Sequence and Structural Analysis of the BTB Domain

by

Peter John Stogios

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Medical Biophysics
University of Toronto

© Copyright by Peter John Stogios 2008
Sequence and Structural Analysis of the BTB Domain

Doctor of Philosophy, 2008

Peter John Stogios

Department of Medical Biophysics
University of Toronto

Abstract

The BTB domain is a eukaryotic protein-protein interaction motif found in variety of proteins. This thesis describes an investigation into the general and specific properties of the sequence, structure and self-association properties of this domain.

The work is divided by two complementary approaches. Chapter 2 describes computational work in assembling a collection of BTB domain sequences from completely sequenced eukaryotic genomes. This chapter describes analyses on this collection including the genomic distribution, domain architectures, identification of putative novel domains and predictions of interactions.

Chapters 3, 4 and 5 are founded on experimental analyses on BTB domains from human BTB-ZF proteins. Chapter 3 describes the structure of the BTB domain from Leukemia/Lymphoma Related Factor (LRF). The structure closely resembles the previously determined structures of BTB domains. The structure showed a large number of sequence substitutions on the surface of the LRF BTB domain that is equivalent to the surface involved in an interaction between the BTB domain from B-Cell Lymphoma 6 (BCL6) and a peptide derived from the SMRT co-repressor (the SMRT-BBD). We show the LRF BTB domain does not interact with this peptide.
Chapter 4 describes the structures of the BTB domains from FAZF and Miz-1. These proteins conserve most of the BTB fold but show some unexpected changes. The BTB domain from FAZF lacks domain swapping which is a novel feature. The BTB domain from Miz-1 contains a naturally truncated N-terminus and a novel movement of 10 residues away from a conserved three-stranded β-sheet. We show these BTB domains are dimeric within a specific concentration range and that they do not interact with the SMRT-BBD.

Chapter 5 describes the structure of the BTB domain from Kaiso. This structure showed interactions between Kaiso BTB domain dimers that extend through the crystal. We identified similar interactions between dimers in a number of other structures of other BTB domains which suggested a common mode of oligomerization.
Acknowledgements

The work in this thesis could not have been possible without the time, effort, support, advice and recommendations from so many people. First and foremost, I’d like to thank the support of my family. In this world filled with so much change, having the unconditional support of my family is the most comforting and stabilizing feeling possible. Thank you to my grandmother and grandfather, whose courage in starting a new life across the world and whose undying love gives me a source of inspiration I will always draw upon. Thank you to my mother and father, who have worked so hard in their lives so that my sister and I could have a world of opportunity. Thank you to my sister Christine of whom I’m so proud and protective of because she deserves nothing but the best in life.

Thank you to my supervisor Gil Privé for his mentorship, friendship and advice. Thanks for teaching me so much scientific but also how to be a leader, manager, presenter, writer and an honest and caring person at the same time.

Much of this thesis relied on the work of Neil Pomroy and Lu Chen, who are the solid foundation the lab is built upon. Thanks to Sylvia Ho for performing the ultracentrifugation experiments, helping with data analysis and advice. Thanks also to Denise Jaworsky and Eden Fussner for reagents and their involvement in numerous wet lab aspects and Greg, Jimmy and Sukhjeen for their involvement in numerous computational aspects. Thank you to all those labmates over the years that have become great friends, especially Farid, Vickie, Eden, Alex, Kosta, Doug, Dona, Tine and Jose. Graduate school is at times a curious endeavour and sharing lab experiences, career aspirations, jokes, gossip and stories with you all made the lab a fun place to be. While the world of science is often cold and impartial, it is actually making the acquaintance of people along the way that makes it all worthwhile.
# Table of Contents

List of Tables ................................................................................................................... x

List of Figures ................................................................................................................. xi

List of Appendices ......................................................................................................... xv

List of Abbreviations ......................................................................................................xvi

## Chapter 1 - Introduction and Thesis Overview ......................................................... 1

1.1. Introduction .............................................................................................................. 2

1.2. Sequence analysis of BTB domains ......................................................................... 6
    1.2.1. Previous sequence analyses of the BTB domain ............................................. 6
    1.2.2. Eukaryotic genome browsers ........................................................................ 8
    1.2.3. Protein sequence comparison and remote homolog detection ....................... 10
    1.2.4. New domain identification.......................................................................... 13

1.3. The BTB fold ........................................................................................................... 17
    1.3.1. Preface ....................................................................................................... 17
    1.3.2. Structure of the BTB domain from BTB-ZF proteins.................................... 18
    1.3.3. T1 domain.................................................................................................. 22
    1.3.4. Skp1 and Elongin C .................................................................................... 23
    1.3.5. The BTB-Cullin 3 interaction ................................................................... 26

1.4. BTB-ZF transcription factors ................................................................................... 29
    1.4.1. Zinc finger proteins ..................................................................................... 29
    1.4.2. Introduction to BTB-ZF proteins ................................................................ 31
    1.4.3. Transcription regulation by BTB-ZF proteins .............................................. 33
        1.4.3.1. Background: histone modifications................................................ 33
        1.4.3.2. Background: mechanisms of transcription repression...................... 35
        1.4.3.3. BTB domains from BTB-ZF proteins recruit co-repressors and
                           histone deacetylases ................................................................. 37
    1.4.4. Dimerization of BTB domains .................................................................... 38
    1.4.5. Oligomerization of BTB domains ............................................................... 40
    1.4.6. Roles of BTB-ZF proteins in cancer ............................................................. 44

1.5. Thesis overview ...................................................................................................... 49

## Chapter 2 - Sequence-structure Relationships of BTB Domain-containing Proteins .......................................................... 54
2.1. Abstract ................................................................................................................ 55

2.2. Introduction ........................................................................................................ 56

2.3. Methods ............................................................................................................. 59
  2.3.1. Strategy for the identification of BTB domains ...................................................... 59
  2.3.2. Structural superpositions of BTB fold-containing proteins ......................................... 61
  2.3.3. Hidden Markov Models (HMM’s) ................................................................ 62
  2.3.4. Sequence alignment .................................................................................... 63
  2.3.5. Genome collection and sequence searches ................................................ 64
  2.3.6. Sequence clustering and most probable sequence detection ................................. 65
  2.3.7. New domain identification ........................................................................ 65
    2.3.7.1. The BACK domain ....................................................................... 65
    2.3.7.2. The PHR domain .......................................................................... 66
  2.3.8. Secondary structure and disorder searches .................................................... 67
  2.3.9. Tertiary structure/fold-recognition predictions ............................................ 68

2.4. Results ................................................................................................................ 69
  2.4.1. BTB fold comparisons ................................................................................ 69
  2.4.2. Representation of BTB domains in fully sequenced genomes ................................. 77
  2.4.3. Evolutionary relationships and sequence clusters ......................................... 81
  2.4.4. New domain identifications ........................................................................ 83
    2.4.4.1. The BACK domain ....................................................................... 83
    2.4.4.2. The PHR domain .......................................................................... 87
  2.4.5. Survey of subfamilies of BTB proteins ......................................................... 90
    2.4.5.1. Long form of the BTB domain ....................................................... 90
    2.4.5.2. BTB-ZF proteins ........................................................................... 91
    2.4.5.3. BTB-BACK-Kelch proteins ........................................................... 92
    2.4.5.4. Skp1 ............................................................................................. 93
    2.4.5.5. ElonginC ...................................................................................... 94
    2.4.5.6. The T1 domain in Kv channels ...................................................... 95
    2.4.5.7. MATH-BTB proteins .................................................................... 95
    2.4.5.8. BTB-NPH3 proteins ...................................................................... 96
    2.4.5.9. BTB-bZip proteins ......................................................................... 97
    2.4.5.10. RhoBTB proteins .......................................................................... 98
    2.4.5.11. BTB-ankyrin proteins .................................................................. 99
    2.4.5.12. BTB domains with no other identified domain ..................................... 99
  2.4.6. BTB domains in Cullin complexes ............................................................. 100
  2.4.7. A model of the ubiquitin-E2-Cul3-Rbx1-BBK complex .................................. 101

2.5. Conclusions ....................................................................................................... 105

Chapter 3 – Crystal structure of the LRF BTB domain and insights into discrimination for binding a peptide from the SMRT co-repressor ..............................................................106

3.1. Abstract .................................................................................................................107

3.2. Introduction ...........................................................................................................108

3.3. Methods .................................................................................................................110
  3.3.1. Cloning of the LRF BTB domain ...............................................................110
  3.3.2. Expression and purification of the LRF BTB domain ..................................110
  3.3.3. Crystallization of the LRF BTB domain .................................................111
  3.3.4. Structure determination and refinement of the LRF BTB domain ..............112
  3.3.5. Peptide binding assay .............................................................................112

3.4. Results and Discussion ............................................................................................113
  3.4.1. Expression and purification of the LRF BTB domain ..................................113
  3.4.2. Crystallization of the LRF BTB domain .....................................................114
  3.4.3. Structure determination of the LRF BTB domain .......................................115
  3.4.4. General description of the LRF BTB domain structure ...............................117
  3.4.5. The lateral grooves of the LRF and BCL6 BTB domains showed many residue differences .........................................................................................................................120
  3.4.6. Comparison of the charged pocket of BTB domains ..................................122
  3.4.7. LRF BTB domain does not interact with the SMRT-BBD .........................124

3.5. Conclusions ............................................................................................................126

Chapter 4 – Domain swapping and the BTB domain: insights from the structures of the BTB domains from FAZF and Miz-1 ........................................................................127

4.1. Abstract .................................................................................................................128

4.2. Introduction ...........................................................................................................129
  4.2.1. Preface ......................................................................................................129
  4.2.2. FAZF is evolutionarily and functionally linked with PLZF .........................130
  4.2.3. Miz-1 is both a transcription activator and repressor ..................................131

4.3. Methods .................................................................................................................132
  4.3.1. Cloning, expression and purification of the FAZF, mutant and Miz-1 BTB domains .................................................................................................................................132
  4.3.2. Crystallization of the FAZF and Miz-1 BTB domains ..................................132
  4.3.3. Structure determination and refinement of the FAZF and Miz-1 BTB domains .................................................................................................................................133
  4.3.4. Conformation of N-terminus in solution via cysteine cross-linking ..........134
  4.3.5. Molecular weight measurement by sedimentation equilibrium analytical ultracentrifugation .........................................................................................................................134
4.3.6. Molecular weight measurement by analytical size exclusion chromatography..................................................................................................135
4.3.7. Peptide binding assay...................................................................................................................................................................................135
4.3.8. Structure comparisons and visualization.................................................................................................................................................135
4.3.9. Structure characters and phylogeny construction........................................................................................................................................136

4.4. Results and Discussion..................................................................................................................................................................................137
4.4.1. Expression and purification of the FAZF, mutant and Miz-1 BTB domains.........................................................................................137
4.4.2. Crystallization of the FAZF and Miz-1 BTB domains.................................................................................................................................139
4.4.3. Structure determination and refinement of the FAZF and Miz-1 BTB domains.........................................................................................140
4.4.4. The FAZF BTB domain was not domain swapped..................................................................................................................................143
4.4.5. The hinge loop region formed interactions unique to the FAZF BTB domain.................................................................................................151
4.4.6. The FAZF BTB domain dimerization interface ...............................................................................................................................151
4.4.7. General description of the Miz-1 BTB domain structure...........................................................................................................................155
4.4.8. The dimerization interface of all BTB domains contain a set of conserved residues at the core...........................................................................................................158
4.4.9 Verification of conformation of N-terminus of FAZF and PLZF BTB domains.................................................................................................159
4.4.10. Oligomeric state of the FAZF and Miz-1 BTB domains..........................................................................................................................162
4.4.11. The lateral groove is not conserved: the FAZF and Miz-1 BTB domains do not interact with the SMRT-BBD......................................................170
4.4.12. Structure-guided evolutionary history of the BTB fold..........................................................................................................................173

4.5. Conclusions..............................................................................................................................................................................................180

Chapter 5 - Crystal structure of the Kaiso BTB domain and a possible mode of oligomerization of BTB domains..................................................................................181

5.1. Abstract .................................................................................................................................................................................................................182

5.2. Introduction ..................................................................................................................................................................................................183
5.2.1. Kaiso is a dual specificity transcription repressor.................................................................................................................................183
5.2.2. Crystal structure of the Kaiso BTB domain suggested a basis for oligomerization of BTB-ZF proteins..................................................184

5.3. Methods.........................................................................................................................................................................................................186
5.3.1. Cloning, expression and purification of the Kaiso BTB domain..............................................................................................................186
5.3.2. Crystallization of the Kaiso BTB domain.................................................................................................................................................186
5.3.3. Structure and refinement of the Kaiso BTB domain.................................................................................................................................187
5.3.4. Analytical size exclusion chromatography..............................................................................................................................................187

5.4. Results ........................................................................................................................................................................................................188
5.4.1. Purification of the Kaiso BTB domain..................................................................................................................................................188
List of Tables

Table Description

2.1 Collection of BTB domain HMM’s used for genome searches ....................... 63

3.1 X-ray data collection and refinement statistics for the LRF BTB domain ........ 115

4.1 X-ray data collection and refinement statistics for the FAZF and Miz-1
   BTB domains........................................................................................................ 140

4.2 Summary of sedimentation equilibrium ultracentrifugation data for the
   FAZF, Miz-1 and PLZF BTB domains................................................................. 164

4.3 23 structural features that describe the BTB fold........................................... 173

5.1 X-ray data collection and refinement statistics for the Kaiso BTB domain...... 190
List of Figures

Figure Description

1.1. Neighbor-joining phylogenetic tree constructed from BTB domain sequences ........ 9
1.2. Growth of unique folds in the PDB ................................................................. 14
1.3. Classification scheme of the BTB fold at SCOP ............................................. 18
1.4. Structure of the PLZF BTB domain ............................................................... 20
1.5. Structure of the BCL6 BTB domain/SMRT-BBD complex ......................... 21
1.6. Structure of the T1 domain ......................................................................... 22
1.7. Structure of the SCF1 complex .................................................................... 25
1.8. Structure of the VCB complex .................................................................... 26
1.9. Examples of BTB domain-containing proteins that interact with Cullin 3 .... 28
1.10. Domain architectures of the main families of ZFP’s .................................... 30
1.11. The 42 BTB-ZF proteins in the human genome ........................................ 32
1.12. Schematic of histone modifications .......................................................... 35
1.13. Model of transcription repression by dimeric BTB-ZF proteins ............... 41
1.15. Dimer-dimer associations in the two structures of PLZF\textsuperscript{BTB} ....... 45

2.1. Flow chart of BTB domain identification pipeline ........................................... 60
2.2. Defining the core BTB fold ......................................................................... 70
2.3. Sequence comparisons of BTB fold-containing proteins .............................. 71
2.4. Majority-rule consensus sequences of BTB domain subfamilies .................. 74
2.5. Protein-protein interaction surfaces in BTB domains ................................. 76
2.6. Representation of BTB domains in eukaryotic genomes ........................................ 80
2.7. BTB sequence clusters and domain architectures .................................................. 82
2.8. Multiple sequence alignment and distribution of the BACK domain ...................... 85
2.9. Multiple sequence alignment of the PHR domain ............................................... 88
2.10. Domain architectures of PHR domain-containing proteins .................................. 90
2.11. Structural model of the ubiquitin-E2-Cul3-Rbx1-BBK complex .............................103

3.1. Expression of 22 BTB domains as thioredoxin-6xhis tag fusions by
Denise Jaworsky .............................................................................................................114
3.2. Native crystals of the LRF BTB domain, grown by Lu Chen ..............................115
3.3. Structure of the LRF BTB domain ......................................................................118
3.4. Comparison of the LRF, PLZF and BCL6 BTB domains ................................... 119
3.5. Comparison of the lateral grooves of the LRF and BCL6 BTB domains .......... 121
3.6. Comparison of the charged pocket of the LRF, PLZF and BCL6 BTB domains 123
3.7. Binding of the SMRT-BBD to the LRF and BCL6 BTB domains ...................... 125

4.1. SDS-PAGE gel of purification of Miz-1 BTB domain ...........................................137
4.2. Gel filtration profiles of Thx-6xhis-Miz-1 fusion and Miz-1 BTB domain ..........138
4.3. Crystals of the FAZF BTB domain, grown by Neil Pomroy ..............................139
4.4. Crystals of the Miz-1 BTB domain .....................................................................140
4.5. Structures of the FAZF and Miz-1 BTB domains ...............................................145
4.6. Electrostatic surface representation of FAZF and Miz-1 BTB domains ............146
4.7. Comparison of the hinge loop regions of the FAZF and PLZF BTB domains ....147
A.2. Crystals of methylated Miz-1 BTB domain ..........................................................255
A.3. Self-interaction chromatography with Kaiso BTB domain as bait .........................257
A.4. Self-interaction chromatography with no bait protein...........................................258
A.5. Location of Kaiso BTB domain mutations S3E and K5D.................................259
List of Appendices

Appendix

1....................................................................................................................................251
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation / Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>APL</td>
<td>Acute Promyelocytic Leukemia</td>
</tr>
<tr>
<td>BACK</td>
<td>BTB And C-terminal Kelch</td>
</tr>
<tr>
<td>BAZF</td>
<td>BCL6-Associated Zinc Finger</td>
</tr>
<tr>
<td>BBD</td>
<td>BCL6 Binding Domain</td>
</tr>
<tr>
<td>BBK</td>
<td>BTB-BACK-Kelch</td>
</tr>
<tr>
<td>BCL6</td>
<td>B-Cell Lymphoma 6</td>
</tr>
<tr>
<td>BCOR</td>
<td>BCL6 Corepressor</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>BTB</td>
<td>Bric-a-brac Tramtrack Broad-complex domain</td>
</tr>
<tr>
<td>BTB-ZF</td>
<td>Proteins containing a BTB domain and zinc finger motifs, also known as POK</td>
</tr>
<tr>
<td>bZip</td>
<td>Basic region leucine Zipper</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-Dependent Kinase</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complimentary DNA</td>
</tr>
<tr>
<td>CNS</td>
<td>Crystallography and NMR System</td>
</tr>
<tr>
<td>COG</td>
<td>Clusters of Orthologous Groups</td>
</tr>
<tr>
<td>CtBP</td>
<td>C-terminal binding protein</td>
</tr>
<tr>
<td>Cul3</td>
<td>Cullin 3</td>
</tr>
<tr>
<td>DBC2</td>
<td>Deleted in Breast Cancer 2</td>
</tr>
<tr>
<td>DLBCL</td>
<td>Diffuse Large B-Cell Lymphoma</td>
</tr>
<tr>
<td>ECS</td>
<td>Elongin C - Cullin2 - SOCS box</td>
</tr>
<tr>
<td>ETO</td>
<td>Eight twenty one protein</td>
</tr>
<tr>
<td>Fc</td>
<td>Calculated structure factor</td>
</tr>
<tr>
<td>Fo</td>
<td>Observed structure factor</td>
</tr>
<tr>
<td>FAZF</td>
<td>Fanconi Anemia Zinc Finger, also known as PLZP</td>
</tr>
<tr>
<td>FBI-1</td>
<td>Factor that binds to IST</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>FL</td>
<td>Follicular Lymphoma</td>
</tr>
<tr>
<td>FYVE</td>
<td>Fab1, YOTB/ZK632.12, Vac1, and EEA1 zinc finger</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone Acetylase</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deacetylase</td>
</tr>
<tr>
<td>HLH</td>
<td>Helix-Loop-Helix</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Pressure Liquid Chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized Metal Affinity Chromatography</td>
</tr>
<tr>
<td>KBS</td>
<td>Kaiso Binding Site</td>
</tr>
<tr>
<td>KOG</td>
<td>cluster of euKaryotic Orthologous Groups</td>
</tr>
<tr>
<td>KRAB</td>
<td>Krüppel Associated Box</td>
</tr>
<tr>
<td>LOV</td>
<td>Light, oxygen or voltage domain</td>
</tr>
<tr>
<td>LRF</td>
<td>Leukemia/lymphoma Related Factor, also known as FBI-1, OCZF, or Pokemon</td>
</tr>
<tr>
<td>LRR</td>
<td>Leucine Rich Repeat</td>
</tr>
<tr>
<td>LSE</td>
<td>Lineage Specific Expansion</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix Assisted Laser Desorption/Ionization - Time of Flight</td>
</tr>
<tr>
<td>MATH</td>
<td>Meprin And Traf Homology</td>
</tr>
<tr>
<td>MGC</td>
<td>Mammalian Gene Collection</td>
</tr>
<tr>
<td>Miz-1</td>
<td>Myc Interacting Zinc finger 1</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular Replacement</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>mSin3A</td>
<td>mammalian Swi-independent 3A</td>
</tr>
<tr>
<td>MTA3</td>
<td>Metastasis-associated protein 3</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>N-CoR</td>
<td>Nuclear Corepressor</td>
</tr>
<tr>
<td>NHL</td>
<td>Non-Hodgkin’s Lymphoma</td>
</tr>
<tr>
<td>NPH3</td>
<td>Nonphototropic hypocotyl 1</td>
</tr>
<tr>
<td>OCZF</td>
<td>Osteoclast derived Zinc Finger</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol</td>
</tr>
<tr>
<td>PHR</td>
<td>PAM, Highwire, RPM1</td>
</tr>
<tr>
<td>PLZF</td>
<td>Promyelocytic Leukemia Zinc Finger</td>
</tr>
<tr>
<td>PLZP</td>
<td>PLZF-like Protein, also known as FAZF</td>
</tr>
<tr>
<td>POK</td>
<td>Proteins contain a BTB/POZ domain and Krüppel type zinc finger motifs, also known as BTB-ZF</td>
</tr>
<tr>
<td>POU</td>
<td>Pit-1, Oct-1, Unc-86 domain</td>
</tr>
<tr>
<td>POZ</td>
<td>Poxvirus Zinc finger</td>
</tr>
<tr>
<td>PSI-BLAST</td>
<td>Position-Specific Iterative BLAST</td>
</tr>
<tr>
<td>PSSM</td>
<td>Position-Specific Scoring Matrix</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>Crystallographic R-factor, free set</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;</td>
<td>Crystallographic R-factor, working set</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>R-factor, intensity merging</td>
</tr>
<tr>
<td>Rho</td>
<td>Ras Homology</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean-square Deviation</td>
</tr>
<tr>
<td>RPS-BLAST</td>
<td>Reversed Position Specific BLAST</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Squared Deviation</td>
</tr>
<tr>
<td>SCAN</td>
<td>SRE-ZBP, CTfin51, AW-1, Number 18 cDNA domain</td>
</tr>
<tr>
<td>SCF</td>
<td>Skp1 - Cullin - F-box</td>
</tr>
<tr>
<td>SCOP</td>
<td>Structural Classification of Proteins</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>Skp1</td>
<td>S-phase kinase-associated protein 1A</td>
</tr>
<tr>
<td>Skp2</td>
<td>S-phase kinase-associated protein 2</td>
</tr>
<tr>
<td>SMRT</td>
<td>Silencing Mediator of Retinoid and Thyroid receptor</td>
</tr>
<tr>
<td>SOCS</td>
<td>Suppressor Of Cytokine Signaling</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface Plasmon Resonance</td>
</tr>
<tr>
<td>STAT</td>
<td>Signal Transducers and Activators of Transcription</td>
</tr>
<tr>
<td>TCEP</td>
<td>tris(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td>Thx</td>
<td>Thioredoxin</td>
</tr>
<tr>
<td>TPR</td>
<td>Tetratricopeptide repeat</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)-aminomethane</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>VAST</td>
<td>Vector Alignment Search Tool</td>
</tr>
<tr>
<td>VCB</td>
<td>VHL - ElonginC - ElonginB</td>
</tr>
<tr>
<td>VHL</td>
<td>von-Hippel Lindau</td>
</tr>
<tr>
<td>ZF</td>
<td>Zinc Finger</td>
</tr>
<tr>
<td>ZFP</td>
<td>Zinc Finger Protein</td>
</tr>
<tr>
<td>ZID</td>
<td>Zinc finger with Interaction Domain</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction and Thesis Overview
1.1. Introduction

The BTB domain is a ~120 residue protein-protein interaction motif that was initially identified as a N-terminal region with significant sequence homology between the *Drosophila melanogaster* transcription regulatory proteins Bric-a-brac (BAB), Tramtrack (TTK) and Broad-Complex (BR-C) (DiBello et al., 1991, Godt et al., 1993). It is also known in the scientific literature as the POZ domain as homology was detected between poxvirus proteins and various zinc finger proteins (Bardwell & Treisman, 1994). The BTB domain is found in eukaryotes only, with no known prokaryotic examples. Proteins containing the BTB domain contain a variety of domain architectures and perform many cellular functions. There has been no enzymatic activity associated with the domain and the main role of the domain appears to be in protein-protein interactions, most notably self-association. The two largest families of BTB domain-containing proteins are those that also contain classical C\textsubscript{2}H\textsubscript{2}/Krüppel-type zinc finger motifs, and proteins that also contain Kelch motifs.

BTB-Zinc Finger (BTB-ZF) proteins are often involved in the transcriptional regulation of a number of genes that play important roles in regulating cell differentiation and proliferation. For example, PLZF regulates the expression of *Hox* genes, which are well-known to be involved in processes such as embryogenesis, limb development and hematopoiesis (reviewed in Zakany & Duboule, 2007, Eklund, 2007). It is therefore not surprising that aberrant regulation of the expression of BTB-ZF genes alters the transcriptional programme of cells and often results in tumourigenesis (Kelly & Daniel, 2006).

The BTB domain in BTB-ZF proteins has a variety of functions: homodimerization, interactions with different components of chromatin remodeling complexes (such as nuclear
co-repressors), nuclear localization and transcription repression. The BTB domain is possibly involved in heterodimerization or oligomerization of BTB-ZF proteins. To begin to understand the structural basis for these various activities of the BTB domain, work in the Privé lab that preceded my studies focused on crystallographic determination of the PLZF and BCL6 BTB domains, as well as characterization of interactions between the BTB domain and the nuclear co-repressor SMRT. The first reported BTB domain structure was from PLZF (Ahmad et al., 1998, Li et al., 1999), revealing a novel mixed α/β fold with an extensive dimerization interface comprised of association of the N-terminus of the domain with the bulk of the partner chain. A charged pocket found at the top of the dimerization interface, large enough to accomodate a peptide, was hypothesized to provide a nuclear co-repressor binding surface. As well, both PLZF\textsuperscript{BTB} structures showed interactions between homodimers via formation of a four-stranded β-sheet between equivalent two-stranded β-sheets at the “floor” of the homodimer. The biological relevance of this interface was not clear, as the PLZF BTB domain behaved as a homodimer in solution for both groups, but was suggested as a possible mode of oligomerization of BTB-ZF proteins.

The structure of the complex between the BCL6 BTB domain and a 17-residue fragment from SMRT showed the two peptides symmetrically interacting with a single homodimer. The peptide, termed the BCL6 Binding Domain or BBD, interacts with BCL6\textsuperscript{BTB} with a dissociation constant of 11 μM as measured by isothermal titration calorimetry (ITC) (Ahmad et al., 2003). The N-terminus of each peptide adopts a β-sheet conformation and interacts with the N-terminal β1 secondary structure element, followed by an extended conformation along a groove (termed the lateral groove) formed at the dimer interface. Muta-
tional and functional analyses identified two residues in the BCL6 BTB domain that were critical for the interaction and for transcription repression function by full-length BCL6.

It is known that the expression of genes coding for BTB-ZF proteins varies according to a variety of factors, including cell type and cell cycle. As well, BTB-ZF proteins interact with different DNA sites and therefore regulate the transcription of different gene targets. Furthermore, BTB domains have diverse interaction partners, including different components of the transcription repression machinery. From a sequence perspective, multiple analyses have shown that there are regions within the BTB domain that are more conserved than others.

Taking these observations into account, it is unclear what features of BTB domain-containing proteins are specific to the individual examples, and which are general features of the family. Therefore, the basic premise of my graduate studies was that a family-wide perspective was required to address this question. Experimentally, additional crystal structures of the BTB domain could shed light on the mechanisms of dimerization, oligomerization, and co-repressor interactions. Chapters 3, 4 and 5 of this thesis describe the crystal structures of the BTB domains from the BTB-ZF proteins FAZF, Miz-1, Kaiso and LRF. These structures revealed conserved features across all BTB domain structures, such as the core of the dimerization interface, and specific features, such as the lack of domain swapping in FAZF\textsuperscript{BTB}. As well, characterization of their oligomeric states and interactions with the SMRT-BBD were performed and each BTB domain was found to be exclusively homodimeric in solution, and only the BCL6 BTB domain was able to interact with the SMRT-BBD peptide.
It has long been appreciated that the BTB domain is found in a large number of proteins in many genomes. The extensive efforts to sequence eukaryotic genomes has lead to the rapid growth of genome sequence databases, and we undertook an analysis of the genomic distribution, evolution, and sequence conservation of the BTB domain. The focus of Chapter 2 of this thesis was to describe our generation of a high quality, manually curated, web-accessible sequence database of BTB domain-containing proteins and a variety of analyses performed on this sequence collection. Our analysis illustrated that the BTB domain evolved in a species-specific manner in different protein types, while the fold itself is maintained across widely divergent proteins. The domain is known to mediate a variety of protein-protein interactions and this study showed that residues at the surface of the fold are extremely variable. We also identified two new domains in BTB domain-containing proteins, one of which may play a vital role in BTB-Cullin 3 E3 ubiquitin ligase complexes.
1.2. Introduction to sequence analysis of BTB domains

1.2.1. Previous sequence analyses of the BTB domain

At the time of the initial identification of the BTB domain by sequence comparisons, there was little known in the way of structure or function of this motif. Thus, early studies focused on identifying additional proteins that contained the BTB domain. Genomic Southern blot analysis of *Drosophila melanogaster* was performed using nucleotide sequences derived from the gene coding for *Bric-a-brac* (Zollman *et al.*, 1994), revealing an estimate of 40 BTB genes in the fly.

Concurrent with the identification of the BTB domain in *Drosophila* proteins, BLAST searches revealed a similarity between the N-terminal regions of the poxvirus proteins A55R and C2L, *Drosophila* BR-C and TTK, and human KUP which is a zinc-finger containing protein (Chang-Yeh *et al.*, 1991, Koonin *et al.*, 1992). Five non-zinc finger repeats of 47-50 residues towards the C-terminus of A55R and C2L were noticed, indicating the BTB domain was found in proteins other than zinc finger proteins. Around the same time, groups began implicating the human PLZF and BCL6 proteins in the cancers APL and NHL, respectively (Chen *et al.*, 1993, Kerckaert *et al.*, 1993, Ye *et al.*, 1993, Miki *et al.*, 1994). Due to this heightened disease relevance, and the continued growth of public sequence databases, BTB domains were identified in increasingly greater numbers. Similarity was noticed between the N-termini of the human proteins ZID, PLZF, BCL6, KUP, mouse ZF5, *Drosophila* GAGA factor, and a number of poxvirus proteins (Bardwell & Treisman, 1994). This group coined the term “POZ” domain, for poxvirus and zinc finger domain, which is a name still in use today. The terms BTB, POZ, or BTB/POZ are used interchangeably,
while the terms BTB-ZF, POZ-ZF or POK are all used to describe proteins containing BTB domains and \( \text{C}_2\text{H}_2/\text{Krüppel} \) zinc finger motifs.

The development of iterative database searching using PSI-BLAST allowed for further identification of more remote homologs of the BTB domain (Aravind & Koonin, 1999, Altschul et al., 1997). Using different BTB domain sequences as queries to exhaustively search the Genbank database, statistically significant sequence similarity was discovered between proteins of widely different domain architectures in many different organisms. BTB domains were found in proteins containing the MATH domain, ankyrin repeats, kelch repeats and leucine zipper motifs, in organisms including *Caenorhabditis elegans*, *Aplysia californicum*, *Dictyostelium discoideum*, *Drosophila*, *Rattus norvegicus* and *Homo sapiens*. The most unexpected result of this study was similarity between the BTB domain and the T1 tetramerization domain of potassium channel proteins, which appeared after multiple PSI-BLAST iterations. The BTB/POZ domain was known to self-associate and this group predicted the structure of the BTB domain using the tetrameric T1 crystal structure (Kreusch et al., 1998). Based on the structural model presented in the Aravind and Koonin paper, number of predictions were made that were verified by subsequent structure determinations of BTB domains in the Privé lab and in this work. These included that most of the conserved residues are found in the hydrophobic core of the BTB domain, and that the extended region before the last two helices of the BTB domain could accept a partner \( \beta \)-strand. As well, they made some predictions that proved to be incorrect, including that the polar residues involved in homo-tetramerization interactions in the T1 domain would be involved in inter-subunit interactions in all BTB domains. Another incorrect prediction was that the concave three-stranded \( \beta \)-sheet in the BTB domain could serve as a peptide-binding surface, perhaps for
the co-repressor SMRT or Sin3. Phylogenetic analysis performed by Aravind and Koonin showed that the BTB domain sequence could be grouped into a variety of clades defined by the species and the identity of the non-BTB domain in the full-length protein. For example, the *Drosophila* zinc finger proteins formed a single monophyletic cluster, while animal T1 domains formed a separate clade (Figure 1.1).

The study by Aravind and Koonin was a seminal paper in the study of BTB domains and established many of the themes regarding the sequence, structure and function of BTB domains that I hoped to build upon during my graduate studies. These themes include: that the BTB domain is widely distributed in many genomes; is found in many different protein types; mediates a variety of homophilic and heterophilic interactions; and that the BTB domain sequence has diverged greatly over time and extensive iterative database searches are necessary to retrieve the full collection of sequence homologs.

### 1.2.2. Eukaryotic genome browsers

Genomic sequence browsers were developed to interrogate the large volume of genome sequence data. The three main genome browsers are the UCSC Genome Browser (*Kent et al.*, 2002), the NCBI Entrez Genome service and Ensembl. We chose the Ensembl browser because its convenient and intuitive data scheme, as well as its application programming interface (API) written in Perl and therefore easily accessible with the Bioperl package.
FIGURE 1.1.
Neighbor-joining phylogenetic tree constructed from BTB domain sequences, adapted from (Aravind & Koonin, 1999). Tree reveals that the BTB domain sequences can be separated into organism- and domain-architecture specific clades. Fans show groups of individual sequences that were not resolved further. Numbers at nodes indicate bootstrap support values out of 100 generated trees.

Ensembl is based on an automated eukaryotic gene annotation system that provides evidence-based gene predictions, from protein and cDNA homology (Birney et al., 2004, Curwen et al., 2004). The NCBI Builds are utilized for the reference human and mouse genomes, with other genomes provided by their species-specific genome sequence and annotation projects, such as Flybase for the Drosophila genome (Grumbling & Strelets, 2006). En-
sembl uses Uniprot/Swissprot and NCBI’s Refseq, databases of fully manual and partially manual curated protein sequences, respectively, to align all known protein sequences to the reference genomes. Protein sequences from multiple species are used to increase the coverage of predicted genes. From the initial alignment, the pipeline adds intron-exon, splice sites, untranslated regions (UTR’s) and other genomic features to build gene structures. Annotations from other sources are added to genes, transcripts or peptide sequences, such as SNP’s, regulatory regions, InterPro and PFAM domains, Gene Ontology (GO) terms, microarray data.

Perhaps the greatest strength of the Ensembl database from a bioinformatics perspective is the Perl API. The API provides modules that interface with the Ensembl MySQL database. In our case, we used the API to access peptide translations of entire genomes, which we then searched with our collection of profile Hidden Markov Models (HMM’s) that describe the sequence properties of the BTB domain. As well, the annotations from other sources, especially the locations of non-BTB domains from Interpro and PFAM, were retrieved. The pipeline designed for the detection and storage of BTB domain sequences is described in section 1.2.4.

1.2.3. Protein sequence comparison and remote homolog detection

The major focus of my graduate work was to gain insights into the sequence, structure, and evolutionary relationships of BTB domain-containing proteins. As suggested by earlier studies and BLAST searches, the BTB domain is found in a large variety of protein types and the sequence itself has high diversity when compared within and between these different protein types. Therefore, approaches that allowed for the detection of remote homologs were needed to maximize the sensitivity of searches. Importantly, we wished to
keep a low false-positive detection rate, which we accomplished by taking advantage of the structural conservation of the BTB fold (see section 1.3).

A major focus of bioinformatics research is devoted to the detection of remote protein homology (reviewed in Fariselli et al., 2007, Wan & Xu, 2005). Although similar protein structures could have evolved by evolutionary divergence or convergence, i.e. they could be homologous or analogous (Russell et al., 1997), conservation of protein structure is greater than sequence conservation. Similarity of structure is the best indicator of functional homology and structural alignment is the gold standard of protein comparison. Unfortunately, structures for all proteins are not known and functional and/or evolutionary relatedness must be inferred from protein sequence comparisons. Research in detecting remote protein homology has evolved from pairwise sequence comparisons, to profile-sequence comparisons, to methods that incorporate structural features such as secondary structure and 3D profiles, to machine-learning approaches.

Early pairwise sequence comparisons algorithms include the Smith-Waterman/Needleman-Wunsch alignment algorithms and the heuristic database search methods BLAST and FASTA. Both use amino acid substitution matrices such as BLOSUM (Henikoff & Henikoff, 1992) to quantify amino-acid similarity. These methods and matrices are well established but are limited in their sensitivity since they rely on the use of single sequences as queries. Importantly, structural similarity can be missed.

The development of profiles and position-specific scoring matrices (PSSM’s) were a major step forward in the detection of remote protein homology (Gribskov et al., 1987, Park et al., 1998). Profiles take into account that different regions of proteins vary in their selective pressure. Profiles also take into account insertions/deletions. Therefore, profiles better
account for structural similarity of proteins. Generating a profile requires a pre-computed alignment of a group of proteins, preferably sampling a reasonable level of sequence divergence. A table or matrix is generated against which individual sequences from a sequence database can be compared and scored. Iterative searching and re-generation of profiles was implemented in PSI-BLAST, which remains today a widely used and effective algorithm for detecting protein similarity which previously could only be detected by structural superposition (Altschul et al., 1997, Aravind & Koonin, 1999).

Most recently, machine-learning approaches have been adapted to detect remote protein homologies. Hidden Markov Models (HMM’s), neural networks (NN’s) and support vector machines (SVM’s) are the main types of approaches utilized. Like PSI-BLAST, HMM’s utilize position-specific profiles, but are based on a more rigorous probabilistic framework and therefore are useful for identifying statistically relevant hits (Eddy, 1998, Krogh et al., 1994). The most popular application of HMM’s is the PFAM database, which provides a single, expert-tested HMM for each structural protein domain (Bateman et al., 2004).

We chose to use HMM’s to capture the sequence diversity of the BTB domain. Since the BTB domain was previously shown to have evolved along species and domain architecture-dependent lines (Aravind & Koonin, 1999), it was necessary to develop more than a single HMM. Instead, a collection of HMM’s that describe these various sequence subfamilies was created. The PFAM database contains HMM’s for the BTB, T1 and Skp1 variants of the BTB fold, which covers most but not all of the diversity in species and domain-architectures of BTB domain containing proteins. As initiating BLAST or PSI-BLAST queries with different query sequences allows for exploration of different regions of
“sequence space”, having a family of HMM’s allowed us to increase the scope of our search. Importantly, we based our HMM’s on sequence alignments derived from structural superpositions, ensuring a high specificity for the retrieval of true BTB domains.

