

Georgia State University

ScholarWorks @ Georgia State University

Chemistry Dissertations

Department of Chemistry

8-10-2021

A reporter platform for characterizing transcription factor-DNA interactions in live cells

Hye Mi Kim
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

Recommended Citation

Kim, Hye Mi, "A reporter platform for characterizing transcription factor-DNA interactions in live cells." Dissertation, Georgia State University, 2021.
doi: <https://doi.org/10.57709/24099907>

This Dissertation is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

A Reporter Platform for Characterizing Transcription Factor-DNA Interactions in Live Cells

by

Hye Mi Kim

Under the Direction of Gregory M. K. Poon, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Reporters are genetic constructs that place the biosynthesis of an exogenous probe, such as an indicator enzyme or fluorescent protein, under the control of a cloned promoter. Benefiting from advances in molecular cloning, both the regulatory and indicator elements can be highly customizable to meet specific experimental needs. Reporters are therefore flexible tools for investigating mechanisms of transcriptional regulation in live cells, where detailed characterization of protein/DNA complexes remains challenging. In addition, reporters are attractive cell-based indicators in library screening and downstream pre-clinical characterization of hit compounds. To obtain a better understanding of the mechanism of action of PU.1, a master transcription regulator of hematopoiesis from the ETS family, as well as to establish a cell-based platform for translational efforts, we engineered a series of highly engineered reporters for PU.1. We designed a dual-color fluorescent reporter system for interrogating mammalian cell lines that do not natively express PU.1. The system generates spectrally distinct fluorescent markers that are coupled with the exogenous expression of PU.1 and the transactivation of a PU.1-responsive enhancer/promoter element. We used this reporter system to investigate the evolutionary basis of target selection by the DNA-binding domain of ETS proteins toward PU.1-specific recognition. Further, we designed synthetic enhancer structures varying in number and spacing of PU.1 binding sites to test mechanistic hypotheses of PU.1 auto-regulation via a negatively cooperative DNA-bound dimer. Finally, by (single-cell) flow cytometry and (batch) plate-based fluorimetry, we used the reporter to characterize a panel of novel DNA-targeted heterocyclic diamidines with therapeutic potential as PU.1 inhibitors in acute myeloid leukemia. In summary, fluorescence-based reporters are powerful tools for correlating functional and biophysical properties of PU.1 in living cells. We expect that further development of our reporters to yield useful molecular

tools for addressing mechanistic inquiries in transcriptional regulation and quantitative indicators in high-throughput screening applications.

INDEX WORDS: Sequence-specific transcription factor, PU.1, DNA-binding domain, Cell-based reporter platform, TF-DNA recognition, DNA minor groove binders

Copyright by
Hye Mi Kim
2021

A Reporter Platform for Characterizing Transcription Factor-DNA Interactions in Live Cells

by

Hye Mi Kim

Committee Chair: Gregory M. K. Poon

Committee: W. David Wilson

Kathryn B. Grant

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2021

DEDICATION

I dedicate this dissertation to my family, Dr. Choong-Yeol Kim, Dr. Kyung-Ok Kim, and Dr.

Soomi Kim for all unconditional love and support.

In the memory of my Godfather Dr. Abraham Park.

ACKNOWLEDGEMENTS

From the bottom of my heart, I would like to express my thanks to my patient and supportive supervisor, Dr. Gregory Poon (truly the GOAT), who supported me throughout this research project. I am incredibly grateful for his support in my growth as a scientist. His endless endeavors from the beginning to the end of my scientific journey have motivated me to improve my shortcomings in every corner of my thoughts. His tireless guidance, positive encouragement, and comprehensive insights have made this an inspiring experience for me. Without his support, I certainly cannot complete this journey and move forward with my future. Throughout my time in his lab, Dr. Poon has always emphasized to me the paramount importance of academic integrity. I therefore claim as my own any and all problems with this dissertation and assume personal responsibility for them.

I am so grateful for the committee members Dr. David Wilson and Dr. Kathryn Grant, who have been great sources of support and directed me to the final line. I would give my special thanks to Dr. Wilson, who provided me excellent insights and advice to understand fascinate sciences of small molecules and DNA. Even I was very naïve to step into the field in my first year, his deep understanding guided me in opening my eyes to new scientific worlds and working with the projects. Further, I would like to thank Dr. Grant for this dissertation's thoughtful comments, understanding, encouragement, and recommendations. Moreover, I would like to express my thank Dr. David Boykin for providing great opportunities to explore his small compounds, which are long-life scientific fruits with great potentials. Even though I could not have a chance to have my committee members because of his retirement, I could not stop thanking him for his great resources.

I would express my sincere gratitude to Dr. Susan Mooring, Dr. Suri Iyer, and Dr. Donald Hamelberg. With their solid supports for my graduate students' life, I was able to complete my journey. Their detailed advice and guidance directed me to follow the right path and complete the race.

I want to express my gratitude to the MBD fellowships, GSU scholarship program for sponsored me, and let me be part of this incredible multidisciplinary young enthusiastic scientist network.

I am also thankful to Dr. Wilson's lab members, especially Dr. Ananya Paul, for all the thoughtful guidance. She always encouraged, supported, and cheered me. She helped me to enrich my knowledge of small compounds and broaden my perspectives on medicinal chemistry.

To Dr. Poon's lab members, I would like to thank you for inspiring me in the research. Dr. Sangchoon Lee, Dr. Shingo Esaki, Dr. Suela Xhani, Dr. Amanda Albrecht, Noa Erlitzki, Van Ha, and Ken Huang gave insights, advice, and helps on my projects. Without their help, cares, and supports, I could not arrive here. They would be remembered as such a lucky charm for their kindness, gentleness, and strong encouragement to complete the journey throughout my tough road in the journey. I am so grateful to have them in the part of my life-long journey.

Finally, I could not stop expressing tremendous thanks to my family for unconditional supporting me during the compilation of this dissertation. My Godfather, Dr. Abraham Park, although he could not see my degree award and share my joy here, I believe that he has been cheering me in heaven, and now he is pleased by my accomplishment. My dad Dr. Choong-Yeol Kim, my mom Dr. Kyung-ok Jung, and my sister Dr. Soomi Kim have persistently shown inexplicable love and solid supports. Even though I fell, made full of mistakes, and was almost about to give up my journey, your unchanging love and trust were my motivation to be back on

my track to keep moving forward. Thank you for all that you do. I will never be here without your never-ending prayers, encouragements, dedications, and sacrifices. Now it is time for me to give back to you in my new chapter of life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
LIST OF TABLES	XVI
LIST OF FIGURES	XVII
LIST OF ABBREVIATIONS	XX
1	GENERAL INTRODUCTION.....	21
1.1	Inherited traits transfer from one generation to the next	21
1.2	Transcriptional process and control	22
1.2.1	<i>Complexity of tight regulation in gene expression.....</i>	<i>22</i>
1.2.2	<i>Epigenetic controls for gene expression.....</i>	<i>24</i>
1.2.3	<i>Transcriptional regulation</i>	<i>29</i>
1.2.4	<i>Specific DNA sequences recognition by TFS.....</i>	<i>35</i>
1.2.5	<i>Significance of DNA target recognition by transcription factors.....</i>	<i>36</i>
1.3	The ETS family of transcription factors	38
1.3.1	<i>The importance of understanding the DBD of the TF family.....</i>	<i>38</i>
1.3.2	<i>The ETS transcription factor's redundant binding on DNA in living cells.....</i>	<i>41</i>
1.3.3	<i>ETS family members and spectrum of diseases</i>	<i>43</i>
1.4	Transcription factor PU.1 as an experiment model.....	50
1.4.1	<i>The rationale for transcription factor PU.1 as an experiment model.....</i>	<i>50</i>
1.4.2	<i>PU.1 in structural and molecular biology</i>	<i>51</i>

1.4.3	<i>PU.1 and diseases occurrences</i>	53
1.4.4	<i>Special features of transcription factor PU.1</i>	54
1.5	Biochemical and biophysical techniques as a tool for TF-DNA interaction studies	58
1.5.1	<i>Approaches to studying transcription factor activation</i>	58
1.5.2	<i>The electrophoretic mobility shift (EMSA) assay</i>	59
1.5.3	<i>Footprinting assay</i>	60
1.5.4	<i>Chromatin Immunoprecipitation (Chip) assay</i>	61
1.5.5	<i>Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) assay</i>	62
1.5.6	<i>Surface plasmon resonance (SPR) assay</i>	64
1.5.7	<i>Systematic evolution of ligands by exponential enrichment (SELEX) assay</i>	65
1.5.8	<i>Fluorescence resonance energy transfer techniques (FRET)</i>	66
1.5.9	<i>Nuclear magnetic resonance (NMR)</i>	66
1.5.10	<i>The widely used techniques for detecting protein expression: functional TF transactivation</i>	67
1.6	Reporter gene platform to characterize TF-DNA interaction	73
1.6.1	<i>Chloramphenicol acetyltransferase (CAT) assay</i>	75
1.6.2	<i>Alkaline phosphatase (AP)</i>	76
1.6.3	<i>Secreted alkaline phosphatase (SEAP) assay</i>	77
1.6.4	<i>β-galactosidase (β-gal) assay</i>	77

1.6.5	<i>β-lactamase (BLA) assay</i>	79
1.6.6	<i>Luciferase assay (Bioluminescent reporter gene assays)</i>	81
1.6.7	<i>Quantum dot reporters</i>	83
1.6.8	<i>Summary of the reporter gene assay</i>	84
1.7	Green fluorescent protein (GFP) reporter assay	86
1.7.1	<i>Discovery and major milestones</i>	86
1.7.2	<i>GFP structure</i>	87
1.7.3	<i>Applications of GFP reporter in our research</i>	88
1.8	Objectives	92
1.8.1	<i>Needs for the research</i>	92
1.8.2	<i>Research aims</i>	95
2	DESIGN AND CONSTRUCTION OF CELLULAR REPORTERS FOR PU.1 ACTIVITY	97
2.1	Introduction	97
2.1.1	<i>Synthetic reporter gene assay for characterizing PU.1 activities</i>	97
2.1.2	<i>Strengths of our cellular reporter gene platform Versatility: Easy adaptation for research aims</i>	99
2.1.3	<i>Objectives of this chapter</i>	103
2.2	Experimental design Versatility: Easy adaptation for research aims	104
2.2.1	<i>Principle of cellular reporter's design for PU.1 activities</i>	104

2.2.2	<i>Plasmid design and construction</i>	113
2.2.3	<i>Cell culture, transfection, RT-PCR</i>	116
2.2.4	<i>Lentiviral generation transduction for constructing stable PU.1-dependent reporter cell line</i>	118
2.3	Results	120
2.3.1	<i>Confirmation of piRFP-2A-PU.1 and its expression in HEK293 cells</i>	120
2.3.2	<i>Confirmation of pentameric pλB-D2EGFP</i>	125
2.3.3	<i>Functional expression confirmation of piRFP-2A-PU.1 and pλB-D2EGFP in live cells</i>	127
2.3.4	<i>Design and construction of dominant-negative mutants for characterization functional PU.1 transactivation at TF/DNA level</i>	128
2.3.5	<i>Intrinsic biophysical properties of PU.1 autoregulation and reporter gene assay</i>	132
2.3.6	<i>Development and confirmation of PU.1-dependent reporters for functional cell reporter assay</i>	135
2.4	Summary	143
2.4.1	<i>Advantages of our reporter assay to study PU.1-DNA interaction</i>	143
3	CHARACTERIZATION OF PU.1 ACTIVITY IN LIVE CELLS	146
3.1	Introduction	146
3.1.1	<i>Functional characterization of transcription factor attributes in living cells</i>	146
3.1.2	<i>The transcription factor PU.1 as a model system for reporter characterization</i>	147

3.1.3	<i>Experimental configurations for functional characterization of PU.1 activity using reporters</i>	147
3.1.4	<i>Objectives of this chapter</i>	150
3.2	Materials and methods	151
3.2.1	<i>Plasmid design/Molecular Cloning</i>	151
3.2.2	<i>Cell culture</i>	151
3.2.3	<i>Transfection assay</i>	152
3.2.4	<i>Immunoblotting analysis and qPCR</i>	154
3.3	Results	156
3.3.1	<i>Functional Inhibition of PU.1 by evolutionarily conservative ETS mutants</i>	156
3.3.2	<i>Functional manifestation of PU.1 self-regulation: negative feedback</i>	161
3.4	Discussion	168
3.4.1	<i>Functional approaches to probing molecular hydration in transcription factor function</i>	169
3.4.2	<i>Self-regulatory mechanisms of PU.1</i>	172
3.5	Conclusion	176
4	CHARACTERIZATION IN LIGAND SCREENING AGAINST PU.1	177
4.1	Introduction	177
4.1.1	<i>Drug screening for targeting transcription factor PU.1</i>	177
4.1.2	<i>Sequence-specific DNA minor groove binding small molecules</i>	178

4.1.3	<i>DNA binding preference of the ETS family TFs</i>	180
4.1.4	<i>DNA recognition of PU.1</i>	182
4.1.5	<i>The pharmacological strategy of targeting the PU.1-DNA interaction by sequence-specific binding molecules</i>	183
4.2	Materials and methods	186
4.2.1	<i>Cell culture</i>	186
4.2.2	<i>Plasmid construction</i>	186
4.2.3	<i>Compounds</i>	187
4.2.4	<i>Cell viability assay: Sulforhodamine B (SRB) assay</i>	187
4.2.5	<i>Fluorescence microscopic examination</i>	188
4.2.6	<i>Liquid Chromatography/Mass Spectrometry</i>	188
4.2.7	<i>Flow cytometric analysis</i>	191
4.2.8	<i>Statistics</i>	191
4.3	Results	191
4.3.1	<i>Biophysical characteristics of DB2115 and DB2313</i>	192
4.3.2	<i>Functional PU.1 inhibition by heterocyclic diamidines in live cells</i>	193
4.3.3	<i>Functional inhibition of PU.1 by non-toxic heterocyclic diamidines</i>	195
4.3.4	<i>Nuclear uptake of the heterocyclic diamidines in live unfixed HEK 293 cells.</i>	197
4.3.5	<i>DB2313 in plasma levels decrease by 50% in 24 h</i>	199

4.3.6	<i>Functional inhibition of PU.1-DNA complex in endogenously PU.1 expressing cells using PU.1-dependent genes' reporter system</i>	202
4.3.7	<i>Functionally relevant outcomes of the small compounds in the new reporter assay</i>	206
4.3.8	<i>Different extent of functional PU.1's inhibition of DB2115 and DB2313 on the native promoters</i>	208
4.4	Discussion	212
4.4.1	<i>Feasibility of the DNA minor groove binders as PU.1 inhibitors</i>	213
4.4.2	<i>Advanced cell-based platform for screening ligands against PU.1-DNA complex</i>	214
4.5	Conclusions	218
5	CONCLUSION	220
5.1	Significance of studying for TF-DNA interaction in live cells	221
5.2	Highlights of research	222
5.2.1	<i>Suitability for reporter assay in our study: providing the knowledge between biophysical experiments to biological functions</i>	223
5.2.2	<i>Versatile cell reporter platform for characterization of PU.1 activity in live cells</i>	225
5.2.3	<i>Dominant-negative reporter gene platform supported biological relevance to biophysical and biochemical research</i>	226

5.2.4	<i>Various λB synthetic reporter gene configurations were suited for monitoring the direct dosage effect on transactivation at the protein/DNA level</i>	227
5.2.5	<i>Enhancer (promoter) activity reporter assays for characterizing the functional consequences of TF binding</i>	227
5.2.6	<i>Cell- based platform to screening pharmacological compounds with detection sensitivity and reproducibility</i>	228
5.3	Prospects	230
5.3.1	<i>Practical recommendations.....</i>	230
5.3.2	<i>Future directions and applications</i>	231
	APPENDIX.....	236
	REFERENCES.....	253

LIST OF TABLES

Table 1.1 ETS family of transcription factor and a variety of cancers	44
Table 1.2 Summary of the most commonly used techniques for TF-DNA binding characterization	58
Table 1.3 Common methods for detection of protein expression	67
Table 1.4 Characteristics of gene reporter systems: advantages and disadvantages for aspect of cell-based assay development and screening	84
Table 2.1 Detailed features of pcDNA 3.1(+)	105
Table 2.2 Mammalian expression plasmid utilized in the research	106

LIST OF FIGURES

Figure 1.1 Overview of the flow of genetic information from DNA to protein in eukaryotic cells.	22
Figure 1.2 Complexity of gene regulation in a eukaryotic cell.....	30
Figure 1.3 Schematic diagram of PU.1 dosage effect on leukemogenesis.	53
Figure 1.4 Schematic representation of a cell-based reporter gene assay.....	71
Figure 1.5 Steps in constructing and generating a recombinant DNA plasmid.....	73
Figure 1.6 Structures of the substrates for alkaline phosphatase (AP)	76
Figure 1.7 Structure and reaction scheme for in vitro and in vivo β -galactosidase assay	78
Figure 1.8 Illustration of the β -lactamase reporter system	80
Figure 1.9 Bioluminescence substrates and enzymatic reactions of several common luciferases	83
Figure 1.10 Proposed Mechanism for the intramolecular biosynthesis of the GFP chromophore.	87
Figure 1.11 Overview of the cell-based reporter	91
Figure 1.12 The principle of in vitro PU.1 transactivation assay for screening ligands against PU.1 activity.	94
Figure 2.1 Illustration of a reporter vector.....	104
Figure 2.2 Letiviral transduction: Delivery of gene transfer vector into a mammalian cell in the research.	112
Figure 2.3 Schematic representation of the design of pcDNA iRFP-2A-PU.1.	121
Figure 2.4 A sequence of iRFP-2A-PU.1 in pcDNA3.1(+) and recognition site of the restriction enzymes for cloning.....	124

Figure 2.5 Identification of PU.1 expression level of RNA and iRFP marker in the PU.1 transfected HEK293 cells.	124
Figure 2.6 A schematic diagram of the synthetic enhancer reporter gene vector construct.	126
Figure 2.7 Sequence confirmation of a reporter gene expression vector containing IgλB2-4, a murine enhancer element bound by PU.1.	127
Figure 2.8 Flow cytometric confirmation of piRFP-2A-PU.1 and pλBx5 expression in transfected HEK293 cells.	128
Figure 2.9 Schematic diagram for characterization of PU.1/DNA complex.	132
Figure 2.10 A schematic diagram of a synthetic PU.1-dependent EGFP reporters for characterization of PU.1 self-regulation mechanism.	134
Figure 2.11 Schematic representation of the reporter gene constructs for characterizing ligands against PU.1 activity.	138
Figure 2.12 Flow cytometric data to confirm the lentiviral particles from the collection.	140
Figure 2.13 PU.1 transactivation on its specific gene confirmed by lentiviral particles, LV-λBx5-D2-EGFP.	141
Figure 2.14 PU.1 transactivation on its reporter gene confirmed by lentiviral particles, LV-Cd11b-D2-EGFP and LV-Csflr-D2EGFP.	142
Figure 2.15 The Stable cell line confirmation of the second-generation PU.1-dependent reporter in THP-1 cell line.	143
Figure 3.1 Sequence variation of critical residue 236 in the most conserved structural element, recognition helix H3 among the ETS family transcription factors.	157
Figure 3.2 Evolutionarily sensitive residue Asn ²³⁶ in H3 PU.1 ETS domain evaluated in cellular reporter gene platform.	159

Figure 3.3 Functional inhibition of PU.1 transactivation by dominant-negative mutants, evolutionarily conservative H3 chimeras in live cells.	160
Figure 3.4 Schematic representation of negative feedback in PU.1 transactivation.....	162
Figure 3.5 Adjustment of transgenic PU.1 dosage for cellular titration experiments in HEK293 cells to native PU.1 expressing myeloid cells.....	164
Figure 3.6 Functional PU.1 trans-regulation by negative feedback.....	165
Figure 4.1 Schematic diagram strategy of the DNA minor groove ligands against PU.1 for AML therapeutics.	183
Figure 4.2 Chemical structure of the DNA minor groove binders.	185
Figure 4.3 Functional PU.1 inhibition by heterocyclic diamidines.	193
Figure 4.4 Cytotoxicity of all heterocyclic diamidines against PU.1 activity.	196
Figure 4.5 Subcellular localization of heterocyclic diamidines in live unfixed HEK 293 cells.	198
Figure 4.6 Levels of DB2313 in the mouse serum after intraperitoneal (I.P.) injection of wild- type mice at different time points (n=4).	201
Figure 4.7 The second-generation PU.1-dependent reporter system for characterizing ligands against PU.1 activity.	203
Figure 4.8 Functional characterization of PU.1 inhibition by DB2115 and DB1976 by cell reporter assay (THP-1 λ Bx5-D2-EGFP).....	207
Figure 4.9 Inhibition of cellular EGFP fluorescence by DNA minor-groove binder.	209
Figure 4.10 Comparison of two different techniques for measurement of functional characterization screening ligands against PU.1.....	211

LIST OF ABBREVIATIONS

Deoxyribonucleic acid (DNA)
Ribonucleic acid (RNA)
Transcription factor (TF)
cis-regulatory elements (CREs)
DNA binding domain (DBD)
Acute Myeloid Leukemia (AML)
Helix-tern-helix (HTH)
Transcription factor binding sites (TFBS)
Chromatin Immunoprecipitation coupled (ChIP)
ETS (E26 transformation-specific) transcription factor family
PU.1(Purine rich Box 1)
Nuclear magnetic resonance (NMR)
Full-length (FL)
Irf-4 (interferon regulatory family transcription factor)
Macrophage-colony-stimulating factor receptors (M-CSFR)
Granulocyte-macrophage colony-stimulating factor receptors (GM-CSF)
Electrophoretic mobility shift assay (EMSA)
Systematic Evolution of Ligands by Exponential Enrichment (SELEX)
Polymerase chain reaction (PCR)
Reverse transcription PCR (RT-PCR)
Fluorescence resonance energy transfer techniques (FRET)
Western blot (WB)
Enzyme-linked immunosorbent assay (ELISA) assay
Chloramphenicol acetyltransferase (CAT) assay
Alkaline phosphatase (AP)
Secreted alkaline phosphatase (SEAP)
Infra-Red Fluorescence Protein (iRFP)
Enhanced Green Fluorescence Protein (EGFP)
ETS binding site (EBS)

1 GENERAL INTRODUCTION

1.1 Inherited traits transfer from one generation to the next

Cells that are the minimal unit of life respond appropriately to environmental signals. The cellular responses in a cell display proliferation, development, differentiation, metabolic process, and apoptosis. In 1958 Francis Crick conceptualized essential life, saying that the Central Dogma of molecular biology postulates a transfer process of the coded genetic information from nucleic acids to proteins in cells (1). The genetic information in DNA is inscribed on a strand of messenger RNA (mRNA) and uses this information to string amino acids together into a protein (1). Thus, DNA retains the complete genetic information that defines the structure and function of an organism. A gene contains the genetic information unit to generate a functional gene product, proteins. In the first of these processes, transcription, DNA sequences are transcribed into mRNA. Then, during translation, the mRNA is translated into its particular protein. During replication, the conservation process, a double-stranded nucleic acid is duplicated to give identical copies (2).

The transcription is the initial regulating point for expressing a gene; it should be highly specified for its function and fate in cells. Therefore, each step of transcription tightly controls (Figure 1).

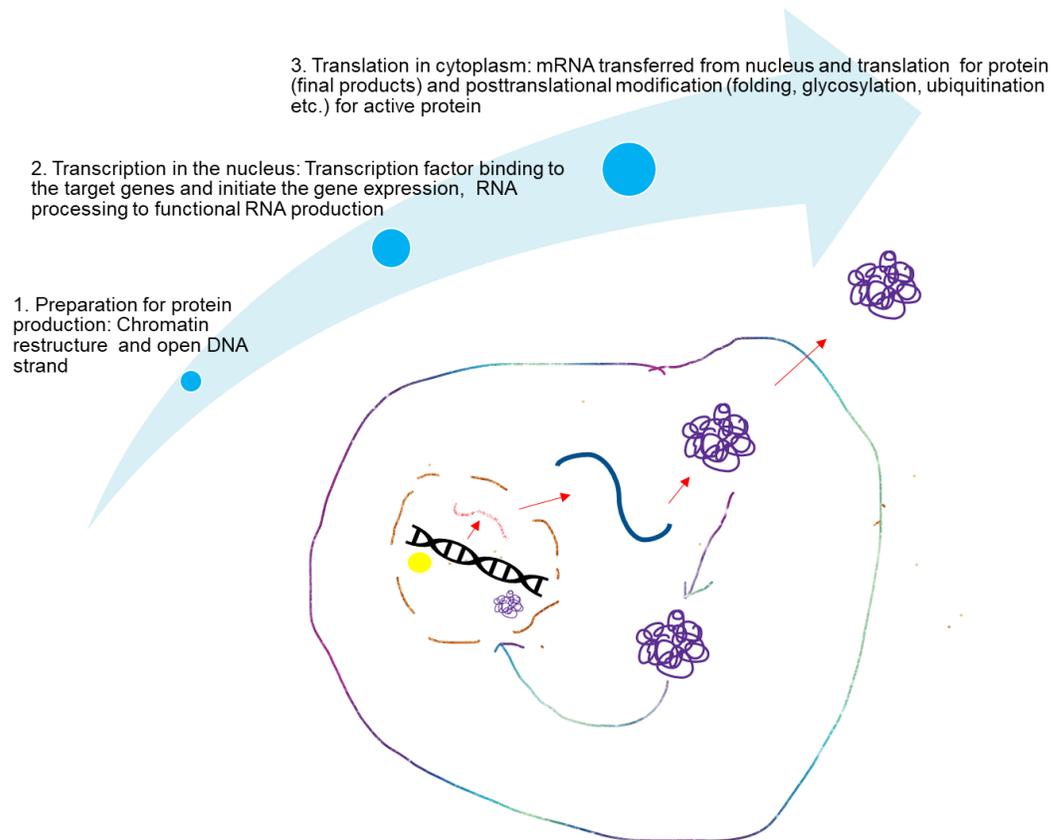


Figure 1.1 Overview of the flow of genetic information from DNA to protein in eukaryotic cells.

First, chromatin remodeling to open DNA for preparation of transcription. Both coding and noncoding region of DNA are transcribed into pre-mRNA during the initial mRNA processing by transcriptionionary regulatory complex (yellow circle). After RNA processing, mature mRNA is prepared for export out of the nucleus through addition of an endcap and a poly A tail. Once mRNA is transported into the cytoplasm, the mRNA is translated into pre-mature protein. Through post-translational modification, an active protein is formed and functions. The active protein is delivered to the target site for its function. The figure was inspired by @2010 Nature Education.

1.2 Transcriptional process and control

1.2.1 Complexity of tight regulation in gene expression

Transcription, responsible for initiating the expression of genetic information in cells, is tightly regulated to ensure that genetic programs are adapted to cell requirements (3). Gene expression produces the specified functional product, which is called protein, based on the cells'

genetic information encoded in a gene. Since the proteins are generated in the specified and crucial for the identity of the cells, gene expression is a sophisticated process; many factors participate in the gene expression for coordinating multi-level regulation. In living organisms, regulation of gene expression is vital because the function and structure of a specific cell are defined and developed into particular organisms. The gene expression is tightly regulated by multiple molecular systems, including transcription factors that bind to specific DNA sequences and chromatin remodeling that enable or block transcriptional machinery from navigating to their binding sites on DNA. Histone or the proteins in chromatin around which DNA is packaged determined whether the transcription factor can access their target DNA or not. Epigenetic control, the process of modifying histones, controls either downregulation or upregulation of gene transcription. Either methylation or acetylation mainly occurred for the prevention or promotion of transcription and translation. As a result, particular external stimuli, particular genes, and proteins are expressed, and cells are differentiated into various mature cell types. Foremost, all the control processes of gene expressions in a cell are entirely different combinations of transcription factor activities at the initial stage of transcription.

The gene expression is controlled at multiple levels of cells. The expression of genes is upregulated or downregulated at the cell level at a specific time and particular situation from the environmental signals so that the processes drive cells to differentiate and develop, respond, and adapt to environmental stimuli. Gene regulation was classically studied, focusing on regulatory units of protein, transcription factors, cofactors, or interactions between them and cis-regulatory units, the DNA binding sites, called promoters or enhancers. Nevertheless, the studies have intensively highlighted the complexity of the gene regulation revealed due to the advanced research technology. For example, the restructuring of chromatins for opening up the

transcription factors for its specific DNA binding sites, or the histone, which protein protects DNA strands, are modified with chemical changes, including methylation or acetylation. Both the chemical changes and structural changes add the highly organized regulation process into the complexity. Therefore, the process for the regulation of gene expression is highly complicated with multilevel control points.

This introductory chapter will briefly overview the transcriptional and post-transcriptional regulation, list the primary database resources that can be used for transcriptional and/or post-translational regulation data, and list the main tools to predict TF and miRNA gene targets (4).

1.2.2 Epigenetic controls for gene expression

Epigenetics is defined as DNA modifications at the molecular level. Through epigenetic control, gene activity is regulated independently of DNA sequence and mitotically stable. In epigenetics, the genetic information is not altered or changes. The biochemical and structural changes happen and open the accessibility for gene expression in a local area where it is needed to be (3). As a result, epigenetic mechanisms influence various biological and biochemical processes (5). In addition, there has been increasing attention to epigenetic landmarks due to health and diseases. Consequently, epigenetic techniques have rapidly expanded to detect chromatin conditions from specific analysis in a local spot to global view through the genome-wide sequencing analysis (6) .

1.2.2.1 DNA methylation in epigenetic control

DNA methylation is the most critical mechanism in epigenetic control. DNA methyltransferases (DNMTs) are the primary process for methylation. Many DNA methylation processes are classified into different modifications performed by the different

methyltransferase, including 5-methylcytosine (5mC), N6-methyladenine (6mA), and 4-methylcytosine (4mC) (7). The 5mC modification happens by DNMTs in which the enzymes transfer a methyl group to the C5 position of cytosine within CpG dinucleotides by DNMTs (8, 9). Many CpG dinucleotides are also found in the *cis*-regulatory regions to establish CpG islands, for involving gene regulation (Figure 1.2a) (10). Thus, DNA methylation is considered a critical process for disease progression. Abnormal DNA methylation is often linked with developing the disease in various ways. The most well-known process of DNA methylation happens for the regulation of the gene expression by controlling the methylation position of CpG islands in the area of a particular gene where the gene expression is regulated. In the promoter, a chromatin structure is closed or compacted by hypermethylation of CpG. The hypermethylation prevents transcription activation. Interestingly, hypomethylation of CpG islands causes the chromatin structure opening for transactivation of the specific gene.

1.2.2.2 The developmental significance of DNA methylation

DNA methylation is also essential in altering many biological processes such as embryonic development, cellular differentiation, and proliferation. Gene expression and DNA methylation are inversely correlated in many genes that must be developmentally controlled following cell-type specification in early development (7). Genome-wide studies characterizing the *de novo* hypermethylation of promoters during differentiation in mouse models represented 5% of the CpG islands as hypermethylated and, consequently, silenced in somatic tissues, but not in germline cells (10-16). Also, DNA methylation establishes patterns of a monoallelic gene expression (17-19). In the example of the X chromosome, inactivation on one of the two X chromosomes in female cells is repressed. In genomic imprinting, epigenetic chromosomal modification enables differential gene expression according to which parent transmitted the

chromosome to the progeny. Tissue-specific DNA methylation in somatic cells is significantly differentially methylated between normal tissues and cell types. Therefore, aberrations of the DNA methylation system have a vital role in human disease (20). Aberration in the promoter regions of critical oncogenes directly leads to tumorigenesis by doing that tumor suppressor gene silencing and/or oncogenes activate (21, 22).

1.2.2.3 Chromatin modifications in epigenetic control

A crucial epigenetic mechanism in regulating gene expression is chromatin modification (23). During post-translation, histone is modified through chemical processes such as acetylation and methylation. It is known that the process of acetylation and methylation alters the particular lysine residues of core histone tails (24). Histone modification gives a distinctive identity to the nucleosome that controls activity in transcription during various biological phases (25, 26). Generally, histone acetyltransferases (HATs) are the enzyme for acetylated histone. The HATs work acetylated histones to increase for gene transactivation; however, the deacetylation process by histone deacetylases (HDACs) is associated with turning on the gene and silencing (27-29). Also, the residue of the histone methylation is considered for turning on or off the expression of a gene by histone methyltransferases (HMTs). For example, gene expression is repressed as lysine residue K9 of histone H3 is methylated. Lysine residue K4 of histone H3 is methylated as turning on the gene expression. The idea is that histone modification causes the alteration of chromatin structure. It triggers the modification of the DNA polymer's spatial conformation and openness of DNA binding proteins and/or epigenetic modifiers to the cis-regulatory region, and consequently, gene transactivation is changed (25). They also work in keeping chromatin structures that allocate epigenetic heritage (30).

1.2.2.4 Non-coding RNA (ncRNA) in epigenetic control and its clinical importance

Non-coding RNAs are not like mRNA, which enables the production of functional protein. Nc RNA is not coding RNA transcripts (31-33). Even though they are not translating to the functional product, ncRNAs are known for their critical role in regulating epigenetic mechanisms, significantly modifying chromatic configuration through many ncRNAs. It results in whether a gene is upregulated or downregulated at gene expression level at transcription and post-transcription; the control allows cells to function normally (34, 35). Generally, two ncRNA are classified by size: short ncRNA (<30 nt) and long ncRNA (lncRNA, >200 nt). The kinds of short ncRNA consist of microRNA (miRNA), small interfering RNA (siRNA), and piwi-interacting RNA (piRNA) (34).

Types of ncRNA

MicroRNAs (miRNA) target a complementary strand of mRNA and attach to it. It leads to cleave, degrade or prevent the mRNA from being translated into the protein. For instance, miRNA genes mir-127 and mir-136 were shown to control the Rtl1 gene, in which its genetic information is transferred to the next generation in the developmental process in mice (36). No Rtl1 protein expression due to improper epigenetic modification might lead to fetal death in mice. Short interfering RNA (siRNA) also functions to mediate post-transcriptional gene slicing resulting from mRNA degradation. In addition, siRNA induces heterochromatin formation. SiRNA promotes H3K3 methylation and chromatin condensation through an RNA-induced transcriptional silencing complex (37-42). Piwi-interacting RNAs (piRNA) are called because they interact with the piwi family of proteins. They mainly function in the germline or somatic cells, control chromatin, and containment of activity of transposons (43, 44). The piRNAs split the transposon and PIWI-proteins complex and several lncRNAs and chromatin-modifying

proteins and engage in other proteins. This enables manipulating gene expression from altering chromatin conditions (34, 45). Inc RNA participates in reconstructing chromatin and controls the transcription and post-transcription process, particularly in development processes. NcRNAs work with mRNA, ncRNAs, DNAs, and proteins to control gene expression in epigenetics (46).

The role of the ncRNAs affects determining gene function in its specificity for cellular activities and identity; consequently, ncRNAs act as a crucial controller of disease progression in many types of cancer (31). Tumor progression involves genomic modifications, rearranging chromosomes that trigger a fusion of two genes in the genome. The fusion between two oncogenes results in genetic alterations and makes unusual circular RNAs (circRNA). This abnormal circRNA has been found in various leukemias and cancers (47). In addition, altered adenosine-to-inosine editing by dsRNA-specific adenosine deaminases (ADAR) is considered to alter the structure of transcripts and an important mechanism for tumorigenesis (48-50). The miRNA is often discovered in the type of tumorigenic drivers due to its missing or addition and low expression of ADAR that regulate miRNA targets, and miRNA precursor was reported in the metastatic ability of melanoma.

Contrastingly, the ADAR in leukemic stem cells (LSC) promotes LSC self-renewal catalytic (51). Therefore, the ADAR might be a good target for therapeutics as well. The *cis*-regulatory sequence in the 3'UTR variation in tumors changes miRNA aiming in ncRNA systems because miRNA is lost and 3'UTRs turns into the oncogenic transcripts (52). They are also sensitive to single nucleotide polymorphism (SNPs), and it may affect miRNA-binding sites associated with human cancers such as lung cancers and pancreatic cancers.

Clinical importance of ncRNA

Understanding the molecular process of the ncRNAs engagement extensively in gene regulatory networks becomes essential (46, 53). A genome-wide scale assay can provide the chance to understand the critical mechanism of developing the disease; single-targeted therapy often proves insufficient as an effective cancer treatment. To comprehensively understand the molecular level of epigenetic mechanism, which is often associated with progressive disease, it is essential to determine complexes of genetic regulation and network in each pinpoint to improve the therapeutic intervention center (54).

1.2.3 Transcriptional regulation

Transcriptional regulation of gene expression is an essential part of tissue-restricted gene expression reacting to environmental factors from the outside. Transcription factors (TFs), the central regulators of gene expression. TFs bind their target DNA sequences to control transcription and are engaged in forming the pre-initiation complex and recruiting RNA polymerase. Many TFs interact with their cognate DNA motifs in a eukaryotic living organism and work with cofactors to modify the chromatic environments. Thus, many factors influence their evolution. They exhibit their evolutionary traits in gene regulatory elements, molecular cofactors, and epigenetic mechanisms. All the regulatory factors shape a system for gene regulation (Figure 1.2).

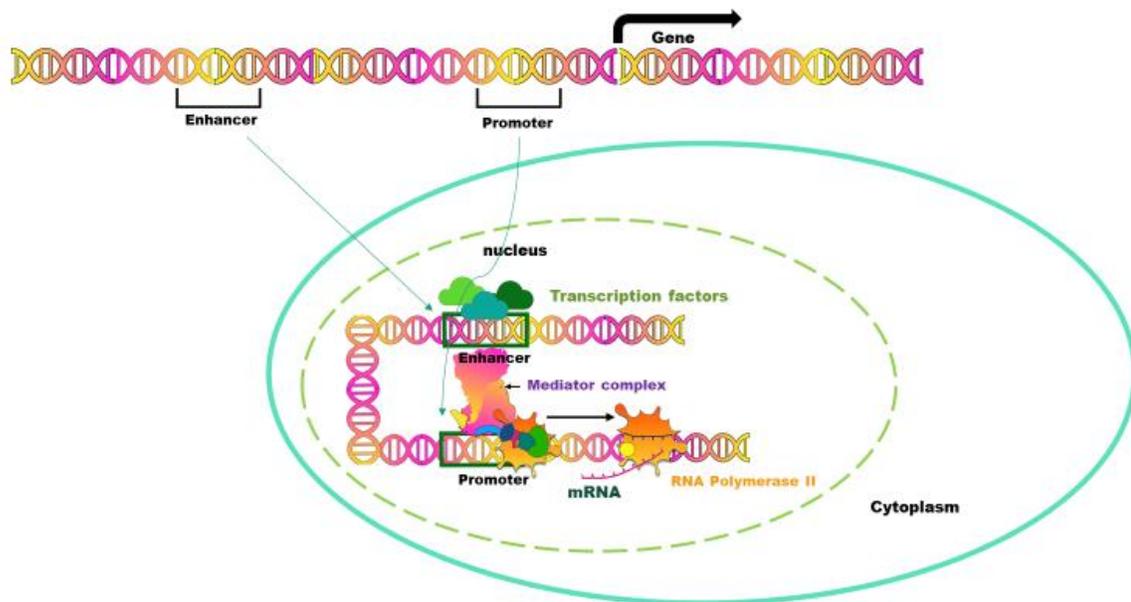


Figure 1.2 Complexity of gene regulation in a eukaryotic cell

In eukaryotic cells, the transcription process is mediated the transcriptional regulatory complex. In the nucleus, transcription factors (transacting molecules) bind specific DNA sequence regions (cis-acting regulatory regions, promoters or enhancers) for gene expression. An enhancer is a distal region for gene regulational point and bound by transacting molecules then DNA looping is formed by mediator compelx in order to activate the target gene expression (55).

Discovering the molecular mechanisms for the regulation can provide valuable knowledge on evolutionary processes and the progress of phenotypic adaptations.

1.2.3.1 Transcriptional regulatory elements

The *cis*-regulatory elements (CREs) consist of promoters, enhancers, and silencers (56). The DNA binding domain (DBD) of TFs enables TF to bind DNA. TFs utilize various DNA-binding structural motifs to identify their target sequences. TFs are categorized using such DBD, including helix-turn-helix (HTH). The interaction between DNA and TFs goes beyond the structural and sequence level since several other factors involve complicated gene expressions for combinatorial factors such as epigenetic modifiers and other TFs or cofactors. Thus, gene

regulation involves many molecular mechanisms. Therefore, an intensive analysis of the TFs, which considers the interaction with others, and how TFs influence other cofactor proteins' evolution, is vital to insight into the phylogenetic relationship among organisms.

Transcriptional regulation of genes in humans relies on promoters and nearby CREs and distal regulatory elements such as enhancers, insulators, locus control regions, and silencing elements. For example, transcriptional enhancers are elements that activate or deactivate gene expression upon cellular environment signals, including time and location of gene expression. In such a sense, enhancer instructs tissue-specific gene expression, development, homeostasis, and cellular integrity. Several DNA regulatory elements interact with chromatin and transcription factors for controlling gene expression.

1.2.3.2 Cis-acting regulatory elements (CRE)

Most of the human genome consists of non-coding regions containing *cis*-regulatory elements (CREs), including promoters, enhancers, silencers, insulators, etc. CREs, the binding site for the TFs, are placed adjacent in a gene necessary for expressing at the proper time and space. CREs have essential roles in cell developments such as specific cell types, conditions, and development stages. Abruption to the CREs could weigh into gene expression alteration; therefore, precisely classifying regulatory elements is vital to unveiling the fundamental procedures for gene regulation.

The CREs functionally diverged (56). The divergence has been shown by minor nucleotide (nt) changes, additions of the nt, and/or missing nts. The mutations disturb TFs' interacting with their target DNA sites. epistatically. The diverging effect of the evolution is founded in enhancer activities working with TFBS displaying in ancient enhancers. Therefore, illuminating the specific genetic alterations accountable for a *cis*-regulatory variance has

presented perceptiveness into the relative impacts of genetic alterations and the effects of the changes. *Cis*-regulatory happenings are intricate activities that include access for the chromatin, TFs' interaction, epigenetic modification. Promoters are defined as DNA sites less than 100 bps of initiation at transcription; they include regulatory elements (DNA motifs) required to be bound by RNA polymerase.

Enhancers are another type of *cis*-regulatory element. Enhancers are often located more distal to the transcription start site. The enhancer cooperates with site-specific TFs to determine the cell-type character and control gene expression. The complex of TFs that attach enhancers alleviate or prevent the RNA polymerase II's binding activity. Numerous enhancers can perform redundantly to control transcriptional activation. Also, enhancers may work individually to form DNA looping at distal regions from the target genes. Therefore, it is challenging to discover enhancers and link them to their target genes based on DNA sequence alone. Super-enhancers exist that groups of enhancers TFs dominate the regions.

The DNA sequence of the enhancer encodes its function in the way of enhancer syntaxes such as the order, orientation, and spacing of TF's binding sites (57). The enhancer syntax is essential for tissue-specific gene expression (55). Traditionally known enhancer grammar, physical interaction between factors and the DNA, and interactions of TFs and their cofactors are thought to create a functional unit of enhancer configuration. The features of enhancer grammar include TFBS type within the enhancer sequence, number, affinity, order, spacing, and orientation. Specifically, the enhancer syntax, including particular spacing between the DBD, revealed its significance for mouse hematopoiesis.

Enhancer interacts with transcription factors and control genes

Enhancers that TFs weakly bind to the site can facilitate robust particular gene expression patterns in tissue when they are arranged with ideal enhancer syntax. The significance of sequence restrictions within enhancers in the development of particular tissues suggested the evolution of the enhancers depending on the selection of TFBS (58). Also, ideal TFBS is identified as genes having proper spacing and strongly bind concerning expression levels determine enhancer function (55). For example, native GATA (ectoderm determinant) and ETS (FGF signaling, pleiotropic fibroblast growth factor) binding sites containing imperfect matches to consensus motifs showed optimized expression combining these motifs with optimal spacing robust expression.

Constraints on the spacing between TFBSs may ensure all TFs within an enhancer can access their sites or spacing. The spacing could be essential to allow both TF-DNA interactions along with protein-protein interaction such as binding dimers in the example of dimeric Notch transcription complexes on promoters with paired sites for a requirement to transcription activation (59). In IFN- β enhancers and several developmental enhancers, the helical phasing of sites within enhancers is essential for gene expression. Even single base-pair changes in spacing influence the level of expression. Further, spacing between binding sites can govern the location of the gene expression in studies of the growth hormone expression in the somatotroph cells in the developing mouse pituitary gland (57, 60).

Genome-wide studies and comparative genomic approaches to study enhancer grammar are functionally supported by reporter gene assay. For example, by the reporter gene assay of the altered DNA specificity on Bicoid mutants, bicoid distinguished among related DNA targeting sites by particular interaction between residue 9 of its recognition alpha-helix and base pairs 7 of

the same helix sites (61). From the studies, the distance between binding sites is critical to DNA recognition.

1.2.3.3 *Trans-acting regulatory elements*

Transcription factors are primary mediators of gene regulation (62). TFs identify and bind to their own cognate DNA site in TFBS. The functional outcomes characterize the transcriptional systems for cells' growth or reaction to their environment. However, the doctrines that control how TFs pick functional TFBS are needed to research further. Furthermore, the connection between TF bindings and their functional outcomes is also vague in which TF binding may influence the expression of a nearby gene or the directionality of such a change.

The complication of gene regulation at the transcription level can be demonstrated by assessing the gene's number and its locations of *cis*-regulatory elements in eukaryotes (62-65). In *Drosophila*, quite a few enhancers are founded and utilized. In contrast, yeast has no enhancer and uses one UAS sequence for a gene. Distal regulation of *cis*-regulatory elements indicates a higher level of organism use this mechanism in developing or differentiating a cell. The yeast contains a few hundred TFs, whereas humans contain about ten times TFs higher than that of the yeast. TFs per gene utilization are highly complex and have vast numbers of translating their functional outcomes of gene expression, granting a remarkable upsurge in various organisms. Restructure of chromatin governs the steadiness of post-translational products, and its control also added the intricacy of the gene expression to the living organism. Therefore, various genes that are distinct, complex, and essential organisms from each other should be considered (66, 67).

1.2.4 *Specific DNA sequences recognition by TFS*

The evolution of elaborated systems might originate from the complexity of cell differentiation in gene regulation. The transcriptional regulatory complex comprises quite a small *cis*-regulatory DNA sequences or motifs less than eight base pairs. Most eukaryotic genes have promoters containing TATA-box close to the 5' end of the gene. Also, many genes for their expression need to have enhancers. Enhancers involve a critical transcriptional process the upstream or downstream. Consequently, specific RNA polymerase II promoters can be affected by several enhancers and by multiple co-factors bound to the CREs.

DNA recognition by TFs occurs in the CREs region. Often, a shift in the TF configuration will go along with DNA binding, in which the recognition motif (about 6 to 10 base pairs) is a pivotal spot to interact to shape the conformation. After TFs form the transcriptional complex, they work together with other bound TFs or coactivators for RNA polymerase II's binding. TFs control many essential parts of progress for cell differentiation; therefore, the effects from a deletion of a TF gene in the organism are detrimental.

Transcription factors exert combinatorial control

Many TFs are known to facilitate transcription at hundreds of different promoters. The intricacy of DNA expression in eukaryotes comes from combinatorial effects (65). The combinatory effects of chromatin and TFs rather than the individual TFs are screened for transcription initiation. Once they are combined, changes to the functional domains of a TF and/or communicating with cofactors can regulate whether a gene is expressed or not or whether RNA polymerase II binds or not. Also, an enhancer is located up to five hundred bps upstream and contains multiple TFBS. It allows different TFs binding as a form of dimers, and DNA is bent. Structural conformation of chromatin lets the transcriptional regulatory complex form as

activators. Therefore, TFs are considered a binding element between CREs by complexation to tie with cofactors.

1.2.5 Significance of DNA target recognition by transcription factors

Gene expression is highly controlled during cellular development, growth regulation, proliferation, and even disease occurrences. By assuring the accurate expression of particular genes, the transcriptional regulatory system governs such biological processes. In the central regulatory system, transcription factors govern the principal mechanisms to maintain life activities. Sequence-specific TFs are critical controllers of the critical biological processes by binding to CREs such as enhancers or promoters at the initial step, transcription. To initiate the vital function is how TFs bind to the target DNA sequence. Therefore, understanding the functional characterization of the TF-DNA binding is the first step to shape the gene regulatory network.

1.2.5.1 Transcription factor and spectrum of diseases

TFs are often involved in the oncogenic process, including initiation, development, and invasion. As a result, targeting these proteins is a theoretically ideal avenue for cancer treatment research. Unfortunately, only about 20% of TFs are established as anti-cancer therapeutics because most TFs are undruggable targets (68). However, with the development of new pharmaceuticals and new delivery processes, it is possible to develop viable therapies targeting new transcription factors.

1.2.5.2 Targeting strategies of transcription factors for drug development

Three approaches to targeting TFs currently have the most potential.

First, a treatment could be developed that directly targets the functional pocket of the protein. For example, this concept is applied to chemical reagent All-Trans Retinoic Acid

(ATRA) for treating alpha-rearranged acute promyelocytic leukemia (APL). In this cancer type, the cell differentiation process is blocked. Therefore, ATRA can target the RAR α hormone/retinoid receptors, unblocking the cell differentiation process (69).

An alternative is to target protein-protein interactions (70). This can be seen in the treatment of human malignancies. For example, the treatment for more than 50% of cancers, including breast cancers, involves well-known oncoprotein p53 (71-73). In targeting these proteins, the treatment aims to interfere with TF's interaction with other coactivators. However, a TF acting as a monomer or having no known heterodimer or cofactor cannot be targeted at the protein-protein interaction level. Another option is to directly block the interface between protein and DNA where DBDs of TFs interact with their target DNA as a competitor for the bindings.

A final approach to discuss is the potential to target TF- DNA binding domains. By itself, DNA is a drug target for cancer therapy (74-77). Following the first DNA alkylating compound, DNA intercalators or major/minor groove DNA ligands were developed, and a "targeted chemotherapy" against DNA was prevalently researched based on the concept of "DNA as a drug target". Proliferation and differentiation blockade are two outcomes of the DNA targeting for disturbing DNA-binding proteins complex such as TFs associated with cancer development. One example of sequence-specific targeting is a small compound targeting the DNA binding site of transcription PAX2 (78-80). This approach of blocking targeting sequence-specific sites is to design small molecules to bind the TF's binding sites to compete with TFs. The minor groove DNA binding ligands are based on the idea. Mithramycin, one of the binders for the sequences in the DNA minor groove, is one of the successful examples in the stage of preclinical screening, and diamidine heterocyclic derivatives are actively researched for targeting transcription factors. Heterocyclic diamidines interrupt the TF that bind to their target DNA sites, such as DB293

targeting TF Pit-1. ERG, one of the ETS family TF, is targeted by DB1255 (81, 82). HOXA9 and PU.1 also have been reported for their potential pharmacological controls by the small molecules (83, 84).

Interestingly, some DNA minor groove ligands exhibited engaging antileukemic activities by targeting sequence-specific DNA recognized by TFs relating to AML progression: the acylating agent-conjugated pyrrole-imidazole polyamide and heterocyclic diamidines (85, 86). Pyrrole-imidazole polyamides recognize the minor groove of specific DNA sequences noncovalently through the interaction between their pyrrole and imidazole pairs interlocked by a hairpin linking. Once Chb-M binds to RUNX-binding sequences, the downregulation of various RUNX signaling pathways induces apoptosis in tumor cells. HOXA9 is a TF involved in differentiation blockade so that it is associated with leukemic occurrence; therefore, up to 70% of AML patients show HOXA9 overexpression and accumulation of leukemic blasts (83). Depauw and their teams reported that DB818 and DB1055 interrupted HOXA9 to bind its specific DNA (83). Munde and their colleagues showed a successful example of heterocyclic diamidines to TF PU.1 (Purine rich Box 1), a member of ETS family TF, and controlled their transcriptional activity (84). In the study conducted by Depauw and Munde, the therapeutic possibility of transcription protein interference by binding the heterocyclic diamidines' target binding sites was confirmed. The small compounds' pharmacological applications were demonstrated as well (81, 87-89).

1.3 The ETS family of transcription factors

1.3.1 The importance of understanding the DBD of the TF family

Gene expression is highly controlled during cellular development, growth regulation, proliferation, and even disease occurrences. By assuring the accurate expression of particular

genes, the transcriptional regulatory system governs such biological processes. In the central regulatory system, transcription factors govern the principal mechanisms to maintain life activities. Sequence-specific TFs are critical controllers of the critical biological processes by binding to CREs such as enhancers or promoters at the initial step, transcription. To initiate the vital function is how TFs bind to the target DNA sequence. Therefore, understanding the functional characterization of the TF-DNA binding is the first step to shape the gene regulatory network.

TFs are often involved in the oncogenic process, including initiation, development, and invasion. As a result, targeting these proteins is a theoretically ideal avenue for cancer treatment research. Unfortunately, only about 20% of TFs are established as anti-cancer therapeutics because most TFs are undruggable targets (68). However, with the development of new pharmaceuticals and new delivery processes, it is possible to develop viable therapies targeting new transcription factors.

1.3.1.1 Targeting strategies of transcription factors for drug development

Three approaches to targeting TFs currently have the most potential.

First, a treatment could be developed that directly targets the functional pocket of the protein. For example, this concept is applied to chemical reagent All-Trans Retinoic Acid (ATRA) for treating alpha-rearranged acute promyelocytic leukemia (APL). In this cancer type, the cell differentiation process is blocked. Therefore, ATRA can target the RAR α hormone/retinoid receptors, unblocking the cell differentiation process (69).

An alternative is to target protein-protein interactions (70). This can be seen in the treatment of human malignancies. For example, the treatment for more than 50% of cancers, including breast cancers, involves well-known oncoprotein p53 (71-73). In targeting these

proteins, the treatment aims to interfere with TF's interaction with other coactivators. However, a TF acting as a monomer or having no known heterodimer or cofactor cannot be targeted at the protein-protein interaction level. Another option is to directly block the interface between protein and DNA where DBDs of TFs interact with their target DNA as a competitor for the bindings.

A final approach to discuss is the potential to target TF- DNA binding domains. By itself, DNA is a drug target for cancer therapy (74-77). Following the first DNA alkylating compound, DNA intercalators or major/minor groove DNA ligands were developed, and a "targeted chemotherapy" against DNA was prevalently researched based on the concept of "DNA as a drug target". Proliferation and differentiation blockade are two outcomes of the DNA targeting for disturbing DNA-binding proteins complex such as TFs associated with cancer development. One example of sequence-specific targeting is a small compound targeting the DNA binding site of transcription PAX2 (78-80). This approach of blocking targeting sequence-specific sites is to design small molecules to bind the TF's binding sites to compete with TFs. The minor groove DNA binding ligands are based on the idea. Mithramycin, one of the binders for the sequences in the DNA minor groove, is one of the successful examples in the stage of preclinical screening, and diamidine heterocyclic derivatives are actively researched for targeting transcription factors. Heterocyclic diamidines interrupt the TF that bind to their target DNA sites, such as DB293 targeting TF Pit-1. ERG, one of the ETS family TF, is targeted by DB1255 (81, 82). HOXA9 and PU.1 also have been reported for their potential pharmacological controls by the small molecules (83, 84).

Interestingly, some DNA minor groove ligands exhibited engaging antileukemic activities by targeting sequence-specific DNA recognized by TFs relating to AML progression: the acylating agent-conjugated pyrrole-imidazole polyamide and heterocyclic diamidines (85,

86). Pyrrole-imidazole polyamides recognize the minor groove of specific DNA sequences noncovalently through the interaction between their pyrrole and imidazole pairs interlocked by a hairpin linking. Once Chb-M binds to RUNX-binding sequences, the downregulation of various RUNX signaling pathways induces apoptosis in tumor cells. HOXA9 is a TF involved in differentiation blockade so that it is associated with leukemic occurrence; therefore, up to 70% of AML patients show HOXA9 overexpression and accumulation of leukemic blasts (83). Depauw and their teams reported that DB818 and DB1055 interrupted HOXA9 to bind its specific DNA (83). Munde and their colleagues showed a successful example of heterocyclic diamidines to TF PU.1 (Purine rich Box 1), a member of ETS family TF, and controlled their transcriptional activity (84). In the study conducted by Depauw and Munde, the therapeutic possibility of transcription protein interference by binding the heterocyclic diamidines' target binding sites was confirmed. The small compounds' pharmacological applications were demonstrated as well (81, 87-89).

1.3.2 The ETS transcription factor's redundant binding on DNA in living cells

Genome-wide analyses of ETS-factor occupancy revealed that individual ETS proteins in living cells have high similarity in DNA binding regions. Genomic technology such as chromatin immunoprecipitation combined with either microarray (ChIP-chip) or massively parallel sequencing (ChIP-Seq) was employed to determine TF-bound DNA regions (109). Although the assay caveats must be considered to understand and interpret the data (i.e., cell types or growth conditions and analysis of quality of the peak), the redundancy for the bound regions is observable with statistically significant frequency. Overlapping binding within the resolution could result from clusters of binding sites specific for each family member. Still, DNA regions bound by many ETS proteins were shown in a short distance upstream of transcription start sites

and are correlated with attributes of active promoters. Also, these regions are near the housekeeping-types genes, indicating different ETS factors regulate the genes in different cell types.

Although all ETS DBDs are relatively highly conserved, not the core site of DBD determined the ETS TFs' specificity. Different functional ETS TFs gain their own specificity for their identity from the preferential binding on the different DNA flanking sequences (116, 117). For example, it has been found that the members of Class II are sensitive to swap A and T at position 7. In the case of PU.1, an A-rich track of sequences is its target in the 5'-flanking sequences near the core consensus sequences. NMR structural studies displayed that ELK decided its specificity detecting from structural configuration changes enacted by tertiary contacts of conserved residues in the DNA binding interface with nearby non-conserved residue. The specificity of Elk-1 sets a constraint on binding variants of the core binding site (118).

The subfamily of the ETS factor's specific binding on the DNA is tightly related to one or a few more DNA binding near the core (108, 119). From genome-wide study, ETS proteins ERG and FLI1 bind to a largely redundant set of targets, whereas PU.1 is not sharing a commonality with the targets of other ETS family members. In vitro analysis for DNA-sequence preference, Ets-1 in Jurkat cells is highly occupied by GABPA. However, PU.1 in myeloid cells and SPDEF in prostate cells were divergent from Ets-1 and other members. Therefore, even the ETS TFs are classified into the subfamily groups; the functional specificity and its DNA binding specificities differ within the members in the subfamily of ETS family members.

The significance of all genomic occupancy should be tested to understand more details by functional tests, especially ones that can differentiate its own physiological function and what are determinants for grants their identity. Also, many ETS DBD sites and their role for the

specificity determinants are projected based on understanding from biophysical and biochemical studies. Therefore, the ETS TFs' specificity is associated with the other proteins in the biological system and cooperation with them and research with the genome-wide assay.

ETS TFs display various cooperative partnerships for transcriptional machinery. For instance, PU.1 works with the Interferon regulatory family transcription factor (Irf-4) at the location of enhancers of the immunoglobulin light-chain gene ($Ig\lambda$) (120-122). As PU.1 exists near the enhancer genes, the DBD of Irf-4 for the λ B enhancer increase. The crystallography analysis displayed the structure of the two DBD on a composite enhancer sequence. The contact of the protein-protein interactions creates the DNA conformational changes, and the interactions is called cooperativity. In the case of Ets-1, its cooperativity arises from the interaction with RUNX1, and it releases from the autoinhibition state (123). In addition to the cofactors, other TFSs that enable binding multiple DNA sites have a partnership with the ETS TFS to control gene expression and shape their own cooperativity. The additional specificity of larger composite DNA-binding sites, gained through the cooperative binding of two factors, may signify the most remarkable biological influence of such partnership and cooperativity.

1.3.3 ETS family members and spectrum of diseases

The ETS factors regulate genes' positive or negative expression in all types of function for proliferating, developing, differentiating, influencing signaling cascades, dying in cells, even tightly involving in the oncogenesis. All ETS TFs are functionally distributed across the organisms and tightly regulated various physiological functions in the cells. PU.1 is a distinct role of the self-renewal of hematopoietic stem cells and specialized work in myeloid cells' differentiation. Therefore, abnormality of the ETS TFs and its dysfunction like PU.1

significantly influences diseases' progression, such as proliferative signaling, resisting cell death, evading growth suppressors, and activating invasion and metastasis (Table 1.1).

Table 1.1 ETS family of transcription factor and a variety of cancers

Class	Subfamily	Official gene symbol	Other names	Disease association	Aberration
I	ETS	ETS1	EWSR2	Breast, colorectal, gastric, lung, ovarian cancer	Overexpression; protein stabilization (through COP1); binding site mutation in TERT promoter
		ETS2		Hematological malignancies, breast, colon, head and neck, liver, pancreas, prostate and thyroid, glioblastoma, osteosarcoma and prostate	Amplification; overexpression; binding site mutation in TERT promoter; protein-protein interaction (p53)
I	ERG	ERG		Hematological malignancies melanoma, Ewing sarcoma, myeloid leukemia, prostate	Overexpression; TMPRSS2-ERG fusion; EWS-ERG fusion; FUS-ERG fusion
		FLI1		Prostate cancer, breast cancer, Ewing sarcoma, leukemia	Overexpression; loss of tumor-suppressive function; EWS-FLI1 fusion
		FEV	PET1	Ewing sarcoma	FWS-FEV fusion
I	PEA3	ETV4	E1AF, PEA3	Breast, Ewing sarcoma, pancreas, prostate	Overexpression, protein stabilization (through COP1), TMPRSS2-ETV4 fusion, DDX1-ETV4 and CANT1-ETV4 fusions, EWS-ETV4 fusion
		ETV5	ERM	Breast, prostate	Overexpression, protein stabilization (through COP1), TMPRSS2-ETV5 fusion
		ETV1	ER81	Breast, Ewing sarcoma, prostate	Amplification, overexpression, protein stabilization (COP1; KIT and MARK), TMPRSS2-ETV1 fusion, ASCL#-ETV1, C15ORF21-ETV1, FLJ35294-ETV1, ERVK6-ETV1, HNRPA2B1-ETV1 and

					SLC45A3-ETV1 fusions, and EWS-ETV1 identified
I	ETV2	ETV2	ER71, ETSRP71	Glioblastoma	
I	TCF	ELK1		Breast, prostate	Overexpression
		ELK4	SAP	Prostate	Overexpression
		ELK3	NET, SAP2	Breast, liver	Overexpression
I	ERF	ERF		Prostate	
		ETV3	METS, PE1	Breast	Amplification
		ETV3L			
IV	PDEF	SPDEF	PDEF, PSE	Breast, prostate, colon	Overexpression; confers endocrine resistance; loss of tumor-suppressive function
II	ELF	ELF1		Lung, prostate, endometrial and ovarian cancer	Overexpression
		ELF2	NERF	Breast	
		ELF4	ELFR, MEF	Ovarian cancer, hematological malignancies, and lung cancer	Overexpression; loss of tumor-suppressive function
II	ESE	ELF3	ESE1, ERT, ESX, JEN	Breast, colon, lung, prostate	Amplification; overexpression.
		ELF5	ESE2	Breast, prostate and urothelial cancer	Overexpression; confers endocrine resistance; loss of tumor-suppressive function
		EHF	ESE3, ESE3B, ESEJ	Ovary, pancreas, prostate	Overexpression, loss of tumor-suppressive function
II	TEL	ETV6	TEL, TEL1	Breast, gastric, head and neck, thyroid, hematological malignancies, T cell Leukemia	ETV6-NTRK3 fusion, ETV6-NTRK3, ETV6-PDGFRB, ETV6-RUNX1 and PAX5-ETV6, ETV6 mutations identified
		ETV7	TEL2, TELB, TREF	Hematological malignancies, Leukemia	Overexpression, loss of tumor-suppressive function
III	SPI	SPI1	PU.1	Hematological malignancies and Leukemia	Overexpression and PU.1 mutations identified
		SPIB		Lung, hematological malignancies	Overexpression

		SPIC			
I	ELG	GABP α		Prostate, Glioblastoma	Overexpression; binding site mutation in TERT promoter

The ETS family TFs members are associated with multiple types of human cancer in various ways (90). Higher ETS TFs have been reported in many types of cancers, including breast, lung, colon, pancreatic, and thyroid cancer. The overexpression of ETS TFs, including Ets-1 or Ets-2 in breast, prostate, and hematological cancers, has been documented (91). In AML, ERG overexpression has been found in a highly complicated manner associated with other factors in chromosomes (92). Moreover, in the case of ETV6 in B lymphoblastic leukemia, ETS TFs' expansion is often reported (93, 94). The ETS fusion has been shown in prostate cancer and leukemia. There is a spectrum of cancers where chromosomal translocation results in fusion proteins that contain the ETS DBD. In prostate cancers, ETS fusion genes causing by chromosomal translocation of 5 'TMPRSS2 to other ETS TFs' DBD genes. One good example of TEL, the variant from ETS translocation, displays in many hematological cancers..

ETS members and oncogenesis

Oncogenesis is associated with ETS TFs binding to CREs. The human telomerase reverse transcriptase (hTERT) gene expression is a representative example of mechanisms of ETS-driving and controlling tumor occurrences (95). The hTERT plays the role of mediating the ETS TFs' transactivation. The CREs of hTERT provide for their binding site where Ets-1 interacts with c-Myc; it is necessary for breast cancer cells' proliferation. The ETS TFs as a driving role is often found as the one base pair of mutation leading to the gain of function and provoke the cancer initiation. The oncogenetic transformation from the ETS TFS binding to their CREs in ERG or ETV6 is also documented in prostate cancers and Ewing sarcoma cancers.

The ETS TFs' cooperativity is often disrupted and leading to occurring cancer. For example, the ETS and c-Jun have partners and are functionally released from the autoinhibition state, and the disturbance of the partnership is associated with oncogenesis (96, 97). Interaction of AML-1 with Ets-1 is an example where TF interaction leads to an escalation in ETS TFs' potency and possibly involves carcinogenesis, such as the cause of myeloid and lymphoid leukemias (98). The rearrangement of the AML-1 gene often influences the partnership of AML-1 with Ets-1 and decreases the control of Ets-1 autoinhibition. The gene alterations and chromosomal translocations in AML-1 result in the disturbance of AML-1's target gene transactivation with PU.1. It negatively influences repress abnormal gene expression. It critically affects to differentiation of myeloid cells, like macrophage and granulocyte-macrophage colony-stimulating factor receptors (M-CSFR and GM-CSF), and consequently, it triggers into cells into pre-leukemic stages.

The PU.1 can act as a repressor or activator, and its dysregulation is associated with leukemia. PU.1 binds to the purine-rich sequence near the CREs of target genes and activates gene expression during myeloid and B-lymphoid cell differentiation in conjunction with other TFs and cofactors. PU.1 also serves as a repressor by interacting with GATA-1, a zinc finger TF mandatory for erythroid differentiation (99). GATA-1 and PU.1 bind together on the target sequence and govern cells' identity during hematopoiesis, but a stoichiometric balance between the two is broken, then it could lead to AML in mice (100). Interestingly, the genes are direct targets for PU.1-mediated activation, and GATA-1 mediate repression ($C/EBP\alpha$) is found the association in murine erythroleukemia (MEL) these target genes are downregulated in MEL by this repression and lead to MEL differentiation (99). Decreased PU.1 function is evident in acute promyelocytic leukemia (APL). In APL, the associated fusion protein promyelocytic

leukemia/retinoic acid receptor α (PML/RAR α) can reduce PU.1 serine phosphorylation and promoter binding (101). This causes a decrease in PU.1-dependent transcriptional activation and promotes leukemogenesis. In human AML, mutated RUNX or translocated RUNX1 and wildtype PU.1 interact and recruit co-repressors (102).

Considering that the ETS family of the transcription factor is the feature of oncogenesis, the ETS factors have been implicated in developing and progressing various types of cancer (103). The studies of the ETS TF's etiological role in cancers started with Ewing sarcoma and leukemia to extended various types of cancers, including prostate, breast, and hematological cancers, focusing on therapeutic developments (104-106). While it is challenging to develop ETS-directed drugs due to the lack of well-defined binding pockets on those proteins, many small molecule inhibitors have been found from screening efforts for computational-based design or biophysical characterization (107, 108). For example, the 3D structure analysis has been done of ERG or Ets-1; the crystallographic structural knowledge enables them to offer a basis for designing small molecules, such as the structure-based drug (81). The intervention strategy is targeting the interaction between ERG and DNA at the TF-DNA complex. The designed small compound binds on the ETS DBD of ERG and intervenes in the interaction between the ERG-DNA. Also, the 3D structures of the DNA-binding pockets are all shaped differently with unique polar and hydrophobic patches and amino acid substitutions and indels. Hence, all of them can enable the development of small-molecule inhibitors specific to each ETS factor. Thus, the available protein structural information that has already been set can be utilized to discover potent and selective anti-ETS chemicals through a structure-based drug design method (109). In the case of DB1255 interacts not with the ERG protein but rather the ETS recognition site on the DNA (GGAA). The small compound is known for preferentially bind in a

sequence-specific manner. The DB1255 targets the DNA binding site of ERG and perturb the protein-DNA interaction so that it enables implications for ERG as a drug target in abnormal ERG associating diseases.

The small molecules could be developed for targeting the ETS autoinhibition. The ETS TFs at the level of binding interaction with DNA binding sites display an autoinhibition mechanism. Since the autoinhibition is based on the structural components and its own inhibitory states for controlling the ETS TFs' mechanism, small molecules which target the structural components of the DNA binding interactions for autoinhibition can be exploited. The concept of autoinhibition has been suggested from biochemical experiments and still are needs to study further and intensively the molecular mechanism in cell physiology. However, because many ETS TFs are pervasive and constitutively expressed in the biological system, implying that the autoinhibition is relieved by the surrounding cofactors and others. If small molecule inhibitors can be developed to utilize autoinhibition mechanisms toward securing the ETS factors into an autoinhibited state, it might be another possible intervention for the ETS TFs-mediated diseases.

Considered findings of the biophysical and biochemical research and pharmacological implications of the ETS family of TFs, PU.1, was selected as an experimental model in this dissertation. PU.1 retains the most common characteristics with the ETS members in the structural homolog, the Ets-1, the ancient protein of the family; yet the functional characteristics of PU.1 are unique among the other ETS orthologs. Therefore, Understanding and characterizing PU. 1's activity enables knowledge of gene regulation by transcription factors at protein/DNA level in structural biology to pharmacological application in drug development.

1.4 Transcription factor PU.1 as an experiment model

Of the ETS TFs, PU.1 has a great representation to study in molecular recognition for developing target-specific drugs such as small molecules. PU.1 has well known for a target DNA sequence preference of its DNA-binding specificity, and high affinity requires for its target gene transactivation (84, 110-118). Also, as a master of gene regulators for hematopoietic stem cells, PU.1 has been well-documented in normal hematopoietic development and self-renewal of the hematopoietic stem cells (HSC) (119, 120). The PU.1 acts in a lineage-restricted and dosage-dependent manner (121). PU.1 interacts with multiple cofactors and other nuclear proteins to regulate an extensive array of gene expression. PU.1 directly controls to differentiate to major myeloid and lymphoid lineage. In addition, depending on the cellular concentration of PU.1, this directs the fate of the hematopoietic stem cells, with the high-level expression of PU.1 required for macrophage maturation and low-level expressions for B-cell and granulocytic cells.

1.4.1 The rationale for transcription factor PU.1 as an experiment model

We focused on transcription factor PU.1 for several reasons. PU.1 is a critical master in regulating developmental processes to decide immune cell fates, especially in B- and T-cell development. PU.1 is essential for life and directly leads to disease occurrence when its function shows abnormalities. It has also been shown to regulate various genes unrelated to the immune system from previously several studies. Several conserved Ets-binding sites are known for key regulation sites in the ETS TFs containing sites of autoinhibition except PU.1. To discover the intrinsic selectivity of PU.1 on their target DNA, the functional analysis of PU.1 transactivation on PU.1-dependent gene has demonstrated a critical ETS domain site that is required for transactivation of the core promoter. We chose to investigate the B-cell-specific enhancer, the $Ig\lambda B_{2-4}$ gene because the essential regulatory elements for B-cell development and Ig gene

expression have been defined *in vitro* (122). We used this B-cell-specific enhancer when mutations in the enhancer prevent the PU.1 and eliminate the enhancer function.

1.4.2 *PU.1 in structural and molecular biology*

Previously the structures of several ETS family members that conform to the wHTH motif and the ETS-domain are involved in multiple protein-protein interactions in addition to its well-defined role in mediating DNA binding. The ETS domain is an evolutionarily conserved DBD that regulates the expression of various cellular genes by binding to a purine-rich GGAA/T core sequence. PU.1, Spi-1 is most phylogenetically far from Ets-1, an ancient member of ETS, and functionally distinct from others; yet the structural domain of the PU.1 displays highly structural similarity of other ETS members.

PU.1 epitomizes a good model of ETS TFs in their family model. The standard configuration of PU.1 contains DBD (~90 residues) and is well-characterized likewise other ETS TFs. Also, PU.1 has flanking regions that are well distinctive disordered regions in eukaryotic factors. Although PU.1 shares the commonality of the structural domains and its similarity with other ETS paralogs, PU.1 differentiate itself from other members for its unique function in a biological system, even its role as a master regulator in the lineage-specific cell commitments cannot be replaced with other factors that have not been discovered. Further, even with the conserved and high homology in DBD structure among other family members, the target DNA selectivity is more stringent among the ETS orthologs. Ultimately, all the distinctive characteristics of PU.1 stand out for their significances in the researches.

Although the ETS family of TFs contains a large evolutionarily conserved gene characterized by sequence homology within the ETS DBD, the ETS members functionally involve various cellular gene activation and control of its expression (123). The ETS TFs are

involved in regulating a variety of biological processes and are present in all of the metazoan genomes analyzed to date. The development of myeloid cells is controlled by PU.1. Remarkably, the unique role in the common myeloid cell development of PU.1 cannot be substituted by the same subfamily of ETS members, Spi-B and Spi-C. Therefore, the non-redundant functional roles of PU.1 are an important target for research in immunology and cell development.

Interestingly, all ETS TFs bind the same core sites, and several ETS family members may be present in a particular tissue at a specific time in development; each ETS TFs play distinctive functions (123, 124). The studying mechanism of functional specificity to ETS proteins at the molecular level is critical because the knowledge grants us to understanding ETS-related cell functional development and disease. Specifically, understanding the DNA-binding activity of ETS DBD of ETS TFS will provide the most initial step of the life-requiring decision-making process. The higher rigidity by PU.1 toward its DNA target recognition is intensively studied from biophysical and biochemical experiments *in vitro*. Thus, PU.1 is more target-selective than Ets-1, and this stringency is accrued from intrinsic differences in DNA recognition between their ETS domains. For such a reason, it is of great value to comprehend TF PU.1 activity for its DNA recognition and selection by the protein-DNA interaction.

Indeed, PU.1 contains partnership and cooperativity for its target gene regulation. The level of PU.1 expression is coordinated with GATA-1. Furthermore, unlike other ETS TFs, PU.1 is identified as a lack of autoinhibition, allosterically decreasing DNA-binding affinity. Consequently, it is essential for characterizing PU.1 activity in target DNA identification and the contributing factor of DNA binding specificities at the TF-DNA.

1.4.3 PU.1 and diseases occurrences

The disruption of PU.1 activity is often associated with various diseases, including Alzheimer's, rheumatism, fibrotic tissue diseases, liver-related diabetes, hematologic cancers, and leukemia. Research has been developed intervention targeting PU.1 for AML. Considered PU.1 as the potential of the therapeutic intervention and research in developmental genetics and immunology, understanding the molecular mechanism of PU.1 has been stressed. PU.1 has received considerable interest because more than 40% of AML patients showed down-regulated PU.1 levels (121, 125-127) (Figure 1.3).

