2017

Heterologous Expression and Purification of Full-Length Human Polybromo-1 Protein

Sarah Hopson
Michigan Technological University, sehopson@mtu.edu

HETEROLOGOUS EXPRESSION AND PURIFICATION OF FULL-LENGTH HUMAN POLYBROMO-1 PROTEIN

By

Sarah E. Hopson

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Chemistry

MICHIGAN TECHNOLOGICAL UNIVERSITY

2017

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Chemistry.

Department of Chemistry

Dissertation Advisor: *Dr. Martin J. Thompson*

Committee Member: *Dr. Marina Tanasova*

Committee Member: *Dr. Tarun Dam*

Committee Member: *Dr. Caryn Heldt*

Department Chair: *Dr. Cary F. Chabalowski*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

Over the past decade, it has become apparent that the human polybromo-1 protein (BAF180) has a critical role in cancer. BAF180 is known to be a driver mutation in clear cell renal cell carcinoma, where it has been found to be mutated in approximately 40% of cases. Mutations have also been found in several other cancers, including intrahepatic cholangiocarcinomas and epithelioid sarcomas. BAF180 is the chromatin targeting subunit of the PBAF (Polybromo-associated BRG1-associated factor) chromatin remodeling complex, a role facilitated by its nine domains: six bromodomains, which recognize and bind to acetylated lysines on histones; two BAH (bromo-adjacent homology) domains, found to be critical for PCNA ubiquitination following DNA damage; and one HMG (high mobility group) box, the DNA binding component. Furthermore, proper expression of BAF180 has also been linked to cardiac development and cell cycle regulation.

Despite these associations, the molecular level interactions of full-length BAF180 have yet to be studied; only the phenomenological effects of BAF180 deficiency/mutation have been studied. It is crucial that we understand the binding dynamics and specificities of wild type and mutated BAF180, since it is the recognition component of PBAF.

Expression of the recombinant full-length BAF180 protein has been difficult because of the complex nature of this protein, its unusual codon usage, and size. After *E. coli* expression failed, other expression systems were investigated and the yeast *Pichia pastoris* was chosen. *Pichia* was chosen for several reasons: its codon usage is similar to

that of BAF180 and it is a eukaryotic system possessing eukaryotic protein folding machinery and capable of performing post-translational modifications. Under control of the *GAP* promoter, full-length BAF180 has been successfully expressed in *Pichia pastoris*. This is the first time that full-length BAF180 has been cloned and expressed in a heterologous host. It was purified using anion exchange chromatography.

The ability to express and purify full-length BAF180 is a huge first step towards increasing the understanding of the molecular mechanisms of this protein and its association with cancer development.

Chapter 1 : Background

1.1 Introduction

Though discovered in the early 2000s, the complex interconnecting roles of human polybromo-1 or BAF180 are just starting to be understood. Though a lot of unknowns remains, it has been established that BAF180 plays roles in the development of the heart and the thymus (Sections 1.4.2.2 and 1.4.2.3), genomic stability and DNA repair (Section 1.4.3), and cell cycle regulation (Section 1.4.4) (Figure 1.1). Mutations in BAF180 are commonly observed in cancers (Section 1.5); most notably, it is mutated in 41% of clear cell renal cell carcinoma cases.[1] When BAF180 is mutated or its expression is affected, there are deleterious effects on normal cell function, such as increased cell proliferation and impaired DNA repair (Figure 1.2).

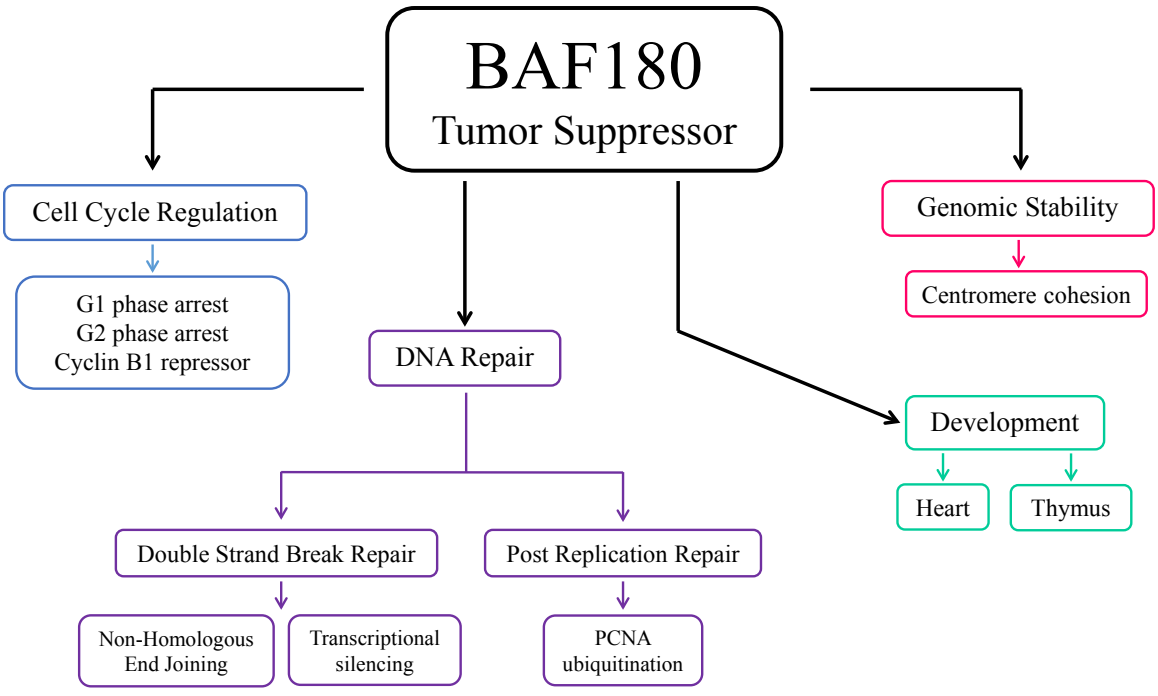


Figure 1.1: Diagram illustrating the roles of BAF180 in normal cells.

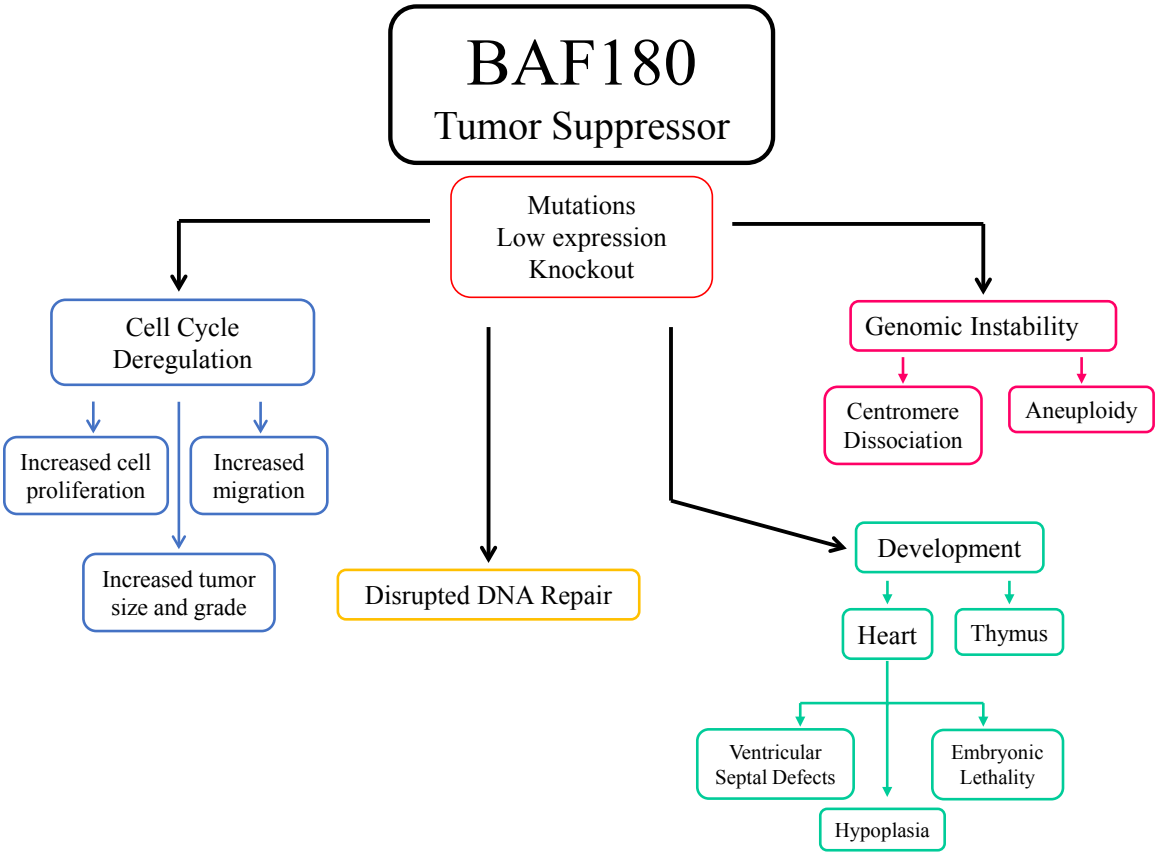


Figure 1.2: Diagram illustrating the consequences of mutations or alteration of expression of BAF180. The consequences of cell cycle deregulation, genomic instability, and disrupted DNA repair has the potential to cause cancer.

1.2 Mutation Susceptibility

When analyzing the DNA sequence of *PBRM1* (the gene that encodes BAF180) some interesting trends are seen. There are a lot of repeats, particularly A- and T-tracts. Base repeats have been known to increase the chances of point mutations during replication.[2-6] Poly-A tracks have been found to influence gene transcription both by stopping translation and creating frameshift mutations.[7] The *PBRM1* gene has 4905 base pairs (isoform two, GenBank ID: AAG48933.1); within the gene, there are numerous

two, three, and four repeats of all the bases (Table 1.1). BAF180 has nine alternate splicing isoforms.

Table 1.1: The number of two or more base pair repeats found in the *PBRM1* gene (GenBank ID: AAG48933.1). Numbers above the table refer to the number of consecutive of bases. The numbers in the box refer to how many times that base-tract is found in the mRNA sequence.

	Number of Sequential Bases					
Base	**Two**	**Three**	**Four**	**Five**	**Six**	**Seven**
A	394	135	54	21	7	1
U	245	69	16	9	1	0
C	232	48	11	4	1	0
G	207	39	7	0	1	0

1.3 Domain Organization

The human polybromo-1 protein (BAF180) is encoded by the *PBRM1* gene (Figure 1.3) and is considered to be the chromatin-recognition subunit of PBAF (polybromo-associated BRG1-associated factor). PBAF is involved in gene transcription. The *PBRM1* gene is located at 3p21.1. BAF180 is comprised of nine domains: six bromodomains, two bromo-adjacent homology domains, and one HMG-box. It is considered to be homolog to yeast Rsc1, 2, and 4 (or a gene fusion product of them).[8-9]

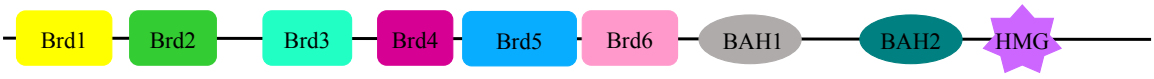


Figure 1.3: Domain organization of BAF180. BAF180 has 6 bromodomains (brd), the most of any known bromodomain-containing protein; 2 BAH (bromo-adjacent homology domains); and 1 HMG (high mobility group) box. The full-length protein has never been expressed in a heterologous host. The domain sizes and distances are approximately to scale [AAG48933.1]. BAF180 is encoded by the *PBRM1* gene.

1.3.1 Bromodomains

Within chromatin, DNA is wound around a core of eight histone proteins (two each of H2A, H2B, H3, and H4). In this state, DNA is highly compacted and therefore less accessible to replication- and transcription-mediating proteins. For replication and transcription to occur, the DNA needs to become accessible. This is believed to occur when post-translational modifications (PTMs) on the histone tails cause the interactions between the negatively-charged DNA backbone and the positively-charged amino acid resides on the histones weaken. PTMs refer to the modification of the side chain of an amino acid by the addition of a molecular group, such as phosphate or ubiquitin. Of interest in this work is the acetylation of lysine residues. The N-terminal tails of histones are rich in lysine residues and the acetylation of the ε-amino group on the side chain is generally associated with transcriptional activation.[10-13] It is hypothesized that the neutralization of the positive charge of the amino group on the lysine side chain weakens the DNA-histone interactions. Acetylation is carried out by histone acetyltransferases, a group of proteins which transfer an acetyl group from acetyl-CoA to the ε-amino group of lysine.

Bromodomains were first identified in the 1990s as a domain found in the Drosophila *brahma* gene and are the only known recognition module that bind acetylated lysines. Within the human proteome, there are 46 proteins which contain at least one bromodomain.[10,12-14] Of these, 15 function as transcriptional regulators and 8 function as chromatin remodelers (Figure 1.4).[10,14] Bromodomain-containing proteins are divided into eight groups based on their structural similarity; the bromodomains of BAF180 are

classified into group 8.[10,14] Figure 1.5 shows all known bromodomains grouped into families based on structural similarity; the functions of each are denoted by a colored dot. Functions within a family can vary, though consensus is seen in some families (as in families 2 and 8; see Figure 1.5).

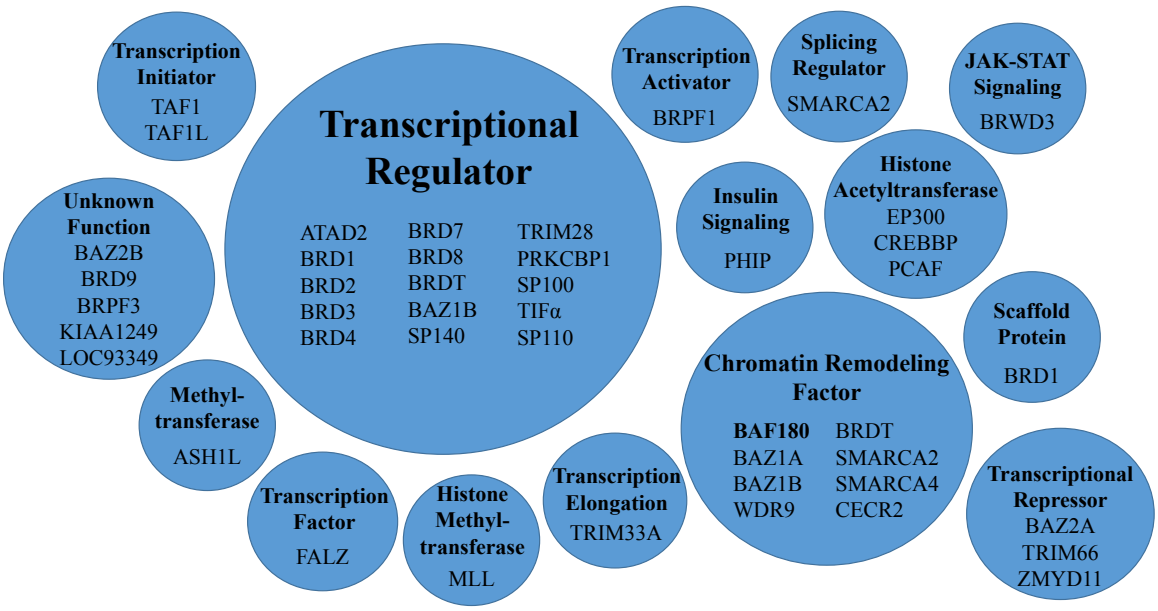


Figure 1.4: Functions of bromodomain-containing proteins. The majority are transcriptional regulators or chromatin remodeling factors.

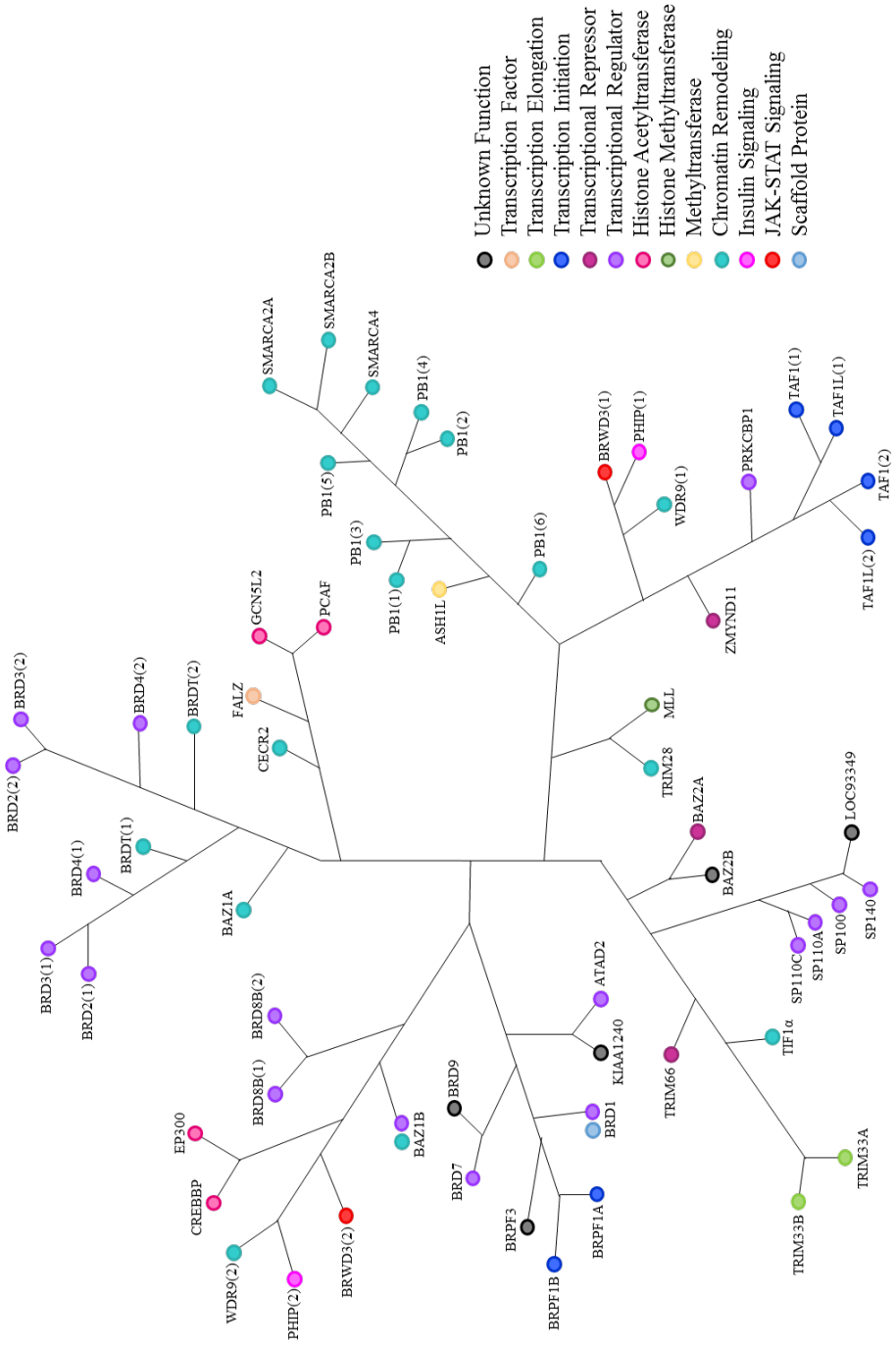


Figure 1.5: Phylogenic tree of the bromodomain-containing proteins based on structural similarity. Bromodomain-containing proteins are divided into 8 families. The molecular functions of each bromodomain within a particular family varies, though some similarities can be observed (family 1 – transcriptional regulator, family 8 – chromatin remodeling). This figure is based on figure 1B from Filippakopoulos, 2012.[14] Functional information from Filippakopoulos, 2012[14] and Muller, 2011.[12]

The ability of bromodomains to recognize and bind to acetylated lysines is critical for several cellular functions, including transcription, replication, and DNA repair.[10,13,15] The importance of this is demonstrated in the number of diseases caused by mutation or incorrect expression of different bromodomains. Among them are numerous cancers, mental disorders, immune disorders, and metabolic diseases.[16] BRD4 (of the BET family of bromodomains) forms a fusion protein with the NUT protein, which causes NUT midline carcinoma (NMC), a rare aggressive cancer with no available treatments.[10,16] Other examples of bromodomains involved in cancers include ASH1L (hepatocellular carcinoma), BRD7 (breast cancer), ATAD2 (breast cancer, colon cancer, lung cancer, uterine cancer, stomach cancer, and lymphoma), and BPTF (neuroblastomas and lung cancers).[16] A mental disorder, Rubinstein-Taybi syndrome, is often associated with mutations of CREBBP.[16] BRD1, SMARCA2, and BAF180 are linked to schizophrenia and bipolar disorders.[16] SP140 is linked to Crohn's disease and BRD2 is linked to rheumatoid arthritis. BAZ1B has associations with type II diabetes and obesity.[16] This is not a comprehensive list, but the relationship between bromodomain dysfunction and diseases is clear.

Epigenetic based treatments for various diseases, such as cancer, are becoming more popular and the inhibition of bromodomains has emerged as a treatment option. Most of the small molecules developed so far target proteins in the BET family (bromodomain and extra-terminal domain family) and all of the bromodomain inhibitors currently in clinical trials target the BET family (BET bromodomains were classified as

"most druggable").[17] Table 1.2 shows the status of the bromodomain inhibitors currently in clinical trials, as well as the targeted conditions.

The other seven families of bromodomains have generally been ignored, as early studies predicted that bromodomains outside of the BET family would be harder to target.[17] Vidler et al. scored the bromodomains of BAF180 "intermediate" to "difficult" druggability, with brd1, 3, and 4 considered "difficult" targets and brd2 and 5, considered "intermediate" targets; brd6 did not have a score.[17]

This analysis has recently been called into question when two groups created an inhibitor that selectively targets brd5 of BAF180.[18-19] Vidler et al. based their analysis on the assumption that the water molecules in the binding pocket remained. Both Myrianthopoulos and Sutherell created compounds which were able to bind to bromodomain 5 selectively with good affinity, while displacing water molecules.[18-19] Sutherell et al. published their compound in April 2016 and showed that it had a K_D = 124 nM for BRD5.[18] Myrianthopoulos et al. published in September 2016 and their compound had a K_D = 3.3 μM for BRD5.[19] Both compounds were able to displace BAF180 from chromatin.[18-19] Structural analysis conducted by Vidler et al. demonstrated that the binding pocket of the fifth bromodomain of BAF180 is unusual when compared to other bromodomain structures; it appears to have a shorter ZA loop and some key residues located in the binding pocket are shifted, resulting in a smaller binding site.[17] This may be a reason why newly developed molecules selectively target brd5.

Table 1.2: Bromodomain inhibitors currently in clinical trials. All of the drugs listed target the BET family of bromodomains. Of note is Resverlogix Corp.'s RVX000222, which has progressed to phase three clinical trials for diabetes and cardiovascular diseases. Data gathered from clinicaltrials.gov on 16 January 2017.[20]

Drug	Company	NCT Database Number	Phase	Status	Structure Disclosed?	Condition
OTX105/MK-8628	Ocoethix GmbH	NCT02259114	1B	Active	Yes	Solid tumors: NUT midline carcinoma, triple negative breast cancer, non-small cell lung cancer with ALK fusion protein or KRAS mutation, castrate-resistant prostate cancer, pancreatic ductal adenocarcinoma
OTX105/MK-8628	Ocoethix GmbH	NCT01713582	1	Completed	Yes	Acute myeloid leukemia, diffuse large B-cell lymphoma, acute lymphoblastic leukemia, multipple myeloma
OTX105/MK-8628	Ocoethix GmbH	NCT02296476	1/2	Terminated	Yes	Glioblastoma multiforme
OTX105/MK-8628	Ocoethix GmbH	NCT02303752	1b/2	Withdrawn	Yes	AML
OTX105/MK-8628	Merck Sharp & Dohme Corp.	NCT02698189	1	Recruiting	Yes	AML
OTX105/MK-8628	Merck Sharp & Dohme Corp.	NCT02698176	1B	Recruiting	Yes	NUT midline carcinoma, triple negative breast cancer, non-small cell lung cancer, castration-resistant prostate cancer
CPI-0610	Constellation Pharmaceuticals	NCT01949883	1	Recruiting	Yes	Lymphoma
CPI-0610	Constellation Pharmaceuticals	NCT02157636	1	Recruiting	Yes	Multiple myeloma
CPI-0610	Constellation Pharmaceuticals	NCT02158858	1	Recruiting	Yes	Acute myelocytic leukemia, myelodysplastic syndrome, myelodysplastic/myeloproliferative neoplasm, myelofibrosis
CPI-0610	University of Texas Southwestern Medical Center	NCT02986919	2	Not yet recruiting	Yes	Peripheral nerve tumors
RVX000222	Resverlogix Corp	NCT02586155	3	Recruiting	Yes	Diabetes mellitus, type 2, coronary artery disease, cardiovascular diseases
BMS-986158	Bristol-Myers Squibb	NCT02419417	1/2	Recruiting	No	Multiple indications cancer
ABBV-075	AbbVie	NCT02391480	1	Recruiting	No	Advanced cancer, bresat cancer, non-small cell lung cancer, AML, multiple myeloma
GSK2820151	GlaxoSmithKline	NCT02630251	1	Recruiting		Cancer
GSK525762	GlaxoSmithKline	NCT01587703	1/2	Recruiting	Yes	NUT midline carcinoma
GSK525762	GlaxoSmithKline	NCT01943851	1/2	Recruiting	Yes	Cancer
GSK525762	GlaxoSmithKline	NCT02964507	2	Not yet recruiting	Yes	Cancer
GSK525762	GlaxoSmithKline	NCT02706535	1	Recruiting	Yes	Drug interactions
RO6870810/TEN-010	Hoffman-La Roche	NCT02308761	1	Recruiting	No	Myelodysplastic syndromes, AML
RO6870810/TEN-010	Hoffman-La Roche	NCT01987362	1	Recruiting	No	Solid tumors, advanced solid tumors
BAY1238097	Bayer	NCT02369029	1	Terminated	No	Neoplasms
ZEN003694	Zenith Epigenetics	NCT02711956	1b	Recruiting	No	Metastic castration-resistant prostate cancer
ZEN003694	Zenith Epigenetics	NCT02705469	1	Recruiting	No	Metastic castration-resistant prostate cancer
INCB054329	Incyte Corporation	NCT02431260	1/2	Recruiting	No	Advanced cancer
INCB054329	Incyte Corporation	NCT02711137	1/2	Recruiting	No	Advanced cancer
FT-1101	Forma Therapeutics	NCT02543879	1	Recruiting	No	AML, acute myelogenous leukemia, myelodysplastic syndrome

1.3.1.1 Structure and Binding Mode

The bromodomain structure consists of a left-handed helix bundle composed of four α-helices (αZ, αA, αB, αC) connected by three loops (ZA-, BC-, AB-loops) (Figure 1.6A). Specificity is conferred by the ZA and BC loops[10,13-14,16,21-22] which form a hydrophobic binding pocket. The acetylated lysine residue forms a hydrogen bond with a highly conserved asparagine residue located near the end of the αB helix (Figure 1.6B).[10,13-14,21-23] Water residues located within the hydrophobic binding pocket also form hydrogen bonds with the acetylated lysine (Kac).[19,21]

Factors influencing acetyl-lysine specificity and affinity are (1) amino acid residues neighboring the binding pocket and (2) histone post-translational modification patterns. Amino acid residues surrounding the hydrophobic pocket have been demonstrated to influence acetyl-lysine recognition by interacting with residues that surround the acetylated lysine on the histone.[10,19,21-22] Different post-translation modification (PTMs) patterns on the histones have been shown to influence binding affinity at a specific acetyl-lysine residue; surrounding PTMs may increase or decrease a bromodomain's affinity for a specific acetylated lysine.[10,14,22] Interestingly, some bromodomains have demonstrated an ability to bind to two acetylated lysines at a time within the same binding pocket.[14,24] For example, the first bromodomain of the BET protein BRD4 has been shown to bind both acetylated lysines of H4K5/8ac at once; a crystal structure shows that both acetylated lysines are in the binding pocket (PDB ID: 3UVW).[14]

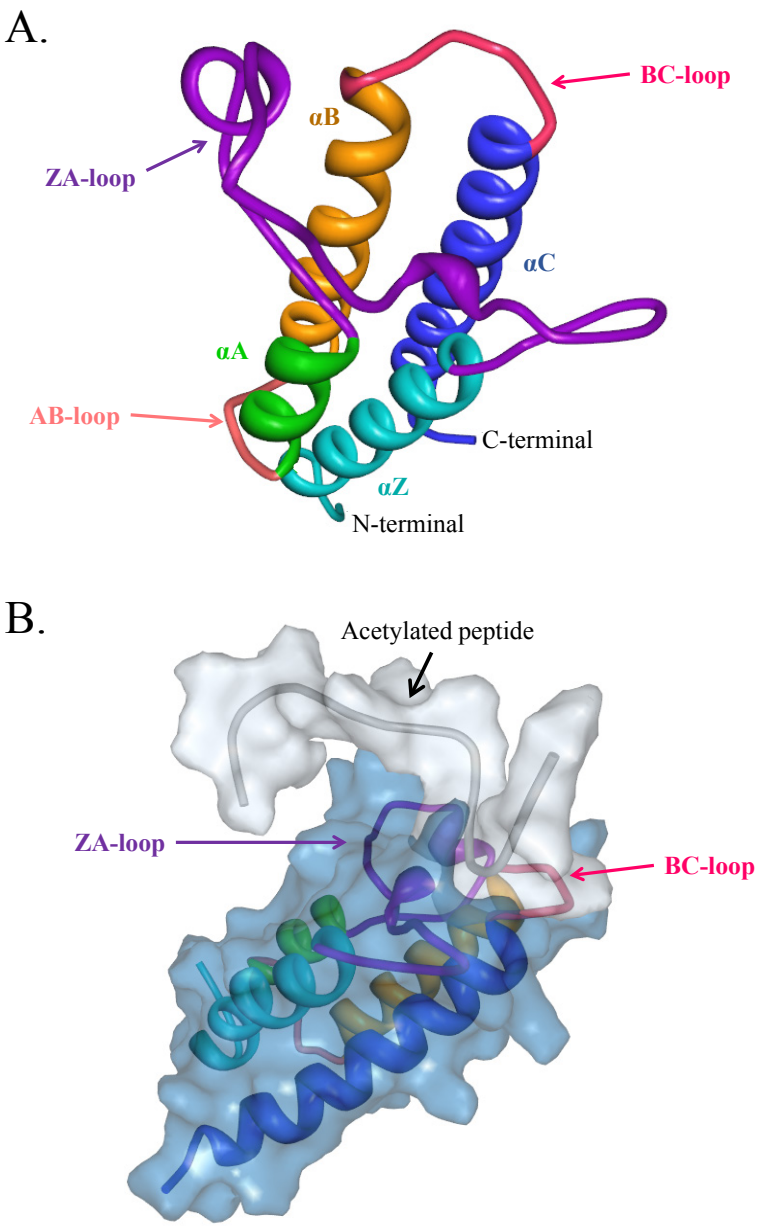


Figure 1.6: Crystal structure of a bromodomain. (A) The crystal structure of bromodomain 3 of BAF180 [PDB ID: 3K2J; structure not yet published: Filippakopoulos, P., Picaud, S., Keates, T., Chaikuad, A., Pike, A.C.W., Krojer, T., Sethi, R., von Delft, F., Arrowsmith, C.H., Edwards, A., Weigelt, J., Bountra, C., Knapp, S., Structural Genomics Consortium]. Labeled are the structural segments found in every bromodomain. The αZ-helix is located at the N-terminal end of the protein. (B) The crystal structure of bromodomain 2 (of BAF180) bound to an acetylated peptide [PDB ID: 2KTB].{Charlop-Powers, 2010 #616} The binding pocket is located between the ZA- and BC-loops of the structure. The αZ-helix is located at the N-terminal end of the protein. Light blue: αZ-helix. Purple: ZA-loop. Green: αA-helix. Pale pink: AB-loop. Orange: αB-helix. Hot pink: BC-loop. Dark blue: αB-helix.{Berman, 2000 #1292;Xu, 2009 #1085} Images manipulated using Protein Workshop.{Moreland, 2005 #1293}

Domains adjacent to the bromodomain (BRD) often include PHD zinc fingers, other bromodomains, and bromo-adjacent homology domains (BAH).[10,14] These neighboring domains have been shown to act together to increase the binding affinity for a bromodomain-containing protein to an acetylated lysine site. For example, BRD2 (bromodomain-containing protein 2), which has two tandem bromodomains, has a much higher affinity for H4K5/12ac (Kd=360 μM) than for the singly-acetylated H4K12ac (Kd=2.9 mM).[10] TRIM33, which has a PHD (plant homology domain) zinc finger followed by a bromodomain, has been shown to bind H3K9me3 and H3K14ac simultaneously.[10,13]

1.3.1.2 The Bromodomains of BAF180

BAF180 has six bromodomains, the most of any known bromodomain-containing protein. Of particular interest is the question of why BAF180 has six bromodomains. At most, other bromodomain-containing proteins contain two. It is accepted that bromodomains bind to acetylated lysines, but the specific interactions of the BAF180 bromodomains are unknown. It is possible that the bromodomains work in concert (as the tandem bromodomains of BRD4 do) and it is possible that they work independently (further discussion in Sections 1.3.1.1).[10,14,23-24]

Filippakopoulos et al. performed a very thorough SPOT array of histone H3 peptides with various combinations of post-translational modifications (acetylation, tri-methylation, phosphorylation, etc.). The data indicates that modifications surrounding the acetylated lysine do have an impact on binding affinity. For example, brd4 (of BAF180)

had “strong” affinity for H3K4ac, but “very strong” affinity for H3K4/9ac. This suggests that brd4 of BAF180 may be able to bind two acetyl-lysines simultaneously, in a manner similar to the BET protein BRD4(1).[14,24] Brd4 of BAF180 had “strong” affinity for H3K9ac, but “weak” affinity for a peptide containing additional acetylation sites on K4 and K14, and two phosphorylation sites on S10 and T11 (H3K4/9/14acS10/T11p, where ‘p’ means phosphorylation). It had no affinity for a peptide containing a tri-methylation mark on K4, acetylation at K9 and K14, and phosphorylation on S10 and T11 (H3K4me3K9/14acS10/T11p, where ‘me3’ refers to tri-methylation). The data also demonstrates that brd4 of BAF180 is a fairly ubiquitous binding partner, as it showed ‘very strong’ affinity for about 74% of peptides tested.[14] Data presented in that publication also supports the hypothesis that the bromodomains work together, as some multiply modified peptides were recognized by more than one bromodomain.

The bromodomains of BAF180 have been expressed individually and their affinities for specific acetyl-lysine sites on histones has been studied.[12,14,27-30] However, there has been some divergence in results, in terms of recognition patterns and affinities (Table 1.3). This may be due to several factors: different lengths of the bromodomains, differing peptide lengths, and differing peptide locations may be influencing the results. In addition, the affinities are hard to compare because different methods were used to quantitate the affinities (nuclear magnetic resonance, isothermal titration calorimetry, fluorescence anisotropy, dot blot, peptide/SPOT array).[12,14,23,27-30]

Table 1.3: Comparison of the individual acetylated lysine sites on histones 1-4 recognized by the bromodomains of BAF180. The specificity data between publications varies, with little consensus between them. This may be a reflection of the different techniques used, thus emphasizing the need to a comprehension study of all the individual bromodomains' binding sites.[14,27-28]

	Histone H1-4	Histone H2A		Histone H2B		Histone H3			Histone H4	
	Reference	Reference		Reference		Reference			Reference	
Bromodomain	**14**	**14**	**27**	**14**	**27**	**14**	**27**	**28**	**14**	**27**
PB1(1)	K33, K80, K84, K89	K15				K56, K64, K122	K36	K4	K44, K91	
PB1(2)		K36		K85		K14	K14	K9		
PB1(3)	K62, K74, K80, K84	K15	K15	K23	K15	K18, K36, K56, K64	K115	K9		K12
PB1(4)						K37, K122	K14, K115	K23		
PB1(5)				K85			K36, K112	K14		
PB1(6)	K84			K43	K24, K116	K56, K115				

A study was recently published wherein the researchers investigated the effect of point mutations of individual bromodomains. Porter and Dykhuizen reported that endogenous BAF180 isolated from HeLa lysate specifically recognizes H3K14/18/23/27ac, as evidenced during a peptide pull-down assay.[23] Porter and Dykhuizen suggest that the second bromodomain of BAF180 is the most critical for BAF180's tumor suppressor properties. They created six constructs wherein a single point mutation was introduced into the binding pocket of a different bromodomain and introduced the constructs into Caki2 cells (clear cell renal cell carcinoma (ccRCC) cells lacking BAF180). They used immunoprecipitation with BAF180 and BRG1 antibodies to confirm that the mutants were integrated into the PBAF complex, though with weaker binding than is seen for wild type.[23] Then they examined the effects of these mutations on cell growth and gene expression.

When wild type BAF180 was introduced into Caki2 cells, the growth rate decreased significantly. In contrast, introduction of the mutant brd1, 2, 4, and 5 BAF180

constructs had no effect on cellular growth, meaning the cells continued to grow at a rate similar to control cells that lacked BAF180. Analysis of gene expression patterns for all the cells demonstrated that the mutant brd1, 2, 4, 5, and 6 BAF180 proteins were unable to bring about the up-regulation of apoptotic genes that is usually observed when wild type BAF180 is introduced. The mutation in brd3 had no effect on gene expression as compared to wild type BAF180.[23]

The crystal structures of the first five bromodomains (BRDs) of BAF180 are similar to the structures of other bromodomains. Bromodomain 6 diverges from this pattern slightly as it has a fifth α-helix (Figure 1.7).[14] This is the only published crystal structure of brd6 (PDB ID: 3IU6). However, Filippakopoulos et al. included about 35 extra amino acids on the C-terminal end past where brd6 is predicted to stop.[31] In the same article, brd6 was shown to bind strongly (via SPOT peptide array) to H2AK36ac and H2BK85ac.[14] Another study by Charlop-Powers et al. used a dot blot array and saw that brd6 interacted with H2BK24ac and H2BK116ac.[27] This difference might be attributable to a few things: first, the cloned brd6 domains were of different lengths in the two publications (Filippakopoulos et al. used amino acids 741-885 and Charlop-Powers et al. encompassed amino acids 743-821).

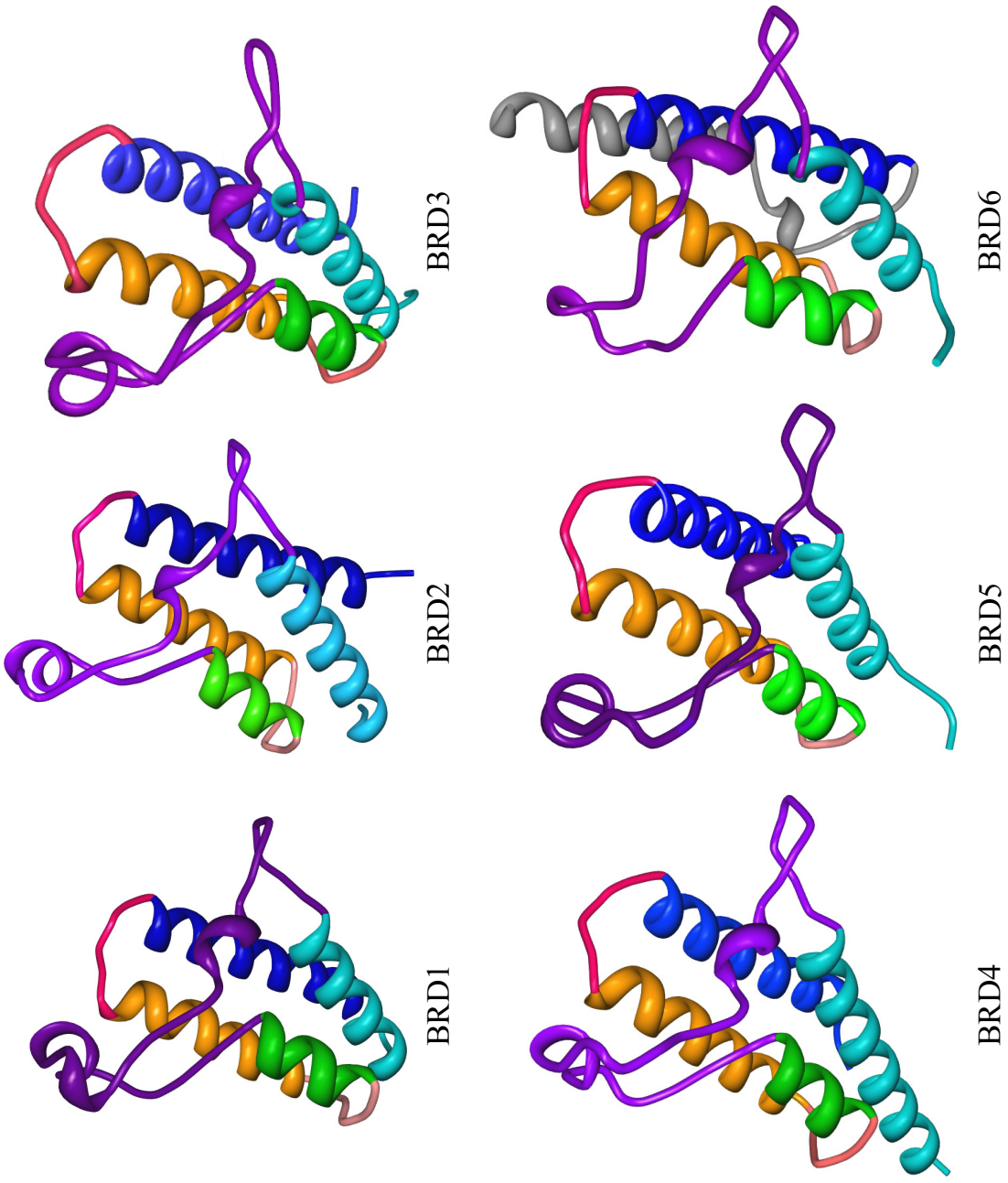


Figure 1.7: Comparison of the crystal structures of each of the six bromodomains from BAF180. Notably, brd6 contains an extra C-terminal helix (grey). The αZ-helix is located at the N-terminal end of the protein. Light blue: αZ-helix. Purple: ZA-loop. Green: αA-helix. Pale pink: AB-loop. Orange: αB-helix. Hot pink: BC-loop. Dark blue: αB-helix. [PDB IDs: brd1 – 3IU5; brd2 – 3HMF; brd3 – 3K2J; brd4 – 3TLP; brd5 – 4Q0N; brd6 – 3IU6].{Filippakopoulos, 2012 #627;Berman, 2000 #1292} Images manipulated using Protein Workshop.{Moreland, 2005 #1293}

By comparison of the same reference sequence and analysis by multiple domain prediction software, brd6 is located around amino acids 743-850 (CDD:743-850; ProSite:785-830; Interpro:743-851; SMART:743-849; MOTIF Search:649-706; Pfam:751-835).[31-36] Filippakopoulos added about 35 extra amino acids to the C-terminal end and Charlop-Powers did not actually clone the whole domain. As Filippakopoulos et al. had 35 extra amino acids past the boundary of predicted brd6, it is possible that their structure is flawed. Studying the sequence data of the crystal structure reveals that the fifth α-helix of brd6 begins at Thr863 (PDB ID: 3IU6); this is outside of the predicted domain boundaries (Figure 1.8).

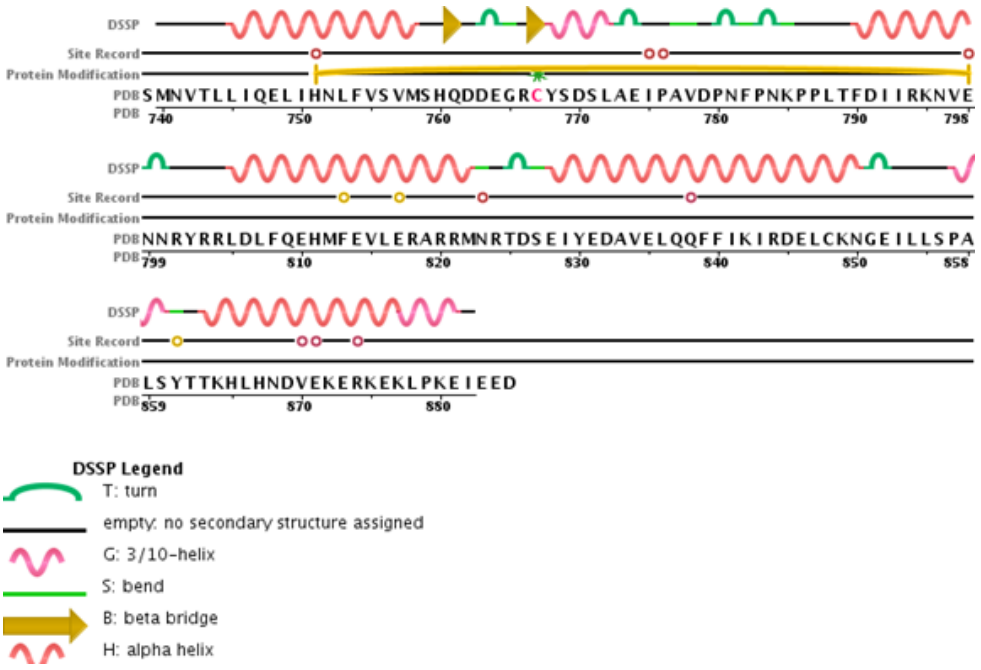


Figure 1.8: Sequence overlay of the brd6 structure. Brd6 is predicted to occur at amino acids 743-850 (CDD prediction). Filippakopouplos et al. cloned AA741-885, resulting in 35 extra amino acids. The fifth α-helix (the anomaly separating it from other bromodomain structures) begins at Thr863, outside of the predicted domain boundary. Sequence display image from PDB ID: 3IU6.[14,37]

Additionally, the peptides used for testing interactions were of different lengths and different positions; this has been shown to influence the recognition of a specific acetyl-lysine site by the bromodomain.[14,21-22] Third, they used different techniques for evaluating binding partners. Filippakopoulos et al. used SPOT arrays[14] and Charlop-Powers et al. used dot blot arrays.[27] It would be interesting to test whether the different constructs have different acetylated lysine recognition patterns, when using the same qualitative and quantitative methods.

1.3.2 Bromo-Adjacent Homology Domains

Bromo-adjacent homology (BAH) domains are believed to mediate protein-protein interactions.[8,32,38] The BAH domain was first identified in BAF180 and was named as such because it was adjacent to the bromodomains.[8] Since that initial discovery, BAH domains have been found in a wide variety of proteins, only some of which contain an adjacent bromodomain.[8] Generally, BAH domains are classified into two groups: Sir3-like BAH domains, which bind to nucleosomes, and RSC-like. The RSC-like BAH domains have not been extensively studied and their binding partners still need to be determined.[9]

BAF180 contains two BAH domains, proximal to the six bromodomains. Based on sequence and structure comparisons, these BAH domains fall into the RSC-like class. Indeed, the BAF180 protein is believed to be the human homolog or fusion of the yeast Rsc1, Rsc2, and Rsc4 proteins.[8-9]

The BAH domains in BAF180 are still relatively unexplored and the exact function is still unknown; however, some inferences can be drawn from other BAH domain-containing proteins. The mammalian ORC1 (origin recognition complex subunit 1) protein contains a BAH domain which has been shown to recognize and bind to a methylated lysine on histone H4.[38-39] Analyses of the structures indicate that the first BAH domain of BAF180 may also recognize methylated lysines, as the methyl-lysine binding pocket is fairly conserved between ORC1 and BAH1 of BAF180.[38] Pull-down assays of BAH1 of BAF180 demonstrated preferential binding to histone H3, but not to any of the other histones.[9] Additionally, the BAH domain of yeast Rsc2 was able to independently associate with histone H3 *in vivo*, albeit at lower levels than full-length Rsc2.[9] Experiments further demonstrated that BAH1 domain of BAF180 is capable of interacting with H3 as well.[9]

BAH domains can also bind to chromatin: the two BAH domains of the enzyme DNA cytosine-5 are localized at replication foci during S-phase.[8] Likewise, BAF180 is found to localize at kinetochores[40] (the region of chromosomes that connect to the mitotic spindle).[41] The BAH domains of BAF180 may play a role in maintaining genomic stability, since BAF180 is found at kinetochores and is known to be important for DNA repair and genomic stability. The roles of the BAH domains in BAF180 remain unknown.

A crystal structure of the first bromo-adjacent homology domain of chicken polybromo was determined by X-ray diffraction in 2005 (Figure 1.9) The BAH domain is composed primarily of β-strands. The ten β-strands are connected by a few helices and form a barrel structure.[42]

Using the crystal structures as a guide, two conserved residues near the predicted H3 binding pocket were mutated in Rsc2 (W436A/W436L and K437E). All three of the mutations impacted histone H3 binding in a negative manner both *in vivo* and *in vitro*.[9]

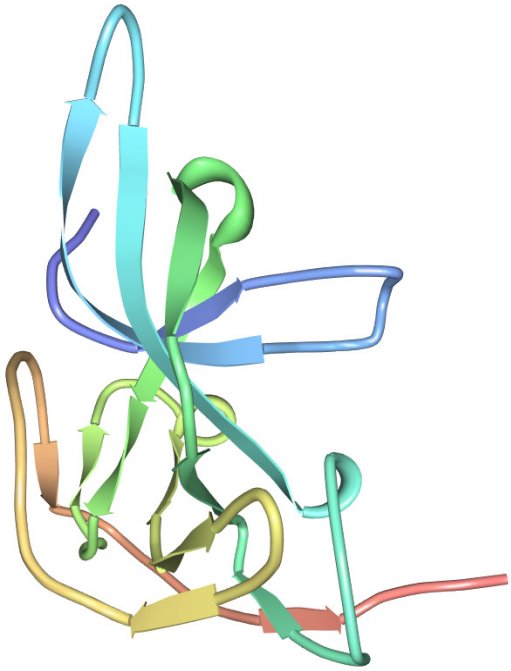


Figure 1.9: Crystal structure of the first bromo-adjacent homology (BAH) domain of chicken polybromo, as determined by X-ray diffraction. The structure is composed of 10 β-strands that form a barrel structure, connected by helices [PDB ID: 1W4S].{Oliver, 2005 #1055;Berman, 2000 #1292} Image manipulated using Protein Workshop.{Moreland, 2005 #1293}

The BAH domains are critical for proper ubiquitination of PCNA during post-replication repair.[43] Data suggests that the BAH domains are not capable of associating with the PBAF complex themselves, as a construct containing just the BAH domains was unable to bind with BAF200 or BRG1, both components of PBAF. Therefore, its role in PCNA ubiquitination is likely independent of the ATPase activity of PBAF.[43] The same study also found that the BAH domains were localized throughout the cytoplasm and nucleus,[43] although this may be a consequence of experimental conditions and the exact

amino acid sequences used in the tested construct. More experiments need to be done for clarification, as the de-localization of the BAH domains of BAF180 seemingly contradicts the data demonstrating that the BAH1 of BAF180 binds to histone H3.[9]

1.3.3 High Mobility Group Box

High mobility group (HMG) domains were first identified in 1973 by Goodwin et al., who noted the unusual electrophoretic mobility of these proteins in Triton-urea gels.[44-46] There are three structurally distinct classes of HMG domains: HMG-nucleosomal binding family (HMGN), HMG-AT-hook family (HMGA) and HMG-box family (HMGB). HMG-box domains mediate the interactions between certain proteins and DNA in either a sequence- or non-sequence-specific manner.[44,47-48] Generally, proteins with sequence-specific motifs bind to promoter DNA and those with non-sequence-specific bind to genomic DNA.[48] HMG box-containing proteins bind to the minor groove of B-type DNA at the entry/exit points of nucleosomes, resulting in distortion of the DNA strand; they also recognize already distorted DNA, such as four-way junctions, bulges, and kinks (Figure 1.11).[44-45,47,49] The DNA bending and unwinding that occurs when an HMG-box binding is a result of intercalation of hydrophobic residues into the minor groove of DNA; these hydrophobic residues are flanked by basic residues which bind to the phosphodiester bond of DNA, thus stabilizing the DNA:protein complex. Sequence-specific HMG-box domains usually only intercalate with DNA at one site.[47] This bending may aid in "opening" the DNA-histone complex to allow for transcription factors to bind.[44-45,50-51]

HMG-box domains also promote non-homologous end joining (NHEJ) by bending the ends of the DNA so that they have more energetically favorable interactions[52] and influence transcription by distorting the nucleosomal complex.[45] Additionally, when the HMG-box binds to bent DNA, the conformation changes such that the acidic tail of the domain is available to interact with histones, thus weakening the binding of the histone to the nucleosome.[47] This may be one mechanism by which BAF180 is involved in chromatin remodeling. PTMs may influence the specificity and binding capabilities of the HMG-box. In BAF180, two serines have been found to be phosphorylated (S1405 and S1453, isoform 1).[53] PTMs on other HMG-box-containing proteins have been known to alter the biological interactions of protein:protein and protein:DNA type.[45]

BAF180 has an HMG-box domain (50 amino acids in isoform 2: AAG48933.1) located on the C-terminus (Figure 1.3). The function of BAF180's HMG-box is unknown though several inferences can be made. The particular domain of BAF180 (of the HMGB type) likely binds in a sequence-specific manner, since it most closely resembles HMGB-containing transcription factors (has only one HMGB domain).[44,47-48] Though the general structure of HMG-box domains is three α-helices in an L-shaped configuration (Figure 1.10),[44,47] the HMG-box of BAF180 differs in that it has a shorted third α-helix and an extended basic N-terminus.[44,46] This difference likely plays a role in its sequence-specificity, as the N-terminal strand and the third helix (also called the 'minor wing') are responsible for the sequence specificity of an HMG-box.[47] The assembly of BAF180 into the PBAF complex has been demonstrated to be dependent on the HMG-box and/or the

C-terminus of BAF180.[43] It is possible that the PTMs present in the HMG-box impact the ability of BAF180 to assemble into the PBAF complex.[54] The HMG-box domain may affect the role of BAF180 in embryonic development, as other HMG-box containing proteins have been found to.