We used the software package HMMER which provides a number of convenient tools for manipulating HMM’s. HMM’s can be generated from a structure-based multiple sequence alignment and calibrated. They can be used to rapidly search sequence flat files and to generate multiple sequence alignments. As well, HMMER can generate the most probable/majority-rule sequence encoded in an HMM. For the purposes of our BTB domain detection pipeline, there are Bioperl modules to interface with HMMER.

1.2.4. New domain identification.

Eukaryotic proteins are usually comprised of one or more domains, defined as a region that has a compact structure, folds autonomously, has a specific function, and is evolutionarily conserved. One use of our high quality collection of protein sequences is the identification of regions of conservation that lie outside of known domains that could correspond to novel domains. The identification of novel domains in proteins can facilitate functional discovery. Statistical analysis of the PDB has shown that deposition of structures with new folds is still occurring at a significant pace, with an average of 101 new folds deposited per year from 2000-2004 (Figure 1.2). This suggests that exploration of “protein fold space” is not near completion. Therefore, there likely remain a significant number of new domains to be discovered.
Traditional biochemical studies are often focused on domains, and structural studies in particular require knowledge of the precise domain boundaries, or at least regions of fixed structure, for the expression of high levels of soluble protein. Domains can be identified computationally, via conservation analysis with traditional PSI-BLAST searches against known proteins, or searches against domain profiles or HMM’s using RPS-BLAST, PFAM or SMART, for example.
Ideally, the sequence of a putative new domain should retrieve no significant hits when searched against profiles or HMM's of known domains. Further detailed structural analysis is required to conclusively show the sequence may correspond to a novel domain. These include secondary structure prediction, disorder prediction, and fold recognition. Taking into account that these predictions have varying levels of accuracy, these predictions would be compared against the PDB and domain databases to look for the absence of matches to known domains. We undertook each of these analyses with novel domains found in BTB domain-containing proteins.

Of each level of protein structure prediction, secondary structure prediction algorithms are the most accurate. In the context of novel domain identification, the secondary structure content and order may provide a unique signature. There are a number of algorithms and web-based services for secondary structure prediction, including PredictProtein (based on the PHD and PROF algorithms) (Rost et al., 2004, Rost, 1996), PSI-PRED (McGuffin et al., 2000) and JPred (Cuff et al., 1998). PredictProtein, PSIPRED and the latest version of JPred (JNet) all use neural networks to predict alpha helix, beta strand and random coil regions. Each also utilizes multiple sequence alignment and profiles to increase prediction accuracy. The published accuracy rates for PredictProtein (PHD), PSIPRED and JPred are 72 and 76 and 76%, (McGuffin et al., 2000, Rost & Sander, 1994, Cuff & Barton, 2000). In addition, each algorithm returns a per-residue confidence score to aid the user.

Due to realizations that intrinsically disordered regions of proteins often mediate protein-protein interactions and can even exist as hubs in protein-protein interaction networks, disorder prediction and analysis is a growing field (reviewed in Tompa, 2002). A putative new domain should have a low level of predicted disorder. We utilized the
PONDR (Romero et al., 2001) and DISOPRED2 (Ward et al., 2004) services. These algorithms utilize neural networks that were trained on short sequences that showed missing electron density in crystal structures. The prediction accuracy rate of PONDR is reasonable, with 78% success in predicting order, and 60% for predicting disorder (PONDR website, http://www.pondr.com/pondr-tut3.html).

Tertiary structure prediction has been the focus of a great deal of research. As a protein’s structure provides a rich amount of information regarding its function, it is thought that the gap between the extraordinary amount of sequence data generated by genome sequencing projects and functional assignment can be bridged by protein structure prediction. At present, it is not considered possible to predict the high-resolution (<5 Å) structure of a protein from its amino acid sequence without some level of comparison to a library of protein structures. Therefore, for our purposes, we did not attempt to predict the structures of novel domains found in BTB domain-containing proteins. Instead, we submitted our putative novel domain sequences to publicly available fold recognition/threading services that attempt to match a sequence with a library of experimentally determined structures. If no significant hits were found, this suggested the novel domain could adopt a novel fold. We used the Phyre fold recognition server, previously known as 3D-PSSM (Bennett-Lovsey et al., 2008), and the mGenTHREADER servers (McGuffin et al., 2000).
1.3. The BTB fold

1.3.1. Preface

Structure determinations of Skp1, ElonginC and the T1 domain of potassium channels unexpectedly revealed the BTB fold. These proteins have little mutual sequence similarity and apparently no functional similarity with the BTB domain from the BTB-ZF and BBK families at the time of their initial structure determinations. The full description grouping the structures of the BTB domain from BTB-ZF proteins, the T1 domain, Skp1 and Elongin C proteins at SCOP (Structural Classification of Proteins, (Murzin et al., 1995) is illustrated in Figure 1.3. They are all considered to have the “POZ domain” Fold.

SCOP groups protein structures into a hierarchy that takes into account structural similarity, evolutionary history and sequence identity. The “common fold” definition is met since these BTB domain, Skp1, ElonginC and T1 domains contain nearly identical arrangement and connectivity of secondary structures; differences in structure occur from small insertions or peripheral elements added to the N- or C-termini. The formal definition for common fold suggests these proteins may not have obvious common ancestry and could have arisen from physical principles, such as favourable packing arrangements of secondary structures. However, since the structures are all grouped in the “POZ domain” Superfamily, SCOP suggests a common evolutionary origin is probable.

For our purposes, we use the term “BTB fold” when comparing the structures of different proteins and highlighting their shared features, and the terms BTB domain, T1 domain, Skp1 and Elongin C when describing the individual proteins or domains.
1.3.2. Structure of the BTB domain from BTB-ZF proteins

The crystal structures of the BTB domains from PLZF and BCL6 were the first two structures determined from human BTB-ZF proteins, both solved by Farid Ahmad in our lab. The structures are very similar, with average RMSD of equivalent Cα atoms of 1.0, the main difference being the size of the loop between secondary structure elements α3 and β4. The structures were homodimeric and showed that the two N-terminal secondary structure elements, α1 and β1, were exchanged with the partner chain (Figure 1.4). The PLZF BTB domain structure was described as “suggestive of a 3D domain-swapped dimer”, or a “candidate for 3D domain swapping” (Ahmad et al., 1998, Liu & Eisenberg, 2002, Bennett et al., 1995). Domain swapping is defined as “a mechanism for two or more protein molecules to form a dimer or higher oligomer by exchanging an identical structural element (‘domain’)
(Bennett et al., 1995). An additional requirement of the formal definition is evidence of both a dimeric molecule with the domain-swapped conformation and a monomer with the non-swapped conformation. The structure of the BTB domain was not considered truly domain-swapped as there was no evidence of a monomeric BTB domain with the swapped β1 and α1 associated with their own chain.

The dimerization interface of the PLZF BTB domain was comprised of two parts: the “closed interface”, involving β1, β5 and α6, and the “open interface” comprised of α1, α2 and α3. The closed interface involves a short two-stranded β-sheet, comprised of four residues from each of β1 and β5, and burial of residues in α6 against this β-sheet. The face of the β-sheet opposite to α6 and one edge of β1 are exposed to solvent. The open interface is comprised of many hydrophobic residues from α2 and α3 that become buried. Domain swapping involves mainly the closed interface. In total, the entire dimerization interface buries 1972 and 1876 Å² in the PLZF and BCL6 BTB domains, respectively.

A possible “hinge loop” sequence in the PLZF BTB domain, or residues that could alter their conformation to switch the swapped elements to associate with their own or the partner chain, was identified as residues H16, P17, T18 and G19 (Ahmad et al., 1998). These residues are found between β1 and α1. These residues are involved in a number of bridging contacts between the chains in the PLZF BTB homodimer and take part in both the open and closed dimerization interfaces.
The structural basis of the interaction between the BCL6 BTB domain and the SMRT co-repressor was revealed in a co-crystal structure (Figure 1.5) (Ahmad et al., 2003). The 17-residue long BCL6 Binding Domain (or BBD, comprising residues SMRT 1414-1430) interacts with BCL6 BTB via a short \( \beta \)-sheet that is added to the solvent-exposed edge.
of β1 in the β1-β5 sheet from the BTB domain, continues along a surface on the “front” of the dimer (named the lateral groove) and crosses the dimerization interface. Thus, the BCL6 BTB dimer is necessary for SMRT interaction. Two peptides bind to a single BCL6 BTB dimer with a dissociation constant of approximately 15 μM.

The interaction buries 1020 Å² between a dimer and one SMRT-BBD peptide. Some residues in the BCL6 BTB domain were shown to be necessary for the interaction, including H116, which forms a clasp over the peptide, and N21, which forms three hydrogen bonds with the peptide and a bridging water molecule. Importantly, mutations of these two residues in the lateral groove greatly reduces the transcription repression function of BCL6 in cell-based transcription repression assays (Ahmad et al., 2003).

**FIGURE 1.5.**

A) Structure of the BCL6 BTB domain/SMRT-BBD complex shown in ribbon representation. BTB domain dimer is coloured blue and red, SMRT-BBD peptides are coloured green and yellow. N-termini of the peptides are labeled. B) Surface representation of the BCL6 BTB domain, with sticks shown for the SMRT-BBD. Residue H116, which was shown to be crucial for the interaction, is labeled. Taken with permission from Ahmad et al., 2003.
1.3.3. T1 domain

The T1 domain is the amino-terminal cytoplasmic domain that precedes the transmembrane helices in voltage-gated potassium channels. This domain mediates specificity of tetrameric channel assembly between the multiple subfamilies of potassium channels. Tetramerization is mediated by a polar interface that buries 20% of the total solvent-accessible surface area (Figure 1.6) (Kreusch et al., 1998). The interface is conserved within but not between the channel subfamilies. After PSI-BLAST searches detected remote homology between BTB domain sequences and the T1 domain, it was predicted that the BTB domain would adopt a similar structure to the T1 domain (Aravind & Koonin, 1999). The prediction was verified when the PLZF BTB domain structure was

**FIGURE 1.6.**

A) Structure of the T1 domain of the Shaker voltage-gated potassium channel. Three subunits of the tetramer are shown. Secondary structure elements are labeled. Adapted from Kreusch et al., 1998).
determined and indeed showed that the T1 domain is comprised exclusively of the core BTB fold (RMSD of 1.5 Å over 90 Cα atoms), lacking α1 and β1 from the BTB-ZF proteins.

1.3.4. Skp1 and Elongin C.

Skp1 and Elongin C function in the ubiquitin ligase system. The ubiquitin ligase system is involved in post-translational regulation of the levels of proteins in eukaryotic cells. Ubiquitin is a small, 76-residue protein that becomes covalently attached to lysines on proteins to target them for degradation by the 26 S proteosome.

A three-step process mediates the attachment of ubiquitin to proteins. The first step is catalyzed by an ubiquitin-activating enzyme, or E1, that uses ATP hydrolysis to produce a thioester linkage between the E1 active site cysteine residue and the C-terminus of ubiquitin. The second step, or ubiquitin transfer, is catalyzed by E2 enzymes, which transfer the ubiquitin from the E1 to a cysteine residue in the E2. The last step, or ubiquitin ligation, is catalyzed by E3 enzymes, of which there are different classes and numerous examples. The E3 enzyme brings the E2-ubiquitin complex and the target protein into close proximity to allow for ubiquitin transfer to the target. One class of E3’s is the SCF complex (for Skp1-Cullin-F-box). In human cells, there is one Skp1 protein, seven Cullins and approximately 44 F-box proteins. F-box proteins contain the F-box motif, a small conserved three-helix structure that is recognized by Skp1, and an substrate recognition module that can be many different protein-protein interaction domains, such as leucine-rich repeats (LRR’s), the WD40 domain, or ankyrin repeats. Thus, Skp1 is a called an “adaptor” for linking the Cullin 1-E2-ubiquitin backbone of the complex to the substrate-recognizing F-box proteins.
The structure of a SCF1 complex has been determined and it was noticed that Skp1 adopts the BTB fold (Figure 1.7, Schulman et al., 2000). Skp1 contains the core BTB fold (1.4 Å over 90 Cα atoms), a helical insertion within the core fold and a C-terminal extension of two helices. The C-terminal extension and the last two helices of the BTB fold interact with the F-box. A non-overlapping surface on Skp1 interacts with the first Cullin repeat of Cullin 1.

Another E3 ubiquitin ligase is the VCB (VHL-Elongin C-Elongin B) complex. VHL is a tumour suppressor protein and tumour-derived mutations destabilize the VHL-Elongin C interaction. The structure of this complex revealed that Elongin C adopts the BTB fold, but it lacks the last helix (Figure 1.8, Stebbins et al., 1999). This structure showed the α-domain of VHL has a structure very similar to the F-box, and the VHL β-domain is thought to provide a substrate-recognition module for the VCB complex (Schulman et al., 2000).

As well, Elongin C interacts with Cullin 2 or Cullin 5 (Kamura et al., 1998, Yu et al., 2004). Therefore, Elongin C is considered an adaptor protein for the VCB or ECS (Elongin C - Cullin 2 - SOCS box) ubiquitin ligase complex.
A) Structure of the full SCF1 complex. The grey dashed box is the Skp1-Skp2 F-box protein region of the structure and is zoomed in part B. Taken from Zheng et al., 2002. B) The Skp1-Skp2 interaction. The LRR repeats are the substrate-recognition module, while the F-box mediates interaction with Skp1. Skp1 adopts the BTB fold, with a C-terminal helical extension, as shown below the structure. Taken from Schulman et al., 2000.
1.3.5. The BTB-Cullin 3 interaction

A number of studies were published that identified proteins that contain a BTB domain can interact with Cullin 3 (reviewed in Willems et al., 2004, Pintard et al., 2004). These proteins often contain a N-terminal BTB domain and C-terminal Kelch motifs, but other types of BTB domain-containing proteins were shown to interact with Cullin 3. These
other proteins contain a BTB domain C-terminal to a MATH domain, or ankyrin repeats followed by a double BTB domain tandem (reviewed in Pintard et al., 2004) (Figure 1.9). This connection made sense, given that Skp1 and Elongin C adopt the BTB fold and interact with Cullin 1 and Cullin 2, respectively, while domains such as the β-propeller Kelch domain, ankyrin repeats and the MATH domain are all known to be protein-protein interaction domains. Notably, these findings showed the adaptor and substrate-recognition functionalities of the BTB-Cullin 3 complex are found in a single polypeptide. These studies brought up further questions that have yet to be solved. What region of the BTB domain interacts with Cullin 3 and how conserved is it across the many subfamilies of BTB domains? Do all BTB domains interact with Cullin 3, including those families with no previous connections to the ubiquitin ligase system, such as BTB-ZF proteins? As well, it was unclear if there was a region in these proteins that is analogous to the F-box/F-box recognition system involving the C-terminal helical extension to the BTB fold in Skp1, or the VHL α-domain. The structure of the SCF1 complex implied rigidity in the structure of the complex was necessary for activity, likely for the orientation between the E2-ubiquitin and the substrate. Are the Cullin 3-interacting proteins also structurally rigid, and how can this be reconciled with the domain architecture of these proteins?

Chapter 2 describes our identification of two novel domains in these proteins: the highly conserved BACK (BTB And C-terminal Kelch) domain that is found in nearly all proteins with BTB and Kelch motifs, and the PHR (Pam, Highwire, RPM-1) domain which could be another substrate recognition module in other BTB-BACK proteins.
FIGURE 1.9.

A) Domain architecture cartoons of examples of BTB domain-containing proteins that interact with Cullin 3 and are likely adaptor + substrate-targeting modules for ubiquitination by SCF3 complexes. Adapted from Furukawa et al., 2003.  B) Models of the SCF1, SCF2 and SCF3 ubiquitin ligase complexes. Skp1, Elongin C (EloC) and the BTB domain are the adaptor molecules for interaction with the Cullin scaffold. Adapted from Pintard et al., 2004.
1.4. BTB-ZF transcription factors

1.4.1. Zinc finger proteins

A significant number of DNA-binding proteins contain classical, C$_2$H$_2$/Krüppel-type zinc finger motifs. These zinc fingers are characterized by a 30 amino acid sequence containing two conserved cysteine and two conserved histidine residues. These residues coordinate a zinc atom, which stabilizes the small α/β fold. These motifs can be found in a single copy or in as many as 60 tandem repeats. Zinc fingers sequence-specifically recognize nucleotide bases by protruding into the major groove of B-DNA. Although there is no simple code for zinc finger-DNA recognition, artificial zinc fingers are often generated in the interest of creating tailor-made transcription factors (reviewed in Klug, 2005, Lee et al., 2003).

Similar to most transcription regulatory proteins, zinc finger proteins (ZFP's) are modular, and contain additional domains involved in transcription activation, repression, self-association, protein-protein interactions and sub-cellular localization. A search of the human genome using the SMART database revealed 778 ZFP's, with the largest fraction, 282 or 36%, also containing the KRAB domain (Krüppel associated box). The next-highest represented domains in ZFP's are found in smaller numbers and include the BTB-ZF families (42 or 5.4%), SCAN-ZF proteins (24 or 3%), and KRAB-SCAN-ZF proteins (20 or 2.6%) (reviewed in Collins et al., 2001). The domain architecture in many of these families is similar: the BTB, KRAB or SCAN is found at or near the N-terminus, followed by a region without any conserved domains and end with a series of C$_2$H$_2$-ZF motifs towards the C-terminus of the protein (Figure 1.10). It is possible that this conserved architecture of
There are many functional similarities between these N-terminal domains. The KRAB and BTB domains are both transcription repression modules and recruit co-repressor proteins to mediate repression (Moosmann et al., 1996, Friedman et al., 1996, Kim et al., 1996). The SCAN domain, however, is thought to mediate transcription activation, although there are examples of SCAN-ZF transcription repressors (Gu et al., 2007, Porsch-Ozcurumez et al., 2001). Expression of BTB-ZF, KRAB-ZF and SCAN-ZF proteins are developmentally regulated and may be important for cell fate determination, often in the hematopoietic system (reviewed in Edelstein & Collins, 2005, Urrutia, 2003). Both the SCAN and BTB domains mediate self-association; the SCAN domain has been better characterized in terms of its ability to mediate specific heterodimerization.
From a genomic perspective, each of the BTB-ZF, KRAB-ZF and SCAN-ZF gene families have undergone lineage-specific expansions (LSE's). For example, there is an expansion of BTB-ZF proteins in *Drosophila melanogaster* and these proteins are not well related in sequence with mammalian BTB-ZFs (Stogios *et al.*, 2005). KRAB-ZF proteins can be found in the frog and chicken genomes, but are found in much greater numbers in mammalian genomes (Shannon *et al.*, 2003, Huntley *et al.*, 2006, Looman *et al.*, 2002). SCAN-ZF proteins are found only in vertebrates and likely underwent rapid expansion in this lineage (Sander *et al.*, 2003).

### 1.4.2. Introduction to BTB-ZF proteins

The human genome encodes 42 BTB-ZF proteins, many of which are regulators of key genes involved in cell proliferation and differentiation (Stogios *et al.*, 2005). In typical BTB-ZF proteins, the BTB domain is found at the extreme N-terminus, followed by a central section predicted to have a high degree of structural disorder (this linker often contains stretches of low complexity sequence) and is terminated by a C-terminal region containing multiple tandem ZF repeats (*Figure 1.11*). In these proteins, the BTB domains have at least two functional roles: first, they drive the dimerization, and possibly oligomerization, of the protein, and second, they provide an interface for recruitment of components of chromatin remodeling complexes (Ahmad *et al.*, 2003, Ahmad *et al.*, 1998, Li *et al.*, 1999).
The 42 BTB-ZF proteins in the human genome, as discovered in this work. Domain architecture cartoons were taken from the SMART database, with the BTB domain shown as a purple triangle, C$_2$H$_2$ ZF motifs as blue rectangles, low complexity sequence as pink, coiled coil regions as green bars. Each protein is labeled by its BTB database identifier number (see section 1.2.4) and its name. Where a research group named the protein, that name is included, otherwise the Uniprot naming convention of ZBTBX is used. Names coloured in red and blue indicate structure of the BTB domain was solved during or previous to this work, respectively, in the Privé lab.
1.4.3. Transcription regulation by BTB-ZF proteins

Many BTB-ZF proteins are transcription repressors and this activity is usually due to recruitment of transcriptional co-repressors, such as N-CoR (nuclear co-repressor), SMRT (silencing mediator of retinoic acid and thyroid hormone receptor, also known as N-CoR2), BCOR, mSin3A, and histone deacetylases to promoter regions (David et al., 1998, Melnick et al., 2002, Ahmad et al., 2003, Yoon et al., 2003, Ghetu et al., 2008).

1.4.3.1. Background: histone modifications

The regulation of transcription initiation is the primary mode of gene expression control in eukaryotic cells. Transcription of eukaryotic mRNA is facilitated by RNA polymerase II (Pol II). Initiation of transcription by Pol II requires many other proteins, called general transcription factors, that serve to position the polymerase and unwind DNA at the start site. The functional, processive Pol II holoenzyme contains dozens of proteins and forms a large complex whose components are well conserved from yeast to man.

Higher order structure is necessary to package DNA into a cell’s nucleus. The basic repeating unit of chromatin is the nucleosome core particle, a 10-nm wide structure that contains 146 base pairs of DNA wrapped around a complex of two copies each of the histone proteins H2A, H2B, H3 and H4. The crystal structure of an intact nucleosome has been solved and revealed that the N-terminal tails of each histone molecule are unstructured and extend out from the nucleosome. Nucleosomes can condense into a 30-nm fiber or solenoid, with histone H1 bound to DNA in the interior of the fiber. Chromatin that is not actively being transcribed (heterochromatin) exists primarily as this condensed, 30-nm fiber,
while actively transcribed chromatin (euchromatin) assumes a “beads-on-a-string” form, comprised of separated nucleosomes.

Regulation of transcription initiation can be achieved by dynamic regulation of the higher-order structure of chromatin (chromatin remodeling) and therefore the accessibility of genes for transcription by the Pol II complex (Figure 1.12). It has long been known that histone tails can be covalently modified by acetylation, phosphorylation, methylation, ubiquitination, sumoylation and other modifications. These modifications can have a number of effects, including altering the strength of histone-histone and histone-DNA interactions and acting as sites of recruitment of proteins. These effects are often associated with alterations in the condensation state and transcription accessibility of chromatin. The reversible process of acetylation is one of the most well-studied histone modifications. Acetylation of the N-terminal tails of histones, especially H3 and H4, reduces nucleosome condensation by neutralizing the positive charge of lysines and induces recruitment of proteins that ultimately result in the activation of transcription (Struhl, 1998). The acetylation reaction is catalyzed by three main families of enzymes: GNAT, MYST and CBP/p300 (Sterner & Berger, 2000), collectively referred to as histone acetyltransferases (HAT’s). Conversely, deacetylation of histones results in chromosome condensation and transcription repression. Deacetylation is catalyzed by histone deacetylases (HDAC’s), of which there are three families: class I, class II and class III or sirtuins (Gallinari et al., 2007, Verdone et al., 2006).

Methylation of histone lysines is also an important modification affecting transcriptional status. Methylation of histones is thought to be involved in the maintenance of transcriptional status over longer periods of time, a process known as gene silencing. Methylation is also reversible due to the activity of demethylase enzymes.
There is evidence that different combinations and/or cross-talk between histone modifications leads to distinct functional outcomes, which has been termed the “histone code” (Jenuwein & Allis, 2001, Strahl & Allis, 2000). Consistent with this hypothesis, modification of histone tails provides docking sites for chromatin-interacting proteins, such as bromo or chromo domain-containing proteins, which specifically recognize acetylated and methylated lysines, respectively (Yang, 2004, Jones et al., 2000).

**FIGURE 1.12.**

Schematic of histone modifications. Histone deacetylation is catalyzed by HDAC’s, acetylation is catalyzed by HAT’s. Deacetylation is associated with chromatin condensation and transcription inactivation. Histones can also be modified by methylation, which is also associated with transcription inactivation. Adapted from Gallinari et al., 2007).

1.4.3.2. Background: mechanisms of transcription repression

Transcription repressors interfere with transcription initiation by a variety of mechanisms. Repressors can function by producing steric hindrance for access to DNA by inter-
fering with transcription activators or by recruiting other factors that induce a repressive chromatin environment (Cowell, 1994).

An example of transcription factors that affect transcription via steric hindrance are the Sp1 family of transcription factors, which bind to GC boxes but have strikingly different transcription regulatory activities, including both repression and activation (Li, He et al., 2004). Another example includes the Ets family transcription repressor Yan, which is regulated by another Ets family member, Mae (Qiao et al., 2004). Yan polymerizes on DNA via self-association of its SAM domain, and the SAM domain of Mae can disturb Yan oligomerization by interacting with the polymerization surface.

Transcription factors can directly interact with histone modifying enzymes such as HDAC’s or histone methyltransferases, or indirectly, by recruiting co-repressor proteins which themselves provide interactions with the histone-targeting enzymes. The co-repressors N-CoR, SMRT, mSin3A, ETO (Eight twenty one protein), MTA3 (Metastasis-associated protein 3) and CtBP (C-terminal binding protein) are recruited by many diverse transcription repressors and serve as scaffolds for the formation of large repression complexes. The N-CoR and SMRT co-repressors are large proteins (2440 and 2517 residues, respectively) that are predicted to be mostly disordered. To list a few examples, the N-CoR co-repressor is involved in the function of such diverse transcription factors as steroid hormone receptors (Wu et al., 2006) and the POU family (Remenyi et al., 2002); the mSin3a co-repressor is involved in the function of the Myc-Mad heterodimer, p53 (Ho et al., 2005) and the Ets-family member PU.1 (Suzuki et al., 2003).

Often transcription factors function by alterations in their subunit compositions. An example of this phenomenon occurs within the Myc transcription factor network. The Myc-
Max heterodimer interacts with E-box sequences and recruits the SWI/SNF nucleosome remodeling complex and histone acetylases (Luscher, 2001). The Mad-Max heterodimer interacts with the same E-box sequences, but instead recruits mSin3 and histone deactetylases.

1.4.3.3. BTB domains from BTB-ZF proteins recruit co-repressors and histone deacetylases

Certain BTB-ZF proteins are known to selectively interact with co-repressors, via contacts with the BTB domain or within areas of the linker and the ZF motifs. Interactions with co-repressors serves to recruit HDAC-containing complexes into high molecular weight (HMW) complexes. Many BTB-ZF proteins have been observed to form a punctate pattern in the nucleus when visualized by immuno-fluorescence, suggesting HMW-complex formation at specific loci recognized by the ZF motifs.

BTB-ZF proteins vary in transcription repression strength. A few examples, such as Miz-1 (Peukert et al., 1997, Herold et al., 2002, Phan et al., 2005, Patel & McMahon, 2007) and ZFP161 (Lee et al., 2004, Wang et al., 2005), act as both transcription repressors and activators by virtue of their interactions with other transcription factors. In terms of transcription repression, the transcription repression strength of BTB-ZF proteins is thought to vary according to their affinities for co-repressors and HDAC-containing complexes. The BCL6 BTB domain is a strong repressor and strongly interacts with SMRT (Dhordain et al., 1998, Huynh & Bardwell, 1998, Ahmad et al., 2003). The FAZF BTB domain is a weak repressor and likely does not interact with SMRT (Melnick et al., 2002). Kaiso\textsuperscript{BTB} selectively interacts with N-CoR but not SMRT (Yoon et al., 2003), while the Hic1 BTB domain does not interact with either (Deltour et al., 1999).
1.4.4. Dimerization of BTB domains

BTB domains have long been known to mediate dimerization between BTB-ZF proteins (Figure 1.13). A dimeric BTB-ZF protein would have two sets of DNA-binding ZF’s and could therefore bind to repeated recognition sites. From a structural perspective, the mechanism of homodimerization became clear when the structures of the BTB domains from PLZF and BCL6 were determined. Homodimerization of the BCL6 BTB domain was shown to be necessary for SMRT recruitment, as the peptides bind across the BCL6BTB dimerization interface (Figure 1.5, Ahmad et al., 2003). Since each BTB domain purified by our lab and others to date is homodimeric (data in this thesis and Li et al., 1999, Ahmad et al., 2003, Ahmad et al., 1998, Stogios et al., 2007, Schubot et al., 2006), we presume that all BTB domains from the BTB-ZF class, and likely the BBK class, are obligate homodimers.

A survey of the literature surrounding BTB-ZF proteins suggests BTB domains can mediate heterodimerization. For reasons attributable to the techniques and protein constructs utilized by other groups studying BTB domain dimerization, our understanding of heterodimerization is not clear. Do the hetero-interactions occur directly between BTB-ZF proteins? What regions of the proteins mediate heterodimerization? What is the basis for specificity? To illustrate the ambiguity, the pairs BCL6 and Miz-1 (Phan et al., 2005), and BCL6 and BAZF (Okabe et al., 1998) have been shown to interact, but it is not clear if this occurs between the BTB domains. Another example is the interaction between FAZF and PLZF, which were shown to interact by a yeast two-hybrid assay. Due to the nature of the assay, the interaction between the proteins may not be direct but could be bridged by other
factors; as well, the interaction could occur outside of the BTB domain since this study utilized the full-length proteins.

The functional consequence of heterodimerization is not clear either. In principle, dimerization of BTB-ZF proteins would dramatically increase the range of DNA sequences these proteins could recognize, as different combinations of ZF motifs could be combined to form new transcription factors. Due to the range of affinity for co-repressors and repression strength by BTB domains, heterodimerization of the BTB domain could be utilized for regulation of recruitment of chromatin remodeling machinery. Potentially, a BTB domain that has low affinity for co-repressor binding could act as a modulator for another BTB domain that has high affinity. Alternatively, transcription repression by a BTB domain with low affinity for co-repressors could be “rescued” by interaction with a different BTB domain that has higher affinity for co-repressors.

The recent discovery that the repertoire of target genes recognized by BCL6 can be altered by interaction with Miz-1 was the first demonstration of a functional consequence of BTB-ZF heterodimerization (Phan et al., 2005). The BCL6-Miz-1 complex interacts with genes lacking the BCL6 binding site, but instead bound to genes normally recognized by Miz-1. This complex represses these genes, including the cell cycle arrest gene CDKN1A. It is unclear if heterodimerization is a general phenomenon with BTB-ZF proteins, although this is a common theme in other families of transcription factors, including the steroid hormone receptors, leucine zipper transcription factors, and the STAT (signal transducers of and activators of transcription) proteins.

The Privé lab has some biochemical data from the work of a previous graduate student (Eden Fussner) about heterodimerization of BTB domains. Eden’s work showed that
one BTB domain, from FAZF, can spontaneously form heterodimers. Two other pairs, BCL6-LRF and BCL6-Miz-1, can form heterodimers but only after a mix of these BTB domains are unfolded and then refolded. However, it is unclear how these results can be reconciled with the fact that all BTB domains studied to date in our lab exist as strong homodimers.

1.4.5. Oligomerization of BTB domains

The BTB domain is thought to mediate the higher-order oligomerization of proteins. As in the study of heterodimerization of BTB domains, oligomerization has not been extensively characterized with purified BTB domains and much of what is known about this phenomenon comes from cellular studies. A number of studies have shown that BTB-ZF proteins form high molecular weight (HMW) complexes (Ball et al., 1999, Espinas et al., 1999, Igarashi et al., 1998) and stain with a punctate pattern in the nucleus (Melnick et al., 2000, Dhordain et al., 1995). These studies suggest many copies of the protein assemble at a specific genetic locus. These activities are dependent on the BTB domain, as deletions and specific mutations have been shown to affect HMW complex formation and nuclear localization behaviour.
A) BTB-ZF homodimer

B) BTB-ZF heterodimer

**FIGURE 1.13.**

Schematic of transcription repression by A) homodimeric, and B) heterodimeric BTB-ZF proteins. The BTB domain and ZF motifs are labeled, with the linker between these domains drawn with a dotted line. Selected BTB-ZF proteins recruit the co-repressors SMRT, N-CoR or B-CoR, mSin3, along with histone deacetylases (HDAC) (the precise connectivity and stoichiometry of this complex is unknown). This complex removes acetyl groups (labeled with an “Ac”) from histone tails. A heterodimeric BTB-ZF protein would have potentially different DNA-binding specificity, as indicated by the red colouring.
The BTB domains from GAGA factor and Hic1 have been shown to inhibit binding to DNA containing a single copy of the target binding site, but promote cooperative binding to DNA containing multiple copies (Katsani et al., 1999, Pinte et al., 2004). These data suggest the functional unit of a BTB-ZF protein is an oligomer, formed via interactions between BTB domains.

The functional consequence of oligomerization of BTB domain-containing transcription factors on DNA is thought to be architectural changes in chromatin structure. This is based on the observations that GAGA factor and a complex including the BTB-bZip transcription factor Bach1 can both induce bending or looping of DNA (Katsani et al., 1999, Yoshida et al., 1999). As well, it is possible that a BTB-ZF oligomer provides a novel, modified, or multiple copies of a binding site for nuclear co-repressors and/or HDAC’s (Figure 1.14). It is conceivable that a complex with many copies of the BTB-ZF protein would recruit more co-repressors and HDAC’s and therefore provides a stronger transcription repression effect on the surrounding chromatin environment. However, this has not been experimentally demonstrated and it possible that a dimeric BTB-ZF protein bound to DNA is sufficient for repression function.

Oligomerization has been observed only for specific BTB domain-containing proteins, most especially the BTB-ZF proteins PLZF, Hic1, GAGA and the BTB-bZip protein Bach1. As well, there has been no clear demonstration of an oligomeric BTB domain when isolated from the context of the full-length protein, such as a size exclusion chromatography experiment showing higher-order species.
It is possible that oligomerization only occurs in the context of the full-length protein bound to DNA. Also, it is possible that each BTB domain has its own oligomerization propensity and this is not a general property for the domain.
Nonetheless, the two crystal structures of the BTB domain from PLZF provided tantalizing clues as a possible structural basis for higher-order association of BTB-ZF proteins (Ahmad et al., 1998, Li et al., 1999). Each crystal was of a different space group, but in both structures the crystal lattice was built by association of BTB domain dimers via interactions between the N-terminal \( \beta_1 \) secondary structure elements (Figure 1.15). The two structures are very similar and superpose with a RMSD of 1.1 Å over all atoms.

The Li et al study thought the dimer-dimer associations explained the biological data. No evidence was provided for this oligomerization in solution, however. It remains to be seen if oligomerization is only a phenomenon observed in crystals and if the \( \beta_1-\beta_1 \) packing is a true oligomerization interface.

1.4.6. Roles of BTB-ZF proteins in cancer

Consistent with their important role in transcription regulation, many BTB-ZF proteins, including BCL6, PLZF, LRF, Kaiso, FAZF and Miz-1 are implicated in oncogenesis. Expression of genes coding for BTB-ZF proteins have been identified to be deregulated in specific cancers. The biological effects of deregulation of the expression of BTB-ZF proteins and whether they are considered candidate proto-oncogenes or tumour suppressors is dependent on the functions of their particular target genes. BTB-ZF proteins regulate a variety of genes that themselves play important roles in proliferation, differentiation or apoptosis, therefore, deregulation of genes coding for BTB-ZF proteins
FIGURE 1.15.

*Dimer-dimer interactions in either of the two PLZF\textsuperscript{BTB} structures (PDB codes 1BUO, 1CS3). Cartoon of association between dimers, with each dimer shown as an “inverted butterfly”, with the β1-β5 sheet in each dimer shown. Figure adapted from (Li et al., 1999).*

often have serious biological consequences. The roles of each of the BCL6, PLZF, LRF, Kaiso, FAZF and Miz-1 proteins in cancer is summarized below.
BCL6: The most prominent example of deregulation of a BTB-ZF protein involved in cancer is the Bcl6 gene, which is strongly implicated in the transformed phenotype in diffuse large B-cell lymphoma (DLBCL), a common and aggressive form of Non-Hodgkin’s Lymphoma (NHL) (Albagli-Curiel, 2003). 40% of DLBCL cases are characterized by genetic rearrangements upstream of the Bcl6 gene at 3q27 (Ohno, 2006; Jardin & Sahota, 2005). The 5’ region of the first intron of the Bcl6 gene is involved in regulation of its expression; chromosomal translocations replace this region with heterologous promoters. As well, somatic mutations (SM) within the 5’ untranslated region (UTR) also affect the expression of the gene. BCL6 represses the expression of proteins that are required for the terminal differentiation of germinal center B-cells, including but not limited to Blimp1 (B-lymphocyte-induced maturation protein 1) (Schebesta et al., 2002). Inappropriate expression of this protein as in DLBCL may trap these cells in a proliferative state (Calame et al., 2003; Shaffer et al., 2000).

**PLZF**: In a rare form of acute promyelocytic leukemia (APL), the Plzf gene undergoes a chromosomal translocation that codes for a fusion protein between PLZF and retinoic acid receptor α (RAR-α) (Chen et al., 1993). This creates a dominant negative inhibitor of wild-type RAR-α and represses genes that normally respond to retinoic acid signaling. This affects normal DNA repair, apoptosis and cell cycle regulation (Alcalay et al., 2003; Muller et al., 2000). All-trans retinoic acid (ATRA) exists as a therapy to rescue the transformation phenotype for tumours expressing a similar chromosomal fusion protein, PML-RAR (Slack, 1999). ATRA dissociates RAR-α from binding to target genes. However, ATRA cannot disassociate PLZF-RAR-α fusion proteins from retinoid-responsive genes.
• **LRF**: The *Zbtb7* gene, which codes for the BTB-ZF protein LRF (also known as FBI-1, OCZF or Pokemon), was shown to be a critical regulator of the gene *p19Arf*, an important tumour suppressor that regulates the levels of p53 by sequestering Mdm2 in the nucleus (reviewed in Lozano & Zambetti, 2005, Maeda, Hobbs & Pandolfi, 2005). Forced overexpression of *Zbtb7* leads to oncogenic transformation in mice and the *Zbtb7* gene was indeed over-expressed in DLBCL and FL samples (Maeda, Hobbs, Merghoub et al., 2005).

• **Kaiso**: Kaiso’s role in oncogenesis has not been fully established, but is indicated by a number of factors (reviewed in van Roy & McCrea, 2005). Kaiso binds to methylated CpG dinucleotides, which are commonly observed at transcriptionally quiescent or silenced genes, a phenomenon commonly observed in cancer (Daniel et al., 2002, Yoon et al., 2003, Prokhortchouk et al., 2001, Aranyi et al., 2005). Kaiso represses expression of genes regulated by canonical Wnt-signalling pathways, such as *c-Fos*, *c-Jun*, *Ccnd1* (which encodes cyclin D1), and *Mmp7*. Kaiso is thought to be involved in a transcriptional response to extracellular cues as Kaiso-mediated repression is relieved by interaction with the armadillo family member p120 catenin (Daniel & Reynolds, 1999, Daniel et al., 2002, Kelly, Spring et al., 2004, Kelly, Otchere et al., 2004, Kim et al., 2004, Spring et al., 2005, Park et al., 2005). As well, Kaiso expression and cellular localization show dependence on growth conditions, tissue, and tumour microenvironment (Daniel & Reynolds, 1999, Daniel et al., 2001, Soubry et al., 2005).