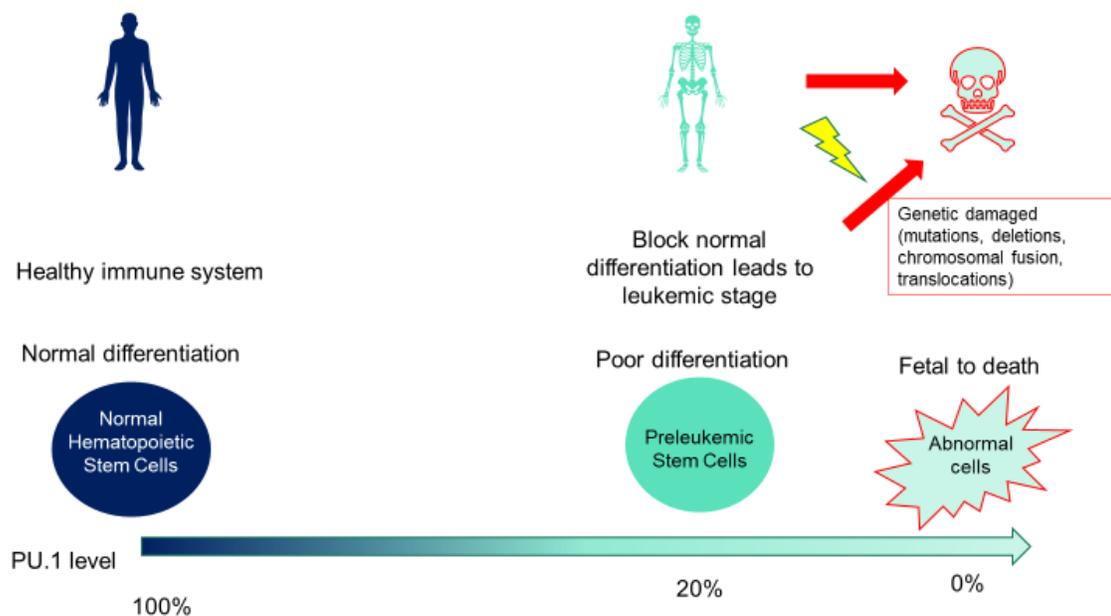


Figure 1.3 Schematic diagram of PU.1 dosage effect on leukemogenesis.

The PU.1 expression level ranged from 50% (haplo-insufficient) to 100% (normal) and assisted hematopoietic stem cells in normal differentiation. However, genetic deletion including PU.1 URE or AML1/ETO, FLT-ITD, PML/RAR α influences damages of PU.1 expression level and below 20% of the PU.1 expression are sufficient to enter the preleukemic phase, in which poorly differentiated cells accumulate myeloid progenitors. When the secondary cooperative mutation occurs, they develop into leukemic stem cells and finally are led to the leukemic phase (121, 128).

Moreover, with a considerable amount of molecular and functional mechanism of PU.1 in cells, it is well established that PU.1 binds to the DNA regulatory region and achieves its target sequence specificity and affinity; however, PU.1 also binds tightly to AT-specific sequences in the flanking regions of DNA minor grooves near the core consensus sequence. The binding of PU.1 to AT-specific sequences introduces a new avenue of research for exploration (129).

Recent research of Mune and our labs proved that the small molecules targeting the PU.1 interacting sites of DNA could modulate the PU.1 transactivation. Because the small DNA binding molecules bind preferably to AT-rich sequence regions, if we can pharmacologically modulate the PU.1 transactivation targeting the AT-rich lying on 5'-flanking region of the interaction between PU.1 and DNA for the binding, then we are able to target low PU.1-expressing cells in AML patients. Because of the potential of the small DNA minor groove binding agents to bind the target DNA sites and specifically inhibit PU.1-DNA interaction, the final outcome would be the breakdown of those pre-leukemic cells due to the further down-regulation in PU.1 low cells and the cells' removal. This can then trigger the body's response to produce normal, healthy cells instead. Suppose the idea of the strategy to inhibit PU.1 transactivation by small compounds and trigger self-renewal of healthy hematopoietic cells can be proved. In that case, the efforts of targeting PU.1 and eliminating preleukemia cells to develop therapeutics have advantages in the development of new types of drugs and agents for biotechnology.

1.4.4 Special features of transcription factor PU.1

Transcription factor PU.1 has its biological and functional importance as a master regulator in transcription affecting immune cell differentiation. Also, considering its inducibility in lineage-specific, and dosage-specific manner, understanding the functional regulatory

mechanisms of PU.1 and PU.1-DNA binding motif structural properties would address shaping gene regulatory network.

PU.1 is a lineage- and dosage- restricted member classified into Class IV of the ETS TFs. PU.1 shares a structurally conserved DBD with other members but binds many cognate sequences in specific tissues, including myeloid and lymphoid lineage cells. PU.1 binds to its target DNA, interacting with the core consensus sequence and flanking sequences adjacent to a central 5'-GGAA-3' consensus. The sites of the flanking regions display significant variations. Even with interaction with other TFs and cofactors are documented, PU.1 has intrinsic heterogeneity for its functional mechanism at the protein-DNA level like other ETS members. Insights of the physical contributing factors for the intrinsic selectivity of the ETS TFs provide to determine the sequence-specific binding protein-DNA interactions and benefits for gaining knowledge for mechanisms of other ETS members.

1.4.4.1 Interfacial hydration is one of the properties of PU.1 -DNA binding mechanism

Water plays an indispensable role in mediating TFs to specifically recognize its cognate DNA site with high affinity (130). At the protein/DNA interfaces, hydration differences in the molecular interactions come from water in the medium relative to those in the surrounding area of large molecules. Therefore, interfacial hydration is considered a crucial determinant to mediate an interaction between TF and DNA. Our previous research exhibited that the ETS protein PU.1 is sensitive to osmotic changes for recognizing its target DNA through site discrimination (113).

The hydration sensitivity of PU.1 manifests PU.1 itself differentiated from other ETS TFs. Interestingly, its structural homolog Ets-1 showed insensitive to water molecule changes. Even though the DBD of Ets-1 and PU.1 bind their optimal sites with similar affinities under

physiological conditions, their ability of site recognition differs in view of hydration. Our data suggest that Ets-1 follows waterless machinery that scatters sites through electrostatic interactions and direct protein-DNA contacts, whereas PU.1 exploits hydration to probe DNA target sequences and form the DNA-binding complex. Further, PU.1 activity is essential to the development and function of macrophages and lymphocytes and is mainly expressed in the myeloid cells, where the environment is osmotically varying. Immune cells like macrophages and activated lymphocytes are susceptible to osmotic pressure from exposure to hyperosmotic conditions, including lymphoid tissues and regional inflammation or active cell proliferation (131-133). As a consequence, hydration-dependent specificity of PU.1 for its DNA recognition may characterize an essential regulatory mechanism distinct to Ets-1 in the biological function in the immune system. The characterization of PU.1 activity at the protein-DNA level is helpful in understanding the developmental immunology in which PU.1 plays a significant role.

1.4.4.2 Self-regulation of PU.1 in negative feedback at protein/DNA level

PU.1 levels in cells are various in the types of cells. With cofactors and other TFs, we have questioned the intrinsic regulatory mechanism of PU.1. Even though PU.1 does not exhibit autoinhibition as a self-regulatory mechanism, unlike other ETS TFs, the tightly regulated level of PU.1, from understanding biological context, should contain the molecular mechanism for self-regulation. Previously we found that negatively cooperative binding to a single DNA site, as PU.1 ETS domain can execute. Our thermodynamic data displayed that PU.1 shapes a reversible, negatively cooperative 2:1 complex with site-specific DNA, suggesting that "self-titration" is a potential negative feedback mechanism (112, 134). Although the 2:1 complex (PU.1: DNA) maintains the functional activity of the 1:1 complex, elimination of PU.1 would diminish PU.1's transactivation. The results displayed the unbound of PU.1 to its target DNA and bound of PU.1

as the formation of 2:1 complex or 1:1 complex under the presence of the site-specific DNA exists. Also, the dynamics between the two forms for achieving equilibrium and exchanges are controlled by PU.1 dose-dependent manner observed by titration calorimetry, DOSY-NMR, and biochemical characterization. The dimerization of bounded PU.1 to target DNA with excess unbound PU.1 suggests the dimerization of PU.1 could provide the negative feedback mechanism when they cannot find their binding DNA site. It might be one explanation for the functional relevance of the self-association mechanism in biology as considering the dosage-sensitive manner for versatile PU.1's roles in a biological context. Since the formation of the 2:1 complex has not been documented yet, the autoinhibitory mechanism of PU.1 could be of high value of further research as a regulatory mechanism at the protein-DNA protein level.

1.4.4.3 Significance of understanding functional mechanism of transcription factor PU.1

Although PU.1 is well known for its biological importance and therapeutic application, the molecular mechanism of PU.1/DNA interaction is needed for more studies. To understand the comprehensive mechanism of PU.1 activity, the fundamental knowledge, especially transcriptional control by PU.1, is the most crucial step. A crucial current consideration initially focuses on how transcription factor PU.1 interacts with its specific DNA site. For the pharmacological application as therapeutics, modulation of transcription factor activity requires disruption of DNA-protein or protein-protein interactions. Because this type of class of transcription factor proteins lacks binding pockets for drug targets (i.e., small molecule design), TFs activity depends on association with other proteins. These interactions occur over large surfaces that generally contact at multiple points and lack hydrophobic fold.

For understanding the interaction, it is essential to characterize the molecular mechanism of the TFs in structure and function. We used PU.1 as a model system to understand the DNA

recognition and determinants of the DNA binding by the PU.1 and provide the basis of knowledge at TF-DNA interaction. Because of the significance of the knowledge of the functional activity of PU.1 interacting with the target DNA site in the biological context, we developed a functional reporter assay to test the biophysical and biochemical properties of PU.1 in live cells. This dissertation attempt to develop a versatile, functional assay for characterizing TF activity at DNA binding and supported biophysical and biochemical characteristics (135).

1.5 Biochemical and biophysical techniques as a tool for TF-DNA interaction studies

Over the last decades' research techniques provides insights into understanding how transcription factors regulate complicated gene expression. Since one of the critical control steps for specific gene expression by TFs is binding their target DNA, many research techniques have been developed to measure transcriptional activities between the protein-DNA complexes and interactions.

1.5.1 Approaches to studying transcription factor activation

Many research techniques have been developed to study the TF-DNA complex interactions. Historically, TFs studies have been conducted using EMSAs/gel shift assays, western blots, and reporter-based assays with plasmid transfections, which are all time-consuming, do not accept the high-throughput analysis, and only deliver only semi-quantitative results. (Table 1.2)

Table 1.2 Summary of the most commonly used techniques for TF-DNA binding characterization

Assay	Advantage(s)	Disadvantage(s)
EMSA	Simple to perform	The results based on the complex analyses, no direct information on DNA binding sites or specific proteins bounded
Footprinting assay	Single base-pair information was provided.	Unclear footprint data hampering for analysis of results coming from incomplete bindings
SPR	Real-time coordination and dissociation for the monitoring.	Cannot link to High throughput (HTP) analysis
SELEX	In vitro, optimal binding selection can be performed.	Selection is limited to only the best complex
Yeast one-hybrid system	It can be conducted in vivo	An experiment is used in the complex system
DNA microarrays	The results can be semi-quantitative, DNA binding sites can be detected. Can be linked to HTS	Dedicated and time-consuming analysis for the individual data
Atomic force microscopy, X-ray crystallography, nuclear magnetic resonance	Structural information can be provided	Neither interaction of the pair or genomic locations are provided

1.5.2 The electrophoretic mobility shift (EMSA) assay

The EMSA technique is the method used to detect protein-DNA interactions (136). The assay is used to identify protein complexes with nucleic acids to characterize the interacting systems. The principle of EMSA for detection is that protein-DNA complexes travel slower than free linear DNA fragments in gel electrophoresis. Mobility-shift assays are often utilized for qualitative purposes to assess binding stoichiometry, affinities, and kinetics under testing conditions. In the assay process, protein and nucleic acid solutions are mixed, and the mixtures run through the electrophoresis under native polyacrylamide (or agarose) gel. The complex is defined as the analysis of radioactive labeled nucleic acid (usually ^{32}P) or the fluorescent label probe. Usually, the protein-DNA complexes slowly move than the corresponding free nucleic

acid. One important requirement for gel shift assays is that the gel-running conditions in which protein-DNA interactions should not be disturbed.

The mobility shift assay has many strengths. The EMSA is simple to carry out, yet it is robust enough to accommodate a wide range of binding conditions (136). The temperature is ranged up to 60°C, the pH range is between 4 and 9.5, and salt concentrations can vary from 1mM to 300 nM. The assay is high sensitivity that the concentrations of the nucleic acids and proteins can be below 0.1nM. As it is an electrophoresis-based assay, the volume of the sample can be less than 20 μ L. It can also be used for studying RNA: protein interactions.

Although there are benefits, there are also limitations. One of the possible limitations of samples is not at chemical equilibrium during the electrophoresis step and leads to prevent detection of complexes for slow-moving due to their density. In addition, the assay cannot detect specific nucleotide sequences to which the protein binds. The EMSA shift does not provide a straightforward measure of proteins' molecular weights or identities present in the complex. To overcome the issue, the assay is linked to mass spectrometry or immunoblotting. Another limitation is that the EMSA cannot provide information on the nucleic acid sequences occupied by protein. Thus, footprinting assays should be independently performed or following with the EMSA. Traditionally EMSA necessitates the use of radioactive; even new probes (biotins or fluorescent probes) are developed. Overall, EMSA is a slow and low-throughput, and non-quantitative assay.

1.5.3 Footprinting assay

Footprinting assays exploit the observation that a protein bound to a specific nucleic acid sequence will interfere with the chemical or enzymatic modification of that sequence (137, 138). In the traditional method, a radioisotope probe is used to label one stand of the nucleic acid

target. After cleaving the nucleic acid at modification sites, the fragments run through a polyacrylamide gel. The resulting ladder of bands is analyzed for seeking gaps in the array. This represents the protein/DNA binding complex for protection sites in the test mixture. The assay provides identification of nucleic acid sequences where the proteins bind.

Footprinting assays require real-time optimization to produce data for the binding reaction of the proteins and the nucleic acid modification. As a result, it is less simple to perform than EMSA. Also, the binding detection from the wide distributed radioactive probes makes the assay less sensitive than EMSA. Additionally, incomplete binding of DNA and proteins lead the unclear footprinting results. Finally, sample proteins can randomly bind to the nucleic acids to result in unspecific data outcomes.

1.5.4 Chromatin Immunoprecipitation (*Chip*) assay

One of the widely used techniques for examining the transcriptional regulatory mechanisms is chromatin immunoprecipitation (ChIP) assay through DNA-protein interactions (139-144). The assay allows the analysis of the association of regulatory molecules to specific CREs and histone modifications. Unlike EMSA, for chemical equilibrium during gel electrophoresis, the assay does not require the chemical equilibrium and provides a tool to determine the protein-DNA complex is happening using specific antibodies (Ab) as *in vivo* application. As a result, the assay delivers information into the chromatin changes occurring in response to extra stimuli or cell differentiation.

ChIP is a method to identify protein-DNA binding events in their natural and intracellular context (145). It is based on detecting a specific protein located in a chromatin fraction that has been selectively augmented. The protein's relative abundance at one or more locations in the genome can be defined using qPCR. It can be additionally applied to identify the specific binding

sequences in the genome (ChIP-seq). Major strengths of this assay work *in vivo* and show interactions with other proteins for studying TF interaction changes over time. But, it requires a specific ChIP-grade antibody to the transcription factor of interest and time- and cost-intensive optimizations for the requirement. The assay is a multistep process, and each step needs to be standardized for obtaining optimum results. Whole cells are treated with a chemical reagent to form a protein-DNA complex; the complexes in the cells are then breaking down by enzymatic reaction or mechanic tools. The sample is run through immunoprecipitation, and then by PCR amplification with primers, target DNA is detected.

The limitation of ChIP is its intrinsic variability due to adaptabilities in the crosslinking. Also, the analysis is based on the standard experiment control; even the variabilities cannot be interpreted as functional data if strict control is set up. Other limitations are the availability of a particular antibody for requiring a high level for the data collection should hamper the assay. Lastly, the protein expression level is challenging if the level is low and under the limit of detection.

1.5.5 Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) assay

The ChIP-seq is a crucial technique in epigenomic research (146-148). The method is used for determining the locations and biochemical modifications of bound proteins genome-wide. The ChIP-seq method uses an antibody for a specific DNA-binding protein or a histone modification to identify loci within a genome (149). The specific ab-directed ChIP assay is a beneficial technique in studying DNA-protein interactions because it enables the chromatin structure surrounding a particular DNA sequence to be analyzed.

Histone modifications are used in the ChIP-seq analysis field to dissect the characteristics and the biological functions of epigenetic signature, especially in elucidating the interaction of

histone modifications with other chromatin regulators at the specific loci. Moreover, the 3D structure of chromatin that links chromatin structure to biological functions can be determined using chromosome conformation capture technology. Advances in next-generation sequencing (NGS) technology and computational analysis enable us to systematically understand how the epigenomic landscape impacts cell identity, development, lineage specification, cancer, and other diseases (148). In addition to genome-wide identification of specific epigenome marks (e.g., enhancers) in a particular cell line, core histone mark enrichment profiles are used to segment and mark whole-genome regions into individual chromatin states to provide more detailed epigenetic features (150). High-quality ChIP-seq databases of chromatin state identify and classified by cell lines and tissues (151).

In ChIP-seq analysis, two significant steps are conducted: sample preparation and sequencing and then computational analysis. From the sample preparation and sequencing step, fragmented DNAs from ChIP-seq samples are sequenced as reads. After multiple scanning steps for paired-end reads can obtain the fragment size distribution. Then, computational analysis is following for mapping. The mapfiles go through a normalization process and represent reading distribution visualization. The normalization for comparative analysis is commonly used, which scales the sample read number to be the same. The underlying assumption is that the difference in mapped reads among samples is sufficiently smaller than the total read number. The computational analysis step is to check the peak callings for motif analysis, which investigates the sequence specificity inherent in called peaks or specific epigenome regions and estimates the potential transcription factor binding sites within identified regions. Or, from the read distribution, chromatin state annotation and chromatin state clustering are performed. ChIP-seq

data quality assessment is critical to judge whether sequencing data are of high quality and suitable for further analyses; therefore, various quantitative QC measures have been developed.

Genome-wide analysis of histone modifications enables a systematic study of how to epigenomic landscape contributes to cell identity, development, lineage specification, and disease (152). Because the machine-learning approaches are used for the advanced ChIP-seq, the methods have limitations for collecting many samples from each cell line to develop training data. As practical limitations lie in the ChIP-seq assay, a particular antibody should be developed before the Chip-Seq assay. Moreover, the limit of a single factor or modification per the experiment is required. Furthermore, a functional understanding of a dynamic cell should be provided to understand the whole biological process.

1.5.6 *Surface plasmon resonance (SPR) assay*

The SPR analysis provides accurate and sensitive kinetics and affinity measurement (153-155). SPR is a label-free technique with no requirement of additional reagents, including abs or radioisotopes. Even the assay is simple for preparation. Also, the small amount of samples is sufficient for the assay. The primary benefits associated with this technique are that it simply responds to changes in refractive index induced by the binding events of molecular reactions. The ligand-of-interest for the study is immobilized on the surface of the biosensor chip. Then, the analyte is injected to flow through the surface of the chip, and detecting the reflective index of the changes is recorded. The SPR angle change resulting from forming a complex of the ligand and the analyte can be interpreted as the binding affinity. However, the SPR assay can perform limited on the *in vitro* experiment so that cannot perform in live cells.

1.5.7 Systematic evolution of ligands by exponential enrichment (SELEX) assay

SELEX is a universal method and widely used for isolating high-affinity single-stranded DNAs or RNAs to bind target protein and selected from a vast library consisting of random oligonucleotide sequences (156-160). The SELEX-derived DNAs and RNAs, aptamers, can be selected against a broad range of targets (e.g., proteins, cells, viruses, microorganisms, toxins, and chemical compounds). Aptamers retain high affinity and specificity to their target molecules, similar to antigen-antibody interaction; therefore, they are called "chemical antibodies." Aptamers offer advantages over antibodies, including small size, higher pH and thermal stability, lower immunogenicity and toxicity, better tissue penetration, lower synthesis costs, and more straightforward conjugation or modifications (161).

The conventional SELEX runs many cycles to obtaining oligonucleotides (aptamers) to bind the target molecules. Target molecules are incubated with a random library of the different oligonucleotides. After washing unbounded molecules, the eluted oligonucleotides are amplified by PCR or RT-PCR to amplify the library. The aptamers are then applied for the next round for selection.

Generally, the isolated aptamers through SELEX possess high affinity and low dissociation constant like antibodies. Consequently, aptamers are seen as chemical abs and become popular for the attractive option instead of abs, resulting in immunogenic reactions. Many therapeutic aptamers have been developed, and many alterations from the types of SELEX method have been made in this process to increase the efficacy and reduce a time for generating aptamer; therefore, aptamers are applicable for many other purposes, including diagnosis of disease and viral infection, protein detection in Western blot, SPR assays, microarrays, and biosensors.

1.5.8 Fluorescence resonance energy transfer techniques (FRET)

FRET is a method to detect the fluorescence energy transfer from an excited donor fluorophore to a ground-state acceptor (162, 163). A fluorescence donor molecule is excited by a light source, and the energy transfer to the closed donor molecule, often the protein, produces the emission signal for detection. This technique can detect the protein's interaction in the cells occurring close distances. FRET provides structural and kinetic information of protein-DNA interactions by preparing dye-labeled nucleic acids and proteins and increased optical sensitivity. The principle of FRET depends on the site-specific labeling with a donor and an acceptor dye, with FRET dyes in each interacting partner or both in the same biomolecule. Direct optical excitation of the donor dye results in fast energy transfer to the FRET acceptor, emitting fluorescence at a longer wavelength.

Intramolecular FRET assays, where both dyes are located on the same molecule, are extensively used to monitor protein-induced conformational changes in the DNA substrate and determine the overall structure and assembly dynamics of various nucleoprotein complexes. FRET method depends on its continuous character so that the cleavage reaction can be monitored from the initial steps in real-time with no need for extensive sample handling. FRET-FLIM is used in situ imaging for protein-DNA interactions in the cell nucleus. This approach lets *in situ* interaction visualization between a GFP-fusion protein and DNA in the cell nucleus, using FRET. To develop a FRET-based method to visualize DNA-protein interaction in situ, a DNA binding fluorescent dye suitable as a FRET acceptor if GFP is the donor must be used.

1.5.9 Nuclear magnetic resonance (NMR)

NMR is used to examine the interactions of DNA with proteins (164, 165). NMR provides dynamic and structural information on the changes in conformation and molecular

flexibility and enables the formulation of mechanistic models of DNA-protein interactions. Some sample preparation steps need to be followed. For the NMR assay, the sample needs to be labeled such that the protein is ^{15}N or ^{13}C labeled while the DNA is unlabeled or vice versa. Sample precipitation needs to be taken care of as there is a strong electrostatic interaction involved within the complex.

The sample can be analyzed by chemical shift mapping where Heteronuclear Single Quantum Coherence (HSQC) spectra of a labeled molecule is analyzed separately for a bound and free state. Chemical shifts are sensitive to alterations in the chemical environment. Thus causing a shift in the spectra as compared to unbound molecules is interpreted as the protein-DNA complex. Another technique that is employed in NMR is the solvent accessibility test which helps in the quantitative analysis of the amide proton exchange rates of the free and the bound protein. However, there are certain intermolecular restraints to the NMR spectroscopy, such as residual dipolar couplings make the assay hamper the precision and accuracy of techniques.

1.5.10 The widely used techniques for detecting protein expression: functional TF transactivation

The importance of DNA-binding proteins has motivated the continued development of experimental and analytic methods to identify better and characterize these interactions. Protein detection methods such as western blot (WB), immunoassay linked with radiolabel probes or enzymes, and reporter genes have been intensively used for targeting detection of sequence-specific binding proteins interacting with DNA complex (Table 1.3).

Table 1.3 Common methods for detection of protein expression

Technique	Description
Western blot	Allows for the detection of a protein in a sample, with measurements of relative concentration between samples
Enzyme-linked immunosorbent assay	Allows for measurement of protein concentration in a sample. More sensitive than a western blot. Quantifying specific amount of protein, rather than determine the relative amount of protein within cells at a very high resolution
Radioimmunoassay	Precise measuring protein concentration in samples that contain dilute proteins
Immunohistochemistry	Characterizing the spatial distribution of the expression of a protein in cells and tissues. Not very quantitative
Immunoelectron microscopy	Characterizes the spatial distribution of the expression of a protein within cells at very high resolution
Reporter proteins	characterize the spatial distribution of the expression of a protein within the need to use immunohistochemistry. Can be performed in live cells or tissues

1.5.10.1 Western blot (WB)

The WB has various applications for examining molecular regulatory events underpinning protein turnover and chronic physiological adaptations (166-168). For instance, the WB can be used to measure protein abundance, kinase activity, cellular localization, protein-protein interactions, or monitoring of post-translational modifications. WB is often used in a study to separate and identify TF proteins or signal molecules. A proteins' mixture is separated based on molecular weight and type using gel electrophoresis in this method. Later, the gel is transferred to a membrane producing a band for each protein. The membrane is then incubated with labeled antibodies specific to the protein of interest. The unbound antibody is washed off,

and only the bound antibody to the protein of interest remains in the membrane. The only bound antibodies are then detected by developing the film. As antibodies only bind to the proteins of interest, only one band should be detected. The band's thickness parallels the amount of protein existing; thus, doing a standard can specify the amount of protein present.

1.5.10.2 Enzyme-linked immunosorbent assay (ELISA) assay

DNA-binding assays have utility in measuring the ability of TFs to interact with DNA (9, 169). ELISA is a plate-based assay technique devised for identifying peptides, proteins, antibodies, and hormones. In ELISA, an antigen is immobilized to a solid surface in the microplate at first and then treated an ab that is coupled to an enzyme for antigen (protein)-ab complex. The ELISA detection is carried out by evaluating the conjugated enzyme activity through incubation with a substrate. This provides a quantifiable product for detection. The ELISA often is available as a commercial kit to quantify activated transcription factors' activity that facilitates the study of transcription-factor activation in mammalian cells.

The first ELISA involved chromogenic reporter molecules and substrates in generating observable color change that monitors the presence of antigen. Then, advancement in the ELISA technique leads to fluorogenic, quantitative PCR, and electro-chemiluminescent reporters to create signals. This assay determines the binding of TFs to DNA bound to a microplate. It can be utilized to determine the effect of mutations or pharmaceutical compounds in the activation of the signal.

ELISA has various advantages, including high throughput and quick to get results. Also, this essay can be quantitative if a standard curve is created with purified active protein. However, it requires a specific antibody to the TFs of interest. The techniques are susceptible and convenient for operation. Also, the experiment outcomes obtain quickly and ease of analysis by

which colorimetric readout enables simple quantitative analysis with spectrophotometry at 450 nm. Furthermore, the assay is flexible format 96-strip well format so that it enables both high low-throughput assays (9, 170, 171).

ELISA is performed in 96-well plates coated to bind protein very strongly. Depending on the ELISA type, testing requirements for the kinds of abs or enzymes vary. The four typical steps are performed for the ELISA immunoassay. First, antigens or antibodies are coated. Then, BSA (bovine serum albumin) solution is used for blocking. After washing the unbound reactants, samples are detected, and data of the enzymatic reaction (color changes) is readout by plate reader.

The experiment outcomes are detected by the addition of a substrate that can generate a color. Mostly the substrates used for ELISA are horseradish peroxidase (HRP) or alkaline phosphatase (ALP) results in a blue color change. In ELISA protocol, a serial dilution of concentrations is placed in the wells of the plate, and the results are calculated. A standard curve is plotted from the serial dilutions data with a concentration on the x-axis and absorbance on the y-axis. The data from ELISA tests can be quantitative, qualitative, or semiquantitative. The quantitative concentration results are plotted and compared to a standard curve. The qualitative results confirm or deny the existence of a particular antigen/antibody in a sample. The semiquantitative results compared the intensity of the signals, comparing relative antigen levels in a sample. Once color changes are detected from the assay, the results are graphed either on paper or software. Usually, the plot evaluates optical density to log concentration, which gives a sigmoidal curve. Known concentrations give the standard curve, and measurement of unknowns can then happen when sample values are compared to the linear fraction of the graphed standard curve.

1.5.10.3 Reporter gene assay

Reporter gene assay is a helpful technique for biomedical and pharmaceutical researchers to monitor cellular events related to gene expression, regulation, and signal transduction. A reporter gene is an exogenous coding region joined to a promoter sequence or element in an expression vector introduced into cells to provide the tools for determining the promoter activity (172). Most physiological phenomena, including cell communication, cellular development, growth regulation, proliferation, proliferation, can be controlled at the TF-DNA level. The differential gene expression is tightly regulated in response to intrinsic developmental programs and extrinsic signals. In order to gain insightful knowledge about the relationship between the activation/inhibition of different pathways and their effects on gene expression, specific response elements are fused to genes encoding reporter proteins (Figure 1.4).

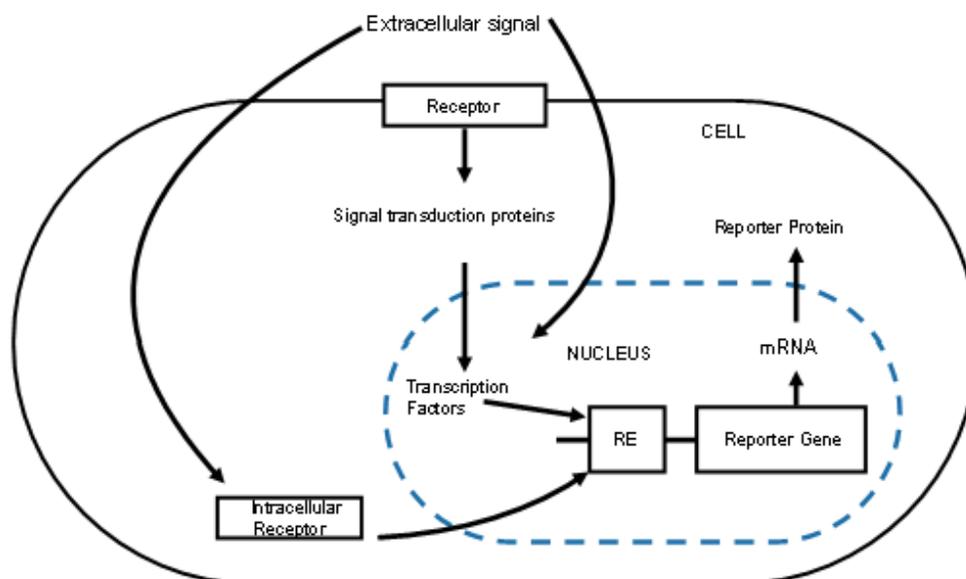


Figure 1.4 Schematic representation of a cell-based reporter gene assay.

Extracellular stimuli can trigger cellular signal cascades, or TFs activate DNA binding sequences as response elements that modulate the reporter gene transactivation, and the expression of reporter proteins are detected. The reporter protein is readout signals as an indirect measurement for the transcriptional activity

Once reporter proteins are expressed in cells, the reporter is assayed by either directly measuring the reporter protein itself or assessing its activity, correlating the strength of the fused DNA response element with the amount of reporter produced. Based on the alternations in reporter gene activities mediated by attaching response elements to the reporter genes, an assay that is sensitive, reliable, and convenient can efficiently report the activation of specific signaling cascades and their effects on gene expression and regulations inside cells living subjects. Generally, reporter genes should have the advantages of reduced background activity in cells and amplify the signal from the cell surface to create a rapid, highly sensitive, reproducible, and easily detectable response. Currently, reporter genes are used in both *in vitro* and *in vivo* applications. In particular, reporter systems are utilized to study the promoter and enhancer sequences or trans-acting mediators for transcription, mRNA processing, and translation. The technique can also be utilized for localization, transfection efficiencies, protein-protein interactions, subcellular protein localization, and recombination events, and screen genome-wide libraries for essential genetic regulatory elements. All these applications signify the enormous contribution to research in biology and biotechnology. In the next section, reporter gene assay will be described in-depth to understand the principles of the reporter gene technology, its applications, advantages, and disadvantages. In addition, rationales of the reporter gene platform in this dissertation will present to understand the functional regulation of transcription factor activity at the protein/DNA level.

1.6 Reporter gene platform to characterize TF-DNA interaction

Most physiological processes, including cell communication, development, proliferation, and apoptosis, can be controlled at the TF-DNA level. The differential gene expression is tightly regulated in response to environmental signals. In order to gain insights into the relationship between the activation or inhibition of different pathways and their effects on gene expression, particular DNA responsive elements are fused to genes of reporter proteins (173) (Figure 1.5).

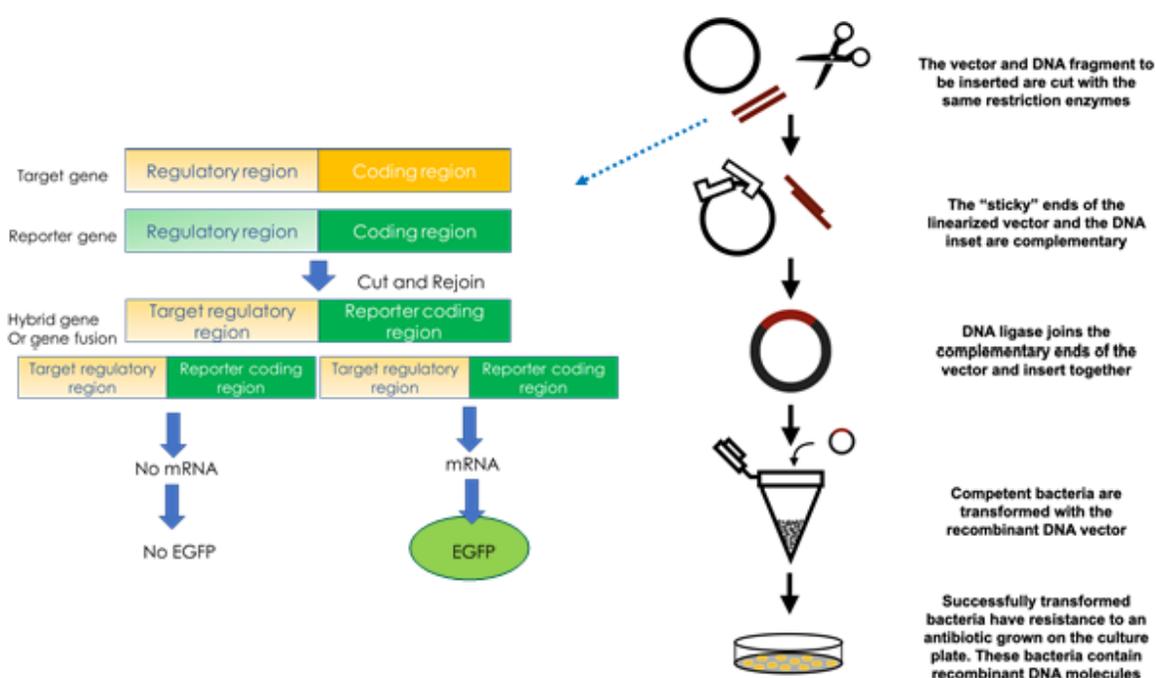


Figure 1.5 Steps in constructing and generating a recombinant DNA plasmid

To investigate the gene of interest (e.g., regulatory component of gene expression including transcription factor or cis-regulatory elements), the target gene of regulatory regions is cut by the proper restriction enzymes first. Then, the cutting gene segments are fused to the gene segment to the reporter coding region, and the hybrid gene or gene fusion product is formed. The hybrid gene is inserted into the appropriate expression vector and transformed into bacterial competent cells to yield the recombinant DNA molecules. After extracting DNA from the amplified cells, the recombinant DNA molecules are purified and ready for the next step, chemically transfected into the host cells. The figure is inspired by Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. *Isolating, Cloning, and Sequencing DNA*.

Reporter genes can indicate events throughout the entire genetic process because their measurable parameter is an or directly measuring the fluorescence of the marker protein expression under the control of the interest promoter or enhancers. Therefore, reporter gene assay provides valuable information on monitoring transcriptional activity in cells. Further studies on gene expression have broadened to include the entire process of physiology regulation and phenotype expression. As a result, our use of genetic reporters has expanded from the analysis of *cis*-acting elements to the study of downstream events such as RNA processing and protein expression and upstream events, such as the biochemical mechanisms preceding DNA transcription.

Reporter assays can also provide a real-time *in vivo* readout of translational activity for a promoter of interest because the reporter genes are fusions of a target promoter DNA sequence and a reporter gene DNA sequence. The DNA promoter sequence is personalized and fused to the reporter gene for a protein with detectable properties. These reporter genes produce fluorescence proteins only when the promoter of interest is activated. The fluorescent protein produces fluorescent light that can be detected by flow cytometry or microplate reader, spectroscopic instrumentation. The signal from the reporter gene is treated as an indirect determinant for the translation of endogenous proteins driven from the same promoter. Experimental conditions are established such that events other than transcriptional regulation is presumed to be constant. Changes in reporter expression are thus coupled to differences in transcriptional activity. Even so, alternative experimental strategies can reveal other stages within the broader genetic event.

Reporter assay has developed because of the advanced cloning technology to identify regulatory elements of cloned genes with the functional assay. Since the gene cloning technique

advanced and developed sophisticated cloning procedures to isolate selected eukaryotic genes, reservoirs of sequenced genomic segments have enormously expanded. Comparative sequence from the genomic segments revealed that many genes share common consensus sequences upstream of the transcription start site. These include the TATAA elements, an AT-rich sequence that appears to set the 5' initiating nucleotide for transcription 25-30 bases downstream, a GC-rich oligomer associated with certain viral and cellular genes, and another element with a consensus sequence CCAAT approximately 80 bases upstream of many transcriptions start sites. After characterization of enhancers in both viral and cellular genomes, regulatory elements to be assigned for their role. As a result, a functional assay system starts developing in which cloned sequences can be reintroduced into eukaryotic cells in culture and tested for transcriptional activity. The popularity of cell culture systems for the analysis of cloned gene expression derives principally from the relative simplicity of the experimental procedures involved and the reproducibility of the assays compared to techniques of transgenic mice. With an appropriate cell line in hand, it is possible to demonstrate relatively promptly with an expression assay that a particular set of nucleotides functions as a transcriptional regulatory element. The reporter gene assays were rapidly developed with the advanced development of expression vectors into culture cells (172).

1.6.1 Chloramphenicol acetyltransferase (CAT) assay

The chloramphenicol acetyltransferase (CAT) gene was the first reporter gene to monitor transcriptional activity in mammalian cells (174, 175). CAT is a bacterial enzyme that enables detoxifying chloramphenicol, an inhibitor of prokaryotic protein synthesis, by catalyzing the transfer of acetyl groups from acetyl CoA to the 3-hydroxyl position of chloramphenicol. The CAT consists of three subunits of 25 kDa protein each and is relatively stable in mammalian

cells. CAT was widely used as an indicator gene because of lack of endogenous proteins for interfering with CAT activity in mammalian cells. As the first gene reporter, CAT has been used to monitor the delivery, location, and pattern of transgene expression. Due to the reliance on radioisotopes, the CAT assay could not employ in live cells.

1.6.2 Alkaline phosphatase (AP)

AP has become a representative marker that participates in most biological processes (176, 177). AP consists of two similar monomers that contain five cysteine residues, two zinc atoms, and one magnesium atom for its critical function. AP serves as a functional role of a catalyzer for biological dephosphorylation and becomes active in ideally active in an alkaline environment (178-182) (Figure 1.6).

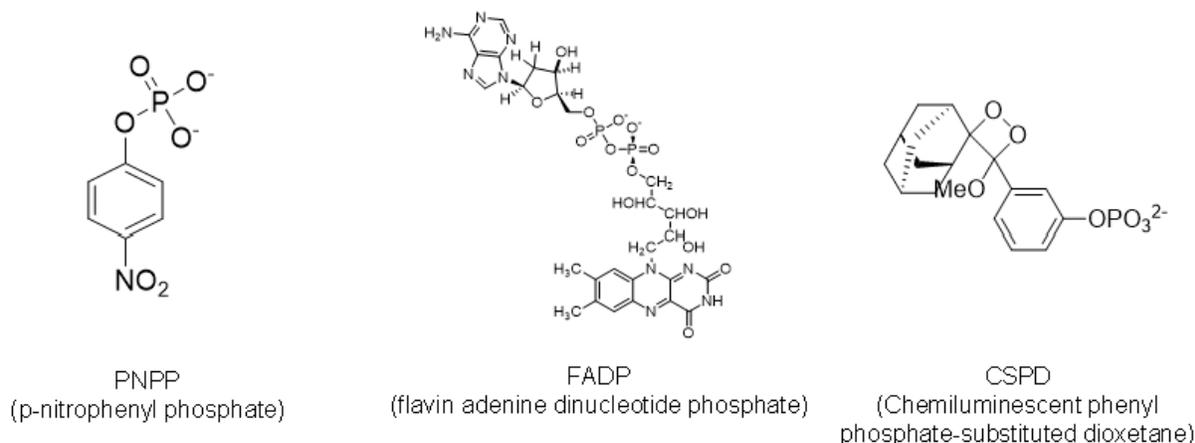


Figure 1.6 Structures of the substrates for alkaline phosphatase (AP)

The substrates are hydrolyzed by AP and generate the absorbance changes and detect the signals by spectrophotometer. PNPP, FADP and CSPD are commonly used as the assay substrates.

Because of the physiologically vital role of AP for living organisms, AP is optimized its activity at pH above 8. AP is a relatively stable protein ubiquitously expressed in bacteria and mammals. The AP hydrolyzed its substrates to produce the absorbance difference and the

signal detected. AP has been used to analyze a culture medium from stably transformed cells, with the variety of easy and sensitive assays letting for a convenient and versatile reporter system. It has also been demonstrated to monitor promoter activity, detect growth factors, and identify inducers/repressors of gene expression. Unluckily, AP is expressed practically in all cell types; therefore, the cells' metabolic states can interfere with monitoring TFs activity in mammalian cells.

1.6.3 Secreted alkaline phosphatase (SEAP) assay

SEAP is a reporter used to study promoter activity or gene expression (176). The SEAP reporter gene encodes a truncated form of the human placental AP gene. Due to no anchoring domain of the enzyme, AP can be secreted from transfected cells to detect reporter gene activity without cell lysis. Using a secreted reporter protein has advantages over traditional reporter assays such as CAT assay because the assay does not require cell lysis. SEAP assay is a type of enzymatic assay that AP dephosphorylates the chemiluminescent alkaline phosphatase substrate (CSPD) into an unstable dioxetane anion and detects the producing light. The assays can be carried out in both single tube and microplate format. SEAP is a stable enzyme, which stays active in the typical cell culture media for 72 hours.

1.6.4 β -galactosidase (β -gal) assay

β -galactosidase is a type of bacterial enzyme, constituting one tetramer with a large subunit size of 1023 amino acids (the monomer is 116 kDa) (183, 184). Each monomer peptide chain in each β -gal tetramer structure can fold into five sequential domains. The five domains contain an extended segment of ~50 amino acid residues at the amino terminus. About 26 amino acids can be replaced with genes from various other proteins with no affecting catalytic activity. In the meantime, other coding regions producing active chimeric β -galactosidase can still be

produced because of the redundant coding region. The expressed β -gal can facilitate the hydrolysis of various β -galactosidase, which has been used as an internal control for normalizing variability in reporter protein activity. Transfection efficiency or cell extracts enable to make variability of the experiment conditions (Figure 1.7).

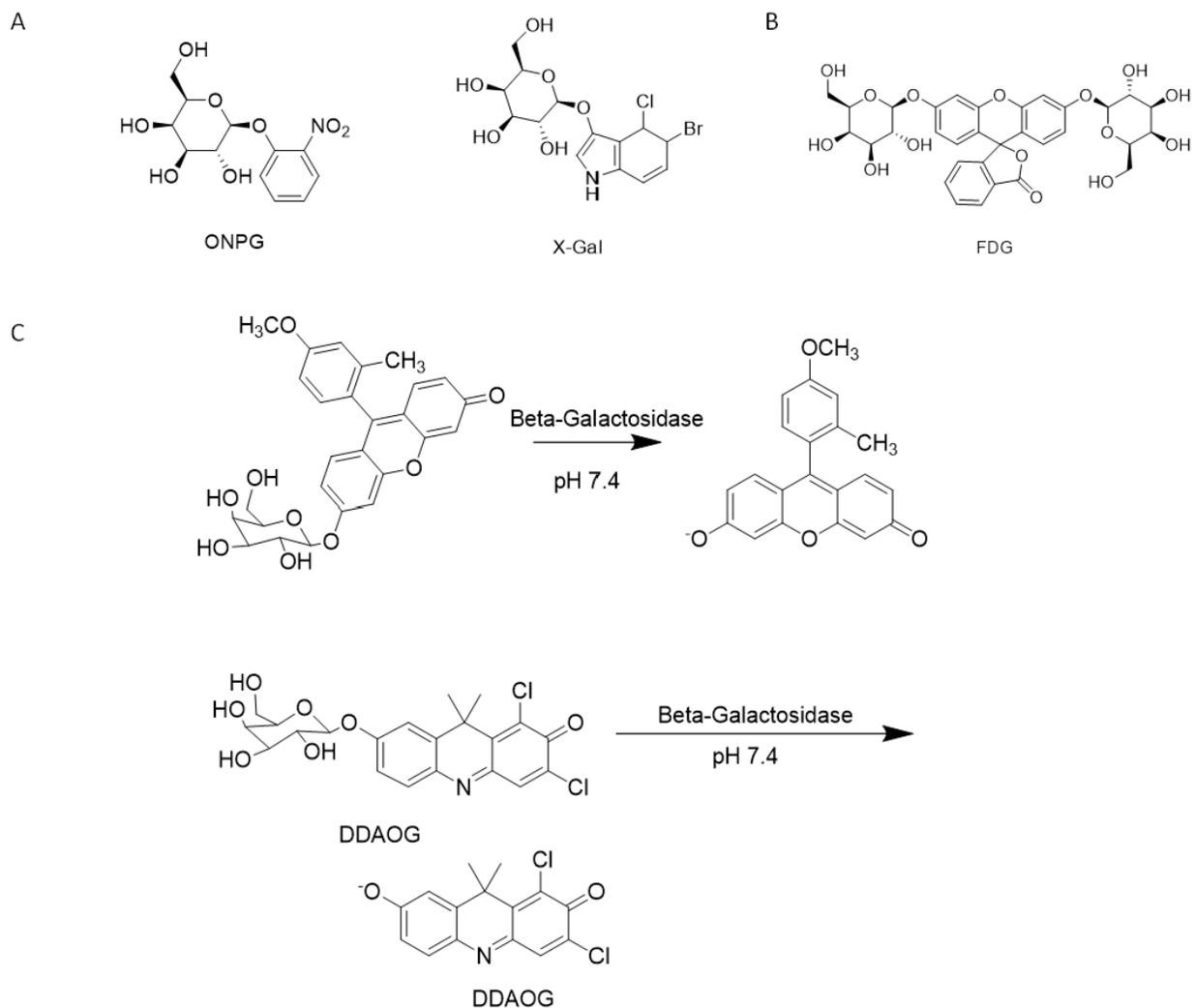


Figure 1.7 Structure and reaction scheme for in vitro and in vivo β -galactosidase assay (A) Structures of o-nitrophenyl β -D-galactopyranoside (ONPG) and 5-bromo-4-chloro-3-indolyl Galactosidase (X-Gal), which are substrates for β -gal expression assay. They have been commonly used for in vitro and ex vivo because of their chromogenic features; however, despite of their simplicity on the colorimetric assay feature, the assay in use of a single cell is insufficient due to weak sensitivity and narrow dynamic range. (B) Fluorescein-di- β -D-galactopyranoside (FDG) has better substrate replacement for assay in the mammalian cells.

FDG exhibits fluorescence after reaction with β -gal. The β -galactosidase-FDG system has commonly used in the reporter assay because one fluorescein exhibits intense fluorescence, particularly for a single cell application. (C) Using substrates of TG- β Gal and DDAOG (9H-(1,3-dichloro-9,9-dimethylacridin-2-one 7-yl) β -galactopyranoside (DDAOG), far-red compound), β -galactosidase activity with TG- β Gal.

A direct comparison can be thru between the cells' number that the cells are transfected and the strength of the promoter by using colorimetric enzymatic assays to identify the transfection efficiency, cell specificity, or preference of infection and evaluation of the promoter expression level. This method is predominantly valuable for recognizing specific cells expressing the promoter-lacZ construct within a primarily heterogeneous cell population. Nevertheless, the X-Gal and ONPG have known for narrow dynamic range and poor sensitivity.

1.6.5 β -lactamase (BLA) assay

The β -lactamase reporter assay is a type of enzymatic assay use of β -lactamase. It utilizes a fluorometric readout on live cells because they do not require cell lysis. An additional benefit of the β -lactamase reporter assay is that the detection is based on FRET substrate, CCF4-AM, which is cell-permeant. The FRET substrates emit two different lights so that the difference of the light emission based on the enzymatic reaction can be detected by a fluorescent light detector such as a flow cytometer. It is beneficial in screening drugs because the detection signal is quantified from numbers of samples (185-190) (Figure 1.8).

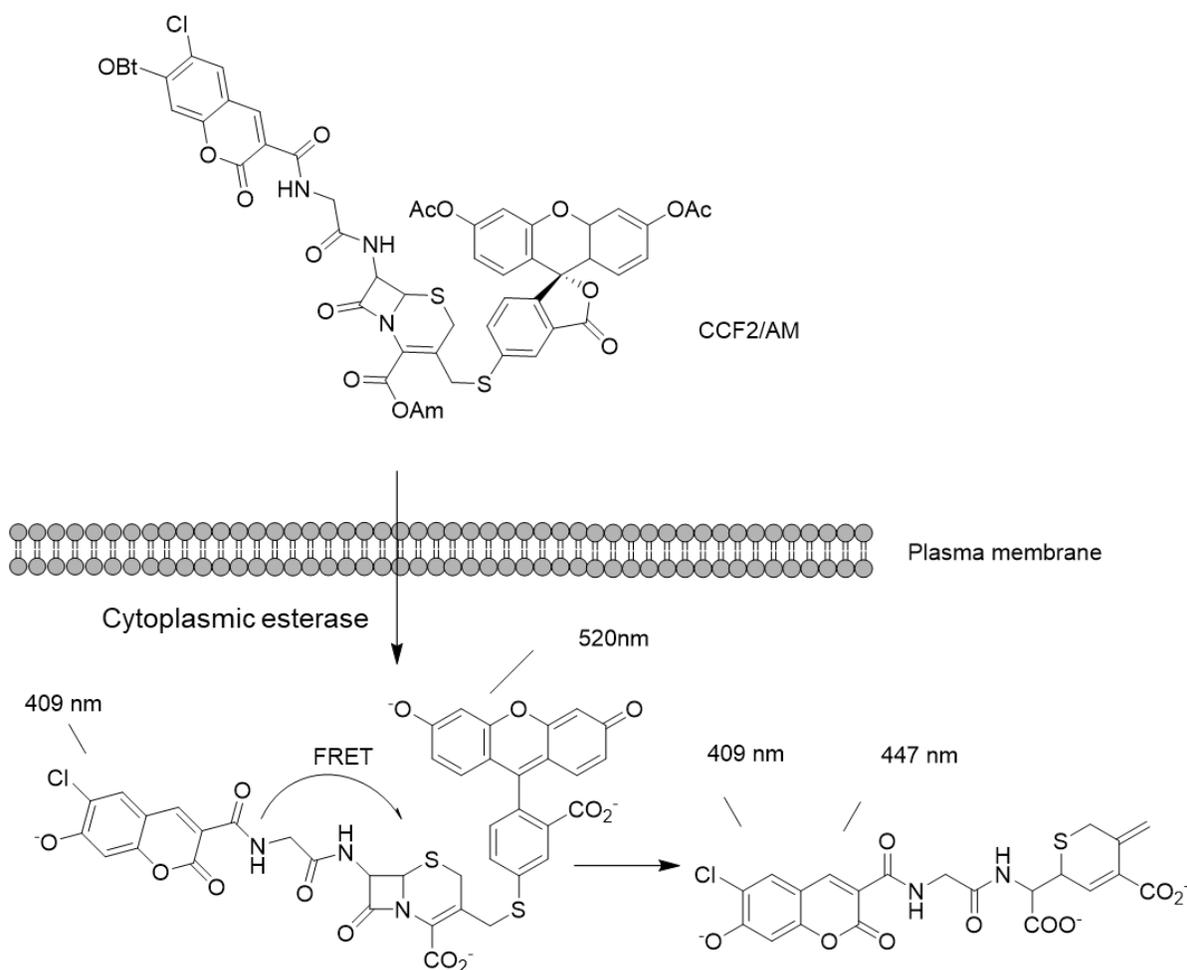


Figure 1.8 Illustration of the β -lactamase reporter system

After the esterified (acetoxymethylated; AM) form of this substrate CCF2/4 enters a cell, endogenous esterase converts it to CCF2/4, thereby catching it inside the cell. Exciting CCF2/4 at 409 nm results in efficient FRET from the coumarin moiety to the fluorescein derivative and creates green fluorescence detectable at 520 nm. After β -lactamase cleaves CCF2/4, the two fluorophores separate, causing loss of FRET; excitation at 409 nm results in blue fluorescence detectable at 447 nm.

The β -lactamase reporter assays employ the enzyme TEM-1 β -lactamases (BLA), which lacks the N-terminal 23 amino acid periplasmic secretory signal sequence. BLA is a 29-kDa enzyme encoded by the ampicillin resistance gene that acts as a monomer or when fused N- or C-terminus to a heterologous protein. It can cleave β -lactam-containing molecules like penicillin and cephalosporins with simple kinetics and high catalytic efficiency. Any ortholog of BLA does

not exist in eukaryotes. Also, overexpression of BLA is not harmful to toxicity in eukaryotic cells. Even though the substrate of BLA and its enzymatic products are nontoxic in cells, the assay is not convenient for monitoring TF activities targeting DNA binding interactions due to the use of enzyme and substrates for the assay.

1.6.6 Luciferase assay (Bioluminescent reporter gene assays)

A luciferase assay is used to determine whether the gene of interest is upregulated or downregulated in the cells from the signals of luciferase activity. The assay can be quantifiable for interpretation under the specific protein existence and the specific gene expression. The detection signal is quantified to provide the functional connection between the protein and the particular gene of interest fused to the luciferase gene. In the case of studying transcriptional factors, the assay cannot uncover whether the protein directly interacts with DNA. The functional connection is limited to measuring the effect of the specific protein for transcriptional activity indirectly.

Luciferase is derived from the firefly and utilized for functional reporter assay (191-193). Luciferase assay involves genetic engineering for the manipulation process. The promoter of the target gene of interest, expecting the DNA binding targets of TFs, is fused with the DNA coding region of the luciferase gene (194). Also, individually target genes for coding the expecting protein, such as transcription factors, are encoded in the expression plasmids. The two constructs are transfected into a cell culture system, such as HEK 293 cells. The expression of luciferase is detected 48 to 72 hours after post-transfection. Depending on the activity of the protein (TFs), the luciferase activity is increased (upregulated gene expression) or decreased (downregulated) after the lysis of the HEK293 cells and the addition of the proper substrates. The catalyzed luciferases' activity is transferred to signal and quantified based on the amount of light. It reflects

on the activity of the protein of interest on the gene of interest. Because of variability in conditions within each well and sample in a cell culture plate, a luciferase assay experiment should repeat the experiments several times for obtaining its reproducible data for each condition at least three times. The assay should have proper negative control, such as conditions in the absence of protein of interest and the absence of target gene for appropriate data interpretation.

Luciferase produces light emission from the enzyme reaction on the proper substrate under the presence of oxygens (Figure 1.9).

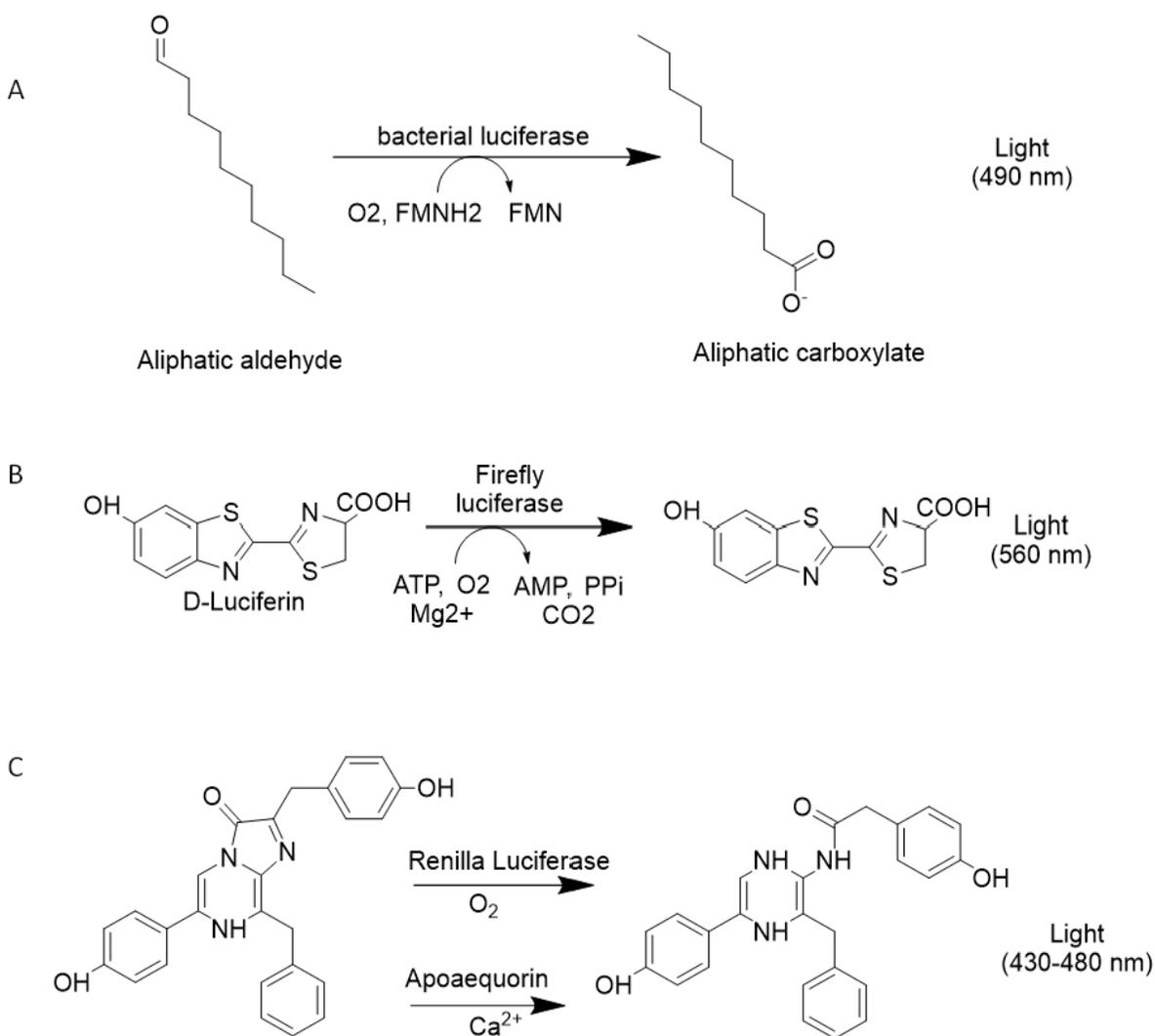


Figure 1.9 Bioluminescence substrates and enzymatic reactions of several common luciferases

(A) The aliphatic aldehyde substrate of luciferase. The luciferase enzyme oxidizes reduced flavin mononucleotide (FMNH₂) and a long-chain aliphatic aldehyde in the presence of oxygen to release visible light; (b) Luciferin is catalyzed by firefly luciferase. The reaction requires oxygen and ATP. During the reaction, the chemical energy is released as light (spectra, 530nm-640 nm); (c) Renilla luciferase (Rluc) is another type of luciferase without the requirement of ATP for its reaction. Rluc catalyzes the oxidation of coelenterazine (substrate).

Luciferases are the heterodimeric enzymes comprising of one 40 kD α -subunit (*luxA*) and one 37 kD β -subunit (*luxB*) and oxidize flavin mononucleotide (FMNH₂) to emit light. The system is limited application. FMNH₂ can be fast oxidized and cannot use the system in a mammalian cell due to the hard to express in the culture cells. Luciferase is an enzymatic protein for the requirement of appropriate substrates, oxygen, or ATP as energy sources, as figure 1.10 stated. The condition for the reaction can be influenced by the cellular metabolic states as well. Moreover, during the detection, the cells should be lysed so that this assay cannot provide real-time monitoring for dynamics of TFs activity. Consequently, luciferase assay is not suitable for characterizing TF-DNA interactions for such reasons.

1.6.7 Quantum dot reporters

Inorganic nanocrystal, Quantum Dots (QDs) have been used as a fluorescent probe as a biological experiment (195-197). Because QDs provide strong brightness and long-lasting photostability for researchers, it is mainly used for mainly in vivo imaging applications. Depending on the size of QDs, many different spectrums can be generated and tuned, it is the popular use of combining with microarray technology for quantitative assay. QDs also have conjugated with small molecules, including chemicals, peptides, small ligands, or antibodies, to reduce the nonspecific molecules' interaction in the cell environment. Overall, QDs are popular

for use in a multiplex fluorescent system for tracking proteins' biological systems in visualization and screening small ligands for therapeutics development.

1.6.8 Summary of the reporter gene assay

The genes that encode CAT, β -galactosidase, luciferase assay have been employed as a merit of quantitative reporters to detect or monitor the gene of interest (promoter) activity. However, the conventional reporter gene expression shares two disadvantages. First, most reporter assays determine based on the total gene activity from entire sample cells so that the gene varies in an individual sample. Also, the measuring time for the conventional assay cannot provide real-time monitoring due to the reliance on the enzymatic reactions, in which the enzymatic reaction is fast. The standard reporter gene techniques are summarized in their characteristics (172, 198) (Table 1.4).

Table 1.4 Characteristics of gene reporter systems: advantages and disadvantages for aspect of cell-based assay development and screening

Reporter gene	Function	Advantage	Disadvantage
Chloramphenicol acetyltransferase (CAT)	Transfers acetyl group from acetyl coA molecule to chloramphenicol, causing its detoxification	Genetic product is stable and detectable at attomolar concentration; suitable for mammalian system; no endogenous activity; automated ELISA	Non suitable for high throughput studies; Narrow linear range; use of radioisotopes; stable
β -Galactosidase (lacZ, bacterial, mammalian cells)	First reported in 1980. In E.Coli, hydrolysis lactose to glucose and galactose	Can act on many substrates; well characterized; stable; simple colorimetric readouts; sensitive bio- or chemiluminescent assays available.	Costly and potentially toxic chemical for assay and lysis of cells; not useful for real time detection systems; endogenous activity (mammalian cells).

<p>β-Lactamase (BLA)</p>	<p>Use of the enzyme TEM-1 β-lactamase (BLA), selectively detects of fluorescence color changes when the fluorogenic BLA substrate CCF2/4-AM are esterified to CCF2/4 and detected by FRET by esterase in the cell and cleaved by BLA.</p>	<p>Enzyme substrate is easy to deliver inside cells; radiometric data available so that eliminate issues from well-to-well variations in cell number and fluorescence signal.</p>	<p>Poor dynamic range; its substrate CCF2/4 has high molecular weight and poor solubility in aqueous solution</p>
<p>Secreted Alkaline Phosphatase (SEAP)</p>	<p>The expressed secreted AP is detected when it activates a dioxetane derivative and decomposes. During the decomposition process, light is emitted and detected.</p>	<p>Secreted protein; inexpensive colorimetric and highly sensitive luminescent assays available</p>	<p>Endogenous activity in some cells; interference with compounds being screened; time-consuming and labor-intensive serial dilution required for the stable cell line.</p>
<p>Luciferase</p>	<p>Proteins that generate luminescence biologically. Can be eukaryotic or bacterial (lux). Firefly luciferase is one of the most common reporter genes.</p>	<p>[Firefly] High sensitivity, tight coupling of Luc protein with luminescence output, protein requires no posttranslational modification; no endogenous activity; broad dynamic range; convenient assays; [Bacterial luciferase], good for measuring/analysing prokaryotic gene transcription.</p>	<p>Firefly luciferase requires addition of costly substrate luciferin to monitor activity; it requires substrate (luciferin) and presence of O₂ and ATP; the substrate for bacterial luciferase is produced endogeneously, but not very active in eukaryotic systems (less sensitive).</p>

1.7 Green fluorescent protein (GFP) reporter assay

GFP is the most widely used and developed reporter applicable in biochemistry and cell biology. After discovering GFP and its mutants, because of the stable fluorescence offers and independence of the enzymatic reactions, even many variants of spectra of fluorescence protein, GFP-based reporter assay become popular for many applications. GFP is employed as a protein fusion reporter in various applications in diverse cells and detects particular biological processes, such as fusion gene expression. The fusion protein produced by the engineered genes maintains its natural function along with fluorescent property acquired from GFP expressing gene.

1.7.1 *Discovery and major milestones*

Shimmonura and his colleges discovered GFP as a companion protein to aequorin, the famous chemiluminescent protein from *Aequorea* jellyfish (199-205). Morin & Hastings discovered GFPs coelenterates *Obelia* from a hydroid and *Renilla* from sea pans. Moris's team purified and crystallized GFP, assessed its absorbance spectrum and fluorescence quantum yield, and demonstrated that aequorin could efficiently transfer its luminescence energy to GFP when the two adsorbed onto cationic support. Prendergast and Mann obtained the first precise estimate for the monomer molecular weight. Shimomura proteolyzed denatured GFP, analyzed the peptide that retained visible absorbance, and correctly projected that the chromophore is a 4-(p-hydroxybenzylidene) imidazolidin-5-one attached to the peptide backbone through the 1- and 2-positions of the ring. Prasher and his colleagues cloned the GFP genes, and Chalfie and his teams and Inouye & Tsuji proved that expression of the genes in other organisms emits fluorescent lights. From the gene cloning works, GFP was able to be used widely for reporter gene assay. GFP has a much higher extinction coefficient, resistance to pH-induced conformational changes and denaturation, and tendency to dimerize. Among these GFPs, the GFP gene from jellyfish

Aequorea victoria has been cloned. The cloned GFP gene has been the most studied green fluorescent protein as a biotechnological marker.

1.7.2 GFP structure

Cubitt suggested the proposed mechanism for the intramolecular biosynthesis of the GFP chromophore (203-205) (Figure 1.10).

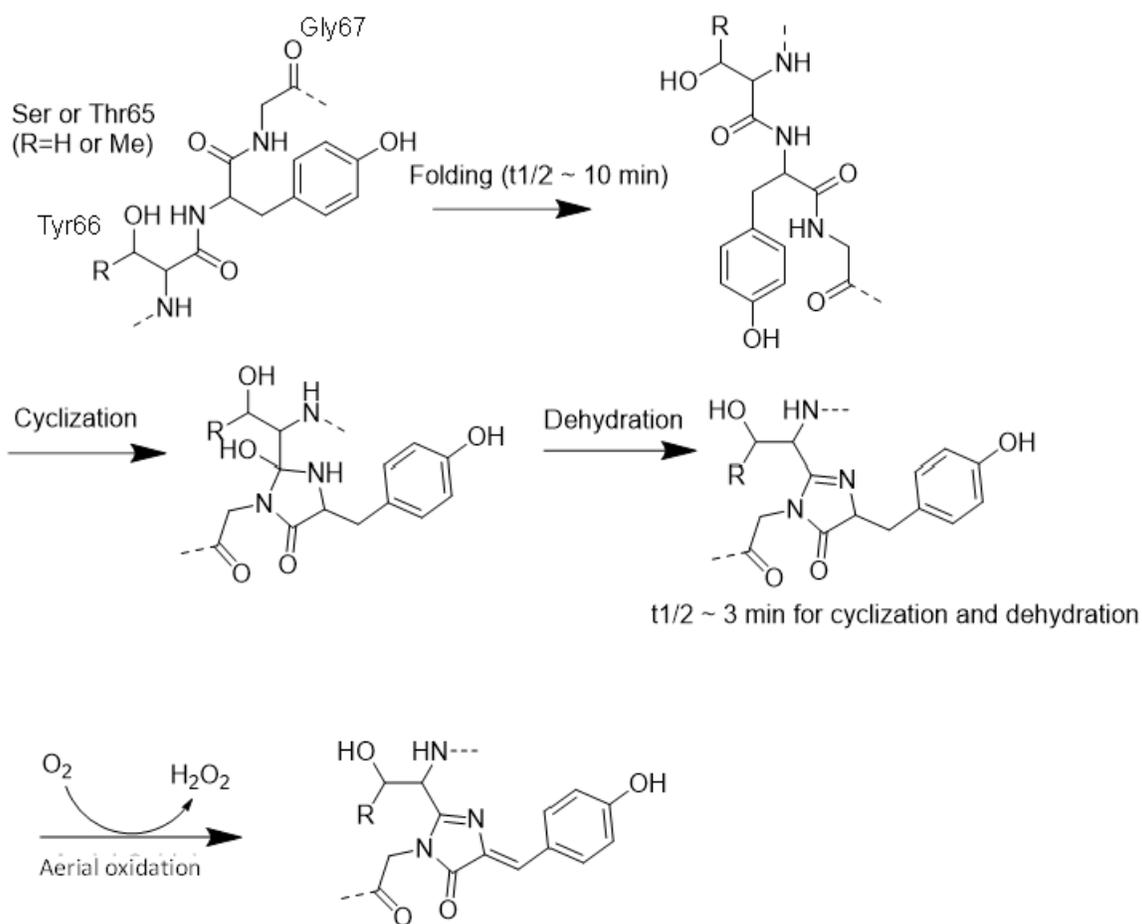


Figure 1.10 Proposed Mechanism for the intramolecular biosynthesis of the GFP chromophore.

First, GFP folds into a nearly native conformation, the imidazolinone is then formed by nucleophilic attack of the amide of Gly67 on the carbonyl of residue Ser65, followed by hydration. Next, molecular oxygen dehydrogenates the α - β bond of residue Tyr66 to put the aromatic group into conjugation with the imidazolinone and therefore the chromophore acquires visible absorbance and fluoresce (206).

The chromophore is a p-hydroxybenzylideneimidazolinone formed from residues 65-66, which are Ser-Tyr-Gly in the native protein. The currently accepted mechanism for chromophore formation is illustrated in the figure. First, GFP folds into a native conformation, and then a nucleophilic attack of the amide of Gly67 on the carbonyl of residue 65 produces imidazoline. After the dehydration process, molecular oxygen dehydrogenates the α - β bond of residue 66 to position its aromatic group into conjugating with the imidazolinone.

GFP is an 11-stranded β -barrel threaded by an α -helix running up the axis of the cylinder. The chromophore is connected to the α -helix and is buried almost perfectly in the center of the cylinder, which has been called a β -can. Almost all the primary sequence is employed to build the β -barrel and axial helix so that there are no prominent places where one could design large deletions and reduce the size of the protein by a significant fraction. Residue deletions and reduce the size of the protein by a significant fraction. GFP forms a dimer. The same wild-type GFP also forms as a monomer, isomorphous to the monomeric crystals.

1.7.3 Applications of GFP reporter in our research

In tagging applications, GFP fluorescence reflects gene expression or subcellular localization caused by targeting domains or host proteins to which GFP is fused. As an indicator, GFP fluorescence can be modulated posttranslationally by its chemical environment and protein-protein interactions. The success of cell-based assays relies on the quality of the reporter gene. While valuable, most commercially available reporter technologies are not versatile enough to cover every aspect of cell-based assay development and screening. For example, although luciferase-based assays offer acceptable sensitivity and a dynamic range, they cannot be used for flow cytometry due to the transience of its enzymatic byproducts and light signals. A preferred reporter for us to study the TF-DNA complex should mimic the properties of the protein of

interest and should produce a stable, readily measurable signal under physiological conditions such as in mammalian cells (172). Also, Except GFP reporter assay, another reporter assay requires the use of an enzymatic process for the detection requires appropriate substrates and lysis of cells. Detection of intracellularly expressed enzymes and their catalyzed process in living cells can be inadequate because of the instability of the substrate and low permeability.

In contrast, the GFP reporter does not require any additional enzyme substrates for the experiment and monitors the GFP expression as a response to target gene activation in live cells. Even GFP is a stable protein under physiological conditions in pH and temperatures. Therefore, for monitoring TFs -DNA interaction in live cells, a GFP reporter assay was selected.

Multiplexing of GFP reporter system in physiological conditions

GFP reporter and the detector instrument of the fluorescence reporter gene are suitable for our experimental model system. Especially, fluorometry is the most common technique used in biological applications because of its high sensitivity, simplicity, and ability for multiplexing. The GFP reporter assay is well-known for its noninvasive and inherently fluorescent. Also, employing different spectrally resolved mutants of GFP enables us to track the expression of two or more genes in the same cell, a multiplexing system.

Real-time monitoring and visualization of gene expression

Because of the brightness of fluorescent reporters, they can be visualized by employing a fluorescent microscope. Therefore, they are often employed in Spatio-temporal and visual imaging, as well as *in vivo* applications. The fluorescent proteins do not require substrates or reagents for detection; however, most need oxygen to form their chromophores. Besides, without lysing cells, fluorescent proteins can be detected, while this is difficult for luciferase. One of the most significant disadvantages of fluorescent reporters is that they need to overcome cellular

auto-fluorescence. However, in general, the cellular autofluorescence is lower for the emission wavelength of the red channel, such as EGFP or iRFP.

Since the cloning and enhancement of the GFP, GFP has been popular as a reporter gene. Certainly, GFP has been used extensively to visualize spatial and temporal patterns of gene expression *in vitro* and *in vivo* and study intracellular patterns of protein localization and gene trafficking in live cells. The main benefits of GFP are that it is noninvasive and inherently fluorescent so that it does not require substrate for detection. Further, using different spectrally resolved mutants of GFP enables an investigator to track two (or more genes) expressions in the same cells (multiplexing). Therefore, this is extensively useful for characterizing the activity of transcription factor binding to its cognate sequences.

Furthermore, GFP fluorescence could be detected in individual eukaryotic cells by flow cytometry. Therefore, GFP expression could be simultaneously analyzed in a mixed population of cells, or analysis of reporter gene expression could be limited to a distinct subpopulation of cells such as only transfected cells. Furthermore, the capacity to monitor GFP accumulation in living cells before flow cytometry has the potential to simplify kinetic analyses of reporter gene expression greatly (Figure 1.11).

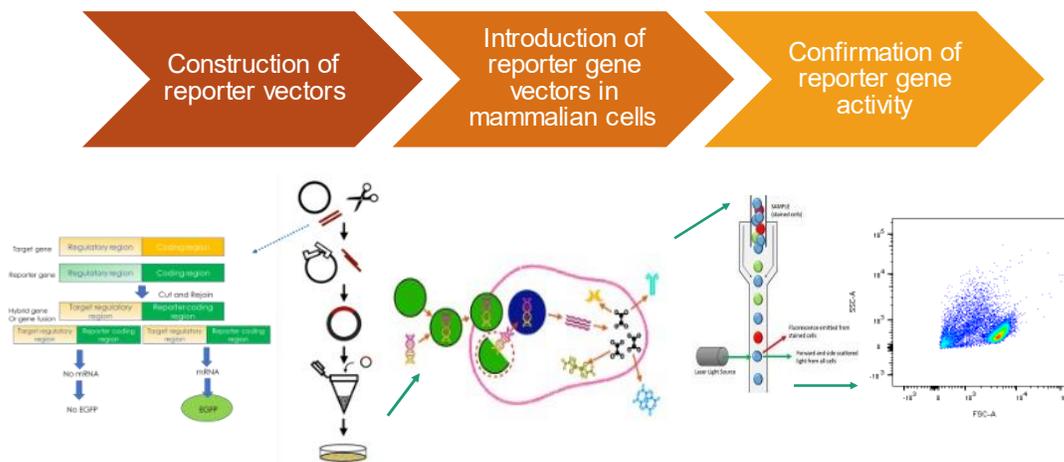


Figure 1.11 Overview of the cell-based reporter

First, we constructed our interest of target, *PU.1*-expressing plasmid and its responsive gene element expressing plasmid. Then, the plasmids were amplified for the delivery into mammalian cells. Using chemical reagents (here in the scheme), two expression vectors were transfected. Once the two expression vectors were successfully introduced and expressed, the two reporters as indicators of TF activity were detected by flow cytometer and projected as a flow cytometric diagram. Depending on the FSC (forward scatter) and SSC (Side scatter) as the laser lights hit, cells are represented its characteristics. Each blue dot denotes a single cell. Highly dense cells represent as red and are considered as live cell population.

