(1.2.4) multiple sequence alignment

AdiC59      GGGAGACAAGAAUAAACGCUCAAUUUUCUCUUUA----UUUAAU---AUUUAU-----AC     48
AdiC56      GGGAGACAAGAAUAAACGCUCAAUGCUCUCUAUA-----UUAAU---AUUUAUCAUUUAU     52
AdiC10      GGGAGACAAGAAUAAACGCUCAAUUUUUUUCA------GUUUAU-CUCUUUGUCAUUUAU     53
AdiC9       GGGAGACAAGAAUAAACGCUCAAUAUAACUCGCU----UUUCUU---UC---UUAUUUCU     50
noorC       GGGAGACAAGAAUAAACGCUCAAUGAGCCCCUAU----UUUUCU---ACCAACAACUUGG     53
AdiC83      GGGAGACAAGAAUAAACGCUCAAUUCUUUUACCUU----UAUUU---AUUA---AUUUAC     50
AdiC73      GGGAGACAAGAAUAAACGCUCAAUUUCCUCAAUUAA--UUUUAU---ACUGAUUAGCUGC     55
AdiC36      GGGAGACAAGAAUAAACGCUCAAGCGCCUUGCAUC---UUUACG---CCGUCCUAUUUUC     54
AdiC99      GGGAGACAAGAAUAAACGCUCAAGCCGAGUACUUC------UGU---CGGUUUACAUUAU     51
AdiC95      GGGAGACAAGAAUAAACGCUCAAUCCUCCAAAAUC----UUUGU---ACAGUCUAUUUAC     53
AdiC72      GGGAGACAAGAAUAAACGCUCAACUCUAACUGUG-A--UUU--C---ACUUUCUCUUUAU     52
AdiC58      GGGAGACAAGAAUAAACGCUCAAUAUUUCAACCC----UUUCUU---AUUUCC----UAU     49
AdiC5       GGGAGACAAGAAUAAACGCUCAAUCUUUGUUGCACA--UGUUUU---AACUAUUGG--UC     53
AdiC46      GGGAGACAAGAAUAAACGCUCAACAUAUCU----AA--UUUUAU---CUCUCUUAUUUAC     51
AdiC19      GGGAGACAAGAAUAAACGCUCAACCUGUGUAUUC-----UCUUU---AUACUCUCUUUAU     52
AdiC85      GGGAGACAAGAAUAAACGCUCAACCCAUAUUUUU-----ACCAU---ACAAAUUCAUUUU     52
AdiC33      GGGAGACAAGAAUAAACGCUCAACCUGUUAUUCAA----UUCAU---ACAAAUAAUUUCU     53
AdiC75      GGGAGACAAGAAUAAACGCUCAACCCAUU-----CA--UUUCUU---U------C-----     39
AdiC80      GGGAGACAAGAAUAAACGCUCAACCGUUAUUUCACA--UUUCUU---U------AUCUUU     49
AdiC53      GGGAGACAAGAAUAAACGCUCAAGUACGUGCU------------ACUGUUUAG-ACCUCC     47
AdiC92      GGGAGACAAGAAUAAACGCUCAAUUUUGC---------UUAAAUACUUUUUCUUACCUUC     51
AdiC18      GGGAGACAAGAAUAAACGCUCAAGUUUAG-CUGGCGCAUUUAUU---AU-----ACUCCG     51
AdiC41      GGGAGACAAGAAUAAACGCUCAAGUGCCG-CUGU-----GAGUGUGUACUGCAAACU---     51
AdiC79      GGGAGACAAGAAUAAACGCUCAAUUUAAUCUUUUCAUGU----UACCACAG---AGU---     50
AdiC69      GGGAGACAAGAAUAAACGCUCAAUUGUGA-------UAUUCUUUCUGAUUCGUAUUCCAC     53
AdiC26      GGGAGACAAGAAUAAACGCUCAAUAUCUA-UU----UUUUAGUUAUUAU-----AUCAUA     50
AdiC8       GGGAGACAAGAAUAAACGCUCAAUUUUUC-CUGU--CAU----UAUUUCUGUAUACCUAG     53
AdiC71      GGGAGACAAGAAUAAACGCUCAACUC----------CGUUAUUUACUAUUUAUAAUUAUA     50
AdiC12      GGGAGACAAGAAUAAACGCUCAAUUCAUU--UGU--CCUU---UAUUAUUUUAAACCUAA     53
noorA       GGGAGACAAGAAUAAACGCUCAAGUGU-C--UGU--UCAUGAUUGCAUUUUUAUAUUAUA     55
AdiC3       GGGAGACAAGAAUAAACGCUCAAUGGG---------CAUUUGUUCCUUUUUGUAAUUCAU     51
AdiC43      GGGAGACAAGAAUAAACGCUCAAUCCU-A--CAUCACAUUAAUUACUAUUUAUGAUUUUU     57
AdiC25      GGGAGACAAGAAUAAACGCUCAAUAA----------UGUU--UUCCUCUUUUUC------     42
noorB       GGGAGACAAGAAUAAACGCUCAACGA----------UUGUGAUGCCUCUUUUUUAUUUUA     50
            **********************
AdiC59      AAUCAUUUUCUUAAAC--UUUCCACAACAUUUCGACAGGAGGCUCACAACAGGC 100
AdiC56      UAUU---UAAUCUAUA--U-UAUUUUGUGUUUCGACAGGAGGCUCACAACAGGC 100
AdiC10      AAUUU--UCUAUUA----UUCAUACUAUUUUUCGACAGGAGGCUCACAACAGGC 101
AdiC9       UUUU--UUCACCCAUG--CUGUUAUUGGUCUUCGACAGGAGGCUCACAACAGGC 100
noorC       UAUUU--UAUUUUA----U-AUAUUUUCAAUUCGACAGGAGGCUCACAACAGGC 100
AdiC83      UAUU--UACAUUCUUU--G-CAGUGUUUAUUUCGACAGGAGGCUCACAACAGGC 99