• **FAZF**: This protein, also known as PLZP, has been shown to interact with FANCC, the gene that is defective in bone marrow failure that occurs in Fanconi Anemia complementation group C patients (Hoatlin et al., 1999). FA patients have genomic instability, are sensitive to DNA-damaging agents, increased apoptosis, defective p53 induction and stem cell
defects, leading to an increased risk of developing leukemia (reviewed in Grompe & D'Andrea, 2001). FAZF is expressed in early hematopoietic progenitor cells (Dai et al., 2002), and FAZF-deficient mice show defects in cell cycle control and cytokine production in T-cells (Piazza et al., 2004).

- **Miz-1**: Miz-1 has strong connections with cancer through its ability to act as a transcription activator of the *p15INK4B* gene, which encodes a cyclin-dependent kinase (CDK) inhibitor, leading to cell cycle arrest (Staller et al., 2001, Herold et al., 2002, Seoane et al., 2001). However, Miz-1 is more widely known through its ability to potentiate Myc-dependent transcription repression. Miz-1 interacts with Myc, a potent oncogene, through interactions between a region around the ZF motifs in Miz-1 and the HLH region of Myc. This complex inhibits Miz-1 activation of *p15INK4B*, relieves cell cycle arrest and allows cell proliferation. Furthermore, DNA methylation appears to play a role in repression by the Myc-Miz-1 complex, as a group found evidence of a Myc-Miz-1-Dnmt3a (Dnmt3a is a DNA methyltransferase) ternary complex (Brenner et al., 2005). This complex represses another gene coding for a CDK inhibitor, *p21CIP1*. This repression blocks cell differentiation and also allows progression through the cell cycle (Wu, Cetinkaya et al., 2003). An alternative mechanism of repression of *p21CIP1* was also observed and involves formation of a complex between the BTB-ZF proteins BCL6 and Miz-1 (Phan et al., 2005). In the absence of BCL6, Miz-1 activates the *p21CIP1* gene; upon its expression, BCL6 is able to interact with Miz-1 at this gene and represses its expression. The BCL6-Miz-1 interaction was dependent on the presence of the BCL6 BTB domain, hinting at the presence of a heterodimeric BTB-ZF transcription factor.
1.5. Thesis overview

At the time of the onset of this work, it was known that the BTB domain was widely distributed in eukaryotic genomes and was found in a variety of protein types with many functions. It was known that there was low sequence similarity between BTB domains in these various subfamilies and not much higher sequence identity within the mammalian BTB-ZF subfamily. From a structural perspective, there was little known: only two nearly identical structures of BTB domains in the BTB-ZF subfamily of proteins were solved. Additionally, a number of structures of the T1 domain and Skp1 proteins which contain the BTB fold were known. In terms of co-repressor interactions, the BCL6 BTB domain was known to interact with the SMRT-BBD, but whether other BTB domains interacted with this peptide was not known. Based on these observations, the main questions we sought to address during the course of my graduate studies were:

- What is the complete repertoire of BTB domains in fully sequenced eukaryotic genomes?
- What types of proteins contain the BTB domain? Can these subfamilies be clustered or grouped in terms of their sequence evolution?
- Are there any sequence signatures in the various subfamilies of the BTB domain that can be connected with the known self-association and/or interaction states? Do these signatures provide any predictive value?
- How similar are the structures of the BTB domain in the BTB-ZF subfamily?
- What are the self-association states of BTB domains from the BTB-ZF subfamily?
- Do other BTB domains interact with the SMRT-BBD peptide?
This work is divided by two approaches. The first is computational and is detailed in Chapter 2. This work comprised a bioinformatics analysis into identifying the genomic distribution, kinds of protein architectures, sequence characteristics and evolutionary relationships of BTB domain-containing proteins. This work was published in *Genome Biology* in 2005. This work began by identifying and aligning 18 available structures of the BTB fold in the PDB. These structures represent the three subfamilies corresponding to the BTB domain from BTB-ZF proteins, the T1 domain, and the Skp1 protein, representing very few of the many identified sequence subfamilies of BTB domains. These structures showed the BTB domain is comprised of a highly conserved “core BTB fold”, plus additional secondary structure elements that are added in specific subfamilies.

We compared the self-association states of the solved BTB domain structures. There were family-specific sequence signatures that often take part in specific protein-protein interactions. We found very little overlap in the residues mediating dimerization in BTB-ZF proteins, tetramerization in the T1 domain, and interactions with Cullin1 or Cullin2 in the Skp1 and Elongin C subfamilies, respectively. Therefore, the BTB domain appears to have evolved different interaction surfaces and acquired additional secondary structure elements, while maintaining a highly conserved core structure.

We found that the BTB domain is typically found in a single copy in proteins that contain only one or two other types of domain, and this served to identify a set of subfamilies. We found numerous lineage-specific expansions of specific classes of BTB proteins, as seen in the relatively large number of BTB-ZF and BBK proteins in vertebrates (85 to 183), MATH-BTB proteins in *C. elegans* (46), and BTB-NPH3 proteins in *A. thaliana* (21). Strikingly, the types and numbers of proteins containing BTB domains in vertebrates from fish to
mammals is very similar, probably indicating that by the time of the evolution of metazoa, the specialization of BTB domain-containing proteins was complete.

Efforts to cluster BTB domain sequences using phylogenetic methods failed due to the low sequence identity across subfamilies. The best relationship of the sequence subfamilies of BTB domains was constructed using pairwise BLAST score clustering. This clustering showed the sequence of the BTB domain varies according to speciation events and the domain architectures. BTB domain-containing proteins should be thought of in terms of their subfamilies: for example, BTB domains from *Drosophila* BTB-ZF proteins have different sequence signatures than BTB domains from mammalian BTB-ZF proteins and therefore may perform different functional roles. This indicated selective pressure on the sequence of the BTB domain could have been due to the functional specialization of these proteins in a species and architecture-specific manner.

During the course of this analysis, a conserved region immediately C-terminal to the BTB domain was identified in a subset of proteins, and we called this region the BACK (BTB and C-terminal Kelch) domain. This published work from *Trends in Biochemical Sciences* is presented in Chapter 2.

Due to the differences in the residues represented in the dimerization and Cullininteraction surfaces we identified by comparisons of the structures of the BTB domains from BTB-ZF proteins and Skp1, we generated a model of the structure of a BBK-Cul3 ubiquitin ligase complex. This model is dimeric via the dimerization interface present in the BTB domains from BTB-ZF proteins, which we predict exists in BBK proteins.

Concurrently with the above computational studies, members of the lab including Denise Jaworsky, Lu Chen and Neil Pomroy, were involved in cloning, expressing and
crystallizing a large set of BTB domains from human BTB-ZF proteins. I undertook the structure determination of three of these BTB domains: FAZF, Kaiso and LRF, as well as the crystallization and structure determination of the Miz-1 BTB domain. The LRF BTB domain structure was published in *Protein Science* in 2007. This work is described in Chapter 3. The crystal structure of LRF showed a canonical domain-swapped BTB domain dimer without any obvious “interesting features” in terms of crystal packing or altered domain swapping state. However, we did notice that the LRF BTB structure showed extensive residue substitutions in the lateral groove region relative to the BCL6 BTB domain. Therefore, we predicted this would have an effect on co-repressor interactions and that the LRF BTB domain would not interact with the SMRT peptide and I verified this experimentally. This structure further highlighted that the core BTB fold provides a conserved structural scaffold and that the surface-exposed residues of the BTB domain have evolved to fulfill different protein-protein interaction states.

FAZF was the first BTB domain structure I determined and it showed an unexpected novel conformation of the N-terminal extension to the core BTB fold that resulted in a non-domain swapped dimer. We immediately thought this conformation would allow for separation of the homodimer into monomers, and we spent a great deal of time verifying the domain swapping state and looking for monomeric BTB domain in solution. Much later in my Ph.D. studies, I successfully expressed, purified and crystallized the BTB domain from Miz-1. This BTB domain was crystallized as it naturally contains a shorter N-terminus. Therefore, we thought that packaging the structure of the FAZF and Miz-1 BTB domains into one manuscript would make sense and would be an excellent opportunity to study the dimerization of the BTB domain. Chapter 4 illustrates the crystallization, structure deter-
mination and solution studies of the FAZF and Miz-1 BTB domains from the human BTB-ZF proteins FAZF and Miz-1.

Chapter 5 illustrates the crystallization and structure determinations of the BTB domain from Kaiso. The Kaiso BTB domain is a canonical domain-swapped dimer, but unexpectedly showed that homodimers interact via extension of the two-stranded β1-β5 sheet with an equivalent sheet from another dimer. We were unable to detect evidence of oligomerization of the Kaiso BTB domain in solution, but we did observe that the interdimer interaction across β1-β5 sheets is strikingly similar to that seen in other BTB domain structures. This includes the two crystal forms of the PLZF BTB domain, the BTB domain from the BTB-bZip protein Bach1, and the BBK protein Gigaxonin. These similarities suggest that β1 provides a common interaction site for contacts between BTB domain dimers, either in crystal structure, or biochemical/biological context. The Kaiso BTB oligomeric structure could reflect the tendency for BTB domains from human BTB-ZF proteins to mediate high-molecular weight complexes on DNA, but this awaits solution verification.

While this thesis provided a number of insights into the sequence, structure and function of the BTB domain, much remains to be learned. Chapter 6 presents ideas for further studies that build upon the results of this work.
Chapter 2

Sequence-Structure Relationships of BTB Domain-containing Proteins

This chapter has been reformatted with permission from the original publications:

Stogios PJ, Downs GS, Jauhal JS, Nandra SK, Privé GG. “Sequence and Structural Analysis of BTB Proteins.”, *Genome Biology* 6:R82 (2005);

Copyright © Peter John Stogios


Reprinting permission granted from Elsevier Science (USA).
2.1. Abstract

The BTB/POZ domain is a versatile protein-protein interaction motif that participates in a wide range of cellular functions. Several BTB domain structures have been solved by x-ray crystallography. In this chapter, we surveyed the protein architecture, genomic distribution and sequence conservation of BTB domain proteins. We took advantage of the strong structural conservation of the BTB fold across proteins with very low sequence identities to train a family of Hidden Markov Models (HMM’s). These HMM’s were utilized to search seventeen fully sequenced eukaryotic genomes in a highly specific and sensitive manner, with BTB domain hits stored in a self-updating and web-accessible database. We observed that the BTB domain is typically found as a single copy in proteins that contain only one or two other types of domain, and this defines the BTB-ZF, BTB-BACK-kelch (BBK), voltage-gated potassium channel T1 (T1-Kv), MATH-BTB, BTB-NPH3 and BTB-BACK-PHR (BBP) families of proteins, among others. We identified two novel domains, the BACK and PHR domains, that likely play important roles in the function of BTB-Cul3 SCF E3 ubiquitin ligase (SCF3) complexes. There are numerous lineage-specific expansions of BTB proteins, as seen by the relatively large number of BTB-ZF and BBK proteins in vertebrates, MATH-BTB proteins in C. elegans, and BTB-NPH3 proteins in A. thaliana. Using the structural similarity between Skp1 and the PLZF$^{\text{BTB}}$ homodimer, we present a model of a SCF3 complex that shows that the BTB dimer or the T1 tetramer is compatible in this complex.
2.2. Introduction


A number of BTB structures have been determined by x-ray crystallography, establishing the structural similarity between different examples of the fold. We use the SCOP terminology of fold to describe the set of BTB sequences that are known or predicted to share a secondary structure arrangement and topology, and the term family to describe more highly related sequences that are likely to be functionally similar (Murzin *et al.*, 1995). Thus, the BTB domain in BTB-ZF, Skp1, ElonginC and T1-Kv proteins all contain the BTB
fold, even though some of these differ in their peripheral secondary structure elements and are involved in different types of protein-protein associations. For example, BTB domains from the BTB-ZF family contain a N-terminal extension and form homodimers (Ahmad et al., 2003, Ahmad et al., 1998), while the Skp1 proteins contain a family-specific C-terminal extension and occur as single copies in heterotrimeric SCF complexes (Schulman et al., 2000, Zheng et al., 2002, Wu, Xu et al., 2003, Orlicky et al., 2003). The ElonginC proteins are also involved in protein degradation pathways, however these proteins consist only of the core BTB fold and are typically less than 20% identical to the Skp1 proteins (Botuyan et al., 2001, Stebbins et al., 1999). Finally, T1 domains in T1-Kv proteins consist only of the core fold and associate into homotetramers (Kreusch et al., 1998, Nanao et al., 2003). Thus, while the structures of BTB domains show good conservation in overall tertiary structure, there is little sequence similarity between members of different families. As a result, the BTB fold is a versatile scaffold that participates in variety of types of family-specific protein-protein interactions.

Given the range of functions, structures and interactions mediated by BTB domains, we undertook a survey of the abundance, protein architecture, conservation and structure of this fold. To undertake our analysis, we constructed a database of BTB domain sequences that was based on the generation of Hidden Markov Models trained on structural superpositions of BTB fold representatives. An early study is consistent with many of the results presented here (Aravind & Koonin, 1999), and we contribute an expanded structure and genome-centric analysis of BTB domain proteins, with an emphasis on the scope of protein-protein interactions in these proteins. As well, we identify two novel domains in proteins that are known to interact with Cul3. Our results should be useful for the structural and
functional prediction by analogy for some of the less-well characterized BTB domain families.
2.3. Methods

2.3.1. Strategy for the identification of BTB domains

Combining the strengths of the Ensembl database, structural alignment, multiple HMM’s and iterative searches, we devised an effective pipeline for the detection of BTB domains with high sensitivity and specificity. Our approach is summarized in Figure 2.1, and can be divided into three main sections: 1) structure comparisons, 2) BTB sequence identification, 3) database update. The organization of these steps is briefly introduced here described in detail throughout Chapter 2.3.

This work was assisted by a number of talented summer and volunteer students, including Sukhjeen Nandra, Jimmy Jauhal and Gregory Downs. Initial searches for BTB domains, using manually downloaded genomes and construction of a MySQL database was performed by Sukhjeen. Jimmy improved on her work by scripting the entire process in Perl: retrieval of sequences from Ensembl, searches using a family of HMM’s, storage in a non-redundant Oracle database, a web interface and some search statistics. Gregory expanded on Jimmy’s script by introducing a number of user interface tools. He created an administration panel that allowed users to easily upload new HMM’s or initiate new searches, and he automated the database update process to regularly check the Ensembl database for new genome builds or annotation updates.

Using superpositions based on solved BTB domain structures, a family of HMM’s was generated describing the entire collection of structures or subsets thereof. These HMM’s were used to search the genomes downloaded from Ensembl, Uniprot and Dictybase.
FIGURE 2.1.
Flow chart of BTB domain identification pipeline. Each step from the three main sections: structure comparisons, BTB sequence identification, database update, are described in section 2.3.1.
To maintain the non-redundant nature of the BTB Domain Database, full-length protein sequences with new candidate BTB domains were compared with the current contents of the Database, using BLAST. Any identical hits resulted in the disposal of the new candidate BTB domain, as it already existed in our database. Other hits proved to be the most problematic, as inexact hits could correspond to a number of different scenarios, including a completely new BTB domain, a BTB domain already in our database whose sequence, intron-exon structure, or identifier codes have been altered at Ensembl and must be updated in our database, or a BTB domain in our database that has become an orphan due to its deletion at Ensembl. These scenarios had to be independently addressed to ensure our database best mirrored the current state of Ensembl. Each unique BTB domain was given its own BTB identifier (“BTB id”), a one to four digit number. A number of annotations and descriptors were added or retrieved from Ensembl, including the position of the BTB domain, the HMM used to identify it and the e-value score, the position of other BTB domains in the full-length protein, Ensembl gene, transcript and peptide identifiers, and more. The database is publicly accessible at http://btb.uhnres.utoronto.ca, and analysis of its contents formed the basis of the analysis published in our *Genome Biology* (2005) study and is described in Chapter 2.4.

2.3.2. Structural superpositions of BTB fold-containing proteins.

The Protein Data Bank (PDB) was rigorously searched with a variety of solved BTB domain structures as queries, using the DALI (http://www.ebi.ac.uk/dali), CE (Shindyalov & Bourne, 1998), and VAST (http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml) structure superposition algorithms. Results from each search were combined after manual
inspection for high Z-score/bit score values and length of alignment across at least 90 amino acids. Corresponding amino acid sequences were collected and structural alignments were generated using CE, SwissPDBViewer (http://www.expasy.ch/spdbv) and the Superpose server (Maiti et al., 2004). Alignments were cross-referenced between the superposition algorithms and visually inspected. Final RMSD values were calculated from SwissPDBViewer.

2.3.3. Hidden Markov Models (HMM’s).

Structural alignments were used as seed alignments for the generation of a panel of HMM’s using the HMMER 2.3.2 package (http://selab.janelia.org). The hmmbuild and hmmcalibrate commands were run with default parameters. Each HMM was trained using at least 200 sequences, if this number was available from the non-uniform distribution of sequences of BTB domain subfamilies. Some HMM’s were generated and calibrated iteratively, using an initial seed structural alignment for the generation of the first HMM, followed by a search of genome databases using this HMM, and another cycle of HMM building and calibrating. Full parameters for genome searches are described in section 2.3.4. Each HMM corresponded to a different level of organism, sequence and/or structural similarity of the BTB domain. The final collection of HMM’s represented BTB domains from the following proteins (also indicated is how they compare to HMM’s at Pfam):
Table 2.1. Collection of BTB domain HMM’s used for genome searches

<table>
<thead>
<tr>
<th>Protein subfamily</th>
<th>Sequence region of BTB domain used in training HMM*</th>
<th>Pfam HMM</th>
<th>Sequence region of BTB domain used in training HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>All BTB domain-containing proteins</td>
<td>(\beta_1) through A5</td>
<td>PF00651</td>
<td>C-terminal half of (\alpha_1) through A5</td>
</tr>
<tr>
<td>Mammalian BTB-ZF proteins – with N-terminal extension</td>
<td>(\beta_1) through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td>Mammalian BTB-ZF proteins – without N-terminal extension</td>
<td>(\alpha_1) through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em> BTB-ZF proteins</td>
<td>(\beta_1) through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td>Mammalian BTB-BACK-Kelch proteins</td>
<td>(\beta_1) through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td><em>C. elegans</em> MATH-BTB proteins</td>
<td>(\beta_1) through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td>T1 domain from Kv potassium channels</td>
<td>(B_1) through A5</td>
<td>PF02214</td>
<td>B1 through N-terminal half of A5</td>
</tr>
<tr>
<td>Skp1</td>
<td>(B_1) through (\alpha_8)</td>
<td>PF03931</td>
<td>PF03931: (B_1) through A3;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01466</td>
<td>PF01466: (\alpha_4) through A5.</td>
</tr>
<tr>
<td>Elongin C</td>
<td>(B_1) through A4</td>
<td>PF03931</td>
<td>(B_1) through A3</td>
</tr>
<tr>
<td>RCC1-BTB proteins</td>
<td>(\alpha_1) through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td>BTB-NPH3 proteins</td>
<td>A1 through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td>BTB-TPR proteins</td>
<td>A1 through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
<tr>
<td>BTB-WD40 proteins</td>
<td>A1 through A5</td>
<td>Not in Pfam</td>
<td>-</td>
</tr>
</tbody>
</table>

*See section 2.4.1 for labels of secondary structures.

2.3.4. Sequence alignment.

Due to the high level of sequence divergence across the BTB domain subfamilies listed above, sequence alignments were generated using the mature structure-based HMM’s with the hmmalign command, using default parameters, from the HMMER 2.3.2 package. For subfamilies with a relatively higher level of sequence similarity (average 30% identity, such as the BTB domains from mammalian BTB-ZF proteins), ClustalW (http://bips.u-
strsbg.fr/fr/Documentation/ClustalX) with default parameters was sufficient for generation of sequence alignment. In each case, sequence alignments were manually inspected and adjusted for the positioning of gap sequences, especially in the β5 region of the alignment.

2.3.5. Genome collection and sequence searches.

All peptides from the translations of all known and predicted transcripts in the genomes of *Anopheles gambiae,Apis mellifera, Caenorhabditis elegans, Canis familiaris, Danio rerio, Drosophila melanogaster, Gallus gallus, Homo sapiens, Mus musculus, Pan troglodytes, Rattus norvegicus, Takifugu rubripes and Xenopus tropicalus* were retrieved from the latest version of Ensembl (http://www.ensembl.org). *Arabidopsis thaliana, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* protein sequences were retrieved from Uniprot (http://www.pir.uniprot.org). *Dictyostelium discoideum* protein sequences were retrieved from Dictybase (“primary features”) (http://dictybase.org). Proteins containing BTB domains were identified using hmmsearch from the HMMER package, with e-value cutoff of 10, using each HMM from our panel. BTB domains scoring in the e-value range 0.1 – 10 were manually inspected. Searches were repeated until no new hits were retrieved (convergence). Peptide sequences, identifiers, names and aliases, domain boundaries of the non-BTB domains (from Pfam annotations included in the Ensembl peptide features) were retrieved for storage in the BTB database.
2.3.6. Sequence clustering and most probable sequence detection.

Family-specific HMM’s were utilized to generate multiple sequence alignments, which were then merged into larger alignments for clustering. Phylogenetic clustering was attempted with the distance, maximum parsimony and maximum likelihood algorithms in the PAUP*4.0 (http://paup.csit.fsu.edu), MEGA 2.0 (http://www.megasoftware.net), Clustal and PHYLIP 3.63 (http://evolution.genetics.washington.edu/phylip.html) software packages. BLASTCLUST was utilized for clustering based on pairwise BLAST e-value scores. The most probable/”signature” sequences were retrieved using the hmmemit program from the HMMER package. The source code for hmmemit was modified to emit consensus sequences with a probability of 0.4, 0.6 and 0.8 from HMM’s.

2.3.7. New domain identification.

2.3.7.1. The BACK domain.

In proteins containing a BTB domain and a series of Kelch motifs, the strongly conserved sequence features (approximately 130 residues) in the region between these regions were not previously thought to contain a structured domain. Searches of domain databases with various examples of sequences in this region (i.e. residues 48-180 of human Keap1, Uniprot id KEAP_HUMAN) revealed only PFAM-B (Bateman et al., 2004) conserved motifs (automatically generated alignments), no conserved domains at SMART (Letunic et al., 2004) and only KOG4441 (”Proteins containing BTB/POZ and kelch domains, involved in regulatory/signal transduction processes”) at the COG database (http://www.ncbi.nlm.nih.gov/COG). There were no significant matches found in BLAST searches of the Protein Data Bank, nor in fold recognition searches at 3D-PSSM/Phyre.
(Kelley et al., 2000) or mGenTHREADER (McGuffin et al., 2000) (see section 2.3.8). A protein disorder analysis on representative sequences revealed no regions of extended disorder using PONDR (Romero et al., 2004) and DISOPRED2 (Ward et al., 2004). The putative new domain is predicted to be all \( \alpha \)-helical by Jpred (Cuff et al., 1998).

These results, along with the high level of sequence conservation, implied this intervening region in BTB-Kelch proteins could represent a novel domain. We named this new domain the BACK domain, for Btb and C-terminal Kelch. A multiple sequence alignment was submitted to the Pfam database, and a BACK domain HMM was included in Pfam release 15.0 with the accession number PF07707.

### 2.3.7.2. The PHR domain.

After the identification of the BACK domain, further sequence analysis was performed on proteins that contained the BTB-BACK domain architecture, with no third domain as a replacement for the Kelch motifs in BTB-BACK-Kelch proteins. Sequence analysis on the BTBD1 (NCBI accession number AAK25825) and BTBD2 (AAH83320) proteins, which contain BTB and BACK domains, revealed the presence of a novel 170-residue long conserved region at their C-terminus. BLAST (blastp) searches revealed excellent matches between BTB-BACK proteins in mammals and insects and the Pam (AAC39928), highwire (AAF76150) and RPM-1 (NP_505267.1) proteins. Searches against Pfam and Prodom matched the automatically generated alignments Pfam-B_6072 and PD039495, respectively, while searches against SMART, KOG and COG did not match any known domains. BLAST searches against the PDB and fold recognition searches with 3D-PSSM and mGenTHREADER did not reveal any significant hits. No sequence features such as coiled-coil,
proline-rich, low complexity, transmembrane or nuclear localization signal were identified with the Predictprotein server. Searches for inherent protein disorder with DISOPRED2 (Ward et al., 2004) did not show any regions of disorder. Secondary structure predictions with PHD (Rost et al., 2004) and PSI-PRED (McGuffin et al., 2000) were consistent with each other and showed this region is all-β. Taken together, these searches suggested this region is a novel conserved sequence motif, predicted to be an all-β fold. We named this region the “PHR domain”. This domain is now represented in the Pfam database with the accession number PF08005.

2.3.8. Secondary structure and disorder searches.

Secondary structure content predictions in the analysis of BTB and BACK domain sequences were performed using a combination of the PredictProtein, Jpred and PSIPRED services. Predictions using the PredictProtein service were performed using the PHD and Prof algorithms, with no PSI-BLAST or multiple sequence alignment corrections performed to eliminate bias from other protein sequences. Predictions using the Jpred service were also completed using single sequences, with no BLAST search or multiple sequence alignment performed. Searches with the PSIPRED v.2.5 service were performed using no masks for low-complexity, transmembrane helices, or coiled-coil regions. High-confidence results for secondary structure prediction (corresponding to confidence threshold values of 8 over at least 3 amino acids with Jpred or PSIPRED) were combined into a “consensus” result.

Protein disorder analysis was performed using the PONDR service, with the VL-XT, XL1-XT, and VSL1 algorithms utilized, and the DISOPRED2 server.
2.3.9. Tertiary structure/fold-recognition predictions.

Tertiary structure predictions were performed using the 3D-PSSM/Phyre Protein Fold Recognition Server, and the mGenTHREADER service. Hits with e-values <1e-10 and estimated precision values >90% at Phyre or hits with high (p-value < 0.001) or certified (p-value < 0.0001) matches at mGenTHREADER were accepted as valid matches for a successful fold-recognition. To qualify as a non-successful fold-recognition run (i.e. no known structure detected), we used the criteria that no hits with e-values <1e-3 at Phyre, and hits at mGenTHREADER with only low (p-value <0.1) or guess (p-value >=0.1) matches only.
2.4. Results

2.4.1. BTB fold comparisons.

I began my graduate studies with a survey of the solved structures of BTB domains from the Protein Data Bank (PDB), which included examples from BTB-ZF proteins, Skp1, Elongin C and T1 domains (Figures 2.2 and 2.3). The 3D superposition showed a common region of approximately 95 amino acids consisting of a cluster of five α-helices made up in part of two α–helical hairpins (A1/A2 and A4/A5), and capped at one end by a short solvent-exposed three stranded β-sheet (B1/B2/B3) (Figure 2.2). An additional hairpin-like motif consisting of A3 and an extended region links the B1/B2/A1/A2/B3 and A4/A5 segments of the fold. Because of the presence or absence of secondary structural elements in certain examples of the fold, in this chapter we use the designation A1—A5 for the five conserved α-helices, and B1—B3 for the three common β-strands. We refer to this structure as the core BTB fold. When present, other secondary structure elements are named according to the labels assigned to the original structures. Thus, the BTB-ZF family members PLZF and BCL6 contain additional N-terminal elements, which are referred to as β1 and α1, Skp1 protein contains two additional C-terminal helices labeled α7 and α8, Elongin C is missing the A5 terminal helix, and the T1 structures from Kv proteins are formed entirely of the core BTB fold (Figures 2.2 and 2.3).
The core BTB fold. A) Superposition of representative BTB fold structures from four major subfamilies of BTB-fold containing proteins (BTB-Zinc Finger proteins, Skp1, Elongin C and T1 domain). B) Schematic of the architecture of the BTB fold. The core BTB fold contains the coloured elements, while subfamily-specific N- and C-terminal extensions are coloured grey: BTB-ZF proteins contain β1 and α1, Skp1 contains α7 and α8.

Using the structure superposition, we next investigated the critical sequence determinants for specifying the BTB fold, along with identifying residues that were allowed to diverge through the course of evolution. Sequence comparisons based on the structure superposition of the BTB fold members showed less than 10% identity between examples from different families, except for Skp1 and Elongin C, which is in the range of 14-22%. However, all structures show remarkable conservation with RMSD values of 1.0-2.0 Å over at least 95 residues (Figure 2.3B).
FIGURE 2.3.
Sequence comparisons of representative BTB-fold containing proteins.  A) Structure-based multiple sequence alignment of representative BTB domains from each of the BTB-ZF, Skp1, Elongin C and T1 families.  The core BTB fold is boxed.  Secondary structure is indicated by red shading for α-helices and yellow for β-strands, with the N- and C-terminal extensions shaded in grey.  The low complexity sequences, which are disordered in the Skp1 structures, are indicated by open triangles.  B) Pairwise sequence and structure comparisons of BTB structures.  Cells contain the percent identity and RMSD (Å) value for each structure pair.  Representative structures from the PDB are labeled as follows: a1buo:A and 1cs3:A; 1inex:a; 1ldk:D, 1p22:b, 1fqv:B, 1fs1:B, 1fs2:B; 1hv2:a; 1vcb:B, 1m8:C, 1lqb:B; 1a68_: 1eod:A, 1eoe:A, 1eof:A, 1t1d:A, 1exb:E (rat Kv1.1); 1islg:A; 1r28:A.
Despite these very low levels of sequence relatedness, 15 positions show significant conservation across all of the structures, and 12 of these correspond to residues that are buried in the monomer core (Figure 2.5). Most of these highly conserved residues are hydrophobic and are found between B1 and A3, with some examples in A4. In addition to this common set, conserved residues are also found within specific families. Therefore, from this initial analysis of structures of the BTB domain in the PDB, we concluded that along the full length of the BTB fold, there is a very low level of sequence conservation but a very high level of structure conservation. Where there is higher sequence identity, these residues are often situated in the hydrophobic core of the protein and are likely necessary to maintain the core BTB fold.

Given that there is a low level of sequence identity in the surface-exposed residues of the BTB fold, we proceeded to identify and compare the protein-protein interaction roles these residues play. The four known structural classes of BTB domains show different oligomerization or protein-protein interaction states involving different surface-exposed residues (Figures 2.4 and 2.5). There is little overlap between the interaction surfaces of the homodimeric, heteromeric and homotetrameric forms of the domain, which are represented here by examples from the BTB-ZF, Skp1/Elongin C and T1 families, respectively. Contacts involving the N-terminal extensions of the BTB-ZF class and the C-terminal elements of the Skp1 families form a significant fraction of the residues involved in protein-protein interaction in each of those respective systems, but additional contributions from the 95
residue core BTB fold are involved. There are multiple examples of conserved surface-exposed residues that participate in family-specific protein-protein interactions. For example, the B1/B2/B3 sheet is found in all BTB structures and therefore is part of the core BTB fold, but participates in very different protein interactions in the T1 homotetramers, the Elongin C/ElonginB and Skp1-Cul1 structures. Inspection of T1 residues in this area shows sequences such as the “FFDR” motif in B3 have diverged from the other BTB families to become important components of the tetramerization interface (Figures 2.4 and 2.5) (Nanao et al., 2003). In Skp1, B3 has a distinctive “PxPN” motif that is involved in Cul1 interactions (Zheng et al., 2002).

The connection between A3 and A4 (drawn as a dashed line in Figure 2.2B) is variable in length and in structure, and makes key contributions to several different types of protein-protein interactions. The region adopts an extended loop structure in the T1 domain and Elongin C, where it makes important contributions to the homotetramerization and to the VHL interfaces, respectively (Figures 2.4 and 2.5). In PLZF and BCL6, this segment forms strand β5 and associates with β1 from the partner chain to form a two-stranded antiparallel sheet at the “floor” of the homodimer (Ahmad et al., 2003, Ahmad et al., 1998). In Skp1, this region includes a large disordered segment followed by a unique helix α4, but it is not involved in any protein-protein interactions (Schulman et al., 2000, Zheng et al., 2002, Wu, Xu et al., 2003, Orlicky et al., 2003).

Thus, after analyzing the different self-association and protein-protein interactions of the solved BTB domain structures, we concluded that the solvent-exposed surface of the BTB fold is extremely variable and different residues are utilized for different subfamily-specific interactions.
The most probable sequences (majority-rule consensus sequences) from each of seven different family-specific HMM's were generated with HMMER hmmemit. Residues positions with probability score (P(s)) of less than 0.6 are variable and are indicated by dots, residues with 0.6 < P(s) < 0.8 have intermediate levels of sequence conservation and are indicated by lower case letters, and residues with a P(s) > 0.8 are highly conserved and are indicated by capital letters. Grey shading indicates positions that are similar in at least four of the seven families shown, and selected “signature sequences” that are particular to a specific family are boxed in cyan. Gaps are indicated by blank spaces. Residue positions that are buried in the core of the BTB fold are indicated with circles, and contact sites for four known protein-protein interaction surfaces are shown in the grid below the alignment. The secondary structure elements β1, α1, α4, β5, α7 and α8, occur only in some of the families, and are discussed in the text.

Protein-protein interaction surfaces in BTB domains. Left column: the BTB monomer is shown in the same orientation for each of four structural families with the core fold in black, and the N and C terminal extensions in blue. Middle column: the monomers are shown with the protein-protein interaction surfaces shaded. Right column: the monomers are shown in their protein complexes.
2.4.2. Representation of BTB domains in fully sequenced genomes.

We were interested in identifying all BTB domains in fully sequenced eukaryotic genomes. In order to effectively eliminate redundant sequences and partial fragments and to reduce sampling bias due to uneven database representation, we limited our search to the known and predicted transcripts from seventeen fully sequenced genomes. We searched the Ensembl, Uniprot and Dictybase databases. We generated Hidden Markov Models (HMM’s) to capture the sequence characteristics of each of the four families of BTB structure that were described above. As expected from the low sequence similarities, searches with family-specific HMM’s could not retrieve sequences from the other families in a single iteration. For example, the HMM trained on the BTB domains from BTB-ZF proteins could not immediately retrieve sequences from T1-Kv proteins. Additional sequences were added to each of the family-specific HMM’s in several cycles and the results were compiled into final multiple sequence alignments. The retrieved sequences were manually inspected. As more BTB domains were identified from proteins with different domain architectures than the four families with solved structures, we generated more HMM’s to capture the increasingly large space of sequence diversity of BTB domains. In this fashion, we increased the sensitivity of our search for BTB domains to find remote homologs, while maintaining a high degree of specificity due to the structure-based alignments we used to initiate the search. We have assembled this collection of over 2200 non-redundant BTB domain sequences in a publicly available database at http://btb.uhnres.utoronto.ca.

In addition to the genome-centric analyses, we searched the NCBI nr database with PSI-BLAST (Benson et al., 2004, Altschul et al., 1997). Beginning with the sequence of the BTB domain from the BTB-ZF protein PLZF, T1 sequences were retrieved with e-values
below 10 after four PSI-BLAST iterations carried out with a generous inclusion threshold of 0.1, as previously reported (Aravind & Koonin, 1999). Skp1 and Elongin C sequences could not be retrieved with e-values below 10 starting with BTB-ZF or T1 sequences, even with a PSI-BLAST inclusion threshold 1.0. Based on searches with representative BTB sequences from each of the major families, BTB sequences were consistently retrieved from eukaryotes and poxviruses, but no examples from bacteria or archaea were found, with the remarkable exception of 43 BTB-leucine-rich repeat proteins in the Parachlamydia-related endosymbiont UWE25 (Horn et al., 2004). In general, plant and animal genomes encode from 70 to 200 distinct BTB domain proteins, while only a handful of examples are found in the unicellular eukaryotes. We identified an intermediate number, 41, in the social amoeba *Dictyostelium discoideum* (Eichinger et al., 2005) (Figure 2.6).

Based on these observations, the domain most likely underwent domain shuffling followed by lineage specific expansion (LSE) during speciation events. The most commonly observed architecture across several different families consists of a single N-terminal BTB domain, a middle linker region, and a characteristic C-terminal domain that is often present as a set of tandem repeats (Figure 2.7). Along with domain shuffling and domain accretion, LSE is considered one of the major mechanisms of adaptation and generation of novel protein functions in eukaryotes, and is frequently seen in proteins involved in cellular differentiation and in the development of multi-cellular organisms (Lespinet et al., 2002). For example, both BTB-ZF proteins and KRAB-ZF proteins have essential roles in development and tissue differentiation and have undergone LSE in the vertebrate lineage (Aravind & Koonin, 1999, Shannon et al., 2003, Collins et al., 2001).
A) Representation of BTB proteins in selected sequenced genomes. Twelve of the seventeen genomes we searched are represented, showing of each type of BTB protein architecture as bar segments. Data for Apis mellifera, Canis familiaris, Gallus gallus, Pan troglodytes, and Xenopus tropicalis are available at http://btb.uhnres.utoronto.ca. Several lineage-specific expansions are evident: BTB-ZF and BBK proteins in the vertebrates, the MATH-BTB proteins in the worm, the BTB-NPH3 proteins in the plant, the Skp1 proteins in the plant and worm, and the T1 proteins in worm and vertebrates. In the D. discoideum genome, a single star indicates five BTB-kelch proteins that do not contain the BACK domain, and a double star indicates two MATH-BTB proteins that also contain ankyrin repeats. (B) Phylogenetic relationship of analyzed genomes. Adapted from Eichinger et al., 2005. The total number of BTB proteins is shown above each genome.
2.4.3. Evolutionary relationships and sequence clusters

Classifying BTB domain sequences according to sequence identity would be a valuable exercise in beginning to connect the structure and function of the domain with sequence properties and to allow predictions to be made. For example, if BTB domains that are known to be dimeric in solution cluster together in a well-defined clade, a dimeric structure could be predicted of members of the same clade without this experimental knowledge. We attempted to construct a phylogeny based on BTB domain sequences, but we could not consistently cluster the entire collection. Due to the very low levels of sequence similarity between some of the families (Figure 2.4), we were unable to support phylogenies with significant bootstrap values despite many attempts with several different approaches and algorithms, including distance, maximum parsimony or maximum likelihood methods.

We eventually turned to BLASTCLUST as a more appropriate tool to subdivide this highly divergent set of sequences (Figure 2.7). BLASTCLUST creates clusters of sequences with similar pairwise BLAST e-values that are calculated by an all-vs.-all BLAST search on a database of sequences. Many runs of BLASTCLUST with different e-value cutoff scores were necessary to define successively finer sequence clusters with higher sequence similarities. In general, we noticed BTB domain sequence/structure families correlated with the protein architectures, and the BTB-NPH3, T1, Skp1 and Elongin C families could be distinguished at an identity threshold of 30% with this method. Domain sequences from BTB-ZF, BBK, MATH-BTB and RhoBTB proteins formed distinct clusters only at higher cutoffs, and are thus more closely related (Figure 2.7).
FIGURE 2.7.