Academic interests to pharmacological application

Inevitably, as the gene project progresses, the focus will shift from discovering genes to monitoring their *in vivo* function. Systems sensitive to transcriptional changes within cells and are amenable to HTS will be pivotal in this role. The reliability, reproducibility, sensitivity, and adaptability of reporter gene technology to HTS have made cell-based assays an increasingly attractive alternative to *in vitro* biochemical assays. The principal advantage of these cell-based assays is that they are robust and can provide information about ligand-receptor interactions. They also represent model biological systems that mock physiological conditions and provide

important information about the bioavailability and cytotoxicity of compounds. Moreover, the ability to maintain these cells in culture for several weeks allows for long-term observations of any adaptive changes associated with drug resistance and side effects.

1.8 Objectives

1.8.1 Needs for the research

A suitable and versatile reporter platform to functionally characterize PU.1 activity in live cells is needed for translating knowledge from biophysical research into a biological context. Cellular reporter gene platform enables us to understand molecular control mechanism by PU.1 to bind its target DNA at TF-DNA level in which most focal point of gene regulation occurs.

The rationale for studying the molecular mechanism of transcription factors

The study of transcription factors comprises the analysis of their role in regulating gene expression and how these changes in gene expression affect cell/organism functions. Transcription factors interact at promoters to modulate the transcription of genes. Because the initial control step of regulation happens at which the TFs interact with its cognate binding sites. In that sense, a complete view of transcription factor action at the molecular level requires the knowledge of target genes and binding partners, the effect these interactions have on gene expression, and the mechanisms involved in recognizing specific DNA sequences, among others. Part of the reason for these difficulties is that TFs appear to act in a highly complex manner. Many TFs bind cooperatively, and we are far from having a complex description of which TFs interact with one another or how they select their binding sites when they interact. Even TFs interact with one another, or how they choose their binding sites when they do interact. Even TFs that bind DNA independently may recruit transcriptional machinery in a combinatorial fashion after binding to influence gene expression.

As we mentioned above, PU.1 exhibits specific features for its preference on interaction with primary DNA sequence motifs that modulate binding specificity. For example, ETS proteins were functionally governed by the intrinsic properties of the ETS domain in terms of sequence usage of each ETS protein as well as interactions with other cofactors. The high-throughput microarray and Chip-seq analysis demonstrated that the ETS proteins targeted DNA specificity and stressed the intrinsic selectivity of the ETS domain. Also, biophysical studies of the PU.1/DNA complex revealed that PU.1 is sensitive to osmotic conditions for recognition of its binding sites. Furthermore, unique structural features of PU.1 that lack of structural autoinhibitory domain of PU.1 displayed that PU.1's inhibitory mechanism is based on the interaction with its specific DNA binding sites. Therefore, we need experimental tools to study quantitative gene regulation, allow for the rapid analysis of many PU.1-specified regulatory sequences, and be easily multiplexed to explore several different TFs.

A cell-based reporter gene platform for pharmacological application

The previous finding demonstrated the viability of transcription factor PU.1 as a pharmacological target and opened new possibilities for antileukemic therapeutics through small molecule inhibitors (84). The small compounds, which bind to DNA minor grooves and have a high affinity and specificity on AT-sequences in flanking regions, are attractive to develop as effective allosteric inhibitors of PU.1-DNA complexes. The ability of PU.1-dependent gene inhibition without significant cytotoxicity and nucleus localization in the live cell has potential for antileukemia therapeutics. This strategy proposes the complete loss of low PU.1-expressed leukemia stem cells using the DNA minor groove inhibitors because their removal might lead to leukemia stem cell failures. In an effort to find a variety of DNA-targeted transcription factor inhibitors, we need to develop a cell-based platform measuring the feasibility of PU.1 inhibition.

This dissertation attempts to evaluate DNA minor groove binders and understand their structure-function relationships. It is also to establish a quantitative analytical method to measure the dications' levels in biological samples. From these aims, the biological behavior characteristics of these dications are examined in two types of cell-based reporter technologies (Figure 1.12).

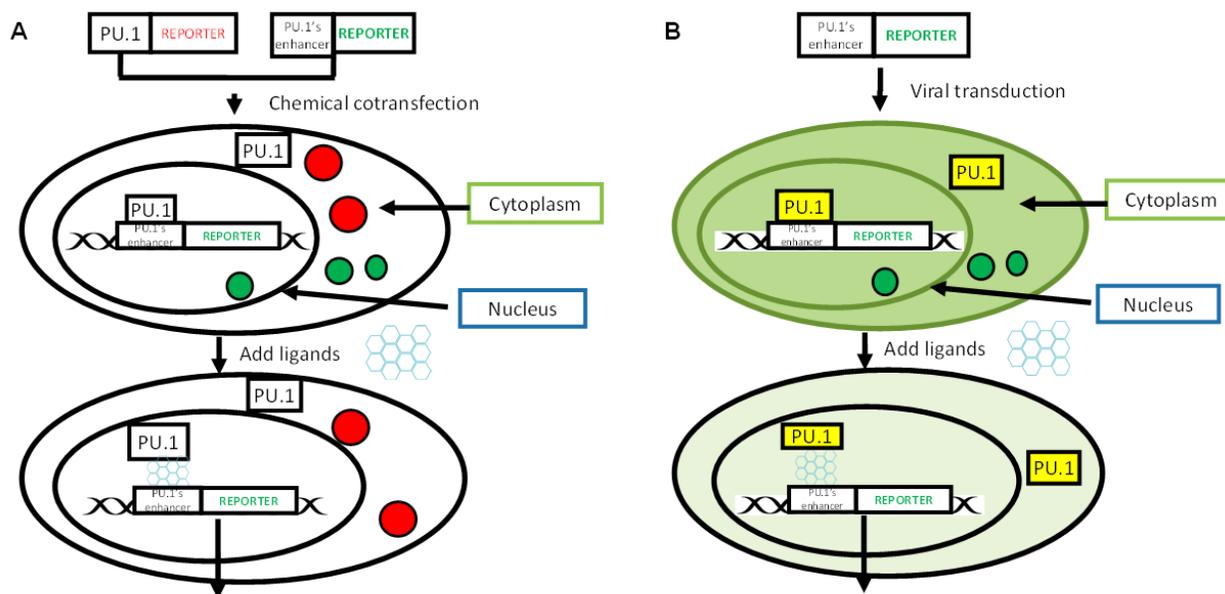


Figure 1.12 The principle of *in vitro* PU.1 transactivation assay for screening ligands against PU.1 activity.

(A) Two expression vectors are co-transfected into HEK293 cells. After ligands bind the PU.1's enhancer, reducing the expression of PU.1-enhancer's reporter gene due to the interference by the ligands is detected. (Red circles denote iRFP fluorescence, and green circles represent EGFP marker expression) (B) Using lentiviral transduction technology, the lentiviral expression vectors coding PU.1 cognitive enhancer fused to reporter marker are introduced into native PU.1-expressing cells (THP-1). Once the ligands bind to the PU.1 enhancer, the EGFP reporter expression is reduced (The green fluorescence is decreased and detected by flow cytometer).

Firstly, we developed a dual-color reporter platform to characterize DNA-sequence-specific binding ligands and proved the three pioneering compounds already characterized by biophysical studies (SPR) that enabled interfering ectopic PU.1 activity to bind its cognate binding site. Then, we further developed the reporter gene platform to use a disease-relevant

experiment model that natively PU.1 is expressed. The current translation rate of successful molecules between biophysical studies and preclinical trials remains low. We hypothesize that structure-function investigations through a cell-based platform would enable us to couple fundamental structural properties of heterocyclic cations to their corresponding cellular properties. This would enable the design of a more efficient screening process for potential therapeutics by allowing the screening of a more abundant compound set. The result would be an ability to design more effective treatments with better PU.1 transactivation inhibition. This dissertation also provides an understanding of the compounds' structural and functional properties to design more valuable compounds.

1.8.2 Research aims

Reporter assays for TFs and their promoter interactions have been used to gain insights into their function and transcriptional regulation pathways. As significant differences exist between species concerning ligand recognition, it is necessary to adapt these tools for transcription factors and other species. In the present work, we describe a PU.1 -specific cell-based assay adapted to analyze its transactivation activity. We employed reporter two dual-colored reporter genes coupling one for measuring transcription factor PU.1 activity depending on its target DNA binding and form the TF-DNA complex and the other for PU.1 dependent enhancer (Ig λ 2-4) as a study model. By measuring the reporter protein expression using flow cytometry, we characterized transcription factor PU.1 activity in living cells to understand its cellular and molecular control mechanism in terms of function. We detected the product of the PU.1 dependent enhancer gene response relative to the PU.1 expression for correlating with PU.1 transactivation activity. A dose-dependent response was detected upon the function of PU.1 transactivation with good correlation to design of enhancer construct (enhancer syntax in terms

of density and spacing) or to compete with dominant-negative mutants as inhibitors of PU.1 activity. Our cell reporter gene assay demonstrated that PU.1's behaviors in mechanistic studies from biophysical characterization were biologically relevant in the cellular environment.

Further, our versatile reporter gene platform was utilized for pharmacological implications in drug screening. We attempted to develop a cell-based platform to enable possible antileukemic agents and screen and investigate structure-function relationships of compounds to understand better and design drugs to target DNA ligands. To accomplish this, we measured the inhibitory properties of heterocyclic dications on PU.1 inhibition *in vitro* and evaluated the compounds' respective potency to assist in developing future compounds. First, we assessed the compounds' viability as potential inhibitors of PU.1-dependent gene transactivation. Afterward, employing the compounds' intrinsic fluorescence, the compounds were examined further to correlate specific structural elements to biological properties. The goal was to investigate the respective target specificity. Fluorescently labeled proteins in the cell allowed us to track the compounds' distribution, providing insight into off-target activities to correlate them with any resulting cytotoxicity. After that, we linked uptake kinetics and cell entry mechanisms to the compounds' structural features, which can significantly help design the compounds.

- a. We designed and constructed cellular reporter gene constructs for investigating PU.1 activity
- b. We functionally characterized PU.1 transactivation in live cells in the perspective of two crucial biophysical characteristics: molecular hydration of DNA recognition by PU.1 and PU.1's trans-regulation in a negative feedback manner.
- c. We characterized ligand screening against PU.1 in the perspective of the pharmacological application.

Our research aims to provide a foundation for analyzing the mechanism that modulates the PU.1 transactivation, which would help us design new ways to manipulate PU.1 transactivation in disease-related function using small molecules knowledge directly used for designing small molecules drugs.

2 DESIGN AND CONSTRUCTION OF CELLULAR REPORTERS FOR PU.1 ACTIVITY

2.1 Introduction

We made efforts to investigate the characterization of the TF-DNA complex in live cells and applied the knowledge for seeking pharmacological agents as a viable approach to target the complex. To achieve the aims, we have developed a cell-based reporter platform. This chapter presents the design and construction of a cell-based reporter assay for characterizing functional PU.1 activity at the TF-DNA level.

2.1.1 Synthetic reporter gene assay for characterizing PU.1 activities

Utilization of cell-based reporter assay in biological research

Cell-based reporter gene assays are increasingly utilized to study complex biological systems. As transcription factor activity, reporter constructs provide the dynamic outcome of its activity upon the DNA binding sites in living cells. The data can enable tracking cellular processes during cell fate decisions, and reporters designed toward a particular outcome to research its research questions. Furthermore, the method can be multiplexed into arrays that provide dynamic, live-cell counterparts to the more traditional techniques that are inherently destructive and thus report on a snapshot in time. Herein, this study sought to develop design

parameters for cell-based reporter assay, which were applied to investigate the dynamics of a well-known pathway, PU.1 activity, in a model cell line.

Cell-based reporter system to study functional activity at the TF-DNA level

Reporter assay for TF activity generally informs on the complex's connectivity on binding its target; however, the TF assays have not been effectively linked with gene expression due to the complexity of full promoters. Here we primarily used λ B (it is known to be activated solely by PU.1 binding) to understand PU.1's DNA recognition and its binding activity. Genetic reporters have been developed in which a reporter is inserted alongside a gene of interest, such that the expression of the gene leads to the production of the reporter. This strategy thus reports on the activity of promoters under the control of TF, PU.1.

We designed and constructed a dynamic live cell monitoring system for characterizing PU.1 transactivation at the protein-DNA level. Transcription factor PU.1 and its integration with genetic reporters to analyze TF-DNA interaction were used as a model of the TFs to provide dynamic information on TFs' intracellular gene regulation. As PU.1-dependent *cis*-regulating elements, the B-cell-specific enhancers, λ B sites, a target of PU.1, provided an excellent target for investigating the syntax of a synthetic enhancer-reporter, as it has a well-defined PWM and is stimulated by PU.1 to produce its gene product Immunoglobulin (Ig) heavy- and light-chain. The myeloid integrin CD11b and M-CSF1R are mainly expressed selectively on the surface of mature monocytes and macrophages, which PU.1 plays an essential role in the cells' differentiation (207, 208). The promoters of human Cd11b and Csf1r were selected as the PU.1-dependent genes tagged with reporter genes to examine the PU.1 transactivation. The chosen PU.1-specific gene contains a purine-rich sequence and the core binding site 'GGAA/G.' The length and sequence of spacer elements of the λ B site were primarily utilized and varied between

PU.1 binding sites. The observed dependence of PU.1-mediated activity on spacer composition was further investigated from the context of competitive binding with other, more active factors. We then applied existing enhancer elements for endogenous PU.1 and its integration with genetic reporters that indicate changes in gene expression associated with cell phenotypes. Our results suggested design principles for TF reporter constructs and expanded native PU.1-dependent genes that integrate TF and genetic reporters in a disease-relevant experiment model for therapeutic application.

2.1.2 Strengths of our cellular reporter gene platform Versatility: Easy adaptation for research aims

2.1.2.1 Versatility: Easy adaptation for research aims

Characterize the strength of promoters and enhancers

In order to study the strength of the promoter or enhancer, the reporter gene is cloned downstream of the promoter in a vector, and then after transfection, its activity is measured by monitoring the expression of the reporter protein. If the reporter gene expression is high, then the promoter is strong, and if the expression is low, then the promoter is weak. Similarly, the strength of the enhancer region, which is upstream of the promoter, is determined. If the transcription factors, when bound to the enhancer, activate or increase the reporter protein expression, then the enhancer is strong.

Assist in defining or characterizing the role of various transcription factors

Transcription factors play an important role in signal transduction as they are trans-acting factors that bind to the enhancer regions of the DNA upon receiving a signal from the cell and increase or decrease the expression of genes. The effects of the transcription factor are multifaceted, and it is difficult to study its role in cells without using any molecular tool. Reporter

gene assay assists in alleviating this as one can study the effect of a transcription factor on desired regulatory sequence or gene expression by cloning the reporter gene downstream of regulatory DNA sequence in a vector. And then, transfecting vector in cell type or cell line where the target transcription factor is expressed. The cells used for developing cell-based reporter assays should fulfill two requirements: (1) express the PU.1 and (2) carry a reporter system that allows measurement of response to PU.1 binding activity. The reporter gene assays are established to believe that when a PU.1 expresses, the PU.1 expression can be measured by its reporter gene construct. Also, the expressed PU.1 traveled and bind to its specific sequence target enhancer; then, the enhancer initiates and activates the following its reporter gene product. We chose three distinct PU.1 specific enhancers to understand and characterize PU.1 transactivation in live cells.

Selected cis-acting elements of PU.1 activity

At λ B motif of E(λ 2-4) in vitro, a B-cell-specific protein complex containing PU.1, a member of the Ets proto-oncogene family, transcriptional activator whose expression is restricted to hematopoietic stem cells. This enhancer is also necessary for *in vivo* activity because mutations in λ B motif, which prevent this interaction, eliminate enhancer function. The complex which binds to λ B motif appears to contain an additional factor that attaches explicitly to a region nearby to the PU.1 binding site. Interestingly, an additional factor cannot bind to the λ B motif autonomously; it appears to require direct interaction with the TF PU.1 to stabilize its association with the DNA. Therefore, in our system to monitor PU.1 activity without additional required cofactors, the λ B motif enables us to determine PU.1 transactivation in live cells solely.

The myeloid-specific CD11b promoter is selectively expressed on the surface of mature monocytes, macrophages, granulocytes, and natural killer cells. The CD11b surface expression is

increased during cell differentiation of the myeloid cell lines U937 and HL-60 directing by PU.1. Tissue-specific regulated expression of CD11b promoter contain that the first 92 bp of 5'-flanking DNA are sufficient to direct the tissue-specific expression. PU.1 binds the sequence '5'-AAAAGGAGAAG-3' at base-pair -20 of the CD11b promoter and regulates its gene expression.

As with other myeloid promoters, the CSF1R gene contains putative PU.1 binding sites, and PU.1 regulates the promoter. It is also known as PU.1; the macrophage transcription factor directs tissue-specific expression of the macrophage colony-stimulating factor (M-CSF) receptor. PU.1 binds to a specific site in the M-CSF receptor promoter upstream from the primary transcription initiation site and transactivates the M-CSF receptor promoter in nonmacrophage cells. The PU.1 binding sequence 5'-AAAGGGGAAGAA-3' is employed for our cell-based reporter system. The M-CSF receptor is limited to monocyte/macrophage lineage in the hematopoietic system. Recently CSF1R gene is essential for AML or Alzheimer's occurrence related to PU.1.

By the three PU.1 binding enhancers/promoters, we characterized functional PU.1 activity in live cells at the TF-DNA level. Unlike Ig λ B motif, as in the case of Cd11b and Csf1r, PU.1 binds to a site just upstream of the major transcriptional start site in a promoter that lacks a consensus TATAA box or initiator sequence. Further, the study of the promoter recognition by PU.1 (e.g., Csf11r) and its transactivation will provide knowledge of PU.1 activity in the structural insights and its pharmacological application.

2.1.2.2 Dual-colored monitoring system for PU.1 activity in live cells

We systemized the monitoring of both PU.1 and its dependent gene transactivation by fusing two different reporter genes of PU.1 and its specific target promoters. The cell-based reporter is adaptable for multiple expression vectors in cells by chemical transfection. Since

many fluorescent protein indicators are developed by mutation of GFP and are available, two different fluorescent proteins enable us to track PU.1 and enhancer activity. The reporter expression in a live cell can be measured by flow cytometry and fluorescence microscope.

Multi-dimensional flow cytometry

Flow cytometry is the science of measuring individual cells' molecular, physical, and chemical properties as they pass through a flow chamber and are interrogated by one and more lasers. The flow cytometer simultaneously and quantitatively measures parameters such as size, granularity, and surface molecule expression for each cell. Flow cytometers have developed capable of performing high-throughput screening and high-level content analysis, assessing many different samples' features in a single run up to 1536 formats on multiple cell populations. The launch of imaging flow cytometry has filled the gap between flow cytometry and conventional high content imaging screening, putting flow cytometry at the center of many laboratories.

Flow cytometry's informative potential has beneficial our cell reporter system. Currently, multicolor analysis of immune cells is the standard in immunology, and this provides a multi-dimensional view of all markers expressed on individual cells. The dual-color reporter for tracking PU.1 and its dependent gene expression using flow cytometry provides us with large multi-dimensional data sets of individual cells' physical and molecular characteristics. Among the various choice of spectrally fluorescent reporter proteins, iRFP and EGFP were chosen for our application, in which the spectrally distinct each other.

Multimodality reporter gene imaging: construction strategies and application

Molecular imaging has played an essential role in the noninvasive exploration of multiple biological processes (207-209). Reporter gene imaging is a crucial part of molecular imaging. By

combining with a reporter probe, a reporter protein can produce the accumulation of specific signals detectable by an imaging device to deliver indirect information of reporter gene expression in living subjects. Fused reporter genes or single reporter genes with products detectable by multiple imaging modalities can pay for the weakness and potentiate the benefits of each modality.

Multicolor tracking PU.1 activity on its dependent gene

Dual-color reporter system in live cells enables to monitor TF activity on its specific target gene by visualization of biological process, characterization, and measurement of biological processes at the molecular and cellular levels in living systems. It allows longitudinal, non-invasive, quantitative, and repetitive imaging of targeted biological processes at both the molecular levels. Molecular imaging technique allows numerous applications, including monitoring endogenous transcriptional regulation, gene transfer analysis, or characterization of disease, evaluating treatment, and expeditions drug discovery.

2.1.3 Objectives of this chapter

This chapter described the design and construct cell-based reporter platform to identify the molecular mechanisms governing lineage- and tissue-specific cell fates' decisions through transcription factor PU.1 and its DNA recognition. Molecular information from living cells can be gained from reporter constructs that provide activities for either individual transcription factors or multiple factors binding to the full promoter following the expression of reporter genes. An overview of the design criteria to obtain reporters that are specific and responsive to TF binding is presented. The integration of TF and genetic reporters recognized the major hubs directing responses enhancer activities, and this method provided information to investigate the regulation of TF activity at the TF-DNA level. Moreover, targeting the AT-rich preference of

PU.1's binding characteristics is utilized, and strategies for pharmacological application of screening ligands against PU.1 activity. For the therapeutic implication, we have developed a cell-reporter platform to characterize screening small molecules to modulate PU.1 transactivation in live cells. Ultimately, we delineated the design and construction of the reporter assay platform.

2.2 Experimental design Versatility: Easy adaptation for research aims

2.2.1 Principle of cellular reporter's design for PU.1 activities

2.2.1.1 The expression vectors

The pcDNA3.1 vectors are employed for their high-level, constitutive expression in various mammalian cell lines (Figure 2.1).

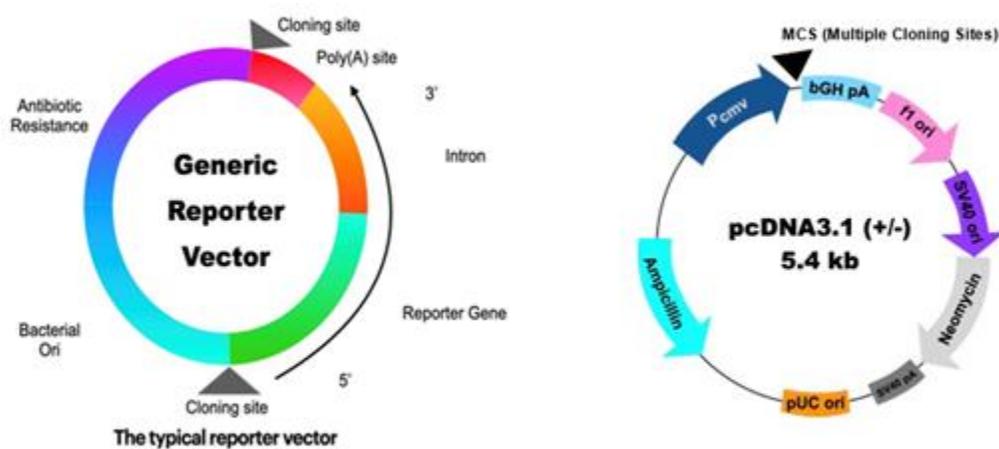


Figure 2.1 Illustration of a reporter vector

For the design of the reporter vector, the backbone of the vector should produce protein production. No regulatory binding sites or sequences other than one should not exist. Also, extra control elements should include antibiotic resistance for the selection. 'Ori' denotes the prokaryotic origin of replication so that amplifying the reporter vector plasmid in the bacterial competent cells and yields plasmid DNA. Multiple cloning sites (MCS) contain two locations upstream of the reporter gene for cloning putative promoter/ enhancer. Another MCS is the

plasmid for cloning regulatory elements like acts as a distance. Polyadenylation signals (Poly(A) site) enhances mRNA stability and translation in mammalian cells. Poly A is located immediately following the reporter gene that the signal additions of 200-250 adenylate residues to the 3' -end of an RNA transcript. The arrow is the direction of transcription of the reporter gene. Illustration of our primary reporter vector (pcDNA3.1(+)) is in the right. Each detail of the components is described in table 1.

Table 2.1 Detailed features of pcDNA 3.1(+)

Feature		Function
P cmv	Human cytomegalovirus (CMV) immediate-early promoter/enhancer	Permits efficient, high-level expression of recombinant protein
T7 promoter	T7 promoter/priming site	Allows for <i>in vitro</i> transcription in the sense orientation and sequencing through the insert
MCS	Multiple Cloning Site in forward or reverse orientation	Allowing insertion of the interested gene and facilitates cloning
BGH pA	Bovine growth hormone polyadenylation signal	Efficient transcription termination and polyadenylation of mRNA
f1 ori		Allows rescue of single-stranded DNA
SV40 ori	SV40 early promoter and origin	Allow efficient, high-level expression of the neomycin resistance gene and episomal replication in cells expressing SV40 large T antigen
Neomycin	Neomycin resistance gene	Selection of stable transfectants in mammalian cells
SV40 pA	SV40 early polyadenylation signal	Efficient transcription termination and polyadenylation of mRNA
pUC ori		High-copy number replication and growth in <i>E. coli</i>
Ampicilin	Ampicillin resistance gene (β -lactamase)	Selection of vector in <i>E. coli</i>
Ampicilin (<i>bla</i>)	Ampicillin (<i>bla</i>) resistance gene (β -lactamase)	Allows selection of transformants in <i>E. coli</i>
*Kozak sequence		vertebrate consensus sequence for strong initiation of translation

*Source from invitrogen

Table 2.2 Mammalian expression plasmid utilized in the research

Name	Description / vector	Study used
pcDNA 3.1(+)	Full-length (FL) human PU.1, hPU.1 ETS domain	Dominant negative mutants
pD2EGFP	Enhancer: λ B, cd11b, csf1r	Reporter for screening ligands against PU.1 PU.1 characterization at PU.1/enhancer DNA binding
piRFP	A phytochrome-based near-infrared fluorescent protein, also known as iRFP713	Control iRFP expression plasmid
pRRLSIN.cPPT.PGK.MCS.wPRE	The 3 rd generation lentiviral backbone. Mammalian expression The enhancer of transcription factor PU.1: λ B, cd11b, csf1r	Reporter gene platform for screening ligands against PU.1
pMD2.G	VSV-G envelop expressing plasmid. Mammalian expression, lentiviral. Envelop. Gene/Insert name: VSV G	Reporter gene platform for screening ligands against PU.1
pRSV-Rev	The 3 rd generation lentiviral packaging plasmid; Contains Rev; also requires pMDLg/pRRE and envelope expressing plasmid Mammalian expression, lentiviral; packaging. Gene/Insert name: Rev (a.k.a HIV1gp6)	Reporter gene platform for screening ligands against PU.1
pMDLg/pRRE	The 3 rd generation lentiviral packaging plasmid; Contains Gag and Pol; also requires pRSV-Rev and envelop expressing plasmid. Mammalian expression, lentiviral; packaging. Gene/insert name: HIV-1 GAG/POL	Reporter gene platform for screening ligands against PU.1

pRRLSINcPPT.PGK.GFP.wPRE	Control EGFP expression plasmid for the 3 rd lentiviral backbone plasmid	Control GFP expression plasmid
--------------------------	---	--------------------------------

2.2.1.2 Choice of fluorescent reporter

The choice of the fluorescent marker was considered on two points: photostability and sensitivity of detection for our experimental purpose. Fluorescence probes, EGFP (enhanced green fluorescent protein), and iRFP were used for our cell-based reporter system. Both proteins are intracellular stable (photostable). Also, the signal-to-background ratio is low, indicating that both proteins are bright and suitable for detection. The EGFP (488nm/511 nm) and iRFP (640/670>nm) are spectrally well separated so that no overlap of the spectrum resulted in increased sensitivity of signal detection. It should have both excitation and emission maxima within a near-infrared window from ~650 nm to 900 nm because the tissue has the lowest absorbance and less light scattering than in the shorter wavelength part of the spectrum (210).

Further, we incorporated self-cleavage peptide and destabilized sequence for improving detection sensitivity because we examined the dynamic transcription factor PU.1 's activity in live cells, and it might be affected by proper folding of PU.1 and the half-life of EGFP protein.

iRFP marker for PU.1

iRFP is a phytochrome-based near-infrared fluorescent protein (iRFP) with excitation and emission maxima at 690 nm and 713 nm(211). Respectively. The fluorescent properties of phytochromes are photosensory receptors that absorb light in the red and far-red parts of the spectrum. Valuable protein reported for *in vivo* imaging, a near-infrared fluorescent mutant of the DrBphP bacteriophytochrome from *Deinococcus radiodurans*, named IFP1.4 (comparison control). RpBphP2, bacteriophytochrome RpBphP2 from the photosynthetic bacterium *Rhodospseudomonas palustris*. Use it for template and genetically engineered to make iRFP.

iRFP in mammalian cells are characterized in photostability, and no intracellular aggregates are found. Degradation kinetics of iRFP with maturation rate at 37°C, a maturation half-time of 2.8 h, has been reported. The iRFP protein is pH stable with a pKa value of 4.0. The protein stability indicates that degradation time-course throughout 20 h. Further, non-cytotoxicity has been documented by that iRFP expressing cells behaved similarly to the cells expressing noncytotoxic control cells. iRFP does not need an exogenous supply of the chromophore biliverdin (BV) and has higher adequate brightness, intracellular stability, and earlier photostability than phytochrome-derived fluorescent probes. Compared with far-red GFP-like proteins, iRFP has a higher signal-background ratio in a mouse model due to its infrared-shifted spectra. iRFP is the brightest phytochrome-based and the fluorescent protein whose spectrum is shifted most to the infrared. Stable, noncytotoxic, and low concentrations of endogenous BV are sufficient to brightly fluorescent cells, tissues, and whole animals.

2.2.1.3 Self-cleavage 2A peptide for PU.1 expression: the need for multicistronic vectors for multigene delivery

We incorporated our PU.1 reporter gene with the 2A-self cleavage peptide (212). The strategy to overcome both highly variable expression levels of transgene and transgene stability in expression due to loss, re-arrangement, or silencing of transgenes. Silencing is exacerbated by increasing transgenic loci numbers and repeated use of homologous promoter or coding sequences. We integrated the licensing strategy to overcome both these issues by expressing multiple genes' products from a single promoter in the form of a self-processing polyprotein. On translation, virus-encoded proteinases usually facilitate intramolecular (*cis*) cleavages of the polyprotein to produce separate protein products. However, in the foot-and-mouth disease virus (FMDV), a key cleavage at the C-terminus of the 2A region appears to be facilitated by just 16-

20 amino acids of the 2A sequence itself, a unique and entirely mechanism. Therefore, incorporating the 2A sequence between the iRFP reporter gene and FL-human PU.1 gene functional products are independently folding correctly and not interfering with the PU.1 to interact with its target gene for transactivation. Most target vectors are typically used to replace a part of the coding sequence of an endogenous gene or all of it. This leads to loss of function. With 2A-sites, it is possible to preserve the gene function and integrate and co-express multiple reporter genes in one gene locus. For this purpose, reporter genes must be placed in a frame with the endogenous genetic code before the endogenous stop codon. The endogenous gene and the reporter genes need to be split with 2A-sites. About 15-20 amino acids, with the addition of a 2A-site, are inserted into the C-terminus of the prior gene.

Since widely used viral promoters, the cytomegalovirus (CMV), the choice of the promoter is of particular importance for PU.1. Therefore, tissue-specific promoters or tetracycline-regulated promoters are the promoters of choice. Restriction sites can be included, flanking each genetic element of the repair vector to replace the resistance gene and/or the reporter genes. With our particular vector design, we conserve the gene function of target genes. The transgenes do not replace the target genes but are expressed along with the host gene under the control of the endogenous promoter. The reporter genes are connected in frame with the endogenous gene divided by 2A-sites. The own promoter drive for its own resistance gene. The transgenes are incorporated at the end of the gene before its stop codon.

2.2.1.4 The utility of d2EGFP as a transcription reporter to that of EGFP

GFP is a widely used reporter in the research of gene expression and protein localization. GFP fluorescence does not involve any substrate or cofactor; therefore, it allows us to use it in many species for live cell detection. We used Enhanced GFP (EGFP) reporter for PU.1-

responsive gene because it is spectrally distinctive from iRFP. Also, to overcome one of the limitations for the long-lasting stability for GFP protein resulting from compact structure, we introduced D2EGFP to increase detection for rapid transcriptional activity in live cells.

The PEST sequence in protein degradation plays a role in several short-lived proteins, including that at the C terminus of MODC. The metabolic enzyme ornithine decarboxylase (ODC) is a critical regulatory control point in polyamine biosynthesis and is regulated by 26S proteasome degradation, albeit without ubiquitination, mediated by two PEST regions. D2EGFP has utilized the concept to incorporate EGFP in the expression vector to facilitate the EGFP protein degradation shorten.

The degradation domain of mouse ornithine decarboxylase was fused to the C terminus of EGFP and used for all PU.1-dependent enhancer reporter construct. The D2EGFP decreased the half-life of EGFP in mammalian cells to 2 hours (213). The mutation of key amino acids in the PEST sequence of the fusion protein and identified several mutants with different half-lives. The use of destabilized EGFP as a transcription reporter makes gene induction study possible in real-time with living cells. Analysis with flow cytometry suggested that the half-life of dEGFP is approximately 2 hours.

The rapid turnover of dEGFP has three advantages. First, the quick turnover of dEGFP allows its application involving destabilizing GFP, such as circadian rhythms studies. Second, its rapid turnover leads to less accumulation in cells to have lower toxicity in the cells when stably expressed. Thus, establishing a stable cell line using this genetic fusion in place of unmodified EGFP will be at ease. Lastly, the destabilized version of EGFP can be used as a transient reporter to study *cis*-acting regulatory elements or transcriptional induction.

2.2.1.5 Workflow for the lentiviral delivery for developing cell reporter platforms

One of our main research interests, PU.1, ETS transcription factors regulate the self-renewal and differentiation of hematopoietic cells and highly involved in a spectrum of diseases. From the understanding of the principles that govern the interactions between PU.1 and DNA to pharmacological implication, the lentiviral gene delivery system has been introduced to develop a stable cell line because we examined the endogenous PU.1 activity focusing on the disease experiment model system in which most of the cell lines are hard to chemically transfected. Chemical transfection has limitations for the purpose.

Weakness of transfection for delivery of genetic materials

Even with extreme care, the amount of DNA taken up by different populations (tissue culture dishes) of cells during transfections often varies significantly, making comparative analysis difficult. The availability of multiple reporter genes provides a convenient means of controlling for the variability in transfection efficiency. Typically, cells are co-transfected with a DNA mixture composed of two separate plasmids, each harboring a different reporter gene: an experimental reporter gene whose expression is controlled by the test DNA and a second reporter gene whose expression is driven by a standard (e.g., viral promoter/enhancer) regulatory element. The resultant activity of the reporter enzyme in each cell extract is normalized to the activity of the control reporter enzyme with that extract. It should be noted that standardization of data to a second reporter enzyme activity controls the inter-assay variables due to transfection efficiency and extract collection. In contrast, standardization against the total protein content of the extract only controls for the latter.

When transfecting an inducible vector, it is possible to end up with so many copies of the vector in a cell that the leakiness associated with all inducible vectors is magnified, which may mask its effect doxycycline on the cells. Further, when it is optimized, the amount of plasmid

DNA used to minimize this effect, with transduction, can control the number of viral particles going into the cells to achieve a more accurately inducible system. Lentiviral vectors provide the strengths for the genetic material delivery systems (Figure 2.2).

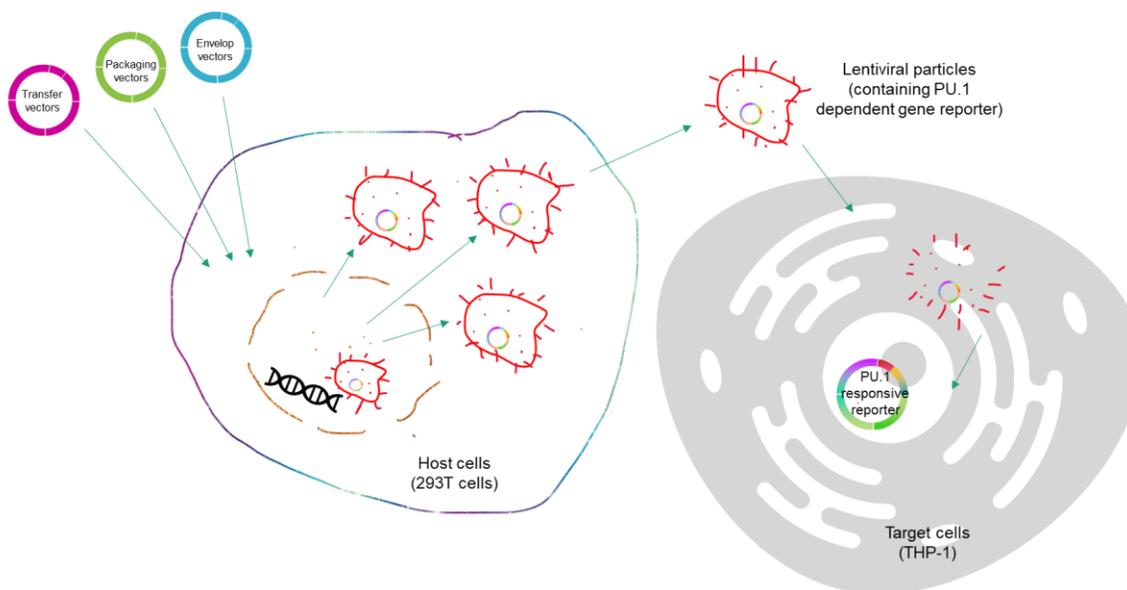


Figure 2.2 Lentiviral transduction: Delivery of gene transfer vector into a mammalian cell in the research.

Lentiviral transduction is a method using a lentivirus carrier to deliver the interest of a gene. The chemical carrier transiently opens pores into the cell membrane to permit the entry of genetic materials. The chemical transfection is helpful to introduce two or more genetic materials into the target cells with marker genes and selectable advantages such as resistance or labeling. The chemical transfection is efficient on adherent immortalized cells such as HEK293 cells, but suspension cells such as THP-1, primary, and stem cells require transduction. Lentiviral transduction is the transfer of DNA using a lentivirus as a vector. Lentiviral transduction is useful for mammalian cells that are hardly transfected by chemical reagents. Also, use a lentiviral vector to transduce target cells that naturally integrate their DNA into the host cell genome, enabling us to develop stable cells. To our experiment model, HL-60 or THP-1, endogenous PU.1-expressing cells, the lentiviral transduction technique was used to develop the stable reporter cell lines to characterize screening ligands against PU.1 activity (214).

The lentiviral vehicles can transduce a wide range of cell types and integrate into the host genome in both dividing and post-mitotic cells, resulting in long-term expression of the

transgene. This is helpful to screen pharmacological reagents against PU.1 activity. Our experiment model cell line HL-60 and THP-1 endogenously express a high level of PU.1, derived from the myeloid cells in leukemic patients. The cell lines are helpful to screen ligands before preclinical screenings. Further, a cell-based reporter platform has been widely used to integrate a High-Throughput-Screening system to identify "Hit" molecules. We have designed the lentiviral vehicles containing PU.1-dependent genes tagging with fluorescent markers with strength in our minds.

2.2.2 *Plasmid design and construction*

2.2.2.1 *piRFP-2A-PU.1*

An expression PU.1 plasmid was previously reported(84). The Human full-length PU.1 sequence (Full name: Transcription factor PU.1, size 31 kD, Sequence ID, P17947.2) was cloned by inserting a fragment encoding the human full-length PU.1 fused to an infraRFP (iRFP) reporter between restriction enzyme NheI and BamHI sites of pcDNA3.1(+). In between PU.1 and iRFP, a fragment encoding P2A (5'-GGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAA CCCTGGACCT-3') self-cleavage peptide was inserted and flanked by AgeI restriction enzyme sites to the PU.1 sequence. Then, an expression plasmid encoding full-length PU.1 was cloned between the NheI/BamHI sites of pcDNA3.1(+) with iRFP tagged. All constructs were validated by DNA sequencing (Macrogen, MD).

2.2.2.2 *pλBx5-d2EGFP*

The expression plasmid pλBx5-EGFP was previously reported for the construction. The CMV promoter of a commercial enhanced green fluorescent protein (EGFP) reporter plasmid (pd2EGFP, Clontech) was replaced with a synthetic enhancer element consisting of five tandem

repeats of the λ B site, spaced one helical-turn apart and followed by a minimal TATA-box promoter. The fragment encoding pentameric λ B site (5'-AATAAAAGGAAGTG-3') was cloned into restriction enzyme sites AgeI and BglII in pd2EGFP. First, the DNA fragment encoding λ B motif was amplified by PCR, followed by digestion with the restriction enzymes. Then, the fragments were ligated with T4 ligase into the AgeI/BglII digested vector pD2EGFP (Clontech, CA). Plasmid DNA was obtained by propagating DH5 α competent cells followed by plasmid DNA extraction with a plasmid miniprep kit according to the manufacture's protocol. The construct was verified by Sanger sequencing (Macrogen, Bethesda, MD).

2.2.2.3 Dominant-negative mutants for characterizing evolutionarily significant site of PU.1 ETS domain

In the research of various dominant-negative mutants of PU.1, the PU.1 mutant DNA construct was synthesized by IDT DNA technologies (Midland, IA, USA). Each DNA fragment of PU.1 mutant used in the study was codon-optimized, amplified by PCR, and subcloned into restriction enzyme site NheI/HindIII of pcDNA3.1(+) eukaryotic expression. Briefly, the hPU.1 ETS dominant-negative mutant DNA fragments were mutated one amino acid codon Asn (N, AAT) to Tyr (Y, TAC), His (H, CAT), and Gln (Q, CAA). The insert fragment is 318 bp and cloned into multiple cloning sites after the CMV promoter of the pcDNA3.1(+). The DNA sequencing confirmed all constructs (Macrogen, Bethesda, MD).

2.2.2.4 Synthetic enhancer PU.1-dependent EGFP reporters for characterizing PU.1's self-association

For the study of the effect of enhancer structure on PU.1 transactivation, several PU.1-sensitive enhancer sequences as described in the text were also purchased from IDT. Briefly, the DNA fragment consists of a minimal PU.1-cognate binding site derived from the murine Ig λ 2-4

enhancer (GenBank X54550). The EGFP reporter protein expression is under the control of the minimal PU.1-dependent enhancer harboring various tandem of a native cognate site for PU.1. Each PU.1-specific ETS binding site (EBS), λ B motif, was spaced by 20 base pairs (bp) or two helical turns. The fragment was encoded in the sequence '5'-AATAAAAAGGAAGTG-3,' and various copies of the site (1x, 2x, 3x, 3x-alt, and 5x) were amplified PCR, following digestion by the restriction enzymes. Then, the fragments were ligated with T4 ligase into the AgeI/BglII digested vector pD2EGFP (Clontech, CA). Plasmid DNA was obtained by propagating DH5 α competent cells followed by plasmid DNA extraction with a plasmid miniprep kit according to the manufacture's protocol. By DNA Sanger sequencing, all constructs were verified (Macrogen, MD).

2.2.2.5 *Lentiviral reporter constructs for characterizing to screen ligands against PU.1 activity*

The 3rd generation lentiviral transfer vector pRRLSIN.cPPT.PGK.MCS.wPRE was used for cloning pRRL-iRFP-2A-PU.1, pRRL- λ B-D2EGFP, pRRL-Cd11b-D2EGFP, and pRRL-Csf1r-D2EGFP. Briefly, each DNA fragment was obtained from piRFP-2A-PU.1, p λ B-D2EGFP, pCd11b-D2EGFP, pCsf1r-D2EGFP by digesting corresponding restriction enzymes. The fragment of PU.1 in piRFP-2A-PU.1 was digested by BamHI/SalI and transfer into the TOPO vector (Invitrogen) first. Then, The TOPO-PU.1 vector was amplified by PCR and digested by BamHI/SalI again. The insert was cloned into the site after the promoter Human Glycerate Kinase (hgk) of pRRLSIN.cPPT.PGK.MCS.wPRE. The insert and the lentiviral transfer vector were ligated with T4 ligase. According to the manufacturer's protocol, the cloned vector pRRL-iRFP-2A-PU.1 was transformed into Stbl3 competent cells, and a plasmid miniprep kit extracted the DNA. The PU.1-dependent gene constructs in pD2EGFP (λ Bx5, Cd11b, and Csf1r) were digested by EcoRV/SalI, where the site was replaced with promoter site hgk of the lentiviral

transfer vector. Ligation was carried out with T4 for 1 h incubation for 22°C, and the ligated construct was transformed to Top10 competent cells with 42 C for 30 s heat shock process following them for 1 h in the 37°C shaker. Then, the cloned construct cells were plated and incubated at 37°C for 14-16 hours. Lastly, the single colony was selected and cultured for DNA extraction by the DNA miniprep kit according to the manufacture's direction. Each construct was verified by Sanger sequencing.

2.2.3 *Cell culture, transfection, RT-PCR*

2.2.3.1 *Cell cultures and maintenance*

The adherent cell lines, Human Embryonic Kidney (HEK) 293 and HEK293T, were cultured in DMEM medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 1% (v/v) Penicillin and Streptomycin (P/S). The HEK293 cells and HEK293T cells were maintained at 37°C in humid air with 5% CO₂. Before reaching 80-90% confluence, cells were subcultured using 0.25% (v/v) trypsin. The trypsin was neutralized with cell culture media (above). The medium was refreshed every 2-3 days.

Additionally, human monocytic leukemia cell line, HL-60, and THP-1 cells were obtained from American Type Culture Collection (ATCC) and cultured in RPMI1640 medium supplemented with 10% (v/v) heated-inactivated FBS. Every 3 to 4 days, they were maintained by replenishing the fresh culture media and keeping the confluency below 8×10^5 /mL in a T-75 flask. Where indicated, cells were induced with a single dose of PMA at 16nM (THP-1) or 160 nM (HL-60) for 72 hours (final dimethyl sulfoxide concentration: 0.1% v/v).

All cell lines were maintained at 37°C under a 5% CO₂ incubator.

2.2.3.2 *Transient transfection for functional characterization of PU.1 activity*

HEK293 cells were seeded in 24-well plates (7×10^4 cells /well) and transfected with 500 ng of each pcDNA-FL-PU.1 and p λ Bx5-d2EGFP after 24 h (JetPRIME, Polyplus-Transfection) according to the manufacturer's instructions. Briefly, 500 ng of each plasmid of DNA was diluted in 50 μ l of transfection buffer and vortex for 10 seconds and spun down. 1 μ l of transfection reagent was added into the mixture, vortex for 1 second, spun down and incubated for 10 mins at room temperature. Each prepared transfection mixture was added to the cells in seeded well and incubated for 24 hours before checking fluorescence microscope or flow cytometry. The cells were trypsinized for the flow cytometric analysis (0.25% Trypsin EDTA) before acquiring Accuri (Becton Dickinson, BD) or BD FortessaTM instruments. Cellular fluorescence was quantified by flow cytometry (excitation/emission = 488/510 nm and 640/>670 nm). Counts were collected to >20 000 per sample, corresponding to a CV of <5 % for a 2% event frequency, and gated against untransfected controls to eliminate the background. The procedure of cell acquisition by flow cytometry was conducted using Becton Dickinson's instruction. The FACS Vantage with a Coherent Enterprise II argon laser producing 60 mW of 351-364 nm multiline ultraviolet excitation. The flow cytometer was equipped with pulse processing and the Macrosoft flow cells. Cells were loaded, and fluorescence emission was detected via 488/510 (green) and 640/70 (red) emission filters, separated by a 490 long-pass dichroic mirror. Analysis of flow cytometric data was carried out using the BD FACSDiva (Becton Dickinson) or FlowJo (Treestar) software.

2.2.3.3 *Q-PCR*

Following extraction of total RNA using a spin column kit (Omega) and RT (Thermo Fisher Scientific), RT-PCR reactions were performed on a QuantStudio 3 instrument (Applied

Biosystems) with SYBR Green PCR Master Mix (Thermo Fisher Scientific). Expression levels of each gene was normalized to *gapdh*. The primer sequences used for *pu.1*, *csf1r*, *e2f1*, and *gapdh* are referred to in the figure.

2.2.4 *Lentiviral generation transduction for constructing stable PU.1-dependent reporter cell line*

The procedure to transduce expression cells is straightforward; a lentivirus producer cell line (HEK293T) is transiently co-transfected with transfer (encoding the genes of interest), envelope, and packaging plasmids to generate lentiviral particles. The lentivirus-containing supernatants are then concentrated and used to transduce the THP-1 target cell line. After genomic integration of the proviral DNA and limiting dilution, these cells are expanded for large-scale protein expression and are ready for use. The procedure can be broken down into three stages. The lentivirus producer cell line is transfected in the first stage, and lentiviral particles encoding the transgene are collected. In the second stage, the viral particles are confirmed and verified. In the third stage, target cells are infected, stably transduced, and expanded.

2.2.4.1 *Generation of lentiviral particles*

HEK293T cells were used as the host cell line to produce the virus particles. They were maintained at 37°C and under 5% CO₂ in DMEM medium supplemented with 10% (v/v) FBS and 1% (v/v) P/S (growth media). They were plated overnight before being chemically transfected in 10 cm dishes. Cells were 3-50% of confluency in the plate on the day of transfection. 1.25 µg to 5 µg, of a lentiviral expression plasmid encoding PU.1 enhancer genes (λ Bx5, *cd11b*, and *csf1r*), lentiviral packaging plasmids (Rev, Tet) and lentiviral envelope plasmid (pMD.2G) were transfected into the cells using the transfection reagent Jetprime

PolyPlus, following the manufacturer's instructions. The ratio of the mixture of each lentiviral plasmid for the production (Transfer vector: lentiviral envelope plasmid lentiviral packaging plasmid = 5:2.5:1.25) was followed by the instruction. After 24 hours of the transfection, the fresh growth media were replaced. The media containing lentivirus particles were collected 48 hours, 60 hours, and 72 hours after the transfection started. All collected media were spun down at 500 g and filtered in a 0.45 μm PES filter to remove cell debris. Then, the particles were concentrated using the ultracentrifuge concentrator (100X centrifugal concentrator, PALL laboratory). The concentrated virus particles were aliquoted in a microcentrifuge tube, and they were stored at -80°C until they were used.

2.2.4.2 Confirmation of lentiviral particles

We conducted the verification experiments to confirm the lentiviral particle worked as we expected, that the EGFP fluorescence protein expressed solely depending on PU.1 transactivation. Also, the same method was used for the virus titering. Because HEK293 cells were non-PU.1 expression, we performed piRFP-2A-PU.1 transfection first, then the virus particles were inoculated into the PU.1- transfected HEK293 cells. pPU.1-2A-iRFP transfection as previously described. After 24 hours of post pPU.1-2A-iRFP transfection into HEK293 cells, 0.1 μl , 1 μl , 5 μl , 10 μl , 25 μl of viral particles were used to infect the HEK293 cells. The transduced cells and the viral particles were confirmed first by fluorescence microscope, then were transferred to analysis using a flow cytometer. The formula for calculating virus titer, the below equation, was used.

$$\text{Titer (TU/ml)} = (F) \times DF$$

F: The frequency of GFP-positive cells decided by flow cytometry

Cn: The total number of target cells that infected

V: The volume of the inoculum

DF: The virus dilution factor

2.2.4.3 *Establishment of stable PU.1 reporter gene cell lines: THP-1-LV- λ Bx5-EGFP, THP-1-LV-Cd11b-EGFP, and THP-1-LV-Csf1r-EGFP*

THP-1 target cell lines were seeded in a 24-well plate with a density of 5×10^4 cells per well. The transducing RPMI1640 culture media were prepared to contain polybrene (Hexadimethrine bromide, final concentration 8 $\mu\text{g}/\text{mL}$) and diluted virus stocks accordingly after gently thawing the lentiviral stock in the ice. Then, the media were added into the seeded THP-1 cells and incubated in a 5% CO_2 incubator at 37°C . The next day, 100 μl of fresh culture media were added to the transduced cells. After 4 days of the transduction, the lentivirus transduced cells were confirmed using fluorescence microscopy and flow cytometry. On the 8th day post-transduction, stable cells were selected by limiting dilution. These cells were expanded for 12-16 weeks. The stably transduced reporter gene THP-1 cell lines were maintained in the complete culture media. For the compound testing, the reporter cells were seeded at a density of 2×10^5 cells/well in 24-well plates with a culture medium containing the indicated concentration of the drug compounds (ranged from 10^{-8} M to 10^{-5} M). Then the cells were divided into two groups, one incubated at 37°C for 24 hours and the other for 48 hours.

2.3 Results

2.3.1 *Confirmation of piRFP-2A-PU.1 and its expression in HEK293 cells*

Full-length PU.1 used in cell-based experiments was the full length of wild type (Figure 2.3).

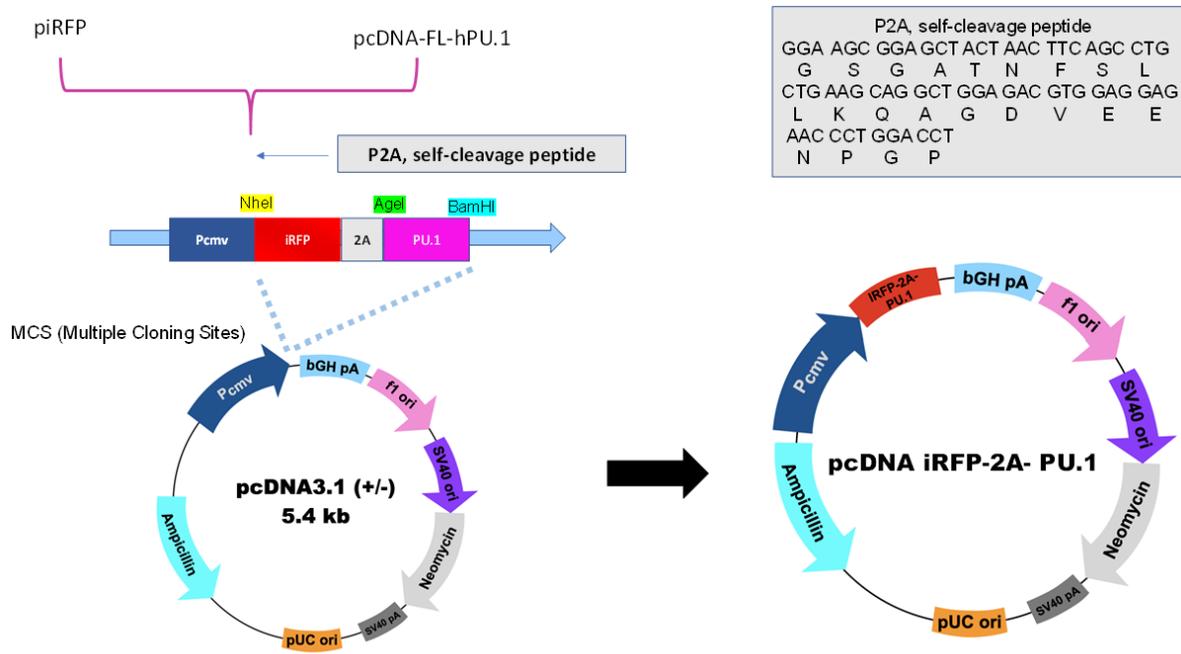


Figure 2.3 Schematic representation of the design of pcDNA iRFP-2A-PU.1.

*An expression plasmid of full-length human PU.1 was cloned between *NheI* and *BamHI* sites of pcDNA 3.1(+). With iRFP and P2A, self-cleavage peptide. The P2A self-cleavage peptide sequence was inserted between the iRFP reporter gene and the FL-hPU.1 gene. The P2A gene functions highly efficiently self-cleaving, the first identified in the foot-and-mouth disease virus. Here we used the porcine teschovirus-1 (P2A), 22 amino acids in length. To prevent different expression of two genes, PU.1 and iRFP, due to the large size, the 2A peptide was introduced Under the human cytomegalovirus (CMV) promoter, in which an element initiates the transcription of the insert, the marker iRFP and PU.1 gene are expressed*

An expression plasmid of PU.1 was cloned by inserting a fragment encoding full-length PU.1 joined to an infrared RFP (iRFP) reporter between the *NheI*/*BamHI* sites of pcDNA3.1(+). Between the full-length PU.1 sequence and the iRFP reporter gene, a sequence encoding a self-cleaving 2A peptide is inserted and linked the two. The design and construction of 2A-peptide-linked multicistronic vectors in our full-length PU.1 expression plasmid overcame the limitations of conventional approaches to express more than two proteins that imbalanced protein expression and large size. The full-length PU.1 transgene was cloned with an iRFP construct via a co-translating 2A peptide (Figure 2.4). The self-cleavage 2A peptide was inserted into between

iRFP gene and full-length PU.1 gene, and it enable isolation of PU.1-expressing cells by flow cytometry. Detecting the iRFP indicator's expression demonstrate the confirmation of the PU.1 gene's expression.

1RFP-2A-PU.1 in pcDNA3.1

Forward Frame 1:

```

1   A S A T G L R S A A T M A E G S V A R Q
1   GCTAGC GCTACCGGACTCAGATCTGCCGCCACCATGGCTGAAGGATCCGTCGCCAGGCAG
121  P D L L T C D D E P I H I P G A I Q P H
61  CCTGACCTCTTGACCTGCGACGATGAGCCGATCCATATCCCCGGTGCATCCAACCGCAT
41  G L L L A L A A D M T I V A G S D N L P
121  GGACTGCTGCTCGCCCTCGCCGCCGACATGACGATCGTTGCCGGCAGCGACAACCTTCCC
61  E L T G L A I G A L I G R S A A D V F D
181  GAACTCACCGGACTGGCGATCGGCGCCCTGATCGGCGCCTCTGCGGCCGATGTCTTGAC
81  S E T H N R L T I A L A E P G A A V G A
241  TCGGAGACGCACAACCGCTGACGATCGCCTTGGCCGAGCCCGGGCGGCCGCTCGGAGCA
101  P I T V G F T M R K D A G F I G S W H R
301  CCGATCACTGTCGGCTTCACGATGCGAAGGACCGCAGGCTTCATCGGCTCCTGGCATCGC
121  H D Q L I F L E L E P P Q R D V A E P Q
361  CATGATCAGCTCATCTTCCTCGAGCTCGAGCCTCCCCAGCGGGACGTGCGCCGAGCCGAC
141  A F F R R T N S A I R R L Q A A E T L E
421  GCGTTCTTCCGCCCGCACCAACAGCGCCATCCCGCCCTGCAGGGCCGCCGAAACCTTGGAA
161  S A C A A A A Q E V R K I T G F D R V M
481  AGCGCCTGCGCCCGCGCGGCAAGAGGTGCGGAAGATTACCGCCTTCGATCGGGTGTAG
181  I Y R F A S D F S G E V I A E D R C A E
541  ATCTATCGCTTCGGCTCCGACTTCAGCGGGCAAGTATCGCAGAGGATCGGTGCGCCGAG
201  V E S K L G L H Y P A S T V P A Q A R R
601  GTCGAGTCAAACCTAGGCTGCACTATCCTGCTCAACCGTCCAGCGGCGCAGCCCGCTCGG
221  L Y T I N P V R I I P D I N Y R P V P V
661  CTCTATACCATCAACCCGGTACGGATCATTCCTCGATATCAATTATCGGCCGGTGCCTGTC
241  T P D L N P V T G R P I D L S F A I L R
721  ACCCCAGACCTCAATCCGGTCACCGGGCGGGCGATTGATCTTAGCTTCGCCATCCTGCGC
261  S V S P V H L E F M R N I G M H G T M S
781  AGCGTCTCGCCCGTCCATCTGGAATTCATGCGCAACATAGGCATGCACGGCAGCATGCG
281  I S I L R G E R L W G L I V C H H R T P
841  ATCTCGATTTTGGCGGGCAGCGACTGTGGGGATTGATCGTTTGCCATCACCGAACGCCG
301  Y Y V D L D G R Q A C E L V A Q V L A W
901  TACTACGTCGATCTCGATGGCCGCAAGCCTGCGAGCTAGTCGCCAGGTCTCTGAGCTGG
321  Q I G V M E E M E E G S G A T N F S L L
961  CAGATCGGCGTGATGGAAGAGATGGAAGAGGGAAGCGGAGCTACTAACTTCAGCCTGCTG
341  K Q A G D V E E N P G E V A T M E G F P
1021  AAGCAGGCTGGAGACGTGGAGGAGAACCTTGGACCGCCACCATGGAAAGGTTTCCC
361  L V P P P S E D L V P Y D T D L Y Q R Q
1081  CTCGTCCACCTCCATCAGAAGACCTGGTGCCTATGACACGGATCTATACCAACGCGCAA
381  T H E Y Y P Y L S S D G E S H S D H Y W
1141  ACGCAGGATATTACCCCTATCTCAGCAGTGTATGGGGAGAGCCATAGCGACCATTACTGG
401  D F H P H H V H S E F E S F A E N N F T
1201  GACTTCCACCCACCACGTCACAGCGAGTTCGAGAGCTTCGCCGAGAACCACTTCACG
421  E L Q S V Q P P Q L Q Q L Y R H M E L E
1261  GAGCTCCAGAGCGTGCAGCCCCCGCAGCTGCAGCAGCTTACC GCCACATGGAGCTGGAG
441  Q M H V L D T P M V P P H P S L G H Q V
1321  CAGATGCACGTCCTCGATACCCCATGGTGCACCCCATCCAGTCTTGGCCACCAAGGTC
461  S Y L P R M C L Q Y P S L S P A Q P S
1381  TCCTACCTGCCCGGATGTGCCTCCAGTACCATCCCTGTCCCCAGCCAGCCAGCTCA
481  D E E E G E R Q S P P L E V S D G E A D
1441  GATGAGGAGGAGGGCGAGGGCAGAGCCCCCTACTGGAGGTGTCTGACGGCGAGGGCGAT
501  G L E P G P G L L P G E T G S K K K I R
1501  GGCCTGGAGCCCGGGCCTGGCTCCTGCTGGGGAGACAGGCAGCAAGAAGAAGATCCGC
521  L Y Q F L L D L L R S G D M K D S I W
1561  CTGTACCAGTTCCTGTGGACCTGCTCCGACGCGGCACATGAAGGACAGCATCTGGTGG
541  V D K D K G T F Q F S S K H K E A L A H
1621  GTGGACAAGGACAAGGGAACTTCCAGTTCTCGTCCAAGCACAAGGAGGCGCTGGCGCAC
561  R W G I Q K G N R K K M T Y Q K M A R A
1681  CGCTGGGGCATCCAGAAGGGCAACCGCAAGAAGATGACCTACCAGAAGATGGCGCGCGG
581  L R N Y G K T G E V K K V K K K L T Y Q
1741  CTGCGCAACTACGGCAAGACGGGCGAGGTCAAGAAGGTGAAGAAGAAGCTCACCTACCAG
601  F S G E V L G R G G L A E R R H P P H -
1801  TTCAGCGCCGAAGTGTGGGCCCGGGGGCCCTGGCCGAGCGGGCCACCCGCCCACTAG
621  G S
1861  GGATCC

```

Figure 2.4 A sequence of iRFP-2A-PU.1 in pcDNA3.1(+) and recognition site of the restriction enzymes for cloning.

The 2A self-cleavage peptides are inserted between the iRFP gene and hPU.1 (Colored in gray). Right after the 2A gene, the AgeI (colored in green), hPU.1 gene is following. NheI sites in the MCS after the pCMV site are used and colored in yellow. (RecName: Full=Transcription factor PU.1; Alt Name: Full=31 kDa-transforming protein [Homo sapiens] Sequence ID: P17947.2 Length: 270 bp).

Determination of RNA abundance of transfected PU.1 in HEK293 cells

In order to measure the mRNA expression and its amount of PU.1 in the transfected cells, 10 μ g of expression piRFP-2A-PU.1 was transfected in 1×10^6 HEK293 cells (Figure 2.5).

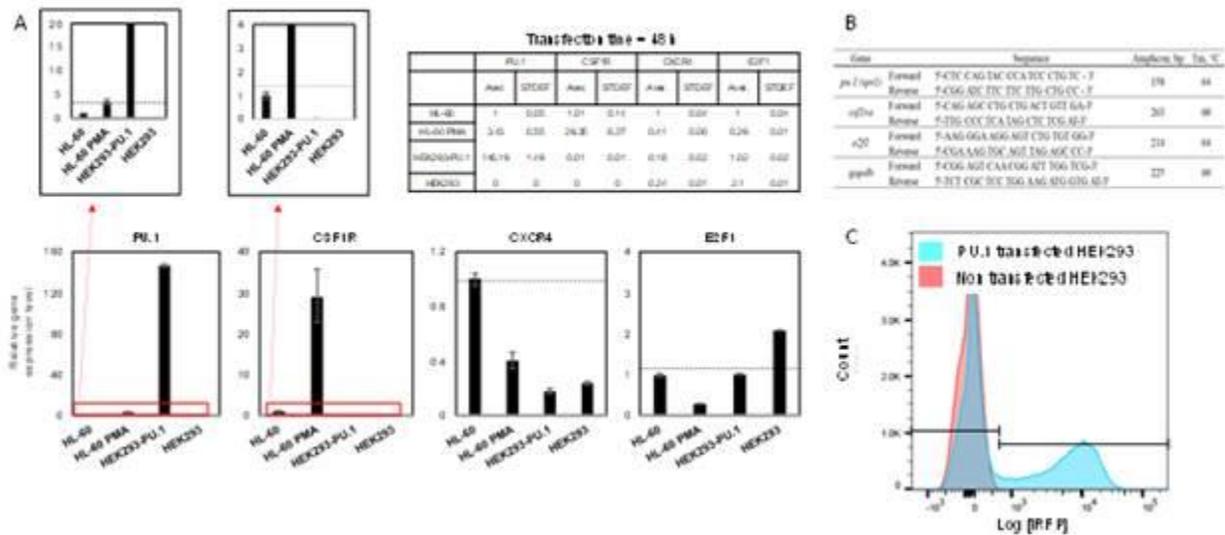


Figure 2.5 Identification of PU.1 expression level of RNA and iRFP marker in the PU.1 transfected HEK293 cells.

(A) RNA expression level in HL-60, HL-60 PMA (Phorbol 12-myristate 13-acetate, PMA), HEK293, and HEK293-PU.1. Native PU.1 expressing HL-60 cells was used as a positive control for the expression level comparison. PMA is an inducing chemical for HL-60 by activating PKC (Protein Kinase C) to differentiate macrophage-like cells, which contains high levels of PU.1 expression. Extraction of total mRNA abundance using an RT-PCR technique represented the expression levels of genes were normalized to housekeeping gene gapdh. Gene *csf1r* is a native PU.1 binding gene (1×10^6 cells PU.1 transfected HEK293). *CSF1R*, known as

macrophage colony-stimulating factor receptor, M-CSFR (target gene products of PU.1), and CXCR4 and E2F1 (negative regulated), are detected. (B) Table of primers used for RT-PCR detection. (c) Representative flow cytometric data of HEK293 and HEK293-PU.1 cells. 1×10^6 HEK293 cells were transfected with 2 μ g of piRFP-2A-PU.1. The RT-qPCR data is credited to Dr. Lee, Sang Choon.

The mRNA abundance from RT-PCR displayed that the expression level of the transfection after 48 hours is about 140 times higher than HL-60, natively PU.1 expressing cell line, and more than 40 times higher than the PMA-induced HL-60 cells, which is macrophage-like cell type. Control expression of Csf1r in HL-60 and the PMA-induced HL-60 indicated that the Csf1r gene is silenced in HEK293 cells and transfected HEK293 cells. Contrastingly, the activity of E2F1, a negative regulator of PU.1, was confirmed in PU.1 transfected HEK293 cells compared with non-transfected HEK293 cells and the PMA-induced HL-60 cells. Therefore, PU.1 in HEK293 cells was functionally expressed and worked in downstream gene activation or repression.

2.3.2 Confirmation of pentameric p λ B-D2EGFP

To measure functional PU.1 binding activity on its cognate sequence, p λ B-D2EGFP was developed (Figure 2.6)

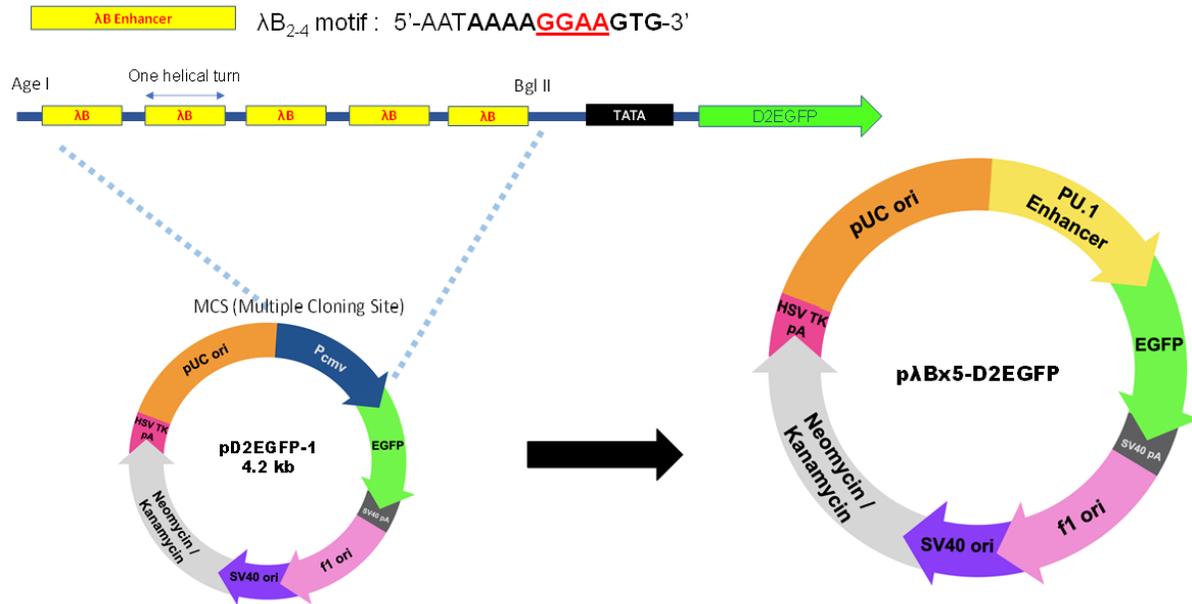


Figure 2.6 A schematic diagram of the synthetic enhancer reporter gene vector construct.

A vector map of pD₂EGFP was used for the reporter construct. D₂EGFP denotes destabilized variants of wild-type GFP. D₂EGFP was developed for overcoming the limitation of long and stable GFP protein because it might interfere in studies that require rapid reporter turnover. In order to characterize PU.1 activity focusing on its target DNA recognition, the utilization of D₂EGFP limits the drawback of GFP stability. Murine Immunoglobulin λB₂₋₄ enhancer was fused with D₂EGFP so that this EGFP is expressed under the control of a minimal promoter downstream from a synthetic λB –based enhancer.

By replacing the CMV promoter in front of the EGFP reporter gene with the λB motif, EGFP expression was detected only under the PU.1 binding on the λB site. The enhancer element consists of five tandem repeats of the λB site, spaced one helical-turn apart and followed by a minimal TATA-box promoter. Upon the functional expression of PU.1 and binding on the enhancer elements, EGFP reporter was expressed and detected by flow cytometer or fluorescence microscope. The expression plasmid of λB-D₂EGFP was confirmed by Sanger sequencing (figure 2.7)

LB5-D2EGFP-1_EGFP-NR.ab1	GGGGCACTCCGTGGTGACGCTCCTCGCCCTTGTCCACCATGGTGGCGACCGGTAGCGCTA	60
LB5-D2EGFP-6_EGFP-NR.ab1	-GGGGACCCCGTGGTGAAGCTCCTCGCCCTTGTCCACCATGGTGGCGACCGGTAGCGCTA	59
	*** ** *****	
LB5-D2EGFP-1_EGFP-NR.ab1	GCGGATCTGACGGTTCACTAAACCAGCTCTGCTTATATAGACCTCCACCGTACACGCCT	120
LB5-D2EGFP-6_EGFP-NR.ab1	GCGGATCTGACGGTTCACTAAACCAGCTCTGCTTATATAGACCTCCACCGTACACGCCT	119

LB5-D2EGFP-1_EGFP-NR.ab1	ACCTGCAGAAATAAAAGGAAATGAAATAAATAAAAGGAAGTGATCCAAATAAAAGGAAGT	180
LB5-D2EGFP-6_EGFP-NR.ab1	ACCTGCAGAAATAAAAGGAAAGTGAATAAATAAAAGGAAGTGATCCAAATAAAAGGAAGT	179

LB5-D2EGFP-1_EGFP-NR.ab1	GAAATAAATAAAAGGAAGTGATCCAAATAAAAGGAAGTGAGATCTGAGTCCGGTAGCGCT	240
LB5-D2EGFP-6_EGFP-NR.ab1	GAAATAAATAAAAGGAAGTGATCCAAATAAAAGGAAGTGAGATCTGAGTCCGGTAGCGCT	239

Figure 2.7 Sequence confirmation of a reporter gene expression vector containing *IgλB2-4*, a murine enhancer element bound by PU.1.

Comparative sequence analysis was confirmed by Sanger's sequence method, and the alignment was performed by Clustal Omega (EMBL-EBI). The minimal enhancer harboring only cognate binding sites for PU.1 was confirmed I lined in yellow. (Red line represented start codon).

2.3.3 Functional expression confirmation of *piRFP-2A-PU.1* and *pλB-D2EGFP* in live cells

Two expression plasmids encoding full-length PU.1 fused with iRFP reporter conjugated with a self-cleavage peptide 2A and a pentameric PU.1-specific enhancer fused with d2EGFP marker were confirmed by flow cytometric analysis (Figure 2.8). Untransfected HEK293 cells and upon transfection with the same dose of the plasmid of control reporter plasmid were aligned with the expression level (fluorescence intensity) of each construct of the plasmid. As we confirmed, the λ B-D2EGFP did not exhibit its expression level without PU.1 expression. The 250 ng of each *piRFP-2A-PU.1* and *pλB×5-D2EGFP* were co-transfected into HEK293 cells, and the reporter positive cells both iRFP and EGFP were observed in Q2 only. The functional PU.1 activity has analyzed the expression of EGFP level among all iRFP expression cells (Q2+Q3). Each dot denotes a single cell. The fluorescence intensity of each cell indicates the level of fluorescence marker protein.