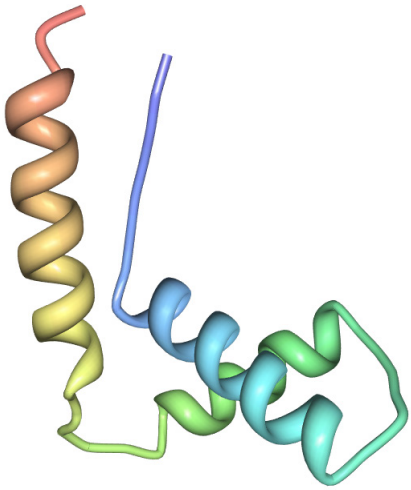


Figure 1.10: The crystal structure of HMG domain A of the HMGB1 protein. Note the L-shaped configuration of the α-helices. The blue helix is at the N-terminus and the red at the C-terminus [PDB ID: 1CKT].{Ohndorf, 1999 #1108;Berman, 2000 #1292} Image manipulated using Protein Workshop.{Moreland, 2005 #1293}

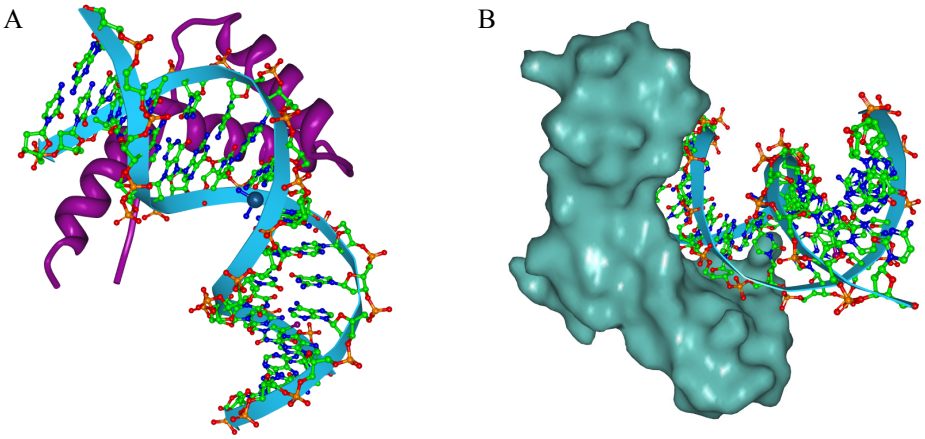


Figure 1.11: The HMG-box domain of the HMGB1 protein of *Rattus norvegicus* bound to cisplatin-modified DNA. (A) Ribbon view; (B) Surface view [PDB ID: 1CKT].{Ohndorf, 1999 #1108;Berman, 2000 #1292;Xu, 2009 #1085} Images manipulated using Protein Workshop.{Moreland, 2005 #1293}

BAF180 is known to be critical for proper heart development in human embryos (see Section 1.4.2.2 below) and normal development has been demonstrated to require the proper expression of various other HMG-box-containing proteins.[44] HMG-box-containing proteins have also been shown to be important for maintaining genomic stability in yeast; absence of several different proteins of this type were associated with increased genomic instability and hypersensitivity to DNA damage.[44] These same effects are seen in BAF180 KO cells (discussed in Section 1.4.4). This leads to the supposition that some of BAF180's roles in maintaining genomic stability and proper DNA damage repair may stem from its HMG-box. BAF180's roles in embryonic development and NHEJ may be a consequence of the HMG-box domain.

1.3.4 BAF180 Binding Hypotheses

The binding mechanism by which BAF180 binds to nucleosomes is currently unknown and is made even more complicated because BAF180 has three-types of domains (6 bromodomains, 2 BAH domains, 1 HMG-box). With regard to the bromodomains, recognition of an acetyl-lysine site may be a singular event or may be cooperative. Bromodomains have been shown to be able to recognize and bind more than one acetyl-lysine at a time.[14,24] The di-acetylated histone sequence has a stronger bond to the bromodomain than a single acetylated site. It is also possible that more than one bromodomain binds to the nucleosome at a time; they may recognize sites on the same histone, on multiple histones, or on adjacent nucleosomes. For example, the bromodomains might bind to more than one H3 protein (Figure 1.12A). This could be a result of sterics, which would prevent more than one bromodomain from binding to the

same histone. If only one bromodomain is bound at a time, the bromodomains may function as some sort of signaling mechanism; for example, bromodomain 1 binds to its specific acetylated lysine site and results in transcription of protein X. This hypothesis is supported by the large variety of PBAF subunits involved in maintaining cell state.[56-62]

There are a few possibilities for the overall binding mechanism of BAF180. Hypothesis 1: The HMG box binds to the DNA surrounding the histone core and BAF180 changes conformation such that the bromodomains are bound to their preferred acetylated lysine sites (Figure 1.12B). Hypothesis 2: The bromodomain recognizes the acetylated lysine site first and then the HMG box binds to the DNA surrounding the histone core, creating a stabilizing effect (Figure 1.12C). Again, it is possible that more than one bromodomain binds at a time. Hypothesis 3: the HMG box binds to something besides the nucleosome-associated DNA. This is less likely than the others, as the HMG box domain is found primarily in chromatin remodeling and transcription-related proteins.[44,47-48]

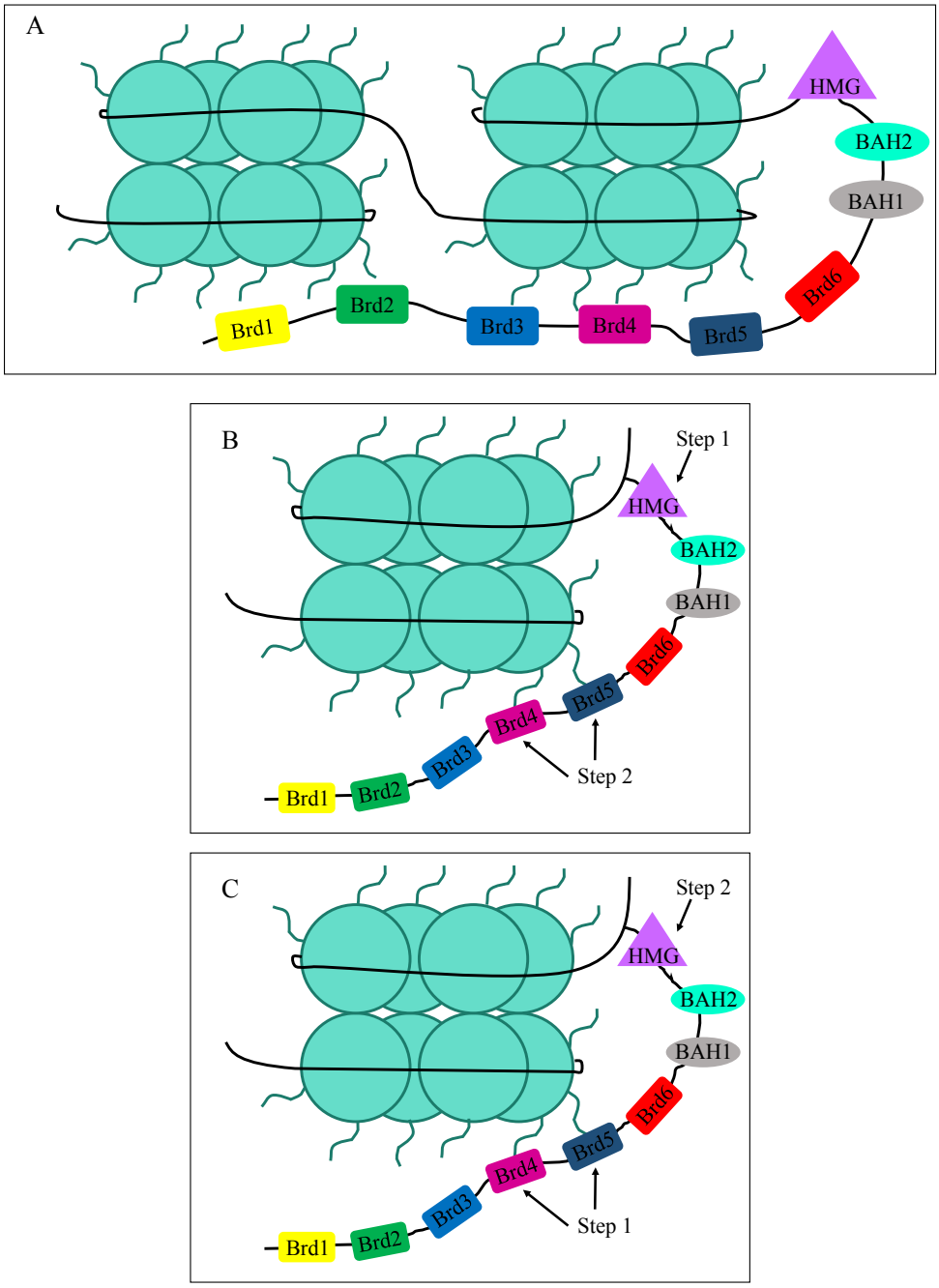


Figure 1.12: Diagram of possible binding mechanisms between BAF180 and nucleosomes. (A) The bromodomains might bind to more than one histone and/or more than one nucleosome at any given time. The pattern of bound bromodomains might have an influence on which gene is transcribed. (B) The HMG box of BAF180 may bind to the DNA surrounding the nucleosome first, followed by the binding of the bromodomain(s). Again, perhaps more than one bromodomain is bound at a time. (C) The bromodomains first bind to acetylated lysines and then the HMG box binds to the DNA, creating a stabilizing effect

1.4 Known Biological Functions of BAF180

As part of the PBAF complex, BAF180 is believed to be the chromatin-recognition subunit, allowing targeted gene transcription. In addition, BAF180 has numerous individual roles in development and DNA repair and it is often mutated in cancers (Figure 1.13).

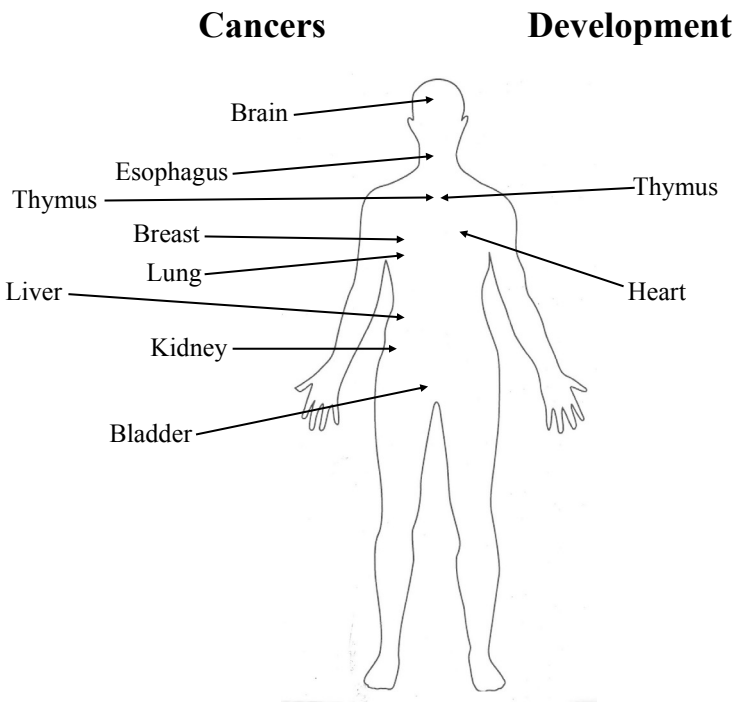


Figure 1.13: Visual listing of the roles of BAF180 in development and its presence in cancers.

1.4.1 PBAF and Transcription

The BAF (BRG-associated factor) and PBAF (polybromo-associated BRG1 factor) complexes control gene expression in the human body by remodeling chromatin to alter the accessibility of DNA to transcription factors. They is accomplished by sliding, displacing, or replacing nucleosomes.[56,58-59,62-65] The BAF and PBAF complexes are the

homologs of yeast SWI/SNF complexes and as such are generally referred to as SWI/SNF. Both complexes (BAF and PBAF) have a core catalytic subunit that possesses ATPase activity, the energy from which is used to remodel nucleosomes. In addition, they also contain various other subunits, some that are common to both complexes (such as BRG1) and some that are unique (such as BAF180) (Figure 1.14 and Table 1.4).[56,58-59,62-65]

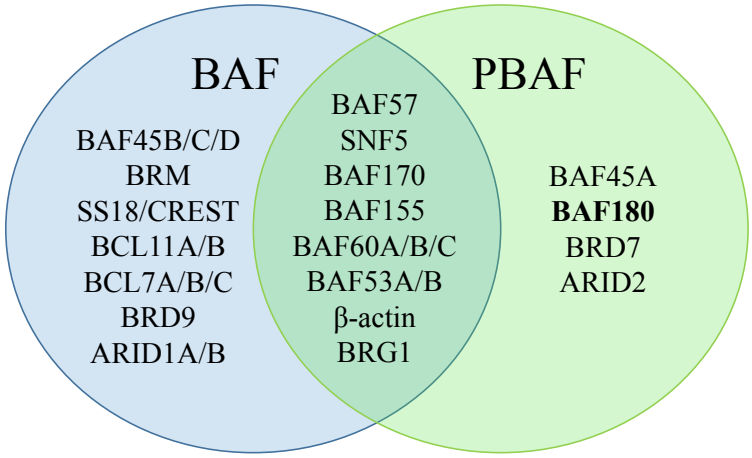


Figure 1.14: The unique and shared subunits of the BAF and PBAF complexes.[15]

It was recently reported that components of the SWI/SNF complexes are mutated in 20% of all cancers.[58] The BAF and PBAF complexes have distinct roles in gene transcription; for example, PBAF is required for the ubiquitination of H2AK119 at sites of DNA damage.[66] The exact roles of the individual subunits of the complexes is beginning to be understood. For example, BAF180 is believed to be the chromatin-recognition component. However, the precise protein(s) involved in binding to the DNA wrapped around nucleosomes have not yet been identified.

Table 1.4: The subunits found in the BAF and PBAF complexes and their individual domains.[15,39,67]

Subunit	Present in	Domains
BAF45B/C/D	BAF	Tandem PHD domains, zinc finger
BRM	BAF	DNA-dependent ATPase, bromodomain, HAS, BRK, SnAC
SS18/CREST	BAF	N-terminal conserved region
BCL11A/B	BAF	Zinc finger
BCL7A/B/C	BAF	N-terminal conserved region
BRD9	BAF	Bromodomain
ARID1A/B	BAF	ARID/BRIGHT, DNA binding domain
BAF57	Both	HMG box
SNF5	Both	SNF5 domain
BAF170	Both	Chromatin related domain, SWIRM, SANT, leucine zipper
BAF155	Both	Chromatin related domain, SWIRM, SANT, leucine zipper
BAF60A/B/C	Both	SWIB/MDM2 domain
BAF53A/B	Both	Actin fold
β-actin	Both	Actin fold, ATPase
BRG1	Both	DNA-dependent ATPase, bromodomain, HAS, BRK, SnAC
BAF45A	PBAF	Tandem PHD domains, zinc finger
BAF180	PBAF	Bromodomain, BAH, HMG-box
BRD7	PBAF	Bromodomain
ARID2	PBAF	ARID/BRIGHT, DNA binding domain

BAF180 interacts with at least BRG1 and BAF200 when assembled in the PBAF complex. The bromodomains and BAH domains were not able to associate with BRG1 or BAF200 on their own, but the HMG domain was able to.[43] This suggests that the HMG domain is the site through which BAF180 interacts with other components of the PBAF complex. This has interesting implications for the mutant BAF180 proteins in cancers. Most of the mutations result in the loss of all or part of the HMG domain, implying that these mutants are unable to associate and form the PBAF complex, which would impair transcription since the PBAF and BAF complexes are shown to interact

with nucleosomes differently. This hypothesis is in accordance with results from Gao et al. demonstrating that a mutant BAF180 protein (without the HMG-box) was unable to associate with other subunits of the PBAF complex.[68] This raises an important question: If the mutant proteins of BAF180 are expressed and not immediately degraded by proteases or the ubiquitination degradation pathway, and they cannot associate with PBAF, what are the mutant BAF180 proteins doing in the cell? Does the mutant bind with something else, causing a deleterious effect or is the problem merely that it does not associate with PBAF? And since BAF180 is the chromatin-targeting subunit of PBAF, if BAF180 doesn't associate with the other subunits of PBAF, maybe PBAF cannot bind to the DNA site of transcription, or at least it cannot pick the right site. The PBAF complex can still form in the absence of BAF180.[69-70] When BAF200 was knocked down, BAF180 levels also decreased, suggesting that BAF200 recruits BAF180 to the PBAF complex or influences its stability. When BAF180 was knocked down, levels of BAF200 did not change.[71] We still do not understand how this data relates to the binding of BAF180 to nucleosomes.

1.4.2 Development

1.4.2.1 Retinoic Acid Pathway

Retinoic acid (RA) mediated transcription is controlled by two nuclear proteins: a retinoic acid receptor (RAR) and a retinoid X receptor (RXR). By themselves, RAR and RXR are not active, but when they come together to form a heterodimeric complex, they are able to bind to specific DNA sequences called RA response elements (RAREs). These elements consist of two repeats of a conserved hexameric motif (PuG(G/T)TCA, where

"Pu" represents a purine base) separated by 1, 2, or 5 base pairs. For repeats separated by 2 or 5 bases (referred to as DR2 or DR5), the RXR subunit interacts with the 5' motif, while RAR subunit interacts with the 3' repeat. In DR1 elements, the interactions are reversed (Figure 1.15).[72]

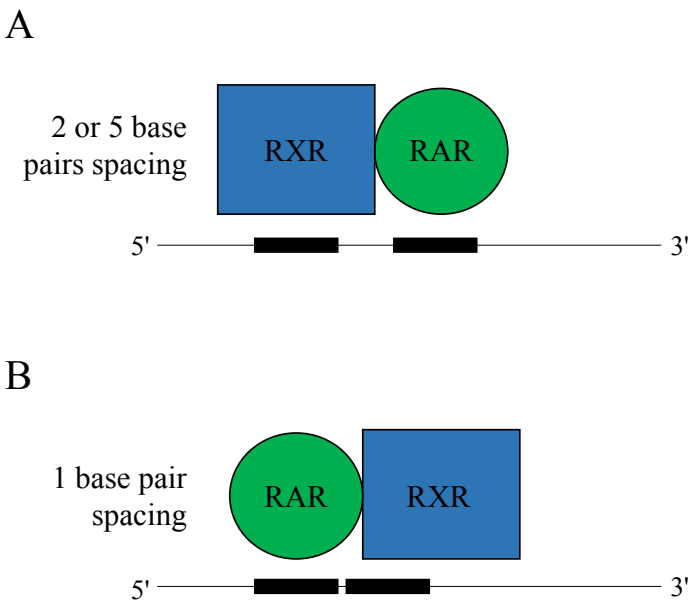


Figure 1.15: The RAR/RXR complex binds to a hexameric repeat in the DNA. If the spacing between repeats is 2 or 5 base pairs, the RXR protein binds on the 5' end of the sequence and the RAR protein binds on the 3' end. If the spacing is only 1 base pair, the orientation is flipped: the RAR protein binds at the 5' end and the RXR protein binds on the 3' end.

In the absence of retinoic acid, the RAR/RXR complex forms and retains binding to the RARE sequences, but performs a repressive role. The RAR/RXR complex recruits the co-repressors NCoR (nuclear co-repressor) and SMRT (silencing mediator for retinoid and thyroid hormone receptor) complexes, which contain histone deacetylase proteins, thus repressing transcription by compacting chromatin.[72-73] The full pathway is detailed in Figure 1.16.

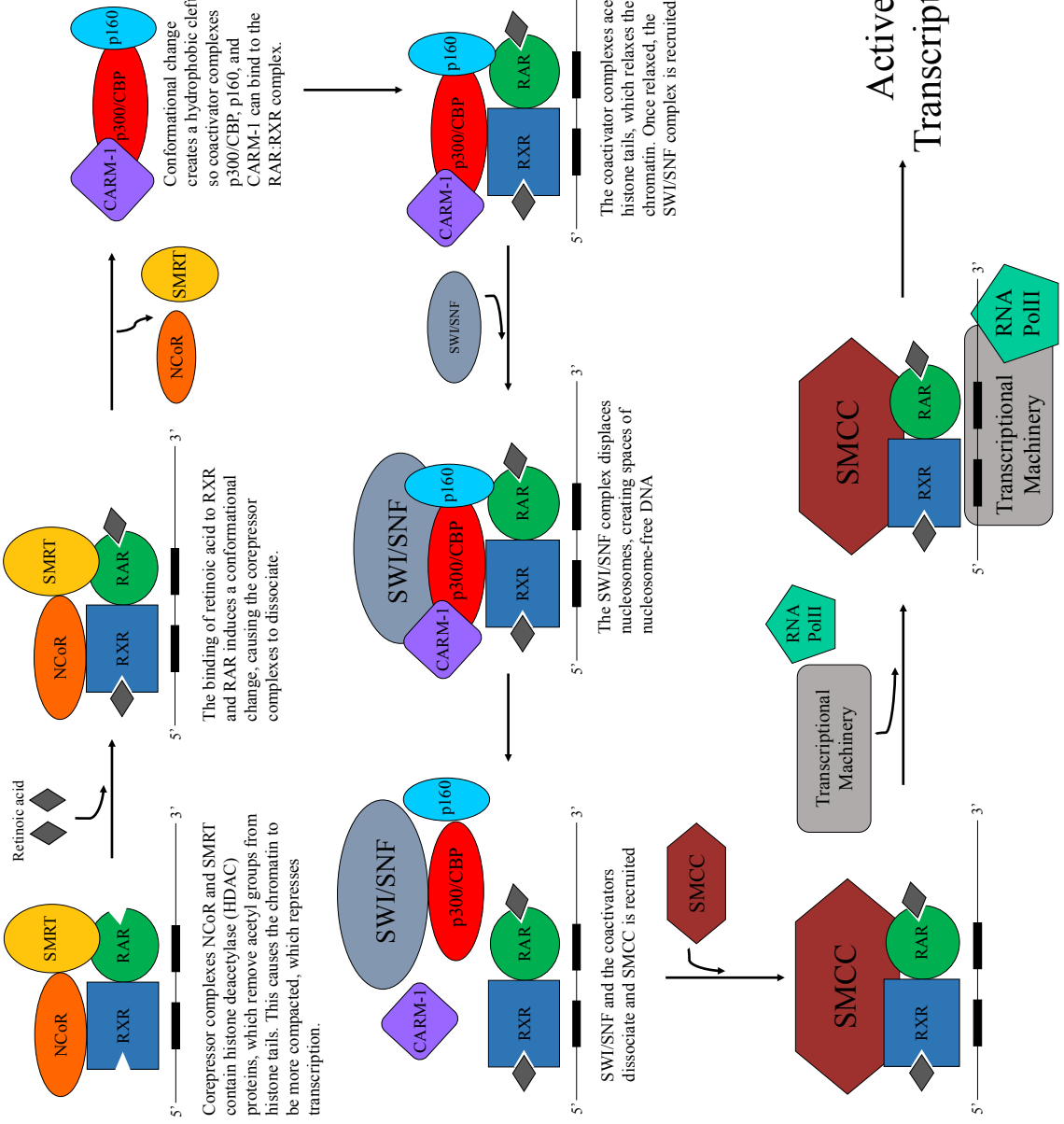


Figure 1.16: The retinoic acid-mediated transcription pathway. The presence of retinoic acid induces structural changes, allowing transcription factors (such as BAF180 in PBAF) to bind.

In the presence of retinoic acid, the RAR/RXR complex switches from a transcriptional repressor to a transcriptional activator. Retinoic acid is transported to the nucleus by CRABPII (cellular retinoic acid binding protein), where it can then bind to the RAR/RXR complex.[74] When retinoic acid binds to the ligand-binding site, RAR and RXR undergo a conformational change which increases the affinity of the complex for the DNA RARE sequences.[72] The conformational changes create a new hydrophobic cleft which causes the co-repressors to dissociate. The coactivators p160, p300/CBP, and CARM-1 now bind and proteins possessing histone acetyltransferase (HAT) activity are recruited. Once the histones have been acetylated, SWI/SNF complexes are recruited, sliding the nucleosomes and allowing for transcriptional machinery to bind to the gene promoter. The coactivators then dissociate and the SMCC complex (Srb and Mediator protein containing complex) is recruited. The SMCC complex then recruits the transcriptional machinery, such as RNA PolII.[73,75]

Studies have demonstrated that the retinoic acid pathway is crucial for heart development, as absence of retinoic acid in quail and mouse embryos resulted in impaired heart development.[76-77] The role of BAF180 in the retinoic acid pathway was elicited by Wang et al., who investigated the effects on heart development in BAF180 knockout mice. Mice which lacked BAF180 showed a decrease in the expression of CRABPII and RARβ2, two important genes involved in the retinoic acid pathway: BAF180 was demonstrated to be present at the promoter sites of these genes through ChIP (chromatin immunoprecipitation) assays.[78] More details can be found in the heart development section (Section 1.4.2.2).

1.4.2.2 Heart

Wang et al. first elicited the role of the BAF180 protein in heart development in 2004 when it was found that BAF180 knockout mice (KO) died mid-gestation. Further investigation revealed heart defects to be the likely cause. The absence of BAF180 resulted in severe hypoplasia (underdevelopment, thinning) of the cardiac ventricular free walls (outer walls of the heart) and ventricular septal defects (hole in the wall separating right and left ventricles).[78]

Several experiments were conducted to determine if the observed heart defects were a direct or an indirect result of BAF180 KO. The possibility of inappropriate cell death during development of the cardiac tissues was ruled out after experiments indicated that apoptotic cells were not present in the developing heart of either normal or BAF180 KO mice.[78] Incorrect cell differentiation was eliminated as a possibility after cardiac biomarker analysis revealed normal expression in BAF180 KO hearts.[78]

Microarray analysis revealed that the absence of BAF180 had an impact on the expression levels of numerous genes, many of which were structurally- or metabolically-related. In particular, two growth arrest-specific genes were up-regulated, supporting the theory that the heart defects may be a result of growth arrest and that loss of BAF180 is directly responsible for the observed cardiac defects.[78]

Building on the observation that placenta in BAF180 KO mice was abnormal, Wang et al. set out to determine if this could be the cause of the heart defects. Two types of experiments were performed to determine if the defective placenta had an impact on

the developing embryo's heart. Studies wherein the placenta was derived from wild type (WT) cells and the embryo was derived from mutant BAF180 KO cells demonstrated that the heart defects still occurred even when placental development was normal. The inverse experiment, wherein the placenta was derived from mutant BAF180 KO cells and the embryo was derived from WT cells, showed that no heart defects were present, even though the placenta was abnormal. This suggested that the placenta defects did not play a role in cardiac defects.[78]

Comparing the phenotypic similarities of BAF180, RXRα, and PPARγ null mice led to the investigation of BAF180's role in the retinoic acid pathway (RXRα and PPARγ are involved in the retinoic acid signaling pathway). The PBAF complex is known to mediate the transcriptional activity of RXRα and PPARγ *in vitro*,[79] so the researchers set out to determine if BAF180 specifically is required for the transcription of some RA-related genes *in vivo*. Quantitative RT-PCR indicated that elimination of BAF180 resulted in a 5-8 fold down-regulation of the retinoic acid transporting protein CRABPII, and a slight down-regulation in the expression of RARβ2, a retinoic acid binding protein. In normal cells, exposure to retinoic acid will increase the expression of RARβ2 (~100 fold) and CRABPII (5-8 fold). However, after treatment with retinoic acid, BAF180 KO cells exhibited no response for CRABPII expression and decreased effect on RARβ2 expression. ChIP analysis further linked BAF180 to the transcription of RARβ2 and CRABPII. BAF180 was found to be present at the RARβ2 and CRABPII promoters, where its occupancy level at the RARβ2 promoter increased in the presence of retinoic

acid. This data suggests that the observed heart defects in BAF180 KO mice may be partially linked to the role of BAF180 in the retinoic acid pathway.[78]

A later study by the same lab revealed that these heart defects might be due to improper coronary vessel formation.[80] Mice lacking BAF180 showed underdevelopment of coronary vessels in the epicardial and myocardial layers of the developing heart.[80] It is possible that the hypoplasia observed in their previous study may be related to this, as underdevelopment of the coronary vessels results in reduced blood circulation, impeding the growth of cardiac cells. Another study, by Singh et al., showed that the expression of BAF180, along with two other SWI/SNF subunits, was higher in the heart as compared to the rest of the embryo during early development, again supporting the critical role of BAF180 in cardiac development.[81]

Taken together, these studies strongly suggest that the cardiac defects observed in BAF180 KO mice are a direct result of the loss of BAF180, at least partially due to the role of BAF180 in coronary vessel development and the retinoic acid pathway.

1.4.2.3 Thymus

Proper immune system function is critical to everyday health. Improper immune function not only leaves the person more vulnerable to pathogens, but also makes them more susceptible to auto-immune diseases, wherein the body's own T cells attack specific tissues (examples include rheumatoid arthritis and Crohn's disease). The thymus is an organ in the lymphatic system which produces the T cells.[82] T cells target and destroy infected cells.[83-84]

Wurster et al. selectively deleted BAF180 in late development T cells to discern the effect on thymus and T cell development. A cre recombinase cassette was used to specifically delete the BAF180 gene in T cells, leaving normal BAF180 levels throughout the rest of the embryo, including the placenta and heart. This is important as previous studies demonstrated that lack of BAF180 caused placental and heart defects, causing embryos to die mid-gestation.[78] Due to the nature of the cre recombinase cassette, BAF180 is not deleted until later in development, allowing normal BAF180 levels in T cells during early embryonic development.[85]

While overall T cell and thymic development were unaffected, a few changes were observed in peripheral CD4 T cells. A general down-regulation of the CD44 antigen (a surface glycoprotein involved in cell-cell interactions, adhesion, and migration)[86] was seen, though it was not determined if this was a direct or indirect result of the loss of BAF180. Additionally, a small increase in the expression of a few cell cycle inhibitors was observed, corresponding with a slight decrease in the proliferative response to stimuli. A decrease in the expression of proteins involved in nucleosome assembly, transcription, and DNA binding was observed, along with an increase in expression of proteins involved in ribosome biogenesis and translation. Interestingly, the gene most affected by BAF180 deletion was IL-10, which was significantly up-regulated.[85] IL-10 is an anti-inflammatory cytokine produced in many types of immune cells. It can regulate the innate and adaptive immune responses by inhibiting T cell activation and differentiation in lymph nodes and suppress pro-inflammatory responses in tissues.[87] So overall, IL-10 hinders the immune response to pathogens. When IL-10 expression

increases as a result of BAF180 deletion, the immune response decreases. It was determined that BAF180 binds directly around the IL-10 locus, suggesting that BAF180 directly affects the expression levels of IL-10. As can be imagined, in the absence of BAF180, the PBAF complex is replaced by BAF complex, which likely has an impact on binding patterns around the IL-10 locus. When PBAF was replaced with BAF, increased acetylation of histones was observed, suggestive of increased transcription; further tests were not performed to investigate this intriguing data.[85]

1.4.3 DNA repair

Malfunctions in DNA repair and genomic instability are hallmarks of cancer. BAF180 has a multitude of roles in double-strand break repair[66] (via non-homologous end joining and homologous recombination) and post-replication repair.[43] In addition, BAF180 was found to be important for maintaining genomic stability and preventing aneuploidy.[88]

1.4.3.1 Genomic Instability

In mice, genomic instability was shown to be caused by a lack of cohesion between the centromeres of chromosomes. This lack of cohesion can also lead to aneuploidy, where a cell contains an abnormal number of chromosomes. When BAF180 was depleted in mouse embryonic stem cells, the distance between centromeres increased, suggesting less cohesion between them. This appears to be exclusive to centromeres, as the distances between telomeres and chromosomal arms showed no significant change. To determine the route by which BAF180 affects the centromeric cohesion in chromosomes, levels of other cohesion genes were tested in BAF180-

deficient cells. It was found that the levels of the cohesion genes were normal, indicating that BAF180 did not affect the transcription of these genes.[88]

Cancer-associated mutants of BAF180 were unable to restore cohesion to BAF180-depleted cells or reduce the cell's sensitivity to DMSO-induced DNA damage. Three different point mutations were introduced in Rsc2, a homolog of BAF180: Rsc2-H458P (which corresponds to H1205P in BAF180 isoform 9, located in second BAH domain); Rsc2-T67P (corresponding to T232P in BAF180, brd2); Rsc2-M280I (corresponding to M523I in BAF180, brd4). Yeast cells lacking Rsc2 are temperature sensitive and are sensitive to DMSO-induced DNA damage. Mutant Rsc2-H458P showed patterns of sensitivity similar to Rsc2 KO cells, but mutants Rsc2-T67P and Rsc2-M280I showed reduced sensitivity, indicating that these mutants were able to partially restore normal function. None of the mutants were able to restore cohesion between centromeres.[88]

Insufficient cohesion between centromeres can have several repercussions, including mis-segregation of chromosomes, formation of micronuclei, and lagging chromosomes. Mis-segregation of chromosomes during mitosis results in acentric chromosomes, chromosomes which lack a centromere.[89] Because they lack a centromere, they cannot attach to the spindle during mitosis and as such are lost during replication.[90] An acentric chromosome that has become enclosed by a nuclear membrane is called a micronuclei;[90] micronuclei are commonly seen in cells with defective cohesion. Lagging chromosomes are another common characteristic of centromeric cohesion. Lagging chromosomes are chromosomes that are left near the spindle equator after anaphase (due

to improper segregation), while other chromosomes move toward the spindle poles. These can also become micronuclei.[91-92] Both acentric chromosomes and lagging chromosomes can cause aneuploidy in a cell,[93] wherein the individual cell has a different number of chromosomes than a wild type cell. Aneuploidy, micronuclei, and lagging chromosomes were each identified in BAF180 deficient cells.[88]

Loss of cohesion negatively impacts recombination-based DNA repair. Therefore, loss of cohesion and chromosomal instability may be two ways in which BAF180 helps prevent tumorigenesis. When BAF180 is depleted from a cell, cohesion between centromeres greatly decreases, and as a result, chromosomal instability increases.[88] This lack of cohesion and increase in chromosomal instability suggests that BAF180 may have a role in homologous recombination (HR), as other proteins, such as BRCA1, have dual roles in cohesion and HR.[63]

1.4.3.2 DNA Double-Strand Break Repair

BAF180 also plays a role in double-strand break (DSB) repair by silencing transcription near the break sites. When transcription is not silenced, DSB repair is delayed; since most irradiation-induced DSBs are repaired rapidly,[66] this delay can lead to chromosomal translocations, as well as deletions.[52] PBAF was found to silence transcription around DNA breaks. When BAF180 was depleted in reporter cells, transcription of the reporter gene (they used yellow fluorescent protein) was not silenced, suggesting that BAF180 is needed to silence transcription around DNA breaks. If transcription isn't repressed, rapid DSB repair is impaired.[66] Additionally, when BAF180-depleted cells were exposed to irradiation, the non-homologous end joining

(NHEJ) DNA repair pathway, a key mechanism by which DSBs are repaired,[66] was compromised.[94]

By treating siBAF180 cells with DRB (5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside), a transcription inhibitor, Kakarougkas et al. determined that the role of BAF180 in DSB repair is dependent on active transcription. In siBAF180 cells, DNA damage levels were high at 20 minutes post-IR, as indicated by a high number of γH2AX; when BAF180 was re-introduced, levels of γH2AX were similar to control cells.[94] γH2AX is used as a biomarker to measure the quantity of DNA double-stranded breaks.[95] Wherever double-stranded breaks form, H2AX is immediately phosphorylated, becoming γH2AX; therefore, the presence of γH2AX can be used to determine the extent of double-stranded breaks in a 1:1 ratio.[95] The phosphorylation of γH2AX is the first step in recruiting DNA repair proteins.[95] When siBAF180 cells (which were unable to stop transcription) were treated with DRB (which stops transcription), the DNA damage levels decreased to around that of the control cells, indicating that stopping transcription restored DNA damage repair function. This means that when transcription is stopped, BAF180 is not needed for double strand break repair.[94]

A protein's function can be influenced by the presence of PTMs. S963 (number corresponding to isoform 8 of BAF180) was found to be phosphorylated by ATM (ataxia telangiectasia mutated).[96] ATM is a kinase that is critical for DSB repair.[96] This phosphorylation site is located between the sixth brd and the first BAH domain. Two mutant constructs (a phosphomutant S963A and a phosphomimic S963E) were created by Kakarougkas et al. to determine if phosphorylation was critical to BAF180's role in DSB

repair. When the phosphomutant was added to siBAF180 cells, no significant change was seen, indicating that the phosphomutant did not help DNA repair, meaning that phosphorylation is important to BAF180's role in DNA DSB repair. When the phosphomimic was added, levels of damaged DNA were similar to control cells, indicating that proper DNA repair function was restored. This suggests that phosphorylation of BAF180 at S963 is a crucial step in DSB repair.[94]

Some of the tumor suppressor properties of BAF180 may be related to its functions in DNA repair. Cancer-associated mutations were unable to restore normal DSB repair function in BAF180 KO cells,[94] suggesting that cancer cells with mutations in BAF180 have a dysfunction in mediating DSB repair. Mutations which eliminated the phosphorylation site S963 were also unable to restore normal DSB repair functioning.[94]

1.4.3.3 Ubiquitination of PCNA

The BAH domains of BAF180 are required to ubiquitinate PCNA (proliferating cell nuclear antigen).[43] PCNA regulates replication past sites of DNA damage and its ubiquitination is a necessary step in re-priming past damaged DNA forks during post-replication repair (PRR).[97]

When BAF180 was depleted, cells were unable to produce the proper levels of ubiquitinated PCNA, which leads to an inability to replicate past DNA lesions.[98] After UV-irradiation, BAF180-depleted cells had far lower levels of chromatin-bound PCNA (both ubiquitinated and non-ubiquitinated) and decreased levels of Rad18, the E3 ligase that ubiquitinates PCNA, when compared to control cells (where BAF180 was present at

normal levels). It has been reported that BAF180 does play a role in regulating the cell cycle,[99] therefore experiments were conducted to determine if the reduced levels of chromatin-bound PCNA were an indirect consequence of BAF180 impacting the cell cycle. Ubiquitination of PCNA and its binding to chromatin occurs during the S-phase; post-IR, the proportion of cells in S-phase actually increased, indicating that the lower levels of chromatin-bound PCNA and Rad18 are not an indirect consequence of cell cycle changes occurring upon depletion of BAF180.[98] This suggests that BAF180 is directly responsible for the observed changes in chromatin-bound PCNA and Rad18, and therefore indirectly responsible for the decreased ubiquitination levels. The BAF180 orthologue Rsc2 was found to be located near replication forks,[98] strongly suggesting that BAF180 may be as well.

Additionally, PCNA remained ubiquitinated longer in flag-BAF180 FL cells than in control cells during S-phase;[43] there was very little ubiquitinated PCNA detected in G1 and G2 phase cells. Interestingly, the levels of BAF180 were highest in S-phase cells,[43] suggesting that the S-phase specific function of BAF180 (i.e. ubiquitination of PCNA) requires higher levels of BAF180 than other phases, which has interesting implications since PBAF (and hence BAF180) is involved in transcription and has not been found to be limited to one phase of the cell cycle. It would be interesting to perform FACS (fluorescence-activated cell sorting) analysis on several components of the PBAF and BAF complexes to see how the protein levels change throughout the different phases of the cell cycle.

To determine which domain(s) of BAF180 are involved in PCNA ubiquitination, Niimi et al. created flag-tagged constructs of the different domains (bromodomains, the BAH domains, and the HMG domain). They found that levels of ubiquitinated PCNA between full-length flag-tagged BAF180 and flag-BAHs were similar; cells with the flag-brds and flag-HMG constructs were unable to ubiquitinate PCNA, which suggests that the BAH domains are required for PCNA ubiquitination and that the other domains of BAF180 are not necessary. Additionally, BAF180 did not need to be in complex with PBAF in order for the BAH domains to facilitate ubiquitination.[43]

1.4.3.4 DNA Repair Conclusion

The tumor suppressor properties of BAF180 might stem from its roles in various DNA repair mechanisms, as discussed above. Cancer cells with a mutation in or decreased expression of BAF180 could be compromised with regards to DNA repair in four ways:

(1) **Genomic instability:** An increased distance between centromeres leads to an increase in genomic instability, which may in turn prevent recombinatorial-based DNA repair mechanisms.[88]

(2) **Inability to ubiquitinate PCNA:** Ubiquitination of PCNA is a key step in the post-replication repair pathway. The BAH domains of BAF180 are required for proper ubiquitination levels. Lack of the BAH domains likely leaves much of the DNA unrepaired.[43,98]

(3) **Silencing of transcription at DNA double-strand breaks:** At sites of active transcription near DNA DSBs, BAF180 is needed for the proper repair of these breaks. BAF180 silences transcription at sites of DSBs, allowing for the rapid repair of damaged DNA.[94]

(4) **HMG box and NHEJ:** As mentioned in an above section (Section 1.3.3), an HMG-box binds to DNA and facilitates NHEJ by bending the ends of the DNA. Assuming this domain is the DNA-binding component of BAF180, deletion of this box likely hinders NHEJ (and transcription) because BAF180 can no longer bind to DNA. Many cancer mutants lack the HMG box.[44,52]

1.4.4 p21 and the Cell Cycle

BAF180 has been shown to influence the cell cycle progression in many types of cells. Evidence has demonstrated that the loss of BAF180 caused cells to replicate and grow faster than control cells (see Section 1.5.2 on ccRCC). Reintroduction of BAF180 into these cells reduced cell proliferation and induced G1[100] and G2[99] arrest, showing a direct link between BAF180 and cell cycle progression. The role of BAF180 in G1 cell cycle arrest is due, at least in part, to its regulation of p21 expression. In the absence of BAF180, the mRNA levels of p21 decreased, resulting in fewer cells in G1 arrest.[100] p21, also known as cyclin-dependent kinase inhibitor 1, is used by cells to enter G1 and G2 phase arrest.[39] Flow cytometry and subsequent cell cycle analysis revealed that, upon re-introduction of BAF180, more cells were entering G1 arrest than control cells which lacked BAF180.[100] BAF180 has also been found to be critical for cellular senescence.[101]

Treatment with nutlin-3, which is used to induce p53-dependent transcription,[101-103] usually leads to a substantial increase in p21 levels; however, when BAF180 was knocked down, the levels of p53-induced p21 were significantly lower.[101] Chromatin immunoprecipitation experiments demonstrated that BAF180 regulates the expression of p21 directly by binding to its promoter *in vivo*.[100]

Cellular damaging agents, such as TGF-β and γ-irradiation, have been shown to up-regulate p21 expression, causing subsequent entry of the cell into G1 or G2 arrest, respectively (Figure 1.17). Their effect on up-regulation of p21 was diminished in BAF180-deficient cells. Interestingly, while the up-regulation of p21 was similarly diminished in both TGF-β and γ-irradiation treated cells, there was no effect on G2 arrest (Figure 1.17),[100] indicating that BAF180's role in G2 arrest may be through another pathway.[99] The role in p21 expression is believed to be part of the mechanism by which BAF180 influences the cell cycle. When BAF180 is mutated, it cannot bind to the p21 promoter and therefore G1 cell cycle arrest is prevented. Interestingly, a recent study by Chowdhury et al. showed that changes in BAF180 expression did not affect p21 levels in ccRCC cell lines A704 or 786-O.[68] Xia et al.'s study was performed in breast cancer cells[100] and Burrows et al., whose data supports that of Xia et al., performed their analysis in human fibroblasts and mammary cells.[101] It is plausible that the effect of BAF180 on p21 is tissue specific.

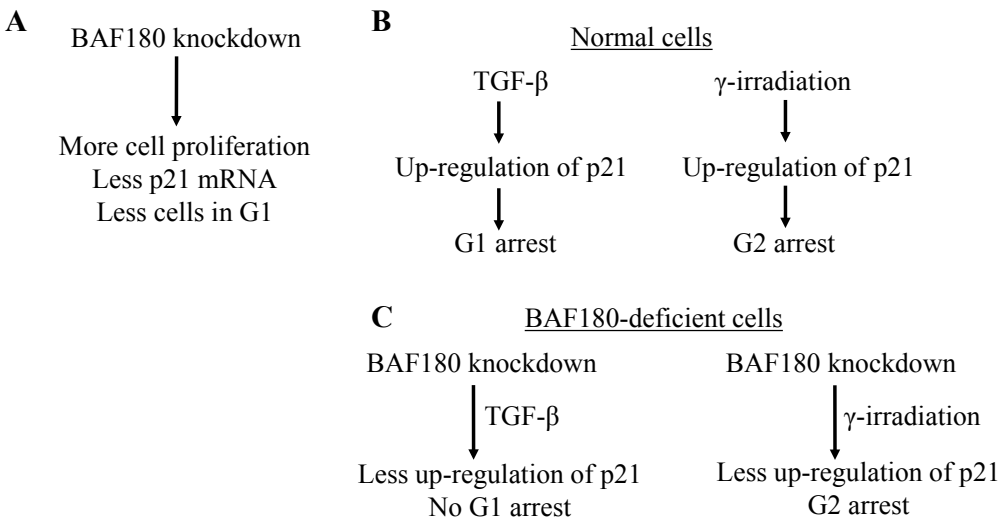


Figure 1.17: Diagram of the role of BAF180 in p21 expression and cell cycle arrest. (A) When a cell is deficient in BAF180, through mutation or knockdown, there is deregulation of the cell cycle, resulting in a high rate of cell proliferation. Additionally, less p21 is produced, as BAF180 is required for its transcription. (B) and (C) Flow chart showing how different treatments, TGF-β and γ-irradiation, affect the cell cycle. In normal cells, upon treatment, there is an increase in p21 expression, resulting in G1 and G2 phase arrest, respectively. In BAF180-deficient cells that are treated with TGF-β, there is less p21 and no G1 phase arrest. But if a cell is exposed to γ-irradiation, G2 arrest can still occur, though there is less p21.[100]

Expression of p21 is regulated primarily by p53;[104] however, p21 is also known to be regulated by the retinoic acid (RA) pathway in a manner independent of p53.[101,105] BAF180 binds to the promoters of CRABPII and RARβ2, both key components of the RA pathway; when there is less BAF180, there is less CRABPII and RARβ2.[78] So BAF180 could affect p21 in an indirect manner by affecting the retinoic acid pathway. Additionally, BAF180 is known to bind to the p21 promoter, thereby affecting p21 transcription directly.[100] So BAF180 may influence p21 transcription in an indirect and direct manner (Figure 1.18).

In bladder cancer cells, *PBRM1* was found to be required for G2 phase cell cycle arrest by inhibiting the expression of cyclin-B1.[99] This does not contradict the results that

Xia et al. published, as they did not examine cyclin-B1; they focused solely on p21 and p53.

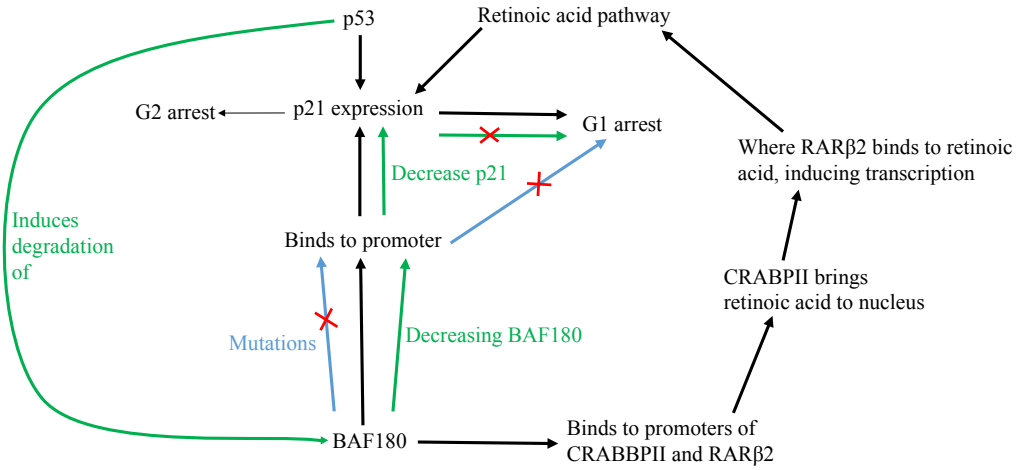


Figure 1.18: Plausible overall pathway for the regulation of p21 expression by BAF180 and p53.[74,78,100-101] For clarity, different colored arrows have been used to denote the different consequences. Blue arrows demonstrate the consequences of BAF180 mutations, while green arrows show the consequences of decreasing the amount of BAF180 in the cell.

Additionally, BAF180 was found to be required for replication senescence,[101] a state different from cell cycle arrest (temporary); in senescence, cell growth is permanently stopped.[106] Once again, the role of BAF180 in senescence is related to its role in regulating the expression of some p53-controlled genes.[101]

1.4.5 p53 Induced Degradation of BAF180

BAF180 levels have been shown to be reduced upon activation of p53 by nutlin-3.[102] This activation of p53 led to a decrease in BAF180 protein levels. The reduction of BAF180 levels was demonstrated to be at the protein level, as RT-PCR revealed little change in the mRNA levels. Experiments showed that proteasomes were responsible for the degradation of BAF180, as the treatment of cells with proteasome inhibitors reduced

the level of degradation (i.e. more BAF180 remained). Further experimentation determined that an E1 ubiquitin ligase is responsible. Therefore, BAF180 degradation is controlled by a p53-dependent ubiquitin related pathway.[102] Connecting this to section 1.4.4 on p21, the reduction of BAF180 by a p53-dependent pathway leads to a decrease in p21 expression and a loss of G1 phase cell cycle arrest (Figure 1.18).

1.4.6 Psychiatric Disorders

Medications for schizophrenia and bipolar disorder have targeted the same gene for more than 60 years. These treatments are only effective for a small proportion of patients.[107] In recent years, the lack of effective treatment options has led to numerous genome-wide association studies. Several of these studies identified *PBRM1* as a significant risk locus for schizophrenia, bipolar disorder, major depressive disorder, and psychosis.[107-117] A study by McMahon et al. discovered that *PBRM1* is overexpressed in the prefrontal cortex of people with bipolar disorder.[113] Unfortunately, the clinical impact of these discoveries remains undetermined.

1.4.7 HIV and Tat protein

During an HIV infection, the Tat protein promotes HIV virus replication and controls transcription occurring at the LTR promoter (long terminal repeat: a type of retroviral promoter).[21,53,118-119] At the LTR, PBAF is recruited by the Tat protein, rearranging nucleosomes to promote transcription.[120] BAF180 interacts with the Tat protein through acetylation sites at K50/K51, resulting in suppression of Tat protein expression.[121-122] BAF180 association with Tat was increased in the presence of p300, as shown in co-immunoprecipitation of BAF180 and Tat.

1.5. Cancers

Since Varela et al. published that *PBRM1* is mutated in ~40% of clear cell renal cell carcinoma cases,[1] the number of studies relating *PBRM1* mutations to cancers has exponentially increased. *PBRM1* is now linked to more than 30 types of cancers (Table 1.5).[123-125] Mutations of every type (frameshift, substitution (synonymous and missense), and truncation) have been identified, though truncation mutations (presence of a premature stop codon) are the most frequent (Figure 1.19).[123-125] Mutations have been seen to affect all nine domains, though mutations are also found between domains. Figure 1.20 displays the domain organization of wild-type BAF180 (A) and several mutants seen in cancers (B). It has been speculated that cancer-associated mutations of BAF180 may impair the ability of the PBAF complex to bind to nucleosomes, as BAF180 is believed to be the histone targeting component. Each of the six bromodomains of BAF180 are capable of binding to acetylated lysines present on histones, though it is unknown whether this occurs singularly or in tandem. Deletion of any of the bromodomains would affect the ability of BAF180 to bind histones, either by weakening the binding interactions or eliminating them completely (depending on the severity of the truncation). The inability of BAF180 to bind precludes the other subunits of the PBAF complex from being recruited; therefore, PBAF-mediated transcription cannot occur.

Discussed in upcoming sections are some of the cancers for which specific studies on *PBRM1* (BAF180) have been conducted. Most of the studies published have noted that *PBRM1* was mutated, but they did not explore further. Therefore, an overall discussion of BAF180 in cancers is not possible.