BTB sequence clusters and protein architectures. Family-specific N- and C- terminal extensions to the core BTB fold are indicated. Dashed lines indicate regions with no predicted secondary structure, while ordered regions are indicated with either domain notations or thick solid lines. The Uniprot code for a representative protein is indicated. Clustering by BLASTCLUST was based on the average pairwise sequence identity for all BTB domain sequences from our database, except for the RhoBTB proteins, where only the C-terminal BTB domain was used. Domain names are from Pfam.
The BTB domain sequences from vertebrate BTB-ZF and BBK proteins are more closely related, and cannot be separated by BLASTCLUST.

2.4.4. New domain identifications

2.4.4.1. The BACK domain

A very important use of our in-house collection of BTB domain-containing sequences was the identification of additional regions of sequence conservation outside of identified domains. In typical BTB domain proteins, the C-terminal domain is separated from the N-terminal BTB domain by several hundred residues. In the BTB-ZF proteins, the central intervening region has little sequence conservation and shows rapid sequence divergence between orthologs. For example, BCL6 from human and zebrafish have 78%, 37% and 85% pairwise sequence identity over the BTB, central and ZF regions, respectively. However, in the BTB-kelch proteins, we noticed strongly conserved sequence features in the intervening region. The length of this region is approximately 130 residues, and the sequence similarity is highest in the first 70 residues that immediately follow the BTB domain (Figure 2.8). We named this region the BACK domain as it is found predominantly in proteins with **BTB** And **C**-terminal **K**elch repeats, although we have identified it in proteins that have replaced the kelch repeats with other protein-protein interaction domains. Conserved features of the BACK domain include the N-terminal NCLGI sequence and a VR[L/M/F]PLL sequence, several hydrophobic positions, two arginines and four glutamic acids. Most of the conserved residues are non-polar, implying a hydrophobic core.
Multiple sequence alignment, domain architecture, and distribution of BACK domains.  

A) Multiple sequence alignment of examples of the BACK domain from a BTB-BACK-kelch(BBK) proteins. Protein identifier codes include the organism abbreviation, the common protein name and the SWISSPROT/TrEMBL accession number. Organism abbreviations are: Hs=Homo sapiens, Mm=Mus musculus, Rn=Rattus norvegicus, Dm=Drosophila melanogaster, Ag=Anopheles gambiae, Ce=Caenorhabditis elegans, At=Arabidopsis thaliana. Positions that are >95% identical in the alignment are shaded black, while positions that are >75% identical are shaded grey. The C-terminal residues from the leading BTB domain are shaded in blue, and beginning of the first kelch repeat is shaded in green. The predicted secondary structure of the BACK domain is shown, and the known secondary structures of the BTB domain and kelch repeat are indicated below the alignment. The alignment was generated with ClustalX with manual adjustments, and formatted with Alscript. 

B) The BACK domain is found predominantly in proteins containing a N-terminal BTB domain and five to seven tandem C-terminal kelch repeats. 

C) Schematic showing the distribution of proteins with BTB, BACK and kelch domains from three selected genomes. The distribution of human BACK proteins is representative of those observed in other vertebrate genomes. A listing of the proteins from these and other genomes can be found at http://xtal.uhnres.utoronto.ca/prive/BACK.html.
Sequence searches with various examples of the BACK domain implied this domain is novel, as no known domains or fold could be assigned to it. Interestingly, a protein disorder analysis on representative BACK domain sequences revealed no regions of extended disorder and it is predicted to be all \( \alpha \)-helical; in comparison, similar analyses on the intervening region of BTB-ZF proteins suggest that the central region of most of these proteins is disordered and have little or no regular secondary structure.

After publication of the identification of the BACK domain (Stogios & Prive, 2004), the NMR structure of the domain was solved (unpublished, PDB id 2eqx). The fold may be novel based on the lack of significant hits to other protein structures using the DALI and VAST structure similarity servers.

A BACK domain–specific HMM was trained using the BACK region of the multiple sequence alignment shown in Figure 2.8. Two rounds of an iterative HMMER search against the NCBI nr database with an E–value cutoff of 0.005 retrieved 662 hits, most of which were metazoan or poxvirus sequences. The BACK HMM was also used to search SWISS-PROT/TrEMBL transcripts (Botuyan et al., 2001) from 12 completely sequenced genomes with an E–value cut-off of 0.1 (a listing of the retrieved sequences is available at http://xtal.uhnres.utoronto.ca/prive/BACK.html). The majority of BACK domains were found in proteins that also contain BTB and kelch motifs, which we name BBK proteins for BTB-BACK-kelch. BBK proteins are found predominantly in vertebrates (there are 53 examples in humans), although some are found in \textit{D. melanogaster} and in \textit{C. elegans}. A few examples of BACK domains were found in proteins with only BTB or kelch repeats, including eleven BTB–BACK proteins in \textit{A. thaliana}, and in a small number of \textit{Drosophila} proteins without either the BTB or kelch repeats. BACK domains were always found as a single
copy per protein, except in one *G. gallus* predicted transcript. We found 54 BBK proteins in the poxvirus sequences from the NCBI nr database, but there are no known examples of BTB- or BACK-containing proteins in prokaryotes or archaea.

### 2.4.4.2. The PHR domain

Most BACK domains are found in BBK proteins, but we were interested in further studying the sequences of proteins that contain the BTB-BACK tandem but without kelch motifs. A very small number of these were found to contain a conserved region of sequence C-terminal to the BACK domain that replaces the kelch motif. These proteins, including BTBD1 and BTBD2, contain what we found to be a novel conserved domain that is predicted to be all-β structure, which we could not match to known all-β folds. We named this domain the “PHR domain” as it was noticed by another group as a conserved region in the *Pam, Highwire* and *RPM-1* proteins (Xu, Yang *et al.*, 2003).

There are some notable sequence features of the PHR domain (*Figure 2.9*). There are many highly conserved glycines (six are 100% identical, two are 75% identical). This is reminiscent of the glycine-rich Kelch repeat, but its distinctive GG sequence is missing in the PHR domain. The N-terminus of the PHR domain has an invariant RF motif, with a 100% identical tryptophan about 10 residues away. There is a highly conserved region with alternating hydrophobic residues, I-X-F-X-V, indicative of β-strand structure. There is a G-hydrophobic-G-hydrophobic-YG-[GS]-X-G motif, and a similar G-X-X-[ST]-X-X-G-X-X-G motif. The C-terminus of the domain is characterized by a highly conserved NGT-X-X-X-GQIP-X-hydrophobic-hydrophobic-[FY] motif.
Multiple sequence alignment of the PHR domain. Proteins are labeled by their Uniprot ID's and the region of the PHR domain is indicated. Positions that are similar in 90% and 65% of sequences are shaded black and grey, respectively. Predicted secondary structure is indicated under the alignment, showing the domain is predicted to be all-β structure. Alignment was retrieved from Pfam.

Using the sequence alignment in Figure 2.9, we generated a profile HMM and used it to search the NCBI nr and Uniprot sequence databases. Using a conservative e-value cut-off of 10, we found 78 PHR domains in the NCBI nr database, all of which were found in eukaryotes (mammals, fish, insects, nematodes, chicken, frog), and 32 in Uniprot (7 in Swissprot, 25 in Trembl). Unique PHR-containing proteins from Uniprot are summarized in Figure 2.9.

We identified the PHR domain in proteins with a variety of domain compositions, with the largest representation in proteins with the architectures BTB-BACK-PHR (BBP
proteins) and the very large proteins that contain 2 RCC1 repeats, 2 PHR domains and ZF(B-box)-ZF(C3H4) motifs. The PAM, Highwire and RPM-1 proteins fall into the latter class. In all, there are not a large number of PHR-containing proteins, but mammals conserve four members (BTBD1, BTBD2, BTBD3 and BTBD6), and other organisms contain at least one of the two major PHR classes.

Interestingly, BBP proteins share many similar features with BTB-BACK-Kelch (BBK) proteins. The BTB domain in BBP and BBK proteins is not at the extreme N-terminus of the protein, which separates them from the BTB-ZF class of proteins. The BACK domain in BBP proteins is a distant (e-values of $10^4$ to $10^6$ at Pfam) but recognizable relative of the equivalent region in BBK proteins and they show similar predicted secondary structure. The BACK domain in both BBP and BBK proteins is separated by very few residues from the BTB domain (as little as five and two residues in BBP and BBK proteins, respectively). There is a short linker (~10 residues) between the BACK and the Kelch domain, and there is a short linker (~20 residues) between the BACK and the PHR domain. The PHR domain is predicted to be an all-β fold, as is the kelch domain. We attempted to align the predicted secondary structure of the PHR domain with the known secondary structure of various β-propeller structures, such as kelch and RCC1, but we were unsuccessful. We cannot conclude that the PHR domain folds into a β-propeller structure. Nonetheless, the overall protein architecture of BBP and BBK proteins is very similar.
Domain architecture of selected PHR domain-containing proteins. There are other architectures of proteins that contain PHR domains, but only those with BTB or BACK domains are shown in this figure.

2.4.5. Survey of subfamilies of BTB proteins

Using the crude clustering of BTB proteins according to the domain architecture of the full-length proteins, we proceeded to analyze the important sequence and structure features of the major subfamilies of BTB proteins. We predicted secondary structure content of the BTB domain and looked for conserved signature sequences. As well, we used the BTB fold structure superposition to predict protein-protein interaction surfaces (dimerization, tetramerization and Cullin interaction) in many of the major subfamilies of BTB proteins.

2.4.5.1. Long form of the BTB domain

The majority of BTB domains from the BTB-ZF, BBK, MATH-BTB, RhoBTB and BTB-bZip proteins contain a conserved region N-terminal to the core region, which likely forms a $\beta_1$ and $\alpha_1$ structure as seen in PLZF (Ahmad et al., 1998, Li et al., 1999) and BCL6 (Ahmad et al., 2003). We refer to this as the “long form” of the BTB domain, which has a total size of approximately 120 residues. Note that many of the protein domain databases, such as Pfam, SMART and Interpro, recognize only the 95 residue core BTB fold, and do not detect all of these additional elements, even though at least half of the metazoan BTB
domains correspond to the long form. The long form BTB domain sequences also are more highly related to each other than to the BTB-NPH3, T1, Skp1 and Elongin C families, as based on the BLASTCLUST analysis (Figure 2.7). These groupings were consistently observed even when only the residues from the core fold were included in the analysis, and so the sequence clustering is not simply due to the presence or absence of the N-terminal elements. We predict that most long form BTB domains are dimeric, and that several of these associate into higher order assemblies via interdimer sheets involving $\beta_1$, as discussed below.

### 2.4.5.2. The BTB-ZF proteins

In the BTB-ZF setting, the BTB domain mediates dimerization, as shown by crystallographic studies of the BTB domains of PLZF (Ahmad et al., 1998, Li et al., 1999) and BCL6 (Ahmad et al., 2003). This is confirmed in numerous solution studies (Ahmad et al., 2003, Ahmad et al., 1998, Li et al., 1999, Kim, Fang et al., 2002, Li et al., 1997, Deltour et al., 2001). An important component of the hydrophobic dimerization interface in PLZF and BCL6 is the association of the long form elements $\beta_1$ and $\alpha_1$ from one monomer with the core structure of the second monomer. We believe that most BTB domains from human BTB-ZF proteins can dimerize, since 34 of these 43 domains are predicted contain all of the necessary structural elements in the swapped interface including $\beta_1$, $\alpha_1$ and $\beta_5$. As well, many highly conserved residues are found in predicted dimer interface positions (Ahmad et al., 1998). Nine human BTB-ZF proteins lack $\beta_1$, and thus cannot form the $\beta_1$-$\beta_5$ inter-chain antiparallel sheet, and we expect that these domains are also dimeric due to the presence of $\alpha_1$ and the conservation of interface residues.
In nearly all BTB-ZF proteins, the long form BTB domain is at or very near the N-terminus of the protein, and the C\(_2\)H\(_2\) zinc fingers are found towards the C-terminus of the protein. These two regions are connected by a long (100-375 residues) linker segment (Figure 2.7). Except for proteins that are highly related over their full lengths, the linker regions do not identify significant matches in sequence searches of the NCBI nr set. This architecture suggests a model in which the dimeric BTB domain connects the DNA binding regions from each chain via long, mostly unstructured tethers.

In all organisms studied, BTB domains from BTB-ZF proteins show high conservation of the residues Asp35 and Arg/Lys49 (PLZF numbering). These residues are found in a “charged pocket” in the BTB structures of PLZF and BCL6, and have been shown to be important in transcriptional repression (Melnick et al., 2002, Melnick et al., 2000). However, the structure of the BCL6\(^{BTB}\) / SMRT co-repressor complex did not show interactions between this region and the co-repressor (Ahmad et al., 2003). Mutation of Asp35 and Arg49 disrupts the proper folding of PLZF (Melnick et al., 2000), and these residues are thus important for the structural integrity of the domain. Interestingly, Asp35 and Arg/Lys49 are also conserved in the BTB domains from BBK, MATH-BTB and BTB-NPH3 proteins (Figure 2.4).

2.4.5.3. The BBK proteins

Many members of this widely represented family of proteins are implicated in the stability and dynamics of actin filaments (Robinson & Cooley, 1997, Lecuyer et al., 2000, Chen et al., 2002, Hara et al., 2004). With few exceptions, all of the 515 BTB-kelch proteins in our database also contain the BACK domain. These BBK proteins are comprised of a long-form BTB domain, the 130 residue BACK domain (Stogios & Prive, 2004), and a C-
terminal region containing 4-7 kelch motifs (Adams et al., 2000, Bork & Doolittle, 1994, Prag & Adams, 2003). Most BBK proteins have a region of approximately 25 residues that precede the BTB domain, unlike BTB-ZF proteins where BTB is positioned much closer to the N-terminus (Figure 2.7). We predict that this N-terminal region in the BBK proteins is unstructured, although it is shown to have a functional role in some proteins (Robinson & Cooley, 1997). Notably, the distribution of BBK proteins parallels that of the BTB-ZF proteins across genomes. We did not find BBK proteins in A. thaliana or the yeasts.

The sequences of BTB domains from BBK proteins are most closely related to those from BTB-ZF proteins (Figure 2.7), suggesting that they adopt similar structures. BTB domains from BBK proteins have been shown to mediate dimerization (Robinson & Cooley, 1997, Soltysik-Espanola et al., 1999, Sasagawa et al., 2002) and have conserved residues at positions equivalent to those at the dimer interface in BTB-ZF proteins. Indeed, a very recent structure of the BTB domain from the BBK protein Gigaxonin (PDB id 2PPI, unpublished structure) shows remarkable similarity with the structures from BTB-ZF proteins. There are reports of BTB-mediated oligomerization in BBK proteins, consistent with the role of some these proteins as organizers of actin filaments (Chen et al., 2002, Robinson & Cooley, 1997, Sasagawa et al., 2002). Because most of the BTB sequences from BBK proteins are predicted to contain the β1, α1 and β5 long form elements, oligomerization of these proteins may occur via dimer-dimer associations involving the β1 sheet, as proposed for the BTB-ZF proteins. However, there are no strongly characteristic sequences or enrichment of hydrophobic residues in the β1 region.
2.4.5.4. Skp1

Skp1 is a critical component of Cul1-based SCF complex (also known as SCF1, where the number refers to the identity of the Cullin component), and forms the structural link between Cul1 and substrate recognition proteins (Bai et al., 1996, Feldman et al., 1997, Skowyra et al., 1997). Skp1 proteins are only distantly related to other BTB families (Figures 2.3 and 2.7), and are comprised of the core BTB fold with two additional C-terminal helices. These latter helices form the critical binding surface for the F-box region of substrate-recognition proteins. Many Skp1 sequences have low complexity insertions after A3, which are disordered in several crystal structures, followed by helix α4 that is unique to this family (Figure 2.3) (Schulman et al., 2000, Zheng et al., 2002, Wu, Xu et al., 2003, Orlicky et al., 2003). Skp1 proteins are found in all organisms studied, with significant expansions in *C. elegans* and *A. thaliana* (Figure 2.6). Interestingly, the Cul1-interacting surface of Skp1 does not overlap with the dimerization surface seen in BTB-ZF structures, and is mostly separate from the tetramerization surface in the T1 domains (Figures 2.4 and 2.5). Therefore, a unique surface of the BTB fold in the Skp1 proteins has adapted to mediate interactions with Cul1.

2.4.5.5. Elongin C

Elongin C is an essential component of Cul2-based SCF complexes, also known as VCB or ECS E3 ligase (Lonergan et al., 1998, Pause et al., 1997). This protein serves as an adaptor between ElonginB and the VHL tumor suppressor protein, which interacts with HIF-1α and targets it for degradation (Pause et al., 1997, Iwai et al., 1999, Lisztwan et al., 1999, Ohh et al., 2000). In any given organism, the sequence identity between Elongin C and Skp1 is approximately 30% or less, but these proteins are nonetheless more closely re-
lated to each other than to other BTB sequences (Figure 2.3). The structure of Elongin C showed that it is comprised entirely of the core BTB fold, but lacks the terminal A5 helix (Botuyan et al., 2001, Stebbins et al., 1999, Min et al., 2002, Hon et al., 2002). We found Elongin C proteins in all organisms studied. Like Skp1, Elongin C is significantly similar to other BTB sequence classes only in the buried positions of the monomer core (Figure 2.4). A β-strand in the A3/A4 connecting region participates in the Elongin C-VHL interaction, and the sequence in this region is characteristic of Elongin C (Stebbins et al., 1999).

2.4.5.6. The T1 domain in Kv channels

The T1 domain from voltage-gated potassium channels modulates channel gating and assembly (Minor et al., 2000, Nanao et al., 2003, Strang et al., 2001). This domain is a distant homolog to all other BTB domains, and segregates into a unique cluster at less than 30% sequence identity with BLASTCLUST. The T1 domain is found in a large number of voltage-gated potassium channel proteins in all metazoan genomes surveyed (Figure 2.6). Despite the very low levels of sequence similarity to the other BTB domain families, several of the characteristic buried residues are conserved (Figure 2.4). It is striking that most of the residues found in the polar tetramerization contact surface in the T1 structures do not overlap with those residues involved in dimerization in the BTB-ZF structures. Of the 24 residues that are found in the T1 tetramer surface, only 6 are common to the BTB-ZF dimer interface (Figure 2.4). Thus, a unique set of residues has evolved in the T1 domain to mediate tetramerization.

2.4.5.7. The MATH-BTB proteins

A large expansion of MATH-BTB proteins occurred in C. elegans, where 46 of 178 total BTB proteins belong to this family, whereas other genomes contain many fewer of
these proteins (Figure 2.7). MATH proteins as a whole are largely expanded in *C. elegans*, with 95 examples present in the Pfam database (Bateman et al., 2004). The MATH domain is thought to be a substrate recognition module in SCF3 complexes (Pintard et al., 2003, Xu, Wei et al., 2003).

MATH-BTB proteins differ from most other BTB families in that the BTB domain is found C-terminal to the partner domain. Typically, there are an additional 75-100 amino acids following the BTB domain that are likely to be structured and rich in α-helices. In contrast to the BTB-ZF proteins, but similar to the BBK proteins, MATH-BTB sequences are highly conserved across the full lengths of the proteins. As a result of this conservation, phylogenetic clustering of the full-length protein sequences can be done with reasonable bootstrap values and shows a clear demarcation between proteins from *C. elegans* and those from all other species (data not shown). The domain in the *C. elegans* proteins lacks several BTB signature sequences, such as the “AH[RK]XVLAA” signature in the B2-A1 region seen in many other long form BTB families (Figure 2.4). The majority of MATH-BTB proteins from all organisms are predicted to contain the long form elements β1, α1 and β5 and we predict that these BTB domains are dimeric. Indeed, biochemical and biological evidence suggest that BTB-mediated dimerization of the MATH-BTB protein MEL-26 is required for its function (Pintard et al., 2003, Dow & Mains, 1998).

**2.4.5.8. The BTB-NPH3 proteins**

Another large expansion is found in *Arabidopsis*, which contains 21 BTB-NPH3 proteins, or over 25% of the BTB proteins in this genome. BTB-NPH3 proteins are not found in any of the other genomes that we considered, and could represent a plant-specific adaptation of the BTB domain. BTB-NPH3 proteins are involved in phototropism in *A. thaliana*
and are thought to be adaptor proteins that bring together components of a signal transduction pathway initiated by the light-activated serine/threonine kinase NPH1 (Motchoulski & Liscum, 1999, Sakai et al., 2000). Heteromerization of BTB-NPH3 proteins have been observed, and the BTB domains of RPT2 and NPH3 have been shown to interact (Motchoulski & Liscum, 1999, Sakai et al., 2000). In addition, the BTB domain from RPT2 can interact with a region of PHOT1 that contains LOV protein-protein interaction domains (Inada et al., 2004). These proteins consist of an N-terminal BTB domain and an NPH3 domain (Figure 2.7). The BTB domains in this family are only distantly related to other examples of the fold, and appear to have two leading β-strands in a region preceding the core fold, with an additional β-strand between A1 and A2.

2.4.5.9. BTB-bZip proteins

Each of the vertebrate genomes considered here contain genes for two BTB-bZip proteins, named Bach1 and Bach2 (Ohira et al., 1998, Oyake et al., 1996), except for Danio rerio, which has three. These proteins are transcription factors and most closely resemble the BTB-ZF proteins in terms of the BTB sequence and overall protein architecture. The proteins consist of a long form BTB domain, a central region of approximately 400 residues, and a C-terminal basic leucine zipper region (Figure 2.7). The close similarity of the BTB sequences between the BTB-ZF and BTB-bZip proteins suggest that these domains are likely to be similar in structure. Notably, the long form elements and β5 are predicted, and dimerization residues are similar to the ZF class (data not shown). Accordingly, the Bach proteins have been shown to dimerize and oligomerize in a BTB-dependent manner (Igarashi et al., 1998). bZip domains themselves are known to dimerize and interestingly, the majority of bZip-containing proteins (550 of 738 Pfam bZip_1 domain) contain no other
identified domains in the full-length protein (Bateman et al., 2004). Therefore, the domain composition and sequence properties of BTB-bZip proteins are unusual in the context of all bZip proteins, but are compatible with dimeric, and most likely oligomeric, BTB transcription factors.

2.4.5.10. The RhoBTB proteins

These proteins have an unusual architecture, and contain a Rho GTPase domain near the N-terminus, two tandem long form BTB domains, and an approximately 100 residue C-terminal tail with predicted α-helical content (Figure 2.7). These proteins are highly conserved across their full-lengths, and three examples (RhoBTB1, RhoBTB2/DBC2, RhoBTB3) are found in each of the vertebrates included in this study (Salas-Vidal et al., 2005, Ramos et al., 2002, Rivero et al., 2001). One RhoBTB protein is also present in the insects. The first BTB domain of human RhoBTB2 has been shown to interact with Cul3 (Wilkins et al., 2004) and contains a large 115 residue insertion between A2 and B3, while the second domain is more typical and most closely resembles BTB domains from BBK proteins. The tandem domains are immediately adjacent and may form an intramolecular dimer.

Mutations have been identified in lung cancer patients that do not disrupt the RhoBTB2-Cul3 interaction (Wilkins et al., 2004), and these map to regions outside of the predicted Cul3-interacting region (see below). However, we predict that the Y284D cancer mutation is found in the dimerization interface of the first BTB domain and prevents the proper folding of the domain. This would be analogous to mutants in the dimer interface of PLZF that abrogate function by affecting the folding of the domain (Melnick et al., 2000).
The PLZF and BCL6 BTB domains are obligate dimers, and cannot fold as stable monomers (unpublished observation and Li et al., 1999).

2.4.5.11. The BTB-ankyrin proteins

Ankyrin repeats are common protein-protein interaction motifs that are found in proteins of very diverse function, such as transcription regulators, ion transporters and signal transduction proteins (Mosavi et al., 2004, Breeden & Nasmyth, 1987). We found examples of BTB-ankyrin proteins in each species that we considered, however, unlike other BTB domain families, these proteins do not fit a single canonical arrangement. For example, some BTB-ankyrin proteins are comprised of an N-terminal BTB domain, a central helical region, 19 ankyrin repeats and a C-terminal FYVE domain, while other examples contain two ankyrin repeats followed by a linker region, two tandem BTB domains, and a 300 residue C-terminal helical region. The three BTB-ankryin proteins from \textit{S. pombe} (Btb1p, Btb2p, Btb3p) are components of a SCF ubiquitin ligase complex and interact with Pcu3p, a Cul3 homolog (Geyer et al., 2003). Both BTB domains of Btb3p are necessary for this interaction. The BTB sequences from these proteins are only distantly related to other BTB domains, and we thus cannot reliably predict the nature of their interaction surfaces.

2.4.5.12. BTB proteins with no other identified domain

A significant number of BTB proteins do not contain other identified sequence motifs (Figure 2.7). Excluding the Skp1 and Elongin C proteins, 52% of the \textit{C. elegans} BTB proteins, but only 17% of the human proteins, belong to this family. There may be additional domains in some of these proteins that have yet to be identified.
2.4.6. BTB domains in Cullin complexes

Several members of the BTB families described here have been found to interact with SCF3 complexes including BTB-ZF (Furukawa et al., 2003), BBK (Furukawa et al., 2003) (Kobayashi et al., 2004, Cullinan et al., 2004), MATH-BTB (Furukawa et al., 2003, Pintard et al., 2003, Xu, Wei et al., 2003), RhoBTB (Wilkins et al., 2004), BTB-ankyrin (Geyer et al., 2003), BTB-only (Furukawa et al., 2003, Geyer et al., 2003) and T1-Kv (Xu, Wei et al., 2003) proteins. The roles of Skp1 and Elongin C as integral components of SCF1 and SCF2 complexes, respectively, have long been established (Feldman et al., 1997, Deshaies, 1999). In SCF1 complexes, F-box proteins such as Cdc4 form precise complexes with Skp1 helices α7 and α8 via their F-box, thus positioning their ligand-binding C-terminal WD40 β-propeller domain such that bound substrate is ubiquitinated by the E3 ligase (Wu, Xu et al., 2003, Orlicky et al., 2003).

Nine of the 49 human BBK proteins have been identified as components of Cul3 SCF complexes (Kobayashi et al., 2004, Furukawa et al., 2003), and in several cases, the BTB domain is necessary and sufficient for interaction with Cul3. We propose that the BBK proteins are structurally analogous to the two-chain Skp1/Fbox or Elongin C/SOCS box complexes (Stogios & Prive, 2004). In these cases, the central BACK domain would serve to position the C-terminal β-propeller kelch repeats for substrate recognition (Li, Zhang et al., 2004). We expect a similar situation in the BBP proteins, where the PHR domain would act at the substrate recognition module.

BTB domains of five of the 46 MATH-BTB proteins from C. elegans have been shown to interact with Cul3. As in the BBK proteins, the MATH-BTB proteins are conserved over much of their entire length, and are likely to be internally rigid. In this scenario, the sub-
strate-recognizing MATH domain is found N-terminal to the BTB domain, but since the N- and C-termini are very close to each other in the long form BTB domain dimer (Furukawa et al., 2003, Pintard et al., 2003, Xu, Wei et al., 2003), the MATH domain in these proteins may occupy a similar spatial position relative to the BTB dimer as the BACK-Kelch region of BBK proteins.

Some BTB-ZF proteins, including PLZF, have also been shown to bind to Cul3, presumably in a BTB-dependent mode (Furukawa et al., 2003). The role of these proteins in SCF3 complexes pose a puzzle, since we do not expect that downstream ZF domains maintain a fixed orientation relative to the BTB domain due to the structurally disordered central region. Further work will be required to understand the structure and function of BTB-ZF proteins in SCF3 complexes.

### 2.4.7. A model of the ubiquitin-E2-Cul3-Rbx1-BBK complex.

To aid in understanding the role of the BTB domain in the SCF3 complex, we generated a structural model of a BBK protein dimerized via its BTB domain in a complex with Cul3, Rbx1, E2 and ubiquitin (Figure 2.11). Three different structures of Skp1 complexes are known (Zheng et al., 2002, Wu, Xu et al., 2003, Orlicky et al., 2003), including a Cul1-Skp1 complex (Zheng et al., 2002). We generated a homology model of human Cul3 based on the structure of Cul1 and placed the PLZF BTB dimer by superposing one chain of the dimer with Skp1. Residues in Skp1 that interact with Cul1 are found at positions that do not involve the dimer interface residues in PLZF (Figures 2.4 and 2.5). The BTB domain from the BTB-ZF, BBK and MATH-BTB and BTB-bZip families are closely related (Figure 2.7) and contain mostly the long form of the domain, as discussed above. We predict these
to form obligate dimers, similar to those observed in PLZF and BCL6 (Ahmad et al., 2003, Ahmad et al., 1998, Li et al., 1999). Proteins from each of these families have been shown to interact with Cul3, therefore, it is reasonable to postulate that these BTB domains drive the dimerization of Cul3 complexes. Indeed, dimerization of adaptor proteins is known to occur (Maniatis, 1999). The resulting model is similar to models presented for the ubiquitin-E2-SCF1Cdc4 (Orlicky et al., 2003) and E2-SCFβ-TrCP1 complexes (Wu, Xu et al., 2003), except that two ligand-binding kelch/WD40 domains and two E2-ubiquitins localize to the same face of the dimeric complex. In each BBK protein, the BACK domain is between the N-terminal BTB domain and the C-terminal ligand-binding domain and is likely to be important for positioning the substrate in the complex.

A more precise model for a dimeric SCF3 complex will require the structure of the BACK domain. There is recent evidence that a SCF1 complex is dimeric via a dimerization motif in the F-box protein (Tang et al., 2007). Dimerization of this SCF1 complex allows for accommodation of multiple different acceptor lysines for ubiquitination. Strikingly, the model for this complex strongly resembles the model presented here for the dimeric SCF3 complex. Since the BTB domain of BBK proteins likely mediates dimerization, our model for a dimeric SCF3 complex is strongly plausible.
Structural model of the ubiquitin-E2-Cul3-Rbx1-BBK complex. The complex is dimerized by the self-association of the BTB domain in the BBK protein. Each full-length BBK protein is shown in red, with the BTB dimer shown in the darkest shading in surface representation, the two BACK domains in pink.
surface, and the two Kelch β-propellers shown in pink cartoon representation. The Cul3 homology model is shown in green cartoon, Rbx1 is in grey cartoon, E2 Ubch7 is in yellow cartoon, and ubiquitin is shown in blue surface. Stars indicate the position associated with substrate binding (Li, Zhang et al., 2004).
2.5. Conclusions

This study illustrates the diversity in the abundance, distribution, protein architecture and sequence characteristics of BTB proteins in seventeen eukaryotic genomes. We surveyed public databases and fully sequenced genomes and identified several lineage-specific expansions. The BTB domain is found in a wide variety of proteins, but it most often occurs as a single copy at or near the protein N-terminus. Residues exposed at the surface of the BTB fold are highly variable across sequence families, reflecting the large number of self-association and protein-protein interaction states seen in solved BTB structures. Most BTB-ZF, BBK and MATH-BTB proteins contain a long form of the domain that has an additional conserved N-terminal region, and these are predicted to form stable dimers. We identify the BACK domain as a novel domain found in nearly all BTB and Kelch motif containing proteins, and the PHR domain as a novel domain that substitutes for the Kelch motif and could function as a protein-protein interaction motif. In BBK proteins, the C-terminus of the BTB domain and the N-terminus of the BACK domain are separated by as few as two residues, suggesting the domains are closely associated or may actually form one continuous structure. Secondary structure predictions suggest the BACK domain adopts an all-α conformation, which we suggest could be similar to the C-terminal extension of Skp1 and the F-box motif. Based on structural superpositions, we show that the Cul3 interaction surface on many BTB proteins does not overlap with the dimerization interface, and therefore these BTB proteins may drive the dimerization of SCF3 complexes.
Chapter 3

Crystal structure of the LRF BTB domain and insights into discrimination for binding a peptide from the SMRT co-repressor

This chapter has been reformatted from the original publication:


Reprint courtesy of Protein Science.
3.1. Abstract

In this chapter, we describe the structure of the LRF BTB domain. The structure revealed a typical homodimer that closely resembles the first two structures of BTB domains, those from PLZF and BCL6. However, close examination of the structure sheds light on the co-repressor binding properties of BTB domains. LRFBTB has novel features on the surface of the homodimer, including differences in the lateral groove and charged pocket regions. The residues that line the lateral groove have little similarity with the equivalent residues from the BCL6 BTB domain and we show that the 17-residue BCL6 Binding Domain (BBD) from the SMRT co-repressor does not bind to the LRF BTB domain. This suggests that while the BTB homodimeric structure is often well conserved, substitutions in residues on the surface of the domain result in diversity of co-repressor interactions in BTB-ZF proteins.
3.2. Introduction

The product of the *Zbtb7* gene codes for a BTB-ZF protein that we referred to as LRF but is known by several names. LRF and its orthologs have been characterized in a variety of functional roles. The protein was first named FBI-1 or Factor Binding to IST-1 due to its ability to bind to sites on the HIV-1 genome known as the inducer of short transcripts (IST) (Pessler et al., 1997). FBI-1 was shown to interact with itself and the HIV-1 viral activator Tat (Pendergrast et al., 2002, Morrison et al., 1999). FBI-1 was also involved in the differentiation of preadipocyte cells (Laudes et al., 2004). The mouse ortholog, LRF or Leukemia/Lymphoma Related Factor, was identified as a localization and heterodimerization partner of BCL6 (Davies et al., 1999). The rat ortholog, OCZF or Osteoclast-derived zinc finger, was shown to regulate differentiation of osteoclast cells and to be a transcription repressor that localizes to discrete nuclear foci (Kukita et al., 1999). LRF was shown to bind and regulate the expression of many genes containing the LRF Binding Element (LBE) sequence. The LBE sequence G(A/G)GGG(T/C)(T/C)(T/C)(T/C) or the related GACCCCCCC, is often repeated in different orientations and spacing in many genes (Pessler & Hernandez, 2003), including those encoding extracellular matrix collagen type I, II, IX, X and XI, aggrecan, fibronectin, elastin and human cartilage oligomeric matrix protein (COMP) (Widom et al., 2001, Liu et al., 2004), alcohol dehydrogenase ADH5/FDH (Lee et al., 2002), the ARF tumour suppressor (Maeda, Hobbs, Merghoub et al., 2005) and the c-fos and c-myc oncoproteins (Pessler & Hernandez, 2003). Furthermore, LRF interacts with other important transcription factors. LRF interferes with GC box recognition by SP-1, dependent on an interaction between LRF$^{BTB}$ and the ZF region of SP-1 at the ADH5/FDH
gene (Lee et al., 2002). LRF enhanced transcription of NF-κB responsive genes by facilitating nuclear import, nuclear stabilization and blocking nuclear export of this transcription factor (Lee et al., 2005). An interaction between LRF\textsuperscript{BTB} and the Rel Homology Domain (RHD) of the p65 subunit of NF-κB or IκB was necessary for this activity.

LRF is a repressor of the ARF tumour suppressor gene (p19\textsuperscript{Arf} in the mouse, and p14\textsuperscript{ARF} in humans) and is a central regulator in oncogenesis (Maeda, Hobbs, Merghoub et al., 2005, Maeda, Hobbs & Pandolfi, 2005). LRF overexpression lead to reduced levels of ARF, resulting in the degradation of nuclear p53 and oncogenic transformation. Conversely, reduced levels of LRF results in senescence and apoptosis. Notably, LRF levels are often elevated in many human cancers, often in association with high levels of BCL6. In fact, LRF expression was required for transformation by classical oncogenes such as Ras and Myc. This indicated LRF plays a critical role in oncogenesis, likely due to its involvement in pathways regulating p53 expression.

The structure determinations of the BTB domains from the PLZF and BCL6 transcription regulators revealed a variety of insights. Likewise, we expected that determination of the structure of the BTB domain from LRF could shed light on the function of this important transcription factor, particularly with respect to interactions with components of the histone deacetylase complex. Therefore, we crystallized this BTB domain and tested for an interaction with a well-characterized peptide from the SMRT nuclear co-repressor, previously shown to be the minimally necessary interaction region between BCL6\textsuperscript{BTB} and the SMRT protein (Ahmad et al., 2003).
3.3. Methods

3.3.1. Cloning of the LRF BTB domain.

cDNA encoding full-length LRF (MGC id 99631) was obtained as a Mammalian Gene Collection (MGC)-compliant clones from American Type Culture Collection. The region corresponding to the BTB domain (residues 1-131) was amplified by PCR with primers that created 5’ BamHI and 3’ HindIII sites. The amplified and doubly-digested fragment was subcloned into a modified pET-32(a) (Novagen) T7-expression vector, coding for a fusion protein with a N-terminal thioredoxin domain followed by a six-histidine tag, a 56-amino acid linker, a thrombin protease cleavage site, followed by the BTB domain. Cloning for LRF BTB domain was completed by Denise Jaworksy.

3.3.2. Expression and purification of the LRF BTB domain.

The pET-32(a) based construct was transformed into E.coli BL21(DE3) codon plus (Stratagene) cells, which were grown as 2 L cultures at 37°C in LB medium in the presence of 100 mg/L ampicillin. Culture was grown to an OD_{600} of 0.5, when protein expression was induced with 100 µM IPTG for 3 hours at 22°C. Cells were harvested by centrifugation, and the pellet was resuspended in 40 mL lysis buffer (500 mM NaCl, 20 mM Tris pH 8.5, 10 mM imidazole and 5 mM β-mercaptoethanol), then rigorously vortexed. Resuspended cells were lysed by pressure using an Emulsiflex (Avestin). The lysate was centrifuged at 35,000 rpm for 35 minutes in an ultracentrifuge (Beckman) to remove membranes and insoluble material. The thioredoxin-six-His-BTB fusion protein in the soluble supernatant was purified by immobilized metal affinity chromatography (IMAC) with a Ni^{2+}-
sepharose affinity column (GE) pre-equilibrated with lysis buffer, washed extensively with lysis buffer, then eluted with elution buffer (lysis buffer but with 100 mM imidazole). The peak elution fractions were purified by size exclusion chromatography on a Superdex 200 column (Amersham Pharmacia Biotech) equilibrated in thrombin cleavage buffer (300 mM NaCl, 20 mM Tris pH 8, 10 mM β-mercaptoethanol). The fusion protein was digested with 4 units thrombin (human plasma thrombin, high activity; Calbiochem) per mg protein for 24-72 hours at RT or until thrombin cleavage was complete as monitored by SDS-PAGE. Cleavage was stopped by addition of 50 μL benzamidine-sepharose beads (GE Healthcare), followed by a 15-minute incubation. Uncleaved fusion protein and the thioredoxin-six-His tag were separated from free BTB by reverse IMAC (immobilized metal affinity chromatography) using two Ni²⁺-sepharose columns in series, with the flow-through passed over the columns twice to ensure separation of the BTB domains. LRF BTB domain was further purified by size exclusion chromatography on a Superdex 200 column equilibrated with the final crystallization buffer (300 mL NaCl, 20 mM Tris pH 8.0, 1 mM TCEP). Protein concentration of purified LRF BTB domain was determined by UV spectrophotometry using an extinction coefficient of 4470 M⁻¹ cm⁻¹. LRF BTB domain was stored at 4 °C until use, or flash frozen into liquid nitrogen and stored at –70° C in the presence of 20% glycerol for longer periods of time.