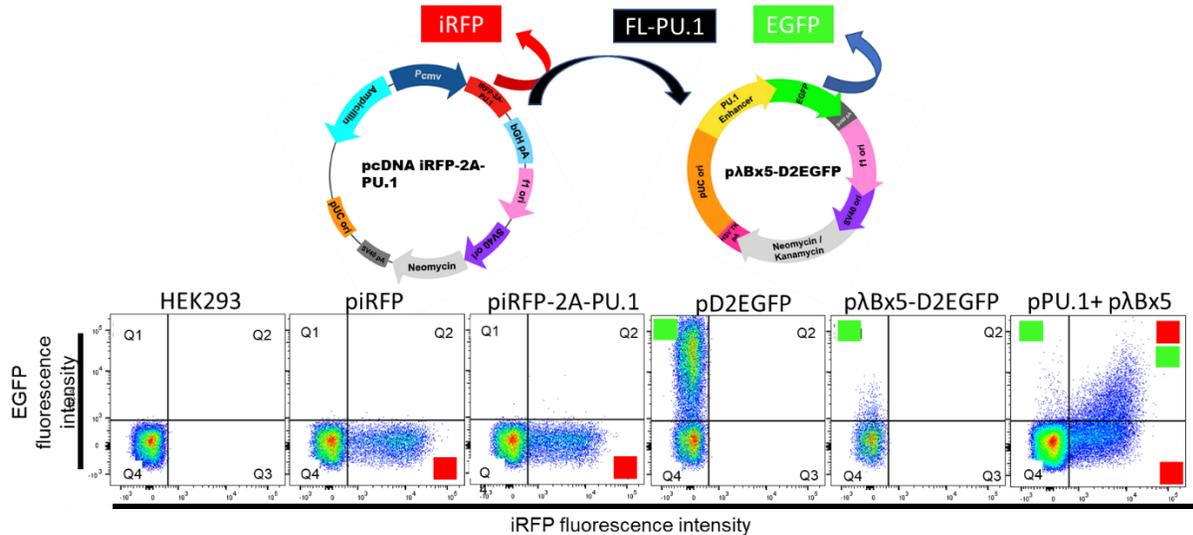


Figure 2.8 Flow cytometric confirmation of piRFP-2A-PU.1 and pλBx5 expression in transfected HEK293 cells.

The control plasmid expression of piRFP and pD2EGFP represented the successful confirmation of chemical transfection and its marker protein expression. The dual color monitoring system for the experiment is illustrated in the flow cytometric diagram (X-axis, fluorescence intensity of iRFP expression, Y-axis, the fluorescence intensity of EGFP expression). The gating was based on the non-transfected HEK293 cells (The leftist) and transfectant of each piRFP and pd2EGFP, which each control marker protein expression. Q2 contains only cells expressing EGFP positive cells with PU.1 expressing cells. In Q3, without any transcriptional requirement in HEK293 cells, iRFP is expressed under the CMV promoter and is the indicator to express PU.1 (left panel) and monitored by the fluorescence signal via flow cytometry (middle). As we expected, pλBx5-D2EGFP is expressed and illustrated in Q1. Every 50 μg of the reporter plasmid construct per sample was used for the transient transfection in non-PU.1 expressing cells in 24- well plates according to the transient transfection protocol.

2.3.4 Design and construction of dominant-negative mutants for characterization functional PU.1 transactivation at TF/DNA level

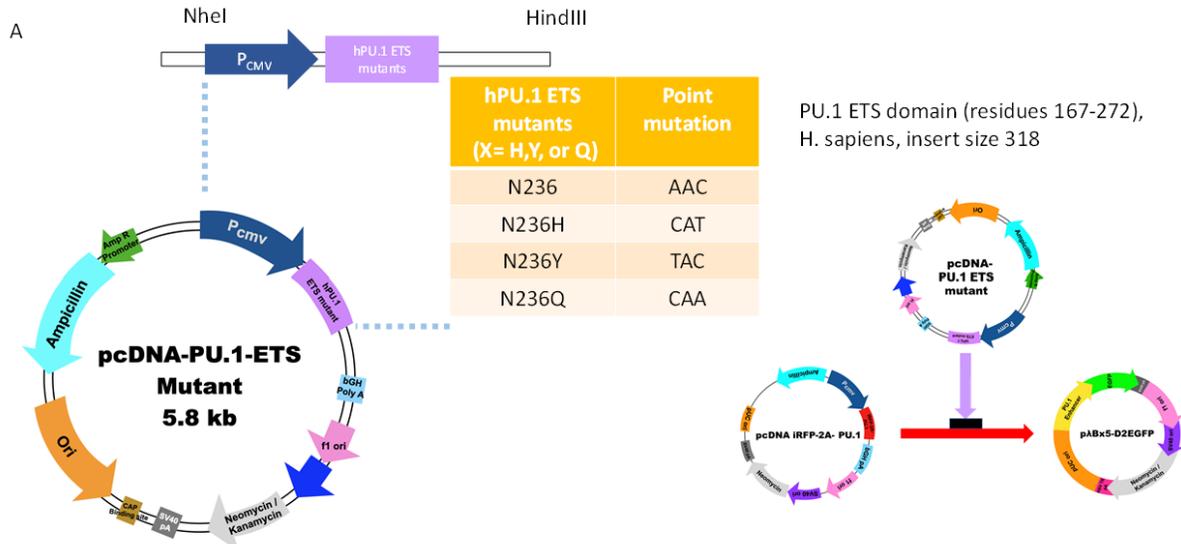
To study the function of the wild-type genes, mutants are used as a powerful tool. Dominant-negative mutants' inhibition happens in which the function of a wild-type gene product is damaged by a co-expressed mutant variant of the same gene product. Most proteins comprise more than one structural domain, and these domains assemble together to form the

native structure. With this perspective, the assembly of the native structure of a monomeric and an oligomeric protein share features in common.

We determined the functional significance of residue in PU.1 linked to interfacial hydrations effect in perspective of the interaction of PU.1 and its site-specific DNA to form the complex through cell-based reporter assay. The dominant-negative mutants of hPU.1 ETS were designed from a sequence alignment of ETS domains with a view of maintaining the register of residues that are most conserved among ETS paralogs. The secondary structural elements that comprise the DNA contact surface in order of primary structure are H2, loop, H3, H3/S3, S3, wing, and S4, the last of which we had previously examined. These elements together comprise the conserved ‘winged helix-loop-helix motif. H3 is the groove-binding recognition helix that interacts with the conserved 5’—GGA(A/T)-3’ core consensus in ETS binding sites. The extended loop and short wing connect H2 and H3, and S3 and S4, respectively. The H3/S3 segment is a sharp turn between H3 and S3 that includes several residues assigned as part of H3 in the PU.1/DNA complex. Here, we follow the assignment specified by the Ets-1/DNA co-crystal structure. From the study of the binding affinity and osmotic sensitivity, the H3 chimera has lost significantly in binding affinity more than 50-fold greater than those of wild-type PU.1 (WT PU.1), even in the biological level of reduction was significant compared to that of WT PU.1 to the low-affinity specific sites (113, 215).

By determining the effect of the dominant-negative effects over the wild-type protein, the evolutionarily functional significance of variations of the critical residue of H3 PU.1 ETS domain; therefore, our cell-based reporter assay using the mutant constructs as the dominant-negative inhibitor of PU.1 transactivation proved itself as a helpful application in support of biophysical experiments and computational research.

One of the previous biochemical characteristics of PU.1's DNA recognition revealed that the H3 recognition helix of PU.1 is sensitive to the osmotic stress on its high-affinity DNA sequence, contrasting to its structural homolog of Ets-1 and the findings led to finding N236 in PU.1 ETS domain for the responsible residue in H3 recognition helix. Interestingly, residue 236 is the most various residue among the conserved ETS domain across the ETS paralogs. For understanding its evolutionary importance, we utilized the idea of dominant-negative mutants to examine the effect of the residue on PU.1 transactivation (Figure 2.9).



B

```

N236H   GGAGATATGAAAGACTCAATTTGGTGGGTTGACAAAGACAAAGGCACCTTTTCAATTCAAGT 238
N236Y   GGAGATATGAAAGACTCAATTTGGTGGGTTGACAAAGACAAAGGCACCTTTTCAATTCAAGT 240
N236Q   GGAGATATGAAAGACTCAATTTGGTGGGTTGACAAAGACAAAGGCACCTTTTCAATTCAAGT 235
WT      GGAGATATGAAAGACTCAATTTGGTGGGTTGACAAAGACAAAGGCACCTTTTCAATTCAAGT 238
*****

N236H   TCAAAGCATAAGGAAGCTCTGGCCCATCGTTGGGGAATTCAAAAAGGTAAACCGCAAGAAA 298
N236Y   TCAAAGCATAAGGAAGCTCTGGCCCATCGTTGGGGAATTCAAAAAGGTAAACCGCAAGAAA 300
N236Q   TCAAAGCATAAGGAAGCTCTGGCCCATCGTTGGGGAATTCAAAAAGGTAAACCGCAAGAAA 295
WT      TCAAAGCATAAGGAAGCTCTGGCCCATCGTTGGGGAATTCAAAAAGGTAAACCGCAAGAAA 298
*****

N236H   ATGACTTACCAGAAAATGGCGCGTGCCTTGCGTACACGGAAAAACAGGGGAGGTGAAG 358
N236Y   ATGACTTACCAGAAAATGGCGCGTGCCTTGCGTTACACGGAAAAACAGGGGAGGTGAAG 360
N236Q   ATGACTTACCAGAAAATGGCGCGTGCCTTGCGTCAAACGGAAAAACAGGGGAGGTGAAG 355
WT      ATGACTTACCAGAAAATGGCGCGTGCCTTGCGTAACACGGAAAAACAGGGGAGGTGAAG 358
*****

N236H   AAAGTAAAGAAAAAATTACATATCAATTCAGCGGGGAAAGTTTTAGGACGCGGTGGCTTG 418
N236Y   AAAGTAAAGAAAAAATTACATATCAATTCAGCGGGGAAAGTTTTAGGACGCGGTGGCTTG 420
N236Q   AAAGTAAAGAAAAAATTACATATCAATTCAGCGGGGAAAGTTTTAGGACGCGGTGGCTTG 415
WT      AAAGTAAAGAAAAAATTACATATCAATTCAGCGGGGAAAGTTTTAGGACGCGGTGGCTTG 418
*****

N236H   GCTGAACGCCGTCACCCACCTCATTGAAAGCTTGGTACCGAGCTCGGATCCACTAGTCCA 478
N236Y   GCTGAACGCCGTCACCCACCTCATTGAAAGCTTGGTACCGAGCTCGGATCCACTAGTCCA 480
N236Q   GCTGAACGCCGTCACCCACCTCATTGAAAGCTTGGTACCGAGCTCGGATCCACTAGTCCA 475
WT      GCTGAACGCCGTCACCCACCTCATTGAAAGCTTGGTACCGAGCTCGGATCCACTAGTCCA 478
*****

N236H   GTGTGGTGGAAATTCGCAGATATCCAGCACAGTGGCGGCCGCTCGAGCTAGAGGGCCCG 538
N236Y   GTGTGGTGGAAATTCGCAGATATCCAGCACAGTGGCGGCCGCTCGAGCTAGAGGGCCCG 540
N236Q   GTGTGGTGGAAATTCGCAGATATCCAGCACAGTGGCGGCCGCTCGAGCTAGAGGGCCCG 535
WT      GTGTGGTGGAAATTCGCAGATATCCAGCACAGTGGCGGCCGCTCGAGCTAGAGGGCCCG 538
*****

N236H   TTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC 598
N236Y   TTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC 600
N236Q   TTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC 595
WT      TTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC 598
*****

```

Figure 2.9 Schematic diagram for characterization of PU.1/DNA complex.

The Helix 3 (H3) DNA recognition domain of PU.1 ETS domain revealed the importance of DNA-contact surface area under the absence of osmotic stress from DNA biophysical and biochemical studies. To examine the essential functional consequence of the sequence variation in the domain, we found the residue 236 revealed its only and phylogenetic analysis among the corresponding residues in the ETS TF paralogs first. (a) Design of dominant-negative mutant construct. We designed the domain mutants to evaluate the sequence variation. Residue 236 (Asparagine) in wild-type human PU.1 was mutated to histidine (ETV-6), tyrosine (Ets-1), and glutamate (PDEF). The PU.1 ETS mutant interferes with the PU.1 binding to λ B enhancer, thereby the EGFP expression upon its binding response decreases. (b) Sequence confirmation of dominant-negative mutants of human PU.1 ETS motif for characterization of PU.1/DNA complex in the red box. Sanger's sequence method was carried out for the comparative sequence analysis, and Clustal W O (version 1.2.4) was used for the alignment.

The construct of dominant-negative mutants was co-transfected with piRFP-2A-PU.1 and p λ B-D2EGFP in HEK293 cells and measured the level of expression of EGFP, in which the effects of dominant-negative mutants intervened PU.1 transactivation (Figure 2.9 A) compared with wild-type of hPU.1 ETS. If exogenous PU.1 would be blocked or inhibited, the reduction of the EGFP marker is detected. The degree of reduction might represent the phylogenetic relationship across the ETS TF members. Each mutant was verified by Sanger's sequencing (Figure 2.9B).

2.3.5 Intrinsic biophysical properties of PU.1 autoregulation and reporter gene assay

PU.1 is inducible for a high expression level for the cells to be differentiated and terminated for their identity in blood cells. In previous research, one explanation for its upregulation or downregulation demonstrated the inhibitory function of the ETS domain in structural biology. The self-regulatory function in the ETS family of TFs has autoinhibition in the structural DNA binding domain. The ETS domains formed a 1:1 complex as a *trans*-regulatory complex for autoinhibition. Contrastingly, PU.1, a member of the ETS TF family, has

demonstrated a complex form, including a unique form of 2:1 complex that no other ETS family member contains.

Our cell-based reporter platform was designed to determine whether transcription factor PU.1 is functionally inhibited by self-association mechanisms (Figure 2.10).

A

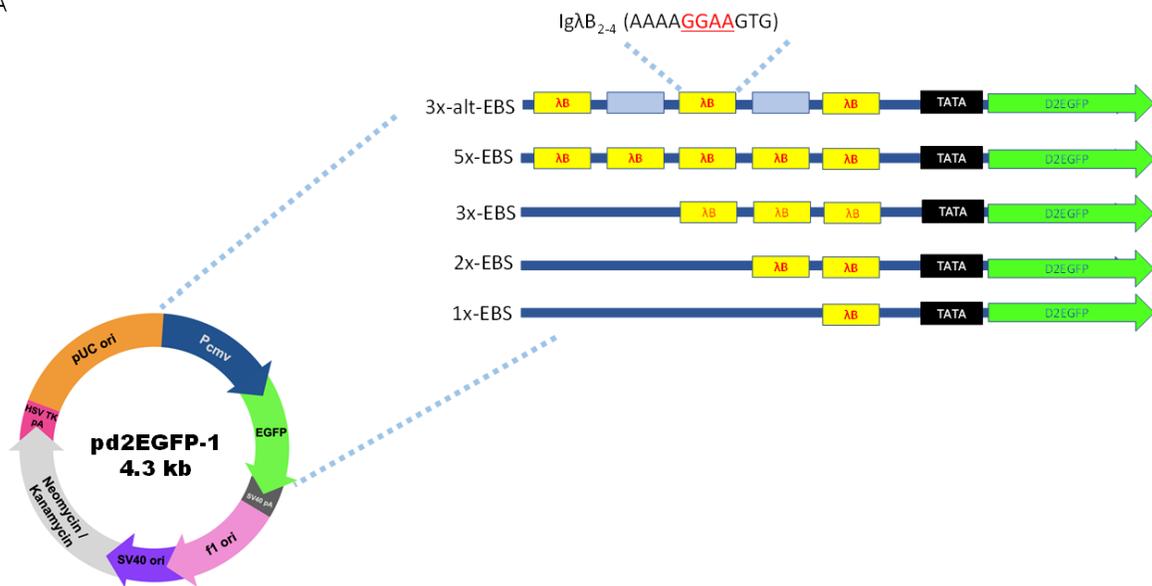




Figure 2.10 A schematic diagram of a synthetic PU.1-dependent EGFP reporters for characterization of PU.1 self-regulation mechanism.

(A) A minimal TATA box was driven by enhancers composed only of tandem spaced 20 bp apart. The light blue denotes mutated sites. We manipulated the number of a PU.1-specific ETS binding site (EBS) $I\gamma B_{2-4}$ (5'-AAAAGGAAGTG-3') and inserted it into the expression vector, each pD₂EGFP at the restriction enzyme site AgeI and BglII. The ability of strength of reporter expression is based on the PU.1 binding, and the structural changes of numbers of elements and spacing between them were designed for readily suitable for PU.1/DNA complex formation and characterization of PU.1 negative feedback mechanism at the TF-DNA level. The hypothesis is that the increasing number of binding sites is detecting by the increasing EGFP reporter signals as PU.1 dose manner. The functional outcome of an inactive, negatively

cooperative complex would be a bell-shaped reporter dose-response as the enhancer, which varies in density and spacing of the PU.1-specific ETS binding sites. (B) Sequence alignment confirmation of the synthetic enhancer reporter constructs by Clustal W. The sequence in red boxes is confirmed in 1x-EBS, 2x-EBS, 3x-EBS, 5x-EBS, and 3x-alt-EBS.

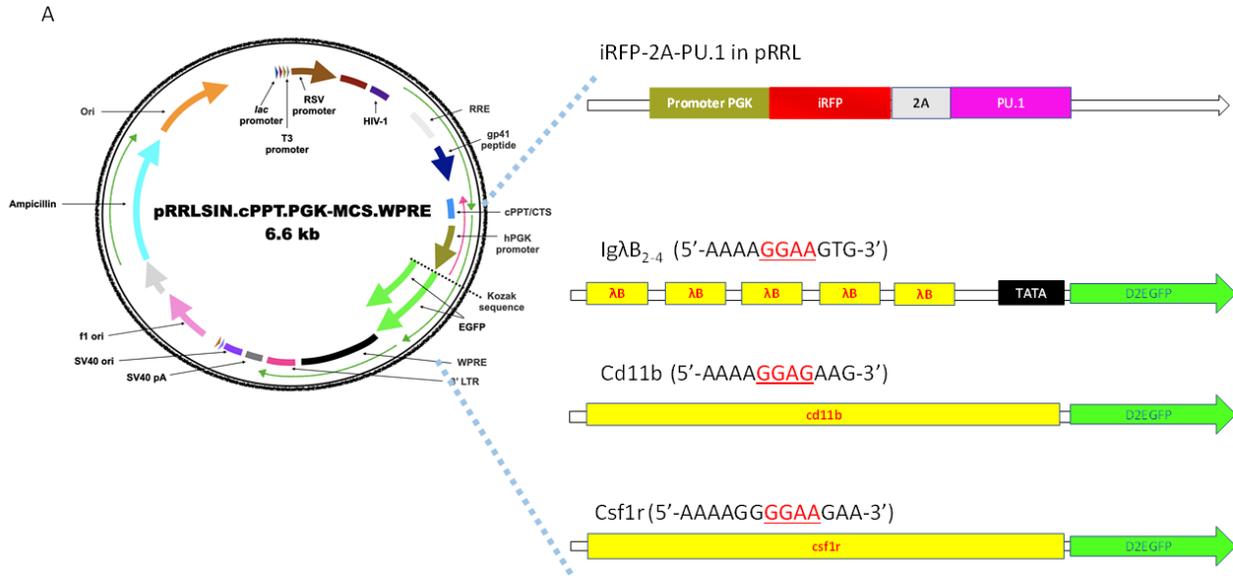
To understand this mystery, we designed structural variants of λ B enhancers for PU.1 transactivation in live cells. The manipulation was designed in terms of enhancer syntax of numbers and space between the λ B enhancers. Because the constrained transcription factor spacing is prevalent and essential for transcriptional control of blood cells in development, we utilized the idea and examined the enhancer structure's effect giving the spacing between its cognate sequence. The enhancers were composed only of tandem EBS spaced 20 bp apart. Therefore, by manipulating enhancer structure such as density and spacing, the PU.1 transactivation in live cells would demonstrate its function concerning the structural changes solely at the DNA as PU.1 dose-dependent manner. From the measuring PU.1 transactivation, if inactive complex 2:1 activity work in a negatively cooperative manner, the level of expression in the graph would illustrate as bell-shape reporter dose-response to the enhancer. If nonproductive 2:1 complex of PU.1 bound and reached the saturated level depending on the level of the 2:1 complex retaining the activity relative to the 1:1 complex as PU.1 dose increased.

2.3.6 Development and confirmation of PU.1-dependent reporters for functional cell reporter assay

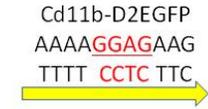
2.3.6.1 Sequence confirmation of the LV-reporter constructs

The special DNA recognition of PU.1 led us to develop a cell-based reporter platform to screen ligands against PU.1. PU.1 preferentially recognizes the AT-rich sequence additionally 5'-flanking region harboring the core cognate 5'-GGA/G-3' sequences. The AT-preference were brought the idea to modulate PU.1 activity using small molecules pharmacologically. Our

previous research attested the hypothesis to functional PU.1 inhibition via small DNA minor groove binders targeting AT-rich sequences, and the compounds' binding allosterically interfered with DNA recognition by PU.1. To screen a large library set, we designed our PU.1-dependent genes fused with an EGFP reporter for an efficient lentivirus delivery vector (Figure 2.11).



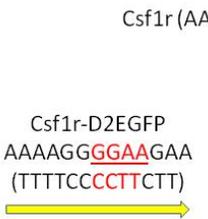
C



Reverse complementary

cd11b	-----GCAGATCTATTTTTTTGTAGAGACAGGG	29
cd11b_pEGFP	CCATGCATTAGTTATTACTACGCTACCGACTCAGATCTATTTTTTTGTAGAGACAGGG	780
cd11b_pRRL_EGFP	-----GCAGATCTATTTTTTTGTAGAGACAGGG	29
cd11b	TCTCTATGTGCCAGGCTGGTTCAAACCTCCAGGCTCAAGCAATCTCTGCCTTG	89
cd11b_pEGFP	TCTCTATGTGCCAGGCTGGTTCAAACCTCCAGGCTCAAGCAATCTCTGCCTTG	840
cd11b_pRRL_EGFP	TCTCTATGTGCCAGGCTGGTTCAAACCTCCAGGCTCAAGCAATCTCTGCCTTG	89
cd11b	GCCTCCCAAAGTGCTGGCATTACAGGCTGAGCCACTGCGCTGGCCGTATAATGTTT	149
cd11b_pEGFP	GCCTCCCAAAGTGCTGGCATTACAGGCTGAGCCACTGCGCTGGCCGTATAATGTTT	900
cd11b_pRRL_EGFP	GCCTCCCAAAGTGCTGGCATTACAGGCTGAGCCACTGCGCTGGCCGTATAATGTTT	149
cd11b	AGAACACGAATCCAGGAGGCAAGCTAAGTCTATTCAAGCTTGTTCATAGCTTGGCCAA	209
cd11b_pEGFP	AGAACACGAATCCAGGAGGCAAGCTAAGTCTATTCAAGCTTGTTCATAGCTTGGCCAA	960
cd11b_pRRL_EGFP	AGAACACGAATCCAGGAGGCAAGCTAAGTCTATTCAAGCTTGTTCATAGCTTGGCCAA	209
cd11b	CCCAAGAAACAAGTGGGTGACAAATGGCACCTTTGGATAGTGGTATTGACTTTGAAAGT	269
cd11b_pEGFP	CCCAAGAAACAAGTGGGTGACAAATGGCACCTTTGGATAGTGGTATTGACTTTGAAAGT	1020
cd11b_pRRL_EGFP	CCCAAGAAACAAGTGGGTGACAAATGGCACCTTTGGATAGTGGTATTGACTTTGAAAGT	269
cd11b	TTGGGTCAGGAAGCTGGGGAGGAAGGGTGGGAGGCTGTGGGCACTCTGGCGGAAGAC	329
cd11b_pEGFP	TTGGGTCAGGAAGCTGGGGAGGAAGGGTGGGAGGCTGTGGGCACTCTGGCGGAAGAC	1080
cd11b_pRRL_EGFP	TTGGGTCAGGAAGCTGGGGAGGAAGGGTGGGAGGCTGTGGGCACTCTGGCGGAAGAC	329
cd11b	CAGGCAGGGCTATGTGCTCACTGAGCTCCGCCCTCTTCTTTGAATCTCGATAGACTT	389
cd11b_pEGFP	CAGGCAGGGCTATGTGCTCACTGAGCTCCGCCCTCTTCTTTGAATCTCGATAGACTT	1140
cd11b_pRRL_EGFP	CAGGCAGGGCTATGTGCTCACTGAGCTCCGCCCTCTTCTTTGAATCTCGATAGACTT	389
cd11b	CTGCCTCTACTTCTCCTTTCTGCCTTCTTTGCTTTGACCGGTAT-----	437
cd11b_pEGFP	CTGCCTCTACTTCTCCTTTCTGCCTTCTTTGCTTTGACCGGTAT-----	1200
cd11b_pRRL_EGFP	CTGCCTCTACTTCTCCTTTCTGCCTTCTTTGCTTTGACCGGTAT-----	437

D



csf1r	GGAAGTGGCAGGCAAGTGCAGGGCTGCAAGGTGCCAGGGCACAGAGCTCTCAGCTACTA	77
csf1r-pEGFP	GGAAGTGGCAGGCAAGTGCAGGGCTGCAAGGTGCCAGGGCACAGAGCTCTCAGCTACTA	119
csf1r-pRRL_EGFP	GGAAGTGGCAGGCAAGTGCAGGGCTGCAAGGTGCCAGGGCACAGAGCTCTCAGCTACTA	120
csf1r	GCTCCGACAGGATCGGGCACTGGACACAGTTCTCTCTCTGCACCTGGCTGTTGTCT	137
csf1r-pEGFP	GCTCCGACAGGATCGGGCACTGGACACAGTTCTCTCTCTGCACCTGGCTGTTGTCT	179
csf1r-pRRL_EGFP	GCTCCGACAGGATCGGGCACTGGACACAGTTCTCTCTCTGCACCTGGCTGTTGTCT	180
csf1r	TGTTTTCTCTCTCTCTCTCTCTGGCTGATCCTCTCTCCCTTTTGGTAAAGCAAGGC	197
csf1r-pEGFP	TGTTTTCTCTCTCTCTCTCTCTGGCTGATCCTCTCTCCCTTTTGGTAAAGCAAGGC	239
csf1r-pRRL_EGFP	TGTTTTCTCTCTCTCTCTCTCTGGCTGATCCTCTCTCCCTTTTGGTAAAGCAAGGC	240
csf1r	AACCACAGAGTTTGGAAATCTTGGGTCTTTAAGAAAGAAACAAAGCCTTGAACCAAGAG	257
csf1r-pEGFP	AACCACAGAGTTTGGAAATCTTGGGTCTTTAAGAAAGAAACAAAGCCTTGAACCAAGAG	299
csf1r-pRRL_EGFP	AACCACAGAGTTTGGAAATCTTGGGTCTTTAAGAAAGAAACAAAGCCTTGAACCAAGAG	300
csf1r	GGAGGGTTCGACGCTGGGGCAGGGCCAGCTGAGCTTGGGGGAGCTAGCTAAGTTTGG	317
csf1r-pEGFP	GGAGGGTTCGACGCTGGGGCAGGGCCAGCTGAGCTTGGGGGAGCTAGCTAAGTTTGG	359
csf1r-pRRL_EGFP	GGAGGGTTCGACGCTGGGGCAGGGCCAGCTGAGCTTGGGGGAGCTAGCTAAGTTTGG	360
csf1r	GGCCCTTCAAAGCCATCTTCAAGGGTCTTGGGAGAGAGACATGTCCAAGAGCAGCCA	377
csf1r-pEGFP	GGCCCTTCAAAGCCATCTTCAAGGGTCTTGGGAGAGAGACATGTCCAAGAGCAGCCA	419
csf1r-pRRL_EGFP	GGCCCTTCAAAGCCATCTTCAAGGGTCTTGGGAGAGAGACATGTCCAAGAGCAGCCA	420
csf1r	GACCGTTAGTCAGTACCAAAATGAGGGTCTTCAGAGTATCAGGTGGCTTGGCAGATC	437
csf1r-pEGFP	GACCGTTAGTCAGTACCAAAATGAGGGTCTTCAGAGTATCAGGTGGCTTGGCAGATC	479
csf1r-pRRL_EGFP	GACCGTTAGTCAGTACCAAAATGAGGGTCTTCAGAGTATCAGGTGGCTTGGCAGATC	480
csf1r	CCAAAGAGAAGTTAGGTTGCATGATATTGTGACATATATATTTTTTCTTACCTGATGT	497
csf1r-pEGFP	CCAAAGAGAAGTTAGGTTGCATGATATTGTGACATATATATTTTTTCTTACCTGATGT	539
csf1r-pRRL_EGFP	CCAAAGAGAAGTTAGGTTGCATGATATTGTGACATATATATTTTTTCTTACCTGATGT	540
csf1r	CCTGGCTTACAACCTCTAAAATCTTGAATCTCCAAGTAAATGCATATCTTTAAAAAT	557
csf1r-pEGFP	CCTGGCTTACAACCTCTAAAATCTTGAATCTCCAAGTAAATGCATATCTTTAAAAAT	599
csf1r-pRRL_EGFP	CCTGGCTTACAACCTCTAAAATCTTGAATCTCCAAGTAAATGCATATCTTTAAAAAT	600

Figure 2.11 Schematic representation of the reporter gene constructs for characterizing ligands against PU.1 activity.

The 3rd generation of the lentiviral system is used for efficiently delivery of PU.1-dependent gene reporters into endogenous PU.1 cells to detect PU.1 activity modulated by ligands that target and bind the AT-rich sequences. The transfer vector has a carrying capacity of ~ 8 kb, which is sufficient for most applications. (A) A schematic diagram of the pRRL-iRFP-2A-PU.1 and the enhancer element reporter vector construct. The lentiviral expression vector of PU.1 was developed and used for confirming the PU.1-dependent gene reporter construct in HEK293 cells. A vector map of each PU.1-specific responsive cis-acting element pRRL-λBx5-D2EGFP, pRRL-Cd11b-D2EGFP, and pRRL-Csf1r-D2EGFP. Each minimal sequence is sufficient for PU.1 transactivation, and EGFP activity of each vector is detected when PU.1 binds to its cognate sequence. Each PU.1 specific enhancer element was inserted into the MCS site between restriction enzyme site AgeI and BglII. The original hPGK promoter of the vector was replaced with a minimal promoter/enhancer so that they are necessary for PU.1 binding to express EGFP reporter. Each minimal element was previously cloned in the pd2EGF-1 vector and transferred into the lentiviral transfer vector. (B) Sequence confirmation of synthetic PU.1-dependent EGFP reporters for the lentiviral system. pRRL-λBx5-D2EGFP was cloned, and the sequence of each synthetic enhancer was confirmed by Sanger sequencing. The five copies of a tandem sequence harboring a cognate sequence are bound by PU.1 and express EGFP marker. (C) Sequence confirmation of synthetic PU.1-dependent EGFP reporters for the lentiviral system. pRRL-cd11b-D2EGFP was cloned, and the sequence was compared with the pCd11b-d2EGFP. The comparison sequence was used for the available 5' primer of EGFP NR and sequenced in a reverse complementary manner. The red box is the cd11b promoter harboring core consensus sequence (5'-GGAG-3') bound by PU.1. The myeloid-specific CD11b promoter binds PU.1, and the sequence 'AAAGAGGAAG' at base-pair -20 of the CD11b promoter has been known to sufficient to direct the expression of a reporter gene EGFP under PU.1 binds to the site. The myeloid integrin CD11b is selectively expressed on the surface of mature monocytes, macrophages, granulocytes, and natural killer cells and regulated its expression by PU.1. (D) Sequence confirmation of synthetic PU.1-dependent EGFP reporters for the lentiviral system. The tissue-specific expressed macrophage colony-stimulating factor (M-CSF) receptor is directly controlled by PU.1 and regulates a key macrophage growth factor. The M-CSF receptor promoter is located just upstream from the major transcription initiation. The pRRL-CSF1R-D2EGFP was cloned, and the sequence was compared with the pCSF1R-d2EGFP. The red box in the sequence denotes the 5'-flanking sequence near a core cognate binding sequence (5'-GGAA-3'). The comparison sequence was used for the available 5' primer of EGFP NR and sequenced in a reverse complementary manner. The yellow arrow indicates the direction of the sequence reading.

The lentiviral expression vectors of PU.1, λB, Cd11b, and Csf1r were fused with the reporter (Figure 2.11A). The promoter of hPGK was used for expressing iRFP-2A-PU.1. The PU.1-specific gene reporters were replaced with hPGK promoter so that EGFP reporter expression was controlled by PU.1 binding (Figure 2.11A). Coupled with the availability of cell

lines expressing PU.1, the reporters can be used to identify and measure the activation of specific enhancers upon PU.1 binding and its reporter in highly sensitive, non-radioactive, and easy-to-perform functional assays. The relevant disease model THP-1 expresses endogenous PU.1 in this research.

Further examination on monitoring native PU.1 activity in the AML disease-relevant model, THP-1 cells, we developed the “gen2” reporter assay. Under native PU.1 activity, we inserted three PU.1-dependent gene constructs (λ Bx5, Csf1r, and Cd11b) into the genomic DNA of THP-1. These genes contain homologs of 5'-GGAA/G-3' of PU.1 binding sites tagged fluorescent protein EGFP. This design aimed to monitor PU.1 transactivation on their target specific gene only by interrupting any other signaling in native PU.1 expressing cells, THP-1. The λ B motif of the Ig2-4 enhancer was previously known to a high-affinity PU.1 binding site with AT-rich tracks flanking both sides of the ETS consensus and used for our cell reporter assay to evaluate the AT sequence-specific binding compounds to PU.1 inhibition. Csf1r (receptor for macrophage colony-stimulating factor) was upregulated by PU.1 in the monocytic development and implicated with leukemic stem cells. According to Aikawa et al., cells expressing high amounts of CSF1R showed potent leukemia-initializing activity in the MOZ-TIF2 interacting with PU.1 and were known to target PU.1-mediated upregulation of CSF1R expression might be a practical therapeutic approach. Each sequence of the construct for the lentiviral expression vector was verified with Sanger sequencing (Figure 2.11 B, C, and D).

2.3.6.2 *Flow cytometric confirmation of LV-reporter constructs*

After the production and collection of each virus particle containing PU.1-dependent gene reporter of λ Bx5, Csf1r, and Cd11b, we confirmed the lentiviral particles were not

expressed under non-expressing PU.1 cell lines, HEK293 cells and 293T cells compared with the positive control pgk-EGFP constructs (Figure 2.12).

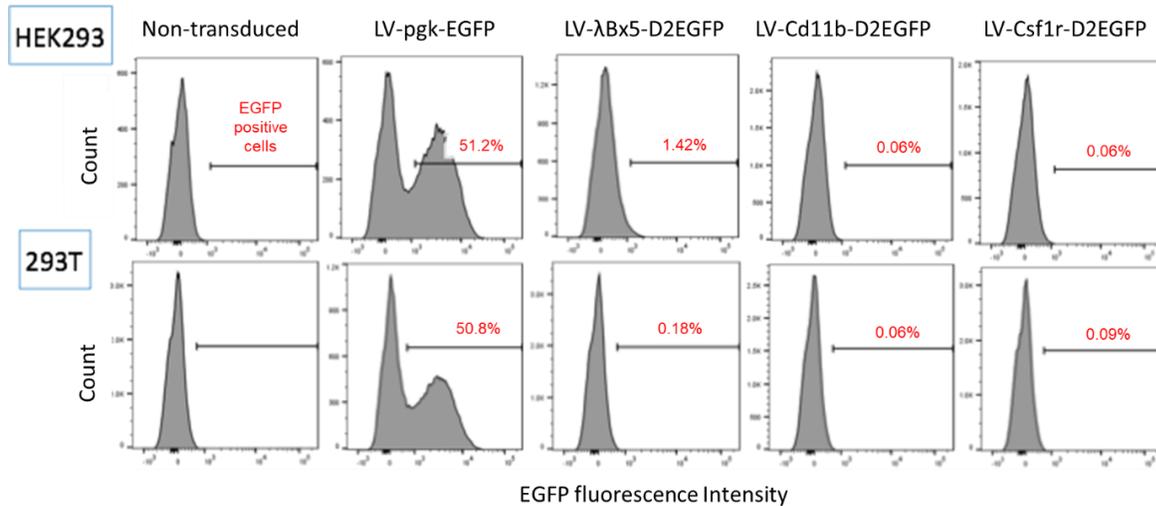


Figure 2.12 Flow cytometric data to confirm the lentiviral particles from the collection. Non-PU.1 expressing cell lines, HEK293, and 293T cells, were not shown its EGFP expression by lentivirus transduction of each PU.1 dependent reporter gene compared with the EGFP expression hPGK-EGFP (human phosphoglycerate kinase gene promoter tagged EGFP).

1 μ l of concentrated particles were inoculated onto 5×10^4 cells per well, and after 4 days of the transduction, no detectable EGFP expression exhibited, indicating that the constructs other *trans*-acting molecules in the HEK293 and 293 T cells could not transactivate the constructs.

Along with the validation procedure, we verified the EGFP expression by the lentiviral reporter system under the PU.1 existence (Figure 2.12 and Figure 2.13). We examined virus particles with either chemically transfected piRFP-2A-PU.1 or transduced with pRRL-iRFP-2A-PU.1 in HEK293 cells; then, we transduced the lentivirus particles of PU.1-specific genes. Compared with control HEK293 cells, the iRFP and EGFP positive cells (Q2) were detected of LV-λBx5-D2EGFP (Figure 2.12). More than 65% of positively iRFP expressing cells, about

23% of EGFP reporter expression among the iRFP marker expressing cells were detected. We detected residual detection of EGFP expression on the Q1, suggesting that solid promoter CMV might overreact on the lentiviral $\lambda B \times 5$ particles.

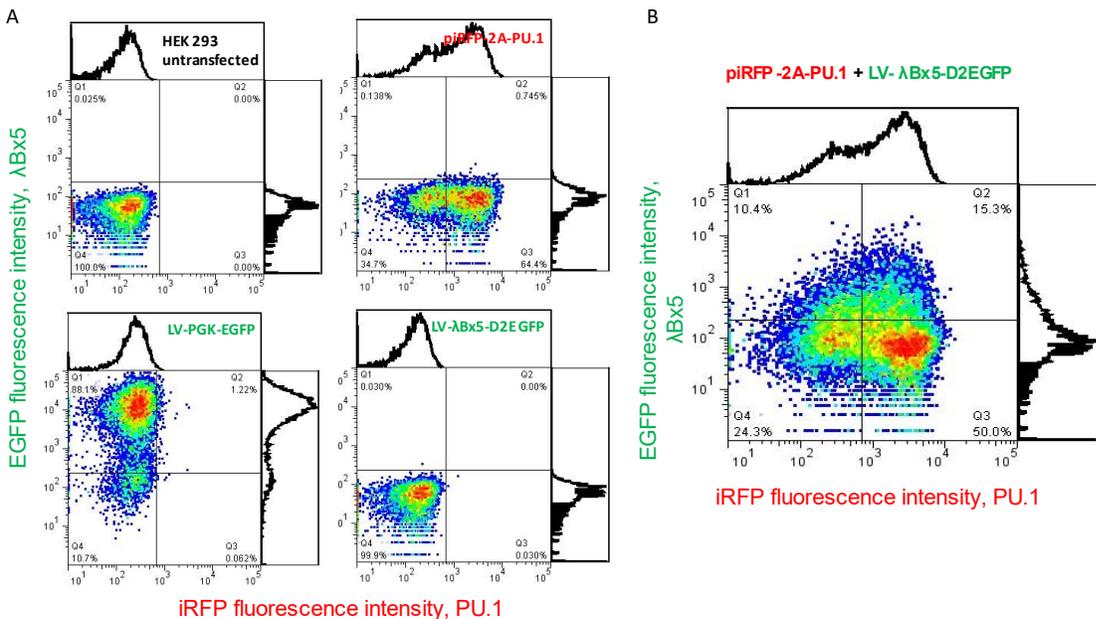


Figure 2.13 PU.1 transactivation on its specific gene confirmed by lentiviral particles, LV- $\lambda B \times 5$ -D2-EGFP.

(a) Flow cytometric analysis of PU.1 marker (iRFP) and EGFP reporter. (b) The upper right quadrant (Q2) represents PU.1 dependent reporter ($\lambda B \times 5$) activation.

LV-Cd11b-D2EGFP was verified with its EGFP expression by PU.1 expressing HEK293 cells (Q2) in Figure 2.13A. The LV- $\lambda B \times 5$, Csf1r, exhibited EGFP expression under the LV-iRPF-2A-PU.1 transduction as well (Figure 2.13B).

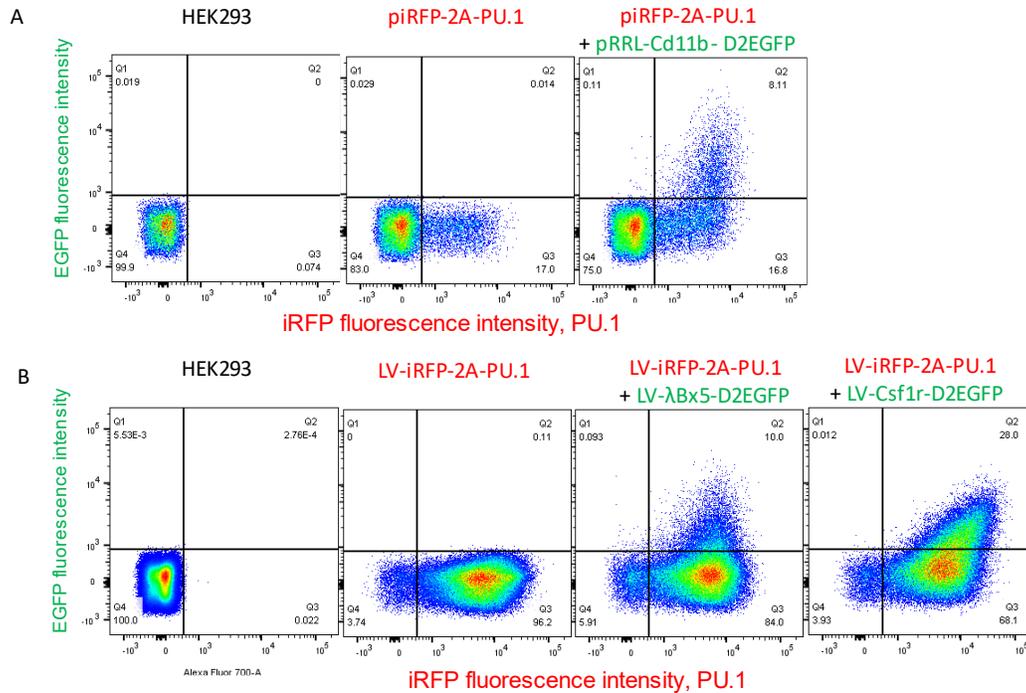


Figure 2.14 PU.1 transactivation on its reporter gene confirmed by lentiviral particles, LV-Cd11b-D2-EGFP and LV-Csf1r-D2EGFP.

(a) Flow cytometric analysis of pRRL-Cd11b-D2EGFP reporter. After 72 h of transduction with the LV-Cd11b-D2EGFP particles to HEK293 cells, the HEK293 cells were transfected with piRFP-2A-PU.1. Q2 represents cells that EGFP positive and iRFP positive, indicating that the Cd11b-D2EGFP particles were functionally responded to PU.1. (b) The upper right quadrant (Q2) represents PU.1 dependent reporter (λ Bx5) activation. PU.1 expressed HEK293 cells transduced by LV- λ Bx5-D2-EGFP and LV-Csf1r-D2-EGFP, and all lentiviral constructs were functionally working as we expected.

After confirming the lentiviral particles of the PU.1-dependent enhancer elements' EGFP expression by exogenously expressing PU.1 in HEK293, we developed stably expressing reporter cell lines in THP-1 (Figure 2.14). Three *cis*-acting elements were successfully transduced and selected only EGFP expressing cells for characterizing cellular PU.1 inhibition by small DNA minor groove binders.

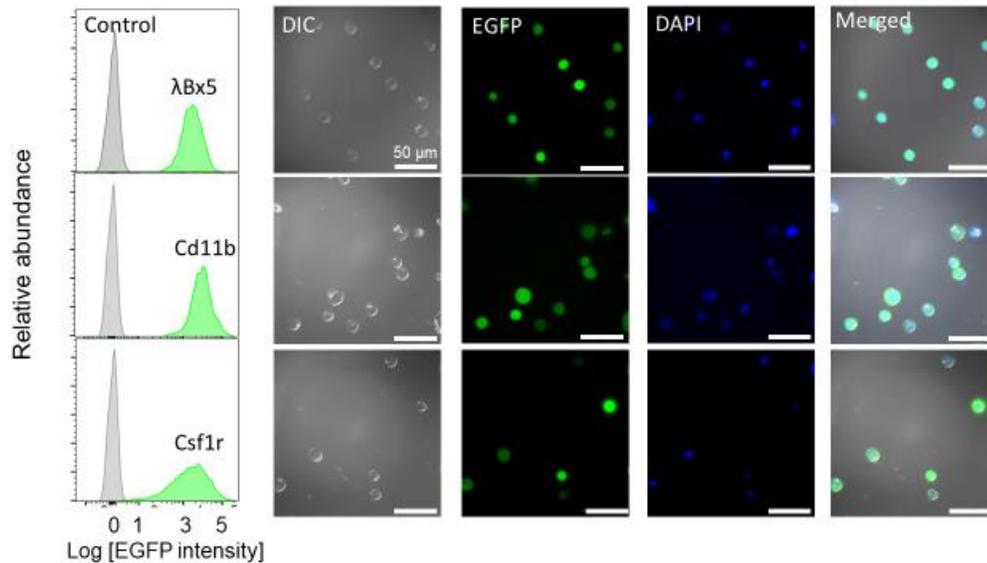


Figure 2.15 The Stable cell line confirmation of the second-generation PU.1-dependent reporter in THP-1 cell line.

Flow cytometric data represent only THP-1 cells expressing EGFP based upon endogenous PU.1 transactivated were selected and developed as stably constructed. Fluorescence images were confirmed the EGFP were colocalized with DAPI, indicative of nuclear stain.

2.4 Summary

2.4.1 Advantages of our reporter assay to study PU.1-DNA interaction

Given these limitations of ChIP and EMSA to study TF-DNA interaction, the biological reporter technology has significant advantages. The reporter gene technology is widely used to monitor the cellular events associated with signal transduction and gene expression. Based on working of transcriptional control elements to various reporter genes with easily measurable phenotypes, it documents the effects of a cascade of signaling events on gene expression inside cells. The main advantage of these assays is their high sensitivity, reliability, convenience, and adaptability to large-scale measurements. Gene reporter technology here we used EGFP and

iRFP reporter is a popular method for non-invasive monitoring of gene expression in living cells. Such methods are essential in defining the molecular events associated with gene transcription to provide a basis for understanding the molecular basis of disease.

A cell-based reporter system has been widely evolving for various study outcomes. Multiple applications have been identified for studying transcriptional activity, biological imaging systems, and molecular signaling pathways. For these aims, especially characterization of the ETS transcription factor PU.1's activity, the reporter's choice has been developed so that the reporter system's versatility enables adaptation of the study outcomes. The compatibility of reporter genes provides a great advantage for multimodality of study of functional activity of PU.1 related to its specific multiple genes of interest, λ B, Cd11b, and Csf1r and their consequences of regulation in gene expression by utilizing the choice of variant reporters. Further, the non-invasive reporter genes offer the dynamics of PU.1 transactivation and monitor its activity by visualization in live cells. The assay involves the expression of fluorescence protein of interest in target cells so that it is preferentially retained within the cells. Such benefits have attractive for researching PU.1's activity at the protein-DNA level.

In this study, the cell-based reporter platform has been adapted for supporting evidence of biophysical characterization of PU.1 in a biologically relevant context. We developed two expression plasmid vectors for studying PU.1 transactivation at the TF-DNA level. *Trans*-acting elements PU.1 were fused with an iRFP marker under common promoter CMV used for mammalian cells and connected with a self-cleavage 2A peptide for the two proteins to function in the system properly. By monitoring iRFP expression, we tracked the PU.1' expression and monitor its activity by flow cytometry. Another *cis*-acting element for the PU.1 cognate site, Ig2-4 λ B, was fused with an EGFP reporter for monitoring PU.1 transactivation on PU.1's binding

sites. The level of EGFP expression under the control of functional PU.1 activity was detected and quantified. With the two reporter constructs, we additionally designed dominant-negative mutants for studying evolutionarily significant residue of PU.1 ETS motif to identify how the mutants influence the PU.1' activity at TF-DNA level.

Further, by designing the variants enhancer constructs in terms of the copy numbers of a λ B motif and spacing, we supported biophysical characteristics of autoinhibitory regulation of PU.1 in a negative feedback manner in a biological context in live cells. Ultimately, with successful characterizing special features of the DNA recognition by PU.1 of the transcription factor from a group of ETS family of transcription factors, the reporter assay can be further utilized to screen more efficient ligands from a more extensive library of DNA minor groove binding agents for the therapeutic applications. In this chapter, we presented how each reporter expression vector of the cell-based platform assay was designed and constructed for aiming with highly sensitive, specific, rapid, and can be performed in a single-step manner without the need for any expensive equipment (131). The following two chapters will present how these reporter gene platforms were employed and the significant findings from the studies.

3 CHARACTERIZATION OF PU.1 ACTIVITY IN LIVE CELLS

3.1 Introduction

3.1.1 Functional characterization of transcription factor attributes in living cells

To examine the key attributes suggested by biophysical studies in the context of biology, all the biophysical parameters tested in the experimental system cannot align with the same conditions in the biological system. In biology, it is impossible to define all parameters affecting a living system, such as a cell; instead, the goal can be said to build models and hypotheses for the function. Through experiments, collect supporting data. In this research for functional characterization of PU.1 activity in live cells, we defined the behavior of PU.1 in dilute solution, which was used for the biophysical models as same as the PU. 1's behavior in living cells. Due to the experimental conditions for preserving transcription factor protein folding and DNA, the biophysical and biochemical research was performed in a diluted solution in which actual biological experiments in live cells differ. Here we supposed that the transcription factor's behaviors of the biophysical context (measured in diluted solution) are consistent with its behaviors in the living cells under the assumption. The conditions that remain under the macromolecular crowding and confinement model are averagely constant. No external environmental changes happen; therefore, testing our biophysical and biochemical hypothesis in living cells lies on the assumption. We inferred the macromolecular bulk effects in the living cells are constant without any external effects as the experimental biophysical conditions in dilution solution even though the intracellular activities to maintain homeostasis against external stimuli or changes constantly exist. Thus, although it primarily behaves similarly and adopts some changes, a cell has the tendency or driving force of getting back to its initial statement, homeostatic mechanisms (131).

3.1.2 The transcription factor PU.1 as a model system for reporter characterization

Transcription factor PU.1, a member of the ETS family transcription factor, is a lineage-specific and dosage-specific regulator in the differentiation of immune cells and decides the cells' fate. PU.1 and ETS family members recognize and bind to cognate sequences harboring a central 5'-GGAA/G-3' consensus. With a combination of transcriptional controls, the intrinsic selectivity of the ETS domains intrinsic selectivity has been studied, and the inherent selectivity of the ETS transcription factors is biologically relevant. Physical determinants of the ETS protein for the intrinsic selectivity and specificity at the genomic levels are essential to understanding higher-order protein-protein and protein-DNA interactions found in the transcription regulatory complex. The knowledge of the regulatory mechanism of transcription factor PU.1 activity in the biological system remains unclear; however, we found two crucial biophysical attributes of PU.1 activity: down-regulation of PU.1 in cells and unique hydration properties PU.1. These two properties are biophysically and biochemically studies in our interests.

3.1.3 Experimental configurations for functional characterization of PU.1 activity using reporters

Here we utilized the strength of the cell-based reporter assay to characterize PU.1 transactivation activity at the protein/DNA level. Previous research demonstrated two critical features of transcription factor PU.1-DNA binding recognition: the hydration mediation role of sensitivity and self-regulated by negative feedback upon the 2:1 formation of PU.1-DNA complex.

3.1.3.1 *Dominant-negative inhibition by ETS domain mutants*

In addition to studying the macro-level structure of a promoter, studies of the single mutation have received attention because they can have profound biological significance. To investigate the functional importance of the single residue of transcription regulatory factor PU.1 in the phylogenetic family among its ETS paralogs, we introduced our cell-based reporter gene assay. Functional inactivation of genes by dominant-negative mutations was used to assign a function to genes that have been cloned by manipulation of the cloned gene to create what is known as dominant-negative mutations. The mutant that, when overexpressed, disrupt the activity of the wild-type gene. The gene function is blocked at the protein level: a mutant gene product is created that inhibits the cell's wild-type products and thus causes the cell to be deficient in that function. A mutation with the effects is termed dominant due to its phenotype exhibited in the presence of the wild-type gene. Because it inactivates wild-type gene function, it is termed dominant negative (216-218).

We designed an ETS mutant whose gene product adversely affects the standard, wild-type gene product within the same cells. The mutants usually occur if the product can still interact with the same elements as the wild type of product but block some aspects of its functions. For instance, a mutation in a transcription factor removes that the activation domain but still contains the DNA binding domain. The ETS mutant product can then block the wild-type transcription factor, full-length PU.1, from binding the DNA site leading to decreased levels of gene activation. By doing so, we investigate the significance of the H3 recognition helix in the ETS motif. H3 is the most important structural element to show a sensitivity of osmotic pressure in the PU.1 ETS domain when swapping with the Ets-1 ETS domain. In addition to that, this domain is the most conserved element in the ETS domain. Because residue 236 of H3

recognition in the ETS domain is highly varied among the ETS paralogs, we expected that the pointed mutants work as an inhibitor of PU.1 transactivation and measure the reduction of PU.1 binding activity on PU.1-dependent gene fused to reporter protein expression. Therefore, our functional PU.1 reporter assay provides quantitative and functional information within a short period permitting them one of the most relevant and vital assays for evaluating evolutionarily significant residue. Further, the functional assay using a PU.1-dependent EGFP reporter confirmed the structural prediction from biophysical experiments and Molecular Dynamic (MD) simulation (215). The use of dominant-negative mutation made us possible to design inhibitory products based on the principle that the activities of a protein can be separately mutated.

3.1.3.2 Probing PU.1 self-regulation through enhancer structure

Detection of effect from enhancer syntax (number of genes and spacing) is crucial for transcription factors in development (60). Transcription factors, especially master regulator PU.1 that play a significant role in the differentiation of the immune cells, are susceptible to such detection for its regulation. Furthermore, transcription factor PU.1 is also known for its inducibility depending on the specific stages of the cells, including macrophages. Consequently, it is critical to understand how the PU.1 transactivation activity positively or negatively interacts with its binding DNA. Our reporter gene assay employing PU.1-dependent specific enhancers and suitable markers enables us to achieve the goal.

We manipulated the design of enhancer structure in the principle of syntax spacing and number in the reporter constructs and observed the functional PU.1 inhibition by measuring the reporter activities. The PU.1 transactivation in live cells would demonstrate its function concerning the structural changes solely at the DNA as PU.1 dose-dependent manner. From the

measuring PU.1 transactivation, the functional outcome of inactive complex 2:1 activity would represent as bell-shape reporter dose-response as the enhancer if PU.1 inactive complex works negatively cooperative (112, 134). If nonproductive 2:1 complex of PU.1 bound and reached the saturated level depending on the status of the 2:1 complex retaining the activity relative to the 1:1 complex as PU.1 dose increased (219). Therefore, our reporter gene platform to assess the effect of the enhancer syntax enabled us to characterize functional PU.1 activity. This method proved to us that the idea of PU.1 self-association is biologically compatible.

3.1.4 Objectives of this chapter

Chapter 3 provided understandings of functional PU.1 transactivation in live cells using our reporter platforms to characterize two signatures of its DNA target selectivity that are biophysically determined. We employed the dominant-negative inhibitors of ETS domain mutants to prove the evolutionary significance of a critical residue in PU.1 interacting with its target DNA. Second, to support biophysical evidence of the intrinsic PU.1's autoregulatory function, we utilized the enhancer syntax rule by designing PU.1-dependent enhancer repetition and by the spacing between the gene copies. We probed that the inactive 2:1 complex functionally acts as negative feedback to control PU.1 transactivation in the biological system. Our reporter platform supplied an adequate assay platform to gain insights into valuable PU.1 inhibition data that the two biophysical determinants were functionally feasible and compatibles in live cells.

3.2 Materials and methods

3.2.1 *Plasmid design/Molecular Cloning*

3.2.1.1 *Dominant-negative inhibitors*

In the research of various dominant-negative mutants of PU.1, the variants were described in chapter 2.

3.2.1.2 *Synthetic PU.1-dependent EGFP reporters*

For the study of the effect of enhancer structure on PU.1 transactivation, several PU.1-sensitive enhancer sequences as described in the previous chapter.

3.2.2 *Cell culture*

The HEK293 cells were purchased from American Type Culture Collection (ATCC) and cultured in Dulbecco's Modification of Eagle's Medium (DMEM) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 1% (v/v) Penicillin and Streptomycin (P/S) as the ATCC recommended. The HEK293 cells were maintained in 10 cm culture dishes at 37°C in humid air with 5% CO₂. Before reaching 80%-90% confluence, cells were sub-cultured using 0.25% (v/v) trypsin. The trypsin was neutralized with cell culture media (above). The medium was refreshed every 2-3 days. For characterizing PU.1 transactivation in live cells, we used Human Embryonic Kidney (HEK) 293. HEK293 cells do not natively express PU.1 to monitor the transcription factor activity without any intrusive intercellular component. Therefore, the cells enable us to monitor PU.1 activity non-invasive manner. Minimal eukaryotic transcription machinery in HEK293 cells was employed to express the functional gene products.

3.2.3 Transfection assay

3.2.3.1 Transient transfection assay for dominant-negative mutants

To identify the effect of cognate DNA binding by dominant-negative mutants of PU.1 ETS motif, PU.1 transactivation in live cells was determined using a PU.1 dependent EGFP reporter construct as previously described. In brief, the EGFP reporter is under the control of a minimal PU.1-dependent enhancer harboring a 5 × tandem of a native cognate site for PU.1. Furthermore, in non-PU.1 expressing HEK293 cells, the reporter was transactivated in the expression plasmid encoding WT full-length PU.1 and a co-translating iRFP marker.

7×10^4 HEK293 cells were seeded in 24-well plates. After 12-18 hours of the cell were seeded and incubated at 37°C 5% CO₂ incubator for 12-16 hours before the transfection assay. According to the manufacturer's instructions, transfection was performed using JetPrime reagent (Polyplus, Illich, France). The seeded cells were co-transfected with a mixture consisting of the EGFP reporter plasmid (300 ng) and expression plasmids for full-length PU.1 (50 ng) and WT/mutant ETS domains (150 ng). One or both expression plasmids were replaced by empty pcDNA3.1(+) vector in negative control samples (total 500 ng of DNA per well). The transfected HEK293 cells were incubated at 37°C 5% CO₂ incubator. After eighteen hours post-transfection, the cells were confirmed for PU.1 expression and EGFP expression under the fluorescence microscope (Leica DMI 800) at first. Then, the cells were trypsinized and analyzed by flow cytometry, an LSRFortessa™ instrument (BD) as recommended by the manufacture's instruction. Data were recorded in BD FACSDiva™ software. The population of interest (live cells) was gated for iRFP and EGFP fluorescence using reporter- and full-length PU.1-only

controls, respectively. The raw data were exported to FlowJo (BD) before computing the total fluorescence of the dually fluorescent population.

3.2.3.2 Transient transfection assay for synthetic PU.1-dependent EGFP reporters

For investigating the effect of enhancer spacing and structure, we designed various reporter gene constructs. As previously mentioned, the PU.1-dependent EGFP reporter constructs were only expressed under the control of a minimal enhancer harboring only cognate binding sites for PU.1; therefore, the reporter gene expression levels were measured as PU.1 transactivation. To non-invasively monitor, two genes were co-transfected into HEK293 cells. The reporter gene was transactivated in the presence of an expression plasmid encoding wild-type full-length PU.1 and a co-translating iRFP marker.

The HEK293 cells (7×10^4) were seeded in 24-well plates about 12-16 hours before transfection was performed. The seeded cells were co-transfected with a mixture comprising the EGFP reporter plasmid (250 ng) and up to 25 ng of full-length PU.1 expression plasmid using jetPRIME reagent (Polyplus, Illkirch, France) according to the manufacturer's instructions. Briefly, the total amount of plasmid was made up to 500 ng with empty pcDNA3.1(+) vector and added into the seeded cells in each well. 24 hours post the transfection, the cells were confirmed the expression by the fluorescence microscope (Leica, DMI800). Then, cells were trypsinized and analyzed by flow cytometer, FCS LSRFortessa instrument (BD Biosciences). The recorded data exported from the BD FACSDiva software (BD) and analyzed by Flowjo (BD) software. Live cells were gated for iRFP and EGFP fluorescence using reporter and full-length PU.1 expressing cells only as controls, respectively, in Flowjo (BD) before computing the total fluorescence of the dually fluorescent population.

3.2.4 Immunoblotting analysis and qPCR

3.2.4.1 Western blot for dominant-negative mutants

For determining the expression of the transfected genes in HEK293 cells, equivalent amounts of whole-cell extract (20 μ g) were prepared using RIPA buffer to lyse the transfected HEK293 cells after 24 h post-transfection. Each lysate was separated by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). After electrophoresis, proteins were transferred to the PVDF membrane (Invitrogen) in the transfer buffer containing 30mM Tris, 200 mM glycine, and 20% methanol for four h (Please check the buffer composition, check the running conditions). Subsequently, the membrane was blocked by incubation in TBST (0.2M Tris, 1.37M NaCl, 0.1% tween-20, pH 7.6) containing 5% Bovine Serum Albumin (BSA) for 1 hour and then probed with a polyclonal anti-PU.1 antibody (GeneScript A01692) with GAPDH as loading control in 5% BSA-TBST at a dilution of 1:1000. These incubations were done at 4C overnight. Then, the next day after three 10-min washes with TBST and were probed with HRP-conjugated secondary anti-rabbit antibody (GE Healthcare #AS014) at a dilution of 1:5000. The bounded antibody was identified by enhanced chemiluminescence (Fisher Scientific) and then exposed to GE image Quant LAS-4000 analyzer as recommended by the manufacturer. Images were taken from the analyzer software, and the band intensity on the western blots, expression of the ETS domains was quantified with ImageJ software.

3.2.4.2 Western blot for the effect of enhancer structure

For quantifying the expression of the transfected PU.1 gene in HEK293 cells, equivalent amounts of whole-cell extract (30 μ g) were prepared using RIPA buffer to lyse the transfected HEK293 cells after 24 h post-transfection. Each lysate was separated by 10% sodium dodecyl

sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). After electrophoresis, proteins were transferred to PVDF membrane (Invitrogen) in the transfer buffer containing 30mM Tris, 200 mM glycine, and 20% methanol for four h (40V Please check the buffer composition). Subsequently, the membrane was blocked by incubation in TBST (0.2M Tris, 1.37M NaCl, 0.1% tween-20, pH 7.6) containing 5% BSA for 1 hour and then probed with a polyclonal anti-PU.1 antibody (Cell Signaling, rabbit, cat #2266) with beta-actin as loading control in 5% BSA-TBST at a dilution of 1:1000. These incubations were done at 4C overnight. Then, the next day after three 10-min washes with TBST and were probed with HRP-conjugated secondary anti-rabbit antibody (GE Healthcare #AS014) at a dilution of 1:100,000 for PU.1. For loading control detection for beta-actin, HRP-conjugated secondary anti-mouse IgG (GE Healthcare #AS003) was used at a dilution of 1:2000. The bounded antibody was exposed by enhanced chemiluminescence (Fisher Scientific) and then exposed to Amersham Imager 600 as recommended by the manufacturer. Images were taken from the analyzer software, and the band intensity on the western blots, expression of the ETS domains was quantified with ImageJ software.

3.2.4.3 *QPCR for PU.1 dosage calibration*

Following extraction of total RNA using a spin column kit (Omega) and RT (Thermo Fisher Scientific), RT-PCR reactions were performed on a QuantStudio 3 instrument (Applied Biosystems) with SYBR Green PCR Master Mix (Thermo Fisher Scientific). Expression levels of each gene was normalized to *gapdh*. The primer sequences used for *pu.1*, *csf1r*, *e2f1*, and *gapdh* are referred to in the figure.

3.3 Results

We have been interested in determinants of PU.1 binding its cognate sequence site with high specificity and affinity and its regulation mechanism. Here we developed two new reporter gene platforms for characterizing transcription factors' activity in live cells to provide critical information that the biophysical properties of the transcription factor PU.1 and support the biophysical attributes of TF's DNA recognitions biologically compatible. First, we developed a cellular reporter platform for dominant-negative mutants of the critical residue in molecular hydration properties of PU.1 to recognize its' specific DNA sites. Second, the other reporter technology is utilized for detecting PU.1's self-regulation. For the aim, we constructed different synthetic PU.1-dependent reporters to determine functional PU.1 activity in live cells. Our previous research proposed that although transcription factor PU.1 contains structural elements for auto-inhibition, the regulation mechanism of the unbounded state of PU.1 might be caused by PU.1 formed 2:1 complex with the cognate sequence. We measured the effect of enhancer structure on PU.1 transactivation using tandem copies of DNA binding sites with 20 bp of spacings; therefore, we proved negative feedback in PU.1 *trans*-regulation in the form of 2:1 complex of PU.1-DNA.

3.3.1 Functional Inhibition of PU.1 by evolutionarily conservative ETS mutants

We determined the functional significance of residue in PU.1 linked to interfacial hydrations effect in perspective of the interaction of PU.1 and its site-specific DNA to form the complex through cell-based reporter assay. The dominant-negative inhibitors of PU.1 were designed from a sequence alignment of murine ETS domains with a view of maintaining the register of residues that are most conserved among ETS paralogs.

3.3.1.1 Rationale for a chimeric approach to studying DNA site recognition by PU.1

In a comparative sequence study, we found this helix recognition is the most conserved element among the ETS domain, except residue 236 (Figure 3.1).

```

sp|Q6P307|112-195|SPIC|Mus_musculus      LRLFEYLFESLCSN-EMVSCIQWVDKARGIFQFISKNKETLAEHLWGQRKGNRPMTYQKMARALRNYARTGEIIRKIR--RKLTYQFS 84
sp|P17433|172-255|SPI1|Mus_musculus      IRLYQFLDLLRSG-DMKDSVWVVDKDKGTFFQSSKHKEALAHRWGIQKGNRKMTYQKMARALRNYAGTGEIRKVK--RKLTYQFS 84
sp|O35906|174-257|SPIB|Mus_musculus      LRLYQFLGLLRRG-DMRECVWVPEGAGVQFSSKHKELLARRWGGQKGNRKMTYQKLARALRNYAKTGEIRKVK--RKLTYQFD 84
sp|Q9WTP3|239-322|SPDEF|Mus_musculus     IHLWQFLKELLKPHSYGRFRIRWLNKEKGIKIED--SAQVARLWGVRR-NRPAMNYDKLSRSIRQYKGGIIRKPDISQRLVYQFV 84
sp|Q69775|288-290|ELF1|Mus_musculus     IYLWFLALLQDKATCPKYIKWTQREKGIKFLVD--SKAVSRLWGGKK-NKPDNNYETMGRALRYVYQGGILAKVE-GQRLVYQFK 83
sp|Q922U4|288-290|ELF4|Mus_musculus     IYLWFLALLQDRMTCPKYIKWTQREKGIKFLVD--SKAVSRLWGGKK-NKPDNNYETMGRALRYVYQGGILAKVE-GQRLVYQFK 83
sp|Q93HC9|288-290|ELF2|Mus_musculus     TYLWFLDLLQDKMTCPRYIKWTQREKGIKFLVD--SKAVSRLWGGKK-NKPDNNYETMGRALRYVYQGGILAKVE-GQRLVYQFK 83
sp|Q8VDK3|161-242|ELF5|Mus_musculus     SHLWFEVVDLDSPEENCGILWEDREQIGFRVVK--SEALAKMGGQKK-NKDRMTYEKLSRALRYVYKGTGLERVD--RRLVYVFG 82
sp|O78273|287-289|EHF|Mus_musculus     THLWFEIRDILLSPKMPGLTKWEDRSEGIKFRFLK--SEVAQLWGGKK-NNSNMTYEKLSRAMRYVYKREILERVD-GRRLVYVFG 82
sp|Q3UPW2|293-375|ELF3|Mus_musculus     THLWFEIRDILLHPELNEGLMKENRHEGVKFRFLR--SEVAQLWGGKK-NNSNMTYEKLSRAMRYVYKREILERVD-GRRLVYVFG 82
sp|P97360|335-416|ETV6|Mus_musculus     RLLWDVYVQLLSDS-RYENFIRWEDKESKIFRIVD--PNGLARLWGNKK-NRTNMTYEKMSRALRHYKLNIRKEP-GQRLVFRFM 82
tr|G17VM5|311-392|ETV7|Oryctolagus_cuniculus  RLLWDVYVQLLSDP-RYEPYIRWEDKDAKIFRIVD--PNGLARLWGNKK-NRTNMTYEKLSRALRHYKLNIRKEP-GQRLVFRFM 82
sp|P41969|5-86|ELK1|Mus_musculus        VTLWQFLQLLREQ-GMGIISWTSRDGGEFKLVLD--AEEVARLWGLRK-NKTNNYDKLSRALRYVYDKNIIRKVI-GQKRVYKVF 82
sp|P41971|5-85|ELK3|Mus_musculus        ITLWQFLHLLLDQ-KHEHLIWTNSD-GEFKLLK--AEEVARLWGLRK-NKTNNYDKLSRALRYVYDKNIIRKVI-GQKRVYKVF 81
sp|P41158|5-85|ELK4|Mus_musculus        ITLWQFLQLLQEP-QNEHMICWTSNN-GEFKLLQ--AEEVARLWGLRK-NKPNMNYDKLSRALRYVYDKNIIRKVI-GQKRVYKVF 81
sp|P78459|27-107|ERF|Mus_musculus       IQLWHFILLELRKE-EYQGVIAWQ-GDYGEFVIKD--PDEVARLWGVRR-CKPQNNYDKLSRALRYVYDKNIIRKVI-GKRFYKFN 81
sp|Q8R424|35-116|ETV3|Mus_musculus      IQLWHFILLELRKE-EFRHVIWQQGEYGEFVIKD--PDEVARLWGVRR-CKPQNNYDKLSRALRYVYDKNIIRKVI-GKRFYKFN 82
tr|A03PP2|40-121|ETV3L|Rattus_norvegicus    IQLWHFILLELRKE-EFRHVIWQQGEYGEFVIKD--PDEVARLWGVRR-CKPQNNYDKLSRALRYVYDKNIIRKVI-GKRFYKFN 82
sp|P41164|335-415|ETV1|Mus_musculus      IQLWQFLVALLDDP-SNHFIAWT-GRGMEFKLIE--PEEVARRWGIQK-NRPAMNYDKLSRALRYVYKGGIMQKVA-GERVYKVF 81
sp|P28322|342-422|ETV4|Mus_musculus      IQLWQFLVALLDDP-TMAHFIAWT-GRGMEFKLIE--PEEVARRWGIQK-NRPAMNYDKLSRALRYVYKGGIMQKVA-GERVYKVF 81
sp|Q9CX91|368-448|ETV5|Mus_musculus      IQLWQFLVLLDDP-AMAHFIAWT-GRGMEFKLIE--PEEVARRWGIQK-NRPAMNYDKLSRALRYVYKGGIMQKVA-GERVYKVF 81
sp|P41163|234-314|ETV2|Mus_musculus      IQLWQFLLELLHDG-ARSSCIWT-NSRREFQLCD--PKEVARLWGERK-RKPGMNYEKLRSRLRYVYDGNIMKVIH-KGRYVYRFG 81
sp|Q88422|328-400|GABPA|Mus_musculus     IQLWQFLLELLTDK-DARDCTSWV-GDGWEFKLNQ--PELVAQKMGQKK-NKPTMNYEKLRSRALRYVYDGNIMKVIH-KGRYVYRFG 81
sp|P27577|335-415|ETS1|Mus_musculus      IQLWQFLLELLTDK-SCQSFISWT-GDGWEFKLSD--PDEVARRWGIQK-NKPKMNYEKLRSRLRYVYDKNIIRKVIH-KGRYVYRFG 81
sp|P15037|362-442|ETS2|Mus_musculus      IQLWQFLLELLSDK-ANAGCITWE-GDGWEFKLAD--PDEVARRWGIQK-NKPKMNYEKLRSRLRYVYDKNIIRKVIH-KGRYVYRFG 81
sp|Q8QZM2|47-127|FEV|Mus_musculus       IQLWQFLLELLADR-ANAGCITWE-GDGWEFKLTD--PDEVARRWGERK-SKPNMNYDKLSRALRYVYDKNIIRKVIH-KGRYVYRFG 81
sp|P81270|318-398|ERG|Mus_musculus      IQLWQFLLELLSDS-SMNSCITWE-GTNGEFKMTD--PDEVARRWGERK-SKPNMNYDKLSRALRYVYDKNIMTKVH-KGRYVYRFG 81
sp|P26323|281-361|FLI1|Mus_musculus     IQLWQFLLELLSDS-ANASCITWE-GTNGEFKMTD--PDEVARRWGERK-SKPNMNYDKLSRALRYVYDKNIMTKVH-KGRYVYRFG 81
*.:.: *      : *      * .      : : * : *      * : : * : * : * : * : : : * : : *

```

Figure 3.1 Sequence variation of critical residue 236 in the most conserved structural element, recognition helix H3 among the ETS family transcription factors.

Comparative analysis for multiple alignments of the murine ETS proteins in the *M. musculus* illustrated the variation of the vital Asn²³⁶ in PU.1. It is unique compared to the corresponding residue of ETS orthologs. Only closest ETS family members of PU.1, Spi-B, and Spi-C represented Asn. Each sequence was retrieved from UniProt and aligned by Cluster Omega. ETV7 and ETV3L from rabbit (*O. cuniculus*) and rat (*R. norvegicus*) were used due to their availability. PU.1 (Spi-1) and Ets-1 are shown in black. The orthologous residue (Tyr, His, and Gln) corresponding to Asn²³⁶ in PU.1 are colored red and boxed in magenta (215).

The residue 236 has Asn in the PU.1, Q in the SPDEF, H in the ETV6 and ETV7, and Y in Ets-1 and most other ETS members. Asn236 is unique to WT PU.1 and its closest few others (SPIC and SPIB) in the phylogenetic tree of the ETS family of transcription factors, and this residue in the WT PU.1/DNA complex participates in water-mediated contacts with the core consensus in the major groove. The corresponding position to Asn236 in WT PU.1 is residue Tyr 395 in Ets-1.

Tyr 395 is engaged in a direct H-bond with the exocyclic N6 of 5'-GGA-3' in the consensus. The residue is directly mediated to H-bond with the exocyclic N6 of 5'-GGA-3' in the consensus site. We mutated Asn to Tyr to examine the significance of the residue, and interestingly the point mutation from Asn236 to Tyr itself retrieved the ability of binding profile of the full H2 chimera without conformationally conserved with WT PU.1. From the secondary structural association of the mutant, the point chimera N236Y is identical to the WT in the CD spectrum, indicating that this residue did not alter the protein structure in non-bound states. However, the mutant showed a loss of high-affinity binding such as that of the H3 chimeras. This experiment was confirmed by computational analysis of molecular dynamics (MD) simulations in terms of the interfacial hydration in the WT complex that the WT complex strongly favored the hydration of Asn236 over direct DNA contact. (Data is not shown) (215). Therefore, the residue Asn²³⁶ in the recognition helix PU.1 plays a significant role in binding to the core consensus. The residue at 236 in the H3 is evolutionarily non-conserved among the ETS families. We also asked whether Asn²³⁶ is evolutionarily sensitive or not. We mutated Asn²³⁶ to His and Gln in which they are the correspondent of ETV6 and PDEF in co-crystal structure and phylogenetic intermediates between PU.1 and Ets-1. His³⁹⁶ of ETV6 in co-crystal structure behaves similar to Tyr in Ets-1 by making a direct H-bond of the core consensus.

In contrast, Gln311 of PDEF co-crystal structure makes only water-mediated contacts such as Asn²³⁶ in PU.1. (Data is now shown) (215). We asked whether the substitution of Asn²³⁶ in PU.1 with His or Gln, which are phylogenetically intermediate, would behave less interrupting than Tyr in Ets-1 in terms of the interfacial hydration (Figure 3.2 a).

