MICROARRAY AND GENOME-WIDE SEQUENCING APPROACHES TO CHARACTERIZING DNA BINDING MOLECULES

Thesis by

James William Puckett

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended May 12, 2009)

Acknowledgements

I would like to thank my graduate research advisor, Professor Peter Dervan, for providing laboratory space, research funding, and helpful advice when asked. Thank you for providing me with a wonderful environment in which to learn as a researcher. I would like to thank the members of my thesis committee, Professor Robert Grubbs, Professor Stephen Mayo, and Professor Linda Hsieh-Wilson for their support.

I have had the opportunity to meet and work alongside many wonderful people during my tenure in the Dervan laboratories. I would like to acknowledge my collaborators in lab: Katy Muzikar, Carey Hsu, Nicholas Nickols, and Fei Yang, whose strong work ethics, scientific insight, and perseverance have made much of this work possible. Working in Church 316 has afforded the opportunity to interact daily with Carey Hsu, Sherry Tsai, Fei Yang, and Dan Gubler, all wonderful coworkers. I would like to thank all of the members of the Dervan lab for their friendship, especially my classmates Carey Hsu and Justin Cohen, and a more senior student, Nicholas Nickols, with whom the graduate student journey has been a true pleasure. I would like to thank John Phillips for carpooling with me from West Los Angeles to Pasadena during my final year of graduate study.

I would like to acknowledge the graduate students of the Dervan group during my tenure: Timothy Best, Michael Brochu, David Chenoweth, Raymond Doss, Benjamin Edelson, Michelle Farkas, Eric Fechter, Claire Jacobs, Benjamin Li, David Montgomery, Katy Muzikar, Julie Poposki, Adam Poulin-Kerstien, and Jim Sanchez; and postdoctoral scholars Hans-Dieter Arndt, Christian Dose, Mareike Goeritz, Bogdan Olenyuk, and Anne Viger.

Much of the research performed was in collaboration with the Gottesfeld Laboratory at the Scripps Research Institute in La Jolla, California (TSRI), the Ansari Laboratory at the University of Wisconsin, Madison, the Wells Laboratory at Texas A&M University, and the Pandolfo Laboratory at Université Libre de Bruxelles—Hôpital Erasme. From these laboratories I would like to thank the principal investigators as well as Ryan Burnett

(TSRI), Myriam Raï (Brussels), Leslie Son (Texas), Josh Tietjen (Wisconsin), and Christopher Warren (Wisconsin). Their collaborative efforts have greatly expanded the scope of research that could be studied with polyamides.

I would like to thank members of the Wold laboratory at Caltech for helpful discussions on ChIP-Seq studies: Ali Mortazavi, Anthony Kirilusha, Ken McCue, Brian Williams, and Diane Trout. Dr. Ryan Baugh and Oren Schaedel of the Sternberg lab were kind to show me how to use their fluorometer and allow me use of it. Thanks to Lorian Schaefer of the Millard and Muriel Jacobs Genetics and Genomics Laboratory for processing the ChIP-Seq experiments.

I would like to thank the Caltech staff, especially Professor Dervan's administrative assistant, Ms. Lynne Martinez. Steve Gould and Lillian Kremar have ensured prompt placement of orders. Thanks to Joe Drew and Ron Koen for their prompt delivery of received shipments.

for Joanne and my parents

Abstract

Hairpin and linear beta-alanine linked pyrrole-imidazole (Py-Im) polyamides are programmable oligomers that bind the DNA minor groove sequence-specifically with affinities comparable to those of DNA-binding proteins found in nature. These small molecules have been observed to localize within the nucleus of living cells and modulate endogenous gene expression. Herein, we demonstrate the utility of a linear beta-alanine linked pyrrole-imidazole polyamide to upregulate *frataxin* mRNA and protein expression in a cell line derived from a Friedreich's Ataxia patient. We examine the binding affinities and specificities of additional linear beta-alanine linked polyamides. We examine binding specificities of a Cy3-fluorescently labeled version of the *frataxin* expression modulating-polyamide and a Cy3-labeled polyamide known to downregulate expression of the Vascular Endothelial Growth Factor using DNA microarrays composed of hairpins harboring all 524,800 unique 10 bp DNA sequences. We experimentally verify the correlation of Cy3 fluorescence intensity with quantitative DNase I footprint derived binding affinities. We find that Cy3 dye placement on the polyamide tail versus labeling of an internal pyrrole does not significantly alter DNA specificity. Finally, we examine the genome-wide binding preferences of Androgen Receptor (AR) in LNCaP cells using ChIP-Seq (chromatin immunoprecipitation followed by high-throughput DNA sequencing). We observe the canonical ARE motif to be present in a majority of the immunoprecipitated binding regions. We observe a secondary sequence motif that may be the dimerization of AR with a forkhead protein, an interaction known in the literature but without a defined sequence motif. We also define AR occupancy with respect to location in and about known genes. We correlate gene expression profiles from mRNA microarray data with the ChIP-Seq data.

Table of Contents

Page

Acknowledgements……iii

Dedication……...v

Abstract……..vi

List of Figures and Tables……... viii

Chapter 1 Introduction…… 1

Chapter 2A Linear *β*-linked Polyamides Target GAA Repeats and Alleviate Transcription Repression in Friedreich's Ataxia Cell Culture……..24

Chapter 2B Footprinting Studies of Polyamides Targeting the $[AAG]_3$ Triplet Repeat and Related Sequences……. 47

Chapter 3 Quantitative Microarray Profiling of DNA Binding Molecules……... 69

Chapter 4 Genome-Wide Binding Profile of Androgen Receptor in Dihydrotestosterone-Induced LNCaP Cells……... 100

Appendix A X-ray Crystal Structure of methyl 3-chlorothiophene-2-carboxylate... 120

Appendix B DNase I Footprinting and Alkylation Assays of Linear *β*-linked Compounds and Hairpin Polyamides Targeted to GAA Repeats……... 129

Appendix C CSI Microarray and Quantitative DNase I Footprinting Data for an Internally Cy3-labeled Linear *β*-linked Polyamide Targeting GAA Repeats……138

Appendix D ChIP-Seq Enriched Regions……... 158

Appendix E Chromatin Immunoprecipitation Protocol for LNCaP Cells Used to Generate Initial ChIP-Seq Dataset……..193