3.3.3. Crystallization of the LRF BTB domain.

Crystallization screens using Hampton Crystal Screens I and II were performed by the hanging drop vapour diffusion technique in 24-well culture plates (Linbro). Diffraction-quality crystals were grown by hanging drop set up in 24-well Linbro plates with each drop
containing 1 μL LRF BTB domain at 6.5 mg/mL plus 1 μL reservoir solution. The 600 μL reservoir solution that produced crystals was 0.7 M sodium citrate pH 5.44. Crystallization of LRF BTB domain was performed by Lu Chen.

### 3.3.4. Structure determination and refinement of the LRF BTB domain.

The LRF BTB domain crystal was cryoprotected with 20% glycerol prior to flash freezing in a liquid nitrogen stream to 100 K. Native diffraction data collected at beamline A1 at the Cornell High Energy Synchrotron Source (CHESS) using a ADSC Quantum Q210 detector. Data were collected using 10 second exposures, 0.5° oscillation width over 90° in total with detector at 180 mm. Diffraction data were reduced with the HKL package, SCALEPACK and DENZO (Table 3.1).

### 3.3.5. Peptide binding assay.

LRF and BCL6 BTB domains were mixed with Thioredoxin-SMRT (SMRT residues 1414-1430, the BBD, expressed and purified as described in (Ahmad et al., 2003) and run at 4° C on a 5% native PAGE gel in Tris-glycine buffer at pH 8.8. Equal amounts of LRF and BCL6 BTB domains were loaded (10 μg) in all lanes, while the amount of added Thx-SMRT ranged from 2.5 to 13 μg. The calculated isoelectric points are 4.4 and 6.3 for LRF$^{BTB}$ and BCL6$^{BTB}$, and 5.4 for Thx-SMRT.
3.4. Results and Discussion

3.4.1. Expression and purification of the LRF BTB domain.

Due to the success of the expression of the PLZF and BCL6 BTB domains as thioredoxin-6-histidine-tagged fusion proteins, a panel of 22 BTB domains were cloned by the lab technician Denise Jaworsky into the corresponding modified pET-32(a) expression vector. As well, the precise N- and C-terminal boundaries of each BTB domain in the panel could be predicted based on the successfully solved structures of the PLZF and BCL6 BTB domains. Denise carried out an expression and solubility screen with the panel of BTB domains, which showed that the Miz-1 (BTB identifier = 5), FAZF (identifier = 38), Kaiso (identifier = 31) and LRF (identifier = 46) BTB domains were among the most highly expressed and soluble proteins as fusion proteins (Figure 3.1). Therefore, these proteins were purified in large scale for crystallization. This chapter discusses characterization of the LRF BTB domain, while subsequent chapters focus on the FAZF, Miz-1 and Kaiso BTB domains.

A single 2 L *E. coli* culture yielded 80-90 mg of the thioredoxin-6xhis-tagged LRF BTB domain fusion protein. After the four chromatographic steps, the final yield for the purified BTB domain was approximately 10-20 mg. The SDS-PAGE gel and gel filtration profiles showing a purification of the Miz-1 BTB domain (Figures 4.1 and 4.2) is representative of the purifications of the LRF BTB domain and all BTB domains studied in this thesis.
3.4.2. Crystallization of the LRF BTB domain.

LRF BTB domain readily crystallized with the hanging drop vapor diffusion method (Figure 3.2) in the condition described in section 3.3.3.
3.4.3. Structure determination of the LRF BTB domain.

Native diffraction data for the LRF BTB domain crystal were collected at the CHESS A1 beamline. Data collection is summarized in Table 3.1.

Table 3.1. Data collection and refinement statistics for LRF BTB domain crystal.

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P4_2_2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>72.14, 72.14, 116.05</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>61.31-2.10</td>
</tr>
<tr>
<td>R_sym (%)</td>
<td>6.7 (44.2)</td>
</tr>
<tr>
<td>I / σI</td>
<td>15.45 (3.08)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.7 (95.6)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Refinement
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>61.3 - 2.10 Å</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>16966</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>21.6 / 26.7</td>
</tr>
<tr>
<td>Number of atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1745</td>
</tr>
<tr>
<td>Water</td>
<td>137</td>
</tr>
<tr>
<td>Solvent (SO&lt;sub&gt;4&lt;/sub&gt;&lt;sup&gt;2-&lt;/sup&gt;)</td>
<td>N/A</td>
</tr>
<tr>
<td>Average B-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>35.3</td>
</tr>
<tr>
<td>Water</td>
<td>44.8</td>
</tr>
<tr>
<td>Solvent (SO&lt;sub&gt;4&lt;/sub&gt;&lt;sup&gt;2-&lt;/sup&gt;)</td>
<td>N/A</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.017</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.642</td>
</tr>
</tbody>
</table>

Values in brackets correspond to the highest resolution shell of 2.15 - 2.10 Å.  
\begin{align*}
R_{sym} & = \frac{\sum_{hkl} |I(hkl) - <I(hkl)>|}{\sum_{hkl} I(hkl)}, \\
R_{work} & = \frac{\sum |F_o - F_c|}{\sum |F_o|}
\end{align*}

The asymmetric unit of the LRF<sub>BTB</sub> crystal contained two chains, with a calculated Matthew’s coefficient of 2.67 Å<sup>3</sup>/Da and a solvent content of 54%. Phaser was used for molecular replacement with the search model the full BTB dimer from PLZF<sub>BTB</sub> (PDB id 1BUO, Ahmad <em>et al.</em>, 1998) with no residue substitutions or deletions of loops or turns. The final LRF<sub>BTB</sub> model was rebuilt using ARP/wARP (Perrakis <em>et al.</em>, 2001), Refmac5 (Murshudov <em>et al.</em>, 1997) and O (Jones <em>et al.</em>, 1991). Initial rounds of refinement were carried out using non-crystallographic symmetry restraints due to the presence of two chains in the asymmetric unit during simulated annealing using CNS. The final atomic model includes residues 7-129 of LRF, except residues 66-71, with missing density for the sidechains of R129 of chain A and Q72 of chain B. The structure was validated for ideal geometry using Procheck, which indicated 92% and 8% of all residues lie within the most favoured and additionally allowed regions of the Ramachandran plot, respectively.
3.4.4. General description of the LRF BTB domain structure.

The BTB domain from LRF showed a strand-exchanged homodimer (Figure 3.3A). The N-terminal β1 region of each monomer interacted exclusively with its partner chain, forming an interchain β1-β5 sheet. The second major contribution to the dimer interface was the tight packing of helices α1, α2 and α3 between the two subunits. Using the nomenclature of domain-swapped proteins (Liu & Eisenberg, 2002), dimerization was mediated by an “open interface”, involving α1 from one chain plus α2 and α3 from the other, and a “closed interface”, comprised of β1 and α6 from one chain and β5 from the other.

The extensive dimer interface involves 44 residues and 1647 Å² in buried surface area. This is similar to what is observed in the other BTB domain structures. We have not observed the presence of monomers or any inter-subunit exchange between LRF_BTB dimers in solution, and as with PLZF and BCL6 (Ahmad et al., 1998, Li et al., 1999, Ahmad et al., 2003), we described the LRF BTB domain as an obligate homodimer. This is consistent with the result that LRF interacts with DNA as at least a dimeric molecule (Pessler & Hernandez, 2003).
There was a negatively charged patch on each side of the homodimer in a region that is far from the dimerization interface (Figure 3.3B). This region is close to but separate from the predicted surface by which some BTB domains may interact with Cul3 (Stogios et al., 2005). The role, if any, of this patch is unknown, but could represent a binding surface for proteins such as the RHD domain or the ZFs of SP-1 (Lee et al., 2002, Lee et al., 2005).

No electron density was observed in either of the two chains in the loop connecting $\alpha3$ and $\beta4$ (Figure 3.3A). Based on the sequence, GAVVDQQ, the loop was not expected
to be disordered (i.e. does not have a high proportion of glycine or serine residues, for example) and forms the lining of the charged pocket region of the domain. It was interesting that there is a lack of electron density for these residues and suggested that this region was likely to be flexible and unstructured in solution. It may adopt a fixed conformation if involved in a protein binding interaction.

The LRF\textsuperscript{BTB} structure closely resembled the other known BTB domains, and the RMSD between equivalent C\textalpha atoms is 1.5 Å between LRF and PLZF (37\% identical at sequence level), and 2.2 Å between LRF and BCL6 (32\% identical) (Figure 3.4). There were minor differences in the structures, including the missing loop in LRF\textsuperscript{BTB}, the length of the α3-β4 loop, the β4-α4 turn and a slight rotation of α6 at the C-terminus of the domain.

**Figure 3.4.**

Superposition of C\textalpha atoms of the LRF, PLZF and BCL6 BTB domains (PDB id codes 2nn2, 1buo, 1r29).

*LRF is shown in red, PLZF in blue, BCL6 in green.*
3.4.5. The lateral grooves of the LRF and BCL6 BTB domains showed many residue differences.

Two BBD peptides of the SMRT corepressor were shown to interact with the lateral grooves of the BCL6\textsuperscript{BTB} homodimer (Ahmad et al., 2003). This groove is at the dimer interface and involves residues from both chains. At the time of my studies it was not clear if all BTB domains interact with this peptide, therefore we compared the residues of the lateral grooves of the BCL6 and LRF BTB domains to gain insights into the possibility of SMRT-BBD binding to LRF. Interestingly, of the 30 BCL6 residues involved in the BCL6\textsuperscript{BTB} / SMRT-BBD interface, only 9 residues were identical in LRF\textsuperscript{BTB}, and 17 positions showed non-conservative substitutions including many charge-reversal changes (Figure 3.5). This resulted in a very different residue composition and charge distribution in the LRF lateral groove. Notably, the most important BTB residues in the BCL6\textsuperscript{BTB} / SMRT-BBD interface were residues Q10, R13, R24 and H116, and the equivalent residues in LRF are P12, D15, E26 and A118. The absence of the H116 sidechain was notable as this residue formed a clasp in the BCL6 BTB domain over the SMRT peptide, making many interactions (Ahmad et al., 2003). As well, \(\beta\)1 was primarily hydrophobic in LRF\textsuperscript{BTB}, but polar in BCL6\textsuperscript{BTB}. The many differences suggested that the SMRT-BBD does not bind to the LRF BTB domain, at least in the same mode as seen in the BCL6\textsuperscript{BTB} / SMRT-BBD complex.
Comparison of the lateral grooves of the LRF and BCL6 BTB domains. A sequence alignment of \( \text{LRF}^{\text{BTB}} \) and \( \text{BCL6}^{\text{BTB}} \) is shown, with residues located in the lateral groove region indicated with asterisks. Lateral groove residues that show non-conservative substitutions between the two BTB domains are shaded by residue type: blue=basic, red=acidic, grey=hydrophobic, cyan=polar, yellow=sulfur-containing. Conservative substitutions are not shaded, except for His-116/Ala-118, which is boxed. Residues are numbered according to the LRF sequence. The region of the disordered loop in the LRF structure is underlined. The structures of the LRF and BCL6 BTB domains are shown in surface representation below the sequence alignment, with the lateral groove residues colored as in the alignment. Residues that have the largest buried surface areas in the \( \text{BCL6}^{\text{BTB}} / \text{SMRT-BBD} \) complex, and their equivalents in \( \text{LRF}^{\text{BTB}} \), are labeled. Residues not involved in the lateral groove are colored in shades of yellow. The charged pocket region of the BTB domains is indicated with an arrow.
3.4.6. Comparison of the charged pocket of BTB domains.

The loop between α3 and β4 of the BTB domains formed the edge of a charged pocket, containing two aspartate and two arginine residues, that was thought to be functionally important for transcription repression even though it did not contact the SMRT-BBD (Melnick et al., 2002, Melnick et al., 2000, Ahmad et al., 2003, Puccetti et al., 2005). The charged pocket was different in depth and contains different residues in each BTB domain structure (Figure 3.6). In all three structures, the bottom of the charged pocket was made up of D33/35 and R47/49 in nearly identical positions. The LRF and PLZF BTB domain charged pockets most closely resembled each other, as D35 and R49 were the only charged features and there were neutral features that extend away from the pocket. The BCL6<sup>BTB</sup> charged pocket showed additional basic features and the width of the pocket was larger than the other BTB domains. Notably, the PLZF BTB domain interacted only weakly with SMRT via regions distinct from the SMRT-BBD. Therefore, the residue composition in the charged pockets of PLZF<sup>BTB</sup> and LRF<sup>BTB</sup> was consistent with weak or no SMRT interactions.
The lack of electron density for the α3-β4 loop was not unusual given that other BTB structures have unresolved loops. The recently solved structure of the BTB domain from Bach1, a BTB-leucine zipper transcription factor, did not show electron density for this same loop at the charged pocket (unpublished, PDB code 2ihc). Furthermore, the BTB-ZF protein Hic-1, whose structure has not been solved, contains a 13 residue alanine-rich insertion in another region of the BTB domain that was predicted to be disordered (Deltour et al., 1999). The significance of flexibility of the α3-β4 loop, the variability in the depth and residue composition of the pocket and the Hic-1BTB insertion have not been experimentally clarified, but could be important selectively determinants for binding of corepressors or HDAC’s to
BTB domains. Given that the LRF BTB domain does not interact with the SMRT-BBD, other regions of this BTB domain, such as the charged pocket, could be important for co-repressor interactions.

The LRF<sub>BTB</sub> structure was deposited in the PDB with the accession code 2NN2. While the manuscript was in preparation, a structure of another LRF BTB domain was published (Schubot et al., 2006). The features we observed are consistent with this structure, including missing density in the α3-β4 loop and residue changes in the lateral groove.

3.4.7. LRF BTB domain does not interact with the SMRT-BBD.

To investigate the effect of the extensive sequence differences in the lateral groove of the LRF BTB domain, we tested for binding to the SMRT-BBD (Figure 3.7) via native PAGE. As predicted, the SMRT-BBD fusion protein (Thx-SMRT) did not form a complex with LRF<sub>BTB</sub>. The positive control of BCL<sub>6</sub>BTB did form a complex, as evidenced by the loss of the BCL<sub>6</sub>BTB band and the appearance of a new band due to the complex. Therefore, if LRF<sub>BTB</sub> is to recruit the SMRT co-repressor for transcription repression, it does so at a region distinct from the BBD. This observation suggests that the BBD peptide is specific for the BCL<sub>6</sub>BTB (Polo et al. 2004) and does not directly interfere with the functions of the LRF<sub>BTB</sub>.

As LRF does not interact with the SMRT-BBD, the successful inhibition of the BCL<sub>6</sub>-SMRT interaction (Polo et al. 2004) by peptides mimicking the SMRT-BBD or small molecules is predicted to have no effect on the function of LRF. This would be presumably true as well in any putative BCL<sub>6</sub>BTB / LRF<sub>BTB</sub> heterodimer, since the lateral groove is located at the dimer interface, with contributions from both BTB chains. Therefore, treatment
of tumours expressing high levels of both LRF and BCL6 would require additional targeted therapies that interfere with the interactions and function of LRF.

**FIGURE 3.7.**

Binding of SMRT-BBD to LRF and BCL6 BTB domains. LRF and BCL6 BTB domains were titrated with increasing amounts of Thioredoxin-SMRT (Thx-SMRT) fusion protein and separated by native gel electrophoresis. Lanes 1-3 contained Thx-SMRT alone (2.5, 8 and 13 μg), lanes 4-7 and 8-11 contained 10 μg LRF<sup>BTTB</sup> or BCL6<sup>BTTB</sup>, respectively, and 0, 2.5, 8 or 13 μg of Thx-SMRT. Three minor impurities in the Thx-SMRT sample are indicated with asterixes.
3.5. Conclusions.

The crystal structure of the BTB domain from LRF closely resembles the previously determined structures of domain-swapped BTB homodimers. The extensive dimerization interface is primarily hydrophobic. However, the surface-exposed residues of the LRF BTB homodimer differ from the other BTB structures, as expected from our earlier analysis of the sequences of BTB domains (Stogios et al., 2005). There is little sequence conservation in the lateral groove, a region experimentally shown to interact with co-repressor proteins in BCL6, and as a consequence, the LRF BTB domain does not interact with the SMRT-BBD. As well, there are some changes in the vicinity of the charged pocket of LRF-BTB, a region thought to be important for transcription repression function of other BTB domains. In light of recent research findings showing the significant and central role of LRF in oncogenesis, the structure of the LRF BTB domain will be important for understanding the molecular basis of transcription repression and for the rational design of therapies.
Domain swapping and the BTB domain: insights from structures from FAZF and Miz-1
4.1. Abstract

The BTB domain is a protein-protein interaction motif found in 43 human BTB-zinc finger transcription factors. Previous crystal structures of BTB domains have revealed tightly intertwined homodimers with a strand-exchanged N-terminus. In this chapter, we report the crystal structures of the BTB domains from FAZF and Miz-1. Unlike the PLZF, BCL6 and LRF BTB domains, the FAZF BTB domain was a non-swapped dimer with its N-terminus associated with its own chain. The buried surface in the FAZF\textsuperscript{BTB} dimer was about half as large as that in the domain-swapped dimers. Miz-1\textsuperscript{BTB} resembled a swapped dimer although its N-terminus was truncated. Both BTB domains were dimeric in solution and retain a set of highly conserved residues at the center of the dimerization interface. We also present a phylogeny of the BTB fold that incorporates structural features. This showed that the common ancestor to the BTB-ZF proteins encoded a domain-swapped dimer, while the N-terminal changes seen in FAZF\textsuperscript{BTB} and Miz-1\textsuperscript{BTB} were later evolutionary developments.
4.2. Introduction

4.2.1. Preface

There are 43 BTB-ZF (Bric-à-brac, Tramtrack, Broad-complex-Zinc Finger) proteins in the human genome, representing approximately 3% of all C2H2-type Zinc Finger proteins at Pfam (Bateman et al., 2004), and 24% of all the BTB domain-containing proteins (Stogios et al., 2005). BTB-ZF proteins contain a N-terminal BTB domain and ZF motifs near the C-terminus, with a poorly conserved middle linker region that is predicted to be unstructured.

Structure determination of the BTB domains from PLZF, BCL6 and LRF revealed tightly intertwined domain-swapped homodimers (Ahmad et al., 2003, Ahmad et al., 1998, Stogios et al., 2007, Li et al., 1999, Schubot et al., 2006). Each structure contained the “core BTB fold” (Stogios et al., 2005) plus elements at the N-terminus (α-helix 1 and β-strand 1) that associated with the core fold of the partner chain. The strand-exchanged homodimer (N-terminus is in a domain swapped conformation but there is no evidence of monomeric BTB domain) is thought to be the minimally functional unit of the BTB domain. Truncation of a portion of the N-terminus of PLZFBTB resulted in misfolded protein (Melnick et al., 2000). Two peptides of the SMRT-BBD interacted with BCL6BTB through interactions with β1 and the “lateral groove” region that was formed at the dimer interface (Ahmad et al., 2003). To date, there has been no evidence of a BTB domain with the same core fold but with β1 and/or α1 associated with its own chain or evidence of monomers.

While the structures of these three BTB domains showed remarkable structural similarity, in general, the residues on the solvent-exposed surface were poorly conserved (Stogios et al., 2005). The differential affinity for components of the nuclear co-repressors
and transcription repression potency was likely based on the specific sequence and structural features on the surface of each BTB domain (Stogios et al., 2007).

In order to further our understanding of BTB-ZF proteins, we determined the crystal structures of the BTB domains from FAZF and Miz-1 to 2.0 and 2.6 Å, respectively. Each structure showed unexpected features in the BTB homodimer.

4.2.2. FAZF is evolutionarily and functionally linked with PLZF

The FAZF protein, encoded by the Zbtb32 gene (also known as TZFP, Tang et al., 2001; PLZP, Piazza et al., 2004 and ROG, Omori et al., 2003, Miaw et al., 2000) was initially identified as a sequence homolog of PLZF (Lin et al., 1999) and in a screen for interaction partners of the Fanconi anemia group C protein (FANCC) (Hoatlin et al., 1999). FAZF was implicated in the regulation of T cell activation (reviewed in Bilic & Ellmeier, 2007). FAZF possesses transcription repression activity, although there are conflicting results about whether this function is dependent on the BTB domain (Miaw et al., 2000, Tang et al., 2001). It is also unclear whether repression is dependent on a SMRT/N-CoR-HDAC complex (Melnick et al., 2002, Tang et al., 2001, Omori et al., 2003). There may be direct interactions between FAZF and HDAC2 (Omori et al., 2003).

It has been suggested that the Zbtb32/Fazf and Zbtb16/Plzf genes are evolutionarily linked by syntenic gene duplication (Zhang et al., 1999). FAZF and PLZF share 35% sequence identity in their BTB domains, making them mutual closest sequence homologs. There is 68% identity between the three ZF motifs of FAZF and the last three ZF motifs of PLZF. PLZF<sup>BTB</sup> was shown to interact with FAZF (it was not investigated whether FAZF<sup>BTB</sup> is sufficient for this interaction) (Hoatlin et al., 1999). Consistent with the high de-
gree of sequence similarity in the ZF region, FAZF was thought to bind and regulate many of the same target genes as PLZF, such as *IL3R* (Hoatlin et al., 1999). FAZF has been shown to regulate expression of the genes *Aie1* (Tang et al., 2001), *IL13* (Omori et al., 2003), *CBFA1*, *collagen 1A1*, *osteocalcin*, and *alkaline phosphatase* (Ikeda et al., 2007).

### 4.2.3. Miz-1 is both a transcription activator and repressor

The *Zbtb17* gene encodes the BTB-ZF protein Miz-1. Miz-1 is a regulator of a variety of genes; it is a transcription activator by virtue of its ability to recruit the histone acetyltransferase p300 and a repressor when it forms a complex with c-Myc. p300 and c-Myc compete for interaction with a region between ZF motifs 12 and 13 of Miz-1 (Peukert et al., 1997, Staller et al., 2001). Miz-1 is one of the main targeting factors for c-Myc-mediated transcription repression and the c-Myc-Miz-1 complex was shown to repress the transcription of a variety of genes including the cell cycle-dependent kinase (CDK) inhibitor genes *p15INK4b* and *p21CIP1* (Peukert et al., 1997, Staller et al., 2001, Wu, Cetinkaya et al., 2003).

Miz-1 was shown to be involved in a number of other protein-protein interactions. Miz-1\textsuperscript{BTB} interacted with topoisomerase binding protein IIb (TopBP1) (Herold et al., 2002). Miz-1\textsuperscript{BTB} also interacted with HectH9, an E3 ubiquitin ligase that ubiquitinates c-Myc (Adhikary et al., 2003). The BTB domain and a C-terminal region of Miz-1 interacted with host cell factor 1 (HCF-1) (Piluso et al., 2002). Miz-1 interacted with BCL6 and recruits it to the *p21CIP1* promoter in germinal center B-cells, which lack c-Myc expression (Phan et al., 2005). BCL6\textsuperscript{BTB} was necessary for this interaction, but it was not clear if this interaction occurred via heterodimerization with Miz-1\textsuperscript{BTB}. 
4.3. Methods

4.3.1. Cloning, expression and purification of the FAZF, mutant and Miz-1 BTB domains.

Cloning from cDNA’s encoding full-length FAZF (MGC 21109), and Miz-1 (MGC id 161441) were cloned for expression as thioredoxin-his-tagged fusion proteins by Denise Jaworsky as described in section 3.3.1.

\( \text{FAZF}^{\text{BTB-2C}} \) (FAZF\(^{\text{BTB}}\) S2C, A117C) and PLZF\(^{\text{BTB-2C}}\) (PLZF\(^{\text{BTB}}\) K5C, Q126C) mutants (to be used in the assay to verify domain swapping) were cloned in two steps of single point mutations, using the Quikchange mutagenesis kit (Stratagene). The following primers were used: FAZF S2C forward = cgcggatccatgTGCctgccccccataagactgc, S2C reverse = gcagtcttatggggggcagGCAcatggatccgc; FAZF A117C forward = ggctcgaggggacaggTGTtaaaagcttgcgg, A117C reverse = ccgcaagcttttaACAcctgtcccctcgagcc; PLZF K5C forward = cggatccatgatctgacaTGTatgggcatgatcc, K5C reverse = ggtcgctgggagcccatACAtgtcagatccatggatcc; PLZF Q126C forward = gctggagaccatcTGCgcctcagacgacaatgactaaaagc; Q126C reverse = gcttttagcttcttgtctgaggcGCAgtggtctcagc.

Wild-type FAZF and Miz-1, plus the two mutant FAZF\(^{\text{BTB-2C}}\) and PLZF\(^{\text{BTB-2C}}\) BTB domains were expressed and purified as described for the LRF BTB domain in section 3.3.2. Protein concentration of purified BTB domains were determined by UV spectrophotometry using extinction coefficients of 14060 and 2980 M\(^{-1}\) cm\(^{-1}\) for FAZF (wild-type or mutant) and Miz-1 BTB domains, respectively.
4.3.2. Crystallization of the FAZF and Miz-1 BTB domains.

Diffraction-quality crystals were grown by hanging drop set-up in 24-well Linbro plates with 1 μL BTB domain (FAZF$^{\text{BTB}}$ at 4.5 mg/mL, Miz-1$^{\text{BTB}}$ at 6 mg/mL) with 1 μL reservoir solution. The 600 μL reservoir solutions that produced crystals were the following: FAZF$^{\text{BTB}} = 0.2$ M ammonium sulfate, 0.35 M sodium citrate pH 5.5 and 30% (w/v) PEG 3350; Miz-1$^{\text{BTB}} = 75$ mM zinc acetate pH 4.6). Crystallization of the FAZF BTB domain was performed by Lu Chen.

4.3.3. Structure determination and refinement of the FAZF and Miz-1 BTB domains.

FAZF$^{\text{BTB}}$ crystals were cryoprotected with mother liquor plus 20% glycerol prior to flash freezing in a liquid nitrogen stream at 100 K. Native diffraction data were collected at the Cornell High Energy Synchrotron Source (CHESS) beamline F2, using an ADSC Q210 detector. Data were collected in 1° oscillations for a total of 152°, with the detector distance at 200 mm. Diffraction data were reduced with the HKL2000 package (Table 4.1).

Miz-1$^{\text{BTB}}$ crystals were cryoprotected in mother liquor plus 20% glycerol prior to flash freezing by dunking into liquid nitrogen solution at 100 K (flash freezing into a liquid nitrogen stream dried out the crystals). Crystals were looped such that the needle would stick out of the loop, and diffraction data were collected at a point on the needle outside of the loop (diffraction could not be collected from the needle inside the cryosolution in the loop due to high background). Native diffraction data were collected at the Advanced Photon Source (APS) beamline 19-ID, using an ADSC Q315 detector. Data were collected in 0.5° oscillations for a total of 180°, with the detector distance at 250 mm. Extensive radiation damage
was observed after 100° of diffraction, as measured by an increase in the R_{merge} per frame. Diffraction data were reduced with the HKL2000 package (Table 4.1).

4.3.4. Conformation of N-terminus in solution test via cysteine cross-linking.

FAZF^{BTB-2C} and PLZF^{BTB-2C} mutants were purified as were the wild-type but in the presence of 20 mM β-mercaptoethanol throughout. Aliquots at 1 mg/mL were put over PD-10 columns (GE Healthcare) to exchange into an equivalent buffer without reducing agent and samples were let sit at room temperature for cysteine oxidation. Samples at various timepoints were run on a 14% SDS-PAGE gel, with sample buffer containing or missing reducing agent as appropriate. After 24 h, samples were desalted using C4 ZipTips (Millipore) samples and were submitted for ESI-MS (Hospital For Sick Children Advanced Protein Technology Centre) using a QStar XL mass spectrometer (Applied Biosystems). Data were analyzed using BioAnalyst software.

4.3.5. Molecular weight measurement by sedimentation equilibrium analytical ultracentrifugation.

Three protein concentrations each of FAZF, Miz-1 and PLZF BTB domains ranging from 10 to 150 μM were spun at three different rotor speeds (25000, 30000, 35000 rpm) in an Optima XL-1 analytical ultracentrifuge (Beckman) at room temperature. Curves of A280 versus radius, or ln(A280) versus radius squared where appropriate were fit using the least squares method to the models described in Table 4.2 and Figure 4.15. Partial specific volumes were estimated by the method of Edelstein and Schachman (Edelstein & Schachman,
1967). Ultracentrifugation was performed by Sylvia Ho in the Avi Chakrabartty Lab at UHN.

4.3.6. Molecular weight measurement by analytical size exclusion chromatography.

100 μL samples of wild-type FAZF, PLZF and Miz-1 BTB domains at concentrations between 2 and 200 μM in 300 mM NaCl, 20 mM Tris pH 8.0, 1 mM TCEP were loaded onto a Shodex KW-803 analytical size exclusion column. Values of Abs(280 nm) were measured using a BioCAD HPLC system (Perceptive Biosystems). The column was calibrated with RnaseA (13.7 kDa), chymotrypsin (20.4 kDa), ovalbumin (49.1 kDa), albumin (67.0 kDa), aldolase (158.0 kDa) and catalase (232.0 kDa). Exclusion and total column volumes were measured with blue dextran 2000 and vitamin B12, respectively.

4.3.7. Peptide binding assay.

10 μg of FAZF\textsuperscript{BTB} and Miz-1\textsuperscript{BTB}, and 6 μg BCL6\textsuperscript{BTB} mixed with 0-13 μg of Thioredoxin-SMRT and run on a 5% native PAGE gel at 4°C, with a Tris-glycine pH 8.8 running buffer. The calculated isoelectric points are 6.79, 5.29, 6.3 and 5.4 for FAZF\textsuperscript{BTB}, Miz-1\textsuperscript{BTB}, BCL6\textsuperscript{BTB} and Thioredoxin-SMRT, respectively.

4.3.8. Structure comparisons and visualization.

Structure superpositions and torsion angles were calculated with Coot or SwissPDBViewer (Guex & Peitsch, 1997, Emsley & Cowtan, 2004). Buried surface area was calculated using the program NACCESS with default probe size
4.3.9. Structure characters and phylogeny construction.

Structure-based multiple sequence alignment of 21 BTB fold-containing structures was constructed by superimposing the core BTB fold. The 23 structure characters described in Table 4.3 were identified by visual inspection and treated as standard morphological characters in the Bayesian phylogenetic reconstruction program Mr. Bayes. Mr. Bayes was used to generate phylogenies from sequence only, structure only, and the mixed sequence+structure alignment. Each type of data were treated with an independent gamma distribution of substitution rates. The sequence data were treated with prior substitution rates using the BLOSUM model. Structure data were modeled using the default substitution model for standard characters. Each analysis was run for 100000 generations, which was sufficient for convergence in all runs as measured by values of the average standard deviation of split frequency values between 0.02 and 0.06 and PRSF values between 1.000 and 1.004. Trees were sampled every 50 generations, producing 2000 trees. As recommended in the Mr. Bayes documentation, the first 500 were discarded using the burn in parameter. Alignments and Mr. Bayes input scripts are available upon request. The consensus tree with posterior probability values labeled at each node was visualized with TreeviewX (Page, 1996) and MEGA4 (Tamura et al., 2007). Trees were rooted in the tree visualization software to properly arrange taxa.
4.4. Results and Discussion

4.4.1. Expression and purification of the FAZF, mutant and Miz-1 BTB domains.

FAZF and Miz-1 BTB domains expressed as thioredoxin-6xhis-tagged fusion proteins yielded 80-100 mg per 2 L E. coli culture. After four chromatographic steps (Ni-NTA column, gel filtration, thrombin protease digest, Ni-NTA column, gel filtration) (shown for Miz-1\textsuperscript{BTB} in Figures 4.1 and 4.2), the yield was 20-30 mg purified BTB domain. The mutants FAZF\textsuperscript{BTB-2C} and PLZF\textsuperscript{BTB-2C} had similar expression levels and were purified in the same manner.

\textbf{FIGURE 4.1.}

SDS-PAGE gel showing the purification of Miz-1\textsuperscript{BTB}. This purification is representative of the purification of all BTB domains. Lane 1 = molecular weight markers, 2 = cell lysate high speed spin supernatant, 3 = First Ni-NTA column total elution, 4 = First gel filtration (Superdex 200) major peak fractions, 5 through 8 = thrombin digest time course until 60 hours at room temperature, 9 = Second Ni-column total flow-through, 10 = sample 9 concentrated, 11 = Second gel filtration (Superdex 200) major
peak fractions, 12 = final product. Graphics on the right are references for molecular weights of uncleaved and thrombin cleavage products (thioredoxin-6xHis-Miz-1^{BTB} = 32.9 kDa, thioredoxin-6xHis = 19.8 kDa, Miz-1^{BTB} = 13.0 kDa).

**FIGURE 4.2.**

Gel filtration profiles of Thioredoxin-6xhis-Miz-1^{BTB} and Miz-1^{BTB} run a Superdex S200 column (GE Healthcare). Molecular weight standards (in kDa) are indicated. These profiles are representative of the purification of all Thioredoxin-6xhis-tagged BTB domains.
4.4.2. Crystallization of the FAZF and Miz-1 BTB domains.

The FAZF BTB domain was crystallized by a technician in our lab, Neil Pomroy, using the hanging drop vapour diffusion method.

I crystallized the Miz-1 BTB domain. This BTB domain crystallized as very thin, long needles.

**Figure 4.3.**

*Native crystals of the FAZF BTB domain grown by Neil Pomroy. Scalebar is indicated.*
4.4.3. Structure determination and refinement of the FAZF and Miz-1 BTB domains.

I collected native diffraction data for the FAZF\textsuperscript{BTB} crystal at the CHESS F2 beamline, and native data for the Miz-1\textsuperscript{BTB} crystal at the APS 19-ID beamline. Data collection statistics for both crystals are summarized in Table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>FAZF\textsuperscript{BTB}</th>
<th>Miz-1\textsuperscript{BTB}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2</td>
<td>P2, 2, 2, 1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a, b, c (Å)</td>
<td>49.08, 32.81, 69.43</td>
</tr>
<tr>
<td></td>
<td>α, β, γ (°)</td>
<td>90, 108.2, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0 - 2.0</td>
<td>50.0 - 2.60</td>
</tr>
<tr>
<td>Rsym(%)</td>
<td>6.3 (24.5)</td>
<td>9.5 (33.2)</td>
</tr>
<tr>
<td>I / σI</td>
<td>16.94 (3.77)</td>
<td>10.94 (4.16)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.2 (96.8)</td>
<td>99.1 (100.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.93 (2.39)</td>
<td>3.67 (3.78)</td>
</tr>
</tbody>
</table>
### Refinement

<table>
<thead>
<tr>
<th></th>
<th>FAZF&lt;sup&gt;BTB&lt;/sup&gt;</th>
<th>Miz-1&lt;sup&gt;BTB&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>26.0 - 2.0</td>
<td>14.7 - 2.60</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>13728</td>
<td>7892</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt;(%)</td>
<td>18.8 / 24.8</td>
<td>23.4 / 30.4</td>
</tr>
<tr>
<td>Number of atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1658</td>
<td>1804</td>
</tr>
<tr>
<td>Water</td>
<td>171</td>
<td>45</td>
</tr>
<tr>
<td>Solvent (zinc, acetate)</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>Average B-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>23.8</td>
<td>35.0</td>
</tr>
<tr>
<td>Water</td>
<td>32.3</td>
<td>27.9</td>
</tr>
<tr>
<td>Solvent (zinc, acetate)</td>
<td>N/A</td>
<td>32.1</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.023</td>
<td>0.022</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.868</td>
<td>1.951</td>
</tr>
</tbody>
</table>

Values in brackets correspond to the highest resolution shells of: 2.05 - 2.0 and 2.65 - 2.60 Å for FAZF<sup>BTB</sup> and Miz-1<sup>BTB</sup>, respectively.  

R<sub>sym</sub> = \( \frac{\sum_{hkl}|I(hkl)-<I(hkl)>|}{\sum_{hkl}|I(hkl)|} \),  

R<sub>work</sub> = \( \frac{\sum|F_o-F_c|}{\sum|F_o|} \)

The FAZF<sup>BTB</sup> crystal contained two chains in the asymmetric unit as predicted from a Matthew’s coefficient of 2.07 Å<sup>3</sup>/Da and a solvent content of 41%. Molecular replacement was performed using Phaser. The successful search model was one chain of the PLZF<sup>BTB</sup> dimer (PDB code 1buo, Ahmad et al., 1998), comprising residues 6-126 as a polyalanine model, except where the residues were identical between FAZF and PLZF BTB domains, and all turn and loop regions were omitted from the model. Phaser searched for two copies of the search model and the solution contained two separated protein chains, with one of them forming a rough homodimer after crystallographic symmetry and the other chain forming another homodimer with a chain from a neighbouring asymmetric unit.
As was evident from the final FAZF\textsuperscript{BTB} structure and the altered domain swapping and dimerization interface, refinement from the MR solution was challenging. The FAZF BTB domain was expected to be a domain-swapped homodimer since FAZF\textsuperscript{BTB} is a close sequence homolog of PLZF\textsuperscript{BTB} and that the first BTB domain structures (from PLZF and BCL6) were very similar domain-swapped homodimers. However, the homodimers in the MR solution had some features that were different from the PLZF\textsuperscript{BTB} homodimer, especially the dimerization interface, conformation of $\beta_1$ and the N-terminal residues of $\alpha_1$. Ultimately, the key bottleneck at the early stages of the refinement was rationalizing the positions of $\beta_1$ and the N-terminal residues of $\alpha_1$ in the strand-exchanged conformation of PLZF\textsuperscript{BTB}. Once these regions were removed, the core BTB folds could be built. Refinement completed smoothly until unambiguous electron density at the N-terminus of each chain could later be modeled by $\beta_1$ and $\alpha_1$.

Structure refinement proceeded with an iterative process of manual or automated rebuilding using ARP/wARP, simulated annealing, group and individual b-factor refinement using CNS. The progress of refinement was monitored with drops in the crystallographic and free R-factors, along with inspection of $2F_o-F_c$ and $F_o-F_c$ electron density maps using O. The final rounds of refinement for FAZF\textsuperscript{BTB} were completed using Refmac5, which dropped the R and R-free factors by an additional two percent. Water molecules were built using ARP/wARP.

The final atomic model of FAZF\textsuperscript{BTB} comprised residues 5-113. The structure was validated for ideal geometry (bond lengths, angles, planar peptide bonds, suitable $\Phi$ and $\Psi$ angles/Ramachandran plot, etc.) using Procheck. This indicated 93% of residues fell within
the most favoured regions of the Ramachandran plot, with the other 7% within additionally allowed regions.