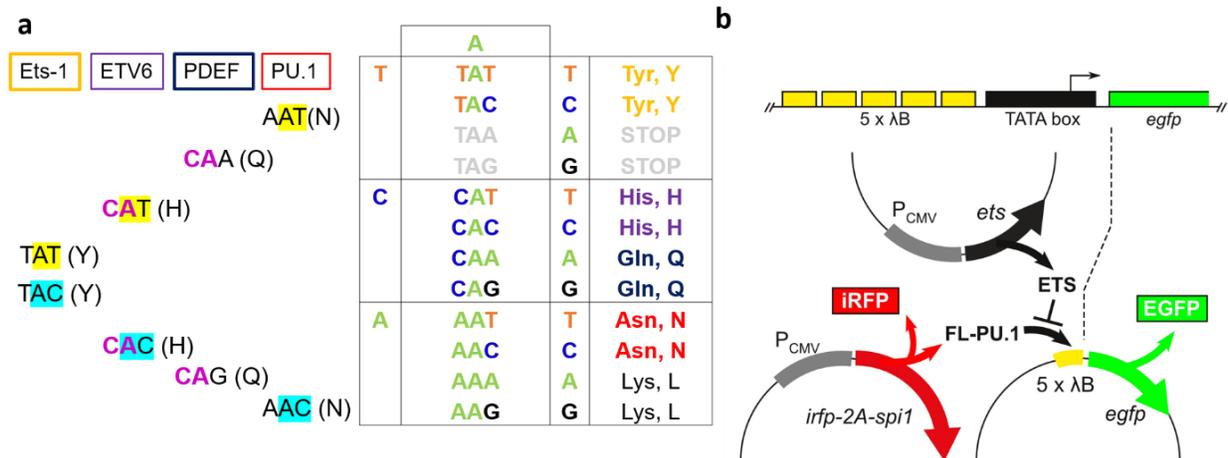


Figure 3.2 Evolutionarily sensitive residue Asn²³⁶ in H3 PU.1 ETS domain evaluated in cellular reporter gene platform.

(a) Residue 236 is the only H3 position that exhibits significant variation. All residues (Y, H, Q, and N) in position 236 are pertinent to single codon changes. (b) Schematic design of dominant-negative inhibitors of functional PU.1 transactivation in live cells. PU.1 ETS mutants N236X (X=Y, H, or Q) were examined for their cognate DNA target as a function of PU.1 inhibition by measuring EGFP reporter expression (colored in green). PU.1-dependent EGFP reporter under the control of its cognitive enhancer consisting of a 5X tandem λ B motif (colored in yellow) was used. iRFP (colored in red) was fused of PU.1 expression to monitor PU.1 expression in HEK293 cells (215).

3.3.1.2 Chimeric perturbation of PU.1 activity by evolutionary Asn 236 chimeras

To address the question, we designed the N236Y/H/Q mutants as dominant-negative inhibitors of PU.1 transactivation in cells. We monitored the effect of them over the wild-type protein on PU.1 transactivation in live cells (Figure 3.2 and Figure 3.3).

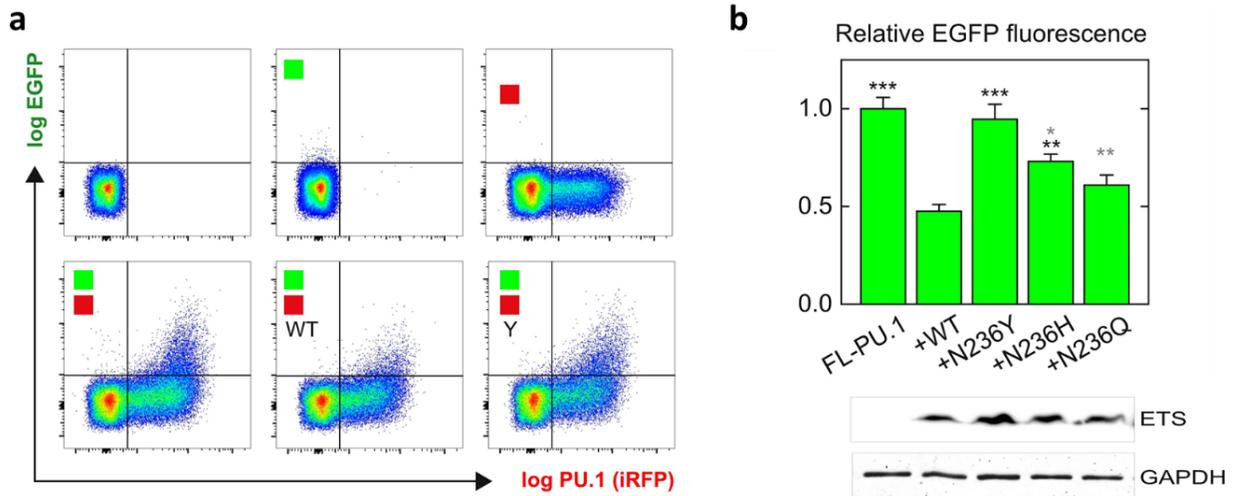


Figure 3.3 Functional inhibition of PU.1 transactivation by dominant-negative mutants, evolutionarily conservative H3 chimeras in live cells.

(a) Representative flow cytometric data in transiently transfected HEK293 cells and controls. Cells expressing ectopic full-length (FL) PU.1 were gated via co-translating iRFP marker expression. Colored boxes illustrate co-transfection with FL-PU.1 expression plasmid with EGFP marker plasmid. WT or Y (N236Y) mutant of the PU.1 ETS domain was co-transfected with all marker expression plasmid. Each axis denoted logarithmic fluorescence intensities of iRFP reporters (x-axis) and EGFP markers (y-axis). (b) Histogram of Relative EGFP fluorescence in each mutant. The EGFP fluorescence was adjusted to express the ETS domains (by immunoblots) and normalized relative to the intensity of the negative control sample as \pm S.D. ($N=3$). The statistical analysis was conducted by one-way ANOVA with Tukey's post-hoc test *, **, ***: $p < 0.05$, 0.005 , 5×10^{-4} (black, versus WT ETS; gray versus N236Y) (215).

In our cell reporter platform, the importance of the residue Asn236 was determined by which mutant N236Y, N236H, and N236Q could exert a dominant-negative effect (Figure 3.2a). We co-transfected HEK293 cells with expression plasmids encoding full-length PU.1 and WT or mutant ETS domain, together with a PU.1-dependent reporter plasmid (Figure 3.2b). The reporter expressed EGFP under the control of a synthetic enhancer consisting of a PU.1-binding λ B motif originated from the Ig λ 2-4 enhancer. In non-PU.1 expressing HEK293 cells, the

reporter was negligibly activated by endogenous transcription factors in control transfectants without ectopic full-length PU.1 (Figure 3.3).

PU.1-dependent EGFP fluorescence was determined in the subpopulation of iRFP-expressing cells (i.e., upper-right quadrant). Co-transfection of constant amounts of full-length and the ETS-encoding genes, the dominant-negative inhibitors, demonstrated that the N236Y mutant failed to effectively inhibit full-length PU.1 relative to the WT ETS domain (Figure 3.3b). The N236H and N236Q mutants were both more effective inhibitors than N236Y, with the N236Q being statistically similar to WT ($p=0.05$; Figure 1C). We confirmed that the PU.1 activity did not affect the dominant-negative mutant plasmid expression level using immunoblot assay. The results of our functional studies in cells supported MD and biophysical characterizations that mutation of Asn236 in the recognition helix altered DNA binding by PU.1 in an evolutionarily sensitive manner as phylogenetic conservation.

3.3.2 Functional manifestation of PU.1 self-regulation: negative feedback

Previous research revealed that the PU.1 ETS domain forms a 1:1 complex with its cognate sequence site. The establishment of a 1:1 complex is the basis of the regulatory mechanism of ETS transcription factors such as Ets-1, ERK, and ETV members as autoinhibition (220, 221). Autoinhibition is defined as helices adjacent to the ETS DNA-binding domain allosterically reduced DNA-binding affinity. PU.1 does not contain the structural part involving in autoinhibition. Instead, the formation of a 2:1 complex is suggested as an alternative role for self-regulated inhibitory mechanisms. The functional nature of the 2:1 complex, for which no ETS analog is known, is needed to further research. In biophysical studies, the 2:1 complex

works as its negative cooperative relationship with the 1:1 complex suggest an inactive species (Figure 3.4).

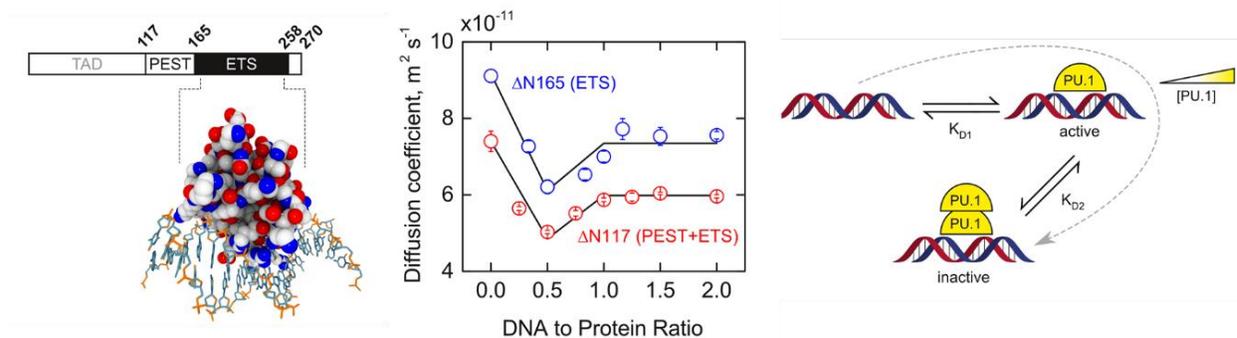


Figure 3.4 Schematic representation of negative feedback in PU.1 transactivation.

The DNA binding domain of PU.1 is the only minimal structural domain unit to bind its target DNA sites (left panel). The ETS domain bind to its cognate DNA target and forms a 1:1 complex as an active form. However, in our previous biophysical studies (DOSY NMR titration of PU.1 with cognate DNA (middle panel, Dr. Shingo Esaki performed a DOSY NMR experiment. DOSY NMR data displayed that the minimal PU.1 DNA binding domain with its target DNA reached equivalence points at DNA: Protein=0.5 and 1.0, corresponding to 2:1 and 1:1 complexes) exhibited that PU.1 formed a 2:1 complex at a single DNA site with different negative cooperativity as defined by the ratio of the two sequential dissociation constants K_{D1} and K_{D2} . Reduced PU.1 activity under conditions of an inactive 2:1 complex was predicted between dimerization and negative feedback (right panel) (219).

3.3.2.1 Design of the synthetic enhancer structure

Hypothesis

We hypothesized that the functional outcome of an inactive 2:1 complex exhibited a bell-shaped curve as PU.1 doses increased. As a result of the PU.1 binding sites saturated with (2:1) bounded PU.1, the complex does not produce reporter signals. In the alternative, the reporter signal would dose-dependently reach a saturable level. Thus, it relies on the level at which the 2:1 complex keeps possessing activity relative to the 1:1 complex. Since the enhancer varied in

density and spacing between the cognate sequences (i.e., *cis*-regulatory syntax), we determined the effects based on the variants of PU.1 dependent regulatory elements. Therefore, the synthetic λ B reporters were well suited to systemically measure for appropriate responses of PU.1 activity in live cells. Furthermore, our cell-based reporter gene assay does not require any other cofactors. Thus, the cell reporter platform is a non-invasively technology to monitor the cellular PU.1 activity on its cognate DNA sequence.

PU.1 dosage setup

To investigate a feasible dose range of PU.1 expression plasmid, we established it by measuring the amount of PU.1 expression in native PU.1 expressing myeloid cells (Figure 3.5).

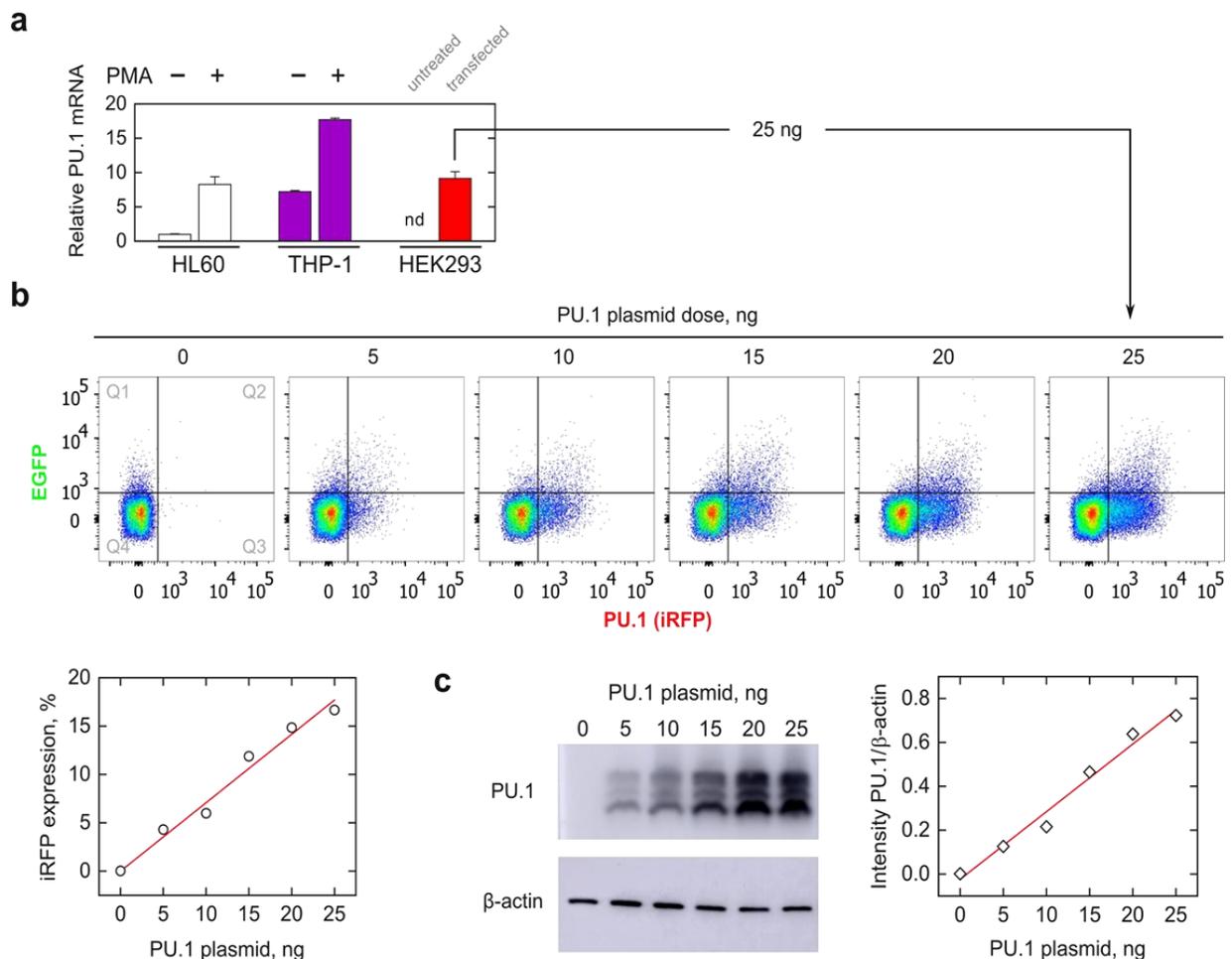


Figure 3.5 Adjustment of transgenic PU.1 dosage for cellular titration experiments in HEK293 cells to native PU.1 expressing myeloid cells.

(a) 25 ng as a maximum dose for an amount of PU.1 transgene was determined using qPCR. Relative PU.1 mRNA levels were quantified and adjusted to PU.1 levels in normal proliferate HL-60 and THP-1 and in Phorbol 12-myristate (PMA) induced cells (16 nM). Data was shown in means \pm SE of at least three biological replicates relative to gapdh. (b) Representative flow cytometric data for series of doses of PU.1 transgenes up to 25 ng per sample in a 24-well plate were transfected in HEK293 cells. PU.1 expression was quantified by co-translating iRFP marker expression. Cells were gated by the control no PU.1 plasmid transfected samples and divided into each quadrant. Cells in Q2 and Q3 represented PU.1-expressing cells, and the total counts of Q2 and Q3 were plotted against the plasmid dosage. Redline illustrates a linear fit to the data. (c) The amounts of PU.1-expression were validated by immunoblot assay. Each transfected HEK293 lysate (10 μ g) was quantified, normalized by housekeeping protein expression (β -actin), and plotted (219).

The maximum expression of PU.1 plasmid was limited to the level of PU.1 expression found in the myeloid cells to be physiologically inducible. For example, THP-1 cells, myeloid cells can be inducible by PMA and differentiated into macrophages. We purposely set up the amount of PU.1 expression plasmid to evaluate our functional PU.1 inhibition result to interpret in a physiologic context (Figure 3.5a). To determine whether 25 ng of the PU.1-expression plasmid would yield a linear dosing range for the λ B reporters following transient transfection in HEK293 cells, we measured the reporter gene expression level of each transfected HEK293 cell samples by flow cytometer. We confirmed the linear variation in abundance PU.1 plasmid expression in HEK293 cells by flow cytometry (Fig 3.5b). When transiently transfected into PU.1-negative HEK 293 cells, we confirmed that the reporters were negligibly activated by endogenous transcription factors, including other ETS family proteins (Figure 3.5b). The full-length PU.1 plasmid expression was independently tracked by quantifying as detecting iRFP positive cells by a flow cytometer and represented it in the linear relationship between PU.1 plasmid dose and iRFP expression. For validating the flow cytometric approach, PU.1 abundance was also determined by immunoblot assay (Figure 3.5c). We enable to quantify that PU.1 protein

expression in transfected HEK293 cells was increased as an increased dose of PU.1 expression plasmid.

Design of synthetic PU.1-dependent enhancer reporter structure

For probing PU.1's *trans*-regulation of negative feedback mechanism, marker gene expression was detected from various synthetic enhancer elements consisting only of tandem copies of the λ B motif (Figure 3.6).

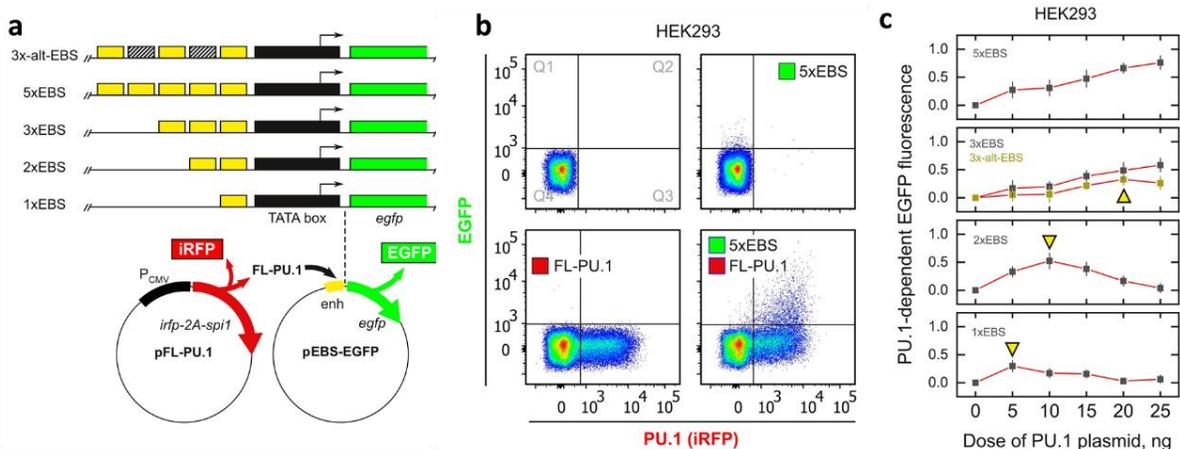


Figure 3.6 Functional PU.1 trans-regulation by negative feedback.

(a) Schematic diagram of synthetic PU.1-dependent EGFP reporters. A minimal TATA box (black blocks) was operated by enhancers composed only of tandem EBS (yellow blocks) spaced 20 bp apart upon functionally binding by PU.1. Mutated sites were denoted as hatched blocks. (b) Representative flow cytometric data of untreated HEK293 cells and transfected HEK293 with a constant dose of the 5xEBS reporter and/or up-to 25 ng of an expression plasmid encoding full-length PU.1. Quadrant Q2 contained the EGFP marker-expressing cells counted out of all PU.1-expressing cells (Q2+Q3). (c) Upon PU.1 dosages, PU.1-dependent EGFP marker expression in each type of synthetic enhancer construct was plotted. EGFP fluorescence in Q2 taken over the summed fluorescence in Q2 + Q3 at 24 hours after co-transfection of the EGFP reporter plasmid and the indicated dose of PU.1 expression plasmid. Each data point is the means \pm SE of triplicate or more samples. The yellow triangle denotes the peak of the EGFP reporter expression (219).

The motif was used for a PU.1-specific ETS binding site (EBS) derived from the lymphoid Ig λ 2-4 enhancer. We constructed various expression plasmids in terms of enhancer syntax (numbers of copies and spacing two helical turns, 20 base pairs (bp) between each consecutive site) and measured reporter gene expressions reflected by the effects of enhancer structures. Because the bound proteins were arrayed on the same helical face, this structure assisted in facilitating the recruitment of the transcriptional machinery (Figure 3.6a).

Further, the spacing enhancer would amplify site-site interaction and DNA perturbations as the 2:1 complex was known to require an extended site size relative to the monomer, presenting the bound protein along one helical face. By doing so, therefore, it would permit evidently to show functional effects of the 2:1 complex. In this design, we quantified PU.1-dependent transactivation as the fraction of iRFP-positive cells that were also EGFP-positive (Figure 3.6b).

3.3.2.2 Dosage effect of PU.1 trans-regulation supports negative feedback

As we expected a multivalent impact concerning PU.1 binding site, all enhancer constructs illustrated graded reporter expression in step with the density of EBS at each enhancer at identical PU.1 dose (Figure 3.6c). However, the dose-dependent curves of EGFP expression were only shown on the enhancers harboring tandem the 3x and 5x EBS, not 1x and 2x EBS. Upon peaking at intermediate PU.1 dose, the 1x and 2x enhancers were going down by further increases in PU.1. Therefore, the curves from the two constructs displayed the non-functional formation and decreased the marker signals, indicating that the dose-dependently settled to a saturated level relying on the 2:1 complex retained activity relative to the 1:1 complex. However, enhancers harboring tandem 3x and 5x EBS showed increased EGFP expression. Therefore, we

decided to ascertain whether PU.1 reversibly interacts with the enhancer in transactivation. To do that, a 3x-EBS variant (3x-alt-EBS) mutated in even-numbered sites in the 5x-EBS reporter was constructed. In the construct of 3x-alt-EBS, it generated the cognate sites doubled in spacing (Figure 3.6a). Compared to 3x-EBS constructs, the 3x-alt-EBS reporter exhibited decreased PU.1 transactivation and its reporter signal also no longer increased. The effect of enhancer space between the cognate sites explains that the functional reversal could not be caused by PU.1 interaction away from the DNA only. The spacing effect demonstrated that the functional reversal would be inactive to syntax changes at the DNA (Figure 3.6c).

Enhancer syntax variations (density and spacing) demonstrated negative feedback in PU.1 transactivation in live cells. The PU.1 activity illustrated the bell-shaped dose responses for the 1x and 2x enhancers, but not the 3x or 5x-EBS enhancers (Figure 3.6c). It indicates that the helical spacing of the sites amplified the local DNA structure and additive perturbations exhibited as we observed previous DNA footprinting of the PU.1 ETS domain. We might suspect that PU.1 bind at the higher-density sites might exhaust a required co-repressing factor for the 2:1 complex. However, this possibility was minimal because the different dose responses showed in functional outcomes of enhancer construct of the 3x-EBS and the 3x-alt-EBS, which had the same site density.

Further, the events were observed in a cell line (HEK293) that does not natively express PU.1 in gene regulation. Because net transactivation activity was decreased under conditions correlated with the population of the 2:1 complex, the data suggested that the 2:1 complex lost activity relative to the 1:1 complex. The functional outcomes and its effect of the synthetic λ B reporter on PU.1 dose-dependent manner supported the biophysically observed 2:1 complex as a functionally relevant in the cells. Thus, our cell reporter platform of enhancer syntax (site density

and spacing) provided compelling evidence of the intrinsic PU.1 self-regulatory activity in biological systems.

3.4 Discussion

Sequence-specific transcription factor PU.1 is a lineage- and dosage-specific member of the ETS family of transcription factors that share a conserved DNA-binding domain with overall low sequence homology among the ETS family members but structurally homologous (222-225). All ETS members execute in hematopoietic cells for blood cells' differentiation in multi-steps with complicated controls in which the control steps happen at transcription levels. Here we presented two reporter platforms to characterize the functional status of the control process for PU.1 transactivation. The use of cell-based reporter gene platform, critical determinants of transcription factor PU.1 activity at TF-DNA level was functionally characterized. We determined functional inhibition of PU.1 for binding its cognate DNA target sites in live cells and proved the biological relevance of the functional heterogeneity in DNA target selection and control by PU.1. The evidence was suggested from biophysical aspects: molecular hydration and self-association. Dominant-negative mutants in the reporter assay were utilized for understanding functional heterogeneity adapted from environmental changes and attested significance of critical residue in PU.1 ETS binding domain evolutionarily. Various enhancer structures fused to reporter genes were employed to gain insights into intrinsic regulatory mechanisms of PU.1 negatively feedback in efforts to elucidate how PU.1 is inducible in a stage-specific and dosage-specific manner.

3.4.1 Functional approaches to probing molecular hydration in transcription factor function

Interfacial hydration is a mechanism by which proteins may recognize specific DNA sequences. We explored the essential strategies of molecular hydration as a role of particular DNA recognition by PU.1 ETS proteins in sequence specificity. The sequence-specific recognition of PU.1 has been achieved from the evolutionary adaptation process through the single transition of amino acid codon and conferred the PU.1 reacted more sensitive on osmotic pressures. In cooperation with water, the molecules led PU.1 to increase its DNA target specificity to discriminate other sequences. Previous reports studying NFAT5 protein, known for the responsive gene of osmotic stress, were expressed in the thymus and activated in lymphocytes (132, 226-229). The observed dependence of lymphocyte function on the NFAT5 osmotic stress response pathway suggests that mechanisms responded to osmotic stress in the thymus where cells for innate immune responses are generated (131, 132, 227-232).

NFAT5 mediates protection from adverse effects of hypertonicity (226). Osmotic stress activates NFAT5 via osmolyte-induced disorder-to-order conformational rearrangement. Therefore, osmotic sensitivity influences the transcription factors' activity in cells. In the context of sophisticated differentiation pathways in mammalian cells, molecular hydration plays a critical role in the master regulator of PU.1 on evolution. Sequence-specific transcription factor PU.1 attained its DNA target specificity in adaption in osmotic sensitivity by DNA recognition of sequence discrimination.

3.4.1.1 Residue Asn236 integrates hydration contributions of DNA binding by PU.1

Water molecules play a vital role in transcription factor-DNA interfaces. Structural and biochemical data have suggested that water-mediated interactions are as meaningful as direct

hydrogen bonds in the stability and specificity of recognition (130, 233-235). From our previous analysis of structural studies of PU.1 ETS domain, ETS domain of PU.1 is sensitive on osmotic pressure to bind cognate DNA targets compared to the ETS domain of Ets-1. We have mapped the critical residue in charge of the interfacial hydration in transcription factor PU.1, which the Ets-1, the same ETS family member of the transcription factor, does not contain (113).

Interestingly, although Ets-1 and PU.1 are structural homologs, only PU.1 exhibits sensitivity to osmotic pressure for detecting its specific DNA sequences and its binding. To assess the interfacial water contact area of PU.1 ETS domain, we made chimera to substitute the several structural parts of Ets-1 into PU.1 ETS domain. The H3 domain exposed the most prominent elements of the PU.1 ETS domain through biochemical and biophysical assessments and showed a significant loss of normo-osmotic pressure and sensitivity. To gain insights on the essential structural residue to provide the osmotic sensitivity of PU.1, we first did sequence alignment of H3 motifs across all the ETS paralogs; all residues in this motif are very conserved except the residue 236. Indeed, N236Y, in which we mutated Asn to Tyr, corresponding to the same deposition of Ets-1, alone exhibited recovery of the identical effect of H3 chimera. Consistently, N236Y is functionally not sensitive to inhibit PU.1 transactivation using our dominant-negative inhibition assay in live cells.

The functional outcomes in live cells were aligned with predicted results from Molecular Dynamic simulation. The simulation exhibited that the H236Y chimera in ETS/DNA complex configuration was distorted by a re-orientation of the protein relative to the target DNA site. Furthermore, the H236Y chimera exhibited the most significant impaired function to bind its cognate sequence site as Ets-1 (215). In this study, we attested evolutionarily the importance of a critical amino acid residue in the helix 3 domain of PU.1 DNA binding motif that biophysical

experiment suggested its importance of molecular hydration. Also, N236H and N236G work as similar effects more likely on its evolutionary position in the phylogenetic tree.

3.4.1.2 Asn236 is a marker in the evolution of the ETS domain

Residue 236 is the only H3 position that illustrates a significant variation in the DNA-binding domains of ETS transcription factors among ETS orthologs. The sequence divergences (Asn, His, Gln, and Tyr) in the residues of the ETS proteins are often adapted from a codon variation from Tyr to Asn progressively. The variation might evolutionarily happen among the phylogenetic intermediate between Tyr (Ets-1) and Asn (PU.1) to interchangeable A to T and C to G or vice versa. The single codon transition was shown the residue Asn236 in PU.1 from Tyr in Ets-1, the ancient ETS TF family member. His (ETV6) closed to Tyr (Ets-1) and Gln (SPDEF) adjacent to Asn (PU.1), where water mediates contents, exhibited similar effects in our functional PU.1 inhibition assay as well as MD simulation. The functional extent of the PU.1 inhibition exhibited as phylogenetical ranks in the ETS family: N236Y (Ets-1) > N236H (ETV6) > N236Q (SPDEF) > N236 (PU.1) (236). A single transition of codon also exhibits evolutionarily separate from one another in the ETS family phylogenetic tree.

Further, the relative hydrophobicity of amino acid residues of Tyr, His, Gln, and Asn at near-neutral pH exhibited the fact that the interfacial hydration properties are evolutionarily moved toward PU.1 that more stringent to be sensitive on hydration (237). Thus, the most recent ETS member PU.1 was evolved to acquire its traits to gain from more hydrophilic residue toward more hydrophobic attributes so that PU.1 is highly sensitive to bind its DNA recognition to control transcription upon environmental changes or stimulations. Therefore, the significance of H3 in PU.1/DNA binding is also closely attached to interfacial hydration and indicates an

evolutionary transition in the essential recognition helix toward the acquisition of interfacial hydration.

3.4.1.3 Significance of studying molecular hydration in transcription factors

For transcription factors, hydration plays an essential role in their target DNA recognition. The *Escherichia Ecoli* tryptophan repressor is iconic in this respect. Both the crystal structure and experiments in solution show that it recognizes operator DNA primarily via water-mediated contacts. More recently, we have found that hydration is strongly coupled with site discrimination by the ETS-family protein PU.1, a master transcription factor in hematopoiesis. Typical of this family of transcription factors, PU.1 recognizes cognate DNA sites ~10 bp in length that contains a 5'-GGA(A/T)-3' central consensus. In addition to serving as an experimental probe for studying hydration changes, osmotic stress is a physiological condition in hematopoietic tissues. We analyzed gene expression data in murine macrophages and found that target genes for PU.1 are significantly over-represented in osmotically responsive genes. Recently, in agreement with their osmotic profiles in binding experiments, hypo-osmotic stress in K562 leukemia cells was reported to drive PU.1 binding to promoter sites in significant excess over Ets-1. Therefore, the differential interfacial hydration in DNA binding also establishes the two ETS members' roles in the cellular response to physiologic osmotic stress.

3.4.2 Self-regulatory mechanisms of PU.1

3.4.2.1 Inducible PU.1 in the immune system and the molecular mechanism of PU.1's regulation

One of the fundamental biological features of TF PU.1 is a lineage-specific and dosage-specific manner for controlling differentiation of the hematopoietic stem cells. In macrophage or

erythrocytes, PU.1 is a markedly inducible transcription factor during hematopoiesis and immune stimulation. Depending on the combination of a cell line, physiology, and stimulatory ligands, cellular PU.1 abundance varies. As previously well-known of biological features, PU.1 works in lineage-specific and dosage-restricted manners. With combinatorial regulation with cofactors, PU.1 should increase the differentiation of B-cells and macrophages, contrary to the differentiation of T-cells. Disrupted PU.1 activity directly or indirectly relates to various disease occurrences, including neurologic disruption, fibrotic disease, Type II diabetes-related hepatic disease, hematological cancers, and leukemias. Acute myeloid leukemia has one of the most manifest examples of the dosage-effect to PU.1 abruption such that 20 % or below of PU.1 expression has associated with the AML in human patient samples and murine models. The dosage effects of PU.1 have been extensively studied in physiology and molecular biology.

3.4.2.2 Real-time regulation of PU.1 at the protein/DNA level

During hematopoiesis and immune stimulation, the cellular and molecular regulatory mechanisms of PU.1 activity have been intensively studied. PU.1 negatively cooperates with cofactors (GATA-1) or positive way with IRF4. The PU.1 gene expression level was also regulated by *cis*-acting regulatory elements, including URE distal elements (238, 239) (240-244). However, the mechanism by the *trans*-regulatory complex of PU.1 relatively little is understood at the protein/DNA level. The cellular abundance of PU.1 should be real-time regulated at target DNA recognition by PU.1. The DNA binding (ETS) domain of PU.1 is known as its only structured domain to interact with its cognate DNA site for gene regulation. The ETS domain is structurally conserved in the ETS family of transcription factors and forms a 1:1 complex with cognate DNA.

In contrast with other ETS members, the ETS domain of PU.1 forms a 2:1 complex at single DNA cognate sites supported by thermodynamic and biophysical studies. It is considered to serve intrinsically self-regulated by negative feedback based on biophysical and biochemical research data. To functionally assess the complex to serve as a self-inhibitory mechanism in cells, we designed several reporter gene plasmid constructs using PU.1-dependent gene, lymphoid Ig λ 2-4 enhancer with two helical turns spacing between each cognate gene segment. Therefore, the effects of enhancer structures on PU.1 transactivation demonstrated PU.1 *trans*-regulation in negative feedback in live cells (245). Depression of negative cooperate manner of PU.1 activity on its target DNA by an inactive 2:1 complex relative to a 1:1 complex exhibited the self-regulatory mechanism of the inducible PU.1 in real-time at the protein/DNA level.

3.4.2.3 Functional relevance of self-titration as a potential negative feedback mechanism for PU.1 transactivation

Many ETS family transcription factors, such as Ets-1, ERG, and members of the ETV subfamily, are regulated at the protein/DNA level by inhibitory helices that pack against their DNA-binding domain in the unbound state (219). Perturbing these helices imposes an energetic penalty on DNA-binding that maintains, by default, an autoinhibited state. Binding partners that disrupt the autoinhibited interactions thus induce a transcriptionally permissive state. ETS paralogs, such as PU.1, that lack this mechanism would thus be locked in a permissive state in the absence of negative feedback fashion: even if the 2:1 complex retains the functional activity of the 1:1 complex, removal of circulating PU.1 alone will attenuate transactivation of target genes. Consistent with this notion, we observed self-titration only with site-specific DNA and not nonspecific DNA.

Further, we did not keep dimer formation with the structural homolog Ets-1, with or without its autoinhibition helices, when its cysteines were maintained in a reduced state. Interestingly, a 2:1 Ets-1/DNA complex was reported under non-reducing conditions, reflecting the strong propensity for its two cysteine residues (which are not present in the PU.1 ETS domain) to form non-native disulfide linkages (112, 246).

To our knowledge, an inactive 2:1 complex form of transcription-DNA has not been studied before. In the ETS complex, most transcription factors are bound to 1:1 with its cognate DNA. To understand the biological role of the 2:1 complex formation in terms of transcriptional machinery at the level of protein/DNA, we employed a cellular enhancer-reporter platform to determine the effect of changes in enhancer syntax on PU.1 activity. Enhancers with low-affinity binding sites can mediate robust tissue-specific gene expression patterns when they are organized with optimal syntax. The importance of sequence constraints within developmental enhancers for tissue-specific gene expression patterns suggested the evolution of developmental enhancers depending on the selection of submaximal binding sites. Also, optimal binding sites defined as genes with the highest affinity and optimal spacing for expression levels determine enhancer function (58). As we predicted, using various tandem copies of PU.1-dependent gene, lambda, fused to EGFP reporter gene, the EGFP expression followed as a bell-shaped curve as PU.1-dose dependent manner. A bell-shaped reporter dose-response as the enhancer varying in density and spacing of EBS was designed for measuring the functional outcome of an inactive, negatively cooperative 2:1 complex. The unique feature of PU.1's functionality with regard to *trans*-regulation of negative feedback will be studied further.

3.5 Conclusion

With combinatorial interaction with other cofactors, DNA target selectivity is acquired from the intrinsic heterogeneity among DNA binding domain, primarily ETS domain for ETS family of transcription factors. This study aimed to characterize the transcriptional regulation of the ETS family of transcription factor PU.1 activity related to its cognate DNA binding sites. Using the PU.1-dependent enhancer; we proved two critical regulatory elements of PU.1 transactivation in molecular hydration and self-association as a function of negative feedback regulation. Our cell reporter technology of dominant-negative mutants enables us to understand functional PU.1 transactivation was abolished. The evolutionarily critical residue of PU.1 ETS domain H3, known as the importance of interfacial hydration, was considered necessary mutated with corresponding residue of ETS paralogs. Also, by constructing various enhancer-reporter gene platforms, PU.1-DNA complex interactions were regulated by a negative feedback mechanism in live cells. Our results demonstrated that our versatile reporter gene platform was utilized to functionally support data in the context of biology for biophysical research. The signatures of target DNA selection by transcription factor PU.1 suggested by biophysical and computational research were biologically relevant proven by reporter gene platform manipulated relatively easily and suitable to characterize functional PU.1 activity in live cells according to our research goals.

4 CHARACTERIZATION IN LIGAND SCREENING AGAINST PU.1

4.1 Introduction

4.1.1 *Drug screening for targeting transcription factor PU.1*

The process of drug development is time-consuming and cost expensive. Understanding a drug's function within biological systems can shorten production times and lower the risks that a drug will have extreme, unforeseen side effects. Therefore, it is crucial to obtain the maximum amount of information about the biological activity, toxicological profile, biochemical mechanisms, and off-target interactions of drug-candidate leads in the earliest drug discovery stages. Due to their advantages of predictability, the possibility of automation, multiplexing, and miniaturization, cell-based assays seem a valuable tool for the high demands of the early stages of the drug discovery process. They should, consequently, be considered as a viable research avenue. This is not to say other research strategies should become obsolete. Relying on different strategies ranging from reporter gene technology to protein fragment complementation assays remain necessary and present various challenges.

TF PU.1 and AML

While recent advances regarding the pathogenesis and treatment of AML have improved clinical outcomes, conventional chemotherapy remains the standard treatment choice. This is due to the complexity of AML, as the disease results from numerous genetic aberrations. To address the genetic variety of AML, alternative therapeutic approaches for targeting commonly dysregulated pathways present in multiple AML subtypes have arisen. Despite the genetically heterogeneous nature of AML, the de-regulation of the PU.1 transcription factor is a recurrent event found in most subtypes of AML (247).

PU.1 is an ETS family transcription factor encoded by the SPI1 gene and is primarily expressed in myeloid cells and B-lymphocyte cells (248). PU.1 specifically recognizes purine-rich DNA sequences containing 5'-GGAA/T-3' at the target gene promoters (249). Cells with hypo-PU.1 activity (20% of wildtype) lead to the onset of AML. Inhibition of residual PU.1 activity results in cell death of only cancerous cells. More than 40% of AML patients have low expression of PU.1 resulting from genetic disruption. Homozygous knockout of an enhancer located -14 kb upstream of PU.1 (URE) causes an 80% reduction of PU.1 expression and leads to AML. Conditional knockout of PU.1 (URE^{-/-}) leads to stem cell failure (128, 250-254). Collectively, this data demonstrates that PU.1 is critical for stem cell generation and homeostasis. Suppressing PU.1 activity in leukemic stem cells with low PU.1 activity may be a viable approach for targeting leukemic cells over healthy hematopoietic stem cells.

4.1.2 Sequence-specific DNA minor groove binding small molecules

TFs regulate gene expression to control protein synthesis, increase or decrease gene transcription, and alter cell function. This is achieved by TF-binding regulatory DNA sequences called enhancers or promoters in the transcription process's first step. The features of TF containing DNA-binding domains enable them to bind to their specific sequences of DNA. Among thousands of sequences, TFs' specific recognition of their target site, enhancer, or promoter is not easy because many TFs recognize similar consensus DNA-binding sites and genomes harboring numerous consensus. For instance, the ETS transcription factor family shares a conserved DNA-binding domain, containing 85-residues, explicitly recognizing the consensus core of 5'-GGA(A/T)-3'. The conserved domain shows that the functions of the ETS domains in TF proteins are critical. While the ETS family proteins contact the major groove of the core sequence, their loop also interacts with flanking bases via backbone contacts at the

minor groove. As a result, the flanking sequences are frequently conserved. This is especially true for PU.1, a protein of the ETS family TF PU.1 knows its target binding specificity, and its affinity is achieved by contacting the harboring sequences of the PU.1 (113, 114, 116, 118).

PU.1 binds to the consensus site in the major groove because of its binding selectivity for AT-rich flanking sequences. The major groove harbors a 5'-GGA(A/T)-3' and gains its DNA-binding affinity through additional contact with the minor groove of the nearest AT-rich region. This AT-rich region flanks the major groove either up to three bases upstream or downstream of the core consensus. PU.1's signature is its strong preference for AT-rich sequences, highlighting a specificity on its target binding from thermodynamic studies. This is a unique feature from other ETS proteins.

Small molecules (102 Da), such as organic dications, have been studied for pharmacological applications on cancers, microbial, and viral diseases (255, 256). Compounds that bind DNA minor groove regions are targeted, especially by transcription factors, because DNA minor groove binders can activate or inhibit transcription, where gene expression is mainly controlled. The study of compound binding sequence-specific mode helps us investigate how genomes are controlled and selectively target diseased cells through the sequence-specific, minor groove-targeted compounds' binding.

By using the concept of compounds' sequence binding preferentially, pentamidine or furamidine (DB75) pharmacologics have been developed. These chemicals are antiparasitic therapeutics for African trypanosomiasis and sleeping sickness (257-262). The structural aspect of chemicals such as pentamidine and furamidine to bind to DNA was pioneered from biophysical research and proves that DNA is a biological target for drug development. In their research, Neidle's teams used X-ray crystallography to reveal that pentamidine could be bound to

a four base pair-AATT-sequence in d(CGCGAATTCGC)₂ (263). The research showed that (a) phenylamine offers minor groove recognition units in AT sequences, (b) H-bond formation with their acceptors of the base pair at the edges in the minor groove effectively stack in the narrow minor groove, and (c) the groove has electrostatic potential.

DB75 and most heterocyclic dications have similar properties (257, 258, 264, 265). The DB75 compound enables the formation of H-bonds as the compound bonds to the acceptors on the base pair edges at the floor of the groove stack of narrow minor groove walls with AT sequences. The compound's positive charge can then offset the phosphate charges and negative electrostatic potential of the groove. The H-bonds are the result of non-covalent binding to target DNA and its inherited sequence specificity. The DB75 compounds mostly preferentially target AT-rich sequences of the minor groove, and this tendency shows the potential for similar compounds to serve as target-specific drugs in cancer treatments. Specifically, the biophysical study of DB75's properties provides solid foundations for designing and developing new compounds which can target the DNA minor groove in various complexes to enhance the bioavailability, especially of cell uptake with limited toxicity.

4.1.3 DNA binding preference of the ETS family TFs

Transcription factors play crucial roles in gene expression for controlling and maintaining cellular activities: specifically, how cells function and respond to the environment. Misregulation of the gene expression from the abruption or alteration of transcription factors or associated cofactors often provokes disruption of cellular activities and leads to diseases such as cancers, autoimmune disorders, neurological disorders, developmental syndromes, diabetes, cardiovascular diseases, and obesity. The ETS transcription factors are particularly implicated in the development of most of these diseases. Therefore, the ETS proteins have received significant

attention as potential drug targets (68). Traditionally, the ETS proteins and other transcription factors were considered "undruggable" due to lack of target sites for drugs, called binding pockets in the protein or protein interface; however, the conventional rationale is changing as new approaches that target the interface of proteins or DNA are developed.

The ETS family of proteins comprises 28 transcription factors, each of which contains a highly conserved DNA binding domain. The structure of the conserved ETS domain is called the winged helix architecture consisting of three alpha-helices on a small four-stranded, antiparallel beta-sheet scaffold. This structure can be viewed using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. The structure of the conserved ETS domain contains about 85 residues. This domain is necessary for DNA binding and is involved in multiple protein-protein interactions in addition to its well-defined role in mediating DNA binding. The ETS family transcription factors bind to various protein partners that inhibit or enhance DNA binding. Additionally, they bind to coactivators and corepressors that modulate their transcriptional regulatory properties. The proteins bind to DNA over a region spanning 12-15 base pairs (bp), but they display sequence preference for ~9 bp with a core motif of '5-GGAA/T-3,' regardless of the proteins' functions.

By inserting the proteins into the major grooves of target DNA regions that harbor the core consensus of 5'-GGAA/T-3', the ETS proteins bind to the DNA. Additionally, the members of the ETS family seem to bind with flanking sequences along the DNA's backbone. The sequence variants determine the ETS protein's specificity at the flanking regions. This sequence-specific recognition shares a typical characteristic across the ETS family of transcription factors. However, the effects vary among different proteins. PU.1, for example, shows a unique use of

this feature. PU.1 is one of the most restricted ETS class III proteins; however, it strongly favors binding to adenine-rich sequence near to the core consensus.

4.1.4 DNA recognition of PU.1

The ETS family transcription factor PU.1 is also known as a Spleen focus forming virus Proviral Integration site-1 (Spi1). The transcription factor, PU.1, plays a significant role as a lineage-specific regulator of the hematopoietic system's gene expression for development, differentiation, proliferation, and self-renewal. PU.1 binds to purine-rich sequences with several myeloid- and lymphoid-specific promoters or enhancers near target genes. Specifically, PU.1 binds to AT-rich DNA sequences containing 5'-GGAA/T- 3' at the promoters of target genes. Different effects are observed in coordination with other transcription factors (129, 266-271). PU.1 expression is dynamically controlled throughout hematopoiesis to direct appropriate lineage specification. During myeloid cell (macrophages and neutrophils) and lymphoid cell (B- and T-lymphoid) development, the expression of PU.1. increases in B cells, macrophages, and neutrophils; however, it decreases in T-lymphoid cells. Therefore, the expression of PU.1 is a crucial feature of functional cellular differentiation. Disruption of PU.1, though, is involved in diseases such as Alzheimer's disease, Hodgkin's disease, fibrosis, liver-associated diabetes, and hematological cancers, including leukemias (101, 272-297). For this research, we look at the role of the disruption of PU.1 levels in humans and its association with AML. Specifically, hypo-functioning PU.1 (20% of wild-type activity) leads to the progression of AML in several subtypes of the disease(121, 128, 298). Therefore, if we modulate the expression levels of PU.1 using small chemical compounds to bind PU. 1's target DNA sites, the advancement of diseased cells can be decreased. The approach will offer a new type of therapeutic development for AML.

4.1.5 The pharmacological strategy of targeting the PU.1-DNA interaction by sequence-specific binding molecules

According to Munde and his colleagues, some heterocyclic diamidines have been shown to have the potential of pharmacological control in the inhibition of PU.1. They hypothesized that the DNA minor groove binders inhibit the interaction of PU.1 and their target DNA site via an allosteric mechanism (299). It is known that PU.1 binding is disrupted by diamidines that target AT-rich sequences, where PU.1 gains its binding specificity through the 5' flanking minor groove as well as the core consensus sequence (Figure 4.1).

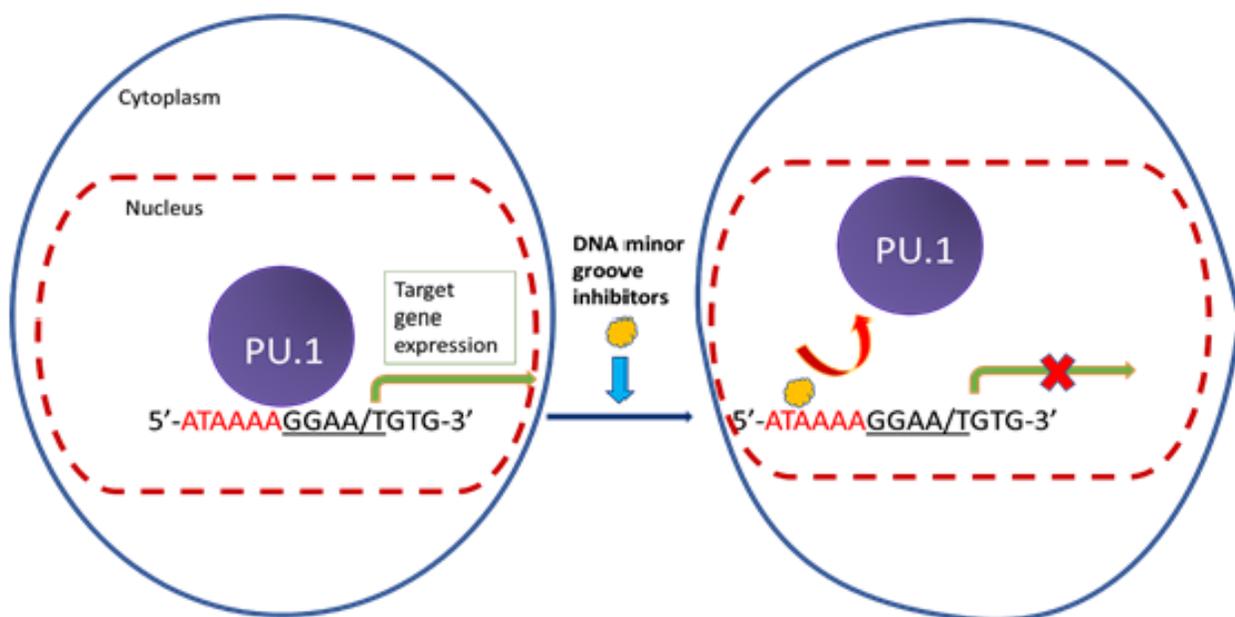


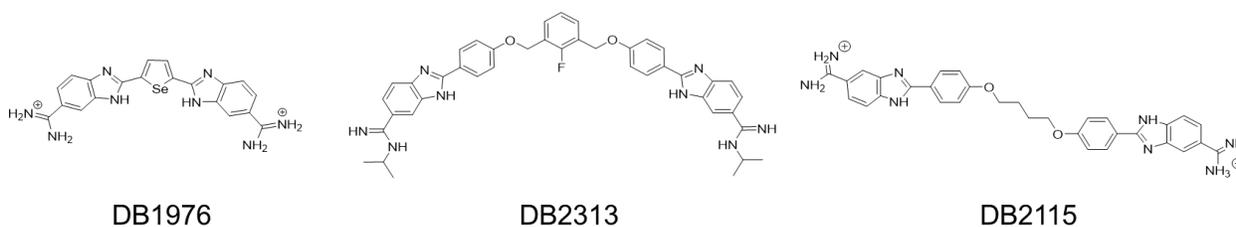
Figure 4.1 Schematic diagram strategy of the DNA minor groove ligands against PU.1 for AML therapeutics.

To modulate PU.1 transactivation, by adding to small DNA minor groove inhibitors, PU.1 fails to achieve its binding sequence selectivity from AT-rich flanking sequences in the minor groove. Targeting low PU.1 expressing leukemia stem cells to PU.1 inhibition can lead to blocking its binding to major groove via an allosteric mechanism. As complete loss of PU.1 results in leukemic stem cell failure, it may remove low PU.1 expresses AML cells and triggers regenerate healthy normal hematopoietic stem cells in AML patients.

Munde and his colleagues showed that the diamidines targeted the AT-rich sequences using the λ B motif of Ig2-4 enhancer (5'-ATAAAAAGGAAGTG-3'); these regions are high-affinity PU.1 binding sites with AT-rich tracks flanking both sides of the ETS consensus (84, 111). As a result, these diamidines disrupted the allosteric interaction of PU.1 and the DNA sites *in vitro* and live cells. The research illustrated that the compounds that inhibited PU.1 were accumulated and successfully localized in the nucleus, the compounds' target sites.

Interestingly, they found the structure-dependent differences in PU.1 inhibition and nuclear localization among the compounds using biosensor-surface plasmon resonance (SPR) screening that targets the PU.1- λ B complex. The research proved that pharmacological inhibition of PU.1 could be achieved via an allosteric mechanism. The approach has opened a new pathway of developing treatments targeting AML. By inhibiting residual PU.1 activity by heterocyclic diamidines, it is possible to trigger stem cell failure to restore normal hematopoietic stem cells, effectively resetting the system.

Our research progressed using the heterocyclic diamidines DB2115, DB2313, and DB1976 as first-test compounds to ascertain whether this theory holds potential for application. (Figure 4.2)



λ B motif: 5' – CCAAATAAAAGGAAGTGAAACCAAGCT
3' – GGTTTATTTTCCTTCACTTTGGTTCTC

Figure 4.2 Chemical structure of the DNA minor groove binders.

The heterocyclic diamidines preferably bind to AT-rich sequences in the 5'-flanking sequence, harboring the ETS core consensus sequence (GGAA) in the λB motif bound by PU.1. The three compounds were selected by SPR analysis as candidates for characterizing the ligands' properties against PU.1 in live cells.

These compounds were shown to have significant potential as anti-leukemic agents through their ability to bind to PU.1 target sites and inhibit PU.1 transactivation via an allosteric mechanism. In live cells, we evaluated their potential as PU.1 inhibitors to determine their potential as anti-leukemic agents; further, their target specificity in nuclear localization was assessed in the hopes of finding minimal off-target cytotoxicity.

Furthermore, we recorded the kinetics of pharmacological action in mice of several heterocyclic dicationic compounds that are biophysically characterized and selected from a large set of the library. We included the diamidines DB2115, DB2313, and DB1976 in these tests but looked at others as well. Using a quantitative analytical method to measure dication levels in biological samples, we found that the candidate compounds were significantly deposited in the bone marrow through our established assay. This allowed us to assess the abilities of the possible allosteric PU.1 inhibitors quantitatively. Further, the biological characteristics of these dicationic compounds were determined, allowing us to narrow down potential treatment compounds.

The rest of this chapter will explain and examine the three steps to develop a cell-based reporter screening system. The three steps include assessing cellular viability, PU.1 specific targeted genes' reporter assay, and subcellular compartment localization. Throughout this process, we documented the initial findings of the potency of heterocyclic compounds using the developed cell-based reporter platforms.

4.2 Materials and methods

4.2.1 Cell culture

The adherent cell lines, Human Embryonic Kidney (HEK) 293 and HEK293T, were cultured in DMEM medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 1% (v/v) Penicillin and Streptomycin (P/S). The HEK293 cells and HEK293T cells were maintained at 37°C in humid air with 5% CO₂. Before reaching 80%-90% confluence, cells were subcultured using 0.25% (v/v) trypsin. The trypsin was neutralized with cell culture media (above). The medium was refreshed every 2-3 days.

Additionally, human monocytic leukemia cell line, THP-1 cells were obtained from American Type Culture Collection (ATCC) and cultured in RPMI1640 medium supplemented with 10% (v/v) heated-inactivated FBS. These cells were maintained at 37°C under an atmosphere of 5% CO₂. Every 3-4 days they were maintained by replenishing the fresh culture media and keeping the confluency below 8×10^5 /mL in a T75 flask.

4.2.2 Plasmid construction

4.2.2.1 Lentivirus 3rd generation plasmids were obtained from Addgene.

pRRLSIN.cPPT.PGK.MCS.WPRE and pRRLSIN.cPPT.PGK.EGFP.WPRE (Addgene plasmid # 21316) was obtained. The lentiviral mammalian expression vector pRRL.SIN lentiviral Enhanced Green Fluorescence Protein (EGFP) gene was cloned. The EGFP protein expression is driven by a ubiquitous PGK promoter. The pRRL-SIN lentiviral vector requires Tat cDNA for lentiviral production and other packaging cDNAs (Gag/Pol, Rev, and VSV-G). Control lentiviral plasmid, human Phosphoglycerate Kinase (pgk)-EGFP lentivirus plasmid was purchased from Addgene. The 3rd generation lentiviral packaging plasmids pMDLg/pRRE (Addgene plasmid #12251), pRSV-Rev (Addgene plasmid #12253), and VSV-G envelope

expressing plasmid pMD2.G (Addgene plasmid #12259) were a gift from Didier Trono. The transfer plasmid, pRRL.SIN.cPPT.PGK.MCS.wPRE was used as a backbone, and the PU.1-specific reporter genes were inserted into the backbone plasmid. λ Bx5 (a synthetic enhancer element consisting of five tandem repeats of the λ B site, spaced one helical-turn apart and followed by a minimal TATA-box promoter) was inserted into the transfer plasmid. Likewise, the PU.1-specific promoter gene, *cd11b* (a native promoter of CD11b, MAC-1 receptor alpha chain, myeloid-cell-specific gene expressed late during the course of monocytic differentiation) and *csf1r* (native M-CSF receptor promoter in non-macrophage cells, which is transactivated by PU.1, Macrophage colony-stimulating factor, M-CSF) genes were cloned in the same method. DNA sequencing was conducted to confirm each reporter gene plasmid construct. Please refer to details already provided in the previous chapter.

4.2.3 *Compounds*

DB1976, DB2313, DB2115, and DB1476 were synthesized by Dr. Boykin's group. (Figure 4.2) The compounds were dissolved in either 100% dimethyl sulfoxide (DMSO) or water to prepare a concentrated stock. Depending on the compounds' structure, some were stored at 4°C; the rest were stored at -20°C. Additional dilutions were prepared in the culture medium. The compounds were added to the plates in 1 μ l volume for the desired final concentrations with a final DMSO concentration of 0.1% or 0.2% present in all samples.

4.2.4 *Cell viability assay: Sulforhodamine B (SRB) assay*

The sulforhodamine B (SRB) assay determines cell density based on the measurement of cellular protein content. This process involves treating cells with the test compounds and then adding pink aminoxanthine dye, sulforhodamine B (SRB), after an incubation period. This method is used for the cytotoxicity screening of compounds to adherent cells in the HEK293

cells (2×10^4 cells/well, 100 μ L) using a 96-well plate. Specifically, after two incubation periods (24 hours and 48 hours) after treatments of the compounds, cell monolayers were fixed with 10% (w/v) trichloroacetic acid and stained with the SRB for 30 min, after which the excess dye was removed by repeated washing with a 1% (v/v) acetic acid solution.

The SRB assay depends on the uptake of the negatively SRB by basic amino acids in the cells. Healthy cells will take up the pink SRB. Unhealthy cells will struggle to do so, leaving the surrounding area pinker. So, the number of pink cells is a proportionate display of cell health. Ultimately, cell density was determined when the protein-bound dye was dissolved in 10 mM Tris base solution for O.D. determination at 510 nm using a microplate reader (Molecular Devices). Cells that processed the test compounds successfully were healthier and were able to process the SRB. As a result, healthier cells showed a more intense color, indicating greater absorption abilities and demonstrating that the compounds were less toxic.

4.2.5 Fluorescence microscopic examination

The green fluorescent protein (GFP) could be observed using a fluorescence microscope (Leica DMI8). The PU.1 activity on the target reporter gene and its transactivation of fluorescence protein was measured 48 hours post-transfection.

4.2.6 Liquid Chromatography/Mass Spectrometry

4.2.6.1 Test sample conditions

Dr. Steidl's team from Albert Einstein College of Medicine prepared biological mouse serum samples. For DB2313, mouse (C57BL6) blood was collected 2 hours, 6 hours, and 24 hours after the IP injection (20 mg/kg of DB2313). C57BL6 mouse serum was commercially purchased and used to prepare the control samples. The samples were also injected with DB2373

as a control compound. This compound's results were used for internal reference in the mass spectrometry.

4.2.6.2 *Solid Phase Extraction*

Each serum sample was taken into new sterile e-tubes, and 2 μL of each concentration of DB2313 solution were added. Then the samples were mixed in pipettes. 2 μL of 5 pmol (2.5 μM) of the DB2373 compound solution were added as an internal reference to each test sample. The samples went through C18 Column extraction. After C18 column extraction, a 0.22 μm centrifugal filter was used to purify the collections, and then samples were analyzed with an LC-MS/MS.

For DB2313 extraction, we used the Solid Phase Extraction by C18 column procedure. To prepare C18 columns, the columns were tapped to settle the resin, then they were placed into a receiver tube, and 200 μL of activation solution (50% MeOH) was added into the rinsing walls of the spin column and to wet the resin. The columns were centrifuged at 1,500 x g for 1 minute; then, the flow-through was discarded. The centrifugation process was repeated, and the flow-through was again discarded. Then 200 μL of equilibration solution (0.5% TFA in 5% ACN) were added to complete preparation. After the C18 columns' preparation, they were vortexed for 30 seconds to make sure that they mixed well. Then the samples were centrifuged twice for 1 minute each at 1,500 x g. After the samples were collected in the receiver tube, 100 μL of washing solution (0.5% TFA in 50% ACN/ 0.5% TFA in ultra-qualified water) was added. Then the centrifuge was run at 1,500 x g for 1min. Then all of the collected samples were dried in the vacufuge. After drying, the samples should be stored in the freezer (-20°C) until they are processed for MS reading.

4.2.6.3 *Liquid Chromatography/Mass Spectrometry (LC/MS) Operating Conditions*

API-MS detection was accomplished using an AB SCIEX API 3200™ LC/MS/MS triple quadrupole mass spectrometer equipped with an orthogonal Turbo electrospray ion source and Agilent 1200 Series HPLC System. The samples were delivered to the ionization source by the autosampler of the Agilent 1200 HPLC system. A C18 column (3 μm particle size, 3×100 mm, Phenomenex, Torrance CA, USA) was used for the chromatographic separation. The injection volume was 5 μL. A generic chromatographic method was employed in this study.

The HPLC eluents were H₂O with 0.1% formic acid (A) and methanol with 0.1% formic acid (B). The initial gradient condition was 90% A and 10% B, holding on 10% B for 1 min, changing linearly to 60% B in 4 min, standing on 60% B for 2 min, then changing linearly to 90% B in 2 min, standing on 90% B for 2 min, decreasing linearly to 10% B within 1 min and equilibrating at 10% B for 5 min, giving a total analysis time of 17 min. After analysis, the column was washed with the above procedure, giving background chromatographic spectra. The eluent flow rate was 0.3 mL/min. The resulting data were processed using the Analyst v1.5 software from AB SCIEX.

All compounds were detected using positive ionization in multiple reaction monitoring (MRM) mode (ion pair 355/294 for DB2313). The analyte-dependent and ESI source parameters were selected to optimize the detection of DB2313 (ion pair 355/294). The parameters in the experiments were set as follows: ion spray voltage 4500 V, ion source gas GS1 45 psi, ion source gas 2 (GS2) 35 psi, ion source temperature was 450°C, with the interface heater turned ON. The declustering potential (DP) was optimized at 40 V and entrance potential (EP) was 6 V. Collision energy (CE) was 30 V and collision gas was 6 psi. The deflector was preset to 200 V and CEM

was operated at 2400 V. The dwell-time was 200 ms. The system was operated under the control of Analyst 1.5 software.

4.2.7 *Flow cytometric analysis*

We analyzed the intracellular EGFP expression using BD Fortessa; the FACS flow cytometer (BD Bioscience, San Diego, CA).

4.2.8 *Statistics*

Statistical Analyses were performed and represented using OriginPro 9.1 (Northampton, MA). All data using flow cytometry analysis were exported from B.D. Diva software and analyzed using FlowJo 10. Subcellular colocalization assay was performed by LAX2.2 (Leica) as described after collecting images with a Hamamatsu ORCA-AG CCD camera.

4.3 Results

Our team's research on the biophysical characterizations of PU.1-DNA complex inhibition and the potentiality of heterocyclic diamidines have previously been discussed. Surface Plasmon Resonance (SPR) analysis showed that PU.1 sequence-specific binding to the ETS domain could be inhibited using the minor groove-binding distamycin as a model compound; as a result, PU.1 binding is allosterically prevented and functionally inhibited. This was demonstrated using the λ B motif of the murine Ig λ 2-4 enhancer (5'-ATAAAAGGAAGTG-3') as a model (122). Munde and his colleagues screened several designed AT-targeting heterocyclic diamidines and proved that DB1976 efficiently competed with PU.1 for DNA binding at the localized target site, nuclei in live cells (300). On the foundation of the general structural features of our initial inhibitors, we designed and prepared a focused library of large derivatives for screening using the solution and cellular analytical method. From this library, DB2115 and DB2313 were selected for further testing.

4.3.1 *Biophysical characteristics of DB2115 and DB2313*

Both DB2115 and DB2313 illustrated outstanding characteristics for the AT recognition module. The two diamidines harbor the amidine-benzimidazole-phenyl platform that offers exceptional recognition for expanded AT sites. The idea for the units of the compounds is that the amidine cation imparts solubility, interacts favorably with phosphate groups that are close to the narrow DNA minor groove, and can form hydrogen bond (H-bond) with acceptor groups on AT base pairs at the floor of the DNA minor groove. The benzimidazole -N.H. points into the minor groove and can also H-bond with AT base pair acceptors. In addition, all three groups (amidine, benzimidazole, and phenyl) make van der Waals contacts with the walls of the minor groove and edges of AT base pairs at the floor of the groove. In addition to that, the compounds have the proper curvature to match the DNA minor groove in AT sequences.

Further, using molecular docked modeling, for instance, DB2313 was inserted in the DNA minor groove along the AT-track (5'-AAATAAAA-3') of the λ B motif, located upstream of the core consensus sequence (5'-GGAA-3') (301). In the model, the DB2313 bound well to the AT-tracks; this shows the compound's potential for further study. Taken together, DB2115 and DB2313, along with DB1976, showed proactivity for expanded and robust AT-rich sequence recognition; as a result, these initial compounds are the best choices to evaluate in more detail. The pharmacological control of the PU.1 transcription factors by the small heterocyclic diamidines has been proven using an integrated screening strategy utilizing biosensor SPR, DNA footprinting, cell-based dual-color PU.1 reporter assays, and ChIP assay in PU.1-low AML cells (301). As part of the published results of this dissertation, we present the highlighted results of DB2313, DB2115, and DB1976 (as a control) focusing on the functional cell-based assay, cell viabilities, subcellular colocalization, and pharmacokinetics. Further, several candidates of PU.1

inhibitors are presented to determine the level of heterocyclic diamidines in the biological samples, plasma, cell pellets, and bone marrow to expand the candidates' library.

4.3.2 Functional PU.1 inhibition by heterocyclic diamidines in live cells

With the biophysical basis of PU.1 inhibition by DB2313, DB2115, and DB1976, we proceeded to define their cellular properties and potential to inhibit PU.1-dependent gene transactivation by assessing the potential of these compounds to inhibit PU.1-specific gene transactivation. To evaluate the potential for functional inhibition of PU.1 transactivation by the three diamidines, we used the SPR screen to test the compounds' effects on the expression of a cellular EGFP reporter under the control of a minimal PU.1-dependent promoter harboring a triple tandem copy of the same λ B enhancer site (Figure 4.3).

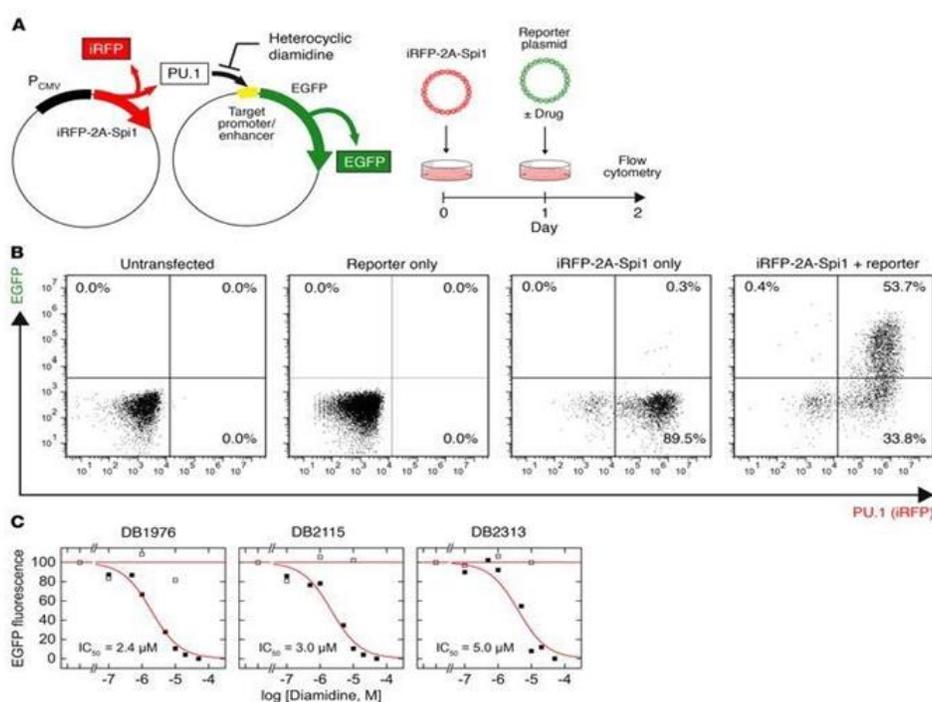


Figure 4.3 Functional PU.1 inhibition by heterocyclic diamidines.

Ectopic PU.1 activity (an iRFP reporter expression) in live HEK293 cells was detected through an expression of an EGFP marker. (A) The schematic diagram for characterizing ligands against PU.1 activity. Two plasmids were transfected with or without compound, under

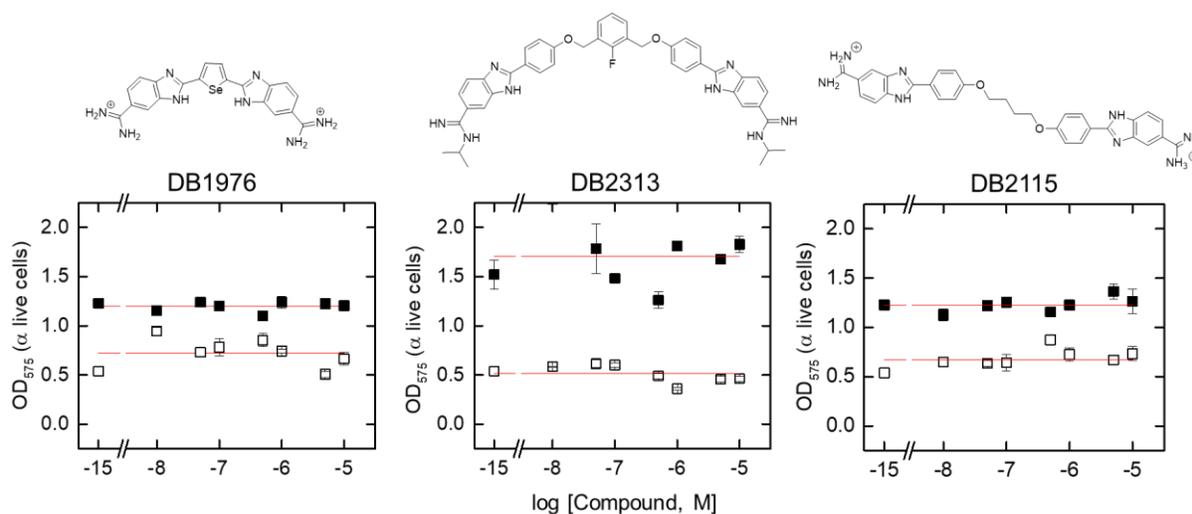
the control of a minimal PU.1-dependent, a tandem five copies of a λ B-based promoter. (B) Flow cytometric analysis of the PU.1 reporter (iRFP) and EGFP reporter, in which the upper right quadrant displays PU.1-dependent reporter activation. Notice the lack of EGFP reporter activation by endogenous non-PI.1 ETS paralogs, representing specificity of the λ B reporter. (C) Cellular EGFP fluorescence Inhibited by the compounds. Black symbols denote the λ B-based reporter, which was vulnerable to diamidine inhibition in a dose-dependent manner, and white symbols stand for the mutated SC1-based reporter, which was non-responsive. A Hill-type equation was used for fitting of the curve. The IC_{50} values for all heterocyclic diamidines compounds ranged between 2 μ M and 5 μ M (301).

This cell-based system was created such that ectopic PU.1 and reporter expression could be independently traced by fluorescence to separate out any background activation of the reporter. In non-PU.1-expressing HEK293 cells, which express several endogenous ETS paralogs such as ETS2, GABPA, and members of the ELF subfamily (Human Protein Atlas), transfection of the λ B-based reporter was silent in flow cytometry, indicating specificity of the λ B enhancer to PU.1 (Figure 4.3B). All three compounds inhibited PU.1-dependent transactivation of the reporter in a dose-dependent manner, with similar IC_{50} values between 2 and 5 μ M (Figure 4.3C). The relatively high IC_{50} values should be interpreted in light of the strong CMV promoter applied to make expressing ectopic PU.1.

To determine the specificity of diamidine action to the inhibition of PU.1, we assessed whether the diamidines could inhibit transactivation at an analogous minimal promoter in which the λ B enhancer sites were replaced by the SC1 sequence. Although PU.1 can bind the SC1 site, it is not found in native PU.1 target genes and is functionally nonspecific. Therefore, none of the diamidines significantly inhibited transactivation at the SC1-based enhancer (Figure 4.3C). This is consistent with the lack of PU.1 binding to the SC1 site we observed by SPR. Thus, heterocyclic diamidines appeared to maintain their strong AT selectivity in live cells and specifically inhibited PU.1 activity at functional DNA sites associated with PU.1 target genes.

4.3.3 Functional inhibition of *PU.1* by non-toxic heterocyclic diamidines

To prove that the decrease of EGFP fluorescence did not result from compound-induced cytotoxicity, we performed a cell viability assay by measuring the amount of bounded dye of SRB to cellular proteins under mildly acidic conditions, which could then be extracted using basing conditions. The SRB assay results were linear, with the number of cells and values for cellular protein measured by both the Lowry and Bradford assays at densities ranging from sparse subconfluence to multilayered supra confluence. The O.D. values correlated with total protein content from the treatment of the diamidines for 24 h and 48 h. All three heterocyclic diamidines did not show cytotoxicity either after 24 hours or 48 hours regardless of dose (all doses were in the 10^{-5} to 10^{-8} M range) (Figure 4.4).