Table 1.5: The frequency of *PBRM1* alterations in different types of cancers. *PBRM1*, the gene for BAF180, has been found to be mutated in more than 30 types of cancers. Data was collected from cBioPortal.[123-124]

Cancer	Frequency of alterations in *PBRM1*
Clear cell renal cell carcinoma (ccRCC)	39%
Neuroendocrine Prostate Carcinoma	24%
Cholangiocarcinoma	18-23%
Cutaneous Squamous Cell Carcinoma	10%
Breast Cancer	2-14%
Bladder Cancer	2-10%
Uterine Carcinomas	4-9%
Stomach Adenocarcinoma	3-8%
Diffuse Large B-Cell Lymphoma	3-8%
Esophageal Carcinoma	2-8%
Pancreatic Carcinomas	1-7%
Prostate Adenocarcinoma	1-7%
Lung Carcinoma	2-6%
Melanomas	1-6%
Mesothelioma	5%
Papillary RCC	4%
Colorectal Adenocarcinoma	3-4%
Head and Neck Carcinoma	2-4%
Thymic Carcinomas	2-4%
Hepatocellular Carcinoma	1-4%
Thyroid Carcinomas	3%
non ccRCC	3%
Gallbladder Carcinoma	3%
Cervical Squamous Cell Carcinoma	3%
Sarcoma	1-3%
Adenoid Cystic Carcinoma	2%
Ovarian Carcinoma	2%
Neuroblastoma	2%
Chromophobe Renal Cell Carcinoma	2%
Gliomas	1-2%
Adrenocortical Carcinoma	1%
Chronic Lymphocytic Leukemia	1%
Pediatric Ewing Sarcoma	1%
Multiple Myeloma	1%

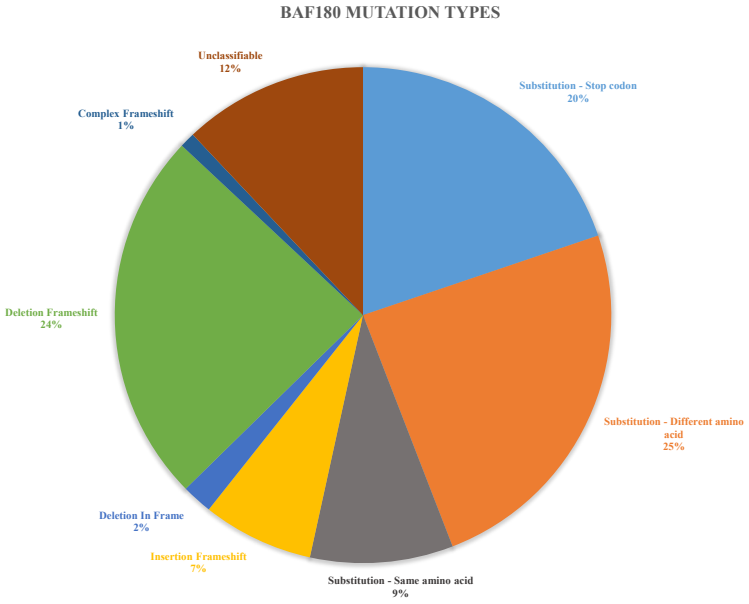


Figure 1.19: Chart of the frequency of different types of mutations of *PBRM1* (BAF180) in cancers. Data was collected from COSMIC database.[125]

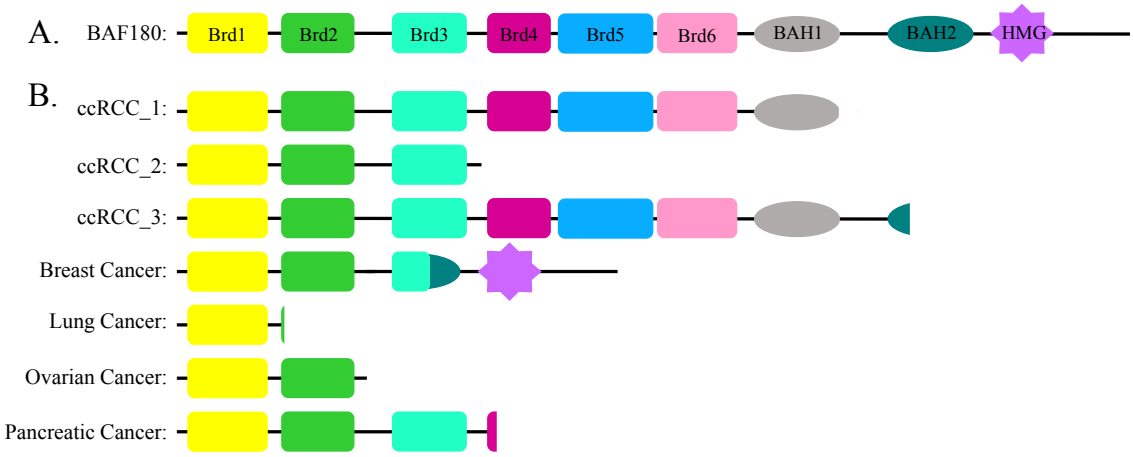


Figure 1.20: (A) The domain organization of wild-type BAF180 and (B) various truncated mutant proteins observed in cancers.

1.5.1 Cell Proliferation

The effect of BAF180 on cell proliferation appear to be cell line specific. In some cases, knockdown of BAF180 resulted in increased proliferation[1,68-69,99-100,126] while in others, knockdown resulted in decreased proliferation[68,127] or no effect[1,68-69] (Table 1.6). Similarly, re-expression of BAF180 into BAF180-deficient cell lines showed either decreased proliferation[68-69,99-100,126-127] or no had effect on proliferation[68] (Table 1.7). For example, Varela et al., Gao et al., and Xiao et al. demonstrated that knockdown of BAF180 results in increased proliferation in cell line 786-O,[1,68,126] but Murakami et al. reported that BAF180 knockdown decreased proliferation in this cell line,[127] suggesting that more studies need to be conducted in order for any true conclusions to be drawn.

Interestingly, Gao et al. found that in some cell lines, for re-expression or knockdown of BAF180 to have an effect on cell proliferation, the cells needed to have BRG1, the ATPase subunit of the PBAF complex.[68] Murakami et al. found that cells needed HIF1α in order for BAF180 to have a tumor-suppressor effect[127] though Gao et al. did not find any correlation to levels of HIF factors.[68]

Gao et al. demonstrated that a tumor-associated BAF180 mutation (missing the HMG-box and C-terminus) was not able to suppress proliferation in A704 cells (kidney adenocarcinoma cells), but was able to suppress growth in Caki2 cells (clear cell renal cell carcinoma cells). They also showed that a mutant lacking the first two bromodomains was able to reduce proliferation, though a mutant lacking all six bromodomains was not able to.[68] Along that line, a recently published study by Porter and Dykhuizen demonstrated that single point mutations in BAF180 were able to abrogate its tumor-

suppressive effects. Mutations in the first, second, fourth, and fifth bromodomain were unable to reduce proliferation in Caki2 cells, though wild type BAF180 could (Section 1.3.1.2).[23]

Table 1.6: Differing results on the effect of BAF180 knockdown on cell proliferation. The origin of the cell lines are as follows: bladder carcinoma: UM-UC-3; renal cell carcinoma: ACHN, 786-O, SN12C, TK10, UMRC2, KC-10, 769-P, RCC10, A498, A704; Clear cell renal cell carcinoma: Caki1

Effects of BAF180 Knockout on Cell Lines			
Knockdown BAF180, increased proliferation			
Author	Cell Line	Cell line +/- for BAF180	KD method
Huang[99]	Mice in vivo injection of UM-UC-3	+	transfection, injection into mice
Varela[1]	ACHN	+	siRNA
Varela[1]	786-O	+	siRNA
Varela[1]	SN12C	+	siRNA
Varela[1]	TK10	+	siRNA
Gao[68]	Mice in vivo implant of 786-O	+	lentiviral, implant into mice
Xiao[126]	ACHN	+	miRNA targeting PBRM1
Xiao[126]	786-O	+	miRNA targeting PBRM1
Knockdown BAF180, decrease proliferation			
Gao[68]	UMRC2	+	shRNA
Murakami[127]	786-O	+	shRNA
Murakami[127]	KC-12	+	shRNA
Murakami[127]	769-P	+	shRNA
Murakami[127]	RCC10	+	shRNA
Murakami[127]	A498	+	shRNA

Knockdown BAF180, no effect on proliferation			
Author	Cell Line	Cell line +/- for BAF180	KD method
Gao[68]	786-O (ex vivo)	+	shRNA
Gao[68]	SN12C	+	
Chowdhury[69]	Caki1	+	
Chowdhury[69]	A498	+	lentiviral
Varela[1]	A704	-	siRNA

Table 1.7: Differing results of the effect of BAF180 re-expression on cellular proliferation. The origin of the cell lines are as follows: bladder carcinoma: UM-UC-3, 5637; renal cell carcinoma: ACHN, 786-O, A704, SKRC20, SLR24; Clear cell renal cell carcinoma: Caki2, RCC4; breast cancer: HCC1143; endometrial adenocarcinoma: EJ.

Effects of Reintroduction of BAF180 into Cell Lines

Re-express BAF180, decrease proliferation			
Author	Cell Line	Cell line +/- for BAF180	Method
Xia[100]	HCC1143	-	Transfection
Gao[68]	A704	-	Lentiviral
Gao[68]	Caki2	-	Lentiviral
Chowdhury[69]	A704	-	Lentiviral
Chowdhury[69]	Caki2	-	Lentiviral
Huang[99]	UM-UC-3	+	Up-regulation
Huang[99]	EJ	+	Up-regulation
Huang[99]	5637	+	Up-regulation
Xiao[126]	ACHN	+	miRNA targeting *PBRM1*
Xiao[126]	786-O	+	miRNA targeting *PBRM1*
Re-express BAF180, no effect on proliferation			
Gao[68]	SKRC20	-	Lentiviral
Gao[68]	SLR24	-	Lentiviral
Gao[68]	RCC4	-	Lentiviral

It should be noted that Varela,[1] Gao,[68] Xiao,[126] Chowdhury,[69] and Murakami[127] performed their studies in renal cells, while Xia[100] and Burrows[101] used mammary cells, Burrows[101] used fibroblasts, and Huang used bladder cells.[99] The discrepancies in BAF180's tumor suppressor capabilities between different cell lines should be a focus of future studies; experiments in all these cell lines using a singular technique performed by the same lab would go a long way towards addressing this disagreement. The reasons behind the different proliferation responses of cell lines to BAF180 is still unknown, though Gao et al. presents several plausible explanations.[68]

1.5.2 Clear Cell Renal Cell Carcinoma

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cell carcinoma (~92% of cases).[128-129] Of these, ~40% of cases carry a mutation in the *PBRM1* gene,[1] though mutations of *PBRM1* are rarely found in other types of renal carcinomas.[130] This makes it the second most commonly mutated gene in ccRCC, after *VHL* (mutated in 75-80% of cases).[131-132]

In cancer, there are "driver" and "passenger" mutations. A "driver" mutation is considered to promote tumorigenesis, whereas a "passenger" mutation does not contribute to cancer development.[133-138] Additionally, a "driver" gene is one in which mutations occur at a rate higher than expected by chance alone.[131] Loss of *PBRM1* is considered to be a driver event in ccRCC.[128,131,139]

Interestingly, in ~90% of ccRCC cases, the 3p arm of one copy of chromosome 3 is deleted, which encompasses four tumor suppressor genes: *VHL*, *PBRM1*, *SETD2*, and

BAP1.[128,131-132,139-143] This deletion results in loss of heterozygosity. Knudson's hypothesis proposes that for tumor suppressor genes, "two hits" are required to bring about cancer. In this case, the "first hit" would be loss of heterozygosity (LOH). The "second hit" is the mutation of the remaining copy of the gene; this mutation may result in the production of a protein which has lost its function or is only partially functional, or result in loss of protein expression altogether.[23,68-69,131,144] *PBRM1*, as a tumor suppressor gene, follows this "two hit" approach.[139]

1.5.2.1 Mutations in BAF180

Mutations in the *PBRM1* gene are found in approximately 40% of ccRCC tumors.[1] Of these, approximately 82% of them are truncating mutations,[145] either by frameshifts or nonsense mutations. Interestingly, mutations in BAF180 are mutually exclusive with mutations of BAP1 (BRCA1-associated protein 1), a nuclear deubiquitinase protein.[139-140,144,146-148] Most mutations are thought to be destabilizing, resulting in no protein expression. Some mutations might be expressed but are predicted to have detrimental functional effects.[131,140,144-145,149]

In one study by Pena-Llopis, 53% of samples tested were negative for *PBRM1*. Of these, 90% had a mutation in the gene. In the 47% of samples that were positive for *PBRM1*, 90% had wild-type BAF180/*PBRM1*.[144] It should be noted that the researchers used an antibody recognizing an epitope on the C-terminus, which is often absent in BAF180 mutant proteins.[144] Actually, most publications use the same antibody from Bethyl Labs (Bethyl A301-591A), which recognizes an epitope at the C-terminus. It is possible that the antibody fails to recognize the mutant proteins, as the C-terminus is

often missing from the mutant BAF180 proteins. A study by da Costa et al. used an antibody that recognized an epitope on the N-terminus of BAF180 (Sigma-Aldrich HPA015629) and showed that in 112 immunostained ccRCC samples, 34 (30.4%) showed negative expression and 78 (69.6%) showed positive expression of BAF180.[150] More experiments need to be conducted using both N- and C-terminus recognizing antibodies on the same samples. Nonsense-mediated mRNA decay likely prevents most truncated proteins from being expressed, though it is possible that some escape this cellular checkpoint.

Silencing of the *PBRM1* promoter via hypermethylation has also been proposed as a possible mechanism for *PBRM1* inactivation.[145] Promoter hypermethylation of tumor-suppressor genes is commonly observed in cancer.[151] However, Ibragimova and colleagues found that hypermethylation of BAF180 is actually a rare event in ccRCC, as is the hypermethylation of *BAP1*, *SETD2*, and some other chromatin modifying genes. The authors point out this is not unusual, as some tumor-suppressor proteins commonly have promoter hypermethylation, while other do not. They propose that the deletion of the 3p21 region (which affects four tumor-suppressor genes) is more favorable to oncogenesis than the inactivation of one gene by hypermethylation.[151] Some consequences of the frequent mutations of BAF180 in ccRCC are discussed in other sections (Sections 1.4.1 and 1.5.2).

1.5.2.2 Gene Expression Signature in BAF180^{-} ccRCC Cells

The gene expression profile of BAF180-deficient cells is substantially different from that of BAF180-proficient cells. Many of the genes differentially expressed upon

BAF180 re-expression are involved in cell adhesion, cell proliferation, chromosomal instability, and hypoxia response.[69,139,149]

Most of the genes affected by loss of BAF180 were down-regulated (1910/2235 genes in Kapur et al.[139] and 485/613 genes in Wang et al.)[149] Both groups looked at gene expression in clear cell renal cell carcinoma (ccRCC) cases, where *VHL*, *PBRM1*, *BAP1*, *SETD2*, *PTEN*, and *KDM5C* are known "driver" genes (see Section 1.5.2). The significant differences in the number of genes affected may be because Wang et al. selected samples that had only *PBRM1* mutations (no *VHL*, *BAP1*, *SETD2*, *PTEN*, and *KDM5C* mutations),[149] while Kapur et al. did not report any such selection criteria.[139] Additionally, Wang et al. investigated methylation of CpG sites in the genome and found that, of the 1405 differently methylated sites, 1229 were hyper-methylated, which would result in gene silencing. They do point out, though, that not all of the genes may be directly regulated by *PBRM1*. The Wang study also investigated effects in miRNA expression; again, most of the affected miRNAs were down-regulated.[149]

Re-expression of BAF180 into Caki2 cells led to the up-regulation of genes involved in cell adhesion and actin cytoskeleton.[69] This is in agreement with Wang et al. and Kapur et al., who reported that BAF180-deficient cells had down-regulation of the same genes, as compared to control samples.[139,149] This is also in accordance with phenomenological data demonstrating that knockdown of BAF180 results in increased cell migration.[1] Re-expression of BAF180 also resulted in the up-regulation of apoptotic processes, carbohydrate metabolic processes and response to hypoxia genes. Down-regulation of genes involved in the cell cycle was observed in the same cells.[69]

BAF180-deficient cells have been noted as having a hypoxia signature[1,69] meaning the genes that are expressed when a cell is hypoxic are expressed when a cell is normoxic (i.e. normal oxygen levels) if BAF180 is absent. Gao et al. found that re-expressing BAF180 resulted in decreased expression of HIF-target genes (hypoxia inducible factor),[68] though it had no effect on HIF1α and HIF2α themselves. Chowdhury et al. found that re-expressing BAF180 increased the expression of HIF-target genes.[69] Also, Murakami et al. found that knockdown of BAF180 in hepatitis B cells and ccRCC cells resulted in down-regulation of HIF-target genes,[127] correlating with data from Chowdhury et al.[69]

1.5.2.3 Clinical Significance of BAF180 Mutations

Multiple groups have attempted to correlate the presence of BAF180 mutations to clinical relevance. However, the studies provide conflicting data with regards to correlation between BAF180 mutations and tumor size, differentiation, and patient prognosis.[128,131-132,139,145,147-148,150,152-153] It is generally accepted that patients with mutations in *BAP1* have a worse prognosis than patients with a mutation in BAF180.[131-132,139,147,153] Additionally, it has been proposed that BAF180 be used as a biomarker, but more investigations need to be conducted before that can happen.[68,139,142,145,148,154-155]

1.5.3 Epithelioid Sarcoma

Epithelioid sarcoma is a rare sarcoma that occurs in the soft tissue of the extremities (referred to as "conventional type"), most often in young adults.[156-158] It has a high metastatic rate. A second type of epithelioid sarcoma, referred to as "proximal-type", occurs in the pelvic area.[156-158] Loss of *PBRM1* expression was found 82.6% of

epithelioid sarcoma cases. It should be noted, though, that the sample size was relatively small (23 cases), so this percentage may be artificially high. There was no difference in *PBRM1* expression between the two locations of epithelioid sarcoma (conventional and proximal) nor between the three types of tumors (spindle, epithelioid, rhaboid); loss was observed in all of them. The researchers did not notice a correlation between *PBRM1* expression and cellular morphology, nor did there appear to be a correlation between *PBRM1* and prognosis. However, there was a correlation between *PBRM1* mutations and mutations in *INI1* (a core subunit of SWI/SNF), which frequently occurred together.[156]

1.5.4 Bladder Carcinoma

In bladder cancer, *PBRM1* acts as a tumor suppressor by repressing gene expression of cyclin B1, a cell cycle checkpoint protein.[99] Huang et al. found that mRNA and protein levels of *PBRM1* were significantly lower in bladder cancer cells than in normal cells. Additionally, lower levels of *PBRM1* were associated with poor patient survival rates in bladder cancer.[99,150] Cancer cell proliferation, migration, and tumor sizes all increased when expression levels of *PBRM1* were reduced. As proposed in other studies, *PBRM1* may prove to be an effective biomarker of tumor grade. Overall, *PBRM1* is mutated in 2-10% of bladder cancers.[123-125]

1.5.5. Breast Carcinoma

Several studies have analyzed potential mutations and expression levels in breast cancer. Pereira et al. found that, in ER- (estrogen receptor negative) tissues, 25% of the samples had *PBRM1* mutations and most of these resulted in loss of heterozygosity (LOH).[159] More in-depth examination of *PBRM1* mutations was performed by Xia et al.

Duplications in various exons (as revealed by PCR analysis) were common and 25 of the 52 tumor samples had LOH for BAF180. Further analysis revealed that BAF180 plays a role in the regulation of p21 expression and induces G1 arrest upon reintroduction into BAF180-deficient breast cancer cell lines.[100] Real time PCR and western blotting have demonstrated that breast cancer tumor samples and cell lines have reduced expression of the BAF180 protein, as compared to paired normal tissues.[51,160] Additionally, low BAF180 expression had a strong correlation with higher tumor stage and worse survival prognosis.[160]

1.5.6 Malignant Mesothelioma

PBRM1 has been found to be deleted in several cases of malignant mesothelioma (MM). Loss of chromosome 3p21 is common in epithelial MM, resulting in the deletion of nine genes. This region includes *PBRM1* (BAF180) as it is located at 3p21.1. Interestingly, this deletion also encompasses *BAP1*, which is sometimes deleted in ccRCC. Mutations in these chromatin modifiers are mutually exclusive in ccRCC, though that does not appear to be the case in MM.[161] *PBRM1* was found to be deleted in both types of MM (pleural and peritoneal), though the losses were more common in pleural MM.[162-163] Two point mutations in *PBRM1* were found[161] and determined to be hemizygous mutations, because there was deletion of the other copy of the gene.[164] Monoallelic gene loss was observed in 12/33 cases (36%) and biallelic deletion in 3/33 cases (9%).[164] They also confirmed that depletion of BAF180 caused increased proliferation in a malignant mesothelioma cell line.[164] cBioPortal reports deletion of *PBRM1* gene in a frequency of 5% in patients with pleural MM.[123-124]

1.5.7 Cancers of the Biliary Tract

Cancers of the biliary tract include intrahepatic (ducts within the liver) and extrahepatic (ducts outside the liver) cholangiocarcinomas and gallbladder carcinoma. Like in different types of renal carcinomas, the mutational frequency of *PBRM1* differs between biliary tract locations. *PBRM1* seems to be most frequently mutated in intrahepatic cholangiocarcinoma (ICC), ranging from 11-14% of samples.[165-167] *PBRM1*, *ARID1A*, *BAP1*, and *SMARCB1* seem to be tied for third most mutated gene in ICC.[167] *PBRM1* has been suggested to be a driver mutation for ICC and patients had a worse prognosis when *PBRM1* was mutated.[165-166] In extrahepatic cholangiocarcinomas, the frequency was 3.5-5%.[165,167] BAF180 is under-expressed in 53% of gallbladder carcinomas.[168]

1.5.8 Lung Carcinoma

In evaluating EGFR/KRAS/ALK-negative lung adenocarcinoma tissue samples, Ahn et al. found one nonsense mutation, E830* in *PBRM1*.[169] EGFR, KRAS, and ALK are driver mutations in lung cancer. In 2005, Sekine et al. analyzed non-small-cell and small-cell lung cancer cell lines and did not unearth any mutations or abnormal expression patterns. However, it should be noted that they used cancer cell lines and used only RT-PCR and western blot to draw their conclusions.[170] Izumchenko and colleagues investigated mutations found in adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA) and found *PBRM1* to be mutated in one AIS patient's samples.[171]

1.6. References

1. Varela, I.; Tarpey, P.; Raine, K., et al., Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **2011,** *469* (7331), 539-542.

2. Schaaper, R. M.; Danforth, B. N.; Glickman, B. W., Mechanisms of spontaneous mutagenesis: An analysis of the spectrum of spontaneous mutation in the Escherichia coli lacI gene. *Journal of Molecular Biology* **1986,** *189* (2), 273-284.

3. Fan, H.; Chu, J.-Y., A Brief Review of Short Tandem Repeat Mutation. *Genomics, Proteomics & Bioinformatics* **2007,** *5* (1), 7-14.

4. Gragg, H.; Harfe, B. D.; Jinks-Robertson, S., Base Composition of Mononucleotide Runs Affects DNA Polymerase Slippage and Removal of Frameshift Intermediates by Mismatch Repair in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **2002,** *22* (24), 8756-8762.

5. Goula, A.-V.; Merienne, K., Abnormal Base Excision Repair at Trinucleotide Repeats Associated with Diseases: A Tissue-Selective Mechanism. *Genes* **2013,** *4* (3), 375-387.

6. Lang, G. I.; Parsons, L.; Gammie, A. E., Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3 (Bethesda, Md.)* **2013,** *3* (9), 1453-65.

7. Yirka, B. Researchers find short tracks of DNA may aid in regulating human gene expression. https://phys.org/news/2015-07-short-tracks-dna-aid-human.html (accessed 3 April).

8. Goodwin, G. H.; Nicolas, R. H., The BAH domain, polybromo and the RSC chromatin remodelling complex. *Gene* **2001,** *268* (1-2), 1-7.

9. Chambers, A. L.; Pearl, L. H.; Oliver, A. W., et al., The BAH domain of Rsc2 is a histone H3 binding domain. *Nucleic acids research* **2013,** *41* (19), 9168-82.

10. Filippakopoulos, P.; Knapp, S., The bromodomain interaction module. *FEBS Letters* **2012,** *586* (17), 2692-2704.

11. Filippakopoulos, P.; Knapp, S., Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov* **2014,** *13* (5), 337-356.

12. Muller, S.; Filippakopoulos, P.; Knapp, S., Bromodomains as therapeutic targets. *Expert Rev Mol Med* **2011,** *13*, e29.

13. Josling, G. A.; Selvarajah, S. A.; Petter, M., et al., The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes* **2012,** *3* (2), 320-343.

14. Filippakopoulos, P.; Picaud, S.; Mangos, M., et al., Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012,** *149* (1), 214-231.

15. Brownlee, P. M.; Chambers, A. L.; Oliver, A. W., et al., Cancer and the bromodomains of BAF180. *Biochem Soc Trans* **2012,** *40* (2), 364-9.

16. Chung, C.-w.; Tough, D. F., Bromodomains: a new target class for small molecule drug discovery. *Drug Discovery Today: Therapeutic Strategies* **2012,** *9* (2–3), e111-e120.

17. Vidler, L. R.; Brown, N.; Knapp, S., et al., Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *Journal of Medicinal Chemistry* **2012,** *55* (17), 7346-7359.

18. Sutherell, C. L.; Tallant, C.; Monteiro, O. P., et al., Identification and Development of 2,3-Dihydropyrrolo[1,2-a]quinazolin-5(1H)-one Inhibitors Targeting Bromodomains within the Switch/Sucrose Nonfermenting Complex. *Journal of Medicinal Chemistry* **2016,** *59* (10), 5095-5101.

19. Myrianthopoulos, V.; Gaboriaud-Kolar, N.; Tallant, C., et al., Discovery and Optimization of a Selective Ligand for the Switch/Sucrose Nonfermenting-Related Bromodomains of Polybromo Protein-1 by the Use of Virtual Screening and Hydration Analysis. *Journal of Medicinal Chemistry* **2016,** *59* (19), 8787-8803.

20. NIH Bromodomain Inhibitor Clinical Trials. https://clinicaltrials.gov/ct2/home (accessed 16 January 2017).

21. Zeng, L.; Zhou, M.-M., Bromodomain: an acetyl-lysine binding domain. *FEBS Letters* **2002,** *513* (1), 124-128.

22. Mujtaba, S.; Zeng, L.; Zhou, M. M., Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **2007,** *26*, 5521.

23. Porter, E. G.; Dykhuizen, E. C., Individual Bromodomains of Polybromo-1 Contribute to Chromatin Association and Tumor Suppression in Clear Cell Renal Carcinoma. *Journal of Biological Chemistry* **2017**.

24. Moriniere, J.; Rousseaux, S.; Steuerwald, U., et al., Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **2009,** *461* (7264), 664-8.

25. Xu, D.; Zhang, Y., Generating Triangulated Macromolecular Surfaces by Euclidean Distance Transform. *PLOS ONE* **2009,** *4* (12), e8140.

26. Moreland, J. L.; Gramada, A.; Buzko, O. V., et al., The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **2005,** *6* (1), 21.

27. Charlop-Powers, Z.; Zeng, L.; Zhang, Q., et al., Structural insights into selective histone H3 recognition by the human Polybromo bromodomain 2. *Cell Res* **2010,** *20* (5), 529-538.

28. Chandrasekaran, R.; Thompson, M., Polybromo-1-bromodomains bind histone H3 at specific acetyl-lysine positions. *Biochemical and biophysical research communications* **2007,** *355* (3), 661-6.

29. Chandrasekaran, R.; Thompson, M., Expression, purification and characterization of individual bromodomains from human Polybromo-1. *Protein expression and purification* **2006,** *50* (1), 111-7.

30. Kupitz, C.; Chandrasekaran, R.; Thompson, M., Kinetic analysis of acetylation-dependent Pb1 bromodomain–histone interactions. *Biophysical Chemistry* **2008,** *136* (1), 7-12.

31. Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R., et al., CDD: NCBI's conserved domain database. *Nucleic acids research* **2015,** *43* (Database issue), D222-6.

32. Sigrist, C. J.; de Castro, E.; Cerutti, L., et al., New and continuing developments at PROSITE. *Nucleic acids research* **2013,** *41* (Database issue), D344-7.

33. Finn, R. D.; Attwood, T. K.; Babbitt, P. C., et al., InterPro in 2017—beyond protein family and domain annotations. *Nucleic acids research* **2016,** *45* (D1), D190-D199.

34. Letunic, I.; Doerks, T.; Bork, P., SMART: recent updates, new developments and status in 2015. *Nucleic acids research* **2015,** *43* (Database issue), D257-D260.

35. MOTIF: Searching Protein Sequence Motifs - GenomeNet. http://www.genome.jp/tools/motif/ (accessed 26 December).

36. Finn, R. D.; Coggill, P.; Eberhardt, R. Y., et al., The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* **2016,** *44* (D1), D279-D285.

37. Berman, H. M.; Westbrook, J.; Feng, Z., et al., The Protein Data Bank. *Nucleic acids research* **2000,** *28* (1), 235-242.

38. Yang, N.; Xu, R. M., Structure and function of the BAH domain in chromatin biology. *Crit Rev Biochem Mol Biol* **2013,** *48* (3), 211-21.

39. The UniProt Consortium, Activities at the Universal Protein Resource (UniProt). *Nucleic acids research* **2014,** *42* (D1), D191-D198.

40. Xue, Y.; Canman, J. C.; Lee, C. S., et al., The human SWI/SNF-B chromatin-remodeling complex is related to yeast rsc and localizes at kinetochores of mitotic chromosomes. *Proc Natl Acad Sci U S A* **2000,** *97* (24), 13015-20.

41. Cleveland, D. W.; Mao, Y.; Sullivan, K. F., Centromeres and Kinetochores. *Cell 112* (4), 407-421.

42. Oliver, A. W.; Jones, S. A.; Roe, S. M., et al., Crystal structure of the proximal BAH domain of the polybromo protein. *Biochem J* **2005,** *389* (Pt 3), 657-64.

43. Niimi, A.; Hopkins, S. R.; Downs, J. A., et al., The BAH domain of BAF180 is required for PCNA ubiquitination. *Mutat Res* **2015,** *779*, 16-23.

44. Štros, M.; Launholt, D.; Grasser, K. D., The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cellular and Molecular Life Sciences* **2007,** *64* (19), 2590-2606.

45. Reeves, R., HMG Nuclear Proteins: Linking Chromatin Structure to Cellular Phenotype. *Biochimica et biophysica acta* **2010,** *1799* (1-2), 3.

46. Goodwin, G. H.; Sanders, C.; Johns, E. W., A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. *Eur J Biochem* **1973,** *38* (1), 14-9.

47. Štros, M., HMGB proteins: Interactions with DNA and chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2010,** *1799* (1–2), 101-113.

48. Malarkey, C. S.; Churchill, M. E. A., The high mobility group box: the ultimate utility player of a cell. *Trends in biochemical sciences* **2012,** *37* (12), 553-562.

49. Reeves, R.; Adair, J. E., Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA Repair* **2005,** *4* (8), 926-938.

50. Ueda, T.; Yoshida, M., HMGB proteins and transcriptional regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2010,** *1799* (1–2), 114-118.

51. Decristofaro, M. F.; Betz, B. L.; Rorie, C. J., et al., Characterization of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. *J Cell Physiol* **2001,** *186* (1), 136-45.

52. Downs, J. A., Chromatin structure and DNA double-strand break responses in cancer progression and therapy. *Oncogene* **2007,** *26* (56), 7765-72.

53. UniProt: a hub for protein information. *Nucleic Acids Research* **2014,** *43* (D1), D204-D212.

54. Zhang, Q.; Wang, Y., HMG Modifications and Nuclear Function. *Biochimica et biophysica acta* **2010,** *1799* (1-2), 28.

55. Ohndorf, U. M.; Rould, M. A.; He, Q., et al., Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **1999,** *399* (6737), 708-12.

56. Biegel, J. A.; Busse, T. M.; Weissman, B. E., SWI/SNF chromatin remodeling complexes and cancer. *Am J Med Genet C Semin Med Genet* **2014**.

57. Helming, Katherine C.; Wang, X.; Roberts, Charles W. M., Vulnerabilities of Mutant SWI/SNF Complexes in Cancer. *Cancer Cell* **2014,** *26* (3), 309-317.

58. Kadoch, C.; Hargreaves, D. C.; Hodges, C., et al., Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature genetics* **2013,** *45* (6), 592-601.

59. Masliah-Planchon, J.; Bieche, I.; Guinebretiere, J. M., et al., SWI/SNF chromatin remodeling and human malignancies. *Annu Rev Pathol* **2015,** *10*, 145-71.

60. Wang, X.; Haswell, J. R.; Roberts, C. W., Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer--mechanisms and potential therapeutic insights. *Clin Cancer Res* **2014,** *20* (1), 21-7.

61. Weissman, B.; Knudsen, K. E., Hijacking the chromatin remodeling machinery: impact of SWI/SNF perturbations in cancer. *Cancer Res* **2009,** *69* (21), 8223-30.

62. Wilson, B. G.; Roberts, C. W., SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **2011,** *11* (7), 481-92.

63. Brownlee, P. M.; Meisenberg, C.; Downs, J. A., The SWI/SNF chromatin remodelling complex: Its role in maintaining genome stability and preventing tumourigenesis. *DNA Repair* **2015,** (0).

64. Marquez, S.; Thompson, K. W.; Lu, L., et al., Beyond Mutations: Additional Mechanisms and Implications of SWI/SNF Complex Inactivation. *Frontiers in Oncology* **2015,** *4*.

65. Lu, P.; Roberts, C. W. M., The SWI/SNF tumor suppressor complex: Regulation of promoter nucleosomes and beyond. *Nucleus* **2013,** *4* (5), 374-378.

66. Kakarougkas, A.; Downs, J. A.; Jeggo, P. A., The PBAF chromatin remodeling complex represses transcription and promotes rapid repair at DNA double-strand breaks. *Molecular & Cellular Oncology* **2015,** *2* (1), e970072.

67. Finn, R. D.; Bateman, A.; Clements, J., et al., Pfam: the protein families database. *Nucleic acids research* **2014,** *42* (D1), D222-D230.

68. Gao, W.; Li, W.; Xiao, T., et al., Inactivation of the PBRM1 tumor suppressor gene amplifies the HIF-response in VHL−/− clear cell renal carcinoma. *Proceedings of the National Academy of Sciences* **2017**.

69. Chowdhury, B.; Porter, E. G.; Stewart, J. C., et al., PBRM1 Regulates the Expression of Genes Involved in Metabolism and Cell Adhesion in Renal Clear Cell Carcinoma. *PLOS ONE* **2016,** *11* (4), e0153718.

70. Yan, Z.; Cui, K.; Murray, D. M., et al., PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes & development* **2005,** *19* (14), 1662-7.

71. Yan, Z.; Cui, K.; Murray, D. M., et al., PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes & development* **2005,** *19* (14), 1662-1667.

72. Bastien, J.; Rochette-Egly, C., Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* **2004,** *328*, 1-16.

73. Kishimoto, M.; Fujiki, R.; Takezawa, S., et al., Nuclear Receptor Mediated Gene Regulation through Chromatin Remodeling and Histone Modifications. *Endocrine Journal* **2006,** *53* (2), 157-172.

74. Das, B. C.; Thapa, P.; Karki, R., et al., Retinoic acid signaling pathways in development and diseases. *Bioorganic & Medicinal Chemistry* **2014,** *22* (2), 673-683.

75. Sigma-Aldrich Retinoic Acid and Gene Expression. http://www.sigmaaldrich.com/technical-documents/articles/biofiles/retinoic-acid-and.html (accessed 11 November 2016).

76. Kam, R. K. T.; Deng, Y.; Chen, Y., et al., Retinoic acid synthesis and functions in early embryonic development. *Cell & Bioscience* **2012,** *2*, 11-11.

77. Niederreither, K.; Vermot, J.; Messaddeq, N., et al., Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. *Development* **2001,** *128* (7), 1019-1031.

78. Wang, Z.; Zhai, W.; Richardson, J. A., et al., Polybromo protein BAF180 functions in mammalian cardiac chamber maturation. *Genes & development* **2004,** *18* (24), 3106-3116.

79. Lemon, B.; Inouye, C.; King, D. S., et al., Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. *Nature* **2001,** *414* (6866), 924-928.

80. Huang, X.; Gao, X.; Diaz-Trelles, R., et al., Coronary development is regulated by ATP-dependent SWI/SNF chromatin remodeling component BAF180. *Dev Biol* **2008,** *319* (2), 258-66.

81. Singh, A. P.; Archer, T. K., Analysis of the SWI/SNF chromatin-remodeling complex during early heart development and BAF250a repression cardiac gene transcription during P19 cell differentiation. *Nucleic acids research* **2014,** *42* (5), 2958-75.

82. Sargis, R. M. An Overview of the Thymus. http://www.endocrineweb.com/endocrinology/overview-thymus (accessed 28 November 2016).

83. Cardiff University. T Cell Activation. http://www.tcells.org/scientific/tcell_activation/ (accessed 6 December 2016).

84. Cardiff University. Helper T Cells. http://www.tcells.org/scientific/helper/ (accessed 6 December 2016).

85. Wurster, A. L.; Precht, P.; Becker, K. G., et al., IL-10 transcription is negatively regulated by BAF180, a component of the SWI/SNF chromatin remodeling enzyme. *BMC Immunol* **2012,** *13*, 9.

86. GeneCards. CD44. http://www.genecards.org/cgi-bin/carddisp.pl?gene=CD44 (accessed 6 December 2016).

87. Couper, K. N.; Blount, D. G.; Riley, E. M., IL-10: The Master Regulator of Immunity to Infection. *The Journal of Immunology* **2008,** *180* (9), 5771-5777.

88. Brownlee, Peter M.; Chambers, Anna L.; Cloney, R., et al., BAF180 Promotes Cohesion and Prevents Genome Instability and Aneuploidy. *Cell Reports* **2014,** *6* (6), 973-981.

89. Griffiths, A.; Miller, J.; Suzuki, D., Glossary. In *Introduction to Genetic Analysis*, 7th ed.; W. H. Freeman: New York, 2000.

90. Fenech, M.; Kirsch-Volders, M.; Natarajan, A. T., et al., Molecular mechanisms of micronucleus, nucleoplasmic bridge and nuclear bud formation in mammalian and human cells. *Mutagenesis* **2011,** *26* (1), 125-132.

91. Bakhoum, S. F.; Silkworth, W. T.; Nardi, I. K., et al., The mitotic origin of chromosomal instability. *Current biology : CB* **2014,** *24* (4), R148-R149.

92. Cimini, D.; Howell, B.; Maddox, P., et al., Merotelic Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic Mammalian Tissue Cells. *The Journal of Cell Biology* **2001,** *153* (3), 517-528.

93. Griffiths, A.; Miller, J.; Suzuki, D., Aneuploidy. In *Introduction to Genetic Analysis*, 7th ed.; W. H. Freeman: New York, 2000.

94. Kakarougkas, A.; Ismail, A.; Chambers, Anna L., et al., Requirement for PBAF in Transcriptional Repression and Repair at DNA Breaks in Actively Transcribed Regions of Chromatin. *Molecular Cell* **2014,** *55* (5), 723-732.

95. Kuo, L. J.; Yang, L.-X., γ-H2AX - A Novel Biomarker for DNA Double-strand Breaks. *In Vivo* **2008,** *22* (3), 305-309.

96. Matsuoka, S.; Ballif, B. A.; Smogorzewska, A., et al., ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. *Science* **2007,** *316* (5828), 1160-1166.

97. Kannouche, P. L.; Wing, J.; Lehmann, A. R., Interaction of Human DNA Polymerase η with Monoubiquitinated PCNA: A Possible Mechanism for the Polymerase Switch in Response to DNA Damage. *Molecular Cell* **2004,** *14* (4), 491-500.

98. Niimi, A.; Chambers, A. L.; Downs, J. A., et al., A role for chromatin remodellers in replication of damaged DNA. *Nucleic acids research* **2012,** *40* (15), 7393-7403.

99. Huang, L.; Peng, Y.; Zhong, G., et al., PBRM1 suppresses bladder cancer by cyclin B1 induced cell cycle arrest. *Oncotarget* **2015**.

100. Xia, W.; Nagase, S.; Montia, A. G., et al., BAF180 Is a Critical Regulator of p21 Induction and a Tumor Suppressor Mutated in Breast Cancer. *Cancer Research* **2008,** *68* (6), 1667-1674.

101. Burrows, A. E.; Smogorzewska, A.; Elledge, S. J., Polybromo-associated BRG1-associated factor components BRD7 and BAF180 are critical regulators of p53 required

for induction of replicative senescence. *Proceedings of the National Academy of Sciences* **2010,** *107* (32), 14280-14285.

102. Macher-Goeppinger, S.; Keith, M.; Tagscherer, K. E., et al., The PBRM1 (BAF180) protein is functionally regulated by p53-induced protein degradation in renal cell carcinomas. *The Journal of Pathology* **2015**, n/a-n/a.

103. Nutlin-3a http://www.selleckchem.com/products/nutlin-3a.html (accessed 23 January 2017).

104. el-Deiry, W. S.; Tokino, T.; Velculescu, V. E., et al., WAF1, a potential mediator of p53 tumor suppression. *Cell* **1993,** *75* (4), 817-25.

105. Tanaka, T.; Suh, K. S.; Lo, A. M., et al., p21WAF1/CIP1 Is a Common Transcriptional Target of Retinoid Receptors: Pleiotropic Regulatory Mechanism Through Retinoic Acid Receptor (RAR)/Retinoid X Receptor (RXR) Heterodimer and RXR/RXR Homodimer. *Journal of Biological Chemistry* **2007,** *282* (41), 29987-29997.

106. Blagosklonny, M. V., Cell cycle arrest is not senescence. *Aging (Albany NY)* **2011,** *3* (2), 94-101.

107. Schizophrenia Working Group of the Psychiatric Genomics, C.; Ripke, S.; Neale, B. M., et al., Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature* **2014,** *511* (7510), 421-427.

108. Kondo, K.; Ikeda, M.; Kajio, Y., et al., Genetic Variants on 3q21 and in the *Sp8* Transcription Factor Gene as Susceptibility Loci for Psychotic Disorders: A Genetic Association Study. *PLoS ONE* **2013,** *8* (8), e70964.

109. Lee, K. W.; Woon, P. S.; Teo, Y. Y., et al., Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: What have we learnt? *Neuroscience & Biobehavioral Reviews* **2012,** *36* (1), 556-571.

110. Goes, F. S.; Hamshere, M. L.; Seifuddin, F., et al., Genome-wide association of mood-incongruent psychotic bipolar disorder. *Translational Psychiatry* **2012,** *2* (10), e180.

111. Ripke, S.; O'Dushlaine, C.; Chambert, K., et al., Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature genetics* **2013,** *45* (10), 1150-9.

112. Sklar, P.; Ripke, S.; Scott, L. J., et al., Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics* **2011,** *43* (10), 977-983.

113. McMahon, F. J.; Akula, N.; Schulze, T. G., et al., Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nature genetics* **2010,** *42* (2), 128-31.

114. Williams, H. J.; Craddock, N.; Russo, G., et al., Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. *Hum Mol Genet* **2011,** *20* (2), 387-91.

115. Green, E. K.; Hamshere, M.; Forty, L., et al., Replication of bipolar disorder susceptibility alleles and identification of 2 novel genome-wide significant associations in a new bipolar disorder case-control sample. *Molecular psychiatry* **2013,** *18* (12), 1302-1307.

116. Vassos, E.; Steinberg, S.; Cichon, S., et al., Replication study and meta-analysis in European samples supports association of the 3p21.1 locus with bipolar disorder. *Biol Psychiatry* **2012,** *72* (8), 645-50.

117. Gurung, R.; Prata, D. P., What is the impact of genome-wide supported risk variants for schizophrenia and bipolar disorder on brain structure and function? A systematic review. *Psychological Medicine* **2015,** *45* (12), 2461-2480.

118. Gautier, V. W.; Gu, L.; O'Donoghue, N., et al., In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology* **2009,** *6*, 47-47.

119. Nolan, G. Long Terminal Repeats: The Retroviral Promoter. https://web.stanford.edu/group/nolan/_OldWebsite/tutorials/retcl_3_ltrs.html (accessed 7 December 2016).

120. Mahmoudi, T., The BAF complex and HIV latency. *Transcription* **2012,** *3* (4), 171-176.

121. Boehm, D.; Conrad, R. J.; Ott, M., Bromodomain proteins in HIV infection. *Viruses* **2013,** *5* (6), 1571-86.

122. Rafati, H.; Parra, M.; Hakre, S., et al., Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency. *PLoS Biol* **2011,** *9* (11), e1001206.

123. Cerami, E.; Gao, J.; Dogrusoz, U., et al., The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* **2012,** *2* (5), 401-404.

124. Gao, J.; Aksoy, B. A.; Dogrusoz, U., et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **2013,** *6* (269), pl1.

125. Forbes, S. A.; Beare, D.; Gunasekaran, P., et al., COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research* **2015,** *43* (D1), D805-D811.

126. Xiao, X.; Tang, C.; Xiao, S., et al., Enhancement of proliferation and invasion by MicroRNA-590-5p via targeting PBRM1 in clear cell renal carcinoma cells. *Oncol Res* **2013,** *20* (11), 537-44.

127. Murakami, A.; Wang, L.; Kalhorn, S., et al., Context-dependent role for chromatin remodeling component PBRM1/BAF180 in clear cell renal cell carcinoma. *Oncogenesis* **2017,** *6*, e287.

128. Pawlowski, R.; Muhl, S. M.; Sulser, T., et al., Loss of PBRM1 expression is associated with renal cell carcinoma progression. *International journal of cancer* **2013,** *132* (2), E11-7.

129. The Cancer Genome Atlas Clear Cell Kidney Carcinoma. https://cancergenome.nih.gov/cancersselected/kidneyclearcell (accessed 1 February 2017).

130. Ho, T. H.; Kapur, P.; Joseph, R. W., et al., Loss of PBRM1 and BAP1 Expression Is Less Common in Non–Clear Cell Renal Cell Carcinoma Than in Clear Cell Renal Cell Carcinoma. *Urologic oncology* **2015,** *33* (1), 23.e9-23.e14.

131. Pena-Llopis, S.; Christie, A.; Xie, X. J., et al., Cooperation and antagonism among cancer genes: the renal cancer paradigm. *Cancer Res* **2013,** *73* (14), 4173-9.

132. Gossage, L.; Murtaza, M.; Slatter, A. F., et al., Clinical and pathological impact of VHL, PBRM1, BAP1, SETD2, KDM6A, and JARID1c in clear cell renal cell carcinoma. *Genes Chromosomes Cancer* **2014,** *53* (1), 38-51.

133. Stratton, M. R.; Campbell, P. J.; Futreal, P. A., The cancer genome. *Nature* **2009,** *458* (7239), 719-724.

134. Roy, D. M.; Walsh, L. A.; Chan, T. A., Driver mutations of cancer epigenomes. *Protein Cell* **2014,** *5* (4), 265-96.

135. Merid, S.; Goranskaya, D.; Alexeyenko, A., Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. *BMC Bioinformatics* **2014,** *15* (1), 308.

136. Vandin, F.; Upfal, E.; Raphael, B. J., Finding driver pathways in cancer: models and algorithms. *Algorithms for Molecular Biology* **2012,** *7* (1), 1-10.

137. Leiserson, M. D. M.; Blokh, D.; Sharan, R., et al., Simultaneous Identification of Multiple Driver Pathways in Cancer. *PLoS Comput Biol* **2013,** *9* (5), e1003054.

138. Vandin, F.; Upfal, E.; Raphael, B. J., De novo discovery of mutated driver pathways in cancer. *Genome Research* **2012,** *22* (2), 375-385.

139. Kapur, P.; Pena-Llopis, S.; Christie, A., et al., Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol* **2013,** *14* (2), 159-67.

140. Brugarolas, J., PBRM1 and BAP1 as novel targets for renal cell carcinoma. *Cancer J* **2013,** *19* (4), 324-32.

141. The Cancer Genome Atlas Research, N., ,, Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **2013,** *499* (7456), 43-49.

142. Czarnecka, A. M.; Kornakiewicz, A.; Kukwa, W., et al., Frontiers in clinical and molecular diagnostics and staging of metastatic clear cell renal cell carcinoma. *Future Oncol* **2014,** *10* (6), 1095-111.

143. Duns, G.; Hofstra, R. M.; Sietzema, J. G., et al., Targeted exome sequencing in clear cell renal cell carcinoma tumors suggests aberrant chromatin regulation as a crucial step in ccRCC development. *Hum Mutat* **2012,** *33* (7), 1059-62.

144. Peña-Llopis, S.; Vega-Rubín-de-Celis, S.; Liao, A., et al., BAP1 loss defines a new class of renal cell carcinoma. *Nature genetics* **2012,** *44* (7), 751-759.

145. Piva, F.; Santoni, M.; Matrana, M. R., et al., BAP1, PBRM1 and SETD2 in clear-cell renal cell carcinoma: molecular diagnostics and possible targets for personalized therapies. *Expert review of molecular diagnostics* **2015,** *15* (9), 1201-10.

146. Joseph, R. W.; Kapur, P.; Serie, D. J., et al., Clear Cell Renal Cell Carcinoma Subtypes Identified by BAP1 and PBRM1 Expression. *The Journal of urology* **2016,** *195* (1), 180-7.

147. Hakimi, A. A.; Chen, Y. B.; Wren, J., et al., Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *Eur Urol* **2013,** *63* (5), 848-54.

148. Kluzek, K.; Bluyssen, H. A.; Wesoly, J., The epigenetic landscape of clear-cell renal cell carcinoma. *2015* **2015,** *2* (3), 15.

149. Wang, Y.; Guo, X.; Bray, M. J., et al., An integrative genomics approach for identifying novel functional consequences of PBRM1 truncated mutations in clear cell renal cell carcinoma (ccRCC). *BMC genomics* **2016,** *17 Suppl 7*, 515.

150. da Costa, W. H.; Rezende, M.; Carneiro, F. C., et al., Polybromo-1 (PBRM1), a SWI/SNF complex subunit is a prognostic marker in clear cell renal cell carcinoma. *BJU Int* **2014,** *113* (5b), E157-63.

151. Ibragimova, I.; Maradeo, M. E.; Dulaimi, E., et al., Aberrant promoter hypermethylation of PBRM1, BAP1, SETD2, KDM6A and other chromatin-modifying genes is absent or rare in clear cell RCC. *Epigenetics* **2013,** *8* (5), 486-93.

152. Nam, S. J.; Lee, C.; Park, J. H., et al., Decreased PBRM1 expression predicts unfavorable prognosis in patients with clear cell renal cell carcinoma. *Urol Oncol* **2015**.

153. Joseph, R. W.; Kapur, P.; Serie, D. J., et al., Clear Cell Renal Cell Carcinoma Subtypes Identified by BAP1 and PBRM1 Expression. *The Journal of urology* **2015**.

154. Fay, A. P.; de Velasco, G.; Ho, T. H., et al., Whole-Exome Sequencing in Two Extreme Phenotypes of Response to VEGF-Targeted Therapies in Patients With Metastatic Clear Cell Renal Cell Carcinoma. *Journal of the National Comprehensive Cancer Network : JNCCN* **2016,** *14* (7), 820-4.

155. Gulati, S.; Martinez, P.; Joshi, T., et al., Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. *Eur Urol* **2014,** *66* (5), 936-48.

156. Li, L.; Fan, X. S.; Xia, Q. Y., et al., Concurrent loss of INI1, PBRM1, and BRM expression in epithelioid sarcoma: implications for the cocontributions of multiple SWI/SNF complex members to pathogenesis. *Hum Pathol* **2014,** *45* (11), 2247-54.

157. Chbani, L.; Guillou, L.; Terrier, P., et al., Epithelioid sarcoma: a clinicopathologic and immunohistochemical analysis of 106 cases from the French sarcoma group. *American journal of clinical pathology* **2009,** *131* (2), 222-7.

158. Hosseinzadeh, P.; Cheung, F. Epithelioid Sarcoma. http://sarcomahelp.org/epithelioid-sarcoma.html (accessed 25 January).

159. Pereira, B.; Chin, S.-F.; Rueda, O. M., et al., The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications* **2016,** *7*, 11479.

160. Mo, D.; Li, C.; Liang, J., et al., Low PBRM1 identifies tumor progression and poor prognosis in breast cancer. *Int J Clin Exp Pathol* **2015,** *8* (8), 9307-13.

161. Yoshikawa, Y.; Sato, A.; Tsujimura, T., et al., Biallelic germline and somatic mutations in malignant mesothelioma: Multiple mutations in transcription regulators including mSWI/SNF genes. *International journal of cancer* **2015,** *136* (3), 560-571.

162. Borczuk, A. C.; Pei, J.; Taub, R. N., et al., Genome-wide analysis of abdominal and pleural malignant mesothelioma with DNA arrays reveals both common and distinct regions of copy number alteration. *Cancer Biology & Therapy* **2016,** *17* (3), 328-335.

163. Yoshikawa, Y.; Sato, A.; Tsujimura, T., et al., Frequent deletion of 3p21.1 region carrying semaphorin 3G and aberrant expression of the genes participating in semaphorin

signaling in the epithelioid type of malignant mesothelioma cells. *Int J Oncol* **2011,** *39* (6), 1365-74.

164. Yoshikawa, Y.; Emi, M.; Hashimoto-Tamaoki, T., et al., High-density array-CGH with targeted NGS unmask multiple noncontiguous minute deletions on chromosome 3p21 in mesothelioma. *Proceedings of the National Academy of Sciences* **2016,** *113* (47), 13432-13437.

165. Churi, C. R.; Shroff, R.; Wang, Y., et al., Mutation Profiling in Cholangiocarcinoma: Prognostic and Therapeutic Implications. *PLoS ONE* **2014,** *9* (12), e115383.

166. Jiao, Y.; Pawlik, T. M.; Anders, R. A., et al., Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nature genetics* **2013,** *45* (12), 1470-3.

167. Simbolo, M.; Fassan, M.; Ruzzenente, A., et al., Multigene mutational profiling of cholangiocarcinomas identifies actionable molecular subgroups. *Oncotarget* **2014,** *5* (9), 2839-2852.

168. Sicklick, J. K.; Fanta, P. T.; Shimabukuro, K., et al., Genomics of gallbladder cancer: the case for biomarker-driven clinical trial design. *Cancer metastasis reviews* **2016,** *35* (2), 263-75.

169. Ahn, J. W.; Kim, H. S.; Yoon, J.-K., et al., Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers. *Genome Medicine* **2014,** *6* (2), 18-18.

170. Sekine, I.; Sato, M.; Sunaga, N., et al., The 3p21 candidate tumor suppressor gene BAF180 is normally expressed in human lung cancer. **2005,** *24* (16), 2735-2738.

171. Izumchenko, E.; Chang, X.; Brait, M., et al., Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nature Communications* **2015,** *6*, 8258.

Chapter 2: Motive and Challenges

2.1 Motive

Ever since Varela et al. reported that *PBRM1* is mutated in 41% of clear cell renal cell carcinoma (ccRCC) cases,[1] the human polybromo-1 protein (gene: *PBRM1*; protein: BAF180) has garnered the attention of medical researchers. *PBRM1* is the second most mutated gene in ccRCC (the first being *VHL*), as well as a known driver mutation of that cancer.[2-3] It has been found to be mutated (at lower frequencies) in more than 35 other cancers, such as cholangiocarcinomas,[4] metastasizing pleomorphic adenomas,[5] esophageal squamous cell carcinoma,[6] and breast cancer.[7-8] As of November 2016, COSMIC (Catalogue of somatic mutations in cancer) has more than 1,100 unique *PBRM1* mutations in its database, many of which are nonsense mutations.[9] Nonsense mutations often result in the elimination one or more domains; in extreme cases, the mutated *PBRM1* may only code for 2 or 3 domains.[1]

The tumor suppressor properties of BAF180 likely stems from its many roles in DNA repair. BAF180 is critical for proper centromeric cohesion[10] (important during recombinatorial repair) and silencing transcription at double-strand breaks.[11] The BAH domains of BAF180 have been found to be critical for proper ubiquitination of PCNA during post-replication repair (PRR).[12] The HMG box, though not well-studied in this protein, likely promotes non-homologous end-joining (NHEJ) by its interactions with DNA.[13-15] Therefore, mutant BAF180 cancer cells could be compromised with regards to DNA repair in four ways: the chromosomes are further apart and have less cohesion,

which leads to genomic instability and prevents recombinatorial DNA repair; BAF180-deficient cells cannot stop transcription at double-stranded breaks, thereby hindering DSB repair; mutant BAF180 without BAH domains cannot ubiquitinate PCNA, which obstructs most of the PRR pathway; and deletion of the HMG box likely hinders NHEJ and transcription because BAF180 can no longer bind to DNA.