The Miz-1\(^{\text{BTB}}\) crystal contained two chains in the asymmetric unit, as predicted from a Matthew’s coefficient of 2.46 Å\(^3\)/Da and a solvent content of 50%. Crystallographic phases for Miz-1\(^{\text{BTB}}\) were determined by molecular replacement using the program EPMR. The successful search model was the coordinates of the BCL6\(^{\text{BTB}}\) dimer (PDB id 1r29, Ahmad et al., 2003), comprising residues 7-128 as a poly-alanine model except where residues were identical between Miz-1 and BCL6 BTB domains. EPMR searched for one dimeric molecule in the asymmetric unit over 500 cycles. The MR solution was refined using positional restrained refinement using Refmac5 using non-crystallographic symmetry (NCS) and TLS restraints at early stages. Model validation and building was completed with Coot (Emsley & Cowtan, 2004). The processes of MR, initial model building and refinement of Miz-1\(^{\text{BTB}}\) were completed with much help from the post-doctoral fellow Jose Antonio Cuesta-Seijo.

The final atomic model of Miz-1\(^{\text{BTB}}\) comprised residues 1-115, with a GS sequence before residue 1 due to the BamHI cloning site (in the PDB file as residues “-1” and “0”). Analysis by Procheck indicated 89% of residues fall within the most favoured regions of the Ramachandran plot, with the other 11% within additionally allowed regions.

**4.4.4. The FAZF BTB domain was not domain-swapped.**

The asymmetric unit of the FAZF\(^{\text{BTB}}\) crystal contained two crystallographically independent chains, each forming symmetric homodimers through crystallographic two-fold axes (Figure 4.5A, B and Table 4.1). The two chains had very similar structures with a
RMSD of 0.65 Å between all Cα atoms. The core fold and the N-terminal extensions, β1 and α1, that are typical of BTB-ZF proteins were clearly evident in both chains. There were no prominent features on the electrostatic surface of the dimer (Figure 4.6).

A striking feature of the FAZF BTB homodimer is that it was not strand-exchanged, in contrast to the BTB dimers from PLZF, BCL6 and LRF. In FAZFBTB, β1 formed an intra-chain two-stranded antiparallel β-sheet with β5, whereas in the other structures β1 and β5 interacted across the two chains (i.e. β1 from chain A forms an antiparallel β-sheet with β5 from chain B). In the PLZF, BCL6 and LRF BTB domains, α1 was 15 to 16 residues long and contributed many hydrophobic interactions in the centre of the dimerization interface. However, in FAZFBTB, α1 consisted of residues 16-24 and was notably shorter due to differences in residues G13, S14 and D15 (Figure 4.5B). The sidechain of residue D15 made extensive contacts, including a charge interaction with the sidechain of R16 from the partner chain and two hydrogen bonds with the backbone of α1 of the same chain. Due to their large deviations from the conformation of their equivalent residues in the other BTB domains (Figures 4.7 and 4.8), we designated this region (residues 13-15 in FAZF; 16-18 in PLZF) as the “hinge loop” sequence, or the sequence that adopts a different conformation in the swapped versus non-swapped states (Liu & Eisenberg, 2002, Bennett et al., 1995).
FIGURE 4.5.

Structures of the FAZF and Miz-1 BTB domains. A) Multiple sequence alignment of the PLZF, Miz-1 and FAZF BTB domains. The N-terminus of Miz-1$^{\text{BTB}}$ is naturally truncated. B, C) Structures of FAZF$^{\text{BTB}}$ and Miz-1$^{\text{BTB}}$. Each chain in the homodimer is separately coloured. The dashed line at the N-terminus of Miz-1$^{\text{BTB}}$ represents the GS sequence as a cloning artifact.
FIGURE 4.6.

Electrostatic surface representations of FAZF and Miz-1 BTB domains. Three views of each BTB domain are shown. Shading scale from -10 to 10 $k_BT$ is indicated.
FIGURE 4.7.

Comparison of the hinge loop region of the FAZF and PLZF BTB domains. The zoomed in views show the N-terminus of the FAZF and PLZF BTB domains, with the hinge loop sequence (residues 13-15 in $FAZFB_{BTB}$ and 16-18 in $PLZFB_{BTB}$) labeled.
FIGURE 4.8.

Definition of the hinge loop sequence from a comparison of the backbone dihedral angles and Cα coordinates of FAZF and PLZFBTB domains. The structures of backbone residues in the hinge loop region of FAZF and PLZFBTB domains are shown, with the hinge loop residues coloured red. The table summarizes comparison of these residues, with the hinge loop sequence coloured red, indicating secondary structure, backbone $\phi$, $\psi$ angles, change in $\phi$, $\psi$ angles ($\Delta\phi$, $\Delta\psi$). Cα RMSD values correspond to distances between chain 1 of FAZF$^{BTB}$ and chain 1 of PLZFBTB, or chain 1 of FAZF$^{BTB}$ and chain 2 of PLZFBTB. Residues with the greatest $\Delta\phi$, $\Delta\psi$ and the crossover point of the Cα coordinates defined the hinge loop sequence.
Figure 4.9.

Averaged main-chain B-factors for FAZF\textsuperscript{BTR}. B-factor values are averaged over all main-chain atoms per residue for both chains in asymmetric unit of FAZF\textsuperscript{BTR}. The hinge loop sequence (residues 13-15) for each chain is coloured red.
FIGURE 4.10.
Simulated annealing omit map of FAZF\textsuperscript{BTB} hinge loop sequence. Omit map was calculated by first removing residues 11-18 from both chains of the final FAZF\textsuperscript{BTB} refined structure, following by simulated annealing with CNS using default parameters, ending with map calculation. Map shown is a $F_o-F_c$ map contoured at 1.0$\sigma$ over the hinge loop sequence (residues 13-15), plus residue 16, of both chains. For reference $C\alpha$ atoms of residues 9-12 and 17-25 are shown.
The hinge loop residues in FAZF were well ordered, have crystallographic B-factors similar to those in other regions of the protein (Figure 4.9), and simulated annealing omit maps showed good electron density (Figure 4.10). Similarly, the equivalent region in PLZF$^{\text{BTB}}$ was well defined in the crystal structure (Ahmad et al., 1998).

4.4.5. The hinge loop region formed interactions unique to the FAZF BTB domain.

The swapped elements in true domain-swapped proteins are found in essentially identical environments in both the swapped and non-swapped states, but the hinge loop may form new interactions due to its conformational change (Liu & Eisenberg, 2002, Bennett et al., 1995). This was true when comparing the swapped and non-swapped BTB domains: both structures contain the $\beta_1$-$\beta_5$ sheet, but the region around the hinge loop provided interactions that were unique to FAZF$^{\text{BTB}}$. The other BTB domains lack the FAZF$^{\text{BTB}}$ D15-R16 salt-bridge or two oppositely charged residues in this region. FAZF$^{\text{BTB}}$ Y12 was displaced by one position in the linear sequence relative to the other BTB domains; in FAZF$^{\text{BTB}}$ this residue packed against its own chain, while in the other BTB domains this residue (a histidine) was well buried within the partner chain. As well, FAZF$^{\text{BTB}}$ had an additional salt-bridge between residues in $\beta_1$ and $\beta_5$ (R7-E84); the other BTB domains did not contain salt-bridges between residues in $\beta_1$ and $\beta_5$.

4.4.6. The FAZF BTB domain dimerization interface.

Single chains from either of the FAZF$^{\text{BTB}}$ dimers superimposed with a RMSD of $\sim$1.3 Å over all C$\alpha$ atoms of the core BTB fold with the single chains of the strand-exchanged BTB domains from PLZF, BCL6 or LRF (i.e. FAZF$^{\text{BTB}}$ residues 16-113 super-
impose with PLZF\textsuperscript{BTB} residues 19-122, Figure 4.11A). Thus the core BTB fold was preserved in FAZF\textsuperscript{BTB}. However, the relative orientation of the two chains (the diad axis) in the FAZF\textsuperscript{BTB} dimer was significantly different relative to the swapped BTB domains. The diad axis of FAZF\textsuperscript{BTB} was rotated by 12° relative to the second chain of the other BTB homodimers (Figure 4.12A), possibly as a result of its non-swapped N-terminus. The lack of strand-exchange and the rotation of the diad axis in FAZF\textsuperscript{BTB} resulted in a dramatic reduction in total buried surface area in the dimerization interface. The PLZF, BCL6 and LRF BTB domains bury between 1647 and 1973 Å\textsuperscript{2}, involving ~42 residues from β1, α1, α2, α3, α4, β5 and α6, while FAZF\textsuperscript{BTB} buries 821 Å\textsuperscript{2}. Only 21 residues participated in dimerization of FAZF\textsuperscript{BTB}, because β1, β5 and α6 do not participate in dimer interactions.
Comparison of single chains structures of BTB domains. A) Superposition of one chain of FAZF and PLZF BTB domains. 

PLZF\textsuperscript{BTB} (PDB id 1buo, Ahmad et al., 1998) is shown as representative of BCL6\textsuperscript{BTB} and LRF\textsuperscript{BTB} as these have previously been shown to have very similar domain-swapped structures (Ahmad et al., 2003, Stogios et al., 2007, Schubot et al., 2006). The core BTB fold is preserved in FAZF\textsuperscript{BTB}, but the main structural is non-swapped conformation of the N-terminus. The FAZF\textsuperscript{BTB} hinge loop sequence is coloured in red.

B) Superposition of one chain of the Miz-1 and PLZF BTB domains. The main structural changes in Miz-1\textsuperscript{BTB} are the shorter N-terminus and the conformation of the “β4” region. Dashed line in Miz-1\textsuperscript{BTB} corresponds to GS sequence as a cloning artifact.

C) Superposition of one chain of Miz-1\textsuperscript{BTB} (this structure), and two chains of Miz-1\textsuperscript{BTB} (PDB code 2q81, Stead et al., 2007), shown as Cα trace and cartoon representations. For clarity, chain A of our Miz-1\textsuperscript{BTB} structure is shown as chain A and chain B superpose with a RMSD of 0.83 Å over all Cα atoms. Also for clarity, only chains A and B of 2q81 are shown: chains A and C both contain the displaced β4 and superpose with a RMSD of 0.46 Å over all Cα atoms, and chains B and D (both containing an intact β2-β3-β4 sheet) superpose with a RMSD of 0.81 Å over all Cα atoms. This superposition shows a rotation in the position of α1 relative to the rest of the BTB fold, and that the path of the displaced “β4” region in our structure resembles the path of the equivalent region in 2q81 chain A.
4.4.7. General description of the Miz-1 BTB domain structure.

Of the 43 BTB-ZF genes in the human genome (Stogios et al., 2005), 11 of these including Miz-1 encode proteins with a truncated N-terminus. Miz-1\textsuperscript{BTB} crystallized with a dimer in the asymmetric unit (Figure 4.5C, Table 4.1). At the single chain level, either chain of the Miz-1\textsuperscript{BTB} dimer superimposed well with the swapped BTB dimers over all C\textsubscript{\textalpha} atoms (Figure 4.11B). Notably, the hinge loop in Miz-1\textsuperscript{BTB} (residues 6-8) followed the same path as the hinge loop residues in PLZF\textsuperscript{BTB} (residues 16-18). At the dimer level, the Miz-1\textsuperscript{BTB} dimer resembled the canonical strand-exchanged BTB dimers, with a rotation of the diad axes between PLZF\textsuperscript{BTB} and Miz-1\textsuperscript{BTB} of only 5° (Figure 4.12A). Despite the truncated N-terminus, the dimer interface in Miz-1\textsuperscript{BTB} was similar to that seen in the PLZF, BCL6 and LRF BTB domains and there was nearly as much surface area burial in the Miz-1\textsuperscript{BTB} dimer interface as in the other BTB domains (1607 Å\textsuperscript{2}).

All of the previously known BTB structures, including FAZF\textsuperscript{BTB}, showed a three-stranded β-sheet at the “top” of the dimer comprised of β2, β3 and β4. However, in Miz-1\textsuperscript{BTB} the residues equivalent to β4 in both chains were displaced relative to the β2 and β3 strands (referred to as the "β4" region; Figures 4.5C and 4.11B).
CHAPTER 4 – STRUCTURES OF THE FAZF AND MIZ-1 BTB DOMAINS

FIGURE 4.12.

Comparison of homodimeric structures of BTB domains. A) Comparison of homodimers from PLZF\textsuperscript{BTB} (PDB id 1buo, Ahmad et al., 1998), FAZF\textsuperscript{BTB}, Miz-1\textsuperscript{BTB} (this work) and Miz-1\textsuperscript{BTB} (PDB id 2q81, Stead et al., 2007). PLZF\textsuperscript{BTB} is representative of the BCL6\textsuperscript{BTB} (Ahmad et al., 2003) and LRF\textsuperscript{BTB} (Stogios et al., 2007) structures, as the angles between the diad axes of these BTB domains are less than 1°. Diad axes are shown in solid black lines, with the dashed lines corresponding to the reference position of the PLZF\textsuperscript{BTB} diad axis. Note that the superpositions were based on single chains (chain A) of each structure, as shown in Figure 3A.
B) Comparison of the dimerization interface of the BTB domains from PLZF, BCL6, LRF, Miz-1 (this work) and FAZF. All BTB domains fully conserve a set of 6 residues at the core of the dimer interface. A ribbon representation of PLZFBTB is shown as representative of the BTB fold with the N-terminal extension, sidechains from each BTB domain shown in stick representation. Sidechains are coloured according to structural conservation, where a 5 corresponds to 5/5 sequence identity and good spatial overlap. Residues scoring a 1/5 are not shown. Multiple sequence alignment: residues in the dimer interface shaded by conservation as in the structure figure (residues scoring 1/5 are black with no shading), residues not involved in either interface coloured light grey. The FAZFBTB hinge loop sequence is boxed.

A structure of Miz-1\textsuperscript{BTB} was reported by another group (Stead \textit{et al.}, 2007) (PDB id 2q81). Stead \textit{et al} describe a Miz-1\textsuperscript{BTB} tetramer involving interactions of the “top” β3-β2-β4 sheets of two dimers. In structure 2q81 the two dimers were labeled AB and CD, and two 5-stranded β-sheets are formed with the topologies β3\textsuperscript{A}-β2\textsuperscript{A}-β4\textsuperscript{D}-β2\textsuperscript{D}-β3\textsuperscript{D} and β3\textsuperscript{B}-β2\textsuperscript{B}-β4\textsuperscript{B}-β2\textsuperscript{C}-β3\textsuperscript{C} (i.e. β4\textsuperscript{A} and β4\textsuperscript{C} are displaced in the A and C sheets by β4\textsuperscript{D} and β4\textsuperscript{B}, respectively). Thus, each 2q81 Miz-1\textsuperscript{BTB} dimer contained one normal β4 and one altered β4, with this β4 instead adopting α-helical conformation (denoted αx in Stead \textit{et al.}, 2007). In contrast, our structure did not contain any dimer-dimer interactions. Interestingly, the displaced “β4” in both chains of our structure closely followed the path of the displaced β4/αx region in chains A and C in the Stead \textit{et al} structure (Figure 4.11C).

Zinc acetate was essential for our Miz-1\textsuperscript{BTB} crystallization and we located six Zn\textsuperscript{2+} atoms coordinated by histidine, aspartate, and/or backbone carbonyl oxygen atoms. Some of
these Zn$^{2+}$ atoms were found at crystal contacts including interactions at the displaced "β4" region, where the sidechains of D55 and D58 and the carbonyl oxygen of V54 coordinate two Zn$^{2+}$ atoms. However, these crystal contacts did not result in Miz-1$^{\text{BTB}}$ tetramers.

In addition to the differences involving the top sheet, the superposition showed that at the single chain level most of the two Miz-1$^{\text{BTB}}$ structures superimposed well, but there was a rotation of $\alpha1$ (Figure 4.11C). At the dimer level, the diad axes of the dimers differ by a large rotation of 20° (Figure 4.12A). Therefore, the two Miz-1$^{\text{BTB}}$ structures had slightly different dimerization interfaces, possibly as a result of their crystal packing environments.

**4.4.8. The dimerization interface of all BTB domains contain a set of conserved residues at the core.**

I investigated the sequence conservation of the dimer interface across all five BTB domain structures. There were six invariant residues at the core of the interface (Figure 4.12B, indicated in red). Residues 20, 21 33 and 51 (PLZF$^{\text{BTB}}$ numbering) were hydrophobic and formed many interactions involving $\alpha1$, $\alpha2$ and $\alpha3$ between the two chains. Notably, these residues were preserved in the non-swapped FAZF$^{\text{BTB}}$ dimer and the Miz-1$^{\text{BTB}}$ dimer interfaces. The next most highly conserved residues (indicated in brown in Figure 4.12B) surrounded the set of the invariant residues and comprised a number of hydrophobic residues and a conserved basic residue in the charged pocket. The long side-chain of this basic residue (R49 in PLZF$^{\text{BTB}}$ and K47 in BCL6$^{\text{BTB}}$) was buried in the dimer interface, with the head group pointing into solvent, and has been shown to play a role in transcription repression of PLZF$^{\text{BTB}}$ and BCL6$^{\text{BTB}}$ (Melnick et al., 2002). The presence of a serine in
FAZF$^{\text{BTB}}$ at this position may be correlated with its weaker transcription repression function (Melnick et al., 2002). Notably, residues in $\beta 1$ and $\beta 5$ were the least well conserved.

This type of distribution of sequence conservation (highest in centre, less on periphery) has been noted as a feature in many protein interfaces (Keskin et al., 2005, Res & Lichtarge, 2005, Bogan & Thorn, 1998). Thus, based on this analysis, the strongest determinants for BTB domain dimerization in the BTB-ZF proteins were most likely the hydrophobic residues in helices $\alpha 1$, $\alpha 2$ and $\alpha 3$, while contacts involving residues from the $\beta 1$ and $\beta 5$ were less important.

4.4.9. Verification of conformation of N-terminus of FAZF and PLZF BTB domains.

To verify the conformation of the N-terminus of FAZF$^{\text{BTB}}$ in solution, I took advantage of the fact that the N- and C-termini of the FAZF and PLZF BTB domains were close together in space. I introduced cysteine residues at the N- and C-termini of both BTB domains (referred to as FAZF$^{\text{BTB-2C}}$ and PLZF$^{\text{BTB-2C}}$, respectively), reasoning that spontaneous disulfide bond formation under oxidizing conditions would result in a cross-linked monomer species if FAZF$^{\text{BTB-2C}}$ is indeed not swapped and a dimeric species PLZF$^{\text{BTB-2C}}$ is swapped. The mutants were purified in reducing conditions, then buffer-exchanged to oxidizing conditions. The cross-linked products were separated via non-reducing SDS-PAGE (Figure 4.13A). As expected, FAZF$^{\text{BTB-2C}}$ ran as a monomer, while PLZF$^{\text{BTB-2C}}$ showed a higher molecular weight species at the size of a dimer (approximately 30 kDa). As verified by electrospray mass spectrometry, complete oxidation occurred in each mutant after 24 h (Figure 4.13B). Therefore, it was concluded that the domain swapping state observed in the
FAZF and PLZF BTB domain crystal structures accurately reflected their conformations in solution.

**FIGURE 4.13 (NEXT PAGE).**

Solution verification of domain-swapped conformations. A) Timepoints during the oxidation process of $\text{FAZF}^{\text{BTB-2C}}$ and $\text{PLZF}^{\text{BTB-2C}}$ were run on a non-reducing SDS-PAGE gel.

B) Electrospray mass spectrometry of $\text{FAZF}^{\text{BTB-2C}}$ and $\text{PLZF}^{\text{BTB-2C}}$ reduced and oxidized (after 24 h) samples. The raw ESI-MS spectra are shown, with charged species labeled. Mass reconstructions with molecular weight (MW) in Da shown as insets, with cartoons showing the corresponding structural species (reduced cysteines represented by SH, disulfide-bond formation represented by S-S). Observed MW match calculated MW from sequence: $\text{FAZF}^{\text{BTB-2C}}$ (reduced) = 12858.7, $\text{FAZF}^{\text{BTB-2C}}$ (oxidized) = 12856.7, $\text{PLZF}^{\text{BTB-2C}}$ (reduced) = 15214.4, $\text{PLZF}^{\text{BTB-2C}}$ (oxidized) = 30424.8.
4.4.10. Oligomeric state of the FAZF and Miz-1 BTB domains.

To corroborate the crystal structures of the FAZF and Miz-1 BTB domains, the apparent molecular weights (MW\text{app}) of the proteins were measured in solution by sedimentation equilibrium ultracentrifugation (Table 4.2 and Figure 4.14). The MW\text{app} of the strand-exchanged PLZF\textsuperscript{BTB} was measured as a comparison, which has previously been shown to be dimeric above 200 nM (Li et al., 1997). The data for each BTB domain was fit in a global manner to a single dimeric species model and also to a monomer-dimer equilibrium model. For each BTB domain, the average residuals for the single species and monomer-dimer fits were almost equal (Table 4.2), and the MW\text{app} values were consistently close to double the sequence molecular weight. Therefore, the monomer-dimer equilibrium models did not explain the measured data any better than a single dimeric species model; at the loading concentrations of these experiments, each BTB domain is exclusively dimeric. I verified this observation by analytical size exclusion chromatography (Figure 4.15) and dynamic light scattering (data not shown), and we did not unambiguously detect monomers down to 2 μM loading concentration.

The suggestion that Miz-1\textsuperscript{BTB} formed tetramers was based on crystal contacts and analytical ultracentrifugation data (Stead et al., 2007). These authors reported a dimerization K\text{d} of 110 μM by sedimentation equilibrium assuming a dimer-tetramer association model. However, their loading concentrations were lower than their K\text{d} (43 and 82 μM) and no other possible models were reported in their fits to the data. They also report velocity sedimentation data showing monodisperse dimers without evidence of tetramers. Since my Miz-1\textsuperscript{BTB} crystal structure did not have the equivalent crystal contacts but was almost exactly the same construct, I analyzed our sedimentation equilibrium centrifugation data in terms of
different association models involving tetrameric species (Table 4.2). A global fit of the data to a dimer-tetramer model could not be completed as the MW$_{\text{app}}$ was less than that of a dimer (i.e. the model was not mathematically possible), and a global fit of the data to a monomer-tetramer model resulted in a higher variance than the single species fit. This was also the case when the data for only the highest loading concentration (355 μM) were fit to a monomer-tetramer equilibrium, at which concentration we expected to have observed tetramers if the K$_d$ of 110 μM was accurate. Overall, no evidence for Miz-1$^{\text{BTB}}$ tetramers was found, even though my studies involved a very similar construct and very similar conditions. Tetramerization of Miz-1$^{\text{BTB}}$ in solution, or any other BTB domain, has not been described in the literature or observed through our routine purifications of the any BTB domain. This does not prove that Miz-1$^{\text{BTB}}$ tetramers never exist biologically, but if they do, the association would be extremely weak.
Table 4.2. Summary of sedimentation equilibrium ultracentrifugation data

<table>
<thead>
<tr>
<th>Model</th>
<th>MW\textsubscript{app} (Da)</th>
<th>MW\textsubscript{app}/MW\textsubscript{seq}</th>
<th>Variance</th>
<th>K\textsubscript{a} (AU)</th>
<th>K\textsubscript{d} *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miz-1\textsuperscript{BTB} model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single species</td>
<td>25469.0 ± 868</td>
<td>1.95 ± 0.07</td>
<td>4.669</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer-Dimer</td>
<td>4.649 344.7</td>
<td>2.1 µM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer-Tetramer</td>
<td>10.346 1.71 8.84 x 10\textsuperscript{-11} M\textsuperscript{3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer-Tetramer (highest concentration only)</td>
<td>8.893 1.73 8.74 x 10\textsuperscript{-11} M\textsuperscript{3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAZF\textsuperscript{BTB} model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single species</td>
<td>24602.0 ± 382</td>
<td>1.92 ± 0.03</td>
<td>1.440</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer-Dimer</td>
<td>1.464 193.0</td>
<td>0.7 µM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLZF\textsuperscript{BTB} model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single species</td>
<td>28198.0 ± 858</td>
<td>1.85 ± 0.05</td>
<td>0.612</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer-Dimer</td>
<td>0.608 2172.6</td>
<td>0.1 µM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(MW_{\text{app}}\) = apparent molecular weight, \(MW_{\text{seq}}\) = sequence molecular weight, \(K_{\text{a}}\) (AU) = dimerization association constant in absorbance units, \(K_{\text{d}}\) (µM) = dimerization disassociation constant in molarity. For each protein, fits were global fits of data at three concentrations and three speeds, except where indicated for of Miz-1\textsuperscript{BTB} at highest concentration only at three speeds. \(MW_{\text{seq}}\) values: Miz-1\textsuperscript{BTB}: 13051.1 Da, FAZF\textsuperscript{BTB}: 12810.7 Da, PLZF\textsuperscript{BTB}: 15264.4 Da. Sample loading concentrations: Miz-1\textsuperscript{BTB}: 75, 182, 355 µM; FAZF\textsuperscript{BTB}: 12, 20, 40 µM; PLZF\textsuperscript{BTB}: 45, 150, 150 µM. * = \(K_{\text{d}}\) values listed should be considered as upper boundaries, as the sample loading concentrations were much higher than the calculated \(K_{\text{d}}\) values.
Sedimentation equilibrium ultracentrifugation data for Miz-1$^{B_{TB}}$ (next page). A) Data plotted as $A_{280}$ vs. $R$ (cm). Plots are separated by the three loading concentrations of 75 μM / 0.9 mg/mL (left), 182 μM / 1.8 mg/mL (center) and 355 μM / 3.6 mg/mL (right). Samples were run at the three speeds 25000 rpm (red), 30000 rpm (blue), and 35000 rpm (green). Solid black lines represent data expected for a single species model (see Table 4.2).

B) Plot of residuals of single-species model fits in A), plotted from $-9\times10^3$ to $9\times10^3$ vs. $R$ (cm). Residuals for each of the nine curves in A) are shown.
FIGURE 4.15 (NEXT TWO PAGES).

Molecular weight measurement by analytical size exclusion chromatography of (A) FAZF, (B) Miz-1 and (C) PLZF BTB domains. Each sample was applied at three concentrations. BTB domain dimers and molecular weight markers are indicated. Note that FAZF BTB was run with a small amount of vitamin B12 as an internal calibration.
Figure C: Absorbance at 280 nm versus elution volume (mL) for different concentrations of PLZF. The peak labeled "PLZF BTB dimer" indicates the elution profile of the BTB domain at various concentrations.
4.4.11. The lateral groove is not conserved: the FAZF and Miz-1 BTB domains do not interact with the SMRT-BBD.

As BTB domains from BTB-ZF proteins have broad functional and obvious structural similarity, I was interested in identifying whether transcription repression function of FAZF and Miz-1 would involve recruitment of the SMRT protein via its BCL6 Binding Domain (BBD). Surface representations of BTB domains clearly showed the presence of the lateral groove, the region that interacts with the BBD (Stogios et al., 2007, Ahmad et al., 2003, Schubot et al., 2006). I noticed the 16 residues that comprised the lateral groove were not conserved across the structures of five BTB domains (Figure 4.16A). Only two positions were conserved across more than three of the five structures and eight positions are not conserved at all. The two residues N21 and H116 that were shown to be critical for SMRT-BBD binding in BCL6$^{\text{BTB}}$ (Ahmad et al., 2003) were among the least well conserved residues in the lateral groove. Importantly, the FAZF and Miz-1 BTB domains lacked β1 in the conformation where the SMRT-BBD forms an interaction with BCL6$^{\text{BTB}}$.

I tested whether the FAZF or Miz-1 BTB domains interacted with the SMRT-BBD. The FAZF, Miz-1 and BCL6 (as a positive control) BTB domains were mixed with increasing amounts of Thioredoxin-tagged SMRT-BBD and resolved via native PAGE (Figure 4.16B). BCL6$^{\text{BTB}}$ clearly formed a complex with Thioredoxin-SMRT, while neither the FAZF nor Miz-1 BTB domains interacted with this peptide by this assay. It should be noted that the $K_d$ of the BCL6$^{\text{BTB}}$-SMRT interaction is 15 μM; any interaction between SMRT and either of the FAZF or Miz-1 BTB domains would have a higher $K_d$ value.

It was unclear whether the lack of β1 in the FAZF and Miz-1 BTB domains was sufficient to result in the observed reduction in the association with the SMRT-BBD. However, we previously observed that LRF$^{\text{BTB}}$, which contained β1 in the domain-swapped con-
formation, was also not able to interact with this peptide (Stogios et al., 2007). Thus, it is likely that the sequence and structure of the lateral groove plays a more important role for SMRT-BBD binding.

**FIGURE 4.16.**

Lateral groove is not conserved, and FAZF\(^{\text{BTB}}\) and Miz-1\(^{\text{BTB}}\) did not interact with SMRT-BBD.

A) Comparison of the lateral groove of PLZF, BCL6, LRF, Miz-1 and FAZF BTB domains. Refer to legend for Figure 3.13C to describe colouring of multiple sequence alignment and struc-
feature representation. Residues N21 and H116 in BCL6$^{\text{BTB}}$, previously shown to be important for SMRT-BBD binding (Ahmad et al., 2003), are boxed.

B) Native PAGE. Lanes 1-3: increasing amounts of Thioredoxin-SMRT alone (Thx-SMRT) (0-13 μg). Lane 4: FAZF$^{\text{BTB}}$ alone (6 μg). Lanes 5-7: same amount of FAZF$^{\text{BTB}}$ mixed with increasing amounts of Thx-SMRT. Lane 8: Miz-1$^{\text{BTB}}$ alone (6 μg). Lanes 9-11: same amount of Miz-1$^{\text{BTB}}$ mixed with increasing amounts of Thx-SMRT. Lane 12: BCL6$^{\text{BTB}}$ alone. Lanes 13-15: positive control for interaction with Thx-SMRT, same amount of BCL6$^{\text{BB}}$ mixed with increasing amounts of Thx-SMRT.


To place the structures of the FAZF and Miz-1 BTB domains in context, we attempted to construct a phylogeny of the BTB fold that reflects the evolution of the various structural features. Previously, using structure-guided sequence alignments alone, we were unable to generate a bootstrap-supported phylogeny that represented the superfamily-level relationships of BTB fold-containing proteins (Chapter 2 and Stogios et al., 2005). The sequence identity between BTB domains from human BTB-ZF proteins are typically in the range of 25-40%, and much lower between the BTB-ZF, T1 and Skp1 subfamilies (Stogios et al., 2005). Facing a similar situation with protein kinase-like sequences, Scheeff et al recently described a method to combine sequence and structure properties to construct robust phylogenies of distant, but clearly related, proteins (Scheeff & Bourne, 2005). These authors supplemented multiple sequence alignments with structural features identified by expert analysis. As in that study, we identified structural features that characterize BTB fold structures, explicitly encoded these as a matrix appended to the sequence alignment, and used
the Bayesian phylogenetic inference program Mr. Bayes (Ronquist & Huelsenbeck, 2003) to construct a phylogeny.

We began with 21 PDB entries of BTB fold-containing proteins, and chose a non-biased set of 23 structural features (Table 4.3) that encode distinct characteristics observed in the members of this set. For example, the features include the strand-exchanged state of the protein, the presence of β1, the conformation of β4, the proportion of hydrophobic residues in the dimer interface and features previously highlighted that differ between the T1, Skp1/Elongin C and BTB-ZF subfamilies (Stogios et al., 2005). Using the structure characters only, a clear division could be detected between these subfamilies (Figure 4.17B).

Table 4.3. 23 structural features that describe the BTB fold were chosen in 21 structures from the PDB.

<table>
<thead>
<tr>
<th>Structure Character</th>
<th>Possible States</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Homodimeric in structure</td>
<td>yes = state 1, no =</td>
<td>Burial of greater than 800 Å², to accommodate the FAZF&lt;sub&gt;BTB&lt;/sub&gt; dimer was used as a cutoff.</td>
</tr>
<tr>
<td></td>
<td>state 0</td>
<td></td>
</tr>
<tr>
<td>2. Homo-tetrameric in structure</td>
<td>yes = 1, no = 0</td>
<td>Only the T1 domain was tetrameric.</td>
</tr>
<tr>
<td>3. Monomeric / heteromeric in structure</td>
<td>yes = 1, no = 0</td>
<td>Skp1 and Elongin C do not self-associate but interact with other proteins in most of their crystal structures.</td>
</tr>
<tr>
<td>4. Length of β1 secondary structure element</td>
<td>0 residues = state 0,</td>
<td>Hydrogen bonding with a partner β-strand was necessary to be deemed a true β1.</td>
</tr>
<tr>
<td></td>
<td>3-5 residues = state 1, 6+ residues = state 2</td>
<td>Miz-1 and HKR3 BTB domains, plus T1, Skp1 and Elongin C lack β1.</td>
</tr>
<tr>
<td>5. Length of α1 secondary structure element</td>
<td>0 residues = state 0,</td>
<td>T1, Skp1 and Elongin C lack α1. α1 is</td>
</tr>
<tr>
<td></td>
<td>&lt;10 residues =</td>
<td>significantly shorter in FAZF&lt;sub&gt;BTB&lt;/sub&gt; relative to</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>State 1, &gt;10 residues</th>
<th>Other BTB domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Presence of β4 in 3-stranded β-sheet</td>
<td>yes = 1, no = 0</td>
<td>Both Miz-1&lt;sup&gt;BTB&lt;/sup&gt; structures contain a displacement of β4.</td>
</tr>
<tr>
<td>7. Domain-swapping state</td>
<td>yes = 1, no = 0, not applicable due to lack of β1, α1 = 2</td>
<td>Only FAZF&lt;sup&gt;BTB&lt;/sup&gt; is not swapped, while T1, Skp1 and Elongin C lack β1, α1 and this criteria is inapplicable.</td>
</tr>
<tr>
<td>8. Nature of structure in β5 region</td>
<td>β-stranded in 2-stranded β-sheet = 1, random coil secondary structure = 0, α-helix conformation = 2</td>
<td>In most structures, β5 exists in 2-stranded β-sheet. In the T1 domain and Elongin C, it does not adopt a fixed structure. In Skp1, this region adopts α-helix.</td>
</tr>
<tr>
<td>9. C-terminal extension</td>
<td>yes = 1, no = 0</td>
<td>Skp1 contains two helices C-terminal to the core BTB fold.</td>
</tr>
<tr>
<td>10. Buried surface area in homodimer</td>
<td>no burial = state 0, burial of 1-500 Å² = state 1, burial of 500-1000 Å² = state 2, burial of 1001+ Å² = state 3</td>
<td>FAZF&lt;sup&gt;BTB&lt;/sup&gt; buries 821 Å², compared to an average of about 1600 Å² in the other BTB domains from the BTB-ZF subfamily. T1, Skp1 and Elongin C are not homodimeric.</td>
</tr>
<tr>
<td>11. Buried surface area in homo-tetramer</td>
<td>burial of 0-500 Å² = state 0, burial of 501+ Å² = state 1</td>
<td>The T1 domains bury 600-900 Å². Every other structure is not homo-tetrameric.</td>
</tr>
</tbody>
</table>
| 12. Presence of Asp + Arg pair of residues at “charged pocket” region at the top of the BTB homodimer | yes = 1, no = 0 | Many of the BTB domains from the BTB-ZF subfamily contain a basic residue pointing straight up from the core of the dimer, with the aliphatic region of the Arg sidechain buried and charged group exposed to solvent. An Asp residue is often
<table>
<thead>
<tr>
<th>No.</th>
<th>Feature Description</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Length of α3-β4 loop</td>
<td>no residues/deletion = state 0, 4 residues with good electron density = state 1, 5+ residues with good electron density = state 2, disordered/missing electron density = state 3, not applicable due to movement of β4 = state 4</td>
</tr>
<tr>
<td>14</td>
<td>Length of α4</td>
<td>11-13 residues = state 1, &lt;11 residues = state 0</td>
</tr>
<tr>
<td>15</td>
<td>Presence of residues N-terminal to β1 in primary sequence</td>
<td>yes = 1, no = 0</td>
</tr>
<tr>
<td>16</td>
<td>Length of α6</td>
<td>10-15 residues = state 1, &lt;10 residues = state 2, distorted = state 0</td>
</tr>
<tr>
<td>17</td>
<td>Presence of salt-bridges between sidechains in β1 and β5</td>
<td>yes = 1, no = 0</td>
</tr>
</tbody>
</table>

found associated with the Arg. These residues have previously been thought to be important for transcription repression. This region forms part of the “charged pocket” region and commonly shows variability between structures. The loop is deleted in FAZF<sup>BTB</sup>, but longer and with good electron density in BCL6<sup>BTB</sup> and the T1 domain. LRF, Bach1, GAN, BTBD6 BTB domains all show disordered sequence here. This region is moved in Miz-1<sup>BTB</sup>. This helix is significantly shorter in our Miz-1<sup>BTB</sup> structure and the HKR3<sup>BTB</sup> structure. Some BTB domains are found at the extreme N-terminus of proteins, others are not. This helix is significantly shorter in the GAN structure and distorted in the Bach1 structure. These interactions, seen in only FAZF, Miz-1 and Bach1 BTB domains, could provide additional stability for the con-
<table>
<thead>
<tr>
<th>18. Hinge loop conformation</th>
<th>α-helical = state 0, not helical = state 1, moved = state 2, not applicable due to missing $\alpha_1$ = state 3</th>
<th>formation of the N-terminus. Only FAZF\textsuperscript{BTB} shows a non-helical conformation here. This region is moved away from $\alpha_1$ in HKR3\textsuperscript{BTB}, and missing in BTBD6, Skp1, Elongin C and T1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>19. Proportion of hydrophobic residues in the dimer interface</td>
<td>47-53% = state 1, 60-65% = state 2, 100% = state 3, not applicable = state 0</td>
<td>LRF, FAZF, Miz-1 and GAN BTB domains have higher proportion of hydrophobic residues in their dimer interfaces and are state 2. A pseudo-dimer interface in BTBD6 could be generated by symmetry operations, is not extensive, and 9/9 residues here are hydrophobic. T1, Skp1 and Elongin C do not have a dimer interface.</td>
</tr>
<tr>
<td>20. Salt bridges in dimer interface</td>
<td>yes = 1, no = 0</td>
<td>FAZF, Miz-1 and HKR3 each have one salt-bridge in their dimer interface. T1, Skp1 and Elongin C do not have a dimer interface.</td>
</tr>
<tr>
<td>21. Missing electron density for $\alpha_1$-$\beta_2$ loop</td>
<td>yes = 1, no = 0</td>
<td>Only GAN\textsuperscript{BTB} structure shows missing density here.</td>
</tr>
<tr>
<td>22. Conformation of residue equivalent to Miz-1\textsuperscript{BTB} Tyr48</td>
<td>superposes with Miz-1\textsuperscript{BTB} Tyr48 = state 0, different residue type or conformation = state 1</td>
<td>In both Miz-1\textsuperscript{BTB} structures, this residue forms a hydrogen bond with the backbone of the displaced $\beta_4$. Some BTB domains have a Tyr residue in the same conformation, others do not.</td>
</tr>
<tr>
<td>23. Interaction partner of aromatic sidechain in $\alpha_1$</td>
<td>interacts with partner chain = state 1, interact with own chain = state 1</td>
<td>This conserved residue interacts with its own chain only FAZF\textsuperscript{BTB}. BTBD6, T1, Skp1 and Elongin C lack $\alpha_1$.</td>
</tr>
</tbody>
</table>
However, the structure-only tree was not well resolved at the level of individual proteins, especially in the BTB-ZF subfamily.