AdiC73      UGCU---UAUUU-A-A--U-UUUUAUACAUUUCGACAGGAGGCUCACAACAGGC 101
AdiC36      UUA----UAUUUU--A--C-GAUUCUUUUUUUCGACAGGAGGCUCACAACAGGC 99
AdiC99      UAAUUUUUCGCGCAUG--C----ACUUUUUUUCGACAGGAGGCUCACAACAGGC 99
AdiC95      UUAUUUAUUAUUUAUU--A----AAUUUUAUUCGACAGGAGGCUCACAACAGGC 101
AdiC72      CAUUCUUGAUUU---U--C-UCUUUGGAGCUUCGACAGGAGGCUCACAACAGGC 100
AdiC58      UACGCUUUCUUUUAAG--C-UUUUUGGCAUUUCGACAGGAGGCUCACAACAGGC 100
AdiC5       UAA-CCAUAAUU---C--U-UAUUUAUAUUUUCGACAGGAGGCUCACAACAGGC 100
AdiC46      AUUUAUUUAAAG---C--G-AAUUUAUUUUUUCGACAGGAGGCUCACAACAGGC 99
AdiC19      UAAUUUUUGGUA---U--U-UAUUUUUGAUUUCGACAGGAGGCUCACAACAGGC 100
AdiC85      ACUUCUUUGAUC---A--U-AAUUUUUUAUUUCGACAGGAGGCUCACAACAGGC 100
AdiC33      UAUU---AUUUU---U--G-CUGUUUUUUAUUCGACAGGAGGCUCACAACAGGC 98
AdiC75      ------UUCUUU---U--C-UUUAUCUUAUUUCGACAGGAGGCUCACAACAGGC 81
AdiC80      UAUCCUUUCCUU---U--U-UUUUUAUUUUUUCGACAGGAGGCUCACAACAGGC 97
AdiC53      GGU-UUGUCGGCAAUGUGCUGGCGUAAAAUUUCGACAGGAGGCUCACAACAGGC 100
AdiC92      AUU-AUAC--UUUCGUG--UGCCAGCUAGUUUCGACAGGAGGCUCACAACAGGC 100
AdiC18      ACUGUAG----AUACUGA-UAUCGCGUUUAUUCGACAGGAGGCUCACAACAGGC 100
AdiC41      -UUUUUAUUA-UUAU----UUGACUAUUUAUUCGACAGGAGGCUCACAACAGGC 99
AdiC79      -CUCUCAU---UUAUUGCCAUACCUUUUAUUUCGACAGGAGGCUCACAACAGGC 100
```

```
AdiC69      CGUUUUGAA--UUUUU-----GAUGAUUUUUUCGACAGGAGGCUCACAACAGGC 100
AdiC26      UUUUUUAGAG-UUCUA-A-UUUAAUUUUUAUUCGACAGGAGGCUCACAACAGGC 101
AdiC8       CAUUUUU----UUAU----UAUUUAGAUUCUUCGACAGGAGGCUCACAACAGGC 99
AdiC71      CAUUUUGU---UUACUGG-AUUUUAUCUUAUUCGACAGGAGGCUCACAACAGGC 100
AdiC12      AUU-UUAC---UUAUU---AAUCUCAGGUCUUCGACAGGAGGCUCACAACAGGC 100
noorA       UUU-UCAC--CUUGGU------GUUUUUAAUUCGACAGGAGGCUCACAACAGGC 100
AdiC3       UUC-GCGCG-CUUGGA---AGAUCUGUUAUUUCGACAGGAGGCUCACAACAGGC 100
AdiC43      -UU-AUGU---UUUUA---C--CUCAUUAAUUCGACAGGAGGCUCACAACAGGC 101
AdiC25      -UU---GUU--UUAUA------CCUUUUUAUUCGACAGGAGGCUCACAACAGGC 84
noorB       UUU-GUGUUGUUUAUA---UUGGGAUUUUAUUCGACAGGAGGCUCACAACAGGC 100
                                          ************************
```

N. Predicted Thermodynamic Stabilities (Kcal/mol) of the Five Aptamer Candidates

Proto-aptamer	ΔG *RNAfold* (MFE)	ΔG *RNAfold* (Ensemble)	ΔG *Kinefold*
noorA	− 13.30	− 15.00	− 13.00
noorB	− 18.30	− 19.60	− 20.40
noorC	− 15.90	− 17.65	− 17.90
AdiC19	− 12.30	− 14.14	− 15.60
AdiC72	− 21.10	− 23.26	− 23.60

O. Computer-generated Model of an AdiC PDL Complex Solubilized and Purified in OM.

- The model below is based on the AdiC dimeric structure (PDB: 3LRB) and the experimental data that describes the number of detergent and lipid molecules bound to the complex purified in 1.5% OM (and solubilized in 5% OM). OM bears physicochemical properties that are distinct from DDM. 109 and 17 molecules of OM and POPE are shown, respectively. The AdiC dimer is shown as the blue cartoon and transparent spheres. POPE and OM molecules, and oxygen and nitrogen atoms are shown as yellow and black spheres, and red and blue spheres, respectively. The figure is reprinted from Biophysical Journal, 106, İlgü et al., Variation of the Detergent-Binding Capacity and Phospholipid Content of Membrane Proteins When Purified in Different Detergents, 1660–1670, Copyright (2014), with permission from Elsevier.