Phenomenological studies of BAF180 in mammalian cells have been conducted and provide valuable information on "cause and effect" of BAF180 in a cell. On the other end of the spectrum, several groups have expressed the bromodomains individually and studied their interactions with acetylated peptides.[16-21] But these studies do not provide insight into the exact mechanisms/interactions of BAF180.

A critical gap in knowledge still exists between the two: how do the domains of BAF180 work together to bind to the nucleosome and what changes result from the binding of different domains? It has been shown that domains within the same protein have the ability to work together to recognize multiple post-translational modification (PTM) sites, thus increasing the binding affinity of the protein to the nucleosome.[17,22-25] Therefore, it is logical to hypothesize that the domains of BAF180 might do the same thing. The ability to study distinct interactions is crucial to increasing our understanding of the role of BAF180 in cancer development and its progression. Only by studying the mechanisms of full-length BAF180 can we begin to develop better treatment options for cancer patients.

2.2 Challenges

BAF180, the chromatin-remodeling subunit of the PBAF complex, is a complicated protein with 1634 amino acids (isoform 2) and 9 domains: 6 bromodomains, 2 BAH (bromo-adjacent homology) domains, and 1 HMG (high mobility group) box (Figure 2.1).

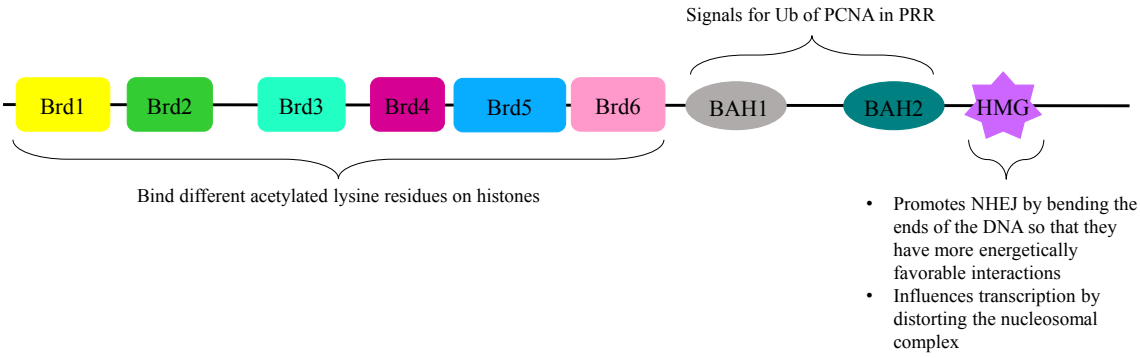


Figure 2.1: Diagram of the 9 domains of BAF180 with likely functions. Sizes of the domains and distances between are approximately accurate for isoform 2 (Genbank: AAG48933.1). The overall size is 1634 amino acids.

Recombinant expression of full-length BAF180 has proven to be extremely difficult and frustrating. There could be many reasons why this is so. First, it must be recognized that BAF180 is a very large protein with 9 domains. Large, multidomain proteins are inherently difficult to express in any heterologous system and have a higher likelihood of folding improperly. Proteins are usually expressed by cloning and expressing the individual domains separately. Secondly, it has numerous rare codons, even compared to the overall usage of *Homo sapiens*. Third, its DNA can form multiple complex secondary structures, making it difficult for PCR modification to occur. Fourth,

its DNA sequence has many base runs (A-tracts, etc.), thus increasing the probability that polymerase might skip bases during PCR.

2.3 Conclusion

This is the first time full-length BAF180 has been successfully expressed and purified in a heterologous host. The creation of recombinant BAF180 will enable researchers to conduct detailed mechanistic studies of the precise binding interactions of the entire protein. The ability to study distinct interactions is crucial to increasing the understanding of the role of BAF180 in cancer development and its progression. In studying the exact mechanisms by which BAF180 performs its many roles, scientists can begin to develop better treatment options for cancer patients.

2.4 References

1. Varela, I.; Tarpey, P.; et al., Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **2011,** *469* (7331), 539-542.

2. Pawlowski, R.; Muhl, S. M.; et al., Loss of PBRM1 expression is associated with renal cell carcinoma progression. *International journal of cancer* **2013,** *132* (2), E11-7.

3. Pena-Llopis, S.; Christie, A.; et al., Cooperation and antagonism among cancer genes: the renal cancer paradigm. *Cancer Res* **2013,** *73* (14), 4173-9.

4. Jiao, Y.; Pawlik, T. M.; et al., Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nature genetics* **2013,** *45* (12), 1470-3.

5. Mariano, F. V.; Gondak Rde, O.; et al., Genomic copy number alterations of primary and secondary metastasizing pleomorphic adenomas. *Histopathology* **2015,** *67* (3), 410-5.

6. Nakazato, H.; Takeshima, H.; et al., Early-Stage Induction of SWI/SNF Mutations during Esophageal Squamous Cell Carcinogenesis. *PLoS One* **2016,** *11* (1).

7. Xia, W.; Nagase, S.; et al., BAF180 Is a Critical Regulator of p21 Induction and a Tumor Suppressor Mutated in Breast Cancer. *Cancer Research* **2008,** *68* (6), 1667-1674.

8. Decristofaro, M. F.; Betz, B. L.; et al., Characterization of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. *J Cell Physiol* **2001,** *186* (1), 136-45.

9. Forbes, S. A.; Beare, D.; et al., COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research* **2015,** *43* (D1), D805-D811.

10. Brownlee, Peter M.; Chambers, Anna L.; et al., BAF180 Promotes Cohesion and Prevents Genome Instability and Aneuploidy. *Cell Reports* **2014,** *6* (6), 973-981.

11. Kakarougkas, A.; Ismail, A.; et al., Requirement for PBAF in Transcriptional Repression and Repair at DNA Breaks in Actively Transcribed Regions of Chromatin. *Molecular Cell* **2014,** *55* (5), 723-732.

12. Niimi, A.; Hopkins, S. R.; et al., The BAH domain of BAF180 is required for PCNA ubiquitination. *Mutat Res* **2015,** *779*, 16-23.

13. Downs, J. A., Chromatin structure and DNA double-strand break responses in cancer progression and therapy. *Oncogene* **2007,** *26* (56), 7765-72.

14. Štros, M.; Launholt, D.; et al., The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cellular and Molecular Life Sciences* **2007,** *64* (19), 2590-2606.

15. Reeves, R., Nuclear functions of the HMG proteins. *Biochimica et biophysica acta* **2010,** *1799* (1-2), 3-14.

16. Muller, S.; Filippakopoulos, P.; et al., Bromodomains as therapeutic targets. *Expert Rev Mol Med* **2011,** *13*, e29.

17. Filippakopoulos, P.; Picaud, S.; et al., Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012,** *149* (1), 214-231.

18. Charlop-Powers, Z.; Zeng, L.; et al., Structural insights into selective histone H3 recognition by the human Polybromo bromodomain 2. *Cell Res* **2010,** *20* (5), 529-538.

19. Chandrasekaran, R.; Thompson, M., Polybromo-1-bromodomains bind histone H3 at specific acetyl-lysine positions. *Biochemical and biophysical research communications* **2007,** *355* (3), 661-6.

20. Chandrasekaran, R.; Thompson, M., Expression, purification and characterization of individual bromodomains from human Polybromo-1. *Protein expression and purification* **2006,** *50* (1), 111-7.

21. Kupitz, C.; Chandrasekaran, R.; et al., Kinetic analysis of acetylation-dependent Pb1 bromodomain–histone interactions. *Biophysical Chemistry* **2008,** *136* (1), 7-12.

22. Filippakopoulos, P.; Knapp, S., The bromodomain interaction module. *FEBS Letters* **2012,** *586* (17), 2692-2704.

23. Josling, G. A.; Selvarajah, S. A.; et al., The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes* **2012,** *3* (2), 320-343.

24. Porter, E. G.; Dykhuizen, E. C., Individual Bromodomains of Polybromo-1 Contribute to Chromatin Association and Tumor Suppression in Clear Cell Renal Carcinoma. *Journal of Biological Chemistry* **2017**.

25. Moriniere, J.; Rousseaux, S.; et al., Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **2009,** *461* (7264), 664-8.

Chapter 3: Recombinant Expression of BAF180 in *E. coli*

Expression of heterologous proteins is a time-consuming and difficult task, as the specific characteristics of the individual protein must be considered alongside the characteristics of the expression system. *Escherichia coli* (*E. coli*) is the most commonly used heterologous protein host, because it is cheap and easy.[26-30] Additionally, the fast cell growth of *E. coli* makes it an efficient factory for the production of large amounts of recombinant protein; large quantities can often be produced in one day and the recombinant protein can comprise up to 50% of the total cellular protein content.[26] However, this fast production can be a disadvantage as well, especially when expressing eukaryotic proteins. It can often result in misfolded, insoluble, and/or inactive proteins, which are often sequestered in inclusion bodies.[26,28] Proteins found in inclusion bodies can be resolubilized and/or refolded after cell lysis, but this is a difficult process and is not always successful. Furthermore, as is increasingly understood, the production of an active protein does not depend solely on proper folding; often, a specific post-translational modification (PTM) is required in order for a protein to be active. *E. coli* is unable to perform post-translational modifications.

The successful expression of heterologous proteins in *E. coli* is often dependent on the protein size and the presence of rare codons. Large proteins are more difficult to express than smaller proteins, and often larger proteins form inclusion bodies.[26] Another problem is rare codons and codon bias. Codon bias is the preference for a particular codon of an amino acid. These codon biases affect protein translation. For example, translation of glutamine is highly biased in humans: 26.5% of codons for glutamine are

CAA, whereas 73.5% are CAG. The codon usage is often different between hosts, as demonstrated when comparing the codon usage for arginine between humans and *E. coli* B strains. In humans, the AGA codon is used 21.5% of the time, but in *E. coli* it is only used 5% of the time. The CGC codon is used 18% of the time in humans, but used 40% of the time in *E. coli*. These so-called "rare codons" (such as those for arginine) can lead to translational stalling, premature termination, frameshifts, and mis-incorporation of amino acids.[26-28]

One of the most commonly used promoters for *E. coli* expression is the T7 promoter.[26] T7 actually refers to the use of T7 RNA polymerase, which has transcription rates five times higher than the native RNA polymerase (of *E. coli*).[26,28] Cell strains denoted with "DE3" have the T7 gene present in its genome. The T7 gene itself is controlled by the *lacUV5* promoter. When IPTG (isopropyl β-D-1-thiogalatopyranoside) is added, the *lacUV5* promoter is induced, resulting in expression of the T7 polymerase, which in turn causes expression of the heterologous protein.[26,28]

The pET30a vector and Rosetta(DE3) cells were used for expression of BAF180 in *E. coli*. The pET30a vector from Novagen was chosen for expression of BAF180 because the pET30 vectors were readily available. The pET30 system of vectors possess kanamycin resistance and a T7 promoter; thus, expression of the recombinant gene is inducible by the addition of IPTG. For expression, the Rosetta(DE3) *E. coli* cell strain was chosen. It is a derivative of BL21(DE3), which is deficient in *lon* and *ompT* proteases and has the T7 gene on a chromosome. Rosetta possesses a plasmid with chloramphenicol resistance, which codes for 7 rare tRNAs: AUA, AGG, AGA, CGG,

CUA, CCC, and GGA.[26,28] There are a few other cell strains that have the ability to make some rare tRNAs, but Rosetta is capable of producing more rare tRNAs than other strains. The BAF180 gene sequence contains a lot of rare codons, so it was reasoned that expression in the Rosetta(DE3) strain would give the highest probability of success.

3.1 Materials

The BAF180 gene sequence (pBabepuroBAF180; isoform 2; Genbank: AF177387.1) was a gift from Ramon Parson's lab (Addgene #41078).[7] EcoRI-HF (cat. R3101), TriDye 1 kb ladder (cat. N3272), Instant Sticky-end Master Mix (cat. M0370) and Calf Intestinal Phosphatase (cat. M0290) were purchased from New England BioLabs (Ipswich, MA). Zyppy™ Plasmid Miniprep Kit was purchased from Zymo Research Corp. (Irvine, CA). Wizard® SV Gel and PCR Clean-Up Kit (cat. A9281) was purchased from Promega Corp. (Madison, WI). Micropulser electroporator and 0.2 cm electroporation cuvettes (cat. 1652086) were from Bio-Rad, Inc. (Hercules, CA). Tris hydrochloride and kanamycin sulfate were purchased from Amresco (VWR: Radnor, PA). Rosetta(DE3) cells (cat. 70954), pET30 vectors (cat. 69909-3) and NaCl were purchased from EMD Millipore (Billerica, MA). DH5α cells (cat. 18259012), lysozyme (cat. BP535), BME (betamercaptoethanol, cat. BP176), LB broth Miller (cat. BP1426), and ampicillin (cat. BP1760) were purchased from ThermoFisher Scientific (Waltham, MA). Chloramphenicol (cat. 227920250), IPTG (Isopropyl β-D-1-thiogalactopyranoside, cat. 302790250), and PMSF (phenylmethylsulfonyl fluoride, cat. 215740050) were purchased from Acros Organics (NJ). Urea, DTT (dithiothreitol, cat. 43815), and TCEP (Tris(2-carboxyethyl)phosphine hydrochloride, cat. 646547) were from Sigma-Aldrich

(St. Louis, MO). DTE (dithioerythritol, cat. 2185D) was from Research Organics (Cleveland, OH). The Digital Sonifier 450 and accessories were from Branson Ultrasonics (Dansbury, CT).

3.2 BAF180 Expression in *E. coli*

3.2.1 Cloning

BAF180 was cloned into the *E. coli* strain Rosetta(DE3) using the pET30a vector. The BAF180 gene sequence (pBabepuroBAF180; isoform 2; GenBank: AF177387.1) was a gift from Ramon Parson's lab (Addgene #41078).[7] As received, the BAF180 gene had been inserted into the pBabepuro vector using EcoRI on both the 5' and 3' ends. The 5' EcoRI site was immediately upstream of the start codon and only 50 base pairs separated the 3' EcoRI site and the stop codon. For time efficiency, the 3' restriction enzyme site was not changed. Both pBabepuroBAF180 and pET30a were digested with EcoRI and purified using the Promega Wizard® SV Gel and PCR Clean-Up Kit. pET30a was dephosphorylated using Calf Intestinal Phosphatase to prevent re-ligation and gel purified again. The BAF180 gene (4961 bp) was ligated into pET30a (5422 bp) using a 5:1 (insert:vector) ratio with Instant Sticky-end Master Mix. The resultant plasmid (BAF180-pET30a) was 10,381 base pairs (Figure 3.1). 2 µL of the ligation reaction was immediately transformed into chemically competent DH5α cells using the protocol outlined in the pET30 manual.[7] Cells were then grown on LBK plates (50 µg/mL kanamycin) overnight at 37 °C. 12 colonies were produced. Each colony was grown in 15 mL LBK overnight at 37 °C and 250 rpm. 5 mL of this was used to prepare freezer stock

and 10 mL was miniprepped. The minipreps were used to confirm the plasmid size (Figure 3.2A).

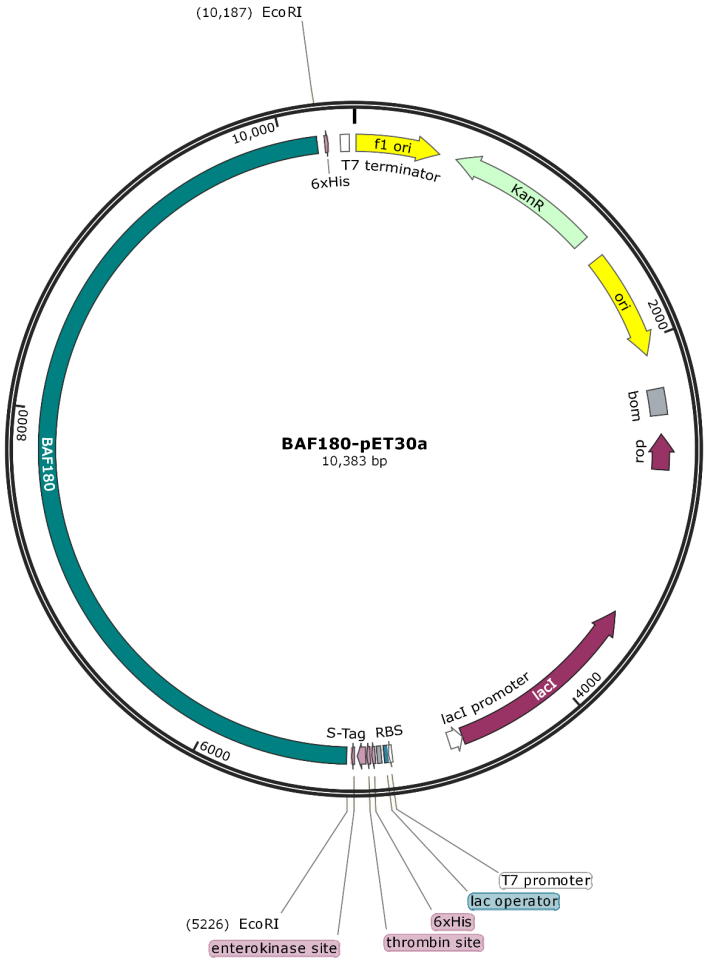


Figure 3.1: Plasmid map of BAF180-pET30a. The BAF180 gene sequence is highlighted in teal in the vector.

When a gene is inserted into a vector using the same restriction enzyme site on both the 5' and 3' ends, it is possible for the gene to be inserted in the 3' to 5' direction, instead of the correct 5' to 3'. BAF180 was inserted into pET30a using EcoRI on both the 5' and 3' ends of the gene (Figure 3.1), so the 12 transformation products were screened for insert direction using PCR and restriction enzyme digests. For PCR, a forward primer

for the T7 promoter (on the vector) and a reverse primer for bromodomain 2 were used to confirm that BAF180 was inserted in the correct direction. The correct product size (1064 bp) was seen on the gel (Figure 3.2B). Additional PCR reactions were performed to confirm that full-length BAF180 was inserted into the pET30a vector. To perform diagnostic digests, a restriction enzyme was chosen that cut twice within the BAF180-pET30a plasmid. The product sizes would change depending on the insert direction (Figure 3.3). Unfortunately, results were inconsistent, possibly due to mixed-transformation colonies (where some of the cells had the insert and some did not).

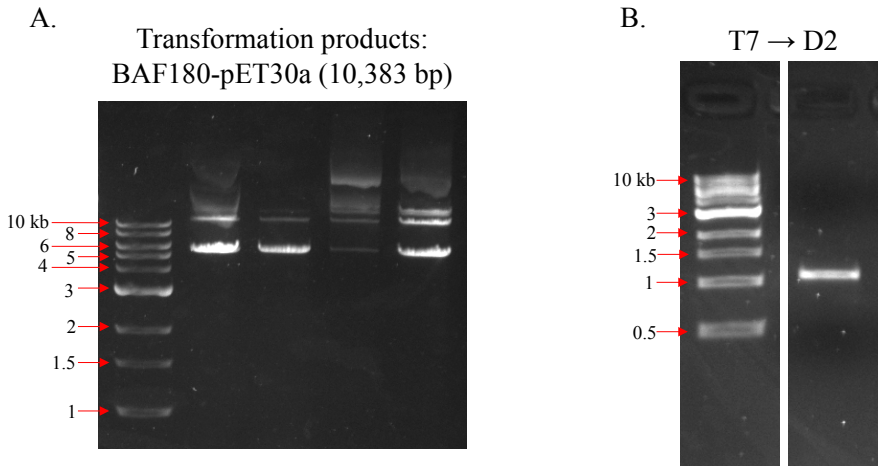


Figure 3.2: Screening of the BAF180-pET30a transformants. (A) Four of the transformants were miniprepped and run on a 1% TBE agarose gel. The expected plasmid size was 10,383 bp. The bands slightly above the 10 kb band on the ladder were believed to be BAF180-pET30a. It was thought that the other bands seen were supercoiled plasmids. (B) PCR screen to confirm the correct insert size. The T7 → D2 product was 1064 bp.

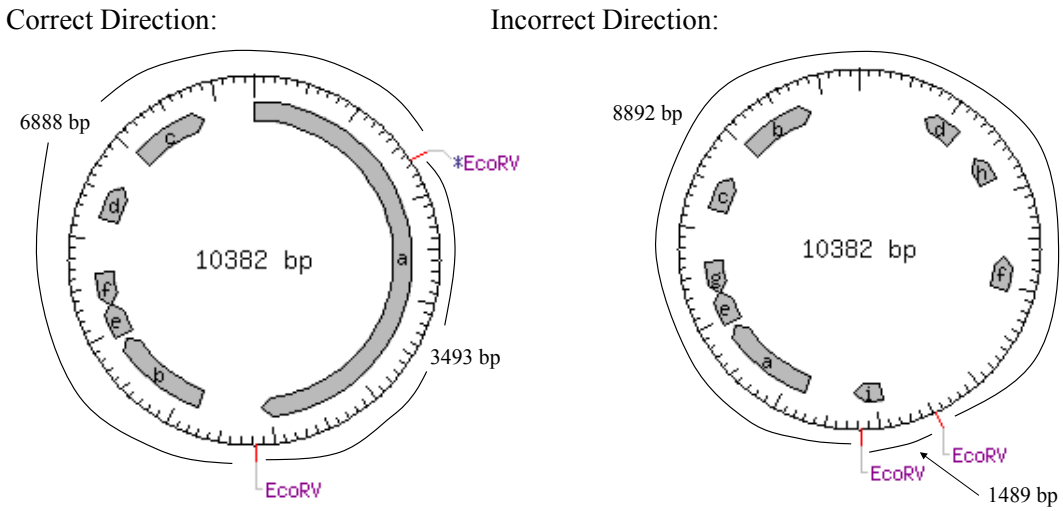


Figure 3.3: Theory behind using restriction enzyme diagnostic digests. If BAF180 had been inserted in the 5' to 3' direction ("correct" direction), digestion of BAF180-pET30a with EcoRV would have resulted in a product approximately 6900 base pairs (bp) and 3500 bp. Insertion of BAF180 in the 3' to 5' direction ("incorrect") would have resulted in products approximately 8900 bp and 1500 bp.

After it was confirmed that BAF180 had been inserted in the 5' to 3' direction, one of the miniprepped BAF180-pET30a plasmids was transformed into electrocompetent *E. coli* Rosetta(DE3) cells using the Bio-Rad Micropulser electroporator on the "ec2" setting. A 0.2 cm electroporation cuvette was used. The electroporation procedure outlined in the Micropulser manual was followed.[31] The cells were then grown overnight on LBCK plates (35 µg/mL chloramphenicol and 50 µg/mL kanamycin).

3.2.2 Expression

To test for BAF180 expression, the following variables were manipulated: growth temperature (30 °C and 37 °C); lysis conditions (native and denaturing); incubation temperature during lysis (4 °C to 25 °C); incubation time during lysis (30 minutes to overnight); IPTG concentration (0.25 mM, 0.50 mM, 1.0 mM, 1.5 mM); induction time (1-24 hours); elimination of possible protein oxidation during lysis by addition of

reducing agents (5% BME, 100 mM DTT, 100 mM DTE, 20 mM TCEP); and inhibition of proteases using PMSF (Table 3.1). No protein band around 190 kDa was ever seen on SDS-PAGE gels.

Table 3.1: Variables manipulated while attempting to express BAF180 in *E. coli* cells.

Variables Tested:	**Actions Taken:**
Growth Temperature	30 °C and 37 °C
Lysis Conditions	Native and Denaturing
Lysis Incubation	Temperature (4 °C to 25 °C) and Time (30 minutes to overnight)
IPTG Concentration	0.25 mM, 0.50 mM, 1.0 mM, 1.5 mM
Induction Time	1-24 hours
Reducing Agents	5% BME, 100 mM DTT, 100 mM DTE, 20 mM TCEP
Proteases	PMSF

3.3 Other Considerations

When efforts to express BAF180 in *E. coli* failed, several different options were explored. Often, large proteins are found in inclusion bodies, where they are insoluble and inactive. The presence of certain protein tags has been found to increase the solubility, thus preventing a recombinant protein from being included in an inclusion body. Fusion tags are polypeptide sequences (such as a 6x His-tag) or proteins (such as GST) which are expressed in-frame with the recombinant protein, thus resulting in a fusion protein which contains both the recombinant protein of interest and the tag. The use of fusion tags has long been known to increase the solubility of some recombinant proteins expressed in *E. coli*.[26,32-33]

Some of the most commonly used tags are 6x His, GST (glutathione S-transferase), and MBP (maltose binding protein).[26] Vectors are available which allow for in-frame expression of one of these tags with the recombinant protein gene. Fusion tags also have another benefit in that purification of the fusion protein is simplified because affinity methods can be employed (such as IMAC for a 6x His-tag). However, there are drawbacks to using tags during recombinant protein expression. The most notable is that the fusion tag can disrupt the biological functions of the recombinant protein. For mechanistic studies, enzymatic cleavage of the fusion tag from the recombinant protein is often needed.[26] Additionally, protein solubility is not equivalent with protein quality; sometimes the soluble proteins are biologically inactive, which essentially negates the use of a solubility tag.[34-35]

Possible co-expression of chaperone proteins was also explored. Chaperone proteins aid in properly folding newly synthesized proteins into their native state. Overexpression of recombinant proteins in *E. coli* often leads to the production of inclusion bodies, where the desired protein is improperly folded and isolated. Co-expression of prokaryotic chaperone proteins, particularly DnaK and GroEL, has been used to increase the yield of some recombinant proteins.[26-27,33-35] DnaK and GroEL are already present in *E. coli* cells; however, when a recombinant protein is overexpressed, there are insufficient quantities of these chaperones and the recombinant protein is misfolded. The chaperones can be co-expressed in the same cell as the recombinant protein by placement on a separate vector (such as those available from Takara Bio Inc.) or by placement on the same vector as the recombinant gene (Duet system by Novagen).

However, the over-expression of these chaperone proteins can have detrimental effects on protein production and on the cell overall, such as cell growth inhibition and enhanced proteolysis of the protein.[34-38]

3.4 Conclusion

While fusion tags and co-expression of chaperone proteins were considered, these were not tried for a few reasons. One, there was doubt whether the fusion tag would be enough to overcome the size of BAF180. They work well with smaller proteins, but the relatively small size of the fusion tags (28 kDa for GST and 43 kDa for MBT)[26] in relation to BAF180 (190 kDa) made it doubtful that the tag would be enough to encourage the proper folding of BAF180. Additionally, there was doubt whether the prokaryotic chaperones would be able to properly fold all of the BAF180 protein; there was a significant possibility that the chaperones would be insufficient, resulting in partial folding of BAF180.

There are many possible reasons why BAF180 failed to express in *E. coli* cells. A few of the most likely are the protein size (190 kDa), presence of multiple domains (9 domains), and the presence of rare codons. Both size and the presence of several domains can influence the expression of recombinant proteins.[26] Large proteins and multi-domain proteins often fail to express or are found in inclusion bodies. Additionally, *E. coli* and other prokaryotic expression systems are limited in their ability to create post-translational modifications (PTMs) on recombinant proteins. Though PTMs on BAF180 are relatively unexplored, there are a few known sites of phosphorylation, which may influence the activity (see Section 1.3.3). Possibly the biggest obstacle to BAF180

expression in *E. coli* (and other bacterial systems) is the codon usage. Though Rosetta(DE3) cells were used for its several rare tRNAs, BAF180 has so many rare codons that this may not have been sufficient; 19 codons have a usage difference greater than 10% (Table 3.2). Though other bacterial systems are available, such as *B. subtilis* and *L. lactis*, the same concerns exist, particularly about codon usage and size (Table 3.3).

Table 3.2: Table showing the codons with a high difference in usage between *E. coli* and BAF180. % difference refers to how much more often a codon is used in BAF180 versus the host. Number refers to how many times that codon is used in BAF180.

Amino Acid	Codon	% Difference	Number
Ala	GCA	13.0	31
Ala	GCT	24.2	32
Arg	AGA	25.0	26
Arg	AGG	17.5	19
Arg	CGA	11.8	18
Asn	AAT	23.8	49
Cys	TGT	21.4	14
Glu	GAG	10.3	57
Gly	GGA	15.0	23
Ile	ATA	15.0	17
Leu	CTT	12.9	36
Lys	AAG	13.3	56
Phe	TTT	13.1	37
Pro	CCA	28.3	48
Pro	CCT	15.0	37
Ser	TCT	11.3	27
STOP	TAA	35.7	1
Thr	ACA	32.3	25
Thr	ACT	10.9	15

Table 3.3: Codon usage comparisons of BAF180 to bacterial expression systems. The number of codons with a certain percentage difference are noted. % difference refers how much more often a codon is used in BAF180 versus the host.

Bacteria							
% Difference	E. coli B	E. coli K12	C. crescentus	S. lividans	L. lactis	B. subtilis	B. megaterium
Codons 10-14.99%:	9	10	5	4	9	5	9
Codons 15-19.99%:	5	2	6	6	4	2	3
Codons 20-29.99%:	3	5	6	8	2	1	0
Codons 30-39.99%:	2	2	5	1	1	1	2
Codons 40-40.99%:	0	0	3	4	1	0	0
Codons ≥ 50%	0	0	4	6	0	0	0
Total codons ≥10% difference:	19	19	29	29	17	9	14
STOP codons % diff:	36.36	35.71	80	90.97	39.99	39.08	32.87
	P. fluorescens	P. haloplanktis	P. putida	M. smegmatis	H. elongata	C. salexigens	L. lactis subsp cremoris MG1363
Codons 10-14.99%:	5	8	10	2	3	3	6
Codons 15-19.99%:	4	5	2	5	5	6	4
Codons 20-29.99%:	3	0	2	8	10	10	4
Codons 30-39.99%:	2	0	1	4	5	3	1
Codons 40-40.99%:	0	2	3	6	2	3	0
Codons ≥ 50%	1	0	1	3	2	2	1
Total codons ≥10% difference:	15	15	19	28	27	27	16
STOP codons % difference:	66.46	40	77.87	93.64	65.22	84.05	31

3.6 References

1. Xia, W.; Nagase, S.; Montia, A. G., et al., BAF180 Is a Critical Regulator of p21 Induction and a Tumor Suppressor Mutated in Breast Cancer. *Cancer Research* **2008,** *68* (6), 1667-1674.

2. Francis, D. M.; Page, R., Strategies to optimize protein expression in E. coli. *Current protocols in protein science* **2010,** *Chapter 5*, Unit 5.24.1-29.

3. Khow, O.; Suntrarachun, S., Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine* **2012,** *2* (2), 159-162.

4. Terpe, K., Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **2006,** *72* (2), 211-222.

5. Chen, R., Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances* **2012,** *30* (5), 1102-1107.

6. Ferrer-Miralles, N.; Villaverde, A., Bacterial cell factories for recombinant protein production; expanding the catalogue. *Microbial Cell Factories* **2013,** *12*, 113-113.

7. Bio-Rad Inc., High Efficiency Electrotransformation of *E. coli*. In *MicroPulser Electroporation Apparatus Operating Instructions and Applications Guide*, Rev B ed.; p 12.

8. Esposito, D.; Chatterjee, D. K., Enhancement of soluble protein expression through the use of fusion tags. *Current opinion in biotechnology* **2006,** *17* (4), 353-8.

9. Peleg, Y.; Unger, T., Resolving bottlenecks for recombinant protein expression in E. coli. *Methods in molecular biology (Clifton, N.J.)* **2012,** *800*, 173-86.

10. Haacke, A.; Fendrich, G.; Ramage, P., et al., Chaperone over-expression in Escherichia coli: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. *Protein expression and purification* **2009,** *64* (2), 185-93.

11. Kolaj, O.; Spada, S.; Robin, S., et al., Use of folding modulators to improve heterologous protein production in Escherichia coli. *Microbial Cell Factories* **2009,** *8* (1), 9.

12. Martínez-Alonso, M.; García-Fruitós, E.; Ferrer-Miralles, N., et al., Side effects of chaperone gene co-expression in recombinant protein production. *Microbial Cell Factories* **2010,** *9* (1), 64.

13. Martínez-Alonso, M.; Toledo-Rubio, V.; Noad, R., et al., Rehosting of Bacterial Chaperones for High-Quality Protein Production. *Applied and Environmental Microbiology* **2009,** *75* (24), 7850-7854.

14. Platas, G.; Rodríguez-Carmona, E.; García-Fruitós, E., et al., Co-production of GroELS discriminates between intrinsic and thermally-induced recombinant protein aggregation during substrate quality control. *Microbial Cell Factories* **2011,** *10* (1), 79.

Chapter 4: The Use of *Pichia pastoris* as an Expression System

The yeast *Pichia pastoris* (*Pichia*) has become increasingly popular as a recombinant host over the last decade. *Pichia* is a methylotrophic budding yeast that is often found in its vegetative state (i.e. not mating).[39] A methylotrophic yeast is capable of utilizing methanol as its sole carbon source. Similar to *E. coli*, *Pichia* has a fast doubling time (approximately 90 minutes in rich medium like YPD)[40] and is relatively inexpensive to grow. But unlike *E. coli*, *Pichia* is capable of performing post-translational modifications and it has eukaryotic protein machinery, which makes it more likely that the recombinant protein will be folded correctly.[41] Expression in *Pichia* has become preferred over other yeast heterologous systems, such as *S. cerevisiae*, partly because *Pichia* has a strong preference for respiratory growth over fermentative growth. This allows *Pichia* cultures to be grown to extremely high cell densities, as toxic ethanol (the product of fermentation) is not present.[40-44]

However, for all its advantages, *Pichia* has one major drawback. The glycosylation pattern produced in *Pichia* is different from that produced in human. Glycosylation of proteins often occurs during secretion. Though proteins produced in *Pichia* are often less hyperglycosylated than recombinant proteins produced in *S. cerevisiae*,[45] *Pichia* produces glycosylation patterns that are different enough from human glycosylation patterns that an immune response can be triggered in a patient. This is especially important in recombinant pharmaceutical production. To overcome this

problem, new strains have recently been engineered that mimic the human glycosylation pattern (such as YSH44 and the GlycoSwitch system).[42-43,45]

4.1 Choosing a Promoter

Several promoters are commercially available for recombinant protein expression in *P. pastoris*. The two most common promoters are the *GAP* promoter (constitutive) and *AOX1* promoters (inducible by methanol). Both of these are strong promoters capable of similarly high levels of recombinant protein expression.[40-48]

The *AOX1* promoter is inducible by addition of methanol. An inducible promoter is exactly what it sounds like: it promotes gene expression when induced. This can be an advantage for the production of some proteins, particularly toxic ones, as the protein is not expressed during the growth phase. A culture is first grown to a particular density before the *AOX1* promoter is induced by addition of methanol. Using an *AOX1* promoter, the recombinant protein can reach up to 30% of the cell biomass.[42] However, there are disadvantages to using the *AOX1* promoter. Because it uses methanol as an inducer, it is unsuitable for large scale expression, because methanol is highly flammable.[40-52] Additionally, the fast production of recombinant protein may result in unfolded or inactive proteins, as an inducible promoter makes a lot of protein all at once.

On the other hand, the *GAP* promoter is a constitutive promoter. A constitutive promoter allows for the constant production of the recombinant protein. This type of promoter does not require an inducer, as it is always “on”. Some researchers have

reported higher recombinant protein production with the *AOX1* promoter than the *GAP* promoter, while others have reported the opposite.[42-46,50,52-55]

Considering the size of the BAF180 protein (190 kDa), it was hypothesized that continuous expression of the protein (under control of the *GAP* promoter) might be more favorable than inducible expression, in terms of cellular stress and protein folding.[44,46,56] Synthesis of a lot of protein all at once might be toxic to the cell or impede growth or there might not be sufficient levels of chaperone proteins available for folding the protein.[40-48,50,53-55,57] The *GAP* promoter is found in pGAPZ vectors.

4.2 Choosing an Expression Strain

There are several *Pichia* strains commercially available. Wild-type (WT) strains include Y-11430 and X33. Other strains have mutations in a biosynthesis pathway (such as *his4* mutants), *AOX1/2* mutants, protease-deficient strains, and glycosylation mutants.[43,45] Strains with a mutation in a biosynthesis pathway, such as *his4*, are only able to grow if the vector (containing the replacement HIS4 gene) has been integrated.[45] Other strains have mutations in one or both of the *AOX* genes, because cells with AOX mutations can sometimes produce larger quantities of recombinant protein.[43,45]

Most of the protease activity in yeast is contained in the membrane bound vacuoles.[56,58] Upon cell lysis, the organelles release their contents. Cell lysis can happen during cell growth, as the cell density increases. The proteases are released into the media wherein secreted proteins are targeted. Therefore, protease-deficient strains have been developed. Most often, these strains have been engineered to be deficient in PEP4. PEP4

is a gene that activates carboxypeptidase Y and protease B1 in vacuoles, therefore mutants without PEP4 do not have either of these proteases.[40-41,43-46,48,58] The use of protease-deficient strains has improved extracellular expression in about 10% of cases,[41] but there is no reported success of using a protease-deficient strain to increase yield of an intracellularly produced protein.[40-41] Protease-deficient strains have several problems, such as much slower growth, faster death on agar plates, lower transformation efficiencies, lower levels of expression, and difficulty utilizing the methanol-inducible *AOX1* promoter.[40-41,43,45] Use of a protease-deficient strain is generally not recommended unless absolutely necessary.[45] The WT strain X33 was chosen for expression because it was readily available.

4.3 Integration of A Recombinant Gene into the *Pichia pastoris* Genome

Expression of a recombinant gene into *Pichia* involves the integration of the gene into the *Pichia* genome through homologous recombination. The plasmid containing the recombinant gene is linearized in a location which shares homology to a location in the *Pichia* genome. Usually this is accomplished by cutting a plasmid within the promoter and then transforming this linearized construct ("expression cassette") into *Pichia* cells. The free DNA ends then promote recombination of the plasmid into the genome, usually with a single-crossover event.[45-46] Incorporation of the recombinant gene is far more stable[41-44,46] than the *E. coli* method of simply inserting a plasmid with its own replication system into the cell. Actually, there are no autonomous plasmids available for *Pichia*,

because *Pichia* seems to spontaneously integrate the plasmid into its genome anyway.[41,44,59]

Due to homologous recombination, multiple copies of the expression cassette can sometimes be integrated into the genome. These "jack-pot" clones are very rare, occurring spontaneously approximately 1% of the time.[60-61] This percentage can be increased by constructing a vector that has multiple copies of the gene in the vector itself.[45] Identification of high-copy number clones is time-consuming and laborious, as it usually requires screening of hundreds of clones. One of the commonly used techniques to isolate multi-copy strains is step-wise increase in antibiotic concentration, where colonies are plated on subsequently higher and higher concentrations of the antibiotic.[40-41,45] Interestingly, while logic would dictate that more copies of the gene would equal more protein, this is not always the case. Sometimes, integration of more than one copy actually leads to a decrease in the recombinant protein production.[43,45,60]

4.4 Choosing to Use *Pichia* as the heterologous system for BAF180 Production

Pichia was chosen as the expression system for BAF180, because the *Pichia* codon usage is similar to the usage for BAF180. BAF180 contains a lot of rare codons, even when compared to the usage in humans (15 codons have a usage difference greater than 10%; see Table 4.1). As such, once *E. coli* failed to express BAF180, other expression systems were investigated. Complete codon analysis was performed on 12 different systems to identify which system had codon usage closest to BAF180: 4 bacterial, 6 yeast, and 2 insect. The results are shown in Table 4.2.

Table 4.1: Codon usage comparisons of BAF180 to humans (native host). The number of codons with a percentage difference greater than 10 are shown. Note that % difference refers to how much more often a codon is used in BAF180 versus the host. Number refers to the number of times that codon is used in the BAF180 gene sequence.

Amino Acid	Codon	% Difference	Number
Ala	GCA	10.88	31
Asn	AAT	23.98	49
Asp	GAT	16.65	65
Cys	TGT	18.02	14
Glu	GAA	17.89	86
His	CAT	18.67	23
Ile	ATT	11.94	37
Leu	CTT	11.49	36
Lys	AAA	16.60	84
Phe	TTT	23.40	37
Pro	CCA	14.45	48
STOP	TAA	70.30	1
Thr	ACA	21.17	25
Tyr	TAT	11.67	42
Val	GTT	14.76	28

Table 4.2: Codon usage comparisons of BAF180 to 14 different expression systems. A positive percentage means the codon usage is biased towards BAF180 (i.e. that particular codon is used more often in BAF180). A negative percentage means the codon usage is biased towards the host organism (i.e. that particular codon is used more often in the host organism).

Codon Difference Comparison of BAF180 to Host Organism

Amino Acid	Codon	Homo sapiens	P. pastoris	S. cervisae	S. pombe	Y. lipolytica	K. lactis	Z. bailii	C. crescentus	S. lividans	L. lactis	P. putida	E. coli K12	E. coli	S. frugiperda	T. ni
Ala	GCA	10.88	10.27	4.84	8.24	22.21	7.94	16.80	31.37	25.44	-1.49	21.36	13.01	12.25	15.52	16.95
Ala	GCC	-9.54	4.71	7.99	12.07	-15.09	10.82	-3.35	-27.77	-23.08	14.36	-20.75	-0.60	2.61	0.99	-3.49
Ala	GCG	-9.53	-4.90	-9.91	-7.51	-7.11	-7.01	-8.99	-32.40	-31.70	-9.53	-24.48	-36.65	-34.57	-16.43	-12.42
Ala	GCT	8.19	-10.08	-2.92	-12.80	-0.02	-11.75	-4.45	28.81	29.34	-3.35	23.87	24.25	19.71	-0.07	-1.04
Arg	AGA	6.19	-20.26	-20.45	4.69	14.27	-30.44	-15.14	26.49	26.37	-1.05	25.89	24.99	22.60	9.00	4.96
Arg	AGG	-0.90	4.41	-0.64	9.82	15.43	6.07	1.14	17.23	15.36	14.46	16.67	17.55	15.72	0.79	0.18
Arg	CGA	8.27	9.18	12.38	2.79	-32.79	14.16	14.06	14.06	13.49	1.54	14.02	11.82	14.09	10.78	9.09
Arg	CGC	-2.42	10.83	10.08	3.67	8.79	12.38	7.48	-42.68	-32.16	6.22	-38.70	-28.38	-23.93	-8.17	-2.53
Arg	CGG	-10.58	4.94	5.65	3.46	-2.05	6.06	5.76	-11.48	-22.11	3.02	-6.02	2.57	-1.10	3.79	2.87
Arg	CGT	-0.56	-9.09	-7.01	-24.45	-3.65	-8.24	-13.32	-3.62	-0.95	-24.19	-11.85	-28.55	-27.38	-16.20	-14.57
Asn	AAC	-23.98	-22.56	-12.04	-5.37	-54.04	-15.72	-21.54	-51.36	-63.58	4.08	-48.81	-23.76	-13.87	-39.21	-41.31
Asn	AAT	23.98	22.56	12.04	5.37	54.04	15.72	21.54	51.36	63.58	-4.08	48.81	23.76	13.87	39.21	41.31
Asp	GAC	-16.65	-5.15	1.93	7.66	-28.39	5.78	-11.41	-40.00	-55.35	11.02	-31.90	1.78	3.36	-23.51	-23.23
Asp	GAT	16.65	5.15	-1.93	-7.66	28.39	-5.78	11.41	40.00	55.35	-11.02	31.90	-1.78	-3.36	23.51	23.23
Cys	TGC	-18.02	0.11	-0.66	-1.86	-17.98	11.15	-1.99	-49.36	-53.57	14.60	-44.29	-21.38	-21.53	-24.49	-19.39
Cys	TGT	18.02	-0.11	0.66	1.86	17.98	-11.15	1.99	49.36	53.57	-14.60	44.29	21.38	21.53	24.49	19.39
Gln	CAA	6.29	-28.21	-36.47	-38.84	13.84	-38.59	-14.58	18.77	24.91	-49.38	9.09	2.39	-2.63	-10.03	-11.76
Gln	CAG	-6.29	28.21	36.47	38.84	-13.84	38.59	14.58	-18.77	-24.91	49.38	-9.09	-2.39	2.63	10.03	11.76
Glu	GAA	17.89	3.82	-10.19	-7.69	35.67	-14.72	9.66	31.47	39.22	-20.65	12.34	-10.30	-1.87	14.70	3.40
Glu	GAG	-17.89	-3.82	10.19	7.69	-35.67	14.72	-9.66	-31.47	-39.22	20.65	-12.34	10.30	1.87	-14.70	-3.40
Gly	GGA	2.72	-4.93	6.17	-3.95	-0.65	7.85	13.55	22.86	18.74	-8.87	27.08	15.04	17.45	-0.30	-0.40
Gly	GGC	2.43	22.35	16.81	19.52	1.83	26.43	14.18	-36.15	-27.84	23.43	-57.78	-9.95	-5.25	5.25	6.48
Gly	GGG	-10.53	4.60	2.55	5.65	9.08	5.40	5.50	1.81	-2.00	2.82	12.19	2.60	-3.42	7.19	8.06
Gly	GGT	5.38	-22.02	-25.54	-21.22	-10.26	-39.67	-33.23	11.48	11.10	-17.38	18.51	-7.69	-8.78	-12.13	-14.14
His	CAC	-18.67	-3.96	3.14	11.65	-26.66	6.60	-12.91	-31.19	-49.87	13.42	-26.26	-5.80	-4.07	-24.74	-26.05
His	CAT	18.67	3.96	-3.14	-11.65	26.66	-6.60	12.91	31.19	49.87	-13.42	26.26	5.80	4.07	24.74	26.05
Ile	ATA	5.17	4.02	-5.26	0.03	19.18	-2.51	5.63	20.62	19.49	2.02	17.51	14.99	14.88	6.10	9.17
Ile	ATC	-17.12	-1.62	3.50	9.22	-23.13	-2.72	-13.08	-62.10	-58.93	10.56	-44.52	-4.83	-14.45	-24.48	-24.38
Ile	ATT	11.94	-2.41	1.76	-9.25	3.94	5.23	7.45	41.49	42.45	-12.58	27.01	-10.16	-0.43	18.38	15.21
Leu	CTA	3.14	-0.88	-3.85	1.38	5.48	-2.73	-4.04	8.87	9.36	-0.03	7.89	5.09	7.43	1.71	1.90
Leu	CTC	-2.43	9.20	11.39	9.73	-12.14	12.34	5.99	0.13	-18.69	9.99	4.07	6.76	5.53	-3.83	-5.53
Leu	CTG	-16.97	7.07	11.56	16.03	-14.86	16.93	11.01	-45.60	-33.55	17.29	-39.92	-23.46	-22.69	-6.98	-2.07
Leu	CTT	11.49	8.18	11.75	-1.12	6.83	11.56	4.88	18.45	21.39	2.40	18.29	12.95	14.15	12.83	9.70
Leu	TTA	-0.12	-8.63	-20.02	-19.31	5.95	-18.49	-2.69	7.19	7.33	-28.32	6.22	-7.44	-6.91	-1.98	-2.57
Leu	TTG	4.90	-14.94	-10.82	-6.71	8.73	-19.62	-15.15	10.97	14.16	-1.34	3.45	6.10	2.49	-1.75	-1.43
Lys	AAA	16.60	13.04	2.39	-1.89	44.34	7.39	15.05	52.43	50.35	-19.49	31.17	-13.28	-16.92	24.74	23.55
Lys	AAG	-16.60	-13.04	-2.39	1.89	-44.34	-7.39	-15.05	-52.43	-50.35	19.49	-31.17	13.28	16.92	-24.74	-23.55
Met	ATG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Phe	TTC	-23.40	-15.85	-57.41	1.59	-32.25	-24.49	-22.62	-57.86	-66.28	5.65	-49.17	-13.07	-9.26	-42.96	-45.04
Phe	TTT	23.40	15.85	57.41	-1.59	32.25	24.49	22.62	57.86	66.28	-5.65	49.17	13.07	9.26	42.96	45.04

Amino Acid	Codon	Homo sapiens	P. pastoris	S. cervisae	S. pombe	Y. lipolytica	K. lactis	Z. bailii	C. crescentus	S. lividans	L. lactis	P. putida	E. coli K12	E. coli	S. frugiperda	T. ni
Pro	CCA	14.45	0.43	0.39	15.01	31.68	-7.70	2.84	38.02	38.01	-0.40	27.60	28.34	23.80	14.44	16.35
Pro	CCC	-13.05	4.33	3.85	2.05	-28.20	10.38	-4.45	-14.99	-21.46	10.41	-4.01	5.94	13.42	-8.58	-9.00
Pro	CCG	-5.18	-2.52	-5.92	-3.57	-3.52	-3.71	1.12	-49.06	-43.37	-3.09	-43.67	-49.33	-55.30	-9.76	-8.76
Pro	CCT	3.78	-2.24	1.68	-13.49	0.04	1.03	0.49	26.03	26.82	-6.92	20.08	15.05	18.08	3.90	1.41
Ser	AGC	-3.16	11.70	9.86	10.93	9.66	13.13	6.85	-11.38	-6.28	10.51	-19.96	-11.86	-4.17	2.67	1.56
Ser	AGT	5.87	5.83	4.91	4.78	13.63	5.35	6.85	18.16	16.49	-2.62	11.60	6.60	5.09	7.15	6.20
Ser	TCA	7.45	4.35	1.48	2.97	14.64	2.75	9.11	19.13	18.87	-6.48	17.13	7.12	11.85	5.18	5.90
Ser	TCC	-9.30	-7.26	-3.50	-0.61	-17.53	-2.86	-4.17	-4.38	-24.00	6.81	-2.87	1.73	-1.39	-8.20	-6.01
Ser	TCG	-4.61	-7.96	-8.80	-7.90	-15.11	-8.46	-8.39	-40.65	-24.54	-4.59	-23.73	-14.94	-19.07	-11.84	-9.34
Ser	TCT	3.74	-6.66	-3.95	-10.18	-5.29	-9.91	-10.24	19.11	19.46	-3.62	17.83	11.35	7.69	5.05	1.70
STOP	TAA	70.30	49.26	52.88	40.93	54.02	45.85	62.50	80.01	90.97	35.99	77.87	35.71	36.36	36.62	33.33
STOP	TGA	-46.62	-19.85	-22.57	-20.91	-15.63	-18.27	-37.50	-47.26	-17.42	-22.37	-11.67	-35.71	-9.09	-18.31	-28.89
STOP	TAG	-23.67	-29.41	-30.31	-20.03	-38.39	-27.57	-25.00	-32.76	-73.55	-13.62	-66.20	0.00	-27.27	-18.31	-4.44
Thr	ACA	21.17	21.17	15.21	19.27	31.70	18.99	20.11	42.72	40.45	6.00	37.94	32.25	34.07	22.23	26.47
Thr	ACC	-4.90	-0.07	3.78	5.84	-23.55	3.15	-1.28	-37.27	-34.38	11.26	-38.78	-21.35	-21.58	-7.00	-9.75
Thr	ACG	-22.46	-8.85	-11.74	-10.20	-9.27	-9.48	-11.03	-29.99	-29.53	-10.80	-16.66	-21.78	-25.41	-15.97	-11.51
Thr	ACT	6.19	-12.26	-7.25	-14.91	1.12	-12.67	-7.80	24.54	23.46	-6.46	17.50	10.87	12.92	0.74	-5.22
Trp	TGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tyr	TAC	-11.67	-9.12	-0.04	9.28	-38.41	-1.74	-18.25	-10.01	-49.97	19.36	-27.74	-2.58	12.63	-26.84	-31.73
Tyr	TAT	11.67	9.12	0.04	-9.28	38.41	1.74	18.25	10.01	49.97	-19.36	27.74	2.58	-12.63	26.84	31.73
Val	GTA	3.64	-0.17	-5.58	-5.20	9.96	-2.34	-0.21	13.68	10.71	-7.89	4.99	-2.06	1.78	-1.80	-2.69
Val	GTC	-2.65	-2.09	0.29	3.52	-15.01	-1.79	-3.14	-30.28	-33.69	4.33	-8.32	3.53	2.93	-7.36	-10.18
Val	GTG	-15.75	11.40	11.50	16.80	-2.31	13.01	0.80	-9.19	-6.35	16.29	-19.32	-9.12	-12.66	-3.75	1.42
Val	GTT	14.76	-9.14	-6.21	-15.12	7.36	-8.88	2.55	25.79	29.33	-12.72	22.65	7.65	7.94	12.91	11.44

BAF180 codon usage was similar to that of *Pichia*; only 11 codons had a usage difference greater than 10%, meaning that only 11 codons were used more than >10% more frequently in BAF180 than in *Pichia* (Table 4.3).

Table 4.3: Codon usage comparisons of BAF180 to *Pichia pastoris*. The number of codons with a percentage difference greater than 10 are shown. Note that % difference refers to how much more often a codon is used in BAF180 versus the host. Number refers to how many of that particular codon are found in the BAF180 gene sequence.

Amino Acid	Codon	% Difference	Number
Ala	GCA	10.3	31
Arg	CGC	10.8	15
Asn	AAT	22.6	49
Gln	CAG	28.2	41
Gly	GGC	22.3	30
Lys	AAA	13.0	84
Phe	TTT	15.9	37
Ser	AGC	11.7	25
STOP	TAA	49.3	1
Thr	ACA	21.2	25
Val	GTG	11.4	26

Pichia pastoris was chosen because it is a eukaryotic expression system, therefore it is capable of performing post-translational modifications and it has eukaryotic protein machinery. This makes it more likely that the recombinant protein will be folded correctly.[41] *Pichia* also has codon usage similar to BAF180, making it more likely that the protein will be transcribed from the DNA.

4.5 References

1. Tolstorukov, I.; Cregg, J., Classical Genetics. In *Pichia Protocols*, Cregg, J., Ed. Humana Press: 2007; Vol. 389, pp 189-201.

2. Cregg, J.; Tolstorukov, I.; Kusari, A., et al., Expression in the Yeast Pichia pastoris. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 169-189.

3. Cregg, J., Distinctions Between Pichia pastoris and Other Expression Systems. In *Pichia Protocols*, second ed.; Cregg, J., Ed. 2007; pp 1-10.

4. Brondyk, W., Selecting an Appropriate Method for Expressing a Recombinant Protien. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 135-136.

5. Felber, M.; Pichler, H.; Ruth, C., Strains and Molecular Tools for Recombinant Protein Production in Pichia pastoris. In *Yeast Metabolic Engineering: Methods and Protocols*, Mapelli, V., Ed. Springer New York: New York, NY, 2014; pp 87-111.