Using both sequence and structural characters, a statistically robust phylogeny could be produced (Figure 4.17A). This phylogeny showed that the structural characters were useful to clearly demarcate the T1, Skp1/Elongin C and BTB-ZF subfamilies, while sequence similarity allowed resolution between structures within subfamilies. The T1 and Skp1/Elongin C structures lacked \( \beta \)1 and \( \alpha \)1, separating them from the BTB-ZF clade. In turn, these clades are separated by the fact that the T1 domain forms tetramers and Skp1 contained a C-terminal extension of two helices. BTBD6 (PDB code 2kvp) appears to represent a transitory form of the BTB domain between the T1 and BTB-ZF clades; it contained only the core BTB fold and thus resembles T1, but it is not tetrameric in its crystal structure.

Each structure of BTB domains from BTB-ZF, BTB-bZip and the BBK proteins solved to date are domain-swapped dimers; the phylogeny suggested the last common ancestor of the BTB-ZF subfamily was dimeric with a N-terminal extension to the core BTB fold in a domain-swapped conformation. Conversely, the phylogeny suggested the non-swapped conformation of FAZF\( ^{\text{BTB}} \) was a late evolutionary development that is unique to this protein. The gene ancestral to \( FAZF \) and \( PLZF \) likely underwent a duplication, as was suggested previously (Hoatlin et al., 1999, Zhang et al., 1999) and supported by our phylogeny. We propose that this FAZF ancestor was strand-exchanged as in PLZF, and mutations in the N-terminus and hinge loop region resulted in the altered non-exchanged conformation (i.e. FAZF “unswapped” its N-terminus). Concurrently, FAZF\( ^{\text{BTB}} \) retained
the set of residues in the core of the dimer interface that kept the dimer intact. Due to the few functional studies of FAZF, it is unclear whether the acquired mutations in this BTB domain were the result of positive selection for a novel function provided by the lack of domain swapping or relaxed selection due to functional overlap between the FAZF and PLZF BTB domains.

Similarly, the placement of Miz-1\textsuperscript{\textit{BTB}} in the phylogeny close to HKR3\textsuperscript{\textit{BTB}}, which also lacks β1, suggests truncation at the N-terminus was a late event that occurred in the ancestor to the Miz-1 and HKR3 BTB domains. These BTB domains also retain the residues at the core of the dimer interface that was sufficient to maintain the integrity of these homodimers. As in the case of the FAZF\textsuperscript{\textit{BTB}} structural alterations, the functional significance of the N-terminal truncation is unclear.

\textbf{Figure 4.17 (Next page).}

\textit{Proposed phylogenies of BTB fold-containing proteins. Structures are labeled by their PDB id’s, species (two letter abbreviations, Ac = Aplysia californica, Hs = Homo sapiens, Rn = Rattus norvegicus, Sc = Saccharomyces cerevisiae) and protein name. The subjects of this work, FAZF\textsuperscript{\textit{BTB}} and Miz-1\textsuperscript{\textit{BTB}}, are coloured green. Nodes are labeled with posterior probabilities. Trees were routed a priori at the position indicated by the two asterisks.}
4.5. Conclusions

In this study, we reported further insights into the structure and function of the BTB domain with the crystal structures of FAZF$^{\text{BTB}}$ and Miz-1$^{\text{BTB}}$. These structures both revealed the homodimer BTB fold but each contained conformational changes. FAZF$^{\text{BTB}}$ was the first demonstration of a non-swapped BTB homodimer. Miz-1$^{\text{BTB}}$ resembled a domain-swapped dimer even though it lacked the \( \beta_1 \) element and also showed a displacement of ten residues corresponding to \( \beta_4 \). Even with these movements, the FAZF and Miz-1 BTB domains retained highly conserved residues in the core of the dimer interface and these BTB domains were homodimeric in solution. In contrast, the N-terminus was not conserved and appeared to play a minor role in dimerization. Lastly, neither BTB domain interacted with the SMRT-BBD, reinforcing the idea that the sequence and conformation of the N-terminus and the lateral groove play important roles in co-repressor interactions of BTB domains.
Chapter 5

Crystal structure of the Kaiso BTB domain and a possible mode of oligomerization of BTB domains
5.1. Abstract

The BTB domain provides a scaffold for a variety of protein-protein interactions, including dimerization, oligomerization, and interactions with nuclear co-repressor proteins. In this chapter, we showed that the crystal structure of the BTB domain from Kaiso, a bimodal DNA binding, genome-wide transcription repressor, suggested a possible mode for the oligomerization of BTB domains. Intriguingly, this structure closely resembled the oligomeric structure of the BTB domains from PLZF (in two crystal forms), the BTB-leucine zipper protein Bach1 and the BBK protein Gigaxonin. The oligomeric structure is compatible with formation of higher order oligomers that could spread over large distances.
5.2. Introduction

5.2.1. Kaiso is a dual-specificity transcription repressor.

Kaiso was first identified as an interaction partner of p120 catenin (Daniel & Reynolds, 1999). This interaction occurred in a non-contiguous sequence around Kaiso’s three C2H2 ZF motifs and p120 catenin’s Armadillo repeats 1 to 7, which also interacted with E-cadherin in a mutually exclusive manner (Daniel & Reynolds, 1999). p120 catenin relieved Kaiso-mediated transcription repression, which was dependent on functioning nuclear localization sequences in both proteins (Kim et al., 2004, Daniel & Reynolds, 1999, Daniel et al., 2002, Kelly, Spring et al., 2004, Kelly, Otchere et al., 2004, Park et al., 2005, Spring et al., 2005). Kaiso also interacted with δ-catenin, a brain-specific p120 homolog (Rodova et al., 2004). Kaiso expression and localization showed dependence on growth conditions, cell and tissue microenvironment (Daniel et al., 2001, Daniel & Reynolds, 1999, Soubry et al., 2005). Kaiso knockout mice showed some resistance to intestinal tumours, indicating Kaiso played a role in tumourigenesis (Prokhortchouk et al., 2006).

Kaiso is interesting among BTB-ZF proteins in that it has bi-model DNA-binding activity. The ZF motifs have been shown to interact with at least two consecutive methylated CpG nucleotides (mCpG), found in the S100A4/metastasin, Rb, Xist, E-cadherin, Pgk, MTA2 and tyrosine hydroxylase genes (Prokhortchouk et al., 2001, Daniel et al., 2002, Yoon et al., 2003, Aranyi et al., 2005) or in a sequence-specific manner with the CTGCNA sequence/Kaiso binding site (KBS), found in the MMP-7/matrilysin, Rapsyn, xWnt11, PPAR-δ, Siamois, c-Fos, c-Myc, Cyclin-D1 and the beta-globin genes (Daniel et al., 2002, Park et al., 2005,
Kim et al., 2004, Rodova et al., 2004, Defossez et al., 2005). Both the BTB and ZF domains are required for repression of promoters containing multiple dispersed copies of mCpG’s, with the ZF motifs mediating direct DNA binding. Kaiso appears to have greater affinity for the KBS rather than mCpG’s, but the \textit{in vivo} relative importance of the different binding modes is unknown (Daniel et al., 2002).

Like many other BTB-ZF proteins, Kaiso has been shown to be a transcription repressor, with this activity dependent on the BTB domain (Kim, Fang et al., 2002, Prokhortchouk et al., 2001). Kaiso’s BTB domain interacted with the RD1 region of N-CoR (nuclear corepressor), forming a large nuclear repression complex that contains histone deacetylase activity (Yoon et al., 2003). In fact, it was shown that the majority of purified N-CoR-containing complexes from HeLa cells contained Kaiso. Interestingly, Kaiso was not found to interact with SMRT (silencing mediator of retinoic acid), which is related with N-CoR. Kaiso has also been shown to interact with CTCF (Defossez et al., 2005), an enhancer blocker protein that plays essential roles in global transcription regulation (Dunn & Davie, 2003). This direct interaction occurred via the BTB domain of Kaiso and a C-terminal region of CTCF (Defossez et al., 2005).

\textbf{5.2.2. Crystal structure of the Kaiso BTB domain suggested a basis for oligomerization of BTB-ZF proteins.}

Biochemical experiments by various groups have suggested oligomerization by the BTB domain and formation of high-molecular weight (HMW) complexes on DNA are important components of the function of BTB-ZF proteins. PLZF has been shown to form BTB domain-dependent HMW complexes (Ball et al., 1999). The BTB domain of Hic-1 was
necessary for optimal and cooperative binding to five copies of its DNA binding site and the BTB domain inhibited binding to a single site (Pinte et al., 2004). The BTB domain from the *Drosophila melanogaster* BTB-ZF protein GAGA factor mediated cooperative DNA binding and inhibits binding to a single site (Katsani et al., 1999). GAGA\textsuperscript{BTB} also mediates DNA bending and formation of large complexes on DNA that were visualized with electron microscopy. The Kaiso BTB domain has been shown to self-associate, likely in a dimeric fashion (Daniel & Reynolds, 1999, Kim, Fang et al., 2002), but it is unknown if Kaiso forms higher order oligomers. The Kaiso BTB domain may also inhibit binding to single copies of the Kaiso binding site (Daniel et al., 2002).

The focus of this study was to identify the structural basis for the oligomerization of the BTB domain from a human BTB-ZF protein. A possible mode of oligomerization was revealed in the structure of the Kaiso BTB domain. The structure showed the typical BTB fold as a quasi-domain swapped homodimer that was very similar to previously determined structures. Notably, each Kaiso\textsuperscript{BTB} homodimer formed interactions with other homodimers in the crystal via formation of interdimer β-sheets.
5.3. Methods

5.3.1. Cloning, expression and purification of the Kaiso BTB domain.

Cloning of the Kaiso BTB domain (residues 1-122) was completed into a pET-32(a) (Novagen) T7-expression vector in parallel with the cloning of the other BTB domains studied in this thesis, as described in section 3.3. Cloning was completed by Denise Jaworksy.

The Kaiso BTB domain was purified using the standard purification method of BTB domains as described in section 3.3. Protein concentration of purified Kaiso BTB domain was determined by UV spectrophotometry using an extinction coefficient of 5960 M⁻¹ cm⁻¹.

5.3.2. Crystallization of the Kaiso BTB domain.

Diffraction-quality crystals were grown by hanging drop set up in 24-well Linbro plates with 1 µL Kaiso BTB domain at 5.05 mg/mL with 1 µL reservoir solution. The volume of reservoir solution that produced crystals was 600 µL of 0.25 M sodium sulfate, 0.1 M sodium acetate pH 4.3, 13% (w/v) PEG 3350. The Kaiso BTB domain could also be crystallized with a reservoir solution of 0.1 M imidazole pH 8.0, 0.2 M lithium sulfate and 20% (w/v) PEG 6000, but this crystal formed in the same space group and unit cell as the first condition and therefore diffraction data was not collected. Crystallization of the Kaiso BTB domain was performed by Lu Chen.
5.3.3. Structure determination and refinement of the Kaiso BTB domain.

Kaiso BTB domain (pH 4.3 condition) crystals were cryoprotected with 15% ethylene glycol prior to flash freezing in a liquid nitrogen stream to 100 K. Native diffraction data were collected at the National Synchrotron Light Source (NSLS) beamline X8C, using a ADSC Quantum 4R detector. Data were collected using 20 second exposures, 0.5° oscillation width over 180° in total with a crystal-detector distance of 150 mm. Diffraction data were reduced with the HKL package, SCALEPACK and DENZO (Table 5.1).

5.3.4. Analytical size exclusion chromatography.

A 100 µL sample of Kaiso BTB domain at 370 µM in 300 mM NaCl, 20 mM Tris pH 8.0, 1 mM TCEP was loaded onto a Shodex KW-803 analytical size exclusion column at room temperature. Values of Abs(280 nm) were measured using a BioCAD HPLC system (Perceptive Biosystems). The column was calibrated with RnaseA (13.7 kDa), chymotrypsin (20.4 kDa), ovalbumin (49.1 kDa), albumin (67.0 kDa), aldolase (158.0 kDa) and catalase (232.0 kDa). Exclusion and total column volumes were measured with blue dextran 2000 and vitamin B12, respectively.
5.4. Results

5.4.1. Purification of the Kaiso BTB domain.

Purifications of Kaiso\textsuperscript{BTB} followed the standard protocol for purification of BTB domains, as described in section 3.3. A single 2 L E. coli culture yielded 30-50 mg of the thioredoxin-6xhis-tagged Kaiso BTB domain fusion protein. After the four chromatographic steps, the final yield for the purified Kaiso\textsuperscript{BTB} was approximately 6-12 mg. The SDS-PAGE gel and gel filtration profiles (Figures 4.2 and 4.3) showing a purification of Miz-1\textsuperscript{BTB} is also representative of purifications of Kaiso\textsuperscript{BTB}.

5.4.2. Size exclusion chromatography of the Kaiso BTB domain.

To identify the oligomeric state of Kaiso\textsuperscript{BTB} in solution, the protein at high concentration (370 \(\mu\)M, or 5 mg/mL) was applied to a Shodex analytical size exclusion chromatography column (Figure 5.1). The protein eluted as a dimer, with no evidence of a higher order molecular weight species. The protein was also applied to a Superdex S200 column at 4 \(^\circ\)C, which also did not reveal species other than a dimer (data not shown).
FIGURE 5.1.

Analytical size exclusion chromatography of Kaiso\textsuperscript{BTB} on a 15 mL Shodex column. BTB domain dimer and molecular weight markers are indicated. Results at room temperature or 4°C were identical.

5.4.3. Crystallization of the Kaiso BTB domain.

Kaiso\textsuperscript{BTB} readily crystallized by hanging drop vapor diffusion (Figure 5.2).
5.4.4. Structure determination and refinement of the Kaiso BTB domain.

Data collection for the Kaiso$^{\text{BTB}}$ crystal at the NSLS beamline X8C is summarized in Table 5.1.

Table 5.1. Data collection and refinement statistics for Kaiso$^{\text{BTB}}$ crystal.

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P6$_2$22</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>117.21, 117.21, 38.41</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00-2.05</td>
</tr>
<tr>
<td>R$_{sym}$ (%)</td>
<td>6.5 (34.7)</td>
</tr>
<tr>
<td>I / σI</td>
<td>51.8 (8.43)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.6 (97.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>20.95</td>
</tr>
</tbody>
</table>

| Refinement          |                  |
| Resolution (Å)      | 50.00-2.05       |
### Chapter 5 – Structure of the Kaiso BTB Domain

<table>
<thead>
<tr>
<th>Number of reflections</th>
<th>9611</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{work} / R_{free} (%)$</td>
<td>17.9 / 23.6</td>
</tr>
<tr>
<td>Number of atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>885</td>
</tr>
<tr>
<td>Water</td>
<td>100</td>
</tr>
<tr>
<td>Solvent ($SO_4^{2-}$)</td>
<td>2</td>
</tr>
<tr>
<td>Average $B$-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>29.2</td>
</tr>
<tr>
<td>Water</td>
<td>39.7</td>
</tr>
<tr>
<td>Solvent ($SO_4^{2-}$)</td>
<td>37.9</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths ($\AA$)</td>
<td>0.023</td>
</tr>
<tr>
<td>Bond angles ($^\circ$)</td>
<td>1.904</td>
</tr>
</tbody>
</table>

Values in brackets correspond to the highest resolution shell of 2.12 – 2.05 $\AA$.  $R_{sym} = \sum_{hkl} |I(hkl)|<I(hkl)> |\Sigma_{hkl}|I(hkl)|$.

The Kaiso$^{BTB}$ crystal contained one chain in the asymmetric unit, as suggested from a Matthew’s coefficient of 2.85 $A^3$/Da and a solvent content of 57%. Phaser easily obtained a molecular replacement solution using a search model of one chain of the PLZF$^{BTB}$ dimer (PDB id 1BUO, Ahmad et al., 1998), comprising residues 16-123 as a poly-alanine chain except where residues were identical between Kaiso and PLZF, and all turn and loop regions were omitted. The initial search solution was used to initiate rounds of automated model building and refinement using ARP/wARP (Perrakis et al., 2001) and Refmac5 (Murshudov et al., 1997), resulting in final $R_{cryst}$ and $R_{free}$ values of 17.9 and 23.6, respectively. $2F_o-F_c$ and $F_o-F_c$ maps were generated by Refmac5/fft and models were visualized with O (Jones et al., 1991) and Coot (Emsley & Cowtan, 2004). The final atomic model comprises residues 3-117 of Kaiso, with missing electron density for part of the Arg4 sidechain. Analysis by Procheck indicated 89% and 10% of non-glycine and prolines residues fall within the most favoured and additionally allowed regions of the Ramachandran plot, respectively, and one residue lies in a generously allowed region.
5.4.5. General description of the Kaiso BTB domain structure.

Kaiso$^{\text{BTB}}$ crystallized with a single chain in the asymmetric unit, therefore crystallographic symmetry operations were necessary to form larger complexes, including the homodimer. The homodimer showed the topology of the core BTB fold plus the N-terminal extension elements $\beta_1$ and $\alpha_1$ (Figure 5.3). Dimerization occurred through an extensive interface comprised of 42 residues (28 non-polar) that buries 1757 Å$^2$ total surface area. Like the structures of the BTB domains from PLZF, BCL6, LRF and Miz-1, but different from FAZF, Kaiso$^{\text{BTB}}$ was domain-swapped.

![Figure 5.3](image)

**Figure 5.3.**

*Structure of the Kaiso$^{\text{BTB}}$ homodimer in cartoon representation. Each chain is coloured in a different shade of red. N- and C-termini of the two chains are labeled.*

Interestingly, an electrostatic surface representation showed a charge asymmetry between the faces of the homodimer that lead to half of the dimer face showing charged fea-
tures, and the other half as hydrophobic (Figure 5.4). The charged half contains a patch of positively-charged amino acids, and a groove of negatively-charged residues that extends around to the “side” of the dimer. The hydrophobic surface was likely dominated by the fact that there are two aromatic sidechains exposed to solvent, Y56 and F72.

Figure 5.4.

_Electrostatic surface representation of the KaisoBTB homodimer, calculated by GRASP2. Red is negative, white is neutral, blue is positively charged. Surfaces were contoured between -10 and +10 kBТ/e and -10 kBТ/e where kB is the Boltzmann constant, T is temperature and e is the electronic charge._
5.4.6. The Kaiso BTB domain dimer-dimer interaction.

The Kaiso\textsuperscript{BTB} crystal had a high degree of symmetry, resulting in a large number of protein chains in the unit cell. Each homodimer formed a $\beta$-sheet with two other homodimers, a phenomenon that propagated through the crystal (Figure 5.5A). This “interdimer” $\beta$-sheet involved the $\beta1$-$\beta5$ sheet from each dimer extending into a four-stranded $\beta$-sheet. The length of $\beta1$ was longer than the other structures (about three residues) and its electron density was clear (Figure 5.5B). A longer $\beta1$ was probably stabilized by interactions between dimers. The interdimer $\beta1$-$\beta5$ $\beta$-sheet was less extensive (four hydrogen bonds between backbone atoms) than the intradimer one (six hydrogen bonds between backbone atoms) (Figure 5.5C). As well, some sidechains of residues in $\beta1$, $\beta5$, $\alpha5$ and $\alpha6$ contributed to the oligomerization interface (Figure 5.5D). The oligomerization interface buried 465 \AA$^2$, comprised of 13 total residues, 8 of which were hydrophobic in nature. L116, I7, A9, and K5 were the most extensively buried residues in the interface, with 99, 90, 88 and 83% of their surface area buried. The I7 residues from the two dimers interacted with each other at the core of the interface. K5 and A9 flanked the I7 residues, with the C$\beta$ through C$\epsilon$ atoms of K5 interacting with the C$\alpha$ of A9 from the opposite dimer. The interface is closed at either side by L116 and F112 interacting with A9. Other residues contributed sidechain burial outside of the central hydrophobic burial of I7-I7, such as L98 and V118. While the interface buried a modest surface area, each dimer oligomerizes equivalently on each side, leading to a substantial combined burial of 930 \AA$^2$ buried of 26 residues per dimer.
The Kaiso\textsuperscript{BTB} dimer-dimer interface. A) Cartoon representation of two dimers. Two chains in one dimer are coloured in two shades of red, two chains in the second dimer are coloured in two shades of blue. The region involved in dimer-dimer interactions, \( \beta 1-\beta 5 \) “interdimer” sheet, is circled. B) Electron density \((2F_o-F_c)\) map contoured at 1.9\(\sigma\), showing part of the \( \beta 1-\beta 5 \) interdimer \( \beta \)-sheet. N-terminus is labeled. C) Zoomed-in view of the dimer-dimer interface. Only the backbone atoms are shown in stick representation. Dashes indicate hydrogen bonds, coloured green for “intradimer” \( \beta 1-\beta 5 \) sheet, magenta for the interdimer \( \beta 1-\beta 5 \) sheet. C) Sidechains involved in dimer-dimer interaction, no backbone atoms are shown. Residues are coloured by atom type (grey=carbon, blue=nitrogen, red=oxygen).
5.4.7. Interactions between oligomers

As another layer of complexity, the crystal structure showed multiple chains of interacting dimers that wind around each other (Figure 5.6). The dimensions of this higher order structure were approximately 70 x 45 Å. These chains wind around each other in a highly polar interface, involving 14 total residues per Kaiso\textsuperscript{BTB} chain, including nine charged and two polar residues. There was a substantial 588 Å\textsuperscript{2} buried in each interface between oligomers, including four salt bridges. The C-termini of the BTB domains were positioned such that the remainder of the full-length protein (linker plus zinc finger motifs) would extend away from the structure.

An electrostatic surface representation of the Kaiso\textsuperscript{BTB} oligomer showed three noteworthy features (Figure 5.7). First, a 25 Å deep positively charged pocket was prominently displayed on the outer surface of the higher order oligomer structure, perpendicular to its propagation axis. The pocket was formed between two dimers that were found on different dimer-dimer chains but was comprised of the equivalent residues of Kaiso\textsuperscript{BTB}. Secondly, there was a negatively charged surface close to the basic surface, but on the opposite face of the oligomer. This surface was comprised mostly from a single chain of Kaiso\textsuperscript{BTB}, but some of the equivalent residues on the next dimer could be found in this surface. Lastly, there was a noticeable hydrophobic face of the oligomer.

Interestingly, the interactions between dimers in the oligomeric structure occurred on the charged face of the Kaiso\textsuperscript{BTB} homodimer, leaving the hydrophobic face prominently exposed.
FIGURE 5.6 (ABOVE LEFT).

Higher order structure of Kaiso\textsuperscript{BTB}. Two chains of interacting dimers are shown, in two views. One chain of dimers shown in cartoon representation, each dimer coloured differently. Second chain of dimers shown as transparent surface representation, coloured grey. Arrows indicate path of remainder of full-length Kaiso from C-termini of BTB domains. The two chains of interacting dimers wind around each other and extend for the length of the crystal. Scalebar is shown.

FIGURE 5.7 (ABOVE RIGHT).

Electrostatic representation of the Kaiso\textsuperscript{BTB} oligomer. Shown are surface representations of two chains of interacting dimers, coloured by electrostatic charge. Three different surface features are labeled: positively charged pocket, negatively charged groove and hydrophobic face.
5.4.8. The Kaiso BTB domain dimer-dimer interaction is seen in other BTB domain structures.

Strikingly, the Kaiso\textsuperscript{BTB} oligomer was very similar to the PLZF\textsuperscript{BTB} oligomer that has been previously solved from two crystal forms. C\textalpha{} atoms from two interacting dimers of Kaiso\textsuperscript{BTB} superimposed with two interacting dimers of PLZF\textsuperscript{BTB} with a RMSD of 1.7 Å (Figure 5.8A). All three structures showed successive homodimers interacting via the intermolecular $\beta$1-$\beta$5 sheet propagating along the length of the crystal. Kaiso\textsuperscript{BTB} and one structure of PLZF\textsuperscript{BTB} showed the phenomenon of two chains of interacting dimers. Due to the imperfect superposition of the BTB domains, the axis and path of propagation of the chains varied (Figure 5.8B). Nonetheless, both Kaiso\textsuperscript{BTB} and PLZF\textsuperscript{BTB} utilized the same interaction surface to form oligomers, and the overall resemblance in the appearance of the oligomers was striking.

The recently determined structures of the BTB domains from the BTB-bZip protein Bach1 (unpublished, PDB id 2ihc) and the BBK protein Gigaxonin (unpublished, PDB id 2ppi) showed arrangements of dimers in the crystal packing that was reminiscent of the crystal packing in the Kaiso and PLZF BTB domain structures (Figure 5.8B). Bach1\textsuperscript{BTB} contained two homodimers in the asymmetric unit, and crystallographic symmetry operations generated an oligomeric structure. Overall, the oligomeric structure was similar to the Kaiso and PLZF oligomers, but there are differences in the contacts between dimers. The two dimers in the asymmetric unit (ASU) were more tightly associated than in Kaiso or PLZF, and the dimers across successive ASU’s were less tightly associated.

Gigaxonin\textsuperscript{BTB} showed interactions across dimers, including the $\beta$1-$\beta$1 H-bonding, and packing of the N-terminal tyrosine residue into the space between $\beta$1 and $\alpha$6 from an-
other dimer (Figure 5.8B). The \( \beta_1 \) sheet in this structure was very well ordered and was quite long (8 residues compared with 7, 5 and 5 in Kaiso, PLZF and Bach1, respectively). Interestingly, there were more hydrogen bonds in the interactions between \( \beta_1 \) elements from different dimers in this structure relative to the other BTB domains: there were 8 main-chain H-bonds compared with 4 H-bonds in the Kaiso and PLZF BTB domain structures. The angle between dimers differed from the other BTB-BTB interactions, resulting in a different axis of propagation of the oligomer.

While there were differences in the number of residues involved in the dimer-dimer interaction and the axis of propagation of the oligomer, it is striking that five crystal forms of four different BTB domains formed very closely related oligomeric structures, all involving interactions between \( \beta_1 \) sheets of the homodimer.

Since multiple BTB domains showed similar oligomer structures, we looked for conserved sequence features. There were only weakly conserved features in the \( \beta_1 \) and \( \beta_5 \) regions (Figure 5.8C). One residue in \( \beta_1 \) was a conserved hydrophobic residue (I, F or M). In \( \beta_5 \), one position was a small residue (S, A or G) due to the turn between \( \alpha_4 \) and \( \beta_5 \), and one residue is a conserved isoleucine or leucine residue. This observation suggested there is no common sequence that mediates oligomerization. The low sequence identity in this region was probably due to the fact that the \( \beta_1-\beta_1 \) interactions were mainly mediated by mainchain H-bonding.
Comparison of BTB domain oligomers. A) Superposition of Ca traces of two PLZF$^{\text{BTB}}$ (PDB code 1buo, representative of PDB code 1cs3 as well) interacting dimers with two Kaiso$^{\text{BTB}}$ interacting dimers. In each structure, the dimers interact via the $\beta1$-$\beta5$ intermolecular sheet. This region is boxed and zoomed in. B) Surface representations of the oligomer chains generated by symmetry operations from five BTB domain structures, from PLZF, Kaiso, Bach1 and Gigaxonin BTB domains (PDB codes are indicated). Each dimer is coloured differently. The axis of propagation of the oligomeric chains is indicated by black lines, with thicker black lines coming out of the page towards the reader, thinner/greyed lines away from the reader. C) Multiple sequence alignment of Kaiso, PLZF, Bach1 and Gigaxonin BTB domains in the $\beta1$ and $\beta5$ regions. Yellow shading indicates $\beta$-sheets.
5.5. Discussion

The BTB domain is found in a large number of eukaryotic proteins and is involved in diverse functions. BTB domains have long been considered promiscuous interaction modules, as they have been shown to self-associate in homotypic and heterotypic fashion and to mediate formation of large protein complexes. There is little structural detail of this interaction diversity, as only homodimeric structures of the BTB domain exist. The two different crystal forms of PLZF\textsuperscript{BTB} showed interactions between homodimers that suggested a mechanism for the formation of higher order oligomers, but the biological importance was unclear (Ahmad \textit{et al.}, 1998, Li \textit{et al.}, 1999). In this chapter, we showed the BTB domain from Kaiso forms an oligomer. As a function of the crystal packing, Kaiso\textsuperscript{BTB} homodimers interacted in the same region of the domain that was observed in the structures of the BTB domains from PLZF. We also identified this interaction in Bach1 (unpublished, PDB id 2ihc) and Gigaxonin (unpublished, PDB id 2ppi). Each of these four BTB domains crystallized in a different crystal lattice but each showed interactions at the same region of the domain.

Solvent-exposed β-sheets are a common interaction surface, such as in sandwich and stacking interactions, or formation of β-barrels and β-propellers. Addition of β-strand elements to β-sheets is also a common mode of protein-protein interaction (Remaut & Waksman, 2006). The BTB homodimer contained two such sites, the “top” β2-β3-β4 sheet and the “bottom” β1-β5 sheet. The β1-β5 sheet was more solvent-exposed and was utilized in the BTB domain to form the interdimer β-sheet. This dimer-dimer interface was not extensive, involving only four hydrogen bonds and burial of 465 Å\textsuperscript{2} per Kaiso\textsuperscript{BTB} dimer, but
this is greater than the value of 400 Å² burial per chain that is considered an upper limit for non-biological/crystal contacts (Henrick & Thornton, 1998). The formation of this interface was likely driven by satisfying the hydrogen-bonding capacity of the exposed edge of β1, and burial of a few hydrophobic side chains. There were few conserved sequence features in β1 and β5 between Kaiso, PLZF, Bach1 and Gigaxonin, and it is likely that the identity of sidechains at this region were not important as the dimer-dimer interaction primarily involves mainchain hydrogen-bonding.

The location and curvature of the interdimer β-sheet allowed this interaction to be open-ended, forming long chains through the crystal. Energetic favourability was likely gained when additional dimers extend the oligomer. The Kaiso<sup>BTB</sup> structure showed that each of these chains of successively interacting dimers themselves interact. The two chains interacted in a polar interface with extensive surface area burial. Satisfying of hydrogen bonding, salt bridges and surface area burial in this interface may have also contributed energetic stability as the oligomer grew.

Kaiso<sup>BTB</sup> behaved as a dimer through the course of our purification, and we were not able to observe any higher order species in solution by size exclusion chromatography or dynamic light scattering.

Given these observations, two possibilities arose. First, the observed type of dimer-dimer interactions was favourable for merely crystallization of these BTB domains; the solvent-exposed β1 secondary structure element provided a convenient interaction site for crystal packing. Second, the dimer-dimer interactions of BTB domains were a biochemically relevant structure. In the absence of solution evidence showing oligomerization of these BTB domains, we were unable to differentiate between these two possibilities. It is possible
that oligomerization of Kaiso requires additional elements of the protein and/or DNA binding. Possible strategies for detecting Kaiso BTB domain oligomerization in solution are discussed in Section 5.4.

It could be suggested that the high protein concentration in the crystal allowed weak dimer-dimer interactions to be artificially encouraged (Ponstingl et al., 2005). However, a powerful argument in favour of the likelihood of the oligomer is the striking resemblance between the modes of dimer-dimer interactions observed between the four BTB domains (five crystal forms). Similar crystallographic behaviour of four proteins in five crystallographic environments was statistically unlikely.

If oligomerization of the Kaiso BTB domain is shown to be more than a crystallographic phenomenon, oligomerization of this transcription factor could be functionally important at a number of levels. Oligomerization of Kaiso could increase its affinity for specific genes containing multiple copies of the Kaiso Binding Site or mCpG's. Genes such as \textit{wnt11}, \textit{PPAR-\delta}, \textit{c-Myc}, \textit{cyclinD1} and \textit{matrilysin} have been shown to have multiple Kaiso recognition sites (Daniel et al., 2002, Kim et al., 2004, Park et al., 2005). After initial recognition of the recognition site, additional binding events may be favoured by association of BTB domains, also known as cooperative binding.

The BTB domain-containing transcription factors GAGA and Bach1 are considered architectural transcription factors due to their ability to lead to higher order chromosomal architecture effects, such as bending and steric hindrance of access to DNA of transcription machinery. These properties were dependent on the presence of a BTB domain (Katsani et al., 1999, Igarashi et al., 1998). These properties have not been observed with Kaiso or other human BTB-ZF proteins and should be addressed with future experiments. The Kaiso$^{\text{BTB}}$
structure we present here could be compatible with these properties seen in other BTB domain-containing proteins.

The Kaiso oligomer could provide new or modified sites for protein-protein interactions; oligomerization of other transcription factors has been shown to be necessary for recruitment of corepressors (Nitta et al., 2005). There are charged and hydrophobic features observed in the Kaiso\textsuperscript{BTB} oligomer that could represent novel protein-protein interaction sites. The D33N and K47Q mutations in Kaiso\textsuperscript{BTB} resulted in loss of interaction with N-CoR without affecting dimerization (Park et al., 2005), and both of these residues were accessible on the outer surface of the oligomer. Interestingly, the lateral groove was not buried in the oligomer, except for β1 and its involvement in the dimer-dimer contacts, and could be accessible for co-repressor interactions. Another possible site of interaction is the hydrophobic surface. The CTCF insulator interacted specifically with the BTB domain from Kaiso and not PLZF, BCL6, FAZF or Hic-1 (Defossez et al., 2005); the structure presented here should be important for studying this interaction.

The oligomeric structure of Kaiso\textsuperscript{BTB} shed light on the mode of heteromeric interactions mediated by the BTB domain. While BTB domains have been shown to interact in a combinatorial fashion, often by yeast two-hybrid methods, it is not known if these interactions occur within a single dimer, or between two homodimers. The structure of Kaiso\textsuperscript{BTB} suggested heteromeric interactions between BTB domains could occur between different homodimers via formation of the intermolecular β5-β1-β1-β5 sheet. This has been suggested for the GAGA-Ttk interaction (Pagans et al., 2002). The ability for BTB domains to form the β1-β5 sheet could be an important determinant on the capacity and specificity of
heterodimerization. This requires experimental validation with recombinantly purified BTB domains.

Many of the themes described here have been observed for the SAM domain, which mediates oligomerization of ETS transcription factors and Polycomb group (PcG) proteins, allowing complex formation over large distances (Kim et al., 2005, Tran et al., 2002, Kim, Gingery et al., 2002, Kim et al., 2001, Kim & Bowie, 2003). Early structures of the SAM domain showed ambiguity regarding the oligomerization state of this domain. Two structures of the SAM domain from EphB2 have been solved, with one showing an oligomeric organization in the asymmetric unit, and the other showing only a monomeric domain (Thanos, Goodwill et al., 1999, Thanos, Faham et al., 1999). Solution studies of this SAM domain showed a 6 mM monomer-dimer dissociation constant, implying self-association was not relevant (Thanos, Faham et al., 1999). However, oligomerization was unambiguously shown in solution and in crystal structure with the SAM domains from TEL and polyhomeotic, whose dimerization dissociation constants were measured as 190 and 1.7 nM, respectively (Kim et al., 2001, Kim, Gingery et al., 2002). As well, factors such as pH and time have been shown to play roles in the formation of SAM domain polymers (Bhattacharjya et al., 2005). By analogy to the BTB domain, it is possible that an oligomeric BTB domain with a nM dimer-tetramer dissociation constant can be measured only with a specific BTB domain, and under the appropriate solution conditions.
5.6. Conclusions

The structure of the BTB domain from Kaiso was determined. The structure is a strand-exchanged homodimer. The structure also contained interactions between dimers involving contacts between the N-terminal β1 strands, forming an oligomeric structure through the crystal. This observation resembles similar interactions seen in other crystal structures of BTB domains. However, we did not detect oligomeric Kaiso BTB domain in solution. Therefore, the interactions between dimers remains a purely crystallographic phenomenon.
Chapter 6

Future Directions
6.1. Preface

This graduate work provided many insights into the sequence distribution, conservation patterns, structure and function of the BTB domain. This work has contributed a high-quality sequence collection stored in a searchable database to the Privé lab and the wider scientific community. This database enabled some discoveries to be made, including two new domains in BTB domain-containing proteins, and a greater understanding of the evolution and relationships between BTB domain sequences. As well, this work contributed four crystal structures of BTB domains, each of which has generated novel insights into this domain’s function.

Based on these findings, this work generated many questions that need to be addressed to fully understand the evolution, structure, and function of the BTB domain. These are divided into five topics: further sequence analyses; studies on dimerization and domain-swapping properties; studies on oligomerization and self-association; co-repressor interactions and other protein-protein interactions; and additional crystal structures of BTB domains.
6.2. Sequence analysis of BTB domains

6.2.1. Identify a successful clustering scheme.

A major part of my graduate studies was attempting to group, or cluster, BTB domain sequences into meaningful groups that correlate with the function of the domain in terms of structure and protein-protein interactions. We attempted to group sequences using phylogenetic criteria by various algorithms (distance, maximum parsimony, maximum likelihood methods) but the high level of sequence divergence of BTB domains precluded generation of statistically valid clustering schemes. As described in Chapter 2, the best scheme that could be derived was grouping of BTB domain sequences according to the domain composition/architecture of the full-length proteins. However, sequence divergence was high even within obvious subfamilies, such as within the BTB-ZF subfamily where pairwise sequence identities were no higher than 45%, and even higher between subfamilies. Therefore, the clustering scheme that was presented lacks statistical backing and cannot be presented as the evolutionary history of the divergence of the BTB domain sequence.

Future work should focus on clustering BTB sequences using non-evolutionary methods. Methods such as principal component analysis (PCA) and support vector machines could be employed. PCA is a technique to reduce multidimensional data to lower dimensions, by eliminating those dimensions that are not responsible for a majority of the statistical variation. PCA has often been applied to the classification of proteins (Zhao et al., 2005, Grognux & Reymond, 2004, Casari et al., 1995). Starting from a multiple sequence alignment of BTB domains from various types of proteins, PCA could be used to identify
the positions in the alignment that account for the majority of the variation in the sequence. The sequences could then be clustered according to the changes in these positions. As described in Chapter 2, we did notice that the hydrophobic core positions showed the most conservation across divergent BTB domains, therefore the surface-exposed amino acids account for the majority of the variation. Since the number of surface-exposed amino acids is large and shows a great deal of sequence diversity, the major issue is identifying a reasonable and user-interpretable number of positions to use as "dimensions" in the PCA analysis. This requires identifying positions in the alignment of interest. For example, specific residues in the dimerization interface could be visualized in the PCA analysis.

In Chapter 4 we presented a structure-guided evolutionary history of the BTB domain that was calculated using the Bayesian inference program Mr. Bayes (Ronquist & Huelsenbeck, 2003). Discovering this program presented a minor breakthrough (although very late in my graduate studies) in the search for an effective means of clustering BTB domains. However, this method required experimentally determined structures of the BTB domains whose sequence we were interested in clustering in order to create a set of structure characters. Therefore, clustering BTB domains whose structure is not known would be difficult and would require some level of homology modeling to build up the set of structure characters necessary. For BTB domains within subfamilies with experimentally determined structures, such as the BTB-ZF and BBK subfamilies, “low resolution” homology model generation would be simple and could allow inference of structure characters.
6.2.2. Novel HMM's describing BTB and novel domains.