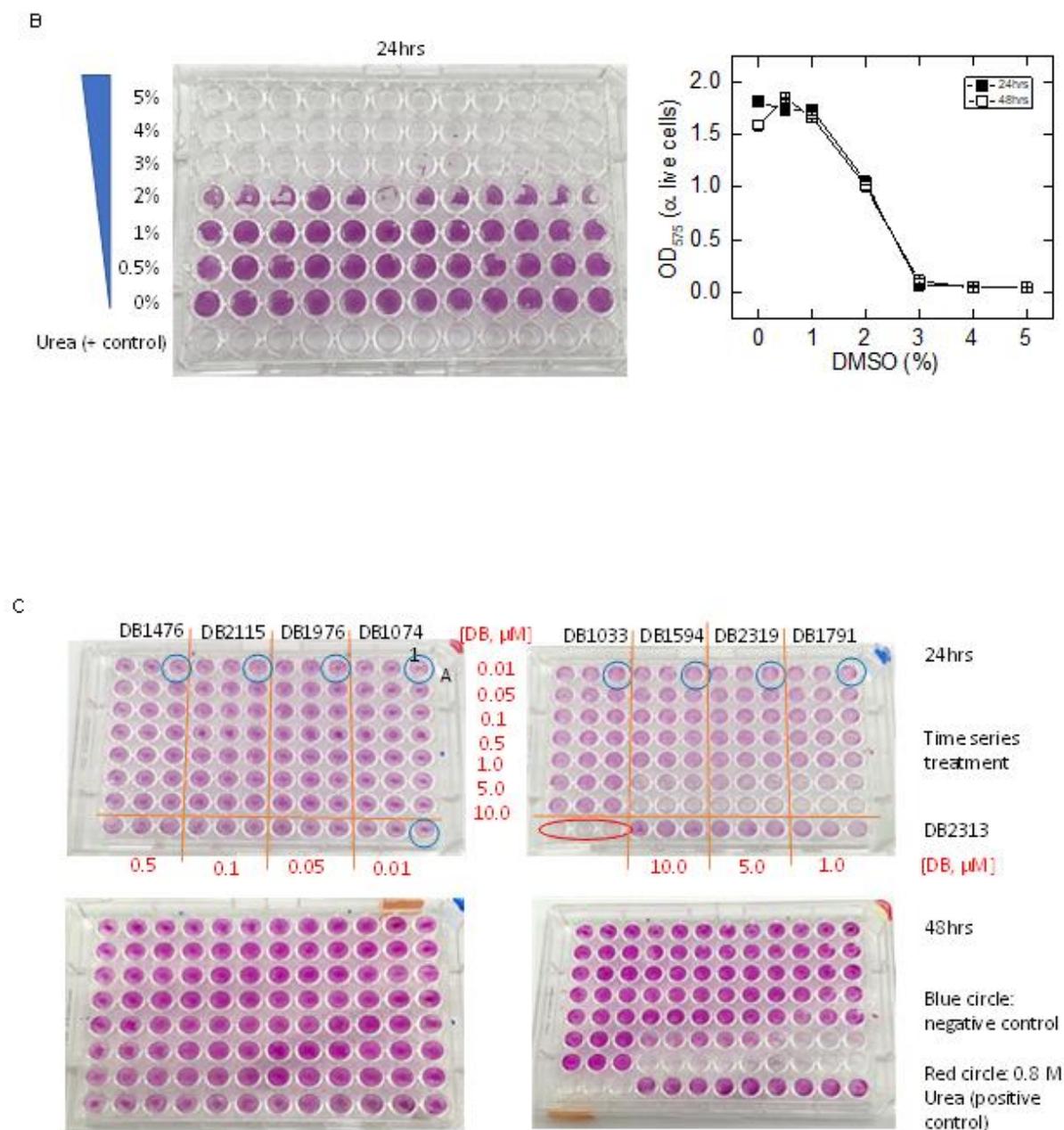


Figure 4.4 Cytotoxicity of all heterocyclic diamidines against PU.1 activity.

Sulforhodamine B colorimetric assay in HEK293 cells measured the compounds' toxicity based on the staining the dye intensity of Optical Density (575 nm), in which the dye stained the cellular protein mass of cultured cells to basic amino acids of cellular protein. The colorimetric examination provides an estimate of total protein mass, which is related to cell number. (A) All compounds were treated for 24 h (empty square) and 48 h (black square) in concentrations

ranged from 10^{-8} to 10^{-5} M. Each data point represents the mean \pm SE of triplicate or more samples. (B) Cytotoxicity of levels of DMSO in HEK293 cells were measured by SRB. As increasing level of DMSO, the O.D. value is close to zero ($n=12$). (C) SRB cytotoxic assay plates' image for samples' staining after various compounds treated displayed. As a positive control for the assay, 0.8M urea for 30 mins were used (Red circle). Blue circle denotes the nontreated HEK293 cells.

These results indicate that the reduction of the EGFP signal was not due to indirect effects related to cytotoxicity.

4.3.4 Nuclear uptake of the heterocyclic diamidines in live unfixed HEK 293 cells

To further understand the biological basis of these compounds' effects on target specificity in PU.1 inhibition among the compounds, we evaluated the cellular uptake of the compounds using fluorescence microscopy (Figure 4.5).

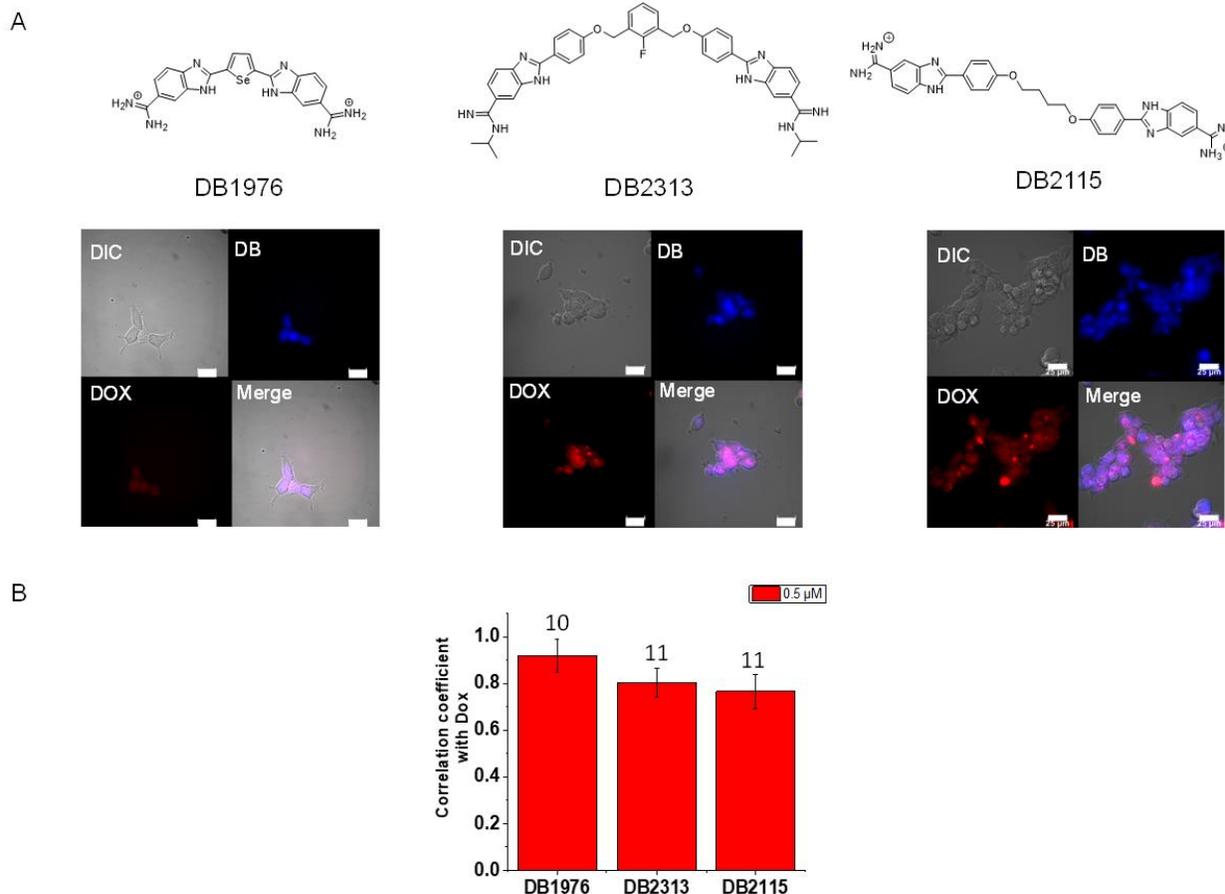


Figure 4.5 Subcellular localization of heterocyclic diamidines in live unfixed HEK 293 cells.

(A) The uptake of DB1976, DB2313, and DB2115 by HEK293 cells and colocalization with DOX were visualized by fluorescence microscopy as the details described in the method sections (Blue, 365/445 nm, Red, 570/630 nm) (B) Quantitative colocalization analysis of the compounds and DOX fluorescence in terms of special co-distribution. Each pixel of the fluorescence was calculated based on the Pearson correlation, indicating that two-colored pixels correlated well as closed to the value 1. All D.B.s were treated 0.5 μM for 24 h, and the representative images of the treated HEK293 cells were displayed.

We performed two types of subcellular compartment localization assays to compare our results to the previous results reported from Munde et al. In both our test and Munde's, the compounds absorbed 350-400 nm and emitted a blue fluorescence (84). HEK293 cells were

incubated with 0.5 μ M compounds for 24 hours at 37°C. Before visualization, the samples were incubated with 2 μ g/mL Doxorubicin (DOX), following the exact method used by Munde's group to highlight the presence of the cell nuclei. DOX was used to stain the DNA within the nucleus and spectrally separated it from our compounds. The compounds used (DB1976, DB2115, and DB2313) fluoresce as blue. Further, these three colors can be easily differentiated using fluorescence microscopy. Nuclear uptake of the heterocyclic diamidines was visualized in fluorescence microscopy and measure in terms of colocalization with DOX (Figure 4.5A). All three compounds co-distributed with DOX, indicative of nuclear selectivity. Quantitative analysis also exhibited that three compounds were deposited in nuclear with DOX marker strongly in signal intensity ($r > 0.83$).

4.3.5 *DB2313 in plasma levels decrease by 50% in 24 h*

After successful confirmation that DB1976, DB2115, and DB2313 inhibited PU.1 transactivation and deposited in the nucleus, the target site in live cells, our collaborators, Dr. Antony-Debre and Steidle's team in Albert Einstein College of Medicine, New York, investigated the potency of heterocyclic diamidines in a murine PU.1-low AML model (URE PU.1 $-/-$). Antony-Debre and her colleagues found that treatment with the three different compounds inhibited AML cell growth (301). More specifically, the compounds preferentially affected PU.1-low AML cells; the compounds decreased proliferation, lowered colony-forming capacity, increased apoptosis, and disrupted serial replating capacity of PU.1-low AML cells and a majority of primary AML cell samples.

Specifically, the researchers assessed the effects of PU.1 inhibitors on the function of *in vitro* self-renewal using a serial replating assay. For a long-term clonogenic capacity of PU.1 URE $^{-/-}$ AML cells, treatment with DB2313 led to a significant decrease in clonogenicity in the

second and third rounds of plating, and clonogenic capacity was wholly disrupted in the fourth and higher rounds of plating. The clonogenic capacity of DB2313 and DB2115 on primary human AML cells from 13 patients' samples showed more than 45% (DB2115) and 60% (DB2313) of a mean decrease compared with vehicle-treated cells. Further, the apoptosis-inducing effect of the two compounds was confirmed in PU.1 low induced AML cell lines (about an average 2-fold of increased activity). Dr. Steidle showed that transcript levels of *Csf1r* and *Junb*, which are positively regulated by PU.1 in myeloid cells, upon treatment of DB2115 and DB2313, decreased in PU.1 URE^{-/-} AML cells and increased *E2f1* expression, which PU.1 represses.

To provide insight into the molecular mechanisms mediating the anti-leukemic effects of pharmacological PU.1 inhibition, Dr. Steidl's teams used targeted ChIP and expression analysis in PU.1-low AML cells after treatment to show that the compounds directly interrupt PU.1 binding to chromatin of *Junb* and *Csf1r* in primary AML cell samples. Further, the treatments of PU.1 inhibitors in mouse and human AML (Xeno) transplantation models decreased tumor burden and resulted in increased survival.

Here, along with Dr. Steidle's promising results in preclinical AML models, we attempted to determine the levels of DB2313 in biological samples, including mouse serum and ascites. DB2313 was chosen because it is the lowest IC₅₀ and had the most negligible effect on wild-type cells. Plasma was collected 2 hours, 6 hours, and 24 hours after intraperitoneal administration (I.P.) to healthy mice (Figure 4.6).

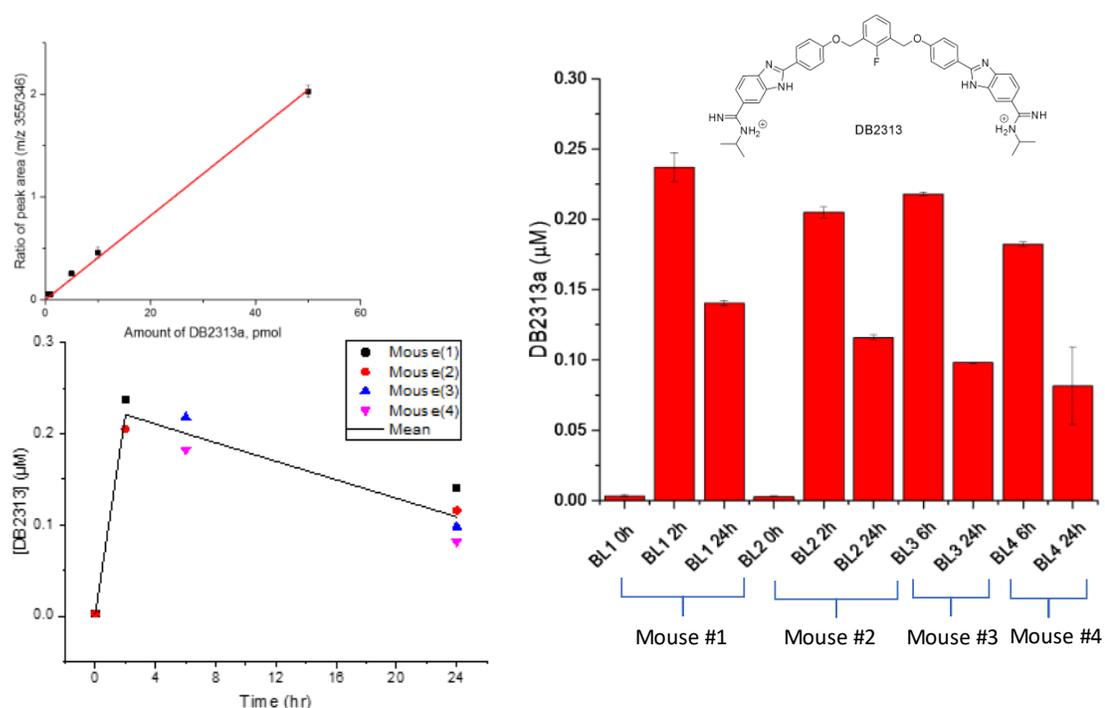


Figure 4.6 Levels of DB2313 in the mouse serum after intraperitoneal (I.P.) injection of wild-type mice at different time points ($n=4$).

Standard Curve of DB2313 (upper right panel), quantity measure of mouse samples treated by DB2313 (bottom right panel), and its PK profile (left panel) (301).

Mice were dosed with DB2313 at 20 mg/kg I.P. Mouse plasma concentrations for DB2313 were measured over 24 hours. In mice samples ($n = 4$), maximal DB2313 level reached peak concentration 2 hours after I.P. administration (mean of $C_{max} = 0.21 \mu\text{M}$). Supporting this finding, the two samples were taken 6 hours post-I.P. showed a similar or lower concentration than 2 hours post-IP. The maximum concentration remained for a few hours then started decaying in the plasma serum.

At 24 hours, all samples represented below half of the level in mice. Based on this experiment set, we cannot state definitively if this is a terminal elimination phase. DB2313 showed a terminal biological half-life of approximately 20 hours in mice. The low concentration of DB2313 in the serum might be caused by the injection method or result from the compound's

distribution to other tissues. Further research will be needed to identify the bioavailability, absorption, distribution, metabolism, and clearance of DB2313 in detail. Overall, the attempt of the heterocyclic diamidines to quantify levels of DB2313 in plasma enabled us to establish the quantification method and assess the effectiveness of drugs in biological samples at micromolar or lower concentrations with specific targets engagements.

Although there is no prototype drug similar to DB2313 with which we can compare the levels in plasma, a comparison between DB2313 and DB829 provides a starting framework (302, 303). DB829 is an extraordinary example of one of the heterocyclic diamidines that completely cured first- and second-stage HAT in animal models. DB289 showed higher systemic exposure, less tissue binding, and greater CNS penetration, contributing to its excellent *in vivo* efficacy in animal models of both first- and second-stage HAT. Contrary to DB829, DB2313 has proven a better candidate for AML cell model testing because of its fewer side effects. The detection levels of DB2313 in mouse serum were low in micromolar concentrations. Considering the potential of bioavailability of small organic compounds and the ability of PU.1 inhibition by heterocyclic diamidines including DB2313, DB2115, and DB1976 in animal models, we proceeded to detect possible PU.1 inhibitors that were already biophysically characterized from SPR assay. Therefore, we have developed further advanced cell-based reporter assays for characterizing screening ligands against PU.1 activity.

4.3.6 Functional inhibition of PU.1-DNA complex in endogenously PU.1 expressing cells using PU.1-dependent genes' reporter system

We have previously established the biophysical basis of PU.1 inhibition by our compounds; we aimed to define their cellular properties and potential to inhibit PU.1-dependent gene transactivation. We chose THP-1 as our AML disease-relevant model. The cell line was

established from AML patients. They natively express PU.1, allowing us to evaluate the compounds' efficacy in a relevant disease state without invasively monitoring PU.1 activity *in vitro*. Doing so provided a metric to the pharmacological effects of the compounds. To monitor the compounds' effects on-target PU.1-inhibitory activity, we have developed reporter assays for PU.1 (Figure 4.7).

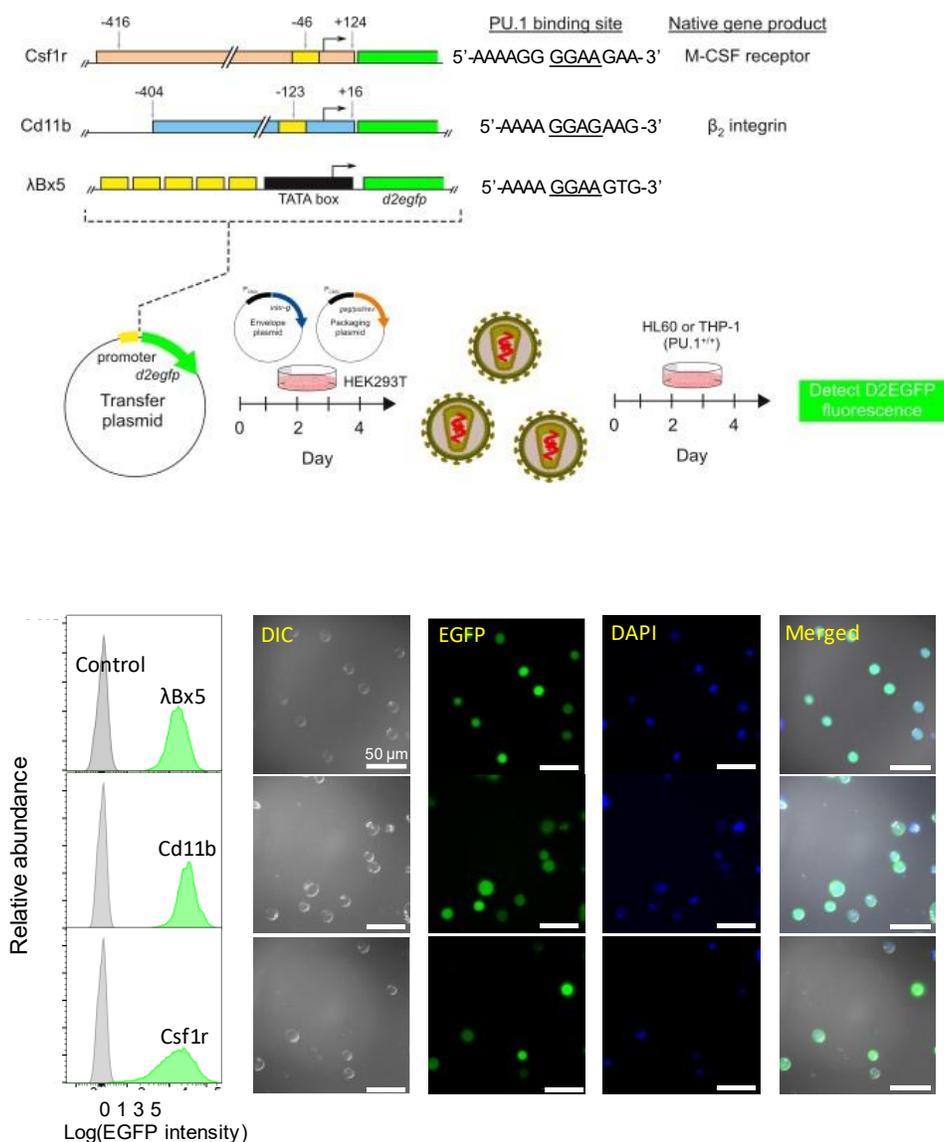


Figure 4.7 The second-generation PU.1-dependent reporter system for characterizing ligands against PU.1 activity.

(A) Schematic diagram of the lentiviral of the PU.1-specific reporter cell line generation. Each key native target binding site for PU.1 was tagged EGFP marker. After lentiviral expression plasmid and other envelope and packaging plasmids were co-transfected to the host 293T cells, the lentiviral particles containing each PU.1-specific gene were collected to transduce the target cell (THP-1, native PU.1 expressing cells). The expression of EGFP fluorescence under PU.1 binding on the promoter was measured by flow cytometry. After 7 days of transduction on THP-1 cells, only EGFP fluorescence cells were selected by limiting dilution, and the second generation stably expressing reporter cell lines were developed. (B) Confirmation of PU.1-dependent reporter in endogenously PU.1 expressing cell lines. THP-1, the human monocyte/macrophage model, was used as an experimental model. LV- λ Bx5-EGFP, LV-Cd11b-EGFP, and LV-Csf1r-EGFP in THP-1 were confirmed by flow cytometry and fluorescence microscopy.

Our previous reporter assay aimed to measure PU.1 activities on its target specific genes chemically transfected on HEK 293 cells, which PU.1 is not expressed. Two plasmids of PU.1 expressing plasmid tagged with iRFP fluorescence protein and the other plasmid with its target specific promoter gene tagged EGFP fluorescence protein was introduced into not PU.1 expressing cells from our previous attempt, HEK293. By monitoring its PU.1 transactivation modulated by the small compounds, we detected the two fluorescent proteins' expression by flow cytometer. This platform was adequate to investigate PU.1 transactivation modulated by our small compound inhibitors. However, the method was required to chemically transfect for expressing PU.1 and its specific enhancer genes to test the compounds every time; therefore, we developed PU.1 functional reporter assay to efficiently allocate time and resources for an extensive library of the candidates.

Further examination on monitoring native PU.1 activity in the AML disease-relevant model, THP-1 cells, we developed the "gen2" reporter assay (Figure 4.8A). Under native PU.1 activity, we inserted three PU.1-dependent gene constructs (λ Bx5, Csf1r, and Cd11b) into the genomic DNA of THP-1. These genes contain homologs of 5'-GGAA/G-3' of PU.1 binding sites tagged fluorescent protein EGFP. This design aimed to monitor PU.1 transactivation on their

target specific gene only by interrupting any other signaling in native PU.1 expressing cells, THP-1. The λ B motif of the Ig2-4 enhancer was previously known to a high-affinity PU.1 binding site with AT-rich tracks flanking both sides of the ETS consensus and used for our cell reporter assay to evaluate the AT sequence-specific binding compounds to PU.1 inhibition. Csf1r (receptor for macrophage colony-stimulating factor) was upregulated by PU.1 in the monocytic development and implicated with leukemic stem cells. According to Aikawa et al., cells expressing high amounts of CSF1R showed potent leukemia-initializing activity in the MOZ-TIF2 interacting with PU.1 and were known to target PU.1-mediated upregulation of CSF1R expression might be a functional therapeutic approach (277). In the previous report from Ileana and her colleagues, the mRNA level expression of Csf1r in URE^{-/-} AML cells was significantly decreased upon the response of DB2115, and DB2313, along with a direct effect on PU.1 transcriptional activity (301). Also, ChIP assay for checking PU.1 occupancy on Csf1r and other PU.1-regulated gene Junb and E2f1 promoters in AML cells decreased, indicating that the compounds interfering with PU.1 binding to chromatin *in vivo*. Therefore, the PU.1 inhibition of Csf1r also represents another pathway for anti-leukemia therapeutics. Cd11b is a target gene of PU.1. It is well known for immune cell surface markers, primarily implicated in various adhesive interactions of mature monocytes, macrophages, granulocytes, and natural killer cells. PU.1 binding plays a critical role in the myeloid-cell-specific expression of CD11b; a gene expressed late during monocytic differentiation. Therefore, the three reporter gene constructs enabled us to identify and determine our PU.1 inhibitor's sequence specificity to disturb the target inhibition. PU.1 also regulated the expression of Cd11b or Csf1r in the tissue-specific and developmentally manner; therefore, studying the markers' expression changes by our inhibitor could imply the effects on determinants in myeloid cells. For example, Antony-Debré *et al.* revealed that

DB1976, DB2115, and DB2313 were more impactful on mature granulocytic differentiation; therefore, our reporter constructs can directly monitor downstream effects by compounds.

By observing the EGFP expression response of PU.1- dependent Csf1r and Cd11b modulated through various dications, we can classify them by their target-specific activity and impacts on downstream cells. Our PU.1 inhibition assay allowed us to directly monitor EGFP expression dependent on PU.1 transactivation and the compounds' effects so that a broad set of PU.1 inhibitors can be screened in a time and cost-efficient manner.

Three promoter sites for PU.1 binding were placed into the lentivirus plasmid and then transduced into our experiment models (HL-60 and THP-1) to develop stable AML- relevant disease cell models with the desired reporters (Figure 4.7B). We successfully developed an experimental model on THP-1 showing fluorescence EGFP reporter by PU.1 transactivation (Figure 4.7B). Our empirical model was then designed to characterize cellular PU.1 inhibition by heterocyclic dications.

4.3.7 Functionally relevant outcomes of the small compounds in the new reporter assay

After constructing the stable cell lines for each PU.1-dependent reporter, we tested whether our cellular reporter assays are reliable for screening anti-leukemia therapeutics candidates (Figure 4.8).

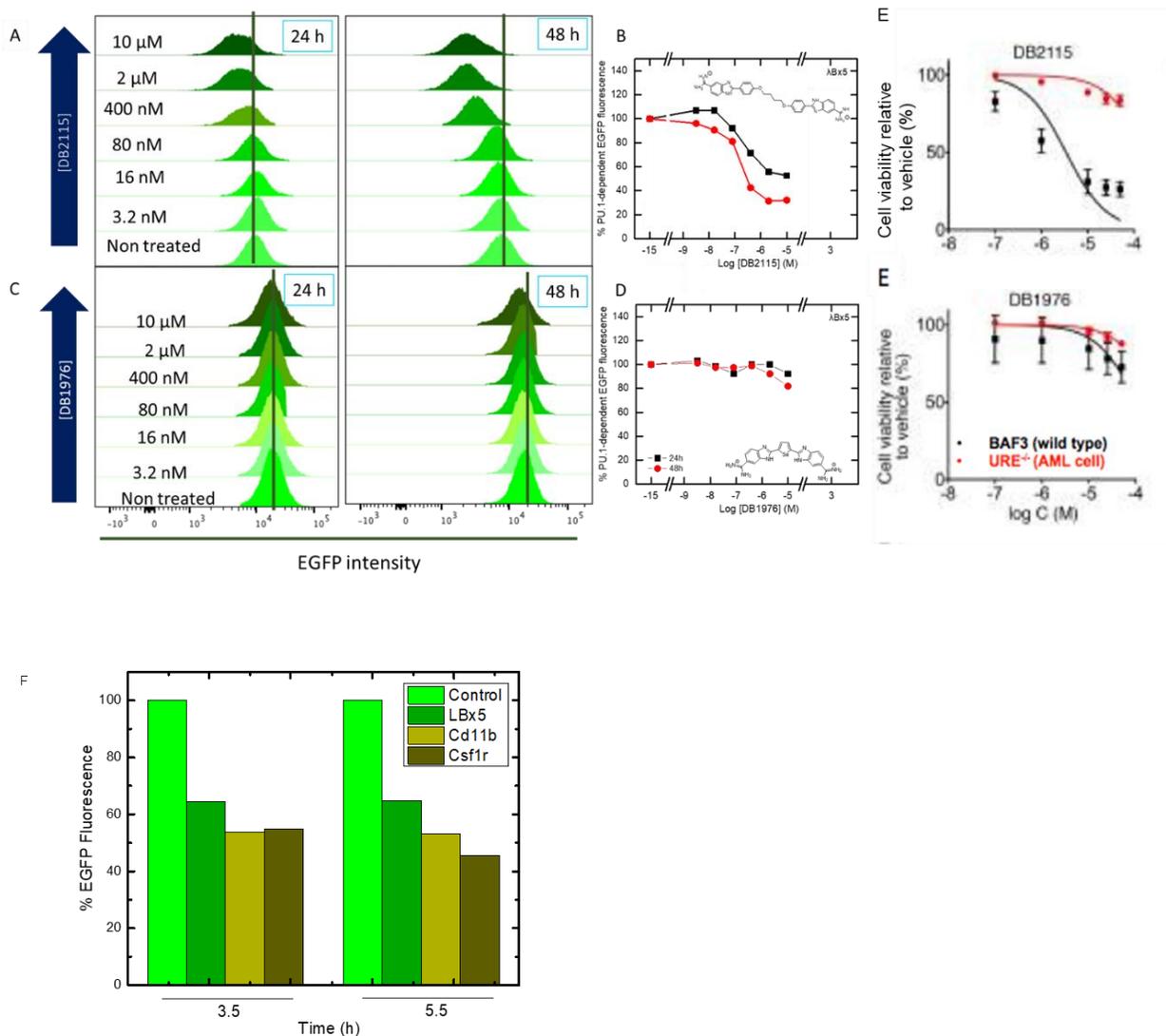


Figure 4.8 Functional characterization of PU.1 inhibition by DB2115 and DB1976 by cell reporter assay (THP-1 λ Bx5-D2-EGFP).

The previously known PU.1 inhibitor showed consistent results in their preclinical studies for the target PU.1 transactivation on cognate enhancer λ Bx5. (A, C) Flow cytometric histogram of EGFP expression changes to dose-response of heterocyclic diamidines. (B, D) Inhibition of cellular EGFP fluorescence by the compounds. (E) Cell viability of PU.1 URE^{-/-} AML cells and WT Bone marrow cells after treatment with different doses of small molecules (Antony-Debré et al.) (f) Fluorescence signal of EGFP expression of all Stable THP-1 cell lines for PU.1-dependent reporter gene were decreased by half around 5 h after treating 100 μ g/ml cycloheximide, a inhibitor of protein synthesis in eukaryotes (301).

To do that, we treated DB2115 and DB1976 with doses ranging from 10^{-9} M to 10^{-5} M to PU.1 target gene reporter, THP-1 $\lambda Bx5$ -D2-EGFP, and measured the potency of PU.1 inhibition after 24 h and 48 h. Compared to the previous results from Antony-Debré *et al.* (IC_{50} for DB2115: 3.4 μ M; IC_{50} for DB1976: 105 μ M) after 48 h, our result showed that IC_{50} for DB 2115 was 0.16 μ M and IC_{50} for DB1976 was 2.5 μ M (Figure 4.8B and D). Compared to the fractional PU.1 inhibition for the λB enhancer from biophysical studies (IC_{50} for DB2115: 2.3 nM; IC_{50} for DB1976: 18 nM), data from all these studies consistently showed for the order of magnitude in the potency of the two heterocyclic diamidines on PU.1 inhibition. Therefore, our cell-based reporter assay is a reliable screening system to define PU.1 inhibitors' cellular properties and the potential to inhibit PU.1-dependent gene transactivation.

4.3.8 Different extent of functional PU.1's inhibition of DB2115 and DB2313 on the native promoters

After functional test confirmation of the second-generation reporter platform (THP-1- $\lambda Bx5$ -D2EGFP), the other two reporter cell lines (THP-1 Cd11b-D2EGFP and THP-1 Csf1r-D2EGFP) were tested by DB2115 and DB1976 (Figure 4.9).

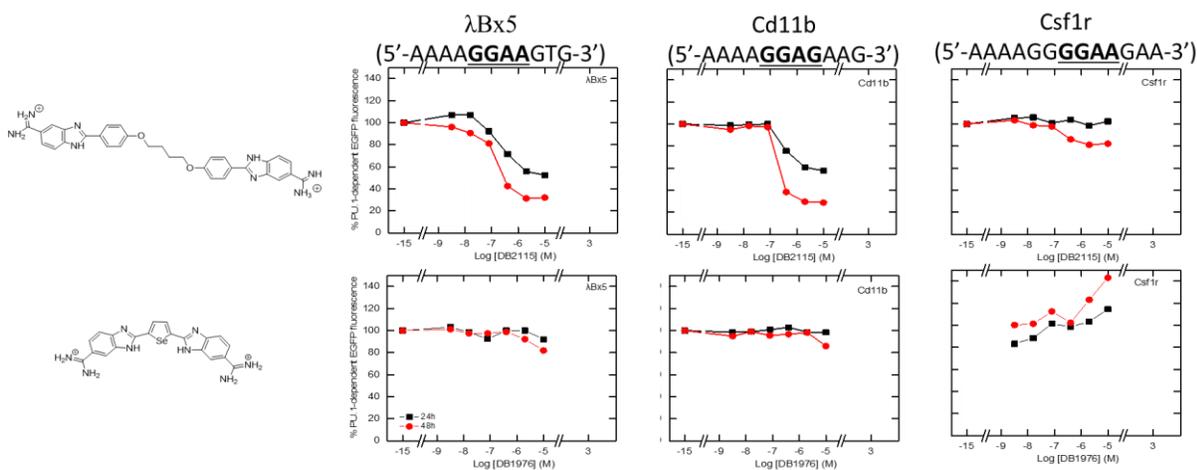


Figure 4.9 Inhibition of cellular EGFP fluorescence by DNA minor-groove binder.

Cell property and functional inhibition of DB2115 and DB1975 were determined by the PU.1-dependent enhancer reporter, λ Bx5, Cd11b, and Csf1r in THP-1 using flow cytometry. Data were expressed as percent inhibition of the ligands as median EGFP fluorescence intensity. The EGFP expression level upon the PU.1 binding decrease as PU.1 activity was interfering by heterocyclic diamidines in a dose-response manner. The median EGFP fluorescence intensity of changes was analyzed. The IC_{50} values for DB2115 on the λ Bx5 ($IC_{50}=0.24 \mu M$ for 24h and $0.16 \mu M$ for 48h), and Cd11b ranged between $0.2 \mu M$ and $0.3 \mu M$ after 24 h (red) and 48 h (black) treatment ($IC_{50}=0.35 \mu M$ for 24 h and $0.26 \mu M$ for 48 h). The black square denotes data for 24h and the red square for 48 h.

DB2115 functionally inhibited the PU.1 activity against the interaction with Cd11b; however, DB1976 did not effectively interfere with the PU.1-DNA complex for corresponding dose ranges. Interestingly, both of the compounds did not successfully inhibit PU.1 activity. From the results among different target DNA promoter sequences, the location of their target "AT-rich" DNA sequence may play an important role for heterocyclic dications to modulate PU.1 transactivation. PU.1 inhibitors that we tested were preferentially binding to AT-rich DNA sequences. They work targeting the AT-rich region near the PU.1 direct binding site. It leads to allosteric changes on sequence geometry when they indirectly interrupt PU.1 binding to its target specific DNA sequence at major groove at the 5'-GGAA-3' consensus and an additional requirement of a minor groove backbone. PU.1 is distinguishable from the paralog by the former's strong preference for AT-rich sequences upstream of the 5'-GGAA-3' consensus. DNA footprinting experiments verified that the compounds targeted the flanking sequences found in cognate PU.1 site, using the Ig λ B motif of the murine Ig λ 2-4 enhancer (5'-ATAAAAGGAAGTG-3') as a model. Based on the general structural features of our candidate PU.1 inhibitor, they were screened and selected for the test same enhancer sequence for cellular reporter assays. On Cd11b promoter (5'-AAAAGGAGAAG-3') has the same sequences as the

λ Bx5 enhancer sequence on the 5' site for heterocyclic cations to the binding. Therefore, the potency of the small molecules was shown consistent results on both. Accessibility of "AAAA" allows heterocyclic cations to bind on the minor groove and expect to interfere with the geometry configure for PU.1 binding on their target sequence to express genes so that EGFP fluorescence signal changes. However, the sequence differences on "Csf1r -EGFP (5'-AAAAGGGGAAGAA-3') may influence on the PU.1 inhibition by heterocyclic cations. Even though AT-rich sequences were provided for heterocyclic dications and their binding in flanking region of the minor groove, two sequences "GG." between the PU.1 target sequence '5-GGAA-3' and their compounds' binding site may not sufficiently enough to inhibit PU.1 transactivation. To prove that, we need to do a DNA footprint experiment.

Overall, heterocyclic cations action on PU.1 inhibition targeting its enhancer λ Bx5 and PU.1 native binding site Cd11b shows similar IC₅₀ values where they both contain AT-rich sequences in flanking sequence adjacent to PU.1 binding site. Contrary to that, another PU.1 native binding site, "csf1r," offers the same "A" rich sites to the compounds, the space from PU.1 binding site may play a role to show less efficiently PU.1 inhibition.

One primary reason for developing the stable cell line for PU.1-dependent gene reporter assay was intended to an intermediate screening system using a microplate reader and application for the suitable HTS system to characterize screening ligands against PU.1 activity. With this in our mind, we have attempted DB2115, which showed significant inhibition on PU.1 transactivation by a microplate reader (Figure 4.10).

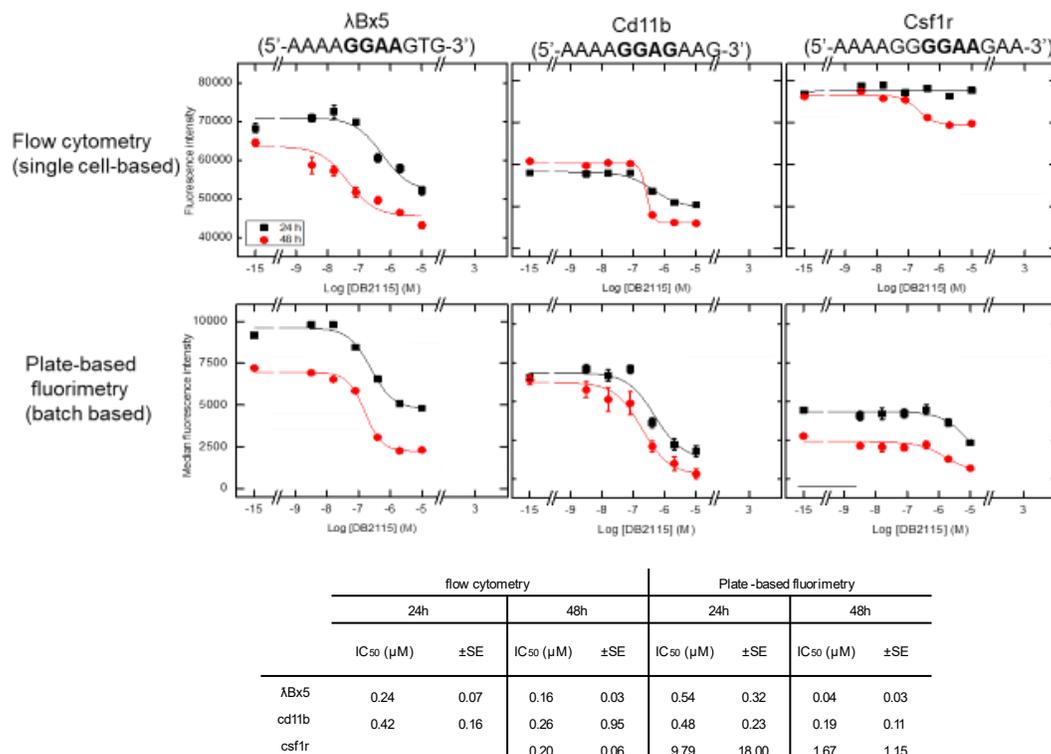


Figure 4.10 Comparison of two different techniques for measurement of functional characterization screening ligands against PU.1.

By using three THP-1 stable reporter systems, cellular property and functional inhibition of PU.1 by DB2115 were determined. Functional inhibition of PU.1 upon λ Bx5, Cd11b, and Csf1r in THP-1 by DB 2115 was determined by flow cytometer (top panel) and a microplate reader (middle panel). Data from the microplate reader are the mean \pm standard error of triplicate samples. On the bottle panel, the IC₅₀ was computed based on the Hill equation. Triplicate data of samples were obtained from microplate readers. The black square denotes data for 24h and the red square for 48 h.

We identified the suitability for adapting the cell-reporter assay by determining the PU.1 inhibition potency of DB2115. Similar outcomes of PU.1 inhibitor DB2115 obtained with the microplate reader from flow cytometric analysis (Figure 4.10A) were detected (Figure 4.10B). The effectiveness of PU.1 inhibition by DB2115 on λ Bx5 enhancer and Cd11b promoter were ranged from 0.1 μ M to 0.3 μ M from both measurements. The same inhibitory effects of DB2115 on the Csf1r promoter from flow cytometry and microplate reader were detected. The data is

preliminary results of the data; therefore, a larger set of tests with extended compounds and dose ranges is needed for the future.

4.4 Discussion

Small compounds that bind in the DNA minor groove regions have gained significant biotechnology and potential pharmacological applications. These include molecules that control gene expression by binding to only specific genes and interacting with and inhibiting the DNA-binding proteins such as transcription factors. Particularly interesting are the heterocyclic diamidines developed by Boykin, which bind sequence-specifically in the minor groove of DNA. DB75, a representative clinical example for HAT that binds preferentially targets AT-specific sequence, has opened a pathway for developing small molecules for various drug applications along with oral availability and ease of use compared with pentamidine (302-305). The concept of drug targeting TF-DNA complex was introduced for PU.1, whose favorable recognition sites could be a potential target by the DNA minor groove compounds. Munde et al. have reported that heterocyclic diamidines, such as DB1976, successfully targeted the PU.1 transactivation (84). The feasibility of the PU.1 inhibition by the small molecules with submicromolar activity without noticeable cytotoxic effects has led us to screen ligands against PU.1 activity. PU.1 specifically recognizes purine-rich DNA sequences containing 5'-GGAA/T-3' at the promoters of target genes. Compared to other ETS TF family members, PU.1 is more strongly selective for AT flanking sequence binding. Small heterocyclic dications that bind to the AT-rich minor grooves of DNA inhibit PU.1-specific DNA recognition at consensus sites harboring a 5'-GGAA/T-3' motif. Upon binding, the DNA minor groove binders, as allosteric inhibitors of TF-DNA complexes, downregulate PU.1 activity as they target gene promoters (specifically, the λ B motif of the Ig2-4 enhancer).

4.4.1 Feasibility of the DNA minor groove binders as PU.1 inhibitors

Small compounds such as DB2115 and DB2313 have proven to be PU.1 inhibitors along with DB1976 that compound has been previously reported. By our cell-based reporter assay in live cells, we provide the ligand's ability to inhibit PU.1 by binding the I γ B sequence first. All three compounds DB1976, DB2313, and DB2115 inhibited PU.1 activity with IC₅₀ in the sub- to the low-micromolar range (301). These values were 10²-fold higher than those of SPR of biophysical characterization, in which the IC₅₀ should be considered that the targeted DNA biologically are protected by proteins and other factors. None of the compounds tested represented detectable cytotoxic effects. This is a critical feature of compound biological activity that is often neglected. We note that our use of DOX as a nuclear counterstain was purposed at preserving the biological disposition of the test compounds by precluding potential artifacts of fixation, which requires the use of propidium (PI) on DNA binding ligands. The absence of cytotoxicity and perturbation on cellular function due to short-term (30 min.) exposure to DOX is well-established. Therefore, we consider our results to be biologically reliable.

We explored the dynamics and kinetics of heterocyclic diamidines as potential therapeutics of AML by inhibiting PU.1 transactivation and detecting diamidine levels in the biological samples. In our studies, 50% of Inhibition Concentration (IC₅₀) of DB2313 on the PU.1 activity in cell-based model, (5.38 μ M) is higher than serum concentration of DB2313 in mice. Compared to similar previous research, 50% of the Inhibition Concentration (IC₅₀) of DB829 is .02 μ M (303). According to Wang, a single dose of DB829 (20 mg/kg) administered intraperitoneally clears its target, parasites, from mouse blood within 2 to 5 days. It might also be due to low drug absorption after IP injection or distribution to other tissues. The biological half-life of DB2313 in the mouse serum is about 24 hours in our study. The gap in levels of

heterocyclic diamidines *in vivo* might be influenced by various factors, including dosage amounts, routine schedules of administration, dosage methods, drug distribution effects on a biological system, and sample collection method. Moreover, DB2313 might be absorbed by cells such as red blood cells in the blood; thus, it might not be detected in the biological fluids such as serum and plasma that we analyzed.

4.4.2 *Advanced cell-based platform for screening ligands against PU.1-DNA complex*

Using DB2115 and DB2313 as a control, we tested the efficiency of our advanced cell reporter system, the second-generation reporter assay in natively PU.1 expressing cells. With the need for an intermediate screening system for identifying candidates, a stable reporter expressing cell line has been established. The PU.1-dependent gene of Cd11b and Csf1r were added into the λ B enhancer and extended for the screening system for candidate searching efforts. The stable reporter cell lines were verified with the similar or consistent outcome of DB2115 with the first reporter assay. The data showed that DB2115 effectively inhibited PU.1 activity on PU.1-dependent gene, Ig λ B and Cd11b, but not Csf1r. The data suggested that the compounds preferentially bind to AT-rich sequences; however, the “GC” sequence nearby PU.1-binding cognate site in the Csf1r influenced their binding the target sites for perturbing PU.1 activity.

Cytotoxic assay of heterocyclic compounds allowed us to investigate compounds' toxicity profiles in the cells. We examined their cellular properties and the potential for functional PU.1 inhibition. Functional inhibition was provided by a cell-reporter assay that responded to native PU.1 transactivation at designated enhancer or native promoter sites. By measuring fluorescence protein changes of PU.1-dependent promoter gene expression modulated by the small compounds aided in explaining the potency of the compounds in the cell and the target specificity of PU.1 inhibitors.

These heterocyclic dications indicated that the compounds' structural features were tied to PU.1 inhibition and subcellular localization. From Munde's research, it was determined that single or multiple modifications enhance the binding affinity of the functional groups in heterocyclic cations on their target site in biosensor Surface Plasmon Resonance (SPR) assays. However, the compound may have differing effects in the target inhibition and distribution of PU.1 in live cells, dependent upon the structural differences of those cells. More interestingly, compounds with a difference of a single heteroatom (such as DB270 and DB1976) were correlated with contrasting inhibitory efficacy on their target transcription factor PU.1. These dicationic compounds demonstrate the importance of the structural and conformational heterogeneity concerning PU.1 inhibition. It is beneficial to determine the relationships between structure and activity with candidates' feasibility in live cells to provide direction to find more effective treatments for AML.

Additionally, subcellular routing was determined by quantitative colocalization with genetically encoded markers tagging with fluorescent proteins to distinguish the nucleus, and mitochondria for target specificity. It enabled us to examine their target specificity and correlated off-target activities. Further, the subcellular compartment colocalization offered us to understand the mechanisms driving the compounds' distribution and the cellular responses within the cell. Doing so provided insight into how these different classes of compounds behaved and localized within the cellular environment dictated by specific structural differences between them, including linkers of two dications, terminal dication's modifications, or even two isosteres. Specially, we classified the preferentiality of target specificity into a group of analogs among the compounds.

Our study attempted to provide proof-of-principle for ETS transcription factor PU.1 inhibition by heterocyclic diamidines with the intention of their serving as anti-leukemic agents. Our new approach for treatment is to target PU.1 because it is downregulated in most AML patients. By further inhibiting the levels of PU.1 pharmacologically, we can trigger the failure of the preleukemic stem cells and stimulate self-renewal in healthy hematopoietic stem cells. We presented the small compounds (including DB1976, DB2115, and DB2313) as first-in-class PU.1 inhibitors which successfully bound to the target-specific sequences and inhibited the PU.1 binding at the target sites via the allosteric mechanism. Our findings supported the alternative strategy of developing anti-leukemic drugs utilizing three small DNA minor groove binders. The compounds interfere with the PU.1 transactivation process effectively; further, this study enabled the process to be well characterized from biophysical research, providing an understanding of the cellular properties, characterizing the properties of PU.1 inhibitors, and suggesting avenues for the design of potential therapeutics.

The previous finding demonstrated the viability of transcription factor PU.1 as a pharmacological target and opened new possibilities for anti-leukemic therapeutics through small molecule inhibitors. The small compounds, which bind to DNA minor grooves and have a high affinity and specificity on AT-sequences in flanking regions, are attractive to develop as effective allosteric inhibitors of PU.1-DNA complexes (299). The ability of PU.1-dependent gene inhibition without significant cytotoxicity and nucleus localization in the live cell has potential for anti-leukemia therapeutics (84). This strategy proposes the complete loss of low PU.1-expressed leukemia stem cells using the DNA minor groove inhibitors because their removal might lead to leukemia stem cell failures (120). In an effort to find a variety of DNA-targeted transcription factor inhibitors, we need to develop a cell-based platform measuring the

feasibility of PU.1 inhibition. This dissertation attempts to evaluate DNA minor groove binders and understand their structure-function relationships. It is also to establish a quantitative analytical method to measure the dications' levels in biological samples. From these aims, the biological behavior characteristics of these dications are examined.

The current translation rate of successful molecules between biophysical studies and preclinical trials remains low. We hypothesize that structure-function investigations through a cell-based platform would enable us to couple fundamental structural properties of heterocyclic cations to their corresponding cellular properties. This would enable the design of a more efficient screening process for potential therapeutics by allowing the screening of a more abundant compound set. The result would be an ability to design more effective treatments with better PU.1 transactivation inhibition. This research also provides an understanding of the compounds' structural and functional properties to design more valuable compounds.

Using our cell-based assay, we monitored PU.1 transactivation's modulation as impacted by DB1976, DB2115, and DB2313. Then we used a dual-color reporter gene assay to determine the potency of each compound and its ability to cause PU.1 inhibition. We confirmed the compounds' effectiveness solely dependent on PU.1 transactivation and the compounds' target specificity in localization at nuclei in live cells. Our methods and cell-based approaches here provide our preliminary research. Our findings regarding the biophysical characteristics of the cell systems and tested compounds provide the basis for our future hypotheses. Especially, this research provides the foundational understanding necessary prior to moving toward animal models with which we can test AT-specific binding compounds in SPR research. The highlighted results from cell-based dual-color reporter assays provide knowledge of monitoring PU.1-dependent transactivation without any background activation of the reporters closer to PU.1 low

expression cell model: PU.1 (URE^{-/-} AML cells. The potency of DB2115 and DB2313 (IC₅₀ for DB2115: 3.0 μM < IC₅₀ for DB2313 = 5.0 μM) was consistent with the AML disease models (IC₅₀ for DB2115:3.4 μM < IC₅₀ for DB2313:7.1 μM) from Dr. Steidl's study (301). This shows DB2115 and DB2313 as the potential anti-leukemic agents, which we can now test further.

These small compounds target AT-rich sequences specifically but were also found to cause no toxicity in the cells. These findings suggested that the ability of PU.1 inhibition, as measured by the reduction of reporter gene expression, was not the cause of the compound's toxicity. DB2115 and DB2313 did not show the cytotoxic effect across the dosage ranges of 10⁻⁸ (M) and 10⁻⁵ (M) using SRB cell viability assay. Moreover, these results had consistently similar effects on wild-type cells. These findings confirmed slight or moderate effects on normal hematopoietic stem cells, supporting our hypothesis and proving that DNA minor groove compounds target preleukemic stem cells (PU.1 low expressed AML cells) without interfering with normal hematopoietic stem cell proliferation and self-renewal.

4.5 Conclusions

In this chapter, compounds capable of binding to small DNA minor groove binders to cause PU.1 inhibition were assessed for their potential as anti-leukemic agents. Aspects that were examined included the compounds' potency, cytotoxicity, and target specificity in live cells. Our findings were consistent with biophysical studies from Dr. Wilson's team of biophysicists who identified and screened possible compound candidates using biosensor-SPR. Also, our outcomes from the cellular properties of the PU.1 inhibitors were closely aligned with the preclinical studies of our collaborator, Dr. Steidl, whose team tested AML disease models. Moreover, we attempted to develop a cell-based platform to enable anti-leukemic agents and screen and investigate structure-function relationships of compounds to understand better and design drugs

to target DNA ligands. To accomplish this, we measured the inhibitory properties of heterocyclic dications on PU.1 inhibition *in vitro* and evaluated the compounds' respective potency to assist in developing future compounds.

Ultimately, our results supported proof-of-principles for transcription factor PU.1 inhibition as a possible target to AML therapeutics and that this can be pharmacologically achieved using DNA minor groove binders via an allosteric mechanism. This new alternative approach is supported by the effective PU.1 inhibition, the effective functionality of the heterocyclic diamidines that were selected from the biophysical compound library, the well-characterized cellular properties of the compounds' effects on cellular function, and the conformation of these compounds' functionality through preclinical AML disease models. Our findings confirm the potential of heterocyclic diamidines as a targeted therapeutic in AML through quantification on both a cell-based assay system and a liquid chromatography/mass spectrometry system. These findings open new therapeutic strategies to support the further development of PU.1 inhibitors as potential anti-leukemic agents. The next step in developing the potential treatment is to confirm these compounds' effectiveness through more cell-based testing.

5 CONCLUSION

A reporter gene technology is widely used for understanding how transcriptional activities can be achieved in the biological system. The reporter gene assay is very suitable for monitoring the TF activities by measuring the level-changes of reporter genes expression directly linked to interest genes (i.e., TF targeted promoters or enhancers) when the genes turn on or off. A reporter gene platform in live cells is helpful not only for the sensitive monitoring in a non-invasive manner and quantitatively detecting the TF activity but also is versatile to design the gene expression platform according to the proposed research questions and outcomes from the biological function of the genes' identification to the biophysical characterization of the transcription factor activity in cells. Ultimately, the cell-based reporter assay can provide a research bridge of knowledge between biophysical and biochemical research to biological studies. Therefore, we utilized these benefits of gene reporter platforms to characterize transcription factor PU.1 activity in live cells.

Our cell-based reporter platform is intended as an assay platform to characterize transcription factor-DNA interaction in live cells. We sought to understand the controlling function of the transcription factor related to its specific cognate sequence sites through reporter gene assays for identifying transcriptional activity. Our reporter assay focuses on the study of how transcription factors recognize their cognate sequence in high-affinity and specificity that has been suggested from biophysical and biochemical research to investigate the ideas in the context of the biological system in live cells. Even furthermore, from the knowledge of transcription factor-DNA interaction, the cell-based reporter platform is utilized for pharmacological application to target undruggable transcription factors. In this dissertation, we present the process of the development of a cell-based reporter platform and its various research

applications. This study outlines an approach to using reporter markers in probing functional characteristics of transcription factors in live cells at the TF-DNA level.

5.1 Significance of studying for TF-DNA interaction in live cells

During immune cell development, spatiotemporal regulation is essential for the generation of blood cell-type diversity. The identity of cell type is driven by combinatorial interaction of specific *trans*-acting factors, mainly transcription factors *cis*-regulatory element presents in noncoding DNA sequences within or near any given gene. In the context of transcription regulation by PU.1, as a representative model of studying TF-DNA interaction, a *trans*-acting factor is a regulatory protein that binds to specific DNA. The binding of a *trans*-acting factor to a *cis*-regulatory element in DNA can cause changes in transcriptional expression levels. The regulatory function that is formed by ETS domains represents that established *trans*-regulatory complex for ETS transcription factors. The ETS TF proteins form complexes, bind regulatory DNA sequences on enhancers and promoter regions, and help to recruit the basic transcriptional machinery to control the expression of genes. The interaction between TFs and DNA and form the complex are essential to regulate specific gene expression. One example to be known for the molecular mechanism of the regulation is that enhancer controls expression of the interferon-beta gene, which precise combination of TFs, and the orientation and spacing between TFs works in synergetic effects. Also, understanding how TFs recognize their DNA binding sites forms the basis for understanding transcriptional regulation and how this process involves disease occurrences, as numerous reports have been suggested.

It is a very complex aim because the TF-DNA binding preferences to gene expression are complicated, and TF activity and phenotype have lied with the complexity. We understand TF-

DNA binding, gene expression, and associated functional consequences for phenotypes toward these goals.

Combinatorial TF binding is essential for cell-type-specific gene regulations. The mechanisms of TF interactions are crucial for regulatory element activity, including constrained spacing and orientation of interacting TFs. Especially in the hematopoietic system, the TF mediate and decide the cell fate. The ETS family TFs PU.1 is a master regulator of transcription factors in innate and adaptive immune cell differentiation and has attracted more and more attention because of their research importance in biophysical, biochemical, and molecular biology and their application in medicine. The activity of PU.1 directly impacts the gene regulation in hematopoiesis and cell fates in the development of the immune system.

5.2 Highlights of research

We have been seeking to understand how the ETS transcription factors interact with its specific DNA. Although they all contain a core ETS binding domain, all 28 ETS proteins display various biological functions. We have reported critical findings from biophysical and biochemical properties to determine ETS TFs-DNA interactions with the coordination of cofactors and other cellular environmental factors in the biological system. For example, Ets-1, a structural homolog that is co-expressed with PU.1 during hematopoiesis, revealed remarkable different DNA-binding properties concerning interfacial hydration. Even though the ETS domains of PU.1 and Ets-1 are divergent in amino acid sequence, they fold into superimposable structures and bind their optimal DNA targets with similarly high affinity as PU.1 in the absence of osmotic stress. However, unlike PU.1, Ets-1 is minimally sensitive to osmotic pressure regardless of DNA sequence. As a functional consequence, we noticed that the two ETS relatives also differ in their levels of site selectivity. Analysis of DNA sequence motifs for the two

homologs by information theory shows that PU.1 is significantly more selective in target readout than Ets-1.

Furthermore, unlikely Ets-1 containing structural domains for autoinhibition to self-regulate for interaction with its cognate sequence to regulate gene expression, PU.1 has not reported the corresponding motif to regulate its DNA binding. Our biophysical and biochemical research reported PU.1's self-association to form a 2:1 complex with its high-affinity DNA binding site, and the data have suggested the 2:1 complex served as PU.1's self-inhibitory function. In connection with such vital findings, we have developed a versatile cell-based reporter platform to determine the suggesting ideas biologically relevant in the biological system context. Through our cell-based reporter assay, we enabled probing the functional characterization of PU.1 in perspective to interact with its cognate DNA sequence in high affinity and specificity. We pursue even further research application in the context of pharmacology; the characteristics of PU.1's preference for AT-rich sequence binding were utilized to develop an intermediate cell-based reporter platform to screen pharmacological agents against PU.1's binding activity in live cells.

5.2.1 Suitability for reporter assay in our study: providing the knowledge between biophysical experiments to biological functions

Our PU.1 functional reporter platform enabled us to monitor PU.1 activity in real-time detection of the gene expression at the single-cell level using a sensitive instrument, including a flow cytometer. Further, with the development of advanced technology in gene expression, we can observe the gene expression using optical modalities such as high magnification fluorescence microscopy.

Since we introduce exogenous DNA to study functional PU.1 activity into cells, we can utilize versatile our research interests. We were interested in the biophysical and structural characterization linked to the functional PU.1 property in a cell. From our previous research, we characterized one particular feature of the interfacial hydration attributes on PU.1-DNA binding complex among the ETS paralogs and its conservational importance in PU.1 activity using ETS domain mutant constructs fusing a reporter gene. Therefore, our functional data from the reporter gene platform directly supported our efforts to investigate the mechanism of PU.1 and other ETS TFs members' DNA recognition in biophysical characterization (e.g., MD simulation and fluorescence anisotropy) to the cellular environment.

Considering the strengths of various uses, we utilized our cell reporter assay to study transcription factor PU.1 expression and its transactivation measuring by two fluorescent proteins' expression. As a reporter technology, fluorescence finds its most significant potential by using negative feedback to characterize functional transcriptional activity control in PU.1 inhibition and control mechanism. Since the genetic reporter works as indicators to characterize PU.1 binding to its cognate sequence, we coupled reporter genes with a full-length PU.1 gene and PU.1-dependent enhancer elements.

For investigating biophysical and biochemical properties of ETS transcription factor PU.1, it is critical to establish experimental conditions in the agreement of structural biological studies and molecular cellular studies. Our biophysical experiment for PU.1 characterization was performed under the previously proven conditions for the thermodynamic characterization of the sequence-specific association of the DNA-binding domain (DBD) of murine PU.1 to the central 23 bp regions of the λ B site of the murine Ig λ B2-4 enhancer and conducted in the identical conditions (in phosphate buffer containing 150mM Na⁺, pH 7.4 at 25°C). From our previous

studies, various salt and pH conditions (with buffers featuring widely different heats of ionization) and the temperature dependence of KD remained the same under each of the conditions. We observed a remarkable insensitivity of the stability of the specific PU.1 DBD-DNA complex to temperature up to about 37°C and minimum changes around 50°C from previously reported. Therefore, we assumed that our experimental biophysical conditions would be identical to physiological conditions for our cellular reporter gene platforms, indicating that functional outcomes for characterization of PU.1 activity in live cells would support data from biophysical and biochemical studies.

In addition to the biological functional characterization of PU.1 activity, two signature biophysical properties of PU.1 activity that we found in target DNA recognition and intrinsic heterogeneity is biologically relevant events. Structural studies have demonstrated a highly homologous structural paradigm in sequence-specific binding by different ETS proteins; our previous results may find applications in understanding the thermodynamics of general site recognition by this family of transcription factors.

5.2.2 Versatile cell reporter platform for characterization of PU.1 activity in live cells

In this study, we have learned DNA-binding specificity as determinants of ETS transcription factor PU.1 for its interaction and activity. In the ETS binding motif, transcription factor PU.1 contains DNA target selectivity in the perspective of TF interacting and binding to its target cognate sequence at the protein/DNA level. Two intrinsic factors of heterogeneities, including molecular hydration and antagonistic dimeric state, were investigated using a cellular reporter gene platform. By utilizing the strength of versatility of the assay platform, application of dominant-negative mutants of PU.1 ETS domain and manipulation of enhancer syntax enable to measure functional PU.1 activity in live cells at the level of protein-DNA interaction. Finally,

functional outcomes and observations presented here will be valuable and reliable resources for broad research into biophysical and biochemical research to molecular biology. It would be possible to use the repertoire computationally to assess how the TFs themselves are regulated through transcriptional, post-transcriptional, and post-translational mechanisms to transform incoming signals into a regulatory output.

5.2.3 Dominant-negative reporter gene platform supported biological relevance to biophysical and biochemical research

Transcription factor PU.1 has a high sensitivity to osmotic pressures to recognize cognate DNA targets contrasting to Ets-1, the homolog structure of PU.1 in the ETS family. The interfacial hydration of PU.1 is unique among the ETS-DNA binding domain family of transcription factors and was proven by biophysical and biochemical molecular properties. Our reporter gene assay using dominant-negative inhibitors exhibited that the structural elements of PU.1 ETS domain for molecular hydration is essential for functional PU.1 interacting with the cognate target sequence as suggested outcomes and predictions from molecular dynamics and biophysical research. Also, our reporter gene platform supports chimeras as a rational approach to map evolutionary relationships of ETS transcription factors. The reporter gene assay, the critical residue 236 of the H3 recognition helix of PU.1 ETS domain that provides a sensitivity of PU.1 osmotically, exhibited an indicative residue of the evolution of the ETS domain. Molecular hydration of PU.1-DNA interaction is a representative example of the significance of interfacial hydration on DNA binding recognition by the transcription factor. Many biological examples of restriction enzymes or the *Escherichia coli* tryptophan repressor are iconic in perspective. The structural analysis by crystal structure and biochemical experiments in solution exhibited that it recognizes operator DNA primarily via water-mediated contacts. Hypo-osmotic stress in K562

leukemia cells was documented to motivate PU.1 to bind specific promoter sites in significant excess over Ets-1. Therefore, the differential interfacial hydration in DNA binding also establishes the two ETS members' roles in the cellular response to physiologic osmotic stress.

5.2.4 Various λ B synthetic reporter gene configurations were suited for monitoring the direct dosage effect on transactivation at the protein/DNA level

Our synthetic PU.1-dependent EGFP reporter platform was designed to characterize how inducible PU.1 expression in cells was governed at the TF-DNA level. To understand the functional nature of the 2:1 complex, the reporter gene expression plasmid is comprised only of tandem copies of enhancer elements, λ B, with two helical turns in each consecutive site. To do so, we monitored whether EGFP reporter signals represented a bell-shaped dose-response as the various enhancer reached saturated signals with non-productively bound PU.1. As we detected, the reporter signal that exhibited dose-dependently settled at a saturated level, depending on the level at which the 2:1 complex retained activity relative to the 1:1 complex without interruption of other cofactors or invasive manner since the PU.1 activity was monitored in non-PU.1 expressing cells, HEK293. Therefore, the effect of enhancer syntax structure in density and spacing demonstrated negative feedback in PU.1 transactivation in agreement with self-titration of the transcriptionally active 1:1 complex by an inactive dimer bound to DNA.

5.2.5 Enhancer (promoter) activity reporter assays for characterizing the functional consequences of TF binding

In parallel with emerging knowledge on feature modulate TF recognition, naturally occurring genetic variants have been informative in assessing the functional roles of TF binding site motifs. Many disease-associated variants that disrupt or introduce TF binding site motifs have been studied in detail, providing mechanistic insights into pathogenesis. Over 70% of the

thousands of noncoding variants found to be associated with common diseases or traits in genome-wide association studies overlap TF binding motifs' inaccessible chromatin. Integration of genotype information with chromatin accessibility, ChIP-seq, and gene expression data has begun to link motif-disrupting variants with altered TF binding and target gene expression. A more accurate and complete TF binding motif will be helpful in facilitating the identification and prioritization of damaging TF binding motif.

Our synthetic enhancer reporter gene assay has linked variation in TF binding sites with changes in reporter gene expression by simultaneously measuring the activity of synthetic *cis*-regulatory elements in a particular tissue or cell type. The real-time monitoring of PU.1's activity depending on the variants of the PU.1-dependent reporter system revealed that the functional relevance of PU.1's self-regulation in a negative feedback manner at the TF-DNA level in live cells. The enhancer reporter gene assay has allowed for the targeted characterization of TF and TF binding motif to obtain our study outcome.

5.2.6 Cell-based platform to screening pharmacological compounds with detection sensitivity and reproducibility

The cell reporter assay offers an effective method of *in vitro* screening for ligands targeting TFs for therapeutics. Considering the importance of PU.1 as a potential target of therapeutics (e.g., AML), our cell-based reporter system focusing on the interaction of PU.1 and its specific gene provides a rapid, sensitive, and reproducible detection system of screening ligands against PU.1 activity.

Our cell-based reporter platform is purposed for an intermediate screening before *in vivo* studies to correlate the structural properties of the compounds with their cellular behavior. Using a tool of studying a drug's mechanism of action in live cells relevant to our disease model will

elucidate fundamental knowledge of drug design. The *in vitro* nature of the screening platform provides a biologically and physiologically relevant environment while reducing the demands for *in vivo* tests. These platforms aim to deliver vital data regarding functional target inhibition, which can then be used to direct compound design. By directly monitoring PU.1 transactivation as affected by compounds, our screening assay will evaluate the potency or efficacy of the selected compounds. Measuring PU.1 dependent gene expression and downstream molecular events provides further insight into the effects of the compounds. Furthermore, we will suggest a mechanical system to investigate the compounds' behavior inside live cells. Overall, our cell-based assay enables us to screen candidate compounds in a biologically relevant setting while providing insight into the structure-activity relationship of the compounds.

Compounds such as DB1976 and DB2313 have proven to be PU.1 inhibitors as a pioneer compound of AML therapeutics and verified the concept of proof. The idea to target PU.1 low expressing AML cells and further PU.1 inhibition to trigger the function of self-renewal in hematopoiesis has been proven by the first-in-class compounds. Through our functional cell-based reporter assay, sequence-specific compounds heterocyclic diamidines targeting PU.1 binding sites enable us to bind the PU.1 specific sequence and inhibit PU.1 transactivation via an allosteric mechanism. We further developed a stable reporter assay using native PU.1 expressing cell line and tested the system using the first-in-class compounds as a control to test the efficiency of our system. Cytotoxic assay of heterocyclic compounds allowed us to investigate compounds' toxicity profiles in the cells. We examined their cellular properties and the potential for functional PU.1 inhibition. Functional inhibition was provided by cell- reporter assay that responded to naive PU.1 transactivation at designated enhancer or native promoter sites. By measuring fluorescence protein changes of PU.1- dependent promoter gene expression

modulated by the small compounds aided in explaining the potency of the compounds in the cell and the target specificity of PU.1 inhibitors.

Additionally, subcellular routing was determined by quantitative colocalization with genetically encoded markers tagging with fluorescent proteins to distinguish the nucleus (including the nucleolus) and mitochondria for target specificity. It enabled us to examine their target specificity and correlated off-target activities. Further, the subcellular compartment colocalization offered us to understand the mechanisms driving the compounds' distribution and the cellular responses within the cell. Doing so provided insight into how these different classes of compounds behaved and localized within the cellular environment dictated by specific structural differences between them, including linkers of two dications, terminal dication's modifications, or even two isosteres. Specially, we classified the preferentiality of target specificity into a group of analogs among the compounds.

5.3 Prospects

5.3.1 *Practical recommendations*

A cell-based reporter platform has developed to characterize transcription factor and DNA interaction in live cells. The multiplex of the integral fluorescence systems (piRFP-2A-PU.1 and p λ Bx5-d2EGFP) was designed to evaluate the functional PU.1 transactivation. The reporter system could be improved in practical choices of the reporter marker. For example, due to the spectra of iRFP, examination by eyepieces of fluorescence microscope often be obstacles to track its expression. iRFP proteins are primarily developed for improving the detection sensitivity *in vivo*. The temporal expression of iRFP as an indicator for monitoring dynamics of transcription factor interacting on its cognate DNA sequence might not be the best choice of the marker. Also, considering the half-life of iRFP (up to 31 h), we might measure the half-life of

PU.1 in our cell-based reporter system. Because we measure iRFP protein as an indicator of PU.1 expression, the half-life of iRFP might not reflect and capture the whole rapid dynamics of PU.1 activity.

In order to identify and characterize screening ligands against PU.1 activity, we developed the lentiviral delivery system to construct a stable enhancer reporter expressing cell lines using the myeloid cell line, which THP-1 cells constitutively express PU.1. Although THP-1 is derived from a myeloid leukemia patient and has many significant advantages for studying PU.1-associated disease and molecular mechanisms to differentiate macrophage, mRNA abundance of PU.1 is relatively high compared to other cell lines such as MOLM-13 or K562. This might lead to interpret the potency of DNA minor groove binders to target PU.1 binding sites as downgraded. Our targets for the AML therapeutics target downregulated PU.1 expressing cells and completely functional inhibition of PU.1; the results might be under evaluation efficacy of the compounds. Therefore, further validate its feasibility is needed to improve our cell reporter gene platform in the relatively high expressing PU.1 cell lines. Also, we are considering further development of the high throughput platform; therefore, more intensive validation of the reporter platform's performance in the microplate and flow cytometry process should be followed with other candidates of compounds such as DB2313.

5.3.2 Future directions and applications

5.3.2.1 Characterization PU.1-iRF4-DNA interaction

Future work with the design of expression vectors using different fluorescence color to measure PU.1 activity on other cofactors, such as Irf4- binding interaction with cofactors, will also provide fruitful knowledge for understanding normal and pathological cell behavior-specific

contexts. Auto-inhibition of PU.1 and other cofactors association effects on binding also can be applied and evaluated through a cell-based reporter platform.

5.3.2.2 *Characterization PU.1's transactivation under osmotic conditions*

By use of cell reporter platform, reporter gene expression under osmotic sensitivity can be determined by the expression of the iRFP reporter fused with transient PU.1 transfected cell under osmotic stressed conditions (Hypo-osmotic stress vs. hyperosmotic stress) or how environmental changes such as osmolytes affects PU.1' activity on its cognate site. Since the biological importance of transcription factors' response to the osmotic stress has been highlighted, such insights will be valuable for understanding TF-DNA interaction at the protein-DNA level and the influence of macromolecules on gene regulation to determine cell fates.

5.3.2.3 *High Throughput System (HTS) development*

HTS screening has been widely applied to evaluate diverse compound libraries for rapid screening. However, maximizing throughput has not directly translated into an expected increase in new drug approvals, primarily due to biochemical screening assays being unable to reflect the living system's complexity. Thus, more sophisticated biological assays are needed to evaluate compound libraries' biological activity and potential toxicity accurately. To address this challenge, cell-based assays have been increasingly used in HTS campaigns. *In vitro*, cell-culture models can provide substantive information on various cellular responses from exposure to a compound. Integrating advanced microscopy with cell-based assay is attractive because cellular events can be monitored with a spatiotemporal resolution by visualization. This has facilitated the use of fluorescence-based reporter assays in HTS campaigns to measure multiple cellular parameters and define relationships among molecular target activities, subcellular localization, and morphologic features. In addition to microscopy, flow cytometry can be used to measure

multiple events within single cells simultaneously. It is challenging to image individual cells. For example, detailed cellular processes such as subcellular localization of proteins are unobservable. To overcome this limitation, an imaging flow cytometry method has been described that combines the precise electronic tracking of moving cells associated with flow cytometry with a high-resolution multispectral imaging system.

5.3.2.4 Hit identification for antileukemic agents

Compounds such as DB1976, DB2115, and DB2313 have shown to be PU.1 inhibitors; therefore, we are using them as a control to test the efficiency of our system. Additionally, there are 16 more compounds we are examining for PU.1 inhibition. They all have been screened on PU.1 inhibition by SPR (Surface Plasmon Resonance) assay. We will test their cellular properties and the potential for functional inhibition. Functional inhibition is provided by cell reporter assay that responds to PU.1 transactivation at designated enhancer or native promoter sites. By measuring fluorescence protein changes of PU.1- dependent promoter gene expression modulated by the small compounds, it will explain the potency of the compounds in the cell and the target specificity of PU.1 inhibitors. By utilizing the intrinsic fluorescence of these compounds, it is possible to measure cellular uptake by flow cytometry. This assay will allow us to elucidate the chemically relevant moieties which facilitate cell entry.

5.3.2.5 Structure-Function relationship of ligands for design compounds

Additionally, subcellular routing is determined by quantitative colocalization with genetically encoded markers to distinguish the nucleus (including the nucleolus), mitochondria, endoplasmic reticulum, and the plasma membrane for target specificity. It will enable us to examine their target specificity and understand the mechanisms driving the distribution of the compounds and the cellular responses within the cell. Doing so will provide insight into how

these different classes of compounds behave and localize within the cellular environment as dictated by specific structural differences between them, including linkers of two dications, terminal dication's modifications, or even two isosteres. Specially, we will classify the preferentiality of target specificity into a group of analogs among the compounds.

5.3.2.6 Downstream effects of PU.1 transactivation to determine cell fates by compounds

HL-60 and THP-1 are well known as in vitro study models for cellular and molecular events involved in the proliferation and differentiation of granulocytes/monocytes/macrophages. HL-60 is stimulated by retinoic acid and induced granulocytic differentiation in vitro. Also, THP-1 can be differentiated by stimulation of PMA treatment into macrophage-like cells and valuable model for studying the mechanism of macrophage differentiation. Therefore, this approach can see how heterocyclic dication impacts the differentiation process on our stable reporter expressing cell lines. According to Antony-Debré et al., heterocyclic diamidines primarily inhibit later stages of granulocytic differentiation. Our disease models can be differentiated from other immune cells such as macrophages or neutrophils by chemical induction (PMA, Phorbol 12-myristate 13-acetate); hence, it allows to explore PU.1 inhibitors' effect on more mature, differentiating cells, whether dications impact differentiation or any changes of cellular responses in the differentiated macrophages or neutrophils, which related to PU.1 transactivation. To test the effects of heterocyclic dications for PU.1 transactivation at the stage on macrophage-like cells, we can treat our inhibitors on the cells before PMA induced and after PMA induced. If our inhibitor affects the differentiation, the phenotypical changes after their treatment and induction could not occur. This allows us to determine whether our compounds modulate PU.1 transactivation and involve differentiation. It leads us to know whether our inhibitors' effects on downstream cascade changes or not.