6. Li, P.; Anumanthan, A.; Gao, X. G., et al., Expression of recombinant proteins in Pichia pastoris. *Applied biochemistry and biotechnology* **2007,** *142* (2), 105-24.

7. Lin-Cereghino, J.; Lin-Cereghino, G., Vectors and Strains for Expression. In *Pichia Protocols*, Cregg, J., Ed. Humana Press: 2007; Vol. 389, pp 11-25.

8. Daly, R.; Hearn, M. T. W., Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition* **2005,** *18* (2), 119-138.

9. Cos, O.; Ramón, R.; Montesinos, J. L., et al., Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: A review. *Microbial Cell Factories* **2006,** *5* (1), 17.

10. Ahmad, M.; Hirz, M.; Pichler, H., et al., Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* **2014,** *98* (12), 5301-17.

11. Wang, A.; Wang, S.; Shen, M., et al., High level expression and purification of bioactive human α-defensin 5 mature peptide in Pichia pastoris. *Applied Microbiology and Biotechnology* **2009,** *84* (5), 877-884.

12. Xu, J.; Wang, L. N.; Zhu, C. H., et al., Co-expression of recombinant human prolyl with human collagen alpha1 (III) chains in two yeast systems. *Letters in applied microbiology* **2015,** *61* (3), 259-66.

13. Li, J.; Zhang, J.; Lai, B., et al., Cloning, Expression, and Characterization of Capra hircus Golgi α-Mannosidase II. *Applied biochemistry and biotechnology* **2015,** *177* (6), 1241-1251.

14. Maity, N.; Thawani, A.; Sharma, A., et al., Expression and Control of Codon-Optimized Granulocyte Colony-Stimulating Factor in Pichia pastoris. *Applied biochemistry and biotechnology* **2016,** *178* (1), 159-72.

15. Zhang, A.-L.; Luo, J.-X.; Zhang, T.-Y., et al., Recent advances on the GAP promoter derived expression system of Pichia pastoris. *Molecular Biology Reports* **2009,** *36* (6), 1611-1619.

16. Pepeliaev, S.; Krahulec, J.; Cerny, Z., et al., High level expression of human enteropeptidase light chain in Pichia pastoris. *J Biotechnol* **2011,** *156* (1), 67-75.

17. Hohenblum, H.; Gasser, B.; Maurer, M., et al., Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris. *Biotechnology and bioengineering* **2004,** *85* (4), 367-75.

18. Puxbaum, V.; Mattanovich, D.; Gasser, B., Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. *Applied Microbiology and Biotechnology* **2015,** *99* (7), 2925-2938.

19. He, D.; Luo, W.; Wang, Z., et al., Combined use of GAP and AOX1 promoters and optimization of culture conditions to enhance expression of Rhizomucor miehei lipase. *J Ind Microbiol Biotechnol* **2015,** *42* (8), 1175-82.

20. Gleeson, M. A.; White, C. E.; Meininger, D. P., et al., Generation of protease-deficient strains and their use in heterologous protein expression. *Methods in molecular biology (Clifton, N.J.)* **1998,** *103*, 81-94.

21. Cregg, J., DNA-Mediated Transformation. In *Pichia Protocols*, Cregg, J., Ed. Humana Press: 2007; Vol. 389, pp 27-42.

22. Aw, R.; Polizzi, K. M., Can too many copies spoil the broth? *Microbial Cell Factories* **2013,** *12* (1), 128.

23. Invitrogen. pGAPZ A, B, and C; pGAPZa A, B, and C. *Pichia* expression vectors for constitutive expression and purification of recombinant proteins 2010, p. 44.

Chapter 5: Cloning and Expression of BAF180 in *Pichia pastoris*

After attempts to express BAF180 in *E. coli* were unsuccessful (see Chapter 3), the yeast *Pichia pastoris* was chosen. There were two primary reasons why *P. pastoris* was chosen (more details in Chapter 4):

(1) Eukaryotic system: *P. pastoris* is a eukaryotic organism, capable of performing post-translational modifications in a manner similar to humans. It also possesses eukaryotic transcriptional machinery.

(2) Codon Usage: BAF180 has unusual codon usage. The codon usage of *P. pastoris* is similar to BAF180 and has only 11 codons have a usage difference greater than 10% (as opposed to 19 for *E. coli*).

The pGAPZA vector was chosen for expression because it contains the constitutive *GAP* promoter. A constitutive promoter permits constant low level expression of a recombinant gene; in other words, the promoter is always "on." It was hypothesized that constant low-level production of BAF180 would be more favorable in terms of cellular stress and protein production. Continuous low level synthesis of BAF180 might be less toxic to the cell than the synthesis of a large amount of protein at one time (as happens with inducible promoters). Additionally, continuous low level expression may increase the probability of BAF180 folding correctly. BAF180 is such a large protein that the protein folding machinery (such as chaperone proteins) may be more likely to fold BAF180 correctly if the machinery is doing so constantly and not being bombarded all at

once as happens with an inducible promoter. In this chapter, several topics are discussed including the cloning of BAF180 into *P. pastoris*. Cell growth conditions and lysis procedures were also investigated to determine which conditions resulted in the highest expression and recovery levels of BAF180. The expression of BAF180 was confirmed with western blot.

5.1. Materials

Pichia pastoris wild-type strain X33 and pGAPZA vector were purchased from Invitrogen (Carslbad, CA). Zeocin (cat. ant-zn-1p) was purchased from Invivogen (San Diego, CA). The BAF180 gene sequence (pBabepuroBAF180; isoform 2; GenBank: AF177387.1) was a gift from Ramon Parson's lab (Addgene #41078).[7] Primers were purchased from Integrated DNA Technologies (IDTDNA: Coralville, IA). Q5 polymerase (M0491), T4 ligase (M0202), all restriction enzymes, TriDye 1 kb ladder (N3272), and Color Prestained Broad Range Protein Marker (P7712) were purchased from New England BioLabs (Ipswich, MA). PCR mutagenesis was performed on the G-Storm GS482 Thermocycler (Gene Technologies, UK). Tryptone, peptone, yeast extract, and dextrose were purchased from Amresco (VWR: Radnor, PA). Urea, sodium phosphate monobasic (NaH_2PO_4), Lyticase from *Arthrobacter luteus*, phenol:chloroform:isoamyl alcohol (25:24:1), chloroform:isoamyl alcohol (24:1), sodium acetate, and 100% ethanol were purchased from Sigma Aldrich (St. Louis, MO). Zirconia/silica disruption beads (0.5 mm) were purchased from Research Products International Corp. (Mt. Prospect, IL). DNA purification kits (Zyppy™ Plasmid Miniprep Kit, Zymo Gel Extraction Kit, and Zymo DNA Clean & Concentrator™) were purchased

from Zymo Research Corp. (Irvine, CA). The Wizard® SV Gel and PCR Clean-Up Kit (cat. A9281) was purchased from Promega Corp. (Madison, WI). Samples were concentrated with the Savant Refrigerator Condensation Trap and Savant Speed Vacuum Concentrator (ThermoFisher: Waltham, MA). *E. coli* subcloning strain DH5α, Ampicillin (cat. BP1760), 0.45 μm nitrocellulose membrane (cat. 77010), 0.45 μm PVDF membrane (cat. 88585), blotting paper (cat. 8860), NuPAGE Tris-Acetate gels 3-8% (cat. EA0375), NuPAGE LDS Sample Buffer 4x (cat. NP0007), NuPAGE Tris-acetate running buffer (LA0041), NuPAGE transfer buffer (cat. NP0006), NuPAGE antioxidant (cat. NP0005), NuPAGE reducing agent (cat. NP0004) and Pierce Fast Western Blot Kit, ECL substrate (cat. 35050) were purchased from ThermoFisher (Waltham, MA). Anti-*PBRM1* antibody (cat. 12563-1-AP, rabbit polyclonal) was purchased from Proteintech (Rosemont, IL). Western blot transfer was performed using a Hoefer TE22 transfer tank (Holliston, MA). Ponceau staining solution (cat. AC2125) was purchased from Azure Biosystems (Dublin, CA). Dialysis membranes were purchased from Spectrum Labs (Rancho Dominguez, California): Spectra/Por 1 (6-8 kD MWCO membrane; part 132660) and Spectra/Por 6 (50 kD MWCO membrane; part 132544). SDS, glycine, Tris hydrochloride, Tris base, glycerol, guanidine hydrochloride, and Protease Inhibitor Cocktail (cat. 97063-970) were purchased from Amresco (VWR: Radnor, PA). EDTA was purchased from Mallinckrodt Baker Inc. (Phillipsburg, NJ). NaCl was purchased from EMD Millipore (Billerica, MA). MicroPulser electroporator and 0.2 cm electroporation cuvettes (cat. 1652086) were from Bio-Rad, Inc. (Hercules, CA). PMSF (phenylmethylsulfonyl fluoride, cat. 215740050) was purchased from Acros Organics.

5.2 Construction and Transformation of Expression Plasmid

Before BAF180 could be cloned into *Pichia pastoris*, a few modifications needed to be made to the gene sequence. (1) A Kozak consensus sequence was added to facilitate gene expression; (2) Insertion of a spacer between 5' restriction enzyme site and the Kozak consensus sequence; (3) The 3' restriction enzyme site was changed to allow for directional cloning; (4) The endogenous stop codon was deleted; (5) An enterokinase site was inserted.

A Kozak consensus sequence was added immediately upstream of the start codon, for use as a transcription initiation start site (Figure 5.1). The BAF180 sequence contained a start codon, but a yeast-specific Kozak consensus sequence needed to be added because the chosen vector did not contain one. To make the Kozak sequence, three adenosine residues were added upstream of the BAF180 start codon (AAAATGG). Literature indicated that higher initiation rates are observed when adenosine residues are located at the -3 to -1 positions and when a guanine is located at the +4 position relative to the start codon (the guanine was already present).[61-62] A spacer of 6 base pairs was added between the 5' restriction enzyme site and the Kozak consensus sequence (Figure 5.1). Ligation efficiency is decreased when a restriction enzyme site and a start codon are immediately adjacent to each other.

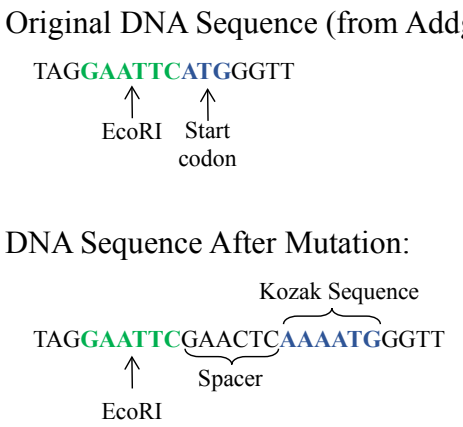


Figure 5.1: Insertion of the Kozak consensus sequence and a spacer into the BAF180 sequence, to facilitate expression in the yeast *Pichia pastoris*. A spacer was added between the restriction enzyme site (EcoRI) and the Kozak consensus sequence.

Additionally, the 3' restriction enzyme site was changed to allow for directional cloning. The group who constructed the parent vector (pBabepuroBAF180) used EcoRI on the 5' and 3' ends of the gene. When the same restriction enzyme site is used on both ends of the gene, there is a possibility that the gene will be inserted in the 3' to 5' direction instead of the correct 5' to 3' direction. To facilitate directional cloning, the 3' site was changed to KpnI (GGTACC) (Figure 5.2). Analysis of the BAF180 sequence revealed limited options for the choice of restriction enzyme; of all commercially available enzymes, only 67 restriction enzymes did not cut within the BAF180 sequence. When comparing those to the available sites in pGAPZA, only five enzymes remained: EcoRI, PmlI, KpnI, SacII, and NotI. PmlI was eliminated as it creates blunt ends. EcoRI was eliminated as it is already being used on the 5' end. KpnI is downstream of EcoRI and creates hanging ends. KpnI was chosen partially because the Thompson lab has had success with this enzyme in the past.

As mentioned, the pGAPZA vector was chosen for expression of BAF180. The pGAPZA vector has a 6x His-tag downstream of the gene. The stop codon already present in the BAF180 gene sequence was deleted so that a His-tag could be included on the recombinant protein, because His-tags can be used for affinity purification of a recombinant protein. Additionally, an enterokinase site (GAC GAC GAC GAC AAG) was included to cleave off the His-tag after purification, as tags can sometimes interfere with mechanistic studies (Figure 5.2).

Figure 5.2: Modifications on the 3' end of the BAF180 DNA sequence. Deletion of stop codon was done to allow expression of 6x His-tag on the C-terminus of BAF180. The His-tag would be cleaved off after purification using enterokinase. The 3' restriction enzyme was changed to KpnI to allow for directional cloning of BAF180 into the pGAPZA vector.

The PCR modifications (insertion of Kozak consensus sequence, insertion of spacer, alteration of 3' restriction enzyme site, and insertion of an enterokinase site) were accomplished using the following primers: (Color key: blue – GC clamp; red – EcoRI site; italics – Kozak sequence; black – spacer; purple – BAF180 gene; green – KpnI site; pink – enterokinase site; dark orange – pGAPZA vector):

Forward:
GCGCTGGTACGTAGGAATTCGAACTC*AAA*ATGGGTTCCAAGAGAAGAAG

Reverse:
GCGCTAATGATGTGAGGTACCCTTGTCGTCGTCGTCAACATTTTCTAGGTT
GTA

PCR conditions were as recommended by NEB for the Q5 Polymerase: 98 °C for 30 seconds, 30 cycles of 98 °C for 10 seconds, 76.3 °C for 30 seconds, 72 °C for 2 minutes followed by a final elongation time of 2 minutes at 72 °C. The PCR product was purified and concentrated before digesting. The modified BAF180 (i.e. PCR product) and the pGAPZA vector were digested with EcoRI and KpnI; 2000 ng of modified BAF180 and 2000 ng of pGAPZA vector were digested with 10 units (2 µL) of each enzyme. The digests were purified and pGAPZA and BAF180 were ligated together in a 5:1 ratio (insert:vector) using T4 ligase. The pGAPZA-BAF180 construct (Figure 5.3) was transformed into *E. coli* DH5α cells (for propagation) via electroporation. Electrocompetent *E. coli* DH5α cells were prepared according to the Bio-Rad MicroPulser manual; the electroporation procedure was from the manual as well.[31] It should be noted that pGAPZA contains the *sh ble* gene, conferring zeocin resistance. Therefore, zeocin was used for antibiotic selection. Because of this, low-salt LB at pH 7.5 had to be used for all procedures in *E. coli*, as zeocin is only active at low-salt concentrations and at that pH.[61] Transformants were grown on low-salt LB plates with 25 µg/mL zeocin overnight at 37 °C. The positive transformants were selected and cultured in low-salt LB with 25 µg/mL zeocin at 37 °C and 250 rpm. BAF180-pGAPZA plasmid was isolated from these cells using a miniprep kit.

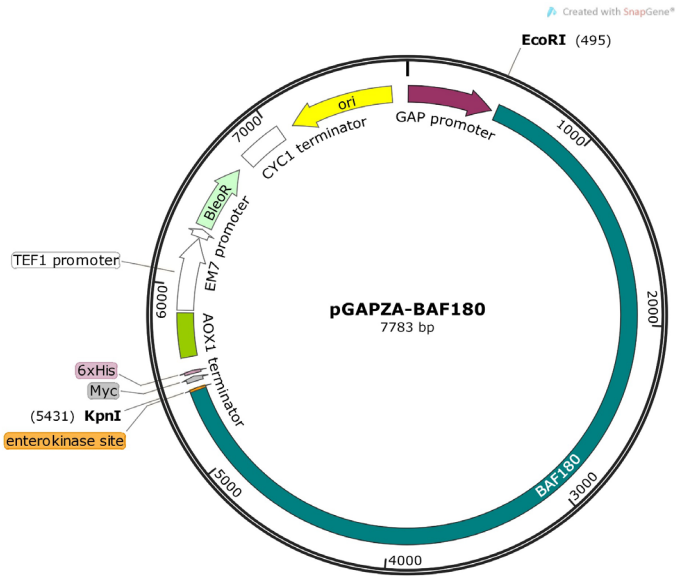


Figure 5.3: Plasmid map of the pGAPZA-BAF180 construct. BAF180 is highlighted in teal. The EM7 promoter allows for replication of the plasmid in *E. coli*. The *GAP* promoter and *AOX1* terminator are for use in yeast. The restriction enzyme sites EcoRI and KpnI are noted, as is the enterokinase site and 6x His-tag. Image created with SnapGene software.

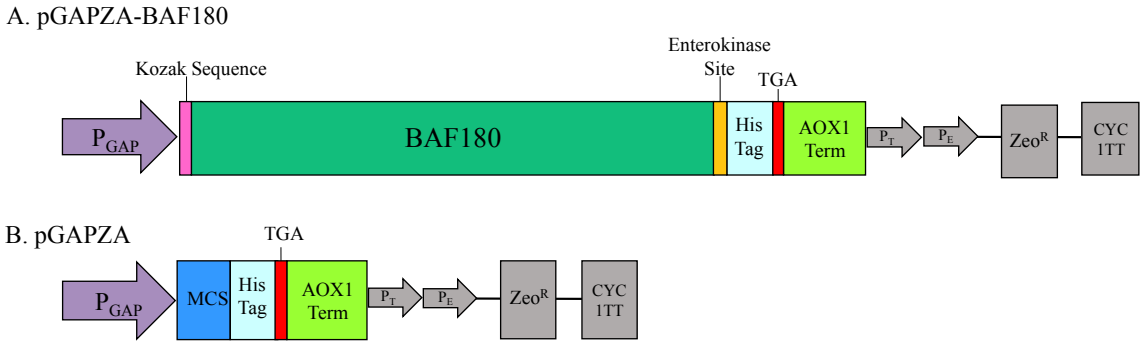


Figure 5.4: Linearized pGAPZA-BAF180 and control pGAPZA plasmids. (A) The linearized pGAPZA-BAF180 plasmid. pGAPZA-BAF180 was linearized with AvrII, which cuts within the GAP promoter. This linearization creates an expression cassette which is incorporated into the *P. pastoris* genome via homologous recombination. (B) The linearized pGAPZA vector. pGAPZA was linearized with AvrII and used as a positive control for transformation. Abbreviations: CYC1 TT: CYC1 transcription terminator region; GAP: constitutive GAP promoter; Zeo^R: *Sh ble* gene, conferring zeocin resistance; PT: TEF1 promoter, controls expression of the *Sh ble* gene in *P. pastoris*; P_E: EM7 promoter, controls expression of the *Sh ble* gene in *E. coli*; AOX1 Term: AOX1 transcription terminator region.

Before transformation into *P. pastoris* X33 cells, 4000 ng of pGAPZA-BAF180 was linearized with 20 units (1 µL = 5 units) of AvrII for 16 hours at 37 °C. AvrII cuts within the *GAP* promoter, thereby allowing for homologous recombination at that same

site in the *P. pastoris* genome (Figure 5.4A).[45,56,61] Homologous recombination results in incorporation of the linearized plasmid into the *P. pastoris* genome.[45,61] As a positive control, pGAPZA was linearized using the same restriction enzyme (Figure 5.4B) and transformed into *P. pastoris* cells as well. Electrocompetent X33 cells (80 µL) and 10 µg of linearized pGAPZA-BAF180 (or control pGAPZA) were combined in a microcentrifuge tube and subsequently transferred to a cold 0.2 cm electroporation cuvette. The cells were incubated on ice for 5 minutes.[61] A Bio-Rad MicroPulser (under the setting "Pic") was used to transform the cells.[63] Immediately after pulsing the cell/plasmid mixture, 1 mL of ice-cold 1 M sorbitol was added to the cuvette; sorbitol is used to stabilize the electroporated cells, because they can be sensitive to osmotic pressure.[61] It was very important that the sorbitol be added immediately, as transformation efficiency decreases exponentially with time delay.[63] The mixture of cells/plasmid/sorbitol was transferred to a 15 mL tube and incubated at 30 °C for 2 hours without shaking. Various amounts of the transformation mixture (5 µL, 10 µL, 15 µL, 25 µL, 50 µL, 75 µL, 100 µL, 125 µL, 150 µL, 175 µL and 200 µL) were spread on YPDS (YPD with sorbitol) plates with 100 µg/mL zeocin and grown for 3 days at 30 °C.[61] Colonies possessing the zeocin resistance gene grow better at lower densities, which is why various amounts of the transformation mixture was used.[61] The transformation of pGAPZA-BAF180 resulted in 6 very small colonies (~1 mm) and the transformation of pGAPZA resulted in 2 colonies. Consultation with another lab experienced in *P. pastoris* expression suggested that the recommended zeocin concentration (100 µg/mL) is too high and noted that it is not used in that lab. This lab uses 25 µg/mL, 50 µg/mL, and 75

μg/mL. Therefore, the low transformation efficiency may be a result of the high zeocin concentration.

The transformation products were purified to ensure that there were no mixed colony transformants, as mixed colony transformants have a tendency to lose the foreign DNA.[61] The six colonies were streaked on YPD plates with 25 μg/mL zeocin, and grown for approximately 42 hours. Each of these colonies were removed and re-streaked on plates with 100 μg/mL zeocin and grown for 3 days at 30 °C. Isolating single colonies was difficult as the cell density was very high. After consultation with the other lab, a new method for plating for single colonies was used, wherein a single colony was picked up and added to 200 μL YPD. A toothpick was dipped in the media and used to streak new plates (YPD with 100 μg/mL zeocin). These resulting single colonies were used to create freezer stock.

5.3 Preliminary Test for Protein Expression

An initial test for BAF180 expression was performed using cultures grown at 4 time-points. Zeocin selection was only used for plates; liquid YPD cultures did not have any zeocin added. A pGAPZA-BAF180 transformant (BAF180 X33) was grown for 3 days at 30 °C on a YPD plate with 100 μg/mL zeocin. A single colony was used to inoculate 10 mL of YPD and was grown overnight at 30 °C and 225 rpm. Four separate cultures were created from this initial stock, each culture having 2 mL of overnight culture added to 50 mL fresh YPD. Consequently, all the cultures stem from the same colony, thereby eliminating the possibility of expression differences between colonies. Each of the 4 cultures were grown at 20.5 °C and 225 rpm, and 1 culture was removed

every 24 hours. The temperature of 20.5 °C was chosen because it was hypothesized that this lower temperature would increase the probability of BAF180 production. These cells were harvested by centrifuging at 12,000 g for 15 minutes at 4 °C and lysed using the method presented by Johnson et al.[64] Harvested cells were re-suspended in lysis buffer (50 mM Tris-Cl, pH 7.4, 1 mM PMSF) at 2 mL/gram cell (PMSF is highly unstable in aqueous solutions and therefore needed to be added immediately prior to use). The cell suspension was then added to 1.5 mL microcentrifuge tubes along with 0.5 mm zirconia/silica disruption beads. The quantities used for each are shown in Table 5.1. Volumes of zirconia/silica beads were measured by displacement. Each tube was vortexed on maximum speed for 30 seconds, followed by a 30 second rest on ice; this cycle was repeated a total of 6 times. The samples were then centrifuged at 12,000 g for 15 minutes at 4 °C and the lysate collected. 8 M urea was added to the cell debris to recover any insoluble protein.

Table 5.1: Ratios of cell suspension and zirconia/silica disruption beads used in the initial screening for BAF180 production.

Time Point	Cell Suspension	Zirconia/Silica Beads
0 hour	~500 µL	~250 µL
24 hour	~500 µL	~250 µL
48 hour	~650 µL	~150 µL
72 hour	~680 µL	~320 µL
96 hour	~680 µL	~320 µL

All samples were visualized on SDS-PAGE to determine if BAF180 production occurred (Figure 5.5A & B). Additionally, the media was tested to confirm that BAF180 was not being secreted from the cell (Figure 5.5C). Results suggested that there was no BAF180 present in the soluble fraction (lysate), but a band around 190 kDa (the size of BAF180) was seen for the 96 hour insoluble fraction. Surprisingly, bands ~190 kDa were seen in the 72 hour media and 96 hour media samples. *P. pastoris* secretes very low levels of native proteins[43,45,65-67] and the pGAPZA vector does not have a secretion signal, therefore no protein should be seen in the media. It was suspected that these bands are present because some of the cells were being lysed during cell growth.[43]

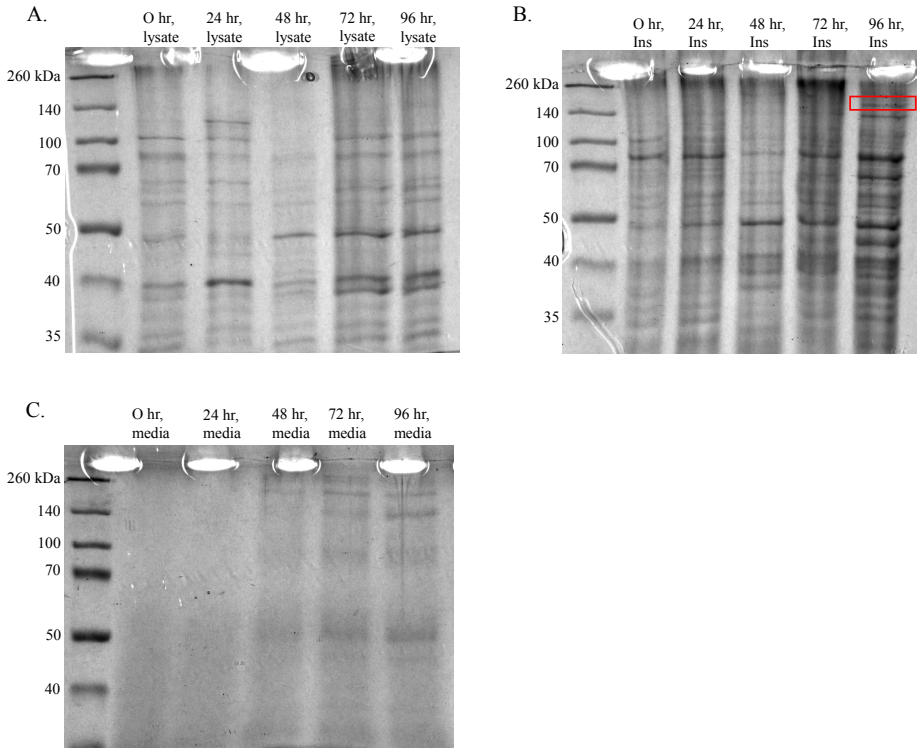


Figure 5.5: SDS-PAGE results from the initial screening for BAF180 production in *Pichia pastoris* X33 cells; BAF180 is approximately 190 kDa. (A) The soluble fraction (i.e. lysate); no band was seen at 190 kDa. (B) The insoluble fraction; a band about 190 kDa was seen for the 96 hour sample. (C) The media samples; bands at ~190 kDa were seen for the 72 hour and 96 hour media samples. The band for BAF180 is found in the red box.

When using yeast as a heterologous host, it is possible for multiple copies of the recombinant gene (i.e. BAF180) to be integrated into the genome. Though these are rare occurrences, the presence of multiple copies can increase recombinant protein yield.[45,56,60] Both BAF180 low-producing (likely 1 copy) and high-producing (likely more than 1 copy) clones were identified via SDS-PAGE of BAF180 X33 lysate (Figure 5.6). On YPD plates, WT X33 cells consistently formed larger colonies than high-producing BAF180 clones (Figure 5.7A) grown under the same conditions (30 °C, 72 hours) and WT X33 cells appeared slightly more white in color than high-producing BAF180 X33 cells (Figure 5.7B). Colonies of high-producing clones were noticeably "stickier" on solid media as well. In liquid media, there were no significant changes in the cellular masses of WT X33 cultures and BAF180 X33 cultures. However, high-producing BAF180 cell cultures were significantly more viscous than cultures of low-producing BAF180 cells or WT X33 cells. Additionally, low-producing BAF180 cells and WT X33 both compacted much better during centrifugation than high-producing cells did, so a higher centrifuge speed was required. High-producing BAF180 cells were also harder to re-suspend during the lysis procedure, which resulted in increasing the buffer:cell ratio (3mL/gram from 2 mL/gram).

High-producing cells were only discovered after visualizing the lysate on SDS-PAGE. Therefore, the high-producing cells needed to be obtained from the soluble lysate. 200 µL of soluble lysate was added to 3 mL of YPD and grown at 30 °C for 20-24 hours. Luckily, this culture grew and plates were made. Various volumes of the high-producing culture were added to YPD plates, as decreasing the cell density on plates can improve

growth.[61] The plates were grown at 30 °C for 3 days and then used to re-streak for single colonies. A single colony was used to inoculate a 5 mL YPD culture, which was grown overnight at 30 °C with 225 rpm. Freezer stock was obtained from this culture.

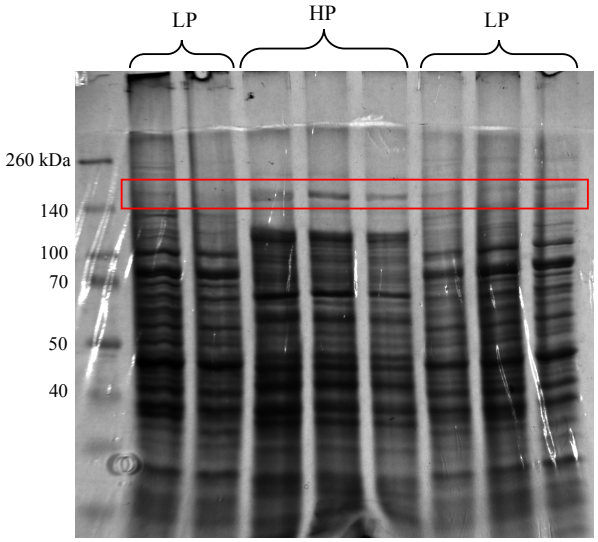


Figure 5.6: SDS-PAGE comparison of low-producing (LP) and high-producing (HP) clones of BAF180 X33 cells. The high-producing clones were identified by visual comparison between the soluble lysates of multiple clones. All cells were lysed using the same method. BAF180 (190 kDa) is within the red box.

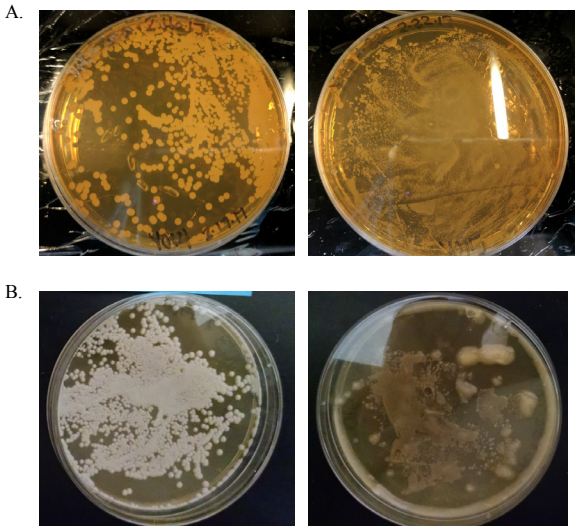


Figure 5.7: Direct visual comparison of colonies of WT X33 cells (left) and colonies of high-producer BAF180 X33 cells (right). (A) Differences in colony size between the two cell types. (B) Differences in color of the colonies.

5.4 Optimization of Growth Conditions

In order to obtain maximum expression levels of BAF180, growth conditions were optimized. This involved the manipulation of growing temperature, time, and growing media.[64,68-70]

5.4.1 Investigating the Role of Temperature and Time on BAF180 Production

The temperature used during heterologous cell growth has been demonstrated to change the yield and activity for several proteins.[64,68-70] Three temperatures were tested: 30 °C, 25 °C, and 20.5 °C to study the influence of temperature on BAF180 production. While yeast is normally grown at 30 °C, lowering the temperature can result in a higher quantity of "active" protein; this is especially important for proteins destined for mechanistic studies. As mentioned previously, the constitutive *GAP* promoter was chosen because it allows for continuous protein production during cell growth. Lowering the temperature slows cell growth which would subsequently result in a slower rate of BAF180 production. It was hypothesized that a slower rate of protein production would give the cell more time to properly fold the recombinant protein, because the quantity of protein present would be less and the protein would not overwhelm the cell's protein folding machinery. Therefore, a lower temperature might result in more functional BAF180 being produced.

The growing time is related to the cell density as it takes a certain amount of time for a culture to reach exponential growth, where it is growing the fastest. Because lowering the temperature slows down cell growth, the time that the cultures require to

reach exponential growth changes; therefore, temperature and time were tested in tandem. Since the *GAP* promoter was chosen (constitutive, not inducible), the exact point of exponential cell growth was not needed. The time to reach exponential cell growth is important with inducible promoters (such as *T7* and *AOX1*), as it is best to add the inducer (IPTG or methanol) at the beginning of this phase. To determine if the stage of cell growth, and thus cell density, affected the expression of BAF180 different growing times were tested: 24 hours, 48 hrs, 72 hrs, 96 hrs, and 120 hrs. It was hypothesized that a higher cell density would result in more BAF180. However, at higher cell densities the cells can burst and release proteases, which would degrade BAF180.

To test the effects of temperature and time on the production of BAF180, BAF180-containing X33 cells were streaked on YPD plates containing 100 μg/mL zeocin and grown at 30 °C for 3 days. A 10 mL YPD culture was inoculated with 1 colony and grown overnight at 30 °C and 225 rpm. YPD cultures (50 mL each) were inoculated with 2 mL of the overnight culture and grown at 30 °C for 24, 48, 72, 96, or 120 hours. Every 24 hours, a different culture was removed and spun down at 12,000 g for 15 minutes at 4 °C. The cells for each time point were lysed and the lysate visualized on SDS-PAGE (Figure 5.8). This procedure was repeated for tests at 20.5 °C and 25 °C; the overnight culture conditions remained the same. At least 6 replicates of each time and temperature combination were compared to determine changes in protein content. The results shown in Figure 5.8 are representative of the results obtained. Results repeatedly indicated that growing the cells at 25 °C was ideal (Figure 5.9); 24 and 48 hour time spans did not exhibit any significant difference, so 24 hours was chosen for brevity.

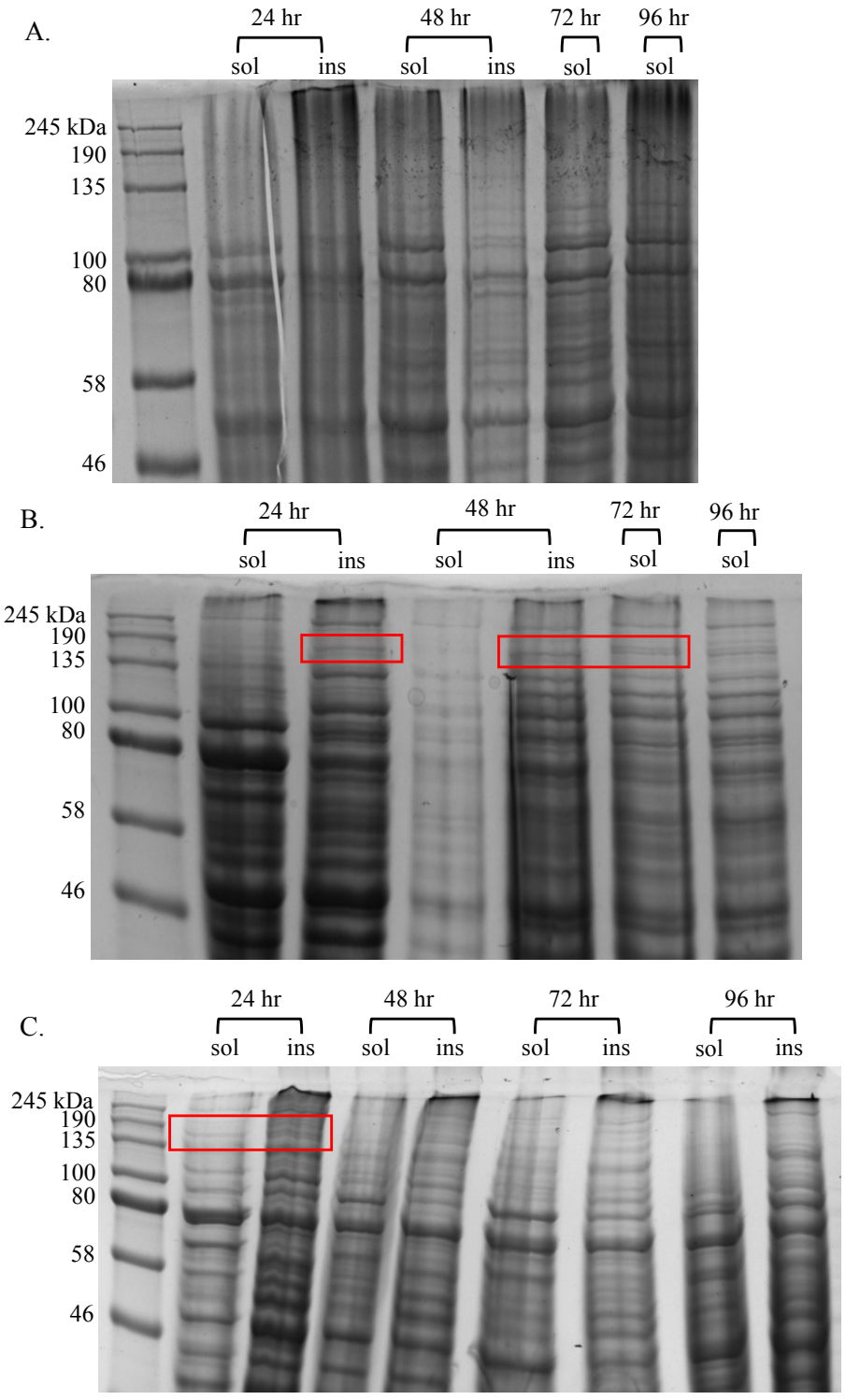


Figure 5.8: SDS-PAGE gels displaying the soluble (s) and insoluble (i) lysate fractions for BAF180 X33 cells. Cells were grown at different temperatures and times and the BAF180 content was compared. Results repeatedly demonstrated that 25 °C was optimal. (A) 20.5 °C; (B) 25 °C; (C) 30 °C. The band for BAF180 is found in the red box.

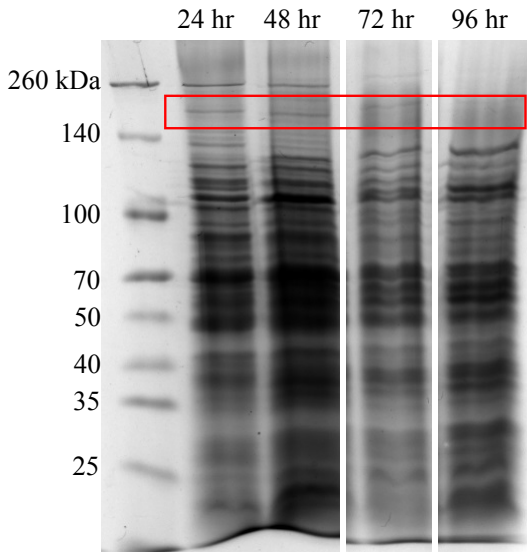
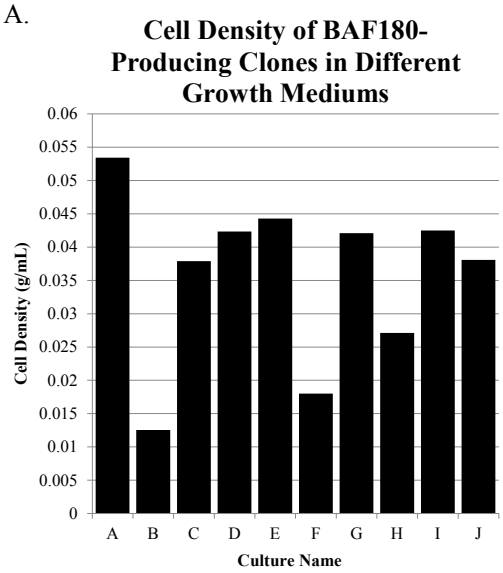


Figure 5.9: SDS-PAGE of lysates from BAF180 X33 cells grown at 25 °C. All lysates shown are the soluble fractions. Results repeatedly demonstrated that optimal growing time was 24 or 48 hours. There was no significant difference in BAF180 content between the 2 time spans; therefore, 24 hours was chosen for brevity. The band for BAF180 is found in the red box.

5.4.2 Influence of Media Composition

Experiments were also conducted to determine if pH or media composition would have a significant impact on cellular growth and BAF180 production.[64,68-70] Each 50 mL culture was started from 2 mL of the same 20 mL overnight culture (25 °C, 225 rpm, ~16 hrs). The 50 mL cultures were grown at 25 °C for 24 hours. YPD was used as the base media for all the tests; pH was varied (4 or 6.7) and various additives (glycerol, sorbitol, NH_4OH, dextrose) were included (Figure 5.10). Interestingly, the pH of the culture had the most significant impact on cell density; normal YPD had approximately 4x higher cell growth than YPD at pH 4. However, YPD+G results demonstrated the opposite effect; pH 4 grew better than pH 6. Results indicate that growing the cells in YPD results in the highest expression levels of BAF180 (Figure 5.11).

A.



KEY:

A. YPD (1% yeast extract, 2% peptone, 2% dextrose)
B. YPD, pH 4
C. YPD, added 5% NH_4OH at 5 hrs
D. YPD, fed-batch, 5 mL added 20% dextrose at 5 hrs
E. YPD+G, pH 4 (YPD + 1% glycerol)
F. YPD+G, pH 6
G. YPDS (YPD + 18.2% sorbitol)
H. YPD, 125 mL baffled shake flask
I. YPD
J. YPD

B.

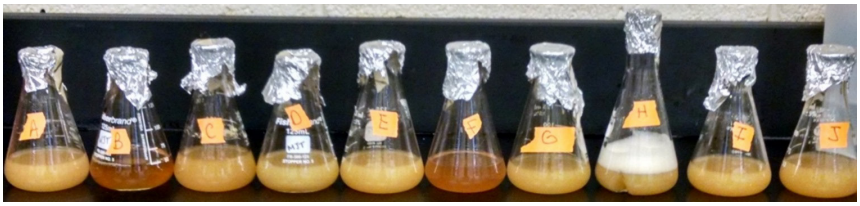


Figure 5.10: Comparison of growth mediums. All cells were grown at 25 °C for 24 hours and lysed in a 5 mL tube with a 1:1 ratio of lysate:beads. (A) Graphical representation of the cell densities observed for the different cultures; (B) Visual comparison of cultures. Growth mediums were: A: YPD (normal pH of 6.7); B: YPD, pH 4; C: YPD, 5% NH_4OH at 5 hours; D: YPD, 5 mL 20 % dextrose at 5 hours; E: YPD+G, pH 4 (YPD + 1 % glycerol); F: YPD+G (normal pH); G: YPDS (YPD + 18.2 % sorbitol); H: YPD in baffled flask; I and J: YPD.

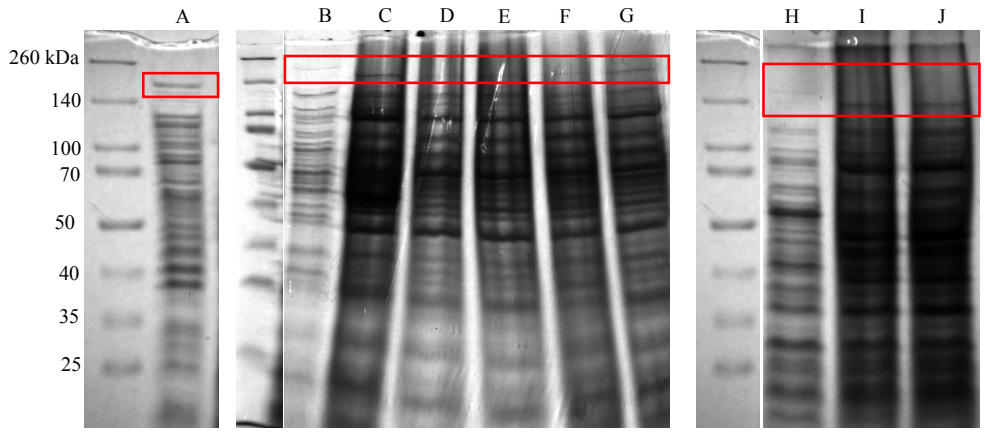


Figure 5.11: Optimization of media composition and pH. All gels were run using the same ladder as seen on the far left side. Some media compositions resulted in more or less BAF180 being produced, though the general protein composition of the cells did not change significantly. All cells were grown at 25 °C for 24 hours and lysed in a 5 mL tube with a 1:1 ratio of lysate:beads. Growth mediums were: A: YPD (normal pH of 6.7); B: YPD, pH 4; C: YPD, 5% NH_4OH at 5 hours; D: YPD, 5 mL 20 % dextrose at 5 hours; E: YPD+G, pH 4 (YPD + 1 % glycerol); F: YPD+G (normal pH); G: YPDS (YPD + 18.2 % sorbitol); H: YPD in baffled flask; I and J: YPD. The band for BAF180 is found in the red box.

5.4.3 Influence of Culture Volume on BAF180 Production

After preliminary tests were performed to determine optimal growth conditions, culture scale-up conditions were tested. Because yeast growth can be highly dependent on available oxygen,[71-77] it was unknown whether the increase in volume would impede growth and/or BAF180 production. Though the *GAP* promoter does not rely on oxygen as much as the *AOX1* promoter, tests were still performed. In tandem with culture scale-up experiments, different time spans were tested to determine if the larger culture volumes required a longer growing period. All initial tests were done with 50 mL cultures that had been inoculated with 2 mL of overnight culture. For scale-up experiments, the amount of overnight culture added to a new culture varied depending on the final culture volume. For culture volumes of 100 mL, 3.5 mL of overnight culture was used. Larger culture volumes (200 mL and 500 mL) were inoculated by taking aliquots from a 50 mL

24 hour culture; 5 mL was used to inoculate a 200 mL culture and 10 mL was used for 500 mL. Conditions tested are shown in Table 5.2. Each culture was harvested and lysed in the same manner and the soluble fractions were visualized on SDS-PAGE. As shown in Figure 5.12, the amount of BAF180 produced did not vary significantly with culture volume or time. It appeared that the quantity of oxygen did not affect BAF180 production, as increasing the volume did not significantly affect BAF180 content. Therefore, all subsequent culture volumes were increased to 500 mL so a larger volume of lysate could be obtained.

Once high-producing cells were discovered, a brief scale-up test was performed again. In this case, a 15 mL overnight culture was used. Culture volumes of 50 mL, 100 mL, and 500 mL were investigated; as before, the amount of overnight culture used to inoculate the fresh cultures varied depending on the volume. 50 mL cultures were inoculated with 2 mL of overnight, while 3.5 mL was used for the 100 mL culture and 5 mL of overnight was used for 500 mL culture. Again, scale-up did not significantly affect BAF180 production (Figure 5.13). Results indicate that ideal growing conditions for BAF180-producing cells are: 25 °C, 225 rpm, 24 hours, 500 mL YPD inoculated with 5 mL of overnight culture.

Table 5.2: The combinations of culture volume and growth time used during scale-up testing. Also displayed are the various volumes of overnight culture added to the larger culture.

Culture volume	Volume of overnight culture added	Growth time (hours)
50 mL	2 mL	24
100 mL	3.5 mL	24
100 mL	3.5 mL	48
200 mL	5 mL	24
500 mL	10 mL	24

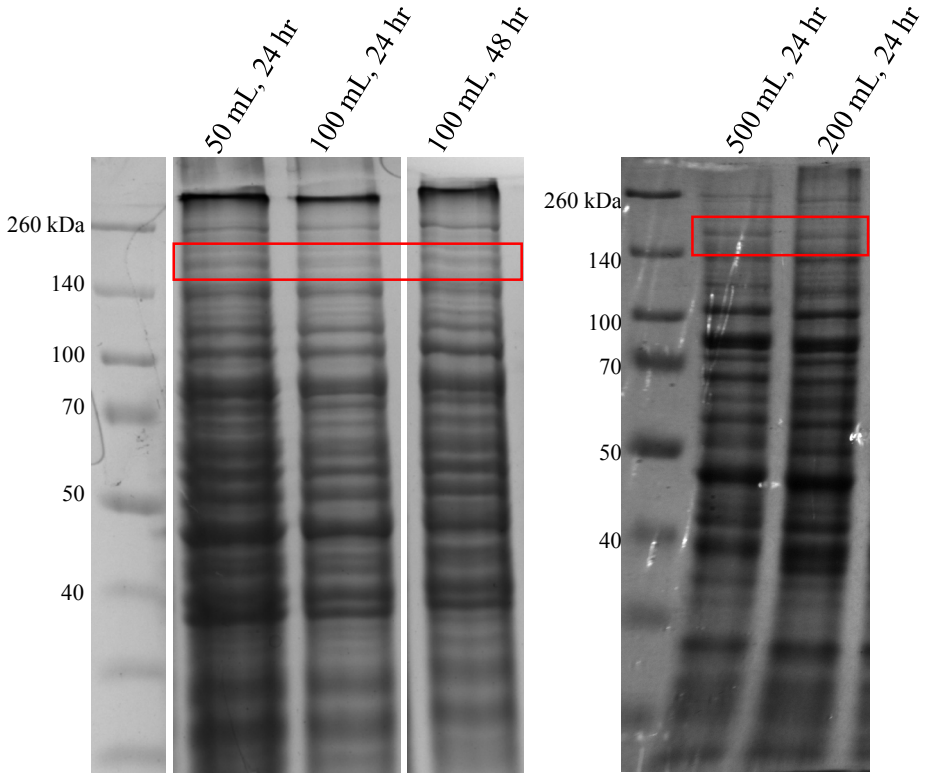


Figure 5.12: Culture scale-up testing. To determine how expression of BAF180 changes when the culture volume changes, multiple experiments were run and the lysates compared. Above are representative samples of the different combinations; all combinations were grown with low-producing cells at 25 °C and 225 rpm shaking. All fractions shown are the soluble fractions and all cells were lysed in the same manner. As can be seen, culture volume did not significantly affect BAF180 production, suggesting that scaling-up the growth of BAF180-producing cells was possible. The band for BAF180 (~190 kDa) is in the red box.

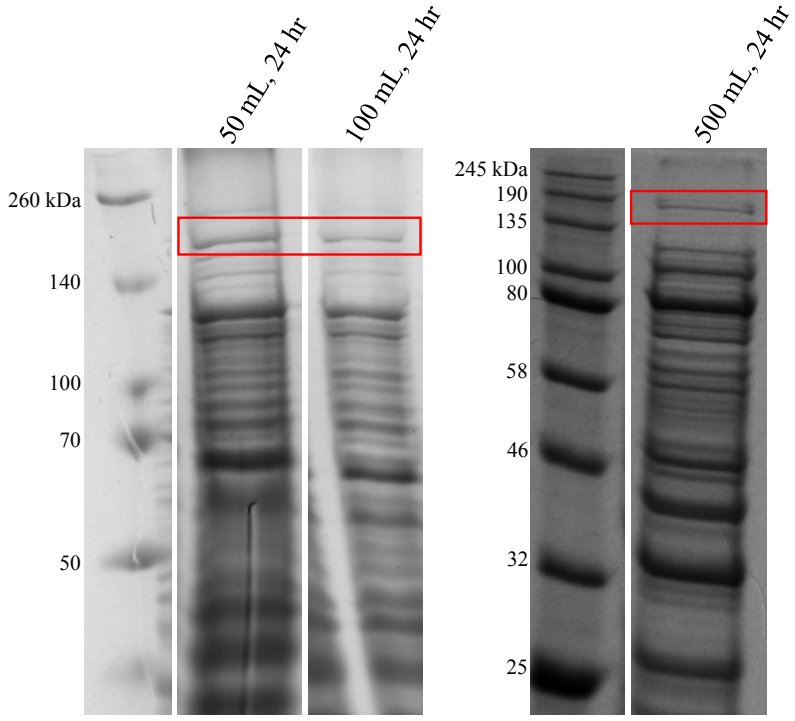


Figure 5.13: Scale-up testing of high-producer cells. Once the high-producing strains were discovered and isolated, scale-up conditions were tested again. There was no significant difference in BAF180 content between the different culture volumes. Cells were grown at 25 °C with 225 rpm shaking. The fractions shown are the soluble fractions and all cells were lysed in the same manner. As can be seen, culture volume did not significantly affect BAF180 production, suggesting that scaling-up the growth of BAF180-producing cells was possible. The band for BAF180 is found in the red box.

5.5 Optimization of Lysis Conditions

To ensure the highest recovery of BAF180, lysis conditions were manipulated and optimized. Lysis of *P. pastoris* cells involves harvesting the cells from the culture and re-suspending them in a lysis buffer. The cell suspension is combined with 0.5 mm beads wherein they are subjected to repeated mechanical agitation. Buffer composition, lysis mixture volume, agitation time, and the ratio of cell suspension to beads were investigated in an effort to maximize the recovery of BAF180.

The initial lysis conditions were based on the procedures outlined in the Invitrogen manual[61] and in publications from Dr. P. Murthy's lab.[64,68] These protocols involved harvesting the *P. pastoris* cells via centrifugation (12,000 g for 15 minutes at 4 °C), followed by wash and subsequent re-suspension in a lysis buffer at a ratio of 2 mL buffer per gram cell. A volume of cell suspension was combined in a tube along with zirconia/silica beads (0.5 mm) and the mixture of cells and beads were vortexed multiple times for a duration of 30 seconds, followed by 30 second rest on ice.[61,64,68] The lysate was recovered after further centrifugation (soluble fraction). Insoluble proteins were extracted by addition of 8 M urea to the cell debris (insoluble fraction).

Three lysis buffers were examined to determine if the buffer composition affected the lysis efficiency and BAF180 protein stability (Figure 5.14). The buffers were as follows: Buffer 1 – 50 mM Tris-Cl, pH 7.4 with 1 mM PMSF;[64,68] Buffer 2 – 50 mM NaH_2PO_4, pH 7.4; 1 mM PMSF; 1 mM EDTA; 5 % glycerol;[61] Buffer 3 omitted EDTA from buffer 2. Since BAF180 is a very large protein (190 kDa), attempting to refold it into its native state after denaturation during would be very difficult, therefore the goal was to determine which buffer composition resulted in the largest amount of BAF180 in the soluble fraction of the lysate. When using buffer 1, BAF180 was primarily found in the insoluble fraction. For buffer 2, BAF180 repeatedly appeared in the soluble fraction. Therefore, buffer 2 was chosen for all future experiments. It was hypothesized that the EDTA and glycerol found in buffer 2 had a stabilizing effect on the protein: glycerol slows down the kinetic energy of the solution and EDTA inhibits metalloproteases.