The BTB Domain Database contained HMM's describing twelve classes of BTB domain-containing proteins. This was not an exhaustive collection; there were additional subfamilies that could be aligned and used for novel HMM's. The BTB domain sequence could show important characteristics in these additional subfamilies. Currently, the HMM trained in a non-subfamily specific manner (the "catch all" HMM, huge_alignment.hmm) successfully identifies the BTB domains in most subfamilies, but is not useful for generating multiple sequence alignments or identifying domain boundaries in a subfamily-specific manner for these not well studied subfamilies. These include but are not limited to proteins with the domain composition BTB-Ankyrin and BTB-RCC1.

A large number of proteins in the database, especially within the C. elegans, D. rerio, and A. gambiae genomes, appear to contain only the BTB domain in a longer polypeptide (Figure 2.5). Given that the majority of BTB domain-containing proteins contain additional functional motifs, it is not likely that these proteins truly contain only a BTB domain. It would be worth closer analysis of these proteins, beginning with sequence clustering of the non-BTB region, followed by possible identification of novel domains.

We spent some time studying the Germ Cell Less proteins, found in Drosophila and mammals, that appears to contain a novel domain we tentatively described as the GCL domain. These proteins have the architecture BTB-BACK-GCL. The GCL domain should be further studied as was done for the BACK and PHR domains in Chapter 2. This family could be interesting to study given the domain architecture resembles proteins that interact with Cullin3-based E3 ubiquitin ligases.
There is a family of proteins containing what appears to be a domain resembling the T1 domain, but do not contain the potassium channel transmembrane domain or no other identified domains. These proteins have been automatically named KCTD by various databases. It is unknown if this T1-like domain adopts the BTB fold and whether it tetramersizes. Further sequence analysis of these proteins, as well as generation of HMM's, should be completed.

6.2.3. Search more genomes.

Genome sequencing efforts constantly produce additional eukaryotic genomes. The BTB Domain Database was far from exhaustive, therefore it would be worthwhile expanding the database to fully explore the eukaryotic evolution. The capability to add more genomes from Ensembl already exists in the administration area of the Database and this should be explored. Possible additional genomes to add from the Superorder Laurasiatheria, such as the cat (*Felis catus*) and the cow (*Bos taurus*); the Order Cingulata, such as the armadillo (*Dasypus novemcintus*); marsupials such as the opossum (*Monodelphis domestica*); and other chordates such as the sea squirt (*Ciona intestinalis*). Populating the database with additional genomes would allow many additional analyses, including targeted genomic comparisons to look at the evolution of specific subfamilies of BTB domain-containing proteins.
6.2.4. Targeted genomic comparisons.

A common theme in current genomics studies is targeted genomic comparisons as a method to gain a better understanding of evolutionary processes. This type of comparison could be completed with BTB domain-containing proteins, especially for BTB-ZF proteins within mammalian genomes. For example, the six proteins of interest in our lab (PLZF, BCL6, FAZF, Miz-1, LRF and Kaiso) should be compared across all sequenced mammalian genomes. This would allow identification of the most highly conserved regions that are likely to be functionally important.

6.3. Relevance of domain swapping and establishing heterodimerization

The structure of the FAZF BTB domain revealed a unique conformation of the N-terminus of the domain, such that the structure is not domain swapped. In other systems, domain swapping is considered a mechanism towards the formation of oligomeric structures that have a selective benefit relative to their monomeric counterparts (Liu & Eisenberg, 2002). As well, domain swapping is thought to be part of the mechanism of the formation of higher order structures, such as amyloids, that are biologically harmful.

In the context of BTB domains, the functional and evolutionary bases for domain swapping are unclear. With the structure determination of FAZF$^{BTB}$ revealing a non-domain swapped structure with a much smaller dimerization interface, we initially thought that this BTB domain could separate into folded monomers. Monomer-dimer exchange may make it more likely to generate heterodimers. As detailed in Chapter 4, we were unable to identify monomeric FAZF$^{BTB}$, suggesting that the less-extensive dimerization inter-
face observed in the crystal structure is as competent for dimerization as the interface in the
domain-swapped BTB domains. There are examples of homodimers with nM $K_d$ values
that nonetheless exchange subunits. Unpublished data by Eden Fussner, a previous gradu-
ate student in the lab, showed that the FAZF BTB domain could indeed spontaneously het-
erodimerize via an experiment involving mixing a Thioredoxin-tagged FAZF BTB domain
with a non-tagged FAZF BTB domain. In fact, of the six BTB domains our lab focuses on
(PLZF, BCL6, FAZF, Kaiso, LRF and Miz-1), only FAZF$^{\text{BTB}}$ could exchange spontane-
ously while the other five homodimers could not exchange without scrambling of the tagged
and untagged BTB domains by unfolding and refolding.

Since FAZF$^{\text{BTB}}$ is the only non-swapped BTB domain, taking into account all the
data suggested that the lack of domain-swapping allows exchange; conversely, domain-
swapping prevents exchange between subunits of the BTB homodimer. Perhaps the N-
terminus of the BTB domain provides a switch between an exchangeable form of the do-
main, when the N-terminus is not domain-swapped, and a non-exchangeable form of the
domain, when the N-terminus is domain-swapped. Further experiments could be designed
to prove this model. Firstly, dynamics in the conformation of $\beta1$ would be necessary in
BTB domains that heterodimerize. Movement of $\beta1$ could be studied with NMR. Sec-
ondly, the FAZF BTB domain could be converted into a domain-swapped version that
would be predicted not to spontaneously exchange. As well, the PLZF BTB domain could
be engineered to be truly non-domain-swapped, which would be predicted to enable sponta-
neous exchange. These experiments would require screening mutations in the hinge-loop
and possibly the closed dimerization interface of the FAZF and PLZF BTB domains for
proper folding. Converting the hinge-loop sequence of FAZF into that in PLZF, and vice versa, would be a good place to begin. If these mutations were completed in the context of the FAZF\textsuperscript{BTB-2C} and PLZF\textsuperscript{BTB-2C} variants (Section 4.4.9, Figure 4.13), the simple domain-swapping assay could be performed to test the swapping state. The swapping mutants could then be tested by the size exclusion heterodimerization assay developed by Eden Fussner.

As discussed in Chapter 4, it is possible that the non-domain-swapped state of the FAZF BTB domain is related to the evolutionary history between FAZF and PLZF and is a phenomenon specific only to these two proteins. Identifying additional BTB domains that are not domain-swapped would suggest domain-swapping plays a more general role in the function of BTB-ZF proteins.

As well, domain swapping could play a role in co-repressor binding specificity. The conformation of FAZF\textsuperscript{BTB} alters the dimerization interface, which forms part of lateral groove region that binds the SMRT-BBD peptide and the charged pocket region that has been shown to play a role in transcription repression. Would a domain-swapped variant of FAZF\textsuperscript{BTB} have greater affinity for SMRT, or would it be a stronger repressor? These could be tested using the mutants discussed above.

The dimerization interface of FAZF\textsuperscript{BTB} is much smaller than the dimerization interface of the other BTB domains. It would be interesting to mutate the residues in the dimer interface to test their importance in stabilizing the dimer. It could be possible to generate a monomeric FAZF BTB domain. These mutants could then be tested for heterodimerization/exchange propensity, which would suggest their role in the homo- vs. heterodimer equilibrium.
The range of heterodimerization of BTB domains needs to be established, and the structures presented in this thesis can be used to understand the specificity of interactions. The data generated by Eden Fussner suggests that along with the spontaneous exchange of subunits in the FAZF$^{BTB}$ dimer, the following pairs of BTB domains heterodimerize after unfolding and refolding: Miz-1$^{BTB}$ + LRFB$^{BTB}$, Miz-1$^{BTB}$ + BCL6$^{BTB}$. Why do these pairs heterodimerize and not the other pairs? How is the dimerization interface of the Miz-1, LRF and BCL6 BTB homodimers able to reorganize to accommodate the other BTB domains? It is interesting to note that these three proteins are all expressed in B-cells, thus it is possible, in a cellular setting, for heterodimerization to occur.

In Chapter 4, the comparison of the dimerization interfaces of the six BTB domains presented suggests the dimer interfaces share a set of highly conserved residues in most of six structures. If the dimer interfaces are so similar, why are only specific pairs of heterodimers observed? Is it a question of co-expression in the same cell? It is worth mutating the non-conserved residues in the dimerization interface. These mutants should then be tested for heterodimerization. As well, a heterodimeric BTB domain should be crystallized. These findings could be used to predict additional heterodimers and to understanding the “rules” of heterodimerization of BTB domains.

In a larger context, the functional/cellular role of heterodimerization of BTB domains needs to be clearly established. Three main conceivable consequences for heterodimerization of BTB domains would be to affect a) co-repressor recruitment, b) other protein-protein interactions, and c) DNA binding. Does a BTB heterodimer have different affinity or specificity for co-repressors? For example, the Miz-1 and BCL6 proteins interact and
may form a heterodimer via their BTB domains (Phan et al., 2005). Does the Miz-1\textsuperscript{BTB} + BCL6\textsuperscript{BTB} heterodimer form, and does it have different affinity for SMRT-BBD binding and/or repression potency relative to the Miz-1 or BCL6 homodimers? Does a heterodimer recruit different proteins than a homodimer? Does the heterodimer have different gene specificity than the homodimers? In the cell, is there dynamic exchange between BTB-ZF homodimers and heterodimers, and does this play a functional role in transcription regulation? Is the balance between homo- and heterodimers affected by signaling cascades, as occurs in other families of transcription factors?

6.4. Validation of BTB domain oligomerization

The crystal structures of the BTB domains from Kaiso and PLZF showed a very well conserved pattern of self-association via interactions of the β1-β5 sheet from different dimers. These interactions occurred on both sides of the dimer in a linear fashion such that the oligomer perpetuated throughout the crystal. As well, there was similarity with the self-association of the BTB domains from Bach1 and Gigaxonin (Section 5.4.8). The striking similarity of the dimer-dimer interaction interface across these different proteins in many different crystal forms suggested, but did not prove, this interface is functionally relevant for the oligomerization of BTB-ZF proteins.

Support for the biological evidence of BTB oligomerization would be strengthened by the observation of a higher molecular weight species in solution. We do not observe species other than dimers in our routine purifications of any human BTB domain, and size exclusion chromatography and dynamic light scattering of the Kaiso BTB domain at reasonably high concentration (5 mg/mL) also showed exclusively a dimeric species. Presumably,
the association between Kaiso BTB domains is very weak, and is only manifested at the very high concentrations (estimated to be about 500 mg/mL) in the crystal.

Future experiments should focus on establishing a solution assay for oligomerization. A possible experiment could be testing the effect of solution crowding agents on the molecular weight of Kaiso\(^\text{WTB}\). Crowding agents have been shown to affect the stability, association and function of proteins and are thought to better mimic the cell (reviewed in Minton, 2006). Agents such as PEG, polyvinyl alcohol, Dextran and BSA, generate excluded volume and could increase the probability of association of BTB domains. In the absence of a functional assay for Kaiso BTB, identifying an appropriate biophysical detection method is the limiting factor for detecting self-association.

Another experiment that is utilized for studying weak protein-protein interactions is self-interaction chromatography (SIC). This experiment has been utilized for the measurement of the osmotic second virial coefficient (\(B_{22}\)) of proteins, which is a measure of a weakly attractive protein interactions (reviewed in Valente et al., 2005). The \(B_{22}\) is considered by some as a useful indicator of crystallizability (reviewed in Tessier & Lenhoff, 2003). In the context of the Kaiso BTB domain, the experiment would involve immobilizing the Kaiso BTB domain on a column to create a large number of interaction sites and flowing the same protein over the column. The retention time is measured and should increase in the presence of interactions between the BTB domains. The result can be compared to mutant versions of the BTB domain to identify critical residues for the association.

I did attempt this experiment but it was not included in this work as it was abandoned due to non-specific interactions between the Kaiso BTB domain and the Ni-
sepharose resin used as the column matrix (see Appendix 1). The experiment is very simple but requires a significant amount of protein (approximately 150 µg per run). The experiment is worth optimizing, and should be repeated with a suitable column resin to minimize non-specific interactions. It is important to ensure the Ni-sepharose column is fully saturated with Thioredoxin-6xhis-Kaiso\textsuperscript{BTN} or with a blocking agent such as BSA. Alternatively, the Kaiso BTB domain (not Thioredoxin-6xhis tagged) could be immobilized to another column matrix with lower non-specific affinity for proteins, such as EAH Sepharose 4B (GE Healthcare). The terminal carboxyl group on the Kaiso BTB domain would be coupled to the matrix, leaving the N-terminus and β1 of the BTB domain, where the putative dimer-dimer associations occur, free to interact with the soluble phase.

An alternative method to SIC is surface plasmon resonance (SPR). This experiment relies on a similar principle, where Kaiso\textsuperscript{BTN} would be immobilized on a dextran surface (via a His-tag) and Kaiso\textsuperscript{BTN} solution is flowed over. The change in refractive index upon self-interaction of the BTB domain can be measured and fit to an association curve. The major issue with this experiment would be whether the association affinity is high enough to detect an association.

It could be possible to detect dimer-dimer interactions between Kaiso BTB domains by introducing mutations that increase the association of dimers, or by using a cross-linking strategy. The N-terminus of each chain in the dimer closely approaches the N-terminus (forming the β1-β1 interdimer sheet) and the C-terminus of another dimer (Chapter 5.4.6). These close approaches could be taken advantage of to introduce salt-bridge interactions, cysteine point mutations that could form interdimer disulfide bonds, or covalent linkages by
using chemical cross-linking agents. For example, I7 could be mutated to an acidic residue in one pool of Kaiso$^{\text{BTB}}$, which would be mixed with a pool of Kaiso$^{\text{BTB}}$ with I7 mutated to a basic residue. Alternatively, I7 could be mutated to cysteine and allowed to oxidize and form a disulfide across multiple dimers. These strategies would require screening many mutations to find the best set that increase the association of dimers and will likely require multiple mutations to increase the binding affinity. These mutations would be designed based on the Kaiso$^{\text{BTB}}$ structure, and the close approach of pairs of residues across the dimer-dimer interaction would need to be identified. An increase in the affinity between Kaiso$^{\text{BTB}}$ dimers could be detected during the purification of the protein, where a higher molecular weight species could form under gel filtration. In the case of cross-linking, Kaiso$^{\text{BTB}}$ would be purified as normal, with the cross-link introduced after purification, via spontaneous oxidation to form disulfides, or addition of cross-linking agents that would introduce specific covalent links. Covalently linked dimers could be detected via SDS-PAGE or mass spectrometry, where a laddering pattern of species with increasing oligomerization state could be observed. Of course, it could argued that these mutations artificially induce interactions between Kaiso$^{\text{BTB}}$ dimers, therefore the proper controls would need to be undertaken, such as introducing similar mutations or cross-links at sites away from the dimer-dimer interface to rule out non-specific protein association.

Since Kaiso has not yet been shown to oligomerize by our lab or other labs, studies can be undertaken with a BTB domain from a BTB-ZF protein that has been shown to oligomerize, such as Hic1. We have not purified or characterized this BTB domain in the lab.
at all; Hic1^{BTB} can be crystallized and could show the dimer-dimer interactions. As well, the characterizations described could be done with this BTB domain.

It is possible that oligomerization occurs only in the context of the full-length BTB-ZF protein bound to a DNA template. This could explain why we have not been able to detect oligomerization of a BTB domain in solution. If multiple copies of the KBS or methylated CpG’s exist on a DNA template, multiple Kaiso proteins could assemble and the local concentration of the BTB domains would be high enough to make the weak dimer-dimer associations relevant. That is, as multiple Kaiso molecules encounter their binding sites on DNA that are within a certain proximity to each other, the self-association of the BTB domains would have a higher probability of occurring, forming the oligomer seen in the crystal. An appropriate experiment to study oligomerization of full-length Kaiso in the context of DNA binding and to detect cooperativity would be electrophoretic mobility shift assay (EMSA). If Kaiso self-associates on a DNA template, at constant Kaiso concentration an increase in the number of DNA binding sites, whether the KBS or methylated CpG’s, would result in an increasing population of Kaiso-DNA complexes. If the model proposed above is true, the DNA-binding affinity should dramatically increase as the number of DNA binding sites increases. In fact, this phenomenon has been observed with GAGA factor, where successive binding events are favoured (Katsani et al., 1999).

It is possible the presence of such a Kaiso oligomer would have an architectural effect on chromatin conformation, and could provide many docking sites for co-repressors and HDAC’s.
If a positive result is observed showing a correlation between the number of binding sites, DNA-binding affinity and oligomerization, mutations at the dimer-dimer interface can be generated and tested in the same system. I have successfully produced the well-behaved Kaiso$^{BTB}$ S3E+K5D mutant, with the mutations designed to dramatically alter two residues that provide significant surface area burial in the dimer-dimer interaction (31 and 101 Å$^2$, respectively, out of a total of 465 Å$^2$). This mutant, or other mutations in this region or truncations of the N-terminus of Kaiso$^{BTB}$, will be used by our collaborators and could be incompetent for oligomerization and co-operative DNA binding. This would conclusively show the dimer-dimer interaction is relevant to the oligomerization and DNA-binding functions of Kaiso.

From a larger perspective, the functional consequence, if any, of oligomerization of Kaiso or other BTB-ZF proteins on transcription should be better studied. Firstly, computational searches for multiple copies of the KBS or mCpG’s at or near promoter regions should be identified in the human genome. I have designed a simple Perl script to perform this function, and preliminary evidence suggests there are indeed multiple copies in the 1 kb upstream of numerous genes in the human genome. For example, the script detected that the gene *HHAT* (Ensembl gene ID ENSG0000054392), which codes for the protein Hedgehog acyltransferase, contains six copies of the KBS within 1 kb of its promoter. As a positive control, the script successfully detects the KBS within 1 kb of the promoters of genes that have been characterized in the literature to be bound by Kaiso, such as *matrilysin/MMP7* (ENSG00000137673) (Daniel NAR 2002), *rapsyn* (ENSG00000165917) (Rodova MCB 2004) and *beta-globin* (ENSG00000188170) (Defossez JBC 2005). It should
be appreciated that detection of transcription factor binding sites upstream of genes is common but does not guarantee their \textit{in vitro} occupation by the TF or \textit{in vivo} relevance. Therefore, this computational approach needs to have appropriate statistical significance and controls in place before it can be used as a predictive tool. However, there has not been a systematic detection of BTB-ZF binding sites, or the presence of multiple copies of these binding sites, that could accommodate an oligomer and is an approach worth considering.

Once a suitable gene with multiple binding sites is detected, BTB-dependent oligomerization and co-operative binding DNA-binding of Kaiso using the EMSA assay described above, should be demonstrated. Following this, experiments could be designed to observe the effect of oligomerization on DNA conformation. This could involve electron microscopy, as was done for GAGA factor (Katsani \textit{et al.}, 1999) to look for DNA looping or bending effects by Kaiso. As well, the effect of Kaiso self-association on transcription repression should be determined. This experiment could involve a standard repression assay (example in Ahmad \textit{et al.}, 2003). If BTB-dependent oligomerization and co-operative binding is occurring, there should be an exponential increase of repression strength with a sequential increase of DNA binding sites.

These experiments should provide insights into the \textit{in vitro} existence of BTB domain-mediated oligomerization of a BTB-ZF protein. It is hoped that the dimer-dimer interactions observed in our structure of Kaiso$^{\text{BTB}}$ and conserved in the PLZF, Bach1 and Gigaxoninin BTB domains, would help shed light on this phenomenon.
6.5. Protein-protein interactions mediated by BTB-ZF proteins

6.5.1. Determination of BTB/co-repressor interaction pairs.

It has long been known that certain BTB-ZF proteins are transcription repressors and are found in complexes containing nuclear co-repressors and HDAC’s. The BTB-ZF protein interacts with the co-repressor and this function is often, but not always, attributed to the BTB domain. A further complication is that some BTB-ZF proteins, such as Hic1, are transcription repressors but function in a SMRT/N-CoR- and HDAC-independent manner. Therefore, BTB-ZF proteins may interact with other cellular components to facilitate repression and chromatin remodeling.

In the case when BTB-ZF proteins do interact with the SMRT or N-CoR type of co-repressors, there remain many open questions in understanding the basis of selectivity in a direct interaction between the two proteins. Does the necessary functional interaction occur between the co-repressor and the BTB domain, or between the co-repressor and another region of the BTB-ZF protein, or both simultaneously? For a given BTB-ZF protein, what is the range of interacting co-repressors? Is there competition between co-repressors to bind a specific BTB-ZF protein? Likewise, for a given co-repressor, what are the range of BTB-ZF interaction pairs? Is there competition between BTB-ZF proteins for interaction with a specific co-repressor?

This work showed that the interaction between the BTB domain and the BBD is specific for the BCL6 BTB domain. The BTB domains of PLZF, Kaiso, LRF, FAZF and Miz-1 are necessary for transcription repression. Therefore, the possibility exists that these BTB
domains interact with non-BBD regions of co-repressors and this should be addressed. Do these interactions occur at the lateral groove, or elsewhere on the BTB domain? For example, Kaiso\textsuperscript{BTB} interacts with N-CoR; does this interaction occur via the lateral groove or another region of the BTB domain? Does this interaction occur with a BBD-like peptide in N-CoR?

FAZF is a weak repressor in a GAL4-DBD fusion transcription repression assay (Melnick \textit{et al.}, 2002) and does not interact with the SMRT-BBD. A serine residue (S44) in the charged pocket, a region found at the “top” of the BTB dimer, is postulated to be responsible for precluding binding to SMRT, as this residue is a lysine (K47) in the strongly interacting BCL6 BTB domain (Melnick \textit{et al.}, 2002). When S44 was mutated to lysine, the in vivo repression strength of FAZF BTB domain increased almost 4-fold. However, the structure of the BCL6\textsuperscript{BTB} + SMRT-BBD complex showed that K47 is not involved in the interaction. Therefore, what role does S44 play in repression? Does it play a role in interacting with SMRT outside of the BBD, or with another component of a transcription repression complex? It should be kept in mind that the discussed mutational analysis of the FAZF BTB domain occurred in a cellular setting, therefore any direct FAZF - SMRT interactions would be masked by other potential interactions. An interaction between FAZF\textsuperscript{BTB} S44K and SMRT should be tested in our lab with purified protein.

Miz-1\textsuperscript{BTB} does not interact with the SMRT-BBD. Again, does this protein interact with the SMRT outside of the BBD? Does interaction with co-repressors require the presence of c-Myc in a complex with Miz-1? Another major question is what are the proteins,
besides p300, recruited to potentiate transcription activation by Miz-1? Does this BTB do-
main play a role in activation, possibly by taking part in protein-protein interactions?

6.5.2. Other interactions involving the BTB domain.

The BTB domain is known to mediate protein-protein interactions besides self-
association and interactions with nuclear co-repressors. Currently, interactions are identi-
fied in a low-throughput manner, where specific laboratories are interested in a specific
BTB-ZF protein and identify novel interactions using methods such as yeast two-hybrid or
co-immunoprecipitation. There are a number of problems with these methods, including
their efficiency and speed of interaction identification, problems with ambiguity of a direct
interaction, and false positives (Mackay et al., 2007).

Furthermore, identifying interactions in this manner is not systematic or exhaustive.
Except for a co-immunoprecipitation/mass-spectrometry study identifying interaction part-
ners of BCL6, a systematic identification of the full range of interactions has not been com-
pleted for the BTB-ZF proteins of interest in our lab. Developing a high-throughput, sys-
tematic method to identify novel interactions using the collection of purified BTB domains
in our lab should be a priority going forward. One possible method could be tandem affinity
purification (TAP). In this system, BTB domains are expressed with two affinity tags, such
as calmodulin-binding peptide (CBP) and protein A, although other tags can be utilized.
The fusion protein is purified and mixed with cell lysates from cells that are known to ex-
press that BTB-ZF protein. The mix is immobilized on a first affinity column (IgG beads,
which bind protein A), washed, eluted, immobilized on a second affinity column
(calmodulin beads, which bind CBP) and washed. The final elution, containing the BTB-
ZF protein complex, can then be subjected to tandem mass spectrometry to identify specific peptides. This method is amenable to high-throughput analysis with multiple BTB domains. It is possible that many false positive interactions could be identified, therefore interactions must be verified by methods such as reciprocal co-immuncoprecipitation. The ultimate test for an interaction, but the most difficult, would be to express and purify the interaction partners and test for interaction via biophysical methods such as ITC or SPR. At this point, the necessary and sufficient interaction regions between the two proteins can be identified by minimizing the protein constructs, and the minimal components could be crystallized.

With respect to BTB-ZF proteins studied in this thesis, FAZF is known to interact with FANCC (Hoatlin et al., 1999). It is unknown if this interaction involves the BTB domain of FAZF and it is worthwhile mapping this interaction. As well, FAZF interacts with GATA transcription factors. The ZF motifs of FAZF interact with GATA-2 and GATA-3. While our laboratory does not explicitly study the structure, function or interactions involving the ZF motifs of BTB-ZF proteins, it is worthwhile to study the interaction of the ZF motifs of FAZF and the GATA proteins to understand the function of the full-length FAZF protein.

The Miz-1 BTB domain is known to interact with Topoisomerase binding protein IIb (TopBP1) (Herold et al., 2002) and the E3 ubiquitin ligase HectH9 (Adhikary et al., 2005). These interactions should be as these interactions are known to play roles in regulating the function of Miz-1. Does the large pocket formed by the movement of the β4 region of Miz-1\textsuperscript{BTB} play a role in any interactions? Regions of Miz-1 outside of the BTB domain, such as
the ZF’s or the C-terminal region, are involved in other interactions such as with c-Myc or p300 (Peukert et al., 1997, Staller et al., 2001) and HCF-1 (Piluso et al., 2002) and would be worth studying.

The Kaiso BTB domain is known to interact with the transcription factor CTCF. Like was done with the BCL6-SMRT interaction (Ahmad et al., 2003), it is important to identify the minimally interacting region of CTCF that interacts with Kaiso\textsuperscript{BTB} and to crystallize this complex. It would be very interesting to identify if the lateral groove of Kaiso\textsuperscript{BTB} is involved in this interaction and the conformation of CTCF.

6.6. Additional crystal structures of BTB domains

It is evident from the structures presented in this thesis that determination of crystal structures of BTB domains has revealed many insights into the dimerization, oligomerization and peptide-binding properties of BTB domains. The average pairwise sequence identity of BTB domains in the BTB-ZF subfamily is in the 20-40\% range, with most of this identity found in the hydrophobic core of the protein. Therefore, it is reasonable to suggest that additional crystal structures of BTB domains should reveal additional insights, such as domain-swapping conformation, oligomerization interfaces in the crystal and the nature and conformation of residues in the BBD-interacting lateral groove.

As was shown in Figure 4.1, BTB domains can be well expressed as Thioredoxin-6xhis tagged fusions. The Privé lab has also shown BTB domains can be well expressed GST fusions (unpublished results). There are a number of these constructs that should provide enough soluble BTB domain (at least 5 mg) for crystallographic studies. These include
BTB id’s 8 (RP58), 39 (ZBTB26), 40 (ZBTB25), 41 (ZBTB2) and 42 (ZBTB4). We chose to crystallize the FAZF, Miz-1, Kaiso and LRF BTB domains due to their biological relevance. This criteria should be used when deciding which additional BTB domains to study, and of the five highly expressed clones listed above, RP58 has the greatest amount of biological information. Like the c-Myc/Miz-1 complex, RP58 has been shown to associate with the DNA methyltransferase Dnmt3a (Fuks et al., 2001). Dnmt3a acts as a co-repressor for RP58 and Dnmt3a associates with HDAC1. The DNA-binding site for RP58 is known (Aoki et al., 1998), RP58 is localized in heterochromatin (Meng et al., 2000) and is thought to be involved in the development of the mouse brain (Ohtaka-Maruyama et al., 2007).

From a sequence perspective, its closest homologs of the six BTB-ZF proteins with solved BTB domain structures are Miz-1 and LRF with 47 and 37% identity between their BTB domains. However, the PLZF, BCL6 and Kaiso BTB domains are significantly less similar. Therefore, solving the structure of RP58-BTB could explore a novel region of sequence diversity of this subfamily. The lateral groove of this BTB domain appears to be more similar to LRF than the BCL6 BTB domain, but there are some conserved residues between RP58 and BCL6. It would be interesting to test for binding to the SMRT-BBD.

Other BTB domains from BTB-ZF proteins that are worth crystallizing are ZBTB4, also known as Kaiso-like 1, and ZBTB38. These proteins contain zinc fingers, three of which are highly similar with the three zinc fingers of Kaiso (Filion et al., 2006). These proteins have similar function with Kaiso in that they interact with methylated CpG nucleotides. Once a solution assay is established for oligomerization of BTB domains, it would be
interesting to determine if ZBTB4 or ZBTB38 oligomerize in solution and in crystal structure.

The structure of the Miz-1 BTB domain presented in Chapter 4 showed the β4 region has swung away from the β3-β2-β4 sheet that is conserved in all other structures. This phenomenon is also observed in another group's structure of the Miz-1 BTB domain (Stead et al., 2007). In our structure, the β4 region participates in crystal contacts via coordination of zinc atoms, therefore the crystallization condition may have affected the structure of this region. As well, our structure was relatively low resolution at 2.6 Å and the crystals were difficult to work with. To clarify whether the movement of β4 is not a crystallographic effect and to generate a higher resolution structure another crystal form of the Miz-1 BTB domain should be obtained. Miz-1\textsuperscript{BTB} was very difficult to crystallize: we obtained one crystal form of the wild-type Miz-1 BTB domain and one form which required reductive methylation, even after screening multiple protein solution conditions. We made some progress with the methylated crystal form (see Appendix 1). These data were not included in this thesis as the resulting crystals, while much better formed and larger than the needles of the wild-type protein that we used to successfully determine the structure, did not diffract x-rays sufficiently beyond 8.5 Å. These crystals were extensively optimized to try to increase the resolution, but additional optimizations could be attempted, such as additive screening. Alternatively, different constructs of the BTB domain could be designed, such as truncations or extensions at the C-terminus. We attempted to express truncations of Miz-1\textsuperscript{BTB}, (1-112) and (1-114), but these BTB domains were not monodisperse in gel filtration after their Thioredoxin-6xhis-tagged fusions were cleaved with thrombin; longer constructs should be generated. NMR studies of Miz-1\textsuperscript{BTB} could be attempted to identify whether the position of
erated. NMR studies of Miz-1\textsuperscript{BTB} could be attempted to identify whether the position of β4 is dynamic or zinc dependent.

The crystal structure of a heterodimeric BTB domain would be informative to reveal how the dimerization interface can adjust to accommodate another BTB chain. Clear biochemical evidence would be required for the existence of a BTB heterodimer to choose the right pair. There is functional data suggesting the BCL6 and Miz-1 BTB domains interact (Phan \textit{et al.}, 2005) and this could be the best candidate for crystallization. Formation of the heterodimer may require denaturing a mix of a tagged and a non-tagged BTB domain, which would separate the homodimers. Renaturation of the mixture would allow subunit exchange and formation of the heterodimer. The heterodimer would be of intermediate molecular weight from the tagged homodimer and non-tagged homodimer, thus it could be purified by gel filtration. The major potential road-block would be the yield after renaturation, which would need to be at least 5 mg for crystallization trials.

Structures of BTB domains from other subfamilies would be extremely valuable. As illustrated by my graduate studies, multiple structures could yield many functional insights and it would be helpful to have more than one reference structure from each of these major classes: BBK, MATH-BTB, BTB-bZip and BTB-NPH3 proteins. There is large interest in BTB domains as adaptor modules for interaction with Cullin 3 E3 ubiquitin ligase complexes and many groups design mutations and experiments based on the structure of the PLZF\textsuperscript{BTB} domain as a reference. It would be valuable to have a structure of a \textit{bona fide} Cullin 3-interacting BTB domain, especially for detailed mutational analysis.
Recently, the Structural Genomics Consortium (SGC) solved the first structures of BTB domains in classes other than the BTB-ZF subfamily. In Chapter 4 we made the prediction that the BTB domain from the BBK and BTB-bZip families would be dimeric and this prediction was validated in the structures of the BTB domains from the BBK protein Gigaxonin, and the BTB-bZip protein Bach1 (PDB id codes 2ppi and 2ihc, unpublished). Interestingly, each of these structures shows interactions between dimers that are similar to the dimer-dimer interactions we observed in the Kaiso-BTB structure. It is possible that Bach1\textsuperscript{BTB} and Gigaxonin\textsuperscript{BTB} oligomerize in solution; there is functional evidence suggesting Bach1 and BBK proteins function as oligomers (Igarashi \textit{et al.}, 1998, Robinson & Cooley, 1997).

The work presented in this thesis identified a number of general and specific features of the BTB domain superfamily. Combined with the proposed experiments presented here, insights could be made to be utilized in the design of specific therapeutic interventions to treat diseases with aberrant expression and activity of BTB domain-containing proteins.
**Bibliography**


A.1. Reductive methylation of Miz-1 BTB domain.

Crystallization screening with Miz-1^{BTB} resulted in needle-like crystals that were difficult to handle. The structure was determined to 2.6 Å but the "swung-out" conformation of the β4 region may have been a crystallization artifact (see Chapter 4). To improve diffraction resolution and rule out the role of zinc in the structure, we searched for a new crystallization condition. We turned to reductive methylation. Reductive methylation of primary amino groups (lysines and N-termini) is a simple and effective technique for modifying the crystallizability of proteins (Walter et al., 2006).

The reductive methylation of Miz-1^{BTB} was achieved and was reproducible using the following procedure. Purified Miz-1^{BTB} was extensively exchanged into a buffer containing 50 mM NaCl, 50 mM HEPES pH 7.5 using a centrifugal concentrator to a final volume of 100 µL and a concentration of 7 mg/mL. Each of the following steps was performed at 4°C and in the dark. Ultrapure formaldehyde was added at a molar ratio of 20 per lysine residue (6.4 µL). DMAB (dimethylamine borane complex) was added at a molar ratio of 10 per lysine residue (3.2 µL). The sample was incubated for two hours. The addition of formaldehyde and DMAB and the incubation was repeated for a second two hour incubation. After the second incubation, DMAB was added at a molar ratio of 5 per lysine residue (1.6 µL) or and the solution was incubated overnight (about 18 hours). The reaction was then quenched with 16 µL 1 M glycine and incubated for one hour. The solution was extensively exchanged into the final crystallization buffer, 300 mM NaCl, 20 mM Tris pH 8,
MALDI-TOF spectra of wild-type and fully methylated Miz-1 BTB domain. The measured molecular weights perfectly match the expected molecular weights from sequence (13051 Da and 13247 Da), showing the BTB domain methylated at seven reactive groups.

1 mM TCEP using a centrifugal concentrator, and then spun at high speed to remove aggregates.

The success of the methylation reaction was verified by mass spectrometry, which showed an increase of 196 Da after the reaction, corresponding to a net gain of 14 methyl groups (Figure A.1). Miz-1BTB contains six lysines plus the N-terminus, therefore the reaction successfully added two methyl groups to each of these seven reactive groups.

A.2. Crystallization of methylated Miz-1 BTB domain.

The fully methylated Miz-1BTB was used for crystallization screening against Hampton Crystal Screens I and II. This resulted in a new condition that was previously not observed with wild-type Miz-1BTB. As well, there were no crystals in the condition that crystallized the wild-type protein. These observations indicated the protein’s surface chemistry and crystallization properties were successfully altered by the reductive methylation reaction. The lead condition was optimized to two slightly different conditions: 1) 0.2 M Mg acetate, 0.1 M HEPES pH 7.0, 5% PEG 6k; 2) 0.1 M Mg acetate, 0.1 M Tris pH 7.5, 13% PEG 6k (Figure A.2). Unfortunately, these crystals did not diffract x-rays to better than 8.5 Å, even when mounted at room temperature. Further optimization of these crystals is worthwhile and could improve diffraction quality (see Section 5.6).

As a method of detecting interactions between Kaiso BTB domain molecules in solution to verify the oligomeric crystal structure presented in Chapter 4, we began developing a self-interaction chromatography assay.

Approximately 40 mg wild-type thioredoxin-6xhis-tagged Kaiso^{BTB} (Thx-Kaiso WT, the "bait") was loaded onto a 1 mL Ni-sepharose column (GE). Both the column and protein solution were equilibrated in binding/lysis buffer (300 mM NaCl, 20 mM Tris pH 8, 1 mM TCEP, 10 mM imidazole). Using the Biocad chromatography system, lysis buffer was flowed at a rate of 0.5 mL/min, with free BTB domains at approximately 1-2 mg/mL applied to the column after a 1 mL wash.

The chromatograms showed the presence of two main peaks: the peak with a lower retention time was interpreted as being unretained by the column (as shown by the location of the negative control of BSA), and the peak with the higher retention time interpreted as proteins that interacted with the bait protein and/or column matrix (Figure A.3). The Kaiso WT BTB domain appeared to be retained based on the ratio of the heights of the retained to unretained peaks, while the mutants Kaiso S3E and Kaiso S3E+K5D (location of mutations in Kaiso^{BTB} structure shown in Figure A.5) were not retained as effectively.
negative control of BSA was not retained at all. As well, the FAZF and Miz-1 BTB domains showed a higher proportion of non-retained vs. retained protein.

**FIGURE A.3.**

*Self-interaction chromatography with Bait = Thx-Kaiso WT*

*Preys*

- Kaiso WT
- Kaiso S3E
- Kaiso S3E+K5D
- FAZF
- Miz-1
- BSA

This suggested the free Kaiso BTB domain was interacting with some part of the immobilized Thx-Kaiso WT protein (the Thx-6xhis tag or the BTB domain), or the column matrix. As well, since the S3E+K5D mutant was not retained as effectively, this indicated
the interaction of the Kaiso BTB domain with the bait or column occurred at the N-terminus of the domain, or at the oligomeric dimer-dimer interface. To test for specific binding, a new 1 mL Ni-sepharose column was washed with lysis buffer without applying any bait protein and the experiment was repeated (Figure A.4).

**Figure A.4.**

Self-interaction chromatography of BTB domains with no bait protein (bare column. Plot shows Absorbance at 280 nm vs. retention volume. BTB domains flowed over column ("preys") are shown in the legend. Positions of unretained and retained peaks are indicated.
In this experiment, retention of all Kaiso BTB domains was noticed. This indicated that retention of the Kaiso BTB domain (WT or mutants) was due to non-specific interactions with the column matrix and therefore the experiments were not pursued further. It was encouraging that the Kaiso S3E+K5D double mutant had different retention properties than Kaiso WT, therefore the experiment should be improved to reduce the level of non-specific binding to the column matrix. See section 6.4 for discussion about repeating and improving the experiment.

**Figure A.5.**

*Location of Kaiso BTB domain mutants S3E and K5D. The mutations are localized to the dimer-dimer interaction surface.*