Synthetic reporter assays are increasingly employed to study complex biological systems. Understanding the regulation of transcription factors to bind its cognate site is a powerful model to study disease onset and progression. Using reporter gene technology, transgenes can be integrated into the gene loci of developmental marker genes. As a Result, cells with reporter gene expression can be selected or sorted to monitor the transcription factor's activity. In our opinion, modulation of PU.1's transactivation provides a possible strategy that withholds an excellent potential for future development of targeted therapeutic approaches either by the impact on cellular function, cellular fate determination, and metabolism but by its ability to influence it transcription. Ultimately, utilization of multimodality and multiplex of live cell-based reporter platform will extend our knowledge for PU.1 basis of DNA affinity by sequence and its unique feature of DNA recognition along with pharmacological applications.

APPENDIX

Signatures of DNA target selectivity by ETS transcription factors

This work is a reprint of a review that appeared as Poon, G.M.K. and Kim, H.M. (2017) *Transcription*. **8**: 193-2. Permission is granted by the publisher, Taylor & Francis Online, to reproduce the text in the co-author's dissertation.

ABSTRACT

The ETS family of transcription factors is a functionally heterogeneous group of gene regulators that share a structurally conserved, eponymous DNA-binding domain. DNA target specificity derives from combinatorial interactions with other proteins as well as intrinsic heterogeneity among ETS domains. Emerging evidence suggests molecular hydration as a fundamental feature that defines the intrinsic heterogeneity in DNA target selection and susceptibility to epigenetic DNA modifications. This perspective invokes novel hypotheses in the regulation of ETS proteins in physiologic osmotic stress, their pioneering potential in heterochromatin, and the effects of passive and pharmacologic DNA demethylation on ETS regulation.

Keywords: ETS transcription factors, PU.1, Ets-1, molecular hydration, phylogeny, epigenetics, DNA methylation

The ETS family of transcription factors binds site-specific DNA via eponymous, structurally conserved DNA-binding domains that share low overall sequence homology [**Figures 1A and 1B**]. Although ETS members are not numerous (28 paralogs in humans), they are ubiquitously distributed in the metazoan (306), and most are indispensable to life. Regardless of function, all ETS proteins show a highly conserved binding mode in which a recognition helix of their ~80-residue ETS domain is inserted into the major groove of target DNA harboring the core consensus 5'-GGAA/T-3', with additional interactions along the DNA backbone at flanking, sequence-variable minor groove positions [**Figure 1C**]. The structural homology among ETS domains is remarkable in the context of the choreography many ETS transcription factors execute in hematopoiesis, the multi-step differentiation of blood cell lineages that is intricately controlled at the transcriptional level (222, 307). Differentiation of the hematopoietic stem cell and fate determination of downstream progenitors is driven by precise ebbs and flows of activity by ETS paralogs in conjunction with other transcription factors in a stage-specific and dosage-specific manner (308).

Of the hematopoietic ETS-family regulators, PU.1 (Spi-1) and Ets-1 draw one of the sharpest lines of contrasts [**Table 1**]. Their ETS domains represent the most phylogenetically distant ETS sub-families, yet they are structurally superimposable when bound to DNA targets (309). The two ETS paralogs drive cell fate specification coordinately, and often in opposing directions (310-315). In T-cell differentiation, for example, an obligatory drop in PU.1 activity is concomitantly balanced by a surge in the Ets-1 activity (315-317), and both are differentially required in the peripheral Th subsets (318-320). Aberrant activity in either paralog is linked to a spectrum of diseases including rheumatism (321), cancer (322-325) and Alzheimer's disease (326-328). Functional heterogeneity occurs even among very close ETS sequence homologs, as in the

case of PU.1 and Spi-B, wherein Spi-B cannot compensate for the absence of PU.1-mediated B-cell signaling in PU.1-null mice (329).

Specificity determinants of ETS transcription factors

Given the significant overlap in expression of ETS proteins,(330) their DNA sequence preference (236), and overall structure of their ETS domains on the one hand, and their general functional non-interchangeability on the other, the basis of their specificity has long been a subject of major interest. Currently, ETS proteins are grouped into classes (I to IV) according to their relative sequence preferences (236), which correspond to their phylogenetic relatedness. The sharp and well-conserved delineation of major groove contacts at the 5'-GGAA/T-3' core consensus, and minor groove backbone (sugar and phosphate) contacts at flanking bases where sequence variation occurs, has led to the notion of indirect readout. In contrast with direct readout of nucleobases at the core consensus, contacts with the DNA backbone are presumably parsed on the basis of some sequence-dependent DNA shape or propensity to adopt a preferred conformation (or ensemble of such). While indirect readout of bases flanking the core consensus has been unequivocally demonstrated (129), its functional significance in target gene transactivation confirmed (331) and its thermodynamic basis understood in some detail (115, 116), the structural (and probably also dynamic) origin of indirect readout in ETS/DNA site recognition remains elusive. Most frustratingly, no predictive capability of how the structure of given ETS domain relates to its distinct spectrum of flanking sequence preferences has yet been achieved.

Combinatorial routes to ETS target specificity

A pervasive, though not universal, feature of ETS proteins is the presence of elements immediately proximal to their ETS domains that are structured in the absence of DNA but unfold

upon DNA binding. The energetic overhead to unfold these appending elements, which are most extensively characterized in Ets-1 (Class I) (332-334) and ETV6 (Class II) (335-337), result in reduced affinity to any given DNA site on the order of an order of magnitude relative to truncated constructs. This auto-inhibitory mechanism serves as a handle for combinatorial control of ETS proteins through interactions with protein partners that displace the extra-ETS appendages. For example, Runx1/AML1/CBF α 2/PEBP2 positively regulates Ets-1/DNA binding by displacing and destabilizing an extended inhibitory segment N-terminal to the ETS domain (98, 338, 339). In ETS paralogs that are not auto-inhibited, such as PU.1, specific interactions with binding partners that positively regulate (e.g., IRF4/Pip) (340-342) or antagonize DNA binding (e.g., GATA-1) (343, 344) are known. Another combinatorial strategy that modifies DNA site targeting is to couple binding to intrinsically low-affinity or nonspecific sites to specific interactions with binding partners. Such mechanisms are illustrated by the ability of Pax5 to recruit Ets-1 to a nonspecific sequence (5'-GGAG-3') (345, 346). Similar recruitment of PU.1 to intrinsically low-affinity sites has also been reported (266). These and other interactions, all of which are functionally linked to cell fate specification or the regulation of lineage-specific target genes, have been well reviewed (123).

In recent years, high-throughput microarray and sequencing technologies have elevated investigations of ETS/DNA interactions to the genome-wide level. Detailed information on the localization, sequence characteristics of DNA targets, and associated binding partners is now available for ETS transcription factors in a range of cell types and developmental contexts (236, 347-352). Although various levels of redundancy and specificity are observed that correlate with the ontology of the genes involved, one recurring feature is the close correspondence between *in vivo* and *in vitro* DNA sequence preferences. Moreover, in the case of PU.1 and Ets-1, the

information contents (a direct informatic measure of target specificity) of the *in vivo* sequences preferences shown by both proteins are over 15% higher (>3 bits over a 10-bp sequence space) than their *in vitro* counterparts (309). Given the vast number of sequence reads in the *in vivo* data, these two observations indicate that, integrated over the whole genome, combinatorial control refines, rather than usurps, the intrinsic selectivity of their ETS domains. Thus, while the target gene specificity of ETS proteins is functionally controlled by an inter-related web of interactions *in vivo* [Figure 2], the intrinsic properties of ETS domains per se remain a keystone in understanding sequence usage of ETS transcription factors in the genome.

A deeper look into ETS/DNA recognition

To date, structures of ETS/DNA complexes, most of which involve high-affinity cognate sequences, have provided physical models of optimal interactions in DNA target recognition by representative ETS paralogs in each class. Do fundamental mechanisms exist that could explain DNA target selectivity across the spectrum of ETS paralogs? Such “molecular phenotypes” would reasonably reflect the selection pressures operating on functionally distinct ETS paralogs, and in turn provide insight into the biological environment in which the proteins operate. To this end, we have identified two aspects that structural and biochemical data suggest unusual potential for insight: molecular hydration accompanying DNA recognition by ETS domains and their sensitivity to epigenetically modified DNA targets. We have been focusing our attention on Ets-1 and PU.1, which are attractive model systems for two reasons. First, they are archetypal representatives of the most phylogenetically most distant classes of ETS proteins (123), so heterogeneity in their molecular phenotype should directly reflect the selection pressures in their evolutionary paths, even if the biological basis of these pressures are not necessarily known. Second, these two ETS

paralogs bind optimal DNA targets with indistinguishably high affinity (118), so that heterogeneity between the two homologs that contribute to their DNA binding affinity and specificity would be biologically relevant.

Role of molecular hydration in DNA recognition by ETS proteins

Transcription factors interact with and regulate their DNA targets in an aqueous milieu. Rather than a passive bathing medium, water molecules participate in protein/DNA interactions and can act as a major determinant of binding affinity and specificity, for example, by forming water-mediated contacts at the protein/DNA interface. One of the most intriguing differences between the co-crystal structures of PU.1 and Ets-1 is the abundance of interfacial water-mediated contacts in the PU.1 complex and the sparsity of such contacts in the Ets-1 complex (346, 353). The water-mediated contacts made by PU.1 are effectively replaced by direct protein-to-DNA contacts in Ets-1. To determine whether the crystallographic data indeed reflect a differential role for interfacial hydration in DNA recognition by the two proteins in solution, we interrogated DNA binding by the two proteins through osmotic stress (118). Using physiologically compatible osmolytes to modulate the osmotic environment (water activity), the data indeed showed that high-affinity DNA binding by PU.1 was osmotic sensitive while binding by Ets-1 was not [**Table 1**]. Moreover, the osmotic sensitivity of PU.1, wherein affinity was reduced by osmotic pressure, was dependent on the sequence context, inferring a direct role of hydration in the specificity mechanism of PU.1.

The strikingly different responses of the two homologs to their osmotic environment, which is quantitatively compensated to yield indistinguishable binding affinities under normo-osmotic conditions, is a provocative observation. Higher-order organisms maintain a homeostatic environment in which intracellular parameters are controlled (354). Osmotic pressure is one such

parameter. Hyperosmotic stress and its attendant perturbation on cell volume trigger signaling pathways mediated by the transcription factor NFAT5/TonEBP that restore isotonicity through the accumulation of compatible osmolytes (355, 356). As a result, the compensated (isotonic) but now hyperosmolar conditions would necessarily perturb biomolecular interactions that involve sufficiently large hydration changes. The affinity of PU.1 for its optimal cognate sequence is reduced ~10-fold across 0.5 osmolal (113), a physiologically relevant level of osmotic stress in lymphocytes (357). We have found through analysis of microarray data that PU.1 target genes are disproportionately represented in osmotically sensitive (NFAT5-dependent) genes in primary murine macrophages (113). Significantly, other transcription factors that are co-expressed with PU.1, such as the interferon regulatory factors, NF- κ B2, and Stat proteins, show no such overlap. While PU.1 may interact with NFAT5 by direct contact or via post-translational modification, no such interactions are currently known. A more intriguing scenario is that PU.1 and NFAT5 targets overlap by virtue of the osmotic sensitivity of their regulators. It would therefore be possible for genetic networks to interact through “fields” and in a manner that requires no direct contact or post-translational modification of the macromolecular components involved.

Significance of molecular hydration in ETS activity under normo-osmotic conditions

In addition to perturbing affinity under hyperosmotic stress, the osmotic sensitivity of DNA target recognition provides insight into differences in binding behavior under normo-osmotic conditions. Specifically, we have observed that the different disposition of hydration water directly impacts on the mechanisms of DNA target recognition as manifest in the kinetics of association and dissociation [**Table 1**] (113). Under normo-osmotic conditions, PU.1 engages sequence-specific target sites about ~100 more slowly than Ets-1, but once formed, the complex is

correspondingly more persistent than Ets-1. The starkly different kinetic profiles establish that interfacial hydration defines different mechanisms of target recognition by the two ETS homologs. In addition, the persistence of the PU.1/DNA complex against dissociation is consistent with PU.1 as a strong pioneer transcription factor (358), by anchoring target genes in chromatin during recruitment of other transcription factors and remodeling proteins (such as histone acetyltransferases) (359). Currently, the pioneer status of Ets-1 is controversial: although it appears to co-localize with nucleosomes in enhancer regions in developing thymocytes (349), it does not exhibit functional pioneer activity in a defined reporter assay.(360) An intrinsic mechanism for resisting nucleosomal dynamics by the ETS domain to secure accessible proximal binding sites for other proteins represents an intriguing component of pioneering activity of “master” transcription factors such as PU.1.

Molecular hydration as a unifying feature in ETS evolution

We considered whether variation in molecular hydration within the broader ETS family. A survey of binary and ternary co-crystal structures of ETS domains shows a range in both the density and pattern of interfacial hydration, as may be expected from the amino acid diversity in their DNA-binding surfaces. To examine this heterogeneity systematically, we considered the correlation between the number of water-mediated contacts in DNA co-crystal complexes of ETS paralogs as a function of their evolutionary relatedness. Taking Ets-1 as reference, we found that the density of water-mediated contacts for ETS paralogs is positively correlated with its pairwise phylogenetic distance from Ets-1 (**Figure 3**). This is a remarkable correlation. The physicochemical diversity of the crystals (e.g., symmetry, packing, co-solvents, overall hydration) strongly discount against the observed correlation as a crystallographic artefact. There is also no

systematic differences in the resolution of the structures that would account for a bias in discernable hydration. Beyond several water-mediated contacts involving the sidechains and backbone of absolutely or highly conserved residues that are observed in all the structures, a significant diversity in bridging pattern is observed at all levels of hydration, suggesting that interfacial hydration is highly adaptive. While the evolution of ETS paralogs is undoubtedly subject to different selection pressures, which are not universally shared, it appears that as a general feature interfacial hydration is incrementally incorporated in the evolution of the ETS family. The biophysical and biological implications this relationship is currently unknown and ripe for hypothesis.

Differential tolerance to CpG methylation

While the importance of epigenetic regulation of ETS-dependent transcription is well established in hematopoiesis and in cancer (352, 361-363), the mechanisms by which ETS activity are modulated at epigenetically modified DNA, with or without nucleosome, are not well understood. Genomic surveys have found that hematopoietic ETS transcription factors are over-represented in hypermethylated regions (350). While several close Class I ETS paralogs (such as Ets-1 and GABP α) have been reported to be inhibited by CpG methylation at their cognate sites, whether inhibition is a universal property of ETS proteins remains unknown. We have directly studied the binding properties of PU.1 and Ets-1 to hemi- and fully methylated cognate DNA harboring a site-specific CpG dinucleotide (5'-CGGAA-3') that frequently occurs in cognate ETS binding sites (364). While any CpG methylation affected binding, PU.1 and Ets-1 responded qualitatively differently to hemi-methylated sites. Hemi-methylation of the sense (5'-GGAA-3') strand was strongly inhibitory to Ets-1 binding, but hemi-methylation of the anti-sense strand (5'-

GGAA-3') was not, and vice versa for PU.1. In addition, auto-inhibition was operative in Ets-1 with respect to binding to CpG-methylated sites. Finally, PU.1 was significantly more robust than Ets-1 in binding fully methylated DNA. Overall, our targeted studies showed that significant heterogeneity exists in the intrinsic sensitivity to CpG methylation among ETS transcription factors. They confirmed the strong inhibition of Ets-1 (and by extension, other Class I ETS members) by full CpG methylation and a basis for the genomic-wide observation of PU.1 to autonomously engage methylated DNA *in vivo* (266).

Mechanistically, we found by molecular simulations that the asymmetric effect of hemimethylation on the DNA-binding affinity of ETS paralogs may be explained by structural perturbations on DNA backbone geometry. While hemimethylation of either strand significantly perturbs backbone geometry out of the unmethylated configuration, full methylation produces a compensatory effect that brings backbone geometry back closer to unmethylated DNA. In light of the plasticity in interfacial hydration, we speculate that hydration waters serve as adapters that moderate the perturbative effects of DNA methylation on binding for hydration-rich ETS paralogs such as PU.1. In addition, the compensatory relationship between hemi- and full methylation on DNA backbone structure suggests new biological implications in view of the semi-conservative nature of DNA replication. Immediately following DNA synthesis with unmethylated nucleotides, the DNA daughter strands are hemimethylated until re-methylated by DNA methyltransferase I (DNMT1). The exact same site in a passively de-methylated genome, therefore, presents a heterogeneous substrate for ETS paralogs (and probably other DNA-binding proteins) depending on the stage in the cell cycle or exogenous treatment with DNMT1 inhibitors ("hypomethylating agents" such as azacytidine and decitabine). Interest in this area is heightened by the advent of

hypomethylating agents as clinical drugs in hematologic cancers, such as azacytidine in in myelodysplastic syndrome (365) and a growing list of other malignancies.

Chemical biology of ETS proteins

Target-specific control of transcriptional pathways has long been a goal in experimental research and therapy. Despite the success and ubiquity of gene-based approaches to knock-in, known-down, and knock-out specific genes *in vitro* and *in vivo*, as well as the intense efforts to deliver genetic and other macromolecular payloads efficiently and without toxicity into cells and tissues, low-molecular weight molecules (i.e., chemical control) remain the preferred modality of intervention. With few exceptions, the clinically successful pharmacology of nuclear receptors has not been reprised for the vast majority of other transcription factors, particularly wildtype forms which lack endogenous ligands as templates for drug development (255, 366-368). A fruitful avenue in the case of inhibition is to target the cognate DNA site to which the transcription factor binds. Thanks to the considerable advance over the past two decades in sequence-specific targeting of DNA-binding ligands, particularly the hairpin polyamides (369) and heterocyclic diamidines (370), proofs of concept have been achieved for several ETS proteins. They include the inhibition of Ets-1 with a designed polyamide (371) and ERG using designed heterocyclic diamidines (81). We have demonstrated the inhibition of PU.1 using diamidines of a different class (300, 372). Viable chemical biology of ETS proteins is a challenge in need of actionable targets, and identification of the molecular heterogeneity among ETS domains could clarify such targets and strategies for control.

Concluding remarks

Although all ETS paralogs display highly homologous backbone structures and engage target DNA in an essentially identical conformation, they harbor a spectrum of distinct physical chemistry that is likely reflected in their functional phenotypes. Targeted studies of interactions by ETS domains in new areas such as molecular hydration and epigenetically modified DNA are sparking novel perspectives and opportunities for new insights into the diversity of this important family of transcriptional regulators.

ACKNOWLEDGEMENT

Our work on ETS transcription factors is supported by NSF grant MCB 15451600 and NIH grant R21 HL129063. H.M.K. is supported by a GSU Molecular Basis of Disease Fellowship. We are grateful to Drs. W. David Wilson (GSU), David W. Boykin (GSU), and James K. Bashkin (University of Missouri – St. Louis), and their research personnel whose collaboration make our work possible.

Table 1.
Biological and biochemical comparison of the ETS paralogs PU.1 and Ets-1

	Ets-1	PU.1
Functional properties		
Expression(330)	Widespread	Lineage-restricted to cells of hematopoietic origin
High dosage requirement	T-cell	B-cell
Low dosage requirement	B-cell	T-cell
Auto-inhibition	Yes	No
Pioneer transcription factor	No (360)	Yes (358)
General role in hematopoietic cancers	Oncogene (373)	Tumor suppressor (325, 374, 375)
Biochemical properties (ETS domain)		
Domain topology	C-terminal	C-terminal
Intrinsic affinity for optimal cognate sites(309)	10^{-10} M	10^{-10} M
Relative sequence discrimination(309)	Low	High
Osmotic sensitivity	Low (118)	High (113)

Kinetics of association(118)	Fast (10^7 $M^{-1} s^{-1}$)	Slow ($10^5 M^{-1} s^{-1}$)
Kinetics of dissociation(118)	Slow (10^{-2} s^{-1})	Very slow ($10^{-3} s^{-1}$)
Relative tolerance to CpG methylation(364)	Low	High
Dimerize in solution without DNA	No (220)	Yes (112, 376)

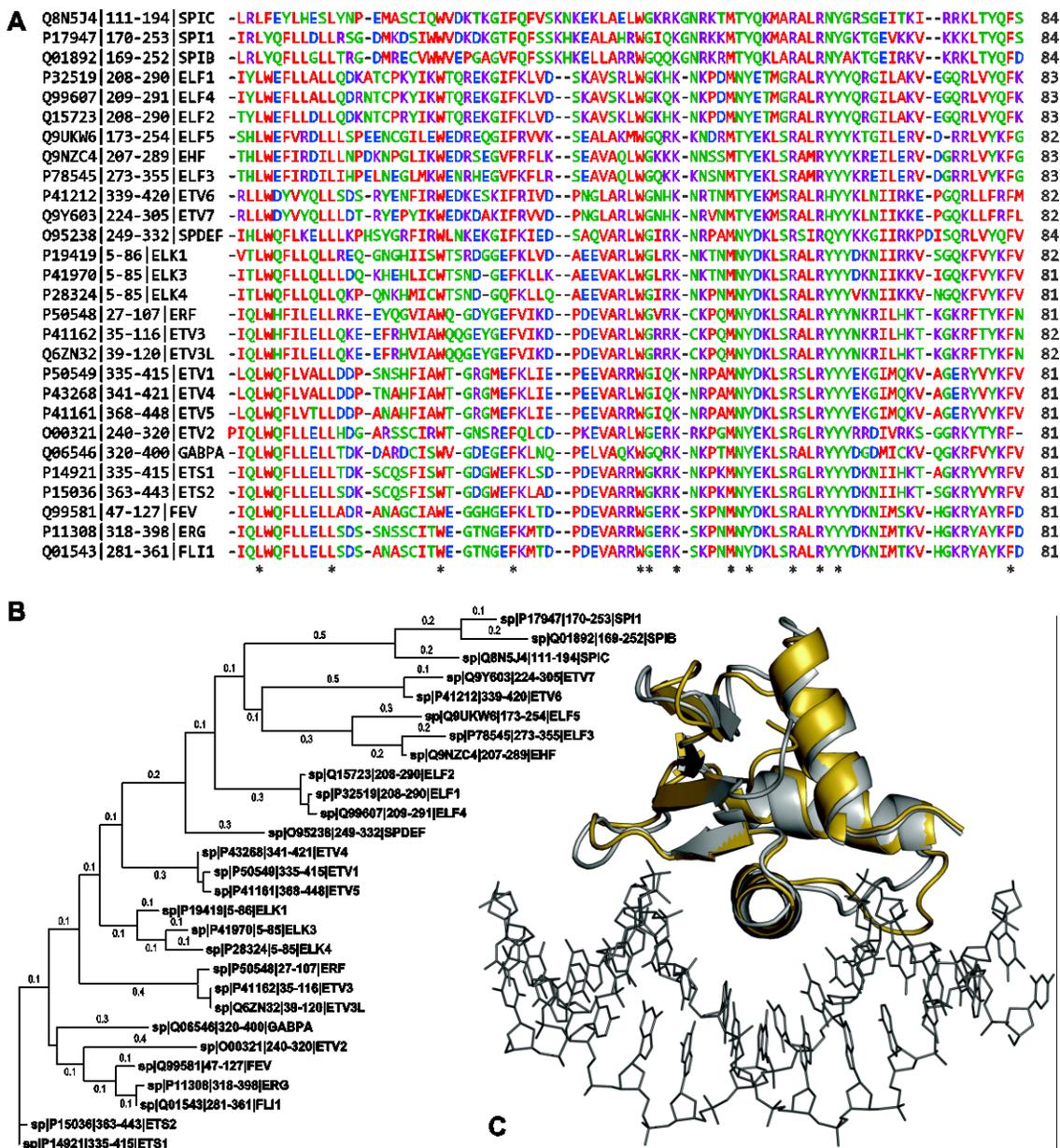


Figure 1. The DNA-binding domains of ETS transcription factors are sequence- and phylogenetically divergent, but strongly conserved in structure. *A* Sequence alignment for the 28 paralogous ETS domains. Proteins were identified by UniProtKB identifier | residue numbering | protein name. Residues are colored by amino acid types. Asterisks denote positions with amino acid identity. *B*, Phylogenetic tree constructed by the maximum-likelihood method, arranged with Ets-1 and PU.1 (Spi-1) at the ends. The horizontal distance (branch length) denotes phylogenetic distance defined as number of substitutions per position. *C*, Structural alignment of the ETS domains of Ets-1 (silver) and PU.1 (gold) from their co-crystal structures with DNA. The root-mean-square deviation is 1.4 Å, well below the experimental resolution of the models themselves. The target DNA from the Ets-1 co-crystal structure is shown to orient the viewer.

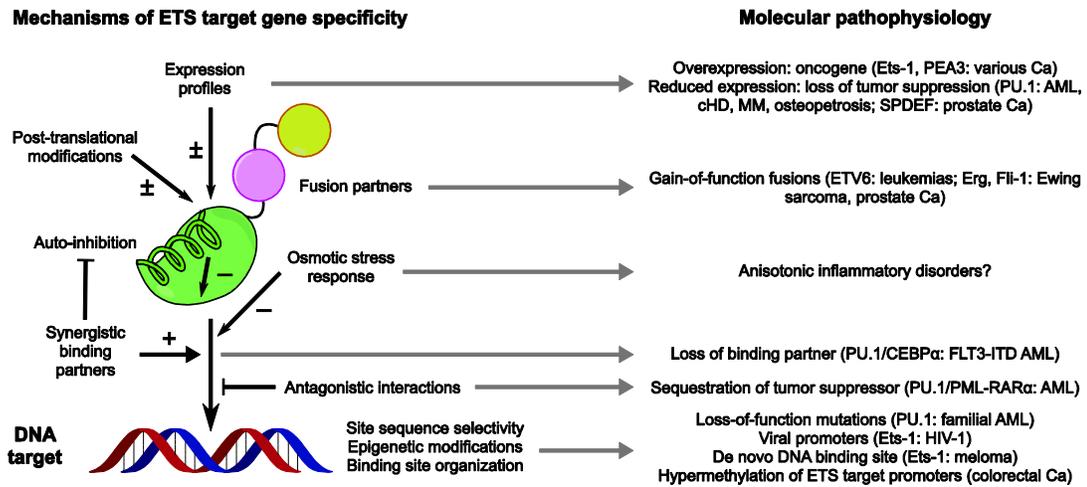


Figure 2. **Selected mechanisms of ETS target gene specificity.** ETS-dependent transcription is regulated at multiple levels, all of which can operate in a combinatorial fashion. The established molecular pathophysiology associated with some of these mechanisms are listed. The literature on ETS proteins is vast and this summary is only intended to be illustrative; readers interested in specific aspects or paralogs mentioned in this figure are referred to reviews and studies such as the following (96, 123, 330, 348, 350, 361, 377-382). Note that auto-inhibition (indicated by the cartoon helix) is not a universal feature of ETS proteins; several paralogs, such as PU.1, are not auto-inhibited. Abbreviations: AML, acute myeloid leukemia; cHD, classical Hodgkin's disease; MM, multiple myeloma; Ca, cancer.

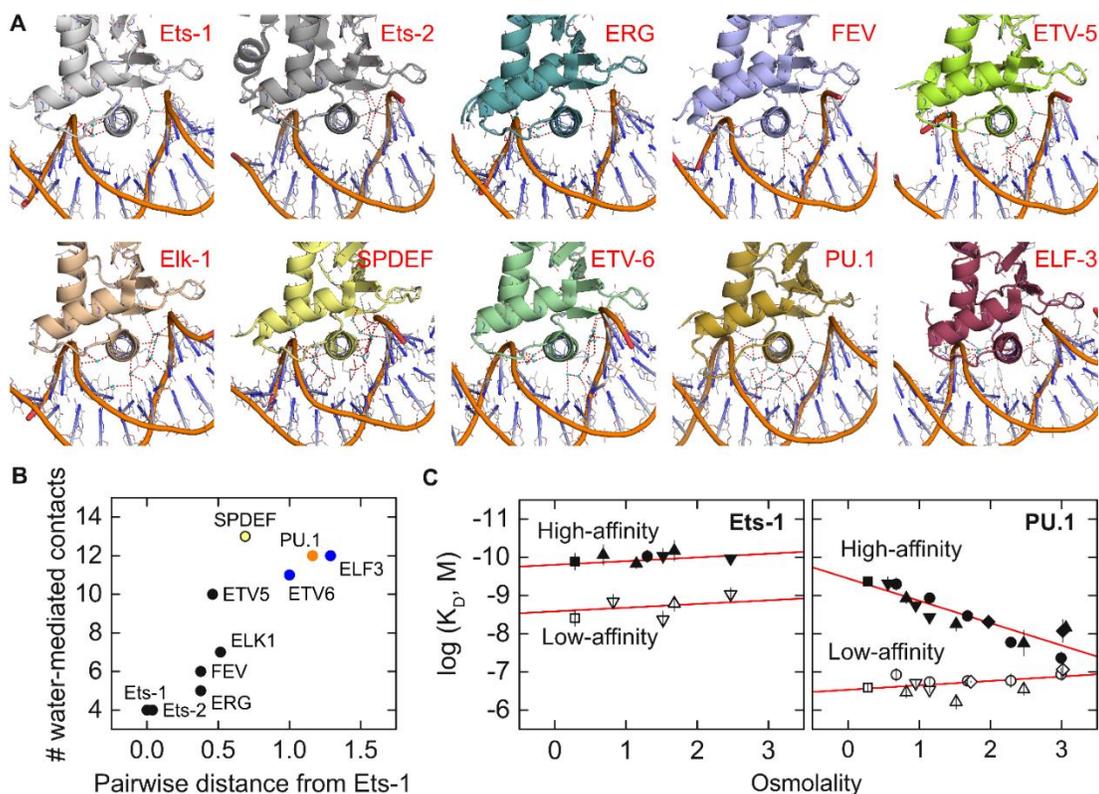


Figure 3. Crystallographic interfacial hydration correlates positively with phylogenetic relatedness among the ETS domains of paralogs. *A*, Co-crystal structures of the ETS domains of nine ETS paralogs, oriented identically with the recognition helix perpendicular to the plane of the page. Water-mediated contacts are shown as cyan spheres, defined operationally as crystallographic water within hydrogen-bonding distance (red dashes, ≤ 3.4 Å) of a protein and DNA contact, or another interfacial water that meets this criterion. To avoid ambiguity, water-mediated contacts involving only three or more consecutive bridging water are not counted. Interfacial water density is weakly correlated with overall hydration of the asymmetric unit, and there is no significant difference in interfacial hydration density (± 1) between different biological units where reported. *B*, Relationship between crystallographic interfacial hydration and pairwise phylogenetic distance from Ets-1, chosen as reference. The primary sequences of the 28 human ETS paralogs were analyzed by ClustalW using the neighbor-joining method. The results were expressed as a distance matrix from whose elements are pairwise distances (number of substitutions per position). ETS paralogs are formally categorized into Class I to IV (236) by color in order from black, blue, orange, to yellow. *C*, Differential sensitivity to osmotic pressure in site-specific binding by the ETS domains of PU.1 and Ets-1 as reported by Wang et al (118). The measured *in vitro* affinity is expressed as the logarithm of the dissociation constant (K_D). High- (solid symbols) and low-affinity DNA (open) refer to defined cognate (not nonspecific) sequences harboring the 5'-GGAA-3' consensus. The different symbols refer to the set of physiologically compatible osmolytes used to exert osmotic stress. The osmotic insensitivity of Ets-1 is not modified by the presence of auto-inhibition (118).

REFERENCES

1. Crick F. Central dogma of molecular biology. *Nature*. 1970;227(5258):561-3. Epub 1970/08/08. doi: 10.1038/227561a0. PubMed PMID: 4913914.
2. Crick FH. On protein synthesis. *Symp Soc Exp Biol*. 1958;12:138-63. Epub 1958/01/01. PubMed PMID: 13580867.
3. Holliday R. Epigenetics: a historical overview. *Epigenetics*. 2006;1(2):76-80. Epub 2007/11/14. doi: 10.4161/epi.1.2.2762. PubMed PMID: 17998809.
4. Buchberger E, Reis M, Lu TH, Posnien N. Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies. *Genes (Basel)*. 2019;10(7). Epub 2019/07/03. doi: 10.3390/genes10070492. PubMed PMID: 31261769; PMCID: PMC6678813.
5. Bird A. Perceptions of epigenetics. *Nature*. 2007;447(7143):396-8. Epub 2007/05/25. doi: 10.1038/nature05913. PubMed PMID: 17522671.
6. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer*. 2006;6(2):107-16. Epub 2006/02/24. doi: 10.1038/nrc1799. PubMed PMID: 16491070.
7. Chen K, Zhao BS, He C. Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem Biol*. 2016;23(1):74-85. Epub 2016/03/05. doi: 10.1016/j.chembiol.2015.11.007. PubMed PMID: 26933737; PMCID: PMC4779186.
8. Jayasena SD, Johnston BH. Intramolecular triple-helix formation at (PunPyn).(PunPyn) tracts: recognition of alternate strands via Pu.PuPy and Py.PuPy base triplets. *Biochemistry*. 1992;31(2):320-7. Epub 1992/01/21. doi: 10.1021/bi00117a002. PubMed PMID: 1731890.
9. McCue J, Freed B. Measuring lymphocyte transcription factor activity by ELISA. *Curr Protoc Toxicol*. 2005;Chapter 18:Unit18 5. Epub 2005/01/01. doi: 10.1002/0471140856.tx1805s22. PubMed PMID: 23045117.
10. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature*. 1986;321(6067):209-13. Epub 1986/05/15. doi: 10.1038/321209a0. PubMed PMID: 2423876.
11. Carlone DL, Skalnik DG. CpG binding protein is crucial for early embryonic development. *Mol Cell Biol*. 2001;21(22):7601-6. Epub 2001/10/18. doi: 10.1128/MCB.21.22.7601-7606.2001. PubMed PMID: 11604496; PMCID: PMC99931.
12. Cyr AR, Hitchler MJ, Domann FE. Regulation of SOD2 in cancer by histone modifications and CpG methylation: closing the loop between redox biology and epigenetics. *Antioxid Redox Signal*. 2013;18(15):1946-55. Epub 2012/09/06. doi: 10.1089/ars.2012.4850. PubMed PMID: 22946823; PMCID: PMC3624766.
13. Migeon BR, Holland MM, Driscoll DJ, Robinson JC. Programmed demethylation in CpG islands during human fetal development. *Somat Cell Mol Genet*. 1991;17(2):159-68. Epub 1991/03/01. doi: 10.1007/BF01232973. PubMed PMID: 2011794.
14. Nakatsuka S, Takakuwa T, Tomita Y, Miwa H, Matsuzuka F, Aozasa K. Role of hypermethylation of DAP-kinase CpG island in the development of thyroid lymphoma. *Lab Invest*. 2000;80(11):1651-5. Epub 2000/11/25. doi: 10.1038/labinvest.3780175. PubMed PMID: 11092525.
15. Skelton-Stroud P, Mull R. Positioning, labelling, and medical information control of co-artemether tablets (CPG 56697): a fixed novel combination of artemether and benflumetol. Novartis Co-Artemether International Development Team. *Med Trop (Mars)*. 1998;58(3 Suppl):77-81. Epub 1999/04/23. PubMed PMID: 10212907.

16. Tate P, Skarnes W, Bird A. The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nat Genet.* 1996;12(2):205-8. Epub 1996/02/01. doi: 10.1038/ng0296-205. PubMed PMID: 8563762.
17. Bensaada M, Kiefer H, Tachdjian G, Lapierre JM, Cacheux V, Niveleau A, Metezeau P. Altered patterns of DNA methylation on chromosomes from leukemia cell lines: identification of 5-methylcytosines by indirect immunodetection. *Cancer Genet Cytogenet.* 1998;103(2):101-9. Epub 1998/06/06. doi: 10.1016/s0165-4608(97)00409-3. PubMed PMID: 9614907.
18. Bernardino J, Lamoliatte E, Lombard M, Niveleau A, Malfoy B, Dutrillaux B, Bourgeois CA. DNA methylation of the X chromosomes of the human female: an in situ semi-quantitative analysis. *Chromosoma.* 1996;104(7):528-35. Epub 1996/04/01. doi: 10.1007/BF00352117. PubMed PMID: 8625741.
19. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet.* 2006;38(12):1378-85. Epub 2006/10/31. doi: 10.1038/ng1909. PubMed PMID: 17072317; PMCID: PMC3082778.
20. Dolinoy DC, Weidman JR, Jirtle RL. Epigenetic gene regulation: linking early developmental environment to adult disease. *Reprod Toxicol.* 2007;23(3):297-307. Epub 2006/10/19. doi: 10.1016/j.reprotox.2006.08.012. PubMed PMID: 17046196.
21. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature.* 2004;429(6990):457-63. Epub 2004/05/28. doi: 10.1038/nature02625. PubMed PMID: 15164071.
22. Robertson KD, Wolffe AP. DNA methylation in health and disease. *Nat Rev Genet.* 2000;1(1):11-9. Epub 2001/03/27. doi: 10.1038/35049533. PubMed PMID: 11262868.
23. Berger SL. The complex language of chromatin regulation during transcription. *Nature.* 2007;447(7143):407-12. Epub 2007/05/25. doi: 10.1038/nature05915. PubMed PMID: 17522673.
24. Bhattacharjee D, Shenoy S, Bairy KL. DNA Methylation and Chromatin Remodeling: The Blueprint of Cancer Epigenetics. *Scientifica (Cairo).* 2016;2016:6072357. Epub 2016/04/28. doi: 10.1155/2016/6072357. PubMed PMID: 27119045; PMCID: PMC4826949.
25. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* 2011;21(3):381-95. Epub 2011/02/16. doi: 10.1038/cr.2011.22. PubMed PMID: 21321607; PMCID: PMC3193420.
26. Marmorstein R, Trievel RC. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim Biophys Acta.* 2009;1789(1):58-68. Epub 2008/08/30. doi: 10.1016/j.bbagr.2008.07.009. PubMed PMID: 18722564; PMCID: PMC4059211.
27. Falbo L, Costanzo V. Epigenetic regulation of replication origin assembly: A role for histone H1 and chromatin remodeling factors. *Bioessays.* 2021;43(1):e2000181. Epub 2020/11/10. doi: 10.1002/bies.202000181. PubMed PMID: 33165968.
28. Kishimoto M, Fujiki R, Takezawa S, Sasaki Y, Nakamura T, Yamaoka K, Kitagawa H, Kato S. Nuclear receptor mediated gene regulation through chromatin remodeling and histone modifications. *Endocr J.* 2006;53(2):157-72. Epub 2006/04/19. doi: 10.1507/endocrj.53.157. PubMed PMID: 16618973.
29. Krogan NJ, Baetz K, Keogh MC, Datta N, Sawa C, Kwok TC, Thompson NJ, Davey MG, Pootoolal J, Hughes TR, Emili A, Buratowski S, Hieter P, Greenblatt JF. Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex,

- and the histone acetyltransferase NuA4. *Proc Natl Acad Sci U S A*. 2004;101(37):13513-8. Epub 2004/09/09. doi: 10.1073/pnas.0405753101. PubMed PMID: 15353583; PMCID: PMC518788.
30. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128(4):693-705. Epub 2007/02/27. doi: 10.1016/j.cell.2007.02.005. PubMed PMID: 17320507.
31. Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. *Nat Rev Cancer*. 2018;18(1):5-18. Epub 2017/11/25. doi: 10.1038/nrc.2017.99. PubMed PMID: 29170536; PMCID: PMC6337726.
32. Liang Y, Zhang C, Ma MH, Dai DQ. Identification and prediction of novel non-coding and coding RNA-associated competing endogenous RNA networks in colorectal cancer. *World J Gastroenterol*. 2018;24(46):5259-70. Epub 2018/12/26. doi: 10.3748/wjg.v24.i46.5259. PubMed PMID: 30581274; PMCID: PMC6295837.
33. Yang S, Zhang H, Guo L, Zhao Y, Chen F. Reconstructing the coding and non-coding RNA regulatory networks of miRNAs and mRNAs in breast cancer. *Gene*. 2014;548(1):6-13. Epub 2014/07/01. doi: 10.1016/j.gene.2014.06.010. PubMed PMID: 24979338.
34. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet*. 2015;16(2):71-84. Epub 2015/01/03. doi: 10.1038/nrg3863. PubMed PMID: 25554358; PMCID: PMC4376354.
35. Li J, Liu C. Coding or Noncoding, the Converging Concepts of RNAs. *Front Genet*. 2019;10:496. Epub 2019/06/11. doi: 10.3389/fgene.2019.00496. PubMed PMID: 31178900; PMCID: PMC6538810.
36. Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, Cockett N, Georges M, Charlier C. RNAi-mediated allelic trans-interaction at the imprinted *Rtl1/Peg11* locus. *Curr Biol*. 2005;15(8):743-9. Epub 2005/04/28. doi: 10.1016/j.cub.2005.02.060. PubMed PMID: 15854907.
37. Bhuyan PK, Kariko K, Capodici J, Lubinski J, Hook LM, Friedman HM, Weissman D. Short interfering RNA-mediated inhibition of herpes simplex virus type 1 gene expression and function during infection of human keratinocytes. *J Virol*. 2004;78(19):10276-81. Epub 2004/09/16. doi: 10.1128/JVI.78.19.10276-10281.2004. PubMed PMID: 15367593; PMCID: PMC516413.
38. Doi N, Zenno S, Ueda R, Ohki-Hamazaki H, Ui-Tei K, Saigo K. Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr Biol*. 2003;13(1):41-6. Epub 2003/01/16. doi: 10.1016/s0960-9822(02)01394-5. PubMed PMID: 12526743.
39. Kishida T, Asada H, Gojo S, Ohashi S, Shin-Ya M, Yasutomi K, Terauchi R, Takahashi KA, Kubo T, Imanishi J, Mazda O. Sequence-specific gene silencing in murine muscle induced by electroporation-mediated transfer of short interfering RNA. *J Gene Med*. 2004;6(1):105-10. Epub 2004/01/13. doi: 10.1002/jgm.456. PubMed PMID: 14716682.
40. Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, Fesik SW. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci U S A*. 2003;100(11):6347-52. Epub 2003/05/15. doi: 10.1073/pnas.1131959100. PubMed PMID: 12746500; PMCID: PMC164449.
41. Vanitharani R, Chellappan P, Fauquet CM. Short interfering RNA-mediated interference of gene expression and viral DNA accumulation in cultured plant cells. *Proc Natl Acad Sci U S A*. 2003;100(16):9632-6. Epub 2003/07/30. doi: 10.1073/pnas.1733874100. PubMed PMID: 12886005; PMCID: PMC170969.
42. Zhao TY, Zou SP, Alimova YV, Wang G, Hauser KF, Ghandour MS, Knapp PE. Short interfering RNA-induced gene silencing is transmitted between cells from the mammalian central

nervous system. *J Neurochem.* 2006;98(5):1541-50. Epub 2006/08/23. doi: 10.1111/j.1471-4159.2006.03974.x. PubMed PMID: 16923165; PMCID: PMC4030304.

43. Grivna ST, Pyhtila B, Lin H. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc Natl Acad Sci U S A.* 2006;103(36):13415-20. Epub 2006/08/30. doi: 10.1073/pnas.0605506103. PubMed PMID: 16938833; PMCID: PMC1569178.

44. Kim VN. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev.* 2006;20(15):1993-7. Epub 2006/08/03. doi: 10.1101/gad.1456106. PubMed PMID: 16882976.

45. Chen MT, Lin HS, Shen C, Ma YN, Wang F, Zhao HL, Yu J, Zhang JW. PU.1-Regulated Long Noncoding RNA Inc-MC Controls Human Monocyte/Macrophage Differentiation through Interaction with MicroRNA 199a-5p. *Mol Cell Biol.* 2015;35(18):3212-24. Epub 2015/07/08. doi: 10.1128/MCB.00429-15. PubMed PMID: 26149389; PMCID: PMC4539372.

46. Pennisi E. Massive project reveals complexity of gene regulation. *Science.* 2020;369(6509):1286-7. Epub 2020/09/12. doi: 10.1126/science.369.6509.1286. PubMed PMID: 32913084.

47. Wang Y, Fu L, Sun A, Tang D, Xu Y, Li Z, Chen M, Zhang G. C/EBPbeta contributes to transcriptional activation of long non-coding RNA NEAT1 during APL cell differentiation. *Biochem Biophys Res Commun.* 2018;499(2):99-104. Epub 2017/11/08. doi: 10.1016/j.bbrc.2017.10.137. PubMed PMID: 29111326.

48. Hoopengardner B. Adenosine-to-inosine RNA editing: perspectives and predictions. *Mini Rev Med Chem.* 2006;6(11):1213-6. Epub 2006/11/15. doi: 10.2174/138955706778742812. PubMed PMID: 17100632.

49. Jaikaran DC, Collins CH, MacMillan AM. Adenosine to inosine editing by ADAR2 requires formation of a ternary complex on the GluR-B R/G site. *J Biol Chem.* 2002;277(40):37624-9. Epub 2002/08/07. doi: 10.1074/jbc.M204126200. PubMed PMID: 12163487.

50. Keller W, Wolf J, Gerber A. Editing of messenger RNA precursors and of tRNAs by adenosine to inosine conversion. *FEBS Lett.* 1999;452(1-2):71-6. Epub 1999/06/22. doi: 10.1016/s0014-5793(99)00590-6. PubMed PMID: 10376681.

51. Ko NL, Birlouez E, Wain-Hobson S, Mahieux R, Vartanian JP. Hyperediting of human T-cell leukemia virus type 2 and simian T-cell leukemia virus type 3 by the dsRNA adenosine deaminase ADAR-1. *J Gen Virol.* 2012;93(Pt 12):2646-51. Epub 2012/09/21. doi: 10.1099/vir.0.045146-0. PubMed PMID: 22993189; PMCID: PMC4091295.

52. Tomaselli S, Bonamassa B, Alisi A, Nobili V, Locatelli F, Gallo A. ADAR enzyme and miRNA story: a nucleotide that can make the difference. *Int J Mol Sci.* 2013;14(11):22796-816. Epub 2013/11/22. doi: 10.3390/ijms141122796. PubMed PMID: 24256817; PMCID: PMC3856091.

53. White MA, Myers CA, Corbo JC, Cohen BA. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci U S A.* 2013;110(29):11952-7. Epub 2013/07/03. doi: 10.1073/pnas.1307449110. PubMed PMID: 23818646; PMCID: PMC3718143.

54. Hughes TR, Lambert SA. Transcription factors read epigenetics. *Science.* 2017;356(6337):489-90. Epub 2017/05/06. doi: 10.1126/science.aan2927. PubMed PMID: 28473550.

55. Belver L, Albero R, Ferrando AA. Deregulation of enhancer structure, function, and dynamics in acute lymphoblastic leukemia. *Trends Immunol.* 2021;42(5):418-31. Epub 2021/04/17. doi: 10.1016/j.it.2021.03.005. PubMed PMID: 33858773; PMCID: PMC8091164.
56. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet.* 2011;13(1):59-69. Epub 2011/12/07. doi: 10.1038/nrg3095. PubMed PMID: 22143240.
57. Jindal GA, Farley EK. Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev Cell.* 2021;56(5):575-87. Epub 2021/03/11. doi: 10.1016/j.devcel.2021.02.016. PubMed PMID: 33689769.
58. Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS. Suboptimization of developmental enhancers. *Science.* 2015;350(6258):325-8. Epub 2015/10/17. doi: 10.1126/science.aac6948. PubMed PMID: 26472909; PMCID: PMC4970741.
59. Nam Y, Sliz P, Pear WS, Aster JC, Blacklow SC. Cooperative assembly of higher-order Notch complexes functions as a switch to induce transcription. *Proc Natl Acad Sci U S A.* 2007;104(7):2103-8. Epub 2007/02/08. doi: 10.1073/pnas.0611092104. PubMed PMID: 17284587; PMCID: PMC1892977.
60. Farley EK, Olson KM, Zhang W, Rokhsar DS, Levine MS. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc Natl Acad Sci U S A.* 2016;113(23):6508-13. Epub 2016/05/08. doi: 10.1073/pnas.1605085113. PubMed PMID: 27155014; PMCID: PMC4988596.
61. Hanes SD, Riddihough G, Ish-Horowicz D, Brent R. Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen. *Mol Cell Biol.* 1994;14(5):3364-75. Epub 1994/05/01. doi: 10.1128/mcb.14.5.3364-3375.1994. PubMed PMID: 7909358; PMCID: PMC358702.
62. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet.* 2007;8(2):93-103. Epub 2007/01/19. doi: 10.1038/nrg1990. PubMed PMID: 17230196.
63. Hochschild A, Irwin N, Ptashne M. Repressor structure and the mechanism of positive control. *Cell.* 1983;32(2):319-25. Epub 1983/02/01. doi: 10.1016/0092-8674(83)90451-8. PubMed PMID: 6218886.
64. Levine M, Tjian R. Transcription regulation and animal diversity. *Nature.* 2003;424(6945):147-51. Epub 2003/07/11. doi: 10.1038/nature01763. PubMed PMID: 12853946.
65. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature.* 2001;409(6819):533-8. Epub 2001/02/24. doi: 10.1038/35054095. PubMed PMID: 11206552.
66. Aegerter-Wilmsen T, Bisseling T. Biology by numbers--introducing quantitation into life science education. *PLoS Biol.* 2005;3(1):e1. Epub 2005/01/22. doi: 10.1371/journal.pbio.0030001. PubMed PMID: 15660151; PMCID: PMC544540.
67. Matsubara K. Molecular biology and life science. *J Periodontal Res.* 1993;28(6 Pt 2):459-61. Epub 1993/11/01. doi: 10.1111/j.1600-0765.1993.tb02101.x. PubMed PMID: 8263713.
68. Dang CV, Reddy EP, Shokat KM, Soucek L. Drugging the 'undruggable' cancer targets. *Nat Rev Cancer.* 2017;17(8):502-8. Epub 2017/06/24. doi: 10.1038/nrc.2017.36. PubMed PMID: 28643779; PMCID: PMC5945194.
69. Noguera NI, Piredda ML, Taulli R, Catalano G, Angelini G, Gaur G, Nervi C, Voso MT, Lunardi A, Pandolfi PP, Lo-Coco F. PML/RARa inhibits PTEN expression in hematopoietic cells

- by competing with PU.1 transcriptional activity. *Oncotarget*. 2016;7(41):66386-97. Epub 2016/09/15. doi: 10.18632/oncotarget.11964. PubMed PMID: 27626703; PMCID: PMC5341808.
70. Fontaine F, Overman J, Francois M. Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. *Cell Regen*. 2015;4(1):2. Epub 2015/04/08. doi: 10.1186/s13619-015-0015-x. PubMed PMID: 25848531; PMCID: PMC4365538.
71. Furtek SL, Backos DS, Matheson CJ, Reigan P. Strategies and Approaches of Targeting STAT3 for Cancer Treatment. *ACS Chem Biol*. 2016;11(2):308-18. Epub 2016/01/06. doi: 10.1021/acscchembio.5b00945. PubMed PMID: 26730496.
72. Leeman-Neill RJ, Bhagat G. BCL6 as a therapeutic target for lymphoma. *Expert Opin Ther Targets*. 2018;22(2):143-52. Epub 2017/12/22. doi: 10.1080/14728222.2018.1420782. PubMed PMID: 29262721.
73. Tisato V, Voltan R, Gonelli A, Secchiero P, Zauli G. MDM2/X inhibitors under clinical evaluation: perspectives for the management of hematological malignancies and pediatric cancer. *J Hematol Oncol*. 2017;10(1):133. Epub 2017/07/05. doi: 10.1186/s13045-017-0500-5. PubMed PMID: 28673313; PMCID: PMC5496368.
74. Burchenal JH, Rhoads CP. Clinical evaluation of 6-mercaptopurine in the treatment of leukemia. *Trans Assoc Am Physicians*. 1954;67:268-75. Epub 1954/01/01. PubMed PMID: 13216836.
75. Demarsh QB. The treatment of leukemia with 6-mercaptopurine. *Ann N Y Acad Sci*. 1954;60(2):483-91. Epub 1954/12/06. doi: 10.1111/j.1749-6632.1954.tb40045.x. PubMed PMID: 14350560.
76. Hall BE, Richards MD, Willett FM, Feichtmeir TV, Aggeler PM. Evaluation of 6-mercaptopurine in the treatment of leukemia; a preliminary report. *Stanford Med Bull*. 1953;11(4):237-40. Epub 1953/11/01. PubMed PMID: 13122302.
77. Lozner EL. 6-Mercaptopurine in acute leukemia. *Ann N Y Acad Sci*. 1954;60(2):466-7. Epub 1954/12/06. doi: 10.1111/j.1749-6632.1954.tb40041.x. PubMed PMID: 14350556.
78. Grimley E, Liao C, Ranghini EJ, Nikolovska-Coleska Z, Dressler GR. Inhibition of Pax2 Transcription Activation with a Small Molecule that Targets the DNA Binding Domain. *ACS Chem Biol*. 2017;12(3):724-34. Epub 2017/01/18. doi: 10.1021/acscchembio.6b00782. PubMed PMID: 28094913; PMCID: PMC5761330.
79. Bouhrel MA, Lambert M, David-Cordonnier MH. Targeting Transcription Factor Binding to DNA by Competing with DNA Binders as an Approach for Controlling Gene Expression. *Curr Top Med Chem*. 2015;15(14):1323-58. Epub 2015/04/14. doi: 10.2174/1568026615666150413154713. PubMed PMID: 25866275.
80. Leung CH, Chan DS, Ma VP, Ma DL. DNA-binding small molecules as inhibitors of transcription factors. *Med Res Rev*. 2013;33(4):823-46. Epub 2012/05/03. doi: 10.1002/med.21266. PubMed PMID: 22549740.
81. Nhili R, Peixoto P, Depauw S, Flajollet S, Dezitter X, Munde MM, Ismail MA, Kumar A, Farahat AA, Stephens CE, Duterque-Coquillaud M, David Wilson W, Boykin DW, David-Cordonnier MH. Targeting the DNA-binding activity of the human ERG transcription factor using new heterocyclic dithiophene diamidines. *Nucleic acids research*. 2013;41(1):125-38. doi: 10.1093/nar/gks971. PubMed PMID: 23093599; PMCID: PMC3592449.
82. Peixoto P, Liu Y, Depauw S, Hildebrand MP, Boykin DW, Bailly C, Wilson WD, David-Cordonnier MH. Direct inhibition of the DNA-binding activity of POU transcription factors Pit-1 and Brn-3 by selective binding of a phenyl-furan-benzimidazole dication. *Nucleic Acids Res*.

2008;36(10):3341-53. Epub 2008/04/29. doi: 10.1093/nar/gkn208. PubMed PMID: 18440973; PMCID: PMC2425483.

83. Depauw S, Lambert M, Jambon S, Paul A, Peixoto P, Nhili R, Marongiu L, Figeac M, Dassi C, Paul-Constant C, Billore B, Kumar A, Farahat AA, Ismail MA, Mineva E, Sweat DP, Stephens CE, Boykin DW, Wilson WD, David-Cordonnier MH. Heterocyclic Diamidine DNA Ligands as HOXA9 Transcription Factor Inhibitors: Design, Molecular Evaluation, and Cellular Consequences in a HOXA9-Dependant Leukemia Cell Model. *J Med Chem*. 2019;62(3):1306-29. Epub 2019/01/16. doi: 10.1021/acs.jmedchem.8b01448. PubMed PMID: 30645099; PMCID: PMC6561105.

84. Munde M, Wang, S., Kumar, A., Stephens, C. E., Farahat, A. A., Boykin, D. W., & Poon, G. M. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. *Nucleic acids research*. 2014;42:1379-90. doi: 10.1093/nar/gkt955.

85. Bando T, Sugiyama H. Synthesis and biological properties of sequence-specific DNA-alkylating pyrrole-imidazole polyamides. *Acc Chem Res*. 2006;39(12):935-44. Epub 2006/12/21. doi: 10.1021/ar030287f. PubMed PMID: 17176032.

86. Daifu T, Mikami M, Hiramatsu H, Iwai A, Umeda K, Noura M, Kubota H, Masuda T, Furuichi K, Takasaki S, Noguchi Y, Morita K, Bando T, Hirata M, Kataoka TR, Nakahata T, Kuwahara Y, Iehara T, Hosoi H, Takita J, Sugiyama H, Adachi S, Kamikubo Y. Suppression of malignant rhabdoid tumors through Chb-M'-mediated RUNX1 inhibition. *Pediatr Blood Cancer*. 2021;68(2):e28789. Epub 2020/11/13. doi: 10.1002/pbc.28789. PubMed PMID: 33180377.

87. Munde M, Ismail MA, Arafa R, Peixoto P, Collar CJ, Liu Y, Hu L, David-Cordonnier MH, Lansiaux A, Bailly C, Boykin DW, Wilson WD. Design of DNA minor groove binding diamidines that recognize GC base pair sequences: a dimeric-hinge interaction motif. *J Am Chem Soc*. 2007;129(44):13732-43. Epub 2007/10/16. doi: 10.1021/ja074560a. PubMed PMID: 17935330; PMCID: PMC3865524.

88. Munde M, Lee M, Neidle S, Arafa R, Boykin DW, Liu Y, Bailly C, Wilson WD. Induced fit conformational changes of a "reversed amidine" heterocycle: optimized interactions in a DNA minor groove complex. *J Am Chem Soc*. 2007;129(17):5688-98. Epub 2007/04/12. doi: 10.1021/ja069003n. PubMed PMID: 17425312; PMCID: PMC2547086.

89. Munde M, Poon GM, Wilson WD. Probing the electrostatics and pharmacological modulation of sequence-specific binding by the DNA-binding domain of the ETS family transcription factor PU.1: a binding affinity and kinetics investigation. *J Mol Biol*. 2013;425(10):1655-69. Epub 2013/02/19. doi: 10.1016/j.jmb.2013.02.010. PubMed PMID: 23416556; PMCID: PMC3762459.

90. Alipov G, Nakayama T, Ito M, Kawai K, Naito S, Nakashima M, Niino D, Sekine I. Overexpression of Ets-1 proto-oncogene in latent and clinical prostatic carcinomas. *Histopathology*. 2005;46(2):202-8. Epub 2005/02/08. doi: 10.1111/j.1365-2559.2005.02059.x. PubMed PMID: 15693893.

91. Gavrilov D, Kenzior O, Evans M, Calaluce R, Folk WR. Expression of urokinase plasminogen activator and receptor in conjunction with the ets family and AP-1 complex transcription factors in high grade prostate cancers. *Eur J Cancer*. 2001;37(8):1033-40. Epub 2001/05/04. doi: 10.1016/s0959-8049(01)00077-6. PubMed PMID: 11334730.

92. Baldus CD, Liyanarachchi S, Mrozek K, Auer H, Tanner SM, Guimond M, Ruppert AS, Mohamed N, Davuluri RV, Caligiuri MA, Bloomfield CD, de la Chapelle A. Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes. *Proc Natl Acad Sci U S A*. 2004;101(11):3915-

20. Epub 2004/03/10. doi: 10.1073/pnas.0400272101. PubMed PMID: 15007164; PMCID: PMC374344.
93. Rovigatti U, Watson DK, Yunis JJ. Amplification and rearrangement of Hu-ets-1 in leukemia and lymphoma with involvement of 11q23. *Science*. 1986;232(4748):398-400. Epub 1986/04/18. doi: 10.1126/science.3457468. PubMed PMID: 3457468.
94. Chae H, Kim M, Lim J, Kim Y, Han K, Lee S. B lymphoblastic leukemia with ETV6 amplification. *Cancer Genet Cytogenet*. 2010;203(2):284-7. Epub 2010/12/16. doi: 10.1016/j.cancergencyto.2010.08.004. PubMed PMID: 21156245.
95. Xu D, Dwyer J, Li H, Duan W, Liu JP. Ets2 maintains hTERT gene expression and breast cancer cell proliferation by interacting with c-Myc. *J Biol Chem*. 2008;283(35):23567-80. Epub 2008/07/01. doi: 10.1074/jbc.M800790200. PubMed PMID: 18586674; PMCID: PMC3259778.
96. Hollenhorst PC, Ferris MW, Hull MA, Chae H, Kim S, Graves BJ. Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev*. 2011;25(20):2147-57. Epub 2011/10/21. doi: 10.1101/gad.17546311. PubMed PMID: 22012618; PMCID: 3205585.
97. Wasylyk B, Hagman J, Gutierrez-Hartmann A. Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. *Trends Biochem Sci*. 1998;23(6):213-6. Epub 1998/06/30. doi: 10.1016/s0968-0004(98)01211-0. PubMed PMID: 9644975.
98. Kim WY, Sieweke M, Ogawa E, Wee HJ, Englmeier U, Graf T, Ito Y. Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *EMBO J*. 1999;18(6):1609-20. Epub 1999/03/17. doi: 10.1093/emboj/18.6.1609. PubMed PMID: 10075931; PMCID: PMC1171248.
99. Yamada T, Abe M, Higashi T, Yamamoto H, Kihara-Negishi F, Sakurai T, Shirai T, Oikawa T. Lineage switch induced by overexpression of Ets family transcription factor PU.1 in murine erythroleukemia cells. *Blood*. 2001;97(8):2300-7. Epub 2001/04/06. doi: 10.1182/blood.v97.8.2300. PubMed PMID: 11290591.
100. Burda P, Curik N, Kokavec J, Basova P, Mikulenková D, Skoultchi AI, Zavadil J, Stopka T. PU.1 activation relieves GATA-1-mediated repression of Cebpa and Cbfb during leukemia differentiation. *Mol Cancer Res*. 2009;7(10):1693-703. Epub 2009/10/15. doi: 10.1158/1541-7786.MCR-09-0031. PubMed PMID: 19825991; PMCID: PMC3193075.
101. Zou D, Yang X, Tan Y, Wang P, Zhu X, Yang W, Jia X, Zhang J, Wang K. Regulation of the hematopoietic cell kinase (HCK) by PML/RARalpha and PU.1 in acute promyelocytic leukemia. *Leuk Res*. 2012;36(2):219-23. Epub 2011/10/14. doi: 10.1016/j.leukres.2011.09.012. PubMed PMID: 21993313.
102. Hu Z, Gu X, Baraoidan K, Ibanez V, Sharma A, Kadkol S, Munker R, Ackerman S, Nucifora G, Sauntharajah Y. RUNX1 regulates corepressor interactions of PU.1. *Blood*. 2011;117(24):6498-508. Epub 2011/04/27. doi: 10.1182/blood-2010-10-312512. PubMed PMID: 21518930; PMCID: PMC3123020.
103. Hsing M, Wang Y, Rennie PS, Cox ME, Cherkasov A. ETS transcription factors as emerging drug targets in cancer. *Med Res Rev*. 2020;40(1):413-30. Epub 2019/03/31. doi: 10.1002/med.21575. PubMed PMID: 30927317.
104. Mohamed AA, Tan SH, Mikhalkevich N, Ponniah S, Vasioukhin V, Bieberich CJ, Sesterhenn IA, Dobi A, Srivastava S, Sreenath TL. Ets family protein, erg expression in developing and adult mouse tissues by a highly specific monoclonal antibody. *J Cancer*. 2010;1:197-208. Epub 2010/11/10. doi: 10.7150/jca.1.197. PubMed PMID: 21060730; PMCID: PMC2974237.
105. Pop R, Shearstone JR, Shen Q, Liu Y, Hallstrom K, Koulis M, Gribnau J, Socolovsky M. A key commitment step in erythropoiesis is synchronized with the cell cycle clock through mutual

inhibition between PU.1 and S-phase progression. *PLoS Biol.* 2010;8(9). Epub 2010/09/30. doi: 10.1371/journal.pbio.1000484. PubMed PMID: 20877475; PMCID: PMC2943437.

106. Butler MS, Roshan-Moniri M, Hsing M, Lau D, Kim A, Yen P, Mroczek M, Nouri M, Lien S, Axerio-Cilies P, Dalal K, Yau C, Ghaidi F, Guo Y, Yamazaki T, Lawn S, Gleave ME, Gregory-Evans CY, McIntosh LP, Cox ME, Rennie PS, Cherkasov A. Discovery and characterization of small molecules targeting the DNA-binding ETS domain of ERG in prostate cancer. *Oncotarget.* 2017;8(26):42438-54. Epub 2017/05/04. doi: 10.18632/oncotarget.17124. PubMed PMID: 28465491; PMCID: PMC5522078.

107. Santos J, Mesquita D, Barros-Silva JD, Jeronimo C, Henrique R, Morais A, Paulo P, Teixeira MR. Uncovering potential downstream targets of oncogenic GRPR overexpression in prostate carcinomas harboring ETS rearrangements. *Oncoscience.* 2015;2(5):497-507. Epub 2015/06/23. doi: 10.18632/oncoscience.142. PubMed PMID: 26097883; PMCID: PMC4468336.

108. Ban F, Dalal K, Li H, LeBlanc E, Rennie PS, Cherkasov A. Best Practices of Computer-Aided Drug Discovery: Lessons Learned from the Development of a Preclinical Candidate for Prostate Cancer with a New Mechanism of Action. *J Chem Inf Model.* 2017;57(5):1018-28. Epub 2017/04/26. doi: 10.1021/acs.jcim.7b00137. PubMed PMID: 28441481.

109. Mohamed AA, Xavier CP, Sukumar G, Tan SH, Ravindranath L, Seraj N, Kumar V, Sreenath T, McLeod DG, Petrovics G, Rosner IL, Srivastava M, Strovel J, Malhotra SV, LaRonde NA, Dobi A, Dalgard CL, Srivastava S. Identification of a Small Molecule That Selectively Inhibits ERG-Positive Cancer Cell Growth. *Cancer Res.* 2018;78(13):3659-71. Epub 2018/05/02. doi: 10.1158/0008-5472.CAN-17-2949. PubMed PMID: 29712692.

110. Liu B, Bashkin JK, Poon GMK, Wang S, Wang S, Wilson WD. Modulating DNA by polyamides to regulate transcription factor PU.1-DNA binding interactions. *Biochimie.* 2019;167:1-11. Epub 2019/08/25. doi: 10.1016/j.biochi.2019.08.009. PubMed PMID: 31445072.

111. Munde M, Poon, G. M., & Wilson, W. D. Probing the electrostatics and pharmacological modulation of sequence-specific binding by the DNA-binding domain of the ETS family transcription factor PU.1: a binding affinity and kinetics investigation. *Journal of molecular biology.* 2013;425:1655-69.

112. Poon GM. DNA binding regulates the self-association of the ETS domain of PU.1 in a sequence-dependent manner. *Biochemistry.* 2012;51(20):4096-107. doi: 10.1021/bi300331v. PubMed PMID: 22533913.

113. Poon GM. Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. *The Journal of biological chemistry.* 2012;287(22):18297-307. doi: 10.1074/jbc.M112.342345. PubMed PMID: 22474303; PMCID: PMC3365747.

114. Poon GM, Gross P, Macgregor RB, Jr. The sequence-specific association of the ETS domain of murine PU.1 with DNA exhibits unusual energetics. *Biochemistry.* 2002;41(7):2361-71. Epub 2002/02/14. doi: 10.1021/bi015572q. PubMed PMID: 11841229.

115. Poon GM, Macgregor RB, Jr. Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. *J Mol Biol.* 2003;328(4):805-19. Epub 2003/05/06. doi: 10.1016/s0022-2836(03)00362-0. PubMed PMID: 12729756.

116. Poon GM, Macgregor RB, Jr. A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1. *J Mol Biol.* 2004;335(1):113-27. Epub 2003/12/09. doi: 10.1016/j.jmb.2003.09.046. PubMed PMID: 14659744.

117. Stephens DC, Kim, H. M., Kumar, A., Farahat, A. A., Boykin, D. W., & Poon, G. M. . Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis. *Nucleic acids research*. 2016;44(9):4005-13. doi: 10.1093/nar/gkw229.
118. Wang S, Linde MH, Munde M, Carvalho VD, Wilson WD, Poon GM. Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1. *J Biol Chem*. 2014;289(31):21605-16. Epub 2014/06/24. doi: 10.1074/jbc.M114.575340. PubMed PMID: 24952944; PMCID: PMC4118120.
119. Scott EW, Simon, M. C., Anastasi, J., & Singh, H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science*. 1994;265(5178):1573-7. Epub 1994/09/09. doi: 10.1126/science.8079170. PubMed PMID: 8079170.
120. Iwasaki H, Somoza, C., Shigematsu, H., Duprez, E. A., Iwasaki-Arai, J., Mizuno, S., Arinobu, Y., Geary, K., Zhang, P., Dayaram, T., Fenyus, M. L., Elf, S., Chan, S., Kastner, P., Huettner, C. S., Murray, R., Tenen, D. G., & Akashi, K. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood*. 2005;106(5):1590-600. doi: 10.1182/blood-2005-03-0860.
121. Rosenbauer F, Wagner, K., Kutok, J. L., Iwasaki, H., Le Beau, M. M., Okuno, Y., Akashi, K., Fiering, S., & Tenen, D. G. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet*. 2004;36(6):624-30. Epub 2004/05/18. doi: 10.1038/ng1361. PubMed PMID: 15146183.
122. Eisenbeis CF, Singh, H., & Storb, U. . PU.1 is a component of a multiprotein complex which binds an essential site in the murine immunoglobulin lambda 2-4 enhancer. *Molecular and cellular biology*. 1993;13(10):6452-61. doi: 10.1128/mcb.13.10.6452.
123. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem*. 2011;80:437-71. Epub 2011/05/10. doi: 10.1146/annurev.biochem.79.081507.103945. PubMed PMID: 21548782; PMCID: PMC5568663.
124. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev*. 2007;21(15):1882-94. Epub 2007/07/27. doi: 10.1101/gad.1561707. PubMed PMID: 17652178; PMCID: PMC1935027.
125. Will B, Vogler, T.O., Narayanagari, W., Bartholdy, B., Todorova, T., Ferreira, M. S., Chen, J., Yu, Y., Mayer, J., Barreyro, L., Carvajal, L., Neriah, D.B., Roth, M., Oers, J., Schaezlein, S., McMahon, C., Edelmann, W., Verma, A., and Steidl, U. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. . *Nature medicine*. 2015;21(10):1172-81. doi: 10.1038/nm.3936.
126. Staber PB, Zhang P, Ye M, Welner RS, Levantini E, Di Ruscio A, Ebralidze AK, Bach C, Zhang H, Zhang J, Vanura K, Delwel R, Yang H, Huang G, Tenen DG. The Runx-PU.1 pathway preserves normal and AML/ETO9a leukemic stem cells. *Blood*. 2014;124(15):2391-9. Epub 2014/09/05. doi: 10.1182/blood-2014-01-550855. PubMed PMID: 25185713; PMCID: PMC4192750.
127. Sive JI, Basilico S, Hannah R, Kinston SJ, Calero-Nieto FJ, Gottgens B. Genome-scale definition of the transcriptional programme associated with compromised PU.1 activity in acute myeloid leukaemia. *Leukemia*. 2016;30(1):14-23. Epub 2015/07/02. doi: 10.1038/leu.2015.172. PubMed PMID: 26126967; PMCID: PMC4705427.

128. Rosenbauer F, Koschmieder, S., Steidl, U., & Tenen, D. G. . Effect of transcription-factor concentrations on leukemic stem cells. *Blood*. 2005;106(5):1519-24. doi: 10.1182/blood-2005-02-0717.
129. Szymczyna BR, Arrowsmith CH. DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. *The Journal of biological chemistry*. 2000;275(37):28363-70. Epub 2000/06/27. doi: 10.1074/jbc.M004294200. PubMed PMID: 10867009.
130. Jayaram B, Jain T. The role of water in protein-DNA recognition. *Annu Rev Biophys Biomol Struct*. 2004;33:343-61. Epub 2004/05/14. doi: 10.1146/annurev.biophys.33.110502.140414. PubMed PMID: 15139817.
131. Ho SN. Intracellular water homeostasis and the mammalian cellular osmotic stress response. *J Cell Physiol*. 2006;206(1):9-15. Epub 2005/06/21. doi: 10.1002/jcp.20445. PubMed PMID: 15965902.
132. Go WY, Liu X, Roti MA, Liu F, Ho SN. NFAT5/TonEBP mutant mice define osmotic stress as a critical feature of the lymphoid microenvironment. *Proc Natl Acad Sci U S A*. 2004;101(29):10673-8. Epub 2004/07/13. doi: 10.1073/pnas.0403139101. PubMed PMID: 15247420; PMCID: PMC489993.
133. Lima AF, May G, Diaz-Colunga J, Pedreiro S, Paiva A, Ferreira L, Enver T, Iborra FJ, Pires das Neves R. Osmotic modulation of chromatin impacts on efficiency and kinetics of cell fate modulation. *Sci Rep*. 2018;8(1):7210. Epub 2018/05/10. doi: 10.1038/s41598-018-25517-2. PubMed PMID: 29740078; PMCID: PMC5940679.
134. Esaki S, Evich MG, Erlitzki N, Germann MW, Poon GMK. Multiple DNA-binding modes for the ETS family transcription factor PU.1. *J Biol Chem*. 2017;292(39):16044-54. Epub 2017/08/10. doi: 10.1074/jbc.M117.798207. PubMed PMID: 28790174; PMCID: PMC5625037.
135. Minderjahn J, Schmidt A, Fuchs A, Schill R, Raithel J, Babina M, Schmidl C, Gebhard C, Schmidhofer S, Mendes K, Ratermann A, Glatz D, Nutzelt M, Edinger M, Hoffmann P, Spang R, Langst G, Imhof A, Rehli M. Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. *Nat Commun*. 2020;11(1):402. Epub 2020/01/23. doi: 10.1038/s41467-019-13960-2. PubMed PMID: 31964861; PMCID: PMC6972792.
136. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc*. 2007;2(8):1849-61. Epub 2007/08/19. doi: 10.1038/nprot.2007.249. PubMed PMID: 17703195; PMCID: PMC2757439.
137. Brenowitz M, Senear DF, Shea MA, Ackers GK. Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol*. 1986;130:132-81. Epub 1986/01/01. doi: 10.1016/0076-6879(86)30011-9. PubMed PMID: 3773731.
138. Connaghan-Jones KD, Moody AD, Bain DL. Quantitative DNase footprint titration: a tool for analyzing the energetics of protein-DNA interactions. *Nat Protoc*. 2008;3(5):900-14. Epub 2008/05/03. doi: 10.1038/nprot.2008.53. PubMed PMID: 18451798.
139. Crane-Robinson C, Myers FA, Hebbes TR, Clayton AL, Thorne AW. Chromatin immunoprecipitation assays in acetylation mapping of higher eukaryotes. *Methods Enzymol*. 1999;304:533-47. Epub 1999/06/18. doi: 10.1016/s0076-6879(99)04031-8. PubMed PMID: 10372380.
140. Ma H, Shang Y, Lee DY, Stallcup MR. Study of nuclear receptor-induced transcription complex assembly and histone modification by chromatin immunoprecipitation assays. *Methods Enzymol*. 2003;364:284-96. Epub 2003/11/25. doi: 10.1016/s0076-6879(03)64016-4. PubMed PMID: 14631851.