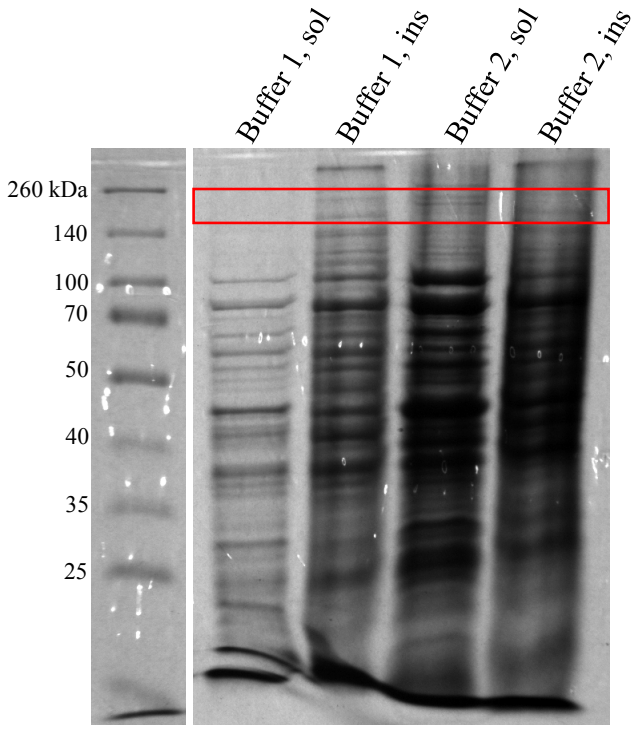


Figure 5.14: Comparison of the soluble (sol) and insoluble (ins) fractions for lysis with buffer 1 and with buffer 2. Though the bands are faint, the gel shows that most of the BAF180 (190 kDa) protein is found in the insoluble fraction (8 M urea) when using buffer 1, but found in the soluble fraction when using buffer 2. Since the BAF180 protein is so large, attempting to refold it into its native state after denaturation during lysis would be very difficult. Therefore, the buffer that had more BAF180 in the soluble fraction was chosen for all future work. Each fraction shown was lysed under the same conditions (1.5 mL tube, 1:1 ratio), except for the buffer composition. The lysis fractions shown here were from low-producing cells. The band for BAF180 is found in the red box.

Various volumes of lysis mixture (total volume of cell suspension and beads) were tested to determine if the total volume impacted recovery of BAF180. It was considered that perhaps the amount of dead space in the tube might affect how the cells and beads interacted (such as number of collisions). In a 1.5 mL tube, total volumes of 0.75 mL, 1 mL, and 1.25 mL were tested. In a 5 mL tube, total volumes of 4 mL and 4.5 mL were tested. There was no significant difference in the amount of BAF180 recovered when comparing lysis volumes (Figure 5.15). Lysis in 5 mL tubes was most time efficient, though recovery of the lysate was more difficult than for 1.5 mL tubes;

however, this trade-off was considered to be worth it because it reduced the overall time required for cell lysis. A volume of 4 mL lysis mixture in a 5 mL tube was chosen for future experiments.

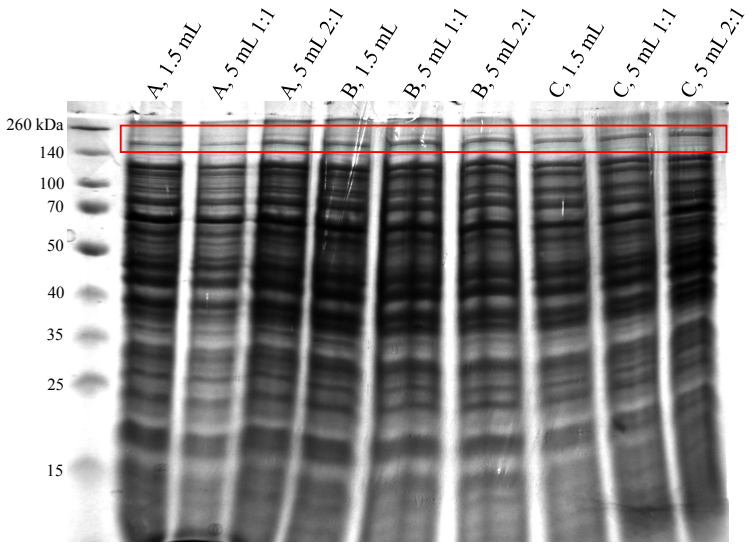


Figure 5.15: Effects of total volume and cell:bead ratio. Three cultures (A, B, C) were grown at 25 °C for 24 hours, all started from the same seed culture. All cultures were lysed in buffer 2 (50 mM NaH_2PO_4, 1 mM PMSF, 1 mM EDTA, 5 % glycerol, pH 7.4); cells were re-suspended at 2 mL/g cell pellet. Different volumes and different ratios of lysate:beads were tested. The differences in lysate content indicate that total volume and ratio of cells:beads did not a have a significant impact. All fractions shown are soluble fractions. BAF180 (~190 kDa) is the intense band between 260 and 140 kDa marker (in red box).

The ratio of cell suspension:beads was tested to determine if the ratio affected BAF180 recovery. It was hypothesized that there might be an optimal ratio to ensure the most favorable interactions between the cells and beads (such as number of collisions). Ratios of 1:1, 2:1, 3:1, and 4:1 were tested (Table 5.3), but no significant difference in BAF180 content was observed for each of the ratios used (Figure 5.15 and Figure 5.16). A ratio of 3:1 was chosen for future work because it was easiest to remove the lysate after centrifugation, meaning that it was easier to separate the supernatant from the beads at this ratio.

Table 5.3: Volumes used when testing different volumes of cells and beads, in both a 1.5 mL tube and a 5 mL tube. All ratios are cells:beads.

Ratio	1.5 mL Tube	5 mL Tube
1:1	500 µL:500 µL	2 mL: 2 mL
2:1	500 µL:250 µL	3 mL:1.5 mL
3:1		3 mL:1 mL
4:1	800 µL:200 µL 1 mL:250 µL	

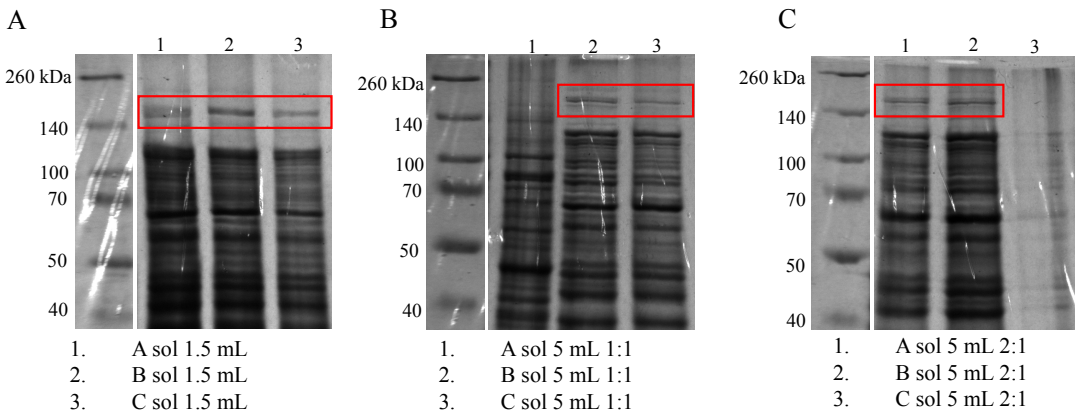


Figure 5.16: Lysis of high-producing cells. Three cultures were grown at 25 °C for 24 hours, all started from the same seed culture. All cultures were lysed in buffer 2 (50 mM NaH_2PO_4, 1 mM PMSF, 1 mM EDTA, 5 % glycerol, pH 7.4); cells were re-suspended at 2 mL/g cell pellet. Different volumes and different ratios of lysate:beads were tested. The differences in lysate content indicated that volume and ratio did not a have a significant impact on BAF180 recovery. These gel images also demonstrate the variation in BAF180 content between cultures, possibly due to inactivation of PMSF over time. The band for BAF180 is found in the red box.

The established methods for lysing yeast cells are enzymatic treatment and mechanical agitation. Mechanical agitation was chosen for use because it is more cost-efficient than enzymatic lysis. Zirconia/silica beads are added to the cell suspension and the mixture is subjected to vortexing for 30 seconds, followed by a 30 second rest on ice. The number of cycles was varied to determine how it affected BAF180 recovery. Repeats

of 6, 8, 10, 12, and 16 were tested (Figure 5.17). It was determined that 12 cycles consistently yielded the most BAF180 in the soluble fraction.

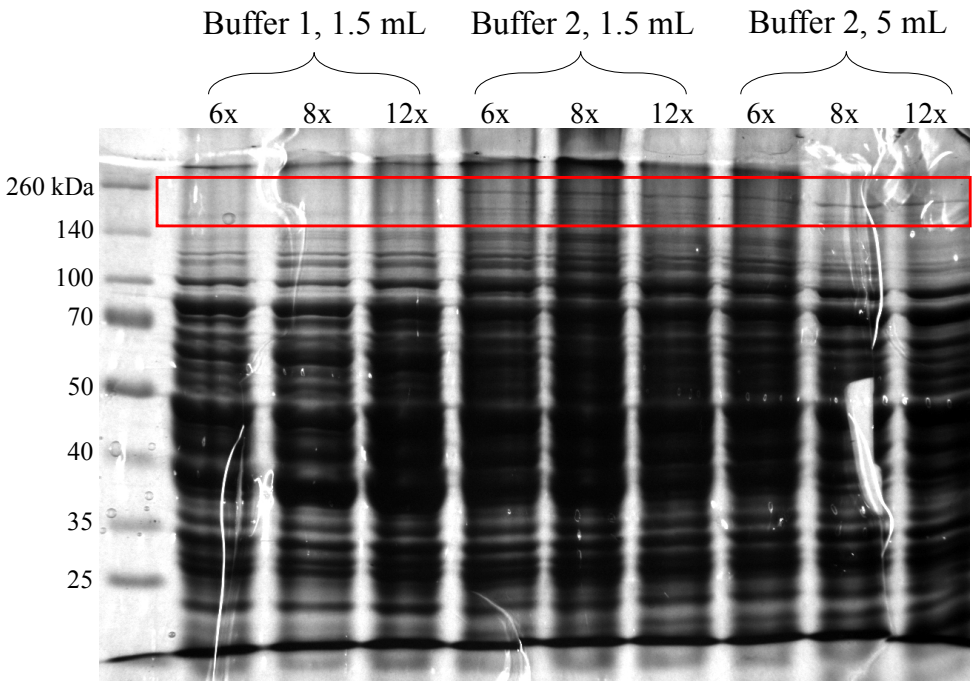


Figure 5.17: Experimental results comparing buffers 1 and 2, the total volume, and the number of cycles. All samples shown are soluble fractions and were lysed using a 1:1 ratio. Note that the total volume did not make a difference in lysate content. However, buffer type and number of cycles did. It can be seen that BAF180 (190 kDa) is more prevalent in the buffer 2 soluble lysate samples than in buffer 1. This is likely due to the presence of glycerol and EDTA in buffer 2. Additionally, the number of cycles did not significantly affect lysate content, though there was a small, but reproducible increase in BAF180 content for 12 cycles. The band for BAF180 is found in the red box.

Once the high-producing clones were discovered, volume and ratio experiments were performed again to verify that the results would not be different (Figure 5.16). High-producing cells were harvested by centrifuging at 17,000 g for 30 minutes at 4 °C. Both centrifugation speed and time were increased due to the increased culture viscosity of high-producing cells. After decanting the media, the cells were re-suspended in breaking buffer at a ratio of 3 mL/gram cell (this was increased from the previous 2 mL/gram cell because a larger volume was necessary to re-suspend the BAF180 high-

producing cells as intercellular adhesion was high, resulting in cells that were noticeably "stickier").

Additionally, the effect of the protease inhibitor composition was examined in relation to BAF180 recovery. It was thought that perhaps a more stable protease inhibitor would yield more consistent results. Use of PMSF gave inconsistent results, likely due to its instability and short half-life in aqueous solutions (30-60 minutes). The amount of time between starting and finishing lysis did vary and this likely impacted BAF180 recovery. Comparisons were done between PMSF and a protease inhibitor cocktail purchased from VWR. PMSF inhibits only serine proteases, while the protease inhibitor cocktail from VWR inhibits serine proteases (AEBSF, Aprotinin, and Leupeptin), cysteine proteases (E-64, Leupeptin), and aminopeptidases (Bestatin). Also, the lysis buffer contains EDTA, a metalloprotease inhibitor; EDTA was present for both tests. Larger yields of BAF180 (i.e. more intense bands on SDS-PAGE) were consistently observed when using the protease inhibitor cocktail (Figure 5.18). All subsequent cell lysis was performed using the cocktail from VWR instead of PMSF.

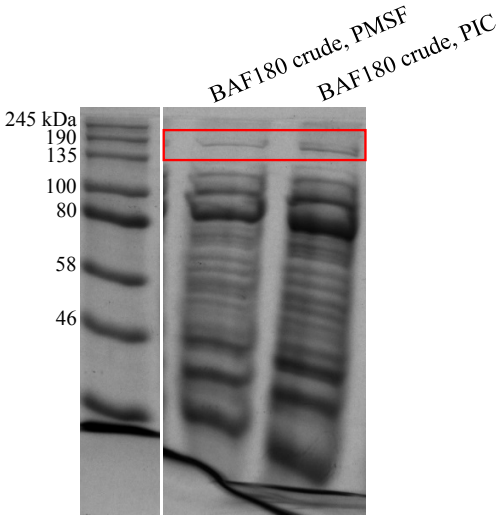


Figure 5.18: Choice of protease inhibitor type. Both PMSF and a protease inhibitor cocktail were used during lysis and the results were compared. A slightly more intense BAF180 band was consistently seen when using the protease inhibitor cocktail (PIC). The band for BAF180 is found in the red box.

After testing numerous combinations of the above variables, the optimal lysis procedure was established as follows: Cells were harvested by centrifuging at 17,000 g for 30 minutes at 4 °C. After decanting the media, the cells were washed by re-suspending them in 100 mL of buffer 2 containing 1 mM PMSF. The cells were spun down again, decanted, and re-suspended at a ratio of 3 mL/gram cell in buffer 2 with protease inhibitor cocktail (at the recommended concentration). The cell suspension was combined with 0.5 mm silica/zirconia beads at a ratio of 3:1 (3 mL cells/1 mL beads, total volume 4 mL). The cell/bead mixture was vortexed on maximum speed for 12 cycles (30 seconds on vortex/30 sec on ice). The lysate was centrifuged at 17,000 g for 30 minutes at 4 °C and the supernatant was removed (soluble fraction).

5.6 Confirmation of BAF180

To determine if the 190 kD band seen was an endogenous protein, wild-type X33 (transformed with pGAPZA for zeocin resistance) soluble and insoluble fractions were compared to the soluble fraction of BAF180 X33 cells (transformed with pGAPZA-BAF180) (Figure 5.19). Only the soluble lysate of BAF180 cells is shown, because this experiment was just to demonstrate that a band at the approximate molecular weight of BAF180 (~190 kDa) is not normally present in X33 cells. The 2 cultures, WT X33 and BAF180 X33, were treated in the same manner: growth was conducted for 96 hours at 20.5 °C and lysed using the same method. Comparison of lysates and insoluble fractions of WT X33 did not show a band around 190 kDa. When compared side by side, a band, approximately 190 kDa was seen only in BAF180 X33 cells (Figure 5.19). This experiment was also repeated several times after ideal growth conditions were established (25 °C for 24 hours).

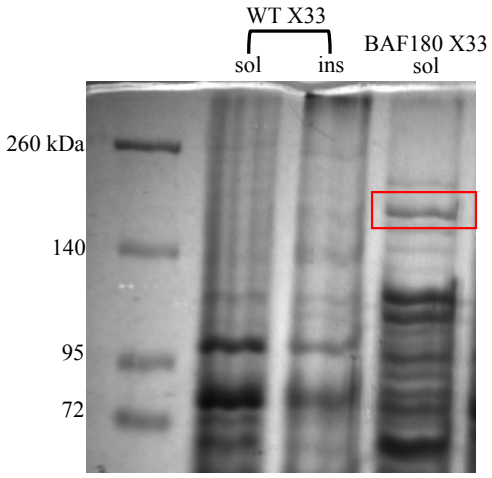


Figure 5.19: A comparison of the lysates WT X33 cells to the lysate of BAF180 X33 cells. SDS-PAGE gels (7.5%) showed a band around 190 kDa (expected molecular weight of BAF180) only in the BAF180 X33 fraction. Neither the soluble (sol) nor insoluble (ins) fractions of WT X33 cells displayed a band of the correct BAF180 size, so it was concluded that BAF180 is not normally expressed in wild type cells. The band for BAF180 is found in the red box.

Confirmation of BAF180 was also done using an anti-*PBRM1* antibody via western and dot blots. A dot blot is the same concept as a western blot, except the protein sample is spotted directly onto the membrane. For the dot blot, 2 μL of protein sample was spotted onto a 0.45 μm nitrocellulose membrane and allowed to dry; the membrane was hydrated in wash buffer for 5 minutes. A Pierce Fast Western ECL kit was used; kit instructions were followed with one exception: a 5 minute wash step was added between primary and secondary antibody incubations, to remove azide present in the primary antibody storage solution. Anti-*PBRM1* antibody was used at a ratio of 1:1000 (10 μL of stock antibody was added to 10 mL of Fast Western Antibody Diluent). After incubation in ECL (enhanced chemiluminescence) substrate, the membrane was imaged with Bio-Rad XRS+ imager (CDD camera). Crude lysates of BAF180 X33 cells and of WT X33 cells (both soluble and insoluble fractions) were spotted onto the membrane and tested against the antibody. The results (displayed in Figure 5.20) indicate that BAF180 is not present in WT X33 cells, as no spot was seen. This test served two purposes: one, to confirm that the protein being produced was BAF180 and two, that this protein is not endogenous to WT X33 cells.

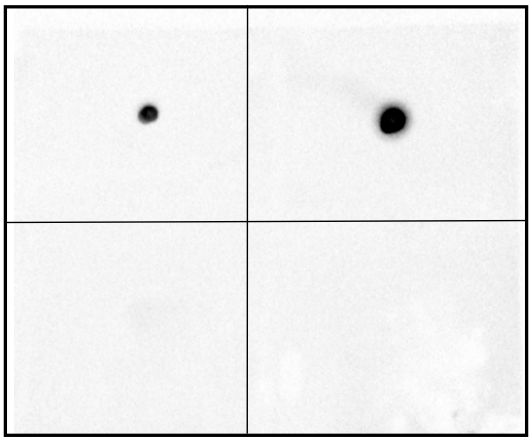


Figure 5.20: Confirmation of BAF180 production via dot blot. Crude lysates of BAF180 X33 and WT X33 cells were spotted onto a nitrocellulose membrane and allowed to dry. The membrane was then exposed to anti-*PBRM1* antibody and imaged using a CDD camera. As can be seen, signals were detected for the crude lysate of BAF180 X33 cells (boxes 1 and 2), but not for WT X33 cells (boxes 3 and 4). Both soluble and insoluble lysate fractions of each cell were used. Box 1: insoluble fraction BAF180 X33; box 2: soluble fraction BAF180 X33; box 3: insoluble fraction WT X33 cells; box 4: soluble fraction WT X33 cells.

Western blot experiments were performed using Tris-glycine and Tris-acetate SDS-PAGE gels. All attempts with the Tris-glycine system were unsuccessful. However, western blots were successfully carried out using pre-cast NuPAGE Tris-Acetate gels (3-8%). Tris-acetate gels were chosen because they are thought to be the best option when transferring large molecular weight proteins.[78] 1x NuPAGE Tris-acetate running buffer was prepared with NuPAGE antioxidant. NuPAGE antioxidant is a proprietary formula purchased from ThermoFisher which is added to the upper chamber of the gel assembly to ensure that the samples remain reduced; in the neutral pH of Tris-acetate gels, reducing agents do not travel with the sample and therefore the samples have the potential to be re-oxidized during electrophoresis.[79] 10 μg of sample (quantified via bicinchoninic acid assay) was prepared and loaded into each well. The gel was run for 1 hour at 150 V. After running, the gel was removed and incubated for 10 minutes in 100 mL of 2x NuPAGE transfer buffer with 0.02% SDS and NuPAGE antioxidant. 1x NuPAGE transfer buffer

was prepared with 0.01% SDS, 10% methanol, and NuPAGE antioxidant. 0.45 μm PVDF membrane was activated in 100% methanol for 30 seconds, rinsed with water, and incubated in 1x transfer buffer. Once the gel and PVDF membrane were prepared, the gel sandwich was prepared and wet transfer was performed at 70 V for 1 hour.[80] The PVDF membrane was air dried and stored overnight at -20 °C.[79] The western blot procedure was carried out the same as for dot blots. Exposure for 55 seconds with a CCD camera showed multiple bands being recognized (Figure 5.21). The most intense band appears at approximately 190 kDa, the expected molecular weight of BAF180. The other bands were a lower weight than this. There are a few possible reasons why multiple protein bands were recognized: degradation of the protein by proteases is possible, as is premature termination during DNA translation. The anti-*PBRM1* antibody used recognizes an epitope on the C-terminus of BAF180. For a band to be recognized by the antibody, the C-terminus of BAF180 must be present, eliminating premature termination as an option and suggesting that degradation occurs. Further investigations using other anti-*PBRM1* antibodies that recognize epitopes on the N-terminus and in the middle of the protein are needed before the hypothesis can be confirmed. Regardless, the concentration of the protease inhibitor cocktail will be increased for future experiments.

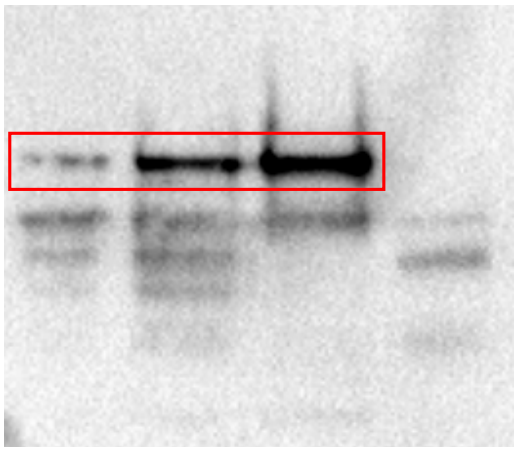


Figure 5.21: Western blot performed on BAF180 lysate samples. The most intense band (top) is thought to be full-length BAF180 (190 kDa). The reasons that bands of lower molecular weight are present are currently unknown. The antibody used recognized an epitope on the C-terminus of BAF180.

The recognition epitope of the antibody used was located at the C-terminus. Therefore, for a signal to be seen, this epitope must be present. Proteintech (the manufacturer of the antibody) used an immunogen 306 amino acids in length, encompassing the HMG box to the end of the protein. After researching the relationship between immunogens and epitopes, analysis was performed using SVMTriP online software[81] to determine the most probable epitope sequences (Figure 5.22).

A.

Proteintech Immunogen Sequence:

MGEEDSEVIEPPSLPQLQTPLASELDLMPYTPPQSTPKSAKGSAKKEGSKRKINMSGYILFSSE
MRAVIKAQHPDYSFGELSRLVGTEWRNLETAKKAEYEGMMGGYPPGLPPLQGPVDGLVSMG
SMQPLHPGGPPPHHLPPGVPGLPGIPPPGVMNQGVAPMVGTPAPGGSPYGQQVGVLGPPGQQ
APPPYPGPHPAGPPVIQQPTTPMFVAPPPKTQRLLHSEAYLKYIEGLSAESNSISKWDQTLAAR
RRDVHLSKEQESRLPSHWLKSKGAHTTMADALWRLRDLMLRDTLNIRQAYNLENV

Possible Epitopes: (SVMTriP)

Underlined is HMG box sequence

KYIEGLSAESNSISKWDQTL Score: 1.00 Position: 178-197

FGELSRLVGTEWRNLETAKK Score: 0.709 Position: 27-46

RKINMSGYILFSSEMRAVIK Score: 0.619 Position: -1-19

QPTTPMFVAPPPKTQRLLHS Score: 0.538 Position: 154-173

B.

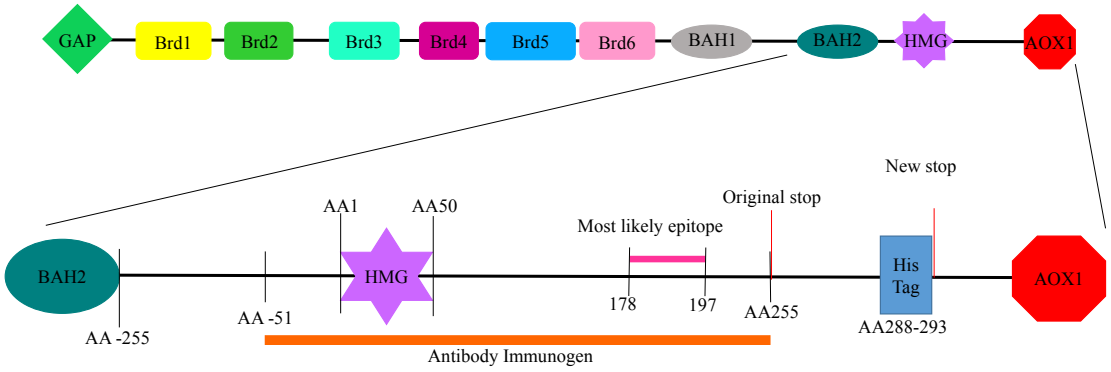


Figure 5.22: The complete immunogen sequence used by Proteintech to create the anti-*PBRM1* antibody. (A) The most likely antibody recognition epitopes are colored and the HMG box sequence is underlined. (B) The most probable recognition epitope is located 128 amino acids downstream of the HMG box. Numbering is relative to the HMG box, where the first amino acid of the HMG box is denoted as 1. There is no 0 position (-2, -1, 1, 2, etc.).

Additionally, PCR mapping of the domains was done to confirm the integration of the full *PBRM1* gene (i.e. BAF180) into the *P. pastoris* genome (Figure 5.23). Genomic DNA (gDNA), isolated from the yeast cells as described in the appendix, was used as a template. Each fragment was generated individually. The fragments cover most of the gene sequence. The fragments were staggered so that each fragment overlaps with at least one other to confirm the gene was integrated in the proper order. All PCR reactions had 500 ng of gDNA.

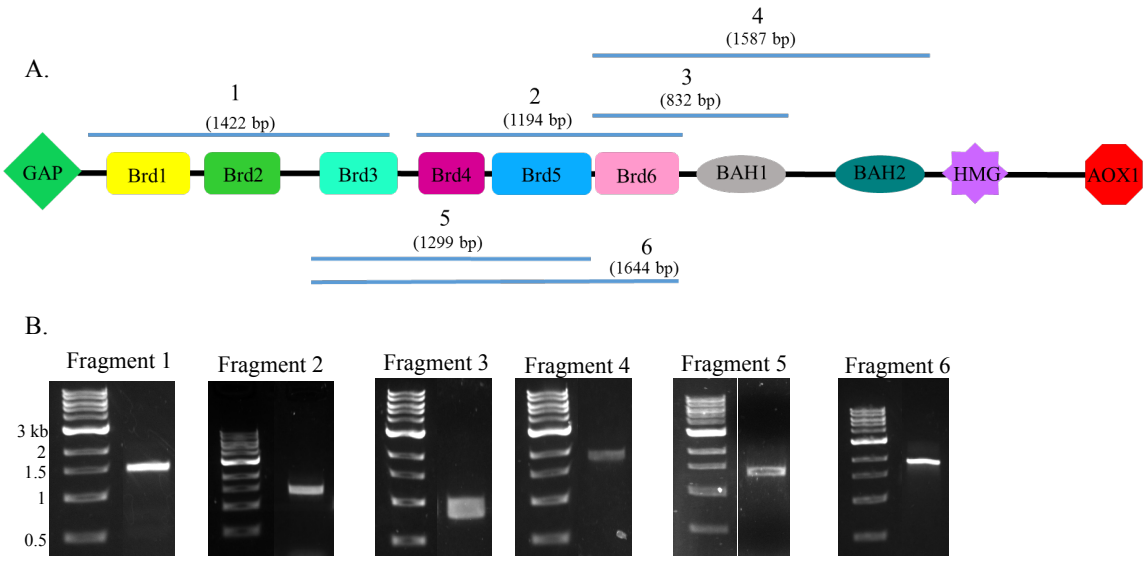


Figure 5.23: Genomic DNA was isolated from BAF180 X33 cells and domain-specific primers were used to confirm integration of the *PBRM1* sequence in *P. pastoris*. (A) Map of the *PBRM1* gene with fragments numbered and expected sizes (inclusive of primers). (B) PCR products for domain mapping.

5.7 Conclusion

For the first time, full-length BAF180 was successfully produced in a heterologous host. BAF180 was an enormously difficult protein to clone, likely owing to the size, presence of multiple domains, presence of many rare codons, frequent base tracks in the DNA, and instability. Regardless, after much effort, BAF180 was successfully produced in the yeast *Pichia pastoris*. To recover the most BAF180, growth and lysis conditions were optimized. Ideal growth conditions were determined to be 24 hours at 25 °C in 500 mL YPD. The optimal lysis procedure was established as follows: the cells were harvested by centrifuging at 17,000 g for 30 minutes at 4 °C. The media was decanted and the cells were washed by re-suspending them in 100 mL of buffer 2 containing 1 mM PMSF. The cells were spun down again, decanted, and re-suspended at a ratio of 3 mL/gram cell in lysis buffer 2 with protease inhibitor cocktail. The cell suspension was combined with 0.5 mm silica/zirconia beads at a ratio of 3:1 (3 mL

cells/1 mL beads, total volume 4 mL). The cell/bead mixture was vortexed on maximum speed for 12 cycles (30 seconds on vortex/30 sec on ice). The lysate was centrifuged at 17,000 g for 30 minutes at 4 °C and the supernatant was recovered (soluble fraction).

Following growth and lysis optimization, the presence of BAF180 was confirmed in multiple ways. Lysates of BAF180 X33 cells and WT X33 cells were compared to confirm that the 190 kD band seen on SDS-PAGE gels of BAF180 X33 lysate was not endogenous to X33 cells. The possibility of an endogenous X33 protein not appearing on the gel was considered and dot blot experiments were performed. After probing with an anti-*PBRM1* antibody, comparison of BAF180 X33 and WT X33 lysates (control) proved that BAF180 was not present in WT X33 cells. Western blot experiments were successfully carried using the Tris-acetate gel system. Probing the membrane with the anti-*PBRM1* antibody showed multiple protein bands being recognized. This is likely due to degradation of the protein; more protease inhibitor cocktail will be added in an attempt to minimize this. Integration of the *PBRM1* gene into the *Pichia pastoris* genome was confirmed using PCR with genomic DNA and domain-specific primers. Results supported the hypothesis that the *PBRM1* gene was successfully integrated into the host genome.

5.8 References

1. Xia, W.; Nagase, S.; Montia, A. G., et al., BAF180 Is a Critical Regulator of p21 Induction and a Tumor Suppressor Mutated in Breast Cancer. *Cancer Research* **2008,** *68* (6), 1667-1674.

2. Robbins-Pianka, A.; Rice, M. D.; Weir, M. P., The mRNA landscape at yeast translation initiation sites. *Bioinformatics* **2010,** *26* (21), 2651-2655.

3. Invitrogen. pGAPZ A, B, and C; pGAPZa A, B, and C. *Pichia* expression vectors for constitutive expression and purification of recombinant proteins 2010, p. 44.

4. Bio-Rad Inc., High Efficiency Electrotransformation of *E. coli*. In *MicroPulser Electroporation Apparatus Operating Instructions and Applications Guide*, Rev B ed.; p 12.

5. Lin-Cereghino, J.; Lin-Cereghino, G., Vectors and Strains for Expression. In *Pichia Protocols*, Cregg, J., Ed. Humana Press: 2007; Vol. 389, pp 11-25.

6. Puxbaum, V.; Mattanovich, D.; Gasser, B., Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. *Applied Microbiology and Biotechnology* **2015,** *99* (7), 2925-2938.

7. Bio-Rad Inc., Electroporation of *Pichia pastoris*. In *MicroPulser Electroporation Apparatus Operating Instructions and Applications Guide*, Rev B ed.; pp 19-20.

8. Johnson, S. C.; Yang, M.; Murthy, P. P. N., Heterologous expression and functional characterization of a plant alkaline phytase in Pichia pastoris. *Protein expression and purification* **2010,** *74* (2), 196-203.

9. Fickers, P., Pichia pastoris: a workhorse for recombinant protein production. *Current Research in Microbiology and Biotechnology* **2014,** *2* (3), 354-363.

10. Felber, M.; Pichler, H.; Ruth, C., Strains and Molecular Tools for Recombinant Protein Production in Pichia pastoris. In *Yeast Metabolic Engineering: Methods and Protocols*, Mapelli, V., Ed. Springer New York: New York, NY, 2014; pp 87-111.

11. Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B., et al., Heterologous protein production using the Pichia pastoris expression system. *Yeast (Chichester, England)* **2005,** *22* (4), 249-70.

12. Cereghino, J. L.; Cregg, J. M., Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiology Reviews* **2000,** *24* (1), 45-66.

13. Aw, R.; Polizzi, K. M., Can too many copies spoil the broth? *Microbial Cell Factories* **2013,** *12* (1), 128.

14. Yang, M.; Johnson, S. C.; Murthy, P. P. N., Enhancement of alkaline phytase production in Pichia pastoris: Influence of gene dosage, sequence optimization and expression temperature. *Protein expression and purification* **2012,** *84* (2), 247-254.

15. Mao, R.; Teng, D.; Wang, X., et al., Optimization of expression conditions for a novel NZ2114-derived antimicrobial peptide-MP1102 under the control of the GAP promoter in Pichia pastoris X-33. *BMC Microbiology* **2015,** *15*, 57.

16. Guan, B.; Chen, F.; Lei, J., et al., Constitutive Expression of a rhIL-2-HSA Fusion Protein in Pichia pastoris Using Glucose as Carbon Source. *Applied Biochemistry and Biotechnology* **2013,** *171* (7), 1792-1804.

17. Baumann, K.; Carnicer, M.; Dragosits, M., et al., A multi-level study of recombinant Pichia pastoris in different oxygen conditions. *BMC Systems Biology* **2010,** *4* (1), 141.

18. Looser, V.; Bruhlmann, B.; Bumbak, F., et al., Cultivation strategies to enhance productivity of Pichia pastoris: A review. *Biotechnology Advances* **2015,** *33* (6, Part 2), 1177-1193.

19. Rebnegger, C.; Vos, T.; Graf, A. B., et al., Pichia pastoris Exhibits High Viability and a Low Maintenance Energy Requirement at Near-Zero Specific Growth Rates. *Applied and Environmental Microbiology* **2016,** *82* (15), 4570-4583.

20. Çalık, P.; İnankur, B.; Soyaslan, E. Ş., et al., Fermentation and oxygen transfer characteristics in recombinant human growth hormone production by Pichia pastoris in sorbitol batch and methanol fed-batch operation. *Journal of Chemical Technology & Biotechnology* **2010,** *85* (2), 226-233.

21. Cos, O.; Ramón, R.; Montesinos, J. L., et al., Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: A review. *Microbial Cell Factories* **2006,** *5*, 17-17.

22. Vassileva, A.; Chugh, D. A.; Swaminathan, S., et al., Expression of hepatitis B surface antigen in the methylotrophic yeast Pichia pastoris using the GAP promoter. *Journal of Biotechnology* **2001,** *88* (1), 21-35.

23. Güneş, H.; Çalık, P., Oxygen transfer as a tool for fine-tuning recombinant protein production by Pichia pastoris under glyceraldehyde-3-phosphate dehydrogenase promoter. *Bioprocess and Biosystems Engineering* **2016,** *39* (7), 1061-1072.

24. Brown, V. Five Fast Tips for Blotting Large Proteins. http://bitesizebio.com/24228/five-fast-tips-for-blotting-large-proteins/ (accessed 3 April 2017).

25. Invitrogen NuPAGE FAQs. https://www.thermofisher.com/search/results?query=EA0375BOX&persona=DocSupport&type=Product+FAQs (accessed 5 April 2017).

26. Invitrogen, NuPAGE® Technical Guide. 2010.

27. Yao, B.; Zhang, L.; Liang, S., et al., SVMTriP: A Method to Predict Antigenic Epitopes Using Support Vector Machine to Integrate Tri-Peptide Similarity and Propensity. *PLOS ONE* **2012,** *7* (9), e45152.

Chapter 6: Purification of BAF180 in *Pichia pastoris*

Purification of recombinant BAF180 was the next step after cloning and optimization of the expression and lysis conditions. As stated in a previous chapter, BAF180 has not been produced in a heterologous host and therefore has never been purified from a heterologous host. Researchers utilizing mammalian cells generally "purify" BAF180 by pull-down assay; this yields only small amounts of BAF180 and as such is unsuitable for larger studies. The properties of BAF180 were studied before attempting purification. Multiple purification methods were tried including IMAC (His-tag affinity), strong and weak anion exchange (AEX), CHT hydroxyapatite (multi-mode), hydrophobic interaction chromatography (HIC), size-exclusion (SEC), ultrafiltration, and dialysis. Only strong AEX was able to purify BAF180 from the crude lysate.

6.1 Materials

SDS, glycine, Tris hydrochloride, Tris base, and glycerol purchased from Amresco (VWR: Radnor, PA). NaCl was purchased from EMD Millipore (Billerica, MA). EDTA was purchased from Mallinckrodt Baker Inc. (Phillipsburg, NJ). Dialysis membranes were purchased from Spectrum Labs (Rancho Dominguez, California): Spectra/Por® 1 (6-8 kD MWCO membrane; part 132660), and Spectra/Por 6® (50 kD MWCO membrane; part 132544). Macro-Prep High Q® Support (cat. 1580040), and Econo® glass columns were purchased from Bio-Rad, Inc. (Hercules, CA). Sodium phosphate monobasic (NaH_2PO_4) and Whatman Puradisc 25 syringe filters – 1.0 µm PES membrane (cat. WHA67812510) were purchased from Sigma Aldrich (St. Louis, MO). Blotting paper (cat. 8860), 0.45 µm PVDF membrane (cat. 88585), NuPAGE Tris-

Acetate gels 3-8% (cat. EA0375), NuPAGE LDS Sample Buffer 4x (cat. NP0007), NuPAGE Tris-acetate running buffer (cat. LA0041), NuPAGE transfer buffer (cat. NP0006), NuPAGE antioxidant (cat. NP0005), NuPAGE reducing agent (cat. NP0004) and Pierce Fast Western Blot Kit, ECL substrate (cat. 35050) were purchased from ThermoFisher (Waltham, MA). Anti-*PBRM1* antibody (cat. 12563-1-AP, rabbit polyclonal) was received as part of the antibody exchange program at Proteintech (Rosemont, IL). Western blot transfer was performed using a Hoefer TE22 transfer tank (Holliston, MA).

6.2 Purification of BAF180 by Anion Exchange Chromatography

Purification of a recombinant protein by ion exchange chromatography uses the pI of the protein (i.e. its charge at a certain pH) to separate it from other cellular proteins. Above the pI, the protein is negatively-charged and can be separated by anion exchange resins; below the pI, the protein is positively-charged and can be separated by cation exchange resins.[82-83] The pI of recombinant BAF180 is calculated to be 6.29. Anion exchange chromatography (AEX) was chosen because of the known stability of BAF180 at pH 7.4, in lysis buffer 2 (see Section 5.5); since this is above the pI, BAF180 would carry a net negative charge. A strong anion exchanger was chosen, because the effectiveness of the resin is not affected by the pH of the buffer system, as weak anion exchangers are.[82,84-85] MacroPrep® High Q was chosen as the strong AEX resin because it is noted as being good for separating large biomolecules and acidic and neutral proteins.[82,86-89] MacroPrep® High Q is a methacrylate polymer with a quaternary ammonium group attached.[88,90]

A well designed buffer system is critical for anion exchange to be successful. When designing a buffer system for AEX, some general rules must be considered:

- The pH of the buffer system should be 0.5-1.5 pH units above the pI of the protein.[83,85,91] Therefore, the pH of the buffer system needed to be between 6.8 and 7.8.
- The pKa of the buffer system should be within 0.5 pH units of the buffer's pH.[91]
- The protein should be stable in the buffer system. Several different buffer systems are possible, depending on the pH needed[83,85,87-88,91] but Tris was chosen because previous experiments demonstrated that BAF180 was stable in it. Tris buffers have an approximate buffering range pH 7.5-8.[83,85,87-88]
- An effective buffer system should be greater than 10 mM concentration.[83,91-92]

Calculating pKa values of various concentrations of Tris at room temperature led to the decision to use 17 mM Tris at pH 7.6.[92] A pH of 7.6 is 1.3 pH units above the pI of BAF180 (6.29). Elution off an anion exchange column is accomplished by increasing the salt (NaCl) concentration. Buffers with 0 M, 0.1 M, 0.5 M, and 1.0 M NaCl were recommended as starting points to determine the salt concentration at which BAF180 elutes.

Samples were prepared for purification by dialyzing crude lysate overnight in 17 mM Tris, pH 7.6. Samples were collected the next day and stored at -20 °C until needed. Samples were defrosted at room temperature and centrifuged at 17,000 g for 20 minutes

at 4 °C to reduce viscosity and remove unwanted particulates. The supernatant was collected and run through a 1.0 μm PES syringe filter before loading onto a column. Preliminary results indicated that BAF180 was eluting from the resin at 0.5 M NaCl, with only a few contaminant proteins (Figure 6.1). More precise buffers (0.1 M, 0.2 M, 0.3 M, 0.4 M, 0.5 M NaCl) were used to narrow down this range. BAF180 co-eluted with two contaminant proteins (~80 kDa and ~70 kDa) at 0.4 M NaCl (Figure 6.2). A buffer with 0.35 M NaCl was used in an attempt to separate these bands, but this did not show any significant difference (Figure 6.3). More buffers were made at concentrations between 0.35 M and 0.4 M NaCl (0.36 M, 0.37 M, 0.38 M, 0.39 M). The results indicated that BAF180 was eluting at 0.36 M, 0.37 M, 0.38 M, 0.39 M, and 0.40 M; the contaminants remained and did not separate from BAF180 (Figure 6.4). Though BAF180 eluted at 0.36 M NaCl and up, 0.4 M NaCl was chosen as the elution concentration in order to maximize recovery and reduce the dilution factor.

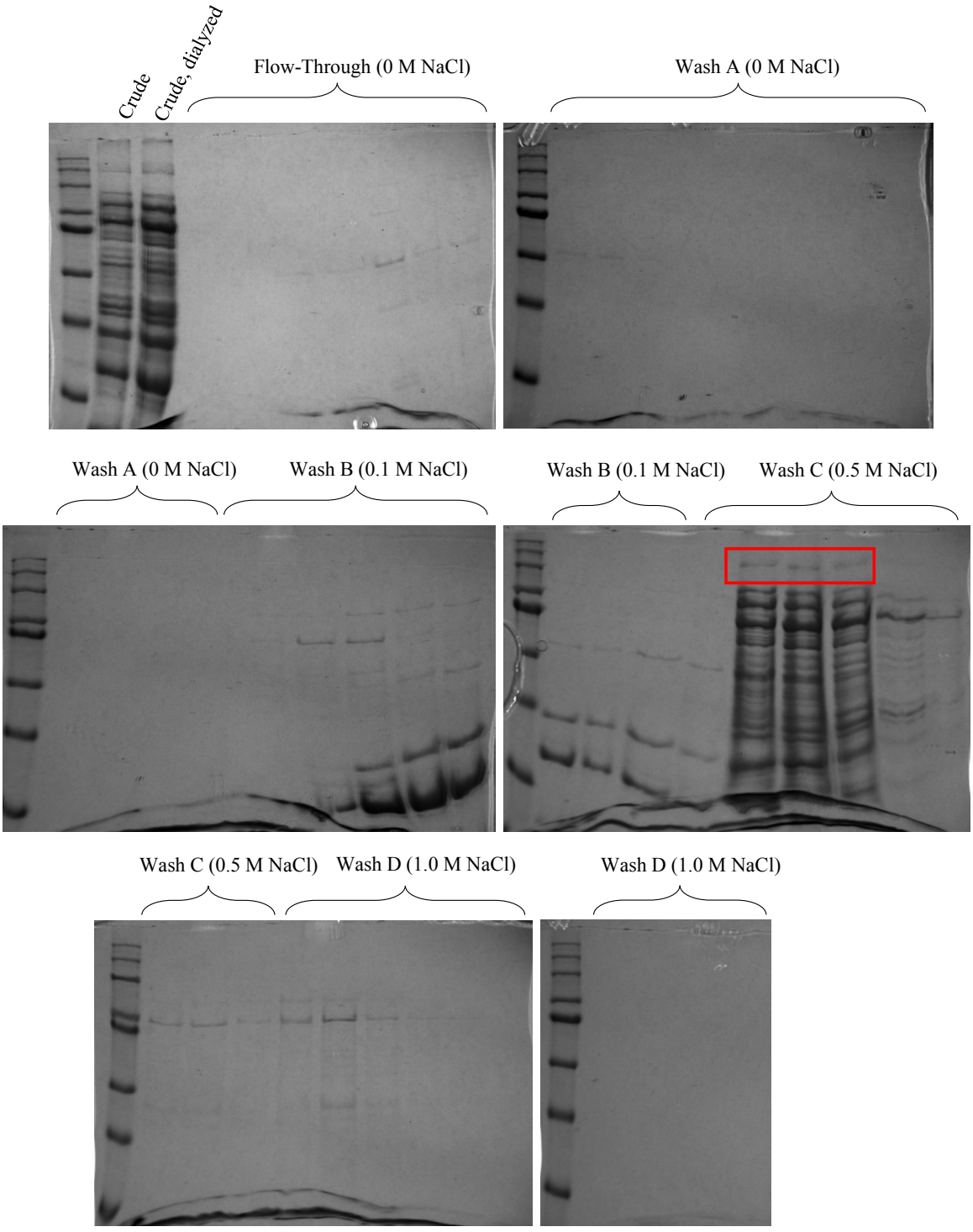


Figure 6.1: First attempt at anion exchange purification. Buffers of 0 M, 0.1 M, 0.5 M, and 1.0 M NaCl were used to elute BAF180 from MacroPrep® High Q resin. This preliminary test indicated that BAF180 was eluting at 0.5 M NaCl. The band for BAF180 is found in the red box.

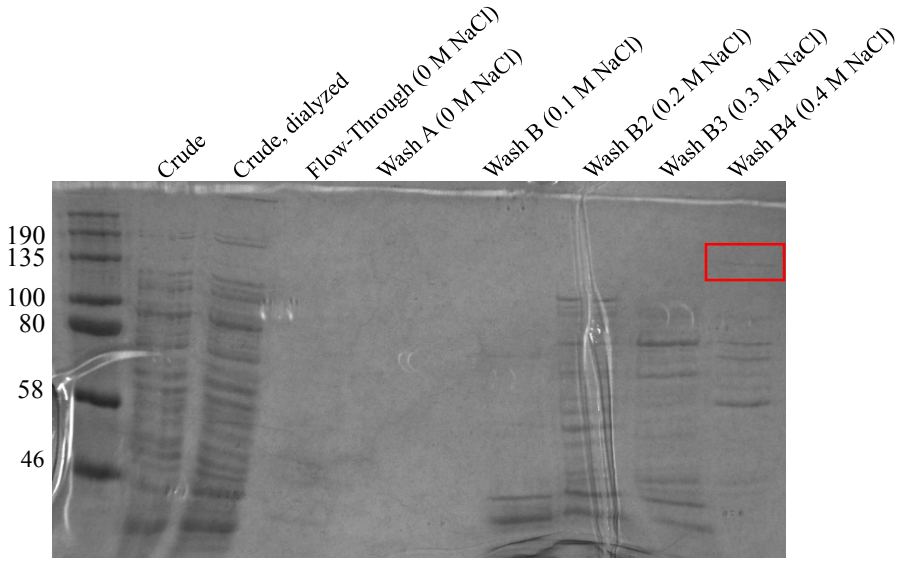


Figure 6.2: Anion exchange purification in 0.1 M NaCl increments. More precise buffers were used to determine the exact NaCl concentration required to elute BAF180 from the anion exchange resin; elution occurred at 0.4 M NaCl. The band for BAF180 is found in the red box.

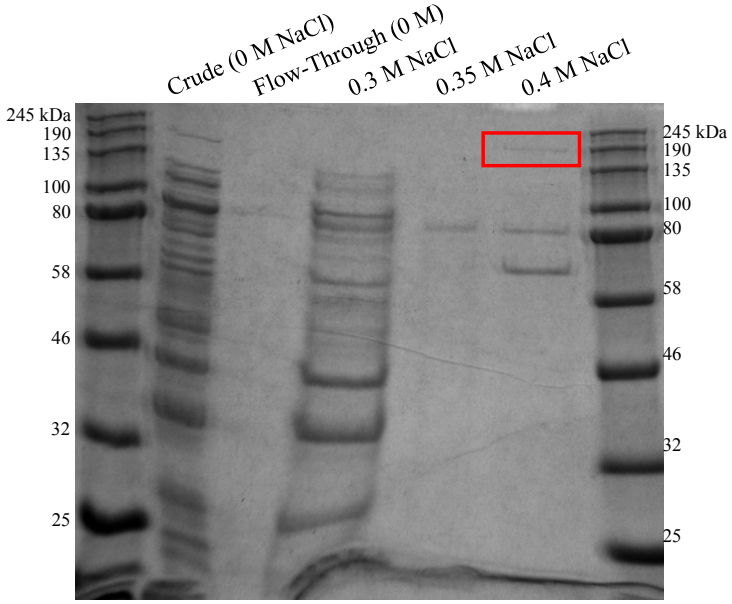


Figure 6.3: Anion exchange purification with Tris buffers with 0.3 M, 0.35 M, and 0.4 M NaCl. Buffers with 0.3 M, 0.35 M, and 0.4 M NaCl were used to zone in on the exact NaCl concentration required to elute BAF180 from the contaminant proteins that remained after washing the column with 0.3 M NaCl. Results indicated that BAF180 eluted after 0.35 M NaCl along with a few other contaminant proteins. The band for BAF180 is found in the red box.

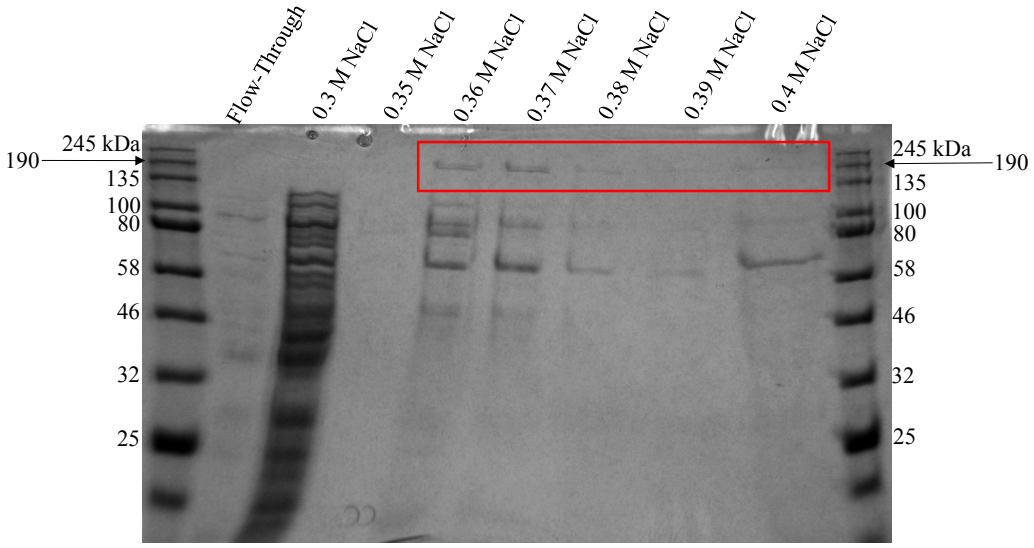


Figure 6.4: Anion exchange elution with 0.01 M increments of NaCl. Though BAF180 co-eluted at 0.36 M NaCl and up, this also decreased the final concentration. To maximize recovery, elution at 0.4 M NaCl was chosen. The BAF180 band is found in the red box.

Full-length BAF180 always eluted at 0.4 M NaCl, however the sizes of the contaminant bands sometimes changed. It was thought that these contaminant bands were other proteins. However, a recent western blot reveals (see Section 5.6) that full-length BAF180 is being produced but much of it is degraded (Figure 6.5). The antibody used recognizes an epitope on the C-terminus (see Section 5.6); therefore this epitope must be present in all the recognized bands. Premature termination during translation was eliminated as a cause because this would leave fragments containing the N-terminus but not the C-terminus, where the epitope is located. The use of different antibodies which recognize epitopes on the N-terminus or in the middle of the protein could be used to confirm that BAF180 is being degraded and that the other bands seen are fragments of it (see Appendix A). Proteolytic degradation could explain why BAF180 content varied

after lysis and why purification results were sometimes inconsistent. Increasing the protease inhibitor cocktail concentration above the recommended level might result in less proteolytic BAF180 degradation.

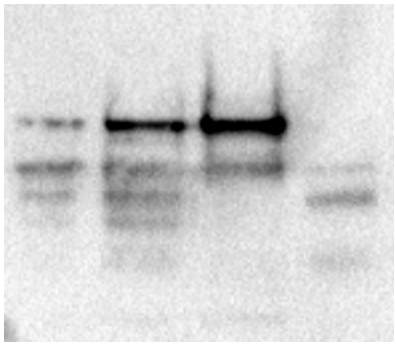


Figure 6.5: Western blot performed on BAF180 lysate samples. The most intense band is thought to be full-length BAF180 (190 kDa). The antibody used recognized an epitope on the C-terminus of BAF180. Lanes 1 and 3 were crude samples. Lanes 2 and 4 were purified fractions. It is believed that the lower molecular weight bands are the result of BAF180 degradation by proteases. Exposure was 55 seconds with a CDD camera and ECL detection.

It is believed that during lysis the BAF180 protein is being degraded by proteases, even though protease inhibitors are added to the lysis mixture. Proteolytic degradation in yeast mostly occurs in vacuoles which are membrane bound,[56,58] therefore, the protein is only exposed to them during lysis. Increasing the protease inhibitor cocktail concentration or changing cocktail type altogether might reduce proteolytic degradation and result in a higher BAF180 yield.

6.3 Conclusion

For the first time, BAF180 was successfully purified from a heterologous host using strong anion exchange chromatography. Recent western blot results revealed that the bands seen on SDS-PAGE gels that were initially thought to be contaminants are actually BAF180 fragments (likely a result of degradation). The concentration and/or

composition of the protease inhibitor cocktail should be manipulated in order to eliminate as much protease activity as possible and increase the yield of BAF180.

6.4 Supplementary Information

The following experiments were run before it was discovered that the other bands seen in the AEX purified fraction were actually fragments of BAF180. The "contaminant" proteins, now known to be fragments of BAF180, eluted with full length BAF180 every time. It was thought that perhaps a protein complex was forming in the cell or that the "contaminant" proteins had a pI too close to that of BAF180 to be separated.

Materials

SDS, glycine, Tris hydrochloride, Tris base, glycerol, guanidine hydrochloride, and imidazole were purchased from Amresco (VWR: Radnor, PA). NaCl and Amicon® Ultra-15 100 kDa Centrifugal Filter Units (cat. UFC910024) were purchased from EMD Millipore (Billerica, MA). EDTA was purchased from Mallinckrodt Baker Inc. (Phillipsburg, NJ). Dialysis membranes were purchased from Spectrum Labs (Rancho Dominguez, California): Spectra/Por® 1 (6-8 kD MWCO membrane; part 132660), Spectra/Por 6® (50 kD MWCO membrane; part 132544) and Float-A-Lyzer G2 Dialysis Device CE (100 kDa MWCO; part G235035). Bio-Scale™ Mini Macro-Prep High Q® cartridge (cat. 7324122), Bio-Scale™ Mini CHT™ Type I cartridge (cat. 7324322), Bio-Scale™ Mini Macro-Prep DEAE® Cartridge (cat. 7324142), Macro-Prep® t-Butyl HIC Support (1580090), Macro-Prep High Q® Support (cat. 1580040), and Econo® glass

columns were purchased from Bio-Rad, Inc. (Hercules, CA). HisPur™ Ni-NTA Resin (cat. 88222), 0.45 µm nitrocellulose membrane (cat. 77010), Western Blotting Filter Paper (cat. 88600), Pierce Fast Western ECL kit (cat. 35050), Pierce Fast Western Blot Kit, ECL substrate (cat. 35050) and anti-His-tag antibody (cat. PA1-983B) were purchased from ThermoFisher (Waltham, MA). Urea, sodium phosphate monobasic (NaH_2PO_4), Sephadex G-50 Fine (cat. 65080-106), Sephacryl 400-HR (cat. S400HR), Whatman Puradisc 25 syringe filters – 1.0 µm PES membrane (cat. WHA67812510), and Corning Spin-X UF 20 concentrators – 100 kDa MWCO (cat. CLS431491-12EA) were purchased from Sigma Aldrich (St. Louis, MO).

Ni-NTA Affinity Chromatography

Histidine tags (His-tags) are commonly used to purify recombinant proteins via IMAC (immobilized metal affinity chromatography). Recombinant BAF180 was cloned with a 6xHis-tag on the C-terminus (see Section 5.2). Ni-NTA affinity purification was tried several times, both batch and column under native and denaturing conditions. Initially, a 50 mM Tris buffer, pH 7.4 was tried; the cells were lysed in 50 mM Tris, so this removed a dialysis step. Both batch and column modes were tried and an imidazole gradient was used, from 50-250 mM. For batch mode, 10 mL of HisPur® resin slurry was added to a 50 mL tube and equilibrated with 50 mM Tris, pH 7.4. 15 mL of crude lysate (in the same buffer) was added to the tube and the mixture was incubated on an orbital shaker for 30 minutes. The resin was allowed to settle and the supernatant was collected. 15 mL of wash buffer 1 (50 mM imidazole) and 15 mL wash 2 (100 mM imidazole) were added and collected using the same procedure. 15 mL of elution buffer (250 mM

imidazole) was added and incubated for 1.5 hours before being collected. Results demonstrated that BAF180 was not binding to the column and was leaving with the flow-through.

The next several attempts used buffers recommended by the manufacturer of the Ni-NTA resin.[93] The samples were denatured to ensure that the conformation of the protein was not hindering the binding of the His-tag; many experiments have shown that the addition of denaturation agents does not affect the ability of a His-tag to bind with the Ni-NTA resin.[94] In a column (6 cm bed height, 0.7 cm diameter), approximately 2 mL of resin was equilibrated with 4 mL equilibration buffer (PBS, 6 M GuHCl, 10 mM imidazole, pH 7.4). Imidazole was included in the equilibration buffer to reduce non-specific binding.[94-99] Approximately 2 mL of crude lysate was mixed with 2 mL of equilibration buffer. The column was washed with 4 mL of wash buffer (PBS, 6 M GuHCl, 25 mM imidazole, pH 7.4) and eluted with 4 mL of elution buffer 1 (PBS, 6 M GuHCl, 250 mM imidazole, pH 7.4). No BAF180 band was observed in any fraction, so a second elution buffer with a higher imidazole concentration was used in the next trial (elution buffer 2: PBS, 6 M GuHCl, 500 mM imidazole, pH 7.4). Again, no BAF180 band was observed in any fraction. It was concluded that perhaps BAF180 was degrading prior to purification, during purification, or that the BAF180 band was too faint to observe.

After high-producing cells were discovered, Ni-NTA purification was tried again. New columns were prepared with 3 mL resin slurry (~1.5 mL resin) and equilibrated with 6 mL equilibration buffer (PBS, 6 M GuHCl, 10 mM imidazole, pH 7.4). For this

attempt, five different columns were prepared, each loaded with a different crude sample. For each, 500 µL sample was mixed with 1 mL equilibration buffer and loaded onto the column. Bound proteins were eluted in 3 mL of elution buffer 1 (PBS, 6 M GuHCl, 250 mM imidazole, pH 7.4). BAF180 was found in the flow-through fraction (Figure 6.6). No wash step was performed; this was a simple test to determine if BAF180 was binding to the Ni-NTA column at all.

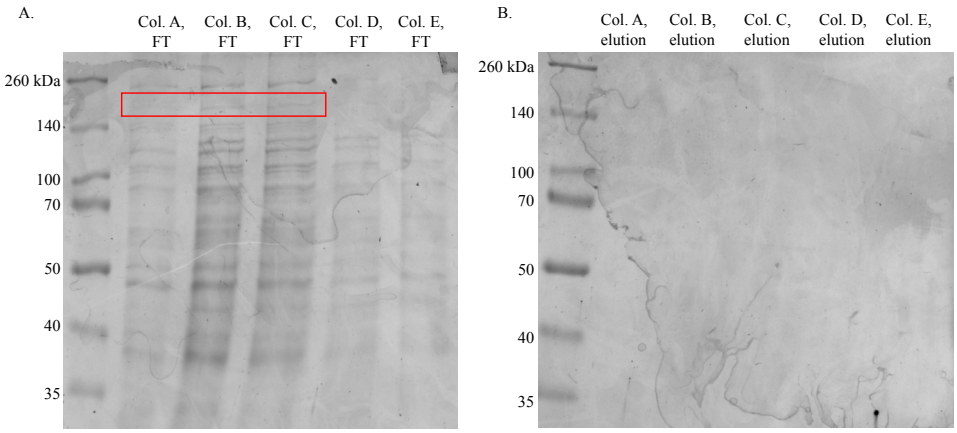


Figure 6.6: Denaturing Ni-NTA purification. BAF180 was found in the flow-through fraction (A), indicating that it did not bind to the column. No proteins were seen in the elution fractions (B). Both soluble and insoluble lysate fractions were used. FT stands for flow-through. The BAF180 band is denoted by the red stars.

Final attempts used a third and fourth buffer system. The third buffer system was 20 mM NaH_2PO_4, 300 mM NaCl, 6 M GuHCl, pH 7.4. 2 mL of resin slurry was added to a large column (2 x 23 cm) and equilibrated with equilibration buffer (20 mM NaH_2PO_4, 300 mM NaCl, 6 M GuHCl, pH 7.4). Imidazole was excluded from this wash buffer to be sure that the low 10 mM imidazole concentration used in previous equilibration buffers was not preventing BAF180 from binding. Multiple soluble lysate fractions (of high-producing cells) were combined (~32.5 mL) and mixed with 20 mL equilibration buffer

before being loaded onto the column. The flow-through was collected and a 5 mL aliquot was saved; the rest of the flow-through was added back into the column, so the sample would pass over the Ni-NTA resin bed twice. The column was washed twice with equilibration buffer. Samples were eluted with ~38 mL elution buffer (20 mM NaH_2PO_4, 300 mM NaCl, 6 M GuHCl, 300 mM imidazole, pH 7.4) and 2 mL of 1 M imidazole. 1 M imidazole was added to the column in an effort to elute anything that might be bound. All the samples were concentrated before being run on an SDS-PAGE gel. After concentration, samples were re-suspended and dialyzed in 50 mM Tris, pH 7.4 and run on an SDS-PAGE gel (Figure 6.7).

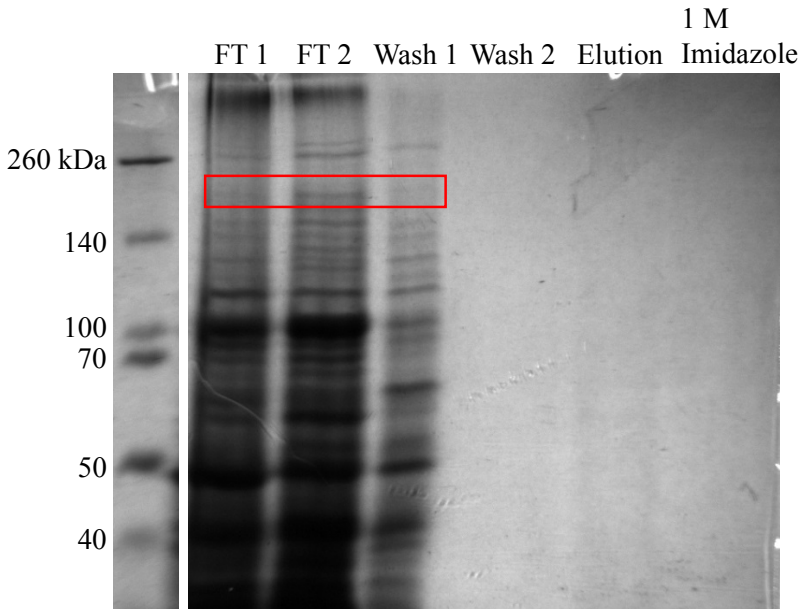


Figure 6.7: Denaturing Ni-NTA purification. Soluble fractions were combined and added to a Ni-NTA column. The flow-through was collected and re-run over the resin bed, so that the sample would pass over the Ni-NTA resin twice. The column was washed twice with equilibration buffer and elution buffer was added (300 mM imidazole). A final elution with 2 mL of 1 M imidazole was loaded onto the column to elute everything that might be bound.

The fourth buffer system was the same as the third buffer system, except the 6 M GuHCl was replaced with 8 M urea. A 1 mL Ni-NTA resin bed was equilibrated with

equilibration buffer (20 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, pH 7.4). Insoluble lysate fractions (in 8 M urea) were pooled and combined with equilibration buffer and loaded onto the prepared column. The column was washed with 40 mL of equilibration buffer before 4 – 1 mL portions of 1 M imidazole was used to elute any bound proteins. BAF180 was found in all fractions (flow-through, wash, and 1 M imidazole elutions), which seemed to indicate that some of the His-tagged BAF180 was binding to the Ni-NTA resin (Figure 6.8). Additional experiments were conducted, adding in a wash step with 30 mM imidazole and another elution series with 3 M imidazole. The results were unable to be replicated (Figure 6.9).

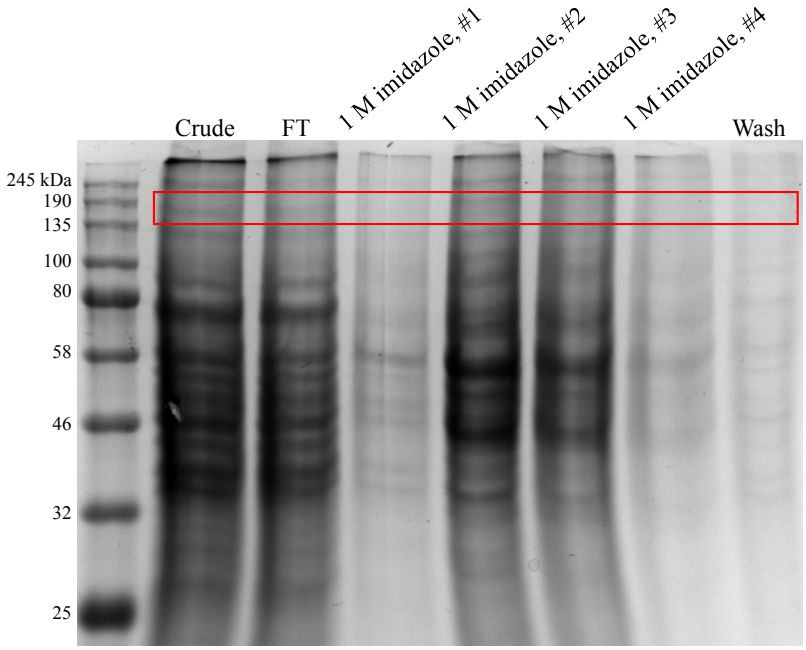


Figure 6.8: Denaturing Ni-NTA purification with 8 M urea. Insoluble lysate fractions were pooled and equilibrated with equilibration buffer. The column was washed with equilibration buffer before being treated to 1 M imidazole. BAF180 was found in most fractions, suggesting that some BAF180 was binding to the Ni-NTA resin.

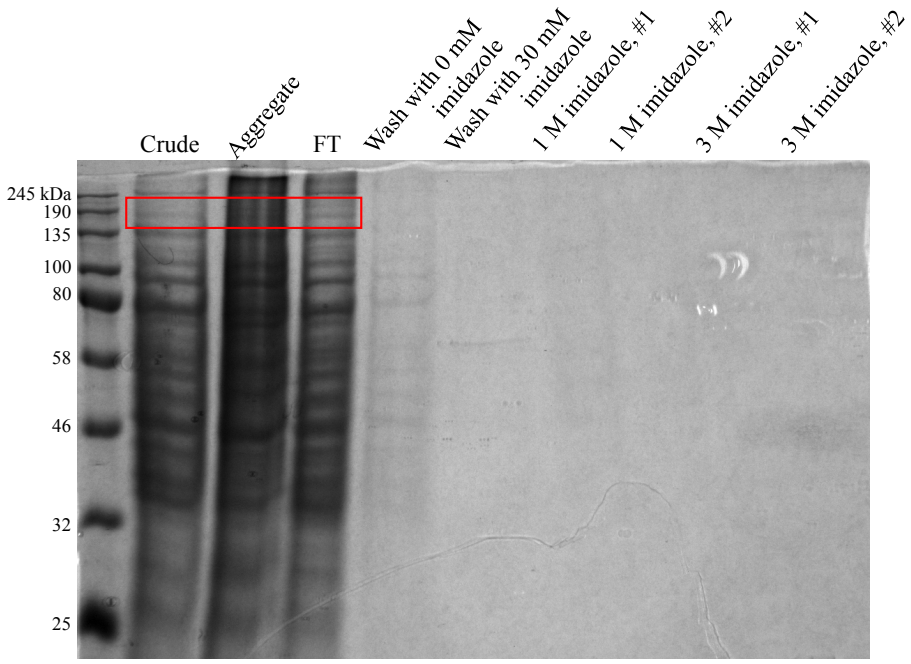


Figure 6.9: Further Ni-NTA attempts with 8 M urea. For this particular experiment and a few others, aggregates formed during purification. Nevertheless, BAF180 appears to be in the flow-through fraction. As can be seen, very few proteins are found in the wash or elution steps. The experiment indicated that BAF180 was not binding to the Ni-NTA resin.

Determination of the Presence of a His-tag

After native and denatured protocols for Ni-NTA purification indicated that BAF180 was not binding to the Ni-NTA resin, it was thought that this might be because the protein is so big (190 kDa) and the His-tag is so small (1 kDa), that the protein might have trouble staying attached to the resin. Later, it was considered that perhaps the His-tag was not actually present on the recombinant BAF180 protein. A dot blot was done to determine if the recombinant BAF180 protein was expressed with a His-tag. The negative control was WT X33 crude lysate (for which it had already been demonstrated that BAF180 was not present). The positive control was hPB1-1 (BAF180 protein, bromodomain 1 only) constructed by previous members of the Thompson lab and

successfully used in Ni-NTA affinity purification.[20] The protein was expressed in *E. coli* Rosetta(DE3) cells in the pET30b construct.[20] After growth and expression of the protein, the cells were lysed using a denaturing protocol, as only the His-tag was needed and this would ensure that the His-tag was exposed. The denaturing protocol was as follows: Harvested cells were re-suspended in *E. coli* denaturing lysis buffer (100 mM NaH_2PO_4, 10 mM Tris-Cl, 300 mM NaCl, 8 M urea, pH 8.0) at 5 mL buffer per gram cell. The cell suspension was incubated on ice for 1 hour on an orbital shaker. Lysate was collected by centrifuging the cell suspension at 10,000 g for 25 minutes at 4 °C.[20] The sample was dialyzed overnight in *E. coli* lysate dialysis buffer (50 mM Tris-Cl, 100 mM NaCl, pH 7.6). This was done primarily to remove urea; urea removal was deemed necessary in order to run the crude lysate sample on an SDS-PAGE gel and also to ensure that the urea could not cause any problems during the blot protocol.

In addition, to ensure that the His-tag was not buried within the protein structure, both samples of BAF180 (crude and anion exchange fraction, see Section 6.2) were denatured under harsh conditions. Approximately 100 µL of crude BAF180 was added to 500 µL of *E. coli* denaturing lysis buffer and incubated for 30 minutes at room temperature. The sample was then incubated for 10 minutes at 95 °C (chosen because this temperature is used for SDS-PAGE sample preparation) and dialyzed overnight in anion exchange dialysis buffer (17 mM Tri, pH 7.6). This protocol was repeated for 100 µL of purified BAF180 sample from anion exchange.

A dot blot was performed using an anti-His-tag antibody. Pierce Fast Western Blot Kit, ECL substrate was used. A grid was drawn in pencil and 2 µL of each sample

was spotted onto a nitrocellulose membrane (0.45 μm) and allowed to dry. The nitrocellulose membrane was then hydrated in 1x Fast Western Wash Buffer for 5 minutes with gentle shaking. The primary antibody solution (10 mL Fast Western Antibody Diluent, 2 μL anti-His-tag antibody) was incubated with the membrane for 1 hour. The membrane was washed for 5 minutes in wash buffer. The secondary antibody solution (10 mL Fast Western Antibody Diluent, 1 mL Fast Western Optimized HRP reagent) was added and incubated with the membrane for 10 minutes. The membrane was washed 4 times in wash buffer for 5 minutes each, followed by brief incubation in detection solution (5 mL Pierce ECL Detection Reagent 1, 5 mL Pierce ECL Detection Reagent 2). The membrane was visualized using a CDD camera. The results (shown in Figure 6.10) confirm that the positive control (hPB1-1) contains a His-tag and that the negative control (X33 WT crude) did not contain a His-tag. Most interestingly, none of the BAF180 samples showed a positive result, indicating, that in both native and denatured states, the recombinately expressed BAF180 did not contain a His-tag.

During the original construction of the expression plasmid, the purchased BAF180 DNA was modified using PCR mutagenesis to remove the stop codon present in the original sequence. The reverse PCR primer was designed to remove the stop codon, add an enterokinase site, and add a KpnI site. It is most likely that the stop codon was not removed during PCR modification, therefore causing termination of protein translation before the His-tag sequence (Figure 6.11). However, the KpnI site was definitely added, because this restriction enzyme was successfully used to ligate BAF180 into pGAPZA. The difference in base pair length between deleting the stop codon and inserting the

enterokinase is very small from that with just the stop codon. This slight difference would not have been observed on a gel.

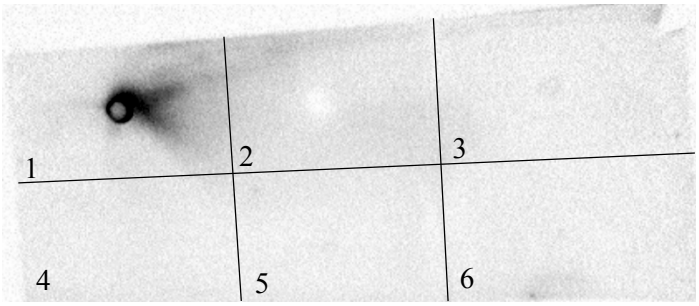


Figure 6.10: The results of the His-tag detection dot blot. A grid was drawn on nitrocellulose membrane and the samples were spotted in different boxes. The box assignments are as follows: (1) hPB1-1, positive control; (2) X33 WT, negative control; (3) BAF180 crude (native); (4) Anion exchange purified BAF180 (native); (5) Denatured BAF180 crude; (6) Denatured anion exchange purified BAF180. Only box 1 shows a positive result, indicating that hPB1-1 is the only sample that contained a His-tag.

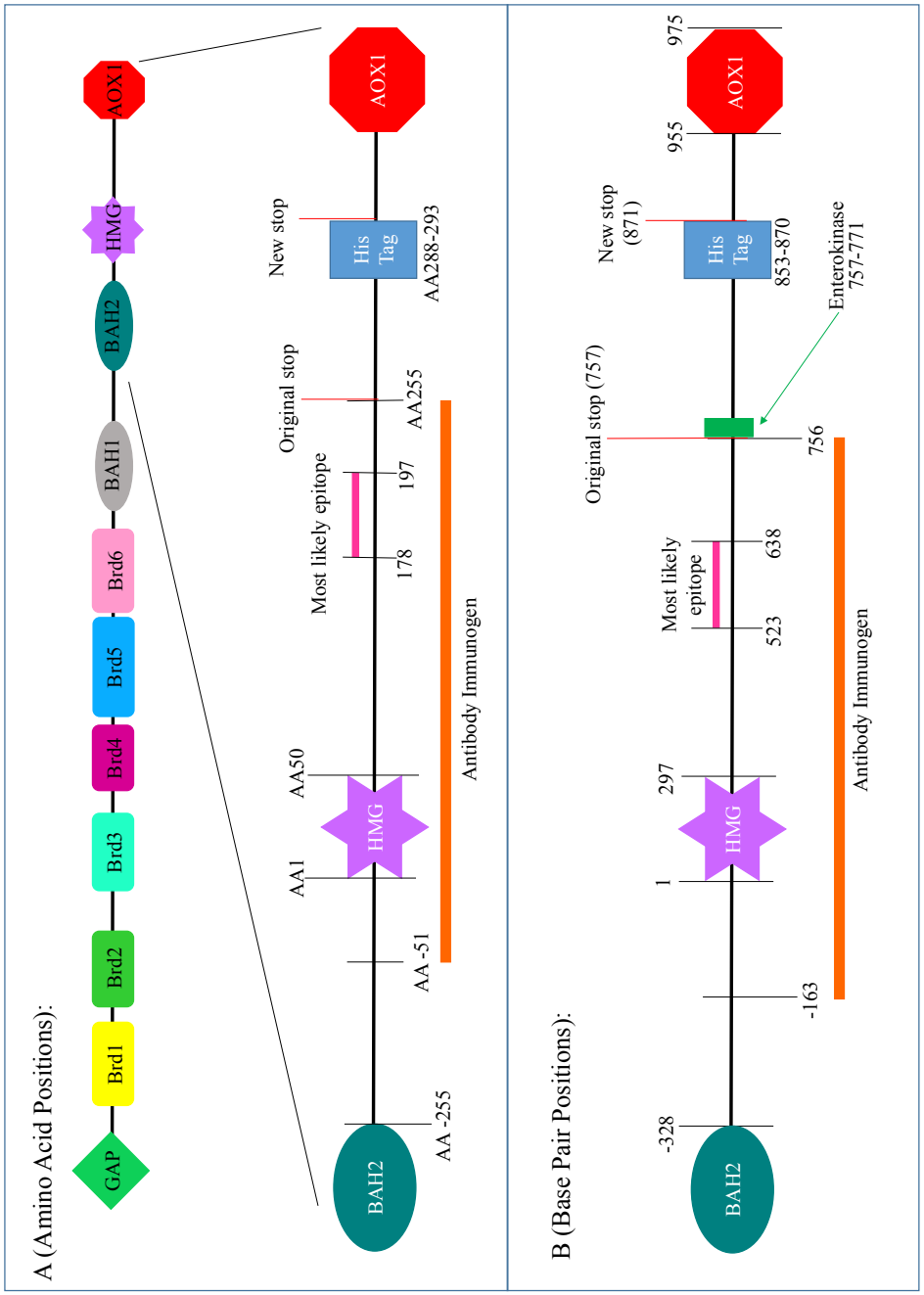


Figure 6.11: Possible explanation for why there was no His-tag detected with an anti-His-tag antibody. All numbering is relative to the HMG box, with the first amino acid or base pair of the HMG box numbered as 1. There is no 0 position (-2, -1, 1, 2, etc). The positions of the antibody immunogen and the most likely epitope are noted. It was thought that the original stop codon was not eliminated and therefore protein translation was terminated at this point, preventing a His-tag from being included on the expressed protein. (A) Amino acid positioning; (B) Base pair positioning.

Strong Anion Exchange Resin: Bis-Tris and Bis-Tris Propane Buffers

Other buffers were tried during AEX when it was still believed that the co-eluting proteins in the purified fraction were contaminants. Decreasing the difference between BAF180's pI and the buffer's pH would result in a smaller negative charge on the protein, therefore it might elute at a lower NaCl concentration, possibly separate from all other cellular proteins. The buffer pH was still kept relatively close to pH 7.4, because BAF180 was shown to be stable at this pH; it was reasoned that a buffer pH further away from 7.4 would result in decreasing stability of BAF180.

Decreasing the difference between BAF180's pI and the buffer's pH required the use of different buffer systems with lower pKa values. The pKa values of Tris buffers limited the pH values that could be used; 10 mM Tris has a pKa of 8.10 at 25 °C, and the pKa increases as the concentration increases. Therefore, the lower limit for the pH of a Tris buffer was 7.6, which was used. The structures of the different buffer salts was also considered; it was reasoned that molecules with a structure closer to that of Tris would be better, since it was already known that the recombinant BAF180 was stable in Tris. Organic molecules were avoided because of the unknown effect on protein stability and the likelihood of denaturation.

After analysis of different buffer systems at different concentrations and temperatures, it was decided to use Bis-Tris next. Bis-Tris has an effective buffering range of 5.8-7.2.[85] A buffer system of 50 mM Bis-Tris, pH 7.3 was tried first; this buffer needed to be used at 4 °C in order for the pH and pKa values to be in agreement (i.e. keep

the pKa within 0.5 pH units of buffer pH). A step-wise elution from 0 – 1.0 M NaCl was used. Results show that most of the cellular proteins and BAF180 were eluting at 0.5 M NaCl (Figure 6.12). Therefore, this buffer system is essentially useless, because it did not separate BAF180 from any other proteins.

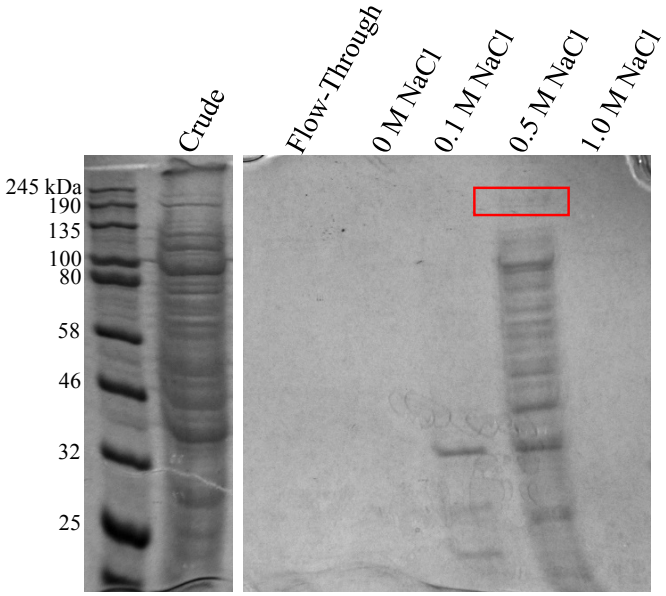


Figure 6.12: Anion exchange with 50 mM Bis-Tris, pH 7.3. As can be seen in the gel, most of the cellular proteins and BAF180 eluted at 0.5 M NaCl. This buffer system did not separate BAF180 from the other proteins. The purification experiment was conducted at 4 °C.

When the SDS-PAGE results showed that pH 7.3 did not separate BAF180 from other cellular proteins, the pH was changed to 7.0. With a 50 mM Bis-Tris buffer at pH 7.0, anion exchange was conducted again at room temperature (~20 °C). BAF180 did elute in a different fraction than many other cellular proteins, though quite a few cellular proteins eluted in the same fraction as BAF180 (Figure 6.13). BAF180 eluted primarily at 0.4 M and 1.0 M NaCl, though a small amount was seen in 0.3 M NaCl and the flow-through. The BAF180 elution fraction was not as pure as that achieved with 17 mM Tris

buffer, pH 7.6. Additionally, BAF180 appears to be eluting at a higher concentration of NaCl. Increasing the NaCl concentration increases the risk of protein denaturation.

pH values below 7.0 were not attempted, for concern that the pH would start to denature the BAF180 protein. pH 7.0 is 0.4 pH units away from known stability at pH 7.4 (lysis buffer 2). In fact, proteins can be seen aggregating in the wells at pH 7.0 (Figure 6.13).

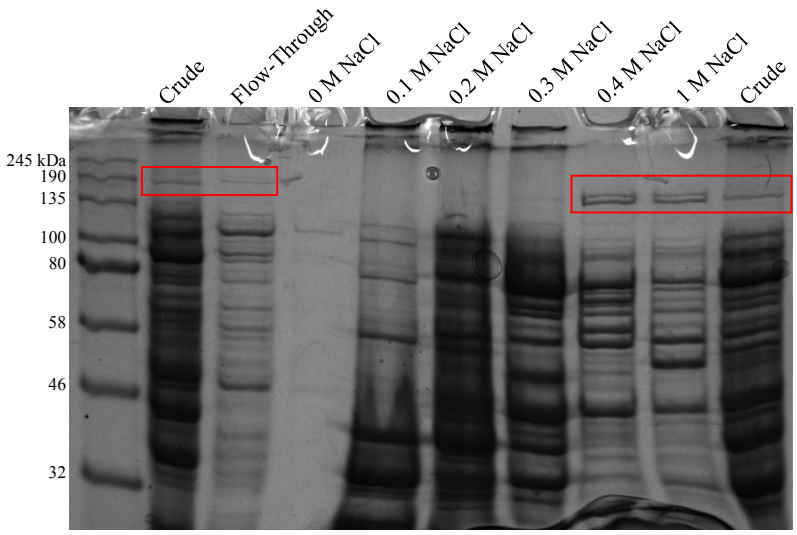


Figure 6.13: Anion exchange with 50 mM Bis-Tris, pH 7.0. BAF180 did separate from many of the cellular proteins, though quite a few eluted with BAF180. BAF180 eluted at 0.4 M NaCl and 1.0 M NaCl, with a small amount in the 0.3 M NaCl fraction. The BAF180 elution is not as clean as the elution achieved with 17 mM Tris buffer, pH 7.6 and requires a higher concentration of NaCl to elute. BAF180 is denoted by the red star.

A Bis-Tris propane (BTP) buffer system was tried next. The pKa of 50 mM BTP is 7.12 at 20 °C. The effective buffering range for BTP is 6.4-7.3;[100] pH 7.4 was chosen. BAF180 eluted at 0.3 M NaCl, along with several other cellular proteins. Comparing these results (Figure 6.14) to those from AEX with 17 mM Tris, pH 7.6 (Figure 6.3), BAF180 eluted at a lower NaCl concentration (0.3 M vs. 0.4 M, respectively). However,

it eluted with many more contaminant proteins in BTP buffer than in Tris buffer. Since BAF180 appears to be stable at pH 7.6 (Tris buffer) and elutes with fewer contaminants, anion exchange was conducted with Tris buffer for all future experiments. The tradeoff (eluting at a lower NaCl concentration but with more contaminants) was not considered to be advantageous. It was interesting that the two proteins that co-eluted with BAF180 using the Tris buffer also co-eluted here. It was puzzling at the time why those proteins always eluted together. It is now understood, however, that the other bands are fragments of BAF180.

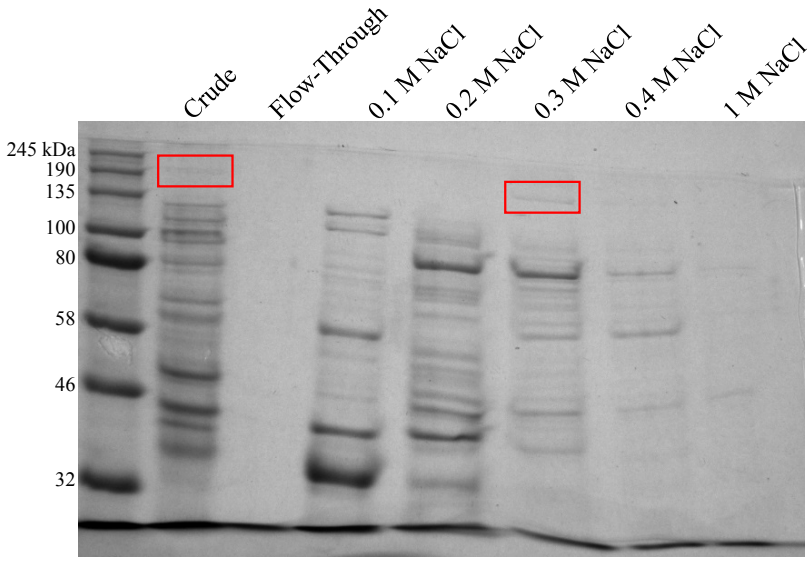


Figure 6.14: Anion exchange with 50 mM Bis-Tris propane buffer, pH 7.4. Though faint, BAF180 can be seen eluting at 0.3 M NaCl, along with several other cellular proteins. Even though BAF180 eluted at a lower concentration of NaCl (0.3 M vs. 0.4 M in 17 mM Tris, pH 7.6), so did other cellular proteins. BAF180 is denoted by the red star.

Weak Anion Exchange: DEAE Resin and Tris Buffer

In addition to strong anion exchange media, the weak anion exchanger DEAE was tried. Bio-Rad's MacroPrep® DEAE is a weak anion exchanger that is a methacrylate polymer with a DEAE (diethylaminoethyl) cellulose group attached.[82,101] Weak and strong ion exchangers have different selectivities,[84-85] so it seemed worthwhile to try a weak anion exchanger as well.

The same buffer as for strong AEX was used (17 mM Tris, pH 7.6). Crude lysate was run through the DEAE cartridge and eluted using various concentrations of NaCl (Figure 6.15). BAF180 eluted at 0.5 M NaCl, along with many contaminants, though a faint amount can be seen in the 0.3 M fraction.

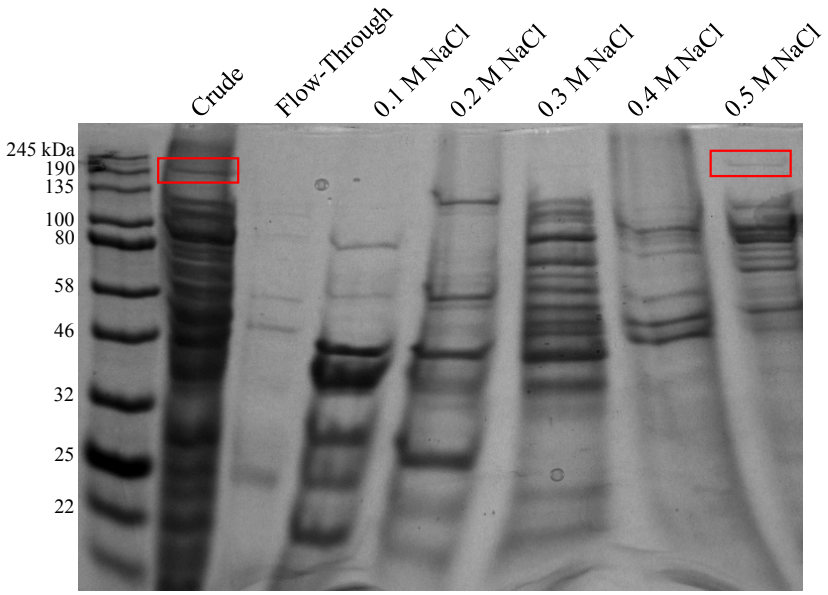


Figure 6.15: DEAE anion exchange. Using 17 mM Tris buffers at pH 7.6, crude lysate was run through the media and eluted at varying concentrations of NaCl. The BAF180 band (denoted by the red star) was seen in the 0.5 M NaCl fraction, along with several contaminant proteins.

Anion Exchange: Prepacked Cartridges vs. Self-Packed Columns

Initially, prepacked cartridges were used. Manipulation of NaCl concentrations were unable to separate the "contaminant proteins" from BAF180. After possible variations in buffers were explored, a longer column was tried. In chromatography, resolution often increases with column length, so there was a possibility that increasing the column length of Bio-Rad High Q® resin would separate out the two contaminant proteins (approximately 70 and 80 kDa, see Figure 6.3). Approximately 10 mL resin slurry was packed into a 0.7 x 20 cm Econo® glass column. Unfortunately, elongating the column did not separate the three bands (results not shown). The new finding that the "contaminants" are fragments of BAF180 explain these results.

Hydroxyapatite (Mixed-Mode) Chromatography

Hydroxyapatite chromatography is not often utilized for protein purification, even though it was first published in the 1950s.[84,102] Hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$) is a mixed-mode (also called multimodal) media (or resin), meaning that its interactions with the proteins appear to be a combination of electrostatic and metal affinity interactions.[84,102-109]

Bio-Rad's Bio-Scale™ Mini CHT™ ceramic hydroxyapatite type I pre-packed cartridges were chosen for analysis. CHT is a ceramic hydroxyapatite that is much more stable than the traditional forms[102-109] and CHT can reportedly separate mixtures that other methods cannot.[103-107] Type I is recommended for acidic proteins (the pI of BAF180 is 6.29).[103-107]

CHT™ Type I Bio-Scale™ mini cartridge was prepared as directed in manual.[110] NaH_2PO_4 buffers at pH 7.2 were used with increasing concentration to elute the proteins. The loading buffer had 10 mM NaH_2PO_4, which was increased step-wise to 400 mM (10 mM, 50 mM, 100 mM, 200 mM, 300 mM, 400 mM). Acceptable purity results were not obtained with this method. Results demonstrate that BAF180 is eluting in 200 mM NaH_2PO_4 along with many other cellular proteins (Figure 6.16).

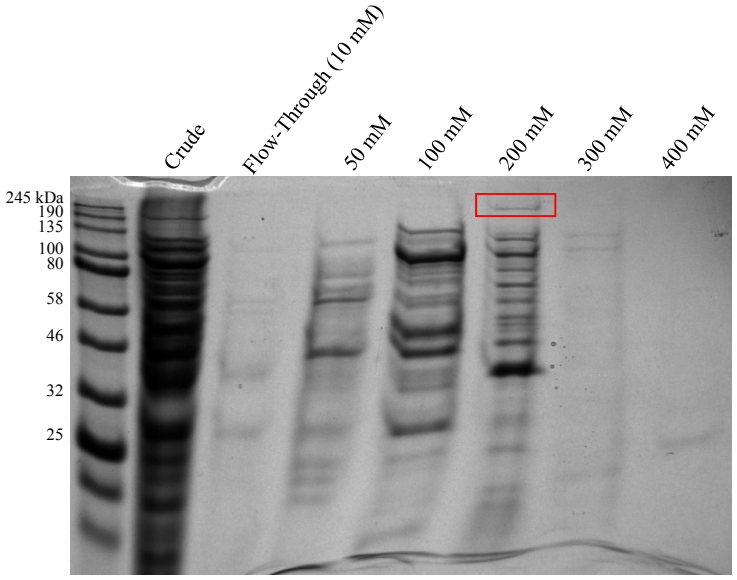


Figure 6.16: CHT Type I chromatography. BAF180 elutes at 200 mM NaH_2PO_4 along with many other cellular proteins. All concentrations are referring to NaH_2PO_4 concentration.

Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography (HIC) takes advantage of the structure of a protein molecule and its interactions with water. Decreasing salt gradients are used to elute a bound protein off a column.[111-115]

Bio-Rad's Macro-Prep® t-butyl resin was chosen because it has strong interactions with proteins that are weakly hydrophobic.[116-118] Because BAF180 is soluble in aqueous solutions, it was reasoned that the degree of hydrophobicity of the protein cannot be too high, otherwise it would not be soluble in water. Macro-Prep® t-butyl resin consists of a methacrylate bead with t-butyl and carboxyl groups attached.[116-118] Following manufacturer's instructions, the t-butyl resin was packed into a 1.5 mm x 10 cm Econo® glass column, with the bed height approximately three quarters of the column height. The elution fraction from anion exchange chromatography (17 mm Tris, 0.4 M NaCl, pH 7.6) was equilibrated to 1 M NaCl by addition of a 2 M buffer (17 mM Tris, 2 M NaCl, pH 7.6) and loaded onto the prepared HIC column. Flow-through was collected and subsequent washes were carried out with a decreasing salt gradient (1 M to 0 M NaCl). The samples were then concentrated and run on an SDS-PAGE gel. Results were inconclusive (not shown).

Size Exclusion Chromatography

Size exclusion chromatography (SEC), also called gel filtration chromatography, is a fairly simple method. Using porous resin, proteins are separated by molecular weight. Large proteins cannot enter the pores and are thus eluted first. Smaller proteins are able to travel through the pores (the degree to which they can enter the pores depends on their size), and are thus slower to elute.[119-123]

Sephadex G-50

Sephadex G-50 has an approximate fractionation range of 1-30 kDa, meaning it can separate proteins with molecular weights of 1-30 kDa;[120-122] it was chosen because it

was readily available. The Sephadex G-50 column was prepared by swelling the resin particles in 50 mM Tris-Cl, pH 7.4 on a hotplate for approximately 1 hour. The slurry was then poured into a 2 x 23 cm glass column; the resultant bed height was approximately 10 cm. Crude lysate was added and 8 – 2 mL fractions were collected (Figure 6.17). Separation did not occur.

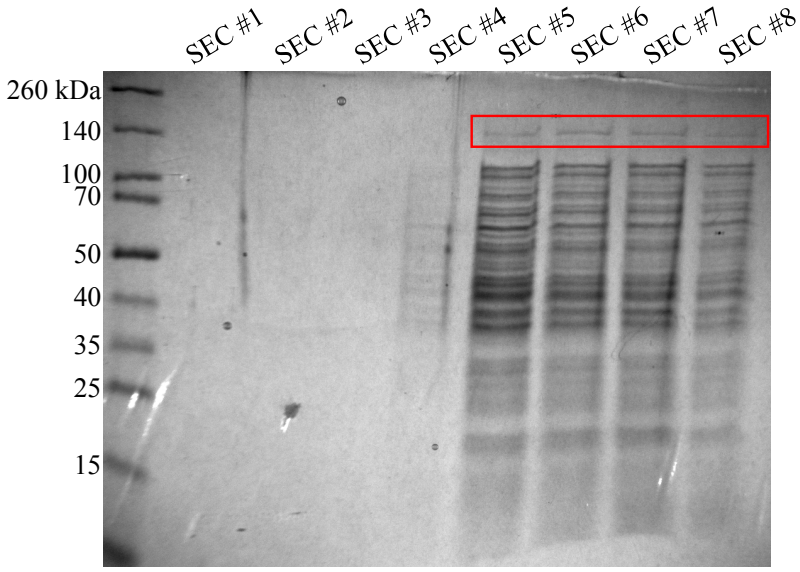


Figure 6.17: Size-exclusion chromatography of crude BAF180 lysate. 8 – 2 mL fractions were collected and concentrated. No separation was seen, though dilution was occurring. Note that the lysate used was from low-producing cells and so the BAF180 band (~190 kDa) was very faint. BAF180 eluted in fractions 5-8. The band for BAF180 is found in the red box.

SEC was also tried with partially purified samples. Anion exchange (AEX) was able to remove all but 2 contaminant proteins (now known to be degradation products) from BAF180. This eluent fraction (0.4 M NaCl) was concentrated and run through the column. 11 – 2 mL fractions were collected and concentrated. Though the bands are faint, BAF180 still eluted with contaminant proteins (Figure 6.18).

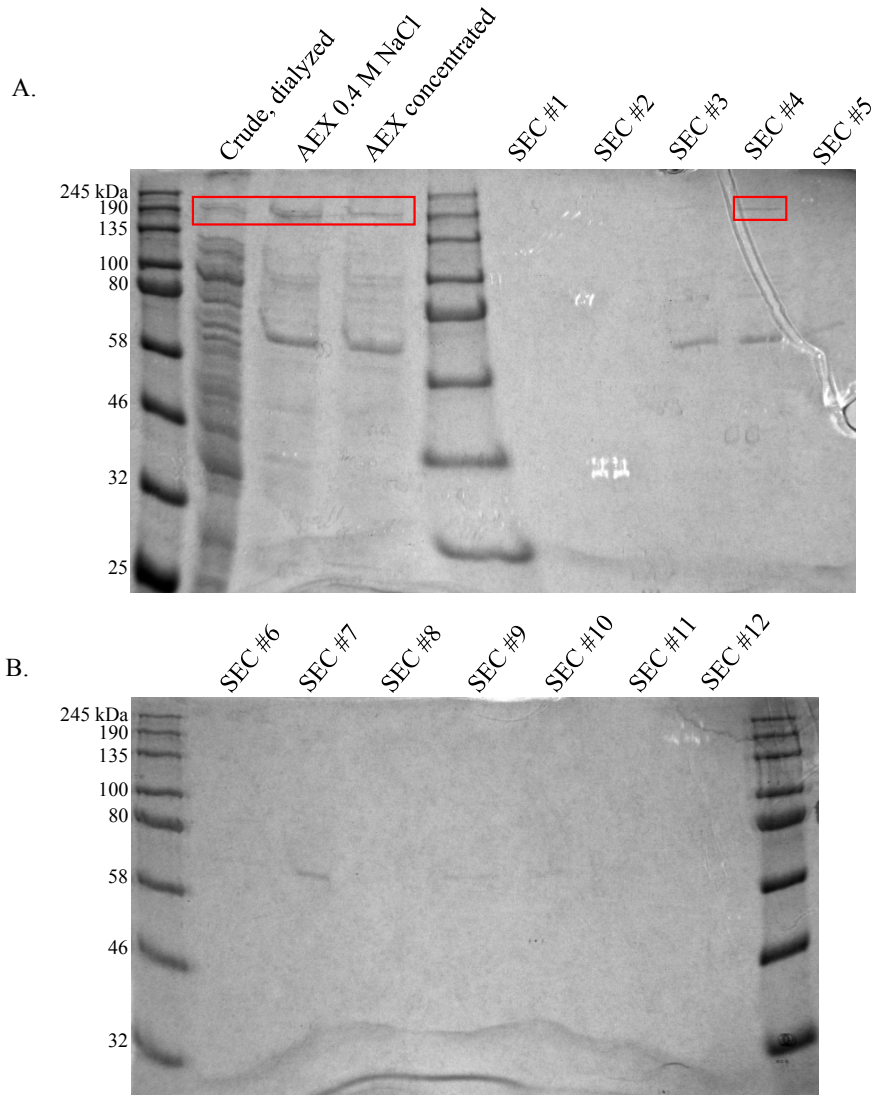


Figure 6.18: Size-exclusion chromatography of partially purified BAF180. AEX fraction B4 (0.4 M NaCl) was concentrated and then run through a Sephadex G-50 column. BAF180 still eluted with the other contaminants (fraction 4), though partial separation did occur. BAF180 is denoted by the red star.

Longer SEC Column Packed with Sephadex G-50

A longer SEC column (0.7 cm x 20 cm) was packed to a bed height approximately 17 cm (Sephadex G-50) and the 0.4 M AEX fraction was run through. In chromatography, resolution is dependent on column length.[121] 30 – 2 mL fractions were collected. Though the gels were not great, faint bands at ~70 and ~80 kDa (the size of the

contaminants) can be seen eluting in the same fractions (Figure 6.19). This led to the supposition that this longer column would not be able to separate BAF180 either.

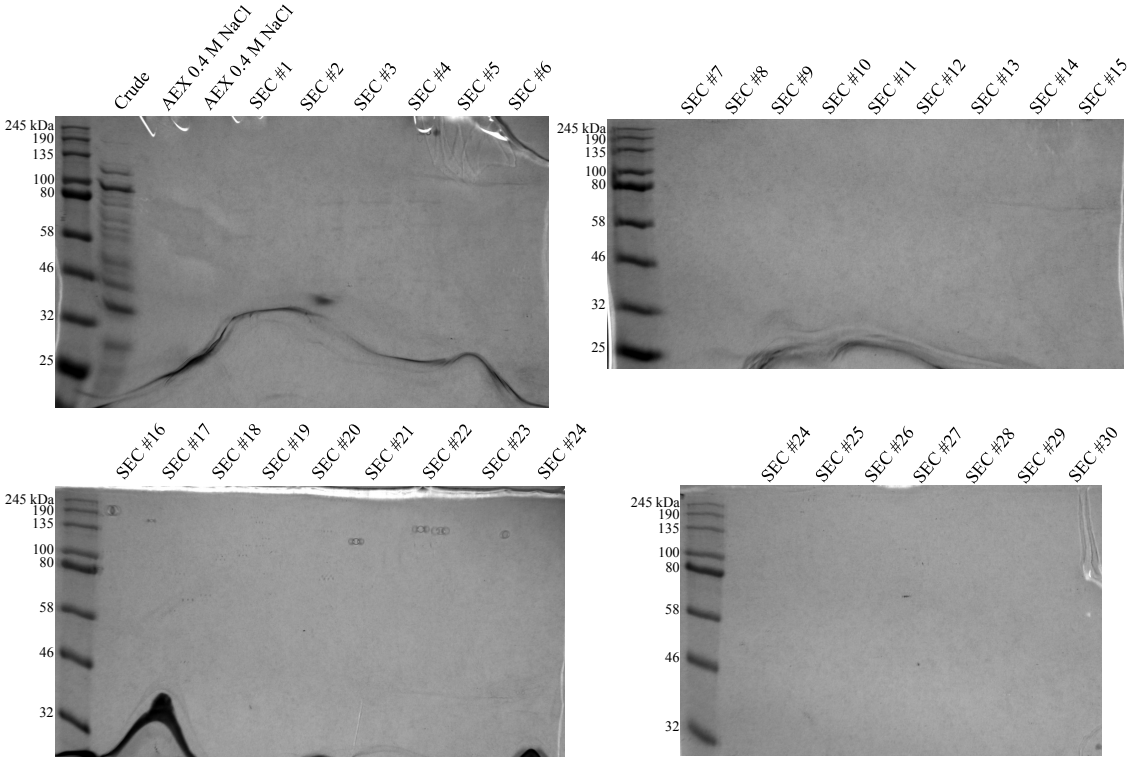


Figure 6.19: Size-exclusion chromatography of partially purified BAF180 on a longer column. Though the bands are very faint, the two contaminant proteins can be seen eluting together in fractions 2-4.

SEC with Sephacryl 400-HR

Another SEC media was tried. Sephacryl 400-HR has a fractionation range of 20-8,000 kDa. Pre-swelled resin was packed into a 15 x 750 mm column, with a bed height approximately 500 mm. Concentrated AEX 0.4 M NaCl fraction was loaded onto the column. The first 23 mL were collected, followed by 34 fractions of 1 mL each. The results were inconclusive, as no bands were seen (results not shown).

Ultrafiltration and Dialysis

Attempts were also made to purify by size directly from a crude sample. Both ultrafiltration and dialysis utilize membranes with distinct molecular weight cut offs (MWCO), allowing proteins of smaller molecular weight to flow out of the membrane, while proteins of higher molecular weight are retained.

Ultrafiltration devices are able to dialyze and concentrate samples during centrifugation. Proteins smaller than the MWCO are able to flow-through the membrane, as does buffer. For subsequent spins (centrifugation), additional buffer is added. Two types of ultrafiltration devices were tried, both with a 100 kDa MWCO, but with a different membrane composition. Merck Millipore's Amicon® Ultra used a regenerated cellulose membrane[124] and Corning's® Spin-X® used a PES (polyethylsulfone) membrane.[125]

The Amicon® Ultra ultrafiltration devices were tried on early samples; this was during growth and lysis optimization, so the protein content is not the same as after optimization. However, after four centrifugation cycles, the sample appears to be concentrating and the purity is slightly improved (Figure 6.20). Unfortunately, these results could not be replicated.

A PES membrane has lower protein binding than other membranes.[126-127] The Corning® Spin-X® ultrafiltration devices have a PES membrane. Unfortunately, they did not serve as an efficient means of purifying BAF180 from endogenous cellular proteins (Figure 6.21). No significant difference in protein content was observed. MWCO can be

relative, so the retention of proteins around 100 kDa is not unexpected; however, proteins 50 kDa and below should have been removed. The reason that proteins below 50 kDa were not removed is not understood.

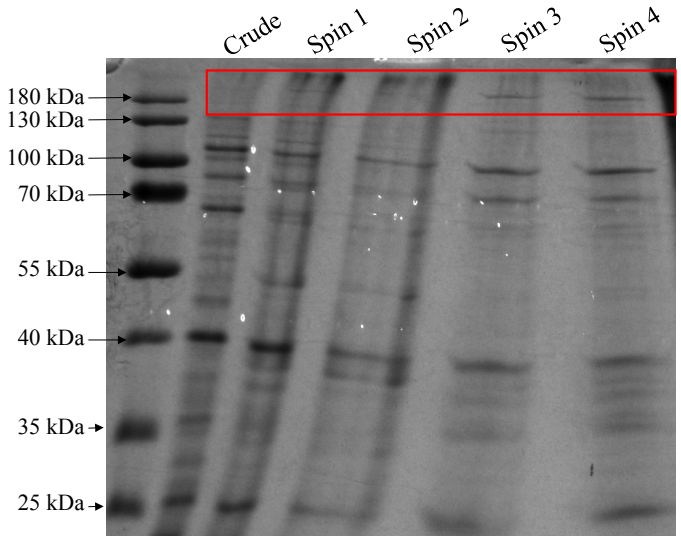


Figure 6.20: Ultrafiltration with Amicon® Ultra device. These early samples had not yet been optimized, so the protein content is different than it is currently. However, after four centrifugation cycles, the sample is slightly more concentrated and slightly purer.