141. Nie L, Vazquez AE, Yamoah EN. Identification of transcription factor-DNA interactions using chromatin immunoprecipitation assays. *Methods Mol Biol.* 2009;493:311-21. Epub 2008/10/08. doi: 10.1007/978-1-59745-523-7_19. PubMed PMID: 18839356; PMCID: PMC2822710.
142. Perez-Romero P, Imperiale MJ. Assaying protein-DNA interactions in vivo and in vitro using chromatin immunoprecipitation and electrophoretic mobility shift assays. *Methods Mol Med.* 2007;131:123-39. Epub 2007/07/28. doi: 10.1007/978-1-59745-277-9_10. PubMed PMID: 17656780.
143. Pillai S, Dasgupta P, Chellappan SP. Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications in vivo. *Methods Mol Biol.* 2009;523:323-39. Epub 2009/04/22. doi: 10.1007/978-1-59745-190-1_22. PubMed PMID: 19381928.
144. Pillai S, Dasgupta P, Chellappan SP. Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications in vivo. *Methods Mol Biol.* 2015;1288:429-46. Epub 2015/04/02. doi: 10.1007/978-1-4939-2474-5_25. PubMed PMID: 25827895.
145. Gade P, Kalvakolanu DV. Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. *Methods Mol Biol.* 2012;809:85-104. Epub 2011/11/25. doi: 10.1007/978-1-61779-376-9_6. PubMed PMID: 22113270; PMCID: PMC3891665.
146. Davies JO, Oudelaar AM, Higgs DR, Hughes JR. How best to identify chromosomal interactions: a comparison of approaches. *Nat Methods.* 2017;14(2):125-34. Epub 2017/02/01. doi: 10.1038/nmeth.4146. PubMed PMID: 28139673.
147. Qi Y, Zhang B. Predicting three-dimensional genome organization with chromatin states. *PLoS Comput Biol.* 2019;15(6):e1007024. Epub 2019/06/11. doi: 10.1371/journal.pcbi.1007024. PubMed PMID: 31181064; PMCID: PMC6586364.
148. Di Pierro M, Cheng RR, Lieberman Aiden E, Wolynes PG, Onuchic JN. De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci U S A.* 2017;114(46):12126-31. Epub 2017/11/01. doi: 10.1073/pnas.1714980114. PubMed PMID: 29087948; PMCID: PMC5699090.
149. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet.* 2012;13(12):840-52. Epub 2012/10/24. doi: 10.1038/nrg3306. PubMed PMID: 23090257; PMCID: PMC3591838.
150. Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform.* 2017;18(2):279-90. Epub 2016/03/17. doi: 10.1093/bib/bbw023. PubMed PMID: 26979602; PMCID: PMC5444249.
151. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shores N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813-31. Epub 2012/09/08. doi: 10.1101/gr.136184.111. PubMed PMID: 22955991; PMCID: PMC3431496.
152. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* 2009;10(10):669-80. Epub 2009/09/09. doi: 10.1038/nrg2641. PubMed PMID: 19736561; PMCID: PMC3191340.

153. Gonzales NR, Schuck P, Schlom J, Kashmiri SV. Surface plasmon resonance-based competition assay to assess the sera reactivity of variants of humanized antibodies. *J Immunol Methods*. 2002;268(2):197-210. Epub 2002/09/07. doi: 10.1016/s0022-1759(02)00205-3. PubMed PMID: 12215388.
154. Schuck P. Reliable determination of binding affinity and kinetics using surface plasmon resonance biosensors. *Curr Opin Biotechnol*. 1997;8(4):498-502. Epub 1997/08/01. doi: 10.1016/s0958-1669(97)80074-2. PubMed PMID: 9265731.
155. Schuck P. Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annu Rev Biophys Biomol Struct*. 1997;26:541-66. Epub 1997/01/01. doi: 10.1146/annurev.biophys.26.1.541. PubMed PMID: 9241429.
156. Bayat P, Nosrati R, Alibolandi M, Rafatpanah H, Abnous K, Khedri M, Ramezani M. SELEX methods on the road to protein targeting with nucleic acid aptamers. *Biochimie*. 2018;154:132-55. Epub 2018/09/09. doi: 10.1016/j.biochi.2018.09.001. PubMed PMID: 30193856.
157. Urak KT, Shore S, Rockey WM, Chen SJ, McCaffrey AP, Giangrande PH. In vitro RNA SELEX for the generation of chemically-optimized therapeutic RNA drugs. *Methods*. 2016;103:167-74. Epub 2016/03/15. doi: 10.1016/j.ymeth.2016.03.003. PubMed PMID: 26972786; PMCID: PMC4921298.
158. Djordjevic M. SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. *Biomol Eng*. 2007;24(2):179-89. Epub 2007/04/13. doi: 10.1016/j.bioeng.2007.03.001. PubMed PMID: 17428731.
159. Djordjevic M. Inferring protein-DNA interaction parameters from SELEX experiments. *Methods Mol Biol*. 2010;674:195-211. Epub 2010/09/10. doi: 10.1007/978-1-60761-854-6_12. PubMed PMID: 20827593.
160. Djordjevic M, Sengupta AM. Quantitative modeling and data analysis of SELEX experiments. *Phys Biol*. 2005;3(1):13-28. Epub 2006/04/04. doi: 10.1088/1478-3975/3/1/002. PubMed PMID: 16582458.
161. Ozer A, Pagano JM, Lis JT. New Technologies Provide Quantum Changes in the Scale, Speed, and Success of SELEX Methods and Aptamer Characterization. *Mol Ther Nucleic Acids*. 2014;3:e183. Epub 2014/08/06. doi: 10.1038/mtna.2014.34. PubMed PMID: 25093707; PMCID: PMC4221594.
162. Sekar RB, Periasamy A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol*. 2003;160(5):629-33. Epub 2003/03/05. doi: 10.1083/jcb.200210140. PubMed PMID: 12615908; PMCID: PMC2173363.
163. Blouin S, Craggs TD, Lafontaine DA, Penedo JC. Functional studies of DNA-protein interactions using FRET techniques. *Methods Mol Biol*. 2009;543:475-502. Epub 2009/04/21. doi: 10.1007/978-1-60327-015-1_28. PubMed PMID: 19378182.
164. Russu IM. Studying DNA-protein interactions using NMR. *Trends Biotechnol*. 1991;9(3):96-104. Epub 1991/03/01. doi: 10.1016/0167-7799(91)90032-d. PubMed PMID: 1367095.
165. Campagne S, Gervais V, Milon A. Nuclear magnetic resonance analysis of protein-DNA interactions. *J R Soc Interface*. 2011;8(61):1065-78. Epub 2011/03/11. doi: 10.1098/rsif.2010.0543. PubMed PMID: 21389020; PMCID: PMC3119875.

166. Liu ZQ, Mahmood T, Yang PC. Western blot: technique, theory and trouble shooting. *N Am J Med Sci.* 2014;6(3):160. Epub 2014/04/18. doi: 10.4103/1947-2714.128482. PubMed PMID: 24741558; PMCID: PMC3978942.
167. Mahmood T, Yang PC. Western blot: technique, theory, and trouble shooting. *N Am J Med Sci.* 2012;4(9):429-34. Epub 2012/10/11. doi: 10.4103/1947-2714.100998. PubMed PMID: 23050259; PMCID: PMC3456489.
168. Bass JJ, Wilkinson DJ, Rankin D, Phillips BE, Szewczyk NJ, Smith K, Atherton PJ. An overview of technical considerations for Western blotting applications to physiological research. *Scand J Med Sci Sports.* 2017;27(1):4-25. Epub 2016/06/07. doi: 10.1111/sms.12702. PubMed PMID: 27263489; PMCID: PMC5138151.
169. Aydin S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides.* 2015;72:4-15. Epub 2015/04/25. doi: 10.1016/j.peptides.2015.04.012. PubMed PMID: 25908411.
170. Alhajj M, Farhana A. Enzyme Linked Immunosorbent Assay. *StatPearls. Treasure Island (FL)2021.*
171. Underwood KF, Mochin MT, Brusgard JL, Choe M, Gnatt A, Passaniti A. A quantitative assay to study protein:DNA interactions, discover transcriptional regulators of gene expression, and identify novel anti-tumor agents. *J Vis Exp.* 2013(78). Epub 2013/09/12. doi: 10.3791/50512. PubMed PMID: 24022461; PMCID: PMC3857129.
172. Jiang T, Xing B, Rao J. Recent developments of biological reporter technology for detecting gene expression. *Biotechnol Genet Eng Rev.* 2008;25:41-75. Epub 2008/01/01. doi: 10.5661/bger-25-41. PubMed PMID: 21412349.
173. Rosenthal N. Identification of regulatory elements of cloned genes with functional assays. *Methods Enzymol.* 1987;152:704-20. Epub 1987/01/01. doi: 10.1016/0076-6879(87)52075-4. PubMed PMID: 2889129.
174. Prost E, Moore DD. CAT vectors for analysis of eukaryotic promoters and enhancers. *Gene.* 1986;45(1):107-11. Epub 1986/01/01. doi: 10.1016/0378-1119(86)90138-1. PubMed PMID: 3023200.
175. Bronstein I, Fortin J, Stanley PE, Stewart GS, Kricka LJ. Chemiluminescent and bioluminescent reporter gene assays. *Anal Biochem.* 1994;219(2):169-81. Epub 1994/06/01. doi: 10.1006/abio.1994.1254. PubMed PMID: 8080073.
176. Cullen BR, Malim MH. Secreted placental alkaline phosphatase as a eukaryotic reporter gene. *Methods Enzymol.* 1992;216:362-8. Epub 1992/01/01. doi: 10.1016/0076-6879(92)16033-g. PubMed PMID: 1479908.
177. Becerra-Arteaga A, Mason HS, Shuler ML. Production, secretion, and stability of human secreted alkaline phosphatase in tobacco NT1 cell suspension cultures. *Biotechnol Prog.* 2006;22(6):1643-9. Epub 2006/12/02. doi: 10.1021/bp060151r. PubMed PMID: 17137313.
178. Coleman JE. Structure and mechanism of alkaline phosphatase. *Annu Rev Biophys Biomol Struct.* 1992;21:441-83. Epub 1992/01/01. doi: 10.1146/annurev.bb.21.060192.002301. PubMed PMID: 1525473.
179. Coleman JE, Gettins P. Alkaline phosphatase, solution structure, and mechanism. *Adv Enzymol Relat Areas Mol Biol.* 1983;55:381-452. Epub 1983/01/01. doi: 10.1002/9780470123010.ch5. PubMed PMID: 6312783.
180. Grunwald SK, Krueger KJ. Improvement of student understanding of how kinetic data facilitates the determination of amino acid catalytic function through an alkaline phosphatase

structure/mechanism bioinformatics exercise. *Biochem Mol Biol Educ.* 2008;36(1):9-15. Epub 2008/01/01. doi: 10.1002/bmb.120. PubMed PMID: 21591153.

181. Kim A, Benning MM, OkLee S, Quinn J, Martin BM, Holden HM, Dunaway-Mariano D. Divergence of chemical function in the alkaline phosphatase superfamily: structure and mechanism of the P-C bond cleaving enzyme phosphonoacetate hydrolase. *Biochemistry.* 2011;50(17):3481-94. Epub 2011/03/04. doi: 10.1021/bi200165h. PubMed PMID: 21366328; PMCID: PMC3102601.

182. Tibbitts TT, Xu X, Kantrowitz ER. Kinetics and crystal structure of a mutant *Escherichia coli* alkaline phosphatase (Asp-369-->Asn): a mechanism involving one zinc per active site. *Protein Sci.* 1994;3(11):2005-14. Epub 1994/11/01. doi: 10.1002/pro.5560031113. PubMed PMID: 7703848; PMCID: PMC2142653.

183. Fowler AV, Zabin I. Purification, structure, and properties of hybrid beta-galactosidase proteins. *J Biol Chem.* 1983;258(23):14354-8. Epub 1983/12/10. PubMed PMID: 6417139.

184. Holmes ML, Dyll-Smith ML. Sequence and expression of a halobacterial beta-galactosidase gene. *Mol Microbiol.* 2000;36(1):114-22. Epub 2000/04/12. doi: 10.1046/j.1365-2958.2000.01832.x. PubMed PMID: 10760168.

185. Broome-Smith JK, Spratt BG. A vector for the construction of translational fusions to TEM beta-lactamase and the analysis of protein export signals and membrane protein topology. *Gene.* 1986;49(3):341-9. Epub 1986/01/01. doi: 10.1016/0378-1119(86)90370-7. PubMed PMID: 3552888.

186. Focareta T, Manning PA. Genetic analysis of the export of an extracellular DNase of *Vibrio cholerae* using DNase-beta-lactamase fusions. *Gene.* 1991;108(1):31-7. Epub 1991/12/01. doi: 10.1016/0378-1119(91)90484-s. PubMed PMID: 1761228.

187. Mueller KE, Fields KA. Application of beta-lactamase reporter fusions as an indicator of effector protein secretion during infections with the obligate intracellular pathogen *Chlamydia trachomatis*. *PLoS One.* 2015;10(8):e0135295. Epub 2015/08/11. doi: 10.1371/journal.pone.0135295. PubMed PMID: 26258949; PMCID: PMC4530969.

188. Tan R, Jiang X, Jackson A, Jin P, Yang J, Lee E, Duggan B, Stuve LL, Fu GK. *E. coli* selection of human genes encoding secreted and membrane proteins based on cDNA fusions to a leaderless beta-lactamase reporter. *Genome Res.* 2003;13(8):1938-43. Epub 2003/07/19. doi: 10.1101/gr.1000903. PubMed PMID: 12869575; PMCID: PMC403786.

189. Zhang J, Barquera B, Gennis RB. Gene fusions with beta-lactamase show that subunit I of the cytochrome bd quinol oxidase from *E. coli* has nine transmembrane helices with the O₂ reactive site near the periplasmic surface. *FEBS Lett.* 2004;561(1-3):58-62. Epub 2004/03/12. doi: 10.1016/S0014-5793(04)00125-5. PubMed PMID: 15013751.

190. Zlokarnik G. Fusions to beta-lactamase as a reporter for gene expression in live mammalian cells. *Methods Enzymol.* 2000;326:221-44. Epub 2000/10/19. doi: 10.1016/s0076-6879(00)26057-6. PubMed PMID: 11036645.

191. Auld DS, Lovell S, Thorne N, Lea WA, Maloney DJ, Shen M, Rai G, Battaile KP, Thomas CJ, Simeonov A, Hanzlik RP, Inglese J. Molecular basis for the high-affinity binding and stabilization of firefly luciferase by PTC124. *Proc Natl Acad Sci U S A.* 2010;107(11):4878-83. Epub 2010/03/03. doi: 10.1073/pnas.0909141107. PubMed PMID: 20194791; PMCID: PMC2841876.

192. Thorne N, Inglese J, Auld DS. Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chem Biol.* 2010;17(6):646-57. Epub

2010/07/09. doi: 10.1016/j.chembiol.2010.05.012. PubMed PMID: 20609414; PMCID: PMC2925662.

193. Thorne N, Shen M, Lea WA, Simeonov A, Lovell S, Auld DS, Inglese J. Firefly luciferase in chemical biology: a compendium of inhibitors, mechanistic evaluation of chemotypes, and suggested use as a reporter. *Chem Biol.* 2012;19(8):1060-72. Epub 2012/08/28. doi: 10.1016/j.chembiol.2012.07.015. PubMed PMID: 22921073; PMCID: PMC3449281.

194. Wurdinger T, Badr C, Pike L, de Kleine R, Weissleder R, Breakefield XO, Tannous BA. A secreted luciferase for ex vivo monitoring of in vivo processes. *Nat Methods.* 2008;5(2):171-3. Epub 2008/01/22. doi: 10.1038/nmeth.1177. PubMed PMID: 18204457; PMCID: PMC2699561.

195. Michalet X, Pinaud FF, Bentolila LA, Tsay JM, Doose S, Li JJ, Sundaresan G, Wu AM, Gambhir SS, Weiss S. Quantum dots for live cells, in vivo imaging, and diagnostics. *Science.* 2005;307(5709):538-44. Epub 2005/02/01. doi: 10.1126/science.1104274. PubMed PMID: 15681376; PMCID: PMC1201471.

196. Bruchez M, Jr., Moronne M, Gin P, Weiss S, Alivisatos AP. Semiconductor nanocrystals as fluorescent biological labels. *Science.* 1998;281(5385):2013-6. Epub 1998/09/25. doi: 10.1126/science.281.5385.2013. PubMed PMID: 9748157.

197. Fu A, Gu W, Boussert B, Koski K, Gerion D, Manna L, Le Gros M, Larabell CA, Alivisatos AP. Semiconductor quantum rods as single molecule fluorescent biological labels. *Nano Lett.* 2007;7(1):179-82. Epub 2007/01/11. doi: 10.1021/nl0626434. PubMed PMID: 17212460; PMCID: PMC3984543.

198. Zuverink M, Barbieri JT. From GFP to beta-lactamase: advancing intact cell imaging for toxins and effectors. *Pathog Dis.* 2015;73(9):ftv097. Epub 2015/10/27. doi: 10.1093/femspd/ftv097. PubMed PMID: 26500183; PMCID: PMC4732026.

199. Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M, Remington SJ. Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc Natl Acad Sci U S A.* 1997;94(6):2306-11. Epub 1997/03/18. doi: 10.1073/pnas.94.6.2306. PubMed PMID: 9122190; PMCID: PMC20083.

200. Dickson RM, Cubitt AB, Tsien RY, Moerner WE. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature.* 1997;388(6640):355-8. Epub 1997/07/24. doi: 10.1038/41048. PubMed PMID: 9237752.

201. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A.* 1994;91(26):12501-4. Epub 1994/12/20. doi: 10.1073/pnas.91.26.12501. PubMed PMID: 7809066; PMCID: PMC45466.

202. Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol.* 1996;6(2):178-82. Epub 1996/02/01. doi: 10.1016/s0960-9822(02)00450-5. PubMed PMID: 8673464.

203. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science.* 1996;273(5280):1392-5. Epub 1996/09/06. doi: 10.1126/science.273.5280.1392. PubMed PMID: 8703075.

204. Tsien RY. The green fluorescent protein. *Annu Rev Biochem.* 1998;67:509-44. Epub 1998/10/06. doi: 10.1146/annurev.biochem.67.1.509. PubMed PMID: 9759496.

205. Wachter RM, King BA, Heim R, Kallio K, Tsien RY, Boxer SG, Remington SJ. Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. *Biochemistry.* 1997;36(32):9759-65. Epub 1997/08/12. doi: 10.1021/bi970563w. PubMed PMID: 9245407.

206. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci.* 1995;20(11):448-55. Epub 1995/11/01. doi: 10.1016/s0968-0004(00)89099-4. PubMed PMID: 8578587.
207. Pahl HL, Scheibe RJ, Zhang DE, Chen HM, Galson DL, Maki RA, Tenen DG. The proto-oncogene PU.1 regulates expression of the myeloid-specific CD11b promoter. *J Biol Chem.* 1993;268(7):5014-20. Epub 1993/03/05. PubMed PMID: 8095266.
208. Zhang DE, Hetherington CJ, Chen HM, Tenen DG. The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. *Mol Cell Biol.* 1994;14(1):373-81. Epub 1994/01/01. doi: 10.1128/mcb.14.1.373-381.1994. PubMed PMID: 8264604; PMCID: PMC358386.
209. Hoffman RM. The multiple uses of fluorescent proteins to visualize cancer in vivo. *Nat Rev Cancer.* 2005;5(10):796-806. Epub 2005/10/01. doi: 10.1038/nrc1717. PubMed PMID: 16195751.
210. Ntziachristos V. Going deeper than microscopy: the optical imaging frontier in biology. *Nat Methods.* 2010;7(8):603-14. Epub 2010/08/03. doi: 10.1038/nmeth.1483. PubMed PMID: 20676081.
211. Filonov GS, Piatkevich KD, Ting LM, Zhang J, Kim K, Verkhusha VV. Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat Biotechnol.* 2011;29(8):757-61. Epub 2011/07/19. doi: 10.1038/nbt.1918. PubMed PMID: 21765402; PMCID: PMC3152693.
212. Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, Kim MK, Shin BA, Choi SY. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One.* 2011;6(4):e18556. Epub 2011/05/24. doi: 10.1371/journal.pone.0018556. PubMed PMID: 21602908; PMCID: PMC3084703.
213. Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang CC, Kain SR. Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem.* 1998;273(52):34970-5. Epub 1998/12/18. doi: 10.1074/jbc.273.52.34970. PubMed PMID: 9857028.
214. Tiscornia G, Singer O, Verma IM. Production and purification of lentiviral vectors. *Nat Protoc.* 2006;1(1):241-5. Epub 2007/04/05. doi: 10.1038/nprot.2006.37. PubMed PMID: 17406239.
215. Albrecht AV, Kim HM, Poon GMK. Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach. *Nucleic Acids Res.* 2018;46(20):10577-88. Epub 2018/10/09. doi: 10.1093/nar/gky894. PubMed PMID: 30295801; PMCID: PMC6237740.
216. Billaud M, Isselbacher KJ, Bernards R. A dominant-negative mutant of Max that inhibits sequence-specific DNA binding by Myc proteins. *Proc Natl Acad Sci U S A.* 1993;90(7):2739-43. Epub 1993/04/01. doi: 10.1073/pnas.90.7.2739. PubMed PMID: 8464883; PMCID: PMC46171.
217. Christian JL, Edelstein NG, Moon RT. Overexpression of wild-type and dominant negative mutant vimentin subunits in developing *Xenopus* embryos. *New Biol.* 1990;2(8):700-11. Epub 1990/08/01. PubMed PMID: 1704253.
218. Sakurai A, Miyamoto T, Refetoff S, DeGroot LJ. Dominant negative transcriptional regulation by a mutant thyroid hormone receptor-beta in a family with generalized resistance to thyroid hormone. *Mol Endocrinol.* 1990;4(12):1988-94. Epub 1990/12/01. doi: 10.1210/mend-4-12-1988. PubMed PMID: 2082193.
219. Xhani S, Lee S, Kim HM, Wang S, Esaki S, Ha VLT, Khanezarrin M, Fernandez GL, Albrecht AV, Aramini JM, Germann MW, Poon GMK. Intrinsic disorder controls two functionally

- distinct dimers of the master transcription factor PU.1. *Sci Adv.* 2020;6(8):eaay3178. Epub 2020/03/05. doi: 10.1126/sciadv.aay3178. PubMed PMID: 32128405; PMCID: PMC7034988.
220. Garvie CW, Pufall MA, Graves BJ, Wolberger C. Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J Biol Chem.* 2002;277(47):45529-36. Epub 2002/09/11. doi: 10.1074/jbc.M206327200. PubMed PMID: 12221090.
221. Graves BJ, Cowley DO, Goetz TL, Petersen JM, Jonsen MD, Gillespie ME. Autoinhibition as a transcriptional regulatory mechanism. *Cold Spring Harb Symp Quant Biol.* 1998;63:621-9. Epub 1999/06/29. doi: 10.1101/sqb.1998.63.621. PubMed PMID: 10384327.
222. Zhu J, Emerson SG. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene.* 2002;21(21):3295-313. Epub 2002/05/29. doi: 10.1038/sj.onc.1205318. PubMed PMID: 12032771.
223. Orkin SH. Embryonic stem cells and transgenic mice in the study of hematopoiesis. *Int J Dev Biol.* 1998;42(7):927-34. Epub 1998/12/16. PubMed PMID: 9853823.
224. Dakic A, Metcalf D, Di Rago L, Mifsud S, Wu L, Nutt SL. PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. *J Exp Med.* 2005;201(9):1487-502. Epub 2005/05/04. doi: 10.1084/jem.20050075. PubMed PMID: 15867096; PMCID: PMC2213186.
225. Dakic A, Wu L, Nutt SL. Is PU.1 a dosage-sensitive regulator of haemopoietic lineage commitment and leukaemogenesis? *Trends Immunol.* 2007;28(3):108-14. Epub 2007/02/03. doi: 10.1016/j.it.2007.01.006. PubMed PMID: 17267285.
226. Kumar R, DuMond JF, Khan SH, Thompson EB, He Y, Burg MB, Ferraris JD. NFAT5, which protects against hypertonicity, is activated by that stress via structuring of its intrinsically disordered domain. *Proc Natl Acad Sci U S A.* 2020;117(33):20292-7. Epub 2020/08/05. doi: 10.1073/pnas.1911680117. PubMed PMID: 32747529; PMCID: PMC7443939.
227. Ito T, Fujio Y, Takahashi K, Azuma J. Degradation of NFAT5, a transcriptional regulator of osmotic stress-related genes, is a critical event for doxorubicin-induced cytotoxicity in cardiac myocytes. *J Biol Chem.* 2007;282(2):1152-60. Epub 2006/11/16. doi: 10.1074/jbc.M609547200. PubMed PMID: 17105721.
228. Lopez-Rodriguez C, Aramburu J, Jin L, Rakeman AS, Michino M, Rao A. Bridging the NFAT and NF-kappaB families: NFAT5 dimerization regulates cytokine gene transcription in response to osmotic stress. *Immunity.* 2001;15(1):47-58. Epub 2001/08/04. doi: 10.1016/s1074-7613(01)00165-0. PubMed PMID: 11485737.
229. Neuhofer W. Role of NFAT5 in inflammatory disorders associated with osmotic stress. *Curr Genomics.* 2010;11(8):584-90. Epub 2011/06/02. doi: 10.2174/138920210793360961. PubMed PMID: 21629436; PMCID: PMC3078683.
230. Kino T, Takatori H, Manoli I, Wang Y, Tiulpakov A, Blackman MR, Su YA, Chrousos GP, DeCherney AH, Segars JH. Brx mediates the response of lymphocytes to osmotic stress through the activation of NFAT5. *Sci Signal.* 2009;2(57):ra5. Epub 2009/02/13. doi: 10.1126/scisignal.2000081. PubMed PMID: 19211510; PMCID: PMC2856329.
231. Lorgen M, Jorgensen EH, Jordan WC, Martin SA, Hazlerigg DG. NFAT5 genes are part of the osmotic regulatory system in Atlantic salmon (*Salmo salar*). *Mar Genomics.* 2017;31:25-31. Epub 2016/06/23. doi: 10.1016/j.margen.2016.06.004. PubMed PMID: 27330039; PMCID: PMC5292104.
232. Timucin AC, Bodur C, Basaga H. SIRT1 contributes to aldose reductase expression through modulating NFAT5 under osmotic stress: In vitro and in silico insights. *Cell Signal.*

- 2015;27(11):2160-72. Epub 2015/08/25. doi: 10.1016/j.cellsig.2015.08.013. PubMed PMID: 26297866.
233. Morton CJ, Ladbury JE. Water-mediated protein-DNA interactions: the relationship of thermodynamics to structural detail. *Protein Sci.* 1996;5(10):2115-8. Epub 1996/10/01. doi: 10.1002/pro.5560051018. PubMed PMID: 8897612; PMCID: PMC2143271.
234. Reddy CK, Das A, Jayaram B. Do water molecules mediate protein-DNA recognition? *J Mol Biol.* 2001;314(3):619-32. Epub 2002/02/16. doi: 10.1006/jmbi.2001.5154. PubMed PMID: 11846571.
235. Schwabe JW. The role of water in protein-DNA interactions. *Curr Opin Struct Biol.* 1997;7(1):126-34. Epub 1997/02/01. doi: 10.1016/s0959-440x(97)80016-4. PubMed PMID: 9032063.
236. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, Yan J, Talukder S, Turunen M, Taipale M, Stunnenberg HG, Ukkonen E, Hughes TR, Bulyk ML, Taipale J. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.* 2010;29(13):2147-60. Epub 2010/06/03. doi: 10.1038/emboj.2010.106. PubMed PMID: 20517297; PMCID: PMC2905244.
237. Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J Pept Sci.* 1995;1(5):319-29. Epub 1995/09/01. doi: 10.1002/psc.310010507. PubMed PMID: 9223011.
238. Kueh HY, Champhekar A, Nutt SL, Elowitz MB, Rothenberg EV. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science.* 2013;341(6146):670-3. Epub 2013/07/23. doi: 10.1126/science.1240831. PubMed PMID: 23868921; PMCID: PMC3913367.
239. Rhodes J, Hagen A, Hsu K, Deng M, Liu TX, Look AT, Kanki JP. Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev Cell.* 2005;8(1):97-108. Epub 2004/12/29. doi: 10.1016/j.devcel.2004.11.014. PubMed PMID: 15621533.
240. Li Y, Okuno Y, Zhang P, Radomska HS, Chen H, Iwasaki H, Akashi K, Klemsz MJ, McKercher SR, Maki RA, Tenen DG. Regulation of the PU.1 gene by distal elements. *Blood.* 2001;98(10):2958-65. Epub 2001/11/08. doi: 10.1182/blood.v98.10.2958. PubMed PMID: 11698277.
241. Chen H, Ray-Gallet D, Zhang P, Hetherington CJ, Gonzalez DA, Zhang DE, Moreau-Gachelin F, Tenen DG. PU.1 (Spi-1) autoregulates its expression in myeloid cells. *Oncogene.* 1995;11(8):1549-60. Epub 1995/10/19. PubMed PMID: 7478579.
242. Mittrucker HW, Matsuyama T, Grossman A, Kundig TM, Potter J, Shahinian A, Wakeham A, Patterson B, Ohashi PS, Mak TW. Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science.* 1997;275(5299):540-3. Epub 1997/01/24. doi: 10.1126/science.275.5299.540. PubMed PMID: 8999800.
243. Mittrucker HW, Matsuyama T, Grossman A, Kundig TM, Potter J, Shahinian A, Wakeham A, Patterson B, Ohashi PS, Mak TW. Pillars Article: Requirement for the Transcription Factor LSIRF/IRF4 for Mature B and T Lymphocyte Function. *Science.* 1997. 275: 540-543. *J Immunol.* 2017;199(11):3717-20. Epub 2017/11/22. PubMed PMID: 29158347.
244. Pongubala JM, Nagulapalli S, Klemsz MJ, McKercher SR, Maki RA, Atchison ML. PU.1 recruits a second nuclear factor to a site important for immunoglobulin kappa 3' enhancer activity. *Mol Cell Biol.* 1992;12(1):368-78. Epub 1992/01/01. doi: 10.1128/mcb.12.1.368. PubMed PMID: 1729611; PMCID: PMC364131.

245. Ng FS, Schutte J, Ruau D, Diamanti E, Hannah R, Kinston SJ, Gottgens B. Constrained transcription factor spacing is prevalent and important for transcriptional control of mouse blood cells. *Nucleic Acids Res.* 2014;42(22):13513-24. Epub 2014/11/28. doi: 10.1093/nar/gku1254. PubMed PMID: 25428352; PMCID: PMC4267662.
246. Bornstein C, Winter D, Barnett-Itzhaki Z, David E, Kadri S, Garber M, Amit I. A negative feedback loop of transcription factors specifies alternative dendritic cell chromatin States. *Mol Cell.* 2014;56(6):749-62. Epub 2014/12/03. doi: 10.1016/j.molcel.2014.10.014. PubMed PMID: 25453760; PMCID: PMC4412443.
247. Annese V. Genetics and epigenetics of IBD. *Pharmacol Res.* 2020;159:104892. Epub 2020/05/29. doi: 10.1016/j.phrs.2020.104892. PubMed PMID: 32464322.
248. Le Bouc Y, Rossignol S, Azzi S, Steunou V, Netchine I, Gicquel C. Epigenetics, genomic imprinting and assisted reproductive technology. *Ann Endocrinol (Paris).* 2010;71(3):237-8. Epub 2010/04/07. doi: 10.1016/j.ando.2010.02.004. PubMed PMID: 20362968.
249. Acevedo N, Alashkar Alhamwe B, Caraballo L, Ding M, Ferrante A, Garn H, Garssen J, Hii CS, Irvine J, Llinas-Caballero K, Lopez JF, Miethe S, Perveen K, Pogge von Strandmann E, Sokolowska M, Potaczek DP, van Esch B. Perinatal and Early-Life Nutrition, Epigenetics, and Allergy. *Nutrients.* 2021;13(3). Epub 2021/03/07. doi: 10.3390/nu13030724. PubMed PMID: 33668787; PMCID: PMC7996340.
250. Steidl U, Rosenbauer, F., Verhaak, R. G., Gu, X., Ebralidze, A., Otu, H. H., Klippel, S., Steidl, C., Bruns, I., Costa, D. B., Wagner, K., Aivado, M., Kobbe, G., Valk, P. J., Passegue, E., Libermann, T. A., Delwel, R., & Tenen, D. G. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. *Nat Genet.* 2006;38(11):1269-77. Epub 2006/10/17. doi: 10.1038/ng1898. PubMed PMID: 17041602.
251. Steidl U, Steidl, C., Ebralidze, A., Chapuy, B., Han, H. J., Will, B., Rosenbauer, F., Becker, A., Wagner, K., Koschmieder, S., Kobayashi, S., Costa, D. B., Schulz, T., O'Brien, K. B., Verhaak, R. G., Delwel, R., Haase, D., Trumper, L., Krauter, J., Kohwi-Shigematsu, T., Griesinger, F., & Tenen, D. G. A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. *J Clin Invest.* 2007;117(9):2611-20.
252. Steidl U, Rosenbauer F, Verhaak RG, Gu X, Ebralidze A, Otu HH, Klippel S, Steidl C, Bruns I, Costa DB, Wagner K, Aivado M, Kobbe G, Valk PJ, Passegue E, Libermann TA, Delwel R, Tenen DG. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. *Nat Genet.* 2006;38(11):1269-77. Epub 2006/10/17. doi: 10.1038/ng1898. PubMed PMID: 17041602.
253. Steidl U, Steidl C, Ebralidze A, Chapuy B, Han HJ, Will B, Rosenbauer F, Becker A, Wagner K, Koschmieder S, Kobayashi S, Costa DB, Schulz T, O'Brien KB, Verhaak RG, Delwel R, Haase D, Trumper L, Krauter J, Kohwi-Shigematsu T, Griesinger F, Tenen DG. A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. *J Clin Invest.* 2007;117(9):2611-20. Epub 2007/08/19. doi: 10.1172/JCI30525. PubMed PMID: 17694175; PMCID: PMC1937499.
254. Steidl U RF, Verhaak RG, Gu X, Ebralidze A, Otu HH, Klippel S, Steidl C, Bruns I, Costa DB, Wagner K, Aivado M, Kobbe G, Valk PJ, Passequé E, Libermann TA, Delwel R, Tenen DG. Essential role of Jun family transcription factors in PU.1 knowdown-induced leukemic stem cells. *Nature genetics.*38(11):1269-77.
255. Hagenbuchner J, Ausserlechner MJ. Targeting transcription factors by small compounds--Current strategies and future implications. *Biochem Pharmacol.* 2016;107:1-13. Epub 2015/12/22. doi: 10.1016/j.bcp.2015.12.006. PubMed PMID: 26686579.

256. Wilson WD, Nguyen B, Tanious FA, Mathis A, Hall JE, Stephens CE, Boykin DW. Dications that target the DNA minor groove: compound design and preparation, DNA interactions, cellular distribution and biological activity. *Curr Med Chem Anticancer Agents*. 2005;5(4):389-408. Epub 2005/08/17. doi: 10.2174/1568011054222319. PubMed PMID: 16101490.
257. Nguyen B, Hamelberg D, Bailly C, Colson P, Stanek J, Brun R, Neidle S, Wilson WD. Characterization of a novel DNA minor-groove complex. *Biophys J*. 2004;86(2):1028-41. Epub 2004/01/30. doi: 10.1016/s0006-3495(04)74178-8. PubMed PMID: 14747338; PMCID: PMC1303896.
258. Nguyen B, Lee MP, Hamelberg D, Joubert A, Bailly C, Brun R, Neidle S, Wilson WD. Strong binding in the DNA minor groove by an aromatic diamidine with a shape that does not match the curvature of the groove. *J Am Chem Soc*. 2002;124(46):13680-1. Epub 2002/11/15. doi: 10.1021/ja027953c. PubMed PMID: 12431090.
259. de Souza EM, Oliveira GM, Boykin DW, Kumar A, Hu Q, De Nazare CSM. Trypanocidal activity of the phenyl-substituted analogue of furamidine DB569 against *Trypanosoma cruzi* infection in vivo. *J Antimicrob Chemother*. 2006;58(3):610-4. Epub 2006/07/21. doi: 10.1093/jac/dkl259. PubMed PMID: 16854954.
260. Ismail MA, Brun R, Easterbrook JD, Tanious FA, Wilson WD, Boykin DW. Synthesis and antiprotozoal activity of aza-analogues of furamidine. *J Med Chem*. 2003;46(22):4761-9. Epub 2003/10/17. doi: 10.1021/jm0302602. PubMed PMID: 14561095.
261. Nguyen B, Tardy C, Bailly C, Colson P, Houssier C, Kumar A, Boykin DW, Wilson WD. Influence of compound structure on affinity, sequence selectivity, and mode of binding to DNA for unfused aromatic dications related to furamidine. *Biopolymers*. 2002;63(5):281-97. Epub 2002/03/06. doi: 10.1002/bip.10073. PubMed PMID: 11877739.
262. Wilson WD, Tanious FA, Mathis A, Tevis D, Hall JE, Boykin DW. Antiparasitic compounds that target DNA. *Biochimie*. 2008;90(7):999-1014. Epub 2008/03/18. doi: 10.1016/j.biochi.2008.02.017. PubMed PMID: 18343228; PMCID: PMC2515095.
263. Neidle S. DNA minor-groove recognition by small molecules. *Nat Prod Rep*. 2001;18(3):291-309. Epub 2001/07/31. doi: 10.1039/a705982e. PubMed PMID: 11476483.
264. Miao Y, Lee MP, Parkinson GN, Batista-Parra A, Ismail MA, Neidle S, Boykin DW, Wilson WD. Out-of-shape DNA minor groove binders: induced fit interactions of heterocyclic dications with the DNA minor groove. *Biochemistry*. 2005;44(45):14701-8. Epub 2005/11/09. doi: 10.1021/bi051791q. PubMed PMID: 16274217.
265. Trent JO, Clark GR, Kumar A, Wilson WD, Boykin DW, Hall JE, Tidwell RR, Blagburn BL, Neidle S. Targeting the minor groove of DNA: crystal structures of two complexes between furan derivatives of berenil and the DNA dodecamer d(CGCGAATTCGCG)₂. *J Med Chem*. 1996;39(23):4554-62. Epub 1996/11/08. doi: 10.1021/jm9604484. PubMed PMID: 8917643.
266. Pham TH, Minderjahn J, Schmidl C, Hoffmeister H, Schmidhofer S, Chen W, Langst G, Benner C, Rehli M. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res*. 2013;41(13):6391-402. Epub 2013/05/10. doi: 10.1093/nar/gkt355. PubMed PMID: 23658224; PMCID: PMC3711439.
267. Bonadies N, Neururer C, Steege A, Vallabhapurapu S, Pabst T, Mueller BU. PU.1 is regulated by NF-kappaB through a novel binding site in a 17 kb upstream enhancer element. *Oncogene*. 2010;29(7):1062-72. Epub 2009/12/08. doi: 10.1038/onc.2009.371. PubMed PMID: 19966852.

268. Bonadies N, Pabst T, Mueller BU. Heterozygous deletion of the PU.1 locus in human AML. *Blood*. 2010;115(2):331-4. Epub 2009/11/06. doi: 10.1182/blood-2009-03-212225. PubMed PMID: 19890096.
269. Mueller BU, Pabst T, Osato M, Asou N, Johansen LM, Minden MD, Behre G, Hiddemann W, Ito Y, Tenen DG. Heterozygous PU.1 mutations are associated with acute myeloid leukemia. *Blood*. 2002;100(3):998-1007. Epub 2002/07/20. doi: 10.1182/blood.v100.3.998. PubMed PMID: 12130514.
270. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, Hiddemann W, Zhang DE, Tenen DG. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat Med*. 2001;7(4):444-51. Epub 2001/04/03. doi: 10.1038/86515. PubMed PMID: 11283671.
271. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, Behre G, Hiddemann W, Tenen DG. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet*. 2001;27(3):263-70. Epub 2001/03/10. doi: 10.1038/85820. PubMed PMID: 11242107.
272. Rustenhoven J, Smith AM, Smyth LC, Jansson D, Scotter EL, Swanson MEV, Aalderink M, Coppeters N, Narayan P, Handley R, Overall C, Park TIH, Schweder P, Heppner P, Curtis MA, Faull RLM, Dragunow M. PU.1 regulates Alzheimer's disease-associated genes in primary human microglia. *Mol Neurodegener*. 2018;13(1):44. Epub 2018/08/21. doi: 10.1186/s13024-018-0277-1. PubMed PMID: 30124174; PMCID: PMC6102813.
273. Marafioti T, Mancini C, Ascani S, Sabattini E, Zinzani PL, Pozzobon M, Pulford K, Falini B, Jaffe ES, Muller-Hermelink HK, Mason DY, Pileri SA. Leukocyte-specific phosphoprotein-1 and PU.1: two useful markers for distinguishing T-cell-rich B-cell lymphoma from lymphocyte-predominant Hodgkin's disease. *Haematologica*. 2004;89(8):957-64. Epub 2004/09/02. PubMed PMID: 15339679.
274. Torlakovic E, Tierens A, Dang HD, Delabie J. The transcription factor PU.1, necessary for B-cell development is expressed in lymphocyte predominance, but not classical Hodgkin's disease. *Am J Pathol*. 2001;159(5):1807-14. Epub 2001/11/07. doi: 10.1016/S0002-9440(10)63027-1. PubMed PMID: 11696441; PMCID: PMC1867045.
275. Wohlfahrt T, Rauber S, Uebe S, Lubber M, Soare A, Ekici A, Weber S, Matei AE, Chen CW, Maier C, Karouzakis E, Kiener HP, Pachera E, Dees C, Beyer C, Daniel C, Gelse K, Kremer AE, Naschberger E, Sturzl M, Butter F, Sticherling M, Finotto S, Kreuter A, Kaplan MH, Jungel A, Gay S, Nutt SL, Boykin DW, Poon GMK, Distler O, Schett G, Distler JHW, Ramming A. PU.1 controls fibroblast polarization and tissue fibrosis. *Nature*. 2019;566(7744):344-9. Epub 2019/02/01. doi: 10.1038/s41586-019-0896-x. PubMed PMID: 30700907; PMCID: PMC6526281.
276. Kim HG, de Guzman CG, Swindle CS, Cotta CV, Gartland L, Scott EW, Klug CA. The ETS family transcription factor PU.1 is necessary for the maintenance of fetal liver hematopoietic stem cells. *Blood*. 2004;104(13):3894-900. Epub 2004/08/26. doi: 10.1182/blood-2002-08-2425. PubMed PMID: 15328162.
277. Aikawa Y, Katsumoto T, Zhang P, Shima H, Shino M, Terui K, Ito E, Ohno H, Stanley ER, Singh H, Tenen DG, Kitabayashi I. PU.1-mediated upregulation of CSF1R is crucial for leukemia stem cell potential induced by MOZ-TIF2. *Nat Med*. 2010;16(5):580-5, 1p following 5. Epub 2010/04/27. doi: 10.1038/nm.2122. PubMed PMID: 20418886; PMCID: PMC3039870.
278. Aikawa Y, Yamagata K, Katsumoto T, Shima Y, Shino M, Stanley ER, Cleary ML, Akashi K, Tenen DG, Kitabayashi I. Essential role of PU.1 in maintenance of mixed lineage leukemia-

- associated leukemic stem cells. *Cancer Sci.* 2015;106(3):227-36. Epub 2014/12/23. doi: 10.1111/cas.12593. PubMed PMID: 25529853; PMCID: PMC4373983.
279. Cook WD, McCaw BJ, Herring C, John DL, Foote SJ, Nutt SL, Adams JM. PU.1 is a suppressor of myeloid leukemia, inactivated in mice by gene deletion and mutation of its DNA binding domain. *Blood.* 2004;104(12):3437-44. Epub 2004/08/12. doi: 10.1182/blood-2004-06-2234. PubMed PMID: 15304397.
280. D'Alo F, Di Ruscio A, Guidi F, Fabiani E, Greco M, Rumi C, Hohaus S, Voso MT, Leone G. PU.1 and CEBPA expression in acute myeloid leukemia. *Leuk Res.* 2008;32(9):1448-53. Epub 2008/03/01. doi: 10.1016/j.leukres.2008.01.007. PubMed PMID: 18308386.
281. Fernandez-Nestosa MJ, Monturus E, Sanchez Z, Torres FS, Fernandez AF, Fraga MF, Hernandez P, Schvartzman JB, Krimer DB. DNA methylation-mediated silencing of PU.1 in leukemia cells resistant to cell differentiation. *Springerplus.* 2013;2:392. Epub 2013/09/07. doi: 10.1186/2193-1801-2-392. PubMed PMID: 24010046; PMCID: PMC3758488.
282. Gu X, Hu Z, Ebrahim Q, Crabb JS, Mahfouz RZ, Radivoyevitch T, Crabb JW, Sauntharajah Y. Runx1 regulation of Pu.1 corepressor/coactivator exchange identifies specific molecular targets for leukemia differentiation therapy. *J Biol Chem.* 2014;289(21):14881-95. Epub 2014/04/04. doi: 10.1074/jbc.M114.562447. PubMed PMID: 24695740; PMCID: PMC4031539.
283. Kastner P, Chan S. PU.1: a crucial and versatile player in hematopoiesis and leukemia. *Int J Biochem Cell Biol.* 2008;40(1):22-7. Epub 2007/03/22. doi: 10.1016/j.biocel.2007.01.026. PubMed PMID: 17374502.
284. Lim M, Batista CR, de Oliveira BR, Creighton R, Ferguson J, Clemmer K, Knight D, Iansavitchous J, Mahmood D, Avino M, DeKoter RP. Janus Kinase Mutations in Mice Lacking PU.1 and Spi-B Drive B Cell Leukemia through Reactive Oxygen Species-Induced DNA Damage. *Mol Cell Biol.* 2020;40(18). Epub 2020/07/08. doi: 10.1128/MCB.00189-20. PubMed PMID: 32631903; PMCID: PMC7459267.
285. Metcalf D. Loss of PU.1: one of the many roads to myeloid leukemia. *Cell Cycle.* 2006;5(7):673-4. Epub 2006/04/04. doi: 10.4161/cc.5.7.2619. PubMed PMID: 16582633.
286. Mueller BU, Pabst T, Fos J, Petkovic V, Fey MF, Asou N, Buergi U, Tenen DG. ATRA resolves the differentiation block in t(15;17) acute myeloid leukemia by restoring PU.1 expression. *Blood.* 2006;107(8):3330-8. Epub 2005/12/15. doi: 10.1182/blood-2005-07-3068. PubMed PMID: 16352814; PMCID: PMC1895760.
287. Ojima M, Hirouchi T, Etani R, Ariyoshi K, Fujishima Y, Kai M. Dose-Rate-Dependent PU.1 Inactivation to Develop Acute Myeloid Leukemia in Mice Through Persistent Stem Cell Proliferation After Acute or Chronic Gamma Irradiation. *Radiat Res.* 2019;192(6):612-20. Epub 2019/09/29. doi: 10.1667/RR15359.1. PubMed PMID: 31560640.
288. Pang SH, Minnich M, Gangatirkar P, Zheng Z, Ebert A, Song G, Dickins RA, Corcoran LM, Mullighan CG, Busslinger M, Huntington ND, Nutt SL, Carotta S. PU.1 cooperates with IRF4 and IRF8 to suppress pre-B-cell leukemia. *Leukemia.* 2016;30(6):1375-87. Epub 2016/03/05. doi: 10.1038/leu.2016.27. PubMed PMID: 26932576; PMCID: PMC5179358.
289. Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, Akashi K, Fiering S, Tenen DG. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet.* 2004;36(6):624-30. Epub 2004/05/18. doi: 10.1038/ng1361. PubMed PMID: 15146183.
290. Savickiene J, Treigyte G, Vistartaite G, Tunaitis V, Magnusson KE, Navakauskiene R. C/EBPalpha and PU.1 are involved in distinct differentiation responses of acute promyelocytic

leukemia HL-60 and NB4 cells via chromatin remodeling. *Differentiation*. 2011;81(1):57-67. Epub 2010/09/25. doi: 10.1016/j.diff.2010.08.003. PubMed PMID: 20864248.

291. Seki M, Kimura S, Isobe T, Yoshida K, Ueno H, Nakajima-Takagi Y, Wang C, Lin L, Kon A, Suzuki H, Shiozawa Y, Kataoka K, Fujii Y, Shiraishi Y, Chiba K, Tanaka H, Shimamura T, Masuda K, Kawamoto H, Ohki K, Kato M, Arakawa Y, Koh K, Hanada R, Moritake H, Akiyama M, Kobayashi R, Deguchi T, Hashii Y, Imamura T, Sato A, Kiyokawa N, Oka A, Hayashi Y, Takagi M, Manabe A, Ohara A, Horibe K, Sanada M, Iwama A, Mano H, Miyano S, Ogawa S, Takita J. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. *Nat Genet*. 2017;49(8):1274-81. Epub 2017/07/04. doi: 10.1038/ng.3900. PubMed PMID: 28671687.

292. Walter MJ, Park JS, Ries RE, Lau SK, McLellan M, Jaeger S, Wilson RK, Mardis ER, Ley TJ. Reduced PU.1 expression causes myeloid progenitor expansion and increased leukemia penetrance in mice expressing PML-RARalpha. *Proc Natl Acad Sci U S A*. 2005;102(35):12513-8. Epub 2005/08/23. doi: 10.1073/pnas.0504247102. PubMed PMID: 16113082; PMCID: PMC1188022.

293. Will B, Vogler TO, Narayanagari S, Bartholdy B, Todorova TI, da Silva Ferreira M, Chen J, Yu Y, Mayer J, Barreyro L, Carvajal L, Neriah DB, Roth M, van Oers J, Schatzlein S, McMahan C, Edelmann W, Verma A, Steidl U. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat Med*. 2015;21(10):1172-81. Epub 2015/09/08. doi: 10.1038/nm.3936. PubMed PMID: 26343801; PMCID: PMC5144917.

294. Xu LS, Sokalski KM, Hotke K, Christie DA, Zarnett O, Piskorz J, Thillainadesan G, Torchia J, DeKoter RP. Regulation of B cell linker protein transcription by PU.1 and Spi-B in murine B cell acute lymphoblastic leukemia. *J Immunol*. 2012;189(7):3347-54. Epub 2012/09/08. doi: 10.4049/jimmunol.1201267. PubMed PMID: 22956576.

295. Yang XW, Wang P, Liu JQ, Zhang H, Xi WD, Jia XH, Wang KK. Coordinated regulation of the immunoproteasome subunits by PML/RARalpha and PU.1 in acute promyelocytic leukemia. *Oncogene*. 2014;33(21):2700-8. Epub 2013/06/19. doi: 10.1038/onc.2013.224. PubMed PMID: 23770850.

296. Zhou J, Wu J, Li B, Liu D, Yu J, Yan X, Zheng S, Wang J, Zhang L, Zhang L, He F, Li Q, Chen A, Zhang Y, Zhao X, Guan Y, Zhao X, Yan J, Ni J, Nobrega MA, Lowenberg B, Delwel R, Valk PJ, Kumar A, Xie L, Tenen DG, Huang G, Wang QF. PU.1 is essential for MLL leukemia partially via crosstalk with the MEIS/HOX pathway. *Leukemia*. 2014;28(7):1436-48. Epub 2014/01/22. doi: 10.1038/leu.2013.384. PubMed PMID: 24445817; PMCID: PMC4410691.

297. Zhu X, Zhang H, Qian M, Zhao X, Yang W, Wang P, Zhang J, Wang K. The significance of low PU.1 expression in patients with acute promyelocytic leukemia. *J Hematol Oncol*. 2012;5:22. Epub 2012/05/10. doi: 10.1186/1756-8722-5-22. PubMed PMID: 22569057; PMCID: PMC3407792.

298. Rosenbauer F, Owens BM, Yu L, Tumang JR, Steidl U, Kutok JL, Clayton LK, Wagner K, Scheller M, Iwasaki H, Liu C, Hackanson B, Akashi K, Leutz A, Rothstein TL, Plass C, Tenen DG. Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. *Nat Genet*. 2006;38(1):27-37. Epub 2005/11/29. doi: 10.1038/ng1679. PubMed PMID: 16311598.

299. Chenoweth DM, Dervan PB. Allosteric modulation of DNA by small molecules. *Proc Natl Acad Sci U S A*. 2009;106(32):13175-9. Epub 2009/08/12. doi: 10.1073/pnas.0906532106. PubMed PMID: 19666554; PMCID: PMC2726366.

300. Munde M, Wang S, Kumar A, Stephens CE, Farahat AA, Boykin DW, Wilson WD, Poon GM. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. *Nucleic Acids Res.* 2014;42(2):1379-90. Epub 2013/10/26. doi: 10.1093/nar/gkt955. PubMed PMID: 24157839; PMCID: PMC3902942.
301. Antony- Debré I, Paul, A., Leite, J., Mitchell, K., Kim, H. M., Carvajal, L. A., Todorova, T. I., Huang, K., Kumar, A., Farahat, A. A., Bartholdy, B., Narayanagari, S. R., Chen, J., Ambesi-Impiombato, A., Ferrando, A. A., Mantzaris, I., Gavathiotis, E., Verma, A., Will, B., Boykin, D. W., Wilson, W. D., Poon, G. M., & Steidl, U. . Pharmacological inhibition of the transcription factor PU.1 in leukemia. *The Journal of clinical investigation.* 2017;12:4297–313.
302. Wenzler T, Boykin DW, Ismail MA, Hall JE, Tidwell RR, Brun R. New treatment option for second-stage African sleeping sickness: in vitro and in vivo efficacy of aza analogs of DB289. *Antimicrob Agents Chemother.* 2009;53(10):4185-92. Epub 2009/07/22. doi: 10.1128/AAC.00225-09. PubMed PMID: 19620327; PMCID: PMC2764217.
303. Yang S, Wenzler T, Miller PN, Wu H, Boykin DW, Brun R, Wang MZ. Pharmacokinetic comparison to determine the mechanisms underlying the differential efficacies of cationic diamidines against first- and second-stage human African trypanosomiasis. *Antimicrob Agents Chemother.* 2014;58(7):4064-74. Epub 2014/05/07. doi: 10.1128/AAC.02605-14. PubMed PMID: 24798280; PMCID: PMC4068519.
304. Powell IJ, Carpten J, Dunston G, Kittles R, Bennett J, Hoke G, Pettaway C, Weinrich S, Vijayakumar S, Ahaghotu CA, Boykin W, Mason T, Royal C, Baffoe-Bonnie A, Bailey-Wilson J, Berg K, Trent J, Collins F. African-American heredity prostate cancer study: a model for genetic research. *J Natl Med Assoc.* 2001;93(12 Suppl):25S-8S. Epub 2002/01/19. PubMed PMID: 11798061; PMCID: PMC2719991.
305. Royal C, Baffoe-Bonnie A, Kittles R, Powell I, Bennett J, Hoke G, Pettaway C, Weinrich S, Vijayakumar S, Ahaghotu C, Mason T, Johnson E, Obeikwe M, Simpson C, Mejia R, Boykin W, Roberson P, Frost J, Faison-Smith L, Meegan C, Foster N, Furbert-Harris P, Carpten J, Bailey-Wilson J, Trent J, Berg K, Dunston G, Collins F. Recruitment experience in the first phase of the African American Hereditary Prostate Cancer (AAHPC) study. *Ann Epidemiol.* 2000;10(8 Suppl):S68-77. Epub 2001/02/24. doi: 10.1016/s1047-2797(00)00194-0. PubMed PMID: 11189095.
306. Degnan BM, Degnan SM, Naganuma T, Morse DE. The ets multigene family is conserved throughout the Metazoa. *Nucleic acids research.* 1993;21(15):3479-84. Epub 1993/07/25. PubMed PMID: 8346026; PMCID: 331448.
307. Orkin SH. Transcription factors and hematopoietic development. *The Journal of biological chemistry.* 1995;270(10):4955-8. Epub 1995/03/10. PubMed PMID: 7890597.
308. Ciau-Uitz A, Wang L, Patient R, Liu F. ETS transcription factors in hematopoietic stem cell development. *Blood Cells Mol Dis.* 2013;51(4):248-55. Epub 2013/08/10. doi: 10.1016/j.bcmd.2013.07.010. PubMed PMID: 23927967.
309. He G, Tolic A, Bashkin JK, Poon GM. Heterogeneous dynamics in DNA site discrimination by the structurally homologous DNA-binding domains of ETS-family transcription factors. *Nucleic Acids Res.* 2015;43(8):4322-31. Epub 2015/04/01. doi: 10.1093/nar/gkv267. PubMed PMID: 25824951; PMCID: PMC4417174.
310. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science.* 1994;265(5178):1573-7. Epub 1994/09/09. PubMed PMID: 8079170.

311. Nutt SL, Metcalf D, D'Amico A, Polli M, Wu L. Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. *J Exp Med*. 2005;201(2):221-31. Epub 2005/01/20. doi: 10.1084/jem.20041535. PubMed PMID: 15657291; PMCID: PMC2212785.
312. Dahl R, Simon MC. The importance of PU.1 concentration in hematopoietic lineage commitment and maturation. *Blood Cells Mol Dis*. 2003;31(2):229-33. Epub 2003/09/16. doi: 10.1016/s1079-9796(03)00152-9. PubMed PMID: 12972030.
313. DeKoter RP, Singh H. Regulation of B Lymphocyte and Macrophage Development by Graded Expression of PU.1. *Science*. 2000;288(5470):1439-41. doi: 10.1126/science.288.5470.1439.
314. Mak KS, Funnell AP, Pearson RC, Crossley M. PU.1 and Haematopoietic Cell Fate: Dosage Matters. *Int J Cell Biol*. 2011;2011:808524. Epub 2011/08/17. doi: 10.1155/2011/808524. PubMed PMID: 21845190; PMCID: PMC3154517.
315. Anderson MK, Hernandez-Hoyos G, Diamond RA, Rothenberg EV. Precise developmental regulation of Ets family transcription factors during specification and commitment to the T cell lineage. *Development*. 1999;126(14):3131-48. Epub 1999/06/22. PubMed PMID: 10375504.
316. Mouly E, Chemin K, Nguyen HV, Chopin M, Mesnard L, Leite-de-Moraes M, Burlen-defranoux O, Bandeira A, Bories JC. The Ets-1 transcription factor controls the development and function of natural regulatory T cells. *J Exp Med*. 2010;207(10):2113-25. Epub 2010/09/22. doi: 10.1084/jem.20092153. PubMed PMID: 20855499; PMCID: 2947068.
317. Anderson MK, Weiss AH, Hernandez-Hoyos G, Dionne CJ, Rothenberg EV. Constitutive expression of PU.1 in fetal hematopoietic progenitors blocks T cell development at the pro-T cell stage. *Immunity*. 2002;16(2):285-96. Epub 2002/03/01. doi: 10.1016/s1074-7613(02)00277-7. PubMed PMID: 11869688.
318. Chang HC, Sehra S, Goswami R, Yao W, Yu Q, Stritesky GL, Jabeen R, McKinley C, Ahyi AN, Han L, Nguyen ET, Robertson MJ, Perumal NB, Tepper RS, Nutt SL, Kaplan MH. The transcription factor PU.1 is required for the development of IL-9-producing T cells and allergic inflammation. *Nat Immunol*. 2010;11(6):527-34. Epub 2010/05/01. doi: 10.1038/ni.1867. PubMed PMID: 20431622; PMCID: PMC3136246.
319. Chang H-C, Zhang S, Thieu VT, Slee RB, Bruns HA, Laribee RN, Klemsz MJ, Kaplan MH. PU.1 Expression Delineates Heterogeneity in Primary Th2 Cells. *Immunity*. 2005;22(6):693-703. doi: <http://dx.doi.org/10.1016/j.immuni.2005.03.016>.
320. Grenningloh R, Kang BY, Ho I-C. Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. *The Journal of Experimental Medicine*. 2005;201(4):615-26. doi: 10.1084/jem.20041330.
321. Dozmorov MG, Wren JD, Alarcon-Riquelme ME. Epigenomic elements enriched in the promoters of autoimmunity susceptibility genes. *Epigenetics*. 2014;9(2):276-85. doi: 10.4161/epi.27021. PubMed PMID: 24213554; PMCID: 3962538.
322. Gilliland DG. The diverse role of the ETS family of transcription factors in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2001;7(3):451-3. Epub 2001/04/12. PubMed PMID: 11297232.
323. Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, Akashi K, Fiering S, Tenen DG. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet*. 2004;36(6):624-30. doi: http://www.nature.com/ng/journal/v36/n6/supinfo/ng1361_S1.html.

324. Tatetsu H, Ueno S, Hata H, Yamada Y, Takeya M, Mitsuya H, Tenen DG, Okuno Y. Down-regulation of PU.1 by methylation of distal regulatory elements and the promoter is required for myeloma cell growth. *Cancer Res.* 2007;67(11):5328-36. Epub 2007/06/05. doi: 10.1158/0008-5472.CAN-06-4265. PubMed PMID: 17545613.
325. Yuki H, Ueno S, Tatetsu H, Niiro H, Iino T, Endo S, Kawano Y, Komohara Y, Takeya M, Hata H, Okada S, Watanabe T, Akashi K, Mitsuya H, Okuno Y. PU.1 is a potent tumor suppressor in classical Hodgkin lymphoma cells. *Blood.* 2013;121(6):962-70. Epub 2012/12/06. doi: 10.1182/blood-2012-05-431429. PubMed PMID: 23212521.
326. Gjonjeska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai L-H, Kellis M. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature.* 2015;518(7539):365-9. doi: 10.1038/nature14252
<http://www.nature.com/nature/journal/v518/n7539/abs/nature14252.html#supplementary-information>.
327. Gomez-Nicola D, Fransen NL, Suzzi S, Perry VH. Regulation of microglial proliferation during chronic neurodegeneration. *J Neurosci.* 2013;33(6):2481-93. doi: 10.1523/JNEUROSCI.4440-12.2013. PubMed PMID: 23392676.
328. Jantarantotai N, Ling A, Cheng J, Schwab C, McGeer PL, McLarnon JG. Upregulation and expression patterns of the angiogenic transcription factor ets-1 in Alzheimer's disease brain. *J Alzheimers Dis.* 2013;37(2):367-77. doi: 10.3233/JAD-122191. PubMed PMID: 23948889.
329. Garrett-Sinha LA, Su GH, Rao S, Kabak S, Hao Z, Clark MR, Simon MC. PU.1 and Spi-B are required for normal B cell receptor-mediated signal transduction. *Immunity.* 1999;10(4):399-408. Epub 1999/05/06. doi: 10.1016/s1074-7613(00)80040-0. PubMed PMID: 10229183.
330. Hollenhorst PC, Jones DA, Graves BJ. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic acids research.* 2004;32(18):5693-702. Epub 2004/10/23. doi: 10.1093/nar/gkh906. PubMed PMID: 15498926; PMCID: 524310.
331. Li SL, Schlegel W, Valente AJ, Clark RA. Critical flanking sequences of PU.1 binding sites in myeloid-specific promoters. *J Biol Chem.* 1999;274(45):32453-60. Epub 1999/11/05. doi: 10.1074/jbc.274.45.32453. PubMed PMID: 10542290.
332. Desjardins G, Okon M, Graves BJ, McIntosh LP. Conformational Dynamics and the Binding of Specific and Nonspecific DNA by the Autoinhibited Transcription Factor Ets-1. *Biochemistry.* 2016;55(29):4105-18. doi: 10.1021/acs.biochem.6b00460.
333. Lee GM, Donaldson LW, Pufall MA, Kang HS, Pot I, Graves BJ, McIntosh LP. The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J Biol Chem.* 2005;280(8):7088-99. Epub 2004/12/14. doi: 10.1074/jbc.M410722200. PubMed PMID: 15591056.
334. Pufall MA, Lee GM, Nelson ML, Kang H-S, Velyvis A, Kay LE, McIntosh LP, Graves BJ. Variable Control of Ets-1 DNA Binding by Multiple Phosphates in an Unstructured Region. *Science.* 2005;309(5731):142-5. doi: 10.1126/science.1111915.
335. Coyne HJ, 3rd, De S, Okon M, Green SM, Bhachech N, Graves BJ, McIntosh LP. Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J Mol Biol.* 2012;421(1):67-84. doi: 10.1016/j.jmb.2012.05.010. PubMed PMID: 22584210; PMCID: PMC3392548.
336. De S, Chan AC, Coyne HJ, 3rd, Bhachech N, Hermsdorf U, Okon M, Murphy ME, Graves BJ, McIntosh LP. Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6. *J Mol Biol.* 2014;426(7):1390-406. doi: 10.1016/j.jmb.2013.11.031. PubMed PMID: 24333486; PMCID: PMC4278593.

337. De S, Okon M, Graves BJ, McIntosh LP. Autoinhibition of ETV6 DNA Binding Is Established by the Stability of Its Inhibitory Helix. *Journal of Molecular Biology*. 2016;428(8):1515-30. doi: <http://dx.doi.org/10.1016/j.jmb.2016.02.020>.
338. Shrivastava T, Mino K, Babayeva ND, Baranovskaya OI, Rizzino A, Tahirov TH. Structural basis of Ets1 activation by Runx1. *Leukemia*. 2014;28(10):2040-8. doi: 10.1038/leu.2014.111.
339. Goetz TL, Gu TL, Speck NA, Graves BJ. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol*. 2000;20(1):81-90. Epub 1999/12/14. PubMed PMID: 10594011; PMCID: 85055.
340. Yee AA, Yin P, Siderovski DP, Mak TW, Litchfield DW, Arrowsmith CH. Cooperative interaction between the DNA-binding domains of PU.1 and IRF4. *J Mol Biol*. 1998;279(5):1075-83. Epub 1998/06/27. doi: 10.1006/jmbi.1998.1838. PubMed PMID: 9642085.
341. Groß P, Yee AA, Arrowsmith CH, Macgregor RB, Jr. Quantitative hydroxyl radical footprinting reveals cooperative interactions between DNA-binding subdomains of PU.1 and IRF4. *Biochemistry*. 1998;37(27):9802-11. Epub 1998/07/10. doi: 10.1021/bi9731448. PubMed PMID: 9657694.
342. Escalante CR, Brass AL, Pongubala JM, Shatova E, Shen L, Singh H, Aggarwal AK. Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol Cell*. 2002;10(5):1097-105. Epub 2002/11/28. PubMed PMID: 12453417.
343. Liew CW, Rand KD, Simpson RJY, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba M, Nerlov C, Poulsen FM, Mackay JP. Molecular Analysis of the Interaction between the Hematopoietic Master Transcription Factors GATA-1 and PU.1. *The Journal of biological chemistry*. 2006;281(38):28296-306. doi: 10.1074/jbc.M602830200.
344. Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, Narravula S, Torbett BE, Orkin SH, Tenen DG. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood*. 2000;96(8):2641-8. Epub 2000/10/07. PubMed PMID: 11023493.
345. Fitzsimmons D, Lukin K, Lutz R, Garvie CW, Wolberger C, Hagman J. Highly cooperative recruitment of Ets-1 and release of autoinhibition by Pax5. *J Mol Biol*. 2009;392(2):452-64. doi: 10.1016/j.jmb.2009.07.028. PubMed PMID: 19616560; PMCID: PMC2763353.
346. Garvie CW, Hagman J, Wolberger C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell*. 2001;8(6):1267-76. Epub 2002/01/10. PubMed PMID: 11779502.
347. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P, Chilarska PM, Kinston S, Ouwehand WH, Dzierzak E, Pimanda JE, de Bruijn MF, Gottgens B. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 2010;7(4):532-44. Epub 2010/10/05. doi: 10.1016/j.stem.2010.07.016. PubMed PMID: 20887958.
348. Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, Graves BJ. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet*. 2009;5(12):e1000778. Epub 2009/12/19. doi: 10.1371/journal.pgen.1000778. PubMed PMID: 20019798; PMCID: PMC2787013.
349. Cauchy P, Maqbool MA, Zacarias-Cabeza J, Vanhille L, Koch F, Fenouil R, Gut M, Gut I, Santana MA, Griffon A, Imbert J, Moraes-Cabe C, Bories JC, Ferrier P, Spicuglia S, Andrau JC. Dynamic recruitment of Ets1 to both nucleosome-occupied and -depleted enhancer regions mediates a transcriptional program switch during early T-cell differentiation. *Nucleic acids research*. 2016;44(8):3567-85. doi: 10.1093/nar/gkv1475. PubMed PMID: 26673693; PMCID: PMC4856961.

350. Hogart A, Lichtenberg J, Ajay SS, Anderson S, Margulies EH, Bodine DM. Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. *Genome research*. 2012;22(8):1407-18. Epub 2012/06/12. doi: 10.1101/gr.132878.111. PubMed PMID: 22684279; PMCID: 3409254.
351. Boros J, Donaldson IJ, O'Donnell A, Odrowaz ZA, Zeef L, Lupien M, Meyer CA, Liu XS, Brown M, Sharrocks AD. Elucidation of the ELK1 target gene network reveals a role in the coordinate regulation of core components of the gene regulation machinery. *Genome research*. 2009;19(11):1963-73. doi: 10.1101/gr.093047.109. PubMed PMID: 19687146; PMCID: PMC2775591.
352. Pham TH, Benner C, Lichtinger M, Schwarzfischer L, Hu Y, Andreesen R, Chen W, Rehli M. Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood*. 2012;119(24):e161-71. Epub 2012/05/03. doi: 10.1182/blood-2012-01-402453. PubMed PMID: 22550342.
353. Kodandapani R, Pio F, Ni CZ, Piccialli G, Klemsz M, McKercher S, Maki RA, Ely KR. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature*. 1996;380(6573):456-60. Epub 1996/04/04. doi: 10.1038/380456a0. PubMed PMID: 8602247.
354. Burg MB, Ferraris JD, Dmitrieva NI. Cellular Response to Hyperosmotic Stresses. *Physiol Rev*. 2007;87(4):1441-74. doi: 10.1152/physrev.00056.2006.
355. Burg MB, Ferraris JD. Intracellular organic osmolytes: function and regulation. *The Journal of biological chemistry*. 2008;283(12):7309-13. Epub 2008/02/08. doi: 10.1074/jbc.R700042200. PubMed PMID: 18256030; PMCID: 2276334.
356. Ferraris JD, Burg MB. Tonicity-Regulated Gene Expression. In: Dieter H, Helmut S, editors. *Methods in enzymology*: Academic Press; 2007. p. 279-96.
357. Bortner CD, Scoltock AB, Sifre MI, Cidlowski JA. Osmotic stress resistance imparts acquired anti-apoptotic mechanisms in lymphocytes. *The Journal of biological chemistry*. 2012;287(9):6284-95. doi: 10.1074/jbc.M111.293001. PubMed PMID: 22228768; PMCID: PMC3307331.
358. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes & development*. 2014;28(24):2679-92. doi: 10.1101/gad.253443.114. PubMed PMID: 25512556; PMCID: 4265672.
359. Goswami R, Kaplan MH. Gcn5 Is Required for PU.1-Dependent IL-9 Induction in Th9 Cells. *The Journal of Immunology*. 2012;189(6):3026-33. doi: 10.4049/jimmunol.1201496.
360. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nature biotechnology*. 2014;32(2):171-8. doi: 10.1038/nbt.2798. PubMed PMID: 24441470; PMCID: 3951735.
361. van Riel B, Rosenbauer F. Epigenetic control of hematopoiesis: the PU.1 chromatin connection. *Biol Chem*. 2014;395(11):1265-74. doi: 10.1515/hsz-2014-0195. PubMed PMID: 25205721.
362. Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell*. 2012;149(2):467-82. doi: 10.1016/j.cell.2012.01.056. PubMed PMID: 22500808; PMCID: PMC3336965.

363. Alvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function. *Nature reviews Immunology*. 2015;15(1):7-17. doi: 10.1038/nri3777. PubMed PMID: 25534619.
364. Stephens DC, Poon GM. Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. *Nucleic acids research*. 2016;44(18):8671-81. doi: 10.1093/nar/gkw528. PubMed PMID: 27270080.
365. Gurion R, Vidal L, Gafter-Gvili A, Belnik Y, Yeshurun M, Raanani P, Shpilberg O. 5-azacitidine prolongs overall survival in patients with myelodysplastic syndrome - a systematic review and meta-analysis. *Haematologica*. 2010;95(2):303-10. doi: 10.3324/haematol.2009.010611.
366. Darnell JE, Jr. Transcription factors as targets for cancer therapy. *Nat Rev Cancer*. 2002;2(10):740-9. Epub 2002/10/03. doi: 10.1038/nrc906. PubMed PMID: 12360277.
367. Koehler AN. A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol*. 2010;14(3):331-40. Epub 2010/04/17. doi: 10.1016/j.cbpa.2010.03.022. PubMed PMID: 20395165; PMCID: PMC3248789.
368. Ghosh D, Papavassiliou AG. Transcription factor therapeutics: long-shot or lodestone. *Current medicinal chemistry*. 2005;12(6):691-701. Epub 2005/03/26. PubMed PMID: 15790306.
369. Melander C, Burnett R, Gottesfeld JM. Regulation of gene expression with pyrrole-imidazole polyamides. *J Biotechnol*. 2004;112(1-2):195-220. doi: 10.1016/j.jbiotec.2004.03.018. PubMed PMID: 15288953.
370. Reddy BS, Sharma SK, Lown JW. Recent developments in sequence selective minor groove DNA effectors. *Curr Med Chem*. 2001;8(5):475-508. PubMed PMID: 11281837.
371. Dickinson LA, Trauger JW, Baird EE, Dervan PB, Graves BJ, Gottesfeld JM. Inhibition of Ets-1 DNA binding and ternary complex formation between Ets-1, NF-kappaB, and DNA by a designed DNA-binding ligand. *The Journal of biological chemistry*. 1999;274(18):12765-73. Epub 1999/04/23. PubMed PMID: 10212261.
372. Stephens DC, Kim HM, Kumar A, Farahat AA, Boykin DW, Poon GMK. Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis. *Nucleic acids research*. 2016;44(9):4005-13. doi: 10.1093/nar/gkw229. PubMed PMID: 27079976; PMCID: PMC4872103.
373. Testoni M, Chung EY, Priebe V, Bertoni F. The transcription factor ETS1 in lymphomas: friend or foe? *Leuk Lymphoma*. 2015;56(7):1975-80. doi: 10.3109/10428194.2014.981670. PubMed PMID: 25363344.
374. Cook WD, McCaw BJ, Herring C, John DL, Foote SJ, Nutt SL, Adams JM. PU.1 is a suppressor of myeloid leukemia, inactivated in mice by gene deletion and mutation of its DNA binding domain. *Blood*. 2004;104(12):3437-44. doi: 10.1182/blood-2004-06-2234.
375. Okuno Y, Yuki H. PU.1 is a tumor suppressor for B cell malignancies. *Oncotarget*. 2012;3(12):1495-6. Epub 2013/03/05. doi: 10.18632/oncotarget.800. PubMed PMID: 23455457; PMCID: PMC3681485.
376. Jia X, Lee LK, Light J, Palmer AG, 3rd, Assa-Munt N. Backbone dynamics of a short PU.1 ETS domain. *J Mol Biol*. 1999;292(5):1083-93. Epub 1999/10/08. doi: 10.1006/jmbi.1999.3123. PubMed PMID: 10512704.
377. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Basinko A, De Braekeleer M. ETV6 fusion genes in hematological malignancies: a review. *Leukemia research*. 2012;36(8):945-61. doi: 10.1016/j.leukres.2012.04.010. PubMed PMID: 22578774.

378. Oikawa T. ETS transcription factors: possible targets for cancer therapy. *Cancer science*. 2004;95(8):626-33. Epub 2004/08/10. PubMed PMID: 15298723.
379. Blair DG, Athanasiou M. Ets and retroviruses - transduction and activation of members of the Ets oncogene family in viral oncogenesis. *Oncogene*. 2000;19(55):6472-81. Epub 2001/02/15. doi: 10.1038/sj.onc.1204046. PubMed PMID: 11175363.
380. Camoes MJ, Paulo P, Ribeiro FR, Barros-Silva JD, Almeida M, Costa VL, Cerveira N, Skotheim RI, Lothe RA, Henrique R, Jeronimo C, Teixeira MR. Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma. *PloS one*. 2012;7(11):e49819. Epub 2012/11/28. doi: 10.1371/journal.pone.0049819. PubMed PMID: 23185447; PMCID: 3501462.
381. Tondravi MM, McKercher SR, Anderson K, Erdmann JM, Quiroz M, Maki R, Teitelbaum SL. Osteopetrosis in mice lacking haematopoietic transcription factor PU.1. *Nature*. 1997;386(6620):81-4. Epub 1997/03/06. doi: 10.1038/386081a0. PubMed PMID: 9052784.
382. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science*. 2013;339(6122):957-9. Epub 2013/01/26. doi: 10.1126/science.1229259. PubMed PMID: 23348506; PMCID: PMC4423787.