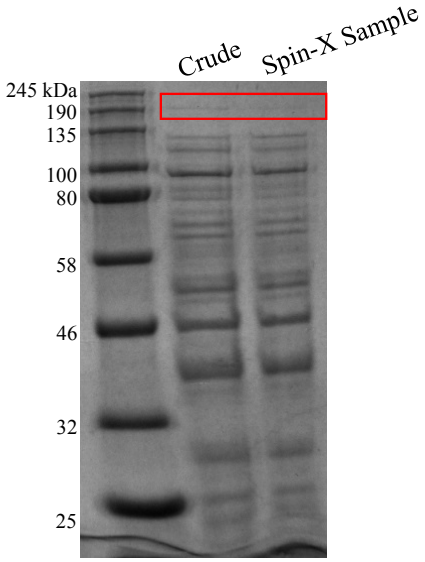


Figure 6.21: Ultrafiltration with Corning Spin-X device. The difference in the crude and Spin-X samples is negligible.

Additionally, dialysis membranes with 100 kDa (Figure 6.22A) and 50 kDa (Figure 6.22B) MWCO were tried. Oddly, the protein composition of the BAF180 sample prior to dialysis was nearly the same as after dialysis; it just appeared slightly less concentrated, though this could have been due to variations in SDS-PAGE loading. It is unclear why larger pore dialysis membranes were unable to remove the smaller proteins.

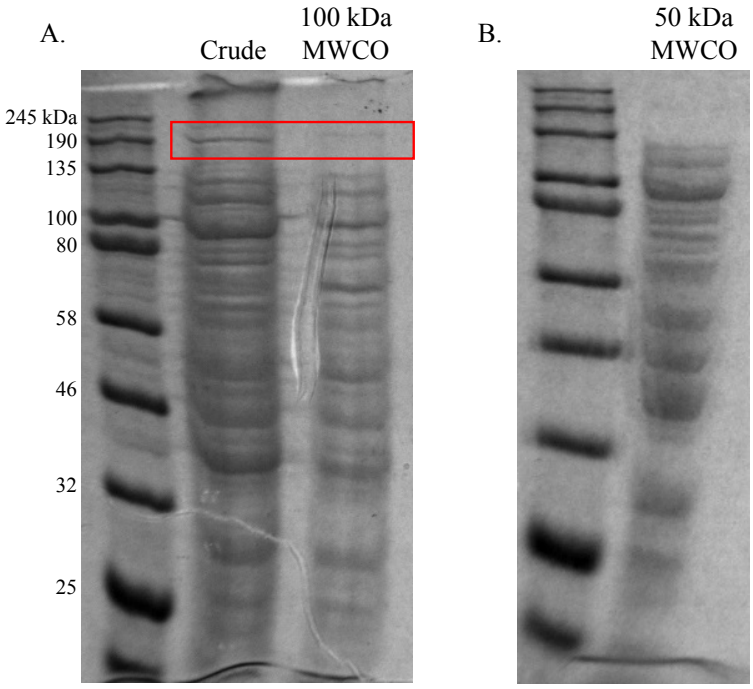


Figure 6.22: Dialysis with high MWCO membranes. (A) Overnight dialysis with 100 kDa MWCO membrane did not appear to improve the purity of crude BAF180 sample. It appears that the dialyzed sample is slightly less concentrated. (B) Sample purity was not improved with use of a 50 kDa membrane.

6.5 References

1. Bio-Rad Inc. Anion Exchange Chromatography. http://www.bio-rad.com/en-us/applications-technologies/anion-exchange-chromatography (accessed 18 March 2017).

2. Selkirk, C., Ion-Exchange Chromatography. In *Protein Purification Protocols*, second ed.; Cutler, P., Ed. Humana Press: Totowa, NJ, 2004; Vol. 244, pp 125-131.

3. Karlsson, E.; Hirsh, I., Ion Exchange Chromatography. In *Protein Purification: Principles, High Resolution Methods, and Applications*, third ed.; Janson, J.-C., Ed. John Wiley & Sons, Inc.: 2011; pp 93-133.

4. Bio-Rad Inc. Ion Exchange Chromatography. http://www.bio-rad.com/en-us/applications-technologies/liquid-chromatography-principles/ion-exchange-chromatography (accessed 18 March 2017).

5. Bio-Rad Inc., Macro-Prep® Ion Exchange Media: Instruction Manual. Rev D ed.

6. Bio-Rad Inc., Instruction Manual, Bio-Scale™ Mini Macro-Prep® High Q, High S, and DEAE Cartridges. Rev C ed.

7. Bio-Rad Inc., Macro-Prep® High Q Support Bulletin, Rev A. Rev A ed.

8. Khandelwal, P., Practical Guide: Selecting the Optimal Resins for the Process Purification of Native and Recombinant Proteins, Ver A. Ver A ed.; Bio-Rad Inc.

9. Bio-Rad Inc., Achieve High Producuctivity Using Macro-Prep High Q Anion Exchange Support. Rev A ed.

10. Jungbauer, A.; Hahn, R., Ion-Exchangy Chromatography. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 349-371.

11. REACH Devices Recommended Buffers for Anion Exchange Chromatography. http://www.reachdevices.com/Protein/buff_anion.html (accessed 18 March 2017).

12. Puxbaum, V.; Mattanovich, D.; Gasser, B., Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. *Applied Microbiology and Biotechnology* **2015,** *99* (7), 2925-2938.

13. Gleeson, M. A.; White, C. E.; Meininger, D. P., et al., Generation of protease-deficient strains and their use in heterologous protein expression. *Methods in molecular biology (Clifton, N.J.)* **1998,** *103*, 81-94.

14. Thermo Scientific, HisPur™ Ni-NTA Resin. 2012.

15. Block, H.; Maertens, B.; Spriestersbach, A., et al., Immobilized-Metal Affinity Chromatography (IMAC): A Review. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Elsevier: 2009; pp 439-473.

16. Kågedal, L., Immobilized Metal Ion Affinity Chromatography. In *Protein Purification: Principles, High Resolution Methods, and Applications*, third ed.; Janson, J.-C., Ed. John Wiley & Sons, Inc.: 2011; pp 183-201.

17. Yip, T.-T.; Hutchens, T. W., Immobilized Metal-Ion Affinity Chromatography. In *Protein Purification Protocols*, Cutler, P., Ed. Humana Press Inc.: Totowa, NJ, 2004; pp 179-190.

18. novagen, Ni-NTA His-Bind® Resins. Rev A ed.

19. GE Healthcare Life Sciences, Purification of histidine-tagged recombinant proteins. In *Recombinant Protein Purification Handbook: Principles and Methods*, 2009; pp 25-41.

20. GE Healthcare Life Sciences, Immobilized metal ion affinity chromatography (IMAC). In *Strategies for Protein Purification Handbook*, 2010; pp 20-21.

21. Chandrasekaran, R.; Thompson, M., Expression, purification and characterization of individual bromodomains from human Polybromo-1. *Protein expression and purification* **2006,** *50* (1), 111-7.

22. Lebendiker, M. Recommended Buffers for Anion Exchange Chromatography. http://wolfson.huji.ac.il/purification/buffers/Buffers_AEIC.html (accessed 18 March 2017).

23. Navvab, M.; Aberin, C.; Olech, L., et al., Macro-Prep DEAE Support for Chromatography of Biomolecules. Rev C ed.; Bio-Rad Inc.

24. Cummings, L.; Snyder, M.; Brisack, K., Protein Chromatography on Hydroxyapatite Columns. In *Guide to Protein Purification*, 2nd ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; Vol. 463, pp 387-404.

25. Bio-Rad Inc. CHT™ Ceramic Hydroxyapatite Type I Media. http://www.bio-rad.com/en-us/product/cht-ceramic-hydroxyapatite-type-i-media?pcp_loc=catprod (accessed 20 March 2017).

26. Ng, P.; He, J.; Cohen, A., How CHT Ceramic Hydroxyapatite Works. Ver A ed.; Bio-Rad Inc.

27. Bio-Rad Inc., CHT™ Ceramic Hydroxyapatite Product Information Sheet. Rev B ed.

28. Gagnon, P.; Frost, R.; Tunón, P., et al., CHT Ceramic Hydroxyapatite - A New Dimension in Chromatography of Biological Molecules. Rev C ed.; Bio-Rad Inc.

29. Bio-Rad Inc. Multimodal or Mixed-Mode Chromatography. http://www.bio-rad.com/en-us/applications-technologies/liquid-chromatography-principles/multimodal-or-mixed-mode-chromatography (accessed 20 March 2017).

30. Doonan, S., Chromatography on Hydroxyapatite. In *Protein Purification Protocols*, second ed.; Cutler, P., Ed. Humana Press: Totowa, NJ, 2004; pp 191-194.

31. GE Healthcare Life Sciences, Overview of Multimodal Chromatography. In *Multimodal Chromatography Handbook*, 2013; pp 17-22.

32. Bio-Rad Inc., Bio-Scale™ Mini CHT™ Ceramic Hydroxyapatite Cartridges, 5 mL: Instruction Manual. Rev A ed.

33. O'Farrell, P. A., Hydrophobic Interaction Chromatography. In *Protein Purification Protocols*, second ed.; Cutler, P., Ed. Humana Press: Totowa, NJ, 2004; pp 133-138.

34. McCue, J., Theory and Use of Hydrophobic Interaction Chromatography in Protein Purification Applications. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 405-414.

35. Tomaz, C. T.; Quieroz, J. A., Hydrophobic Interaction Chromatography. In *Liquid Chromatography: Fundamentals and Instrumentation*, Fanali, S.; Haddad, R.; Poole, C.; Schoenmakers, P.; Lloyd, D., Eds. Elsevier Inc.: 2013; pp 122-141.

36. Cummins, P.; O'Conner, B., Hydrophobic Interaction Chromatography. In *Protein Chromatography: Methods and Protocols*, Walls, D.; Loughran, S., Eds. Springer: 2011; pp 431-437.

37. Eriksson, K.-O.; Belew, M., Hydrophobic Interaction Chromatography. In *Protein Purification: Principles, High Resolution Methods, and Applications*, third ed.; Janson, J.-C., Ed. John Wiley & Sons, Inc.: 2011; pp 165-181.

38. Bio-Rad Inc., Instruction Manual, Macro-Prep® HIC Supports. Rev C ed.

39. Bio-Rad Inc., Macro-Prep® HIC Resin Product Information Sheet. Ver F ed.

40. Bio-Rad Inc. Macro-Prep® t-butyl HIC Resin http://www.bio-rad.com/en-us/product/macro-prep-t-butyl-hic-resin (accessed 21 March 2017).

41. Stellwagen, E., Gel Filtration. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 373-385.

42. Hagel, L., Gel Filtration: Size Exclusion Chromatography. In *Protein Purification: Principles, High Resolution Methods, and Applications*, third ed.; Janson, J.-C., Ed. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2011; pp 51-91.

43. Cutler, P., Size-Exclusion Chromatography. In *Protein Purification Protocols*, second ed.; Cutler, P., Ed. Humana Press: Totowa, NJ, 2004; pp 239-252.

44. Sigma-Aldrich Gel Filtration Chromatography. http://www.sigmaaldrich.com/life-science/proteomics/protein-chromatography/gel-filtration-chromatography.html (accessed 17 March 2017).

45. Amersham Biosciences, Gel Filtration: Principles and Methods. 2002.

46. Merck Millipore Ltd., Amicon® Ultra-15 Centrifugal Filter Devices. Rev A ed.; Merck Millipore Ltd.: 2013.

47. Corning® Life Sciences, Corning® Spin-X® UF 6 and 20 mL Concentrators - Technical Data and Operating Instructions. Rev 2 ed.; 2010.

48. Nalgene Rapid-Flow Sterile Disposable Filter Units with PES Membrane. https://www.thermofisher.com/order/catalog/product/124-0045 (accessed 22 March 2017).

49. Whatman Polyethersulfone (PES) Membrane Filter. http://www.gelifesciences.com/webapp/wcs/stores/servlet/productById/en/GELifeSciences-us/28417655 (accessed 21 March 2017).

Chapter 7: Conclusion and Future Work

For the first time, BAF180 has been successfully expressed in a heterologous host. Though attempts to express BAF180 in *E. coli* were unsuccessful, expression in the yeast *Pichia pastoris* was successful. Using the *GAP* promoter, BAF180 was cloned and expressed in wild-type X33 cells. Growth and lysis conditions were optimized and high-producing strains were discovered. BAF180 was successfully purified from crude lysate using anion exchange chromatography. It was initially believed that, during anion exchange chromatography, two contaminant proteins were eluting with BAF180 (as seen in SDS-PAGE gel), so more purification methods were tried. However, western blot results indicate that the other bands are actually fragments of BAF180. Using an antibody that recognized an epitope on the C-terminus, multiple protein bands were recognized, indicating that the so-called "contaminant" proteins are actually fragments of BAF180. These results suggest that significant amounts of BAF180 are being degraded by proteases during cell lysis. A higher concentration of protease inhibitor cocktail will be tried in an effort to increase the yield of full-length BAF180.

In the future, the recombinant BAF180 should be studied using thermodynamic and kinetic methods such as isothermal titration calorimetry and fluorescence anisotropy. All possible binding partners should be explored via use of a peptide array featuring post-translationally modified histone peptides. The results will indicate whether neighboring PTMs impact the binding of BAF180 to acetylation sites. Neighboring PTMs have been known to impact the binding of many proteins, sometimes increasing the affinity and

sometimes decreasing. Kinetic studies (such as ITC) could be performed using the peptides that had the strongest interactions.

It would also be interesting to determine the structural content of full-length BAF180 by circular dichroism (% α-helices, % β-sheet, and % random coil); 8 of the individual domains (not HMG-box) have been cloned and crystallized, but it would be interesting to see if any of the structural characteristics change due to neighboring amino acids. Other proposed experiments include determining the role of the HMG-box in BAF180, confirming proteolytic degradation, determination of gene copy number, and increasing the yield of full-length BAF180 by using a protease-deficient strain, testing other promoters, or extracellular expression. These experiments are briefly discussed in appendix A.

Expression of the individual domains of BAF180 and *in vivo* experiments have provided some insight into the roles of BAF180. However, a significant gap in our knowledge remains between singular domain interactions and cellular impacts of BAF180: the molecular-level interactions of the complete BAF180 protein are unknown. The success of finally expressing and purifying full-length BAF180 in a heterologous host will provide us with the opportunity to study the fine nuances of the protein *in vitro*. The intricate interactions with individual cellular components can now be quantified and fully explored. With this information, small-molecule drugs that target the BAF180 protein can be developed.

Appendix A: Other Proposed Experiments

Exploring the Role of the HMG-box Domain

Referring back to the introductory chapter, the role of BAF180 in heart development was mentioned. While phenomenological consequences of BAF180 knock-out are known, the exact interactions that cause these events are unknown. The HMG-box domain of BAF180 may be involved in proper embryonic development. BAF180 is known to be critical for proper heart development in embryos (see Section 1.4.2.2) and proper expression of various other HMG-box-containing proteins have been demonstrated to be required for normal embryonic development.[1] It would be interesting to determine if the role of BAF180 in heart development is reliant on its HMG-box. BAF180 knockout mice, generated as per Wang et al., could be used to create mice with different constructs (instead of wild-type BAF180) in the embryo and wild-type BAF180 in the placenta.[2] The different constructs would contain different types of domains[3] (i.e. bromodomains, BAH domains, HMG-box domain) and a partial-sequence of BAF180 (all bromodomains, both BAH domains, but no HMG-box). Each embryo would be studied to determine if each of them are able to survive past embryonic day E12.5-E15.5, when BAF180 KO mice died.[2] Gene expression of each of the embryos could also be analyzed to see if the same genes that were up-regulated upon BAF180 KO are also up-regulated in the mutants.[2] The role of BAF180 in maintaining genomic stability and proper DNA damage repair may also stem from its HMG-box. This theory could be tested by generating constructs, like those described above, and transfecting them into BAF180 KO cancer cells. The results of the construct containing just the bromodomains

and the BAH domains (i.e. lacking the HMG-box) could be used to infer if cancer-associated mutations of BAF180, which often lack the HMG-box, result in genomic instability and aneuploidy.

Proteolytic Degradation

Western blot experiments were successfully carried out using the Tris-acetate gel system. When probed with an anti-*PBRM1* antibody that recognized an epitope on the C-terminus of BAF180, multiple bands were seen. The most intense band appears at approximately 190 kDa, the expected molecular weight of BAF180. The other bands were a lower molecular weight than this. It is believed that degradation of the protein is causing these small bands. Further western blot experiments should be performed with an antibody recognizing an epitope on the N-terminus and an antibody recognizing an epitope near the middle of BAF180 to confirm this hypothesis. Four antibodies are available from ThermoFisher which recognize epitopes in these regions (Table A.1). The antibody currently used recognizes an epitope somewhere between amino acids 1329-1689 (full-length BAF180 is 1689 amino acids long). Experiments using these other antibodies will provide additional evidence that BAF180 is being degraded.

Table A.1: *PBRM1* antibodies available from ThermoFisher.

Antibody Catalog Number	Epitope Region
PA5-41591	AA31-80
PA5-27268	AA105-374
730027	AA501-637
730076	AA773-917

Determination of Copy Number via qPCR

As mentioned in chapter 4, a recombinant gene is integrated into the *Pichia* genome via homologous recombination. Because of this, multiple copies of the gene can sometimes be inserted (~1% chance). qPCR should be performed to determine the exact copy number of BAF180 in the high-producing cells. It would also be interesting to test the low-producing cells as well, since multiple gene copies can sometimes result in decreased expression levels.[4-6]

Increasing Yield by Expression in Protease-Deficient Strains

The possible expression of BAF180 in a protease-deficient *Pichia pastoris* strain could also be explored. The protease-deficient strains are not recommended for initial expression attempts as they have some potentially serious problems, such as poor growth and low transformation efficiency.[5-8] However, because so much of the recombinant BAF180 is being degraded, these strains might be worth a try. It should be noted that these strains improve the yield only 10% of the time for secreted proteins. No reports using these to increase intracellular expression levels are available (see Section 4.2).

Increasing Yield by Extracellular Expression

A secretion signal could be added to the N-terminus of BAF180 which would allow for the secretion of the protein into the growth media. Generally it is recommended that expression of a protein in a heterologous host be done in the same manner as in its native host (i.e. if a protein is produced intracellularly in its native host, it should be intracellularly expressed in the heterologous host.

Increasing Yield with the *AOX1* Promoter

Expression of BAF180 with the inducible *AOX1* promoter could also be tried. The *GAP* (used here) and *AOX1* promoters are generally thought of as most efficient. Some proteins are expressed at higher quantities when using the GAP promoter than with the AOX1 promoter, and visa-versa.[5-6,9-16]

References

1. Štros, M.; Launholt, D.; Grasser, K. D., The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cellular and Molecular Life Sciences* **2007,** *64* (19), 2590-2606.

2. Wang, Z.; Zhai, W.; Richardson, J. A., et al., Polybromo protein BAF180 functions in mammalian cardiac chamber maturation. *Genes & development* **2004,** *18* (24), 3106-3116.

3. Niimi, A.; Hopkins, S. R.; Downs, J. A., et al., The BAH domain of BAF180 is required for PCNA ubiquitination. *Mutat Res* **2015,** *779*, 16-23.

4. Aw, R.; Polizzi, K. M., Can too many copies spoil the broth? *Microbial Cell Factories* **2013,** *12* (1), 128.

5. Felber, M.; Pichler, H.; Ruth, C., Strains and Molecular Tools for Recombinant Protein Production in Pichia pastoris. In *Yeast Metabolic Engineering: Methods and Protocols*, Mapelli, V., Ed. Springer New York: New York, NY, 2014; pp 87-111.

6. Lin-Cereghino, J.; Lin-Cereghino, G., Vectors and Strains for Expression. In *Pichia Protocols*, Cregg, J., Ed. Humana Press: 2007; Vol. 389, pp 11-25.

7. Cregg, J., Distinctions Between Pichia pastoris and Other Expression Systems. In *Pichia Protocols*, second ed.; Cregg, J., Ed. 2007; pp 1-10.

8. Cregg, J.; Tolstorukov, I.; Kusari, A., et al., Expression in the Yeast Pichia pastoris. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 169-189.

9. Daly, R.; Hearn, M. T. W., Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition* **2005,** *18* (2), 119-138.

10. Li, P.; Anumanthan, A.; Gao, X. G., et al., Expression of recombinant proteins in Pichia pastoris. *Applied biochemistry and biotechnology* **2007,** *142* (2), 105-24.

11. Brondyk, W., Selecting an Appropriate Method for Expressing a Recombinant Protien. In *Guide to Protein Purification*, second ed.; Burgess, R.; Deutscher, M., Eds. Academic Press: 2009; pp 135-136.

12. Zhang, A.-L.; Luo, J.-X.; Zhang, T.-Y., et al., Recent advances on the GAP promoter derived expression system of Pichia pastoris. *Molecular Biology Reports* **2009,** *36* (6), 1611-1619.

13. Maity, N.; Thawani, A.; Sharma, A., et al., Expression and Control of Codon-Optimized Granulocyte Colony-Stimulating Factor in Pichia pastoris. *Applied biochemistry and biotechnology* **2016,** *178* (1), 159-72.

14. Pepeliaev, S.; Krahulec, J.; Cerny, Z., et al., High level expression of human enteropeptidase light chain in Pichia pastoris. *J Biotechnol* **2011,** *156* (1), 67-75.

15. Hohenblum, H.; Gasser, B.; Maurer, M., et al., Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris. *Biotechnology and bioengineering* **2004,** *85* (4), 367-75.

16. Xu, J.; Wang, L. N.; Zhu, C. H., et al., Co-expression of recombinant human prolyl with human collagen alpha1 (III) chains in two yeast systems. *Letters in applied microbiology* **2015,** *61* (3), 259-66.

Appendix B: Protocols

Preparation of samples for SDS-PAGE on Tris-glycine gels

Protein samples were separated via SDS-PAGE and samples were prepared as follows:

1. Combine in a tube: 20 µL protein sample
   5.4 µL 4x SDS-PAGE Loading Buffer
   1.4 µL betamercaptoethanol
2. Incubate at 95 °C for 10 minutes
3. Centrifuge 13k g for 1 minute
4. Incubate on ice for 5 minutes

Preparation of Samples for SDS-PAGE on Tris-acetate gels

Protein samples for SDS-PAGE with Tris-acetate gels were prepared differently than for Tris-glycine gels. The sample preparation was based on guidelines in the NuPAGE manual.[1]

1. Combine in a tube: 10 µg protein sample
   6.25 µL NuPAGE LDS Sample Buffer
   2.5 µL NuPAGE Reducing Agent
   H_2O to 25 µL
2. Incubate samples for 10 minutes at 70 °C

Creation of *E. coli* Freezer Stock

1. Plate cells on LB with the correct antibiotic and grow overnight at 37 °C.
2. Inoculate a 15 mL LB culture (with correct antibiotic) with one colony. Grow overnight at 37 °C and 250 rpm.
3. Transfer the cultures to 30 mL oakridge tubes and centrifuge at 3000 rpm for 10 minutes at 4 °C.
4. Decant out most of the media, leaving only a little left in the tube (~1-2 mL). Re-suspend cells gently.
5. Add 1 mL of re-suspended cells to cryogenic freezer tubes. Add 1 mL cold 50% glycerol in water. Invert 2-3 times to mix. Store at -80 °C.

Preparation of $CaCl_2$ Chemically Competent DH5α *E. coli* Cells

This procedure is not original but the origin is unknown.

1. Streak LB plates with DH5α cells. Don't use an antibiotic as DH5α does not have any resistance. Grow overnight at 37 °C.
2. Inoculate 5 mL of LB with 1 DH5α colony. Grow overnight at 37 °C and 250 rpm.
3. Transfer 5 mL culture to a fresh 100 mL culture. Incubate ~2 hours.
4. Remove cells from incubator and place on ice for 10 minutes. Keep the cells cold during the rest of the procedure.
5. Centrifuge for 3 minutes at 6000 rpm and 4 °C. Decant media.
6. Gently re-suspend cells in 10 mL cold 0.1 M $CaCl_2$. Incubate on ice for 20 minutes.

7. Centrifuge cells again. Decant supernatant and gently re-suspend cells in 5 mL cold 0.1 M $CalCl_2$/15% glycerol. Aliquot 100 µL into microcentrifuge tubes and store at -80 °C.

Transformation with Chemically Competent *E. coli* cells

The procedure outlined in the pET30 manual was modified slightly for transformation of chemical competent *E. coli* cells.[2]

1. Remove prepared chemical competent cells (aliquots of 20 µL) from the -80 °C freezer. Thaw tubes on ice 2-5 minutes.
2. Add 2 µL ligation reaction directly to cells. Pipet to mix. Incubate the tubes (cells/DNA) on ice for 5 minutes.
3. Heat the tubes at 42 °C for 30 seconds. Remove immediately and place on ice for 2 minutes.
4. Add 80 µL of LB and plate on LB plates with correct antibiotic

Preparation of Electrocompetent *E. coli* Cells

Electrocompetent *E. coli* cells were prepared following the protocol outlined in the "Micropulser Electroporation manual" from Bio-Rad, Inc.[3]

1. Begin a DH5α/Rosetta(DE3) culture in 5 mL LB. Incubate overnight at 37 °C and 250 rpm.
2. Transfer all 5 mL of the DH5α culture into 500 mL fresh LB. Incubate at 37 °C and 250 rpm until cells reach an OD_{600} = 0.5-0.7.
3. Chill cells on ice for approximately 20 minutes.

4. Transfer cells to a cold 500 mL centrifuge bottle and centrifuge at 4000 g for 15 minutes at 4 °C.
5. Decant the media. Resuspend the cell pellet in 500 mL ice-cold 10 % glycerol. Centrifuge at 4000 g for 15 minutes at 4 °C.
6. Decant the media. Resuspend the cell pellet in 250 mL ice-cold 10 % glycerol. Centrifuge at 4000 g for 15 minutes at 4 °C.
7. Decant the media. Resuspend the cell pellet in 20 mL ice-cold 10 % glycerol. Centrifuge at 4000 g for 15 minutes at 4 °C.
8. Decant the media. Resuspend the cell pellet in 500 mL ice-cold 10 % glycerol. Centrifuge at 4000 g for 15 minutes at 4 °C.
9. Resuspend cell pellet with 10 % glycerol to a final volume of approximately 2 mL. The cell density should be 1-3 x 10^{10} cells/mL.
10. Prepare 40 µL aliquots and store at -80 °C.

Native Lysis of *E. coli* Cells

The native lysis procedure was as follows: cells were re-suspended in native lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl, pH 8.0) at 5mL/gram cell and 100 mg/mL lysozyme was added (final concentration: 3 mg/mL). The cell suspension was incubated on ice for 30-45 minutes before being sonicated at 25% amplitude for 80 seconds (10 sec on/10 sec off). The lysed cells were centrifuged for 10 minutes at 10,000 g and the supernatant recovered. The cell pellet/debris was re-suspended in 5 mL native lysis buffer.

Denaturing Lysis of *E. coli* Cells

For the denaturing lysis protocol, the cells were re-suspended in denaturing lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl, 8 M urea, pH 8.0) at 5 mL/gram cell and incubated on ice for ~1 hour. The suspension was centrifuged at 10,000 g for 10 minutes and the supernatant recovered. To the cell pellet/debris, 5 mL of denaturing lysis buffer was added and the cells re-suspended.

Electroporation/Transformation of pGAPZA into *E. coli* DH5α Cells

The parent vector pGAPZA and the plasmid construct pGAPZA-BAF180 were transformed into E. coli DH5α cells for propagation. The procedure was slightly modified from the protocol outlined in the "MicroPulser Electroporation manual" from Bio-Rad, Inc.[3] Note that low-salt LB is used instead of regular LB.

1. Thaw electrocompetent *E. coli* cells on ice. Wrap an electroporation cuvette in aluminum foil and place on ice. Place a microcentrifuge tube on ice as well.
2. Add DNA to 40 µL of electrocompetent cells. Mix and incubate on ice for 1 minute.
    - Note: For transformation of pGAPZA, 1 µL of 10 ng/µL was used
    - Note: For transformation of pGAPZA-BAF180, all 20 µL of the ligation reaction was used
3. Transfer DNA-cell mixture to 0.2 µm electroporation cuvette. Place cuvette in micropulser slide.
4. With micropulser set to "Ec2', pulse the electroporation cuvette.

5. Immediately add 1 mL of low-salt LB to the cuvette. Pipet up and down with a PAGE gel loading tip.
6. Transfer the cell suspension to a 15 mL tube and incubate for 1 hour at 37 °C and 225 rpm.
7. Pour the cell suspension onto pre-warmed low-salt LB plates with 25 µg/mL zeocin and incubate overnight at 37 °C. (Note: Zeocin is inactive at high salt concentrations, hence the use of low-salt LB).

Linearization of pGAPZA-Pb1 Construct Using AvrII

1. Combine 1 µL of AvrII (5 units) per 1 µg DNA (WHICH?). In total, digested 12 µg of pGAPZA-Pb1 construct. Performed 3 – 4 µg digested because the DNA recovery limit with the Zymo kit in 5 µg.

DNA	X µL (4000ng)
CutSmart	5 µL
AvrII	4 µL
H_2O	to 50 µL

2. Incubated at 37 °C for 16 hours, then purified the DNA digest using the Zymo DNA Clean and Concentrator kit. Recovered 7380 ng DNA.
3. Used the vacuum centrifuge (Savant Refrigerator Condensation Trap and Savant Speed Vac Concentrator) to concentrate the recovered DNA.
4. Decided to do another AvrII digest, this time incubated at 37 °C for 2 hours. The mixture was the same as for the 16 hour digest.
5. Combined both purified digests to final volume of ~8 µL.

Preparation of Electrocompetent *Pichia pastoris* X33 Cells

Electrocompetent X33 cells (*Pichia pastoris*) were prepared as per protocol found in the pGAPZA manual from Invitrogen.[4]

1. Begin 5 mL culture of X33 in YPD; incubate at 30 °C and 225 rpm overnight
2. Prepare 500 mL of fresh YPD in a 2 L flask. Inoculate this culture with 200 µL

Alternatively:

1. Begin 5 mL culture of X33 in YPD; incubate 8 hours and 30 °C and 225 rpm
2. Prepare 500 mL of fresh YPD in a 2 L flask. Inoculate this culture with all 5 mL.

Both methods:

3. Centrifuge the cells for 5 minutes at 1500 g and 4 °C. Decant media and re-suspend cells in 500 mL ice-cold sterile distilled deionized water.
4. Centrifuge the cells for 5 minutes at 1500 g and 4 °C. Decant media and re-suspend cells in 250 mL ice-cold sterile distilled deionized water.
5. Centrifuge the cells for 5 minutes at 1500 g and 4 °C. Decant media and re-suspend cells in 20 mL ice-cold sterile 1 M sorbitol.
6. Centrifuge the cells for 5 minutes at 1500 g and 4 °C. Decant media and re-suspend cells in 1 mL ice-cold sterile 1 M sorbitol. Approximate volume of sorbitol cell suspension should be 1.5 mL.
7. Keep the cells on ice and use that day. Do not freeze (transformation efficiency greatly decreases with frozen prepped cells).

Electroporation/Transformation of *Pichia pastoris* Cells

A micropulser apparatus from Bio-Rad, Inc. was for the transformation of pGAPZA and pGAPZA-BAF180 into *Pichia pastoris* cells. The following procedure, adapted from the "MicroPulser Electroporation manual" was used.[5]

1. Combine 80 µL of electrocompetent X33 cells and 10 µg of linearized pGAPZA-Pb1 construct in a 1.5 mL microfuge tube. Transfer the mixture to an ice-cold 0.2 cm electroporation cuvette using a gel loading tip.
2. Incubate the cuvette on ice for 5 minutes
3. Pulse the cells with the Bio-Rad Micropulser under "Pic" setting. The pulse for this transformation was 2.00 kV and 5.00 ms.
4. Immediately add 1 mL ice-cold 1 M sorbitol to the cuvette. Pipet gently with a gel loading tip to mix. Transfer the mixture to a cold, sterile 15 mL tube.
5. Incubate the tube at 30 °C without shaking for 2 hours.
6. Spread 5 µL, 10 µL, 15 µL, 25 µL, 50 µL, 75 µL, 100 µL, 125 µL, 150 µL, 175 µL and 200 µL onto YPDS (__% yeast extract, __% peptone, __% dextrose, __% sorbitol) plates with 100 µg/mL zeocin. Grow at 30 °C for 3 days. (Note: Sorbitol was added to stabilize the transformants).

Purification of Zeocin Resistant Transformants and Selection of Multicopy Integrants

Transformants resulting from the successful transformation of pGAPZA-Pb1 into X33 cells needed to be purified to ensure that there were no mixed colony transformants. The mixed colony transformants have a tendency to lose the foreign DNA.

1. After successful transformation, streaked the 6 colonies onto 25 µg/mL zeocin plates. Grew ~42 hours.
2. Remove colonies from 25 µg/mL plates and streak on 100 µg/mL plates. Grow for 3 days.
3. Create freezer stocks of the clones that grew on 100 µg/mL. Start 5 mL YPD overnight cultures and prep for storage at -80 °C.

Isolation of High-Producer Strains

The high-producer strains were only identified after lysis using SDS-PAGE visualization. Since the colonies had already been lysed, creation of new colonies was necessary to form a stock of high-producer cells.

1. Take 200 µL of the soluble lysate and add to 3 mL YPD. Grow at 30 °C for 20-24 hours.
2. Pour culture onto YPD + zeocin plates. 4 plates were made, each with a different amount of starter culture added: 150 µL, 200 µL, 250 µL, and 500 µL. Cells grow better on zeocin plates agar when there is lower cell density.[4]
3. Grow for 3 days at 30 °C.
4. Make freezer stock using the protocol described in this chapter

Plating for Single Colonies

1. Streak *Pichia pastoris* clone on YPD agar + 100 µg/mL zeocin. Grow 3 days at 30 °C.
2. Pick up one colony from this plate using a toothpick.

3. In a 1.5 mL epi tube, add 200 µL YPD. Gently swirl the toothpick in the media to dislodge the colony. The purpose of this step is to dilute the number of cells that get plated in the next step.
4. Using a new toothpick, streak on YPD agar + 100 µg/mL zeocin. Grow 3 days at 30 °C.

Creation of Freezer Stock from *Pichia pastoris* Cells

After successfully plating for single colonies (see notes), freezer stocks needed to be made. Single colonies from the same plate were grow up in a series of 5, 7, and 10 mL cultures. They were grown overnight at 30 °C and 225 rpm. The three different volumes were grown to ensure that at least one of the cultures would be in exponential mode. Colonies from 5 separate plates (therefore 5 separate parent strains) were grown in this manner. The following day, 800 µL of culture (do not spin down) and 200 µL of 100 % glycerol, sterilized were combined in a 2 mL cryogenic vial and flash frozen in liquid nitrogen and immediately transferred to storage at -80 °C.

Isolation of genomic DNA from *Pichia pastoris*

Lysis of Yeast Cells

Reagent needed: Lyticase

1. The desired *P*. pastoris strain was streaked onto YPD plates and grown for 2-3 days at 30 °C.
2. 25 U of lyticase solution was added to a PCR tube and one colony, approximately 2 mm, was suspended in the solution.

3. The samples were incubated in the PCR machine for 30 minutes at 30 °C.
4. The sample was frozen at -80 °C for 15 minutes, then thawed at room temperature, and vortexed on medium speed to release the intracellular contents.
5. This freeze, thaw, and vortex step was repeated a second time.

Phenol-Chloroform Extraction

Reagents needed: Phenol:Chloroform:Isoamyl alcohol (P:C:I, 25:24:1)

Chloroform:Isoamyl alcohol (C:I, 24:1)

TE Buffer: 50 mM Tris-HCl, 0.5 mM EDTA, pH 7.5

1. The lysed cell suspension was diluted in 250 µL TE buffer and 250 µL P:C:I was added.
2. The solution was vortexed gently then centrifuged at 17,000 g for 2 minutes at 4 °C.
3. The top layer (aqueous phase, DNA containing) was carefully transferred to a clean tube without disturbing the interface. The volume was approximately 250 µL.
4. To the original tube, 250 µL TE buffer was added.
5. Steps 2 and 3 were repeated and the water phases were combined.
6. To the tube with the aqueous phases, 250 µL C:I was added. This removed the phenol layer from the remaining aqueous solution.
7. The solution was vortexed gently and centrifuged at 17,000 g for 2 minutes at 4 °C.
8. Any remaining aqueous solution was removed.

Ethanol Precipitation

Reagents needed: 3 M sodium acetate, pH 5.2

100 % Ethanol, ice cold

70 % Ethanol

Sterile ddiH_2O

1. In a 1.5 mL microcentrifuge tube, 400 µL of the aqueous solution from above (DNA-containing solution) was combined with 40 µL 3M sodium acetate, pH 5.2 (final concentration: 0.3 M) and 880 µL 100% ice-cold ethanol. This solution was vortexed.
2. The solution from step 1 above was incubated at -80 °C for 45 minutes and then centrifuged at 40,000 g for 15 minutes at 4 °C.
3. The DNA pellet was very small, indistinguishable to the naked eye. The supernatant that contained salts and organic contaminants was removed, leaving a little (~10-30 µL) remaining.
4. The DNA pellet was resuspended in 1 mL 70 % ethanol and then centrifuged at 40,000 g for 15 minutes at 4 °C.
5. The supernatant was carefully removed, again leaving a small amount in the tube. In this step, a faint opaqueness was noticed on the bottom of the tube.
6. The DNA pellet was dried at 40 °C for approximately 30 minutes. There was still some supernatant left.
7. Sterile ddiH_2O, enough to bring the volume to approximately 50 µL, was added and the DNA pellet was resuspended.
8. The DNA concentration and purity was determined using the Nanodrop 1000 spectrophotometer (ThermoFisher).

Optimization of Growth Time

1. Grow up 4 separate 5 (or 7, or 10) mL YPD cultures at 30 °C and 225 rpm overnight. The 4 cultures were all started from the same parent strain (i.e. single colony on same plate).
2. Transfer a 2 mL aliquot of the 5 (or 7, or 10) mL culture to fresh 50 mL YPD (in 125 mL flask). Grow at 20.5 °C and 225 rpm over four days. There will be a total of 4 cultures.
3. Combine the remaining culture (from each of the 4 separate) into one tube and spin down at 12,000 g for 15 min at 4 °C.
4. Every 24 hours, remove one of the cultures and spin it down. Decant the supernatant into a separate tube and freeze both at -20 °C. There will be a time point sample for 0 hours, 24, 48, 72 and 96 hours.
5. Once all the sample are acquired, lyse the yeast cells with the following procedure:

Lysis buffer:

- Tris-HCl 50 mM, pH 7.4
- PMSF 10 mM

Add 2 mL lysis buffer/gram cell pellet.

Aliquots of 1 mL of re-suspended cell pellet were added to 0.5 mm zirconia/silica disruption beads (0.5 mL) and lysed with vortex for 30 sec by vigorous mixing followed by 30 sec incubation on ice. This mix/freeze cycle was repeated a total of 6 times.

Samples were collected by centrifugation at 12,000 g for 15 min at 4 °C and the supernatant was removed for further analysis on an SDS-PAGE gel.

Lysis of *Pichia pastoris* Cells – Procedure

Breaking (Lysis) buffer: 50 mM NaH_2PO_4, pH 7.4

1 mM EDTA

5% glycerol

1 mM PMSF (add immediately prior to use)

Cells were re-suspended in breaking buffer at 2 mL breaking buffer/gram cell. This suspension (wash to remove any residual media) was then centrifuged 3,000 g for 5 minutes at 4 °C. The supernatant was decanted and breaking buffer was added to re-suspend cells (2 mL breaking buffer/gram cell). In a 1.5 mL tube, 1 mL cell suspension was combined with 250 µL (measured by displacement) zirconia/silica disruption beads (0.5 mm). The samples were vortexed on maximum speed for 30 seconds and then rested on ice for 30 sec. This vortex/rest cycle was repeated a total of 12 times. The lysis mixture was centrifuged at 12,000 g for 15 minutes at 4 °C. The lysate (soluble fraction) was transferred to a new 1.5 mL tube. To the tube with cells/beads, 1 mL 6 M urea was added to extract any insoluble proteins. A 200 µL pipet tip was used to loosen the cell debris from the wall of the tube and then the tube was vortexed to re-suspend completely. Both the lysate and the urea/cells/beads tubes were centrifuged at 12,000 g for 15 minutes at 4 °C. The lysate was transferred to a third tube and the urea extraction was transferred to a different tube. Both tubes were centrifuged at 20,000 g for 15 minutes at 4 °C to

clear them. The cleared lysate and urea extraction were transferred to different tubes for storage.

Aliquots of 1 mL of re-suspended cell pellet were added to 0.5 mm zirconia/silica disruption beads (0.5 mL) and lysed with vortex at max speed for 30 sec followed by 30 sec incubation on ice. This mix/freeze cycle was repeated a total of 6 times. Samples were collected by centrifugation at 12,000 g for 15 min at 4 °C.

Dot Blot

These experiments were performed with 0.45 µm nitrocellulose membrane and Pierce Fast Western Kit, ECL substrate.

1. Spot 2 µL of protein sample onto 0.45 µm nitrocellulose membrane. Allow to dry completely. Hydrate the membrane in 1x wash buffer for 5 minutes.
2. Prepare primary antibody solution by combining 10 mL of antibody diluent with 10 µL of anti-BAF180. Incubate membrane in solution for 1 hour at room temperature with gentle shaking.
3. Remove membrane from antibody solution and wash for 5 minutes with 1x wash buffer.
4. Prepare secondary antibody solution by combining 10 mL of antibody diluent with 1 mL Optimized HRP Reagent. Incubate membrane in solution for 10 minutes.
5. Remove membrane from antibody solution and perform 3 consecutive washes, 5 minutes each, with 1x wash buffer

6. Prepare ECL detection reagent by combining 5 mL ECL detection reagent 1 and 5 mL ECL detection reagent 2. Incubate membrane briefly (5 minutes max).
7. Take images using a CDD camera

Western Blot Transfer

1. Transfer SDS gel from gel box to 2x NuPAGE transfer buffer with 0.2% SDS. Equilibrate for 10 minutes
2. Cut a notch on corner of membrane to indicate direction
3. Active the PVDF membrane by submerging it in 100% methanol for 30 seconds. Briefly wash the membrane in water to decrease the alcohol concentration, then place in 1x NuPAGE transfer buffer with 0.1% SDS
4. Place blotting paper and sponges in 1x NuPAGE transfer buffer with 0.1% SDS
5. Assemble the cassette (see) [6-7]

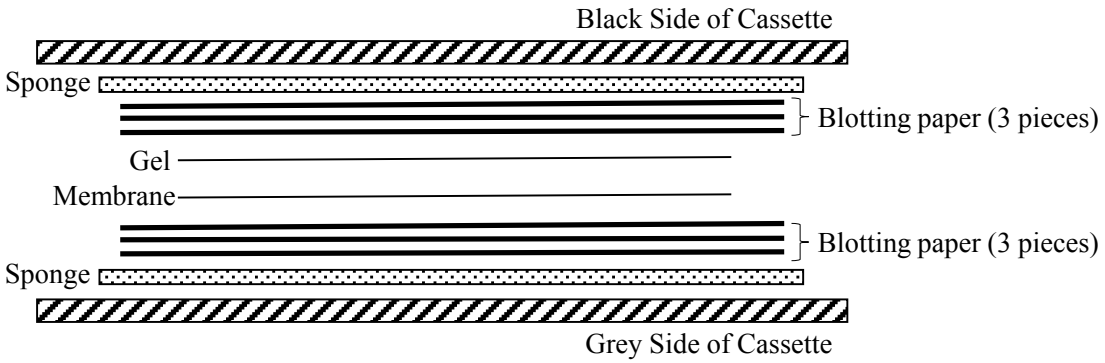


Figure B.1: Assembly of a western blotting "sandwich".

6. Place cassette in tank with the grey side facing the anode (black).[7-8]
7. Run at 70 V for 1 hour

References

1. Invitrogen, NuPAGE® Technical Guide. 2010.

2. Novagen®, pET System Manual. 11th ed.

3. Bio-Rad Inc., High Efficiency Electrotransformation of *E. coli*. In *MicroPulser Electroporation Apparatus Operating Instructions and Applications Guide*, Rev B ed.; p 12.

4. Invitrogen. pGAPZ A, B, and C; pGAPZa A, B, and C. *Pichia* expression vectors for constitutive expression and purification of recombinant proteins 2010, p. 44.

5. Bio-Rad Inc., Electroporation of *Pichia pastoris*. In *MicroPulser Electroporation Apparatus Operating Instructions and Applications Guide*, Rev B ed.; pp 19-20.

6. Scientific, T., User Guide: Nitrocellulose Transfer Membranes. 2011.

7. Hoefer TE22, Hoefer TE22 User Manual. 2012.

8. Abcam, General western blot protocol.

Appendix C: PCR Programs and Primers

Testing for Presence of Individual Domains

Table C.1: PCR conditions for BAF180 KpnI domain test. Testing Bromodomain 1 (D1) → D3; D3 → D5, D3 → D6.

Program Name	Domains	For Use with Primers	Product Size		Temp (°C)	Time
BAF180 KpnI domain test	D1 -> D3	92, 97	1323	Initial Denaturation:	95	30 sec
	D3 -> D5	96, 101	1138	# Cycles: 30		
	D3 -> D6	96, 103		Denat:	95	30 sec
				Anneal:	59	30 sec
				Elong:	68	1 min 20 sec
				Final Elongation	68	5 min
				Hold	4	

Table C.2: PCR conditions for BAF180 domain test 1. Testing start codon to end of bromodomain domain 3 (D3), D1 → D3, D3 → D6, and D6 → stop codon.

Program Name	Domains	For Use with Primers	Product Size		Temp (°C)	Time
BAF180 Domain Test 1	Start -> D3	322, 97	1452	Initial Denaturation:	95	30 sec
	D1 -> D3	92, 97	1323	# Cycles: 30		
	D3 -> D6	251, 103	1500	Denat:	95	30 sec
	D6 -> Stop	102, 47	2583	Anneal:	48	20 sec
				Elong:	68	90 sec
				Final Elongation	68	5 min
				Hold	4	

Table C.3: PCR conditions for BAF180 domain test 2. Testing from start codon to end of D2, D1 → D2, D2 → D3, D3 → D4, D4 → D5, D5 → D6.

Program Name	Domains	For Use with Primers	Product Size		Temp (°C)	Time
BAF180 Domain Test 2	Start -> D2	322, 295	852	Initial Denaturation:	95	30 sec
	D1 -> D2	92, 95	723	# Cycles: 30		
	D2 -> D3	94, 97	909	Denat:	95	20 sec
	D3 -> D4	138, 99	693	Anneal:	49	15 sec
	D4 -> D5	98, 101	732	Elong:	68	45 sec
	D5 -> D6	253, 254	747	Final Elongation	68	5 min
				Hold	4	

Table C.4: PCR conditions for BAF180 domain test 3. Testing for presence of D1 → D3 and D3 → D6.

Program Name	Domains	For Use with Primers	Product Size		Temp (°C)	Time
BAF180 Domain Test 3	D1->D3	92, 97	1323	Initial Denaturation:	95	30 sec
	D3->D6	96, 103	1500	# Cycles: 30		
				Denat:	95	20 sec
				Anneal:	50-60	15 sec
				Elong:	68	1 min 40 sec
				Final Elongation	68	5 min
				Hold	4	

PCR Program for Modification of BAF180 Gene Sequence (insertion of Kozak consensus sequence, deletion of stop codon, insertion of enterokinase site, 3' KpnI site)

Table C.5: PCR conditions for modification of BAF180 gene sequence: insertion of Kozak consensus sequence, deletion of stop codon, insertion of enterokinase site, and modification of 3' restriction enzyme site to KpnI.

Program Name	For Use with Primers	Product Size		Temp (°C)	Time
SH 328/329	328/329	4985	Initial Denaturation:	98	30 sec
			# Cycles: 30		
			Denat:	98	10 sec
			Anneal:	76.3	30 sec
			Elong:	72	2 min
			Final Elongation	72	2 min
			Hold	4	

Table C.6: BAF180 primer sequences.

Archive Number	Sequence	Name
092	GGTGGTCATATGCCAAGCAGGAAAAGGAGGA	BAF180 Bromo1 - forward
093	GCTGCTGGTACCATTGTCTTGCCCATCTTC	BAF180 Bromo1 - reverse
094	GGTGGTCATATGGATGGGCAAGACAATCAGG	BAF180 Bromo2 - forward
095	GCTGCCGGTACCTGAGCCATATTGAAGTGCCAT	BAF180 Bromo2 - reverse
096	GGTGGTCATATGGCACTTCAATATGGCTCA	BAF180 Bromo3 - forward
097	GCTGCTGGTACCGGTGGCTGAAGAGATCATGC	BAF180 Bromo3 - reverse
098	GGTGGTCATATGAGCATGATCTCTTCAGCCAC	BAF180 Bromo4 - forward
099	TCCTCCGGTACCGAGTTTGGGAGAAGCCATGTC	BAF180 Bromo4 - reverse
100	GGTGGTCATATGGCTTCTCCCAAACTCAA	BAF180 Bromo5 - forward
101	GCTGCTGGTACCCTCTTGAATCAGCAAAGTCACATT	BAF180 Bromo5 - reverse
102	GGTGGTTCTAGA*ATG*CCAAATGTGACTTTGCTGATTC	BAF180 Bromo6 - forward
103	GCTGCTGGTACCTGTGGTATAGCTGAGTGCCG	BAF180 Bromo6 - reverse
136	GGTGGTATTAATATGCCAAGCAGGAAAAGGAGGA	BAF180-bd1 fwd (Ase 1)
137	GGTGGTATTAATATGGATGGGCAAGACAATCAGG	BAF180-bd2 fwd (Ase 1)
138	GGTGGTATTAATATGGCACTTCAATATGGCTCA	BAF180-bd3 fwd (Ase 1)
139	GGTGGTATTAATATGAGCATGATCTCTTCAGCCAC	BAF180-bd4 fwd (Ase 1)
140	GGTGGTATTAATATGGCTTCTCCCAAACTCAA	BAF180-bd5 fwd (Ase 1)
141	GGTGGTATTAATATGCCAAATGTGACTTTGCTGATTC	BAF180-bd6 fwd (Ase 1)
142	GGTGGTAAGCTTCCAAATGTGACTTTGCTGATTC	BAF180-bd6 fwd (hind3)
143	CGTGCTGCGGCCGCAGTTGTGGTATAGCTGAGTGCCG	BAF180-bd6 rev (Not1)
146	GGTGGTATTAAT ATGTCCATGGGTTCCA	BAF180 fwd (Ase1)
147	GCTGCTGGTACCAACATTTTCTAGGTTGTATGC	BAF180 rev (Kpn1)

249	GGTGGTGCTCTTCCAACCCAAGCAGGAAAAGGAGGA	BAF180 Bromo1 - forward
250	GGTGGTGCTCTTCCAACGATGGGCAAGACAATCAGG	BAF180 Bromo2 - forward
251	GGTGGTGCTCTTCCAACGCACTTCAATATGGCTCA	BAF180 Bromo3 - forward
252	GGTGGTGCTCTTCCAACAGCATGATCTCTTCAGCCAC	BAF180 Bromo4 - forward
253	GGTGGTGCTCTTCCAACGCTTCTCCCAAACTCAA	BAF180 Bromo5 - forward
254	GGTGGTCTGCAGTTATGTGGTATAGCTGAGTGCCG	BAF180 Bromo6 - reverse
255	GGTGGTGGTCATATGCCCAAGCAGGAAAAGGAGGA	BAF180 Bromo1 - forward
256	GGTGGTGGTCATATGCGATGGGCAAGACAATCAGG	BAF180 Bromo2 - forward
257	GGTGGTGGTCATATGCGCACTTCAATATGGCTCA	BAF180 Bromo3 - forward
258	GGTGGTGGTCATATGCAGCATGATCTCTTCAGCCAC	BAF180 Bromo4 - forward
259	GGTGGTGGTCATATGCGCTTCTCCCAAACTCAA	BAF180 Bromo5 - forward
308	GGCGAAGCTT TTCCATGGGTTCC	BAF180 fwd_outer (whole protein)
309	CGCGCCGGATCC GTGATTAAAC	BAF180 rev_outer (whole protein)
322	CGGGTACC ATGGGTTCCAAGAGAAGA	BAF180, full length, forward
323	CCGGCGAATTC CTGATCCACAACTCTTTATG	BAF180, full length, reverse
324	GGCGCC GTCCCTATTTCAATCAATTGAA	5' GAP
328	gcgctggtacgtagGAATTCGAACTCAAAatgggttccaagagaagaag	Modifications of BAF180 gene sequence, fwd
329	GCGCTAATGATGTGAGGTACCCTTGTCGTCGTCGTCAACATTTTCTAGGTTGTA	Modifications of BAF180 gene sequence, rev
336	GCG CGT TGC ATA ATG ATG TGG AGA AAG	BAH1 forward
337	GCG CCT GGG CAT AAC TTA AAG TAT TCC TT	BAH1 reverse
338	GCG CGA GCA GCT CCA TTA CAA T	BAH2 forward
339	GGC CCT CCT TCT GAG GAA CAA TTG	BAH2 reverse
340	GCG CAG TGC AAA GAA GGA AGG	HMG forward
341	GCG CCC TTG GTT CAT CAC ACC	HMG reverse

Copyright Permissions

Regarding Figure 1.5:

Regarding Figure 1.13:

The sketch of the outline of the human body is in the public domain. All notations are original. [http://diagrampic.com/printable-human-body-outline/]

Regarding Crystal Structures:

All crystal structures were found on Protein Data Bank (PDB).[1-2] As such, they are in the public domain. Images were manipulated using Protein Workshop.[3]

1. Berman, H. M.; Westbrook, J.; Feng, Z., et al., The Protein Data Bank. *Nucleic acids research* **2000,** *28* (1), 235-242.

2. Rose, P. W.; Prlić, A.; Altunkaya, A., et al., The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic acids research* **2017,** *45* (D1), D271-D281.

3. Moreland, J. L.; Gramada, A.; Buzko, O. V., et al., The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **2005,** *6* (1), 21.